Université d'Ottawa · University of Ottawa

Université d'Ottawa • University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Ian Gregory BREUKELAAR

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M. A. Sc. (Electrical Engineering)

GRADE - DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

Surface Plasmon-Polaritons in Thin Metal Strips and Slabs : Waveguiding and Mode Cutoff

P. Berini

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

R. Gauthier

J. Yao

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Surface Plasmon-Polaritons in Thin Metal Strips and Slabs: Waveguiding and Mode Cutoff

by

Ian G. Breukelaar, B. Sc., B. A. Sc.

A thesis submitted to the
School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Applied Science
In Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

September 15, 2004

Your file Votre référence
ISBN: 0-494-01426-1
Our file Notre référence
ISBN: 0-494-01426-1

Abstract

Using a fully vectorial formulation, the normal mode analysis method is used to model for what is believed to be the first time, the radiative spreading of surface plasmon-polariton modes input into regions where the bound surface mode is cutoff, or radiative. The analysis was performed at $\lambda = 1.55\mu m$ for embedded infinite-width slabs and finite-width strip metal waveguides where the metal was Au and the surrounding dielectrics were SiO_2 and $LiNbO_3$. Mode cutoff is induced by an asymmetry between the index of refraction of the top clad and the bottom clad. Comparisons between predictions and experimental results for cutoff finite-width waveguides show very good agreement. The analysis was performed for a range of waveguide dimensions and results show that a dynamic range of better than 20dB is possible for an appropriate waveguide and dielectric asymmetry. Practical applications are outlined to make use of this cutoff effect.

Acknowledgments

I would first like to thank my supervisor Professor Pierre Berini for allowing me to investigate this very interesting topic and for giving me a large amount of freedom to pursue various curiosities along the way. His encouragement and guidance were invaluable.

I am also grateful to my former colleague, Dr. Philip Wort for the many enlightening discussions and for suggesting the normal mode analysis approach as a method for analyzing mode cutoff.

I am indebted to Robert Charbonneau for performing measurements to compare with the theoretical prediction from this work. Along this line I would like to thank all of my colleagues at Spectalis for their input into this. I also thank Spectalis itself for financially assisting a portion of this work and for providing technical resources as needed.

The final version of this work has benefited from the close scrutiny and helpful suggestions of the two official examiners, Professor Robert Gauthier from Carleton University and Professor Jianping Yao from the University of Ottawa.

I would also like to thank Robert Charbonneau and Dr. Philip Wort for proof reading this document and providing valuable feedback. Any remaining errors were well disguised or added by me afterwards.

Finally, I would like to thank my beautiful wife for her continual support and limitless patience. Muchas gracias.

Contents

Chapter 1

Introduction

Simply put, surface plasmon-polaritons (SPPs) are electromagnetic waves propagating at the surface of a metal and a dielectric, where the metal has a negative dielectric constant. This class of modes has been studied since the turn of the century in the context of radio waves travelling along the surface of the earth.[1] In the 1940's Fano[2] studied these modes at optical frequencies to explain the Wood's anomalies in metal diffraction gratings. Later, Ritchie,[3] Stern and Ferrell[4] and others contributed theoretically to the understanding of these surface modes. SPPs were first observed in electron energy loss experiments[5] and were later observed optically in attenuated total reflection (ATR) experiments devised by Otto[6] and Kretschmann.[7] These methods are still used today in surface characterization and measuring the optical constants of metals.[8]

Once the surface modes at one interface were understood, research moved on to thin metal slab waveguides[9–13] and it was found that if the slab was made thin enough that a special class of coupled SPPs emerged with longer propagation distances. This led to a number of device applications such as modulators,[14–16] sensors,[17] and polarizing devices.[18] Plasmons have been associated with a number of other recent ideas, from optical transistors[19] to plasmon lasers[20,21] and light emitters.[22] The broad term *plasmonics* is

now used popularly to describe plasmon-based technologies.

In recent years, as fabrication techniques continue to improve, this field has seen a number of exciting and active branches develop. These include photonic band gap waveguides,[23] plasmons in nanoparticle chains,[24] channel plasmons,[25] and nano-wires.[26] Recently, finite width SPP waveguides were predicted[27,28] and experimentally realized.[29] These are analogous to the low loss slab waveguides, but have mode confinement in two dimensions rather than one and are a candidate for a new integrated optics technology. The slab and finite width SPP waveguides will be the primary subject of this work.

The bound modes of these waveguides can be cut off if there is a sufficiently asymmetric index of refraction between the top and bottom clads surrounding the metal.[12,30,31] SPP optical modulators in thin slabs[14–16] are often based on, or related to this concept of mode cutoff. In these modulators the SPP waveguide is not the fundamental guiding element in the devices, but is used to couple the light out of the main dielectric waveguide. The high loss in the SPP waveguides or the coupling to SPPs extinguishes the signal in the off-state of the modulator. These surface plasmon resonance or attenuation based modulators often require bulk items like prisms to couple the light into the metal and are not necessarily suited to planar lightwave circuits. In this work the SPP waveguide is the main guiding element. Light is coupled into and out of a region where an asymmetry is switched off and on in the cladding index of refraction so that mode cutoff is induced and the mode radiates away from the core. The input and output source can be other planar SPP waveguides in parallel, or edge coupled fibers. If the asymmetric region is long enough a large dynamic range can be found between the off and on states. Mode cutoff is analyzed for a range of slab and finite width waveguides using the normal mode analysis method. Radiation spreading loss is predicted in order to find lengths necessary for sufficient dynamic range. To our knowledge an investigation of SPP mode spreading in cutoff waveguides has not been performed in slabs or in strips.

This document is set up as follows: chapter 2 outlines the requisite theory and background for SPP waveguides. New results are given for silica and lithium niobate clad Au waveguides. The current basic knowledge of SPPs waveguides is put into a consistent formalism and most of it is generalized for diagonally anisotropic, lossy media.

In chapter 3, the mode cutoff effect is introduced and analyzed, and new cutoff results are presented. The normal mode analysis method is then introduced and validated and radiation loss curves are given for infinite width (slab) waveguides in silica and lithium niobate.

Chapter 4 is similar to chapter 3 and here we look at finite width strip waveguides, as opposed to the infinite width slabs. We also compare the theoretical predictions with some recent experimental results and discuss some practical device applications.

Chapter 5 has some concluding remarks and suggestions for future work.

Chapter 2

Plasmon Waveguide Theory

2.1 Introduction

We now look at the necessary theory behind plasmon based waveguides. A plasma is an neutrally charged group of positively and negatively charged particles (approximately the same number of each). A plasma oscillation[32] is when one group of charges moves relative to the other, such as in a metal with electrons moving with respect to the positive ions. The quantum of the plasma oscillation is a quasi-particle referred to as a plasmon. It is a boson,[33] so it follows Bose-Einstein statistics and has integer spin, although we will not look at it in that much detail in waveguide analysis.

At optical frequencies, the Drude model of the permittivity of the plasma (or metal) predicts a negative permittivity. This is not quite the same concept as negative refraction, which requires both negative permittivity and negative permeability[34] to create an effectively negative index of refraction. At certain frequencies light may freely propagate through the metal while at others it will be sharply attenuated and surface modes may exist at the interface with a dielectric.

The field of plasmons has matured to the point where an introduction to the optical

properties of metals and sometimes even surface plasmons is included in many standard texts in solid-state physics,[35–39] optics[40–44] and electrodynamics.[45–47] Some of the material in this chapter is thus found scattered in one place or another in varying notations and systems of units, and not specifically focused on waveguiding. A number of works[48–51] are also devoted to SPPs themselves, but these works are now out of print and of limited availability. The goal here was to bring together the background necessary for proper introduction to SPP waveguides, and place it into an updated and consistent notation generalized for lossy, anisotropic media.

2.2 Relative Permittivity of Metals

As mentioned earlier an SPP propagates at the interface of two materials of opposite sign of permittivity. In this section we look at the origin of the relative permittivity in metals and show how it may change signs depending on the frequency. We consider the metal to be a free electron gas where conduction electrons are free to move about in the metal with no restoring force from a host atom. Assume now that these electrons are subject to a time harmonic electric field (optical wave) of the form $\mathbf{E} = E_0 e^{-i\omega t}\hat{y}$ where ω is the angular frequency of oscillation. If treated classically[52] as a free particle, the equation of motion of the electron is

$$\frac{d^2y}{dt^2} + \frac{1}{\tau}\frac{dy}{dt} = -\frac{e}{m}E_0 e^{-i\omega t}, \tag{2.1}$$

where τ is the average time between collisions for the electron, and its reciprocal $\omega_c = \tau^{-1}$ is referred to as the collision frequency.[35] m is the mass of the electron and e is the elementary charge. The following derivation follows Kittel[36] and adds damping courtesy of the second term in (2.1). The end result can also be found from the Lorentz model for

the optical properties of lossy dielectrics by setting the resonant frequency ω_0 to zero.[37]

Equation 2.1 has the solution $y(t) = y_0 e^{-i\omega t}$, and upon substitution we find the position of the electron at a time t to be

$$\mathbf{y}(t) = \frac{e\mathbf{E}}{m(\omega^2 + i\omega/\tau)}. \tag{2.2}$$

The polarization $\mathbf{P}$ is defined as the induced dipole moment ($\mathbf{p} = -e\mathbf{y}$ for an electron at position $\mathbf{y}$) per unit volume so if N is the number of free electrons per unit volume the polarization is

$$\begin{aligned} \mathbf{P} &= -Ne\mathbf{y} \\ &= -\frac{e^2 N}{m(\omega^2 + i\omega/\tau)}\mathbf{E} \end{aligned} \tag{2.3}$$

We also define the polarization for linear dielectrics in terms of the permittivity of free space ϵ_0, and the electrical susceptibility χ_e as

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} = \epsilon_0 \left(\epsilon_m - 1\right) \mathbf{E}. \tag{2.4}$$

ϵ_m the relative permittivity, or dielectric constant of the metal, is found by equating (2.3) and (2.4)

$$\epsilon_m = 1 - \frac{\omega_p^2}{\omega(\omega + i/\tau)} = 1 - \frac{\omega_p^2}{\omega^2 + \tau^{-2}} + i\frac{\omega_p^2 \tau}{\omega\left(1 + \omega^2\tau^2\right)}, \tag{2.5}$$

where $\omega_p = \sqrt{\frac{Ne^2}{m\epsilon_0}}$ is the plasma frequency. This is commonly called the Drude model for the dielectric constant of a metal since it is based on the Drude theory of electrical conductivity.[35]

Often it is useful to work backwards from (2.5) and find the plasma frequency and

relaxation time in terms of the dielectric constant $\epsilon_m = \epsilon_m^R + i\epsilon_m^I$, resulting in

$$\tau = \frac{1 - \epsilon_m^R}{\omega \epsilon_m^I}, \; \omega_p = \sqrt{1 - \epsilon_m^R}\sqrt{\omega^2 + \tau^{-2}}. \tag{2.6}$$

Note that a heavily doped semiconductor will optically have the same behavior as a metal and a similar derivation can be done for its dielectric constant. It turns out to be similar to that of a metal with the only alteration being the replacement of the '1' in the first real term in (2.5) by the relative permittivity of the undoped semiconductor.[53]

In the following sections we will discuss the consequences of a relative permittivity of this form.

2.3 Plane Waves in a Metal

2.3.1 Dispersion Relation

We first look at plane waves propagating freely in a metal, away from any boundary or surface. This can be more simply understood as an optical wave propagating through a material with relative permittivity ϵ_m.

We assume a plane wave travelling in the z-direction with wavenumber (or propagation constant) β and frequency ω with an electric field of the form

$$\mathbf{E} = E_0 e^{i(\beta z - \omega t)} \hat{y}, \tag{2.7}$$

where

$$\beta = \omega\sqrt{\epsilon\mu}. \tag{2.8}$$

μ is the permeability which from now on is assumed to be the permeability of free space $\mu = \mu_0$ (nonmagnetic media) unless otherwise noted and $\epsilon = \epsilon_m \epsilon_0$. If the relaxation

time $\tau \rightarrow \infty$ (the lossless approximation) then the relative permittivity of the metal simplifies to

$$\epsilon_m = 1 - \frac{\omega_p^2}{\omega^2}. \tag{2.9}$$

β simplifies to

$$\beta = \frac{\omega}{c}\sqrt{1 - \frac{\omega_p^2}{\omega^2}}, \tag{2.10}$$

where c is the speed of light in vacuum.

For $\omega > \omega_p$, ϵ_m is real and positive and β is real and the wave propagates along z. However, if $\omega < \omega_p$, ϵ_m is real and negative, β is imaginary and the fields attenuate as $e^{-\beta z}$ along z. The wave is evanescent along z and therefore does not propagate.

Rearranging (2.10) we find the dispersion relation

$$\omega = \sqrt{c^2\beta^2 + \omega_p^2}, \tag{2.11}$$

from which we see that for all real β, $\omega > \omega_p$. The consequences of (2.11) are discussed in the next section.

2.3.2 Phase and Group Velocity

From the dispersion relation (2.11) we can easily find the phase and group velocity remembering that $\omega > \omega_p$ for propagation. For the plane wave described in (2.7) the phase velocity is

$$v_p = \frac{\omega}{\beta} = c\sqrt{1 + \frac{\omega_p^2}{c^2\beta^2}} = c\left(1 - \frac{\omega_p^2}{\omega^2}\right)^{-1/2}, \tag{2.12}$$

and is always greater than the speed of light in vacuum c. The group velocity

$$\begin{aligned} v_g &= \frac{\partial \omega}{\partial \beta} \\ &= c\sqrt{1 - \frac{\omega_p^2}{\omega^2}}, \end{aligned} \tag{2.13}$$

and is always less than c.[46]

In the limit that $\beta \to \infty$ the frequency approaches the light line $\omega = c\beta$ and the wave becomes photonlike as it propagates identically to a photon in free space, i.e. the phase and group velocity converge to c. As $\beta \to 0$ the group velocity falls to zero and the wave reaches a resonance at the plasma frequency ω_p, the natural frequency of oscillation of the free electrons in the material. Between these extremes there is coupling between the two excitations (the *plasma* oscillation and the *photon*) and we refer to the excitation as a *plasmon*, although commonly it is referred to as a plasmon for all β. This curve is plotted later in comparison with the dispersion curves for SPPs (see Figure 2.3).

2.4 Surface Plasmon-Polariton at One Interface

2.4.1 Surface Polaritons vs. Surface Plasmon-Polaritons

The situation of surface plasmon-polaritons is much different from that of plane waves in metals. A surface polariton is a surface electromagnetic wave at the interface between a material with a negative dielectric constant (surface active medium), and a material with a positive dielectric constant (passive medium) which induces a polarization in the active medium and couples to it. When the polarized medium is a metal the surface polariton is termed a surface plasmon-polariton, because of the plasma contribution to the negative dielectric constant. Other types of surface polaritons exist such as exciton-polaritons[54]

and phonon-polaritons.[55]

In the development of this field there have been several naming conventions where 'surface plasmon', 'surface plasma wave', 'surface electromagnetic wave', 'surface polariton', and 'surface plasmon-polariton' all referred to the same excitation. Today the situation has improved and most authors now refer to these waves by the less ambiguous term *surface plasmon-polaritons*, for the case of waveguides.

2.4.2 Dispersion Relation for TM Surface Plasmon-Polaritons

We begin by solving for the dispersion relation of a transverse magnetic (TM, sometimes called p-polarized) surface polariton (ie. independent of the cause of the negative permittivity) at one interface, and then apply it to the case of SPPs. For a TM mode only the fields H_x, E_y and E_z exist. Modes are solutions to Maxwell's equations which maintain their shape when propagating through a geometry constant in the direction of propagation.

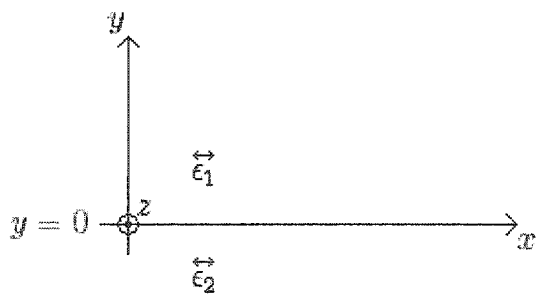

Figure 2.1: Single interface structure for analysis of TM SPP. Interface at y=0.

We assume the coordinates as shown in Figure 2.1 with the z-axis out of the page and two infinite half spaces joined at $y = 0$ with $\overleftrightarrow{\epsilon_1}$ for all $y > 0$ and $\overleftrightarrow{\epsilon_2}$ for all $y < 0$. We solve for bound modes, that is square integrable or normalizable modes whose fields go to zero at $y = \pm\infty$. Modes typically have discrete β values for a given geometry, in

contrast to plane waves which take on a continuum of values.

The TM mode is solved in terms of the H-field and we assume z-propagating fields of the form

$$\begin{aligned} \mathbf{H_1} &= H_0 e^{i(\beta z - \omega t) - k_1 y}\hat{x} \qquad y > 0, \\ \mathbf{H_2} &= H_0 e^{i(\beta z - \omega t) + k_2 y}\hat{x} \qquad y < 0, \end{aligned} \tag{2.14}$$

where β is the wavenumber or propagation constant along z. k_1 and k_2 are the transverse wavenumbers in media 1 and 2 respectively and the real parts of k_1 and k_2 are greater than zero to maintain evanescent fields into regions 1 and 2.

The fields must satisfy Maxwell's equations for source-free, time-harmonic fields in anisotropic media[45]

$$\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}, \tag{2.15}$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0 \overleftrightarrow{\epsilon_j}\, \mathbf{E}, \tag{2.16}$$

where $\overleftrightarrow{\epsilon_j}$ is the relative permittivity tensor. We assume the materials are oriented so that $\overleftrightarrow{\epsilon_j}$ will be diagonal with the form

$$\overleftrightarrow{\epsilon_j} = \begin{pmatrix} \epsilon_{xx_j} & 0 & 0 \\ 0 & \epsilon_{yy_j} & 0 \\ 0 & 0 & \epsilon_{zz_j} \end{pmatrix}, \tag{2.17}$$

where the elements are in general complex to account for material losses and the subscript j denotes the layer.

Substituting (2.14) into (2.16) gives the electric field components in terms of the H-field. The $e^{-i\omega t}$ time dependance is understood.

$$\begin{aligned} \mathbf{E}_{\mathbf{y_1}} &= -\frac{H_0\beta}{\omega\epsilon_0\epsilon_{yy_1}} e^{i\beta z - k_1 y}\hat{y} \\ \mathbf{E}_{\mathbf{z_1}} &= \frac{iH_0 k_1}{\omega\epsilon_0\epsilon_{zz_1}} e^{i\beta z - k_1 y}\hat{z} \end{aligned} \qquad y > 0, \tag{2.18}$$

$$\begin{aligned} \mathbf{E}_{\mathbf{y_2}} &= -\frac{H_0\beta}{\omega\epsilon_0\epsilon_{yy_2}} e^{i\beta z + k_2 y}\hat{y} \\ \mathbf{E}_{\mathbf{z_2}} &= -\frac{iH_0 k_2}{\omega\epsilon_0\epsilon_{zz_2}} e^{i\beta z + k_2 y}\hat{z} \end{aligned} \qquad y < 0. \tag{2.19}$$

H_0 is a normalization constant. The transverse wavenumbers are found by substituting the fields $\mathbf{E}$ and $\mathbf{H}$ into (2.15). We find

$$k_j = \sqrt{\epsilon_{zz_j}\left(\frac{\beta^2}{\epsilon_{yy_j}} - k_0^2\right)} \qquad j = 1, 2. \tag{2.20}$$

$k_0 = \frac{2\pi}{\lambda_0}$ is the free space wavenumber.

The boundary conditions at $y = 0$ require continuity of the tangential fields. For the case $\mathbf{E}_{\mathbf{z_1}}(0^+) = \mathbf{E}_{\mathbf{z_2}}(0^-)$, where $y = 0^+$ and $y = 0^-$ are the points just above and just below the interface, respectively, we find

$$\frac{k_1}{\epsilon_{zz_1}} + \frac{k_2}{\epsilon_{zz_2}} = 0, \tag{2.21}$$

and since $k_1 > 0$ and $k_2 > 0$ this shows that the surface polariton can only exist at the interface between media with permittivity of opposite sign.

Now substituting for k_1 and k_2 in (2.21) and solving for β leads to a new dispersion relation for single interface surface polaritons.

$$\beta = k_0\sqrt{\frac{\left(\epsilon_{zz_2} - \epsilon_{zz_1}\right)\epsilon_{yy_1}\epsilon_{yy_2}}{\epsilon_{zz_2}\epsilon_{yy_2} - \epsilon_{zz_1}\epsilon_{yy_1}}}. \tag{2.22}$$

To our knowledge this is the first time the dispersion relation for the surface polariton at the interface between two anisotropic media has been reported. This may be use-

ful for anisotropic metals and other composite nano-materials behaving like anisotropic metals.[56] The consequences of this new form may be quite broad and heavily specific to individual materials. We will look at results for some anisotropic dielectrics later. If one of the media is isotropic this reduces to the result reported by Stegeman[57]

$$\beta = k_0 \sqrt{\frac{\epsilon_{yy_1}\epsilon_2\left(\epsilon_2 - \epsilon_{zz_1}\right)}{\epsilon_2^2 - \epsilon_{zz_1}\epsilon_{yy_1}}}. \tag{2.23}$$

Now allowing both media to be isotropic leads to the more common result[58]

$$\beta = k_0 \sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_1 + \epsilon_2}}. \tag{2.24}$$

Inserting the permittivity function for a metal into (2.22), (2.23), or (2.24) results in the dispersion relation for surface *plasmon*-polaritons.

These dispersion relations place more conditions on ϵ_1 and ϵ_2 for lossless media, namely that in addition to the requirement that $\epsilon_1\epsilon_2 < 0$ it is also required that $\epsilon_1 + \epsilon_2 < 0$ so that $\mathbf{R}\{\beta\} \neq 0$ and the mode can propagate along z. This result holds approximately for the lower loss modes we are most interested in. Workers have classified four different surface plasmon type modes,[51] three of which only exist because of the damping in the metal, which allows (2.24) to pick up a real part so that a mode can still propagate. By that classification, the modes we will be looking at are termed 'Fano' modes although Fano[2] himself referred to them as 'Sommerfeld[1] waves'. The existence of the Fano modes is not dependant on having loss in the metal.

The range of possible β values can be found from (2.24) assuming $\epsilon_1 > 0$ and setting $\epsilon_2 = -(\epsilon_1 + \delta)$ with $\delta > 0$ since $\epsilon_1 + \epsilon_2 < 0$. We have

$$\beta = k_0 \sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_1 + \epsilon_2}}$$

$$= k_0\sqrt{\frac{\epsilon_1^2}{\delta}+\epsilon_1}. \tag{2.25}$$

In the limit of $\delta \to 0^+$, $\beta \to \infty$ and in the limit of $\delta \to \infty$, $\beta \to \sqrt{\epsilon_1}k_0$. Therefore, for bound surface polaritons $\beta > \sqrt{\epsilon_1}k_0$.

Figure 2.2 shows the propagation constant $\beta = \beta^R + i\beta^I$ for a range of values of ϵ_1 above a gold[59] layer with $\epsilon_2 = -131.95 + i12.65$ for the wavelength $\lambda_0 = 1.55\mu m$. ϵ_1 is given in terms of its associated index of refraction n_1 where the index of refraction is given by

$$n_1 = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}}, \tag{2.26}$$

where $\epsilon = \epsilon_0\epsilon_1$. For nonmagnetic materials this reduces to $n_1 = \sqrt{\epsilon_1}$. We tend to work with the relative permittivity for metals since for lossless metals a real negative relative permittivity seems conceptually less abstract than an imaginary index of refraction. It is often convenient to switch between the relative permittivity and the index of refraction, and subscripted variables ϵ_j and n_j should generally be taken to be related by (2.26) throughout this work, if not explicitly stated.

The mode power attenuation (MPA) shown in curve (i) is the power loss per unit length and is found from the imaginary part of the propagation constant β^I.

$$MPA = 0.02\log_{10}(e)\beta^I \quad (\text{dB/mm}), \tag{2.27}$$

where $\beta^I \geq 0$ for attenuating modes by our convention and has units of m^{-1}.

The phase part of the propagation constant shown in curve (ii) is given in the usual normalized form referred to as the *effective index of refraction* β^R/k_0 or n_{eff}. Curve (iii) is the curve $\beta^R/k_0 = n_1$ and is plotted as a reference. β^R/k_0 can usually be compared to the cladding index of refraction to determine a relative confinement where higher

relative β^R/k_0 typically means higher confinement to the waveguide core, or in this case the interface. Hence for increasing n_1, the confinement increases and as curve (i) shows, the MPA also increases. MPA is more than 10dB/mm for all n_1, so single interface SPPs are very lossy, and have limited range.

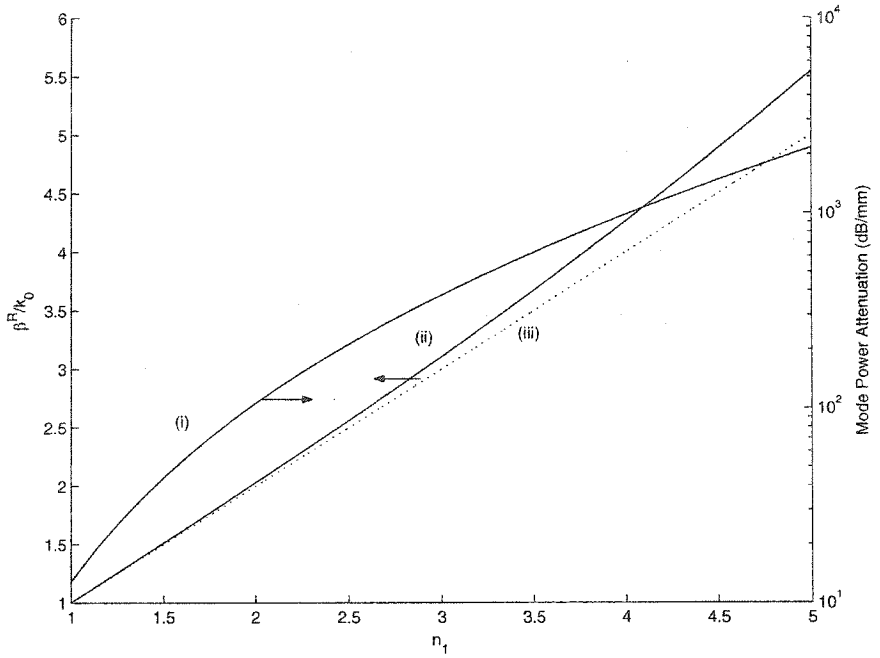

Figure 2.2: Propagation constant of a Au single interface SPP vs. index of refraction of dielectric. (i) MPA, (ii) effective index of mode, (iii) $\beta^R/k_0 = n_1$.

2.4.3 Phase and Group Velocity

As mentioned, the dispersion equation derived above holds for all surface polaritons. To apply it to SPPs we insert the relative permittivity value for the metal.

Substituting the lossless version of ϵ_m (2.9) into (2.24) we solve for the dispersion

equation in a form with ω a function of β,

$$\omega = \left[\frac{1}{2}\left(\omega_p^2 + \beta^2 c^2(1 + 1/\epsilon_1) - \sqrt{\omega_p^4 + 2\omega_p^2\beta^2c^2(1 - 1/\epsilon_1) + \beta^4c^4(1 + 2/\epsilon_1 + 1/\epsilon_1^2)}\right)\right]^{1/2}. \tag{2.28}$$

Note that ϵ_1 is also generally a function of ω for dispersive dielectrics. For a metal-vacuum interface $\epsilon_1 = 1$ for all frequencies and (2.28) reduces to

$$\omega = \left[\frac{1}{2}\left(2\beta^2c^2 + \omega_p^2 - \sqrt{4\beta^4c^4 + \omega_p^4}\right)\right]^{1/2}. \tag{2.29}$$

Using (2.9), (2.24), and the constraint $\epsilon_1 + \epsilon_2 < 0$, we find an upper bound on the range of possible frequencies for lossless, single-interface SPPs

$$\omega < \frac{\omega_p}{\sqrt{\epsilon_1 + 1}}. \tag{2.30}$$

This upper limit is the so called surface plasmon frequency, ω_{sp}.

Note that for doped semiconductors the upper limit is

$$\omega < \frac{\omega_p}{\sqrt{\epsilon_1 + \epsilon_{sc}}}, \tag{2.31}$$

where ϵ_{sc} is the relative permittivity of the undoped semiconductor.

Equations (2.28) and (2.29) are better understood after plotting them versus β. Figure 2.3 shows generic dispersion curves for plane waves and surface plasmons. Curve (i) shows the dispersion curve for the plane wave in a metal (as discussed earlier) and we see that as β increases the curve converges to that of the dispersion of light in free space (the 'light line' $\omega = c\beta/\sqrt{\epsilon_1}$), curve (iii). For decreasing β the result converges to the plasma frequency ω_p as discussed earlier. Curve (ii) is (2.29) or the dispersion equation of an

SPP at a metal-vacuum interface. At low β it converges to the light line in free space while for high β it converges to $\omega_p/\sqrt{2}$ as expected from (2.30). Curves (iv) and (v) are the dispersion curves for the SPP at an interface between a metal and a dielectric with $\epsilon_1 > 1$ and the light line in that dielectric, respectively. Notice that this curve always lies below that of the metal-vacuum SPP case. We have left ϵ_1 independent of ω here for this example, but since $\epsilon > 1$ for all dielectrics the curve will remain below the free space curve for any value.

We see from Figure 2.3 that the phase velocity, $v_p = \omega/\beta$, and the group velocity $v_g = \partial\omega/\partial\beta$ for SPPs are always less than c. At low β they both converge to the speed of light in the material ϵ_1 and at high β they both converge to zero with the group velocity converging much more quickly.

2.4.4 Excitation

From wave-particle duality we can say that the momentum of a photon with wavenumber $k = \beta$ is $p = \hbar k$, where $\hbar$ is Planck's constant divided by 2π. During any excitation, energy and momentum (or proportionally the wavenumber) must be conserved.[38] Referring again to Figure 2.3, we see that at a given frequency, the wavenumber of the SPP is always to the right of the light line. This means that SPPs are not excitable by light sources out of the plane of propagation or reciprocally that an SPP is a non-radiative or bound mode, as we have shown. SPPs can however be launched by end-fire excitation,[12,29,60] attenuated (ATR) or frustrated (FTR) total reflection,[6] and grating or roughness coupling.[48,49]

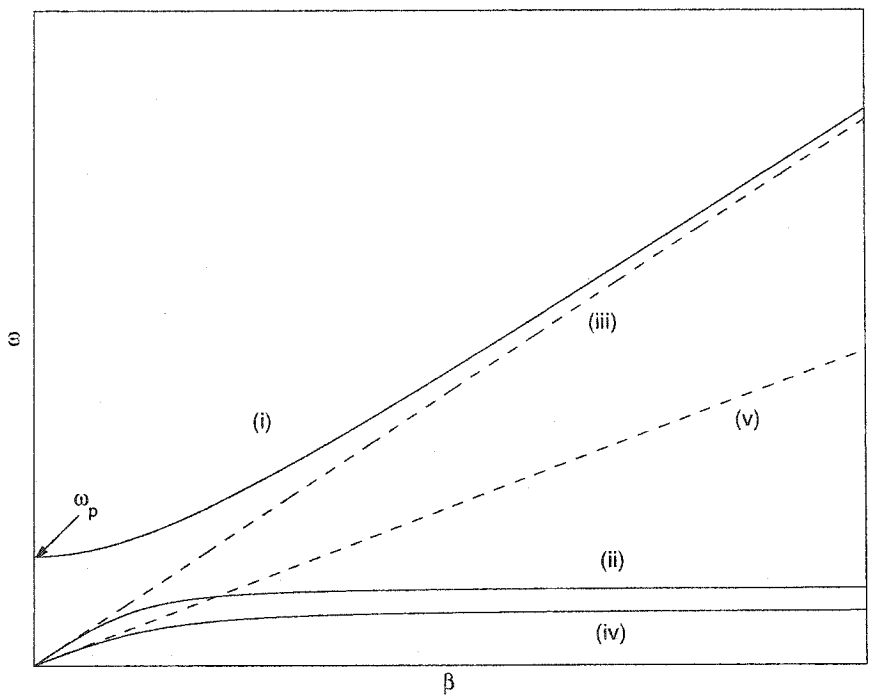

Figure 2.3: Generic dispersion curves for plane waves and surface plasmons. (i) plane wave in a metal, (ii) metal-vacuum SPP, (iii) light line in vacuum, (iv) metal-ϵ_1 SPP, (v) light line in ϵ_1.

2.4.5 Plasmon Band Gap

From Figure 2.3 we see that for SPPs with small β, $\omega \rightarrow c\beta/\sqrt{\epsilon_1}$ while for large β, $\omega \rightarrow \omega_p/\sqrt{\epsilon_1+1}$. Surface plasmon-polaritons are then limited to the frequencies $\omega < \omega_p/\sqrt{\epsilon_1+1}$ while plane waves (or radiating modes) are limited to the frequencies $\omega > \omega_p$. There exists then a so called 'plasmon band gap' in the region $\omega_p/\sqrt{\epsilon_1+1} < \omega < \omega_p$ in which no plasmon related fields can propagate. As ϵ_1 increases the band gap widens as seen by the differences between curves (ii) and (iv). This is an idealization for lossless metals. If loss is included we find that damped excitations are possible within the gap.[61]

2.4.6 Power Flow

Power flow in optical waveguides can be found from the real part of the Poynting vector which gives the time averaged energy flux density.[45]

$$\begin{aligned} \mathbf{S} &= \frac{1}{2}\Re\left\{\mathbf{E}\times\mathbf{H}^*\right\} \\ &= \frac{1}{2}\Re\left\{\hat{x}\left(E_yH_z^* - E_zH_y^*\right) - \hat{y}\left(E_xH_z^* - E_zH_x^*\right)\right. \\ &\quad \left. +\hat{z}\left(E_xH_y^* - E_yH_x^*\right)\right\}, \end{aligned} \tag{2.32}$$

where $\Re\{\}$ represents the real part of the quantity in the brackets and * is the complex conjugate operator acting on the field to its left. For TM modes this reduces to

$$\mathbf{S} = \frac{1}{2}\Re\left\{\hat{y}E_zH_x^* - \hat{z}E_yH_x^*\right\}. \tag{2.33}$$

For the single interface fields, using the fields (2.14), (2.18) and (2.19) gives the power flow on each side of the interface

$$\begin{aligned} \mathbf{S}_{\mathbf{y}_1} &= \tfrac{1}{2}\Re\left\{\frac{iH_0^2 k_1}{\omega\epsilon_0\epsilon_{zz_1}} e^{-2k_1^R y}\right\}\hat{y} \\ \mathbf{S}_{\mathbf{z}_1} &= \tfrac{1}{2}\Re\left\{\frac{H_0^2\beta}{\omega\epsilon_0\epsilon_{yy_1}} e^{-2k_1^R y}\right\}\hat{z} \end{aligned} \qquad y > 0, \tag{2.34}$$

$$\begin{aligned} \mathbf{S}_{\mathbf{y}_2} &= \tfrac{1}{2}\Re\left\{\frac{-iH_0^2 k_2}{\omega\epsilon_0\epsilon_{zz_2}} e^{2k_2^R y}\right\}\hat{y} \\ \mathbf{S}_{\mathbf{z}_2} &= \tfrac{1}{2}\Re\left\{\frac{H_0^2\beta}{\omega\epsilon_0\epsilon_{yy_2}} e^{2k_2^R y}\right\}\hat{z} \end{aligned} \qquad y < 0, \tag{2.35}$$

where k_j^R is the real part of $k_j = k_j^R + ik_j^I$. $\mathbf{S}_{\mathbf{y}_j}$ is zero in lossless media since k_j and ϵ_{zz_j} are both real. Taking the ratio of $\mathbf{S}_{\mathbf{z}_1}(y = 0^+)$ to $\mathbf{S}_{\mathbf{z}_2}(y = 0^-)$ referring to the fields just above and below the interface leads to (for lossless media)

$$\mathbf{S}_{\mathbf{z}_1} = \frac{\epsilon_{yy_2}}{\epsilon_{yy_1}}\mathbf{S}_{\mathbf{z}_2}, \tag{2.36}$$

which shows that the power flows in opposite directions in region 1 and region 2, specifically power flows backwards (negative z) in the metal region because $\epsilon_{yy_1}\epsilon_{yy_2} < 0$. This result holds for all values of y for lossless modes.

For lossy modes this result still holds near the interface, provided the loss is not too high. If the metal is in region 2 then the ratio of $\mathbf{S}_{\mathbf{z}_1}(y = 0^+)$ to $\mathbf{S}_{\mathbf{z}_2}(y = 0^-)$ leads to the requirement:

$$\beta^R\epsilon_2^R + \beta^I\epsilon_2^I < 0, \tag{2.37}$$

which holds well for most cases such as an Au/SiO_2 SPP at $\lambda_0 = 1.55\mu m$ with $\epsilon_2 = -131.95 + i12.65$ and $\epsilon_1 = 2.085$ (values from Palik[59]) for which the left side of (2.37) is -778.

2.4.7 Normalization and Orthogonality

Modes are normalized according to

$$\begin{aligned} \frac{1}{2}\int_A \mathbf{E}\times\mathbf{H}^*\cdot\hat{z}dA &= 1 \\ \frac{1}{2}\int_A \left(E_xH_y^* - E_yH_x^*\right)dA &= 1, \end{aligned} \tag{2.38}$$

so that the mode carries unit power along z. The integration is over the entire mode cross section A. For 1D slab modes the integral is only along y, while for 2D modes the integral is over the xy plane. For TM modes (2.38) reduces to

$$\frac{1}{2}\int_A -E_yH_x^*dA = 1. \tag{2.39}$$

For lossless modes the power is constant along the length of the waveguide, but for lossy modes the power attenuates as $e^{-2\beta^I z}$. The unit power interpretation of the normalization therefore does not hold for lossy modes.

Normalized fields $\{\mathbf{E},\mathbf{H}\}$ are related to the original fields $\{\mathbf{E}',\mathbf{H}'\}$ by $\{\mathbf{E},\mathbf{H}\} = \{\mathbf{E}',\mathbf{H}'\}/\sqrt{N}$, where the normalization constant N is found from[62]

$$N = \frac{1}{2}\left|\int_A \mathbf{E}'\times\mathbf{H}'^*\cdot\hat{z}dA\right|. \tag{2.40}$$

Orthogonality between the normalized modes is then found from

$$\frac{1}{2}\int_A \mathbf{E_m}\times\mathbf{H_n}^*\cdot\hat{z}dA = \frac{1}{2}\int_A \mathbf{E_n}\times\mathbf{H_m}^*\cdot\hat{z}dA = \delta_{mn}, \tag{2.41}$$

The indexes m and n refer to different modes and δ_{mn} is the Kronecker delta,[63] which is 0 if $m \neq n$ and 1 if $m = n$.

For lossy modes the unconjugated versions of (2.40) and (2.41) are used although in

practice to a good approximation the conjugated versions can be used for low-loss modes if the complex part of the propagation constant is taken into account when calculating the fields and the field propagation.[62] The modes we deal with in this report are in the low-loss regime since for Au based waveguides, $\beta^I \ll \beta^R$, even for very thick layers.

2.4.8 Mode Spot Size

The mode spot size is usually defined as the width where the fields outside of the waveguide core diminish to $1/e$ of their maximum value. With a surface mode this is easily found since the fields decay exponentially away from the surface according to the transverse wavenumbers k_1 and k_2. For low loss modes the mode spot size is

$$W_{1/e} = \frac{1}{k_1^R} + \frac{1}{k_2^R}. \tag{2.42}$$

For very high loss modes the imaginary part of $k_{1,2}$ is large and the fields are damped oscillations moving away from the interface. Mode spot size and also the concept of a propagating mode begin to break down in this regime.

The fields typically decay very quickly into the metal region so the mode spot size is dominated by the field in the dielectric. Figure 2.4 shows that the spot size decreases rapidly for the case considered previously with Au ($\epsilon_2 = -131.95 + i12.65$) in contact with an increasing index of refraction n_1.

2.4.9 Example: Gold on Silica SPP

We now give an example of a gold (Au) on silica (SiO_2) single interface SPP. The relative permittivity values are $\epsilon_{Au} = -131.95 + i12.65$ and $\epsilon_{SiO_2} = 2.085$ at $\lambda_0 = 1.55\mu m$, giving a propagation constant $\beta = 5.8999 + i0.004499\ \mu m^{-1}$ so that MPA=39dB/mm.

Figure 2.5 shows the real ($\Re\{\}$) and imaginary ($\Im\{\}$) components of the fields. The

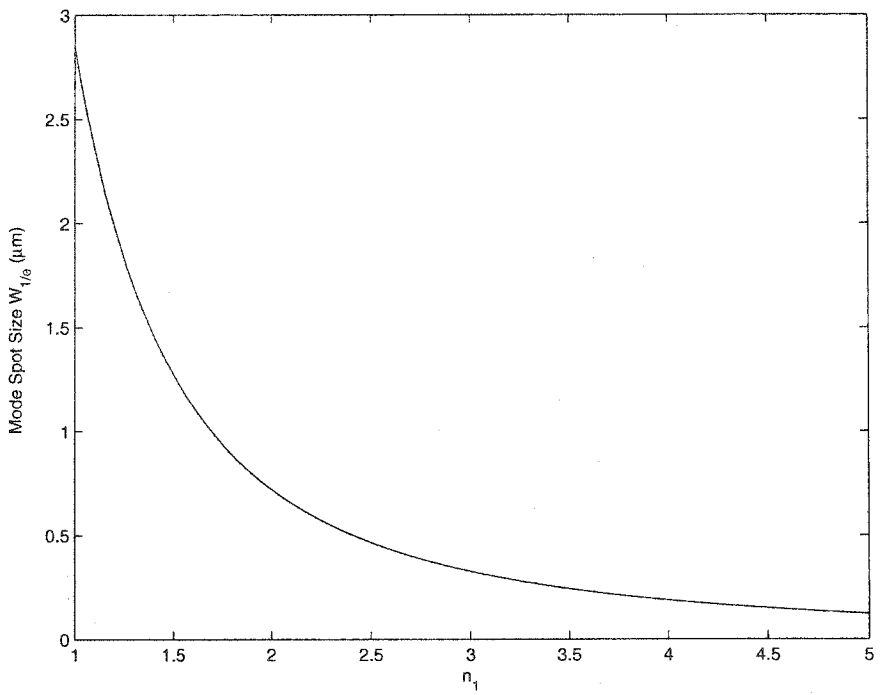

Figure 2.4: Mode spot size of a single interface SPP with Au as the metal, as a function of index of refraction of the dielectric layer.

fields spread much farther into the silica ($y > 0$) than into the metal ($y < 0$). Note also that E_y is much larger than E_z in the silica. Although not visible on this scale the E_z component is larger in the metal.

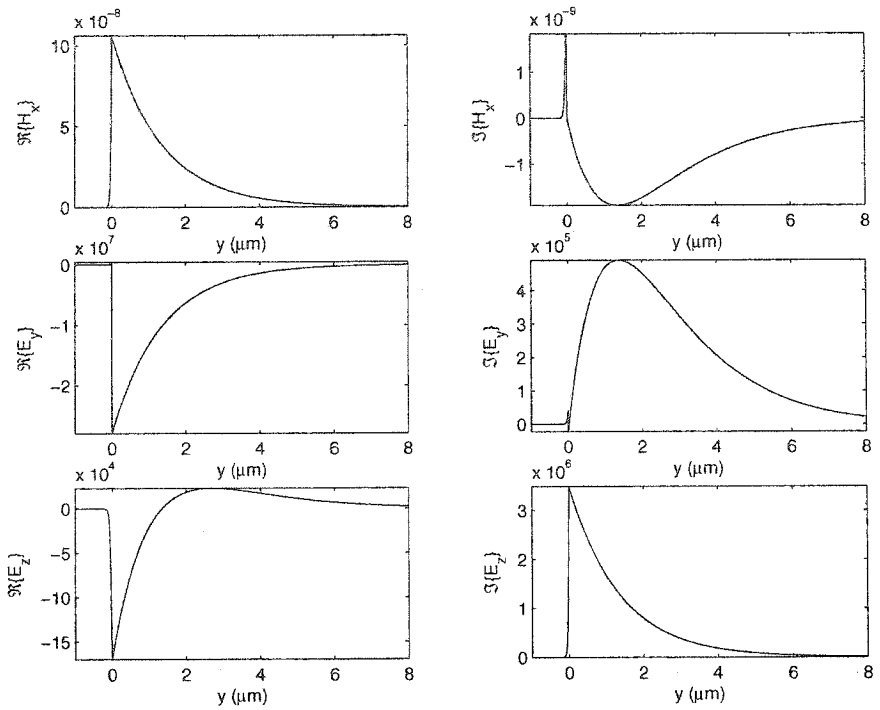

Figure 2.5: Fields for a Au-silica single interface SPP. Silica is in the region $y > 0$. The H-fields are in units of $A/\mu m$ and the E-fields are in $V/\mu m$.

Figure 2.6 shows the power flow for this case where S_y in (a) is more than 2 orders of magnitude smaller than S_z and is negative for all y indicating energy being directed slightly into the metal. For lossy modes the mode shape stays the same, but the amplitude diminishes. Since the metal is the only lossy medium, energy must be directed into it from the lossy clads to maintain the distribution of energy in the mode.[62] For lossless modes $S_y = 0$ for all y. In the z component we see the difference in signs between the Au ($y < 0$), and the silica ($y > 0$) regions, as discussed earlier. The inset (c) shows a

close up of the Au region and the negative sign of S_z.

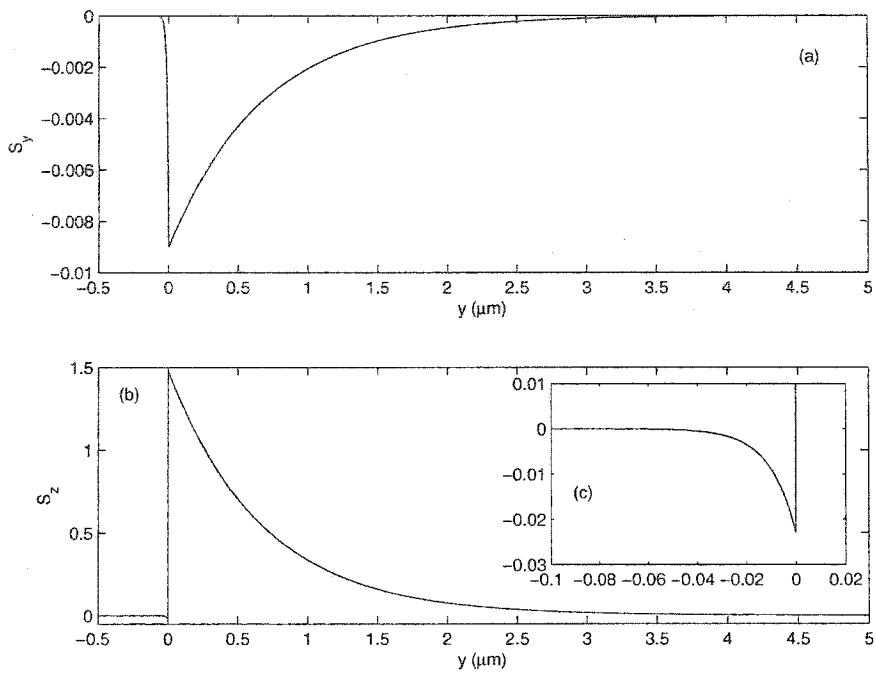

Figure 2.6: Power flow for Au-silica interface. Silica is in the region $y > 0$. S is in units of W/μm^2 and the axes in (c) are identical to (b).

2.4.10 Note on TE Surface Modes

A similar derivation to above can be made for the case of transverse electric (TE, also called s-polarized) surface modes. TE modes consist of only the fields E_x, H_y, and H_z. We begin by assuming electric fields of the form

$$\begin{aligned} \mathbf{E_1} &= E_0 e^{i(\beta z - \omega t) - k_1 y}\hat{x} \qquad y > 0, \\ \mathbf{E_2} &= E_0 e^{i(\beta z - \omega t) + k_2 y}\hat{x} \qquad y < 0, \end{aligned} \tag{2.43}$$

eventually leading to a continuity condition for the fields similar to (2.21)

$$\frac{k_1}{\mu_1} + \frac{k_2}{\mu_2} = 0. \tag{2.44}$$

However, for nonmagnetic media we have $\mu_1 = \mu_2 = \mu_0$ so (2.44) reduces to $k_1 + k_2 = 0$ which cannot be satisfied for bound modes since $k_1 > 0$ and $k_2 > 0$. As well naturally occurring materials have $\mu > 0$, so bound TE surface modes cannot be supported whether the material is magnetic or not. Recent progress in negative refraction and left handed materials[34] has led to composite materials which effectively have $\mu < 0$ and this may open up the possibility of TE surface modes.

2.4.11 Polarization

We have seen that SPPs are exclusively TM polarized for single interfaces and within each mode, Yariv and Yeh have noted that the fields above and below the interface are elliptically polarized and orthogonal to each other.[41,42] This highly polarizing property has led to applications in fiber optic polarizing devices.[18] We will see next that for double interface SPPs, the modes are again TM polarized. Later on when looking at finite width guides, the situation becomes more complicated.

2.5 Surface Plasmon-Polariton Coupled Between Two Interfaces

2.5.1 Introduction

Imagine now an optically infinite (thick) metal having interfaces with a dielectric at $y = \pm\infty$. Let two single interface SPPs then also exist at $y = \pm\infty$. If the thickness of

the metal is decreased, eventually the fields of these two modes begin to couple through the metal, creating supermodes. We can guess that there will be two bound surface modes for this sort of structure, a symmetric supermode corresponding to both modes having fields with the same sign of amplitude (polarity) and an antisymmetric supermode corresponding to the modes having opposite polarity at the interfaces. These modes are denoted as s_b and a_b respectively, with the subscript referring to the bound nature of the modes (leaky modes are also known to exist[12]).

We now have a three layer structure with a metal slab between two semi-infinite dielectrics. Quite often SPPs can be investigated using only three layers and we outline the modal solutions for that case later. First though we look at the more general case of a multilayered structure. This will be used in the analysis of radiation in chapter 3.

2.5.2 Modal Methods for Multilayered Slab Structures

To analyze multilayered waveguides we use the common transfer matrix method.[64,65] A transfer matrix is used to calculate the field strength at each interface, relative to some other point in the structure. The method gives a transcendental equation for the propagation constant β.

The formulation used closely follows Chen[65] so it will only be summarized here. The structure is shown in Figure 2.7 where the layers are infinite along x and along the propagation direction z.

Within each layer $\mathbf{j}$ we assume TM fields of the form

$$\mathbf{H_j} = \hat{x} H_{x_j}(y) e^{i(\beta z - \omega t)}, \tag{2.45}$$

$$\mathbf{E_j} = \left[\hat{y} E_{y_j}(y) + \hat{z} E_{z_j}(y)\right] e^{i(\beta z - \omega t)}. \tag{2.46}$$

Putting these into Maxwell's equations (2.15) and (2.16) gives the wave equation

$$\frac{d^2 H_{x_j}(y)}{dy^2} + k_j H_{x_j}(y) = 0, \tag{2.47}$$

where k_j within layers 1 to r has the form

$$k_j = \sqrt{\epsilon_{zz_j}}\sqrt{k_0^2 - \beta^2/\epsilon_{yy_j}}. \tag{2.48}$$

We also find the electric field components

$$E_{z_j}(y) = \frac{-i}{\omega\epsilon_0\epsilon_{zz_j}}\frac{dH_{x_j}(y)}{dy}, \tag{2.49}$$

$$E_{y_j}(y) = \frac{-\beta}{\omega\epsilon_0\epsilon_{yy_j}}H_{x_j}(y). \tag{2.50}$$

H_{x_j} and its y-derivative must be solutions to the wave equation (2.47). For the intermediate layers we assume a sine and cosine dependance of the fields and after some algebra we may relate the fields at the bottom of one layer to the fields at the bottom of

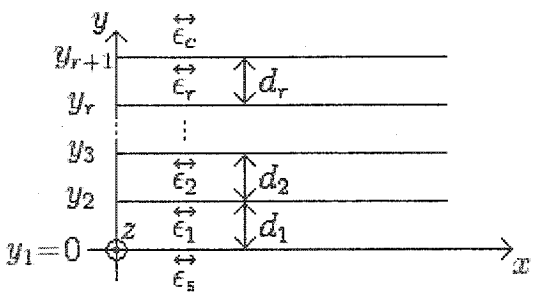

Figure 2.7: Architecture for multilayered slab structure.

the layer above it through

$$\begin{bmatrix} H_{x_j}(y_j) \\ \frac{1}{\epsilon_{zz_j}}\frac{d}{dy}H_{x_j}(y_j) \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(k_j d_j) & -\frac{\epsilon_{zz_j}}{k_j}\sin(k_j d_j) \\ \frac{k_j}{\epsilon_{zz_j}}\sin(k_j d_j) & \cos(k_j d_j) \end{bmatrix}}_{\mathbf{M_j}} \begin{bmatrix} H_{x_{j+1}}(y_{j+1}) \\ \frac{1}{\epsilon_{zz_{j+1}}}\frac{d}{dy}H_{x_{j+1}}(y_{j+1}) \end{bmatrix}. \quad (2.51)$$

In the substrate and cover layers we assume fields which exponentially decay away from the waveguide core

$$\begin{aligned} H_{y_s}(y) &= A_s e^{k_s y} & y < 0, \\ H_{y_c}(y) &= A_c e^{-k_c(y-y_{r+1})} & y > 0, \end{aligned} \quad (2.52)$$

where y_{r+1} is the position of the bottom of the cover layer. A_s and A_c are constants of which only one of them will be independent and is determined by normalization of power in the fields. The transverse wavenumbers k_s and k_c are given by

$$k_s = \sqrt{\epsilon_{zz_s}\left(\frac{\beta^2}{\epsilon_{yy_s}} - k_0^2\right)}, \quad (2.53)$$

$$k_c = \sqrt{\epsilon_{zz_c}\left(\frac{\beta^2}{\epsilon_{yy_c}} - k_0^2\right)}. \quad (2.54)$$

The fields at the top of the substrate ($y = 0$) can be related to the fields at the bottom of the cover ($y = y_{r+1}$) through repeated applications of (2.51)

$$\begin{bmatrix} A_s \\ \frac{A_s k_s}{\epsilon_{zz_s}} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} A_c \\ -\frac{A_c k_c}{\epsilon_{zz_c}} \end{bmatrix}, \quad (2.55)$$

where

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \prod_{j=1}^{r} \mathbf{M_j}. \quad (2.56)$$

$\mathbf{M}_i$ is given in (2.51).

Removing A_s and A_c from (2.55) leads to an equation for the propagation constant β

$$F(\beta) = \frac{k_s}{\epsilon_{zz_s}} m_{11} - \frac{k_s k_c}{\epsilon_{zz_s}\epsilon_{zz_c}} m_{12} - m_{21} + \frac{k_c}{\epsilon_{zz_c}} m_{22} = 0. \tag{2.57}$$

If the waveguide is symmetric we can exploit this property by separately solving for symmetric and antisymmetric modes. We have previously found simplified equations for this case.[66] For symmetric modes we place a so-called *electric* wall halfway through the structure parallel to the other layers where $\hat{n} \times \mathbf{E} = 0$ ($\hat{n}$ is the normal vector to the interface, pointing in the y-direction). The equation (2.57) then simplifies to

$$F(\beta) = m_{11} - \frac{\epsilon_{zz_s}}{k_s} m_{21} = 0. \tag{2.58}$$

For antisymmetric modes we place a *magnetic* wall halfway through the structure where $\hat{n} \times \mathbf{H} = 0$. For this case (2.57) simplifies to

$$F(\beta) = \frac{\epsilon_{zz_s}}{k_s} m_{22} - m_{12} = 0. \tag{2.59}$$

After looking at the simplified case with three layers we will look at modal solutions to these equations.

2.5.3 Modal Methods for Three Layered Slab Structures

For the special case of a three layer structure the problem is greatly simplified. Although broadly known for *isotropic* layers, solutions for *anisotropic* three layered slabs are not given in the standard texts on waveguides,[41,42,62,67–71] nor in the standard work on SPP waveguides with the exception of Stegeman who assumed anisotropic clads in one work[57] and Zhizhin who assumed anisotropic metal.[58] The formulation for three anisotropic

layers is now derived and given here. The basic structure is shown in Figure 2.8 and the layers are again infinite along x and z.

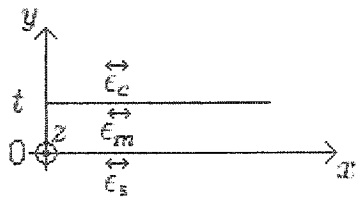

Figure 2.8: Architecture for three layer slab structure.

We begin by looking at the three layer slab where $\overleftrightarrow{\epsilon_r}$ is generally different in each layer. TM fields within each layer must again be solutions to the wave equation (2.47), and are generally of the form

$$\mathbf{H} = \hat{x} H_x(y) e^{i(\beta z - \omega t)}, \tag{2.60}$$

$$\mathbf{E} = \left[\hat{y} E_y(y) + \hat{z} E_z(y)\right] e^{i(\beta z - \omega t)}. \tag{2.61}$$

The y-distribution of the H-fields for the bound modes take the form

$$H_x(y) = \begin{cases} H_0 \left(\cos\left(k_m t\right) + \frac{k_s}{k_m} \frac{\epsilon_{zz_m}}{\epsilon_{zz_s}} \sin\left(k_m t\right)\right) e^{-k_c(y-t)} & y > t \\ H_0 \left(\cos\left(k_m y\right) + \frac{k_s}{k_m} \frac{\epsilon_{zz_m}}{\epsilon_{zz_s}} \sin\left(k_m y\right)\right) & 0 < y < t \\ H_0 e^{k_s y} & y < 0 \end{cases} \tag{2.62}$$

where k_c and k_s are the same as (2.53) in the previous section and k_m is found from (2.48) with $j = m$. H_0 is a normalization constant. The electric field components can again be found from (2.50) and (2.49)

Enforcing continuity of tangential fields at the boundaries leads to the coefficients for the fields in (2.62) above, and gives a transcendental equation for the propagation

constant

$$\tan(k_m t) = \frac{k_m\left(\frac{\epsilon_{zzm}}{\epsilon_{zzc}}k_c + \frac{\epsilon_{zzm}}{\epsilon_{zzs}}k_s\right)}{k_m^2 - \frac{\epsilon_{zzm}^2}{\epsilon_{zzc}\epsilon_{zzs}}k_c k_s}. \tag{2.63}$$

There is a simplification if the waveguide is symmetric ($\overleftrightarrow{\epsilon_c}=\overleftrightarrow{\epsilon_s}$) where (2.63) becomes

$$\tan(k_m t) = \frac{2k_c k_m}{\frac{\epsilon_{zzc}}{\epsilon_{zzm}}k_m^2 - \frac{\epsilon_{zzm}}{\epsilon_{zzc}}k_c^2}. \tag{2.64}$$

For symmetric modes of symmetric structures (2.64) becomes

$$\tan(k_m t/2) = \frac{k_c}{k_m}\frac{\epsilon_{zzm}}{\epsilon_{zzc}}, \tag{2.65}$$

and for antisymmetric modes (2.64) becomes

$$\tan(k_m t/2) = -\frac{k_m}{k_c}\frac{\epsilon_{zzc}}{\epsilon_{zzm}}. \tag{2.66}$$

These results reduce to the common results referenced above, for isotropic media.

2.5.4 Mode Spot Size

The mode spot size is defined and found similar to the case of a single interface SPP, however this time the thickness of the core must also be accounted for. In lossless modes k_s and k_c are real, but in lossy modes they is complex. The imaginary part of k_j $(j = s, c)$ will also contribute to the mode spot size if it is large enough. If $(k_j^R)^2 >> (k_j^I)^2$ where $k_j = k_j^R + ik_j^I$, the mode spot size is

$$W_{1/e} = \frac{1}{k_s^R} + \frac{1}{k_c^R} + t \tag{2.67}$$

where t is the thickness of the core. If t is large enough to significantly contribute to the mode size, the interfaces will usually be decoupled.

2.5.5 Validation of Mode Solver

To validate the model we compare with an often cited result[12] for a three layer, asymmetric slab structure with ϵ_m=-19+i0.53 (silver), ϵ_c=4.0, and ϵ_s=3.61 at a free space wavelength of λ_0=0.633μm. We solve for the two bound modes: s_b and a_b (we will discuss these modes further in the next section), using the multilayer slab model and compare with values extracted from the referenced curve. Figure 2.9 shows that the current implementation of the model is in agreement with the reported results. The new equations derived for the three layer cases concur with these results.

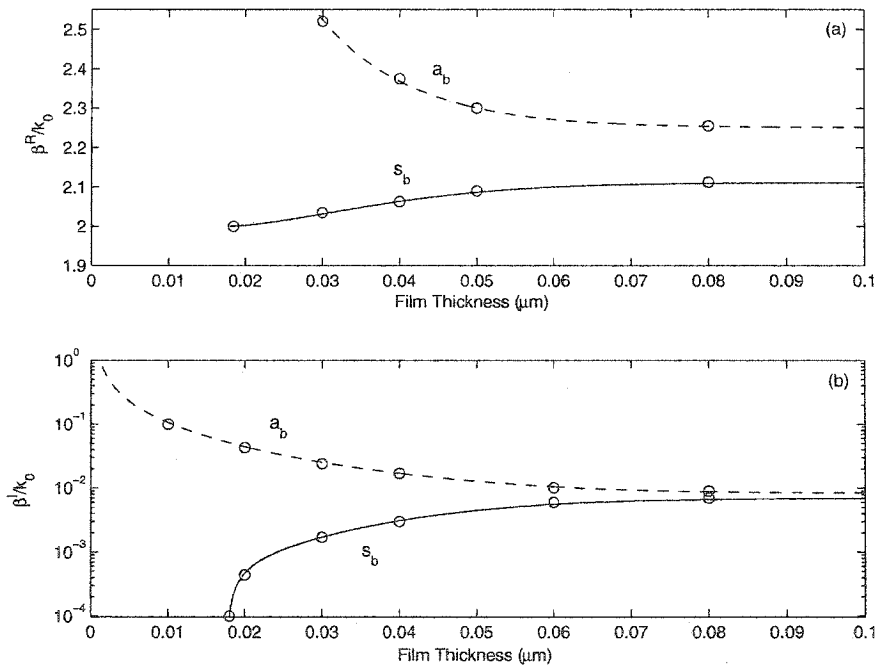

Figure 2.9: Validation of mode solver for 3-layer SPP modes. Real (a) and imaginary (b) parts of the normalized β. Circles are approximate values taken from curves reported by Burke.[12]

2.5.6 Bound Modes of Symmetric Slabs: Silica and Lithium Niobate

The topic of bound modes in symmetric slabs has been thoroughly documented,[10–13,72,73] so we will only briefly go over the fundamental modes for the two materials we are interested in, those being silica (SiO_2) and lithium niobate ($LiNbO_3$). We use Au for the metal throughout.

Dispersion curves (ω vs. β) for thin slabs have been reported[9,10,49] and are reminiscent of the curves for the single interface case only with the plasmon band-gap smaller. These curves differ considerably above the surface plasmon frequency when material damping is considered.[61] The approach here, however, will be to look at design curves for a single frequency, well below the plasma frequency.

The equations for the propagation constant are transcendental, so they must be solved numerically. They are rewritten with all terms moved to one side of the equality and set equal to zero as in (2.57). Zeros of these complex functions are then found using Muller's method.[74]

Figure 2.10 shows the propagation constant for a Au film of thickness t surrounded above and below by silica at $\lambda_0 = 1.55\mu m$. The material properties are the same as before. Only two bound modes exist for this case in contrast to dielectric slabs where many higher order modes exist for large t. This makes sense remembering that plane waves in metals do not propagate in the frequency range that surface plasmons do, so oscillations within the metal are not expected. The symmetric mode is labelled s_b while the antisymmetric mode is labelled a_b, based on the symmetry of the H_x or E_y fields about the center of the metal layer.

The a_b mode should not be thought of as a higher order mode than the s_b since they have quite distinct behavior. As t decreases the MPA for the s_b mode decreases to zero

in the limit that $t \to 0$. In this limit the mode becomes a plane wave in the surrounding material as we see the effective index $\beta^R/k_0 \to 1.444$, the index of refraction of the silica. Just before this limit at $t \simeq 0$ the mode is infinite in extent as seen in Figure 2.11. The a_b mode does exactly the opposite as shown in both figures. As $t \to 0$ the mode becomes tightly confined to the metal core region. Neither mode has a cutoff thickness and for t large enough they share the same propagation characteristics as a single interface SPP.

Because the s_b mode has relatively lower MPA for thin films it is sometimes called the "long-range SPP" with the a_b mode then being termed the "short-range SPP".

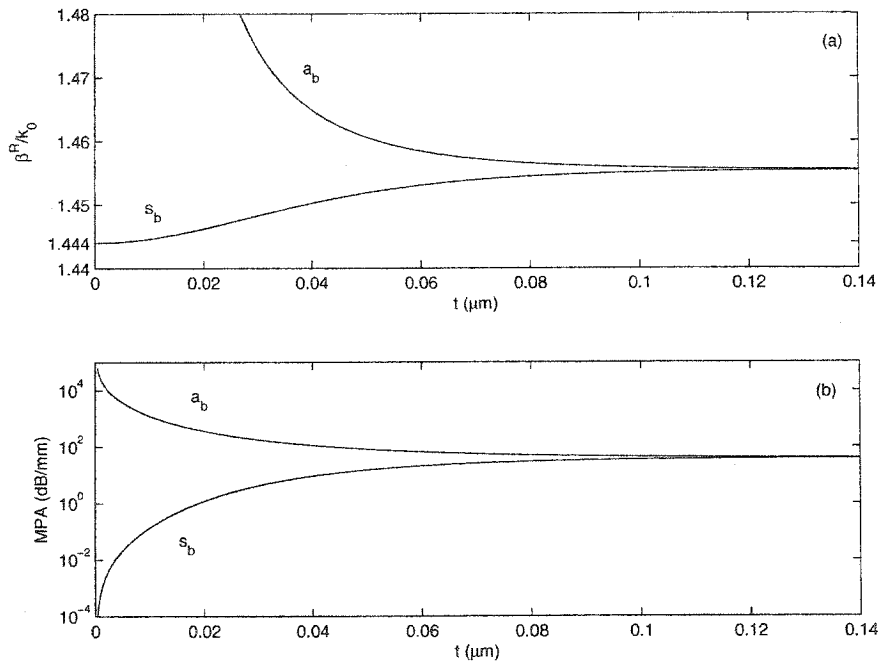

Figure 2.10: Propagation constants for s_b and a_b SPPs for a Au film with silica clads.

Figure 2.12 shows the shape of the H-fields for (a) the s_b and (b) the a_b modes at $t = 20nm$. The a_b mode is significantly more confined to the waveguide core than the s_b mode and hence has a larger amplitude for the same power normalization.

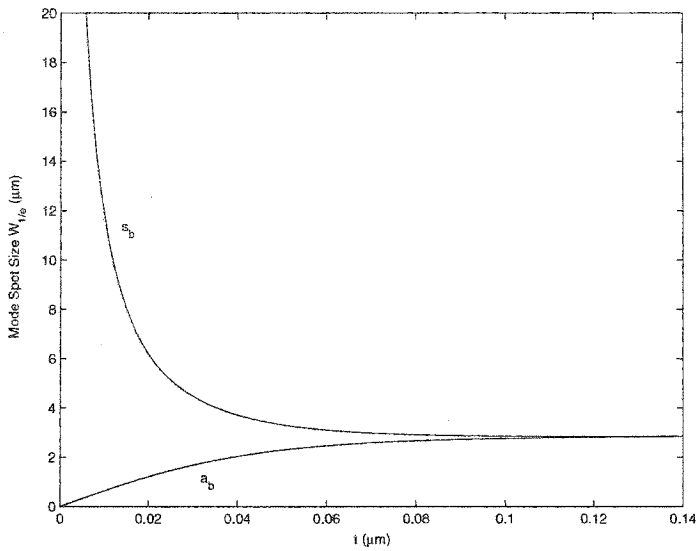

Figure 2.11: Mode spot size for a_b and s_b SPPs for Au film with silica clads.

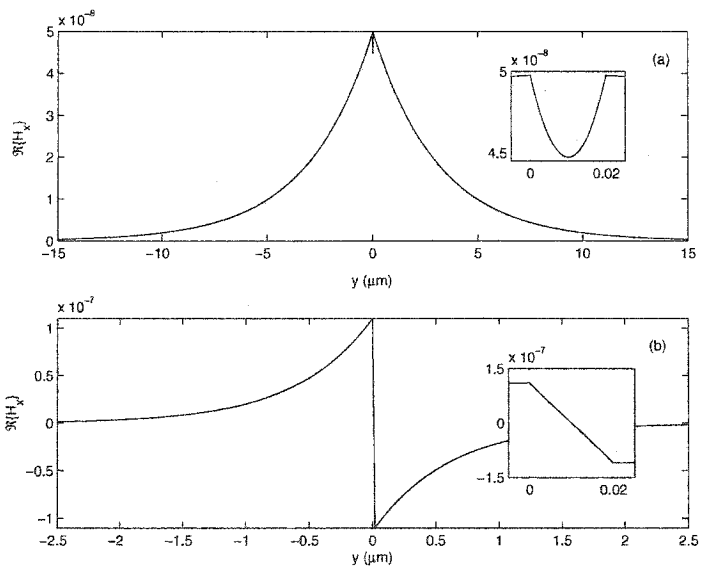

Figure 2.12: Real part of the H-fields in $A/\mu m$ for (a) s_b and (b) a_b SPPs for a Au film $(0 < y < 20nm)$ with silica clads. Insets show closeup of field in the metal (same axes).

Similar calculations were done for lithium niobate (LN). LN is a uniaxial anisotropic crystal,[41,75] meaning that the relative permittivity along one crystal axis (the extraordinary axis, denoted with a subscript e) is different from the other two axes (the ordinary axes, denoted with subscripts ord). As in silica, there are two bound modes for this structure, but in this case they will have different properties depending on the orientation of the LN crystal axes relative to the interface. Since we are only considering symmetric structures in this section, we assume that the crystals are oriented the same above and below the Au film. At $\lambda_0 = 1.55\mu m$ $\epsilon_e = 2.1377^2 = 4.5698$ and $\epsilon_{ord} = 2.2125^2 = 4.8952$.

Figure 2.13 shows the propagation constants for the three crystal orientations. The s_b and a_b modes show analogous behavior to the silica case, but we see some significant differences within each mode depending on the crystal orientation. The solid curves correspond to $\epsilon_{zz} = \epsilon_e$, the dashed curves $\epsilon_{yy} = \epsilon_e$, and the dotted curves $\epsilon_{xx} = \epsilon_e$. In the limit as $t \to 0$, $E_z(y) \to 0$ (since $dH_{x_i}(y)/dy \to 0$) and E_y becomes the only electric field component. The effective index of the mode then converges to that of a transverse electromagnetic mode polarized along the y-direction, $n_{eff} = n_{yy} = \sqrt{\epsilon_{yy}}$ which we see from the s_b curves in Figure 2.13. We can also see this analytically from (2.65) if we take the limit as $t \to 0$.

The solid and dotted curves are very similar as β is much less dependant on the value of ϵ_{zz} (both curves have $\epsilon_{yy} = \epsilon_{ord}$). On this scale MPA is the same for each mode.

Where β was more dependant on ϵ_{yy}, the mode spot size is more dependant on ϵ_{zz} as shown in Figure 2.14. This difference is more subtle than for β and will probably be harder to verify experimentally than the β curves. In these figures the dotted and dashed curves are overlapping. The inset shows a zoom in on the spot size for values less than $2\mu m$ and for a larger thickness range (with the same units as the main figure). For $t > 0.15\mu m$ the spot size begins to increase linearly due to the increasing core size. At this point the fields at each interface are essentially uncoupled.

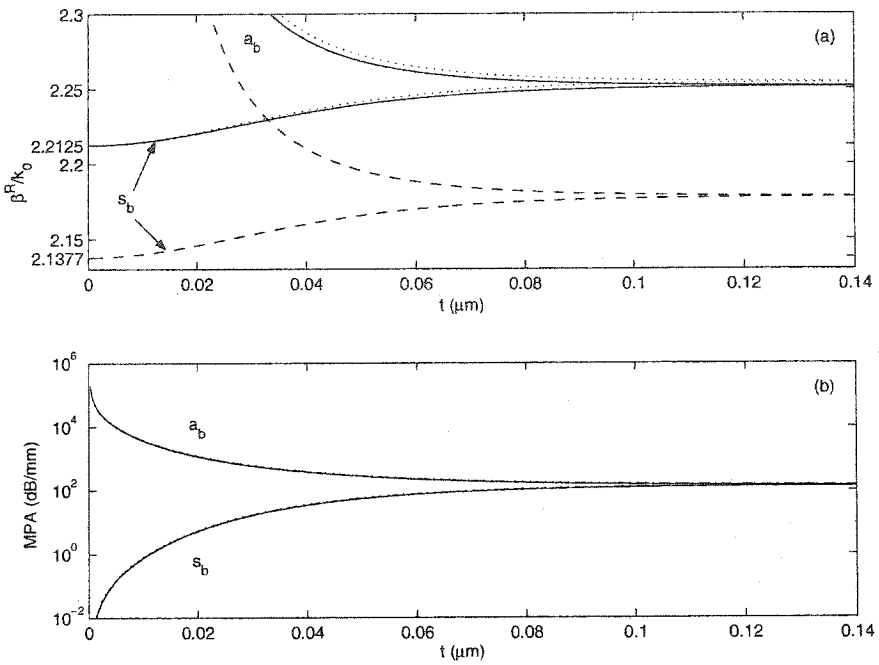

Figure 2.13: Propagation constants for a_b and s_b SPP modes for a Au film with $LiNbO_3$ clads. Solid curves: $\epsilon_{zz} = \epsilon_e$, dashed curves: $\epsilon_{yy} = \epsilon_e$, dotted curves: $\epsilon_{xx} = \epsilon_e$.

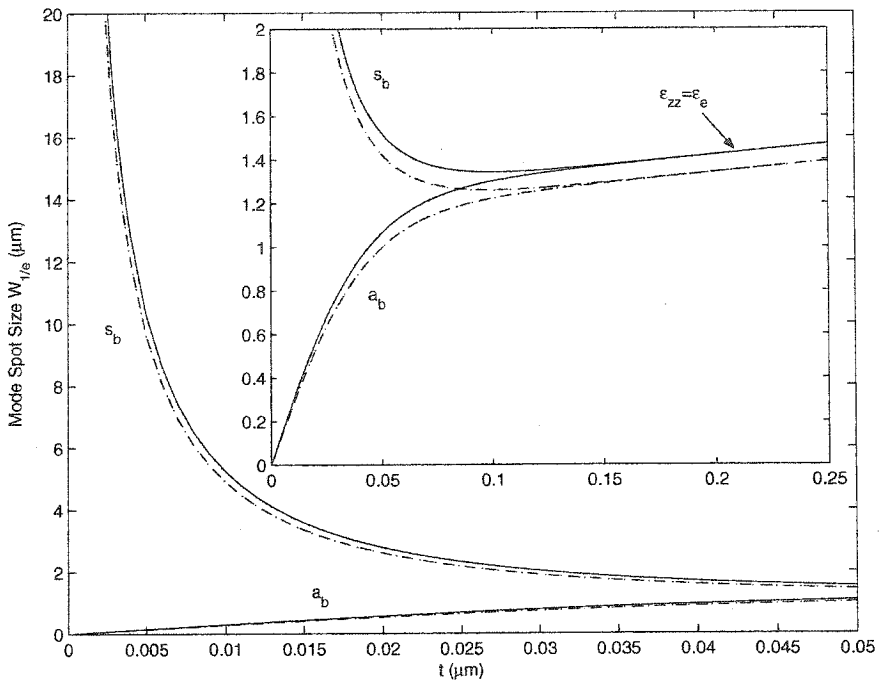

Figure 2.14: Mode spot size for a_b and s_b SPP modes for a Au film with $LiNbO_3$ clads. Inset: same units as main figure. Solid curves: $\epsilon_{zz} = \epsilon_e$, dashed curves: $\epsilon_{yy} = \epsilon_e$, dotted curves: $\epsilon_{xx} = \epsilon_e$.

2.5.7 Metal-Insulator-Metal Slabs

The converse of the three layer structure described in the introduction is one where the core is the dielectric and the clads are metal, known as a metal-insulator-metal (MIM) structure (also known as metal-oxide-metal). Using the same model we can investigate modes of this sort. There again exists one symmetric and one antisymmetric surface mode for the symmetric waveguide case. Traditional dielectric core modes with the field maximum in the core rather than at the surface are also supported. The H-fields for the two surface modes are shown in Figure 2.15 for a 5μm thick silica core with Au clads at $\lambda_0 = 1.55\mu m$. The propagation constants are shown in Figure 2.16 and we see that these modes are less useful for waveguiding since the loss never decreases below that of the single interface mode. The antisymmetric mode seems to have a resonance around $t = 0.5\mu m$ and below this thickness, the propagation constant approaches zero. This mode still appears to be guided with the same shape for a thin dielectric core, but the term guided mode is only used loosely here as the mode has attenuation greater than 1000dB/mm.

These structures have been shown to emit light in what is called a tunnel junction where the dielectric layer is very thin (of the order of 10nm) and a bias is applied to the top and bottom metal.[22] We mention these modes here for completeness only and will not discuss them further as they have been well documented elsewhere.[10,25,73,76]

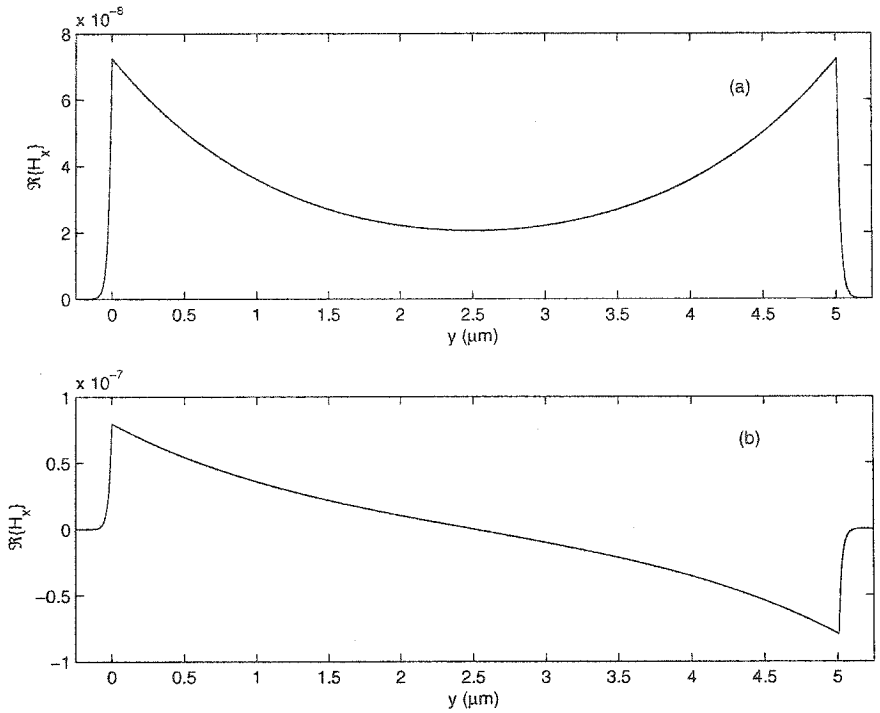

Figure 2.15: Real part of H_x fields in $A/\mu m$ for (a) symmetric and (b) antisymmetric modes of MIM waveguide.

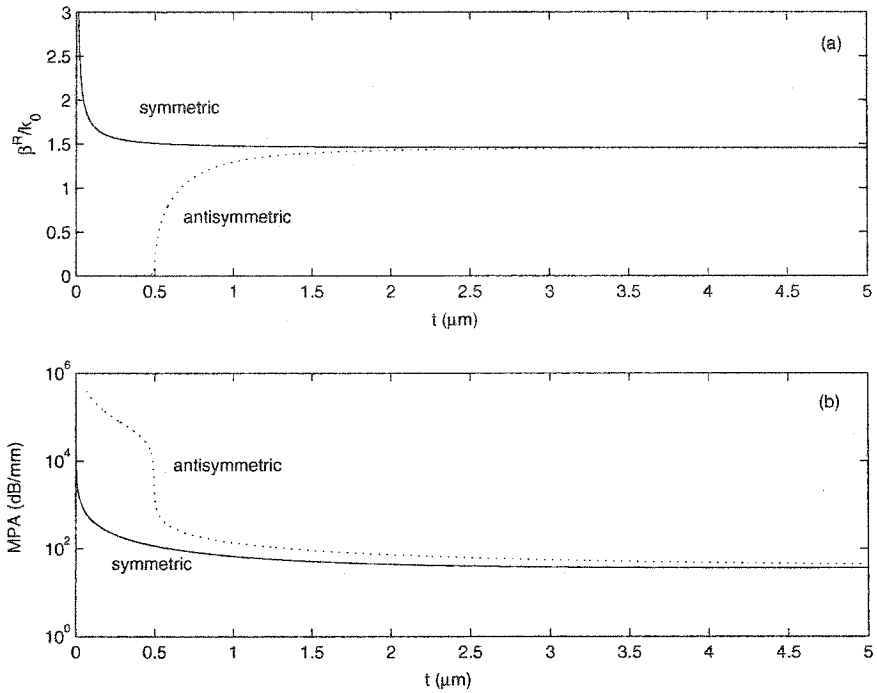

Figure 2.16: Propagation Constants for MIM surface modes. Silica core with Au clads.

Chapter 3

Mode Cutoff in Infinite Width SPP Waveguides

3.1 Introduction

In this chapter we look at mode cutoff in SPP waveguides, specifically cutoff of the s_b mode outlined in the last chapter for the three layer dielectric-metal-dielectric waveguides. We look at cutoff caused by an asymmetry in the relative permittivities of the layers above and below the metal layer, so that $\overleftrightarrow{\epsilon}_c \neq \overleftrightarrow{\epsilon}_s$.

For small asymmetries the mode is still supported while if the asymmetry is too large the mode becomes cutoff and no low loss bound modes are supported.[12,30] We will see that the a_b mode is still supported, but recall that it has a high loss. Light input into this structure then radiates away from the core.

This feature of SPP waveguides has been employed in the modulation of light[14–16] where the existence of the mode, or coupling to it, can be controlled by index of refraction changes caused by the thermo-optic or electro-optic effects. In these methods the SPP modes are used as the 'off' state, where light is coupled into them to produce a diminished

output. The opposite approach is used here, with the s_b mode being the input and output state and the asymmetric region being used to radiate the energy away from the metal (rather than into it). We seek in the next two sections to have a better understanding of the length required to sufficiently attenuate an input mode in a region where the bound mode is asymmetric or cutoff.

3.2 Cutoff of Bound Modes

3.2.1 Antisymmetric Mode

We begin our study of mode cutoff by looking at the antisymmetric a_b mode. The bound modes of asymmetric structures keep the same names for simplicity and they are identified as the modes evolving from the symmetric and antisymmetric modes of a symmetric structure.

Figure 3.1 shows the effect of a changing index of refraction of the cover n_c on the propagation constant. The bottom clad is silica (n_s=1.444) and the metal is an idealized lossless Au layer with ϵ_m=-131.95. The free space wavelength for all cases in this chapter is $\lambda_0 = 1.55\mu m$. Three metal thickness values are shown and the dotted curve represents the higher of the index values of the substrate and cover layers. We see that β^R/k_0 is greater than the index of either clad for the large range of n_c shown, and for all thickness. The a_b mode then does not have a cutoff asymmetry and remains a bound mode. This mode is therefore not of interest for this cutoff application.

3.2.2 Symmetric Mode in Isotropic Media

As mentioned in the introduction, the symmetric s_b mode is susceptible to cutoff. Figure 3.2 shows the same structure as above with n_s=1.444 and ϵ_m=-131.95 and changing n_c,

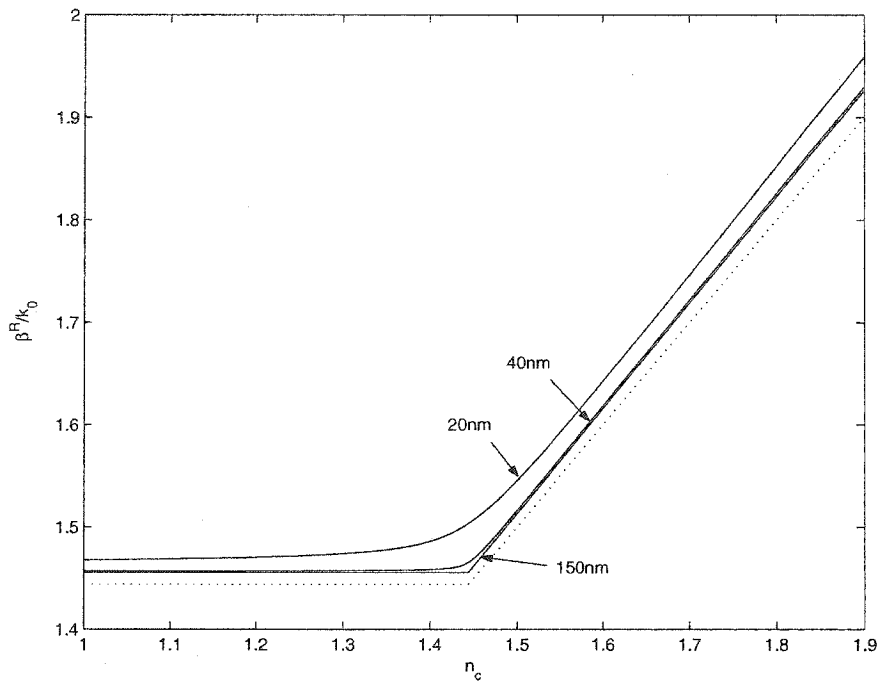

Figure 3.1: Propagation constant of the lossless a_b mode for labelled metal thickness with n_s=1.444 and ϵ_m=-131.95 and with changing cover index of refraction n_c. The dotted curve represents the larger of n_s and n_c.

for several metal layers. The dotted curve again represents the higher of the two cladding index values. In this case we see cutoff occurring where β^R/k_0 is equal to either 1.444 or n_c for certain index asymmetries. Notice that the asymmetries are much smaller than those applied to the a_b mode. As the thickness is increased a larger asymmetry is needed to cut off the mode. This also means that for a given asymmetric structure there will be a thickness cutoff, which is something we did not see in symmetric structures.

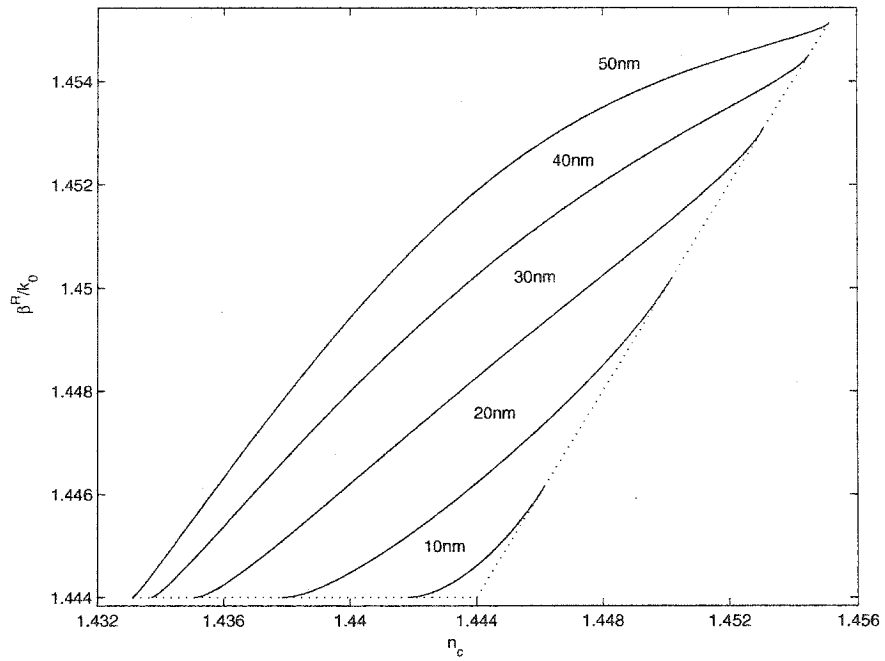

Figure 3.2: Cutoff curves for the s_b mode of a lossless Au layer of labelled thickness, n_s=1.444 and ϵ_m=-131.95. The dotted curve represents the larger of n_s and n_c.

Recall that the fields above and below the core exponentially decay away from the interface according to the transverse wavenumbers k_c and k_s which for isotropic media reduce to

$$k_{c,s} = \sqrt{\beta^2 - \epsilon_{c,s} k_0^2}. \tag{3.1}$$

At the cutoff points $\beta^2 \rightarrow \epsilon_i k_0^2$ where ϵ_i is the higher of ϵ_s or ϵ_c so the corresponding

transverse wavenumber goes to zero. At this point the mode is infinitely large into the higher index media and we consider the mode cutoff. It is essentially a plane wave travelling in that medium and this mode can no longer be excited by a finite source.

Figure 3.3 shows the spot size for the above case. Near the cutoff points the spot size increases sharply so that the mode is only of a practical size for a limited range of n_c.

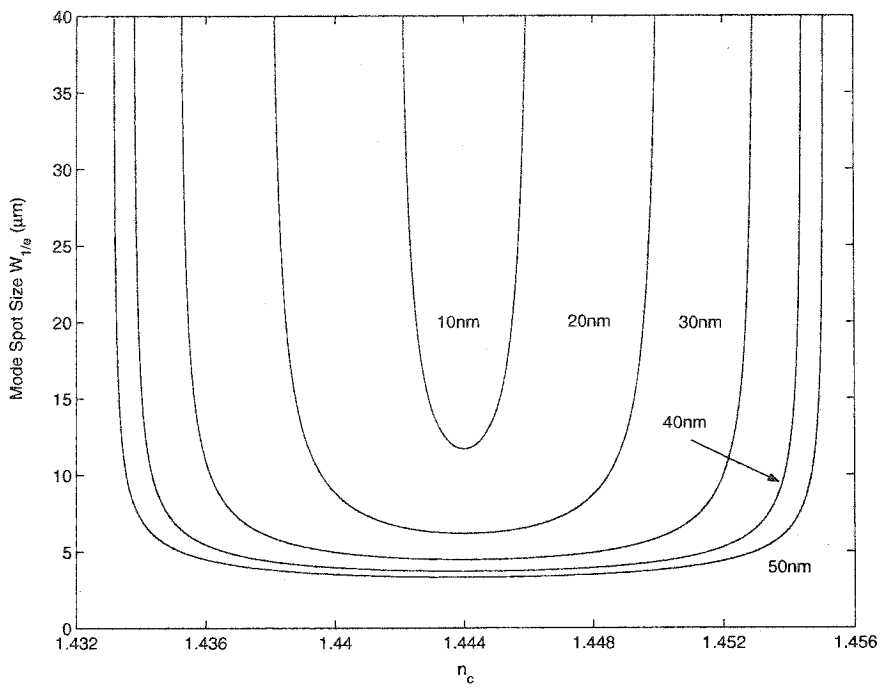

Figure 3.3: Spot size of the s_b mode for changing cladding index n_c and labelled Au thickness. n_s=1.444 and ϵ_m=-131.95.

In Figure 3.4 we see an example of the evolution of the s_b mode as the index of refraction asymmetry is increased from 0 to 18×10^{-4}, which is very close to the cutoff point. In this example, the waveguide is a 9nm Au slab ($\epsilon_m = -131.95 + i12.65$) with $n_s = 1.444$ and a varying cover layer ($n_c = 1.444 + \delta n$). The normalization of unit power (at $z = 0$) is consistent between the four plots, and we see the mode spreading into the higher index region (y>9nm). Notice the horizontal scale on the fourth plot compared

with the others.

Near the metal itself, the field peak at the interface with the low index ($y = 0$) side has a higher amplitude than the field at the interface with the high index region, $y = 9nm$, as seen in the inset in the fourth plot.

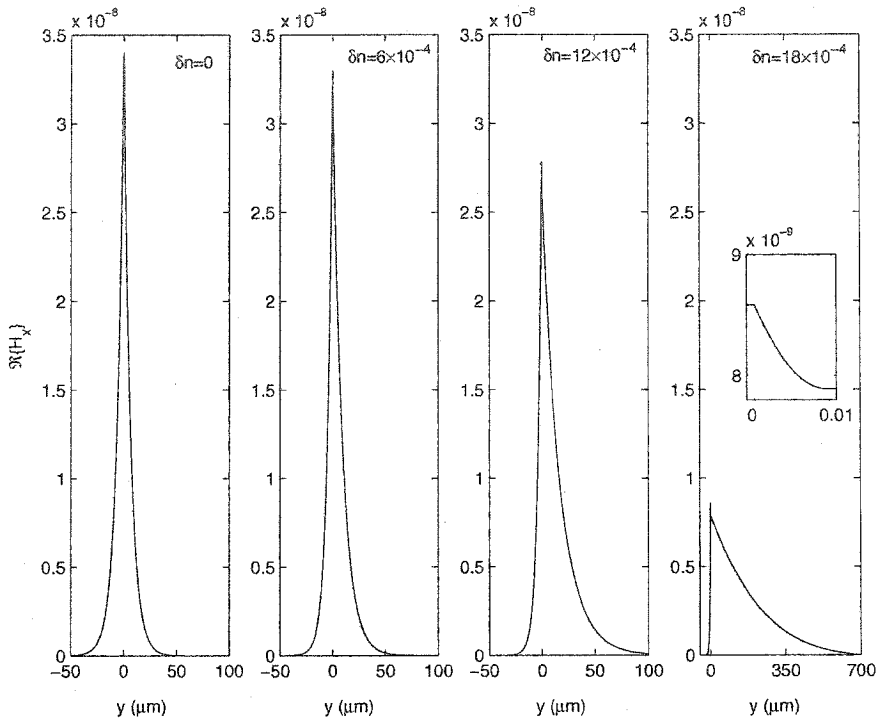

Figure 3.4: H_x-field profiles for a 9nm Au film in SiO_2 for successive index asymmetry δn, where $n_c = n_s + \delta n$. All vertical axes are the same. Inset in fourth picture is a closeup of the metal region with the same axes.

We now look at cutoff curves for lossy Au (ϵ_m=-131.95+i12.65) and see that the results are very similar for thin Au, but the cutoff points differ for thicker Au. Figure 3.5 shows these results. Assuming that n_c is increased to the point where β^R/k_0=n_c (the cutoff point for the lossless modes), for isotropic, lossy media we have

$$k_c = \sqrt{\beta^2 - n_c^2 k_0^2}$$

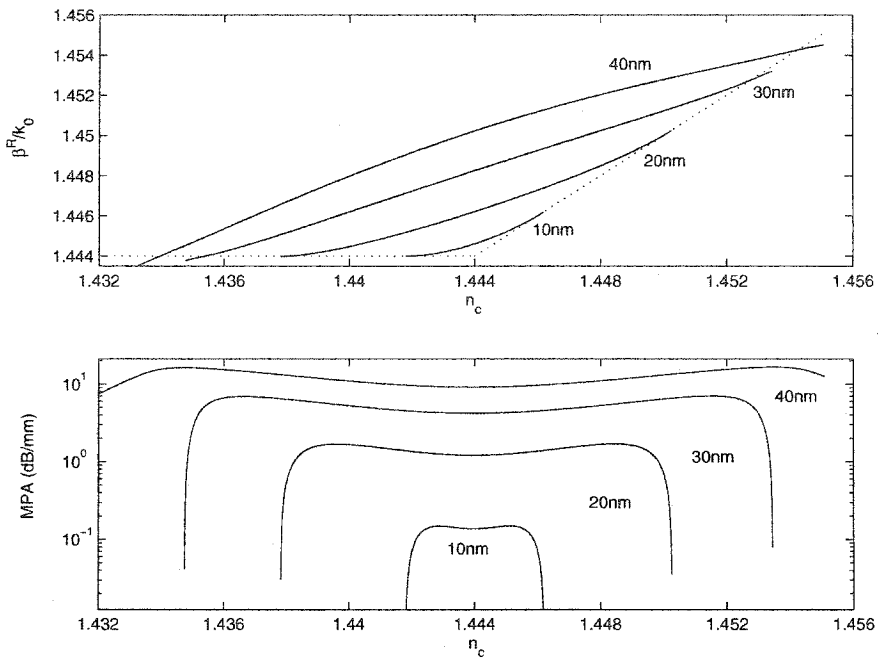

Figure 3.5: Cutoff curves and propagation constants for the s_b mode of a lossy Au film of labelled thickness with an SiO_2 substrate. n_s=1.444 and ϵ_m=-131.95+i12.65. The dotted curve represents the larger of n_s and n_c.

$$= \sqrt{(\beta^R + i\beta^I)^2 - n_c^2 k_0^2}$$
$$= \sqrt{(\beta^R)^2 + (\beta^I)^2 - n_c^2 k_0^2 + i2\beta^R\beta^I}$$

but $\beta^R = n_c k_0$ and $\beta^R \gg \beta^I$ so

$$k_c \simeq \sqrt{\beta^R\beta^I} + i\sqrt{\beta^R\beta^I}.$$

Here we have used that $\sqrt{z} = \pm\left[\sqrt{(|z|+x)/2} + (sign(y))i\sqrt{(|z|-x)/2}\right]$ for $z = x + iy$ courtesy of Kreyszig.[77]

The real part of the transverse wavenumber k_c is not quite zero, because of the small imaginary part of β so the mode is not cutoff. From Figure 3.5 we see that a slightly larger n_c is needed for cutoff. This is more noticeable in thicker metals (higher loss, larger β^I), as shown by the t=30nm curve where the mode cuts off just past the dotted line. The t=40nm curves cuts off for much larger asymmetry and the cutoff point is not shown here. This blurring of the cutoff point occurs at a point on the curve where the mode is very large and the imaginary part of the wavenumber causes the fields to be oscillatory into the high index media rather than constant as in the lossless case.

At the cutoff points, the attenuation goes to zero as all of the power resides in one of the clads, and is not absorbed by the metal. Some authors have interpreted this as a means to increase the propagation length by many orders of magnitude.[30] This may work in some experimental techniques such as ATR, but we caution that for integrated optics applications which use an end-fire excitation scheme the mode will be difficult to excite since at the cutoff or low attenuation point it is impractically large (infinite in the limit), while typical input and output fiber modes have spot size of the order of $10\mu m$.

3.2.3 Symmetric Mode in Anisotropic Media

In this section we present the results for $LiNbO_3$ (LN). LN is electro-optic, meaning that its permittivity tensor varies with an externally applied electric field. LN has its largest change in index of refraction along the extraordinary axis.[41] It will be shown that cutoff only occurs for changes to n_{yy} so we orient the LN above and below the Au layer to have $n_{yy} = n_e$, to make the results applicable to the optimal LN orientation. This is the so-called "z-cut LN" or "z-LN" referring to the crystal axis of the LN (not to be confused with the z-axis in our coordinate system).

Figure 3.6 shows the results where we use the value for lossy Au given above and λ_0=1.55μm. In the cover n_{yy} varies as shown (corresponding to an externally applied E-field along y) while n_{zz}=2.2125 remains constant. In practice n_{zz} also varies with this electric field, but to a lesser degree. In the substrate $n_{yy} = 2.1377$ and $n_{zz} = 2.2125$ to match the cover. We will look at how this E-field can be applied in the next chapter.

The behavior here is similar to the silica case and a comparison is made between the two in the next section. The full cutoff curve is again not shown for the t=40nm case since its cutoff points are much farther out.

Recalling the transverse wavenumber now in anisotropic media

$$k_{s,c} = \sqrt{\epsilon_{zz_{s,c}} \left(\frac{\beta^2}{\epsilon_{yy_{s,c}}} - k_0^2 \right)}. \tag{3.2}$$

From Figure 3.6 the cutoff points (lossless) are when $\beta^R/k_0 = n_{yy} = \sqrt{\epsilon_{yy}}$ so that either k_s or k_c (corresponding to the high index clad) again go to zero.

For completeness we look at cutoff based on a change in the n_{zz} component of the tensor for the cover with n_{yy}=2.1377 in the cover, and all other values as above. As seen from Figure 3.7 no cutoff point exists even for the large range of n_{zz} and small thickness of 2nm. We conclude that mode cutoff is achieved by varying the component of the

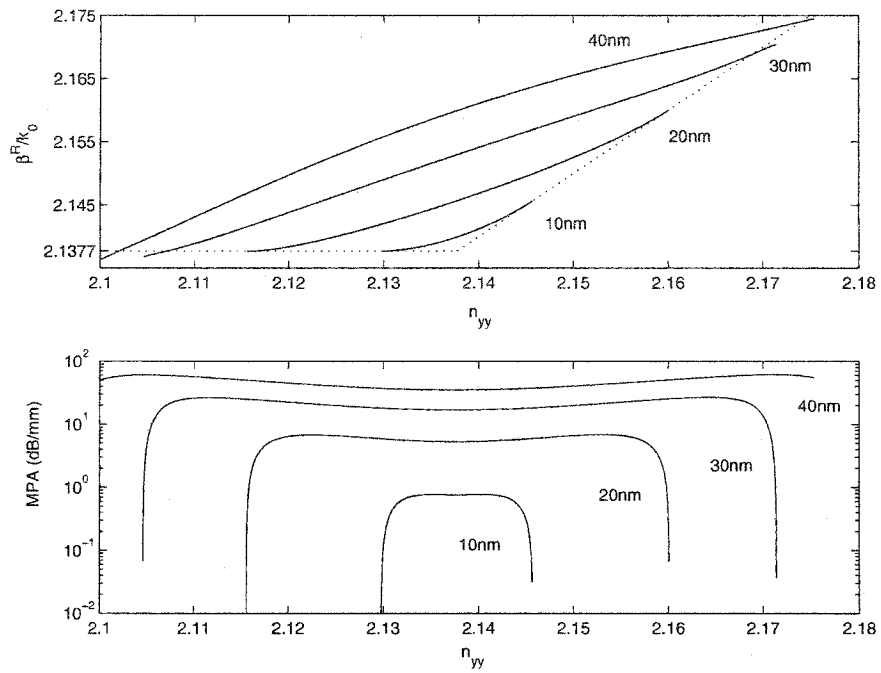

Figure 3.6: Cutoff curves and propagation constants for the s_b mode of a Au layer of labelled thickness with an LN substrate. Varying n_{yy} in top clad. The dotted curve represents the larger of n_{yy} in the substrate and the cover.

index of refraction tensor normal to the interface, the n_{yy} component. Recalling that the electric field E_y was much larger than E_z it makes sense that the mode is more sensitive to the n_{yy} component.

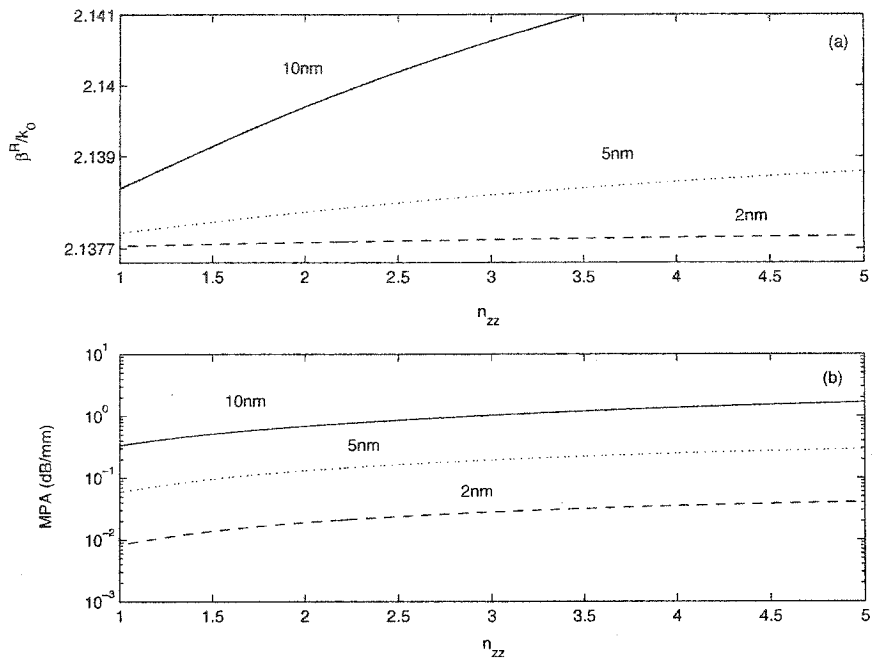

Figure 3.7: Cutoff curves for the s_b mode of Au with an LN substrate. Varying n_{zz} in the cover. Curves labelled with Au thickness.

3.2.4 Comparison of Silica and Lithium Niobate Modes

Modes in each media (SiO_2 and LN) are now compared based on their relative confinement (compared to the background index) which we define as $\delta n_{eff} = n_{eff} - n_{yy}$. For silica then: $\delta n_{eff} = n_{eff} - n = n_{eff} - 1.444$, and for LN: $\delta n_{eff} = n_{eff} - n_{yy} = n_{eff} - 2.1377$. This has been found to be a good indicator of cut off sensitivity since this is approximately the amount of change in $n_{eff} = \beta^R/k_0$ required for the mode to cut off. Figure

3.8 shows that for a given thickness, the LN modes are much more confined than the SiO_2 modes. This agrees with the earlier analysis of mode properties in Figures 2.10 to 2.14 where we saw a decrease in mode spot size, and an increase in attenuation for a given Au thickness when switching from silica to LN. This is due to, among other things, the smaller guided wavelength in LN where the constant metal thickness is now a larger percentage of the wavelength.

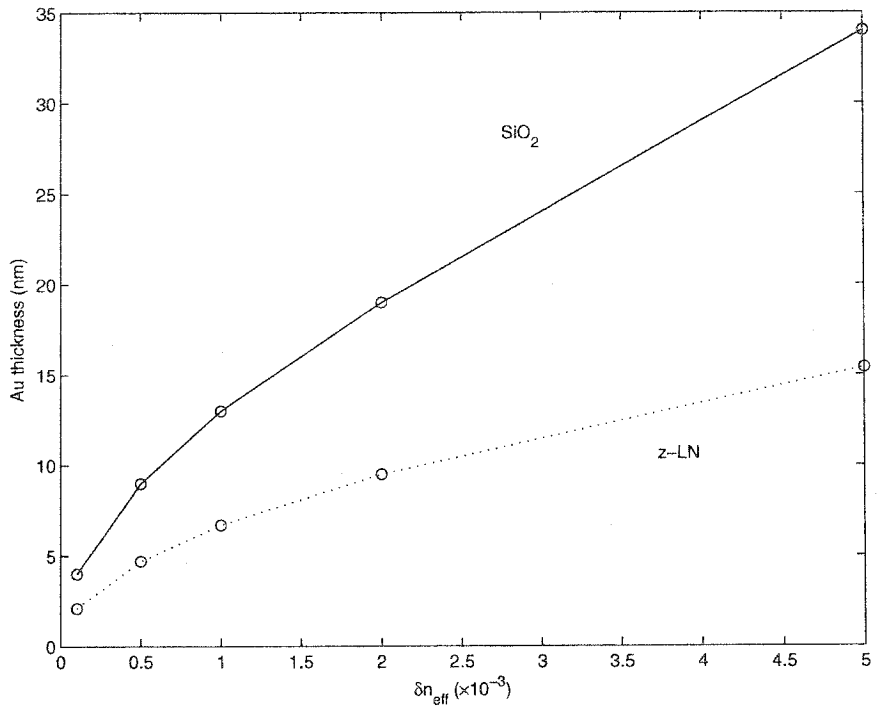

Figure 3.8: Comparison of mode confinement $\delta n_{eff} = n_{eff} - n_{yy}$ for various Au thicknesses in SiO_2 and LN. $n_{yy} = 1.444$ for SiO_2 and $n_{yy} = 2.1377$ for LN.

Because of the higher δn_{eff}, a larger change in the index n_{yy}, defined as δn_{yy} is required for cutoff of the s_b mode in LN than in a similarly thick metal in SiO_2, as shown in Figure 3.9. For modes with similar δn_{eff}, the δn_{yy} required for cutoff converges as both values tend to zero for low confinement. The asymmetry required for cutoff diverges

rapidly for higher confinement with the s_b mode more easily cutoff in SiO_2. Cutoff points are generally very similar for positive and negative δn_{yy} (Figure 3.9 shows the average) and are larger than the actual δn_{eff}.

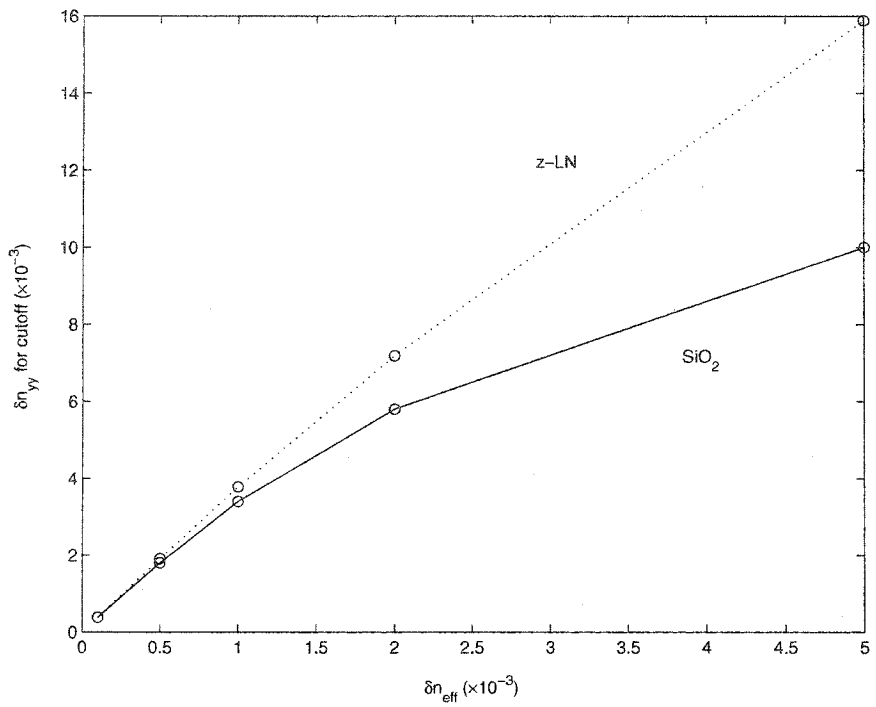

Figure 3.9: Positive and negative δn_{yy} cutoff points for s_b modes of different confinement in SiO_2 and LN.

A number of mode parameters are summarized in tables 3.1 and 3.2. δn_{yy}^{CO} is the index asymmetry for cutoff. The three middle columns are mode properties for symmetric structures. There are no parameters which generally correspond to modes in both materials. For example, modes with similar δn_{eff} have different MPA and cutoff points, while modes with similar MPA have different spot size.

The metal thickness in table 3.2 for the first two cases (2.1nm and 4.7nm) are probably not currently physically realizable as a continuous film, but are used for comparison with

δn_{eff} ($\times 10^{-4}$)	t (nm)	MPA (dB/mm)	Spot size (μm)	δn_{yy}^{CO} ($\times 10^{-4}$)
1.0	4	0.01	29	4
5.0	9	0.1	13	18
8.0	11.8	0.2	10	29
10	13	0.3	9.1	34
20	19	1	6.5	59

Table 3.1: Thickness, attenuation, spot size and δn_{yy}^{CO} values of specific s_b modes for Au in SiO_2 for a range of δn_{eff}.

δn_{eff} ($\times 10^{-4}$)	t (nm)	MPA (dB/mm)	Spot size (μm)	δn_{yy}^{CO} ($\times 10^{-4}$)
1.0	2.1	0.023	23	4
5.0	4.7	0.13	10	20
10	6.7	0.28	9.6	39
20	9.5	0.67	5.2	73

Table 3.2: Thickness, attenuation, spot size and δn_{yy}^{CO} values of specific s_b modes for Au in LN for a range of δn_{eff}.

SiO_2 guides with similar confinement. In the next section we will look at what happens when light is propagated in a region where these specific s_b modes are cutoff.

3.3 Normal Mode Analysis of Radiation

3.3.1 Introduction

If the conditions for cutoff of the s_b mode are met as described above, and the cutoff portion of the waveguide is sufficiently long we can be confident that no energy will be coupled back out of this waveguide, since there are no other low loss bound modes. The question remains about what happens to the energy in the intermediate region closer to the excitation point. In this region we must consider all of the bound and radiative normal modes of the waveguide.[45]

Although no low-loss bound modes are supported when the s_b mode is cutoff, radiative modes will still be supported. Radiative modes will be oscillatory in the regions above and/or below the waveguide. These modes are approximated by enlarging the waveguide system with thicker regions above and below the metal and additional free space clads outside of these thicker regions as shown in Figure 3.10 (we will continue to refer to the regions above and below the metal layer as the cover and substrate). As the thickness of the ϵ_c and ϵ_s regions in section 1 increases, the number of normal modes supported tends to infinity and approaches a radiation continuum. Since the radiation continuum is not practically realized, the modes are discrete, with each mode having a unique number of oscillations in the substrate and cover regions. To model radiation of our input mode we decompose it into a basis of these normal modes. The basis is said to be complete (for this application) if it captures all of the input energy. Finding a complete basis that contains all possible modes is typically unnecessary and inefficient since coupling into many of these modes is negligible. The modes of this structure are now bound not to the metal region, but to the region between the free space layers. The system is closed, but it is possible to model free radiation over a finite distance if the system is large enough.

This method is analogous to Fourier optics where free space radiation is calculated

by decomposing the input excitation into a basis of plane waves. However plane waves are not supported by this structure since even though the bound mode is cutoff, there still exists a waveguide and two cladding regions and any optical wave must still be a solution to Maxwell's equations in this region.

Electric or magnetic walls could also be used at the outer boundaries rather than free space layers, and would be necessary if radiating into free space, but for this application are unnecessary. Using free space layers, the multilayer slab model used previously may be employed by simply adding two thick layers, and a new equation for β does not need to be derived. Either method would produce identical results for this application, as the outer walls are moved far enough away from the core so as not to impact the results.

This common method is known as *normal mode analysis*, but is also referred to as the *eigenmode expansion method*.[78] Related applications to this one were to model end-fire excitation, reflection, transmission and radiation of single interface SPPs at abrupt geometry boundaries[79] and modelling cutoff of higher order dielectric modes.[80] Another

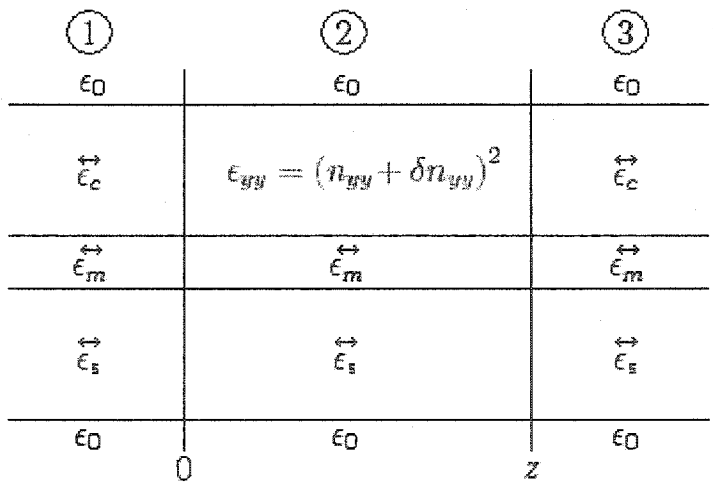

Figure 3.10: Geometry for normal mode analysis. 1: input region, 2: asymmetric region, 3: output region.

common method for modelling waveguide radiation is leaky-mode analysis, but Smith and Houde-Walter showed recently that this approximate method fails for waveguides near and beyond cutoff, since incorrect results are predicted.[80]

To our knowledge, neither the normal mode analysis technique nor any other method has been used to calculate the radiation of cutoff SPP modes in slabs or in strips.

3.3.2 Transmission Between Regions

We will consider three regions for this analysis, the input region 1, the radiating region 2, and the output region 3, as shown in Figure 3.10.

Strictly speaking, a portion of the input field in region 1 is transmitted into region 2 while the remainder is reflected back into region 1. To calculate the transmission we assume a local plane wave like behaviour at each point in the mode cross section and use the transmission and reflection coefficients for plane waves. The transverse electric field of the input just to the right of the interface ($z = 0^+$) in terms of the field just to the left ($z = 0^-$) is

$$E_{2,y}\left(x, y, z = 0^+\right) = t_\perp E_{1,y}\left(x, y, z = 0^-\right), \tag{3.3}$$

where $t_\perp = 2n_1/(n_1 + n_2)$ for normal incidence and n_1 and n_2 are the index of refraction values in regions 1 and 2, respectively.[40] The corresponding H-field is then found from (2.50).

The energy coupled across is found from the transmissivity, which for normal incidence (which is the case here) is

$$T = \frac{4n_1 n_2}{(n_1 + n_2)^2}. \tag{3.4}$$

The corresponding reflectivity is

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2, \tag{3.5}$$

so that energy is conserved with $R + T = 1$.

For our analysis we will be looking at the input mode from a symmetric waveguide coupled into an asymmetric waveguide. For much of the mode cross section, $n_1 = n_2$ so $T = 1$. In the remaining region we increase the index of refraction by an amount δn. Putting $n_2 = n_1 + \delta n$ into (3.4) gives

$$T = \frac{4n_1\,(n_1 + \delta n)}{4n_1\,(n_1 + \delta n) + (\delta n)^2}. \tag{3.6}$$

Now since $\delta n \ll n_1$, $T \simeq 1$. Specifically for $n_1 = 1.444$ (SiO_2) and $\delta n < 0.094$, $T > 0.999$ and for $n_1 = 2.1377$ (LN) and $\delta n < 0.14$, $T > 0.999$. Since we will not be using asymmetries larger than this, we may safely assume $T = 1$ and ignore reflections from interfaces. This simplifies the analysis as we only have to account for forward propagating modes and can ignore coupling to backwards propagating modes.

3.3.3 Modal Decomposition

We now represent the input field in region 1 in terms of a basis of orthonormal modes of region 2.

$$H_{x,1}(z = 0) = \sum_{k=1}^{n} c_k H_{x,2}^{(k)}(z = 0), \tag{3.7}$$

where

$$c_k = \frac{1}{2}\int_A \mathbf{E_1} \times \left[\mathbf{H_2^{(k)}}(z = 0)\right]^* \cdot \hat{z} dA. \tag{3.8}$$

As alluded to in Section 2.4.7, an unconjugated version of (3.8) would be used for very lossy modes, but since we are in the low-loss regime the conjugated version works to a good approximation.

It should be checked that the basis is complete enough to capture all of the input mode.

$$\sum_{k=1}^{n} |c_k|^2 = 1. \tag{3.9}$$

A minimum of 0.99 was set, but usually the value is within 1±0.001. In lossless and low-loss modes $|c_k|^2$ can be thought of as the fractional power or percentage power in the k-th mode.[62] In the low-loss modes this is the power distribution at $z = 0$ only, since as the basis modes propagate they in general have different power attenuation.

This basis is propagated according to the discrete propagation constants β_k of each mode a length $z = L$, and then coupled into the bound mode of region 3, which is set to be the same as in region 1. We then find a coupling coefficient for the whole system

$$\begin{aligned} c_L &= \frac{1}{2}\int_A \mathbf{E_3} \times \left[\sum_{k=1}^{n} c_k \mathbf{H_2^{(k)}}(z=L)\right]^* \cdot \hat{z} dA \\ &= \frac{1}{2}\int_A \mathbf{E_3} \times \left[\sum_{k=1}^{n} c_k e^{i\beta_k L} \mathbf{H_2^{(k)}}(z=0)\right]^* \cdot \hat{z} dA \\ &= \sum_{k=1}^{n} \left[\left(c_k e^{i\beta_k L}\right)^* \frac{1}{2}\int_A \mathbf{E_3} \times \left[\mathbf{H_2^{(k)}}(z=0)\right]^* \cdot \hat{z} dA\right] \\ &= \sum_{k=1}^{n} |c_k|^2 \left(e^{i\beta_k L}\right)^*, \end{aligned} \tag{3.10}$$

where the last step was found using $E_3 = E_1$ and (3.8). The $|c_k|^2$ factor in the last line looks similar to the fractional power described above, but is here caused by the input and output modes being the same and thus having the same coupling coefficient c_k to the k-th mode.

The insertion loss (IL) for the whole system is now found from

$$IL = 20\log_{10}\left(|c_L|\right). \tag{3.11}$$

Since a minimum of 99% of the energy is found there is an approximate error in the insertion loss of $\pm 20\log_{10}(0.99) = \pm 0.09dB$.

3.3.4 Validation of Normal Mode Analysis Method

Although the method is accepted and well known, we validate it here to highlight the importance of the outer boundary position and the accuracy of the method and also to validate this implementation of method. We do this by computing the free-spreading of a Gaussian beam. Analytically this is computed using a common formula and the derivation is found in many references.[41,67]

Using the notation from Saleh and Teich,[67] the $\mathbf{H_x}$-field of a two dimensional (2D) Gaussian beam varying spatially in x and y and propagating along z has the form

$$H_x(x,y,z) = H_0\frac{W_0}{W(z)}\exp\left[\frac{-\left(x^2+y^2\right)}{W^2(z)}\right]\exp\left[-ikz - \frac{ik\left(x^2+y^2\right)}{2R(z)} + i\zeta(z)\right], \tag{3.12}$$

where

$$\begin{aligned} W(z) &= W_0\left[1+\left(\frac{z}{z_0}\right)^2\right]^{1/2} \\ R(z) &= z\left[1+\left(\frac{z_0}{z}\right)^2\right] \\ \zeta(z) &= \tan^{-1}\left(z/z_0\right) \\ W_0 &= \left(\frac{\lambda z_0}{\pi}\right)^{1/2}. \end{aligned}$$

The last exponential factor of (3.12) contains the phase with the first term represent-

ing a plane wave, the second term accounting for wavefront bending and the third term adding phase retardation.[67] z_0 is known as the Rayleigh range. The spot size at $z = 0$ is $2W_0$ and H_0 is a normalization constant. Also $\lambda = \lambda_0/n$, where n is the background index, and $k = 2\pi/\lambda$.

The one dimensional (1D) Gaussian beam form keeping in line with our slab model can be found from (3.12) by removing the x dependency.

$$H_x(y,z) = \frac{W_0}{W(z)} \exp\left[\frac{-y^2}{W^2(z)}\right] \exp\left[-ikz - \frac{iky^2}{2R(z)} + i\zeta(z)\right]. \tag{3.13}$$

For both cases we will need the y component of the electric field which can be found from (2.50)

$$E_y(y,z) = -\sqrt{\frac{\mu_0}{\epsilon_r\epsilon_0}} H_x(y,z). \tag{3.14}$$

$\beta = k_0\sqrt{\epsilon_r}$ with the Gaussian beam spreading into a medium with relative permittivity ϵ_r.

The insertion loss (IL) of the Gaussian beam coupled into its original state after travelling a distance L is

$$IL_{Gauss} = 20\log_{10}(|c|), \tag{3.15}$$

where c is

$$c = \frac{1}{2}\int_A \mathbf{E}(z=0) \times \mathbf{H}^*(z) \cdot \hat{z} dA. \tag{3.16}$$

Modes are normalized according to (2.39).

Figure 3.11 shows the analytical curve as obtained from (3.15) and values computed using the normal mode analysis approach for three different separations of the outer boundaries. The input mode is a 1D, 10μm spot size Gaussian beam and the medium is SiO_2 (n=1.444) at λ_0=1.55μm. In the normal mode analysis the Gaussian beam is centered in the middle of the thick slab of SiO_2 (with the free space boundaries above

and below the slab).

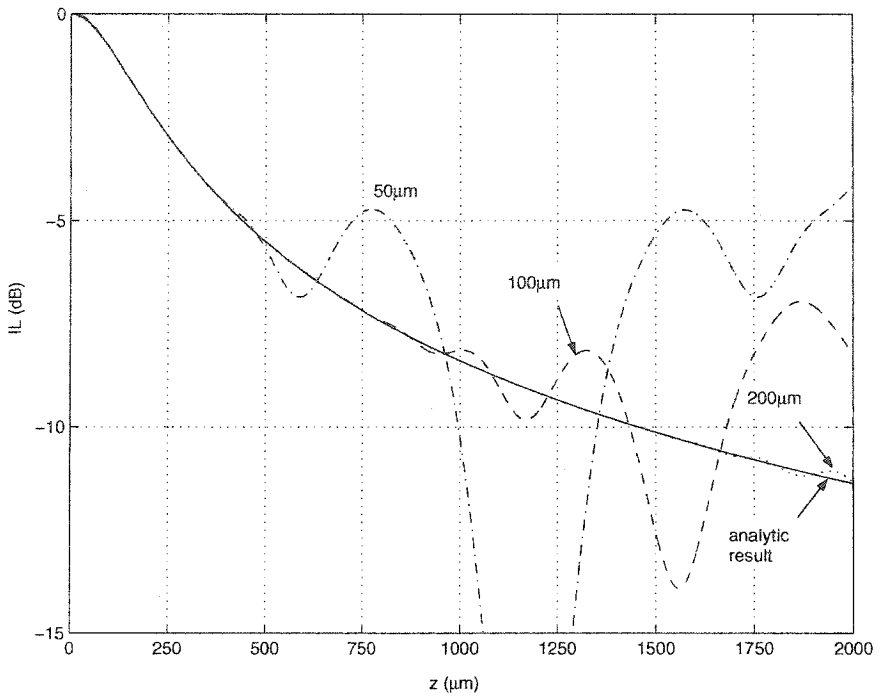

Figure 3.11: Spreading of a 10μm Gaussian beam in silica after a distance z. Analytical spreading (solid curve) vs. normal mode analysis for three sizes of boundary separation (50, 100 and $200\mu m$).

The method predicts the correct results up until a certain propagation length, depending on the outer boundary separation, and better results are obtained for larger separations. For a 200μm separation accurate results for the beam spreading are obtained up to about 1700μm. It is very apparent when the boundary positions begin to influence the results as the curves begin oscillating rapidly.

Note that the spreading IL loss for a 2D Gaussian beam is twice that of the 1D case, as spreading is then occurring in the x *and* y directions and since the Gaussian spreads the same in all directions. An SPP excited on a slab in 2D will not necessarily have twice the spreading loss of the values given here however, and we should keep in mind that the

results in this chapter are only for 1D confinement and a comparison with experiment will need to account for the distribution of the source with an additional factor. This can probably be done to a good approximation by assuming a given source profile in x, e.g. Gaussian, and determining analytically the additional spreading factor in the x-direction.

3.4 Radiation of Surface Plasmon-Polaritons

3.4.1 Introduction

Using the method described in the last section, radiation of SPPs in varying degrees of asymmetry and cutoff are calculated, for a range of propagation distances.

The loss in the waveguide in the symmetric state must be subtracted from the amount of IL added with a change in index to accurately determine the additional loss, or dynamic range due to the asymmetry. Doing this, the same value of dynamic range was found using a lossy and lossless version of the same SPP waveguide, within the basis error stated above. If the bound mode is not entirely cutoff, a portion of the input will continue to propagate without radiating over the length of the asymmetric region at a similar attenuation to the symmetric case. In this case the loss due just to the asymmetry is approximately the coupling loss into the asymmetric bound mode. We have used a lossless approximation for all guides, implemented by assuming a purely real dielectric constant for the metal. From a design standpoint, if two waveguides have the same dynamic range for a given asymmetry, it would be best to choose the one with the lowest loss for no asymmetry, in order to have the lower on-state IL.

All index of refraction changes are done by adding an amount δn_{yy} to n_{yy} of the cover region immediately above the metal for the anisotropic media or an analogous amount

δn to n of the cover for isotropic media. The results will be essentially the same for subtracting this amount because globally the refractive index is about the same in both cases (lowered only by an amount $\delta n_{yy} \ll n_{yy}$). We will see more of a difference when comparing the two materials, SiO_2 and LN.

The asymmetries used may not all be achievable in SiO_2 and LN, but the results are not meant to be entirely applicable only to these materials. They were chosen in order to first of all, use two common materials, and second of all to cover a range of refractive indexes. Many polymers and crystals have similar refractive indexes, while having vastly different electrooptic or thermooptic coefficients.[41]

3.4.2 Estimation of Boundary Separation

Approximate outer boundary separations can be determined using a ray tracing method where the total internal reflection condition, used in reverse (for the index values of the clads), determines the angle of refraction away from the surface of the metal. If a ray reaches the outer boundary it will be reflected back towards the core region and unreliable results will be found. This method gives a good estimate of the size required, but the method shown for the Gaussian beam spreading example with successively larger boundary separations was used to determine it more accurately. All results in this section were obtained using 500μm of cladding above and below the metal layer for a total outer boundary separation of 1000μm. The thickness of the lower clad is not as critical since the radiation is preferentially directed upwards, however a portion of the uncoupled light spreads into the lower clad. 500μm was found to be a good approximation of the radiation continuum in each clad, for the asymmetries considered and the propagation lengths used.

3.4.3 Radiation in Silica

Looking at the case of spreading in SiO_2, in the asymmetric region we use $n_s = 1.444$, $n_c = 1.444 + \delta n$, $\lambda_0 = 1.55\mu m$ and the lossless Au is approximated with $\epsilon_m = -131.95$. The input and output regions are identical with $\delta n = 0$ and the s_b mode of the symmetric region is the input and output field. Results for four metal thickness are given in Figure 3.12.

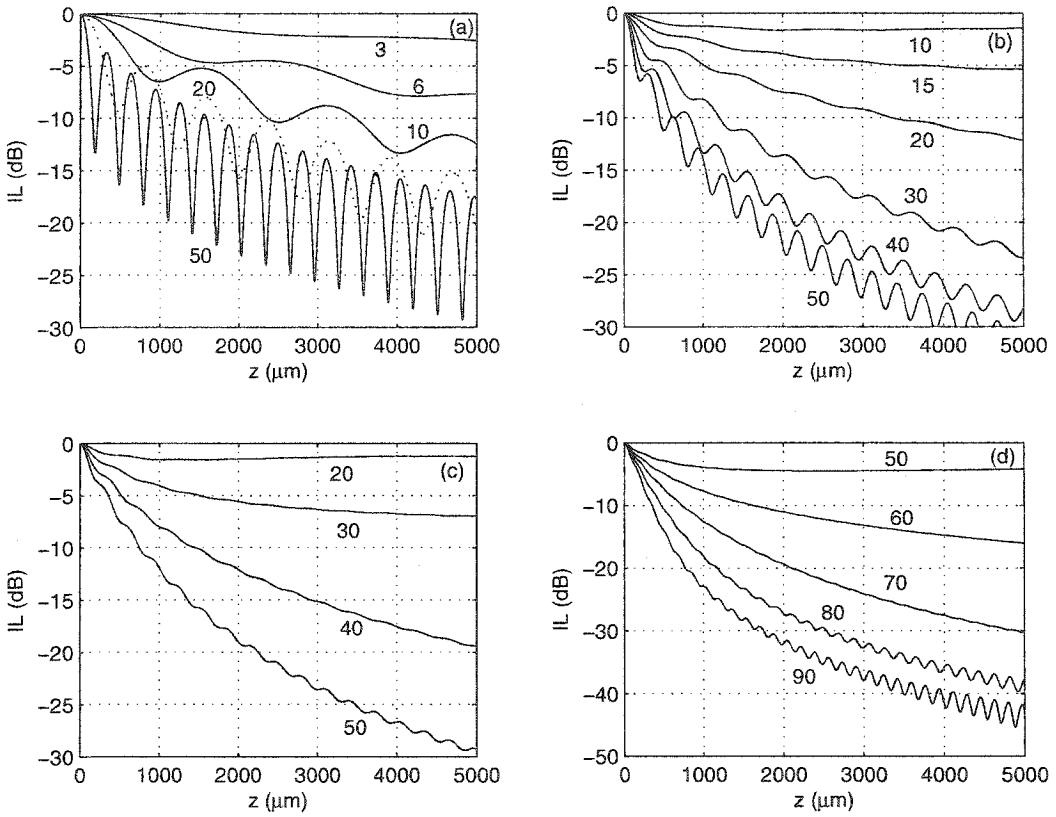

Figure 3.12: Radiation loss curves for SiO_2. Curves labelled with δn with a factor of 10^{-4} suppressed. Metal thickness and s_b cutoff point: (a) 4nm, $\delta n^{CO} = 4 \times 10^{-4}$; (b) 9nm, $\delta n^{CO} = 18 \times 10^{-4}$; (c) 13nm, $\delta n^{CO} = 34 \times 10^{-4}$; (d) 19nm, $\delta n^{CO} = 59 \times 10^{-4}$.

Ripples are seen in the curves for large δn. This is interpreted as interference of the z-directed plane waves in the substrate with those in the cover. The radiation modes

are excited over a range of $n_{eff} = \beta^R/k_0$ values as seen in Figure 3.13, but are typically centered near the refractive index values in each clad. The plane wave in the cover will have an effective index $n_{eff,c} \approx n + \delta n$ and the plane wave in the substrate will have $n_{eff,s} \approx n$. Constructive interference will occur when there is a phase difference of 2π between the waves. This leads to a length for the first constructive interference point of

$$l \approx \lambda_0/(n_{eff,c} - n_{eff,s}) = \lambda_0/\delta n. \tag{3.17}$$

For plot (a) in Figure 3.12, the $\delta n = 50 \times 10^{-4}$ curve has its first peak at 318μm with the predicted peak at $l = 1.55/50 \times 10^{-4} = 310\mu m$ while the $\delta n = 20 \times 10^{-4}$ curve has its first peak at 795μm with the predicted peak at $l = 1.55/20 \times 10^{-4} = 775\mu m$. The small difference in values is because a large number of plane waves with a range of individual n_{eff} values are excited to varying degrees in each region. Ripple is more pronounced for thinner films and for larger δn. This suggests that the interference is suppressed by the metal film, as seen in plots (a)-(d) with the $\delta n = 50 \times 10^{-4}$ curves. This could be because more of the mode energy is in the metal for thicker films so the interference of the energy in the clads is less pronounced. This ripple should be experimentally realizable and for thin metal with large asymmetries any modulating device based on this should account for a possible 12dB swing in radiation IL caused by choosing the wrong propagation length.

It is now useful to recall the bound mode cutoff δn values calculated earlier for each of the curves in Figure 3.12: (a) 4×10^{-4}, (b) 18×10^{-4}, (c) 34×10^{-4}, (d) 59×10^{-4}. In all cases the IL at the cutoff δn point is no greater than -15dB, even after 5mm. For case (a) the IL is only about -5dB after 5mm. This case was for the largest mode so it is expected to spread less rapidly than the smaller modes. Along this line, the slope of the cutoff curves decreases for larger z since as the mode expands it spreads less rapidly. The

low loss after 5mm means that although the bound mode is cutoff the radiative modes are not always radiating very rapidly and a considerable portion of the energy can still be coupled out of the system. An asymmetry significantly larger than the cutoff value is necessary to reach the -25 or -30dB IL level.

In each case, asymmetries that do not cut off the bound mode are also analyzed. For these curves there is a transient region near the source after which the IL oscillates around the coupling loss into the asymmetric bound mode eg. the $\delta n = 10 \times 10^{-4}$ curve in (b).

In the other extreme where the bound mode is well past cutoff ($\delta n = 50 \times 10^{-4}$ in (a)-(c)) similar IL of about -30dB after 5mm is found in each case. This means that changes to the metal between 4nm and 13nm only slightly perturb the radiation for this asymmetry. Figure 3.13 shows the fractional power coupled into each mode in

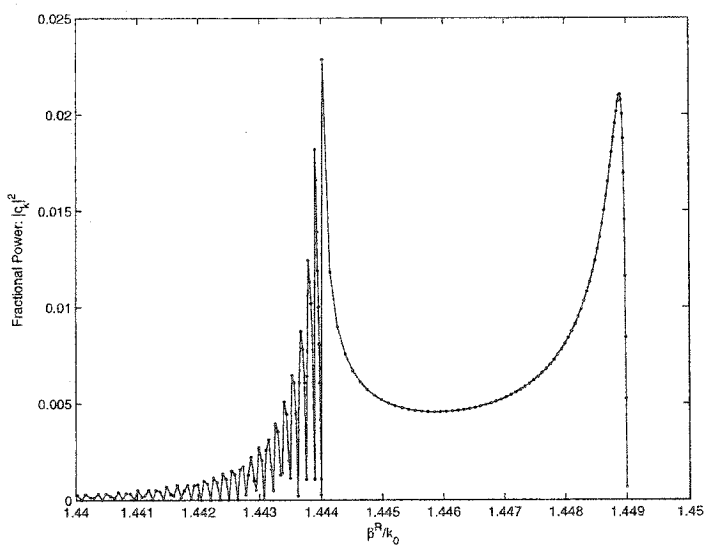

Figure 3.13: Fractional power in each mode in the basis for cutoff in SiO_2, with t=4nm and $\delta n = 50 \times 10^{-4}$.

the basis for the case t=4nm, $\delta n = 50 \times 10^{-4}$. We see several distinct regions in this

curve. As discussed earlier, bound SPPs must have β^R/k_0 greater than the larger of the two cladding regions, so no bound SPP modes exist for this case. The curve is smooth between 1.444 and 1.449 since normal modes in this region are bound to the high index upper clad and are exponentially decaying into the lower clad, so they always overlap to some extent with the input mode which has an exponential tail into both clads. Below 1.444, the normal modes are oscillatory in both clads and do not necessarily overlap well with the input. The asymmetric bound mode a_b has a much larger n_{eff} and was never needed to complete the basis. Its overlap with the s_b mode in a symmetric structure is identically zero and this does not change considerably for asymmetric structures.

The modes necessary to form a complete basis are typically contained in a small region of the complex β-space. The equation for β is then plotted on a very fine grid over this region and a simple algorithm finds the minima of the equation. These minima are then refined using Muller's method as before. A tolerance of 10^{-8} away from a minimum is set for refined eigenvalues.

The number of modes needed depended greatly on the asymmetry of the structure with small asymmetries only requiring the bound mode plus a few radiation modes to capture the input, while large asymmetries may require 200-300 radiation modes. Good basis orthogonality was found with cross coupling power coefficients between any two modes in the basis typically below 10^{-7}.

Figure 3.14 gives a visualization of the H-field intensity, $|H_x|^2$, spreading into the top clad ($y > 500\mu m$) for t=4nm, $\delta n = 50 \times 10^{-4}$. For $y < 500\mu m$ a portion of the light slowly spreads into the lower clad. Near $y = 500\mu m$ the complicated interference behavior manifesting itself as ripples in the cutoff curves is seen. For $y > 500\mu m$, light is leaving the core at roughly the total internal reflection angle, throughout the length. At $z \approx 5000\mu m$ the intensity is very oscillatory in the top clad, characteristic of a leaky mode.

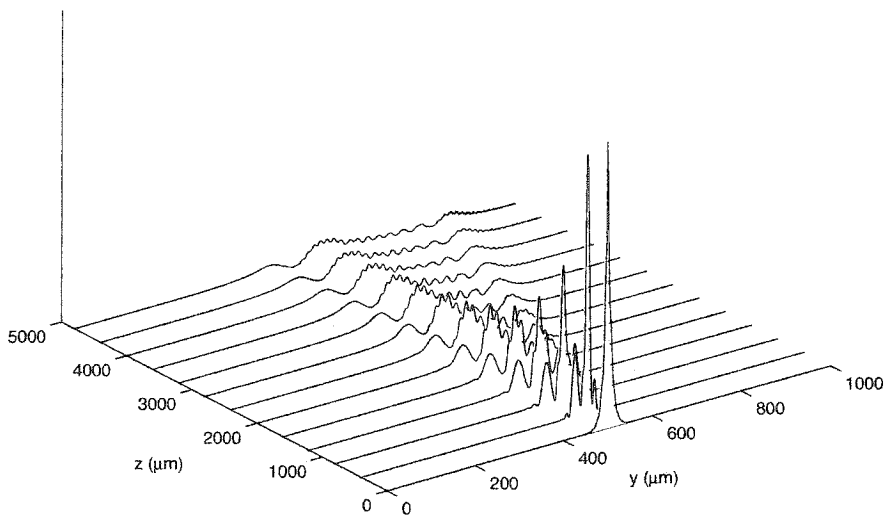

Figure 3.14: Evolution of the H-field intensity $|H_x|^2$ along z in SiO_2, with t=4nm and $\delta n = 50 \times 10^{-4}$.

3.4.4 Radiation in Lithium Niobate

To compare with silica, the spreading in a medium with a much different index of refraction, namely LN, is computed. This will give an idea of the variability of changing material platforms. The next section compares the materials directly.

For LN we use $n_{zz} = 2.2125$ in all substrate and cladding regions, thus ignoring any variation in this parameter, seen before to have minimal effect. In the substrate $n_{s,yy} = 2.1377$ and in the cover $n_{c,yy} = 2.1377 + \delta n_{yy}$. Au keeps the same lossless value as before. In the input and output regions $\delta n_{yy} = 0$, with the input and output mode again being the s_b mode of these symmetric regions.

Figure 3.15 shows the results for LN. A similar ripple pattern is found again for large asymmetry and thinner guides.

The cutoff points δn_{yy}^{CO} for the s_b modes in LN for these cases were (a) 4×10^{-4}, (b) 20×10^{-4}, (c) 39×10^{-4}, (d) 73×10^{-4}. In case (a) there is only about -5dB of IL at 5mm so the mode energy is not significantly attenuated.

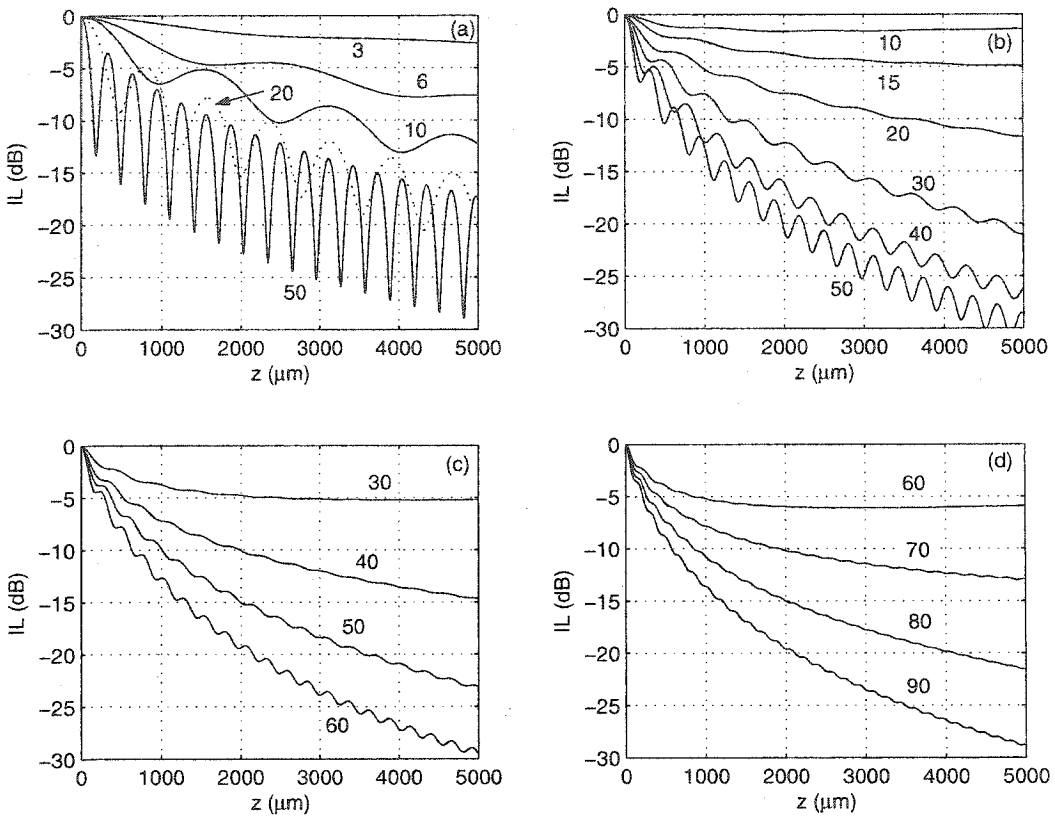

Figure 3.15: Radiation loss curves for LN. Curves labelled with δn with a factor of 10^{-4} suppressed. Metal thickness and s_b cutoff point: (a) 2.1nm, $\delta n^{CO} = 4 \times 10^{-4}$; (b) 4.7nm, $\delta n^{CO} = 20 \times 10^{-4}$; (c) 6.7nm, $\delta n^{CO} = 39 \times 10^{-4}$; (d) 9.5nm, $\delta n^{CO} = 73 \times 10^{-4}$.

3.4.5 Comparison of Radiation in Lithium Niobate and Silica

It is difficult to find comparable waveguides in LN and SiO_2. Generally if one parameter such as loss, confinement, or spot size is the same, the others are different. The waveguides analyzed in the above two sections were chosen with the idea of comparing waveguides with similar $\delta n_{eff} = n_{eff} - n_{yy}$. Plots (a)-(d) in Figure 3.12 for SiO_2 have the same δn_{eff} as the corresponding plots in Figure 3.15 for LN. Figure 3.9 showed that LN and SiO_2 s_b modes have similar cutoff points for small δn_{eff}, but that the cutoff points diverge for larger δn_{eff}, with the SiO_2 modes cuting off easier. The same is found when looking at the radiation loss curves. Plot (a) for both is very similar, while in plot (d) the IL is more than 10dB higher after 5mm in SiO_2.

Comparison of modes with $10\mu m$ spot size is shown in plot (a) of Figure 3.16. The solid curves are for a t=11.8nm thick Au film in SiO_2 and the dotted curves are for a t=4.7nm thick Au film in LN. For low δn (10-30$\times 10^{-4}$) the LN modes have a much deeper cutoff, but for 50×10^{-4}, the cutoff curves have nearly converged. The higher δn_{eff} in SiO_2 seems to make it more difficult to radiate this mode away from the core.

Plot (b) of Figure 3.16 compares two of the more similar waveguides, the t=6.7nm Au guide in LN (dotted curves) and the t=13nm Au guide in SiO_2 (solid curves). Each have loss of about 0.3 dB/mm and $\delta n_{eff} = 10 \times 10^{-4}$. The IL curves show larger attenuation for the SiO_2 case for the three asymmetries, which may be expected since the SiO_2 mode will be more cutoff than the LN mode when they both have the same δn_{eff}.

In the next chapter we repeat this normal mode analysis for finite width SPP waveguides.

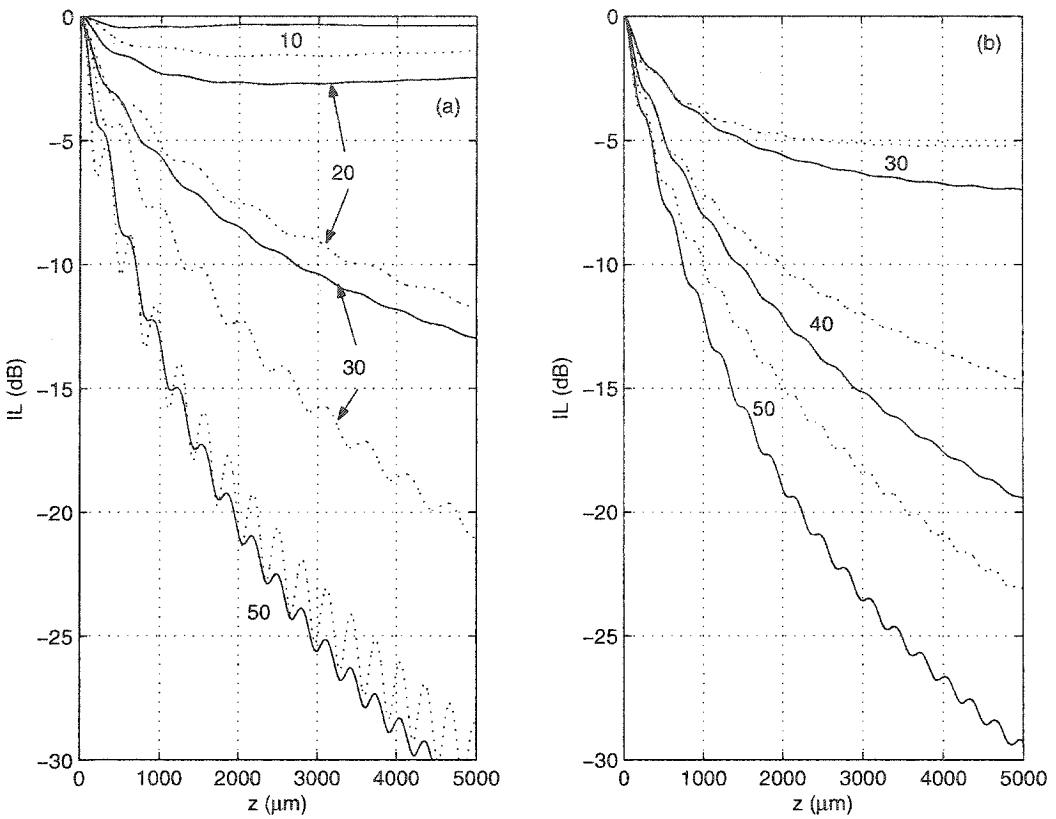

Figure 3.16: Comparison of radiation loss for Au waveguides in SiO_2 (solid curves) and LN (dotted curves) for (a) modes with $10\mu m$ spot size and (b) modes with similar attenuation (0.3dB/mm) and $\delta n_{eff} = 10 \times 10^{-4}$. Curves labelled with index asymmetry δn with a factor of 10^{-4} suppressed.

Chapter 4

Mode Cutoff in Finite Width SPP Waveguides

4.1 Introduction

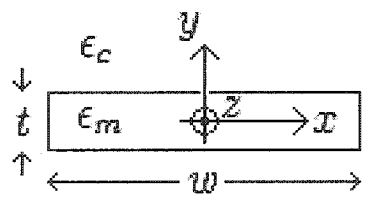

Figure 4.1: Geometry for finite width SPP waveguide. Metal waveguide ϵ_m surrounded by dielectric ϵ_c.

Finite width SPP waveguides are similar to the three layer slab SPP waveguides studied in chapters 2 and 3 only with the extent of the metal limited to a fixed width along the x-axis, with cladding filling the region all around the metal core as shown in Figure 4.1. The modes of these structures were theoretically predicted recently by Berini[27,28] and were subsequently experimentally verified by Charbonneau *et al.*[29] for Au films embedded in SiO_2. More recently Nikolajsen *et al.*[81] have demonstrated these

modes for Au films in polymers.

These waveguides confine the mode both vertically and horizontally as opposed to the slab modes which only provide vertical confinement. The main long range (low-loss) mode is the ss_b^0 mode.[28] It is symmetric along the horizontal and vertical directions allowing it to couple well to optical fiber for certain geometries. This suggests the use of this mode in integrated optics and it will be the mode we focus our attention on in this chapter.

4.2 Modal Solutions for 2D Hybrid-Modes

4.2.1 Formulation

In finite width waveguides we must now include all six electric and magnetic field components. These modes are termed hybrid-modes since they are no longer TE or TM but a combination of the two.

The fields take the general form

$$\mathbf{E}(x,y,z,t) = \mathbf{E_0}(x,y)e^{i(\beta z-\omega t)}, \tag{4.1}$$

$$\mathbf{H}(x,y,z,t) = \mathbf{H_0}(x,y)e^{i(\beta z-\omega t)}, \tag{4.2}$$

where

$$\mathbf{E_0} = E_x\hat{i} + E_y\hat{j} + E_z\hat{k}, \tag{4.3}$$

$$\mathbf{H_0} = H_x\hat{i} + H_y\hat{j} + H_z\hat{k}. \tag{4.4}$$

Removing $\mathbf{E}$ from Maxwell's equations leads to the wave equation for the $\mathbf{H}$-fields

$$\nabla \times \left(\epsilon_{\mathrm{r}}^{-1} \nabla \times \mathbf{H}\right) = k_0^2 \mathbf{H}. \tag{4.5}$$

ϵ_{r} is a function of x and y but is independent of z, the propagation direction. It is assumed to be a diagonal tensor as before.

Putting the above fields into Maxwell's equations leads to

$$\begin{aligned}
\frac{\partial E_z}{\partial y} &= i\beta E_y + i\omega\mu_0 H_x && (4.6)\\
\frac{\partial E_z}{\partial x} &= i\beta E_x - i\omega\mu_0 H_y && (4.7)\\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= i\omega\mu_0 H_z && (4.8)\\
\frac{\partial H_z}{\partial y} &= i\beta H_y - i\omega\epsilon_0\epsilon_{xx} E_x && (4.9)\\
\frac{\partial H_z}{\partial x} &= i\beta H_x + i\omega\epsilon_0\epsilon_{yy} E_y && (4.10)\\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= -i\omega\epsilon_0\epsilon_{zz} E_z && (4.11)
\end{aligned}$$

In addition to satisfying the above equations, the fields are also constrained by the boundary conditions arising from Maxwell's equations[45] as used in the slab model.

If we assume ϵ_{r} is a scalar constant (isotropic) rather than a tensor, a new wave equation in terms of the two transverse components of the $\mathbf{H}$-field is found from (4.5) to be

$$\nabla_t \times \left(\epsilon_r^{-1} \nabla_t \times \mathbf{H}_t\right) - \epsilon_r^{-1} \nabla_t \left(\nabla_t \cdot \mathbf{H}_t\right) - \left(k_0^2 - \beta^2 \epsilon_r^{-1}\right) \mathbf{H}_t = \mathbf{0}, \tag{4.12}$$

where the transverse nabla operator is $\nabla_t = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y}$ and $\mathbf{H}_t = \left(H_x\hat{i} + H_y\hat{j}\right) e^{i(\beta z - \omega t)}$.

This is an eigenvalue equation for $\lambda = -\beta^2$ which is solved using the commercially available software package FEMLAB.[82] FEMLAB implements the finite element method (FEM)[83] specifically to solve optical waveguide problems and is an accepted method in

optics.[56] In the next section we compare the results from FEMLAB with other methods and test the convergence of the software for SPP waveguides.

4.2.2 Convergence

When using a mode solver for a waveguide it was perhaps not intended for, such as ultrathin, high aspect ratio SPP waveguides, it is wise to test its convergence and compare it with other methods.

We begin by checking the appropriateness of the initial adaptive mesh generated by FEMLAB. There are two regions of interest here: the mesh across the thickness of the metal, and the mesh near the corners of the metal.

We begin with the thickness of the metal by modelling a simple three-layer slab using the same hybrid-mode package that is used for the finite width simulations. In general the initial mesh in the metal has two triangles between the interfaces for metal thickness between 10nm and 30nm.

Propagation constants were generated using FEMLAB and the transfer matrix model (discussed earlier) for Au slab waveguides of thickness between 4nm and 19nm in SiO_2 at λ_0=1.55μm using the same material properties as before, recreating Table 3.1. The values of the real part of the propagation constant differed by at worst 3×10^{-5}% and at best by 2×10^{-8}%. The imaginary parts of β were different by at worst 0.9% and at best 8×10^{-4}%. We then conclude that the default mesh generated by FEMLAB inside the metal away from the corners is sufficient.

The question then remains if the mesh is adequate near the corners since some of the field components can be very localized there.[28] Triangles in the default mesh may be refined so that there is a greater density of triangles in a sensitive area than FEMLAB automatically generates. For this test we refine the density of triangles in the immediate

vicinity of the corners, along the height of the metal waveguide, as shown in the cross section in Figure 4.2. Refinement of these triangles spreads to other triangles in close proximity. Away from the corners we see the default mesh density in the metal of two triangles through the thickness, determined above to be sufficient in this region. (Note that for symmetric modes the optical properties of the substrate and cover regions are identical, as shown in Figure 4.1, and the regions are only distinguished in this model in preparation for cutoff simulations later.)

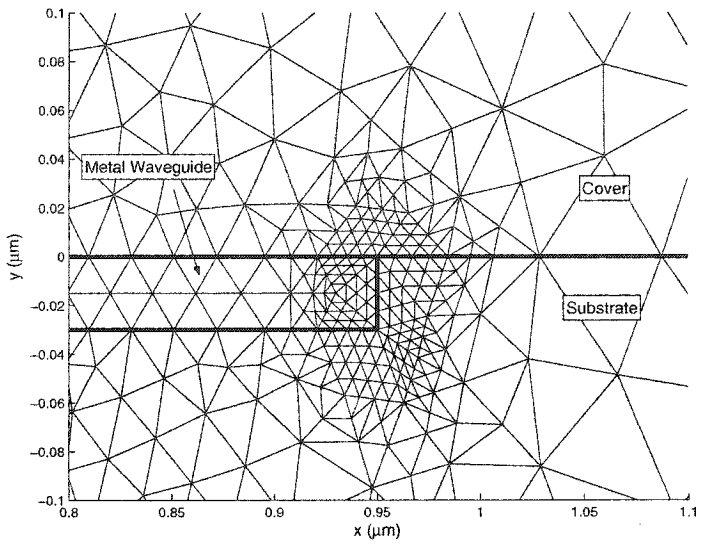

Figure 4.2: Example of finite element mesh refinement near the corners of a metal waveguide.

We test for convergence based on the height of the individual triangles along the vertical wall of the metal. Table 4.1 shows the convergence of the propagation constants for a thin, wide waveguide and a thicker, narrower waveguide. Generally, the mesh can be refined four times, as shown, after which the limits of the computers memory are reached. This is typically around 15,000 triangles, but can sometimes be more. The values listed for the triangle height of zero are obtained using Richardson's extrapolation

algorithm.[84] This method can be used for smoothly converging functions where the independent parameter is successively halved, as we have with the height of the triangles. For the 15nm thick waveguide, the percentage differences between the last computed value and the extrapolated values are $10^{-4}\%$ for n_{eff} and 1% for MPA. For the 30nm thick waveguide the differences are $8 \times 10^{-4}\%$ for n_{eff} and 0.2% for MPA. From this we conclude that refining the mesh density four times in the immediate vicinity of the corners gives sufficiently converged values.

One last check of the accuracy of the model is done by recreating some reported values for finite width SPP waveguides[28] which were generated with the method of lines (MoL), a different numerical technique. We recreate the values from Figure 11 of that work for the ss_b^0 mode of a 1μm wide silver film ($\epsilon_m = -19 + i0.53$) surrounded by dielectric ($\epsilon_c = 4$) at a free space wavelength of $\lambda_0 = 0.633\mu m$. The values are given in Tables 4.2 and 4.3. The values for n_{eff} agree well and the attenuation values (β^I/k_0) differ by less than 3%, similar to what we had found above for the difference between the converging and extrapolated values. FEMLAB and the MoL model seem to generate

Waveguide	Triangle Height (nm)	n_{eff}	MPA (dB/mm)
t=15nm, w=8.52μm	15	1.4444893	0.3675
	7.5	1.4444962	0.3510
	3.75	1.4445005	0.3378
	1.875	1.4445029	0.3306
	0.9375	1.4445044	0.3265
	0	1.4445061	0.3220
t=30nm, w=1.9μm	30	1.4440360	0.3109
	15	1.4440603	0.3795
	7.5	1.4440806	0.4084
	3.75	1.4440945	0.4165
	1.875	1.4441035	0.4179
	0	1.4441145	0.4170

Table 4.1: Convergence of FEMLAB results with increasing mesh density near corners for a Au strip embedded in SiO_2.

comparable results since the error estimated in the latter was between 1% and 6%.

The model implemented in FEMLAB generates results in good agreement with two well known methods (the transfer matrix and the method of lines) and so should produce reliable design data for new waveguides as well.

Thickness (nm)	This model	Referenced work	% difference
8	2.0029098	2.004133	0.06
10	2.0062267	2.00729510	0.05
15	2.01949	2.02008816	0.03
20	2.038109	2.03857848	0.02

Table 4.2: Comparison of FEMLAB model with referenced MoL model: n_{eff}.

Thickness (nm)	This model	Referenced work	% difference
8	5.715×10^{-5}	5.738×10^{-5}	0.4
10	1.102×10^{-4}	1.073×10^{-4}	2.7
15	3.161×10^{-4}	3.184×10^{-4}	0.7
20	6.336×10^{-4}	6.486×10^{-4}	2.3

Table 4.3: Comparison of FEMLAB model with referenced MoL model: β^I/k_0.

4.2.3 Waveguide Results for Silica and Lithium Niobate

We now present new results for the ss_b^0 mode of a Au waveguide embedded in SiO_2 with $\epsilon_c = 1.444^2$, $\epsilon_m = -131.95 + i12.65$, and $\lambda_0 = 1.55\mu m$.

Figure 4.3 shows the effective index, the attenuation and the mode spot size. The curves in (a) and (b) show the propagation constant tending to that of a slab waveguide as the width increases. These modes are y-polarized, with only a small electric field component along x. As the width increases the E_x component diminishes to zero (the TM limit).

With the width of the waveguide available as a parameter, it is now possible to have modes for different metal thickness with the same β^R/k_0. This will again be an important point of comparison in the radiation curves. While this parameter may be the same, the attenuation and spot size are generally different.

MPA lower than 0.1dB/mm appears possible for spot sizes smaller than $15\mu m$, for a range of realizable thicknesses. The 15nm curve shows less sensitivity to width changes at the 0.1dB/mm point than does the 30nm curve.

The spot size is found numerically from mode plots and is measured along horizontal (solid curve) and vertical (dotted curve) cuts through the center of the waveguide. Along the vertical cut the spot size converges to the slab value as changes to the fields by increasing the width no longer affect the field near the center of the guide.

In the horizontal, starting from a small width the mode is weakly confined, so widening the guide causes a rise in the confinement, and a reduction in the spot size. As the width continues to increase, the spot size must eventually stop decreasing and meet the size of the guide, as the spot size can never be less than the width of the metal. For large widths the spot size increase almost linearly with width.

When $\beta^R/k_0 \approx 1.444$ the mode becomes very large and the $1/e$ point is far away

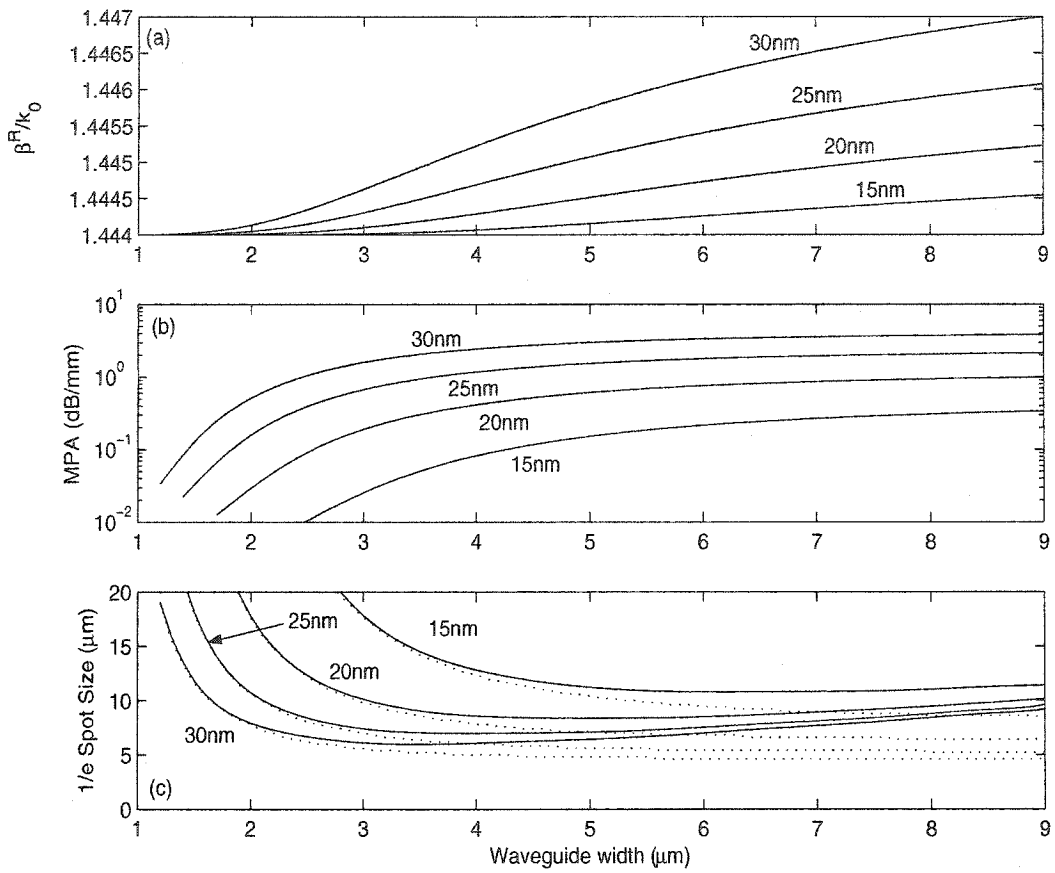

Figure 4.3: Propagation constants and spot size for the ss_b^0 SPP of a Au strip embedded in silica vs. the strip width. Curves labelled with Au thickness. In (c) the solid curve represents the spot size along x and the dotted curve represents the spot size along y, both from slices through the center of the waveguide. Horizontal axis is the same for all three plots.

from the core. The actual waveguide core shape is then less important and so the spot size along the horizontal and vertical directions converge.

A sample field is shown in Figure 4.4. In (a) the real part of the H_x field for a t=20nm, w=4.95μm guide is shown, plotted on a coarse grid. The peak is not well resolved on this scale, but the circularity of the mode away from the peak is very visible. In (b), cuts along x (solid) and y (dotted) through approximately the center of the waveguide on a finer grid better resolve the peak of the mode. Along x the mode has a more Gaussian like shape, while along y it retains the exponentially decaying fields characteristic of a surface mode. Away from the peak at about the $1/e$ point (=0.37) the mode is essentially the same size in each direction. Later, we will look at the cutoff characteristics of this waveguide.

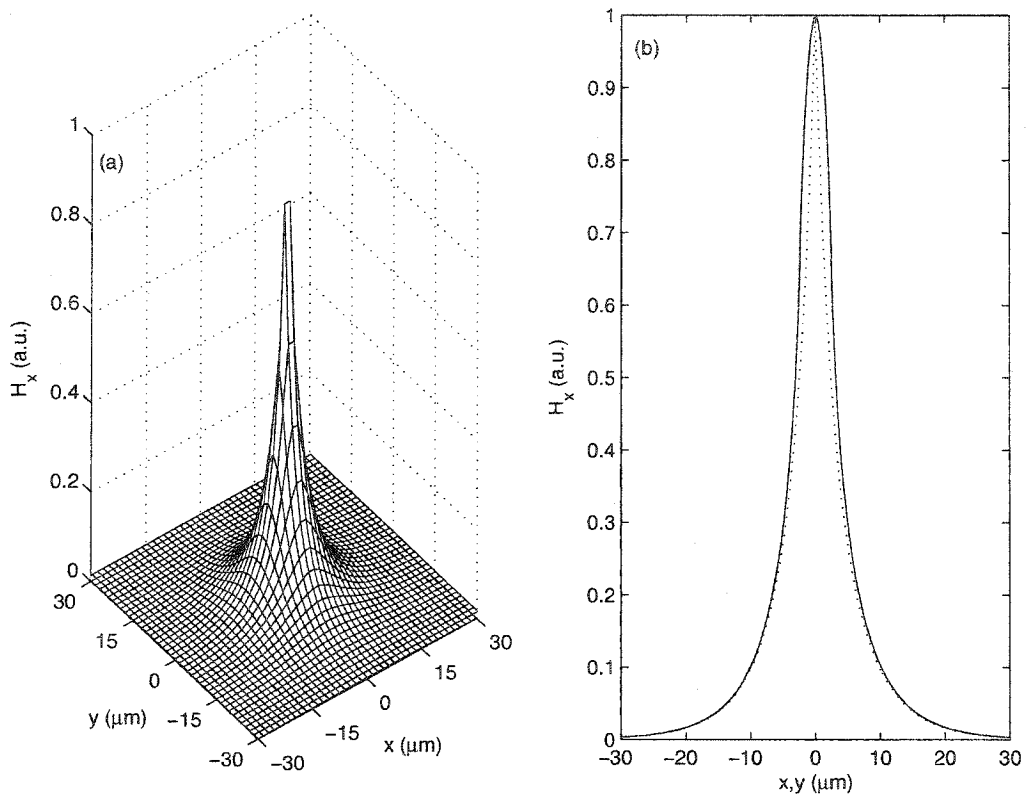

Figure 4.4: Real part of H_x fields for ss_b^0 mode of a Au strip in SiO_2 for t=20nm, w=4.95μm. (b) solid: slice along x, $(y = 0)$; dotted: slice along y, $(x = 0)$.

Similar calculations were performed for a Au strip embedded in Lithium Niobate (LN). Here we treat the LN more simply by assuming it is isotropic with $\epsilon_c = 2.1377^2$. This simplification is not expected to impact the results since changes to the tensor elements other than ϵ_{yy} were seen in chapters 2 and 3 to have little impact on the propagation constants, especially for the lower confinement modes we are interested in.

The curves shown in Figure 4.5 have similar shapes and much of the discussion for the silica case applies here as well. Low loss propagation ($< 0.1dB/mm$) is also possible in LN for modes smaller than $10\mu m$. The 25nm curve was removed from plot (c) for clarity, but it would fall roughly halfway between the 20nm and 30nm curves.

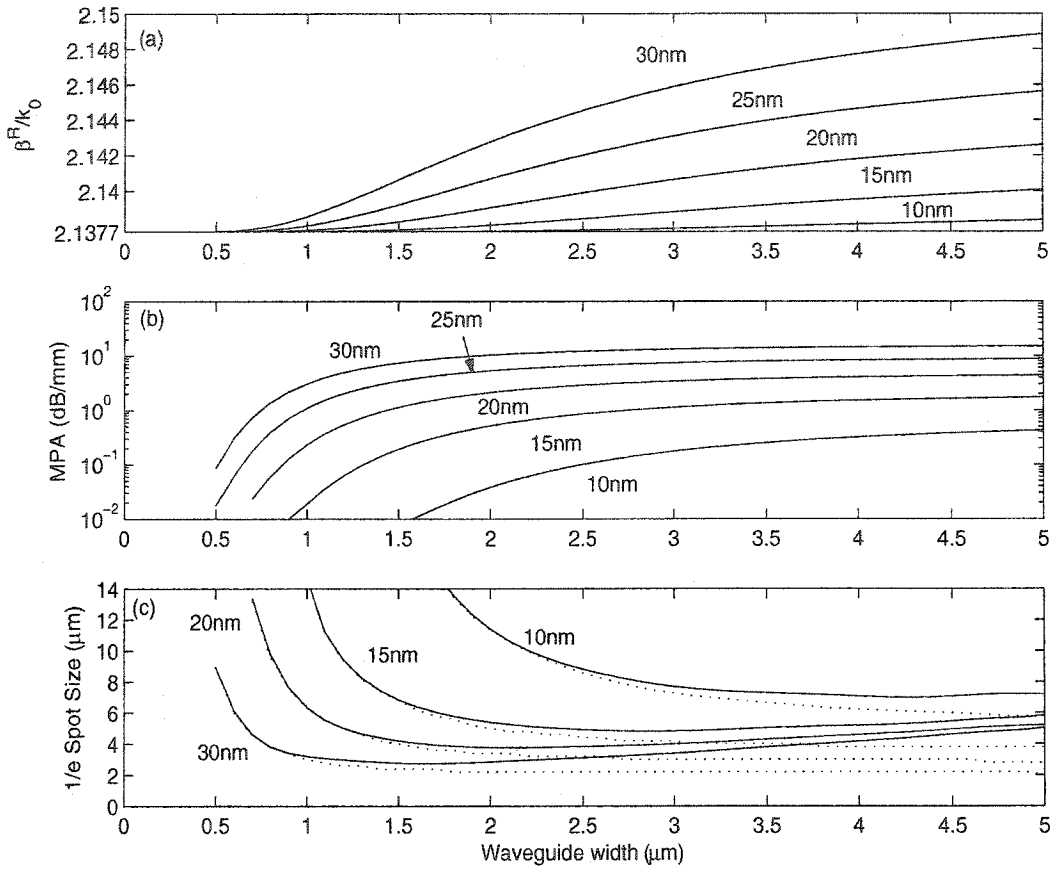

Figure 4.5: Propagation constants and spot size for the ss_b^0 SPP of a Au strip embedded in Lithium Niobate vs. the strip width. Curves labelled with Au thickness. In (c) the solid curve represents the spot size along x and the dotted curve represents the spot size along y, both from slices through the center of the waveguide.

t (nm)	w (μm)	δn_{eff} ($\times 10^{-4}$)	MPA (dB/mm)	Spot size (μm) [x,y]
15	4.45	1.0	0.114	12.0, 11.4
15	5.5	2.0	0.186	10.8, 9.9
15	8.52	5.0	0.327	11.4, 8.7
20	3.0	1.0	0.190	10.2, 10.2
20	3.6	2.0	0.186	9.0, 8.7
20	4.95	5.0	0.603	8.4, 7.5
30	1.9	1.0	0.418	8.4, 8.4
30	2.18	2.0	0.700	7.5, 7.2
30	2.78	5.0	1.362	6.3, 6.0

Table 4.4: Propagation constants and mode spot sizes through the center of the waveguide for ss_b^0 modes of a Au strip in SiO_2.

t (nm)	w (μm)	δn_{eff} ($\times 10^{-4}$)	MPA (dB/mm)	Spot size (μm) [x,y]
15	1.5	1.0	0.193	6.9, 6.75
15	1.73	2.0	0.333	6.0, 5.7
15	2.2	5.0	0.660	5.1, 4.8
20	1.05	1.0	0.299	6.3, 6.3
20	1.19	2.0	0.520	5.4, 5.1
20	1.46	5.0	1.051	4.2, 4.2
30	0.665	1.0	0.561	5.7, 5.4
30	0.745	2.0	1.002	4.8, 4.5
30	0.885	5.0	2.048	3.6, 3.6

Table 4.5: Propagation constants and mode spot sizes through the center of the waveguide for ss_b^0 modes of a Au strip in LN.

Specific results are compiled in Tables 4.4 and 4.5 for each material and waveguide radiation will be compared for these eighteen waveguides based again on $\delta n_{eff} = n_{eff} - n_c$. This time δn_{eff} values of 1, 2 and 5×10^{-4} are used with three waveguides of each found for each material system.

4.3 Cutoff in Finite Width Waveguides

In Chapter 3 it was shown that for waveguide asymmetries larger than a specific amount that the low loss s_b mode was cutoff. An analogous situation occurs for the ss_b^0 mode of a finite width waveguide where the cladding above the plane of the metal is changed relative to the cladding below it.[31] Recall that as the mode was cutoff it spread very far into the high index clad. Cutoff curves are therefore not generated here since the model in FEMLAB has a finite box size and any cutoff curve would inevitably be influenced by the size of the box. Similarly, any device measuring this cutoff phenomenon is also finite sized so the cutoff point in practice is influenced somewhat by the size of the device. While cutoff curves have some importance, we saw in Chapter 3 that the actual cutoff points were not indicative of the loss after a given length since the radiative IL of a mode at cutoff could be anywhere from -5dB to -15dB in silica after 5mm.

4.4 Normal Mode Analysis of Mode Cutoff in Finite Width SPP Waveguides

4.4.1 Introduction

The normal mode analysis follows the same theory as in chapter 3 and the waveguide structure is shown in cross section in Figure 4.6. For this case the entire cladding around the metal is surrounded by a free space outer region.

Two approximations are made for this analysis. The lossless material approximation used and validated in Chapter 3 for the radiation IL is again used here. Also, the dominantly x-polarized basis modes are neglected. Recall that slab modes were solely TM polarized, meaning the only transverse electric field component was E_y. For the

SPP of a finite width waveguide all three E-fields exist, but the E_y component is still the dominant one for the large aspect ratio waveguides we are looking at. The cladding region however, supports normal modes polarized along x and y although the fractional power coupling between the input SPP mode and the x-polarized normal modes is typically well below 0.0001. The computer memory and computational time are halved by discarding the x-polarized modes.

In the next sections we look at radiation IL results for SiO_2 and LN, grouped by thickness as outlined above in Tables 4.4 and 4.5. The asymmetry is $\delta n = 25 \times 10^{-4}$ or smaller because of the reduced box size. Larger asymmetries in the same boxes would not be accurate for the propagation length of $2400\mu m$ used.

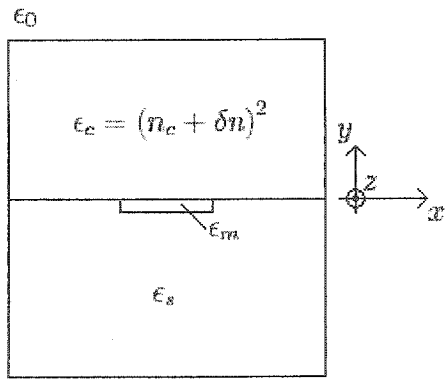

Figure 4.6: Geometry for radiation analysis of finite width SPP waveguide.

4.4.2 Radiation in Silica

The results for radiation in silica are shown in Figures 4.7, 4.8, and 4.9, for metal thickness 15nm, 20nm, and 30nm respectively. The material parameters are $\epsilon_s = 1.444^2$, $\epsilon_c = (1.444 + \delta n)^2$, and $\epsilon_m = -131.95$ at a free space wavelength $\lambda_0 = 1.55\mu m$. For each figure, plots (a), (b), and (c) have δn_{eff} equal to 1, 2, and 5×10^{-4} respectively.

For the lower δn_{eff} plots (plots (a) and (b) in each figure), the strong ripple from the interference of the normal modes near the waveguide is again seen and the location of the peaks again agrees with the predicted location. For example, the 25×10^{-4} curve in plot (a) of Figure 4.7 has its first peak at $z = 627\mu m$ and the predicted peak is at $z \approx \lambda_0/\delta n = 620\mu m$. Thicker waveguides again suppress this interference even though the waveguide is now only finite in width. While strong interference may still occur between normal modes away from the waveguide, the input and output modes are still highly concentrated near the core and the interference away from this region is not noticed.

The ripple should be accounted for in any design. Comparing the $\delta n = 20$ and 25×10^{-4} curves in plots (a) and (b) (of each figure) shows that because of the ripple a larger dynamic range is found for the higher confinement modes at $z = 2400\mu m$, while a slightly shorter or longer asymmetric region would yield the opposite result.

The results for the three thicknesses are practically identical, for a given δn_{eff}, with less than 1dB difference in the IL after $z = 2400\mu m$ for any particular case. It is best then to use a thinner waveguide of 15-20nm in an application since the dynamic range from the radiation loss is the same, but the on-state attenuation is lower.

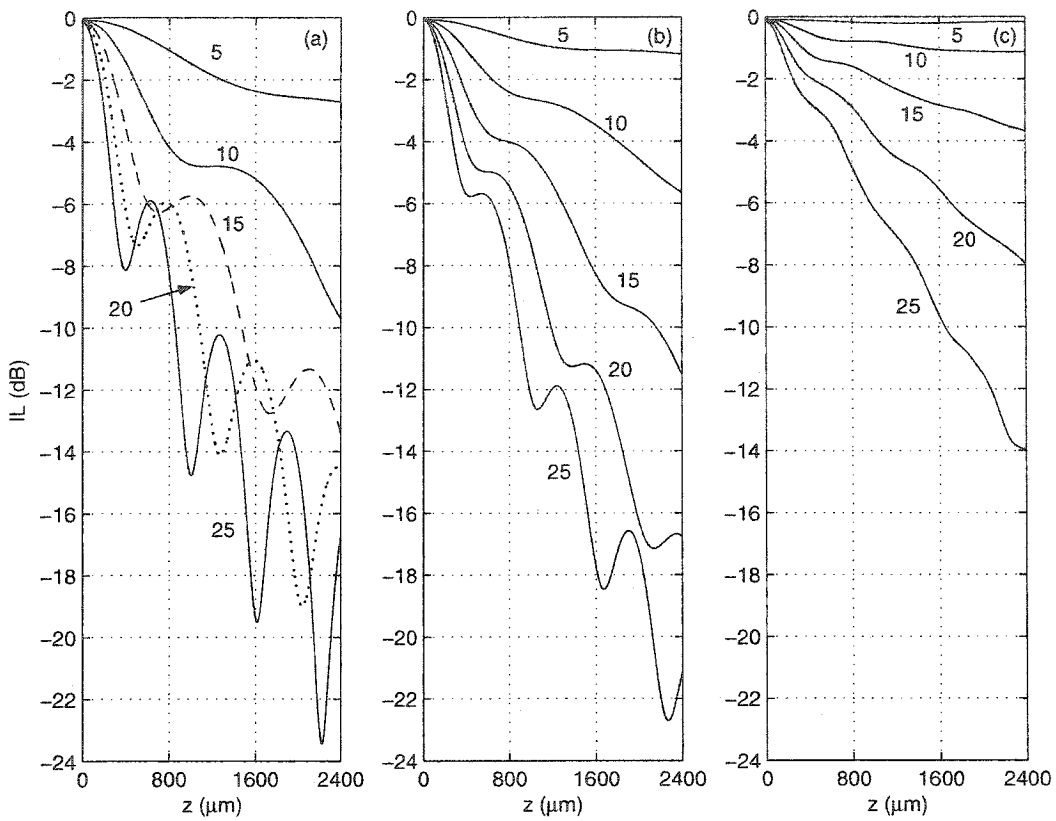

Figure 4.7: Radiation loss curves for SPPs of 15nm thick Au in SiO_2. (a) w=4.45μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=5.5$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=8.52$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

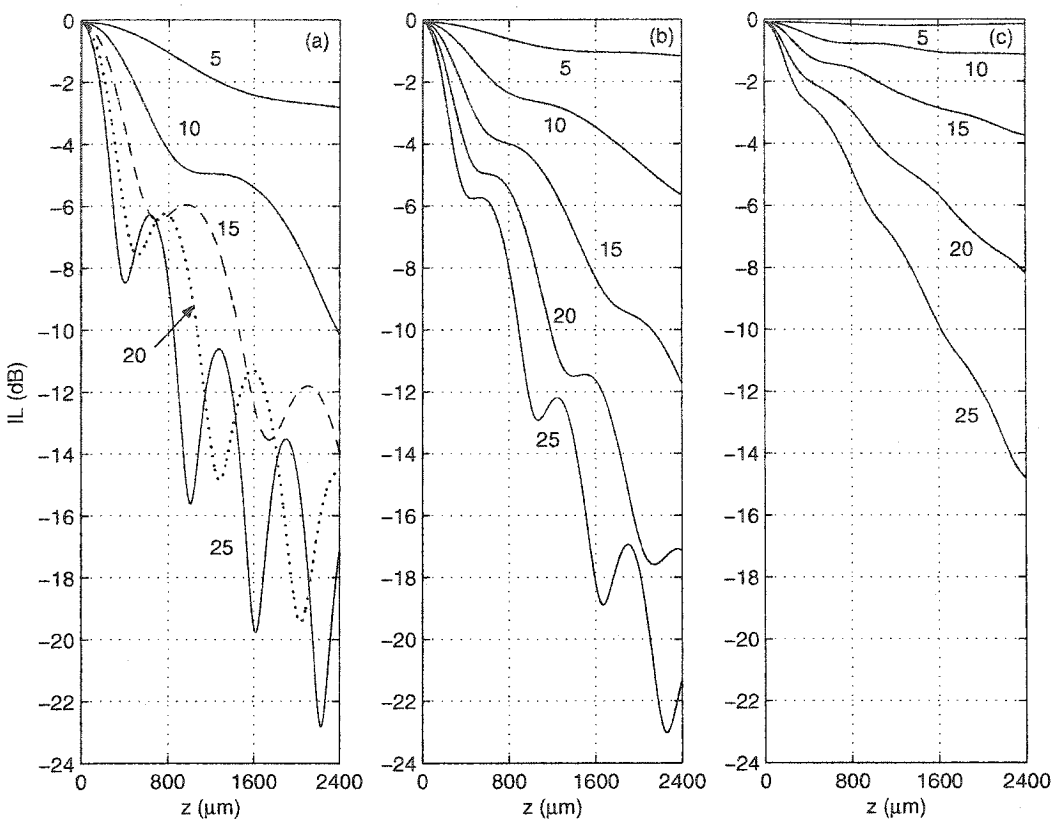

Figure 4.8: Radiation loss curves for SPPs of 20nm thick Au in SiO_2. (a) w=3.0μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=3.6$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=4.95$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

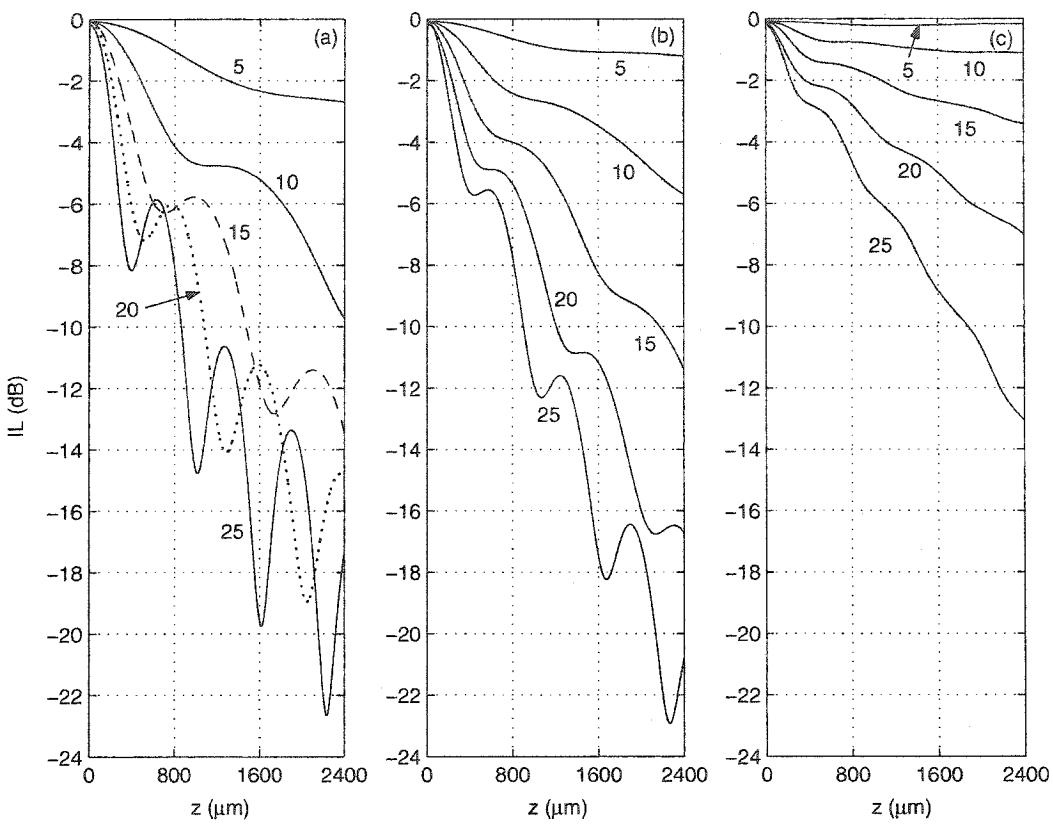

Figure 4.9: Radiation loss curves for SPPs of 30nm thick Au in SiO_2. (a) w=1.9μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=2.18$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=2.78$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

4.4.3 Radiation in Lithium Niobate

Next we compute the radiation loss in LN using $\epsilon_s = 2.1377^2$, $\epsilon_c = (2.1377 + \delta n)^2$, and $\epsilon_m = -131.95$ at a free space wavelength $\lambda_0 = 1.55\mu m$. Figures 4.10, 4.11, and 4.12 show the results for metal thicknesses 15nm, 20nm, and 30nm respectively. Plots (a), (b), and (c) have δn_{eff} equal to 1, 2, and 5×10^{-4}, respectively.

The results for the three thicknesses are again nearly identical for a given δn_{eff}, with differences less than 1dB at $z = 2400\mu m$ so thinner metal should be used in application to have the lowest on-state IL.

The interference phenomenon also occurs in these results and the discussion on this topic for silica applies here as well.

Since the LN and SiO_2 results are each nearly the same for a given δn_{eff} it is easy to compare the two materials. Looking for example at the t=20nm figures for each (Figures 4.8 and 4.11), the results are within about 1dB again. This similarity between materials was seen for the slab results for $\delta n_{eff} = 1$ and 5×10^{-4}, and the earlier discovery that the s_b mode cutoff points are very similar for low δn_{eff}. Evidently the ss_b^0 modes chosen here also fall into this regime.

Applications of the results in SiO_2 and LN are discussed in Section 4.6.

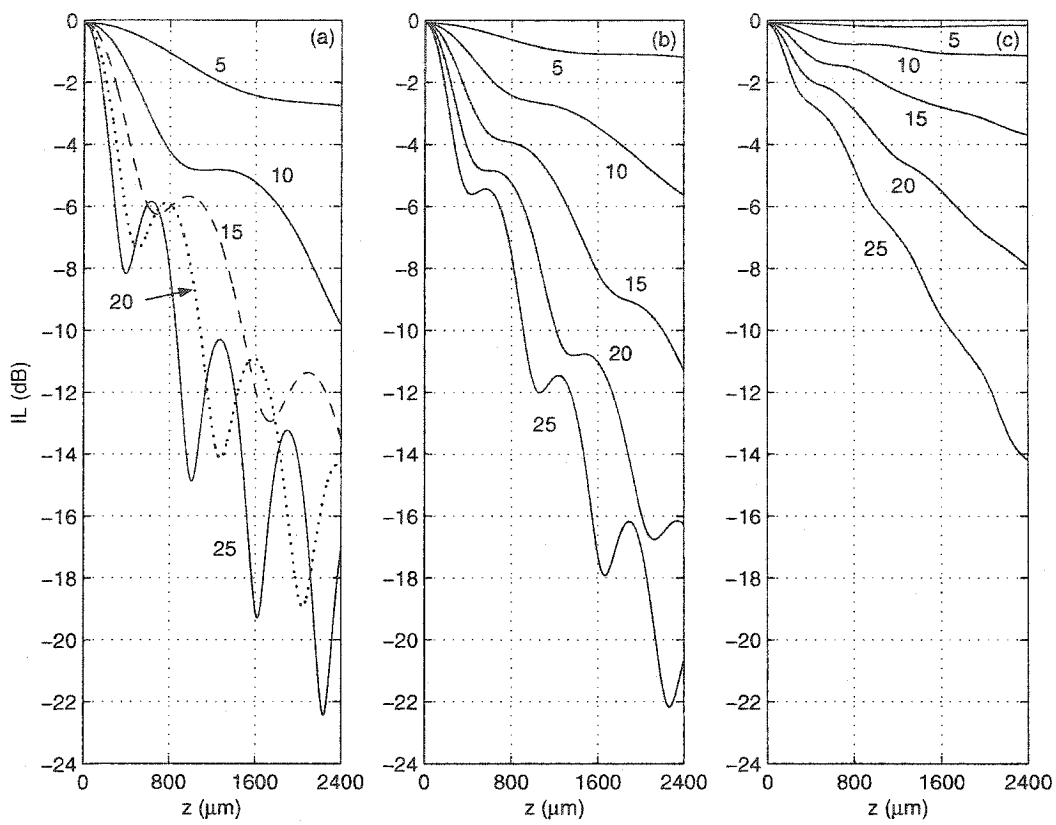

Figure 4.10: Radiation loss curves for SPPs of 15nm thick Au in LN. (a) w=1.5μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=1.73$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=2.2$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

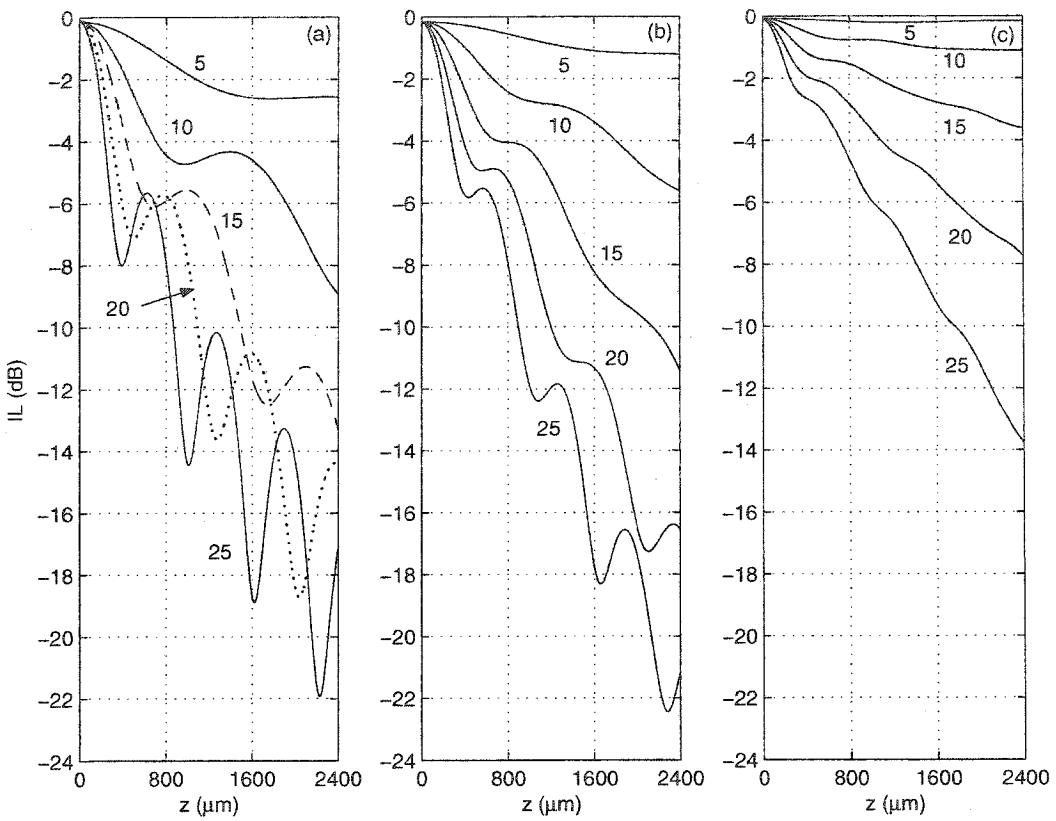

Figure 4.11: Radiation loss curves for SPPs of 20nm thick Au in LN. (a) w=1.05μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=1.19$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=1.46$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

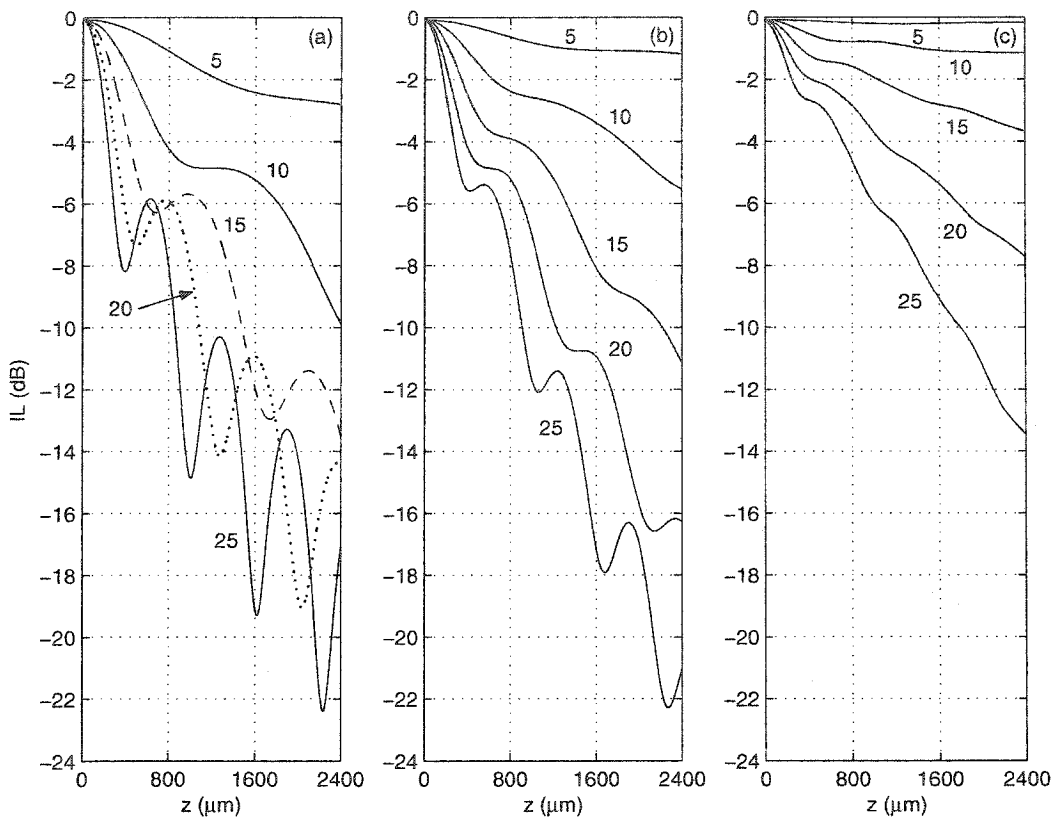

Figure 4.12: Radiation loss curves for SPPs of 30nm thick Au in LN. (a) w=0.665μm, $\delta n_{eff} = 1 \times 10^{-4}$; (b) w=0.745$\mu m$, $\delta n_{eff} = 2 \times 10^{-4}$; (c) w=0.885$\mu m$, $\delta n_{eff} = 5 \times 10^{-4}$. Curves labelled with δn asymmetry, with a factor of 10^{-4} suppressed.

4.5 Comparison With Experimental Results for Silica

Experiments were performed[85] to test the predictions of the radiation loss in SiO_2. Samples were prepared and tested by staff of Spectalis Corp. in the Micro-Electronics Laboratory of Carleton University. A thin gold layer of about 20nm of thickness was e-beam evaporated onto an Si wafer with $15\mu m$ of SiO_2 on top of the Si. The Au was then patterned using contact lithography into waveguide strips with the remaining metal being lifted off.

The top clad used to provide the index change was an index matching, thermo-optic oil (Light-span LS-5246). The temperature of the entire sample was controlled using a thermo-electric cooler and monitored by a micro-thermistor. Temperature changes were converted to index changes using the dn/dT value supplied by Light-span ($dn/dT = -3.55 \times 10^{-4}/^{o}C$).

The waveguides were end-fire excited by polarization maintaining fiber and the output was captured by an SMF-28 single mode fiber.

The structure used for the normal mode analysis of these waveguides is nearly the same as in the last sections. The only exception is that the bottom free space area is replaced by Si, $15\mu m$ below the plane of the metal. The top clad is $130\mu m$ thick and the box is $200\mu m$ wide. We use $\epsilon_s = 1.448^2$, $\epsilon_c = (1.448 + \delta n)^2$, and $\epsilon_m = -131.95 + i12.65$ at a free space wavelength $\lambda_0 = 1.55\mu m$. For the silicon layer $\epsilon_{Si} = 12.085$ was used.[59] The metal strip is $4\mu m$ wide and results are compared for two propagation lengths: 1.20mm and 2.55mm.

Some approximations were necessarily made since mode attenuation and coupling loss measurements were not made on these samples, and their respective values in the on-state could not be separated. Coupling loss in this case is input light not coupled into the metal

waveguide that propagates through the system and then does not couple to the SMF-28 output fiber. We assume that this coupling loss remains relatively constant throughout the cutoff curve so that the total on-state IL and coupling loss can be subtracted from the results and the additional radiation loss due to the asymmetry can be found.

As well, an accurate theoretical model of the polarization maintaining input fiber used in the experiments is necessary to predict the input coupling loss, but this would require mode profiling experiments, beyond the scope of this work. In the absence of models of the input and output fiber we use the same approach as before with a symmetric waveguide used as the input and output condition. The coupling loss in the on-state for the theoretical predictions is then zero.

Figure 4.13 shows the results for a 1.20mm long waveguide. The circles are the results from the normal mode analysis computations, the solid curves are measurements from sample A and the dotted curves are from sample B. Measurements were repeated to account for any drift of the input and output fibers. The two curves (from each sample) are the best and worst fits from about six trials with each sample. There is good agreement between the predictions and the measurements with a difference of a little more than 1dB in the worst case, at the maximum asymmetry.

Figure 4.14 shows results for longer 2.55mm waveguides from two different samples. As expected the radiation loss is higher for the longer sample. The results in this case are better centered on the predicted values, and are still within about 1dB.

The small discrepancies between the predicted and the measured results can probably be accounted for by imperfections in the waveguide shape and size and changes to the coupling loss as the asymmetry is changed. From these results though, the theory is in good agreement with experiment.

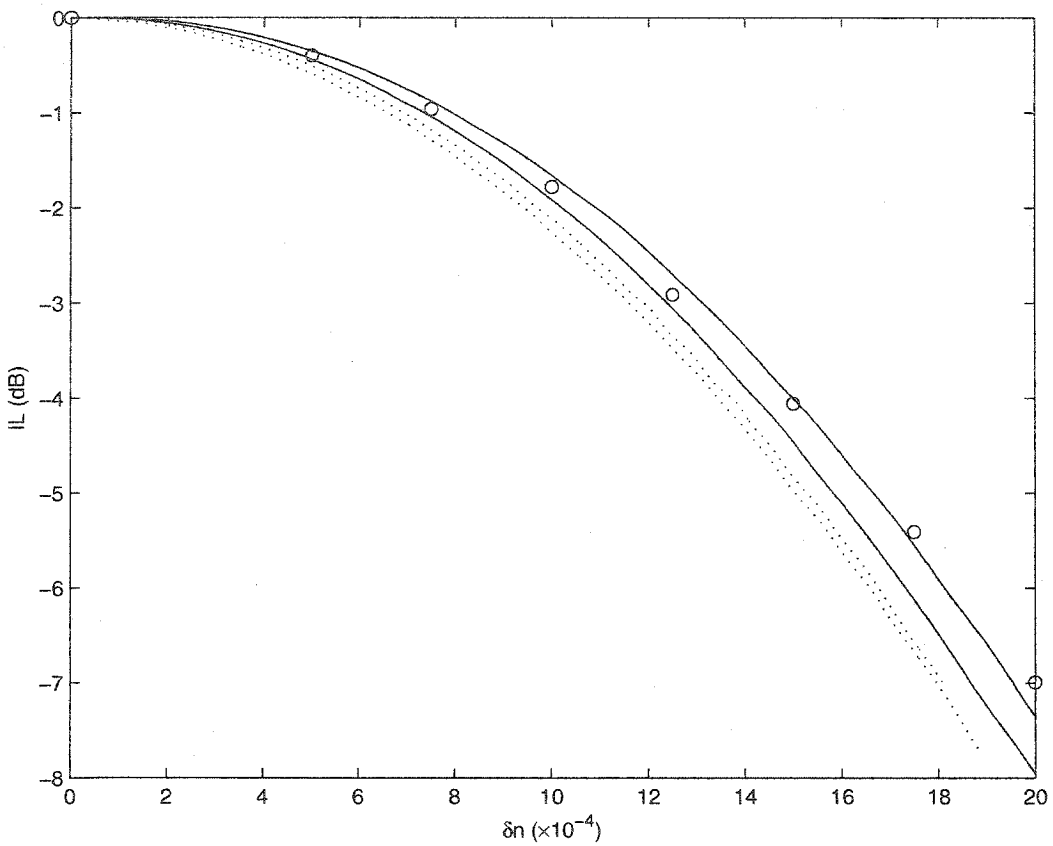

Figure 4.13: Comparison of experimental radiation loss for a 20nm thick, $4\mu m$ wide Au waveguide in SiO_2 for two samples (solid and dotted curves) with predicted values (circles) for a 1.20mm long sample.

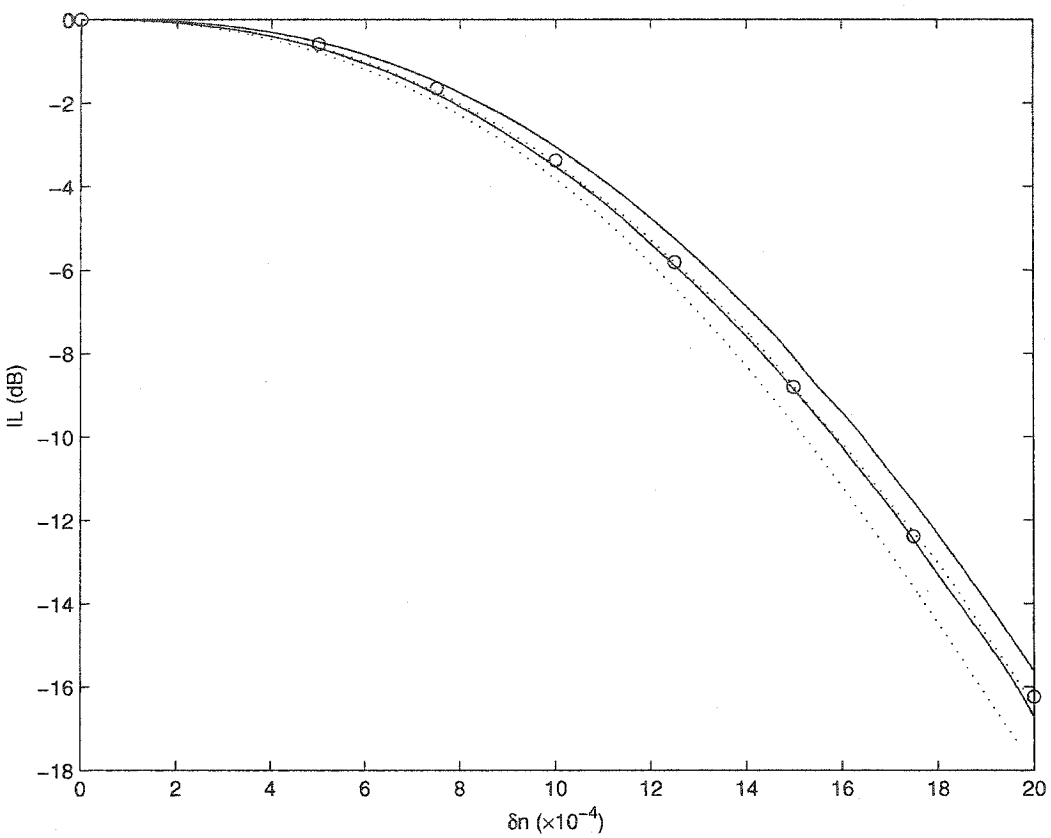

Figure 4.14: Comparison of experimental radiation loss for a 20nm thick, $4\mu m$ wide Au waveguide in SiO_2 for two samples (solid and dotted curves) with predicted values (circles) for a 2.55mm long sample.

4.6 Applications of Mode Cutoff

4.6.1 Introduction

We saw that waveguides in LN and silica with similar δn_{eff} had very similar radiation loss curves. A choice between these two materials (or other active materials with similar index of refraction) based then on the on-state mode attenuation favors the lower index SiO_2, since identical dynamic range can be obtained for lower loss. It is important to note that virtually identical radiation performance is found in these vastly different materials. A designer then does not sacrifice radiation performance by using one particular material over the other (assuming identical index changes are possible). In practice other issues such as device speed may be more important than small differences in on-state attenuation. In this section we briefly discuss some possible applications for slower thermo-optic based devices and faster electro-optic based modulators.

4.6.2 Thermo-optic Applications

Deep extinction will be difficult in actual silica with $IL < -20dB$ requiring $\delta n \approx 25 \times 10^{-4}$ and silica having a thermo-optic coefficient of only $dn/dT \sim 10^{-5}\ K^{-1}$. However, some thermo-optic polymers have coefficients more than an order of magnitude greater so deep extinction should be possible in other materials.

Figure 4.15 shows the cross-section of a possible realization of a slow modulator or variable optical attenuator (VOA) based on the mode cutoff effect. The layout is similar to that of the experiments compared with in the last section.

Heating elements are placed away from the metal waveguide so that they do not contribute additional loss. They can be placed above and/or below the cladding material, as shown. Two heaters may be required to reduce thermal gradients within the device.

One of the cladding materials would be thermo-optic while the other passive materical would have a negligible thermo-optic coefficient. Heating of this structure would result in an index asymmetry leading to radiation loss.

Some materials have coefficients such that heating reduces the index while others have the reverse situation where heating increases the index. If materials are chosen such that the index of refraction increases with temperature, then more care needs to be taken in designing the width of the heaters. If the heaters are too narrow a thermally induced waveguide will be created between the heater and the metal waveguide and light will not radiate away from the core effectively, but will move into the new waveguide.

When choosing the waveguide it is best to choose dimensions so that the lowest attenuation possible is obtained for the on-state as it was seen that waveguides with the same relative confinement δn_{eff} (but different attenuation) have nearly identical radiation insertion loss in the off-state.

Another heating scenario is to use the metal waveguide as a heater. This is a natural advantage of the metal waveguide and it eliminates the need for external heaters. The

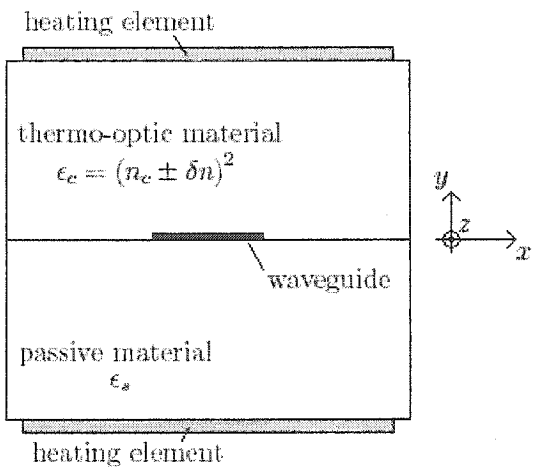

Figure 4.15: Schematic cross-section of thermo-optic based mode cutoff device.

physical lifetime of the very thin waveguide could be an issue with this scenario.

A third related possibility is to again use the waveguide as a heater, but to have both clads comprised of thermo-optic material where the coefficient is such that the index of refraction reduces when heated. This does not cut off the mode as we have discussed in this report, but would create an anti-guiding region around the waveguide where the light will leak away from the core into the higher index clads away from the heated region.

4.6.3 Electrooptic Applications

Large dynamic ranges should be possible in LN since the largest index change in this electro-optic crystal is around $10-15\times10^{-4}$ leading to radiation IL of about -14dB. Other electro-optic crystals and polymers have larger electro-optic coefficients,[41] and it should be feasible to have better than -20dB dynamic range in a short length (a few millimeters). The dynamic range here is comparable to much longer LN Mach-Zehnder modulators. These are based on Ti-indiffused waveguides and are typically several centimeters long.[86] Extrapolating the results obtained from the normal mode analysis in LN to $z = 5000\mu m$ would give much improved performance and much shorter lengths than traditional LN modulators.

Figure 4.16 shows the schematic for a possible implementation of a cutoff modulator in an electro-optic material such as LN. Electrodes are placed above and below the clads as shown, far enough away to be optically non-intrusive to the mode.

Several material scenarios are possible here. If the crystal is aligned in the top clad such that the index of refraction increases it is necessary to orient the crystal in the lower clad anti-parallel to the top so that the index decreases in the bottom clad (for linear electro-optic materials). The lower clad could also consist of a passive material with no index change occurring, however this discards half of the possible effect.

Care must be taken in designing the electrodes as an electro-optically induced waveguide can also be formed above or below the metal waveguide in these scenarios.

Again the natural advantage of the metal waveguide is its use as an electrode. In this scenario the crystals in each clad would be aligned parallel since the fields above and below the waveguide would be oppositely directed. The outer electrodes in Figure 4.16 would in this case both be grounds if the waveguide is raised to a potential.

Another scenario is to have the electrodes placed vertically on each side of the waveguide (a sufficient distance away) since some materials such as KDP have changes to the y component (in our coordinate system) of their index tensor resulting from fields applied along x. Yariv[41] has a very good discussion on this and electro-optic crystals in general.

Clearly there are many design considerations for applications to thermo-optic and electro-optic devices, not the least of which is the electro-static and thermal analysis of the devices, especially if the metal waveguide is used as an electrode or heater.

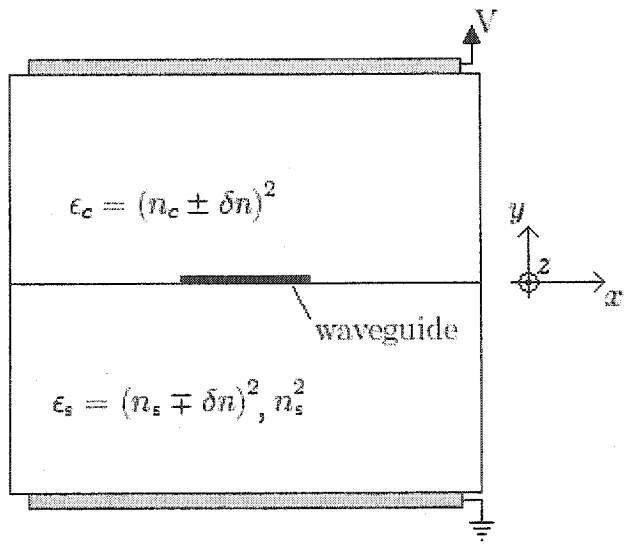

Figure 4.16: Schematic cross-section of electro-optic based mode cutoff device.

Chapter 5

Conclusion

5.1 Thesis Contributions

The concept of cutoff for the bound modes of a slab SPP waveguide was revisited and analyzed. The effect of this mode cutoff on the propagation characteristics of an input excitation was analyzed using the normal mode analysis, and the radiative cutoff characteristics of SPP modes were computed for the first time in both slab waveguides and new finite-width waveguides. Results for finite width waveguides agree with experimental results. All modes were computed using a full vectorial analysis and waveguide attenuation was accounted for. New design curves were also calculated for the important materials silica and lithium niobate. Applications of these results for practical devices are proposed and discussed.

A secondary contribution was the collection and development of all necessary background for a proper understanding of SPP mode waveguiding in thin metal strips and films into a consistent formalism.

5.2 Suggestions for Future Work

An initial round of experiments were made to verify the predicted radiation loss. This analysis could be improved with a better understanding of the experimental waveguide attenuation and coupling loss. The predictions could also be improved by accurately modelling the input and output fibers.

An important investigation would be a practical implementation of a modulator based on mode cutoff as discussed briefly in this work. This would need to account for a practically sized region of index asymmetry, placement of electrodes or heaters near the waveguide to induce the asymmetry, or possibly using the waveguide as an electrode or heating element since it is metal.

The normal mode analysis approach could be applied to analysis of cutoff in other SPP waveguide structures such as edge-couplers and broadside couplers. In a coupler, the antisymmetric supermode is more sensitive to horizontal and vertical asymmetries than the symmetric supermode and can be cutoff.[87] A proper normal mode analysis would show whether this could be used effectively in a directional coupler type modulator.

Because the SPP waveguide platform is typically lossy, the length of components is important. It would be useful to be able to predict the optimal length of integrated optics components such as y-junctions, tapers and multimode-type splitters using SPP waveguides. The normal mode analysis technique is also able to handle this.[78]

List of Figures

List of Tables

Bibliography

[1] A. Sommerfeld. *Ann. Physik*, 28:665–736, 1909.

[2] U. Fano. The theory of anomalous diffraction gratings and of quasi-stationary waves on metallic surfaces (Sommerfeld's waves). *J. Opt. Soc. Am.*, 31:213–222, 1941.

[3] R. H. Ritchie. Plasma losses by fast electrons in thin films. *Phys. Rev.*, 106(5):874–881, 1957.

[4] E. A. Stern and R. A. Ferrell. Surface plasma oscillations of a degenerate electron gas. *Phys. Rev.*, 120(1):130–136, 1960.

[5] C. J. Powell and J. B. Swan. Origin of the characteristic electron energy losses in aluminum. *Phys. Rev.*, 115(4), 1959.

[6] A. Otto. Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. *Zeitschrift fur Physik*, 216:398–410, 1968.

[7] E. Kretschmann and H. Raether. *Z. Naturf.*, 23:2135, 1968.

[8] R. A. Innes and J. R. Sambles. Optical characterisation of gold using surface plasmon-polaritons. *J. Phys. F: Met. Phys.*, 17:277–287, 1987.

[9] K. L. Kliewer and R. Fuchs. Collective electronic motion in a metallic slab. *Phys. Rev.*, 153(2):498–153, 1967.

[10] E. N. Economou. Surface plasmons in thin films. *Phys. Rev.*, 182(2):539–554, 1969.

[11] D. Sarid. Long-range surface-plasma waves on very thin metal films. *Phys. Rev. Lett.*, 47(26):1927–1930, 1981.

[12] J. J. Burke, G. I. Stegeman, and T. Tamir. Surface-polariton-like waves guided by thin, lossy metal films. *Phys. Rev. B*, 33(8):5186–5201, 1986.

[13] F. Yang, J. R. Sambles, and G. W. Bradberry. Long-range surface modes supported by thin films. *Phys. Rev. B*, 44(11):5855–5872, 1991.

[14] C. Jung, S. Yee, and K. Kuhn. Integrated optics waveguide modulator based on surface plasmon resonance. *J. Lightwave Tech.*, 12(10):1802–1806, 1994.

[15] N. A. Janunts and Kh. V. Nerkararyan. Modulation of light radiation during input into waveguide by resonance excitation of surface plasmons. *Appl. Phys. Lett.*, 79(3):299–301, 2001.

[16] J. S. Schildkraut. Long-range surface plasmon electrooptic modulator. *Appl. Opt.*, 27(21):4587–4590, 1988.

[17] K. Welford. Surface plasmon-polaritons and their uses. *Optical and Quantum Electronics*, 23:1–27, 1991.

[18] J. Johnstone, G. Stewart, T. Hart, and B. Culshaw. Surface plasmon polaritons in thin metal films and their role in fiber optic polarizing devices. *J. Lightwave Tech.*, 8(4):538–543, 1990.

[19] J. Tominaga *et al.* Local plasmon photonic transistor. *Appl. Phys. Lett.*, 78:2417–2419, 2001.

[20] A. Tredicucci *et al.* Single-mode surface-plasmon laser. *Appl. Phys. Lett.*, 76:2164–2166, 2000.

[21] D. J. Bergman and M. I. Stockman. Can we make a nanoscopic laser? *Laser Physics*, 14(3):409–411, 2004.

[22] J. Lambe and S. L. McCarthy. Light emission from inelastic electron tunneling. *Phys. Rev. Lett.*, 37(14), 1976.

[23] S. I. Bozhevolnyi, J. Erland, K. Leosson, P. M. W. Skovgaard, and J. M. Hvam. Waveguiding in surface plasmon polariton band gap structures. *Phys. Rev. Lett.*, 86(14):3008–3011, 2001.

[24] M. Quinten, A. Leitner, J. R. Krenn, and F. R. Aussenegg. Electromagnetic energy transport via linear chains of silver nanoparticles. *Opt. Lett.*, 23(17):1331–1333, 1998.

[25] I. V. Novikov and A. A. Maradudin. Channel polaritons. *Phys. Rev. B*, 66(035403):923–925, 2002.

[26] J. P. Kottmann and O. J. F. Martin. Plasmon resonant coupling in metallic nanowires. *Opt. Express*, 8(12):655–663, 2001.

[27] P. Berini. Plasmon-polariton modes guided by a metal film of finite width. *Opt. Lett.*, 24(15):1011–1013, 1999.

[28] P. Berini. Plasmon-polariton waves guided by thin lossy metal films of finite width: Bound modes of symmetric structures. *Phys. Rev. B.*, 61(15):10484–10503, 2000.

[29] R. Charbonneau, P. Berini, E. Berolo, and E. Lisicka-Shrzek. Experimental observation of plasmon-polariton waves supported by a thin metal film of finite width. *Opt. Lett.*, 25(11):844–846, 2000.

[30] L. Wendler and R. Haupt. Long-range surface plasmon-polaritons in asymmetric layer structures. *J. Appl. Phys.*, 59(9):3289–3291, 1986.

[31] P. Berini. Plasmon-polariton waves guided by thin lossy metal films of finite width: Bound modes of asymmetric structures. *Phys. Rev. B.*, 63(125417):1–14, 2001.

[32] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics II.* Addison-Wesley, Massachusetts, 1964.

[33] M. C. Steele and B. Varul. *Wave Interactions in Solid State Plasmas.* McGraw-Hill, New York, 1969.

[34] J. B. Pendry and D. R. Smith. Reversing light with negative refraction. *Physics Today*, 57(6), 2004.

[35] N. W. Ashcroft and N. D. Mermin. *Solid State Physics.* Saunders College, Fort Worth, 1976.

[36] C. Kittel. *Introduction to Solid State Physics.* Wiley, New York, 7th edition, 1996.

[37] M. S. Rogalski. *Solid State Physics.* Overseas, Amsterdam, 2000.

[38] H. Ibach and H. Lüth. *Solid-State Physics, An Introduction to Principles of Materials Science.* Springer, Berlin, 3rd edition, 2003.

[39] O. Madelung. *Introduction to Solid-State Theory.* Springer-Verlag, Berlin, 1978.

[40] M. Born and E. Wolf. *Principles of Optics.* Cambridge, Cambridge, 7th edition, 1999.

[41] A. Yariv and P. Yeh. *Optical Waves in Crystals.* Wiley, New Jersey, 2003.

[42] P. Yeh. *Optical Waves in Layered Media.* Wiley, New York, 1988.

[43] G. R. Fowles. *Introduction to Modern Optics.* Dover, New York, 2nd edition, 1975.

[44] E. Hecht. *Optics.* Addison-Wesley, 1987.

[45] J. D. Jackson. *Classical Electrodynamics.* Wiley, New York, 3rd edition, 1999.

[46] D. J. Griffiths. *Introduction to Electrodynamics.* Prentice-Hall, New Jersey, 2nd edition, 1989.

[47] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii. *Electrodynamics of Continuous Media.* Butterworth-Heinenann, Jordan Hill, Oxford, 2nd edition, 1984.

[48] H. Raether. *Excitation of Plasmons and Interband Transitions by Electrons.* Springer-Verlag, Berlin, 1980.

[49] H. Raether. *Surface Plasmons on Smooth and Rough Surfaces and on Gratings.* Springer-Verlag, Berlin, 1988.

[50] V. M. Agranovich and A. A. eds. Maradudin. *Surface Polaritons, Electromagnetic Waves at Surfaces and Interfaces.* North-Holland, Amsterdam, 1982.

[51] A. D. Boardman. *Electromagnetic Surface Modes.* Wiley, Chichester, 1982.

[52] A. P. French. *Vibrations and Waves.* Norton, New York, 1971.

[53] R. G. Hunsperger. *Integrated Optics.* Springer, Berlin, 5th edition, 2002.

[54] J. Lagois and B. Fischer. *Surface Polaritons, Electromagnetic Waves at Surfaces and Interfaces*, chapter 2. edited by V.M. Agranovich and A.A. Maradudin. North-Holland, Amsterdam, 1982.

[55] D. N. Mirlin. *Surface Polaritons, Electromagnetic Waves at Surfaces and Interfaces*, chapter 1. edited by V.M. Agranovich and A.A. Maradudin. North-Holland, Amsterdam, 1982.

[56] G. Shvets. Photonic approach to making a material with a negative index of refraction. *Phys. Rev. B.*, 67(035109):1–8, 2003.

[57] G. I. Stegeman. Long range surface plasmons in birefringent media. *Applied Optics*, 22(15):2243–2245, 1983.

[58] G. N. Zhizhin, M. A. Moskalova, E. V. Shomina, and V. A. Yakovlev. *Surface Polaritons, Electromagnetic Waves at Surfaces and Interfaces*, chapter 3. edited by V.M. Agranovich and A.A. Maradudin. North-Holland, Amsterdam, 1982.

[59] E. D. Palik. *Handbook of Optical Constants of Solids.* Academic Press, Orlando, 1985.

[60] G. I. Stegeman, R. F. Wallis, and A. A. Maradudin. Excitation of surface polaritons by end-fire coupling. *Optics Letters*, 8(7):386–388, 1983.

[61] R. W. Alexander, G. S. Kovener, and R. J. Bell. Dispersion curves for surface electromagnetic waves with damping. *Phys. Rev. Lett.*, 32(4):154–157, 1974.

[62] A. W. Snyder and J. D. Love. *Optical Waveguide Theory.* Kluwer, Great Britain, 2000.

[63] E. Butkov. *Mathematical Physics.* Addison-Wesley, Massachusetts, 1968.

[64] J. Chilwell and I. Hodgkinson. Thin-films field-transfer matrix theory of planar multilayer waveguides and reflection from prism-loaded waveguides. *J. Opt. Soc. Am. A*, 1(7):742–753, 1984.

[65] CK. Chen, P. Berini, DZ. Feng, S. Tanez, and V. P. Tzolov. Efficient and accurate numerical analysis of multilayer planar optical waveguides in lossy anisotropic media. *Opt. Exp.*, 7(8):260–272, 2000.

[66] I. Breukelaar. Implementation of a versatile model for numerical analysis of multilayer planar optical waveguide structures: Verification of model and results for new structures. 4th year undergraduate thesis (unpublished), University of Ottawa, April 2001.

[67] B. E. A. Saleh and M. C. Teich. *Fundamentals of Photonics.* Wiley, New York, 1991.

[68] D. Marcuse. *Theory of Dielectric Optical Waveguides.* Academic Press, New York, 1974.

[69] D. L. Lee. *Electromagnetic Principles of Integrated Optics.* Wiley, New York, 1986.

[70] A. B. Buckman. *Guided-Wave Photonics.* Saunders College, Fort Worth, 1982.

[71] K. Okamoto. *Fundamentals of Optical Waveguides.* Academic Press, San Diego, 2000.

[72] F. A. Burton and S. A. Cassidy. A complete description of the dispersion relation for thin metal film plasmon-polaritons. *J. Lightwave. Technol.*, 8(12):1843–1849, 1990.

[73] B. Prade, J.Y. Vinet, and A. Mysyrowicz. Guided optical waves in planar heterostructures with negative dielectric constant. *Phys. Rev. B*, 44(24):13556–13572, 1991.

[74] C. F. Gerald and P. O. Wheatley. *Applied Numerical Analysis.* Addison-Wesley, Massachusetts, 5th edition, 1994.

[75] R. S. Weis and T. K. Gaylord. Lithium niobate: Summary of physical properties and crystal structure. *Appl. Phys. A*, 37:191–203, 1985.

[76] F. Villa, T. Lopez-Rios, and L. E. Regalado. Electromagnetic modes in metal-insulator-metal structures. *Phys. Rev. B*, 63(165103), 2001.

[77] E. Kreyszig. *Advanced Engineering Mathematics.* Wiley, New York, 7th edition, 1993.

[78] D. F. G. Gallagher and T. P. Felici. Eigenmode expansion methods for simulation of optical propagation in photonics: pros and cons. *Proc. SPIE, Int. Opt.: Devices, Materials, and Technologies VII*, 4987:69–82, 1995.

[79] G. I. Stegeman, A. A. Maradudin, and R. F. Wallis. The optics of surface plasmons. *Journal de Physique Colloque C5*, 4(45):233–241, 1984.

[80] Smith R. E. and S. N. Houde-Walter. Failure of the leaky-mode representation near the waveguide mode cutoff. *Opt. Lett.*, 20(10):1133–1135, 1995.

[81] T. Nikolajsen, K. Leosson, I. Salakhutdinov, and S. I. Bozhevolnyi. Polymer-based surface-plasmon-polariton stripe waveguides at telecommunication wavelengths. *Appl. Phys. Lett.*, 82(5):668–670, 2003.

[82] COMSOL AB, Sweden. *FEMLAB Electromagnetics Module Manual*, 2.3 edition, 2002.

[83] J Jin. *The Finite Element Method in Electromagnetics.* Wiley, New York, 2nd edition, 2002.

[84] R. C. Boonton. *Computational Methods for Electromagnetics and Microwaves.* Wiley, New York, 1992.

[85] R. Charbonneau. Experimental and theoretical correlation of the mode cutoff behaviour of thin metal waveguides finite in width. (unpublished), Spectalis Corp., December 2003.

[86] E. L. Wooten. A review of lithium niobate modulators for fiber-optic communications systems. *IEEE J. Selec. Top. Quantum Elec.*, 6(1):69–82, 2000.

[87] I. Breukelaar. Unpublished.