Field Theory of
Optical Guided Waves

Jerome K Butler
Southern Methodist University
Dallas, TX 75275
Contents

Chapter 1 Fields, Power and Energy .............................................. 1
  1.1 Introduction ...................................................................... 1
  1.2 Harmonically Time-Varying Signals ........................................ 2
  1.3 Spectral Densities .............................................................. 4
  1.4 Maxwell’s Equations ............................................................ 5
  1.5 Complex Poynting Vector ....................................................... 7
  1.6 Boundary Conditions .......................................................... 7
  1.7 Complex Dielectric Constant .................................................. 8
  1.8 General Constitutive Relations ............................................... 10
  1.9 Narrow-Band Processes ....................................................... 14
  1.10 Wave Equation .................................................................. 18
  1.11 Field Analysis with Potential Functions ................................. 23
    1.11.1 Magnetic Vector Potential ............................................. 24
    1.11.2 Electric Vector Potential ............................................... 29
    1.11.3 Electromagnetic Fields in Waveguides ............................... 29
Chapter 2 Static and Quasi-Static Fields ....................................... 31
  2.1 Introduction ...................................................................... 31
  2.2 Embedded Dielectric Sphere ................................................. 32
  2.3 Arbitrary Polarization Direction ............................................ 35
  2.4 Isotropic Spheres in an Isotropic Medium ............................. 36
  2.5 Anisotropic Spheres in an Isotropic Medium ......................... 38
  2.6 Absorption in Embedded Nanospheres .................................. 45
Chapter 3 Plane Waves ................................................................. 47
  3.1 Introduction ...................................................................... 47
  3.2 Formalism ........................................................................ 47
  3.3 Phase and Group Velocities .................................................... 50
  3.4 Polarization ....................................................................... 53
  3.5 Reflection and Transmission at Plane Boundaries .................. 54
  3.6 Stratified Media .................................................................. 58
    3.6.1 Single-Layer Transfer Matrix ....................................... 58
    3.6.2 Single-Layer Fields ....................................................... 62
    3.6.3 Transfer Characteristics of Double Layers ...................... 62
    3.6.4 Double-Layer Dispersion .............................................. 67
    3.6.5 Double-Layer Fields ..................................................... 68
  3.7 Periodic Stacks .................................................................. 68
    3.7.1 Phase and Group Velocities ........................................... 70
    3.7.2 Reflection and Transmission ........................................... 70
  3.8 Field Distributions for Quarter-Wavelength Layers ................ 71
  3.9 Matrix Formulation of Scattering ......................................... 73
Chapter 4 Waves in Layered Waveguides ...................................... 79
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Waveguide Equations</td>
<td>79</td>
</tr>
<tr>
<td>4.2</td>
<td>Wave Definitions</td>
<td>81</td>
</tr>
<tr>
<td>4.3</td>
<td>Slab Waveguides</td>
<td>82</td>
</tr>
<tr>
<td>4.4</td>
<td>Transverse Electric Modes: Formulation</td>
<td>84</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Case I: Bounded Modes</td>
<td>86</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Case II: Radiation Modes</td>
<td>92</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Case III: Radiation Modes</td>
<td>97</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Case IV: Evanescent Radiation Modes</td>
<td>102</td>
</tr>
<tr>
<td>4.5</td>
<td>Mode Completeness and Orthogonally</td>
<td>102</td>
</tr>
<tr>
<td>4.6</td>
<td>Spectral Power Flux</td>
<td>107</td>
</tr>
<tr>
<td>4.7</td>
<td>Generalized Notation</td>
<td>108</td>
</tr>
<tr>
<td><strong>Chapter 5</strong></td>
<td>Radiation Modes from Plane Waves</td>
<td>111</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>111</td>
</tr>
<tr>
<td>5.2</td>
<td>Construction of Case A Radiation Modes</td>
<td>112</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Properties of Case A Fields</td>
<td>113</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Orthonormalization of Case A Radiation Modes</td>
<td>117</td>
</tr>
<tr>
<td>5.3</td>
<td>Construction of Case B Radiation Modes</td>
<td>119</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Properties of Case B Fields</td>
<td>120</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Orthonormalization of Case B Radiation Modes</td>
<td>122</td>
</tr>
<tr>
<td>5.4</td>
<td>Degeneracy of A and B Modes Sets</td>
<td>123</td>
</tr>
<tr>
<td><strong>Chapter 6</strong></td>
<td>Transitions in Layered Waveguides</td>
<td>125</td>
</tr>
<tr>
<td>6.1</td>
<td>Waveguide Geometries</td>
<td>125</td>
</tr>
<tr>
<td>6.2</td>
<td>Single Discontinuity</td>
<td>126</td>
</tr>
<tr>
<td>6.3</td>
<td>Scattering Formulation</td>
<td>127</td>
</tr>
<tr>
<td>6.4</td>
<td>Single-Mode Approximation</td>
<td>128</td>
</tr>
<tr>
<td>6.5</td>
<td>Generalized Cost Function</td>
<td>131</td>
</tr>
<tr>
<td><strong>Chapter 7</strong></td>
<td>Modes in Complex Structures</td>
<td>137</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>137</td>
</tr>
<tr>
<td>7.2</td>
<td>Fundamental Layers</td>
<td>138</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Homogeneous Layers</td>
<td>138</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Graded Layers</td>
<td>140</td>
</tr>
<tr>
<td>7.3</td>
<td>Superlayers</td>
<td>144</td>
</tr>
<tr>
<td><strong>Chapter 8</strong></td>
<td>Waveguides Characteristics near the 2nd Bragg Condition</td>
<td>153</td>
</tr>
<tr>
<td>8.1</td>
<td>Grating Layer</td>
<td>154</td>
</tr>
<tr>
<td>8.2</td>
<td>Floquet-Bloch Waves</td>
<td>156</td>
</tr>
<tr>
<td>8.3</td>
<td>Grated/Non-Grated Waveguide Interfacing</td>
<td>173</td>
</tr>
<tr>
<td>8.4</td>
<td>Properties Grating Etched in a Laser Waveguide</td>
<td>177</td>
</tr>
<tr>
<td><strong>Chapter 9</strong></td>
<td>Numerical Solution of Layered Structures</td>
<td>185</td>
</tr>
<tr>
<td>9.1</td>
<td>Branch Cuts</td>
<td>185</td>
</tr>
<tr>
<td>9.2</td>
<td>Mode Solutions</td>
<td>186</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Method of Stationary Phase</td>
<td>191</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Mode Orthonormalization</td>
<td>195</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Overlap Integrals</td>
<td>197</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Tooth Geometries for Boundary Element Methods</td>
<td>203</td>
</tr>
<tr>
<td>Appendix E</td>
<td>Laplace Transform Table</td>
<td>207</td>
</tr>
</tbody>
</table>
Chapter 1
Fields, Power and Energy

This fundamental behavior of classical electromagnetic fields is governed by Maxwell's equations. This chapter focuses on the fundamental characteristics of electromagnetic fields derived from Maxwell's equations. The classical description of the fields is applicable to contemporary optical fields and to optical devices that are governed by field theory as well as from quantum mechanics. Most of the devices discussed here are designed to operate at optical wavelengths which covers a broad range of frequencies.

1.1 Introduction
The electromagnetic field is characterized by the electric and magnetic field strengths, and in the most general case these fields are a function of both position and time,

\[
e = e(x, y, z; t), \quad (1.1.1)
\]
\[
h = h(x, y, z; t). \quad (1.1.2)
\]

The other field quantities appearing in Maxwell’s equations include the electric, \( \mathbf{d} \), and magnetic, \( \mathbf{b} \) flux densities. Typically, for static or slowly time varying fields (radio frequencies), the electric flux density, \( \mathbf{d} \) is connected to the electric field in a nonlinear fashion such as \( \mathbf{d} = g(e) \), where the function \( g \) depends on the material. However, for most practical photonic devices, the fields strengths are relatively small, so that the relationship is usually linear.
Harmonically time-varying fields are useful because mathematical tools such as Fourier analysis can be applied to an arbitrary time-varying signal. The Fourier transform and its inverse, are defined as

\[ \mathcal{F} \{ f(t) \} = F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} \, dt, \quad (1.1.3) \]

\[ \mathcal{F}^{-1} \{ F(\omega) \} = f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} \, d\omega, \quad (1.1.4) \]

so that \( f(t) \) and \( F(\omega) \) are designated as Fourier transform pair \( f(t) \Leftrightarrow F(\omega) \).

Without loss of generality the harmonic time varying signals can be written as

\[ e(x, y, z; t) = \Re[\mathcal{E}(x, y, z; \omega)e^{j\omega t}] \]

\[ = \frac{1}{2}[\mathcal{E}(x, y, z; \omega)e^{j\omega t} + \mathcal{E}^*(x, y, z; \omega)e^{-j\omega t}], \quad (1.1.5) \]

where \( \mathcal{E}^* \) is the complex conjugate of \( \mathcal{E} \). It is useful to note that the field \( e(x, y, z; t) \) must be a real physical quantity as is dictated by the above expression. Typically, the first expression is more convenient to use because the \( \Re \) designator can be dropped or hidden in the expression. The actual expression for field quantities can be obtained taking the real part of the expression, \textit{i.e.} \( \Re \) becomes visible again.

### 1.2 Harmonically Time-Varying Signals

The complex notation as illustrated here has been very useful in the analysis of systems, however, some caution must be exhibited when multiplying two complex field quantities, for example, when the electromagnetic field power or energy is computed. Figure 1.1 illustrates the connection of time varying current source to a tuned circuit, parallel, \( R, L \) and \( C \).

![Figure 1.1. An arbitrary time-varying current source feeding a tuned circuit](image)

The input current source has an instantaneous time variation, \( i(t) \), and must equal to the sum of the instantaneous currents \( i_R(t) \), \( i_L(t) \), and \( i_C(t) \) in each of the respective elements \( R, L, \) and \( C \). The instantaneous power delivered to the tuned load is \( p(t) = v(t) \cdot i(t) \). If the current source has a time dependence of \( i(t) = I_0 \cos(\omega t) \) and voltage across the load is \( v(t) = V_0 \cos(\omega t) + \)
1.2 HARMONICALLY TIME-VARYING SIGNALS

The instantaneous power is easily evaluated. The time average of the instantaneous power, representing the “power drain” from the current source, is

\[
\langle p(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} p(t) \, dt = \frac{1}{2} I_0 V_0 \cos \phi. \tag{1.2.1}
\]

It should be noted that the concept of average power is useful only for periodic time signals. A signal that exhibits only a single pulse, in time, has zero average power. Specifically, signals with a spectral density that does not include impulses at specific frequencies have zero average power, but the signals can have an average energy. Thus, signals can be divided into “power” or “energy” signals.

Instead of using (1.2.1) to calculate the time average power, manipulations of complex expressions of the current and voltage can be performed to also yield average powers. In particular, using the complex notation, \( I(\omega) = I_0 \exp j\omega t \) and \( V(\omega) = V_0 \exp j(\omega t + \phi) \). The expression, \( V(\omega)I^*(\omega) \), useful for finding the average power drain from the current source, is

\[
V(\omega)I^*(\omega) = I_0 V_0 \cos \phi + jI_0 V_0 \sin \phi. \tag{1.2.2}
\]

Note that

\[
\langle p(t) \rangle = \frac{1}{2} \text{Re}[V(\omega)I^*(\omega)]. \tag{1.2.3}
\]

It is instructive to determine additional information that can be obtained from \( VI^* \). The current to the load is

\[
I(\omega) = \frac{V(\omega)}{R} + I_L(\omega) + j\omega CV(\omega).
\]

Noting that \( V(\omega) = j\omega L I_L(\omega) \), the above expression can be used to calculate the quantity \( VI^* \)

\[
\frac{1}{2} V(\omega)I^*(\omega) = \frac{1}{2} \frac{V(\omega)V^*(\omega)}{R} + \frac{j\omega}{2} \left[ L I_L(\omega)I_L^*(\omega) - C V(\omega)V^*(\omega) \right].
\]

The time-average power dissipation in the resistor is \( P_R = (1/R)|V(\omega)|^2/2 \), whereas, the time-average energies stored in an inductor and a capacitor are \( U_L = L |I_L(\omega)|^2/4 \) and \( U_C = C |V(\omega)|^2/4 \), respectively, so that the desired form of the above is

\[
\frac{1}{2} V(\omega)I^*(\omega) = \text{Ohmic Losses} + j2\omega(U_L - U_C). \tag{1.2.4}
\]

At resonance, time-average energies stored in the inductor and capacitor are equal while off resonance they are unequal. This expression illustrates the fact that for a generic resonant system, there can be a large circulation or transfer of energy between two passive components and system losses occur because that oscillating energy incurs losses during its transfer. The coefficient \( \omega \) contains the inverse oscillation period \( T = 1/f \) which is the transfer time of the
energy between the inductor and capacitor at resonance. Thus, there is a flux or “power” flow. Although a storage element may store a given amount of energy that is proportional to the voltage/current values, the power flow is proportional to frequency. Off resonance there will be energy oscillation between the current source and the energy storage elements. Thus, power flows in one direction for half the period and flows in the opposite direction for the other half. If the storage elements have frequency-dependent values, the above equations must be modified. A similar expression obtained in the derivation of the complex Poynting vector will be illustrated later in this chapter.

1.3 Spectral Densities

When signals have simple harmonic time variation as in Section 1.2 the computation of the power or energy is relatively simple. If the signal is not harmonically varying with time then the computation of power or energy must be modified from that of Section 1.2. A technique can be used to determine time-average powers using the power spectral densities of cross- and autocorrelation functions. Further, the technique is applicable to many signals whose variation in time is stochastic. For example, the cross-correlation function, \( R_{iv}(\tau) \) is (for power signals)

\[
R_{iv}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} i(t + \tau)v(t) \, dt,
\]

and using the Weiner-Khinchine relation, its power spectral density is

\[
S_{iv}(\omega) = \mathcal{F}[R_{iv}(\tau)].
\]

Using the differential equation for the node voltage in Fig. 1.1, the relation between the various spectral densities is

\[
S_{iv}(\omega) = \frac{1}{R} S_{v}(\omega) + j2\omega \left[ \frac{1}{2} L S_{i_{L}}(\omega) - \frac{1}{2} C S_{v_{C}}(\omega) \right],
\]

where \( S_{i_{L}}(\omega) \) is the spectral density of the current through the inductor whereas, \( S_{v_{C}}(\omega) \) is the spectral density of the voltage across the capacitor \( (v(t) = v_{C}(t)) \). The power spectral densities are real functions that are symmetric about \( \omega = 0 \) and the total power dissipation is (in R)

\[
P_L = \frac{1}{R} \int_{-\infty}^{\infty} S_{v}(\omega) \, d\omega.
\]

The time-average energy storage in \( L \) and \( C \), respectively, are

\[
U_L = \frac{1}{2} L \int_{-\infty}^{\infty} S_{i_{L}}(\omega) \, d\omega,
\]

\[
U_C = \frac{1}{2} C \int_{-\infty}^{\infty} S_{v_{C}}(\omega) \, d\omega.
\]
1.3 SPECTRAL DENSITIES

The spectral density of real time signals is an even function of $\omega$ because auto-correlation functions are even about $\tau = 0$. The spectral density of a cross-correlation function may not be symmetric and is generally complex, having both a real part and imaginary part. Because the imaginary part of $S_{iv}(\omega)$ is an odd function of $\omega$, the integration of $S_{iv}(\omega)$ over the interval $(-\infty, \infty)$ is a real quantity and represents the time-average power drain on the current source.

1.4 Maxwell’s Equations

Many treatments of the electromagnetic theory deduce Maxwell’s equations from Faraday’s law and from Ampere’s circuital theorem, while other treatments start with Maxwell’s equation as first principals.[30] In some sense, using Maxwell’s equations as a starting point is completely logical as long as their limitations are understood. Using the MKS units, the field equations are

$$\nabla \times \mathbf{e} = -\frac{\partial \mathbf{b}}{\partial t},$$

(1.4.1)

$$\nabla \times \mathbf{h} = \mathbf{j} + \frac{\partial \mathbf{d}}{\partial t},$$

(1.4.2)

$$\nabla \cdot \mathbf{d} = \rho,$$

(1.4.3)

$$\nabla \cdot \mathbf{b} = 0.$$  

(1.4.4)

As mentioned earlier, the quantities $\mathbf{e}$ and $\mathbf{h}$ are the electric and magnetic field intensities, respectively, while $\mathbf{b}$ and $\mathbf{d}$ are the corresponding magnetic and electric flux densities. In (1.4.2), $\mathbf{j}$ is the conduction current while $\partial \mathbf{d}/\partial t$ is the displacement current. The current density $\mathbf{j}$ can be split into two terms: (1) that due to external sources (current generators) and (2) that which is induced by the fields themselves. For example, an induced current density, due to the electric field, can be expressed by the linear relation such as that given in 1.4.7 (Ohm’s Law). The attenuation or growth of a propagating electromagnetic field can be attributed to the conductivity as discussed in Section 1.7. For material that attenuates a wave, the conductivity is positive, while the conductivity is negative for a material that amplifies a wave.

The electromagnetic field sources $\mathbf{j}$ and $\rho$ are related via the continuity equation

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0,$$

(1.4.5)

that can be obtained from Eqs. (1.4.2) and (1.4.3), by multiplying both sides by divergence operator.

Maxwell’s equations in their present form are incomplete until the relations between the various field quantities, $\mathbf{e}$, $\mathbf{d}$, $\mathbf{b}$ and $\mathbf{h}$ are connected to the medium under consideration. The physics of the medium or material must be investigated to find the relation between the various quantities. (The relationships are generally referred to as the constitutive relations.) Before we discuss
some of the more general constitutive relations, first consider mathematical relations related to the instantaneous Poynting vector, \( s = \mathbf{e} \times \mathbf{h} \),

\[
\mathbf{e} \cdot \nabla \times \mathbf{h} - \mathbf{h} \cdot \nabla \times \mathbf{e} = -\nabla \cdot \mathbf{s} = \mathbf{e} \cdot \mathbf{j} + \mathbf{e} \cdot \frac{\partial \mathbf{d}}{\partial t} + \mathbf{h} \cdot \frac{\partial \mathbf{b}}{\partial t}
\]

(1.4.6)

If the electric field \( \mathbf{e} \) is not out of phase with the current density \( \mathbf{j} \), then the time average \( \langle \mathbf{e} \cdot \mathbf{j} \rangle \) will produce the ohmic losses per unit volume. In an isotropic, non-dispersive medium, the constitutive relations are

\[
\mathbf{j} = \sigma \mathbf{e}, \quad \mathbf{d} = \varepsilon \mathbf{e}, \quad \mathbf{b} = \mu \mathbf{h}.
\]

(1.4.7)-(1.4.9)

In an anisotropic, dispersive medium, the constitutive relations are considerably more complicated as discussed later in this chapter. Nevertheless, using the simple constitutive relations, (1.4.7)-(1.4.9), (1.4.6) becomes,

\[
-\nabla \cdot \mathbf{s} = \sigma \mathbf{e} \cdot \mathbf{e} + \frac{\partial}{\partial t} \left( \frac{1}{2} \varepsilon \mathbf{e} \cdot \mathbf{e} + \frac{1}{2} \mu \mathbf{h} \cdot \mathbf{h} \right).
\]

(1.4.10)

As noted earlier, the first term on the right-hand-side of the above equation represents the instantaneous ohmic losses per unit volume. The remaining two terms represent the time rate of change of the instantaneous energy stored per unit volume in the electric and magnetic fields respectively. Integrating both sides of (1.4.10) over a given volume, the integral of \( \sigma \mathbf{e} \cdot \mathbf{e} \) is the instantaneous power dissipated within the volume; the time average represents the average power dissipation. Applying Gauss’s Law to \( -\nabla \cdot \mathbf{s} \), the resulting quantity \( \mathbf{n} \cdot \mathbf{s} \) is interpreted as the instantaneous power flux flowing across the surface into the surrounding the volume, where \( \mathbf{n} \) is surface normal (externally directed).

Using Fourier analysis, all field quantities, such as that given in (1.1.5), can be assumed to vary harmonically as \( \exp j\omega t \), so that Maxwell’s equations can be written as

\[
\nabla \times \mathbf{E} = -j\omega \mathbf{B}, \quad \nabla \times \mathbf{H} = \mathbf{J} + j\omega \mathbf{D}, \quad \nabla \cdot \mathbf{D} = \rho, \quad \nabla \cdot \mathbf{B} = 0.
\]

(1.4.11)-(1.4.14)

The corresponding constitutive relations become[23]

\[
\mathbf{J} = \sigma \mathbf{E}, \quad \mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}.
\]

(1.4.15)-(1.4.17)
1.4 Maxwell’s Equations

The continuity equation for these harmonically time varying fields can be obtained from (1.4.13).

1.5 Complex Poynting Vector

The Poynting vector $s(t)$ represents the instantaneous flow of power “flux”. If the electromagnetic fields have harmonic time variation, the use of the complex notation for the field quantities are useful for finding the time-average quantities as discussed in Section 1.2. In this discussion, the constitutive relations are assumed for an isotropic material that is non-dispersive. Using the two curl equations given by (1.4.11) and (1.4.12),

$$\mathbf{H}^* \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H}^* = \nabla \cdot (\mathbf{E} \times \mathbf{H}^*) \equiv \nabla \cdot \mathbf{S},$$  (1.5.1)

where $\mathbf{S}$ is the complex Poynting vector and satisfies

$$-\frac{1}{2} \nabla \cdot \mathbf{S} = \frac{1}{2} \mathbf{E} \cdot \mathbf{J}^* + \frac{1}{2} j\omega (\mathbf{E} \cdot \mathbf{D}^* - \mathbf{H}^* \cdot \mathbf{B}),$$

$$= \frac{1}{2} \sigma \mathbf{E} \cdot \mathbf{E}^* + 2j\omega \left(\frac{1}{4} \mu \mathbf{H} \cdot \mathbf{H}^* - \frac{1}{4} \varepsilon \mathbf{E} \cdot \mathbf{E}^*\right).$$  (1.5.2)

The term $\sigma \mathbf{E} \cdot \mathbf{E}^*/2$ is the time-average ohmic losses per unit volume, while $U_e = \varepsilon \mathbf{E} \cdot \mathbf{E}^*/4$ and $U_m = \mu \mathbf{H} \cdot \mathbf{H}^*/4$ are the time-averages of the stored electric and magnetic energies respectively. The integration of the real part of $-\mathbf{S} \cdot \hat{n}$ over a surface represents the time-average power dissipation in the volume surrounded by the surface. The integration of the imaginary part of $-\mathbf{S} \cdot \hat{n}$ over a surface represents the of the difference in time-average stored electric energy $U_e$ and the time-average stored magnetic energy $U_m$ in the volume surrounded by the surface.

1.6 Boundary Conditions

The solution of Maxwell’s equations in a given region of space is matched to the solution in another region of space by appropriately fitting the fields at the boundary. The material characteristics of the two regions are illustrated in Fig. 1.2 where each region has its corresponding material constants. For the sake of simplicity, the material constants are assumed to be scalar quantities. The unit vector which is normal to the surface is designated as $\hat{n}$. The direction has been arbitrarily selected as pointing from region 1 to region 2.

![Figure 1.2](image_url)

**Figure 1.2.** The boundary between two different materials. The tangent vector lies parallel to the interface whereas the normal vector is perpendicular to the interface.
The normal component of the magnetic flux must be continuous, while the difference in the normal components of the electric flux accounts for a surface charge density at the interface. On the other hand, the continuity of the tangential electric field is due to the fact that the electric field vector is conservative. Differences in the tangential components of the magnetic field yields the surface current density on the interface. The boundary conditions satisfy the following vector equations

\[ \hat{n} \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0, \]  
\[ \hat{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \rho_s, \]  
\[ \hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_s, \]  
\[ \hat{n} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0, \]

where \( \rho_s \) is the surface charge density and \( \mathbf{J}_s \) is the corresponding surface charge density, which are connected via the continuity equation. More specifically, (1.6.1) indicates that the normal component of the magnetic field is continuous across the boundary, and (1.6.2) indicates that the normal component of the electric flux density \( \mathbf{D} \) is discontinuous by an amount equivalent to the surface electric charge density. The tangential field components are governed by (1.6.3) indicating that the tangential magnetic field intensity is discontinuous by an amount equivalent to the surface charge density, while (1.6.4) dictates that the tangential electric fields must be continuous.

For the case in which the surface charge density and surface current densities are nonexistent, the normal boundary fields \( \mathbf{B}_n \) and \( \mathbf{D}_n \) and tangential boundary fields \( \mathbf{E}_t \), and \( \mathbf{H}_t \) are all continuous from one medium to another.

1.7 Complex Dielectric Constant

The Maxwell equation, Eq. (1.4.12), can be written in a form for lossy or conducting material which is similar to the case for a reactive medium, one without losses. Instead of being real, the dielectric constant will be assumed to be complex.[29] The equivalent form is useful for obtaining solutions to various boundary value problems with lossy regions from those solutions obtained for a lossless reactive material. Using the constitutive relations, Eq. (1.4.12) becomes

\[ \nabla \times \mathbf{H} = (\sigma + j \omega \varepsilon) \mathbf{E}, \]
\[ = j \omega \varepsilon_0 (\kappa_e - j \sigma / \omega \varepsilon_0) \mathbf{E}, \]  

where \( \varepsilon_0 \) is the free-space permittivity (The subscript is “o”, not zero. All free-space constants will be designated with an “o” subscript.) and \( \kappa_e \) is the relative dielectric constant. The term \( \kappa_e \) can be a real function of frequency provided the conductivity \( \sigma \) is frequency dependent. However, if \( \sigma \equiv \sigma_{dc} \), the “dc” conductivity, then relative dielectric constant must be assumed to be complex in order to account for losses. Therefore, \( \kappa_e \) must be replaced by complex value \( \kappa_e \) whose imaginary part vanishes at \( \omega = 0 \) because \( \sigma_{dc} \) accounts for
ohmic losses at \( \omega = 0 \). The relative complex dielectric constant is then defined as

\[
\kappa'_e(\omega) + j \kappa''_e(\omega) = \varepsilon_e(\omega) - j \varepsilon_{dc} / \omega \varepsilon_0 \equiv 1 + \chi_e(\omega), \tag{1.7.2}
\]

where \( \chi_e(\omega) \) is the complex electric susceptibility.

Alternatively, one can define complex conductivity as

\[
\sigma(\omega) = \sigma'(\omega) + j \sigma''(\omega) = \sigma_{dc} + j \omega \varepsilon_0 \tilde{\varepsilon}_e(\omega). \tag{1.7.3}
\]

For the case of a lossy material, \( \sigma > 0 \), the imaginary part of \( \sigma_e \) is negative. (In the definition of \( \tilde{\varepsilon}_e(\omega) \) in (1.7.2), \( \kappa''_e(\omega) < 0 \) for lossy material. This occurs because of the form of the time dependence \( \exp(j\omega t) \).) The corresponding complex dielectric constant is \( \varepsilon(\omega) = \varepsilon_0 \tilde{\varepsilon}_e(\omega) \), where the free-space permittivity is \( \varepsilon_0 \approx 10^{-9} / 36\pi \) farads/meter. If the conductivity, \( \sigma_{dc} \neq 0 \), then the complex dielectric constant and susceptibility have simple poles at \( \omega = 0 \).

The optical properties of a semiconductor can be described by the complex dielectric constant and the nature of the electromagnetic wave propagation in a semiconductor is dependent on the optical constants of the material. The two optical constants totally responsible for the propagation characteristics in the material discussed here are (1) the index of refraction \( n = \sqrt{\kappa'_e} \), and wave attenuation coefficient \( \alpha \). \( \uparrow \) The two optical constants \( n \) and \( \alpha \) are more commonplace than the real and imaginary parts of the complex dielectric constant.

Consider for the moment, the propagation of a plane wave in a lossy dielectric material. Assume the field is propagating in the positive \( z \) direction and has the form \( \exp(j\omega t - \gamma z) \). The complex propagation constant

\[
\gamma = \alpha + j \beta = j \omega \sqrt{\mu_0 \varepsilon_0}, \tag{1.7.4}
\]

where \( \alpha > 0 \) because of the lossy material. Substituting the expression for the complex dielectric constant into (1.7.4), the propagation constant becomes

\[
\gamma = j k_0 \sqrt{\kappa'_e - j \kappa''_e}, \tag{1.7.5}
\]

where \( k_0 = \omega \sqrt{\mu_0 \varepsilon_0} \) is the free-space wave number. The value of \( \alpha \) becomes

\[
\alpha = k_0 \frac{\sqrt{\kappa'_e}}{\sqrt{2}} \left( \sqrt{1 + (\kappa''_e / \kappa'_e)^2} - 1 \right)^{1/2}, \tag{1.7.6}
\]

while \( \beta \) becomes

\[
\beta = k_0 \frac{\sqrt{\kappa'_e}}{\sqrt{2}} \left( \sqrt{1 + (\kappa''_e / \kappa'_e)^2} + 1 \right)^{1/2}. \tag{1.7.7}
\]

\( \uparrow \)

Generally, the magnetic permeability \( \mu \) may be complex; however, when \( \mu = \mu_0 = 4\pi \times 10^{-7} \) henrys/meter, the material is said to be nonmagnetic. On the other hand, some materials that are used for the fabrication of optical isolators have permeabilities that are different from that of free-space. Indeed, the permeability may also be described by a complex quantity.
If $\gamma$ is written as
\[ \gamma = \alpha + j\beta = k_0(A + jB), \] (1.7.8)
then the normalized attenuation constant $A = \alpha/k_0$ and normalized propagation constant $B = \beta/k_0$ can be approximated for $|\varepsilon''/\varepsilon'| \ll 1$

\[ B = \sqrt{\varepsilon'}, \] (1.7.9)
\[ A = \frac{\varepsilon''}{2 \sqrt{\varepsilon'}} = \frac{\sigma Z_0}{2 k_0 \sqrt{\varepsilon'}}, \] (1.7.10)

where $Z_0 = 120\pi$ Ohms is the free-space wave impedance. Another common term is the complex refractive index

\[ \bar{n} = \sqrt{k_0} = n - jk, \] (1.7.11)

where $n = B$ is the index of refraction and $k = A$ is the extinction coefficient.

The important results here is the relationship between the components of the complex dielectric constant and the familiar optical constants $n$ and $\alpha$. Consequently, if the values of $n$ and $\alpha$ are known, the complex dielectric constant can be determined. It should be noted that $\alpha$ represents the wave attenuation whereas, $2\alpha$ is the power attenuation, sometimes called the absorption coefficient.

The conductivity, $\sigma$, the permittivity, $\varepsilon = \varepsilon_0 \varepsilon_r$, and the permeability, $\mu = \mu_0 \mu_r$, in (1.4.15)-(1.4.17) are the optical constants of the material defined at the optical wavelength or frequency, $\omega/2\pi$. It is uncommon to discuss the optical conductivity of a material because, experimentally, the complex index of refraction, $n - jk$, is usually an experimentally measured quantity and $\sigma$ is deduced from the complex index of refraction. At low frequencies (rf frequencies) the conductivity of a semiconductor is related to the flow of electrons in the crystal. When an electric field is applied to the material, the free-electrons and holes migrate in opposite directions as determined by the polarization of the applied electric field. Because free-electrons/holes scatter off atomic sites in the crystal, the free-carriers cannot have an infinite momentum and have a velocity that is proportional to the applied electric field strength. Thus, conductivity is finite. The resulting “ohmic” losses are due to the scattering, producing crystal vibrations (phonons), and the corresponding thermal heating of the semiconductor lattice. Consequently, as a wave propagates through the material, the wave experiences attenuation because of the ohmic losses. At optical wavelengths, the losses are much more complicated. Inter-band transitions are a major source of wave attenuation and propagating fields experience large attenuation at wavelengths below the band-gap wavelength of the material. Introducing the complex dielectric constant is the vehicle for accounting for the losses.

A final comment on the complex dielectric constant formulation is that it simplifies Maxwell’s equation by dropping the induced conduction current. Alternatively, the displacement current could be dropped by making the conduction current complex. In terms of analysis, one is not better than the other.
1.8 General Constitutive Relations

In many cases, optical properties of a semiconductor may be described by the simple complex dielectric constant as discussed earlier. However, the fundamental relations of the various field quantities are modified for high-frequency signals. Because the nature of the electromagnetic wave propagation in a semiconductor is dependent on the optical constants of the material, the field behavior must be characterized for the signal frequencies of interest.

The electronic structure of an atom can only be understood using proper quantum mechanical analysis; otherwise, the electron would spiral into the nucleus, due to the attractive force. For a single electron hydrogen atom, the electron’s quantum mechanical wave function indicates the electron is located about $0.5 \times 10^{-10}$ meters (Bohr radius) from the nucleus. Crystals such a Si and Ge have several electrons. As an example, Fig. 1.3 shows an atom with $N = 10$ electrons. In the absence of an applied field, electron clouds are distributed about the nucleus and typically have their “center-of-charge” located at the nucleus as in Fig. 1.3(a). When a static field is applied the nucleus and electrons are pulled in opposite directions and the electron charge center

$$
\mathbf{r}_e = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i
$$

is displaced from the nucleus. The resulting dipole moment is

$$
\rho = e N (\mathbf{r}_{nu} - \mathbf{r}_e),
$$

where $e$ is the electron charge. The displacement $\mathbf{r}_{nu} - \mathbf{r}_e$ is a function of the electric field and for weak fields the resulting dipole moment $\rho$ may be written as

$$
\rho = \alpha e,
$$

(1.8.1)

where $\alpha$ represents the polarizability of the atom (the ability of an applied electric field to redistribute the charge distribution). If $\rho$ and $e$ are parallel, the material formed by the atoms is said to be isotropic, otherwise, it is anisotropic. It should be obvious that in a crystal structure formed by atoms, atoms at different sites may have different values of $\rho$ (magnitude and/or direction) but the aggregate behaves in a specific manner.

![Figure 1.3](image)

Figure 1.3. (a) shows the free atom and (b) the charge displacements with an applied electric field.
In a cubic lattice structure as shown in Fig. 1.4 where the atoms are separated by \( d \), the “volume” of polarized atom is \( V = d^3 \) and the resulting static polarization in the crystal becomes

\[
P = \frac{1}{V} \rho.
\]

where \( P \) is the dipole moment per unit volume. The units of \( P \) are the same as for \( D \), coulombs/m\(^2\) in SI. At the boundary of the crystal, there is a surface polarization charge density with \( P_n = \rho_{sp} \), where \( P_n \) is the component of \( P \) that is normal to the surface.

Even in the absence of an applied field the microscopic electric field will have large spatial variations due to the electric charges within the crystal. Also, the microscopic electric field is not constant in time at a point in the crystal because of the movement of electrons as well as atomic sites because of lattice vibrations. As a result, the individual dipole moments at the atomic sites in Fig. 1.4 are moving and well as changing value because \( r_{nu} \) and \( r_c \) are changing with time. However, the average value of the microscopic field over a volume much greater than \( V \) is constant with time and in the absence of an applied field, the average field \( \langle E \rangle = 0 \)

The macroscopic fields, are average field distributions in the crystal and the macroscopic static field quantities are related (a constitutive relation)

\[
D = \varepsilon_0 E + P.
\]

It goes without saying that \( P \) is a function of the resulting macroscopic electric field \( E \) which is written as

\[
P = \varepsilon_0 \chi_e E,
\]

where \( \chi_e \) is the electric susceptibility that depends on the atomic structure of the atoms forming the crystal.
When the applied electric field has rapid changes, the polarization of the atom will not follow instantaneously. For slow variations of $e(t)$, the polarization will lag an inconsequential amount. In any event the resulting dipole produces a dipole moment so that in a crystal the aggregate of induced dipoles produces polarization $P$ that is a function of the applied electric field. In some instances, the polarization vector points in a direction that is different from that of the electric field. In this case, the material is said to be anisotropic, otherwise it is isotropic. Because the dynamic response of the field in a dielectric medium is the topic of the present discussion, for the moment, the material will assumed to be isotropic.

Causality can be defined as a response to an event such that the response cannot occur before the event. In particular, the time dependent polarization $p(t)$ can be written as a function of a time-dependent variable $X_e(t)$ and the electric field for all times in the interval $(-\infty, t)$ as

$$p(t) = \varepsilon_0 \int_{0}^{\infty} X_e(\tau)e(t - \tau) \, d\tau$$

(1.8.2)

where $X_e(t)$ is the response of the polarization due to an impulse of the electric field occurring at $t = 0$, i.e., $e(t) = E_0\delta(t)$.

The Fourier transform of $p(t)$ is

$$P(\omega) = \varepsilon_0 \chi_e(\omega)E(\omega),$$

where $e(t) \leftrightarrow E(\omega)$ is a Fourier-transform pair while

$$\chi_e(\omega) = \int_{0}^{\infty} X_e(\tau) e^{-j\omega\tau} \, d\tau.$$  

(1.8.3)

Since $X_e(t) = 0$ for $t < 0$, the lower limit can be extended to $-\infty$. The complex variable $\chi_e = \chi'_e + j\chi''_e$ satisfies the condition $\chi_e(\omega) = \chi'_e(-\omega)$, so that $\chi'_e(\omega) = \chi'_e(-\omega)$ while $-\chi''_e(-\omega)$.

The integral in (1.8.3) is evaluated over the real variable $\omega$, but, $\omega$ could be replaced with a complex variable $\xi$ and the integral in (1.8.3) would be along the real axis. The path of integration could be modified and yet retain its value as defined in (1.8.3). By replacing $\omega$ with the complex variable $\xi = \xi + j\eta$, the variable on the real axis is $\xi$. The resulting complex function $\chi_e(\xi)$ becomes an “entire function” (An entire function has no poles in the entire $\xi$ plane. Furthermore, the integral of an entire function over a closed path always vanishes.), except when $\sigma(0) = \sigma_{dc} \neq 0$ (see Eq. 1.7.2). With $\sigma_{dc} \neq 0$, there is a simple pole at $\xi = 0$. Furthermore, a new function

$$f(\xi) = \frac{\chi_e(\xi)}{\xi - \xi_0},$$

(1.8.4)

In keeping with earlier notations, time varying quantities are designated in lower case letters, while the transformed variables are in upper case letters. In the discussion on dipole moments, the moment of a single atom was designated by $p$, and not $p$ which represents the time-varying polarization.
has an additional simple pole at $\xi = \xi_0$. The contour in Fig. 1.5 avoids both poles so that a counter-clockwise integral over $C$ vanishes,

$$
\int_C f(\xi) d\xi = \int_{-R}^{-\rho} f(\xi) d\xi + \int_{\rho}^{2\pi} f(\rho e^{j\theta}) j\rho e^{j\theta} d\theta + \int_{\rho}^{\xi_0 - \rho} f(\xi) d\xi
+ \int_{\pi}^{2\pi} f(\xi_0 + \rho e^{j\theta}) j\rho e^{j\theta} d\theta + \int_{\xi_0 + \rho}^{R} f(\xi) d\xi + \int_{\pi}^{2\pi} f(R e^{j\theta}) jRe^{j\theta} d\theta = 0
$$

Figure 1.5. The function $f(\xi) = \chi_{\xi}(\xi)/(\xi - \xi_0)$ has no poles within the region enclosed by the contour. The large semi circle has a radius $R$ while $\rho$ is the radius of the two small semi circles.

Integration on the large semi circle vanishes as $R \to \infty$ because the exponential term in (1.8.3) becomes

$$
e^{-j\xi\tau} = e^{R\tau(-j\cos\theta + \sin\theta)} \to 0,$$

as $R \to \infty$ because $\sin \theta < 0$.

Similarly, the relation between the magnetic flux density $b$ and the magnetic field intensity $h$ can be written as a convolution. First, the magnetization $m(t)$ is related to $h(t)$ as

$$m(t) = X_m(t) \ast h(t),$$

and the corresponding magnetic flux density is

$$b(t) = \mu_0 [h(t) + m(t)] = \mu_0 K_m(t) \ast h(t).$$

The Fourier transform pairs are $\chi_m(\omega) \leftrightarrow X_m(t)$ and $K_m(\omega) \leftrightarrow K_m(t)$. The quantity $K_m(\omega)$ is the frequency-dependent relative permeability of the medium while $\chi_m(\omega)$ is the frequency-dependent magnetic susceptibility.

### 1.9 Narrow-Band Processes

The previous discussion of the Poynting theorem in Section 1.4 and Section 1.5 was developed for a specific medium (dispersionless and isotropic). A general discussion requires detailed information regarding properties of the material.
Without knowledge of the material, used to tie the various field vectors by way of the constitutive relations, the general Poynting theorem contains little information regarding details of power flow in the material and how energy is stored in the electromagnetic fields. In free-space the permittivity is a constant and polarization is absent. Further, when the material is non-dispersive, the polarization tracks the electric field and the magnetization tracks the magnetic field strength with no time delays; the permittivity and permeability of the material are simple constants. (The previous discussion of the complex Poynting vector was specifically for a simple non-dispersive medium.) Dispersive material is commonplace in photonic devices and their performance in systems is critically dependent on dispersion properties. Therefore, attention will now be focused on the Poynting theorem for those materials.

The general thrust of the Poynting deviation, given in (1.4.6) produces several vector products that need to be addressed for a proper understanding of power flow and energy of an electromagnetic field. For example, the time derivative of the electric flux density must first be computed to determine the cross correlation between \( e \) and \( \frac{\partial \mathbf{d}}{\partial t} = \mathbf{d} \).

A general discussion of systems with dispersion requires a detailed description of the frequency response of the materials over all frequencies. Accordingly, discussion here will be limited to signals that have a specific type of time variation that is common to contemporary photonic communication systems. Namely, the photons will have wavelengths concentrated in a narrow part of the wavelength spectrum centered about \( \omega_0 \), as shown in Fig. 1.6. The power spectral densities of various terms in (1.4.6) will be computed for the narrow-band signals. First, the inner product \( e \cdot \mathbf{d} \) which relates to the energy stored in the electric field, will be computed specifically for the narrow-band signal.

Consider two frequency components, \( \omega_0 + \delta \omega \) and \( \omega_0 - \delta \omega \), so that the time dependence of \( e(t) \) is

\[
e(t) = \frac{1}{2} E_0 [\cos(\omega_0 + \delta \omega)t + \cos(\omega_0 - \delta \omega)t].
\]

†

For the sake of compactness, the spatial dependencies of \( e(t) \) and \( E_0 \) are understood, i.e., \( E_0 = E_0(x, y, z) \).
where the modulation function \(m(t)\) is slowly varying with time. The corresponding magnetic field intensity is

\[
h(t) = H_0 m(t) \cos(\omega_0 t + \phi),
\]

(1.9.2)

where \(\phi\) is the phase difference between the electric and magnetic fields.

While the power spectral density of the electric field \(e(t)\) is shown in Fig. 1.6, the power spectral density of \(m(t)\) is shown in Fig. 1.7. For the example function \(m(t) = \cos \delta \omega t\), its power spectral density is

\[
S_m(\omega) = \frac{1}{4} \delta(\omega - \delta \omega) + \frac{1}{4} \delta(\omega + \delta \omega),
\]

and it total time-average power \(\langle m^2(t) \rangle = 1/2\).

The derivative of (1.9.1) with respect to time is

\[
\frac{\partial \mathbf{d}(t)}{\partial t} = \mathbf{d}(t) = \varepsilon_0 \int_{-\infty}^{\infty} \frac{\partial K_\varepsilon(t - \tau)}{\partial t} \mathbf{e}(\tau) \, d\tau.
\]

(1.9.3)

The Fourier transform of (1.9.3) is

\[
\mathcal{F}\left[\mathbf{d}(t)\right] = \varepsilon_0 j \omega K_\varepsilon(\omega) \mathbf{E}(\omega),
\]

(1.9.4)

where \(K_\varepsilon(\omega)\) is the frequency-dependent relative dielectric constant. Since we are interested in only the frequencies near \(\omega_0\) and because of the narrow-band spectrum of the electric field, \(\omega K_\varepsilon(\omega)\) can be approximated with a two-term Taylor series

\[
\omega K_\varepsilon(\omega) \approx (\omega_0 K_\varepsilon) + \left. \frac{d(\omega K_\varepsilon)}{d\omega} \right|_{\omega = \omega_0} (\omega - \omega_0),
\]

\[
\equiv \omega_0 K_\varepsilon + \left(\omega_0 K_\varepsilon\right)_1 (\omega - \omega_0).
\]

(1.9.5)
Substituting (1.9.5) back into (1.9.4), and performing the inverse Fourier transform, the derivative of the impulse response becomes

$$\frac{\partial K_e(t)}{\partial t} = j(\omega k_e) \delta(t) + (\omega k_e) \hat{\delta}(t) e^{j\omega_0 t}. \quad (1.9.6)$$

Accordingly, the time derivative of the electric flux becomes, after using the narrow-band expression of the electric field given in (1.9.1),

$$\dot{d}(t) = \varepsilon_0 \mathbf{E}_0 \left[ (\omega \varepsilon_0) m(t) e^{j(\omega_0 t + \pi/2)} + (\omega \varepsilon_1) \dot{m}(t) e^{j\omega_0 t} \right]. \quad (1.9.7)$$

Upon inserting \( \mathfrak{M} \) into (1.9.7), the time derivative of the electric flux density becomes

$$\dot{d}(t) = \varepsilon_0 \mathbf{E}_0 \left[ (\omega \varepsilon_0) m(t) \sin \omega_0 t + (\omega \varepsilon_1) \dot{m}(t) \cos \omega_0 t \right]. \quad (1.9.8)$$

The cross-correlation function of the vectors \( \mathbf{e}(t) \), (1.9.1), and \( \dot{d}(t) \) becomes

$$R_{\mathbf{e}, \dot{d}}(t) = \frac{\varepsilon_0}{2} \mathbf{E}_0 \cdot \mathbf{E}_0 \left[ (\omega \varepsilon_0) R_m(\tau) \sin \omega_0 \tau + (\omega \varepsilon_1) R_{m\dot{m}}(\tau) \cos \omega_0 \tau \right]. \quad (1.9.9)$$

where \( R_m(\tau) \) is the auto-correlation function of \( m(t) \) and \( R_{m\dot{m}}(\tau) \) is the cross-correlation of \( m(t) \) and \( \dot{m}(t) \). The time average of \( \mathbf{e} \cdot \dot{d} \) is zero as the cross-correlation function \( R_{m\dot{m}}(0) = 0 \) because \( R_{m\dot{m}}(\tau) \) is an odd function about \( \tau = 0 \) while the auto-correlation function \( R_m(\tau) \) is an even function. Nevertheless, the power spectral density \( S_{\mathbf{e}, \dot{d}}(\omega) \) of the cross-correlation function \( R_{\mathbf{e}, \dot{d}}(\tau) \) can be written as

$$S_{\mathbf{e}, \dot{d}}(\omega) = -j \frac{\varepsilon_0}{4} \mathbf{E}_0 \cdot \mathbf{E}_0 \left\{ (\omega \varepsilon_0) \left[ S_m(\omega - \omega_0) - S_m(\omega + \omega_0) \right] \right.$$  

$$+ (\omega \varepsilon_1) \left[ \omega \left( S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right) \right.$$  

$$- \omega_0 \left( S_m(\omega - \omega_0) - S_m(\omega + \omega_0) \right) \left\} \right.$$,  

$$= -j \frac{\varepsilon_0}{4} \left[ \mathbf{E}_0 \right]^2 \left( \omega \varepsilon_0 \right) \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right], \quad (1.9.10)$$

which is only a function of the power spectral density of the base-band signal \( m(t) \). It should be noted that the spectral density of the cross-correlation function has only an imaginary part which therefore represents reactive power spectral density. Further, the integral of that reactive power spectral density, \( S_{\mathbf{e}, \dot{d}}(\omega) \), over the interval \((-\infty, \infty)\) is zero. The central term of (1.9.10) has a coefficient that is linear in \( \omega \), so that its coefficient is interpreted as the energy spectral density that is stored in the electric field. Thus, the time average of the energy per unit volume per hertz stored in the electric field is

$$\frac{d \varepsilon_0}{d \omega} = \frac{\varepsilon_0}{4} \left[ \mathbf{E}_0 \right]^2 \left( \omega \varepsilon_0 \right) \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right]. \quad (1.9.11)$$

In a similar manner, the power spectral density of the cross-correlation function, between the \( \mathbf{b}(t) \) and \( \mathbf{h}(t) \) can be expressed as

$$S_{\mathbf{b}, \mathbf{h}}(\omega) = -j \frac{\mu_0}{4} \mathbf{H}_0 \cdot \mathbf{H}_0 \left\{ (\omega \mu_0) \left[ S_m(\omega - \omega_0) - S_m(\omega + \omega_0) \right] \right.$$  

$$+ (\omega \mu_0) \left[ \omega \left( S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right) \right.$$  

$$- \omega_0 \left( S_m(\omega - \omega_0) - S_m(\omega + \omega_0) \right) \left\} \right.$$,  

$$= -j \frac{\mu_0}{4} \left[ \mathbf{H}_0 \right]^2 \left( \omega \mu_0 \right) \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right]. \quad (1.9.12)$$
The time average of energy per unit volume per hertz stored in the magnetic field is

$$\frac{d w_m}{d \omega} = \frac{\mu_0}{4} |H_0|^2 (\omega \kappa_m) \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right]. \tag{1.9.13}$$

For the sake of completeness, the power spectral density of the cross correlation of $e(t)$ and $j(t)$ is

$$S_{e-j}(\omega) = \frac{1}{4} E_0 \cdot E_0 \left\{ \sigma_0 \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right] + \sigma_1 \left[ (\omega - \omega_0) S_m(\omega - \omega_0) - (\omega + \omega_0) S_m(\omega + \omega_0) \right] \right\}. \tag{1.9.14}$$

The terms $\sigma_0$ and $\sigma_1$ are the expansion coefficients of $\sigma(\omega)$ about $\omega_0$. Because of the symmetry of $S_m(\omega)$ about $\omega = 0$, the integral over all frequencies of the second and third terms is zero and there is no contribution to the average power. The average power lost due to heating is

$$P_L = \frac{\sigma_0}{4} |E_0|^2 \int_{-\infty}^{\infty} \left[ S_m(\omega - \omega_0) + S_m(\omega + \omega_0) \right] d\omega. \tag{1.9.15}$$

The Poynting theorem given by (1.4.6) can now be expressed in terms of the spectral density of the base-band signal $m(t)$ which has only a narrow band of frequencies concentrated near $\pm \omega_0$, as

$$-\nabla \cdot S_0 \cdot S_{e \times h}(\omega) = S_{e \cdot e}(\omega) + S_{h \cdot h}(\omega) + S_{e \cdot j}(\omega), \tag{1.9.16}$$

where $S_0 = E_0 \times H_0$, characterizes the direction of the power flux, and $S_{e \times h}(\omega)$ is the power spectral density of the cross correlation of the vector product of $e(t)$ and $h(t)$.

### 1.10 Wave Equation

The equation defining the wave phenomena of the electromagnetic fields can be derived from the basic time-dependent Maxwell’s equations. Without loss of generality, the current density will be dropped from (1.4.2), because losses due to current flow that produces ohmic heating, can be included in later discussions of the complex dielectric constant formalism. First, the general wave equation will be derived for arbitrary time variations and then consider the case for harmonic time variation $\exp j \omega t$. Using the general form of the time-dependent equations, the curl operator is applied to both sides of (1.4.1),

$$\nabla \times (\nabla \times e) = -\frac{\partial}{\partial t} (\nabla \times b), \tag{1.10.1}$$

where the space and time derivatives have been interchanged. Using the relationship of $b$ to $h$, (1.8.5), the above equation becomes

$$\nabla \times (\nabla \times e) = -\mu_0 \frac{\partial}{\partial t} K_m(t) \ast (\nabla \times h),$$

$$= -\mu_0 \varepsilon_0 \frac{\partial^2}{\partial t^2} K_m(t) \ast K_e(t) \ast e(t). \tag{1.10.2}$$
where the constitutive relation in (\textit{ef. 51}) was used to tie $\mathbf{d}$ to $\mathbf{e}$. The time derivatives have been grouped because they commute with the convolution operators. The right-hand-side of (1.10.2) can be expanded as follows

$$\nabla \times (\nabla \times \mathbf{e}) = \nabla (\nabla \cdot \mathbf{e}) - \nabla \cdot \nabla \mathbf{e}.$$  \hspace{1cm} (1.10.3)

If the material is homogeneous, the convolution operators are independent of position so that in a charge-free region,

$$\nabla \cdot \mathbf{d} = 0 = \varepsilon_0 K_e(t) \ast [\nabla \cdot \mathbf{e}(t)].$$  \hspace{1cm} (1.10.4)

Applying the inverse operator $K_e^{-1}(t)$ to the above expression, gives $\nabla \cdot \mathbf{e}(t) = 0$ for all $t$ values. The final form of the time dependent wave equation is now given as

$$\nabla^2 \mathbf{e}(x, y, z; t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} K_m(t) \ast K_e(t) \ast \mathbf{e}(x, y, z; t) = 0,$$  \hspace{1cm} (1.10.5)

where $c = 1/\sqrt{\mu_0 \varepsilon_0}$ is the velocity of light in free space. A similar wave equation for the magnetic field $\mathbf{h}(x, y, z; t)$ can be obtained by simply exchanging $\mathbf{e}$ with $\mathbf{h}$ and interchanging the two convolution operators involving $K_m(t)$ and $K_e(t)$.

The wave equation for harmonic time variation at the frequency $\omega$, can be found by taking the Fourier transform of the time-dependent wave equation

$$\nabla^2 \mathbf{E}(x, y, z; \omega) + k_0^2 \kappa_m(\omega) \kappa_e(\omega) \mathbf{E}(x, y, z; \omega) = 0,$$  \hspace{1cm} (1.10.6)

where $\omega/c$ has been replaced with the free-space wave number, $k_0$. The Fourier transform of the double convolution of the three terms $K_m(t)$, $K_e(t)$ and $\mathbf{e}(x, y, z; t)$ in (1.10.5), is the product of their individual transforms.

The vector wave equation in (1.10.6) has three scalar equations and the corresponding wave equation for the magnetic field has three scalar equations. As a result, all electromagnetic field components satisfy the same differential equation so that it can be generalized. Further, to account for ohmic losses due to the conductivity and other losses in the polarization and magnetization vectors, magnetic permeability and the dielectric permittivity can be assumed to be complex quantities. Accordingly, we write the generalized differential equation as

$$\nabla^2 \psi(x, y, z; \omega) + k_0^2 \bar{\kappa}_m(\omega) \bar{\kappa}_e(\omega) \psi(x, y, z; \omega) = 0,$$  \hspace{1cm} (1.10.7)

where $\bar{\kappa}_m$ and $\bar{\kappa}_e$ are complex constitutive constants. From this point on, the tilde signs will be dropped and it will be understood that the constitutive constants are complex. Also, the frequency dependence of the field solutions as well as constitutive parameters will not be denoted. For lossy material the imaginary parts of $\kappa_e$ and $\kappa_m$ are negative while if the material has gain due to population inversion, the imaginary parts will be assumed as positive. Transparency point in a material is obtained with the optical gain is sufficient to off-set the normal internal losses.
A general solution of (1.10.7) can be obtained from the separation of variables technique with a typical solution of the form

\[ \psi(x, y, z; t) = e^{-j(k_x x + k_y y + k_z z)} e^{j\omega t} \]  

(1.10.8)

where \( k_x, k_y \) and \( k_z \) are the separation constants. (Equation (1.10.8) represents a plane wave propagating in the direction \( \mathbf{k} = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z \).) The constants are constrained by the equation obtained by substituting the solution, (1.10.8), into the differential equation given in (1.10.7),

\[ k_x^2 + k_y^2 + k_z^2 = k_0^2 \varepsilon \mu. \]  

(1.10.9)

Equation (1.10.9) shows that the four parameters or variables, \( k_x, k_y, k_z \) and \( k_0 \) (related to \( \omega \)) are not independent. A solution of the wave equation can be written as a linear combination of the functions given in (1.10.8) where each of the \( k \)s may be variable, but there can be only three independent variables. The value of \( k_0(\omega) \) is a variable when the signal is modulated with information and thus consists of many different frequencies. The following cases are:

1. Assuming the signal has multiple frequencies and three space dimensions, there are three independent variables, which may be a combination of \( \{k_x, k_y, k_z, k_0\} \). The choice of variables depends on the initial conditions of the fields. For example, if the initial value of the field is specified at \( t = 0 \) in space then the proper choice of variables is \( \{k_x, k_y, k_z\} \) so that the value of the initial field is

\[ \psi(x, y, z; 0) = \iiint_{-\infty}^{\infty} \Psi(k_x, k_y, k_z) e^{-j(k_x x + k_y y + k_z z)} d k_x d k_y d k_z, \]

where

\[ \Psi(k_x, k_y, k_z) = \frac{1}{8\pi^3} \iiint_{-\infty}^{\infty} \psi(x, y, z; 0) e^{j(k_x x + k_y y + k_z z)} d x d y d z, \]

is the 3-dimensional inverse Fourier transform of the initial field. With \( \Psi(k_x, k_y, k_z) \) specified, the solution may be written as

\[ \psi(x, y, z; t) = \iiint_{-\infty}^{\infty} \Psi(k_x, k_y, k_z) e^{-j(k_x x + k_y y + k_z z)} e^{j\omega t} d k_x d k_y d k_z, \]

where

\[ \omega = \frac{c}{\sqrt{\varepsilon \mu k_0}} \sqrt{k_x^2 + k_y^2 + k_z^2}. \]

If the initial value of the field in a plane, say \( z = 0 \) is specified as a function of time, then the proper choice of variables is \( \{k_x, k_y, k_0\} \).

2. If the field is harmonically varying, \textit{i.e.}, it has only one frequency component, then the general solution can be written as a linear combination of
two running variables, taken as a combination of \{k_x, k_y, k_z\}. The proper choice of the two variables depends on the initial conditions.

3. If the field has symmetry conditions or is independent of one space coordinate, say \(y\), and has only one frequency component, then the running variable is chosen from \(\{k_x, k_z\}\), where the proper choice depends on the initial conditions. When the field is independent of \(y\), \(k_y = 0\) and is therefore fixed.

The spatial argument of the exponential function in (1.10.8) may be written as \(\mathbf{k} \cdot \mathbf{r}\). A more general solution of the wave equation will be composed of a superposition of the individual waves defined by the direction of the \(\mathbf{k}\) vector. If the frequency is assumed to be fixed, then only two components of \(\mathbf{k}\) may vary, such as \(k_x\) and \(k_y\), while \(k_z\) is constrained by (1.10.8). The two variable solution in \(k_x\) and \(k_y\) is

\[
\psi(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(k_x, k_y) \exp(-j \mathbf{k} \cdot \mathbf{r}) dk_x dk_y, \tag{1.10.10}
\]

where \(\Psi(k_x, k_y)\) is the weighting function. The form of the above solution is a double Fourier integral in the \(x\) and \(y\) spatial dimensions. The weighting coefficient \(\Psi(k_x, k_y)\) can be determined from a specified field in the plane defined by a given \(z\) value. For example, if we assume the field is specified in the \(x\)-\(y\) plane at \(z = 0\) as \(\psi_0(x, y)\), then the weighting factor \(\Psi(k_x, k_y)\) is given by the double inverse Fourier transform,

\[
\Psi(k_x, k_y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_0(x, y) \exp(j(k_x x + k_y y)) dx dy. \tag{1.10.11}
\]

The calculation of the electromagnetic field at a large distance from the aperture can be easily found from (1.10.11). Typically, there are three regions associated with distances from the aperture: near-field region, Fresnel region, and Fraunhofer region (far-field). The Fraunhofer region is usually associated with the aperture size, \(D\), where \(D\) represents the rough aperture diameter, and it is given by a distance \(R > 4D^2/\lambda\). Assuming the aperture is at the center of a hemisphere of radius \(R\), the Cartesian coordinates can be transformed to spherical coordinates, \((x, y, z) \rightarrow (R, \theta, \phi)\). Assuming the material in the hemisphere is characterized by the wave number \(k_o\), then the argument of the exponential in (1.10.10) is

\[
\mathbf{k} \cdot \mathbf{r} = R(k_x \sin \theta \cos \phi + k_y \sin \theta \sin \phi + k_z \cos \theta).
\]

Upon normalizing the axial wave numbers with \(k_o\), as \(k_x = k_o u\), \(k_y = k_o v\) and \(k_z = k_o w\),

\[
\mathbf{k} \cdot \mathbf{r} = k_o R(\alpha u + \beta v + \gamma w) \equiv k_o R g(u, v),
\]

where \(w^2 = 1 - u^2 - v^2\), and \(\alpha = \sin \theta \cos \phi\), \(\beta = \sin \theta \sin \phi\), and \(\gamma = \cos \theta\). The Fraunhofer region is characterized by \(k_o R \gg 1\). An approximate evaluation of (1.10.10) can be obtained using the stationary-phase method as
discussed in Appendix A. Stationary points occur at $u_0 = \alpha$, $v_0 = \beta$, and $w_0 = \sqrt{1-u_0^2-v_0^2} = \gamma$. The second-order partial derivatives of $g(u, v)$ are

$$g_{uu}(u_0, v_0) = -(\alpha^2 + \gamma^2)/\gamma^2$$
$$g_{uv}(u_0, v_0) = -\alpha \beta/\gamma^2$$
$$g_{vv}(u_0, v_0) = -(\beta^2 + \gamma^2)/\gamma^2$$

Using the results of (A.8), the radical $g_{xx}g_{yy} - g_{xy}^2 = 1/\gamma = 1/\cos \theta$, the approximate value of (1.10.10) is

$$\psi(r, \theta, \phi) = j 4\pi \cos \theta \Psi(k_o \cos \phi \sin \theta, k_o \sin \phi \sin \theta) e^{-j k_o r} \frac{e^{-j k_o r}}{k_o r}.$$

Example 1.1: A circular aperture of radius $a$ is located in the $y$-$y$ plane at $z = 0$. The aperture is uniformly illuminated from the negative $z$ direction. As a first approximation, assume the field in the aperture, $x^2 + y^2 < a^2$ has a constant value while the field field outside the aperture is zero. The electric field is polarized along $x$, $E = E_0$, and the material in $z > 0$ is free space ($\epsilon_r, \mu_r = 1$). This implies that the initial value $\psi_0(x, y) = E_0$ in the aperture while $\psi_0(x, y) = 0$ outside the aperture. First, coordinates are transformed to

$$x = r \cos \phi,$$
$$y = r \sin \phi,$$

while $k_x$ and $k_y$ are transformed to

$$k_x = k_r \cos \phi,$$
$$k_y = k_r \sin \phi.$$

According to (1.10.10) the variables $k_x$ and $k_y$ cover the entire plane which implies the variable $k_r$ satisfies $0 < k_r < \infty$, while $\phi$ lies in the interval $(0, 2\pi)$ or any interval that spans a length of $2\pi$. Equation (1.10.11) becomes

$$\Psi(k_r, \phi) = \frac{1}{(2\pi)^2} \int_0^a \int_0^a E_0 \exp j[k_r r \cos(\phi - \phi)] r \, dr \, d\phi.$$

The integral is independent of $\phi$ because the circular integration can be taken over any period of $2\pi$. Therefore, take $\phi = 0$. The integral over $\phi$ yields a Bessel function of zero order,

$$\Psi(k_r) = \frac{1}{2\pi} \int_0^a r J_0(k_r r) \, dr,$$
$$= \frac{a^2}{2\pi} \frac{J_1(k_r a)}{k_r a},$$

where $J_1$ is a Bessel function of first order. Substituting the previous expression back into (1.10.10), the expression for the field in the region $z > 0$ becomes

$$\psi(x, y, z) = \frac{a^2}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{J_1(a \sqrt{k_x^2 + k_y^2})}{a \sqrt{k_x^2 + k_y^2}}$$
$$\exp -j(k_x x + k_y y + \sqrt{k_x^2 - k_y^2} z) \, dk_x \, dk_y$$
1.10 WAVE EQUATION

This integral cannot be evaluated in closed form, however, there are useful approximations that can be used for the far-field behavior, as given by (1.10.12). In this example, the field has no azimuthal dependence because the transform depends only on $k_x^2 + k_y^2 = k_0^2 \sin^2 \theta$. Thus, the electric field in the Fraunhofer region becomes

$$E(r, \theta) = j2 \hat{k} E_0 a^2 \cos \theta \left( \frac{J_1(k_0 \sin \theta)}{k_0 \sin \theta} \right) \frac{e^{-jk_0r}}{k_0r}$$

The unit vector $\hat{k}$ can be written in terms of the base vectors in spherical coordinates as $\hat{k} = \hat{r} \sin \theta \cos \phi + \hat{\theta} \cos \theta \cos \phi - \hat{\phi} \sin \phi$. In the far-field region the power is flowing the the radial direction so that only the unit vector components of $\hat{\theta}$ and $\hat{\phi}$ determine the radial flux.

Several points should be made regarding the solution given by (1.10.10). The solution is obtained by integrating over the two components of the $k$ vector, $k_x$ and $k_y$. The third component of $k$ is determined because the frequency $\omega$ is fixed. (If the frequency was a variable then there could be three independent variables in the Fourier transform.) Although the Fourier transform indicates integration for real values of $k_x$ and $k_y$, the integration could take a circuitous path in the complex $k_x$ and $k_y$ planes.

1.11 Field Analysis with Potential Functions

Instead of solving Maxwell’s equation directly, it is often convenient to transform the field quantities into auxiliary functions, a combination of vector and/or scalar functions. For waves that are guided by a metallic or dielectric structure along a single direction, the fields can be characterized in terms modes which are eigenfunctions of Maxwell’s equations. The modes are characterized by field shapes that are invariant, in the transverse direction, along the direction of propagation, say the $z$ direction. Typically, waves are characterized by a vector field such as the electric or magnetic field that has no component in the axial direction, the direction of propagation. When the mode has only an axial component of the magnetic field, it is called a transverse electric (TE) field, while if the mode has only an axial electric field, it is called a transverse magnetic (TM) field. If the mode requires both electric and magnetic axial components, it is called a hybrid mode. (Section 4.2 gives a more detailed explanation of wave definitions.) Although TE, TM and hybrid modes are typically used to classify the waves, there are of course many other possible wave definitions that can be defined via potential functions.

In this section, a more specific focus of the field equations will be addressed. Namely, Maxwell’s equations describing the electromagnetic fields in a homogeneous isotropic region with arbitrary time variation are

$$\nabla \cdot \mathbf{b} = 0, \quad (1.11.1)$$
$$\nabla \times \mathbf{e} = -\frac{\partial \mathbf{b}}{\partial t}, \quad (1.11.2)$$
$$\nabla \times \mathbf{h} = \frac{\partial \mathbf{d}}{\partial t} + \mathbf{j}, \quad (1.11.3)$$
$$\nabla \cdot \mathbf{d} = \rho, \quad (1.11.4)$$
1.11.1 Magnetic Vector Potential

Since the \( \nabla \cdot \mathbf{b} = 0 \), the magnetic field can be expressed in terms of another vector field, \( \mathbf{a} \), called the magnetic vector potential. The vector field, \( \mathbf{a} \), may be separated into a solenoidal part, \( \mathbf{a}_s \), whose divergence is zero, \( \nabla \cdot \mathbf{a}_s = 0 \), and an irrotational part, \( \mathbf{a}_i \), whose curl is zero, \( \nabla \times \mathbf{a}_i = 0 \). Thus,

\[
\mathbf{b} = \nabla \times \mathbf{a} = \nabla \times (\mathbf{a}_s + \mathbf{a}_i) = \nabla \times \mathbf{a}_s.
\]

so that (1.11.5) does not designate \( \mathbf{a}_i \) from \( \mathbf{b} \); only its solenoidal part is specified. Substituting (1.11.5) into (1.11.2) gives

\[
\nabla \times (\mathbf{e} + \frac{\partial \mathbf{a}}{\partial t}) = 0,
\]

which implies the term in parenthesis can be written as the gradient of a scalar since the curl of the gradient of an arbitrary scalar is the null vector, giving

\[
\mathbf{e} = -\frac{\partial \mathbf{a}}{\partial t} - \nabla \phi.
\]

The two remaining equations, (1.11.3) and (1.11.4) can be used to find the coupled differential equations for the vector potential \( \mathbf{a} \) and scalar potential \( \phi \). Substituting \( \mathbf{e} \) and \( \mathbf{h} \) into (1.11.3),

\[
\nabla^2 \mathbf{a} - \mu \epsilon \frac{\partial^2 \mathbf{a}}{\partial t^2} + \mu \mathbf{j} = \nabla (\nabla \cdot \mathbf{a} + \mu \epsilon \frac{\partial \phi}{\partial t}),
\]

and into (1.11.4),

\[
\nabla^2 \phi - \mu \epsilon \frac{\partial^2 \phi}{\partial t^2} + \frac{\rho}{\epsilon} = -\frac{\partial}{\partial t} (\nabla \cdot \mathbf{a} + \mu \epsilon \frac{\partial \phi}{\partial t}).
\]

The two coupled equations may now be solved, note however, the irrotational part of the the vector potential \( \mathbf{a} \) has not been specified. Assuming the electromagnetic field quantities \( \mathbf{e} \), \( \mathbf{h} \), \( \mathbf{b} \), and \( \mathbf{d} \), are measurable field quantities (so-called observables) then introducing \( \mathbf{a} \) and \( \phi \) is simply a mathematical excursion. There are two common methods (Lorentz and Coulomb Gauges) that tie down the definition of \( \mathbf{a} \), but nevertheless, in each method \( \mathbf{a} \) and \( \phi \) are not unique and the transformation between different values, say, \( \mathbf{a}_1 \rightarrow \mathbf{a}_2 \) and \( \phi_1 \rightarrow \phi_2 \) is called a gauge transformation, but, both the pair \( \{ \mathbf{a}_1, \phi_1 \} \) and \( \{ \mathbf{a}_2, \phi_2 \} \) yield identical electromagnetic field quantities as is the case for the different gauges.

**Lorentz Gauge**

There are many ways to specify the irrotational part of the vector potential, however, one approach to solving for \( \mathbf{a} \) and \( \phi \) is to specify the irrotational part of \( \mathbf{a} \) that uncouples the differential equations by placing

\[
\nabla \cdot \mathbf{a} + \mu \epsilon \frac{\partial \phi}{\partial t} = 0,
\]

(1.11.10)
so that the irrotational part of \( a (\nabla \cdot a = \nabla \cdot a_i) \) is now defined in terms of the scalar potential \( \phi \) and the two coupled equations become

\[
\nabla^2 a - \mu \varepsilon \frac{\partial^2 a}{\partial t^2} = -\mu j, \tag{1.11.11}
\]

\[
\nabla^2 \phi - \mu \varepsilon \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\varepsilon}. \tag{1.11.12}
\]

The specification of the irrotational part of \( a \) in (1.11.10) is known as the Lorentz condition. The solutions for (1.11.11) and (1.11.12) are the retarded potentials

\[
a(r, t) = \frac{\mu}{4\pi} \int \int \int \frac{[j(r', t')]}{R} dV, \tag{1.11.13}
\]

\[
\phi(r, t) = \frac{1}{4\pi \varepsilon} \int \int \int \frac{[\rho(r', t')]}{R} dV, \tag{1.11.14}
\]

where the retarded time, \( t' \), must be replaced by \( t - R/v; r = \hat{x} x + \hat{y} y + \hat{z} z \) is the vector from the origin to the observation point while \( r' = \hat{x} \xi + \hat{y} \eta + \hat{z} \zeta \) is the vector from the origin to the source point. \( R = |r - r'|, v = 1/\sqrt{\mu \varepsilon} \) is the propagation velocity, and \( dV = d\xi \, d\eta \, d\zeta \).

Although the electromagnetic fields, \( e \) and \( h \), are unique for given current and charge densities, it should be noted that the Lorentz condition still does not uniquely specify the vector, \( a \), and scalar, \( \phi \), potentials for if \( a \) and \( \phi \) are replaced by

\[
a \rightarrow a' - \nabla \varphi, \tag{1.11.15}
\]

\[
\phi \rightarrow \phi' + \frac{\partial \varphi}{\partial t}, \tag{1.11.16}
\]

the new potentials satisfy the same equations as \( a \) and \( \phi \), i.e.,

\[
\nabla^2 a' - \mu \varepsilon \frac{\partial^2 a'}{\partial t^2} = -\mu j, \tag{1.11.17}
\]

\[
\nabla^2 \phi' - \mu \varepsilon \frac{\partial^2 \phi'}{\partial t^2} = -\frac{\rho}{\varepsilon}. \tag{1.11.18}
\]

The Lorentz condition for \( a' \) and \( \phi' \), will be satisfied provided

\[
\nabla^2 \varphi - \mu \varepsilon \frac{\partial^2 \varphi}{\partial t^2} = 0.
\]

Example 1.2: Calculate the vector and scalar potentials for an applied current density given by

\[
j(r, t) = 2 \int_0^\delta (x) \delta(y) \delta(z) \left[ t [u(t) - u(t - \tau)] + (2\tau - t) [u(t - \tau) - u(t - 2\tau)] \right].
\]

\[
= 2 \int_0^\delta (x) \delta(y) \delta(z) \begin{cases} 
0, & t < 0, \\
\tau, & 0 \leq t \leq \tau, \\
2\tau - t, & \tau < t \leq 2\tau, \\
0, & t > 2\tau,
\end{cases}
\]
where \( u(t) \) is unit step function at \( t = 0 \). The current density is driven by an external source that produces a triangular pulse that “turns on” at \( t = 0 \) and “turns off” at \( 2\tau \). In this example, the Laplace transform pair is useful for finding appropriate field quantities. Namely,

\[
J(r,s) = \mathcal{L}[j(r,t)],
\]

so that

\[
J(r,s) = \int_0^\infty j(r,t) e^{-st} \, dt = \hat{\delta} J_o \delta(y) \delta(z) \frac{1}{s^2} (1 - 2e^{-st} + e^{-2st}),
\]

and the inverse

\[
j(r,t) = \mathcal{L}^{-1}[J(r,s)] = \frac{1}{2\pi i} \int_{Br} J(r,s) e^{st} \, ds,
\]

where \( Br \) represents the Bromwich integral. The Laplace transform of the continuity equation gives

\[
s \tilde{\rho}(r,s) - \rho(r,0) + \nabla \cdot J(r,s) = 0
\]

where \( \tilde{\rho}(r,s) = \mathcal{L}[\rho(r,t)] \). Assuming the initial condition \( \rho(r,0) = 0 \),

\[
\rho(r,t) = \mathcal{L}^{-1}[\tilde{\rho}(r,s)] = -J_o \delta(x) \delta(y) \delta_z \left( \frac{1}{2\pi i} \int_{Br} \frac{1}{s^2} (1 - 2e^{-st} + e^{-2st}) \right) e^{st} \, ds,
\]

\[
= \frac{J_o}{2} \delta(x) \delta(y) \delta_z \left[ t^2 u(t) - 2(t - \tau)^2 u(t - \tau) + (t - 2\tau)^2 u(t - 2\tau) \right],
\]

where \( \delta_z = \partial \delta(z) / \partial z \). As a side point, note that the source has a dipole moment

\[
p = \int \int \int \rho(r,t) \, dx \, dy \, dz = \int \int \int (\hat{k} x + \hat{j} y + \hat{z} z) \rho(r,t) \, dx \, dy \, dz
\]

\[
= \hat{z} \frac{J_o}{2} \left[ t^2 u(t) - 2(t - \tau)^2 u(t - \tau) + (t - 2\tau)^2 u(t - 2\tau) \right].
\]

The retarded potentials can now be computed from the inverse transforms of \( J(r,s) \) and \( \tilde{\rho}(r,s) \). The retarded term of the inverse Laplace transform \( \exp(st) \to \exp(s(t - R/v)) \exp(st) \), so that

\[
a(r,t) = \frac{\mu J_o}{4\pi} \hat{z} \frac{1}{r} \frac{1}{2\pi i} \int_{Br} \left[ \frac{1}{s^2} (1 - 2e^{-st} + e^{-2st}) e^{-sr/v} \right] e^{st} \, ds,
\]

\[
= \frac{\mu J_o}{4\pi} \hat{z} \frac{1}{r} \mathcal{L}^{-1} \left[ \frac{1}{s^2} (1 - 2e^{-st} + e^{-2st}) e^{-sr/v} \right]
\]

\[
= \frac{\mu J_o}{4\pi} \hat{z} \frac{1}{r} \left[ (t - r/v) u(t - r/v) - 2(t - r/v - \tau) u(t - r/v - \tau) + (t - r/v - 2\tau) u(t - r/v - 2\tau) \right],
\]

where \( r = |r| \). In spherical coordinates the unit vector \( \hat{z} = \hat{r} \cos \theta - \hat{\theta} \sin \theta \). In time \( a(r,t) \) is a single triangular pulse that starts from zero value at \( t = r/v \), reaches a peak at \( t = r/v + \tau \), and decreases linearly to zero at \( t = 2\tau \). (The magnetic field, \( b = \nabla \times a \), has only a \( \hat{\phi} \) component but it has 2 terms: one that decreases as \( 1/r^2 \).
1.11 FIELD ANALYSIS WITH POTENTIAL FUNCTIONS

and another that decreases as \(1/r\). Because the integrand in (1.11.14) contains \(\delta_\zeta(\zeta)\), the scalar potential requires the computation

\[
\frac{\partial}{\partial \zeta} \left( \frac{e^{-iR/v}}{R} \right)_{\zeta=0} = (1 + sr/v) \frac{z}{r^3} e^{-sr/v},
\]

so that,

\[
\phi(r,t) = \frac{J_0}{4 \pi \varepsilon} \frac{z}{R^3} \frac{1}{2 \pi j} \int_{Br} \left[ \frac{1}{\delta^3} \left( 1 - 2 e^{-2\tau} + e^{2\tau} \right)(1 + rs/v) \right] e^{i(t-r/v)} \, ds,
\]

\[
= \phi_N(r, t) + \phi_F(r, t),
\]

where \(\phi_N(r, t)\) represents the near field while \(\phi_F(r, t)\) represents the far field. In spherical coordinates, \(z = r \cos \theta\), so that

\[
\phi_N(r, t) = \frac{J_0}{4 \pi \varepsilon} \frac{\cos \theta}{r^3} \frac{1}{2 \pi j} \int_{Br} \left[ \frac{1}{\delta^3} \left( 1 - 2 e^{-2\tau} + e^{2\tau} \right) \right] e^{i(t-r/v)} \, ds,
\]

\[
= \frac{J_0}{8 \pi \varepsilon} \frac{\cos \theta}{r^2} \left[ (t-r/v)^2 u(t-r/v) - 2(t-r/v-\tau)^2 u(t-r/v-\tau) \right.
\]

\[
+ (t-r/v-2\tau)^2 u(t-r/v-2\tau) \bigg],
\]

\[
\phi_F(r, t) = \frac{J_0}{4 \pi \varepsilon v} \frac{\cos \theta}{r^2} \frac{1}{2 \pi j} \int_{Br} \left[ \frac{1}{\delta^3} \left( 1 - 2 e^{-2\tau} + e^{2\tau} \right) \right] e^{i(t-r/v)} \, ds,
\]

\[
= \frac{J_0}{4 \pi \varepsilon v} \frac{\cos \theta}{r} \left[ (t-r/v)^2 u(t-r/v) - 2(t-r/v-\tau)^2 u(t-r/v-\tau) \right.
\]

\[
+ (t-r/v-2\tau)^2 u(t-r/v-2\tau) \bigg].
\]

The electric field can now be computed using (1.11.7).

In a source-free region the field quantities may be simplified. For harmonic time variation, \(\exp(j\omega t)\), the differential equation for the vector potential, \(\mathbf{a} = \mathbf{A}(x, y, z) \exp(j\omega t)\), is

\[
\nabla^2 \mathbf{A} + k^2 \mathbf{A} = 0, \quad (1.11.19)
\]

where the wavenumber \(k^2 = \omega^2 \mu \varepsilon\), and the dielectric constant \(\varepsilon = \varepsilon_0 \varepsilon\). Harmonic time varying fields can be determined from only the vector potential \(\mathbf{A}\) using (1.11.5) and (1.11.3),

\[
\mathbf{H} = \frac{1}{j\omega \mu} \nabla \times \mathbf{A}, \quad (1.11.20)
\]

\[
\mathbf{E} = \frac{1}{j\omega \varepsilon} \nabla \times \mathbf{H}
\]

\[
= \frac{1}{j\omega \mu \varepsilon} \nabla \times \nabla \times \mathbf{A}. \quad (1.11.21)
\]

Example 1.3: Assume a waveguide structure has a field that represents a mode propagating in the positive \(z\) direction according to \(\exp(j\omega t - \gamma_m z)\). The fields have
a transverse variation on both the $x$ and $y$ directions. Assume the vector potential has only a $z$ component and may be written as

$$ A = i A_m \psi_m(x, y) e^{-vmz}. $$

Since $A$ satisfies (1.11.19), $\psi_m(x, y)$ is governed by

$$ \nabla_{xy}^2 \psi_m + (\gamma_m^2 + \vec{k}^2) \psi_m = 0, $$

and it is normalized according to

$$ \iint \psi_m^* \psi_m \, dx \, dy = 1. $$

The double integral is over the area of the waveguide. The magnetic field $H_z = 0$, while the transverse components satisfy

$$ H_x = \frac{A_m}{\mu} \frac{\partial \psi_m}{\partial y}, \quad H_y = -\frac{A_m}{\mu} \frac{\partial \psi_m}{\partial x}. $$

The electric field becomes

$$ E_x = \frac{\gamma_m A_m}{j\omega \varepsilon} \frac{\partial \psi_m}{\partial x}, \quad E_y = \frac{\gamma_m A_m}{j\omega \varepsilon} \frac{\partial \psi_m}{\partial y}, \quad E_z = \frac{\gamma_m^2 + \vec{k}^2}{j\omega \varepsilon} A_m \psi_m $$

Note that $E_z$ and $A$ have a common transverse dependence, i.e.,

$$ A_m \psi_m(x, y) = \frac{j\omega \mu \varepsilon}{\gamma^2 + \vec{k}^2} E_z(x, y), $$

and therefore, boundary conditions on $\psi_m$ are identical to those on $E_z$.

### Coulomb Gauge

There are an infinite number of conditions on the vector and scalar potentials that may define a gauge that simplifies the computation of the electric and magnetic fields [13, 18, 24]. The Lorentz gauge was defined by (1.11.10) which specifies the irrotational part of the vector potential in terms of the scalar potential which produces a symmetry of the differential equations, (1.11.11) and (1.11.12). Still, the vector and scalar potentials are not unique and may be transformed according to (1.11.15) and (1.11.16).

The coulomb gauge is obtained by placing

$$ \nabla \cdot \mathbf{a} = 0, \quad (1.11.22) $$

so that the vector potential is only solenoidal, i.e., $\nabla \cdot a = 0$, similar to the magnetic flux density. According to (1.11.9), the scalar potential must satisfy

$$ \nabla^2 \phi = \frac{\rho}{\varepsilon}, \quad (1.11.23) $$
which shows no propagation effects, i.e., $\phi$ instantaneously varies the same at each point in space. The vector potential in this gauge satisfies

$$\nabla^2 a - \mu \frac{\partial^2 a}{\partial t^2} = -\mu \left( j - \epsilon \nabla \frac{\partial \phi}{\partial t} \right).$$  \hspace{1cm} (1.11.24)

First the solution of (1.11.23) is obtained (The solution of the potential function is identical to that of a static potential where the time variation mirrors the charge density.) so that the right-hand side of (1.11.24) represents the “source” of the vector potential. By separating the current density into its solenoidal and irrotational parts, the continuity equation becomes

$$\nabla \cdot \mathbf{J}_i + \frac{\partial \rho}{\partial t} = 0,$$  \hspace{1cm} (1.11.25)

so that (1.11.23) reduces to

$$\nabla \cdot \left( \epsilon \nabla \frac{\partial \phi}{\partial t} - \mathbf{J}_i \right) = 0$$  \hspace{1cm} (1.11.26)

### 1.11.2 Electric Vector Potential

The solution of the field equations in a charge-free region may also be obtained using an approach similar to that given in Section 1.11.1, using the fact that $\nabla \cdot \mathbf{d} = 0$. Thus, the electric flux density may be written as

$$\mathbf{e} = -\nabla \times \mathbf{a}_e.$$  \hspace{1cm} (1.11.27)

The quantity $\mathbf{a}_e$ is called the electric vector potential and satisfies the differential equation

$$\nabla^2 \mathbf{a}_e - \mu \frac{\partial \mathbf{a}_e}{\partial t} = \mu \frac{\partial^2 \mathbf{a}_e}{\partial t^2} = 0.$$  \hspace{1cm} (1.11.28)

Note that both $\mathbf{a}$ and $\mathbf{a}_e$ satisfy the same differential equation however, they may have different boundary conditions at the different interfaces.

For harmonic time variation, the electric vector potential $\mathbf{a}_e$, for propagating waves in the $z$ direction, have the form $\mathbf{a}_e = \mathbf{A}_e(x, y) \exp(j \omega t - yz)$, and the electromagnetic fields may be written as

$$\mathbf{E} = -\frac{1}{\varepsilon} \nabla \times \mathbf{A}_e,$$  \hspace{1cm} (1.11.29)

$$\mathbf{H} = \frac{1}{j \omega \varepsilon \mu} \nabla \times \nabla \times \mathbf{A}_e.$$  \hspace{1cm} (1.11.30)

### 1.11.3 Electromagnetic Fields in Waveguides

The electromagnetic fields may be determined from the magnetic and electric vector potentials that have only one component. In particular, if the component is chosen along the $z$ direction, the propagation direction, the magnetic field due to the magnetic vector potential, $\mathbf{A}$, has no $z$ component; the field
is transverse magnetic (TM). Similarly, if the electric vector potential, $A_e$ has only a $z$ component, the electric field is transverse electric (TE). The electromagnetic fields listed in Table 1.1, show the fields in terms of the magnetic and electric vector potentials. The propagation constants of the fields are $\gamma_m$ and $\gamma_e$ for the two types of polarization.

Table 1.1. Components of the electric and magnetic fields in terms of vector potentials are listed below. The constants $A_m$ and $A_e$ represent the amplitude of the magnetic and electric vector potentials respectively. Because both $\psi_m$ and $\psi_e$ are assumed to be normalized over the cross section of the waveguide as illustrated in Example 1.3, they have dimensions of $\text{Imeter}$ in MKS units. Accordingly, $A_m$ has dimensions of $\text{Henry-Ampere}$, while $A_e$ has dimensions $\text{Farad-Volt}$.

<table>
<thead>
<tr>
<th></th>
<th>$A = \hat{z}A_m\psi_m(x, y)e^{-\gamma_m z}$</th>
<th>$A_e = \hat{z}A_e\psi_e(x, y)e^{-\gamma_e z}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_x$</td>
<td>$-\frac{\gamma_m A_m}{j\omega \mu \epsilon} \frac{\partial \psi_m}{\partial x}$</td>
<td>$-\frac{A_e}{\epsilon} \frac{\partial \psi_e}{\partial y}$</td>
</tr>
<tr>
<td>$E_y$</td>
<td>$-\frac{\gamma_m A_m}{j\omega \mu \epsilon} \frac{\partial \psi_m}{\partial y}$</td>
<td>$\frac{A_e}{\epsilon} \frac{\partial \psi_e}{\partial x}$</td>
</tr>
<tr>
<td>$E_z$</td>
<td>$\frac{\gamma_e^2 + k^2}{j\omega \mu \epsilon} A_m \psi_m$</td>
<td></td>
</tr>
<tr>
<td>$H_x$</td>
<td>$\frac{A_m}{\mu} \frac{\partial \psi_m}{\partial y}$</td>
<td>$-\frac{\gamma_e A_e}{j\omega \epsilon \mu} \frac{\partial \psi_e}{\partial x}$</td>
</tr>
<tr>
<td>$H_y$</td>
<td>$-\frac{A_m}{\mu} \frac{\partial \psi_m}{\partial x}$</td>
<td>$-\frac{\gamma_e A_e}{j\omega \epsilon \mu} \frac{\partial \psi_e}{\partial y}$</td>
</tr>
<tr>
<td>$H_z$</td>
<td>$\frac{\gamma_e^2 + k^2}{j\omega \epsilon \mu} A_e \psi_e$</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2
Static and Quasi-Static Fields

Many properties of harmonically time-varying electromagnetic fields can be deduced from characteristics of static fields provided time variations are slow, implying that the wavelength of the fields is large compared to the dimensions of the structure under consideration. This condition is satisfied if a metallic dielectric sphere of radius $a$ is placed in a time varying field with $a \ll \lambda$. In terms of the vector and scalar potential functions, the static analysis of charge and current densities is applicable when the retardation time $t - R/v$ is replaced by $t$, when the propagation time $R/v$ is small compared to the time period $T = 2\pi/\omega$.

2.1 Introduction
Effective analyses of many integrated optical and nano-structured devices can be accomplished with static and quasi-static field analysis. The development of integrated isolators is dependent on the inclusion of iron or similar magnetically sensitive materials such as nickel or cobalt in semiconductors. The dielectric characteristics of material with a mix of a different material that is usually “clustered” is governed by the Maxwell-Garnett mixing formula.[7] While the semiconductor material is isotropic the magnetic material is anisotropic so that mixture needs to exhibit anisotropy. This chapter discusses a new analysis that accounts for the macroscopic nature of a semiconductor/magnetic material mixture.

Section 2.2 calculates the electric field in a single isotropic dielectric sphere embedded in an isotropic medium. This analysis is typically found many places.[30],[11].
Section 2.3 calculates the electric field for an arbitrary polarization that may occur in typical dielectric waveguides and Section 2.4 develops the Maxwell-Garnett mixing formula for isotropic materials. Section 2.5 develops the Maxwell-Garnett mixing formula for anisotropic material in an isotropic medium. Finally, the losses in such a mixture are derived in Section 2.6.

2.2 Embedded Dielectric Sphere

An embedded dielectric sphere with a dielectric constant \( \varepsilon_1 = \varepsilon_0 \kappa_1 \) and a radius \( a \) will be placed in a medium with a dielectric constant \( \varepsilon_2 = \varepsilon_0 \kappa_2 \) as illustrated in Fig. 2.1. The solution of the electric field is obtained by solving the Laplace equation for the electrostatic potential \( \phi \) in spherical coordinates. \([11],[30]\). Laplace’s equation in spherical coordinates is

\[
\nabla^2 \phi = \frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \phi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \phi}{\partial \phi^2} = 0, \tag{2.2.1}
\]

and can be solved using the separation of variables technique, which has a general solution

\[
\phi(r, \theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} P_n^{|m|}(|\cos \theta|) e^{im\varphi} \left\{ \begin{array}{ll}
A_{mn} r^n + B_{mn} r^{n+1}, & r > a, \\
C_{mn} r^n + D_{mn} r^{n+1}, & r < a,
\end{array} \right. \tag{2.2.2}
\]

where the associated Legendre functions are given by

\[
P_n^l(x) = (1 - x^2)^{l/2} \frac{d^l P_n(x)}{dx^l}, \tag{2.2.3}
\]

and \( P_n(x) \) are Legendre polynomials. Since \( n \) represents the maximum order of the polynomials, \( d^l P_n(x)/dx^l = 0 \) for \( l > n \). The coefficients \( A_{mn}, B_{mn}, C_{mn}, \) and \( D_{mn} \) are determined from the boundary conditions. First, note that the general solution will grow without bound as \( r \to \infty \) if the coefficients \( A_{mn} \) are not zero; similarly, the solution will have a singularity at \( r = 0 \) if the \( D_{mn} \) coefficients are not zero. To avoid unlimited field growth as \( r \to 0 \) and \( r \to \infty \), the general solution is

\[
\phi(r, \theta, \varphi) = \left\{ \begin{array}{ll}
\phi^+(r, \theta, \varphi), & r > a; \text{ where } A_{mn} = 0, \\
\phi^-(r, \theta, \varphi), & r < a; \text{ where } D_{mn} = 0.
\end{array} \right. \tag{2.2.4}
\]

(When an applied field is included in the potential function (2.2.4), the condition \( A_{mn} = 0 \) must be modified accordingly.) The normal and tangential electric fields determined from \( \phi^+ \) and \( \phi^- \) are matched at the boundary \( r = a \).

In the present framework a static electric field is applied to the mixture with the embedded sphere located at the origin. The embedded sphere will induce a dipole moment that affects the static field so that there are 2 fields: (1) the applied static field and (2) the induced field produced by the dipole moment of the sphere. Both the applied and induced fields satisfy Laplace’s equation and thus the total field satisfies Laplace’s equation and \( \phi \) represents the total solution.
2.2 EMBEDDED DIELECTRIC SPHERE

The applied field is accounted for by applying a condition at distances $r \gg a$ or as $r \rightarrow \infty$. For an applied field directed along the $z$ axis, the potential function will have azimuthal symmetry so the series of $m$ values has only 1 term, $m = 0$. The associated Legendre functions for $m = 0$ are $P_n^0(\cos \theta) = P_n(\cos \theta)$ are the Legendre polynomials. As $r \rightarrow \infty$ the potential $\phi^+ \rightarrow -E_0 z = -E_0 r \cos \theta$, which produces a static electric field $E_0 = -\nabla \phi^+ = z E_0$, where $E_0$ is the static field strength. The coefficients $A_n \equiv A_{0n}$ have $A_n = 0$, for all $n$ save the $n = 1$ term, where $A_1 = -E_0$ and $P_1(\cos \theta) = \cos \theta$. It is interesting to note that the first term represents the applied potential field while the second term with $B_n \equiv B_{0n}$ coefficients represents the induced field.

All $A_n$ coefficients for $r > a$ and all $D_n \equiv D_{0n}$ coefficients for $r < a$ are determined and the solution becomes

$$
\phi(r, \theta, \varphi) = \phi(r, \theta) = \sum_{n=0}^{\infty} P_n(\cos \theta) \begin{cases} 
E_0 \delta_{1n} r^n + B_n \frac{1}{r^{n+1}}, & r > a \\
C_n r^n, & r < a,
\end{cases}
$$

(2.2.5)

where $C_n \equiv C_{0n}$. The electric field calculated from $\phi$ is

$$
E = -\mathbf{r} \frac{\partial \phi}{\partial r} - \hat{\theta} \frac{1}{r} \frac{\partial \phi}{\partial \theta}.
$$

The radial component becomes

$$
E_r = \sum_{n=0}^{\infty} P_n(\cos \theta) \begin{cases} 
E_0 \delta_{1n} r^{n+1} + (n + 1) B_n / r^{n+2}, & r > a \\
C_n r^{n-1}, & r < a,
\end{cases}
$$

(2.2.6)

whereas the angular component is

$$
E_\theta = \sum_{n=0}^{\infty} \sin \theta P_n'(\cos \theta) \begin{cases} 
E_0 \delta_{1n} r^{n+1} + B_n / r^{n+2}, & r > a \\
C_n r^{n-1}, & r < a.
\end{cases}
$$

(2.2.7)

$P_n'$ is the derivative of $P_n$ with respect to its argument. $\delta_{1n}$ is the Kronecker delta. Matching the normal and tangential components of $D_r$ and $E_\theta$ at the surface of the sphere gives

$$
\kappa_2 \left( E_0 P_1(\cos \theta) + \sum_{n=0}^{\infty} P_n(\cos \theta) \left[ B_n (n + 1)/a^{n+2} \right] \right) = \kappa_1 \left( \sum_{n=0}^{\infty} P_n(\cos \theta) C_n a^{n-1} \right)
$$

$$
E_0 \sin \theta P_1'(\cos \theta) + \sum_{n=0}^{\infty} \sin \theta P_n'(\cos \theta) B_n / a^{n+2} = \sum_{n=0}^{\infty} \sin \theta P_n'(\cos \theta) C_n a^{n-1}
$$

The Legendre polynomials form an orthogonal set over a range $(-1, 1)$ of its argument so that multiplying the first equation by $\sin \theta P_k(\cos \theta)$ and integrating on $\theta$ from $0$ to $\pi$ gives

$$
\kappa_2 E_0 + 2\kappa_2 B_1 / a^3 + \kappa_1 C_1 = 0, \quad k = 1,
$$

(2.2.8)

$$
(k + 1)\kappa_2 B_k / a^{k+2} + k\kappa_1 C_k a^{k-1} = 0, \quad k = 0, 2, 3, \ldots
$$

(2.2.9)
Determining the coefficients in the equations obtained from matching the tangential components of $E_0$ requires some what more work because of the derivatives of the Legendre polynomials. By multiplying (2.2.9) by $P_{k+1} - P_{k-1}$ and integrating by parts (to eliminate the derivatives) and using the resulting orthogonality properties of Legendre polynomials gives

$$E_0 - B_1/a^3 + C_1 = 0, \quad k = 1, \quad (2.2.10)$$
$$B_k/a^{k+2} - C_k a^{k-1} = 0, \quad k = 0, 2, 3, \ldots. \quad (2.2.11)$$

The potentials $\phi$ must be continuous across the boundary at $r = a^+$ and $r = a^-$. Conditions on $B_k$ and $C_k$ for $k = 0, 2, 3, \ldots$, can be satisfied only if $C_k = B_k = 0$, while the $k = 1$ term yields

$$B_1 = \frac{\kappa_1 - \kappa_2}{2\kappa_2 + \kappa_1} a^3 E_0, \quad (2.2.12)$$
$$C_1 = -3 \frac{\kappa_2}{2\kappa_2 + \kappa_1} E_0. \quad (2.2.13)$$

The resulting potential outside the sphere is

$$\phi^+(r, \theta) = -E_0 r \cos \theta - \frac{\kappa_2 - \kappa_1}{2\kappa_2 + \kappa_1} a^3 E_0 \frac{\cos \theta}{r^2}. \quad (2.2.14)$$

(The polynomial $P_1(\cos \theta) = \cos \theta$.)

Figure 2.1. The electric field produced by an embedded dielectric sphere of radius $a$ and a dielectric constant $\varepsilon_1 = \varepsilon_0 \kappa_1$, in a dielectric medium, $\varepsilon_2 = \varepsilon_0 \kappa_2$.

Interpretation of (2.2.14) is important as it leads to understanding polarizability of an atom or a group of atoms (such as a sphere) in an environment. Even though the sphere may occupy a large space, behavior of the induced potential as seen in the exterior region ($r > a$) has the form of a single electric dipole whose moment is

$$\rho = \hat{z} 4\pi \varepsilon_0 \frac{\kappa_2(\kappa_1 - \kappa_2)}{2\kappa_2 + \kappa_1} a^3 E_0 = \hat{z} \frac{1}{3} \varepsilon_0 \frac{\kappa_2(\kappa_1 - \kappa_2)}{2\kappa_2 + \kappa_1} V_1 E_0. \quad (2.2.15)$$

where $V_1$ is the volume of the sphere. Figure 2.1 shows the dipole moment whose polarization is along the positive $z$ direction which occurs when $\kappa_1$ is larger than $\kappa_2$, i.e., there is a net positive charge at the right side of the sphere and corresponding net negative charge at the left side. This condition occurs
2.2 EMBEDDED DIELECTRIC SPHERE

when the surface polarization density \( \rho_{sp} \) is larger for the \( \kappa_1 \) material. At the right side of the sphere there are more positive charges in the exterior region of the sphere than negative charges in the interior region.

Most of the interest of the present configuration concerns the field external to the sphere, however, it is interesting to note that the internal field has a simple form

\[
\phi^{-}(z) = -3 \frac{\kappa_2}{2 \kappa_2 + \kappa_1} E_0 z, \tag{2.2.16}
\]

which represents a constant electric field directed in the positive \( z \) direction, whose value is \( E_z = 3\kappa_2 E_0 / (2 \kappa_2 + \kappa_1) = 3E_0 / (2 + \kappa_1 \kappa_2) \). When \( \kappa_1 > \kappa_2 \) the internal field is smaller than the applied field which implies an induced field, directed from right to left (meaning a net positive charge on the right side of the sphere and corresponding net negative charge on the left side) that weakens the applied electric field. Conversely, when \( \kappa_1 < \kappa_2 \), the internal field is larger than the applied field and the induced field strengthens the applied field.

The static electric field calculated from the electrostatic potential is

\[
E = E_0 \begin{cases} 
\hat{z} + \frac{\kappa_1 - \kappa_2}{2 \kappa_2 + \kappa_1} (\hat{r} 2 \cos \theta + \hat{\theta} \sin \theta) \frac{a^3}{r^3}; & r > a, \\
\hat{z} \frac{3 \kappa_2}{2 \kappa_2 + \kappa_1}; & r < a.
\end{cases} \tag{2.2.17}
\]

\( E_0 \) is the field strength at large distances from the sphere or it can be regarded as the total field in the crystal in the absence of a sphere, \( a = 0 \).

2.3 Arbitrary Polarization Direction

When the electric field is polarized in an arbitrary direction, say \((\theta_0, \varphi_0)\) as illustrated in Fig. 2.2, the solution must exhibit azimuthal variation. The electrostatic potential that produces the electric field

\[
E = \hat{x} E_1 + \hat{y} E_2 + \hat{z} E_3 = E_0 (\hat{x} \sin \theta_0 \cos \varphi_0 + \hat{y} \sin \theta_0 \sin \varphi_0 + \hat{z} \cos \varphi_0) \tag{2.3.1}
\]

is given by

\[
\phi = -(E_1 x + E_2 y + E_3 z). \tag{2.3.2}
\]

where \( E_0 \) is the field strength.

---

Figure 2.2. The spherical coordinate system with the electric field polarized along the \( \theta_0 \) and \( \varphi_0 \) direction.
At an arbitrary observation point defined by \((r, \theta, \varphi)\) the field will be a function of the angular variables and the field polarization direction. When the electric field is polarized along \(z\), it is symmetric about the \(z\) axis; thus, it depends only on the angular variable \(\theta\), the angle between the \(z\) axis and the observation point. Similarly, in Fig. 2.2, the field should be dependent only on the angular variable \(\gamma\), i.e., it is symmetric about the vector \(E_0\). The electrostatic potentials inside and outside the sphere for polarization along the \(z\) axis are given by (2.2.14) and (2.2.16) where \(P_1(\cos \theta)\) has been replaced by \(\cos \theta\), so that the new potential for arbitrary polarization can be written as

\[
\phi(r, \theta, \varphi) = -E_0 P_1(\cos \gamma) R(r),
\]  

(2.3.3)

where

\[
R(r) = \begin{cases} 
  r + \frac{\kappa_1 - \kappa_2}{2\kappa_2 + \kappa_1} \frac{a^3}{r^2}, & r > a; \\
  \frac{\kappa_2}{2\kappa_2 + \kappa_1} r, & r < a.
\end{cases}
\]  

(2.3.4)

The Legendre addition theorem\(^{[30]}\)

\[
P_n(\cos \gamma) = P_n(\cos \theta) P_n(\cos \theta_0) + 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_m^n(\cos \theta) P_m^n(\cos \theta_0) \cos(\varphi - \varphi_0),
\]  

(2.3.5)

allows the calculation of \(P_1(\cos \gamma)\) in terms of the angular variables \((\theta, \varphi)\) and \((\theta_0, \varphi_0)\) as

\[
P_1(\cos \gamma) = \cos \gamma = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0),
\]  

(2.3.6)

which yields the potential in terms of the angular variables \(\theta\) and \(\varphi\). The corresponding electric fields are

\[
E(r, \theta, \varphi) = \left( \hat{r} \cos \gamma \frac{\partial R}{\partial r} + \hat{\theta} \frac{R}{r} \frac{\partial \cos \gamma}{\partial \theta} + \hat{\varphi} \frac{R}{r \sin \theta} \frac{\partial \cos \gamma}{\partial \varphi} \right) E_0,
\]  

(2.3.7)

where

\[
\frac{\partial \cos \gamma}{\partial \theta} = -\sin \theta \cos \theta_0 + \cos \theta \sin \theta_0, \cos(\varphi - \varphi_0)
\]

\[
\frac{\partial \cos \gamma}{\partial \varphi} = -\sin \theta \sin \theta_0 \sin(\varphi - \varphi_0).
\]

### 2.4 Isotropic Spheres in an Isotropic Medium

In a mixture of nano-spheres of radius \(a\) in a dielectric medium as illustrated in Fig. 2.3, the spheres are placed in a “cubic lattice” configuration. The spacing between embedded spheres is \(d\) so that the spheres have a volume \(V_1 = 4/3 \pi a^3\) while the volume of each unit cell is \(V = d^3\). The larger sphere has a radius of \(R = d/2\).
Figure 2.3. A mixture of multiple embedded dielectric spheres of radius $a$ and a dielectric constant $\varepsilon_1 = \varepsilon_0 \kappa_2$, in a dielectric medium, $\varepsilon_2 = \varepsilon_0 \kappa_1$. The unit cells have a volume of $V = d^3$.

The mixture with embedded spheres acts as a medium with an effective dielectric constant that is a function of the volume fractions of materials with dielectric constants $\kappa_1$ and $\kappa_2$. The electric field governed by the mixture is denoted as

$$\mathbf{D} = \varepsilon_0 \kappa_{\text{eff}} \mathbf{E},$$

where $\mathbf{E}$ is the super macroscopic field $\mathbf{E} = \langle \mathbf{E} \rangle$, which is the average of the “microscopic field,” $\mathbf{E}$, which replaces $\mathbf{E}$ in (2.2.17). (The field calculations in Section 2.2 are now the “microscopic fields.”) $\kappa_{\text{eff}}$ is the homogenized dielectric constant. To calculate the “macroscopic” fields, it is assumed that interaction between spheres is negligible or that $d/2 \gg a$ so that the fields in each unit cell can be approximated using (2.2.17). The super macroscopic field must be calculated for all rectangular components given in

$$\mathbf{\hat{r}} = \hat{x} \sin \theta \cos \varphi + \hat{y} \sin \theta \sin \varphi + \hat{z} \cos \theta,$$

$$\mathbf{\hat{\theta}} = \hat{x} \cos \theta \cos \varphi + \hat{y} \cos \theta \sin \varphi - \hat{z} \sin \theta.$$

Note that the electric field in (2.2.17) can be split into (1) the constant field, denoted as $\mathbf{E}_c$, and (2) the variable field, denoted as $\mathbf{E}_v$. The main part of the $z$ component of the super macroscopic field is obtained by averaging over the constant $\mathbf{E}_c$ fields in the unit cell. The $E_{zv}$ component due to the $\mathbf{\hat{r}}$ and $\mathbf{\hat{\theta}}$ components in (2.2.17) is given below. Integration of the constant terms yields

$$E_{zc} = \frac{1}{V} \iiint_V E_{zc} dV = \frac{1}{V} \left( \frac{3\kappa_2}{2\kappa_2 + \kappa_1} V_1 + (V - V_1) \right) E_0,$$

$$= \left( \frac{3\kappa_2}{2\kappa_2 + \kappa_1} f + 1 - f \right) E_0. \quad (2.4.1)$$

where $f = V_1 / V$ is the fraction of the volume of the embedded sphere and the volume of the unit cell. The two remaining $z$ components obtained from
\( \mathbf{\hat{r}} \) and \( \mathbf{\hat{\theta}} \) components is approximated by integrating over the larger sphere, excluding the embedded sphere, giving

\[
E_{zv} = \iiint E_{zv} r^2 \sin \theta dr d\theta d\varphi,
\]
\[
= \frac{4\pi}{3} \frac{\kappa_1 - \kappa_2}{2\kappa_2 + \kappa_1} \ln(a/R) \int_0^\pi (2 \cos^2 \theta - \sin^2 \theta) \sin \theta d\theta \]
\[
= 0.
\]

Because of symmetry the \( E_x \) and \( E_y \) components are zero resulting from integration of \( \varphi \) over the range \( (0, 2\pi) \). Thus, the super macroscopic field of the mixture is

\[
E = \hat{z} \frac{2\kappa_2 + \kappa_1 + 2 f(\kappa_1 - \kappa_2)}{2\kappa_2 + \kappa_1} E_0. \tag{2.4.2}
\]

The effective dielectric constant is computed from \((\kappa_1 \mathcal{E})_{V_1}\), the average over \( V_1 \), and \((\kappa_2 \mathcal{E})_{V_2}\), the average over \( V_2 \).

\[
\kappa_{\text{eff}} E = \frac{1}{V} (\langle \kappa_1 \mathcal{E} \rangle_{V_1} + \langle \kappa_2 \mathcal{E} \rangle_{V_2})
\]
\[
= \hat{z} \kappa_2 \frac{2\kappa_2 + \kappa_1 + 2 f(\kappa_1 - \kappa_2)}{2\kappa_2 + \kappa_1} E_0 \tag{2.4.3}
\]

where \( V_2 = V - V_1 \) is the volume of the region outside the embedded sphere. Substituting the value of \( E_0 \) in terms of the super macroscopic field, the effective dielectric constant becomes

\[
\kappa_{\text{eff}} = \kappa_2 \left( 1 + 3 f \frac{\kappa_1 - \kappa_2}{2\kappa_2 + \kappa_1 - f(\kappa_1 - \kappa_2)} \right). \tag{2.4.4}
\]

Equation (2.4.4) is known as the Maxwell-Garnett mixing recipe,[12][28] first derived by J. C. Maxwell Garnett.[7]

### 2.5 Anisotropic Spheres in an Isotropic Medium

When the embedded spheres are anisotropic then the Maxwell-Garnett mixing formula in Section 2.4 must be modified to account the dielectric tensor of the spheres. The polarizability of the anisotropic spheres will be somewhat different from isotropic spheres in Section 2.2.

In a charge-free region, the equation for the electrostatic potential is

\[
\nabla \cdot (\mathbf{\kappa} \cdot \nabla) \phi = 0, \tag{2.5.1}
\]

where the dyad \( \mathbf{\kappa} \) is the relative dielectric constant of the sphere. For a general dielectric dyad

\[
\mathbf{\kappa} = \begin{bmatrix}
\kappa_{11} & \kappa_{12} & \kappa_{13} \\
\kappa_{21} & \kappa_{22} & \kappa_{23} \\
\kappa_{31} & \kappa_{32} & \kappa_{33}
\end{bmatrix}, \tag{2.5.2}
\]
and the operator in (2.5.1) becomes

\[ \nabla \cdot (\kappa \cdot \nabla) = \kappa_{11} \frac{\partial^2}{\partial x^2} + \kappa_{22} \frac{\partial^2}{\partial y^2} + \kappa_{33} \frac{\partial^2}{\partial z^2} \]

\[ + (\kappa_{12} + \kappa_{21}) \frac{\partial^2}{\partial x \partial y} + (\kappa_{13} + \kappa_{31}) \frac{\partial^2}{\partial x \partial z} + (\kappa_{23} + \kappa_{32}) \frac{\partial^2}{\partial y \partial z}, \]  

(2.5.3)

where it is assumed the partial derivatives commute. For dyads with asymmetric off-diagonal terms, \( \kappa_{ij} = -\kappa_{ji}, \ i \neq j \), the terms with cross derivatives in (2.5.3) cancel, and Laplace’s equation becomes

\[ \kappa_{11} \frac{\partial^2 \phi}{\partial x^2} + \kappa_{22} \frac{\partial^2 \phi}{\partial y^2} + \kappa_{33} \frac{\partial^2 \phi}{\partial z^2} = 0. \]  

(2.5.4)

Equation (2.5.4) depends only on the diagonal terms of (2.5.2). Nevertheless, the off-diagonal terms are required in applying boundary conditions.

As an example of a sparse dyad assume the matrix form to be

\[ \bar{\kappa} = \begin{bmatrix} \kappa_v & 0 & -\kappa_o \\ 0 & \kappa_c & 0 \\ -\kappa_o & 0 & \kappa_v \end{bmatrix}, \]  

(2.5.5)

which represents magnetic-type material with a static magnetic field biased along the \( x \) direction. The dyad may be written as

\[ \bar{\kappa} = (\hat{x} \hat{x} + \hat{y} \hat{y})\kappa_v + \hat{y} \hat{y} \kappa_c + (\hat{z} \hat{z} - \hat{x} \hat{x})\kappa_o. \]  

(2.5.6)

The operator in (2.5.3) reduces to

\[ \nabla \cdot \bar{\kappa} \cdot \nabla = \kappa_v \frac{\partial^2}{\partial x^2} + \kappa_c \frac{\partial^2}{\partial y^2} + \kappa_v \frac{\partial^2}{\partial z^2}, \]  

(2.5.7)

where \( \partial_x = \partial/\partial x \), etc. Although (2.5.6) represents a specific anisotropic dyad, it should be noted that when \( \bar{\kappa} \) has the off-diagonal asymmetric form, the electrostatic potential obtained from (2.5.1) will depend only on the diagonal elements \( \kappa_c \) and \( \kappa_v \).

By transforming the Cartesian coordinates in (2.5.4) using

\[ \bar{x} = \frac{x}{\sqrt{\kappa_{11}}}; \quad \bar{y} = \frac{y}{\sqrt{\kappa_{22}}}; \quad \bar{z} = \frac{z}{\sqrt{\kappa_{33}}}, \]  

(2.5.8)

the electrostatic potential satisfies Laplace’s equation

\[ \nabla^2 \phi(\bar{x}, \bar{y}, \bar{z}) = 0. \]  

(2.5.9)

The spherical surface in \((x, y, z)\) becomes an ellipsoid in \((\bar{x}, \bar{y}, \bar{z})\) or an oblate spheroid with \( \kappa_{11} = \kappa_{22} = \kappa_v \) and \( \kappa_{33} = \kappa_c \), provided the dielectric constants are real with \( \kappa_c > \kappa_v \).
The components of the arbitrary dielectric dyad, \( \kappa_{ij} \), in rectangular coordinates can be transformed to spherical coordinates, \( \kappa^s_{ij} \). Table 2.1 lists the corresponding spherical components in terms of rectangular components. (The off-diagonal elements are asymmetric.) Thus, transforming (2.5.6) to spherical coordinates gives

\[
\begin{align*}
\kappa_{11}^s &= (\kappa_{11} \cos^2 \varphi + \kappa_{22} \sin^2 \varphi) \sin^2 \theta + \kappa_{33} \cos^2 \theta \\
\kappa_{22}^s &= (\kappa_{11} \cos^2 \varphi + \kappa_{22} \sin^2 \varphi) \cos^2 \theta + \kappa_{33} \sin^2 \theta \\
\kappa_{33}^s &= \kappa_{11} \sin^2 \varphi + \kappa_{22} \cos^2 \varphi \\
\kappa_{12}^s &= (\kappa_{11} \cos^2 \varphi + \kappa_{22} \sin^2 \varphi - \kappa_{33}) \cos \theta \sin \theta \\
\kappa_{21}^s &= (\kappa_{11} \cos^2 \varphi + \kappa_{22} \sin^2 \varphi - \kappa_{33}) \cos \theta \sin \theta \\
\kappa_{13}^s &= -\left(\kappa_{11} - \kappa_{22}\right) \sin \theta \cos \varphi \sin \varphi \\
&\quad + \kappa_{12} \sin \theta + \kappa_{13} \cos \theta \sin \varphi - \kappa_{23} \cos \theta \cos \varphi \\
\kappa_{31}^s &= -\left(\kappa_{11} - \kappa_{22}\right) \sin \theta \cos \varphi \sin \varphi \\
&\quad - \kappa_{12} \sin \theta - \kappa_{13} \cos \theta \sin \varphi + \kappa_{23} \cos \theta \cos \varphi \\
\kappa_{23}^s &= -\left(\kappa_{11} - \kappa_{22}\right) \cos \theta \cos \varphi \sin \varphi \\
&\quad + \kappa_{12} \cos \theta - \kappa_{13} \sin \theta \sin \varphi + \kappa_{23} \sin \theta \cos \varphi \\
\kappa_{32}^s &= -\left(\kappa_{11} - \kappa_{22}\right) \cos \theta \cos \varphi \sin \varphi \\
&\quad - \kappa_{12} \cos \theta + \kappa_{13} \sin \theta \sin \varphi - \kappa_{23} \sin \theta \cos \varphi 
\end{align*}
\]

In spherical coordinates, relevant transformations from (2.5.8) include

\[
\begin{align*}
\tilde{r} &= q(\vartheta, \varphi) \tilde{r} \\
\cos \tilde{\vartheta} &= \frac{1}{q(\vartheta, \varphi) \sqrt{\kappa_{33}}} \cos \vartheta, \quad \sin \tilde{\vartheta} = \frac{g(\varphi)}{q(\vartheta, \varphi)} \sin \vartheta, \\
e^{\pm j\tilde{\vartheta}} &= \frac{1}{2g(\varphi)} \left[ e^{j\vartheta} \left( \frac{1}{\sqrt{\kappa_{11}}} \pm \frac{1}{\sqrt{\kappa_{22}}} \right) + e^{-j\vartheta} \left( \frac{1}{\sqrt{\kappa_{11}}} \mp \frac{1}{\sqrt{\kappa_{22}}} \right) \right]
\end{align*}
\]
2.5 ANISOTROPIC SPHERES IN AN ISOTROPIC MEDIUM

where

\[ g^2(\phi) = \frac{\cos^2 \phi}{\kappa_{11}} + \frac{\sin^2 \phi}{\kappa_{22}} \]

\[ q^2(\theta, \phi) = g^2(\phi) \sin^2 \theta + \frac{\cos^2 \theta}{\kappa_{33}}. \]

Note that for azimuthal symmetry, \( \kappa_{11} = \kappa_{22}, \) \( g \) is a constant while \( q \) depends only on \( \theta. \)

The solution of Laplace’s equation will be expressed in terms of spherical harmonics because there may not be azimuthal symmetry when the electric field is biased in an arbitrary direction as in Section 2.3. Inside the anisotropic sphere

\[ \phi^- (\tilde{r}, \tilde{\theta}, \tilde{\phi}) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} C_{mn} Y_{mn}(\tilde{\theta}, \tilde{\phi}) \tilde{r}^n. \] (2.5.12)

Since the \( n = 0 \) term represents a constant potential, it has been dropped from the series. The spherical harmonics \( Y_{mn}(\theta, \phi) \) are

\[ Y_{mn}(\theta, \phi) = P^m_n(\cos \theta)e^{jm\phi}. \] (2.5.13)

and they are orthogonal over the surface \( S \) of a sphere

\[ \oint_S Y^*_{m'n'}(\theta, \phi)Y_{mn}(\theta, \phi)d\Omega = N_{mn} \delta_{mm'} \delta_{nn'}. \] (2.5.14)

where \( d\Omega = \sin \theta d\theta d\phi \) and the normalization coefficients are

\[ N_{mn} = \frac{4\pi (n + m)!}{2n + 1 (n - m)!}. \] (2.5.15)

The potential outside the sphere \( (r > a) \) for an arbitrary polarization along \( (\theta_0, \phi_0) \) is given by

\[ \phi^+(r, \theta, \phi) = -E_0 P_1(\cos \gamma) r + \sum_{n=1}^{\infty} \sum_{m=-n}^{n} B_{mn} Y_{mn}(\theta, \phi)/r^{n+1}. \] (2.5.16)

The \( n = 0 \) term has been dropped as it represents the potential of a monopole, not present in the charge-free region. From the orthogonal relations for the spherical harmonics, \( P_1(\cos \gamma) \) can be expressed as

\[ P_1(\cos \gamma) = \sin \theta_0 e^{j\phi_0} Y_{-11}(\theta, \phi) + \cos \theta_0 Y_{01}(\theta, \phi) - \frac{1}{2} \sin \theta_0 e^{-j\phi_0} Y_{11}(\theta, \phi) \] (2.5.17)

The coefficients \( B_{mn} \) and \( C_{mn} \) are determined from the boundary conditions on the potential and radial flux density at \( r = a. \) The boundary conditions for the electrostatic potential requires

\[ \sum_{m,n} C_{mn} Y_{mn}(\tilde{\theta}, \tilde{\phi}) q^n d^n = -E_0 P_1(\cos \gamma) a + \sum_{m,n} B_{mn} Y_{mn}(\theta, \phi)/a^{n+1}. \] (2.5.18)
The radial flux inside the sphere is
\[ D_r = -\varepsilon_0 \mathbf{r} \cdot \mathbf{k} \cdot \nabla \phi^+ = -\varepsilon_0 \mathbf{r} \cdot \mathbf{k} \left( r \frac{\partial \phi^-}{\partial r} + \hat{\theta} \frac{1}{r} \frac{\partial \phi^-}{\partial \theta} + \hat{\phi} \frac{1}{r \sin \theta} \frac{\partial \phi^-}{\partial \phi} \right) \]  
(2.5.19)

The gradient of the potential inside the sphere requires derivatives with respect to \( r, \theta \) and \( \varphi \). Thus, \( \phi^- (\mathbf{r}, \hat{\theta}, \hat{\phi}) \) must be evaluated using (2.5.11) so that matching the radial flux inside the sphere to that outside yields
\[ - \sum_{m,n} C_{mn} d^{n-1} \left\{ \kappa_{11} q^n Y_{mn}(\hat{\theta}, \hat{\phi}) + \kappa_{12} \frac{\partial [q^n Y_{mn}(\hat{\theta}, \hat{\phi})]}{\partial \theta} + \kappa_{13} \frac{1}{\sin \theta} \frac{\partial [q^n Y_{mn}(\hat{\theta}, \hat{\phi})]}{\partial \phi} \right\} = \kappa_2 \left[ E_0 P_1 (\cos \gamma) + \sum_{m} (n+1) B_{mn} Y_{mn}(\theta, \varphi) / a^{n+2} \right] \]  
(2.5.20)

Two sets of linear equations obtained by multiplying (2.5.18) and (2.5.20) by \( Y_{m'n'}(\theta, \varphi) \) and integrating over the spherical surface are somewhat complicated because \( q(\theta, \varphi) Y_{mn}(\hat{\theta}, \hat{\phi}) \) and the associated derivatives are not generally orthogonal to \( Y_{mn}(\theta, \varphi) \). Ordering the double series with a single index \( l = n(n + 1) + m = 1, 2, \ldots, l_{\max} \rightarrow \infty \), with \( C_l \equiv C_{mn} \), and so forth, the double series can be reduced to a single series. For \( n = 1, l = 1, 2, 3, \) corresponding to \( m = -1, 0, 1; \) for \( n = 2, l = 4, 5, 6, 7, 8, \) corresponding to \( m = -2, -1, 0, 1, 2. \) The value \( l_{\max} = n_{\max}(n_{\max} + 2) \). Equation (2.5.18) becomes
\[ \sum_{l=1}^{l_{\max}} L_{l'l} C_l = \]  
(2.5.21)

\[ - E_0 a^3 \left( \sin \theta_0 e^{i \varphi_0} N_1 \delta_{l'1} + \cos \theta_0 N_2 \delta_{l'2} - \frac{1}{2} \sin \theta_0 e^{-i \varphi_0} N_3 \delta_{l'3} \right) + N_{l'} B_{l'} \]
for \( l' = 1, 2, \ldots, l_{\max} \), where
\[ L_{l'l} = a^{n'+n+1} \oint Y^*_l(\theta, \varphi) q^n(\theta, \varphi) Y_l(\hat{\theta}, \hat{\phi}) d\Omega. \]  
(2.5.22)

Similarly, (2.5.20) can be reduced to
\[ \sum_{l=1}^{l_{\max}} M_{l'l} C_l = \]  
(2.5.23)

\[ - E_0 \frac{a^3}{2} \left( \sin \theta_0 e^{i \varphi_0} N_1 \delta_{l'1} + \cos \theta_0 N_2 \delta_{l'2} - \frac{1}{2} \sin \theta_0 e^{-i \varphi_0} N_3 \delta_{l'3} \right) - N_{l'} B_{l'} \]
where
\[ M_{l'l} = \frac{a^{n'+n+1}}{\kappa_2 (n' + 2)} \oint Y^*_l(\theta, \varphi) \times \]
\[ \left[ \kappa_{11} q^n Y_l(\hat{\theta}, \hat{\phi}) + \kappa_{12} \frac{\partial [q^n Y_l(\hat{\theta}, \hat{\phi})]}{\partial \theta} + \kappa_{13} \frac{1}{\sin \theta} \frac{\partial [q^n Y_l(\hat{\theta}, \hat{\phi})]}{\partial \phi} \right] d\Omega. \]  
(2.5.24)
2.5 ANISOTROPIC SPHERES IN AN ISOTROPIC MEDIUM

Equations (2.5.21) and (2.5.23) can be written as two matrix equations

\[ LC = -b + N B, \quad MC = -b/2 - N B, \]

where \( L \) and \( M \) are square matrices whose elements are given by (2.5.21) and (2.5.23), respectively, the vectors \( B \) and \( C \) containing the coefficients \( B_l \) and \( C_l \), respectively, \( N \) is a diagonal matrix containing the normalization coefficients given by (2.5.15), and the column vector

\[
b = E_0 \begin{bmatrix}
e^{j\varphi_0} N_1 \sin \theta_0 \\
N_2 \cos \theta_0 \\
-\frac{1}{2} e^{-j\varphi_0} N_3 \sin \theta_0 \\
0 \\
\vdots \\
0
\end{bmatrix},
\]

has only 3 non-zero terms, those corresponding \( l = 1, 2, 3 \) \( (n = 1, m = -1, 0, 1) \). The solution for \( C \) is given as

\[ C = -\frac{3}{2} (M + L)^{-1} b \]

However, note that two sets of homogeneous linear equations of the \( B_{mn} \) and \( C_{mn} \) can be obtained by partitioning the matrices \( G = M + L, C \) and \( b \) with the result

\[
\begin{bmatrix}
\Theta_{11} & \Theta_{12} \\
\Theta_{21} & \Theta_{22}
\end{bmatrix}
\begin{bmatrix}
C_1 \\
C_2
\end{bmatrix} = \frac{3}{2} \begin{bmatrix}
b_1 \\
b_2
\end{bmatrix},
\]

where \( \Theta_{11} \) is \( 3 \times 3 \) and \( \Theta_{12} \) is \( 3 \times l_{\text{max}} - 3 \). It can be shown that \( \Theta_{21} = 0 \) and since \( b_2 = 0, C_2 = 0 \). Solving for the unknown coefficients allows computation of the potential inside the sphere

\[
\phi^-(x, y, z) = -\frac{3 \kappa_2}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} [(2 \kappa_2 + \kappa_v)E_x + \kappa_0 E_z] x
\]
\[
-\frac{3 \kappa_2}{2 \kappa_2 + \kappa_c} E_y y
\]
\[
-\frac{3 \kappa_2}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} [-\kappa_0 E_x + (2 \kappa_2 + \kappa_v) E_z] z.
\]

The electric field inside the sphere is polarized in a different direction compared to the applied field external to the sphere and is given by

\[
E^- = 3 \kappa_2 \begin{bmatrix}
\frac{2 \kappa_2 + \kappa_v}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} & 0 & \frac{\kappa_0}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} \\
0 & \frac{1}{2 \kappa_2 + \kappa_c} & 0 \\
-\frac{\kappa_0}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} & 0 & \frac{2 \kappa_2 + \kappa_v}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2}
\end{bmatrix} \cdot E_0. \quad (2.5.26)
\]
where the matrix notation is used to express the dyad in Cartesian coordinates. In general, (2.5.26) may be written as

\[ \mathbf{E}^* = 3 \kappa_2 \left( \mathbf{k} + 2 \kappa_2 \mathbf{i} \right)^{-1} \cdot \mathbf{E}_0, \]  
(2.5.27)

where \( \mathbf{k} \) is given by (2.5.5) and \( \mathbf{i} \) is the unit dyad.

The potential outside the anisotropic sphere is

\[ \phi^+(x, y, z) = -\mathbf{E}_0 \cdot \mathbf{r} - \frac{a^3}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} \left\{ \left[ \left( \kappa_2 - \kappa_v \right) (2 \kappa_2 + \kappa_v) - \kappa_0^2 \right] E_z - 3 \kappa_0 \kappa_2 E_3 \right\} \frac{\cos \theta}{r^2} \]
\[ - \frac{a^3}{\kappa_0^2 + (2 \kappa_2 + \kappa_v)^2} \left\{ \left[ \left( \kappa_2 - \kappa_v \right) (2 \kappa_2 + \kappa_v) - \kappa_0^2 \right] E_x - 3 \kappa_0 \kappa_2 E_3 \right\} \frac{\sin \theta \cos \varphi}{r^2} \]
\[ - \frac{a^3}{2 \kappa_2 + \kappa_v} \left( \kappa_2 - \kappa_v \right) E_y \frac{\sin \theta \sin \varphi}{r^2} \]  
(2.5.28)

where \( \mathbf{E}_0 \), is given in (2.3.1) and \( \mathbf{r} = \hat{x} x + \hat{y} y + \hat{z} z \).

The electric field inside the sphere is constant while on the outside there is the constant applied field and the induced field due to the polarization caused by the sphere. The effective dielectric constant of a mixture of anisotropic spheres in an isotropic medium can be computed similar to that discussed in Section 2.4. In the spirit of the earlier discussion, we denote \( \mathcal{E} \) as the “microscopic field” and calculate the “macroscopic field” by averaging the “microscopic field” over regions in proximity to the anisotropic sphere. The “microscopic fields” are computed from the potentials given by (2.5.25) and (2.5.28). First, the “macroscopic field” of the anisotropic mixture is obtained in the same manner as that of the isotropic mixture given by (2.4.1). As before, the induced field, in the region \( r > a \), will not affect the average electric field. Only the constant internal and external fields produce the macroscopic field, where the constant external field is

\[ \mathcal{E}^+ = \mathbf{E}_0 = \hat{x} E_1 + \hat{y} E_2 + \hat{z} E_3 \]  
(2.5.29)

while the internal field (in terms of the external field) is

\[ \mathcal{E}^- = 3 \kappa_2 \left( \mathbf{k} + 2 \kappa_2 \mathbf{i} \right)^{-1} \cdot \mathbf{E}_0, \]  
(2.5.30)

so that the macroscopic field is

\[ \mathbf{E} = (1 - f) \mathcal{E}^+ + f \mathcal{E}^-, \]  
(2.5.31)

where \( f \) is the volume fraction of the sphere in the unit cell. Using the constant fields

\[ \mathbf{E} = (1 - f) \mathbf{E}_0 + 3 f \kappa_2 \left( \mathbf{k} + 2 \kappa_2 \mathbf{i} \right)^{-1} \cdot \mathbf{E}_0, \]  
(2.5.32)

so that the applied field, in terms of the “macroscopic field” can be expressed as

\[ \mathbf{E}_0 = \left( \mathbf{k} + 2 \kappa_2 \mathbf{i} \right) \cdot \left[ \left( \mathbf{k} + 2 \kappa_2 \mathbf{i} \right) - f \left( \mathbf{k} - \kappa_2 \mathbf{i} \right) \right]^{-1} \cdot \mathbf{E}. \]  
(2.5.33)
2.5 ANISOTROPIC SPHERES IN AN ISOTROPIC MEDIUM

The product of the effective dielectric constant and the electric field of mixture is given by

\[ \kappa_{\text{eff}} \cdot E = (1 - f)\kappa_2 E_0 + f 3 \kappa_2 \kappa \cdot (\kappa + 2 \kappa_2 i)^{-1} \cdot E_0. \] (2.5.34)

Substituting \( E_0 \) from (2.5.33) gives

\[ \kappa_{\text{eff}} = \kappa_2 \left[ (\kappa + 2 \kappa_2 i) + 2 f (\kappa - \kappa_2 i) \right] \cdot \left[ (\kappa + 2 \kappa_2 i) - f(\kappa - \kappa_2 i) \right]^{-1} \]
\[ = \kappa_2 \left\{ \frac{i}{3} + 3 f (\kappa - \kappa_2 i) \cdot \left[ (\kappa + 2 \kappa_2 i) - f(\kappa - \kappa_2 i) \right]^{-1} \right\}, \] (2.5.35)

which is similar to that given in reference[28]. If the medium is composed of spheres of different sizes then \( \kappa_{\text{eff}} \) is a function of the random variable \( f \). The expected value of the effective dielectric dyad would then be calculated from the probability density function \( p(f) \).

\[ \kappa_{\text{eff}} = \int_0^1 \kappa_{\text{eff}}(f) p(f) df \]

2.6 Absorption in Embedded Nanospheres

If the embedded sphere is a metallic nanospheres such as iron then the dielectric constant of the sphere is complex. The static field analysis above gives a good approximation of the time varying field provided the free-space wavelength of the field \( \lambda_0 \gg a \sqrt{\kappa_1} \).

At optical frequencies the dielectric constant of the sphere is \( \kappa_2 = \kappa_2' - j\kappa_2'' \), and the above analysis is applicable when \( \kappa_2 \) is replaced by \( \kappa_2' \). The conductivity of the spheres is \( \sigma(\omega) = \omega \kappa_0 \kappa_2''(\omega) \). The total absorption in the sphere is

\[ P_{\text{loss}} = 6\pi \omega \kappa_0 a^3 \frac{\kappa_2'^2 + \kappa_2''^2}{(\kappa_1 + 2\kappa_2')^2 + 2\kappa_2''^2} E_0^2 \]
Chapter 3
Plane Waves

Plane waves play a key role in understanding the propagation of more complex waves in waveguides as well as in radiation systems. In Chapter 1 the fundamental solution of Maxwell’s equations was discussed in terms of an integral or summation of plane waves.

3.1 Introduction
Since the solution of the wave equation can be written as a linear combination of plane waves, it is important to understand some of the basic properties of plane waves. A plane wave is defined as one that is dependent only on one space variable, under a suitable transformation of the coordinate system, and time.

3.2 Formalism
Generally, a plane wave can be defined as one propagating in a direction defined by the vector \( \mathbf{k} = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z \). Using a Cartesian coordinate transformation, the direction of propagation can be transformed to a direction along one of the coordinate axes, say the \( x \) axis. Thus, without loss of generality, assume that all field quantities vary only with respect to \( x \) and \( t \), so that the field quantities have a functional dependence of the form \( e(z,t) \) and \( h(z,t) \). The two Maxwell curl equations become

\[
0 = \frac{\partial b_x}{\partial t}, \quad (3.2.1)
\]
\[
\frac{\partial e_y}{\partial x} = -\frac{\partial b_z}{\partial t}, \quad (3.2.2)
\]
\[
\frac{\partial e_z}{\partial x} = \frac{\partial b_y}{\partial t}, \quad (3.2.3)
\]
and

\begin{align}
0 &= \partial d_x / \partial t, \quad (3.2.4) \\
\partial h_y / \partial x &= \partial d_z / \partial t, \quad (3.2.5) \\
\partial h_z / \partial x &= -\partial d_y / \partial t. \quad (3.2.6)
\end{align}

From (3.2.1) and (3.2.4), it is recognized that the \( x \) components of the electromagnetic field, \( d_x \) and \( b_x \) are independent of time, i.e., \( x \) components are static if they exist. Since dynamic fields are the focus of this discussion, assume a static fields are zero, \( d_x, b_x = 0 \). The remaining four equations above can be combined to describe the propagation of plane waves. The remaining field components are \( e_y, d_y, h_y, h_z, b_y \) and \( b_z \) are all perpendicular to the \( x \) direction. The relationship between \( d \) and \( e \) and the relationship between \( b \) and \( h \) are determined from (\textit{ef. 51}) and (1.8.5) respectively.

For the remainder of this section, the dielectric constant and permeability will assumed to be frequency independent. (This assumption will assist in the understanding of forward and backward waves.) A more general development later in this chapter will allows the constitutive parameters to be frequency dependent.

Now \( h_z \) will be eliminated from two of the equations; the resulting equation governing the electric field is

\[ \frac{\partial^2 e_y}{\partial x^2} = \mu \varepsilon \frac{\partial^2 e_y}{\partial t^2}. \quad (3.2.7) \]

This equation governs the behavior of the \( y \) component of the electric field. Similarly, the equations governing the remaining field components have identical form and, a generic partial differential equation obeyed by all components can be written as

\[ \frac{\partial^2 \phi}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 \phi}{\partial t^2}, \quad (3.2.8) \]

where \( v = 1/\sqrt{\mu \varepsilon} \). The general solutions of (3.2.8) have the following form

\[ \phi^+(x,t) = f(t - x/v), \quad (3.2.9) \]
\[ \phi^-(x,t) = g(t + x/v). \quad (3.2.10) \]

where the two function \( f() \) and \( g() \) are arbitrary. The first solution represents a wave traveling in the positive \( x \) direction while the second represents a wave traveling in the negative \( x \) direction. For example, if the argument of \( f \) is a constant, then the derivative of its argument indicates that \( dx/dt = v \) with the quantity \( v \) in (3.2.8) representing the \textit{velocity of propagation}.

To simplify the discussion, assume there is only a single component of the electric field, \( e_y \), with the corresponding magnetic field, \( h_z \). The \( y \) component of the electric field has the solution

\[ e^+_y(x,t) = E_f f(t - x/v), \]
the magnetic field associated with $e_y$ is

$$h^+_z(x,t) = -\frac{E_t}{\sqrt{\mu/\epsilon}} f(t-x/v).$$

Note that the added superscript $^+$ is used to designate propagation in the positive $x$ direction for both field components. The field components and $e_y$ and $h_z$ are tied together in a linear relation

$$\frac{e^+_y}{h^+_z} = \sqrt{\frac{\mu}{\epsilon}} = -Z,$$

is called the wave impedance. The wave admittance $Y \equiv 1/Z$. For free-space the wave impedance, $Z_0$, is $120\pi$ ohms. If the $x$ component of the electric field exists, then the $y$ component of the magnetic field is tied to $e_x$, as determined from Maxwell’s equations,

$$\frac{e^+_x}{h^+_y} = Z$$

Similarly, for the backward propagating wave

$$e^-_y(x,t) = E_b g(t + x/v),$$

while the associated magnetic field is

$$h^-_z(x,t) = Y E_b g(t - x/v),$$

and the ratio of electric and magnetic field components is

$$\frac{e^-_y}{h^-_z} = Z.$$

It is seen that the ratio of the electric and magnetic field strengths have different signs depending on whether propagation is in the positive or in the negative $x$ directions. The sign of these ratios can also be understood by investigating the Poynting vector and the noting the direction of power flow. When the wave is traveling in the positive $x$ direction, $\hat{x} \cdot (e \times h)$ is positive whereas it is negative when the wave is traveling in the negative $x$ direction. For example the $x$ components of the instantaneous Poynting vectors are

$$s^+_x = \hat{x} \cdot (e \times h) = E_t^2 \left[ f(t-x/v) \right]^2 / Z,$$

for positive-traveling waves and

$$s^-_x = \hat{x} \cdot (e \times h) = -E_b^2 \left[ g(t-x/v) \right]^2 / Z,$$

for backward-traveling waves.
Another property of plane waves is the fact that the electric and magnetic fields are perpendicular. General polarized fields, \( e^+ = \hat{x}e^+_x + \hat{y}e^+_y \) and \( h^+ = \hat{x}h^+_x + \hat{y}h^+_y \), are perpendicular for
\[
e^+ \cdot h^+ = e^+_x h^+_x + e^+_y h^+_y = 0.
\]

### 3.3 Phase and Group Velocities

The previous section pertained to arbitrary time variation signals, while this section will focus on harmonic time variation. When the field is harmonically varying with time, the manipulation of the electromagnetic expressions are relatively simple and time averages of energy storage and power flux are easy to compute. The focus of this section will be on the development of phase and group velocities and the relation of group velocity and energy storage to power flux. The material will assumed to be homogeneous and isotropic with frequency-dependent constants \( \mu \) and \( \epsilon \). Assuming time variation with a single frequency \( \exp(j\omega t) \), (3.2.8) becomes
\[
\frac{d^2\phi}{dx^2} + k^2 \phi = 0. \tag{3.3.1}
\]
where \( k^2 = \omega^2 \mu \epsilon \). A forward traveling wave will be described assuming the electric field has an \( E_y \) component and the field can be written as
\[
e^+_y(x, t) = E^+_y(x, \omega) e^{j\omega t} = E_f e^{j(\omega t - kx)}, \tag{3.3.2}
\]
while the wave traveling in the negative \( x \) direction is
\[
E^-_y(x, \omega) = E_b e^{j(\omega t + kx)}. \tag{3.3.3}
\]

Magnetic fields associated with electric fields above are (the time variation is understood)
\[
H^+_z(x, \omega) = -YE_f e^{-j kx}, \tag{3.3.4}
H^-_z(x, \omega) = YE_b e^{j kx}. \tag{3.3.5}
\]
The time average Poynting vectors becomes
\[
S^+_x = \frac{1}{2} Y |E_f|^2, \tag{3.3.6}
S^-_x = -\frac{1}{2} Y |E_b|^2. \tag{3.3.7}
\]
Note that \( S^\pm = \hat{x} \cdot S^\pm \).

For a slightly more complicated signal, assume that there are two frequencies present, and the propagation will be in the positive \( x \) direction. If the frequencies are \( \omega_0 + \delta \omega \) and \( \omega_0 - \delta \omega \) the value of \( k \) will be modified because it is defined with an explicit \( \omega \) dependence and an implicit dependence due to the
Fact that the permeability and permittivity may be frequency dependent, i.e., the material is dispersive. A formal Taylor expansion of $k$ about $\omega_0$ is

$$k(\omega) = k(\omega_0) + \left. \frac{d k(\omega)}{d \omega} \right|_{\omega=\omega_0} (\omega - \omega_0) + \frac{1}{2} \left. \frac{d^2 k(\omega)}{d \omega^2} \right|_{\omega=\omega_0} (\omega - \omega_0)^2 + \cdots$$

$$= (k_0) + (k_1) \delta \omega + (k_2) (\delta \omega)^2 + \cdots \quad (3.3.8)$$

Characteristics of the two-frequency component propagating field will be derived using only the first two terms of (3.3.8) which is a good approximation of $k(\omega)$ when $\delta \omega$ is small. The two frequency components have corresponding wave numbers designated as $(k_0) + \delta k$ and $(k_0) - \delta k$, where $\delta k = (k_1) \delta \omega$. (If the material is non-dispersive, there are only two terms in the series expansion.) For the linear medium, the forward propagating fields become

$$e^+(y, t) = \frac{1}{2} E_y \left[ e^{j[(\omega_0 - \delta \omega) t - ((k_0) - \delta k) y]} + e^{j[(\omega_0 + \delta \omega) t - ((k_0) + \delta k) y]} \right],$$

$$= E_y \cos(\delta \omega t - \delta k y) e^{j(\omega_0 t - (k_0) y)}. \quad (3.3.9)$$

The magnetic field associated with the propagating electric field is

$$h^+(z, t) = -\frac{1}{2} E_z \left[ Y(\omega_0 - \delta \omega) e^{j[(\omega_0 - \delta \omega) t - ((k_0) - \delta k) y]} + Y(\omega_0 + \delta \omega) e^{j[(\omega_0 + \delta \omega) t - ((k_0) + \delta k) y]} \right],$$

The variation of the wave admittance from its value at $\omega_0$ is a "second-order" effect on the fields. On the other hand, small changes of the wave number with frequency can have large effects on the field expressions because the variations occur in the exponent, i.e., the small variations produce phase changes. Accordingly, the wave impedance at $\omega_0 \pm \delta \omega$ will be approximated with its value at $\omega_0$. The resulting magnetic field is

$$h^+(z, t) = -Y(\omega_0) E_z \cos(\delta \omega t - \delta k y) e^{j(\omega_0 t - (k_0) y)}. \quad (3.3.10)$$

At $x = 0$, the electric, magnetic and time average Poynting vector are

$$e^+(t) = E_y \cos \delta \omega t e^{j\omega_0 t} = E_y m(t) e^{j\omega_0 t}, \quad (3.3.11)$$

$$h^+(t) = -Y(\omega_0) E_z m(t) e^{j\omega_0 t}, \quad (3.3.12)$$

$$S^+_x = \frac{1}{4} Y(\omega_0)|E|^2. \quad (3.3.13)$$

The propagation term is similar to that of (3.3.2), but its amplitude slowly varies with time, similar to the narrow-band processes discussed in Section 1.9. Figure 3.1 illustrates the propagating wave where the envelope is represented by the dashed line. If the third term in the series expansion of $k(\omega)$ was included in the field expressions, then the envelope would have a phase variation term of $\exp[-j(k_2)(\delta \omega)^2 x]$. 
When the argument of the propagating wave is a constant, i.e.

\[ \omega_0 t - (k)_0 x = \text{const}, \]

its derivative with respect to time yields

\[ \frac{dx}{dt} = \frac{\omega_0}{(k)_0} \equiv v_p, \quad (3.3.14) \]

which is defined as the phase velocity, while the velocity of the envelope is determined by differentiating the argument of the envelope. The resulting envelope velocity is known as the group velocity, and is given by

\[ v_g(\omega_0) \equiv \frac{dz}{dt} = \frac{1}{\delta k/\delta \omega} \rightarrow \frac{1}{(k)_1}, \quad (3.3.15) \]

where \((k)_1\) is defined in (3.3.8). The concept of group velocity has physical meaning only when applied to narrow-band signals; it represents the propagation of a packet of frequencies clustered about \(\omega_0\). Since the group velocity is the velocity of the envelope, it represents the velocity of energy. In a dispersionless medium, the phase and group velocities are identical.

The relationship between the Poynting vector, electric and magnetic field storage densities, and the group velocity are fundamental concepts. Namely, the Poynting vector is equal to the product of the sum of the electric and magnetic energy storage densities per unit volume and wave group velocity. The sum of the energy storage in the electric and magnetic fields per unit volume can be obtained using the results of (1.9.11) and (1.9.13),

\[ w_e + w_m = \frac{\varepsilon_0 |E_l|^2}{4} \frac{2 \kappa_e \kappa_m + \omega_0 (\kappa_e \kappa_m)^'}{\kappa_m} \int_{-\infty}^{\infty} [\delta_m(\omega - \omega_0) + \delta_m(\omega + \omega_0)] d\omega, \]

where the \((\kappa_e \kappa_m)^'\) is the derivative of the product \(\kappa_e \kappa_m\) with respect to \(\omega\) and evaluated at \(\omega_0\), and \(\delta_m\) is the power spectral density of \(m(t)\) in (3.3.11). The group velocity as calculated from (3.3.15) is

\[ v_g = c = \frac{2 \sqrt{\kappa_e \kappa_m}}{2 \kappa_e \kappa_m + \omega_0 (\kappa_e \kappa_m)^'}. \]
where \( c = 1/\sqrt{\varepsilon_0/\mu_0} \) is the velocity of light in free space. The desired result is

\[
(w_c + w_m) v_g = \frac{1}{2} Y(\omega_0) |E_2|^2 \int_{-\infty}^{\infty} \left[ \delta_m(\omega - \omega_0) + \delta_m(\omega + \omega_0) \right] d\omega.
\]

The integral over the power spectral densities yields the same result as given in (3.3.13). Specifically,

\[
S_x = (w_c + w_m) v_g.
\]  

(3.3.16)

for the plane-wave solutions to Maxwell’s equations: the component of the Poynting vector in the direction of propagation is equal to flux of the energy density.

### 3.4 Polarization

In many instances it is important to understand the nature of the polarization of the electric field as it propagates along a given direction. For optical systems it is often easy to measure the polarization of the field using simple devices known as polarizers. For wave propagation along the positive \( x \) axis, the nature of the polarization is described by the path traced by the electric field vector in the \( y-z \) plane as time elapses. For example, Fig. 3.2(a) shows the electric field traces a path along the \( x \) axis, indicating linear polarization while in Fig. 3.2(b), the tip of the field vector traces an ellipse, elliptical polarization.

![Polarization Diagram](image)

**Figure 3.2.** The polarization of two different types of plane waves. (a) Linear Polarization: the electric field lies in the \( y-z \) plane with the tip of the electric field vector tracing a line along the \( y \) axis. (b) Elliptical Polarization: the electric field vector is perpendicular to the \( x \) axis and the tip of the vector traces an ellipse projected on the \( y-x \) plane. As the wave propagates in the positive \( x \) direction, the vector in the \( x = 0 \) plane rotates in a counter-clockwise direction in the figure. If the field is circularly polarized, the ellipse will have equal major and minor lengths.

Figure 3.2 illustrates the two main types of polarization: linear and elliptical. The elliptical-polarized field has vectors that change in length and direction while for the linear polarized field the vector lengths vary from zero to
some maximum value but their directions do not change. For circular polarization, the polarization path is a circle. For this special case the length of the electric field vectors as illustrated in (b) are equal, i.e., the length of the vector is a constant. The vector in Fig. 3.2(b) shows one complete revolution.

An analytical expression for an elliptically-polarized wave may be given by the fields

\[
e_y(x,t) = E_y \cos(\omega t - \beta x)
\]

\[
e_z(x,t) = E_z \cos(\omega t - \beta x + \theta(t))
\]

where \(\theta\) is a function of time, and \(E_y\) and \(E_z\) are the amplitudes of the corresponding \(y\) and \(z\) components. For linear polarization, \(\theta = 0\). Elliptical polarization occurs then \(E_y \neq E_z\) and the phase term is not zero. When \(\theta\) varies with time, the axes of the ellipse rotates with time.

### 3.5 Reflection and Transmission at Plane Boundaries

Many fundamental and useful parameters can be developed using the simple formulations of electromagnetic field continuities at plane boundaries. Indeed, reflection and radiation characteristics of semiconductor injection lasers can be understood from the wave reflection and transmission properties of plane waves incident upon a plane boundary separating two dielectric media. Figure 3.3 shows an incident wave on an interface with the reflected and transmitted waves. Note that the figure does not indicate the polarization of the electric or magnetic fields. The incident wave is directed at the interface at an angle \(\theta_i\) with respect to the normal of the interface. (In the figure the \(x\) axis is normal to the interface of the two media.) The reflected wave propagates away from the interface in the direction defined by the wave vector \(k_r\) while the transmitted wave “refracts” into the second medium and propagates at an angle \(\theta_t\).

![Figure 3.3](image)

There are two types of electric-field polarization associated with a plane wave that is obliquely incident on a plane interface between two media:
1. the electric field of the incident wave is perpendicular to the x-z plane (labeled $\perp$ polarization) or the plane of incidence and
2. the electric field is parallel to the x-z plane (labeled $\parallel$ polarization).

For $\perp$ polarization the electric field points in the y direction whereas for $\parallel$ polarization, the electric field has both x and z components. The incident electric field for case 1 can be expressed as

$$E_i(x, z) = \hat{y} E_i e^{-j k_1 r}.$$  \hfill (3.5.1)

whereas the magnetic field of the incident plane wave is

$$H_i(x, z) = Y_0 E_i (\hat{x} \sin \theta_i + \hat{z} \cos \theta_i) e^{-j k_1 r},$$  \hfill (3.5.2)

where the position vector $r = \hat{x} x + \hat{y} y + \hat{z} z$, and the incident wave vector is $k_i = k_0 (\hat{x} \cos \theta_i + \hat{z} \sin \theta_i)$. The ratio $|E_i|/|H_i|$ is the intrinsic wave impedance of medium 0, $Z_0 = 1/Y_0$. The reflected fields are given by

$$E_r(x, z) = \hat{y} E_r e^{-j k_r r},$$  \hfill (3.5.3)

$$H_r(x, z) = -Y_0 E_r (\hat{x} \sin \theta_r + \hat{z} \cos \theta_r) e^{-j k_r r},$$  \hfill (3.5.4)

where the wave vector of the reflected wave is $k_r = k_0 (\hat{x} \cos \theta_r + \hat{z} \sin \theta_r)$. Similarly, the transmitted fields are

$$E_t(x, z) = \hat{y} E_t e^{-j k_t r},$$  \hfill (3.5.5)

$$H_t(x, z) = Y_1 E_t (\hat{x} \sin \theta_t + \hat{z} \cos \theta_t) e^{-j k_t r},$$  \hfill (3.5.6)

where $k_t = k_1 (\hat{x} \cos \theta_t + \hat{z} \sin \theta_t)$. The boundary conditions for the three waves at $x = 0$ are

$$\hat{y} \cdot (E_i + E_r) = \hat{y} \cdot E_t,$$  \hfill (3.5.7)

$$\hat{z} \cdot (H_i + H_r) = \hat{z} \cdot H_t,$$  \hfill (3.5.8)

which are obtained by matching the tangential components at the interface.

The two equations have several unknown quantities: $E_r$ and $E_t$ which are the electric field strengths of the two waves generated by the incident wave, and the reflection and transmission angles $\theta_r$ and $\theta_t$. The first detailed equation obtained with matching tangential electric fields at $x = 0$ is

$$E_i e^{-j (k_1 \sin \theta_i) z} + E_r e^{-j (k_1 \sin \theta_r) z} = E_t e^{-j (k_0 \sin \theta_t) z}.$$  

This equation is a function of the space variable $z$ and it must be true for all $z$ which can only be satisfied by matching phase terms and coefficients. The phase matching produces two results: (1) $\theta_r = \theta_t$ and (2) $k_0 \sin \theta_i = k_1 \sin \theta_t$ (Snell’s Law). Matching the coefficients for both the tangential electric and magnetic fields gives

$$E_i + E_r = E_t,$$  \hfill (3.5.9)

$$(E_i - E_r) Y_0 \cos \theta_i = E_t Y_1 \cos \theta_t.$$  \hfill (3.5.10)
The “perpendicular” wave admittances (incident/transmitted regions) are defined as,

\[ Y_{\perp i} = Y_0 \cos \theta_i, \quad Y_{\perp t} = Y_1 \cos \theta_t, \quad (3.5.11) \]

whereas the “perpendicular” wave impedances are

\[ Z_{\perp i} = \frac{Z_0}{\cos \theta_i}, \quad Z_{\perp t} = \frac{Z_1}{\cos \theta_t}, \quad (3.5.12) \]

and the reflection, \( \rho \) and transmission \( \tau \) coefficients are defined as \( \rho = E_t/E_i \) and \( \tau = E_t/E_i \). It is easy to show that

\[
\rho_{\perp} = \frac{Y_{\perp i} - Y_{\perp t}}{Y_{\perp i} + Y_{\perp t}} = \frac{Z_{\perp i} - Z_{\perp t}}{Z_{\perp i} + Z_{\perp t}}, \quad (3.5.13)
\]

\[
\tau_{\perp} = \frac{2Y_{\perp i}}{Y_{\perp i} + Y_{\perp t}} = \frac{2Z_{\perp i}}{Z_{\perp i} + Z_{\perp t}}, \quad (3.5.14)
\]

The wave admittance \( Y_{\perp} \) or impedance \( Z_{\perp} \) govern the power flux toward and away from the interface. The scattering, reflection and transmission, of the fields is determined only by the wave admittance, \( Y_{\perp} \) or impedance \( Z_{\perp} \), given by (3.5.11) and (3.5.12) respectively.

There are similar results for parallel polarization which has field vectors

\[
E_i(x, z) = E_i(\hat{x} \cos \theta_i - \hat{z} \sin \theta_i) e^{-k_i \cdot r}, \quad (3.5.15)
\]

\[
H_i(x, z) = \hat{y} Y_0 E_i e^{-k_i \cdot r}. \quad (3.5.16)
\]

The reflected fields are

\[
E_r(x, z) = E_r(-\hat{x} \cos \theta_t - \hat{z} \sin \theta_t) e^{-k_r \cdot r}, \quad (3.5.17)
\]

\[
H_r(x, z) = \hat{y} Y_0 E_r e^{-k_r \cdot r}. \quad (3.5.18)
\]

and the transmitted fields are

\[
E_t(x, z) = E_t(\hat{x} \cos \theta_t - \hat{z} \sin \theta_t) e^{-k_t \cdot r}, \quad (3.5.19)
\]

\[
H_t(x, z) = \hat{y} Y_1 E_t e^{-k_t \cdot r}. \quad (3.5.20)
\]

Matching the fields at the boundaries yields results similar to those for the previous polarization. In this case, the “parallel” wave admittances are defined for the left/right regions as

\[ Y_{\parallel i} = \frac{Y_0}{\cos \theta_i}, \quad Y_{\parallel t} = \frac{Y_1}{\cos \theta_t}, \quad (3.5.21) \]

while the parallel wave impedances are given by

\[ Z_{\parallel i} = Z_0 \cos \theta_i, \quad Z_{\parallel t} = Z_1 \cos \theta_t. \quad (3.5.22) \]
3.5 REFLECTION AND TRANSMISSION AT PLANE BOUNDARIES

The reflection and transmission coefficients can be found using (3.5.13) and (3.5.14), simply by replacing $Z_\perp$ with $Z_\parallel$. For example, the reflection coefficient

$$\rho_\parallel = \frac{Y_{||} - Y_{||t}}{Y_{||} + Y_{||t}} = \frac{Z_{||t} - Z_{||||}}{Z_{||t} + Z_{||||}}. \quad (3.5.23)$$

The reflection of an oblique-incident wave will be considered using the above results. In this example, a plane wave inside a semiconductor will be directed toward a semiconductor/air interface. For the sake of simplicity, assume the both regions are non-magnetic material, i.e. the relative permeabilities, $\kappa_{m0} = \kappa_{m1} = 1$. Assume that Region 1 is free space while Region 2 has a refractive index of 3.6, such as that of GaAs at a free-space wavelength of $\lambda_0 = 1 \mu m$.

![Figure 3.4. The power reflectivity, $R = |\rho|^2$, of a plane wave incident upon a GaAs-air interface. At normal incidence, the parallel, $R_\parallel$, and perpendicular, $R_\perp$, reflectivities are equal at about 0.31. At angles larger than the critical angle, $16.1^\circ$, both reflectivities are 1.](image)

The power reflectivity for both the parallel and perpendicular polarizations is shown in Fig. 3.4 as a function of incident angle. In going from a high-index medium to low-index medium such as air, the rays refract away from the normal to the interface so that for rays whose incident angle, with normal, is greater than the critical angle, $\theta_c = \theta_i = 16.1^\circ$, there will be total internal reflection of power. An important distinction between $R_\perp$ and $R_\parallel$ is that there is an angle $\theta_i$ which gives $R_\parallel = 0$, provided the media are lossless, i.e., $\kappa_\epsilon0$ and $\kappa_\epsilon1$ are real. The point where total transmission occurs is when the wave impedances are identical, $Z_{||i} = Z_{||t}$. When the condition is satisfied, all the power incident on the interface is transmitted into Region 0; the angle $\theta_i \equiv \theta_B$, is called the Brewster angle.

In the process of extracting light from either a laser or LED, the photons must exit the semiconductor to air. In some cases it might be desirable to transfer the light form the semiconductor to glass-like materials which have refractive indices of about 1.5. In any event, the light must transfer from a
high-index material to a low-index material. Consequently, there is considerable interest in estimating the emission power from sources embedded in the semiconductor.

Finally, the wave impedances given above are valid complex quantities when the permittivity and permeability are complex. The reflection and transmission coefficients are also valid complex quantities.

### 3.6 Stratified Media

In the previous section the reflection and transmission properties of a plane wave incident on a simple plane interface was developed. When there are multiple layers of different materials, the reflection and transmission properties are more complex. However, with the development of a proper formalism, the complex characteristics can be easily understood. For the initial development, assume the plane wave is propagating perpendicular to the stratified layers. The boundaries of the different layers will lie in planes parallel to the $z$-$y$ plane, and propagation will be along the $x$ axis.

#### 3.6.1 Single-Layer Transfer Matrix

First, consider the three-layer structure shown in Fig. 3.5. The incident field in Region 0 produces a reflected field in Region 0 and a transmission to Region 2.

![Figure 3.5](image)

Figure 3.5. A plane wave is incident on a slab region sandwiched between two semi-infinite regions. To simplify the discussion below, all regions will be assumed to be non-magnetic and the permeability of all regions is assumed to be that of free space. The refractive index of the $i$th layer is $n_i = \sqrt{\varepsilon_i}$. The field components for harmonic time variation in a non-magnetic material satisfies

\[ -j \omega \mu_0 H_z = \frac{dE_y}{dx}. \]  
(3.6.1)

\[ j \omega \varepsilon E_y = -\frac{dH_z}{dx}. \]  
(3.6.2)

These two equations can be combined to produce a second-order differential equation for $E_y$ or $H_z$, of the from given in (3.3.1). A two-dimensional field
vector is now defined to represent the electric and magnetic fields. The electric field $E_y$ is a solution of the second-order differential equation while $Z_o H_x$ is determined using (3.6.1). Putting

$$E_y = \psi(z), \quad Z_o H_x = \frac{j}{k_o} \frac{d\psi(z)}{dz},$$

where $\omega \mu_o$ in (3.6.1) has been replaced with $k_o Z_o$, the two-dimensional field vector is defined as

$$\Psi(z) = \begin{pmatrix} \psi_0(z) \\ \psi_1(z) \end{pmatrix} \equiv \begin{pmatrix} E_y \\ Z_o H_z \end{pmatrix}. \quad (3.6.3)$$

Because $E_y$ and $H_y$ must be continuous across boundaries, this two-dimensional vector must be continuous across boundaries. By defining the first component as $E_y$ and the second component by $Z_o H_y$, both components have the same units of measure.

In Region 0, the incident wave components can be written as

$$E_{y0}^+ = E_i e^{-j k_o (x - x_1)}, \quad Z_o H_{z0}^+ = n_0 E_{y0}^+,$$

whereas the reflected wave components are

$$E_{y0}^- = \rho E_i e^{j k_o (x - x_1)}, \quad Z_o H_{z0}^- = -n_0 E_{y0}^-.$$ 

The variable $\rho$ is the electric field reflection coefficient. The electric and magnetic fields in Region 0 is obtained by adding the incident and reflected fields so that the two-dimensional field vector in Region 0 at the point $x = x_1^+$ is

$$\Psi(x_1^+) = E_i \begin{pmatrix} 1 + \rho \\ n_0(1 - \rho) \end{pmatrix}. \quad (3.6.4)$$

In Region 2, the transmitted wave components are

$$E_{y2}^+ = \tau E_i e^{-j k_2 (x - x_2)}, \quad Z_o H_{z2}^+ = n_2 E_{y2}^+.$$ 

where $\tau$ is the electric field transmission coefficient, and the two-dimensional field vector in Region 2 is

$$\Psi(x_2) = \tau E_i \begin{pmatrix} 1 \\ n_2 \end{pmatrix}. \quad (3.6.5)$$

In Region 1, the electric field can be written as the solution of the second-order differential equation as

$$\psi_1(x) = A \cos k_1 (x - x_1) + \frac{B}{jn_1} \sin k_1 (x - x_1).$$

The point $x = x_1^+$ is the point just to the left of $x_1$, i.e. the point is in Region 0, whereas the point $x = x_2^+$ is a point just to the right of $x_2$, i.e. is in Region 2.
where \( A \) and \( B \) must be determined from the initial condition of \( \Psi(x) \) at the left side of the layer. The two-dimensional field vector in Region 1 is

\[
\Psi(x) = \begin{pmatrix}
\cos k_1(x - x_1) & \frac{1}{jn_1} \sin k_1(x - x_1) \\
-jn_1 \sin k_1(x - x_1) & \cos k_1(x - x_1)
\end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}.
\]

At \( x = x_1 \), the 2×2 matrix is the unit matrix so that the column vector having components \( A \) and \( B \), represents \( \Psi(x_1) \). Thus,

\[
\Psi(x) = \begin{pmatrix}
\cos k_1(x - x_1) & \frac{1}{jn_1} \sin k_1(x - x_1) \\
-jn_1 \sin k_1(x - x_1) & \cos k_1(x - x_1)
\end{pmatrix} \Psi(x_1) = T_1(x - x_1) \Psi(x_1),
\]

and the field vector at \( x = x_2 \) is

\[
\Psi(x_2) = \begin{pmatrix}
\cos k_1 d_1 & \frac{1}{jn_1} \sin k_1 d_1 \\
-jn_1 \sin k_1 d_1 & \cos k_1 d_1
\end{pmatrix} \Psi(x_1) = T_1(d_1) \Psi(x_1). \tag{3.6.6}
\]

Thus, \( T_1 \) transfers \( \Psi(x_1) \equiv \Psi_1 \), given by (3.6.4) to \( \Psi(x_2) \equiv \Psi_2 \), given by (3.6.5). Since the determinant of \( T_1 \), \( \det T_1 = 1 \), and the diagonal terms are identical, the inverse of \( T_1 \) can be found by simply changing the sign of the two off-diagonal terms.

The transfer matrix \( T_1 \) can be factored (See Section 7.4),

\[
T_1 = V_1 X_1 V_1^{-1}, \tag{3.6.8}
\]

where the diagonal elements of \( X_1 \) contain the eigenvalues of \( T_1 \),

\[
X_1 = \begin{pmatrix}
e^{-j\theta_1} & 0 \\
0 & e^{j\theta_1}
\end{pmatrix}, \tag{3.6.9}
\]

where \( \theta_1 = k_1 d_1 \) is the film thickness in radians. The eigenvectors of \( T_1 \) are the column elements of \( V_1 \),

\[
V_1 = \begin{pmatrix}
1 & -1/n_1 \\
n_1 & 1
\end{pmatrix}. \tag{3.6.10}
\]

The reflection/transmission equation can be written as

\[
V_1^{-1} \Psi_2 = X_1 V_1^{-1} \Psi_1, \tag{3.6.11}
\]

so that the two vectors \( \Psi_1 \) and \( \Psi_2 \) are transformed using the inverse of the eigenvector matrix. More generally, the mapping \( \Psi \rightarrow \Phi \) uses the linear transformation \( \Psi = V_1 \Phi \). This transformation uncouples the vector components of \( \Phi \) in the new transfer matrix equation which can now be written in the canonical form

\[
\Phi_2 = X_1 \Phi_1. \tag{3.6.12}
\]
The final canonical form of the reflection/transmission equation is

\[ \tau \left( \frac{n_2 + n_1}{n_2 - n_1} \right) = X_1 \left( \frac{(n_1 + n_0) + (n_1 - n_0)\rho}{(n_0 - n_1) - (n_1 + n_0)\rho} \right). \quad (3.6.13) \]

The reflection and transmission coefficients become

\[ \rho = \frac{(n_2 + n_1)(n_0 - n_1) + (n_0 + n_1)(n_1 - n_2)e^{-j2\theta_1}}{(n_2 + n_1)(n_0 + n_1) + (n_0 - n_1)(n_1 - n_2)e^{-j2\theta_1}}, \quad (3.6.14) \]

\[ \tau = \frac{4n_1 n_0 e^{-j\theta_1}}{(n_2 + n_1)(n_0 + n_1) + (n_0 - n_1)(n_1 - n_2)e^{-j2\theta_1}}. \quad (3.6.15) \]

It should be noted that the reflection and transmission coefficients can also be written in terms of the elements of the transfer matrix as illustrated in Problem 3–7.

The resulting reflection and transmission coefficients for a quarter-wave film, \( e^{j2\theta_1} = -1 \), and when the refractive index of the film is \( n_1^2 = n_2 n_0 \), the reflection coefficient \( \rho = 0 \) so that there is a complete transmission of incident power into Region 2. Generally, it is impossible to match a film’s refractive index, \( n_1 = \sqrt{n_0 n_2} \), to the materials in Regions 2 and 0. For example, if \( n_2 = 1 \) and \( n_0 = 3.6 \), the matching index of the film is \( n_1 \approx 1.89737 \), however, it is frequently possible to engineer a material whose refractive index is near that of the calculated index.

![Figure 3.6](image)

Figure 3.6. The power reflectivity, \( R = |\rho|^2 \), of a plane wave incident upon a GaAs-air interface at normal incidence with a film between the two regions. The quarter-wave film has \( \theta_1 = \pi/2 \), while the half-wave film has \( \theta_1 = \pi \).

Figure 3.6 shows that the power reflection about the minimum point is rather soft so that there is a broad range of indices that will give results close to the optimum value. The calculated reflection coefficient in the figure is obtained\footnote{Final expressions were simplified via Mathematica. See Problem 3–2.}
as follows: as the refractive index \( n_1 \) varies, the film thickness tracks the variation in order to maintain the quarter- and half-wavelength thicknesses for the curves shown in the figure. Note that the half-wavelength film does not alter the reflection coefficient from its value with no film present.

### 3.6.2 Single-Layer Fields

The electromagnetic fields at the inter-layer boundaries is simply determined from the fields at another boundary. For example, the fields at \( x = x_2 \) are determined from the fields at \( x = x_1 \) and vice versa, using (3.6.11) or its canonical form (3.6.13). The transforming of the fields from the semi-infinite “input” layer to semi-infinite “output” layer, is equivalent to solving a boundary-value problem; the unknown variables are \( \rho \) and \( \tau \). Now the fields inside the single layer can be determined using (3.6.6) after factoring

\[
\begin{pmatrix}
\cos k_1(x - x_1) & \frac{1}{jn_1} \sin k_1(x - x_1) \\
-jn_1 \sin k_1(x - x_1) & \cos k_1(x - x_1)
\end{pmatrix} = V_1 \begin{pmatrix}
e^{-jk_1(x-x_1)} & 0 \\
0 & e^{jk_1(x-x_1)}
\end{pmatrix} V_1^{-1},
\]

which can be written as

\[
\Phi(x) = \begin{pmatrix}
e^{-jk_1(x-x_1)} & 0 \\
0 & e^{jk_1(x-x_1)}
\end{pmatrix} \Phi(x_1).
\] (3.6.16)

The first element of the vector represents forward wave while the second represents a backward wave in the layer.

### 3.6.3 Transfer Characteristics of Double Layers

Vertical surface emitting laser structures are fabricated with several alternating layers of two different materials to form a highly reflecting mirror. To understand the reflection characteristics of multi-layer stacks, it is necessary to evaluate the performance of a simple two-layer region (ham and cheese) surrounded by two semi-infinite regions (the bread). Figure 3.7 shows the four-layer structure.

**Figure 3.7.** A plane wave is incident on a two-layer slab region. For non-magnetic material, the optical thicknesses of Layers 1 and 2 are \( \theta_1 = k_0 n_1 d_1 \) and \( \theta_2 = k_0 n_2 d_2 \) respectively. The two-layer thickness is \( d_1 + d_2 = \Lambda \).
3.6 STRATIFIED MEDIA

The transfer matrix of the two-layer combination which is labeled the unit cell, has \( T_u = T_2 T_1 \),

\[
T_u = \begin{pmatrix} A_u & B_u \\ C_u & D_u \end{pmatrix},
\]

which has the following elements (See Table 7.5.),

\[
A_u = \cos \theta_1 \cos \theta_2 - \frac{n_1}{n_2} \sin \theta_1 \sin \theta_2, \quad B_u = \frac{1}{j n_1} \sin \theta_1 \cos \theta_2 + \frac{1}{j n_2} \cos \theta_1 \sin \theta_2, \\
C_u = -j n_1 \sin \theta_1 \cos \theta_2 - j n_2 \cos \theta_1 \sin \theta_2, \quad D_u = \cos \theta_1 \cos \theta_2 - \frac{n_2}{n_1} \sin \theta_1 \sin \theta_2.
\]

Using the transfer matrix of the unit cell, the two-layer structure can be treated as a single layer, and factoring the transfer matrix produces

\[
T_u = V_u X_u V_u^{-1}.
\]

The eigenvalues of \( T_u \) are elements of the diagonal matrix \( X_u \), while the associated eigenvectors are the columns of \( V_u \). The eigenvalue matrix can be written as

\[
X_u = \begin{pmatrix} \chi_1 & 0 \\ 0 & \chi_2 \end{pmatrix} = \begin{pmatrix} e^{-j \theta_u} & 0 \\ 0 & e^{j \theta_u} \end{pmatrix},
\]

where \( \chi_1 \) and \( \chi_2 \) are the eigenvalue of \( T_u \). The eigenvalues can also be written in terms of \( \theta_u \equiv \beta \Lambda \), defined as the the optical thickness of the unit cell which satisfies

\[
\cos \theta_u = \frac{1}{2} (A_u + D_u).
\]

The term \( \beta \) can be interpreted as the propagation constant of an electromagnetic wave in the two-layer structure. Thus, \( \beta \Lambda \) is the phase change from \( z_1 \) to \( z_2 \). Some of the most important features and characteristics of the unit cell are dependent only on its optical thickness, \( \theta_u \).

When the matrix elements \( A_u \) and \( D_u \) are given by (3.6.18) and (3.6.21) respectively, the equation defining the optical thickness becomes

\[
\cos \theta_u = \cos \theta_1 \cos \theta_2 - \frac{1}{2} (\frac{n_2}{n_1} + \frac{n_1}{n_2}) \sin \theta_1 \sin \theta_2.
\]

For two half-wavelength layers \( \cos \theta_u = 1 \) whereas two quarter-wavelength layers, \( \cos \theta_u = -(n_2/n_1 + n_1/n_2)/2 < -1 \) which results in a complex value of \( \theta_u \). (See Problem 3–3.) It is interesting to note that \( \cos \theta_u \) has a minimum (maximum negative value) for \( \theta_1 = \theta_2 = \pi/2 \), i.e., both layers have a quarter-wavelength thickness. On the other hand, \( \cos \theta_u \) has a maximum value when both layers have half-wavelength thicknesses.

In practice, layer thicknesses are designed for a certain wavelength such as for a quarter wavelength or a half wavelength. For example, if layers are designed to have quarter-wavelength optical thicknesses at \( \lambda_q = 2\pi/k_{aq} \), then
\( \theta_1 = k_o n_1 d_1 = \pi/2 \) and \( \theta_2 = k_o n_2 d_2 = \pi/2 \). (These conditions imply \( n_1 d_1 = n_2 d_2 \)) It is interesting to let the optical thicknesses of the two layers be identical, i.e., \( \theta_1 = \theta_2 = \theta = \pi/2 + \delta \), where \( \delta \) represents the change in optical thickness from its quarter-wavelength value. Thus, as the operational wavelength varies from its quarter-wavelength condition, the optical thickness of the unit cell, \( \theta_u \), varies, and optical thicknesses of the individual layers change in unison. For the discussion below, the refractive index \( n_2 \) will be referenced to \( n_1 \) as 

\[ n_2 = n_1(1 + \delta_u), \quad (3.6.26) \]

where \( \delta_u \) represents the fractional difference (relative to \( n_1 \)) between the two values of the unit cell. Now the elements of the transfer matrix can be expressed in terms of \( \delta_u \) and \( n_1 \), however, the optical thickness of the unit cell, \( \theta_u \), depends only on \( \delta_u \), and of course, the optical thickness \( \theta \) of the individual layers. Since, \( d_1 = (1 + \delta_u) d_2 \), the optical thickness, \( \theta \), of the individual layers can be written as 

\[ \theta = k_o n_1 \frac{1 + \delta_u}{2 + \delta_u} \Lambda. \quad (3.6.27) \]

The transfer matrix elements given in (3.6.18)-(3.6.21), become 

\[ A_u = \cos^2 \theta - \frac{1}{1 + \delta_u} \sin^2 \theta, \quad (3.6.28) \]

\[ B_u = \frac{1}{jn_1} \left( 1 + \frac{1}{1 + \delta_u} \right) \sin \theta \cos \theta, \quad (3.6.29) \]

\[ C_u = -jn_1(2 + \delta_u) \sin \theta \cos \theta, \quad (3.6.30) \]

\[ D_u = \cos^2 \theta - (1 + \delta_u) \sin^2 \theta. \quad (3.6.31) \]

The optical thickness of the unit cell in terms of the optical thickness of the individual layers satisfies 

\[ \cos \theta_u = \cos^2 \theta - \frac{1}{2} \left( 1 + \delta_u + \frac{1}{1 + \delta_u} \right) \sin^2 \theta, \quad (3.6.32) \]

while the eigenvector matrix can be written as 

\[ V_u = \begin{pmatrix} 1 & v_x/n_u \\ v_y/n_u & 1 \end{pmatrix}, \quad (3.6.33) \]

where \( n_u = \sqrt{A_u/B_u} = \sqrt{n_1 n_2} = n_1 \sqrt{1 + \delta_u} \) is independent of \( \theta \) whereas, the mantle functions \( v_x(\theta) \) and \( v_y(\theta) \) are dependent on \( \theta \) and are discussed later in this section.

The cosine of the optical thickness of the unit cell \( \theta_u \) can be computed from (3.6.32) when the optical thicknesses, \( \theta \), of the cell layers and the fractional index step in the unit cell are specified. The value of \( \theta_u \) will be real when the relation between the optical layer thicknesses, \( \theta \), and \( \theta_c \) satisfies 

\[ \sin \theta < \sin \theta_c, \quad (3.6.34) \]
where the value of $\theta$ at “cutoff” satisfies
\[
\sin \theta_c = \frac{2\sqrt{1 + \delta_u}}{2 + \delta_u}.
\] (3.6.35)

In the “stop” band, $\theta_u$ has both real and imaginary parts when the layer thickness $\theta$ satisfies $\theta_c < \theta < \pi - \theta_c$. This condition occurs when the RHS of (3.6.32) is more negative than $-1$.

Figure 3.8 shows the mapping of the complex optical thickness $\theta_u = \xi + j\eta$ where the optical thickness $\theta$ is parameter. Starting at the origin in the $\theta_u$ plane, $\theta = 0$. As $\theta$ increases from 0 to $\theta_c$, $\theta_u$ varies along the real axis from $0 + j0$ to $\pi + j0$. As $\theta$ varies from $\theta_c$ to $\pi/2$, $\theta_u$ varies from $\pi + j0$ to $\pi + j\eta_{\text{min}}$, where $\eta_{\text{min}} = -\cosh^{-1}\{[1 + (1 + \delta_u)^2]/[2(1 + \delta_u)]\}$. 

The region specified by (3.6.34) will be designated as the first “pass-band” region while the region defined by $\theta_c < \theta < \pi - \theta_c$ will be designated as the first “stop-band” region. Figure 3.9 illustrates the two regions. The boundary between the regions is defined as $\theta_c$, the optical thickness at cutoff, so that the stop-band extends from $\theta_c$ to $\pi/2$ (the central point) to $\pi - \theta_c$. Thus, the width of the stop-band is given by
\[
\theta_{\text{SB}} = \pi - 2\theta_c = 2 \arccos\left(\frac{2\sqrt{1 + \delta_u}}{2 + \delta_u}\right)
\] (3.6.36)

Note that as the index step $\delta_u$ increases, the width of the stop-band increases. For small values of $\delta_u$, the width of the stop-band becomes
\[
\theta_{\text{SB}} \rightarrow \delta_u \quad \text{as} \quad \delta_u \rightarrow 0.
\] (3.6.37)

\[\text{Figure 3.8. A map of the optical thickness of the unit cell in the } \theta_u\text{-Plane where } \theta \text{ is a parameter.}
\]

The value of “$\eta_{\text{min}}$” can be either positive or negative, however, the negative value is chosen so that a propagation condition is satisfied as in Problem 3–3. Thus, the imaginary part of $\theta_u$ must be negative.

\[\text{\dag}\]
The mantle functions \( v_x^2 \) and \( v_y^2 \) in (3.6.33) are given by

\[
v_x^2(\theta) = \frac{\chi_2 - D_u}{\chi_2 - A_u} \quad v_y^2(\theta) = \frac{\chi_1 - A_u}{\chi_1 - D_u}.
\] (3.6.38)

The mantle functions \( v_x^2 \) and \( v_y^2 \) are identical (See Problem 3–5.) so that the single mantle function \( v^2 = v_x^2 = v_y^2 \) will be used in the discussion below. It is interesting to note that \( v^2(\theta) = 0 \) at optical layer thicknesses of \( \theta = \pi/2, 3\pi/2 \ldots \), while \( v^2 = 1 \) at \( \theta = 0, \pi \ldots \) Thus, at the quarter-wave thicknesses, the eigenvector matrix \( V_u \) is the unit matrix and the transformation (3.6.22) becomes \( T_u = X_u \). When, \( \theta \) is small, the expression under the radical sign in (3.6.36) is positive, however, when

\[
\theta = \theta_c = \arcsin \left( \frac{2\sqrt{1 + \delta_u}}{2 + \delta_u} \right).
\] (3.6.39)

The functional dependences of \( v^2(\theta) \) and \( v(\theta) \) can be described from (3.6.38), however, Fig. 3.10 illustrates their behavior in the complex plane.

Table 3.1 gives values of \( v^2 \) at specific points that shows the trajectories of \( v^2 \) and \( v \) in the complex plane. These specific points can be obtained from the
3.6 STRATIFIED MEDIA

The complex function \( v^2 \) progresses, as \( \theta \) increases, in a clockwise direction when the index parameter \( \delta_u \), given by (3.6.26), is positive, i.e., \( n_2 > n_1 \), however, the trajectories are counter clockwise when \( n_2 < n_1 \).

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( \theta_u )</th>
<th>( v^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0^\pi )</td>
<td>( 0^\pi )</td>
<td>( 3{v^2} &lt; 0 )</td>
</tr>
<tr>
<td>( \theta_C )</td>
<td>( \pi )</td>
<td>( -1 )</td>
</tr>
<tr>
<td>( \pi/2 )</td>
<td>( \pi + j\eta_{\text{min}} )</td>
<td>0</td>
</tr>
<tr>
<td>( \pi - \theta_C )</td>
<td>( \pi )</td>
<td>( -1 )</td>
</tr>
<tr>
<td>( \pi )</td>
<td>( 2\pi )</td>
<td>1</td>
</tr>
</tbody>
</table>

† For \( \delta_u > 0 \).

Many applications use the “quarter-wavelength” designs to form the unit cells. This occurs when optical thicknesses of the layers satisfy \( D = 2C \).

To study the wavelength characteristics of these quarter-wavelength structures about their design target, put \( D = 2C + \Delta \), where \( \Delta \) is the change in optical thickness about the design point. The transform matrix elements over the unit cell are

\[
A_u(\theta) = \sin^2 \theta - \frac{1}{1 + \delta_u} \cos^2 \theta, \tag{3.6.40}
\]

\[
B_u(\theta) = \frac{1}{j\eta_1}(1 + \frac{1}{1 + \delta_u}) \cos \theta \sin \theta, \tag{3.6.41}
\]

\[
C_u(\theta) = -j\eta_1(2 + \delta_u) \cos \theta \sin \theta, \tag{3.6.42}
\]

\[
D_u(\theta) = \sin^2 \theta - (1 + \delta_u) \cos^2 \theta. \tag{3.6.43}
\]

The thickness of the unit cell satisfies

\[
\cos \theta_u = \sin^2 \theta - \frac{1}{2}(1 + \delta_u + \frac{1}{1 + \delta_u}) \cos^2 \theta, \tag{3.6.44}
\]

and the mantle function used to define the eigenvectors is obtained by transforming the independent variable in (3.6.38). The point corresponding to \( \theta = 0 \) is at the origin of Fig. 3.10 (b). Notice the symmetry of the \( v(\theta) \) from the figure such as \( v(\theta) = -v(-\theta) \), in the stop band, however, more specifically, \( v^*(-\theta) = v(\theta) \).

3.6.4 Double-Layer Dispersion

As seen from Fig. 3.8 the relation between \( \theta \) and \( \theta_u \) is repetitive with stopbands whose centers occur at \( \theta = \pi/2, 3\pi/2, 5\pi/2, \ldots \). Thus, the dispersion curve, \( \theta \) vs \( \theta_u \). In the limiting case of \( \delta_u = 0 \), the dispersion curve is a straight line through the center of the stop band region. Figure 3.11 illustrates the dispersion characteristics for the step parameter \( \delta_u = 0.1 \). When the refractive index step between layers 1 and 2 is small, i.e. \( (n_2 - n_1)/n_1 = \delta_u << 1 \), then the matrix eigenvalue at \( \theta = \pi/2 \) becomes

\[
e^{-j\theta_u} = e^{-j\pi + \eta_u} \approx -e^{-\delta_u}.
\]
For example, the minimum value of $\eta_u$ in Fig. 3.11 is slightly larger than the approximate value of $-0.1$.

Figure 3.11. The dispersion curve, $\theta$ vs $\theta_u = \xi_u + j\eta_u$ for the double-layer structure, near the quarter-wavelength condition. The fractional index step in the unit cell is $\delta_u = 0.1$. The stop-band width is given by (3.6.36). The imaginary part of $\theta_u$, $\eta_u$, is zero outside the stop band.

Using the technique developed earlier for the single layer sandwiched between two semi-infinite regions, the transfer equation becomes

$$V_u^{-1}\Psi_3 = X_u V_u^{-1}\Psi_1,$$

which results in transferring the field vector at $x = x_1$ to the field vector at $x = x_3$. The canonical form of the transfer equation is

$$\Phi_2 = X_u \Phi_1.$$  

The resulting reflection and transmission coefficients computed from (3.6.45) become

$$\rho = \frac{(vn_3 + n_u) (n_0 - vn_u) + (vn_0 + n_u) (vn_u - n_3) e^{-j2\theta_u}}{(vn_3 + n_u) (n_0 + vn_u) + (vn_0 - n_u) (vn_u - n_3) e^{-j2\theta_u}},$$

$$\tau = \frac{2 (1 + v^2) n_0 n_u e^{-j\theta_u}}{(vn_3 + n_u) (n_0 + vn_u) + (vn_0 - n_u) (vn_u - n_3) e^{-j2\theta_u}}.$$  

3.6.5 Double-Layer Fields

As mentioned above, the fields at the inter-layer boundaries can be determined by appropriately solving the boundary-value problem by matching the “input” fields to the “output fields.” In the case of the unit cell with two layers formed by layers 1 and 2 of Fig. 3.7, the field vectors $\Psi_1$ and $\Psi_3$ are known after $\rho$ and $\tau$ have been computed. The fields in layer regions are

$$\Psi(x) = \begin{cases} 
T_1(x - x_1) \Psi(x_1), & \text{for } x_1 < x < x_2, \\
T_2(x - x_2) T_1(d_1) \Psi(x_1), & \text{for } x_2 < x < x_3. 
\end{cases}$$
3.7 Periodic Stacks

Attention is now focused on the solution of Maxwell’s equations with multiple dielectric layers where alternating layers have identical dielectric constants. These “double layers” form an optical stack that has some unique reflection/transmission characteristics. Vertical cavity surface emitting lasers (VCSELS) use optical stacks to enclose the active region of the laser. With proper design of the stacks, the power reflectivity can approach almost 100 percent.

The major parameter that governs the effectiveness of the periodic stack is the optical thickness of the unit cell that is formed by two adjacent layers. The two layer stack is illustrated in Fig. 3.7, while the multilayered stack is shown in Fig. 3.12.

For non-magnetic material, the optical thicknesses of Layers 1 and 2 are \( \theta_1 = k_\alpha n_1 d_1 \) and \( \theta_2 = k_\alpha n_2 d_2 \) respectively. There are \( N \) layers in the total structure, and \( N_u \) unit cells composed of double layers.

The stack structure shown Fig. 3.12 has \( N_u = (N - 2)/2 \) unit cells since there are two layers per unit cell. The transfer of the field across \( N_u \) unit cells, starting at \( x = x_1 \) is given by

\[
V_u^{-1} \Psi_{N-1} = X_u^{N_u} V_u^{-1} \Psi_0, \tag{3.7.1}
\]

while the canonical form of the transfer equation is

\[
\Phi_{N-1} = X_u^{N_u} \Phi_0. \tag{3.7.2}
\]

The difference between transferring the field across a single unit cell and \( M \) unit cells is the exponent of the eigenvalue matrix. Indeed, the eigenvalue matrix becomes

\[
X_u^{N_u} = \begin{pmatrix}
    e^{-jN_u \theta_0} & 0 \\
    0 & e^{jN_u \theta_0}
\end{pmatrix}, \tag{3.7.3}
\]

and \( e^{j\theta_u} \) can be recognized as an equivalent “Floquet-Bloch” multiplier.

Equation (3.7.2) implies that the field in each unit cell is reproduced and that the overall field is modified by the Floquet-Bloch multiplier. In the typical solution of Floquet-Bloch modes in a periodic structure, the modal fields are determined for an infinite periodic structure while in this formulation, the
fields are computed for a finite-length periodic structure. The fields (components of \( \Phi \)) at some intermediate point that transforms across \( M \) unit cells, say from \( x = x_2(M+m) \) to \( x = x_2m \), can be easily written as

\[
\begin{align*}
\phi_1(x_2(m+M)) &= \phi_1(x_2m) e^{-jM\theta_u}, \\
\phi_2(x_2(m+M)) &= \phi_2(x_2m) e^{jM\theta_u}.
\end{align*}
\]

The first equation represents a Floquet-Bloch field propagating in the positive \( x \) direction while the second characterizes the field propagating in the negative \( x \) direction. Thus, \( \phi_1(x) \equiv \phi^+(x) \), while \( \phi_2(x) \equiv \phi^-(x) \). The field distribution in each unit cell is repeated across each unit cell, i.e., the fields are periodic with a period of \( \Lambda = d_1 + d_2 \). The Floquet-Bloch coefficient, \( e^{-j\theta_u} \), modifies the overall field distribution in each unit cell.

### 3.7.1 Phase and Group Velocities

Although the designation of the unit cell as described has boundaries located at the interfaces of the layers as in Fig. 3.12, the unit cell could be designated as one over any period of length \( \Lambda \). The choice of the unit cell will affect the eigenvector matrix, \( V_\mu \), however, the eigenvalue matrix \( X_\mu \) is independent of the unit cell’s starting point. Accordingly, the vectors \( \hat{\Phi}_f \) are independent of the unit-cell position. In the spirit of Floquet-Bloch waves, phase and group velocities can be determined from

\[
\frac{\omega \Lambda}{\theta_u},
\]

while the group velocity is

\[
v_g = \frac{1}{d_\beta/d\omega} = \frac{\omega (1 + \delta_u)}{(2 + \delta_u)^2} \frac{\sin \theta_u}{\theta \cos \theta \sin \theta}. \tag{3.7.7}
\]

For \( \theta_u < \pi \), and as \( \theta_u \to \pi \), \( v_g \to 0 \). The dispersion curve shown in Fig. 3.11, illustrates that \( v_g = 0 \) at the edges of the stop band. While the group velocity is associated with power flow, in the pass-band regions, both phase and group velocities lose their meaning in the stop band regions.

### 3.7.2 Reflection and Transmission

The reflection and transmission coefficients can now be determined using the input and output vectors given by

\[
\Psi_1 = E_i \left( \frac{1 + \rho}{n_0(1 - \rho)} \right) , \quad \Psi_{N-1} = \tau E_i \left( \frac{1}{n_{N-1}} \right) . \tag{3.7.8}
\]

have forms almost identical to those given in (3.6.47) and (3.6.48):

\[
\begin{align*}
\rho &= \frac{(v n_{N-1} + n_u) (n_0 - v n_u) + (v n_0 + n_u) (v n_u - n_{N-1}) e^{-j2N_u\theta_u}}{(v n_{N-1} + n_u) (n_0 + v n_u) + (v n_0 - n_u) (v n_u - n_{N-1}) e^{-j2N_u\theta_u}} \tag{3.7.9} \\
\tau &= \frac{2 (1 + v^2) n_0 n_u e^{-jN_u\theta_u}}{(v n_{N-1} + n_u) (n_0 + v n_u) + (v n_0 - n_u) (v n_u - n_{N-1}) e^{-j2N_u\theta_u}} \tag{3.7.10}
\end{align*}
\]
In the “stop-band” region, \( \theta_u \) has an imaginary part that forces an exponential growth/decay of the Floquet-Bloch multipliers. Specifically, \(|e^{-jN_u\theta_u}| \to 0\) as \( N_u \to \infty \) and (3.7.9) implies \( \rho \to 1 \).

The peak value of the reflection coefficient \( R > 0.9 \), when 20 unit cells are used while \( R \approx 0.8 \) for 15 cells.

When layers 1 and 2 have quarter-wavelength thicknesses, \( \nu = 0 \). The unit-cell optical thickness \( \theta_u = \pi + j\eta_{\text{min}} \). (See Problem 3–3.) so that the reflection coefficient becomes

\[
\rho = \frac{n_0 - n_{N-1}e^{2N_u\eta_{\text{min}}}}{n_0 + n_{N-1}e^{2N_u\eta_{\text{min}}}}, \tag{3.7.11}
\]

where \( \eta_{\text{min}} \approx -\delta_u \). Conversely, it is possible to approximate the number of layers that will yield a particular value of reflection in the stop band. The value of \( N_u \) obtained from (3.7.11) can be written as

\[
N_u := \frac{1}{2\delta_u} \ln \left( \frac{n_{N-1}}{n_0} \right) + \ln \left( \frac{1 + \rho}{1 - \rho} \right). \tag{3.7.12}
\]

(The operator := can be read as “integer value of.”) For example, if \( n_0 = n_{N-1} \), and \( \delta_u = 0.1 \), and reflection of \( R = 0.9 = \rho^2 \), the number of unit cells \( N_u > 19 \). The value of \( N_u = 19 \) gives a value of \( R > 0.9 \). For the examples shown in Fig. 3.13 the reflection coefficient \( R = 0.92 \) for 20 unit cells.

It should be noted that small values of \( \theta_{\text{SB}} \) require a large number of unit cells in order to achieve large values of reflectivity, i.e., for a given value of reflectivity,

\[ \theta_{\text{SB}}N_u = \text{constant}, \]

so that \( N_u \) must be large for small \( \theta_{\text{SB}} \) and vice versa.

### 3.8 Field Distributions for Quarter-Wavelength Layers
As noted in the previous section, the field distribution in each unit cell is identical except for the Floquet-Bloch coefficient. The input field to the stack region obtained from (3.7.6) where the values of \( \rho \) is given in (3.7.11), is

\[
\Psi_1 = \frac{2n_0 E_i}{n_0 + n_{N-1} e^{2N_0 \eta_{\text{min}}}} \left( n_{N-1} e^{2N_0 \eta_{\text{min}}} \right).
\]

(3.8.1)

The electric and magnetic fields in the 1st layer is (For the sake of brevity, the \( x \) and \( y \) subscripts on the electric and magnetic fields will be replaced by the layer number.),

\[
\begin{pmatrix}
E_1(x) \\
Z_0 H_1(x)
\end{pmatrix} = \begin{pmatrix}
\cos k_1(x - x_1) & \frac{1}{j n_1} \sin k_1(x - x_1) \\
-j n_1 \sin k_1(x - x_1) & \cos k_1(x - x_1)
\end{pmatrix} \Psi_1.
\]

(3.8.2)

while the field distribution in the next layer is

\[
\begin{pmatrix}
E_2(x) \\
Z_0 H_2(x)
\end{pmatrix} = \begin{pmatrix}
\cos k_2(x - x_2) & \frac{1}{j n_2} \sin k_2(x - x_2) \\
-j n_2 \sin k_2(x - x_2) & \cos k_2(x - x_2)
\end{pmatrix} \begin{pmatrix}
0 & 1/j n_1 \\
-j n_1 & 0
\end{pmatrix} \Psi_1.
\]

(3.8.3)

Note that for very large values of \( N_u \), the electric field has its peak value at input (\( x = 0 \)) whereas, the magnetic field at input has its smallest value. Thus in the limit \( N_u \to \infty \), (3.8.1) shows \( E(x_1) \to 2 E_i \) and \( Z_0 H(x_{N-2}) \to 0 \). On the other hand, because of the second matrix in (3.8.3), the values of \( E \) and \( Z_0 H \) switch. Thus, the electric field has peak values at the input of each unit cell but has a null value at the interface between the two layers forming the unit cell. The magnetic field has null values at the input of each unit cell but is peak values at the interface between the two layers forming the unit cell. Since the magnetic field has null values at the beginning of each unit cell, the Poynting vector is zero so that the power flux is zero which implies that the energies stored in the electric and magnetic fields are oscillating.

For finite-length stacks, there is a transmission of power from the left-hand side of the stack to the right-hand side. Accordingly, the \( x \) component of the time-average Poynting vector at the input, \( x = x_1 \), is

\[
S_x = \frac{2n_0^2 n_{N-1} E_i^2 e^{2N_0 \eta_{\text{min}}}}{Z_0(n_0 + n_{N-1} e^{2N_0 \eta_{\text{min}}})^2}.
\]

(3.8.4)

Because the material in the layers has no ohmic losses, the Poynting vector should be independent of position along the \( x \) direction, so that (3.8.4) represents the power transmitted to the output layer.

The time-average energy stored in the electric and magnetic fields in the input unit cell can be computed using field distributions given in (3.8.2) and (3.8.3). For the electric field

\[
W_e = \frac{\pi \varepsilon_o n_1 (2 + \delta_u)}{4 k_\infty (1 + \delta_u)} \left[ 1 + (1 + \delta_u) \left( \frac{n_{N-1}}{n_1} \right)^2 e^{2N_0 \eta_{\text{min}}} \right] \left( \frac{n_0 E_i}{n_0 + n_{N-1} e^{2N_0 \eta_{\text{min}}}} \right)^2.
\]

(3.8.5)
The time-average energy stored in the magnetic field is $W_m = W_e$, so that the average per unit volume over the unit cell is $w_{avg} = 2W_e/\Lambda$. In the limit, $N_u \to \infty$, the Poynting vector $S_x \to 0$, and the energy stored per unit area in the first cell is

$$W_1 = \frac{\pi n_1 (2 + \delta_u) \rho_1}{\omega n_0 (1 + \delta_u)} P_i,$$

(3.8.6)

where $P_i$ is the incident power.

The total energy stored in the semi-infinite stack region can be determined by summing the energy storage in each of the unit cells. Because the input vector to the $m$th unit cell is

$$\Phi_m = X_u^m \Phi_1,$$

the energy stored in the $n$th cell is

$$W_m = W_1 e^{2m \eta_{\text{min}}}. $$

Thus, geometric sum gives the total energy stored in the semi-infinite stack

$$W_s = \frac{W_1}{(1 - e^{2 \eta_{\text{min}}})}.$$
Figure 3.14. Two plane waves incident on a central slab region. Waves traveling toward the central region are designated as + while those traveling away the central region are designated as −. Although the central region may consist of multiple layers, the figure indicates a single layer.

In the absence of $E_2^+$ the exciting field $E_0^+$ produces the reflected field $E_0^-$ and the transmitted field $E_0^-$. The reflection and transmission coefficients which were calculated earlier are $\rho_0 = E_0^- / E_0^+$, while the transmission coefficient is $\tau_2 = E_2^- / E_0^+$. In the absence of $E_0^+$ the exciting field $E_2^+$ produces the reflected field $E_2^-$ and the transmitted field $E_0^-$, where $\rho_2 = E_2^- / E_2^+$ and $\tau_0 = E_0^- / E_2^+$. If all the layers are linear with respect to the field intensity, the two processes are additive.

The electric and magnetic fields in different layers will be normalized normalized according the the wave impedance of the layer. For the sake of simplicity, the electric field will have only an $x$ component while the associated magnetic field has only a $y$ component. In particular, for the $\ell$th layer ($\ell = 0, 2$)

$$\mathcal{E}_\ell^\pm = \sqrt{Y_\ell} E_\ell^\pm, \quad (3.9.1)$$
$$\mathcal{H}_\ell^\pm = \sqrt{Z_\ell} H_\ell^\pm, \quad (3.9.2)$$

where $Y_\ell$ and $Z_\ell$ are the admittance and impedance respectively of the $\ell$th layer. As a result of the normalization, $\mathcal{E}_\ell^+ \mathcal{E}_\ell^-$ represents the power flow toward the central slab region. Assume the incident fields carrying power toward the central slab region are given by

$$\mathcal{E}_0^+ = a_0 e^{-jk_0(x-x_1)}, \quad (3.9.3)$$
$$\mathcal{E}_2^+ = a_2 e^{jk_2(x-x_2)}, \quad (3.9.4)$$

while the fields carrying power away from the central slab region are given by

$$\mathcal{E}_0^- = b_0 e^{jk_0(x-x_1)}, \quad (3.9.5)$$
$$\mathcal{E}_2^- = b_2 e^{-jk_2(x-x_2)}, \quad (3.9.6)$$
then the vector $a = \text{col}(a_0, a_2)$ is related to the vector $b = \text{col}(b_0, b_2)$ by the linear equation

$$b = S a, \quad (3.9.7)$$

where $S$ is the scattering matrix

$$S = \begin{pmatrix} s_{00} & s_{02} \\ s_{20} & s_{22} \end{pmatrix}. \quad (3.9.8)$$

Since $a^H \cdot a$ represents the power flowing toward the central slab region while $b^H \cdot b$ represents the power flowing away from the central slab region, then if the layers are lossless, $S^H S = I$, the unit matrix. Thus

$$S^H S = \begin{pmatrix} s_{00} & s_{02} \\ s_{02} & s_{22} \end{pmatrix} \begin{pmatrix} s_{00} & s_{02} \\ s_{20} & s_{22} \end{pmatrix} = \begin{pmatrix} (s_{00}^* s_{00} + s_{20}^* s_{02}) & (s_{00}^* s_{02} + s_{20}^* s_{22}) \\ (s_{02}^* s_{00} + s_{22}^* s_{20}) & (s_{02}^* s_{02} + s_{22}^* s_{22}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.9.9)$$

The scattering coefficients are related to the reflection and transmission coefficients calculated earlier. For example, if $a_2 = 0$ and $a_0 = 1$, then

$$\frac{E_0^-}{E_0^+} = s_{00} = \frac{E_0^-}{E_0^+} = \rho_0, \quad (3.9.10)$$

$$\frac{E_2^-}{E_2^+} = s_{20} = \frac{Y_2}{Y_0} \tau_2, \quad (3.9.11)$$

whereas, placing $a_2 = 1$ and $a_0 = 0$,

$$\frac{E_0^-}{E_2^+} = s_{02} = \frac{Y_0}{Y_2} \tau_0, \quad (3.9.12)$$

$$\frac{E_2^-}{E_2^+} = s_{22} = \rho_2. \quad (3.9.13)$$

The off-diagonal term in (3.9.9) gives the following relation between the reflection and transmission coefficients

$$\frac{|Y_0|}{|Y_2|} \rho_0^* \tau_0 + \tau_2^* \rho_2 = 0, \quad (3.9.14)$$

while the unit value of the diagonal elements expresses conservation of power.

When the outside layers are also lossless, the wave admittances are real. In addition, the ratio of the outside admittances are equivalent to the ratio of the cladding wave numbers. Namely,

$$\frac{Y_0}{Y_2} = \frac{k_0}{k_2}.$$
Accordingly, the scattering parameters satisfy
\[ k_0 \rho_0^2 \tau_0 + k_2 \tau_2^2 \rho_2 = 0. \] \hfill (3.9.15)

Although the computed reflection and transmission coefficients for the various multi-layer structures of this chapter are relatively complicated, the simple expression given in (3.9.15) can be used to verify the computations. It should be noted that the reflection and transmission coefficients, \( \rho_0 \) and \( \tau_2 \) were computed only for left-to-right incident fields. The coefficients \( \rho_2 \) and \( \tau_0 \) are pertinent to a right-to-left incident field.

---

**Problems**

3–1 In Section 3.4, a general polarized field may have more than one transverse component. For example, when the field is propagating in the positive \( z \) direction, both \( x \) and \( y \) components of the electric and magnetic fields may be present at any point along the \( z \) axis. Show that the electric and magnetic fields are orthogonal.

3–2 Derive the canonical form of the reflection/transmission matrix as given in (3.6.13). Solve for the reflection and transmission coefficients. Assuming a film thickness such that \( e^{i\theta} = 1 \), show that the reflection and transmission coefficients are
\[ \rho = \frac{n_2 - n_0}{n_2 + n_0}, \quad \tau = \frac{2n_0}{n_2 + n_0}. \]

The listing below is a Mathematica sequence for solving for \( \rho \) and \( \tau \), (3.6.14) and (3.6.15).

```
in[1]:= T /EThetaual /LeftBrace2/RightBrace2Cos/LeftBracket1Theta/RightBracket1, ... /Slash1. tr/RightBracket1/RightBracket1/RightBracket1, /Expand 2 /ImaginaryTheta, Factor/RightBracket1

in[2]:= Ev = Eigenvalues[T]
in[3]:= X = {{Exp[-ITheta], 0}, {0, Exp[ITheta]}}, MatrixForm[X]
in[4]:= Ev = Eigenvectors[T]
in[5]:= Vi = Transpose[{N[Ev[[1]], Ev[[2]]]}], MatrixForm[Vi]
in[6]:= Vi = ((2, 0), (0, 2)).Inverse[Vi], MatrixForm[VI]
in[7]:= S0 = Collect[Vi.(1 + rho. (1 - rho) n0), rho]; MatrixForm[S0]
in[8]:= S2 = Collect[Vi.(tau. n2), tau]; MatrixForm[S2]
in[9]:= tr = Solve[S0 = X.S0, (tau, rho)]
in[10]:= rho = Collect[Expand[Numerator[Simplify[rho / tr]]], e^2PiTheta, Factor]  
           Collect[Expand[Denominator[Simplify[rho / tr]]], e^2PiTheta, Factor]
```

3–3 Let \( \theta_u \) be the optical thickness of the unit cell and assume it satisfies the condition \( \cos \theta_u = a \) where \( a \) is real and \( a > 1 \). In terms of the fractional index of refraction step in the unit cell, \( \delta_u \),
\[ a = \frac{1}{2} \left( 1 + \delta_u + \frac{1}{1 + \delta_u} \right). \]

(a.) If
\[ \cos \theta_u = \cos \theta_1 \cos \theta_2 - a \sin \theta_1 \sin \theta_2 \]
Problems

Show that the maximum and minimum values of \( \cos \theta_u \) occur at

\[ \theta_1 = \theta_2 = 0, \pi/2, \pi, \ldots. \]

(b.) By defining \( \theta_u = \xi_m + j \eta_m \), show that \( \xi_m = \pi \) and the imaginary part, \( \eta_m \), satisfies

\[ \eta_m = -\ln(a + \sqrt{a^2 - 1}) < 0. \]

Show that

\[ e^{i \theta_u} = -e^{-\eta_m} = -(a + \sqrt{a^2 - 1}), \]
\[ e^{-j \theta_u} = -e^{\eta_m} = -(a - \sqrt{a^2 - 1}), \]

are the two eigenvalues of the transfer matrix of the unit cell.

3–4 Using properties of transfer matrix over several layers such as the determinant of \( T \), \( \det T = 1 \), show that the eigenvalues of \( T \), \( \chi_1, \chi_2 \), are

\[ \chi_1, \chi_2 = \left( \frac{A + D}{2} \right) \pm j \sqrt{1 - \left( \frac{A + D}{2} \right)^2}. \]

where the negative sign is for \( \chi_1 \) and positive sign is for \( \chi_2 \). The eigenvectors \( v_1 \) and \( v_2 \) satisfy

\[ Tv_1 = \chi_1 v_1, \quad Tv_2 = \chi_2 v_2. \]

Show that if

\[ v_1 = \begin{pmatrix} 1 \\ y_1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} x_2 \\ 1 \end{pmatrix} \]

then

\[ y_1^2 = \frac{C}{B} (x_1 - A), \quad x_2^2 = \frac{B}{C} (x_2 - D). \]

Note: When there is reflection symmetry about the center of the unit cell, the elements \( A \) and \( D \) are identical and the corresponding expressions for \( y_1 \) and \( x_2 \) are simplified.

3–5 Show that the two mantle functions

\[ v_1^2 = \left( \frac{\chi_1 - A}{\chi_1 - D} \right) \quad v_2^2 = \left( \frac{\chi_2 - D}{\chi_2 - A} \right), \]

are identical. \( \text{Hint:} \) Calculate the value of \( v_1^2 - v_2^2 \) and use \( \chi_1 \chi_2 = 1. \)

3–6 Derive the eigenvector matrix given in (3.6.33) where \( v \) is given in (3.6.38).

Show that as \( \theta \to \pi/2 \), that \( v \to 0. \)

3–7 The transfer matrix over a single layer (or multiple layers) can be used to determine the reflection and transmission coefficients, \( \rho \) and \( \tau \) such as that given in (3.6.14) and (3.6.15). Putting

\[ T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \]
show that the reflection and transmission coefficients are

\[ \rho = \frac{n_0(A + n_2 B) - (C + n_2 D)}{n_0(A + n_2 B) + (C + n_2 D)}, \]
\[ \tau = \frac{2n_0}{n_0(A + n_2 cB) + (C + n_2 D)}. \]

Here \( n_2 \) is the refractive index of the output layer so that for multiple layers, the refractive index \( n_2 \) would be replaced by the refractive index of the output layer.
Chapter 4
Waves in Layered Waveguides

The application of Maxwell’s equation to dielectric waveguides is rich in physical phenomena. The analysis of “open” dielectric waveguides is considerably more complicated than that of “closed” metallic waveguides where the electromagnetic fields are confined in space. The modes in closed metallic waveguides form a discrete spectrum of propagating and evanescent modes while open dielectric wave guides have both a discrete spectrum and a continuous spectrum of propagating (radiation modes) and a continuous spectrum of evanescent modes. While the secular equation characterizing modes of both closed and open waveguides have an infinite number of roots, the modes (defined by the roots) of open waveguides form two types of electromagnetic field distributions: (1) a finite number of roots that characterize “bounded” or proper modes and (2) an infinite number of roots that characterize “unbounded” or improper modes. Oddly enough, the improper mode, often referred to as a leaky mode, plays a major role in characterizing field losses produced by the leakage of energy from a “propagating” field.

4.1 Waveguide Equations
Maxwell’s equations can be written in a simplified form when they are applied to systems of cylindrical waveguide structures. In particular, consider the cylindrical structures with wave propagation along the axis of the cylinder. The electromagnetic fields have a functional dependence of the form

\[ e(x, y, z, t) = E_\nu(x, y) e^{j\omega t - \gamma z}, \quad (4.1.1) \]
\[ h(x, y, z, t) = H_\nu(x, y) e^{j\omega t - \gamma z}, \quad (4.1.2) \]
where \( \gamma = \alpha + j\beta \) is the complex propagation constant. The subscript \( \nu \) pertains to field polarization. The attenuation coefficient \( \alpha \) represents the field amplitude attenuation while \( \beta \) is the wave propagation constant. When the solution to Maxwell’s equations is written as above, it should be noted that this is a special form of solutions to Maxwell’s equations and the form of the solution is true only for specific values of the complex propagation constant \( \gamma \) at a given radian frequency \( \omega \). The value of the complex propagation constant \( \gamma \) can be determined only after the geometry of the waveguide structure is specified. The mathematical technique for determining \( \gamma \) is obtained through application of the boundary conditions of the electromagnetic field quantities.

When the fields of the above equation are substituted into Maxwell’s equations, the following differential equations are obtained:

\[
\frac{\partial E_z}{\partial y} + \gamma E_y = -j\omega \mu H_x, \\
-\gamma E_x - \frac{\partial E_z}{\partial x} = -j\omega \mu H_y, \\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega \mu H_z, 
\]

and

\[
\frac{\partial H_z}{\partial y} + \gamma H_y = j\omega \varepsilon E_x, \\
-\gamma H_x - \frac{\partial H_z}{\partial x} = j\omega \varepsilon E_y, \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega \varepsilon E_z. 
\]

The fields above have two independent solutions. The two solutions are designated by the subscript \( \nu \) corresponding to the two different polarizations of the electromagnetic field. For the open-type guided structures discussed here, the two independent solutions are the transverse electric (TE) and transverse magnetic (TM) fields. Transverse electric fields have no longitudinal component of the electric field whereas the transverse magnetic field has no longitudinal component of the magnetic field. The two independent solutions mean that only two field components are needed to completely specify all six components of the electromagnetic field. Assuming that \( E_z \) and \( H_z \) are specified or determined from a solution of the second order wave equation, the remaining field components become:

\[
E_x = \frac{-1}{k_0^2 \kappa + \gamma^2} \left( \gamma \frac{\partial E_z}{\partial x} + j\omega \mu_\alpha \kappa \mu \frac{\partial H_z}{\partial y} \right), \\
E_y = \frac{1}{k_0^2 \kappa + \gamma^2} \left( -\gamma \frac{\partial E_z}{\partial y} + j\omega \mu_\alpha \kappa \mu \frac{\partial H_z}{\partial x} \right), \\
H_x = \frac{1}{k_0^2 \kappa + \gamma^2} \left( j\omega \varepsilon_\alpha \kappa \varepsilon \frac{\partial E_z}{\partial y} - \gamma \frac{\partial H_z}{\partial x} \right), \\
H_y = \frac{-1}{k_0^2 \kappa + \gamma^2} \left( j\omega \varepsilon_\alpha \kappa \varepsilon \frac{\partial E_z}{\partial x} + \gamma \frac{\partial H_z}{\partial y} \right). 
\]
where, $k_o$ is the free-space wavenumber, and $\kappa = \kappa_e \kappa_m$ is the product of the relative permittivity and permeability. When the material has a non-zero conductivity, the form of the solution will not change from that given by the above equations. However, $\kappa$ will be complex. Most photonic devices are fabricated with non-magnetic materials so that typically, the magnetic permeability $\kappa_m = 1$.

The longitudinal field components $E_z$ and $H_z$ are determined from the solution of the wave equation with appropriate boundary conditions. In addition, all field components including the longitudinal ones can be determined from the solution of the wave equation with appropriate boundary conditions. For example, all field components satisfy the scalar Helmholtz equation

$$\nabla_t^2 \psi(x, y) + \left(k_o^2 \kappa + \gamma^2\right) \psi(x, y) = 0,$$

(4.1.9)

where $\nabla_t^2$ is the transverse Laplacian operator. The transverse wavefunction $\psi(x, y)$ is combined with the propagation term giving the wavefunction $\Psi(x, y, z) = \psi(x, y) e^{-j\gamma z}$. Upon putting $E_z = E_0 \Psi(x, y, z)$, (4.1.9) must be solved with boundary conditions appropriate for $E_z$, and the field for the TM modes can be obtained. On the other hand, putting $H_z = H_0 \Psi(x, y, z)$, the TE fields can be generated.

### 4.2 Wave Definitions

For most geometrical structures, the solution to (4.1.9) is independent for the two different polarizations. For example, many of the modes of an optical fiber must have both an $E_z$ and $H_z$ component. This resulting mode is called a 'hybrid' mode. On the other hand, there is one type of mode that has neither an $E_z$ nor an $H_z$ field. This type of mode is commonly excited in two-wire transmission systems, and called a transverse electromagnetic (TEM) mode.

1. Transverse electromagnetic waves (TEM) have $E_z$ and $H_z = 0$. The total field quantities lie in the transverse (direction of propagation) plane. These waves are usually associated with twin conductors such as coaxial cables.
2. Transverse magnetic waves (TM) have $H_z = 0$ and $E_z \neq 0$.
3. Transverse electric waves (TE) have $E_z = 0$ and $H_z \neq 0$.
4. Hybrid waves occur when both $E_z$ and $H_z$ are non zero. This condition is required because both electric and magnetic field components are necessary to satisfy the boundary conditions. It should be noted that hybrid modes should be distinguished from the case when both $E_z$ and $H_z$ components are present in the field solutions. For hybrid modes, both $E_z$ and $H_z$ are necessary, while for the TE and TM combination modes, the modes are independent.

The TEM waves have neither $E_z$ nor $H_z$ field components so that the transverse components can exist only if $k_o^2 \kappa + \gamma^2 = 0$. This implies the propagation constant $\gamma = j\beta$ where $\beta = k_o \sqrt{\kappa}$. The nature of TEM modes are such that they are not found in optical waveguide systems including two dimensional dielectric waveguides or the three dimensional guides such as the common optical fiber with arbitrary cross sections.
The transverse electric and magnetic waves are common in both hollow metallic structures such as pipes, as well as in dielectric waveguides. The waves can also exist in two conductor metallic cables such as parallel wires or coaxial cables, however, TEM waves are most common for these structures. For two dimensional planar dielectric waveguides, the TE and TM waves are common. In these structures the field components do not vary in the transverse direction parallel to the planar slab.

Finally, hybrid waves which are combination of TE and TM waves are found in the more complicated waveguide systems. For example, in a rectangular waveguide which is partially loaded with a dielectric material, both $E_z$ and $H_z$ are present in the modal field. Hybrid waves exist in optical waveguides such as optical fibers, optical stripe lines, or in any structure which has a cylindrical region with a high dielectric constant, embedded in a region that has a lower dielectric constant. Since optical fibers play an important role in optical communication systems, it is imperative that hybrid modes be well understood.

### 4.3 Slab Waveguides

The three layer slab waveguide is mathematically simple and easy to understand compared to the multiple-layer waveguides and optical fibers. However, the physical waveguide phenomena of the three-layer structure is applicable to the more complicated waveguides. The three-layer slab is formed by three layers of dielectric material with the center layer having the largest index of refraction.

Before discussing the mathematical and physical concepts, we need to mention the various types of modes associated with slab waveguides. A \textit{trapped} or \textit{bounded} mode is one whose field energy is located in the neighborhood of the waveguide central slab region. A propagating mode implies that the electromagnetic energy travels along the waveguide. The electromagnetic energy stays predominantly in the central dielectric layer. The other type of mode satisfying Maxwell’s equations is called a radiation mode. Radiation modes have field intensities that do not vanish at large distances from the slab.

![Figure 4.1](image_url)
The fundamental slab geometry shown in Fig. 4.1, illustrates an optical mode propagating in the axial $z$ direction. (Layer 0 is bounded by $x_0 < x < \infty$, layer 1 is bounded by $x_1 < x < x_0$ and the layer 2 is bounded by $-\infty < x < x_1$. Note that layer 1 starts at $x_1$.) The electromagnetic fields of the slab waveguide can be configured in terms of plane waves propagating along the positive $z$ direction and cocked at an angle with respect to the direction of propagation. For example, in Fig. 4.2 the plane wave is propagating at the angle $\theta$ with respect to the $z$ axis. When the value $\theta' = \pi/2 - \theta$ is larger than the critical angle, the plane wave will be totally reflected at the boundary between layers 1 and 2. This reflection in turn sets up a standing wave along the $x$ direction; however, there will be a traveling wave in the $z$ direction. For a given plane wave inclination $\theta$ and a waveguide width, the plane wave will be reflected from the boundary at the layer 0-1 interface, thus confining the majority of the wave energy to layer 1. There will be a field extension into both layers 0 an 2, but it decays exponentially since the refracted angle is imaginary with a corresponding imaginary propagation constant.

![Layer 0 Layer 1 Layer 2](image)

**Figure 4.2.** The bounded waveguide mode representation using ray optics and the superposition of plane waves.

The waveguide modes of the two-dimensional dielectric structure can be obtained by first solving the differential equation for the $z$ component of the electric or magnetic field depending upon whether the field is TM or TE respectively. The longitudinal fields are found by solving the wave equation with appropriate boundary conditions. Frequently, it is convenient to solve the differential equation of the transverse field components directly. Assuming the fields are constant along the $y$ direction, Maxwell’s equations become

\begin{align}
\gamma E_y &= -j\omega \mu_0 \kappa_m H_x, \\
-\gamma E_x - \frac{\partial E_z}{\partial x} &= -j\omega \mu_0 \kappa_m H_y, \\
\frac{\partial E_y}{\partial x} &= -j\omega \mu_0 \kappa_m H_z
\end{align}

and

\begin{align}
\gamma H_y &= j\omega \epsilon_0 \kappa_e E_x, \\
-\gamma H_x - \frac{\partial H_z}{\partial x} &= j\omega \epsilon_0 \kappa_e E_y, \\
\frac{\partial H_y}{\partial x} &= j\omega \epsilon_0 \kappa_e E_z.
\end{align}
It is clear that TE waves do not have $E_x$ and $H_y$ components while the TM waves do not have $E_y$ and $H_x$ components. Generally, the longitudinal component is first calculated and the transverse fields are written in terms of the axial fields. Frequently, it is advantageous to write the fields in terms of the dominant transverse field component. In particular the TE waves have the $E_y$ component as dominant, whereas the TM wave has the $H_y$ component as dominant. Note that the main field component lies along the $y$ axis. Since all components of the electromagnetic field satisfy the wave equation, it is noticed that the fields $E_y$ and $H_y$ which each satisfy the wave equation, are used for finding the fields of the corresponding TE and TM waves respectively. For the sake of simplicity, the field equations will be developed for non-magnetic material so that the relative permeability $\mu_m = 1$.

### 4.4 Transverse Electric Modes: Formulation

As discussed above, the $E_y$ field component satisfies the the wave equation:

$$\frac{\partial^2 E_y}{\partial x^2} + (k_0^2 k_x + \gamma^2)E_y = 0 \quad (4.4.1)$$

The remaining field components are:

$$H_x = -\frac{j\omega\mu_0}{\gamma} E_y, \quad (4.4.2)$$
$$H_z = -\frac{1}{j\omega\mu_0} \frac{\partial E_y}{\partial x}. \quad (4.4.3)$$

For wave propagation, $\gamma = j\beta$. The quantity $\omega\mu_0 = k_0 Z_o$, where $Z_o = 120\pi$ ohms, is the free-space wave impedance. If the propagation constant $\gamma = j\beta$ is purely imaginary, then the modes represent traveling waves in the positive $z$ direction. The value of the normalized propagation constant $\beta/k_0$ lies between $n_2$ and 0, whereas the attenuation coefficient $\alpha$ lies between 0 and $\infty$.

For propagating modes there are three types of field distribution, as presented in Fig. 4.3:

- **Radiation Mode:** Case II
- **Bounded Mode:** Case I
- **Evanescent Modes:** Case IV

![Figure 4.3. The various types of modes supported by a lossless three-layer waveguide. (a) A map of the complex propagation constant for the various types of modes. The bounded modes are represented by points on the axis, whereas the two types of radiation modes are represented by a continuum of values. (b) The field distribution in the transverse direction.](image)
4.4 TRANSVERSE ELECTRIC MODES: FORMULATION

**Case I** The mode is bounded in positive and negative x directions. The fields decay exponentially in both layers 0 and 2. For this case there are only a finite set of bounded modes as indicated by the filled circles on the $j\beta$ axis. Assuming that the refractive index values satisfy $n_0 < n_2 < n_1$, the bounded modes have a normalized propagation constant $\beta/k$ that lies between $n_0$ and $n_2$, $n_0 < \beta/k < n_2$.

**Case II** The mode is not bounded in the negative x direction. The normalized propagation constant can take on a set of continuous values that lie in the range from $n_1$ to $n_2$.

**Case III** The mode is not bounded on either the positive or negative x directions. The normalized propagation constant lies between $n_0$ and 0.

**Case IV** Evanescent modes have an x dependence similar to Case III radiation modes, however, the propagation constant $\gamma = \alpha$.

While the fields may be unbounded in space, in some cases, the field amplitudes have finite values at all points in space. It is the boundedness of the field amplitudes that allows for mode normalization and thus for arbitrary field representations in terms of the modes.

Since $E_y$ acts as a scalar potential, i.e. the remaining field components are written in terms of $E_y$, the component will be written in terms of the generalized scalar function $\psi(x)$ so that

$$E_y = E_o \psi(x) e^{-\gamma z},$$  \hspace{2cm} (4.4.4)

where $\psi(x)$ satisfies the differential equation

$$\frac{d^2\psi}{dx^2} + (k_o^2 + \gamma^2)\psi = 0.$$  \hspace{2cm} (4.4.5)

Note that the time dependence has not been included in the field expression.

In the solution of multilayer waveguides it is convenient to write the differential equations given by (4.4.1)–(4.4.3) in terms of a set of first-order differential equations. The two first-order equations will be built around the two independent variables $\phi_1$ and $\phi_2$ which are defined as

$$\phi_1(x) = \psi(x), \hspace{0.5cm} \phi_2(x) = \frac{d\psi}{dx}.$$  \hspace{2cm} (4.4.6)

The two components are combined to form the vector

$$\Phi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix}.$$  \hspace{2cm} (4.4.7)

The advantage of this formalism lies in the representation of the fields in the various layers and the transformation of the solution across the various boundaries of multilayer structures. For example, the $E_y$ component must be continuous across each layer; this implies that $\phi_1$ must be continuous. The magnetic field $H_z$ must be continuous across the various layers; this implies that $\phi_2$ must be continuous.
Using the appropriate components of the reduced (assumed no variation along the y direction) Maxwell’s equations given by (4.3.1) and (4.3.2), the two differential equation describing the vector $\Phi$, are given by

\[
\frac{d\phi_1}{dx} = \phi_2 \\
\frac{d\phi_2}{dx} = -[k^2\kappa(x) + \gamma^2] \phi_1
\] (4.4.8)

The above equation can be written in matrix notation as

\[
\frac{d\Phi}{dx} = A(x) \Phi,
\] (4.4.10)

where the matrix $A(x)$ is

\[
A(x) = \begin{pmatrix}
0 & 1 \\
-[k^2\kappa(x) + \gamma^2] & 0
\end{pmatrix} = \begin{pmatrix}
0 & 1 \\
-h^2(x) & 0
\end{pmatrix}
\] (4.4.11)

Notice that the matrix $A$ is a function of $x$, or the refractive index values of the various layers. If the refractive index is constant in a layer, then $A$ is a constant, however, if the refractive index varies with position in a layer such as for a linearly graded refractive index layer then $A$ will reflect that dependence as specified by (4.4.9).

A solution of (4.4.9) can not be implemented until the dielectric constant is specified, however, information pertaining to the solution can be obtained if the $A(x)$ matrix is factored as

\[
A(x) = VXV^{-1} = 
\frac{1}{2} \begin{pmatrix}
1 & 1/j \hat{h}(x) \\
-j \hat{h}(x) & 1
\end{pmatrix} \begin{pmatrix}
-j \hat{h}(x) & 0 \\
0 & +j \hat{h}(x)
\end{pmatrix} \begin{pmatrix}
1 & -1/j \hat{h}(x) \\
j \hat{h}(x) & 1
\end{pmatrix}.
\] (4.4.12)

The elements of the central matrix, $X$ are the eigenvalues of $A$, while the columns of the first matrix $V$ are the eigenvectors of $A$. The matrix on the right, including the $1/2$ coefficient, is the inverse of $V$.

The main advantage to using the description of the system of equations as described by (4.4.10) is that numerical techniques for solving a set of first order differential equations with non-constant coefficients can be easily implemented. Furthermore, it will be easy to implement the solutions for multilayer waveguides using this vector formalism.

4.4.1 Case I: Bounded Modes

The solution of (4.4.10) will now be implemented for the simple three–layer slab structure. The field solution will be determined in each of the three layers.

**CENTRAL LAYER**

Consider the case for *bounded modes* where the dielectric constant of layer 1 is given as $\kappa_1 = n_1^2$. The modes form a a discrete spectrum (Case I) and have
a normalized propagation constant that lies between \( n_1 \) and \( n_2 \). The \( A \) matrix is given by (\( A_1 \) denotes the \( A \) matrix of layer 1.)

\[
A_1 = \frac{1}{2} \begin{pmatrix}
1 & 1/j \hat{h}_1 \\
-j \hat{h}_1 & 1
\end{pmatrix}
\begin{pmatrix}
-j \hat{h}_1 & 0 \\
0 & +j \hat{h}_1
\end{pmatrix}
\begin{pmatrix}
1 & -1/j \hat{h}_1 \\
j \hat{h}_1 & 1
\end{pmatrix}.
\] (4.4.13)

The constant \( \hat{h}_1^2 = k^2 \kappa_1 + \gamma^2 \). The complex propagation constant \( \gamma = j \beta \) has only an imaginary part for the bounded modes. The eigenvalues and the corresponding eigenvectors determine the solution of the set of equations specified by (4.4.10), are

\[
\lambda_{1,a} = -j \hat{h}_1, \quad \lambda_{1,b} = j \hat{h}_1,
\]

for the eigenvalues, and the corresponding eigenvectors are

\[
e_{1,a} = \begin{pmatrix} 1 \\ -j \hat{h}_1 \end{pmatrix}, \quad e_{1,b} = \begin{pmatrix} 1/j \hat{h}_1 \\ 1 \end{pmatrix}.
\]

The eigenvalues and eigenvectors have two subscripts: the first denotes the layer number, whereas the second identifies the eigenvalue/eigenvector. Note that the second order system yields two eigenvalues (denoted as \( a \) and \( b \)) with the associated eigenvectors that satisfy:

\[
A_1 e_{1,a} = \lambda_{1,a} e_{1,a}, \quad A_1 e_{1,b} = \lambda_{1,b} e_{1,b}.
\]

The two characteristic solutions to (4.4.10) are given by:

\[
\Phi_{1,a} = e_{1,a} e^{\lambda_{1,a}(x-x_1)}, \quad \Phi_{1,b} = e_{1,b} e^{\lambda_{1,b}(x-x_1)},
\]

where the solutions of the layer, layer 1, have been referenced to \( x_1 = -d_1 \). The general solution is given by the linear combination

\[
\Phi_1 = C_1 \Phi_{1,a} + D_1 \Phi_{1,b},
\] (4.4.14)

where \( C_1 \) and \( D_1 \) are arbitrary constants that are determined from the boundary conditions.

The constants \( C_1 \) and \( D_1 \) can be written in terms of the initial value at \( x = x_1 \). In particular (4.4.14) can be written in the form:

\[
\Phi_1(x) = \begin{pmatrix}
1 & 1/j \hat{h}_1 \\
-j \hat{h}_1 & 1
\end{pmatrix}
\begin{pmatrix}
e^{-j \hat{h}_1(x-x_1)} & 0 \\
e^{j \hat{h}_1(x-x_1)} & 0
\end{pmatrix}
\begin{pmatrix}
C_1 \\
D_1
\end{pmatrix}.
\] (4.4.15)

The first matrix \([e_{1,a} e_{1,b}]\) is composed from the eigenvectors \( e_{1,a} \) and \( e_{1,b} \). The constants \( C_1 \) and \( D_1 \) are determined from the initial condition \( \Phi(x_1) \). From the above equation

\[
\Phi_1(x_1) = \begin{pmatrix}
1 & 1/j \hat{h}_1 \\
-j \hat{h}_1 & 1
\end{pmatrix}
\begin{pmatrix}
C_1 \\
D_1
\end{pmatrix}.
\]
Upon inverting the matrix containing the eigenvectors, and substituting the vector \( \text{col}(C_1, D_1) \) into (4.4.15), one gets, after some simplification,

\[
\Phi_1(x) = T_1(x - x_1)\Phi_1(x_1),
\]

where the displacement or transfer matrix \( T_1(x - x_1) \) is given by:

\[
T_1(x - x_1) = \begin{pmatrix}
1 & 1/j \hat{h}_1 \\
-j \hat{h}_1 & 1
\end{pmatrix}
\begin{pmatrix}
e^{-j \hat{h}_1(x - x_1)} & 0 \\
0 & e^{j \hat{h}_1(x - x_1)}
\end{pmatrix}
\begin{pmatrix}
1 & 1/j \hat{h}_1 \\
-j \hat{h}_1 & 1
\end{pmatrix}^{-1}
\]

\[
= \begin{pmatrix}
\cos \hat{h}_1(x - x_1) & \sin \hat{h}_1(x - x_1)/\hat{h}_1 \\
-\hat{h}_1 \sin \hat{h}_1(x - x_1) & \cos \hat{h}_1(x - x_1)
\end{pmatrix}.
\]

The displacement operator \( T_1(x - x_1) \) displaces (or maps) the initial vector \( \Phi_1(x_1) \) forward to various points within the layer. Indeed, the boundary vector \( \Phi_1(x_1) \) can be displaced across the layer using the operator \( T_1(d) \) where \( d \) is the layer thickness. The displacement operator has the following properties:

1. The determinant of the operator is unity, so that it is unimodular.
2. The inverse \( T_1^{-1}(x) = T_1(-x) \). Thus, \( T_1(-x) \) can be used to displace \( \Phi_1(0) \) to points to the left. In particular, \( \Phi_1(x_1) = T_1(-d)\Phi_1(0) \).
3. The eigenvalues of the \( T_1(x) \) operator are:
   \[
e^{-j \hat{h}_1(x - x_1)}, \quad e^{j \hat{h}_1(x - x_1)}.\]
4. The eigenvectors \( e_a \) and \( e_b \) are identical to the eigenvectors of the \( A_1 \) operator in (4.4.13),

\[
e_{1,a} = \begin{pmatrix} 1 \\ -j \hat{h}_1 \end{pmatrix}, \quad e_{1,b} = \begin{pmatrix} 1/j \hat{h}_1 \\ 1 \end{pmatrix}.
\]

**SUPERSTRATRE LAYER**

The solution of the three-layer waveguide can be obtained using the vector methods described above. In layer 0 the matrix \( A_0 \) (corresponding to the \( A \) matrix with \( x \) lying in layer 0) is

\[
A_0 = \begin{pmatrix} 0 & 1 \\ -\hat{h}_0^2 & 0 \end{pmatrix},
\]

where \( \hat{h}_0^2 = k_0^2 \kappa_0 + p^2 \). The eigenvalues are

\[
\lambda_{0,a} = -j \hat{h}_0, \quad \lambda_{0,b} = j \hat{h}_0
\]

and the corresponding eigenvectors are

\[
e_{0,a} = \begin{pmatrix} 1 \\ -j \hat{h}_0 \end{pmatrix}, \quad e_{0,b} = \begin{pmatrix} 1/j \hat{h}_0 \\ 1 \end{pmatrix}.
\]

The field solution in layer 0 is

\[
\Phi_0(x) = C_0 e_{0,a} e^{-j \hat{h}_0 x} + D_0 e_{0,b} e^{j \hat{h}_0 x}
\]
Since the fields must vanish as $x \to \infty$ (bounded mode), one of the coefficients, either $C_0$ or $D_0$, must be zero. One solution produces exponential decay whereas, the other solution produces exponential growth. After choosing a solution, the imaginary part of $\hat{h}_0$ dictates boundedness of the field. The key point is that both solutions may not exist simultaneously. Therefore, $D_0 = 0$ is arbitrarily chosen so that

$$
\Phi_0(x) = C_0 e^{0x}.
$$

As discussed above, the solution in any semi-infinite layer has only one bounded solution. Either solution may be chosen as the bounded solution, however, the calculation of the transverse propagation constant $h_0 = h'_0 + j h''_0$, which is obtained from a complex square root, is an issue that needs discussion. The root must be chosen such that the boundary condition at $x = 0$ is satisfied. The choice requires some knowledge of the type of material of layer 0 relative to that of the remaining layers of the waveguide structure. In addition, the fields must vanish at large lateral distances from the waveguide. First focus on the real part of $\hat{h}_0$. We consider three cases, and require self consistency in the solutions.

1. Layers 0, 1 and 2 are lossless. For this case $h'_0 = 0$. The solution requires $h_0 = -j |h''_0|$, i.e., the imaginary part of $h_0$ is negative. This condition assures that the field will decay.

2. Layer 0 is lossy, and layers 1 and 2 are lossless. For this case $h'_0 > 0$. The solution is $h_0 = |h'_0| - j |h''_0|$. The real part of $h_0$ must be positive because power must be transferred to layer 0, i.e., the lateral component of the Poynting vector has to indicate power flow in the positive $x$ direction at $x = 0$.

3. Layer 0 is lossless, and layers 1 and 2 are lossy. In this case $h'_0 < 0$. The solution is $h_0 = -|h'_0| - j |h''_0|$. The real part of $h_0$ must be negative because power must be transferred from layer 0, i.e., the lateral component of the Poynting vector has to indicate power flow in the negative $x$ direction at $x = 0$.

SUBSTRATE LAYER

Attention is now turned to the substrate layer. Similarly, the solution in layer 2 is

$$
\Phi_2(x) = D_2 e^{2x} e^{j h'_2 (x-x_1)},
$$

where the eigenvectors are

$$
e_{2,a} = \begin{pmatrix} 1 \\ -j h'_2 \end{pmatrix}, \quad e_{2,b} = \begin{pmatrix} 1/j h'_2 \\ 1 \end{pmatrix}.
$$

The form of the solution in layer 2 represents a traveling wave propagating in the negative $x$ direction. For a bounded mode, the imaginary part of $\hat{h}_2$ must be negative ($h''_2 < 0$).

CHARACTERISTIC EQUATION

Now that the form of the field solutions are known in each of the three layers, we proceed with matching the fields across the boundaries at $x = 0$ and
$x = x_1 = -d_1$. At $x = x_1$, $\Phi_2(x_1) = \Phi_1(x_1)$, while at $x = 0$, $\Phi_1(0) = \Phi_0(0)$. The resulting equation that transforms the fields across the central guiding layer is

$$\Phi_0(0) = T_1(d_1) \Phi_2(0),$$

where the transfer matrix $T_1(d_1)$ is given as

$$T_1(d_1) = \begin{pmatrix} \cos \theta_1 & \sin \theta_1/h_1 \\ -h_1 \sin \theta_1 & \cos \theta_1 \end{pmatrix} \equiv \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \quad (4.4.18)$$

where $\theta_1 = h_1 d_1$ is defined as the optical thickness of layer 1. The above condition assures that the necessary field and flux will be continuous across the two boundaries at $x = x_1$ and at $x = x_0 = 0$. Now, using the eigenvectors $e_{2,b}$ and $e_{0,a}$, the above equation can be written as

$$C_0 e_{0,a} = D_2 T_1(d_1) e_{2,b},$$

which can be simplified to

$$(T_1 e_{2,b} e_{0,a}) \begin{pmatrix} D_2 \\ C_0 \end{pmatrix} = 0.$$

For the above system of equations to have a non-trivial solution, the matrix operating on the coefficient vector must be singular, or, the determinant

$$\det (T_1 e_{2,b} e_{0,a}) = 0. \quad (4.4.19)$$

Substituting the elements of the displacement operator, the above equation becomes

$$\begin{vmatrix} A_1/j h_2 + B_1 & 1 \\ C_1/j h_2 + D_1 & -j h_0 \end{vmatrix} = 0.$$  

The resulting dispersion equation becomes\(^\dagger\)

$$j h_0 (A_1 + j h B_1) + C_1 + j h D_1 = 0. \quad (4.4.20)$$

where the substrate wavenumber

$$h \equiv h_2.$$ 

\(^\dagger\) This is the most general equation governing the discrete modes. The matrix elements of $T_1$ represent the transfer matrix of a single layer and its elements reflect the dispersion characteristics of a single layer. For a more complicated core layer composed of several layers, the dispersion characteristics of the transfer matrix will be considerably more complicated.
(For waveguide structures with more than three layers, the substrate wavenumber will be denoted as \(h\).) Noting that both \(h_1\) and \(h_0\) can be written as functions of \(h\), (4.4.20) is only a function of \(h\) so that it will be treated as the independent variable. Indeed, all modes and their characteristics will be identified by the independent variable \(h\), the transverse wave number of the substrate layer.

After substitution of the elements of the transfer matrix, the above equation reduces to

\[
(h_0 + h)h_1 \cos h_1d_1 + j(h_1^2 + h_0h) \sin h_1d_1 = 0
\]

(4.4.21)

The above secular equation governs the modes of the slab waveguide structure. The above characteristic equation can be modified to a more familiar form. For the bounded modes of a lossless structure, the characteristic values of the transverse wave numbers \(jh_0 = p\), \(h_1 = q\), and \(jh = r\) are expressed in terms of real positive numbers \(p\), \(q\), and \(r\) whose values are computed by some numerical method. The modified dispersion equation is [1]

\[
\tan qd_1 = \frac{q(p + r)}{q^2 - pr}
\]

(4.4.22)

The modes whose transverse wave numbers satisfy (4.4.20) (or the equivalent (4.4.22)) are bounded provided \(j\) and \(jh_0\) are positive real numbers (assuming the layers are lossless). Also, the propagation constant \(\gamma = j\beta\), i.e., the mode propagates without attenuation. The finite number of bounded modes are defined as “proper.” If the layers are lossy, the propagation constant will have a real part, \(\alpha > 0\), and \(\Re\{jh_0\},\Re\{jh\} > 0\). There exists another set (infinite in number) of modes that are solutions of (4.4.20), that have \(\alpha > 0\) and \(\Re\{jh_0\},\Re\{jh\} < 0\). These modes are defined as “improper,” or “leaky.” In Fig. 4.4, improper modes are represented by points in right-half of the \(\gamma\)-Plane, and in the upper-half of the \(h\)-Plane. The field distribution of leaky modes diverges away from the core of the waveguide, however, the field is finite in the core region. Practical information concerning the performance of many contemporary waveguide structures can be deduced using the leaky mode solutions. For example, \(\alpha\) represents the “loss of power” per unit length along the \(z\) direction as the mode propagates. This power loss can also be computed from the lateral Poynting vector using the fields at the core boundaries. However, characterizing leaky-mode fields outside the core is impractical.

For proper Case I modes, the electric field \(E_y\) is determined in each of the three layers and is given by

\[
\psi_i(h,x) = D_2 \begin{cases} 
(A_1/jh + B_1)e^{-jh_0(x-x_0)}, & x_0 < x, \\
\cos h_1(x-x_1)/jh + \sin h_1(x-x_1)/h_1, & x_1 < x < 0, \\
(1/jh) e^{jh(x-x_1)}, & x < x_1.
\end{cases}
\]

(4.4.23)

where the constants \(h_0\), \(h_1\), and \(h\) satisfy the secular equation given by (4.4.22) and

\[
(jh_0)^2 = k_0^2\kappa_0 - \beta^2, \\
(h_1)^2 = k_0^2\kappa_1 - \beta^2, \\
(jh)^2 = k_0^2\kappa_2 - \beta^2
\]

(4.4.24)
The field solution is generated by starting with the substrate eigenvector and working up to the superstrate. Note that for the bounded modes, the constants \( p \) and \( r \) are real and positive so that the field decays exponentially, i.e., the mode is bounded. The coefficient \( D_2 \) can be calculated from a normalization process. A most useful normalization is that obtained from the power calculation of the \( z \) component of the Poynting vector. For example, the time-average power flow along the \( z \) direction is

\[
P_z = \frac{1}{2} \text{Re} \int_{-\infty}^{\infty} (\mathbf{E} \times \mathbf{H}^*) \cdot \hat{z} \, dx.
\]

Substituting the \( E_y \) and \( H_x \) components of the fields, the axial flow of the power becomes

\[
P_z = \frac{1}{2} \frac{\beta}{Z_0} \int_{-\infty}^{\infty} |\psi_1(h, x)|^2 \, dx.
\]

For the normalization equation, we calculate \( D_2 \) in (4.4.23) such that

\[
\int_{-\infty}^{\infty} |\psi_1(x)|^2 \, dx = 1.
\]

This gives

\[
D_2 = \sqrt{\frac{2\rho \mu}{\mu + \rho + d_1 \frac{\mu}{\rho} \frac{q^2 + q^2}{q^2 + \rho^2}}}
\]

As a final note on the discrete set of modes, the improper modes cannot be normalized and are not a part of the complete set of modes that are used to characterize the fields of dielectric waveguides. Only the proper modes of Case I, and the proper modes of Cases II, III and IV are used to describe the field structure.

### 4.4.2 Case II: Radiation Modes

The radiation modes of Case II have the field extensions into the lower half \( x-z \) plane or into the negative \( x \) layer, whereas the fields exponentially decay in the superstrate. The mode amplitudes remain finite at all points in the \( x-z \) plane. While the power in the trapped modes is concentrated in the vicinity of the central waveguide layer, the power in the radiation modes of this section is concentrated in the substrate. In fact, the power is almost uniformly distributed along the negative \( x \) values. It is easy to conclude that “all of the power” is propagating the semi-infinite layer 2. Again, it should be emphasized that the trapped modes have an effective propagation constant that lies between \( n_1 \) and \( n_2 \), whereas, the radiation modes in the lower half-plane have normalized propagation constants that lie between \( n_2 \) and \( n_0 \). The trapped modes had transverse propagation constants \( (h_i) \) that were real in the inside layer 1, but imaginary in the outside layers 0 and 2. The lower half-plane radiation modes have real transverse propagation constants in layers 1 and 2, but it is imaginary in layer 0.

The form of the solution in layers 0 and 1 remain identical to that obtained earlier. However, the solution in layer 2 becomes

\[
\Phi_2(x) = C_2 e_2, a e^{-j(h(x-x_1) + D_2 e_2, b e^{j(h(x-x_1)}), \quad (4.4.25)
\]
where the arbitrary constants $C_2$ and $D_2$ govern the field propagation in the transverse direction. In (4.4.25), the transverse wavenumber in layer 2 is denoted by the variable $h$. (It will be treated as the independent variable.) Recall that this infinite or continuous (on $\beta_0$) set of modes propagates without attenuation or gain, assuming all layers are reactive and without ohmic losses. Hence, the lateral Poynting vector at the boundaries $x = 0$ and $x = x_1$ must be null. This implies that $D_2/jh$ is the complex conjugate of $C_2$, i.e., $D_2/jh = C_2^*$. The lateral wavenumber $h$ of layer 2 lies in the interval $(0, k_o \sqrt{\kappa_2^{\prime} - \kappa_0})$.

The dependent variables related to the independent variable $h$ are

$$\gamma = j \sqrt{k_0^2 \kappa_2 - h^2},$$

$$h_0 = -j \sqrt{k_0^2 (\kappa_2 - \kappa_0) - h^2}.$$  \hspace{1cm} (4.4.26)

$$h_1 = \sqrt{k_0^2 (\kappa_1 - \kappa_2) + h^2}.$$

![Figure 4.4. Mapping of points from the $\gamma$-plane to the $h$-plane. In the $h$-plane bounded modes are represented by points lying on the imaginary axis while radiation and evanescent modes are represented by points (lines) lying on the real axis.](image)

The transformation from the $\gamma$-plane to the $h$-plane is illustrated in Fig. 4.4. Note that both radiation modes and the evanescent modes are represented in the $h$-plane as $h$ values lying on the positive real axis.

The solution in the substrate as given by (4.4.25) can be written in the form:

$$\Phi_2(x) = \begin{pmatrix}
\cos h(x - x_1) & \sin h(x - x_1)/h \\
-h \sin h(x - x_1) & \cos h(x - x_1)
\end{pmatrix}
\begin{pmatrix}
Q_2(h) \\
P_2(h)
\end{pmatrix}$$

where $Q_2(h)$ and $P_2(h)$ can be expressed in terms of $C_2$ and $D_2$. The elements $Q_2(h)$ and $P_2(h)$ are real quantities owing to the fact that $D_2 = C_2^*$.

For the bounded modes, the fields were matched starting in layer 2, while in this case, the process of matching the fields across the various layers will be started at layer 0. In layer 0, the solution is given by (4.4.17). The coefficients at $x = x_0$ in terms of the coefficients at $x = x_1$ is

$$C_0(h) \begin{pmatrix} 1 \\ -j h_0 \end{pmatrix} = \begin{pmatrix} \mathcal{A}_1(h) & \mathcal{B}_1(h) \\ \mathcal{C}_1(h) & \mathcal{D}_1(h) \end{pmatrix} \begin{pmatrix} Q_2(h) \\ P_2(h) \end{pmatrix}.$$
(The matrix elements $A_1, B_1, C_1, D_1$ are also given by (4.4.18). However, here the transverse wavenumber $h_1 = h_1(h)$ is a continuous function of $h$ for the radiation modes while $h_1$ takes on discrete values for the trapped modes.) Thus, the coefficients of layer 2 are obtained using the inverse of $T_1(h)$,

$$
\begin{pmatrix}
Q_2(h) \\
P_2(h)
\end{pmatrix} = C_0(h) \begin{pmatrix}
D_1(h) & -B_1(h) \\
-C_1(h) & A_1(h)
\end{pmatrix} \begin{pmatrix} 1 \\ -j h_0 \end{pmatrix}
$$

Multiplying the matrices, one gets the following values of the coefficients of layer 2 as:

$$
\begin{align*}
Q_2(h) &= C_0(h)[D_1(h) + j h_0 B_1(h)], \\
P_2(h) &= -C_0(h)[C_1(h) + j h_0 A_1(h)].
\end{align*}
$$

The field distribution in the various layers becomes:

$$
\psi_2(h, x) = C_0(h) \begin{cases}
\begin{aligned}
& e^{-j h_0 x}, & 0 = x_0 < x, \\
& [D_1(h) + j h_0 B_1(h)] \cos h_1(x - x_1) - \left[ C_1(h) + j h_0 A_1(h) \right] \sin h_1(x - x_1)/h_1, & x_1 < x < x_0, \\
& [D_1(h) + j h_0 B_1(h)] \cos h(x - x_1) - \left[ C_1(h) + j h_0 A_1(h) \right] \sin h(x - x_1)/h, & x < x_1.
\end{aligned}
\end{cases}
$$

The $h$ variable has been added as an argument to the wave function. Equation (4.4.27) may be written in exponential form as

$$
\psi_2(h, x) = \begin{cases}
\begin{aligned}
& C_0(h) e^{-j h_0 x}, & 0 = x_0 < x, \\
& C_1(h) e^{-j h_1(x-x_1)} + D_1(h) e^{j h_1(x-x_1)/j h_1}, & x_1 < x < x_0, \\
& C_2(h) e^{-j h(x-x_1)} + D_2(h) e^{j h(x-x_1)/j h}, & x < x_1.
\end{aligned}
\end{cases}
$$

where

$$
\begin{align*}
C_1(h) &= [Q_2(h) + j P_2(h)/h_1]/2, & D_1(h) &= j h_1[Q_2(h) - j P_2(h)/h_1]/2, \\
C_2(h) &= [Q_2(h) + j P_2(h)/h]/2, & D_2(h) &= j h[Q_2(h) - j P_2(h)/h]/2.
\end{align*}
$$

The coefficient $C_0(h)$ can be found using the normalization procedure for the continuous modes as

$$
\delta(h - \tilde{h}) = \int_{-\infty}^{\infty} \psi_2(h, x) \psi_2(\tilde{h}, x) \, dx.
$$

For the Case II radiation modes, the field extension in the layer $x < x_1$, is infinite. This implies that the fraction of mode power in layer 2 is unity. (The overlap integrals, developed in 'apo1', can be used to calculate the normalization coefficient simply by replacing $\phi(h, x)$ with $\psi(\tilde{h}, x)$.) After substitution of the fields in the substrate,

$$
\int_{-\infty}^{\infty} \psi_2(h, x) \psi_2(\tilde{h}, x) \, dx = \frac{\pi}{2} [Q_2^2(h) + P_2^2(h)/h^2] \delta(h - \tilde{h}),
$$
so that according to (4.4.29), the coefficient of the delta function is unity. After substituting the coefficients $Q_2$ and $P_2$, one gets the following value for the normalization coefficient:

$$C_0(h) = \frac{2}{\pi} \frac{h}{\sqrt{h^2(D_1 + j h_0 B_1)^2 + (C_1 + j h_0 A_1)^2}}.$$  \hspace{1cm} (4.4.30)

The term under the radical sign in the denominator can be factored so that the singular points of $C_0(h)$ occur when

$$j h_0(A_1 - j h B_1) + (C_1 - j h D_1) = 0,$$ \hspace{1cm} (4.4.31)
$$j h_0(A_1 + j h B_1) + (C_1 + j h D_1) = 0.$$ \hspace{1cm} (4.4.32)

After substitution of the values of $A_1, B_1, C_1, D_1$ in terms of the layer 1 parameters, the factored terms show that there are singular points of $C_0(h)$ when

$$j h_0(h_1 \cos h_1 d_1 - j h \sin h_1 d_1) + h_1(h_1 \sin h_1 d_1 - j h \cos h_1 d_1) = 0;$$ \hspace{1cm} (4.4.33)
$$j h_0(h_1 \cos h_1 d_1 + j h \sin h_1 d_1) + h_1(h_1 \sin h_1 d_1 + j h \cos h_1 d_1) = 0.$$ \hspace{1cm} (4.4.34)

It is interesting to note that the normalization coefficient given by (4.4.30) has singular points at $h$ (or $\beta$) values that are solutions of (4.4.32) and (4.4.34) which is also recognized as representing the eigenvalues of both the proper and improper modes of the three layer waveguide as given by (4.4.21). The second equation, (4.4.31), indicates singular points at $-h$, so that the singular points come in pairs at $\pm h$. For bounded modes in a lossless medium, the singular points occur on the imaginary axis of $h$, and in conjugate pairs. For improper modes, the singular points lie above the real axis in the right-half of the $h$–Plane, while the associated singular points lie below the real axis in the left-half of the $h$–Plane. Figure 4.7(a) shows an example of the partial set of both the proper and improper modes.

Figure 4.5 shows the mode amplitude, $C_0(h)$ of (4.4.30), as a function of $h/k_o$, the normalized lateral wavenumber of the substrate, for several core widths. The refractive indices of the waveguide are identical for the different core widths.

\[†\]

Generally, $h$ is a complex variable but it is restricted to “real” values in the figure.
As the width of the waveguide increases, these normalized characteristics modes have oscillating amplitudes. The sequence of curves which represents the normalized field amplitude of the Case II modes at the boundary between the core region and the superstrate, are for different core widths, but identical dielectric constants. As a result, the range of \( h/k_o \) for the Case II modes is from 0 to approximately 0.8. If \( \lambda_o = 1 \mu m \), the core widths are 1, 2, 4 and 8 \( \mu m \). The 1 \( \mu m \) core supports only a single trapped waveguide mode, while the structure with an 8 \( \mu m \) core supports 13 trapped modes.

The modes of Case II have field distributions that are dependent upon the substrate wavenumber. Figure 4.6 shows the amplitude of the superstrate coefficient \( C_0(h) \) as a function of \( h/k_o \) for the waveguide width \( d_1 = 4\lambda_o \). The associated electric fields at the two specific points is illustrated.
4.4 TRANSVERSE ELECTRIC MODES: FORMULATION

Fig. 4.5. The electric field distribution is illustrated at two different points, one at a local maximum while the other is at a local minimum for $C_0$.

It is instructive to understand the primary cause of the oscillations of $C_0(h)$. Consider the $8 \, \mu m$ waveguide structure. Notice that the amplitude of $C_0(h)$ oscillation in Fig. 4.5, damps as $h$ increases. The location and magnitude of the peaks are related to the location of the leaky modes in the complex $h$-plane. Figure 4.7(a) shows the location of both the trapped and leaky modes in the the $h$-plane, while (b) illustrates the amplitude of the superstrate coefficient $C_0(h)$ as a function of $h$ on the real axis. The circles on the abcissa indicates the nearest leaky mode roots in (a), the vertical line points to the neighboring peak of $C_0$.

Although it is difficult to decipher from Fig. 4.7(a), the displacement of the leaky mode from the real axis increases as the mode number increases (say from left to right). For example, the first leaky mode is at $h/k_0 = 0.284 + j 0.0201$, while the last is at $h/k_0 = 0.754 + j 0.0227$. Thus, the effect of the singularity on $C_0$ decreases the oscillation amplitude.

4.4.3 Case III: Radiation Modes

Before analyzing the Case III radiation modes, consider the steps taken in describing the discrete bounded modes of Case I, and the Case II radiation modes (infinite field extension only in the substrate, layer 2). The Case III modes will have extension in both the substrate and superstrate layers. The fields in both the semi-infinite layers involve two independent solutions of the second-order differential equation, or two first-order equations. The solutions in the outside layers are coupled via the displacement of transfer matrix. This coupling process yields two equations. (In our method of modal analysis, the bounded modes have only two field amplitude constants, namely, $C_0$ and $D_2$.) In addition, the complex propagation constant $\gamma$ can be considered as undetermined. The non-trivial solution for the set of linear equations in the coefficients $C_0$ and $D_2$ produced the secular equation yielding the $\gamma$ eigenvalues. The coefficients $C_0$ and $D_2$ are connected by the linear set (2 equations) which has been reduced to rank 1. The normalization process determines the value of the remaining coefficient.
For the Case II radiation modes, there is an extra coefficient due to the fact that the field solution in the substrate, layer 2, has two amplitude coefficients that must be determined by matching the solutions at the various boundaries. For the Case III radiation modes, the fields in layer 0 have an added coefficient with a corresponding solution of the form

\[ \Phi_0(x) = C_0 \, e^{-j k_0 x} + D_0 \, e^{+j k_0 x}. \]  

Again, since the field propagates without loss or gain, the transverse Poynting vector at the \( x = 0 \) boundary must be null. This implies that \( C_0 \) and \( D_0 \) do not differ in magnitude. The independent variable \( \gamma \), \( h_0 \), and \( h_1 \) become

\[ \gamma = j \sqrt{k_0^2 \kappa_2 - h^2}, \]

\[ h_0 = \sqrt{h^2 - k_0^2 (\kappa_0 - \kappa_2)}, \]

\[ h_1 = \sqrt{h^2 + k_0^2 (\kappa_1 - \kappa_2)}. \]

The solution in layer 0 takes the familiar form

\[ \Phi_0(x) = \begin{pmatrix} \cos h_0 (x - x_0) & \sin h_0 (x - x_0) / h_0 \\ -h_0 \sin h_0 (x - x_0) & \cos h_0 (x - x_0) \end{pmatrix} \begin{pmatrix} Q_0(h) \\ P_0(h) \end{pmatrix}. \]

The field distribution in the various layers becomes

\[ \psi_{ii}(\hat{h}, x) = \begin{cases} Q_0(h) \cos h_0 (x - x_0) + P_0 \sin h_0 (x - x_0) / \hat{h}_0, & x_0 < x, \\ Q_2(h) \cos h_1 (x - x_1) + P_2(h) \sin h_1 (x - x_1) / \hat{h}_1, & x_1 < x < x_0, \\ Q_2(h) \cos h (x - x_1) + P_2(h) \sin h (x - x_1) / \hat{h}, & x < x_1, \end{cases} \]

and the normalization of the Case III modes satisfies

\[ \int_{-\infty}^{\infty} \psi_{ii}(\hat{h}, x) \psi_{ii}(\tilde{\hat{h}}, x) \, dx = \frac{\pi}{2} \left( Q_2^2(h) + \frac{P_2^2(h)}{h^2} + [Q_0^2(h) + \frac{P_0^2(h)}{h_0^2}] \delta(\hat{h} - \tilde{\hat{h}}) \right). \]

which implies that the coefficient of the delta function is unity. Coefficients in the central layer, \( Q_1(h) \) and \( P_1(h) \), are identical to the coefficients of the bottom layer, \( Q_2(h) \) and \( P_2(h) \), because the components of the \( \Phi \) are continuous across the boundaries. (Both layers have their field expressions referenced to \( x_1 \).) The vector containing the components of \( Q_0 \) and \( P_0 \) are obtained by the transformation

\[ \begin{pmatrix} Q_0(h) \\ P_0(h) \end{pmatrix} = \begin{pmatrix} \cos (h_1 d_1) & \sin (h_1 d_1) / h_1 \\ -h_1 \sin (h_1 d_1) & \cos (h_1 d_1) \end{pmatrix} \begin{pmatrix} Q_2(h) \\ P_2(h) \end{pmatrix}. \]
Thus, there remains two coefficients, either \( \{Q_0, P_0\}, \{Q_2, P_2\} \), or a combination of them. The normalization procedure will take care of one coefficient, so that the other must be determined by some other means. It is important to realize that there are “two sets” of modes with the same eigenvalue \( \lambda \). Accordingly, these degenerate modes can be obtained by arbitrarily specifying \( P_0 \) in terms of \( Q_0 \) or vice versa. For example, one set could have \( Q_0 = 0 \), while the second set could have \( P_0 = 0 \). The normalization procedure could then be developed similar to that for the Case II radiation modes. Although there is an infinite number of ways to break degeneracy, such as that discussed above, it is useful to make the two sets orthogonal. For example, consider the case when the waveguide is “symmetric” or when the two refractive indices, \( n_0 \) and \( n_2 \) are identical. (A simple three-layer waveguide is symmetric when it has reflection symmetry about the mid-point of the central slab layer.) With the dielectric profile symmetric about the center of the waveguide, the functional dependence of the waveguide modes may be either “even” or “odd” functions about the midpoint of the central slab.

The electromagnetic field functions as discussed in this chapter are written in terms of the reference points, the layer interfaces. According to Fig. 4.1, the layer interfaces are \( x_0 = 0 \) and \( x_1 = -d_1 \). On the other hand, the central-layer fields given by (4.4.38), may be referenced to the center of Layer 1 according to

\[
\psi_{int}(h, x) = \frac{\hat{h}_1 Q_2 \cos(h_1 d_1/2)}{\hat{h}_1} \cos(h_1 (x + d_1/2))
- \frac{\hat{h}_1 Q_2 \sin(h_1 d_1/2)}{\hat{h}_1} \sin(h_1 (x + d_1/2)).
\]

An “even” mode will be designated as one satisfying

\[
P_2^e = \hat{h}_1 \tan(h_1 d_1/2) \frac{Q_2^e(h)}{Q_2},
\]

which implies the coefficient of the second term in the above equation is zero. This condition only forces the “spatial” phase of the field to zero at the center of the core region so in fact, the field is symmetric about the center of the core in the core region. The “even” modes of Case III can now be written as

\[
\psi_{int}^e(h, x) = \begin{cases} 
Q_0^e(h) \cos(h_0 (x - x_0)/h_0) + P_0^e(h) \sin(h_0 (x - x_0)/h_0), & x_0 < x, \\
Q_2^e(h) \cos(h_1 (x - x_1)/\hat{h}_1) + P_2^e(h) \sin(h_1 (x - x_1)/\hat{h}_1), & x_1 < x < x_0, \\
Q_2^e(h) \cos(h (x - x_1)/h) + P_2^e(h) \sin(h (x - x_1)/h), & x < x_1.
\end{cases}
\]

On the other hand, an “odd” mode satisfies

\[
P_2^o = -\hat{h}_1 \cot(h_1 d_1/2) \frac{Q_2^o(h)}{Q_2},
\]

and the field expressions are similar to that of (4.4.42), where “e” is replaced with “o.” The product of an even and an odd mode produces an odd function whose integral is zero, so that the “even” modes are orthogonal to the “odd” modes.
Instead of using the process of characterizing the two sets of modes, as discussed above for the three-layer waveguide, consider the more general approach of using the “optical center” of the “core” region. If the fields in the outside cladding layers are referenced to the optical center of the core region (see Problem 4–2), \( \{Q_{\text{oc}}, P_{\text{oc}}\} \), the normalization process can be simplified. The coefficients of the fields in the cladding layers can be written as (for the “even” modes)

\[
\begin{pmatrix}
  Q_{\text{e}}^2 \\
  P_{\text{e}}^2
\end{pmatrix} =
\begin{pmatrix}
  D_{\text{oc}} & -B_{\text{oc}} \\
  -C_{\text{oc}} & A_{\text{oc}}
\end{pmatrix}
\begin{pmatrix}
  Q_{\text{oc}}^2 \\
  0
\end{pmatrix} =
\begin{pmatrix}
  D_{\text{oc}} \\
  -C_{\text{oc}}
\end{pmatrix} Q_{\text{oc}}^e, \tag{4.43}
\]

\[
\begin{pmatrix}
  Q_{\text{e}}^0 \\
  P_{\text{e}}^0
\end{pmatrix} =
\begin{pmatrix}
  A_{\text{oc}} & B_{\text{oc}} \\
  C_{\text{oc}} & D_{\text{oc}}
\end{pmatrix}
\begin{pmatrix}
  Q_{\text{oc}}^e \\
  0
\end{pmatrix} =
\begin{pmatrix}
  A_{\text{oc}} \\
  C_{\text{oc}}
\end{pmatrix} Q_{\text{oc}}^e, \tag{4.44}
\]

whereas the “odd” modes have

\[
\begin{pmatrix}
  Q_{\text{o}}^2 \\
  P_{\text{o}}^2
\end{pmatrix} =
\begin{pmatrix}
  -B_{\text{oc}} \\
  A_{\text{oc}}
\end{pmatrix} P_{\text{o}}^e, \tag{4.45}
\]

\[
\begin{pmatrix}
  Q_{\text{o}}^0 \\
  P_{\text{o}}^0
\end{pmatrix} =
\begin{pmatrix}
  B_{\text{oc}} \\
  D_{\text{oc}}
\end{pmatrix} P_{\text{o}}^e. \tag{4.46}
\]

The inner product of an “even” and “odd” mode becomes

\[
\int_{-\infty}^{\infty} \psi_{\text{e}}^e(h, x) \psi_{\text{o}}^o(h', x) \, dx = \frac{\pi}{2} \left\{ Q_{\text{e}}^2(h) Q_{\text{o}}^2(h') + P_{\text{e}}^2(h) P_{\text{o}}^2(h') / hh' \right\} + [Q_{\text{e}}^0(h) Q_{\text{o}}^0(h') + P_{\text{e}}^0(h) P_{\text{o}}^0(h') / hh'] \delta(h - h') \neq 0. \tag{4.47}
\]

so that the two sets are not orthogonal. To make the fields “look” symmetrical about the optical center of the core, the field vectors for the two sets can be written as

\[
\Phi_{\text{oc}}^e = m(h) \begin{pmatrix} 1 \\ \varrho \end{pmatrix} \quad \text{for even modes;} \quad \Phi_{\text{oc}}^o = m(h) \begin{pmatrix} \varrho \\ 1 \end{pmatrix} \quad \text{for odd modes;} \tag{4.48}
\]

where \( \varrho(h) \) can be found from orthogonally, while \( m(h) \) can be calculated from normalization.

Another procedure which may be used to separate the degenerate modes can be obtained by making \( d \ln \psi / dx \) at one cladding equal to its negative value at the opposite cladding. Note that this condition can be applied to either the even or odd modes in the case of symmetric structures. In this procedure, the following terms help to simplify the field expressions in terms of the “spatial phase,” \( \varphi \), and field magnitude \( M(h) \),

\[
\tan \varphi_2 = \frac{P_2(h)}{h Q_2(h)}, \quad M_2^2(h) = Q_2^2(h) + P_2^2(h) / h^2, \tag{4.49}
\]

and

\[
\tan \varphi_0 = \frac{P_0(h)}{h_0 Q_0(h)}, \quad M_0^2(h) = Q_0^2(h) + P_0^2(h) / h_0^2. \tag{4.50}
\]
As discussed in Problem 4.4, the angle $\varphi_2 = -\varphi_0$ satisfies

$$
\tan \varphi_2 = -\frac{(h_0 A_1 + h D_1) \pm \sqrt{(h_0 A_1 - h D_1)^2 + 4 h h_0}}{2 h h_0 B_1},
$$

(4.4.51)

where $i = e$ uses the $+$ sign while $i = o$ uses the $-$ sign. The field amplitude coefficient $M_2$ satisfies

$$
M_2^i(h) = \sqrt{\frac{2}{\pi}} \frac{\sqrt{h} \cos \varphi^i_2}{\sqrt{h \cos^2 \varphi^i_2 + h_0(A_1 \cos \varphi^i_2 + h B_1 \sin \varphi^i_2)^2}},
$$

(4.4.52)

and $M_0$ is tied to $M_2$ by the normalization relation

$$
h_0 [M_0^i(h)]^2 + h [M_2^i(h)]^2 = \frac{2 h}{\pi}.
$$

(4.4.53)

Finally, the field distribution for the Case III radiation modes is now given by

$$
\psi^i_{hi}(h, x) = M_2^i(h) \begin{cases} 
(A_1 \cos \varphi^i + B_1 h \sin \varphi^i) \cos h_0(x - x_0) + \\
(C_1 \cos \varphi^i + D_1 h \sin \varphi^i) \sin h_0(x - x_0), & 0 \leq x_0 < x, \\
\cos \varphi^i \cos h_1(x - x_1) + \\
\sin \varphi^i \sin h_1(x - x_1), & x_1 \leq x < x_0, \\
\cos \varphi^i \cos h(x - x_2) + \sin \varphi^i \sin h(x - x_0), & x < x_1.
\end{cases}
$$

(4.4.54)

where $i$ designates the subset mode group, $e$ or $o$.

Figure 4.8. The amplitude coefficient $M_2(h)$ for the Case III “$e$" and “$o$” modes as a function of the normalized substrate wavenumber, $h/k_0$, for the core width of $d_1 = 4\lambda_0$. The associated electric fields for the two modes at the two specific points is illustrated.

Figure 4.8 shows the mode amplitude coefficient for the $e$ and $o$ modes as a function of the designated spectrum parameter $h$, the transverse wavenumber of the wave in the substrate. (The points illustrated on the abscissa show the
location of the nearest leaky mode.) These Case III modes show that $M_2(h)$ has a damped oscillation that is dependent on the proximity of the leaky modes to the real axis of $h$, $\Re h$. The amplitudes $M_2^e$ and $M_2^o$ have local peak values that depend on every other leaky mode number.

### 4.4.4 Case IV: Evanescent Radiation Modes

Finally, going to the Case IV evanescent modes, where $\gamma = \alpha$, the substrate wavenumber $h$, satisfies $k_0\kappa_2 < h < \infty$, and the amplitude coefficients $M_2^e$ and $M_2^o$ satisfy (4.4.52) and (4.4.53). While the coefficients of the Case II radiation modes cannot be analytically continued to the Case III radiation modes, the Case III radiation modes are analytically continued to the Case IV evanescent modes. The amplitude coefficients $M_2^e$ and $M_2^o$ of ‘w.f. 7’ continue their damped oscillation as a function of $h$, and will converge to $1/\sqrt{\pi}$ as $h \to \infty$. The leaky mode locations in the complex $h$–Plane become more displaced from the real axis.

### 4.5 Mode Completeness and Orthogonally

Beginning in this section, modes will be identified by the substrate wavenumber $h$ which uniquely identifies the mode. The wave numbers $h^k$ and $h^l$ refer to different modes and should not be confused with the layer wave numbers discussed earlier in this chapter.

Generally, the characteristic modes of enclosed metallic waveguides are orthogonal. Furthermore, the characteristic modes are discrete and infinite in number. Dielectric waveguides have a somewhat more complicated modal structure and the spectrum is divided into a discrete spectrum, the bounded modes of Case I, and a continuous spectrum, modes of infinite extent of Case II, Case III and Case IV. To address the problems of radiation from dielectric waveguides, the scattering of light at waveguide discontinuities or waveguide transitions, it is necessary to have a complete set of modes in order to characterize arbitrary fields.

When possible it is useful to make $\psi(h, x)$ orthogonal with its complex conjugate $\psi^*(h, x)$. If the material used to form the waveguide layers is lossless, the fields are real quantities so that $\psi(h, x) = \psi^*(h, x)$. This implies that the power in a mode will remains unchanged as it propagates along the guided structure. When the layers are lossy, power orthogonally cannot be met. An equivalent condition for power orthogonally is that the operator in (4.5.2) be Hermitian, i.e., $L_{op} = L_{op}^*$. Nevertheless, it is always possible to have an orthogonal condition between all modes as discussed below.

The second-order differential equation governing the fields has a Strum-Liouville form [4], [25] given by

$$
\frac{d^2\psi}{dx^2} + k_0^2\kappa(x) - \kappa_2 \psi + h^2\psi = 0, \quad (4.5.1)
$$

where $\kappa_2$ is the dielectric constant of the substrate. The term

$$
L_{op} = -\frac{d^2}{dx^2} - k_0^2\kappa(x) - \kappa_2, \quad (4.5.2)
$$
can be recognized as a linear operator and $h^2$ as its eigenvalue. By assuming two different solutions $\psi_\mu$ and $\psi_\nu$ with corresponding eigenvalues $h_\mu^2$ and $h_\nu^2$, with $h_\mu \neq h_\nu$, it can be shown that the two eigenfunctions $\psi_\mu$ and $\psi_\nu$ are orthogonal over the interval $-\infty < x < \infty$. (The two solutions may be in the same case or in different cases.) The integration over the region containing the core

\[ (h_\mu^2 - h_\nu^2) \int_{x_0+X}^{x_1-X} \psi_\mu(h_\mu, x) \psi_\nu(h_\nu, x) \, dx = \]

\[ \int_{x_0+X}^{x_1-X} \{ \psi_\mu(h_\mu, x) L_{op} \psi_\nu(h_\nu, x) - \psi_\nu(h_\nu, x) L_{op} \psi_\mu(h_\mu, x) \} \, dx. \]  

Upon integrating the right-hand-side by parts, (4.5.3) becomes

\[ (h_\mu^2 - h_\nu^2) \int_{-\infty}^{\infty} \psi_\mu(h_\mu, x) \psi_\nu(h_\nu, x) \, dx = \]

\[ \lim_{X \to \infty} \{ \psi_\mu(h_\mu, x_0 + X) \psi'_\nu(h_\nu, x_0 + X) - \psi_\mu(h_\mu, x_1 - X) \psi'_\nu(h_\nu, x_1 - X) \}

- \psi_\nu(h_\nu, x_0 + X) \psi'_\mu(h_\mu, x_0 + X) + \psi_\nu(h_\nu, x_1 - X) \psi'_\mu(h_\mu, x_1 - X) \} \]  

If either $\psi_\mu$ or $\psi_\nu$ is a bounded mode, then the bounded fields and their derivative vanish as $X \to \infty$, so that the right-hand-side is zero, implying the modes are orthogonal. If either $\psi_\mu$ or $\psi_\nu$ are Case II modes while the other is Case II, III or IV, then, the above integral becomes

\[ \lim_{X \to \infty} \{ -\psi_\mu(h_\mu, x_1 - X) \psi'_\nu(h_\nu, x_1 - X) + \psi_\nu(h_\nu, x_1 - X) \psi'_\mu(h_\mu, x_1 - X) \} \]  

because the Case II fields and their derivatives vanish in the superstrate. After substituting the fields and the appropriate derivatives, (4.5.4) becomes,

\[ (h_\mu^2 - h_\nu^2) \int_{-\infty}^{\infty} \psi_\mu(h_\mu, x) \psi_\nu(h_\nu, x) \, dx = \lim_{X \to \infty} \frac{k_0}{2} \{

- (h_\mu + h_\nu)(Q_2^\mu P_2^\nu / h_\nu - Q_2^\nu P_2^\mu / h_\mu) \cos(h_\mu - h_\nu)X

+ (h_\mu - h_\nu)(Q_2^\mu P_2^\nu / h_\nu + Q_2^\nu P_2^\mu / h_\mu) \cos(h_\mu + h_\nu)X

+ (h_\mu + h_\nu)(Q_2^\nu Q_2^\nu - P_2^\mu P_2^\nu / h_\mu P_2^\nu / h_\nu) \sin(h_\mu - h_\nu)X

+ (h_\mu - h_\nu)(Q_2^\nu Q_2^\nu - P_2^\mu P_2^\nu / h_\mu P_2^\nu / h_\nu) \sin(h_\mu + h_\nu)X \} \]  

When $h_\mu \neq h_\nu$, the “residual” terms on the right-hand-side of (4.5.6) do not vanish, but their frequency of spacial oscillation becomes infinite as $X \to \infty$. This condition effectively makes $\psi_\mu$ orthogonal to $\psi_\nu$ because when the residual terms are included in any integration over any part of the continuous spectrum the result is zero. Thus, it is concluded that the modes of Case II are orthogonal and are orthogonal to the modes of Cases III and IV.
Finally, consider when $\psi_\mu$ and $\psi_\nu$ are modes of either Case III or Case IV. A form similar to that obtained in (4.5.6) can be obtained after retaining the superstrate terms of (4.5.4)

$$\left(h_\mu^2 - h_\nu^2\right) \int_{-\infty}^{\infty} \psi_\mu(h_\mu, x) \psi_\nu(h_\nu, x) \, dx = \lim_{X \to \infty} \frac{1}{2} \left\{ \right. \left. \right. \right. \right. \right.
\begin{align*}
&+ (h_\mu h_\nu - h_\nu h_\mu)(Q_0^\mu P_0^\nu / h_\nu - Q_0^\nu P_0^\mu / h_\mu) \cos(h_\mu h_\nu)X \\
&- (h_\mu h_\nu - h_\nu h_\mu)(Q_0^\mu P_0^\nu / h_\nu + Q_0^\nu P_0^\mu / h_\mu) \cos(h_\mu h_\nu)X \\
&+ (h_\mu h_\nu + h_\nu h_\mu)(Q_0^\mu Q_0^\nu + P_0^\mu / h_\mu P_0^\nu / h_\nu) \sin(h_\mu h_\nu)X \\
&+ (h_\mu h_\nu - h_\nu h_\mu)(Q_0^\mu Q_0^\nu - P_0^\mu / h_\mu P_0^\nu / h_\nu) \sin(h_\mu h_\nu)X \\
&- (h_\mu h_\nu - h_\nu h_\mu)(Q_0^\mu P_0^\nu / h_\nu - Q_0^\nu P_0^\mu / h_\mu) \cos(h_\mu h_\nu)X \\
&+ (h_\mu h_\nu + h_\nu h_\mu)(Q_0^\mu Q_0^\nu + P_0^\mu / h_\mu P_0^\nu / h_\nu) \cos(h_\mu h_\nu)X \\
&+ (h_\mu h_\nu - h_\nu h_\mu)(Q_0^\mu Q_0^\nu - P_0^\mu / h_\mu P_0^\nu / h_\nu) \sin(h_\mu h_\nu)X \\
&+ (h_\mu h_\nu + h_\nu h_\mu)(Q_0^\mu Q_0^\nu + P_0^\mu / h_\mu P_0^\nu / h_\nu) \cos(h_\mu h_\nu)X \right\}.
\end{align*}

(4.5.7)

When $h_\mu \neq h_\nu$, all terms on the right-hand-side of (4.5.7) rapidly oscillate with respect to $h$.

Before leaving this section, the normalization as discussed earlier in this chapter will also be obtained from (4.5.6). Upon dividing both sides of (4.5.6) by $h_\mu^2 - h_\nu^2$, and letting $h_\nu \to h_\mu$, the coefficient of $\cos(h_\mu - h_\nu)X$ vanishes, the terms $\cos(h_\mu + h_\nu)X$ and $\sin(h_\mu + h_\nu)X$ have rapid oscillation. The remaining term is

$$\int_{-\infty}^{\infty} \psi_\mu(h_\mu, x) \psi_\nu(h_\mu, x) \, dx = \pi \delta(h_\mu - h_\nu)$$

so that

$$\int_{-\infty}^{\infty} \psi_\mu(h_\mu, x) \psi_\nu(h_\mu, x) \, dx = \frac{\pi}{2} \left( Q_0^\mu Q_0^\nu + P_0^\mu P_0^\nu \right) \delta(h_\mu - h_\nu).$$

(4.5.8)

This expression is identical to the earlier normalization. A similar procedure can be developed for the Case III and Case IV radiation modes.

With the complete set of modes, Case I through Case IV, an arbitrary field distribution, $E_\nu(x, z) = \Psi(x, z)$ can be represented in terms of the spectral content of the operator $L_{op}$ as

$$\Psi(x, z) = \sum_{\mu=0}^{M-1} A_i(h_\mu) \psi_i(h_\mu, x) e^{-\gamma(h_\mu)z} + \int_{k_0 \sqrt{x^2-k_0^2}} A_{i\nu}(h) \psi_i(h, x) e^{-jz(h)z} \, dh$$

$$+ \int_{k_0 \sqrt{x^2-k_0^2}} \left[ A_{iii}(h) \psi_{i\nu}(h, x) + A_{i\nu}(h) \psi_{ii}(h, x) \right] e^{-jz(h)z} \, dh$$

$$+ \int_{-\infty}^{\infty} \left[ A_{iv}(h) \psi_{iv}(h, x) + A_{iv}(h) \psi_{iv}(h, x) \right] e^{-jz(h)z} \, dh$$

(4.5.9)
The above expression for the fields includes the dielectric constant in the limits of the integrals over the continuous spectrum. For the discrete spectrum, the propagation constant is written as \( \gamma_n (= \alpha_n + j\beta_n) \) to allow for possible losses in the layers. In this instance, the eigenfunctions \( \psi_1(h_{\mu}, x) \) are complex functions. In the event the medium is totally reactive, then \( \Re\{\gamma\} = \alpha = 0 \). In addition, the eigenfunctions \( \psi_1(h^n, x) \) are real, i.e. \( \psi_1(h_{\mu}, x) = \psi_1^*(h_{\mu}, x) \).

In a totally reactive cladding structure, dielectric constants are real so that the integration path is along the real part of the \( h \) axis. In the event the claddings are lossy, the integration path may be along the real part of the \( h \) axis, but it must be modified to satisfy the end points which are complex numbers. The magnetic field requires \( \partial_z \psi(x, z) \) which is given by

\[
- \frac{\partial \psi(x, z)}{\partial z} = \sum_{\mu=0}^{M-1} A_{\mu}(h_{\mu}) \gamma(h_{\mu}) \psi_1(h_{\mu}, x) e^{-\gamma(h_{\mu}) z} \tag{4.5.10}
\]

\[
+ \int_{0}^{k_o \sqrt{k_2^2 - k_1^2}} j\beta(h) A_{\mu}(h) \psi_{\mu}(h, x) e^{-j\beta(h) z} \, dh
\]

\[
+ \int_{k_o \sqrt{k_2^2 - k_1^2}}^{k_2} j\beta(h) \left[ A_{\mu}(h) \psi_{\mu}(h, x) + A_{\mu}(h) \psi_{\mu}^*(h, x) \right] e^{-j\beta(h) z} \, dh
\]

\[
+ \int_{k_o \sqrt{k_2^2 - k_1^2}}^{\infty} \alpha(h) \left[ A_{\mu}(h) \psi_{\mu}(h, x) + A_{\mu}(h) \psi_{\mu}^*(h, x) \right] e^{-\alpha(h) z} \, dh
\]

Because \( \beta(h) = \sqrt{k_1^2 k_2^2 - h^2} \), it should be noted that the integrands have branch points at \( h = \pm k_o k_2 \). Finally, all integration paths may be modified using Cauchy’s integral theorem.

**Example:** The waveguide structure of Fig. 4.5 is excited in such a fashion so that electric field is uniform in core region and zero in the substrate and superstrate claddings. Assume that the field is specified at \( z = 0 \), i.e. \( \psi(x, z = 0) = E_o \) only in the core. Determine the spectral density of the Case II radiation modes and the corresponding fields at large distances from the excitation point at \( z = 0 \).

Using the general field expression of (4.5.9), the expansion coefficient \( A_{\mu}(h) \) can be calculated from initial fields at \( z = 0 \), as follows

\[
A_{\mu}(h) = E_o \int_{x_1}^{x_0} \psi_{\mu}(h, x) \, dx = E_o \left[ Q_1 \frac{\sin h_1 d_1}{h_1} + 2 P_1 \left( \frac{\sin h_1 d_1/2}{h_1} \right)^2 \right]
\]

The above expression for \( A_{\mu} \) shows that it is dependent on \( C_0(h) \) and thus will be influenced by the location of the leaky modes that are in the neighborhood of \( h \) values in the interval \( 0 < h < k_o (\sqrt{k_2^2 - k_1^2}) \).

For the waveguide width \( d_1 = 4 \lambda_o \), the figure on the left shows \( A_{\mu}(h) \) as a function of the normalized substrate wavenumber. The extreme values of \( A_{\mu}(h) \) occur at \( h = 0.32k_o \), \( 0.58k_o \) and \( 0.76k_o \), which are identical to the peak values of \( C_0 \) as shown in Fig. 4.5.
To determine the Case II radiation mode field, it is necessary to evaluate the integral

$$\Psi_{ii}(x, z) = \int_{k_0 \sqrt{\kappa_2 - \kappa_0}}^{k_0 \sqrt{\kappa_2 - \kappa_0}} A_{ii}(h) e^{-j\beta(h)z} dh$$

The Case II modes have field values in all three layers, however, only the fields in the substrate layers will be evaluated. Thus, the field in the substrate layer, $x < x_1$, is given by

$$\Psi_{ii}(x, z) = \int_{k_0 \sqrt{\kappa_2 - \kappa_0}}^{k_0 \sqrt{\kappa_2 - \kappa_0}} A_{ii}(h) \left[ Q_2(h) \cos h(x - x_1) + \frac{P_2(h)}{h} \sin h(x - x_1) \right] e^{-j\beta(h)z} dh$$

$$= \int_{k_0 \sqrt{\kappa_2 - \kappa_0}}^{k_0 \sqrt{\kappa_2 - \kappa_0}} \frac{1}{2} A_{ii}(h) e^{-j\beta(h)z} \left\{ \left[ Q_2(h) + jP_2(h)/h \right] e^{-j(h(x-x_1))} [Q_2(h) - jP_2(h)/h] e^{j(h(x-x_1))} \right\} dh$$

At $z = 0$ the total power flow in the $z$ direction is contained within the core region, and as the wave propagates in the $z$ direction, the Case II radiation modes will drain the power flow outside the core. Accordingly, the terms in the above integral can be identified as propagation in the positive and negative $x$ direction. The term $\exp - jh(x-x_1)$ identifies propagation in the positive $x$ direction, while $\exp jh(x-x_1)$ identifies that in the negative $x$ direction. In this example, only the propagation in the negative (substrate) direction should contribute to the integral because the field in the core should migrate to the substrate.

At large distances from the excitation point, the integrals for $\Psi_{ii}(x, z)$ can be approximated by the stationary phase method as discussed in 'apA'. Using the transformation

$$x = R \sin \theta, \quad z = R \cos \theta,$$

the first integral becomes

$$\int_{k_0 \sqrt{\kappa_2 - \kappa_0}}^{k_0 \sqrt{\kappa_2 - \kappa_0}} \frac{1}{2} A_{ii}(h) [Q_2(h) + jP_2(h)/h] e^{-j(R(h \sin \theta + \beta \cos \theta))} dh$$

The stationary phase point satisfies

$$h \cos \theta - \beta(h) \sin \theta = 0.$$

Because $h^2 + \beta^2 = k_0^2 \kappa_2$, the stationary point is easily identified by writing

$$h = k_0 \sqrt{\kappa_2} \sin \theta, \quad \beta = k_0 \sqrt{\kappa_2} \cos \theta,$$

which yields, $\theta = \theta$, so that the stationary point occurs at $h = k_0 \sqrt{\kappa_2} \sin \theta$. In the substrate $\theta$ is negative so that the stationary point lies outside the integration interval and accordingly, the first integral may be dropped. This conclusion agrees with the physical observation discussed discussed. The principal value of $\Psi_{ii}$ in the substrate layer is determined from the second integral because it exhibits a stationary point within the integration interval $(0, k_0 \sqrt{\kappa_2 - \kappa_0})$ and is approximated by

$$\Psi(x, z) \approx \sqrt{\frac{j \pi}{2k_0 \sqrt{\kappa_2} R}} \cos \theta \left\{ A_{ii}(h) \left[ Q_2(h) - \frac{jP_2(h)}{h} \right] \right\} e^{-j\beta h \cos \theta}.$$

Recall that the functional dependence of $A_{ii}(h)$ is influenced by the roots representing the leaky modes of the waveguide. As the approximate solution of $\Psi_{ii}(x, z)$ mirrors the influence of the leaky modes, the field at large $R$ values will exhibit major...
radiation lobes at $\theta$ values determined by the proximity of the leaky mode roots to the integration path.

<table>
<thead>
<tr>
<th>$h/k_o$</th>
<th>$\theta = \arcsin \left( \frac{h}{\sqrt{2} k_o} \right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3253</td>
<td>$-5.3^\circ$</td>
</tr>
<tr>
<td>0.5776</td>
<td>$-9.5^\circ$</td>
</tr>
<tr>
<td>0.7633</td>
<td>$-12.6^\circ$</td>
</tr>
</tbody>
</table>

According to Fig. 4.5, the waveguide with a core width of $4 \lambda_o$ has three peak values of $C_0(h)$ as illustrated in the table on the left. The corresponding values of $\theta$ are also given in the table.

The fields may be computed by evaluating the integral representing $\Psi_{II}(x, y)$. The figure below shows the magnitude of the field as a function of $x$ at various $z$ values. To illustrate the integral evaluation of $\Psi_{II}(x, y)$, the lateral fields in the substrate will be calculated at various distances from the excitation point located at $z = 0$. This exact field calculation is contrasted with the approximate method as discussed above which is only valid for large values of $R$. The fields will be calculated as a function of $x$ (in the substrate) at the axial points $z = 100 \lambda_o$, $200 \lambda_o$, $400 \lambda_o$, and $600 \lambda_o$.

The negative $x$ values correspond to points in the substrate and the plot points vary from $x = x_1 = -d_1$ to $-200 \lambda_o$. At excitation point $z = 0$, the total power flux in the waveguide is confined to the core, which has a width of only $4 \lambda_o$, however. The figure shows that almost all of the Case II power flux is in the substrate layer after a distance of about $100 \lambda_o$. At $z = 200 \lambda_o$, the radiation fields are forming the three major lobes that are positioned according to the leaky mode poles.

### 4.6 Spectral Power Flux

The power flow can be determined using the spectral expansion of the allowed modes of the structure. If the $y$ component of the electric field is given by (TE modes)

$$E_y = \Psi(x, z),$$

then the components of the magnetic field are

$$H_x = \frac{1}{jk_o Z_o} \frac{\partial \Psi(x, z)}{\partial z},$$

$$H_z = -\frac{1}{jk_o Z_o} \frac{\partial \Psi(x, z)}{\partial x}. $$

The total power flowing (per unit width in $y$) in the positive $z$ direction is computed from the Poynting vector

$$P_z = -\frac{1}{2} \Im \int_{-\infty}^{\infty} E_y H_x^* \, dx. $$
When the waveguide structure is reactive, and owing to the fact that the eigenfunctions are real,

\[
P_z = \frac{1}{k_0 Z_0} \left[ \sum_{n=0}^{M-1} \beta_n |A_i(h_n)|^2 + \int_0^{\frac{k_0}{\delta}} \frac{k_0^2 k_2 - h^2}{k_0^2 k_2 - \delta h^2} |A_{ii}(h)|^2 \, dh \right] + \int_0^{\frac{k_0}{\delta}} \frac{k_0^2 k_2 - h^2}{k_0^2 k_2 - \delta h^2} \left[ |A_{iii}^e|^2 + |A_{iii}^o|^2 \right] \, dh \tag{4.6.5}
\]

The Case IV radiation modes do not contribute to the Poynting vector and do not appear in (4.6.5).

### 4.7 Generalized Notation

The identification of the various mode sets for the different cases has been via the subscripts, i, ii, iii and iv. The Case I modes are discrete (and finite) while the remaining cases are composed of modes belonging to a continuous spectrum. Because the modes of the various cases can be identified by the substrate wavenumber \( h \), it is convenient to identify the modes by their appropriate \( h \) value. The modes of the degenerate spectrum, Case III and IV modes, it will be frequently necessary to distinguish the modes by the superscripts \( e \) and \( o \). Accordingly, an arbitrary function \( f(x) \) can be expanded as

\[
f(x) = \sum_{\mu} F(h_{\mu}) \psi(h_{\mu}, x). \tag{4.7.1}
\]

where the expansion coefficients \( F(h_{\mu}) \) are written as

\[
F(h_{\mu}) = \int f(x) \psi(h_{\mu}, x) \, dx. \tag{4.7.2}
\]

For the modes of the continuous spectrum, the summation symbol, \( \sum \) implies integration. For example, the Case III “even modes” of (4.7.1) are represented by

\[
\int_{k_0 \sqrt{\delta}}^{k_0 \sqrt{\kappa_2}} F_{iii}^e(h) \psi_{iii}^e(h, x) \, dh,
\]

while the expansion coefficient \( F_{iii}^e(h) \) is given by

\[
F_{iii}^e(h) = \int f(x) \psi_{iii}^e(h, x) \, dx,
\]

and is recognized as a generalized “Fourier” coefficient, whose kernel is \( \psi_{iii}^e(h, x) \).

### Problems

4–1 For TM slab modes in a non–magnetic medium, the \( y \)–component of the magnetic field satisfies

\[
\frac{\partial^2 H_y}{\partial x^2} + (k_0^2 \kappa_e + \gamma^2) H_y = 0.
\]
where the electric field components are given by

\[ E_x = \frac{\gamma}{\omega \varepsilon_0 \kappa_x} H_y, \quad E_z = \frac{1}{\omega \varepsilon_0 \kappa_x} \frac{\partial H_x}{\partial x}. \]

Determine the secular equation for the trapped modes using the vector techniques where

\[ \phi_1(x) = \psi(x), \quad \phi_2(x) = \frac{1}{\kappa_x} \frac{d\psi}{dx}, \]

and the magnetic field for the trapped mode is \( H_y(x, z) = H_o \psi(x) e^{-\gamma z}. \) Write the transfer matrix across a single layer.

4-2 Symmetric three-layer structures have two semi-infinite cladding layers with \( n_0 = n_2. \) Because of symmetry about the center of the core layer, modes may be either even or odd functions about the optical center of the core. (The optical center of the core is always the geometric center when the core consists of a single layer.)

The transfer matrix from \( x_1 \) to \( x_{oc} \) satisfies

\[ T_{oc} = \sqrt{T_1}, \]

where

\[ T_1 = \begin{pmatrix} \mathcal{A}_1 & \mathcal{B}_1 \\ \mathcal{C}_1 & \mathcal{D}_1 \end{pmatrix} = V_1 X_1 V_1^{-1}. \]

The matrices \( V_1 \) and \( X_1 \) are the eigenvectors and eigenvalues of \( T_1. \) The eigenvalue matrix is

\[ X_1 = \begin{pmatrix} e^{-j \theta_1} & 0 \\ 0 & e^{j \theta_1} \end{pmatrix}, \]

where \( \cos \theta_1 = (\mathcal{A}_1 + \mathcal{D}_1)/2. \)

(a.) Show that

\[ T_{oc} = V_1 \begin{pmatrix} e^{-j \theta_1/2} & 0 \\ 0 & e^{j \theta_1/2} \end{pmatrix} V_1^{-1}. \]

(Note that the eigenvectors of \( T_{oc} \) are identical to the eigenvectors of \( T_1 \) while the eigenvalues of \( T_{oc} \) have a simple relationship with the eigenvalues of \( T_1. \)) Find the matrix elements \( \mathcal{A}_{oc}, \mathcal{B}_{oc}, \mathcal{C}_{oc}, \) and \( \mathcal{D}_{oc}. \)

(b.) Show that the field coefficients at \( x_1 \) and \( x_0 \) can be written in terms of the field coefficients at \( x_{oc} \) as

\[ \Phi_0 = T_{oc} \Phi_{oc}, \quad \Phi_1 = T_{oc}^{-1} \Phi_{oc}. \]

where the even modes use the condition

\[ \Phi_{oc}^e = \begin{pmatrix} Q_{oc}^e \\ 0 \end{pmatrix}, \]
and the odd modes use the condition
\[ \Phi_{oc}^o = \begin{pmatrix} 0 \\ P_{oc}^o \end{pmatrix}. \]

(c) Show that the secular equation for even proper modes is
\[ jh_0 = h_1 \tan \theta_{oc}. \]
and for odd proper modes is
\[ jh_0 = -h_1 \cot \theta_{oc}. \]

4.3 The Case III and IV radiation modes are degenerate (two sets of modes have identical eigenvalues, in the continuous spectrum). According to (4.4.48), the two sets of modes can be made orthogonal by appropriate choice of \( \varrho \). Transform \( \Phi_{oc}^e \) and \( \Phi_{oc}^o \) to the superstrate and substrate cladding layers, and use the orthogonally relation
\[ Q_{ei}^2 Q_{oi}^2 + P_{ei}^2 P_{oi}^2 / h^2 + (Q_{ei}^2 Q_{oi}^2 + P_{ei}^2 P_{oi}^2 / h_0^2) h_0 / h = 0, \]
to calculate \( \varrho \). Show that \( \varrho \to 0 \) when \( h_0 \to h \), the “symmetric waveguide.”

4.4 The degenerate modes for the Case III and Case IV radiation modes can be separated by specifying an additional relation between the various coefficients \( Q_{ei}(h) \), \( P_{ei}(h) \), \( Q_{oi}(h) \) and \( P_{oi}(h) \). For example, a relation can be written in which (4.4.39) as
\[ 1 = \frac{\pi}{2} \left\{ Q_{ei}^2(h) \left[ 1 + \frac{P_{ei}^2(h)}{h^2 Q_{ei}^2(h)} \right] + Q_{oi}^2(h) \left[ 1 + \frac{P_{oi}^2(h)}{h_0^2 Q_{oi}^2(h)} \right] \delta(h - h') \right\} \]
is simplified by using
\[ \tan \varphi_2(h) = \frac{P_{ei}(h)}{h Q_{ei}(h)} = -\frac{P_{oi}(h)}{h_0 Q_{oi}(h)} = -\tan \varphi_0(h). \]
Show that the two sets of modes have
\[ \tan \varphi_2(h) = -\frac{(h_0 A + h D) \pm \sqrt{(h_0 A - h D)^2 + 4 h h_0}}{2 h h_0 B}, \]
where the different signs correspond to different sets. Show that the two sets are orthogonal. \( \text{(Note: For a three–layer structure, these two sets reduce to the even and odd modes when } n_0 = n_2. \)
Chapter 5
Radiation Modes from Plane Waves

In Chapter 4 the TE radiation modes were calculated by solving Maxwell’s equations with the assumption that the modes were propagating along a single direction, viz. the $z$ direction. The radiation modes were then characterized using various types of cases determined from the value of the propagation constant. For example, the Case II radiation modes consisted of fields that extend to infinite distances in the substrate region while the fields exponentially decay in the superstrate. (Without loss of generality the refractive index of the superstrate is assumed to be larger than that of the substrate.) Another case, Case III radiation modes were obtained by allowing the fields to extend in both substrate and superstrate regions. In all cases, the fields remain finite as $|x| \to \infty$.

The radiation modes of layered waveguides may also be calculated from plane waves that are incident on the central slab region. These waves are incident from sources in both the superstrate and substrate regions of the waveguide.

The important results of this chapter are that the structure of the new radiation modes will allow for two single wave functions to characterize the modes over the total range $(0, \infty)$ of the transverse wave number $\beta$ as described in Chapter 4.

5.1 Introduction

While the radiation modes discussed in Chapter 4 were characterized by three cases:

II Field extensions only in the substrate region. The modes propagate according to $\exp(-j\beta z)$.
III Field extensions in both the substrate and superstrate regions. The modes propagate as $\exp(-j\beta z)$.

IV Field extensions in both the substrate and superstrate regions. The modes are evanescent and propagate as $\exp(-\alpha z)$.

The complex propagation constant $\gamma = \alpha + j\beta$ is continuous and takes on a range of values that characterize the different cases. The attenuation constant $\alpha = 0$ for cases II and III, while $\beta = 0$ for IV.

The structure of the radiation modes composed of plane waves as discussed in this chapter will be a combination of the radiation modes in cases II, III and IV. The resulting radiation modes will be designated as Case A and Case B. Since Cases II, III and IV form a complete set of radiation modes, the Case A and B radiation modes are obviously a linear combination of the radiation modes of Chapter 4.

5.2 Construction of Case A Radiation Modes

The radiation modes designated as Case A will be composed of planes waves whose source is located at $x = -\infty$. Without loss of generality, the refractive index of Layer 0 is smaller than the refractive index of Layer 2. A specific example is one where the substrate is a semiconductor, while the superstrate is air. The incident waves

![Figure 5.1. Case A radiation modes are composed of the set of plane waves whose electric and magnetic fields are $E_i$ and $H_i$. Although a single layer is shown of the central waveguide structure, any central waveguide region can be characterized using the transfer matrix methods as given in Chapter 3. The layers are characterized only by their refractive indices, $n_0$, $n_1$ and $n_2$. Without loss of generality, it is assumed that $n_2 > n_0$.](image-url)
5.2 CONSTRUCTION OF CASE A RADIATION MODES

Electric and magnetic fields have the following components

\[ E_{1y} = a e^{-jk_2[(x-x_1) \cos \theta_2 + z \sin \theta_2]} \]  \hspace{1cm} (5.2.1)
\[ H_{1x} = -a \frac{n_2}{Z_0} \sin \theta_2 e^{-jk_2[(x-x_1) \cos \theta_2 + z \sin \theta_2]} \]  \hspace{1cm} (5.2.2)
\[ H_{1z} = a \frac{n_2}{Z_0} \cos \theta_2 e^{-jk_2[(x-x_1) \cos \theta_2 + z \sin \theta_2]} \]  \hspace{1cm} (5.2.3)

where the complex parameter \( a \) is a constant, \( Z_0 \) is the free-space wave impedance, \( k_2 = k_0 n_2 \) is the wave number of layer 2. (In the present discussion, the coefficient \( a \) has an arbitrary phase which can be assumed to be zero, however its amplitude can be determined from normalization.) It should be noted that only the field components that are perpendicular to \( x \) affect the transmission fields \( E_t, H_t \), and reflected fields \( E_r, H_r \). Thus, to calculate the reflected and transmitted fields, use the procedures of Chapter 3 and replace the refractive indices \( n_0 \) with \( n_0 \cos \theta_0, n_1 \) with \( n_1 \cos \theta_1 \) and \( n_2 \) with \( n_2 \cos \theta_2 \). The angles \( \theta_0, \theta_1, \) and \( \theta_2 \) are related via Snell’s law

\[ \frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1}, \quad \frac{\sin \theta_0}{\sin \theta_1} = \frac{n_1}{n_0}, \quad \text{so that} \quad \frac{\sin \theta_0}{\sin \theta_2} = \frac{n_2}{n_0}. \]  \hspace{1cm} (5.2.4)

Since only the \( z \) component of the magnetic field is used to calculate the reflected and transmitted fields, its \( x \) component will be dropped in the discussion below. The pertinent reflected substrate fields can be written as

\[ E_{r_y} = a \rho_A e^{-jk_2[-(x-x_1) \cos \theta_2 + z \sin \theta_2]}, \]  \hspace{1cm} (5.2.5)
\[ Z_0 H_{rz} = -a \rho_A n_2 \cos \theta_2 e^{-jk_2[-(x-x_1) \cos \theta_2 + z \sin \theta_2]}, \]  \hspace{1cm} (5.2.6)

while the transmitted fields have the form

\[ E_{t_y} = a \tau_A e^{-jk_0[(x-x_0) \cos \theta_0 + z \sin \theta_0]}, \]  \hspace{1cm} (5.2.7)
\[ Z_0 H_{t_z} = a \tau_A n_0 \cos \theta_0 e^{-jk_0[(x-x_0) \cos \theta_0 + z \sin \theta_0]}, \]  \hspace{1cm} (5.2.8)

where Snell’s law relates the angle \( \theta_0 \) to \( \theta_2 \)

5.2.1 Properties of Case A Fields

With the general construction of the Case A radiation modes established as discussed above, their fields will now be characterized. A specific mode is characterized by the angle \( \theta_2 \), the incident angle of the plane wave in the substrate. In Chapter 4, the modes were characterized by substrate wave number which ranged from 0 to \( \infty \). The Case A modes may also be characterized by the substrate wave number by an appropriate mapping. Accordingly, the total fields in the substrate may be written as

\[ E_{2y} = a[e^{-jk_2(x-x_1)} + \rho_A e^{j\hat{k}_2(x-x_1)}]e^{-\gamma z}, \]  \hspace{1cm} (5.2.9)
\[ k_0 Z_0 H_{2z} = a \hat{k}_2[e^{-jk_2(x-x_1)} - \rho_A e^{j\hat{k}_2(x-x_1)}]e^{-\gamma z}, \]  \hspace{1cm} (5.2.10)
where \( h_2 = k_2 \cos \theta_2 \) and \( \beta = \Im\{\gamma\} = k_2 \sin \theta_2 \). The total fields in the superstrate become

\[
E_{0y} = a \tau_A e^{-j h_0 (x-x_0)} e^{-\gamma z}, \tag{5.2.11}
\]
\[
k_0 Z_0 H_{0z} = a \tau_A h_0 e^{-j h_0 (x-x_0)} e^{-\gamma z}. \tag{5.2.12}
\]

where \( h_0 = k_0 \cos \theta_0 \). The field equations now have a functional dependence similar to that discussed in Chapter 4 where the variable \( h_2 = k_2 \cos \theta_2 \equiv h \).

When \( h = 0 \), \( \theta_2 = \pi/2 \). Assuming the critical angle \( \theta_c = \arcsin(n_0/n_2) \), the power in the incident wave will be totally reflected when when \( \pi/2 > \theta_2 > \theta_c \). In terms of transverse wave number \( h \), total reflection occurs for \( 0 < h < k_2 \cos \theta_c \) or \( 0 < h < k_0 \sqrt{n_2^2 - n_0^2} \). Also note that when \( \theta_2 > \theta_c \), the angle \( \theta_0 \) is complex as illustrated in Fig. 5.2. Under these conditions, fields will exponentially decay with \( x \) in Layer 0. It should be noted that the Case A radiation modes contain the Case II modes, i.e., Case II modes form a subset of Case A modes.

When \( \theta_2 > \theta_c \) the incident wave in Layer 2 will transfer power to Layer 0 producing waves that propagate without decay, however, the resulting radiation modes are not equivalent to the Case III radiation modes because the transverse propagation of the fields have only waves traveling in the positive \( x \) direction in Layer 0.

The \( x \) component of the complex Poynting vector \( S_x \) can be evaluated at the interfaces, \( x = x_1 \) and \( x = x_0 \). At substrate boundary, the time-average power flux is

\[
\frac{1}{2} \Re\{E_{2y} H_{2z}^*\}_{x=x_0} = \frac{1}{2 Z_0 k_0} |a|^2 (1 - |\rho_A|^2) h. \tag{5.2.13}
\]

The complex variable \( h \) may assumed to be a real positive quantity for incident plane waves with \( \pi/2 > \theta_2 > 0 \). At the superstrate boundary,

\[
\frac{1}{2} \Re\{E_{2y} H_{2z}^*\}_{x=x_0} = \frac{1}{2 Z_0 k_0} |a|^2 |\tau_A|^2 \Re\{h_0\}. \tag{5.2.14}
\]

When the plane waves from a source at \( x = -\infty \) are incident on the central slab region, the angle \( \theta_2 \) represents the incident angle relative to the layer interface normal. However, in general the angle may be a complex quantity. Specifically, if \( \theta_2 = \theta_2' + j \theta_2'' \), then on the imaginary axis, \( \cos \theta_2 = \cosh \theta_2'' \) is real, whereas, \( \sin \theta_2 = j \sinh \theta_2'' \) is imaginary.

For all \( \theta_2 \) angles, the conservation of power produces the relation between \( \rho_A \) and \( \tau_A \) as

\[
1 = \begin{cases} 
|\rho_A|^2 + \frac{h_0}{h_2} |\tau_A|^2, & \theta_2 < \theta_c; \\
|\rho_A|^2, & \theta_c < \theta_2 < \pi/2. 
\end{cases} \tag{5.2.15}
\]

The complex propagation constant \( \gamma = \alpha + j \beta \) so that for propagation along \( z \), \( \alpha = 0 \).
In Chapter 4 the radiation modes were characterized by $h$, the transverse wave number of the substrate. Here, the radiation modes are characterized by the plane wave propagating at an angle $\theta_2$ relative to the transverse direction $x$, as illustrated in Fig. 5.1. The Case II and Case III radiation modes can be partially composed of the Case A modes where $\theta_2$ varies from $\pi/2$ to 0. However, the evanescent modes of Case IV cannot be obtained from $\theta_2$ values when it is a real number. The evanescent modes must be determined from complex $\theta_2$ values that are imaginary. Mapping of complex values of $\theta_2$ to the transverse substrate wave number $h$, and to the $\theta_0$ plane are illustrated in Fig. 5.2. When $\theta_2 = \pi/2$, $\theta_0 = \pi/2 + j \cosh^{-1}(n_2/n_0)$.

When the electric field of the Case A modes is written as $E_y = \Psi_A(h, x, z) = \psi_A(h, x) \exp -jz$, the transverse wave function $\psi_A(h, x)$ is given by

$$\psi_A(h, x) = a(h) \begin{cases} e^{-j\theta_0(x-x_0)}, & x_0 < x \\ q_A \cos h_1(x-x_1) + \frac{\rho_A}{j\theta_1} \sin h_1(x-x_1), & x_1 < x < x_0, \\ e^{-j\theta_1(x-x_1)} + \rho_A e^{j\theta_1(x-x_1)}, & x < x_1. \end{cases}$$

The coefficient $a(h)$ is a function of independent variable $h$ which will be calculated by normalizing the fields.) The axial magnetic field of the Case A modes, $H_z$, satisfies

$$Z_0 H_z = \frac{j}{k_0} \frac{\partial \Psi_A}{\partial x}. \quad (5.2.17)$$

By matching $E_y$ and $H_z$ at the $x_0$ and $x_1$ interfaces, the transmission coefficient $\tau_A$ is

$$\tau_A(h) = \frac{2h_1 h}{(h + h_0) h_1 \cos h_1 d_1 + j(h_1^2 + h_0 h) \sin h_1 d_1}; \quad (5.2.18)$$

whereas, the reflection coefficient $\rho_A$ is

$$\rho_A(h) = \frac{(h - h_0) h_1 \cos h_1 d_1 - j(h_1^2 - h_0 h) \sin h_1 d_1}{(h + h_0) h_1 \cos h_1 d_1 + j(h_1^2 + h_0 h) \sin h_1 d_1}. \quad (5.2.19)$$
These coefficients, which are written as functions of the transverse wave number, \( h \), have poles at the roots characterizing the proper and improper modes (leaky modes). In particular, the denominators of (5.2.19) and (5.2.18) are identical to (4.4.34). (The coefficients \( q_1 \) and \( p_1 \) are determined from field continuity at \( x = x_1 \).) In the limit \( h \to \infty \), \( |\tau_A| \to 1 \) and \( |\rho_A| \to 0 \).

To illustrate the Case A modes, the transmission coefficient \( \tau_A \) and reflection coefficient \( \rho_A \) will be calculated for the waveguide structure identical to that shown in Fig. 4.7. The refractive indices and waveguide width are shown in the figure. Figure 5.3 shows that the functional dependence of \( |\tau_A| \) with respect to \( h \). The functional dependence of the coefficient \( C_0 \) in the Case II radiation modes is identical to that of \( |\tau_A| \) because they represent the field strength at the \( x = x_0 \) interface, provided the normalization coefficient \( a(h) \) in (5.2.16) is a constant. (The value of \( a(h) \) is calculated in Section 5.2.2.) Note also that the figure shows the real and imaginary parts of \( \tau_A \) indicating the peak values of \( |\Re\{\tau_A\}| \) are identical to the peak values of \( |\tau_A| \). Figure 5.4 shows the reflection coefficient as a function of the substrate transverse wave number \( h \), normalized to the free-space wave number. As noted earlier, these oscillations are dependent on the proximity of the roots characterizing improper modes. In addition to the singular points defined by the improper modes, the functions \( \tau_A(h) \) and \( \rho_A \) have an additional singularity at

\[
h = k_0 \sqrt{k_2 - \kappa_0}
\]

because \( h_0 = 0 \) at that point. In particular the singularity here is a branch point which must be addressed in contour integrations involving the Case A modes.

As illustrated in Figs. 5.3 and 5.4 the Case A radiation modes are strongly influenced by the roots of the improper modes in the interval \( 0 < h < k_0 \sqrt{k_2 - \kappa_0} \). When \( h > k_0 \sqrt{k_2 - \kappa_0} \), the roots due to improper modes have little or no influence on the oscillation of \( \tau_A \). It appears the behavior of \( \tau_A \) on \( h \) has settled...
out for \( h > 1.5k_0 \), however, \( \rho_A \) is more sensitive to the roots of the improper modes.

![Figure 5.4](image)

Figure 5.4. The transmission coefficient \( \rho_A \) as a function of the substrate transverse wave number \( h/k_0 \). The three-layer waveguide is shown in Fig. 4.7.

The damping of \(|\rho_A|\) occurs because the value of \( \rho_A \) will approach a small value in the present example of about \((n_2 - n_0)/(n_2 + n_0) = 0.0145\). Larger asymptotic values of \( \rho_A \) occur when there is a larger refractive index difference between the superstrate and substrate.

### 5.2.2 Orthonormalization of Case A Radiation Modes

The orthonormalization of the Case A radiation modes can be performed using only the fields in the superstrate and substrate regions. Assuming two different wave functions \( \Psi(h, x, z) \) and \( \Psi(\hat{h}, x, z) \), orthonormalization will be accomplished using only the transverse fields \( \psi(h, x) \) and \( \psi(\hat{h}, x) \).

While the fields in Chapter 4 were real wave functions \( \psi(h, x) \), the fields here are complex functions so the procedure of ‘apor’ will be used. In particular,

\[
\langle \psi_A(\hat{h}, x), \psi_A(h, x) \rangle_X = \int_{x_1-X}^{x_0+X} \psi_A^*(\hat{h}, x) \psi_A(h, x) \, dx, \tag{5.2.21}
\]

will be considered for two ranges of the independent variables \( h \) and \( \hat{h} \): (1) when either \( h \) or \( \hat{h} \) (or both) lie in the interval \((0, k_0\sqrt{n_2^2 - n_0^2})\) and (2) when neither lie in the interval \((0, k_0\sqrt{n_2^2 - n_0^2})\).

When \(|\rho_A| = 1\) either (or both) \( \psi(h, x) \) or \( \psi(\hat{h}, x) \) decay exponentially with \( x \) in the superstrate so that

\[
\psi_0^*(\hat{h}, X) \psi_0(h, X) - \psi_0^*(\hat{h}, X) \psi_0(h, X) \to 0 \quad \text{as} \quad X \to \infty.
\]
Using the substrate field expressions in (5.2.16), (5.2.21) becomes (See (B.3.).)

\[
\langle \psi_A(h, x), \psi_A(h, x) \rangle_X = |a|^2 \left\{ -j \frac{\cos(h - \tilde{h})X}{h - \tilde{h}} \left[ 1 - \rho_A(h) \rho_A^*(\tilde{h}) \right]
+ \frac{\sin(h - \tilde{h})X}{h - \tilde{h}} \left[ 1 + \rho_A(h) \rho_A^*(\tilde{h}) \right]
+ \rho_A^*(\tilde{h}) e^{j(h + \tilde{h})X} - \rho_A(h) e^{j(h + \tilde{h})X} \right\}
\]

(5.2.22)

Without loss of generality, assume \( \tilde{h} \in (0, k_0 \sqrt{n_2^2 - n_0^2}) \). Since the oscillation frequency of the third term on the RHS of (5.2.22), increases without limit as \( X \to \infty \) and \( h + \tilde{h} \) is not zero, the second term may be dropped because the integration with respect to \( h \), of product of this function with any well-behaved function of \( h \) will be zero. The first term also has an oscillation frequency that increases without limit except at the singularity \( h = \tilde{h} \), however, integration about the simple singularity is of principle value. Thus, the first term may be discarded.

On the other hand, the second term has an impulse singularity at \( h = \tilde{h} \) and because \( \rho_A(h) \rho_A^*(\tilde{h}) \to 1 \) as \( h \to \tilde{h} \), (5.2.22) becomes in the limit \( X \to \infty \),

\[
\langle \psi_A(\tilde{h}, x), \psi_A(h, x) \rangle = \delta(h - \tilde{h}),
\]

(5.2.23)

provided \( a = 1/\sqrt{2\pi} \). (The phase of \( a \) is arbitrary and is assumed zero.)

The value of the constant \( a \) will now be determined when both \( h \) and \( \tilde{h} \) lie outside the interval \( (0, k_0 \sqrt{n_2^2 - n_0^2}) \). To evaluate \( a \), the superstrate term which will be added to the brackets in (5.2.22), is

\[
j |a|^2 \tau_A(h) \tau_A^*(\tilde{h}) e^{-j(h_0 - \tilde{h}_0)X} =
\]

\[
|a|^2 \tau_A(h) \tau_A^*(\tilde{h}) \left[ \frac{j \cos(h_0 - \tilde{h}_0)X}{h_0 - \tilde{h}_0} + \frac{\sin(h_0 - \tilde{h}_0)X}{h_0 - \tilde{h}_0} \right]
\]

The first term on the right-hand-side of the above equation will be dropped because its frequency of oscillation increases with \( X \) when \( h_0 \neq \tilde{h}_0 \), and the singularity at \( h = \tilde{h} \) is simple and will be dropped owing to the fact that all later integrations will be of principle value type. Accordingly, as \( X \to \infty \),

\[
j |a|^2 \tau_A(h) \tau_A^*(\tilde{h}) \frac{e^{-j(h_0 - k_0)X}}{h_0 - \tilde{h}_0} \to \pi |a|^2 |\tau_A(h)|^2 \delta(h_0 - \tilde{h}_0)
\]

\[
= \pi |a|^2 |\tau_A(h)|^2 \frac{h}{h_0} \delta(h - \tilde{h}).
\]

(5.2.24)

When \( |\rho_A| \neq 1 \) the second bracketed term in (5.2.22) must be added to (5.2.24), and in the limit \( X \to \infty \),

\[
\langle \psi_A(\tilde{h}, x), \psi_A(h, x) \rangle = \pi |a|^2 \left( 1 + |\rho_A(h)|^2 + \frac{h}{h_0} |\tau_A(h)|^2 \right) \delta(h - \tilde{h}).
\]

(5.2.25)
Owing to the conservation of power, (5.2.15), the constant \(a\) is given by
\[
a = \frac{1}{\sqrt{2\pi}}.
\] (5.2.26)

The value of \(a\) given is (5.2.26) is valid for \(0 < h < \infty\). Thus, the transverse Case A radiation modes are orthogonal and can be written as
\[
\psi_A(h, x) = \frac{1}{\sqrt{2\pi}} \begin{cases} 
\tau_A(h) e^{-j\frac{h_0(x-x_0)}{}} & x_0 < x \\
q_1(h) \cos h_1(x-x_1) + \frac{p_1(h)}{j h_1} \sin h_1(x-x_1), & x_1 < x < x_0, \\
e^{-j h(x-x_1)} + \rho_A(h) e^{j h(x-x_1)}, & x < x_1.
\end{cases}
\] (5.2.27)

These radiation modes are continuous in the variable \(h\) in the range \((0, \infty)\), and are normalized according to (5.2.23). The coefficients \(\tau_A(h), q_1(h), p_1(h)\), and \(\rho_A(h)\) have a branch point, at \(h = k_0 \sqrt{k_2 - k_0}\) and simple poles at points characterizing proper and improper modes of the structure.

Finally, the transverse components of the Case A electromagnetic fields propagating in the positive \(z\) direction are
\[
E_y(h, x, z) = \psi_A(h, x) e^{-j\gamma z},
\] (5.2.28)
\[
Z_0 H_x(h, x, z) = \frac{j \gamma}{k_0} \psi_A(h, x) e^{-j\gamma z}.
\] (5.2.29)

### 5.3 Construction of Case B Radiation Modes

The Case B radiation modes will be constructed similar to that for the Case A modes. While the Case A modes are constructed with a plane wave incident from the substrate region on the waveguide core, the Case B modes will be constructed from a plane wave that is incident on the core region of the waveguide from the superstrate. Here the incident wave produces a forward and backward (with respect to \(x\)) wave in the superstrate (Fig. 5.5) while there will only be a backward wave in the substrate.

![Figure 5.5](image_url)

**Figure 5.5.** Case B radiation modes are composed of the set of plane waves whose electric and magnetic fields are \(E_i\) and \(H_i\). As in Fig. 5.1 the layers are characterized only by their refractive indices, \(n_0\), \(n_1\) and \(n_2\).
The incident fields are given by

\[ E_{ix} = b e^{j k_0[(x-x_0) \cos \theta_0 - z \sin \theta_0]}, \quad (5.3.1) \]
\[ Z_0 H_{ix} = -b n_0 \sin \theta_0 e^{j k_0[(x-x_0) \cos \theta_0 - z \sin \theta_0]}, \quad (5.3.2) \]
\[ Z_0 H_{iz} = -b n_0 \cos \theta_0 e^{j k_0[(x-x_0) \cos \theta_0 - z \sin \theta_0]} \quad (5.3.3) \]

The incident waves in the superstrate region are similar to the incident fields in the substrate region of the Case A radiation modes. The relation between the angles \( \theta_0, \theta_1 \) and \( \theta_2 \) satisfies Snell’s law as given in (5.2.4).

Relevant reflected fields in Layer 0 are

\[ E_{rx} = b \rho_B e^{-j k_0[(x-x_0) \cos \theta_0 - z \sin \theta_0]}, \quad (5.3.4) \]
\[ Z_0 H_{rz} = b \rho_B n_0 \cos \theta_0 e^{-j k_0[(x-x_0) \cos \theta_0 + z \sin \theta_0]} \quad (5.3.5) \]

while the relevant transmitted fields in Layer 2 are

\[ E_{tx} = b e^{j k_0[(x-x_1) \cos \theta_2 - z \sin \theta_2]}, \quad (5.3.6) \]
\[ Z_0 H_{tz} = -b n_2 \cos \theta_2 e^{j k_0[(x-x_1) \cos \theta_2 - z \sin \theta_2]} \quad (5.3.7) \]

The Case B radiation modes have propagating fields transmitted (Layer 0 to Layer 2) through the central core region, similar to that (Layer 2 to Layer 0) of the Case A modes, however, the Case B modes with plane wave sources at \( x = \infty \) cannot produce evanescent fields (exponential decay along the \( x \) direction) in the substrate because \( n_2 > n_0 \). In Fig. 5.2, Case B radiation modes start at \( \theta_0 = \pi/2 \), and \( \theta_0 \) varies from \( \pi/2 \) to 0 on the real axis and continues to vary along the negative imaginary axis. The angle \( \theta_2 \) varies along the real axis in the figure, from \( \theta_c = \arcsin(n_0/n_2) \) to 0 and continues along the negative imaginary axis.

### 5.3.1 Properties of Case B Fields

Although the Case B radiation modes are physically characterized by the angle \( \theta_0 \), they may be equally characterized by \( \theta_2 \) because of Snell’s law. Furthermore, \( h = h_2 = k_0 n_2 \cos \theta_2 \) can be considered the independent variable. Similar to the development of the Case A TE modes in Section 5.2.1, the Case B TE modes are now defined.

The fields are now written in terms of the complex propagation constant \( \gamma = \alpha + j \beta \) and the vertical wave numbers as

\[ E_{0y} = b [e^{j h_0(x-x_0)} + \rho_B e^{-j h_0(x-x_0)}] e^{-\gamma z}, \quad (5.3.8) \]
\[ k_0 Z_0 H_{0z} = -b h_0 [e^{j h_0(x-x_0)} - \rho_B e^{-j h_0(x-x_0)}] e^{-\gamma z}, \quad (5.3.9) \]

for the superstrate while the relevant substrate fields are given by

\[ E_{2y} = a \tau_B e^{j h_2(x-x_1)} e^{-\gamma z}, \quad (5.3.10) \]
\[ k_0 Z_0 H_{0z} = a \tau_B h_0 e^{-j h_0(x-x_0)} e^{-\gamma z}, \quad (5.3.11) \]
5.3 CONSTRUCTION OF CASE B RADIATION MODES

For all real angles, \( \pi/2 > \theta_2 > 0 \), the Case B radiation modes propagate along \( z \), however, in the range \( 0 > \Im\{\theta_0\} < -\infty \), the Case B radiation modes are evanescent along \( z \).

The conservation of power in the \( x \) direction yields

\[
1 = |\rho_B|^2 + \frac{\hat{h}_2}{\hat{h}_0} |\tau_B|^2, \quad \text{all allowed } \theta_0 \text{ values,} \tag{5.3.12}
\]

where the allowed \( \theta_0 \) values are illustrated in Fig. 5.6.

![Figure 5.6. The mapping from the complex \( \theta_0 \)-plane to the \( \theta_2 \)-plane and the \( h \)-plane. The Case B modes form a continuous set determined by allowed points in the \( \theta_0 \) plane and for \( h' > h_c \).]

The lateral component of the electric field for the Case B radiation modes are

\[
E_y = \Psi_B(h, x, z) = \psi_B(h, x) \exp(-\gamma z), \quad \text{where } \psi_B(h, x) \text{ satisfies}
\]

\[
\psi_B(h, x) = b(h) \begin{cases} 
    e^{j\h_0(x-x_0)} + \rho_B e^{-j\h_0(x-x_0)}, & x_0 < x \\
    q_B \cos \h_1(x-x_0) + \frac{\rho_B}{j\h_1} \sin \h_1(x-x_0), & x_1 < x < x_0, \\
    \tau_B e^{-j\h_2(x-x_1)}, & x < x_1. 
\end{cases} \tag{5.3.13}
\]

The axial magnetic field for the Case B radiation modes satisfies (5.2.17). Matching the fields at the various layer interfaces produces the transmission and reflection coefficients for the Case B radiation modes as follows

\[
\tau_B(h) = \frac{2\h_0 \h_1}{(h + \h_0) \h_1 \cos \h_1 d_1 + j(h_1^2 + \h_0 h) \sin \h_1 d_1}, \tag{5.3.14}
\]

\[
\rho_B(h) = \frac{(\h_0 - h) \h_1 \cos \h_1 d_1 - j(h_1^2 - \h_0 h) \sin \h_1 d_1}{(h + \h_0) \h_1 \cos \h_1 d_1 + j(h_1^2 + \h_0 h) \sin \h_1 d_1}. \tag{5.3.15}
\]

The transmission and reflection coefficients for the Case B modes are similar to the coefficients obtained for the Case A modes. In going from Case A to Case B, simply exchange the values of \( h \) and \( \h_0 \). Figure 5.7 shows the transmission coefficient is valid only for \( \tau_B \) for \( h > h_c \).
5.3.2 Orthonormalization of Case B Radiation Modes

Similar to the Case A radiation modes, the Case B radiation modes are orthogonal for functions defined with different \( h \) values. In particular, using the wave functions given in (5.3.13) the normalization procedure in ‘apox’ yields

\[
\langle \psi_B(\tilde{h}, x), \psi_B(h, x) \rangle = \lim_{X \to \infty} \int_{x_1 - X}^{x_0 + X} \psi_B^*(\tilde{h}, x) \psi_B(h, x) \, dx, \tag{5.3.16}
\]

\[
= |b|^2 \pi \left[ (1 + |\rho_B|^2) \delta(h_0 - \tilde{h}_0) + |\tau_B|^2 \delta(h - \tilde{h}) \right],
\]

\[
= |b|^2 \pi \left[ 1 + |\rho_B|^2 + \frac{\tilde{h}_0}{\tilde{h}} |\tau_B|^2 \right] \frac{\tilde{h}}{\tilde{h}_0} \delta(h - \tilde{h}).
\]
Using the conservation of power for the Case B modes, given in (5.3.12), the coefficient \( b(h) \) can be written as
\[
b(h) = \sqrt{\frac{1}{2\pi} \frac{\tilde{h}_0}{h}}.
\]

The Case B radiation modes have the functional dependence
\[
\psi_B(h,x) = \sqrt{\frac{\tilde{h}_0}{2\pi h}} \begin{cases} 
e^{rac{ih_0(x-x_0)}{h}} + \rho_B e^{-\frac{ih_0(x-x_0)}{h}}, & x_0 < x \\
q_B \cos \frac{h_1(x-x_0)}{h} + \frac{\rho_{B_1}}{j\tilde{h}_1} \sin \frac{h_1(x-x_0)}{h}, & x_1 < x < x_0, \\
\tau_B e^{-\frac{ih_2(x-x_1)}{h}}, & x < x_1.
\end{cases}
\]

The modes are orthogonal and normalized to unity. The coefficients in the core region are obtained by making the fields and their derivatives continuous across the boundaries.

### 5.4 Degeneracy of A and B Modes Sets

The radiation modes defined in Case A are orthogonal. Similarly, the modes defined in Case B are orthogonal,
\[
\langle \psi_A(h,x), \psi_A(h,x) \rangle = 0,
\]
\[
\langle \psi_B(h,x), \psi_B(h,x) \rangle = 0,
\]
provided \( \tilde{h} \neq h \). Further, the plane wave radiation mode sets are orthogonal to the discrete modes.

As discussed in Chapter 4, there was mode degeneracy in the Case III and Case IV radiation modes, i.e., two modes with the same propagation constant have different shapes. The Case A and Case B radiation modes are not degenerate in the range \( 0 < h < h_c \), however, they are degenerate in the range \( h > h_c \). Although the modes are degenerate, it will be shown that they are orthogonal, i.e., \( \langle \psi_B(h,x), \psi_A(h,x) \rangle = 0 \) for all \( h, \tilde{h} > h_c \).

Since both \( \psi_A(h,x) \) and \( \psi_B(h,x) \) satisfy the same differential equation, replace \( \psi(h,x) \) and \( \psi(h,x) \) in (B.3) with \( \psi_B(h,x) \) and \( \psi_A(h,x) \) respectively, which produces the following
\[
(h^2 - \tilde{h}^2) (\psi_B(h,x), \psi_A(h,x))_X = \psi^*_B(\tilde{h},X) \psi_A(\tilde{h},X) - \psi^*_A(\tilde{h},X) \psi_A(\tilde{h},X) - \psi^*_B(\tilde{h},X) \psi_A(\tilde{h},X) - \psi^*_A(\tilde{h},X) \psi_A(\tilde{h},X),
\]
where \( \psi_A(\tilde{h},X) \) and \( \psi_B(\tilde{h},X) \) represent the Case A and Case B fields in the superstrate region and \( \psi_A(\tilde{h},X) \) and \( \psi_B(\tilde{h},X) \) are the corresponding fields in the substrate. The superstrate term in the RHS of (5.4.1) can be written as
\[
jab^* [(\tilde{h}_0 - \tilde{h}) \tau_A e^{-j(h_0+\tilde{h}_0)X} + (\tilde{h}_0 + \tilde{h}) \tau_A \rho_A^* e^{-j(h_0-\tilde{h}_0)X}].
\]
where \( a \) and \( b \) are given by (5.2.26) and (5.3.17) respectively. The substrate term on the RHS of (5.4.1) reduces to

\[
jab^*[-(h - \tilde{h})\tau_B^*e^{j(h+\tilde{h})X} + (h + \tilde{h})\tau_B^*\rho_Ae^{-j(h-\tilde{h})X}].
\]

Dividing both sides of (5.4.1) by \( h^2 - \tilde{h}^2 = \tilde{h}^2 - \tilde{h}_0^2 \), and allowing \( X \to \infty \),

\[
\langle \psi_B(h, x), \psi_A(h, x) \rangle = jab^*[\tau_A\rho_A^*\delta(h_0 - \tilde{h}_0) + \tau_B^*\rho_A\delta(h - \tilde{h})].
\]

Owing to the relation between the scattering parameters as given in (3.9.15), the coefficient of the delta function is 0. Thus, the Case A radiation modes are orthogonal to the Case B radiation modes when each has a different eigenvalue as well as when the eigenvalues are identical.

The base system of modes are all orthogonal. The system now includes only the discrete trapped modes, the two sets of continuous (Case A and Case B) modes.
Chapter 6
Transitions in Layered Waveguides

This chapter will develop the fundamental formulation for analyzing discontinuities in layered waveguides. The techniques will rely on the spectral content of the waveguides as discussed in Chapter 4. The immediate applications of this material will be the simple joining to two different waveguide geometries, periodic structures, and radiation characteristics of optical waveguides.

6.1 Waveguide Geometries

In earlier chapters, development of the field equations were based on simple three-layer structures, a substrate cladding, a core, and a superstrate cladding. In many instances, the results were based on the single-layer transfer matrix, however, it should be noted that all results could be easily extended to multiple-layer (more than three) structures simply by modifying the elements of the transfer matrix.

Many contemporary photonic circuits are fabricated on wafers that consist of many layers of differing materials that are grown by some epitaxial process. Different waveguide structures can be formed by processing two regions differently. For example, the two waveguides shown in Fig. 6.1 contain different materials in the first layer, (top core layer), while the remaining layers are identical. Note that both waveguides are simple three-layer structures but are represented by four-layer structures. In this representation all layers have identical thicknesses which would naturally fit the geometries of actual epitaxial wafers.
6.2 Single Discontinuity

The structure in Fig. 6.1 shows a discontinuity at $z = 0$. In the region $z < 0$, the fields will be described by a set of modes, while the field in the region $z > 0$ will be described by a different set. The problem at hand is to assume a single bounded mode is incident (from the left) on the waveguide discontinuity and induces a set of modes transmitted into the $z > 0$ region and a set of reflected modes propagating into the $z < 0$ region. Figure 6.2 indicates the processes involved for the scattering problem. The propagation constants, $\gamma$ and $\bar{\gamma}$ are assumed complex to allow for lossy material.

![Figure 6.1. Schematic cross section of the four-layer optical waveguide with thickness discontinuity at $z = 0$. (a) The waveguide structures, and (b) the refractive index as a function of the $x$ coordinate for the two different sections.](image)

![Figure 6.2. The scattering processes occurring at the interface of two different waveguide structures.](image)
The modes of the left waveguide are called the $\psi$-modes while the modes on the right are called the $\phi$-modes. As illustrated in the figure, there are $M$ bounded $\psi$-modes and $N$ bounded $\phi$-modes. The $\psi$-modes are characterized by propagation constants $\gamma$, while the $\phi$-modes are characterized by propagation constants $\tilde{\gamma}$. (Although Fig. 6.1 shows the dielectric constant of the various layers as being continuous across the discontinuity, the $\psi$-modes and $\phi$-modes may have different layer parameters or even different layer thicknesses.)

As illustrated in the figure, the input waveguide, on the left, has a single bounded mode (the fundamental $\psi$-mode, Case I) incident on the discontinuity which produces an array of mode generation, Case I–IV modes that are reflected from the discontinuity and Case I–IV modes that are transmitted into the second waveguide region. This process is similar to that of an $M_i + M_f + 6$ port network where $M_i$ ports pertain to the bounded $\psi$-modes, $M_f$ ports pertain to the bounded $\phi$-modes and the remaining ports pertain the the Case II, III, IV radiation modes. In reality, it is an infinite port network, because the Case II, III and IV modes have continuous spectra.

The major issue here is to determine how the power from the input mode is distributed across the the array of scattered modes.

### 6.3 Scattering Formulation

The problem of determining the distribution of power in the scattered modes has been extensively discussed in the literature. For example, [4] has used impedance and variational arguments to find the scattering coefficients in closed waveguide system, which have sets of modes with discrete spectra. A slightly different approach, [6] focuses on minimizing the difference in the electric and magnetic fields on either side of the discontinuity. This latter approach can produce exact results when the calculation uses all of the modes of the left and right structures. However, in practical situations, the number of modes used in the calculation is always finite.

For the sake of mathematical simplicity, assume all layers are reactive and the electric field distribution (using the electric field $E_L = E_y(x, z < 0)$) in the left-hand waveguide is given by

$$E_L(x, z) = \psi_i(h_0, x)e^{-j\beta_0 z} + \sum_m \rho_i(h_m)\psi_i(h_m, x)e^{j\beta_m z}$$

$$+ \int_{II} \rho_{ii}(h)\psi_{ii}(h, x)e^{j\beta(h) z} dh + \int_{III} \rho_{iii}(h)\psi_{iii}(h, x)e^{j\beta(h) z} dh$$

$$+ \int_{IV} \rho_{iv}(h)\psi_{iv}(h, x)e^{\alpha(h) z} dh,$$

where $\beta(h) = \sqrt{k_0\kappa_s - h^2}$, $\kappa_s$ being the dielectric constant of the substrate, and $\alpha(h) = \sqrt{h^2 - k_0\kappa_s}$. The electric field in the right-hand-waveguide is given by

$$E_R(x, z) = \sum_n \tau_i(h_n)\phi_i(h_n, x)e^{-j\tilde{\beta}_n z} + \int_{II} \tau_{ii}(h)\phi_{ii}(h, x)e^{-j\tilde{\beta}(h) z} dh$$

$$+ \int_{III} \tau_{iii}(h)\phi_{iii}(h, x)e^{-j\tilde{\beta}(h) z} dh + \int_{IV} \tau_{iv}(h)\phi_{iv}(h, x)e^{-\tilde{\alpha}(h) z} dh.$$
Although the subscripts for the continuous versions of $\rho$ and $\tau$ are unnecessary because the value of $h$ uniquely identifies the case, they are included here for convenience. Even though modes for the left and right waveguides are different, it is interesting to note that waveguide structures with identical sub- and superstrate materials, have identical integration limits (for $h$ and $\tilde{h}$) for all continuous modes.

The corresponding transverse component of the magnetic fields, $H_x$, satisfy

$$k_o Z_o H_L(x, z) = -\beta_0 \psi_i(h_0, x) e^{-j\beta_0 z} + \sum_m \beta_m \rho_i(h_m) \psi_i(h_m, x) e^{j\beta_m z}$$

$$+ \int_{II} \beta(h) \rho_{ii}(h) \psi_{ii}(h, x) e^{j\beta(h) z} \, dh + \int_{III} \beta(h) \rho_{iii}(h) \psi_{iii}(h, x) e^{j\beta(h) z} \, dh$$

$$- j \int_{IV} \alpha(h) \rho_{iv}(h) \psi_{iv}(h, x) e^{\alpha(h) z} \, dh,$$

and

$$k_o Z_o H_R(x, z) = -\sum_n \tilde{\beta}_n \tau_i(h_n, x) e^{-j\tilde{\beta}_n z} - \int_{II} \tilde{\alpha}(\tilde{h}) \tau_{ii}(\tilde{h}) \phi_{ii}(\tilde{h}, x) e^{-j\tilde{\alpha}(\tilde{h}) z} \, d\tilde{h}$$

$$- \int_{III} \tilde{\alpha}(\tilde{h}) \tau_{iii}(\tilde{h}) \phi_{iii}(\tilde{h}, x) e^{-j\tilde{\alpha}(\tilde{h}) z} \, d\tilde{h} + j \int_{IV} \tilde{\alpha}(\tilde{h}) \tau_{iv}(\tilde{h}) \phi_{iv}(\tilde{h}, x) e^{-\tilde{\alpha}(\tilde{h}) z} \, d\tilde{h}.$$ (6.3.3)

The evanescent fields of Case IV attenuate in both the positive and negative $z$ directions while the remaining cases are composed of modes that propagates power away from the discontinuity. Although the evanescent fields store energy near the discontinuity they are not involved in the distribution of power, from the incident mode, that carries the total power away from the junction. Using the results of Section 4.7, the total power propagating from the junction can be computed.

### 6.4 Single-Mode Approximation

For proper matching of the electromagnetic fields at $z = 0$, fields at $z = 0^-$ must match those at $z = 0^+$. There are several approaches. The impedance/admittance variational principle can be applied, however, with the advances in computational power, it is useful to consider the field matching technique. For example, a collocation technique uses a set of points along the $x$ axis and matches the electric, magnetic or both on either side of the discontinuity. The number of collocation points determines the number of unknown quantities which are composed of the reflection and transmission coefficients. Although there is a finite number of parameters for the bounded modes, there is an infinite number for the continuous modes. By choosing to approximate the integrals in the Case II, III and IV modes, the number of unknowns can be finite. Another approach is to minimize the mean-square difference between the left and right fields at the discontinuity, and if that difference is zero, the fields will be perfectly matched. This latter approach will be developed in this section.
Before developing a solution for all reflection and transmission coefficients, a simple case is addressed first. Assume the simple case of only one propagating mode in both waveguides, and as a first approximation, all radiation modes will be ignored. The integral for the mean-square difference is

\[ J_e(\rho, \tau) = k_0 \int_{-\infty}^{\infty} [E_L(x, 0) - E_R(x, 0)]^2 \, dx \]

\[ = k_0 \int_{-\infty}^{\infty} [(1 + \rho)\psi_i(h_0, x) - \tau \phi_i(\tilde{h}_0, x)]^2 \, dx \]

\[ = 1 + 2\rho + \rho^2 - 2u_{00}\tau - 2u_{00}\rho\tau + \tau^2, \quad (6.4.1) \]

where the overlap integral \( u_{00} \) is

\[ u_{00} = k_0 \int_{-\infty}^{\infty} \psi_i(h_0, x) \phi_i(\tilde{h}_0, x) \, dx. \]

The mean-square difference can be written as

\[ J_e(x) = x^T \mathcal{H}_e^e \, x + 1, \quad (6.4.2) \]

where

\[ x = \begin{pmatrix} \rho \\ \tau \end{pmatrix}, \quad \mathcal{H}_e^e = \begin{pmatrix} 1 & -u_{00} \\ -u_{00} & 1 \end{pmatrix}, \quad \mathbf{b}_e = -2 \begin{pmatrix} -1 \\ u_{00} \end{pmatrix}. \]

The minimum of \( J_e \) with respect to \( x \) can be obtained by taking the variation of \( J_e \), given by

\[ \delta J_e = (\mathbf{b}_e^T + 2x^T \mathcal{H}_e^e)\delta x = 0, \]

which yields a minimum at \( x = -(\mathcal{H}_e^e)^{-1}\mathbf{b}_e^2/2 \). When the overlap integral is unity, \( \mathcal{H}_e^e \) is singular so that there is no unique solution.

Next consider matching the magnetic field at the boundary using

\[ J_m(x) = k_0 \int_{-\infty}^{\infty} [Z_0 H_L(x, 0) - Z_0 H_R(x, 0)]^2 \, dx \]

\[ = k_0 \int_{-\infty}^{\infty} [\beta\rho(1 - \rho)\psi_i(h_0, x) - \tilde{\beta}_0 \tau \phi_i(\tilde{h}, x)]^2 \, dx. \quad (6.4.3) \]

The value of \( J_m \) could be minimized with respect to \( x \) and may yield a different solution so that a decision on which is best must be debated. Another possibility is to minimize \( J_e \) with a constraint on \( J_m \) or vice-versa. The approach used here is to make the objective function a linear combination of \( J_e \) and \( J_m \), but weighting the importance of one relative to the other must be addressed. Note that the coefficients of \( \rho^2 \) and \( \tau^2 \) in (6.4.1) are unity, but the coefficients in (6.4.3) are \( \beta_0^2 \) and \( \tilde{\beta}_0^2 \). Thus, to put "equal" footing for the electric and magnetic fields, \( J_m(x) \) will be defined as

\[ J_m(x) = k_0 \frac{\beta_0^2}{\rho_0^2} \int_{-\infty}^{\infty} [Z_0 H_L(x, 0) - Z_0 H_R(x, 0)]^2 \, dx \]

\[ = k_0 \int_{-\infty}^{\infty} [(1 - \rho)\psi_i(h_0, x) - \frac{\tilde{\beta}_0}{\beta_0} \tau \phi_i(\tilde{h}, x)]^2 \, dx. \quad (6.4.4) \]
The mean-square difference becomes
\[ J_m(x) = x^T H^m x + b^m T x + 1, \] (6.4.5)
where
\[ H^m = \begin{pmatrix} 1 & \frac{\bar{n}_{ef0}}{n_{ef0}} u_{00} \\ \frac{\bar{n}_{ef0}}{n_{ef0}} u_{00} & \frac{\bar{n}_{ef0}^2}{n_{ef0}^2} \end{pmatrix}, \quad b^m = -2 \begin{pmatrix} 1 \\ \frac{\bar{n}_{ef0}}{n_{ef0}} u_{00} \end{pmatrix}. \]

The effective indices of the left and right waveguide modes are \( n_{ef0} = \beta_0/k_0 \) and \( \bar{n}_{ef0} = \bar{\beta}_0/k_0 \), respectively.

The variation of \( J \) gives the optimum solution as
\[ x = -\frac{1}{2} H^{-1} b, \] (6.4.6)
where
\[ H = H^e + H^m = \begin{pmatrix} 2 & -(1 - \frac{\bar{n}_{ef0}}{n_{ef0}}) u_{00} \\ -(1 - \frac{\bar{n}_{ef0}}{n_{ef0}}) u_{00} & 2 \end{pmatrix}. \] (6.4.7)
and
\[ b = b^e + b^m = -2 \begin{pmatrix} 0 \\ (1 + \frac{\bar{n}_{ef0}}{n_{ef0}}) u_{00} \end{pmatrix}. \] (6.4.8)

The optimum value of \( \rho \) is
\[ \rho = \frac{(n_{ef0} - \bar{n}_{ef0})(n_{ef0} + \bar{n}_{ef0}) u_{00}^2}{(2 - u_{00}^2)(n_{ef0}^2 + \bar{n}_{ef0}^2) + 2 n_{ef0} \bar{n}_{ef0} u_{00}^2}, \] (6.4.9)
while the optimum value of \( \tau \) becomes
\[ \tau = \frac{2 n_{ef0} (n_{ef0} + \bar{n}_{ef0}) u_{00}}{(2 - u_{00}^2)(n_{ef0}^2 + \bar{n}_{ef0}^2) + 2 n_{ef0} \bar{n}_{ef0} u_{00}^2}. \] (6.4.10)

The quality of the approximation correlates with the value of \( J(x) \) which is given by
\[ J(x) = 2 - \frac{1}{4} b^T H^{-1} b. \] (6.4.11)
For the single-mode approximation whose solution is given in (6.4.9) and (6.4.10), \( J(x) \) becomes
\[ J(x) = \frac{4 (1 - u_{00}^2)(n_{ef0}^2 + \bar{n}_{ef0}^2)}{(2 - u_{00}^2)(n_{ef0}^2 + \bar{n}_{ef0}^2) + 2 n_{ef0} \bar{n}_{ef0} u_{00}^2}. \] (6.4.12)
As can be seen, the quality of the approximation is directly related to the overlap integral and when \( u_{00} = 1 \), the approximation is exact! In fact, when \( u_{00} = 1 \) (6.4.9) and (6.4.10) give the reflection and transmission coefficients of a plane wave that is perpendicularly incident on a plane interface between two dielectric media.

When the overlap integral of the two fundamental modes, \( u_{00} \) decreases, the single-mode approximation starts to degrade and in fact the degradation is proportional to \( u_{00}^2 \) so that a 90 percent overlap extrapolates to \( 1 - u_{00}^2 \approx 0.2 \). As the optimum value of \( J(x) \) in the simple two-mode approximation increases, there is implied scattering into other modes and if the left and right structures are indeed single-mode waveguides, then the generation of radiation modes is implied. Finally, in the event that \( n_{ef} \approx n_{ef}^0 \), the field transmission coefficient \( \tau \approx u_{00} \) and power transmission across the discontinuity is proportional to \( u_{00}^2 \).

### 6.5 Generalized Cost Function

In the previous section the cost function was minimized with respect to the reflection and transmission coefficients of only the fundamental modes in the left-hand and right-hand waveguides. This process will usually reflect a minimum cost, however, it is the goal of this section to drive the cost function to 0, and this can only be accomplished by including all the proper modes of both waveguides.

Without loss of generality or usefulness, the fundamental \( \Psi \)-mode will be directed toward the discontinuity at \( z = 0 \). The portion of the cost function for the electric field \( J_e(x) \), as given in (6.4.1), will be expanded to include all bounded and radiation modes using the electric fields given in (6.3.1) and (6.3.2), while the portion of the cost function \( J_m(x) \), as given in (6.4.4), will use the magnetic fields given in (6.3.3) and (6.3.4). These cost functions will be considered as generalized because they include all possible mode expansions. Because the set of modes is complete, it is possible to drive the generalized cost function to zero.

The \( \mathbf{x} \) vector now contains all reflection and transmission coefficients, including the those of the Case II, III and IV radiation modes where

\[
\mathbf{x} = \begin{pmatrix} \rho_l \\ \rho_{ii} \\ \rho_{iii} \\ \rho_{iv} \\ \tau_i \\ \tau_{ii} \\ \tau_{iii} \\ \tau_{iv} \end{pmatrix}.
\]

(6.5.1)

While \( \rho_i \) and \( \tau_i \) represent the reflection and transmission coefficients for the discrete bounded modes, the remaining terms are continuous functions of the
independent variables \( h \) and \( \tilde{h} \), the transverse wave numbers of the substrate regions. For example, the block vectors \( \mathbf{\rho}_i \) and \( \mathbf{\tau}_i \) are

\[
\mathbf{\rho}_i = \begin{pmatrix}
\rho_i(h_0) \\
\rho_i(h_1) \\
\vdots \\
\rho_i(h_N)
\end{pmatrix}, \quad \mathbf{\tau}_i = \begin{pmatrix}
\tau_i(\tilde{h}_0) \\
\tau_i(\tilde{h}_1) \\
\vdots \\
\tau_i(\tilde{h}_N)
\end{pmatrix}.
\]

The symmetric matrix \( \mathcal{H}^e \) can be written as a block matrix

\[
\mathcal{H}^e = \begin{pmatrix}
\mathcal{H}^e_{11} & \mathcal{H}^e_{12} \\
\mathcal{H}^e_{21} & \mathcal{H}^e_{22}
\end{pmatrix}, \quad \text{(6.5.2)}
\]

where the block matrices are composed of smaller block matrices. The two diagonal components are given by

\[
\mathcal{H}^e_{11} = \begin{pmatrix}
I_{i,i} & 0_{i,i} & 0_{i,i} & 0_{i,v} \\
0_{T,i} & I_{i,i} & 0_{i,i} & 0_{i,v} \\
0_{T,i} & 0_{T,i} & I_{i,i} & 0_{i,i} & 0_{i,v} \\
0_{T,i} & 0_{T,i} & 0_{T,i} & I_{i,v} & 0_{i,v}
\end{pmatrix}, \quad \text{(6.5.3)}
\]

and

\[
\mathcal{H}^e_{22} = \begin{pmatrix}
\tilde{I}_{i,i} & \tilde{0}_{i,i} & \tilde{0}_{i,i} & \tilde{0}_{i,v} \\
\tilde{0}_{T,i} & \tilde{I}_{i,i} & \tilde{0}_{i,i} & \tilde{0}_{i,v} \\
\tilde{0}_{T,i} & \tilde{0}_{T,i} & \tilde{I}_{i,i} & \tilde{0}_{i,i} & \tilde{0}_{i,v} \\
\tilde{0}_{T,i} & \tilde{0}_{T,i} & \tilde{0}_{T,i} & \tilde{I}_{i,v} & \tilde{0}_{i,v}
\end{pmatrix}. \quad \text{(6.5.4)}
\]

The elements of \( \mathcal{H}^e_{11} \) are tied to the \( \psi \) modes while the elements of \( \mathcal{H}^e_{22} \) are tied to the \( \phi \) modes. The block matrices for the continuous modes are a function of the substrate wave numbers \( h \) and \( \tilde{h} \). The elements \( 0_{-} \) and \( \tilde{0}_{-} \) are zero “block” matrices while the \( I_{-} \) and \( \tilde{I}_{-} \) are identity matrices. The matrices \( 0_{-} \) and \( I_{-} \) are distinguished from the \( 0_{-} \) and \( I_{-} \) to allow for the fact that the \( \psi \)-modes may be different in number from the \( \phi \)-modes. The block matrices \( \mathcal{H}^e_{11} \) and \( \mathcal{H}^e_{22} \) are unit matrices and their block elements are unit matrices because the different cases are orthogonal and the modes within the cases are orthogonal. The off-diagonal elements, containing the overlap integrals, are

\[
\mathcal{H}^e_{21} = \mathcal{H}^e_{12} = \begin{pmatrix}
-U_{i,i} & -U_{i,i} & -U_{i,i} & -U_{i,v} \\
-U_{i,i} & -U_{i,i} & -U_{i,i} & -U_{i,v} \\
-U_{i,i} & -U_{i,i} & -U_{i,i} & -U_{i,v} \\
-U_{i,v} & -U_{i,v} & -U_{i,v} & -U_{i,v}
\end{pmatrix}. \quad \text{(6.5.5)}
\]

The elements \( U_{i,i} \) represent the overlap integrals of the discrete modes and has a dimension \( M_i \times M_i, \( M_i \) being the number of bound \( \psi \) modes while \( M_i \) is the number of bound \( \phi \) modes. It should be noted that since the dielectric constants of the substrate and superstrates are identical for both waveguides, the range of \( h \) and \( \tilde{h} \) are identical for cases II, III and IV.
The element $m,n$ of $U_{i,i}$ is

$$u_{i,i}(m,n) = \int_{-\infty}^{\infty} \psi_i(h_m,x) \phi_i(\tilde{h}_m,x) \, dx \equiv \langle \psi_i(h_m), \phi_i(\tilde{h}_n) \rangle,$$

whereas the elements of $U_{i,i}$ are $\langle \psi_i(h_m), \phi_i(\tilde{h}_n) \rangle$.

The generalized $b_e$ vector is also partitioned as follows

$$b_e = \begin{pmatrix} b_{e_p}^r \\ b_{e_t} \end{pmatrix}. \quad (6.5.6)$$

where the block vectors $b_{e_p}^r$ and $b_{e_t}$ are further partitioned according to the I, II, III, and IV modes

$$b_{e_p}^r = \begin{pmatrix} b_{e_p_{\rho_1}}^r \\ b_{e_p_{\rho_1}}^r \\ b_{e_p_{\rho_1i}}^r \\ b_{e_p_{\rho_1ii}}^r \\ b_{e_p_{\rho_1iv}}^r \end{pmatrix}, \quad b_{e_t} = \begin{pmatrix} b_{e_{\tau_1}}^r \\ b_{e_{\tau_1}}^r \\ b_{e_{\tau_1i}}^r \\ b_{e_{\tau_1ii}}^r \\ b_{e_{\tau_1iv}}^r \end{pmatrix}.$$

Although the generalized vector $b_e$ has a form that depends on the mode incident on the discontinuity, the fundamental $\psi$-mode, $\psi_i(h_0,x)$, will be used here to excite the mode conversion. This assumption affects only the $b_{e_{\rho_i}}^r$ block vector

$$b_{e_{\rho_i}}^r = 2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (6.5.7)$$

(The location of the non-zero value is determined by the mode incident on the discontinuity. For single bounded-mode waveguides, $b_{e_{\rho_i}}^r$ has only 1 element.)

The remaining block vectors, $b_{e_{\rho_1i}}^r$, $b_{e_{\rho_1ii}}^r$, and $b_{e_{\rho_1iv}}^r$, have zero elements due to the fact that the modes are orthogonal. Thus, $b_{e_{\rho_i}}^r$ has only a single non-zero element.

The elements of $b_{e_{\rho_i}}^r$ contain the overlap integrals of the fundamental $\psi$ mode with the $\phi$ modes of the various cases. Namely,

$$b_{e_{\tau_1}}^r = -2 \begin{pmatrix} u_{i,1}(0,0) \\ u_{i,1}(0,1) \\ \vdots \\ u_{i,1}(0,\tilde{M}_i - 1) \end{pmatrix}, \quad (6.5.8)$$

while the remaining vector elements are

$$b_{e_{\tau_1i}}^r = -2(u_{i,1i}(0,\tilde{h})), \quad (6.5.9)$$

$$b_{e_{\tau_1ii}}^r = -2(u_{i,1ii}(0,\tilde{h})), \quad (6.5.10)$$

$$b_{e_{\tau_1iv}}^r = -2(u_{i,1iv}(0,\tilde{h})). \quad (6.5.11)$$
Finally, the cost function representing the discontinuity of the electric field across the boundary between the $\psi$ and $\phi$ modes can be written in terms of the block matrices and vectors as

$$J_e = 1 + b_b^T \rho + b_e^T \tau + \rho^T H_{11}^e \rho + \rho^T H_{12}^e \tau + \tau^T H_{21}^e \rho + \tau^T H_{22}^e \tau. \quad (6.5.12)$$

This quadratic function contains both series of terms over the bounded modes and integral terms over the continuous modes. For example, the quadratic expression containing the “block matrix” $U_{ii,ij}$ is the double integral

$$\int_{\tilde{h}} \int_{\tilde{h}} \rho_{ii}(h) \langle \psi_i(h), \phi_i(\tilde{h}) \rangle \tau_{ii}(\tilde{h}) \, dh \, d\tilde{h}. \quad (6.5.13)$$

If the integrals are approximated, for example, using a Gauss quadrature method, then the integral expressions will indeed reduce to a proper matrix/vector formulation. However, simplifications of various integral expressions in the quadratic form, (6.5.12), can be made prior to performing numerical calculations. The overlap integrals, given in Appendix B, develops the appropriate simplifications of the integrals used in (6.5.12).

Even though the cost function $J_m$ for the magnetic field as given in (6.4.3), was developed only for the bound modes, the generalized cost function for the magnetic field will be expressed in a slightly different form from that used in the earlier discussion. Note that the electric field $E_L(x, 0^-)$, (6.3.3), and $E_R(x, 0^+)$, (6.3.4), are real functions, provided the reflection and transmission coefficients are real. This implies that the phase of the electric field is either 0 or $\pi$ radians. In fact, the field components for each of the mode cases, have either the 0 or $\pi$ phase condition. However, the magnetic fields as written in (6.3.3) and (6.3.4), have the Case IV modes out of phase by $\pi/2$ radians. The “in phase” (with the electric field) components are

$$k_o Z_o H_L^I(x, 0^-) = -\beta_0 \psi_i(\theta_0, x) + \sum_m \beta_m \rho_i(\theta_m) \psi_i(\theta_m, x) \quad (6.5.14)$$

and

$$k_o Z_o H_R^I(x, z) = -\sum_n \beta_n \tau_i(\theta_n) \phi_i(\theta_n, x) - \int_{\tilde{\theta}} \tilde{\beta}(\tilde{\theta}) \tau_{ii}(\tilde{\theta}) \phi_i(\theta, x) \, d\tilde{\theta} \quad (6.5.15)$$

while the quadrature terms are

$$k_o Z_o H_L^Q(x, 0^-) = \int_{IV} \alpha(h) \rho_i(h) \psi_i(h, x) \, dh, \quad (6.5.16)$$

and

$$k_o Z_o H_R^Q(x, 0^+) = \int_{IV} \tilde{\alpha}(\tilde{h}) \tau_{ii}(\tilde{h}) \phi_i(\theta, x) \, d\tilde{h}. \quad (6.5.17)$$
The cost function for the magnetic fields will now be written as

\[ J_m(x) = \frac{k_o Z_o^2}{\beta_0^2} \int_{-\infty}^{\infty} \left\{ |H_L^I(x,0^-) - H_R^I(x,0^+)|^2 + |H_L^Q(x,0^-) - H_R^Q(x,0^+)|^2 \right\} dx \]  
(6.5.18)

Although it appears the parameters \( \rho_{1v} \) and \( \tau_{1v} \) could have undue weighting in the quadrature component compared to the “in phase” component as expressed in (6.5.18), \( \rho_{1v} \) and \( \tau_{1v} \) also appear in \( J_m(x) \) which acts to balance their overall weight compared to \( \rho_i, \rho_{ii}, \rho_{iii}, \tau_i, \tau_{ii}, \) and \( \tau_{iii} \). By treating the “in phase” and quadrature components in the said manner, the electric and magnetic fields will be matched in time. An alternative approach is to average absolute square value of the field difference between the left-hand and right-hand waveguide. This latter case would amount to minimizing the “time average” difference.

The cost function for the magnetic field is obtained from (6.5.18) by expanding the magnetic fields in terms of their representation using the \( \psi \) and \( \phi \) modes. As illustrated in (6.5.14) and (6.5.15), the terms \( \hat{\beta}_m/\beta_0, \hat{\beta}(\hat{h})/\beta_0, \hat{\beta}_m/\beta_0, \hat{\beta}(\hat{h})/\beta_0, \) etc., will appear in each term involving \( \rho \) and \( \tau \). The ratio \( \beta_m/\beta_0 = n_{efm}/n_{ef0} \) can be associated with the \( \rho \) coefficients while \( \hat{\beta}_m/\beta_0 = \tilde{n}_{efm}/n_{ef0} \) can be associated with the \( \tau \) coefficients. Thus, the diagonal matrix \( N_{ef} \) will be composed of the two diagonal block matrices \( N_{\rho} \) and \( N_{\tau} \) as

\[ N_{ef} = \text{diag}(N_{\rho} \quad N_{\tau}). \]  
(6.5.19)

where

\[ N_{\rho} = \text{diag}(1 \quad n_{ef1}/n_{ef0} \quad \cdots), \quad N_{\tau} = \text{diag}(\tilde{n}_{ef0}/n_{ef0} \quad \tilde{n}_{ef1}/n_{ef0} \quad \cdots). \]

The symmetric matrix \( \mathcal{H}^m \) can be written as a block matrix

\[ \mathcal{H}^m = \begin{pmatrix} \mathcal{H}_{11}^m & \mathcal{H}_{12}^m \\ \mathcal{H}_{21}^e & \mathcal{H}_{22}^e \end{pmatrix}, \]  
(6.5.20)

where \( \mathcal{H}_{11}^m = \mathcal{H}_{11}^e \) and \( \mathcal{H}_{22}^m = \mathcal{H}_{22}^e \). The off-diagonal terms are

\[ \mathcal{H}_{21}^{mT} = \mathcal{H}_{12}^m = \begin{pmatrix} U_{i,i} & U_{i,ii} & U_{i,iii} & \tilde{0}_{i,jv} \\ U_{ii,i} & U_{ii,ii} & U_{ii,iii} & \tilde{0}_{ii,jv} \\ U_{iii,i} & U_{iii,ii} & U_{iii,iii} & \tilde{0}_{iii,jv} \\ \tilde{0}_{i,v,j} & \tilde{0}_{i,v,ii} & \tilde{0}_{i,v,iii} & U_{i,v,jv} \end{pmatrix}. \]  
(6.5.21)

The block matrix \( \tilde{0}_{i,jv} \) has the same dimensions as \( U_{i,v,j} \) as given in (6.5.5).

The generalized \( \mathbf{b}^m \) vector will be specified without using the elements of \( N_{ef} \) as written earlier in (6.4.5) for the simple single mode. Instead, the \( \mathbf{b}^m \) vector will be written similar to that for \( \mathbf{b}^e \), so that

\[ \mathbf{b}^m_\rho = -2 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{b}^m_\tau = \mathbf{b}^m_\tau \]  
(6.5.22)
Now the generalized cost function for the magnetic field can be written as

\[ J_m(x) = 1 + b^m N_{ef} x + x^T N_{ef} \mathcal{H}_N^m N_{ef} x. \]  \hspace{1cm} (6.5.23)

and the overall cost function is

\[ J(x) = J_e(x) + J_m(x) = 2 + (b^e + N_{ef} b^m)^T x + x^T (\mathcal{H}_N^e + N_{ef} \mathcal{H}_N^m N_{ef}) x, \] \hspace{1cm} (6.5.24)

which has the optimum solution

\[ x = -\frac{1}{2} (\mathcal{H}_N^e + N_{ef} \mathcal{H}_N^m N_{ef})^{-1} (b^e + N_{ef} b^m). \] \hspace{1cm} (6.5.25)
Chapter 7
Modes in Complex Structures

Complex structures are formed from simple structures such as layers with constants indices of refraction or from a linearly graded dielectric constant. Waves in layered dielectric media have been investigated for over 100 years and have many applications in integrated optics. Dielectric wave guides can be easily tailored for a specific profile with resulting modes that can have synthesized shapes. Later in this chapter, the concept of a super layer will be introduced. Specifically, a super layer is a group of layers that can be characterized as a single layer that has a “non-simple” dispersion relation.

7.1 Introduction

In this section, the properties of the transfer matrix will be determined relative to the characteristics of layered material that forms the dielectric waveguide. In general optoelectronic devices that are common include optical waveguides, detectors and the optical sources. The fundamental structures are grown from binary, ternary and quaternary materials by some type of epitaxial growth. Generally, the devices are grown as layers of different materials. However, in a single layer, the material is of one type so that its properties such as the dielectric constant is position independent in the layer. If the layer in question is lossy (or has gain in the case of laser active region), then the dielectric constant is complex. On the other hand, if a layer is grown with a “continuously changing” material, then the dielectric constant varies with position. Obviously, this picture is only applicable in a macroscopic sense because the layers themselves are grown with layers of atoms which forms a complex structure.
For transverse electric waves where the electric field is polarized in the $y$ direction, the wave equation is written as

## 7.2 Fundamental Layers

Assume for example, that a layered structure has a dielectric profile as shown in Fig. 7.1.

The geometry can be characterized according to the following list:

1. The total number of layers is $N$.
2. The first layer is labeled as Layer 0 while the last layer is labeled as Layer $N - 1$.
3. The 0th layer starts at $x = x_0$ and extends to $x = \infty$ while the last layer starts at $x = -\infty$ and extends to $x = x_{N-2}$.
4. The $l$th layer starts at $x = x_l$ and extends to $x = x_{l+1}$.
5. A fundamental layer is characterized as “constant” or linearly “graded”.

For example, Layer 1 is graded while Layer 0 is constant.

A specific layer is then defined by its dielectric constant and it thickness, $d_l = x_{l+1} - x_l$.

### 7.2.1 Homogeneous Layers

The dielectric profiles of a constant layer and a graded layer are shown in Fig. 7.2. The constant layer is characterized by a single value while the graded layer is determined from the two values at the beginning and ending of a layer.

The field vector at the end of the layer is determined from its value at the beginning point. The transformation across the $l$th layer is given by

$$\Phi_l(d_l) = T_l \cdot \Phi_l(0) = \Phi_{l-1}(0), \quad (7.2.1)$$
where the transformation matrix $T_l$ is

$$T_l(d_l) = \begin{pmatrix} A_l & B_l \\ C_l & D_l \end{pmatrix}$$

The layer thickness $d_l = x_{l-1} - x_l$ is determined from the layer intersections or vice versa.

The elements of the transfer matrix given in the table below, have characteristics discussed in Chapter 4. The transfer matrix transforms the field vector at that beginning of a layer to the field vector at the end of the layer. It should be noted that since the field vectors are continuous across the layer interfaces, the field vector at the interfaces can simply be denoted as $\Phi_l$ (without an argument). Thus, the layer transfer matrix transforms interface-to-interface.

<table>
<thead>
<tr>
<th>Homogeneous Layer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_l$</td>
<td>$\cos h_1 d_l$</td>
</tr>
<tr>
<td>$B_l$</td>
<td>$(\sin h_1 d_l)/h_1$</td>
</tr>
<tr>
<td>$C_l$</td>
<td>$-h_1 \sin h_1 d_l$</td>
</tr>
<tr>
<td>$D_l$</td>
<td>$\cos h_1 d_l$</td>
</tr>
</tbody>
</table>

Table 7.1. The transfer matrix elements for a single, homogeneous dielectric layer of thickness $d_l$ and transverse wavenumber $h_1$.

The derivative of the transfer matrix elements shown in Table 7.1 with respect to $h_1$ is used to find the modes from the dispersion equation when applying Newton’s method. Thus it is useful to calculate the derivative of the transfer matrix elements with respect to $h_1$ as listed in Table 7.2.

<table>
<thead>
<tr>
<th>Derivative: Homogeneous Layer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$dA_l/dh_1$</td>
<td>$-d_l \sin h_1 d_l$</td>
</tr>
<tr>
<td>$dB_l/dh_1$</td>
<td>$(h_1 d_l \cos h_1 d_l - \sin h_1 d_l)/h_1^2$</td>
</tr>
<tr>
<td>$dC_l/dh_1$</td>
<td>$-h_1 d_l \cos h_1 d_l - \sin h_1 d_l$</td>
</tr>
<tr>
<td>$dD_l/dh_1$</td>
<td>$-d_l \sin h_1 d_l$</td>
</tr>
</tbody>
</table>

Table 7.2. The derivative of the transfer matrix elements with respect to $h_1$ for a single, homogeneous dielectric layer.

The two dielectric distributions as illustrated in Fig. 7.2 show the homogeneous and linearly graded profiles.
7.2.2 Graded Layers

In the previous section, the matrix that transforms interface fields from interface to interface, was concerned with only layers having a homogeneous dielectric constant. If the dielectric constant varies continuously across the layer, then homogeneous layer is inappropriate. However, in many cases the continuous layer can be efficiently modeled using the “stair step” profile. The actual number of layers used to partition the continuous profile must not be too large because numerical accuracy can be compromised.

In this section we will derive the transfer matrix for a structure with linear dielectric profile. In many instances, the dielectric profile may not be linear, however, a complex profile can be partitioned into a series of layers using linear profiles (similar to the “stair step” profile used to convert a continuous layer into a series of constant layers). The approximation of the dielectric profile using linear layers can be very accurate and the number of linear layers may be smaller than the number of “star step” layers required to describe the profile.

In many structures, the linear dielectric profile can be very close to a linear index profile. For example, assume in a specific layer that the refractive is linear, then the dielectric constant is

\[ \kappa(x) = (m_n x + b_n)^2, \]

\[ \approx 2m_n b_n x + b_n^2, \]

where the constants \( m_n \) and \( b_n \) characterize the line describing the index of refraction. (In (7.2.2), the term \( m_n^2 x^2 \) has been dropped because in most dielectric waveguide structures used for integrated optics and injection lasers, the “slope” \( m_n \) is much smaller than \( b_n \).)
The dielectric constant \( \kappa_l(x) \), can be characterized by three parameters, the layer thickness, \( d_l \), and the two values of the dielectric constant at the left side, \( \kappa_{Li} \), and at the right side, \( \kappa_{Rl} \). The linear expression
\[
\kappa_l(x) = m_l x + b_l, \tag{7.2.3}
\]
has \( m_l = (\kappa_{Rl} - \kappa_{Li})/d_l \) and \( b_l = \kappa_{Li} \).

The solution of the wave equation for the linear profile involves Airy functions which are relatively more complicated than the functions obtained from solution of the constant dielectric layer. Also, when the material in any of the layers becomes complex the Airy functions will have complex arguments. However, there is an extensive literature, and even computer programs, that compute Airy functions with both real and complex arguments. In fact the Airy functions can be obtained from Bessel functions of fractional order.

The wave equation for the linear profile is obtained by substituting (7.2.3) into (4.4.5). The resulting expression becomes for the \( l \)th layer
\[
\frac{d^2 \psi_l(x)}{dx^2} + [k_0^2 (m_l x + b_l) + \gamma^2] \psi_l(x) = 0 \tag{7.2.4}
\]
where the free-space wavenumber is \( k_o \). To put the above equation into an Airy type of differential equation, change from the dependent variable \( x \) to the variable \( \xi \), using the relation
\[
\xi = \frac{-k_0^2 (m_l x + b_l) + \gamma^2}{\Delta_l}
\]
where the constant \( \Delta_l \) is to be determined. By making substitutions into (7.2.4), and noting that the operator \( d/dx = (-k_0^2 m_l/\Delta_l) d/d\xi \), the new equation in terms of the independent variable \( \xi \) becomes
\[
k_0^4 |m_l|^2 \frac{d^2 \psi_l(\xi)}{d\xi^2} = \Delta_l^3 \xi \psi_l(\xi).
\]
Note that \( m_l \) has been replaced with \( |m_l| \) since its value has been squared in the equation. Below we will use the cube root of an expression involving \( m_l^2 \).

To get the proper Airy differential equation governing the fields in the \( l \)th layer, the value of \( \Delta_l \) must be chosen such that \( \Delta_l = (k_0^4 |m_l|^2)^{1/3} \), so that
\[
\xi = \frac{-\gamma/k_0 + b_l + (m_l/k_0)x}{(|m_l|/k_0)^{2/3}}. \tag{7.2.5}
\]
It is convenient to write the transformation \( x \to \xi \) in the form given above because when normalizing linear dimensions to the free-space wave number \( k_o \), \( \beta \) and \( m_l \) can be factored to their normalized form.

Finally, the differential equation for the field in the \( l \)th layer takes the Airy differential equation form,
\[
\frac{d^2 \psi_l(\xi)}{d\xi^2} = \xi \psi_l(\xi). \tag{7.2.6}
\]
Before discussing the vector form of the solution, it is useful to understand the behavior of the two independent solutions. In Fig. 7.3 the two independent solutions $A_i(\xi)$ and $B_i(\xi)$, which are solutions of (7.2.6) are shown as a function of the independent variable $\xi$. It is interesting to note that $A_i$ and $B_i$ have “exponential” behavior for $\xi > 0$ while they have “sinusoidal” behavior for $\xi < 0$.

In the constant layer model, the two independent solutions are the two trigonometric solutions that have either a functional value of zero at the origin, or derivative of zero, the Airy functions are both non-zero at the origin. The functions oscillate for negative arguments and have exponential behavior for positive arguments. It is now appropriate to transform the solution to the real space coordinate $x$.

The solution of the field in the $l$th layer in terms of the independent variable $x$ where $x$ is referenced to the value $x_l$ (Also, note that the dielectric constant is reference to $x_l$), is

$$\psi_l(x) = A A_i[\xi(x)] + B B_i[\xi(x)].$$  \hfill (7.2.7)

where $\xi(x)$ is given by the linear relation, (7.2.5). The vector solution in terms of the constants $A$ and $B$, is

$$\begin{pmatrix} \psi_l(x) \\ \psi'_l(x) \end{pmatrix} = \begin{pmatrix} A A_i[\xi(x)] & B B_i[\xi(x)] \\ \xi'_l A_i[\xi(x)] & \xi'_l B_i[\xi(x)] \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}$$ \hfill (7.2.8)

where the expression $\xi'_l$ is the derivative with respect to $x$,

$$\xi'_l = -\frac{m_l}{(|m_l|/k_o)^{2/3}}.$$

$$\begin{pmatrix} A_i(\xi_l) \\ B_i(\xi_l) \\ A'_i(\xi_l) \\ B'_i(\xi_l) \end{pmatrix}$$

Figure 7.3. The Airy functions and their derivatives.
while the expressions \( A_i'(\xi) \) and \( B_i'(\xi) \) are the derivatives with respect to their argument, \( \xi \). Note that the derivative \( \xi_i' \) is independent of position. (When normalizing \( m_l \) with respect to \( k_o \), in (7.2.9) the derivative \( \xi_i' \) should be normalized.)

Attention is now focussed on the calculation of the coefficients \( A_l \) and \( B_l \) in terms of the initial conditions of the field vector at \( x = 0 \), where the dielectric constant \( \kappa(0) = \kappa_{L1} \). Solving for coefficients in terms of the initial conditions gives

\[
\begin{pmatrix}
A_l \\
B_l
\end{pmatrix} = \begin{pmatrix}
\pi B_i'[\xi(0)] & -\pi B_i[\xi(0)]/\xi_i' \\
-\pi A_i'[\xi(0)] & \pi A_i[\xi(0)]/\xi_i'
\end{pmatrix} \cdot \begin{pmatrix}
\psi_l(0) \\
\psi_i(0)
\end{pmatrix}
\]

where we have used the Wronskian \( B_i[\xi] \) \( B_i'(\xi) - A_i'(\xi) B_i(\xi) = 1/\pi \).

Finally, the field vector transfer across a graded layer can be written in the form given by (7.2.1), where the elements of the transfer matrix are given in Table 7.3. The two parameters \( \xi_{0l} \) and \( \xi_{dl} \), which are arguments of the Airy function, in the table are

\[
\begin{align*}
\xi_{0l} &= -\frac{(\gamma/k_o)^2 + \kappa_{L1}}{(m_l/k_o)^{2/3}} \\
\xi_{dl} &= -\frac{(\gamma/k_o)^2 + \kappa_{R1}}{(m_l/k_o)^{2/3}}
\end{align*}
\]

(7.2.10)

<table>
<thead>
<tr>
<th>Graded Layer</th>
<th>( A_l )</th>
<th>( B_l )</th>
<th>( C_l )</th>
<th>( D_l )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_l )</td>
<td>( \pi [A_i(\xi_{dl})B_i'(\xi_{dl}) - A_i'(\xi_{dl})B_i(\xi_{dl})] )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( B_l )</td>
<td>-( \pi [A_i(\xi_{dl})B_i(\xi_{0l}) - A_i'(\xi_{0l})B_i(\xi_{dl})]/\xi_i' )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C_l )</td>
<td>( \pi \xi_i'[A_i'(\xi_{dl})B_i'(\xi_{0l}) - A_i'(\xi_{0l})B_i'(\xi_{dl})] )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( D_l )</td>
<td>-( \pi [A_i'(\xi_{dl})B_i(\xi_{0l}) - A_i(\xi_{0l})B_i'(\xi_{dl})] )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3. Matrix elements for a single, graded-dielectric value layer. The values \( \xi_{0l} \) and \( \xi_{dl} \), as given in (7.2.10), and \( \xi_i' \) is given by (7.2.9).

For the calculation of the derivative of the matrix elements with respect to \( \gamma \), the derivative of each element with respect to \( \xi \) will be tabulated and the chain rule can be used to find the element derivatives with respect to \( \gamma \). Namely, the derivative of \( A_l \) with respect to \( \gamma \) is

\[
\frac{dA_l}{d\gamma} = \frac{dA_l}{d\xi} \frac{d\xi}{d\gamma},
\]

(7.2.11)

where

\[
\frac{d\xi}{d\gamma} = -\frac{\gamma}{k_o^2(m_l/k_o)^{2/3}}.
\]

(7.2.12)

The derivatives of all elements of the transfer matrix across a single graded layer is shown in Table 7.4.

Table 7.4. Derivative of the matrix elements with respect to \( \xi \) for a single, graded-dielectric layer. The values \( \xi_{0l} = \xi(0) \), and \( \xi_{dl} = \xi(d_l) \), as given in (7.2.10), and \( \xi_i' \) is given by (7.2.9).
### 7.3 Superlayers

We are now in a position to define a super layer which will be used as the fundamental building blocks for a complex dielectric structure which will be used as an optical waveguide. A super layer is defined as a group of contiguous layers of the two fundamental layers: the constant dielectric layer and the graded dielectric layer. To solve for the modal characteristics in a dielectric waveguide with multiple layers, it can be advantageous to introduce the concept of the superlayer if the superlayer is reproduced many times. If layer geometries were not repeated, it would be superfluous to develop the algebra of superlayers. In practice, multiple superlayers are used for the design of reflection stacks that may be used in either the superstrate or substrate regions (the cladding regions) surrounding a dielectric waveguide. Also, stacks are generally used in the design of vertical cavity surface emitting lasers. Superlayer repetitions are also used for the design of multi quantum-well lasers. In the more general sense, the super layer is the building blocks for photonic crystals or for artificial dielectrics. Although a super layer can have any number of fundamental layers, attention here will be restricted to only two- and three-layer structures.

When the superlayer repeats an infinite number of times, such as in the two semi-infinite cladding regions, the group of superlayers behaves similar to that of a single layer.

As mentioned above, the super layers may be homogeneous or graded, however, for the sake of simplicity, the layers will be assumed to be homogeneous in the following discussion. Figure 7.3 shows both a two-layer cell and a three layer cell. The solution of the wave equation for each of the layers can be written in terms of the refractive indices and layer thicknesses. The width of the superlayer is $d_s$. 

---

<table>
<thead>
<tr>
<th>Derivative: Graded Layer</th>
<th></th>
</tr>
</thead>
</table>
| $d A_1/d \xi$ | \[ \pi \{ A'_i(\xi_d)B_i'(\xi_o) - A_i'(\xi_d)B_i(\xi_o) \]  
|  | \[ + \xi_o [A_i(\xi_d)B_i(\xi_o) - A_i(\xi_d)B_i'(\xi_o)] \} \]
| $d B_1/d \xi$ | \[ \pi \{ A'_i(\xi_o)B_i(\xi_d) - A_i'(\xi_d)B_i(\xi_o) \]  
|  | \[ + A_i(\xi_o)B_i(\xi_d) - A_i(\xi_o)B_i'(\xi_d)]/\xi_i \]
| $d C_1/d \xi$ | \[ \pi \xi_i [A_i(\xi_d)B_i'(\xi_o) - A_i'(\xi_d)B_i(\xi_o)] \]  
|  | \[ + \xi_o [A'_i(\xi_d)B_i(\xi_o) - A_i(\xi_d)B_i'(\xi_d)] \} \]
| $d D_1/d \xi$ | \[ \pi \xi_v [A_i(\xi_o)B_i(\xi_d) - A_i(\xi_d)B_i(\xi_o)] \]  
|  | \[ + [A'_i(\xi_o)B_i'(\xi_d) - A_i'(\xi_d)B_i'(\xi_o)] \]
7.3 SUPERLAYERS

Figure 7.4. The refractive index profiles of two fundamental building blocks. (a) A two-layer super structure and (b) a three-layer super structure. The width of the superlayer is \( d_s \) and is independent of the number of basic layers that form the superlayer.

In particular, the wave number for the \( i \)st layer is written as

\[
\beta_i^2 = \gamma^2 + k^2 n_i^2
\]

where \( \gamma \) is mode propagation constant while \( n_i^2 = \kappa_i \) is the layer dielectric constant of the first fundamental layer of the superlayer. Of course, if the superlayer contains a single material, the transfer matrix elements are given in Table 7.1 while for the graded layer, they are given in Table 7.3. Note that the “optical thickness” of a layer is \( h_i d_i \) and the secular equation producing modal eigenvalues is only a function of the optical thicknesses of internal layers.

7.3.1 Two-Region Superlayer

The two-region structure is illustrated in Fig. 7.4(a), and it includes the \( i \)st and \( ii \)nd layers (the \( l \)th and \( l + 1 \)th layers. The transfer-matrix elements, used to transfer the fields/fluxes across the superlayer, are listed in Table 7.5.

<table>
<thead>
<tr>
<th>Table 7.5. Matrix elements for a two layer structure. The layers are homogeneous and are shown in 'sf. 3'(a).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Homogeneous Layers</td>
</tr>
</tbody>
</table>
| \( \mathcal{A}_s \) | \[
\cos(h_i d_i) \cos(h_{ii} d_{ii}) - \frac{h_{ii}}{h_i} \sin(h_i d_i) \sin(h_{ii} d_{ii})
\] |
| \( \mathcal{B}_s \) | \[
\frac{1}{h_i} \cos(h_{ii} d_{ii}) \sin(h_i d_i) + \frac{1}{h_{ii}} \cos(h_i d_i) \sin(h_{ii} d_{ii})
\] |
| \( \mathcal{C}_s \) | \[
-h_i \cos(h_{ii} d_{ii}) \sin(h_i d_i) - h_{ii} \cos(h_i d_i) \sin(h_{ii} d_{ii})
\] |
| \( \mathcal{D}_s \) | \[
\cos(h_i d_i) \cos(h_{ii} d_{ii}) - \frac{h_{ii}}{h_i} \sin(h_i d_i) \sin(h_{ii} d_{ii})
\] |

The transfer matrix with elements given in Table 7.5 has a unit determinant \((\mathcal{A}_s \mathcal{D}_s - \mathcal{B}_s \mathcal{C}_s = 1)\) since the determinant of the fundamental is unity. The transfer of the field and flux in the negative direction (i.e. from \( x_{l-2} \) to \( x_{l} \)) can be obtained by replacing \( d_i \) with \(-d_i\) and \( d_{ii} \) with \(-d_{ii}\). Also, note the elements \( \mathcal{A}_i \) and \( \mathcal{D}_i \) are different.
7.3.2 Three-Region Superlayer

The three region superlayer will be the most general discussed. In fact it is easily seen that the two-layer structure can be configured as a three-layer one that would be symmetric. The low value of the refractive index of the structure shown in ‘sf. 3’ can be visualized as the “cladding” while the region with the large value of the refractive index can be “core” region. The three-region superlayer would have two cladding layers on either side of the core region.

Table 7.6. Matrix elements for a three layer structure. The layers are homogeneous and are shown in ‘sf. 3(b). To fit the expressions in the table, the variable names were changed; the thicknesses \(a, b, c\) are \(d_i, d_{ii}, d_{iii}\), and the wave numbers \(p, q, r\) are \(h_i, h_{ii}, h_{iii}\), respectively.

<table>
<thead>
<tr>
<th>Triple-Layer Elements</th>
<th>(4pq A_s)</th>
<th>((-pq - q^2 + pr - qr)) (\cos(pa - qb - rc)) + ((pq + q^2 - pr - qr)) (\cos(pa + qb - rc)) + ((pq - q^2 - pr + qr)) (\cos(pa - qb + rc)) + ((pq + q^2 + pr + qr)) (\cos(pa + qb + rc))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4pqr B_s)</td>
<td>((-pq + q^2 - pr + qr)) (\sin(pa - qb - rc)) + ((-pq - q^2 + pr + qr)) (\sin(pa + qb - rc)) + ((pq - q^2 - pr + qr)) (\sin(pa - qb + rc)) + ((pq + q^2 + pr + qr)) (\sin(pa + qb + rc))</td>
<td></td>
</tr>
<tr>
<td>(-4q C_s)</td>
<td>((pq - q^2 + pr - qr)) (\sin(pa - qb - rc)) + ((pq + q^2 - pr - qr)) (\sin(pa + qb - rc)) + ((pq - q^2 - pr + qr)) (\sin(pa - qb + rc)) + ((pq + q^2 + pr + qr)) (\sin(pa + qb + rc))</td>
<td></td>
</tr>
<tr>
<td>(4qr D_s)</td>
<td>((-pq + q^2 - pr + qr)) (\cos(pa - qb - rc)) + ((-pq - q^2 + pr + qr)) (\cos(pa + qb - rc)) + ((pq - q^2 - pr + qr)) (\cos(pa - qb + rc)) + ((pq + q^2 + pr + qr)) (\cos(pa + qb + rc))</td>
<td></td>
</tr>
</tbody>
</table>

The symmetric three-layer structure has interesting elements that are greatly simplified from those given in Table 7.6. In this configuration, the thicknesses and dielectric constants of the first and third layers are identical, i.e. \(d_i = d_{iii}\) and \(\kappa_i = \kappa_{iii}\).
Table 7.7. Matrix elements for a symmetric three-layer structure. The layers are homogeneous and are shown in 'sf.3'(b).

<table>
<thead>
<tr>
<th>Symmetric Three Layers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_s$</td>
<td>$\cos(2 \hat{h_i} d_i) \cos(h_{ii} d_{ii}) - \frac{\hat{h}<em>i^2 + \hat{h}</em>{ii}^2}{2 \hat{h}<em>i h</em>{ii}} \sin(2 \hat{h}<em>i d_i) \sin(h</em>{ii} d_{ii})$</td>
</tr>
<tr>
<td>$B_s$</td>
<td>$\frac{1}{\hat{h}<em>i} \sin(2 \hat{h}<em>i d_i) \cos(h</em>{ii} d</em>{ii}) - \frac{1}{2 \hat{h}<em>i^2 h</em>{ii}} [k_0^2 \Delta - (\hat{h}<em>i^2 + \hat{h}</em>{ii}^2) \cos(2 \hat{h}<em>i d_i)] \sin(h</em>{ii} d_{ii})$</td>
</tr>
<tr>
<td>$C_s$</td>
<td>$-\frac{1}{\hat{h}<em>i} \sin(2 \hat{h}<em>i d_i) \cos(h</em>{ii} d</em>{ii}) - \frac{1}{2 \hat{h}<em>i} [k_0^2 \Delta + (\hat{h}<em>i^2 + \hat{h}</em>{ii}^2) \cos(2 \hat{h}<em>i d_i)] \sin(h</em>{ii} d</em>{ii})$</td>
</tr>
<tr>
<td>$D_s$</td>
<td>$\cos(2 \hat{h}<em>i d_i) \cos(h</em>{ii} d_{ii}) - \frac{\hat{h}<em>i^2 + \hat{h}</em>{ii}^2}{2 \hat{h}<em>i h</em>{ii}} \sin(2 \hat{h}<em>i d_i) \sin(h</em>{ii} d_{ii})$</td>
</tr>
</tbody>
</table>

The transfer matrix elements for the symmetric superlayer has equal diagonal elements, viz, $A_s = D_s$. When the transverse propagation constants $\hat{h}_i$ and $h_{ii}$ are real numbers, then all matrix elements are also real. Assume for the moment that the dielectric constants $\kappa_i < \kappa_{ii}$ and the modal propagation constant $\gamma = j \beta$, i.e., the layers are formed with reactive material with no losses. Further, assume the the propagation constant lies between $k_0 n_i$ and $k_0 n_{ii}$ so that the transverse wave number $h_{ii} \equiv h'_{ii}$ is real while $h_i \equiv j h''_i$ is imaginary ($h'_i$ and $h''_i$ are the real and imaginary parts of $h_i$). This implies that the fields in the the two cladding regions are evanescent; however, the superlayer matrix elements as given in Table 7.7 remain real for this special case.

Special cases, as well as numerical issues, arise when the layer thicknesses of the cladding regions become too large. This leads to the physical condition where energy is trapped in the super layer and cannot escape. Alternatively, if two super layers are adjacent, then power will be trapped in the individual layers and the fields will be uncoupled or incoherent (In a practical sense, this occurs when the “coupling strength” between the two super layers is less than the background noise.) Numerically, coupling is lost when machine precision produces $\cosh(2 h''_i d_i) = \sinh(2 h''_i d_i)$, or when the numerical value of $|1 - \exp(-4 h''_i d_i)|$ is not different from 1.

### 7.4 Mathematics of Superlayers

In 'w', the wave equation was used to characterize the electromagnetic field over a single layer. In particular, the transfer matrix wave developed into a given form such as in the $A, B, C, D$ matrix form. Specifically, for a single layer, say $l$th layer, the matrix parameters are

$$A_l = \cos h_l d_l, \quad B_l = \sin h_l d_l / h_l, \quad C_l = -h_l \sin h_l d_l, \quad D_l = \cos h_l d_l.$$
The thickness of the layer is $d_l$, while its lateral wave number is $h_l$. The optical thickness of the layer is $H_l = h_l d_l$, so that $H_l / 2\pi$ is the number of lateral wave periods across the layer.

The superlayer transfer matrix translates the field/flux over multiple layers of perhaps differing material as well as optical thicknesses. The thickness of the superlayer is simply the sum of the thicknesses of the individual layers, however, the optical thickness of the superlayer is not obvious. One possible answer is to write the optical thicknesses as the sum of optical thicknesses of the individual layers. A more accurate definition of the optical thickness of the superlayer is associated with the matrix elements of the superlayer transformation and its matrix eigenvalues.

Attention is now focussed on the more complex superlayer transformation which may involve multiple groups of individual basic layers. In this complicated case the transformation matrix elements will be denoted as $A_s$, $B_s$, $C_s$, $D_s$. The general superlayer transformation matrix, $T_s$, has the following eigenvalues, $\chi_1$ and $\chi_2$

$$\chi_1, \chi_2 = (A_s + D_s) / 2 \pm \sqrt{(A_s + D_s)^2 / 4 - 1}$$ (7.4.1)

where $\chi_1, \chi_2$ are associated with the $\pm$ signs. The eigenvectors $v_1$ and $v_2$ tied to the eigenvalues respectively are

$$v_1 = \begin{pmatrix} 1 \\ v_{21} \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ v_{22} \end{pmatrix}$$ (7.4.2)

where the two bottom elements of the eigenvectors are given by

$$v_{21}, v_{22} = -(A_s - D_s) \pm \sqrt{(A_s - D_s)^2 + 4B_s C_s}$$ (7.4.3)

where the $v_{21}$ uses the positive value of the radical while $v_{22}$ uses the negative value of the radical. Now the super matrix satisfies the following matrix equation

$$T_s V = V X$$ (7.4.4)

where the matrix $V$ has columns formed from first eigenvectors.

$$V = \begin{pmatrix} 1 & 1 \\ v_{21} & v_{22} \end{pmatrix}, \quad X = \begin{pmatrix} \chi_1 & 0 \\ 0 & \chi_2 \end{pmatrix}$$

It should be noted that the eigenvectors of $T_s \cdot T_s$ are identical to the eigenvectors of $T_s$, but where the eigenvalues of $T_s$ are the diagonal elements of $X$, the eigenvalues of $T_s \cdot T_s$ are the diagonal elements of $X^2 = X \cdot X$. Whereas, the eigenvalues of $T_s^N$ are $\chi_1^N$ and $\chi_2^N$, the eigenvectors are identical to those of $T_s$.

At this point, the concept of the optical thickness of a superlayer can be addressed. In (7.4.1), the eigenvalues $\chi_1$ and $\chi_2$ can be written as

$$\chi_1, \chi_2 = \exp(\pm j H_s)$$ (7.4.5)
where the superlayer optical thickness $H_s$ satisfies
\[ \cos(H_s) = \frac{1}{2}(A_s + D_s). \] 
(7.4.6)

Accordingly, the transverse lateral wave number $h_s$ in a superlayer can be expressed as
\[ h_s = \frac{H_s}{d_s}. \] 
(7.4.7)

where $d_s = d_i + d_{ii} + d_{iii}$ is the physical thickness of the superlayer, so that the diagonal eigenvalue matrix can be expressed in terms of the wavenumber of the superlayer,
\[ X = \begin{pmatrix} \exp(j h_s d_s) & 0 \\ 0 & \exp(-j h_s d_s) \end{pmatrix}. \] 
(7.4.8)

In general $h_s$ is a complex number, and it depends on the basic geometry of the superlayer and how the superlayer interfaces in complex structures. If superlayers are stacked together such as in a photonic crystal, then the transformation matrix will translate the fields across the stacked superlayers. The lateral wave number $h_s$ corresponds to the Floquet multiplier and the lateral waves will be identical to Bloch waves.

Before leaving this section, it is useful to address the symmetric superlayer. Recall, the symmetric superlayer is characterized by $A_s = D_s$. Thus, the eigenvalues, (7.4.1), become
\[ \chi_1, \chi_2 = A_s \pm j \sqrt{1 - A_s^2}, \] 
(7.4.9)

while the corresponding eigenvectors become
\[ v_1 = \begin{pmatrix} 1 \\ v \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ -v \end{pmatrix}, \] 
(7.4.10)

where $v = v_{21} = v_{22} = \sqrt{C_s/D_s}$. The optical thickness of the superlayer satisfies
\[ h_s d_s = \arccos A_s. \] 
(7.4.11)

Figure 7.5 shows the values of $A_s$ as a function of the wavenumber of the central region for a symmetrical structure. The structure used for the computations has the central region, Region ii, thickness that is twice the size of the each of the cladding layers. The dielectric step is 1.4 which corresponds to an index step of about 0.2 for typical material used for optical systems. As illustrated in the figure, when $h_{ii} \approx 0.65$, the element $A \approx 1$. 
Figure 7.5. The matrix element \( A_s \) for a symmetric superlayer as a function of the wavenumber \( h_{ii} \) of the central layer, Layer ii. The dielectric step is 1.4, and the cladding thicknesses are \( d_i \), \( d_{iii} = d_{ii} / 2 \).

As long as the value of \( A_s \) \(((A_s + D_s)/2 \) for the general superlayer) lies in the range \((-1, 1)\), the optical layer thickness, \( H_s \) and the superlayer wavenumber, \( h_s \) are real quantities, assuming of course that \( A_s \) is real, corresponding to the case of lossless, reactive material in each of the sublayers forming the superlayer. In the event \( A_s \) for the symmetric structure is greater than 1, (or less than -1), the wavenumber \( h_s \) will have an imaginary value which implies exponential growth or exponential decay.

### 7.5 Mode Confinement in Periodic Superlayers

A major advantage of using the superlayer approach discussed above is the fact that repeating or multiple superlayers forming a periodic structure/photonic crystal, can be easily characterized. In practice, superlayers are used in the fabrication of artificial dielectrics, multi quantum-well lasers, the reflecting stacks of VCSEL lasers, and others. Figure 7.6 shows a superlayer that is repeated \( N \) times. The figure shows a repetition of \( N \) layers, there may be an infinite number of such layers.

![Figure 7.6](image.png)

Figure 7.6. A distribution of \( N \) superlayers. The first superlayer is located on the right (between \( x_l - d_s \) and \( x_l \)) while the last superlayer is located on the left (between \( x_l - N d_s \) and \( x_l - (N - 1)d_s \)).

The transformation of the field and flux across a superlayer is is determined from the matrix elements given in Table 7.7. For the first superlayer, the
7.5 MODE CONFINEMENT IN PERIODIC SUPERLAYERS

transformation is $T_s$, and thus the field/flux vector $\Psi_s$ at $x = x_s$ in terms of the field/flux vector, $\Psi_{s+1}$ at $x = x_{s+1}$ is

$$\Psi_{s-1} = T_s \Psi_s$$  \hspace{1cm} (7.5.1)

where the transfer matrix can be written in its factored form

$$T_s = V \begin{pmatrix} \exp(j h_s d_s) & 0 \\ 0 & \exp(-j h_s d_s) \end{pmatrix} V^{-1}. \hspace{1cm} (7.5.2)$$

Assuming there are $L$ superlayers, the transformation of the field/flux across the layers becomes

$$T_s^L = V \begin{pmatrix} \exp(j L h_s d_s) & 0 \\ 0 & \exp(-j L h_s d_s) \end{pmatrix} V^{-1}. \hspace{1cm} (7.5.3)$$

For the sake of simplicity, mode confinement in superlayer architectures will be now limited to modes of “even” and “odd” functional dependence. The secular equation that defines these modes will be discussed below. There are two basic types of symmetric superlayer wave guides, shown in Fig. 7.7, that yield even and odd mode shapes.

![Image of superlayer waveguide architectures](image)

Figure 7.7. Dielectric profiles of superlayer waveguide architectures. In (a), the core is a single uniform layer while the cladding is an infinite number of stacked superlayers, and in (b), the core region is made from a stack of superlayers of thicknesses $t_s$.

Both structures in Fig. 7.7 may have “even” and “odd” bounded modes or there may be no bounded modes present. First, the general solution for modes of the structure in Fig. 7.7(a) will be characterized. In the core region, the even field solutions are given by

$$\Psi_e(x, z) = (a \cos h_c x - b \sin h_c x) \exp(-y z), \hspace{1cm} (7.5.4)$$
where the wavenumber in the core satisfies

$$h_c^2 = \gamma^2 + k^2 \kappa_{ij}.$$ 

In this case, the dielectric constant of the core is identical to that of the second region of the superlayer as illustrated in the figure, and the wave number in the core \( h_c = h_{ii} \). The even fields have \( b = 0 \) while the odd fields have \( a = 0 \). The field/flux at the beginning of the superlayer adjacent to the core is

$$\begin{pmatrix} \psi(t) \\ \psi'(t) \end{pmatrix} = \begin{pmatrix} \cos h_c t & \sin h_c t/h_c \\ -h_c \sin h_c t & \cos h_c t \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \quad (7.5.5)$$

The field/flux will now be transformed across \( L \) superlayers, using (7.5.3),

$$\begin{pmatrix} \psi(t + L d_s) \\ \psi'(t + L d) \end{pmatrix} = V \begin{pmatrix} \exp(j h_s L d_s) & 0 \\ 0 & \exp(-j h_s L d_s) \end{pmatrix} V^{-1} \begin{pmatrix} \psi(t) \\ \psi'(t) \end{pmatrix}$$

$$= V \begin{pmatrix} \exp(j h_s L d_s) [a (\cos h_c t - h_c \sin h_c t/v) + b (\sin h_c t/h_c + \cos h_c t/v)] \\ \exp(-j h_s L d_s) [a (\cos h_c t + h_c \sin h_c t/v) + b (\sin h_c t/h_c - \cos h_c t/v)] \end{pmatrix}. \quad (7.5.6)$$

Recall that when \( \lambda_s > 0 \) (or \( \lambda_s + 2d_v/2 > 0 \) for the general non-symmetric case) the superlayer wave number \( \text{Im} \{ h_s \} > 0 \) so that \( |\exp(-j h_s L d_s)| \to \infty \) as \( L \to \infty \). Accordingly, a bounded mode exists provided the coefficient of \( \exp(-j h_s L d_s) \) vanishes. The secular equation for even modes, \( b = 0 \) is

$$v = -h_c \tan h_c t, \quad (7.5.7)$$

while for the odd modes, \( a = 0 \), it is

$$v = h_c \cot h_c t, \quad (7.5.8)$$
Chapter 8

Waveguides Characteristics near the 2nd Bragg Condition

The periodic dielectric waveguides have electromagnetic characteristics that are considerably different from similarly constructed non-periodic structures. Periodic structures can exhibit significant power transfer between counter travelling waves affecting the propagation characteristics. As a result, standing waves are formed at the so-called Bragg conditions; the point where resonance occurs. By utilizing the resonance, the grated waveguide can be constructed as a band-reject filter or a distributed-Bragg reflector (DBR), usually at the “1st Bragg condition.” At wavelengths near the 2nd Bragg condition, periodic structures may produce transverse radiation due to a “fast wave” or leaky spatial harmonic that has a small longitudinal propagation constant, due to the slowness of the wave. Such grating structures fabricated with with optical sources such a semiconductor laser, may produce a highly directional, narrow band and a small divergent radiation beam. At the 2nd Bragg condition, the radiation beam is normal to the propagation direction. The analysis of the periodic structure is relatively complicated and numerous techniques have been employed for analysis.[9, 16, 31] A major method used to analyze periodic structures employs Floquet’s theorem that indicates the envelope of the propagating wave is periodic with a period identical to the grating period. The theorem was used by Bloch to analyze Schrodinger’s wave equation in crystals.

Dielectric waveguides with special layers whose dielectric constant varies with $z$ (the propagation direction) are usually fabricated with multiple layers, along the $x$ direction, of dielectric materials where the special layer is periodic along one direction, $z$. As a result, a mode propagating in the $z$ direction have
shapes dependent upon both \( x \) and \( z \). In Chapter 3, the propagation characteristics of plane waves in periodic-layered structures were discussed. In this chapter, the characteristics of waveguide modes in a multi-layer structure, with a special layer whose dielectric constant varies along the propagation direction, will be developed. All layers, except the grating layer, are uniform and isotropic.

8.1 Grating Layer

A coordinate system will be chosen such that the grating layer will referenced at the transverse location \( x = 0 \) such as that shown in Fig. 8.1. In the figure, the grating is formed in layer 3. Figure 8.1 shows a periodic grating with two regions fabricated with two different dielectric materials. The dielectric constant of the \( n \)th layer is given by \( \kappa_n \). Although the dielectric materials of the grating layer may be formed with arbitrary materials, it is typically formed with the dielectric materials of the neighboring layers. For example, the central “tooth” region would have a dielectric constant \( \kappa_t = \kappa_4 \) and the “fill” region would have a relative dielectric constant of \( \kappa_f = \kappa_2 \).

In the grating layer, the dielectric constant is independent of \( y \) and \( x \) but it is periodic with respect to \( z \), the propagation direction, and may be expanded in a spatial Fourier series

\[
\kappa_g(z) = \sum_{m=-\infty}^{\infty} \kappa_m e^{-j m K z},
\]

where the \( K = 2\pi/\Lambda \) is the grating wave number, \( \kappa_m \) is the Fourier coefficient

\[
\kappa_m = \frac{1}{\Lambda} \int_{-\Lambda/2}^{\Lambda/2} \kappa_g(z) e^{j m K z} \, dz
\]
When the grating layer thickness, \( t_h \) is small compared to the transverse field extension, the periodic layer will have only a slight influence (perturbation) on the shapes of the transverse fields. A first order analysis of the electromagnetic fields can be obtained by replacing the periodic layer with a material whose dielectric constant is the average value

\[
\kappa_0 = \frac{1}{A} \left[ t_w \kappa_1 + (A - t_w) \kappa_f \right],
\]

which is the 0th Fourier coefficient. If the fraction \( t_w/A = d_c \), the average dielectric constant \( \kappa_0 = d_c \kappa_1 + (1 - d_c) \kappa_f \). Note that all Fourier coefficients are independent of the grating period, \( A \), and the free-space wavelength, \( \lambda_\infty \).

The Fourier expansion of the relative dielectric constant in (8.1.1) can be used to extract information regarding the interaction with an electromagnetic mode that impinges on a region with an embedded grating such as that in Fig. 8.1. For example, if grating period, \( A \), is near the half-wavelength of a propagating mode, \( (\lambda_\infty/2)/n_e \), where \( \lambda_\infty \) is the free-space wavelength and \( n_e \) is the effective index of the mode, then back scatter from the multiple waveguide discontinuities are “in-phase,” and enhances the backscatter. These reflections are called “1st order” Bragg scattering. In general, backscatter can be accomplished with a multitude of grating periods such as for m-th order Bragg scattering,

\[
-\frac{1}{m} A = \frac{\lambda_e}{2},
\]

where \( \lambda_e = \lambda_\infty/n_e \) is the effective wavelength of the mode in the waveguide. When (8.1.4) is satisfied, the negative space harmonic in (8.1.1), \(-|m|\) interacts with the incident mode.

![Figure 8.2. Scattering from each of the scattering centers. When the round trip path from scatterer 2 to scatterer 3 and back to scatterer 2 is a multiple of the mode wavelength, \( \lambda_e \), back scattering is enhanced.](image)

If the distance between the scattering centers is \( \lambda_e/2 \) then there is a phase difference of \( \pi \) radians between each neighboring center; this is similar to an “end-fire” array and produces “back” scattering/radiation. When the distance between centers is \( \lambda_e \), 2nd Bragg condition, then neighboring elements are “in phase;” this is similar to a “broadside” array and produces scattering/radiation at 90° to the array axis. At a spacing of \( 3\lambda_e/2 \), 3rd Bragg condition, neighboring elements are again \( \pi \) radians out of phase, similar to the 1st Bragg condition.

The strength of the Fourier coefficients for the various spatial harmonics determines the amount of interaction of a mode and the grating, assuming the wavelength condition is satisfied. For the rectangular grating shown in Fig. 8.1, the Fourier coefficients are given as

\[
\kappa_m = (\kappa_1 - \kappa_f) \frac{\sin m\pi d_c}{m\pi}, \quad m \neq 0.
\]
When the duty cycle $d_c = 1/2$, the coefficients $\kappa_m = 0$ when $m$ is an even integer so it appears there would be little or no interaction with a propagating mode whose effective wavelength, $\lambda_e$ satisfies the Bragg condition. Similarly, when the duty cycle $d_c = 1/3$, $\kappa_m = 0$ when $m$ is a multiple of 3. However, when $d_c = 1/2$, the Fourier coefficients $\kappa_{\pm 1}$ have maximum values. The magnitude of the Fourier coefficients decreases slowly, $1/|m|$, because of the abrupt changes in the dielectric constant at the tooth edges.

For symmetric gratings, the coefficients in (8.1.2) satisfy $\kappa_m = \kappa_{-m}$. Furthermore, if the dielectric constant of the two regions forming the grating, have no losses, the coefficients are all real, $\Im\{\kappa_m\} = 0$. The first few coefficients for different duty cycles is shown in Table 8.1. The coefficients for $d_c = 2/3$ and $3/4$ are similar to $1/3$ and $1/4$ respectively.

### Table 8.1. The Fourier coefficients for a rectangular tooth for different duty cycles, $d_c$.

<table>
<thead>
<tr>
<th>$d_c$</th>
<th>1/2</th>
<th>1/3</th>
<th>1/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>$\kappa_m/(\kappa_t - \kappa_f)$</td>
<td>$\kappa_m/(\kappa_t - \kappa_f)$</td>
<td>$\kappa_m/(\kappa_t - \kappa_f)$</td>
</tr>
<tr>
<td>$\pm 1$</td>
<td>0.318</td>
<td>0.276</td>
<td>0.225</td>
</tr>
<tr>
<td>$\pm 2$</td>
<td>0</td>
<td>0.138</td>
<td>0.159</td>
</tr>
<tr>
<td>$\pm 3$</td>
<td>$-0.106$</td>
<td>0</td>
<td>0.075</td>
</tr>
<tr>
<td>$\pm 4$</td>
<td>0</td>
<td>$-0.069$</td>
<td>0</td>
</tr>
<tr>
<td>$\pm 5$</td>
<td>0.064</td>
<td>$-0.051$</td>
<td>$-0.045$</td>
</tr>
</tbody>
</table>

A 1st-order Bragg reflector has the largest Fourier coefficient when the duty cycle is $1/2$, while a 2nd-order Bragg reflector has the largest Fourier coefficient for $1/4$ duty cycle, which is half the size of the coefficient of the 1st-order Bragg reflector. The value of the coefficient $\kappa_m$ is calculated by multiplying the factors in the table with the relative dielectric constant difference $\kappa_t - \kappa_f$. The calculations for duty cycles of 2/3 and 3/4 can be obtained using the coefficients in Table 8.1 by interchanging $\kappa_t$ and $\kappa_f$.

### 8.2 Floquet-Bloch Waves

Early studies of periodic waveguides were focused on the problems in antenna design and analysis, such as wave guiding, scattering and radiation pattern [20, 19, 27, 26]. In a dielectric slab waveguide with neither losses nor gain, in any layer, there is no electromagnetic power loss for a bounded propagating mode, unless there is a discontinuity. On the other hand, a grating tooth is a discontinuity and thus produces significant scattering.

Wave phenomena such as that of electromagnetic propagation in periodic waveguides can be described by the superposition of all forward and backward waves which are excited by the periodic structure. According to Floquet’s theorem, the fields can mathematically characterized in terms of phase difference between the propagating fields and is discrete and equal to the multiple of the unit reciprocal grating period, $K = 2\pi/\Lambda$. 


In general, the field distribution of a wave propagating in the positive $z$ direction in a waveguide that is periodic along the propagation direction $z$ can be expressed in terms of Floquet-Bloch analysis as

$$\Psi(x, z) = e^{-\gamma z} R(x, z),$$

where $R(x, z) = R(x, z + \Lambda)$ is a periodic function along $z$ and may be expanded in a Fourier series

$$R(x, z) = \sum_{n=-\infty}^{\infty} \psi_n(x)e^{-jnKz},$$

which produces the field solution

$$\Psi(x, z) = \sum_{n=-\infty}^{\infty} \psi_n(x)e^{-\gamma n z}.$$  \hspace{1cm} (8.2.3)

where $\gamma_n = \gamma + jnK$.

The complex propagation constant $\gamma = \alpha + j\beta$, where $\alpha$ and $\beta$ are the wave attenuation coefficient and propagation constant respectively. The grating wave number $K = 2\pi/\Lambda$. The Fourier coefficient $\psi_n(x)$ is called the $n$th space harmonic. In general, the space harmonics $\psi_n(x)$ and $\psi_m^*$ are not orthogonal, i.e.,

$$\langle \psi_m | \psi_n \rangle = \int_{-\infty}^{\infty} \psi_m^*(x) \psi_n(x) dx \neq 0.$$

Thus, the propagating wave consists of a set of space harmonics called partials that all travel with the same group velocity so that a light pulse injected into the grating will remain intact as it propagates. However, there may be an exchange of power from one partial to another because of their non-orthogonality. When there exists multiple modes, interaction between two modes can occur when $\beta_p + \beta_q = \ell K$, where $p$, $q$, and $\ell$ are integers. The subscripts $p$ and $q$ pertain to different guided modes in a periodic waveguide.

### 8.2.1 TE Modes

In practice, dielectric waveguides with a periodic region embedded is constructed from uniform, isotropic materials. This construction will produce two independent sets of waveguide modes, TE and TM. While modes of either the TE set or the TM set may interact as they propagate, TE and TM modes will not interact because their polarizations are orthogonal. Although TE and TM modes have different wave characteristics, the numerical solutions in periodic waveguide have similar accuracy, stability and convergence issues, especially for a periodic waveguide with a large grating tooth height and a large buffer layer, the layer separating the main waveguide and the grating region. In the section below, an algebraic eigenvector method originally developed by Peng et al.[22] will be used for the analysis and understanding of Floquet-Bloch modes.
This section will be concerned with the propagation characteristics of TE-type waves. The TE waves (electric field has only an $E_y$ component) propagating in positive $z$ direction is governed by three components of Maxwell’s equations

$$-\frac{\partial E_y}{\partial z} = -j\omega \mu \alpha H_x, \quad (8.2.4)$$

$$\frac{\partial E_y}{\partial x} = -j\omega \mu \alpha H_z, \quad (8.2.5)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega \varepsilon_0 \kappa_g(x, z) E_y, \quad (8.2.6)$$

Substituting the first two equations into the third gives the wave equation representing the field component $E_y$.

The Helmholtz equation for the uniform-isotropic layered waveguide can be written as

$$\frac{\partial \Psi^2}{\partial x^2} + \frac{\partial \Psi^2}{\partial z^2} + k_0^2 \kappa_g(x, z) \Psi = 0, \quad (8.2.7)$$

and the resulting field solutions are

$$E_y = \Psi(x, z) \quad (8.2.8)$$

$$H_x = \frac{1}{j\omega \mu_0} \frac{\partial \Psi}{\partial z} \quad (8.2.9)$$

$$H_z = -\frac{1}{j\omega \mu_0} \frac{\partial \Psi}{\partial z} \quad (8.2.10)$$

A set of differential equations for the space harmonics can be obtained by substituting (8.2.3) into the Helmholtz equation, (8.2.7) and using the expansion, (8.1.1)

$$\frac{d^2 \psi_n(x)}{dx^2} + \gamma_n^2 \psi_n(x) + k_0^2 \sum_{m=-\infty}^{\infty} \kappa_{n-m}(x) \psi_m(x) = 0. \quad (8.2.11)$$

The relative dielectric constant $\kappa_g$ in (8.2.7) is a function of $z$ only in the grating layer. The coefficient $\kappa_0(x)$ is the largest coefficient and it is instructive to remove the $n = m$ term from the summation in (8.2.11)

$$\frac{d^2 \psi_n(x)}{dx^2} + \left( \gamma_n^2 + k_0^2 \kappa_0(x) \right) \psi_n(x) + k_0^2 \sum_{m=-\infty}^{\infty} \kappa_{n-m}(x) \psi_m(x) = 0. \quad (8.2.12)$$

The summation acts as a perturbation terms on the set of differential equations while $\kappa_0(x)$ is the average dielectric constant of the grating region.

In Fig. 8.1, the grating “tooth” has a rectangular shape rendering $\kappa_g$ and $\kappa_{n-m}$ independent of $x$. When the grating tooth is not rectangularly shaped, then $\kappa_g$ and the corresponding $\kappa_{n-m}$ Fourier coefficient have transverse $x$ dependence, which results in a set of differential equations with non-constant (with respect to $x$) coefficients. (While differential equations with constant coefficients have well-known solutions, solutions of differential equations with
variable coefficients may not have simple “closed-form” answers and one must rely on the use of numerical methods.)

The “fill” regions of the grating may consist of multiple material compositions. For example, Fig. 8.3 illustrates a single grating period that consists of a “tooth” region and several “fill” regions. Furthermore, the dielectric constant in the grating layer is also independent of the transverse \( x \) dimension.

![Figure 8.3](image)

Figure 8.3. (a) Tooth shape in the grating consists multiple “fill” regions, and (b) the profile of the dielectric constant over a single period. The set of rectangular regions in (b) approximate the actual grating shape.

As the “tooth” and “fill” regions increase in number, the variation of \( \kappa(z) \) tends to a continuous function, with respect to \( z \). Further, the continuously varying dielectric constant is also independent of transverse direction \( x \). Accordingly, the expansion coefficients, \( \kappa_{n-m}(x) \), in (8.2.11) are independent of \( x \). For example, if \( \kappa_{\text{max}} \) and \( \kappa_{\text{min}} \) are the maximum and minimum values of \( \kappa(z) \) in the grating period \( \Lambda \), then a sinusoidal grating might have a functional dependence

\[
\kappa(z) = \left( \frac{\kappa_{\text{max}} - \kappa_{\text{min}}}{2} \right) \cos \frac{Kz}{2} + \frac{\kappa_{\text{max}} + \kappa_{\text{min}}}{2}.
\]

The technique developed by Peng[22] constrains dielectric variation in the grating layers to the propagation direction, i.e., the dielectric constant varies only with \( z \). To allow for arbitrarily shaped profiles, the grating layer must be partitioned into multiple layers with rectangular shaped teeth that approximate the arbitrary profile. Figure 8.4(a) and (b) illustrate the detail of rectangular non-rectangular tooth shapes, respectively. In Fig. 8.4 (b), the triangular-shaped tooth is represented by a sequence of rectangular-shaped regions which may be increased or decreased depending on a desired approximation of the actual shape.[22] Although, modal characteristics of arbitrarily-shaped grating profiles have been developed,[3, 16, 31, 17, 32], the partitioning of the arbitrary shapes into a series of rectangular regions can similarly provide solutions to any desired accuracy.

![Figure 8.4](image)

Figure 8.4. Tooth shapes in the grating region may be rectangular, (a), or arbitrarily shaped such as that illustrated in (b). The set of rectangular regions in (b) approximate the actual grating shape.
Although the material used to form layers in devices have dielectric constants that are uniform and isotropic, devices with non-uniform, anisotropic materials may be used. When non-uniformity is an issue, the rectangular regions may also be partitioned to approximate the non-uniformity.

Outside the grating, the solution of the Helmholtz equation can be reduced to that of solving a second-order differential equation in all layers without a grating. However, to find a solution for the waveguide modes it will be necessary to match solutions outside the grating region to field solutions within the grating region.

8.2.2 Field Solutions in a Periodic Layer
One technique of solving for the fields in the grating is to assume that only a finite number of space harmonics in (8.2.11) play a significant role in the determination of the mode. Assuming that the smallest indexed space harmonic is $\psi_L$ while the largest is $\psi_H$ (the total number of space harmonics is $M = H - L + 1$), then the infinite set of differential equations reduces to a finite set of $M$ second order differential equations that may be written in vector form as

$$\frac{d^2\psi}{dx^2} = -(\Gamma^2 + k_\alpha^2 \hat{K})\psi = -\hat{P} \psi,$$  \hspace{1cm} (8.2.13)

where $\Gamma$ is a diagonal matrix of $\gamma_n$’s, and $\hat{K}$ is a Toeplitz matrix whose entries are obtained from the set of Fourier coefficients $\{\kappa_n(x)\}$ in (8.1.1). The resulting matrix $\hat{P}$ has elements

$$\hat{P}_{mn} = (\gamma + jnK)^2 \delta_{mn} + k_\alpha^2 \kappa_{m-n},$$  \hspace{1cm} (8.2.14)

where $\delta_{mn}$ is the Kronecker delta. The vector $\psi$ is

$$\psi = \begin{pmatrix} \psi_L \\ \psi_{L+1} \\ \vdots \\ \psi_H \end{pmatrix},$$ \hspace{1cm} (8.2.15)

has entries made from the individual space harmonics. It should be noted that the set of $M$ second-order equations may be reduced to a set of $2M$ first-order equations. The $2M$ set will produce a set of $2M$ $x$-dependent solutions.

When the matrix $\hat{P}$ is independent of $x$, the solution of the vector differential equation, (8.2.13), can be written as

$$\psi = ve^{jhx},$$  \hspace{1cm} (8.2.16)

and the substitution of $\psi$ into (8.2.13) gives

$$\hat{P}v = h^2 v \equiv \sigma v.$$ \hspace{1cm} (8.2.17)

The solution of (8.2.11) reduces to the algebraic eigenvalue problem where $h^2 = \sigma$ is the eigenvalue of $\hat{P}$ and $v$ is the corresponding eigenvector.
There are two exponential functions, in the solution vector, composed of \( \exp(j h x) \) and \( \exp(-j h x) \) or a combination of the two exponential functions. The number of different pairs of lateral propagation constants, \( (h_v = \pm \sqrt{\sigma_v}) \), corresponds to the dimension of the \( \hat{P} \) matrix and is equal to \( M \), the number of space harmonics that will be used to form the solution. The set of eigenvectors \( \{\mathbf{v}_L, \ldots, \mathbf{v}_H\} \) form a vector space and the solution to the fields in the grating region must be determined from a linear combination of these eigenvectors.

The vector \( \mathbf{v}_i \) with the corresponding eigenvalue \( \sigma_i \), is called a right-hand eigenvector. There is a left-hand eigenvector \( \mathbf{u}_i \) that has the same eigenvalue and satisfies

\[
\mathbf{u}_i^H \hat{P} = \sigma_i \mathbf{u}_i^H, \tag{8.2.18}
\]

where \( \mathbf{u}_i^H \) is the transpose of the complex conjugate of \( \mathbf{u} \). The vectors in the set \( \mathcal{S} \) form a Hilbert space and they are orthogonal to vectors in the \( \{\mathbf{u}_i\}_{i=1}^M \) \( i.e., \) if the eigenvalues \( \sigma_i \neq \sigma_j \), then the inner product \( \mathbf{u}_i^H \cdot \mathbf{v}_j = 0 \). (In case degeneracy sets in, the Gram-Schmidt process can be employed.) Orthogonality allows for simple evaluation of expansion coefficients of an arbitrary vector and the corresponding coefficients are optimum in a mean-square sense for any vector in a subspace. (Expansions of eigenfunctions are commonly used for various applications. Perturbational analyses of energy eigenvalues and their corresponding wave functions can be accomplished using eigenvector expansions[14]) For example, assume the vector \( \mathbf{e}_i = \text{col}(0, \ldots, 1, \ldots, 0) \), where the non-zero element appears in the \( i \)th row, then

\[
\mathbf{e}_i = \sum_{j=L}^H E_{ji} \mathbf{v}_j, \tag{8.2.19}
\]

and the expansion coefficients become \( E_{ii} = \mathbf{u}_i^H \cdot \mathbf{e}_i \), provided the vectors have been normalized, \( \mathbf{u}_i^H \cdot \mathbf{v}_j = \delta_{ij} \).

### 8.2.3 Vector Expansions for Symmetric Gratings

In the special case when the dielectric constant is symmetric about \( z = 0 \), \( \kappa_e(x, z) = \kappa_e(x, -z) \), then the complex matrix \( \hat{P} \) is symmetric; it can be shown that \( \mathbf{u}_i^H = \mathbf{v}_i^T \), the transpose of \( \mathbf{v}_i \). The expansions of the eigenvectors about the point \( \gamma = \gamma_0 \), in terms of the eigenvectors obtained from (8.2.17) will be developed for gratings that are symmetric about \( z = 0 \). The expansions will be used in the evaluation of the derivatives of \( \mathbf{v} \) and \( \sigma \) with respect to \( \gamma \).

The propagation constant, \( \gamma \) of a Floquet-Bloch mode is determined from the solution of a set of transcendental equations that form a matrix \( G(\gamma, k_0) \), given in (8.2.42). Typically, an iterative approach to finding a solution for \( \gamma \) is obtained by driving the determinant of \( G(\gamma, k_0) \) to zero. An iteration that uses Newton’s method requires calculating the derivative of \( G(\gamma, k_0) \) with respect to \( \gamma \), so that the derivatives of the eigenvalues and eigenvectors in (8.2.17) must
be calculated in the vicinity of \( \gamma = \gamma_0 \). Retaining only the first three terms, the expansion of (8.2.17) about \( \gamma = \gamma_0 \) for the \( i \)th eigenvalue/vector, gives

\[
\left( \hat{p}^{(0)} + \hat{p}^{(1)} \delta \gamma + \frac{1}{2} \hat{p}^{(2)} \delta \gamma^2 \right) \left( v_i^{(0)} + v_i^{(1)} \delta \gamma + \frac{1}{2} v_i^{(2)} \delta \gamma^2 \right) = \left( \sigma_i^{(0)} + \sigma_i^{(1)} \delta \gamma + \frac{1}{2} \sigma_i^{(2)} \delta \gamma^2 \right) \left( v_i^{(0)} + v_i^{(1)} \delta \gamma + \frac{1}{2} v_i^{(2)} \delta \gamma^2 \right),
\]

where \( \delta \gamma = \gamma - \gamma_0 \), and the superscript \((n)\) represents the derivative \( \partial^n / \partial \gamma^n \) at \( \gamma = \gamma_0 \). A first order solution of the eigenvector is \( v_i = v_i^{(0)} + v_i^{(1)} \delta \gamma \), and the corresponding normalization \( v_i^T v_i = 1 \) requires \( v_i^T v_i^{(1)} = 0 \). Hence, the vector \( v_i^{(1)} \) is orthogonal to \( v_i \), i.e., \( v_i^{(1)} \) in \( S \), where \( S \) is the space of all vectors save \( v_i \). The first order results allows for the calculation of \( v_i^{(1)} \), and \( v_i^{(1)} \), given by

\[
\sigma_i^{(1)} = v_i^{(0)^T} \hat{p}^{(1)} v_i^{(0)} \equiv \hat{p}_{ij}^{(1)},
\]

\[
v_i^{(1)} = \sum_{j \neq i} \frac{\hat{p}_{ji}^{(1)}}{\sigma_i - \sigma_j} v_i^{(0)}. \tag{8.2.21}
\]

The element \( \hat{p}_{ji}^{(1)} \) is \( v_i^{(0)^T} \hat{p}^{(1)} v_i^{(0)} \), and not the derivative, with respect to \( \gamma \), of the matrix element in (8.2.13). The matrix with terms \( \hat{p}_{mn}^{(1)} = e_m^T \hat{p}^{(1)} e_n = 2(\gamma + j n K) \delta_{mn} \) is diagonal. Furthermore, \( \hat{p}_{mn}^{(2)} = 2 \delta_{mn} \) and higher order elements are zero.

The second order coefficients are

\[
\sigma_i^{(2)} = \hat{p}_{ii}^{(2)} + 2 \sum_{k \neq i} \frac{(\hat{p}_{ik}^{(1)})^2}{\sigma_i - \sigma_k}, \tag{8.2.23}
\]

\[
v_i^{(2)} = \sum_{j \neq i} \frac{1}{\sigma_i - \sigma_j} \left( \hat{p}_{ji}^{(2)} + 2 \sum_{k \neq i} \frac{\hat{p}_{jk}^{(1)} \hat{p}_{ki}^{(1)}}{\sigma_i - \sigma_k} - 2 \frac{\hat{p}_{ii}^{(1)} \hat{p}_{ji}^{(1)}}{\sigma_i - \sigma_j} \right) v_i^{(0)}. \tag{8.2.24}
\]

8.2.4 Grating Fields

Since, the set of vector equations are derived from the second order differential equation, the general solution will be written as the sum of \( 2M = 2(H - L + 1) \) independent solutions as

\[
\Psi_v = (Q_v \cos h_v x + P_v \sin h_v x / h_v) v_v, \quad v = L, \ldots, H. \tag{8.2.25}
\]

The coefficient of \( v_v \) is the expansion variable that depends only on the index \( v \) so that a general vector in the space \( S \) will be linear combination of the base vectors as in (8.2.19), which is written as

\[
\Psi = \sum_{v=L}^{H} (Q_v \cos h_v x + P_v \sin h_v x / h_v) v_v. \tag{8.2.26}
\]
The complete set of base eigenvectors can be written in matrix form as

\[ V = (v_L, v_{L+1}, \ldots, v_H). \]  

(8.2.27)

A single transverse wave number \( h_v \) does not define a space harmonic, as a space harmonic is obtained from a specific combination of the set of eigenvalues, \( \{ h_n^2 \} \). In particular, rows of \( V \) are associated with the individual space harmonics, \( \text{viz.} \),

\[ \psi_n = \sum_v V_{nv}(Q_v \cos h_v x + P_v \sin h_v x / h_v), \]  

(8.2.28)

where the expansion coefficient is \( Q_v \cos h_v x + P_v \sin h_v x / h_v \). The 2\( M \) coefficients \( Q_v, P_v \) are determined from the process of the matching the individual space harmonics across all layers of the waveguide while the matrix \( V \) is formed from the set \( S \). Coupling between the space harmonics comes from the off-diagonal elements of the Toeplitz matrix whose entries are the Fourier coefficients of the dielectric expansion in the grating region. It is interesting to note the evolution of \( V \) as the duty cycle changes. For duty cycles of 0 and 1, \( V \) is a diagonal matrix or unit matrix with proper vector order and normalization. For the special duty cycles a single vector \( v_v \) represents a space harmonic and the corresponding \( h_v \) represents the transverse wave number of that space harmonic. However, as given in Table 8.1, the 50 percent duty cycle produces the largest magnitude of off-diagonal terms and will strongly affect coupling of all vectors in \( S \).

**Characteristics of the Eigenvalues \( h_v \).**

Before addressing field calculations outside the grating, it is important to estimate rough behavior of the \( h_v \) eigenvalues because they play a major role in forecasting the numerical accuracy of the complex propagation constant \( \gamma = \alpha + j \beta \) where \( \alpha \) can be correlated with the transmission, reflection and radiation characteristics of a finite-length grating fabricated on a waveguide. Also, the number of space harmonics used for the solution affects the accuracy of the solution.

A rough estimate of the eigenvalues can be obtained by dropping the off-diagonal terms of the Toeplitz matrix \( \tilde{K} \), leaving only the diagonal terms, which are all identical, having a value of \( \kappa_0 \), given in (8.1.3). In a typical structure, the main “sub-waveguide” region that carries the majority of the mode power, is separated by a cladding layer whose dielectric constant is \( \kappa_c \). The grating is then etched onto the cladding layer, taking about 10 to 50 percent of cladding thickness for the grating. The refractive indices of the layers forming the region (“sub-waveguide”) carrying most of the optical power are generally larger than the refractive index of the cladding layer. This geometry produces modes whose effective indices are larger than the refractive index of the cladding. Thus, the grating tooth material is the same as that of the cladding, while the “fill” material is typically air or some combination of air/nitride materials. For a duty cycle \( d_c = 0.5 \), the average value \( \kappa_0 \approx \kappa_c / 2 \).
The eigenvalues $h_\nu$ of $\hat{P}$ are solutions of algebraic equation

$$f(h^2) = \prod_{n=L}^{H} [(\gamma + jnK)^2 + k_0^2 \kappa_c/2 - h^2] = 0. \quad (8.2.29)$$

For an arbitrary space harmonic $n$, the corresponding eigenvalue, $h^2$, satisfies

$$h^2 = (\alpha + jk_o n_e + jnK)^2 + k_0^2 \kappa_c/2. \quad (8.2.30)$$

The free-space wavenumber $k_o$ and the effective index of refraction of the propagating mode $n_e$ have replaced the propagation constant $\beta$.

To investigate the solution obtained for various eigenvalues, the solution of $h^2$ for different $n$'s can easily be observed. For example, when $n = 0$,

$$h^2 = k_0^2 \kappa_c/2 + \alpha^2 - k_0^2 n_e^2 + j 2\alpha k_o n_e.$$

Because $k_0^2 n_e^2 > k_0^2 \kappa_c/2$, $h^2$ lies in the second or third quadrant, depending on the value of $\alpha$, which is generally small compared to other terms in the expression. Thus, $h = h_r + j h_i$ which results in a large imaginary part with $|h_i| >> |h_r|$, so that the corresponding field associated with $h$ will be predominately evanescent. On the other hand, near the second Bragg condition, when the grating wave number $K \approx k_o n_e$, the eigenvalue for the $n = -1$ space harmonic becomes

$$h^2 = k_0^2 \kappa_c/2 + \alpha^2,$$

so that $h \approx k_o \sqrt{\kappa_c/2}$ indicating the fields are non-evanescent.

Employing a large number of space harmonics will of course, increase the accuracy yielded by the model. On the other hand, using a large number of space harmonics makes the model vulnerable to numerical inaccuracies associated with machine precision. Thus, as $n$ becomes large,

$$h^2 \to -n^2 K^2 + 2\alpha + k_0^2 \kappa_c/2 + j 2k_o n_e, \quad (8.2.31)$$

which clearly indicates the fields of the space harmonics of large $n$ are evanescent. The expression in (8.2.28) represents the field when the space harmonics are coupled (the off-diagonal terms of the Toeplitz matrix $K$ are not assumed to be zero). However, for the approximate solution, $\psi_n$ may be written as

$$\psi_n(x) \approx Q_n \cosh h_1 x + P_n \sinh h_1 x / h_1,$$

where $h$ in (8.2.31) is replaced by the dominant imaginary part $j h_i$. When the value of $\exp(-2h_1 x)$ is negligible (via machine precision) with respect to 1, then field calculations become ill-conditioned, causing the ansatz of numerical instability.

A loss of numerical accuracy is associated with the product transverse wave number and grating layer thicknesses. When a grating layer thickness, $t_h$, or the space harmonic number $|n|$ becomes too large, $\exp(-2h_1 t_h)$ may be smaller than machine precision, causing numerical instabilities. Furthermore,
8.2 FLOQUET-BLOCH WAVES

Numerical instabilities may occur with a combination of large $|n|$ and thicknesses for non-grating layers if the mode is evanescent there.

In short, at wavelengths near the second Bragg condition, the discussion above points to the fact that all of the space harmonics, save $n = -1$, have electromagnetic fields in the grating region that are predominately evanescent. At wavelengths near the first Bragg condition, it can be shown that all space harmonics have electromagnetic fields of evanescent type.

8.2.5 Field Solutions External to Grating Layers

Exterior to the grating region, space harmonics are uncoupled because the dielectric constant is independent of $z$, and in the $i$th layer (8.2.11) becomes

\[ \frac{d^2 \psi_{in}(x)}{dx^2} + (\gamma_n^2 + k_0^2 k_i) \psi_{in}(x) = 0. \]  

(8.2.32)

(The first and second subscripts represent the layer and space harmonic respectively.) The field in the $i$th layer may be written in terms of the transverse wave number, $\hat{h}_{in}$, a solution of

\[ \hat{h}_{in}^2 = k_0^2 k_i + (\gamma + jnK)^2. \]  

(8.2.33)

\[
\psi_{in}(x) = Q_{in} \cos \hat{h}_{in}(x - x_i) + P_{in} \sin \hat{h}_{in}(x - x_i) / \hat{h}_{in}.
\]  

(8.2.34)

Layer-interface points $x_i$ are illustrated in Fig. 8.1. In the two semi-infinite regions, $i = 1, N$, the solutions will be written specifically as

\[
\psi_{in}(x) = Q_{in} e^{-\hat{h}_{in}(x-x_i)}; \quad \hat{h}_{in}^2 = -(k_0^2 k_i + (\gamma + jnK)^2) \]  

(8.2.35)

\[
\psi_{Nn}(x) = Q_{Nn} e^{\hat{h}_{Nn}(x-x_{N-n})}; \quad \hat{h}_{Nn}^2 = -(k_0^2 k_N + (\gamma + jnK)^2) \]  

(8.2.36)

which has the appearance of exponential field decay if $\Re(h_{in}) > 0$. However, to allow for the possibility of “fast waves” or leaky modes, allowances must be made for power propagation away from the guided structure, “leak-outward” or for power propagation toward the guided structure, “leak-inward” as illustrated in Fig. 8.5. Figures 8.5 (a) and (b) represent two different modes near the 2nd Bragg condition. In (a) the mode absorbs power for grating periods below the 2nd Bragg condition but looses power above the 2nd Bragg condition while (b) shows the opposite. The plane waves in the superstrate and substrate have different incident angles governed by Snell’s law. The “leak-inward” modes are not physically realizable unless there are sources at $x = \pm \infty$ and the sources produce plane waves with the correct angles of incidence. On the other hand, “leak-outward” modes do not require the external sources so that physically realizable modes switch at the 2nd Bragg condition so that the propagating wave exhibits attenuation.
Figure 8.5. (a) Plane waves leak into the waveguide when $\beta A$ is below the 2nd Bragg condition, but leak out of the waveguide above the 2nd Bragg condition and (b) shows the companion mode exhibiting the opposite leaking. At the 2nd Bragg each mode changes their direction of leakage.

Modes having power leak out can be obtained by making the branch cut, used for square-root calculations of $h_{1n}$ and $h_{N_n}$, along the negative imaginary axis. Thus, all square-root calculations will be forced to the first and second quadrants, making $\Im(h_{1n}), \Im(h_{N_n}) > 0$, i.e., propagation away from the waveguide. Similarly, if the branch is taken along the positive imaginary axis, propagation in the clad regions will be toward the central waveguide region.

It should be noted that $Q_{in}$ represents the field at the base of each layer while $P_{in}$ represent field flux or field derivative with respect to $x$. The tangential field components at the layer interfaces are $E_y$ and $H_z$, and (8.2.10) indicates the continuity of $E_y$ and $H_z$ can be accomplished by matching $\psi_{in}$, and its derivative $d\psi_{in}/dx$ between layers. The matrix that transfers the field and its derivative across a layers accomplishes the matching condition, and ties the field coefficients as

$$
\begin{pmatrix}
Q_{l-1,n} \\
P_{l-1,n}
\end{pmatrix} = \begin{pmatrix}
\cos h_{in}d_i & \sin h_{in}d_i / h_{in} \\
-h_{in} \sin h_{in}d_i & \cos h_{in}d_i
\end{pmatrix} \begin{pmatrix}
Q_{in} \\
P_{in}
\end{pmatrix}.
$$

By repeated application of (8.2.37) the field and its derivative can be transferred from the substrate layer, $N$, to the grating layer. To simplify the expressions, the transfer matrix will be written as

$$
\begin{pmatrix}
A_{in} & B_{in} \\
C_{in} & D_{in}
\end{pmatrix} \equiv \begin{pmatrix}
\cos h_{in}d_i & \sin h_{in}d_i / h_{in} \\
-h_{in} \sin h_{in}d_i & \cos h_{in}d_i
\end{pmatrix}.
$$

The above definition will be applicable for all layers including grating layers.

The Secular Equation

Matching of the proper tangential fields, $E_y$ and $H_z$, at the various layer interfaces will produce the secular equation. Recall that the fields are functions
of both $x$ and $z$. At layer boundaries, fields could be point-matched at several $z$ locations, however, as the fields satisfy Floquet’s theorem, it suffices to match the space harmonics so that field components will be matched for all $z$ positions.

Assuming the structure of Fig. 8.1, the matching the fields will produce a set of 10 equations for each space harmonic

$$Q_{1L} = A_{2L}Q_{2L} + B_{2L}P_{2L}. $$

$$- \dot{h}_{1L} Q_{1L} = C_{2L} Q_{2L} + D_{2L} P_{2L}. $$

$$Q_{2L} = \sum_{v} V_{Lv}(A_{3v} Q_{3v} + B_{3v} P_{3v}). $$

$$P_{2L} = \sum_{v} V_{Lv}(C_{3v} Q_{3v} + D_{3v} P_{3v}). $$

$$\sum_{v} V_{Lv} Q_{3v} = A_{4L} Q_{4L} + B_{4L} P_{4L}. $$

$$\sum_{v} V_{Lv} P_{3v} = C_{4L} Q_{4L} + D_{4L} P_{4L}. $$

$$Q_{4L} = A_{5L} Q_{5L} + B_{5L} P_{5L}. $$

$$P_{4L} = C_{5L} Q_{5L} + D_{5L} P_{5L}. $$

$$Q_{5L} = Q_{6L}. $$

$$P_{5L} = \dot{h}_{6L} Q_{6L}. $$

$$\vdots $$

$$P_{5H} = \dot{h}_{6H} Q_{6H}. $$

The first 10, $(2N - 2)$ where $N$ is the number of layers) equations pertain to the $n = L$, the lowest space harmonic and these 10 equations will be reproduced for each $n = L, L + 1, \ldots, H - 1, H$. The set of equations can be represented in matrix form with the grand matrix $G$ with dimension $2MN \times 2MN$ that can be written in block form as

$$G = \begin{pmatrix}
B_{LL} & B_{L,L+1} & \cdots & B_{LH} \\
B_{L+1,L} & B_{L+1,L+1} & \cdots & B_{L+1,H} \\
\vdots & \vdots & \ddots & \vdots \\
B_{HL} & B_{H,L+1} & \cdots & B_{HH}
\end{pmatrix} \quad (8.2.39)$$

The diagonal blocks all have the same band shape (1 lower diagonal, diagonal, and 2 upper diagonals) and are identical for each space harmonic. For exam-
In the context of the Floquet-Bloch theory, the $B_{nn}$ block is

$$B_{nn} = \begin{pmatrix}
-1 & A_{2n} & B_{2n} \\
1 & C_{2n} & D_{2n} & 0 \\
-1 & 0 & V_{nn}A_{3n} & V_{nn}B_{3n} \\
-1 & V_{nn}C_{3n} & V_{nn}D_{3n} & 0 \\
-V_{nn} & 0 & A_{4n} & B_{4n} \\
-V_{nn} & C_{4n} & D_{4n} & 0 \\
-1 & 0 & A_{5n} & B_{5n} \\
-1 & C_{5n} & D_{5n} & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}$$

The off-diagonal block matrices have only 6 non-zero elements or in general 6 times the number of grating layers. For the off-diagonal blocks $B_{mn}$ where $m \neq n$ and are given by

$$B_{mn} = \begin{pmatrix}
0 & 0 \\
V_{mn}A_{3n} & V_{mn}B_{3n} \\
V_{mn}C_{3n} & V_{mn}D_{3n} \\
-V_{mn} & 0 \\
0 & -V_{mn} \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{pmatrix}$$

The left 0 is $10 \times 3$ block of 0's while the right 0 is a $10 \times 5$ block.

The field coefficients $Q_{in}$ and $P_{in}$ can also be partitioned or blocked as follows

$$q_n^T = (Q_{1n} \; Q_{2n} \; P_{2n} \; \cdots \; Q_{6n})$$

so that all coefficients are contained in the vector $q$

$$q^T = (q_L^T \; q_{L+1}^T \; \cdots \; q_H^T)$$

and the complete system of equations are represented by

$$G(\gamma, k_0)q = 0,$$

where $G$ is shown to be a function of the complex propagation constant $\gamma$ and the free-space wave number $k_0$. Equation 8.2.42 is the secular equation for the system and its solution describes a forward-propagating Floquet-Bloch mode.
Assuming all layers that form the waveguide have no ohmic losses, then typically, one specifies the free-space wave number and a value of $\gamma$ near a singular condition on $G$, then $\gamma$ is perturbed in such a fashion as to make $G$ more singular. In waveguides near the 2nd Bragg condition, $\gamma$ has real and imaginary parts, but when all space harmonics reside in the slow-wave region, $\gamma$’s real part is zero. In this latter case, $G$ is a function of two real variables, $\beta$ and $k_0$. Thus, $\beta$ could be specified, then the problem is find the proper value of $k_0$ that makes $G$ singular. However, with proper mathematical methods, $\gamma$ could be a specified condition, then the problem at hand would be to find the free-space wave number that makes $G$ as singular as possible.

Nodes and Layer Interface Locations

A node is defined as a transverse position $x$, where the field is calculated from the solution of (8.2.42). In the derivation of (8.2.42), the nodes are specified as points that lie at the interface of two distinct layers, including the different layers of the grating region. In general, nodes may be placed at any point in the transverse direction $x$.

Figure 8.6. The nodes in (a) are placed at layer interfaces, while in (b) the nodes are placed at arbitrary locations.

Figure 8.6(a) shows 5 node position where the nodes are placed at layer interfaces so that the node position is $\xi_i = x_i$. However, Fig. 8.6(b) has 8 nodes with the addition of 4 new nodes, while 2 nodes are moved from that shown in (a). The two end nodes are anchor nodes and cannot be deleted but they can be moved to cladding regions 1 and 6. For the node positions shown in Fig. 8.6(b), the field quantities $Q$ and $P$ would be computed at all 8 node positions for each space harmonic. When a node does not lie on an interface the transfer matrix will be computed as a product of 2 single layer transfer matrices. For example, the transfer matrix from $\xi_7$ to $\xi_6$ would be the product of 2 transfer matrices, one from $\xi_7$ to $x_4$ and the other from $x_4$ to $\xi_6$.

Solution of Secular Equation

The number and location of nodes affect the accuracy of the calculation of the eigenvalues $\gamma$ and the eigenvectors $Q$ and $P$. The characteristics of the simple dielectric structure[22, 3] with an etched grating will be derived.
The dielectric structure described by Table 8.2 is a basic 4-Layer waveguide where the superstrait, Layer 1, is air. This particular structure has been extensively investigated by many different techniques. A grating with a 50% duty cycle is formed in Layer 2 where the dielectric constant varies, with respect to the propagation direction, between 3 and 1, so that its average value is 2. When the grating period is \( \Lambda = \lambda/2 \), the structure is lossy because of power loss via radiation. The propagation constant \( \gamma \) has \( \alpha \lambda = 0.018716 \) and the mode’s effective index is \( n_e = \beta/k_0 = 1.5808 \).

A profile of the structure given in Table 8.2 is illustrated in Fig. 8.7. The dielectric profiles across the tooth and fill regions are also illustrated. To illustrate the accuracy of the numerical calculations, 2 different sets of nodes will be used. When the number of nodes (properly chosen) increase in number, the calculation time increases as \( \mathcal{O}(N^2) \) where \( N \) is the number of nodes. However, the overall accuracy is better, particularly with increasing number of space harmonics. In fact, the node spacing must be decreased as the number of space harmonics increases.

![Figure 8.7](image)

**Figure 8.7.** Dielectric profile used for calculations below. Calculations are made for 2 different node numbers and placement. The points, (boxes), are placed 0.2 \( \lambda \) apart, while the circles are placed 0.1 \( \lambda \) apart.

For a non-trivial solution of (8.2.42), the grand matrix \( G \) must be singular, or equivalently, its determinant must be zero. Newton’s method can be used to find the propagation constant that makes the the matrix \( G(\gamma, k_0) \) singular. Assuming the free-space wave number is fixed, \( G \) is expanded about \( \gamma = \gamma_0 \) as

\[
G(\gamma, k_0) = G_0(\gamma_0, k_0) + G_1(\gamma_0, k_0)(\gamma - \gamma_0)
\] (8.2.43)
where $\gamma_0$ is the initial guess value of the propagation constant, and $G_1$ is the derivative of $G$ with respect to $\gamma$ at $\gamma = \gamma_0$. The next iteration point is

$$\gamma_1 = \gamma_0 - \frac{1}{\chi},$$

(8.2.44)

where $\chi$ is the maximum eigenvalue of $G_0^{-1}G_1$. However, as the number of space harmonics increases, for a given node distribution, the matrix $G_0$ becomes ill conditioned because of the large imaginary parts of $h_i$ in (8.2.26) and of $h_{in}$ in (8.2.34). When $h_{in} d_i$ in the latter equation, has a large imaginary part, the cosine and sine expressions are identical, for a given precision, and the transfer matrix is singular and the grand matrix (8.2.39) becomes ill-conditioned for $\gamma$ values in regions near the root.

The condition number of a matrix is a measure of its condition.[5, 8, 10, 2] When the condition number is 0, the matrix is well conditioned or non singular, whereas if the condition number is $\infty$, the matrix is singular and its corresponding determinant is 0. Contemporary linear algebra software is used here to estimate the reciprocal condition number (reciprocal of the condition number, $R_c$) of $G$. In particular, the reciprocal condition number of $G$ will be driven from a value close to 1 (for the initial $\gamma_0$ guess) to a value close to 0, at the desired complex root, via Newton’s method.

For the structure in Fig. 8.7, the initial value of $\gamma_0$ has $\lambda a_0 = 0.0004$ and $\beta_0/k_o = 1.57$, which are in the neighborhood of the approximate solution pair (0.002357, 1.581319). Thus, the initial guess for $\gamma_0$ should produce a large value of $R_c$ compared to the machine precision. Starting with small number of space harmonics, the reciprocal condition number will be calculated as a function of space harmonics (SH).

![Figure 8.8](image_url)

Figure 8.8. The reciprocal condition number as a function of space harmonics. Calculations are made for 2 different total node numbers having uniform separation; boxes represent the reciprocal condition numbers for 0.2 $\lambda$ node separation, while the circles are for 0.1 $\lambda$ node separation.
The first iteration of the Newton search involves the calculation of $R_c$ for the initial value $\gamma_0$. For the uniform node spacing of $0.2\lambda$, and 3 SH, the initial $R_c$ is about $10^{-3}$. Using Newton iterations ($\sim 3-4$), $R_c \to 10^{-20}$ and the Newton step (for $\gamma$) is less than $10^{-12}$ which is close to machine precision. On the other hand, as the number of space harmonics is increased above $\sim 20$ the initial computed value of $R_c$ drops below $\sim 10^{-15}$, and consecutive iterations are unsuccessful in reducing $R_c$. Indeed, for space harmonic numbers above about 27, $R_c$ fluctuate and the Newton step sizes $\to 0$.

When the node separation is $0.1\lambda$, initial values of $R_c$ are higher for identical numbers of space harmonics. Indeed, the lower node spacing produces stable results for a number of space harmonics exceeding 45. The cost for the additional accuracy is computation cycles.

The process of adding space harmonics to obtain the “true” solution for the propagation constant is illustrated in Fig. 8.9. (Cauchy convergence, analyzed from the figure will be discussed below.) The blue markers were obtained with the large inter node spacing while the red markers pertain to the small inter node spacing. As noted above, when the number of space harmonics is greater than $\sim 20$, the solutions for $\gamma$ obtained for the $0.2\lambda$ inter node spacing, become unstable as the grand matrix becomes ill conditioned at initial root guesses. The red markers result from the calculations with $0.1\lambda$ inter node spacing. It should be noted that the red markers actually overlap the blue markers so that if the number of space harmonics is less than $\sim 20$, then the large inter node spacings give relatively accurate results.

![Figure 8.9. The convergence of the propagation constant as a function of the number of space harmonics. The value $\gamma_n$ is the propagation constant of the singular matrix $G(\gamma_n)$ computed with $n$ space harmonics. Calculations are made for 2 different total node numbers having uniform separation; blue markers represent solutions for 0.2 $\lambda$ node separation, while the red markers are for 0.1 $\lambda$ node separation.](image-url)
Convergence of $\gamma_n$ will be assumed if for 2 different roots obtained with increasing $n$ and $m$ space harmonics, $|\gamma_n - \gamma_m| \to \varepsilon$, where $\varepsilon$ represents machine precision. There are two sets of points illustrated in Fig. 8.9; the lower set represents $|\gamma_{n-2} - \gamma_{n-6}|$, whereas the upper set represents $|\gamma_n - \gamma_{n-4}|$. The number of space harmonics are $n = 9, 13, 17, \ldots$. The first point in the lower set is $|\gamma_7 - \gamma_3|$ while the last point in the set is $|\gamma_{29} - \gamma_{25}|$. These numbers were used because of the formation of the Toeplitz matrix $\hat{K}$. When 3 space harmonics are used to form $\hat{K}$, the matrix has a diagonal and the 1st upper and lower diagonal elements; for 5 space harmonics, there are 2 upper and lower diagonal entries. However, for the 50% duty cycle, $\kappa_2 = \kappa_{-2} = 0$, as illustrated in Table 8.1. Thus, to get proper root differences, two different cases were computed: (1) when the maximum space harmonic produces a zero value on the diagonal and (2) when the maximum space harmonic produces a non-zero value on the diagonal. Thus, the difference $|\gamma_7 - \gamma_3|$ is more meaningful than say $|\gamma_7 - \gamma_5|$.

### 8.3 Grated/Non-Grated Waveguide Interfacing

A waveguide with a grating is usually “etched” in a section of an optical waveguide that has been fabricated as a multi-layer structure. Typically, the waveguides are designed to operate in only the fundamental mode where the 1st high-order modes are beyond cutoff at the designed wavelength. As a result, the optical fields of the grated waveguide have shapes that almost match the optical fields of the non-grated waveguide. This design allows for almost 100% of power transfer between the two structures.

![Figure 8.10. The cross section of a waveguide with periodic structure etched into a regular waveguide. The core region is typically fabricated with many layers.](image)

The grating has a length $L = 2NA$ and has a symmetric profile with respect to the $z = 0$ plane, i.e., $\kappa(x, z) = \kappa(x, -z)$. The input to the grated region is at $z = -N \Lambda$ while the output plane is at $z = N\Lambda$. As indicated in Fig. 8.10, the top of the grating is “level” with the input waveguide.

Input/Output Wave Guide Fields
An input field of a wave propagating in the positive $z$ direction may experience a reflected field from the grating so that the electromagnetic fields in the region $z < -N\Lambda$ may be written as

\begin{align}
E_y &= \phi(x) \left( e^{-\gamma_i(z+N\Lambda)} + \rho e^{\gamma_i(z+N\Lambda)} \right), \\
jk_o Z_o H_x &= -\gamma_i \phi(x) \left( e^{-\gamma_i(z+N\Lambda)} - \rho e^{\gamma_i(z+N\Lambda)} \right),
\end{align}

where $\rho$ is the electric field reflection coefficient and $Z_o$ is the free-space wave impedance. If the output waveguide has a “matched” termination, i.e., there is no wave propagating in the negative $z$ direction, the fields in the output waveguide may be written as

\begin{align}
E_y &= \tau \phi(x) e^{-\gamma_i(z-N\Lambda)}, \\
jk_o Z_o H_x &= -\tau \gamma_i \phi(x) e^{-\gamma_i(z-N\Lambda)},
\end{align}

where $\tau$ is the electric field transmission coefficient. (The equations (8.3.1)-(8.3.4) are approximate field expressions because the radiation modes are not included.)

**Grating Fields**

In the central grating region, the fields are composed of two independent Floquet-Bloch modes $R(x, z) \exp(-\gamma z)$ and $S(x, z) \exp(\gamma z)$ that are propagating in both the positive and negative $z$ directions respectively. These propagating modes satisfy the two differential equations

\begin{align}
\frac{\partial^2 R}{\partial x^2} + \frac{\partial^2 R}{\partial z^2} + \gamma^2 R - 2\gamma \frac{\partial R}{\partial z} + k_o^2 \kappa(x, z) R &= 0, \\
\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial z^2} + \gamma^2 S + 2\gamma \frac{\partial S}{\partial z} + k_o^2 \kappa(x, z) S &= 0.
\end{align}

Because the relative dielectric constant is symmetric about $z = 0$, i.e., $\kappa(x, z) = \kappa(x, -z)$, the relation between the two fields satisfy

\begin{align}
S(x, z) = R(x, -z), \quad \frac{\partial S}{\partial z} = -\frac{\partial R}{\partial z}.
\end{align}

The general solution of the fields in the grating region may be written as a linear combination of forward and backward waves as

\begin{align}
E_y = a R(x, z) e^{-\gamma z} + b S(x, z) e^{+\gamma z},
\end{align}

for the electric field and

\begin{align}
jk_o Z_o H_x = -\gamma a \left( R(x, z) - \frac{1}{\gamma} \frac{\partial R(x, z)}{\partial z} \right) e^{-\gamma z} + \\
&\gamma b \left( S(x, z) + \frac{1}{\gamma} \frac{\partial S(x, z)}{\partial z} \right) e^{+\gamma z},
\end{align}

where $a$ and $b$ are constants.
for the magnetic field. The quantities $a$ and $b$ are arbitrary and are determined from matching the electric and magnetic fields at the input and output planes. Because the dielectric constant is symmetric about $z = 0$, the right function $R(x, z)$, and the left function $S(x, z)$, are periodic and satisfy

$$R(x, z) = \sum_n \psi_n(x) e^{-jnKz}, \quad (8.3.10)$$

$$S(x, z) = R(x, -z) = \sum_n \psi_n(x) e^{jnKz}. \quad (8.3.11)$$

At $z = m\Lambda$, $m = -N, \ldots, N$,

$$R(x, m\Lambda) = R(x, 0) = \sum_n \psi_n(x),$$

$$R_z(x, m\Lambda) = R_z(x, 0) = -jK \sum_n n \psi_n(x),$$

where $R_z(x, 0) = \partial R(x, z)/\partial z|_{z=0}$.

Matching Electromagnetic Fields

The fact that the shape of the electric fields at the input and output planes are identical will be used to determine the effects of the grating region embedded in a dielectric waveguide. The approximate values of the reflection and transmission coefficients may be determined by matching the input and output fields to those of the grating region.

Assume that the $y$-component of the electric field at the interface of the input waveguide and the grating region is $E_y(x)$, and it is normalized, $\langle E | E \rangle = 1$. Furthermore, assume the transverse electric fields on either side of the input plane is given by $E_y[x, z = -(L/2)-] = \mathcal{E}(x)$ and $E_y[x, z = -(L/2)+] = \mathcal{E}(x)$. [4] The electric field at the output plane is $C \mathcal{E}(x)$, where $C$ is a complex constant. \(^\dagger\) Matching the electric field at the output aperture is obtained by (1) equating the output field to the aperture field, and by (2) equating the field in the grating region to the aperture field.\(^\ddagger\)

Thus,

$$C \mathcal{E}(x) \approx \tau \phi(x)$$

$$C \mathcal{E}(x) \approx R(x, 0)(a e^{-N\Lambda\xi} + b e^{N\Lambda\xi}), \quad (8.3.12)$$

where the complex variable $\xi = \alpha + j(\beta - K)$. Assuming the aperture field shape $\mathcal{E}$ is normalized, $\langle \mathcal{E} | \mathcal{E} \rangle = 1$, the transmission coefficient

$$C \approx \tau \langle \mathcal{E} | \phi \rangle \approx \langle \mathcal{E} | R \rangle(a e^{-N\Lambda\xi} + b e^{N\Lambda\xi}).$$

\(^\dagger\) The transverse shapes of the various fields in the grating region are replicated every period, $\Lambda$.

\(^\ddagger\) The input, output and grating fields are approximate because radiation modes are not included in the solutions.
Using (8.3.8) and (8.3.9) one can reduce the boundary conditions to

\[
\tau(\mathcal{E}|\phi) \approx \langle \mathcal{E}|R \rangle (a e^{-N_{A}\zeta} + b e^{N_{A}\zeta}),
\]
(8.3.13)

\[
Y_1 \tau(\mathcal{E}|\phi) \approx Y \left( \langle \mathcal{E}|R \rangle - \frac{1}{\gamma} \langle \mathcal{E}|R_z \rangle \right) (a e^{-\zeta N_{A}} - b e^{\zeta N_{A}}),
\]
(8.3.14)

where the terms \(Y_1 = \gamma_1/(jk_o Z_o)\) is the characteristic admittance of the input/output waveguide and \(Y = \gamma/(jk_o Z_o)\) is the characteristic admittance of the grating region. Dividing (8.3.14) by (8.3.13) yields

\[
\frac{1 + \frac{b}{a} e^{2\zeta N_{A}}}{1 - \frac{b}{a} e^{2\zeta N_{A}}} = \frac{Y_1}{Y_1} \frac{\langle \mathcal{E}|R \rangle - \frac{1}{\gamma} \langle \mathcal{E}|R_z \rangle}{\langle \mathcal{E}|R \rangle},
\]
(8.3.15)

which allows for the computation of \(b/a\).

Matching the electric and magnetic fields at the grating input plane yields

\[
(1 + \rho)(\mathcal{E}|\phi) \approx \langle \mathcal{E}|R \rangle (a e^{N_{A}\zeta} + b e^{-N_{A}\zeta}),
\]
(8.3.16)

\[
Y_1 (1 - \rho)(\mathcal{E}|\phi) \approx Y \left( \langle \mathcal{E}|R \rangle - \frac{1}{\gamma} \langle \mathcal{E}|R_z \rangle \right) (a e^{\zeta N_{A}} - b e^{-\zeta N_{A}}),
\]
(8.3.17)

which allows computation of the input wave admittance, when the output waveguide has a matched load,

\[
Y_{in} = \frac{1 - \rho}{1 + \rho} \frac{Y_1}{Y_1} \frac{\langle \mathcal{E}|R \rangle - \frac{1}{\gamma} \langle \mathcal{E}|R_z \rangle}{\langle \mathcal{E}|R \rangle} \frac{1 - \frac{b}{a} e^{-2N_{A}\zeta}}{1 + \frac{b}{a} e^{-2N_{A}\zeta}}.
\]
(8.3.18)

The input admittance (at \(z = -N_{A} = -L/2\)) to the finite-length grating may be obtained by solving (8.3.15) for the ratio \(b/a\) and substituting into (8.3.18). The quantity

\[
\mathcal{Q} \equiv \frac{Y}{Y_1} \left( 1 - \frac{\langle \mathcal{E}|R_z \rangle}{\gamma \langle \mathcal{E}|R \rangle} \right) = \frac{\gamma_1}{Y_1} \left( 1 - \frac{\langle \mathcal{E}|R_z \rangle}{\gamma \langle \mathcal{E}|R \rangle} \right),
\]
(8.3.19)

is a function of only the grating parameters and the characteristic admittance of the input/output waveguides. For a finite-length grating, the input admittance can be written in terms of \(\mathcal{Q}\) as

\[
Y_{in} = \mathcal{Q} (\mathcal{Q} + 1) - (\mathcal{Q} - 1) e^{-4N_{A}\zeta},
\]
(8.3.20)

and the corresponding field reflection coefficient at the grating input is

\[
\rho = \frac{(1 - \mathcal{Q}^2)(1 - e^{-4N_{A}\zeta})}{(\mathcal{Q} + 1)^2 - (\mathcal{Q} - 1)^2 e^{-4N_{A}\zeta}}.
\]
(8.3.21)
while the field transmission coefficient at the output of the grating is

\[
\tau = \frac{4\Omega e^{-2N\Lambda\xi}}{(\Omega + 1)^2 - (\Omega - 1)^2 e^{-4N\Lambda\xi}}.
\] (8.3.22)

The field coefficients of the forward and backward waves are obtained from

\[
a \langle \mathcal{E} | R \rangle = \frac{2(\Omega + 1)e^{-\xi N\Lambda}}{(\Omega + 1)^2 - (\Omega - 1)^2 e^{-4\xi N\Lambda}}
\] (8.3.23)

\[
b \langle \mathcal{E} | R \rangle = \frac{2(\Omega - 1)e^{-\xi N\Lambda}}{(\Omega + 1)^2 - (\Omega - 1)^2 e^{-4\xi N\Lambda}}
\] (8.3.24)

The calculation of the aperture field \( \mathcal{E}(x) \) is generally complicated. On the other hand, estimates of the input admittance, to the grating region, may be relatively accurate by approximating the electric field, \( \mathcal{E}(x) \). Furthermore, the first order evaluation of the aperture field, \( \mathcal{E}(x) \), leads to the second order evaluation of the input admittance and the corresponding field reflection coefficient. This “good approximation” occurs at the 2nd Bragg condition because the input admittance is stationary, with respect to small changes in \( \mathcal{E}(x) \), and the resulting expression for \( Y_{\text{in}} \) will be accurate to second order. That is, if \( \mathcal{E}(x) = \phi(x) + \xi \psi(x) \), where \( \xi \) is a small constant, then the input admittance

\[
Y_{\text{in}}(\xi) = Y_{\text{in}}(\xi)|_{\xi=0} + O(\xi^2).
\] (8.3.25)

In addition to \( Y_{\text{in}} \), being stationary, \( \Omega \), \( \rho \) and \( \tau \) are stationary at the 2nd Bragg condition.

### 8.4 Properties Grating Etched in a Laser Waveguide

This section will analyze the characteristics of a laser structure designed to operate at \( \lambda = 0.780 \mu m \) with a reflective grating etched into the waveguide. The grating period \( \Lambda = 0.233 \mu m \) is designed for resonance at the 2nd Bragg condition. The structure’s geometry is given in Table 8.3. There are six parameters that characterize each layer; however, additional parameters are required to specify the grating layers. The material in the grating regions, specified in the table represent the material that forms the “tooth regions.” The material of the “fill regions” are specified in a different table.

The structure given in Table 8.3 has 4 grating layers with the top grating layer an oxide layer whose refractive index is 2.04 with a thickness of 0.07 \( \mu m \).
Table 8.3. A laser structure formed with 16 layers including 4 grating layers. The first column shows the layer number, \( n \) is the refractive index of the layer, \( \alpha \) is the wave absorption coefficient of the material (/cm), \( t \) is the layer thickness (\( \mu \)m), \( G \) is a boolean variable, T (true) or a grating layer, F (false) otherwise, \( N \) identifies a node layer, and the last column describes the layer. (The node point of a layer occurs at the bottom of the layer.)

<table>
<thead>
<tr>
<th>( n )</th>
<th>Refractive Index ( n )</th>
<th>Wave Abs. ( \alpha )</th>
<th>Layer Thickness ( t )</th>
<th>( G )</th>
<th>( N )</th>
<th>Layer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0</td>
<td>( \infty )</td>
<td>F</td>
<td>T</td>
<td>Air</td>
</tr>
<tr>
<td>2</td>
<td>2.01</td>
<td>0</td>
<td>0.07</td>
<td>T</td>
<td>T</td>
<td>Grating Layer 1</td>
</tr>
<tr>
<td>3</td>
<td>3.3179</td>
<td>0</td>
<td>0.0825</td>
<td>T</td>
<td>T</td>
<td>Grating Layer 2</td>
</tr>
<tr>
<td>4</td>
<td>3.3179</td>
<td>0</td>
<td>0.0825</td>
<td>T</td>
<td>T</td>
<td>Grating Layer 3</td>
</tr>
<tr>
<td>5</td>
<td>3.3179</td>
<td>0</td>
<td>0.05</td>
<td>T</td>
<td>T</td>
<td>Grating Layer 4</td>
</tr>
<tr>
<td>6</td>
<td>3.3179</td>
<td>0</td>
<td>0.125</td>
<td>F</td>
<td>T</td>
<td>P-Clad Layer</td>
</tr>
<tr>
<td>7</td>
<td>3.3517</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 1</td>
</tr>
<tr>
<td>8</td>
<td>3.3855</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>T</td>
<td>GRIN Layer 2</td>
</tr>
<tr>
<td>9</td>
<td>3.4193</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 3</td>
</tr>
<tr>
<td>10</td>
<td>3.4531</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 4</td>
</tr>
<tr>
<td>11</td>
<td>3.6301</td>
<td>0</td>
<td>0.008</td>
<td>F</td>
<td>T</td>
<td>Quantum Well</td>
</tr>
<tr>
<td>12</td>
<td>3.4531</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 5</td>
</tr>
<tr>
<td>13</td>
<td>3.4193</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 6</td>
</tr>
<tr>
<td>14</td>
<td>3.3855</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>F</td>
<td>GRIN Layer 7</td>
</tr>
<tr>
<td>15</td>
<td>3.3517</td>
<td>0</td>
<td>0.025</td>
<td>F</td>
<td>T</td>
<td>GRIN Layer 8</td>
</tr>
<tr>
<td>16</td>
<td>3.3179</td>
<td>0</td>
<td>( \infty )</td>
<td></td>
<td></td>
<td>Substrate</td>
</tr>
</tbody>
</table>

Figure 8.5 illustrates the parameters used to describe the grating region which is composed of many layers.

![Geometry of a grating layer](image)

Figure 8.11. The geometry of a grating layer of thickness \( t \) and several “fill” regions that are used to shape the grating. The figure illustrates only half of the grating region. For a symmetric (with respect to \( z \)), the grating off-set \( d_{os} = 0 \).

The off-set of the center of the “tooth region”, \( d_{os} \) is measured relative to the bottom tooth layer which starts at \( x = 0 \). For symmetric gratings, the tooth off-set is 0. (Typical symmetric gratings examples include sinusoidal, rectangular, trapizoid, and triangular gratings. For structures with gratings covered with a protective surface such as silicon dioxide, the “tooth region” of
the top grating layer would be oxide material. The second layer would have
the first “fill region” composed of the oxide; the value $L_2$ represents the oxide
deposition on the “sidewalls.”

Table 8.4 lists the “fill material” for each of the grating layers.

Table 8.4. Grating layer properties are listed for each grating layer.
In the example shown, the tooth material parameters are given in Ta-
ble 8.3 while the material parameters of the “fill material.” † The last
fill layer $L_f$ is calculated using the grating period $A$.

<table>
<thead>
<tr>
<th>Grating Layer</th>
<th>Layer</th>
<th>Fill Regions</th>
<th>Fill Width</th>
<th>$d_{ox}$</th>
<th>Fill $n$</th>
<th>Fill $\alpha$</th>
<th>Fill $L_f$ ($\mu$m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Fill Region</td>
<td>1</td>
<td>2</td>
<td>0.5A</td>
<td>0</td>
<td>2.01</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>2nd Fill Region</td>
<td>2</td>
<td>3</td>
<td>0.5A</td>
<td>0</td>
<td>1.00</td>
<td>0</td>
<td>†</td>
</tr>
<tr>
<td>3rd Fill Region</td>
<td>3</td>
<td>4</td>
<td>0.5A</td>
<td>0</td>
<td>2.01</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>4th Fill Region</td>
<td>4</td>
<td>5</td>
<td>0.5A</td>
<td>0</td>
<td>1.00</td>
<td>0</td>
<td>†</td>
</tr>
</tbody>
</table>

The tooth geometry as illustrated in Table 8.4 is shown in ‘perf.10’.
The 1st grating layer has oxide in the tooth region, while layers 2, 3 and 4 are
formed from the semiconductor p-clad material that has been etched. The ox-
ide is deposited over the etched grating with some being attached to the side of
the etched tooth. The tooth material is given in Table 8.3, but the tooth width
and fill material are given in Table 8.4. For square teeth, grating layers 2 and 3
are identical, however, for a trapezoidal tooth, the width of all “tooth regions”
will be different. Refining the teeth shape can be accomplished by adding ad-
ditional layers.

Figure 8.12. The tooth geometry given in Table 8.4 is formed with
semiconductor material, oxide, and air.
Propagation Characteristics Near 2nd Bragg

The propagation characteristics of the waveguide structure shown in Table 8.3 and 8.4 exhibit 4 different modes. The four modes are split into 2 forward propagation modes and into 2 backward propagation modes. The mode attenuation characteristics are shown in Fig. 8.13 (a), while the “ω − β” are shown in Fig. 8.13 (b).

Figure 8.13. The propagation characteristics of the waveguide structure given in Tables 8.3 and 11. At the 2nd Bragg the grating period is Λ = 0.2336875 μm, the effective indices of all modes is ne = 3.33779, and wave length λ = 0.78 μm.

All four modes have the same complex propagation constant, γ, at the 2nd Bragg condition, i.e., (8.2.44) has a 4th order zero. The duplicity of roots occurs because the four modes meet at the same point, however, at points away from the 2nd Bragg, all four modes are distinct. For example, the forward waves, Fig. 8.13(b), have different attenuation constants, Fig. 8.13(a); the complex propagation constants of the two modes are (αi + jβR) and (αii + jβR).

At the 2nd Bragg condition ζ = α + j(β − K) = 0, the n = −1 space harmonic vanishes while the space harmonics for n > −1 pair off with the space harmonics n < −1.
Figure 8.14. Various space harmonics at the 2nd Bragg condition. (The imaginary parts of the field expressions are zero.) Space harmonics, \( n = 0 \) and \( -2 \), are shown in the \((0,-2)\) frame. The \((0,-2)\) frame also shows the dielectric profile of the input waveguide as well as the nodes (dots), used in the computation. The grating starts at \( x = 0 \), and extends over layers 2, 3, 4 and 5 with a tooth height of 0.285 \( \mu m \).

Figure 8.14 shows only the real parts of the various space harmonics since their imaginary parts are zero. (There are no material losses as indicated in Table 8.4.) Nodes illustrated by the dots in frame \((0,-2)\), show there are 5 nodes in the grating region and 4 nodes outside the grating.

The transverse fields are real at the 2nd Bragg because the attenuation coefficient \( \alpha = 0 \). Furthermore, the overlap function

\[
\langle \mathcal{E} | R \rangle = \sum_{n=1}^{H} \langle \phi | \psi_n \rangle, \tag{8.4.1}
\]

is obtained by approximating \( \mathcal{E}(x) \) with the input/output waveguide field \( \phi(x) \). The 0 and \(-2\) space harmonics dominate at and near the 2nd Bragg, however, other space harmonics are non zero and must be included in the analysis. Typical “coupled-mode” analysis omit all space harmonics except 0 and \(-2\). A more important observations is the pairing of all space harmonics, excluding the \( n = -1 \). Because of the pairing of the space harmonics, \( \langle \phi | R \rangle = 0 \),
which implies that $Q$ has a simple pole at the 2nd Bragg condition. If the simple pole did not occur at the 2nd Bragg condition, the reflection coefficient, $(8.3.21)$, would be zero.

To investigate the reflection coefficient near the 2nd Bragg condition, the transformation $R(\xi) = 1/Q(\xi)$ has $R \to 0$ as $\xi \to 0$. The reflection coefficient becomes

$$ \rho(\xi) = \frac{(R^2 - 1)(1 - e^{-4NA\xi})}{(R + 1)^2 - (R - 1)^2 e^{-4N\Lambda\xi}} = |\rho|e^{i\theta} \quad (8.4.2) $$

Using L'Hopital's rule as $\xi \to 0$, gives

$$ \rho(0) = -\frac{NA}{R_x + NA} = -\frac{L}{2R_x + L}, \quad (8.4.3) $$

where $R_x = dR/d\xi$ and $L$ is the grating length. Assuming $R_x$ is finite at the 2nd Bragg, $\rho = 1$ for a grating of infinite length.

The real, $\Re\{R\}$, and imaginary, $\Im\{R\}$, of $R$ are shown in Fig. 8.15 (a) and the derivatives are shown in (b). It should be noted that abscissas of the 2 figures are plotted relative to the detuned wavelength $\delta \lambda$ while the derivatives of $R$ are taken with respect to $\xi$. Both the real and imaginary parts of $R$ are 0 at the 2nd Bragg, while the derivative $R_x$ is continuous and has a non-zero value. For the structure shown in Tables 8.3 and 8.4, $R_x(0) \approx 41.99 + j119.50$. The phase angle of the reflection coefficient becomes

$$ \theta = -\arctan\left(\frac{R''}{R_x' + NA}\right) \quad (8.4.4) $$

Figure 8.15. (a) The real and imaginary parts of the function $Q = 1/R$ where $Q$ is given by $(8.3.19)$, and (b) $dR/d\xi$.

For a grating length of 1000 $\Lambda$, the reflection coefficient computed from $(8.4.3)$ is $\rho(0) \approx -0.471 + j0.353$. The power reflection coefficient $R \approx 0.347$ and the angle $\theta = 0.795\pi$.

As noted earlier the reflection coefficient $\rho \neq 0$ at the 2nd Bragg even though $\xi = 0$. It is the pairing of the space harmonics that produces a simple pole of $Q$ at the 2nd Bragg so that the product $\xi Q$ is finite and is a well-behaved complex function of the complex variable $\xi$. The dependent variable $R = 1/Q$ can be expanded about the complex zero, giving

$$ \frac{1}{Q(\xi)} = R(\xi) = R_x(0)\xi + \cdots \quad (8.4.5) $$

can be used for the evaluation of $\rho$ at the 2nd Bragg and it indicates that its behavior in the vicinity of the the 2nd Bragg is tied to the differential behavior of $R = 1/Q$. 
Writing the dependent variable in terms of its real and imaginary parts
\( \zeta = \xi + j \eta, \mathcal{R} = U(\xi, \eta) + j V(\xi, \eta), \) and employing the Cauchy-Riemann equations to \( \mathcal{R}, \)
\[
\frac{\partial U}{\partial \xi} = \frac{\partial V}{\partial \eta}, \quad \frac{\partial U}{\partial \eta} = -\frac{\partial V}{\partial \xi},
\]
the partial derivatives \( U_\xi = \frac{\partial U}{\partial \xi}, U_\eta, V_\xi, \) and \( V_\eta \) may be computed. Figure 8.16 shows only the derivatives of the real part of \( \mathcal{R}. \) At \( \zeta = 0, U_\xi = 41.99 \) while \( U_\eta = -119.50 \) The grating length for a given power reflection coefficient at the 2nd Bragg, can be found using (8.4.3),
\[
L = \frac{2}{1 - R} \left( R U_\xi + \sqrt{R^2 U_\xi^2 + R(1 - R) U_\eta^2} \right) \tag{8.4.6}
\]

Reflection and Transmission Coefficients

The reflection and transmission coefficients and radiation losses will be discussed for the structure given in Table 8.3.
Chapter 9
Numerical Solution of Layered Structures

The solution of the wave equation for layered media has been extensively investigated. In Chapter 4 the characteristic equation governing the behavior of trapped waveguide modes was developed. Generally, the trapped modes are found from a single equation; the trapped modes are identified from the roots of the equation. In this chapter, the numerical methods for finding the roots will be discussed.

9.1 Branch Cuts

The trapped modes are typically found from the solution of a single characteristic equation such as that given in Eq. 4.4.22. For the simple three-layer structure, the characteristic equation can be written as a function of the thickness of the central layer, \( d = d_1 \), and the three transverse wave numbers, \( p \), \( q \), and \( r \). Further, the transverse wave numbers can be expressed in terms of the modal propagation constant \( \gamma \), as shown in Eq. 4.4.24. (Note that the original definitions of the transverse wave numbers were defined as \( h_0 \) and \( h_2 \) in ‘wave’.) Accordingly, the roots of the equation

\[
F(\gamma) = \tan q(\gamma)d - \frac{q(\gamma)(p(\gamma) + r(\gamma))}{q^2(\gamma) - p(\gamma)r(\gamma)}.
\]

(9.1.1)

The secular equation can be written in a more numerically friendly form by eliminating the pole at \( qd = \pi/2 \),

\[
F(\gamma) = (q^2 - pr) \sin(qd) - q(p + r) \cos(qd)
\]

(9.1.2)
Equation (9.1.2) is a transcendental equation that has multiple roots. If conditions are placed on the values of the transverse wave numbers, \( p \) and \( r \), the modes governed by the roots of Eq. (9.1.2), may be either trapped, where the electromagnetic field of the mode is confined to the slab region, or leaky, where the electromagnetic field may exponentially grow on either side of the central slab, i.e., the field grows without bound as \( x \to -\infty \), the field grows without bound as \( x \to \infty \), or both. A condition required for trapped modes is \( \text{Re}\{p\}, \text{Re}\{r\} > 0 \), where the values of \( p \) and \( r \) are calculated using Eq. 4.4.24.

For a complex propagation constant, \( \gamma = \alpha + j\beta \), the wave numbers \( p \) and \( q \) satisfy

\[
\begin{align*}
p &= \pm \sqrt{-(\gamma^2 + k^2 \kappa_0)}, \\
r &= \pm \sqrt{-(\gamma^2 + k^2 \kappa_2)}.
\end{align*}
\]

(9.1.3)

This condition is satisfied if the complex values of \( p \) and \( r \) are restricted to the right-half-region of the complex plane, as shown in Fig. 9.1.

With the branch cut along the negative real axis, the values of \( p \) and \( r \) will lie in the right half of the plane, while their values will lie in the left-half of the plane when the negative sign is chosen. It should be noted that the wave equation is satisfied if either sign is chosen. The condition on the field behavior as \( x \to \pm \infty \) determines which sign is chosen. If only the positive sign is used for the determination of \( p \) and \( r \), only trapped will exist as solutions of Eq. (9.1.2). Similarly, if only the negative sign is used for the determination \( p \) and \( r \), only leaky modes will exist as solutions of Eq. (9.1.2). If the mode leaks out Region 0 and not Region 2, say, then the negative sign is taken for \( p \) while the positive sign is taken for \( r \). (In the numerical evaluation of \( q \), the sign is not an issue because both solutions are used in the evaluation of the field in Region 1.)

On the other hand, if the wave numbers \( \hat{h}_0 \) and \( \hat{h}_2 \) were calculated as

\[
\begin{align*}
\hat{h}_0 &= \pm \sqrt{\gamma^2 + k^2 \kappa_0}, \\
\hat{h}_2 &= \pm \sqrt{\gamma^2 + k^2 \kappa_2},
\end{align*}
\]

and the branch cut is taken along the negative real axis, then the positive sign represents waves propagating outward while the negative sign represents waves propagating inward.

9.2 Mode Solutions

The roots of the characteristic equation as given by Eq. 9.1.2 or its equivalent form as given by Eq. 4.4.20 which is expressed in terms of the transfer matrix,
governs the modes of the three-layer waveguide. The characteristic equation of a waveguide structure that has \( N \) layers is given by

\[
F(\gamma) = j \bar{h}_0 (A + j \bar{h}_{N-1} B) + \bar{C} + j \bar{h}_{N-1} D
\tag{9.2.1}
\]

where \( \bar{h}_0 \) and \( \bar{h}_{N-1} \) are the usual transverse wavenumbers of the superstrate and substrate regions respectively. The matrix elements of the multi-layer transformation are \( A, B, C \) and \( D \). The multi-layer transformation matrix is a product of \( N - 2 \) single-layer transformations.

There are limited techniques for computing roots of a complex transcendental function and all methods generally follow an iteration approach. Mueller’s method is a popular algorithm used for finding a single root of the transcendental equation. The algorithm is started by supplying three complex numbers that are close to the real root. The process is usually fast and will yield roots in less than ten iterations. However, if there are multiple roots near the guess point, convergence will be difficult.

The Newton algorithm is a much better root searching technique, however, it requires the computation of a derivative. For example, Eq. 9.2.1 is expanded as

\[
F(\gamma) = F(\gamma_0) + \frac{dF(\gamma_0)}{d\gamma} (\gamma - \gamma_0),
\tag{9.2.2}
\]

where \( dF/d\gamma \) is the derivative of \( F \) with respect to \( \gamma \) evaluated at \( \gamma = \gamma_0 \). The difficulty here is that \( F(\gamma) \) is evaluated as a product of transfer matrices so that there are many products of transcendental function. Thus, Mueller’s technique appears to be a much simpler approach.

If the characteristic equation is expressed in slightly different form, it will be possible to develop a matrix equivalent of the Newton method. First write the solution of the wave equation for the individual layers. Assume the permittivity of the individual layers is constant, so that the field solutions may be written as

\[
\psi_0(x) = Q_0 \cos \bar{h}_0(x - x_0) + P_0 \sin \bar{h}_0(x - x_0)/\bar{h}_0,
\]
\[
\psi_1(x) = Q_1 \cos \bar{h}_1(x - x_1) + P_1 \sin \bar{h}_1(x - x_1)/\bar{h}_1,
\]
\[
\psi_2(x) = Q_2 \cos \bar{h}_2(x - x_2) + P_2 \sin \bar{h}_2(x - x_2)/\bar{h}_2,
\]
\[
\vdots
\]
\[
\psi_{N-1}(x) = Q_{N-1} \cos \bar{h}_{N-1}(x - x_{N-2}) + P_{N-1} \sin \bar{h}_{N-1}(x - x_{N-2})
\tag{9.2.3}
\]

where \( Q_n \) and \( P_n \) are the field and flux values at \( x = x_n \) in the nth layer. (Note that that superstrate and substrate layers are also denoted by “S” and “s” respectively.) Because of the radiation condition of the bounded modes, the superstrate coefficients satisfy

\[
Q_0 + \frac{P_0}{\bar{h}_0} = 0.
\tag{9.2.4}
\]
while the substrate coefficients satisfy
\[
Q_{N-1} - \frac{P_{N-1}}{j h_{N-1}} = 0. \tag{9.2.5}
\]

By matching the field and flux values from layer to layer, there will be a set of \(2N\) linear equations in terms of the canonical coefficients \(Q_n\) and \(P_n\),

\[
\begin{pmatrix}
  j \hat{h}_0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
  -1 & 0 & c_1 & s_1/\hat{h}_1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
  0 & -1 & -\hat{h}_1 s_1 & c_1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
  0 & 0 & 0 & 0 & 0 & \cdots & -1 & 0 & 1 & 0 \\
  0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 0 & 1 \\
  0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -j \hat{h}_{N-1} & 1 \\
\end{pmatrix}
\begin{pmatrix}
  Q_0 \\
  P_0 \\
  Q_1 \\
  P_1 \\
  \vdots \\
  Q_{N-1} \\
  P_{N-1}
\end{pmatrix} = 0, \tag{9.2.6}
\]

where \(c_n\) and \(s_n\) are abbreviations for \(\cos \hat{h}_n d_n\) and \(\sin \hat{h}_n d_n\) respectively. In the present configuration, the lower diagonal elements are \(-1\). Now the value of \(\gamma\) must be chosen such that the determinant of the coefficient matrix is zero. More specifically, the matrix will be singular when its reciprocal condition factor is zero[5]. Denote \(M\) as the coefficient matrix and \(v\) as the vector of field and flux coefficients.

It is instructive to write the matrix \(M\) in terms of the transfer matrices of the individual layers,

\[
M(\gamma) = \begin{pmatrix}
  (j \hat{h}_0 & 1) & 0 & 0 & \cdots & 0_v & 0_v \\
  -I & (A_1 & B_1) & \cdots & 0 & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  0 & 0 & \cdots & -I & I \\
  0_v & 0_v & \cdots & 0_v & (-j \hat{h}_{N-1} & 1)
\end{pmatrix}, \tag{9.2.7}
\]

where \(0\) is a \(2\times 2\) matrix whose elements are zero, \(0_v\) is a \(1\times 2\) vector whose elements are zero, and \(I\) is the \(2\times 2\) unit matrix. It is seen that the transfer matrices of the individual layers are positioned near the diagonal. Specifically, the element \(A_n\) of the \(n\)th layer is \(M_{2n,2n+1}\), \(M_{2n,2n+2} = B_n\), \(M_{2n+1,2n+1} = C_n\), and \(M_{2n+1,2n+2} = D_n\).

The propagation constant, \(\gamma_v\), for each bounded mode satisfies

\[
\det M(\gamma_v) = 0. \tag{9.2.8}
\]

Since (9.2.8) is a transcendental equation, the roots must be calculated numerically. The Newton-Matrix method will be developed for the numerical process, which uses an initial guess, and employs the matrix derivatives to converge to the actual root. The process is developed by expanding \(M\) in a Taylor series about \(\gamma_0\), an initial value that is near the actual root

\[
M(\gamma) = M(\gamma_0) + M'(\gamma_0) \delta + \cdots = 0, \tag{9.2.9}
\]
where $\delta = \gamma - \gamma_0$. Using only the first two terms of the expansion, a bounded mode requires

$$\det M(\gamma) = \det[M_0 + M_1 \delta] = 0,$$  \hspace{1cm} (9.2.10)

where, $M(\gamma_0) = M_0$ while $M'(\gamma_0) = M_1$. If $\gamma_0$ is the propagation constant of one of the waveguide modes, then the matrix $M_0$ is singular then $\delta = 0$. On the other hand, if $\gamma_0$ is not the value of a bounded mode, $M_0$ is non-singular and it has an inverse so that (9.2.10) is equivalent to the eigenvalue problem

$$M_0^{-1}M_1v = \chi v,$$  \hspace{1cm} (9.2.11)

where the eigenvalues are roots of

$$\det[M_0^{-1}M_1 - \chi I].$$  \hspace{1cm} (9.2.12)

The eigenvalue $\chi = -1/\delta$. Accordingly, the smallest Newton step from the approximate root $\gamma_0$ is the negative reciprocal of the largest eigenvalue of $M_0^{-1}M_1$. The next step in the Newton search is given by

$$\gamma = \gamma_0 - 1/\chi_{\text{max}},$$  \hspace{1cm} (9.2.13)

where $\chi_{\text{max}}$ is the largest eigenvalue of (9.2.11). The eigenvector corresponding to the smallest eigenvalue of $M$ represents the field and flux values at the layer boundaries. Note that it is not recommended that $M_0^{-1}$ be computed, but rather, one should solve a set of linear equations

$$M_0A = M_1,$$

for the square matrix $A$. Many linear algebra algorithms will calculate both $A$ and reciprocal condition factor for $M_0$.

The $M_1$ is the derivative of $M$ with respect to $\gamma$ and is given by

$$M_1 = \begin{pmatrix}
(j \hat{h}_0 & 1) & 0_v & \ldots & 0_v & 0_v \\
0 & (A_1 & B_1) & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0_v & 0_v & \ldots & 0_v & (-j \hat{h}_{N-1} & 1)'
\end{pmatrix},$$  \hspace{1cm} (9.2.14)

where the element derivatives include $\hat{h}_n' = \gamma/\hat{h}_n$. The derivative of the elements of the transfer matrix is

$$\left(\begin{array}{c}
A_n & B_n \\
C_n & D_n
\end{array}\right)' = \left(\begin{array}{cc}
dA_n/d\gamma & dB_n/d\gamma \\
dC_n/d\gamma & dD_n/d\gamma
\end{array}\right)_{\gamma=\gamma_0}.$$  \hspace{1cm} (9.2.15)

Finally, upon convergence of the propagation constant, the eigenvector of (9.2.11) gives the field and flux values at the beginning of each layer.
Appendix A
Method of Stationary Phase

Frequently it is advantageous to obtain an approximation of integrals that cannot be evaluated in closed form. Although integrals may be evaluated using numerical techniques, approximate integration methods are useful for understanding characteristics associated with physical attributes dependent on the integrals. Fourier analysis of electromagnetic fields frequently uses integral expressions of the various field components.

In particular, an integral such as

\[ I = \int_{a}^{b} f(x)e^{\lambda g(x)} \, dx, \]  

(A.1)

is difficult to evaluate in closed form if the function \( g(x) \) is different from a second-order polynomial is \( x \), the integration interval is finite and \( f(x) \) is not a constant.

An approximate value of (A.1) can be obtained when the parameter \( \lambda \gg 1 \). The major contribution to the integral's value occurs in the neighborhood of the points where the phase function \( g(x) \) is stationary, the points where \( g'(x) = 0 \). For example, if the function \( f(x) = 1 \) and \( g(x) = x^2 \), the stationary point occurs at \( x = 0 \). The real part of the integrand of (A.1), for \( \lambda = 1 \), becomes \( y(x) = \cos x^2 \) and is shown in Fig. 9.1. As seen from the figure, the function is highly oscillating as \( x \) becomes large, while the function varies slowly at the stationary point at the origin.

![Figure A.1. The function \( y(x) = \cos x^2 \) has a stationary phase point at the origin, \( x = 0 \).](image-url)
The approximation of the integral in (A.1) can be accomplished by: (1) approximating the function \( g(x) \) with a three-term Taylor series and (2) extending the interval of integration from \((a, b)\) to \((-\infty, \infty)\). The stationary point must occur within the \((a, b)\) interval. In regard to the latter condition, interval extensions must be practical in the sense that the stationary point must not be close the integral limits, although, the stationary point may occur on the boundary. Figure 9.2 shows the integral of the function shown in Fig. 9.1 where the stationary point is at \( x = 0 \).

![Figure A.2. The integral of the function \( y(u) = \cos u^2 \) over the interval \((0, x)\).](image)

The exact value of the integral over the interval \((0, \infty)\) is indicated by the horizontal line, so that when the integral over \((0, x)\) is approximated by the integral over \((0, \infty)\), the actual value oscillates about the approximate answer.

For large values of \( \lambda \), contributions to the integral in (A.1) occur only over a small interval about the stationary point. Therefore, \( f(x) \) can be approximated with its value at the stationary point. To evaluate the integral's approximate value, first expand \( g(x) \),

\[
g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2, \]

\[
= g_0 + g_1(x - x_0) + g_2(x - x_0)^2. \tag{A.2}
\]

Assuming \( x_0 \) is the only stationary point, the integral becomes

\[
I \approx f(x_0)e^{i\lambda g(x_0)} \int_{-\infty}^{\infty} e^{i\lambda g' u^2} du, \tag{A.3}
\]

where the integration variable has been changed, \( x \to u \). The integral in Eq. (A.3) exists in closed form, and the and the approximate value for (A.1) is

\[
I \approx \sqrt{\frac{j2\pi}{\lambda g''(x_0)}} f(x_0) e^{i\lambda g(x_0)} \tag{A.4}
\]

The above one-dimensional approximation can be extended to include two-dimensional (2D) integrals that are useful to evaluate 2D Fourier transforms and their inverses. The 2D integral that is equivalent to (A.1) is

\[
J = \int_{a}^{b} \int_{c}^{d} f(x, y)e^{i\lambda g(x, y)} dx dy. \tag{A.5}
\]
The argument of the exponent must be expanded by a Taylor series in two dimensions as

\[
g(x, y) \approx g(x_0, y_0) + \frac{1}{2} g_{xx} (x-x_0)^2 + g_{xy} (x-x_0)(y-y_0) + \frac{1}{2} g_{yy} (y-y_0)^2, \quad (A.6)
\]

where \( g_{xx} = \frac{\partial^2 g(x, y)}{\partial x^2} \) is the second partial derivative of \( g(x, y) \) evaluated at \((x_0, y_0)\). The other terms \( g_{xy} \) and \( g_{yy} \) are similarly defined. The first–order derivatives were not included in the expansion because at the stationary point is located at \((x_0, y_0)\), \(i.e.,\) the stationary points are defined by

\[
g_x(x_0, y_0) = 0; \quad g_y(x_0, y_0) = 0.
\]

When the integration limits are extended to the interval \((-\infty, \infty)\), and integration variables changed to \( u = x - x_0 \) and \( v = y - y_0 \), the approximate integral becomes

\[
J \approx f(x_0, y_0) e^{j\lambda g(x_0, y_0)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\lambda/2)(g_{xx} u^2 + 2g_{xy} uv + g_{yy} v^2)} \, du \, dv \quad (A.7)
\]

One approach is to notice that the argument of the integrand exponent is a standard quadratic form which can be written as

\[
g(u, v) = \begin{pmatrix} u \\ v \end{pmatrix} \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}
\]

The integral in (A.7) can be evaluated without much work by using the properties of the quadratic form. The two dimensional matrix can be diagonalized by finding its eigenvalues and corresponding eigenvectors. The eigenvectors can be combined to form a linear transformation of the \((u, v) \to (\xi, \eta)\) variables. The eigenvectors play no role in the final answer, other than the fact that the transformation Jacobian is unity. The resulting transformation produces an expression that will have no cross–products of the integration variables, \(\xi\) and \(\eta\) and the resulting 2D integral is the product of two 1D integrals. The eigenvalues are

\[
v_1, v_2 = \frac{1}{2} \left( g_{xx} + g_{yy} \pm \sqrt{g_{xx}^2 + 4g_{xy}^2 - 2g_{xx}g_{yy} + g_{yy}^2} \right),
\]

and the integral in (A.7) becomes

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\lambda/2)(g_{xx} u^2 + 2g_{xy} uv + g_{yy} v^2)} \, du \, dv = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\lambda/2)(v_1 \xi^2 + v_2 \eta^2)} \, d\xi \, d\eta = \frac{2\pi}{\lambda \sqrt{v_1 v_2}}.
\]

The approximate value of the 2D integral becomes

\[
J \approx \frac{j 4\pi}{\lambda \sqrt{(g_{xx}g_{yy} - g_{xy}^2)}} f(x_0, y_0) e^{j\lambda g(x_0, y_0)}. \quad (A.8)
\]
Mode Orthonormalization

Orthogonality of the radiation modes in the power sense can be realized when the dielectric constant is real, or equivalently when the dielectric layers are lossless. The characteristic modes of the slab waveguide are solutions to the differential equation

\[ \frac{d^2\psi(h, x)}{dx^2} + k_0^2[k(x) - \kappa_s]\psi(h, x) = -h^2\psi(h, x), \tag{B.1} \]

where \( k(x) \) is the relative dielectric profile and \( \kappa_s \) is the dielectric constant of the substrate. The quantity \(-h^2\), assumed to be real, is the eigenvalue of the operator

\[ L_{op} = \frac{d^2}{dx^2} + k_0^2[k(x) - \kappa_s]. \]

while \( \psi(h, x) \) is the corresponding eigenfunction. For the present case, assume the operator is Hermitian,

\[ L_{op} = L_{op}^*, \]

which implies that the dielectric constant is real.

Assuming a different eigenvalue, \( \tilde{h}^2 \), and corresponding eigenfunction, \( \psi(\tilde{h}, x) \), satisfies a differential equation similar to (B.1). The complex conjugate of that equation is

\[ \frac{d^2\psi^*(\tilde{h}, x)}{dx^2} + k_0^2[k(x) - \kappa_s]\psi^*(\tilde{h}, x) = -\tilde{h}^2\psi^*(\tilde{h}, x). \tag{B.2} \]

Now multiply (B.1) by \( \psi^*(\tilde{h}, x) \) and (B.2) by \( \psi(h, x) \), substract the two equations and integrate the results from \( x = x_s - X \) to \( x = x_S + X \), where \( x_s \) and \( x_S \) represent the substrate and superstrate layer locations bordering the central waveguide region, gives the following

\[ (h^2 - \tilde{h}^2) \langle \psi(h, x), \psi(h, x) \rangle_X = \]

\[ \psi_S^*(\tilde{h}, X) \psi_S(h, X) - \psi_S^*(\tilde{h}, X) \psi_S(h, X) \]

\[ -\psi_S^*(\tilde{h}, -X) \psi_S(h, -X) + \psi_S^*(\tilde{h}, -X) \psi_S(h, -X), \tag{B.3} \]

where

\[ \langle \psi(h, x), \psi(h, x) \rangle_X = \int_{x_s - X}^{x_s + X} \psi^*(h, x) \psi(h, x) \, dx. \]
The subscripts $s$ and $S$ are used in (B.3) to indicate that the wave functions are those in the substrate and superstrate respectively. In the limit $X \to \infty$, orthogonality occurs if

$$\int_{-\infty}^{\infty} \psi^*(\tilde{h}, x) \psi(h, x) \, dx = \delta(h - \tilde{h}),$$

which also implies that the modes are normalized to unity.
Appendix B
Overlap Integrals

The scattering of modes at a waveguide discontinuity such as that illustrated in Fig. 6.1 is addressed here. To simplify the discussion, assume that only one layer is discontinuous while the remaining layers are continuous across the discontinuity. This configuration of layers commonly occurs in photonic devices such as lasers and associated optical waveguides because these devices are fabricated from epitaxial wafers that have several layers grown on a common substrate. Although Fig. 6.1 illustrates Layer 1 with a discontinuity, it is possible to have a different layer, of a multi-layer structure, that is discontinuous. If a layer is grown above the discontinuous or etched layer, the new material acts as a “fill” layer. Subsequent layers may then be grown. Thus, a single layer, say Layer \( \ell \), will be designated as discontinuous while all remaining layers will be continuous.

The overlap integral of the modes of one waveguide with those of a different waveguide such as that of Fig. 6.1 govern the scattering processes occurring at the interface of the two waveguides. These overlap integrals can be computed by several methods. The most obvious method is to calculate the modes and evaluate the integrals. The approach used here is to use the differential equations that govern the modes on the left, called \( \psi \) modes, and the modes on the right, called the \( \phi \) modes. The \( \psi \) modes satisfy

\[
\frac{d^2 \psi}{dx^2} + k_0^2 [\kappa(x) - \kappa_s] \psi + h^2 \psi = 0, \tag{B.1}
\]

while the \( \phi \) modes satisfy

\[
\frac{d^2 \phi}{dx^2} + k_0^2 [\tilde{\kappa}(x) - \kappa_s] \phi + \tilde{h}^2 \phi = 0, \tag{B.2}
\]

where \( \kappa(x) \) is the dielectric profile of the left-hand waveguide and \( \tilde{\kappa}(x) \) is the dielectric profile of the right-hand waveguide. The value \( \kappa_s \) is the dielectric constant of the substrate. Note that \( \kappa(x) \) and \( \tilde{\kappa}(x) \) differ only in the discontinuous layer. The Case I modes are characterized by eigenvalues \( h_n \), while the Case II, III and IV modes have continuous spectra, \( \tilde{h} \), as illustrated in Chapter 4.

To find the overlap integrals, multiply (B.1) by \( \phi \) and (B.2) by \( \psi \) then subtract the two equations and integrate over the interval \( (x_{N-2} - X, x_0 + X) \),
where \( X \) is a positive number and represents the distance from the substrate and superstrate boundaries to points in the superstrate and substrate regions. After an integration by parts of the two second derivative terms, the resulting equation becomes

\[
(h^2 - \tilde{h}^2) \int_{x_{N-2}}^{x_0+X} \psi(h, x) \phi(\tilde{h}, x) dx + k_\text{g}^2 (\kappa_\ell - \tilde{\kappa}_\ell) \int \ell \psi(h, x) \phi(\tilde{h}, x) dx \\
+ \left[ \phi(\tilde{h}, X) \psi'(h, X) - \phi'(\tilde{h}, X) \psi(h, X) \right] \\
- \left[ \phi(\tilde{h}, -X) \psi'(h, -X) - \phi'(\tilde{h}, -X) \psi(h, -X) \right] = 0.
\] (B.3)

For the discussion of the overlap integrals,

\[
\langle h | \tilde{h} \rangle_X = \int_{x_{N-2}}^{x_0+X} \psi(h, x) \phi(\tilde{h}, x) dx,
\] (B.4)

where the \( \psi \) mode is represented by \( \langle \cdot | \cdot \rangle \), and the \( \phi \) mode is represented by \( | \cdot \rangle \).

The overlap integral

\[
\langle h | \tilde{h} \rangle = \lim_{X \to \infty} \langle h | \tilde{h} \rangle_X,
\] (B.5)

is the solution of (B.3). If either \( \psi \) or \( \phi \) is a Case I mode, the terms at \( X \) and \(-X\) vanish in the limit so that

\[
\langle h | \tilde{h} \rangle = -\frac{k_\text{g}^2 (\kappa_\ell - \tilde{\kappa}_\ell)}{h^2 - \tilde{h}^2} \int \ell \psi(h, x) \phi(\tilde{h}, x) dx,
\]

\[
= -\frac{k_\text{g}^2 (\kappa_\ell - \tilde{\kappa}_\ell)}{h^2 - \tilde{h}^2} \langle h | \tilde{h} \rangle_\ell
\] (B.6)

Thus, the mode overlap integral is dependent only on the single layer overlap integral, \( \langle h | \tilde{h} \rangle_\ell \), of the discontinuous \( \ell \) layer. Of course, (B.6) is invalid if \( h = \tilde{h} \), which is unlikely to occur in general. (See Table B.1 for the specific case-to-case overlap integrals.)

Recall that \( h \) and \( \tilde{h} \) are the transverse wave numbers of the substrate of the left-hand and right-hand wave guides respectively, and since the superstrate and substrate are identical for both wave guides, the ranges of \( h \) and \( \tilde{h} \) are identical. For structures that have \( N \) layers, the two bracketed terms in (B.3) depend on the mode characteristics in the superstrate, 0th layer, designated as the “S” layer, and the substrate, (\( N - 1 \))th layer, designated as the “s” layer.

The superstrate term reduces to

\[
\left[ \phi(\tilde{h}, X) \psi'(h, X) - \phi'(\tilde{h}, X) \psi(h, X) \right] = \\
- \frac{\tilde{h}_S + \tilde{h}_S}{2} (\tilde{Q}_S P_S / \tilde{h}_S - Q_S \tilde{P}_S / \tilde{h}_S) \cos(h_S - \tilde{h}_S) X \\
+ \frac{h_S - \tilde{h}_S}{2} (\tilde{Q}_S P_S / h_S + Q_S \tilde{P}_S / h_S) \cos(h_S + \tilde{h}_S) X \\
- \frac{\tilde{h}_S + \tilde{h}_S}{2} (Q_S \tilde{Q}_S + P_S \tilde{P}_S / h_S \tilde{h}_S) \sin(h_S - \tilde{h}_S) X \\
- \frac{h_S - \tilde{h}_S}{2} (Q_S \tilde{Q}_S - P_S \tilde{P}_S / h_S \tilde{h}_S) \sin(h_S + \tilde{h}_S) X.
\] (B.7)
For Case II radiation modes, the superstrate fields tend to zero as $X \to \infty$ so that if the $\psi$ mode is a Case II mode, and if the $\phi$ mode is either a Case II, III or IV mode or vice versa, then

$$\phi(\tilde{h}, X) \psi'(h, X) - \phi'(\tilde{h}, X) \psi(h, X) \to 0, \quad X \to \infty,$$

whereas, the substrate term will exhibit specific singularities. Upon dividing (B.3) by $h^2 - \tilde{h}^2$, the substrate term becomes

$$\frac{1}{h^2 - \tilde{h}^2} \left[ \phi(h, -X) \psi'(h, -X) - \phi'(h, -X) \psi(h, -X) \right] =$$

$$+ \frac{1}{2} \left( \tilde{Q}_s P_s / h - Q_s \tilde{P}_s / \tilde{h} \right) \cos(h - \tilde{h}) X / (h - \tilde{h})$$

$$+ \frac{1}{2} \left( \tilde{Q}_s P_s / h + Q_s \tilde{P}_s / \tilde{h} \right) \cos(h + \tilde{h}) X / (h + \tilde{h})$$

$$+ \frac{1}{2} \left( Q_s \tilde{Q}_s + P_s \tilde{P}_s / h \tilde{h} \right) \sin(h - \tilde{h}) X / (h - \tilde{h})$$

$$+ \frac{1}{2} \left( Q_s \tilde{Q}_s - P_s \tilde{P}_s / h \tilde{h} \right) \sin(h + \tilde{h}) X / (h + \tilde{h}). \quad (B.8)$$

In the limit $X \to \infty$, the sine terms produce impulse singularities at $h = \pm \tilde{h}$, while the two cosine terms oscillate with infinite spacial frequency, except in the neighborhood of the two singular points. Since $h, \tilde{h} > 0$, only the first cosine term on the right-hand side of (B.8) exhibits simple pole behavior in the integration range. The second cosine term will be dropped because the product of rapidly oscillating functions with an arbitrary function of the independent variable, the substrate wavenumber, will integrate (over $h$) to zero. In addition, since $h$ is always positive, the relevant impulse term is

$$\lim_{X \to \infty} \frac{\sin(h - \tilde{h}) X}{h - \tilde{h}} = \pi \delta(h - \tilde{h}),$$

and the second sine term will be dropped. To obtain a contribution from the impulse singularity, the transverse wavenumbers of the substrate must be identical, $h = \tilde{h}$, which implies that $\psi$ and $\phi$ must both be Case II modes. Nevertheless, the overlap integral for a $\psi$ mode of Case II and a $\phi$ mode of Case II, III or IV is

$$(h | \tilde{h})_\ell = - \frac{k_0^2 (\kappa_\ell - \tilde{\kappa}_\ell)}{h^2 - \tilde{h}^2} (h | \tilde{h})_\ell$$

$$+ \frac{1}{2} \left( \tilde{Q}_s(h) P_s(h) / h - Q_s(h) \tilde{P}_s(h) / \tilde{h} \right) \frac{\cos(h - \tilde{h}) X}{h - \tilde{h}}$$

$$+ \frac{1}{2} \left( Q_s(h) \tilde{Q}_s(h) + P_s(h) \tilde{P}_s(h) / (h \tilde{h}) \right) \frac{\sin(h - \tilde{h}) X}{h - \tilde{h}}. \quad (B.9)$$

The third term on the right-hand-side of (B.9) produces an impulse singularity, while the first and second terms appear to have a simple pole at $h = \tilde{h}$, however, the singularity of these terms offset. In particular, if $h = \tilde{h}$, then (B.3)
becomes, assuming the $\psi$ and $\phi$ modes are Case II so that the superstrate term vanishes at large $X$,

$$k_0^2(\kappa_\ell - \tilde{\kappa}_\ell)(\hat{h} \mid \tilde{h})_\ell = \left[ \phi(\tilde{h}, -X)\psi(\tilde{h}, -X) - \phi'(\tilde{h}, -X)\psi'(\tilde{h}, -X) \right],$$

$$= \tilde{Q}_s(\tilde{h}) P_s(\tilde{h}) - Q_s(\tilde{h}) \tilde{P}_s(\tilde{h}). \quad (B.10)$$

which indicates the simple poles in $(B.9)$ offset. The cosine term in $(B.9)$ can be modified using, $\cos(h - \tilde{h})X = 1 - 2 \sin^2(h - \tilde{h})X/2$. The resulting sine term produces an impulse singularity whose coefficient is zero at $h = \tilde{h}$ which can therefore be neglected.

Thus, the proper expression for the overlap integral $(B.9)$ becomes

$$\langle h \mid \tilde{h} \rangle = \frac{k_0^2(\kappa_\ell - \tilde{\kappa}_\ell)}{h^2 - \tilde{h}^2} \langle h \mid \tilde{h} \rangle_\ell +$$

$$\frac{1}{2h^2 - \tilde{h}^2} \left[ \tilde{Q}_s(\tilde{h}) P_s(h) - Q_s(\tilde{h}) \tilde{P}_s(\tilde{h}) \right] +$$

$$\frac{\pi}{2} \left[ Q_s(h) \tilde{Q}_s(h) + P_s(h) \tilde{P}_s(h) \right] \delta(h - \tilde{h}). \quad (B.11)$$

which can also be written as

$$\langle h \mid \tilde{h} \rangle = -\frac{k_0^2(\kappa_\ell - \tilde{\kappa}_\ell)}{h^2 - \tilde{h}^2} \left[ \langle h \mid \tilde{h} \rangle_\ell - \langle \tilde{h} \mid \tilde{h} \rangle_\ell \right]$$

$$+ \frac{1}{2h^2 - \tilde{h}^2} \left[ \tilde{Q}_s(\tilde{h}) \left( \frac{h + \tilde{h}}{2h} P_s(h) - P_s(\tilde{h}) \right) \right]$$

$$- \tilde{P}_s(\tilde{h}) \left( \frac{h + \tilde{h}}{2h} Q_s(h) - Q_s(\tilde{h}) \right)$$

$$+ \frac{\pi}{2} \left[ Q_s(h) \tilde{Q}_s(h) + P_s(h) \tilde{P}_s(h) \right] \delta(h - \tilde{h}). \quad (B.12)$$

The overlap integral given by $(B.11)$ is used, for example, to determine the quadratic expression in $(6.5.13)$, which implies that $(B.11)$ is used as an argument in the computation, producing

$$\int_h \int_{\tilde{h}} \rho_{i\ell}(h) \langle h \mid \tilde{h} \rangle_\ell \tau_{i\ell}(\tilde{h}) \, dh \, d\tilde{h} = \quad (B.13)$$

$$k_0^2(\kappa_\ell - \tilde{\kappa}_\ell) \int_h \int_{\tilde{h}} \frac{\rho_{i\ell}(h) \langle h \mid \tilde{h} \rangle_\ell \tau_{i\ell}(\tilde{h})}{h^2 - \tilde{h}^2} \, dh \, d\tilde{h}$$

$$+ \frac{\pi}{2} \int_h \rho_{i\ell}(h) \left[ Q_s(\tilde{h}) \tilde{Q}_s(\tilde{h}) + P_s(\tilde{h}) \tilde{P}_s(\tilde{h}) \right] \tau_{i\ell}(\tilde{h}) \, d\tilde{h}.$$

The second term of $(B.13)$ is a double over the range $(0, \Delta)$ where $\delta$ is the dielectric step between the substrate and the superstrate, i.e., $\Delta = \kappa_s - \kappa_S$. Note that the integrals are even functions with respect to $h$ and $\tilde{h}$ so that the integrals can be extended over the $(-\Delta, \Delta)$ and divided by 2. Thus, integration over $\tilde{h}$ can be written as an integral along the contour $C_1$ in ‘olif.1′.

The overlap integral of Case III and IV radiation modes is obtained by evaluating the field expressions of $(B.3)$ for the superstrate region, Layer S.
Here, the field expressions at \( x = X \) must be similarly evaluated and included in the overlap integral as

\[
\lim_{X \to \infty} \left\{ \left[ \phi(\tilde{h}, X) \psi'(h, X) - \phi'(\tilde{h}, X) \psi(h, X) \right] / (h^2 - \tilde{h}^2) \right\} = \frac{\pi}{2} \left[ Q_S(h) \tilde{Q}_S(h) + P_S(h) \tilde{P}_S(h) / \tilde{h}_S^2 \right] \delta(hS - \tilde{h}S),
\]

where

\[
\delta(hS - \tilde{h}S) = \frac{\tilde{h}S}{h} \delta(h - \bar{h}).
\]

Note that the quantity \( h^2 - \tilde{h}^2 = \tilde{h}_S^2 - \tilde{h}_S^2 \) because \( \tilde{h}_S^2 = k_o^2 (\kappa_S - \kappa_s) + h^2 \).

| \( \psi \) Modes | \( \phi \) Modes | \( \langle h | \bar{h} \rangle \) |
|----------------|----------------|----------------|
| I              | I              | \(-k_o^2 (\kappa - \bar{\kappa}) \langle h_m | \bar{h}_n \rangle \ell\) |
| I              | II, III, IV    | \(-k_o^2 (\kappa - \bar{\kappa}) \langle h_m | \bar{h} \rangle \ell\) |
| II, III, IV    | I              | \(-k_o^2 (\kappa - \bar{\kappa}) \langle h | \bar{h}_n \rangle \ell\) |
| II             | II             | \(-k_o^2 (\kappa - \bar{\kappa}) \langle \psi_{ii}(h), \phi_{ii}(\bar{h}) \rangle \ell + \frac{(\pi/2)}{2} \left[ Q_S(h) \tilde{Q}_S(h) + P_S(h) \tilde{P}_S(h) / \tilde{h}_S^2 \right] \delta(h - \bar{h})\) |
| II             | III, IV        | \(-k_o^2 (\kappa - \bar{\kappa}) \langle \psi_{ii}(h), \phi_{ii}(\bar{h}) \rangle \ell\) |
| III (IV)       | III (IV)       | \(-k_o^2 (\kappa - \bar{\kappa}) \langle \psi_{iii}(h), \phi_{iii}(\bar{h}) \rangle \ell + \frac{(\pi/2)}{2} \left[ (Q_S(h) \tilde{Q}_S(h) + P_S(h) \tilde{P}_S(h) / \tilde{h}_S^2 \right] \delta(h - \bar{h})\) |
| III            | IV             | \(-k_o^2 (\kappa - \bar{\kappa}) \langle \psi_{iv}(h), \phi_{iv}(\bar{h}) \rangle \ell\) |

For a mode in any of the possible cases, the fields in the \( \ell \)th layer can be written as

\[
\psi(h, x) = Q_{\ell} \cos \bar{h}_{\ell}(x - x_{\ell}) + P_{\ell} \sin \bar{h}_{\ell}(x - x_{\ell}) / \bar{h}_{\ell},
\]

\[
\phi(h, x) = \tilde{Q}_{\ell} \cos \bar{h}_{\ell}(x - x_{\ell}) + \tilde{P}_{\ell} \sin \bar{h}_{\ell}(x - x_{\ell}) / \bar{h}_{\ell}.
\]
and the value of the layer overlap integral is (via Mathematica)

\[
\langle \psi(h), \phi(h) \rangle_\ell = \frac{1}{h_\ell^2 - \hat{h}_\ell^2} \left\{ P_\ell \hat{Q}_\ell - Q_\ell \hat{P}_\ell + \right.
\]

\[
\left. \left[ (-P_\ell \hat{Q}_\ell + Q_\ell \hat{P}_\ell) \cos \hat{h}_\ell d_\ell - (Q_\ell \hat{Q}_\ell \hat{h}_\ell^2 + P_\ell \hat{P}_\ell) \sin \hat{h}_\ell d_\ell / \hat{h}_\ell \right] \cos h_\ell d_\ell + \right.
\]

\[
\left. \left[ (Q_\ell \hat{Q}_\ell \hat{h}_\ell^2 + P_\ell \hat{P}_\ell) \cos \hat{h}_\ell d_\ell + (Q_\ell \hat{P}_\ell \hat{h}_\ell^2 - P_\ell \hat{Q}_\ell \hat{h}_\ell^2) \sin \hat{h}_\ell d_\ell / \hat{h}_\ell \right] \sin h_\ell d_\ell / h_\ell \right\}
\]

Depending on the values of \( h \) and \( \hat{h} \), \( h_\ell \) and \( \hat{h}_\ell \) may be either real or imaginary (reactive material). The layer overlap integral as given by (B.14) can be easily modified to reflect the different possibilities. For example, assume that \( h_\ell \) is imaginary, \( h_\ell = j h_\ell'' \), then \( h_\ell^2 = -(h_\ell'')^2 \) and \( \sin h_\ell d_\ell / h_\ell = \sinh h_\ell'' d_\ell / h_\ell'' \).
Appendix C

Tooth Geometries for Boundary Element Methods

The expression for the electric field is given by the expansion

\[ E_y = \Psi(x, z) = \exp(-y z) \sum_{n=N_f}^{N_f-1} \psi_n(x) \exp(-jKz), \quad (C.1) \]

where the total number of elements is \( N_f = 2N_f \). The \( x \) component of the magnetic field is

\[ H_x = \frac{1}{j\omega\mu} \frac{\partial E_y}{\partial z} \quad (C.2) \]

For reference, the origin, \( z = 0 \) is located at the center of the tooth, and the location of the nodes on the panels along the \( z \) axis is given by

\[ z_l = \frac{\Delta z}{2}(2l + 1 - N_a). \quad (C.3) \]

Figure ‘fapA.1’ illustrates the geometry of the grating with the axial boundary points. The number of axial points (and space harmonics) is

\[ N_l = N_a + N_b = 2N_f = \Lambda/\Delta z. \quad (C.4) \]

Figure C.1. The panel geometry and its relation with the space harmonics.
The electric field given by Eq. (C.1) may be rewritten as
\[
\Psi(x, z) \exp(\gamma z) = F(x, z) = F(x, z + \Lambda) \\
= \sum_{n=N_f}^{N_f-1} \psi_n(x) \exp(-jKz).
\]  
(C.5)

At the axial panel points, \( z_l \), at the base of the grating \((x = 0)\) the periodic part of the electric field can be expressed in terms of the space harmonic amplitudes \( \psi_n \equiv \psi_n(0) \),
\[
F(0, z_l) = \Phi_l = \sum_{n=-N_f}^{N_f-1} \psi_n \exp\left[j \frac{2\pi}{N_l} n (N_a - 1) \right] \exp(-j \frac{2\pi}{N_l} nln). 
\]

The values of \( \Phi_l \) are the discrete Fourier Transform of the field at the base of the grating and are written as
\[
\Phi_l = \sum_{n=-N_f}^{N_f-1} \phi_n W^{nl}, 
\]  
(C.6)

where
\[
W = \exp(-j \frac{2\pi}{N_l}), \\
\phi_n = \psi_n W^{-n(N_a-1)/2}. 
\]

The inverse transform is
\[
\psi_n = \frac{1}{N_l} \sum_{l=0}^{N_l-1} \Phi_l W^{-nl}. 
\]  
(C.7)

Using the expression \( \omega \mu = k \eta \), where \( k \) is the free-space wave number and \( \eta \) is the free-space wave impedance, the transverse magnetic field becomes
\[
\eta H_x = j \frac{1}{k} \left[ \gamma F(x, z) + jK \tilde{F}(x, z) \right] \exp(-\gamma z). 
\]  
(C.8)

where
\[
\tilde{F}(x, z) = \sum_{n=N_f}^{N_f-1} n \psi_n(x) \exp(-jKz). 
\]  
(C.9)

At the axial points \( z = z_l \), and at the bottom of the grating, Eq. (C.9) becomes
\[
\tilde{\Phi}_l = \sum_{n=-N_f}^{N_f-1} n \phi_n W^{nl}, \\
= \frac{1}{N_l} \sum_{l'=0}^{N_l-1} N_l \tilde{\Phi}_{l'}. 
\]  
(C.10)
where \( N_l \) is the discrete Fourier Transform of \( n \),

\[
N_l = -(-1)^l \frac{N_t}{2} \begin{cases} 1, & l = 0; \\ 1 - j \cot(l/N_t), & \text{otherwise} \end{cases}
\]

Although the above vectors with elements \( \psi_n, \phi_l \), and \( \Phi_l \) are identified with the fields at the base of the grating, the above analysis is valid for any lateral \( x \) position, \( i.e. \), the vector elements will be functions of the lateral position.
Appendix D
Laplace Transform Table

The Laplace transform of the function $f(t)$ for $t > 0$ is

$$F(s) = \mathcal{L} [f(t)] = \int_0^\infty f(t) e^{-st} \, dt,$$

and the inverse

$$f(t) = \mathcal{L}^{-1} [F(s)] = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} F(s) e^{st} \, ds.$$

<table>
<thead>
<tr>
<th></th>
<th>$f(t) = \mathcal{L}^{-1} [F(s)]$</th>
<th>$F(s) = \mathcal{L} [f(t)]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Derivative)</td>
<td>$f'(t)$</td>
<td>$sF(s) - f(0)$</td>
</tr>
<tr>
<td>2 (2nd derivative)</td>
<td>$f''(t)$</td>
<td>$s^2 F(s) - sf(0) - f'(0)$</td>
</tr>
<tr>
<td>3 (Integral)</td>
<td>$\int_0^t f(\tau) , d\tau$</td>
<td>$\frac{1}{s} F(s)$</td>
</tr>
<tr>
<td>4 (Scaling)</td>
<td>$f(\alpha t)$</td>
<td>$\frac{1}{\alpha} F\left(\frac{s}{\alpha}\right)$</td>
</tr>
<tr>
<td>5 (Step)</td>
<td>$u(t)$</td>
<td>$\frac{1}{s}$</td>
</tr>
<tr>
<td>6 (Ramp)</td>
<td>$t u(t)$</td>
<td>$\frac{1}{s^2}$</td>
</tr>
<tr>
<td>7 ($n = 1, 2, 5, \ldots$)</td>
<td>$t^n u(t)$</td>
<td>$\frac{n!}{s^{n+1}}$</td>
</tr>
<tr>
<td>No.</td>
<td>Function</td>
<td>Transform</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>8</td>
<td>$f(t - \alpha) u(t - \alpha)$</td>
<td>$F(s) e^{-\alpha s}$</td>
</tr>
<tr>
<td>9</td>
<td>$\delta(t - \alpha)$</td>
<td>$e^{-\alpha s}$</td>
</tr>
<tr>
<td>10</td>
<td>$u(t) \sin \omega t$</td>
<td>$\frac{\omega}{s^2 + \omega^2}$</td>
</tr>
<tr>
<td>11</td>
<td>$u(t) \cos \omega t$</td>
<td>$\frac{s}{s^2 + \omega^2}$</td>
</tr>
<tr>
<td>12</td>
<td>$u(t) e^{-\alpha t}$</td>
<td>$\frac{1}{s + \alpha}$</td>
</tr>
<tr>
<td>13</td>
<td>$u(t) \frac{t^{n-1}}{(n-1)!} e^{-\alpha t}$</td>
<td>$\frac{1}{(s + \alpha)^n}$</td>
</tr>
</tbody>
</table>


