SURFACE PLASMON EFFECTS IN NANO-OPTICS: SUPER-RESOLVED OPTICAL READOUT AND COHERENCE CONVERSION
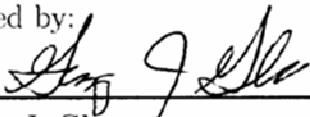
by

Choon How Gan

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
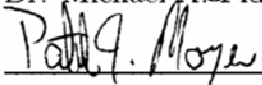Optical Science and Engineering

Charlotte

2009

Approved by:

_____
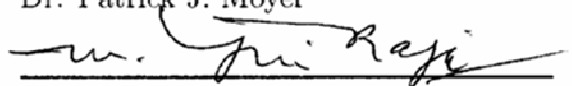Dr. Greg J. Gbur

_____
Dr. Michael A. Fiddy

_____
Dr. Patrick J. Moyer

_____
Dr. M. Yasin Akhtar Raja

_____
Dr. Kent E. Curran

ABSTRACT

CHOON HOW GAN. Surface plasmon effects in nano-optics: super-resolved optical readout and coherence conversion. (Under the direction of Dr. GREG J. GBUR.)

Surface plasmon effects in nano-optical systems are investigated through rigorous numerical simulations and analytical modeling. Several strategies to achieve superresolution in plasmon-assisted optical readout systems are proposed, and simulations reveal that resolution up to $\lambda/5.6$ ($\lambda \equiv$ wavelength of light) is achievable. These promising results can increase dramatically the storage capacity permitted with current optical data storage technologies. It is also shown that surface plasmons can be employed to modulate the spatial coherence of light that emanates from subwavelength apertures. Starting from a simple Young's double-slit geometry, feasibility of a practical plasmon-assisted coherence converting device is demonstrated as progress is made to investigate multiple aperture arrays. Finally, the surface plasmon effects are exploited in multi-layered metallic structures to both impede the field decay in a single metal slab and provide for extraordinary optical transmission. The results presented here suggest a number of ways in which surface plasmons can play important roles for the development of the nano-optical technologies.

ACKNOWLEDGEMENTS

I like to express my appreciation to the people who have helped directly or indirectly in various ways in the accomplished work and research presented in this thesis. Instructors who have imparted to me the fundamental aspects of the optical sciences include Prof. Michael Fiddy, Prof. Yasin Raja, Prof. Raphael Tsu, Dr. Vasily Astratov, Dr. Angela Davies, Dr. Faramarz Farahi, Dr. Tsing-Hua Her, Dr. Eric Johnson, and Dr. Patrick Moyer. Technical support provided by Evan Rowell has allowed me access to various software packages for performing tedious calculations and numerical simulations, which are key ingredients in a thesis reporting theoretical work.

Stimulating discussions with Dr. Miguel Alonso and Dr. Olga Korotkova have offered many useful ideas on the approaches and perspectives to adopt in studying the problems encountered. Many specialized topics have been made illuminating to me through the advice and guidance of Prof. Taco Visser and Dr. Paul Scott Carney, who were collaborators as well. It has been my privilege to have the opportunity to work with them. Yalong Gu, my co-worker, has also provided me with valuable insights through our deliberations on optical coherence theory.

Special thanks must go to my parents, Tiong Meng Gan and Kwee Tin Soh, who supported my decision to give up a stable job five years ago to pursue the doctorate program at the University of North Carolina at Charlotte. They must have *known* it was far from a foolish choice! My wife, Heng-Ching, whom I met in Charlotte, has also been very understanding and encouraging throughout.

Finally, I am indebted to Dr. Greg Gbur, my advisor, for the many constructive suggestions he has given, for his patience in dispelling my numerous obscure perceptions of physics and optics, for being a reliable mentor in times of downs, and for being a great friend whom I shared my joys with. None of the work in this thesis would have been possible if not for his continuous guidance and belief in me. It has been great working with him, even as one of his *dispensable slaves*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: SURFACE PLASMON EFFECTS IN NANO-OPTICS

"Why don't lightning cast a shadow, Jim?"

"Well, I reckon it do, but I don't know."

"Well, it don't. *I* know. The sun does, and a candle does,
but the lightning don't. Tom Sawyer says it don't, and it's so."

"Sho, child, I reckon you's mistaken 'bout dat.
Gimme de gun—I's gwyne to see."

So he stood up the gun in the door, and held it,
and when it lightened the gun didn't cast any shadow. Jim says:

"Well, dat's mighty cur'us—dat's oncommon cur'us. Now dey
say a ghos' don't cas' no shadder. Why is dat, you reckon?
Of course de reason is dat ghosts is made out'n lightnin',
or else de lightnin' is made out'n ghosts—but I don't know
which it is. I wisht I knowed which it is, Huck."

—Mark Twain, *Adventures of Huckleberry Finn*

For centuries, the study of light has generated numerous interesting controversies, and scientific discoveries that at times contradict intuition. As new technologies emerge, even things that were once thought to be possible only in science fiction have been successfully demonstrated with experiments. These include among many others, the teleportation of information carried by light photons (Zeilinger 2000, Gisin and Thew 2007), the complete "stopping" of the propagation of light (Phillips et al. 2002, Chiao and Milonni 2002, Ginsberg et al. 2007), and the development of invisibility "cloaking" devices (Cai et al. 2007, Atwater 2007, Pendry et al. 2006). With these breakthroughs, it seems that light can be manipulated to do nearly whatever we want it to, if only we can figure out how. The possibilities with optics are continu-

ally stretched to the limits as we advance, and one has to be constantly prepared for surprises that almost always raise objections at first sight[1].

Research on surface plasmon effects in nano-optics has also revealed several astounding phenomena when surface plasmons couple with light. Counter-intuition associated with some of these effects has also fueled debates among scientists as they seek to understand and unfold the physics responsible for the observed phenomena. One such surprise was the experimental observation of enhanced optical transmission with a subwavelength hole array in a metal plate (Ebbesen et al. 1998). In that experiment, optical transmission much higher than predicted by standard diffraction theory (Bethe 1944) for subwavelength holes was observed. In fact, the fraction of the light transmitted was found to be greater than the fraction of surface area occupied by the holes on the array, implying that even the light impinging on the metal between the holes contributed to the enhanced transmission.

To explain the unexpected enhancement in transmission, diverse views ranging from mediation by surface plasmons (Ebbesen et al. 1998, Martín-Moreno et al. 2001, Popov et al. 2002, Barnes et al. 2003, Schouten et al. 2005, Liu and Lalanne 2008), resonating Fabry-Perot cavities (Collin et al. 2001, Takakura 2001, Yang and Sambles 2002), waveguiding in subwavelength apertures (Lalanne et al. 2000, Gordon and Brolo 2005, Shin et al. 2005), and diffraction of evanescent waves (Lezec and Thio 2004, Gay et al. 2006) have been proposed. While diffraction is undoubtedly central to the process of transmission, studies have confirmed that the role of surface plamsons is critical for enhanced transmission in metallic structures (Ghaemi et al. 1998, Grupp

---

[1]See for instance, *Suprises in Theoretical Physics* by R. Peierls (Peierls 1979).

et al. 2000, Barnes et al. 2004, Lalanne and Hugonin 2006).

It was also found that enhanced transmission can occur even in the absence of holes or slits, i.e. surface corrugations or structured films alone would suffice (Bonod et al. 2003, Darmanyan and Zayats 2003, Giannattasio et al. 2004, Bai et al. 2005). Besides, the size and shape of the subwavelength holes in the arrays can also strongly influence the optical transmission (van der Molen et al. 2004, Koerkamp et al. 2004, Kim and Moyer 2006). More recently, researchers have found evidence that suggests that the coupling of surface plasmons between very closely-spaced metallic arrays of subwavelength hole and slit arrays can also contribute to enhanced optical transmission (Ye and Zhang 2005, Chan et al. 2006, Cheng et al. 2007, Marcet et al. 2008).

Another surprising discovery related to surface plasmon effects was the strong beaming of light emerging from a single subwavelength aperture surrounded by periodic corrugations on a metallic plate (Lezec et al. 2002, Degiron and Ebbesen 2004). In this instance, light emerges as a very narrow beam from the aperture, as opposed to the usual diffraction effects one would expect as the size of the aperture decreases. The strong beaming effect was explained by the presence of electromagnetic surface resonances arising from the coupling between surface plasmons propagating in the aperture and along the surface corrugations (Martín-Moreno et al. 2003).

To add to the list of surprises associated with surface plasmon effects, it has also been demonstrated that the quantum entanglement between pairs of photons contributing to the plasmon-assisted transmission observed by Ebbesen et al. (1988), can be preserved in the process. In essence, the metallic hole array converts incident photons into surface plasmons, which in turn tunnel through the holes before reradiating as light photons.

It was established experimentally that the entanglement of the photons is preserved in the process of coupling between plasmons and photons (Altewischer et al. 2002). This has opened up possibilities to employ nanostructured arrays in quantum information and computing systems.

With the current trend towards nanotechnology, research and development involving surface plasmons has continually received tremendous attention. Traditionally, surface plasmon resonances have been applied to Surface Enhanced Raman Spectroscopy (SERS) (Nie and Emory 1997, Xu et al. 1999, Cao et al. 2002) for the purpose of biosensing. The resonating effects of plasmons in metal nanoparticles have also made them useful as contrast agents in optical coherence tomography (OCT), and has led to the development of nanoshell cancer therapy (Lal et al. 2007, Loo et al. 2004).

Capable of coupling with light and producing enhanced field effects, surface plasmons have found wide use in other applications such as scanning near-field optical microscopy (SNOM) (Bouhelier et al. 2003, Bouhelier 2006, Hall et al. 2007), nanolithography (Martin 2003, Luo and Ishihara 2004), and near-field optical readout systems (Shi et al. 2002, Tominaga and Tsai 2003, Gbur et al. 2005). A plasmonic version of the LASER, the SPASER (Surface Plasmon Amplifcation by Stimulated Emission of Radiation), which can potentially act as a source of localized high-intensity fields, was proposed recently as a quantum-nanoplasmonic device (Bergman and Stockma 2003). Two-dimensional optical elements and waveguides, as well as nanowires for guiding both light and plasmonic fields, have been investigated and developed by several research groups (Ditlbacher et al. 2002, Krenn et al. 2002, Takahara and Kobayashi 2004, Maier and Atwater 2005, Karalis et al. 2005, Pile et al. 2005). The list is far

from exhaustive.

Amid such diverse applications, our goal in this thesis is to investigate how surface plasmons can be employed in nano-optical systems to overcome diffraction limits, to alter coherence properties of the light they couple with, and to boost the optical transmission with multi-layer structures. We proceed by reviewing in the next chapter the Maxwell equations in macroscopic media, and how the surface plasmon mode emerges at a metal-dielectric interface. The basic properties of surface plasmons and the conditions in which they exist will also be discussed.

In Chapter 3, the Green tensor formalism, which we have used to devise the rigorous simulations to obtain numerical solutions of the Maxwell equations, is covered in detail. In Chapter 4, we discuss strategies for employing surface plasmons in optical readout systems to overcome the diffraction limit and achieve superresolution.

We demonstrate a new role for surface plasmons in nano-optics in Chapter 5: Starting from a simple Young's interferometer setup and working towards realizing a practical device using a hole-array, we show how surface plasmons confined to the near-field can modulate the spatial coherence of light radiated to the far-field. Such a device can be very useful in nano-optical systems since the spatial coherence of a light source determines numerous properties of the field it produces, such as its directionality (Wolf 1978), spectrum (Wolf 1986), and polarization (James 1994).

In Chapter 6, we propose a multi-layered geometry that can potentially impede the decay of the optical field intensity as it would usually experienced in propagating through a single slab of metal. Our numerical results suggest that in impeding the field decay, extraordinary optical transmission is achieved. Finally in Chapter 7, we offer

concluding remarks on these findings.

CHAPTER 2: REVIEW OF THE MAXWELL EQUATIONS AND SURFACE PLASMONS

To begin, it is instructive to devote a section to introduce the Maxwell equations in macroscopic media, which are the governing equations of the behavior of electromagnetic waves in the presence of matter. In this thesis, we will focus on investigation of the behavior of surface plasmons in nano-optical systems through theoretical modeling and numerical simulations of the Maxwell equations, though it is recognized that experimental proofs undeniably provide for a more complete picture. The methodology adopted for the rigorous numerical simulations is elucidated in the next chapter. Numerical analysis inevitably contains assumptions and approximations, which we will set forth in the following section.

We then review in some detail the basic properties of surface plasmons. Surface plasmons are collective oscillations of electrons at a metal-dielectric interface, and can be excited by photons or electrons under conditions that satisfy the law of conservation of momentum. First predicted by Ritchie (1957), their existence were later confirmed experimentally through reflection of high-energy electrons (Powell and Swan 1959), and through optical radiation generated by surface plasmons (Teng and Stern 1967). In the latter experiment by Teng and Stern, the surface plasmons were excited by electrons impinging on metal gratings.

2.1 Maxwell Equations in Macroscopic Media

The fundamental set of equations governing electromagnetic phenomena are the Maxwell equations. Expressed in theoretically convenient Gaussian units, the Maxwell equations in free space take on the form

$$\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0 \tag{2.1a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.1b}$$

$$\nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{J} \tag{2.1c}$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho \tag{2.1d}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields, respectively, and c is the speed of light in vacuum.

Taking the divergence of Eq. (2.1c), and substituting Eq. (2.1d) into the resulting expression, one arrives at the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \,, \tag{2.2}$$

which relates the divergence of the current density $\mathbf{J}$ to the time rate of decrease of the charge density ρ.

When the sources ρ and $\mathbf{J}$ are completely specified, the set of Maxwell equations in (2.1) can be used to solve for the $\mathbf{E}$ and $\mathbf{B}$ fields everywhere in space. However, for macroscopic aggregates of matter, the solution of the equations is almost impossible due to the prohibitively large number of sources that contribute to the currents and charge densities. In such circumstances, it is useful to employ macroscopic quantities

whereby the fields and sources are spatially averaged[1]. We may decompose the sources in the matter according to (Melia 2001, Sec 1.2)

$$\rho = \rho_f + \rho_b \tag{2.3a}$$

$$\mathbf{J} = \mathbf{J}_f + \mathbf{J}_b \tag{2.3b}$$

where the subscripts f and b are used to indicate that these quantities are associated with the "free" and "bound" charges, respectively. The free charges produce the $\mathbf{E}$ and $\mathbf{B}$ fields. The bound charges create electric dipoles that give rise to a macroscopically averaged polarization field $\mathbf{P}$ (average dipole moment per unit volume). Magnetic dipoles[2] on the other hand, induce a macroscopically averaged magnetization field $\mathbf{M}$ (average magnetic dipole moment per unit volume). The fields $\mathbf{E}$, $\mathbf{B}$, $\mathbf{P}$, and $\mathbf{M}$ are related through

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \tag{2.4a}$$

$$\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M} \,, \tag{2.4b}$$

where $\mathbf{D}$ is the dielectric displacement vector, and $\mathbf{H}$ is the portion of the magnetic field produced by the sources. Furthermore, we know from electrostatics and magnetostatics that (Griffiths 1981, Sec 7.3)

$$\rho_b = -\nabla \cdot \mathbf{P} \,, \tag{2.5a}$$

$$\mathbf{J}_b = \nabla \times \mathbf{M} \,. \tag{2.5b}$$

---

[1]Averaging over time is not necessary. See for instance, Section 6.1 of Brau (2004) or Section 6.6 of Jackson 1975.

[2]Magnetic monopoles have yet been found despite an extended search by physicists.

In addition to the currents $\mathbf{J}_f$ and $\mathbf{J}_b$, the polarization $\mathbf{P}$ gives rise to a polarization current $\mathbf{J}_p$ in the time varying case, consistent with the continuity equation (2.2)

$$
\begin{aligned}
\nabla \cdot \mathbf{J}_p &= -\frac{\partial \rho_b}{\partial t} \\
&= \frac{\partial}{\partial t}(\nabla \cdot \mathbf{P}) \\
&= \nabla \cdot \frac{\partial \mathbf{P}}{\partial t} \ ,
\end{aligned}
\tag{2.6}
$$

from which it is straightforward to see that time variation in $\mathbf{M}$ produces no such redistribution of the charges.

In terms of the free charges ρ_f and current $\mathbf{J}_f$, Maxwell equations in macroscopic media are

$$
\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0 \ ,
\tag{2.7a}
$$

$$
\nabla \cdot \mathbf{B} = 0 \ ,
\tag{2.7b}
$$

$$
\nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} = \frac{4\pi}{c}\mathbf{J}_f \ ,
\tag{2.7c}
$$

$$
\nabla \cdot \mathbf{D} = 4\pi\rho_f \ .
\tag{2.7d}
$$

The four Maxwell equations can be divided into two pairs of equations, a homogeneous pair containing $\mathbf{E}$ and $\mathbf{B}$, and an inhomogeneous pair containing $\mathbf{D}$ and $\mathbf{H}$ (Born and Wolf 1999, Sec. 1.1). The inhomogeneous pair contains charges and currents, thereby implying the influence of matter. As such, one can attribute $\mathbf{D}$ and $\mathbf{H}$ to the influence of matter.

For an isotropic, linear medium, the electric polarization $\mathbf{P}$ can be described as

$$
\mathbf{P} = \chi_e \mathbf{E} \ ,
\tag{2.8}
$$

where χ_e is a dimensionless parameter that characterizes (macroscopically) the ability of the material to become polarized, which is also known as the electrical susceptibility of the medium. Similarly, the magnetization $\mathbf{M}$ can be described as

$$\mathbf{M} = \chi_m \mathbf{H} \ . \tag{2.9}$$

In this case, χ_m is known as the magnetic susceptibility, and it is the magnetic counterpart of χ_e. Using Eq. (2.4), one may express the dielectric displacement vector $\mathbf{D}$ and magnetic field $\mathbf{B}$ as

$$\mathbf{D} = (1 + 4\pi\chi_e)\mathbf{E}$$

$$= \epsilon\mathbf{E} \,, \tag{2.10a}$$

$$\mathbf{B} = (1 + 4\pi\chi_m)\mathbf{H}$$

$$= \mu\mathbf{H} \,, \tag{2.10b}$$

where ϵ and μ are the effective electric permittivity and magnetic permeability of the macroscopic medium.

In nature, materials at optical frequencies are typically non-magnetic so that $\mathbf{M} = 0$, and $\mathbf{H} = \mathbf{B}$. If the medium is conducting and obeys Ohm's law, the electric current density $\mathbf{J}_f$ can be described as

$$\mathbf{J}_f = \sigma\mathbf{E}, \tag{2.11}$$

where σ is again a macroscopic quantity defining the specific conductivity of the medium with units of inverse time (Gaussian units).[3] For media exhibiting anisotropy, these material parameters χ_e, χ_m and σ are, in general, tensors of rank 2.

---

[3]In the S. I. unit system, σ has units of siemens per meter (S/m).

For the electromagnetic simulations in this thesis, we will restrict ourselves to non-magnetic, isotropic, and electrically-neutral media whose light-matter interaction can be studied with the set of Maxwell equations in macroscopic media set forth in Eq. (2.7). By electrically-neutral, we refer to either lossless dielectrics where no free charges are present, or good conductors in which any initial free charge density has been redistributed at equilibrium along the surface.[4]

Making the substitutions $\mathbf{B} = \mathbf{H}$ (since $\mu = 1$), $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{J}_f = \sigma\mathbf{E}$, and $\rho_f = 0$ into the inhomogeneous set (Eqs. (2.7c) and (2.7d)), we find that

$$\nabla \times \mathbf{B} = \frac{4\pi\sigma}{c}\mathbf{E} + \frac{\epsilon}{c}\frac{\partial\mathbf{E}}{\partial t} , \tag{2.12a}$$

$$\nabla \cdot (\epsilon\mathbf{E}) = \epsilon\nabla\cdot\mathbf{E} + \mathbf{E}\cdot\nabla\epsilon$$

$$= 0 . \tag{2.12b}$$

Let us assume that $\mathbf{E}$ is a time-harmonic, monochromatic wave of the form

$$\mathbf{E} = \mathbf{E}(\mathbf{r})e^{i\omega t}$$

$$= \mathbf{E_r}e^{i\omega t} , \tag{2.13}$$

with ω its angular frequency.

Taking the curl of Eq. (2.7a) and eliminating $\mathbf{B}$ using Eq. (2.12a), and assuming that inhomogeneities in the media are negligible ($\nabla\epsilon \sim 0$) except at material boundaries,

---

[4]It can be shown that the free charge density ρ_0 in the interior of a good conductor decays exponentially as $\rho_c = \rho_0 e^{-\frac{t}{\tau_0}}$, with the time constant $\tau_0 = \frac{\epsilon}{4\pi\sigma}$ (Born and Wolf 1999, Sec. 14.1). This kind of relaxation phenomenon is also known as a Debye-type relaxation phenomenon, and is frequently encountered in physical processes (Kubo et al. 1998, Sec 3.3).

we obtain the following homogeneous Helmholtz equation for $\mathbf{E_r}$:[5]

$$\nabla^2 \mathbf{E_r} + k^2 \mathbf{E_r} = 0 \,, \tag{2.14}$$

with

$$k^2 = k_0^2 \epsilon_c \,, \tag{2.15}$$

where

$$k_0 = \frac{\omega}{c} \,, \tag{2.16}$$

is the wavevector in free space, and

$$\epsilon_c = \epsilon - \mathrm{i}\frac{4\pi\sigma}{\omega}, \tag{2.17}$$

is the complex permittivity of the medium that takes into account the effects of the electrical susceptibility (χ_e) and conductivity (σ). Similarly, taking the curl of Eq. (2.7c) and eliminating $\mathbf{E}$ yields for the space dependent part of the magnetic field $\mathbf{B_r}$:

$$\nabla^2 \mathbf{B_r} + k^2 \mathbf{B_r} = 0 \,. \tag{2.18}$$

Thus we see that the electric and magnetic fields propagate as waves in the macroscopic medium with a wavevector k, which may or may not be complex depending on the dielectric constant ϵ_r of the medium defined as

$$\epsilon_r = \frac{\epsilon_c}{\epsilon_0}, \tag{2.19}$$

with $\epsilon_0 = 1$ the permittivity of free space.

---

[5]Here we have used the vector identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$.

In arriving at the pair of wave equations (2.14) and (2.18) for the electric and magnetic fields, it is taken[6] that $\epsilon_r \neq 0$ so that the wavevector k assumes a finite value. More importantly, we have shown that the effects of the free and bound charges in a macroscopic medium can be taken into account by the use of a complex dielectric constant of the form

$$\epsilon_r = \epsilon_r' - i\epsilon_r'' \tag{2.20}$$

in the Maxwell equations.

To ensure validity in using the macroscopic Maxwell equations, we need to consider the shortest length scale over which the averaging procedure produces results within acceptable limits. Specifically, the shortest length scale should be much larger compared with the lattice constant, which is the spacing between the atoms in the material. Lattice constants of different materials are usually of the of the order of several angstroms, therefore we impose a minimum length scale of 100Å for the structures in our simulations. Furthermore, it is well-known that x-ray diffraction exposes the atomic structure of matter while light-matter interactions in the visible spectrum can be described by Maxwell equations with a continuous dielectric constant (Jackson 1975, Sec. 6.7). As such, we will only simulate light waves whose free space wavelength is greater than 400 nm in our study.

To summarize, the results of our study will be based primarily on numerical solutions of the Maxwell equations in macroscopic media. The assumptions we make are:

---

[6]The special case $\epsilon_r = 0$ corresponds to the bulk plasmon frequency in metals, where the free charges all oscillate back and forth to achieve equilibrium (Fox 2004, Sec. 7.5).

a) The media we consider are linear, non-magnetic ($\mu = 1$), isotropic, and electrically

   neutral ($\rho_f = 0$).

b) The electric and magnetic fields are time-harmonic with time dependence $e^{i\omega t}$.

c) The complex dielectric constant ϵ_r describes the effects of the current sources $\mathbf{J}_f$ and

   $\mathbf{J}_p$ introduced by the medium. (Note that $\mathbf{J}_b = 0$ due to our assumption $\mu = 1$.)

d) The length scale of structures in our simulations are at least 100Å .

e) The free space wavelength of light is at least 400 nm.

The assumptions listed and the relations presented here will serve as the point of

departure in the ensuing theoretical analysis of surface plasmons and their influence in

subwavelength optical systems.

2.2   Surface Plasmons

In this section, we describe the basic properties of surface plasmons. The characteristics

of the surface plasmons will be demonstrated as we mathematically derive expressions

for surface plasmons on a planar interface separating two media. Let us consider elec-

tromagnetic waves at a planar boundary $z = 0$ that separates two semi-infinite isotropic

media ϵ_1 and ϵ_2, as illustrated in Fig. 2.1. In a two-dimensional geometry, the electro-

magnetic waves can be fully described as a superposition of TE (transverse electric,

$E_x = H_y = E_z = 0$) and TM (transverse magnetic, $H_x = E_y = H_z = 0$) compo-

nents. The TE and TM components are also sometimes referred to as s-polarized and

p-polarized waves, respectively. Surface plasmons, as their name implies, are confined
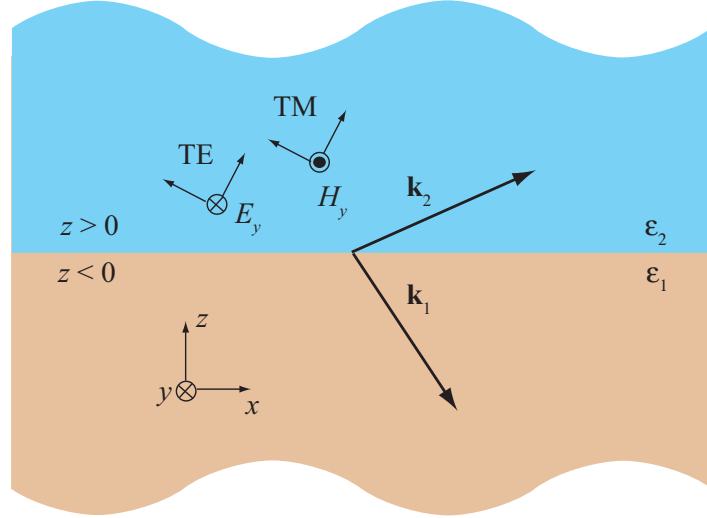
Figure 2.1: Geometry for illustrating the surface plasmons as surface wave confined to the interface between a dielectric (ϵ_1) and a metal (ϵ_2) .

to the boundary at $z = 0$. This confinement is characterized by the exponential decay of the fields with increasing distance away from the boundary.

We define the wave vectors describing these fields as $\mathbf{k}_1 = (k_x, 0, k_{z_1})$ and $\mathbf{k}_2 = (k_x, 0, k_{z_2})$ in medium 1 and 2, respectively. Across the boundary, phase matching requires that the transverse component of the wave vector k_x to be continuous. The exponential decay of the fields along $\pm z$ are described by the longitudinal wave vector component given by

$$
k_{zi} = \begin{cases} \sqrt{k_i^2 - k_x^2} & \text{if } k_i^2 \geq k_x^2 \\ -\mathrm{i}\sqrt{(k_x^2 - k_i^2)} & \text{if } k_i^2 < k_x^2 \end{cases} \tag{2.21}
$$

where k_i is the wavenumber in each medium, and the subscript $i = 1, 2$ is used to denote quantities in medium 1 and 2 respectively. In this case both k_{z_1} and k_{z_2} must be imaginary to ensure the exponential decay of the fields. Omitting from now on the explicit dependence on ω, the electric and magnetic fields in the two media are

$$
\mathbf{E}_{TE} = \begin{cases} E_{y_1}\, \hat{\mathbf{y}}\, e^{-i(k_x x - k_{z_1} z)} & z < 0 \\[2mm] E_{y_2}\, \hat{\mathbf{y}}\, e^{-i(k_x x + k_{z_2} z)} & z > 0 \end{cases} \tag{2.22}
$$

and

$$
\mathbf{H}_{TE} = \begin{cases} \left[H_{x_1}\, \hat{\mathbf{x}} + H_{z_1}\, \hat{\mathbf{z}} \right]\, e^{-i(k_x x - k_{z_1} z)} & z < 0 \\[2mm] \left[H_{x_2}\, \hat{\mathbf{x}} + H_{z_2}\, \hat{\mathbf{z}} \right]\, e^{-i(k_x x + k_{z_2} z)} & z > 0 \end{cases} \tag{2.23}
$$

for the TE case, and

$$
\mathbf{E}_{TM} = \begin{cases} \left[E_{x_1}\, \hat{\mathbf{x}} + E_{z_1}\, \hat{\mathbf{z}} \right]\, e^{-i(k_x x - k_{z_1} z)} & z < 0 \\[2mm] \left[E_{x_2}\, \hat{\mathbf{x}} + E_{z_2}\, \hat{\mathbf{z}} \right]\, e^{-i(k_x x + k_{z_2} z)} & z > 0 \end{cases} \tag{2.24}
$$

and

$$
\mathbf{H}_{TM} = \begin{cases} H_{y_1}\, \hat{\mathbf{y}}\, e^{-i(k_x x - k_{z_1} z)} & z < 0 \\[2mm] H_{y_2}\, \hat{\mathbf{y}}\, e^{-i(k_x x + k_{z_2} z)} & z > 0 \end{cases} \tag{2.25}
$$

for the TM case, with $\hat{\mathbf{x}}, \hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ the respective unit vectors.

The continuity of the tangential components of the electric and magnetic fields across boundaries, when applied to E_x, E_y, B_x and B_y in Eqs. (2.22) - (2.25) at $z = 0$ give us

$$
E_{x_1} = E_{x_2} = E_x\,, \tag{2.26}
$$

$$
E_{y_1} = E_{y_2} = E_y\,, \tag{2.27}
$$

$$
H_{x_1} = H_{x_2} = H_x\,, \tag{2.28}
$$

and

$$
H_{y_1} = H_{y_2} = H_y\,. \tag{2.29}
$$

For the TE case, we express H_{x_1} and H_{x_2} in terms of E_y using Eq. (2.7a), and then using Eq. (2.28), we find that

$$k_{z_1} + k_{z_2} = 0 \,. \tag{2.30}$$

For the TM case, we express E_{x1} and E_{x2} in terms of H_y using Eq. (2.7c) (with $\mathbf{J}_f = 0$), and then using Eq. (2.26), we find that

$$\epsilon_2 k_{z_1} + \epsilon_1 k_{z_2} = 0 \,. \tag{2.31}$$

On substituiting Eq. (2.21) into Eq. (2.30), we obtain

$$\sqrt{k_x^2 - k_1^2} + \sqrt{k_x^2 - k_2^2} = 0 \,, \tag{2.32}$$

which cannot be satisfied for $k_1 \neq k_2$ (noting that $k_i^2 < k_x^2$ in order to have imaginary k_{z_1} and k_{z_2}) and is hence not a physical solution. Directing our attention now to Eq. (2.31) for the TM case, we find that

$$k_x = \sqrt{\frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2}} k_0 \,. \tag{2.33}$$

This relation given by Eq. (2.33) represents the dispersion of the surface plasmon mode at a planar metal-dielectric interface (Raether 1988), with k_x being the wavenumber of the surface plasmon. Substituting for Eq. (2.33) in Eq. (2.21), we find that

$$k_{zi}^2 = \frac{\epsilon_i^2}{\epsilon_1 + \epsilon_2} k_0^2 \,. \tag{2.34}$$

For a wave propagating along the interface, we require that k_x given by Eq. (2.33) is real-valued. This is possible if the product and sum of ϵ_1 and ϵ_2 are either both positive or both negative. However, to satisfy the condition that its amplitude decays exponentially away from the boundary, it is seen from Eq. (2.34) that the sum of ϵ_1

and ϵ_2 has to be negative. Therefore, the product $\epsilon_1\epsilon_2$ is necessarily negative, which means that ϵ_1 and ϵ_2 must differ in sign. Furthermore, the absolute value of the negative permittivity must exceed that of the positive permittivity.

To illustrate with further detail the solution provided by Eq. (2.33), we assign for convenience and without loss of generality, the relative permittivity of the dielectric medium $\epsilon_1 = 1$. The dielectric constant of the metal (medium 2) is

$$\epsilon_2 = \epsilon_2' - i\epsilon_2'', \tag{2.35}$$

with $\epsilon_2' < -1$. At visible and higher frequencies, the conductivity of metals is usually predominantly imaginary, so that $|\epsilon_2'| >> |\epsilon_2''|$ and $\epsilon_2 \sim \epsilon_2'$. When $|\epsilon_2| > \epsilon_1(= 1)$, we see that Eq. (2.33) yields a real k_x (imaginary component of k_x negligible since $|\epsilon_2'| >> |\epsilon_2''|$), and thus a wave that propagates along the interface between the two media is obtained. Furthermore, according to Eq. (2.34), both k_{z_1} and k_{z_2} are imaginary in this case, consistent with the exponentially decaying nature of the fields from the interface. In this sense, the surface plasmon represents a *true* surface mode. [7]

It can be seen from the above derivation that the surface plasmons have electric field components both parallel (E_x) and perpendicular (E_z) to its direction of propagation (cf. Eq.(2.24)). The surface plasmons therefore exhibit a longitudinal wave nature, in contrast to transverse electromagnetic waves that radiate from an oscillating electric dipole. With the exception of the boundary at $z = 0$, the fields in each of the regions in Fig. 2.1 satisfy the divergence equation

$$\nabla \cdot \mathbf{E} = 0. \tag{2.36}$$

[7] The surface plasmon mode is sometimes referred to as the *Fano* mode.

Using the divergence condition in Eq. (2.36), it is found that the electric field components are related through

$$E_z = \begin{cases} \dfrac{k_x}{k_{z_1}} E_x & z < 0 \\[3mm] -\dfrac{k_x}{k_{z_2}} E_x & z > 0 \end{cases} . \tag{2.37}$$

Assming again $\epsilon_1 = 1$, we obtain from Eq. (2.37), for small k_x or when $|\epsilon_2'| >> |\epsilon_2''|, |\epsilon_1|$, that

$$E_z = \begin{cases} i \sqrt{|\epsilon_2'|} \, E_x & z < 0 \\[3mm] -i \dfrac{1}{\sqrt{|\epsilon_2'|}} E_x & z > 0 \end{cases} . \tag{2.38}$$

In this case, the field extends much more into the dielectric (ϵ_1) than the metal (ϵ_2). It will be clear that surface plasmons that possess low wave vectors are typically excited with light photons when we discuss their dispersion characteristics in the next section. On the other hand, for large k_x, we find that the magnitude of E_x and E_z are comparable, i.e.,

$$E_z = \begin{cases} i \, E_x & z < 0 \\[3mm] -i \, E_x & z > 0 \end{cases} . \tag{2.39}$$

Surface plasmons with such large wave vectors are typically excited with electrons.

Along a smooth surface, ohmic losses in the metal will cause the intensity of propagating surface plasmons to decrease exponentially as $e^{-2k_x''}$, where $k_x = k_x' - ik_x''$. These ohmic losses dissipate as heat in the metal, and can be taken into account with the imaginary part of the dielectric function of the metal (ϵ_2''), as discussed in the previous section. The length L_{sp} at which the intensity decreases to $1/e$ is then given by

$$L_{sp} = (2k_x'')^{-1} . \tag{2.40}$$

As an example, L_{sp} for a semi-infinite silver-air interface at a wavelength of 500 nm is typically around $20\,\mu$m.

Let us also briefly consider the behavior of surface plasmons propagating on the surfaces of a metal film of thickness d that is surrounded by a dielectric medium such as free-space. If the metal is thick enough, the characteristics of the surface plasmon modes on the two surfaces approach that of the semi-infinite metal-dielectric interface and are identical. As discussed, the surface plasmon modes decay evanescently both into the metal and the free space. If the metal film is sufficiently thick, the evanescent waves extending into the metal do not overlap, and the surface plasmon modes do not couple.

As the film thickness d decreases, the evanescent waves extending into the metal starts to overlap, and the metal film acts effectively as a waveguide. Upon interaction, the degenerate surface plasmon mode on each surface couples and splits into a symmetric mode and an anti-symmetric one. The anti-symmetric mode is defined here as having a zero in its transverse electric field distribution in the metal film, and for a given plasmon wave vector, is associated with a higher frequency ω (Raether 1988). By solving for the real and imaginary parts of the propagation constants of these modes, it is found that the antisymmetric mode has an imaginary part that decreases sharply towards zero as the film thickness $d \to 0$. For such thin film structures, the surface plasmons can potentially travel for long ranges up to dozens of microns on the surfaces without appreciable attenuation (Fukui et al. 1979, Sarid 1981, Berini et al. 2007). These long propagation lengths can be desirable properties for practical applications.

We have shown that surface plasmons are indeed solutions to the Maxwell equations,

and have derived their dispersion relation (Eq. (2.33)) at a planar dielectric-metal interface.[8] The surface plasmons are TM-polarized modes that can propagate along such an interface, and they possess a longitudinal wave nature. They can travel for long ranges on the surfaces of thin metallic films where the surface plamson modes on each surface are allowed to couple.

Since their wave vector k_x is greater than the wave vector of the light k_0, there is a momentum gap between the two, and therefore, surface plasmons do not radiate on their own. In the next section, we will discuss how surface plasmons can be excited and coupled with light photons when we examine their dispersion relation in detail. Finally, we note that, along the interface separating a right-handed and a left-handed material $(\epsilon, \mu < 0)$, TE-polarized surface plasmons are legitimate solutions of Maxwell equations, in the same way the TM counterparts exist according to Eq. (2.33) (Ruppin 2000).

2.3   Dispersion Relation for the Surface Plasmons

The dispersion relation of the surface plasmons at a planar dielectric-metal interface is as derived in Eq. (2.33)

$$k_{sp} = \sqrt{\frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2}} k_0, \tag{2.41}$$

where k_{sp}, a complex quantity in general, is used to denote the wavenumber of the surface plasmons. As in the previous section, we will assume that $\epsilon_1 = 1$ whenever

---

[8]It is worthwhile noting that, depending on ϵ_1 and ϵ_2, different surface modes may be obtained on the boundary. Interested readers are referred to Sec. 1, Chap. 7 of Boardman (1982) or Yang et al. (1991), for instance.

it is convenient to do so for the sake of discussion. Apparently, the surface plasmon vector $k_{sp} \rightarrow \infty$ as $\epsilon_2 \rightarrow -\epsilon_1$. The frequency at which $\epsilon_2 = -\epsilon_1$ is the surface plasmon frequency, ω_{sp}. This means that a propagating longitudinal wave at ω_{sp} is sustained at the dielectric-metal interface when the total dielectric function vanishes (Peyghambarian et al. 1993, Sec. 3.6), i.e., $\epsilon_1 + \epsilon_2 = 0$. Physically, this situation arises because the electric field $\mathbf{E}$ cancels the effects of the polarization $\mathbf{P}$ ($\mathbf{D} = \epsilon\mathbf{E} = \mathbf{E} + 4\pi\mathbf{P}$), and the electrons then just oscillate backward and forward with respect to the fixed lattice of positive ions (Fox 2004, Sec. 7.5).

Let us estimate the surface plasmon frequency ω_{sp} using the Drude model for oscillations due to free electrons in metals (Reitz et al. 1980). According to the Drude model,

$$\epsilon_2(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - i\omega\gamma} \tag{2.42}$$

where the plasma frequency ω_p is defined as

$$\omega_p = \frac{4\pi N e^2}{m_0} \tag{2.43}$$

with

$$\gamma = \text{damping rate of the respective oscillations,}$$

$$N = \text{number of charges per unit volume,}$$

$$e = \text{electronic charge} = 4.803 \times 10^{-10} \text{ statcoulombs,}$$

$$m_0 = \text{electron rest mass} = 9.109 \times 10^{-28} \text{ g.}$$

For the case where $|\epsilon_2'| >> |\epsilon_2''|$, the Drude model in Eq. (2.42) reduces to

$$\epsilon_2(\omega) = 1 - \frac{\omega_p^2}{\omega^2}.$$ (2.44)

Using Eq. (2.44), we find that the frequency at which $\epsilon_2 = -\epsilon_1$ occurs at

$$\omega_{sp} = \frac{\omega_p}{\sqrt{\epsilon_1 + 1}}.$$ (2.45)

When the lossless dielectric ϵ_2 is free-space, the surface plasmon frequency is then

$$\omega_{sp} = \frac{\omega_p}{\sqrt{2}}.$$ (2.46)

The dispersion relation in Eq. (2.41) holds only for a specific value of the dielectric constant $\epsilon_2(\omega)$ at a specific frequency. It will be useful to find an alternative expression that is a continuous function of ω. To do this, we express the dispersion relation in Eq. (2.41) as

$$\omega^2 = (ck_{sp})^2(\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2}).$$ (2.47)

Substituting for ϵ_2 in Eq. (2.47) through the use of Eq. (2.44), we find, after some algebra that,

$$\omega^2 = \frac{\epsilon_1\omega_p^2 + c^2k_{sp}^2 + \epsilon_1c^2k_{sp}^2 - \sqrt{(\epsilon_1\omega_p^2 + c^2k_{sp}^2 + \epsilon_1c^2k_{sp}^2)^2 - 4\epsilon_1c^2k_{sp}^2\omega_p^2}}{2\epsilon_1}$$ (2.48)

which reduces to, for the case $\epsilon_1 = 1$,

$$\omega^2 = \frac{\omega_p^2}{2} + c^2k_{sp}^2 - \sqrt{\frac{\omega_p^4}{4} + c^4k_{sp}^4}.$$ (2.49)

It is worthwhile noting that the negative root, instead of the positive one for the last term in the numerator of Eq. (2.48), is chosen to arrive at Eq. (2.49) for the dispersion of the surface plasmons. It is straightforward to show that ω in Eq. (2.49) approaches

$\omega_{sp} = \frac{\omega_p}{\sqrt{2}}$ in the asymptotic limit $k_{sp} \to \infty$ as in Eq. (2.46), thus verifying our choice of the negative root. The positive root, on the other hand, leads to the asymptotic limit $\omega \to \sqrt{2}ck_{sp}$, as is the case for the propagating Brewster mode (Boardman 1982, Chap. 1, Sec. 3.2).

Let us define normalized frequency

$$\Omega = \frac{\omega}{\omega_p}, \qquad (2.50)$$

and normalized wave vector

$$q = \frac{k}{k_p}, \qquad (2.51)$$

where $k_p = \frac{\omega_p}{c}$. With these nomalized quantities Ω and q, Eq. (2.49) can be simplified to (Ferrell 1958)

$$\Omega^2 = [1 + \frac{1}{2q^2} + (1 + \frac{1}{4q^2})^{\frac{1}{2}}]^{-\frac{1}{2}}. \qquad (2.52)$$

The dispersion relation of Eq. (2.52) is plotted in Fig. 2.2. It shows that the dispersion curve for the surface plasmons lies to the right of the light line ($\omega = ck_{sp} \to \Omega = q$), and that the two lines do not intersect except at the origin. The wave vector of the surface plasmons is always greater than that of light at a given frequency, and this illustrates why the surface plasmons are non-radiative by themselves - since the wave vector is related to momentum, surface plasmons have higher momentum than light ($k_{sp} > k_0$ or $\lambda_{sp} < \lambda_0$) at a given frequency ω. On a smooth surface, surface plasmons and light do not couple as this momentum gap cannot be bridged.

It is of course possible to couple light with surface plasmons if the additional momentum is provided to the photons. There are several ways in which surface plasmons can be excited with light photons (Zayats and Smolyaninov 2003):
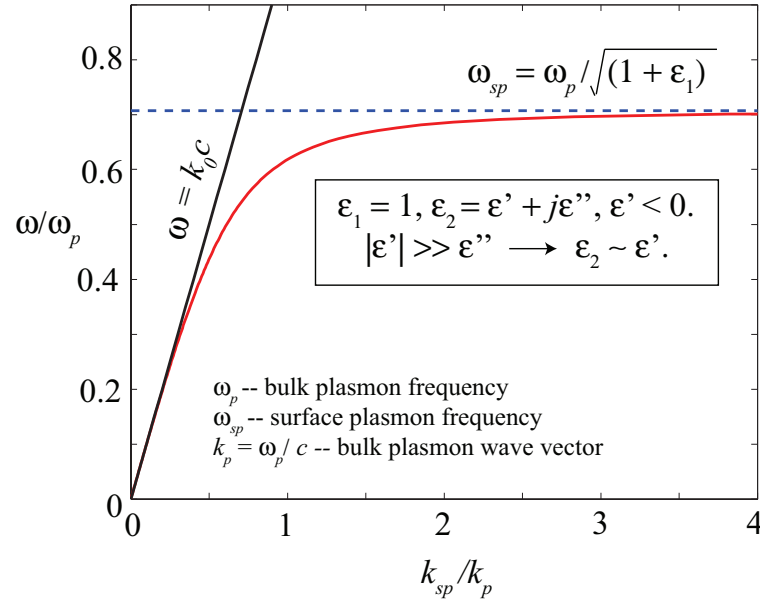
Figure 2.2: Dispersion relation of surface plasmons.

a) The metal surface can be illuminated through attenuated total reflection (ATR) by placing it close to a dielectric prism (ϵ_m) with light incident at an angle θ. In the prism, the wave vector of the light is increased, given by $k_m = \sqrt{\epsilon_m}k_0$. When the in-plane wave vector $k_m \sin\theta$ matches that of the surface plasmon (k_{sp}), surface plasmons are excited by the light that evanescently tunnels to the metal surface. The most commonly employed geometries are the Kretschmann geometry and the Otto geometry (Raether 1988) .

b) Diffraction gratings on a metal plate can be used to excite surface plasmons when illuminated with light. In this case, the surface plasmon modes whose wave vectors

coincide with that of the diffracted orders will be excited.

c) Topological defects (such as protrusions, holes, slits, or surface roughness) on the surface of a metal plate can serve to locally excite surface plasmons when they diffract light incident on them (Hecht et al. 1996).

d) In place of the dielectric prism in (a), a probing tip can be brought very close to a metal surface to locally excite plasmons on a metal surface. The development of this technique is due in large to applications related to scanning naer-field optical microscopy (SNOM).

These methods of exciting surface plasmons with light photons are illustrated in Fig. 2.3. The reverse process of surface plasmons coupling into light and consequently radiating as photons can occur as well, as long as the momentum is conserved.

If we note that within the framework of the effective mass approximation the energy E of an electron is inversely proportional to its effective mass m_e ($E = \hbar\omega = \frac{\hbar^2 k^2}{2\,m_e}$), and that m_e is in turn a measure of the curvature of the electron's dispersion (Tang 2005, Sec. 10.6), it perhaps explains why electrons were employed to excite surface plasmons in the earlier experiments (Powell and Swan 1959, Teng and Stern 1967). When we consider that the surface plasmon frequency ω_{sp} lies typically in the range of $\sim 25 - 120$ eV (clearly beyond the visible spectrum) (Fox 2004, Sec. 7.3.1), it becomes apparent that while electrons are useful for studying the properties of surface plasmons at large k_{sp} values ($k_{sp} > k_p$), light photons ($\sim 1 - 3$ eV) are more suitable for analysis and experiments at the other extreme ($k_{sp} < k_p$).

We have discussed in detail some of the fundamental properties of surface plas-
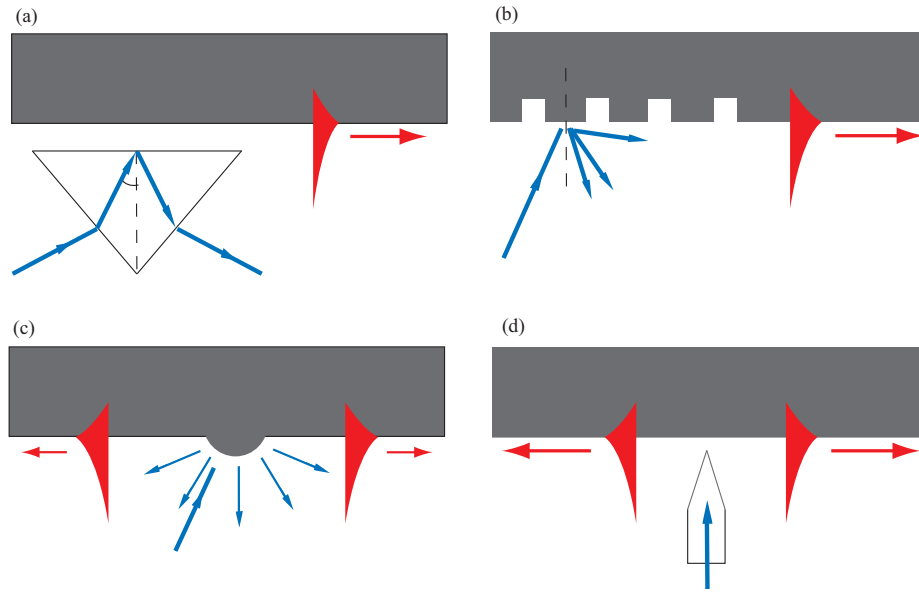
Figure 2.3: Illustrating different methods of exciting surface plasmons with light: (a)attenuated total reflection (ATR) with the Otto geometry, (b)diffraction grating, (c)surface protrusion, (d)probing tip.

mons, and how their interaction with light can occur. To summarize, the important characteristics of the surface plasmons are:

a) They are TM-polarized.

b) For a planar interface separating two semi-infinite isotropic media ϵ_1 and ϵ_2, it is required that both the sum and product of ϵ_1 and ϵ_2 be negative for surface plasmons to propagate along the interface.

c) The surface plasmons possess a longitudinal wave nature, in contrast to transverse electromagnetic waves that radiate from an oscillating electric dipole.

d) They are non-radiative on their own, but can couple with light photons and electrons if the momentum gap is bridged.

e) In the visible spectrum, they can typically travel for long ranges up to dozens of microns along metallic thin films without appreciable attenuation.

f) Their wavelength λ_{sp} is always less than the wavelength λ_0 of light at a given frequency ω. The smaller wavelength means they can be more readily 'squeezed' through subwavelength apertures, which is advantageous in nano-optical systems.

g) They can give rise to field enhancements at surfaces where they are excited.

In particular, their ability to travel long range, their smaller wavelength, and their potential to produce field enhancement effects make them attractive for various nano-optical applications. We are in a better position now to investigate how surface plasmons can be employed in the regime of nano-optics, and will present results of our theoretical research of potential applications in the following chapters.

Chapter 3: The Green Tensor Formalism

We have performed rigorous simulations of the Maxwell equations to investigate the behavior of subwavelength optical systems, keeping in mind the assumptions set out at the end of Section 2.1. The structures that we will be mostly concerned with are systems of thin films that are perforated with subwavelength apertures or corrugated on the surfaces. To simulate the Maxwell equations in such geometries, we have adopted a planar multi-layered geometry, in which the electromagnetic Green tensor can be calculated analytically, up to within a Fourier Transform. The Green tensor $\mathbf{G}(\mathbf{r}, \mathbf{r}')$ describes the field at position $\mathbf{r}$ due to the three orthogonal dipoles radiating at $\mathbf{r}'$, and takes into account the effects of multiple reflections and refractions that occur at the different interfaces. Closely following the formalism by (Visser et al. 1999), details of the derivation of the dyadic Green functions are given in this chapter. Those readers interested primarily in physical results can read Section 3.1 and skip forward to Chapter 4.

Since in the monochromatic case, the magnetic field can be calculated by taking the curl of the electric field, only the Green tensor for the electric field needs to be derived. With the appropriate dyadic Green functions determined, a domain integral equation approach is then used to determine the electromagnetic fields.

The Green tensor method has been widely studied (Tai 1971, Sphicopoulos et al.

1985, Chew 1995, Paulus et al. 2000) and used (Kolk et al. 1990, Bastiaansen et al. 1992, Urbach and Lepelaars 1994, Visser et al. 1999, Schouten et al. 2005) for the simulations of electromagnetic fields. This method is attractive for our study as the dyadic Green functions account for the complex interactions of a vector field in multiple scattering problems in a compact manner.

As opposed to methods such as the FDTD (finite difference time domain) and FEM (finite element method), the Green tensor in principle provides exact solutions to the Maxwell equations, and is discretized only in converting the domain integral equation to a matrix equation for obtaining numerical values of the fields. Furthermore, the application of Sommerfeld's radiation condition at infinity avoids the need to implement complicated matching boundary conditions at the boundaries of the region of interest[1].

There are some limitations associated with the Green tensor method, despite of the advantages it can offer. The geometries we have simulated are restricted to planar layered systems, which may not necessarily be the most suitable in practical situations. As matrix inversion is involved in the solutions, the computational time and memory required increase rapidly as the number of discretization points are increased. Furthermore, the algorithm to invert the matrices can become numerically unstable as the size

---

[1]The Sommerfeld's radiation condition ensures that the field must be outgoing at infinity. For illustration purpose, let us consider the radiation condition for a scalar field ψ, which is given by $\lim_{r \to \infty} r \left(\dfrac{\partial \psi}{\partial r} + \mathrm{i} k \psi \right) = 0$, where $k = \omega/c$. By substituting outgoing and incoming waves proportional to $e^{\mp \mathrm{i} k r}$ into the condition, it will be seen that only the outgoing field satisfies the condition. For more details, see Sec. 2.5 of Ishimaru (1991).

of the matrices are increased.

A final word on units before we proceed. Though we have employed Gaussian units in the previous chapter in view of the physical intuition it offers to theorists, we will adhere to the S.I. unit from now on for convenience in the implementation of the Green tensor as machine-executable code. For this reason, we give here Maxwell equations in macroscopic media in S.I. units

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \ , \tag{3.1a}$$

$$\nabla \cdot \mathbf{B} = 0 \ , \tag{3.1b}$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}_f \ , \tag{3.1c}$$

$$\nabla \cdot \mathbf{D} = \rho_f \ . \tag{3.1d}$$

3.1   Configuration of Multi-layered Geometry

We adopt a two-dimensional multi-layer geometry that consists of n arbitrary layers along the z direction (see Fig. 3.1). Each layer is characterized by its permittivity $\epsilon^B(z)$. In each layer, we allow for an arbitrary number of 'deviant' regions whose permittivity $\epsilon^D(x, z)$ differ from $\epsilon^B(z)$. We will refer to these deviant regions by $\mathcal{D}_m$, where $m = 1, 2, 3, ..., m_0$, where m_0 is the number of deviant regions. In the absence of free charges, time harmonic monochromatic waves of frequency ω given by the form in Eq. (2.13) satisfy the set of Maxwell equations

$$\nabla \times \hat{\mathbf{E}} + i\omega\mu_0\hat{\mathbf{H}} = 0 \ , \tag{3.2a}$$

$$\nabla \times \hat{\mathbf{H}} - i\omega\epsilon(x, z)\hat{\mathbf{E}} = 0 \ , \tag{3.2b}$$
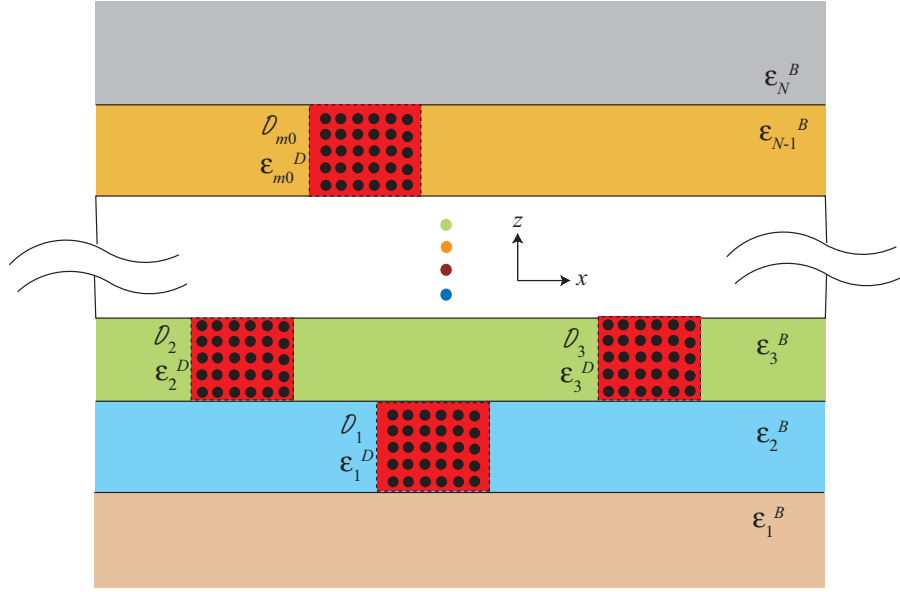
Figure 3.1: Total configuration for Green tensor formalism with multi-layered geometry.

where μ_0 is the permeability in free space. Formally, the total fields $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ in Eq. (3.2) can be expressed as the sum of the incident field and scattered field as

$$\hat{\mathbf{E}} = \hat{\mathbf{E}}^{inc} + \hat{\mathbf{E}}^{sca} \ , \tag{3.3a}$$

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}^{inc} + \hat{\mathbf{H}}^{sca} \ . \tag{3.3b}$$

We define the incident field to be the part of the field that is present in the layered medium without the deviant regions. The incident field then satisfies the following set of Maxwell equations

$$\nabla \times \hat{\mathbf{E}}^{inc} + \mathrm{i}\omega\mu_0\hat{\mathbf{H}}^{inc} = 0 \ , \tag{3.4a}$$

$$\nabla \times \hat{\mathbf{H}}^{inc} - \mathrm{i}\omega\epsilon^B(z)\hat{\mathbf{E}}^{inc} = 0 \ . \tag{3.4b}$$

We shall refer the multi-layered medium without the deviant regions as the 'background' configuration (see Fig. 3.2). Expressing Eq. (3.2) in terms of the background
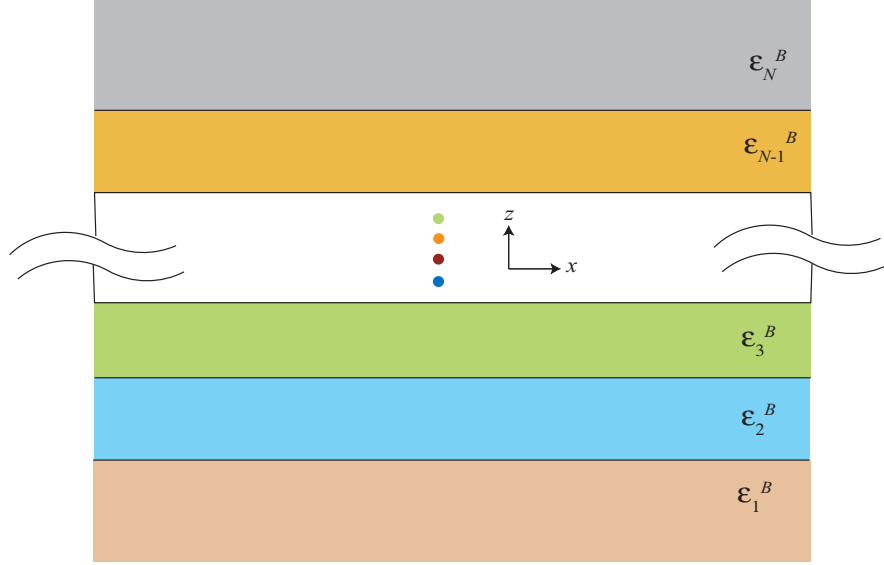
Figure 3.2: Background multi-layer configuration.

configuration, we have

$$\nabla \times \hat{\mathbf{E}} + i\omega\mu_0\hat{\mathbf{H}} = 0 \ , \tag{3.5a}$$

$$\nabla \times \hat{\mathbf{H}} - i\omega\epsilon^B(z)\hat{\mathbf{E}} = \hat{\mathbf{J}}^{con} \ , \tag{3.5b}$$

where

$$\hat{\mathbf{J}}^{con} = \begin{cases} i\omega \left[\epsilon^D(x,z) - \epsilon^B(z)\right] \hat{\mathbf{E}}, & \mathbf{r} \in \mathcal{D}_m, \\ \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

Subtracting Eq. (3.4) from Eq. (3.5), we obtain for the scattered field

$$\nabla \times \hat{\mathbf{E}}^{sca} + i\omega\mu_0\hat{\mathbf{H}}^{sca} = 0 \ , \tag{3.7a}$$

$$\nabla \times \hat{\mathbf{H}}^{sca} - i\omega\epsilon^B(z)\hat{\mathbf{E}}^{sca} = \hat{\mathbf{J}}^{con} \ , \tag{3.7b}$$

which shows that the scattered field is the part of the field due to the 'electric contrast source' $\hat{\mathbf{J}}^{con}$. This contrast source $\hat{\mathbf{J}}^{con}$ vanishes everywhere outside the deviant regions $\mathcal{D}_m$.

Next, we define the so-called Green states $\hat{E}^G$ and $\hat{H}^G$ that satisfy

$$\nabla \times \hat{\mathbf{E}}^G + i\omega\mu_0\hat{\mathbf{H}}^G = 0 \; , \tag{3.8a}$$

$$\nabla \times \hat{\mathbf{H}}^G - i\omega\epsilon^B(z)\hat{\mathbf{E}}^G = \mathbf{a}\,\delta(x - x')\,\delta(z - z') \; , \tag{3.8b}$$

with $\mathbf{a}$ as a unit vector. These Green states describe the scattered electric and magnetic fields generated by a current line source of infinite length in y and directed along $\mathbf{a}$. The electric and magnetic Green tensor, $\hat{G}^E_{ij}$ and $\hat{G}^H_{ij}$, may then be defined through the Green states as

$$\hat{E}^G_i = \hat{G}^E_{ij}a_j \; , \tag{3.9a}$$

$$\hat{H}^G_i = \hat{G}^H_{ij}a_j \; . \tag{3.9b}$$

By superposition of the Green states, the scattered electric and magnetic fields are found to be

$$\hat{E}^{sca}_i(x, z) = \sum_{m=1}^{m_0} \iint_{\mathcal{D}_m} \hat{G}^E_{ij}(x, z; x', z')\hat{J}^{con}_j(x', z')\,dx'\,dz' \; , \tag{3.10a}$$

$$\hat{H}^{sca}_i(x, z) = \sum_{m=1}^{m_0} \iint_{\mathcal{D}_m} \hat{G}^H_{ij}(x, z; x', z')\hat{J}^{con}_j(x', z')\,dx'\,dz' \; . \tag{3.10b}$$

Using Eq. (3.6), the total electric and magnetic fields can then be expressed as

$$\hat{E}_i(x, z) = \hat{E}^{inc}_i + i\omega \sum_{m=1}^{m_0} \iint_{\mathcal{D}_m} \Delta\epsilon(x', z')\,\hat{G}^E_{ij}(x, z; x', z')\hat{E}_j(x', z')\,dx'\,dz' \; , \tag{3.11a}$$

$$\hat{H}_i(x, z) = \hat{H}^{inc}_i + i\omega \sum_{m=1}^{m_0} \iint_{\mathcal{D}_m} \Delta\epsilon(x', z')\,\hat{G}^H_{ij}(x, z; x', z')\hat{E}_j(x', z')\,dx'\,dz' \; . \tag{3.11b}$$

where $\Delta\epsilon\left(x',z'\right) = \epsilon^D(x',z') - \epsilon^B(z')$, with $(x',z') \in \mathcal{D}_m$. These are the so-called *domain integral equations.* It is to be noted that the Green tensors $\hat{E}^G$ and $\hat{H}^G$ pertain to the relatively simple background configuration that does not contain $\mathcal{D}_m$. Eq. (3.11) shows that the electric field $\hat{\mathbf{E}}$ and magnetic field $\hat{\mathbf{H}}$ at any point in the multi-layer medium can be calculated once the fields in the deviant regions are known. By discretizing the field in the deviant regions in Eq. (3.11), we obtain a matrix equation which may be inverted to determine the fields. The field everywhere else may then be determined by substituting the fields in the deviant regions back into Eq. (3.11). Since the magnetic field can be calculated by taking the curl of the electric field, only the Green tensor for the electric field will be derived.

Before tackling the challenging task of determining the dyadic Green functions, we can see from Eq. (3.4) that the incident field satisfies the Maxwell equations for the background configuration ($\mathcal{D}_m$ not included). We show how one can analytically 'construct' the incident field for a plane wave in the next section.

3.2   The Incident Field

For an incident plane wave, the incident field in the multi-layer background configuration can be calculated analytically using a recursive procedure by matching the boundary conditions for the tangential components of $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$. Suppose we have an incident plane wave with wave vector $(k_x, 0, k_{z_1})$ incoming from the bottom of the multi-layer configuration. With the exception of the top layer, the transmission and reflection across the boundaries will result in upward and downward waves 'bouncing' between the layers. Let there be N arbitrary layers, with the Nth layer being the semi-infinite

layer on the top. We denote the amplitude of the upward and downward waves in the nth layer as u_n^{pol} and d_n^{pol} respectively, with the superscript $pol = E$ for the TE-polarized component (electric field transverse to the $x - z$ plane), and $pol = H$ for the TM-polarized component (magnetic field transverse to the $x - z$ plane). In the first layer $(n = 1)$, we take $u_1^{E,H} = 1$. In the Nth layer, boundary conditions dictate that $d_N^{E,H} = 0$. The tangential components in each layer may then be expressed as

$$\hat{E}_y = \begin{cases} \left[e^{-i\,k_{z_1}\,(z-z_1)} + d_1^E\, e^{i\,k_{z_1}\,(z-z_1)} \right] e^{-i\,k_x x} & z < z_1 \\ \left[u_n^E\, e^{-i\,k_{zn}\,(z-z_{n-1})} + d_n^E\, e^{i\,k_{zn}\,(z-z_n)} \right] e^{-i\,k_x x} & z_{n-1} < z < z_n \\ \left[u_N^E\, e^{-i\,k_{zN}\,(z-z_{N-1})} \right] e^{-i\,k_x x} & z > z_{N-1} \end{cases} \tag{3.12}$$

$$\hat{H}_x = -\frac{1}{\omega\mu_0} \begin{cases} \left[e^{-i\,k_{z_1}\,(z-z_1)} - d_1^E\, e^{i\,k_{z_1}\,(z-z_1)} \right] k_{z_1} e^{-i\,k_x x} & z < z_1 \\ \left[u_n^E\, e^{-i\,k_{zn}\,(z-z_{n-1})} - d_n^E\, e^{i\,k_{zn}\,(z-z_n)} \right] k_{zn} e^{-i\,k_x x} & z_{n-1} < z < z_n \\ \left[u_N^E\, e^{-i\,k_{zN}\,(z-z_{N-1})} \right] k_{zN} e^{-i\,k_x x} & z > z_{N-1} \end{cases} \tag{3.13}$$

for the TE case, and

$$\hat{E}_x = \begin{cases} \left[e^{-i\,k_{z_1}\,(z-z_1)} + d_1^H\, e^{i\,k_{z_1}\,(z-z_1)} \right] e^{-i\,k_x x} & z < z_1 \\ \left[u_n^H\, e^{-i\,k_{zn}\,(z-z_{n-1})} + d_n^H\, e^{i\,k_{zn}\,(z-z_n)} \right] e^{-i\,k_x x} & z_{n-1} < z < z_n \\ \left[u_N^H\, e^{-i\,k_{zN}\,(z-z_{N-1})} \right] e^{-i\,k_x x} & z > z_{N-1} \end{cases} \tag{3.14}$$

$$\hat{H}_y = \frac{1}{\omega\mu_0} \begin{cases} \left[e^{-i\,k_{z_1}\,(z-z_1)} - d_1^H\, e^{i\,k_{z_1}\,(z-z_1)} \right] k_1 e^{-i\,k_x x} & z < z_1 \\ \left[u_n^H\, e^{-i\,k_{zn}\,(z-z_{n-1})} - d_n^H\, e^{i\,k_{zn}\,(z-z_n)} \right] k_n e^{-i\,k_x x} & z_{n-1} < z < z_n \\ \left[u_N^H\, e^{-i\,k_{zN}\,(z-z_{N-1})} \right] k_N e^{-i\,k_x x} & z > z_{N-1} \end{cases} \tag{3.15}$$

for the TM case. For each of the layers between the first and last layers, let us define

$$\zeta_n = e^{-i\,k_{zn}\,(z_n-z_{n-1})}, \quad \text{where } (n = 2, 3, 4, ..., N - 1). \tag{3.16}$$

For the first layer at the bottom of the multi-layer medium, we define for convenience

$$\zeta_1 = 1. \tag{3.17}$$

With the definitions (3.16) and (3.17), matching the boundary conditions at the $N-1$ boundaries $z = z_1, z_2, z_3..., z_{N-1}$ yield the following equations

$$\hat{E}_y: \quad u_n^E \zeta_n + d_n^E = u_{n+1}^E + d_{n+1}^E \zeta_{n+1} \,, \tag{3.18a}$$

$$\hat{H}_x: \quad k_{zn}(u_n^E \zeta_n - d_n^E) = k_{z_{n+1}}(u_{n+1}^E - d_{n+1}^E \zeta_{n+1}) \,, \tag{3.18b}$$

$$\hat{E}_x: \quad \frac{k_{zn}}{k_n}(u_n^M \zeta_n + d_n^H) = \frac{k_{z_{n+1}}}{k_{n+1}}(u_{n+1}^H + d_{n+1}^H \zeta_{n+1}) \,, \tag{3.18c}$$

$$\hat{H}_y: \quad k_n(u_n^H \zeta_n - d_n^H) = k_{n+1}(u_{n+1}^H - d_{n+1}^H \zeta_{n+1}) \,. \tag{3.18d}$$

To calculate the amplitudes $u_n^{E,H}$ and $d_n^{E,H}$, we define generalized transmission and reflection coefficients $t_n^{E,H}$ and $r_n^{E,H}$ through the expressions

$$u_{n+1}^{E,H} = t_n^{E,H} \zeta_n u_n^{E,H} \,, \tag{3.19a}$$

$$d_n^{E,H} = r_n^{E,H} \zeta_n u_n^{E,H} \,. \tag{3.19b}$$

By using the relations in Eq. (3.19), and substituting into Eq. (3.18), we obtain for the TE case, the downward recursive relations

$$t_n^E = \frac{k_{zn}(1 - r_n^E)}{k_{z_{n+1}}(1 - r_{n+1}^E \zeta_{n+1}^2)} \,, \tag{3.20a}$$

$$r_n^E = \frac{k_{zn}(1 + r_{n+1}^E \zeta_{n+1}^2) - k_{z_{n+1}}(1 - r_{n+1}^E \zeta_{n+1}^2)}{k_{zn}(1 + r_{n+1}^E \zeta_{n+1}^2) + k_{z_{n+1}}(1 - r_{n+1}^E \zeta_{n+1}^2)} \,, \tag{3.20b}$$

which is initialized by defining $r_N^E = 0$. Similarly for the TM case, we find

$$t_n^H = \frac{k_n(1 - r_n^H)}{k_{n+1}(1 - r_{n+1}^H \zeta_{n+1}^2)} \,, \tag{3.21a}$$

$$r_n^H = \frac{k_{z_{n+1}} \epsilon_n^B(1 + r_{n+1}^H \zeta_{n+1}^2) - k_{zn} \epsilon_{n+1}^B(1 - r_{n+1}^H \zeta_{n+1}^2)}{k_{z_{n+1}} \epsilon_n^B(1 + r_{n+1}^H \zeta_{n+1}^2) + k_{zn} \epsilon_{n+1}^B(1 - r_{n+1}^H \zeta_{n+1}^2)} \,, \tag{3.21b}$$

which is initialized by defining $r_N^H = 0$.

Having determined the amplitude of a single plane wave in the multi-layer medium, it is straightforward to perform the same calculations for a collection of incident plane waves propagating in different directions. In our simulations, we will employ either a plane wave or a Gaussian beam as the incident field. The field distribution of an incident Gaussian beam can be readily synthesized through an angular spectrum of plane waves[2].

3.3   Green Tensor in Homogeneous Medium

In this section, we derive the electric Green tensor for a homogeneous medium with constant permittivity ϵ^B. To begin, we note that the background configuration is invariant in the x direction, and we may therefore employ the spatial Fourier transform of the electric Green tensor

$$\tilde{G}(k_x, z, z') = \int_{-\infty}^{\infty} \hat{G}(x, x', z, z') e^{i\,k_x(x-x')}\, dx\,, \qquad (3.22)$$

with the corresponding inverse transform

$$\hat{G}(x, x', z, z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{G}(k_x, z, z') e^{-i\,k_x(x-x')}\, dk_x\,, \qquad (3.23)$$

where we have dropped the superscript E for notational convenience. Another simplification applicable to our derivation is that the Maxwell equations can be split into two independent sets, namely the TE and TM polarizations, for our two-dimensional problem (Born and Wolf 1999, Sec. 11.4). In the Fourier domain, the Maxwell equations of

---

[2]For a detailed formulation of the angular spectrum pertaining to our two-dimensional geometry, the reader is referred to Appendix A.

Eq. (3.8) for a line source current directed along $\hat{\mathbf{a}}_p$ $(p = x, y, z)$ read

$$-\partial_z \tilde{E}_y^{G,p} + i\omega\mu_0 \tilde{H}_x^{G,p} = 0 \ , \tag{3.24a}$$

$$\partial_z \tilde{E}_x^{G,p} + ik_x \tilde{E}_z^{G,p} + i\omega\mu_0 \tilde{H}_y^{G,p} = 0 \ , \tag{3.24b}$$

$$-ik_x \tilde{E}_y^{G,p} + i\omega\mu_0 \tilde{H}_z^{G,p} = 0 \ , \tag{3.24c}$$

$$\partial_z \tilde{H}_y^{G,p} + i\omega\epsilon^B \tilde{E}_x^{G,p} = -\delta_{px}\delta(z - z') \ , \tag{3.24d}$$

$$-\partial_z \tilde{H}_x^{G,p} - ik_x \tilde{H}_z^{G,p} + i\omega\epsilon^B \tilde{E}_y^{G,p} = -\delta_{py}\delta(z - z') \ , \tag{3.24e}$$

$$ik_x \tilde{H}_y^{G,p} + i\omega\epsilon^B \tilde{E}_z^{G,p} = -\delta_{pz}\delta(z - z') \ , \tag{3.24f}$$

where δ_{ij} denotes the Kronecker delta, and where the substitution $\partial_x = -ik_x$ has been used. Eliminating $\tilde{E}_x^{G,p}, \tilde{E}_y^{G,p}, \tilde{H}_x^{G,p}$, and $\tilde{H}_y^{G,p}$ in Eq. (3.24), we obtain for $z \neq z'$

$$(\partial_z^2 + k_z^2)\{ \tilde{E}_z^{G,p} , \ \tilde{H}_z^{G,p} \} = 0 \,. \tag{3.25}$$

The solutions of these two equations are exponential functions that must vanish for $z \longrightarrow \pm\infty$. We may now write the field as

$$\{ \tilde{\mathbf{E}}^{G,p} , \ \tilde{\mathbf{H}}^{G,p} \} = \begin{cases} \{ \tilde{\mathcal{E}} , \ \tilde{\mathcal{H}} \} (k_x, -k_z) \cdot \mathbf{f}_p^+ \, e^{-ik_z(z-z')} , & z > z' \\ \{ \tilde{\mathcal{E}} , \ \tilde{\mathcal{H}} \} (k_x, +k_z) \cdot \mathbf{f}_p^- \, e^{ik_z(z-z')} , & z < z' \end{cases} \tag{3.26}$$

where $\mathbf{f}_p^+ = [f_{p,H}^+, f_{p,E}^+]^T$ and $\mathbf{f}_p^- = [f_{p,H}^-, f_{p,E}^-]^T$ are the amplitudes of the solutions of Eq. (3.25), corresponding to waves traveling in the positive and negative z direction, respectively. The superscript T denotes matrix transpose. The 3×2 coefficient matrices $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{H}}$ express the other field components in terms of $\tilde{E}_z^{G,p}$ and $\tilde{H}_z^{G,p}$ are given by

$$\tilde{\mathcal{E}}(k_x, k_z) = \begin{bmatrix} \dfrac{k_z}{k_x} & 0 \\ 0 & \dfrac{\omega\mu_0}{k_x} \\ 1 & 0 \end{bmatrix} , \tag{3.27}$$

$$\tilde{\mathcal{H}}\left(k_x, k_z\right) = \begin{bmatrix} 0 & \dfrac{k_z}{k_x} \\ -\dfrac{\omega\epsilon^B}{k_x} & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.28}$$

The fields $\tilde{\mathbf{E}}^{G,p}$ and $\tilde{\mathbf{H}}^{G,p}$ can be calculated with Eq. (3.26) once the four components of $\mathbf{f}_p^+$ and $\mathbf{f}_p^-$ are determined. This can be done by matching the tangential components of the fields at the source level $z = z'$. To this end, we integrate Eqs. (3.24a), (3.24b), (3.24d), and (3.24e) over an infinitesimally small interval around $z = z'$, and eliminate $\tilde{E}_z^{G,p}$ and $\tilde{H}_z^{G,p}$ using Eqs. (3.24c) and (3.24f) to find

$$\begin{bmatrix} \tilde{E}_x^{G,p} \\ \tilde{E}_y^{G,p} \\ \tilde{H}_x^{G,p} \\ \tilde{H}_y^{G,p} \end{bmatrix} (z \downarrow z') \; - \; \begin{bmatrix} \tilde{E}_x^{G,p} \\ \tilde{E}_y^{G,p} \\ \tilde{H}_x^{G,p} \\ \tilde{H}_y^{G,p} \end{bmatrix} (z \uparrow z') \; = \; \begin{bmatrix} \dfrac{k_x}{\omega\epsilon^B}\delta_{pz} \\ 0 \\ \delta_{py} \\ -\delta_{px} \end{bmatrix} \tag{3.29}$$

Substituting for the field components in Eq. (3.29) using Eq. (3.26), we obtain the following results for $\mathbf{f}_p^\pm$:

$$\mathbf{f}_x^+ = -\mathbf{f}_x^- = \begin{bmatrix} \dfrac{k_x}{2\omega\epsilon^B} \\ 0 \end{bmatrix}, \tag{3.30}$$

$$\mathbf{f}_y^+ = \mathbf{f}_y^- = \begin{bmatrix} 0 \\ -\dfrac{k_x}{2k_z} \end{bmatrix}, \tag{3.31}$$

$$\mathbf{f}_z^+ = \mathbf{f}_z^- = \begin{bmatrix} -\dfrac{k_x^2}{2\omega\epsilon^B k_z} \\ 0 \end{bmatrix}. \tag{3.32}$$

In addition, there is a contribution of a singular part $\dfrac{\mathrm{i}}{\omega\epsilon^B}\delta(z - z')$ to $\tilde{E}_z^{G,z}$, derived by similarly integrating Eq. (3.24f) over an infinitesimally small interval around $z = z'$.

For each choice of the direction p of the line current source, the solution for the electric field $\tilde{\mathbf{E}}^{G,p}$ is identically the pth column of the electric green tensor, as given in Eq. (3.9a). The electric Green tensor $\tilde{G}$ can therefore be written as a sum of a singular part $\tilde{G}^{(sin)}$ and a part due to the field vectors associated with the source current $\tilde{G}^{(source)}$:

$$\tilde{G}(k_x, z, z') = \tilde{G}^{(sin)} \delta(z - z') + \tilde{G}^{(source)} e^{-i\,k_z|\,z-z'\,|} \ , \tag{3.33}$$

with

$$\tilde{G}^{(sin)} = \frac{i}{\omega \epsilon^B} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \tag{3.34}$$

and

$$\tilde{G}^{(source)} = -\frac{1}{2\omega\epsilon^B} \begin{bmatrix} k_z & 0 & -Sk_x \\ 0 & \dfrac{k^2}{k_z} & 0 \\ -Sk_x & 0 & \dfrac{k_x^2}{k_z} \end{bmatrix} , \tag{3.35}$$

where

$$S = \text{sign}\,(z - z') \ . \tag{3.36}$$

The magnetic green tensor $\tilde{H}(k_x, z, z')$, related to the curl of the electric green tensor is

$$\tilde{H}(k_x, z, z') = \tilde{H}^{(source)} e^{-i\,k_z|\,z-z'\,|} \ , \tag{3.37}$$

with

$$\tilde{H}^{(source)} = -\frac{1}{2} \begin{bmatrix} 0 & -S & 0 \\ S & 0 & -\dfrac{k_x}{k_z} \\ 0 & \dfrac{k_x}{k_z} & 0 \end{bmatrix} . \tag{3.38}$$

3.4   Green Tensor in Multi-layered Medium

We now derive the Green tensor for the multi-layered background configuration. Assume that a line source current is located at $z = z'$ within one of the layers $\mathcal{B}^{(s)}$. The source layer $\mathcal{B}^{(s)}$ contains two source-free regions $\mathcal{B}^{(s+)}$ and $\mathcal{B}^{(s-)}$, situated above and below the line source at $z = z'$, respectively. In all the other source-free layers $(x, z) \in \mathcal{B}^{(n)}, n \neq s$, the electric and magnetic fields are associated with amplitude vectors $\mathbf{f}_p^{n,+}$ and $\mathbf{f}_p^{n,-}$, as in the case of the homogeneous medium (cf. Eq. (3.26)). These field amplitude vectors, superscripted with an additional label n, are however now, layer-dependent. It turns out that these field vectors can be determined recursively for the multi-layered medium, as we will now show. For layers above the source layer, we will employ a downward recursion scheme starting at $\mathcal{B}^{(N)}$ and ending at $\mathcal{B}^{(s+)}$. For layers below the source layer, we will employ a upward recursion scheme starting at $\mathcal{B}^{(1)}$ and ending at $\mathcal{B}^{(s-)}$.

The electric and magnetic fields in the source-free layers $\mathcal{B}^{(n)} (n \neq s)$ can be expressed as a sum of upward and downward traveling waves through the layer-dependent amplitude vectors as

$$\left\{ \tilde{\mathbf{E}}^{G,p}, \ \tilde{\mathbf{H}}^{G,p} \right\} = \left\{ \tilde{\mathcal{E}}, \ \tilde{\mathcal{H}} \right\} (k_x, -k_{zn}) \cdot \mathbf{f}_p^{n,+} \, e^{-i\,k_{zn}\,(z - z_{n,ref})}$$
$$+ \left\{ \tilde{\mathcal{E}}, \ \tilde{\mathcal{H}} \right\} (k_x, +k_{zn}) \cdot \mathbf{f}_p^{n,-} \, e^{i\,k_{zn}\,(z - z_{n,ref})} \ .$$

$$(3.39)$$

For a layer $\mathcal{B}^{(n)}$ above the the source level $z = z'$, the reference level $z_{n,ref}$ is taken as the lower interface of the layer, i.e., $z_{n,ref} = z_{n-1}$. For a layer $\mathcal{B}^{(n)}$ below the the source level, the reference level $z_{n,ref}$ is taken as the upper interface of the layer, i.e., $z^{n,ref} = z_n$. As in the previous section, the fields in all the other layers $\mathcal{B}^{(n)} (n \neq s)$ can

be calculated with Eq. (3.39) once all the field vectors $\mathbf{f}_p^{n,+}$ and $\mathbf{f}_p^{n,-}$ are determined.

Since the line source does not generate fields incoming from infinity, we write

$$\mathbf{f}_p^{N,-} \quad = \quad \mathbf{f}_p^{1,+} \quad = \quad 0 \,. \tag{3.40}$$

In a similar fashion as for the homogeneous medium, the amplitude vectors can be determined by matching the tangential field components at the material boundaries. To do this we define matrices $\tilde{\mathcal{E}}_T$ and $\tilde{\mathcal{H}}_T$ which are the upper 2×2 sub-matrices of $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{H}}$, respectively. The continuity condition at the interface between the layers $\mathcal{B}^{(n)}$ and $\mathcal{B}^{(n+1)}$ can then be written as

$$\mathbf{Q}_{n+1}(z_n) \begin{bmatrix} \mathbf{f}_p^{n+1,+} \\ \mathbf{f}_p^{n+1,-} \end{bmatrix} \quad = \quad \mathbf{Q}_n(z_n) \begin{bmatrix} \mathbf{f}_p^{n,+} \\ \mathbf{f}_p^{n,-} \end{bmatrix} \,, \tag{3.41}$$

where the 4×4 matrix $\mathbf{Q}_n$ is given by

$$\mathbf{Q}_n(z) \quad = \quad \begin{bmatrix} \dfrac{\tilde{\mathcal{E}}_T(k_x, -k_{zn})}{\sigma} & \tilde{\mathcal{E}}_T(k_x, k_{zn})\,\sigma \\[2ex] \dfrac{\tilde{\mathcal{H}}_T(k_x, -k_{zn})}{\sigma} & \tilde{\mathcal{H}}_T(k_x, k_{zn})\,\sigma \end{bmatrix} \,, \tag{3.42}$$

with $\sigma = e^{i\,k_{zn}\,(z - z_{n,ref})}$. We now define the transmission and reflection coefficients $t_{pol}^{d,n}$, $r_{pol}^{d,n}$, $t_{pol}^{u,n}$, $r_{pol}^{u,n}$, where $pol = E\,(TE), H\,(TM)$, and the superscripts d and u denote downward recursion and upward recursion respectively. For $\mathcal{B}^{(n)}$ above the source level, define

$$f_{p,pol}^{N,+} = t_{pol}^{d,n}\, f_{p,pol}^{n,+} \,, \tag{3.43a}$$

$$f_{p,pol}^{n,-} = r_{pol}^{d,n}\, f_{p,pol}^{n,+} \,, \tag{3.43b}$$

and for $\mathcal{B}^{(n)}$ below the source level, define

$$f_{p,pol}^{1,-} = t_{pol}^{u,n}\, f_{p,pol}^{n,-} \,, \tag{3.44a}$$

$$f_{p,pol}^{n,+} = r_{pol}^{u,n}\, f_{p,pol}^{n,-} \,. \tag{3.44b}$$

In the downward recursion scheme, the reflection and transmission coefficients of layer $\mathcal{B}^{(n)}$ are expressed recursively in those of $\mathcal{B}^{(n+1)}$ by using the continuity condition in Eq. (3.41). The results are

$$r_E^{d,n} = -\frac{\gamma_d^2 \left[k_{z_{(n+1)}}(-1 + r_E^{d,n+1}) + k_{zn}(1 + r_E^{d,n+1}) \right]}{\left[k_{z_{(n+1)}}(-1 + r_E^{d,n+1}) - k_{zn}(1 + r_E^{d,n+1}) \right]} , \tag{3.45a}$$

$$t_E^{d,n} = \frac{2\,\gamma_d\, k_{zn}\, t_E^{d,n+1}}{\left[k_{z_{(n+1)}}(-1 + r_E^{d,n+1}) - k_{zn}(1 + r_E^{d,n+1}) \right]} , \tag{3.45b}$$

$$r_H^{d,n} = -\frac{\gamma_d^2 \left[\epsilon_n\, k_{z_{(n+1)}}(-1 + r_H^{d,n+1}) + \epsilon_{n+1}\, k_{zn}(1 + r_H^{d,n+1}) \right]}{\left[\epsilon_n\, k_{z_{(n+1)}}(-1 + r_H^{d,n+1}) - \epsilon_{(n+1)}\, k_{zn}(1 + r_H^{d,n+1}) \right]} , \tag{3.45c}$$

$$t_H^{d,n} = \frac{2\,\gamma_d\, \epsilon_n\, k_{zn}\, t_H^{d,n+1}}{\left[\epsilon_n\, k_{z_{(n+1)}}(-1 + r_H^{d,n+1}) - \epsilon_{(n+1)}\, k_{zn}(1 + r_H^{d,n+1}) \right]} , \tag{3.45d}$$

with $\gamma_d = e^{-i\,k_{zn}\,(z_n - z_{(n-1)})}$. Similarly for the upward recursion scheme, the reflection and transmission coefficients of layer $\mathcal{B}^{(n+1)}$ are expressed recursively in those of $\mathcal{B}^{(n)}$, and the results are

$$r_E^{u,n+1} = -\frac{\gamma_u^2 \left[k_{z_n}(1 - r_E^{u,n}) - k_{z_{(n+1)}}(1 + r_E^{u,n}) \right]}{\left[k_{z_n}(1 - r_E^{u,n}) + k_{z_{(n+1)}}(1 + r_E^{u,n}) \right]} , \tag{3.46a}$$

$$t_E^{d,n+1} = \frac{2\,\gamma_u\, k_{z_{(n+1)}}\, t_E^{u,n}}{\left[k_{z_n}(1 - r_E^{u,n}) + k_{z_{(n+1)}}(1 + r_E^{u,n}) \right]} , \tag{3.46b}$$

$$r_H^{d,n+1} = -\frac{\gamma_u^2 \left[\epsilon_{n+1}\, k_{z_n}(1 - r_H^{u,n}) - \epsilon_n\, k_{z_{(n+1)}}(1 + r_H^{u,n}) \right]}{\left[\epsilon_{n+1}\, k_{z_n}(1 - r_H^{u,n}) + \epsilon_n\, k_{z_{(n+1)}}(1 + r_H^{u,n}) \right]} , \tag{3.46c}$$

$$t_H^{d,n+1} = \frac{2\,\gamma_u\, \epsilon_{n+1}\, k_{z_{(n+1)}}\, t_H^{u,n}}{\left[\epsilon_{n+1}\, k_{z_n}(1 - r_H^{u,n}) + \epsilon_n\, k_{z_{(n+1)}}(1 + r_H^{u,n}) \right]} , \tag{3.46d}$$

with $\gamma_u = e^{-i\,k_{z_{(n+1)}}\,(z_{(n+1)} - z_n)}$. Consistent with the boundary conditions at infinity, the

downward recursion is initialized with the conditions

$$r_{pol}^{d,N} = 0 \ , \tag{3.47a}$$

$$t_{pol}^{d,N} = 1 \ , \tag{3.47b}$$

and the upward recursion with

$$r_{pol}^{u,1} = 0 \ , \tag{3.48a}$$

$$t_{pol}^{u,1} = 1 \ . \tag{3.48b}$$

The next task is to initiate the recursion schemes in Eqs. (3.43) and (3.44) to enable us
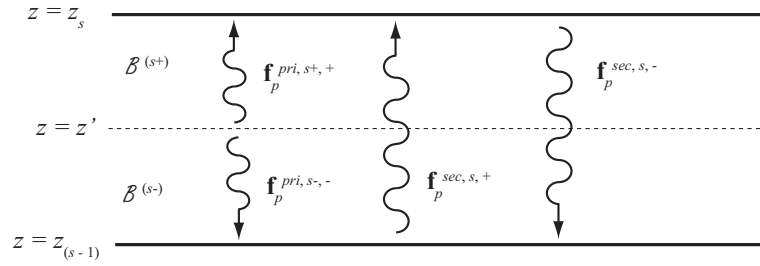
Figure 3.3: The primary and secondary amplitude vectors .

to calculate the field amplitude vectors in all the layers. To do this, we first determine $f_{p,pol}^{N,+}$ and $f_{p,pol}^{1,-}$ by using the field amplitude vectors in the regions $\mathcal{B}^{(s+)}$ and $\mathcal{B}^{(s-)}$. For this purpose, we define, in addition to the primary field vectors that radiates from the

line source, a pair of upward and downward traveling secondary field vectors between $z = z_s$ and $z = z_{(s-1)}$ (Fig. 3.3.). The primary field takes into account the contribution from the line source, and the secondary field takes into account contribution from multiple reflections and refractions at the various material boundaries. The primary field vectors are given by Eqs. (3.30) - (3.32) as for the homogeneous case. In line with the notation for the amplitude vectors in Eq. (3.39), we have the following relations between the primary and secondary fields

$$\mathbf{f}_p^{s+,+} = \mathbf{f}_p^{pri,s+,+} + \mathbf{f}_p^{sec,s,+} \ , \tag{3.49a}$$

$$\mathbf{f}_p^{s+,-} = \mathbf{f}_p^{sec,s,-} \ , \tag{3.49b}$$

$$\mathbf{f}_p^{s-,+} = \mathbf{f}_p^{sec,s,+} \ , \tag{3.49c}$$

$$\mathbf{f}_p^{s-,-} = \mathbf{f}_p^{pri,s-,-} + \mathbf{f}_p^{sec,s,-} \ . \tag{3.49d}$$

From Eqs. (3.43) and (3.44), we find

$$f_{p,pol}^{s+,-} = r_{pol}^{d,s+} f_{p,pol}^{s+,+} \ , \tag{3.50a}$$

$$f_{p,pol}^{s-,+} = r_{pol}^{u,s-} f_{p,pol}^{s-,-} \ . \tag{3.50b}$$

Substituting Eq. (3.50) into Eq. (3.49), we can solve for $f_{p,pol}^{s+,+}$ and $f_{p,pol}^{s-,-}$ in terms of the known primary field vectors. The results are

$$f_{p,pol}^{s+,+} = \frac{f_{p,pol}^{pri,s+,+} + r_{pol}^{u,s-} f_{p,pol}^{pri,s-,-}}{1 - r_{pol}^{d,s+} r_{pol}^{u,s-}} \ , \tag{3.51a}$$

$$f_{p,pol}^{s-,-} = \frac{f_{p,pol}^{pri,s-,-} + r_{pol}^{d,s+} f_{p,pol}^{pri,s+,+}}{1 - r_{pol}^{d,s+} r_{pol}^{u,s-}} \ . \tag{3.51b}$$

Back-substituting into Eq. (3.50), we find

$$f_{p,pol}^{s+,-} = \frac{1}{M_{pol}} \left[r_{pol}^{d\star} \, f_{p,pol}^{pri,s+,+} \, e^{-\mathrm{i}\,2k_{zn}\,(z_c-z')} \, + \, L_{pol} \, f_{p,pol}^{pri,s-,-} \right] \, , \tag{3.52a}$$

$$f_{p,pol}^{s-,+} = \frac{1}{M_{pol}} \left[L_{pol} \, f_{p,pol}^{pri,s+,+} \, + \, r_{pol}^{u\star} \, f_{p,pol}^{pri,s-,-} \, e^{-\mathrm{i}\,2k_{zn}\,(z'-z_c)} \right] \, , \tag{3.52b}$$

where we have made use of the following substitutions

$$r_{pol}^{d,s+} = r_{pol}^{d\star} \, e^{-\mathrm{i}\,2k_{zn}\,(z_c-z')} \, , \tag{3.53a}$$

$$r_{pol}^{u,s-} = r_{pol}^{u\star} \, e^{-\mathrm{i}\,2k_{zn}\,(z'-z_c)} \, , \tag{3.53b}$$

$$r_{pol}^{d\star} = r_{pol}^{d,s+}(z' = z_c) \, , \tag{3.53c}$$

$$r_{pol}^{u\star} = r_{pol}^{u,s-}(z' = z_c) \, , \tag{3.53d}$$

$$z_c = \left(z_s + z_{(s-1)} \right)/2 \, , \tag{3.53e}$$

$$L_{pol} = r_{pol}^{d\star} \, r_{pol}^{u\star} \, , \tag{3.53f}$$

$$M_{pol} = 1 - L_{pol} \, . \tag{3.53g}$$

With Eq. (3.51), the field vectors $f_{p,pol}^{N,+}$ and $f_{p,pol}^{1,-}$ can be determined, and all the field vectors in the other layers $\mathcal{B}^{(n)}, n \neq s$, follow from the recursion relations in Eqs. (3.43) and (3.44).

In the source layer $\mathcal{B}^{(s)}$, the electric Green tensor is given by the sum of the contributions of a singular part $\tilde{G}^{(sin)}$, a source part $\tilde{G}^{(source)}$, and a part $\tilde{G}^{(field\,(sec))}$ due to the secondary field vectors that takes into account multiple reflections and refractions across the various material boundaries

$$\tilde{G}(k_x, z, z') = \tilde{G}^{(sin)} \, \delta(z - z') + \tilde{G}^{(source)} e^{-\mathrm{i}\,k_{zs}|z-z'|} + \tilde{G}^{(field\,(sec))} \, . \tag{3.54}$$

The contributions $\tilde{G}^{(sin)}$ and $\tilde{G}^{(source)}$ are as given in Eqs. (3.34) and (3.35) for the case of the homogeneous media. The contribution $\tilde{G}^{(field\,(sec))}$ in the source layer can be

further broken down into two parts $\tilde{G}^{(field(sec,+))}$ and $\tilde{G}^{(field(sec,-))}$

$$\tilde{G}^{(field(sec))} = \tilde{G}^{(field(sec,+))} \, e^{-i\,k_{zs}(z-z')} + \tilde{G}^{(field(sec,-))} \, e^{i\,k_{zs}(z-z')} \,, \tag{3.55}$$

where

$$\tilde{G}^{(field(sec,+))} = \begin{bmatrix} -\dfrac{k_{zs}}{k_x} & 0 \\ 0 & \dfrac{\omega\mu_0}{k_x} \\ 1 & 0 \end{bmatrix} \times$$

$$\frac{1}{M_{pol}} \left\{ \mathcal{F}^{pri,s+,+} \, L_{pol} + \mathcal{F}^{pri,s-,-} \, r^{u\star}_{pol} \, e^{-i\,2k_{zs}(z'-z_c)} \right\} \,, \tag{3.56}$$

$$\tilde{G}^{(field(sec,-))} = \begin{bmatrix} \dfrac{k_{zs}}{k_x} & 0 \\ 0 & \dfrac{\omega\mu_0}{k_x} \\ 1 & 0 \end{bmatrix} \times$$

$$\frac{1}{M_{pol}} \left\{ \mathcal{F}^{pri,s+,+} \, r^{d\star}_{pol} \, e^{-i\,2k_{zs}(z_c-z')} + \mathcal{F}^{pri,s-,-} \, L_{pol} \right\} \,, \tag{3.57}$$

with the 2×3 matrices

$$\mathcal{F}^{pri,s+,+} = \left[\mathbf{f}_x^+, \mathbf{f}_y^+, \mathbf{f}_z^+ \right] \,, \tag{3.58a}$$

$$\mathcal{F}^{pri,s-,-} = \left[\mathbf{f}_x^-, \mathbf{f}_y^-, \mathbf{f}_z^- \right] \,. \tag{3.58b}$$

Replacing the 3×2 matrices $\tilde{\mathcal{E}}(k_x, -k_{zs})$ and $\tilde{\mathcal{E}}(k_x, k_{zs})$ with $\tilde{\mathcal{H}}(k_x, -k_{zs})$ and $\tilde{\mathcal{H}}(k_x, k_{zs})$ in Eqs. (3.56) and (3.57) respectively, the tensors $\tilde{H}^{(field(sec,+))}$ and $\tilde{H}^{(field(sec,-))}$ for the magnetic field are determined. For all other layers $\mathcal{B}^{(n)}, n \neq s,$, the electric and magnetic fields can be found by substituting the appropriate parameters for each layer in Eq. (3.39).

We have now determined, up to within a Fourier transform, the electric Green tensors $\hat{G}$ in the multi-layered medium. In order to determine the electric fields, the domain integral equation in Eq. (3.11a) is to be solved numerically, a process we describe in the following section.

3.5   Numerical Treatment of the Domain Integral Equations

To obtain a numerically stable solution of the domain integral equation in Eq. (3.11a), we define an approximate solution of the form

$$\check{E}_i(x,z) = \sum_{\mathcal{K}=1}^{\mathcal{K}_0} \alpha_{i\mathcal{K}}\, \beta_\mathcal{K}(x,z)\;, \tag{3.59}$$

where $\alpha_{i\mathcal{K}}$ are expansion coefficients to be determined, and $\beta_\mathcal{K}$ are suitable basis functions.

It is to be noted that $\check{E}_i(x,z)$ is in principle exact as the number of basis functions $\mathcal{K}_0 \longrightarrow \infty$. In practical implementation, the solution $\check{E}_i(x,z)$ is always an approximation since $\mathcal{K}_0$ must be finite. Following the method of weighted residuals[3], we define the residual R_i as

$$
\begin{aligned}
R_i(x,z) = {}& \check{E}_i(x,z) - \hat{E}_i^{inc}(x,z) \\
& - \mathrm{i}\omega \sum_{m=1}^{m_0} \iint_{\mathcal{D}_m} \Delta\epsilon\,(x',z')\,\hat{G}_{ij}^E\,(x,z;x',z')\check{E}_j\,(x',z')\,dx'\,dz' \\
& \neq 0\;,
\end{aligned}
\tag{3.60}
$$

where it is required that the weighted integral of R_i to vanish in the domain Ω such

[3]For detailed discussions on this method, see for instance, Section 9.2 of Grandin (1986) or Section 2.4.3 of Reddy (1993).

that

$$\int_\Omega W_\mathcal{L}(x, z)\, R_i(x, z)\, dx\, dz = 0 \qquad (\mathcal{L} = 1, 2, 3, ..., \mathcal{K}_0)\,, \tag{3.61}$$

with $W_\mathcal{L}$ as the appropriate weight functions. The integral in Eq. (3.60) is evaluated over the deviant regions $\mathcal{D}_m$ only. We employ the collocation method (Reddy 1993, Sec. 2.4.3) in choosing the weight functions $W_\mathcal{L}$'s, resulting in a set of Dirac delta functions for the weight functions, viz

$$W_\mathcal{L} = \delta(x - x_\mathcal{L})\, \delta(z - z_\mathcal{L})\,. \tag{3.62}$$

Using this set of weighting functions in Eq. (3.61) is equivalent to setting the residual R_i at each of the $\mathcal{L}$th points to be zero, i.e.,

$$R_i\,(x_\mathcal{L},\, z_\mathcal{L}) = 0\,. \tag{3.63}$$

Substituting Eq. (3.59) into Eq. (3.60), and making use of Eq. (3.63), we obtain $3\mathcal{K}_0$ linear equations for the $3\mathcal{K}_0$ coefficients $\alpha_{x1}, ..., \alpha_{x\mathcal{K}_0}, \alpha_{y1}, ..., \alpha_{y\mathcal{K}_0}$, and $\alpha_{z1}, ..., \alpha_{z\mathcal{K}_0}$,

$$\sum_{j=x}^{z} \sum_{\mathcal{K}=1}^{\mathcal{K}_0} \left[\, \beta_\mathcal{K}\,(x_\mathcal{L},\, z_\mathcal{L})\, \delta_{ij} - \mathrm{i}\omega\, A_{ij\mathcal{L}\mathcal{K}} \,\right] \alpha_{j\mathcal{K}} \;=\; \hat{E}_i^{inc}\,(x_\mathcal{L},\, z_\mathcal{L})\,, \tag{3.64}$$

where

$$A_{ij\mathcal{L}\mathcal{K}} \;=\; \iint_{\mathcal{D}_m} \Delta\epsilon\,(x', z')\, \hat{G}_{ij}^E\,(x_\mathcal{L}, z_\mathcal{L}; x', z') \beta_\mathcal{K}\,(x', z')\, dx'\, dz' \tag{3.65}$$

Taking the basis functions $\beta_\mathcal{K}$'s as piecewise constant functions, i.e.,

$$\beta_\mathcal{K}\,(x, z) = \begin{cases} 1 & \text{if } (x, z) \in \mathcal{D}_m \\[2ex] 0 & \text{otherwise} \end{cases}, \tag{3.66}$$

Eq. (3.65) reduces to

$$A_{ij\mathcal{L}\mathcal{K}} \;=\; \iint_{\mathcal{D}_m} \Delta\epsilon\,(x', z')\, \hat{G}_{ij}^E\,(x_\mathcal{L}, z_\mathcal{L}; x', z')\, dx'\, dz'\,. \tag{3.67}$$

In our configuration, we will take that $\Delta\epsilon\,(x',z')$ is constant within each deviant region, so that

$$
A_{ij\mathcal{L}\mathcal{K}} = \Delta\epsilon\,(x',z') \iint_{\mathcal{D}_m} \hat{G}_{ij}^E\,(x_\mathcal{L}, z_\mathcal{L}; x', z')\,dx'\,dz'
$$

$$
= \Delta\epsilon\,(x',z')\,C_{ij\mathcal{L}\mathcal{K}}\,. \tag{3.68}
$$

Using the Fourier relationship established in Eq. (3.23), one obtains for $C_{ij\mathcal{L}\mathcal{K}}$

$$
C_{ij\mathcal{L}\mathcal{K}} = \frac{1}{2\pi} \iint_{\mathcal{D}_m} \left[\int_{-\infty}^{\infty} \tilde{G}(k_x, z_\mathcal{L}, z')e^{-i\,k_x(x_\mathcal{L}-x')}\,dk_x \right] dx'\,dz'\,. \tag{3.69}
$$

According to Fubini's theorem, we may change the order of integration to obtain

$$
C_{ij\mathcal{L}\mathcal{K}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int \tilde{G}(k_x, z_\mathcal{L}, z')\,dz' \right] \left[\int e^{-i\,k_x(x_\mathcal{L}-x')}\,dx' \right] dk_x
$$

$$
= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int \tilde{G}(k_x, z_\mathcal{L}, z')\,dz' \right] \Delta x\,\mathrm{sinc}\frac{k_x\Delta x}{2}\,e^{-i\,k_x(x_\mathcal{L}-x')}\,dk_x\,, \tag{3.70}
$$

where Δx is the length of each discretization element along the x direction in each contrast source region $\mathcal{D}_m$. For a contrast region $\mathcal{D}_m$ located in one of the layers $n=s$, we can make use of Eq. (3.34, 3.35, 3.39, 3.54, 3.55), to rewrite Eq. (3.70) for $n=s$

$$
C_{ij\mathcal{L}\mathcal{K}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ \int \left[\tilde{G}^{(sin)}\,\delta(z_\mathcal{L}-z') + \tilde{G}^{(source)}\,(k_x)\,e^{-i\,k_{zs}|z_\mathcal{L}-z'|} \right.\right.
$$

$$
\left.+ \tilde{G}^{(field\,(sec,+))}\,(k_x)\,e^{-i\,k_{zs}(z_\mathcal{L}-z')} + \tilde{G}^{(field\,(sec,-))}\,(k_x)\,e^{i\,k_{zs}(z_\mathcal{L}-z')} \right] dz'
$$

$$
\left. \times\,\Delta x\,\mathrm{sinc}\frac{k_x\Delta x}{2}\,e^{-i\,k_x(x_\mathcal{L}-x')} \right\} dk_x\,, \tag{3.71}
$$

and for $n \neq s$,

$$
\begin{aligned}
C_{ij\mathcal{L}\mathcal{K}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Bigg\{ \int \Bigg[&\tilde{G}^{(field(sec,n,+))}\left(k_x\right) e^{-\mathrm{i}\,k_{zn}\left(z_{\mathcal{L}} - z_{n,ref}\right)} \\
&+ \tilde{G}^{(field(sec,n,-))}\left(k_x\right) e^{\mathrm{i}\,k_{zn}\left(z_{\mathcal{L}} - z_{n,ref}\right)} \Bigg] dz' \\
&\times \Delta x \, \mathrm{sinc}\frac{k_x \Delta x}{2}\, e^{-\mathrm{i}\,k_x(x_{\mathcal{L}} - x')} \Bigg\} dk_x \; ,
\end{aligned}
$$

$$(3.72)$$

where

$$
\tilde{G}^{(field(sec,n,+))}\left(k_x\right) = \tilde{\mathcal{E}}\left(k_x, -k_{zn}\right) \cdot \mathbf{f}_p^{n,+} \; ,
\tag{3.73a}
$$

$$
\tilde{G}^{(field(sec,n,-))}\left(k_x\right) = \tilde{\mathcal{E}}\left(k_x, +k_{zn}\right) \cdot \mathbf{f}_p^{n,-} \; .
\tag{3.73b}
$$

Here we have shown explicitly the dependence of the Green tensor $\left(G^{(source)}, \tilde{G}^{(field(sec,+))}, \right.$ $\tilde{G}^{(field(sec,-))}, \tilde{G}^{(field(sec,n,+))},$ and $\left. \tilde{G}^{(field(sec,n,-))}\right)$ on k_x. Note that $\tilde{G}^{(sin)}$ does not depend on k_x. Integrating the first two terms in square brackets in Eq. (3.71) with respect to z', we find

$$
\int \tilde{G}^{(sin)}\, \delta(z_{\mathcal{L}} - z')\, dz' = \delta_{\mathcal{L}\mathcal{K}} \tilde{G}^{(sin)} \; ,
\tag{3.74}
$$

and

$$
\int \tilde{G}^{(source)}\left(k_x\right) e^{-\mathrm{i}\,k_{zs}\,|z_{\mathcal{L}} - z'|}\, dz \;=\; \frac{-1}{2\omega\epsilon^B}
\begin{bmatrix}
I_s k_{zs} & 0 & -I_a k_x \\[2mm]
0 & I_s \dfrac{k_s^2}{k_{zs}} & 0 \\[2mm]
-I_a k_x & 0 & I_s \dfrac{k_x^2}{k_{zs}}
\end{bmatrix} ,
\tag{3.75}
$$

where

$$
I_s =
\begin{cases}
\dfrac{2}{k_{zs}} \sin \dfrac{k_{zs}\Delta z}{2} e^{-\mathrm{i}\,k_{zs}\,|z_{\mathcal{L}} - z'|} & |z_{\mathcal{L}} - z'| > \dfrac{\Delta z}{2} \\[4mm]
\dfrac{2}{\mathrm{i}k_{zs}} \left[1 - \cos k_{zs}(z_{\mathcal{L}} - z') e^{-\mathrm{i}\,k_{zs}\,\frac{\Delta z}{2}} \right] & |z_{\mathcal{L}} - z'| < \dfrac{\Delta z}{2}
\end{cases} ,
\tag{3.76}
$$

$$I_a = \begin{cases} \dfrac{2}{k_{zs}} \sin \dfrac{k_{zs}\Delta z}{2} e^{-\mathrm{i}\,k_{zs}\,|z_{\mathcal{L}} - z'|} & z_{\mathcal{L}} - z' > \dfrac{\Delta z}{2} \\[2ex] -\dfrac{2}{k_{zs}} \sin \dfrac{k_{zs}\Delta z}{2} e^{-\mathrm{i}\,k_{zs}\,|z_{\mathcal{L}} - z'|} & z_{\mathcal{L}} - z' < -\dfrac{\Delta z}{2} \\[2ex] \dfrac{2}{k_{zs}} \sin k_{zs}(z_{\mathcal{L}} - z') e^{-\mathrm{i}\,k_{zs}\,\frac{\Delta z}{2}} & |z_{\mathcal{L}} - z'| < \dfrac{\Delta z}{2} \end{cases}, \tag{3.77}$$

and Δz is the length of each discretization element along the z direction in each contrast source region $\mathcal{D}_m$. The rest of the terms in the square brackets of Eqs. (3.71) and (3.72), $\tilde{G}^{(field\,(sec,+))}$, $\tilde{G}^{(field\,(sec,-))}$, $\tilde{G}^{(field\,(sec,n,+))}$, and $\tilde{G}^{(field\,(sec,n,-))}$, contribute to upward and downward propagating waves according to Eqs. (3.56), (3.57), and (3.73). The dependence on z' of each upward and downward propagating wave due to these terms is $e^{-\mathrm{i}\,k_{zs}\,z'}$ and $e^{+\mathrm{i}\,k_{zs}\,z'}$ respectively. Therefore, each of these terms are associated with either the factor $\Delta z\, \mathrm{sinc} \dfrac{k_{zs}\Delta z}{2} e^{-\mathrm{i}\,k_{zs}z'}$ or $\Delta z\, \mathrm{sinc} \dfrac{k_{zs}\Delta z}{2} e^{\mathrm{i}\,k_{zs}z'}$.

The coefficients $C_{ij\mathcal{L}\mathcal{K}}$ are then determined numerically by use of the Fast Fourier Transform (Press et al. 2005, Sec. 12). In doing so, two problems need to be addressed. First, the constant term in I_s that has no exponent converges slowly; $I_{s,xx}$ does not tend to zero and both $I_{s,yy}$ and $I_{s,zz}$ tends to infinity as $k_x \longrightarrow \infty$. For these terms, we calculate the transform analytically using the residue theorem. The second problem is that of pole singularities in the system for values of k_x for which M_{pol} vanishes. These poles correspond physically to guided modes in the layers as well as the surface plasmon mode (TM case) along the boundaries of the layers. In the case of surface plasmons, the pole lies close to the real k_x-axis. To avoid the pole at k_x, we may deform the contour of integration in the vicinity of the poles to a semi-circular path with radius r as shown in Fig. 3.4

Once the coefficients $C_{ij\mathcal{L}\mathcal{K}}$ are determined, we may solve for the coefficients $\alpha_{j\mathcal{K}}$

in Eq. (3.64) through matrix inversion, and hence the field in the deviant regions $\mathcal{D}_m$

(cf. Eq. (3.59)). For TE polarization, there is only an $\hat{E}_y$ field, and a $\mathcal{K}_0 \times \mathcal{K}_0$ matrix

has to be inverted. For the TM polarization, there is both an $\hat{E}_x$ and an $\hat{E}_z$ field, and

a $2\mathcal{K}_0 \times 2\mathcal{K}_0$ matrix has to be inverted. The matrix may be inverted using a method

such as Gaussian elimination with partial or maximum pivoting (Press et al. 2005, Sec.

2.1). The field anywhere else can then be determined by substituting the fields solved

through the inversion back into Eq. (3.11), the domain integral equation.
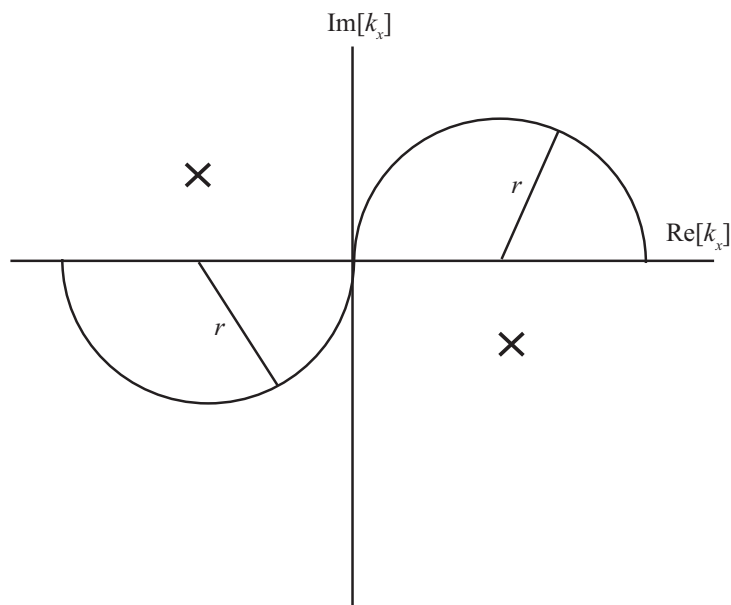
Figure 3.4: Contour deformation with semi-circles of radius r to avoid pole singularities marked at 'x'.

CHAPTER 4: SUPER-RESOLVED OPTICAL READOUT

The ever-growing demand by industries and consumers alike for high capacity digital storage media has stimulated much research interest in near field optical data readout, an approach that offers the capability for operation beyond the diffraction limit. According to Lord Rayleigh[1], the limit of resolution of a diffraction-limited imaging system is (Born and Wolf 1999, Sec 7.6.3, Hecht 1998, Sec 10.2.6)

$$\Delta l_{min} = 0.61\lambda_0/NA \qquad (4.1)$$

where λ_0 is the free space wavelength, and NA is the numerical aperture of the scanning objective. It is immediately clear that one can improve the resolution by using a shorter wavelength or by increasing the NA. From the CD (compact disc) to Blu-ray disc, the application of these strategies to improve the resolution is clear, as can be seen from the data in Table 4.1.

The fact that shorter wavelengths of radiation such as ultraviolet and X-rays are ionizing radiation that could change or damage materials they illuminate suggests that the blue laser is about the shortest wavelength practical for an optical data storage system. In the regime of near field optics, imaging with a subwavelength aper-

---

[1]In Rayleigh's own words: "this rule is convenient on account of its simplicity; and it is sufficiently accurate in view of the necessary uncertainty as to what is meant by resolution" (Strutt JW 1879).

Table 4.1: Physical system parameters for the various optical storage media. Data from van de Nes et al. (2006).

Medium	λ (nm)	NA	Min. spot size (nm)	Capacity(GB/layer)
CD	785	0.45	2110	0.65
DVD	650	0.60	1320	4.7
Blu-ray	405	0.85	580	23

ture (Ash and Nicholls 1972, Pohl et al. 1984), sampling with near field scanning optical microscopy (Betzig et al. 1991, Betzig and Trautman 1992), increasing the NA with solid immersion lens (Terris et al. 1994, Suzuki et al. 1998, Török and Kao 2003), and employing localized plasmon effects in so-called super-RENS structures (Tominaga and Tsai 2003, Her et al. 2003), are methods that have been used to overcome the diffraction limit. In fact, resolution limit close to $\sim \lambda/9$ for near field magneto-optic readout has been demonstrated by Betzig et al. (1992).

The major problem however, is that the amount of light throughput available to perform the readout operation is severely limited by the subwavelength apertures, making the system very susceptible to noise signals. Here we propose strategies to perform super-resolved optical readout with surface plasmons, which could serve as potential alternatives to retrieving the information from the recorded medium. As demonstrated by experiments conducted in several studies (Ebbesen et al. 1998, Thio et al. 2001, Lezec et al. 2002), surface plasmons can couple with light to produce enhanced optical transmission and localized field effects in subwavelength apertures. In contrast

to the super-RENS technique, we manipulate the plasmon effects by means of suitably placed 'plasmon pits' in the optical readout system. Depending on their locations and geometries, these plasmon pits can serve both to enhance or hinder the performance of the readout system (García-Vidal et al. 2003, Martín-Moreno et al. 2003, Gbur et al. 2005).

4.1 Plasmon-assisted Optical Readout System Configurations
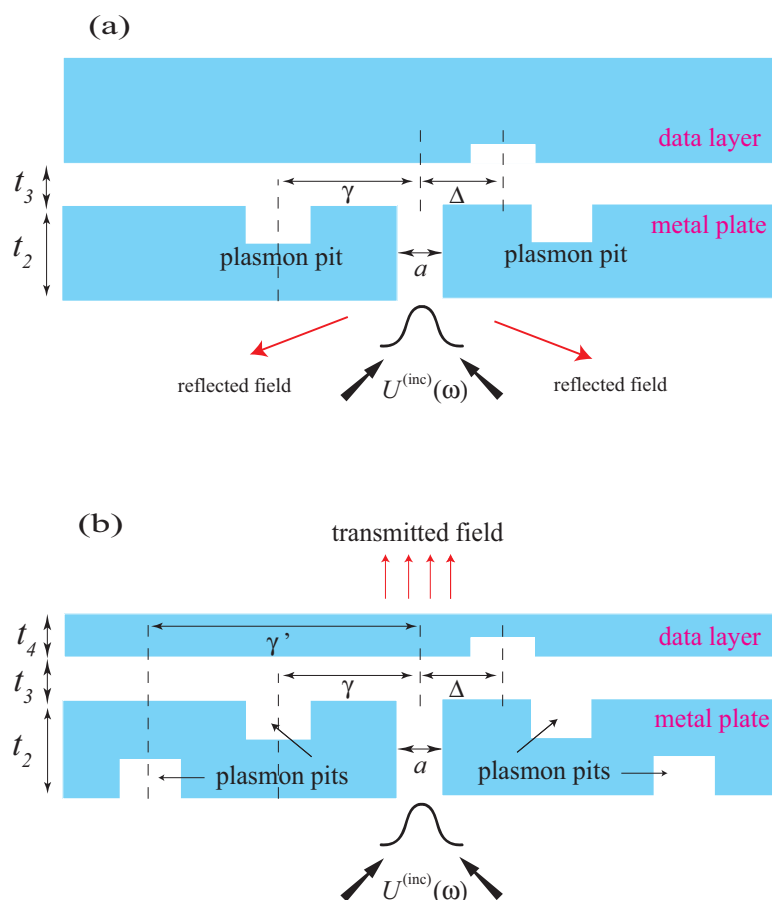
Figure 4.1: Illustrating the geometries of the configurations investigated for plasmon-assisted optical readout.

In this section, we present results on our investigations of suitable strategies for

plasmon-assisted optical readout based on: (a) a reflection readout configuration; and (b) a transmission readout configuration. The geometries of the two configurations are shown in Fig. 4.1. A metal plate (taken to be silver) of thickness $t_2 = 100$ nm that contains a single subwavelength slit of width a is illuminated with a quasi-monochromatic incident field of wavelength $\lambda = 500$ nm. The refractive index of silver is taken to be $n_{ag} = 0.05 - i\,2.87$, following the data of Johnson and Christy (1972).

Plasmon pits whose size and position depend on the strategy considered may be located either on the light (illuminated) or dark side of the silver plate. The separation between each plasmon pit and the subwavelength slit is given by γ on the dark side, and by γ' on the light side. These plasmon pits play several roles in the optical readout system. First, together with the subwavelength slit, they help to bridge the momentum gap between the surface plasmons and the incident field by scattering part of the incident field into a plasmonic channel, as discussed earlier in Section 2.3. Second, surface plasmons that propagate along the surface can be reflected by the edges of the plasmon pits, resulting in increased confinement of the surface plasmons to the region around the slit. Third, the plasmon pits can serve as surface corrugations to enhance optical transmission in the system (García-Vidal et al. 2003). In the reflection configuration, the plasmon pits on the light side of the plate degrade the readout contrast and are therefore excluded (Gbur et al. 2005).

Both the width and height of the plasmon pits in the 100 nm thick silver plate are taken to be 50 nm and 40 nm in the reflection and transmission configuration respectively. As discussed, the plasmon pits can be present on both the dark side and light side in the transmission configuration. By making them smaller, it is ensured that the

two pairs of pits (whose positions are to be determined) do not coincide to form two additional slits in the silver plate. Furthermore, simulations (results to be shown shortly) with only the metal plate and plasmon pits on the illuminated side reveals that more optical power is transmistted with pit size 40×40 nm^2 than with pit size 50×50 nm^2. This finding also demonstrates that the effects of the surface plasmons do not scale linearly with the dimensions of the subwavelength features.

The width a of the subwavelength slit in the silver plate is taken to be 25 nm, except for a particular case in the reflection readout configuration where a slightly different approach to that shown Fig. 4.1a is explored. In that case, only one plasmon pit is present, and the plasmon pit, instead of the slit, acts as the probe for the data structure. The slit width a in that case is taken to be 10 nm to minimize effects in the system due to its presence, as we will see in the next section.

The data layer, taken to be either silver or silicon, is separated from the metal plate by an air gap of $t_3 = 30$ nm, similar to that in commercially available hard disk drives (Mamun and Ge 2005, van de Nes et al. 2006). The refractive index of silicon is taken to be $n_{si} = 4.30 - \mathrm{i}\,0.07$, following the data of Palik (1985).

In the reflection readout configuration, as we are interested in the reflected field, we take the thickness of the data layer t_4 to be infinite for convenience.[2] For the transmission readout configuration, it was found through a series of numerical simulations that a thickness of 30 nm for the data layer t_4 is optimal. The choice of a 30 nm thin data layer helps reduce the effects of absorption, and allows for pos-

---

[2]This assumption is reasonable in the visible spectrum where the noble metals, such as silver and gold, are highly reflective.

sible plasmon-assisted field enhancement effects when the data layer is also metal-lic (Bonod et al. 2003, Giannattasio et al. 2004). Data films thinner than 30 nm could be very challenging for nano-fabrication, especially when including the data structures at the surfaces. The choice for the size of the data pits is slightly different for the different strategies.

To compare the readout performance of each of the strategies discussed below, we define here a contrast parameter η and the resolution criterion. The contrast parameter η quantifies the difference in the signal power P (reflected or transmitted intensity) between the presence and absence of a data structure, i.e.

$$\eta = \frac{P_{(\text{data})}}{P_{(\text{no data})}} \ . \tag{4.2}$$

The signal power $P_{(\text{data})}$ could be a noninverting ($\eta > 1$) or inverting ($0 < \eta < 1$) signal.

For resolution, we have adopted a criterion more stringent than the Rayleigh crite-rion (Born and Wolf 1999, Sec. 7.6.3). Two data pits are considered just resolved when the signal power returns to $P_{(\text{no data})}$ ($\eta \longrightarrow 1$) between two data structures, and this makes our resolution estimates conservative ones. It is to be noted that in our strate-gies, the binary data may be encoded with the data pits, or only with the edges of the data pits.[3]

To demonstrate that the effects in our proposed strategies are due to plasmons rather than other possibilities such as waveguide modes or diffracted evanescent waves,

---

[3]Detection of the edges is useful when binary data is encoded on the edges of the data pit, similar to the method in which conventional compact discs are en-coded (Pohlmann 1992).

we performed simulations in which we replace the silver plate with a silicon plate and compared the readout performance. For the reflection and transmission readout configuration, we simulated a typical geometry with $\gamma = 90$ nm and $\gamma' = 225$ nm (applies only to transmission configuration), and the results are shown in Fig. 4.2. Here we compare results of using a silicon plate instead of a silver plate. The data layer is either silver or silicon. The size of the data pit is taken to be 50 nm wide and 25 nm deep in Fig. 4.2a, and 40 nm wide and 15 nm deep in Fig. 4.2b. All other parameters are given as above.
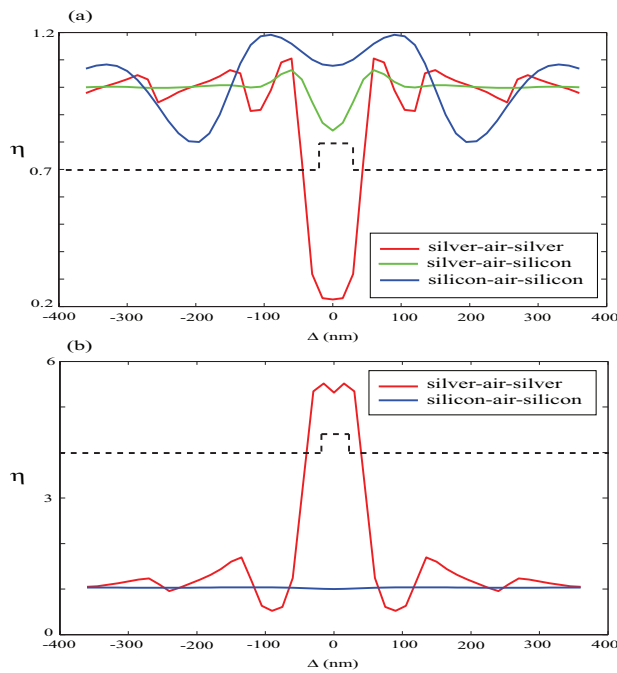
Figure 4.2: Typical readout performance for the (a) reflection and (b) transmission configurations, with $\gamma = 90$ nm and $\gamma' = 225$ nm. The size of the data pit is taken to be 50 nm wide and 25 nm deep in (a), and 40 nm wide and 15 nm deep in (b). The dashed line indicates the location of the data pit.

In Fig. 4.2a, we see that the reflected field intensity η differs significantly for a silver plate - silver data layer and silicon plate - silicon data layer geometry, and the latter

geometry does not support any surface plasmons. Oscillatory behavior in η for the silicon-silicon geometry also suggests that waveguide modes could be responsible for the effects for this case. Moreover, if waveguide modes were responsible for the effects in the proposed readout system, one would expect dramatic differences between a silver plate - silver data layer and a silver plate - silicon data layer geometry, since the effective index of the waveguide mode would be very different (Taylor and Yariv 1974, Prade et. al 1991, Berini 2000). However, the simulation results show that the main difference in the response in these two cases is the magnitude of η. In Fig. 4.2b, it is observed that the transmitted field intensity is almost constant in the case of the silicon plate - silicon data layer geometry. In addition, the significant enhancement in transmission for the silver plate - silver data layer geometry can be related to the presence of surface plasmons in the system. With these observations, we conclude that surface plasmons are responsible for the effects in the optical readout systems proposed in Fig. 4.1.

It is to be noted that the quantities $P_{\text{(data)}}$ and $P_{\text{(no data)}}$ used in Eq. (4.2) for evaluating η, are reflected and transmitted field intensities for the reflection and transmission readout configurations, respectively. Also, the incident field, i.e., the field illuminating the system, while taken to be a normally incident plane wave for the reflection configuration, is taken to be a normally incident Gaussian beam for the transmission configuration. As opposed to a Gaussian beam that carries a finite amount of optical power, a plane wave has infinite energy. By calculating the absolute transmission for the transmission configuration, one can better gauge the response of the performance of the system. Nevertheless, given that a plane wave is a good approximation to a Gaussian beam at least in the paraxial region, one does not expect drastic differences

for a qualitative analysis of the reflection readout system.

Having demonstrated the importance of the presence of surface plasmons in the proposed systems in Fig. 4.1 with typical readout results in Fig. 4.2, and with η and the resolution criterion specified, we proceed to evaluate the performance of the various strategies to achieve superresolved optical readout in the following sections.

4.1.1   Strategies for Reflection Readout

For the reflection readout configuration, our preliminary study on the readout performance as the size and position of the plasmon pits are varied led us to propose three distinct strategies that might be used to perform optical readout, each with its own system configuration:

1. A configuration designed to roughly map the shape of the data structures on the disc.

2. A configuration designed to respond only to the edges of the data structure. This configuration would be used if the binary data is encoded only at the edges of a data pit.

3. A configuration in which a plasmon pit is itself used as the near-field probe, and the slit acts only as a source/receiver of plasmons.

The three configurations are shown in Fig. 4.3. For all three configurations, the size of each plasmon pit is taken to be 50 nm wide and 50 nm deep. Configuration 1 and configuration 2 differ only in the separation between the plasmon pits and the center of

(a) Configuration 1 - pit detection

(b) Configuration 2 - edge detection

(c) Configuration 3 - plasmon pit as probe
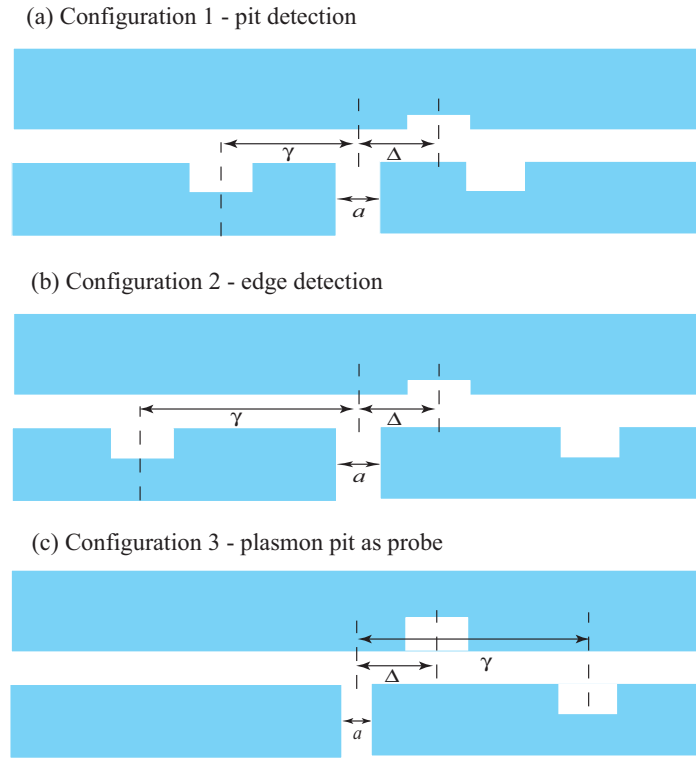
Figure 4.3: Illustration of the three configurations in the reflection readout system, and the relevant system parameters. Configurations 1 (a) and 2 (b) differ only in the position of the plasmon pits, while configuration 3 (c) has a different slit width and only a single plasmon pit.

the subwavelength slit, γ. From results of our simulations, it turns out that the system response can be quite different when γ is varied, which will be discussed shortly. For these two configurations, the width of the subwavelength slit a is taken to be 25 nm, and the size of each data pit is taken to be 50 nm wide and 25 nm deep.

In configuration 3, only one plasmon pit is present, itself acting as the near field probe. The size of the data pit is taken to be 50 nm wide and 50 nm deep, different from that in configurations 1 and 2. The size of the data pit is made identical to that of the plasmon pit to enhance the resonance effects when the plasmon pit and data pit coincide. It appears that the aligned pits act together to form a cavity, causing
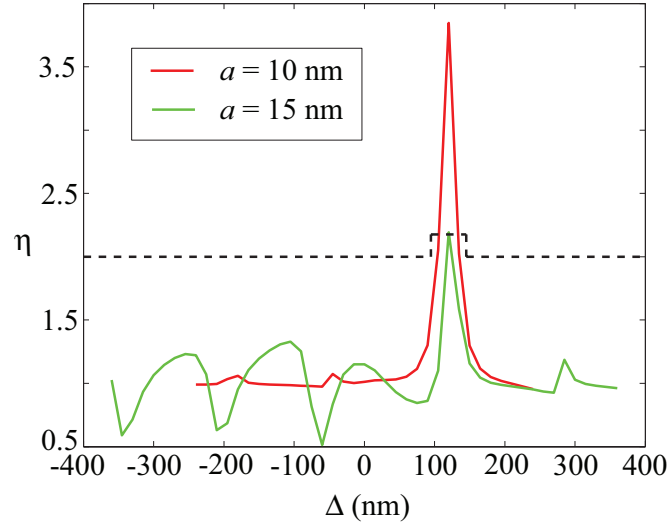
Figure 4.4: Illustrating the increased system oscillations and reduced readout contrast when the slit width is increased from 10 nm to 15 nm for configuration 3. The value of γ is taken to be 120 nm in this case.

the field within to resonate strongly. The slit acts merely as a plasmon source and collector in this case. The slit width a is kept as thin as seems practical, and is taken to be 10 nm. For larger values of a, interactions between the data structure and slit can increase system oscillations as shown in Fig. 4.4, similar to the standing wave like behavior observed in Fig. 4.2a.

The illuminating field is taken to be a normally incident plane wave for these three configurations. For an incident plane wave, the definition of the transmission T consists of two parts: the first is the integral of the normal component of the time-averaged Poynting vector S over the slit, and the second is the difference of the normal components of the time-averaged Poynting vector and that of the Poynting vector in the

absence of the slits, S^{inc}, integrated over the dark side of the plate. The result is normalized by the normal component of $S^{(0)}$. This may be written as

$$T = \frac{\int_{\text{slit}} S_z \, dx + \int_{\text{plate}} (S_z - S_z^{inc}) \, dx}{\int_{\text{slit}} S_z^{(0)} \, dx} \quad . \tag{4.3}$$

The subtraction in the second integral of the numerator corrects for the small part of the incident field which may tunnel directly through the plate itself. A transmission greater than unity roughly indicates more light is passing through the slit than is geometrically incident upon it. On the other hand, a transmission less than zero indicates that the light passing through a plate with a slit is less than that through a plate without a slit.

To help illustrate the roles of the plasmon pits, we show in Fig. 4.5 the effects on η in the absence of the plasmon pits, and when only the plasmon pits on the illuminated side are present in the reflection configuration. With no plasmon pits, the readout contrast η oscillates at a roughly constant amplitude, and is not suitable for performing readout operations. As plasmon pits on the illuminated side can help enhance the optical power reaching the data structure by coupling and reflecting more plasmons back into the slit (see Fig. 4.6), it was intially thought that adding these to the metal plate would improve the readout contrast. However, the results illustrated in Fig. 4.5 shows that not only did the readout contrast not improve, but it was worsened in the presence of the plasmon pits on the illuminated side. The value of γ' is taken to be 200 nm, where the transmission T through the plate is seen in Fig. 4.6 to be ~ 3.5 times higher than that in the absence of any plasmon pits. This inferior performance with only plasmon pits on the illuminated side arises from the fact that even though more light is transmitted to the data structure, the reflected throughput is still relatively weak compared to the
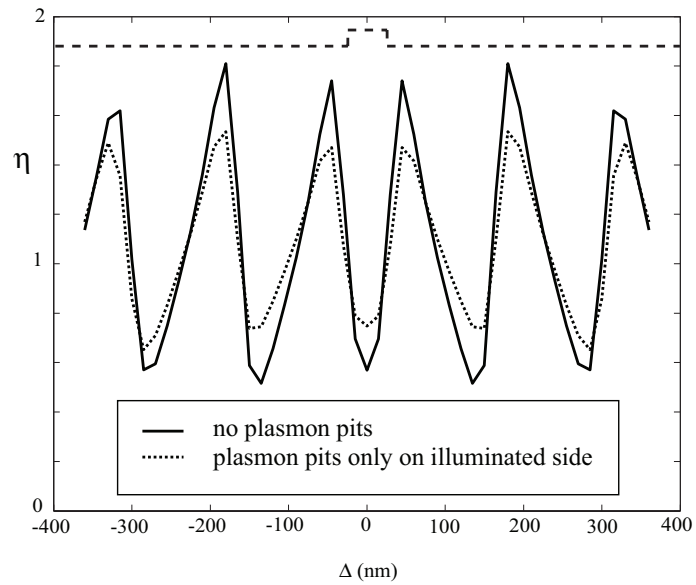
Figure 4.5: Readout performance for the reflection configuration in the absence of the plasmon pits (solid line), and when only the plasmon pits on the illuminated side are present (dotted line). For the latter case, γ' is taken to be 200 nm. The data layer is silver, and the size of the data pit is taken to be 50 nm wide and 25 nm deep. The dashed line indicates the location of the data pit.

backscattering by the plasmon pits themselves (Gbur et al. 2005). While they are of no use in this case for the reflection configuration, these pits on the illuminated side can be beneficial to the system performance for the transmission configuration presented in the next section.

It is fairly obvious by now that to design a readout system that uses surface plasmons optimally, one must consider the variation of a daunting number of parameters: the material of the metal plate, the material of the data layer, the slit width a, the size of the
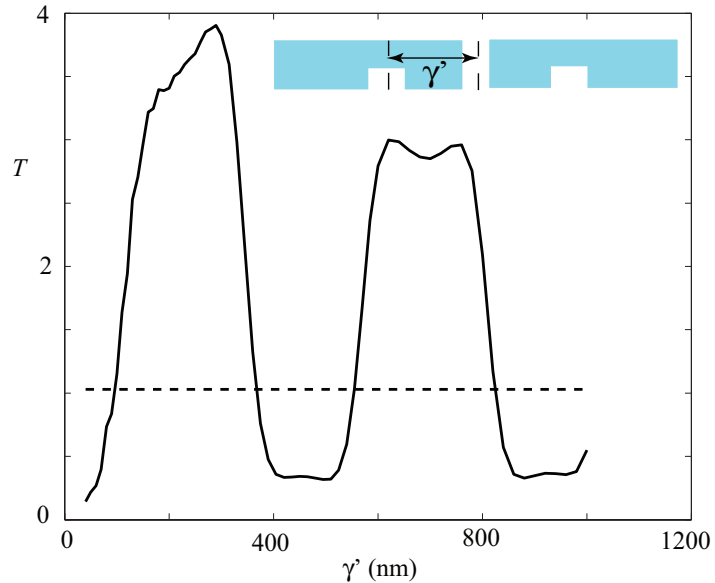
Figure 4.6: Transmission T as a function of γ' with a normally incident plane wave. The dashed line indicates the transmission through the plate in the absence of any plasmon pits.

data structure, and the size and location of the plasmon pits. We focus on optimizing a limited number of system properties that most directly relate to the effectiveness of the selected readout geometries. In particular, the position of the plasmon pits (γ) affects the field distribution in the system and hence the readout performance measured in our case by η and the resolution.

In configuration 1, γ is kept relatively small to suppress oscillations due to propagating plasmon modes and waveguide modes between the plate and data layer. Placing the plasmon pits near to the slit confines the electromagnetic fields to the im-
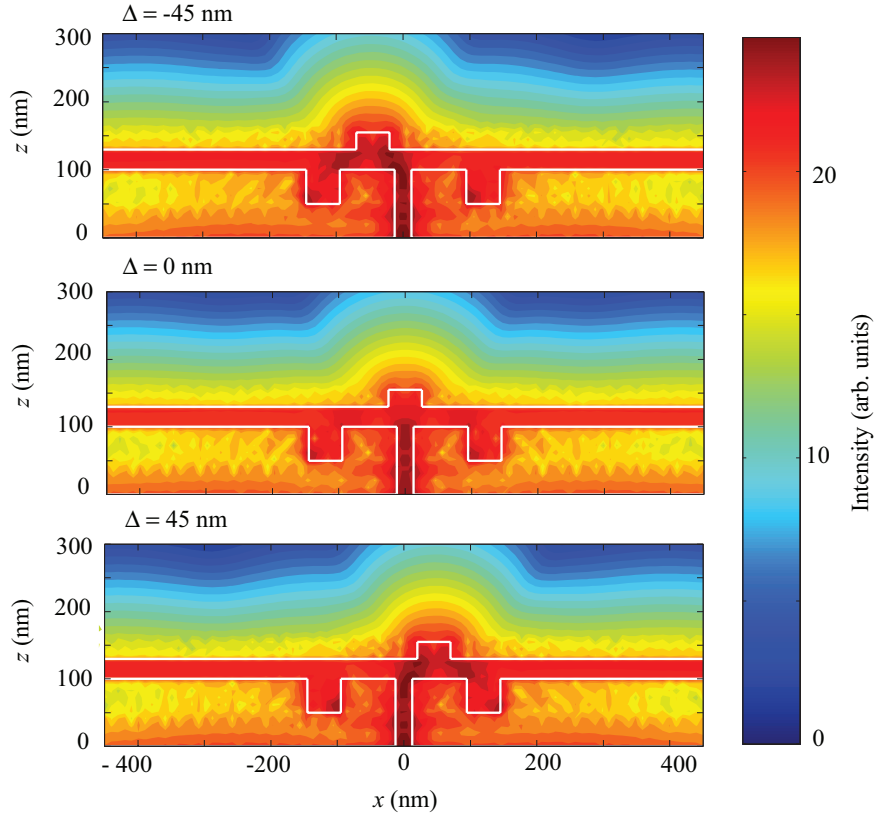
Figure 4.7: The electric field intensity distribution in the readout system when the data pit is in the vicinity of the readout slit, showing the intense field around the edges.

mediate vicinity of the slit, reducing oscillatory system response such as in Fig. 4.5. In configuration 2, the separation γ is extended from that in configuration 1 to allow standing-wave plasmon resonances to develop between the slit and the plasmon pits; because surface plasmons fields are strongest near the edges of the data structure (Schröter et al. 1997, Xie et al. 2004), evidently enhancing the plasmon resonances enhances the response of the system to the edges of the data structure. This can also be seen from the field distribution in the data pit when it is in the vicinity of the readout slit in Fig. 4.7, where is has been taken that $\gamma = 120$ nm.

For both configurations 1 and 2, we have performed numerical simulations of the transmission T through the silver plate as a function of γ without the data layer to find values of γ for which T is optimal. In addition, the readout constrast η is simulated as a function of γ to see how the system response varies with γ. In particular, we seek values of γ that will minimize reflection artifacts and provide accetpable readout contrast η at the same time. The findings from these simulations are discussed now.
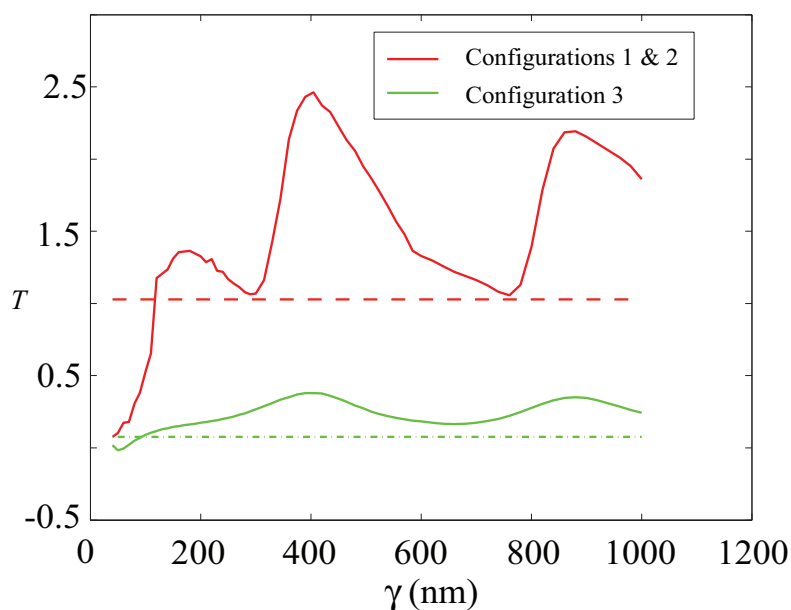
Figure 4.8: Normalized transmission through the metal plate for the different readout configurations. The red curve indicates the transmission for configurations 1 and 2, while the green curve indicates the transmission for configuration 3. The dashed line (1 and 2) and dashed-dot line (3) indicate the transmission in the absence of the plasmon pits.

As can be seen from Fig. 4.8, the transmission T is roughly a periodic function of γ, suggesting the standing wave behavior of the surface plasmons propagating between

the slit and the pits[4]. To prevent a similar standing wave from appearing between the slit and the data structure, the plasmon pits should be kept as close to the slit as possible. The transmission for configurations 1 and 2 has a local maximum in the range 120 nm $< \gamma <$ 240 nm, which suggests that the plasmon pits should be taken to lie within this range for a strong resonant response. In configuration 3, the choice of γ is again to suppress the oscillations between the slit and the data structure due to the plamson and waveguide modes in the system. From Fig 4.8, it is seen that, unlike for configurations 1 and 2, the transmission in this case did not vary substantially as a function of γ. In what follows, the qualitative behavior of the readout system is studied by examining the response of the system to data structures for various values of γ.

The gradual transition from the well-behaved readout reflectivity for configuration 1 to the growth of the oscillations for configuration 2 can be seen from Fig. 4.9, where plots of the system response are given for values of γ between 60 nm to 160 nm.

For small $\gamma(60-100$ nm$)$, the reflected power exhibits dips when the data structures are detected by the readout slit, as expected since the presence of the data pits reduces the amount of backscattering to the readout slit as compared to the case when the pits are not present. An optimal contrast ratio of about 60% is achieved with $\gamma = 80$ nm.

For larger $\gamma(110-140$ nm$)$, the reflected power exhibits strong resonances when the edges of the pits are detected by the readout slit. From Fig. 4.9, an optimal contrast ratio of about 700% is achieved when the edges are detected for $\gamma = 120$ nm. As the plasmon effects are most intense at the edges where the field tends to accumulate, the field resonates strongly when the edges of the data pits coincide with the readout slit.

---

[4]See, for instance, Schouten et al. (2005).

Standing plasmon waves suppressed in configuration 1 however, can be clearly observed in this case. Beyond $\gamma = 160$ nm, either the resonances are too weak to provide useful data detection or additional spurious resonances appear due to plasmon reflections from the data structure.

For configuration 3, we observe again in Fig. 4.10 the growth of oscillations in the regions around the slit, plasmon pit, and data pits as the separation γ is increased from $60-240$ nm. The plots suggest that the optimal γ in this case is 120 nm, with a contrast ratio of about 280%. A resonance in the response of the system can be seen whenever
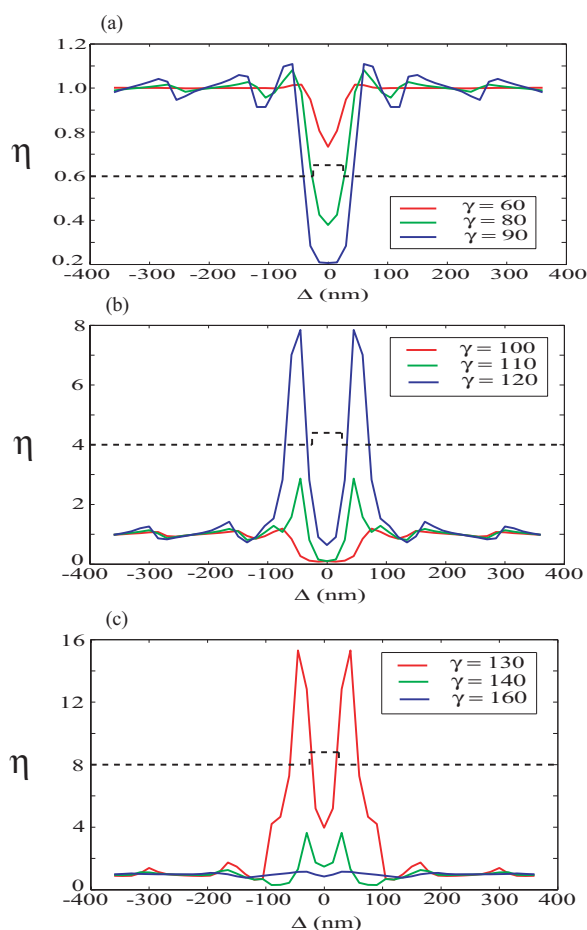
Figure 4.9: Readout η for various values of γ for a single data pit in configurations 1 and 2. The dashed line indicates the location of the data pit.
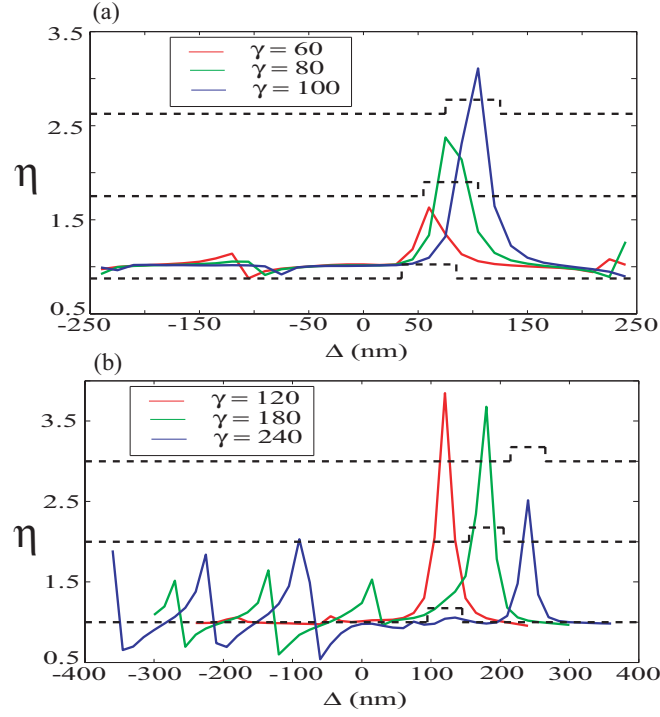
Figure 4.10: Readout η for various values of γ for a single data pit in configuration 3. The dashed line indicates the location of the data pit.

the data structure passes the location of the probing plasmon pit. We have plotted the electric field intensity distribution in the system to illustrate this resonant effect in Fig. 4.11. However, significant oscillations appear for values of γ beyond 180 nm.

Next, we look at the resolution of the three configurations by examining their ability to resolve a pair of spatially separated data pits. For configuration 1, γ is taken to be 80 nm. For configurations 2 and 3, γ is taken to be 120 nm. The simulation results in Fig. 4.12 show two data structures just resolved according to our resolution critierion. It is to be noted that the parameter δ used for the horizontal axis in Fig. 4.12 is defined as the distance *between* the center of the two data pits and the subwavelength slit, i.e., the slit coincides with each of the data pits at

$$\delta = \pm\Delta \ , \tag{4.4}$$
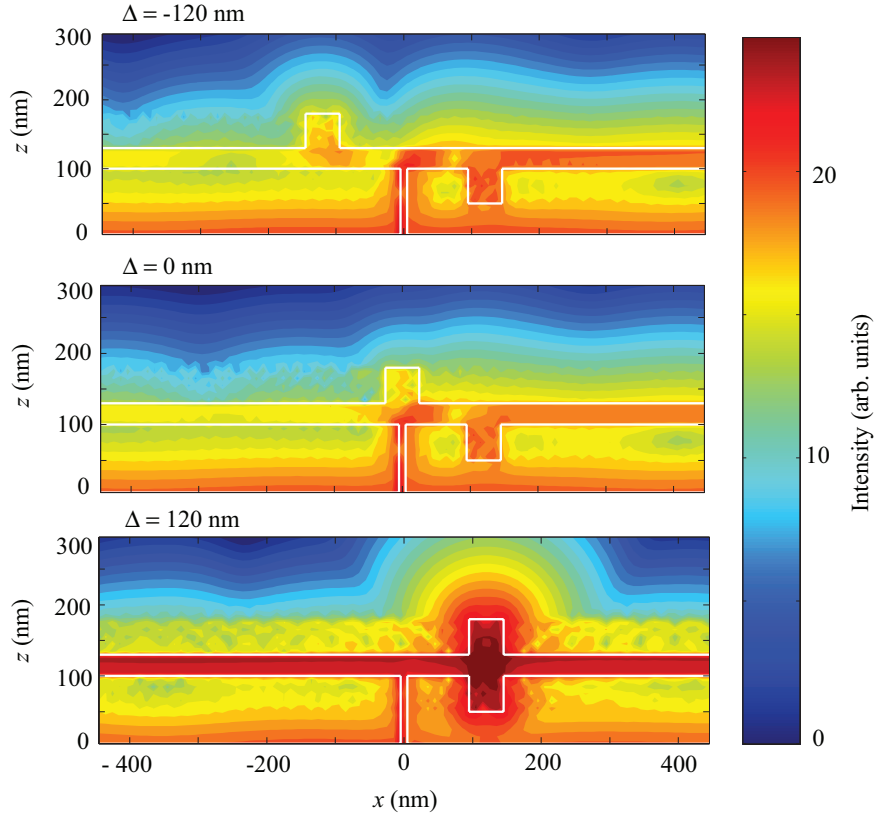
Figure 4.11: The electric field distribution in the readout system as the data pit approaches the probing plasmon pit.

where

$$\Delta = \frac{\text{data pit separation}}{2} \ . \tag{4.5}$$

It can be seen that two data pits separated by 120 nm, 240 nm, and 120 nm, could be resolved by each configuration, respectively. The performance for each of the configurations are summarized in Table 4.2.

Comparison of the results for the three configurations reveal that each is associated with its own advantages and disadvantages. The highest contrast ratio of any system is found in configuration 2, with an excellent η of 700%. This system exhibits significant oscillations of the reflected power when the data structure is away from the vicinity of
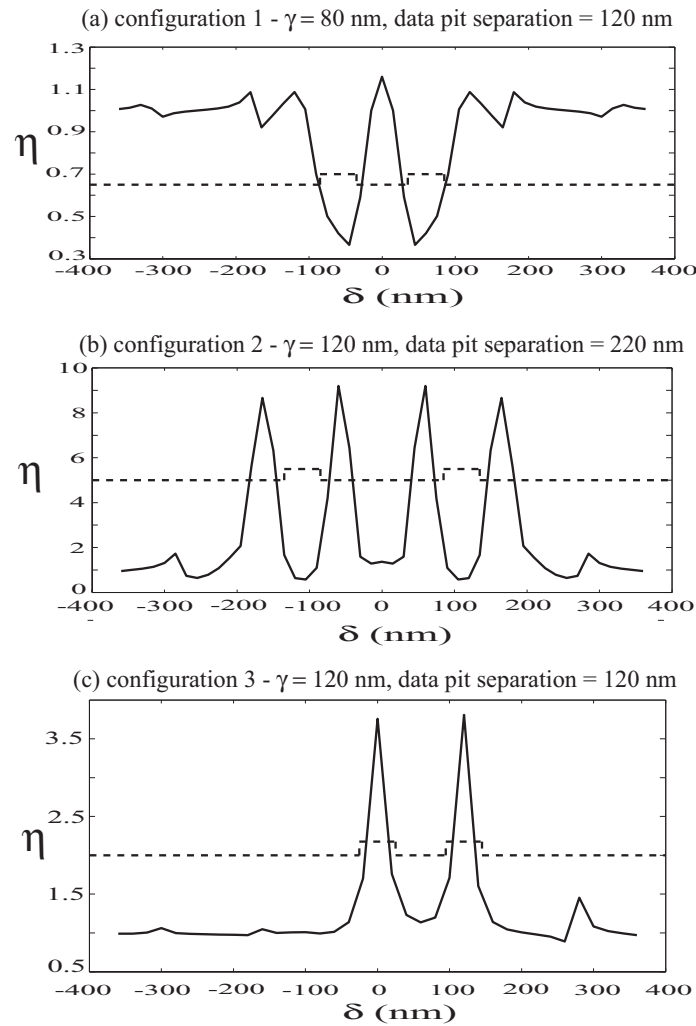
Figure 4.12: Demonstrating the resolution of two data pits for the three configurations.

the slit, as can be seen in Fig. 4.9. But on account of the high contrast ratio, these oscillations are not necessarily a problem. However, its resolution is the worst of the three systems at 220 nm, as can be seen from Fig. 4.12. This is attributed also to the fact that there are two edges associated with a single data pit, and hence four detections instead of two are required to resolve two data pits.

Configuration 3 offers a better resolution, at 120 nm, and also has an excellent η of 280%. Furthermore, as can be seen in Fig. 4.10, values of γ between $80 - 180$ nm

result in a significant η far above the oscillations of the reflected power. Nevertheless, this system uses a very small slit size of 10 nm, and therefore offers the lowest overall reflected power of any systems, which will make it more susceptible to system noise.

Table 4.2: Summary of results for reflection readout.

Configuration	γ (nm)	Number of plasmon pits	$\eta(\%)$	Resolution (in terms of λ)
1	80	2	60	$\lambda/4.2$
2	120	2	700	$\lambda/2.3$
3	120	1	280	$\lambda/4.2$

On the other hand, configuration 1 exhibits minimal oscillations in the reflected power (see Fig. 4.9) while giving the same resolution of 120 nm. By keeping the plasmon pits close to the slit to suppress unwanted oscillations, this system exhibits the best stability among the three. It offers the lowest η of 60% however.

In terms of fabrication, all these systems require rather precise specification of the slit width, and the size and position of the plasmon pits. As any realistic fabrication will involve some imperfections, it is of interest to examine how sensitive the readout is to variation in parameters. We found that a tolerance of ± 10 nm is acceptable for the value of γ in configurations 1 and 2. Configuration 3, with a slit width of only 10 nm, requires any deviation from its specified slit width to be less than 5 nm.

As noted, each of the systems described here has its own advantages and limitations. The choice of an optimal system will depend on what characteristics the designer finds most important (readout contrast, resolution, stability, signal strength). The systems

we have investigated so far rely on the reflected power for readout, as in conventional optical data systems. To exploit directly the plasmon-enhanced transmission effects, we next consider strategies based on a transmission readout configuration.

4.1.2 Strategies for Transmission Readout

We now extend the investigation in the previous section to study near field optical readout by detection of the transmitted field intensity. The aim is to achieve superresolution through the modulation of the surface plasmon-enhanced transmission. There are several reasons a transmission readout system could offer more benefits than one based upon reflection.

First, this clearly allows for the surface plasmon-enhanced transmission to be exploited directly. Second, the amount of useful signal in the reflection configuration is relatively limited as backscattering in the vicinity of the subwavelength slit contributes significantly to the overall reflected field. Third, plasmon pits on the illuminated side of the metal plate, which degrade the readout contrast in the reflection readout configuration (Gbur et al. 2005), can now be incorporated to enhance the transmission (García-Vidal et al. 2003), directly increasing the strength of the signal. Therefore, in addition to the plamson pits on the dark side separated by 2γ, we have also included plasmon pits (separated by $2\gamma'$) on the illuminating side. Finally, the detection geometry for the transmission configuration (see Fig. 4.1) may be more straightforward to implement in practice.

Instead of using an incident plane wave as for the reflection readout case, we take a Gaussian beam as the incident field for this case. Since the optical power carried by a

Gaussian beam is finite and can be computed through its associated Poynting vector, we can calculate the absolute transmission to better quantify the system response. The transmitted power T is normalized to the incident field such that

$$T = \frac{\int_{-\infty}^{+\infty} S_{z1} dx}{Y_0 \int_{-\infty}^{+\infty} |E^{(inc)}(x,z)|^2 dx} , \tag{4.6}$$

where,

$$Y_0 = \sqrt{\frac{\epsilon_0}{\mu_0}} , \tag{4.7}$$

with S_{z1} the normal component of the time-average Poynting vector emerging from the data layer, and $E^{(inc)}(x,z)$ the electric field amplitude of the incident Gaussian beam. For simulation purpose, the beamwidth of the Gaussian beam at full width at half maximum is taken to be 530 nm ($\lambda = 500$ nm).

The integral in the numerator of Eq. (4.6) implies that it is possible to capture the total transmitted power. This can be experimentally problematic if there is a lot of power scattered at highly oblique angles. To address this issue, we have studied also the far-field radiation patterns for the geometries considered. Our simulations revealed that the scattered power lies predominantly in the forward direction. This agrees with the observation of a well-localized field emerging from the data layer, as we shall see in the simulation results shortly.

It is also known through Fourier analysis that the field emerging from an aperture of a very narrow extent diverges very quickly and produces a far-field radiation pattern that varies marginally as a function of the viewing angle. One can understand this effect from the position-momentum uncertainty relation. By localizing photons to the position of the subwavelength aperture, their momentum, and hence their wave vectors ($p = \hbar k$)

are likely to take on a larger range of values. As such, there is a high probability for the photons to be spread evenly over a wide range of angles at the detection plane. Moreover, the subwavelength extent of the localized field at the slit in this case closely resembles a point source. Therefore, we have reason to believe that the system can capture the bulk of the transmitted optical power.

The thickness of the silver data layer is taken to be 30 nm, and the size of each data pit is taken to be 40 nm wide by 15 nm deep. As explained earlier, the choice of such a thin metallic layer reduces the effects of absorption, and allows for possible plasmon-assisted field enhancement effects. We further speculate that it would be possible also to achieve similar field enhancement effects with a dielectric data layer (such as silicon) instead, but with metallic strips in place of the air pits. In the discussion that follows, we will consider the data structure to be either a silver layer with air pits, or a silicon layer embedded with strips of silver as the data pits, referring to them as 'silver data layer' and 'silicon data layer' respectively. Similar to the previous section, we will be concerned with the location of the plasmon pits.

For the plasmon pits on the light side, the choice of γ' is such as to obtain enhanced optical transmission through the slit. As discussed previously, the size of the plasmon pits are taken to be 40×40 nm^2 in the transmission readout configuration, different from the choice of 50×50 nm^2 in the reflection readout. Numerical simulations with only the subwavelength slit and light side plasmon pits in the silver plate (without the dark side plasmon pits and data layer) show that maximum transmission occurs at $\gamma' \sim 225$ nm. Results for plasmon pits of 40×40 nm^2 and 50×50 nm^2 are shown in Fig. 4.13. It is seen that the 40×40 nm^2 plasmon pits can give slightly more transmission than
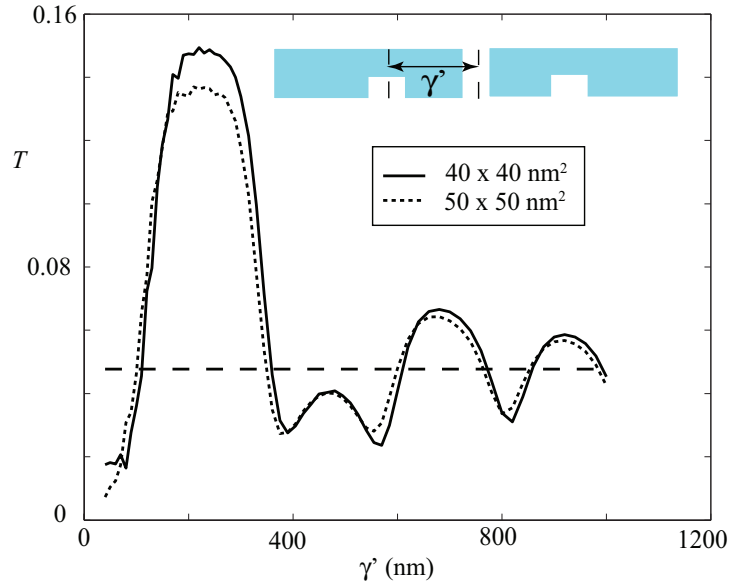
Figure 4.13: Transmission T as a function of γ' with a normally incident Gaussian beam for plasmon pit size of 40×40 nm^2 (solid line) and 50×50 nm^2 (dotted line). The dashed line indicates the transmission through the slit in the plate in the absence of any plasmon pits.

the 50×50 nm^2 plasmon pits. This demonstrates that surface plasmon effects do not scale linearly with the size of the structures, and that rigorous numeric simulations are necessary to study these effects. Also, the value of γ' found here for the incident Gaussian beam is similar to that found for a incident plane wave (as for the reflection readout configuration). In that case, γ' was taken to be ~ 200 nm (Fig. 4.6). This shows that our simulations with an incident plane wave or with a incident Gaussian beam do produce results that are similar qualitatively.

The choice of γ is determined through numerical simulations of the contrast ratio η

for various values of γ with a single data pit. From our findings in the previous section, we expect the dark side plasmon pits to aid both in confining the enhanced field effects to the vicinity around the data structure, and in suppressing undesirable oscillations of the transmission. Fig. 4.14 shows that for both the silver and silicon data layer, the optimal value of γ is ~ 75 nm. It is worth noting that the analysis here yields very similar results to configuration 1 investigated for reflection readout, where the optimal value of γ was found to be ~ 80 nm.

A closer examination of Fig. 4.14 reveals that for the silicon data layer, the transmission dips significantly when the edges of the data pit coincide with the center of the slit. This can arise from increased backscattering at the discontinuity along the silver/silicon/air boundaries. Such a backscattering mechanism, typically undesirable in many applications, offers the possibility to detect the edges of the data pits through the associated transmission dips, i.e., the presence of an edge is indicated with a transmission minima. However, on account of our definition of the contrast ratio η, the maximum achievable η for edge detection in this case is limited to unity; the transmission peaks play no role in edge detection, unlike in pit detection. We like to point out that the strategy to detect the edges differ from that in configuration 2 in the previous section, where plasmon resonances at the edges had been exploited and η of 700% was achieved. The reason we do not employ the same strategy here for edge detection is because the size of the data pit in the transmission configuration is 40×15 nm^2, smaller than the 50×25 nm^2 used in the reflection readout case. With our stringent requirement for resolution described in Sec 4.1, discriminating the plasmon resonances at each of the edges could be significantly more challenging. This is due to the fact that

(a)

(b)
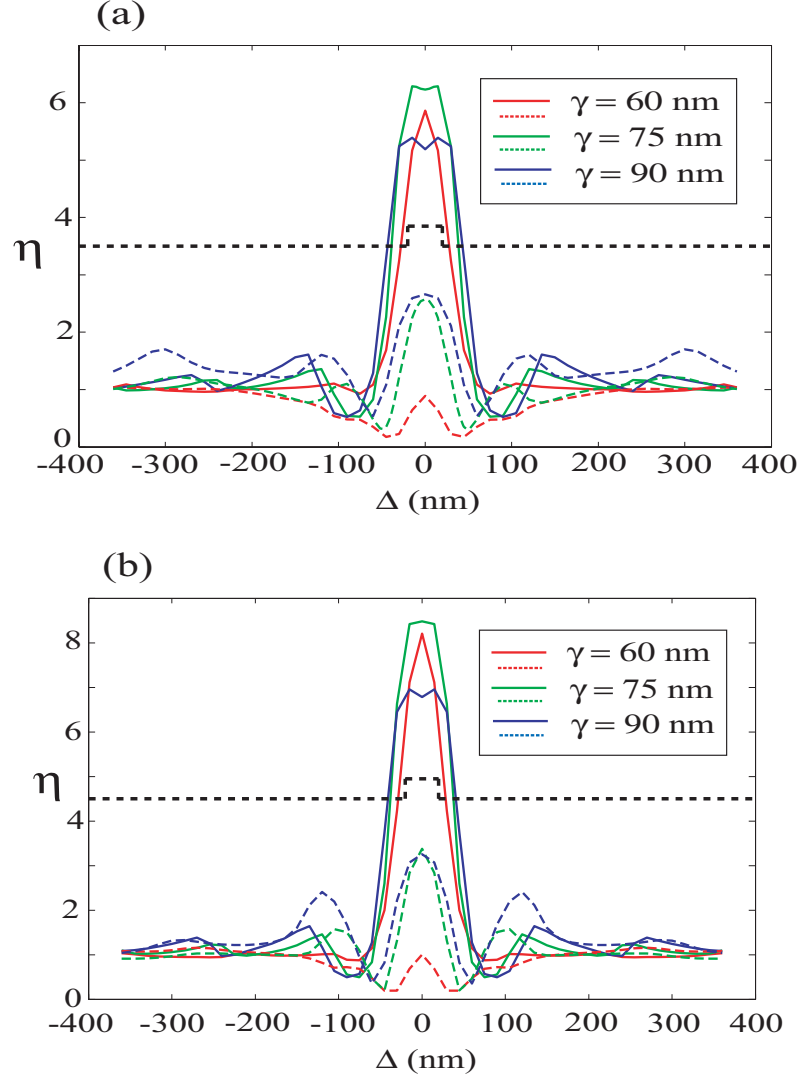
Figure 4.14: Readout η for one data pit: (a) with $\gamma' = 225$ nm, and (b) without light side plasmon pits. Solid lines are for silver data layer and dotted lines are for silicon data layer. The black dashed line indicates the position of the data pit.

$P_{(\text{data})}$ must approach $P_{(\text{no data})}$ in a gradual manner, thus potentially restricting η to only values close to unity.

It is to be noted that the contrast ratio η, as defined in Eq. (4.2), has been normalized to the transmission in the absence of the data structure ($P_{(\text{no data})}$), and does not provide

adequate information on the actual amount of transmission through the system. As such, we have also analyzed the distribution of the electric field intensity in the system (see Fig. 4.15). It can be observed that the field intensity through the silver data layer is more intense than that through the silicon data layer, even though silicon is more transparent than silver at $\lambda = 500$ nm . With $\gamma = 75$ nm and no plasmons pits on the light side, $P_{(\text{no data})}$ through the silver and silicon data layers is 0.31%, and 0.22% respectively. When the light side plasmon pits are present ($\gamma' = 225$ nm), $P_{(\text{no data})}$ increases to 1.33% and 1.06%. These simulations are repeated with the silver plate replaced with a silicon plate, and $P_{(\text{no data})}$ was found to increase from 3.32% to 3.79% (silicon data layer), and from 5.59% to 6.53% (silver data layer). The higher value of $P_{(\text{no data})}$ in the case of the 100 nm silicon plate is expected due to the decrease in absorption, but clearly the transmission increases minutely in the presence of the light side plasmon pits. These results again suggest the presence of surface plasmon-enhanced field effects through the silver sections. While the silicon data layer provides a lower readout contrast, the plots in Fig. 4.15 also show that the silver strip embedded within is more effective in confining the surface-plasmon enhanced transmission in the near field, and has the potential to provide higher resolution.

Based on the above discussion, the following strategies for readout were investigated:

1. Data pit detection with silver data layer.

2. Data pit detection with silicon data layer.

3. Edge detection with silicon data layer.

For all three cases, the readout performance with and without the light side plasmon
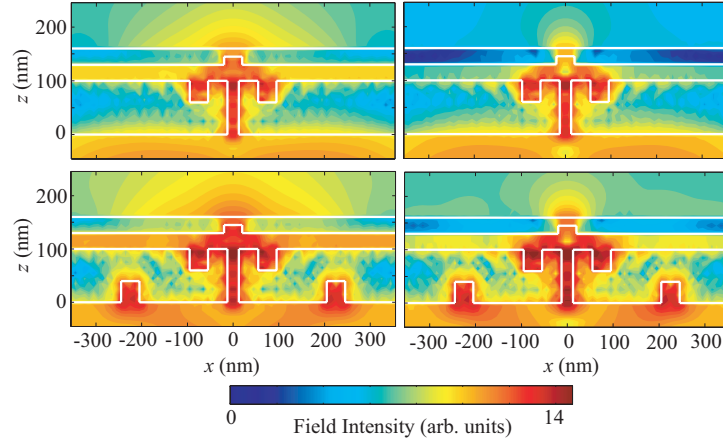
Figure 4.15: Field intensity distribution with and without light side plasmon pits, $\gamma = 75$ nm and $\gamma' = 225$ nm. Plots on the left column and right column for silver and silicon data layer, respectively.

pits are compared to assess the overall improvement when the transmission is enhanced. The simulation results are shown in Fig. 4.16. It is to be noted that the parameter δ used for the horizontal axis in Fig. 4.16 is defined as the distance *between* the center of the two data pits and the subwavelength slit, as with the Eqs. (4.4) and (4.5). For each configuration [(1), (2), and (3)], resolution up to 110 nm$(\lambda/4.5), 90$ nm$(\lambda/5.6)$, and 180 nm$(\lambda/2.8)$ was obtained, with readout contrast of around 600%, 200%, and 60% respectively. These results are summarized in Table 4.3.

As in the previous section for reflection readout, we find that each strategy is associated with its advantages and limitations. The silicon data layer provides better resolution for data pit detection, an effect we attribute to the more effective localiza-
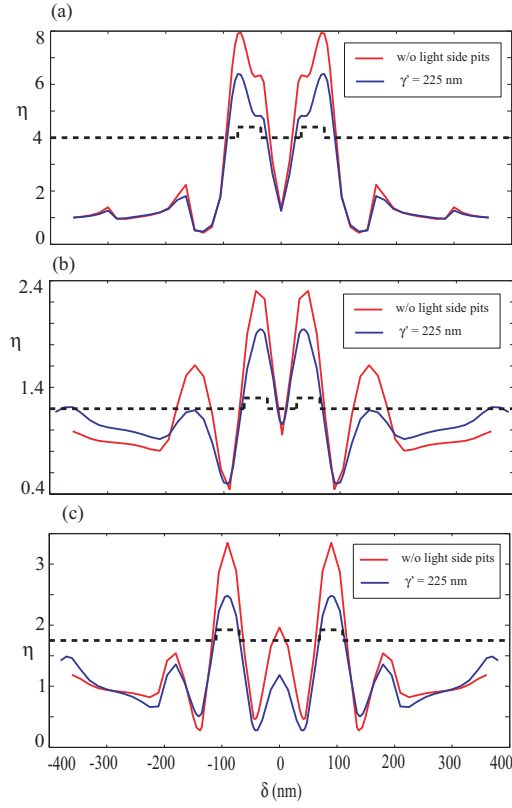
Figure 4.16: Readout η for two data pits: (a) pit detection in silver data layer, (b) pit detection in silicon data layer, and (c) edge detection in silicon data layer. The solid lines show the results without light side plasmon pits. The dashed lines indicate the respective positions of the data pits.

Table 4.3: Summary of results for transmission readout. For all three configurations, $\gamma = 75$ nm, and $\gamma' = 225$ nm.

Configuration	$P_{no\ data}(\%)$	$\eta(\%)$	Resolution (in terms of λ)	Remarks
1	1.33	600	$\lambda/4.5$	$\eta > 1$
2	1.06	200	$\lambda/5.6$	$\eta > 1$
3	1.06	60	$\lambda/2.8$	$0 < \eta < 1$

tion of the field enhancement effects. This better localization is achieved at the expense of the enhanced transmission $P_{(data)}$ however, leading to the lower readout contrast.

Edge detection doubles the features to be detected, and it is therefore not unreasonable for the resolution to be inferior.

4.2 Summary and Future Work

In summary, we have demonstrated numerically that with careful selection of system parameters, surface plasmon-enhanced transmission can be manipulated to achieve superresolution in near field optical readout. It is to be noted that the results here could be extended to a three-dimensional geometry in which the subwavelength-width slit is replaced by a subwavelength-radius hole. Such a configuration is the natural choice for a full experimental realization of a readout system, but is extremely difficult to simulate numerically and computationally intensive. Our results give a qualitative idea of what kind of behaviors to expect in such a system. Our simulations suggest that it is possible to optimize the application of these effects for near field optical read-out. In the strategies simulated, superresolution of data structures up to $\sim \lambda/5$ is achievable. It is hoped that the investigation of these strategies would lead to the effective employment of surface plasmon-enhanced transmission effects for optical readout and other near field applications. We end this chapter by estimating how much improvement in the data storage capacity a plasmon-assisted superresolved optical readout system can offer. For a 120 mm disc with a typical useful storage area (A_s) of 92.7 cm^2 (van de Nes et al. 2006), a rough estimate of the storage capacity C in bytes is

$$C = \frac{1}{8} \frac{A_s}{\Delta l_{min}^2} \; , \tag{4.8}$$

which yields $C = 180$ Gb assuming $\Delta l_{min} = \lambda/5$, and $\lambda = 400$ nm as for the current Blu-Ray technology. This is nearly 8 times more than its current storage capacity of 23 GB per layer. This estimate is conservative in view of the progress in the achievable precision and system tolerancing of modern digital systems. Thus, there is much potential for superresolved optical readout systems to increase data storage capacity tremendously.

CHAPTER 5: PLASMON-ASSISTED OPTICAL COHERENCE MODULATION

The state of coherence of a light field is an important property that affects its interference-causing capability in both classical and quantum optical systems (Mandel and Wolf 1995). In classical systems, variable coherence-optics applications include examples such as speckle reduction (Dainty 1975), optical coherence tomography (Huang et al. 1991), and beam propagation through turbulence (Wu 1990, Gbur and Wolf 2002, Ricklin and Davidson 2002, Korotkova et. al 2004a). In the realm of quantum optics, an understanding of coherence is essential to the entanglement of quantum states[1]. The effect of entanglement allows for teleportation of the quantum states, which offers potential for the development of applications such as quantum cryptography (Ekert 1991, Gisin and Thew 2007) and quantum computing (Spiller 1996, Zeilinger 2000).

Here we focus on the coherence of classical optical fields. There are two concepts used to characterize the coherence of classical optical fields - temporal and spatial coherence (Wolf 2007). For an optical source, the temporal coherence portrays the effects of its finite bandwidth, and the spatial coherence describes the effects arising from its finite spatial extent. Optical sources or fields are never strictly monochromatic

---

[1]It is interesting to note that there is a mathematical duality between partial coherence and partial entanglement (Saleh et al. 2005).

in practice, i.e., they always have a finite effective bandwidth $\Delta\omega$. If $\Delta\omega$ is very narrow, they are described as being *quasi-monocromatic.* A measure of the temporal coherence is the coherence length

$$\Delta l_c = \frac{(\bar{\lambda})^2}{\Delta\lambda} \ , \tag{5.1}$$

where $\bar{\lambda}$ is the mean wavelength and $\Delta\lambda$ the effective wavelength range.

While an optical field is never strictly monochromatic, it may be completely spatially coherent not only at a single frequency but for all frequency components in its spectrum (Wolf 2003). A planar source of finite extent illuminating an aperture plane $\mathcal{A}$ at a distance R away from it has a coherence area of ΔA_c on the plane $\mathcal{A}$ given by (Wolf 2007, Sec 1.3)

$$\Delta A_c = \frac{(\bar{\lambda})^2}{\Delta\Omega} \ , \tag{5.2}$$

with the solid angle

$$\Delta\Omega = \frac{S}{R^2} \ , \tag{5.3}$$

where S is the area of the source. If a pair of pinholes on plane $\mathcal{A}$ are situated within an area ΔA_c, the pinholes will produce visible fringes on an image plane $\mathcal{B}$ sufficiently far from $\mathcal{A}$, as in the Young's interference experiment.

With the coherence length Δl_c and coherence area ΔA_c defined in Eq. (5.1) and Eq. (5.2) respectively, one can further introduce the concept of the *coherence volume* ΔV_c given as (Mandel and Wolf 1995, Sec 4.2.3)

$$\begin{aligned} \Delta V_c &= \Delta l_c \Delta A_c \\ &= \frac{1}{\Delta\Omega} \left(\frac{\bar{\lambda}}{\Delta\lambda} \right) (\bar{\lambda})^3 \ . \end{aligned} \tag{5.4}$$

The coherence volume represents the region of space throughout which photons in the field are indistinguishable from each other. It also has a counterpart, known as the so-called cell of phase space[2] in quantum mechanics.

Here, we will be concerned with modulating the spatial coherence properties of light. The spatial coherence of an optical source is a fundamental characteristic that determines numerous properties of the field it produces, such as its spectrum and directionality (Wolf 1986, Wolf 1978). It is a measure of the "statistical similarity" between any two points within the field's domain (Wolf 2007). Fields that are fully coherent can interfere to form interference fringes with maximum fringe visibility or contrast, while fully incoherent fields only add in intensity. Also, both the spectrum and the polarization of the field that is generated by a partially coherent source may change on propagation (Wolf and James 1996, James 1994, Korotkova et al. 2004b, Salem et al. 2004), and these changes are dependent on the degree of coherence. Therefore any new method to influence the state of coherence is of fundamental importance.

Techniques of variable-coherence optics that have been proposed include diffusion by rotating ground-glass plates (Martienssen and Spiller 1964, Gonsiorowski and Dainty 1983), acousto-optic modulation (Ohtsuka 1986, Turunen et al. 1990), liquid crystal light modulation (Scudieri et al. 1974, Carter and Bertolotti 1978), and Bragg diffraction by holographic gratings (Vahimaa and Turunen 1997). Most of these techniques allow for either the reduction of the spatial coherence with mechanically rotating diffusers (potentially slow), or increasing it by improving the directionality of a diffracted

---

[2]For details on the notion of phase space, see for instance, Agarwal and Wolf (1970) or Sec. 3.7 in Marcuse (1982).

beam.

As an alternative, we propose here to modulate the spatial coherence of light with surface plasmons. In a recent work, surface plasmons were found to be responsible for enhancing and suppressing the optical transmission in a Young's interference experiment (Schouten et al. 2005). Since the coherence properties of fields are intimately related to their ability to interfere, we were thus prompted to investigate if surface plasmons propagating between the slits in Young's interference experiment can also modulate the spatial coherence of light (Gan et al. 2007). Our simulations demonstrate that surface plasmons propagating between the slits can indeed both enhance or suppress the spatial coherence properties of light emanating from the apertures. Further investigations with a three-slit interferometer (Gan and Gbur 2008) and a symmetric three-hole interferometer in an equilateral geometry reveal similar results. This chapter is primarily devoted to the presentation and discussion of these results.

It is worth noting that the ability of surface plasmons to enhance the spatial coherence has been demonstrated experimentally in a Young's interferometer where each slit was separately illuminated by an independent optical source (Kuzmin et al. 2007). In that experiment, it was also found that far-field interference fringes are observable even when only a single slit is illuminated. Both observations can be understood by considering that surface plasmons generated at one slit can propagate to the other slit, where they can couple back into light. This light from the other slit is in turn radiated to the far-field, where it interferes with light directly transmitted from the illuminated slit to form fringe patterns (see illustration in Fig. 5.1).

Modulation of the spatial coherence with surface plamsons may find potential ap-

Interference fringes from spectral intensity $I(\theta)$ in the far-field.
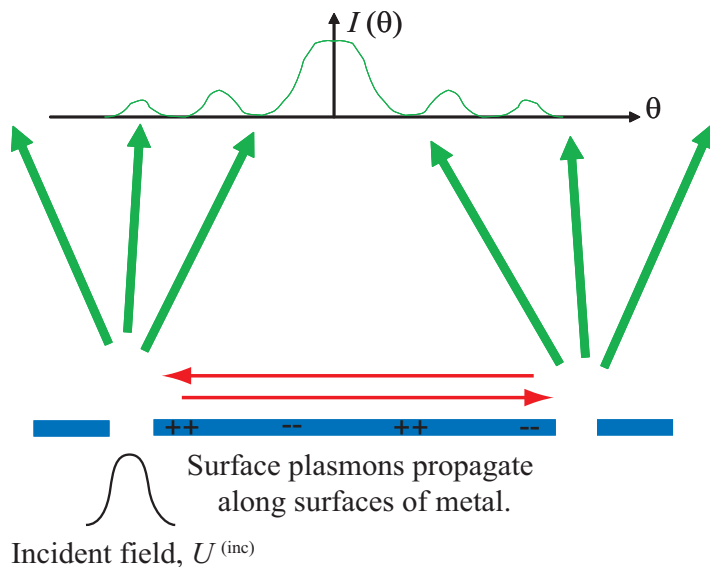
Figure 5.1: Illustrating how illuminating just one of the apertures in a Young's interference experiment can still result in interference fringes in the far field with surface plasmons.

plications in producing coherent light from spontaneous emission as a result of resonant coupling between semicondcutor qunatum wells and surface plasmons (Gontijo et al. 1999), and in plamsonic interferometry devices such as all-optical modulators made with quantum dots (Pacifici et al. 2007).

5.1 Spatial Coherence Conversion with Surface Plasmons in Interferometers

In this section, we demonstrate through theoretical analysis and numerical simulations that surface plasmons propagating between subwavelength apertures on a metal plate can modulate the spatial coherence of light emanating from the apertures. We begin by investigating the Young's interference experiment, and show that the degree of coherence

of the light emanating from the two slits can be greater or lesser than that of the illuminating field, depending on the slit separation.

It is not clear, however, if the results for Young's double-slit configuration extend readily to an array of holes or slits. Because the field scattered from each hole or slit can interact with that scattered from another in the near-field, the interaction between light and surface plasmons cannot be described accurately through a straightforward superposition of the response from each individual aperture. Thus, it is not immediately obvious how the inclusion of additional apertures affects the coherence and radiation properties. Furthermore, unlike the double-slit case where one can examine the coherence properties of light from just two apertures, for an array one would have to study the coherence properties of light as a global effect resulting from contributions from each aperture in the array.

Working towards a coherence converting device with an array of subwavelength apertures, we have therefore analyzed a three-slit interferometer in which similar modulation effects was observed. The simulations indicate that the additional center 'barrier' slit can serve not just to decrease the effects of the plasmons propagating from one slit to the other, but also to preserve, and even enhance these effects. Further extending our analysis to a symmetric three-hole interferometer arranged in an equilateral geometry where the plasmonic contributions at each hole are identical, we simulated how the fringe contrast of the far-field pattern changes as the distance between the holes is varied.
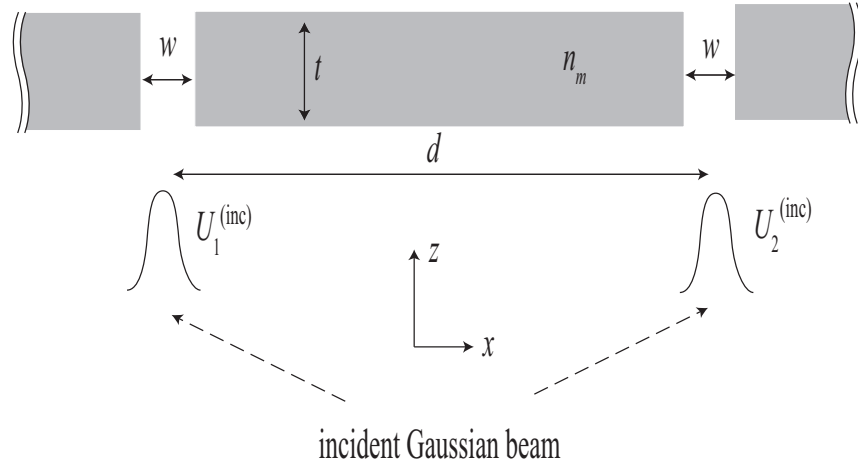
Figure 5.2: Illustrating the geometry for the Young's interferometer.

5.1.1 Young's Interferometer

We consider the Young's interferometer shown in Fig. 5.2. The wavelength and the beamwidth of the incident beam are taken to be 600 nm and 750 nm respectively. The slit separation d is taken to be at least 1000 nm, to avoid joint illumination of the slits with a single beam. The metal plate of thickness $t = 200$ nm is assumed to be silver, whose refractive index at 600 nm is taken to be $n_m = 0.21 - i3.27$, after Johnson and Christy (Johnson and Christy 1972).

The fields U_1 and U_2 represent the x component of the electric field, which is the dominant radiating field component. It is assumed that the spectral densities of the fields incident at each of the slits are identical, i.e., $|U_1^{(\text{inc})}(\omega)|^2 = |U_2^{(\text{inc})}(\omega)|^2 = |U_0(\omega)|^2$,

ω being the angular frequency of light. The spectral degree of coherence between the incident fields and the fields emerging at the two slits are denoted as $\mu_{12}^{(inc)}(\omega) = \mu_0(\omega)$, and $\mu_{12}(\omega)$, respectively. The spectral degree of coherence between the two fields $U_1(\omega)$ and $U_2(\omega)$ is defined as

$$\mu_{12}(\omega) = \frac{W_{12}(\omega)}{\sqrt{S_1(\omega)S_2(\omega)}}, \tag{5.5}$$

where

$$W_{12}(\omega) = \langle U_1(\omega)^* U_2(\omega) \rangle , \tag{5.6}$$

is the cross spectral density, and $S_i(\omega) = W_{ii}(\omega)$ is the spectral intensity of the field. The asterisk denotes complex conjugation, and the brackets indicate averaging over an ensemble of space-frequency realizations (Wolf 2007, Sec 4.1). The modulus of the spectral degree of coherence is confined to the values

$$0 \leq |\mu_{12}(\omega)| \leq 1, \tag{5.7}$$

with zero indicating complete incoherence, and unity indicating complete coherence.

Based on the Green tensor formalism described in Chap. 3, we have performed rigorous numerical simulations for the interferometers in Fig. 5.2. Due to invariance in the y-direction, only TM-polarized light (**H** perpendicular to the $x - z$ plane) will excite surface plasmons, and the simulations in this section are restricted to this case. For brevity, the explicit dependence on ω will be dropped from now on.

As we have assumed that $|U_1^{(inc)}|^2 = |U_2^{(inc)}|^2$, the fringe contrast or visibility $\mathcal{V}$ in the far-field is also equal to $|\mu_{12}|$. The fringe contrast $\mathcal{V}$ is defined as

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{5.8}$$
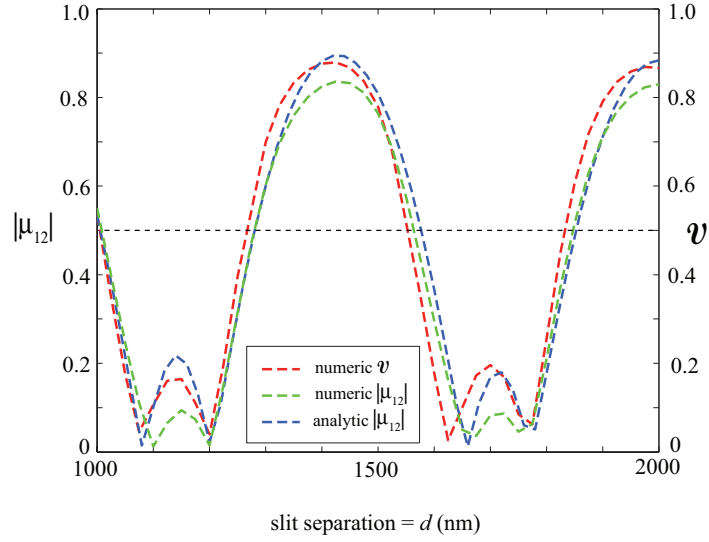
Figure 5.3: Simulation results of $|\mu_{12}|$ and $\mathcal{V}$, with spectral degree of coherence of the incident field $\mu_0 = 0.5$. The analytic result is based on the model described in Eq. (5.9), with the light-plasmon coupling constant $\beta = 0.33\,e^{i\pi}$. The black, dotted, horizontal line indicates the value of $|\mu_0|$.

where $I_{\min}$ and $I_{\max}$ are the minimum and maximum intensity of the fringes. As an example to illustrate the modulation of the coherence, simulations for $|\mu_{12}|$ and $\mathcal{V}$ with partially coherent light ($\mu_0 = 0.5$) illuminating the two slits are shown in Fig. 5.3. The slit width w is taken to be 200 nm. From Fig. 5.3, it is observed that the spatial coherence can indeed be either enhanced or suppressed with respect to the degree of coherence of the incident field. To better understand the modulation, we have derived an analytic expression for μ_{12} using the following model for the fields at each of the two
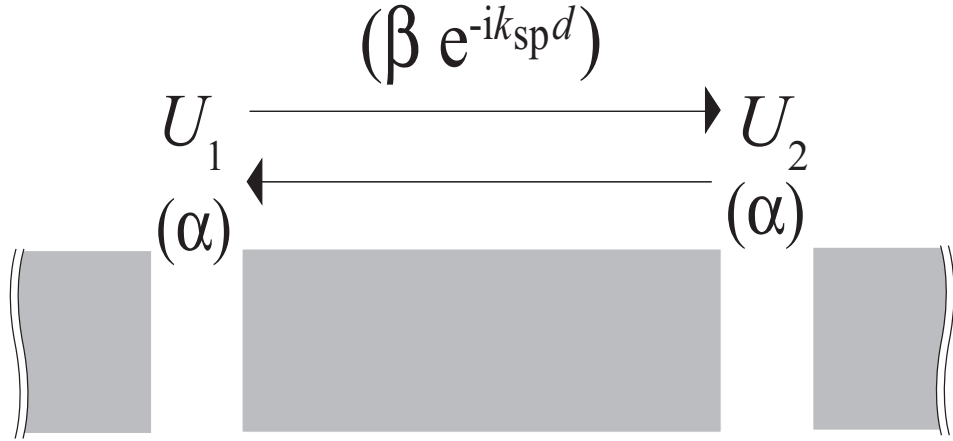
Figure 5.4: Illustrating the plasmonic contributions to the field radiated at each of the slits considered in the model of Eq. (5.9).

slits

$$U_1 = \alpha U_1^{(\text{inc})} + \alpha\beta U_2^{(\text{inc})} e^{-\mathrm{i}k_{\text{sp}}d}, \tag{5.9a}$$

$$U_2 = \alpha U_2^{(\text{inc})} + \alpha\beta U_1^{(\text{inc})} e^{-\mathrm{i}k_{\text{sp}}d}, \tag{5.9b}$$

where $k_{\text{sp}} = k'_{\text{sp}} - \mathrm{i}k''_{\text{sp}}$, is the wavenumber associated with the surface plasmons. The model assumes that a fraction α of the incident field is directly transmitted through the slits, and a fraction $\alpha\beta$ is converted into surface plasmons which travel to the other slit where they re-radiate as a propagating field, as illustrated in Fig. 5.4. The parameters α and β are complex quantities in general.

On substituting from Eq. (5.9) into Eq. (5.6), we find

$$W(\mathbf{r}_1, \mathbf{r}_2) = |\alpha|^2 S^{(\mathrm{inc})}\big\{\, \mu_0^{(\mathrm{inc})} + |\beta|^2 \mu_0^{(\mathrm{inc})*} \exp(-2k_{\mathrm{sp}}''d)$$

$$+ 2\mathrm{Re}\left[\beta \exp(-\mathrm{i}k_{\mathrm{sp}}d)\right]\big\}, \tag{5.10}$$

where

$$S^{(\mathrm{inc})} = S_1^{(\mathrm{inc})} = S_2^{(\mathrm{inc})}, \tag{5.11}$$

and we have have made use of the fact that $\mu_{12}^{(\mathrm{inc})} = \mu_{21}^{(\mathrm{inc})*}$. Here Re denotes the real part. It is to be noted that Eq. (5.10) has the form of an interference law in the frequency domain. But whereas the classical *spectral interference law* (Mandel and Wolf 1995, Sec. 4.3.2) pertains to the spectral density at a single point, Eq. (5.10) pertains to the cross-spectral density of the fields radiated by two slits. It suggests that *surface plasmons propagating from one slit to the other can modulate the state of spatial coherence.* In other words, the cross-spectral density function of the fields emanating from the two slits can be increased or decreased, according to whether at each slit there is constructive or destructive interference between the directly transmitted field and the field which is due to plasmon generation at the the other slit.

Substituting the fields given by Eq. (5.9) into Eq. (5.5) to solve for μ_{12}, it is found that

$$\mu_{12} = \frac{\mu_0 + |\beta|^2 \mu_0^* \exp(-2k_{\mathrm{sp}}''d) + 2\mathrm{Re}\left[\beta \exp(-\mathrm{i}k_{\mathrm{sp}}d)\right]}{1 + |\beta|^2 \exp(-2k_{\mathrm{sp}}''d) + 2\mathrm{Re}\left[\beta \mu_0 \exp(-\mathrm{i}k_{\mathrm{sp}}d)\right]}. \tag{5.12}$$

This formula demonstrates that the spectral degree of coherence of the field that is radiated by the apertures is not equal to the spectral degree of coherence of the incident field. Because of the presence of the oscillating terms in this equation, the modulus

of the former can either be larger or smaller than that of the latter. In other words, varying the distance that separates the two slits will modulate the degree of coherence between the fields emanating from the slits. That this effect is solely due to the action of surface plasmons is easily verified by setting β, their relative contribution strength, equal to zero. In that case formula (5.12) reduces to $\mu_{12} = \mu_0$, i.e., the spectral degree of coherence of the radiated field is then equal to that of the incident field, the expected result from the classical Young's interferometer. While it is recognized that contributions from so-called "creeping waves" can produce similar behaviors as the surface plasmons in very closely spaced apertures, their effects are expected to be negligible at the frequencies and slit spacings considered here (Lalanne and Hugonin 2006).

It is helpful also to consider the effect of the plasmons on the state of coherence in the limiting cases $|\mu_0| = 1$ and $\mu_0 = 0$. For $|\mu_0| = 1$, the numerator and denominator of Eq. (5.12) are equal and we find that $\mu_{12} = |\mu_0|$. In other words, if the incident field illuminating the slits is fully coherent, the plasmons will not modify the state of coherence. For $\mu_0 = 0$, i.e. an incident incoherent field, Eq. (5.12) reduces to

$$\mu_{12} = \frac{2\mathrm{Re}\left[\beta \exp(-ik_{\mathrm{sp}}d)\right]}{1 + |\beta|^2 \exp(-2k_{\mathrm{sp}}''d)}. \tag{5.13}$$

It is to be noted that this formula suggests that not only can the spatial coherence of the output field be greater than that of the input field, it may also switch signs, resulting in the field at the two slits being anti-correlated. Setting the plasmon decay constant k_{sp}'' to zero for the moment, the maximum value of $|\mu_{12}|$ is approximately given by

$$|\mu_{12}^{(\mathrm{max})}| = \frac{2|\beta|}{1 + |\beta|^2} < 1. \tag{5.14}$$

The spatial coherence of the output field increases with increasing value of $|\beta|$ in this

case. Also, $|\mu_{12}|$ never exceeds unity, as required by the range of values allowed for the spectral degree of coherence as given in Eq. (5.7).

Furthermore, by plotting μ_{12} with $|\beta| = 0.33$ and $\arg(\beta) = 180°$ in formula (5.12) against the numerical results in Fig. 5.3, we find excellent agreement between the two methods, reinforcing the evidence of the coherence converting ability of the surface plasmons. It is to be noted that the parameter β describing the light-plasmon coupling strength, is used to fit the analytic expression in the formula (5.12) to the rigorous numeric simulations.

To see how the coupling constant β changes with the slit width, we performed the simulations for slit widths $w = 100$ nm and $w = 50$ nm. The results are shown in Fig. 5.5. When the slit width is 100 nm, the phase shift $\arg(\beta) = 205°$, and $|\beta| \sim 0.33$. When the slit width $w = 50$ nm, the phase shift $\arg(\beta) = 225°$, and $|\beta| \sim 0.21$. From the plots, the maximum value of $|\mu_{12}|$ is approximately 0.60 and 0.40, respectively, in agreement with the prediction of Eq. (5.14). These results also suggest that the argument of β varies non-trivially with the width of the slits (cf. (Lalanne et al. 2006)).

Before proceeding, we offer an explanation of the coherence converting ability of the surface plasmons by drawing an analogy with the van Cittert-Zernike theorem when it is applied to the special case of a fully incoherent source (Goodman 2000, Sec. 5.6). In this special case, the spectral degree of coherence in the far-field is proportional to the Fourier transform of the intensity distribution across the source, and it becomes straightforward to see that the fields produced by an incoherent source can become partially coherent simply through the process of propagation.

Suppose there is a pair of fully incoherent point sources Q_1 and Q_2, i.e., $\mu_0 = 0$.
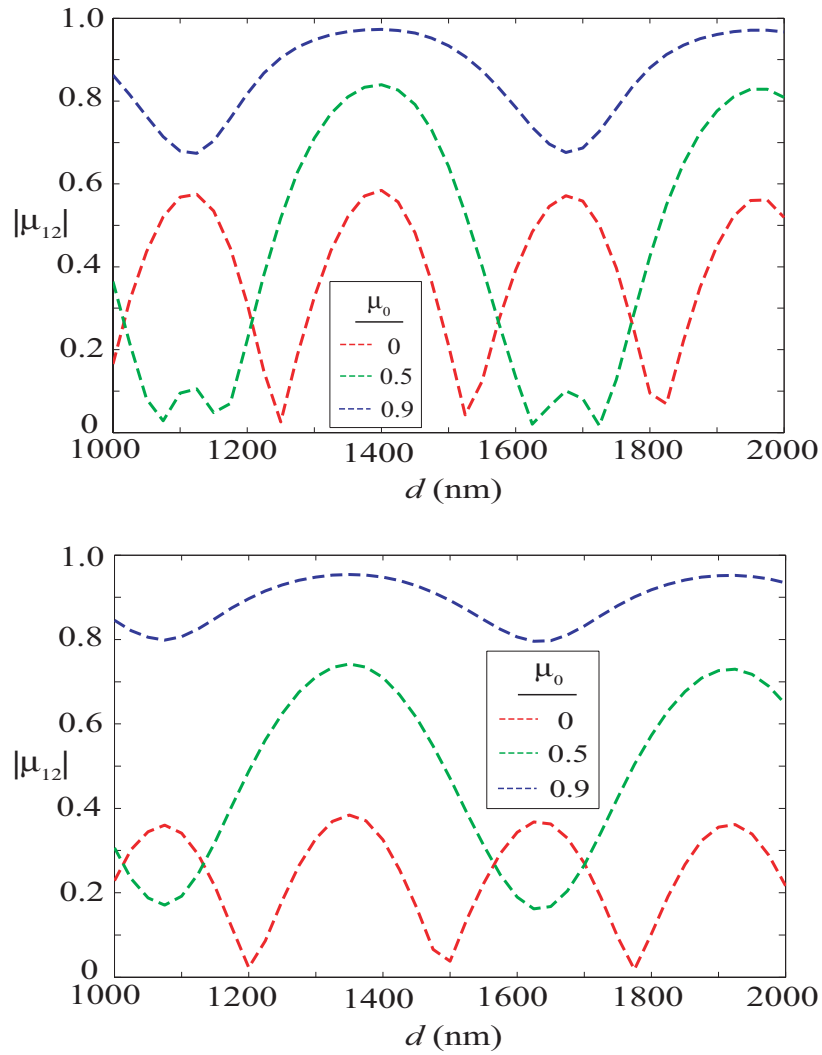
Figure 5.5: The absolute value of the spectral degree of coherence $|\mu_{12}|$ as a function of μ_0, for slit widths: (a) $w = 100$ nm, (b) $w = 50$ nm. For the sake of clarity, only numeric results are shown.

According to the van Cittert-Zernike theorem, the field at two observation points P_1 and P_2 in the far zone, being a superposition of the fields from Q_1 and Q_2, are in general partially correlated. This fact can be appreciated by denoting the wave trains from Q_1 to the points P_1 and P_2 be A_1 and A_2 respectively, and that from Q_2 to the points P_1
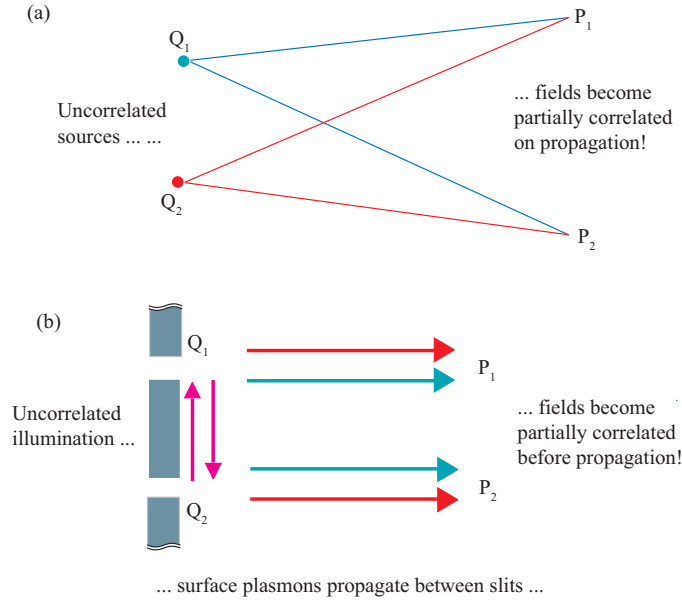
Figure 5.6: An analogy with the van Cittert-Zernike theorem to illustrate the coherence converting ability of the surface plasmons. The situation (a) consists of a pair of fully incoherent point sources Q_1 and Q_2, while the situation (b) consists of a pair of slits in a metal plate illuminated with a fully incoherent field. In the latter situation, surface plasmons excited at each of the slits can propagate along the metal surface.

and P_2 be B_1 and B_2 respectively (Wolf 1978). Since Q_1 and Q_2 are incoherent, the waves they emit will be uncorrelated, i.e.,

$$\langle A_i\, Bj \rangle = 0 \qquad (i, j = 1, 2).$$

(5.15)

If the points P_1 and P_2 are sufficiently far from Q_1 and Q_2, then

$$A_1 \approx A_2\,,$$

(5.16a)

$$B_1 \approx B_2\,.$$

(5.16b)

Therfore, the fields at P_1 and P_2 will be correlated with each other upon superposition of the waves, even though their contributions are uncorrelated individually, according

to Eq. (5.16). Clearly, the change in coherence is a result of superposition in the process of propagation. This situation is illustrated in Fig. 5.6a.

Next let us illuminate a pair of slits Q_1 and Q_2 with fully incoherent light. Suppose the light is allowed to couple with surface plasmons, which can propagate between the slits, as in the case of Fig. 5.6b. The coherence of the fields emanating from the slits is then dependent on the distance between Q_1 and Q_2, as we have demonstrated in the above analysis. The field at the two points P_1 and P_2 are still partially correlated in general, however the change in coherence is now due to the surface plasmons propagating between the slits instead of the propagation of the radiating fields themselves. Thus, one may think of the surface plasmons as performing an 'in-plane' van Cittert-Zernike effect. This analogy demonstrating the coherence converting ability of the surface plasmons is illustrated in Fig. 5.6.

5.1.2 Three-slit Interferometer

It is not clear if the results for the two slits in the previous section apply as well to an array - in the near field region, the optical and plasmonic fields scattered from each of the slits can interact with that scattered from another. As such, their interactions in an array of slits cannot be described accurately through a straightforward superposition of the response from each individual slit. As an intermediate approach to this problem, we have analyzed a three-slit interferometer (Fig. 5.7) to investigate the effects on the modulation of the coherence when an additional slit A is placed between the original pair.

In this case, the spectral degree of coherence between any two fields U_i and U_j, with
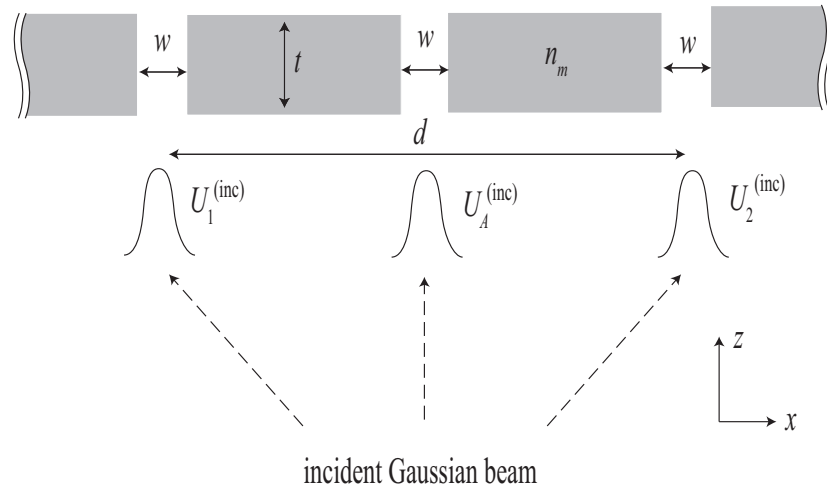
Figure 5.7: Illustrating the geometry for the three-slit interferometer.

$i, j = 1, 2, A$ is

$$\mu_{ij} = \frac{W_{ij}}{\sqrt{S_i \, S_j}}, \tag{5.17}$$

where

$$W_{ij} = \langle U_i^* \, U_j \rangle \,, \tag{5.18}$$

is the cross spectral density, and $S_i = W_{ii}$ is the spectral intensity of the field. As before, the spectral degree of coherence between the incident fields at any of the two slits are denoted as $\mu_{ij}^{(inc)} = \mu_0$.

We show the simulation results of μ_{12} and μ_{1A} for various values of μ_0 as a function of the slits' separation for slit width $w = 200$ nm in Fig. 5.8. It is observed that, similar to what was found in the two-slit case, the degree of coherence μ_{12} and μ_{1A} can be either enhanced or suppressed with respect to the degree of coherence of the incident field μ_0.
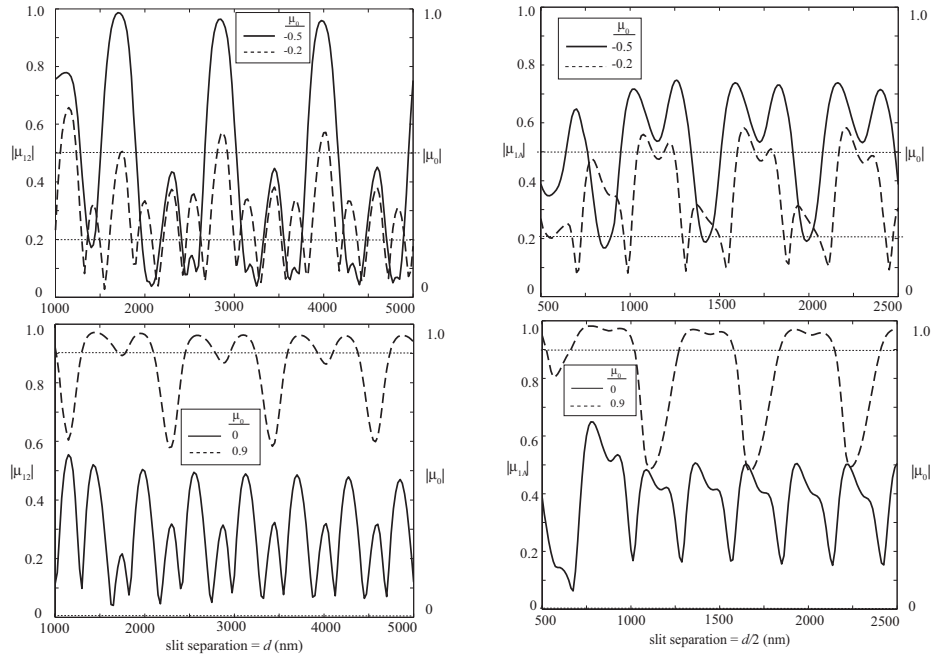
Figure 5.8: Simulation results for μ_{12} (left column) and μ_{1A} (right column) for $w = 200$ nm, with different values of spectral degree of coherence of the incident field μ_0. Top row: μ_0 = -0.5 (solid line), and -0.2 (dashed line). Bottom row: μ_0 = 0 (solid line), and 0.9 (dashed line). Dotted horizontal lines indicate the respective values of $|\mu_0|$.

Furthermore, the results suggest that μ_{12} and μ_{1A} may be periodically modulated as a function of the slits' separation. Since $\mu_{2A} = \mu_{1A}$ by symmetry, simulation results for μ_{2A} will not be shown.

To better relate the effects of the surface plasmons to the numerical results, we extend the model used to calculate the fields for the two-slit case (cf. Eq. (5.9)) to include slit A in the center. At each of the slits, a fraction α of the incident field will be directly transmitted. Part of the incident field will be converted into surface plasmons, which can travel to another slit, whereby the following two situations may take place: (1) part of the plasmonic field couples back to light and reappears as a freely propagating field; and/or (2) part of the plasmonic field is scattered, after which
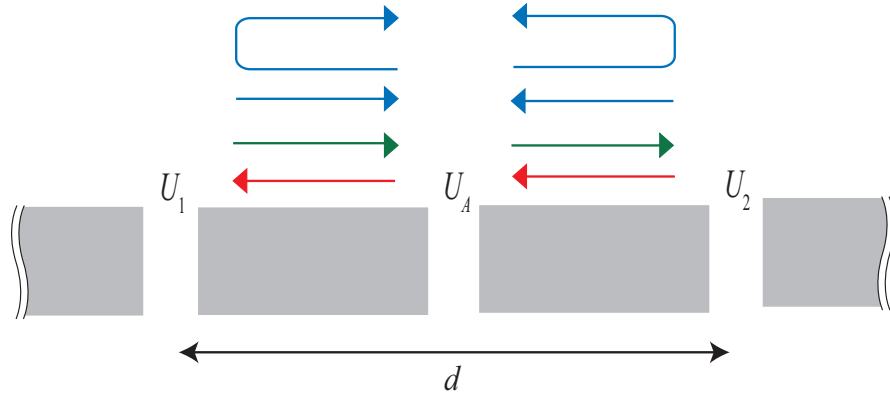
Figure 5.9: Illustrating the plasmonic contributions to the field radiated at each of the slits considered in the model of Eq. (5.19). The red, green and blue colored arrows represent the contributions to the fields U_1, U_2 and U_A respectively.

it continues propagating to the next slit where it couples back to light and reappears as a freely propagating field. These interactions are indicated with the arrows in Fig. 5.9, with the solid, dashed and dotted arrows representing the contributions to the fields U_1, U_2 and U_A respectively.

Let us also define, in addition to the light-plasmon coupling constant β, parameters η and Γ, which describe the forward and backward scattering of the plasmonic field by the slits respectively. Similar to α and β, η and Γ are complex quantities in general. While α and β are related to three-dimensional propagating fields, η and Γ describes the transmission and reflection of the plasmonic field propagating along the surface of the metal plate. In terms of these parameters and the incident fields, the field at each

of the three slits is

$$U_1 = \alpha U_1^{(\text{inc})} + \alpha\beta\eta U_2^{(\text{inc})} e^{-ik_{\text{sp}}d} + \alpha\beta U_A^{(\text{inc})} e^{-ik_{\text{sp}}d/2}, \tag{5.19a}$$

$$U_2 = \alpha\beta\eta U_1^{(\text{inc})} e^{-ik_{\text{sp}}d} + \alpha U_2^{(\text{inc})} + \alpha\beta U_A^{(\text{inc})} e^{-ik_{\text{sp}}d/2}, \tag{5.19b}$$

$$U_A = \alpha\beta U_1^{(\text{inc})} e^{-ik_{\text{sp}}d/2} + \alpha\beta U_2^{(\text{inc})} e^{-ik_{\text{sp}}d/2} + \alpha U_A^{(\text{inc})} [1 + 2\beta\Gamma e^{-ik_{\text{sp}}d}]. \tag{5.19c}$$

The field U_1 at slit 1 in Eq. (5.19a), for instance, consists of three parts: (i) a directly

transmitted part (α); (ii) a part that couples to the plasmons at slit 2 (β), 'jumps'

over the center slit A (η), and coupled to the output (α); (iii) a part that couples to

the plasmons at slit A (β), and coupled to the output (α). It is to be noted that the

contribution from surface plasmons that travel from slit $1(2)$ to slit A, and then are

reflected back to slit $1(2)$ can be expressed as a direct contribution from $U_1^{(\text{inc})}(U_2^{(\text{inc})})$,

and may be taken into account by the parameter α in our model.

As expressed in Eq. (5.19), each of the fields U_i or U_j in Eq. (5.18) is a function of

the incident fields at the three slits. Using a 3×3 matrix $\mathbf{M}$ to describe the degree of

coherence between the incident fields, we can write Eq. (5.18) in matrix form such that

$$W_{ij} = U^{(i)\dagger} M U^{(j)}, \tag{5.20}$$

with

$$U^{(i)} = \begin{bmatrix} U_{i1} \\ U_{i2} \\ U_{iA} \end{bmatrix}, \tag{5.21}$$

where the superscripted $\dagger$ indicates adjoint. The components of the vector $U^{(i)}$, namely

U_{i1}, U_{i2}, U_{iA} denotes respectively the contribution from the incident fields $U_1^{(\text{inc})}, U_2^{(\text{inc})}$
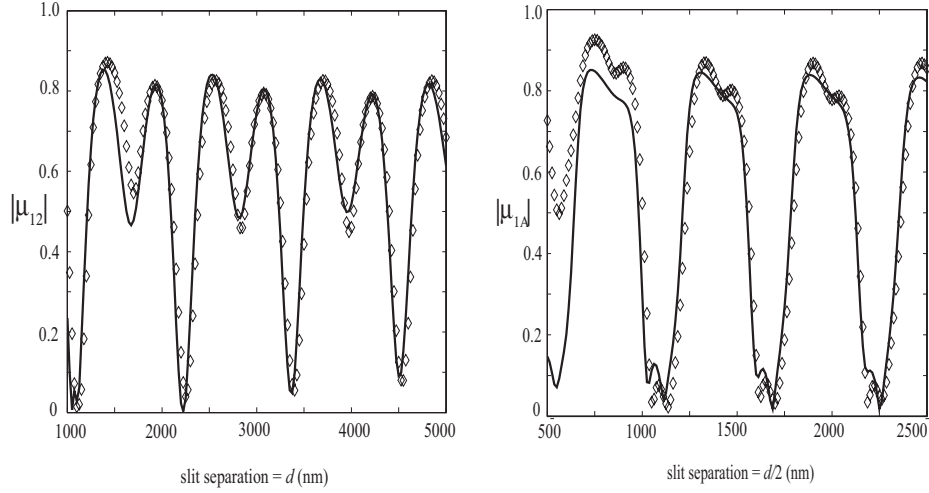
Figure 5.10: Simulation results for μ_{12} (left column) and μ_{1A} (right column) for $w =$ 100 nm, with spectral degree of coherence of the incident field $\mu_0 = 0.5$. Solid line represents analytic results, and diamond markers ($\Diamond$) represent numeric results.

and $U_A^{(\text{inc})}$ to the field U_i as expressed in Eq. (5.19). The matrix $\mathbf{M}$ takes on the form

$$
\mathbf{M} = \begin{bmatrix} 1 & \mu_{12}^{(inc)} & \mu_{1A}^{(inc)} \\ \mu_{21}^{(inc)} & 1 & \mu_{2A}^{(inc)} \\ \mu_{A1}^{(inc)} & \mu_{A2}^{(inc)} & 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & \mu_0 & \mu_0 \\ \mu_0^* & 1 & \mu_0 \\ \mu_0^* & \mu_0^* & 1 \end{bmatrix} \tag{5.22}
$$

where we have made use of the fact that $\mu_{ij}^{(inc)} = \mu_{ji}^{(inc)*}$. Using Eq. (5.5), the degree

of coherence between any two of the three slits may then be calculated. Fig. 5.10

illustrates the agreement between the two sets of results, using the case $w = 100$ nm,

and $\mu_0 = 0.5$ as an example. Therefore, the model we have developed could be used to

analyze the modulation of the degree of coherence as a result of the surface plasmons

propagating between the slits. It should be emphasized that the parameters β, η, and Γ

remain fixed for a given plate index n_m, plate thickness t, and slit width w, independent

of the slits' separation or the choice of the two slits for which the degree of coherence

is to be calculated.

Table 5.1: Choice of values for the parameters β, η, and Γ for different slit widths
(expressed in nm).

slit width (w)	$|\beta|$	$arg(\beta)$	$|\eta|$	$arg(\eta)$	$|\Gamma|$	$arg(\Gamma)$
200	0.33	180°	0.82	0.0°	0.57	230°
100	0.33	190°	0.80	0.0°	0.60	240°
50	0.24	190°	0.78	0.0°	0.60	240°

The analytic results are fitted to the numerical ones for slit width $w = 200, 100$

and 50 nm. The appropriate choice of the values for β, η, and Γ are shown in Ta-

ble 5.1. It is worthwhile noting that the values for $|\beta|$, which describes the strength

of the light-plasmon coupling, are comparable to those used in the case of the Young's

interferometer. In that case, β was found to be 0.33 for slit widths 200 nm and 100 nm,

and 0.21 for slit width 50 nm. The values of η and Γ did not vary appreciably as

the slit width is decreased from 200 nm to 50 nm. This provides support that η and

Γ may be regarded as transmission and reflection coefficients, which should depend

largely on the refractive index n_m of the metal at each metal/air boundary (assuming

the the slits are far away enough from each other) rather than on the slit width. It is

also found that the choice of a purely real, positive η provides a good fit for both μ_{12}
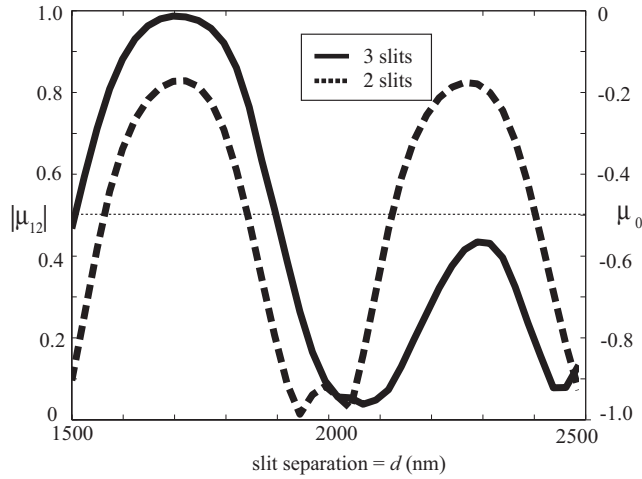
Figure 5.11: Comparison of the modulation of the degree of coherence μ_{12} in the presence and absence of the additional slit A, with degree of coherence of the incident field $\mu_0 = -0.5$ (dotted line). The slit width w is taken to be 200 nm.

and μ_{1A}, which can exhibit quite different behaviors as seen in Figs. 2 and 3. Such a real, positive η resembles the forward scattering efficiency in electromagnetic scattering problems (Bohren and Huffman 1983, Sec. 3.4). Thus we could think of η as describing how strongly the slit scatters the plasmonic field in the forward direction.

Finally, we compare the modulation of μ_{12} in the absence and presence of the additional center slit. The results are shown in Fig. 5.11. It can be seen that for a given μ_0, the range of μ_{12} can be larger or smaller in the case of the three-slit geometry. Intuitively, one might expect the additional slit to serve as a barrier for the surface plasmon interactions between the end slits, thus acting to suppress the modulation of μ_{12}, which is indeed the case depicted by the results for slit separation d between $\sim 2000 - 2500$ nm

in Fig. 5.11, where it is taken that $\mu_0 = -0.5$. However, the plots also show that for d between $\sim 1500 - 2000$ nm, the modulation of the degree of coherence is greater in the three-slit geometry than the two-slit geometry. Thus, the additional 'barrier' slit can serve not just to decrease the effects of surface plasmons propagating from one slit to the other, but also to preserve, and even enhance these effects. From these observations, it can be inferred that the additional center slit can either enhance or suppress the modulation of the coherence properties of the fields emanating from the two end slits. This is promising for the development of coherence converting devices with hole arrays in metal plates.

5.1.3 Symmetric Three-hole Interferometer in an Equilateral Geometry

We have demonstrated the coherence converting ability of surface plasmons coupled by subwavelength apertures on a metallic plate. However, we have been restricted to examining the spatial coherence between any two points of the field, i.e., μ_{ij}. For an array of apertures, we would have to study the coherence properties of light as a global effect resulting from contributions from each aperture in the array. To address this issue, we will now consider a symmetric three-hole interferometer in an equilateral geometry as shown in Fig. 5.12. The holes, which are subwavelength in size, are illuminated with plane waves of wavelength 600 nm, and the metal plate is taken to be gold, as before. Due to the symmetry in the system, we will investigate the relationship between the holes' separation and a 'global' degree of coherence of the fields at the three holes. It is shown that this 'global' spatial coherence directly influences the fringe contrast of the radiation pattern.

Instead of exact three-dimensional simulations that are complex and time consuming, we will again attempt to study the problem using simple analytical modeling methods, justified by success of the analytical models in the earlier sections. Let us assume that the spectral densities of the fields incident at each of the slits are identical and is taken to be U_0. A part of the incident field is directly transmitted (α), while a part couples to surface plasmons and contributes to a plasmonic field, which is multiply scattered by the three holes. To take into account the multiple scattering effects, we employ the Foldy-Lax equations (Foldy 1945) to calculate the field at each hole

$$U_F(\boldsymbol{\rho}_n) = \alpha U_0(\boldsymbol{\rho}_n) + U_S(\boldsymbol{\rho}_n), \tag{5.23}$$

where $\boldsymbol{\rho}_n = \sqrt{X_n^2 + Y_n^2}$ describes the position of the nth hole ($n = 1, 2, 3$) in the aperture plane, $U_F(\boldsymbol{\rho}_n)$ is the total plasmonic field at each hole, and $U_S(\boldsymbol{\rho}_n)$ is the contribution due to the multiple scattering of the plasmonic field. At the nth hole, the scattered field $U_S(\boldsymbol{\rho}_n)$ may be expressed as

$$U_S(\boldsymbol{\rho}_n) = \chi \sum_{m=1, m \neq n}^{3} G(\boldsymbol{\rho}_n, \boldsymbol{\rho}_m) U_F(\boldsymbol{\rho}_m) \tag{5.24}$$

where χ describes the scattering strength of each hole, and

$$G(\boldsymbol{\rho}_n, \boldsymbol{\rho}_m) = \frac{\mathrm{i}}{4} H_0^{(1)} \left(k_{sp} |\boldsymbol{\rho}_n - \boldsymbol{\rho}_m| \right) \tag{5.25}$$

with H being the Hankel function (Arfken 1970, Sec. 16.6). Here we have used the two-dimensional Green function for the propagation of the plasmonic field on the surface of the array, since its amplitude decays exponentially into either of the neighboring media.

A few words on the simple multiple scattering model based on Eqs.(5.23 − 5.25) before we proceed. Similar scalar, multiple scattering approach has been used by others
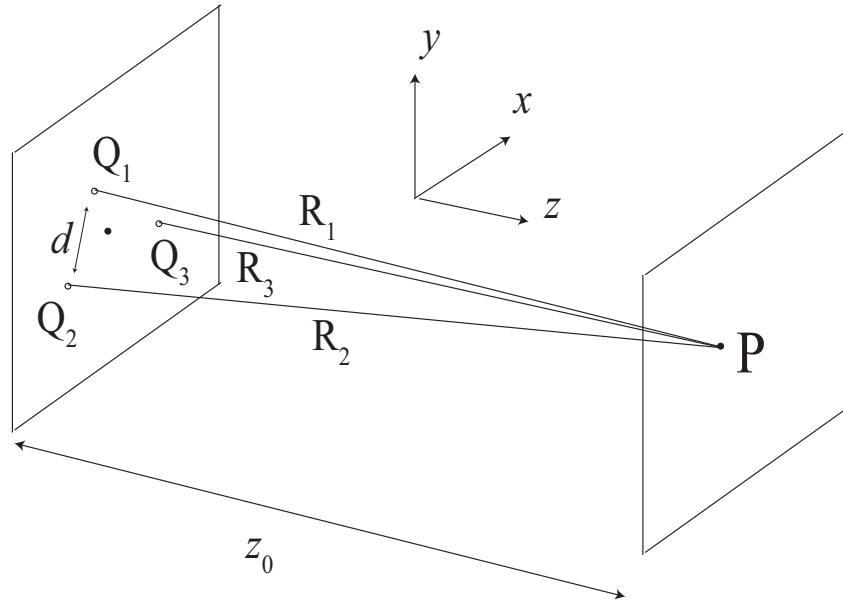
Figure 5.12: The geometry for the three-pinhole interferometer.

to investigate the multiple scattering behavior of surface plasmons by periodic arrays of surface nanoparticles at a metal-dielectric interface. It is worthwhile to note that the models have been reported to have attained some success in reproducing experimental results (Bozhevolnyi and Coello 1998, Bozhevolnyi and Volkov 2001, Søndergaard and Bozhevolnyi 2003). While it is recognized that such a scalar model will not fully describe the effects of the surface plasmons (Visser 2006), its simplicity offers much computing convenience, and can potentially provide insights to a problem of complicated nature like this. Furthermore, results from Sections 5.1.1 and 5.1.2 suggest that such simplified models would work well in providing a qualitative analysis of the problem. In those sections, analytic models that consider only the dominant field component produce results in agreement with rigorous numerical simulations.

Substituting Eqs. (5.24) and (5.25) into Eq. (5.23), the field $U_F(\boldsymbol{\rho}_n)$ at each of

the holes can be determined. For a symmetric structure like this with three pinholes, $U_F(\boldsymbol{\rho}_n)$ in Eq. (5.23) can be calculated by analytically inverting a 3×3 matrix. The spectral degree of coherence μ_{ij} between any two of the three holes may then be found using Eq. (5.5). Due to the symmetry, it is convenient to define μ_F and γ such that

$$\mu_F = \mu_{12} = \mu_{23}\,, \tag{5.26}$$

and

$$\gamma = \chi\, G(\mathbf{r}_n, \mathbf{r}_m)\,. \tag{5.27}$$

In terms of γ, the spectral degree of coherence μ_F for this case of the symmetric three-hole interferometer is

$$\mu_F = \frac{\mu_0|1-\gamma|^2 + (1+3\mu_0)|\gamma|^2 + 2(1+\mu_0)\mathrm{Re}[(1-\gamma)^*\gamma]}{|1-\gamma|^2 + 2(1+\mu_0)|\gamma|^2 + 4\mu_0\mathrm{Re}[(1-\gamma)^*\gamma]}. \tag{5.28}$$

Similar to Eq. (5.12), we find that μ_F is in general not equal to μ_0, as can be seen from the plots in Fig. 5.13. Here it is assumed that the illuminating field is fully incoherent ($\mu_0 = 0$), and $arg(\chi)$, which results in a translational shift of μ_F, is taken to be zero. It can be seen that the amount of modulation of the coherence increases with $|\chi|$, the scattering strength of each hole. It is to be noted that the scattering coefficient χ applies to the two-dimensional plasmonic field U_F and can be greater than unity (Jones 1986, Bohren and Huffman 1983), unlike β, which applies to the radiating field and whose magnitude is restricted to values less than unity to satisfy energy conservation laws. Clearly, the appropriate value of $|\chi|$ determines the precise amount of modulation offered by the array, and this will be investigated as part of future work. In the present case, we will focus on demonstrating the feasibility of a coherence converting device with a suitable array of subwavelength apertures.
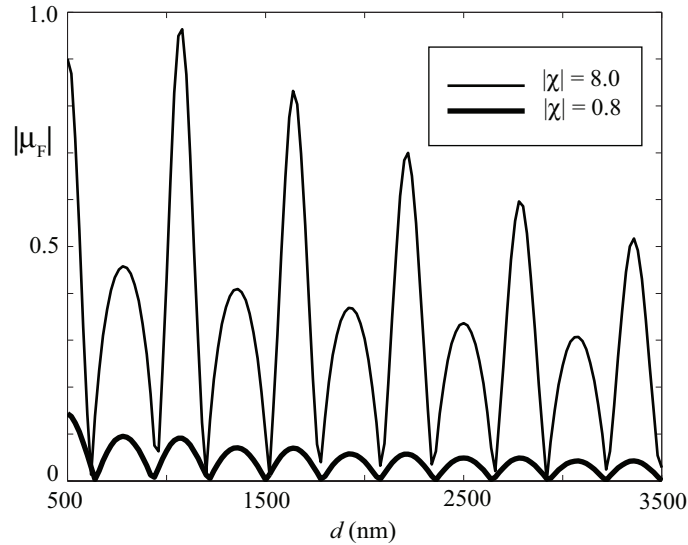
Figure 5.13: Modulation of the degree of coherence in the three-pinhole interferometer as a function of d for $|\chi| = 8.0$ (thin line) and $|\chi| = 0.8$ (thick line), with $\mu_0 = 0$ and $arg(\chi) = 0$.

For this case of the symmetric three-hole interferometer, unlike the three-slit interferometer, each hole separation d is associated with a 'global' value of μ_F. Since highly coherent fields can interfere to form interference fringes with a high fringe contrast, we will calculate the radiation pattern from the three-hole interferometer to see if there is any qualitative agreement between the value of μ_F and the fringe contrast of the radiation pattern. For the case $\mu_F = 0$, one would expect to see no fringes in the radiation pattern, i.e., $\mathcal{V} = 0$ in Eq. (5.8).

Due to the evanescent nature of the surface plasmons, it is reasonable to assume that only the field at each of the holes contributes to the radiation pattern, which we

(a)

(b)

(c)

(d)
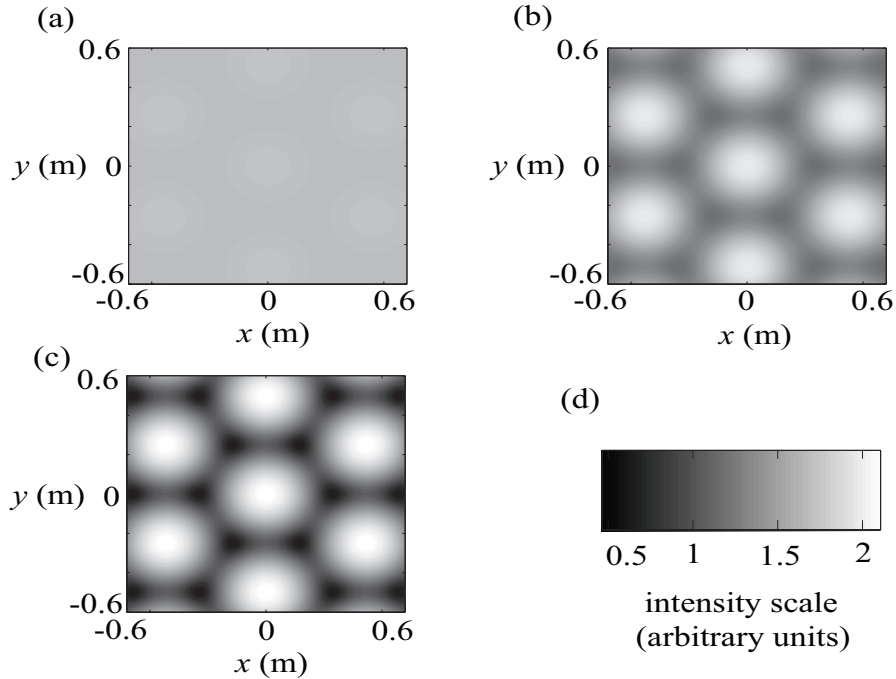
Figure 5.14: Radiation pattern at $z_0 = 2$ m with $\mu_0 = 0$ and $\chi = 8.0$, for hole separation d of (a) 2660 nm, (b) 2720 nm, and (c) 2780 nm. The intensity scale for the three plots is shown in (d). For the three cases, the degree of coherence $\mu_F = 0.02, 0.36$, and 0.60, and the fringe contrasts are $0.03, 0.46$, and 0.67, respectively.

may readily evaluate with the Fresnel diffraction integral (Goodman 2005, Sec. 4.2)

$$U^{(\infty)}(x, y; z_0) = \frac{k_0 \exp\left[\mathrm{i}\, k_0\, z_0\right]}{\mathrm{i}\, 2\pi\, z_0} \quad \times$$

$$\iint U(X, Y) \exp\left[\mathrm{i}\frac{k_0}{2z_0}\{(x - X)^2 + (y - Y)^2\}\right] dX\, dY. \qquad (5.29)$$

where $U^{(\infty)}(x, y; z_0)$ denotes the field on a plane $z = z_0$ in the far zone, and it has been taken that the source plane is located at $z = 0$. With $\mu_0 = 0$ and $\chi = 8.0$, the radiation patterns for $d = 2660$ nm, 2720 nm, and 2780 nm are calculated and shown in Fig. 5.14. The values of μ_F from Fig. 5.13 at these values of d are 0.02, 0.36, and 0.60, respectively. The distance between the metallic plate and the observation plane is taken to be $z_0 = 2$ m. It is found from Fig. 5.14 that the fringe contrast for the

three cases are 0.03, 0.46, and 0.67, given in ascending order of μ_F. Thus the fringe contrast increases with μ_F, and the dependence of the spatial coherence on the distance d between the pinholes is verified.

5.2 Summary and Future Work

We have shown using analytic and numerical modeling that, depending on the separation between subwavelength apertures on a metal plate, surface plasmons propagating between them can increase or decrease the spectral degree of coherence of the emanated fields. Simulations of the three slit-interferometer show that the surface plasmons can either increase or decrease the degree of modulation of the coherence in the presence of additonal apertures. These results suggest that one may develop coherence converting devices with suitable arrays of subwavelength apertures in metal plates. Such coherence converting devices may be very useful in optical systems since the coherence properties of a field strongly affect its propagation characteristics.

Presently, we are using the Foldy-Lax equations presented in the previous section to study the feasibility of a practical coherence converting device by means of a suitable array of subwavelength holes in a metal plate. For a practical choice of the field illuminating the apertures, we adopt a Schell-model to describe the coherence properties of the incident field. The cross spectral density function between two points $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ of a planar, secondary Schell-model source takes on the form (Wolf 2007, Sec 5.3)

$$
\begin{aligned}
W_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) &= \sqrt{S_0(\boldsymbol{\rho}_1)} \sqrt{S_0(\boldsymbol{\rho}_2)} \, \mu_0(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1) \\
&= \sqrt{S_0(\boldsymbol{\rho}_1)} \sqrt{S_0(\boldsymbol{\rho}_2)} \exp[-|\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1|^2/2\delta^2] \,,
\end{aligned}
\tag{5.30}
$$

where δ is a measure of the effective correlation length. Schell-model sources are representative of many real physical sources, and are characterized by the property that their degree of spatial coherence at any two source points $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$, depends only on the difference $\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1$. We hope to assess the performance of such a device by studying the degree of coherence of the fields and the fringe contrast they produce.

CHAPTER 6: EXTRAORDINARY OPTICAL TRANSMISSION IN SYSTEMS OF MULTI-LAYERED METALLIC THIN FILMS

Extraordinary optical transmission (hereafter referred to as EOT) through corrugated thin metal films and subwavelength-aperture arrays in metal plates has evoked great interest since its first experimental demonstration by Ebbesen et al. (1998) . Although there are still continuing discussions on the physical mechanism in play, the critical role of surface plamsons in EOT has been confirmed through several studies (Ghaemi et al. 1998, Grupp et al. 2000, Barnes et al. 2004, Schouten et al. 2005, Lalanne and Hugonin 2006, Visser 2006).

More recently, EOT with plasmonic arrays arranged in cascade have been investigated (Ye and Zhang 2005, Bai et al. 2005, Chan et al. 2006, Cheng et al. 2007). It was found that the optical transmission through the cascaded layers either exceeded or was comparable to the transmission through a single layer. While most of these recent investiagations focused on two cascaded layers, here we will consider up to four cascaded layers. Proposed applications of such cascaded structures include the improvement of surface enhanced Raman scattering, light confinement and guidance at the nanoscale, SNOM capabilities, and control of the phase delay of the transmitted light (Bakker et al. 2004, Marcet et al. 2008).

In this chapter, we show through rigorous numerical simulations that corrugated metallic films arranged in cascade can potentially impede the exponential decay of elec-

tromagnetic fields in metals, and EOT may be achieved, depending on the thickness of

the structure. The simulations are again based on the Green tensor formalism described

in Chapter 3.

6.1 EOT with Corrugated Metallic Films in Cascade
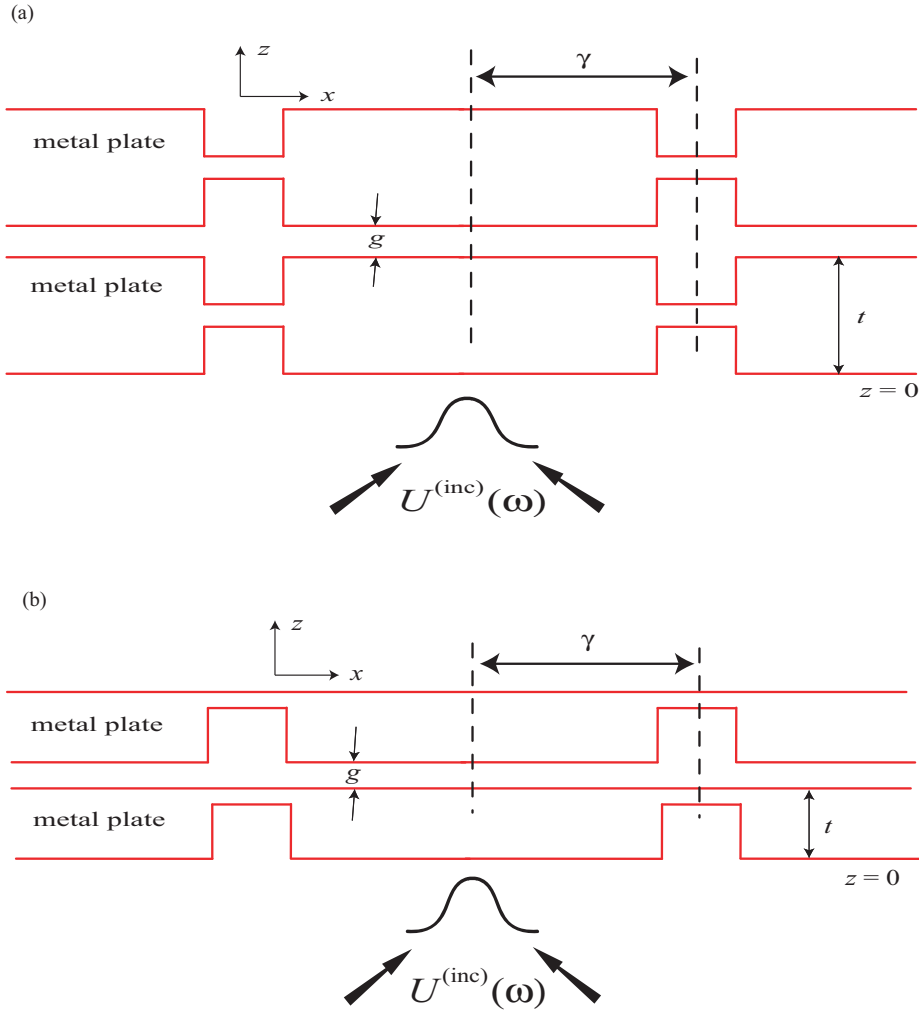
Figure 6.1: Geometry for the multi-layer structure with two layers for (a) plasmon
             pits on both sides of the metal plates, and (b) plasmon pits only on the
             illuminated side of the metal plates. More layers of identical metal plates
             may be cascaded.

The geometry we consider is motivated by our earlier investigations on the plasmon-

assisted super-resolved optical readout effects in Chapter 4. It was found there that the performance of the readout system could be improved or degraded with suitably placed 'plasmon pits' on a metal plate, taken to be silver in that case. Besides enhancement of the coupling between surface plasmons and light, these plasmon pits serve to either increase the transmission and/or confine the electromagnetic fields between them, depending on whether they are on the illuminated or dark side of the metal plate.

Here we consider placing the two pairs of plasmon pits 'back to back' on the two sides of the metal plate. Two or more of these metal plates can then be cascaded to form a multi-layer structure, as shown in Fig. 6.1a. The metal plates have thickness t, and are separated by an air gap g. The plasmon pits are laterally separated by a distance of 2γ. More layers of identical metal plates may be cascaded with the same gap g. As it has been suggested that the grooves on the dark side only contribute weakly to the plasmon-assisted enhanced transmission (García-Vidal et al. 2003, Barnes et al. 2003, Degiron and Ebbesen 2004), we will also investigate structures with plasmon pits only on the illuminated side of the metal plates, as shown in Fig. 6.1b. These structures, which are simpler than those in Fig. 6.1a, can be more easily fabricated with precison, making them more favorable in experiments.

Similar to the case of the transmission readout system discussed in Chapter 4, the optical transmission T is normalized to the incident field such that

$$T = \frac{\int_{-\infty}^{+\infty} S_z dx}{Y_0 \int_{-\infty}^{+\infty} |E^{(inc)}(x,z)|^2 dx},$$

(6.1)

where S_z is the normal component of the time-average Poynting vector emerging from the data layer, $Y_0 = \sqrt{\frac{\epsilon_0}{\mu_0}}$, and $E^{(inc)}(x,z)$ is the electric field amplitude of the incident

Gaussian beam.

For the ensuing discussion, we will take the metal plates to be silver, which typically exhibits low absorption losses in the visible spectrum. It is taken that the illuminating wavelength (λ) is 500 nm, around the middle of the visible spectrum. The beamwidth of the incident Gaussian beam at full width at half maximum is taken to be 530 nm. The refractive index of silver at $\lambda = 500$ nm is taken to be $n_{ag} = 0.05 - \mathrm{i}\,2.87$, following the data of Johnson and Christy (1972).
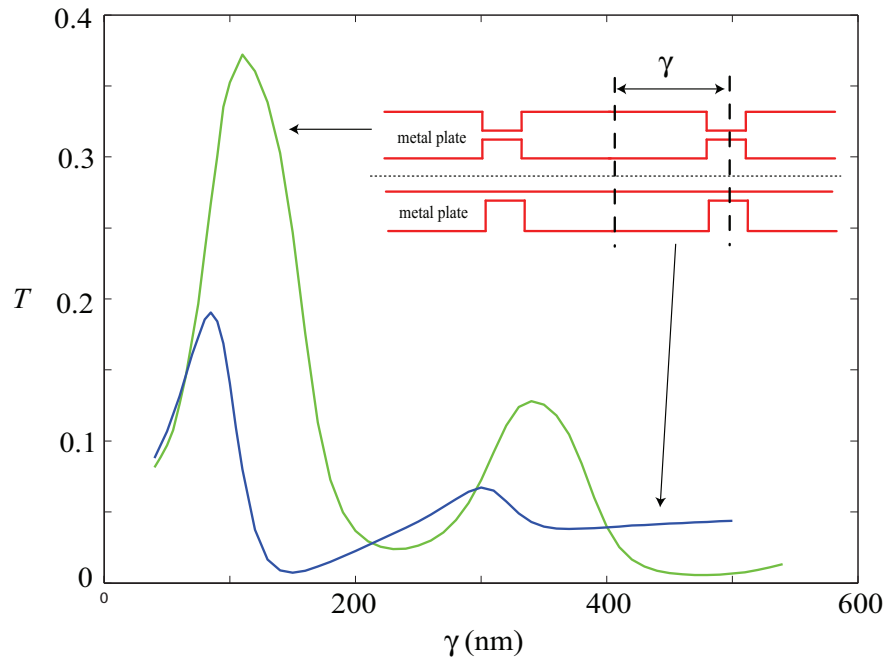
Figure 6.2: Optical transmission T as a function of γ for a single layer of the structures depicted in Fig. 6.1a (green line) and Fig. 6.1b (blue line).

The gap g is kept small in this case so that surface plasmons propagating on surfaces separated by the air gaps can couple resonantly with each other (Ye and Zhang 2005, Martín-Moreno et al. 2001). We take g to be 20 nm and 10 nm for the multi-layer structure in Fig. 6.1a and Fig. 6.1b, respectively. The thickness t of the silver plate is

taken to be 100 nm and 50 nm for the structure in Fig. 6.1a and Fig. 6.1b, respectively. The thickness of the silver plates are chosen to be thin both to decrease absorption losses and to enhance resonant plasmon coupling (Bonod et al. 2003).

It is to be noted that both the separation g and the thickness t for the multi-layer structure in Fig. 6.1b have been taken to be half of that in Fig. 6.1a. As we will see, such a choice would allow us to compare the transmission between structures that have an equal volume of metal but has different number of layers for the two geometries in Fig. 6.1. The value of γ is taken to be 80 nm, as simulations with a single layer of the structures illustrated in Fig. 6.1 show optimal optical transmission for values of γ between 70 and 90 nm for both structures (see Fig. 6.2).

To demonstrate the ability of the multi-layered structure in Fig. 6.1a to impede the exponential decay of the field in a single silver plate, we compare the field distribution for the two cases. Typical results are shown in Fig. 6.3, where it has been taken that $\gamma = 80$ nm. For consistency, the volume of silver is kept equal in both the multi-layered structure and the single slab.

It is clear from Fig. 6.3 that while the field in the single slab decays very quickly as it penetrates through the material (skin depth at $\lambda = 500$ nm is ~ 15 nm), the field decay in the multi-layered structure is relatively insignificant, even in the layer furthest from the incident field. This is true even when the number of layers is increased to four. Moreover, the transmission for the multi-layer structure is much greater than that through the single silver slab, with $T = 0.185$ and $T = 0.096$ for two and four layers respectively. The corresponding values of transmission for the single silver slab are $T = 0.012$ and $T = 0.001$ respectively. This suggests that EOT is achievable with

Figure 6.3: Field distribution plots depicting the impedance to exponential decay of the fields with a multi-layered structure as opposed to a single metal plate with equivalent thickness. The volume of silver is kept equal in both systems. EOT for multi-layered structures is clearly observable from the values of the associated optical transmission T.

the proposed multi-layered structures.

To show that these effects are due to surface plasmons, and not waveguide modes or other possible evanescent modes, we repeated the simulations with tungsten with TM-polarized light, and with silver with TE-polarized light. The refractive index of tungsten at $\lambda = 500$ nm is taken to be $n_w = 3.38 - i\,2.68$, following the data of Palik (1985). At this wavelength, tungsten therefore does not support surface plasmons (since $\epsilon'_w > 0, \epsilon_w = n_w^2 = \epsilon'_w - i\epsilon''_w$). TE-polarized light do not support surface plasmons either,
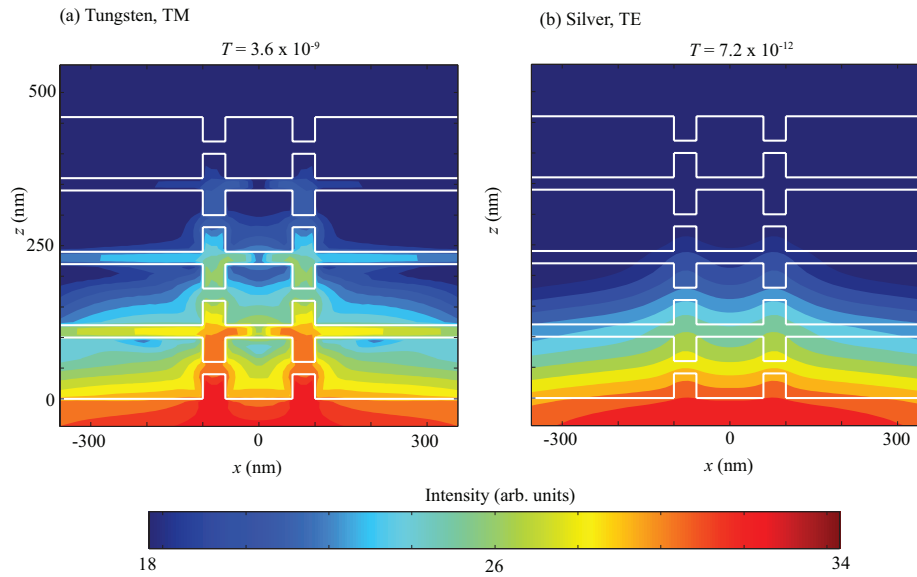
Figure 6.4: Field distribution plots for (a) tungsten multi-layered structure with TM-polarized light, and (b) silver multi-layered structure with TE-polarized light.

as has been explained in Chapter 2. For a subwavelength slit of width a, the cutoff wavelength for the TE dominant waveguide mode is $\lambda_c = 2a$ (Rao 2004, Sec 8.2). Since our plasmon pits are only 40 nm wide, waveguide modes are not supported as well for the TE case[1]. Thererfore for the TE case, one would expect very low optical transmission. Typical results are shown in Fig. 6.4. Clearly, none of the field distribution plots for the multi-layered structures in Fig. 6.4 bear any resemblance to those in Fig. 6.3. Furthermore, the optical transmission T is extremely low, orders of magnitude lower than for the case where surface plasmons are supported.

Though the multi-layered structures proposed in Fig. 6.1a could significantly impede exponential decay of fields in a single metallic slab, it is conceivable that these structures

[1]For the TM case, the TM_0 mode is a special case of the TEM (transverse electro-magentic) mode, having no cutoff. See for instance, Sec 9.2 of Cheng (1994).

present challenges for precise nano-fabrication. In this respect, we investigated simpler structures with only the plasmon pits on the illuminated side of the metal plate as depicted in Fig. 6.1b. Similar impeding of the field decay is still observed, although the effects are less prominent, and the EOT less significant, as seen from Fig. 6.5.

To compare the effectiveness in impeding the field decay and the EOT between the two structures proposed in Fig. 6.1, we may look at (a) a two-layered structure in Fig. 6.1a with a four-layered structure in Fig. 6.1b, and (b) a single-layered struture in Fig. 6.1a with a two-layered structure in Fig. 6.1b. It can be seen that, on account that g and t for the structure of Fig. 6.1b are half that of Fig. 6.1a, the volume of silver in both cases (a) and (b) is equivalent. For both the cases (a) and (b), the results for the field distribution and optical transmission for the two-layered structure are given
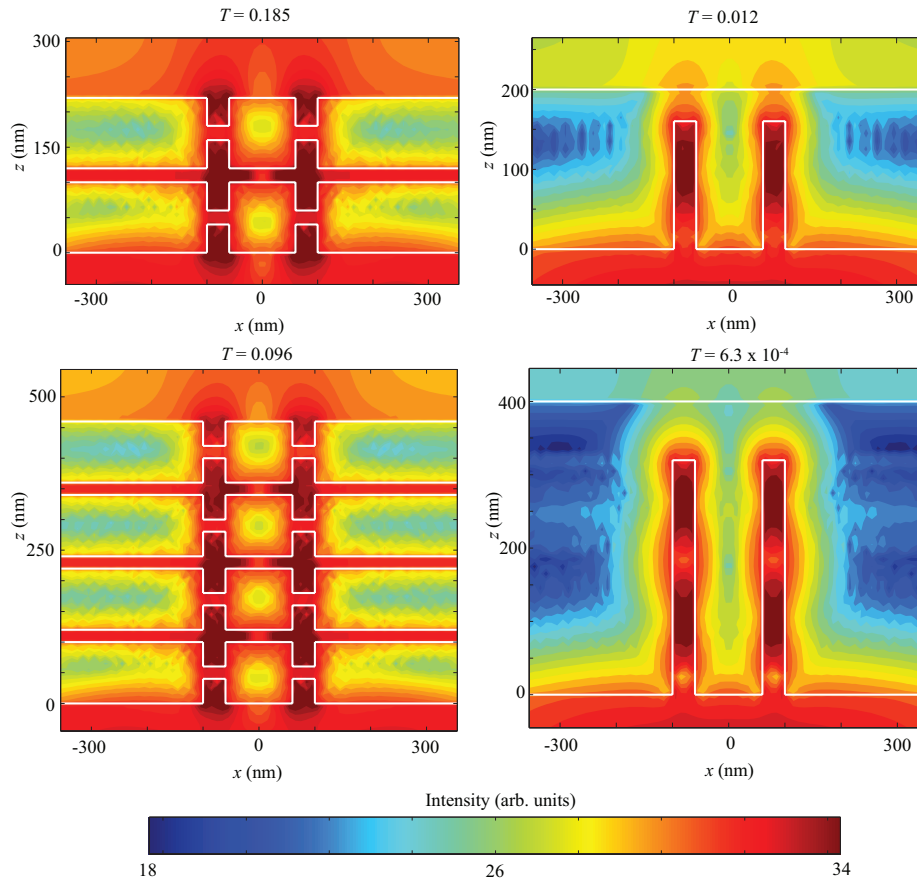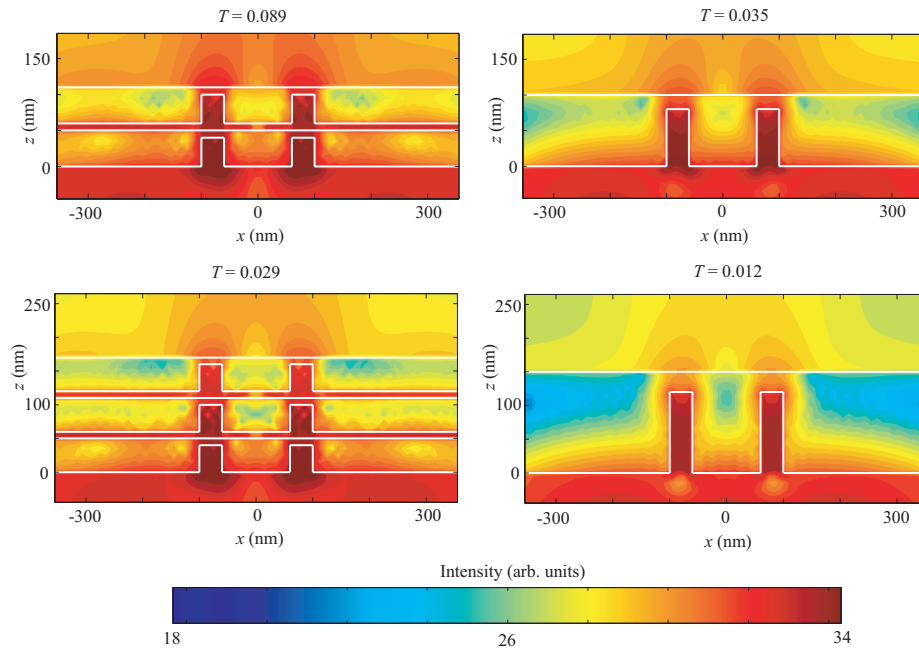

Figure 6.5: Field distribution plots depicting the impedance to exponential decay of the fields with multi-layered structures of Fig. 6.1b, as compared to a single metal plate with equivalent thickness. The volume of silver is kept equal in both systems.

in the top row of Fig. 6.3 and Fig. 6.5, respectively. As such, we show only the results

for the four-layered structure in Fig. 6.1b, and that of the single-layered struture in

Fig. 6.1a in Fig. 6.6. The results suggest that the structure with plasmon pits back to

back on both sides of the silver plate are more effective in impeding the exponential

decay and consequently produces higher EOT. A possible explanation for the better

performance with the structure of Fig. 6.1a is that coupling losses between separate
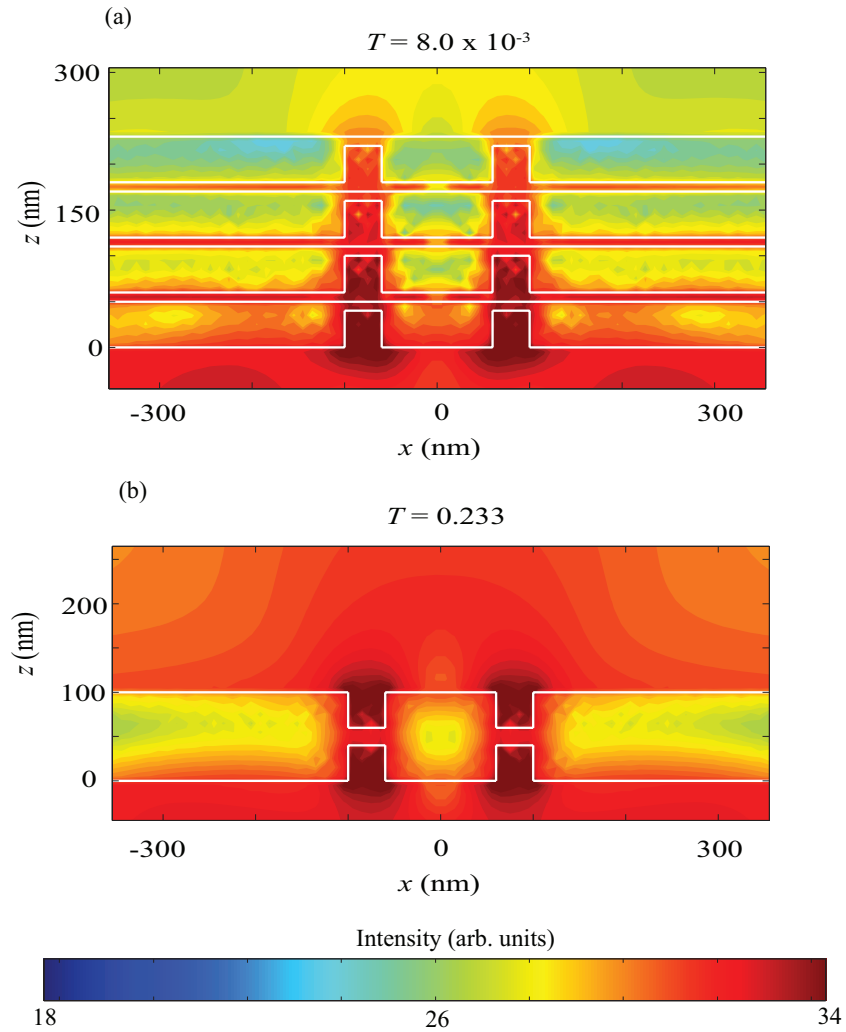
Figure 6.6: Field distribution plots for (a) a four-layered structure in Fig. 6.1b, and (b) a single-layered structure in Fig. 6.1a. These results are compared with those in the top row of Fig. 6.3 and Fig. 6.5, respectively.
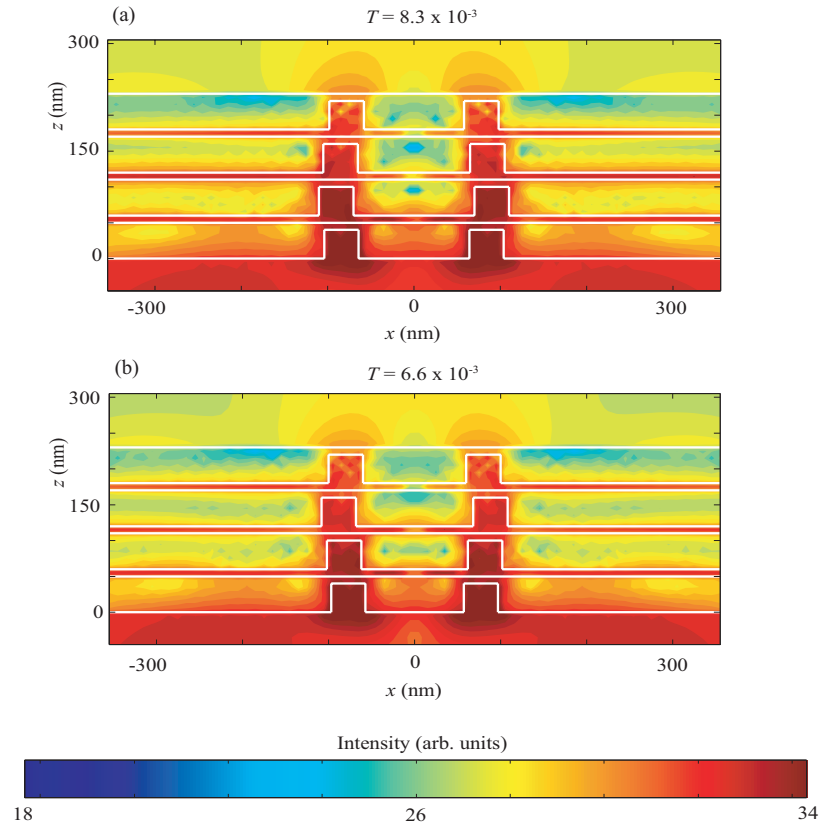
Figure 6.7: Field distribution plots depicting the impedance to exponential decay of the fields with multi-layered structures of Fig. 6.1b when each layer is allowed to be slightly misaligned. For the four layers, starting from the bottom layer, the plasmon pits are separated by $\gamma = 84, 90, 85, 78$ nm in (a), and $\gamma = 77, 82, 88, 80$ nm in (b).

layers can accumulate as the number of layers is increased.

Since the alignment of these multi-structured layers can be a daunting task in actual experiments, we have investigated how sensitive these effects are when there is some misalignment between each layer. The results are shown in Fig. 6.7. It is seen that the effects on the field decay are not dramatically different, and neither does the transmission T vary significantly from the perfectly aligned case ($T = 8.0 \times 10^{-3}$). Therefore, slight misalignments in the multi-layered structure does not cause drastic changes in the system response.

It is also interesting to see how the optical transmission varies for the multi-layered structure as compared to a slab of metal of equivalent volume. Our simulation for four layers of the multi-layered structure in Fig. 6.6a yields a transmission $T = 8.0 \times 10^{-3}$, slightly lesser compared to the transmission $T = 0.012$ for the 200 nm thick slab (see Fig. 6.3), suggesting that impeding the field decay does not always produce a better EOT. This is not unreasonable to expect since the transmission through a subwavelength slit is very much dependent on both its thickness and width (Schouten et al. 2004). In addition, we also like to see how different the optical transmission is for structures that can and cannot excite surface plasmons.

The optical transmission T for up to five layers or a single slab of equivalent volume for the cases: (i) silver$^{*(m,s)}$, TM, (ii) tungsten$^{*(m,s)}$, TM, (iii) silver$^{\dagger(m,s)}$, TM, and (iv) silver$^{\dagger(m,s)}$, TE, are tabulated in Table 6.1. The superscripted symbols $*$ and $\dagger$ are used to refer to the structure with back to back plasmon pits on each layer (Fig. 6.1a), and the structure with plasmon pits on only the illuminated side of each layer (Fig. 6.1b), respectively. The multi-layer structure and the equivalent single slab are represented by the superscripts m and s respectively. It is to be noted that the values for the transmission T in Table 6.1 are based on approximate numerical solutions, and are sufficient for our qualitative analysis in lieu of experimental data.

For the structures that support surface plasmons (silver, TM), it can be seen from Table 6.1 that the transmission for the multi-layered structure can be up to several times higher than that in the single silver slab. On the contrary, for structures that do not support surface plasmons (tungsten, TM, and silver, TE), it is seen that the transmission for the multi-layered structure can be lower or comparable to the case of

Table 6.1: Optical transmission T for the multi-layered structures (up to five layers) and equivalent single slabs. The number in [] to the right of the numeric value of T refers to the number of layers, where applicable. The transmission T is evaluated at z_T(nm), where z_T is the total thickness of the structure. For compact presentation, e^{-n} is used as short form for $\times 10^{-n}$.

	$z_T = 100$	$z_T = 220$	$z_T = 340$	$z_T = 460$	$z_T = 580$
silver[*](m), TM	0.233 [1]	0.185 [2]	0.134 [3]	0.096 [4]	0.066 [5]
tungsten[*](m), TM	0.011 [1]	$7.3e^{-5}$ [2]	$5.2e^{-7}$ [3]	$3.6e^{-9}$ [4]	$2.4e^{-11}$ [5]
	$z_T = 100$	$z_T = 200$	$z_T = 300$	$z_T = 400$	$z_T = 500$
silver[*](s), TM	0.035	0.012	$1.8e^{-3}$	$6.3e^{-4}$	$9.0e^{-5}$
tungsten[*](s), TM	0.012	$9.7e^{-4}$	$1.6e^{-4}$	$2.0e^{-5}$	$2.5e^{-6}$
	$z_T = 50$	$z_T = 110$	$z_T = 170$	$z_T = 230$	$z_T = 290$
silver[†](m), TM	0.185 [1]	0.089 [2]	0.029 [3]	$8.0e^{-3}$ [4]	$2.0e^{-3}$ [5]
silver[†](m), TE	0.064 [1]	$2.1e^{-3}$ [2]	$7.1e^{-5}$ [3]	$2.5e^{-6}$ [4]	$9.0e^{-8}$ [5]
	$z_T = 50$	$z_T = 100$	$z_T = 150$	$z_T = 200$	$z_T = 250$
silver[†](s), TM	0.185	0.035	0.012	0.012	0.021
silver[†](s), TE	0.064	$2.5e^{-3}$	$1.1e^{-4}$	$5.3e^{-6}$	$2.6e^{-7}$

the single slab. In addition, the structures that support surface plasmons are capable of optical transmission much higher than the structures that do not. These observations help support our claim that surface plasmons are responsible for the observed impeding field decay and EOT effects.

To further illustrate the differences in the transmission, the results for each of the
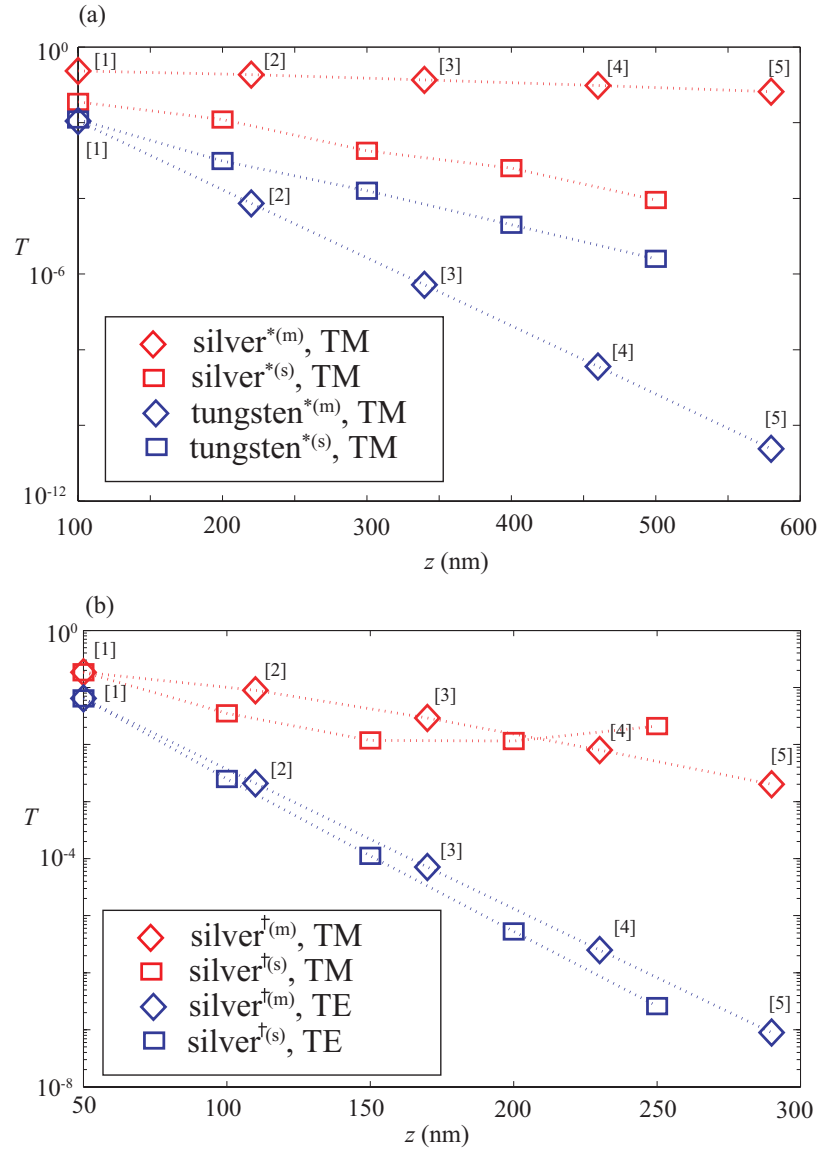
Figure 6.8: Plots of the transmission T as a function of z, for the data in Table 6.1. The data for silver*[(m,s)], TM, and tungsten*[(m,s)], TM, are plotted in (a), and that for silver†[(m,s)], TM, and silver†[(m,s)], TE, are plotted in (b).

structures in Table 6.1 are plotted in Fig. 6.8 . It is seen that when compared to the single slab, the silver*[(m)], TM structure is superior to the silver†[(m)], TM structure in terms of the acheivable EOT. This could be attributed to the much thicker slab for the silver*[(s)] structure, which is twice as stick as the silver†[(s)] structure. As the slab gets thicker, the effects of the skin depth predominates over the effects of the thickness of

Figure 6.9: Field distribution plots depicting increased impedance to exponential decay of the fields and EOT with multi-layered structures of Fig. 6.1a and Fig. 6.1b, with three plasmon pits instead of two in each layer.

the slab on the optical transmission. This behavior is also observable from the field distribution plots in Fig. 6.3 and Fig. 6.5.

Finally, we have performed another series of simulations with more plasmon pits to further confirm our claim that surface plasmons are responsible for these effects of EOT. Each of the plasmon pits can serve to enhance the light-plasmon coupling, and to confine the fields through the reflection of the plasmons at the edges. By having three

plasmon pits on each silver layer, we repeated the simulations, expecting the surface plasmon effects to be enahnced. Typical results are shown in Fig. 6.9. Comparing the results to those in Fig. 6.3 and Fig. 6.7, we see that there is a significant increase in the achievable EOT, and the suppression of the field decay is also improved. For the three-pit structures in Fig. 6.9, the optical transmission T is about twice as much as their corresponding two-pit structures. These results provide evidence that the enhanced coupling between light and surface plasmons is critical to achieve the effects for the multi-layered structures discussed in this section.

6.2   Summary and Future Work

Extrordinary optical transmission (EOT) with a multi-layered structure of corrugated metal plates was demonstrated. It was found that the multi-layered structure can impede the exponential decay of the field that would otherwise take place in a single silver plate. For consistency in our comparisons, the volume of air space due to the plasmon pits are kept equal in both the multi-layered structure and the single slab of silver.

From our simulations, these field effects related to the observed EOT can occur even with four metal slabs in cascade. Simulations with more plasmon pits and geometries that do not support surface palsmons also demonstrate the role of surface plasmons in the EOT and impeded decay of the fields. It is hoped that with advances and improvements to the processes for nano-fabrication of such multi-layered structures, cascaded corrugated metal films will find a host of interesting applications in future nano-optical technologies.

Chapter 7: Conclusions

I respect faith, but doubt is what gets you an education.

— Wilson Mizner

In this thesis, potential applications of surface plasmons to enhance or modulate optical field effects in subwavelength nano-optical systems have been demonstrated. The results presented are based on rigorous simulations of the Maxwell equations, which are useful in providing quantitative analyses for the systems we investigated.

Strategies to achieve super-resolution in near field optical readout systems were proposed, both for reflection and transmission readout configurations. To confirm that effects are due to surface plasmons, prominent local resonances observed in structures supporting surface plasmons were no longer observable in cases where surface plasmons are not excited. The size and position of plasmon pits on both the illuminated and dark side of the metal plates affect dramatically the response of the readout systems. Our simulations also reveal that the surface plasmon effects, which tend to be most intense at the edges of a structure, can be employed to detect the edges of the data structure. Each system has its associated pros and cons, and the choice of an optimal system will depend on what characteristics the system designer finds most important (readout contrast, resolution, stability). Resolution up to almost $\lambda/6$ was achieved in our simulations, which is a conservative estimate due to our stringent requirement for the resolution. Such plasmon-assisted readout systems have the potential to further

increase the data storage capacity offered by the current Blu-ray technology.

A new role for surface plamsons in nano-optics was also demonstrated: Theoretical modeling, verified by numerical simulations, was used to illustrate how surface plasmons in a Young's interference experiment can modulate the spatial coherence of light. Extending our investigations to a three-slit interferometer, it was found that the 'barrier' slit can suppress the degree of modulation as one would expect, but can also serve to increase the degree of modulation between the pair of end slits. These findings are promising to the development of a practical coherence converting device using a metallic hole array. Working towards such a device, the fringe contrast and spatial coherence of fields emerging from a symmetric three-hole interferometer was studied. Due to the simplicity of the geometry, analytical calculations for the spatial coherence as a function of the holes' separation d were readily derivable. It was found that the simulated fringe contrast varied accordingly with the calculated spatial coherence as a function of d. Work on larger arrays of holes illuminated by Schell-model fields is in progress. It is worthwhile to note that the enhancement of the spatial coherence by surface plasmons in a double-slit configuration has been experimentally demonstrated by collaborators in The Netherlands.

With multi-layered systems of metallic thin films, extraordinary optical transmission (EOT) accompanied with enhanced field effects were observed. In comparing the multi-layered systems to their equivalent structures in a single slab (no air gaps in between the layers), it was found that the decay of the field intensity can be significantly impeded or suppressed, leading in some cases to EOT. Total absence of these enhanced field effects in simulations with TE (transverse electric) polarized light and in materials that do not

excite surface plasmons supports our claim that these effects are plasmon-mediated. The striking suppression of the field decay over distances much longer than the skin depth, coupled with the achievable EOT in these multi-layered structures offers potential for application in near field optical systems where enhanced field effects are desired.

Like many others, the ideas investigated occurred to me in various ways; some out of pure imagination, some inspired by others, some as a result of unintended mistakes in running the simulations, and some over doubts that metal can ever effectively transmit light over distances much greater than the skin depth. It is hoped that these investigations have served to put more light to the sometimes questionable influence of surface plasmons in light/matter interactions. These results will hopefully lead to ideas that will unfold more exotic plasmon-assisted optical phenomena.

BIBLIOGRAPHY

Agarwal GS, Wolf E (1970) Calculus for Functions of Noncommuting Operators and General Phase-Space Methods in Quantum Mechanics. II. Quantum Mechanics in Phase Space. *Phys. Rev. D*, 2:2187–2205.

Agrawal GP, Pattanayak DN (1978) Gaussian beam propagation beyond the paraxial approximation. *J. Opt. Soc Am.*, 69:575–578.

Altewischer E, van Exter MP, Woerdman JP (2002) Plasmon-assisted transmission of entangled photons. *Nature*, 418:304–306.

Arfken G (1970) *Mathematical Methods for Physicists*. Academic Press, New York, 2nd edition.

Ash EA, Nicholls G (1972) Super-resolution aperture scanning microscope. *Nature*, 237:510–512.

Atwater HA (2007) The promise of plasmonics. *Scientific American Reports*, 17:56–63.

Bai B, Li L, Zeng L (2005) Experimental verification of enhanced transmission through two-dimensionally corrugated metallic films without holes. *Opt. Lett.*, 30:2360–2362.

Bakker RM, Drachev VP, Yuan HK, Shalaev VM (2004) Enhanced transmission in near-field imaging of layered plasmonic structures. *Opt. Express*, 12:3701–3706.

Barnes WL, Dereux A, Ebbesen TW (2003) Surface plasmon subwavelength optics. *Nature*, 424:824–830.

Barnes WL, Murray WA, Dintinger J, Devaux E, Ebbesen TW (2004) Surface plasmon polaritons and their role in the enhanced transmission of light through periodic arrays of subwavelength holes in a metal film. *Phys. Rev. Lett.*, 92:107401.

Bastiaansen H. J. M., Baken N. H. G., Blok H (1992) Domain-integral analysis of channel waveguides in anisotropic multi-layered media. *IEEE Trans. on Microwave Theory and Tech.*, 40:1918–1926.

Bergman DJ, Stockman MI (2003) Surface plasmon amplification by stimulated emission of radiation: quantum generation of coherent surface plasmons in nanosystems. *Phys. Rev. Lett.*, 90:027402.

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

Berini P, Charbonneau R, Jetté-Charbonneau S, Lahoud N, Mattiussi G (2007) Long-range surface plasmon-polariton waveguides and devices in lithium niobate. *J. Appl. Phys.*, 101:113114.

Bethe HA (1944) Theory of diffraction by small holes. *Phys. Rev.*, 66:163–182.

Betzig E, Trautman JK (1992) Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction Limit. *Science*, 257:189–195.

Betzig E, Trautman JK, Harris TD, Weiner JS, Kostelak RL (1991) Breaking the diffraction barrier: optical microscopy on a nanometric scale. *Science*, 251:1468–1470.

Betzig E, Trautman JK, Wolfe R, Gyorgy G, Finn PL, Kryder MH, Chang CH (1992) Near field magneto-optics and high density data storage. *Appl. Phys. Lett.*, 61:142–144.

Boardman AD (1982) *Electromagnetic Surface Modes*. John Wiley & Sons, Inc., New York.

Bohren CF, Huffman D (1983) *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, Inc., Toronto.

Bonod N, Enoch S, Li L, Popov E, Nevière M (2003) Resonant optical transmission through thin metallic films with and without holes. *Opt. Express*, 11:482–490.

Born M, Wolf E (1999) *Principles of Optics*. Cambridge University Press, Cambrdige, 7th edition.

Bouhelier A (2006) Field-enhanced scanning near-field optical microscopy. *Microsc. Res. Tech.*, 69:563–579.

Bouhelier A, Renger J, Beversluis MR, Novotny L (2003) Plasmon-coupled tip-enhanced near-field optical microscopy. *J. Microsc.*, 210:220–224.

Bozhevolnyi SI, Coello V (1998) Elastic scattering of surface plasmon polaritons: modeling and experiment. *Phys. Rev. B*, 58:10899–10910.

Bozhevolnyi SI, Volkov VS (2001) Multiple-scattering dipole approach to modeling of surface plasmon polariton band gap structures. *Opt. Comm.*, 198:241–245.

Brau CA (2004) *Modern Problems in Classical Electrodynamics*. Oxford University Press, New York.

Cai W, Chettiar UK, Kildishev AV, Shalaev VM (2007) Optical cloaking with metamaterials. *Nature Photonics*, 1:224–227.

Cao YC, Jin R, Mirkin CA (2002) Nanoparticles with raman spectroscopic fingerprints for dna and rna detection. *Science*, 297:1536–1540.

Carter WH (1972) Electromagnetic field of a Gaussian beam with an elliptical cross section. *J. Opt. Soc Am.*, 62:1195–1201.

Carter WH, Bertolotti M (1978) An analysis of the far field coherence from liquid crystals. *J. Opt. Soc. Am.*, 68:329–333.

Chan HB, Marcet Z, Woo K, Tanner DB, Carr DW, Bower JE, Cirelli RA, Ferry E, Klemens F, Miner J, Pai CS, Taylor JA (2006) Optical transmission through double-layer metallic subwavelength slit arrays. *Opt. Lett.*, 31:516–518.

Chen CG, Konkola PT, Ferrera J, Heilmann RK, Schattenburg ML (1972) Analyses of vector Gaussian beam propagation and the validity of paraxial and spherical approximations. *J. Opt. Soc Am. A*, 19:404–412.

Cheng DK (1994) *Fundamentals of Engineering Electromagnetics.* Addison-Wesley, Massachusetts.

Cheng C, Chen J, Wu Q, Ren F, Xu J, Fan Y, Wang H (2007) Controllable electromagnetic transmission based on dual-metallic grating structures composed of subwavelength slits. *Appl. Phys. Lett.*, 91:111111.

Chew WC (1995) *Waves and Fields in Homogeneous Media.* IEEE Press, New York.

Chiao RY, Milonni PW (2002) Fast light, slow light. *Opt. Photonics News*, 13:26–30.

Collin S, Padro F, Teissier R, Pelouard JL (2001) Strong discontinuities in the complex photonic band structure of transmission metallic gratings. *Phys. Rev. B*, 63:033107.

Dainty C (1975) *Laser Speckle and Related Phenomena.* Springer, Berlin.

Darmanyan SA, Zayats AV (2003) Light tunneling via resonant surface plasmon polariton states and the enhanced transmission of periodically nanostructured metal films: An analytical study. *Phys. Rev. B*, 67:035424.

Degiron A, Ebbesen TW (2004) Analysis of the transmission process through single apertures surrounded by periodic corrugations. *Opt. Express*, 12:3694–3700.

Ditlbacher H, Krenn JR, Schider G, Leitner A, Aussenegg FR (2002) Two-dimensional optics with surface plasmon polaritons. *Appl. Phys. Lett.*, 81:1762–1764.

Ebbesen TW, Lezec HJ, Ghaemi HF, Thio T, Wolff PA (1998) Extraordinary optical transmission through sub-wavelength hole arrays. *Nature*, 391:667–669.

Ekert AK (1991) Multiple paths to enhance optical transmission through a single subwavelength slit. *Phys. Rev. Lett.*, 67:661–663.

Ferrell RA (1958) Predicted radiation of plasma oscillations in metal films. *Phys. Rev.*, 111:1214–1222.

Foldy LL (1945) The multiple scattering of waves. *Phys. Rev.*, 67:107–119.

Fox M (2004) *Optical Properties of Solids.* Oxford University Press, New York.

Fukui M, So VCY, Normandin R (1979) Lifetimes of surface plasmons in thin silver films. *Phys. Stat. Sol.*, 91:K61–K64.

Gan CH, Gbur G (2007) Strategies for employing surface plasmons in a near-field transmission optical readout system. *Appl. Phys. Lett.*, 91:131109.

Gan, CH, Gbur G. (2008) Spatial coherence conversion with surface plasmons using a three-slit interferometer. *Plasmonics*, 3:111–117

Gan, CH, Gbur G, Visser TD (2007) Surface plasmons modulate the spatial coherence of light in Young's interference experiment. *Phys. Rev. Lett.*, 98:043908.

García-Vidal F, Lezec HJ, Ebbesen TW, Martín-Moreno L (2003) Multiple paths to enhance optical transmission through a single subwavelength slit. *Phys. Rev. Lett.*, 90:213901.

Gay G, Alloschery O, Lesegno BVD, O'Dwyer C, Weiner J, Lezec HJ (2006) Interaction between optical nano-objects at metallo-dielectric interfaces. *Nature Phys.*, 2:262–267.

Gbur G, Schouten HF, and Visser TD (2005) Achieving superresolution in near-field optical data readout systems using surface plasmons. *Appl. Phys. Lett.*, 87:191109.

Gbur G, Wolf E (2002) Spreading of partially coherent beams in random media. *J. Opt. Soc. Am. A*, 19:1592–1598.

Ghaemi HF, Thio T, Grupp DE, Ebbesen TW, Lezec HJ (1998) Surface plamsons enhance optical transmission through subwavelength holes. *Phys. Rev. B*, 58:6779–6782.

Giannattasio A, Hooper IR, Barnes WL (2004) Transmission of light through thin silver films via surface plasmon polaritons. *Opt. Express*, 12:5881–5886.

Ginsberg NS, Garner SR, Hau LV (2007) Coherent control of optical information with matter wave dynamics. *Nature*, 445:623–626.

Gisin N, Thew R (2007) Quantum communication. *Nature Photonics*, 1:165–171.

Gonsiorowski T, Dainty JC (1983) Correlation properties of light produced by quasi-thermal sources. *J. Opt. Soc. Am.*, 73:234–237.

Gontijo I, Boroditsky M, Yablonovitch E, Keller S, Mishra UK, Denbaars SP (1999) Coupling of InGaN quantum-well photoluminescence to silver surface plasmons. *Phys. Rev. B*, 60:11564–11567.

Goodman JW (2000) *Statistical Optics*. John Wiley & Sons, Inc., New York.

Goodman JW (2005) *Introduction to Fourier Optics*. Roberts & Company, Englewood, 3rd edition.

Gordon R, Brolo AG (2005) Increased cut-off wavelength for a subwavelength hole in a real metal. *Opt. Express*, 13:1933–1938.

Grandin H (1986) *Fundamentals of the Finite Element Method*. Macmillan Publishing, New York.

Griffiths DJ (1981) *Introduction to Electrodynamics*. Prentice Hall, New Jersey.

Grupp DE, Lezec HJ, Ebbesen TW, Pellerin KM, and Thio T (2000) Crucial role of metal surface in enhance transmission through subwavelength apertures. *Appl. Phys. Lett.*, 77:1569–1571.

Hall JE, Wiederrecht GP, Gray SK, Chang S, Jeon S, Rogers JA, Bachelot R, Royer P (2007) Heterodyne apertureless near-field scanning optical microscopy on periodic gold nanowells. *Opt. Express*, 15:4098–4105.

Hecht B, Bielefeldt H, Novotny L, Inouye Y, Pohl DW (1996) Local excitation, scattering, and interference of surface plasmons. *Phys. Rev. Lett.*, 77:1889–1892.

Hecht E (1998) *Optics.* Addison Wesley Longman, 3rd edition.

Her Y, Lan Y, Hsu W, Tsai S (2003) Recording and readout mechanisms of super-resolution near-field structure disk with a silver oxide mask layer. *Appl. Phys. Lett.*, 83:2136–2138.

Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA (1991) Optical coherence tomography. *Science*, 254:1178–1181.

Ishimaru A (1991) *Electromagnetic Wave Propagation, Radiation, and Scattering.* Prentice Hall, New Jersey.

Jackson JD (1975) *Classical Electrodynamics.* John Wiley & Sons, Inc., New York, 2nd edition.

James DFV (1994) Change of polarization of light beams on propagation in free space. *J. Opt. Soc. Am A*, 11:1641–1643.

Johnson PB, Christy RW (1972) Optical constants of the noble metals. *Phys. Rev. B*, 6:4370–4379.

Jones DS (1986) *Acoustic and Electromagnetic Waves.* Oxford University Press, Berlin.

Karalis A, Lidorikis E, Ibanescu M, Joannopoulos JD, Soljačić M (2005) Surface-plasmon-assisted guiding of broadband slow and subwavelength light in air. *Phys. Rev. Lett.*, 95:063901.

Kim J H, Moyer PJ (2006) Transmission characteristics of metallic equilateral triangular nanohole arrays. *Appl. Phys. Lett.*, 89:121106.

Koerkamp KJK, Enoch S, Segerink FB, van Hulst NF, Kuipers L (2004) Strong influence of hole shape on extraordinary transmission through periodic arrays of subwavelength holes. *Phys. Rev. Lett.*, 92:183901.

Kolk EW, Baken NHG, Blok H (1990) Domain integral equation analysis of integrated optical channel and ridge waveguides in stratified media. *IEEE Trans. on Microwave Theory and Tech.*, 38:78–85.

Korotkova O, Andrews LC, Phillips RL (2004) A model for a partially coherent Gaussian beam in atmospheric turbulence with application in Lasercom. *Opt. Eng.*, 43:330–341.

Korotkova O, Salem M, Wolf E (2004) Beam conditions for radiation generated by an electromagnetically gaussian schell-model source. *Opt. Lett.*, 29:1173–1175.

Krenn JR, Lamprecht B, Ditlbacher H, Schider G, Salerno M, Leitner A, Aussenegg FR (2002) Non-diffraction-limited light transport by gold nanowires. *Europhys. Lett.*, 60:663–669.

Kubo R, Toda M, Hashitsume N (1998) *Statistical Physics II*. Springer, Berlin, 2nd edition.

Kuzmin N, 't Hooft G, Eliel E, Gbur G, Schouten HF, Visser TD (2007). Enhancement of spatial coherence by surface plasmons. *Opt. Lett.*, 32:445–447.

Lal S, Link S, Halas NJ (2007) Nano-optics from sensing to waveguiding. *Nature Photonics*, 1:641–648.

Lalanne P, Hugonin JP (2006) Interaction between optical nano-objects at metallo-dielectric interfaces. *Nature Phys.*, 2:551–556.

Lalanne P, Hugonin JP, Astilean S, Palamaru M, Moller KD (2000) One-mode model and airy-like formulae for one-dimensional metallic gratings. *J. Opt. A.: Pure Appl. Opt.*, 2:48–51.

Lalanne P, Hugonin JP, Rodier J (2006) Approximate model for surface-plasmon generation at slit apertures. *J. Opt. Soc. Am. A*, 23:1608–1615.

Lezec HJ, Thio T (2004) Diffracted evanescent wave model for enhanced and suppressed optical transmission through subwavelength hole arrays. *Opt. Express*, 12:3629–3651.

Lezec HJ, Degiron A, Devaux E, Linke RA, Martín-Moreno L, García-Vidal FJ, Ebbesen TW (2001) Beaming light from a subwavelength aperture. *Science*, 297:820–822.

Liu H, Lalanne P (2008) Microscopic theory of the extraordinary optical transmission. *Nature*, 452:728–731.

Loo C, Lin A, Hirsch L, Lee M, Barton J, Halas NJ, West J, Drezek R (2004) Nanoshell-enabled photonics-based imaging and therapy of cancer. *Tech. Cancer Res. & Treatment*, 3:33–40.

Luo X, Ishihara T (2004) Surface plasmon resonant interference lithography technique. *Appl. Phys. Lett.*, 84:4780–4782.

Maier SA, Atwater HA (2005) Plasmonics: Localization and guiding of electromagnetic energy in metal/dielectric structures. *J. Appl. Phys.*, 98:011101.

Mamun AA, Ge SS (2005) Precision control of hard disk drives. *IEEE Control Systems Magazine*, 25:14–19.

Mandel L, Wolf E (1995) *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge.

Marcet Z, Paster JW, Carr DW, Bower JE, Cirelli RA, Klemens F, Mansfield WM, Miner JF, Pai CS, Chan HB (2008) Controlling the phase delay of light transmitted through double-layer metallic subwavelength slit arrays. *Opt. Lett.*, 33:1410–1412.

Marcuse D (1982) *Light Transmission Optics*. Van Nostrand-Reinhold, New York, 2nd edition.

Martienssen W, Spiller E (1964) Coherence and fluctuations in light beams. *Am. J. Phys.*, 32:919–926.

Martin OJF (2003) Surface plasmon illumination scheme for contact lithography beyond the diffraction limit. *Microelect. Engg.*, 67:24–30.

Martín-Moreno L, García-Vidal FJ, Lezec HJ, Pellerin KM, Thio T, Pendry JB, Ebbesen TW (2001) Theory of extraordinary optical transmission through subwavelength hole arrays. *Phys. Rev. Lett.*, 86:1114–1117.

Martín-Moreno L, García-Vidal FJ, Lezec HJ, Degiron A, Ebbesen TW (2001) Theory of highly directional emission from a single subwavelenght aperture surrounded by surface corrugations. *Phys. Rev. Lett.*, 90:167401.

Melia F (2001) *Electrodynamics*. The University of Chicago Press, Chicago.

Nie S, Emory SR (1997) Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *Science*, 275:1102–1106.

Ohtsuka Y (1986) Modulation of optical coherence by ultrasonic waves. *J. Opt. Soc. Am. A*, 3:1247–1257.

Pacifici D, Lezec HJ, Atwater HA (2007) All-optical modulation by plasmonic excitation of cdse quantum dots. *Nature Photonics*, 1:402–406.

Palik E (1985) *Handbook of Optical Constants of Solids*. Academic Press, New York.

Paulus M, Gay-Balmaz P, and Martin OJF (2000) Accurate and efficient computation of the green tensor for stratified media. *Phys. Rev. E*, 62:5797–5807.

Peierls R (1979) *Suprises in Theoretical Physics*. Princeton University Press, New Jersey.

Pendry JB, Schurig D, Smith DR (2006) Controlling electromagnetic fields. *Science*, 312:1780–1782.

Peyghambarian N, Koch SW, Mysyrowicz A (1993) *Introduction to Semiconductor Optics*. Prentice Hall, Englewood Cliffs.

Phillips DF, Fleischhauer A, Mair A, Walsworth RL, and Lukin MD (2002) Storage of light in atomic vapor. *Phys. Rev. Lett.*, 86:783–786.

Pile DFP, Ogawa T, Gramotnev DK, Matsuzaki Y, Vernon KC, Yamaguchi K, Okamoto T, Haraguchi M, Fukui M (2005) Two-dimensionally localized modes of a nanoscale gap plasmon waveguide. *Appl. Phys. Lett.*, 87:261114.

Pohl DW, Denk W, Lanz M (1984) Stethoscopy: image recording with resolution $\lambda/20$. *Appl. Phys. Lett.*, 44:651–653.

Pohlmann KC (1992) *The Compact Disc Handbook*. Oxford University Press, Oxford, 2nd edition.

Popov E, Nevière M, Enoch S, Reinisch R. (2002) Theory of light transmission through subwavelength periodic hole arrays. *Phys. Rev. B*, 62:16100–16108.

Powell CJ, Swan JB (1959) Effect of oxidation on the characteristic loss spectra of aluminum and magnesium. *Phys. Rev.*, 118:640–643.

Prade B, Vinet JY, Mysyrowicz A (1991) Guided optical waves in planar heterostructures with negative dielectric constant. *Phys. Rev. B*, 44:13556–13572.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2005) *Numerical Recipes in C++*. Cambridge University Press, Cambridge, 2nd edition.

Raether H (1988) *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*. Springer, Berlin.

Rao NN (2004) *Elements of Engineering Electromagnetics*. Prentice Hall, New Jersey, 6th edition.

Reddy JN (1993) *An Introduction to the Finite Element Method*. McGraw-Hill, Inc., New York, 2nd edition.

Reitz JR, Milford FJ, Christy RW (1980) *Foundations of Electromagnetic Theory*. Addison-Wesley Publishing Company, Inc., Don Mills, Ontario, 3rd edition.

Ricklin JC, Davidson FM (2002) Atmospheric turbulence effects on a partially coherent Gaussian beam: implications for free-space laser communication. *J. Opt. Soc. Am. A*, 19:1794–1802.

Ritchie RH (1957) Plasma losses by fast electrons in thin films. *Phys. Rev.*, 106:874–881.

Ruppin R (2000) Surface polaritons on a left-handed medium. *Phys. Lett. A*, 277:61–64.

Saleh BEA, Abouraddy AF, Sergienko AV, Teich MC (2005) Duality between partial coherence and partial entanglement. *Phys. Rev. A*, 62:043816.

Salem M, Korotkova O, Dogariu A, Wolf E (2004) Polarization changes in partially coherent EM beams propagating through turbulent atmosphere. *Waves in Random Media*, 14:513–523.

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

Schouten HF, Visser TD, Gbur G, Lenstra D, Blok H (2004) The diffraction of light by narrow slits in plates of different materials. *J. Opt. A.: Pure Appl. Opt.*, 6:S277–S280.

Schouten HF, Kuzmin N, Dubois G, Visser TD, Gbur G, Alkemade PFA, Blok H, 't Hooft GW, Lenstra D, Eliel ER (2005) Plasmon-assisted two-slit transmission: Young's experiment revisited. *Phys. Rev. Lett.*, 94:053901.

Schröter U, Seider S, Tode S, Heitmann D (1997) Surface plasmon reflection at edges and resonance effects in metal bars. *Ultramicroscopy*, 68:223–230.

Scudieri F, Bertolotti M, Bartolino R (1974) Light scattered by a liquid crystal: a new quasi-thermal source. *Appl. Opt.*, 13:181–185.

Shi LP, Chong TC, Yao HB, Tan PK, Miao XS (2002) Super-resolution near-field optical disk with an additional localized surface plasmon coupling layer. *J. Appl. Phys.*, 91:10209–10211.

Shin H, Catrysse PB, Fan S (2005) Effect of the plasmonic dispersion relation on the transmission properties of subwavelength cylindrical holes. *Phys. Rev. B*, 72:085436.

Søndergaard T, Bozhevolnyi SI (2003) Vectorial model for multiple scattering by surface nanoparticles via surface polariton-to-polariton interactions. *Phys. Rev. B*, 67:165405.

Sphicopoulos T, Teodoridis V, Gardiol FE (1985) Dyadic green function for the electromagnetic field in multi-layered isotropic media: an operator approach. *IEE Proc.*, 132:399–403.

Spiller TP (1996) Quantum information processing: cryptography, computation, and teleportation. *Proc. IEEE*, 84:1719–1746.

Stamnes JJ (1986) *Waves in Focal Regions*. Adam Hilger Ltd, Bristol and Boston.

Strutt JW (III Lord Rayleigh) (1879) Investigations in Optics, with special reference to the spectroscope. *Phil. Mag.*, 8:261–274.

Suzuki T, Itoh Y, Birukawa M, Drent WV (1998) Solid immersion lens near field optical approach for high density optical recording. *IEEE Trans. Magnetics*, 34:329–334.

Tai CT (1971) *Dyadic Green Functions in Electromagnetic Theory*. Intext Educational Publishers, Scranton.

Takahara J, Kobayashi T (2004) From subwavelength optics to nano-optics. *Opt. Photon. News*, 15:54–59.

Takakura Y (2001) Optical resonance in a narrow slit in a thick metallic screen. *Phys. Rev. Lett.*, 86:5601–5603.

Tan EL (2002) A robust formulation of saw green functions for arbitrarily thick multilayers at high frequencies. *IEEE Trans. on Ultrasonics, Ferroelectrics, and Freq. Control*, 49:929–936.

Tang CL (2005) *Fundamentals of Quantum Mechanics*. Cambridge University Press, Cambridge.

Taylor HF, Yariv A (1974) Guided wave optics. *Proc. IEEE*, 62:1044–1060.

Teng Y, Stern EA (1967) Plasma radiation from metal grating surfaces. *Phys. Rev. Lett.*, 19:511–514.

Terris BD, Mamin HJ, Rugar D, Studenmund WR, Kino GS (1994) Near field optical data storage using a solid immersion lens. *Appl. Phys. Lett.*, 65:388–390.

Thio T, Pellerin KM, Linke RA, Lezec HJ, Ebbesen TW (2001) Enhanced light transmission through a single subwavelength aperture. *Opt. Lett.*, 26:1972–1974.

Tominaga J, Tsai DP (2003) *The Manipulation of Surface and Local Plasmons*. Springer, Heidelberg.

Török P, Kao F (2003) *Optical Imaging and Microscopy*. Springer, Heidelberg.

Turunen J, Tervonen E, Friberg AT (1990) Acousto-optic control and modulation of optical coherence by electronically synthesized holographic gratings. *J. Appl. Phys.*, 67:49–59.

Urbach HP, Lepelaars ESAM (1994) On the domain integral equation method for anisotropic inhomogeneous waveguides. *IEEE Trans. on Microwave Theory and Tech.*, 42:118–126.

Vahimaa P, Turunen J (1997) Bragg diffraction of spatially partially coherent fields. *J. Opt. Soc. Am. A*, 14:54–59.

van de Nes AS, Braat JJM, Pereira SF (2006) High-density optical data storage. *Rep. Prog. Phys.*, 69:2323–2363.

van der Molen KL, Segerink FB, van Hulst NF, Kuipers L (2004) Influence of hole size on the extraordinary transmission through subwavelength hole arrays. *Appl. Phys. Lett.*, 85:4316–4318.

Visser TD (2006) Surface plasmons at work? *Nature Phys.*, 2:509–510.

Visser TD, Blok H, Lenstra D (1999) Theory of polarization-dependent amplification in a slab waveguides with anisotropic gain and losses. *IEEE Journal of Quantum Electronics*, 35:240–249.

Wolf E (1978) Coherence and radiometry. *J. Opt. Soc. Am*, 68:6–17.

Wolf E (1986) Invariance of the spectrum of light on propagation. *Phys. Rev. Lett.*, 56:1370–1372.

Wolf E (2003) Significance and measurability of the phase of a spatially coherent optical field. *Opt. Lett.*, 28:5–6.

Wolf E (2007) *Introduction to the Theory of Coherence and Polarization of Light.* Cambridge University Press, Cambridge.

Wolf E, James DFV (1996) Correlation-induced spectral changes. *Rep. Prog. Phys.*, 59:771–818.

Wu J (1990) Propagation of a Gaussian-Schell beam through turbulent media. *J. Mod Opt.*, 37:671–684.

Xie Y, Zakharian A, Moloney J, Mansuripur M (2004) Transmission of light through slit apertures in metallic films. *Opt. Express*, 12:6106–6121.

Xu H, Bjerneld EJ, Käll M, Böjesson L (1999) Spectroscopy of single hemoglobin molecules by surface enhanced raman scattering. *Phys. Rev. Lett.*, 83:4357–4360.

Yang F, Sambles JR (2002) Resonant transmission of microwaves through a narrow metallic slit. *Phys. Rev. Lett.*, 89:063901.

Yang F, Bradberry GW, Sambles JR (2002) Long-range surface mode supported by very thin silver films. *Phys. Rev. Lett.*, 66:2030–2032.

Ye Y, Zhang J (2005) Enhanced light transmission through cascaded metal films perforated with periodic hole arrays. *Opt. Lett.*, 30:1521–1523.

Zayats AV, Smolyaninov II (2003) Near-field photonics: surface plasmon polaritons and localized surface plasmons. *J. Opt. A.: Pure Appl. Opt.*, 5:S16–S50.

Zeilinger A (2000) Quantum teleportation. *Scientific American*, 282:32–41.

APPENDIX A: ANGULAR SPECTRUM REPRESENTATION FOR VECTOR GAUSSIAN BEAMS IN SOURCE-FREE REGIONS

We consider here the construction of a vectorial Gaussian beam through an angular spectrum of plane waves. It is assumed that the beam is monochromatic and that it propagates in a medium void of free charges. Such a wave must satisfy the vector wave equation derived in Sec 2.1 (see Eq. (2.14))

$$\nabla^2 \mathbf{E_r} + k^2 \mathbf{E_r} = 0 \,, \tag{A-1}$$

with $k = n k_0$, where $k_0 = \omega/c$ is the wave vector in free space, and n is the refractive index at frequency ω. As the magnetic field can be determined by taking the curl of the electric field in the monochromatic case, it is sufficient to consider the electric field.

In the case of a two-dimensional geometry, a further simplification can be made since the TM (transverse magnetic) and TE (transverse electric) polarizations satisfy independent sets of the Maxwell equations. Let us take our geometry to be uniform along the y-axis, such that $k_y = 0$. In terms of the angular spectrum representation, the electric field components can be expressed as (Carter 1972, Agrawal and Pattanayak 1978, Chen et al. 2002)

$$E_x(x, z) = \int_{-\infty}^{\infty} a_x(k_x) e^{-i(k_x x + k_z z)} \, dk_x, \,, \tag{A-2a}$$

$$E_z(x, z) = -\int_{-\infty}^{\infty} \frac{k_x}{k_z} a_x(k_x) e^{-i(k_x x + k_z z)} \, dk_x \,, \tag{A-2b}$$

for the TM polarization, and

$$E_y(x,z) = \int_{-\infty}^{\infty} a_y\,(k_x) e^{-\mathrm{i}(k_x x + k_z z)}\,dk_x \ , \tag{A-3}$$

for TE polarization, with

$$k_z = \begin{cases} \sqrt{k^2 - k_x^2} & \text{if } k^2 \geq k_x^2 \\[2mm] -\mathrm{i}\sqrt{(k_x^2 - k^2)} & \text{if } k^2 < k_x^2 \end{cases} \ . \tag{A-4}$$

The complex factors $a_x\,(k_x)$ and $a_y\,(k_x)$ are weighting factors that are to be determined. Eqs. (A-2) and (A-3) show that, for a two-dimensional geometry, each field component may be represented as a sum of suitably weighted plane waves propagating in different directions given by k_x and k_z (Stamnes 1986, Goodman 2005). For the waves to be propagating, k_x and k_z must also satisfy the relation

$$k_x^2 + k_z^2 \leqslant k^2 \ . \tag{A-5}$$

Let us assume that at the boundary $z = 0$, the beam has an initial Gaussian profile

$$E_x(x,0) = e^{-(x^2/2\sigma_x^2)} \ , \tag{A-6}$$

for the TM case, and

$$E_y(x,0) = e^{-(x^2/2\sigma_y^2)} \ , \tag{A-7}$$

for the TE case, where σ_x and σ_y are measures of the beamwidth. Further taking the

inverse Fourier transform of Eqs. (A-2a) and (A-3), we solve for the factors $a_x(k_x)$ and

$a_y(k_x)$ as

$$a_{x,y}(k_x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(k_x^2/2\sigma^2)}, \tag{A-8}$$

where depending on the polarization,

$$\sigma = \begin{cases} 1/\sigma_x & \text{for TM} \\ 1/\sigma_y & \text{for TE} \end{cases}. \tag{A-9}$$

Substituting Eq. (A-8) back into Eqs. (A-2) and (A-3), and noting the condition (A-5) for the limits of integration, one obtains the following expressions for the angular spectrum of each of the field components:

$$E_{x,y}(x,z) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{k_x^2+k_z^2 \leqslant k^2} e^{-(k_x^2/2\sigma^2)}\, e^{-i(k_x x + k_z z)}\, dk_x\ , \tag{A-10}$$

and

$$E_z(x,z) = -\frac{1}{\sqrt{2\pi}\,\sigma} \int_{k_x^2+k_z^2 \leqslant k^2} \frac{k_x}{k_z}\, e^{-(k_x^2/2\sigma^2)}\, e^{-i(k_x x + k_z z)}\, dk_x\ . \tag{A-11}$$

VITA

The author was born on October 22, 1976 in the island city-state of Singapore. He received his BEng in Electrical and Electronic Engineering in June 2001 from the Nanyang Technological University, Singapore, majoring in Photonics. Between 2002 and 2004, he worked as a Project Engineer with the Defense Science Technology Agency, during which he consulted for the Republic of Singapore Navy on advanced underwater communication systems. He left for the University of North Carolina at Charlotte in the fall of 2004 to pursue the PhD degree in Optical Science and Engineering. Under the supervision of Dr. Greg Gbur for his doctoral research, his current research interests include surface plasmon effects, optical coherence, optical singularities, and nonradiating scatterers in one-dimensional wave systems. The work he has published in these fields include:

- C. H. Gan and G. Gbur, "Strategies for employing surface plasmons in near field optical readout systems," Optics Express **14**, 2385 (2006).

- C. H. Gan, G. Gbur and T. D. Visser, "Surface plasmons modulate the spatial coherence of light in Young's interference experiment," Phys. Rev. Lett **98**, 043908 (2007).

- D. M. Karabacak, K. L. Ekinci, C. H. Gan, G. Gbur, S. B. Ippolito, M. S. Unlu and P. S. Carney, "Diffraction of evanescent waves and detection of nanometer scale motion," Optics Lett. **32**, 1881 (2007).

- C. H. Gan and G. Gbur, "Phase and coherence singularities generated by the interference of partially coherent fields," Optics Commun. **280**, 249 (2007).

- C. H. Gan and G. Gbur, "Strategies for employing surface plasmons in a near field transmission optical readout system," Appl. Phys. Lett. **91**, 131109 (2007).

- C. H. Gan, G. Gbur and T. D. Visser, "A new role for surface plasmons," Optics and Photonics News, Dec 2007.

- C. H. Gan and G. Gbur, "Spatial coherence conversion with surface plamons using a three-slit interferometer," Plasmonics **3**, 111 (2008).

- D. Moses, C. H. Gan and G. Gbur, "Directional, nonpropagating, and polychromatic excitations in one-dimensional wave systems," Phys. Rev. E **79**, 026606 (2009).