This is to certify that the
thesis entitled

ON THE SIMULATED INTERACTIONS BETWEEN SURFACE
PLASMONS AND BURIED TWO-DIMENSIONAL ELECTRON
GASES

presented by

COLLIN S. MEIERBACHTOL

has been accepted towards fulfillment
of the requirements for the

M.S. degree in ELECTRICAL ENGINEERING

Terence Brown
Major Professor's Signature

5/11/09

Date

ON THE SIMULATED INTERACTIONS BETWEEN SURFACE PLASMONS AND BURIED TWO-DIMENSIONAL ELECTRON GASES

By

Collin S. Meierbachtol

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Electrical Engineering

2009

ABSTRACT

ON THE SIMULATED INTERACTIONS BETWEEN SURFACE PLASMONS AND BURIED TWO-DIMENSIONAL ELECTRON GASES

By

Collin S. Meierbachtol

The surface waves associated with Surface Plasmons (SPs) were derived for one-dimensional electromagnetic scattering. In particular, the surface wave amplitudes stemming from the excitation of SPs in the often-cited Kretschmann-Raether system are provided. Furthermore, standing surface waves introduced via binary gratings were also derived using the Rigorous Coupled-Wave Analysis (RCWA). These electric fields were then simulated at optical and far infrared frequencies via MATLAB. An AlGaN/GaN HEMT was simulated under standard operating variables using the Silvaco ATLAS software package. The formation of a Two-Dimensional Electron Gas (2DEG) layer resulted in a roughly 5nm deep region of free carriers, with a complex permittivity determined citing the phenomenological fluid model of the effective mass equation. Many-layered electromagnetic scattering was then performed including the presence of the AlGaN/GaN HEMT under nominal SP excitation conditions. Various combinations of real and lossy permittivity layers were then analyzed for maximum electric field amplitude within the 2DEG layer. Even in the presence of multiple lossy media, the electric fields within the 2DEG layer were often significant when compared to that of the incident electromagnetic radiation. The exploitation of these interactions between the SP and 2DEG electric fields could lead to the development of novel sensors and high frequency optoelectronic devices.

To my teachers for always guiding me.

To my coaches for always pushing me.

To my parents for always believing in me.

To Krista for always being there for me.

PREFACE

The original goal at the onset of this research project was to improve the overall throughput of Surface Plasmon Resonance (SPR) based biosensors. Although originally proposed more than twenty years ago and utilized widely in the pharmaceutical industry since their inception, the simple physics of these systems have changed very little over the years. Moreover, few improvements have been developed regardless of engineering, physics, and optics breakthroughs during this same time.

Several lengthy discussions regarding the sensing capabilities of present SPR sensor arrays, along with their limitations, followed. Before long, it was well understood that a more in-depth knowledge of the fields associated with surface plasmons (SPs) was required before moving onto the solid state applications. This set the premise for the first several chapters of this thesis.

Returning to the physical applications side, the idea of placing a transistor on the back of an SPR-based biosensor was proposed. The advantage of this configuration is it avoids the indirect measurement of reflection data, and at the same time allows for the decrease in size of the device. Thus, a solid state device was introduced as a data collection mechanism. Familiarity with high electron mobility transistors (HEMTs) led to its inclusion in the device strategy. However, although familiarity with HEMTs with regards to semiconductor physics and engineering was well understood, the interactions between the SPs and HEMTs via electromagnetic fields has been relatively untouched in the scientific field, save the pioneering work by Hashim, et. al. and several others [2, 3] conducted over the past decade. Many these articles either pertain to the interactions produced by interdigitated slow waves, terahertz (THz) radiation, or so-called *plasmons* as related to the 2DEG region. Therefore, in an effort to bridge the gap between the two disciplines and these inherently similar phenomena, this work will primarily explore the interactions fields produced by to SPs for wavelengths within the optical region. However, in order to stay consistent with the referenced material, wavelengths within the far infrared will also be addressed in the later chapters.

Table of Contents

List of Tables

List of Figures

Chapter 1

An Introduction to Surface Plasmons

Surface Plasmons (SPs) have proven to play a vital role in academia and industry within the past century. Their presence has been demonstrated to improve the transmission of electromagnetic fields across complex systems, and provide a critical link to the real-time analysis of biological and molecular interactions. However, their dissemination has proven rather limited, as their presence is often known only within these small academic subcultures. Therefore, an introduction to SPs is provided in the following chapter. A history of SPs in the academic field is first presented, followed by their applications in the pharmaceutical industry. Finally, the present state of plasmonic devices is visited in brief.

1.1 Plasmons, Plasmas, and Polaritons

A complete disambiguation of plasmonic terms is first presented as a courtesy to the reader. Over the past several decades, many different, albeit related, plasmon phenomena have been coined under the same cloud of ambiguity. A list of such likely terms may include surface plasmons, surface polaritons, surface plasmon polaritons, surface plasmon resonance, two-dimensional electron gases, and so on. Before proceeding, these terms must be consistently defined as to avoid such hazy language here.

Plasmons are defined as quanta of plasma, or free charge, oscillations. These can be thought of as an extension of similar oscillatory phenomena due to light and sound waves, or *photons* and *phonons*, respectively. The energy associated with any one plasmon quanta is approximated using the free electron model [4] in metals as

$$E = \hbar\omega_p = \hbar\sqrt{\frac{4\pi n q^2}{m_0 \epsilon_0}} \tag{1.1}$$

where $\hbar = h/2\pi$, h being Planck's Constant, n is the electron density in dimensions of per volume, q is the electron unit charge, m_0 is the rest mass of an electron, and ϵ_0 is the permittivity of free space. A quick calculation shows these energy quanta are often on the order of tens of electron-volts within any real metallic material.

Surface Plasmons (SPs) are defined as coherent oscillations of free charge, or plasmons, bound to the interface separating two adjacent media comprising a metal and a dielectric. To some extent, SPs are present in all wave interactions across an interfacial boundary for complex transmitted wavevectors involving a metal layer. However, their coherent wave intensity is in most cases negligible within the far-field regime. This is due to their evanescent nature in both the longitudinal and tangential directions with respect to the interface.

Surface Plasmon Resonance (SPR) is a unique wavenumber matching condition that results in low, or zero-, reflectance above the critical angle. This occurs whenever the incident wave is completely coupled to the SPs, often described physically as the complete absorption of the incident field into the bounded layers. Coincidentally, this resonance condition often corresponds to an enhanced transmitted wave magnitude. A more detailed derivation of this wavenumber matching relationship will be provided later on.

Surface Plasmon Polaritons (SPPs) are the aptly named polarized oscillations of SPs. Polaritons in general are often defined as a pair of positive and negative states, be these atoms, electron-hole pairs, or even entire molecules. For example, in solid-state engineering, a polariton is defined as a coupled electron-hole pair and an incremental lattice vibration, or phonon. In academia and supported literature, it is often the case that SPs and SPPs are used interchangeably to describe the coherent oscillation of free charge. In an effort to avoid confusion, such free charge oscillations will be henceforth referred to as SPs.

With the ambiguity related to plasmons addressed, the discussion of SPs can proceed uninhibited. This discussion will begin with a short history of their discovery and development.

1.2 History of Surface Plasmons

The first scholarly record outlining what would later be called SPs is attributed to R.W. Wood in 1902 at Johns Hopkins University [5] . His work described in detail the illumination intensity recorded when shining monochromatic light upon a diffraction grating at various incident angles. This work also made note to great detail the dramatic drop in transmitted light over a small change in incident angle. Although Wood had no idea at the time that this dramatic drop of transmitted light intensity was due to SPs, his article did mark the beginning of the SP research era.

The term surface *plasmon* itself dates to 1957 in a Physical Review article published by R.H. Ritchie on the plasma loss associated with fast electrons [6]. This work was originally based on the theories of electron interactions pioneered by Bohm and Pines five years earlier in a trio of Physical Review articles dealing with the "collective plasma-oscillations" of electron gases in thin metal films [7–9]. These theories of oscillating free charges confined to an interface would finally give way to the collective theory of plasmons through experimental verification by Watanabe in 1956 [10].

Meanwhile, the first experimental observance of SPR occurred in 1959, when T. Turbadar at the University of Damascus observed unusually low reflectance from thin metal films at oblique incident angles [11]. Seeking only experimental verification of the metallic thin film reflection theory outlined two years earlier by F. Abeles [12], Turbadar realized that at certain oblique angles, the measured reflectance dropped dramatically. In fact, as it turned out this phenomenon was also predicted by Abeles' theoretical work. However, unlike Woods, Turbadar was using only planar (one-dimensional) metallic thin films. Turbadar correctly concluded, although could not verify, that this drastic decrease in reflectance was due to the wave energy simply being absorbed by the metal. It was not until some nine years later that Andreas Otto correctly identified and experimentally verified the zero reflection phenomenon that Turbadar had reported.

Otto's findings were published in his 1968 paper on non-radiative Surface Plasma Waves (SPWs) [13]. Using a total internal reflection system shown in Figure 1.1 with a prism coupler suspended above a bulk metallic layer, Otto systematically excited these SPWs on smooth surfaces of silver films using TM_y polarization. In fact, it was this experimental setup for which Otto is most commonly known today. It has become one of the two most commonly referenced experimental SP articles and systems.

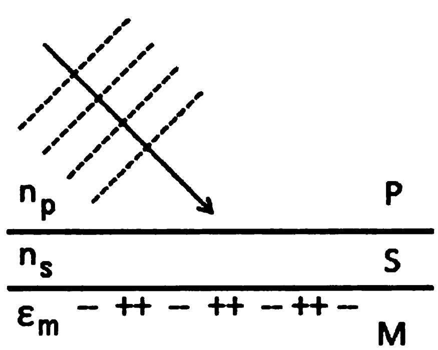

Figure 1.1: The Otto configuration for enhanced SP excitation.

Note in the figure that the metal layer (M) is assumed as a semi-infinite thick slab and is separated from the higher permittivity layer by a thin dielectric slab (S). The location of the SPs is denoted by the alternating plus (+) and minus (-) signs on the interface separating the M and S layers. The necessary condition for total internal reflection of $(n_P) > (n_S)$ must be met to produce these SPs at real oblique incidence. Moreover, this top layer is often a type of glass with a refractive index of roughly 1.5 while the dielectric layer (S) is vacuum or air, with close to unity refractive index.

The standard notation of TM polarization with respect to the y-axis is adopted from here on out, in accordance with the standard literature [14]. More information pertaining to polarization will be presented in the proceeding chapters.

In that same year, Heinz Raether and Erik Kretschmann proposed a similar configuration for launching SPs. However, their setup, later referred to as the Kretschmann-Raether Configuration shown in Figure 1.2, assumed that the metal was the thin layer, while the dielectric was now assumed to be of semi-infinite thickness [15].

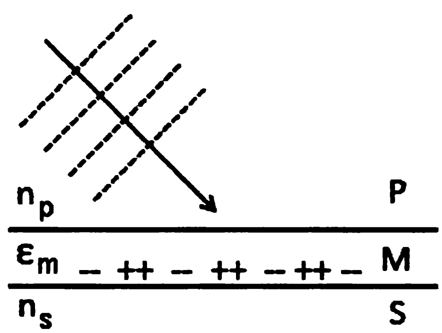

Figure 1.2: The Kretschmann-Raether Configuration for enhanced SP excitation

This assumption of a thin metallic layer [4] was more applicable with the sputtering technologies available at the time, which is still the case today. It also allows for the direct observation of the transmitted SPs, which would otherwise be shielded from view by the semi-finite metallic layer and the incident wave in the Otto Configuration. This turned out to be a main advantage over the Otto Configuration with the birth of the Scanning Near-Field Optical Microscopy (SNOM) over the next decades.

With the demonstration of the Otto and Kretschmann-Raether Configurations, the proven theoretical work on the electromagnetics of thin film layers put forth by Abeles, and the understanding of SPs now in place, various and extensive research of SPs could now be conducted.

1.3 SP-Based Biosensors

A brief introduction on SPR sensors follows. This will include sections on the elementary device physics, current SPR analysis technology, and its advantages and limitations.

SPR-based biosensors have been available to the biomedical and pharmaceutical industries ever since Biacore AB introduced BIAcore SPR–the world's first SPR-based analytical instrument for studying biomolecular interactions–in 1990 [16].

These sensors are primarily used to monitor protein-protein interactions in real-time, a major breakthrough in the development of pharmaceutical industry. Although a breakthrough in its own right, the BIAcore SPR boasted a key distinguishing advantage over various other analysis equipment; its remarkable sensitivity to drug interactions. For instance, BIAcore SPR sensors have reported real-time interaction sensitivities as low as 10 ng/ml of staphylococcal enterotoxin B (SEB) in foods, and even 0.02 ppb of Chlorampenicol in poultry and dairy products [17, 18]. Of course in all actuality, this sensitivity is not based on the protein-protein interactions themselves, but in the minute changes in the refractive index of the sensing layer.

Another distinct advantage of SPR-based sensors is their real-time presentation of results. For instance, the BIAcore SPR, along with corresponding software, can display the real-time binding kinetics of interactions visually in plot form. This was a marked improvement upon traditional sensors that provided only quantifiable information at the conclusion of the analysis period. The interaction data are often displayed as changes in refractance units (RU) [19]. These RUs are defined by Biacore as one one-millionth of a Refractive Index Unit (RIU), whereas RIU is the ratio of change in reflectance per a change in refractive index of the underlayer (i.e. dielectric layer). Mathematically, this becomes [20]

$$RU = \frac{RIU}{10^6} = \frac{\delta R}{\delta n} \tag{1.2}$$

where δR is the change in reflectance per some change in the refractive index δn of the dielectric layer adjacent to the thin metal film. Note this intense sensitivity presented in Equation (1.1). Since reflectance is simply a measure of the ratio of scattered field to incident, it has a range extending from zero to unity with no units. Likewise, the refractive index of any material is unit less and related to the speed of light passing through that material. Refractive indices of common materials range from one up to 2.4, which is roughly the refractive index of diamond [21]. Therefore, with normal operation ranges in rates of 1000 RU/min during the initial *association* phase of the analysis and -100 RU/min during *dislocation*, it is easy to see why such SPR biosensors are heralded for their sensitivity to surface kinematics [20].

Further advantages of such sensors include their reusability and their non-volatile nature. That is, their use of gold as the primary thin film metal avoids any unwanted chemical reactions that may cause unwanted damage to the applied proteins or skewed data because of these reactions. Other usable sensing parameters include bonding affinity, association and disassociation time constants, and percentage of bonding.

First and foremost, this entire process hinges on the extreme sensitivity and reproducibility of reflectance data for any given protein application. Moreover, acquisition period must undergo extensive calibration procedures in order to record the particular reflectance data that will be used in future measurements.

However, these points aside, there are several distinct disadvantages so SPR base biosensors. First and foremost, SPR biosensors have low throughput. *Throughput* is an engineering term defined as a measure of the the amount of data that a system can analyze or transmit over a given time interval. To the reader, it may appear that current SPR biosensors would lend themselves to analyzing billions of individual proteins at a time. This argument is based on each protein occupying a cross sectional area of hundreds of nanometers on a side.

To a point, this is true–the analyzed proteins do in fact occupy cross sectional areas within this characteristic size. However, protein bonding sites are rarely probed individually. That is, it is extremely difficult, albeit nearly impossible, to focus light through an aperture of 100nm radius and emit any usable light out the other side. Instead, the ever-so-critical reflectance measurement is often performed over the numerous samples as a whole. In other words, any given reflectance measurement is only a record of the *average* reflectance from the entire sample region.

Although this would appear to cut down on the analysis run time and lead to a much more precise measurement, it in fact leads to the opposite. Scanning across many sites requires a much longer run time in order to correctly reach an overall sample average. Moreover, it reduces the interactions of thousands of bonding sites to a single data point. This also introduces unwanted noise into the measurements, as the protein-gold bonds are often anything but uniform.

To improve their product in light of these effects, in 2000 Biacore AB introduced the *Flexchip*; an SPR sensor array containing 400 individually measured protein-protein interaction sites. Each one of these mentioned sites can be monitored individually in real time [22].

Although heralded as a tremendous breakthrough upon its introduction, the problem with low throughput still remains a key hurdle in the development of SPR sensing technology. Remember, the spatial location of proteins is only one part of the problem; the other is the time it takes to perform accurate measurements. The Flexchip quotes analysis periods of roughly 10 minutes [23]. Although the Flexchip does increase the spatial throughput of the system, it does little to decrease the analysis run time. In order to increase the run time throughput as a whole, the analysis period would have to be shortened drastically. This is presently not possible with the current Flexchip technology.

Since their introduction to the biosensing industry in 1990, SPR-based biosensors have changed very little in regards to their underlying physics. Thus, it remains paramount that research continues in the SPR biosensing field in order to increase its throughput and remain a volatile choice for the biosensing industry.

1.4 Plasmonic Devices

SP-based sensors are not limited to the biomedical and pharmaceutical industries. Their applications have included the detection of vapors and thin film characterization [24]. Several patents have also recently been awarded involving various SPR sensors [25]. In fact, one such patent involved an SPR sensor that actually sensed the SPR phenomenon itself [26].

Recently, SPs have been introduced within the solid state and optical engineering fields to enhance the photoluminescence of various wide-bandgap Lasers [27–29]. Various other improved performance parameters have also been attributed to the presense of SPs [30, 31]. Finally, the sensor applications for Terahertz radiation has been explored in the past decade [1, 2, 32, 33]. However, many of these cited articles report merely experimental findings. The following chapters will attempt to bridge the gap between electromagnetic *and* solid state engineering involved in the introduction of the field associated with SPs into semiconductor devices.

Chapter 2

Maxwell's Equations, Waves, and Polarization

Before the *interactions* between SPs and two-dimensional electron gases may be explored, they must first be separately well-defined and theoretically explored. This chapter will provide an elementary introduction of Maxwell's Equations, plane waves, and polarization. Finally, a brief section addressing the polarization requirements in order to produce SPs will be discussed.

2.1 Maxwell's Equations

As is the case with most electromagnetic theory systems, Maxwell's Equations mark the beginning point of the discussion. Admittedly, it is a bit disingenuous to coin the following series of expressions as Maxwell's Equations, for Maxwell's only contribution to this set was a single term in a single equation. Nevertheless, as it is standard practice, the following expressions are presented as the point form time-harmonic, or frequency domain, Maxwell's Equations in the absense of magnetic charge and current [34].

$$\nabla\cdot\vec{D} = \rho_e \tag{2.1}$$

$$\nabla \cdot \vec{B} = 0 \tag{2.2}$$

$$\nabla \times \vec{E} = -j\omega \vec{B} \tag{2.3}$$

$$\nabla \times \vec{H} = \vec{J_e} + j\omega \vec{D} \tag{2.4}$$

In Equations 2.1-2.4, $\vec{D}$ is defined as the electric flux density, ρ_e is the electric charge density, $\vec{B}$ is the magnetic flux density, $\vec{E}$ is the electric field, $\vec{H}$ is the magnetic field, and $\vec{J_e}$ is the electric current density. As stated previously, it will be assumed henceforth that no magnetic charge, nor magnetic current, are present (as is the case with most real world systems). The inclusion of these quantities adds extra terms to the above relations and are presented elsewhere [35]. Assuming isotropic media with regards to magnetization and polarization, the relationships between field intensities and flux densities can be further defined as

$$\vec{D} = \epsilon \vec{E} \tag{2.5}$$

$$\vec{B} = \mu \vec{H} \tag{2.6}$$

where ϵ and μ are the total permittivity and permeability within each medium. These two values are often defined as the product of their respective values for free space and their relative counterparts. Thus, the total permittivity is often recorded as

$$\epsilon = \epsilon_r \epsilon_0 \tag{2.7}$$

and for the permeability

$$\mu = \mu_r \mu_0 \tag{2.8}$$

where ϵ_r is the relative permittivity of the medium and ϵ_0 is the permittivity of free space. The relative permeability is μ_r, while μ_0 is permeability of free space.

The previous equations are all presented in the frequency domain. Without loss of generality, corresponding equations are available in the time domain. Likewise, the integral form of Maxwell's Equations in both the time and frequency domains are also available, but are not discussed here.

2.2 Plane Waves

With Maxwell's Equations clearly defined, their application to physical systems may now be discussed. The solutions to Maxwell's Equations may take on many forms; dipole sources, sheet currents, and circular polarized propagating fields are just a few of these solutions to name a few. The following incident fields to be discussed from now on will all take on the form of plane waves with wavevector $\vec{k}$, and frequency ω, as given within the rectangular spatial coordinate system.

Plane waves are spectral solutions to the wave equation. These waves also describe the extreme far-field realization of most point and local field sources. This phenomenon is often referred to as the Sommerfeld Radiation Condition, which in turn poses that the fields radiated from any source must tend toward resembling plane waves as the distance from the source approaches infinity [37]. Plane waves are some of the most applicable, yet simplistic field representations available.

The derivation of such field solutions is quite simple. Without loss of generality, the solution for the electric field intensity will be presented. However, a similar solution exists in the same form for the magnetic field. In fact, it will be obvious that this magnetic field solution is often more applicable in the TM_y polarization required.

With Maxwell's Equations defined in point form within the frequency domain as above, the curl of Equation 2.3 can be performed once more. Moreover, inserting Equation 2.4 into the new expression results in the following

$$\nabla \times \nabla \times \vec{E} = -\omega^2 \mu\epsilon \vec{E} \tag{2.9}$$

where the vector current density, J_e, has been assumed to be negligible. Equation 2.9 may be rewritten as

$$\nabla \times \nabla \times \vec{E} + \omega^2 \mu \epsilon \vec{E} = 0 \tag{2.10}$$

Replacing the coefficient in front of the single vector electric field term with the wavevector, k_n, for any given medium denoted with relative permittivity ϵ_n, is defined by

$$k_n = \omega\sqrt{\mu\epsilon_n} = \frac{\omega}{c}\sqrt{\epsilon_{rn}} \tag{2.11}$$

where c is the speed of light given by $c = 1/\sqrt{\mu\epsilon}$. Equation 2.10 now becomes

$$\nabla \times \nabla \times \vec{E} + k^2 \vec{E} = 0 \tag{2.12}$$

Equation 2.12 is a linear, second-order differential equation in terms of $\vec{E}$, with its solutions taking the form

$$\vec{E} = Ae^{-j\vec{k}\cdot\vec{r}} + Be^{j\vec{k}\cdot\vec{r}} \tag{2.13}$$

where A and B are scalar constants to be determined. Therefore, the solution takes on the form of two plane waves, one travelling along $\vec{k}$, and the other in the opposite direction. In the absense of any change in relative permittivity in the medium through which a plane may propagate, one of these terms is often dropped. This is due to the fact that, without a change in permittivity, the wave will continue to propagate uninhibited toward infinity.

To begin, a three-dimensional plane wave of linear polarization is assumed to be propagating within some medium at an angular frequency, ω. This wave may be represented mathematically as

$$\vec{E}\left(\vec{r}\right) = \vec{u}E_0 e^{-j(\vec{k}\cdot\vec{r} - \omega t)} \tag{2.14}$$

where $\vec{u}$ is the polarization vector, E_0 is the electric field magnitude or intensity, $\vec{k}$ is the wavevector, and $\vec{r}$ is the vector spatial coordinate. This polarization vector will be dealt with in the following section, but for now is spatially related to the polarization angle. Written in terms of rectangular coordinates, the wavevector can be reduced to its spatial components along each coordinate axis

$$\vec{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z} \tag{2.15}$$

where the $\hat{x}$,$\hat{y}$, and $\hat{z}$ are simply the unit vectors along the x-, y-, and z-directions, respectively. Moreover, the magnitude of the wavevector, known as the wavenumber, is directly related to the wavelength by

$$|\vec{k}| = \frac{2\pi}{\lambda} \tag{2.16}$$

where λ denotes the wavelength of the electromagnetic field within the medium of propagation. A similar relation also exists between the radial frequency of the wave, ω, and the frequency of the wave

$$\omega = 2\pi f \tag{2.17}$$

where f is the frequency of the wave, always provided in units of Hertz, or cycles per second.

Therefore, the plane wave electric field may be completely described by

$$\vec{E}(x, y, z, t) = (u_x\hat{x} + u_y\hat{y} + u_z\hat{z})E_0e^{-j(k_xx + k_yy + k_zz - \omega t)} \tag{2.18}$$

Note that it could have been posited that the incident field was completely described via a magnetic field with a similar linear polarization. Likewise, the reader will quickly recognize that a magnetic field *must* accompany the above electric fields in order to satisfy Maxwell's Equations. This accompanying magnetic field intensity is calculated from the applicable Maxwellian Equations 2.3 and 2.4 as

$$\nabla \times \vec{E} = -j\omega\mu\vec{H} \tag{2.19}$$

Solving for the magnetic field intensity in terms of the electric field intensity yields

$$\vec{H} = \frac{j}{\omega\mu\epsilon}\nabla \times \vec{E} \tag{2.20}$$

Moreover, if the electric field intensity is the defining quantity as

$$\vec{E}(\vec{r},t) = \vec{u}E_0 e^{-j(\vec{k}\cdot\vec{r} - \omega t)} \tag{2.21}$$

then the corresponding magnetic field intensity is calculated via

$$\vec{H}(\vec{r},t) = \frac{j}{\omega\mu\epsilon}\nabla \times \vec{E}(\vec{r},t) \tag{2.22}$$

Note here the electromagnetic wave frequency, permeability, and permittivity of the surrounding environment remain outside of the del-operator. This is standard practice when the physical properties of this medium remain spatially invariant. This is assumed to be the case from here on, since this often models the behavior of real world materials.

Thus, the above derivations highlight the inherent relation between magnetic and electric fields associated with any propagating electromagnetic wave.

2.3 Polarization

The polarization of any such electromagnetic wave is defined as the lotus traced by the electric field vector. Since this electric field vector may vary in time, the polarization of the propagating wave may also vary in time. Such time-varying polarizations will not be discussed here, and are presented elsewhere [38].

Linear polarization is characterized by an electric field vector with constant phase angle with respect to time. However, this does not require the magnitude of the vector to be constant. The phase angle associated with linear polarization defines the class of linear polarization for the wave. Two such classes, often referred to as TE and TM polarization, will be used here to classify all plane waves incident upon a boundary. Moreover, a time-varying superposition of the two can be used to represent any such polarization state. However, only these two polarization states, or a linear combination thereof, will be dealt with from here on out. In fact, the last section of this chapter describes the necessary polarization condition to support SPs.

Transverse Electric (TE) polarization is denoted by an electric field vector that is oriented perpendicular to the plane of incidence. Also coined as E-polarization or s-polarization, coming from the German word *senkrecht* meaning "vertical", this polarization is often characterized by an electric field that lies parallel to the interfacial boundary plane.

On the other hand, Transverse Magnetic (TM) polarization refers to an electric field that is oriented parallel to the plane of incidence. TM polarization is sometimes referred to as H-polarization or p-polarization, with its origins in the German word *parallele*, an obvious english cognate meaning "parallel".

One realizes that for the special case of completely orthogonal incidence, the electromagnetic wave may be defined by both TM and TE polarization. This special case is coined as TEM polarization, since the electric field vectors defining TE and TM polarizations becomes ambiguous. This case also results in an angle of incidence, θ, of 0° as measured from the normal with respect to the interface. This is the standard convention found in most scholarly references, and will continue to be referred to as such.

2.4 Plasmons and TM Polarization

The preceding section dealt with the polarization of free waves travelling through space. However, this polarization becomes an important factor in the launching of plasmons. It is often the case in many books and citing articles that the polarization of SPs is simply noted to require TM polarization. However, the discussion is often left at that, with no further information provided.

As stated previously, SPs are essentially charged plasma quanta that have been confined to the surface, or interface, of a particular medium. These plasma quanta are simply electric charge volumes. Moreover, as will be proven in the following several chapters, the fields associated with them must be evanescent, or decay, in both the tangential and normal directions away from the interface. Since these charges are electric in nature, the fields associated with them must also be electric, as opposed to magnetic. Therefore, these charges must have *electric* fields associated with them that vary in both the parallel and perpendicular directions away from the interface. It is this fact that gives rise to a requirement for the polarization of such fields to be transverse magnetic. It is only through TM polarization that these *electric* charges are confined via an *electric* field to the surface of the medium, through the combination of parallel and perpendicular electric fields associated with this surface. For this reason, *only* TM polarization will be considered as the primary linear polarization state of interest from here on out. Moreover, since only the transverse magnetic polarization state will be discussed henceforth, the following derivations will be primarily carried out for the associated magnetic fields, with the resulting electric fields calculated via Equation 2.4.

Chapter 3

Scattering in One-Dimension: Two-Layer Systems

In the previous chapter, both TE and TM linear polarizations were discussed, with the definition of each depending on the field orientations upon interacting with a planar interface. This chapter will deal with such simple planar systems, confined to vary spatially along a single vector. Out of simplicity, this will be assumed as the z-axis. Moreover, a tendency to define the incident field impinging from the negative z-region, or $z < 0$, will be assumed. Since the system varies only in the z-direction, this leads to interfacial planes separating adjacent media lying parallel to the x-y plane.

An incident azimuthal angle of 0° will also be assumed unless otherwise noted, transforming the three-dimensional real space into a two-dimensional real space. However, no information is lost from the original wave during this transformation. Moreover, it can be shown that, for any wave incident from any azimuthal angle ϕ with respect to the x-axis upon a planar boundary, a coordinate axis rotation of the same angle results in a similar two-dimensional system. Therefore, this alignment rotation with the x-axis is justified as equivalent to a three-dimensional system with a coordinate transformation.

3.1 Polarized Plane Waves

As stated in the previous chapter, the magnetic field vector describing any two-dimensional electromagnetic wave above may be expressed in the general sense using its wave vector, $\vec{k}$, and the spatial vector, $\vec{r}$ within the first (incident) medium as

$$\vec{H}(\vec{r},t) = \vec{u}H_0 e^{-j(\vec{k_1}\cdot\vec{r} - \omega t)} \tag{3.1}$$

with its corresponding electric fields expressed as

$$\vec{E}(\vec{r},t) = \left(\frac{-j}{\omega\epsilon_1}\right)\nabla\times\vec{H}(\vec{r},t) \tag{3.2}$$

Since the system is now two-dimensional in the x-z plane, the y-dependence in the exponential term vanishes. Moreover, the time-dependence of the electromagnetic waves will be assumed in each of the following chapters, although not explicitly expressed during the following derivations. The above equation has been written with a positive imaginary component. This is done to ensure the decaying nature of waves associated with complex wavevectors when defined as the positive sum of imaginary components.

Since the propagating wave is assumed to be TM_y polarized, Equation 3.1 may be rewritten as

$$\vec{H}(x,z) = \begin{pmatrix} 0 \\ H_0 \\ 0 \end{pmatrix} e^{-j(k_{1x}x + k_{1z}z)} \tag{3.3}$$

where k_{1x} and k_{1z} are the x- and z-components of the incident field wavevector within the first medium. These are often simplified to the spatial component representation related to the incident angle where $k_{1x} = k_1 \sin\theta$, and $k_{1x} = k_1 \cos\theta$, with $|k_1|$ being the magnitude of the total wavevector in the first medium. However, this representation becomes irrelevant when describing complex wavevectors, resulting in complex transmission angles. Therefore, the x- and z-components of the wavevector will remain in order to preserve the correct methodology in deriving reflection and transmission coefficients later on.

The multiplying factor of H_0 in front of the oscillating wave multiplies only the y-direction, as determined by the polarization, with its magnitude being some constant value, often quoted as unity. Thus, there is only a y-directed magnetic field, which is expressed as

$$H_y(x,z) = H_0 e^{-j(k_{1x}x + k_{1z}z)} \tag{3.4}$$

Per previous derivations, the associated electric field may be expressed as

$$\vec{E}(x,z) = \left(\frac{-j}{\omega\epsilon_1}\right) \nabla \times \vec{H}(x,z) \tag{3.5}$$

Since this magnetic field intensity is only represented in the y-direction, the resulting electric field intensity becomes

$$\vec{E}(x,z) = \left(\frac{1}{\omega\epsilon_1}\right) \begin{pmatrix} -k_{1z} \\ 0 \\ k_{1x} \end{pmatrix} H_0 e^{-j(k_{1x}x + k_{1z}z)} \tag{3.6}$$

It is immediately clear from Equations 3.4-3.6 that the corresponding electric field intensity is always much larger than the electric field intensity, in the standard SI units. This is primarily due to the fact that a relationship between the wavenumber and frequency of the electromagnetic field are related by a factor of the speed of light. Since this quantity is large, the difference in intensities between the magnetic and electric fields is also large.

The complete set of electric and magnetic fields intensities are now known. These are provided in Table 3.1.

Table 3.1: Plane Wave Fields for TM_y Polarization

Component	Electric Field	Magnetic Field
$\hat{x}$	$\left(\frac{-k_{1z}}{\omega\epsilon_1}\right) H_0 e^{-j(k_{1x}x + k_{1z}z)}$	0
$\hat{y}$	0	$H_0 e^{-j(k_{1x}x + k_{1z}z)}$
$\hat{z}$	$\left(\frac{k_{1x}}{\omega\epsilon_1}\right) H_0 e^{-j(k_{1x}x + k_{1z}z)}$	0

3.2 Finite Width Beams

Before the derivation of the reflection and transmission coefficients can take place, the topic of finite width beams will be addressed. Although plane waves do exist in the real world, they are often confined to small areas during analysis. Moreover, this is often the case for much theoretical work since for computational purposes, the area of numerical simulation must remain relatively small.

The fact that plane waves are to be simulated, since they are a solution to Maxwell's Equations, requires the necessity of the finite beam to be composed of plane waves. The standard Gaussian beam often quoted by many scholarly articles and publications is in fact, non-physical. That is, a Gaussian beam in the sense of a standard, normalized field intensity as related to the spatial domain does not itself satisfy Maxwell's Equations. This is due in part to the squared spatial coordinate term within the exponential. Therefore, the use of such an incident beam will be prohibited here.

However, this does allow for the Gaussian-weighted summation of plane waves, as proposed by Tran and Maradudin [39] and later referenced by Jandhyala [40, 41].

This method assumes a sufficient summation of plane waves, each incident from a single angle. Yet, each of these plane waves undergoes a Gaussian weighting scheme of their wavevectors, with the centralized wavevector having a weight of unity. A brief derivation of this finite beam method is as follows.

Consider a TM_y polarized plane wave incident upon a planar surface with zero azimuthal approach angle, and tangential wavenumber k_x. The incident beam is assumed to be incident from a centralized polar angel of θ_0. A The summation of a Gaussian distribution of plane waves is now performed, each weighted with respect to this central polar angle. Mathematically, the magnetic field representation can be expressed as

$$H_y(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_x=-k_1}^{k_1} e^{-j(k_x x + k_y y + k_z z)} \times e^{-((k_x - k_{x0})^2 + k_y^2)W^2/2} \tag{3.7}$$

where k_{x0} is the centralized wavevector in the x-direction, as given by the central polar angle, θ_0, as previous. The x-component wavevector is nominally stepped in constant incraments proportional to the wavelength, $K = 2\pi/\Lambda$, where Λ is often on the order of tens of wavelengths. Thus, the expression for the x-direction wavenumber is often written as

$$k_x \in \{-k_1, \ldots, K, \ldots, k_1\} \tag{3.8}$$

Here, the centralized y-component wavenumber has been suppressed as denoted previously. Moreover, the y-component wavevector is calculated over the remaining range in much the same way, only limited by the remainder available wavenumbers

$$k_y \in \left\{-\sqrt{k_1^2 - k_x^2}, \ldots, \sqrt{k_1^2 - k_x^2}\right\} \tag{3.9}$$

The final wavevector component, normal to the interface, is calculated as the remainder of the available total wavenumber.

$$k_z = \left\{ \sqrt{k_1^2 - k_x^2 - k_y^2} \right\} \tag{3.10}$$

The corresponding electric field components are calculated in much the same way, now with a multiplying coefficient necessary to maintain a field amplitude of unity at the center of the beam intersection with the interface.

$$E_x(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_x = -k_1}^{k_1} \frac{-k_z}{\sqrt{k_x^2 + k_z^2}} e^{-j(k_x x + k_y y + k_z z)} \times e^{-((k_x - k_{x0})^2 + k_y^2)W^2/2} \tag{3.11}$$

$$E_z(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_x = -k_1}^{k_1} \frac{k_x}{\sqrt{k_x^2 + k_z^2}} e^{-j(k_x x + k_y y + k_z z)} \times e^{-((k_x - k_{x0})^2 + k_y^2)W^2/2} \tag{3.12}$$

The fact that this finite width beam is composed of plane waves, which are inherently defined over all space, leads to a periodic Gaussian distribution. That is, the finite width of this system will repeat over every length, L. However, as long as the area of simulation remains smaller than this width, the beam will appear to be completely Gaussian.

A plot of a Gaussian-weighted sum of plane waves is provided in Figures 3.1 and 3.2. The centralized polar angle is set at 45°, and the beam is set to intersect the origin with unity amplitude. Thus, the incident beam appears to have a Gaussian weight, although it really is a summation of single plane waves. This finite beam will be recalled periodically within the following chapter as necessary.

With the incident electric and magnetic fields now well defined, the following sections will deal with the interactions of these TM_y polarized fields with various

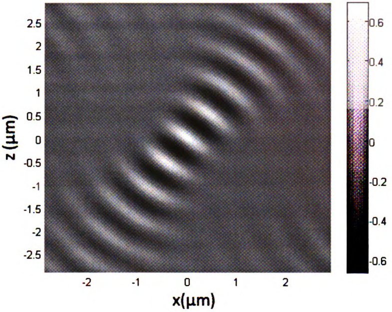

Figure 3.1: Real x-component of electric field for Gaussian-weighted sum of plane waves with maximum intensity of unity at the origin. Wave is assumed incident from 45° with wavelength 632.8nm.

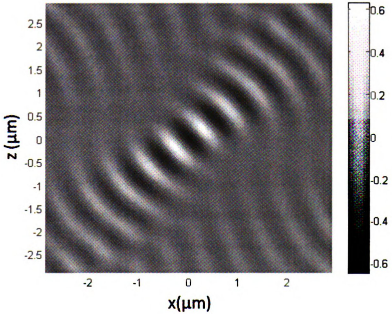

Figure 3.2: Real z-component of electric field for Gaussian-weighted sum of plane waves with maximum intensity of unity at the origin. Wave is assumed incident from 45° with wavelength 632.8nm.

boundaries. This will begin with the derivation of reflection and transmission coefficients for two adjacent regions of real permittivities, ϵ_1 and ϵ_2.

3.3 Reflection and Transmission

To begin, the simplest possible interactions will be explored. This involves investigating an electromagnetic field incident in one medium with permittivity ϵ_1 impinging upon an adjacent medium of different permittivity, ϵ_2. For the time being, it is assumed that both permittivities are completely real, as the derivation for complex permittivities will follow later. An electromagnetic wave with magnetic field vector aligned along the y-direction is now assumed incident on a planar interface separating two unlike permittivity regions, separated by the x-y plane. This system is shown in Figure 3.3. As with any quantum mechanical interactions, this electromagnetic

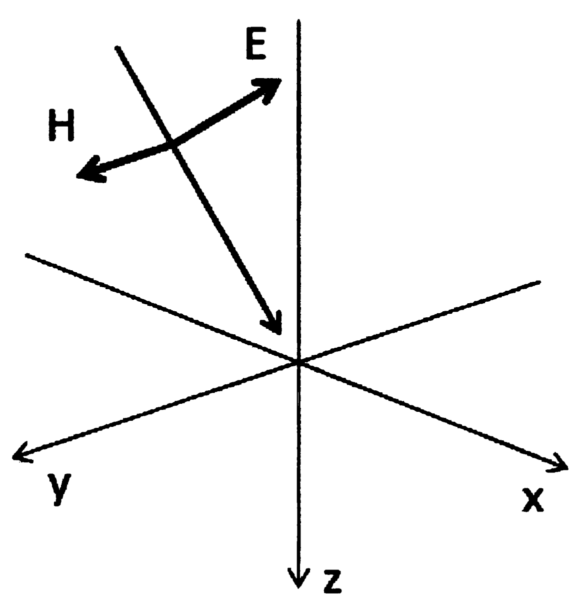

Figure 3.3: TM_y polarized electromagnetic wave incident upon a two-layer planar system with interface lying in the x-y plane.

wave will experience partial reflection and partial transmission through the boundary. Moreover, it is known that the total reflected and transmitted wave energy associated with the boundary must remain equal to the incident beam energy. Thus, the reflected and transmitted waves can be expressed mathematically as

$$\vec{H_{inc}} = \vec{H_R} + \vec{H_T} \tag{3.13}$$

where the magnetic field, H_{inc}, takes the form of equation 3.1 as

$$\vec{H}_{inc,y}(x,z) = H_0 e^{-j(k_{1x}x + k_{1z}z)} \tag{3.14}$$

Moreover, the magnetic fields within each medium can be expressed as the vector sum of their respective components

$$\begin{aligned} \vec{H_1} = \vec{H_{inc}} + \vec{H_R} &= \hat{y}\left(H_0 e^{-j(k_{1x}x + k_{1z}z)} + H_R e^{-j(k_{1x}x - k_{1z}z)}\right) \\ \vec{H_2} = \vec{H_T} &= \hat{y} H_T e^{-j(k_{2x}x + k_{2z}z)} \end{aligned} \tag{3.15}$$

Since there is no variation of the physical system along the x- or y-axes, the electric and magnetic fields lying in the x-y plane must be equal in magnitude, which means

$$\begin{aligned} H_{1y} &= H_{2y} \\ E_{1x} &= E_{2x} \end{aligned} \tag{3.16}$$

These fields can be written out at the $z = 0$ interface as

$$\begin{aligned} H_0 e^{-jk_{1x}x} &= H_0 e^{-jk_{2x}x} \\ \left(\frac{k_{1z}}{\omega\epsilon_1}\right) H_0 e^{-jk_{1x}x} &= \left(\frac{k_{2z}}{\omega\epsilon_2}\right) H_0 e^{-jk_{2x}x} \end{aligned} \tag{3.17}$$

From the Equation 3.16, the requirement that the adjacent tangential wavevectors must be equal may be extrapolated. This follows directly from the condition of equal wavevectors regardless of spatial tangential vector.

From the second part of Equation 3.16, plugging in for the first relationship results in the equation

$$\frac{k_{1z}}{\epsilon_1} = \frac{k_{2z}}{\epsilon_2} \tag{3.18}$$

Both of these expressions will prove vital later on during the derivation of the fields produced by, and conditions necessary for, for the enhanced excitation of plasmons.

Returning to the reflection and transmission coefficients derivation, the tangential magnetic and electric field components within each of the regions may be set equal to each other assuming real reflection and transmission coefficients multiplying magnitudes

$$\begin{gathered} H_0 e^{-jk_{1x}x} + rH_0 e^{-jk_{1x}x} = tH_0 e^{-jk_{2x}x} \\ \left(\frac{-k_{1z}}{\omega\epsilon_1}\right) H_0 e^{-jk_{2x}x} - r\left(\frac{-k_{1z}}{\omega\epsilon_1}\right) H_0 e^{-jk_{2x}x} = t\left(\frac{-k_{1z}}{\omega\epsilon_1}\right) H_0 e^{-jk_{2x}x} \end{gathered} \tag{3.19}$$

which may be simplified to solve for the magnetic filed reflection and transmission coefficients as

$$1 + r = t \tag{3.20}$$

$$\epsilon_2 k_{1z}(1 + r) = \epsilon_1 k_{2z} t \tag{3.21}$$

Solving Equations 3.20 and 3.21 for r and t yields

$$r_{TM} = \left(\frac{\epsilon_2 k_{1z} - \epsilon_1 k_{2z}}{\epsilon_2 k_{1z} + \epsilon_1 k_{2z}}\right) \tag{3.22}$$

$$t_{TM} = \left(\frac{2\epsilon_2 k_{1z}}{\epsilon_2 k_{1z} + \epsilon_1 k_{2z}}\right) \tag{3.23}$$

where the z-component wavevectors can *always* be calculated using the following identities along with Equation 2.11

$$k_{xi} = k_1 \sin\theta_1 \tag{3.24}$$

$$k_{zi}^2 = k_i^2 - k_{xi}^2 \tag{3.25}$$

in the absence of a y-directed wavenumber. Furthermore, the reflectance and transmittance are often calculated to represent the total reflected and transmitted energy across the boundary

$$R_{TM} = r_{TM}^* r_{TM} \tag{3.26}$$

$$T_{TM} = (1 - R_{TM}) = \left(\frac{k_{2z}}{k_{1z}}\right) t_{TM}^* t_{TM} \tag{3.27}$$

Note that the first part of Equation 3.27 is only valid for real dielectric media interfaces (in the absence of absorption). This equation will thus change when complex dielectric materials are addressed.

Below are figures plotting the reflectance and transmittance for various ratios of incident to transmitted layer permittivities ranging from 2 to 100 for an incident wavelength matching a Helium-Neon Laser of 632.8 nm. From the previous reflected

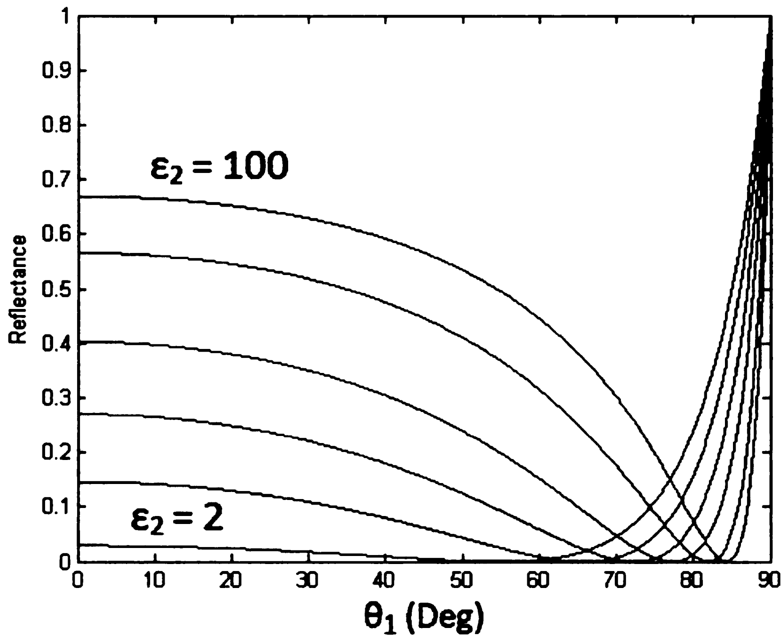

Figure 3.4: Reflectance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization

and transmitted magnetic field quantities, the corresponding electric field vectors may

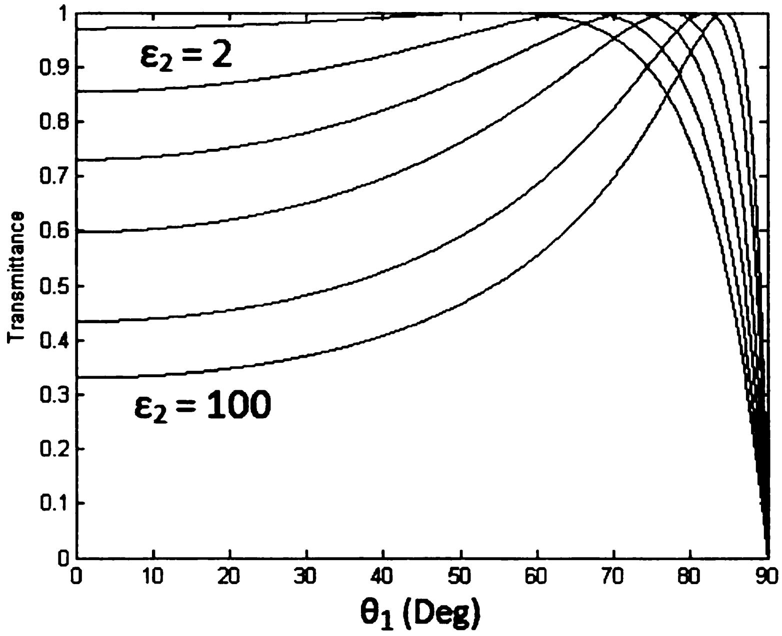

Figure 3.5: Transmittance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization

be derived. Again citing Maxwell's Equations as provided before, the electric fields within each medium can be calculated from their respective magnetic field vectors as

$$\vec{E_1}(x,z) = \begin{pmatrix} \frac{-k_{1z}}{\omega\epsilon_1} H_0 \left(e^{-j(k_x x + k_{1z} z)} + r_{TM} e^{-j(k_x x - k_{1z} z)} \right) \\ 0 \\ \frac{k_x}{\omega\epsilon_1} H_0 \left(e^{-j(k_x x + k_{1z} z)} - r_{TM} e^{-j(k_x x - k_{1z} z)} \right) \end{pmatrix} \tag{3.28}$$

$$\vec{E_2}(x,z) = \begin{pmatrix} \frac{-k_{2z}}{\omega\epsilon_2} H_0 \left(t_{TM} e^{-j(k_x x + k_{2z} z)} \right) \\ 0 \\ \frac{k_x}{\omega\epsilon_2} H_0 \left(t_{TM} e^{-j(k_x x + k_{2z} z)} \right) \end{pmatrix} \tag{3.29}$$

where the subscripts on the electric field vector denote the region number, with one representing the space $z < 0$ and two relating to $z > 0$. Moreover, the layer number subscripts on the x-component wavevectors have been dropped due to their equality across boundaries as proven earlier.

The reflection and transmission of electromagnetic waves is also apparent through the use of the Gaussian weighted sum of plane waves. Below is a visual representation of the x- and z-directed electric fields for a 632.8nm wavelength electromagnetic wave whose angle of incidence is centered about 45°. The relative permittivities of the two media are given by $\epsilon_1 = 1$ and $\epsilon_2 = 2$.

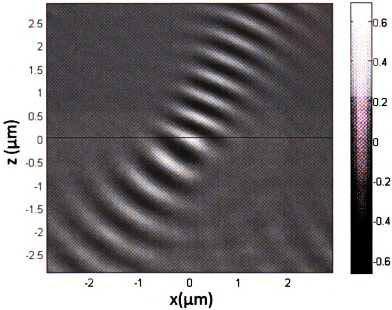

Figure 3.6: Reflection and transmission of real x-cmponent electric field for Gaussian-weighted Sum of Plane Waves for incident angle centered at 45°. $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization.

3.4 Total Internal Reflection

The previous electromagnetic derivations were all valid for systems with a plane wave incident from a medium of lower relative permittivity to a medium of higher permittivity. This results in a completely real transmission angle, θ_2, for any incident angle within the real range $(0^{\circ}, 90^{\circ})$. However, the condition of a wave travelling from a medium of higher permittivity to a lower one must also be dealt with. When this is the case, various combinations of relative permittivities and incident angles result

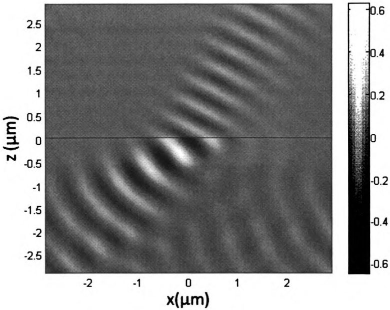

Figure 3.7: Reflection and transmission of real z-cmponent electric field for Gaussian-weighted Sum of Plane Waves for incident angle centered at 45^o. $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization.

in complex transmission angles. The transmitted field intensities will also become complex. Thus, the field derivations under such conditions must be established. Moreover, it is under these conditions in which SPs will be exist.

As with most electromagnetic derivations that start from Maxwell's Equations, the fields expressed are still valid under any circumstances. Therefore, for a TM_y polarized field propagating within a medium of higher permittivity than that post transmission across the adjacent boundary, it still holds that the fields above and below the interface must be equal. What follows will be the field derivations associated with total internal reflection.

Total Internal Reflection occurs when the angle of incidence is greater than the critical angle of the system. This can naturally occur for any system where the incident wave travels from a region of higher relative permittivity to a region of lower relative permittivity. The critical angle is defined as the incident angle resulting in a transmission angle of 90°. This is often stated citing Snell's Law as [38]

$$\theta_c = \sin^{-1}\left(\frac{\sqrt{\epsilon_2}}{\sqrt{\epsilon_1}}\right) \tag{3.30}$$

When the angle of incidence equals the critical angle of the system, the reflectance takes on the value

$$\left(\frac{\epsilon_2 k_{1z} - \epsilon_1 k_{2z}}{\epsilon_2 k_{1z} + \epsilon_1 k_{2z}}\right) = 1 \tag{3.31}$$

since there is no normal wavevector component for the transmitted field in the second region. Therefore, one would expect the transmission to equal zero, since the condition of $1 + R = T$ must hold across any planar boundary. However, by inspection of Equation 3.23, the transmission is determined to be not zero, but two.

This non-physical result hints at the interesting behavior of the fields under this unique condition. Figures 3.8 and 3.9 demonstrate this phenomenon showing the real and imaginary components of the transmittance for various ratios of permittivities over all available angles of incidence. Thus, it is clear from the above figures that the *coherent* transmittance of a plane wave is purely imaginary above the critical angle during total internal reflection, leading to zero real transmittance. Of particular interest is the transmitted field and its behavior within the second region.

3.5 Surface Waves

The preceding derivations dealt with angles of incidence equal to the critical angle of the system. From these, it was determined that the transmitted energy across the interface was imaginary, and resulted in a transmitted wave with magnitude equal to the two. However, the transmitted field is also of interest when the angle of incidence is *greater* than the critical angle.

For angles of incidence greater than the critical angle, one would expect the trans-

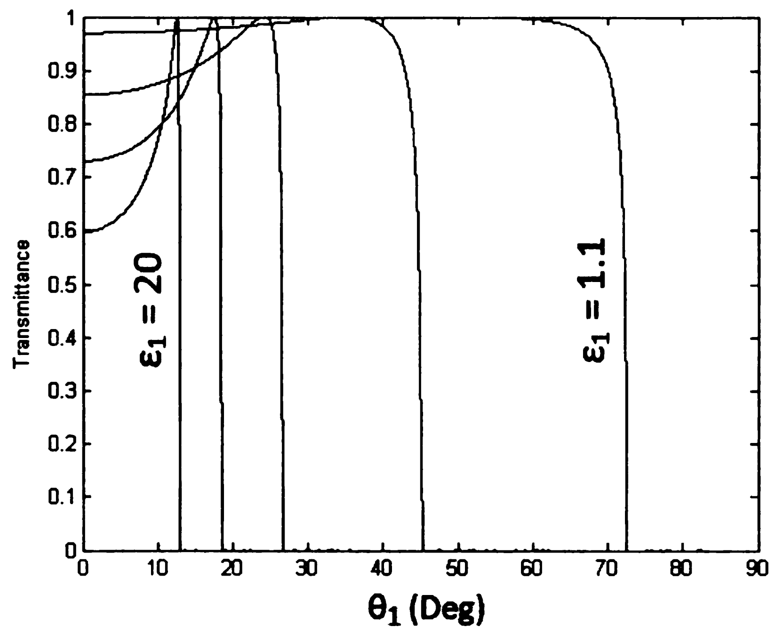

Figure 3.8: Real component of Transmittance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization

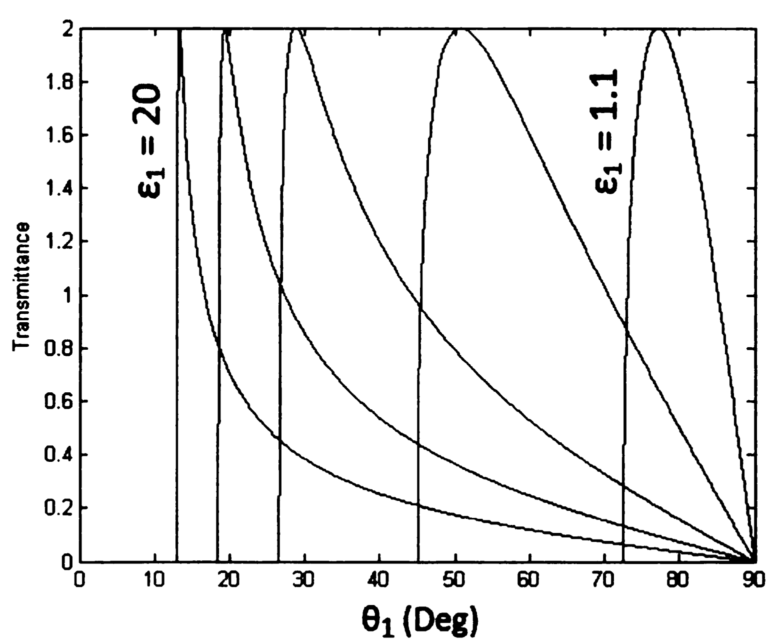

Figure 3.9: Imaginary component of Transmittance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization

mitted field to again result in a surface wave with real field magnitude at the interface and vanishing any finite distance away. However, this is not the case. When the angle of incidence is greater than the critical angle, the calculated z-component wavevector within the second region becomes complex. From this fact, it can be deduced that the field magnitudes of the transmitted waves present must decay away from the interface. Therefore, from the previous derivations, the wavevector components may be calculated in each region.

In the first region, or region of incidence, where $z < 0$, the wavevector components are calculated as before in terms of the plane wave wavenumber, k_1, using Equations 3.24 and 3.25. As is expected, both of these components take on real values, since the angle of incidence is always completely real. Similarly, the z-component of the wavevector within the second region is again calculated citing Equation 3.25, which, writing out in terms of the plane wave frequency, ω, and the relative permittivities becomes

$$k_{2z} = \sqrt{\left(\frac{\omega}{c}\right)^2 \epsilon_2 - \left(\frac{\omega}{c}\right)^2 \epsilon_1 \sin^2 \theta_1} \tag{3.32}$$

Removing the like terms from the root, this can be written as

$$k_{2z} = \frac{\omega}{c}\sqrt{\epsilon_2}\sqrt{1 - \frac{\epsilon_1}{\epsilon_2} \sin^2 \theta_1} \tag{3.33}$$

which becomes completely imaginary under the condition of total internal reflection. Thus, this is often rewritten as the product of the imaginary index and the positive real root

$$k_{2z} = jk_{2z}'' = j\frac{\omega}{c}\sqrt{\epsilon_2}\sqrt{\frac{\epsilon_1}{\epsilon_2} \sin^2 \theta_1 - 1} \tag{3.34}$$

where the positive multiplier is chosen to ensure decaying field magnitude away from the boundary in the positive z-direction.

Just as with the critical angle incidence, the propagation of the magnetic field within the second region is still completely in the x-direction. However, now there is a magnitude dependence of the field normal to the boundary. This is readily apparent when the magnetic field within the second region is expressed

$$\begin{aligned} H_{2y}(x,z) &= 2H_0 e^{-j(k_x x + jk_{2z} z)} \\ &= 2H_0 e^{k_{2z} z} e^{-jk_x x} \end{aligned} \tag{3.35}$$

where the imaginary unit j has been inserted to describe the imaginary nature of the z-component wavevector. As before, the direction of propagation remains parallel to the interface. Thus, the above expression still describes a plane wave. However, it now takes on the form of an *evanescent* plane wave, or one that decays in magnitude away from the interface. The rate at this field magnitude decays is directly related to the imaginary z-component wavevector.

As before, the electric fields which accompany these magnetic fields may also be derived. And once again, these are calculated via applying the correct application of Equation 2.4. A brief derivation of these associated electric fields is presented in the following equations.

From Equation 2.4, it is known that given some vector magnetic field, $\vec{H}$, the electric fields can be calculated as

$$\vec{E} = \left(\frac{-j}{\omega\epsilon}\right) \nabla \times \vec{H} \tag{3.36}$$

Thus, since the exact form of the vector magnetic field is well known, the vector electric field follows as

$$\begin{aligned} E_x(x,z) &= \frac{-2jk_{2z}''}{\omega\epsilon_2} \vec{H}_{2y}(x,z) \\ E_z(x,z) &= \frac{2k_x}{\omega\epsilon_2} \vec{H}_{2y}(x,z) \end{aligned} \tag{3.37}$$

Thus, these surface waves automatically take on the form of propagating waves with *lobes* evanescent behaviors normal to the interface. Moreover, the wavelength and phase velocity associated with these surface waves may be calculated.

The wavelength of the fields is related to the wavevector component lying in the direction of propagation. Thus, since this surface wave is travelling in the x-direction, the wavelength associated with the wave is related to the x-component wavevector

$$\lambda_{SW} = \frac{2\pi}{k_x} = \frac{2\pi c}{\omega\sqrt{\epsilon_1}\sin\theta_1} \tag{3.38}$$

where the subscript SW denotes the wavelength of the *surface wave.* The phase velocity at which the surface wave travels can also be calculated directly from this value and the initial plane wave frequency, ω, as

$$\vec{\nu_{SW}} = \frac{\omega}{k_x}\hat{x} = \frac{2\pi c}{\sqrt{\epsilon_1}\sin\theta_1}\hat{x} \tag{3.39}$$

This phase velocity in Equation 3.39 is slower than the propagation velocity of a similar coherent wave propagating within the second medium. Thus, because of this slower velocity than what is expected within this medium, these surface waves are sometimes referred to as *slow waves.* From the above derivations, it is clear these surface waves exist for *all* angles of incidence above the critical angle for the two-layer system.

3.6 Photonic Dispersion Relation

A commonly-derived expression often appearing in articles citing SPs is the Photonic Dispersion Relation. This equation provides the exact conditions required to couple photons to plasmons at the boundary of a planar system. However, it can be easily derived from the matching of tangential wavevector components across planar boundaries. For completeness, this expression is now derived while the topic of oblique angle incidence is pertinent.

From the above derivations, it is clear the x-component wavevectors must be matched across the interface separating the two media. Moreover, the normal component of the wavevector may be calculated using Equation 3.25. Since the tangential wavenumber must be constant across all boundaries, no indexing subscript is required, and this equation can be written as

$$k_{zi}^2 = k_i^2 - k_x^2 \tag{3.40}$$

where the subscript i denotes the referenced layer number. Equation 3.40 can be written in terms of the relative permittivities and angular frequency as

$$k_{zi}^2 = \left(\frac{\omega}{c}\right)^2 \epsilon_i - k_x^2 \tag{3.41}$$

Finally, from setting the tangential electric and magnetic fields equal to each other across the boundary, it can be proven that

$$\frac{k_{z1}}{k_{z2}} = \frac{\epsilon_1}{\epsilon_2} \tag{3.42}$$

Substituting Equation 3.42 back into Equation 3.40 results in

$$\left(1 - \frac{\epsilon_2^2}{\epsilon_1^2}\right) k_x^2 = \left(\frac{\omega}{c}\right)^2 \left(\epsilon_2 - \frac{\epsilon_2^2}{\epsilon_1}\right) \tag{3.43}$$

which can be solved for k_x as

$$k_x = \frac{\omega}{c}\sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_1 + \epsilon_2}} \tag{3.44}$$

which is often referred to as the plasmon wavevector, or

$$k_{sp} = \frac{\omega}{c}\sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_1 + \epsilon_2}} \tag{3.45}$$

The importance of the above relationship is its location of the enhanced fields associated with SPs. That is, the incident wavevector must be matched to this condition to result in the enhanced launching of plasmons.

After further review, it is clear that the *SP Angle*, or θ_{SP}, may be calculated for any two-layer system by setting the x-component wavevector equal to the SP wavevector as

$$\sqrt{\epsilon_1}\theta_{SP} = \sin^{-1}\left(\sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_1+\epsilon_2}}\right) \tag{3.46}$$

In order to satisfy the condition for real angles of incidence in two-layer systems, the ratio of $\epsilon_2 : (\epsilon_1 + \epsilon_2$ must be less than unity. Hence, this condition is always met for some real angle for every system consisting of two completely real dielectric regions. The relationship between the relative permittivity regions of each layer and θ_{SP} is shown in the figure below. It is also clear from the figure above that the SP Angle

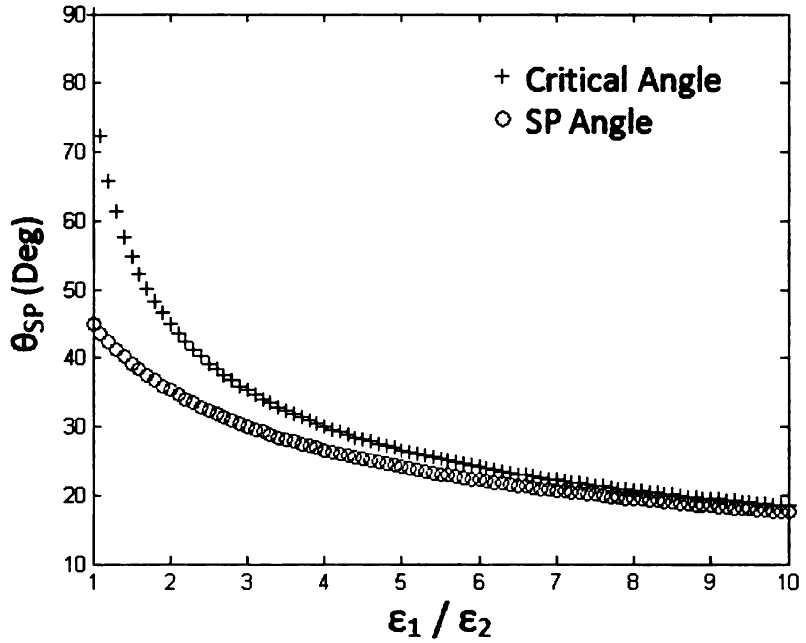

Figure 3.10: Dependence of the SP Angle, θ_{SP}, on the ratio of relative permittivities between the incident and adjacent media.

depends heavily on the critical angle of the system. Moreover, this angle is related to the intrinsic properties of the system, and not those of the incident plane wave.

Upon further inspection, one realizes that θ_{SP} is directly related to the critical angle of the system. Since the existence of surface waves depends on the angle of incidence being greater than this critical angle, the launching of SPs insists $\theta_{SP} > \theta_c$ of the system. This also holds for the sinusoidal function of these two values. Therefore, the condition for launching SPs becomes

$$\sqrt{\frac{\epsilon_2}{\epsilon_1 + \epsilon_2}} > \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{3.47}$$

which means that

$$\sqrt{\epsilon_1} > \sqrt{\epsilon_1 + \epsilon_2} \tag{3.48}$$

which can clearly never be the case for two real dielectric mediums. Therefore, the introduction of lossy dielectrics, which in general contain complex dielectric functions, is required. Moreover, the introduction of complex dielectric constants may result in the inclusion of negative real components of the permittivity, which is also a possible solution to the above expression.

The above surface waves exhibit oscillatory behavior, decay rapidly in magnitude away from the interface, and propagate with a slower velocity than a normally-propogating wave. However, the aforementioned surface waves are not associated with any *free* charge oscillations. Recall that the definition of an SP is the oscillation of bound free charges to a surface or interface. Thus, since completely real dielectrics do not contain any of the said free charges, these waves are not considered true SPs. However, upon the introduction of a lossy, or metallic, dielectric transmission layer, the opportunity for free charge coupling does exist. This system will be explored in the following sections.

3.7 Surface Waves in Lossy Dielectrics

In a two-layer system, the introduction of a lossy, or complex, dielectric must occur within the second, or transmission, region. If this were not introduced in this region, the initially propagating wave would decay to negligible amplitude before reaching the interface, which would result in a trivial field solution of zero everywhere. Thus, the introduction of the complex dielectric into the second region is justified.

As hinted before, the dielectric function of the second region becomes complex. This results in wavevectors which are also complex in all directions, meaning the associated fields will decay in all directions. More on this will be provided later on. For now, the complex dielectric function within the second layer is written as

$$\epsilon_2 = \epsilon_2' - j\epsilon_2'' \tag{3.49}$$

where a single prime has been left to denote the real component, and a double-prime to signal the imaginary component of this dielectric function.

It is often the case that the relative permittivity of any material depends on the frequency of the wave travelling through it. However, this is even more relevant when dealing with a complex permittivity. Therefore, the permittivity function will often become a function of angular frequency

$$\epsilon_2(\omega) = \epsilon_2'(\omega) - j\epsilon_2''(\omega) \tag{3.50}$$

The real component of the above function is often well-defined or experimentally verified for many materials. However, the imaginary component is often determined analytically and can be related to the conductivity of the material, σ. This conductivity term will be visited once more when the discussion turns to semiconductor materials in the later chapters. However, for the time being, the complex permittivity will be calculated via the *dispersion relation*, as is the usual method. A length derivation of this relation is provided elsewhere [35]. Writing the complex permittivity now as a function of frequency and the plasma energy, this becomes

$$\epsilon_2(\omega) = \epsilon_2'(\omega) - j\frac{\sigma}{\omega} = 1 + \frac{\frac{n_e q^2}{\epsilon_0 m_e}}{\left(\omega_0^2 - \omega^2\right) + 2j\omega\alpha} \tag{3.51}$$

where α is the absorption coefficient of the material and the fraction in the numerator is the square of the plasmon energy as derived in the first chapter. ϵ_0 takes on the relative permittivity value of the layer at the DC frequency. This is also known as the static permittivity of the region. With the complex permittivity now defined, the fields associated with the system may be discussed.

As before, the derivations carried out for the magnetic field reflection and transmission coefficients are still valid; the only change is the relative permittivity of the second layer. Thus, the magnetic field reflectance and transmittance will still take on completely real values for angles of incidence below the critical angle. Some of these conditions are shown in the figures below. It is clear from Figures 3.11 and 3.12 that the effect the imaginary component of the dielectric function has on the fields associated with the system comes in the form of a decrease in transmittance and corresponding increase in reflectance. Thus, as the imaginary component of this dielectric function increases, the field within the second, or lossy, region decreases accordingly.

However, as with the previous sections, the fields of interest are the surface waves associated with plane waves incident from oblique angles greater than the critical angle of the system. Yet, the calculation of the critical angle of a dielectric-metallic system is frivolous. That is, since the dielectric function of the second medium is complex, the calculation of the critical angle also results in complex values. Since the critical angle must be a real value, the critical angle calculation becomes non-existent.

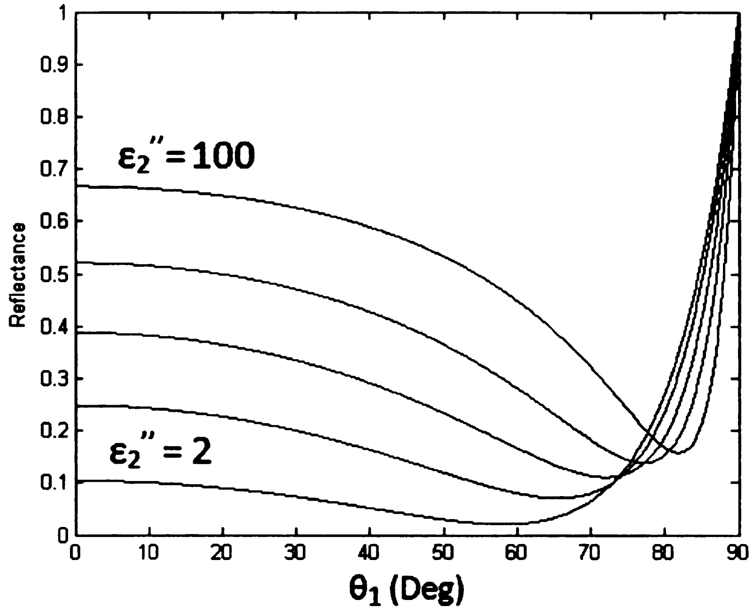

Figure 3.11: Reflectance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization with Complex Layer 2 Permittivity

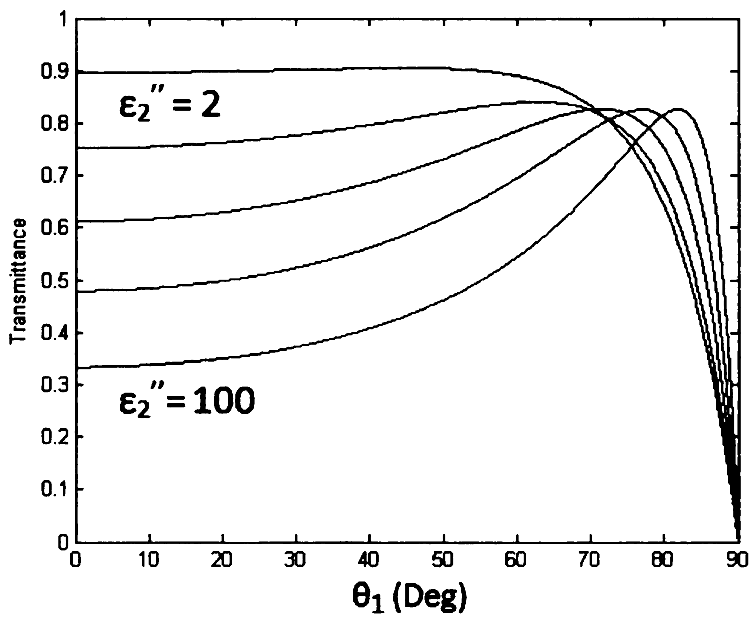

Figure 3.12: Transmittance Over All Incident Polar Angles for $\epsilon_1 = 1$ and $\lambda = 632.8nm$ for TM Polarization with Complex Layer 2 Permittivity

Therefore, the calculation of the transmitted fields within the second medium will proceed making no assumptions regarding the relative permittivities of the first region, nor the angle of incidence.

As before, the transmitted magnetic field within the lossy region regardless of angle of incidence within the first region, takes the form

$$H_{2y}(x,z) = T_{TM} H_0 e^{-j(k_x x + jk_{2z} z)} \tag{3.52}$$

where once again, the z-component of the wavevector in the second region is complex. However, the angle of propagation is no longer equal to 90°. Rather, this angle is calculated from the wavevector components in the x- and z-directions. As always, the x-component wavevectors must remain constant across the boundary. Furthermore, the z-component wavevector is again calculated in much the same way

$$k_{2z}^2 = k_2^2 - k_x^2 = \left(\tfrac{\omega}{c}\right)^2 \epsilon_2 - \left(\tfrac{\omega}{c}\right)^2 \epsilon_1 \sin\theta_1 \tag{3.53}$$

where the relative permittivity in region two, ϵ_2, is complex. Writing this formulation in long form and taking the square room yields

$$k_{2z} = \left(\left(\tfrac{\omega}{c}\right)^2 \left(\epsilon_2' - j\epsilon_2''\right) - \left(\tfrac{\omega}{c}\right)^2 \epsilon_1 \sin\theta_1 \right)^{1/2} \tag{3.54}$$

It is clear that this gives rise to real and imaginary components, due to the complex nature of the lossy dielectric region. Therefore, this may be rewritten as the sum of real and imaginary terms as

$$k_{2z} = k_{2z}' + jk_{2z}'' \tag{3.55}$$

Thus, the magnetic field within the lossy dielectric region is written as

$$H_{2y}(x,z) = T_{TM} H_0 e^{k_{2z}'' z} e^{-j(k_x x + jk_{2z}' z)} \tag{3.56}$$

This above expression results in, once again, a magnetic field intensity that decays exponentially away from the boundary in the perpendicular axis, or z-axis. However, unlike the surface waves described for the completely real dielectric systems, the wavevector associated with these fields contains both tangential and normal components. Thus, the magnetic field propagates at some intermediate angle, θ_2, within the lossy medium as given by

$$\sin\theta_2 = \frac{k_x}{\sqrt{k_x^2 + k_{2z}^* k_{2z}}} \tag{3.57}$$

where k_{2z}^* represents the complex conjugate of the z-component wavevector within the second medium.

Finally, the electric fields associated with the previously-derived magnetic fields may also be calculated. As before, these electric fields can be solved in terms of the magnetic field applying Equation 2.4.

$$E_x(x,z) = \frac{-j}{\omega\epsilon_2}\nabla \times \vec{H}(x,y) \tag{3.58}$$

Substituting in for the complex z-component wavevector, k_{2z}, and the complex dielectric function, the corresponding electric fields can be expressed as the summation of real and imaginary terms. After some algebra, these turn out to be

$$E_x(x,z) = -\left(\frac{E_x'}{\omega|\epsilon_2|^2} + j\frac{E_x''}{\omega|\epsilon_2|^2}\right) t_{TM} H_0 e^{k_{2z}'' z} e^{-j(k_x x + jk_{2z}' z)} \tag{3.59}$$

$$E_z(x,z) = \left(\frac{jk_x\epsilon_2^*}{\omega|\epsilon_2|^2}\right) t_{TM} H_0 e^{k_{2z}'' z} e^{-j(k_x x + jk_{2z}' z)} \tag{3.60}$$

where

$$E_x' = -\epsilon_2'' k_{2z}' - \epsilon_2' k_{2z}'' \tag{3.61}$$

and

$$E_z'' = \epsilon_2' k_{2z}' - \epsilon_2'' k_{2z}'' \tag{3.62}$$

Thus, both the tangential and normal electric field components are complex. This results in complex field magnitudes, which decay in both the x- and z-directions. Therefore, the electric fields associated with a two-layer lossy dielectric system lead to fields that are evanescent in nature in both the tangential and normal directions with respect to the interface.

Chapter 4

Scattering in One-Dimension: N-Layer Systems

The introduction of layers of finite thickness signals the next logical step in the complexity of plane wave propagation in electromagnetics. This will be performed using the matrix method, based upon the magnetic field and its derivative at each interface. This is different from several related matrix methods derived elsewhere, which are often based upon the magnetic field coefficients of independent linear solutions [42].

The matrix method is a rather straight-forward analysis system for solving reflectance and transmittance from many-layered one-dimensional systems. Instead of solving for the magnetic filed reflection and transmission coefficients for the entire system as a whole, the individual fields within each layer are solved recursively. From these, the overall magnetic field reflection and transmission coefficients are derived. This results in a much faster and simpler matrix method, since the recursion is based on known physical observables. This method is discussed in many sources, but the general method outlined here will be taken from Lekner [42].

4.1 Reflection and Transmission

Once again assuming TM_y polarization, the magnetic field consists of only a y-directional amplitude. Furthermore, it depends only on the variable z and satisfies the second order linear differential Equation 4.1 for any layer, n

$$\frac{d^2 H_{ny}(x,z)}{dz^2} + k_{nz}^2 H_{ny}(x,z) = 0 \tag{4.1}$$

where n is the n^{th} layer in the one-dimensional system. This is often referred to as the Helmholtz Equation in one dimension. Once again, this leads to solutions straddling the n^{th} interface of the form

$$H_y(x,z) = \begin{cases} e^{-jk_{nz}z} - r_{TM}e^{jk_{nz}z}, & z < 0 \\ \left(\frac{\epsilon_{(n+1)}}{\epsilon_n}\right)^{1/2} t_{TM}e^{-jk_{(n+1)z}z}, & z > 0 \end{cases} \tag{4.2}$$

Instead of writing the solution to Equation 4.2 as the sum of two exponential terms, a set of two coupled differential equations is used. This will allow the magnetic fields to be solved in terms of the layer thicknesses and physical constants in a recursive manner. Thus, the solutions of Equation 4.2 can be written as

$$\frac{1}{\epsilon_n}\frac{dH_y(x,z)}{dz} = C_y(x,z) \tag{4.3}$$

$$\frac{dC_y(x,z)}{dz} = -\frac{k_{nz}^2}{\epsilon_n} H_y(x,z) \tag{4.4}$$

Thus, both the magnetic field solution, $H_y(x,z)$, and its derivative, $C_y(x,z)$, are defined for each layer within the region (z_n, z_{n+1}).

Solving these equations is performed by setting the magnetic field and its derivative equal to H_{ny} and C_{ny}, respectively at the interface located at z_n. Thus, over the interval (z_n, z_{n+1}), the magnetic field and its derivative take the form

$$H_y(x,z) = H_{ny}(x,z_n)\cos(k_{zn}(z-z_n)) + \frac{C_{ny}(x,z_n)}{K_{zn}}\sin(k_{zn}(z-z_n)) \tag{4.5}$$

$$C_y(x,z) = -K_{zn}H_{ny}(x,z_n)\sin(k_{zn}(z-z_n)) + C_{ny}(x,z_n)\cos(k_{zn}(z-z_n)) \tag{4.6}$$

Thus, Equations 4.5 and 4.6 can be used to solve for the magnetic field and its derivative at the next interface using matrix formalism as

$$\begin{pmatrix} H_y(x,z_{n+1}) \\ C_y(x,z_{n+1}) \end{pmatrix} = \begin{pmatrix} \cos(k_{zn}(z-z_n)) & \frac{\sin(k_{zn}(z-z_n))}{K_{zn}} \\ -K_{zn}\sin(k_{zn}(z-z_n)) & \cos(k_{zn}(z-z_n)) \end{pmatrix} \times \begin{pmatrix} H_y(x,z_n) \\ C_y(x,z_n) \end{pmatrix} \tag{4.7}$$

where the 2×2 matrix described in Equation 4.7 becomes the *layer matrix* associated with each system layer. Thus, the n^{th} layer matrix may be written as

$$M_n = \begin{pmatrix} \cos(k_{zn}(z_{n+1}-z_n)) & \frac{\sin(k_{zn}(z_{n+1}-z_n))}{K_{zn}} \\ -K_{zn}\sin(k_{zn}(z_{n+1}-z_n)) & \cos(k_{zn}(z_{n+1}-z_n)) \end{pmatrix} \tag{4.8}$$

The reflection and transmission coefficients are present within the magnetic field equations for the first and last layers, 1 and N, respectively. Using Equation 4.8 for the magnetic field and its derivative as derived from Equations 4.5 and 4.6, the magnetic field and its derivative take the form within these boundary layers

$$H_{1y}(x,z) = e^{-j(k_x x + k_{1z}z)} - r_{TM}e^{-j(k_x x - k_{1z}z)} \tag{4.9}$$

$$C_{1y}(x,z) = jk_{1z}\left(e^{-j(k_x x + k_{1z}z)} + r_{TM}e^{-j(k_x x - k_{1z}z)}\right) \tag{4.10}$$

$$H_{Ny}(x,z) = t_{TM}e^{-j(k_x x + k_{1z}z)} \tag{4.11}$$

$$C_{Ny}(x,z) = jk_{Nz}t_{TM}e^{-j(k_x x + k_{1z}z)} \tag{4.12}$$

Using the layer matrices defined in Equation 4.8, these two corresponding expressions may be linked by the total product of the layer matrices, M, as given by

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \Pi_{n=1}^{N} M_n \tag{4.13}$$

as

$$\begin{pmatrix} t_{TM}e^{-jk_{Nz}z} \\ jk_{Nz}t_{TM}e^{-jk_{Nz}z} \end{pmatrix} = M \begin{pmatrix} e^{-jk_{1z}z} - r_{TM}e^{jk_{1z}z} \\ jk_{1z}\left(e^{-jk_{1z}z} + r_{TM}e^{jk_{1z}z}\right) \end{pmatrix} \tag{4.14}$$

In equation 4.14, the x-dependence of the spatial fields has been dropped for simplicity. Since the layer matrices are only determined from the physical constants of that specific layer, the reflection and transmission coefficients can be solved using the two linear equations provided above. After much algebra, these magnetic field reflection and transmission coefficients are derived as the following

$$r_{TM} = -e^{-2j(k_{1z}z_1)} \frac{K_{1z}K_{Nz}m_{12} + m_{21} - jK_{Nz}m_{11} + jK_{1z}m_{22}}{K_{1z}K_{Nz}m_{12} - m_{21} + jK_{Nz}m_{11} + jK_{1z}m_{22}} \tag{4.15}$$

$$t_{TM} = e^{-j(k_{1z}z_1 - k_{Nz}z_N)} \frac{2jK_{1z}}{K_{1z}K_{Nz}m_{12} - m_{21} + jK_{Nz}m_{11} + jK_{1z}m_{22}} \tag{4.16}$$

Finally, the reflectance and transmittance are calculated in much the same way as before citing Equations 3.26 and 3.27

$$R_{TM} = r_{TM}r^{*}_{TM} = |r_{TM}|^2 \tag{4.17}$$

$$T_{TM} = \left(\frac{k_{1z}}{k_{Nz}}\right) t_{TM}t^{*}_{TM} = \left(\frac{k_{1z}}{k_{Nz}}\right) |t_{TM}|^2 \tag{4.18}$$

Therefore, once the reflection and transmission coefficients are solved using the total layer matrix, M, the magnetic field intensities within each layer may be completely determined. Since the electric fields are the solutions of importance here, they may be calculated in much the same way within the two extreme layers as

$$\vec{E_1}(x,z) = \left(\frac{-j}{\omega\epsilon}\right)\begin{pmatrix} -jk_{1z} \\ 0 \\ jk_x \end{pmatrix}\left(e^{-j(k_x x + k_{1z}z)} - r_{TM}e^{-j(k_x x - k_{1z}z)}\right) \tag{4.19}$$

within the first layer (incidence layer), and in the last, or N^{th} layer, as

$$\vec{E_N}(x,z) = \left(\frac{-j}{\omega\epsilon}\right)\begin{pmatrix} -jk_{Nz} \\ 0 \\ jk_x \end{pmatrix} t_{TM}e^{-j(k_x x + k_{Nz}z)} \tag{4.20}$$

However, the calculation of the electric fields within the finite layer regions must be performed using the coupled magnetic field equations. Since only the derivative in the z-direction is relevant when using the coupled equations, the z-component of the electric field is calculated in much the same way. This means the z-component electric field takes on the form

$$E_{nz}(x,z) = \left(\frac{k_x}{\omega\epsilon_n}\right)e^{-j(k_x x + k_{nz}z)} \tag{4.21}$$

while the x-component of the electric field is related to the coupled field equations. Thus, the spatial derivative with respect to the z-direction can be written in terms of the expressions in Equations 4.5 and 4.6 as

$$\begin{aligned} E_{nx}(x,z) = & \left(\frac{j}{\omega}\right)K_{zn}H_{yn}(x,z)\sin(k_{zn}(z-z_n)) \\ & + \left(\frac{-j}{\omega}\right)C_{yn}(x,z)\cos(k_{zn}(z-z_n)) \end{aligned} \tag{4.22}$$

For instance, the magnetic field intensity and electric field vectors associated with a planar three-period system, with each period consisting of a 50nm thick layer of relative permittivity, $\epsilon = 3$, surrounded by 100nm-thick layers of relative permittivity ϵ_1 is illustrated in Figures 4.1 and 4.2. Thus, the first and N^{th} layers are assumed to have permittivity of unity and, in this case, $N = 7$. An incident angle of 45° is assumed, along with a free-space wavelength of 632.8nm. Just as was the case for

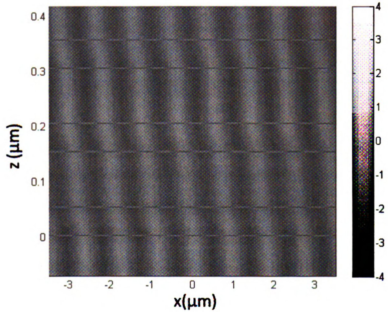

Figure 4.1: Magnetic field intensity for a three period system of ϵ = (1,3,1,3,1,3,1) with finite layer thicknesses of (50,100,50,100,50)nm. Note the x- and z-direction scales are not equal. The incident wave is impinging from the region $z < 0$, with a free space wavelength of 632.8 nm and an incident angle of 45°.

the two and three-layer system derivations, the above derivations are still valid for all angles of incidence and even systems involving complex permittivities. However, unlike the previous chapters, surface waves will now be confined in both the tangential and longitudinal directions due to the finite thickness of system layers.

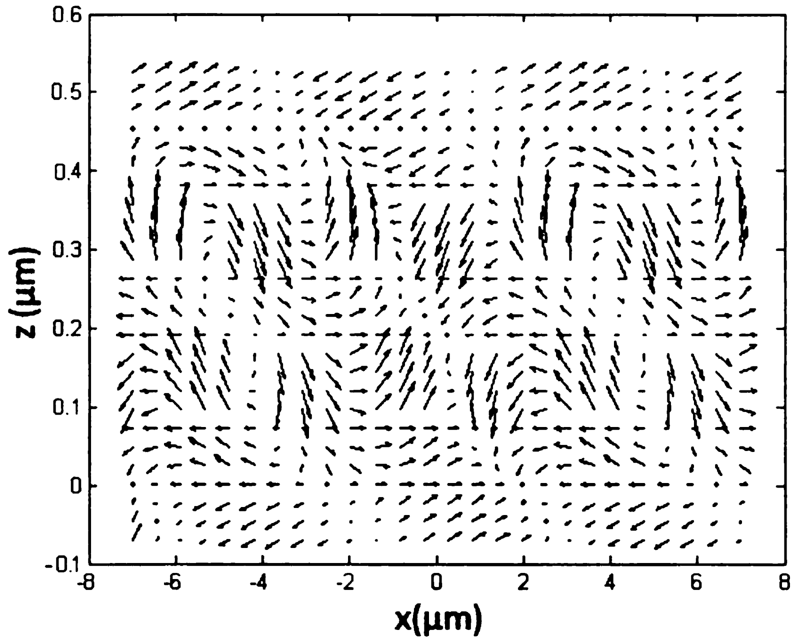

Figure 4.2: Electric field vectors for a periodic system of $\epsilon = (1,3,1,3,1,3,1)$ with finite layer thicknesses of (50,100,50,100,50)nm. Note as with the magnetic field, the x- and z-direction scales are not equal. The incident wave is impinging from the region $z < 0$, with a free space wavelength of 632.8 nm and an incident angle of 45°.

4.2 Finite Width Beams in N-Layer Systems

The introduction of finite width beams incident upon many-layer systems is now briefly discussed. The derivations of the fields associated with the Gaussian weighted sum of plane waves is performed in much the same manner as the previous chapter. However, the overall reflection and transmission coefficients must now first be calculated for all possible indexed wavenumbers and summed accordingly. Moreover, the wavenumbers within each layer for the various combinations of orthogonal wavevectors must be calculated. The system mentioned in the previous section is once again assumed as a periodic layer system consisting of three periods. The electric field component associated with this system are presented in Figures 4.3 and 4.4.

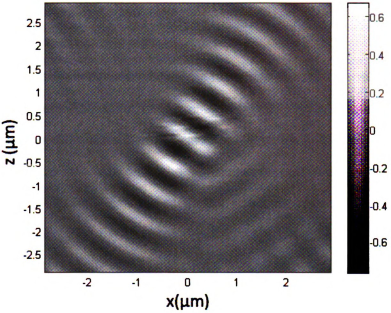

Figure 4.3: Real x-component electric field for Gaussian weighted sum of plane waves for the periodic system presented previously.

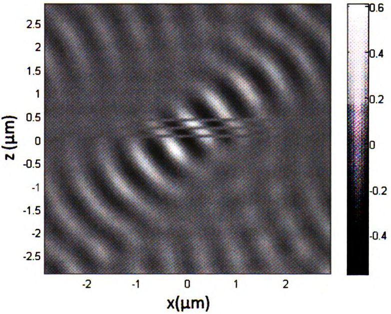

Figure 4.4: Real z-component electric field for Gaussian weighted sum of plane waves for the periodic system presented previously.

4.3 Kretschmann-Raether: Reflection, Transmission, and SPs

With the derivations now complete for the fields associated with many layered planar systems, a complete derivation of the usual reflection and transmission coefficients for one system in particular will be discussed. As proposed by Kretschmann and Raether more than 40 years ago, the so-named system consists of one finite lossy layer sandwiched between free space in the transmission regime, and a higher refractive index region, usually glass or prism, within the reflection regime. Again for reference, this system is observed in Figure 4.5. The often-quoted enhanced transmis-

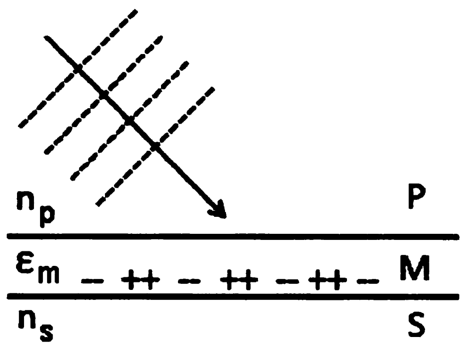

Figure 4.5: The Kretschmann-Raether Configuration for enhanced SP excitation.

sion field and zero reflectance at the SPR angle as given by the photonic dispersion relation are inherently due to the unique characteristics of the materials as much as the system configuration itself. At the Helium-Neon Laser wavelength of 632.8nm, gold has a complex permittivity of $\epsilon = -12.6 + \jmath 1.1$, or in terms of refractive index, $n = 0.16 + \jmath 3.55$. The reflectance and transmittance for this system over all available incident angles is presented in Figure 4.6. Notice that the transmittance drops to zero for all angles above the critical angle of the system, or 41.8°. Also, the reflectance does indeed fall to zero at the SPR angle of 43.6°. However, the enhanced transmission field magnitude that is often quoted using this system is not readily apparent in this plot. This is because the transmitted field amplitude is calculated as the trans-

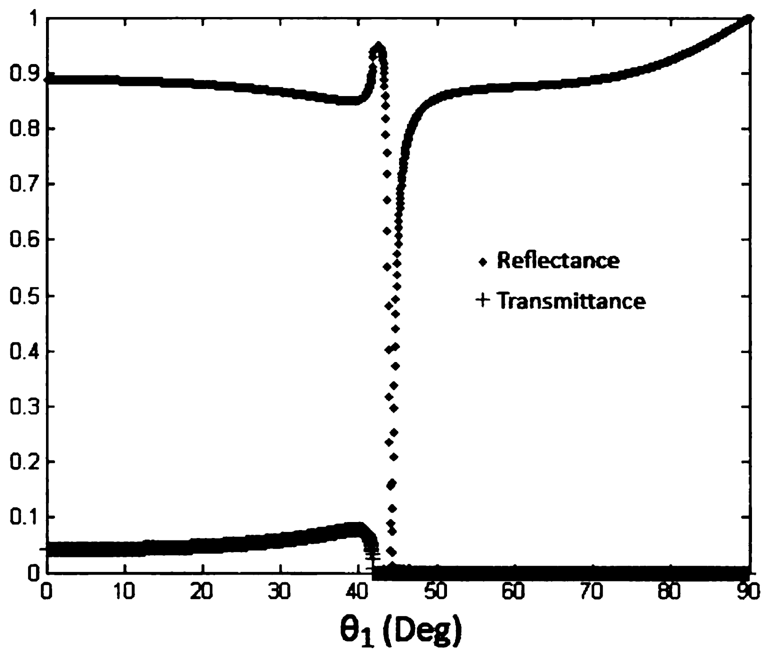

Figure 4.6: Reflectance and Transmittance for the Kretschmann-Raether configuration for all angles at incident wavelength 632.8nm and gold thickness 50nm.

mission coefficient, not the transmittance. The real and imaginary components of this transmitted field amplitude are plotted in Figure 4.7. It is clear from Figure 4.7 that the transmitted field becomes purely imaginary at the SPR angle. Moreover, for all angles above the SPR angle, both the real and imaginary components of the field become negative. Thus, no real energy is transmitted through the system into free space. Rather, the fields become complex and negative.

The Gaussian weighted sum of plane waves may also be carried over to the Kretschmann-Raether Configuration for the excitation of SPs. Moreover, the finite width nature of this excitation will allow for the direct observation of both the tangential and longitudinal extinction of these surface waves. The system presented in Figure 4.7 is once again simulated with the Gaussian weighted sum of plane waves. The centralized polar angle becomes the SPR angle of 43.6°. The electric field plots are presented in Figures 4.8 and 4.9. It is clear from Figure 4.9 that the enhanced transmitted field intensity is present at the immediate surface of the gold layer ex-

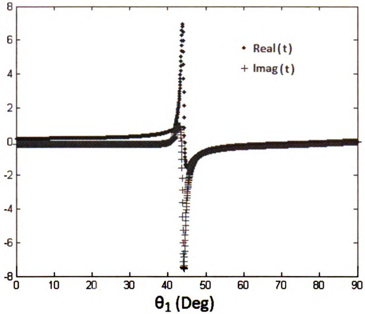

Figure 4.7: Real and Imaginary components of the transmitted electric field in the Kretschmann-Raether configuration for all angles at incident wavelength 632.8nm and gold thickness 50nm

.

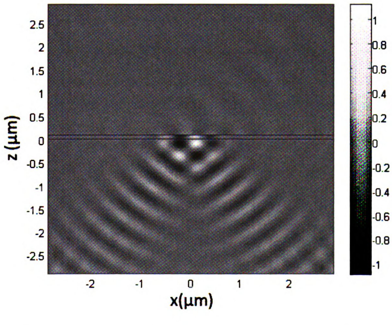

Figure 4.8: Real x-component electric field component for the Gaussian weighted sum of plane waves for system simulated the previous figure.

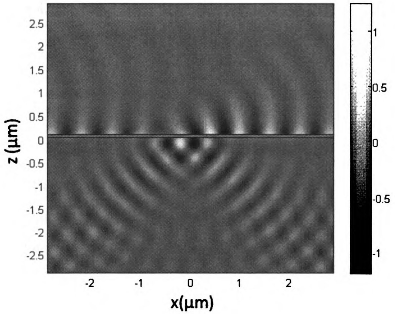

Figure 4.9: Real z-component electric field component for the Gaussian weighted sum of plane waves for system simulated in the previous figure.

tending into free space. Moreover, the x- and z-component evanescent nature of these SPs is also clearly observed.

4.4 Surface Waves in Layered Media

In order to observe surface waves within this N-layer system, such a system supporting surface waves will need to be produced. For simplicity, consider a system of four layers. The relative permittivities and thicknesses for each layer are presented in Table 4.4.

The support of surface waves within the final layer has already been derived and observed in the previous chapter. In order to avoid this same situation, the angle of incidence will be chosen in conjunction with the relative permittivities to result in surface waves that exist within both the second and third layers. These unique waves will be hence termed *intermediate surface waves*, in that they exist within the intermediate layers of the system. The angle of incidence is chosen to be 40°, which

Table 4.1: 4-Layer System supporting intermediate surface waves

Layer	d	ϵ
One	∞	10
Two	50nm	3
Three	50nm	2
Four	∞	1

is greater than the critical angles for both the second and third layer with respect to the first. The magnetic field intensity and electric field vector are observed in Figures 4.10 and 4.11 for this system. Note the difference in the lateral spatial resolution in

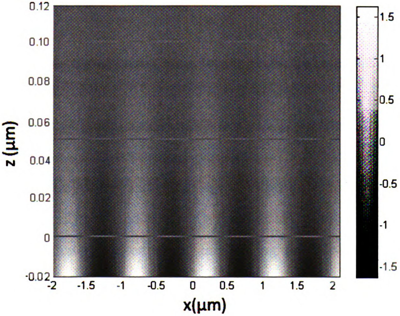

Figure 4.10: 4-Layer system supporting Intermediate Surface Waves - magnetic field intensity.

Figures 4.10 and 4.11. This was chosen in order to more clearly present the vector electric field components of the surface waves.

As is clear from the above plots, the magnetic field does support surface waves, while their evanescent nature remains intact. Thus, the existence of surface waves may exist within layered media at intermediate locations. Moreover, surface waves

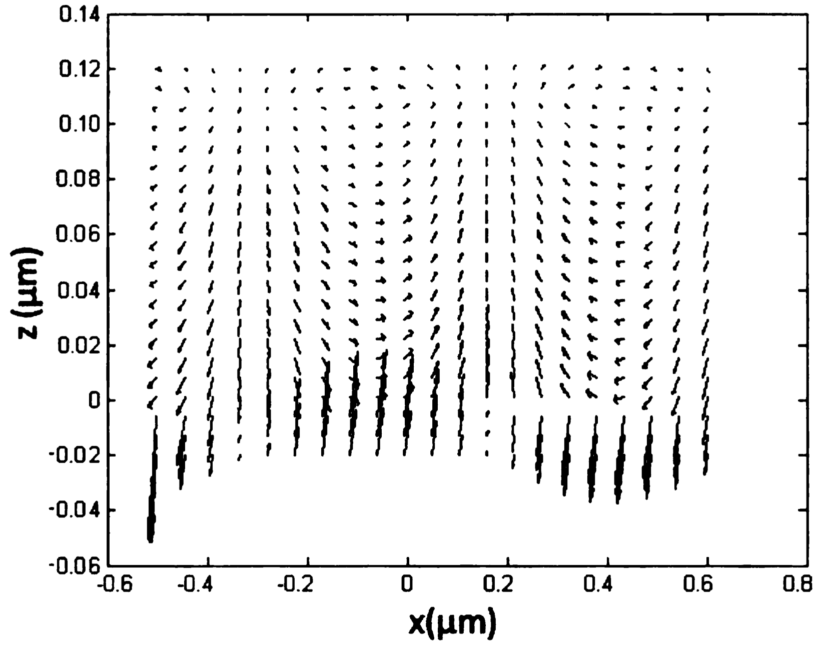

Figure 4.11: 4-Layer system supporting Intermediate Surface Waves - electric field vector.

are even shown to exist within several layers at once. This is clear from the evanescent and oscillatory wave patterns produced via the electric field vectors in Figure 4.11 that exist within the second and third layers. Therefore, the existence of surface waves is not necessarily confined to a single layer within a multi-layered system. This last statement will have elevated importance during the discussion of surface waves that can exist within 2DEGs, to be presented later on.

Chapter 5

Scattering in Two-Dimensions: Rigorous-Coupled Wave Analysis

One of the advantages of creating SPs at the interface of a one-dimensional (planar) system is that the fields produced are rather easily solvable. However, as pointed out earlier, SPs created at the boundary of planar systems leads to travelling slow waves. That is, these waves created will propagate along the interface separating the metal and dielectric at a speed of $v = \omega \sin\theta / k$. Here, θ is the angle of incidence measured from normal of the initial electromagnetic wave, and k is its wavenumber.

This propagation does not pose an immediate threat to device performance. In fact, Hashim, et. al. have proven this to actually enhance the performance of some high frequency devices [1]. However, the travelling nature of this wave does lead to spatially variant fields that propagates in time. This in turn could lead to poor device performance and reliability issues. Thus, in order to overcome this propagation of the wave, a periodic grating may be introduced. This essentially inhibits the travelling nature of the SPs through the creation of standing waves.

Standing waves are aptly named for their non-traveling (or standing) nature, while their creation is due to the discreteness of the grating region introduced. The tangential wavevector associated with any discontinuous permittivity will result in the formation of a discrete set of tangential wavevectors. Moreover, since the tangential wavevectors associated with any system must be continuous across interfaces, any system containing a grating region will result in a finite basis set of these wavevectors. The discrete wavevectors are determined based on the grating period, Λ, along with the free space wavelength and incident tangential wavevector of the incoming plane wave. More on this relationship will be presented later on.

Many derivations on the fields associated with electromagnetic waves in periodic gratings exist. Differential methods, integral methods, and coordinate transformation methods are well documented and are discussed elsewhere [43–45].

Here, the rigorous coupled-wave analysis (RCWA) will be discussed. Sometimes called the Fourier Modal Method (FMM), its dependence and derivation are related to the spectral transform coefficients associated with periodic structures. Here, the periodic structures are assumed to be periodic permittivity layers in the form of binary gratings. As always, a brief introduction into the background information required is provided.

5.1 History of Electromagnetic Diffraction

Lord Rayleigh was the first to introduce a quantifiable explanation for the diffraction observed by Woods in 1907 [46]. Essentially a Floquet mode matching method, the procedure is based on the simple idea of matching the spectral Fourier coefficients of the tangential fields across the grating region.

Since Rayleigh proposed his diffracted field methodology almost a century ago, several major improvements in the methods for solving the reflected and transmitted fields due to diffraction gratings have been proposed. Many of these have come about with the advent of the computer in the middle part of the twentieth century. Also, several of these methods were proposed within a few years of each other in the early 1980s.

One such computational method is the aptly named Chandezon Method, or C-Method (CM). Named for the 1981 paper by Chandezon, et. al. [44], this numerical method uses a coordinate transformation to solve for the fields within the grating system. The advantage to this method is its simplistic and straight forward implementation. It is perhaps today the most widely used numerical method for determining the exact fields due to grating diffraction. However, the *exactness* of the solution often requires tens of thousands of iteration loops, sometimes extending into the millions for TM polarization cases. Thus, unless the presence of a supercomputer or a parallel processor system is available, this method is often avoided due to its unnecessarily long computation times.

Various other matrix methods have been proposed over the years [45, 47–49]. Likewise, several integral and differential forms have provided exact solutions without the need to set an arbitrary number of terms to keep within the solution [50, 51]. However, many of these quoted matrix methods involve large computation times and require the definitions of an arbitrarily large number of terms. Moreover, many similar integral methods are often complex and difficult to implement in numerical practice to a large degree of accuracy and precision. Finally, some of these methods are inherently unstable and lead to large errors in the fields when solved for any set number of terms. Therefore, a simple method which employs the Floquet Mode matching scheme will be used here.

5.2 RCWA for Two-Layer Systems

The Rigorous Coupled-Wave Analysis (RCWA) method employs the use of matrices to solve for the electromagnetic fields. It does in fact promote the use of an arbitrary number of terms to keep in the infinite summation, but it also involves the exact solution to these fields. Therefore, the inclusion of an arbitrary number of terms simply results in the inclusion or exclusion of higher order terms inside the solution form.

The RCWA was first popularized by Moharam and Gaylord [52] in 1981, just one year after Chandezon proposed his C-Method. As hinted previously, these two methods are completely different in solving for the electromagnetic fields present within each diffraction region. RCWA employs the Floquet modal matching condition and matrix formalism to completely solve the electromagnetic fields within each grating region. That is, it essentially slices the grating region into very thin layers which can each be approximated by a rectangular grating. Thus, this methodology will work for *any* periodic grating profile, assuming it can be split into an arbitrary number of rectangular grating regions.

As was followed with the Rayleigh Expansion provided previously, the derivation will first begin assuming a rectangular grating region. The first necessity in the RCWA is to calculate the Fourier coefficients for the rectangular grating region. Since a rectangular grating will first be explored, the Fourier coefficients may be expressed as the permittivity coefficients at any point in the x-direction. Thus, this may be expressed as

$$\epsilon(x) = \sum_{n=-N}^{N} e^{2j\pi n/\Lambda} \tag{5.1}$$

where n is the coefficient to sum over, extending from some arbitrary term -N to N, and Λ is the period of the diffraction grating. Here, the diffraction grating takes on the two values of the permittivities of either region depending on the lateral position of x. Thus, once these Fourier coefficients, ϵ_h are determined, the fields may be solved as follows. The derivation for TM_y polarization will follow, but a similar derivation may be carried out for TE polarization as outlined elsewhere [53].

As with any two-dimensional plane wave, the incident magnetic field below the grating region (z ¡ 0) may be expressed as

$$H_{inc,y} = e^{(-jk_0\sqrt{\epsilon_1}(\sin(\theta)x + \cos(\theta)z))} \tag{5.2}$$

This means the total fields in each layer, above and below the diffraction grating region, respectively, are given by

$$H_{1,y} = H_{inc,y} + \sum_{n=-N}^{N} R_n e^{-j(k_{xn}x - k_{1,zn}z)} \tag{5.3}$$

$$H_{2,y} = \sum_{i=-N}^{N} T_n e^{-j(k_{xn}x - k_{1,zn}z)} \tag{5.4}$$

where the discrete tangential and normal wavevectors are defined as

$$k_{xn} = k_0\sqrt{\epsilon_1}\sin(\theta) - i\frac{\lambda_0}{\Lambda} \tag{5.5}$$

$$k_{l,zn} = k_0\sqrt{\epsilon_l - (\frac{k_{xn}}{k_0})^2}, l = 1, 2 \tag{5.6}$$

respectively, where the negative imaginary root is chosen when $k_0\sqrt{\epsilon_l} < k_{xi}$. Thus, the coefficients R_n represent the normalized reflected wave of the i^{th} order within the first region. Moreover, the coefficients T_n represent the normalized forward-diffracted waves in the second region. From these formula for the magnetic fields, the electric fields are simply calculated via Maxwell's Equations using Equation 2.4.

However, before calculating the electric fields in these two regions, the reflection and transmission coefficients at the various diffraction angles must first be calculated. This will essentially involve solving for the tangential fields within the grating region first. This is necessary for matching the Floquet modes across the grating for the various diffraction orders. Thus, the tangential fields inside the grating region (0 <

$z < d$) now take on the form

$$H_{gy} = \sum_{n=-N}^{N} U_{yn}(z) e^{-jk_{xn}x} \tag{5.7}$$

$$E_{gx} = j\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} \sum_{n=-N}^{N} S_{xi}(z) e^{-jk_{xn}x} \tag{5.8}$$

where the two functions of z, $U_{yn}(z)$ and $S_{xn}(z)$, represent the normalized amplitudes of the n^{th} order space-harmonic fields. It is clear that both of these functions are necessary for both H_{gy} and E_{gx} to satisfy Maxwell's Equations

$$\frac{\partial H_{gy}}{\partial z} = -j\omega\epsilon_0\epsilon(x)E_{gx} \tag{5.9}$$

$$\frac{\partial E_{gx}}{\partial z} = -j\omega\mu_0 H_{gy} + \frac{\partial E_{gx}}{\partial x} \tag{5.10}$$

Now substituting the previous expressions for E_{gx} and H_{gy} into the above Maxwell's Equations, the following matrix formula is resolved

$$\begin{pmatrix} \frac{\partial \vec{U_y}}{\partial z} \\ \frac{\partial \vec{S_x}}{\partial z} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{E} \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \vec{U_y} \\ \vec{S_x} \end{pmatrix} \tag{5.11}$$

where the $\mathbf{0}$ represents a null matrix of size $(1 \times 2N+1)$. Substituting the expression for the $\mathbf{U}_y$ matrix into what would be the second equation for $\mathbf{S}_x$, the matrix is reduced to

$$\left(\frac{\partial^2 \vec{U_y}}{\partial z^2}\right) = (\mathbf{EB})\,(\mathbf{U_y}) \tag{5.12}$$

where the matrix $\mathbf{B}$ is calculated as

$$\mathbf{B} = \mathbf{K_x}\mathbf{E}^{-1}\mathbf{K_x} - \mathbf{I} \tag{5.13}$$

where $\mathbf{K_x}$ is a diagonal matrix of size $(2N+1) \times (2N+1)$ with the diagonal coefficients being equal to

$$K_x(n,n) = \frac{k_{xn}}{k_0} \tag{5.14}$$

and $\mathbf{E}$ is a full matrix, also of size $(2N+1) \times (2N+1)$ with its coefficients defined by

$$E(n,p) = \epsilon_{n-p} \tag{5.15}$$

where p is the coefficient order and i is the diffraction order, and $\mathbf{I}$ is the identity matrix of size $(2N+1) \times (2N+1)$. Before proceeding any further, a note on the relative size of each matrix must be made. It must been assumed that the number of coefficients that have been kept in the field expansion is equal to the number of diffraction order terms kept for the field expansion. This will become readily apparent later in the derivation when the inversion of these matrices is required.

Furthermore, since the size of these matrices are $(2N+1) \times (2N+1)$, and the matrix $\mathbf{E}$ is calculated as the Fourier coefficient expansion of order $i-p$, this means the number of Fourier coefficients kept in the expansion of the binary grating will need to be twice this number, namely $(4N+2)$. Thus, the user must make sure to keep an inherently large number of Fourier coefficient terms during the numerical calculation process to avoid numerical and computational errors in light of accuracy.

The space harmonic electric and magnetic fields can now be written as

$$S_{xn}(z) = \sum_{m=1}^{n} v_{n,m} \left(-c_m^+ e^{-k_0 q_m z} + c_m^- e^{k_0 q_m (z-d)} \right) \tag{5.16}$$

$$U_{yn}(z) = \sum_{m=1}^{n} w_{n,m} \left(c_m^+ e^{-k_0 q_m z} + c_m^- e^{k_0 q_m (z-d)} \right) \tag{5.17}$$

where q_m and $w_{n,m}$ are the eigenvalues and eigenvectors of the matrix $\mathbf{EB}$, respectively and the matrix elements, $v_{n,m}$ are calculated via

$$\mathbf{V} = \mathbf{E}^{-1}\mathbf{W}\mathbf{Q} \tag{5.18}$$

Since $\mathbf{Q}$ is composed solely of eigenvalues, it becomes a diagonal matrix. With these matrices now defined, the coefficients c_m^+ and c_m^- are now calculated by matching the boundary conditions across the grating region.

Matching these tangential fields at both interfaces of the grating region, the system may be solved. At the first boundary of $z = 0$, these equations become

$$\delta_{n0} + R_n = \sum_{m=1}^{n} w_{n,m}\left(c_m^+ + c_m^- e^{-k_0 q_m d}\right) \tag{5.19}$$

$$j\left(\left(\frac{cos\theta}{\sqrt{\epsilon_1}}\right)\delta_{n0} - \left(\frac{k_{1,zn}}{k_0\epsilon_1}\right)R_n\right) = \sum_{m=1}^{n} v_{n,m}\left(c_m^+ - c_m^- e^{-k_0 q_m d}\right) \tag{5.20}$$

while at the second boundary

$$T_n = \sum_{m=1}^{n} w_{n,m}\left(c_m^+ e^{-k_0 q_m d} + c_m^-\right) \tag{5.21}$$

$$\left(j\frac{k_{1,zn}}{k_0 n_2^2}\right)T_n = \sum_{m=1}^{n} v_{n,m}\left(c_m^+ e^{-k_0 q_m d} - c_m^-\right) \tag{5.22}$$

Equations 5.19 and 5.22 are often written in matrix form, combining the respective equations at the two boundaries as

$$\begin{pmatrix} \delta_{n0} \\ j\delta_{n0}\cos\theta/n_1 \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ -j\mathbf{Z_1} \end{pmatrix}(\mathbf{R}) = \begin{pmatrix} \mathbf{W} & \mathbf{WX} \\ \mathbf{V} & -\mathbf{VX} \end{pmatrix}\begin{pmatrix} \mathbf{c}^+ \\ \mathbf{c}^- \end{pmatrix} \tag{5.23}$$

at $z = 0$ and at the second boundary

$$\begin{pmatrix} \mathbf{I} \\ j\mathbf{Z_2} \end{pmatrix} (\mathbf{T}) = \begin{pmatrix} \mathbf{WX} & \mathbf{W} \\ \mathbf{VX} & -\mathbf{V} \end{pmatrix} \begin{pmatrix} \mathbf{c}^{+} \\ \mathbf{c}^{-} \end{pmatrix} \tag{5.24}$$

at $z = d$. The above matrix equations are often solved simultaneously for the column matrix elements in $\mathbf{c}$ comprising the normalized Floquet mode fields within the grating. However, a more intuitive method of solving these equations can be performed by multiplying the first row of Equation 5.23 by $j\mathbf{Z_1}$ and Equation 5.24 by $-j\mathbf{Z_2}$, and performing matrix addition between the two rows. This effectively eliminates the column matrices $\mathbf{R}$ and $\mathbf{T}$ from the resulting two equations. Moreover, these may be combined into Equation 5.25.

$$\begin{pmatrix} j(\mathbf{Z_I}\delta_{n0} + \delta_{n0}\cos(\theta)/n_1) \\ 0 \end{pmatrix} = j \begin{pmatrix} \mathbf{Z_1W} + \mathbf{V} & (\mathbf{Z_1W} - \mathbf{V})\mathbf{X} \\ -(\mathbf{Z_2W} + \mathbf{V})\mathbf{X} & -\mathbf{Z_2W} - \mathbf{V} \end{pmatrix} \times \begin{pmatrix} \mathbf{c}^{+} \\ \mathbf{c}^{-} \end{pmatrix} \tag{5.25}$$

Clearly, the simple inversion of this multiplying matrix in Equation 5.25 will result in the normalized Floquet mode field coefficients within the grating region, given by the matrix $\mathbf{c}$. Using these filed values, the reflection and transmission coefficients for an arbitrary number of diffraction orders may be solved. Finally, from these coefficients, the fields may be matched across the boundaries, and both the tangential magnetic and electric fields are solved over all space.

One problem that quickly arises with RCWA is its numerical instability for large matrices. That is, the numerical precision of many computers introduces erroneous values whenever large, sparse matrices are inverted. One must keep this in mind when dealing with either a very large number of modes, or with sparse matrices with very small coefficients.

Another often witnessed problem with RCWA is the convergence of terms often requires the number of harmonic modes to reach into the tens or hundreds. Although this in itself poses no difficulty, the inversion of large matrices, especially ones which result in exceedingly small or large determinants, is sometimes beyond the capabilities or precision of the average personal computer. A more thorough discussion on the numerical stability of RCWA is clearly presented in the follow up article by Moharam and Gaylord [53]. However, the presentation here poses no such threat as the number of terms is often kept below 100.

As an example, a binary grating region will be simulated separating two layers, one of free space, and the other with a relative permittivity of four. A plane wave with wavelength 632.8nm and from polar angle 45° is assumed. Moreover, the grating constraints have been set to a period and thickness of one micron. Finally, the fill factor, or the time high, for the grating will be set to represent a perfect binary grating, or 0.5. The spatial electric fields components are presented below in Figures 5.1 and 5.2.

As stated previously, the convergence of the RCWA is often of great importance. That is, the system should converge toward single values as the number of Fourier modes kept within the solution increases. The convergence of the RCWA for the above system is given in Figure 5.3. Note that even for the TM_y case, which inherently requires more time to converge, the final answer for the zero order diffraction efficiencies is approximated with just ten iterations kept.

Moreover, the total sum of all diffraction efficiency orders, representing the total energy incident upon the system, must equal unity for completely real permittivities. In the above RCWA iterations, the total sum of all diffraction efficiencies after 50 iterations was calculated as 1.00000003. Thus, the accuracy of this method after 50 iterations produces an answer accurate to better than one part in ten million for the

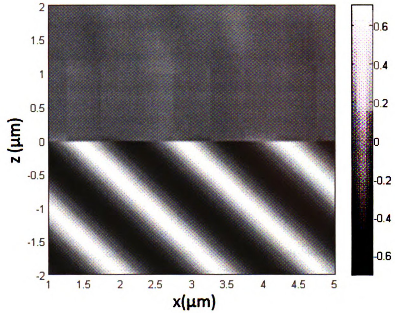

Figure 5.1: Real x-component electric field magnitude for binary grating of period and thickness one micron. Incident wave has 632.8nm wavelength and polar angle 45°.

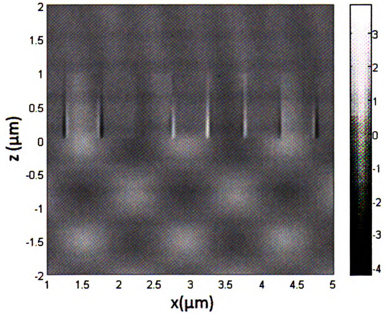

Figure 5.2: Real z-component electric field magnitude for binary grating of period and thickness one micron. Incident wave has 632.8nm wavelength and polar angle 45°.

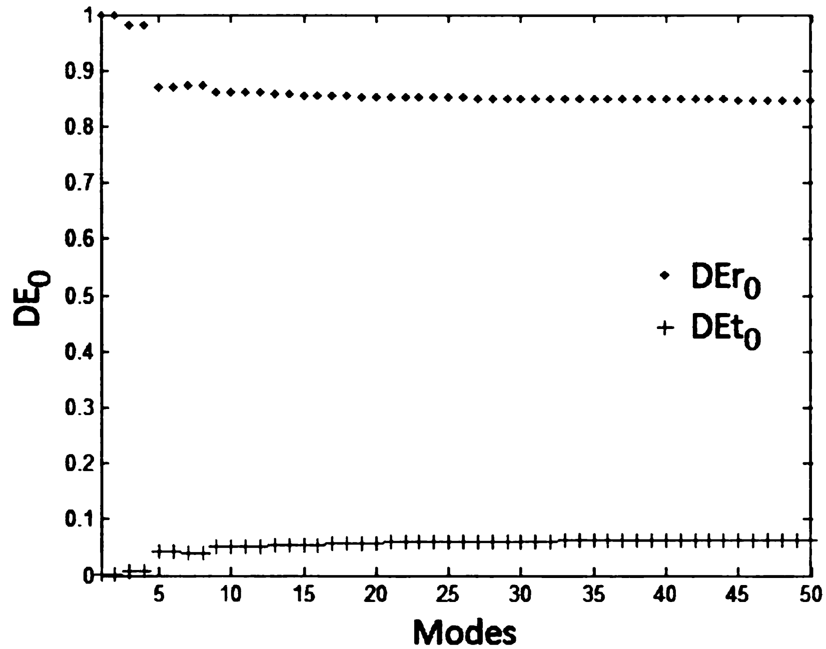

Figure 5.3: Convergence of the RCWA Zero Order Diffraction Efficiencies vs. Number of Fourier Modes kept in the solution. The system is described in the above figures.

above system.

For complex media, this diffraction efficiency sum must be less than unity. This is due to the lossy nature of one of the media within the grating region, resulting in a decreased field intensity across the grating region. For example, the system previously considered is once again simulated, but the second medium now takes on the complex permittivity of gold. The resulting electric fields and convergence of the solution are presented below in Figures 5.4-5.6. For 50 kept Fourier modes, the sum of all diffraction efficiencies for the above system is 0.537. Thus, it is clear that almost half of the energy associated with the incident electromagnetic wave is lost within the grating region.

Another system of interest involves the production of SPs via the total internal reflection. Consider once again the first system simulated for the RCWA method as presented in Figures 5.1 and 5.2. However, the two media are now interchanged, with the electromagnetic wave incident from within the medium of relative permittivity

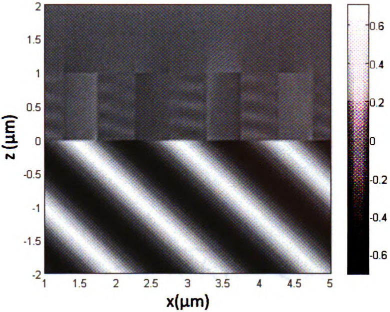

Figure 5.4: Real x-component electric field magnitude for binary grating of period and thickness one micron. Layer 2 is assumed as gold. Incident wave has 632.8nm wavelength and polar angle 45°

.

four, and the transmitted region is free space. Also assume an incident angle of 50° so to ensure total internal reflection. This system is simulated in Figures 5.7-5.9 below. Notice in the x-directed electric field plot the existence of resonance fields situated at the corners of the grating separating the planar free space region from the grating in Figure 5.8. These are attributed to the matching of the zero order tangential wavenumber with that of the grating wavenumber. Moreover, these become standing waves, which do not travel along the free space-grating interface as was presented previously for the planar systems. Thus, these standing waves simply oscillate in time. This is exactly the benefit of introducing gratings into the systems of consideration; they may inhibit the maximum transmission of energy, but they do create standing waves.

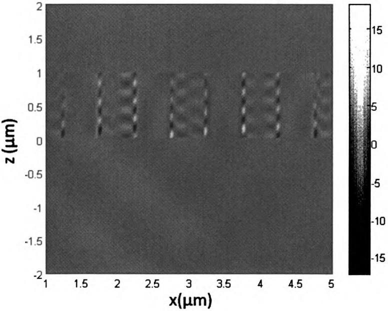

Figure 5.5: Real z-component electric field magnitude for binary grating of period and thickness one micron. Layer 2 is assumed as gold. Incident wave has 632.8nm wavelength and polar angle 45°

.

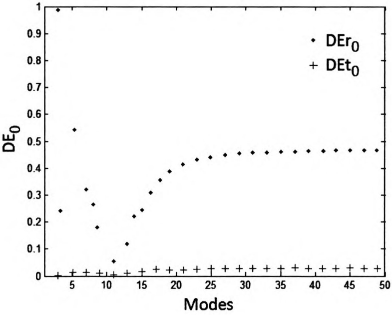

Figure 5.6: Convergence of the RCWA Zero Order Diffraction Efficiencies vs. Number of Fourier Modes kept in the solution. The system is described in the above figures

.

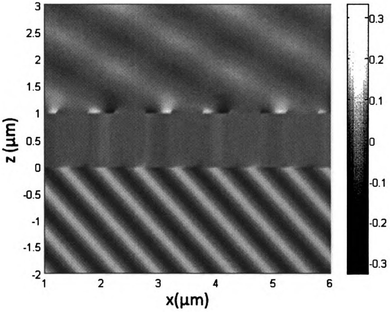

Figure 5.7: Real x-component electric field magnitude for binary grating of period and thickness one micron. Total Internal Reflection. Incident wave has 632.8nm wavelength and polar angle 50°

.

5.3 RCWA and Finite Width Beams

The introduction of standing waves within the grating structure improves the spatial resolution of the system. In order to enhance this even further, the usage of the Gaussian weighted sum of plane waves presented earlier may be applied. This will essentially confine the region of interest.

As previously derived, the Gaussian weighted sum of plane waves involves choosing a set of wavevectors along each orthogonal coordinate within the boundaries of the incident wave. However, the reflected and transmitted fields, along with the fields inside the grating, now take on the discrete values as related to the grating period. Mathematically, this results in a new summation over the tangential wavevectors.

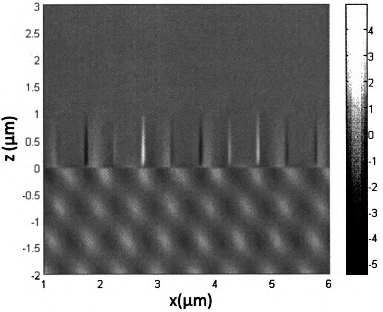

Figure 5.8: Real z-component electric field magnitude for binary grating of period and thickness one micron. Total Internal Reflection. Incident wave has 632.8nm wavelength and polar angle 50°

.

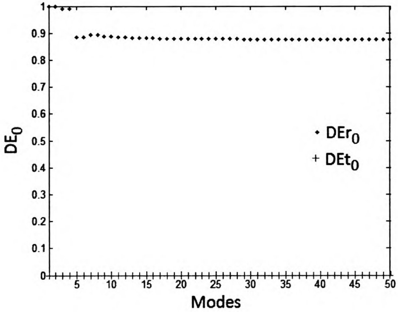

Figure 5.9: Convergence of the RCWA Zero Order Diffraction Efficiencies vs. Number of Fourier Modes kept in the solution. System is described in the above figures

.

While the incident magnetic field takes on the same formulation as before,

$$H_{y0}(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_x=-k_1}^{k_1} e^{-j(k_x x + k_y y + k_z z)} \times e^{-((k_x - k_{x0})^2 + k_y^2)W^2/2} \tag{5.26}$$

the reflected magnetic field is calculated via

$$H_{yr}(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_{xn}} e^{-j(k_{xn} x + k_y y - k_{1zn} z)} \times e^{-((k_{xi} - k_{x0})^2 + k_y^2)W^2/2} \tag{5.27}$$

where these tangential wavenumbers, k_{xi}, are now calculated as

$$k_{xi} = k_x - iK \quad i \in (0, \pm 1, \pm 2, \ldots, \pm N) \tag{5.28}$$

and $K = 2\pi/\Lambda$. The y-directed wavenumber is derived just as before, with the normal component calculated so as to ensure the positive square root for complex values.

In a similar manner, the transmitted electric fields are calculated via

$$H_{yt}(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_{xn}} e^{-j(k_{xn} x + k_y y - k_{2zn} z)} \times e^{-((k_{xn} - k_{x0})^2 + k_y^2)W^2/2} \tag{5.29}$$

where

$$k_{2zn} = \sqrt{k_2^2 - k_{xn}^2 - k_y^2} \tag{5.30}$$

Inside the grating region, the magnetic field is summed over both the diffraction orders and the sum of plane waves. This leads to a two-dimensional sum of the form

$$H_{yg}(x,z) = \frac{2\pi W^2}{L^2} \sum_{k_{xn}} U_y e^{-j(k_{xn}x + k_y y - k2_{zn}z)} \times e^{-((k_{xn} - k_{x0})^2 + k_y^2)W^2/2} \tag{5.31}$$

where the modal field amplitudes, U_{yi} are calculated by summing over all of the diffraction order fields calculated from the eigenmodes of the grating

$$U_{yn} = \sum_{m=1}^{N} W_{n,m} \left(c_m^+ e^{-k_0 q_m z} + c_m^- e^{k_0 q_m (z-d)} \right) \tag{5.32}$$

where the normalized z-component wavenumber, q_m, eigenvalues $W_{n,m}$, and field order amplitudes, c_m^+ and c_m^-, all depend on both the diffraction order along with the wavevector associated with each plane wave.

The corresponding electric fields can be derived using Equation 2.4. A visual representation of the electric fields associated with the initial system simulated in Figures 5.1 and 5.2 are presented below in Figures 5.10 and 5.11. Thus, the fields associated with diffraction gratings have been completely simulated via the RCWA method in conjunction with a Gaussian weighted sum of plane waves. Although the formation of standing waves does occur, the fields associated with these systems involve much lower transmitted field intensities. Therefore, their involvement in the following chapters will not be carried out.

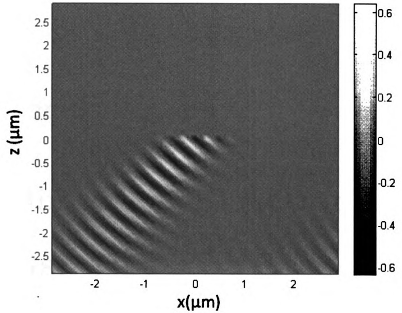

Figure 5.10: Real x-component electric field for system presented in Figures 5.1 and 5.2.

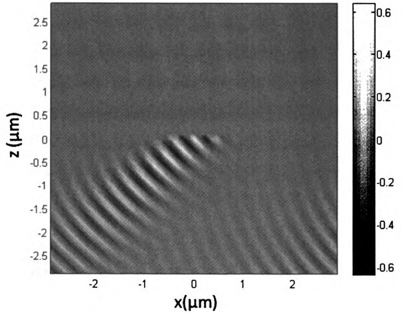

Figure 5.11: Real z-component electric field for system presented in Figures 5.1 and 5.2.

Chapter 6

Two-Dimensional Electron Gases (2DEGs)

With the derivation of the fields associated with SPs now available, the theory behind the formation of electron gases may now be carried out. However, the following derivations will depart, for a time away, from electromagnetic theory. That is, the physics describing the formation of electron gases is based on solid state engineering and quantum physics rather than electromagnetics. Therefore, the derivation of the material properties associated with these systems will be provided from a devices engineering standpoint rather than an electromagnetic perspective. However, once the material properties have been fully derived, the electromagnetic interactions between electron systems and SPs will be explored. This will be performed in the following chapter.

6.1 Defining QWs and 2DEGs

For much the same reasons as quoted within the first chapter, the definitions of various solid state engineering terms are provided as a benefit to the reader unfamiliar with such a lexicon. Moreover, these definitions are also provided in order to distinguish between SPs and 2DEGs, as they are inherently similar.

A Quantum Well (QW) is defined as a nano scale region of space in which the three-dimensional motion of a particle is essentially confined to smaller dimensions. Often quoted for one or three dimensions, these regions of confinement are often only several nanometers long in one-dimension. This confinement is often the result of an external potential experienced by the particle. Often applied to the negatively-charged electron, QWs are often found in specific semiconductor devices and junctions.

Two-Dimensional Electron Gases or Systems (2DEG, 2DES) are often defined as quantum wells for electrons. They encompass very thin (often less than 5nm thick) layers of essentially free electrons residing at the surface, or boundary, of a homogeneous layer. One major difference between 2DEGs and the aforementioned *plasmons*, is that 2DEGs are comprised of completely negative charge, or electrons. Moreover, they do not necessarily oscillate coherently in space nor time. Therefore, there is no wavelength or frequency associated with 2DEGs, unlike their plasmon cousins. Sometimes referred to as *surface plasmas*, these systems are found within certain types of solid state devices and transistors.

One such device is the High Electron Mobility Transistor, often abbreviated as HEMT. Essentially a heterojunction transistor, it is comprised of a junction between a highly donor-doped wide-bandgap material and an intrinsic, or undoped, narrow-bandgap material. HEMTs are often utilized in high frequency or power applications, and have seen wide industrial implementation since their theoretical proposal several decades ago [54]. The formation of the 2DEG within a transistor is due to the conduction band bending upon alignment of its Fermi Levels. More on Fermi Levels and the device characteristics of HEMTs is provided later.

6.2 History of the 2DEGs and HEMTs

The term *electron gas* was first used by Sommerfeld to describe the random nature of electron motion within a metal [55]. However, its first implementation in the world of electromagnetics came twenty years later. As it turns out, it was the same Bohm and Pines who first described the plasmonic oscillations in a metal that used the term *electron gas* to describe the free electrons within a *degenerate* metal [9]. Moreover, their treatment of the system in terms of the relatively new quantum mechanical approach would lay the groundwork for the semiconductor implementation later on.

The first to refer to these free electrons within a two-dimensional slab system was Ferrell in 1957 [56]. However, he did not coin the complete term. This would finally be done by Fowler, et. al. in 1967 at IBM [57]. Fowler's work probing magnetic-based carrier oscillations were performed using a semiconducting material, in this case silicon. Although Ferrell et. al. assigned this term to the surface of the silicon, the identification of the two-dimensional electron gas, along with its plasma-like charge oscillations under external field interaction, was complete.

The following history and development of the HEMT has been paraphrased from Sze and Ng [54]. A more detailed and personal presentation of this history can be found elsewhere [58].

Only two years after the naming of the 2DEG by Ferrell et. al., another team of IBM researchers theorized the transport of electrons and holes parallel to the junction interface of semiconductors. In 1969 it was Esaki and Tsu who proposed the modern theory of what would become the HEMT [59]. This concept was completely unheard of in the field of solid state physics at that time. That is, until then, the transport of carriers within any semiconductor had been almost completely focused on the idea of propagation through the interfaces. For example, the simplest of semiconductor devices, the $p-n$ junction, works via transporting carriers through the boundary separating the two doped materials. However, the technology to produce such a device was not yet available. Over the next decade, the idea remained, while the slow growth technologies of Molecular Beam Epitaxy (MBE) and Metallorganic Chemical Vapor Deposition (MOCVD) were developed and refined.

Thus, it wasn't until 1978 when Dingle et. al. demonstrated experimentally the enhanced mobility in the AlGaAs/GaAs modulation-doped superlattice [60] , as theorized by Esaki and Tsu almost a decade earlier. Moreover, similar results citing a drastic increase in the mobility within a AlGaAs/GaAs semiconductor were presented by Stormer et. al. later that next year [61]. However, both of these studies were conducted using HEMTs that didn't have the benefit of a control gate. That is, both of these articles cited the increased mobility due to current and voltage measurements in the absense of an external voltage, or bias. However, this did not impede the device performance, as nominal AlGaAs/GaAs HEMTs operate in a usually "on" state, citing a negative threshold voltage.

It wasn't until the year 1980 when a full-blown HEMT with a control gate was first produced and analyzed by Mimura et. al. [62], and later by Delagebeaudeuf et. al. [63] The importance of the latter is the recognition of the formation of the 2DEG within the supperlattice. Coincidentally, this is also the direct cause of the increased mobility within the intrinsic material.

The next several decades would witness a dramatic increase in the number of scholarly articles citing the HEMT, as well as new material discoveries implementing its design. A good overview of these applications and discoveries are available elsewhere [64, 65]. Today, commercially available HEMTs are used for various high power, high frequency applications. Moreover, new material introductions including the implementation of the nitride compounds have been explored [66]. In particular is the nitride-based HEMTs, often employing combinations of GaN and its ternary combinations. Finally, HEMTs have even begun replacing MESFETs for high frequency device applications [67].

6.3 Physics of HEMTs

The Fermi Level of any semiconductor denotes the probabilistic energy at which exactly half of the electrons have an energy lying above this level, and half below. The expression governing this theoretical energy level can be derived from the net doping of electrons and electron vacancies, or holes, within the material, and is provided elsewhere [54, 68].

The doping of a wide-bandgap material results in the artificial raising of the Fermi level toward the conduction band. Within a HEMT, this is realized as the highly donor-rich wide-bandgap material. The doping of this layer results in a Fermi Level much higher than that of the narrow-bandgap semiconductor. The energy band diagram for a typical HEMT is provided in Figure 6.1. The image on the left shows

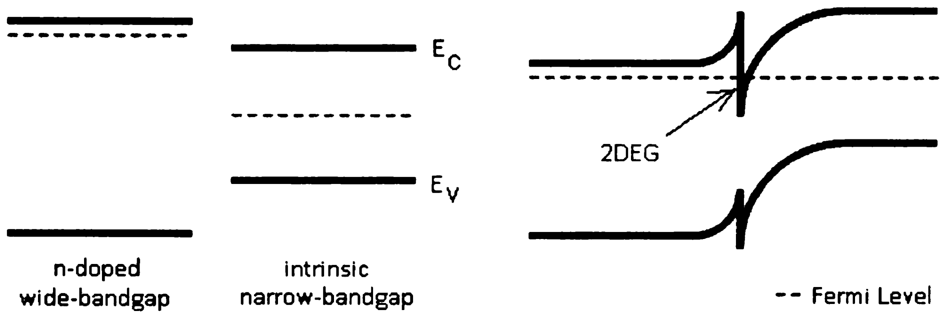

Figure 6.1: A typical energy band diagram for a high electron mobility transistor before (left) and after (right) the formation of the heterojunction.

the energy bands of the two materials when separated. Notice here, that the Fermi levels are distributed at different levels with respect to each other. When sandwiched together, as seen in the image on the right, the two Fermi Levels align, which results in the observed band bending. The narrow region where the Fermi Level lies above the conduction band minimum within the intrinsic semiconductor material results in a region of essentially free electrons, and hence, the formation of the 2DEG.

In any given HEMT, the formation of this free electron layer occurs within the intrinsic narrow-bandgap material. Thus, the production of the 2DEG layer is not addi

simp

2DE

ofte

ext

dis

bili

fiel

nu

ele

to

wh

en

m

re

fi

r

e

t

p

r

a

addressed during the semiconductor growth phase. Rather, this two-dimensional layer simply results via the junction of the two materials.

As was highlighted in the history of such devices, the mobility present within the 2DEG layer is what sets these devices apart from other transistors. This is why it is often utilized within high frequency devices, operating with sufficient output currents extending into the terahertz range [69, 70]. However, unlike the electromagnetic wave discussion presented earlier, there is no simple analytical expression relating the mobility within this region to the physical characteristics of the device nor its external fields. Therefore, an experimental approach citing numerous scholarly articles and numerical simulations, is often explored.

In any case, the mobility within any semiconductor is defined as the ratio of the electric field vector to the drift velocity of an electron. Mathematically, this reduces to

$$\vec{\nu} = \mu \vec{E} \tag{6.1}$$

where μ now defines the mobility within a semiconductor and is completely different from the electromagnetic permeability, μ. However, since the permeability of many real world materials is not of great concern, the Greek letter μ will henceforth represent the mobility.

Thus, the mobility plays a crucial role in the relationship between the electric field applied across any material and the velocity of the electrons within this same region. This also means that, once the mobility is known within a given region experiencing an electric field, the velocity of the electrons is also known. Therefore, the maximum operating frequency of the device can be determined analytically if its physical dimensions are provided.

The *drift velocity* of an electron is simply defined as the average velocity experienced by an electron under the influence of an electric field. Since scattering is always of concern within semiconductors, the average velocity is the best calculation possible for this value. Moreover, it is this drift velocity which is directly related to the operating frequency of the device.

At this point, the total current within the 2DEG region may be derived. Within any semiconductor, the total current is the sum of the electron and hole current densities as

$$\vec{J_n} = q\mu_n n\vec{E} + qD_n \nabla n \tag{6.2}$$

$$\vec{J_p} = q\mu_p p\vec{E} - qD_p \nabla p \tag{6.3}$$

$$\vec{J} = \vec{J_n} + \vec{J_p} \tag{6.4}$$

where J_n and J_p are the electron and hole current densities, respectively, q is the charge associated with a single electron, n and p are the spatial concentrations of electrons and holes, and D is the diffusion constant related with each. Moreover, these current densities presented above represent the impressed current densities, which are directly taken into account in Equation 2.1.

The coefficients in front of the electric field vectors in Equations 6.2 and 6.3 essentially form the total conductivity of that material as

$$\sigma = q(\mu_n n + \mu_p p) \tag{6.5}$$

This, in turn, may be related to the imaginary component of the dielectric constant of a material via

$$\epsilon = \epsilon' - j\frac{\sigma}{\omega} \tag{6.6}$$

This relationship between the solid state device characteristics and its electromagnetic properties will be used later on.

It is clear from Equations 6.2 and 6.3 that mobility is associated with both the electrons and holes. This is the case with most all semiconductor materials, with the electron mobility often greater than that of its hole counterpart.

The diffusion constants can be replaced via the Einstein Relation where [68]

$$\frac{D_{n,p}}{\mu_{n,p}} = \frac{k_B T}{q} \tag{6.7}$$

where k_B is Boltzmann's Constant and T is the material temperature.

The current within any semiconductor for some spatial coordinate has been presented in terms of the electric fields, the mobility, and the charge concentration present at that point. With this brief introduction to semiconductor physics presented, the simulations of such a HEMT device may be discussed.

6.4 HEMT Simulations

Over the past decade, the need to simulate ever-more complex semiconductor transistors and devices has increased dramatically. The original motivation behind numerical simulations was to avoid the costly production of complex systems by simulating the devices first. However, with the advent of modern computers, the numerical simulations of many complex semiconductor devices are now possible for a nominal cost.

Many computer simulation software currently available for device simulation implements a combination of Finite Element Method (FEM) and Monte Carlo simulations to perform many of the required calculations. Moreover, both of these approaches are general enough to allow for the detailed simulation of spatially and electrically complex systems, regardless of the design specifications. One such software package that has grown in popularity over the last several years is the ATLAS Device Simulation Framework from Silvaco [71].

The Silvaco ATLAS software package was used to simulate a $Al_{0.2}Ga_{0.8}N$/GaN modularly-doped HEMT. The physical device schematic with included dimensions, along with its doping levels, are provided in Figures 6.2 and 6.3. The metal contacts are labelled in the structure plot of Figures 6.2 and 6.3 as the three thin regions of

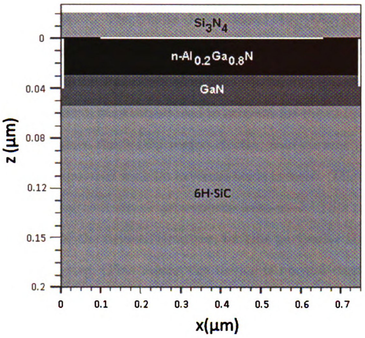

Figure 6.2: Silvaco ATLAS HEMT Simulation - Physical Structure.

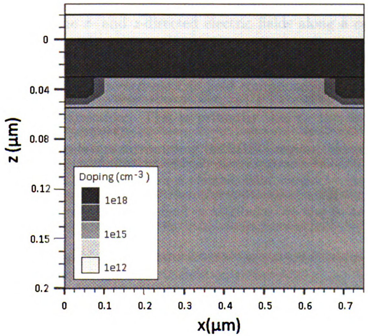

Figure 6.3: Silvaco ATLAS HEMT Simulation - Doping Levels.

white shading. The doping of the structure was taken to be a common doping level of 10^{15} electrons per cm^{-3}. The highly donor-doped $Al_{0.2}Ga_{0.8}N$ concentration was chosen in order to introduce enough carriers into the wide-bandgap layer, but not so many as to introduce erroneous scattering. The doping gradients surrounding the source and drain contacts were introduced to account for the unique structure of these contacts, and to enhance the 2DEG region electric field current during operation.

The device was simulated with no external fields present. That is, no bias voltages between either the source-drain or gate-source were used. This resulted in an "equilibrium" simulation of the device. However, for this particular HEMT, the threshold voltage remains negative. This means the device is normal turned "on" even when no bias voltage is applied. Thus, this application of zero bias voltage will maintain simplicity and also provide a normally-on device.

Since the electric field can only be simulated along a one-dimensional line within this software package, these fields were assumed to be approximately constant throughout the lateral direction of the device. Using this simulation, the finite element mesh simulation of both the x- and z-directed electric fields along a central vertical slice through the device were extrapolated. These are provided in Figures 6.4 and 6.5.

It is clear that the z-directed electric field component is much greater than the component in the x-direction. This is primarily due to the build up of interface charges within the two layers surrounding the 2DEG region. Moreover, in the absence of externally applied fields, no lateral electric field would be expected. Thus, the presence of this x-directed electric field is attributed to the finite element nature of the simulation.

In any semiconductor, the electric field on either side of any boundary composed of a heterojunction must be equal. Therefore, the fact that the z-directed electric field clearly demonstrates a discontinuity between the wide- and narrow-bandgap layers must be inaccurate. However, upon further investigation it is clear that this is

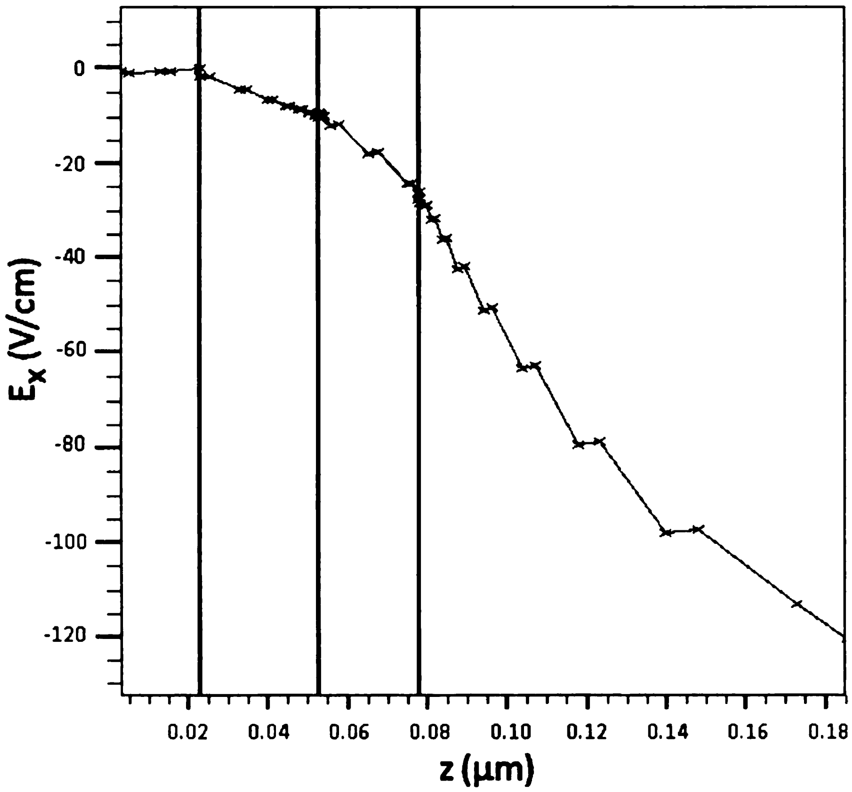

Figure 6.4: Silvaco ATLAS HEMT Simulation - Electric field component in the x-direction.

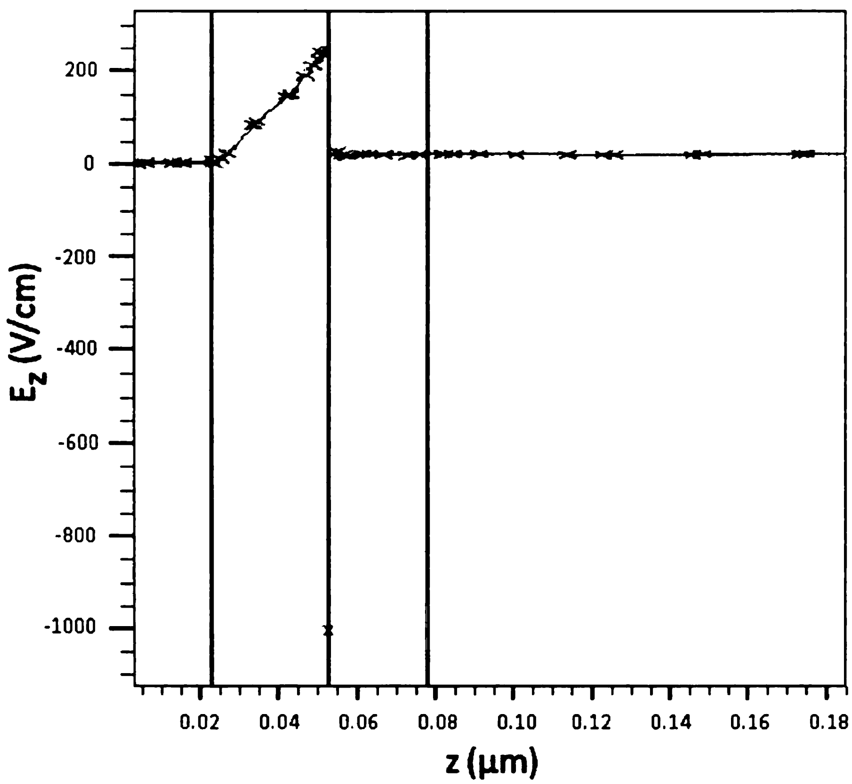

Figure 6.5: Silvaco ATLAS HEMT Simulation - Electric field component in the z-direction.

appearance of a discontinuity of the electric field is misleading. To prove this point, a detailed view of the z-component electric field in the immediate region surrounding the 2DEG layer is provided in Figure 6.6. Thus, it is clear the electric field is in fact continuous. Thus, it is clear from the above figure that even though the HEMT

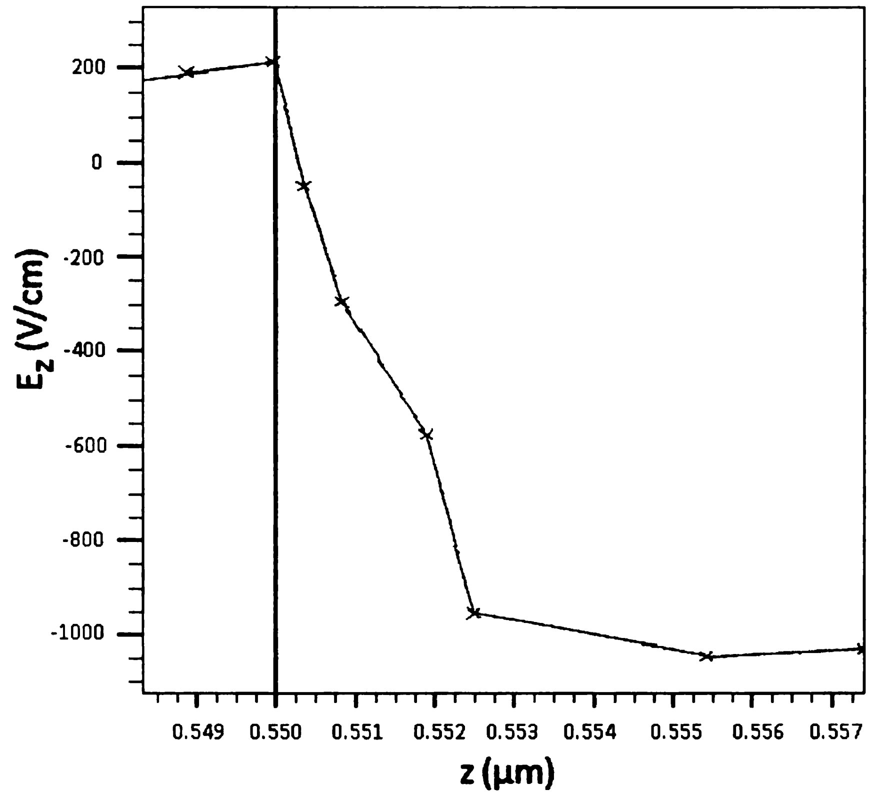

Figure 6.6: Silvaco ATLAS HEMT Simulation - electric field component in the z-direction, detailed view.

device was simulated in the absence of external applied electric fields via a source voltage, the internal electric fields do still maintain noticeable magnitudes within the device region of interest. Moreover, these fields are mainly directed upwards toward the gate as labelled in the above figures. The x-directed electric fields also consist of significant magnitude, yet much smaller than the z-directed.

Since the above device is normally "on" even for zero bias applied voltage, the current density within the device would need to be significant. Therefore, the current density was also simulated and is provided in the Figure 6.7. As expected, the bulk

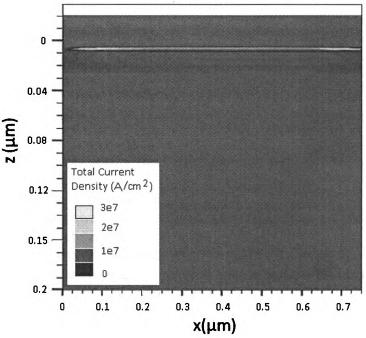

Figure 6.7: Silvaco ATLAS HEMT Simulation - Total current density.

of the total current density resides within the 2DEG region at the interface of the narrow-bandgap region with the wide-bandgap one.

6.5 2DEG Permittivity

Keeping in mind the electromagnetic simulations that are to come, the relative permittivities of most semiconductor materials are known as a function of doping concentrations, temperature, and frequencies. However, the formation of the 2DEG layer within the HEMT device cannot be so easily determined. Thus, another approach must be developed.

Based upon the Phenomenological Fluid Model of the Effective Mass Approximation for a free charge carrier, Hashim et. al. have derived the complex permittivity for a 2DEG layer in terms of the electromagnetic field frequency and tangential wavenumber present within this region [1].

$$\epsilon_{2DEG}(\omega, k) = \epsilon \left[1 - \left(\frac{q^2 n_{so}}{m^* \epsilon_0} \right) \left(\frac{k}{(\omega - k v_d)(\omega - k v_d - j\nu) - (k v_{th})^2} \right) \right] \quad (6.8)$$

where the physical constants cited above are provided in the Table 6.5 below. Thus,

Table 6.1: Physical Constants Used to Calculated Effective Permittivity of 2DEG.

Symbol	Physical Meaning
ϵ	Relative permittivity of narrow-bandgap material
q	Electron charge
n_{so}	Carrier sheet density
m^*	Effective mass of electron within narrow-bandgap material
ϵ_0	Permittivity of free space
k	x-component wavenumber
ω	Frequency of operation
v_d	Electron drift velocity
v_{th}	Electron thermal velocity
nu	Carrier collision frequency

under the operating conditions cited in the simulations above, along with nominal physical constants associated with many 2DEGs [72], the complex permittivity of the 2DEG region within the above AlGaN/GaN HEMT was determined to be $eps_{2DEG} = 5.7 - j0.043$ at an optical wavelength of 632.8nm.

One curious trend observed for this equation is that the inclusion of many nominal electric field strengths, for which the electron drift velocity depends, has little to no effect on this calculated term. Thus, the electric field strength is not required to be known before hand when calculating this value.

Thus, the electric fields and current residing within the HEMT have been numerically simulated using computational software. With this now complete, the interactions between the 2DEG and SPs may now be explored.

Chapter 7

Interactions of SPs and 2DEGs: Optical Frequencies

Now that the electric fields are completely known for both layered planar and periodic grating systems, and within HEMT devices, their interactions are simulated. The vector addition property of electric fields will be heavily used to simulate these interactions. In fact, these so-called interactions really are just the summation of electric fields due to both the internal electric field of the HEMT and the external applied fields due to a source or potential. Moreover, since the electric fields due to the surface waves associated with SPs are simulated to an arbitrary amplitude, their relative magnitude within the HEMT and more importantly, 2DEG layer, will be assumed enough to overcome those already simulated in the preceding chapter. A completely planar HEMT device with one-dimensional variability will be assumed. It will take on the values, both known directly and calculated, for the device.

7.1 Absence of HEMT Gate Contact

To begin, the same structure is assumed as presented in the previous chapter simulations. However, since no bias voltage was used during the solid state simulations, the infinitesimally thin gate region will be temporarily removed. Moreover, the 2DEG region will be considered its own layer with a homogeneous charge distribution, and therefore, complex permittivity. This avoids the need to simulate the layer as a continuously-altering permittivity for the variances in its doping level. This assumption will be carried through the rest of the discussion. Finally, only the electric field amplitudes will be presented, since the magnetic fields add very little usable information to the structure.

In order to launch the SPs, the creation of a surface wave must take place. Thus, the introduction of a prism-coupler above the system will be assumed. This prism coupler consists of a glass layer of semi-infinite thickness backed with a 50nm layer of Gold. The introduction of the Gold layer should result in a dramatic increase in the transmitted field intensity that often corresponds to SPR. An air gap 500nm thick will be assumed initially to reside between the prism coupler and the HEMT device in order to observe the SPs and their evanescent nature.

Thus, a one-dimensional device with physical characteristics presented in Table 7.1 is assumed initially [4, 72–76]. The single angle incidence plane wave magnetic

Table 7.1: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Prism	∞	0	0	0	2.25	0
2	Gold	50	1e23	2e9	4.5e5	-12.6	1.1
3	Air	500	0	0	1	0	
4	Si_3N_4	20	0	0	10^{-12}	7.5	0
5	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.35	0
6	2DEG (GaN)	5	1e15	2000	0.32	5.7	0.043
7	GaN	20	1e15	500	0.08	5.7	0
8	6H-SiC	∞	1e15	350	0.056	6.9	0

field intensity, electric field intensities along both the tangential and longitudinal directions, and the total electric field are observed for the previously labelled system in Figures 7.1 and 7.2. An incident angle of 43.5° was assumed, corresponding to the SPR angle of the prism-coupler system. It is clear from Figures 7.1 and 7.2

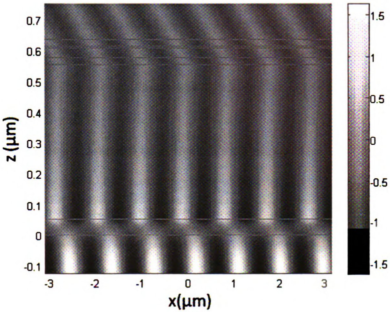

Figure 7.1: Electric field plot along the x-direction for the One-Dimensional System of Prism Coupler suspended 500nm above the HEMT.

that the evanescent nature associated with the SPs is still present. Moreover, their oscillatory nature is also preserved. However, the introduction of finite layers on top of these fields results in magnitudes which behave as sinusoidal functions instead of exponential ones. Thus, the SP intensity actually increases along the z-direction as one approaches the HEMT device.

As noted in Figures 7.1 and 7.2, even after 500nm of attenuation in the air layer, the electric fields associated with the SPs produced still remain of pertinent size within the 2DEG region. However, the z-directed electric field component is observed to reside mainly within the air layer, resulting in a primarily tangential electric field intensity within this 2DEG region. As a figure of merit, the total electric field is observed to fall to roughly ten percent of its original intensity upon incidence.

This air gap thickness was chosen as a rather arbitrary value. In order to further

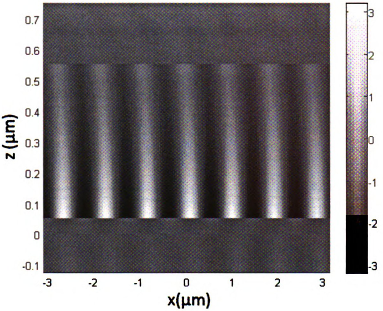

Figure 7.2: Electric field plot along the z-direction for the One-Dimensional System of Prism Coupler suspended 500nm above the HEMT.

explore this variable, this air gap is now allowed to vary between 500nm and 10nm, as its absence will be simulated later on. The electric field intensities along both directions across the prism-coupler, air gap, and HEMT are plotted in Figures 7.3 and 7.4. From these plots, it is clear that an air gap of 400nm corresponds to the greatest magnitude of the electric field within the HEMT along both the tangential and longitudinal directions. However, little variation of the fields with respect to different distances is observed. That is, the electric fields appear to be equal in ratio, with only their magnitude being offset.

Moreover, the x-component electric field is clearly observed to have a much greater intensity, especially within the 2DEG layer, compared to the z-compoennt electric field. This is to be expected when compared to the plots in Figures 7.1 and 7.2.

The air gap thickness has lead to a discovery that a thickness of roughly 400nm corresponds to the greatest electric field magnitude within the 2DEG region. However, this several other critical variables can also be altered. The next most obvious choice

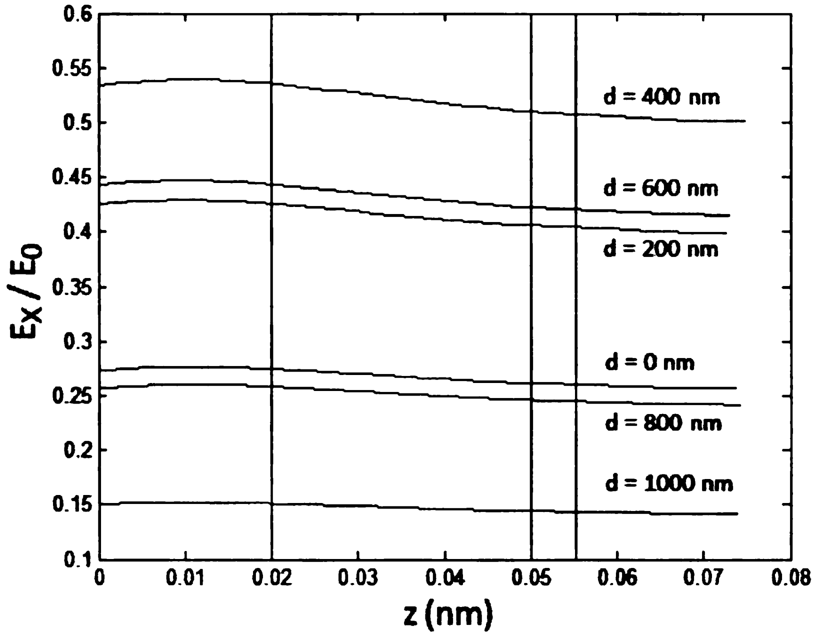

Figure 7.3: Real x-component electric field plots for the One-Dimensional System of Prism Coupler suspended a distance d above the HEMT. Plots are shown across the HEMT device sliced vertically.

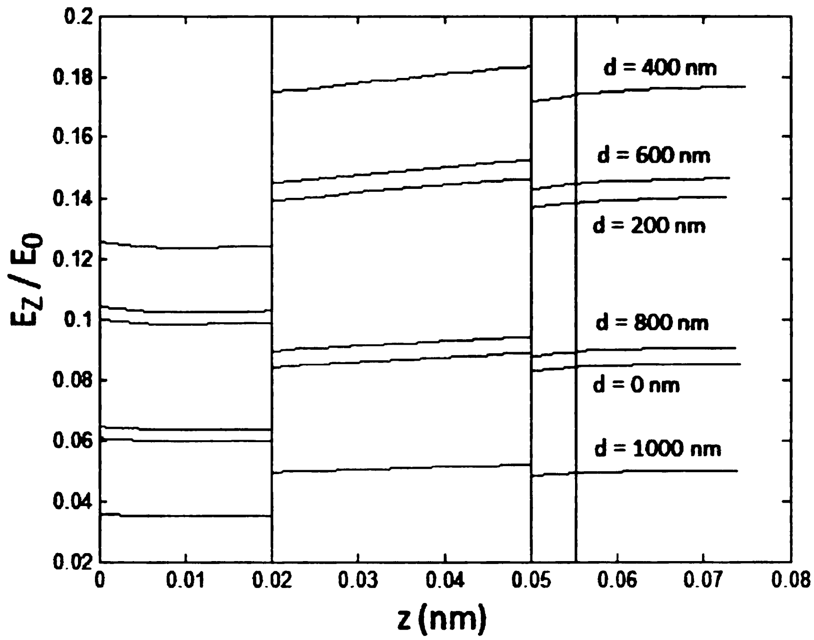

Figure 7.4: Real z-component electric field plots for the One-Dimensional System of Prism Coupler suspended a distance d above the HEMT. Plots are shown across the HEMT device sliced vertically.

is the angle of incidence. The SPR Angle or 43.5° was chosen previously since it corresponded to the maximum field intensity, or SPR, of the 3-layer prism coupler. However, this angle may not correspond to the greatest field magnitude for the present system, taking into account the several new layers. Moreover, the maximum field intensity is sought within the 2DEG region, not the air gap. Thus, the 400nm air gap thickness will be maintained, while the angle of incidence is varied over all of the possible incident angles. The resultant electric field intensities within each layer are presented in Figures 7.5 and 7.6. Although it appears that the resonance angle

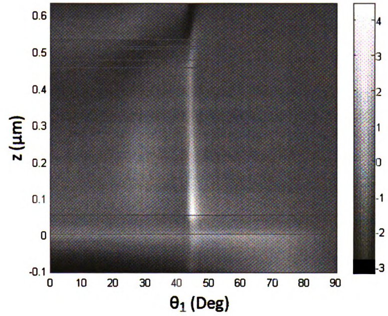

Figure 7.5: Real x-component electric field plots for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT for various angles of incidence. The plot is shown across the HEMT device sliced vertically.

presented in Figures 7.5 and 7.6 shows an SPR angle of 43.5°, the angle corresponding to maximum field amplitude within the 2DEG layer is 44.5°. Therefore, the maximum electric fields in both the x- and z-direction occur for a 400nm air gap system and an angle of incidence of 44.5°.

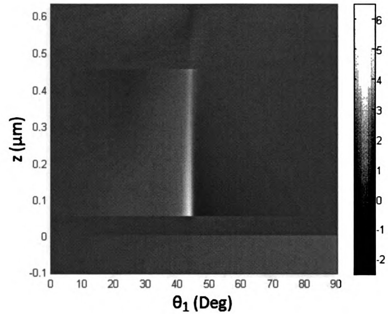

Figure 7.6: Real z-component electric field plots for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT for various angles of incidence. The plot is shown across the HEMT device sliced vertically.

7.2 Absence of Gate Contact and Insulator

The next system that can be simulated is that of the HEMT in the absense of an insulator *and* a gate contact. Thus, this new system is identical to that quoted above, sans the Silicon Nitride insulator layer. The field intensities within the system are presented in Figures 7.7 and 7.8 for all possible incident angles. It is still clear that the majority of the electromagnetic energy is contained within the air gap region. A clear resonance is still observed, this time occurring at an incident angle of 41.7°. The components of the fields within the HEMT device at this angle are provided in Figures 7.9 and 7.10. Notice in both figures that the surface wave on the back of the gold layer exhibits the evanescent nature and enhanced amplitude associated the the SPs present in the gold. Moreover, the fields within the 2DEG region are greatly suppressed compared with those presented earlier for the insulator present. This can be attributed to the fact that the insertion of this Silicon Nitride layer results

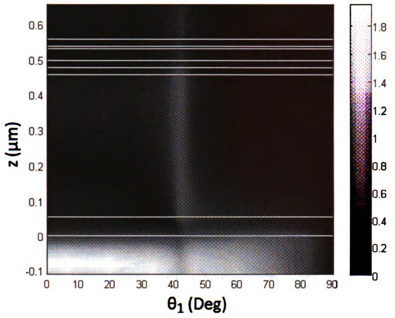

Figure 7.7: Real x-component electric field plot for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT without gate contact or insulation layer.

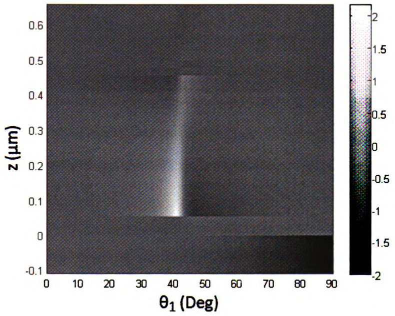

Figure 7.8: Real z-component electric field plot for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT without gate contact or insulation layer.

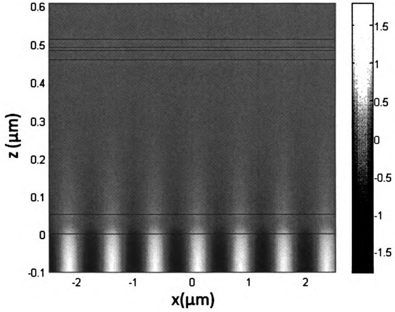

Figure 7.9: Real x-component electric field plot for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT without gate contact or insulation layer.

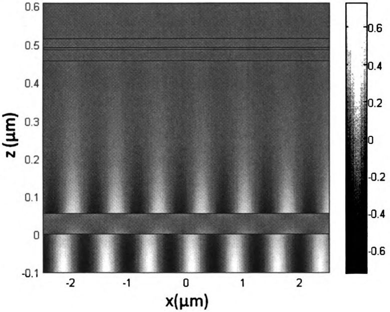

Figure 7.10: Real z-component electric field plot for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT without gate contact or insulation layer.

in a formation of surface waves within the HEMT as well, since the the relative permittivity of this insulator is greater than those of the HEMT layers. Thus, in its absence, the creation of surface waves cannot exist within the HEMT.

7.3 In the Presence of a Gate Contact

In the preceding sections, the gate contact present within most transistor devices, and therefore HEMTs, was not taken into account. The argument that the device, usually in the "on" position even under zero bias voltage was the basis for its removal. However, in order to complete the real world device situations, this gate will now be included within the discussion.

The size and make-up of the gate plays a role of considerable weight. Since this gate will essentially screen the incoming electromagnetic waves from the 2DEG region, its thickness cannot be so great that it absorbs the bulk of the incident energy. Moreover, it must be thick enough to support semi-uniform fields for a given applied voltage, thus resulting in one-dimensional constraint of the 2DEG region.

In order to satisfy both of these criteria, the gate contact will be assumed as a gold layer of 20nm thick. This should allow for the passing of enough electromagnetic energy so as to not interfere with the 2DEG, but also thick enough to support the consistent field requirement. Since the complex permittivity of the gold region is already known, the system provided in Table 7.3 will be simulated, as similar to the above section. Taking into consideration what was previously determined to yield the maximum electric field amplitude within the 2DEG region, the system is simulated for an incident angle of 44.5° and a free space wavelength of 632.8nm, corresponding to a standard Helium-Neon Laser. An air gap is still present, with a thickness of 400nm. The resulting relative electric field intensities are presented in Figures 7.11 and 7.12.

Thus, as was seen previously, the bulk of the electromagnetic energy remains within

Table 7.2: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions with Gold Gate Contact

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Prism	∞	0	0	0	2.25	0
2	Gold	50	1e23	2e9	4.5e5	-12.6	1.1
3	Air	500	0	0	1	0	
4	Si_3N_4	20	0	0	10^{-12}	7.5	0
5	Gold	20	1e23	2e9	4.5e5	-12.6	1.1
6	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.35	0
7	2DEG (GaN)	5	1e15	2000	0.32	5.7	0.043
8	GaN	20	1e15	500	0.08	5.7	0
9	6H-SiC	∞	1e15	350	0.056	6.9	0

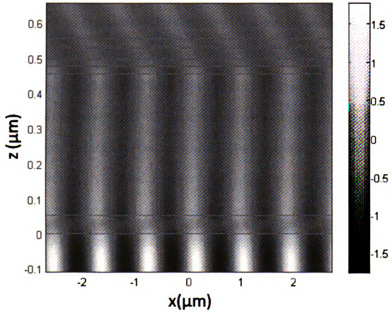

Figure 7.11: Real x-component electric field plot for the One-Dimensional System of Prism Coupler suspended a distance 400 nm above the HEMT with a Gold Gate contact for $\theta_1 = 44.5^{\circ}$. The plot is shown across the HEMT device sliced vertically.

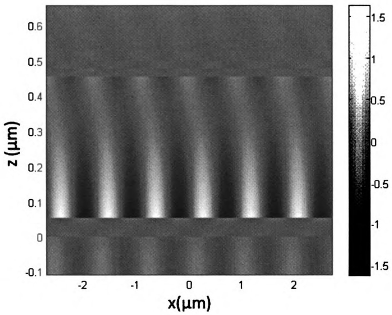

Figure 7.12: Real z-component electric field plots for the One-Dimensional System of Prism Coupler suspended a distance 600 nm above the HEMT with a Gold Gate contact for $\theta_1 = 44.5^{\circ}$. The plot is shown across the HEMT device sliced vertically.

the air gap region. The gold contact layer clearly absorbs more of the electromagnetic energy than in the previous section, without this contact region. Therefore, the electric fields for both tangential and longitudinal directions are vigorously damped compared to the no-gate system. This is more clearly observed via Figures 7.13 and 7.14. It is obvious from Figures 7.13 and 7.14 that the electric fields are damped much more so within the 2DEG region by the gold layer than with no such layer present. With the gold contact layer present, both the x- and z-directed electric fields observe almost a 40-percent decrease in the electric field magnitude compared to the same system sans the gold contact in the previous section. Therefore, although this gold contact is a necessity in order to apply external bias voltage, it does in fact lead to a decrease in field presence within the 2DEG region.

The reason for this is the lossy nature of the metal, whereby a larger proportion of the energy is reflected back from this layer into the air gap. This is easily observed

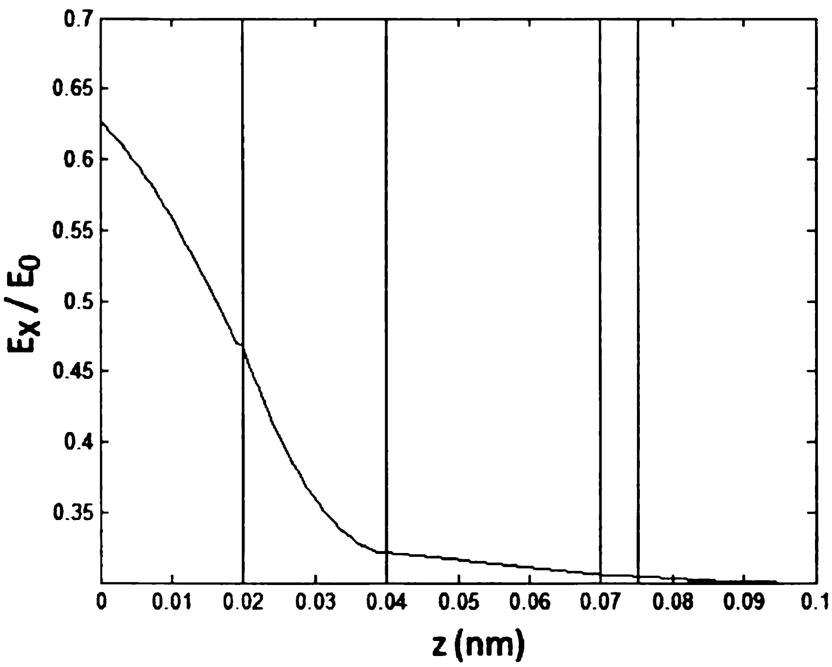

Figure 7.13: Real x-component electric field across the HEMT, sliced vertically. Gold gate contact of thickness 20nm is assumed, along with $\theta_1 = 44.5^{\circ}$.

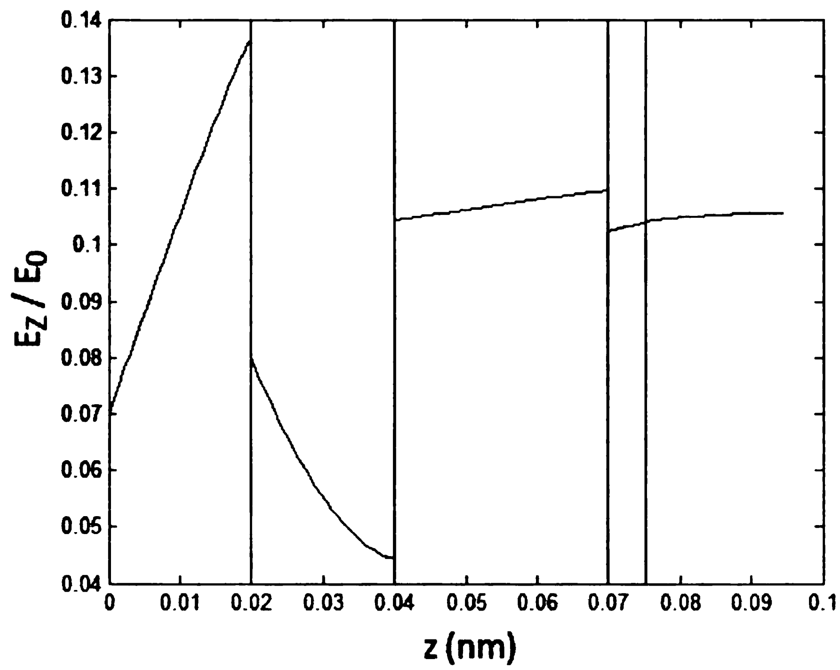

Figure 7.14: Real z-component electric field across the HEMT, sliced vertically. Gold gate contact of thickness 20nm is assumed, along with $\theta_1 = 44.5^{\circ}$.

by the sharp increase in the field intensities within the insulator region, composed of Silicon Nitride. With the gold layer present, the x-directed electric field sees a roughly 25-percent increase in field intensity, while the z-directed component of the electric field is increased by almost a factor of two.

7.4 Sapphire-Backed HEMT

The preceding sections all dealt with a nominal HEMT device assuming a GaN thickness of 20nm. However, this can sometimes result in breakdown of the channel, since the carriers can be depleted rather quickly. Thus, the introduction of a much thicker GaN region would avoid this case. Thus, a one micron GaN thickness will now be assumed. Furthermore, the system will be assumed backed by Sapphire, or Al_2O_3, which is also a common substrate material for the matching of III-nitride semiconductors. The new physical system constants are presented in Table 7.4. The air gap

Table 7.3: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Prism	∞	0	0	0	2.25	0
2	Gold	50	1e23	2e9	4.5e5	-12.6	1.1
3	Air	400	0	0	1	0	
6	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.35	0
7	2DEG (GaN)	5	1e15	2000	0.32	5.7	0.043
8	GaN	1000	1e15	500	0.08	5.7	0
9	Sapphire	∞	0	0	0	3.1	0

will remain at 400nm, since this provides the maximum field enhancement, which is solely due to the prism coupler system. The field magnitudes within each of the layers is presented in Figures 7.15 and 7.16 for all possible incident angles. A clear resonance angle is still observed, this time at the θ_{SPR} angle of 43.9°. Simulating the electric field intensities within the system at this incidence angle results in Figures

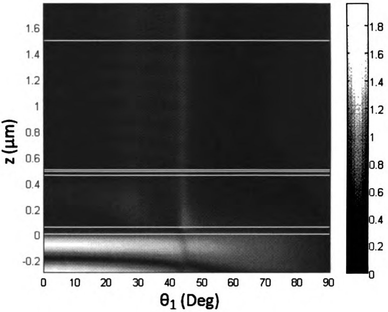

Figure 7.15: x-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate.

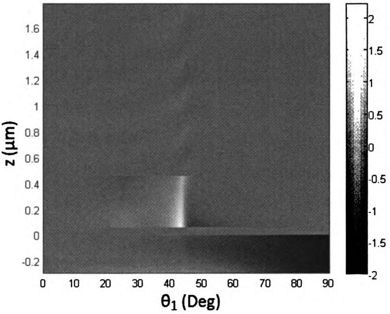

Figure 7.16: Real z-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate.

7.17 and 7.18. Once again, the creation of surface waves within the air gap results

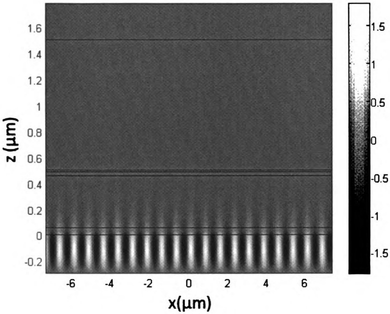

Figure 7.17: x-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate at an incident angle of 43.9°.

in evanescent behavior. However, as with the previous section, the field intensities within the 2DEG region are greatly suppressed, due to the absence of total internal reflection between HEMT layers.

The system presented in Figures 7.15-7.18 may once again be simulated, but with the thin gold contact layer replaced by insulating Silicon Nitride. This might lead to a modestly-improved field intensity within the 2DEG region, due to the absence of a second lossy layer. The field plots over all possible angles of incidence are presented in Figure 7.19 and 7.20. Thus, from Figures 7.15,7.16, 7.19 and 7.20 it is clear that with the inclusion of a large intrinsic Gallium Nitride layer backing the 2DEG region, the fields within this layer are almost completely identical. Therefore, the conclusion that the presence of a very thick layer of bulk intrinsic semiconductor backing the 2DEG region has little effect on the field intensities within the 2DEG is made.

The primary reason for this congruent field magnitude within the 2DEG layer

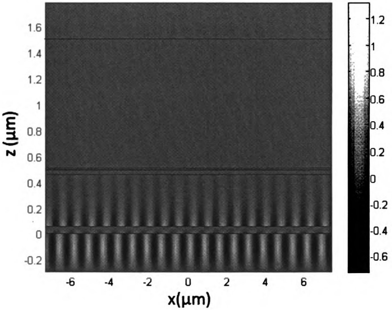

Figure 7.18: Real z-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate at an incident angle of 43.9°.

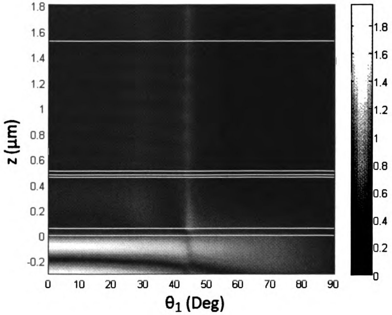

Figure 7.19: x-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate. Gold contact layer is now replaced with Silicon Nitride.

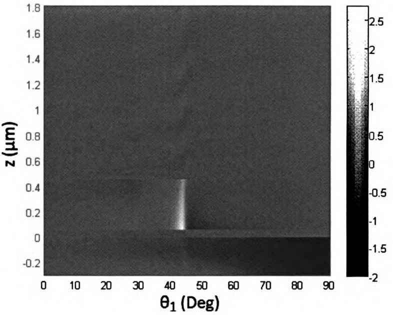

Figure 7.20: x-component electric field magnitude across the Prism-Coupler-HEMT system with a 1μm GaN layer on top of a Sapphire substrate. Gold contact layer is now replaced with Silicon Nitride.

is due to the extremely small thickness of the gold. Undoubtedly, if this layer was increased in thickness, a greater discrepancy between the electric field magnitudes within the 2DEG region would be observed.

7.5 Simple HEMT Device

The previous two sections both dealt with the introduction of an air gap separating a prism-coupler and the HEMT device. The first of these systems allowed for the introduction of an enhanced-field surface wave, while the second introduced the 2DEG region. It was clearly observed that, although this air gap introduced a free space region where the surface wave intensity dropped off evanescently, it did result in a relatively high electric field intensity within this 2DEG layer. This was primarily due to the enhancement from the prism-coupler system.

A much simpler system would involve simply a HEMT in the absense of the prism-coupler system. In order to produce the surface waves within the HEMT itself, it is necessary to create total internal reflection. Thus, since the Silicon Nitride layer has inherently greater relative permittivity than the $Al_{0.2}Ga_{0.8}N$ layer, then total internal reflection may occur for a specific set of angles.

As described earlier, the one-dimensional system under consideration is provided below. The choice of an insulator with a high enough relative permittivity in order to observe total internal reflection was chosen so that this critical angle would not approach grazing conditions. The material and physical characteristics of the new, much simpler HEMT system, are provided in Table 7.5. The system above implies

Table 7.4: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions, simple HEMT device

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Si_3N_4	∞	0	0	10^{-12}	7.5	0
2	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.35	0
3	2DEG (GaN)	5	1e15	2000	0.32	5.7	0.043
4	GaN	20	1e15	500	0.08	5.7	0
5	6H-SiC	∞	1e15	350	0.056	6.9	0

a critical angle between the first and third layers of 57.6°. Thus, surface waves, and thus, SPs, may only exist above this angle of incidence. In order to determine the angle of incidence corresponding to the greatest surface wave intensity, plots of the x- and z-directed electric fields across this device for the various angles of incidence are presented in Figures 7.21 and 7.22. Notice in Figure 7.22 the z-component electric field magnitude within the 2DEG region reaches its maximum intensity at an angle of incidence of roughly 64°. This is in sharp contrast to the x-directed electric field, which has its maximum magnitude at normal incidence, with a much less pronounced local maximum at roughly 75°. Therefore, it would appear the surface waves which result from this system do not yield such high electric field magnitudes as in the previous systems. However, this is not the case, in that both the x- and z-directed

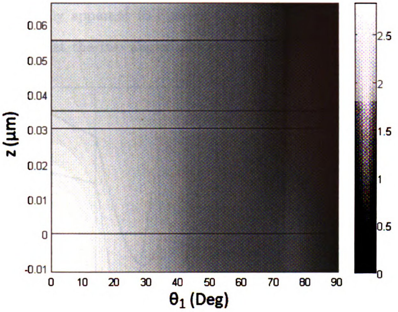

Figure 7.21: x-component electric field Across the simple HEMT, sliced vertically for various angles of incidence.

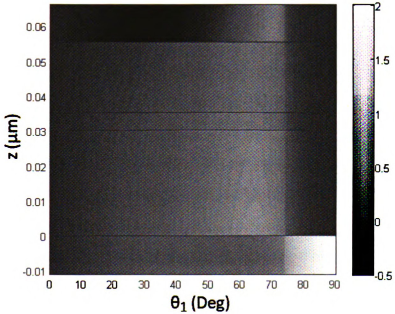

Figure 7.22: Real z-component electric field Across the simple HEMT, sliced vertically for various angles of incidence.

fields do reach as high magnitudes as in the previous systems at both of these angles. This is more readily apparent in Figures 7.23 and 7.24, showing the electric fields across the system at the two angles previously mentioned. Thus, it is clear that

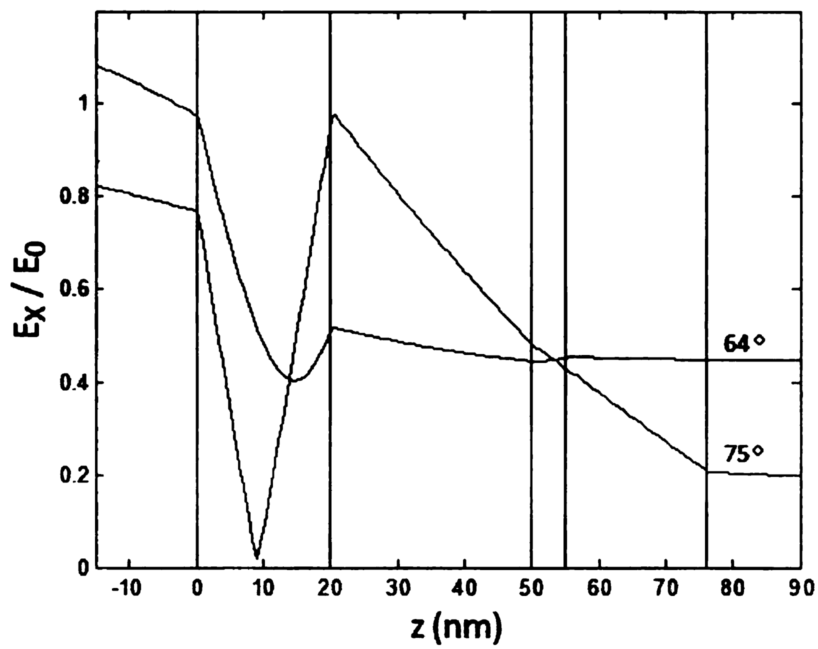

Figure 7.23: x-component electric field Across the simple HEMT, sliced vertically for angles of incidence of 64° and 75°.

an incident angle of 75° produces the greatest z-directed electric field component. This could in turn be used as a control mechanism to act as an external bias field controlling the "on/off" switch of the HEMT device. This would result in a much simpler HEMT device, without the need for a gate contact controlling field. Moreover, the x-directed field is observed to have much the same magnitude as with the previous simulated systems.

Finally, this gold gate contact has once again been avoided in the previous section, since it would its only duty pertaining to the operation of the HEMT is to control this "on/off" device performance. For completeness purposes, this gold gate contact is now replaced and the fields are once again simulated within this simple HEMT system for all possible angles of incidence. Thus, it appears from Figures 7.25 and

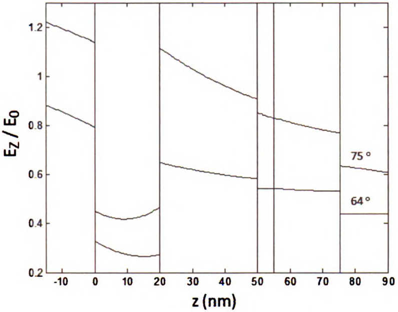

Figure 7.24: Real z-component electric field Across the simple HEMT, sliced vertically for angles of incidence of 64° and 75°.

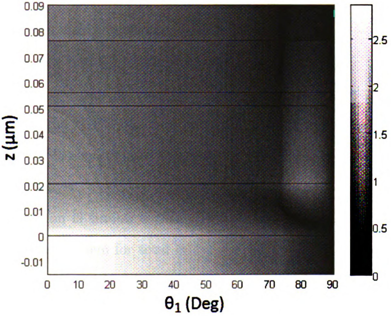

Figure 7.25: x-component electric field Across the simple HEMT, sliced vertically for all angles of incidence. Gold gate contact layer of 20nm is present.

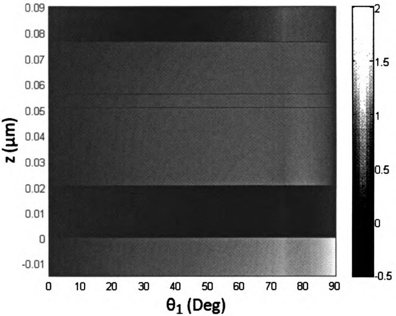

Figure 7.26: Real z-component electric field Across the simple HEMT, sliced vertically for all angles of incidence. Gold gate contact layer of 20nm is present.

7.26 that both the x- and z-directed electric fields have maxima occurring at the same incident angle of exactly 80°. This is more clearly observed in Figures 7.27 and 7.28 showing the electric field magnitude in both directions for this angle of incidence. The previous section highlighted the increased absorption of electromagnetic energy upon the introduction of a gold contact layer, resulting in a decrease electric field magnitude within the 2DEG region. However, as is clear from the above figures, the introduction of a gold contact layer in this case results in an *enhanced* field magnitude within the 2DEG layer.

This phenomenon is similar to the introduction of the gold layer onto the back of the prism-coupler system for total internal reflection. In that case, the maximum electric field magnitude within the transmission region was a factor of two, due to the total internal reflection of the fields and the necessary equality of the fields across the boundary. However, once the metal was introduced, a much greater field intensity resulted.

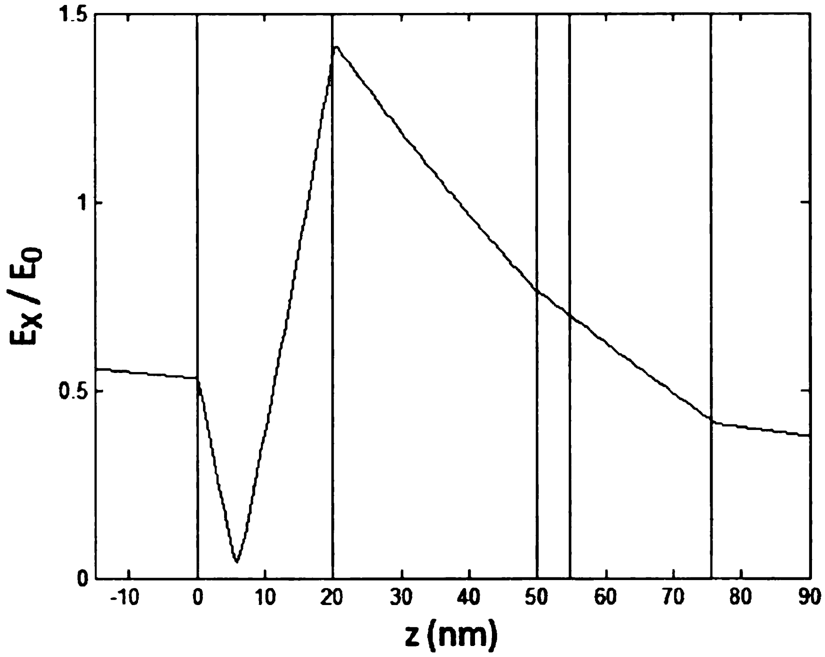

Figure 7.27: x-component electric field Across the simple HEMT, sliced vertically for an angle of incidence of 80°. Gold gate contact layer of 20nm is present.

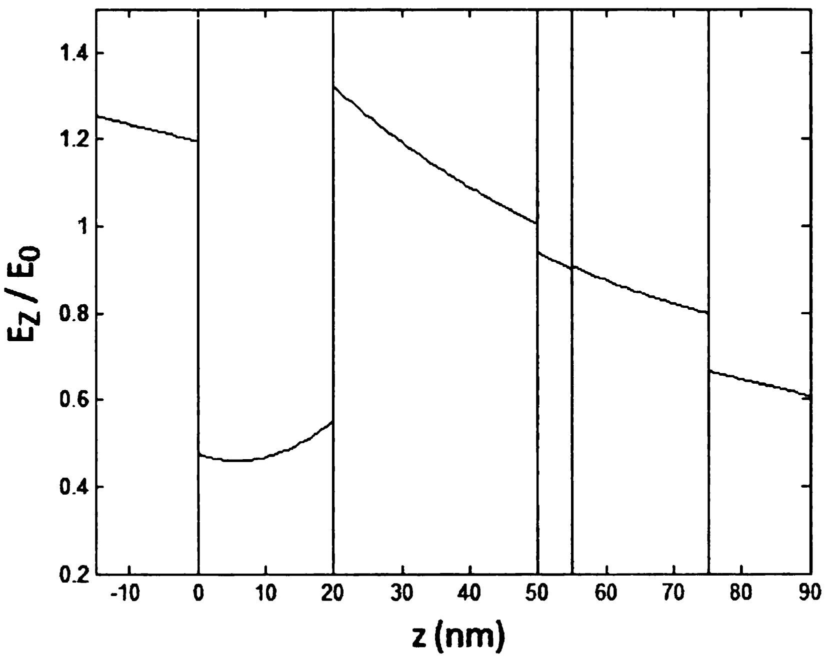

Figure 7.28: Real z-component electric field Across the simple HEMT, sliced vertically for an angle of incidence of 80°. Gold gate contact layer of 20nm is present.

The same phenomenon is occurring here in the HEMT system. The introduction of a metallic layer results in a much greater coupling of the SPs to the back side of this region and, hence, a much greater electric field intensity. Quantitatively, the introduction of this gold contact layer results in an enhanced x-component electric field of approximately 50-percent, while the z-directed electric field magnitude observes an more reasonable increase of almost 10-percent. Thus, the introduction of a gold contact layer greatly enhances the tangential electric field vector, yet has little effect on the z-directed counterpart.

Thus, the argument for the avoidance of the gate contact layer appears valid. Since the only role of the gate contact layer is to provide external biased fields in order to mitigate between the "on" and "off" switching of the device, and that the z-directed electric field can be used to perform the same task to an arbitrary field intensity, it may be removed altogether. Moreover, this z-directed electric field may be used as a *switch* of sorts for the device performance handling.

However, one must remember that every electromagnetic wave carries with it both spatial and time-dependent oscillations. Therefore, to say that the switching of the device may be mediated by the incident electromagnetic wave in the same sense as a contact gate voltage is incorrect. This becomes even more true when the wavelength of the incident wave reaches the optical spectrum. This has to do with the frequency of the radiation.

As was highlighted in the previous chapter, the maximum operating frequency of most HEMTs lies somewhere in the Gigahertz to single Terahertz regime. Thus, the optical switching of such a device is, at this point in history, not reasonable since the frequency of optical electromagnetic radiation lies in the Petahertz range, or more than one thousand Terahertz. Moreover, the switching of such devices is often mediated by a DC signal. Thus, the switching of such a device is assigned to absurdity. However, the interactions of a HEMT device with these high frequencies must also be addressed. The following chapter will deal with the same SP-2DEG interactions, but this time within the Gigahertz-to-Terahertz range.

Chapter 8

Interactions of SPs and 2DEGS: Far Infrared Frequencies

Electromagnetic waves with frequencies in the single Terahertz regime, also referred to as far-infrared radiation, have seen a dramatic increase in their use in medical imaging and academics over the past several decades. Moreover, their increasing use in the digital communications industry had led to much interest in their implementations. The following chapter will again probe the interactions of SPs with 2DEGs within one-dimensional HEMT devices.

In order to study the interactions between SPs and 2DEGs once more, but this time using wavelengths on the order of a millimeter, the physical properties of the layers must be well-known. Thus, the permittivities of each layer must be first developed.

8.1 Permittivity of Gold and 2DEG

The permittivity of various metals has been calculated by Ordal et. al. using the standard Drude model. This results in a complex permittivity of gold of $(-0.98 + j2.60) \times 10^5$ for a wavelength of 118 μm [77].

The choice of wavelength coincides with the radiation condition proposed by Hashim et. al. for a 2.5 THz electromagnetic wave [32].

Right away, the extremely high relative permittivities for both the real and imaginary components signals dramatic signal attenuation. Therefore, one would expect the fields to vanish almost completely through a gold layer only a few nanometers thick. Thus, the prism-coupler system no longer serves its primary purpose of producing enhanced amplitude surface waves, and only the HEMT will be discussed in this chapter.

The relative permittivity of the 2DEG region must also be calculated at this wavelength. Thus, performing similar calculations as previously, the complex permittivity within the 2DEG layer becomes $\epsilon = 4.84 - j0.04$.

To begin, the system under consideration is the same as described by Hashim et. al. consisting of a one-dimensional HEMT structure, but this time composed of AlGaN/GaN instead of the arsenic-based HEMT cited. Although this is not necessarily the same components as in the cited work, the basic device is similar and should pose similar results. The introduction of a space charge layer due to polarization charges lying in the interface between the nitrides will be discussed later on.

8.2 Interactions at the Plasma Frequency

The system now under simulation is provided in the table below. Physical constants and known values are provided where available. Since no physical characteristics other than the total dimensions of the metal contact regions were provided by Hashim, the physical constants assumed from before will be continued here. No insulating layer above the gate region is assumed, in order to stay consistent with the referenced work. Without the presense of the insulator, the total internal reflection case is not necessarily met. However, since the gold gate layer poses such a large relative

Table 8.1: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions, as proposed by Hashim et. al.

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Air	∞	0	0	0	1	0
2	Gold	20	1e23	0	4.7e6	-9.8e4	2.6e5
3	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.06	0
4	2DEG (GaN)	5	1e15	2000	0.32	4.84	-0.04
5	GaN	20	1e15	500	0.08	4.84	0
6	6H-SiC	∞	1e15	350	0.056	6.5	0

permittivity, with both the real and imaginary components being on the order of ten-thousand, much reflection does still occur. Since the angle of incidence appears to be arbitrary at this point, the electric field magnitudes will be plotted for all possible angles of incidence. These appear in Figures 8.1 and 8.2. Thus, it appears that

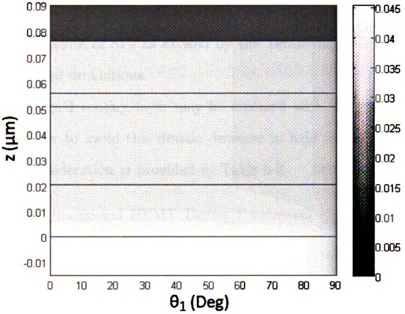

Figure 8.1: x-component electric field Across the Hashim Nitride-Based HEMT, sliced vertically for all angles of incidence. Gold gate contact layer of 20nm is present.

the fields within the 2DEG layer, regardless of angle of incidence will be much less

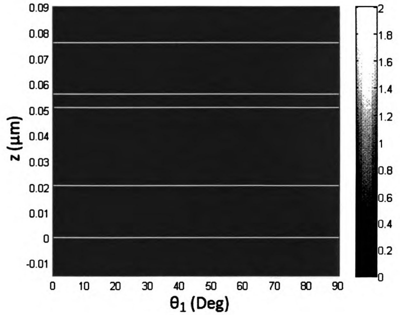

Figure 8.2: Real z-component electricfield Across the Hashim Nitride-Based HEMT, sliced vertically for all angles of incidence. Gold gate contact layer of 20nm is present.

than that compared to the optical excitation radiation from the previous section. Therefore, the presense of SPs as excited by the Terahertz radiation is not readily apparent from these simulations.

However, the gold contact layer may be replaced with the Silicon Nitride insulator layer in order to avoid this drastic decrease in field intensity. Thus, the new system under consideration is provided in Table 8.2. As with most of the previ-

Table 8.2: One-Dimensional HEMT Device Parameters Used to Study SP-2DEG Interactions, as proposed by Hashim et. al.

Layer	Material	d (nm)	$n(cm^{-3})$	$\mu_n(\frac{cm^2}{V-s})$	$\sigma(S/cm)$	ϵ'	ϵ''
1	Air	∞	0	0	0	1	0
2	Si_3N_4	20	0	0	0	4.0	0
3	$Al_{0.2}Ga_{0.8}N$	30	1e18	1000	160	5.06	0
4	2DEG (GaN)	5	1e15	2000	0.32	4.84	-0.04
5	GaN	20	1e15	500	0.08	4.84	0
6	6H-SiC	∞	1e15	350	0.056	6.5	0

Figu
vert
Nitr

F
sl
Si

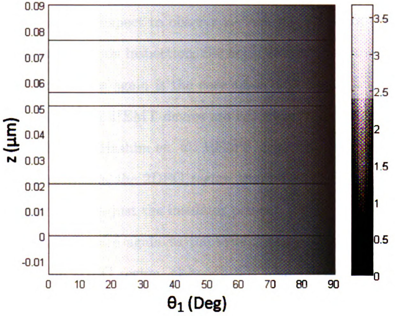

Figure 8.3: x-component electric field Across the Hashim Nitride-Based HEMT, sliced vertically for all angles of incidence. Gold gate contact layer is absent, but a Silicon Nitride layer of 20nm thickness is present.

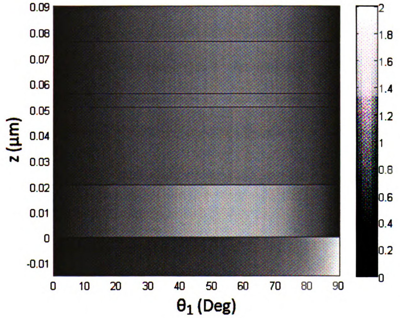

Figure 8.4: Real z-component electric field Across the Hashim Nitride-Based HEMT, sliced vertically for all angles of incidence. Gold gate contact layer is absent, but a Silicon Nitride layer of 20nm thickness is present.

ous instances, the resonance of the z-component electric field enhancement occurs at roughly 58°. One would expect to observe surface waves within the HEMT system at this angle. However, upon inspection, the fields are observed to demonstrate plane wave characteristics. This again is the case since both the x- and z-components of the wavevector within the HEMT device are completely real. Thus, the creation of surface waves within the Hashim et. al. HEMT structure is either not realizable for realistic amplitudes within the 2DEG region assuming a gold contact gate, or they are not created at all with just the insulator present.

However, returning once again to the idea of the air gap region, surface waves can exist within the 2DEG region, although not with the same enhancement factor as previously quoted. Nevertheless, the introduction of this air region does result in surface waves whose amplitudes extend into the HEMT. This is presented in Figures 8.5 and 8.6 for all angles of incidence. Thus, the creation of surface waves within

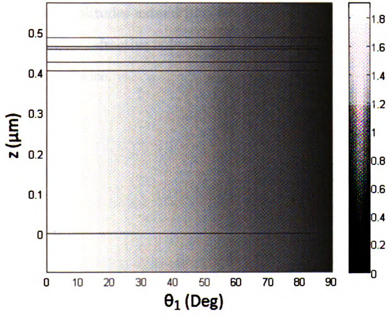

Figure 8.5: x-component electric field magnitude across the Hashim Nitride-Based HEMT with a prism suspended 400nm above , sliced vertically for all angles of incidence. Silicon Nitride layer of 20nm thickness is present.

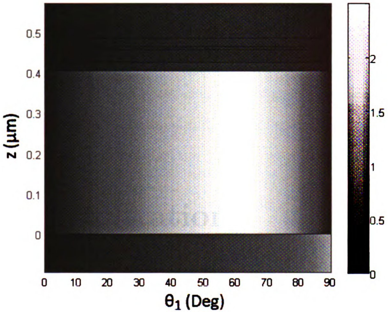

Figure 8.6: Real z-component electric field magnitude across the Hashim Nitride-Based HEMT with a prism suspended 400nm above , sliced vertically for all angles of incidence. Silicon Nitride layer of 20nm thickness is present.

the air gap, whose amplitudes extend into the HEMT device, are possible upon the introduction of a higher refractive index semi-infinite layer suspended a small distance above the transistor device.

Chapter 9

Device Applications

The systems simulated in the previous chapters all dealt with the theoretical interactions between the electric fields produced by SPs and those inherent to 2DEG regions within one-dimensional devices. However, many of these were simulated in the optical regime, most often where SPs are observed, yet the normal operation of any HEMT has yet to climb past the 10 THz cutoff. Thus, the direct incorporation of these fields due to SPs within the 2DEG region has little promise for optical frequencies. This does not exclude, however, the resonant coupling between the optical excitation of SPs or plasmons within the 2DEG region at Terahertz frequencies. Moreover, these interactions have been proven to exist just outside the optical wavelengths citing Meziani and Shaner [2, 3, 78].

The phase velocity associated with electromagnetic fields and surface waves is the primary physical constraint used in many of the following devices. The phase velocity of any electromagnetic wave is given by the ratio of the frequency of the wave and its tangential wavevector, $v_p = \omega/k_x$. Thus, if the channel length and frequency of the incident wave are known, the maximum operational frequency of the device can be calculated. Moreover, if the plasma oscillation frequency is known, this could be used as a coupling mechanism to extend the working frequency of the solid state device above the standard frequency cutoff.

The plasma oscillation frequency is given in terms of the harmonics within the 2DEG channel and can be calculated if the gate length (L_G) and plasma velocity (v_p) are known using [33]

$$f_r = \frac{(2n-1)}{L_G} v_p \tag{9.1}$$

As stated in the previous chapter, this has already been proven by Hashim et. al. and Lu et. al. for a standard GaAs-based HEMTs [32, 33]. However, Hashim cited a noticeable improvement of only several milliamps when the device is excited with THz radiation above the cutoff frequency of the device. Thus, for high power applications, this does not present a clear improvement on the device performance. However, the introduction of the radiation from the back side of the device, where the lossy nature of the gate contact does not impede the surface wave amplitude, was quoted by Lu et. al. Moreover, the removal of this gate contact due to its extremely lossy nature would also avoid this tremendous decrease in field strength. Although, this would also inhibit the internal control of the device performance via an external bias.

9.1 Surface Charge Detection

The fundamental operational mechanism inherent to all SPR-based biosensors is their sensing of a change in refractive index of the system. As proteins bond with one another, a slight change in the SPR angle, due to a change in relative permittivity, is recorded. This signal is then processed along with thousands of others to record a the binding rates, affinities, and characteristics of these proteins. However, as was stated earlier, the overall throughput of these systems is low, due to the normalization of thousands of binding sites. Returning to the premise of this work, the idea of placing a transistor behind each binding site and recording the interactions for every site would dramatically improve the throughput of such a sensor.

A novel HEMT device employing SP-sensing characteristics might follow a similar schematic presented below. An array of HEMTs sans gate contacts would have incident upon them an electromagnetic wave resulting in an enhanced electric field strength within the 2DEG region. An air gap would be present separating the HEMT from a higher refractive index packaging material. Through this air gap would flow some gas or liquid that is meant to interact with the protein present at each site. A schematic of this device is presented in Figure 9.1.

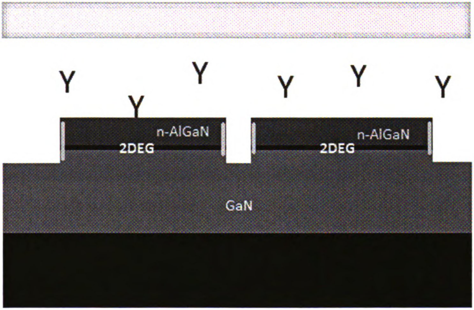

Figure 9.1: Proposed Surface Charge Detection Device

During binding between the molecules and the AlGaN, surface charges would be exchanged. These surface charges would create electron-hole pairs. With an applied voltage between drain and source contacts, these charges would in turn be registered by the each HEMT as an electrical signal. Thus, this device would act as a surface charge detection transistor.

Another related device that utilizes the external electromagnetic radiation could be used to sense the interactions between protein pairs. This device is provided in Figure 9.2. An electromagnetic wave with optical range wavelength is incident

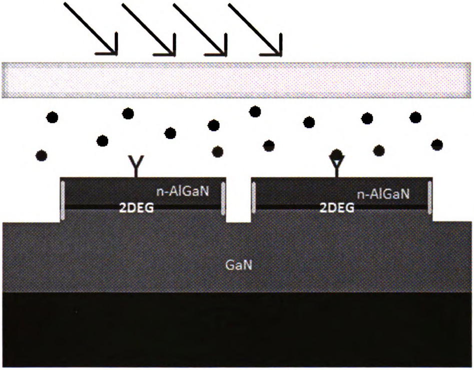

Figure 9.2: Proposed Binding Kinematics Detection Device

from outside of the packaging layer. A thin metal layer is also present just inside the higher refractive index layer essentially forming a Kretschmann-Raether configuration backed by a HEMT. Once again, gas or liquid of some sort is allowed to pass through the air gap region. Instead of registering the optical signal via reflectance of the electromagnetic wave, the enhanced electric fields within the HEMT are registered between the source and drain contacts. A change in field strength between contacts will be registered for varying permittivities present within the air gap layer.

Only within the past decade has the design and production of Terahertz-based

HEMTs been achieved. Moreover, the research tied to the area of SPs has seen continued growth over the past half century. However, it has only recently that the interactions between the two have begun to be studied. Therefore, continued research in this area of electromagnetic and electronic engineering must remain.

Appendix A: MATLAB Codes

A.1 TM Polarized N-Layer Reflection and Transmission

```
1D (Planar) N-Layer P-Polarization (TM)

clear all
close all

eps = [2.25 complex(-12.6,1.1) 1]; %Relative permittivity
d = 10^-9*[0 50 0]; %Layer thickness matrix
[epsr,epsc] = size(eps); %epsc reports number of layers
lambda0 = 632.8*10^-9; %Free space wavelength
k = 2*pi/lambda0; %Free space wavevector
e = exp(1);
i = complex(0,1);

tc = 180*asin(sqrt(eps(epsc)/eps(1)))/pi;

tmin = 0;
tmax = 90;
tdmin = 40;
tdmax = 50;
nt = 2000;
dt = (tmax-tmin)/nt;

tj = 0;
for t = tmin:dt:tmax;
    tj = tj + 1
    Mtot = [1 0;0 1];
    for n = 1:1:epsc;
        kz(n) = k*sqrt(eps(n) -
        eps(1)*(sin(t*pi/180))^2);
        Kz(n) = kz(n)/eps(n);
        if ((n > 1) && (n < epsc))
```

```
            M(:,:,n) = [cos(kz(n)*d(n))
            (sin(kz(n)*d(n)))/(Kz(n));-Kz(n)*
            (sin(kz(n)*d(n))) cos(kz(n)*d(n))];
            Mtot = M(:,:,n)*Mtot;
        end
    end
    rp = -(Kz(1)*Kz(epsc)*Mtot(1,2)+Mtot(2,1)
    -i*Kz(epsc)*Mtot(1,1)+i*Kz(1)*Mtot(2,2))
    /(Kz(1)*Kz(epsc)*Mtot(1,2)-Mtot(2,1)+
    i*Kz(epsc)*Mtot(1,1)+i*Kz(1)*Mtot(2,2));
    tp = 2*i*Kz(1)*sqrt(eps(1,1)/eps(1,epsc))
    /(Kz(1)*Kz(epsc)*Mtot(1,2)-Mtot(2,1)+
    i*Kz(epsc)*Mtot(1,1)+i*Kz(1)*Mtot(2,2));
    Rp = conj(rp)*rp;
    Tp = (kz(epsc)/kz(1))*conj(tp)*tp;

    figure(1)
    plot(t,real(tp),'.',t,imag(tp),'+')
    xlabel('Theta (deg)');
    ylabel('Rp and Tp');
    title('p-Polarization');
    hold on
end
hold off
```

A.2 TM Polarized Gaussian Weighted Sum of Plane Waves for Two-Layer System

```
clear all
close all

%User Input
lambda0 = 632.8*10^-9; %Free space wavelength
k0 = 2*pi/lambda0; %Free space wavenumber
phi = 0; %Azimuthal angle
theta = 45; %Polar angle
eps1 = 1; %Relative permittivity of medium 1
eps2 = 1; %Relative permittivity of medium 2
L = 9.15*lambda0; %Effective size of envelope
W = L/16; %Gaussian beam width

%Physical Constants
```

```
i = complex(0,1);
pi = 3.14159265;
e = exp(1);

%Calculations
k1 = k0*sqrt(eps1);
k2 = k0*sqrt(eps2);
k0x = k1*cos(pi*phi/180)*sin(pi*theta/180);
k0y = k1*sin(pi*phi/180)*sin(pi*theta/180);
k0z = k1*cos(pi*theta/180);

xmin = -L/2;
xmax = L/2;
nx = 100;
dx = (xmax-xmin)/nx;

zmin = -L/2;
zmax = L/2;
nz = nx;
dz = (zmax-zmin)/nz;

xj = 0;
%For loop summing over all kx and ky values
for x = xmin:dx:xmax
    xj = xj + 1
    y = 0;
    zj = 0;
    for z = zmin:dz:zmax
        zj = zj + 1;
        Hysum = 0;
        Hyrsum = 0;
        Hytsum = 0;
        Exsum = 0;
        Exrsum = 0;
        Extsum = 0;
        Ezsum = 0;
        Ezrsum = 0;
        Eztsum = 0;
        for kx = -k1:2*pi/L:k1
            for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
            sqrt((k1)^2-(kx)^2)
                k1z = sqrt((k1)^2-(kx)^2-(ky)^2);
                k2z = sqrt((k2)^2-(kx)^2-(ky)^2);
                if ((k1z*eps2 + k2z*eps1) == 0)
                    rp = 1;
```

```
                    tp = 0;
                else
                    rp = (k1z*eps2 - k2z*eps1)/
                    (k1z*eps2 + k2z*eps1);
                    tp = 2*k1z*eps2/
                    (k1z*eps2 + k2z*eps1);
                end
                if (z <= 0)
                    Hysum = e^(i*(kx*x+ky*y+k1z*z))*
                    e^(-((((kx-k0x)^2+(ky-k0y)^2)*W^2)/2)
                    +Hysum;
                    Hyrsum = rp*e^(i*(kx*x+ky*y-k1z*z))
                    *e^(-((((kx-k0x)^2+(ky-k0y)^2)*W^2)/2)
                    +Hyrsum;
                    Exsum = (k1z/sqrt((kx)^2+(k1z)^2))
                    *Hysum+Exsum;
                    Exrsum= (k1z/sqrt((kx)^2+(k1z)^2))
                    *Hyrsum+Exrsum;
                    Ezsum = (-kx/sqrt((kx)^2+(k1z)^2))
                    *Hysum+Ezsum;
                    Ezrsum = (-kx/sqrt((kx)^2+(k1z)^2))
                    *Hyrsum+Ezrsum;
                elseif (z > 0)
                    Hytsum = tp*e^(i*(kx*x+ky*y+k2z*z))
                    *e^(-((((kx-k0x)^2+(ky-k0y)^2)*W^2)/2)
                    +Hytsum;
                    Extsum = sqrt(eps1/eps2)*
                    (k2z/sqrt((kx)^2+(k2z)^2))*Hytsum+Extsum;
                    Eztsum = sqrt(eps1/eps2)*
                    (-kx/sqrt((kx)^2+(k2z)^2))*Hytsum+Eztsum;
                end
            end
        end
        Hy1(zj,xj) = real(Hysum + Hyrsum);
        Hy2(zj,xj) = real(Hytsum);
        Ex1(zj,xj) = real(Exsum + Exrsum);
        Ex2(zj,xj) = real(Extsum);
        Ez1(zj,xj) = real(Ezsum + Ezrsum);
        Ez2(zj,xj) = real(Eztsum);
    end
end

xj = 0;
for x = xmin:dx:xmax;
    xj = xj + 1;
```

```
    zj = 0;
    for z = zmin:dz:zmax
        zj = zj + 1;
        if (z <= 0)
            Hy(zj,xj) = Hy1(zj,xj);
            Ex(zj,xj) = Ex1(zj,xj);
            Ez(zj,xj) = Ez1(zj,xj);
        elseif (z > 0)
            Hy(zj,xj) = Hy2(zj,xj);
            Ex(zj,xj) = Ex2(zj,xj);
            Ez(zj,xj) = Ez2(zj,xj);
        end
        E(zj,xj) = sqrt((Ex(zj,xj).^2)
        +(Ez(zj,xj).^2));
    end
end

x = 10^6*xmin:10^6*dx:10^6*xmax;
z = 10^6*zmin:10^6*dz:10^6*zmax;
xj = 1:1:(xmax-xmin)/dx+1;
zj = xj;

figure(1)
pcolor(x,z,Ex(zj,xj))
shading interp
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')

figure(2)
pcolor(x,z,Ez(zj,xj))
shading interp
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')

figure(3)
pcolor(x,z,E(zj,xj))
shading interp
title('Etot')
colorbar

figure(4)
```

```
quiver(x,z,Ex(zj,xj),Ez(zj,xj))
shading interp
title('E-vector')
colorbar

figure(5)
pcolor(x,z,Hy(zj,xj))
shading interp
title('Hy')
colorbar
```

A.3 TM Polarized Gaussian Weighted Sum of Plane Waves for N-Layer System

```
clear all
close all
format long

%User Input
eps0 = 8.85*10^-12;
eps = [2.25 complex(-12.6,1.1) 1]; %Relative premittivity
[epsr,epsc] = size(eps);
d = 10^-9*[0 50 0]; %Thickness of each layer
theta = 43.6; %Polar angle
phi = 0; %Azimuthal angle

lambda0 = 632.8*10^-9; %Free space incident wave wavelength
k0 = 2*pi/lambda0; %Incident wavenumber in free space
c = 2.997*10^8;
omega = c*k0;
L = 9.15*lambda0; %Effective size of gaussian incidence
W = L/16; %Gaussian beam width

%Physical Constants
i = complex(0,1);
pi = 3.14159265;
e = exp(1);

%Calculations
k1 = k0*sqrt(eps(1));
kx0 = k1*cos(pi*phi/180)*sin(pi*theta/180);
ky0 = k1*sin(pi*phi/180)*sin(pi*theta/180);
```

```
kz0 = k1*cos(pi*theta/180);

xmin = -L/2;
xmax = L/2;
nx = 100;
dx = (xmax-xmin)/nx;

ztot = sum(d);
zmin = -L/2;
zmax = L/2;
nz = nx;
dz = (zmax-zmin)/nz;

zsum(1) = 0;
for zj = 2:1:epsc;
    zsum(zj) = zsum(zj-1) + d(zj);
end

kj = 0;
for kx = -k1:2*pi/L:k1
    for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:sqrt((k1)^2-(kx)^2)
        kj = kj + 1;
        Mtot(1,1,kj) = 1;
        Mtot(2,1,kj) = 0;
        Mtot(1,2,kj) = 0;
        Mtot(2,2,kj) = 1;
        for nj = 1:1:epsc
            k(nj) = k0*sqrt(eps(nj));
            kz(nj,kj) = sqrt((k(nj))^2-(kx)^2-(ky)^2);
            Kz(nj,kj) = kz(nj,kj)/eps(nj);
            if ((nj > 1) && (nj < epsc))
                M(:,:,nj,kj) = [cos(kz(nj,kj)*d(nj))
                sin(kz(nj,kj)*d(nj))/Kz(nj,kj);
                -Kz(nj,kj)*sin(kz(nj,kj)*d(nj)),
                cos(kz(nj,kj)*d(nj))];
                if Kz(nj,kj) == 0
                    M(1,2,nj,kj) = 1;
                end
                Mtot(:,:,kj) = M(:,:,nj,kj)*Mtot(:,:,kj);
            end
        end
    end
end

for kj = 1:1:kj
```

```
    if (Kz(1,kj)*Kz(epsc,kj)*Mtot(1,2,kj)-Mtot(2,1,kj)+
    i*Kz(epsc,kj)*Mtot(1,1,kj)+i*Kz(1,kj)*Mtot(2,2,kj)) == 0
        rp(kj) = 1;
        tp(kj) = 0;
        tp0(kj) = 0;
    else
        rp(kj) = -(Kz(1,kj)*Kz(epsc,kj)*
        Mtot(1,2,kj)+Mtot(2,1,kj)-i*Kz(epsc,kj)*Mtot(1,1,kj)+
        i*Kz(1,kj)*Mtot(2,2,kj))/(Kz(1,kj)*Kz(epsc,kj)
        *Mtot(1,2,kj)-Mtot(2,1,kj)+i*Kz(epsc,kj)
        *Mtot(1,1,kj)+i*Kz(1,kj)*Mtot(2,2,kj));
        tp(kj) = 2*i*Kz(1,kj)*sqrt(eps(1)/eps(epsc))/
        (Kz(1,kj)*Kz(epsc,kj)*Mtot(1,2,kj)-
        Mtot(2,1,kj)+i*Kz(epsc,kj)*Mtot(1,1,kj)+
        i*Kz(1,kj)*Mtot(2,2,kj));
        tp0(kj) = 2*i*Kz(1,kj)*e^(-j*kz(epsc,kj)*ztot)/
        (Kz(1,kj)*Kz(epsc,kj)*Mtot(1,2,kj)-Mtot(2,1,kj)+
        i*Kz(epsc,kj)*Mtot(1,1,kj)+i*Kz(1,kj)*Mtot(2,2,kj));
    end
    Rp(kj) = real(conj(rp(kj))*rp(kj));
    Tp(kj) = real((kz(epsc,kj)/kz(1,kj))*conj(tp(kj))*tp(kj));
end

y = 0;
xj = 0;
%For loop summing over all potential kx and ky values
for x = xmin:dx:xmax
    xj = xj + 1
    zj = 0;
    for z = zmin:dz:zsum(1)
        zj = zj + 1;
        kj = 0;
        Hysum = 0;
        Exsum = 0;
        Ezsum = 0;
        H0sum = 0;
        E0sum = 0;
        for kx = -k1:2*pi/L:k1
            for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
            sqrt((k1)^2-(kx)^2)
                kj = kj + 1;
                Hysum = (e^(j*kz(1,kj)*z)-rp(kj)*
                e^(-j*kz(1,kj)*z))*e^(i*(kx*x+ky*y))*
                e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
                 + Hysum;
```

```
            Exsum = (kz(1,kj)/sqrt((kx)^2+(kz(1,kj))^2))
            *(Kz(1,kj)/(omega*eps0))*(e^(j*kz(1,kj)*z)
            +rp(kj)*e^(-j*kz(1,kj)*z))*e^(i*(kx*x+ky*y))
            *e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2) + Exsum;
            Ezsum = (-kx/sqrt((kx)^2+(kz(1,kj))^2))*
            (kx/(omega*eps0*eps(1)))*(e^(j*kz(1,kj)*z)
            +rp(kj)*e^(-j*kz(1,kj)*z))*e^(i*(kx*x+ky*y))
            *e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2) + Ezsum;
            H0sum = e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
             + H0sum;
            E0sum = sqrt(conj((kz(1,kj)/sqrt((kx)^2+
            kz(1,kj))^2))*(Kz(1,kj)/(omega*eps0))).*
            (kz(1,kj)/sqrt((kx)^2+(kz(1,kj))^2))*
            (Kz(1,kj)/(omega*eps0)) +
            conj((-kx/sqrt((kx)^2+(kz(1,kj))^2))*
            (kx/(omega*eps0*eps(1)))).*(-kx/sqrt((kx)^2+
            (kz(1,kj))^2))*(kx/(omega*eps0*eps(1))))*
            e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2) + E0sum;
        end
    end
    Hy(zj,xj) = Hysum;
    Ex(zj,xj) = Exsum;
    Ez(zj,xj) = Ezsum;
    E(zj,xj) = sqrt(conj(Ex(zj,xj)).*Ex(zj,xj)
     + conj(Ez(zj,xj)).*Ez(zj,xj));
    H0 = H0sum;
    E0 = E0sum;
end
for nj = 2:1:epsc-1
    for z = zsum(nj-1):dz:zsum(nj)
        zj = zj + 1;
        Hysum = 0;
        Exsum = 0;
        Ezsum = 0;
        kj = 0;
        for kx = -k1:2*pi/L:k1
            for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
            sqrt((k1)^2-(kx)^2)
                kj = kj + 1;
                BC(1,1,1,kj) = 1-rp(kj);
                BC(2,1,1,kj) = j*Kz(1,kj)*
                (1+rp(kj));
                BC(:,:,nj,kj) = M(:,:,nj,kj)*
                BC(:,:,nj-1,kj);
                if Kz(nj,kj) == 0
```

```
                Hysum = (BC(1,1,nj-1,kj)
                *cos(kz(nj,kj)*(z-zsum(nj-1))))
                *e^(i*(kx*x+ky*y))*e^(-(((kx-kx0)^2+
                (ky-ky0)^2)*W^2)/2) + Hysum;
                Ezsum = (-kx/sqrt((kx)^2+(kz(nj,kj))^2))
                *(kx/(omega*eps(nj)*eps0))*
                (BC(1,1,nj-1,kj)*cos(kz(nj,kj)*
                (z-zsum(nj-1))))*e^(i*(kx*x+ky*y))*
                e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
                 + Ezsum;
            else
                Hysum = (BC(1,1,nj-1,kj)*cos(kz(nj,kj)*
                (z-zsum(nj-1)))+(BC(2,1,nj-1,kj)/
                Kz(nj,kj))*sin(kz(nj,kj)*(z-zsum(nj-1))))*
                e^(i*(kx*x+ky*y))*e^(-(((kx-kx0)^2+
                (ky-ky0)^2)*W^2)/2) + Hysum;
                Ezsum = (-kx/sqrt((kx)^2+
                (kz(nj,kj))^2))*
                (kx/(omega*eps(nj)*eps0))*
                (BC(1,1,nj-1,kj)*cos(kz(nj,kj)*
                (z-zsum(nj-1)))+(BC(2,1,nj-1,kj)
                /Kz(nj,kj))*sin(kz(nj,kj)*
                (z-zsum(nj-1))))*e^(i*(kx*x+ky*y))*
                e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
                 + Ezsum;
            end
            Exsum = (kz(nj,kj)/sqrt((kx)^2+
            (kz(nj,kj))^2))*(-j/(omega*eps0))*
            (BC(2,1,nj-1,kj)*cos(kz(nj,kj)*
            (z-zsum(nj-1)))-Kz(nj,kj)*
            BC(1,1,nj-1,kj)*sin(kz(nj,kj)*
            (z-zsum(nj-1))))*e^(i*(kx*x+ky*y))*
            e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
             + Exsum;

        end
    end
    Hy(zj,xj) = Hysum;
    Ex(zj,xj) = Exsum;
    Ez(zj,xj) = Ezsum;
    E(zj,xj) = sqrt(conj(Ex(zj,xj)).*
    Ex(zj,xj) + conj(Ez(zj,xj)).*Ez(zj,xj));
  end
end
for z = zsum(epsc-1):dz:zmax
```

```
        zj = zj + 1;
        kj = 0;
        Hysum = 0;
        Exsum = 0;
        Ezsum = 0;
        for kx = -k1:2*pi/L:k1
            for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
            sqrt((k1)^2-(kx)^2)
                kj = kj + 1;
                Hysum = tp(kj)*e^(j*(kx*x+ky*y+
                kz(epsc,kj)*(z-zsum(epsc))))*
                e^(-(((kx-kx0)^2+(ky-ky0)^2)*W^2)/2)
                 + Hysum;
                Exsum = (kz(epsc,kj)/sqrt((kx)^2+
                (kz(epsc,kj))^2))*(Kz(epsc,kj)
                /(omega*eps0))*tp(kj)*
                e^(j*(kx*x+ky*y+kz(epsc,kj)*
                (z-zsum(epsc))))*e^(-(((kx-kx0)^2+
                (ky-ky0)^2)*W^2)/2) + Exsum;
                Ezsum = (-kx/sqrt((kx)^2+
                (kz(epsc,kj))^2))*
                (kx/(omega*eps(epsc)*eps0))*
                tp(kj)*e^(j*(kx*x+ky*y+
                kz(epsc,kj)*(z-zsum(epsc))))*
                e^(-(((kx-kx0)^2+(ky-ky0)^2)*
                W^2)/2) + Ezsum;
            end
        end
        Hy(zj,xj) = Hysum;
        Ex(zj,xj) = Exsum;
        Ez(zj,xj) = Ezsum;
        E(zj,xj) = sqrt(conj(Ex(zj,xj)).*
        Ex(zj,xj) + conj(Ez(zj,xj)).*Ez(zj,xj));
    end
end

x = 10^6*xmin:10^6*dx:10^6*xmax;
z = 10^6*zmin:10^6*dz:10^6*zmax;
xj = 1:1:(xmax-xmin)/dx+1;
zj = 1:1:(zmax-zmin)/dz+1;

figure(1)
pcolor(x,z,real(Hy(zj,xj))/H0)
shading interp
title('Hy')
```

```
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')

figure(2)
pcolor(x,z,real(Ex(zj,xj))/E0)
title('Ex')
shading interp
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')


figure(3)
pcolor(x,z,real(Ez(zj,xj))/E0)
shading interp
title('Ez')
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')


figure(4)
pcolor(x,z,real(E(zj,xj))/E0)
shading interp
title('Etot')
colorbar
colormap gray
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')
```

A.4 TM Polarization Rigorous Coupled-Wave Analysis

```
clear all
close all
format long
```

```
%Define system physical constants
order = 0;
nhmax = 50;
lambda0 = 1e-6;
k0 = 2*pi/lambda0;
c = 2.997e8;
omega = c*k0;
eps0 = 8.85e-12;
mu0 = 1.25e-7;
theta = 50;
n1 = 2;
eps1 = n1^2;
n2 = 1;
eps2 = n2^2;
ff = 0.5;
d = 1e-6;
Tl = 1e-6;
K = 2*pi/Tl;
kx = k0*n1*sin(pi*theta/180);
k1z = k0*n1*cos(pi*theta/180);
e = exp(1);

%Field visualization parameters
xmin = Tl;
xmax = 6*Tl;
nx = 100;
dx = (xmax-xmin)/nx;
zmin = -2*Tl;
zmax = 3*Tl;
nz = nx;
dz = (zmax-zmin)/nz;

for NH = abs(order)+1:1:nhmax;
    ijmax = floor((NH-1)/2);
    nn = 2*ijmax+1;

    %Setup calculation matrices
    delta = zeros(nn,1);
    k1_zi = zeros(nn,1);
    k2_zi = zeros(nn,1);
    Identity = speye(nn); %Sparse identity matrix

    %Calculations
    ij_prime = -ijmax:ijmax;
    ij2_prime = -2*ijmax:2*ijmax;
```

```
ij = ij_prime';
ij2 = ij2_prime';
delta(ijmax+1,1) = 1;
k_xi = kx-ij*K;
K_xi = spdiags(k_xi,0,nn,nn)./k0;
a1 = find(k_xi.^2<(k0*n1)^2);
b1 = find(k_xi.^2>(k0*n1)^2);
k1_zi(a1) = k0*sqrt(eps1 - (k_xi(a1)./k0).^2);
k1_zi(b1) = -j*k0*sqrt((k_xi(b1)./k0).^2 - eps1);
Z_I = spdiags(k1_zi,0,nn,nn)./(k0*eps1);
a2 = find(k_xi.^2<(k0*n2)^2);
b2 = find(k_xi.^2>(k0*n2)^2);
k2_zi(a2) = k0*sqrt(eps2 - (k_xi(a2)./k0).^2);
k2_zi(b2) = -j*k0*sqrt((k_xi(b2)./k0).^2 - eps2);
Z_II = spdiags(k2_zi,0,nn,nn)./(k0*eps2);

%Calculate and create Fourier coefficients for Eps matrix
epscoef(ij2+nn) = (j./(2*pi.*ij2)).*
(eps1.*(e^(j*pi.*ij2.*ff)-e^(-j*pi.*ij2))+
eps2.*(e^(-j*pi.*ij2.*ff)-e^(j*pi.*ij2.*ff))+
eps1.*(e^(-j*pi.*ij2)-e^(-j*pi.*ij2.*ff)));
epsbar(ij2+nn) = (j./(2*pi.*ij2)).*
((1/eps1).*(e^(j*pi.*ij2.*ff)-e^(-j*pi.*ij2))+
(1/eps2).*(e^(-j*pi.*ij2.*ff)-e^(j*pi.*ij2.*ff))+
(1/eps1).*(e^(-j*pi.*ij2)-e^(-j*pi.*ij2.*ff)));
epscoef(nn) = eps1 + (eps2-eps1)*ff;
epsbar(nn) = 1/eps1 + (1/eps2-1/eps1)*ff;
for aj = 1:1:nn
    for bj = 1:1:nn
        E(bj,aj) = epscoef(nn+aj-bj);
        Ebar(bj,aj) = epsbar(nn+aj-bj);
    end
end
B = Identity-K_xi*E\K_xi;
C = Ebar\B;
if ijmax == 0
    W = 1;
    Q = C;
else
    [W,Q] = eig(C);
end
q = sqrt(diag(Q));
Q = spdiags(q,0,nn,nn);
V = E\W*Q;
X = spdiags(e^(-k0*q*d),0,nn,nn);
```

```
    M = [[j*Z_I*W+V ,(j*Z_I*W-V)*X];
    [(-j*Z_II*W+V)*X, -j*Z_II*W-V]];
    b = [j*full(Z_I*delta+delta.*cosd(theta)/n1);zeros(nn,1)];
    c = M\b;
    cplus = c(1:nn);
    cminus = c(nn+1:2*nn);
    r = W*cplus + W*X*cminus - delta;
    t = (n1/n2)*(W*X*cplus + W*cminus);
    R = conj(r).*r;
    T = conj(t).*t;
    DEr = R.*real(k1_zi./k1z);
    DEt = T.*real(k2_zi./k1z);
    DEstr = num2str([NH DEr(ijmax+1+order)
    DEt(ijmax+1+order) sum(DEr)+sum(DEt)];
    figure(1)
    plot(NH,DEr(ijmax+1+order),'.',NH,
    DEt(ijmax+1+order),'+')
    axis([1 nhmax 0 1])
    title(sprintf(DEstr))
    xlabel('Inteations')
    ylabel('DE(0)')
    hold on
end
hold off

%Loop over visualization grid
xj = 0;
for x = xmin:dx:xmax
    xj = xj + 1
    zj = 0;
    for z = zmin:dz:zmax
        zj = zj + 1;
        %Calculate magnetic fields within eps1 (z < 0) region
        if (z < 0)
            Hysum = zeros(nn,1);
            Exsum = zeros(nn,1);
            Ezsum = zeros(nn,1);
            for nj = 1:1:nn
                Hysum(nj) = r(nj)*e^(-j*(k_xi(nj)*x-
                k0*Z_I(nj,nj)*z));
                Exsum(nj) = (Z_I(nj,nj)/(omega*eps0*eps1))*
                r(nj)*e^(-j*(k_xi(nj)*x-k0*Z_I(nj,nj)*z));
                Ezsum(nj) = (k_xi(nj)/(omega*eps0))*r(nj)*
                e^(-j*(k_xi(nj)*x-k0*Z_I(nj,nj)*z));
            end
```

```
        Hy(zj,xj) = e^(j*(kx*x+k1z*z)) -
        sum(Hysum);
        Ex(zj,xj) = (k1z/(omega*eps0*eps1))*
        e^(j*(kx*x+k1z*z)) + sum(Exsum);
        Ez(zj,xj) = (kx/(omega*eps0))*
        e^(j*(kx*x+k1z*z)) + sum(Ezsum);
        E0 = sqrt(conj((k1z/(omega*eps0*eps1))).*
        (k1z/(omega*eps0*eps1)) +
        conj((kx/(omega*eps0))).*(kx/(omega*eps0)));
    %Calculate magnetic fields within grating region
    elseif (z >= 0 && z <= d)
        Hysum = zeros(nn,1);
        Exsum = Hysum;
        Ezsum = Hysum;
        for nj = 1:1:nn
            Uyisum = zeros(nn,1);
            Syisum = zeros(nn,1);
            for mj = 1:1:nn
                Uyisum(mj) = W(nj,mj)*(cplus(mj)*
                e^(-k0*q(mj)*z)+cminus(mj)*
                e^(k0*q(mj)*(z-d)));
                Syisum(mj) = V(nj,mj)*(-cplus(mj)*
                e^(-k0*q(mj)*z)+cminus(mj)*
                e^(k0*q(mj)*(z-d)));
            end
            Uyi(nj) = sum(Uyisum);
            Syi(nj) = sum(Syisum);
            Hysum(nj) = Uyi(nj)*e^(j*k_xi(nj)*x);
            Exsum(nj) = j*sqrt(mu0/eps0)*Syi(nj)*
            e^(j*k_xi(nj)*x);
            Ezsum(nj) = (k_xi(nj)/(omega*eps0))*
            Uyi(nj)*e^(j*k_xi(nj)*x);
        end
        Hy(zj,xj) = sum(Hysum);
        Ex(zj,xj) = sum(Exsum);
        Ez(zj,xj) = sum(Ezsum);
    %Calculate magnetic fields within eps2 (z > d) region
    else
        Hysum = zeros(nn,1);
        Exsum = Hysum;
        Ezsum = Hysum;
        for nj = 1:1:nn
            if imag(eps2) == 0
                Hysum(nj) = t(nj)*e^(-j*(k_xi(nj)*x+
                k0*Z_II(nj,nj)*(z-d)));
```

```
                    Exsum(nj) = (Z_II(nj,nj)*k0/(omega*eps0))*
                    t(nj)*e^(-j*(k_xi(nj)*x+
                    k0*Z_II(nj,nj)*(z-d)));
                    Ezsum(nj) = (k_xi(nj)/(omega*eps0))*
                    t(nj)*e^(-j*(k_xi(nj)*x+
                    k0*Z_II(nj,nj)*(z-d)));
                else
                    Hysum(nj) = t(nj)*e^(j*(k_xi(nj)*x+
                    k0*Z_II(nj,nj)*(z-d)));
                    Exsum(nj) = (Z_II(nj,nj)*k0/(omega*eps0))*
                    t(nj)*e^(j*(k_xi(nj)*x+
                    k0*Z_II(nj,nj)*(z-d)));
                    Ezsum(nj) = (k_xi(nj)/
                    (omega*eps0))*t(nj)*
                    e^(j*(k_xi(nj)*x+k0*Z_II(nj,nj)*(z-d)));
                end
            end
            Hy(zj,xj) = sum(Hysum);
            Ex(zj,xj) = sum(Exsum);
            Ez(zj,xj) = sum(Ezsum);
        end
    end
end

x = 10^6*xmin:10^6*dx:10^6*xmax;
xj = 1:1:nx+1;
z = 10^6*zmin:10^6*dz:10^6*zmax;
zj = 1:1:nz+1;

figure(2)
pcolor(x,z,real(Hy(zj,xj)))
colorbar
colormap gray
shading interp
xlabel('x (\mu m)')
xlabel('z (\mu m)')

figure(3)
pcolor(x,z,real(Ex(zj,xj))/E0)
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')
```

```
figure(4)
pcolor(x,z,real(Ez(zj,xj))/E0)
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
ylabel('z (\mu{m})')
```

A.5 TM Polarization Rigorous Coupled-Wave Analysis with Gaussian Weighted Sum of Plane Waves

```
clear all
close all
format long

%Define system physical constants
order = 0;
lambda0 = 632.8e-9;
k0 = 2*pi/lambda0;
c = 2.997*10^8;
omega = c*k0;
theta = 45;
phi = 0;
eps0 = 8.85e-12;
n1 = 2;
eps1 = n1^2;
n2 = 1;
eps2 = n2^2;
ff = 0.5;
d = 1e-6;
Tl = 1e-6;
Kl = 2*pi/Tl;
L = 9.15*lambda0; %Effective size of gaussian incidence
Wg = L/16; %Gaussian beam width
k1 = k0*n1;
k2 = k0*n2;
kx0 = k1*cos(pi*phi/180)*sin(pi*theta/180);
ky0 = k1*sin(pi*phi/180)*sin(pi*theta/180);
kz0 = k1*cos(pi*theta/180);
e = exp(1);
```

```
%Field visualization parameters
xmin = -L/2;
xmax = L/2;
nx = 50;
dx = (xmax-xmin)/nx;

zmin = -L/2;
zmax = L/2;
nz = nx;
dz = (zmax-zmin)/nz;

NH = ceil(k1*L/pi);
ijmax = floor((NH-1)/2);
nn = 2*ijmax+1

ij_prime = -ijmax:ijmax;
ij2_prime = -2*ijmax:2*ijmax;
ij = ij_prime';
ij2 = ij2_prime';

%Calculate and create Fourier coefficients for Eps matrix
epscoef(ij2+nn) = (j./(2*pi.*ij2)).*
(eps1.*(e^(j*pi.*ij2.*ff)-e^(-j*pi.*ij2))+
eps2.*(e^(-j*pi.*ij2.*ff)-e^(j*pi.*ij2.*ff))+
eps1.*(e^(-j*pi.*ij2)-e^(-j*pi.*ij2.*ff)));
epsbar(ij2+nn) = (j./(2*pi.*ij2)).*((1/eps1).*
(e^(j*pi.*ij2.*ff)-e^(-j*pi.*ij2))+
(1/eps2).*(e^(-j*pi.*ij2.*ff)-e^(j*pi.*ij2.*ff))+
(1/eps1).*(e^(-j*pi.*ij2)-e^(-j*pi.*ij2.*ff)));
epscoef(nn) = eps1 + (eps2-eps1)*ff;
epsbar(nn) = 1/eps1 + (1/eps2-1/eps1)*ff;
for aj = 1:1:nn
    for bj = 1:1:nn
        E(bj,aj) = epscoef(nn+aj-bj);
        Ebar(bj,aj) = epsbar(nn+aj-bj);
    end
end

disp('Calculating RCWA ...')
kj = 0;
for kx = -k1:2*pi/L:k1
    for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:sqrt((k1)^2-(kx)^2)
        kj = kj + 1;
        kz1(kj) = sqrt((k1)^2-(kx)^2-(ky)^2);
        Kz1(kj) = kz1(kj)/eps1;
```

```
    kz2(kj) = sqrt((k2)^2-(kx)^2-(ky)^2);
    Kz2(kj) = kz2(kj)/eps2;

    %Setup calculation matrices
    delta = zeros(nn,1);
    k_xikj = zeros(nn,1);
    k1_zikj = zeros(nn,1);
    k2_zikj = zeros(nn,1);
    Identity = speye(nn); %Sparse identity matrix

    %Calculations
    delta(ijmax+1,1) = 1;
    k_xikj = kx-ij*Kl;
    K_xi = spdiags(k_xikj,0,nn,nn)./k0;
    a1 = find((k_xikj.^2+ky^2)<(k0*n1)^2);
    a2 = find((k_xikj.^2+ky^2)>(k0*n1)^2);
    k1_zikj(a1) = k0*sqrt(eps1 - (k_xikj(a1)./k0).^2);
    k1_zikj(a2) = -j*k0*sqrt((k_xikj(a2)./k0).^2 - eps1);
    Z_I = spdiags(k1_zikj,0,nn,nn)./(k0*eps1);
    a3 = find((k_xikj.^2+ky^2)<(k0*n2)^2);
    a4 = find((k_xikj.^2+ky^2)>(k0*n2)^2);
    k2_zikj(a3) = k0*sqrt(eps2 - (k_xikj(a3)./k0).^2);
    k2_zikj(a4) = -j*k0*sqrt((k_xikj(a4)./k0).^2 - eps2);
    Z_II = spdiags(k2_zikj,0,nn,nn)./(k0*eps2);

%     B = K_xi*E\K_xi-SPEYE_NH;
    B = Identity-K_xi*inv(E)*K_xi;
    C = Ebar\B;
    if ijmax == 0
        Wkj = 1;
        Q = C;
    else
        [Wkj,Q] = eig(C);
    end
    qkj = sqrt(diag(Q));
    Qkj = spdiags(qkj,0,nn,nn);
    Vkj = E\Wkj*Qkj;
    X = spdiags(e^(-k0*qkj*d),0,nn,nn);
    M = [[j*Z_I*Wkj+Vkj ,(j*Z_I*Wkj-Vkj)*X];
    [(-j*Z_II*Wkj+Vkj)*X, -j*Z_II*Wkj-Vkj]];
    b = [j*full(Z_I*delta+delta.*cosd(theta)/n1);
    zeros(nn,1)];
    c = M\b;
    cplus = c(1:nn);
    cminus = c(nn+1:2*nn);
```

```
        r = Wkj*cplus + Wkj*X*cminus - delta;
        t = (n1/n2)*(Wkj*X*cplus + Wkj*cminus);
        t0 = Wkj*X*cplus + Wkj*cminus;
        R = conj(r).*r;
        T = conj(t).*t;
        DEr = R.*real(k1_zikj./kz0);
        DEt = T.*real(k2_zikj./kz0);
        for aj = 1:nn
            k_xi(aj,kj) = k_xikj(aj);
            k1_zi(aj,kj) = k1_zikj(aj);
            k2_zi(aj,kj) = k2_zikj(aj);
            q(aj,kj) = qkj(aj);
            cp(aj,kj) = cplus(aj);
            cm(aj,kj) = cminus(aj);
            rp(aj,kj) = r(aj);
            tp(aj,kj) = t(aj);
            tp0(aj,kj) = t0(aj);
            for bj = 1:nn
                W(bj,aj,kj) = Wkj(bj,aj);
                V(bj,aj,kj) = Vkj(bj,aj);
            end
        end
    end
end

disp('RCWA Calculations ... done!')

%Loop over visualization grid
xj = 0;
y = 0;
for x = xmin:dx:xmax
    xj = xj + 1
    zj = 0;
    %Calculate fields within eps1 (z < 0) region
    for z = zmin:dz:zmax
        if (z <= 0)
            zj = zj + 1;
            Hysum = 0;
            Exsum = 0;
            Ezsum = 0;
            H0 = 0;
            E0 = 0;
            H0r = 0;
            E0r = 0;
            kj = 0;
```

```
for kx = -k1:2*pi/L:k1
    for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
    sqrt((k1)^2-(kx)^2)
        kj = kj + 1;
        Hyrsum = 0;
        Exrsum = 0;
        Ezrsum = 0;
        for nj = 1:nn
            Hyrsum = -rp(nj,kj)*
            e^(j*(-k_xi(nj,kj)*x+k1_zi(nj,kj)*z))
            *e^(-(((k_xi(nj,kj)-kx0)^2+
            (ky-ky0)^2)*Wg^2)/2)*e^(j*ky*y)
             + Hyrsum;
            Exrsum = (k1_zi(nj,kj)/
            sqrt((k_xi(nj,kj))^2+
            (k1_zi(nj,kj))^2))*
            (k1_zi(nj,kj)/(omega*eps0*eps1))*
            rp(nj,kj)*e^(j*(k_xi(nj,kj)*x+
            k1_zi(nj,kj)*z))*
            e^(-(((k_xi(nj,kj)-kx0)^2+
            (ky-ky0)^2)*Wg^2)/2)*e^(j*ky*y)
             + Exrsum;
            Ezrsum = (-k_xi(nj,kj)/
            sqrt((k_xi(nj,kj))^2+
            (k1_zi(nj,kj))^2))*
            (k_xi(nj,kj)/(omega*eps0*eps1))*
            rp(nj,kj)*
            e^(j*(k_xi(nj,kj)*x+
            k1_zi(nj,kj)*z))*
            e^(-(((k_xi(nj,kj)-kx0)^2+
            (ky-ky0)^2)*Wg^2)/2)*
            e^(j*ky*y) + Ezrsum;
            H0r = e^(-(((k_xi(nj,kj)-kx0)^2+
            (ky-ky0)^2)
            *Wg^2)/2) + H0r;
            E0r = sqrt(conj((k1_zi(nj,kj)/
            sqrt((k_xi(nj,kj))^2+
            (k1_zi(nj,kj))^2))*
            (k1_zi(nj,kj)/
            (omega*eps0*eps1))).*
            (k1_zi(nj,kj)/
            sqrt((k_xi(nj,kj))^2+
            (k1_zi(nj,kj))^2))*
            (k1_zi(nj,kj)/
            (omega*eps0*eps1)) +
```

```
    conj((k_xi(nj,kj)/
    sqrt((k_xi(nj,kj))^2+
    (k1_zi(nj,kj))^2))*
    (k_xi(nj,kj)/
    (omega*eps0*eps1))).*
    (k_xi(nj,kj)/
    sqrt((k_xi(nj,kj))^2+
    (k1_zi(nj,kj))^2))*
    (k_xi(nj,kj)/
    (omega*eps0*eps1)))*
    e^(-(((k_xi(nj,kj)-kx0)^2
    +(ky-ky0)^2)*Wg^2)/2) + E0r;
end
Hysum = e^(j*(kx*x+kz1(kj)*z))*
e^(-(((kx-kx0)^2+
(ky-ky0)^2)*Wg^2)/2)
+Hysum+Hyrsum/H0r;
Exsum = (-kz1(kj)/sqrt((kx)^2+
(kz1(kj))^2))
*(kz1(kj)/(omega*eps0*eps1))
*e^(j*(kx*x+kz1(kj)*z))*
e^(-(((kx-kx0)^2+(ky-ky0)^2)*
Wg^2)/2)+Exsum+Exrsum/E0r;
Ezsum = (kx/sqrt((kx)^2+
(kz1(kj))^2))
*(kx/(omega*eps0*eps1))*
e^(j*(kx*x+kz1(kj)*z))*
e^(-(((kx-kx0)^2+(ky-ky0)^2)*
Wg^2)/2)
+Ezsum+Ezrsum/E0r;
H0 = e^(-(((kx-kx0)^2+(ky-ky0)^2)*
Wg^2)/2)
 + H0;
E0 =  sqrt(conj((kz1(kj)/
sqrt((kx)^2+(kz1(kj))^2))*
(kz1(kj)/(omega*eps0*eps1))).*
(kz1(kj)/sqrt((kx)^2+
(kz1(kj))^2))*
(kz1(kj)/(omega*eps0*eps1)) +
conj((kx/sqrt((kx)^2+
(kz1(kj))^2))*
(kx/(omega*eps0*eps1)))
.*(kx/sqrt((kx)^2+
(kz1(kj))^2))*
(kx/(omega*eps0*eps1)))*
```

```
            e^(-(((kx-kx0)^2+(ky-ky0)^2)*
            Wg^2)/2)+E0;
        end
    end
    Hy(zj,xj) = Hysum/H0;
    Ex(zj,xj) = Exsum/E0;
    Ez(zj,xj) = Ezsum/E0;
elseif (z > 0 && z < d)
    zj = zj + 1;
    kj = 0;
    Hygsum = 0;
    Exgsum = 0;
    Ezgsum = 0;
    for kx = -k1:2*pi/L:k1
        for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
        sqrt((k1)^2-(kx)^2)
            kj = kj + 1;
            for nj = 1:nn
                Uyisum = 0;
                Syisum = 0;
                for mj = 1:nn
                    Uyisum = W(nj,mj,kj)*(cp(mj,kj)*
                    e^(-k0*q(mj,kj)*z)+cm(mj,kj)*
                    e^(k0*q(mj,kj)*(z-d)))+Uyisum;
                    Syisum = V(nj,mj,kj)*(-cp(mj,kj)*
                    e^(-k0*q(mj,kj)*z)+cm(mj,kj)*
                    e^(k0*q(mj,kj)*(z-d)))+Syisum;
                end
                Uyi(nj) = Uyisum;
                Syi(nj) = Syisum;
                Hygsum = Uyi(nj)*e^(-j*k_xi(nj,kj)*x)*
                e^(-(((k_xi(nj,kj)-kx0)^2+
                (ky-ky0)^2)*Wg^2)/2) + Hygsum;
                Exgsum = Syi(nj)*e^(-j*k_xi(nj,kj)*x)*
                e^(-(((k_xi(nj,kj)-kx0)^2+
                (ky-ky0)^2)*Wg^2)/2) + Exgsum;
                Ezgsum = (k_xi(nj,kj)/k1)*Uyi(nj)*
                e^(-j*k_xi(nj,kj)*x)*e^(-(((k_xi(nj,kj)
                -kx0)^2+(ky-ky0)^2)*Wg^2)/2) + Ezgsum;
            end
        end
    end
    Hy(zj,xj) = Hygsum/H0r;
    Ex(zj,xj) = Exgsum/E0r;
    Ez(zj,xj) = Ezgsum/E0r;
```

```
        elseif (z >= d)
            zj = zj + 1;
            Hytsum = 0;
            Extsum = 0;
            Eztsum = 0;
            kj = 0;
            for kx = -k1:2*pi/L:k1
                for ky = -sqrt((k1)^2-(kx)^2):2*pi/L:
                sqrt((k1)^2-(kx)^2)
                    kj = kj + 1;
                    for nj = 1:nn
                        Hytsum = tp(nj,kj)*e^(j*(-k_xi(nj,kj)*
                        x-k2_zi(nj,kj)*z))
                        *e^(-(((k_xi(nj,kj)-kx0)^2+(ky-\\ky0)^2)*
                        Wg^2)/2)
                        *e^(j*ky*y) + Hytsum;
                        Extsum = (k2_zi(nj,kj)/
                        sqrt((k_xi(nj,kj))^2+(k2_zi(nj,kj))^2))*
                        (k2_zi(nj,kj)/(omega*eps0*eps1))
                        *tp(nj,kj)*e^(j*(-k_xi(nj,kj)*
                        x-k2_zi(nj,kj)*z))*
                        e^(-(((k_xi(nj,kj)-kx0)^2
                        +(ky-ky0)^2)*Wg^2)/2)*e^(j*ky*y) +
                        Extsum;
                        Eztsum = (-k_xi(nj,kj)/
                        sqrt((k_xi(nj,kj))^2+(k2_zi(nj,kj))^2))*
                        (k_xi(nj,kj)/(omega*eps0*eps1))*tp(nj,kj)*
                        e^(j*(-k_xi(nj,kj)*x-k2_zi(nj,kj)*z))*
                        e^(-(((k_xi(nj,kj)-kx0)^2+
                        (ky-ky0)^2)*Wg^2)/2)*e^(j*ky*y) + Eztsum;
                    end
                end
            end
            Hy(zj,xj) = Hytsum/H0r;
            Ex(zj,xj) = Extsum/E0r;
            Ez(zj,xj) = Eztsum/E0r;
        end
        Etot(zj,xj) = sqrt(conj(Ex(zj,xj))*Ex(zj,xj)+
        conj(Ez(zj,xj))*Ez(zj,xj));
    end
end

x = 10^6*xmin:10^6*dx:10^6*xmax;
xj = 1:1:nx+1;
z = 10^6*zmin:10^6*dz:10^6*zmax;
```

```
zj = 1:1:nz+1;

figure(1)
pcolor(x,z,real(Hy(zj,xj)))
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
zlabel('z (\mu{m})')

figure(2)
pcolor(x,z,real(Ex(zj,xj)))
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
zlabel('z (\mu{m})')

figure(3)
pcolor(x,z,real(Ez(zj,xj)))
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
zlabel('z (\mu{m})')

figure(4)
pcolor(x,z,Etot(zj,xj))
colorbar
colormap gray
shading interp
xlabel('x (\mu{m})')
zlabel('z (\mu{m})')
```

Bibliography

[1] A.M. Hashim, T. Hashizumea, K. Iizukab, H. Hasegawaa, *Plasma wave interactions in the microwave to THz range between carriers in a semiconductor 2DEG and interdigital slow waves*, Superlattices and Microstructures, Vol. 34, 2003, pgs. 531-537.

[2] E.A. Shaner, M. Lee, M.C. Wanke, A.D. Grine, J.L. Reno, S.J. Allen, *Single-quantum-well grating-gated terahertz plasmon detectors*, Applied Physics Letters, Vol. 87, 2005, pgs. 193507-1-3.

[3] Y. M. Meziani, H. Handa, W. Knap, T. Otsuji, E. Sano, V.V. Popov, G.M. Tsymbalov, D. Coquillat, F. Teppe *Room temperature terahertz emission from grating coupled two-dimensional plasmons*, Applied Physics Letters, Vol. 92, 2008, pgs. 201108-1-3.

[4] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag Publishing, Berlin, Germany, 1988, pg. 11.

[5] R.W. Wood, *On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum*, Proceedings of the Physical Society of London, Vol. 18, 1902, pgs. 269-275.

[6] R.H. Ritchie, *Plasma Losses by Fast Electrons in Thin Films*, Physical Review Letters, Vol. 106, Issue 5, 1957.

[7] D. Bohm and D. Pines, *A Collective Description of Electron Interactions. I. Magnetic Interactions*, Physical Review Letters, Vol. 82, Issue 5, 1951.

[8] D. Bohm and D. Pines, *A Collective Description of Electron Interactions: II. Collective vs Individual Particle Aspects of the Interactions*, Physical Review Letters, Vol. 85, Issue 2, 1952.

[9] D. Bohm and D. Pines, *A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas*, Physical Review Letters, Vol. 92, Issue 3, 1953.

[10] H. Watanabe, *Experimental Evidence for the Collective Nature of the Characteristic Energy Loss of Electrons in Solids: Studies on the Dispersion Relation of Plasma Frequency*, Journal of the Physical Society of Japan, Vol. 11, 1956.

[11] T. Turbadar, *Complete Absorption of Light by Thin Metal Films*, Proceedings of the Physical Society, Vol. 73, Issue 1, 1959, pgs. 40-44.

[12] F. Abeles, *Optical properties of thin absorbing films*, Journal of the Optical Society of America, Vol. 47, Issue 6, 1957, pg. 473.

[13] A. Otto, *Excitation of Nonradiative Surface Plasma Waves in Silver by the Method of Frustrated Total Reflection*, Zeitschrift fr Physik A Hadrons and Nuclei, Vol. 216, Number 4, 1968, pgs. 398-410.

[14] R.F. Harrington, `Time-Harmonic Electromagnetic Fields`, McGraw-Hill, 1961.

[15] E. Kretschmann, H. Raether, *Radiative Decay of Non-Radiative Surface Plasmons Excited by Light*, Naturforschung. A, Vol. 23, 1968, pgs. 2135-1236.

[16] *http : //www.biacore.com*

[17] A. Rasooly, *Surface Plasmon Resonance Analysis of Staphylococcal Enterotoxin B in Food*, Journal of Food Protection, Vol. 64, Issue 1, 2001, pgs. 37-43.

[18] O.P. Karlsson, S. Lofas, *Flow-Mediated On-Surface Reconstruction of G-Protein Coupled Receptors for Applications in Surface Plasmon Resonance Biosensors*, Analytical Biochemistry, Vol. 300, Issue 2, 2002, pgs. 132-138.

[19] http://www.rsc.org/ej/AN/2001/b105252g/b105252g-f2.gif.

[20] Hakan Roos, *Biosensor Systems and Biacore Technology*, Biacore Life Sciences, Presentation, 2001.

[21] J. Fontanella, R.L. Johnston, J.H. Colwell and C. Andeen, *Temperature and Pressure Variation of the Refractive Index of Diamond*, Applied Optics, Vol. 16, Issue 11, 1977.

[22] *Biacore Life Sciences - Flexchip*, Biacore Life Sciences, http://www.biacore.com/lifesciences/products/systems_overview/Flexchip/, Accessed 28 July 2008.

[23] *Biacore Life Sciences - Flexchip*, Biacore Life Sciences, http://www.biacore.com/lifesciences/products/systems_overview/Flexchip/, Accessed 28 July 2008.

[24] *Thin Film Characterization by Surface Plasmon Resonance*, LOT Oriel Group - Europe (Romania), http://www.lot-oriel.com/site/pages_ro_ro/spectrum/european_edition_7/spr.php, Accessed 29 July 2008.

[25] *SPR Sensor Probe*, U.S. Patent no. 6,466,323.

[26] *Integrated opto-electric SPR sensor*, U.S. Patent application no. 20070052049.

[27] K. Okamoto, I. Niki, A. Shvartser, Y. Narukawa, T. Mukai, A. Scherer, *Surface-plasmon-enhanced light emitters based on InGaN quantum wells*, Nature Materials, Vol. 3, 2004, pgs. 601-605.

[28] I. Gontijo, M. Boroditsky, E. Yablonovitch, S. Keller, U.K. Mishra, S.P. DenBaars, *Coupling of InGaN quantum-well photoluminescence to silver surface plasmons*, Physical Review B, Vol. 60, 1999, pgs. 11564 - 11567.

[29] J. Vuckovic, M. Loncar, A, Scherer, *Surface plasmon enhanced light-emitting diode*, IEEE Journal of Quantum Electronics, Vol. 36, Issue 10, 2000, pgs. 1131-1144.

[30] K. Okamoto, I. Niki, A. Scherer, Y. Narukawa, T. Mukai, Y. Kawakami, *Surface plasmon enhanced spontaneous emission rate of InGaN/GaN quantum wells probed by time-resolved photoluminescence spectroscopy*, Applied Physics Letters, Vol. 87, 2005, pgs. 071102-1-3.

[31] A. Neogi, C.W. Lee, H.O. Everitt, T. Kuroda, A. Tackeuchi, E. Yablonovitch, *Enhancement of spontaneous recombination rate in a quantum well by resonant surface plasmon coupling*, Physical Review B, Vol. 66, Issue 15, 2002, pgs. 153305-153308.

[32] A.M. Hashim, S. Kasai, H. Hasegawa *Observation of first and third harmonic responses in two-dimensional AlGaAs/GaAs HEMT devices due to plasma wave interaction*, Superlattices and Microstructures, Vol. 44, 2008, pgs. 754-760.

[33] J.Q. Lu, M.S. Shur, *Terahertz detection by high-electron-mobility transistor: Enhancement by drain bias*, Applied Physics Letters, Vol. 78, Number 17, 2001, pgs. 2587-2588.

[34] J.C. Maxwell, *A Dynamical Theory of the Electromagnetic Field*, Philosophical Transactions of the Royal Society of London, Vol. 155, 1865, pgs. 480-488.

[35] C. Balanis, `Advanced Engineering Electromagnetics`, Wiley, 1989.

[36] G.S. Ohm, *Ueber die Definition des Tones, nebst daran geknpfter Theorie der Sirene und hnlicher tonbildender Vorrichtungen*, Annalen der Physik, Vol. 135, Issue 8, 1843, pgs.513-565.

[37] A. Sommerfeld, `Partial Differential Equations in Physics`, Academic Press, 1949.

[38] E. Hecht, *Optics*, Pearson Education, 2001.

[39] P. Tran and A. A. Maradudin, *The scattering of electromagnetic waves from a randomly rough 2-D metallic surface*, Opt. Commun., Vol. 110, 1994, pgs. 269273.

[40] V. Jandhyala, B. Shanker, E. Michielssen, W.C. Chew, *Fast Algorithm for the Analysis of Scattering by Dielectric Rough Surfaces*, Journal of the Optical Society of America, Vol. 15, Number 7, 1998, pgs. 1877-1885.

[41] V. Jandhyala,E. Michielssen, B. Shanker, W.C. Chew, *A Combined Steepest Descent-Fast Multipole Algorithm for the Fast Analysis of Three-Dimensional Scattering by Rough Surfaces*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, Vol. 36, Number 3, 1998, pgs. 738-748.

[42] J. Lekner, `Theory of Reflection`, Martinus Nijhous Publishers, Dordrecht, Germany, 1987.

[43] ed. R. Petit, `Electromagnetic Theory of Gratings`, Springer-Verlag, Berlin, 1980.

[44] J. Chandezon, D. Maystre, G. Raoult, *A New Theoretical Method for Diffraction Gratings and Its Numerical Application*, Journal of Optics, Vol. 11, Number 4, 1980, pgs. 235-241.

[45] L. Li, *Bremmer Series, R-Matrix Propagation Algorithm, and Numerical Modelling of Diffraction Gratings*, Journal of the Optical Society of America: A, Vol. 11, Number 11, 1994, pgs. 2829-2836.

[46] J.W.S. Rayleigh, *On the dynamical theory of gratings*, Proc. Roy. Soc. Lon. A, Vol. 79, 1907, pgs. 399-416.

[47] H. Kogelnik, *Coupled Wave Theory for Thick Hologram Gratings*, The Bell System Technical Journal, Vol. 24, Number 9, 1969, pgs. 2909-2947.

[48] G. Tayeb, R. Petit, *On the Numerical Study of Deep Conducting Lamellar Diffraction Gratings*, Journal of Modern Optics, Vol. 31, Issue 12, 1984, pgs. 1361-1365.

[49] S.S.Wang, R. Magnusson, J.S. Bagby, M.G. Moharam, *Guided-mode Resonances in Planar Dielectric-Layer Diffraction Gratings*, Journal of the Optical Society of America: A, Vol. 7, Number 8, 1990, pgs. 1470-1474.

[50] D.W. Prather, M.S. Mirotznik, J.N. Mait, *Boundary Integral Methods Applied to the Analysis of Diffractive Optical Elements*, Journal of the Optical Society of America: A, Vol. 14, Number 1, 1997, pgs. 34-43.

[51] A. Meier, T. Arens, S.N. Chandler-Wilde, A. Kirsch, *A Nystroem Method for a Class of Integral Equations on the Real Line with Applications to Scattering by Diffraction Gratings and Rough Surfaces*, Journal of Integral Equations and Applications, Vol. 12, Number 3, 2000, pgs. 281-321.

[52] M.G. Moharam, T.K. Gaylord, *Rigorous Coupled-Wave Analysis of Planar-Grating Diffraction*, Journal of the Optical Society of America, Vol. 71, Issue 7, 1981, pgs. 811-818.

[53] M.G. Moharam, E.B. Grann, D.A. Pommet, T.K. Gaylord, *Formulation for Stable and Efficient Implementation of the Rigorous Coupled-Wave Analysis of Binary Gratings*, Journal of the Optical Society of America: A, Vol. 12, Number 5, 1995, pgs. 1068-1076.

[54] S.M. Sze, K.K. Ng, *Physics of Semiconductor Devices*, Third Edition, Wiley-Interscience, Hoboken, 2007.

[55] A. Sommerfeld, H. Bethe, *Handbuch der Physik*, 2nd ed. Springer-Verlag, Berlin, Vol. 24, 1933.

[56] R.A. Ferrell, *Predicted Radiation of Plasma Oscillations in Metal Films*, Physical Review, Vol. 111, Issue 5, 1958, pgs. 1214-1222.

[57] A.B. Fowler, F.F. Fang, W.E. Howard, P.J. Stiles, *Magneto-Oscillatory Conductance in Silicon Surfaces*, Physics Review Letters, Vol. 16, Issue 20, 1966, pgs. 901-903.

[58] T. Mimura, *The Early History of the High Electron Mobility Transistor (HEMT)*, IEEE Transactions on Microwave Theory and Technology, Vol. 50, Number 3, 2002, pgs. 780-782.

[59] L. Esaki, R. Tsu, *Superlattice and Negative Differential Conductivity in Semiconductors*, IBM Journal of Research Development, RC 2418, 1969.

[60] R. Dingle, H.L. Stormer, A.C. Gossard, W. Wiegmann, *Electron Mobilities in Modulation-Doped Semiconductor Heterojunction Superlattices*, Applied Physics Letters, Vol. 33, Number 7, 1978, pgs. 665-667.

[61] H.L. Stormer, R. Dingle, A.C. Gossard, W. Wiegmann, M.D. Sturge, *Two-Dimensional Electron Gas at Differentially Doped GaAs-$Al_xGa_{1-x}As$ Heterojunction Interface*, Journal of Vacuum Science and Technology, Vol. 16, Issue 5, 1979, pgs. 1517-1519.

[62] T. Mimura, S. Hiyamizu, T. Fujii, K. Nanbu, *A New Field-Effect Transistor with Selectively Doped GaAs/n-$Al_xGa_{1-x}As$ Heterojunctions*, Japanese Journal of Applied Physics, Vol. 19, 1980, pg. L225-L227.

[63] D. Delagebeaudeuf, P. Delescluse, P, Etienne, M. Laviron, J. Chaplart, N.T. Linh, *Two-Dimensional Electron Gas M.E.S.F.E.T. Structure*, Electronics Letters, Vol. 16, Issue 17, 1980, pgs. 667-668.

[64] H. Daembkes, ed. *Modulation-Doped Field-Effect Transistors: Principle, Design, and Technology*, IEEE Press, Piscataway, NJ, 1991.

[65] H. Morkoc, H. Unlu, G. Ji, *Principles and Technology of MODFETs: Principles, Design, and Technology*, Wiley, New York, 1991.

[66] U.K. Mishra, P. Parikh, Y.F. Wu, *AlGaN/GaN HEMTs–An Overview of Device Operation and Applications*, Proceedings of IEEE, Vol. 90, Number 6, 2002, pgs. 1022-1031.

[67] P.H. Ladbrooke, *GaAs MESFETs and High Electron Mobility Transistors (HEMT)*, in `Gallium-Arsenide for Devices and Integrated Circuits`, Peregrinus, London, 1986.

[68] M. Razeghi, *Fundamentals of Solid State Engineering*, Second Edition, Springer, New York, 2006.

[69] N. Dyakonova, A.E. Fatimy, J. Lusakowski, W. Knap, M.I. Dyakonov, M.A. Poisson, E. Morvan, *Room-Temperature Terahertz Emission from Nanometer Field-Effect Transistors*, Applied Physics Letters, Vol. 88, 2006, pgs. 141906-1-3.

[70] I. Mehdi, E. Schlecht, G. Chattopadhyay, P.H. Siegel, *THz Local Oscillator Sources: Performance and Capabilities*, Proceedings of SPIE, Vol. 4855, Number 435, 2003, pgs. 435-446.

[71] $http://silvaco.com/products/device_simulation/atlas.html.$

[72] S. Chakraborty, I.J. Maasilta, S.H. Tessmer, M.R. Melloch, *Imaging a Two-Dimensional Electron System with a Scanning Charged Probe*, Physical Review B, Vol. 69, 2004, 073308-1-3.

[73] http://www.ioffe.ru/SVA/NSM/Semicond/GaAs/basic.html.

[74] D. Brunner, A. Angerer, E. Bustarret, F. Freudenberg, R. Hoepler, R. Dimitrov, O. Ambacher, M. Stutzmann, *Optical Constants of Epitaxial AlGaN Films and their Temperature Dependence*, Journal of Applied Physics, Vol. 82, Number 10, 1997, pgs. 5090-5096.

[75] Y. Zhang, J. Singh, *Charge Control and Mobility Studies for an AlGaN/GaN High Electron Mobility Transistor*, Journal of Applied Physics, Vol. 81, Number 1, 1999, pgs. 587-594.

[76] G.L. Harris, ed., *Properties of Silicon Carbide*, IET, Stevenage, 1995.

[77] M.A. Ordal, L.L. Long, R.J. Bell, S.E. Bell, R.R. Bell, R.W. Alexander, C.A. Ward, *Optical Properties of the Metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the Infrared and Far Infrared*, Applied Optics, Vol. 22, Number 7, 1983, pgs. 1099-1120.

[78] A. Satou, V. Ryzhii, I. Khmyrova, M. Ryzhii, M.S. Shur, *Characteristics of a terahertz photomixer based on a high-electron mobility transistor structure with optical input through the ungated regions* Journal of Applied Physics, Vol. 95, Number 4, 2004, pgs. 2084-2089.