Nonreciprocal and Nonlinear Nanoplasmonic Devices

by

Curtis James Firby

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Photonics and Plasmas

Department of Electrical and Computer Engineering

University of Alberta

Abstract

This thesis explores the incorporation of nonreciprocal magneto-optic effects, as well as nonlinear phenomena into integrated nanoplasmonic devices, and encompasses a wide spectrum of theoretical, numerical, fabrication, and characterization methodologies.

Incorporating magnetic dielectrics into conventional nanoplasmonic waveguide geometries produces unique magnetoplasmonic architectures capable of fulfilling novel nonreciprocal functions. Application of a transverse magnetization across such devices leads to a nonreciprocal phase shift of the propagating mode, while a longitudinal magnetization can induce nonreciprocal polarization rotation via the Faraday effect. Additionally, the magnetization within the material exhibits highly nonlinear temporal dynamics when stimulated by external magnetic fields. These dynamics can be mapped onto the polarization or phase of the guided mode, or in an interferometer configuration, onto the output intensity.

These concepts are employed to demonstrate several novel magnetoplasmonic devices including:

- A high-speed phase shifter capable of up to 0.33rad of phase modulation and encoding data onto the phase of the mode.
- An optical isolator exhibiting 2.51dB insertion loss and a 22.82dB isolation ratio.

- A nanoplasmonic analog to an electrical clock multiplier, with modulation depths up to 16.26dB, a tunable output frequency between 280MHz and 5.6GHz, and multiplication factors of up to 2.5×10^3.

- A nonlinear frequency mixer, which can generate harmonics, frequency splitting, and frequency down-conversion of a driving signal, as well as frequency mixing when two driving signals are present.

- A dynamic plasmonic polarization modulator providing 99.4% mode conversion within only 830μm and operable in either a pulsed-output mode, or a continuous-output mode at up to 10GHz. This is the first demonstration of the Faraday effect within a plasmonic waveguide.

- A versatile, broadband, and polarization-independent Faraday effect optical circulator.

Furthermore, extensive fabrication process development was undertaken to construct a unique silicon-on-insulator waveguide characterization platform, which simultaneously incorporates micromachining of a bulk substrate and nanopatterning of Si-based metal-encapsulated plasmonic waveguides. This process allows nanoplasmonic waveguides only a few microns in length to be directly excited by ultrafast laser pulses via end-fire excitation.

Experimental characterization of these waveguides depicts a complex interplay between the numerous nonlinear processes occurring within silicon. Notably, the infrared exciting pulses are converted into bright green visible light via third-harmonic generation. The conversion efficiency is measured to be 4.9×10^{-4}, which is over an order of magnitude higher than any previously reported result in a silicon-based waveguide structure, and was achieved within a device 100nm wide, 340nm tall, and 2.4μm long, yielding a compact on-chip footprint of only 0.24μm^2.

Further examination of the measured spectra and their power scaling trends, as well as comparison to simulated results, suggests that an additional electric field enhancement of up to 2.2 times is present, due to the granular nature of the Au film. This leads to consequential nonlinear absorption via mechanisms such as two-photon absorption and free-carrier absorption within the silicon. Moreover, evidence of the onset of self-phase modulation is present. All measured results are found to be in good agreement with theoretical predictions.

The findings presented in this thesis open new pathways for development of nonreciprocal and nonlinear nanoplasmonic devices, which are of fundamental importance to the future development of chip-scale integrated nanoplasmonic networks.

Preface

Parts of Chapter 2 were published in the invited review article as S. Sederberg, C. J. Firby, S. R. Greig, and A. Y. Elezzabi, "Integrated nanoplasmonic waveguides for magnetic, nonlinear, and strong-field devices, *Nanophotonics*, **6**(1), 235-257 (2017). For this review article, all authors co-wrote the manuscript, and contributed to editing, figure preparation, and revisions.

Chapter 3 presents work published in the following four journal articles:

[1] C. J. Firby and A. Y. Elezzabi, "High-speed nonreciprocal magnetoplasmonic waveguide phase shifter," *Optica* **2**(7), 598-606 (2015).

[2] C. J. Firby and A. Y. Elezzabi, "Magnetoplasmonic isolators utilizing the nonreciprocal phase shift," *Opt. Lett.* **41**(3), 563-566 (2016).

[3] C. J. Firby and A. Y. Elezzabi, "A magnetoplasmonic electrical-to-optical clock multiplier," *Appl. Phys. Lett.* **108**(5), 051111 (2016).

[4] C. J. Firby and A. Y. Elezzabi, "Magnetoplasmonic RF mixing and nonlinear frequency generation," *Appl. Phys. Lett.* **109**(1), 011101 (2016).

For each article of the above listed articles, I was responsible for the numerical simulations, data analysis, and figure preparations, as well as writing, editing, and revising the manuscripts. A. Y. Elezzabi supervised the project and assisted with the manuscript preparations and revisions.

Chapter 4 describes work that has been published in the following two journal articles:

[1] C. J. Firby, P. Chang, A. S. Helmy, and A. Y. Elezzabi, "Magnetoplasmonic Faraday rotators: enabling gigahertz active polarization control for integrated plasmonics," *ACS Photonics* **3**(12), 2344-2352 (2016).

[2] C. J. Firby, P. Chang, A. S. Helmy, and A. Y. Elezzabi, "Versatile broadband polarization-independent optical circulators for nanophotonic integrated circuits," *J. Opt. Soc. Am. B* **35**(7), 1504-1513 (2018).

I was responsible for the numerical simulations, data analysis, figure preparations, and writing the initial manuscript drafts. P. Chang assisted with the waveguide design and structural optimization for both papers. A. S. Helmy and A. Y. Elezzabi supervised the project. All authors contributed to editing and revising the manuscripts.

Life is too short not to create something
with every breath we draw.

—Maynard James Keenan

Acknowledgements

First and foremost, I would like to thank my PhD supervisor, Prof. Abdul Elezzabi. Without his guidance, insight, pure enthusiasm for science, and faith in me, none of this thesis work would have been possible. His encouragement was invaluable to keep me motivated when the research became difficult. The skills I have acquired and the lessons I have learned while working in his lab will stay with me forever, and have shaped me into the researcher I am today. I cannot thank him enough for taking a chance and hiring me as an undergraduate summer student all those years ago, and for all of his mentorship since.

I would also like to sincerely thank all of my colleagues and lab-mates throughout the years: Shawn Greig, Brett Carnio, Nir Katchinskiy, Katherine Smith, Taylor Robertson, Aiden McDermott, Ryan Boehnke, Liam McRae, Kevan Bell, Eric Hopmann, Amir Badkoobehhezaveh, Wu Zhang, and Dr. Haizeng Li. Whether discussing experiments within the lab or over a drink after work, your friendships have been invaluable and my PhD experience would not have been the same without them. I wish you all the best, and I wholeheartedly hope that we might work together again someday.

Additionally, I'd like to thank Dr. Shawn Sederberg for taking the time to answer my emails and provide advice and guidance on my fabrication and experiments, as well as for laying the foundation in the lab upon which my research is built. I'd also like to extend my gratitude to our collaborators at the University of Toronto,

Prof. Amr Helmy and PoHan Chang. With their assistance and unique approach to problem solving, together we were able to realize things that I had thought physically impossible.

Assistance from the technical support staff in the Department of Electrical and Computer Engineering was essential to the work completed in this thesis. Specifically, I'd like to thank Peng Li, Stephanie Bozic, Scott Munro, Les Schowalter, Aditi Ganji, and Dr. Aaron Hryciw of the nanoFAB. Without their training and process advice, none of my fabrication would have worked. Furthermore, I'd like to thank the machine shop staff, Herbert Dexel, Reiner Schwarze, and Terry Kugler, for always being able to translate my terrible sketches into immaculately machined components.

This PhD would not have been possible without the endless love and support of my friends and family. Thank you all for your words of encouragement and optimism, and for being there for me throughout this experience. None of this would have been possible without you all.

Last but certainly not least, I extend my deepest thanks and infinite appreciation to my wonderful girlfriend, Sarah. Her boundless love and support during this experience helped keep me grounded and on track. Whether it was listening to my frustrations and problems, celebrating my successes, or just expressing genuine interest and curiosity towards my work, she was always there for me. She always knew when to leave me to my studies, and when to pull me away to decompress and relax. Without her, I could not have done any of this.

Contents

List of Tables

List of Figures

List of Fundamental Constants and Abbreviations

Fundamental Constants

$\hbar = 1.055 \times 10^{-34}$J·s	reduced Planck constant
$\mu_0 = 4\pi \times 10^{-7}$H/m	permeability of free space
$\mu_B = 9.274 \times 10^{-24}$J/T	Bohr magneton
$\varepsilon_0 = 8.854 \times 10^{-12}$F/m	permittivity of free space
$c = 2.998 \times 10^{8}$m/s	speed of light in free space
$h = 6.626 \times 10^{-34}$J·s	Planck's constant
$m_e = 9.109 \times 10^{-31}$kg	mass of an electron
$q = 1.602 \times 10^{-19}$C	elementary charge

Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
AM	antisymmetric mode
Bi:YIG	bismuth-substituted yttrium-iron-garnet
BOE	buffered oxide etch
BOX	buried oxide
BPSK	binary phase-shift keying
Ce:YIG	cerium-substituted yttrium-iron-garnet
CM	clock multiplier

CMOS	complimentary metal-oxide-semiconductor
CVD	chemical vapour deposition
CW	continuous-wave
DC	direct-current
DI	deionized
DL	dielectric loaded
DRIE	deep reactive ion etching
EBL	electron beam lithography
EFISHG	electric-field-induced second-harmonic generation
FCA	free-carrier absorption
FDTD	finite-difference-time-domain
FIB	focused ion beam
FMR	ferromagnetic resonance
FWHM	full-width-half-maximum
GGG	gadolinium gallium garnet
HIM	helium ion microscope
HWP	half-wave plate
ICPRIE	inductively coupled plasma reactive ion etch
IMI	insulator-metal-insulator
IPA	isopropyl alcohol
IR	infrared
LLG	Landau-Lifshitz-Gilbert
LRDLSPW	long-range dielectric loaded surface plasmon waveguide
LRSP	long-range surface plasmon
ME	metal-encapsulated
MEMS	micro-electro-mechanical systems
MI	metal-insulator
MIBK	methyl isobutyl ketone
MIM	metal-insulator-metal
MO	magneto-optic
MZI	Mach-Zehnder interferometer
NA	numerical aperture
ND	neutral-density

NRPS	nonreciprocal phase shift
OWPS	one-way phase shift
PECVD	plasma enhanced chemical vapour deposition
PICO	pulsed-input continuous-output
PIPO	pulsed-input pulsed-output
PLD	pulsed laser deposition
PMMA	poly(methyl methacrylate)
POP	plane of precession
RF	radio frequency
RIE	reactive ion etch
RPM	revolutions per minute
SEM	scanning electron microscope
SHG	second harmonic generation
SM	symmetric mode
SOI	silicon-on-insulator
SP	surface plasmon
SPM	self-phase modulation
TE	transverse-electric
THG	third-harmonic generation
TIR	total internal reflection
TL	transmission line
TM	transverse-magnetic
TPA	two-photon absorption
UV	ultraviolet
WDM	wavelength-division-multiplexing
YIG	yttrium-iron-garnet

Chapter 1

Introduction

1.1 Overview

Conventional silicon complimentary metal-oxide-semiconductor (CMOS) electronics are rapidly approaching their fundamental size and speed limits as transistors are continually reduced in size. However, society demands increases in processing power at a much faster rate. Another paradigm is required. Processing signals with light is one such solution that provides orders of magnitude increases in speed and bandwidth.

At the chip scale, light is typically manipulated on the silicon (Si) photonics platform. This solution is inexpensive, as Si is readily available, and can take advantage of fabrication techniques and processes that have been pioneered and meticulously refined over decades by the microelectronics industry. At present, silicon photonic devices are widely used as interconnects between CMOS processors and fiber optic networks for data transmission.

However, realizing a computing platform on silicon photonics is challenging, in part due to the inherent diffraction limit, and the inability of Si to emit light. This forces silicon photonic components to dimensions an order of magnitude larger than transistors, inhibits dense integration, and requires the incorporation of other semi-

conductor lasers. The field of plasmonics has been hailed as a potential successor that can merge the benefits of both platforms, and to this end, has been at the forefront of optical research for the last several decades [1–7].

Plasmonic devices work on the principle of coupling light to the free conduction band electrons within a metal, and exciting oscillations within the charge density at the boundary between the metal and the surrounding dielectric medium. These coherent charge oscillations, which can be either propagating or localized, are highly sensitive to changes in their constituent materials. This property makes them ideal for a variety of sensor and modulator platforms. Moreover, the coupling of an electromagnetic wave to a metal-dielectric interface allows light to be confined to areas below the diffraction limit of the free space radiation, making such structures ideal for dense, on-chip integration. Encoding signals into these confined lightwaves maintains the improved speeds and bandwidth of their photonic counterparts in a significantly compressed form factor, resulting in a platform with the potential to compete with conventional CMOS electronics.

To reach the end goal of developing optical data processing platforms within plasmonic nanocircuitry, components analogous to those of traditional electronic circuits must be developed [8]. Significant work has been done on the development of integrated plasmonic devices, such as waveguides [9], resonators [10], routers [11], lasers and optical sources [12], modulators [13], and even simple logic gates [14]. In order to truly unlock the potential of nanoplasmonic circuitry though, mechanisms of introducing nonreciprocity and nonlinearity within an integrated platform must be thoroughly developed and incorporated.

In optics, reciprocity embodies the idea that one can interchange the positions of a source and a detector within an optical system, and still measure the same fraction of the input power [15]. Introducing nonreciprocity implies that two counter-propagating waves within a structure interact with the medium in a slightly different

manner. This property facilitates the control of light flow, analogous to the operation of electronic semiconductor diodes. It grants the ability to route signals, as well as provides the potential for unique modulation schemes not present in any other material system. This phenomenon is most commonly introduced through magneto-optic effects such as the Faraday effect, the nonreciprocal phase shift, and the magneto-optic Kerr effects, whereby an applied magnetic field is employed to break the time reversal symmetry and induce nonreciprocity. Incorporating such effects into plasmonic structures is the focus of the field of magnetoplasmonics [16, 17]. To date, much of the work in this area has been focused on studying the interplay between plasmonic field enhancement and magneto-optic effects [16, 17]. There have been few reports of magnetoplasmonic waveguide based devices exhibiting these properties suitable for incorporation into chip scale architectures.

Nonlinearity is also crucial within integrated optical systems. Exploitation of the electric field enhancement offered by plasmonic nanostructures has led to many studies of nonlinear effects within such devices [18, 19]. These effects can generate new frequencies of light on chip, as well as provide all-optical modulation of signals propagating through a device. However, one of the fundamental challenges in implementing these devices is the requirement for phase matching, and, in the case of nonlinear light generation, consideration must be made for the absorption of both the fundamental and the generated light. Due to the high optical losses experienced optical signals propagating within plasmonic waveguide structures, this can be difficult to realize.

In order to practically realize nanoplasmonic circuitry, and bring this budding technology to fruition, one requires practical and integrable nonreciprocal and nonlinear components. Just as in electronics, where nonlinear and nonreciprocal devices such as diodes and transistors form the foundation of our computational architecture, so too shall these components form the basis of future optical computing strategies within the nanoplasmonic regime.

1.2 Thesis Objectives

The aim of the work presented within this thesis is to theoretically and experimentally demonstrate the incorporation of nonreciprocity and nonlinearity within a nanoscale plasmonic platform to introduce new functionality within chip-scale integrated optics. The primary goals of this research are:

1. Acquire a fundamental understanding of how magneto-optics and the unique temporal dynamics of magnetization can be utilized within integrated plasmonic architectures.

2. Exploit magneto-optic effects to exert control over the phase, polarization, and intensity of guided modes.

3. Design integrated optical analogs of conventional nonreciprocal components, such as isolators and circulators.

4. Apply this knowledge to develop novel plasmonic modulator geometries that demonstrate new mechanisms of modulation, and can satisfy an unfulfilled niche of applications.

5. Demonstrate the Faraday effect within a plasmonic waveguide for the first time.

6. Develop a flexible fabrication process capable of producing Si-based plasmonic waveguides on a platform that facilitates direct excitation and characterization.

7. Characterize and evaluate the potential for nonlinear light generation within metal-encapsulated plasmonic waveguides on a silicon-on-insulator platform.

8. Examine the complex interplay between the numerous nonlinear phenomena excited in nanoplasmonic waveguides by ultrafast laser pulses.

9. Demonstrate enhanced nonlinear visible light emission efficiency from Si-based nanoplasmonic waveguides.

1.3 Thesis Organization

The seven chapters presented within this thesis encompass an array of fundamental investigations in the design and construction of novel nanoplasmonic devices employing nonreciprocal and nonlinear effects. These studies combine theoretical investigations, computational modeling, fabrication process development, as well as experimental characterization. This thesis is organized as follows:

Chapter 2 presents the broad spectrum of background and theory required for the work contained in the following chapters. This includes a summary of classical electromagnetics, and proceeding on to a description of surface plasmons, as well as various waveguiding geometries to confine and control them. An overview of magneto-optic effects and their physical underpinnings, as well as a derivation of the basic magnetization dynamics relations follows. Finally, this chapter presents a description of the nonlinear phenomena that can be observed within Si-based structures.

Chapter 3 outlines the process of examining the magneto-optic effect known as the nonreciprocal phase shift to manipulate the phase of a guided mode. The first section introduces the concept of dynamic phase modulation with a simple plasmonic waveguide structure. Noting the limitations of the architecture, the second section improves on the design of the waveguide geometry to implement a fundamental integrated optics component: the optical isolator. The developed architecture is applied to exploit the novel dynamic properties of magnetic materials in order to design and assess two completely new types of nanoplasmonic modulators, which have no current photonic analog.

With the foundations of magneto-optic phase manipulation and magnetization dynamics established, Chapter 4 sets to apply this knowledge to dynamically control another property of light: its polarization state. The first section of this chapter details the design of a plasmonic waveguide Faraday rotator, a device which has never before been achieved within the plasmonic regime. Following the demonstration of the

high-speed dynamic polarization modulating properties of this structure, these concepts are extended to develop another fundamental component for optical networks: an optical circulator.

At this point, with the thorough understanding of optical waveguide and modulator engineering, the thesis focus is complemented by solving another standing problem within the field of integrated nanoplasmonics. Namely, efficient generation of light within a Si platform. To achieve this feat, unique structures had to be developed, and correspondingly a novel combination of micro- and nanofabrication processes had to be applied for such ends. Chapter 5 presents the detailed fabrication process development that was undertaken to construct Si-based metal-encapsulated nanoplasmonic waveguides on a micromachined characterization beam, such that their nonlinear properties could be studied.

Chapter 6 details the results of the experiments performed on the fabricated devices. It is shown that a complex interaction of various nonlinear effects manifests within these nanoscale structures. The waveguides are shown to be capable of highly efficient light emission via third-harmonic generation. Correspondingly, these structures have fundamental applications in on-chip plasmonic circuitry and integrated optical networks, as light generation with a silicon platform is difficult to accomplish, and conventionally other material systems must be incorporated to achieve these ends.

Finally, Chapter 7 summarizes the investigations and key results obtained throughout the work presented within this thesis, before discussing the potential studies that could be undertaken in the continuation of this strain of work.

Chapter 2

Background

The results presented within this thesis examine the design, simulation, fabrication, and characterization of a variety of novel active nanoplasmonic devices employing nonreciprocal magneto-optic and nonlinear effects to manipulate the properties light. This chapter presents the necessary background information for the simulations and experiments presented throughout the rest of the thesis. The chapter begins with a brief discussion of Maxwell's Equations, an analysis of wave propagation within a medium, and the interaction of electromagnetic waves with metals. The next section provides an introduction to surface plasmons at planar interfaces, as well as various waveguiding geometries. Subsequently, the following section introduces the origins and fundamentals of magneto-optic effects within various media in both reflection and transmission geometries, before a brief discussion of magneto-optic effects within plasmonic waveguides, and analysis of magnetization dynamics that can be utilized to modulate these effects. Finally, the next sections illustrate nonlinear effects relevant to silicon plasmonic devices, as well as the free carrier effects that can be induced by passing intense laser pulses through nanoscale waveguides.

2.1 Classical Electromagnetics

2.1.1 Maxwell's equations

Only four equations, called Maxwell's Equations, are required to completely define and analyze any classical electromagnetics problem [20–23]:

$$\nabla \cdot \mathbf{D} = \rho \tag{2.1}$$

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

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{2.4}$$

In this set of equations, $\mathbf{E}$ is the electric field, $\mathbf{D}$ is the electric flux density, $\mathbf{H}$ is the magnetic field, $\mathbf{B}$ is the magnetic flux density, $\mathbf{J}$ is the free current density, and ρ is the free charge density within the medium. Equations 2.1-2.4 are denoted as Gauss' Law, Gauss' Law for Magnetism, Faraday's Law of Induction, and Ampère's Law, respectively. The relationship between each field and its corresponding flux density are given by the following constituent relations:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{2.5}$$

$$\mathbf{B} = \mu_0 \mathbf{H} + \mu_0 \mathbf{M} \tag{2.6}$$

where $\mathbf{P}$ is the polarization of the material, $\mathbf{M}$ is the magnetization of the material, $\varepsilon_0 = 8.854 \times 10^{-12}$F/m is the permittivity of free space, and $\mu_0 = 4\pi \times 10^{-7}$H/m is the permeability of free space. In a linear and isotropic medium, the polarization and

magnetization are related to the fundamental fields via:

$$\mathbf{P} = \varepsilon_0 \chi_e \mathbf{E} \tag{2.7}$$

$$\mathbf{M} = \chi_m \mathbf{H} \tag{2.8}$$

where χ_e is called the electric susceptibility and χ_m is the magnetic susceptibility of the medium. Substituting Eqn. 2.7 into Eqn. 2.5, and Eqn. 2.8 into Eqn. 2.6, one can derive simplified relationships for the flux densities in terms of the fundamental fields:

$$\mathbf{D} = \varepsilon \mathbf{E} = \varepsilon_0 \varepsilon_r \mathbf{E} \tag{2.9}$$

$$\mathbf{B} = \mu_0 \mathbf{H} = \mu_0 \mu_r \mathbf{H} \tag{2.10}$$

where $\varepsilon = \varepsilon_0 \varepsilon_r$ is the total permittivity, $\mu = \mu_0 \mu_r$ is the total permeability, and ε_r and μ_r are the relative permittivity and relative permeability of the medium, respectively. From the aforementioned substitution, these useful material quantities are related to the susceptibilities via:

$$\varepsilon_r = 1 + \chi_e \tag{2.11}$$

$$\mu_r = 1 + \chi_m \tag{2.12}$$

At the optical and near infrared frequencies of interest for the work within this thesis, $\mu_r = 1$, and therefore, $\mu = \mu_0$, due to the fact that the temporal evolution of magnetization is inherently a very slow process, typically on the order of hundreds of picoseconds to nanoseconds. At the telecommunication wavelength of $\lambda = 1550$nm, the radiation oscillates with a frequency of 193.4THz, or correspondingly, a period of about 5.2fs. No material's magnetization can respond to radiation at this range of frequencies. In all optical derivations onwards within this thesis, these equivalencies

are assumed.

In general, the material properties are wavelength dependent, due to the electronic structure of the material within which the wave is propagating. This in turn leads to dispersion and absorption. The properties have been assumed scalar in the above derivations because of the assumption of a linear and isotropic medium. However, crystal structure, or the introduction of a magnetic field within a magneto-optic material, can introduce anisotropy. In such cases, as will be discussed in a subsequent section and in Chapters 3-4, properties such as the permittivity must be considered as a full second-order tensor. Additionally, Eqn. 2.7 assumed a linear relationship between the electric fields and the polarization. This assumption can break down at sufficiently high electric field strength (above approximately 10^6V/m [24]), leading to what is known as nonlinear effects. This behaviour will be discussed shortly, and is utilized within the results of Chapter 6.

2.1.2 The Wave Equation

Examining Maxwell's Equations (Eqns. 2.1-2.4), it is evident that Eqn. 2.3 and 2.4 describe the temporal variations of the electric and magnetic flux densities and fields. Thus, one can derive the relation for the propagation of electric and magnetic fields within a medium [20–23]. First, a linear and isotropic medium is assumed, with no free charges ($\rho = 0$) or currents ($\mathbf{J} = 0$). Taking the curl of Eqn. 2.3 yields the following:

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times (-\frac{\partial \mathbf{B}}{\partial t}) \tag{2.13}$$

Applying the vector identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$, where $\mathbf{A}$ is an arbitrary vector, yields:

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) \tag{2.14}$$

Since we are considering a linear, isotropic, and charge-free medium, Eqn. 2.1 implies that $\nabla \cdot \mathbf{E} = 0$. With the substitution of Eqn. 2.10, one can show that $\nabla \times \mathbf{B} = \mu_0(\nabla \times \mathbf{H})$. Thus:

$$\nabla^2\mathbf{E} = \mu_0 \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) \tag{2.15}$$

Substituting in Eqn. 2.4, followed by Eqn. 2.9 gives:

$$\nabla^2\mathbf{E} = \mu_0 \frac{\partial^2 \mathbf{D}}{\partial t^2} \tag{2.16}$$

$$\nabla^2\mathbf{E} = \varepsilon_0 \varepsilon_r \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.17}$$

This can be further simplified by taking $c = 1/\sqrt{\mu_0 \varepsilon_0}$, where $c = 2.998 \times 10^8$m/s is the speed of light in free space. The refractive index of the medium, n, can be defined as $n = \sqrt{\varepsilon_r}$. Therefore, the propagation of the electric field in a medium is given by:

$$\nabla^2\mathbf{E} = \frac{n^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{2.18}$$

Similar analysis starting from Eqn. 2.4 yields the analogous result for the magnetic fields:

$$\nabla^2\mathbf{H} = \frac{n^2}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{2.19}$$

These two equations are known as the wave equations for the electric and magnetic fields, respectively. They describe the spatial and temporal evolution of the fields.

2.1.3 Response of Materials to Electromagnetic Fields

In order to understand how light can be manipulated within nanoscale structures, it is essential to discuss the response of materials to electromagnetic fields. Consider first a linear, isotropic dielectric material in one dimension. In such a material, the negatively charged valence electrons exist in a cloud bound to the positively charged

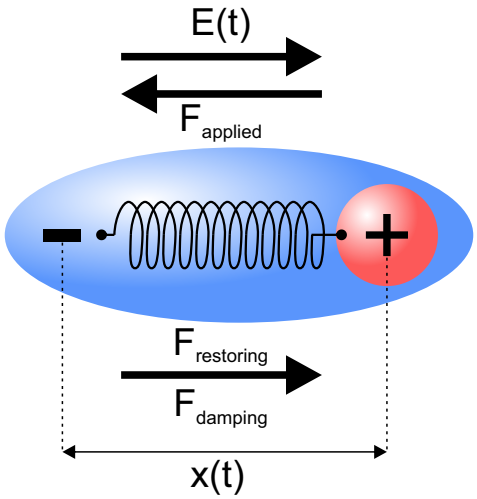

Figure 2.1: Schematic illustration of the Lorentz Oscillator model for material constants. An applied electric field, $E(t)$, exerts a force, $F_{applied}$, on the electron cloud, displacing it from the nucleus. A restoring force, $F_{restoring}$, and a damping force, $F_{damping}$ act against $F_{applied}$. The resulting displacement of the electron cloud is denoted by $x(t)$. This model is analogous to a mass-spring damped harmonic oscillator.

atomic nucleus. Application of an electric field, $E(t)$, will exert a force, $F_{applied}$, on the electrons, displacing them from the nucleus, which is considered fixed. This situation is depicted schematically in Fig. 2.1. Due to the attraction of the electrons to the protons in the nucleus, there exists a restoring force, $F_{restoring}$, acting to return them to their equilibrium position, and there exists a damping force, $F_{damping}$, opposing to the motion due to collisions. The interaction of this system is analogous to that of a mass-spring damped harmonic oscillator, and can be modeled as such [21–23]. The balance of forces for such a scenario in one dimension yields:

$$F_{total} = F_{applied} + F_{restoring} + F_{damping} \tag{2.20}$$

$$m_e \frac{\partial^2 x}{\partial t^2} = -qE(t) - k_{spring}x - m_e\gamma\frac{\partial x}{\partial t} \tag{2.21}$$

$$-\frac{qE(t)}{m_e} = \frac{\partial^2 x}{\partial t^2} + \gamma\frac{\partial x}{\partial t} + \omega_0^2 x \tag{2.22}$$

Here, F_{total} is the total force acting on the electrons, $F_{applied}$ is the force exerted by the applied field, $F_{restoring}$ is the restoring force, $F_{damping}$ is the damping force, $x(t)$ is the displacement, k_{spring} is the spring constant of the restoring force, γ is the damping

coefficient, $m_e = 9.109 \times 10^{-31}$kg is the mass of an electron, and $q = 1.602 \times 10^{-19}$C is the magnitude of the electronic charge. The natural resonant frequency of the oscillator is defined as $\omega_0 = \sqrt{k_{spring}/m_e}$. With a harmonic excitation at a frequency ω, $E(t) = E_0 e^{-i\omega t}$, and assuming the amplitude E_0 is low enough that the electron displacement is proportional to $E(t)$ (i.e. the linear regime), the displacement will take on the form $x(t) = x_0 e^{-i\omega t}$, where x_0 is the displacement amplitude, t is the time, and i is the imaginary unit. Applying this substitution and simplifying yields:

$$x_0(\omega) = -\frac{q}{m_e}\left(\frac{1}{w_0^2 - i\omega\gamma - \omega^2}\right)E_0 \tag{2.23}$$

The dipole moment of a single atom, p_a, is defined by $p_a = -qx_0$, and therefore, the total polarization of a medium containing N atoms is given by $P = Np_a = -Nqx_0$. This implies that:

$$P(\omega) = \frac{Nq^2}{m_e}\left(\frac{1}{w_0^2 - i\omega\gamma - \omega^2}\right)E_0 \tag{2.24}$$

In general, a material can have numerous resonances due to a complex electronic structure. The previously assumed linearity of the material implies that the total response is simply the sum of the contributions of each individual resonance:

$$P(\omega) = \frac{Nq^2}{m_e}\sum_j \frac{f_j}{w_{0,j}^2 - i\omega\gamma_j - \omega^2}E_0 \tag{2.25}$$

where $\omega_{0,j}$ and γ_j are the resonant frequency and damping coefficient associated with the j^{th} resonance, and f_i is a weighting factor indicating the strength of the resonance (oscillator strength). Comparing Eqn. 2.25 to Eqn. 2.7, it is obvious that:

$$\chi_e(\omega) = \frac{Nq^2}{\varepsilon_0 m_e}\sum_j \frac{f_j}{w_{0,j}^2 - i\omega\gamma_j - \omega^2} \tag{2.26}$$

We can define a new useful quantity, ω_p, as the plasma frequency, where $\omega_p = \sqrt{\frac{Nq^2}{\varepsilon_0 m_e}}$. This quantity defines the natural collective oscillation frequency of the electrons. Substituting this and Eqn. 2.26 into Eqn. 2.11 allows us to determine the frequency-dependent relative permittivity of the material:

$$\varepsilon_r(\omega) = 1 + \sum_j \frac{f_j \omega_p^2}{w_{0,j}^2 - i\omega\gamma_j - \omega^2} \tag{2.27}$$

Equation 2.27 represents the Lorentz Oscillator model of a dielectric material's optical properties.

Notably, special consideration must be given to the case of a metal, as metals are a fundamental component in plasmonics. In metals, electromagnetic interactions are dominated by a sea of unbound conduction band electrons, as opposed to bound valence band electrons. A similar derivation can be applied, and the unbound nature of these electrons allows us to neglect the restoring force term, $F_{restoring}$, in Eqn. 2.20. Repeating the analysis yields:

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\gamma} \tag{2.28}$$

This model for the optical properties of metals is denoted as the Drude model.

2.2 Surface Plasmons

Surface plasmon (SP) waves are coherent, collective oscillations of charge density at the boundary between a metal and a dielectric. These excitations can occur when an incident photon couples to the free conduction band electrons at the surface of the metal [1]. Naturally, this SP wave is a bound surface wave, and the electromagnetic fields evanescently decay away from the interface, into the constituent media.

This surface confinement differentiates SPs from bulk, or volume plasmons. Bulk

plasmons are the coherent, collective oscillations of the entire electron gas throughout the bulk metallic film, in relation to the atomic nuclei of the material's crystal structure. While typical SP waves propagate along the metal-dielectric interface, there is a subset of SPs called localized SPs. Localized SPs are dipole-like oscillatory resonances over a subwavelength structure such as a nanoparticle or a nanoantenna. Accordingly, these oscillations are localized to the structure itself, and do not propagate. This thesis addresses propagating SPs, as such waves can be guided and manipulated in waveguide structures and are the most applicable to integrated optics applications.

Due to the surface bound nature of propagating SP waves, the electromagnetic energy can be confined to an area much smaller than the diffraction limit of light, giving these systems obvious advantages over traditional diffraction limited photonic devices. In this way, plasmonic components have the potential to realize high density optical signal processing on scales approaching those of modern CMOS technology.

2.2.1 1D Surface Plasmon Wave

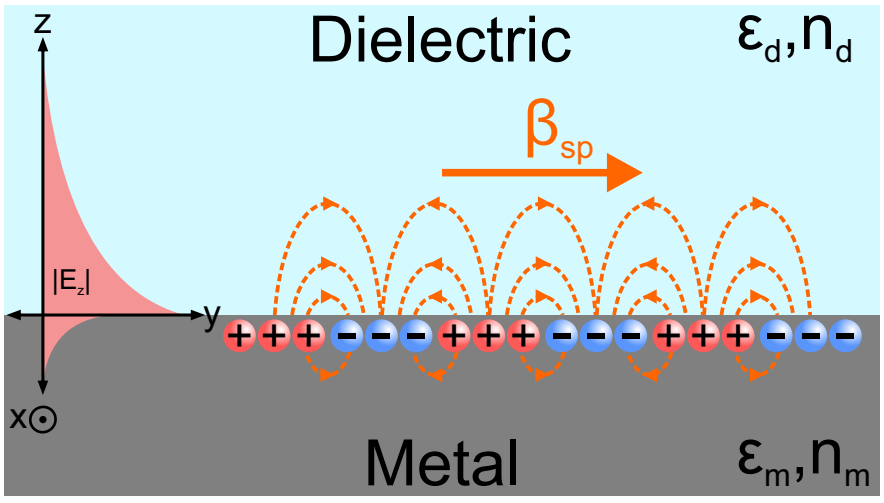

Figure 2.2: Illustration of a 1D surface plasmon wave at the interface between a dielectric, having a relative permittivity, ε_d, and a refractive index, n_d, and a metal, having a relative permittivity, ε_m, and refractive index, n_m. The bound surface wave is characterized by the coupling of light to oscillations in the free charge density of the metal, and as such, the electric field of the wave decays away from the interface evanescently.

To understand the surface plasmon wave in its most basic form, the simplest geometry can first be considered: a one-dimensional (1D) interface between a metal and a linear, isotropic dielectric [1]. In this configuration, depicted in Fig. 2.2, a metal-dielectric interface is located at $z = 0$; the half-space where $z > 0$ is filled with a dielectric having a permittivity ε_d, while the $z < 0$ half-space consists of a metal having a permittivity ε_m. The electric and magnetic fields are assumed to be of the general form $\mathbf{E}(y,z,t) = \langle E_x(y,z,t), E_y(y,z,t), E_z(y,z,t)\rangle$, and $\mathbf{H}(y,z,t) = \langle H_x(y,z,t), H_y(y,z,t), H_z(y,z,t)\rangle$, where E_x, E_y, and E_z are the x-, y-, and z-components of the electric field, respectively, and H_x, H_y, and H_z are the x-, y-, and z-components of the magnetic field, respectively. The fields are further distinguished by region, where subscript d indicates fields in the dielectric, while subscript m indicates fields in the metal. Equations for the temporal and spatial variation of the fields can be written as follows:

$$E_x(y,z,t) = \begin{cases} E_{x,d} e^{i(\beta_{SP} y - \omega t)} e^{-\alpha_d z}, & z > 0, \\ E_{x,m} e^{i(\beta_{SP} y - \omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.29}$$

$$E_y(y,z,t) = \begin{cases} E_{y,d} e^{i(\beta_{SP} y - \omega t)} e^{-\alpha_d z}, & z > 0, \\ E_{y,m} e^{i(\beta_{SP} y - \omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.30}$$

$$E_z(y,z,t) = \begin{cases} E_{z,d} e^{i(\beta_{SP} y - \omega t)} e^{-\alpha_d z}, & z > 0, \\ E_{z,m} e^{i(\beta_{SP} y - \omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.31}$$

$$H_x(y,z,t) = \begin{cases} H_{x,d}e^{i(\beta_{SP}y-\omega t)}e^{-\alpha_d z}, & z > 0, \\ H_{x,m}e^{i(\beta_{SP}y-\omega t)}e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.32}$$

$$H_y(y,z,t) = \begin{cases} H_{y,d}e^{i(\beta_{SP}y-\omega t)}e^{-\alpha_d z}, & z > 0, \\ H_{y,m}e^{i(\beta_{SP}y-\omega t)}e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.33}$$

$$H_z(y,z,t) = \begin{cases} H_{z,d}e^{i(\beta_{SP}y-\omega t)}e^{-\alpha_d z}, & z > 0, \\ H_{z,m}e^{i(\beta_{SP}y-\omega t)}e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.34}$$

In Eqns. 2.29-2.34, the surface plasmon is assumed to propagate in the $+y$-direction with a propagation constant of β_{SP}, and time harmonic oscillations at an angular frequency of ω. Since the SP wave is a bound surface wave, the fields must evanescently decay away from the interface. The fields in the dielectric are assumed to have a decay constant of α_d and the fields in the metal exhibit a decay constant of α_m, where both $\alpha_d > 0$ and $\alpha_m > 0$ must hold to ensure exponential decay of the fields.

Applying the standard electromagnetic boundary conditions, where the tangential magnetic fields and normal magnetic flux density must be continuous at the $z = 0$ interface, as well as enforcing Gauss' Law for Magnetism (Eqn. 2.2), it is evident that $H_{y,d} = H_{y,m} = H_{z,d} = H_{z,m} = 0$, and $H_{x,d} = H_{x,m} = C_1$, where C_1 is a normalization constant defining the amplitude of the wave. Application of Ampere's Law (Eqn. 2.4) to the above fields yields the conclusion that $E_{x,d} = E_{x,m} = 0$. Thus, the only non-zero field components are $E_y(y,z,t)$, $E_z(y,z,t)$, and $H_x(y,z,t)$. This confirms that the surface plasmon wave is a transverse-magnetic (TM) wave, and that surface plasmon cannot exist for transverse-electric (TE) polarization within this 1D slab configuration. Furthermore, the remaining electric field amplitudes can be written in

terms of the constant C_1 as follows:

$$E_{y,d} = -C_1 \frac{i\alpha_d}{\omega\varepsilon_0\varepsilon_d} \tag{2.35}$$

$$E_{y,m} = C_1 \frac{i\alpha_m}{\omega\varepsilon_0\varepsilon_m} \tag{2.36}$$

$$E_{z,d} = C_1 \frac{\beta_{SP}}{\omega\varepsilon_0\varepsilon_d} \tag{2.37}$$

$$E_{z,m} = C_1 \frac{\beta_{SP}}{\omega\varepsilon_0\varepsilon_m} \tag{2.38}$$

With this knowledge, the electric and magnetic field components of the 1D surface plasmon wave can be written as follows:

$$E_x(y,z,t) = 0 \tag{2.39}$$

$$E_y(y,z,t) = \begin{cases} -C_1 \frac{i\alpha_d}{\omega\varepsilon_0\varepsilon_d} e^{i(\beta_{SP}y-\omega t)} e^{-\alpha_d z}, & z > 0, \\ C_1 \frac{i\alpha_m}{\omega\varepsilon_0\varepsilon_m} e^{i(\beta_{SP}y-\omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.40}$$

$$E_z(y,z,t) = \begin{cases} C_1 \frac{\beta_{SP}}{\omega\varepsilon_0\varepsilon_d} e^{i(\beta_{SP}y-\omega t)} e^{-\alpha_d z}, & z > 0, \\ C_1 \frac{\beta_{SP}}{\omega\varepsilon_0\varepsilon_m} e^{i(\beta_{SP}y-\omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.41}$$

$$H_x(y,z,t) = \begin{cases} C_1 e^{i(\beta_{SP}y-\omega t)} e^{-\alpha_d z}, & z > 0, \\ C_1 e^{i(\beta_{SP}y-\omega t)} e^{\alpha_m z}, & z < 0 \end{cases} \tag{2.42}$$

$$H_y(y,z,t) = 0 \tag{2.43}$$

$$H_z(y,z,t) = 0 \tag{2.44}$$

Examining these equations reveals that the surface plasmon wave exhibits elliptical polarization within the y-z-plane, and that the handedness of the elliptical polarization differs between the dielectric and the metal. This elliptical polarization property

will be shown to be invaluable for the realization of magnetoplasmonic devices discussed in subsequent sections and chapters of this thesis.

Further insight into the 1D surface plasmon wave is gained through substitution of the non-zero field components into Eqn. 2.4, which produces the following relation between the material permittivities and the transverse decay constants:

$$\frac{\alpha_d}{\alpha_m} = -\frac{\varepsilon_d}{\varepsilon_m} \tag{2.45}$$

Since the requirement of exponential decaying fields away from the interface invokes that $\alpha_d > 0$ and $\alpha_m > 0$, Eqn. 2.45 demands that ε_d and ε_m have opposite signs, which is only true in the case of a metal-dielectric interface. These SP waves cannot be supported at an interface between two metals or two dielectrics.

With the aforementioned results, one can substitute the electric fields represented in Eqns. 2.29-2.31 into the wave equation for each medium (Eqn. 2.18) and obtain the following relations:

$$\beta_{SP}^2 - \alpha_d^2 = \varepsilon_d \left(\frac{\omega}{c}\right)^2 \tag{2.46}$$

$$\beta_{SP}^2 - \alpha_m^2 = \varepsilon_m (\frac{\omega}{c})^2 \tag{2.47}$$

Combining and rearranging Eqns. 2.45-2.47, one can arrive at the surface plasmon dispersion relation for a metal-dielectric interface:

$$\beta_{SP} = \frac{\omega}{c}\sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} = \beta_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \tag{2.48}$$

Here, β_0 is the free-space wavevector, defined as $\beta_0 = \omega/c = 2\pi/\lambda_0$, where λ_0 is the free space wavelength.

The surface plasmon dispersion relation is plotted in Fig. 2.3 for an air-Au interface, where the optical properties of Au are calculated via the Drude model (Eqn.

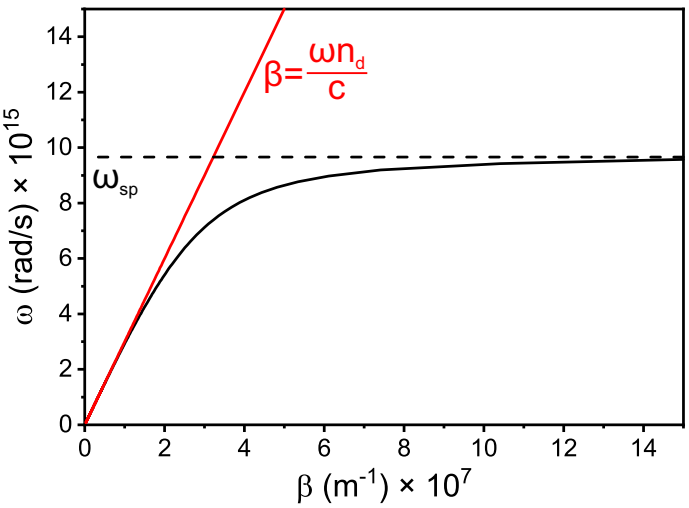

Figure 2.3: Plot of the dispersion relation for a plasmon wave propagating at the interface of air and Au. The light line for air is depicted in red, where $n_d = 1$. There is no intersection between the light line and the plasmon dispersion relation, and the plasmon wave cannot be directly excited without momentum matching.

2.28) with values taken from Ref. [25]. Additionally, the light line for radiation propagating air is plotted as the red line in Fig. 2.3. Two important facts can be noted from these curves. First, the SP dispersion relation approaches the light line at low frequencies and low wavevectors. As the wavevector of the SP increases, it asymptotically approaches the surface plasmon frequency, ω_{sp}, given by $\omega_{sp} = \omega_p/\sqrt{1+\varepsilon_d}$, where $\varepsilon_d = n_d = 1$ in the present example of an air-Au interface. Second, there is no intersection between the SP dispersion relation and the air light line. The surface plasmon dispersion relation lies to the right of the air light line, and as such, surface plasmons require more in-plane momentum than can be physically supported in air. This implies that one cannot directly excite surface plasmons via incidence in air. Alternate methods of excitation must be employed to provide the additional momentum required by the SP wave.

Numerous methods have been developed to impart the additional required momentum and excite surface plasmons. Prominent examples of these coupling schemes are methods which employ a prism to couple light into an SP wave, such as the Kretschmann [26] or Otto [27] configurations, and periodic grating couplers [28]. In

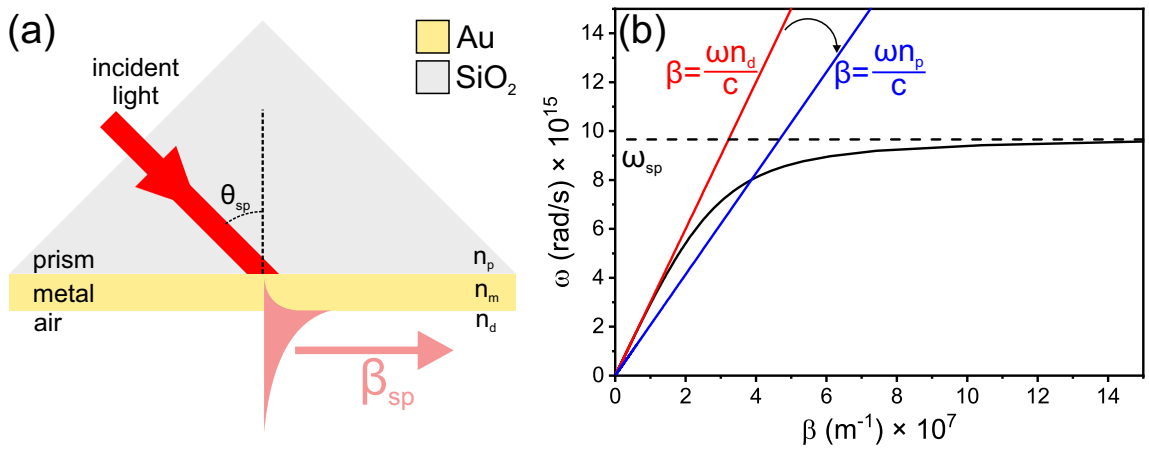

Figure 2.4: (a) Schematic illustration of SP excitation via the Kretschmann configuration. (b) Dispersion relation of a surface plasmon at the air-Au interface. The air light line is plotted in red, and the light line in the prism, taken here to be SiO_2 (refractive index, $n_p = 1.44$), is plotted in blue. Incidence from the prism provides the required additional momentum to excite a surface plasmon at the air-Au interface.

the Kretschmann configuration [26], schematically depicted in Fig. 2.4(a), light impinges on a metallic film from a prism made of a material with a higher refractive index than air, such as SiO_2. The momentum of the incident photon inside the prism is given by $\beta = \omega n_p/c$, where n_p is the refractive index of the prism material (taken to be $n_p = 1.44$ for SiO_2 in this example). As shown in Fig. 2.4(b), when the photon is incident from the prism side, this light coupling scheme effectively increases the photon's wavevector. Thus, the in-plane momentum component of a photon, β_y, incident at an angle of θ_i is given by the following:

$$\beta_y = n_p \beta_0 \sin(\theta_i) \tag{2.49}$$

Notably, to couple to a SP, one must increase θ_i beyond the critical angle, θ_c, for total internal reflection (TIR). The reason for this is as follows. Light incident at an angle below θ_c will refract through the interface and out into the air. Since it radiates in the air, its wavevector will lie to the left of the air light-line, and it inherently will not have enough momentum to couple to a SP. However, increasing θ_i beyond the critical angle induces TIR. In this case, the light outside the prism interface exists as a nonradiative

wave that evanescently decays away from the interface. This nonradiative wave lies to the right of the air light-line, as shown in Fig. 2.4(b), and can preserve the increased momentum from the higher index material. Accordingly, as the transmitted light does not radiate into air and decays away from the interface rapidly, the thickness of the metallic film must be sufficiently thin such that the evanescent fields can penetrate through and couple to the air-metal interface. In this way, one can employ Eqns. 2.48 and 2.49 to determine the specific angle that simultaneously satisfies both equations, and allows for the incident light to be efficiently coupled into a propagating plasmon mode at the air-metal interface. This angle is denoted as the surface plasmon angle, or θ_{sp}, as shown in Fig. 2.4(a).

A variation of the Kretschmann configuration that also employs a prism to couple to a SP wave is the Otto configuration [27]. The difference here is that the metal film is separated from the prism by a nanoscale dielectric gap, such that the evanescent fields from TIR couple through this thin dielectric layer to a SP on the adjacent metal film.

In contrast, the grating coupling method [28] can facilitate coupling to the plasmon with light directly incident on the air-metal interface, and does not require the use of a bulk prism structure. In this scheme, a grating having a period of Λ is constructed directly on the surface of the metallic film, at the air-metal interface. Light impinges onto the grating at an incident angle θ_i, and as such, the in-plane momentum of the incident photon is $\beta_y = \beta_0 \sin(\theta_i)$. Incident light diffracts off of this periodic grating, and this diffraction imparts extra momentum. The total in-plane momentum is given by:

$$\beta_y = \beta_0 \sin(\theta_i) + k\frac{2\pi}{\Lambda} \tag{2.50}$$

where k is an integer representing the diffraction order. In this design, the grating period, Λ, can be designed to increase the in-plane photon momentum to precisely match that required by a SP on the air-metal interface. Since the momentum match-

ing condition can be achieved in this manner, the light can be coupled into a plasmon wave propagating in the y-direction.

2.2.2 Plasmonic Waveguides

While the study of surface plasmon waves excited on 1D interfaces is a useful tool to illustrate the fundamental physics, practical integrated applications require implementation of full three-dimensional (3D) waveguide architectures. In literature, there are a multitude of waveguide designs and material configurations that have been explored. Typically, for use in the telecommunications and optical spectral regimes, metals such as Au and Ag are employed, due to their low optical losses. However, many studies examine materials such as Al and Cu, especially when considering CMOS compatibility. The following sections briefly detail some of the characteristics of the most common plasmonic waveguide geometries.

Metal-Insulator Plasmonic Waveguides

The simplest plasmonic waveguide geometry, and essentially the two-dimensional (2D) extension of the 1D plasmon studied in Section 2.2.1, is the metal-insulator (MI) or dielectric loaded (DL) plasmonic waveguide [29–31]. There are two variations of this waveguide structure. In the first configuration, depicted in Fig. 2.5(a), a high-index dielectric waveguide core is set atop of a low-index dielectric substrate, and a thin metal layer caps the top of the core. In the second configuration, depicted in Fig. 2.5(b), only one dielectric is required, and it is formed into the waveguide core and set atop an optically thick metallic film acting as the substrate.

The fundamental mode at a wavelength of $\lambda = 1550$nm of the first configuration is depicted in Fig. 2.5(c), considering a Si core $w_{Si} = 350$nm wide and $t_{Si} = 340$nm tall, capped with a Au film $t_{Au} = 75$nm tall and $w_{Au} = w_{Si}$ wide, and set atop a SiO_2 substrate. This mode propagates with an effective refractive index of $n_{eff} = 3.33$,

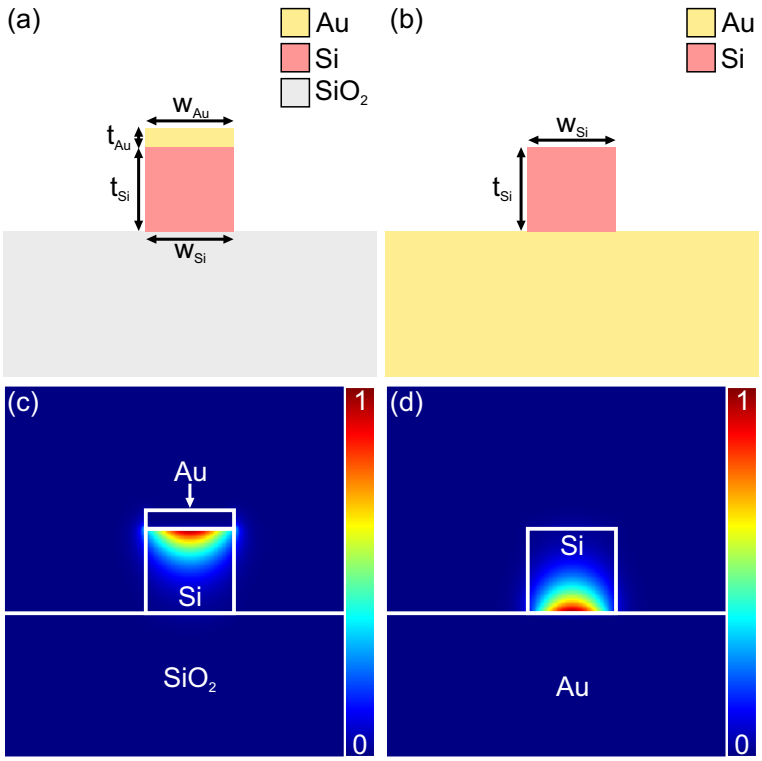

Figure 2.5: (a) MI plasmonic waveguide with a metal cap. (b) MI plasmonic waveguide with a metal substrate. (c) $|E|^2$ profile of the fundamental TM mode of the structure depicted in (a) at $\lambda = 1550$nm. (d) $|E|^2$ profile of the fundamental TM mode of the structure depicted in (b) at $\lambda = 1550$nm. The dimensions of the layers are $w_{Si} = w_{Au} = 350$nm, $t_{Si} = 340$nm, and $t_{Au} = 75$nm.

and propagation losses of 0.97dB/μm, which implies a modal propagation length of 4.48μm. Similarly, the fundamental mode at $\lambda = 1550$nm for the second configuration in illustrated in Fig. 2.5(d) for a Si core with the same dimensions as above, set atop a Au substrate (or optically thick film). Here, the mode propagates with an effective index of $n_{eff} = 3.25$, propagation losses of 0.86dB/μm, and a propagation length of 5.08μm. In both structures, the electric fields are confined tightly to the Au-Si interface, with 94.0% and 96.0% of the power contained in the Si core in each structure, respectively. Note that these calculations employ refractive index data from Ref. [32] for Au, and Ref. [33] for Si and SiO_2.

Since the plasmonic mode propagates at the interface between the dielectric core and the metallic film, these waveguide architectures can keep the modal energy confined to the core as the dimensions are reduced. However, the disadvantage of this

geometry is its simplicity; it provides few degrees of freedom to tailor the modal properties. These are limited only to the material choices and the dimensions of the core.

Both of these geometries are relatively easy to fabricate using common lithographic and nanofabrication techniques. The first configuration is particularly attractive for integration with silicon photonic waveguides, as they can both be formed in a similar manner on a silicon-on-insulator (SOI) wafer, and can be efficiently integrated together [31]. The second configuration has been employed into modulator structures, particularly those exploiting thermo-optic properties of the core material via Joule heating of the underlying metal [34–36].

Hybrid Plasmonic Waveguides

While the previous MI configurations are simple to construct, the SP modes suffer from relatively high losses and short propagation lengths. One proposed solution to mitigate these losses is to incorporate a thin, low-index buffer layer to form a hybrid plasmonic waveguide [37, 38]. Since this is a slight alteration of the MI architecture, there are two variations. The first, depicted in Fig. 2.6(a), consists of a high index dielectric core set atop a lower index dielectric substrate, and the capping metallic layer is separated from the core by the thin, low index buffer layer. Here, both the buffer layer and the substrate are taken to be SiO_2, however these two components do not need to be the same material. The second configuration, depicted in Fig. 2.6(b), consists of a high index dielectric core separated from the metallic substrate by the thin low index buffer layer. The modes of such structures are denoted as hybrid modes, since the electric fields in the low index dielectric at the metal interface are plasmonic, while the fields propagating within the high index core propagate as a photonic mode. The losses are therefore reduced, as the fields within the high index dielectric are separated from the metal.

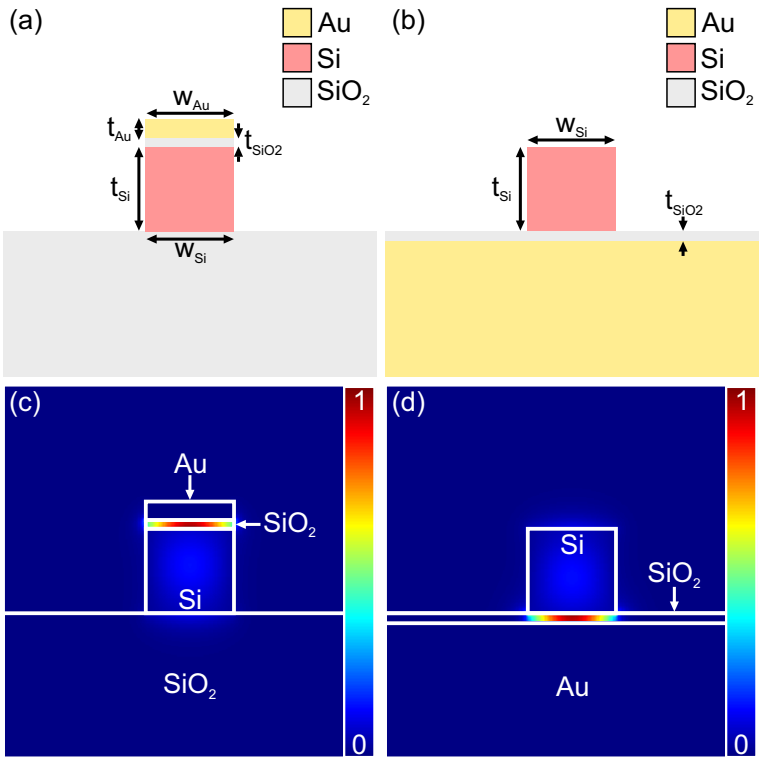

Figure 2.6: (a) Hybrid plasmonic waveguide with a metal cap. (b) Hybrid plasmonic waveguide with a metal substrate. (c) $|E|^2$ profile of the fundamental TM mode of the structure depicted in (a) at $\lambda = 1550$nm. (d) $|E|^2$ profile of the fundamental TM mode of the structure depicted in (b) at $\lambda = 1550$nm. The dimensions of the layers are $w_{Si} = w_{Au} = w_{SiO_2} = 350$nm, $t_{Si} = 340$nm, $t_{Au} = 75$nm, and $t_{SiO_2} = 35$nm.

Exemplary mode profiles for the two geometries are depicted in Fig. 2.6(c) and 2.6(d), respectively. In Fig. 2.6(c), the TM mode at $\lambda = 1550$nm is shown considering a Si core with $w_{Si} = 350$nm wide and $t_{Si} = 340$nm, separated from the $t_{Au} = 75$nm Au film by a $t_{SiO_2} = 35$nm SiO_2 buffer layer, and built on top of a SiO_2 substrate. As well, $w_{Si} = w_{Au} = w_{SiO_2}$. Here, the mode propagates with a lower effective index of $n_{eff} = 2.53$, and reduced losses of 0.12dB/μm, which correspond to $L_{prop} = 37.17\mu$m, while 73.2% of the modal power is confined to the Si core, and 17.6% propagates in the SiO_2 layer. In a similar manner, Fig. 2.6(d) depicts the $\lambda = 1550$nm mode for the second configuration, with an identical Si core separated from the optically thick Au substrate by a $t_{SiO_2} = 35$nm. This mode exhibits $n_{eff} = 2.48$, propagation losses of 0.099dB/μm, and therefore, a propagation length of $L_{prop} = 43.95\mu$m. 73.4% of

the power resides within the Si core, while 18.7% resides in the SiO_2. This scheme improves the propagation length by factors of 8.3 and 8.7 relative to the respective MI configurations presented in the previous section.

Due to their hybrid nature, enhanced field confinement, reduced optical losses, small footprint, and relative ease of fabrication, hybrid plasmonic waveguides have found applications in passive devices such as resonators [39], nanolasers [40], and nonlinear optics [41].

Insulator-Metal-Insulator Plasmonic Waveguides

Another natural extension of the 1D plasmon geometry involves evanescent waves propagating on both sides of a metallic film, leading to what is known as the insulator-metal-insulator (IMI) plasmonic waveguide. In practice, this structure can be as simple as a thin metal strip, several microns wide, sandwiched between two dielectric claddings [42, 43], or a complex layered structure [44, 45]. For a sufficiently thick metal film (i.e. >> the decay length of the SP fields into the metal, which is typically on the order of a few tens of nanometers), the plasmons at each metal-dielectric interface behave independently. As the metal film thickness is reduced (i.e. to the order of a few tens of nanometers and thinner), the evanescent plasmon tails interact through the metal film, and the plasmon waves on either side can couple together. Typically, this coupling leads to the formation of modes with symmetric transverse field profiles (denoted as symmetric modes, or SM), and modes with antisymmetric field profiles (denoted as antisymmetric modes, or AM).

The mode with the symmetric transverse profile exhibits an antisymmetric longitudinal electric field profile across the metal film. Antisymmetry reduces the longitudinal field component within the metal, and since the longitudinal component is primarily responsible for the Ohmic losses experienced by the longitudinal charge oscillations of the SP wave [9], this configuration reduces the losses, and forms a long-

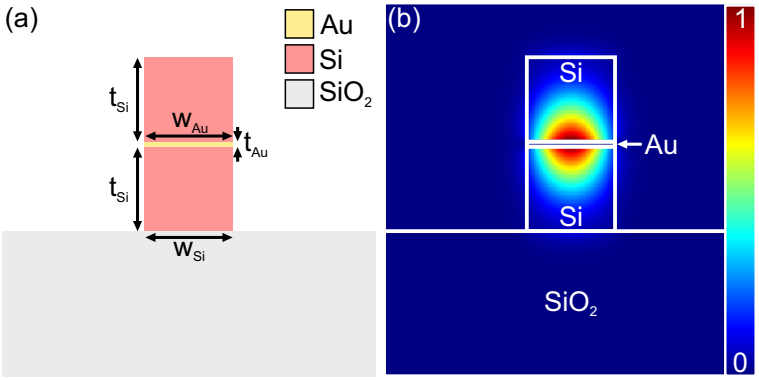

Figure 2.7: (a) IMI plasmonic waveguide having a finite width in a silicon platform. (b) $|E|^2$ profile of the fundamental TM mode of the structure depicted in (a) at $\lambda = 1550$nm. The dimensions of the layers are: $w_{Si} = w_{Au} = 350$nm, $t_{Si} = 340$nm, and $t_{Au} = 25$nm.

range surface plasmon (LRSP) mode [46]. In this respect, these modes can propagate distances up to the order of millimeters with proper material choices. Naturally, these modes cannot be confined as tightly.

An integrable and compact form of an IMI plasmonic waveguide is depicted in Fig. 2.7(a), where a thin Au film is stacked between two Si cladding layers, on an SiO_2 substrate. Note that the two cladding layers do not have to be of the same material, or have the same dimensions. Here, the structure is considered such that $w_{Si} = 350$nm and $t_{Si} = 340$nm for both cladding layers, and a Au film $t_{Au} = 25$nm tall, with $w_{Au} = w_{Si}$. The symmetric mode at $\lambda = 1550$nm is depicted in Fig. 2.7(b), and has an effective index of $n_{eff} = 3.07$, propagation losses of 0.11db/μm, and a corresponding propagation length of $L_{prop} = 39.93\mu$m. The fields are not confined as tightly to the interface, and these structures tend to have large vertical dimensions.

Despite this, IMI waveguides provide several avenues for mode tailoring, as the two cladding regions can be comprised of different materials and different cross sectional dimensions. Supplementary layers can also be added into the structure to further tailor the mode profile for low-loss applications [45].

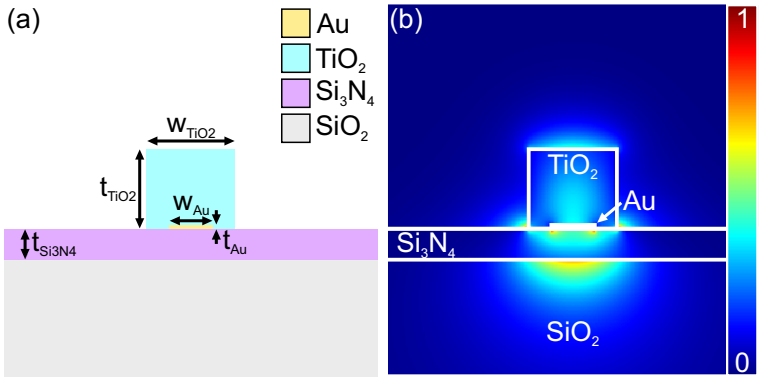

Figure 2.8: (a) LRDLSPW waveguide structure. (b) Fundamental mode of the structure depicted in (a) at $\lambda = 1550$nm. The dimensions of the layers are $w_{TiO_2} = 350$nm, $t_{TiO_2} = 320$nm, $t_{Au} = 25$nm. $w_{Au} = 175$nm, and $t_{Si_3N_4} = 120$nm.

Long-Range Dielectric Loaded Surface Plasmon Waveguide

The merits of the MI/DL and IMI architectures have been merged into a further waveguide design, denoted as the long-range dielectric loaded surface plasmon waveguide (LRDLSPW) [47–50]. This geometry is considerably more complex, but provides a multitude of parameters for tailoring the modal properties, and can achieve both nanoscale confinement and millimeter scale propagation lengths. In this platform, depicted in Fig. 2.8(a), at least three dielectrics are required: one as the substrate, one as the buffer layer on top of the substrate, and one as the waveguide core. A thin metallic layer, typically 15-20nm in thickness and of a width less than the core width, is embedded within the core at the interface between the core and the buffer layer. The reduced width of the metal layer allows the mode to experience less loss. Further, while this is the basic geometry, multiple buffer layers can be employed for further mode tailoring versatility [47].

For example, the mode profile of the structure depicted in Fig. 2.8(a) at $\lambda = 1550$nm is shown in Fig. 2.8(b). Here, a core constructed out of TiO_2 is considered, which has dimensions of $w_{TiO_2} = 350$nm and $t_{TiO_2} = 320$nm, and a refractive index of $n_{TiO_2} = 2.3 + 1.1 \times 10^{-6}i$ [51]. A $t_{Au} = 15$nm strip of Au, $w_{Au} = 175$nm wide, is embedded at the base. The buffer layer incorporates $t_{Si_3N_4} = 120$nm of Si_3N_4 (with

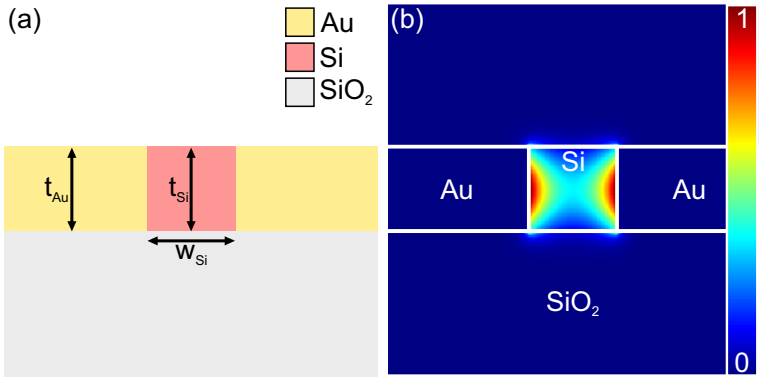

Figure 2.9: (a) MIM waveguide structure on a silicon platform. (b) Fundamental mode of the structure depicted in (a) at $\lambda = 1550$nm. The dimensions of the layers are $w_{Si} = 350$nm, and $t_{Si} = t_{Au} = 340$nm.

$n_{Si_3N_4} = 1.977$ [52]), and all of this is built atop a SiO_2 substrate. In this configuration, the $\lambda = 1550$nm mode has an effective index of $n_{eff} = 1.46$, propagation losses of only 0.0036dB/μm, and correspondingly a propagation length of $L_{prop} = 1209.8\mu$m. Here, 15.8% of the modal power is contained within the TiO_2 layer, 15.2% is contained within the Si_3N_4 layer, and 57.8% propagates within the SiO_2 layer.

This geometry provides the greatest number of degrees of freedom for mode tailoring, including the constituent materials, the number of buffer layers, and the dimensions of the buffer layers, core, and metal.

Metal-Insulator-Metal Plasmonic Waveguides

In contrast to the IMI waveguide geometry, one can analogously consider the metal-insulator-metal (MIM) geometry, where a thin dielectric core is sandwiched between two metal claddings [53]. This configuration is sometimes referred to as a slot waveguide. An example of this structure on a Si platform is depicted in Fig. 2.9(a). Here, we consider a Si core between two Au claddings, on a SiO_2 substrate. Note that the metal claddings do not have to be of the same material. In a similar manner to the IMI geometry, for very large core widths, the surface plasmons on each metal-dielectric interface essentially behave independently. As the width of the core decreases such

that the modal fields begin to interact (i.e. to widths approximately on the order of twice the SP field decay length), the plasmonic fields on each side will couple and form both symmetric and antisymmetric supermodes.

The mode profile at $\lambda = 1550$nm for an exemplary structure with $w_{Si} = 350$nm, and $t_{Si} = t_{Au} = 340$nm is illustrated in Fig. 2.9(b). The mode has a high effective index of $n_{eff} = 3.49$, propagation losses of $0.83802\text{dB}/\mu\text{m}$, and a propagation length of $L_{prop} = 5.18\mu$m. This particular configuration confines 93.4% of the power within the Si core, and unlike the other geometries, the core power fraction of this structure actually increases and w_{Si} decreases. Hence, this geometry can be employed to form nanoscale devices with extremely high field confinement [54]. The high confinement reduces cross talk between adjacent waveguides, making it ideal for plasmonic circuit applications.

The strong field confinement in combination with the plasmonically enhanced fields make the MIM waveguide an ideal platform for nonlinear optics studies [55] and modulator designs [56].

Non-planar Geometries

The aforementioned plasmonic waveguide architectures are straightforward to integrate with planar nanofabrication processes used in microelectronics and silicon photonics, due to their layered structure and simple, rectangular geometry. However, there are a number of additional geometries presented in literature that do not conform to this planar standard. While difficult to merge into existing process flows, these platforms are briefly summarized below.

Two complementary device architectures presented in literature are the wedge [57] and groove (or channel) [58] plasmonic waveguides, whose structures are schematically depicted in Fig. 2.10(a) and 2.10(b), respectively. The wedge waveguide relies on a triangular metallic point in a metallic film, which guides the evanescent mode. In

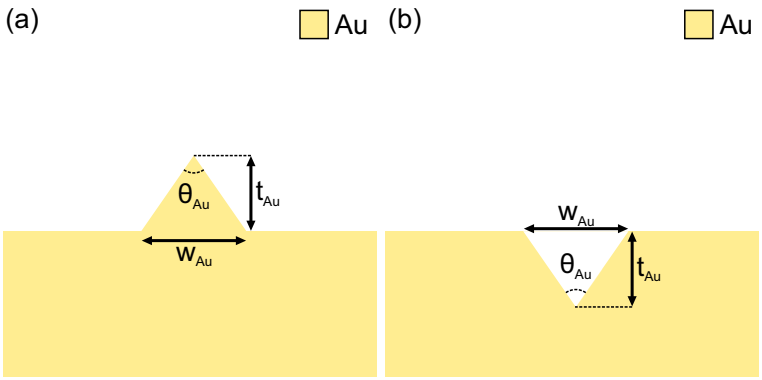

Figure 2.10: Schematic illustration of (a) the wedge plasmonic waveguide, and (b) the groove or channel plasmonic waveguide, both formed from Au.

contrast, the groove geometry consists of a V-shaped groove formed in a metallic film, which guides a mode within its channel. These structures have been utilized to construct optical components such as interferometers and resonators [59], as well as to demonstrate plasmon propagation on the order of 4μm in the visible spectral region [60]. However, their fabrication requires focused ion beam (FIB) milling [57–59] which can be expensive and low-throughput/yield, or metalization of anisotropically etched silicon [60], which only provides limited control over the groove angle, depth, and width. As such, these geometries have yet to find widespread usage in nanoplasmonic circuits and devices.

2.3 Magneto-Optics

Since Michael Faraday first discovered that the polarization of light is rotated as it passed through a piece of glass in a magnetic field in 1845 [61], the influence of magnetic fields and magnetization on the propagation of light has been a fundamental interest in optics. Magneto-optic (MO) effects have the unique property of non-reciprocity, meaning that electromagnetic waves propagating in opposing directions through a magnetized material will interact with the medium in a different manner. The presence of a magnetization and magnetic field breaks the time reversal symme-

try of the electromagnetic wave-material interaction. In fact, incorporating magnetic materials into optical structures and devices is one of the few ways to achieve such nonreciprocal effects.

Nonreciprocity is crucial for controlling the flow of light. Only by constructing devices with this property can one realize optical analogs to electrical diodes, which is fundamentally required to develop complex integrated optical networks and on-chip optical circuitry. MO effects also offer a means of manipulating the intrinsic properties of a lightwave, such as its phase and polarization state, and facilitate unique optical modulator architectures. By extension, it is crucial to miniaturize such components and realize them in the nanoplasmonic regime.

2.3.1 Origin of Magneto-Optical Effects

Microscopic Description

The origin of the various magneto-optic effects that will be examined is rooted in a quantum mechanical phenomenon called the Zeeman Effect. This effect can be explained within the context of classical mechanics and electrodynamics as follows [20, 62–64]: Consider the simple classical picture of an atom having a single electron orbiting the nucleus, illustrated in Fig. 2.11. This orbiting charge possesses orbital angular momentum, $\boldsymbol{L}$, and acts as a small current loop, which can induce a magnetic moment, called the orbital magnetic moment, $\boldsymbol{\mu_L}$. The electron itself is spinning about its axis, and possesses its own spin angular momentum, $\mathbf{S}$, and its associated spin magnetic moment, $\boldsymbol{\mu_s}$. The total magnetic moment of the atom is the sum of these two magnetic moments, but the orbital magnetic moment is typically very weak, and only plays a role in diamagnetic materials, where all electrons are paired up and the net spin magnetic moments cancel out. In a material with unpaired electrons, such as one that exhibits paramagnetism, ferromagnetism, or ferrimagnetism, the spin magnetic moment is the dominate moment, and this mechanism is considered

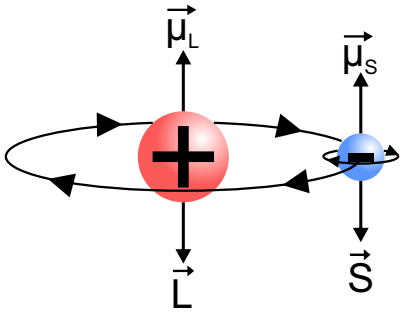

Figure 2.11: Schematic of the atomic magnetic moments

here. While this classical analogy is a useful illustrative tool, it is essential to keep in mind that electrons are not physically spinning objects, as they are more accurately thought of as clouds of charge density. The concepts of spin and the spin magnetic moment are simply intrinsic quantum mechanical properties in that picture, and not actually physical angular motion.

The spin angular momentum of the electron can be defined by a vector, $\mathbf{S}$. Correspondingly, the spin magnetic moment, $\boldsymbol{\mu_s}$, due to the electron spinning about its axis can be written as:

$$\boldsymbol{\mu_s} = -g\frac{q}{2m_e}\mathbf{S} \tag{2.51}$$

where g is the Landé factor, typically taken to be equal to 2; however, rigorous quantum electrodynamics analysis has shown the exact value is 2.00232. Applying a magnetic field, $\mathbf{H}$, or equivalently, a magnetic flux density, $\mathbf{B}$, will exert a torque, $\boldsymbol{\tau}$, on the spin magnetic moment:

$$\boldsymbol{\tau} = \boldsymbol{\mu_s} \times \mathbf{B} \tag{2.52}$$

The corresponding potential energy, U, can be written as:

$$U = -\boldsymbol{\mu_s} \cdot \mathbf{B} \tag{2.53}$$

For simplicity, consider the applied magnetic flux density to be oriented along the $+y$-axis, such that $\mathbf{B} = \langle 0, B_y, 0 \rangle$. Since quantum mechanics dictates that the spin

angular momentum of the electron is quantized, and that the only allowable values for the electron spin quantum number are $\pm\frac{1}{2}$, the y-component of the spin angular momentum is $S_y = \pm\frac{1}{2}\hbar$, where $\hbar = 1.055 \times 10^{-34}$J·s is the reduced Planck constant. Thus, the y-component of the $\boldsymbol{\mu_s}$ is:

$$\mu_{s,y} = -g\frac{q}{2m_e}S_y = \pm\frac{1}{2}g\mu_B \tag{2.54}$$

where $\mu_B = 9.274 \times 10^{-24}$J/T is the Bohr magneton, given by:

$$\mu_B = \frac{q\hbar}{2m_e} \tag{2.55}$$

The potential energy due to the interaction of the applied flux density and the spin magnetic moment is:

$$U = -\mu_{s,y}B_y = \pm\frac{1}{2}g\mu_B B_y \tag{2.56}$$

Remarkably, there are two different energy states for the potential energy. The implication of this is as follows: If the electron exists in a state with energy E_1, the application of a magnetic field splits this energy level into two new energy states separated by an energy difference of $\Delta E = g\mu_B B_y$. Since $g \approx 2$, the two new energy levels will exist at $E_1 \pm \mu_B B_y$, and all of the electrons in each state will possess the same spin. This phenomenon, discovered by Pieter Zeeman, is denoted as Zeeman splitting, and is schematically illustrated in Fig. 2.12. Figure 2.12(a) depicts the case where no magnetic field is applied, where there is only one energy state at E_1 for electrons of both spins. Applying a magnetic field, causes the splitting of the energy levels and spin states, as shown in Fig. 2.12(b). Notably, this effect is also responsible for the splitting of spectroscopic absorption lines [65, 66].

Next, consider the interaction of light with a magnetic medium. For example, a linearly x-polarized plane wave propagating along the y-axis and having an amplitude

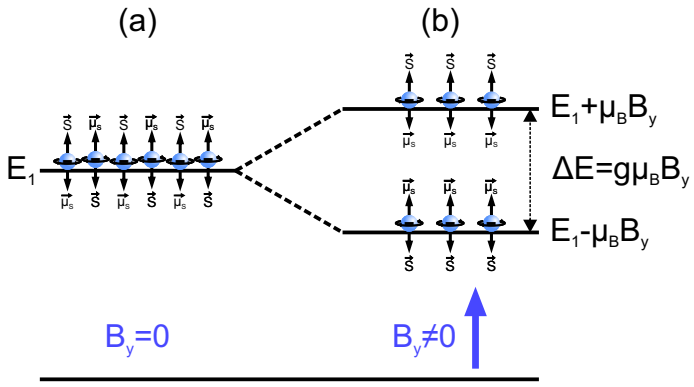

Figure 2.12: Schematic of the Zeeman Effect. (a) With no applied field, electrons having both spin states are present at the same energy level, E_1. (b) With an applied magnetic flux density, B_y, the energy level E_1 splits into two levels, and the electrons are separated by spin state. Image adapted and modified from [63].

E_0 can be written as:

$$\mathbf{E}(y,t) = \mathbf{E}_0 e^{i(\beta y - \omega t)} = E_0 e^{i(\beta y - \omega t)} \hat{\mathbf{x}} \tag{2.57}$$

This wave, as well as any linearly polarized wave, can be decomposed as the superposition of a right-circularly polarized wave ($\mathbf{E}_R$) and a left-circularly polarized wave ($\mathbf{E}_L$). In the present example, the decomposition yields the following:

$$\mathbf{E}(y,t) = \mathbf{E}_R(y,t) + \mathbf{E}_L(y,t) \tag{2.58}$$

$$\mathbf{E}_R(y,t) = \frac{E_0}{2} e^{i(\beta y - \omega t)} \hat{\mathbf{x}} - i \frac{E_0}{2} e^{i(\beta y - \omega t)} \hat{\mathbf{z}} \tag{2.59}$$

$$\mathbf{E}_L(y,t) = \frac{E_0}{2} e^{i(\beta y - \omega t)} \hat{\mathbf{x}} + i \frac{E_0}{2} e^{i(\beta y - \omega t)} \hat{\mathbf{z}} \tag{2.60}$$

Similar to electrons, the photons comprising such a wave are each characterized by a spin angular momentum as well, where the two possible quantized values are related to the handedness of the rotating wave [23]. When this wave interacts with the medium in the presence of an applied magnetic field, conservation of angular momentum dictates that photons of a particular handedness interact with electrons with the same orientation of spin angular momentum. Since these electrons are

contained within two separate and distinct energy levels, the two waves of opposite handedness observe different refractive indices and absorption, and interact with the medium differently. This difference in refractive index and absorption gives rise to magneto-optic effects.

Classical Electromagnetic Treatment

While the microscopic quantum mechanical treatment is useful for understanding the physical mechanism behind MO effects, a macroscopic description is also insightful for analyzing these effects and modeling devices based on Maxwell's equations. In the case of an optically active material, such as one exhibiting magneto-optic effects, the displacement field is given by [64]:

$$\mathbf{D} = \varepsilon_0(\varepsilon_r\mathbf{E} + i[\mathbf{g} \times \mathbf{E}]) \tag{2.61}$$

where $\mathbf{g}$ is denoted as the gyration vector. The additional $i\,[\mathbf{g} \times \mathbf{E}]$ term arises from the inclusion of the Lorentz force acting on the electron in the presence of an applied magnetic field in the balance of forces presented in Eqn. 2.20. Correspondingly, in a MO material, the relative permittivity is no longer a simple scalar value, but rather, it becomes a full anisotropic and asymmetric tensor [64]:

$$\hat{\varepsilon}_r = \begin{bmatrix} \varepsilon_r & -ig_z & ig_y \\ ig_z & \varepsilon_r & -ig_x \\ -ig_y & ig_x & \varepsilon_r \end{bmatrix} \tag{2.62}$$

where ε_r is the diagonal, isotropic permittivity, and $\hat{\varepsilon}_r$ is the full tensor. The gyration vector, $\mathbf{g} = \langle g_x, g_y, g_z \rangle$ and its constituent elements are defined as:

$$g_i = -\frac{\lambda\sqrt{\varepsilon_r}}{\pi}\theta_F \tag{2.63}$$

Here, θ_F is the Faraday rotation, which is related to the magnetization state by [67]:

$$\theta_F = \frac{M_i}{M_S}\theta_{spec} \tag{2.64}$$

where M_i are the components of the magnetization vector, $\mathbf{M} = \langle M_x, M_y, M_z \rangle$, M_S is the saturation magnetization of the material, and θ_{spec} is the specific Faraday rotation, which is a material parameter defined as the Faraday rotation per unit length in a fully magnetized medium.

By orienting the applied magnetic field, and therefore the magnetization, along one of the Cartesian axes, many of the elements of the tensor in Eqn. 2.62 reduce to 0, simplifying the analysis.

2.3.2 Transmission Effects

The Faraday Effect

The Faraday effect, also known as magnetic circular birefringence, was first discovered in 1845 by Michael Faraday [61]. This effect manifests as the rotation of the plane of polarization of light propagating through a medium parallel to the direction of an applied static magnetic field ($\mathbf{H}_{static}$), as schematically illustrated in Fig. 2.13. As discussed in the previous section, any linearly polarized wave can be decomposed into the superposition of a right- and left-circularly polarized wave ($\mathbf{E}_R$ and $\mathbf{E}_L$). Zeeman splitting of the material's energy levels under an applied magnetic field causes these two oppositely handed waves to experience different refractive indices, denoted as n_+ for $\mathbf{E}_R$ and n_- for $\mathbf{E}_L$. A phase difference accrues between the two circularly polarized waves as they propagate through the medium, manifesting as rotation of the linear polarization state. A plane wave having wavelength λ propagating a distance L through a bulk magnetized medium experiences a polarization rotation of θ_{rot}, which can be described as [64]:

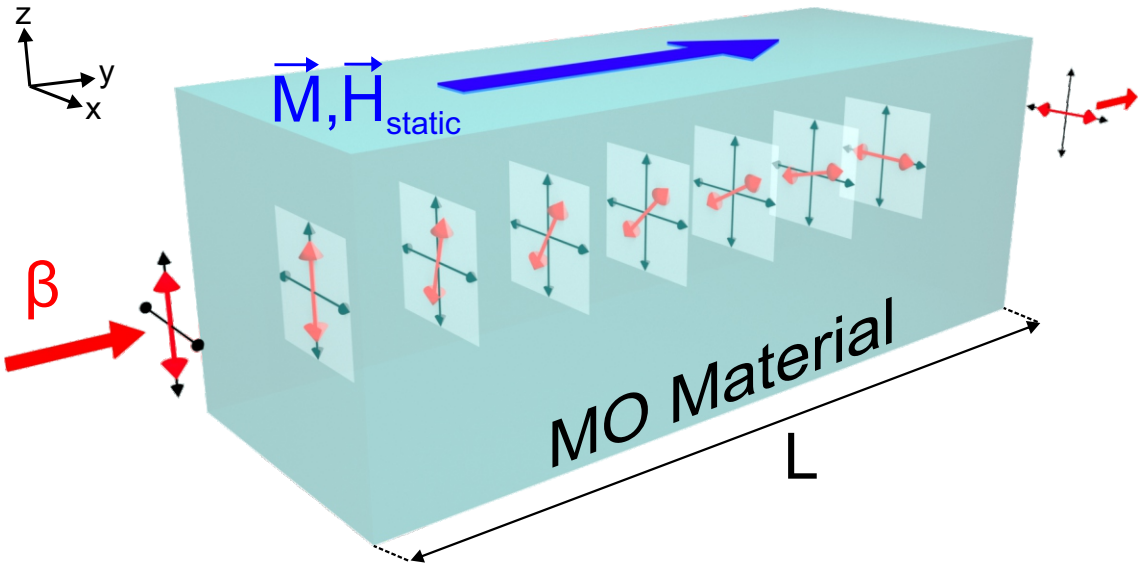

Figure 2.13: Schematic illustration of the Faraday Effect. As linearly polarized light propagates through a MO medium which is magnetized parallel to the applied field, the polarization state (indicated here by the red arrows) rotates.

$$\theta_{rot} = \frac{\pi}{\lambda}\left(n_+ - n_-\right) L \tag{2.65}$$

For diamagnetic and paramagnetic materials, which do not maintain their magnetization state upon removal of the applied magnetic fields, the polarization rotation is proportional to the applied magnetic flux density. For a wave propagating a distance L in such a medium, the polarization rotation can be described by [64]:

$$\theta_{rot} = VBL \tag{2.66}$$

where V is the Verdet constant, or the polarization rotation per unit length per unit flux density, and B is the applied magnetic flux density parallel to the direction of propagation.

In contrast, ferromagnetic and ferrimagnetic materials, upon removal of an applied magnetic field, can maintain long range ordering of their magnetic moments, and Eqn. 2.66 does not apply. In this case, the total rotation is related to the magnetization state of the material, as prescribed by Eqn. 2.64, and $\theta_{rot} = \theta_F L$.

While the above analyses focused on plane wave propagation through bulk media, the situation differs for waveguide geometries. In a waveguide platform, the polarization rotation induced by the Faraday effect manifests as conversion between the TE and TM modes supported by the structure. Such a situation can be treated by coupled mode theory. If an arbitrary waveguide geometry is considered to be composed of MO materials magnetized parallel to the propagation direction (e.g., the y-axis), and a TE mode is launched into the structure at $y = 0$, the resultant power contained within the TE and TM modes (P_{TE} and P_{TM}, respectively) along the length of the waveguide are [68, 69]:

$$\frac{P_{TM}(y)}{P_{TE}(0)} = \frac{4|\kappa|^2}{4|\kappa|^2 + \Delta\beta_{FR}^2} \sin^2\left(y\sqrt{|\kappa|^2 + \frac{\Delta\beta_{FR}^2}{4}}\right) \tag{2.67}$$

$$\frac{P_{TE}(y)}{P_{TE}(0)} = 1 - \frac{4|\kappa|^2}{4|\kappa|^2 + \Delta\beta_{FR}^2} \sin^2\left(y\sqrt{|\kappa|^2 + \frac{\Delta\beta_{FR}^2}{4}}\right) \tag{2.68}$$

Here, $\Delta\beta_{FR} = |\beta_{TE} - \beta_{TM}|$ is the difference in propagation constants between the TE and TM modes of the structure (β_{TE} and β_{TM}, respectively), and it is equivalent to $\Delta\beta_{FR} = \beta_0|n_{eff,TE} - n_{eff,TM}| = \beta_0\Delta n_{eff}$, where $n_{eff,TE}$ and $n_{eff,TM}$ are the effective refractive indices of the TE and TM modes, and Δn_{eff} is the difference between them. κ is a coupling constant related to the overlap of the two modes within the MO material (integrated over the waveguide's cross sectional area, A):

$$\kappa = \iint \mathbf{E}_{TE}^* \Delta\hat{\varepsilon}_r \mathbf{E}_{TM} dA \tag{2.69}$$

and $\Delta\hat{\varepsilon}_r = \hat{\varepsilon}_r(\mathbf{M}) - \hat{\varepsilon}_r(0)$. Examining Eqns. 2.67 and 2.68, it is evident that the maximum mode conversion efficiency (TE to TM) is calculated via:

$$\eta_{FR} = \frac{4|\kappa|^2}{4|\kappa|^2 + \Delta\beta_{FR}^2} \tag{2.70}$$

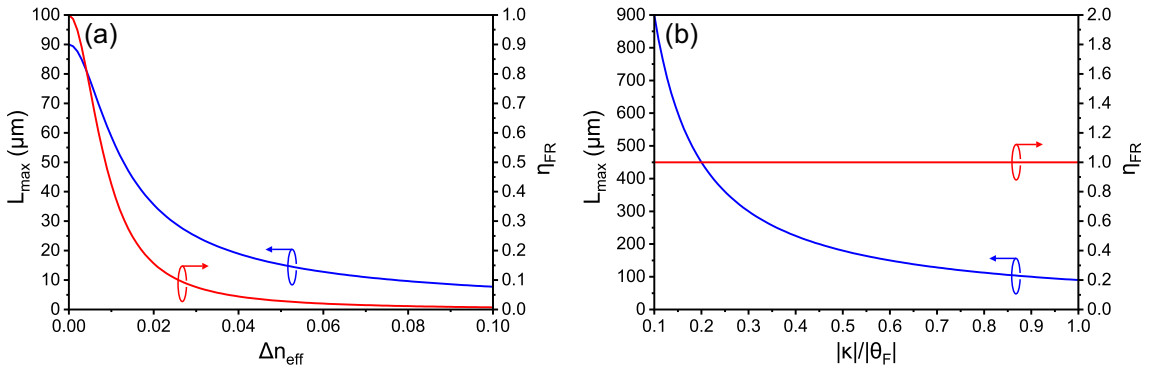

Figure 2.14: (a) Plot of η_{FR} and L_{max} as functions of Δn_{eff}, the difference in effective index between the TE and TM modes. (b) Plot of η_{FR} and L_{max} as functions of the normalized coupling coefficient, $|\kappa|/|\theta_F|$.

The mode conversion varies sinusoidally along the length of the interaction region, and the first mode conversion efficiency maximum occurs at a point $y = L_{max}$:

$$L_{max} = \frac{\pi}{\sqrt{4|\kappa|^2 + \Delta\beta_{FR}^2}} \tag{2.71}$$

Analogous operation can be found for a TM mode input to the device as well. With proper mode amplitude normalization, Eqn. 2.69 and 2.71 show that the coupling constant varies between $0 \leq |\kappa| \leq |\theta_F|$, corresponding to values ranging from no mode overlap ($|\kappa| = 0$) to perfect mode overlap ($|\kappa| = |\theta_F|$). Consequently, Eqns. 2.70 and 2.71 illustrate that there are two key factors which influence the mode conversion within a MO waveguide: (1) phase matching between the TE and TM modes, and (2) the overlap between the two modes within the MO material.

To illustrate the dependence of the mode conversion on these two parameters, consider a TE wave propagating through an exemplary MO material having $\theta_F = 1°/\mu$m. In this bulk material, a plane wave would rotate from TE to TM (90° of rotation) within a length of 90μm. Figure 2.14(a) illustrates η_{FR} and L_{max} as functions of Δn_{eff}, assuming perfect mode overlap ($|\kappa|/|\theta_F| = 1$). Clearly, as the phase difference between the two modes increases, the maximum possible mode conversion efficiency drops rapidly, and the distance to the first maximum of the sinusoidal power variation

decreases as well. In practice, $\Delta n_{eff} = 0$ can be achieved with techniques such as quasi-phase matching [70–72]. Figure 2.14(b) illustrates the same two quantities as a function of the normalized coupling constant, $|\kappa|/|\theta_F|$, assuming perfect phase matching ($\Delta n_{eff} = 0$). In this case, the mode conversion efficiency is always 100%, but the length required to achieve it increases asymptotically as the mode overlap decreases. In a real waveguide, there is a complex interplay between these two parameters. In order to optimize the device's performance, the waveguide structure must be precisely engineered to maximize modal overlap and minimize the phase difference.

The Nonreciprocal Phase Shift

While the Faraday effect described above can occur in bulk material and is polarization independent, there is another MO effect that is unique to guided wave structures. Consider the geometry depicted in Fig. 2.15. In this orientation, the light in a waveguide structure propagates down the waveguide (i.e., along the $+y$-axis) where the applied static magnetic field, $\mathbf{H}_{static}$ is oriented transverse to the direction of propagation (i.e., along the $+x$-axis). Confining light to the waveguide produces a propagating wave having more than just one electric and one magnetic field component. Specifically, the TM mode of the waveguide contains E_y, E_z, and H_x components. The confinement introduces a longitudinal field component, such as E_y, that is phase shifted with respect to the transverse component, E_z. This implies that the guided mode is slightly elliptically polarized within the y-z-plane [73, 74], as shown in Fig. 2.15.

The rotation of this an elliptically polarized field occurs around the same axis as the magnetization of the material (i.e., the x-axis in Fig. 2.15). This creates a situation analogous to that of the Faraday effect. There exists photons with a particular spin angular momentum, which will interact with electrons of the corresponding spin angular momentum due to Zeeman splitting in the material, and this wave will

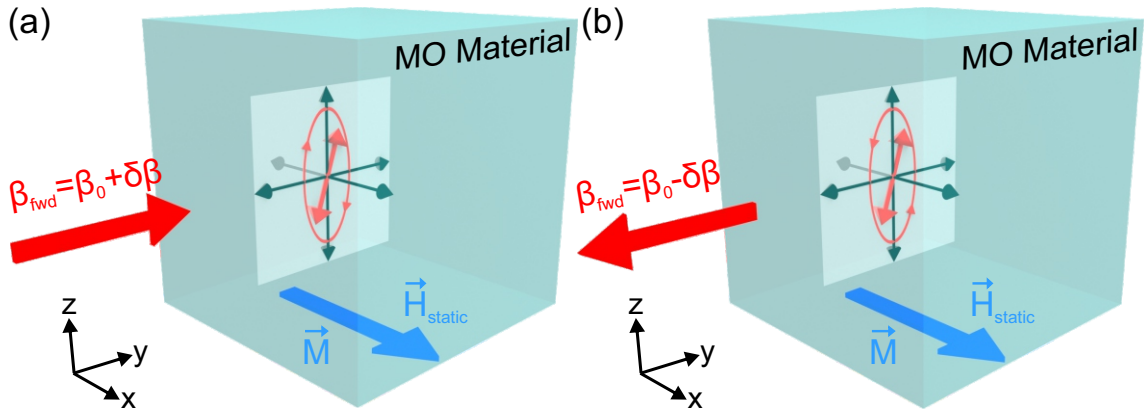

Figure 2.15: Schematic illustration of the NRPS for a wave elliptically polarized in the$y$-z-plane, such as a waveguide mode. (a) Propagation in the forward direction, and (b) propagation in the reverse direction. Note that the handedness of the rotating fields is reversed relative to the applied magnetic field. This leads to different propagation constants for the two directions.

observe a slightly perturbed refractive index. The same wave propagating in the opposite direction (i.e., along the -y-axis in Fig. 2.15) will be rotating with the opposite handedness relative to the material's magnetization, and will interact with electrons in a different energy band and observe a different refractive index. This is schematically depicted in Fig. 2.15(a) for a forward propagating wave, and Fig. 2.15(b) for a backward propagating wave.

If the guided mode propagates in the unmagnetized waveguide with a propagation constant of β_0, the propagation constants of the forward and backward propagating modes (β_{fwd} and β_{bwd}) in the magnetized guide can be described as [68]:

$$\beta_{fwd} = \beta_0 + \delta\beta \tag{2.72}$$

$$\beta_{bwd} = \beta_0 - \delta\beta \tag{2.73}$$

where $\delta\beta$ is the perturbation to the propagation constant introduced by the magnetization. Clearly, there is a difference between β_{fwd} and β_{bwd}, and these two waves will experience a phase difference upon propagating the same length through the

waveguide in opposite directions [68]:

$$\Delta\beta = \beta_{fwd} - \beta_{bwd} = 2\delta\beta \tag{2.74}$$

where $\Delta\beta$ is defined as the difference in the propagation constants of the forward and backward propagating waves, and is known as the nonreciprocal phase shift (NRPS).

The NRPS experienced by a waveguide mode depends on a number of crucial factors. First is the degree of ellipticity, which should be made as large as possible, with the ideal scenario exhibiting circular polarization [74]. However, this might be difficult to achieve in practice. The second key factor is the fraction of power confined within the MO material. If the fields are well confined within the MO material, then the NRPS will be stronger. The third key component is that the guided mode must also propagate for an appreciable distance in order to accumulate a significant NRPS, as this effect is typically much weaker than Faraday rotation.

Finally, to ensure significant NRPS within a waveguide, it is necessary to establish strong spatial asymmetry within the mode's profile [73, 74]. In many geometries, the longitudinal field component of the mode contains a node, and regions of opposite polarity. This is schematically illustrated in Fig. 2.16, which depicts an exemplary waveguide constructed from the magneto-optic garnet, yttrium-iron garnet (YIG), embedded in SiO_2. Figure 2.16(a) depicts the structure, where $t_{YIG} = w_{YIG} = 500$nm, while Figs. 2.16(b) and 2.16(c) depict the profiles of the transverse (E_z) and longitudinal (E_y) field components, respectively, of the fundamental TM mode. In essence, this creates regions within the MO material where the fields are elliptically rotating with one handedness, and regions where they are rotating with the opposite handedness. If there is equal modal power present over these two regions, the net NRPS of the mode will be canceled out [73, 74]. Therefore, one must ensure that the longitudinal field component does not exhibit inverted mirror symmetry. A number of

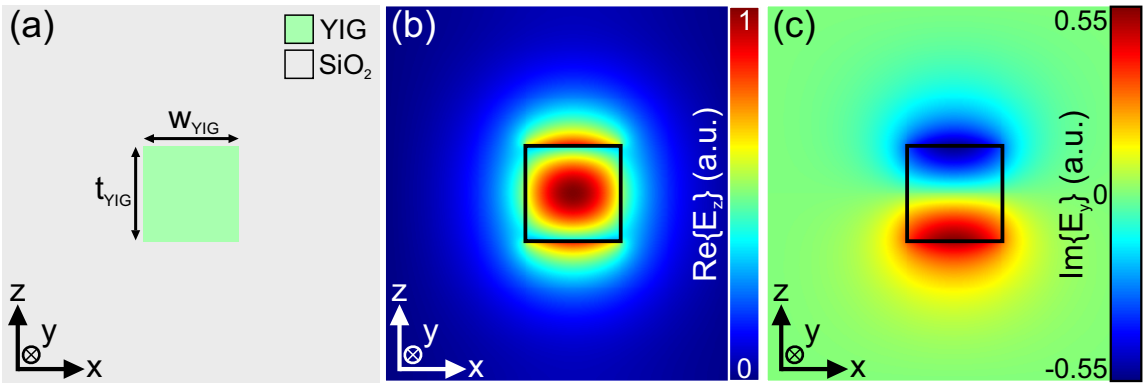

Figure 2.16: (a) Exemplary waveguide geometry consisting of a YIG core having $t_{YIG} = w_{YIG} =$ 500nm, embedded in SiO_2. (b) The profile of the transverse component (E_z) of the fundamental TM mode of the structure. (c) The profile of the longitudinal component (E_y) of the fundamental TM mode of the structure.

solutions to this problem have been proposed, such as employing the MO material as the waveguide cladding and relying on the evanescent tail of the mode [75], incorporating MO materials with intrinsic θ_{spec} properties of opposing sign on each side of the field node [76, 77], adjusting the position of the MO region of the waveguide core [74], or simply employing highly asymmetric field profiles, such as surface plasmons [73].

From the preceding arguments, the NRPS is a highly polarization dependent nonreciprocal effect. The aforementioned example explored the scenario where the NRPS is applied to a TM mode, but in order to observe the effect in a TE mode, the orientation of **M** needs to be changed (i.e., **M** should be oriented along the z-axis). However, the behaviour is the same.

2.3.3 Reflection Effects

While the magneto-optic effects in transmission geometries are the most applicable for integrated optics, it should be noted that a number of MO effects manifest upon the reflection of a lightwave from a magnetized medium. This family of effects are known as the Kerr effects, as they were first discovered by John Kerr in 1877 [78] and 1878 [79]. These effects are delineated by their respective geometries, depicted

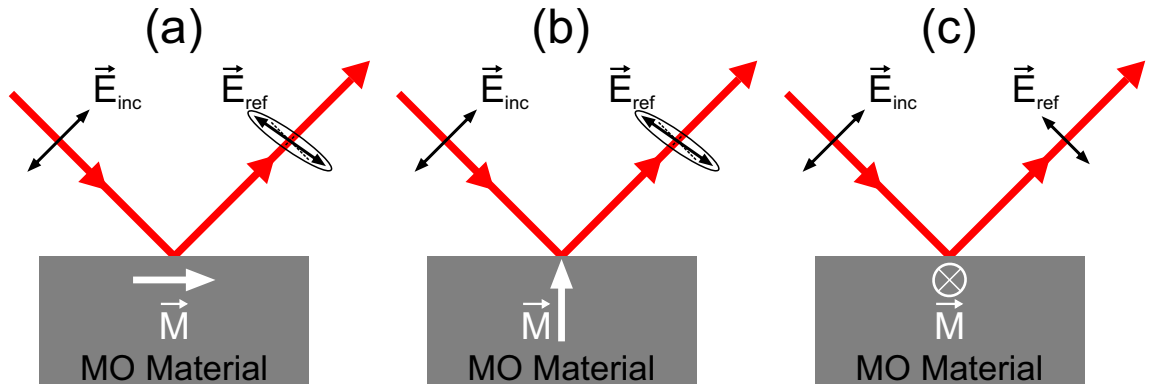

Figure 2.17: Schematic illustration of the geometry for the (a) longitudinal Kerr effect, (b) polar Kerr effect, and (c) transverse Kerr effect. $\mathbf{E}_{inc}$ is the incident electric field, and $\mathbf{E}_{ref}$ is the reflected electric field.

in Fig. 2.17. Due to their being reflective effects, the Kerr effects do not manifest in waveguide geometries. However, they have been the focus of a vast number of studies on the interplay between magneto-optics and plasmonic nanostructures such as gratings [80], magnetoplasmonic crystals [81], and nanoparticle assemblies [82], as well as continuous Au/Co/Au trilayer films excited via the Kretschmann configuration [83] or via arrays of Au nanoantennas [84]. These MO effects are briefly described below for completeness of the discussion.

Longitudinal and Polar Magneto-Optical Kerr Effect

The longitudinal Kerr effect requires the sample to be magnetized parallel to both the sample surface and the plane of incidence, as depicted in Fig. 2.17(a). A linearly polarized wave incident on the surface in this geometry will reflect, and the polarization of the reflected wave will be rotated slightly, and also exhibit a small degree of ellipticity [16, 64].

The polar Kerr geometry, depicted in Fig. 2.17(b) yields an analogous effect. Here, the magnetization is oriented perpendicular to the sample surface, but still within the plane of incidence of the incident linearly polarized wave. Again, this geometry yields a slight rotation of the incident polarization, as well as induces slight ellipticity upon

reflection from the interface [16, 64].

In both configurations, the modifications of the reflected polarization state occur because the incident wavevector has a component along the direction of magnetization. The effect of the magnetization works to couple the TE and TM components of the incident wave, as is evident through the nonzero elements of the permittivity tensor (Eqn. 2.62).

Transverse Magneto-Optical Kerr Effect

The final Kerr effect, denoted as the transverse Kerr effect, is depicted in Fig. 2.17(c). In this configuration, the MO material is magnetized parallel to the sample surface, but perpendicular to the plane of incidence. Here, the incident wavevector does not have any component along the direction of the magnetization. As such, Eqn. 2.62 does not induce coupling between the different field components. Since the TE and TM components are uncoupled, no polarization rotation or induced ellipticity is observed. Instead, this effect manifests as a small change in the reflected intensity of the lightwave [16, 64].

2.3.4 Magnetoplasmonic Waveguides

To realize plasmonic circuitry, reliable methods of modulating optical signals at the nanoscale need to be developed. Since the evanescent plasmonic signals confined at a metal-dielectric interface are inherently sensitive to the properties of the surrounding materials, previous explorations have examined altering the properties of plasmonic signals through the use of carrier accumulation via applied voltages [85], gain media [86], two-photon-absorption [87], thermo-optic effects [88], or electro-optic effects [56], to name just a few. Another method which has received less widespread attention was proposed in 1972 by Chiu and Quinn, whereby MO properties were used to modulate a plasmonic response [89]. This mechanism is intriguing, since it can facilitate both

nonreciprocal phenomena and highly nonlinear temporal dynamics, leading to the development of unique and fundamental optical components in the nanoplasmonic regime.

MO effects cannot be readily observed within typical plasmonic structures, as the metals commonly used for plasmonic applications, such as Au, Ag, Cu, or Al, exhibit extremely weak magnetic responses. Large magnetic flux densities, on the order of several Tesla to several tens of Tesla, are required to induce a response, and these fields are impractical for nanoscale applications. To remedy this, two different material systems are of interest to study these effects: ferromagnetic metals, such as Fe, Ni, and Co, and magnetic garnets such as YIG and its derivatives. Both sets of materials can generate appreciable magneto-optic effects with weak applied magnetic fields, on the order of a few tens of millitesla. However, while Fe, Ni, and Co films are easy to fabricate, these materials impart high optical losses. To mitigate this, ferromagnetic films are often embedded within noble metal films, as shown in Fig. 2.18, such that the evanescent field of the plasmon penetrates into the magnetic film, but the plasmon propagates on the noble metal [90, 91]. Conversely, a YIG crystal is characterized by a transparency window around the telecommunications wavelength band, and exhibits excellent magneto-optic properties. The YIG crystal's MO response can be enhanced by introducing rare-earth ions (e.g. bismuth, cerium, etc.) into its lattice [64]. Fabrication of such films typically requires liquid phase epitaxy on a gadolinium gallium garnet (GGG) substrate, and recent work has demonstrated fabrication with pulsed laser deposition [92] and sputtering [93]. Both material systems have found applications in recent studies.

Plasmonic modes, such as those at the interface of a metal and dielectric where one constituent is magnetic, are ideal for studies of the NRPS [73]. As shown in Section 2.2.1, SP modes consist of two electric field components: one transverse, and one longitudinal. Due to the phase difference between these components, these

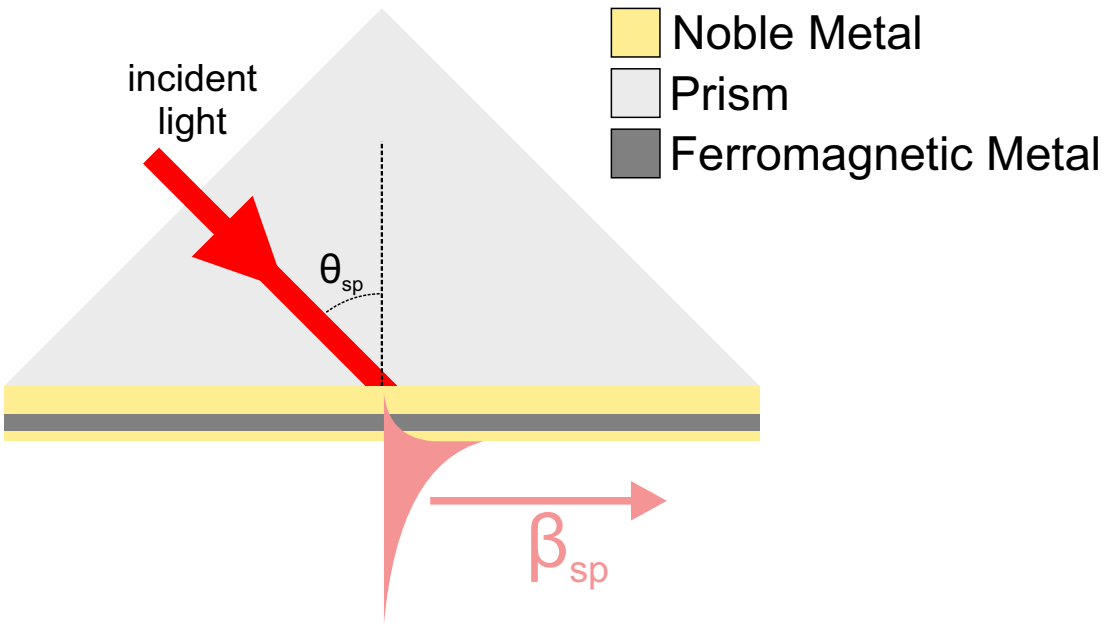

Figure 2.18: Noble metal/ferromagnetic metal/noble metal trilayer structure employed to incorporate ferromagnetic materials into plasmonic structure without significant losses. The trilayer film is shown here on the surface of a prism in the Kretschmann configuration.

modes are elliptically polarized, and if the dielectric is the MO material, a large fraction of the modal power will be contained by this region. The evanescent nature of SPs also ensures a high degree of asymmetry, and a 1D surface plasmon does not exhibit the detrimental inverted mirror symmetry that causes cancellation of the NRPS. Considering these factors, SPs are the quintessential structure for realizing nonreciprocal behaviour in nanoscale environments.

The first work examining propagating plasmons in MO structures was that of Chiu and Quinn, who presented a detailed theoretical analysis of plasmon propagation at the interface of a dielectric and a magnetic metal [89]. One key result of their analysis showed that under the presence of a transverse static magnetic field, the plasmon dispersion relation splits into two branches (i.e. it has two solutions), corresponding to forward and backward propagating modes. One can define a NRPS between them via Eqn. 2.74. The branches are found to interchange upon reversal of the magnetic field direction, or inversion of the structure. This effect breaks the Lorentz reciprocity condition, and is nonreciprocal.

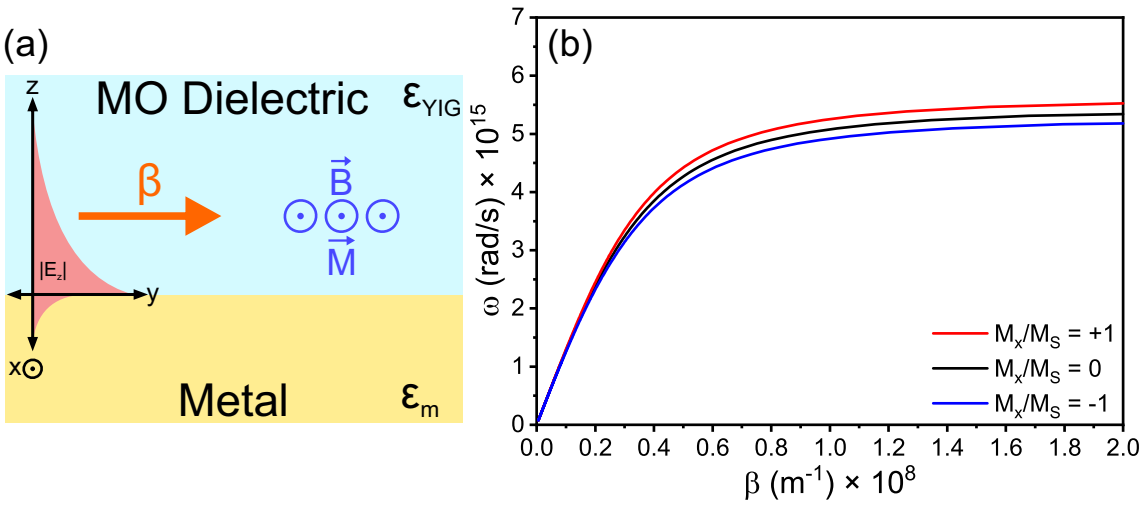

Figure 2.19: (a) Schematic illustration of a plasmon propagating at the interface between a noble metal and a MO dielectric, under a transverse magnetic field, and correspondingly, a transverse magnetization. (b) Dispersion relation of the plasmon supported at the interface between YIG and Au. To illustrate the effect, the value of the Faraday rotation has been exaggerated to $\theta_F = 20°/\mu$m. The dispersion relation splits into two branches when a transverse magnetization is present.

While this was a useful depiction of the concept of implementing MO effects into plasmonic, due to the losses, a more practical structure would consist of a noble metal and a MO dielectric, such as YIG. This configuration has been employed in the design of a reflectivity modulator in the past [94]; however, incorrect assumptions in the dispersion derivation resulted in an overestimation of the attainable reflectivity modulation. The geometry is depicted in Fig. 2.19(a). In this case, the dispersion relation of the plasmon can be derived in a similar manner to that presented in Section 2.2.1. This simple configuration does not facilitate mechanisms of phase matching required to implement the Faraday effect in a longitudinally magnetized configuration. To realize such an effect, a more complicated structure is required, as will be discussed in Chapter 4. Thus, here we consider only a transverse magnetization (i.e., along the x-axis), and the magnetization is assumed to be saturated. In this configuration, the YIG will possess an asymmetric permittivity tensor (Eqn. 2.62), with a diagonal permittivity of ε_{YIG} (and diagonal refractive index n_{YIG}), and where the only nonzero component of the gyration vector is g_x. Following this procedure

outlined in Section 2.2.1, the dispersion can be derived as follows:

$$\begin{aligned}
&\beta_{SP}^2\left[\left(\varepsilon_{YIG}^2-g_x^2\right)^2-\varepsilon_m^2 g_x^2-\varepsilon_{YIG}^2\varepsilon_m^2\right] \\
&\pm 2\beta_{SP}\varepsilon_m g_x\sqrt{\beta_{SP}^2\varepsilon_{YIG}^2\varepsilon_m^2-\varepsilon_{YIG}\varepsilon_m^2\beta_0^2\left(\varepsilon_{YIG}^2-g_x^2\right)} \\
&=\varepsilon_m\beta_0^2\left(\varepsilon_{YIG}^2-g_x^2\right)^2-\varepsilon_{YIG}\varepsilon_m^2\beta_0^2\left(\varepsilon_{YIG}^2-g_x^2\right)
\end{aligned} \tag{2.75}$$

This is the dispersion relation without any assumptions or removal of higher order terms. There are clearly two solutions for the propagation constant when $M_x \neq 0$, and Eqn. 2.75 can be shown to reduce to the standard SP dispersion relation of Eqn. 2.48 with the removal of the magnetization (i.e. setting $g_x = 0$). The solutions to Eqn. 2.75 are plotted in Fig. 2.19(b), considering the interface between a YIG material with $n_{YIG} = 2.3$, and Au, whose optical constants are calculated via the Drude model (Eqn. 2.28), with parameters taken from Ref. [25]. Since the NRPS is inherently a weak effect, and in order to illustrate the effect, an exaggerated value for the Faraday rotation ($\theta_F = 20°/\mu$m) is employed in the calculation. Clearly, the introduction of the magnetization causes the SP dispersion relation to split into two branches around the isotropic plasmon dispersion. These two branches correspond to the propagation constants of forward and backward propagating plasmons, or analogously, plasmons propagating in media where the direction of the magnetization is reversed. This structure will exhibit the NRPS, as there is a difference between these two values. This simple geometry illustrates that SPs are capable of supporting nonreciprocity.

Dispersion relations have been derived for other common plasmonic waveguide geometries under a magnetic bias. A MIM waveguide consisting of Ag/YIG/Ag was analyzed by Nikolova and Fisher [95]. Such a structure supports both a SM and an AM. The propagation constants as a function of YIG thickness, T, and magnetization are depicted in Fig. 2.20(a). The presence of a transverse magnetic field shifts the modal propagation constant for both modes, and the propagation constant can be modulated by switching the magnetic field on and off. However, due to the symmetry

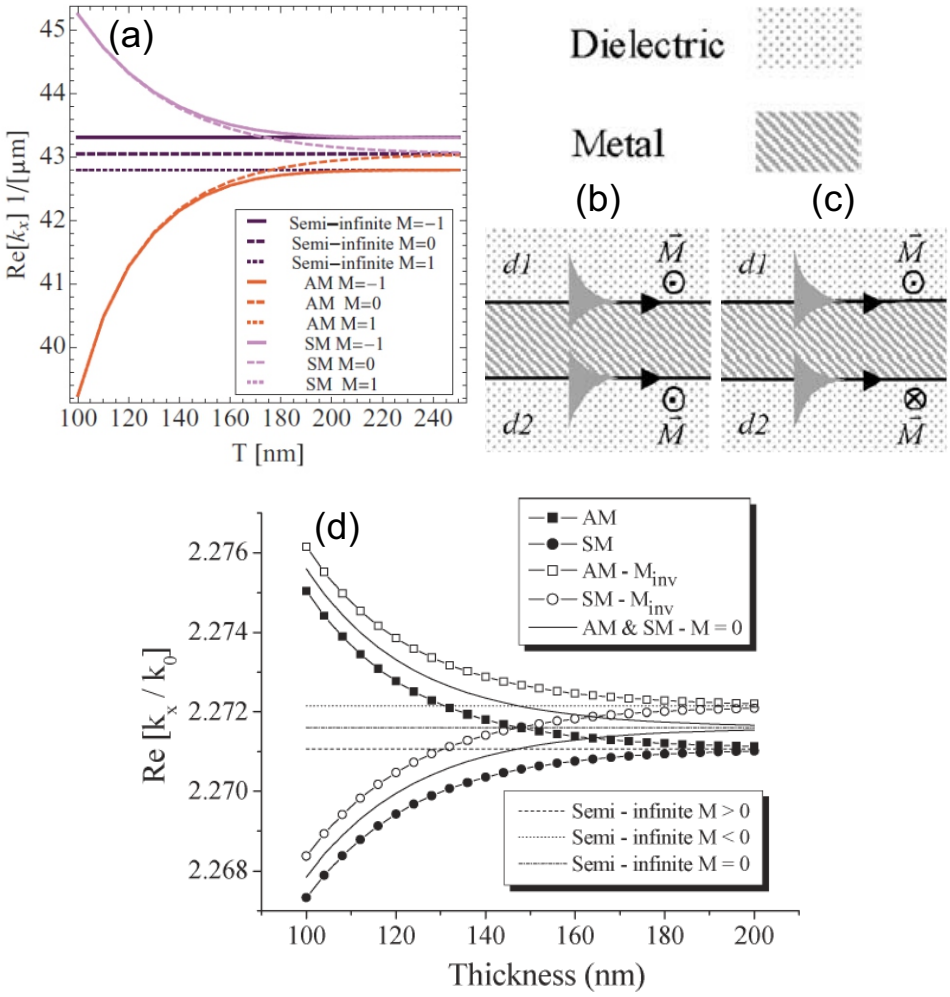

Figure 2.20: (a) Plot of the SM and AM propagation constants in an Ag/YIG/Ag MIM waveguide, as a function of the YIG thickness, T. Note that the curves $M = 1$ and $M = -1$ are indistinguishable for either the SM or AM, due to the symmetry. Image adapted from [95] (b) Schematic of a YIG/Au/YIG IMI waveguide with the same magnetization in each layer. (c) Schematic of a YIG/Au/YIG IMI waveguide with opposite magnetization in the two YIG layers. (d) Plot of the SM and AM propagation constants in a YIG/Au/YIG IMI waveguide, where the YIG layers are oppositely magnetized to break the symmetry. Note that the SM and AM propagation constants vary upon changing M. Images in (b)-(d) were adapted from [96].

of the structure, and both the SM and the AM exhibit a node and polarity change within one of the two electric field components, and the nonreciprocal splitting of the dispersion relation is not present for either mode, i.e., $\beta_{fwd} = \beta_{bwd}$. This is shown in Fig. 2.20(a), where the curves for $\mathbf{M} = 1$ and $\mathbf{M} = -1$ are indistinguishable for both the SM and the AM. Despite the reciprocal nature of the propagation constants in this structure, the application of a magnetic field was shown to modify the field

distribution within the guide, and hence was used to manipulate the emitted far-field radiation pattern.

Nonreciprocity can be achieved in an MIM waveguide if asymmetry is introduced into the structure somehow. One proposed solution incorporates thin Co layers within the Ag films comprising an Ag/SiO_2/Ag MIM waveguide [97]. Here, it was shown that the presence of exactly one Co film, present in either the top or the bottom Ag layer would induce nonreciprocity of the propagation constant. Including Co in both layers cancels out the effect.

In a similar manner, IMI plasmonic waveguides have been examined, and they exhibit the same mode symmetry concerns [96]. Consider a YIG/Au/YIG IMI waveguide: this waveguide supports both SM and AM modes. When both YIG films are magnetized in the same direction (Fig. 2.20(b)), a shift in the propagation constant for the SM and AM modes is realized, but the mode's symmetry cancels out any nonreciprocity. This is remedied by introducing asymmetry in either the magnetization profile or to the structure itself. By oppositely magnetizing the two YIG layers (Fig. 2.20(c)), the cancellation effect can be avoided. This condition is depicted by the plot of propagation constant verses the Ag film thickness, shown in Fig. 2.20(d). Due to the asymmetry, here the SM and AM experience different propagation constants upon inversion of $\mathbf{M}$, and a NRPS can be defined between them. In a similar manner, replacing one of the YIG films with an isotropic dielectric with a different refractive index will alter the field penetration into the different layers, and hence reduce the cancellation of the nonreciprocal effect. Clearly, asymmetry is key in achieving nonreciprocity in plasmonic waveguides.

Another proposed solution extends the concept of the simple metal-dielectric interface. Davoyan and Engheta suggested the use of cylindrical Au nanowires coated in a YIG film for a possible magnetoplasmonic waveguide architecture [98], as shown in Fig. 2.21(a). In this configuration, the nanowire supports the propagation of a

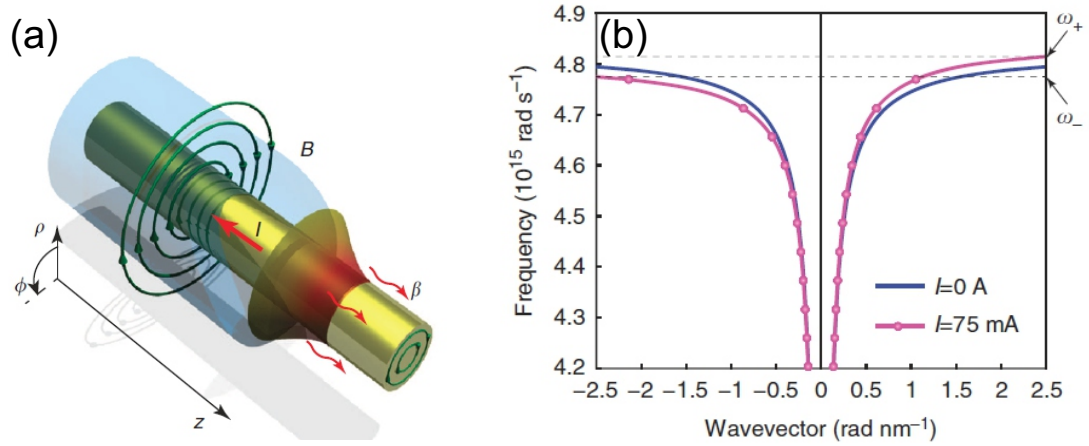

Figure 2.21: (a) Schematic depiction of an Au nanowire waveguide coated in a YIG cladding. (b) Dispersion relation of the waveguide depicted in (a). Images adapted from [98].

radially symmetric plasmonic mode. Passing current through the nanowire generates an azimuthally symmetric magnetic field circulating around the wire, and as such, this field is always transverse to the plasmon. Such a configuration again exhibits the splitting of the plasmon dispersion relation, as depicted in Fig. 2.21(b), and supports nonreciprocal propagation. Additionally, the authors identified a spectral region in which excited plasmons can propagate in only one direction, suggesting interesting applications in on-chip nonreciprocal devices.

2.3.5 Magnetization Dynamics

Magneto-optic effects are unique in that they depend directly on the magnetization state and orientation within a material. As the magnetization of a material is the aggregation of many magnetic moments, one can dynamically modulate the MO effects through the unique temporal characteristics introduced by influencing these magnetic moments with an applied magnetic field. Such attributes can facilitate novel optical modulator designs

It is crucial to understand the temporal evolution of $\mathbf{M}$ within an applied magnetic field. These dynamics can be derived as follows [62, 63, 99]. Equation 2.52 illustrates

that an applied magnetic field exerts a torque on the electron's spin magnetic moment. Classically, this torque defines the time derivative of the spin angular momentum:

$$\frac{d\mathbf{S}}{dt} = \mu_0 \boldsymbol{\mu_s} \times \mathbf{H} \tag{2.76}$$

Equation 2.51 provides a relation between $\mathbf{S}$ and $\boldsymbol{\mu_s}$, and substitution of this relation into Eqn. 2.76 yields the following:

$$\frac{d\boldsymbol{\mu_s}}{dt} = -\mu_0 \frac{gq}{2m_e} \boldsymbol{\mu_s} \times \mathbf{H} \tag{2.77}$$

which is further simplified by defining the gyromagnetic ratio, γ_0, as:

$$\gamma_0 = \frac{gq}{2m_e} \tag{2.78}$$

Analogous to the definition of polarization in terms of electric dipole moments, the magnetization of a medium with N magnetic moments is defined as $\mathbf{M} = N\boldsymbol{\mu_s}$. As such, Eqn. 2.77 can be written in terms of magnetization:

$$\frac{d\mathbf{M}}{dt} = -\mu_0 \gamma_0 (\mathbf{M} \times \mathbf{H}) \tag{2.79}$$

Here, $\mathbf{M} = \langle M_x, M_y, M_z \rangle$ is the three dimensional magnetization vector having a constant magnitude, $|\mathbf{M}| = M_S$, which defines the saturation magnetization of the material. The precessional motion prescribed by Eqn. 2.79 is analogous to that of a spinning top, and this motion is schematically illustrated in Fig. 2.22(a). Here, the applied magnetic field is considered to be a static field oriented along the $+z$-axis, H_z, and $\mathbf{M}$ is initially at an angle θ to the applied magnetic field. Clearly, the torque exerted on $\mathbf{M}$ by the applied $\mathbf{H}$ causes $\mathbf{M}$ to precess along a circular path around the z-axis. The frequency at which this precessional motion occurs is denoted as the

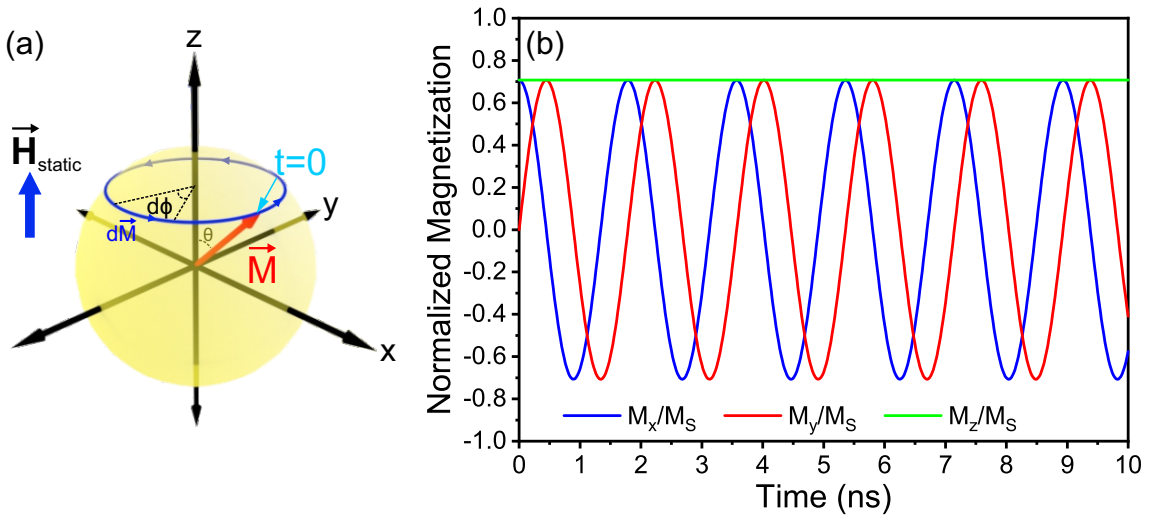

Figure 2.22: (a) Schematic of the trajectory of **M** undergoing basic precessional motion. (b) Plot of the components of **M** comprising the trajectory depicted in (a). Here, $\mu_0 H_z = 20\text{mT}$, and $\theta = 45°$.

Larmor angular frequency, ω_L, or the Larmor frequency, ν_L, where $\omega_L = 2\pi\nu_L$. The value of this frequency can be determined from the geometry depicted in Fig. 2.22(a). The angular velocity of the motion is denoted by:

$$\frac{d\phi}{dt} = \omega_L \tag{2.80}$$

where ϕ is the angle subtended by the motion of **M**. Furthermore, **M** will traverse an arc of length $|d\mathbf{M}|$ over an angle $d\phi$ in a time dt. As such, this arc length is:

$$|d\mathbf{M}| = M_S \sin(\theta) d\phi \tag{2.81}$$

Similarly, one can take the magnitude of Eqn. 2.79 to find:

$$|d\mathbf{M}| = \mu_0 \gamma_0 M_S H_z \sin(\theta) dt \tag{2.82}$$

Equating Eqns. 2.81 and 2.82 and rearranging yield a relation for the Larmor angular

frequency:

$$\frac{d\phi}{dt} = \mu_0 \gamma_0 H_z = \omega_L \tag{2.83}$$

which can be further manipulated to find the Larmor frequency:

$$\nu_L = \frac{\mu_0 \gamma_0 H_z}{2\pi} \tag{2.84}$$

The magnitude of the components of $\mathbf{M}$ are plotted in Fig. 2.22(b) for an applied field of $\mu_0 H_z = 20\text{mT}$, and an initial angle of $\theta = 45°$. The oscillatory precessional motion is evident by the oscillating M_x and M_y components with a phase difference of $\pi/2$, and the frequency of oscillation is $\nu_L = 560\text{MHz}$, in agreement with the value predicted by Eqn. 2.84. Figure 2.22(b) also depicts that the amplitude of the M_x and M_y oscillations does not change, and the M_z component is a constant, implying that this precessional motion will persist indefinitely, and that $\mathbf{M}$ will not align with the applied field, H_z. Clearly, this is not physical as experimental evidence shows that $\mathbf{M}$ will align with the applied field. Some form of damping must be present.

While rigorous quantum mechanical analysis can model the relaxation and damping based on interactions of the spin with other spin magnetic moments and the crystal lattice, it is more instructive to modify Eqn. 2.79 by adding a phenomenological damping term. There have been several forms of the damping term proposed. One of the most commonly employed models for the damped precession of magnetization in ferromagnetic and ferrimagnetic materials is was originally proposed by Landau and Lifshitz in 1935 [100], and modified by Gilbert in 1956 [101]:

$$\frac{d\mathbf{M}}{dt} = -\mu_0 \gamma_0 (\mathbf{M} \times \mathbf{H}) + \frac{\alpha}{M_S} \left(\mathbf{M} \times \frac{d\mathbf{M}}{dt} \right) \tag{2.85}$$

Here, α is the Gilbert damping parameter, which is a dimensionless, material specific parameter describing the rate at which $\mathbf{M}$ relaxes and aligns to the applied magnetic

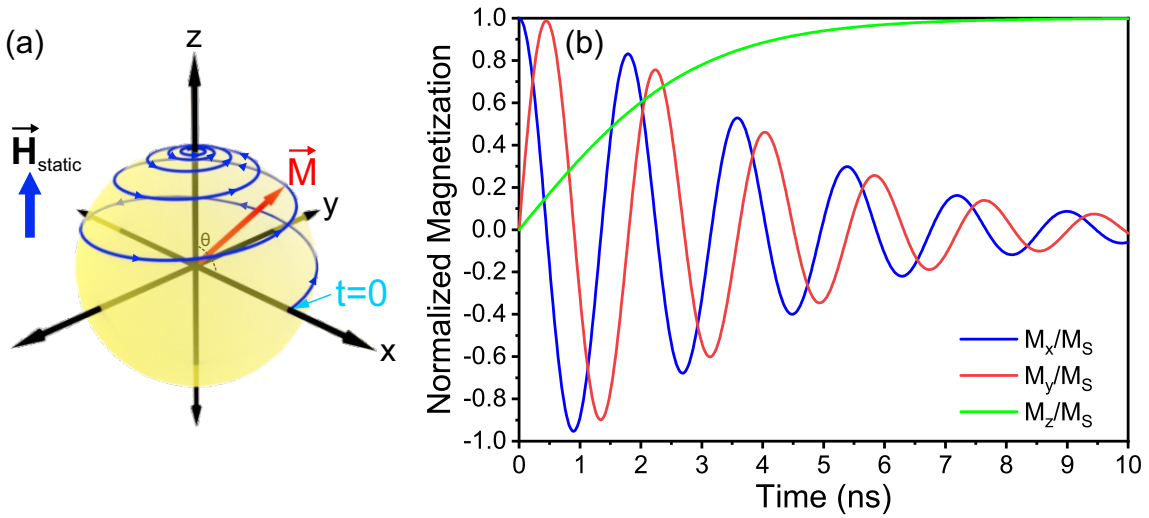

Figure 2.23: (a) Schematic of the trajectory of **M** as prescribed by the LLG model. (b) Plot of the components of **M** comprising the trajectory depicted in (a). Here, $\mu_0 H_z = 20\text{mT}$, $\alpha = 0.1$, and $\theta = 90°$ initially. Note that the value of α is exaggerated by several orders of magnitude to illustrate the phenomenon.

field. Through the application of several vector identities, as well as some algebraic manipulation, Eqn. 2.85 can be simplified to the following form:

$$\frac{d\mathbf{M}}{dt} = -\frac{\mu_0\gamma_0}{1+\alpha^2}(\mathbf{M}\times\mathbf{H}) - \frac{\mu_0\gamma_0\alpha}{M_S(1+\alpha^2)}[\mathbf{M}\times(\mathbf{M}\times\mathbf{H})] \tag{2.86}$$

This equation is known as the Landau-Lifshitz-Gilbert (LLG) model of magnetization dynamics. Equation 2.86 prescribes a precessional trajectory for the magnetization that eventually relaxes and aligns with the applied magnetic field, as is dictated by experiment. An exemplary path for the temporal evolution of **M** is illustrated in Fig. 2.23(a). Again, we can consider the example of an applied magnetic field of $\mu_0 H_z = 20\text{mT}$, and an initial condition of $\theta = 90°$. Taking $\alpha = 0.1$, we see that **M** follows a spiraling path until it aligns with the applied magnetic field. Since the magnitude of **M** is conserved here, this vector traces out its motion on the surface of a sphere of radius M_S.

The temporal evolution of the individual components of the magnetization in this scenario are presented in Fig. 2.23(b). In contrast to the previous example, it is clear

that the oscillatory amplitude of M_x and M_y decrease, corresponding to a reduction of θ as $\mathbf{M}$ aligns with the applied magnetic field. Furthermore, M_z/M_S approaches 1, as is expected when the field and magnetization are aligned. In this example, the Gilbert damping factor of $\alpha = 0.1$ was taken to illustrate the motion prescribed by Eqn. 2.86. In practical magnetic materials, α is lower by several orders of magnitude, causing the relaxation to occur over much longer time scales.

2.4 Nonlinear Optics and Carrier Dynamics in Si

2.4.1 Nonlinear Polarization

In Section 2.1.1, the materials considered were linear and isotropic dielectrics, where the polarization vector was taken to be linearly related to the electric field. In general, and especially when considering strong electric fields, the polarization response of a material can be expanded as a power series of the electric field [102, 103]:

$$\mathbf{P}(t) = \varepsilon_0[\chi^{(1)}\mathbf{E}(t) + \chi^{(2)}\mathbf{E}^2(t) + \chi^{(3)}\mathbf{E}^3(t) + ...] \tag{2.87}$$

where $\chi^{(n)}$ represents the n^{th} order susceptibility. In this notation, $\chi^{(1)} = \chi_e$ for congruence with Eqn. 2.7. The polarization can be decomposed into a linear component, $\mathbf{P}_L(t)$, and a series of higher n^{th} order nonlinear polarization terms, $\mathbf{P}_{NL}^{(n)}(t)$:

$$\mathbf{P}(t) = \mathbf{P}_L(t) + \mathbf{P}_{NL}^{(2)}(t) + \mathbf{P}_{NL}^{(3)}(t) + ... \tag{2.88}$$

where $\mathbf{P}_{NL}^{(n)}(t) = \varepsilon_0\chi^{(n)}\mathbf{E}^n(t)$. These nonlinear terms can give rise to new frequency components, consisting of sums and differences of the input frequencies. The magnitude of $\chi^{(n)}$ decreases rapidly as the order increases, requiring stronger electric fields for higher order terms to become significant.

One of the most prevalent and extensively explored materials for use in integrated

optics is silicon. Silicon provides a number of advantages for photonic applications. Namely, silicon is abundant, cost-effective, and it is virtually lossless at wavelengths within the telecommunications band, making it ideal for incorporation into chip-scale integrated optics. Being widely employed in the microelectronics industry, Si fabrication processes have been extensively developed and optimized. Correspondingly, silicon photonics are the dominant chip-scale platform to date.

Exploiting nonlinear effects within silicon is intriguing since silicon is an indirect band gap semiconductor, and cannot generate light via stimulated emission. The only way to generate coherent light within Si, particularly within the visible spectral range, is via inducing optical nonlinearities. Silicon itself has a centrosymmetric crystal structure. This symmetry implies that all even-ordered nonlinear polarization terms are identically zero, since $\chi^{(2)}, \chi^{(4)}, ... = 0$. While centrosymmetric crystals do not support even-ordered optical nonlinearities, they do exhibit odd-ordered ones, as will any material with a sufficiently strong applied field. Because of this, the most important nonlinear term within the silicon material system is the third order nonlinear polarization, $\mathbf{P}_{NL}^{(3)}$. This is due to the fact that silicon inherently possesses a high $\chi^{(3)}$ coefficient, measured in the range of $1.4 - 3.9 \times 10^{-19}\text{m}^2/\text{V}^2$ [104, 105], which is 670-1884 times greater than that of SiO_2, commonly utilized in nonlinear fibers. Moreover, while higher odd-ordered terms could be present, the optical fields required to observe them would surpass the damage threshold of silicon (reported in the range of $4 \times 10^7\text{V/m}$ [106] to $> 5 \times 10^9\text{V/m}$ [107]), and are difficult to observe. As such, much of the nonlinear optics research involving silicon focuses on the $\chi^{(3)}$ processes.

In nonlinear optical applications, the high electric fields (approximately $> 10^6\text{V/m}$ [24]) required to observe nonlinear effects are difficult to achieve with continuous-wave (CW) laser output. In practice, these fields are achieved by employing laser pulses having durations ranging from a few femtoseconds to several nanoseconds. By

reducing the time in which the optical energy is delivered, peak powers well in excess of several kilowatts are obtained. In this case, the electric field, and correspondingly, the optical intensity, $I_{opt}(t)$, are a function of time. Typically this takes the form of Gaussian or sech^2 shaped pulses. This method allows nonlinear phenomena to be readily observed within a practical setting.

2.4.2 $\chi^{(3)}$ Nonlinear Optical Processes

Since the n^{th} order nonlinear polarization term is produced by the n^{th} power of the optical electric field, this gives rise to polarization terms at new frequencies. Specifically, this process can generate sums, differences, and harmonics of the input frequencies. The nonlinear polarization also be formulated in terms of a sum of the nonlinear polarizations generated at each of these frequencies [102]:

$$\mathbf{P}_{NL}^{(n)}(t) = \sum_j \mathbf{P}_{NL}(\omega_j) e^{-i\omega_j t} + c.c. \tag{2.89}$$

where ω_j are the generated frequencies, $\mathbf{P}_{NL}(\omega_j)$ is the amplitude of the nonlinear polarization at that frequency, and *c.c.* denotes the complex conjugate. In the simplest case, the electric field can be taken as a monochromatic electric field of the form $\mathbf{E}(t) = \frac{1}{2}\mathbf{E}(\omega)e^{-i\omega t} + c.c.$ Accordingly, the $\chi^{(3)}$ interaction only produces nonlinear polarization at two frequencies, ω and 3ω, and can be written as follows:

$$\mathbf{P}_{NL}^{(3)}(t) = \mathbf{P}_{NL}(\omega)e^{-i\omega t} + \mathbf{P}_{NL}(3\omega)e^{-i3\omega t} + c.c. \tag{2.90}$$

Substituting the expression for the electric field into the nonlinear polarization equation (Eqn. 2.87) and simplifying yields $\mathbf{P}_{NL}(3\omega)$ and $\mathbf{P}_{NL}(\omega)$ [102]:

$$\mathbf{P}_{NL}(3\omega) = \frac{1}{4}\varepsilon_0 \chi^{(3)} \mathbf{E}(\omega)^3 \tag{2.91}$$

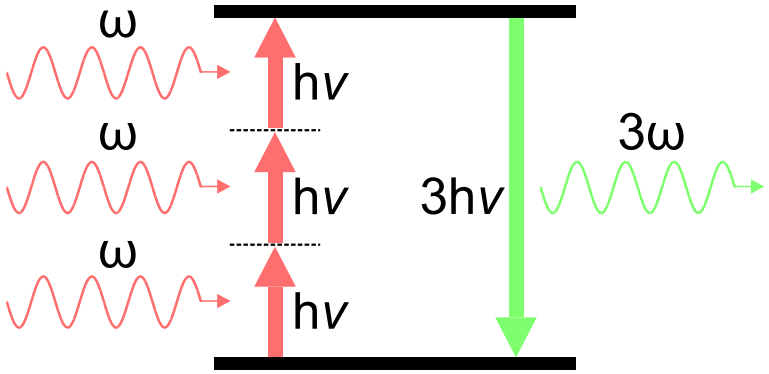

Figure 2.24: Illustration of the third-harmonic generation nonlinear process. In Si, this process is responsible for converting light within the telecom band (i.e. $\lambda = 1550$nm) into light in the visible region of the spectrum ($\lambda = 516.7$nm).

$$\mathbf{P}_{NL}(\omega) = \frac{3}{4}\varepsilon_0\chi^{(3)}|\mathbf{E}(\omega)|^2\mathbf{E}(\omega) \tag{2.92}$$

These two equations describe the dominant nonlinear optical interactions in silicon. Equation 2.91 represents third-harmonic generation (THG), while Eqn. 2.92 describes both the optical Kerr effect and two-photon absorption (TPA).

Third-Harmonic Generation

Third-harmonic generation is one of the 3^{rd} order nonlinear optical processes responsible for generating new frequencies of light within a nonlinear medium. The process is illustrated in Fig. 2.24. Three incident photons, each possessing a frequency ω and an energy $h\nu$ (where $h = 6.626 \times 10^{-34}$J·s is Planck's constant and ν is the frequency), simultaneously excite an electron into a virtual state. When the virtual state immediate decays as the electron relaxes, it results in the emission of a photon having a total energy of $3h\nu$. The emitted photon is characterized by a frequency of 3ω, or accordingly, a wavelength 1/3 of the exciting radiation.

Due to its indirect bandgap, nonlinear processes such as THG are the only mechanism capable of generating coherent visible light within Si. As an example, considering the primary telecommunications wavelength of $\lambda_\omega = 1550$nm as the fundamental excitation, the generated third-harmonic radiation will be in the green spectral region,

at $\lambda_{3\omega} = 516.7$nm. Due to the magnitude of the nonlinear interaction as well as the linear absorption of silicon in the visible regime, this effect typically requires high optical electric fields to observe.

The Kerr Effect, Self-Phase Modulation, and Two Photon Absorption

Equation 2.92 describes nonlinear polarization at the fundamental frequency of ω, where the total polarization is the linear superposition of both the linear and nonlinear polarization terms:

$$\mathbf{P}_{TOT}(\omega) = \mathbf{P}_L(\omega) + \mathbf{P}_{NL}(\omega) \tag{2.93}$$

$$\mathbf{P}_{TOT}(\omega) = \varepsilon_0 \chi^{(1)} \mathbf{E}(\omega) + \frac{3}{4} \varepsilon_0 \chi^{(3)} |\mathbf{E}(\omega)|^2 \mathbf{E}(\omega) \tag{2.94}$$

Noting the similarities with Eqn. 2.7, one can define an effective susceptibility, χ_{eff}, relating the polarization with the oscillating electric field [102, 108]:

$$\chi_{eff} = \chi^{(1)} + \frac{3}{4} \chi^{(3)} |\mathbf{E}(\omega)|^2 \tag{2.95}$$

In general, the $\chi^{(3)}$ coefficient is a complex quantity, $\chi^{(3)} = \chi_R^{(3)} + i\chi_I^{(3)}$, where $\chi_R^{(3)}$ and $\chi_I^{(3)}$ are the real and imaginary components, respectively. The refractive index and absorption coefficient of the medium can be considered as being perturbed by a factor that is proportional to $|\mathbf{E}(\omega)|^2$, or more conveniently, proportional to the intensity, I_{opt}, since $I_{opt} \propto |E|^2$. The refractive index and absorption coefficient take the form [109–111]:

$$n = n_0 + n_2 I_{opt} + i \frac{\lambda}{4\pi} \left(\alpha_0 + \beta_{TPA} I_{opt} \right) \tag{2.96}$$

where n_0 and α_0 are the linear refractive index and absorption coefficient, respectively, n_2 is the nonlinear refractive index, and β_{TPA} is the TPA coefficient. Here,

n_2 is related to $\chi_R^{(3)}$, while β_{TPA} is proportional to $\chi_I^{(3)}$. Further simplification yields expressions for these two quantities:

$$n_2 = \frac{3\chi_R^{(3)}}{4c\varepsilon_0 n_0^2} \tag{2.97}$$

$$\beta_{TPA} = \frac{3\pi\chi_I^{(3)}}{\varepsilon_0 c n_0 \lambda} \tag{2.98}$$

The parameter n_2 is responsible for the optical Kerr effect, also known as the intensity dependent refractive index. As the intensity of the incident wave increases, the refractive index of the medium increases as well. This leads to a self-focusing effect, where the high intensity region at the center of a focused laser beam experiences a higher refractive index than the lower intensity tail of the focused beam. The effect is analogous to producing a graded index lens, and acts to further focus the beam.

Changing the refractive index modifies the phase of the wave propagating through the medium in a process called self-phase modulation (SPM). This phenomenon is especially important when considering the propagation of optical pulses through nonlinear media. Considering an optical pulse, where the phase of the carrier wave is of the form $\varphi(t) = \beta y - \omega t$ and $\beta = 2\pi n/\lambda$ is the propagation constant, the phase of the wave subject to the intensity-dependent refractive index after traveling a distance $y = L$ is [112]:

$$\varphi(t) = n_0 \frac{2\pi}{\lambda} L + \frac{2\pi}{\lambda} n_2 I_{opt}(t) L - \omega t \tag{2.99}$$

This nonlinear modulation of the phase affects the instantaneous frequency of the pulse:

$$\omega_{inst} = \frac{d\varphi}{dt} = \frac{2\pi}{\lambda} n_2 L \frac{dI_{opt}(t)}{dt} - \omega \tag{2.100}$$

Therefore, the effect of the nonlinear refractive index is to introduce time varying modulation of the instantaneous frequency, which leads to both broadening of the pulse spectrum, as well as the introduction of chirp.

The imaginary part of $\chi^{(3)}$ gives rise to TPA within silicon, specifically around the telecommunications wavelength of $\lambda = 1550$nm. The band structure of Si is depicted in Fig. 2.25. TPA occurs here because the indirect bandgap of Si is 1.1eV [113], and incident photons at $\lambda = 1550$nm contain 0.8eV of energy. Two incident photons, in addition to phonons present in the lattice, can excite carriers from the valence band into the conduction band. In Si, the requirement of a phonon for this process is the only reason that TPA can be observed at this wavelength. Otherwise, two photons do not have enough energy to excite carriers into the conduction band via the direct bandgap.

TPA involves the absorption of photons to excite carriers into the conduction band, and this manifests as a reduction in the intensity of a propagating wave. For example, considering a wave propagating along the y-axis in a medium having no TPA, the change in intensity as a function of distance is simply given by:

$$\frac{dI_{opt}}{dy} = -\alpha_0 I_{opt} \tag{2.101}$$

The change in intensity as a function of distance from the TPA only can be described as:

$$\frac{dI_{opt}}{dy} = -\beta_{TPA} I_{opt}^2 \tag{2.102}$$

Both linear absorption and TPA contribute to the attenuation of a wave propagating through the medium. To illustrate this, the intensity as a function of distance can be written as:

$$I_{opt}(y) = \frac{W(I_{opt,0}\beta_{TPA}e^{-(\alpha_0)y}y)}{\beta_{TPA}y} \tag{2.103}$$

where W is known as the Lambert W function, and $I_{opt,0}$ is the incident intensity at $y = 0\mu$m. As an illustrative example, consider a $\lambda = 1550$nm wave propagating through Si. At this wavelength, Si is essentially lossless, and $\alpha_0 = 0$. However, $\beta_{TPA} = 0.68$cm/GW has been observed experimentally [104], and cannot be ignored

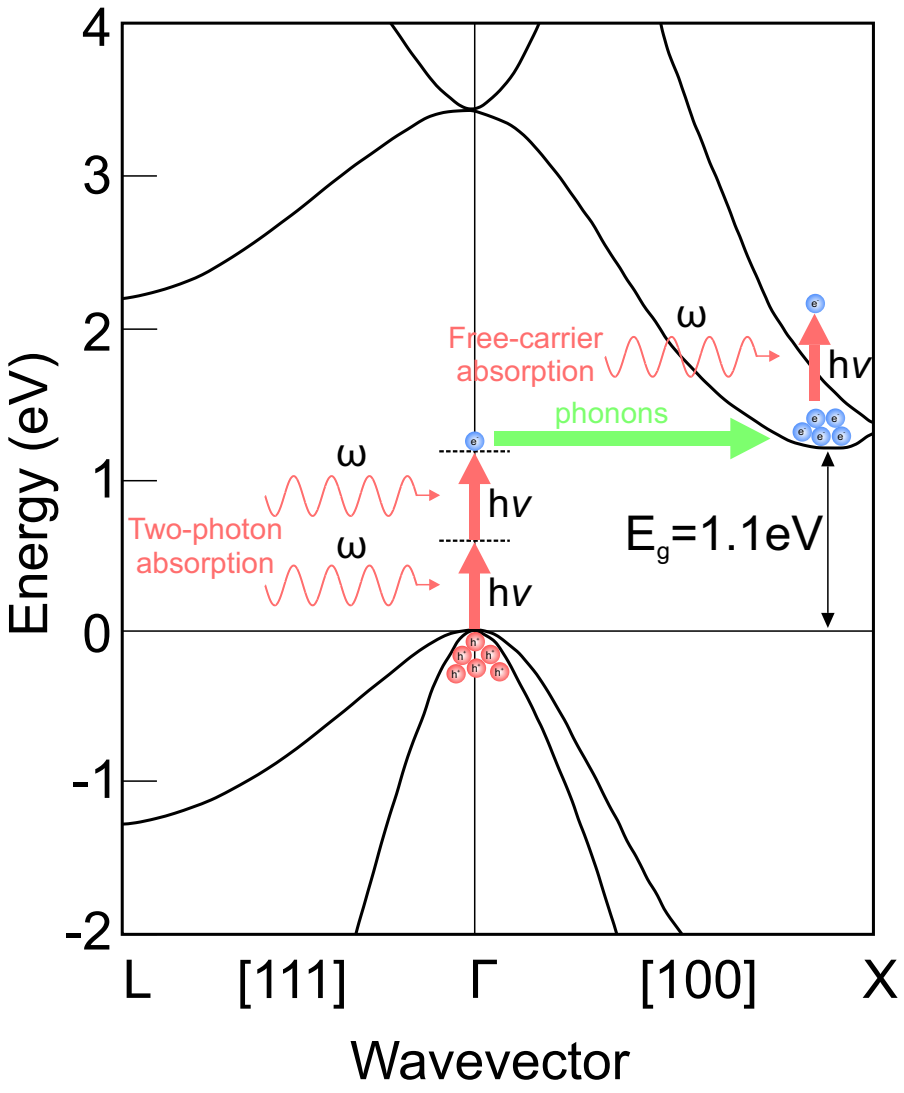

Figure 2.25: The band structure of Si. TPA occurs when two photons at a frequency ω excite an electron, and a phonon provides the additional momentum required for the electron to enter the conduction band via the indirect bandgap. Free carrier absorption occurs when free electrons in the conduction band absorb additional photons. Note that the superimposed carriers are for illustration only, and are not to scale. Image was adapted from and modified from [113].

at high intensities. If the wave propagates a distance $L = 25\mu$m through the material, the intensity of the wave at $y = L$ can be plotted as a function of the incident intensity $I_{opt,0}$, as depicted in Fig. 2.26. Here, the blue curve represents the transmitted intensity with only linear losses, which in this particular scenario, is effectively lossless, leading to a straight line. The red curve indicates the transmitted intensity with the effects of TPA included. At low incident intensities, the two curves behave in a similar

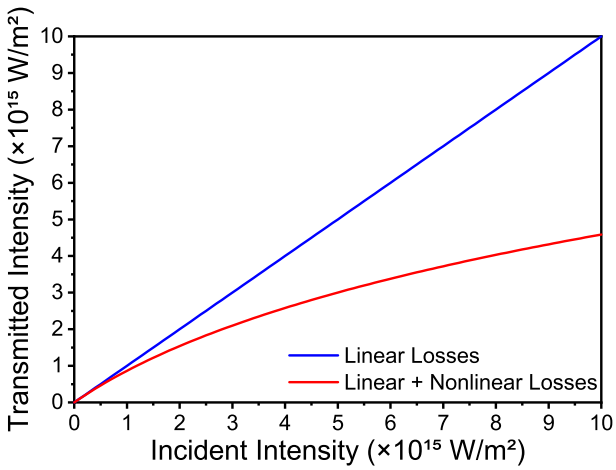

Figure 2.26: Plot of the transmitted light intensity of a propagating wave through 25μm of Si as a function of the incident intensity. The blue curve corresponds to only linear losses present, which are effectively negligible for Si at $\lambda = 1550$nm. The red curve corresponds to the transmitted intensity when TPA-induced nonlinear loss is present.

manner. However, at higher intensities (i.e. $> 1 \times 10^{15}$W/m^2 in this example), the nonlinear absorption dominates and begins to saturate the transmitted intensity, even at relatively short length scales.

2.4.3 Carrier Effects in Silicon

While TPA generates free carriers within silicon, it is important to consider the dynamics of these carriers and the impact they have on any transmitted radiation. Once excited, carriers in silicon recombine within a characteristic recombination time, denoted as τ_r. The generation and recombination of the carriers can be summarized with the following relation [110]:

$$\frac{dN(t)}{dt} = \frac{\beta_{TPA}}{2h\nu} I_{opt}^2(t) - \frac{N(t)}{\tau_r} \tag{2.104}$$

where $N(t)$ is the density of carriers generated as a function of time. The generated carriers as a function of time are depicted in Fig. 2.27(a) for the exemplary case of an exciting pulse with a $\lambda = 1550$nm carrier, a Gaussian envelope with a full-width-half-maximum (FWHM) pulse duration of $\tau_p = 84$fs, and peak intensities of $I_{opt,pk} =$

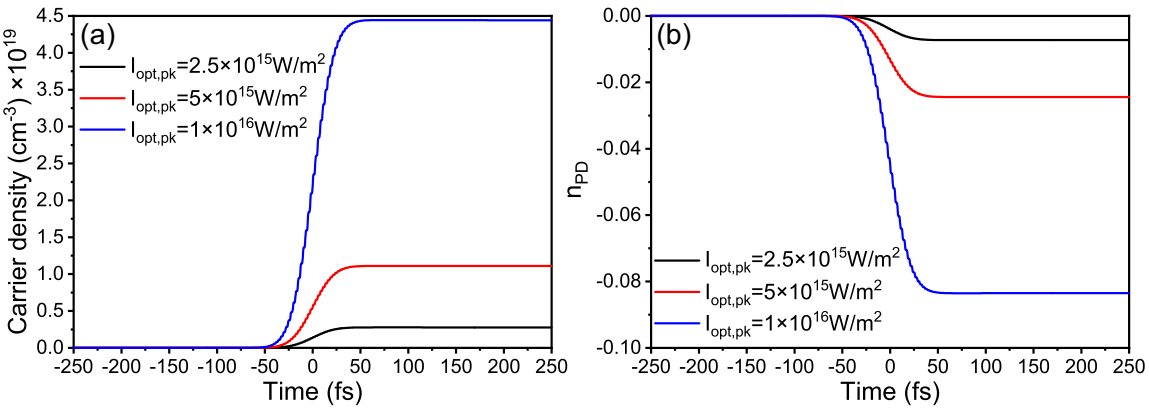

Figure 2.27: (a) Excited carrier density as a function of time for a $\tau_p = 84$fs Gaussian pulse centered at $t = 0$fs with peak optical intensities of $I_{opt,pk} = [2.5 \times 10^{15}, 5 \times 10^{15}, 1 \times 10^{16}]$W/m^2. (b) Change in the refractive index due to plasma dispersion from the carriers generated under the conditions depicted in (a).

$[2.5 \times 10^{15}, 5 \times 10^{15}, 1 \times 10^{16}]$W/m^2, centered at time $t = 0$fs. The recombination time is taken to be $\tau_r = 265$ps [114], and $\beta_{TPA} = 0.68$cm/GW [104], both of which have been measured in Si. Clearly, larger $I_{opt,pk}$ generates more carriers, with maximum carriers generated in each case being $N = [2.8 \times 10^{18}, 1.1 \times 10^{19}, 4.4 \times 10^{19}]$cm^{-3}. The carrier concentration decays back down on the time scale of τ_r.

The extra carriers naturally embody another mechanism by which the refractive index and absorption coefficient of the material are perturbed. The change in refractive index, Δn due to this effect introduces additional dispersion from the free carriers, and is denoted as n_{PD}. One cannot ignore the fact that the free carriers in the conduction band are able to absorb single photons (as depicted in Fig. 2.25), which leads to a process called free-carrier absorption (FCA). This change in absorption, $\Delta\alpha$, is denoted by α_{FCA}.

Exact expressions have been derived from the Drude model for n_{PD} and α_{FCA} induced by the additional free carriers in the conduction band. Experimental evidence has shown that the actual magnitude of the refractive index change is larger than predicted by such equations [106]. Therefore, for wavelengths around $\lambda = 1550$nm,

the empirical formulas shown below are more commonly utilized [109–111, 115]:

$$\Delta n = n_{PD} = -\left[8.8 \times 10^{-22} N_e + 8.5 \times 10^{-18} N_h^{0.8}\right] \left(\frac{\lambda}{1.55 \times 10^{-6}}\right)^2 \tag{2.105}$$

$$\Delta \alpha = \alpha_{FCA} = \left[8.5 \times 10^{-18} N_e + 6.0 \times 10^{-18} N_h\right] \left(\frac{\lambda}{1.55 \times 10^{-6}}\right)^2 \tag{2.106}$$

where N_e is the density of excited electrons in units of cm^{-3}, N_h is the density of excited holes in units of cm^{-3}, and λ is the wavelength in units of meters. The mechanism of TPA exciting an electron from the valence band to the conduction band naturally implies that the number of electrons and holes created this way are equal, or $N_e = N_h = N$. Figure 2.27(b) plots the value of n_{PD} as a function of time for each of the carrier excitations depicted in Figure 2.27(a). Again, since higher $I_{opt,pk}$ generates more carriers, this naturally corresponds to a higher magnitude of n_{PD}.

The free carrier absorption (α_{FCA}) at $\lambda = 1550$nm can be further simplified to $\alpha_{FCA} = \sigma_{FCA} N$, where σ_{FCA} is the free-carrier absorption cross section. The attenuation of the intensity of a wave propagating through the medium due to FCA alone is:

$$\frac{dI_{opt}}{dy} = -\alpha_{FCA} I_{opt} = -\sigma_{FCA} N I_{opt} \tag{2.107}$$

The total attenuation of the intensity as a function of length due to linear absorption, TPA, and FCA, can be written as:

$$\frac{dI_{opt}}{dy} = -(\alpha_0 + \sigma_{FCA} N) I_{opt} - \beta_{TPA} I_{opt}^2 \tag{2.108}$$

Evidently, once the nonlinear loss mechanisms are introduced, the intensity of a wave propagating through the medium decreases rapidly. At low intensities, the nonlinear absorption is not significant, and light propagation is affected only by the linear losses of the material . However, increasing the intensity significantly attenuates the

propagating wave. This is a fundamental factor in integrated optics that sets a limit on the maximum optical power that can be utilized within a Si-based nanophotonic or nanoplasmonic circuit.

Chapter 3

Integrated Plasmonic Devices Utilizing the Nonreciprocal Phase Shift

3.1 Introduction

Magnetic materials are indispensable building blocks for advanced electronics and optical systems. For integrated optics, MO materials possess distinct optical properties that facilitate the implementation of unique components and devices. Externally or inherently magnetized gyrotropic materials break the time reversal symmetry and the Lorentz reciprocity [116] of the interacting optical waves, leading to the realization of nonreciprocal phenomena. Having functionalities unattainable by other materials, a new class of nonreciprocal devices need to be developed to fulfill key applications in optical networks and information processing optical circuits.

To fully realize integrated nanoplasmonic circuitry, plasmonic analogues of nonreciprocal photonic and electronic devices must be developed. Elements such as optical isolators, routers, modulators, and circulators are indispensable in optical signal pro-

cessing systems. Unique magnetic devices must be scaled down to the nanoplasmonic regime, making magnetoplasmonic systems an attractive solution. The nonreciprocal phase shift (NRPS) is an interesting effect that provides a pathway to modulating both phase and intensity of a guided optical mode. Incorporating the NRPS into nanoplasmonic waveguide geometries and structures provides the potential for developing a new paradigm in compact and high-speed nonreciprocal plasmonic devices for signal modulation and encoding.

This chapter explores the design and numerical simulation of a number of unique magnetoplasmonic geometries exploiting the nonreciprocal phase shift. The first section explores the incorporation of the NRPS within a single waveguide, and details how this phase shift can be actively modulated at high speeds with applications in data encoding. The following section improves on the waveguide design in order to construct another crucial optical component: the optical isolator. With this architecture established, the final two sections detail how a similar geometry can be utilized to realize two completely novel lightwave modulator designs, having unique functions that have never previously been realized in an integrated plasmonic platform.

3.2 High-speed Magnetoplasmonic Waveguide Phase Shifter[1]

3.2.1 Phase Shifter Design

A marked advantage of nanoplasmonics is that it can be configured to allow for significant asymmetry in the structure and the confined optical mode. Both criteria are essential for developing NRPS based MO devices [73, 74]. As described in Chapter 2, in a nanoplasmonic waveguide, the TM mode is composed of two elec-

[1]A version of this section has been published as C. J. Firby and A. Y. Elezzabi, "High-speed nonreciprocal magnetoplasmonic waveguide phase shifter," Optica 2(7), 598-606 (2015).

tric field components (i.e. a transverse component and a longitudinal component). The phase difference between these constituent fields implies that the mode exhibits slight elliptical polarization [73, 74]. When the nanoplasmonic waveguide consists of a magnetic dielectric guiding layer magnetized perpendicular to these two electric field components, the rotating modal fields will be influenced by the magnetization of the material, and the NRPS will manifest.

To demonstrate an active class of magnetoplasmonic devices, it is important to start by determining the general characteristics of a nonreciprocal phase shifter under passive operation. The MO material here is a sample of bismuth-substituted yttrium-iron garnet (Bi:YIG) having a chemical composition of $[(Y_{0.6}Bi_{0.4}LuPr)_3(FeGa)_5O_{12}]$, a saturation magnetization ($\mu_0 M_S = 9$mT), a refractive index ($n_{YIG} = 2.3$ at 1550nm), and a specific Faraday rotation ($\theta_{spec} = 0.25°/\mu$m at 1550nm) [117]. The aforementioned Bi:YIG material has shown ultrafast magnetic field interactions in excess of 82GHz [118]. Operation is designed for the telecommunications wavelength of 1550nm, which provides the most practical spectral region for implementing Bi:YIG magnetoplasmonic modulators due to the presence of a high optical transparency window.

The magnetoplasmonic waveguide under consideration is of the MI or DL configuration. It consists of a ridge of Bi:YIG set upon a silver film, as shown in Fig. 3.1(a). The Bi:YIG guiding layer has a width w_{YIG}, height t_{YIG}, and length L_{YIG}, while the underlying Ag has a width w_{Ag}, height t_{Ag}, and length L_{Ag}, as schematically shown in Figs. 3.1(b)-(c). Here, the guided mode propagates in the $+z$-direction, and the plasmonic mode of the waveguide is characterized by the electric fields E_y and E_z, which are depicted in Figs. 3.1(b)-(c) respectively (for an exemplary case of a 270nm×300nm Bi:YIG ridge).

Implementation of this magnetoplasmonic architecture is achievable through a combination of standard nanofabrication techniques. Bi *et. al.* showed that magnetic

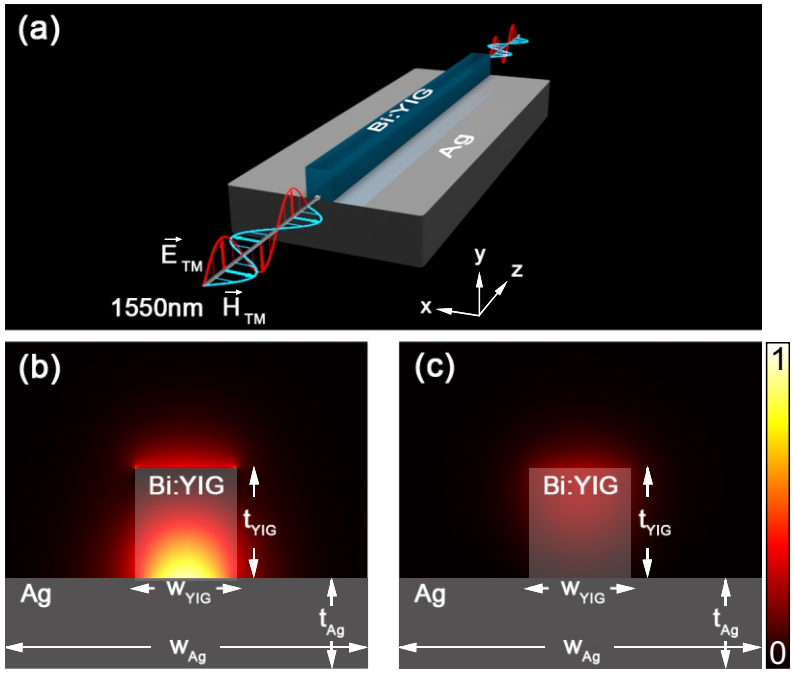

Figure 3.1: Schematic illustration of the magnetoplasmonic waveguide geometry. (b) $|E_y|^2$ profile and (c) $|E_z|^2$ profile for an exemplary waveguide having $w_{YIG} = 270$nm and $t_{YIG} = 300$nm.

garnets can be deposited by pulsed laser deposition (PLD) onto an SOI platform [92]. Practically, to create the present structure, the underlying Ag slab can be fabricated via sputtering/evaporation and photolithography on a standard Si substrate, while the magnetic garnet can be deposited via PLD. Focused ion beam (FIB) milling can precisely define the required waveguide geometry.

Remarkably, this simple nanoplasmonic waveguide configuration satisfies each of the conditions required for inducing the NRPS. The optical mode is tightly confined at the nanoscale, with 70% of the modal power contained within the 0.081μm^2 cross-sectional area of the core. With a Bi:YIG core and cladding layers of air and Ag, the plasmonic excitations are naturally confined to the metal-dielectric interface. Figs. 3.1(b)-(c) show that the dominant transverse field component, E_y, is highly localized near the metal, while the longitudinal component, E_z, is localized near the top of the waveguide. In the limit as the waveguide dimensions are increased, the maximum of the longitudinal field shifts downwards toward the Bi:YIG-metal interface

and eventually couples to the surface (i.e., the mode becomes more like that at a planar interface between two semi-infinite materials). The presence of both E_y and E_z components indicates that the mode is elliptically polarized in the y-z-plane. The maximum magnitude of the two field components are found to be $|E_y| = 0.96$ and $|E_z| = 0.48$. The ratio between the two electric fields is about 2, and the ellipticity condition is inherently satisfied. Finally, this waveguide has a relatively long propagation length of 49.06μm, allowing for considerable buildup of the NRPS along its length.

To ascertain the details of the interaction between the plasmonic mode and the magnetized Bi:YIG material, the effect of the magnetization vector, $\mathbf{M}$, on the optical wave can be expressed through the off-diagonal elements of the material's relative permittivity tensor, given by Eqn. 2.62 and taking $\varepsilon_r = \varepsilon_{YIG} = n_{YIG}^2$. When no static magnetic field ($\mathbf{H}_{static} = 0$) is applied to the Bi:YIG, $\hat{\varepsilon}_r$ becomes an isotropic, diagonal matrix, and a TM mode propagating through the medium is characterized by a reciprocal propagation constant β_0. However, when $\mathbf{H}_{static} \neq 0$, the situation is altered, and the propagation constant becomes nonreciprocal. Due to the presence of the transverse magnetization (i.e. along the x-direction, perpendicular to E_y and E_z), a NRPS will be induced.

It is essential to optimize the phase shifting performance of the magnetoplasmonic waveguide, by employing fully vectorial finite-difference-time-domain (FDTD) calculations to discern the NRPS of the passive waveguide configuration for representative waveguide widths and heights. However, incorporating MO effects within an FDTD algorithm is a complex task, and requires modification to the standard FDTD algorithms to perform correctly. Numerous intricate details have to be accounted for in order to have confidence in the accuracy of the results. For the work presented in this thesis, FDTD calculations were performed with the aid of the commercial software, *Lumerical FDTD Solutions* [119], although additional components

and auxiliary scrips were required to accurately model the MO phenomena such as the NRPS. Substantial work was carried out in close collaboration with the engineers at *Lumerical Inc.* over a period of eight months in order to rigorously test these methods and scripts, and document the intricacies involved.

Magnetoplasmonic devices provide the ideal platform for thoroughly characterizing a numerical modeling system, due to their small size, electric field enhancement, extreme sensitivity to perturbations in their environment, high index contrast between the metals and dielectrics involved, and strong absorption. These simulations push the limits of the FDTD frameworks, and expose the complexities of the modeling process.

For example, it was discovered that modeling a MO material in contact with another material, such that there existed a high refractive index contrast at the border of a mesh cell, would cause the algorithm to generate incorrect results, and occasionally diverge entirely. To remedy this, a single mesh cell border around the perimeter of the MO material had to be made of an isotropic material with the same refractive index as the MO material, but without the MO effects. This single cell could be made as small as 1-2nm in order to produce a negligible impact of the MO performance, while still allowing the algorithm to converge and produce correct results. In a similar manner, it was found that the FDTD algorithm could not handle injecting a source wave within the MO material, so the input fields had to be launched into a few mesh cells of isotropic material first.

Additionally, measurement of the NRPS within FDTD proved to be challenging. The eigenmode solvers could not account for off-diagonal elements in the permittivity matrix, and could not be used to calculate the NRPS directly. Since the magnitude of the phase shift due to the magnetization is typically quite small, grid dispersion, which arises from the discretization of a structure, would often mask or inhibit its accurate measurement as well. While grid dispersion can be reduced and effectively

eliminated with by reducing the mesh size, this procedure can increase the required computational resources substantially, which is not ideal during the initial iterative optimization and design processes. An alternative method was conceived based on an approach that treated the waveguide as a periodic structure with a period of one mesh cell to quickly determine the perturbed propagation constants of the modes of the structure in the presence of a magnetization.

All calculations were meticulously tested and quantified against previously published theoretical and experimental results in conjunction with the developers at *Lumerical Inc.* prior to deployment for the studies in this thesis. The auxiliary script files employed to include the off-diagonal permittivity elements into the FDTD algorithm of *Lumerical FDTD Solutions*, as well as other scripts required for performing the MO calculations are presented in Appendix A for reference.

Figure 3.2(a) depicts the NRPS as a function of Bi:YIG width for waveguide heights of 300nm, 400nm, and 500nm. For any of the examined heights, as the width is increased the NRPS gradually increases as well, due to increased fraction of the optical power confined within the Bi:YIG core. Maximum NRPS of $\Delta\beta$ =7.53rad/mm, 6.39rad/mm, and 5.60rad/mm are obtained at widths of 450nm, 400nm, and 400nm for the three cases, respectively. Beyond these values, the NRPS remains relatively constant, with only a slight decrease resulting from reduced ellipticity of the mode. Likewise, Fig. 3.2(b) depicts the NRPS as a function of Bi:YIG height for fixed widths of 300nm, 400nm, and 500nm. The NRPS increases rapidly, again due to the increase in mode confinement in the Bi:YIG core. Maximum NRPS of 7.44rad/mm, 8.12rad/mm, and 8.40rad/mm are realized for heights of 260nm, 230nm, and 220nm for the three cases respectively. Clearly, the magnitude and overall shape and of the NRPS curves depend on a complex interplay between optical modal power confinement and optical modal ellipticity. This is illustrated in the insets of Fig. 3.2, where the E_y field component is shown to shift away from the Ag-Bi:YIG interface to the

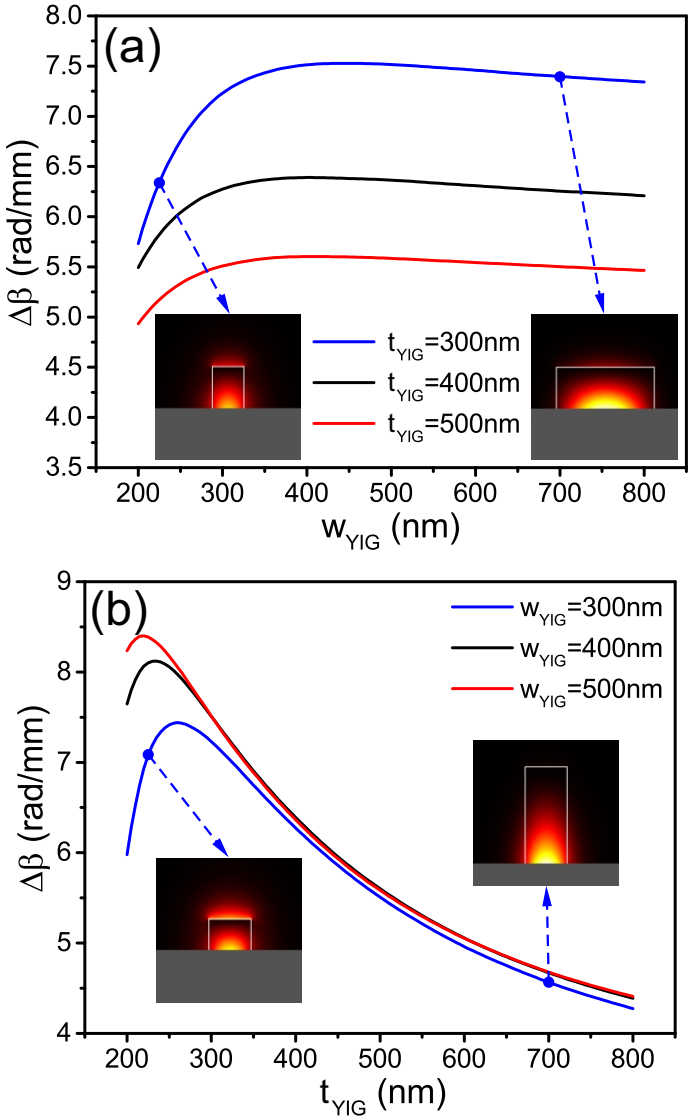

Figure 3.2: NRPS as a function of the waveguide dimensions. (a) shows the NRPS as a function of width for fixed waveguide heights, while (b) shows the NRPS as a function of height for several fixed widths. Both figures are calculated for $M_x = M_S$. The insets in both (a) and (b) show the mode profiles at the indicated dimensions.

air-Bi:YIG interface as the waveguide dimensions are reduced.

For practical implementation, the net accumulated NRPS must be realized over the propagation distance of the plasmonic mode (i.e., the net NRPS is equal to the product of the NRPS and the modal propagation length). In Fig. 3.3, the net NRPS is plotted for the same ranges of waveguide dimensions. At each w_{YIG}, the net NRPS increases monotonically to 0.395rad, 0.379rad, and 0.369rad at heights of 800nm for

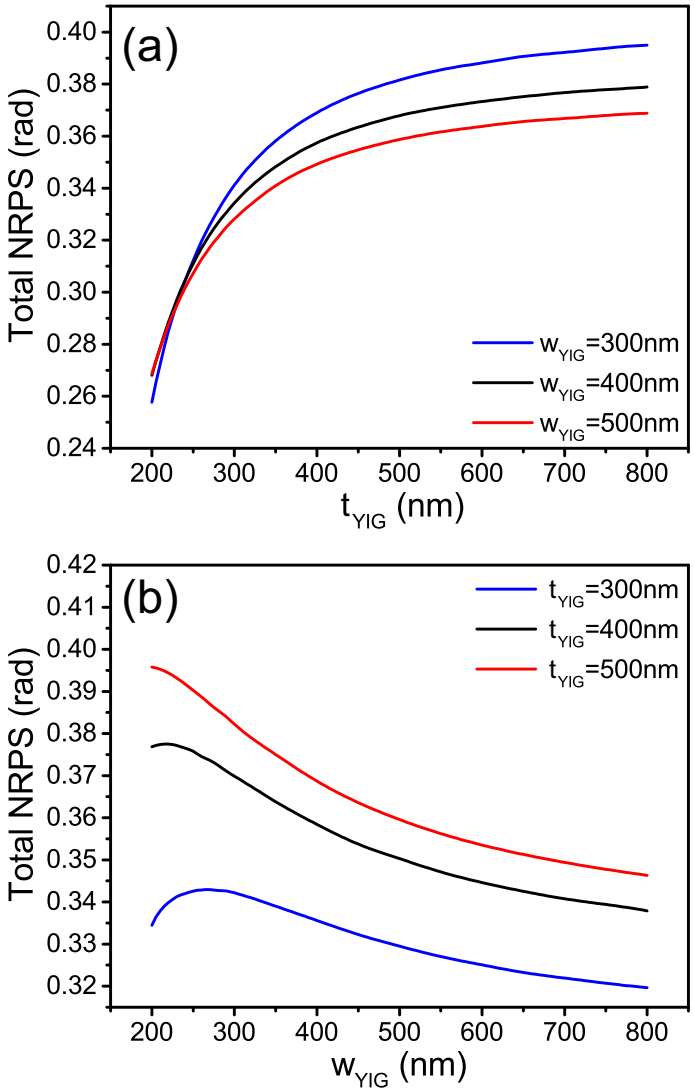

Figure 3.3: The total attainable NRPS as a function of the magnetoplasmonic waveguide's (a) height, and (b) width.

the three widths respectively, as shown in Fig. 3.3(a). However, for a set t_{YIG}, optimum widths are present (see Fig. 3.3(b)), but the increase in total attainable NRPS with height is marginal. To remain in the realm of nanoplasmonic integrated circuitry, smaller device dimensions are desirable. Taking these factors into account, the optimum parameters for the Bi:YIG magnetoplasmonic waveguide are selected to be w_{YIG} = 270nm and t_{YIG} = 300nm, resulting in an NRPS of $\Delta\beta$ =6.99rad/mm and 70% modal power confinement. Over the modal propagation length of 49.06μm, a total NRPS of 0.34rad is realized. The transverse E_y and longitudinal E_z intensity

profiles of this optimal configuration are shown in Fig. 3.1(a) and Fig. 3.1(b). The device is compact, as the waveguide has a footprint of only $13.25\mu\text{m}^2$.

3.2.2 High-Speed Active Operation

The aforementioned passive device can be converted into a high-speed active magnetoplasmonic device by altering the direction of $\mathbf{M}$ via incorporating and modulating an additional time-varying magnetic field, $\mathbf{h}(t)$. Accordingly, the time dynamics can be described through the LLG formalism described by Eqn. 2.86. In this case, the total applied magnetic field becomes the superposition of the static and transient fields, $\mathbf{H}_{static} + \mathbf{h}(t)$. With this arrangement, the LLG equation can be re-cast as:

$$\begin{aligned}\frac{d\mathbf{M}}{dt} =& -\frac{\mu_0\gamma_0}{1+\alpha^2}[\mathbf{M}\times(\mathbf{H}_{static}+\mathbf{h}(t))] \\ &-\frac{\mu_0\gamma_0\alpha}{M_S(1+\alpha^2)}\mathbf{M}\times[\mathbf{M}\times(\mathbf{H}_{static}+\mathbf{h}(t))]\end{aligned} \tag{3.1}$$

where all parameters are as defined previously, and the Gilbert damping parameter, α, is taken to be an experimentally measured value of $\alpha = 10^{-4}$ [120]. The LLG model describes a system of highly nonlinear coupled differential equations relating the three magnetization components to each other, and to the components of the applied magnetic fields. Certainly, there is no analytic solution to Eqn. 3.1, and hence, it must be solved numerically. To this end, a custom program was developed in MATLAB to solve these equations as a function of time for arbitrary input field configurations and initial conditions. The resultant MATLAB program employed for these calculations is presented within Appendix B. Given initial conditions and the applied magnetic field, the temporal evolution of the magnetization is calculated. The degree of magnetization enters the FDTD simulations through the off-diagonal elements of the Bi:YIG permittivity matrix via the auxiliary scripts specifically developed for MO modeling. By combining the in-house developed MATLAB program

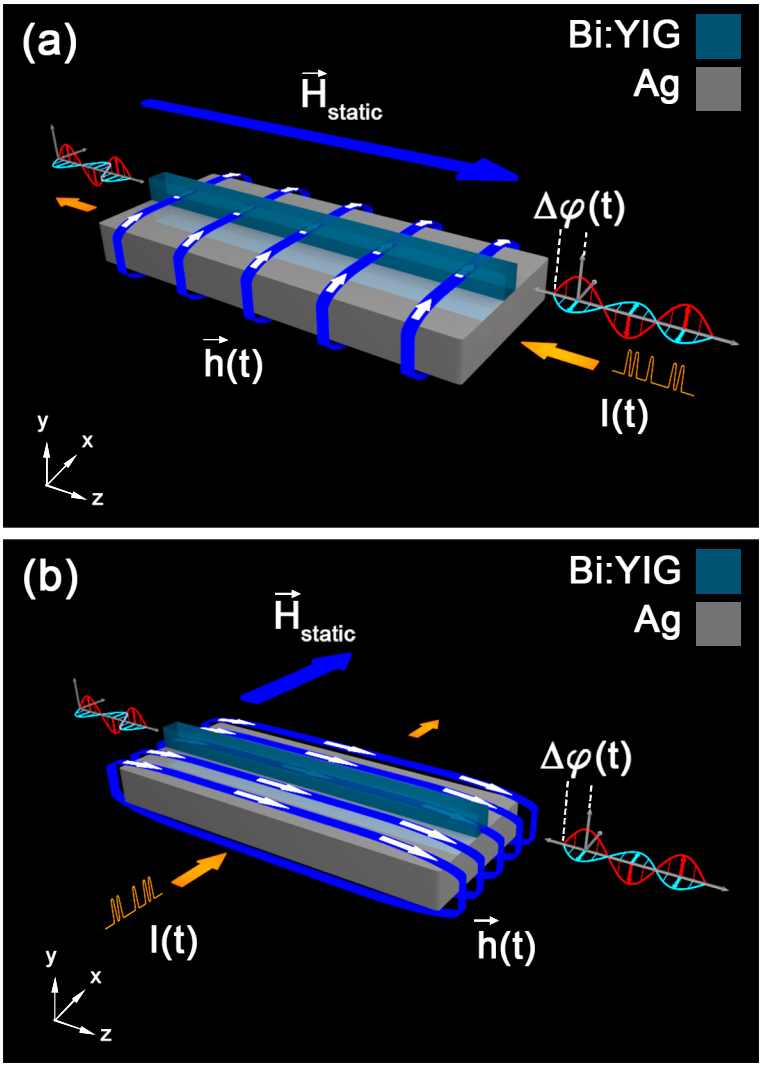

Figure 3.4: The active magnetoplasmonic phase shifter in (a) The longitudinal (L) configuration, and (b) The transverse (T) configuration under consideration.

with the commercial FDTD software and custom scripts for MO simulations, the effects of **M** on the optical wave propagating through the waveguide can be effectively determined.

To study the dynamic response of the active magnetoplasmonic phase shifter, two magnetic fields are required: a transient field (i.e., $\mathbf{h}(t)$) to induce a temporal change in the magnetization, and a static magnetic field (i.e. $\mathbf{H}_{static}$) to define the initial and final magnetization states before and after excitation, respectively. Two arrangements are considered. For the longitudinal configuration (denoted by L), the

$\mathbf{H}_{static} = \langle 0, 0, \pm H_z \rangle$ field is oriented longitudinally along the waveguide's length, and the modulating current pulse, $I(t)$, traveling through the Ag generates a transient magnetic field, $\mathbf{h}(t) = \langle \pm h_x(t), 0, 0 \rangle$. In the transverse configuration (denoted by T), the static $\mathbf{H}_{static} = \langle \pm H_x, 0, 0 \rangle$ field is oriented transverse to the waveguide, and the modulating current pulse, $I(t)$, generates a transient $\mathbf{h}(t) = \langle 0, 0, \pm h_z(t) \rangle$. Different permutations of the field signs and the direction of light propagation induce one-way phase shifts (OWPS) of different polarity, $\Delta\varphi = (\delta\beta)L_{YIG}$, where $\pm\delta\beta = \beta_{fwd,bwd} - \beta_0$. In both arrangements, only the M_x component contributes to the phase shift. When $M_x/M_S = 1$, a maximum phase shift $\Delta\varphi = 0.17$rad is obtained for a propagation length of $L_{YIG} = 49.06\mu$m. One representative manifestation of these permutations for each of the L- and T- configurations is illustrated in Figs. 3.4(a) and 3.4(b), respectively. Figure 3.4(a) depicts the L-configuration with $\mathbf{H}_{static} = \langle 0, 0, +H_z \rangle$ and $\mathbf{h}(t) = \langle +h_x(t), 0, 0 \rangle$ generated by a current pulse traveling anti-parallel to the mode propagation, in the -z direction. Figure 3.4(b) shows the T-configuration with $\mathbf{H}_{static} = \langle +H_x, 0, 0 \rangle$ and $\mathbf{h}(t) = \langle 0, 0, +h_z(t) \rangle$ generated by a current pulse traveling in the $+x$ direction. The initial magnetization conditions are set by the static field, $\mathbf{H}_{static}$ (i.e. $\mathbf{M} = \langle 0, 0, M_S \rangle$ and $\mathbf{M} = \langle M_S, 0, 0 \rangle$ for the L- and T-configurations, respectively).

To illustrate the temporal switching dynamics in response to the modulating current pulse, $I(t)$, with the initial conditions stated above, the device can be shown to operate in either an under-damped or a critically-damped mode. These two modes of operation are defined as conceptually analogous to a driven harmonic oscillator, but not mathematically equivalent. The under-damped mode is defined such that the magnetization component of interest exhibits residual precessional oscillations and does not attain the maximum possible primary peak. Conversely, a critically-damped response is defined such that the magnetization component of interest exhibits minimal precessional oscillations and attains the maximum possible primary response.

Consider first the L-configuration. Applying an $I(t)$ signal characterized by a FWHM pulse width, τ_p, and an amplitude I_{peak}, will administer a torque on $\mathbf{M}$, resulting in precessional motion of $\mathbf{M}$ around $\mathbf{H}_{static}$. Figure 3.5(a) shows the under-damped motion of the magnetization as it precesses around the static field for exemplary parameters of $\tau_p = 500$ps, a pulse peak of 40mT applied along the $+x$ direction, and a static biasing field of 50mT applied in the $+z$ direction (see Fig. 3.4(a)). The torque acting on $\mathbf{M}$ by the transient field pulse tips it away from the $+z$ axis, whereas $\mathbf{H}_{static}$ acts to return $\mathbf{M}$ to its original state. However, due to the under-damped nature of the excitation, harmonic oscillations around $\mathbf{H}_{static}$ at the ferromagnetic resonance (FMR) are induced. Following the time evolution of M_x, this precession manifests itself as residual oscillations following the primary peak, depicted in Fig. 3.5(b). Over time scales of several hundred nanoseconds to several microseconds, the amplitude of these oscillations diminishes. Certainly, this precessing magnetization is not practical for implementation into a fast phase shifting device since the oscillations last for significantly longer than the driving magnetic field pulse duration.

While it is possible to suppress the amplitude of the precession by increasing $\mathbf{H}_{static}$, the magnitude of the primary peak, and correspondingly the OWPS, will be significantly reduced. It is possible to critically-damp the magnetization and obtain the maximum response with minimal precessional oscillations. Such conditions can be realized by examining the magnetization response as functions of the transient and static magnetic fields, (see Appendix C for more details). An example of this critically-damped motion is presented in Fig. 3.5(c) for $\tau_p = 500$ps, a pulse peak of 54mT applied in the $+x$ direction, and a static biasing field of 47.8mT applied in the $+z$ direction (see Fig. 3.4(a)). In this scenario, the transient torque tips $\mathbf{M}$ away from the $+z$ axis, along the path shown in Fig. 3.5(c). Figure 3.5(d) depicts the evolution of M_x and shows that $\mathbf{M}$ is deflected all the way to the $+x$ direction before returning to its initial state, i.e., M_x is completely switched from 0 to $+M_S$. At the

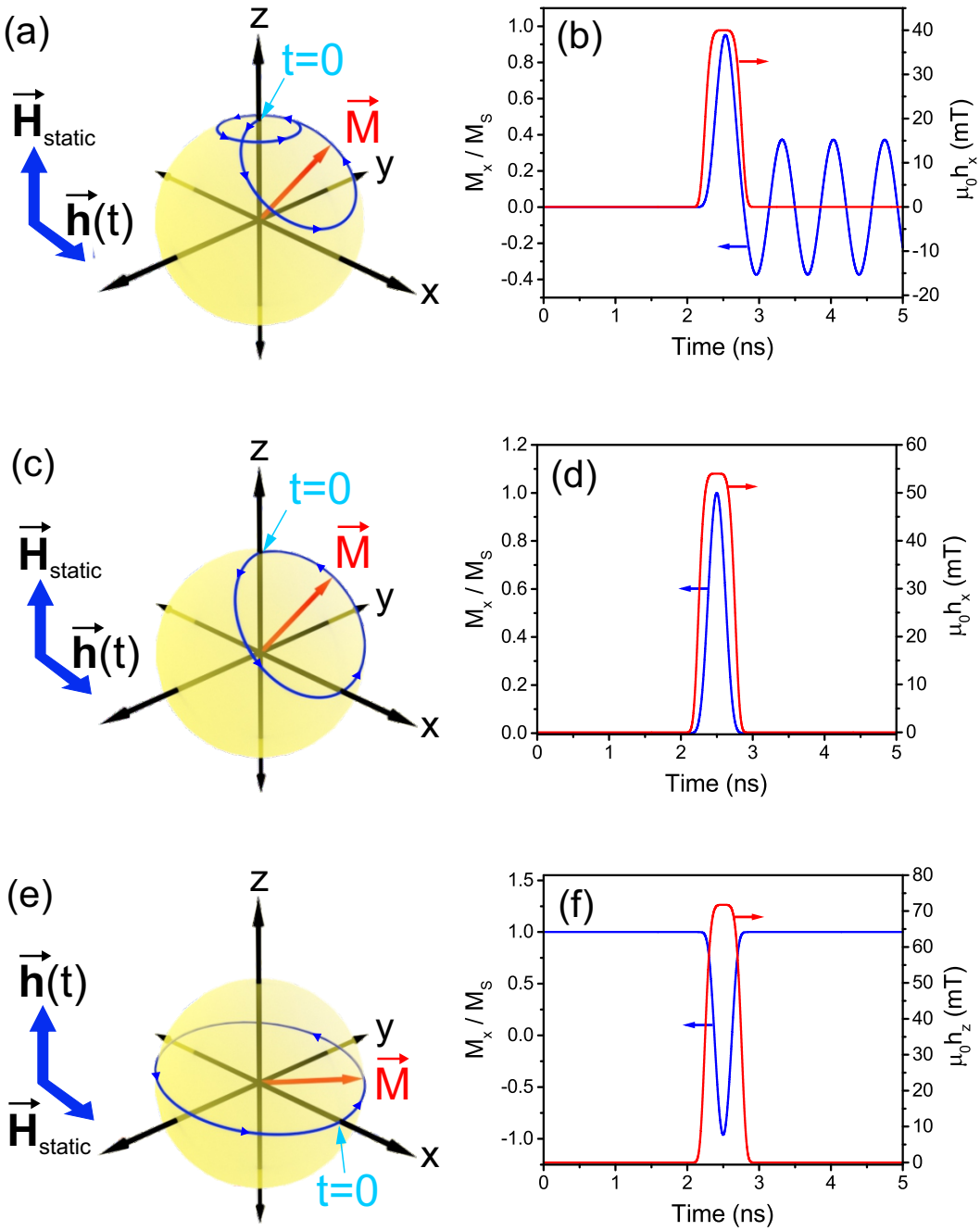

Figure 3.5: Temporal switching dynamics of the magnetization vector **M**. (a) Under-damped trajectory of the magnetization vector under an applied static field of 50mT and a transient field of 40mT with τ_p = 500ps for the L-configuration. (b) Time response of M_x component of the magnetization vector under the same conditions of (a). (c) Critically-damped trajectory of **M** for the L-configuration with a static field of 47.8mT and a τ_p = 500ps dynamic field of 54mT. (d) The critically-damped time response of M_x to the conditions of (c). (e) Critically-damped trajectory of **M** for the T-configuration with a static field of 9mT and a τ_p = 500ps transient field of 71.75mT. (f) The critically-damped time response of M_x for the T-configuration shown in (e).

critical-damping conditions, the amplitude of the precessional oscillations is negligibly small (<0.05%), resulting in an ideal response for high-speed phase shifting.

Similar critical-damping conditions exist for devices in the T-configuration. An exemplary plot of the magnetization motion is shown in Fig. 3.5(e) for τ_p = 500ps, a pulse peak of 71.75mT applied along the $+z$ direction, and a static biasing field of 9mT applied in the $+x$ direction (see Fig. 3.4(b)). Likewise, the transient field pulse tips **M** away from the $+x$ axis. Here, **M** temporally evolves along the path shown in Fig. 3.5(e), overshooting the y-axis and proceeding towards the -x axis before returning to its initial state parallel to $\mathbf{H}_{static}$. M_x is effectively switched from $+M_S$ to nearly -M_S and back, producing almost twice the phase shift compared to the L-configuration. This scenario is schematically illustrated in Fig 3.5(f), which depicts the evolution of M_x. The critical-damping condition ensures that **M** returns nearly to its original position before precession begins.

The optimal magnetic field conditions for critical-damping can be found for each of the L- and T-configurations for any pulse duration, static field, and modulating field. Specifically, the magnetoplasmonic modulator's operation is given here at two exemplary pulse widths of τ_p = 500ps and τ_p = 100ps. For τ_p = 500ps, the L-configuration (Fig. 3.4(a)) requires a $\mu_0 H_z$ = 47.8mT and a $\mu_0 h_{x,pk}$ = 54mT to switch M_x from 0 to $+M_S$ and back, while the T-arrangement (Fig. 3.4(b)) requires a $\mu_0 H_x$ = 9mT and a $\mu_0 h_{z,pk}$ = 71.75mT to switch M_x from $+M_S$ to -0.96M_S and back. To satisfy the same magnitude of switching in M_x, at τ_p = 100ps, the L-configuration requires a $\mu_0 H_z$ = 239mT and a $\mu_0 h_{x,pk}$ = 270mT, while the T-arrangement requires a $\mu_0 H_x$ = 9mT and a $\mu_0 h_{z,pk}$ = 361.5mT to switch from $+M_S$ to M_S. Full modulation can be achieved with shorter magnetic pulse durations, however, this comes at the expense of increasing the magnetic field requirements.

Since M_x is sensitive to the magnitude of $\mathbf{h}(t)$, a spatially uniform field distribution of the corresponding components over the waveguide's Bi:YIG core is essential to

ensure physically accurate performance of the device. The Ag slab dimensions and resulting currents are selected such that the fields are constant across the waveguide volume (see Appendix C). These parameters are summarized in Table 3.1. Note that since some of the required currents amplitudes are fairly large, specific current sources must be used to generate the short pulses (see Appendix C). Furthermore, as it is shown in Appendix C, thermal effects due to Joule heating can be discounted.

With the above parameters, the dynamic operation of the magnetoplasmonic modulator can be examined. Figure 3.6 shows the total phase shift response in all four devices (listed in Table 3.1) to a single unipolar electrical pulse $I(t)$ resembling a square wave in a telecommunication channel. Such a pulse representation is ideal since it is characterized by a high bandwidth, and it is simple to generate, receive, and interpret. Logic states are distinct, and a detection threshold can be easily set at the receiver.

In the L-configuration (Fig. 3.6(a)), the 500ps pulses attain a OWPS of 3.49rad/mm, and correspondingly, a positive $\Delta\varphi = 0.17$rad (or 9.74°). In comparison, for the T-configuration (Fig. 3.6(b)), the default state of the device already has $\Delta\varphi = 0.17$rad phase shift present. Application of the transient field further decreases the phase by -0.33rad to a peak value of $\Delta\varphi = -0.16$rad. The net shift from the initial state is -0.33rad (or 18.91°), nearly twice that of the L-configuration. In both configurations, $\Delta\varphi(t)$ conforms to the modulating $I(t)$ signal. Likewise, for electrical signals with shorter duration ($\tau_p = 100$ps) and higher bandwidth, the behaviour of the modulation is analogous. The phase shifts from the initial states are 0.17rad (9.74°) and 0.34rad

Table 3.1: Ag Dimensions and Peak Current Requirements

τ_p (ps)	Configuration	I_{peak} (A)	w_{Ag} (μm)	t_{Ag} (μm)	L_{Ag} (μm)
100	L	24.19	40.00	10.00	49.06
500	L	1.41	13.20	1.80	49.06
100	T	60.12	1.00	2.00	101.00
500	T	7.23	1.00	1.00	61.00

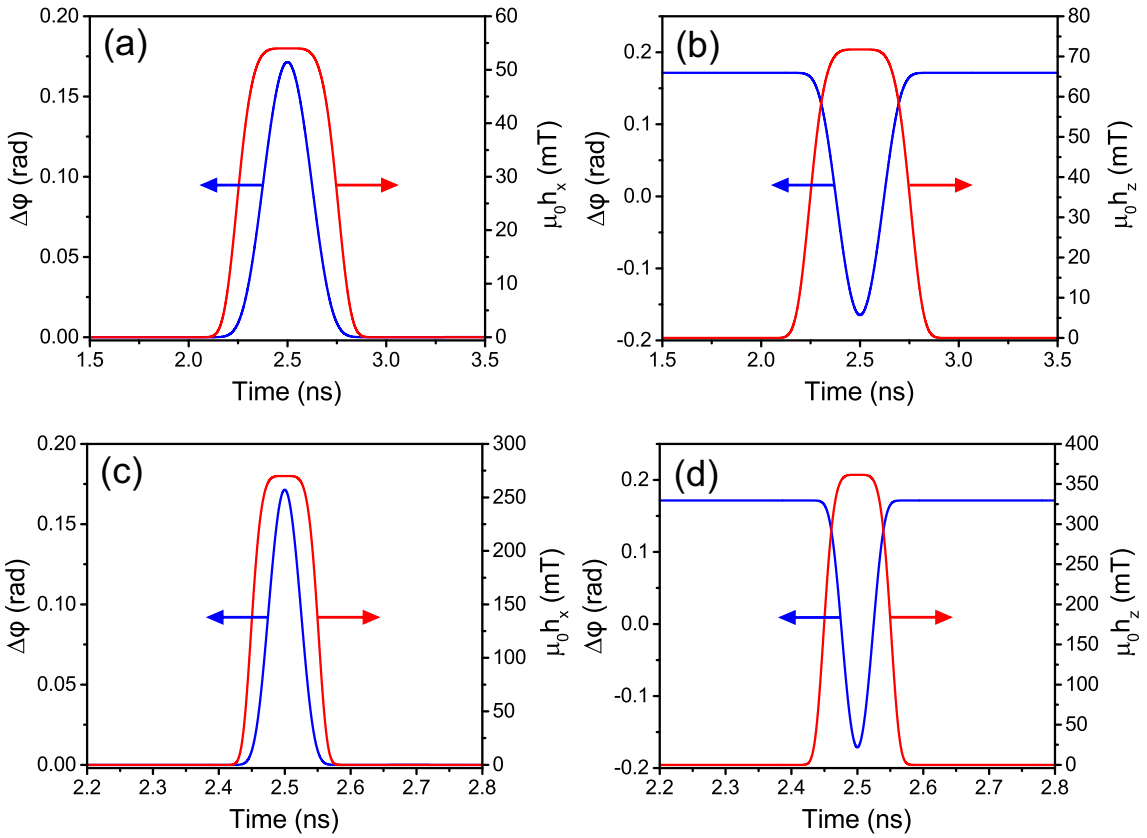

Figure 3.6: The OWPS time response to a single transient $I(t)$ pulse. (a) L-configuration with $\tau_p =$ 500ps, $\mu_0 H_z = 47.8$mT, and $\mu_0 h_{x,pk} = 54$mT. (b) T-configuration with $\tau_p = 500$ps, $\mu_0 H_x = 9$mT, and $\mu_0 h_{z,pk} = 71.75$mT. (c) L-configuration with $\tau_p = 100$ps, $\mu_0 H_z = 239$mT, and $\mu_0 h_{x,pk} = 270$mT. (d) T-configuration with $\tau_p = 100$ps, $\mu_0 H_x = 9$mT, and $\mu_0 h_{z,pk} = 361.5$mT.

(19.48°) for the L- and T-configurations as shown in Fig. 3.6(c)-(d), respectively.

By examining Fig. 3.7, it is evident in that the frequency content of the modulating current/magnetic field signals are well encompassed by the spectral response of the net phase shift signals. As such, with the higher bandwidth response of the Bi:YIG material, all spectral information contained (<4GHz for $\tau_p = 500$ps and <20GHz for $\tau_p = 100$ps) in the electrical signal will be mapped into the phase of the optical wave.

One primary application of this phase shifting magnetoplasmonic platform is encoding data via binary phase-shift keying (BPSK). Here, the phase of the optical signal is shifted by a known amount to encode data. The aforementioned magnetoplasmonic phase modulators can deliver a means of such data encoding in nanoscale plasmonic circuitry. To illustrate this application, a random 8-bit sequence, [11010011] is

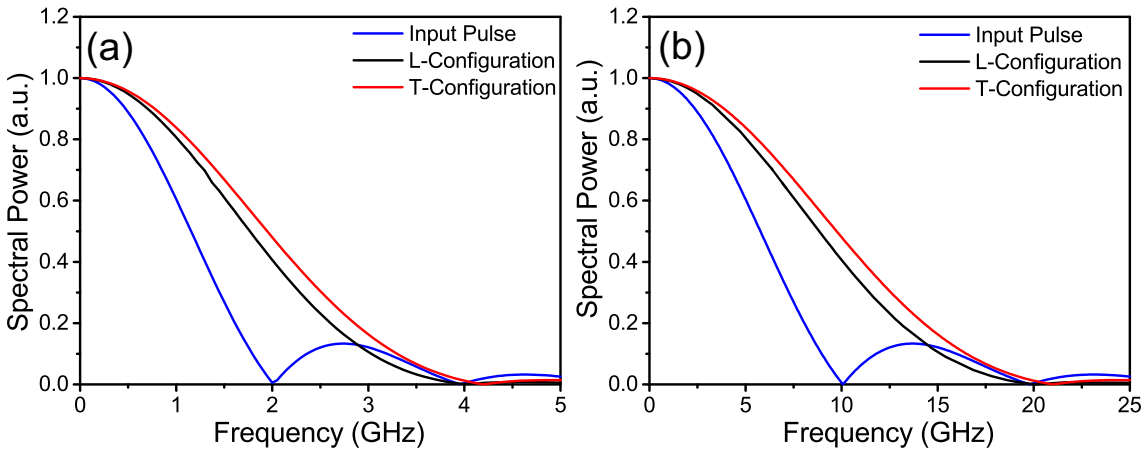

Figure 3.7: Normalized spectral power for the driving electrical signal and the resultant phase shift signal for (a) $\tau_p = 500$ps and (b) $\tau_p = 100$ps.

considered as the data train of current pulses at bit rates of 1.21Gbit/s and 6.06Gbit/s.

Figures 3.8(a)-(d) illustrate the phase shift as the high bit rate data sequence is mapped directly into the phase of the optical wave with virtually no residual phase oscillation noise during the off states. The 8-bit data train can be decoded with a high discrimination level. Higher bit-rates are possible at higher magnetic fields, and hence higher currents through the devices.

The aforementioned devices have several significant benefits over other phase modulators. Due to the nonreciprocal nature of this magnetoplasmonic platform, this class of modulators is bidirectional, where two counter-propagating signals experience phase shifts of opposing sign. This is an effect that cannot be achieved with electro-optic based [121], or other types of phase modulators. Both configurations offer an additional powerful feature, where the response can be tuned to the critical-damping regime for a wide range of pulse widths, static field strengths, and transient field strengths, making the devices extremely versatile. Another interesting feature is the fact that the transmission lines are quite wide (namely in the L-configuration), many devices can be integrated on the same transmission line. It is conceivably possible to utilize various telecommunications band wavelengths to encode the same signal into different channels in a large bus, which is beneficial for parallel signal processing

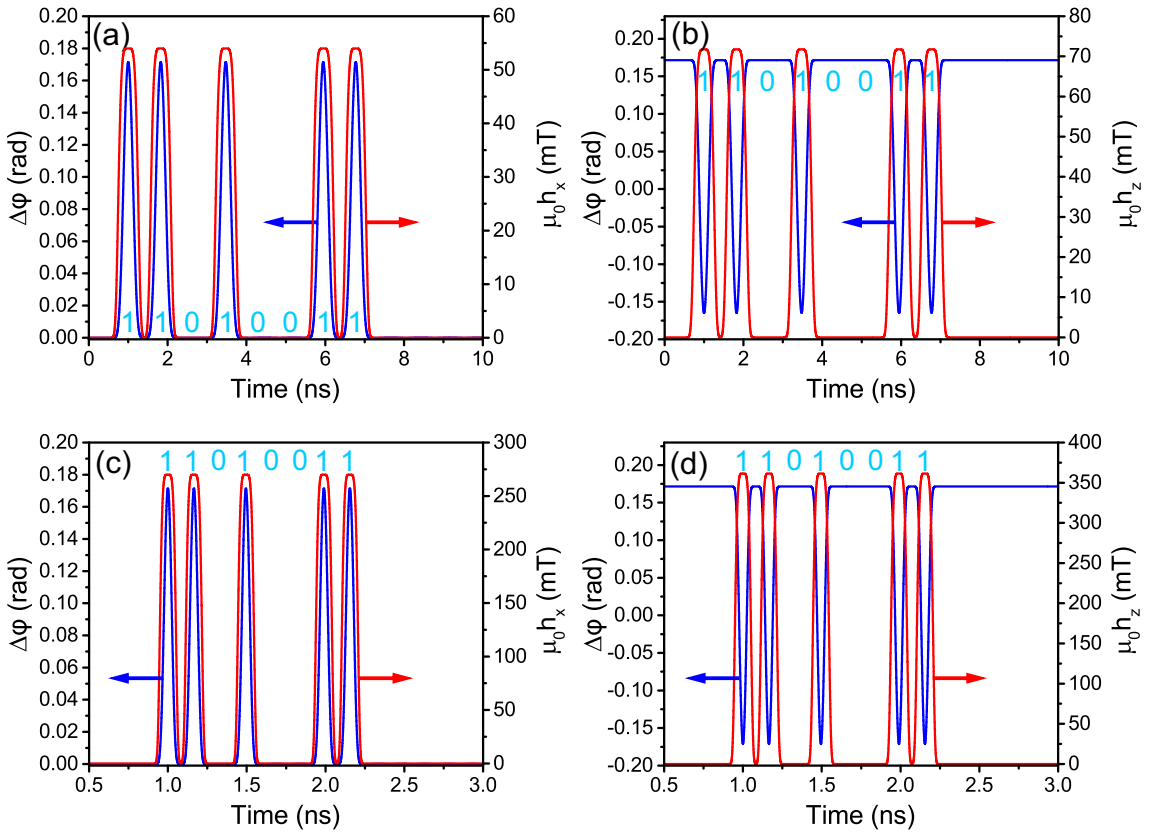

Figure 3.8: Phase shift response of the magnetoplasmonic device to a random bit sequence [11010011]. (a) L-configuration and (b) T-configuration with τ_p = 500ps pulses at a bitrate of 1.21Gbit/s. (c) L-configuration and (d) T-configuration with τ_p = 100ps pulses at a bitrate of 6.06Gbit/s. All devices are operated under critical-damping conditions.

and multiplexing applications.

Finally, it is noted that one of the primary factors limiting the speed of these devices is the increasing magnetic field requirements for operation with shorter pulse widths, and consequently, the mandatory increase in peak currents. With future development in nanoscale magnetic field generation and methods of providing uniform fields over the magnetic core, it is envisioned that the required currents could be dramatically reduced.

Another key application of nonreciprocal optical phase shifting devices is the implementation of isolators, which are crucial for protecting sensitive optical components. At present, both of the examined devices cannot produce the required phase shift for complete isolation. This is primarily due to two main factors. First, with a

modal propagation length of 49.06μm, the guided mode decays before it can acquire a useful NRPS of $\pi/2$. Second, while the Bi:YIG layer is useful to employ in active magnetic devices because it has demonstrated ultrafast switching capabilities, its MO parameters, specifically θ_{spec}, are lower than other variations of YIG. Specifically, substituting lattice sites within YIG with Ce, producing cerium-substituted yttrium-iron-garnet (Ce:YIG), has been shown to exhibit MO parameters almost twice as high as those of Bi:YIG [122]. The natural evolution from the present device is to modify the plasmonic waveguide geometry to reduce the losses, increase the modal propagation length, and incorporate Ce:YIG having higher MO properties.

3.3 Magnetoplasmonic Isolators[2]

Having found applications ranging from high-power laser systems to chip-scale integrated optics, optical isolators are fundamental components in a wide array of advanced optical devices. Such building blocks act as a one-way optical conduit, permitting the flow of electromagnetic energy in one direction, while prohibiting the flow in the reverse direction. In this way, sensitive components are protected from potentially damaging reflections and unintentional backwards propagating radiation. Optical isolation is attainable by breaking the time reversal symmetry of the optical wave, via a direction-dependent (i.e. nonreciprocal) propagation phenomenon.

Utilizing the NRPS in conjunction with an interference structure has previously allowed for the implementation of a series of waveguide-based optical isolators [123–125]. The most common method incorporates the NRPS of a magnetized medium into one or both arms of a Mach-Zehnder interferometer (MZI). Here, the forward propagating light interferes constructively at the output, while the backwards propagating light interferes destructively. A less common incarnation incorporates a multi-mode-

[2]A version of this section has been published as C. J. Firby and A. Y. Elezzabi, "Magnetoplasmonic isolators utilizing the nonreciprocal phase shift," Opt. Lett. 41(3), 563-566 (2016).

interference structure and the self-imaging principle [126]. The waveguide nature of these devices naturally suggests their integrated capabilities in photonic circuits.

In attempts to bring the nonreciprocal nature of magnetic garnets into the realm of silicon photonics, it has recently been demonstrated that Ce:YIG crystals can be effectively bonded to SOI structures, and high isolation of 30dB can be achieved in a MZI [75] configuration. Ce:YIG films have also been deposited onto SOI waveguides to create nonreciprocal isolators out of a racetrack resonator [92]. Unfortunately, these structures rely on the fact that a large fraction of the fields are contained within the magnetic cladding, and hence have poor field confinement in the Si waveguides. Additionally, bonding a millimeter-scale Ce:YIG chip to the surface of a device negates benefits achieved with the nanoscale dimensions of the Si waveguide core.

3.3.1 Isolator Design

In order to create an optical isolator based on the NRPS, one requires an MZI arrangement. This configuration is inherently sensitive to the phase of the confined light. To mitigate the shortcomings of the previous magnetoplasmonic waveguide design, another type of waveguide geometry must be considered. Namely, the LRDLSPW geometry [47–50], having propagation lengths on the order of several millimeters, is ideal.

To reduce the optical loss, the MZI isolator consists of two types of guiding structures: the photonic (acting as an input/output splitter/combiner) and the LRDLSPW guiding arms. Note that here, the waveguide cross-section is taken in the x-z plane, and propagation is assumed in the y-direction. The base photonic waveguide structure consists of a ridge of Ce:YIG having a width $w_{YIG} = 320$nm and a height $t_{YIG} = 400$nm, as shown in Fig. 3.9(a). The Ce:YIG material has a refractive index, $n_{YIG} = 2.21$, a saturation magnetization of $\mu_0 M_S$=1.5mT, and a specific Faraday rotation, $\theta_F = -0.48°/\mu$m at $\lambda = 1550$nm [122]. This material has a higher MO

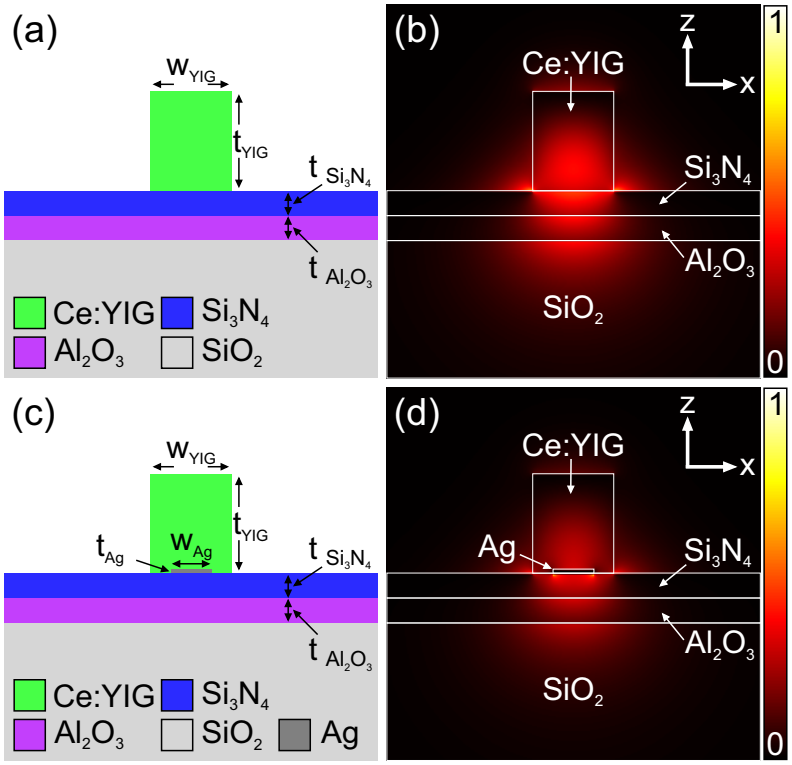

Figure 3.9: (a) Schematic illustration of the photonic waveguide geometry and (b) $|E_z|^2$ profile for the photonic waveguide used for the input/output junctions. (c) Schematic illustration and (d) $|E_z|^2$ profile of the LRDLSPW used in the MZI arms.

properties than Bi:YIG, but has not demonstrated ultrafast switching abilities. In this regard, it is ideal for passive magnetoplasmonic devices such as isolators. This ridge sits atop buffer layers of Si_3N_4 ($n_{Si_3N_4} = 1.977$ [52]) and Al_2O_3 ($n_{Al_2O_3}$ =1.746 [33]) with thicknesses of $t_{Si_3N_4} = t_{Al_2O_3} = 100$nm, on a SiO_2 (n_{SiO_2} =1.44 [33]) substrate. The transverse $|E_z|^2$ component of the fundamental photonic TM mode of the structure is shown in Fig. 3.9(b). This photonic waveguide is easily modified into a LRDLSPW with the addition of a thin metal strip, taken here to be a Ag film (n_{Ag} =0.145+11.438i [32]) $w_{Ag} = 160$nm wide and $t_{Ag} = 15$nm thick, placed at the bottom of the Ce:YIG layer, as shown in Fig. 3.9(c). The transverse $|E_z|^2$ component of the fundamental plasmonic TM mode is presented in Fig. 3.9(d). Interestingly, the TM mode is characterized by a long propagation length, L_{prop}, of 3.14mm.

For the TM mode and geometry presently under consideration, propagating along

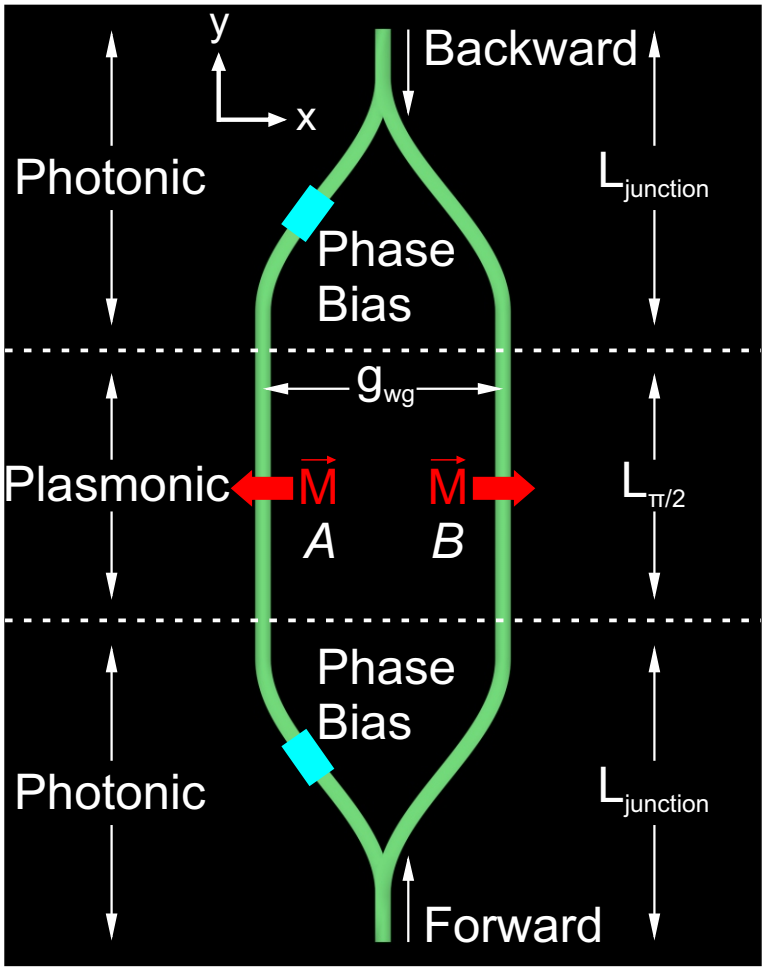

Figure 3.10: Schematic illustration of the MZI geometry (not to scale). A small path length difference creates a $\pi/4$ phase bias at each of the junctions, for a total $\varphi_{bias} = \pi/2$.

the y-axis, the NRPS manifests itself when the magnetization has a component in the $\pm x$ direction. For the calculation of the NRPS (performed via the FDTD simulations), $\mathbf{M}$ is taken to be $\mathbf{M} = \langle +M_S, 0, 0 \rangle$. The values of the NRPS in the plasmonic and photonic waveguide configurations are found to be $\Delta\beta = 3.11\text{rad/mm}$ and $\Delta\beta = 2.93\text{rad/mm}$, respectively. The presence of the metal and the plasmonic nature of the mode enhances the NRPS by 6.14%, and correspondingly, reduces the length to attain a given phase shift by the same fraction. In this regard, the plasmonic mode with sufficiently low loss is advantageous over its analogous photonic counterpart.

A schematic illustration of the MZI setup is depicted in Fig. 3.10. The present design utilizes the phase shifts in the two oppositely magnetized MZI arms in con-

junction with a static phase bias to generate the optical isolation behaviour. Let the $M_x = -M_S$ branch (the left side of Fig. 3.10) be denoted as A, and the $M_x = +M_S$ branch (the right branch of Fig. 3.10) be denoted as B. The two magnetoplasmonic arms are separated by a center-to-center distance of $g_{wg} = 10\mu$m, and are each 505.60 μm long. This length corresponds to the length required to acquire a $\pi/2$ NRPS, $L_{\pi/2} = \pi/(2\Delta\beta)$. Accordingly, a forward propagating mode will experience a phase change of $\delta\beta L_{\pi/2} = \pi/4$ in arm B, and a phase change of $-\delta\beta L_{\pi/2} = -\pi/4$ in arm A. A static phase bias, $\varphi_{bias} = \pi/2$ is built into the device through design of the two branches of the asymmetric adiabatic Y-junctions at the input and output. These junctions divide the input/output coupling waveguide into the two magnetoplasmonic arms over a length $L_{junction} = 55\mu$m to minimize bending losses. In this design, the power coupling between the photonic and plasmonic waveguides is 99.6% due to a high modal overlap.

The operational principal of the device is as follows: In the case of forward propagation, the total phase gained in branch A, φ_A, is a combination of the φ_{bias}, and the $-\pi/4$ shift resulting from the magnetization, resulting in $\varphi_A = \pi/4$. Conversely, the total phase gained in branch B is $\varphi_B = \pi/4$, due to the magnetization only. Hence, at the output $\varphi_A = \varphi_B$ and constructive interference occurs. For a backwards propagating mode, the signs of $\delta\beta$ are reversed. Here, φ_A acquires a $+\pi/4$ contribution from the magnetized magnetoplasmonic section in addition to φ_{bias}, resulting in $\varphi_A = 3\pi/4$. Conversely, in branch B, $\varphi_B = -\pi/4$ due to the magnetized region, and therefore, $\varphi_A - \varphi_B = \pi$. Destructive interference occurs, resulting in optical isolation.

To achieve magnetization reversal in the Ce:YIG while maintaining the integrated nature of this device, thin film SmCo magnets are employed to provide the static fields. To introduce negligible losses into the optical modes, the thin film SmCo magnets are buried $d = 4\mu$m into the SiO_2 substrate. By introducing a gap in the film below one waveguide, opposing magnetic fields can be generated [123, 124]. Considered here is

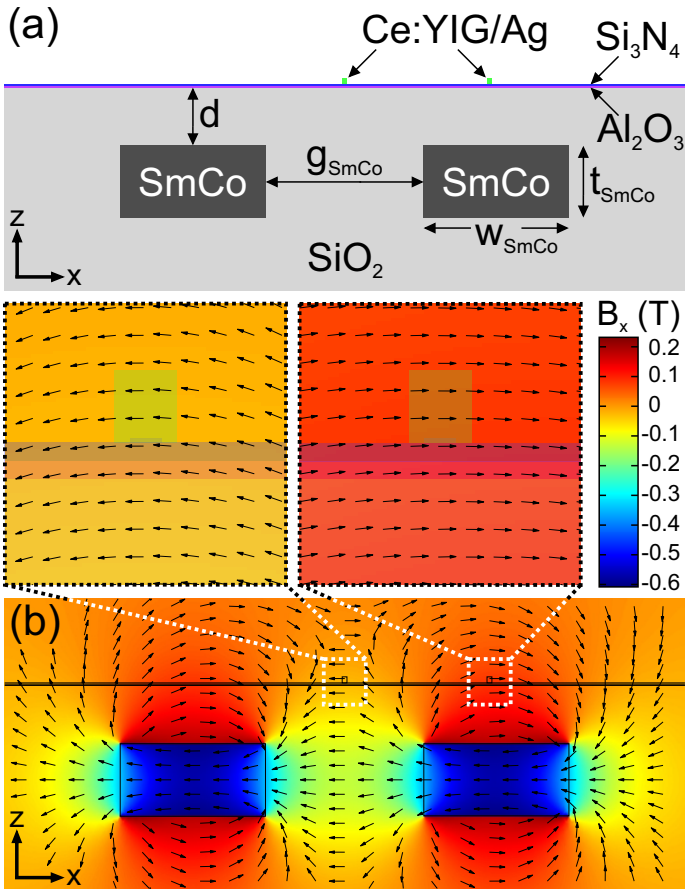

Figure 3.11: (a) Illustration of the SmCo biasing magnet geometry. (b) The B_x component of the flux density generated by the magnets. The vector arrow overlay indicates the direction of **B** in space, while the insets show the magnitude of B_x and the vector direction of **B** over the two waveguides.

a $t_{SmCo} = 5\mu$m thick SmCo film, 31μm wide, and $L_{\pi/2}$ =505.60μm long, positioned under the magnetoplasmonic arms. Typical films of this type are characterized by a remnant flux density of $B_r = 0.8$ T [127] oriented in the -x direction, and such films can be sputtered in the presence of a magnetic field to preferentially orient their in-plane magnetization [128].

Creating a gap of width $g_{SmCo} = 11\mu$m, centered under arm A as shown in Fig. 3.11(a), divides the film into two $w_{SmCo} = 10\mu$m magnets and generates the magnetic field pattern shown in Fig. 3.11(b). Clearly, this configuration establishes opposing magnetic fields over the two MZI arms. Profiles of the magnetic field components through each waveguide are shown in the insets of Fig. 3.11(b). Although

the magnitude of the x-component of the magnetic field is not equal between the two waveguides, it is sufficiently high to completely saturate the Ce:YIG in the $\pm x$ directions, as the Ce:YIG has an extremely low value of $\mu_0 M_S$. Additionally, the uniformity of the fields is of little concern providing that the magnitude is sufficient to saturate the magnetization over the entire cross section of the waveguide. Finally, this configuration is designed to minimize the z-component of the magnetic field over the waveguide.

Fabrication of these magnetoplasmonic isolators is possible using the standard toolkit of nanofabrication processes. A combination of lithography, sputtering in a magnetized environment, and wet etching can define the underlying SmCo magnetic film on an SiO_2 substrate, including the gap required to generate opposing fields. SiO_2 and Si_3N_4 can be deposited through chemical vapour deposition, the Al_2O_3 film can be produced through reactive sputtering of Al in an oxygenated environment, and Ag films can be easily sputtered or evaporated. High-quality Ce:YIG films have recently been shown to be deposited with PLD [92], and utilizing this PLD in conjunction with focused ion beam milling would precisely define the required MZI geometry.

3.3.2 Isolator Performance

The device operation is shown in Fig. 3.12. For a forward propagating wave, constructive interference is observed as expected and depicted in Fig. 3.12(a). Here, the transmission at the output is -2.51dB (measured 10μm from the end of the Y-junction), which is attributed to bending loss in the waveguide as well as absorption loss in the magnetoplasmonic arms. Conversely, destructive interference is observed for the backward propagating wave, as shown in Fig. 3.12(b), where the transmission is -25.33dB (measured 10μm from the Y-junction). The device exhibits a low insertion loss of only 2.51dB, and an effective isolation of 22.82dB between the forward and backwards waves. The choice to implement a low-loss LRDLSPW with a long propa-

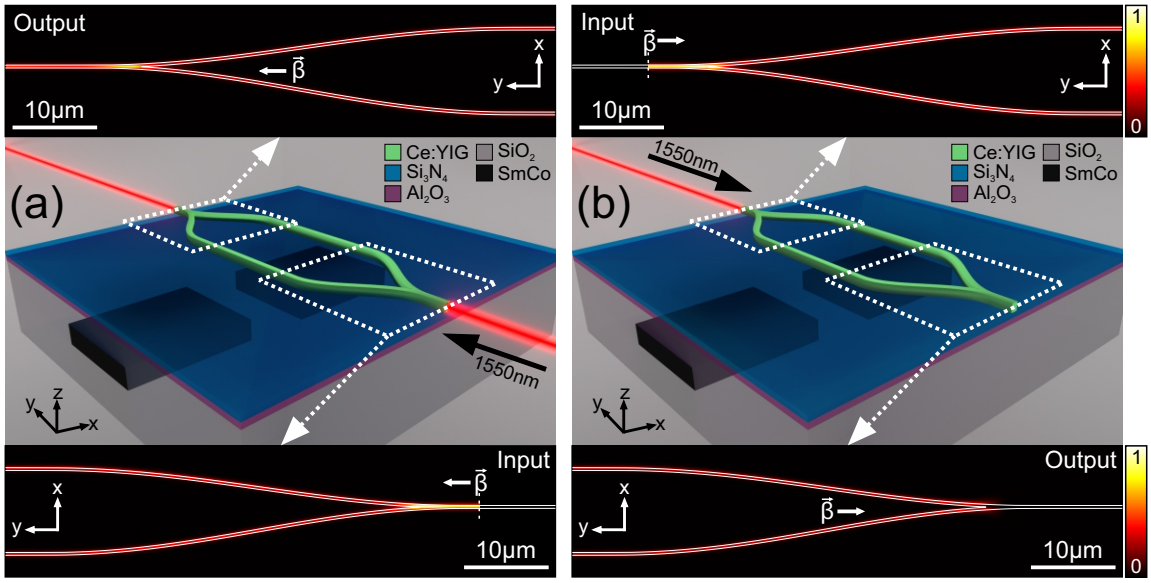

Figure 3.12: (a) An illustration of the device operation with forward propagating light. The bottom inset shows the $|E_z|^2$ for input region of the device, while the top inset shows the $|E_z|^2$ profile for the output region of the device. (b) An illustration of the device operation with backwards propagating light. The top inset shows the $|E_z|^2$ for input region of the device, while the bottom inset shows the $|E_z|^2$ profile for the output region of the device. All of the field profile slices are taken 100 nm above the Ce:YIG-Si_3N_4 interface, and that the devices are not to scale, while the $|E_z|^2$ profiles are depicted to scale.

gation length implies that the absorption losses incurred by the plasmonic mode are only 0.68dB. The remaining 1.83dB of the insertion loss is due to bending/radiative loss in the Y-junctions. For comparison, rib waveguide isolators have shown insertion losses of 15.6dB and extinction ratios of 19dB [125], while bonded Ce:YIG-SOI devices exhibited insertion losses of 13dB and an extinction ratio of 30dB [75].

This device is ideal for implementation into integrated optical circuits, as it occupies a small footprint of $6.4 \times 10^{-3}\text{mm}^2$, which is an improvement compared to the footprint of rib waveguide ($\sim 0.1\text{mm}^2$) [125] or bonded Ce:YIG-SOI ($\sim 0.2\text{mm}^2$) [75] isolators. Employing the NRPS enhancement, offered by the plasmonic arms, allows the arm length to be reduced by 31μm, and the total footprint to be reduced by $318\mu\text{m}^2$ in comparison to a device with purely photonic arms.

While the this MZI based nonreciprocal magnetoplasmonic platform provides efficient passive isolation, it simultaneously opens a new pathway for optical intensity

modulation. Since the NRPS in the MZI depends on M_x, introduction of temporal variations in M_x via a transient magnetic field can be used to modulate $\Delta\beta$ in the MZI. In this way, the output port can be switched between constructive and destructive interference, and the optical intensity can be modulated. Additionally, this geometry allows us to exploit the inherent properties of the magnetization dynamics that were deemed problematic in the operation of the waveguide phase shifter described in Section 3.2, and facilitates the design of two completely new types of optical intensity modulator never before realized in the nanoplasmonic regime.

3.4 A Magnetoplasmonic Electrical-to-Optical Clock Multiplier[3]

One key device that has yet to manifest in the optical regime is the clock multiplier (CM). The role of a CM is to scale up the clock signal frequency at different locations within a system. A nanoplasmonic CM would transfer an electrical signal to the optical domain, upconvert the modulation frequency, and synchronize the electrical and optical logic systems operating at different speeds. A key requirement of an optical material for this task is to facilitate a mechanism of generating regular optical modulation between stimulating electrical pulses, in the absence of excitation. Previously explored material properties, such as electro-optic, thermo-optic, and nonlinear effects are unable to display this characteristic, as they do not exhibit long-lasting, resonantly driven oscillations that can be mapped into the phase of the optical wave.

Resonant magnetic systems do in fact display this property, as well as facilitate a mechanism of transferring the nonlinear dynamics onto an optical wave. Recall from the analysis of the waveguide phase shifter in Section 3.2 that a transient magnetic

[3]A version of this section has been published as C. J. Firby and A. Y. Elezzabi, "A magnetoplasmonic electrical-to-optical clock multiplier," Appl. Phys. Lett. 108(5), 051111 (2016).

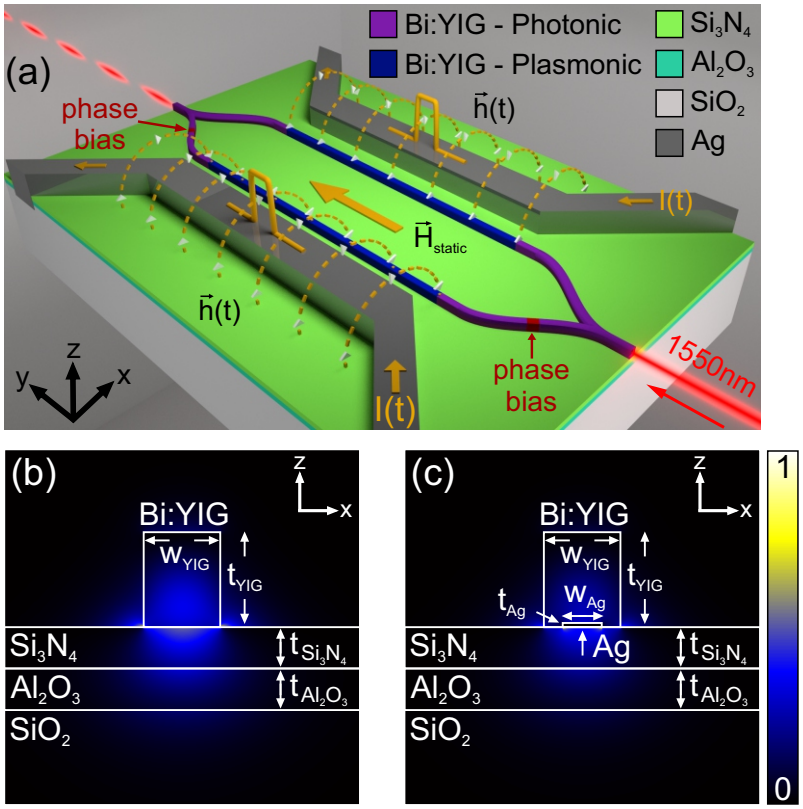

Figure 3.13: (a) Illustration of the electrical-to-optical CM. (b) $|E_z|^2$ profile and geometry of the photonic waveguide. (c) $|E_z|^2$ profile and geometry of the plasmonic waveguide.

field exerts a torque on the magnetization vector and changes the magnetization state. Upon conclusion of the transient pulse, the magnetization precesses around the static magnetic field vector as it relaxes back to its initial state. This precession manifests as an oscillatory response of the perpendicular magnetization components, initiating dynamic actuation of the NRPS. Integrating this into an MZI converts the dynamic NRPS into a sequence of optical clock pulses.

3.4.1 Clock Multiplier Design

To achieve this end, several minor modifications had to be made to the MZI design, where the altered design is illustrated in Fig. 3.13(a). The LRDLSPW platform remains unchanged, as do the Y-junction lengths and distance between the two MZI arms. However, since this is a dynamic modulator, the MO garnet is taken to be

Bi:YIG, due to its verified high speed modulation capabilities [118]. The gyrotropic Bi:YIG core retains the same dimensions as its Ce:YIG predecessor, with dimensions of $w_{YIG} = 320$nm and $t_{YIG} = 400$nm, but due to the slight difference in refractive index between the Ce:YIG and Bi:YIG (2.21 vs 2.3), the Si_3N_4 and Al_2O_3 buffer layer thicknesses were modified to $t_{Si_3N_4} = t_{Al_2O_3} = 175$nm. The Ag dimensions remain the same at $w_{Ag} = 160$nm and $t_{Ag} = 15$nm, the structure remains set atop a SiO_2 substrate, and the SmCo magnets are removed. Despite the change in core material, these minor modifications allow the fundamental mode to maintain a characteristic propagation length of $L_{prop} = 3.0$mm. The transverse $|E_z|^2$ profiles of the photonic and plasmonic modes are shown in Figs. 3.13(b)-(c).

The static biasing field is generated along the y-axis by an external magnet over the magnetoplasmonic arms, i.e., $\mathbf{H}_{static} = \langle 0, +H_y, 0 \rangle$. Transient magnetic fields are applied to the arms as a result of current pulses passing through two Ag microwires (having a 2μm$\times 2\mu$m cross sectional area, length of $L_{\pi/2}$, and a separation of 16μm), as shown in Fig. 3.13(a). The current pulses comprising the clock signal, $I(t)$, resemble practical square wave pulses having finite rise and fall times and a super-Gaussian form. The arrangement of the Ag microwires generates magnetic fields of opposite polarity, $\mathbf{h}(t) = \langle 0, 0, \pm h_z(t) \rangle$, over the two MZI arms.

Since the primary electric field component of the plasmonic mode is E_z, a NRPS will only occur when $\mathbf{M}$ is oriented along the $\pm x$-axis. In calculating the NRPS of the waveguide, $\mathbf{M}$ is taken to be $\mathbf{M} = \langle +M_S, 0, 0 \rangle$. The NRPS is found to be $\Delta\beta = -1.77$rad/mm, and $L_{\pi/2} = 886.1\mu$m. The NRPS is dependent on M_x, where introducing a temporal variation in M_x modulates $\Delta\beta$ and the interference condition. The temporal evolution of the $\mathbf{M}$ can be modeled through the Landau-Lifshitz-Gilbert equation, while the transmission of the magnetoplasmonic CM is determined via the FDTD simulations.

3.4.2 Operation Principle

For the operation of a tunable CM, the magnitude of static magnetic field is crucial in the device operation as it defines initial/final states of the **M**, dictates the FMR frequency, and determines the precession axis of **M**. In this case, $\mathbf{H}_{static} = \langle 0, +H_y, 0 \rangle$ initially saturates **M** along the y-axis, $\mathbf{M} = \langle 0, +M_S, 0 \rangle$. Short, transient magnetic field pulses, $\mathbf{h}(t)$, tip the **M** away from this axis, and it precesses around the y-axis as it relaxes. This precession manifests as a decaying oscillatory response in M_x and M_z with a characteristic FMR at the Larmor frequency, given by Eqn. 2.84, substituting the H_y for the static magnetic field component.

Figure 3.14 depicts the M_x/M_S amplitude of the precessional oscillations that can be excited as functions of the static magnetic field, H_y, and peak transient magnetic field, $h_{z,pk}$, for pulse widths, $\tau_p = 100$ps (Fig. 3.14(a)) and $\tau_p = 500$ps (Fig. 3.14(b)). When $\mathbf{h}(t) \neq 0$, **M** is subject to an effective magnetic field due to the superposition of $\mathbf{h}(t)$ and $\mathbf{H}_{static}$. and it precesses around this resultant vector. When the transient field is turned off (i.e. $\mathbf{h}(t) = 0$), **M** is subject only to $\mathbf{H}_{static}$, and it precesses around the y-axis with a frequency, ν_L. The magnitude of M_x depends on the orientation of the **M** as $\mathbf{h}(t)$ turns off. For maximum amplitude oscillations of M_x, the precession must begin when $M_y = 0$. This condition can be satisfied, as shown in Fig. 3.14(a)-(b), with an appropriate choice of $h_{z,pk}$, H_y, and τ_p.

Since the magnitude of **M** is a constant value (M_S), the tip of the **M** vector moves on the surface of a sphere of radius M_S (denoted as the magnetization sphere). The initial deflection of **M** is set by its orientation and the direction of the effective magnetic field vector. Here, the tip of **M** traverses around the magnetization sphere within the plane whose normal is the effective field vector, and which passes through the initial state (denoted as the plane of precession, or POP). For the condition $M_y = 0$ to be satisfied, **M** must lie within the x-z plane. When $h_{z,pk} = H_y$, the effective field vector makes an angle of 45°with both the y- and z- axes, and the POP

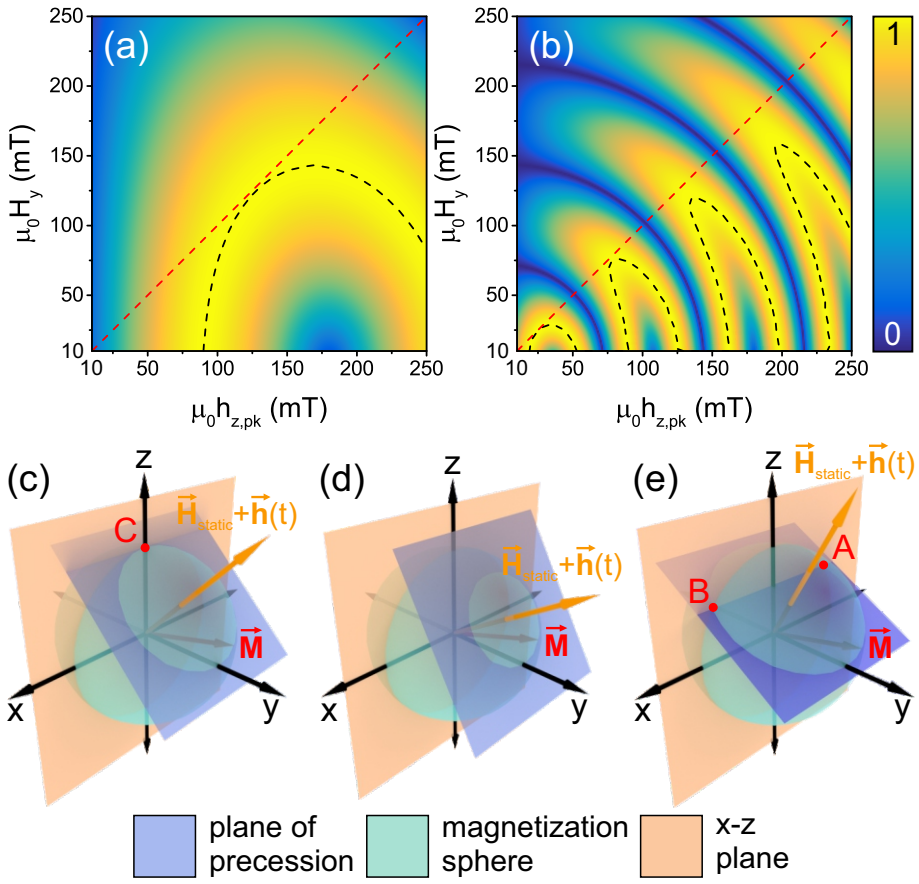

Figure 3.14: Plots of the excited M_x/M_S precession oscillation amplitude as functions of H_y and $h_{z,pk}$ for (a) τ_p = 100ps and (b) τ_p = 500ps. The black dashed lines depict the contour where the M_x/M_S oscillation amplitude is 1 (maximum), while the red dashed line marks the $h_{z,pk} = H_y$ boundary. (c)-(e) show the plane of precession around the effective magnetic field when (c) $h_{z,pk} = H_y$, (d) $h_{z,pk} < H_y$, and (e) $h_{z,pk} > H_y$.

intersects the z-axis at one point (point C in Fig. 3.14(c)) on the magnetization sphere. When $h_{z,pk} < H_y$, the effective field vector lies closer to the y-axis, the POP does not intersect the x-z plane at any point on the sphere (Fig. 3.14(d)). Interestingly, for $h_{z,pk} > H_y$, the effective field vector lies closer to the z-axis, and the POP intersects the x-z plane on the magnetization sphere, as shown in Fig. 3.14(e). Therefore, maximum amplitude oscillations in M_x can be excited.

The arch patterns within Figs. 3.14(a)-(b) reflect such dynamics. When $h_{z,pk} < H_y$, the amplitude of the M_x oscillations is < 1, since the $M_y = 0$ condition can never be satisfied. However, when $h_{z,pk} > H_y$, the trajectory of $\mathbf{M}$ satisfies the $M_y = 0$ condition at two points (within the x-z plane) per revolution (points A and B in Fig.

3.14(e)). If the fields are chosen such that **M** ends at one of these states, M_x can be made to oscillate between ± 1. At a fixed magnitude of H_y, the corresponding $h_{z,pk}$ can be applied such that **M** is deflected from its initial state to point A. Increasing $h_{z,pk}$ further increases the effective precession frequency, allowing **M** to overshoot point A and reach point B by the time $\mathbf{h}(t)$ concludes. These two points manifest as the two branches of the arch patterns in Figs. 3.14(a)-(b). At high values of H_y, the effective field vector shifts more towards the y-axis, and thus the points A and B coalesce into a single point (point C) and the two branches of the arch merge into one. Points A and B should merge into point C when $h_{z,pk} = H_y$. However, deviation from this is observed, especially in Fig. 3.14(b) for $\tau_p = 500$ps pulses. This deviation is the result of the nonzero rise and fall time of the $\mathbf{h}(t)$, where **M** is subject to a time varying effective field vector. Here, **M** is deflected as the effective field vector moves, and when it reaches its peak, **M** no longer resides at the initial state. In this way, the POP shifts and the condition required to maximize the M_x amplitude changes.

As shown in Fig. 3.14(b), a long τ_p allows **M** to complete several revolutions around the effective field vector, which is characterized by multiple arches and coalescing points in the H_y vs $h_{z,pk}$ plot, resulting in multiple frequencies in the output optical train. Clearly, this behaviour is undesirable for a single frequency CM operation. Since the CM frequency is proportional to H_y, higher precession frequencies are more readily attainable when utilizing electrical pulses of shorter durations.

To map the long-lasting M_x oscillations into optical intensity modulation, the MZI arm length must be $L_{\pi/2}$. With an NRPS of $\Delta\beta = -1.77$rad/mm, $L_{\pi/2} = 886.1\mu$m. This interaction length provides the maximum NRPS and intensity modulation. To limit the current pulses ($I(t)$) to a practical range, peak currents of less than 1A are needed, which in the described transmission lines, generates peak magnetic fields of ± 36mT at the waveguides. Applying higher static fields, and accessing higher FMR frequencies, requires compensation for the reduction in M_x oscillation amplitude by

increasing $L_{\pi/2}$. Additionally, the output frequency of the CM is denoted here as ν_{CM}, which is equivalent to the Larmor frequency, ν_L.

3.4.3 Clock Multiplier Performance

An exemplary parameter set for which maximum optical modulation can be attained includes $\tau_p = 500$ps, $\mu_0 h_{z,pk} = 19$mT (i.e., a peak current of 0.53A), and $\mu_0 H_y = 10$mT (corresponding to a CM output frequency of $\nu_{CM} = 280$MHz). Since higher ν_{CM} requires shorter τ_p, the scenario where $\tau_p = 100$ps is considered. As shown in Fig. 3.14(a), for a maximum attainable field of $\mu_0 h_{z,pk} = 36$mT, the CM frequencies between $\nu_{CM} = 280$MHz and $\nu_{CM} = 5.6$GHz are attainable for static field biases of $\mu_0 H_y = 10$mT and $\mu_0 H_y = 200$mT, respectively. These field configurations correspond to reduced M_x/M_S oscillation amplitudes of 0.59 and 0.34, (or NRPS of $\Delta\beta = -1.04$rad/mm and $\Delta\beta = -0.60$rad/mm), respectively. Therefore, the MZI arm length must be increased to $L_{\pi/2} = 1510.7\mu$m and $L_{\pi/2} = 2607.4\mu$m. Note that in each case, L_{prop} is longer than $L_{\pi/2}$, and hence a complete $\pi/2$ NRPS can always be obtained.

Figures 3.15(a)-(c) display the MZI transmission as a function of time. For $\tau_p = 500$ps (Fig. 3.15(a)), $\mu_0 h_{z,pk} = 19$mT and $\mu_0 H_y = 10$mT produce an optical pulse train at $\nu_{CM} = 280$MHz, having transmission between -1.85dB and -18.11dB. Here, a total modulation depth of 16.26dB is attainable when the NRPS is oscillating maximally. Figure 3.15(b) depicts the transmission for $L_{\pi/2} = 886.1\mu$m and $L_{\pi/2} = 1510.7\mu$m in response to a $\tau_p = 100$ps pulse at $\mu_0 h_{z,pk} = 36$mT and $\mu_0 H_y = 10$mT. At $L_{\pi/2} = 886.1\mu$m, the 280MHz output oscillates between -2.30dB and -10.96dB, resulting in a modulation depth of 8.65dB. Optimizing $L_{\pi/2}$ to 1510.7μm improves the modulation depth to 16.26dB. Similarly, Fig. 3.15(c) shows the transmission for a $\tau_p = 100$ps pulse with $\mu_0 h_{z,pk} = 36$mT and $\mu_0 H_y = 200$mT. These parameters produce a $\nu_{CM} = 5.6$GHz signal, modulated between -3.04dB and -7.65dB (a modu-

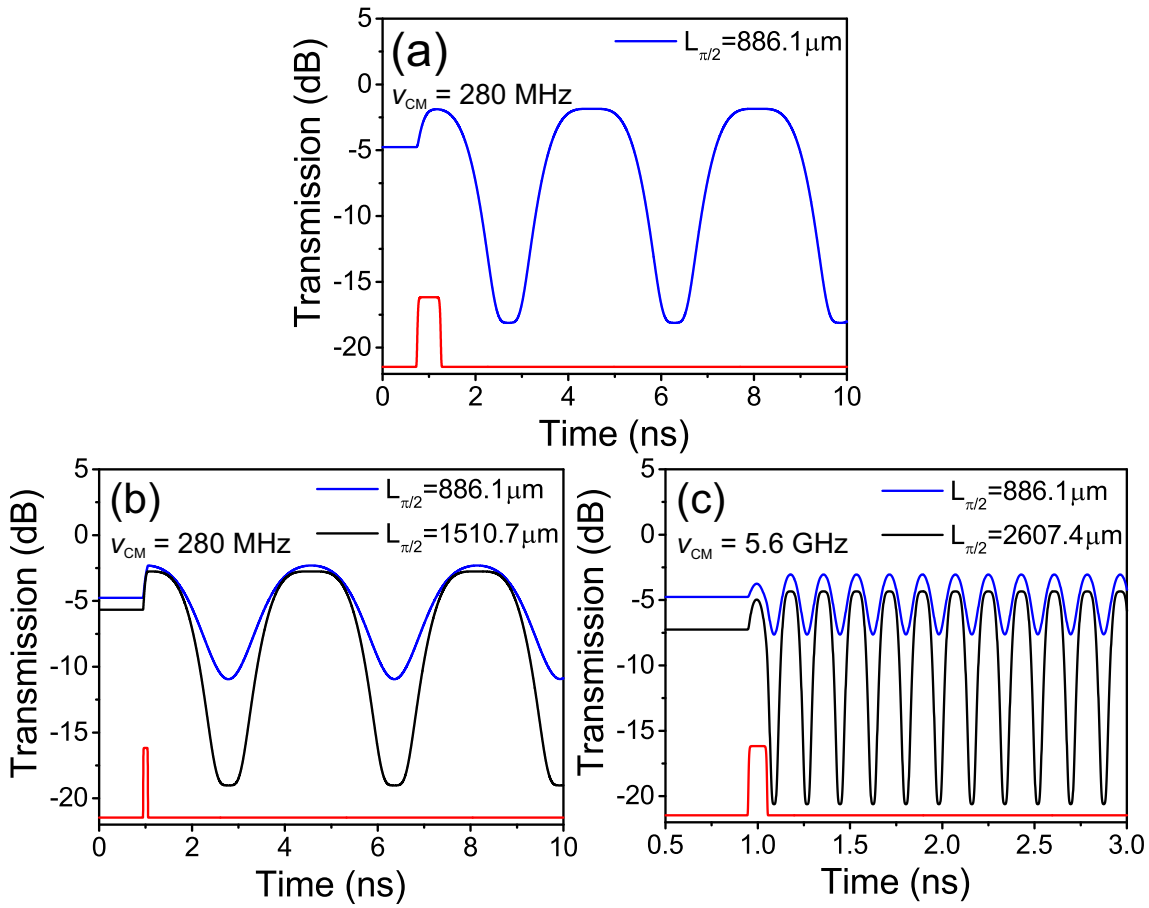

Figure 3.15: MZI transmission versus time for (a) $\tau_p = 500$ps, $\mu_0 h_{z,pk} = 19$mT, and $\mu_0 H_y = 10$mT, (b) $\tau_p = 100$ps, $\mu_0 h_{z,pk} = 36$mT, and $\mu_0 H_y = 10$mT, and (c) $\tau_p = 100$ps, $\mu_0 h_{z,pk} = 36$mT, and $\mu_0 H_y = 200$mT. The red curve represents the triggering pulse (not shown to scale).

lation depth of 4.61dB) for $L_{\pi/2} = 886.1\mu$m. At the optimized $L_{\pi/2} = 2607.4\mu$m, the modulation depth is improved to 16.26dB.

Similar to electronic CMs, the MZI must be able to respond to a continuous train of electrical pulses comprising the triggering clock signal. Since the M_x response decays over the relaxation time of the Bi:YIG, one must set a minimum threshold for the magnetization (and NRPS) decay to initiate the next triggering cycle. As a design figure of merit, the threshold is assigned to the time when the M_x amplitude drops to 50% of its initial value. The clock repetition frequency, f_{rep}, can then be adjusted over a wide range. For synchronous clock operation, one must account for a slight difference between the transit time for the initial tipping of **M** at the onset of

the pulse train, and the tipping required to re-establish the original oscillation after decaying to 50%. As such, the perturbation between the first two triggering pulses and the remaining train is on the order of 0.001% of the triggering clock frequency in the examples presented here.

The three optimized exemplary cases discussed above are depicted under stimulus from a regular clock signal in Fig. 3.16. The $\tau_p = 500$ps electrical pulse train with $\mu_0 h_{z,pk} = 19$mT can be applied at $f_{rep} = 132.0$kHz (Fig. 3.16(a)). At $\nu_{CM} = 280$MHz, the CM exhibits a multiplication factor of 2.1×10^3. Similarly, Fig. 3.16(b) depicts the output due to an electrical pulse train of $\tau_p = 100$ps ($\mu_0 h_{z,pk} = 36$mT) pulses with $f_{rep} = 228.0$kHz, where the input frequency is multiplied by a factor of 1.2×10^3 to $\nu_{CM} = 280$MHz. Despite these two cases possessing the same FMR frequency, this case has a lower multiplier due to the threshold condition and the nonlinear relaxation of the precession. Since the decay process is nonlinear, it takes less time for these oscillations to decay to 50% of their initial value compared to the case in Fig. 3.16(a), and as such, f_{rep} is greater and the multiplication factor is reduced. Figure 3.16(c) shows the output for the higher FMR, $\nu_{CM} = 5.6$GHz. A train of $\tau_p = 100$ps pulses can be applied at $f_{rep} = 4.9$MHz, resulting in a 1.1×10^3 multiplication factor. Clearly, the MZI is capable of large multiplication factors, in excess of 10^3, and is tunable over a wide range of frequencies. The repetition rate can be tailored at the cycle level, and relaxing the set threshold condition and employing detectors with higher sensitivity can lead to much larger multiplication factors. It is interesting to note that, due to the low duty cycle of the electrical clock, the average power dissipation is less than 5mW.

The magnetoplasmonic electrical-to-optical CM is a unique device that currently has no optical equivalent. This novel functionality is facilitated only through the use of MO materials and MO effects. The second unique optical modulator that can be derived from the magnetoplasmonic MZI platform exploits the highly nonlinear

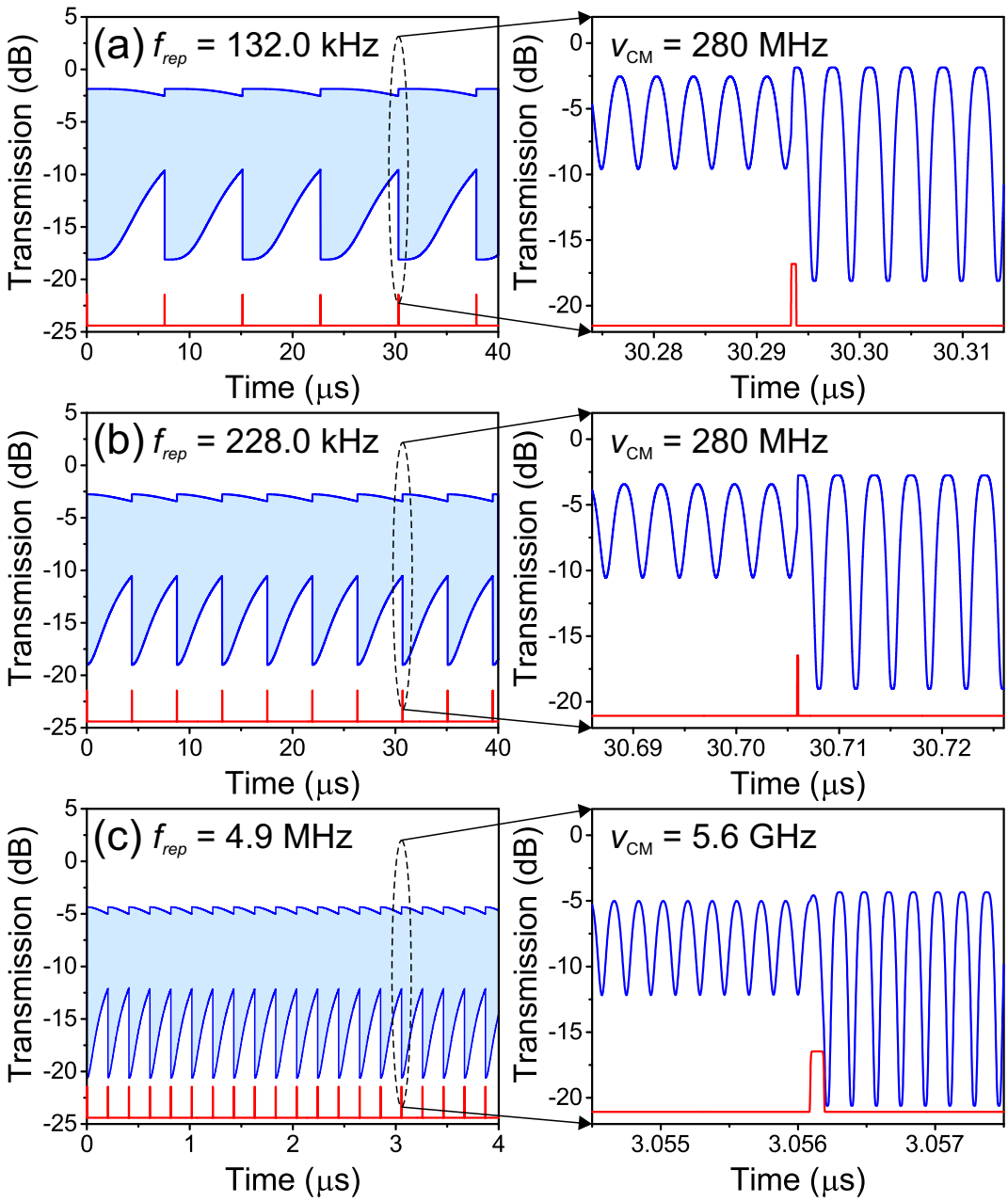

Figure 3.16: Transmission response to a train of pulses with the following properties: (a) τ_p = 500ps, $\mu_0 h_{z,pk}$ = 19mT, $\mu_0 H_y$ = 10mT, and f_{rep} = 132.0kHz; (b) τ_p = 100ps, $\mu_0 h_{z,pk}$ = 36mT, $\mu_0 H_y$ = 10mT, and f_{rep} = 228.0kHz; (c) τ_p = 100ps, $\mu_0 h_{z,pk}$ = 36mT, $\mu_0 H_y$ = 200mT, and f_{rep} = 4.9MHz. These plots depict the overall envelope of the transmission, and on this time scale, the individual optical pulses are indistinguishable. The optical pulse train occurs in the blue band, as shown in the insets. The red curve represents the triggering pulses (not shown to scale).

characteristics of the magnetization dynamics to modulate an optical signal in a unique manner.

3.5 Magnetoplasmonic RF Mixing and Nonlinear Frequency Generation[4]

Electrical control of optical signals is desirable for integrating both CMOS and optical components into a single platform. In recent years, a vast array of device architectures, achieved by altering the properties of a nanoplasmonic waveguide's constituent materials, have been developed to meet this demand [13]. These include the modification of refractive indices via the thermo-optic effect [35], electro-optic effect [56], or free carrier effects [85]. These mechanisms present a fundamental limitation: within the response bandwidth of the material, only the driving source frequency is mapped onto the optical mode. For on-chip applications, this implies that optical modulation at multiple distinct frequencies requires numerous single frequency sources, or a widely tunable on-chip electrical source. These solutions are uninviting, for the reason that the footprint consumed by the electrical components is large. There is a need for a tunable mechanism of generating optical modulation at many distinct frequencies while employing only one or two electrical sources.

This limitation can be surpassed by considering the magnetoplasmonic MZI device described in Section 3.4 in a different mode of operation. By applying a sinusoidal driving signal, the nonlinear dynamics of $\mathbf{M}$ can be accessed in a new way. Previous studies examining ferromagnetic response in the microwave regime have shown that the nonlinear behaviour of $\mathbf{M}$ can facilitate harmonic generation when driven by a radio-frequency (RF) signal [129, 130]. As well, the nonlinearity of the response can induce frequency mixing when $\mathbf{M}$ is driven with two RF signals [130, 131]. Our lab has also demonstrated that nonlinear frequency generation can be enhanced through parametric excitation of the Suhl instability [132]. Given the extreme nonlinearity of the magnetization dynamics, it would be advantageous to map the nonlinearity onto

[4] A version of this section has been published as C. J. Firby and A. Y. Elezzabi, "Magnetoplasmonic RF mixing and nonlinear frequency generation," Appl. Phys. Lett. 109(1), 011101 (2016).

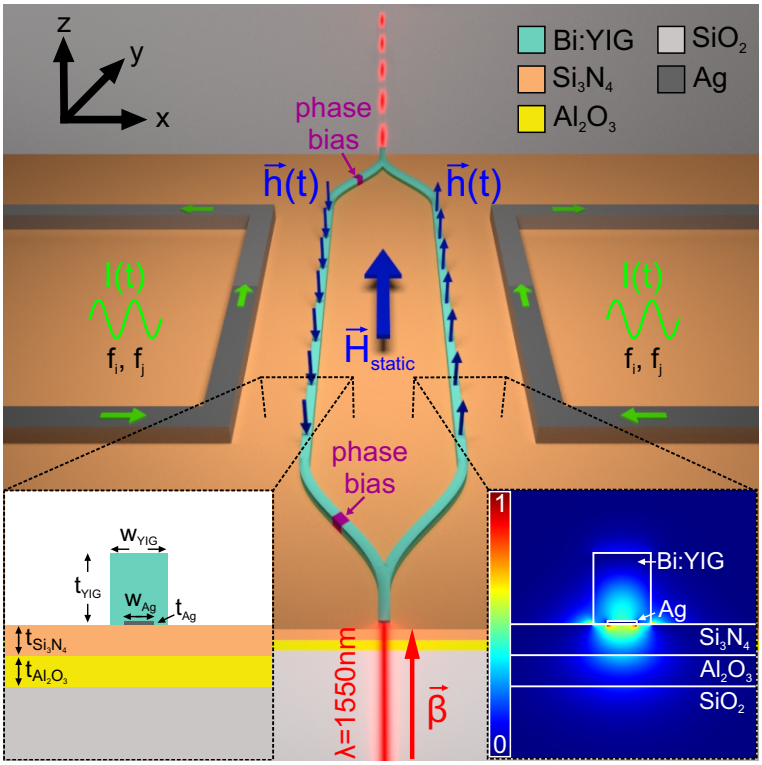

Figure 3.17: Schematic illustration of the magnetoplasmonic MZI geometry for RF mixing and nonlinear frequency generation. The left inset shows a cross section of the LRDLSPW arm structure, while the right inset shows the $|E_z|^2$ profile.

optical intensity modulation to generate these additional spectral components (i.e., harmonics, etc.) within the transmitted signal

3.5.1 RF Mixer Design

To achieve this nonlinear frequency mixing functionality, the MZI modulator geometry described in Section 3.4 is employed, with one minor alteration: the input current pulses that drove the CM are replaced with sinusoidal signals here, as depicted in Fig. 3.17. Propagating a sinusoidal RF current signal $I(t)$ through such transmission lines generates sinusoidally varying magnetic fields, $\mathbf{h}(t) = \langle 0, 0, \pm h_z(t) \rangle$, that nonlinearly drive $\mathbf{M}$ in the LRDLSPW arms. The initial magnetization state, $\mathbf{M} = \langle 0, +M_S, 0 \rangle$, is fixed by an externally applied magnetic field oriented in the +y-direction, $\mathbf{H}_{static} = \langle 0, +H_y, 0 \rangle$, as it was for the previous mode of operation.

Although the applied magnetic fields are not oriented transverse to the waveguide, the nonlinear dynamics induce a transverse magnetization component (M_x), which generates the NRPS within the MZI arms. Driving the magnetization between $M_x/M_S = \pm 1$ modulates the transmission of the MZI between its minimum and maximum values. $\mathbf{M}$ is approximated as uniform over the MZI arm due to the low magnetic field variation over the Bi:YIG core cross section, uniformity of the structure in the propagation direction, and short length of the arms relative to the driving RF wavelengths of interest.

The dynamic magnetization model of Eqn. 2.86 was combined with FDTD simulations of the MZI, and the resultant temporal variations in the transmitted intensity were determined. The frequency spectrum of the signal depicts the nonlinear modulation frequency generation provided by the LLG dynamics.

3.5.2 Response to a Single Driving Frequency

To demonstrate the versatile range of applications of nonlinear frequency generation in the magnetoplasmonic MZI, several exemplary parameter sets are considered. Each case is statically-biased at $\mu_0 H_y = 10\text{mT}$, and is driven at a single frequency: $f_d = 280\text{MHz}$ (Case 1), $f_d = 420\text{MHz}$ (Case 2), and $f_d = 1\text{GHz}$ (Case 3). Here, $h_z(t)$ takes the form, $h_z(t) = h_z \sin(2\pi f_d t)$. The frequency spectra of the modulated intensity at the MZI output port are plotted as a function of driving field amplitude, $\mu_0 h_z$, in Figs. 3.18(a)-(c). In all spectra presented in Fig. 3.18, the DC frequency component, arising from the nonzero transmission occurring when $M_x/M_S = 0$ in the MZI arms, has been filtered out for clarity. Two additional spectral features are worth noting. For low amplitude driving signals, there exists a spectral component at the driving frequency, f_d, with sidebands occurring at the FMR Larmor frequency (given by Eqn. 2.84), and at $f = 2f_d - \nu_L$. Note that $\nu_L = 280\text{MHz}$ in Cases 1-3. As the amplitude of the driving signal at f_d is increased, the frequency of the sidebands shifts in a

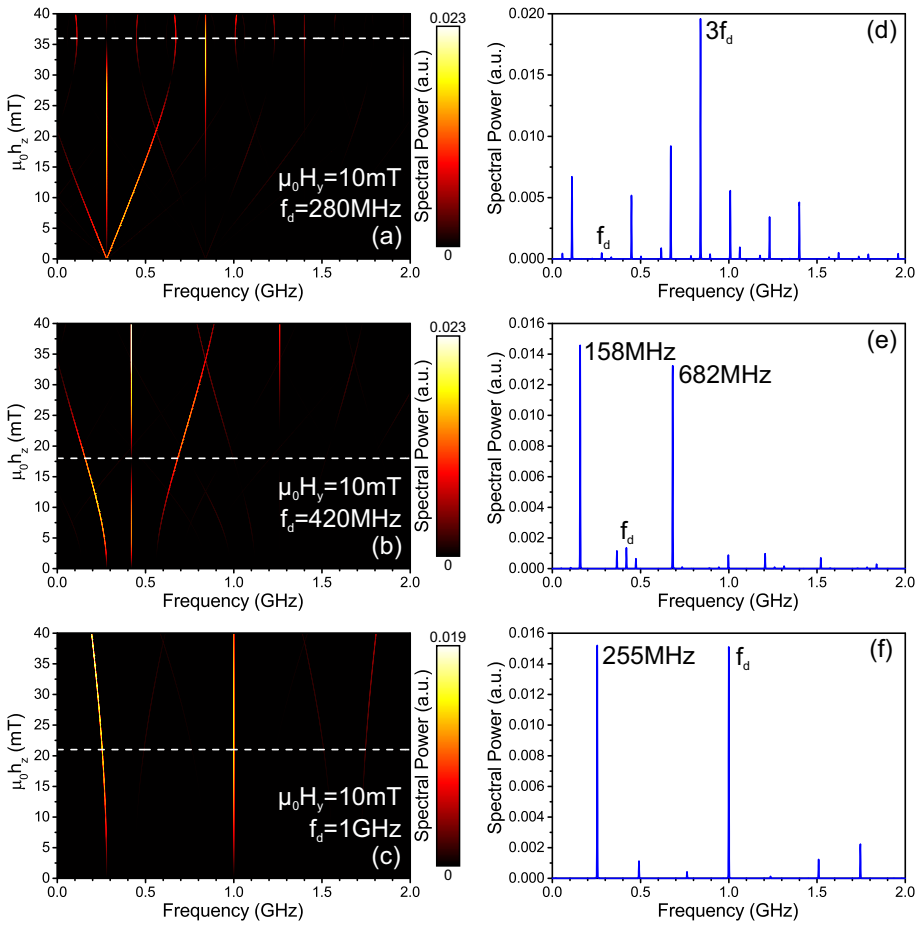

Figure 3.18: Plots of the frequency spectra of the modulated intensity signal at the MZI output versus $\mu_0 h_z$ for $\mu_0 H_y = 10$mT and (a) $f_d = 280$MHz, (b) $f_d = 420$MHz, and (c) $f_d = 1$GHz. (d) depicts the spectral profile denoted by the dashed line in (a), for $\mu_0 h_z = 36$mT, where the $3f_d$ component is enhanced. (e) shows the spectrum marked by the dashed line in (b), when $\mu_0 h_z = 18$mT, where the dominant spectral components are those at 158MHz and 682MHz, due to the nonlinear sideband formation. (f) illustrates the spectrum indicated by the dashed line in (c), for $\mu_0 h_z = 21$mT, where f_d and the downconverted frequency of 255MHz are excited with nearly equal amplitude.

nonlinear manner. This is followed by the appearance of odd harmonics of f_d. Since the Bi:YIG is driven by a transverse linearly polarized magnetic field, the even order mixing effects are only present in the longitudinal component, M_y, while odd order mixing effects are present in the transverse M_x and M_z components. This harmonic response is observed in the intensity signal, as the NRPS in the MZI only maps the temporal variations in M_x onto the optical waveform.

The frequency spectrum for Case 1 is shown in Fig. 3.18(a). Of particular interest

in this situation is the frequency spectrum at a driving amplitude of $\mu_0 h_z = 36$mT (plotted in Fig. 3.18(d)). This configuration suppresses the driving frequency, $f_d = \nu_L = 280$MHz, and the largest spectral component is in fact the third-harmonic of f_d at 840MHz. The spectral amplitude of the $3f_d$ frequency is 40 times greater than that of the driving frequency spectral amplitude at f_d. Other sizable spectral components are produced at 112MHz, 448MHz, 671MHz, 1GHz, 1.231GHz, and 1.4GHz. Filtering schemes can be employed to isolate just the sought after frequency component. Indeed, this platform provides a means of generating the third-harmonic of the input RF signal.

In a similar manner, the frequency spectrum for Case 2 is displayed in Fig. 3.18(b). At an amplitude of $\mu_0 h_z = 18$mT, the driving frequency ($f_d = 420$MHz in this case) is again suppressed in the spectrum, as shown in Fig. 3.18(e). However, with these parameters, the most prominent spectral components are the two peaks at 158MHz and 682MHz, each spaced a distance of 262MHz from f_d. These peaks are excited with nearly equal amplitude, and are the result of the nonlinear sideband formation within the spectrum. Interestingly, the spectral separation is a function of both driving frequency and amplitude, and all other spectral components have amplitudes less than 10% of 158MHz peak. Clearly, the device can be driven at a single frequency, but the output modulation is split between two different and distinct frequency components.

In Case 3, driving the device at a higher frequency of $f_d = 1$GHz is investigated. The resultant spectra as functions of driving amplitude are shown in Fig. 3.18(c). Specifically, the spectrum of the transmission signal driven with a peak field magnitude of $\mu_0 h_z = 21$mT is displayed in Fig. 3.18(f). In this case, two primary spectral peaks are observed: one at f_d, and one at 255MHz, each with equal amplitude. The 255MHz peak represents only a small deviation from the FMR frequency at $\nu_L = 280$MHz. One can conclude that by driving the device at high frequency,

the nonlinear response generates a component near the significantly lower FMR frequency. Such down-conversion response is desirable in complex computing architectures, where the various subsystems operate at different modulation frequencies.

3.5.3 Response to Multiple Driving Frequencies

Simultaneously driving the MZI with two or more RF signals presents a much more complex situation. Here, similar sideband and harmonic generation for each of the two individual frequencies can be realized, however, the strong nonlinear mixing process gives rise to numerous additional spectral components. Two exemplary spectra are displayed in Figs. 3.19(a)-(b) as functions of the driving field amplitude. Again, the DC frequency component has been filtered out for clarity. Both spectra are taken under a static field bias of $\mu_0 H_y = 25\text{mT}$ (i.e., $\nu_L = 700\text{MHz}$), and are driven by two frequencies, f_{d1} and f_{d2}, such that $h_z(t) = h_z[\sin(2\pi f_{d1} t) + \sin(2\pi f_{d2} t)]$. The two exemplary cases consider $f_{d1} = 275\text{MHz}$ and $f_{d2} = 420\text{MHz}$ (Case 4), and $f_{d1} = 700\text{MHz}$ and $f_{d2} = 1\text{GHz}$ (Case 5).

Case 4 is depicted in Fig. 3.19(a). To illustrate the frequency content obtained by driving the MZI with two RF signals, the spectrum obtained at a driving field amplitude of $\mu_0 h_z = 20\text{mT}$ is expanded in Fig. 3.19(c). As stated earlier, the transverse magnetization component, and the intensity transmission from output port of the device, exhibits odd ordered mixing effects. This behaviour is confirmed by a perturbation expansion of the LLG equation, predicting the presence of eight first- and third-order mixed frequencies due to the nonlinear mixing [131]. Each of these frequencies are present in Fig. 3.19(c), as shown by the peaks at $f_{d1} = 275\text{MHz}$, $f_{d2} = 420\text{MHz}$, $3f_{d1} = 825\text{MHz}$, $3f_{d2} = 1.26\text{GHz}$, $2f_{d1} + f_{d2} = 970\text{MHz}$, $2f_{d1} - f_{d2} = 130\text{MHz}$, $2f_{d2} + f_{d1} = 1.115\text{GHz}$, and $2f_{d2} - f_{d1} = 565\text{MHz}$. Employing appropriate filtering techniques at the output port can isolate these distinct frequency components of the modulated signal, for routing to different optical devices.

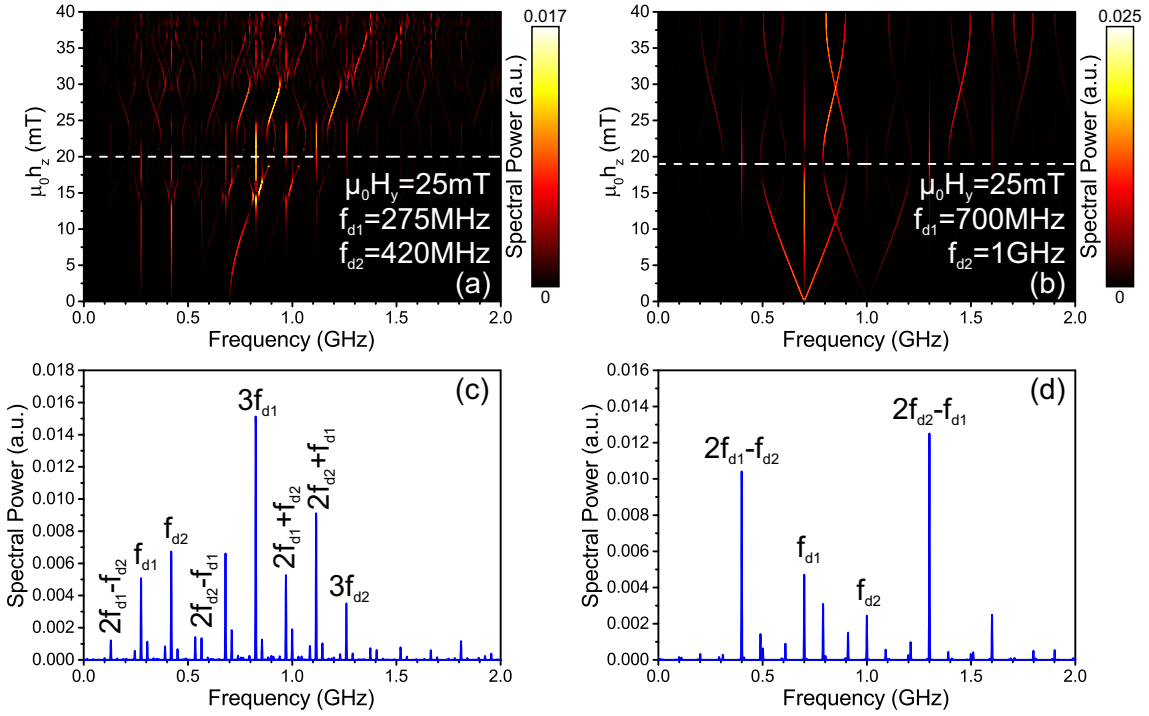

Figure 3.19: Plots of the frequency spectra of the modulated intensity signal at the MZI output versus $\mu_0 h_z$ for $\mu_0 H_y = 25$mT and (a) $f_{d1} = 275$MHz and $f_{d2} = 420$MHz, and (b) $f_{d1} = 700$MHz and $f_{d2} = 1$GHz. (c) depicts the spectral profile denoted by the dashed line in (a), for $\mu_0 h_z = 20$mT. Eight first- and third-order mixed frequencies are present. (d) shows the spectrum marked by the dashed line in (b), when $\mu_0 h_z = 19$mT, where the frequencies $2f_{d1} - f_{d2}$ and $2f_{d2} - f_{d1}$ are enhanced.

Adjusting the operating parameters can enhance or suppress some of these components and increase the power contained in a particular frequency of interest. Case 5, shown in Fig. 3.19(b), illustrates this trend. Here, the device is driven at $f_{d1} = 700$MHz and $f_{d2} = 1$GHz, with an amplitude of $\mu_0 h_z = 19$mT. The resulting spectra is illustrated in Fig. 3.19(d). Markedly, this configuration enhances the third-order mixed frequencies at $2f_{d1} - f_{d2} = 400$MHz and $2f_{d2} - f_{d1} = 1.3$GHz, as these spectral peaks are the dominant features of the spectrum. Power is transferred more efficiently to these two components than to the others.

3.6 Summary

In summary, this chapter has explored a number of unique magnetoplasmonic waveguide based devices exploiting the NRPS. First, a simple waveguide phase shifter was developed. It was demonstrated that passing coded current pulses through the underlying metal layer in the magnetoplasmonic system provides a natural method of generating transient magnetic fields, which, in combination with external static fields, provides a mechanism of actively modulating the magnetization vector of the Bi:YIG guiding layer. Correspondingly, the phase of the output wave could also be actively switched. The inherent shortcoming of such a device was the limited propagation length of the mode, and as such, it was demonstrated that a modification of the waveguide geometry could in fact produce useful phase shifts. These new LRDL-SPWs were incorporated into a fully integrable MZI isolator that demonstrated a low insertion loss of 2.51 dB and an extinction ratio of 22.82 dB between forward and backward propagating light.

This MZI architecture was then shown to provide the framework for two completely new nanoplasmonic modulators which have no contemporary analogue within the integrated optical regime. First was the optical equivalent of an electrical clock multiplier. By exploiting the precessional motion of Bi:YIG within the MZI, a device was developed that could upconvert an input signal repetition rate by factors up to 2.1×10^3, while providing 16.26dB modulation at frequencies ranging from 280MHz to 5.6GHz. The second unique plasmonic modulator exploited the highly nonlinear temporal dynamics of the magnetization and effectively mapped them onto the output intensity of the MZI. Driving the device with a single RF input generates harmonics, frequency splitting, and frequency down-conversion of the driving signals, while the application of two RF frequencies results in nonlinear mixing of the signals.

This extensively developed nanoscale magnetoplasmonic platform is robust, compact, tunable, and facilitates a new class of high-speed, electrically controlled optical

phase shifting devices and intensity modulators. Such devices are envisioned to satisfy crucial applications in the development of hybrid electrical-optical circuitry, and have fundamental applications in optical signal encoding, data transfer, and modulation in integrated optical networks.

Chapter 4

Integrated Faraday Effect-Based Plasmonic and Photonic Devices

4.1 Introduction

While the previous chapter focused on methods of incorporating the NRPS into plasmonic waveguide devices in order to manipulate the phase and intensity of a guided optical mode, there is another property of light that would be advantageous to control within an integrated platform: the polarization state. Polarization control is crucial in optical signal processing and integrated optical networks, specifically those applications employing polarization division multiplexing. Here, optical signals of orthogonal polarization states are passed through a waveguide or a fiber simultaneously, doubling the transmission bandwidth of a communications network [133]. In combination with an external polarizer, altering the polarization state is an effective means of modulating the intensity of an optical signal [134, 135], as well as creating optical isolators [136]. Accordingly, polarization controlling and manipulating devices have been a subject of great interest [137, 138]. As integrated optical technology progresses towards the plasmonic regime, polarization state management is increasingly

critical, due to the sensitivity of nanoplasmonic devices to an incident lightwave's polarization.

The investigations within this chapter examine methods by which polarization can be controlled and manipulated via MO effects within integrated platforms. The first section details the design and characterization of a magnetoplasmonic Faraday rotator, a device which has eluded realization due to the fundamental phase mismatch between photonic and plasmonic modes. In a similar manner to Chapter 3, the magnetization dynamics of the constituent MO garnets are shown to generate active polarization control and high-speed modulators. The following section builds on this component level understanding by taking a nanophotonic Faraday rotator and illustrating how it can be incorporated with Si photonic elements to form a more complex subsystem: an optical circulator. Employing the Faraday effect and its polarization manipulating properties within such a device allows the resulting construction to be polarization-independent and compact, while providing high isolation ratios over a broad bandwidth.

4.2 Magnetoplasmonic Faraday Rotators[1]

4.2.1 Overview

Current polarization manipulating devices utilizing plasmonics can typically be classified into two categories: planar metasurface structures and waveguide-based devices. The metasurface approach involves the precise design of plasmonic nanoantenna arrays or arrays of subwavelength apertures in a metallic film to control the amplitude and phase of the reflected or transmitted electric fields [139–142]. In general, these designs tend to be resonant devices, and are intended as free space components.

[1] A version of this section has been published as C. J. Firby, P. Chang, A. S. Helmy, and A. Y. Elezzabi, "Magnetoplasmonic Faraday rotators: enabling gigahertz active polarization control for integrated plasmonics," ACS Photonics, 3(12), 2344-2352 (2016).

Alternatively, waveguide based designs are much more amenable for integrated operation, and have been extensively examined for silicon photonics [143, 144]. There have been several approaches to this class of plasmonic mode converting device, involving either mode evolution or mode interference. In a plasmonic mode evolution device, the structure is designed to have a metallic feature that gradually changes in size and/or position along the length of the waveguide, such that the input mode couples to a plasmon on the metal and adiabatically evolves into the desired output polarization state. This includes metal tapers [145, 146], shifted metal caps [147], or metallic claddings that wrap around the edge of the waveguide [148, 149], often employed in conjunction with tapered dielectric cores. Plasmonic devices of this type have been used to convert between orthogonal photonic mode polarizations [145, 148, 149], or to convert between a TE photonic mode and a TM plasmonic mode [146, 147].

On the other hand, mode interference based devices are composed of an asymmetric structure that rotates the symmetry axis of the waveguide. A pure TE or TM input mode decomposes into two hybrid modes, which interfere along the length of the device. Examples of this type of device include dielectric cores covered by a metallic film on two adjacent sides [150, 151], or a thin metallic strip covering a fraction of the top of the guide [152, 153].

While both these classes of plasmonic devices can be very short (i.e. $< 10\mu$m), both approaches are inherently passive in nature, as the mode conversion is fixed for a particular device length. More importantly, even though the aforementioned devices can be designed for passive polarization manipulation, these structures are unable to achieve nonreciprocal operation, such as optical isolation or circulation. For chip-scale integrated plasmonics, a versatile polarization manipulation platform that can be dynamically configured for the desired applications, and exhibits nonreciprocal propagation behaviour is of great necessity. Such a platform must facilitate an active

and tunable mechanism of polarization/mode conversion for encoding signals and switching polarization states within an optical network. To meet these demanding performance requirements, structural rotation configurations cannot be utilized, and material-based mechanisms must be considered. A method of achieving this tunable and active functionality can be realized by invoking the Faraday effect.

Within the field of magnetoplasmonics [16, 154], demonstrations of the Faraday effect have been limited. Specifically, there are reports involving the enhancement of polarization rotation for light transmitted through Au nanowire gratings on a magnetic substrate [155, 156], Au nanoparticles embedded in a magnetic garnet matrix [157], or a perforated metallic film incorporating magnetic materials within either the subwavelength apertures [158] or the perforated film itself [159]. To date there has been no report of a waveguide based plasmonic Faraday rotator, due to the fact that the mode conversion efficiency is strongly dependent upon matching the phase velocities of the photonic TE and plasmonic TM modes of the structure. In typical plasmonic waveguides, the effective index difference between the TE and TM modes is significant, and correspondingly, the conversion efficiencies are limited. Additionally, optical modes in close proximity to metallic features are limited by their finite propagation length, while mode conversion due to the Faraday effect typically occurs on a length scale of several hundred microns to millimeters. In this respect, to realize this device, one requires a structure capable of supporting low-loss, long-range, phase-matched TE and TM modes. These stringent constraints have proven to be formidable to simultaneously overcome.

4.2.2 MO Waveguide Design and Passive Operation

In order to overcome this challenge, the waveguide structure for a magnetoplasmonic Faraday rotator must simultaneously facilitate long-range plasmonic and photonic modes, as well as a mechanism of achieving phase matching between them. To attain

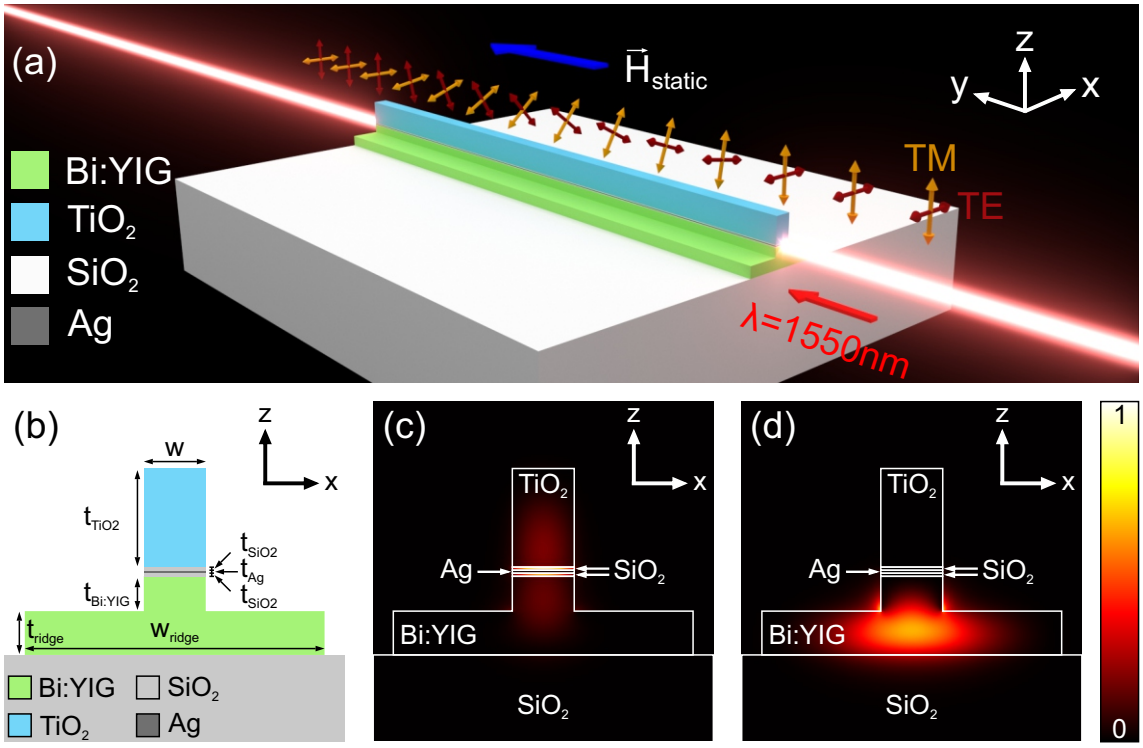

Figure 4.1: (a) Illustration of the magnetoplasmonic Faraday rotator under consideration. The device is capable of rotating the polarization of either a TE or TM input mode. (b) Schematic diagram of the hybrid ridge-plasmonic waveguide structure. (c) $|E_z|^2$ profile for the TM plasmonic mode supported by the structure. (d) $|E_x|^2$ profile for the TE photonic mode supported by the structure.

this, a unique hybrid ridge-plasmonic waveguide design was conceived (schematically depicted in Figs. 4.1(a)-(b)), which combines functionalities of both 5-layer hybrid plasmonic structures and standard ridge waveguides. Through judicial engineering of the hybrid plasmonic waveguide dimensions, one can effectively minimize the modal overlap with the metal and therefore reduce the losses for the TM plasmonic mode [45]. This design does not demand a vertically symmetric geometry or material profile, as the layer thicknesses can be tailored for minimum losses [45]. The hybrid ridge-plasmonic waveguide allows for the customization of a long-range TM plasmonic mode, as depicted in Fig. 4.1(c).

To form this mode within the present structure, the multilayer stack is chosen to be $w = 450$nm wide, the metal is chosen to be a $t_{Ag} = 25$nm thick film of Ag, and the two low-index layers consist of $t_{SiO_2} = 20$nm thick films of SiO_2. The refractive

indices (at $\lambda = 1550$nm) of these layers are taken to be $n_{Ag} = 0.145 + 11.438i$ [32] and $n_{SiO_2} = 1.44$ [33], respectively. The top high-index layer of the structure is chosen to be a $t_{TiO_2} = 725$nm thick layer of amorphous TiO_2, having a refractive index of $n_{TiO_2} = 2.3 + 1.1 \times 10^{-6}i$ at $\lambda = 1550$nm [51]. The high-index bottom film and the ridge sit atop a SiO_2 film, and are constructed from Bi:YIG, where $t_{ridge} = 320$nm and $t_{Bi:YIG} = 260$nm. It is within this Bi:YIG MO region that the TE and TM modes must exhibit spatial overlap for the Faraday rotation to occur.

The operation of this device requires a delicate balance between the modal losses and the modal overlap required for efficient Faraday rotation. To achieve long-range propagation from a TE mode in a structure with a lossy metal film, the overlap of the TE mode with the metal must be minimized, while maintaining overlap with the TM mode fields inside the MO material. This situation is possible with the inclusion of the Bi:YIG ridge at the base of the hybrid plasmonic guide, which shifts the TE mode away from the metal, as shown in Fig. 4.1(d). Increasing the width of the ridge effectively reduces the loss of the TE mode, as illustrated in Fig. 4.2(a). Furthermore, this ridge structure provides a means of tuning the modes to achieve phase matching without requiring complex periodic structures or magnetization reversal. Typically, altering a waveguide dimension impacts the effective index for both TE and TM modes ($n_{eff,TE}$ and $n_{eff,TM}$), and they cannot be tailored individually. However, here the TM plasmonic mode is confined near the Ag film, while the TE photonic mode is contained within the ridge. Since the TM mode fields do not significantly extend into the ridge, adjusting the ridge width, w_{ridge}, allows the effective index of the TE mode to be tuned almost independently from that of the TM mode, as shown in Fig. 4.2(b). In the present configuration, w_{ridge} is chosen to be 2.2μm, corresponding to the crossing point of the curves in Fig. 4.2(b).

The proposed magnetoplasmonic Faraday rotators can be fabricated in any well-equipped nanofabrication facility. The high quality magnetic garnet films can be

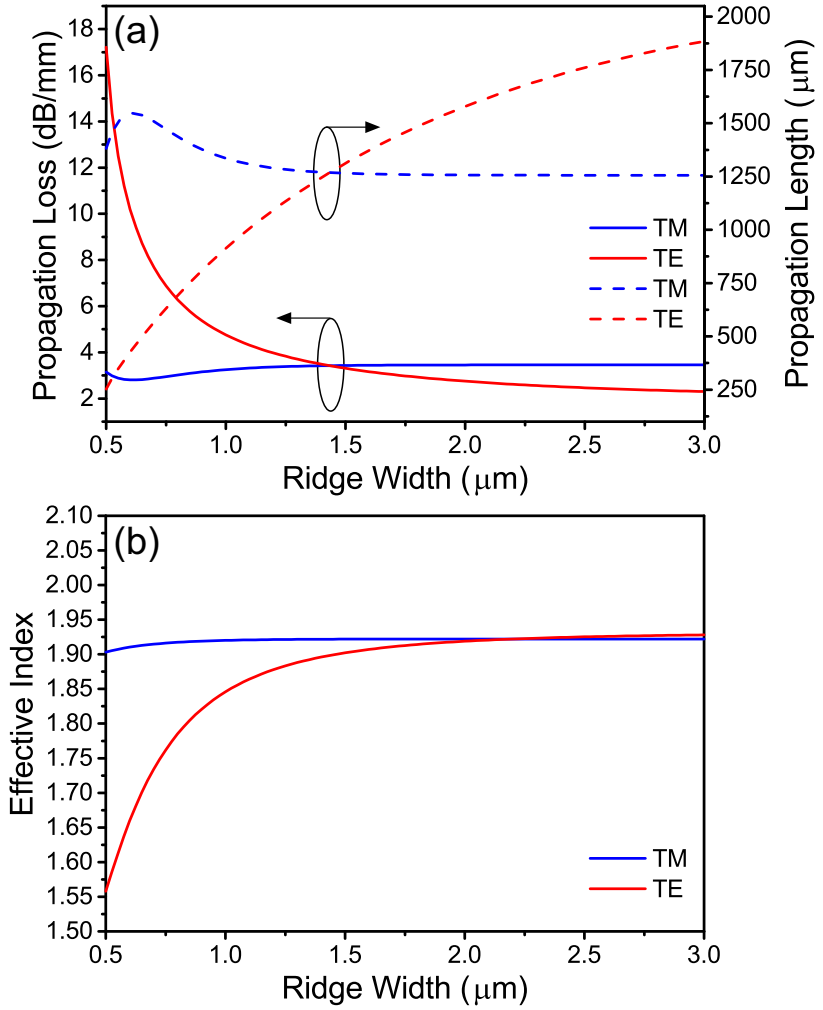

Figure 4.2: (a) Propagation loss and corresponding propagation length for the TE and TM modes as a function of the ridge width, w_{ridge}. (b) Effective index of the TE and TM modes as a function of the ridge width, w_{ridge}. Note the phase-matching condition is satisfied at $w_{ridge} = 2.2\mu$m. In both (a) and (b), $w = 450$nm, $t_{ridge} = 320$nm, $t_{Bi:YIG} = 260$nm, $t_{SiO_2} = 20$nm, $t_{Ag} = 25$nm, and $t_{TiO_2} = 725$nm.

deposited onto SiO_2 via PLD [92], while the thin SiO_2 films and the Ag film can be deposited with plasma enhanced chemical vapour deposition (PECVD) and sputtering, respectively. Amorphous TiO_2 can be reactively sputtered as the top layer, and electron beam lithography along with FIB milling can accurately shape the structure. It is common for devices relying on phase-matched modes to possess strict fabrication tolerances, but this is alleviated in the hybrid ridge-plasmonic design due to the independent tuning of the TE mode effective index. One can fabricate the structure having an exaggerated w_{ridge}, characterize the resulting $n_{eff,TM}$ accounting for fabrication defects in the multilayer stack, and selectively mill the ridge via FIB to achieve precise phase-matching.

Theoretical treatment describing Faraday rotation within waveguides [69] predict that mode separation degrades the power conversion efficiency and increases the conversion length. This drop in conversion efficiency can be remedied with sufficient phase matching. For the TE and TM modes of this waveguide, the effective index difference is $\Delta n_{eff} = |n_{eff,TM} - n_{eff,TE}| = 4 \times 10^{-5}$, and the conversion efficiency and conversion length (L_{conv}) are predicted to be 99.4% and $L_{conv} = 819\mu$m, respectively. Correspondingly, in this configuration, the propagation lengths of the two modes are $L_{prop,TM} = 1257\mu$m and $L_{prop,TE} = 1660\mu$m, as shown in Fig. 4.2(a), leading to relatively low insertion loss (2.5dB) compared to other plasmonic structures [145, 151, 153]. This value can be further optimized by either reducing the thickness of the metal layer, or restructuring the hybrid plasmonic section such that the phase matching condition corresponds to a wider ridge width. Clearly, this novel structure allows the simultaneous implementation of long-range phase-matched eigenmodes, which are ideal for demonstrating the Faraday effect.

To investigate the Faraday rotation of the device, fully-vectorial FDTD simulations are employed. Here, the desired input mode is directly injected into the waveguide, while the magnetization of the Bi:YIG is accounted with a full gyrotropic

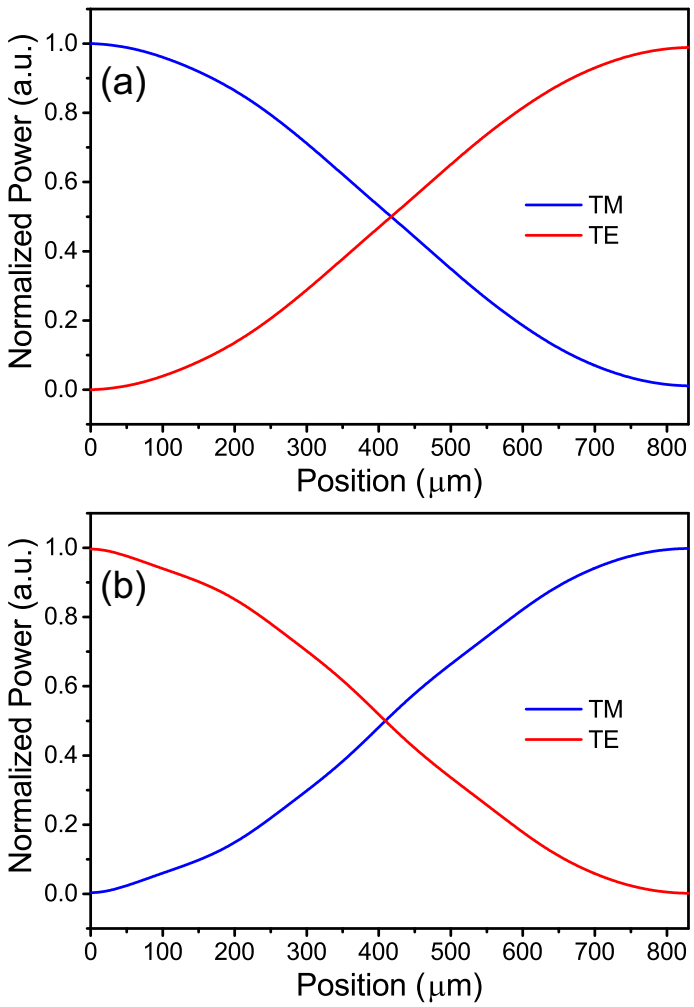

Figure 4.3: Normalized power in each of the TE and TM components along the length of the device for (a) a TM input or (b) a TE input into the device depicted in Fig. 4.1.

permittivity tensor as per the procedure described in Chapter 3. In the present geometry, Faraday rotation is observed when the $\mathbf{M}$ is parallel to the propagation direction (i.e. the y-axis). $\mathbf{M} = \langle 0, +M_S, 0 \rangle$ is induced by a parallel static magnetic field vector ($\mathbf{H}_{static} = \langle 0, +H_y, 0 \rangle$) as shown in Fig. 4.1(a). The local normalized power in each of the $\lambda = 1550$nm modes as a function of device length is plotted in Fig. 4.3(a) for a TM input, and in Fig. 4.3(b) for a TE input. The power contained in the TM and TE modes, P_{TM} and P_{TE} respectively, are the integrated TM and TE contributions to the y-component of the Poynting vector, $S_{Poynting,y} = 0.5(E_z H_x^* - E_x H_z^*)$, over the

cross sectional area, A:

$$\begin{aligned} P_{TM} &= \frac{1}{2}\,\mathrm{Re}\left(\iint E_z H_x^* dA\right) \\ P_{TE} &= -\frac{1}{2}\,\mathrm{Re}\left(\iint E_x H_z^* dA\right) \end{aligned} \tag{4.1}$$

The local normalized power in each cross section along the device length is given by:

$$\begin{aligned} P_{TM,norm} &= \frac{P_{TM}}{P_{TM} + P_{TE}} \\ P_{TE,norm} &= \frac{P_{TE}}{P_{TM} + P_{TE}} \end{aligned} \tag{4.2}$$

As evident from Fig. 4.3, power conversion efficiency of 99.4% is attained within a conversion length of $L_{conv} = 830\mu$m when either a TE or TM mode is launched into the device. These values agree with the theoretical predictions within 0.4% and 1.3%, respectively. The unique hybridization of the multilayered plasmonic structure and conventional ridge waveguides facilitates sufficiently low losses, wherein L_{conv} is significantly less than the propagation lengths of the TE or TM modes, and high efficiency conversion can be attained within the plasmonic device. Cross-sectional profiles of $|E|^2$ within the waveguide at five locations along the device are depicted in Fig. 4.4 (i.e., $y = 0\mu$m, $y = 270\mu$m, $y = 410\mu$m, $y = 554\mu$m, and $y = 830\mu$m). For a plasmonic TM mode input (Fig. 4.4(a)-(e)), one can clearly observe the $|E|^2$ profile evolving from the dominantly E_z distribution around the metal film (see Fig. 4.1(c)) into the E_x distribution contained within the ridge (see Fig. 4.1(d)) at the device output ($y = L_{conv} = 830\mu$m). Conversely, for a photonic TE mode input (Fig. 4.4(f)-(j)), the dominant E_x field (see Fig. 4.1(d)) evolves into a symmetric E_z field distribution (see Fig. 4.1(c)) around the Ag film at the device output ($y = L_{conv} = 830\mu$m).

This magnetoplasmonic device is ideal for polarization diversity circuits, as it is versatile enough to handle either input polarization state (i.e. TE or TM), and can

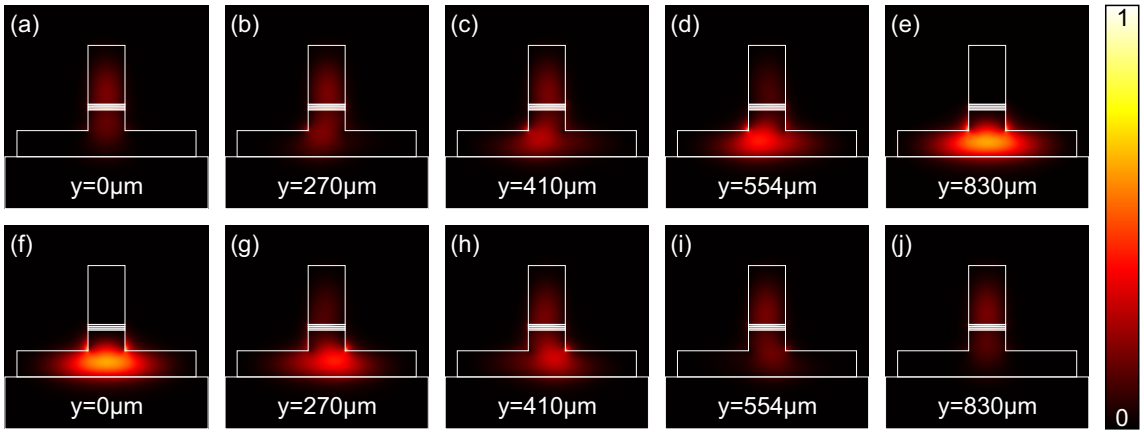

Figure 4.4: $|E|^2$ profiles within the device for (a)-(e) a TM input and (f)-(j) a TE input. (a) and (f) depict the input modes at $y = 0\mu$m, (b) and (g) depict the fields at $y = 270\mu$m, where 25% conversion has occurred, (c) and (h) depict the fields at $y = 410\mu$m, where 50% conversion has occurred, (d) and (i) depict the fields at $y = 554\mu$m, where 75% conversion has taken place, and (e) and (j) depict the output field profiles, at $y = L_{conv} = 830\mu$m. The color scale of each frame is normalized between 0 and 1 to clearly depict the mode evolution.

convert between tightly confined modes ideal for integrated optics applications. With future developments in the fabrication of Bi:YIG films, one will be able to employ films exhibiting higher θ_{spec}, and hence attain complete mode conversion within a much shorter device length.

4.2.3 Active Polarization Control

A powerful and unique feature arising from the implementation of a magnetoplasmonic Faraday rotator is the ability to actively control the fraction of power in each of the TE and TM components at the device output. This is facilitated since the Faraday rotation is induced by the magnetization of the Bi:YIG, and **M** itself can be dynamically modified by transient magnetic field pulses in a similar manner to the demonstrations of Chapter 3. When current signals are made to propagate through a metal feature in close proximity to the Bi:YIG, the generated magnetic fields can be made to actively modulate the polarization state of the lightwave.

In the present device, active modulation is achieved through an Ag transmission line (TL), buried $d = 1.5\mu$m below the surface of the SiO_2. This structure has dimen-

sions of $w_{TL} = 8\mu$m and $t_{TL} = 1.5\mu$m, runs along the length of the device, and is set in contact with a Si substrate for heat dissipation. In this orientation, super-Gaussian current pulses with a peak of 1A will generate transient fields, $\mathbf{h}(t) = \langle h_x(t), 0, 0\rangle$, over the Bi:YIG, where the peak magnetic field is about $\mu_0 h_{x,pk} = 51$mT. This in turn modulates $\mathbf{M}$ in accordance with the LLG model of magnetization dynamics (Eqn. 2.86).

The magnetoplasmonic Faraday rotator can be operated in two different dynamic modes, depending on the orientation of the static biasing magnetic field, $\mathbf{H}_{static}$. If $\mathbf{H}_{static}$ is aligned with the y-axis, $\mathbf{H}_{static} = \langle 0, +H_y, 0\rangle$, the device can be operated in a pulsed-input pulsed-output (PIPO) mode, whereby the appropriate choice of the magnetic field parameters can map individual magnetic field pulses onto the magnetization, and onto the polarization state. Conversely, if $\mathbf{H}_{static}$ is aligned with the z-axis, $\mathbf{H}_{static} = \langle 0, 0, +H_z\rangle$, the device can be operated in a pulsed-input continuous-output (PICO) mode, where a single magnetic field pulse excites an magnetization oscillation within the Bi:YIG, and hence, maps the decaying oscillation onto the output polarization state.

Pulsed-Input Pulsed-Output Operation

A schematic illustration of the geometry required for PIPO operation is shown in Fig. 4.5. A static magnetic field, $\mathbf{H}_{static} = \langle 0, +H_y, 0\rangle$, is externally applied to the device to initially set $\mathbf{M}$ to an initial state, $\mathbf{M}_0 = \langle 0, +M_S, 0\rangle$. When the transient magnetic field, $\mathbf{h}(t) = \langle h_x(t), 0, 0\rangle$, is applied, the magnetization and the magnetic field vectors are not aligned, and hence $\mathbf{M}$ begins to precess around $\mathbf{H}_{static} + \mathbf{h}(t)$. Following this excitation, when $\mathbf{h}(t)$ is turned off, $\mathbf{M}$ will precess around $\mathbf{H}_{static}$ alone. In the present device, this implies that $\mathbf{M}$ is deflected away from the y-axis by $\mathbf{h}(t)$, and then begins to precess around it as it slowly relaxes.

It is the M_y component that induces the Faraday effect, and hence is mapped onto

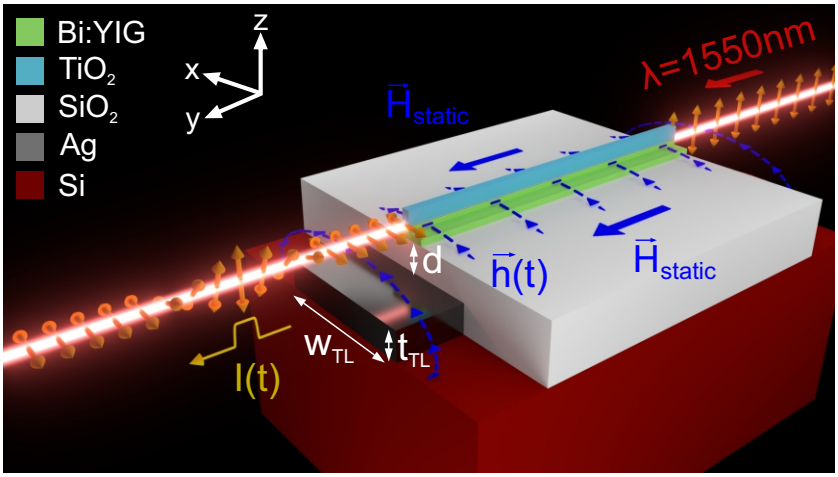

Figure 4.5: Schematic illustration of the geometry required for the pulsed-input pulsed-output (PIPO) dynamic operation mode.

the polarization of the lightwave at the device output. To induce one-to-one pulse mapping between the driving electrical pulse and the optical polarization state, the oscillatory precession amplitude must be minimized, or correspondingly, H_y, $h_{x,pk}$, and the pulse width τ_p must be judicially chosen to return $\mathbf{M}$ to its initial state at the conclusion of $\mathbf{h}(t)$. With these conditions satisfied, the waveguide can switch the polarization rotation from "on"-to-"off", and back again. Note that this mode of operation is analogous to that of the magnetoplasmonic phase shifter presented in Section 3.2, except the geometry allows for the manipulation of the polarization instead of the phase of the lightwave.

As an example, this functionality can be achieved by employing $\tau_p = 500$ps pulses having a peak magnetic field of $\mu_0 h_{x,pk} = 51$mT, and a static biasing field of $\mu_0 H_y = 51$mT. The local normalized power at the device output is shown in Fig. 4.6(a) for a TM input to the device. Since the device acts as an "on"-to-"off" switch, the TM input switches the output polarization state from TE to TM and back. These parameters achieve polarization modulation depth of $> 99\%$ in the output state.

This device can respond to a train of the above specified electrical input pulses to encode binary data onto the polarization state of the guided optical mode. Figure

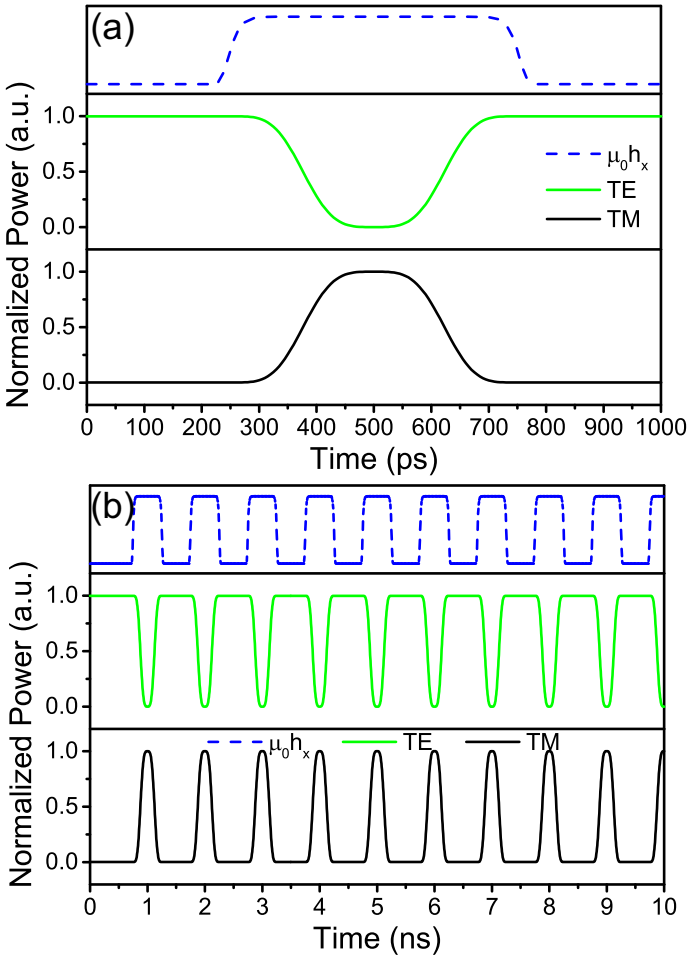

Figure 4.6: (a)-(b) Depict the operation of the device in PIPO mode, with a TM input. (a) illustrates the polarization switching due to a single input pulse with $\tau_p = 500$ps, $\mu_0 h_{x,pk} = 51$mT, and $\mu_0 H_y = 51$mT. The device acts as an "on"-to-"off" switch. (b) illustrates the polarization response to a train of these input pulses at a repetition rate of $f_{rep} = 1$GHz. The $\mu_0 h_x$ curves are not to scale, and are only intended to indicate the timing.

4.6(b) depicts the output of the device under excitation from a train of electrical pulses at a repetition rate of $f_{rep} = 1$GHz. Analogous operation is achieved for a TE input to the device as well. Such polarization modulation can easily be converted to optical intensity modulation by passing the output light signal through a polarizer. Clearly, the polarization rotation can be switched in a regular manner with the above parameters, and as such, provides an efficient means of mapping an electrical excitation onto the polarization state.

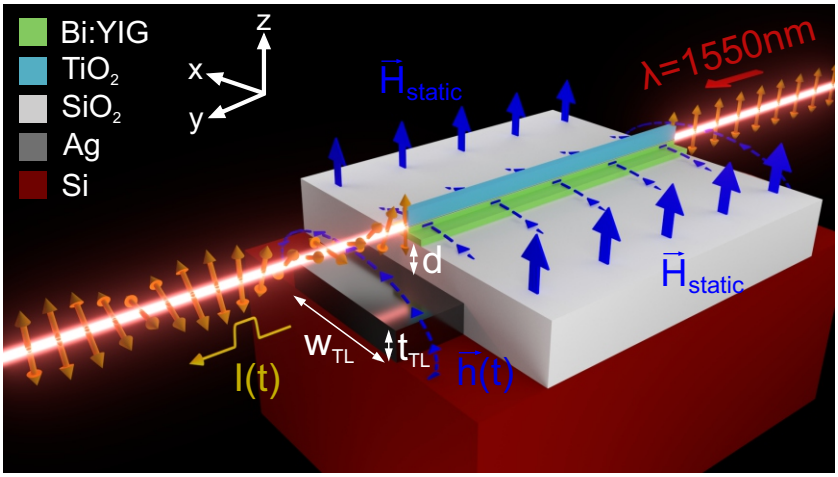

Figure 4.7: Schematic illustration of the geometry required for the pulsed-input continuous-output (PICO) dynamic operation mode.

Pulsed-Input Continuous-Output Operation

While the PIPO mode required suppression of the precessional oscillations within the Bi:YIG, changing the orientation of $\mathbf{H}_{static}$ allows for PICO operation. This mode of operation is analogous to that of the clock multiplier presented in Section 3.4, except that the present device allows for the excitation of high-speed oscillations in the output polarization state. In this configuration, shown in Fig. 4.7, the static biasing magnetic field is oriented along the z-axis, $\mathbf{H}_{static} = \langle 0, 0, +H_z \rangle$, to initially set $\mathbf{M}$ in the same direction. Applying the transient pulse $\mathbf{h}(t) = \langle h_x(t), 0, 0 \rangle$ deflects $\mathbf{M}$ away from the z-axis, around which it begins to precess at the Larmor frequency, ν_L. The Faraday rotation within the device exhibits similar oscillatory behaviour to that of M_y. As $\mathbf{M}$ revolves around the z-axis, the sinusoidal variations in M_y (and by extension θ_F) occur at ν_L. However, since the fraction of power contained in the TE and TM modes at the device output do not depend on the sign of θ_F, the power carried in each of the TE and TM modes oscillate at a frequency of $2\nu_L$. In this way, a single magnetic field pulse can excite long-lasting high frequency modulation of the output polarization state.

In order to maximize the polarization modulation depth, one requires the magnetic

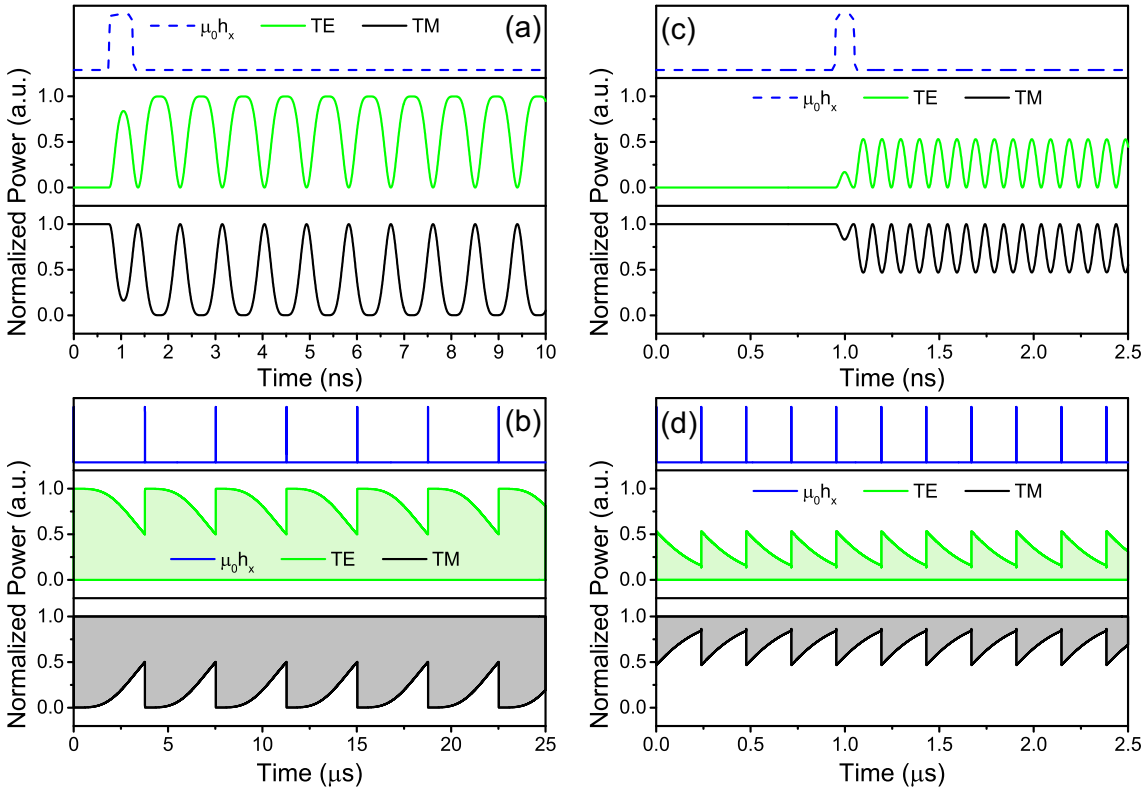

Figure 4.8: (a)-(d) Depict the operation of the device in PICO mode with a TM input. (a) depicts the $2\nu_L = 1.12$GHz oscillating polarization response induced by a single magnetic field pulse with $\tau_p = 500$ps, $\mu_0 h_{x,pk} = 22$mT, and $\mu_0 H_z = 20$mT. The amplitude of this oscillation decays on a microsecond timescale. (b) depicts the decaying envelope of the polarization response due to the aforementioned magnetic field pulses repeated at $f_{rep} = 267$kHz. (c) depicts the $2\nu_L = 10.08$GHz oscillating polarization response induced by a single magnetic field pulse with $\tau_p = 100$ps, $\mu_0 h_{x,pk} = 51$mT, and $\mu_0 H_z = 180$mT. The amplitude of this oscillation decays on a timescale of a few hundred nanoseconds. (d) depicts the decaying envelope of the polarization response due to the aforementioned magnetic field pulses repeated at $f_{rep} = 4.2$MHz. Note that within the shaded region, the polarization oscillations occur on a timescale of nanoseconds in (b) or a few hundred picoseconds in (d), and cannot be observed here. In (a)-(d), the $\mu_0 h_x$ curves are not to scale, and are only intended to indicate the timing.

field pulse to excite the highest amplitude precession about the z-axis possible. Since $|\mathbf{M}| = M_S$ is constant, this can be achieved if the transient pulse excites $\mathbf{M}$ to a state such that $M_z/M_S = 0$. By appropriately selecting of the operating biasing field (H_z), peak transient field ($h_{x,pk}$), and pulse width (τ_p), one can realize this condition. One such example is depicted in Fig. 4.8(a), which depicts the temporal variations in the local normalized power at the device output for $\tau_p = 500$ps, $\mu_0 h_{x,pk} = 22$mT, and $\mu_0 H_z = 20$mT. At this biasing magnetic field, $\nu_L = 560$MHz, but the power contained within the TE and TM components oscillates at $2\nu_L = 1.12$GHz. Such an oscillatory

behaviour decays on a time scale of microseconds. The repetition rate, f_{rep} of the input electrical pulses can be very low to drive this continuous oscillation. Adopting the re-excitation threshold defined in Section 3.4.3, whereby the device is stimulated by a new pulse when the precession amplitude decays to 50% of its initial value, this device can be driven at the low repetition rate of $f_{rep} = 267$kHz (Fig. 4.8(b)). The frequency of the output polarization state oscillations is about 4.2×10^3 times greater than the input repetition rate.

Higher modulation frequencies can be achieved, as the Bi:YIG material has demonstrated responses of 82GHz [118]. However, in order to achieve higher output frequencies in the PICO mode, one must utilize higher static fields (to increase ν_L), and correspondingly, shorter pulse-widths. As a consequence, one must sacrifice modulation depth to avoid unwieldy increases in the current requirements. For example, the output of the device is plotted as a function of time for a device driven with $\tau_p = 100$ps pulses with an amplitude of $\mu_0 h_{x,pk} = 51$mT, under a static bias of $\mu_0 H_z = 180$mT, in Fig. 4.8(c). In this scenario, $\nu_L = 5.04$GHz, and correspondingly, the polarization state oscillates at $2\nu_L = 10.08$GHz, but the modulation depth is reduced to 53%. These rapid oscillations decay much quicker, so the repetition rate required by the aforementioned threshold is $f_{rep} = 4.2$MHz (Fig. 4.8(d)). Correspondingly, the frequency of the oscillating polarization at the output is 2.4×10^3 times greater than the input modulation frequency.

Armed with the findings from this fundamental study illustrating the design methodology, as well as the utility of MO components for manipulating the polarization of a guided lightwave, the next logical step is to incorporate a polarization altering device into a more advanced device architecture. It is essential to examine how these MO elements can be integrated with more typical passive integrated optical components to create a complex subsystem. One such example is an optical circulator, which incorporates nonreciprocal MO elements and passive elements to

manipulate the direction of light flow.

4.3 Versatile Polarization-Independent Nanophotonic Circulators[2]

4.3.1 Overview

Circulators perform the critical function of routing electromagnetic waves sequentially through the various ports on the device, as well as providing isolation from potentially damaging reflected radiation. In these devices, a lightwave entering the device in Port 1 is routed to Port 2, but a lightwave entering the device at Port 2 is directed to Port 3, and not back to Port 1. The operation is schematically depicted in Fig. 4.9 for a generalized circulator with an arbitrary number of ports. This is inherently a nonreciprocal function, and therefore requires MO components.

Due to this unique functionality, circulators have found applications in a many facets of optics and photonics, ranging from integration into laser cavities to increase the pumping efficiency [160], to quantum optics, where the circulation of light is controlled by the quantum state of a single atom [161]. More commonly, circulators are often employed within fiber-optic networks employing wavelength-division-multiplexing (WDM), as they can be utilized to multiplex transmitted optical signals [162]. Circulators enable bi-directional or full duplex operation in a single fiber [163] by separating counter-propagating signals at a transceiver, allowing simultaneous transmission and reception, and increasing the network efficiency. These components are a vital building block for optical networks.

Historically, there have been a vast array of circulator demonstrations presented.

[2]A version of this section has been published as C. J. Firby, P. Chang, A. S. Helmy, and A. Y. Elezzabi, "Versatile broadband polarization-independent optical circulators for nanophotonic integrated circuits," J. Opt. Soc. Am. B 35(7), 1504-1513 (2018).

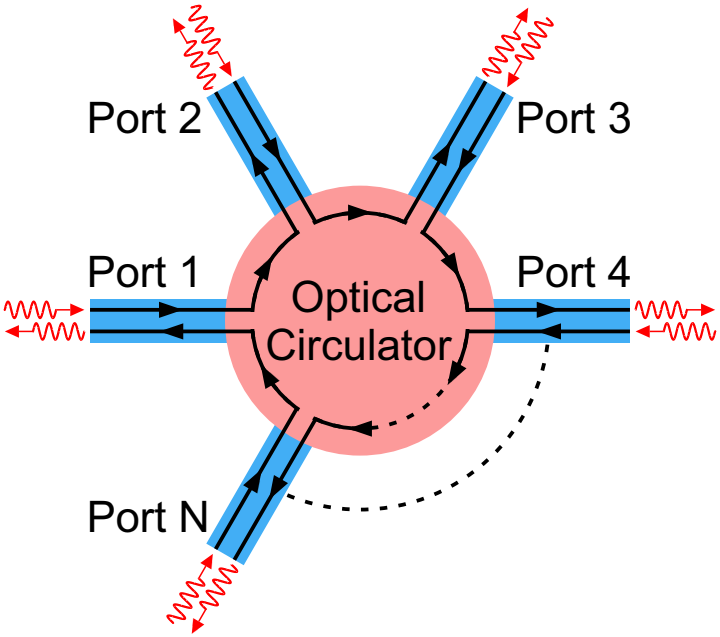

Figure 4.9: Schematic depiction of the operation of a generalized optical circulator.

In the microwave regime, circulators have been realized through the inclusion of ferrite posts [164], disks [165], or spheres [166] at waveguide junction, or through the inclusion of a 45° Faraday rotator within a waveguide with four ports [167]. The early free space optical circulator designs extended the latter concept by incorporating a 45° Faraday rotator between two polarizing prisms [168, 169]. While these free space implementations were promising proof of concept, they were impractical for telecommunications and optical networking due to their large physical size, and the transition to fiber coupled schemes was implemented [170, 171]. These bulk components are still incompatible with the microscale device dimensions demanded by modern nanophotonic chips, and are not suitable for future integrated applications.

For integration into chip-scale devices, circulator designs are required to replace fibers and bulk components with thin film waveguides. One implementation utilized a combination of a silica-based 3dB coupler, to split the power input to a port between two arms where each arm included both a nonreciprocal Faraday rotator in the form of a magnetic rib waveguide, and a polyimide film half-wave plate (HWP) for a

reciprocal rotator [172]. This configuration was shown to be insensitive to the input polarization state and is capable of providing up to 23.7dB isolation and as low as 3dB insertion loss. However, the device was several centimeters in length, the waveguide cross-sectional dimensions were on the order of several microns, and the polyimide HWP films had to be manually inserted into vertically cut slots on the substrate, rendering the device impractical for on-chip integration [172]. An alternative scheme employs the NRPS instead of the optical Faraday effect. Circulators based on the NRPS are limited to operating with only one polarization state, and are inherently polarization dependent. In Ref. [173], a MZI is constructed from magnetic garnet rib waveguides. Passing current through electrodes deposited over a segment of each arm creates a nonreciprocal phase shifter, and in combination with a reciprocal phase shifter introduced into one arm, circulating action was observed with a low isolation ratio of 3dB. However, to achieve this isolation ratio, a total device length of 8.5mm was required.

Nonetheless, this concept was further exploited, both theoretically [174] and experimentally [175], to hybridize the technology with silicon photonics. Here, a magnetic garnet slab was directly bonded onto a similar SOI MZI, where the silicon waveguides where characterized by sub-micron cross-sectional dimensions. While this was the first circulator presented on a silicon platform, it was based on the bonding of a bulk Ce:YIG crystal to the surface in order to experimentally obtain an isolation of 15.3dB [175], and required bi-directional magnetic fields to function. By further modifications to the topology and by using a benzocyclobutene bonding agent, an MZI-type integrated circulator requiring only a unidirectional magnetic field, and yielding an isolation ratio of 22dB was demonstrated [176].

Recently, more exotic circulator designs have emerged. For example, the use of magnetic garnets bonded to SOI ring resonators exhibited circulating action and 11-14.4dB isolation [177, 178]. However, despite the small, 20μm radius of the ring,

exploiting the NRPS within the ring inherently limits the device operation to one polarization state, and the use of a ring resonator limits the device operation to a sub-nanometer bandwidth. Another configuration, composed of a 2D photonic crystal constructed from a magnetic garnet, exhibited similar characteristics, namely, limited polarization response and sub-nanometer bandwidths [179].

Finally, it was proposed that including metal nanoparticles within a magnetic garnet cavity leads to nonreciprocal symmetry breaking, and thus, a magnetic field dependent circulation of power around the embedded nanoparticles [180]. This concept was theoretically realized in a nanoscale three-port circulator having Au nanorods embedded in a triangular magnetic garnet cavity at the junction of three waveguides [181]. Despite the nanoscale dimensions of the device, 63% of the power injected into Port 1 is routed to Port 2, while 3% of the injected power is routed to Port 3 [181]. A separate study examined the effect of inserting magnetic garnet rods into a Si photonic crystal cavity at the junction between three Si waveguides [182]. This configuration yielded a high isolation ratio of 16dB, but large insertion losses of 9dB.

While each of the above arrangements exhibit a specific advantage, none of the discussed designs possess all of the criteria required by an ideal optical circulator for implementation into chip-scale nanophotonic networks. An optimal device would exhibit a number of key characteristics: (1) be miniature in size for dense on-chip integration, (2) provide a large extinction ratio between the device ports, (3) offer efficient circulation over a broad bandwidth, and (4) be able to accept either TE or TM polarization states. To date, while many implementations have been presented, none were able to satisfy all of the criteria for a truly versatile device. Designing a device capable of satisfying these stringent requirements is the focus of the next sections.

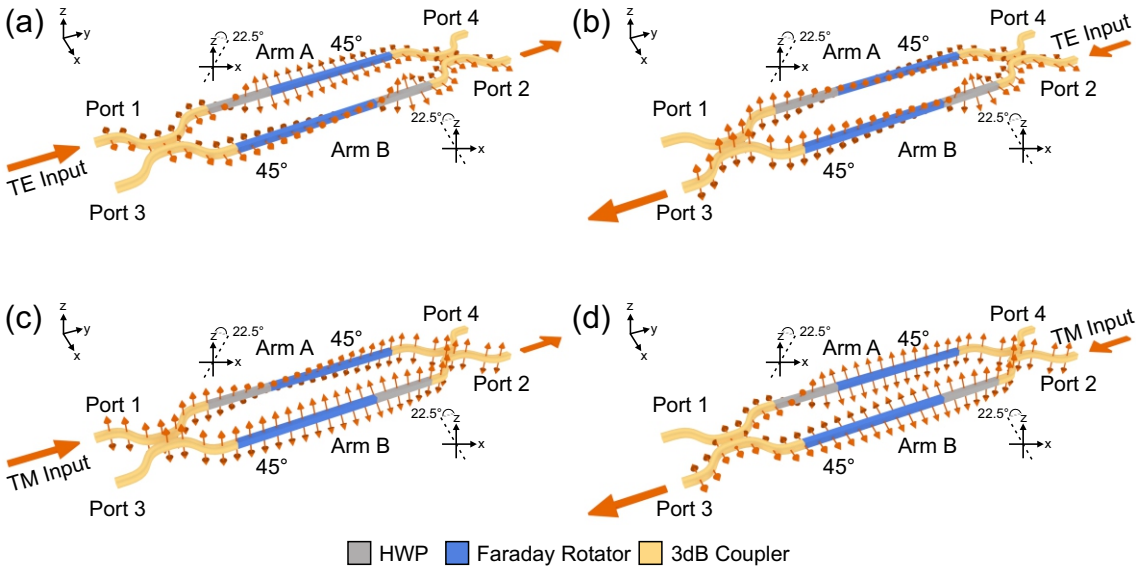

Figure 4.10: Schematic illustration of the principle of operation of a Faraday effect based optical circulator for (a) TE mode input to Port 1, (b) TE mode input to Port 2, (c) TM mode input to Port 1, and (d) TM mode input to Port 2. Input light is split between the two MZI arms via a 3dB power splitter, before passing through a HWP and Faraday rotator in each arm. The HWPs, with their optical axes set at $\pm 22.5°$ from the z-axis, provide reciprocal polarization rotation, while the Faraday rotators provide nonreciprocal polarization rotation. The signals in each arm are appropriately recombined by a second 3dB power combiner at the output. Note that the orange arrows along the structure indicate the orientation and evolution of the polarization state as the lightwave propagates through the device.

4.3.2 Circulator Design

Operation Principle

The proposed optical circulator design consists of three key components: a 3dB power splitter/combiner comprising the input and output ports, two HWPs for reciprocal polarization rotation, and two Faraday rotators for nonreciprocal polarization rotation. The general principle of operation for this device is based on manipulating the polarization state within the two arms of a MZI [172, 183], and is schematically depicted in Fig. 4.10(a)-(d). Consider first a TE mode input to the device at Port 1, as shown in Fig. 4.10(a). The 3dB power splitter is used to equally split the input optical power between the two interferometer arms (denoted arm A and arm B in Fig. 4.10). At the output of the 3dB power splitter, each interferometer arm contains 50% of the optical power with TE polarization, but due to the nature of the 3dB splitter,

the coupled arm has a 90° phase lag relative to the through arm. In arm A, light first passes through a HWP with its axis set at 22.5° clockwise from the z-axis, and then a 45° Faraday rotator, while in arm B, light passes through an identical 45° Faraday rotator first, followed by a HWP with its axis at 22.5° counter clockwise from the z-axis. As depicted in Fig. 4.10(a), after passing through these two components, each arm contains TE polarization, and no additional phase is introduced between the two arms. The 90° phase shift between the two arms is preserved. Upon entering the 3dB power combiner, the signals from the two arms are recombined into the output waveguide corresponding to Port 2, and routing from Port 1 to Port 2 is achieved.

Next, the device operation for a TE mode injected at Port 2 is depicted in Fig. 4.10(b). The input light is again split 50/50 between arms A and B, before entering the Faraday rotators and waveplates. After propagating through the Faraday rotator and waveplate in each arm, the output polarization is rotated to TM, and an additional 180° phase shift is introduced between the two arms. This phase shift, in combination with the 90° shift introduced by the input coupler, effectively switches which arm is leading and which is lagging in phase, and at the output 3dB power combiner, the lightwaves from the two arms are recombined into the output waveguide corresponding to Port 3 instead of Port 1.

Considering the above arguments, it is evident that a lightwave input into Port 3 of the device will be routed to Port 4, and correspondingly a lightwave directed in through Port 4 will emerge at Port 1. Evidently, in this configuration effective power circulation is observed, where input to any port will be routed to the next sequentially numbered port. Analogous functionality is observed for the device under excitation from a TM mode, as depicted in Fig. 4.10(c) for input to Port 1, and in Fig. 4.10(d) for input to Port 2. Regardless of the input polarization state, the output of the device at Ports 2 and 4 will have the same polarization state as the input, while output at Ports 1 and 3 will be characterized by the orthogonal polarization state.

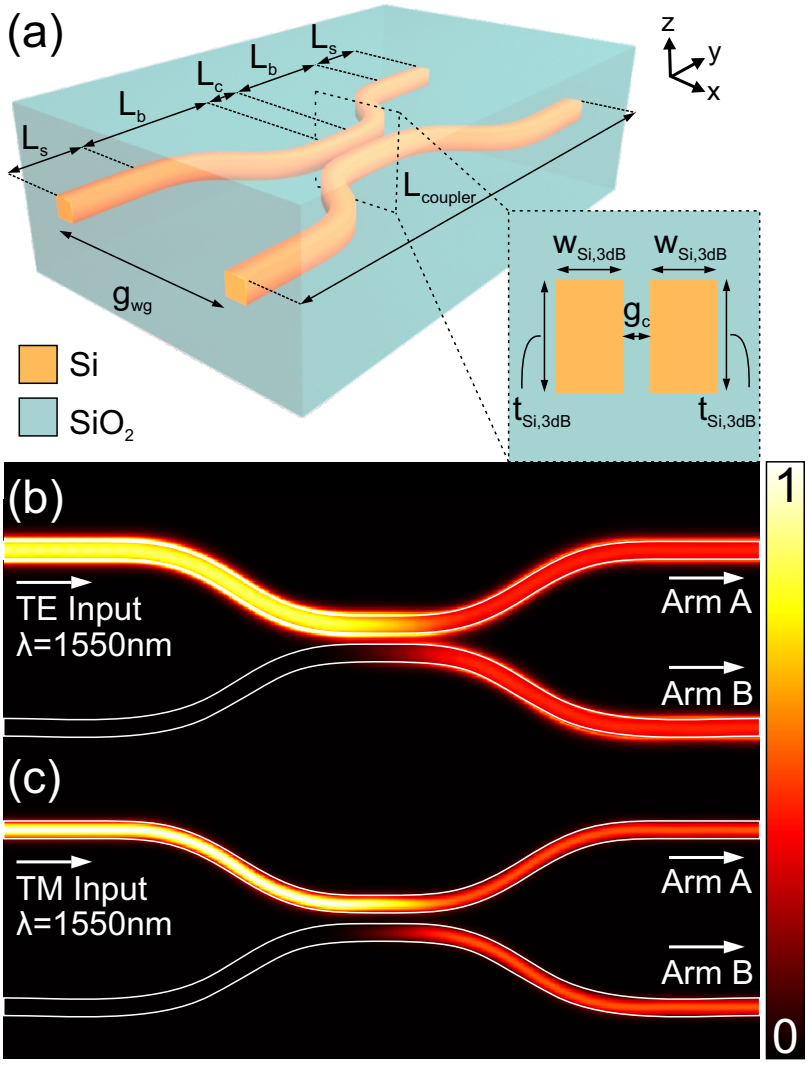

Figure 4.11: (a) Silicon photonic 3dB power splitter/combiner geometry considered here. (b) $|E|^2$ field profile at the center of the device for a TE mode input with $\lambda = 1550$nm. (c) $|E|^2$ profile at the center of the device for a TM mode input with $\lambda = 1550$nm. In both (b) and (c), power is split 50/50 between arms A and B.

3dB Power Splitter

To realize the 3dB power splitter required by a versatile optical circulator, Si photonic waveguides are considered on an SiO_2 substrate, and covered in a SiO_2 cladding. All calculations are undertaken considering an operating wavelength of $\lambda = 1550$nm, and the refractive indices for Si and SiO_2 are taken to be $n_{Si} = 3.48$ and $n_{SiO_2} = 1.44$, respectively [33]. The critical design specifications of this component include 50/50 power splitting for both TE and TM input modes, so that the overall circulator will

be polarization-independent, and a small on-chip footprint.

The optimized component is schematically depicted in Fig. 4.11(a), and is comprised of nanophotonic silicon waveguides having a width of $w_{Si,3dB} = 305$nm and a height $t_{Si,3dB} = 370$nm arranged into three sections: a straight input/output waveguide of length L_s, a straight coupling waveguide of length L_c, and S-bend sections of length L_b connecting the two. To minimize the footprint of the coupler, the straight input/output sections are chosen to be $L_s = 2\mu$m long, and are separated by a center-to-center distance of $g_{wg} = 3\mu$m, while the S-bend sections adiabatically narrow the gap between the waveguides to $g_c = 175$nm over a distance of $L_b = 5\mu$m. Often in the design of directional couplers, parasitic coupling between the S-bend waveguides is ignored, and only coupling in the straight section of length L_c is considered. In a design this compact, and to achieve polarization independent operation, this parasitic coupling is important since it differs for TE and TM input modes due to the differing mode profiles. Because of this, the straight coupling length will deviate from the ideal design length of $L_c = \pi/(2|\beta_a - \beta_s|)$, where β_a and β_s are the propagation constants of the antisymmetric and symmetric eigenmodes in the coupling section, due to the additional coupling in the S-bend waveguides. In fact, this is a key variable for tailoring the polarization independent operation of the splitter/combiner.

Accounting for this, at $L_c = 1.125\mu$m, 50/50 power splitting between the two arms is observed for both polarization states simultaneously. Figures 4.11(b) and 4.11(c) depict a cross section of the fields in the device at this optimum length for TE polarization and TM polarization input respectively. In the TE input case, 49.4% is directed into arm A and 50.2% is directed into arm B, whereas for a TM mode, 49.8% is injected into arm A and 50.1% is injected into arm B. With the aforementioned parameters, the silicon based power splitter can effectively divide the power equally between the two interferometer arms regardless of the input polarization state, and this functionality is offered in a compact device with a total length of $L_{coupler} =$

15.125μm and a footprint of only 13.84μm^2.

Half-Wave Plate

Reciprocal polarization rotation in a Faraday effect optical circulator is provided though the use of a HWP with its axis at $\pm 22.5^\circ$ from the z-axis. For nanophotonic integration, the use of bulk crystals [170, 171] or thin films [172] as HWPs are impractical, and a compact waveguide solution that is compatible with a silicon photonics platform is required. This key challenge can be overcome through the use of an asymmetric waveguide structure. In such a structure, the optic axis of the waveguide is rotated with respect to the standard coordinates, so a TE or TM mode input excites two hybrid modes. As the modes propagate down the guide, beating between the two hybrid modes allows the output polarization state to be tailored. Examples of this include optical waveguides with slanted sidewalls [184, 185], asymmetric nanowire designs [186], or the more recently demonstrated L-slot architecture [143]. The latter design is chosen for incorporation into the present optical circulator implementation due to its lossless dielectric architecture, and capability of providing polarization rotation within a distance of only a few wavelengths.

The waveguide HWP under consideration is schematically depicted in Fig. 4.12(a). It consists of a Si core having a width of $w_c = 220$nm and a height of t_c. An L-shaped Si region with $t_L = 135$nm and $w_L = 130$nm surrounds the top and side of the Si core, separated by a slot of $g_{HWP} = 40$nm. The HWP is embedded into an SiO_2 cladding. The length of the HWP is defined by the point where a π phase retardation is accrued between the two eigenmodes of the HWP, and is given by $L_{HWP} = \pi/\Delta\beta_{HWP}$, where $\Delta\beta_{HWP}$ denotes the difference in propagation constants of the two hybrid HWP modes. Figure 4.12(b) plots the Stokes parameters of the HWP output at the corresponding L_{HWP} on the Poincaré sphere as t_c is varied from 135nm to 235nm for both TE and TM excitation (the red and blue trajectories in

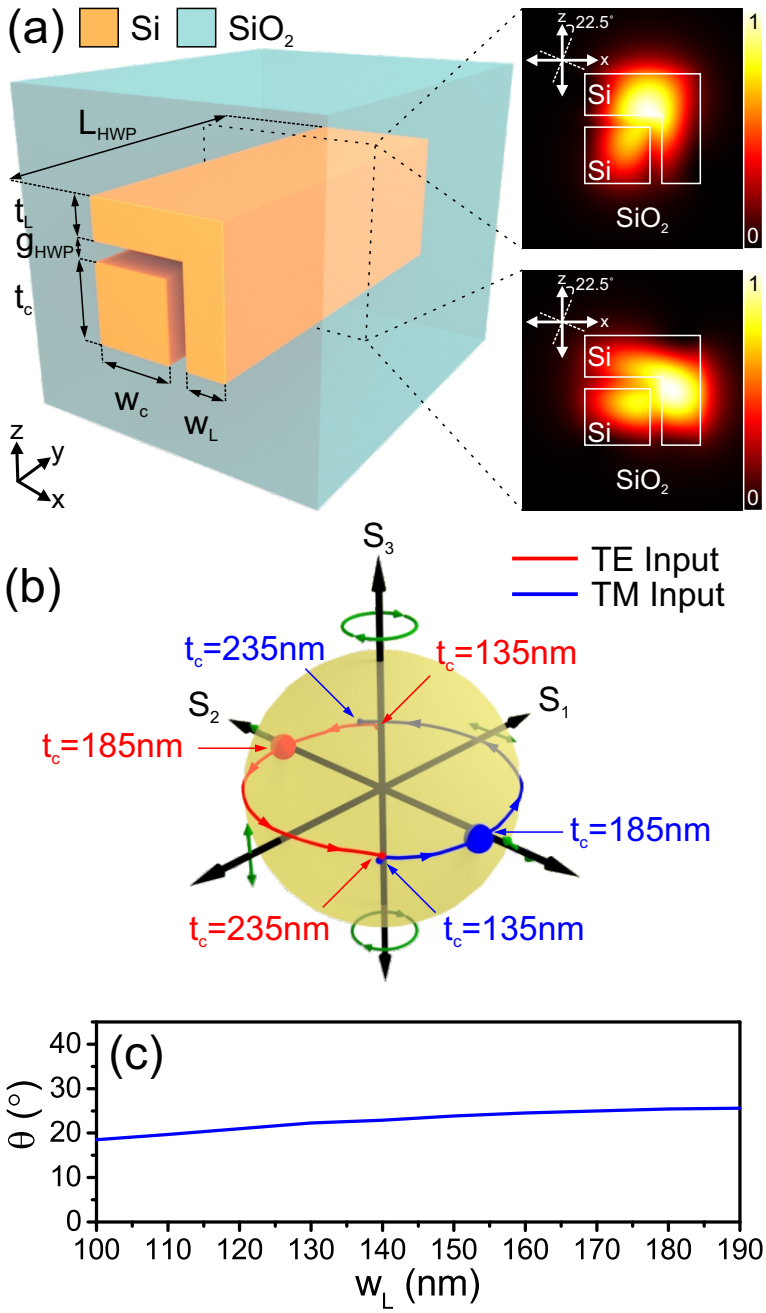

Figure 4.12: (a) Illustration of the silicon L-slot geometry employed as a HWP in the nanophotonic optical circulator. The two insets depict the $|H|^2$ field profiles of the rotated eigenmodes in the structure. (b) Plot of the Stokes parameters of the HWP output having a length of L_{HWP}, as a function of the silicon core height, t_c. The curves for both TE and TM input to the HWP are plotted around the Poincaré sphere. The green arrows indicate the polarization states represented by the poles of the sphere. $\pm 45°$ polarization is achieved for both input polarization states when $t_c = 185$nm. (c) Plot of the eigenmode axis angle as a function of w_L.

Fig. 4.12(b), respectively). As t_c increases, the output Stokes parameters plot a trajectory around the equator of the Poincaré sphere, indicating different linearly

polarized states. As highlighted in Fig. 4.12(b), at $t_c = 185$nm for both polarizations, the Stokes vector is dominated by the S_2 component, indicating that the output polarization state is linearly polarized at $\pm 45°$, as desired. This further implies that the eigenmode axes of the waveguide are rotated by 22.5° from the Cartesian axes, as denoted in the insets of Fig. 4.12(a). With these dimensions, the optimum length is found to be $L_{HWP} = 6.25\mu$m.

This L-slot HWP design can be shown to be extremely robust and tolerant to potential fabrication errors. For example, Fig. 4.12(c) depicts the eigenmode orientation angle, θ, as a function of w_L. Clearly, if one considers a design tolerance of $22.5° \pm 2.5°$ for the waveguide axis, the device will meet the specifications if 110nm$\leq w_L \leq$170nm, which is well within the tolerance of existing fabrication processes.

The preceding analysis considered an L-slot HWP with the Si L-shape on the top and right sides of the core. To realize the other required HWP, whose optical axis is rotated by 22.5° in the opposite direction, one simply needs to mirror the geometry about the z-axis, placing the Si L-shape on the top and left sides of the core. Thus, this design acts as an efficient waveguide HWP compatible with Si-based nanophotonic circuitry.

Faraday Rotator

Nonreciprocal operation within a Faraday effect based optical circulator is provided by a Faraday rotator. Here, with the goals of compact device dimensions as well as passive operation, the MO garnet Ce:YIG is chosen over Bi:YIG, as it exhibits higher MO properties, yielding shorter rotation lengths. Ce:YIG possesses a low saturation magnetization ($\mu_0 M_S = 1.5$mT), allowing devices to function with low applied magnetic fields, and the refractive index of Ce:YIG, $n_{YIG} = 2.21$ [122], provides good index contrast with SiO_2, the cladding material of interest in this design.

As in Section 4.2, the Faraday effect manifests when the Ce:YIG is magnetized

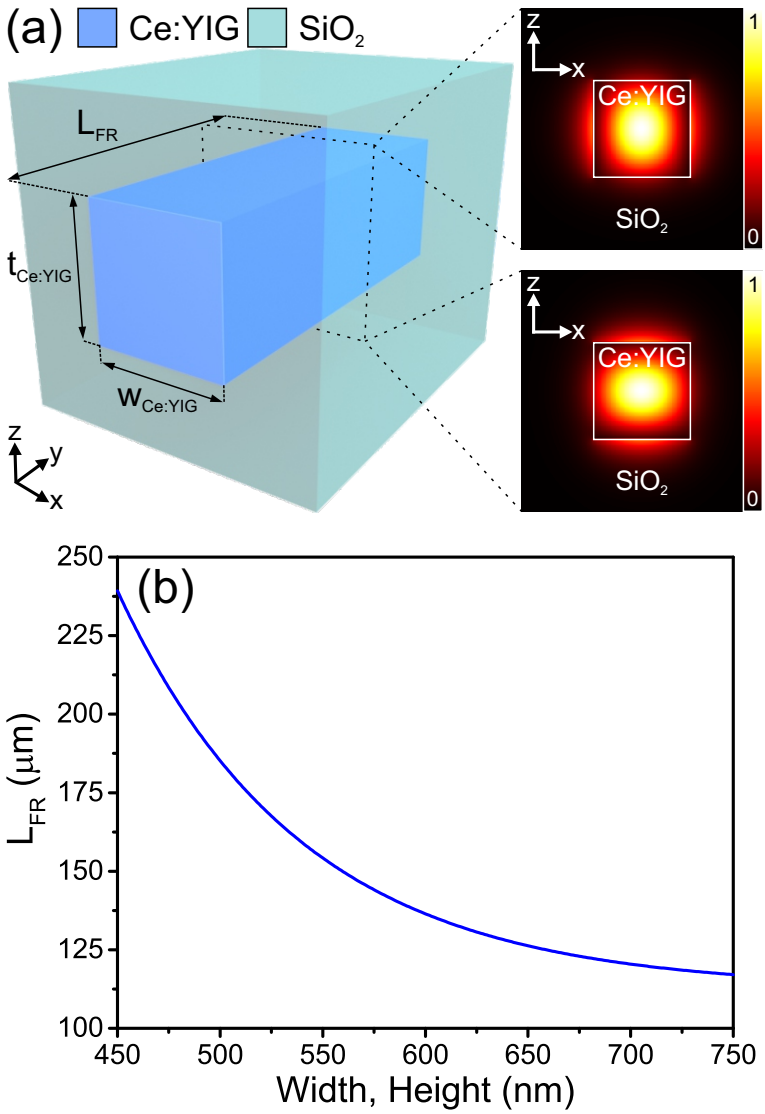

Figure 4.13: (a) Illustration of the Ce:YIG based Faraday rotator geometry employed within the circulator. The top inset indicates the $|E_x|^2$ field profile of the TE mode, while the bottom inset indicates the $|E_z|^2$ field profile of the TM mode supported by the structure. (b) Plot of the 45° rotation length, or correspondingly the 50% conversion length, as a function of the waveguide width and height. Here, a square Ce:YIG waveguide is considered, such that $t_{Ce:YIG} = w_{Ce:YIG}$.

parallel or antiparallel to the direction of light propagation, which here is considered to be along the $+y$-axis. This is accounted for through the full permittivity tensor within the FDTD simulations of the device. In this configuration, a static magnetic field ($\mathbf{H}_{static}$) is externally applied, and oriented along the $+y$-axis (i.e., $\mathbf{H}_{static} = \langle 0, +H_y, 0 \rangle$) that saturates the magnetization vector of the Ce:YIG along the same direction (i.e., $\mathbf{M} = \langle 0, +M_S, 0 \rangle$).

The Ce:YIG Faraday rotator design under consideration is illustrated in Fig. 4.13(a). It consists of a block of Ce:YIG having a width, $w_{Ce:YIG}$, and a height, $t_{Ce:YIG}$, embedded into a SiO_2 cladding. For efficient polarization rotation, there are two crucial factors that must be considered: the phase matching between the TE and TM modes of the structure, and the overlap of the modal fields within the magnetic garnet. Modal birefringence has been a fundamental limitation that has limited the use of Faraday rotators at the nanoscale. Previous studies have employed a quasi-phase matching technique to overcome this [70, 187]. As an alternative method to circumvent that limit, one can consider waveguides with a square cross section, i.e., $w_{Ce:YIG} = t_{Ce:YIG}$. Inherently, due to the symmetry of the cross section, phase matching will be satisfied. Polarization independent square and box-shaped [188] waveguides facilitate phase matching between TE and TM modes, and the symmetric structure offers the potential of phase matching over a broad bandwidth, which is particularly attractive for implementation into magneto-optic devices such as optical circulators. Additionally, due to the high refractive index contrast between the Ce:YIG core and the SiO_2 cladding, most of the optical fields are confined within the core, as indicated in the insets of Fig. 4.13(a), yielding high overlap between the two optical modes.

For implementation in an optical circulator, the Faraday rotator is only required to rotate the state of polarization by 45°. The optimal device length, L_{FR}, is the length at which this condition is met. Considering a pure TE or TM input lightwave to the rotator, this implies that 50% of the optical power is contained in each mode. L_{FR} is calculated via FDTD as a function of the Ce:YIG dimensions, $w_{Ce:YIG} = t_{Ce:YIG}$, and the results are plotted in Fig. 4.13(b). As the waveguide dimensions are reduced, a greater fraction of the optical power is carried outside of the Ce:YIG core, in the SiO_2 cladding, and the modal overlap in the Ce:YIG is reduced while the length required to rotate the polarization increases. Conversely, as the waveguide dimensions are

increased, less of the optical power is carried in the cladding, and the modal overlap improves, decreasing the required length. There is an inherent trade-off between the waveguide dimensions, and the device length. For a sufficient balance between these two parameters, $w_{Ce:YIG} = h_{Ce:YIG} = 600$nm is chosen for the present device implementation, yielding a rotator length of $L_{FR} = 137.25\mu$m. Remarkably, due to the high modal overlap, this is the smallest 45° Faraday rotator presented to date.

4.3.3 Circulator Performance

The complete circulator, composed of all three of the aforementioned components, is depicted in Fig. 4.14(a). The three different waveguide architectures are arranged such that the base of each is set at a common height, implying that they are all fabricated on a common substrate plane. As well, in each arm, the centers of each waveguide are aligned. As described above, arms A and B are separated by a center-to-center distance of $g_{wg} = 3\mu$m. In this arrangement, no crosstalk occurs between the waveguide arms. The total device length is 173.75μm, and the total on-chip footprint is only $625.5\mu\text{m}^2$. As a result, this device is the smallest Faraday effect optical circular presented to date, as circulators of this class are typically several millimeters in length.

The device was simulated to determine operation under excitation from both TE and TM modes, at all four input ports. Cross sectional field profiles are depicted in Figs. 4.14(b)-(e) for TE inputs to Ports 1-4, respectively, and in Figs. 4.14(f)-(i) for TM inputs to Ports 1-4, respectively. In both cases, clear sequential power circulation is observed as expected, with power injected at Port 1 routed to Port 2, Port 2 to Port 3, Port 3 to Port 4, and Port 4 to Port 1. Transmission values at each port for the TE excitation case are presented in Table 4.1, while the corresponding transmission values for TM excitation are tabulated in Table 4.2. During TE mode operation, insertion losses range from 2.2-2.7dB, while the isolation ratio between the output

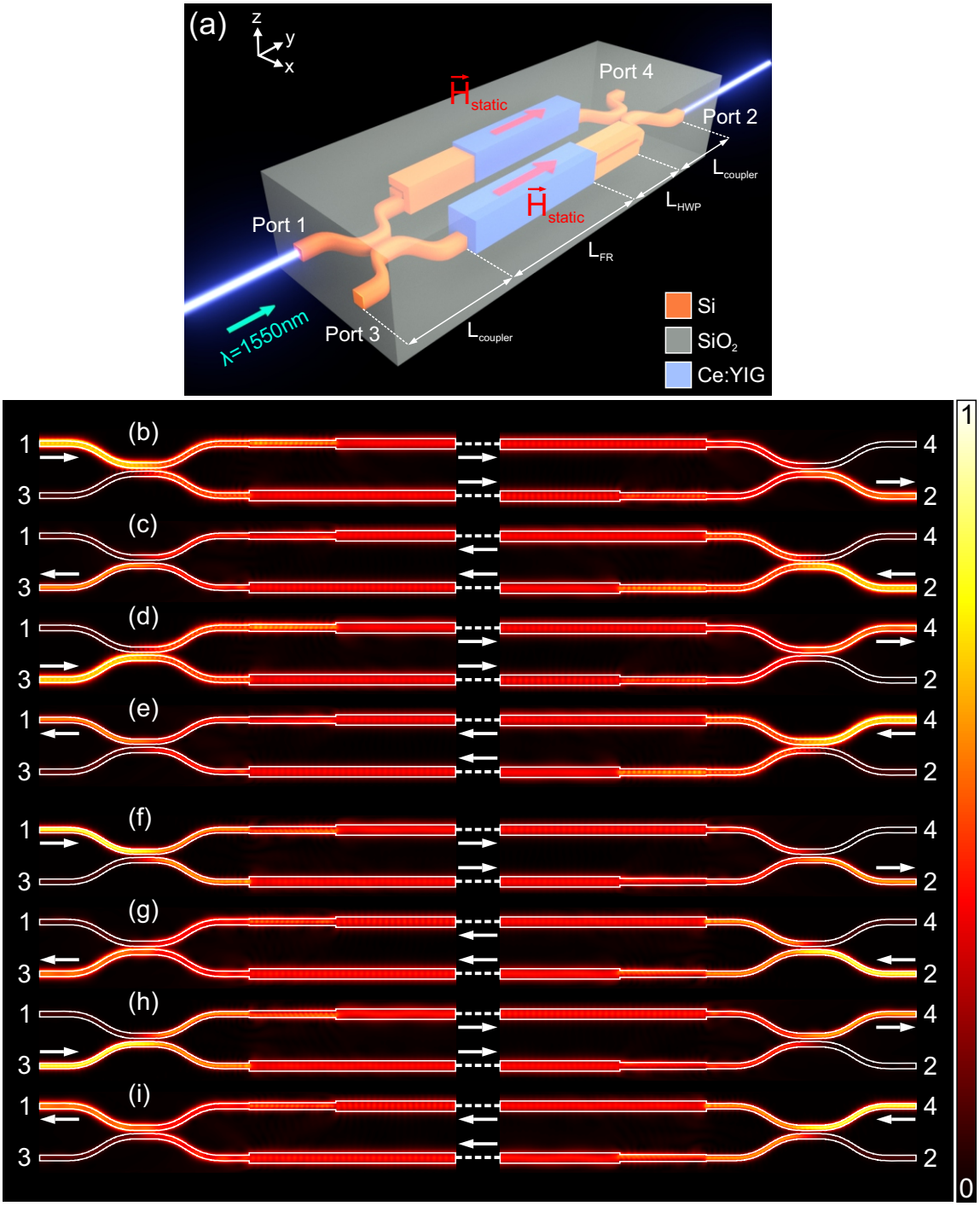

Figure 4.14: (a) Schematic depiction of the complete nanophotonic optical circulator component. (b)-(e) illustrate the $|E|$ field profiles at the input and output of the device for a TE mode injected into (b) Port 1, (c) Port 2, (d) Port 3, and (e) Port 4. (f)-(i) depict the $|E|$ field profiles at the input and output of the device for a TM mode injected into (f) Port 1, (g) Port 2, (h) Port 3, and (i) Port 4. Only the first 30μm from each end of the device are depicted here, and all field profiles are calculated 300nm above the substrate plane.

Table 4.1: Circulator Transmission for TE Mode Input

		Output Port			
		1	2	3	4
Input Port	1	-22.6dB	-2.2dB	-17.2dB	-22.5dB
	2	-16.0dB	-25.0dB	-2.5dB	-17.0dB
	3	-17.0dB	-23.0dB	-22.6dB	-2.3dB
	4	-2.7dB	-17.2dB	-15.0dB	-24.9dB

Table 4.2: Circulator Transmission for TM Mode Input

		Output Port			
		1	2	3	4
Input Port	1	-21.6dB	-2.6dB	-18.9dB	-24.4dB
	2	-15.9dB	-20.4dB	-2.6dB	-19.0dB
	3	-18.9dB	-24.9dB	-21.7dB	-2.5dB
	4	-2.5dB	-18.9dB	-15.7dB	-20.4dB

and input ports range from 20.3-22.5dB. Extinction ratios between the output port and the adjacent forward port range from 12.3-20.7dB. Similarly, for TM excitation, insertion losses in the range of 2.5-2.6dB, isolation ratios between 17.8-19.2dB, and adjacent port extinction ratios between 13.2-22.4dB are observed. From this data, two features are apparent: first, the device efficiently provides high extinction ratios, and second, the device performance is comparable for both polarization inputs. Since all materials are modeled as effectively lossless at $\lambda = 1550$nm, the insertion losses result from reflections and scattering at the interfaces between the various waveguide architectures, due to the inherent impedance mismatch. This could potentially be remedied with the inclusion of an impedance matching structure between each component, such as some variation of an adiabatic taper [189, 190], at the expense of increasing the device length.

To examine the operation of the circulator over broad bandwidth, the device is excited with both TE and TM modes at Port 1 at wavelengths of $\lambda_1 = 1520$nm and $\lambda_2 = 1580$nm. The resultant transmission values are collected in Table 4.3 for TE mode input, and in Table 4.4 for TM mode input. The insertion loss changes by only 0.2dB for TE input and 0.4dB for TM, while the corresponding isolation

Table 4.3: Port Transmission vs λ for TE Input at Port 1

		Wavelength		
		1520nm	1550nm	1580nm
Output Port	1	-22.4dB	-22.6dB	-20.2dB
	2	-2.5dB	-2.2dB	-2.3dB
	3	-14.6dB	-17.2dB	-18.8dB
	4	-16.0dB	-22.5dB	-15.0dB

Table 4.4: Port Transmission vs λ for TM Input at Port 1

		Wavelength		
		1520nm	1550nm	1580nm
Output Port	1	-24.2dB	-21.6dB	-21.8dB
	2	-3.0dB	-2.6dB	-2.8dB
	3	-15.9dB	-18.9dB	-16.1dB
	4	-16.6dB	-24.4dB	-16.9dB

ratios vary by up to 2.5dB for TE and 2.2dB for TM. In this regard, the operation is comparable across this wide wavelength range. The only noticeable change occurs in the extinction ratio between the two forward ports, Ports 2 and 4. Here, the extinction ratio drops from 20.3dB to 13.5dB at λ_1 and 12.7dB at λ_2 for TE input, and from 21.8dB to 13.6dB at λ_1 and 14.1dB at λ_2 for TM input. Both the refractive index and specific Faraday rotation of the Ce:YIG are slowly varying and relatively constant in this wavelength range [191], so the primary bandwidth limitation is the response of the 3dB power splitter, which deviates from the ideal 50/50 as the wavelength moves farther from $\lambda = 1550$nm. The bandwidth limitation of the proposed circulator can be further optimized by employing a more sophisticated wide band coupler [192, 193], rendering the circulator to be implemented in versatile platforms with flexible choices of device components. The primary deviation is observed in the extinction ratio between the two forward ports, i.e., the difference in transmission between Ports 2 and 4. Furthermore, while the worst case drop in extinction ratio observed here is seen for the case of TM input at $\lambda_1 = 1520$nm. The forward extinction ratio drops from 21.8dB at $\lambda = 1550$nm to 13.6dB at $\lambda_1 = 1520$nm, for a change of 8.2dB. However, a significant 13.6dB extinction ratio is still observed. Evidently, the device

is operable over a bandwidth of at least 60nm, making it more versatile and capable of handling wavelength instabilities than the much more compact ring resonator [178] or photonic crystal devices [179].

The presented device exhibits all the characteristics of a truly versatile nanophotonic optical circulator. It provides efficient sequential power routing with low insertion losses and high extinction ratios, operates over a sizable bandwidth for both TE and TM polarization states, and achieves this functionality within a short device length and a small on-chip footprint.

4.3.4 Proposed Fabrication Scheme

Due to the nature of this circulator requiring three separate waveguide geometries, it is crucial to examine how these diverse platforms can be integrated into a single device from a practical standpoint. Despite the architectural complexities, such a device could be fabricated in any well-equipped nanofabrication facility. One potential process flow for this purpose is schematically depicted in Fig. 4.15, where one waveguide from each component is depicted during each step of the process to illustrate the simultaneous construction of each device. Beginning from a standard SOI wafer (Fig. 4.15(a)), photolithography would be performed, exposing a region slightly wider than the width of the Faraday rotator, and a standard Si reactive ion etch (RIE) would remove the device layer from this area (Fig. 4.15(b)). Subsequently, the Ce:YIG would be deposited via PLD or sputtering (Fig. 4.15(c)), and a lift-off procedure would remove the Ce:YIG outside of the exposed region (Fig. 4.15(d)). Electron beam lithography (EBL) would pattern resist over the locations for the power splitter and HWP cores, and a standard Si RIE would define them (Fig. 4.15(e)). After removal of the resist, a brief FIB milling step could adjust the waveguide core dimensions to their required values (Fig. 4.15(f)). Following this, a thin film of SiO_2 would be conformally deposited over the structures via chemical vapour deposition

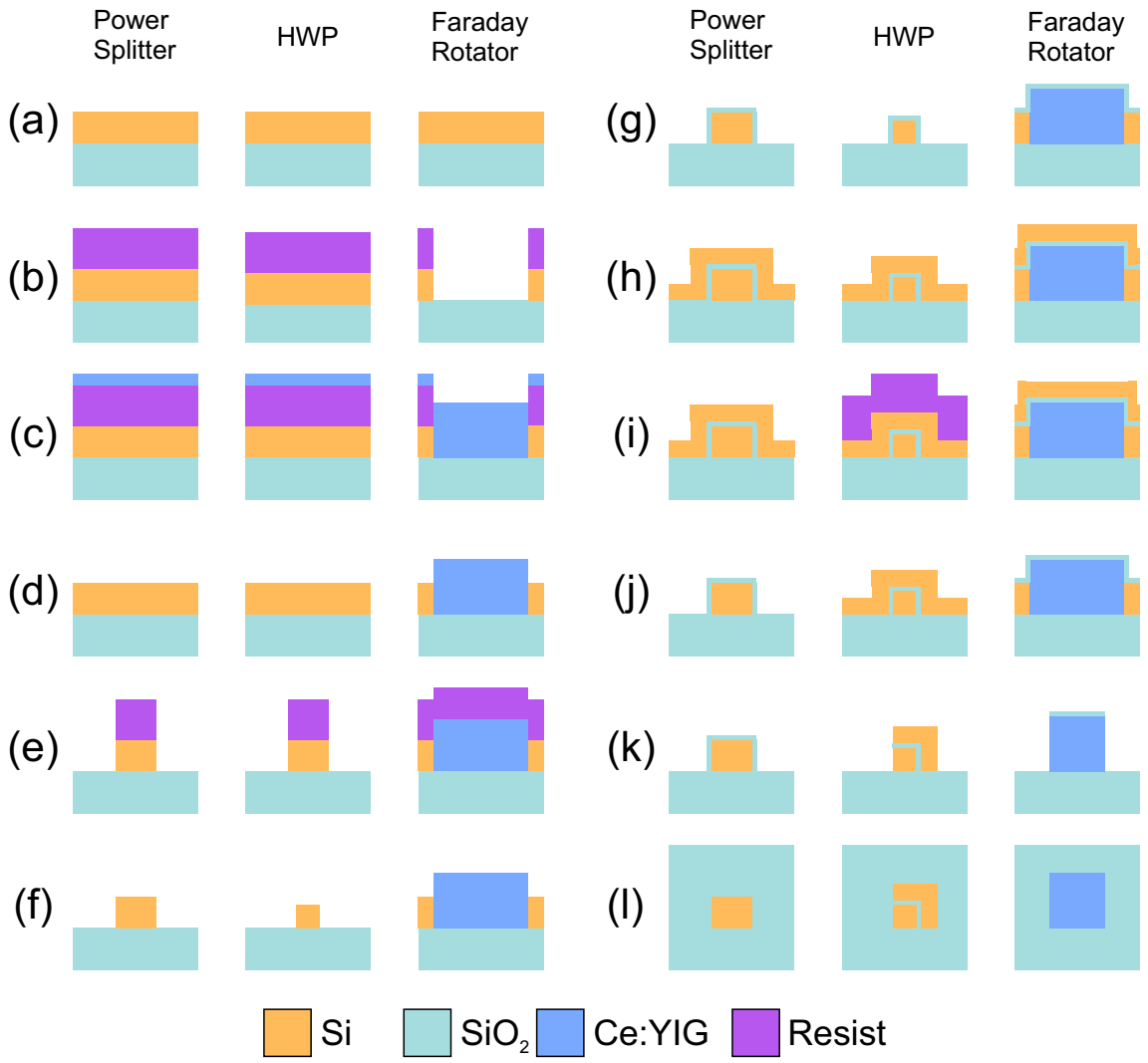

Figure 4.15: Potential process flow for fabricating the optical circulator. (a) The initial SOI chip. (b) Photolithography defines a region slightly larger than the Faraday rotator, and a Si RIE removes the devices layer from this region. (c) Deposition of the Ce:YIG via sputtering or PLD. (d) Lift-off of the resist and garnet. (e) An EBL step allows for the patterning of the HWP and power splitter cores via a Si RIE. (f) Use the FIB to adjust the Si core dimensions to their proper design values. (g) CVD produces a thin SiO_2 spacer layer. (h) CVD of the amorphous Si layer. (i) EBL covers the HWP in resist, but exposes the rotator and power splitter. (j) A Si RIE removes the excess Si covering the rotator and power splitter, and the resist over the HWP is removed. (k) FIB milling is used to pattern the HWP and rotator to the appropriate widths. (l) A thick SiO_2 cladding is deposited via CVD. Note that the cross-sections depicted in this figure are not to scale.

(CVD), as shown in Fig. 4.15(g), to define the gap between the HWP core and L-shaped cladding. A second CVD step would deposit a layer of amorphous Si over the structures (Fig. 4.15(h)), while another lithography step would mask off the HWP, leaving the power splitter and Faraday rotator exposed, as shown in Fig. 4.15(i). A

Si RIE would remove the CVD deposited Si over these two components, and leave it present around the HWP (Fig. 4.15(j)). After the resist is removed, another FIB milling step could accurately define both the HWP, as well as reduce the width of the Ce:YIG deposited for the Faraday rotator to the desired value, as shown in Fig. 4.15(k). Finally, a thick SiO_2 cladding could be deposited via CVD, simultaneously completing the fabrication of all three of the required components, as depicted in Fig. 4.15(l). This exemplary process would be able to construct this versatile optical circulator on a chip scale platform for integration into nanophotonic optical circuitry.

4.4 Summary

This chapter has demonstrated the applicability of the Faraday effect to nanophotonic and nanoplasmonic devices. First, a novel hybrid ridge-plasmonic Faraday rotator was designed to facilitate long-range phase matched photonic TE and plasmonic TM modes. This phase matching allowed the magnetized device to convert between the two modes via the Faraday effect, with 99.4% mode conversion efficiency. With the inclusion of a nearby metallic transmission line to generate transient magnetic fields, the magnetization of the Bi:YIG, and correspondingly the output polarization state, were effectively manipulated. The device was operable in two dynamic modes, mapping electrical pulses onto the output polarization state or generating continuous polarization oscillations at frequencies in excess of 10GHz. As such, this device presents a unique and versatile platform for polarization control in integrated plasmonics, and notably is the first presentation of a plasmonic waveguide supporting the Faraday effect.

With the fundamental understanding of the design principles behind such nonreciprocal polarization modulators, incorporating a Faraday rotator into a more complex platform was examined by designing and characterizing an optical circulator.

Specifically, a versatile, broadband, and polarization-independent device was presented. This nanophotonic circulator employed three distinct waveguide architectures: silicon photonic waveguides for 3dB power splitters/combiners, a compact silicon-based L-slot HWP, and a square shaped Ce:YIG Faraday rotator. The designed component was shown to be both polarization independent and more compact than any other Faraday effect based optical circulator presented to date, while providing high isolation ratios and low insertion losses over a broad bandwidth. These devices fill a crucial niche of applications in on-chip optical signal routing, isolation, signal multiplexing, and enabling bi-directional full duplex operation within a chip scale optical network, as well as polarization division multiplexing and polarization encoding within nanoplasmonic circuitry. These magnetoplasmonic polarization modulators and nanophotonic circulators are envisioned to comprise integral building blocks within future nanoplasmonic optical networks.

Chapter 5

Metal-Encapsulated Nanoplasmonic Waveguide Fabrication Process Development

5.1 Introduction

While the nonreciprocal MO effects and magnetoplasmonic devices thoroughly explored in the previous chapters provide intriguing and versatile mechanisms of manipulating the phase, polarization, and intensity of an optical wave, they cannot provide other functionalities demanded by integrated optical circuitry: namely, light generation. This is difficult to achieve on-chip, as chip-scale devices are based on the well-developed Si platform, and Si itself is an indirect bandgap semiconductor and does not allow for stimulated emission. Currently, options for on-chip optical sources include employing other material systems, or utilizing the intrinsic strong $\chi^{(3)}$ nonlinearity of Si to generate new frequencies via nonlinear optical interactions.

Recent works have shown that an Si-Au MI nanoplasmonic waveguide only a few microns in length is capable of generating record THG conversion efficiencies, and

producing strong green light from excitation via ultrafast $\lambda = 1550$nm laser pulses [107]. The prospect of employing a new hybrid structure, the metal-encapsulated (ME) geometry, is intriguing. This structure, which is a superposition of both the MI and MIM geometries, is essentially a MIM waveguide with a metal top cladding and can provide a pathway for further improving the THG conversion efficiency. This topic will be explored extensively in the following chapters of this thesis.

The design of an optimized nanofabrication process flow was a significant component required for the work in this thesis. The resultant procedure described in the following sections is the culmination of nearly two and a half years of process development endeavors. While basic MI and MIM plasmonic waveguides are conceptually simple, physically realizing a hybrid ME structure in a platform that can be easily excited, practically handled, and simultaneously characterized is a challenging task. Such a technique requires the merger of both microscale and nanoscale fabrication methodologies. Furthermore, the presented fabrication scheme is not limited to SOI architectures. It can conceivably be extended to examine nonlinear effects within a vast array of device geometries and material systems, making it an invaluable tool for nanoplasmonic waveguide characterization, and nonlinear component design.

This chapter describes in detail the design and optimization of the fabrication process flow developed to create ME nanoplasmonic waveguides on a characterization beam architecture. The chapter begins with a brief discussion of current methods of coupling light into such waveguides, before discussing the goals of an ideal characterization architecture and the envisioned platform. Details of each individual fabrication step used to create the samples characterized in Chapter 6 are presented within the subsequent sections.

5.2 Methods of Coupling to Plasmonic Waveguides

As plasmonic waveguide devices have been at the forefront of nanophotonics research for decades, a variety of novel schemes to couple incident light into these structures have been presented in the past. As discussed Chapter 2, one of the first methods of surface plasmon coupling involved TIR of light within prisms in the Kretschmann [26] or Otto [27] configurations. Accordingly, plasmonic slab waveguides have been fabricated on the surface of prisms [194], however this scheme is not compatible with integrated nanoplasmonic circuit applications. Prisms are large, bulky, and are not easily integrated into chip scale devices. The waveguides formed in this manner are wide slab waveguides; they do not provide 3D field confinement. Moreover, due to the reflective excitation geometry, they are often characterized through reflectivity measurements. While this is useful for sensing applications, the characterization scheme does not provide much information regarding nonlinear light generation within the waveguide core. Outcoupling of nonlinearly generated light in this geometry is challenging as well.

The second general method elaborated on in Chapter 2 was the grating coupling scheme, where a periodic grating imparts the additional momentum required to launch a surface plasmon. Simple rectangular gratings are useful for launching plasmons that propagate normal to the grating [28], but these SP waves are unconfined. To confine them, another structural component is required. Structures such as a metallic taper following the grating [55] have been shown to focus the surface plasmons down to a highly confined spot. Modifying the grating from a simple rectangular geometry to a circular metallic grating can focus the excited surface plasmons as well [195]. Unfortunately, these grating structures are plagued by both large on-chip footprints, as well as optical bandwidth limitations. Considering excitations with a broadband pulse, only a narrow spectral range will actually be coupled into the device. Moreover, at the high peak electric fields required for nonlinear optical studies, nonlinearities

can occur within this coupling structure, and it is extremely difficult to isolate the nonlinearities generated in the grating from those generated in the waveguide of interest. For these reasons, grating coupling schemes are not ideal for many nanoscale nonlinear optical investigations.

Tapered Si photonic bus waveguides have been employed to characterize such structures [31, 196], when considering Si-based nanoplasmonics operating in the telecom band specifically. Since Si is essentially lossless around $\lambda = 1550$nm, light can be coupled into a large Si photonic waveguide and focused for coupling into a plasmonic guide via a gradual taper in the Si bus width. An alternative scheme employs momentum matching in an orthogonal junction geometry [197, 198]. While this method eliminates the need for tapering the Si waveguide, it still requires the implementation of a Si bus guide. These schemes are useful for examining the linear and passive properties of the Si-based nanoplasmonic waveguides, but they are not practical for nonlinear optical studies. This is due to the high nonlinearity of Si, as the high peak fields will excite THG within the coupling bus waveguide, and simultaneously attenuate the pulse due to TPA before the light reaches the nanoplasmonic guide to be studied. In this respect, alternate approaches must be taken for nonlinear studies.

The most direct method of coupling to any waveguide is end-fire excitation [199]. In this scheme, focused laser light is incident directly on the end facet of the waveguide to be studied, and coupling is achieved through matching of the field distribution of the focused beam and the modes of the plasmonic waveguide. This mechanism has been employed previously in a number of studies into nonlinear phenomena in Si-based nanoplasmonic waveguides [107, 200, 201], and does not involve the fabrication of any additional coupling structure; it simply requires exposing the end facet of the waveguides under consideration. Any nonlinearities observed are due directly to the structure being tested, and, if one has knowledge of the properties of the exciting laser pulse such as its spectrum, correlations and conclusions can be drawn about the

spectrum of any nonlinearly generated light.

For the studies presented in the Chapter 6 of this thesis, end-fire excitation was selected as the optimal method for coupling laser light into the nanoplasmonic waveguides for the above reasons. In implementing such a scheme, special consideration has to be given to the method of delivering the incident laser pulses to the input facet of the waveguides. Compact methods such as a tapered optical fibers cannot be employed to bring in the laser for nonlinear studies. At the high electric fields and intensities required to excite the nonlinear behaviour within the nanoplasmonic guide, the incident pulse will also induce strong nonlinearities within the silica fiber. Over the considerable length of the fiber, these nonlinearities in the fiber will also modify and reshape the pulse both spectrally and temporally. Thus, in order to preserve the known state of the laser pulses, the light must be directly incident on the waveguides, without propagating through a fiber. Focusing the light through a high numerical aperture (NA) objective lens can provide a diffraction limited spot suitable for directly exciting the structure via end-fire coupling.

5.3 ME Waveguide Characterization Platform Concept and Design

In order to experimentally investigate nonlinear third-harmonic generation within silicon-based ME nanoplasmonic waveguides, and to excite this interaction via end-fire coupling, a unique approach towards sample design and preparation had to be explored. Investigating nonlinear optics within such a platform presents several significant challenges:

(1) Nanoplasmonic waveguides are inherently exceptionally small. Cross-sectional dimensions in this platform are designed to be on the order of $w_{Si} \times t_{Si} =$ 100nm×340nm. Furthermore, the propagation losses of plasmonic modes, due

to optical losses of the adjacent metal films, limit the propagation length of these waveguides to the order of a couple of microns in length. Because of this, the characterization platform must facilitate exciting the waveguide and extracting the transmitted light on this scale. One must also be able to handle and manipulate the waveguides in order to place them within the various fabrication tools as well as the characterization setup. This requires samples (i.e., the platform upon which the waveguides are constructed) on the millimeter scale. The ideal characterization platform must facilitate both of these length scales.

(2) The samples must facilitate direct end-fire excitation via an objective lens. This implies that the sample geometry must allow for the approach of a large, high NA objective lens to within a short focal distance of the sample (on the order of several hundred microns). Due to the relative size of the objective lens, this signifies that the nanoplasmonic waveguides must be fabricated on the edge of a given sample, to provide the required access.

(3) After excitation, the sample platform must provide a method of collecting the transmitted light to be directed to the required characterization equipment, such as power meters and spectrometers. The most compact method of achieving this is via a tapered and lensed optical fiber. An optical fiber can be employed at the waveguide output, as the intensity of the nonlinearly generated light is low enough as to not induce nonlinearities within the fiber.

(4) The fabrication must produce a high yield of devices. The number of samples obtained from a single die with consideration of the above factors must be maximized, and the fabrication method must be versatile and flexible to allow for arbitrary devices configurations to be constructed.

An artistic rendering of the envisioned concept that satisfies all of the aforementioned requirements is depicted in Fig. 5.1. Figure 5.1(a) depicts the envisioned

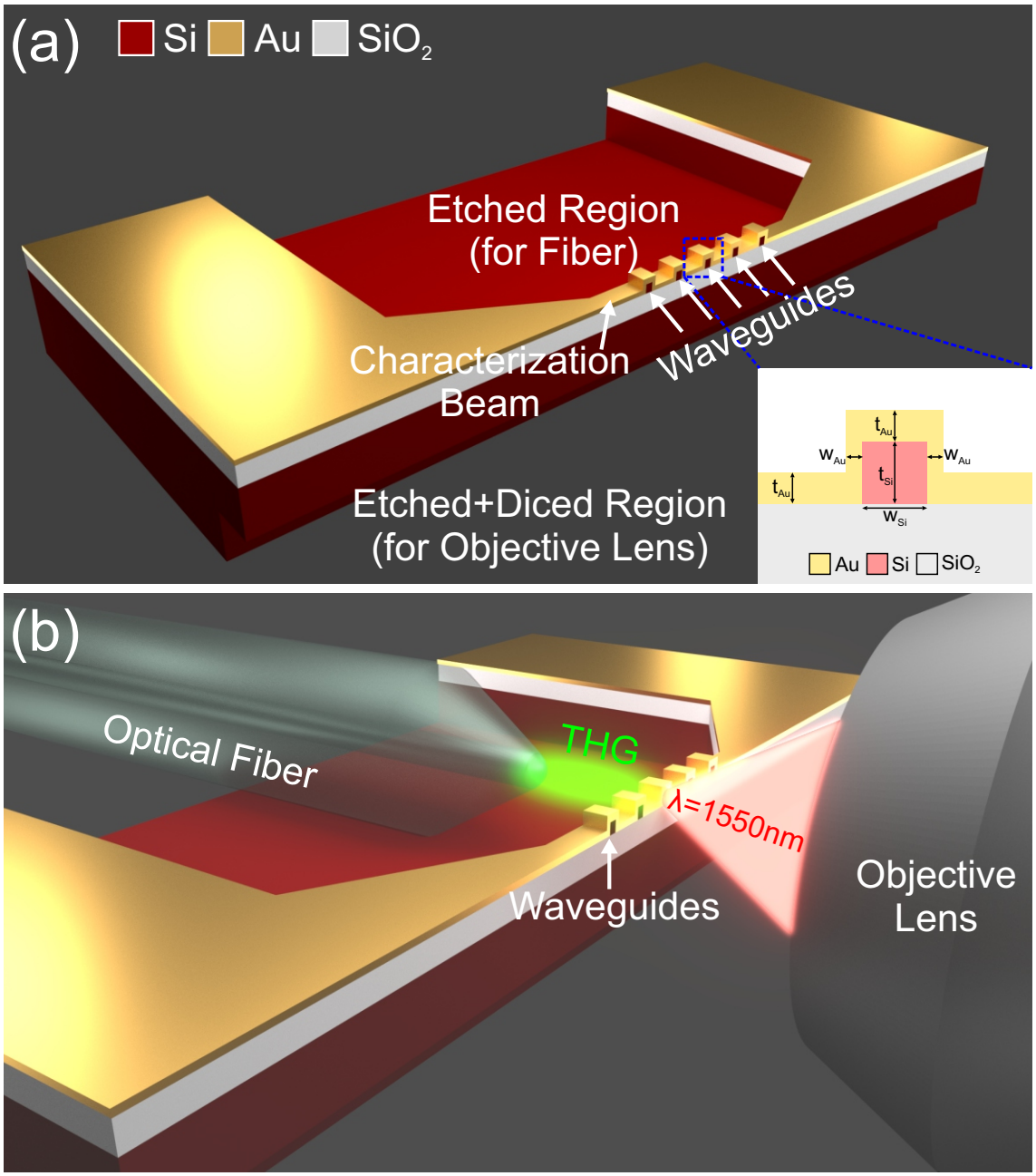

Figure 5.1: (a) Schematic illustration of the conceived design for a ME nanoplasmonic waveguide characterization architecture. Created from an SOI platform, ME waveguides are formed on a micromachined beam. Reactive ion etching creates a region where an optical fiber can approach and collect the transmitted light. A combination of reactive ion etching and dicing allows the waveguides to be formed on the edge of the sample, facilitating end fire excitation from an objective lens. The inset depicts the ME waveguide geometry, which can be considered as the superposition of MI and MIM waveguides, or equivalently, an MIM waveguide with a metal top cladding. (b) Artistic rendering of the envisioned excitation scheme, where an objective lens focuses $\lambda = 1550$nm laser pulses onto the input facet of the nanoplasmonic waveguides, while an optical fiber collects the light produced by THG in the waveguides.

sample, while Fig. 5.1(b) illustrates the excitation scheme. The inset of Fig. 5.1(a) depicts the ME waveguide geometry, which can be considered as the superposition of MI and MIM structures, or comparably, an MIM waveguide with a top metal cladding layer. The sample architecture is based on an SOI die, and the ME nanoplasmonic waveguides with arbitrary widths are fabricated atop a micromachined beam, which defines the length of the guides. The beam is supported between two large contact pads for practical handling. The sub-10μm wide micromachined beam is constructed at the edge of the sample, facilitating the approach of a large objective lens from one side, while the opposing side of the beam has been deeply etched to grant access to an optical fiber. This platform satisfies requirements (1)-(3) above, while challenge (4) will be shown to be satisfied with the subsequent construction of an array of devices with high yield. The following sections detail the process developed to fabricate the structure depicted in Fig. 5.1.

5.4 Fabrication Process Development

5.4.1 Layout and Substrate Preparation

Sample Design and Layout

A top-down schematic illustration of a typical sample is depicted in Fig. 5.2(a), where the blue region denotes the etched area, while the red region denotes the top of the etched mesas. The individual samples are designed to be 5.0mm×2.25mm in size, which is sufficient for manipulating and handling, while compact enough to achieve numerous samples per die. The two side handle regions measure 1.0mm wide and 1.9mm long, while the sub-10μm wide beam, where the nanoplasmonic waveguides are formed, is located at the top edge of the sample, and is taken to be 250μm long. The two side handle regions taper down from 0.5mm to the beam width over a length

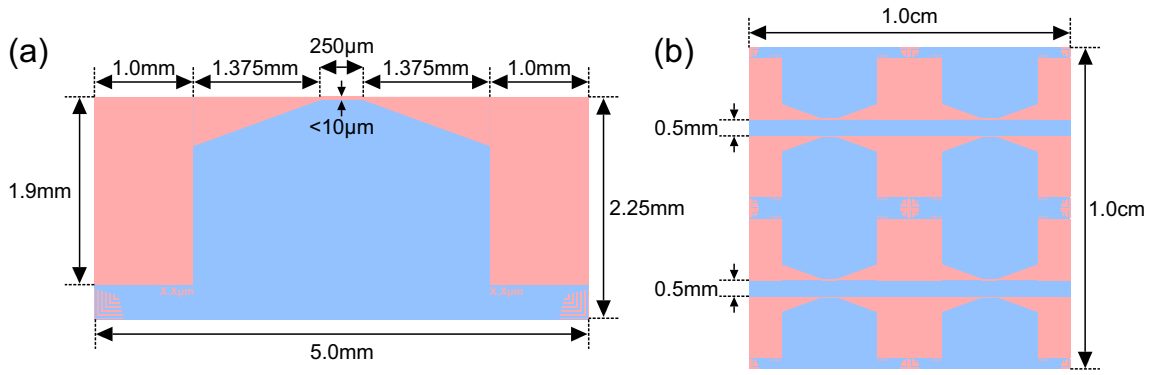

Figure 5.2: (a) Schematic layout of a single sample. (b) Illustration of an array of samples that can be fabricated on a 1cm×1cm die. In both images, the blue regions represent the etched areas, while the red areas represent the top of the structures.

of 1.375mm. The space below the two side handle regions and the bottom edge of the sample is reserved for alignment marks and labels.

The length of the narrow beam, at 250μm, is sufficiently long to host numerous nanoplasmonic waveguides placed side-by-side, while the etched area is large enough to allow for a tapered and lensed single mode optical fiber, having a maximum diameter $d_{fiber} = 125\mu$m, to be scanned along the length of the beam to collect transmitted light from the waveguides. This design maintains the full sample thickness over nearly three full sides of the sample, which yields sufficient mechanical stability for handling the sample with tweezers, and is crucial for protecting and maintaining the structural integrity of the fragile micromachined beam.

These patterns are arrayed onto a standard 1cm×1cm square, assuming a 0.5mm gap between adjacent beams as shown in Fig. 5.2(b), allowing for eight samples to be patterned per standard SOI die. The SOI wafers that form the bases of these devices consist of a $t_{Si} = 340$nm Si device layer, and a $t_{BOX} = 1\mu$m buried oxide (BOX) layer, on a 550μm thick Si substrate. A cross-sectional scanning electron microscope (SEM) image of an unpatterned SOI sample is shown in Fig. 5.3. The Si device layer will form the core of the ME plasmonic waveguides, while the SiO_2 buried oxide acts as the lower cladding. Since there is a significant refractive index contrast between these two materials ($n_{Si} = 3.48$ and $n_{SiO_2} = 1.44$ at $\lambda = 1550$nm), the modes will be

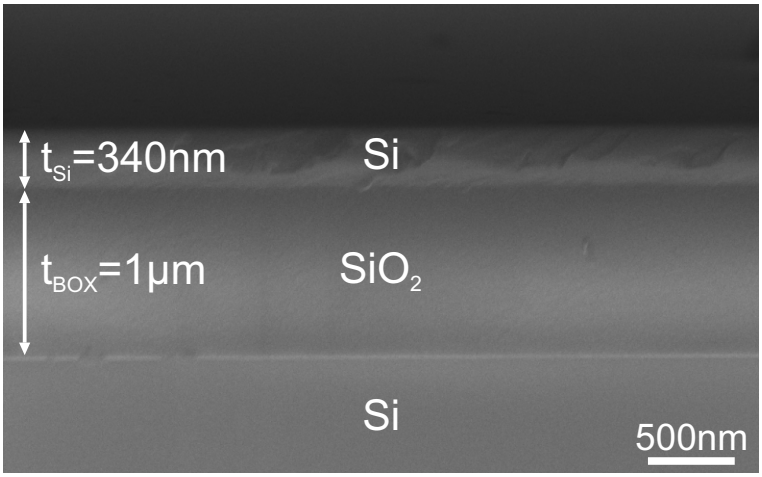

Figure 5.3: Cross-sectional SEM image of the SOI wafer employed as the basis for the device fabrication. The sample consists of a $t_{Si} = 340$nm Si device layer, a $t_{BOX} = 1\mu$m BOX layer, and a 550μm Si handle. Roughness in the device layer observed here is due to the cleaving process.

strongly confined to the Si core, which is desirable for nonlinear applications.

One of the key challenges in fabricating these structures is detaching the beams from the rest of the die. Silicon, with its well-defined crystal planes, is often cleaved into desirable sample sizes. This method has a number of drawbacks though. Namely, while Si will cleave along its crystal plane, there is some inherent randomness to the cleave which prevents it from yielding sub-micron accuracy, and this level of accuracy is required to detach the beams from the rest of the die without damage. Additionally, during cleaving, a significant amount of mechanical stress is present, which can easily destroy the fragile beam. One proposed solution to this problem was to RIE a trench into the silicon to selectively define the cleave location and improve its accuracy, however this method resulted in a low device yield [201]. Dicing is another procedure commonly used to detach samples that can provide high accuracy, but mechanical dicing is a rough process that creates chips, cracks, and damage on a scale of tens of microns at the edge. Here, dicing alone would destroy the fragile micromachined beams. Deep reactive ion etching (DRIE) all the way through the wafer would be another possible solution; however, DRIE through this thickness is time consuming and requires a hard metal etch mask, which was prohibited on the

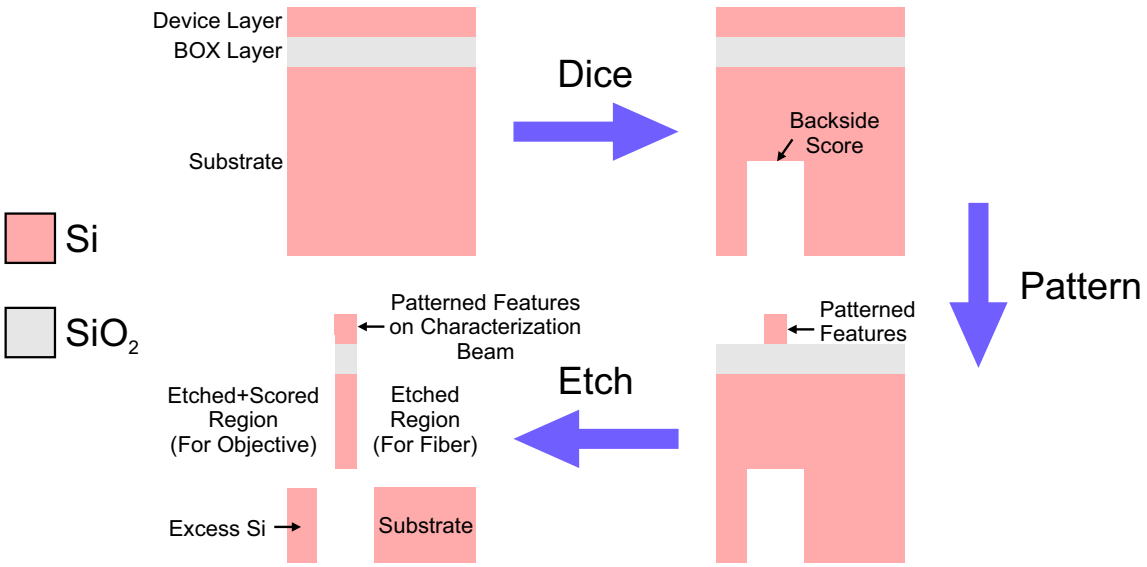

Figure 5.4: Illustration of the general hybrid dicing and DRIE approach for separating the structures from the rest of the die. This method creates the waveguides on the edge of the sample, such that a large objective can be brought in close proximity to couple in light. Depicted here is a cross-sectional profile of the characterization beam. Starting from the SOI wafer, a score is made with a dicing saw in the backside of the chip. The features are subsequently patterned directly above the score. A DRIE step simultaneously etches the region for the fiber and breaks through the score, effectively separating the characterization beam from the rest of the chip and leaving the waveguides on the edge. In the last image, the beam is connected to the substrate out of the plane of the page.

DRIE tools available. Thus, a new approach had to be developed.

An extremely robust solution to this problem involves a hybrid process of dicing and DRIE. The general approach is illustrated in Fig. 5.4. First the chip is inverted, and a dicing saw is employed to score the backside of the wafer, but not cut through it completely. These scores are precisely positioned directly underneath the location where the beam will be patterned on the top surface, and run parallel to the beam along the entire length of the die. Following this, the required patterning is done on the top surface as normal. During this patterning, an etch mask is applied to define the beam, and the sample is placed within a DRIE tool and etched. Due to the alignment between the cut and the lithography, this etch removes the remaining Si between the top surface and the cut, effectively separating the micromachined beam from the bulk sample with high accuracy. This hybrid method avoids the stress and destructive tendencies of dicing the surface, while eliminating the need for

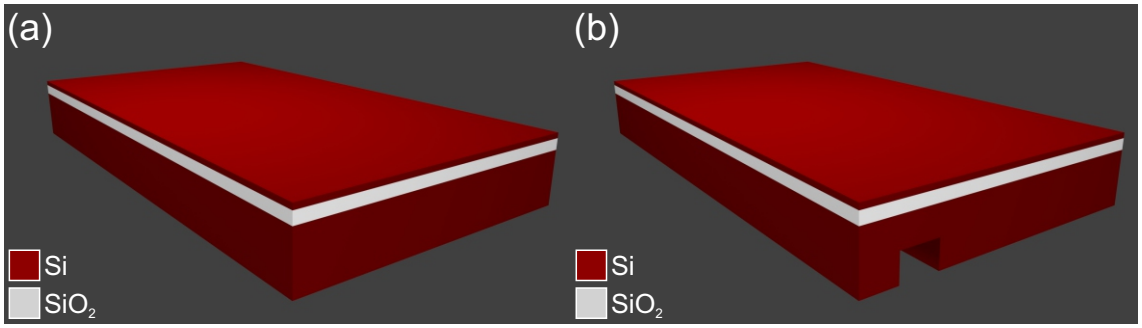

Figure 5.5: Schematic illustration of (a) the SOI platform that forms the basis of the device fabrication, and (b) the SOI sample scored on the backside by the dicing saw. The dicing saw cut does not extend all the way through the sample. Images are not to scale.

extremely deep etches and large-area hard metal etch masks. As will be shown in the following sections, this DRIE can be performed simultaneously with the other RIE steps required to define the structure, and the dicing and lithography tools can provide more than adequate alignment accuracy.

Backside Scoring

The process of fabricating the samples begins with a 100mm diameter SOI wafer, which is diced all the way through to form 1cm×1cm die. It is crucial to note that the width of the dicing saw blade is 50μm, and thus, the actual dimensions of a die become 9.95mm×9.95mm. This must be accounted for in subsequent steps. The SOI starting point is schematically depicted in Fig. 5.5(a). The chip is inverted for the backside scoring, depicted in Fig. 5.5(b). Beginning at the bottom of the chip, the center of the blade is precisely aligned to the edge of the chip. The score pattern is depicted in Fig. 5.6(a), where the scores are aligned to the location of the beams depicted in Fig. 5.2(b), as well as the midpoints of the die. The dicing saw blade is set to make each score 400μm deep into the substrate, at a speed of 15mm/s, leaving 150μm of the silicon substrate intact. This cut depth was found to be robust, as subsequent handling during processing did not break the chips along the scores. Cutting 20μm deeper resulted in numerous chips breaking during handling

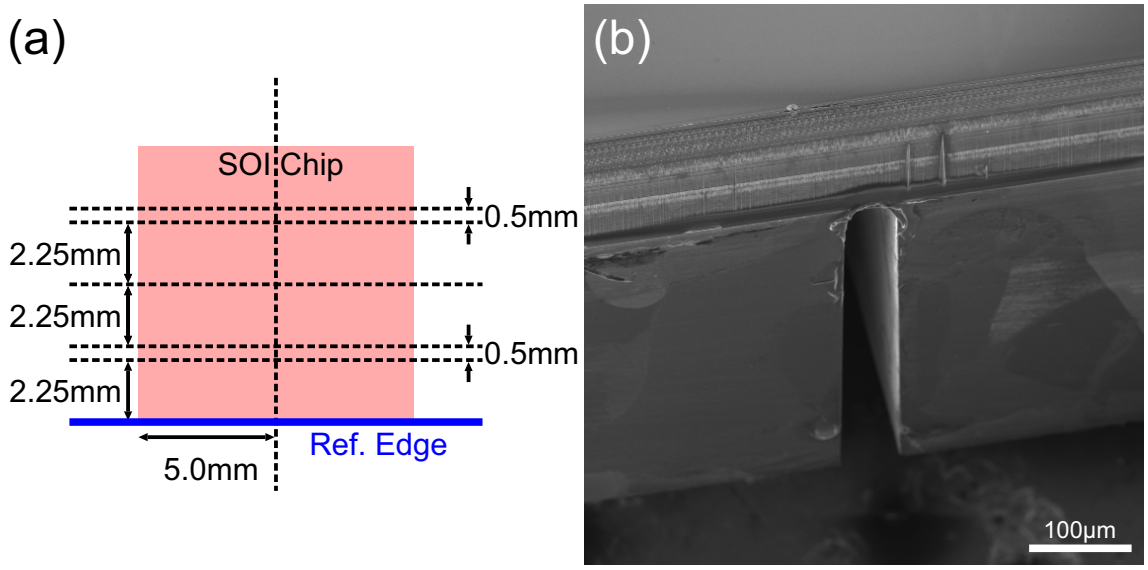

Figure 5.6: (a) Diagram of the dice pattern for scoring the back side of the SOI chip. The black dashed lines indicate the cuts made, while the solid blue line denotes the reference edge. (b) SEM image of one of the scores made in the backside of a chip.

and cleaning steps, and therefore, 150μm of material remaining was deemed optimal. An exemplary SEM image depicting one such score is shown in Fig. 5.6(b).

After the backside scoring was completed, the backside of the chip had to be marked with a diamond scribe to indicate the edge from which the dicing alignment was taken. All further alignment during subsequent lithography steps in this process must be taken with reference to this edge of the chip. Due to the slight reduction in the die's dimensions from the finite width of the blade and the fact that the reference point for the scoring was taken to be the bottom edge of the chip, the pattern is not symmetric on the die (i.e., the scores are 50μm closer to one edge of the die than the other). Accidentally rotating the chip and performing lithography would cause complete misalignment between the scores and the defined patterns. Therefore, this edge must be carefully tracked throughout the fabrication process, and must be maintained as the reference point.

5.4.2 Si Core Mask Definition

Cleaning the Substrate

After the backside of the chip was scored and marked, the samples were thoroughly cleaned prior to further processing. This was done in two stages. First, the samples were placed in a bath of Piranha, which is a 3:1 mixture of 96% H_2SO_4 and 30% H_2O_2, for 15 minutes. This mixture heats up and effectively strips organic contaminants from the surface of the sample. Following this bath, the samples were rinsed in deionized (DI) water.

The second stage of cleaning involved placing the samples in a bath of buffered oxide etch (BOE) for 5 minutes, followed by a rinse in DI water. BOE is a commercially available etchant with comprised of a 10:1 mixture of 40% NH_4F and 49% HF which is used for wet etching SiO_2. This bath removes any native oxide that has formed on the surface of the Si device layer.

Lithographically Defining the Si Core Mask

The next task was to define a mask that would be used to etch the Si cores for the ME nanoplasmonic waveguides. Due to the desired ME core widths $< 1\mu$m, EBL would be required due to its extremely high resolution. The first step in this process was to spin on a layer of resist for the EBL exposure, as schematically depicted in Fig. 5.7(a). A bilayer resist was chosen to improve the results of the following lift-off step. The bilayer was composed of two different poly(methyl methacrylate) (PMMA) films: the lower molecular weight PMMA 495k A2 as the bottom layer, and the higher molecular weight PMMA 950k A2 as the top layer. During development after exposure, the lower molecular weight PMMA 495k A2 layer on the bottom develops faster than the higher molecular weight counterpart on top, leading to a small undercut, as shown schematically in Fig. 5.7(b). This prevents deposited materials from sticking to the

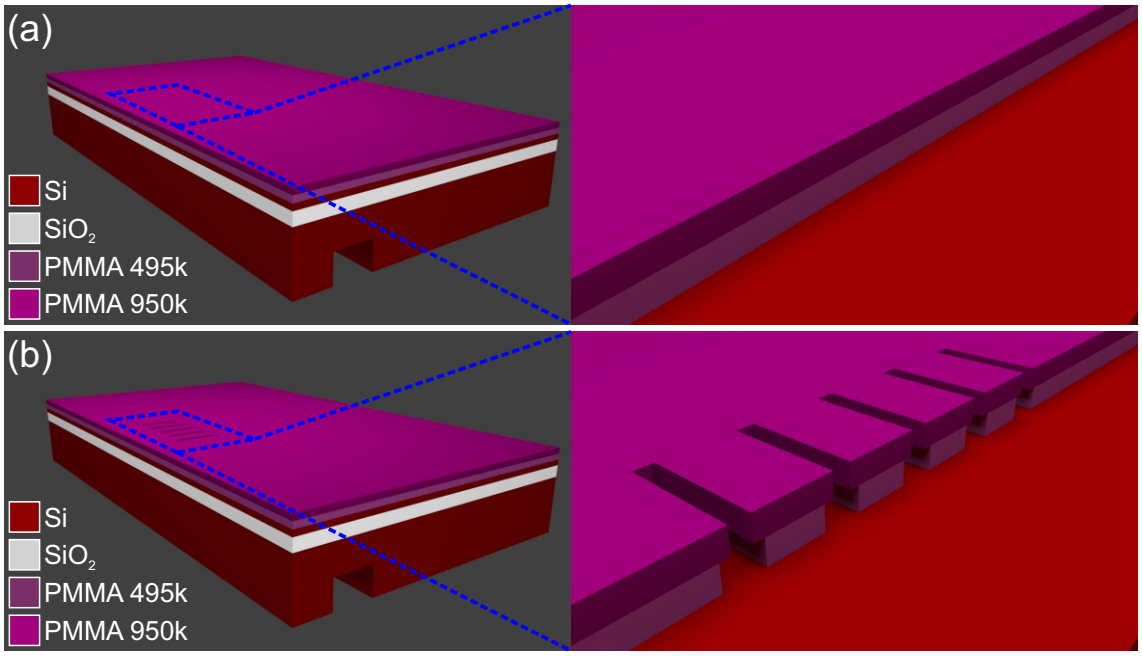

Figure 5.7: Schematic illustration of the EBL process. (a) The bilayer PMMA 495k A2/PMMA 950k A2 is spun onto the substrate. (b) A pattern for the Si core masks is exposed and developed. The enlarged regions on the right depict a cross section of the resist profile for clarity. Images are not to scale.

resist sidewalls, and drastically improves lift-off yield.

The spin process for the bilayer was as follows: First, the die were dehydrated on a hotplate at 180°C for 5 minutes to drive off moisture on the surface and promote adhesion. During a 2 minute cooldown period, a small piece of adhesive tape was applied to the back of the die, which acted to cover up the score marks, allowing the chip to be held onto the spinner's chuck via vacuum pressure. A couple of drops of PMMA 495k A2 were deposited on the sample, and the resist was subjected to a two stage spin process. To first spread the resist, the die was spun at 500 revolutions per minute (RPM) for 10s (which includes a 5s ramp at 100RPM/s). Second, the resist was spun at 4000RPM for 45s (which includes a ramp at 400RPM/s). Following this, the sample was removed from the spinner, the adhesive tape on the back was detached, and the sample was soft-baked for 10 minutes at 180°C to drive off moisture and solvent from the resist. The adhesive covering the scores was reapplied before a few drops of PMMA 950k A2 was applied to the sample, and the sample was spun

with the same parameters from the previous step. After the removal of the tape, a second soft-bake for this top layer was performed at 180°C for 10 minutes, and the die were allowed to cool for 2 minutes.

Thicknesses of the various films were measured on test chips during the spin process. The PMMA 495k A2 bottom layer spun on Si was found to be 60-65nm thick, while the PMMA 950k A2 layer itself spun on Si was measured at about 75nm. However, the total bilayer thickness, measured on the SOI die, was found to be in the range of 110-120nm. This implies that the PMMA 950k A2 top layer was thinner on the bilayer than it was on the bare silicon, which is likely due to a difference in the adhesion of the film to bare silicon (from the test measurement) versus the adhesion of the film to PMMA 450k A2. On the SOI chip, the thickness of the PMMA 950k A2 layer in the bilayer was 50-55nm.

Following the application of the bilayer resist to the die, a pattern defining the waveguide cores had to be exposed into the resist. The actual feature size achieved during EBL is typically not the same as the designed feature size. One of the main factors that influences this is the dosage applied to the sample. Applying too high of a dose will overexpose the resist, resulting in larger feature sizes, while too small of a dose will underexpose the resist, and the features won't be fully formed. As such, extensive dosage and feature size testing had to be performed to be able to effectively correlate design feature size with actual feature size. One exemplary dosage and feature size test is depicted in Fig. 5.8(a). Here, the pattern was exposed with a beam current of 48pA, an accelerating voltage of 10kV, and a 15μm diameter aperture. In this test, the dosage was varied between 200μC/cm^2 and 400μC/cm^2, while the designed line size was varied between 70nm and 200nm. Each linewidth was measured, and the resultant attained feature sizes are plotted in Fig. 5.8(b) as a function of the dosage for each design size, and in Fig. 5.8(c) as a function of the design feature size for each dosage. From the SEM image in Fig. 5.8(a), every feature

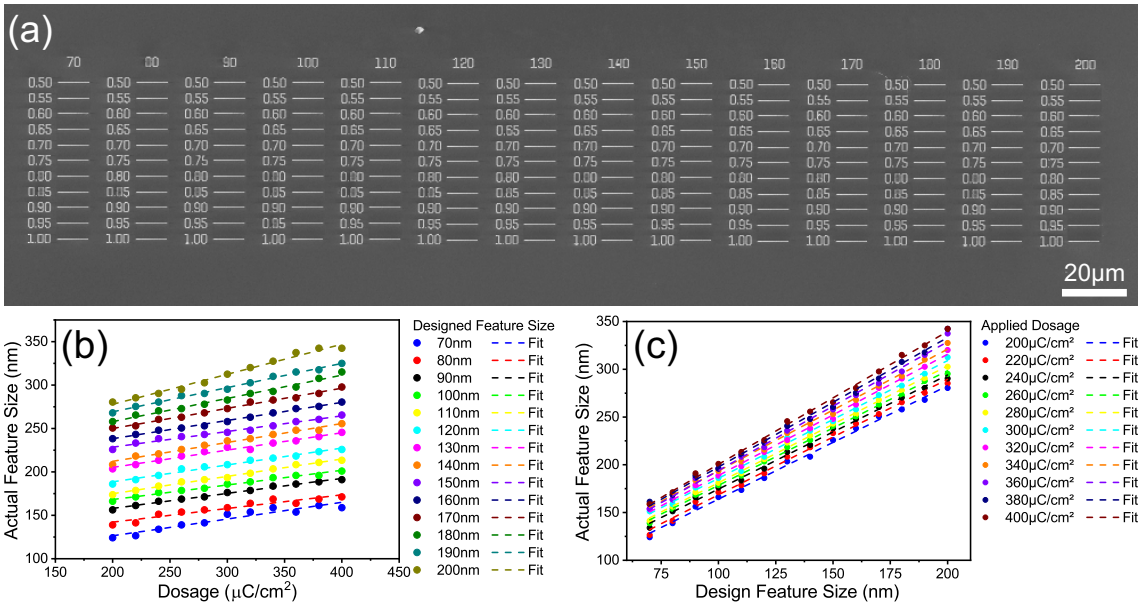

Figure 5.8: (a) SEM image of one exemplary exposure dosage and feature size test performed to calibrate the EBL process. Each exposed line is 10μm long. (b) Plot of the actual feature size as a function of the applied exposure dosage for different design feature sizes. (c) Plot of the actual feature size as a function of the design feature size for different exposure dosages.

was fully formed. None of these applied dosages represents an under exposure, so the lowest dose of 200μC/cm^2 was selected for forming the waveguide mask. From the data collected, the design size could be calibrated as required to achieve the desired feature size.

With the calibration complete, patterns that would define a mask for the Si cores of the ME nanoplasmonic waveguides could be formed. The designed pattern consisted of arrays of 9-11 rectangles, each spaced 5μm apart and designed to produce the desired core width. This spacing between them is sufficiently far that cross-talk will not occur between the waveguides, especially since they will be metal clad. This spacing is larger than the focal spot of the high NA objective lens (with an estimated focal spot radius of $w_0 = 710$nm for a 50×, NA=0.85 objective), so only one nanoplasmonic waveguide can be excited at a time. Eight of these arrays are designed over the die, each intersecting a location where a beam will be formed.

The patterns for the core masks were designed to be 50μm long. This is sig-

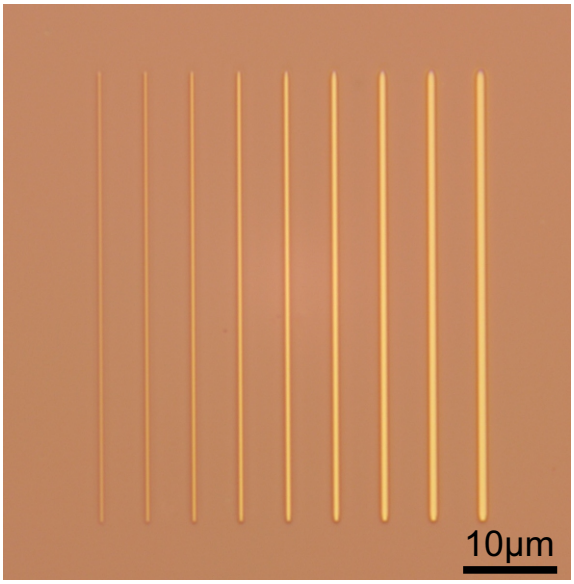

Figure 5.9: Optical microscope image of the pattern transferred into the bilayer PMMA resist via EBL exposure and development for an array of nine Si core masks ranging in width from w_{mask} = 190nm to w_{mask} = 875nm.

nificantly longer than the intended sub-10μm width of the beam on which they are situated; however, this is done to ease alignment tolerances in subsequent steps, and the excess masking material is removed in a later stage. Since EBL inherently has a very high alignment accuracy, the pattern is aligned precisely with reference to one corner of the chip on the same edge taken as the reference during dicing. Ensuring this fact would align the exposed pattern on the top side directly over the scores placed in the backside. As well, since the waveguide cores are designed here to be equal to the width of the dicing saw blade, they should completely cross the location of the scores.

The pattern was exposed with a dosage of 200μC/cm^2, and raster scanned with a 4nm step size at a speed of 6mm/s, resulting in a dwell time of 0.374μs per step. At this rastering rate, each array was exposed in about 16s, resulting in all eight devices on a die being exposed in just over 2 minutes. These exposure parameters allowed for high throughput EBL.

Following exposure, the pattern was developed in three stages. First was a 60s

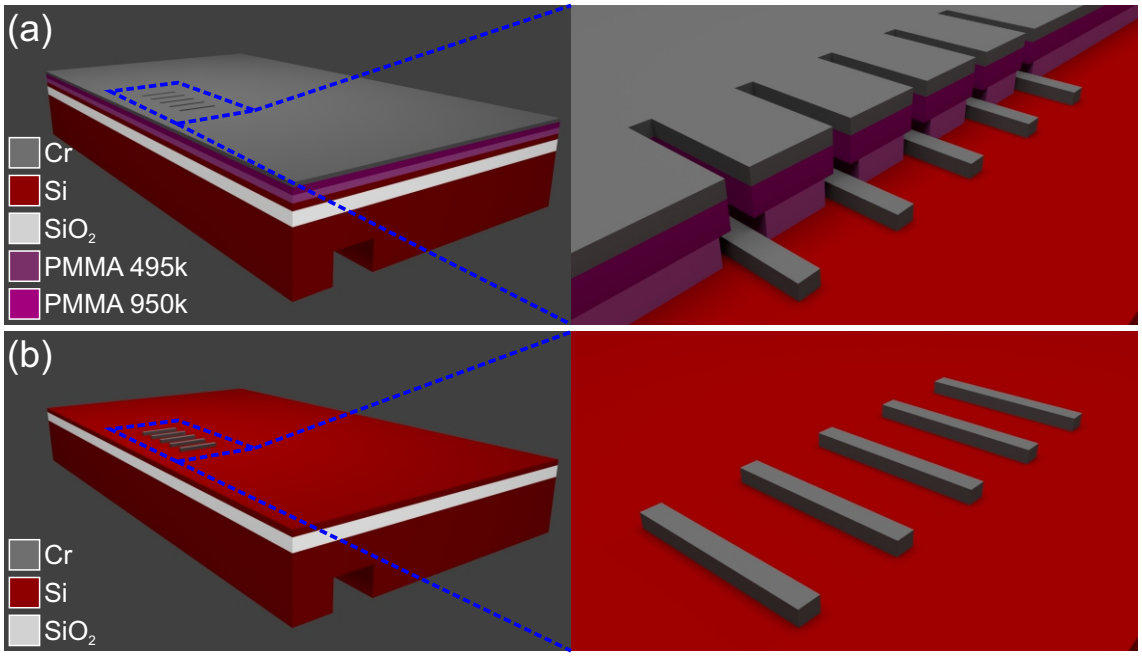

Figure 5.10: Schematic illustration of the lift-off process. (a) A 50nm layer of Cr is evaporated onto the samples. (b) The samples are processed in an ultrasonic acetone bath to remove the resist and transfer the pattern. The enlarged region for (a) depicts a cross section of the resist/Cr profile for clarity. Images are not to scale.

bath in a 1:3 mixture of methyl isobutyl ketone (MIBK):isopropyl alcohol (IPA), followed by a 20s bath in IPA, and a 15s bath in DI water. After drying with N_2, the samples were examined to ensure that the waveguide core array was fully developed. An optical microscope image of one exemplary array is depicted in Fig. 5.9, where it is clear that the array pattern is fully developed.

Metalization and Liftoff

The next step was to transfer the lithographically defined pattern into a Cr film to act as the etch mask for the Si waveguide cores. The patterned chips were placed in an electron-beam evaporator, and a 50nm thick Cr film was evaporated onto the samples, as schematically depicted in Fig. 5.10(a). The 50nm film is thinner than the bottom layer of the bilayer resist, which dramatically improves the lift-off yield. Evaporation was performed at a base pressure of 8.3×10^{-7}Torr, and with a current of 10mA, yielding deposition rates that varied between 0.6-1.0Å/s.

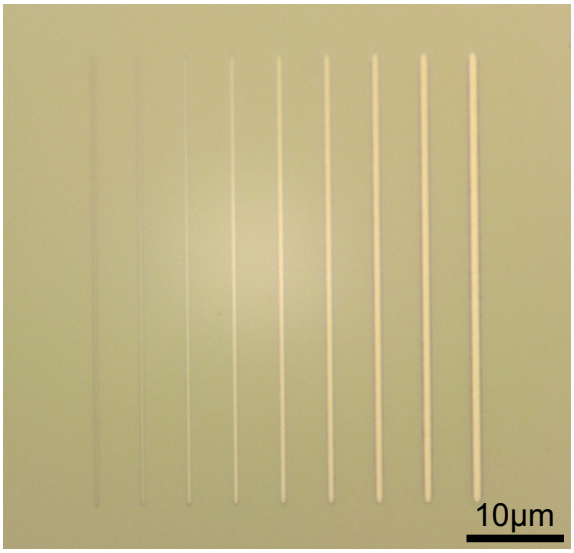

Figure 5.11: Optical microscope image of the 50nm thick Cr lines patterned onto the SOI substrate via liftoff. These Cr lines define the future RIE masks for the Si waveguide cores, with widths ranging from $w_{mask} = 190$nm to $w_{mask} = 875$nm.

With the samples metalized, lift-off was performed, as schematically depicted in Fig. 5.10(b). The chips were first placed in an bath of acetone under ultrasonic agitation for 1 minute, after which, the majority of the Cr had been removed. The samples were placed in a second ultrasonic acetone bath to prevent any of the large Cr flakes in the first bath from sticking to and contaminating the sample, and agitated for an additional 25 minutes to ensure complete lift-off. Subsequently, the chips were placed in an ultrasonic IPA bath for 25 minutes, followed by an ultrasonic bath of DI water for 15 minutes to guarantee that the samples were sufficiently clear of resist and residual particulate. An optical image of one exemplary array after the lift-off procedure was completed if depicted in Fig. 5.11. Clearly, the 50μm long Cr mask lines are well defined, and this lift-off process consistently produced a 100% yield.

5.4.3 Characterization Beam Fabrication

Photolithography Development

At this stage, the SOI chip contains Cr lines that act as a mask for the nanoplasmonic waveguide core definition. The next step in the process was to define a photoresist mask for the beam structure itself via photolithography. However, this presented some challenges that needed to be overcome before devices could be fabricated. First, as this process was intended to mask for several different plasma etch processes required to etch through the Si device layer, BOX, and deep into the Si handle, an appropriate resist thickness was required. From the nominal selectivity specifications of the RIE tools available, the estimated photoresist thickness required was around 1.9μm. From the photoresists available for use in this process, HPR 506 provided the closest match, as it nominally creates films of about 2.2μm thick. However, the actual etch selectivity obtained in practice varies based on a number of factors, including conditions in the chamber, substrate size, substrate type, exposed area, feature size, etc. The spin process for the HPR 506 photoresist had to be modified to produce a thicker film to adequately mask the devices through all of the RIE processes.

The second challenge arises from performing photolithography on a 1cm×1cm die, and is further compounded by the need for a thicker photoresist mask. Photoresists, which are typically applied to larger wafer scale substrates, form much thicker films than EBL resists (several microns versus 50-200nm), which are employed on smaller chip-scale die. This is due to the fact that photoresists are more viscus. When spinning photoresist onto a substrate, the centripetal force causes the resist to spread outward into a thin even layer, as it is pushed to the edges of the substrate. At the edges, surface tension causes it to build up into an edge bead until the centripetal force exceeds the surface tension, and the resist is spun off of the wafer. This edge bead effect causes nonuniformity in the photoresist thickness near the edges of a

wafer, and this effect is amplified when considering substrates with corners. Due to the size of a typical 100-150mm diameter wafer, and the fact that crucial devices are customarily not placed near the wafer edges, this is not commonly an issue. Many photoresist spinners also incorporate components that dispense solvent onto the edge of the wafer as it spins to eliminate this edge bead. When performing photolithography on single 1cm×1cm die, this issue is exceedingly important, as the edge bead can cause nonuniformity over the important device regions. In the present case, the edge bead unintentionally masks the borders of the chip, which after RIE processing, produces a barrier preventing the optical fiber from reaching the waveguides for samples at the edge of the die.

To address these issues, first the spin process was optimized for the single chip lithography. The standard spin process recommended dispensing resist onto the chip, and employing a two stage spinning procedure. The first stage spreads the resist, and the second spins it to the required thickness. The spread was performed at 500RPM for 10s, while the spin stage was recommended to run at 4000RPM for 40s. To optimize this process, the spin time was held constant, but the spin speed was varied between 1000-4000RPM. Two chips were tested in each case, and two measurements of the resist thickness near the center of the chip were taken on each. The results are plotted in Fig. 5.12. From this plot, decreasing the speed from 4000RPM to 2000RPM increased the thickness from 2.2μm to 3.0μm, and the thickness increased rapidly if the spin speed was reduced further. On these square chips, the edge bead manifests as a thick border around the perimeter of the chip, as well as large deposits of resist in the corners. As the spin speed was reduced, the deposits in the corners became larger. In the case of spin speeds of 1000RPM, these deposits encompass nearly a quarter of the chip each. Considering both factors, it was determined that the minimum spin speed, and correspondingly, the maximum film thickness that could be reasonably used to fabricate the devices was 2500RPM, resulting in about

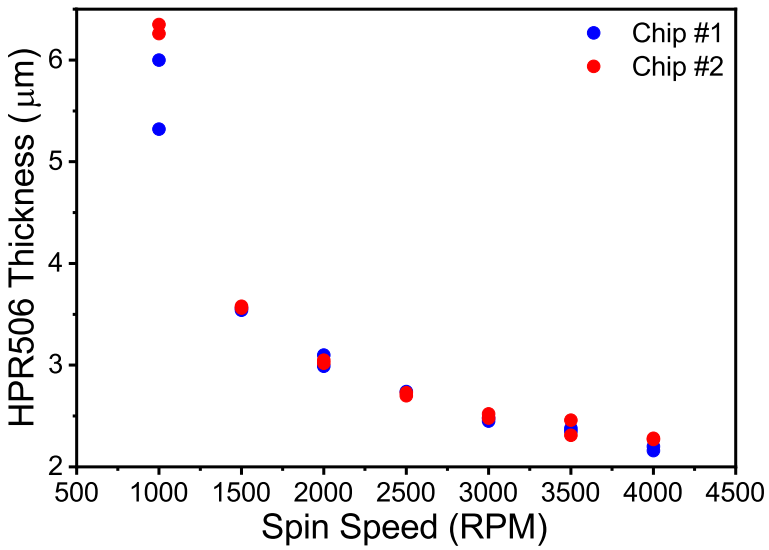

Figure 5.12: Scatter plot of the HPR 506 thickness obtained at the center of a 1cm×1cm die as a function of spin speed. Data was measured at two points on two separate chips.

a 2.75μm thick film. These parameters produced a film that was relatively uniform, except for the region ∼0.5mm from the edge and in the corners. The resultant bead would not cause uniformity issues over the crucial device areas.

The edge bead effect had to be addressed in a different manner. First, the spinners available for this process did not possess the capability of dispensing solvent to remove the resist. Even if this were possible, due to the small size of the die, too much of the resist would be removed. Other methods of removal were attempted, including manual removal of the edge bead with a sharp razor blade or an acetone soaked swab, but neither method was effective. A new approach was taken, and a double exposure method was developed. The concept behind this procedure was to expose only the edge of the chip with a large dosage, and develop to remove the photoresist bead on edges where its masking effect would inhibit the optical fiber alignment. This would also act to reduce the corner resist deposits, which can cause vacuum contact issues during photolithography. To simplify alignment and expedite this double exposure procedure, a unique sample holder was machined, and is depicted in Fig. 5.13. The mechanical drawings for this part are presented in Appendix D.

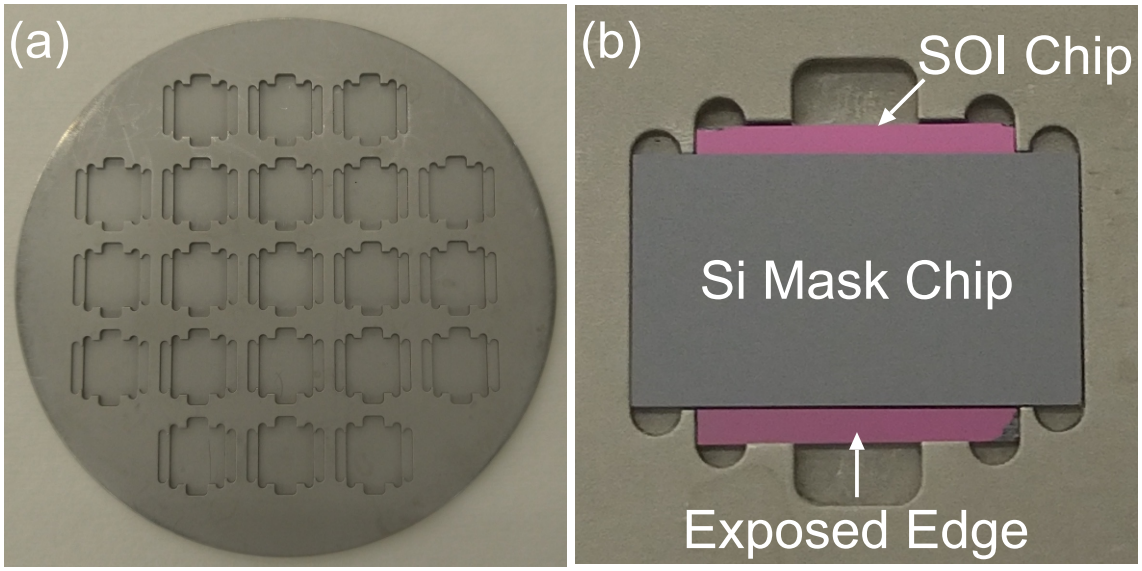

Figure 5.13: (a) Photograph of the custom aluminum holder designed for the double exposure edge bead removal process. (b) Photograph of a 1cm×1cm SOI chip placed in the holder and masked by a 8.0mm×14.0mm Si masking chip. This exposes 1.0mm on the top and the bottom, as well as the four corners for the first, high dosage exposure to remove the photoresist edge bead and corner deposits.

This 100mm diameter holder was designed to sit on the standard 100mm wafer chuck of the optical mask aligner. A square 1cm×1cm pocket was milled 1.0mm deep, for placing the sample die into. A second 0.5mm deep pocket was milled around this, which allowed for placing an 8.0mm×14.0mm Si chip over top of the sample die. This Si chip effectively masks the majority of the sample die, while exposing the region 1.0mm from the edge on two sides. The sample was exposed to an exceptionally high dosage of 1700mJ/cm^2, and overdeveloped in 354 developer for 60s to remove as much of the edge bead and corner deposits as possible. This method effectively removed the edge bead completely, leaving only very small residual resist deposits in the corners. The sample die could then be aligned to the photomask and exposed as normal. Notably, this method removes the edge bead from two sides, and all four corner deposits. The bead on the remaining two sides is actually beneficial in maintaining the chip's structural integrity during the RIE processing.

One final critical consideration for the photolithography process is the alignment

of the chip to the pattern depicted in Fig. 5.2(b). Once more, great care had to be taken to align the reference edge of the chip from the dicing and EBL steps to the bottom of the pattern. Doing this would automatically intersect the pattern for the beam with the Cr mask, and self-align it to the scores on the backside of the chip. Moreover, with the alignment marks depicted in Fig. 5.2, the chip could be aligned to the bottom of the pattern with an accuracy of better than $\pm 5\mu$m. Considering that the backside scores are 50μm wide, and the Cr mask lines are 50μm long, this alignment method all but ensures that the three patterns (the score, the Cr lines, and the photoresist beam) will coincide with each other correctly.

Photolithography to Define the Beams

With these considerations, the photolithography process was performed as follows. First, the samples were baked at 150°C in an oven for 10 minutes to drive off moisture from the surface. This dehydration bake acts to help promote adhesion between the resist and the chip. After a 5 minute cooldown, the chips were placed on the vacuum chuck of the spinner, and a few drops of HPR 506 were applied to the sample before spinning. The spread stage consisted of 10s at 500RPM, while the spin stage was 40s at 2500RPM. This step is schematically illustrated in Fig. 5.14(a). A soft-bake was performed for 90s at 115°C on a hotplate to drive off moisture and solvent from the resist, before allowing the resist to rehydrate for 15 minutes. At this point, the die were placed into the machined holder, and covered with the 8.0mm×14.0mm Si masking chips, taking care to ensure the reference edge was one of the exposed edges. A long exposure yielding a dosage of approximately 1700mJ/cm^2 was applied to thoroughly expose the top and bottom edge of the chip. Subsequently, the chip was developed in 354 Developer for 60s, followed by a 15s rinse in DI water to remove the edge bead and reduce the photoresist deposits in the corners. The chip was then accurately aligned to the photomask and exposed with a dose of 220mJ/cm^2.

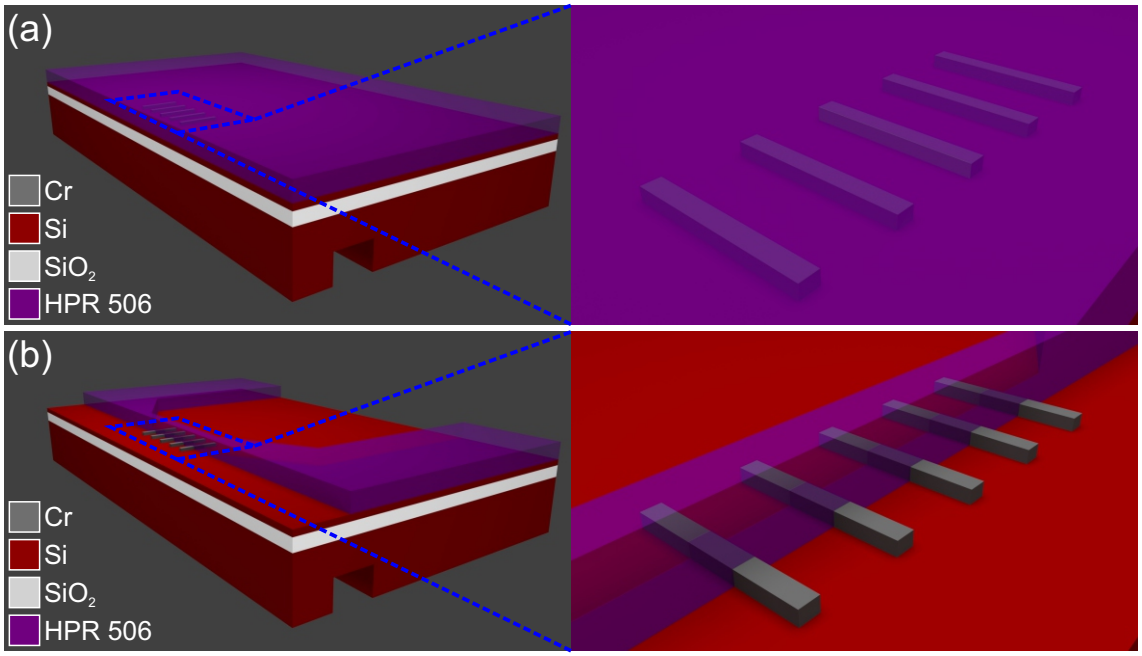

Figure 5.14: Schematic illustrations of the photolithography process. (a) HPR 506 photoresist is spun onto the chip. This covers the Cr mask lines defined previously. (b) After exposure and development, the resist forms a mask for the characterization beams. The Cr mask lines extend from under the resist. Images are not to scale.

The pattern was developed in 354 Developer for 45s, followed by a 15s rinse in DI water. Finally, the photoresist was hard-baked for 120s at 115°C on a hot plate to improve its performance during subsequent plasma etching processes. The final result is schematically depicted in Fig. 5.14(b).

An optical microscope image of one exemplary beam is depicted in Fig. 5.15. Here, it is evident that the photolithography has defined a beam that intersects with the Cr etch mask lines. In the overlapping region, the Cr lines are buried under the HPR 506 beam. This process was effectively used to generate beam widths ranging between 2.0μm and 9.5μm.

Photoresist Descum and Cr Etch

Before the excess Cr could be etched away, it was found that the lithography process left behind a thin residual photoresist layer that need to be removed. This can be seen in the Fig. 5.16(a), which depicts three regions: the photoresist structure, the

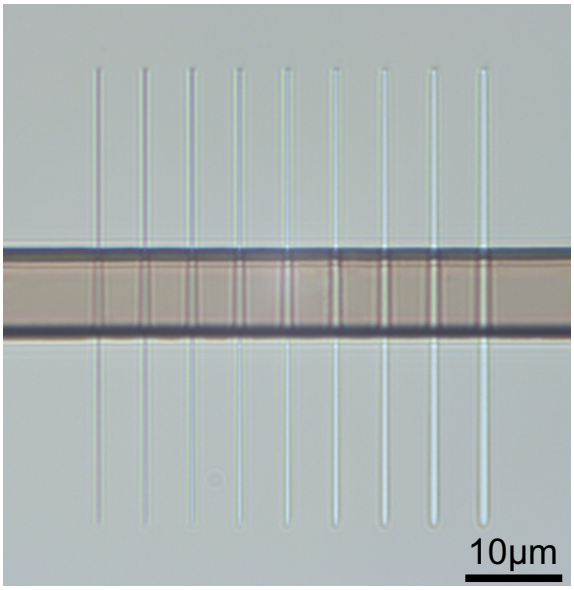

Figure 5.15: Optical microscope of an exemplary 8μm beam formed by photolithography intersecting with an array of 50nm thick Cr mask lines.

edge region which has experienced both exposures, and the patterned region, which was only unmasked during the second exposure. Clearly, a border exists between these regions indicating that there is some residual resist present that needs to be removed. This thin residual resist covers the area defined by the second exposure, and therefore blankets the excess Cr and inhibits its removal. The thickness of this layer is estimated to be < 25nm, as it could not be accurately measured on a profilometer but was clearly visible under a microscope and noticeably inhibited Cr etching during trials. To remedy this, a brief descum process in an O_2 plasma had to be performed. The chips were subjected to 100W O_2 plasma, where the O_2 flow rate was 100sccm and the chamber pressure was 100mTorr for 10s to descum the samples. As depicted in Fig. 5.16(b), this effectively removed the residual resist, while having a negligible impact on the thickness of the photoresist structures.

Application of this brief O_2 plasma removed any residual resist covering the protruding Cr core mask features, and as such, the Cr extending out from under the HPR 506 mask could be wet etched and removed. This process is schematically illustrated in Fig. 5.17(a). The chips were placed in a bath of Cr etchant, which is a mixture of

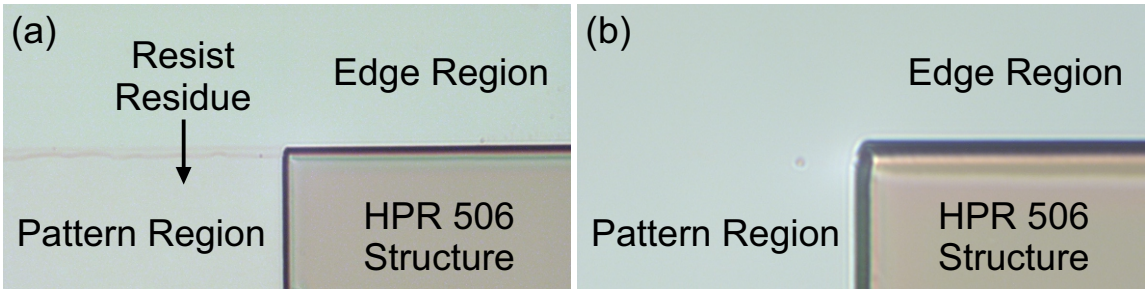

Figure 5.16: Image of the photoresist residue present over the patterned area (a) before the brief descum, and (b) after the descum. A 10s O_2 plasma completely removes the residue

ceric ammonium nitrate ($(NH_4)_2Ce(NO_3)_6$), perchloric acid ($HClO_4$), and water, for 40s, followed by a rinse in DI water for 20s. An optical microscope image of one of the beams following this step is depicted in Fig. 5.17(b). The excess Cr extending out from under the photoresist is removed, leaving only the Cr embedded underneath the photoresist beam remaining. Evidently, this procedure precisely aligns the masks for the nanoplasmonic waveguide cores with the mask for the micromachined beam.

RIE Process Development

At this stage, all of the patterning required to form the characterization beam structure and the waveguides was complete. The designed patterns simply needed to be transfered into the substrate. This would be completed with a six step etching process, which is schematically depicted in Figs. 5.18 and 5.19. To summarize the process before discussing each step in detail: first, the silicon device layer was etched (Fig. 5.18(a)), followed by a RIE of the BOX layer (Fig. 5.18(b)). A DRIE was performed to etch 150μm into the Si substrate (Fig. 5.18(c)). This would break through the scores left by the dicing saw on the backside of the chip, freeing the characterization beam with the waveguides. Subsequently, an O_2 plasma would strip away the remaining photoresist (Fig. 5.19(a)), followed by a second Si RIE to define the waveguide cores (Fig. 5.19(b)). Finally, the Cr core mask would be wet etched away, and the Si waveguide cores would be fully formed on the characterization beam (Fig. 5.19(c)).

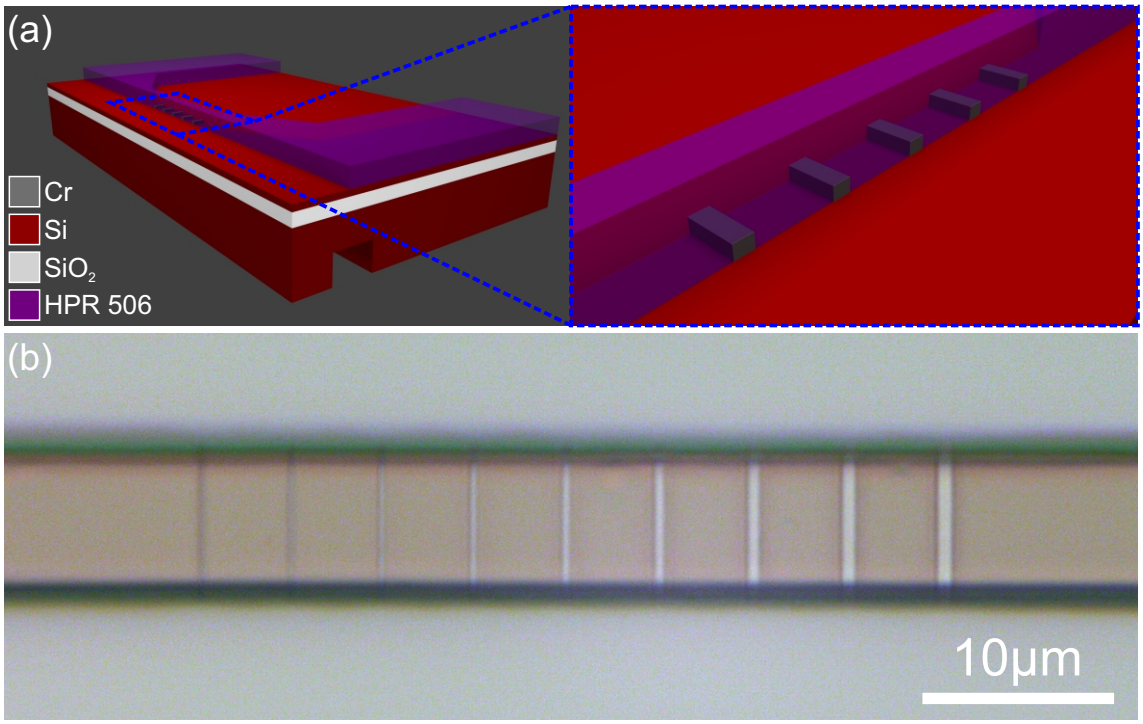

Figure 5.17: (a) Schematic illustration of the device with the excess Cr etched away. The enlarged region shows how the Cr embedded under the HPR 506 beam remains. Image is not to scale. (b) Optical microscope image of one exemplary 8.5μm wide photoresist beam and an array of Cr waveguide masks ranging from $w_{mask} = 190$nm to $w_{mask} = 875$nm. The Cr etch successfully eliminates the protruding Cr.

The first step in the process required etching through the $t_{Si} = 340$nm device layer. Etching silicon is often performed with the Bosch process [202]. Briefly, the Bosch process is an inductively-coupled-plasma reactive ion etching (ICPRIE) process that involves cycling between two different gas plasmas generated above the sample. The first gas is the etch gas, and is typically SF_6, while the second gas is the passivation gas, and is typically C_4F_8. Fluorine radicals and ions generated in the SF_6 plasma etch the silicon both chemically, and through physical sputtering. Applying the passivation gas deposits a thin layer of polymer onto the surface. In the subsequent etch cycle, the physical ion bombardment from the plasma removes the polymer at the bottom of the feature, where the SF_6 plasma can etch the silicon again. However, it does not remove it as quickly from the sidewalls, and therefore a highly anisotropic and directional etch is achieved. As a result of the cycling, this etch procedure typically results in rough,

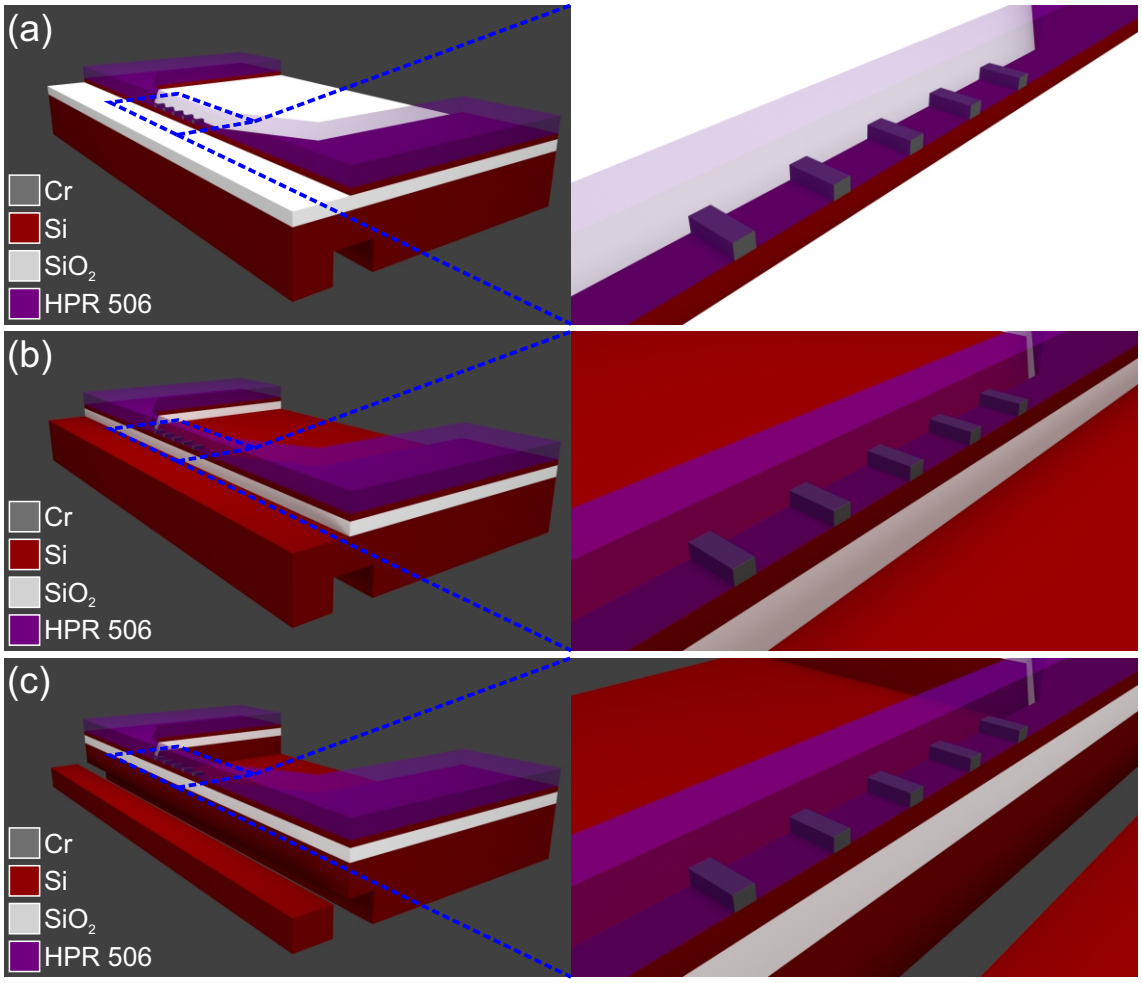

Figure 5.18: Schematic depiction of the first three steps of the etching process. (a) First, an unswitched Bosch etch removes the Si device layer. (b) A SiO_2 RIE removes the BOX layer. (c) The sample is deeply etched via the Bosch process 150μm into the Si handle. This step breaks through into the backside score diced earlier, separating the input side of the characterization beam from the rest of the chip. The enlarged regions depict a close up view of the waveguide formation. Images are not to scale.

scalloped sidewalls. These scalloped sidewalls are not ideal for waveguide formation, as the high surface roughness will cause scattering, significant optical losses, and reduced coupling efficiency. The scalloping can be reduced by decreasing the time between the cycles, and in the limit that the cycle time becomes zero and both gases are applied simultaneously, no scallops are formed. This unswitched Bosch etch is useful for forming straight sidewalls of waveguides.

An unswitched Bosch etch was employed to remove the Si device layer in this first

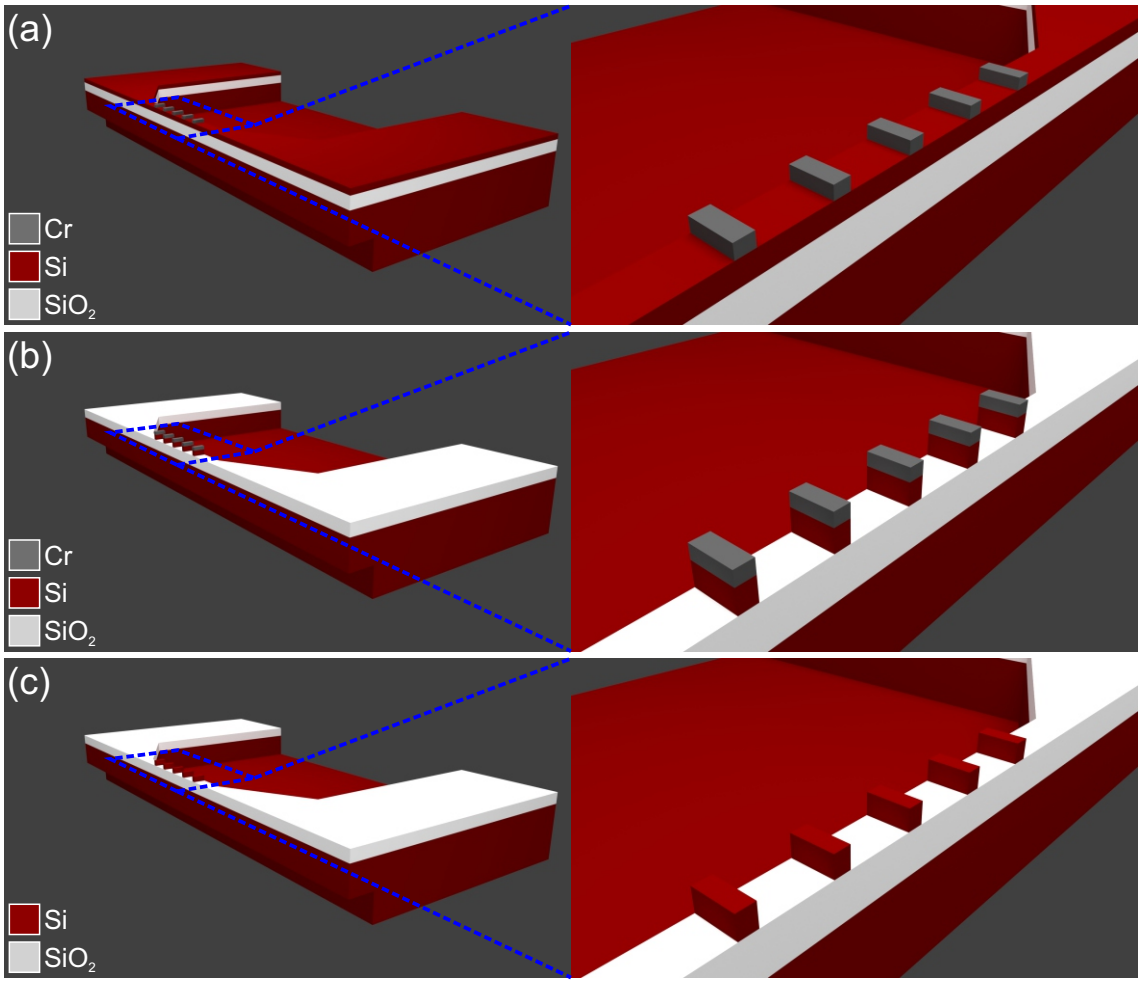

Figure 5.19: Schematic depiction of the last three steps of the etching process. (a) An O_2 plasma removes the remaining HPR 506 photoresist from the surface. (b) A second unswitched Bosch etch removes the remaining device layer, leaving behind the Si waveguide cores masked by the Cr lines. (c) A wet Cr etch removes the Cr layer capping the Si cores. The enlarged regions depict a close up view of the waveguide formation. Images are not to scale.

etch stage. This etch simultaneously applied SF_6 at 30sccm and C_4F_8 at 48sccm, with an ICP power of 350W, a substrate power of 25W, and a chamber pressure of 7.5mTorr. This etch was capable of removing the device layer in 75s for an etch rate of 4.5nm/s.

The second step in the etch process required removal of the exposed $t_{BOX} = 1\mu$m BOX layer. This was achieved with a standard SiO_2 ICPRIE process, whereby 17sccm of C_4F_8 and 15sccm of CH_4 are applied to the chamber, with an ICP power of 2800W, a substrate power of 220W, and a chamber pressure of 9mTorr. The SiO_2 layer

was completely etched within 220s, for an etch rate of about 4.5nm/s. This etch was performed with a 150mm carrier wafer, and the etch time and selectivity were dramatically reduced if a 100mm carrier wafer was used.

The ensuing DRIE process aimed to etch 150μm into the silicon handle to (1) break through the backside score, and (2) etch the region behind the beam for the optical fiber to approach. This phase of the etch process posed significant challenges to forming the desired patterns. First, this pattern requires deep etching of high aspect ratio features (the beam), as well as large open areas. Furthermore, since the edge of the Si device layer is exposed under the photoresist, care had to be taken not to damage this layer significantly. A typical Bosch recipe, having a single stage cycling between C_4F_8 and SF_6 gases was found to be insufficient for this task. While these standard recipes were more than capable of rapidly etching to the required depth, the etch stage was far too aggressive. This resulted in significant undercutting of the Si device layer, as shown in Fig. 5.20(a)-(b). Approximately 13μm of the device layer around the perimeter was removed. In the cases of sub-10μm beams, this completely removed the device layer and rendered the devices useless. The aggressive etch also caused significant damage to the structure's sidewalls, as shown in Fig. 5.20(c), where the signature scallop pattern of the Bosch process is not observed. A more complex DRIE process was required.

In order to alleviate this, an etch process that was specifically designed for high aspect ratio etches was taken as the starting point. After some modifications of the etch and passivation times to improve the etch rate for the present structures, the resultant parameters of the etch are given in Table 5.1. This process is composed of two stages, where the first was repeated 70 times and the second was repeated 100 times, and each stage is comprised of 12 individual steps. These steps act to make the transition between passivation and etching more gradual and provide finer control over the etch process, as well as yielding a process that is less aggressive.

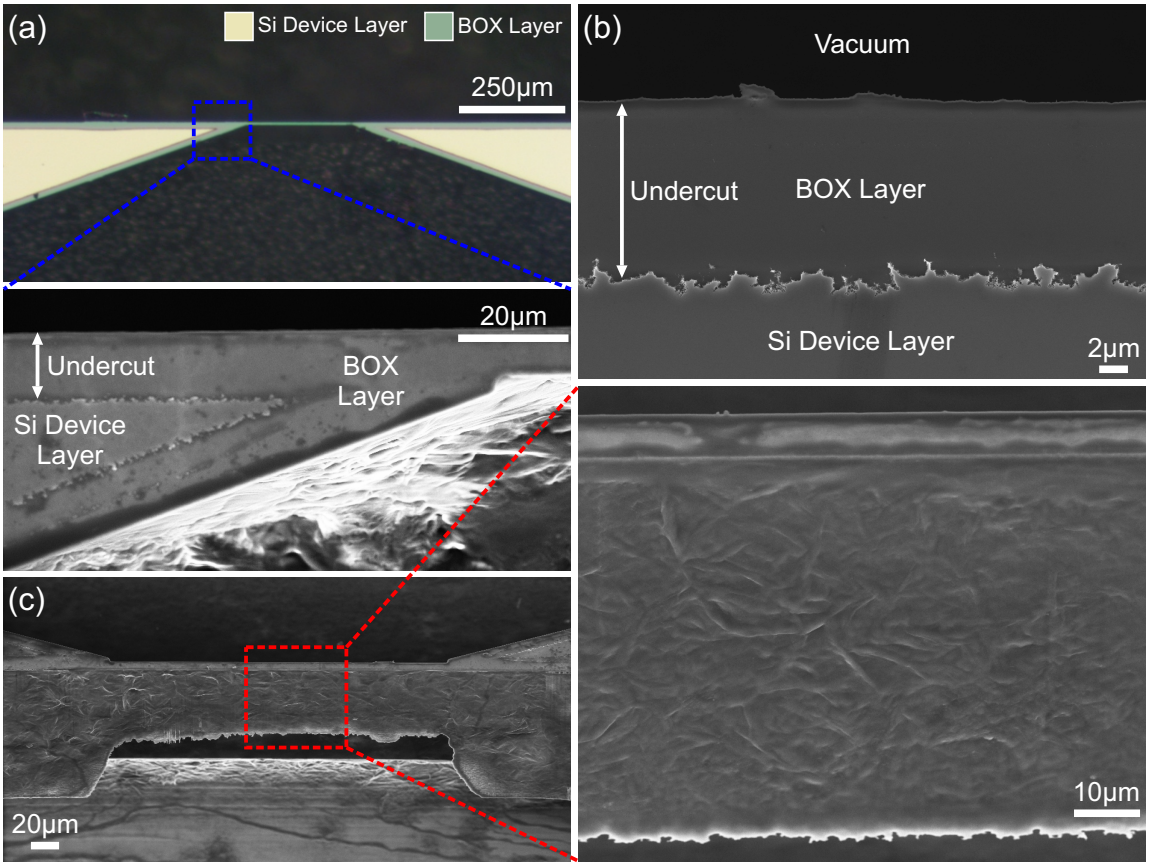

Figure 5.20: (a) Optical microscope image of the device layer undercut caused by the aggressive DRIE process. The enlarged SEM image further depicts the magnitude of the undercut. (b) SEM image of the undercut. Approximately 13μm of the device layer is removed around the perimeter of the device. (c) SEM image of the sidewall profile produced with the aggressive DRIE process. The scallops typical of the Bosch process are not visible, and the sidewall appears severely damaged. In (c), the sample is tilted at a 45° angle.

During each stage, Step 1 is the primary passivation step, while Steps 6 and 10 are the primary etching steps. The other steps act to gradually ramp various parameters to their desired values. The etch phase is divided into two, where the first, having a higher RF power applied to the substrate chuck (see Table 5.1 for exact values during each step), is characterized by more energetic ion bombardment of the substrate, and has a larger physical etch component. This is desirable after the passivation step to remove the deposited polymer. The second etch step, having a lower RF power applied to the substrate chuck but higher ICP power and SF_6 flow rate (see Table 5.1 for exact values during each step), is therefore a stronger chemical etch, for removing

Table 5.1: DRIE Etch Parameters

		Step #	Time (ms)	Pressure (mTorr)	Substrate Power (W)	ICP Power (W)	C_4F_8 (sccm)	SF_6 (sccm)
Repeat 3 Times	Repeat 70 Times	1	175	80	5	1750	280	10
		2	100	80	5	1750	280	160
		3	50	80	5	1750	10	160
		4	75	50	5	2000	10	160
		5	50	30	110	2000	10	160
		6	425	30	110	2000	10	160
		7	25	30	110	2000	60	600
		8	50	80	110	2500	60	600
		9	50	80	5	2500	60	600
		10	700	80	5	2500	60	600
		11	150	80	5	2500	280	10
		12	100	80	5	2250	280	10
	Repeat 100 Times	1	175	80	5	1750	280	10
		2	100	80	5	1750	280	100
		3	50	80	5	1750	10	160
		4	75	50	5	2000	10	160
		5	50	30	110	2000	10	160
		6	465	30	120	2000	10	160
		7	25	30	120	2000	60	600
		8	50	80	120	2500	60	600
		9	50	80	5	2500	60	600
		10	800	80	5	2500	60	600
		11	150	80	5	2500	280	10
		12	100	80	5	2250	280	10

the exposed silicon.

The key difference between the two stages is that the etch steps in the second stage are longer, and thus, this stage will etch deeper. The less aggressive 70 cycle stage is performed at the start of the etch process, where top of the structures is exposed and etched through. Upon implementation, this 170 cycle process was found to etch the sample between $48 - 55\mu$m, and was repeated 3 times to achieve a total etch depth of around 150μm.

As illustrated by the image in Fig. 5.21(a), with this recipe the device layer undercut is not visible on an optical microscope at low magnification, as it was previ-

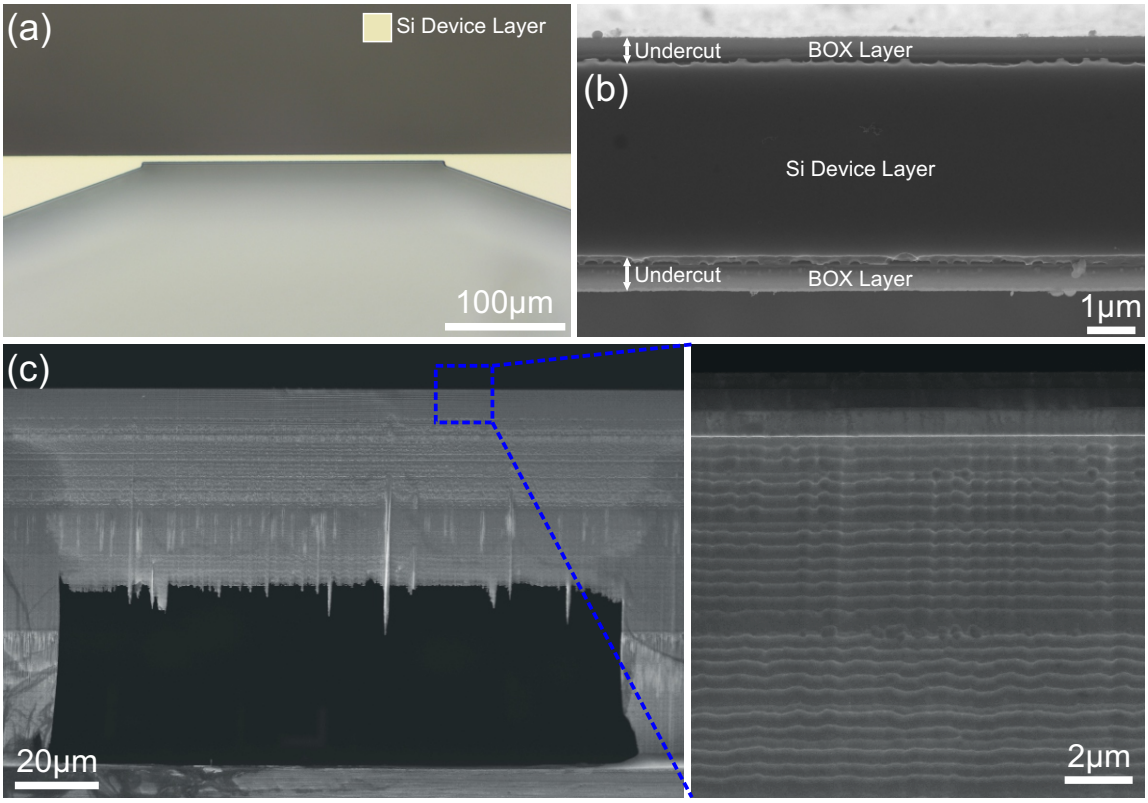

Figure 5.21: (a) Optical microscope image of a beam with the modified Bosch process. The device layer undercut is not visible here. (b) SEM image of an exemplary 5.3μm wide beam. Here, the undercut is reduced to about 600nm on each side. (c) Image of the sidewall profile of the characterization beam. The enlarged region depicts the Bosch scallops present at the surface of the sample. The DRIE undercuts the Si handle after etching about 80μm downwards.

ously. SEM analysis, shown in Fig. 5.21(b) shows that the undercut does still occur; however, measurements on numerous samples show that this undercut to be between 400-900nm (about 600nm undercut is presented in Fig. 5.21(b)), which is a more reasonable value. While this undercutting does cause the edge of the device layer to become rough, this is easily remedied in a subsequent FIB milling step. Figure 5.21(c) depicts the sidewall of the beam, and the enlarged region clearly shows the characteristic Bosch scallop pattern near the surface, with a pitch measured to be about 260nm. The sidewall profile becomes slightly more rough the deeper the etch goes, eventually undercutting the beam completely. However, this occurs 80μm below the BOX, which is a significant distance away from the crucial features. Furthermore, this etch successfully broke through the backside score, allowing the samples to be

separated with the characterization beams correctly positioned at the edge of the samples.

With the DRIE complete, the next step involved the removal of the remaining photoresist mask on the surface of the sample. To achieve this, the samples were again placed in an O_2 plasma for 5 minutes, with identical parameters to the descum process performed earlier. This exposed the Cr core mask lines on the surface of the characterization beam. The samples were subjected to a second unswitched Si RIE for 75s, with the same parameters as the previous unswitched etch step. This effectively removed the device layer everywhere except in the regions defined by the Cr masks, and the Si waveguide cores were constructed directly onto the characterization beam.

Cleaning and Mounting Samples

With the RIE processing complete, the samples needed to be cleaned. First, the Cr masks were etched away in a bath of Cr etchant for 5 minutes. This was significantly longer than the time that should be required to etch through the 50nm Cr mask capping the waveguides, but the additional time was added in order to ensure that all Cr was removed. Next, the samples were placed in two baths of boiling DI water for 5 minutes each to dissolve any of the remaining Crystalbond adhesive that had been used to attach the die to larger carrier wafers for the RIE processing. A bath in acetone for 10 minutes, a bath in IPA for 10 minutes, and a bath in DI water for 5 minutes, followed.

Due to the fragility of the narrow micromachined beams, the individual devices needed to be handled with care. In order to facilitate transfer to the remaining processing tools and characterization setup while minimizing the risk of damage, the samples had to be mounted onto a larger sample holder. A scheme was devised whereby a standard SEM stub was milled down to the exact width of the sample, 5.0mm. The device could be attached to the end of this stub via carbon tape, and the

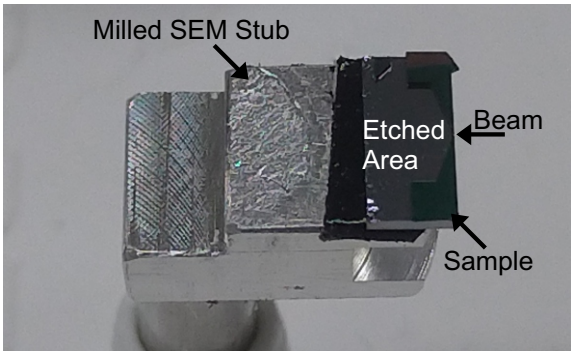

Figure 5.22: Photo of an exemplary sample mounted on one of the milled SEM stub sample holders.

stub itself could be manipulated, reducing the risk of damaging the delicate characterization beam. This arrangement allows the sample to be inserted into SEMs, FIBs, and with the design of some simple mounting components, into the required deposition tools and characterization setups. An image of a mounted sample is depicted in Fig. 5.22, and the mechanical drawings for the part are presented in Appendix D.

SEM images of the completed beams and the Si waveguide cores are presented in Fig. 5.23. Figure 5.23(a) depicts a top down image of one exemplary 5.5μm wide beam containing an array of waveguide cores with widths ranging from 190nm$\leq$ w_{Si} $\leq$875nm. Figure 5.23(b) depicts the same structure tilted at a 45° angle, while Figs. 5.23(c), 5.23(d), and 5.23(e) depict exemplary close up images of the cores for waveguide widths of w_{Si} = 780nm, w_{Si} = 500nm, and w_{Si} = 230nm, respectively. From these images, a couple of key observations can be made. First, the slight device layer undercut is present, and it is about 900nm on this specific sample. The undercut also acted to damage the end facets of the waveguides, but both of these issues are remedied in the ensuing FIB milling step. The waveguide cores are characterized by vertical sidewalls with minimal surface roughness, which is ideal for these optical waveguides as it minimizes scattering and optical losses.

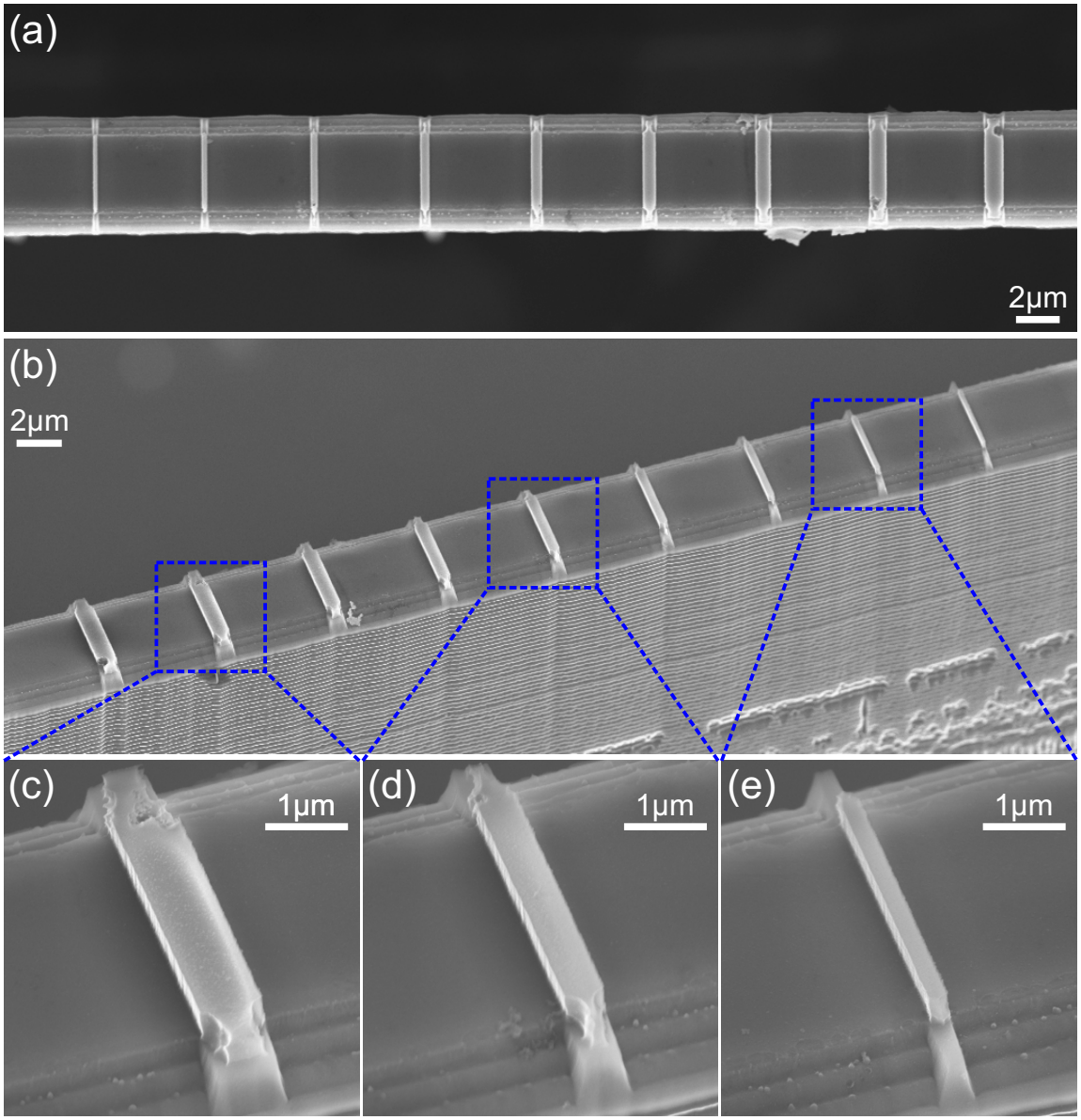

Figure 5.23: (a) Top-down image of an exemplary array of Si waveguide cores on a 5.5μm characterization beam. Widths of the cores range between 190nm$\leq w_{Si} \leq$ 875nm. (b) The same array of Si waveguide cores tilted at an angle of 45°. The enhanced regions depict close up images for Si waveguide cores with (c) w_{Si} = 780nm, (d) w_{Si} = 500nm, and (e) w_{Si} = 230nm. The RIE damage at the end faces can be remedied by a subsequent FIB milling step.

5.4.4 Metalization and Plasmonic Waveguide Formation

Magnetron Sputtering the Au Cladding

In order to convert these Si waveguide cores into ME nanoplasmonic waveguides, two more steps, depicted in Fig. 5.24, must be completed. First, the devices must be clad

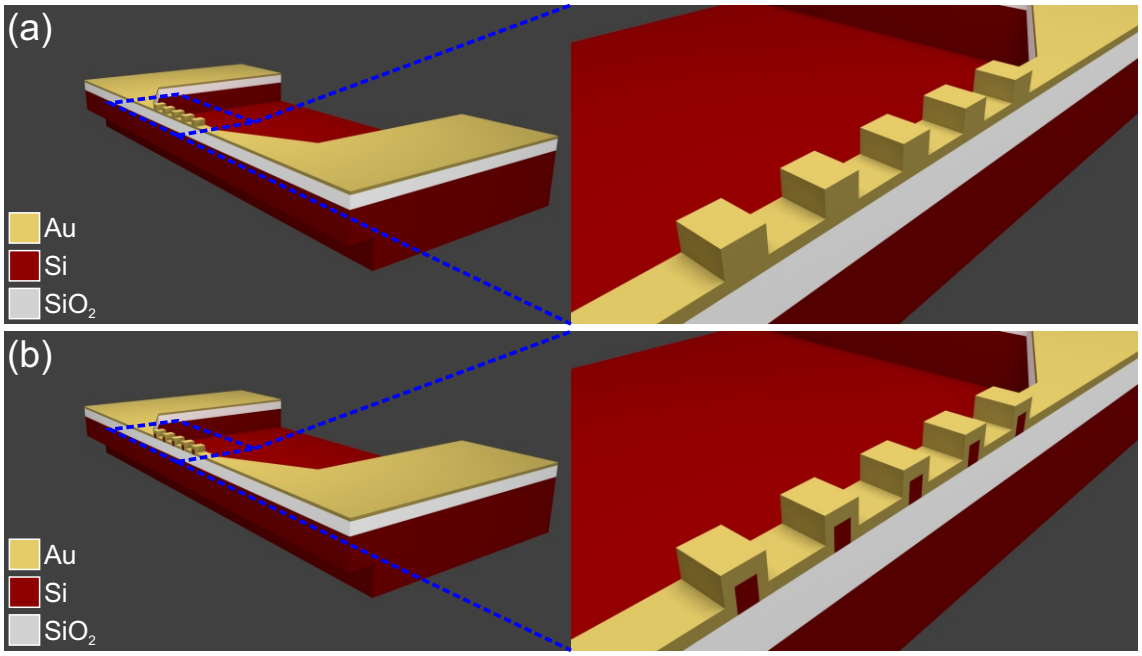

Figure 5.24: Schematic illustration of the plasmonic waveguide formation. (a) Au is sputtered onto the sample, coating the waveguides with excellent sidewall coverage. (b) FIB milling is employed to expose the end facets of the waveguides. Note that for clarity, Au has is only shown as covering the top layer here, whereas in practice, the entire sample is coated in Au. The enlarged regions depict a close up view of the waveguide formation. Images are not to scale.

in a metal, as depicted schematically in Fig. 5.24(a). Here, the metal was chosen to be Au, due to its stability upon exposure to air and low optical losses. While Ag would provide lower waveguide losses, it reacts with sulfur in the air to form AgS_2, which rapidly degrades the optical performance. Materials such as Cu and Al could also be chosen to more closely align with CMOS compatible processes, at the expense of slightly higher losses.

Magnetron sputtering was selected as the method to deposit the Au film, as sputtering is less directional than evaporation, and yields better sidewall coverage. Good sidewall coverage is crucial in this application for the formation of the ME nanoplasmonic waveguides. As such, a 200nm film of Au was sputtered onto the sample at an average rate of 7.6nm/min, under the conditions of a 1.1×10^{-6}Torr base pressure, 7mTorr Ar flow, and 75W of power for the plasma.

A SEM image of the resulting film is shown in Fig. 5.25(a), depicting a very

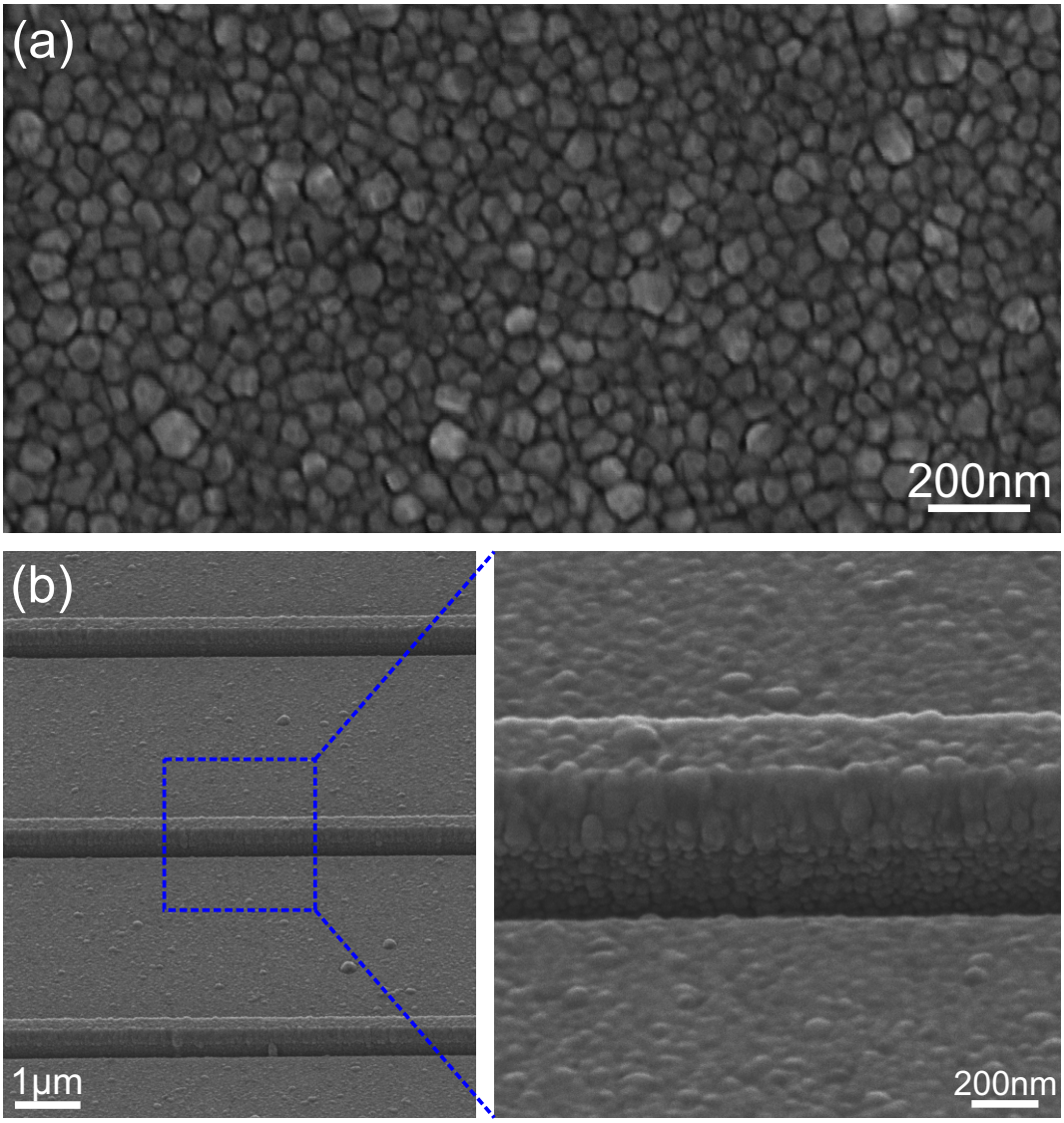

Figure 5.25: (a) SEM image of the Au film deposited onto the device. (b) HIM image of the Au sidewall coverage around the waveguide cores. The enlarged region shows both complete coverage, and the granular structure of the Au cladding. In (b), the sample is tilted at 54°.

granular structure, with grain sizes between 40-100nm. Figure 5.25(b) depicts a helium ion microscope (HIM) image of the sidewall of one exemplary waveguide. The sidewall is completely covered, and ensuing measurements showed the sidewall thickness to be about 100nm, making them optically thick and suitable for ME plasmonic waveguides. The sidewalls exhibit a similar granular structure, which will have two impacts on the ME mode propagation. First, the additional roughness from this film will increases the losses of the optical modes, and second, the added roughness will

also act to enhance the electric fields within the waveguides.

Exposing the End Facets via FIB Milling

Since the sputtered Au film yielded excellent sidewall coverage, naturally the end facets of the Si waveguide cores, which were damaged during the DRIE process, were coated in an optically thick layer of Au. Both defects were easily remedied through the use of a FIB milling to complete the formation of the ME nanoplasmonic waveguides. This FIB milling step is illustrated schematically in Fig. 5.24(b). FIB milling involves accelerating a beam of ions, gallium ions in the present case, and bombarding the sample surface in a highly localized and controlled pattern. As the ions impact the surface, substrate ions are sputtered away, and patterns can be written into the substrate.

The present process required simply exposing the end facets of the waveguide and removal of any damaged silicon. To achieve this, a region 2.5μm wide was milled into both the input and output sides of each waveguide. The distance milled from the edge of the beam into the silicon was chosen manually by inspection, such that it was sufficiently far to remove all damaged silicon and expose a clean end facet. This parameter was held constant for all waveguides on a single beam, such that each ME waveguide on the beam is characterized by the same length. The 2.5μm milled region is larger than the diffraction limited spot size for the high NA objective lens to be used in the characterizations (with a focused beam radius of $w_0 = 710$nm), so this milled region will not inhibit end-fire coupling to the waveguide.

For this process, the gallium ions were accelerated at 30kV, and 100pA of beam current was incident on the sample through a 50μm aperture. This current was high enough to rapidly mill the required areas, while low enough to provide detailed depth resolution. An area dosage of 1nC/μm^2 was sufficient to mill through the Au and nearly all the way through the BOX, generously exposing the waveguide end

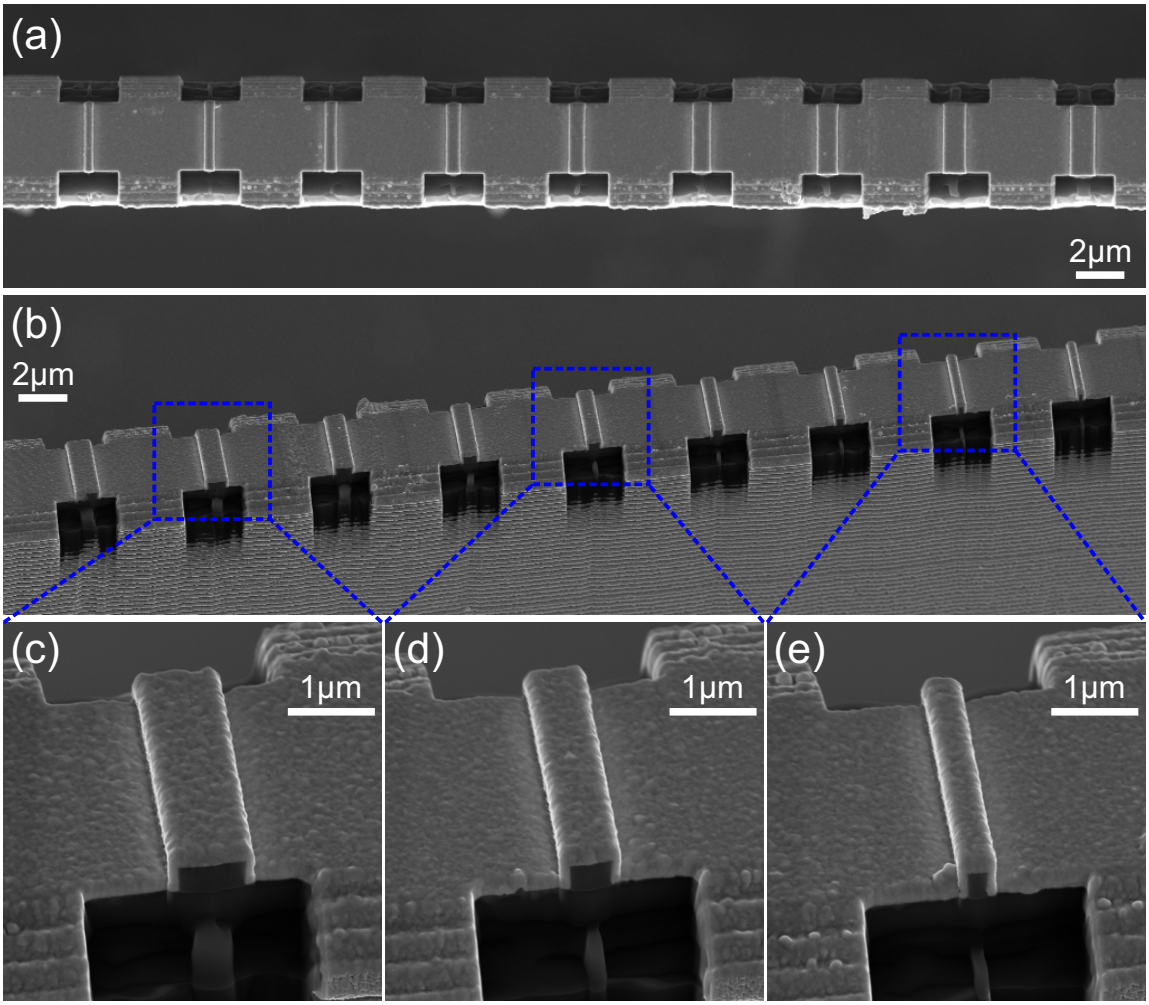

Figure 5.26: (a) Top-down SEM image of an exemplary array of Si-based ME plasmonic waveguides on a 5.5μm characterization beam. Widths of the Si cores range between 190nm$\leq w_{Si} \leq$ 875nm. A FIB has been used to clear the end faces of the waveguides by milling a region 2.5μm wide at the end face of each waveguide. This results in waveguides $L = 3.0\mu$m long. (b) The same array of nanoplasmonic waveguides tilted at an angle of 45°. The enhanced regions depict close up images for ME waveguides with a core width of (c) $w_{Si} = 780$nm, (d) $w_{Si} = 500$nm, and (e) $w_{Si} = 230$nm.

facet in the process. The beam was scanned with a 5nm step size and a 0.2μs dwell time per step. Importantly, the beam was rastered in a double serpentine pattern perpendicular to the waveguides to avoid redeposition of milled material onto the waveguide end facet, which would inhibit coupling light through the devices. With this process, the waveguide end facets were exposed and damaged Si was removed, completing the fabrication of the ME nanoplasmonic waveguides on micromachined characterization beams.

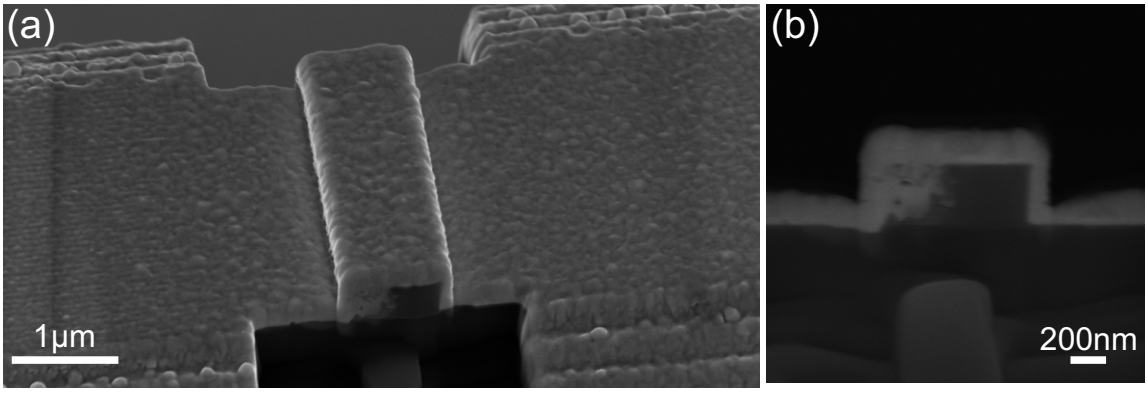

Figure 5.27: Exemplary images of defects caused by inadequate FIB milling of the damaged Si. (a) $w_{Si} = 875$nm, $L = 3.0\mu$m waveguide with defects on the input face. (b) Cross-sectional image of the input facet of the waveguide.

Images of the completed devices are displayed in Fig 5.26. Figure 5.26(a) depicts a top-down SEM image of one exemplary beam, yielding an array of nanoplasmonic waveguides with widths ranging from 190nm$\leq w_{Si} \leq$875nm, and a length of $L = 3.0\mu$m. An image of the same beam tilted at 45° is presented in Fig. 5.26(b), where it is clear that all end faces have been exposed. Figures 5.26(c)-(e) depict zoomed in images of the $w_{Si} = 780$nm, $w_{Si} = 500$nm, and $w_{Si} = 230$nm wide waveguides, respectively. Here, uniform Au coverage of the core along the length of the devices can be observed.

Extreme care had to be taken to ensure that the FIB milled region adequately removed all of the Si damaged from the RIE processing. SEM images of exemplary defects caused by removing an inadequate amount of damaged Si are presented in Fig. 5.27. Here, the $w_{Si} = 875$nm Si waveguide core was not milled back far enough, and the Au filled defects can be observed at the input facet. These defects act to induce scattering and reduce the optical coupling efficiency to the waveguide, and would sporadically occur if one Si core exhibited more extensive RIE damage than its counterparts.

The developed process is extremely versatile in that it facilitates construction of waveguides with dimensions encompassing a vast parameter space. While the

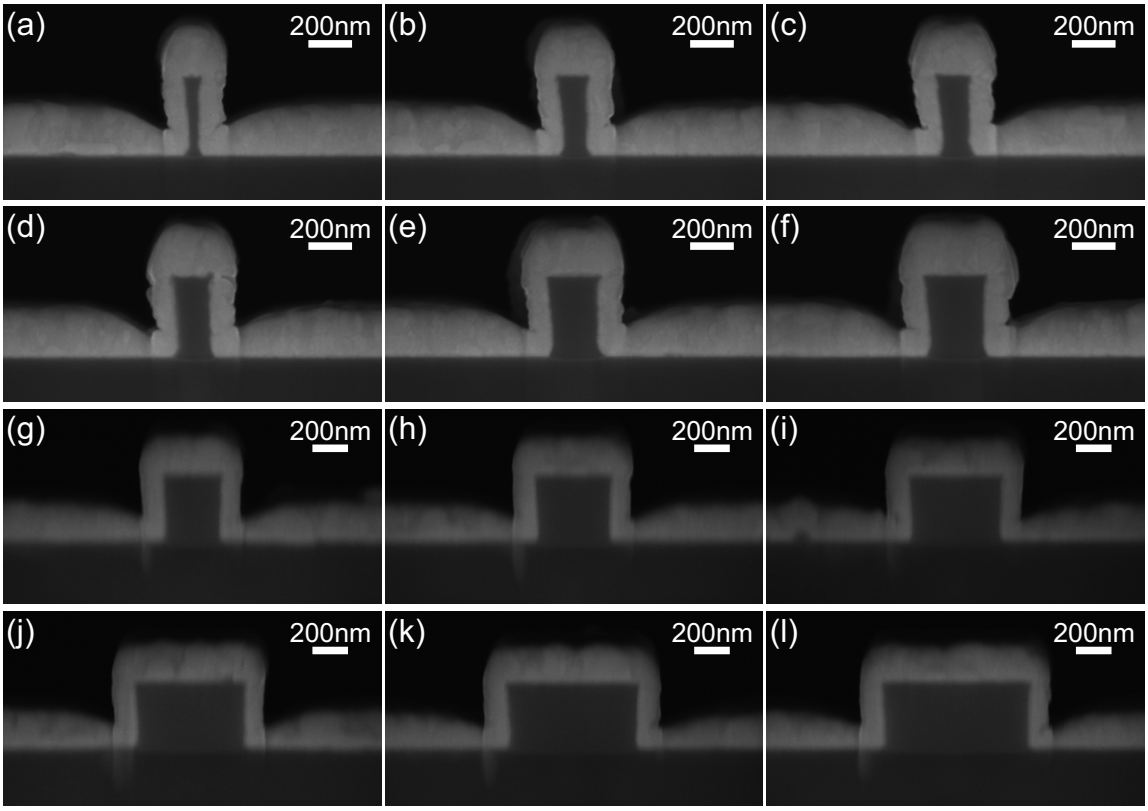

Figure 5.28: Exemplary cross-sectional profiles of the ME nanoplasmonic waveguides fabricated with the process described in this chapter. All Si cores have a height of $t_{Si} = 340$nm. Cross sections are shown for waveguides with core widths of: (a) $w_{Si} = 80$nm, (b) $w_{Si} = 145$nm, (c) $w_{Si} = 160$nm, (d) $w_{Si} = 180$nm, (e) $w_{Si} = 235$nm, (f) $w_{Si} = 275$nm, (g) $w_{Si} = 320$nm, (h) $w_{Si} = 400$nm, (i) $w_{Si} = 500$nm, (j) $w_{Si} = 590$nm, (k) $w_{Si} = 695$nm, (l) $w_{Si} = 780$nm.

photolithography process can define arbitrary waveguide lengths, the EBL procedure can define arbitrary waveguide widths. This process was used to construct waveguides having lengths ranging from $L = 1.7\mu$m to $L = 8.6\mu$m, and widths between $w_{Si} =$ 80nm and $w_{Si} = 875$nm. Figure 5.28 depicts a selection of ME waveguide cross-sections fabricated with this process. In each case, uniform Au sidewall coverage around the Si is evident, with the Au sidewall thickness being about 100nm, or 50% of the sputtered film thickness. This process is robust, and could in the future be used to examine an array of different material systems in the conventionally challenging ME plasmonic geometry.

5.5 Summary

This chapter has detailed the extensive fabrication process development that was undertaken in order to realize Si-based ME nanoplasmonic waveguides within an SOI platform compatible with nonlinear optical studies. The unique challenges presented by the waveguide geometry, integrating nanoscale and microscale structures on the same substrate, and practical considerations for characterizing such structures, required unique approaches and exploitation of every facet of the standard nanofabrication toolkit. The resultant process flow is robust, versatile, and capable of producing devices spanning a wide parameter space. With minor modifications it could easily be extended towards other material systems as well. It is anticipated that the concepts conceived, developed, and advanced for this process will find diverse applications in the fabrication of micro- and nanophotonic devices for both linear and nonlinear optics, as well as for the construction of micro-electro-mechanical systems (MEMS) and other micromachined devices.

Chapter 6

Nonlinear Effects in Si-Based Metal-Encapsulated Nanoplasmonic Waveguides

6.1 Introduction

With the ME nanoplasmonic waveguides fabricated in Chapter 5, the next logical stage was to characterize the various nonlinearities occurring within them. There exists a unique and complex interplay between the various nonlinear phenomena present within a Si-based platform. Moreover, it was essential to examine their potential as on-chip optical light sources via third-harmonic generation, as highly efficient visible light sources currently require the integration of different material systems on-chip. A versatile mechanism of generating coherent visible light locally, in a compact SOI compatible platform, is fundamentally required.

The experimental results presented within this chapter detail the various nonlinear effects that occur within Si-based ME nanoplasmonic waveguides, as well as demonstrate third-harmonic generation within the structures at a high conversion efficiency.

The first section presents a brief overview of harmonic generation within Si-based structures, while the second section describes the experimental setup, calibrations, and simulation methodology. The following sections detail the measurements performed on the waveguides.

6.2 Overview of Nonlinear Light Generation in Si

As discussed in Chapter 2, Si is an indirect bandgap semiconductor, and is incapable of direct light generation via stimulated emission. To produce coherent light within a silicon platform, one must either incorporate different materials on chip, or exploit the inherent nonlinearities of silicon.

Since silicon is centrosymmetric, in the bulk crystal $\chi^{(2)}$ is identically zero. Second order nonlinear effects, such as second-harmonic generation (SHG), are physically prohibited in the bulk. At the surfaces of the crystal however, the symmetry is broken and SHG can occur [203]. As it is a surface effect, nanostructuring the Si with nanopillar arrays has been shown to greatly increase the surface area, and can enhance this SHG [204]. Despite this, the interface-based nature of this effect makes it inherently very weak.

Other studies have examined methods of breaking the centrosymmetry to enhance SHG within silicon structures for use in integrated optics as well. Methods such as straining the crystal with a layer of SiO_2 [205] or by embedding a Si waveguide within a Si_3N_4 cladding [206] were previously accepted as another mechanism of breaking the symmetry and facilitating SHG in Si. Recently though, a number of studies have suggested alternate explanations disproving the idea that strain facilitates SHG. One such work demonstrated that the SHG in a nitride clad Si waveguide was unaffected by the further mechanical stressing, and effectively eliminated upon illumination of the waveguide by ultraviolet (UV) light [207]. Since UV exposure is known to passi-

vate charge defects in at the Si_3N_4-Si interface, the conclusion was reached that the presence of charge defects sets up an internal direct-current (DC) electric field, leading to a process called electric-field-induced second-harmonic generation (EFISHG) [207]. EFISHG is a third-order nonlinear process that is characterized by mixing between two photons at a fundamental frequency and a steady state DC field, producing SHG. EFISHG was directly employed by Timurdogan *et. al.* by applying a voltage across a Si rib waveguide to produce SHG at a conversion efficiency of 13% W^{-1} [208]. However, this scheme required quasi-phase matching, and numerous large electrical contacts to generate this light over the 1.0mm length of the rib waveguide.

A more practical way to generate light within silicon for use within integrated optics is through the inherent $\chi^{(3)}$ of the material. When considering the primary telecommunications wavelength of $\lambda = 1550$nm, the third-harmonic lies in the visible regime, where silicon is strongly absorbing. One approach is to employ nanoscale structures that resonantly enhance the incident fields to improve the third-harmonic generation efficiency, but are small enough to limit the optical absorption. Examples of this approach include Si nanodisks [209], various metasurface designs [210, 211], or plasmonic hybrid Au-Si nanoantennas [212]. The hybrid plasmonic antenna array design was able to achieve high conversion efficiencies of 7×10^{-5} [212]. In general, these arrayed component designs are more suitable for large scale frequency converters, and not integrated, on-chip applications demanded by silicon photonics and plasmonic nanocircuitry.

THG in silicon-based waveguide structures, which are ideal for on-chip integrated applications, have seen sparse experimental demonstration in this regard. One of the earlier demonstrations incorporated a suspended Si photonic crystal waveguide specifically engineered to slow the group velocity of light and enhance both the internal intensity, and the THG [213]. This implementation yielded an 80μm long waveguide that scattered the THG out the top of the waveguide with a conversion efficiency of

1×10^{-7}. This configuration is not ideal for integrated photonic circuits, as the THG is scattered out of plane and is not guided. Photonic crystal waveguides also require strict fabrication tolerances and are inherently only operable over a narrow band. Recently, THG was studied in a much simpler structure: standard SOI nanophotonic waveguides [214]. This study first examined the strong THG emission observed in SOI waveguides with a 340nm×340nm cross sectional area, and lengths ranging between 3.7μm-7.5μm, where it was shown that waveguides 5.2μm long produced the highest THG conversion efficiencies in a Si-based waveguide structure reported to date of 2.8×10^{-5}. The work also investigated THG from longer, 200μm nanophotonic waveguides having widths ranging from 260nm to 4150nm, and showed that significant reshaping of the THG spectrum was observed due to the nonlinear effects manipulating the spectral shape of the pump pulses over this long distance.

Nonlinear light emission was demonstrated at the nanoscale within a plasmonic waveguide architecture consisting of a 95nm wide and 340nm tall Si waveguide core, capped with a 60nm gold layer through two prominent studies [107, 200]. First, it was shown that this compact device, only 5μm long, could produce high THG conversion efficiencies of 2.3×10^{-5}, despite the fact that phase matching could not occur in the structure and the THG was occurring as a localized interaction [107]. It was also demonstrated that this simple nanoplasmonic structure generates broadband white light as well [200]. This was attributed to electrons generated through TPA being ponderomotively accelerated in the steep electric field gradient produced by the surface plasmon mode. As such, these energetic electrons were shown to incite an electron avalanche, with the subsequent particle collisions generating the broadband white light emission.

6.3 Experimental and Simulation Details

This section details the experimental setup, calibration procedures, and simulation methodology employed in characterizing nonlinear effects and THG in ME nanoplasmonic waveguides.

6.3.1 Experimental Setup

The experimental setup employed to characterize the Si-based ME nanoplasmonic waveguides is schematically depicted in Fig. 6.1. Ultrafast laser pulses are generated by an erbium-doped fiber laser (Toptica Photonics FemtoFiber Scientific Laser System). This laser produces optical pulses with a duration of $\tau_p = 84$fs, an average power of 270mW, a pulse energy of $\sim$3nJ per pulse, and a central wavelength of $\lambda = 1550$nm, at a repetition rate of $f_{rep} = 90$MHz. The spectrum of the resultant laser pulses is depicted in Fig. 6.2.

A second CW diode laser, emitting in the red at $\lambda = 652$nm is also employed within the setup for alignment purposes. The CW diode laser is passed through a variable neutral-density (ND) wheel, before reflecting off of mirror M3, while the fiber laser reflects off of mirrors M1 and M2. Here, the red diode laser reflects off of the beam splitter BS1, while the ultrafast laser pulses are transmitted through this component. This allows the red alignment laser and the ultrafast $\lambda = 1550$nm pulses to be perfectly overlapped, and consequently, the ultrafast laser pulse spot can be aligned on the sample visually via the red diode laser. BS1 is mounted on a flip-mount, such that after alignment it can be removed from the path of the incident laser radiation. The red alignment laser and the ultrafast $\lambda = 1550$nm pulses pass through a second ND wheel employed to vary the power within the experiment, before impinging on a 50×, NA=0.85 objective lens (Leitz Wetzlar). This objective lens focuses the incident laser radiation onto the sample, which is mounted on the

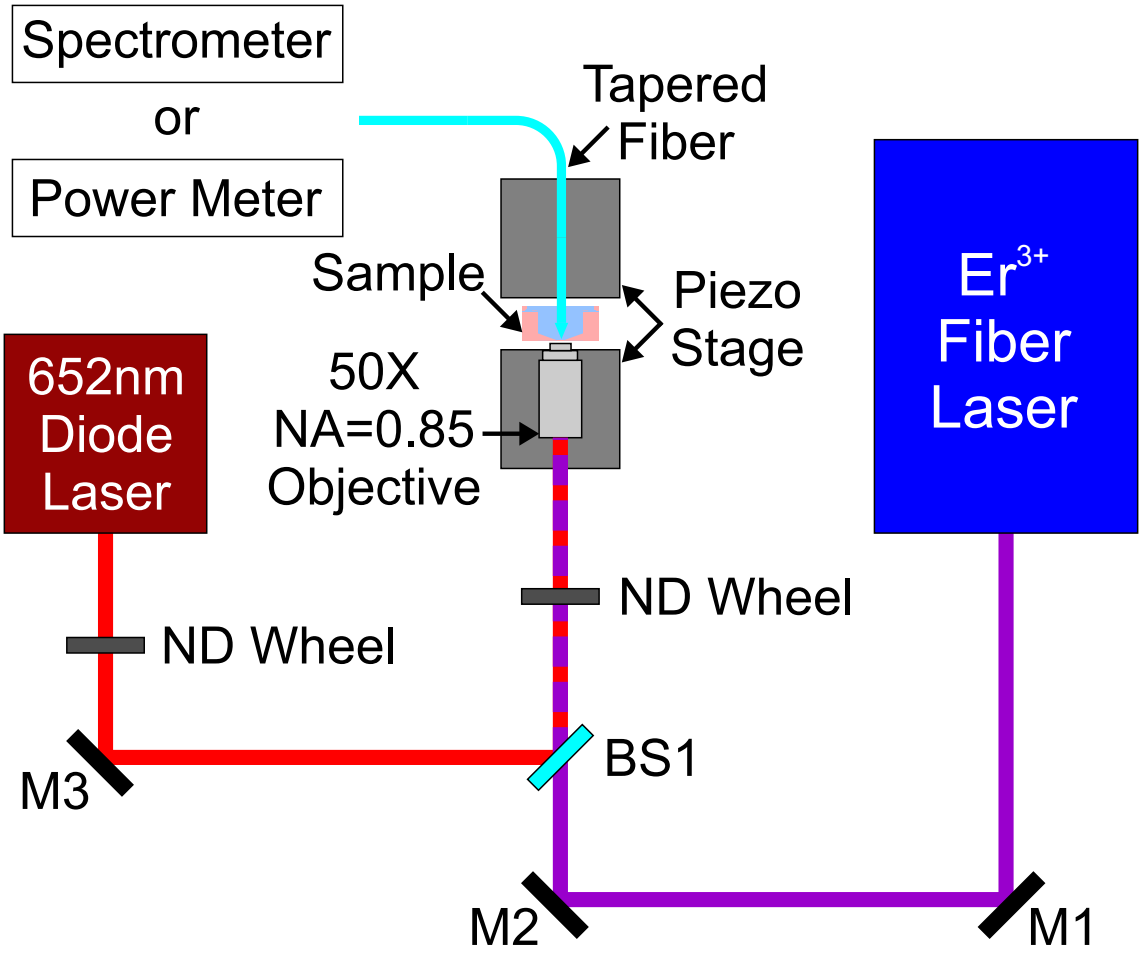

Figure 6.1: Schematic illustration of the experimental setup for characterizing the nonlinear properties of nanoplasmonic waveguides.

milled SEM stubs set into a custom made sample stage (see Appendix D for the mechanical drawings). On the opposite side of the sample, a tapered optical fiber (OZ Optics, Model# TSMJ-3A-1550-9/125-0.25-7-2.5-14-2) collects the transmitted light and routes it to various power meters and spectrometers. Both the objective lens and the optical fiber are mounted on six-axis piezo-controlled stages (Melles Griot NanoMax-HS, Model # 17MAX600R), enabling precise control of their positions via piezo-controllers (Melles Griot APT Piezo Controllers, Model # 17BPC203). A photograph of an exemplary sample mounted within the system is depicted in Fig. 6.3.

A variety of detectors were employed to conduct the experiments. First, the exci-

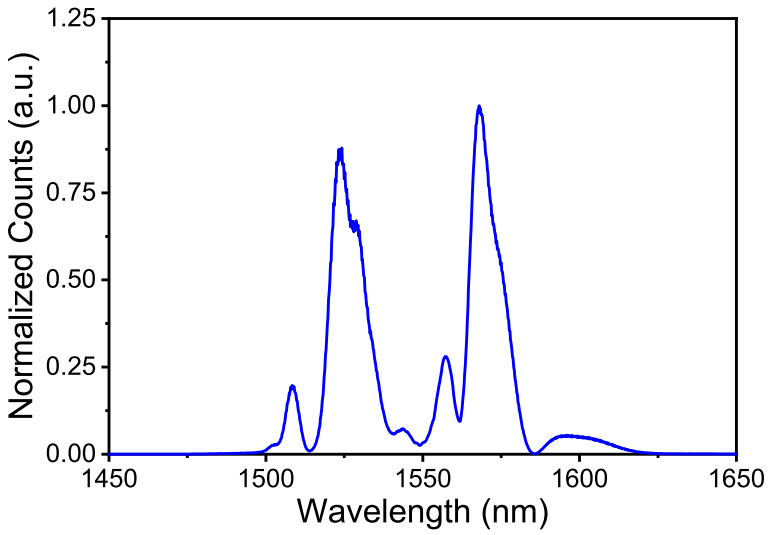

Figure 6.2: Measured spectrum of the ultrafast laser pulses generated by the erbium-doped fiber laser used to excite the nanoplasmonic waveguides.

tation and nonlinear emission processes were monitored and scattered light from the waveguide end facets was imaged from above with a USB camera (AmScope MU1003), through 10× and 20× long working distance objective lenses (Mitutoyo). To measure the spectra, two different spectrometers were required: a Yokogawa AQ6370C Spectrum Analyzer was employed for infrared (IR) signals, while an Ocean Optics USB2000 spectrometer was employed for visible spectrum measurements. Optical power measurements were executed with a Thorlabs PM100D power meter, with a number of different photodiode sensors. A Thorlabs S122C power sensor was employed to monitor the power entering the objective lens, while Thorlabs S154C and S150C fiber photodiodes were utilized to measure IR and visible signals, respectively.

6.3.2 Spectrometer Calibration

For the measurements presented in this chapter, it was noted that the Si fiber power meter could not be used to determine the THG power within the fiber. Since the fiber also collects transmitted and scattered fundamental light, this would induce TPA within the detector, introducing error within the measurement. Absolute power

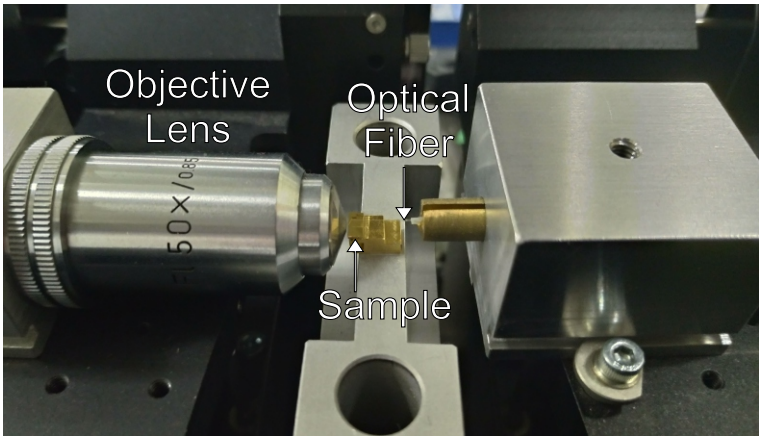

Figure 6.3: Photograph of an exemplary sample in the characterization setup. Ultrafast laser pulses are focused onto the waveguide end facet via the objective lens, and the transmitted light is collected by the tapered and lensed optical fiber.

measurements had to be obtained from the spectrometer, which measures only the number of photon counts. In order to extract the actual third-harmonic power generated, it was necessary to determine the relationship between the photon counts and the corresponding THG power in the green spectral region for the specific spectrometer. To accomplish this, the experimental setup depicted in Fig. 6.4 was employed. Two laser diodes were utilized: one emitting in the red, at $\lambda = 652$nm, was used to ensure alignment, and one emitting in the green, at $\lambda = 532$nm, was used to generate a signal for calibration. Note that the green light emission originated from a $\lambda = 1064$nm diode and a frequency doubling crystal, as is typical in commercial diode lasers to generate green light. The output of the red laser was aligned to reflect off of a beam splitter (BS1 in Fig. 6.4), while the output from the green laser was first passed through a piece of BG39 glass to attenuate any of the $\lambda = 1064$nm radiation present, before being transmitted through BS1. Light from the two diode lasers was overlapped, and then passed through a fixed polarizer (P1) and a second polarizer (P2) affixed in a rotational mount, before entering the objective lens. The angle between the two polarizers, θ_p, was initially set to $\theta_p = 0^\circ$ to ensure maximum transmission, while the objective lens focused the light onto the end of the same tapered lensed fiber used in the THG experiments. The incident power of the lasers could be

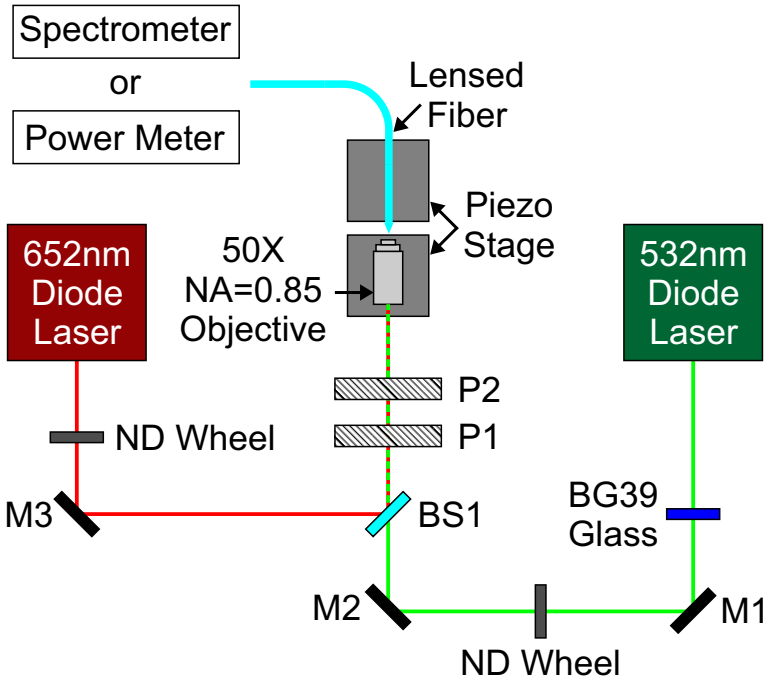

Figure 6.4: Schematic of the experimental setup employed to calibrate the spectrometer photon counts to absolute power values.

adjusted with the use of variable ND wheels inserted into the respective paths.

The power in the fiber, P_{fiber}, required to saturate the spectrometer at $\lambda = 652$nm was $P_{fiber} = 375$pW. With the green laser turned off, the power at $\lambda = 652$nm was adjusted to this value, and the coupling between the fiber and the spectrometer was optimized until the detector was saturated. This alignment method was utilized in the calibration experiment to be consistent with the method used to optimize the fiber-spectrometer coupling in the THG experiments. Subsequently, the red laser diode was switched off, and the green light was turned on. The neutral density wheels were adjusted until the spectrometer signal at $\lambda = 532$nm was just below saturation. Spectra were then recorded as the angle of P2 was increased from $\theta_p = 0°$ to $\theta_p = 90°$, relative to P1. Each spectrum was captured with in integration time of $\tau_{int} = 293$ms, and was averaged over 10 scans, consistent with the settings of the THG experiments.

Next, the fiber was connected to the power meter and the absolute values of the $\lambda = 532$nm power were recorded as the angle of P2 was increased from $\theta_p = 0°$ to

$\theta_p = 90°$. It was found that this crossed polarizer method varied the $\lambda = 532$nm power in the fiber between 0pW $\leq P_{fiber} \leq$ 645pW, where $P_{fiber} = 645$pW corresponded to the spectrum measured just below the spectrometer's saturation level. The fact that the spectrometer became saturated at $\lambda = 652$nm at lower input power than at $\lambda = 532$nm agrees with the fact that a silicon detector has a higher responsivity to red light than it does to green light.

The measured $\lambda = 532$nm power as a function of θ_p is plotted in Fig. 6.5(a). In a similar manner, the corresponding total number of counts registered on the spectrometer above the noise were calculated and plotted as a function of θ_p in Fig. 6.5(b). As expected, both sets of data agree well with the $\cos^2(\theta_p)$ variation predicted by Malus's Law for the transmission through two crossed polarizers. The total counts registered on the spectrometer were plotted as a function of P_{fiber}, yielding a linear relationship as depicted in Fig. 6.5(c). A linear fit to this data reveals a slope of 13.9 counts/pW, or 72.1fW/count. These calibration factors were used to determine the measured power from the spectra measured in the THG experiments.

6.3.3 Simulation Methodology

To validate the experimental results contained within the ensuing sections, fully vectorial 3D FDTD simulations were performed. These simulations were performed with the commercial software, *Lumerical FDTD Solutions* [119]. However, as examined in Section 2.4, Si exhibits a number of nonlinear optical effects and complex carrier dynamics, which are not straightforward to implement within FDTD, and are not standard features of any commercial FDTD product. In order to validate the experimental results presented within this chapter, a custom, in-house developed material plugin for the FDTD algorithm had to be developed to accurately describe the intricate interplay between all of the various nonlinear phenomena, and to precisely model ultrafast pulse propagation through the Si-based nanoplasmonic waveguides.

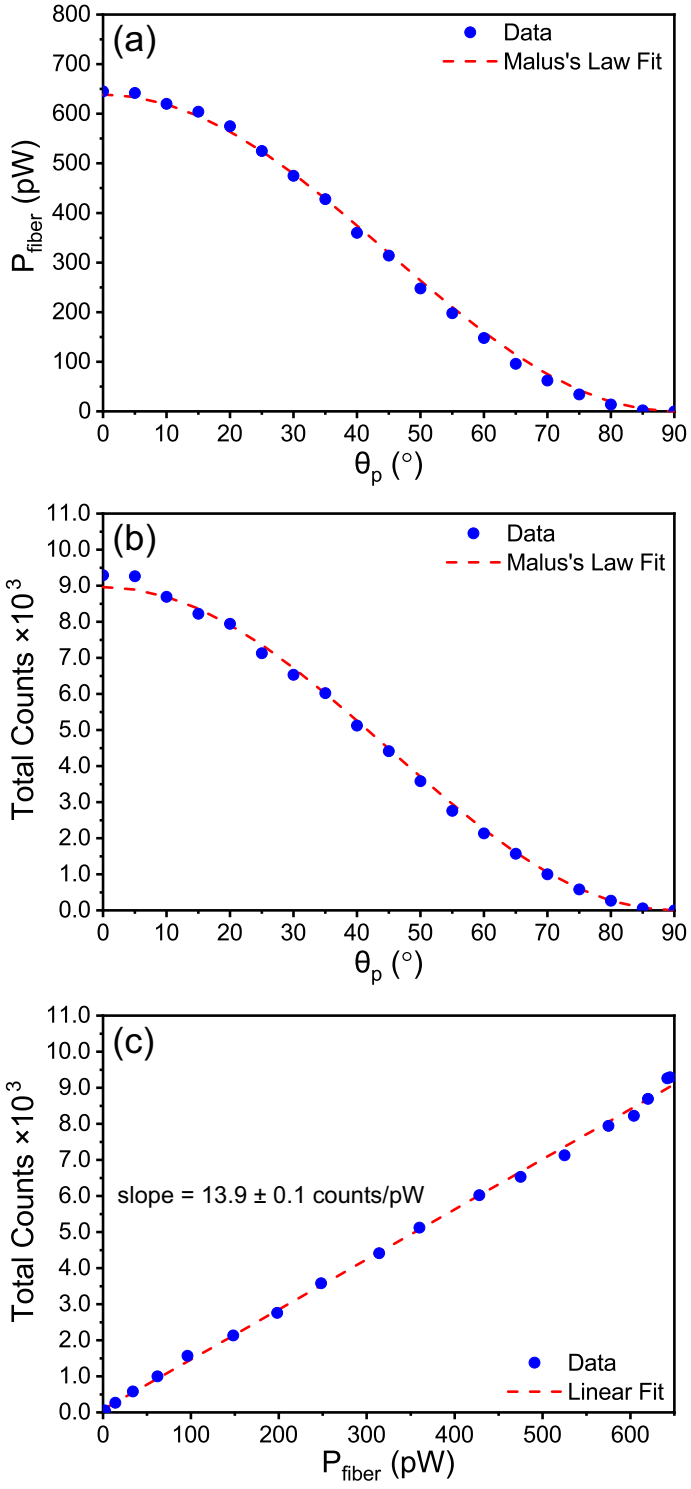

Figure 6.5: (a) Plot of the measured P_{fiber} at $\lambda = 532$nm as a function of the angle between the two polarizers, θ_p. (b) Plot of the total spectrometer counts as a function of the angle between the two polarizers, θ_p. (c) Plot of the total spectrometer counts as a function of the measured P_{fiber}.

This code was based on the algorithm originally developed by Suzuki [215], and is presented in full in Appendix E. The versatile material plugin, written in C++, was developed and meticulously tested against theoretical results over a period of several months to ensure its validity, before being employed to verify the experimental results. Flexibility was a key goal in the development of this plugin, and as such, this code can be applied to characterize any combination of the nonlinearities within an arbitrary Si-based geometry. The nonlinear material parameters of Si utilized within the simulations are given in Table 6.1, while the linear material parameters are from Ref. [33].

Table 6.1: Nonlinear Properties of Si

Parameter	Value	Ref.
n_2	9.0×10^{-14}cm^2/W	[105]
β_{TPA}	0.68cm/GW	[104]
σ_{FCA}	1.45×10^{-17}cm^2	[110]
τ_r	265ps	[114]

In all verifying simulations that follow, the linear material properties of the SiO_2 BOX layer are taken from Ref. [33], and the linear material properties of Au were obtained from Ref. [32]. As both of these materials are well known to have appreciable third-order nonlinearities themselves, these effects must also be accounted for. The nonlinearity of Au is modeled through a $\chi^{(3)} = 0.2 \times 10^{-18}$m^2/V^2 [216], while SiO_2 has demonstrated a nonlinear refractive index of $n_2 = 2.79 \times 10^{-16}$cm^2/W [217].

In each simulation, the nanoplasmonic waveguides are excited via a Gaussian spot focused through a NA=0.85 lens, in congruence with the experimental setup. The incident pulses are taken to have a Gaussian temporal profile, having a $\tau_p = 84$fs pulse duration, and a spectrum centered around $\lambda = 1550$nm. The peak electric field at the focus of the objective was set to align with the experimental conditions, and this methodology allows for accurate comparison with the measured results.

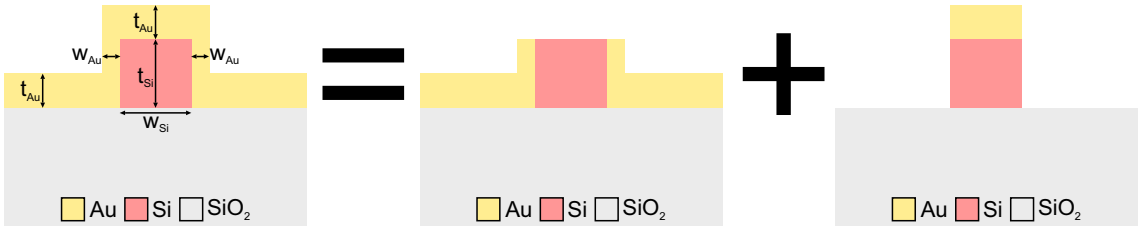

Figure 6.6: Illustration of the cross section of the fabricated ME nanoplasmonic waveguides. The presence of the metal on top of the core implies that this structure is analogous to the superposition of a vertical MI waveguide and a typical MIM waveguide without the Au cap.

6.4 ME Nanoplasmonic Waveguides

To understand nonlinear phenomena occurring within the ME nanoplasmonic waveguides, it is first instructive to examine the basic modal properties at the fundamental wavelength of interest, $\lambda = 1550$nm. A schematic of the fabricated structure is illustrated in Fig. 6.6, which can be thought of as the superposition of the MI and MIM waveguides depicted in Section 2.2.2. In this case, the presence of the top metal layer provides a mechanism for the orthogonal TE and TM polarizations to couple. This can be seen by examining the modal dispersion properties of the waveguide as a function of the Si core width, w_{Si}. The modal effective indices are plotted versus w_{Si} in Fig. 6.7. In these calculations, the core height is $t_{Si} = 340$nm, the Au film thickness is $t_{Au} = 200$nm, and the sidewall coverage is $w_{Au} = 100$nm on each side, in accordance with the fabricated structures. Note that in Fig. 6.7, the typical modal designations of "TE" and "TM" have been purposely avoided, due to the fact that this coupling changes the nature of the modes as w_{Si} is altered. While a mode can be defined as "TE" or "TM" for a single value of w_{Si}, that nomenclature would change as one progresses along the curve. Because of this, the modes are named in descending order of their effective index at $w_{Si} = 1000$nm. For example, the mode profiles are depicted aside the plot for the first four modes at w_{Si} =230nm (left) and $w_{Si} = 1000$nm (right), where the arrows indicate the orientation of the primary

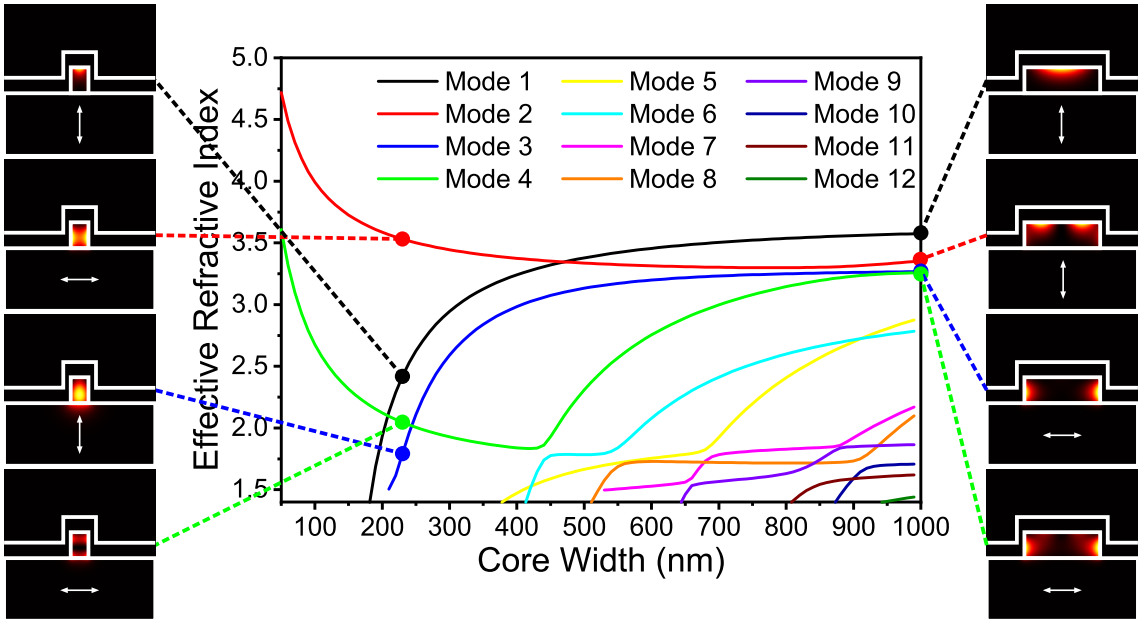

Figure 6.7: Plot of the effective indices of the modes present within the ME waveguide as a function of the core width, w_{Si}. Wide waveguides are highly multimode, with 12 modes existing at $w_{Si} = 1000$nm, while narrow waveguides support only two modes. The presence of the metal cap allows coupling between the different field components, and as the waveguide width is reduced, modes can change their polarization state. The insets depict the $|E|^2$ profiles of the first four modes at $w_{Si} = 230$nm (left) and $w_{Si} = 1000$nm (right), while the arrows indicate the orientation of the dominant electric field component.

electric field component. Clearly, Mode 1 and Mode 2 are TM (with respect to the substrate) at $w_{Si} = 1000$nm (right), while Modes 3 and 4 are TE (with respect to the substrate). As the width is reduced, Mode 1 remains TM until it reaches cutoff, Mode 2 converts from TM to TE due to this coupling as w_{Si} is reduced, Mode 3 converts from TE to TM before reaching cutoff, and Mode 4 converts from TE to TM and back between $w_{Si} = 1000$nm and $w_{Si} = 230$nm. Additional evidence of the coupling can be observed by the anticrossings and inflection points present in the dispersion curves. It is evident from the plot that ME waveguides with wide Si cores are highly multimode in nature with complex dispersion properties, while the modal structure simplifies for narrower structures as the higher order modes become cutoff.

For simplicity, only the first four modes will be considered henceforth, as these are the four lowest order modes. The propagation length, L_{prop}, of these four modes as a function of w_{Si} is plotted in Fig. 6.8(a). At $w_{Si} = 1000$nm, the propagation lengths

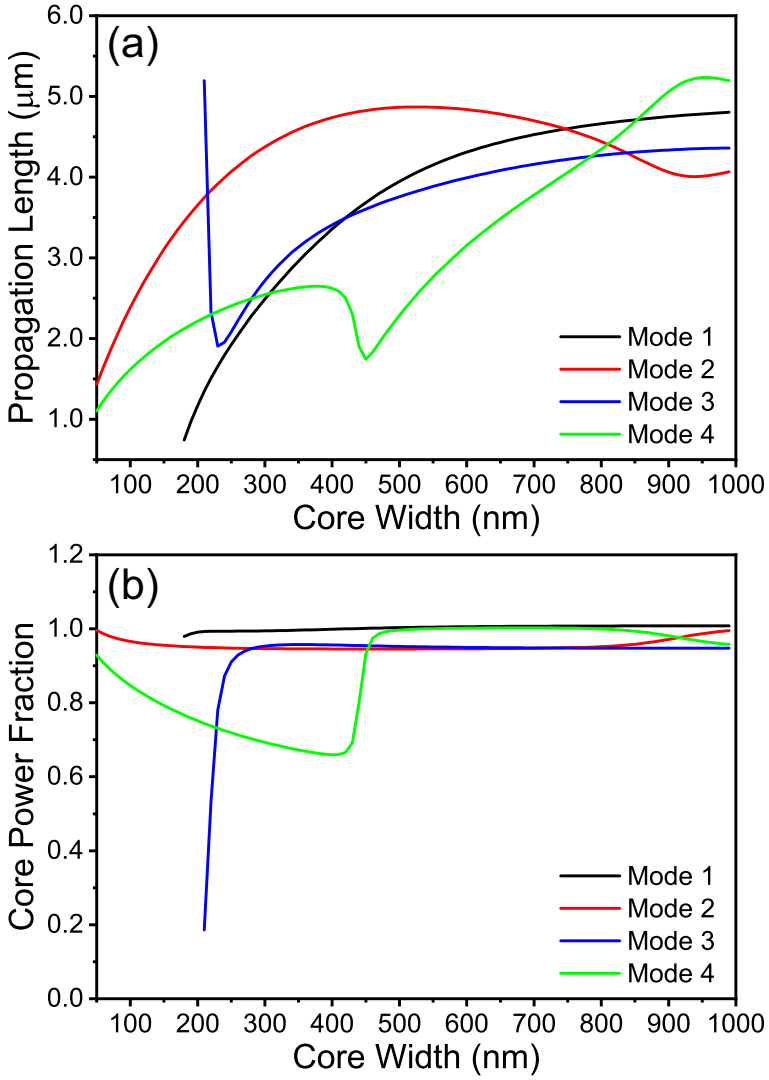

Figure 6.8: Plots of (a) the propagation length of the first four modes as a function of the core width, and (b) the fraction of power confined within the Si core of the first four modes, as a function of the core width.

range between 4.1μm$\leq L_{prop} \leq$ 5.2μm, while at $w_{Si} = 50$nm, the propagation lengths have been reduced to values between 1.1μm$\leq L_{prop} \leq$ 1.4μm. This is indicative of high optical losses, and it also implies that devices based upon this architecture will inherently possess an extremely compact on-chip footprint.

The confinement of optical power to the Si core is examined in Fig. 6.8(b) for the same four modes as a function of w_{Si}. Modes 1-3 consistently exhibit >94% of the power confined to the Si core, although the power confined by Mode 3 begins decreases

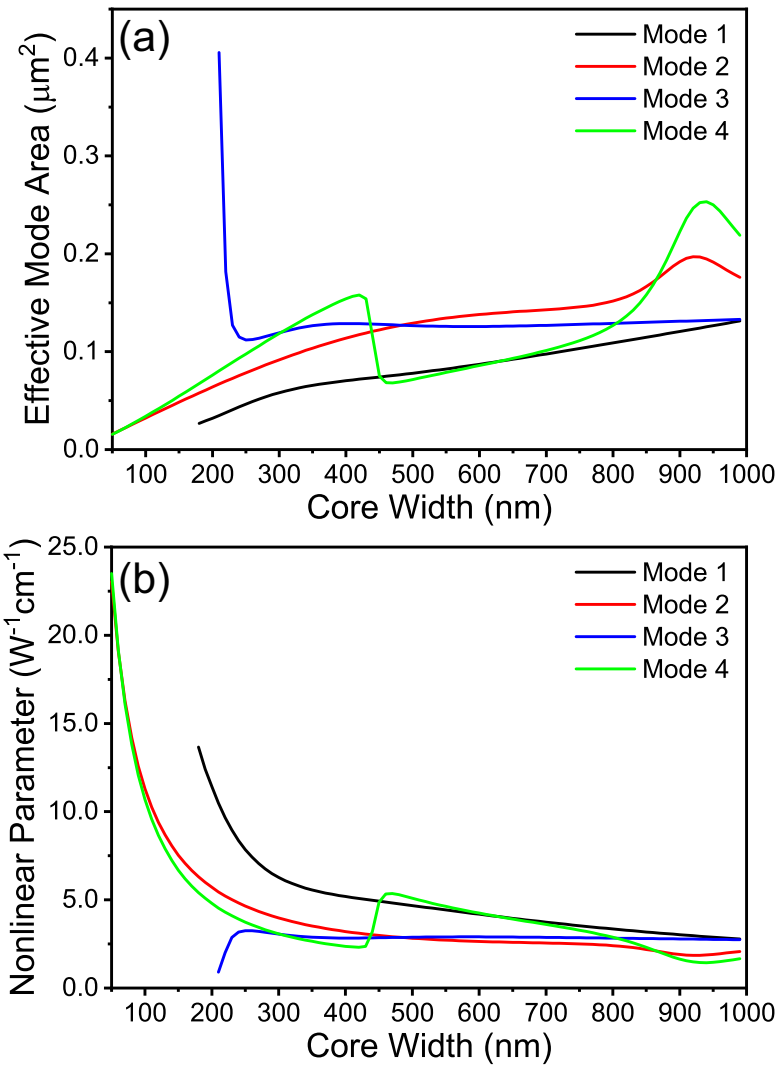

Figure 6.9: Plots of (a) the effective mode area of the first four modes as a function of the core width, and (b) the nonlinear parameter of the first four modes, as a function of the core width.

rapidly as it approaches cutoff. Mode 4 maintains >98% power confinement until w_{Si} is reduced to 460nm. Here, it drops to 66% at $w_{Si} = 400$nm before increasing back to 93% as $w_{Si} = 50$nm. In general, this ME geometry consistently maintains high modal confinement. This is desirable for several applications, including nonlinear optics, where the power is confined to the highly nonlinear core.

Figure 6.9(a) plots the effective mode area, A_{eff}, of each of the four lowest order

modes. The effective mode area is defined as [108]:

$$A_{eff} = \frac{\left(\iint_{-\infty}^{\infty} |E(x,y)|^2 dxdy\right)^2}{\iint_{-\infty}^{\infty} |E(x,y)|^4 dxdy} \tag{6.1}$$

This is another metric indicating the ability of these structures to confine energy to a compact volume. For the wider waveguides, the effective mode area is less than the actual area of the core. For example, at $w_{Si} = 1000$nm, the Si core area is 0.34μm^2, while the effective mode areas range between $0.13\mu\text{m}^2 \leq A_{eff} \leq 0.21\mu\text{m}^2$. Here, the effective mode area is 38-62% of the total available area. This means that nonlinear harmonic generation is only occurring over a small fraction of the core area. If the desired application is nonlinear light generation, the chip space is not used as effectively here. However, considering $w_{Si} = 50$nm, the effective mode area ranges between $0.0155\mu\text{m}^2 \leq A_{eff} \leq 0.0158\mu\text{m}^2$, while the core area is 0.017μm^2. The modes comprise 91-93% of the available core area for nonlinear light generation.

Figure 6.9(b) plots the nonlinear parameter, γ_{NL}, for these modes as a function of w_{Si} as well. This value is defined as [108]:

$$\gamma_{NL} = \frac{2\pi n_2}{\lambda A_{eff}} \tag{6.2}$$

The nonlinear parameter is a metric typically employed to characterize the spectral broadening due to SPM in nonlinear fibers and waveguides. However, it has also been shown that higher values of γ_{NL}, which can be interpreted as a higher nonlinearity in a smaller area, increase the third-harmonic power generated [218], and it is therefore desirable to have high nonlinear parameters. Here, the lowest value of γ_{NL}, excluding the decrease as Mode 3 approaches cutoff, is $\gamma_{NL} = 1.44\text{W}^{-1}\text{cm}^{-1}$. The nonlinear parameters approach $\gamma_{NL} = 23\text{W}^{-1}\text{cm}^{-1}$ and $\gamma_{NL} = 23.5\text{W}^{-1}\text{cm}^{-1}$ for Modes 2 and 4 respectively at $w_{Si} = 50$nm. For reference, a recent study on THG in SOI photonic

waveguides calculated values of γ_{NL} between $0.24\text{W}^{-1}\text{cm}^{-1} \leq \gamma_{NL} \leq 1.68\text{W}^{-1}\text{cm}^{-1}$ for SOI waveguides [214]. The maximum values of this structure are 13.7-14.0× higher, implying a much stronger nonlinearity.

6.4.1 Excitation Parameters

During the ensuing experiments, the average incident laser power was varied within the range $0\text{mW} \leq P_{ave}^{\omega} \leq 30\text{mW}$, where the superscript ω denotes that this is the power at the fundamental frequency of the incident light at $\lambda = 1550\text{nm}$. With the knowledge of the ultrafast pulse duration of $\tau_p = 84\text{fs}$, the pulse repetition rate of $f_{rep} = 90\text{MHz}$, and the fact that the laser pulses exhibit a sech^2 temporal profile, the peak laser power incident upon the waveguides was therefore within the range of $0\text{kW} \leq P_{pk}^{\omega} \leq 3.49\text{kW}$.

Furthermore, after focusing through the NA=0.85 objective lens, the $1/e^2$ focal spot radius is calculated to be $w_0 = 710\text{nm}$, which implies that the focal spot diameter, $2w_0$, is less than the $2.5\mu\text{m}$ FIB milled opening at the end facets of the waveguides. The peak intensities incident on the waveguide end facet are calculated to vary between $0\text{W/m}^2 \leq I_{opt,pk}^{\omega} \leq 2.63 \times 10^{15}\text{W/m}^2$, and correspondingly, the peak electric fields incident on the waveguide end facet are swept between $0\text{V/nm} \leq E_{pk}^{\omega} \leq 1.41\text{V/nm}$.

There are two additional key factors to consider upon exciting the nanoplasmonic ME waveguides. The first parameter is the coupling efficiency between the objective lens and the waveguide, which determines how much power is coupled into the waveguide. This quantity is plotted in Fig. 6.10(a). For the micrometer scale waveguides fabricated in Chapter 5, this parameter must be calculated and cannot be experimentally measured with techniques such as the commonly employed cut-back method. The narrow Si characterization beam inherently implies that the optical fiber is located within a close proximity to the objective focus, and consequently

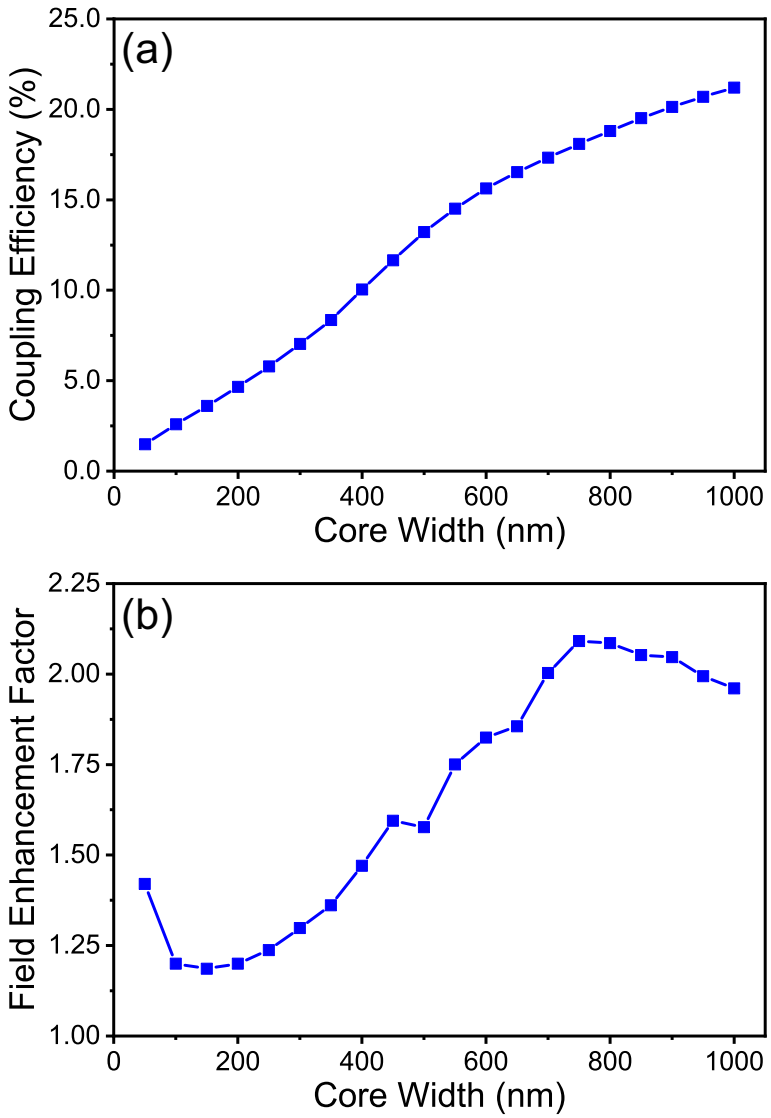

Figure 6.10: Plots of (a) objective-waveguide coupling efficiency as a function of the core width, and (b) the factor by which $|E|$ is enhanced within the nanoplasmonic waveguide, as a function of the core width.

collects a significant portion of incident light scattered over the Si characterization beam. Since this scattered light cannot be differentiated from the light transmitted through the waveguide, the coupling efficiency cannot be determined experimentally from these structures. As shown in Fig. 6.10(a), the objective-waveguide coupling efficiency increases as the waveguide core width increases, and exceeds 20% for the wide waveguides.

The second parameter is the field enhancement produced within the structure. As

the incident radiation couples into the confined surface plasmon mode, the electric fields become enhanced. The factor by which the magnitude of the electric field, $|E|$, is enhanced is plotted in Fig. 6.10(b) as a function of the waveguide core width. The field enhancement is larger for wider waveguides, peaking at about $2.1\times$ for $w_{Si} = 750$nm. This field enhancement is due to the confinement the plasmonic mode within the structure, where smooth sidewalls are considered. Additional enhancement can be observed if significant roughness is present, as is the case in the fabricated structures comprised of a rough, granular Au film.

6.4.2 Device Robustness at High Field Excitation

Upon excitation by the ultrafast laser pulses, it was found that a number of the devices were extremely sensitive to the strength of the excitation, and thus, susceptible to damage. In a typical experiment, first, the laser input from the objective had to be aligned to one end facet of the waveguide, which was achieved by monitoring the IR power in the fiber, as well as observing the above USB camera output for visible THG emission.

However, alignment of the output fiber had to be performed with the THG signal in the visible regime, and not based on power readings of the $\lambda = 1550$nm femtosecond pulse. Due to the narrow width of the Si characterization beams, and the short length of the nanoplasmonic waveguides, the output optical fiber was in extremely close proximity to the objective's focal spot, and could efficiently collect the light scattered off of the characterization beam. Since the scattered light could not be separated from the light transmitted through the guide, the THG had to be utilized. Once the THG was visible on the display, the fiber could be aligned to the waveguide output facet via optimizing its position and monitoring the THG spectrum.

Nanoplasmonic ME waveguides with large core widths exhibit very high coupling efficiencies and field enhancements, as depicted in Fig. 6.10. Considering this, it

was possible for them to be damaged, even with incident laser powers around $P_{ave}^{\omega} =$ 20mW. Furthermore, longer devices were also susceptible to damage. Due to the short propagation lengths of optical modes at the fundamental wavelength, waveguides having lengths longer than their modal propagation length would attenuate the $\lambda =$ 1550nm signal significantly by the time it reached the end of the waveguide. Combined with the high absorption of the THG in the visible regime within Si, this resulted in low THG output signals. In order to align the THG signal for measurement, the incident $\lambda = 1550$nm power had to be increased to boost the THG signal to a level suitable for aligning the fiber, often to the point of damaging the device.

As a result, many specimens were damaged and did not survive during the alignment process and the experiments. Damage was characterized by a sudden drop in the THG signal, which could not be recovered. An example is depicted in Figs. 6.11(a) and 6.11(b), which depict the measured third-harmonic signal as a function of the incident power for waveguides with $w_{Si} = 780$nm and $L = 2.2\mu$m, and $w_{Si} = 500$nm and $L = 4.8\mu$m, respectively. Figure 6.11(c) depicts a SEM image of an exemplary device that sustained damage in this manner, where $w_{Si} = 875$nm, and $L = 1.7\mu$m. The damage occurs at the waveguide input facet and causes a rough, uneven surface that inhibits efficient optical coupling. This damage could feasibly be repaired with FIB milling.

In this regard, characterizing these structures requires a delicate balance of the input power, coupling efficiency, structural field enhancement, and surface roughness field enhancement. It should be noted that this sensitivity is simply a consequence of the laboratory setup and procedures required to collect the signals to characterize the structures. In an on-chip configuration, which doesn't exhibit such limitations, many of these issues would not arise.

The optimum configuration for such a device is a narrow waveguide core and a short length, $\leq L_{prop}$. From the analysis presented in Section 6.4, shorter and

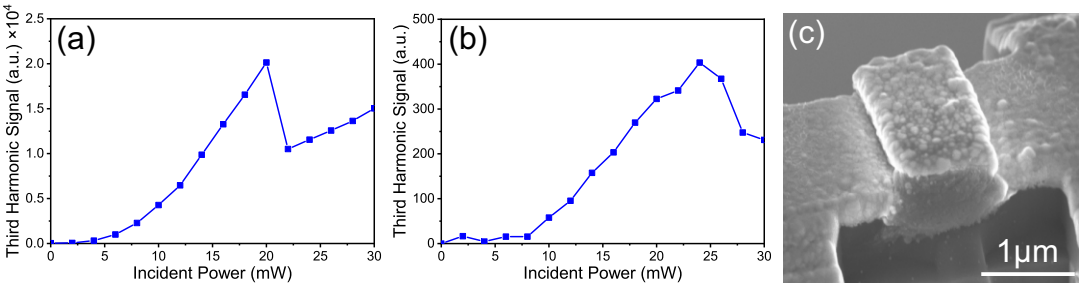

Figure 6.11: Depiction of the damage that can occur within these nanoplasmonic devices. Plot of the measured THG signal for a waveguide with (a) $w_{Si} = 780$nm and $L = 2.2\mu$m, and (b) $w_{Si} = 500$nm and $L = 4.8\mu$m. The sudden drops in the signal are the signature of permanent, irreversible damage to the waveguide. These signals are plotted against incident power, as the calculated value for coupled power becomes inaccurate after the device breaks, due to the damaged input facet decreasing the coupling efficiency. (c) SEM image of an exemplary damaged device with $w_{Si} = 875$nm, and $L = 1.7\mu$m. The input facet is rough and damaged, inhibiting efficient optical coupling.

narrower devices are more desirable because (1) they exhibit simplified mode structure and support fewer modes, (2) they facilitate compact devices which are desirable for dense on-chip integration, (3) they provide high field confinement, and (4), yield small mode areas which are a large fraction of the core area, leading to high nonlinearity.

6.4.3 Device and Modes

The most experimentally robust devices are the narrowest in width and the shortest in length. Herein, nonlinear performance of one exemplary device, depicted in Fig. 6.12, is discussed. This ME nanoplasmonic waveguide consists of a Au clad Si core $t_{Si} = 340$nm tall and $w_{Si} = 100$nm wide, having a length of $L = 2.4\mu$m.

The effective indices of a selection of the modes are depicted in Fig. 6.13(a). Note that here, the standard "TE" and "TM" notation can be employed, as the mode character does not change. As is conventional, TE modes are dominated by a horizontal electric field component, parallel to the substrate, while TM modes are dominated by a vertical electric field component, perpendicular to the substrate. Also, while it is unconventional to denote a surface plasmon mode as "TE", this does not imply any different nature to the plasmon; it is simply the consequence of the

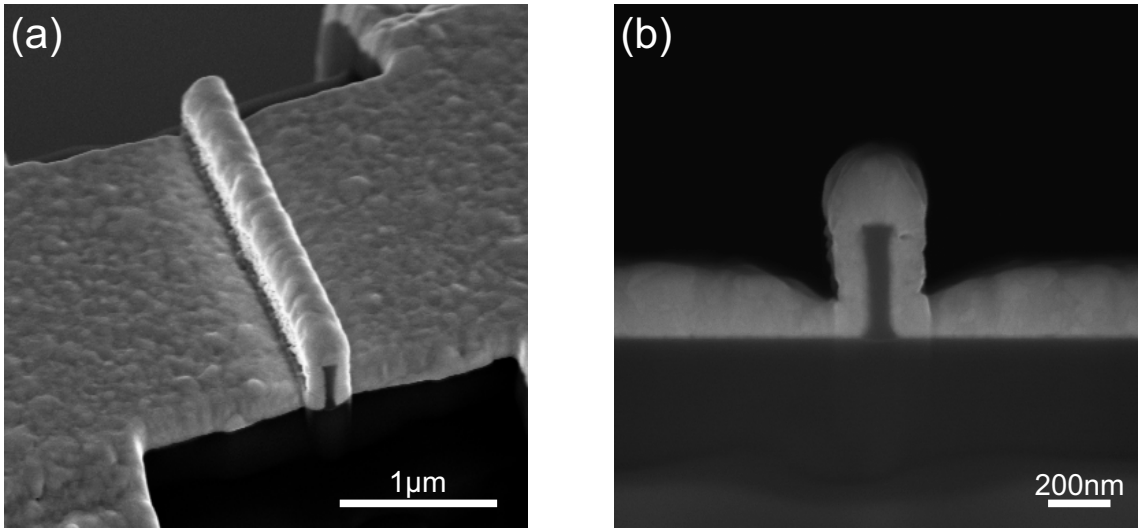

Figure 6.12: SEM images of the optimal device with an Au clad Si core t_{Si} = 340nm tall, w_{Si} = 100nm wide, and $L = 2.4\mu$m long. (a) Depicts a tilted image of the whole device, and (b) depicts the cross-sectional profile of the input facet.

plasmonic fields propagating along the vertical sidewall of the waveguide.

This waveguide supports two modes in the spectral regime around 1550nm, both of which are TE in nature and are depicted in Fig. 6.13(b) and 6.13(c). For convenience, a summary of the properties of these modes is presented in Table 6.2. The third harmonic of these signals will manifest in the green region of the visible spectrum. Henceforth, wavelengths in the IR region around 1550nm will be denoted by λ_ω, while wavelengths in the green part of the visible spectrum produced via THG will be denoted by $\lambda_{3\omega}$. At $\lambda_{3\omega}$, due to the high refractive index of silicon and the shorter wavelength, many modes are present. The discussion can be limited to the three modes having TE nature shown in Fig. 6.13(d)-6.13(f). This is due to the fact that the silicon crystal is of the *m3m* point group, and thus, its $\chi^{(3)}$ tensor only has two independent elements [219, 220]. Since one is significantly larger than the other, the nonlinearities in silicon can be treated by an effective $\chi^{(3)}$ value and is modeled isotropically [219, 220]. As a consequence, the nonlinear polarization generated will be aligned with the exciting field. By orienting the incident polarization as to excite the TE fundamental modes, only TE fields at $\lambda_{3\omega}$ will be produced.

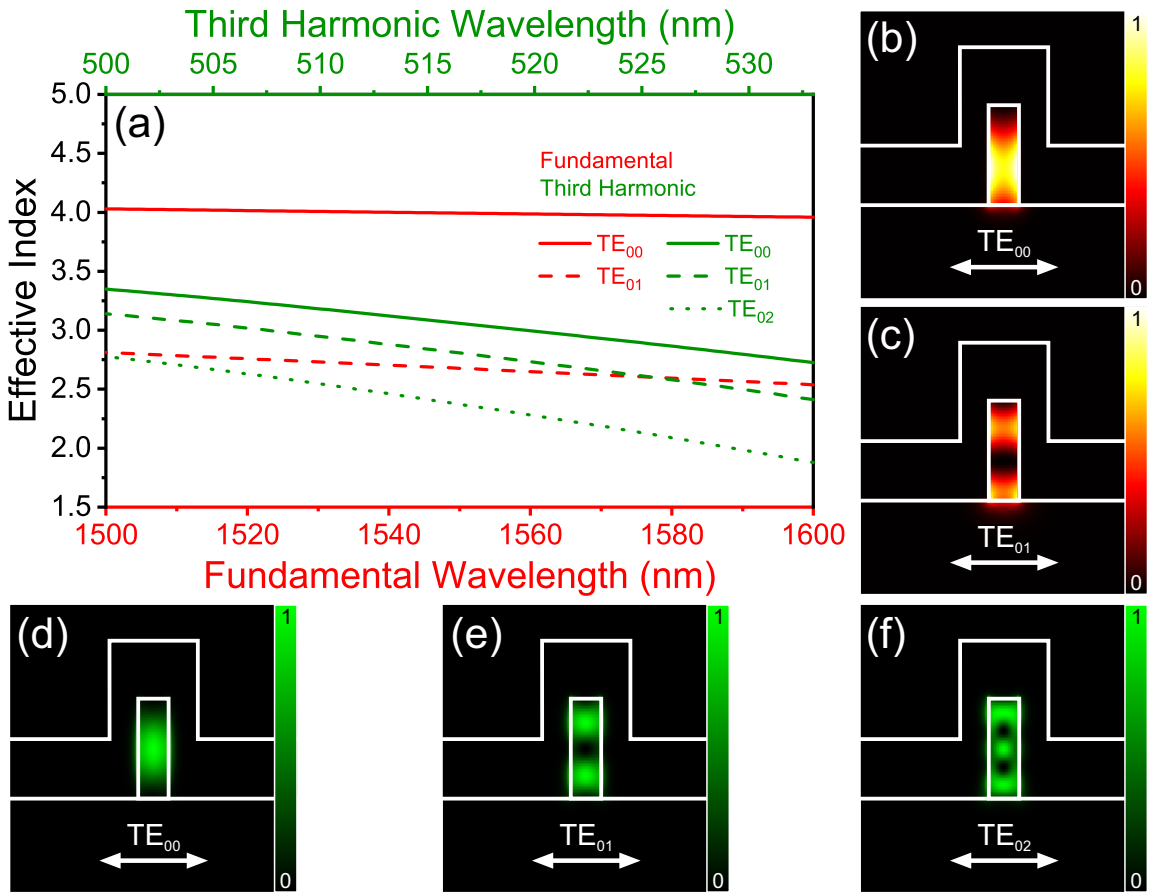

Figure 6.13: (a) Plot of the effective index of the modes within the optimal device. Data is shown for the two fundamental modes, and the three TE third-harmonic modes. (b) and (c) depict the $|E|^2$ field profiles for the TE_{00} and TE_{01} modes at $\lambda_\omega = 1550$nm, respectively. (d)-(f) depict the $|E|^2$ field profiles for the TE_{00}, TE_{01} and TE_{02} modes at $\lambda_{3\omega} = 516.7$nm, respectively. In (b)-(f), the arrows indicate the orientation of the dominant electric field component.

Nonlinear processes, such as THG, typically require phase matching between the fundamental mode at λ_ω and the third harmonic mode at $\lambda_{3\omega}$ for efficient light generation. The dispersion plot shown in Fig. 6.13(a) indicates that phase matching is present within this waveguide between the fundamental TE_{01} mode at $\lambda_\omega = 1576$nm

Table 6.2: Mode Properties at $\lambda_\omega = 1550$nm

Parameter	TM_{00}	TM_{01}
$n_{eff,\omega}$	3.99	2.68
L_{prop}	2.4μm	1.6μm
Core Power Fraction	96.6%	84.7%
A_{eff}	0.0323μm^2	0.0341μm^2
γ	11.3W^{-1}cm^{-1}	10.7^{-1}cm^{-1}

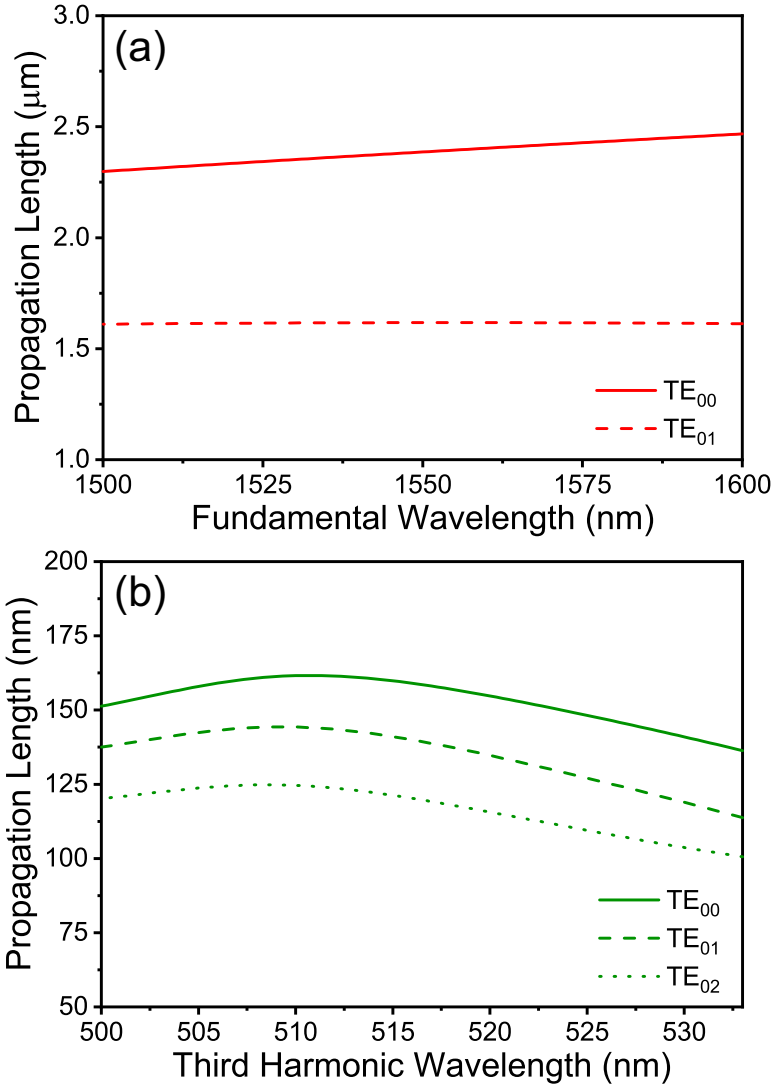

Figure 6.14: (a) Plot of the propagation length of the two fundamental TE modes. (b) Plot of the propagation length of the three third-harmonic TE modes.

and the third harmonic TE_{01} mode at $\lambda_{3\omega} = 525.3$nm, which is encompassed within the spectrum of the laser pulse shown in Fig. 6.2. As will be shown shortly, it is not possible for phase matching to play a substantial role within this structure.

First, the propagation lengths associated with the aforementioned optical modes are plotted in Fig. 6.14(a) (fundamental modes) and Fig. 6.14(b) (third-harmonic modes). While the two fundamental modes have propagation lengths on the micrometer scale, with $L_{prop} = 2.4\mu$m and $L_{prop} = 1.6\mu$m for the TE_{00} and TE_{01} modes at $\lambda_{\omega} = 1550$nm, the third-harmonic modes are significantly more lossy. Considering

the third-harmonic of λ_ω = 1550nm, at $\lambda_{3\omega}$ = 516.7nm, the propagation lengths for the three harmonic modes are L_{prop} = 158nm, 139nm, and 120nm for the TE_{00}, TE_{01}, and TE_{02} modes at this wavelength, respectively. Interestingly, by examining their effective refractive indices, $n_{eff,3\omega}$ =3.06, 2.80, and 2.37, the effective wavelength of these modes can be determined as $\lambda_{3\omega,eff}$ =169nm, 184nm, and 281nm, respectively. Clearly, within a span of less than a wavelength within the waveguide, each of these modes is significantly attenuated, implying that the THG process in this ME nanoplasmonic waveguide is extremely localized.

This can be further demonstrated through examination of the coherence length between the modes, which is calculated as:

$$L_{coh} = \frac{\lambda_\omega}{6|n_{eff,\omega} - n_{eff,3\omega}|} \tag{6.3}$$

Figure 6.15 plots the coherence length between each of the modes as a function of the fundamental wavelength. The coherence lengths between the TE_{00} mode at λ_ω and the third-harmonic modes are all on the order of a couple hundred nanometers, while the coherence lengths of the TE_{01} mode at λ_ω and the third-harmonic modes are on a micrometer scale, except for the phase matched point observed in Fig. 6.15(e), where the coherence length increases asymptotically. The third-harmonic signals will be attenuated well within these length scales, and coherent build up of the THG cannot be observed. Any detected THG is generated within an extremely local region. This suggests that THG measured from a waveguide essentially originates from the local region around the end facet itself.

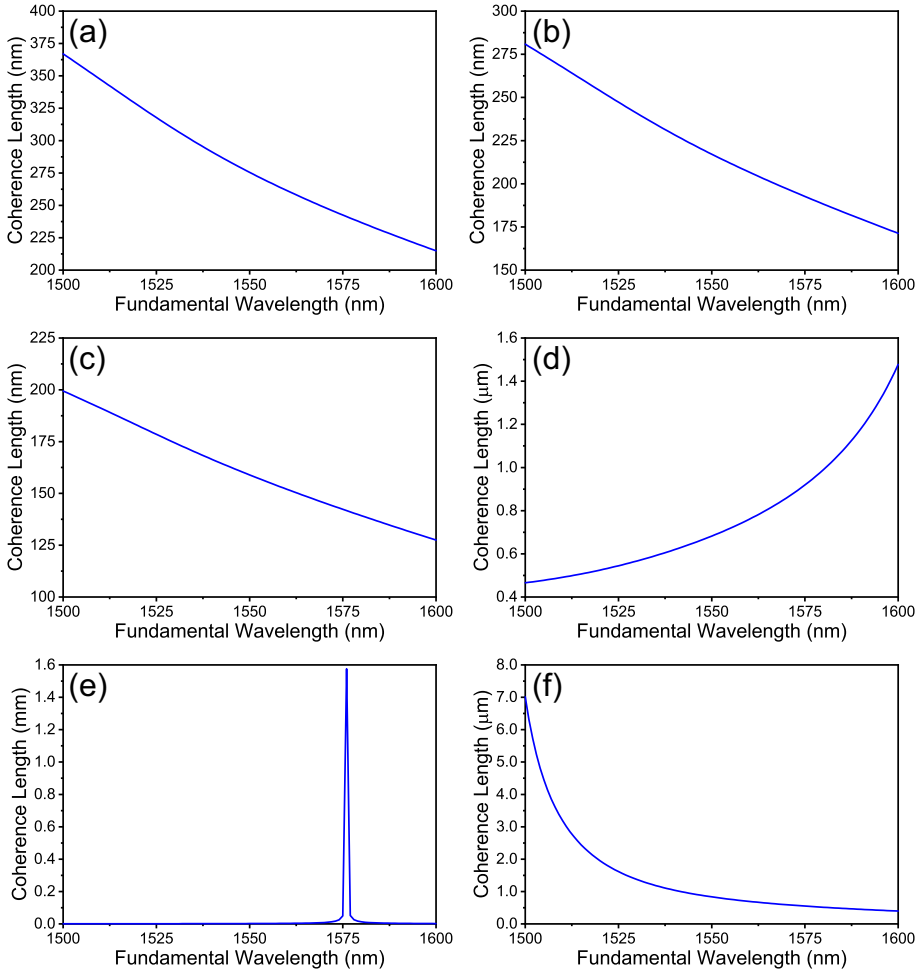

Figure 6.15: Coherence lengths of the modes plotted against the fundamental wavelength for: (a) TE_{00} mode at λ_ω and TE_{00} mode at $\lambda_{3\omega}$, (b) TE_{00} mode at λ_ω and TE_{01} mode at $\lambda_{3\omega}$, (c) TE_{00} mode at λ_ω and TE_{02} mode at $\lambda_{3\omega}$, (d) TE_{01} mode at λ_ω and TE_{00} mode at $\lambda_{3\omega}$, (e) TE_{01} mode at λ_ω and TE_{01} mode at $\lambda_{3\omega}$, (f) TE_{01} mode at λ_ω and TE_{02} mode at $\lambda_{3\omega}$.

6.5 Nonlinear Phenomena

6.5.1 Third-Harmonic Generation

Upon coupling the ultrafast laser pulses into the nanoplasmonic ME waveguides, the devices emitted bright green light from their input and output facets. Photographs of the THG emitted by the $w_{Si} = 100$nm device during characterization are depicted in Fig. 6.16(a)-6.16(b), where the emission is observed from the end faces. Since the THG light is emitted from the output facet in a highly directional manner, and the top of the waveguide is clad in Au, only a small amount of light is scattered

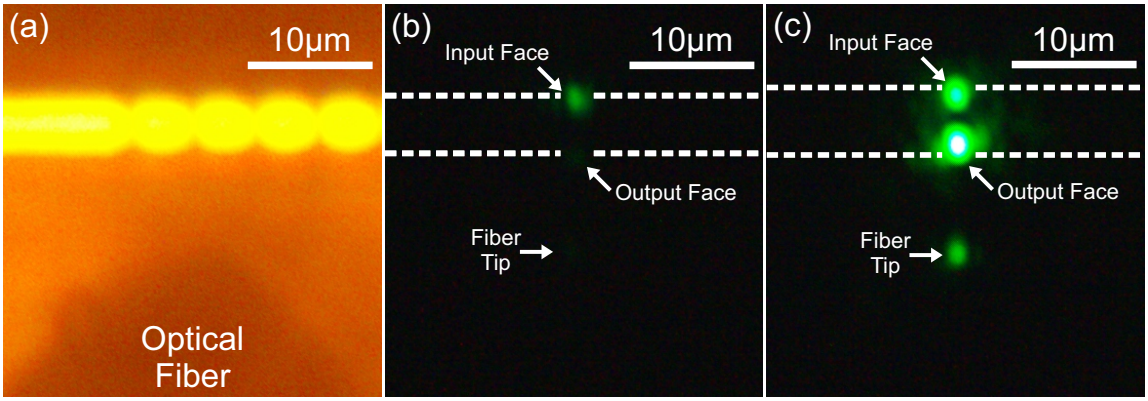

Figure 6.16: (a) Image of the $w_{Si} = 100$nm, $L = 2.4\mu$m Au-coated Si characterization beam. (b) Photograph of the THG generated by this ME nanoplasmonic waveguide. (c) Photograph of the THG emitted from a $w_{Si} = 400$nm waveguide that is $L = 3.0\mu$m long with a defect at its output face such that more light is scattered out of plane and into the above camera. THG is observed at both the input and output face, as well as scattering off of the optical fiber.

towards the camera, which is mounted above the waveguide. For illustration, Fig. 6.16(c) depicts the THG in a $w_{Si} = 400$nm, $L = 3.0\mu$m waveguide with a defect at the output facet such that more light is scattered out of plane to the above camera.

The THG spectra were captured as the incident laser power was increased from $P^{\omega}_{ave} = 0$mW to $P^{\omega}_{ave} = 30$mW. Calculations show that the objective-to-waveguide coupling efficiency was 2.8%, and the corresponding average power coupled into the waveguides was varied between $0\mu\text{W} \leq P^{\omega}_{ave,coupled} \leq 840\mu\text{W}$, which corresponds to coupled peak powers in the range $0\text{W} \leq P^{\omega}_{pk,coupled} \leq 98\text{W}$. A surface plot of the measured spectra is presented in Fig. 6.17, where the surface has been normalized to the maximum value, and is plotted as a function of the coupled average power. Visible green emission was observed by eye and on the spectrometer for coupled average (peak) powers of only $P^{\omega}_{ave,coupled} = 112\mu\text{W}$ ($P^{\omega}_{pk,coupled} = 13\text{W}$). This is much lower than the $P^{\omega}_{pk,coupled} = 37\text{W}$ threshold required for observing THG in SOI photonic waveguides [214]. The green spectrum spans between approximately $505\text{nm} \leq \lambda_{3\omega} \leq 525\text{nm}$, and exhibits a three-peaked structure, with the peaks located at $\lambda_{3w} = 507.6$nm, 512.7nm, and 517.7nm.

For comparison, Fig. 6.18 depicts the maximum measured THG spectrum along-

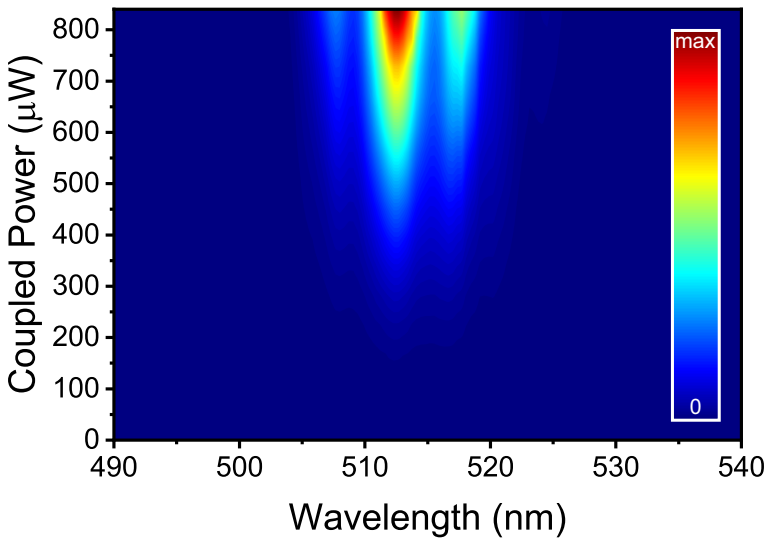

Figure 6.17: Experimentally measured THG spectrum as a function of the power coupled into the waveguide.

side the calculated ideal third-harmonic spectrum of laser pulses in free space. This calculation was performed from the spectrum in Fig. 6.2 and accounts for mixing between any three frequency components within the laser pulse. The measured THG spectrum output from the nanoplasmonic waveguides agrees relatively well with the idealized third-harmonic spectrum of the laser pulses, and this agreement indicates the highly localized nature of the THG within these waveguides. The laser pulses are essentially converted to their third-harmonic locally, and the high absorption of the green radiation does not allow for material or structural properties to influence the spectrum significantly.

Absent from the measured spectra is the broadband white light that had been observed previously within an Au-Si MI nanoplasmonic waveguide [200]. This white light spectrum was attributed to the energy given off by collisions between electrons in an avalanche process generated by the ponderomotive acceleration of TPA excited carriers. While the Si core dimensions are comparable between the MI device studied in [200] and the present waveguide, the spatial electric field gradients produced by the modes within the waveguide cores of the two structures are significantly different.

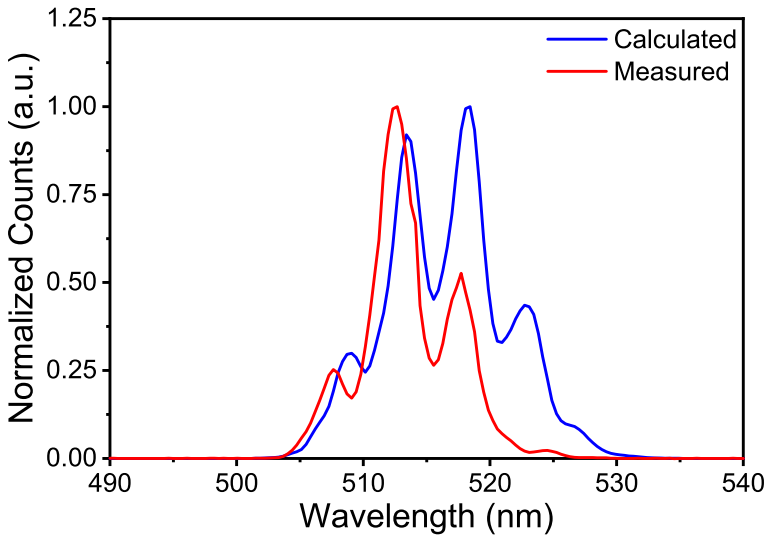

Figure 6.18: Comparison of the measured THG spectrum at $P^{\omega}_{ave,coupled} = 840\mu$W with the ideal third-harmonic spectrum calculated from the incident laser pulse spectrum.

From the fundamental mode distribution depicted in Fig. 6.13(b), it is evident that across the width of the ME waveguide, the fields are much more symmetric, and have a reduced gradient. The force acting to accelerate the electrons is lower in the ME structure, and the electrons cannot gain enough energy to incite an electron avalanche, and emit the broadband white light spectrum. The white light emission cannot be observed within the present ME structure. In fact, this is advantageous in the present geometry, as the electron avalanche process creates additional excess carriers which can further attenuate the fundamental pulse via FCA.

6.5.2 Third-Harmonic Signal Power Scaling

The power scaling trends for the measured THG spectrum, as well as for the three individual peaks are plotted on a logarithmic scale in Fig. 6.19. The linear fits exhibit slopes of 2.70 for the integrated spectrum scaling, and 2.73, 2.79, and 2.87 for the three individual peaks. These values are close to the expected slope of 3, as dictated by the third-order $\chi^{(3)}$ process. However, there is a slight deviation from the ideal value.

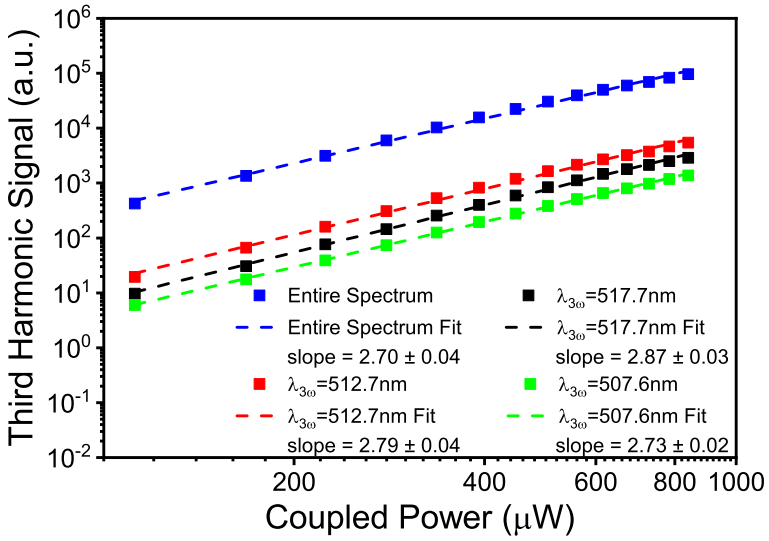

Figure 6.19: Measured power scaling trends of the whole spectrum, and the three peaks plotted as a function of the power coupled into the waveguide.

The best way to measure the fundamental power exciting the THG detected by the spectrometer would be to measure the power at the output of the waveguide. Due to scattering of the incident light from the objective focus into the optical fiber, this value cannot be experimentally realized. To determine the third-harmonic signal power scaling here, the incident power is measured and the power coupled into the waveguide is calculated. For the power scaling order to be less than the expected value of 3, the actual power exciting the measured THG must be lower than expected, and this reduction must be nonlinear. Any linear losses, such as linear absorption or scattering, would not affect the power scaling. Thus, this power reduction must originate from TPA and FCA induced loss, and these effects must be significant on this short 2.4μm length scale. Typically, these loss mechanisms are not a concern for waveguides this short, and are much more prevalent in devices 1-2 orders of magnitude longer. Simulations of this waveguide, having a length of $L = 2.4\mu$m, and a known structural electric field enhancement factor of about 1.2, indicated that the power scaling should be 3. The simulated power scaling for excitation by the same peak electric fields as in the experiment is shown in Fig. 6.20, where it is clear that the

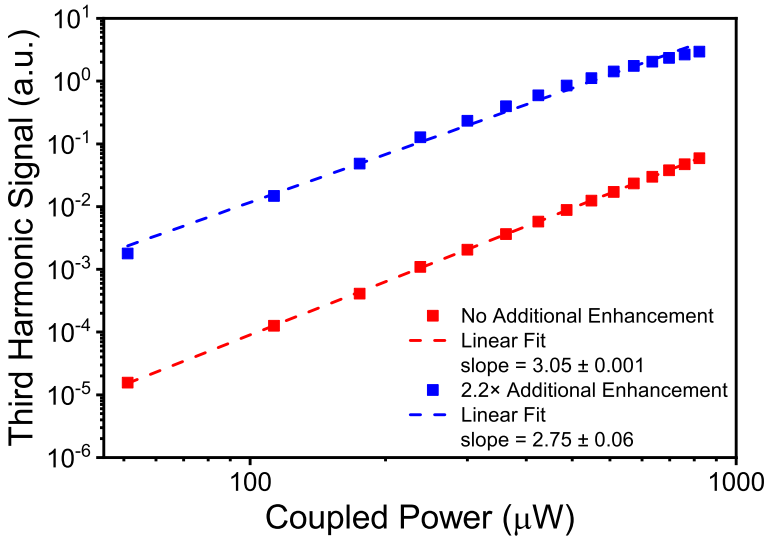

Figure 6.20: Plot of the simulated power scaling trends against the experimentally coupled power when considering no additional electric field enhancement (red), and an extra 2.2× electric field enhancement due to the roughness of the Au cladding (blue).

THG output scales with the expected slope of 3.05±0.001 as expected.

Clearly, in order for TPA and FCA to be significant at these short length scales, the electric fields within the waveguide must be higher than predicted by simulations. This additional electric field enhancement must be due to the roughness of the Au composing the sidewalls of the device, which was observed in Chapter 5. To reproduce the reduced power scaling order of about 2.70, an additional roughness induced electric field enhancement factor of 2.2 has to be accounted for. Fig. 6.20 depicts the power scaling trends with this additional electric field enhancement included in the calculation. The simulated power scaling order is found to be 2.75±0.06, which is in excellent agreement with the experimentally measured value of 2.70±0.04.

The observed reduction in the power scaling order is a consequence of employing the power coupled into the waveguide as a reference. The nonlinear THG process is still third-order, however, the nonlinear loss induced by TPA and FCA attenuate the exciting power significantly. This would eventually lead to an observed saturation in the THG output, as the fundamental power at the output of the waveguide is reduced by TPA and FCA.

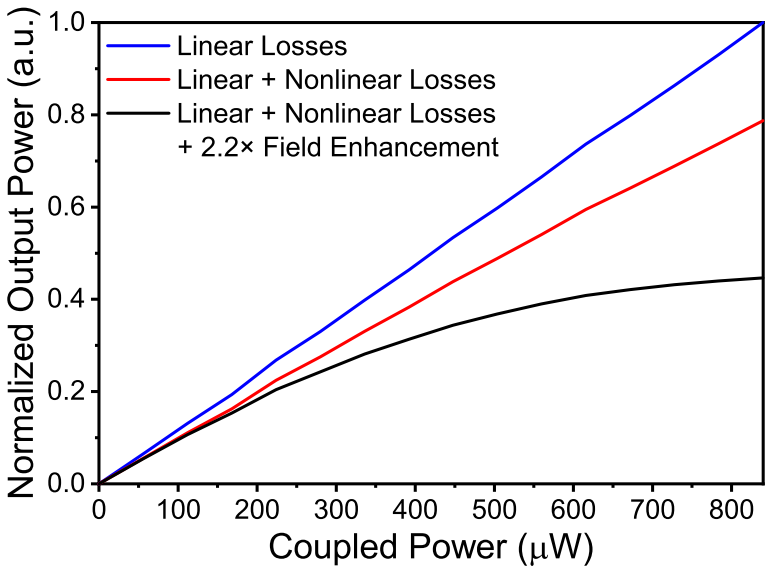

Figure 6.21: Plot of the simulated output fundamental power as a function of the coupled average power for the device considering only linear losses (blue), linear and nonlinear losses (red), and both linear and nonlinear losses in the presence of an additional 2.2× electric field enhancement (black). The data is normalized to the maximum output in the linear case to clearly indicate the magnitude of the nonlinear losses.

To illustrate the magnitude of this additional loss within the present device, the simulated output power is plotted in Fig. 6.21 as a function of the power experimentally coupled to the waveguide for three cases: (1) where no nonlinearity is included and the loss is purely linear, (2) with the inclusion of nonlinear loss and no additional roughness induced electric field enhancement, and (3) with the inclusion of nonlinear loss and the additional 2.2× electric field enhancement introduced by the rough Au film. To indicate the magnitude of the changes, the data is normalized to the output of the linear case at $P^{\omega}_{ave,coupled} = 840\mu$W. Clearly, with only linear losses included, the output power increases linearly with the input power. The introduction of the nonlinearity causes a decrease in the output power of 21.3% at $P^{\omega}_{ave,coupled} = 840\mu$W, but the trend is still predominantly linear. Remarkably, the inclusion of the additional electric field enhancement is found to reduce the output power by 55.4% relative to the linear losses, and the trend is nonlinear. Evidently, effects such as TPA and FCA are clearly significant on this length scale, and cannot be neglected.

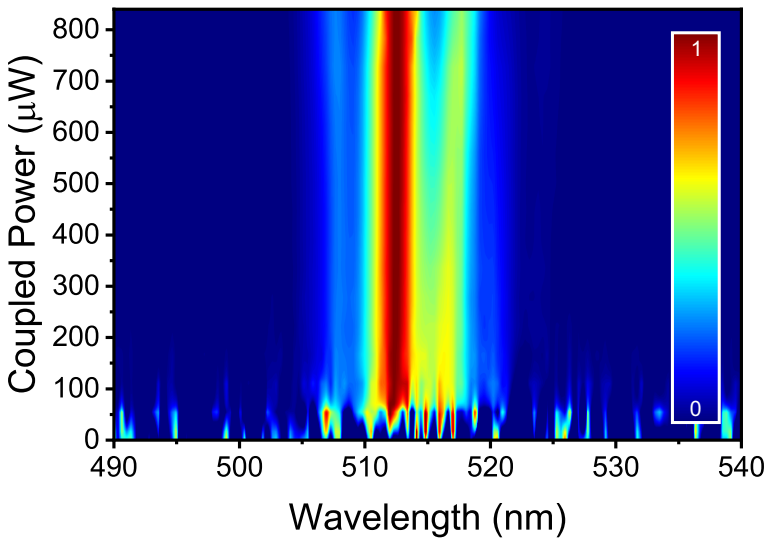

Figure 6.22: Plot of the experimentally measured spectrum as a function of the coupled power, where each spectral slice has been normalized to unity.

6.5.3 Evidence of Self-Phase Modulation

Further insight can be gained by examining the measured THG signal normalized such that the peak of each measured spectral slice is unity. Such a representation would clearly depict any subtle spectral changes. As shown in Fig. 6.22, several interesting details are observed: the spectrum exhibits slight broadening. This can be especially observed by examining the broadening gap between the central peak at $\lambda_{3\omega} = 512.7$nm, and right peak at $\lambda_{3\omega} = 517.7$nm. Second, a slight red-shift is evident in the $\lambda_{3\omega} = 517.7$nm peak.

These spectral characteristics can be further compared with their simulated counterparts. Figure 6.23 depicts the simulated spectrum, including the extra 2.2× electric field enhancement, plotted against the coupled experimental power. Figure 6.23(a) depicts the entire spectrum normalized to its maximum, while Figure 6.23(b) depicts the same spectrum where each spectral slice is normalized to unity. First, the structure of the THG spectrum in the simulations does not share the same three-peaked characteristics as the measurement, due to the fact that the simulated pulse spectrum exhibited a smooth Gaussian distribution, whereas the actual spectrum output by the

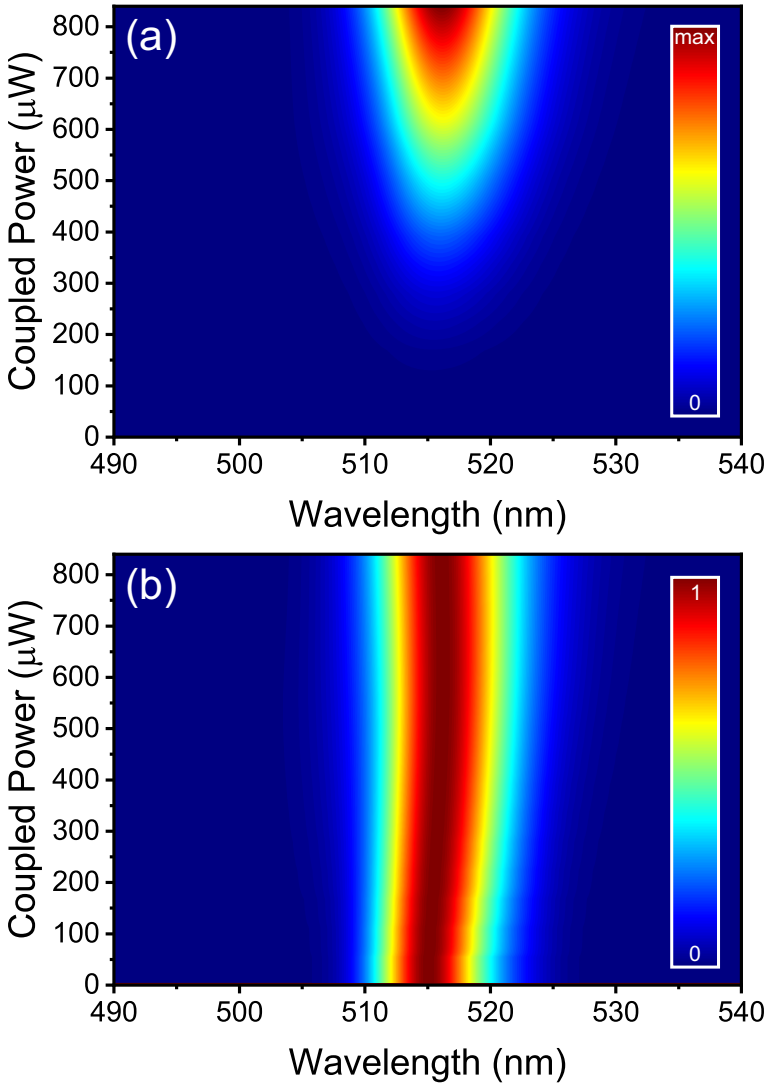

Figure 6.23: Simulated THG spectra plotted against the corresponding coupled power in the experiment, where (a) the spectra normalized to the maximum value, and (b) each spectral slice is normalized to unity.

laser is highly structured. Second, good qualitative agreement is found between the simulated and experimental plots, as the simulated spectra in Figure 6.23(b) also depicts a slight red-shift, as well as a small degree of broadening.

To quantify these characteristics, Fig. 6.24 plots the wavelength of the three peaks of the measured spectrum as a function of the coupled power, as well as the peak of the simulated spectrum. The peak around $\lambda_{3\omega} = 507.6$nm exhibits virtually no change, while the peak around $\lambda_{3\omega} = 512.7$nm shifts 0.36nm from 512.32nm to 512.68nm.

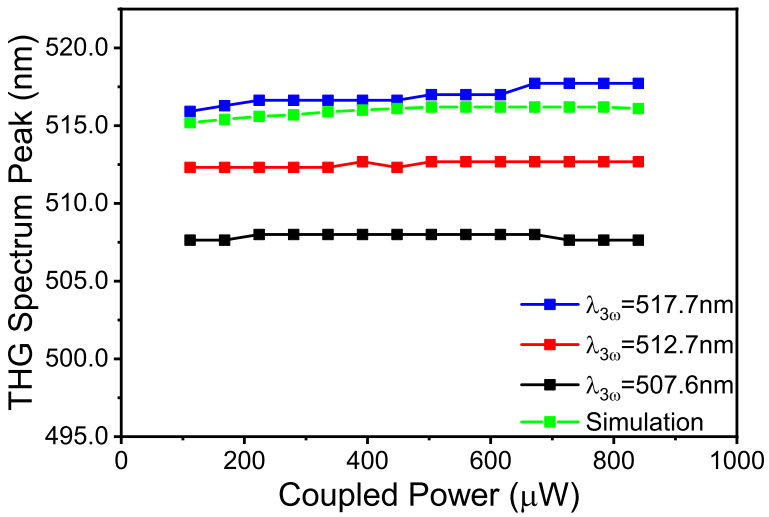

Figure 6.24: Plot of the measured and simulated wavelength of the THG spectral peaks as a function of power coupled into the waveguide.

Since this shift is within the limitation of the resolution of the spectrometer, it is therefore not meaningful. The more dynamic peak is the one residing around $\lambda_{3\omega} =$ 517.7nm, which red-shifts 1.8nm, from 515.92nm to 517.72nm. The simulated values agree most closely with this peak, and exhibit a total red-shift of 1nm from 515.1nm to 516.1nm. This slight red-shift could be due to either spectral modifications of the fundamental pulse that excites the THG, or direct spectral modification of the THG. However, since the THG is absorbed within ∼150nm of propagation (i.e. essentially localized), the latter cannot explain the observed shift. The origin of the red shift must be explained by SPM of the fundamental pulse.

The measured THG spectrum is complicated, and it clearly does not exhibit a Gaussian spectral shape. However, in order to approximate the change in the spectral width as the input power is increased, the FWHM measurement is employed. For a highly structured spectrum such as this, this measurement may not be the most accurate, but it is useful to qualitatively examine any trends. Figure 6.25 displays the FWHM of both the experimentally measured spectra and the simulated spectra. Experimentally, the measured FWHM exhibits a small change, increasing from 6.9nm to 7.1nm as the power is increased. Conversely, the simulations present a much more

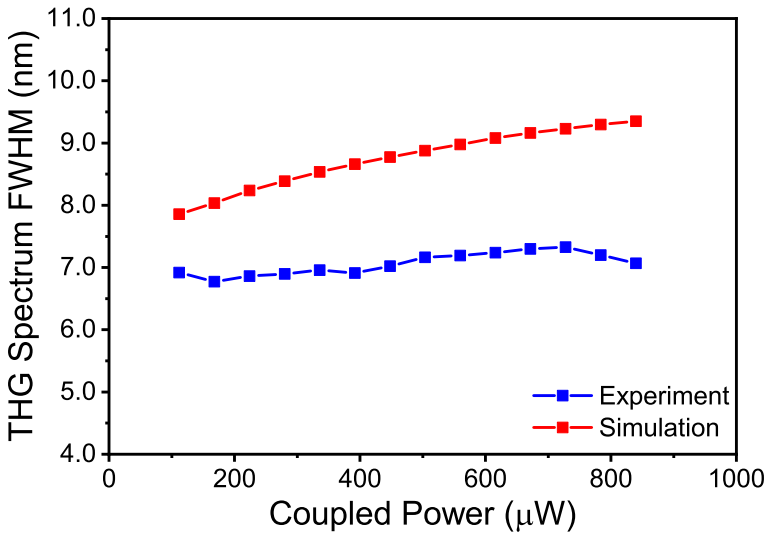

Figure 6.25: Plot of the measured and simulated THG spectrum FWHM as a function of the power coupled into the waveguide.

pronounced spectral broadening, with the THG spectrum FWHM increasing from 7.7nm to 9.4nm.

To summarize this analysis, both the experimental and calculated THG spectra emitted from these waveguides exhibit a small red-shift and slight spectral broadening. This likely arises from the onset of SPM broadening of the fundamental pulses exciting the THG locally within the waveguide. It is therefore important to consider such effects, even in short structures only a couple of micrometers in length.

6.5.4 High Conversion Efficiency Calculation

One of the most crucial metrics in determining the performance of a nonlinear device is the conversion efficiency, η. This indicates the amount of output power generated for a given input power. First, the simulated internal conversion efficiency within the device can be examined as a function of the device length. This is plotted in Fig. 6.26. Remarkably, about 300nm from the input face, for the excitation conditions corresponding to coupled power of $P^{\omega}_{ave,coupled} = 840\mu$W (or correspondingly, a coupled peak power of $P^{\omega}_{pk,coupled} = 98$W), the internal conversion efficiency can reach 1.95%.

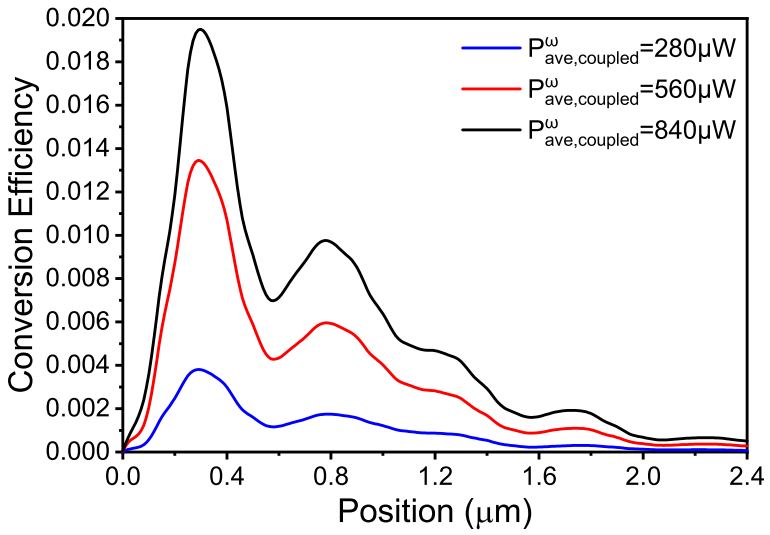

Figure 6.26: Plot of the simulated internal conversion efficiency within the nanoplasmonic waveguide along the length of the device. These simulations account for the additional 2.2× field enhancement.

The conversion efficiency drops off rapidly along the device length, due to the linear and nonlinear attenuation of the fundamental pulse generating the THG, and the high absorption of the visible light. This suggests that the optimal waveguide length for this ME geometry is only about 300nm.

The modulations depicted in the conversion efficiency along the waveguide are related to the coherence length between the fundamental TE_{00} mode excited by the focused fundamental laser pulse, and the third-harmonic modes. For example, the distance between the peak conversion efficiency and the first minimum is 280nm, which agrees well with the $L_{coh} = 276$nm coherence length calculated between the TE_{00} mode at $\lambda_{\omega} = 1550$nm and the TE_{00} mode at $\lambda_{3\omega} = 516.7$nm. Additionally, the distance between the first minimum and the second maximum is 200nm, which agrees with the coherence length $L_{coh} = 217$nm calculated between the TE_{00} mode at $\lambda_{\omega} = 1550$nm and the TE_{01} mode at $\lambda_{3\omega} = 516.7$nm. While the end-fire excitation scheme excites primarily the TE_{00} at $\lambda_{\omega} = 1550$nm, the localized nature of the THG excites several third-harmonic modes.

Since the absorption at $\lambda_{3\omega}$ is high (i.e. the modes only propagate ∼150nm),

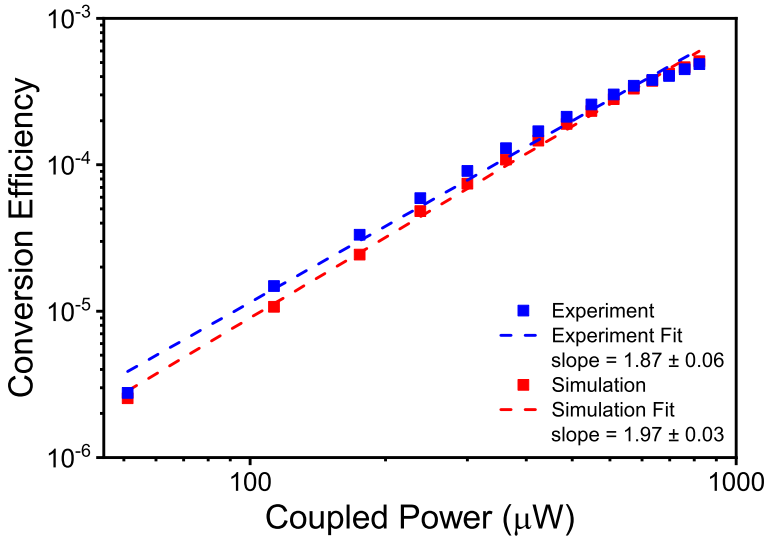

Figure 6.27: Plot of both the measured and simulated internal conversion efficiency at the output of the waveguide.

coherent build up of the THG cannot happen. This is clearly observed in Fig. 6.26 by realizing that the modulations in the conversion efficiency do not go to zero. Such a behaviour is expected to result from the phase mismatch between the fundamental and third-harmonic radiation, when the third-harmonic light is not attenuated as quickly. In this respect, this result reinforces the fact that the THG in this waveguide is a highly localized process.

A plot of the experimentally determined internal conversion efficiency at the waveguide output is shown in Fig. 6.27. Internal conversion efficiencies are quoted, because these are the actual values that would be utilized in an on-chip application. To calculate the measured efficiencies, first the spectrometer counts were integrated, and the calibration factor previously determined in Section 6.3.2 was applied to convert the measured counts to an absolute power of THG radiation within the fiber. The maximum value was found to be $P^{3\omega}_{ave,fiber} = 6.9$nW. Simulations then showed that only 10.2% of the THG light inside the waveguide at the end facet couples out to free space, and only 1.7% of the THG light present inside the waveguide, at the waveguide end facet, would couple into the fiber placed 7.5μm away. Accounting for

these coupling parameters, the THG power generated inside of the waveguide at the end facet was 409.4nW, while the $\lambda_\omega = 1550$nm power coupled into the waveguide to generate green light was $P^{\omega}_{ave,coupled} = 840\mu$W. Since the power is attenuated by both linear and nonlinear means, this is an overestimate of the power that actually excited the measured THG, and therefore presents a conservative estimate of the conversion efficiency. The maximum internal conversion efficiency was found to be $\eta = 4.9\times10^{-4}$. Commonly, the normalized conversion efficiency is quoted to standardize results. The peak-power normalized conversion efficiency is defined as:

$$\eta_{norm} = \frac{\eta}{\left(P^{\omega}_{pk,coupled}\right)^2} \tag{6.4}$$

where here, $P^{\omega}_{pk,coupled} = 98$W. As such, the normalized conversion efficiency is found to be $\eta_{norm} = 5.1 \times 10^{-8}\text{W}^{-2}$. For comparison, the simulated internal conversion efficiency is plotted alongside the experimental results in Fig. 6.27, where good agreement between the two can be seen. Linear fits to both data sets scale with slopes of about 2, as expected. Moreover, at higher powers, the onset of saturation due to TPA and FCA attenuation can be observed in both the experimental and simulated data. The simulations tend to underestimate the measured conversion efficiency at low power, but at high powers η is slightly overestimated. This implies that the Au roughness induced electric field enhancement may be greater than 2.2×. Finally, simulations show that 85.7% of the third-harmonic power produced within this waveguide is generated within the Si core itself.

To place this device's performance within the context of other silicon-based devices presented in literature, the presented conversion efficiency is 7× larger than hybrid-plasmonic nanoantennas [212], 408.3× larger than resonant Si metasurfaces [211], 4900× larger than Si nanodisks [209] and photonic crystal waveguides [213], 17.5× larger than standard SOI waveguides [214], and 21.3× larger than a Au-Si MI

plasmonic waveguide [107]. This is the highest THG conversion efficiency presented within a silicon based structure to date. Furthermore, with a compact core footprint of $0.24\mu m^2$, it is also the most compact of the waveguide based demonstrations.

6.6 Summary

To summarize, the experimental results demonstrated in this chapter represent a significant advancement in nonlinear light generation on a silicon platform. First, through judicial modal engineering, a significant improvement in the THG conversion efficiency was demonstrated, while simultaneously reducing the on-chip footprint of the structure. Such characteristics are vital for integrated optics applications. This unique platform facilitates highly efficient third-harmonic generation within a SOI-based platform, allowing strong visible light signals to be generated. Since the THG is highly localized and does not rely on long-range phase matching, one could conceivably employ excitation sources with different wavelengths, to tailor the third-harmonic emission to a spectral region of interest, without modifications to the waveguide geometry. Efficient nonlinear devices such as these are envisioned to fill critical roles in the generation of light in chip-scale integrated nanoplasmonic networks.

Furthermore, it has been demonstrated that nanoplasmonic waveguides only a few microns in length can still suffer the effects of nonlinear loss mechanisms such as TPA and FCA. Great care has to be taken as devices are miniaturized into the nanoplasmonic regime. These mechanisms suggest a fundamental limit to the amount of light that can be generated within these nanoscale waveguides. Additionally, evidence of the onset of self-phase modulation based spectral broadening was also shown within this compact platform. These provide the potential for compact and unique modulator designs to be explored in the future.

Chapter 7

Conclusions

7.1 Summary

The theoretical and experimental investigations presented within the preceding chapters of this thesis represent crucial steps towards the practical implementation of nonreciprocity and nonlinearity within integrated nanoplasmonic circuits. These studies have examined methods of incorporating magneto-optic effects into nanoplasmonic waveguides, and the unique modulation schemes afforded by the nonlinear temporal dynamics of magnetization. A versatile SOI-based nanoplasmonic platform was fabricated, and nonlinearities within Si-based ME nanoplasmonic waveguides were studied, demonstrating record third-harmonic conversion efficiencies.

The preliminary study of magnetoplasmonic waveguides Chapter 3 demonstrated how magneto-optic effects such as the nonreciprocal phase shift could be introduced within a conceptually simple Bi:YIG-Ag dielectric loaded plasmonic waveguide architecture. This compact architecture was shown to be capable of sizable nonreciprocal phase shifts of 6.99rad/mm, and by passing transient current pulses through the underlying metal, it was demonstrated that the phase could be dynamically modulated by up to 0.33rad, which proved useful for encoding a data signal onto the phase of

the guided mode. The fundamental limitation of this device was the limited propagation length of the mode. The waveguide architecture was improved, and a long-range dielectric loaded surface plasmon waveguide was developed to produce a useful $\pi/2$ nonreciprocal phase shift within its propagation length. This configuration was integrated into an MZI and employed to a small footprint plasmonic isolator, having low insertion losses of 2.51dB and high extinction ratios of 22.82dB.

Combining the established magnetoplasmonic MZI with the insight into dynamic magnetization modulation, the latter sections of Chapter 3 discussed the design of two new optical modulators. The first is analogous to an electrical clock multiplier in digital circuits. This device utilizes the precessional motion of the magnetization within the Bi:YIG core to modulate the output of the interferometer at a rate much faster than the input excitation pulses. It was shown that the clock multiplier modulates the output intensity between 280MHz and 5.6GHz, at a modulation depth of 16.26dB, while simultaneously upscaling the modulation frequency by factors of up to 2.1×10^3 from the input pulse train. The inherent nonlinear temporal dynamics were then exploited by driving the device with sinusoidal signals, resulting in harmonics, frequency splitting, and frequency down-conversion of the RF excitation, as well as frequency mixing when two signals were applied simultaneously. These modulator platforms have no current analog in integrated optics, and bring new functionality to nanoplasmonic circuitry.

With the knowledge and understanding gained by studying nonreciprocal methods of controlling the phase and intensity of a guided nanoplasmonic mode through magneto-optic effects, attention was turned to manipulating another fundamental property of light: the polarization state. The Faraday effect would provide this functionality, however, due to the phase mismatch between photonic and plasmonic modes, this MO phenomenon had previously not been demonstrated in a plasmonic waveguide geometry. The results presented in Chapter 4 demonstrate that through

judicial engineering of a novel hybrid ridge-plasmonic waveguide, this limitation was overcame. Remarkably, 99.4% polarization conversion was achieved within a plasmonic device 830μm in length for the first time via the Faraday effect. Additionally, the previously studied dynamic properties were incorporated to demonstrate dynamic polarization modulation. The device was shown to be operable in two different modes, where application of a train of electrical pulses could either generate pulsed modulation of the polarization state, or continuous modulation at frequencies up to 10GHz.

Finally, the concepts of polarization manipulation via the Faraday effect were applied to design and characterize another more complex, yet equally important optical component: the optical circulator. By combining three distinct waveguide architectures, including standard Si-photonic waveguides, a Si-based L-slot half-wave plate, and a Ce:YIG Faraday rotator, a compact, versatile, and polarization independent optical circulator was demonstrated. It was shown that the optical circulator was capable of providing efficient optical circulation with isolation ratios between 17.8dB-22.5dB, and insertion losses of less than 2.7dB.

With mastery of the principles required to introduce nonreciprocity into plasmonic geometries, focus was allocated to the incorporation of nonlinearity into a similar platform and solving another longstanding challenge within integrated nanoplasmonics: the efficient generation of coherent light within a Si platform. Due to its field enhancement and confinement properties, the metal-encapsulated plasmonic waveguide geometry, which is a superposition of MIM and MI structures, was selected as the architecture of interest. However, due to the inherently high optical losses of such as structure, device lengths had to be kept on the order of a few micrometers. Considering the practical complexities of end-fire excitation and handling of such short waveguides, as well as collecting the transmitted and generated light, presented challenges in the realization and characterization of such structures.

To mediate these challenges, Chapter 5 details the extensive fabrication process de-

velopment undertaken. The developed process utilized nearly all of the available tools within the standard nanofabrication toolkit and combined both micro and nanoscale fabrication techniques. The end product was a versatile process capable of producing ME nanoplasmonic waveguides on thin micromachined characterization beams. The waveguides had lengths of only a few micrometers, and their placement on the characterization beam at the edge of the sample facilitated direct end-fire excitation from a high numerical aperture objective lens, as well as the collection of transmitted light via a tapered and lensed optical fiber. This process was found to be extremely robust and high yield.

Investigations pertaining to the characterization of the nonlinear nanoplasmonic waveguides are detailed within Chapter 6. It was determined that the optimal devices in this scheme are those having narrow widths and short lengths. The most robust device was composed of a Si core only 100nm wide, 340nm tall, and 2.4μm in length, thus making for a total core footprint of only 0.24μm^2. Upon excitation from ultrafast $\lambda_\omega = 1550$nm laser pulses, third-harmonic generation produced bright visible light in the green spectrum, between about 505nm$\leq \lambda_{3\omega} \leq$525nm.

Examining the power scaling trends of the measured spectrum revealed an apparent reduction in the power scaling order. Supporting calculations affirmed that this was due two factors. First, the electric fields within the structure were enhanced by an additional factor of 2.2$\times$ due to the roughness of the granular Au cladding. Second, nonlinear loss attributed to TPA and FCA attenuated the fundamental signal more rapidly than expected within such a short device. Further examination of the measured spectral characteristics showed evidence of the onset of self-phase modulation effects. Simulation results are found to be in very good agreement with experimental results.

Most remarkably, the measured internal THG conversion efficiency of the device was found to be $\eta = 4.9 \times 10^{-4}$, which is over an order of magnitude higher than any

previously presented Si-based waveguide geometries, and was produced in a waveguide device with a more compact footprint. Good agreement with theoretical predictions were reached, and furthermore, simulations predict that the conversion efficiency could reach nearly 2% a distance of 300nm from the input face. Due to the extremely high absorption of the emitted green light within these waveguides, the THG process is extremely localized, and does not rely on phase matching. As such, applying different excitation sources could be utilized to tailor the THG emission to a particular spectral region of interest.

7.2 Future Directions

The research presented in this thesis provides the foundational groundwork for integrated nonreciprocal nanoplasmonic devices, and demonstrates a new architecture for characterizing nonlinearities within Si-based waveguides, as well as improvements in the third-harmonic conversion efficiency. However, there is still numerous investigations worth exploring.

With regards to the magnetoplasmonic devices and modulators, the first extension of the present research would be the experimental demonstration of the designed components and modulators. Specifically, experimental implementation of the unique geometries such as the electrical-to-optical clock multiplier and RF mixer would fulfill crucial niche applications. At present, the facilities available here do not possess the capabilities of processing the MO garnet material required.

While Chapter 4 presents the first implementation of the Faraday effect within a plasmonic waveguide platform, further work on this subject could be undertaken. Improving and further optimizing the waveguide design could lead to higher mode overlap between the TE and TM modes, and reduce the length of the device considerably.

In regards to the nonlinear studies presented within this thesis, further research could be undertaken to improve the conversion efficiency. As suggested by calculations, this configuration may yield unprecedented conversion efficiencies on the order of 2% with lengths of only 300nm. Modification of the waveguide to achieve this would be beneficial.

The fabrication process designed to create these waveguides is flexible in that other material systems could feasibly be employed. First, studies with Ag could be undertaken. Due to the lower optical losses of Ag, this should result in higher fundamental power at the output, and higher third-harmonic power output. Other core materials could conceivably be deposited on top of the micromachined beam platform, which would facilitate studies of nonlinear light generation in a variety of material systems. For example, silicon nitride is a CMOS compatible material that exhibits a well studied intrinsic third-order nonlinearity [221]. Although it is typically considered centrosymmetric and amorphous, and should only exhibit these third-order nonlinear effects, recent studies have shown that by tailoring the deposition parameters, one can observe efficient second-harmonic generation from this material as well [222, 223]. This has been attributed to the formation of small Si nanocrystals within the film that break the centrosymmetry. Due to the localized nature of the harmonic generation in these structures, incorporating silicon nitride into the present architecture could provide a new route to demonstrating efficient second- and third-harmonic generation simultaneously within a more compact platform than previous implementations based on a ring resonator design [223].

7.3 Outlook

The results demonstrated within this thesis open up numerous new pathways for research into nonreciprocal and nonlinear effects within integrated nanoplasmonic

devices. The demonstrated magnetoplasmonic devices offer a new paradigm to control of the fundamental properties of light within a nanoplasmonic architecture. In addition, the micromachined characterization platform that was developed enables nonlinear optical studies within a wide variety of materials and waveguide geometries. Furthermore, the significant enhancement in the third-harmonic conversion efficiency within the Si-based ME nanoplasmonic waveguides provides a means of producing efficient on-chip optical sources, without the need for integration of additional semiconductor material systems. These developments are envisioned to expand the potential of nanoplasmonic devices within integrated optical networks, and facilitate the implementation of future optical computing strategies miniaturized down into the nanoplasmonic regime.

References

[1] S. A. Maier, *Plasmonics : Fundamentals and Applications*, (Springer, New York, 2007).

[2] E. Ozbay, "Plasmonics: Merging photonics and electronics at nanoscale dimensions," *Science* 311(5758), 189–193 (2006).

[3] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface plasmon subwavelength optics," *Nature* 424(6950), 824 (2003).

[4] D. K. Gramotnev and S. I. Bozhevolnyi, "Plasmonics beyond the diffraction limit," *Nat. Photonics* 4(2), 83–91 (2010).

[5] S. Hayashi and T. Okamoto, "Plasmonics: Visit the past to know the future," *J. Phys. D: Appl. Phys.* 45(43), 433001 (2012).

[6] M. I. Stockman, "Nanoplasmonics: Past, present, and glimpse into future," *Opt. Express* 19(22), 22029–22106 (2011).

[7] A. V. Zayats, I. I. Smolyaninov, and A. A. Maradudin, "Nano-optics of surface plasmon polaritons," *Phys. Rep.* 408(3), 131–314 (2005).

[8] Y. Fang and M. Sun, "Nanoplasmonic waveguides: Towards applications in integrated nanophotonic circuits," *Light: Sci. Appl.* 4(6), e294 (2015).

[9] S. I. Bozhevolnyi (Ed.), *Plasmonic Nanoguides and Circuits*, (Pan Stanford, Singapore, 2009).

[10] S. Randhawa, S. Lachèze, J. Renger, A. Bouhelier, R. E. de Lamaestre, A. Dereux, and R. Quidant, "Performance of electro-optical plasmonic ring resonators at telecom wavelengths," *Opt. Express* 20(3), 2354–2362 (2012).

[11] D.-S. Ly-Gagnon, K. C. Balram, J. S. White, P. Wahl, M. L. Brongersma, and D. A. Miller, "Routing and photodetection in subwavelength plasmonic slot waveguides," *Nanophotonics* 1(1), 9–16 (2012).

[12] R.-M. Ma, R. F. Oulton, V. J. Sorger, and X. Zhang, "Plasmon lasers: Coherent light source at molecular scales," *Laser Photon. Rev.* 7(1), 1–21 (2013).

[13] A. Emboras, C. Hoessbacher, C. Haffner, W. Heni, U. Koch, P. Ma, Y. Fedoryshyn, J. Niegemann, C. Hafner, and J. Leuthold, "Electrically controlled plasmonic switches and modulators," *IEEE J. Sel. Top. Quantum Elec* 21(4), 4600408 (2015).

[14] Y. Fu, X. Hu, C. Lu, S. Yue, H. Yang, and Q. Gong, "All-optical logic gates based on nanoscale plasmonic slot waveguides," *Nano Lett.* 12(11), 5784–5790 (2012).

[15] C. Caloz, A. Alù, S. Tretyakov, D. Sounas, K. Achouri, and Z.-L. Deck-Léger, "Electromagnetic nonreciprocity," *Phys. Rev. Appl.* 10(4), 047001 (2018).

[16] G. Armelles, A. Cebollada, A. García-Martín, and M. U. González, "Magnetoplasmonics: Combining magnetic and plasmonic functionalities," *Adv. Opt. Mater.* 1(1), 10–35 (2013).

[17] I. S. Maksymov, "Magneto-plasmonics and resonant interaction of light with dynamic magnetisation in metallic and all-magneto-dielectric nanostructures," *Nanomaterials* 5(2), 577–613 (2015).

[18] M. Kauranen and A. V. Zayats, "Nonlinear plasmonics," *Nat. Photonics* 6(11), 737–748 (2012).

[19] N. C. Panoiu, W. E. I. Sha, D. Y. Lei, and G.-C. Li, "Nonlinear optics in plasmonic nanostructures," *J. Opt.* 20(8), 083001 (2018).

[20] H. D. Young, R. A. Freedman, and A. L. Ford, *Sears and Zemansky's University Physics : With Modern Physics*, (Pearson Addison Wesley, San Francisco, 2008), 12th ed.

[21] E. Hecht, *Optics*, (Addison-Wesley, San Francisco, 2002), 4th ed.

[22] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, (Cambridge University Press, Cambridge, 1999), 7th ed.

[23] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, (Wiley-Interscience, Hoboken, N.J., 2007), 2nd ed.

[24] H. Romer, *Theoretical Optics: An Introduction*, (Wiley-VCH, 2005).

[25] M. A. Ordal, L. L. Long, R. J. Bell, S. E. Bell, R. R. Bell, R. W. Alexander, and C. A. Ward, "Optical properties of the metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the infrared and far infrared," *Appl. Opt.* 22(7), 1099–1119 (1983).

[26] E. Kretschmann and H. Raether, "Radiative decay of non radiative surface plasmons excited by light," *Z. Naturforschung* 23(12), 2135–2136 (1968).

[27] A. Otto, "Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection," *Z. Physik* 216(4), 398–410 (1968).

[28] I. Pockrand, "Resonance anomalies in the light intensity reflected at silver gratings with dielectric coatings," *J. Phys. D: Appl. Phys.* 9(17), 2423–2432 (1976).

[29] T. Holmgaard and S. I. Bozhevolnyi, "Theoretical analysis of dielectric-loaded surface plasmon-polariton waveguides," *Phys. Rev. B* 75(24), 245405 (2007).

[30] A. V. Krasavin and A. V. Zayats, "Silicon-based plasmonic waveguides," *Opt. Express* 18(11), 11791–11799 (2010).

[31] S. Sederberg, V. Van, and A. Y. Elezzabi, "Monolithic integration of plasmonic waveguides into a complimentary metal-oxide-semiconductor- and photonic-compatible platform," *Appl. Phys. Lett.* 96(12), 121101 (2010).

[32] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Phys. Rev. B* 6(12), 4370–4379 (1972).

[33] E. D. Palik, *Handbook of Optical Constants of Solids*, (Academic Press, San Diego, 1998).

[34] A. Kumar, J. Gosciniak, V. S. Volkov, S. Papaioannou, D. Kalavrouziotis, K. Vyrsokinos, J.-C. Weeber, K. Hassan, L. Markey, A. Dereux, T. Tekin, M. Waldow, D. Apostolopoulos, H. Avramopoulos, N. Pleros, and S. I. Bozhevolnyi, "Dielectric-loaded plasmonic waveguide components: Going practical," *Laser Photon. Rev.* 7(6), 938–951 (2013).

[35] J. Gosciniak, L. Markey, A. Dereux, and S. I. Bozhevolnyi, "Efficient thermo-optically controlled Mach-Zhender interferometers using dielectric-loaded plasmonic waveguides," *Opt. Express* 20(15), 16300–16309 (2012).

[36] K. Hassan, J.-C. Weeber, L. Markey, and A. Dereux, "Thermo-optical control of dielectric loaded plasmonic racetrack resonators," *J. Appl. Phys.* 110(2), 023106 (2011).

[37] R. F. Oulton, V. J. Sorger, D. A. Genov, D. F. P. Pile, and X. Zhang, "A hybrid

plasmonic waveguide for subwavelength confinement and long-range propagation," *Nat. Photonics* 2(8), 496–500 (2008).

[38] D. Dai and S. He, "A silicon-based hybrid plasmonic waveguide with a metal cap for a nano-scale light confinement," *Opt. Express* 17(19), 16646–16653 (2009).

[39] M. P. Nielsen and A. Y. Elezzabi, "Silicon-based nanoplasmonic disc resonator on CMOS platform," *Electron. Lett.* 49(19), 1241–1243 (2013).

[40] R. F. Oulton, V. J. Sorger, T. Zentgraf, R.-M. Ma, C. Gladden, L. Dai, G. Bartal, and X. Zhang, "Plasmon lasers at deep subwavelength scale," *Nature* 461(7264), 629–632 (2009).

[41] F. J. Diaz, G. Li, C. M. de Sterke, B. T. Kuhlmey, and S. Palomba, "Kerr effect in hybrid plasmonic waveguides," *J. Opt. Soc. Am. B* 33(5), 957–962 (2016).

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

[43] P. Berini, "Plasmon-polariton waves guided by thin lossy metal films of finite width: Bound modes of asymmetric structures," *Phys. Rev. B* 63(12), 125417 (2001).

[44] L. Chen, X. Li, G. Wang, W. Li, S. Chen, L. Xiao, and D. Gao, "A silicon-based 3-D hybrid long-range plasmonic waveguide for nanophotonic integration," *J. Lightwave Technol.* 30(1), 163–168 (2012).

[45] W. Ma and A. S. Helmy, "Asymmetric long-range hybrid-plasmonic modes in asymmetric nanometer-scale structures," *J. Opt. Soc. Am. B* 31(7), 1723–1729 (2014).

[46] P. Berini, "Long-range surface plasmon polaritons," *Adv. Opt. Photon.* 1(3), 484–588 (2009).

[47] X. Shi, X. Zhang, Z. Han, U. Levy, and S. I. Bozhevolnyi, "CMOS-compatible long-range dielectric-loaded plasmonic waveguides," *J. Lightwave Technol.* 31(21), 3361–3367 (2013).

[48] J. Gosciniak, T. Holmgaard, and S. I. Bozhevolnyi, "Theoretical analysis of long-range dielectric-loaded surface plasmon polariton waveguides," *J. Lightwave Technol.* 29(10), 1473–1481 (2011).

[49] V. S. Volkov, Z. Han, M. G. Nielsen, K. Leosson, H. Keshmiri, J. Gosciniak, O. Albrektsen, and S. I. Bozhevolnyi, "Long-range dielectric-loaded surface plasmon polariton waveguides operating at telecommunication wavelengths," *Opt. Lett.* 36(21), 4278–4280 (2011).

[50] T. Holmgaard, J. Gosciniak, and S. I. Bozhevolnyi, "Long-range dielectric-loaded surface plasmon-polariton waveguides," *Opt. Express* 18(22), 23009–23015 (2010).

[51] J. D. B. Bradley, C. C. Evans, J. T. Choy, O. Reshef, P. B. Deotare, F. Parsy, K. C. Phillips, M. Lončar, and E. Mazur, "Submicrometer-wide amorphous and polycrystalline anatase TiO_2 waveguides for microphotonic devices," *Opt. Express* 20(21), 23821–23831 (2012).

[52] A. Arbabi and L. L. Goddard, "Measurements of the refractive indices and thermo-optic coefficients of Si_3N_4 and SiO using microring resonances," *Opt. Lett.* 38(19), 3878–3881 (2013).

[53] J. A. Dionne, L. A. Sweatlock, H. A. Atwater, and A. Polman, "Plasmon slot waveguides: Towards chip-scale propagation with subwavelength-scale localization," *Phys. Rev. B* 73(3), 035407 (2006).

[54] H. Choo, M.-K. Kim, M. Staffaroni, T. J. Seok, J. Bokor, S. Cabrini, P. J. Schuck, M. C. Wu, and E. Yablonovitch, "Nanofocusing in a metal–insulator–metal gap plasmon waveguide with a three-dimensional linear taper," *Nat. Photonics* 6(12), 838–844 (2012).

[55] M. P. Nielsen, X. Shi, P. Dichtl, S. A. Maier, and R. F. Oulton, "Giant nonlinear response at a plasmonic nanofocus drives efficient four-wave mixing," *Science* 358(6367), 1179–1181 (2017).

[56] C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, and J. Leuthold, "All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale," *Nat. Photonics* 9(8), 525–528 (2015).

[57] D. F. P. Pile, T. Ogawa, D. K. Gramotnev, T. Okamoto, M. Haraguchi, M. Fukui, and S. Matsuo, "Theoretical and experimental investigation of strongly localized plasmons on triangular metal wedges for subwavelength waveguiding," *Appl. Phys. Lett.* 87(6), 061106 (2005).

[58] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, and T. W. Ebbesen, "Channel plasmon-polariton guiding by subwavelength metal grooves," *Phys. Rev. Lett.* 95(4), 046802 (2005).

[59] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J.-Y. Laluet, and T. W. Ebbesen, "Channel plasmon subwavelength waveguide components including interferometers and ring resonators," *Nature* 440(7083), 508 (2006).

[60] O. Lotan, C. L. C. Smith, J. Bar-David, N. A. Mortensen, A. Kristensen, and U. Levy, "Propagation of channel plasmons at the visible regime in aluminum V-groove waveguides," *ACS Photonics* 3(11), 2150–2157 (2016).

[61] M. Faraday, "I. Experimental researches in electricity.—nineteenth series," *Phil. Trans. R. Soc.* 136, 1–20 (1846).

[62] B. Lax and K. J. Button, *Microwave Ferrites and Ferrimagnetics*, (McGraw-Hill, New York, 1962).

[63] R. F. Soohoo, *Microwave Magnetics*, (Harper & Row, New York, 1985).

[64] A. K. Zvezdin and K. A. Kotov, *Modern Magnetooptics and Magnetooptical Materials*, (IOP Publishing, Bristol, 1997).

[65] P. Zeeman, "VII. Doublets and triplets in the spectrum produced by external magnetic forces," *Philos. Mag.* 44(266), 55–60 (1897).

[66] P. Zeeman, "XXXIII. Doublets and triplets in the spectrum produced by external magnetic forces.—(II.)," *Philos. Mag.* 44(268), 255–259 (1897).

[67] G. Scott and D. Lacklison, "Magnetooptic properties and applications of bismuth substituted iron garnets," *IEEE Trans. Magn.* 12(4), 292–311 (1976).

[68] H. Dötsch, N. Bahlmann, O. Zhuromskyy, M. Hammer, L. Wilkens, R. Gerhardt, P. Hertel, and A. F. Popkov, "Applications of magneto-optical waveguides in integrated optics: Review," *J. Opt. Soc. Am. B* 22(1), 240–253 (2005).

[69] S. Yamamoto, Y. Koyamada, and T. Makimoto, "Normal-mode analysis of anisotropic and gyrotropic thin-film waveguides for integrated optics," *J. Appl. Phys.* 43(12), 5090–5097 (1972).

[70] C. Zhang, P. Dulal, B. J. H. Stadler, and D. C. Hutchings, "Monolithically-integrated TE-mode 1D silicon-on-insulator isolators using seedlayer-free garnet," *Sci. Rep.* 7, 5820 (2017).

[71] P. K. Tien, R. J. Martin, R. Wolfe, R. C. Le Craw, and S. L. Blank, "Switching and modulation of light in magneto-optic waveguides of garnet films," *Appl. Phys. Lett.* 21(8), 394–396 (1972).

[72] P. K. Tien, D. P. Schinke, and S. L. Blank, "Magneto-optics and motion of the magnetization in a film-waveguide optical switch," *J. Appl. Phys.* 45(7), 3059–3068 (1974).

[73] J. B. Khurgin, "Optical isolating action in surface plasmon polaritons," *Appl. Phys. Lett.* 89(25), 251115 (2006).

[74] D. Marcuse, "Influence of position of magnetooptic layer on differential phase shift of slab waveguide," *IEEE J. Quantum Electron.* 23(8), 1268–1272 (1987).

[75] Y. Shoji and T. Mizumoto, "Magneto-optical non-reciprocal devices in silicon photonics," *Sci. Technol. Adv. Mater.* 15(1), 014602 (2014).

[76] N. Bahlmann, V. Chandrasekhara, A. Erdmann, R. Gerhardt, P. Hertel, R. Lehmann, D. Salz, F.-J. Schroteler, M. Wallenhorst, and H. Dotsch, "Improved design of magnetooptic rib waveguides for optical isolators," *J. Lightwave Technol.* 16(5), 818–823 (1998).

[77] M. Wallenhorst, M. Niemöller, H. Dötsch, P. Hertel, R. Gerhardt, and B. Gather, "Enhancement of the nonreciprocal magneto-optic effect of TM modes using iron garnet double layers with opposite Faraday rotation," *J. Appl. Phys.* 77(7), 2902–2905 (1995).

[78] J. Kerr, "XLIII. On rotation of the plane of polarization by reflection from the pole of a magnet," *Philos. Mag.* 3(19), 321–343 (1877).

[79] J. Kerr, "XXIV. On reflection of polarized light from the equatorial surface of a magnet," *Philos. Mag.* 5(30), 161–177 (1878).

[80] C. Clavero, K. Yang, J. R. Skuza, and R. A. Lukaszew, "Magnetic-field modulation of surface plasmon polaritons on gratings," *Opt. Lett.* 35(10), 1557–1559 (2010).

[81] V. I. Belotelov, I. A. Akimov, M. Pohl, V. A. Kotov, S. Kasture, A. S. Vengurlekar, A. V. Gopal, D. R. Yakovlev, A. K. Zvezdin, and M. Bayer, "Enhanced magneto-optical effects in magnetoplasmonic crystals," *Nat. Nanotechnol.* 6(6), 370–376 (2011).

[82] J. B. González-Díaz, A. García-Martín, J. M. García-Martín, A. Cebollada, G. Armelles, B. Sepúlveda, Y. Alaverdyan, and M. Käll, "Plasmonic Au/Co/Au nanosandwiches with enhanced magneto-optical activity," *Small* 4(2), 202–205 (2008).

[83] J. B. González-Díaz, A. García-Martín, G. Armelles, J. M. García-Martín, C. Clavero, A. Cebollada, R. A. Lukaszew, J. R. Skuza, D. P. Kumah, and R. Clarke, "Surface-magnetoplasmon nonreciprocity effects in noble-metal/ferromagnetic heterostructures," *Phys. Rev. B* 76(15), 153402 (2007).

[84] G. Armelles, J. B. González-Díaz, A. García-Martín, J. M. García-Martín, A. Cebollada, M. U. González, S. Acimovic, J. Cesario, R. Quidant, and G. Badenes, "Localized surface plasmon resonance effects on the magneto-optical activity of continuous Au/Co/Au trilayers," *Opt. Express* 16(20), 16104–16112 (2008).

[85] J. A. Dionne, K. Diest, L. A. Sweatlock, and H. A. Atwater, "PlasMOStor: A metal-oxide-Si field effect plasmonic modulator," *Nano Lett.* 9(2), 897–902 (2009).

[86] A. V. Krasavin, T. P. Vo, W. Dickson, P. M. Bolger, and A. V. Zayats, "All-

plasmonic modulation via stimulated emission of copropagating surface plasmon polaritons on a substrate with gain," *Nano Lett.* 11(6), 2231–2235 (2011).

[87] S. Sederberg, D. Driedger, M. Nielsen, and A. Y. Elezzabi, "Ultrafast all-optical switching in a silicon-based plasmonic nanoring resonator," *Opt. Express* 19(23), 23494–23503 (2011).

[88] J. Gosciniak, S. I. Bozhevolnyi, T. B. Andersen, V. S. Volkov, J. Kjelstrup-Hansen, L. Markey, and A. Dereux, "Thermo-optic control of dielectric-loaded plasmonic waveguide components," *Opt. Express* 18(2), 1207–1216 (2010).

[89] K. W. Chiu and J. J. Quinn, "Magneto-plasma surface waves in solids," *Il Nuovo Cimento B* 10(1), 1–20 (1972).

[90] V. V. Temnov, G. Armelles, U. Woggon, D. Guzatov, A. Cebollada, A. Garcia-Martin, J.-M. Garcia-Martin, T. Thomay, A. Leitenstorfer, and R. Bratschitsch, "Active magneto-plasmonics in hybrid metal–ferromagnet structures," *Nat. Photonics* 4(2), 107–111 (2010).

[91] V. V. Temnov, "Ultrafast acousto-magneto-plasmonics," *Nat. Photonics* 6(11), 728–736 (2012).

[92] L. Bi, J. Hu, P. Jiang, D. H. Kim, G. F. Dionne, L. C. Kimerling, and C. A. Ross, "On-chip optical isolation in monolithically integrated non-reciprocal optical resonators," *Nat. Photonics* 5(12), 758–762 (2011).

[93] B. J. H. Stadler and A. Gopinath, "Magneto-optical garnet films made by reactive sputtering," *IEEE Trans. Magn.* 36(6), 3957–3961 (2000).

[94] K. J. Chau, S. E. Irvine, and A. Y. Elezzabi, "A gigahertz surface magneto-plasmon optical modulator," *IEEE J. Quantum Electron.* 40(5), 571–579 (2004).

[95] D. Nikolova and A. J. Fisher, "Switching and propagation of magnetoplasmon polaritons in magnetic slot waveguides and cavities," *Phys. Rev. B* 88(12), 125136 (2013).

[96] B. Sepulveda, L. M. Lechuga, and G. Armelles, "Magnetooptic effects in surface-plasmon-polaritons slab waveguides," *J. Lightwave Technol.* 24(2), 945–955 (2006).

[97] E. Ferreiro-Vila, J. M. García-Martín, A. Cebollada, G. Armelles, and M. U. González, "Magnetic modulation of surface plasmon modes in magnetoplasmonic metal-insulator-metal cavities," *Opt. Express* 21(4), 4917–4930 (2013).

[98] A. Davoyan and N. Engheta, "Electrically controlled one-way photon flow in plasmonic nanostructures," *Nat. Commun.* 5, 5250 (2014).

[99] M. Lakshmanan, "The fascinating world of the Landau–Lifshitz–Gilbert equation: An overview," *Phil. Trans. R. Soc. A* 369(1939), 1280–1300 (2011).

[100] L. Landau and E. Lifshitz, "On the theory of the dispersion of magnetic permeability in ferromagnetic bodies," *Physik. Zeits. Sowjetunion* 8, 153–169 (1935).

[101] T. L. Gilbert, "A phenomenological theory of damping in ferromagnetic materials," *IEEE Trans. Magn.* 40(6), 3443–3449 (2004).

[102] R. W. Boyd, *Nonlinear Optics*, (Academic Press, Burlington, 2008), 3rd ed.

[103] P. N. Butcher and D. Cotter, *The Elements of Nonlinear Optics*, (Cambridge University Press, Cambridge, 1990).

[104] A. R. Motamedi, A. H. Nejadmalayeri, A. Khilo, F. X. Kärtner, and E. P. Ippen, "Ultrafast nonlinear optical studies of silicon nanowaveguides," *Opt. Express* 20(4), 4085–4101 (2012).

[105] H. Fukuda, K. Yamada, T. Shoji, M. Takahashi, T. Tsuchizawa, T. Watanabe, J. Takahashi, and S. Itabashi, "Four-wave mixing in silicon wire waveguides," *Opt. Express* 13(12), 4629–4637 (2005).

[106] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.* 23(1), 123–129 (1987).

[107] S. Sederberg and A. Y. Elezzabi, "Coherent visible-light-generation enhancement in silicon-based nanoplasmonic waveguides via third-harmonic conversion," *Phys. Rev. Lett.* 114(22), 227401 (2015).

[108] G. P. Agrawal, *Nonlinear Fiber Optics*, (Elsevier, Burlington, 2013), 5th ed.

[109] M. Borghi, C. Castellan, S. Signorini, A. Trenti, and L. Pavesi, "Nonlinear silicon photonics," *J. Opt.* 19(9), 093002 (2017).

[110] R. Dekker, N. Usechak, M. Först, and A. Driessen, "Ultrafast nonlinear all-optical processes in silicon-on-insulator waveguides," *J. Phys. D: Appl. Phys.* 40(14), R249–R271 (2007).

[111] J. Leuthold, C. Koos, and W. Freude, "Nonlinear silicon photonics," *Nature Photonics* 4(8), 535–544 (2010).

[112] A. M. Weiner, *Ultrafast Optics*, (Wiley, Hoboken, N.J. :, 2009).

[113] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, (Wiley-Interscience, Hoboken, N.J., 2007), 3rd ed.

[114] S. Sederberg and A. Y. Elezzabi, "Nonmonotonic wavelength-dependent power scaling in silicon-on-insulator waveguides via nonlinear optical effect conglomeration," *ACS Photonics* 1(7), 576–581 (2014).

[115] Q. Lin, O. J. Painter, and G. P. Agrawal, "Nonlinear optical phenomena in silicon waveguides: Modeling and applications," *Opt. Express* 15(25), 16604–16644 (2007).

[116] C. Altman and K. Suchy, *Reciprocity, Spatial Mapping and Time Reversal in Electromagnetics*, (Springer, Dordrecht, 2011), 2nd ed.

[117] S. E. Irvine and A. Y. Elezzabi, "Modeling of high-speed magnetooptic beam deflection," *IEEE J. Quantum Electron.* 38(10), 1428–1435 (2002).

[118] A. Y. Elezzabi and M. R. Freeman, "Ultrafast magneto-optic sampling of picosecond current pulses," *Appl. Phys. Lett.* 68(25), 3546–3548 (1996).

[119] "Lumerical Inc." https://www.lumerical.com/.

[120] H. Kurebayashi, O. Dzyapko, V. E. Demidov, D. Fang, A. J. Ferguson, and S. O. Demokritov, "Controlled enhancement of spin-current emission by three-magnon splitting," *Nat. Mater.* 10(9), 660–664 (2011).

[121] A. Melikyan, L. Alloatti, A. Muslija, D. Hillerkuss, P. C. Schindler, J. Li, R. Palmer, D. Korn, S. Muehlbrandt, D. Van Thourhout, B. Chen, R. Dinu, M. Sommer, C. Koos, M. Kohl, W. Freude, and J. Leuthold, "High-speed plasmonic phase modulators," *Nat. Photonics* 8(3), 229–233 (2014).

[122] T. Shintaku, T. Uno, and M. Kobayashi, "Magneto-optic channel waveguides in Ce-substituted yttrium iron garnet," *J. Appl. Phys.* 74(8), 4877–4881 (1993).

[123] M. Levy, M. H. Hu, R. Scarmozzino, R. M. Osgood Jr., H. Hegde, F. J. Cadieu, and V. J. Fratello, "Sputtered-magnet Mach-Zehnder waveguide isolator," in *Lasers and Electro-Optics Society Annual Meeting*, (IEEE, Boston, MA, 1996), vol. 2, pp. 232–233.

[124] J. Fujita, M. Levy, R. M. Osgood Jr., L. Wilkens, H. Dötsch, C. Gutierrez, and M. Matheaus, "Integrated waveguide Mach-Zehnder isolator using the transverse magneto-optical phase-shift," in *Integrated Photonics Research*, (Optical Society of America, Quebec City, Canada, 2000), vol. 45, p. IFF1.

[125] J. Fujita, M. Levy, R. M. Osgood Jr., L. Wilkens, and H. Dötsch, "Waveguide optical isolator based on Mach–Zehnder interferometer," *Appl. Phys. Lett.* 76(16), 2158–2160 (2000).

[126] J. S. Yang, J. W. Roh, W. Y. Lee, S. H. Ok, D. H. Woo, Y. T. Byun, Y. M. Jhon, M. T, and S. Lee, "A magneto-optic waveguide isolator using multimode interference effect," *J. Magn.* 10(2), 41–43 (2005).

[127] H. Hegde, S. U. Jen, K. Chen, and F. J. Cadieu, "Film Sm–Co permanent magnets for the biasing of thin permalloy strips," *J. Appl. Phys.* 73(10), 5926–5928 (1993).

[128] F. J. Cadieu, T. D. Cheung, S. H. Aly, L. Wickramasekara, and R. G. Pirich, "Square hysterisis loop $SmCo_5$ films synthesized by selectively thermalized sputtering," *IEEE Trans. Magn.* 19(5), 2038–2040 (1983).

[129] J. L. Melchor, W. P. Ayres, and P. H. Vartanian, "Microwave frequency doubling from 9 to 18 KMC in ferrites," *Proc. IRE* 45(5), 643–646 (1957).

[130] R. L. Jepsen, "Harmonic generation and frequency mixing in ferromagnetic insulators," *J. Appl. Phys.* 32(12), 2627–2630 (1961).

[131] J. Marsh and R. E. Camley, "Two-wave mixing in nonlinear magnetization dynamics: A perturbation expansion of the Landau-Lifshitz-Gilbert equation," *Phys. Rev. B* 86(22), 224405 (2012).

[132] A. Y. Elezzabi and S. E. Irvine, "Parametrically driven first-order Suhl instability and nonlinearities in bismuth-substituted yttrium iron garnet films," *Appl. Phys. Lett.* 82(15), 2464–2466 (2003).

[133] M. I. Hayee, M. C. Cardakli, A. B. Sahin, and A. E. Willner, "Doubling of bandwidth utilization using two orthogonal polarizations and power unbalancing in a polarization-division-multiplexing scheme," *IEEE Photonic Tech. L.* 13(8), 881–883 (2001).

[134] S. Y. Wang, S. H. Lin, and Y. M. Houng, "GaAs traveling-wave polarization electro-optic waveguide modulator with bandwidth in excess of 20 GHz at 1.3 Mm," *Appl. Phys. Lett.* 51(2), 83–85 (1987).

[135] S. E. Irvine and A. Y. Elezzabi, "Wideband magneto-optic modulation in a bismuth-substituted yttrium iron garnet waveguide," *Opt. Commun.* 220(4), 325–329 (2003).

[136] L. J. Aplet and J. W. Carson, "A Faraday effect optical isolator," *Appl. Opt.* 3(4), 544–545 (1964).

[137] D. Dai, L. Liu, S. Gao, D.-X. Xu, and S. He, "Polarization management for silicon photonic integrated circuits," *Laser Photon. Rev.* 7(3), 303–328 (2013).

[138] D. Dai, J. Bauters, and J. E. Bowers, "Passive technologies for future large-scale photonic integrated circuits on silicon: Polarization handling, light non-reciprocity and loss reduction," *Light: Sci. Appl.* 1(3), e1 (2012).

[139] F. Ding, Z. Wang, S. He, V. M. Shalaev, and A. V. Kildishev, "Broadband high-efficiency half-wave plate: A supercell-based plasmonic metasurface approach," *ACS Nano* 9(4), 4111–4119 (2015).

[140] Z. H. Jiang, L. Lin, D. Ma, S. Yun, D. H. Werner, Z. Liu, and T. S. Mayer, "Broadband and wide field-of-view plasmonic metasurface-enabled waveplates," *Sci. Rep.* 4, 7511 (2014).

[141] A. Roberts and L. Lin, "Plasmonic quarter-wave plate," *Opt. Lett.* 37(11), 1820–1822 (2012).

[142] J. Li, S. Chen, H. Yang, J. Li, P. Yu, H. Cheng, C. Gu, H.-T. Chen, and J. Tian, "Simultaneous control of light polarization and phase distributions using plasmonic metasurfaces," *Adv. Funct. Mater.* 25(5), 704–710 (2015).

[143] P. Chang, C. Lin, and A. S. Helmy, "Polarization engineering in nanoscale waveguides using lossless media," *J. Lightwave Technol.* 34(3), 952–960 (2016).

[144] C. Brooks, P. E. Jessop, H. Deng, D. O. Yevick, and G. Tarr, "Passive silicon-on-insulator polarization-rotating waveguides," *Opt. Eng.* 45(4), 044603 (2006).

[145] J. Zhang, S. Zhu, H. Zhang, S. Chen, G.-Q. Lo, and D.-L. Kwong, "An ultracompact surface plasmon polariton-effect-based polarization rotator," *IEEE Photonic Tech. L.* 23(21), 1606–1608 (2011).

[146] Y.-J. Chang and T.-H. Yu, "Photonic-quasi-TE-to-hybrid-plasmonic-TM polarization mode converter," *J. Lightwave Technol.* 33(20), 4261–4267 (2015).

[147] S. Kim and M. Qi, "Mode-evolution-based polarization rotation and coupling between silicon and hybrid plasmonic waveguides," *Sci. Rep.* 5, 18378 (2015).

[148] J. N. Caspers, M. Z. Alam, and M. Mojahedi, "Compact hybrid plasmonic polarization rotator," *Opt. Lett.* 37(22), 4615–4617 (2012).

[149] J. N. Caspers, J. S. Aitchison, and M. Mojahedi, "Experimental demonstration of an integrated hybrid plasmonic polarization rotator," *Opt. Lett.* 38(20), 4054–4057 (2013).

[150] L. Jin, Q. Chen, and L. Wen, "Mode-coupling polarization rotator based on plasmonic waveguide," *Opt. Lett.* 39(9), 2798–2801 (2014).

[151] K. Hassan, F. Leroy, G. Colas-des-Francs, and J.-C. Weeber, "Dihedron dielectric loaded surface plasmon athermal polarization converter," *Opt. Lett.* 39(3), 697–700 (2014).

[152] S. Kim and M. Qi, "Polarization rotation and coupling between silicon waveguide and hybrid plasmonic waveguide," *Opt. Express* 23(8), 9968–9978 (2015).

[153] L. Gao, Y. Huo, K. Zang, S. Paik, Y. Chen, J. S. Harris, and Z. Zhou, "On-chip plasmonic waveguide optical waveplate," *Sci. Rep.* 5, 15794 (2015).

[154] B. Hu, Y. Zhang, and Q. J. Wang, "Surface magneto plasmons and their applications in the infrared frequencies," *Nanophotonics* 4(4), 383–396 (2015).

[155] J. Y. Chin, T. Steinle, T. Wehlus, D. Dregely, T. Weiss, V. I. Belotelov, B. Stritzker, and H. Giessen, "Nonreciprocal plasmonics enables giant enhancement of thin-film Faraday rotation," *Nat. Commun.* 4, 2609 (2013).

[156] D. Floess, J. Y. Chin, A. Kawatani, D. Dregely, H.-U. Habermeier, T. Weiss, and H. Giessen, "Tunable and switchable polarization rotation with nonreciprocal plasmonic thin films at designated wavelengths," *Light: Sci. Appl.* 4(5), e284 (2015).

[157] H. Uchida, Y. Masuda, R. Fujikawa, A. V. Baryshev, and M. Inoue, "Large enhancement of Faraday rotation by localized surface plasmon resonance in Au nanoparticles embedded in Bi:YIG film," *J. Magn. Magn. Mater.* 321(7), 843–845 (2009).

[158] A. B. Khanikaev, A. V. Baryshev, A. A. Fedyanin, A. B. Granovsky, and M. In-

oue, "Anomalous Faraday effect of a system with extraordinary optical transmittance," *Opt. Express* 15(11), 6612–6622 (2007).

[159] B. Caballero, A. García-Martín, and J. C. Cuevas, "Faraday effect in hybrid magneto-plasmonic photonic crystals," *Opt. Express* 23(17), 22238–22249 (2015).

[160] S.-K. Liaw, W. Y. Jang, C.-J. Wang, and K. L. Hung, "Pump efficiency improvement of a C-band tunable fiber laser using optical circulator and tunable fiber gratings," *Appl. Opt.* 46(12), 2280–2285 (2007).

[161] M. Scheucher, A. Hilico, E. Will, J. Volz, and A. Rauschenbeutel, "Quantum optical circulator controlled by a single chirally coupled atom," *Science* 354(6319), 1577–1580 (2016).

[162] A. V. Tran, W. D. Zhong, R. C. Tucker, and R. Lauder, "Optical add-drop multiplexers with low crosstalk," *IEEE Photonic Tech. L.* 13(6), 582–584 (2001).

[163] H. H. Lu, A. S. Patra, W. J. Ho, P. C. Lai, and M. H. Shiu, "A full-duplex radio-over-fiber transport system based on FP laser diode with OBPF and optical circulator with fiber bragg grating," *IEEE Photonic Tech. L.* 19(20), 1652–1654 (2007).

[164] E. J. Denlinger, "Design of partial height ferrite waveguide circulators," *IEEE Trans. Microw. Theory Tech.* 22(8), 810–813 (1974).

[165] C. E. Fay and R. L. Comstock, "Operation of the ferrite junction circulator," *IEEE Trans. Microw. Theory Tech.* 13(1), 15–27 (1965).

[166] E. K. N. Yung, D. G. Zhang, and R. S. K. Wong, "A novel waveguide Y-junction circulator with a ferrite sphere for millimeter waves," *IEEE Trans. Microw. Theory Tech.* 44(3), 454–456 (1996).

[167] E. A. Ohm, "A broad-band microwave circulator," *IRE Trans. Microw. Theory Tech.* 4(4), 210–217 (1956).

[168] S. Saito, K. Yokoyama, and Y. Fujii, "A light circulator using the Faraday effect of heavy flint glass," *Proc. IEEE* 52(8), 979–979 (1964).

[169] A. Shibukawa and M. Kobayashi, "Compact optical circulator for near-infrared region," *Electronics Letters* 14(25), 816–817 (1978).

[170] Y. Fujii, "High-isolation polarization-independent optical circulator coupled with single-mode fibers," *J. Lightwave Technol.* 9(4), 456–460 (1991).

[171] Y. Fujii, "Compact high-isolation polarization-independent optical circulator," *Opt. Lett.* 18(3), 250–252 (1993).

[172] N. Sugimoto, T. Shintaku, A. Tate, H. Terui, M. Shimokozono, E. Kubota, M. Ishii, and Y. Inoue, "Waveguide polarization-independent optical circulator," *IEEE Photonic Tech. L.* 11(3), 355–357 (1999).

[173] T. Mizumoto, H. Chihara, N. Tokui, and Y. Naito, "Verification of waveguide-type optical circulator operation," *Electron. Lett.* 26(3), 199–200 (1990).

[174] R. Takei and T. Mizumoto, "Design and simulation of silicon waveguide optical circulator employing nonreciprocal phase shift," *Jpn. J. Appl. Phys.* 49(5R), 052203 (2010).

[175] K. Mitsuya, Y. Shoji, and T. Mizumoto, "Demonstration of a silicon waveguide optical circulator," *IEEE Photonic Tech. L.* 25(8), 721–723 (2013).

[176] S. Ghosh, S. Keyvaninia, W. V. Roy, T. Mizumoto, G. Roelkens, and R. Baets, "Adhesively bonded Ce:YIG/SOI integrated optical circulator," *Opt. Lett.* 38(6), 965–967 (2013).

[177] P. Pintus, D. Huang, C. Zhang, Y. Shoji, T. Mizumoto, and J. E. Bowers, "Microring-based optical isolator and circulator with integrated electromagnet for silicon photonics," *J. Lightwave Technol.* 35(8), 1429–1437 (2017).

[178] D. Huang, P. Pintus, C. Zhang, P. Morton, Y. Shoji, T. Mizumoto, and J. E. Bowers, "Dynamically reconfigurable integrated optical circulators," *Optica* 4(1), 23–30 (2017).

[179] Z. Wang and S. Fan, "Optical circulators in two-dimensional magneto-optical photonic crystals," *Opt. Lett.* 30(15), 1989–1991 (2005).

[180] A. R. Davoyan and N. Engheta, "Nonreciprocal rotating power flow within plasmonic nanostructures," *Phys. Rev. Lett.* 111(4), 047401 (2013).

[181] A. R. Davoyan and N. Engheta, "Nanoscale plasmonic circulator," *New J. Phys.* 15(8), 083054 (2013).

[182] K. Yayoi, K. Tobinaga, Y. Kaneko, A. V. Baryshev, and M. Inoue, "Optical waveguide circulators based on two-dimensional magnetophotonic crystals: Numerical simulation for structure simplification and experimental verification," *J. Appl. Phys.* 109(7), 07B750 (2011).

[183] T. R. Zaman, X. Guo, and R. J. Ram, "Proposal for a polarization-independent integrated optical circulator," *IEEE Photonic Tech. L.* 18(12), 1359–1361 (2006).

[184] V. P. Tzolov and M. Fontaine, "A passive polarization converter free of longitudinally-periodic structure," *Opt. Commun.* 127(1), 7–13 (1996).

[185] Z. Huan, R. Scarmozzino, G. Nagy, J. Steel, and R. M. Osgood, "Realization of a compact and single-mode optical passive polarization converter," *IEEE Photonic Tech. L.* 12(3), 317–319 (2000).

[186] Z. Wang and D. Dai, "Ultrasmall Si-nanowire-based polarization rotator," *J. Opt. Soc. Am. B* 25(5), 747–753 (2008).

[187] D. C. Hutchings and B. M. Holmes, "A waveguide polarization toolset design based on mode beating," *IEEE Photonics J.* 3(3), 450–461 (2011).

[188] F. Morichetti, A. Melloni, M. Martinelli, R. G. Heideman, A. Leinse, D. H. Geuzebroek, and A. Borreman, "Box-shaped dielectric waveguides: A new concept in integrated optics?" *J. Lightwave Technol.* 25(9), 2579–2589 (2007).

[189] Y. Fu, T. Ye, W. Tang, and T. Chu, "Efficient adiabatic silicon-on-insulator waveguide taper," *Photon. Res.* 2(3), A41–A44 (2014).

[190] P. Sethi, A. Haldar, and S. K. Selvaraja, "Ultra-compact low-loss broadband waveguide taper in silicon-on-insulator," *Opt. Express* 25(9), 10196–10203 (2017).

[191] M. C. Onbasli, L. Beran, M. Zahradník, M. Kučera, R. Antoš, J. Mistrík, G. F. Dionne, M. Veis, and C. A. Ross, "Optical and magneto-optical behavior of Cerium Yttrium Iron Garnet thin films at wavelengths of 200–1770 nm," *Sci. Rep.* 6, 23640 (2016).

[192] Z. Lu, H. Yun, Y. Wang, Z. Chen, F. Zhang, N. A. F. Jaeger, and L. Chrostowski, "Broadband silicon photonic directional coupler using asymmetric-waveguide based phase control," *Opt. Express* 23(3), 3795–3808 (2015).

[193] P.-H. Fu, Y.-C. Tu, and D.-W. Huang, "Broadband optical waveguide couplers with arbitrary coupling ratios designed using a genetic algorithm," *Opt. Express* 24(26), 30547–30561 (2016).

[194] S. Toyama, N. Doumae, A. Shoji, and Y. Ikariyama, "Design and fabrication of

a waveguide-coupled prism device for surface plasmon resonance sensor," *Sens. Actuator B-Chem.* 65(1), 32–34 (2000).

[195] P. Kumar, V. K. Tripathi, A. Kumar, and X. Shao, "Launching focused surface plasmon in circular metallic grating," *J. Appl. Phys.* 117(1), 013103 (2015).

[196] M. P. Nielsen, A. Ashfar, K. Cadien, and A. Y. Elezzabi, "Plasmonic materials for metal–insulator–semiconductor–insulator–metal nanoplasmonic waveguides on silicon-on-insulator platform," *Opt. Mater.* 36(2), 294–298 (2013).

[197] C. Lin, H. M. K. Wong, B. Lau, M. A. Swillam, and A. S. Helmy, "Efficient broadband energy transfer via momentum matching at hybrid junctions of guided-waves," *Appl. Phys. Lett.* 101(12), 123115 (2012).

[198] B. Lau, M. A. Swillam, and A. S. Helmy, "Hybrid orthogonal junctions: Wideband plasmonic slot-silicon waveguide couplers," *Opt. Express* 18(26), 27048–27059 (2010).

[199] G. I. Stegeman, R. F. Wallis, and A. A. Maradudin, "Excitation of surface polaritons by end-fire coupling," *Opt. Lett.* 8(7), 386–388 (1983).

[200] S. Sederberg and A. Y. Elezzabi, "Ponderomotive electron acceleration in a silicon-based nanoplasmonic waveguide," *Phys. Rev. Lett.* 113(16), 167401 (2014).

[201] S. Sederberg and A. Y. Elezzabi, "Integration of silicon-loaded nanoplasmonic waveguides onto a micro-machined characterization beam for nonlinear optics applications," *Opt. Mater.* 48, 150–155 (2015).

[202] M. J. Madou, *Fundamentals of Microfabrication and Nanotechnology*, (CRC Press, Boca Raton, 2012), 3rd ed.

[203] H. W. K. Tom, T. F. Heinz, and Y. R. Shen, "Second-harmonic reflection from silicon surfaces and its relation to structural symmetry," *Phys. Rev. Lett.* 51(21), 1983–1986 (1983).

[204] B. D. Choudhury, P. K. Sahoo, R. Sanatinia, G. Andler, S. Anand, and M. Swillo, "Surface second harmonic generation from silicon pillar arrays with strong geometrical dependence," *Opt. Lett.* 40(9), 2072–2075 (2015).

[205] W. Daum, H.-J. Krause, U. Reichel, and H. Ibach, "Identification of strained silicon layers at Si-SiO_2 interfaces and clean Si surfaces by nonlinear optical spectroscopy," *Phys. Rev. Lett.* 71(8), 1234–1237 (1993).

[206] M. Cazzanelli, F. Bianco, E. Borga, G. Pucker, M. Ghulinyan, E. Degoli, E. Luppi, V. Véniard, S. Ossicini, D. Modotto, S. Wabnitz, R. Pierobon, and L. Pavesi, "Second-harmonic generation in silicon waveguides strained by silicon nitride," *Nat. Mater.* 11(2), 148–154 (2012).

[207] C. Castellan, A. Trenti, C. Vecchi, A. Marchesini, M. Mancinelli, M. Ghulinyan, G. Pucker, and L. Pavesi, "On the origin of second harmonic generation in silicon waveguides with silicon nitride cladding," *Sci. Rep.* 9(1), 1088 (2019).

[208] E. Timurdogan, C. V. Poulton, M. J. Byrd, and M. R. Watts, "Electric field-induced second-order nonlinear optical effects in silicon waveguides," *Nat. Photonics* 11(3), 200–206 (2017).

[209] M. R. Shcherbakov, D. N. Neshev, B. Hopkins, A. S. Shorokhov, I. Staude, E. V. Melik-Gaykazyan, M. Decker, A. A. Ezhov, A. E. Miroshnichenko, I. Brener, A. A. Fedyanin, and Y. S. Kivshar, "Enhanced third-harmonic generation in silicon nanoparticles driven by magnetic response," *Nano Lett.* 14(11), 6488–6492 (2014).

[210] W. Tong, C. Gong, X. Liu, S. Yuan, Q. Huang, J. Xia, and Y. Wang, "Enhanced third harmonic generation in a silicon metasurface using trapped mode," *Opt. Express* 24(17), 19661–19670 (2016).

[211] Y. Yang, W. Wang, A. Boulesbaa, I. I. Kravchenko, D. P. Briggs, A. Puretzky, D. Geohegan, and J. Valentine, "Nonlinear fano-resonant dielectric metasurfaces," *Nano Lett.* 15(11), 7388–7393 (2015).

[212] T. Shibanuma, G. Grinblat, P. Albella, and S. A. Maier, "Efficient third harmonic feneration from metal–dielectric hybrid nanoantennas," *Nano Lett.* 17(4), 2647–2651 (2017).

[213] B. Corcoran, C. Monat, C. Grillet, D. J. Moss, B. J. Eggleton, T. P. White, L. O'Faolain, and T. F. Krauss, "Green light emission in silicon through slow-light enhanced third-harmonic generation in photonic-crystal waveguides," *Nat. Photonics* 3(4), 206–210 (2009).

[214] S. Sederberg, C. J. Firby, and A. Y. Elezzabi, "Efficient, broadband third-harmonic generation in silicon nanophotonic waveguides spectrally shaped by nonlinear propagation," *Opt. Express* 27(4), 4990–5004 (2019).

[215] N. Suzuki, "FDTD analysis of two-photon absorption and free-carrier absorption in si high-index-contrast waveguides," *J. Lightwave Technol.* 25(9), 2495–2501 (2007).

[216] J. Renger, R. Quidant, N. van Hulst, and L. Novotny, "Surface-enhanced nonlinear four-wave mixing," *Phys. Rev. Lett.* 104(4), 046803 (2010).

[217] D. Milam, "Review and assessment of measured values of the nonlinear refractive-index coefficient of fused silica," *Appl. Opt.* 37(3), 546–550 (1998).

[218] V. Grubsky and A. Savchenko, "Glass micro-fibers for efficient third harmonic generation," *Opt. Express* 13(18), 6798–6806 (2005).

[219] C. Koos, L. Jacome, C. Poulton, J. Leuthold, and W. Freude, "Nonlinear silicon-on-insulator waveguides for all-optical signal processing," *Opt. Express* 15(10), 5976–5990 (2007).

[220] R. M. Osgood, N. C. Panoiu, J. I. Dadap, X. Liu, X. Chen, I.-W. Hsieh, E. Dulkeith, W. M. J. Green, and Y. A. Vlasov, "Engineering nonlinearities in nanoscale optical systems: Physics and applications in dispersion-engineered silicon nanophotonic wires," *Adv. Opt. Photon.* 1(1), 162–235 (2009).

[221] T. Ning, O. Hyvärinen, H. Pietarinen, T. Kaplas, M. Kauranen, and G. Genty, "Third-harmonic UV generation in silicon nitride nanostructures," *Opt. Express* 21(2), 2012–2017 (2013).

[222] T. Ning, H. Pietarinen, O. Hyvärinen, J. Simonen, G. Genty, and M. Kauranen, "Strong second-harmonic generation in silicon nitride films," *Appl. Phys. Lett.* 100(16), 161902 (2012).

[223] J. S. Levy, M. A. Foster, A. L. Gaeta, and M. Lipson, "Harmonic generation in silicon nitride ring resonators," *Opt. Express* 19(12), 11415–11421 (2011).

[224] L. W. Hallman, B. Ryvkin, K. Haring, S. Ranta, T. Leinonen, and J. Kostamovaara, "Asymmetric waveguide laser diode operated in gain switching mode with high-power optical pulse generation," *Electron. Lett.* 46(1), 65–66 (2010).

[225] L. W. Hallman, K. Haring, L. Toikkanen, T. Leinonen, B. S. Ryvkin, and J. T. Kostamovaara, "3 nJ, 100 ps laser pulses generated with an asymmetric waveguide laser diode for a single-photon avalanche diode time-of-flight (SPAD TOF) rangefinder application," *Meas. Sci. Technol.* 23(2), 025202 (2012).

[226] S. Vainshtein, J. Kostamovaara, Y. Sveshnikov, S. Gurevich, M. Kulagina, V. Yuferev, L. Shestak, and M. Sverdlov, "Superfast high-current switching of GaAs avalanche transistor," *Electron. Lett.* 40(1), 85–86 (2004).

[227] W. A. Davis and K. Agarwal, *Radio Frequency Circuit Design*, (Wiley, New York, 2001).

[228] "COMSOL Multiphysics Modeling Software," https://www.comsol.com.

Appendix A

Modeling MO Materials in Lumerical FDTD Solutions

Defining Magneto-Optic Materials

In order to implement magneto-optic materials within *Lumerical FDTD Solutions* [119], auxiliary script files had to be developed since the material models do not inherently account for the full anisotropic permittivity matrix; they can only natively account for diagonal anisotropy. In order to model the full permittivity tensor of the material, $\hat{\varepsilon}_r$, the software implementation requires it to be decomposed into a diagonal permittivity matrix, ε_D and a unitary rotation matrix U, such that $\varepsilon_D = U\hat{\varepsilon}_r U^*$, where U^* is the complex conjugate of the transpose of the matrix U. With these parameters, the diagonal permittivity, ε_D, or correspondingly the diagonal refractive index matrix n_D, can be integrated into the material model. The matrix U then gets added to a grid rotation element, which is used by the structure. For accurate results to be returned, a second isotropic version of the material had to be defined, and overlapped with the MO material to form a single mesh cell wide isotropic border around the MO material. It was found that having MO materials in mesh cells adjacent to

other materials having a high index contrast could cause numerical divergence of the fields and produce erroneous results. This border mesh cell could be reduced to a few nanometers, such that it does not impact the overall results. Furthermore, it was determined that optical sources should not be injected directly into these MO materials. Sources should be placed within a few mesh cells of isotropic material in order to achieve correct output. Over a period of eight months before implementation into the studies contained within this thesis, this method was rigorously tested and validated against theoretical and published results in collaboration with the engineers and developers at *Lumerical Inc.*.

The following script file, *MOmaterials.lsf*, implements this method to create two oppositely magnetized MO materials. It assumes basic parameters such as the wavelength, refractive index, specific Faraday rotation, magnetization axis, and material names have been predefined within the model. The script employs these to calculate the appropriate matrices, and inserts them appropriately. The code is set up for magnetization along a single axis only; however, this is adequate for most simulations.

MOmaterials.lsf

```
# clear the workspace
clear; closeall;

# get parameters from model variables
groupscope("::model");
lambda = get("lambda");
n = get("n_mo");
theta_f = get("theta_f");
mag_axis = get("mag_axis");
isotropic_material = get("isotropic_material_name");
pos_anisotropic_material = get("pos_anisotropic_material_name");
pos_matrix_rotation = get("pos_matrix_rotation_name");
neg_anisotropic_material = get("neg_anisotropic_material_name");
neg_matrix_rotation = get("neg_matrix_rotation_name");
k0 = 2*pi/lambda;
theta = theta_f * pi/180;# deg/m to rad/m
```

```
# reset the material index values
switchtolayout;
setmaterial(isotropic_material, "Refractive index", n);
setmaterial(pos_anisotropic_material, "Refractive index", [n, n, n]);
setnamed(pos_matrix_rotation, "U", [1,0,0;0,1,0;0,0,1]);
setmaterial(neg_anisotropic_material, "Refractive index", [n, n, n]);
setnamed(neg_matrix_rotation, "U", [1,0,0;0,1,0;0,0,1]);


if (mag_axis=="x")
{
        # Define positive yig material
        # set up materials
        xi = -1*2*n*theta/k0;
        eps2d = [  n^2, -1i*xi;
                  1i*xi,   n^2];

        temp = eig(eps2d,3);
        U2d = ctranspose(pinch(temp,3,2));
        eps2dD = pinch(temp,3,1);

        # re-insert the x row/column
        U = [ 1,       0,       0;
             0, U2d(1,1), U2d(1,2);
             0, U2d(2,1), U2d(2,2)];
        epsD = [n^2,         0,          0;
                  0, eps2dD(1,1),        0;
                  0,          0, eps2dD(2,2)];
        nD = sqrt(epsD);

        # test
        eps = [n^2,        0,         0;
                0, eps2d(1,1), eps2d(1,2);
                0, eps2d(2,1), eps2d(2,2)];

        testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

        if(testval > 1e-10)
        {
                ?"Error, positive rotation unitary transformation is not
                   correct!";
                break;
        }
        else
        {
                ?"Positive rotation unitary transformation is correct.";
```

```
        ?"Setting positive anisotropic material properties and U
            transform.";
        setmaterial(pos_anisotropic_material,"Refractive Index",[nD
            (1,1),nD(2,2),nD(3,3)]);
        setnamed(pos_matrix_rotation,"U",U);
}

# Define negative yig material
# set up materials
theta = theta*-1;
xi = -1*2*n*theta/k0;
eps2d = [  n^2, -1i*xi;
1i*xi,   n^2];

temp = eig(eps2d,3);
U2d = ctranspose(pinch(temp,3,2));
eps2dD = pinch(temp,3,1);

# re-insert the x row/column
U = [ 1,      0,       0;
     0, U2d(1,1), U2d(1,2);
     0, U2d(2,1), U2d(2,2)];
epsD = [n^2,         0,          0;
         0, eps2dD(1,1),         0;
          0,          0, eps2dD(2,2)];
nD = sqrt(epsD);

# test
eps = [n^2,        0,         0;
        0, eps2d(1,1), eps2d(1,2);
        0, eps2d(2,1), eps2d(2,2)];

testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

if(testval > 1e-10)
{
        ?"Error, negative rotation unitary transformation is not
            correct!";
        break;
}
else
{
        ?"Negative rotation unitary transformation is correct.";
        ?"Setting negative anisotropic material properties and U
            transform.";
        setmaterial(neg_anisotropic_material,"Refractive Index",[nD
            (1,1),nD(2,2),nD(3,3)]);
```

```
                setnamed(neg_matrix_rotation,"U",U);
        }
}

if (mag_axis=="y")
{
        # Define positive yig material
        # set up materials
        xi = -1*2*n*theta/k0;
        eps2d = [  n^2, 1i*xi;
                 -1i*xi,   n^2];

        temp = eig(eps2d,3);
        U2d = ctranspose(pinch(temp,3,2));
        eps2dD = pinch(temp,3,1);

        # re-insert the x row/column
        U = [ U2d(1,1), 0, U2d(1,2);
                    0, 1,       0;
              U2d(2,1), 0, U2d(2,2)];
        epsD = [eps2dD(1,1), 0,          0;
                        0, n^2,          0;
                     0,   0, eps2dD(2,2)];
        nD = sqrt(epsD);

        # test
        eps = [eps2d(1,1), 0, eps2d(1,2);
                       0, n^2,        0;
               eps2d(2,1),  0, eps2d(2,2)];

        testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

        if(testval > 1e-10)
        {
                ?"Error, positive rotation unitary transformation is not
                    correct!";
                break;
        }
        else
        {
                ?"Positive rotation unitary transformation is correct.";
                ?"Setting positive anisotropic material properties and U
                    transform.";
                setmaterial(pos_anisotropic_material,"Refractive Index",[nD
                    (1,1),nD(2,2),nD(3,3)]);
                setnamed(pos_matrix_rotation,"U",U);
        }
```

```
        # Define negative yig material
        # set up materials
        theta = theta*-1;
        xi = -1*2*n*theta/k0;
        eps2d = [  n^2, 1i*xi;
                -1i*xi,   n^2];

        temp = eig(eps2d,3);
        U2d = ctranspose(pinch(temp,3,2));
        eps2dD = pinch(temp,3,1);

        # re-insert the x row/column
        U = [ U2d(1,1), 0, U2d(1,2);
                    0, 1,       0;
             U2d(2,1), 0, U2d(2,2)];
        epsD = [eps2dD(1,1), 0,          0;
                        0, n^2,          0;
                        0,   0, eps2dD(2,2)];
        nD = sqrt(epsD);

        # test
        eps = [eps2d(1,1), 0, eps2d(1,2);
                      0, n^2,        0;
               eps2d(2,1),  0, eps2d(2,2)];

        testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

        if(testval > 1e-10)
        {
                ?"Error, negative rotation unitary transformation is not
                    correct!";
                break;
        }
        else
        {
                ?"Negative rotation unitary transformation is correct.";
                ?"Setting negative anisotropic material properties and U
                    transform.";
                setmaterial(neg_anisotropic_material,"Refractive Index",[nD
                    (1,1),nD(2,2),nD(3,3)]);
                setnamed(neg_matrix_rotation,"U",U);
        }
}

if (mag_axis=="z")
{
```

```
# Define positive yig material
# set up materials
xi = -1*2*n*theta/k0;
eps2d = [  n^2, -1i*xi;
          1i*xi,   n^2];

temp = eig(eps2d,3);
U2d = ctranspose(pinch(temp,3,2));
eps2dD = pinch(temp,3,1);

# re-insert the x row/column
U = [ U2d(1,1), U2d(1,2), 0;
      U2d(2,1), U2d(2,2), 0;
             0,       0, 1];
epsD = [eps2dD(1,1),        0, 0;
                 0, eps2dD(2,2), 0;
                 0,          0, n^2];
nD = sqrt(epsD);

# test
eps = [eps2d(1,1), eps2d(1,2), 0;
       eps2d(2,1), eps2d(2,2), 0;
                0,         0, n^2];

testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

if(testval > 1e-10)
{
       ?"Error, positive rotation unitary transformation is not
          correct!";
       break;
}
else
{
       ?"Positive rotation unitary transformation is correct.";
       ?"Setting positive anisotropic material properties and U
          transform.";
       setmaterial(pos_anisotropic_material,"Refractive Index",[nD
          (1,1),nD(2,2),nD(3,3)]);
       setnamed(pos_matrix_rotation,"U",U);
}

# Define negative yig material
# set up materials
theta = theta*-1;
xi = -1*2*n*theta/k0;
eps2d = [  n^2, -1i*xi;
```

```
                1i*xi,   n^2];

        temp = eig(eps2d,3);
        U2d = ctranspose(pinch(temp,3,2));
        eps2dD = pinch(temp,3,1);

        # re-insert the x row/column
        U = [ U2d(1,1), U2d(1,2), 0;
             U2d(2,1), U2d(2,2), 0;
                   0,       0, 1];
        epsD = [eps2dD(1,1),        0, 0;
                        0, eps2dD(2,2), 0;
                        0,          0, n^2];
        nD = sqrt(epsD);

        # test
        eps = [eps2d(1,1), eps2d(1,2), 0;
              eps2d(2,1), eps2d(2,2), 0;
                      0,         0, n^2];

        testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

        if(testval > 1e-10)
        {
               ?"Error, negative rotation unitary transformation is not
                  correct!";
               break;
        }
        else
        {
               ?"Negative rotation unitary transformation is correct.";
               ?"Setting negative anisotropic material properties and U
                  transform.";
               setmaterial(neg_anisotropic_material,"Refractive Index",[nD
                  (1,1),nD(2,2),nD(3,3)]);
               setnamed(neg_matrix_rotation,"U",U);
        }
}
```

Calculating the Nonreciprocal Phase Shift

The script file, *NRPScalc.lsf*, presented below, was employed to calculate the NRPS within *Lumerical FDTD Solutions* for various waveguide geometries. Briefly, its operation is as follows. First, it is assumed that the waveguide cross section is in the

x-z plane, that the y-direction is the propagation direction, and that the parameters defining the MO material are initialized within the model. This method treats the waveguide as a periodic structure in the direction of propagation, with a period equal to one mesh cell. To determine the NRPS, the difference in the propagation constants for the forward and backward propagating modes are required. These wavevectors represent a small perturbation from the isotropic value. Recall that $\beta_{fwd} = \beta_0 + \delta\beta$, and $\beta_{bwd} = \beta_0 - \delta\beta$. This is found by setting the material properties for the MO material as desired, and then calculating the excitation spectrum for two estimated wavevectors in close proximity to the wavelength of interest. Knowledge of the perturbed wavelength that corresponds to these two wavevectors allows for linear interpolation to determine the β_{fwd} at the wavelength of interest. To find the propagation constant of the backward propagating mode, the material properties are reset, assuming magnetization in the opposite direction. This works because switching the magnetization direction produces equivalent results to switching the direction of propagation. In the same manner, the β_{bwd} mode can be found, and correspondingly $\Delta\beta$ can be calculated.

NRPScalc.lsf

```
# clear the workspace
clear; closeall;

groupscope("::model");
# define dimensions
lambda = get("lambda");
n = get("n_mo");
theta_f = get("theta_f");

# set the appropriate mode number for updating the source
mode = 2;

# set the material index to isotropic for the mode calculation and reset
    the U matrix
```

```
switchtolayout;
setmaterial("yig - isotropic", "Refractive index", n);
setmaterial("yig - anisotropic", "Refractive index", [n, n, n]);
setnamed("yig rotation", "U", [1,0,0;0,1,0;0,0,1]);

# get source effective index and calculate the ky values
select("source");
updatesourcemode(mode);
source_data = getresult("source", "neff");
neff_mode = source_data.neff;
ky = real(neff_mode)*2*pi /[ 1.525e-6, 1.575e-6 ];

# run calculation for dbeta
f0 = matrix(2,2);

# set direction matrix. 1=forward, -1=backward, 0=isotropic
dirsign = [1,-1];

for(d = 1:2)
{
       for(i = 1:length(ky) )
       {
              switchtolayout;
              setnamed("FDTD","ky",ky(i));

              # Define positive yig material
              # set up materials
              k0 = 2*pi/lambda;
              theta = dirsign(d)*theta_f * pi/180;# deg/m to rad/m
              xi = -1*2*n*theta/k0;
              eps2d = [ n^2, -1i*xi;
                     1i*xi, n^2 ];

              temp = eig(eps2d,3);
              U2d = ctranspose(pinch(temp,3,2));
              eps2dD = pinch(temp,3,1);

              # re-insert the x row/column
              U = [ 1, 0, 0;
                   0, U2d(1,1),U2d(1,2);
                   0, U2d(2,1),U2d(2,2) ];
              epsD = [n^2,0,0;
                       0,eps2dD(1,1),0;
                       0,0,eps2dD(2,2)];
              nD = sqrt(epsD);

              # test
```

```
eps = [n^2,0,0;
        0,eps2d(1,1),eps2d(1,2);
        0, eps2d(2,1),eps2d(2,2)];

?testval = max( abs(epsD - mult(U,eps,ctranspose(U))) );

if(testval > 1e-10)
{
       ?"Error, unitary transformation is not correct!";
       break;
} else
{
       ?"Unitary transformation is correct.";
       ?"Setting material properties for yig - anisotropic,
          and U transform for 'yig rotation'";
       setmaterial("yig - anisotropic","Refractive Index",[
          nD(1,1),nD(2,2),nD(3,3)]);
       setnamed("yig rotation","U",U);
}

run;

runanalysis;
spectrum = getresult("bandstructure","spectrum");
if(d==1)
{
       if(i==1)
       {
              spec_fwd1 = spectrum;
       }
       if(i==2)
       {
              spec_fwd2 = spectrum;
       }
}
if(d==2)
{
       if(i==1)
       {
              spec_bwd1 = spectrum;
       }
       if(i==2)
       {
              spec_bwd2 = spectrum;
       }
}
f = spectrum.f;
```

```
            pos = findpeaks(abs(spectrum.fs)^2);
            f0(i,d) = f(pos);
        }
}

f0_forward = f0(1:2,1);
f0_backward = f0(1:2,2);

plotxy(c/f0_forward,ky,c/f0_backward,ky);
delta_beta = interp(ky,f0_forward,c/lambda)-interp(ky,f0_backward,c/lambda)
    ;
?delta_beta/1000;# for rad/mm

NRPS = delta_beta;
```

Appendix B

Magnetization Dynamics MATLAB Code

The MATLAB code presented in this appendix was employed to calculate the magnetization dynamics for the various modulator designs presented within Chapter 3 and Chapter 4. Correspondingly, it takes input from the user, and solves the LLG equation for magnetization dynamics given appropriate initial conditions and data on the form of the applied magnetic field. Currently, the code is set up to accept input applied fields of the following forms: a single Gaussian pulse, a single super-Gaussian pulse, a single square pulse, a train of super-Gaussian pulses, a sinusoid, or the superposition of two independent sinusoids. The code presents the flexibility to implement driving functions of different forms with minor additions. Furthermore, it can sweep over one or two different variables as desired. The main program file, containing all of the user inputs, is *MagnetizationDynamicsProgram.m*, while the other auxiliary functions are called by the main program and used to calculate the resultant temporal dynamics. Plotting was typically performed manually, but the *plotMagnetization.m* function quickly plots the temporal variations in **M** for a single iteration.

MagnetizationDynamicsProgram.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MAGNETIZATION DYNAMICS SOLVER
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% clear workspace
clear; clc; close all;

%-------------------------------------------------------------------------%
% USER INPUT
%-------------------------------------------------------------------------%
tic;
% set time vector
tstart = 0e-9;
tend =10e-9;
dt = 0.001e-9;
t = tstart:dt:tend;

% set initial magnetization condition [x; y; z], normalized to Ms
% the magnitude of this vector must be 1
M0 = [1;0;0];

%########################
% set static magnetic fields
%########################
staticCompX = 'x';
BstaticX = [0];

staticCompY = 'y';
BstaticY = [0];

staticCompZ = 'z';
BstaticZ = [0];

%########################
% Input the transient field properties
%########################
transientCompX = 'x';
btransientX = [0];
transientTypeX = ''; % or 'SuperGauss', 'Gauss', 'Sinusoid', 'Rectangle', '
    SuperGaussTrain', 'TwoSineSum'
fwhmX = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
t0X = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
NX = []; % required for the 'SuperGauss' and 'SuperGaussTrain'
numPulsesX = []; % required for 'SuperGaussTrain'
pulseSpacingX = []; % required for 'SuperGaussTrain
```

```
freqX = []; % required for 'Sinusoid' and 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
phaseX = []; % required for 'Sinusoid', 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
amp2X = []; % required for 'TwoSineSum'

transientCompY = 'y';
btransientY = [0];
transientTypeY = ''; % or 'SuperGauss', 'Gauss', 'Sinusoid', 'Rectangle', '
    SuperGaussTrain', 'TwoSineSum'
fwhmY = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
t0Y = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
NY = []; % required for the 'SuperGauss' and 'SuperGaussTrain'
numPulsesY = []; % required for 'SuperGaussTrain'
pulseSpacingY = []; % required for 'SuperGaussTrain
freqY = []; % required for 'Sinusoid' and 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
phaseY = []; % required for 'Sinusoid', 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
amp2Y = []; % required for 'TwoSineSum'

transientCompZ = 'z';
btransientZ = [0];
transientTypeZ = ''; % or 'SuperGauss', 'Gauss', 'Sinusoid', 'Rectangle', '
    SuperGaussTrain', 'TwoSineSum'
fwhmZ = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
t0Z = []; % required for 'SuperGauss' and 'Gauss' and 'SuperGaussTrain'
NZ = []; % required for the 'SuperGauss' and 'SuperGaussTrain'
numPulsesZ = []; % required for 'SuperGaussTrain'
pulseSpacingZ = []; % required for 'SuperGaussTrain
freqZ = []; % required for 'Sinusoid' and 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
phaseZ = []; % required for 'Sinusoid', 'TwoSineSum'; must have two
    elements for 'TwoSineSum'
amp2Z = []; % required for 'TwoSineSum'

%######################

% specify components for the data to store
primaryPkComp = 'My'; % specify 'Mx', 'My', or 'Mz'
pkType = 'Min'; % specify 'Max' or 'Min'
amplitudeComp = 'Mx'; % specify 'Mx', 'My', or 'Mz'
endMagComp = 'My'; % specify 'Mx', 'My', or 'Mz'

% store M components for looped calculations
% use only if looping over one variable, not if looping over two
% if only a single calculation is done, leave this as false
```

```
storeMagComp = false;

% specify which variables to loop over
% options are 'Bstatic', 'btransient'
outerLoopVar = 'BstaticY';
innerLoopVar = 'btransientX';

% filename for saving the results.
% NOTE: include '.mat' at the end
filename = 'filename.mat';

%-------------------------------------------------------------------------%
% END USER INPUT
%-------------------------------------------------------------------------%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-------------------------------------------------------------------------%
% FIELD CALCULATIONS
%-------------------------------------------------------------------------%

% combine the transient types into a single vector
transientType = [{transientTypeX} {transientTypeY} {transientTypeZ}];

% predefine matrices for various values
primaryPk = zeros(length(eval(outerLoopVar)), length(eval(innerLoopVar)));
primaryPkTime = zeros(length(eval(outerLoopVar)), length(eval(innerLoopVar)
    ));
amplitude = zeros(length(eval(outerLoopVar)), length(eval(innerLoopVar)));
endMag = zeros(length(eval(outerLoopVar)), length(eval(innerLoopVar)));

% if the option to store the magnetization components is set, predefine the
% matricies for them
if storeMagComp == true
    if( ((length(eval(outerLoopVar))==1) && (length(eval(innerLoopVar))~=1)
        ) || ((length(eval(outerLoopVar))~=1) && (length(eval(innerLoopVar))
        ==1)) )
        MxComp = zeros(length(eval(innerLoopVar)), length(t));
        MyComp = zeros(length(eval(innerLoopVar)), length(t));
        MzComp = zeros(length(eval(innerLoopVar)), length(t));
    else
        error('Cannot store the magnetization components when looping over
            two variables.');
    end
end

% select the appropriate parameters for the transient fields from the
```

```
% complete list of parameters

[paramX1, paramX2, paramX3, paramX4, paramX5] = selectParameters(
   transientTypeX, t0X, fwhmX, NX, numPulsesX, pulseSpacingX, freqX,
   phaseX, amp2X);
[paramY1, paramY2, paramY3, paramY4, paramY5] = selectParameters(
   transientTypeY, t0Y, fwhmY, NY, numPulsesY, pulseSpacingY, freqY,
   phaseY, amp2Y);
[paramZ1, paramZ2, paramZ3, paramZ4, paramZ5] = selectParameters(
   transientTypeZ, t0Z, fwhmZ, NZ, numPulsesZ, pulseSpacingZ, freqZ,
   phaseZ, amp2Z);


% loop through transient fields
for i=1:length(eval(outerLoopVar))

   % define the field matrix for the outer loop variable
   switch outerLoopVar
      case 'btransientX'
         b = fieldMatrix('Transient', transientCompX, [btransientX(i);
            paramX1; paramX2; paramX3; paramX4; paramX5], transientCompY,
             [btransientY; paramY1; paramY2; paramY3; paramY4; paramY5],
            transientCompZ, [btransientZ; paramZ1; paramZ2; paramZ3;
            paramZ4; paramZ5]);
      case 'btransientY'
         b = fieldMatrix('Transient', transientCompX, [btransientX;
            paramX1; paramX2; paramX3; paramX4; paramX5], transientCompY,
             [btransientY(i); paramY1; paramY2; paramY3; paramY4; paramY5
            ], transientCompZ, [btransientZ; paramZ1; paramZ2; paramZ3;
            paramZ4; paramZ5]);
      case 'btransientZ'
         b = fieldMatrix('Transient', transientCompX, [btransientX;
            paramX1; paramX2; paramX3; paramX4; paramX5], transientCompY,
             [btransientY; paramY1; paramY2; paramY3; paramY4; paramY5],
            transientCompZ, [btransientZ(i); paramZ1; paramZ2; paramZ3;
            paramZ4; paramZ5]);
      case 'BstaticX'
         B = fieldMatrix('Static', staticCompX, BstaticX(i), staticCompY,
              BstaticY, staticCompZ, BstaticZ);
      case 'BstaticY'
         B = fieldMatrix('Static', staticCompX, BstaticX, staticCompY,
            BstaticY(i), staticCompZ, BstaticZ);
      case 'BstaticZ'
         B = fieldMatrix('Static', staticCompX, BstaticX, staticCompY,
            BstaticY, staticCompZ, BstaticZ(i));
      otherwise
         error('Invalid outer loop variable type.');
```

```
end

% loop through static fields
for j=1:length(eval(innerLoopVar))
    % print iteration message to the screen
    fprintf('Outer loop interation %g of %g. Inner loop iteration %g of
        %g. Iterations remaining: %g\n', i, length(eval(outerLoopVar)),
        j, length(eval(innerLoopVar)), length(eval(outerLoopVar))*length
        (eval(innerLoopVar))-(i-1)*length(eval(innerLoopVar))-j);

    % define the field matrix for the inner loop variable
    switch innerLoopVar
        case 'btransientX'
            b = fieldMatrix('Transient', transientCompX, [btransientX(j);
                 paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ; paramZ1;
                 paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'btransientY'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY(j); paramY1; paramY2;
                paramY3; paramY4; paramY5], transientCompZ, [btransientZ;
                 paramZ1; paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'btransientZ'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ(j);
                paramZ1; paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'BstaticX'
            B = fieldMatrix('Static', staticCompX, BstaticX(j),
                staticCompY, BstaticY, staticCompZ, BstaticZ);
        case 'BstaticY'
            B = fieldMatrix('Static', staticCompX, BstaticX, staticCompY,
                 BstaticY(j), staticCompZ, BstaticZ);
        case 'BstaticZ'
            B = fieldMatrix('Static', staticCompX, BstaticX, staticCompY,
                 BstaticY, staticCompZ, BstaticZ(j));
        case 'pulseSpacingX'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5(j)],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ; paramZ1;
                 paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'pulseSpacingY'
```

```
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5(j)], transientCompZ, [btransientZ;
                paramZ1; paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'pulseSpacingZ'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ; paramZ1;
                 paramZ2; paramZ3; paramZ4; paramZ5(j)]);
        case 'frequencyX'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1(j); paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ; paramZ1;
                 paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'frequencyY'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1(j); paramY2;
                paramY3; paramY4; paramY5], transientCompZ, [btransientZ;
                 paramZ1; paramZ2; paramZ3; paramZ4; paramZ5]);
        case 'frequencyZ'
            b = fieldMatrix('Transient', transientCompX, [btransientX;
                paramX1; paramX2; paramX3; paramX4; paramX5],
                transientCompY, [btransientY; paramY1; paramY2; paramY3;
                paramY4; paramY5], transientCompZ, [btransientZ; paramZ1(
                j); paramZ2; paramZ3; paramZ4; paramZ5]);
        otherwise
            error('Invalid inner loop variable type.');
    end

    % calcuate the magnetization response
    [Mx, My, Mz] = magnetizationDynamics(t, M0, B, b, transientType, '
        NoPlots');
    % calculate the primary peak and primary peak time
    [primaryPk(i,j), primaryPkTime(i,j)] = primaryPeak(t, eval(
        primaryPkComp), pkType);
    % calculate the amplitude of the precessional oscillations
    amplitude(i,j) = precessionAmplitude(eval(amplitudeComp));
    % calculate the end magnetization state
    endMag(i,j) = endMagnetization(eval(endMagComp));

    % if the option to store the magnetization components is set, store
    % the components in the appropriate matrices
    if( storeMagComp == true )
```

```
            MxComp(j,:) = Mx;
            MyComp(j,:) = My;
            MzComp(j,:) = Mz;
        end

    end
end
toc;
%-------------------------------------------------------------------------%
% END FIELD CALCULATIONS
%-------------------------------------------------------------------------%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%-------------------------------------------------------------------------%
% SAVE DATA AND SETTINGS
%-------------------------------------------------------------------------%

% save the data with the predefined filename
save(filename);

%-------------------------------------------------------------------------%
% END SAVE DATA AND SETTINGS
%-------------------------------------------------------------------------%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% END MAGNETIZATION DYNAMICS SOLVER
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

magnetizationDynamics.m

```
function [ Mx, My, Mz ] = magnetizationDynamics( t, M0, Bstatic, btransient
    , transient_type, gen_plots )
% Function to calculate the temporal evolution of the magnetization vector
%   Inputs:
%   t = time vector (s)
%   M0 = normalized initial magnetization state, ie |M|=1, in a column vec.
%   Bstatic = 3 element vector of the external static fields Bx, By, Bz
%   btransient = transient fields. Row 1 is the peak, row 2 is the center
%   time of the pulse, and row 3 is the FWHM. Columns correspond to x, y, z
%   transient_type = string containing the type of transient pulse.
%   Supported types include 'SuperGauss' and 'Gauss'.
%   gen_plots = optional input. 'NoPlots' will disable the output plots. If
%   not specified, plots will be generated.
%   Outputs:
%   Mx, My, Mz = the normalized components of the magnetization vector.
```

```
% Define constants
mu0 = 4*pi*10^(-7);         % permeability of free space (H/m)
q = 1.602e-19;              % electronic charge (C)
me = 9.109e-31;             % mass of electron (kg)
g = 2;                      % Lande factor
gamma0 = (q*g)/(2*me);      % gyromagnetic ratio (Hz/T)
alpha = 1e-4;               % Gilbert damping parameter
Ms = 9e-3;                  % saturation magnetization (T)

% set options for the ODE solver
options = odeset('RelTol', 1e-12, 'AbsTol', 1e-12);

% check to ensure M0 is a unit vector
if (abs(norm(M0)-1)>1e-6)
    error('Initial magnetization condition is not a unit vector. |M0| = %g
        ', norm(M0));
end

% set initial conditions
M0 = Ms*M0;

% split up the time vector into pieces
timeBlock = 5e-9;

% predefine the M matrix
M = zeros(length(t), 3);

% if the total time is greater than the predefined time block, break it up
if (abs(t(end)-t(1))>=timeBlock)
    % output an error if the total time is not divisible by the time block
    if ( mod(abs(t(end)-t(1)), timeBlock) ~= 0 )

        error('Time vector must be a multiple of 5ns.');
    end

    % set an index
    index = 1;

    dt = t(2)-t(1);
    numSteps = round(timeBlock/dt);
    numBlocks = round(abs(t(end)-t(1))/timeBlock);

    for i=1:numBlocks
        tSplit = t(index:index+numSteps);

        % redefine the initial conditions for each time block
```

```
        if i ~= 1
            M0 = [Mi(end,1); Mi(end,2); Mi(end,3)];
        end

        % calculate the magnetization dynamics with the LLG model
        [tSplit,Mi] = ode45(@(tSplit,M)LLGModel(tSplit, M, gamma0, alpha, Ms
            , Bstatic, btransient, transient_type), tSplit, M0, options);

        % put the calculated values into one M matrix, noting that the last
        % point of one block will also be the first point of the next
        if i == 1
            M(index:index+numSteps,1) = Mi(:,1);
            M(index:index+numSteps,2) = Mi(:,2);
            M(index:index+numSteps,3) = Mi(:,3);
        else
            M(index+1:index+numSteps,1) = Mi(2:end,1);
            M(index+1:index+numSteps,2) = Mi(2:end,2);
            M(index+1:index+numSteps,3) = Mi(2:end,3);
        end

        index = index + numSteps;
    end
end



% split the M matrix into component vectors and normalize to Ms
Mx = M(:,1).'/Ms;
My = M(:,2).'/Ms;
Mz = M(:,3).'/Ms;

% generate plots of the flag is not set, or set incorrectly
if ( (nargin == 6) && (~strcmp(gen_plots, 'NoPlots')) )
    error('Invalid plot toggle command.');
end
if (nargin < 6) || ((nargin == 6) && (~strcmp(gen_plots, 'NoPlots')))
    % plot the three components as a function of time
    figure;
    subplot(1,3,1); plot(t, Mx); title('M_x'); xlabel('Time (s)'); ylabel('
        M_x/M_s');
    subplot(1,3,2); plot(t, My); title('M_y'); xlabel('Time (s)'); ylabel('
        M_y/M_s');
    subplot(1,3,3); plot(t, Mz); title('M_z'); xlabel('Time (s)'); ylabel('
        M_z/M_s');

    % plot the trajectory of the magnetization vector
    figure;
```

```
    subplot(1,2,1);
    plot3(Mx, My, Mz); title('M Trajectory Zoomed'); xlabel('M_x/M_s');
       ylabel('M_y/M_s'); zlabel('M_z/M_s');
    grid on;

    subplot(1,2,2);
    plot3(Mx, My, Mz); title('M Trajectory'); xlabel('M_x/M_s'); ylabel('
       M_y/M_s'); zlabel('M_z/M_s');
    xlim([-1 1]); ylim([-1 1]); zlim([-1 1]); grid on;
    hold on;
    plot3(xlim, [0 0], [0 0], 'k', 'LineWidth', 1.5); % plot the horizontal
        line for x axis
    plot3([0 0], ylim, [0 0], 'k', 'LineWidth', 1.5); % plot the horizontal
        line for y axis
    plot3([0 0], [0 0], zlim, 'k', 'LineWidth', 1.5); % plot the horizontal
        line for z axis
    hold off;
end

end

function [ dM_dt ] = LLGModel( t, M, gamma0, alpha, Ms, Bstatic, btransient
    , transient_type )
% Function defining the LLG equation for calculating magnetization dynamics
%   Inputs:
%   t = time vector (s)
%   M = magnetization vector (quoted in T here)
%   gamma0 = gyromagnetic ratio
%   alpha = Gilbert damping parameter
%   Ms = saturation magnetization (quoted in T)
%   Bstatic = vector of the Bx, By, and Bz static applied fields
%   btransient = matrix with properties of the transient, defined above
%   transient_type = string containing the form of the transient. Supported
%   options include 'SuperGauss' and 'Gauss'.
%   Output:
%   dM_dt = time derivative of the magnetization components

% calculate transient components
if (btransient(1,1) == 0)
    bx = 0;
else
    bx = transient( t, btransient(:,1), transient_type{1} );
end
if (btransient(1,2) == 0)
    by = 0;
else
    by = transient( t, btransient(:,2), transient_type{2} );
```

```
end
if (btransient(1,3) == 0)
    bz = 0;
else
    bz = transient( t, btransient(:,3), transient_type{3} );
end

% split up the Bstatic vector into components
Bx = Bstatic(1);
By = Bstatic(2);
Bz = Bstatic(3);

% build the LLG equation
prefactor1 = -gamma0 / (1+alpha^2);
prefactor2 = -gamma0*alpha / (Ms*(1+alpha^2));
precession_term = [M(2)*(Bz+bz)-M(3)*(By+by)
                   M(3)*(Bx+bx)-M(1)*(Bz+bz)
                   M(1)*(By+by)-M(2)*(Bx+bx)];
damping_term = [M(2)*(M(1)*(By+by)-M(2)*(Bx+bx))-M(3)*(M(3)*(Bx+bx)-M(1)*(
    Bz+bz))
                M(3)*(M(2)*(Bz+bz)-M(3)*(By+by))-M(1)*(M(1)*(By+by)-M(2)*(Bx+
                    bx))
                M(1)*(M(3)*(Bx+bx)-M(1)*(Bz+bz))-M(2)*(M(2)*(Bz+bz)-M(3)*(By+
                    by))];
dM_dt = prefactor1*precession_term + prefactor2*damping_term;

end

function [ b ] = transient( t, btransient, transient_type )
% Function to select the appropriate transient form
%   Inputs:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   transient_type = string with the form of the transient field. Options
%   include 'SuperGauss', and 'Gauss'
%   Output:
%   b = magnetic field (T)

% select appropriate field form
switch transient_type
    case 'SuperGauss'
        b = superGauss( t, btransient );
    case 'Gauss'
        b = gauss( t, btransient );
    case 'Sinusoid'
        b = sineWave( t, btransient );
    case 'TwoSineSum'
```

```
        b = twoSineSum( t, btransient );
    case 'SuperGaussTrain'
        b = superGaussTrain( t, btransient );
    case 'Rectangle'
        b = rectPulse( t, btransient );
    otherwise
        error('Unrecognized transient field type.');
end

end

function [ b ] = superGauss( t, btransient )
% Function to calculate a super Gaussian transient field pulse
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with super Gaussian form (T)

% split up the btransient matrix into peak, center time, FWHM, and order
bpk = btransient(1);
t0 = btransient(2);
fwhm = btransient(3);
N = btransient(4);

% calculate sigma from the FWHM
sigma = nthroot(fwhm^N/(2^(N+1)*log(2)), N);

% calculate the magnetic field
b = bpk.*( exp(-1.*abs(t-t0).^N./(2*sigma^N)) );

end

function [ b ] = gauss( t, btransient )
% Function to calculate a Gaussian transient field pulse
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with Gaussian form (T)

% split up the btransient matrix into peak, center time, and FWHM
bpk = btransient(1);
t0 = btransient(2);
fwhm = btransient(3);

% calculate the standard deviation from the FWHM
```

```
sigma = fwhm/(2*sqrt(2*log(2)));

% calculate the magnetic field
b = bpk.*( exp(-1.*abs(t-t0).^2./(2*sigma^2)) );

end

function [ b ] = sineWave( t, btransient )
% Function to calculate a sine wave transient field
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with a sine wave form (T)

% split up the btransient matrix into amplitude, frequency, and phase
amp = btransient(1);
freq = btransient(2);
phase = btransient(3);

% calculate angular frequency
w = 2*pi*freq;

% calculate the magnetic field
b = amp.*sin(w.*t + phase);

% truncate the field so that the sine wave starts at t=0
b(t<0) = 0;

end

function [ b ] = twoSineSum( t, btransient )
% Function to calculate a transient field that is the sum of 2 sines
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with a sum of 2 sine waves form (T)

% split up the btransient matrix into amplitude, frequency, and phase
amp1 = btransient(1);
freq1 = btransient(2);
phase1 = btransient(3);

amp2 = btransient(4);
freq2 = btransient(5);
phase2 = btransient(6);
```

```
% calculate angular frequencies
w1 = 2*pi*freq1;
w2 = 2*pi*freq2;

% calculate the magnetic field
b = amp1.*sin(w1.*t + phase1) + amp2.*sin(w2.*t + phase2);

% truncate the field so that the sine wave starts at t=0
b(t<0) = 0;

%plot(t, b, 'ro'); hold on;

end

function [ b ] = superGaussTrain( t, btransient )
% Function to calculate a super Gaussian transient field pulse
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with super Gaussian form (T)

% split up the btransient matrix into peak, center time, and FWHM
bpk = btransient(1);
t0 = btransient(2);
fwhm = btransient(3);
N = btransient(4);
numPulses = btransient(5);
pulseSpacing = btransient(6);

% calculate sigma from the FWHM
sigma = nthroot(fwhm^N/(2^(N+1)*log(2)), N);

% calculate the magnetic field
%b = bpk.*( exp(-1.*abs(t-t0).^N./(2*sigma^N)) );

b = 0;
for k = 1:numPulses
    b = b + bpk.*( exp(-1.*abs(t-t0 - (k-1)*pulseSpacing).^N./(2*sigma^N))
        );
end

end

function [ b ] = rectPulse( t, btransient )
% Function to calculate a Gaussian transient field pulse
```

```
%   Input:
%   t = time vector (s)
%   btransient = matrix with the transient field properties
%   Output:
%   b = magnetic field with rectangular pulse form (T)

% split up the btransient matrix into peak, center time, and FWHM
bpk = btransient(1);
t0 = btransient(2);
fwhm = btransient(3);

% calculate the magnetic field
b = bpk.*ones(1, length(t));

% set the field to 0 outside of the pulse
b(t<(t0-fwhm/2))=0;
b(t>(t0+fwhm/2))=0;

%plot(t, b, 'ro'); hold on;

end
```

selectParameters.m

```
function [ param1, param2, param3, param4, param5 ] = selectParameters(
    transientType, t0, fwhm, N, numPulses, pulseSpacing, freq, phase, amp2
    )
% Function to select the required parameters for a given transient input
% field type.
%   Inputs:
%   transientType = a string defining the transient type. This can be
%   either 'SuperGauss', 'Gauss', or 'Sinusoid'
%   t0 = time of the pulse center. Used for 'SuperGauss' and 'Gauss' types.
%   fwhm = pulse FWHM. Used for 'SuperGauss' and 'Gauss' types.
%   freq = frequency in Hz. Used for 'Sinusoid' type.
%   phase = phase of the sinusoid in radians at t=0; Used for the
%   'Sinusoid' type.


% assign the variables param1 and param2 values required by the input field
% type. Inputs that do not correspond to a particular field type are
% ignored.
switch transientType
    case 'SuperGauss'
        param1 = t0;
        param2 = fwhm;
```

```
        param3 = N;
        param4 = 0;
        param5 = 0;
    case 'Gauss'
        param1 = t0;
        param2 = fwhm;
        param3 = 0;
        param4 = 0;
        param5 = 0;
    case 'Sinusoid'
        if ((length(freq)~=1) || (length(phase)~=1))
            error('Error. Sinusoid option must have only one frequency and
                phase input.');
        end
        param1 = freq;
        param2 = phase;
        param3 = 0;
        param4 = 0;
        param5 = 0;
    case 'TwoSineSum'
        if ((length(freq)~=2) || (length(phase)~=2))
            error('Error. TwoSineSum option must have two frequency and
                phase inputs.');
        end
        param1 = freq(1);
        param2 = phase(1);
        param3 = amp2;
        param4 = freq(2);
        param5 = phase(2);
    case 'SuperGaussTrain'
        param1 = t0;
        param2 = fwhm;
        param3 = N;
        param4 = numPulses;
        param5 = pulseSpacing;
    case 'Rectangle'
        param1 = t0;
        param2 = fwhm;
        param3 = 0;
        param4 = 0;
        param5 = 0;
    case ''
        param1 = 0;
        param2 = 0;
        param3 = 0;
        param4 = 0;
        param5 = 0;
```

```
    otherwise
        error('Invalid transient field type.')
end

end
```

fieldMatrix.m

```
function [ b ] = fieldMatrix( type, comp1, fields1, comp2, fields2, comp3,
    fields3 )
% Function to define the field matrix for a desired input. The field matrix
% defines either the transient or the static input.
%   Inputs:
%   type = a string defining either a 'Transient' or 'Static' field
%   comp1 = a string defining the cartesian direction of the field. Options
%   are 'x', 'y', or 'z'
%   fields1 = column vector defining the values to be input to the total
%   field matrix. For a static field, this is one number, for a transient,
%   it is a column vector defining the properties of the transient pulse as
%   outlined in magnetizationDynamics().
%   ** Note that comp1 and fields1 are required. comp2, fields2, comp3, and
%   fields3 are optional inputs allowing the user to define fields with
%   components in all three directions.

% check the number of components, and add the inputs to a vector that can
% easily be indexed
if nargin == 3
    component = comp1;
    fields = fields1;
elseif nargin == 5
    component = [comp1, comp2];
    fields = [fields1, fields2];
    if strcmp(comp1, comp2)
        error('Cannot specify values for a component more than once in
            fieldMatrix()');
    end
elseif nargin == 7
    component = [comp1, comp2, comp3];
    fields = [fields1, fields2, fields3];
    if (strcmp(comp1, comp2) || (strcmp(comp1, comp3) || strcmp(comp2,
        comp3)))
        error('Cannot specify values for a component more than once in
            fieldMatrix()');
    end
else
    error('Invalid number of arguments in fieldMatrix()');
```

```
end

% Define the size of the field matrix. If the input is transient, the
% number of rows must match that specified in the fields input. If the
% input is static, there is only one row. All cases have three columns,
% corresponding to 'x', 'y', and 'z'.
if strcmp(type, 'Transient')
    b = zeros(length(fields1), 3);
elseif strcmp(type, 'Static')
    b = zeros(1,3);
else
    error('Invalid field type for fieldMatrix(). Enter ''Transient'' or ''
       Static''');
end

% set the appropriate column of the field matrix
for i=1:((nargin-1)/2)
    switch component(i)
        case 'x'
            b(:,1) = fields(:,i);
        case 'y'
            b(:,2) = fields(:,i);
        case 'z'
            b(:,3) = fields(:,i);
        otherwise
            error('Invalid component selection for fieldMatrix(). Enter ''x
               '', ''y'', or ''z''');
    end
end

end
```

primaryPeak.m

```
function [ primPk, primPkTime ] = primaryPeak( t, Mi, type, tol )
% Function to calculate the magnitude of the primary peak of a
% magnetization profile, as well as the time at which this peak occurs
%   Inputs:
%   t = time vector (s)
%   Mi = one of the vector components of the magnetization response
%   type = string specifying if the primary peak is a maxima or a minima
%   tol = optional input to specify the tolerance used to find the primary
%   peak. Default value is 1e-10
%   Outputs:
%   primPk = magnitude of the primary peak of the response
%   primPkTime = time that the primary peak occurs
```

```
% check how many arguments were input
if (nargin < 3) || (nargin > 4)
    % output error if less than 3 arguments or more than 4 arguments
    error('Improper number of input arguments to primaryPeak().');
elseif nargin == 3
    % if 3 arguments are provided, set tol to a default value
    tol = 1e-10;
end

% calculate the primary peak when it is specified as a maximum
if (strcmp(type, 'Max'))
    % find all peaks
    [pk, loc] = findpeaks(Mi);
    % search through the peaks and take the first one greater than tol as
    % the primary peak. Determine the corresponding peak time.
    for(i=1:length(pk))
        if pk(i)>tol
            primPk = pk(i);
            primPkTime = t(loc(i));
            break;
        end
    end
% calculate the primary peak when it is specified as a minimum
elseif (strcmp(type, 'Min'))
    % find all peaks. Note, Mi is inverted to find maxima of -Mi
    [pk, loc] = findpeaks(-Mi);
    % convert peak values back to negative
    pk = -pk;
    % search through the peaks and take the first one with magnitude above
    % tol as the primary peak. Determine the corresponding peak time.
    for(i=1:length(pk))
        if abs(pk(i))>tol
            primPk = pk(i);
            primPkTime = t(loc(i));
            break;
        end
    end
else
    % output error if peak type is not properly set.
    error('Invalid specification of peak type. Specify ''Max'' or ''Min''')
        ;
end

end
```

endMagnetization.m

```
function [ endMag ] = endMagnetization( Mi )
% Function to calculate the magnetization state at the end of a given time
% period.
%   Inputs:
%   Mi = one of the vector components of the magnetization response
%   Outputs:
%   endMag = end state of the Mi magnetization profile

% find the end state of the magnetization profile
endMag = Mi(end);

end
```

precessionAmplitude.m

```
function [ amplitude ] = precessionAmplitude( Mi )
% Function to calculate the magnitude of the swing in the precessional
% oscillations of one component of the magnetization
%   Inputs:
%   Mi = one of the vector components of the magnetization response
%   Outputs:
%   amplitude = amplitude of the precessional oscillations in the Mi
%   component calculated at the end of the examined time span.


% find the maxima of the magnetization response
[pkMax, locMax] = findpeaks(Mi);
% find the minima of the magnetization response (maxima of -Mi)
[pkMin, locMin] = findpeaks(-Mi);

% calculate the amplitude swing at the end of the magnetization profile
swing = pkMax(end) - (-1*pkMin(end));
amplitude = swing/2;

end
```

larmorFrequency.m

```
function [ nu ] = larmorFrequency( Bstatic )
%Function to calculate the Larmor frequency given a static magnetic field
```

```
q = 1.602e-19;
me = 9.109e-31;
g = 2;
gamma0 = (q*g)/(2*me);

nu = gamma0*Bstatic/(2*pi);

end
```

plotMagnetization.m

```
function [] = plotMagnetization( t, Mx, My, Mz )
% Function to plot the components of the magnetization vector vs time, and
% the total trajectory of M
%   Inputs:
%   t = vector of time values (s)
%   Mx, My, Mz = vectors containing the magnetization components normalized
%   to the saturation magnetization, Ms
%   Outputs:
%   Plots the components vs time and the magnetization trajectory

% output error if the number of inputs is incorrect
if (nargin ~= 4)
    error('Invalid number of arguments in plotMagnetization().');
end

% plot the individual components of the magnetization vector
figure;
subplot(1,3,1); plot(t, Mx);
title('M_x vs Time'); xlabel('Time (s)'); ylabel('M_x/M_S');
subplot(1,3,2); plot(t, My);
title('M_y vs Time'); xlabel('Time (s)'); ylabel('M_y/M_S');
subplot(1,3,3); plot(t, Mz);
title('M_z vs Time'); xlabel('Time (s)'); ylabel('M_z/M_S');

% plot the total trajectory of the magnetization vector
figure;
subplot(1,2,1);
plot3(Mx, My, Mz); title('M Trajectory Zoomed'); xlabel('M_x/M_s'); ylabel
    ('M_y/M_s'); zlabel('M_z/M_s');
grid on;

subplot(1,2,2);
plot3(Mx, My, Mz); title('M Trajectory'); xlabel('M_x/M_s'); ylabel('M_y/
    M_s'); zlabel('M_z/M_s');
xlim([-1 1]); ylim([-1 1]); zlim([-1 1]); grid on;
```

```
hold on;
plot3(xlim, [0 0], [0 0], 'k', 'LineWidth', 1.5); % plot the horizontal
    line for x axis
plot3([0 0], ylim, [0 0], 'k', 'LineWidth', 1.5); % plot the horizontal
    line for y axis
plot3([0 0], [0 0], zlim, 'k', 'LineWidth', 1.5); % plot the horizontal
    line for z axis
hold off;

end
```

Appendix C

Supplementary Phase Shifter Design Calculations

This appendix presents supplementary sample calculations from the design of the high-speed magnetoplasmonic waveguide phase shifter presented in Section 3.2. This section includes sample calculations to determine the optimum magnetic field parameters for the critically-damped response, calculation of the magnetic field distribution and Ag transmission line design, as well as thermal analysis of the structure.

Calculation of the Magnetic Fields for a Critically-Damped Magnetization Response

As a sample calculation, we consider the L-configuration as shown in Fig. 3.4(a) of the thesis. To determine the magnetic field values required to critically-damp the magnetization response, the LLG model is numerically solved to track the temporal evolution of the magnetization vector in response to a transient pulse, $\mathbf{h}(t) = \langle h_x(t), 0, 0 \rangle$, and a static biasing field $\mathbf{H}_{static} = \langle 0, 0, H_z \rangle$. The pulse is characterized by a FWHM, τ_p, and a peak value, $h_{x,pk}$, such that the transient magnetic field pulse has the following

super-Gaussian form, resembling a practical logic signal:

$$h_x(t) = h_{x,pk} \exp\left[-\frac{|t-t_0|^5}{2\sigma^5}\right] \tag{C.1}$$

where the parameter σ is related to τ_p via:

$$\sigma = \sqrt[5]{\frac{\tau_p^5}{2^6 \ln 2}} \tag{C.2}$$

While all three magnetization components, M_x, M_y, and M_z, are determined simultaneously, only the M_x component contributes to the NRPS. Considering the typical M_x response shown in Fig. 3.5(b) of the thesis, two parameters must be optimized for a critically-damped response. Here, the magnitude of the primary peak of M_x must be maximized, while the amplitude of the residual precessional oscillations must be minimized. Ideally, the peak value of M_x/M_S equals 1 and the oscillation amplitude is equal to 0, corresponding to a critically-damped response.

The LLG vector equation is solved for various combinations of static and transient peak magnetic fields. Both the magnitude of the primary peak and the oscillation amplitude are plotted as functions of H_z for different values of $h_{x,pk}$, as shown in Fig. C.1(a)-(b) for the $\tau_p = 500$ps exemplary scenario. As can be seen in Fig. C.1(a), the primary peak in the response will exhibit a unity maxima that shifts to higher static fields as the transient magnetic field peak is increased. Conversely, the amplitude of the oscillations, shown in Fig. C.1(b), exhibit minima at multiple values of the static magnetic field. Only the first minimum is of interest here. With successive iterations, a parameter set can be found such that both of the observed features occur for the same combination of static/dynamic magnetic field values. Figure C.1(c) depicts both curves for a pulse with $\mu_0 h_{x,pk} = 54$mT. For a static magnetic field of 47.8mT, both features occur simultaneously, and thus the critical damping point has been determined. Similar analysis yields the critical damping parameters for other values

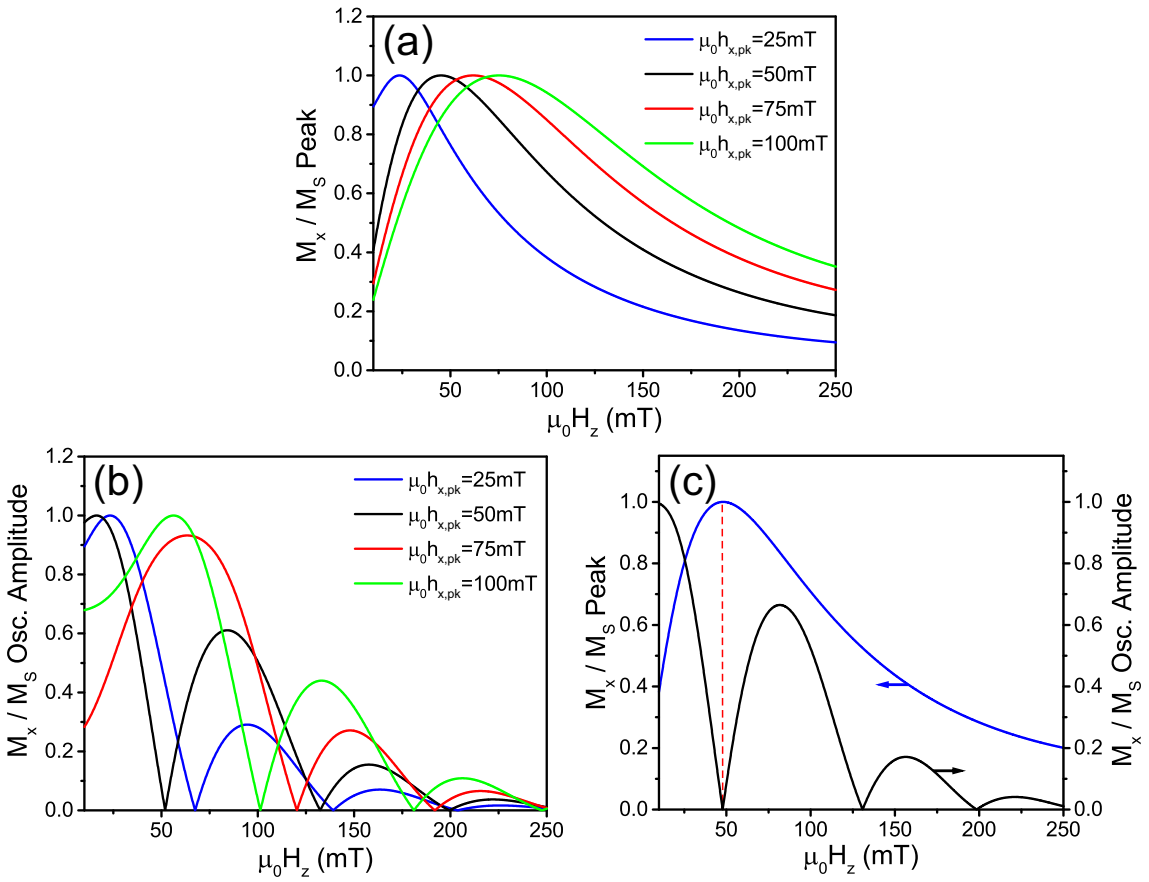

Figure C.1: (a) Magnitude of the primary peak of the M_x/M_S response as a function of static magnetic field for $\tau_p = 500$ps in the L-configuration. (b) Amplitude of the residual precessional oscillations as a function of static magnetic field strength for $\tau_p = 500$ps in the L-configuration. (c) Primary peak magnitude and oscillation amplitude for a $\tau_p = 500$ps pulse with $\mu_0 h_{x,pk} = 54$mT in the L-configuration. Note that the maximum of the primary peak response occurs simultaneously with the first minimum of the oscillation amplitude at a static magnetic field of $\mu_0 H_z = 47.8$mT, as indicated by the dashed red line.

of τ_p in the L-configuration.

When considering the T-configuration, analogous expressions are taken for the appropriate field components, and the LLG vector equation is again numerically solved to determine the motion of the magnetization vector, **M**. The parameters that must be optimized in this configuration differ slightly from the L-configuration. Here, one must consider both the value of the primary peak in the M_x response and the value that M_x returns to immediately following excitation. The final state of M_x is intimately related to the precessional oscillations that will occur in M_z. Since the magnitude of **M** is a constant (M_S) and this particular arrangement forces the precession of

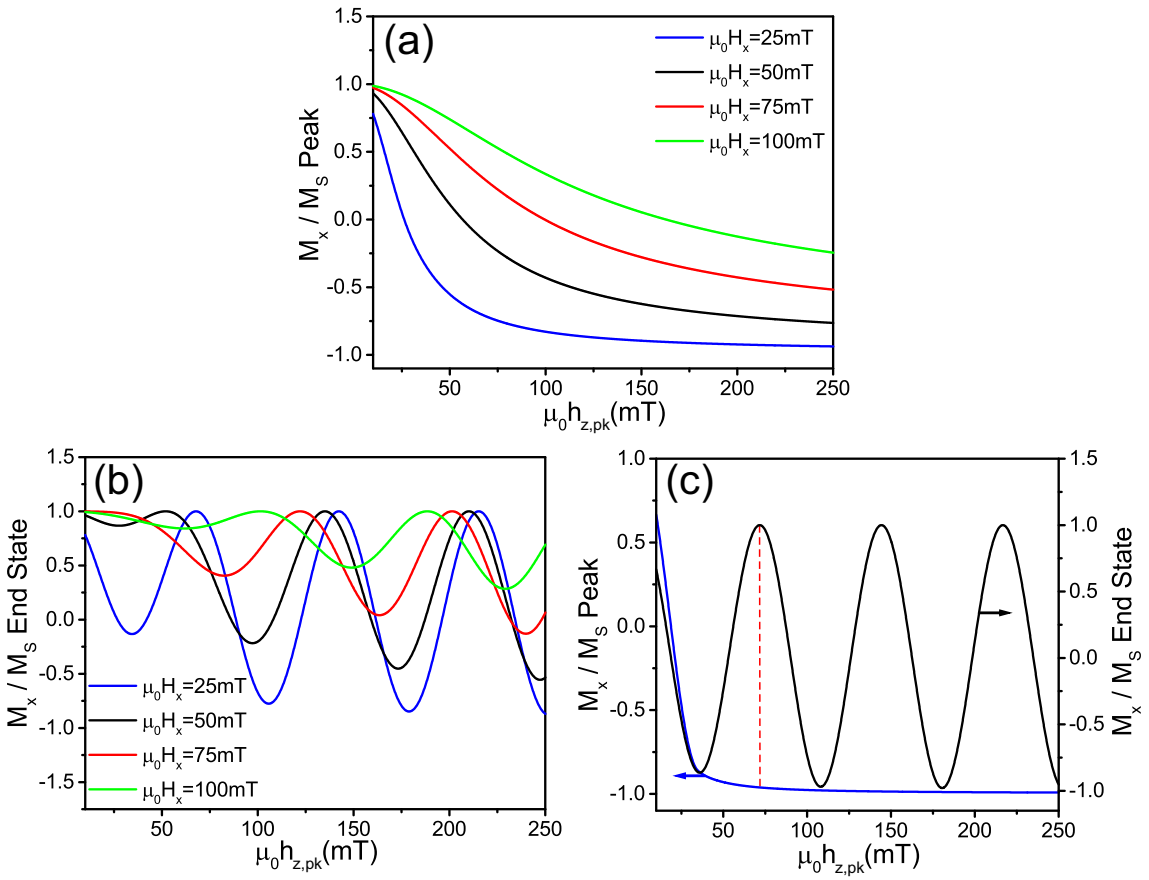

Figure C.2: (a) Primary peak of the M_x/M_S response as a function of the peak transient magnetic field for a $\tau_p = 500$ps pulse in the T-configuration. (b) M_x state achieved after deflection as a function of the peak transient magnetic field for a $\tau_p = 500$ps pulse in the T-configuration. (c) Both the normalized peak and the end state response of M_x for a $\tau_p = 500$ps pulse with a static magnetic field of $\mu_0 H_x = 9$mT in the T-configuration. The dashed red line indicates the optimum operating conditions.

the magnetization around the x-axis, a large amplitude of oscillations in M_z implies that M_x has not returned to its initial state. Note that over a time scale of several hundred nanoseconds, the initial magnetization state will be restored as the oscillations in M_z diminish; however, this is undesirable for efficient operation at high speeds. Thus, in this configuration, it is desirable to have a peak response of $M_x/M_S = -1$, after which M_x returns back to its initial saturated state of $M_x/M_S = 1$, implying minimal precession, and hence a critically-damped response.

In this case, it is illustrative to consider parameters as functions of the peak magnetic field, $h_{z,pk}$. Figure C.2(a) displays the peak in the M_x/M_S response as a function

of $h_{z,pk}$ for various values of the static field, at $\tau_p = 500$ps. For a set static magnetic field, H_x, increasing the peak magnetic field strength deflects the magnetization further toward $M_x/M_S = -1$, as desired. Decreasing H_x further enhances this effect, and allows more deflection to be achieved at a lower transient field. Note that the limiting case of $\mu_0 H_x$ is 9mT, as the static magnetic field must be able to saturate the magnetization. Figure C.2(b) depicts the value that M_x/M_S returns to after deflection, having oscillatory behaviour as a function of $h_{z,pk}$. The response exhibits multiple maxima of $M_x/M_S = 1$. As well, these maxima tend to shift to lower transient field values as the static magnetic field is increased. However, the trends for the two parameters are opposing. To maximize the peak **M** deflection at lower transient magnetic field strengths, H_x must be reduced, while conversely, in order to maximize the end state of M_x at lower values of $h_{z,pk}$, one requires H_x to be increased.

In the current application, achieving the maximum possible phase shift is desirable, and thus, the operating conditions are taken with $\mu_0 H_x = 9$mT. This ensures that the peak deflection is as large as possible. At $\mu_0 H_x = 9$mT, the peak magnetic field is chosen to maximize the M_x end state at the lowest possible value of $h_{z,pk}$, i.e., $h_{z,pk}$ then corresponds to the first maximum in the end state response, as shown in Figure C.2(c). Accordingly, the peak deflection is $M_x/M_S = -0.96$.

Calculation of the Magnetic Field Distribution and Ag Slab Design

The dimensions of the Ag slabs, as well as the required currents, are determined based on the criterion that they must produce a magnetic field distribution that exhibits little variation over the waveguide's volume (i.e., <2.5% variation), and thus, can be considered uniform. The spatial dependence of the magnetic fields is highly dependent upon the current distribution within the Ag wire. At the frequencies or

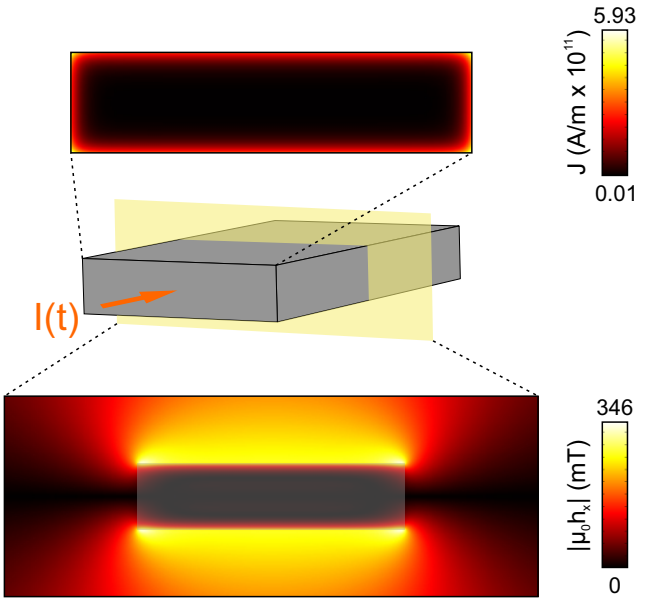

Figure C.3: Current density distribution (top) and $|\mu_0 h_x|$ distribution (bottom) for a device operating in the L-configuration with $\tau_p = 100$ps pulses, $I_{peak} = 24.19$A, and Ag transmission line dimensions of 40μm $\times$ 10μm $\times$ 49.06μm.

bit rates examined in the main text (1.21GHz and 6.06GHz), the current density is concentrated near the surface of the conductor due to the skin effect, as shown in Fig. C.3 for the $\tau_p = 100$ps, L-configuration case. Note that the skin depths at 1.21GHz and 6.06GHz are 1.823μm and 815nm, respectively. Accounting for this, the magnetic fields induced by these currents at any point in space are numerically calculated. In the L- and T- configurations, the transient x- and z-components are of interest, respectively. An example, the spatial distribution of the x-component of the magnetic fields due to the aforementioned current distribution is shown in Fig. C.3.

The magnetic fields present over the cross section of the waveguide normal to the direction of current flow are examined, and the maximum variation over this region is calculated. For devices operating in the L-configuration, this cross section is 270nm$\times$300nm, while in the T-configuration the cross section is 49.06μm$\times$300nm. Dimensions of the transmission line and the total current are then iteratively optimized until the required magnetic field variation around the critically damped conditions

have been achieved, and the resultant parameters for each case are detailed in Table 3.1 in the main text. With the aforementioned parameters, the variation of the magnetic field is 0.74%-2.42%, and thus, the magnetic fields over the waveguide volume can be considered constant.

The required high current pulses can be generated in a variety of ways. In addition to commercial current sources (Montena model GP5-01-05), compact circuits have been developed to drive laser diodes with currents in excess of 10A, and nanosecond pulse lengths [224, 225]. In this manner, GaAs transistors operating in avalanche mode have been shown to produce short current pulses up to 120A [226]. Additionally, common transmission line transformers can be incorporated into the chip design to step currents up to a sufficiently high value [227].

We can estimate the power dissipation by first determining the root-mean-square (RMS) current. This is done by approximating the super-Gaussian pulses as rectangular pulses with a duration, τ_p, and repetition rates, τ_{rep}, of either 825ps (1.21Gbit/s with $\tau_p = 500$ps) or 165ps (6.06Gbit/s with $\tau_p = 100$ps). The RMS current of such a pulse is given by the relation $I_{RMS} = I_{peak}\sqrt{D}$, where D is the duty cycle of the pulse, defined as the ratio of τ_p and τ_{rep}. The effective resistance of each transmission line is then determined considering an effective area within one skin depth from the conductor's edges. The $I_{RMS}^2 R$ power dissipation is estimated to be 3.51W and 22.71mW for the $\tau_p = 100$ps and $\tau_p = 500$ps pulses in the L-configuration. Similarly, in the T-configuration, the estimated power dissipation is 210.74mW and 2.37mW for the two pulse widths respectively. Thus, power dissipation should not be a significant issue with the incorporation of a sufficient sink.

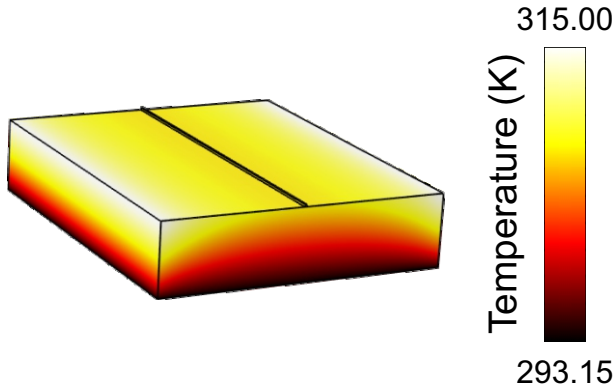

Figure C.4: Temperature profile of the device operating in the L-configuration with $\tau_p = 100$ps pulses, and $I_{peak} = 24.19$A. The maximum temperature increase over the device volume is $\Delta T_{ss,max} = 21.85$K above room temperature.

Thermal Analysis

With currents passing through small metal structures, the temperature of each device needs to be carefully assessed for practical implementation. As such, temperature distributions are calculated via the finite element method in *COMSOL Multiphysics* [228] for the magnetoplasmonic devices with the dimensions and current parameters determined earlier.

Since the phase shifter will typically be operated with multiple pulses representing an encoded data sequence, the steady state temperature that will be reached by the device due to repetitive pulsing must be considered. To calculate the heat dissipated under continuous AC operation, a pulse train is approximated as a sinusoidal signal with a frequency equal to the pulse repetition rate (i.e., 1.21GHz for $\tau_p = 500$ps pulses and 6.06GHz for $\tau_p = 100$ps pulses). Utilizing finite element calculations, the current distribution in the transmission line is determined at the required frequency, allowing for the incorporation of Joule heating due to the nonuniform current distribution caused by the skin effect. The temperature profile is then numerically calculated to determine the maximum temperature in the structure, $T_{ss,max}$, and the maximum difference from a room temperature of 293.15K, $\Delta T_{ss,max}$. Here, standard convective boundary conditions are employed, while the bottom surface is held at room

temperature.

Maximum steady state temperatures of the device are calculated to be $T_{ss,max} =$ 315.00K ($\Delta T_{ss} =$ 21.85K) and $T_{ss,max} =$ 293.29K ($\Delta T_{ss} =$ 0.14K) for $\tau_p =$ 100ps and $\tau_p =$ 500ps pulses respectively in the L-configuration. In the T-configuration, the steady state temperatures are $T_{ss,max} =$ 298.05K ($\Delta T_{ss} =$ 4.90K) and $T_{ss,max} =$ 293.29K ($\Delta T_{ss} =$ 0.14K) for the two pulse widths respectively. Clearly, even in the worst case scenario, the L-configuration with $\tau_p =$ 100ps (as shown in Fig. C.4), this temperature increase is far below any compromising value.

Appendix D

Mechanical Designs

This appendix presents the mechanical drawings for the various mechanical mounts and sample holders designed throughout the development of the fabrication processes in Chapter 5, and for the experiments detailed in Chapter 6.

Figures D.1-D.3 depict the designs for components used in the fabrication of the ME nanoplasmonic waveguides. These include the sample holder used for the double exposure photolithography process (Fig. D.1), the milled SEM stub (Fig. D.2), and for the mount used to place the milled SEM stubs and samples into the magnetron sputtering system (Fig. D.3).

Figures D.4-D.6 illustrate the plans for various components used in the experiments. Figure D.4 illustrates the modifications made to the characterization system sample stage required to insert the samples on milled SEM stubs. Figures D.5 and D.6 contain the designs for mounts design to hold the optical fiber into the spectrometer, and to hold the camera for monitoring the samples.

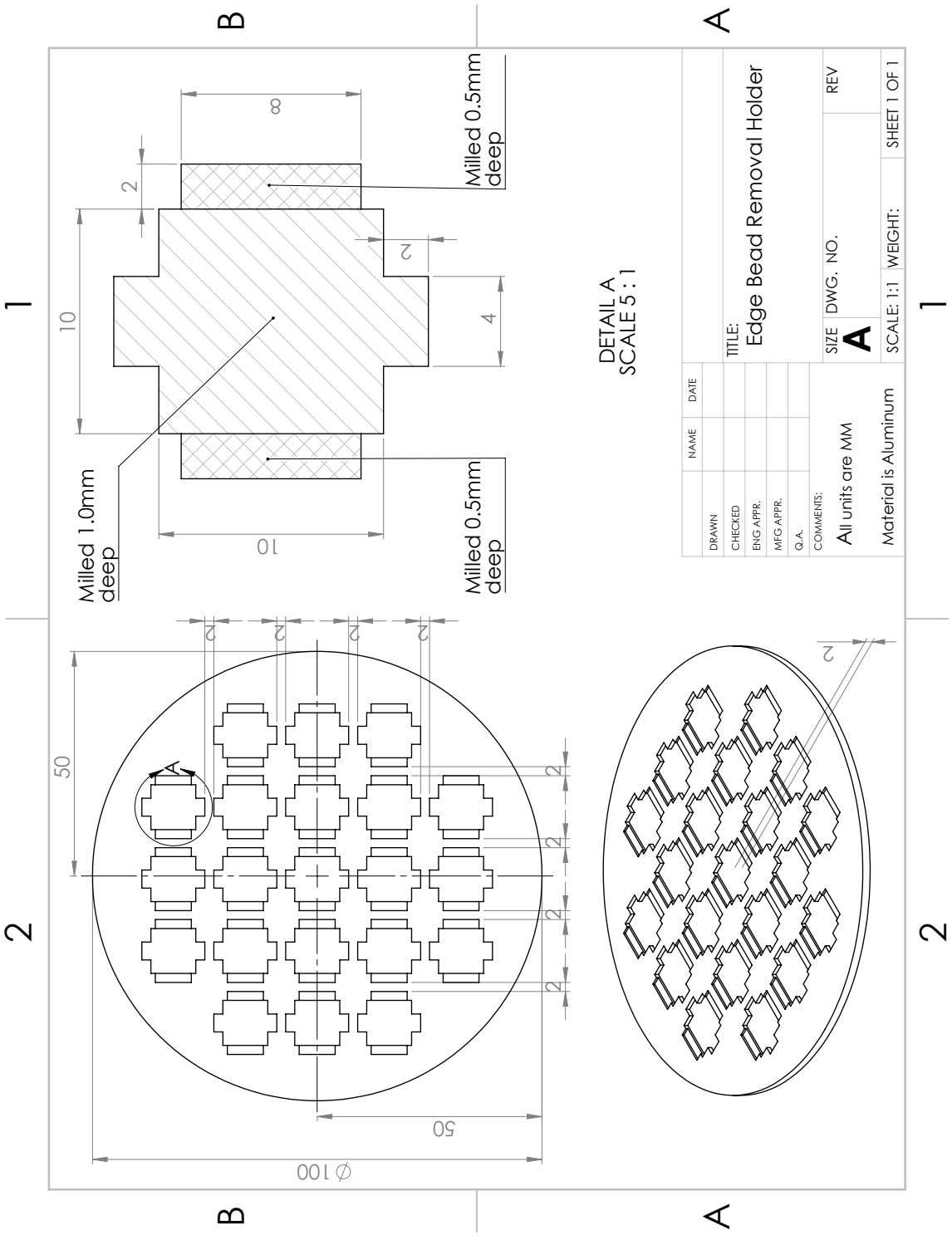

Figure D.1: Mechanical drawing for the sample holder used in the double exposure process to remove the photoresist edge bead.

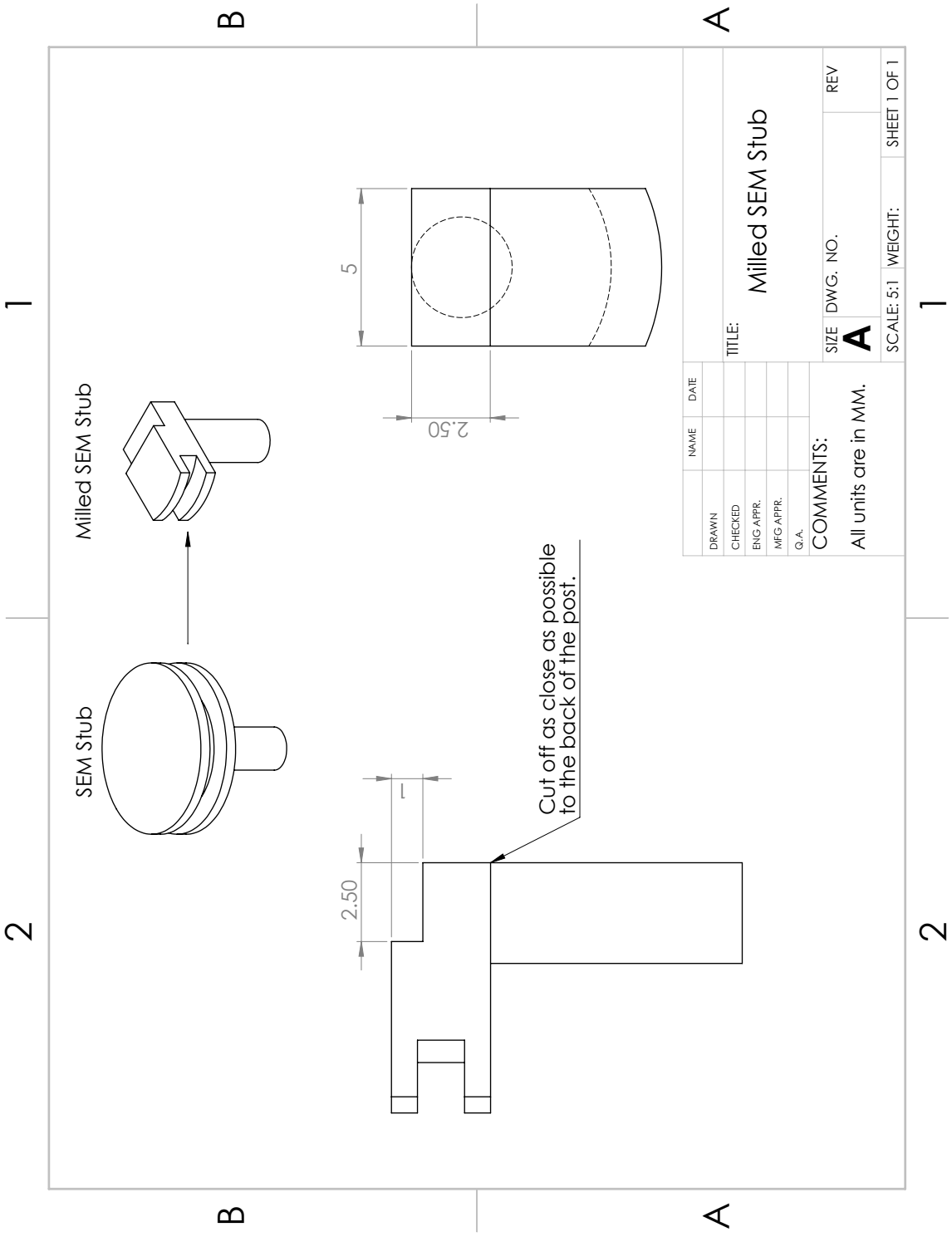

Figure D.2: Mechanical drawing for the milled SEM stub sample holder.

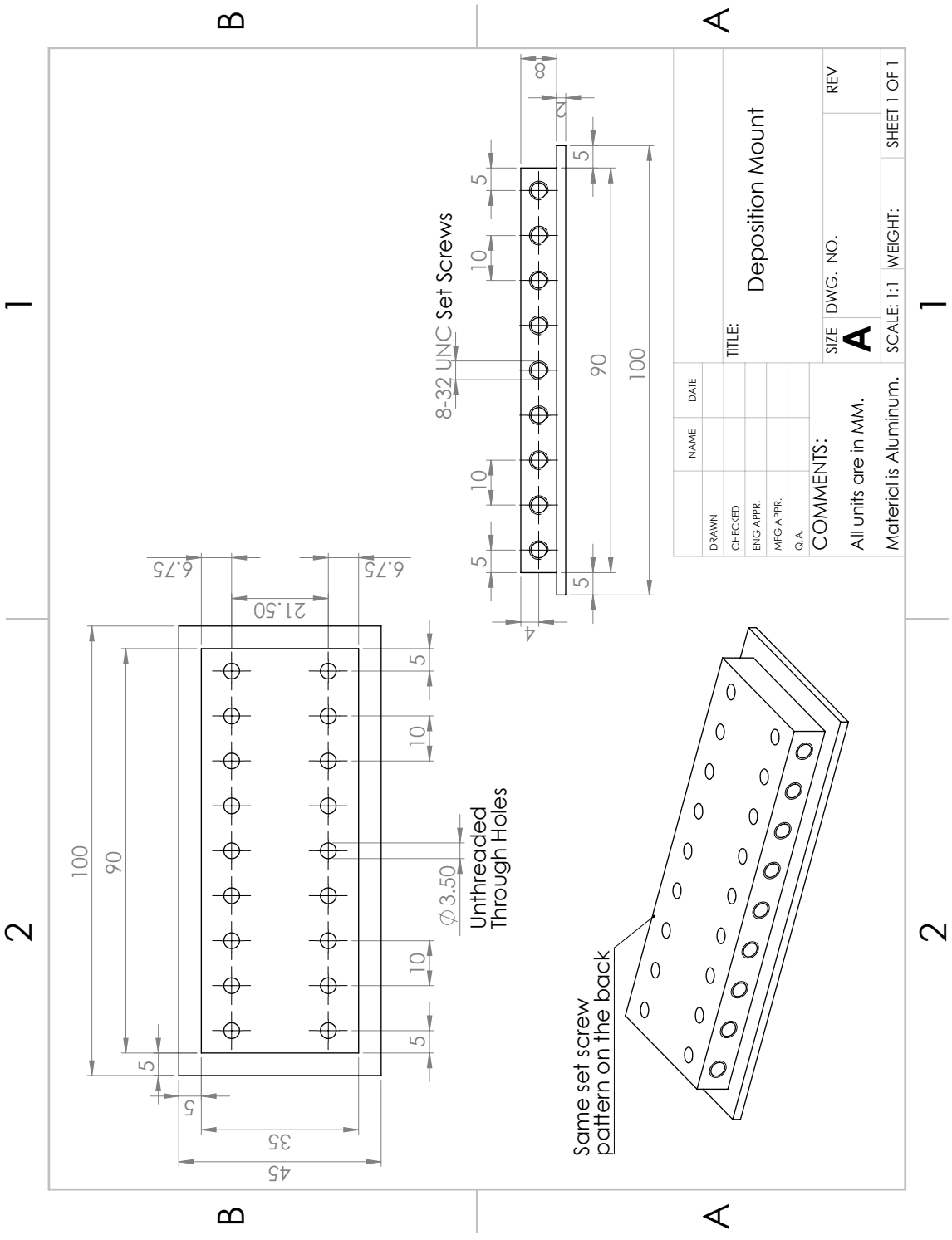

Figure D.3: Mechanical drawing for the mount utilized to hold the milled SEM stubs containing the individual samples in the magnetron sputtering system.

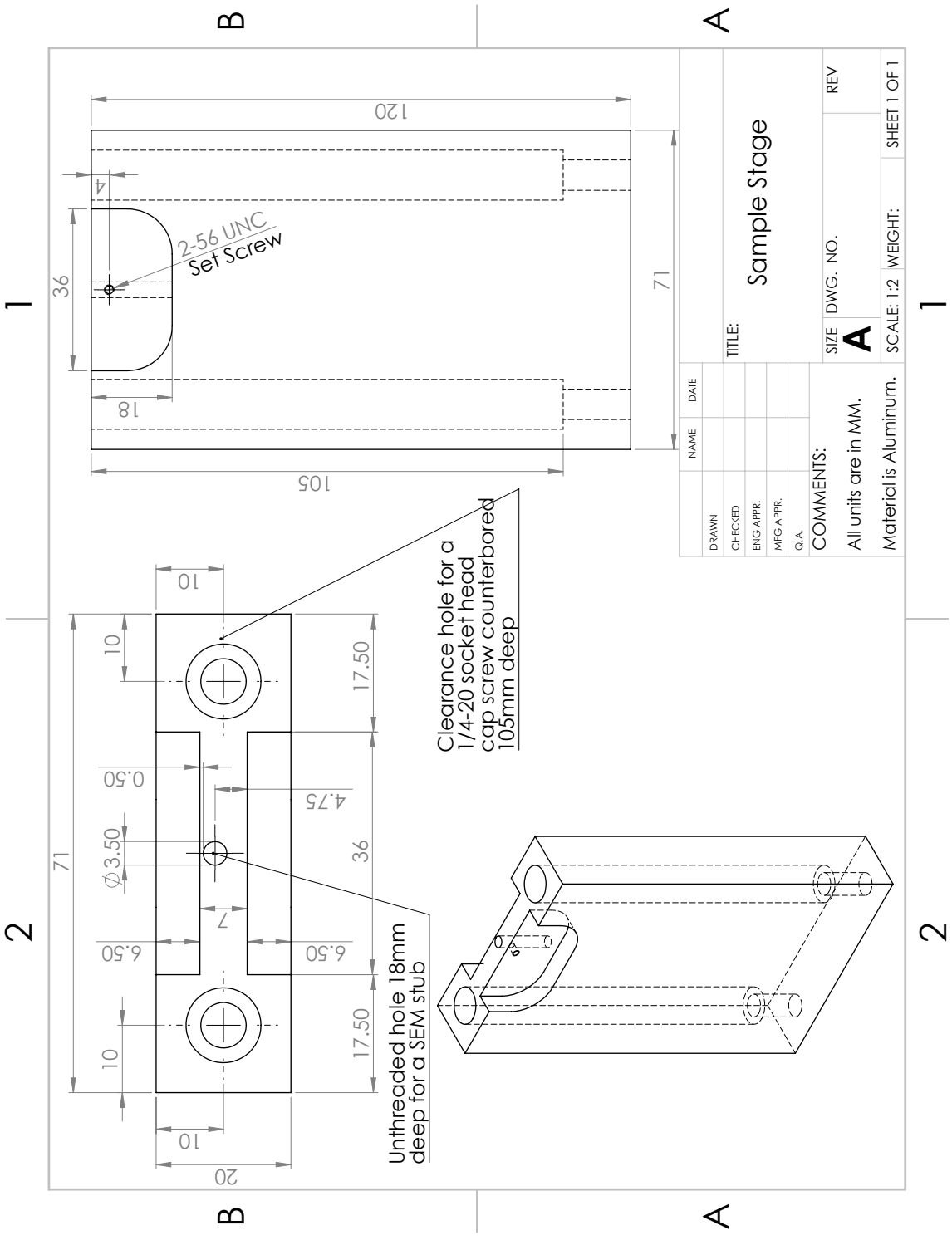

Figure D.4: Mechanical drawing for the modified sample stage designed to hold the milled SEM stubs and samples in the characterization system.

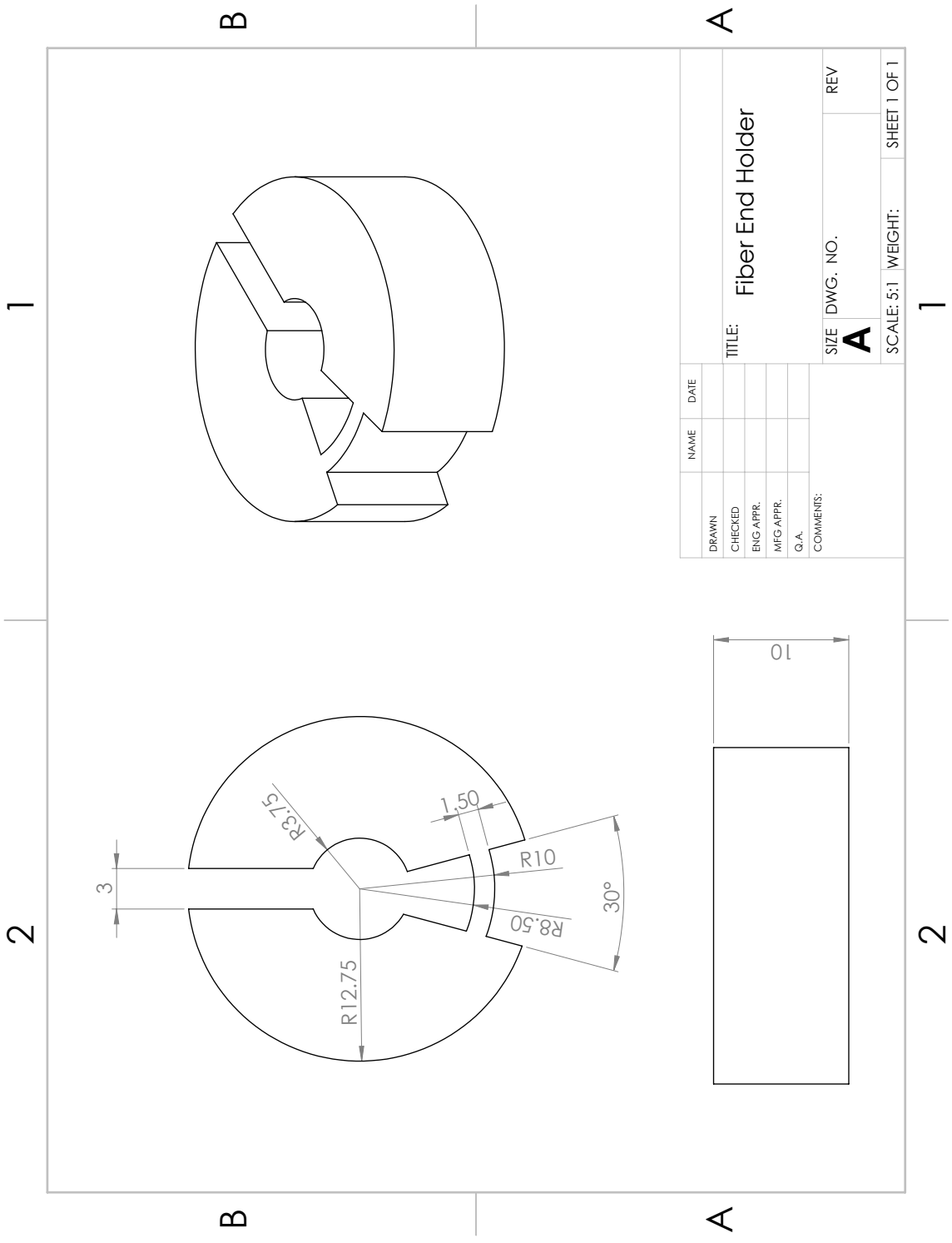

Figure D.5: Mechanical drawing of the mount used to insert the end of the optical fiber into the spectrometer.

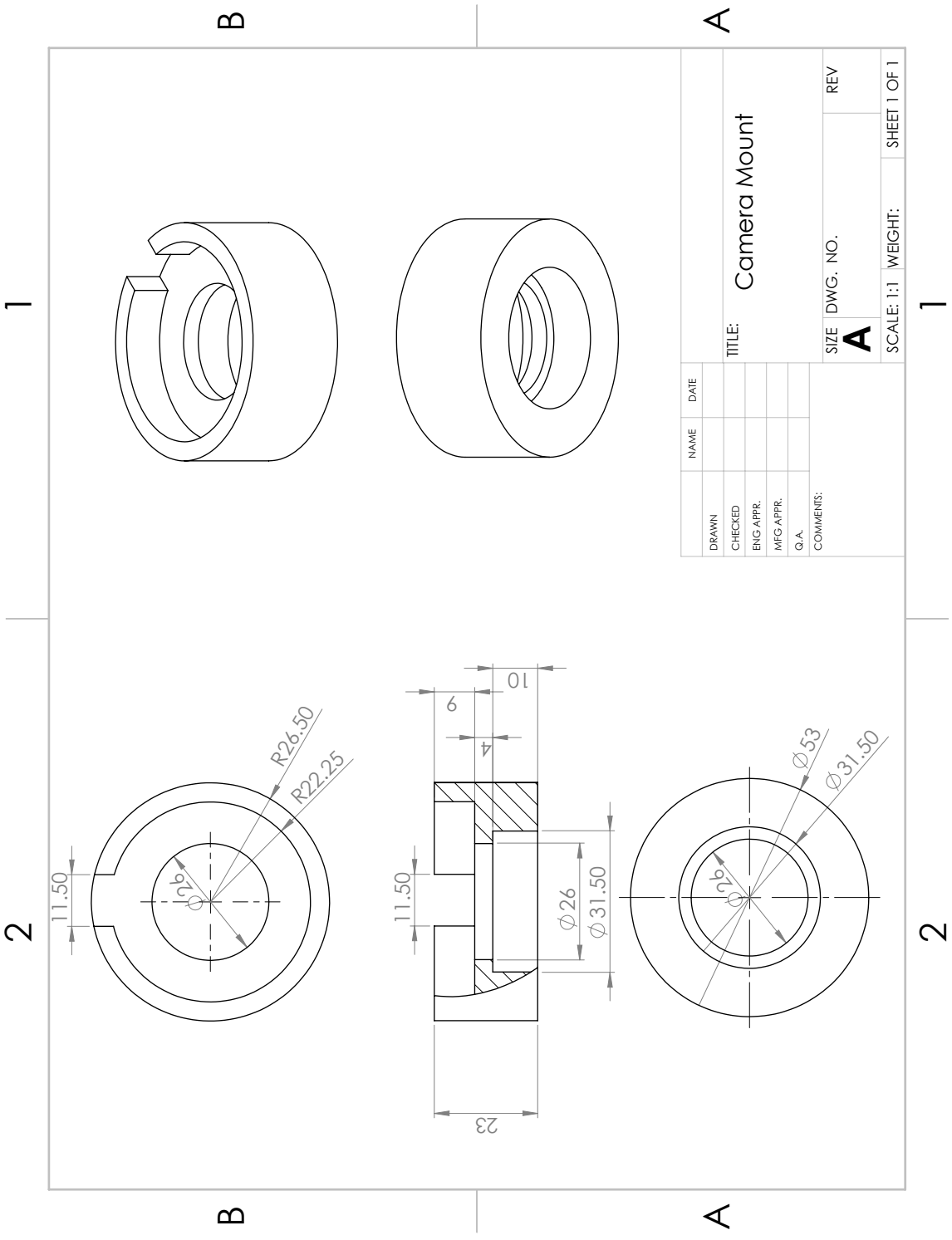

Figure D.6: Mechanical drawing of the mount used to hold the camera in the characterization system for imaging.

Appendix E

Nonlinear Si Material Plugin

In order to accurately compare the experimental results presented in Chapter 6 to theoretical values, the material model implemented in FDTD must account for all of the various nonlinear effects that occur within silicon. Namely, $\chi^{(3)}$ effects such as THG, the optical Kerr effect, and SPM, as well as effects related to the generation and absorption of free carriers, such as TPA, FCA, and plasma dispersion must be included. Unfortunately, a detailed material model is not standard within commercial software such as *Lumerical FDTD Solutions* [119], and thus, must be implemented by the user. However, Lumerical does provide a framework for custom material plugins, providing that the user can develop his/her own algorithm to calculate the nonlinear polarization.

The algorithm employed with the present plugins was originally presented by Suzuki [215]. It accurately accounts for all of the aforementioned nonlinear phenomena. The implementation presented below is versatile, as all of the material parameters can be input through the material interface of *Lumerical FDTD Solutions* without need to recompile the plugin. Moreover, this implementation allows the extraction of various parameters, such as the carrier concentrations or the change in refractive index due to plasma dispersion, through the added storage fields. Ad-

ditionally, this implementation facilitates a mechanism of turning various effects on and off, and thus, can allow for simulation of a subset of the nonlinear effects, which is often useful for debugging divergent simulations or determine their relative contributions. As well, this implementation is designed to take either a $\chi^{(3)}$ value or an n_2 value. This was necessary as both are presented in literature, and there are numerous conventions that can cause discrepancies when converting between the two [103]. The output of the in-house developed plugin was rigorously tested and validated against the theoretical values calculated from the equations presented in Chapter 2 to ensure its accuracy.

The following files were used to generate the plugin for including the nonlinearities in *Lumerical FDTD Solutions* simulations. Parameters such as constants and storage fields are initialized in the *NonlinearSi.h* header file, while the algorithm is implemented in the *NonlinearSi.cpp* source code file. It requires the *imaterialplugin.h* header file provided by Lumerical [119] to compile correctly, which is presented below for reference. Compiling these files together generates a *.dll* file named *NonlinearSi.dll*, which can then be included in the software.

Header File: imaterialplugin.h

```
#ifndef _IMATERIALPLUGIN_H
#define _IMATERIALPLUGIN_H

#include <stddef.h>

/*!
        \brief The interface class definition for a material plugin for FDTD
            Solutions

        This pure abstract class defines the methods that must be
           implemented to create a plugin material in FDTD Solutions. A
           class that inherits from this interface class and implements all
            the methods can be compiled into a dynamic library that can be
           loaded into FDTD Solutions to define new materials.
```

```
*/

class IMaterialPlugin
{
public:
	virtual ~IMaterialPlugin(){};
	virtual const char* name() const = 0;
	virtual const char* uniqueId() const = 0;
	virtual const char** parameterNames() const = 0;
	virtual float calculateEx( float U, float V, float Ex, float*
	    storage) = 0;
	virtual float calculateEy( float U, float V, float Ey, float*
	    storage) = 0;
	virtual float calculateEz( float U, float V, float Ez, float*
	    storage) = 0;
	virtual void initialize(const double** parameters, double dt) = 0;
	virtual void initializeStorageEx(float* storage) = 0;
	virtual void initializeStorageEy(float* storage) = 0;
	virtual void initializeStorageEz(float* storage) = 0;
	virtual size_t storageSizeE() const = 0;
};

/*!
	\brief The interface class for a magnetic material plugin

	This extends the material plugin defined above with a few more
	    methods that need to be defined for a magnetic material
*/
class IMagneticMaterialPlugin : public IMaterialPlugin
{
public:
	virtual float calculateHx( float U, float V, float Ex, float*
	    storage) = 0;
	virtual float calculateHy( float U, float V, float Ey, float*
	    storage) = 0;
	virtual float calculateHz( float U, float V, float Ez, float*
	    storage) = 0;
	virtual void initializeStorageHx(float* storage) = 0;
	virtual void initializeStorageHy(float* storage) = 0;
	virtual void initializeStorageHz(float* storage) = 0;
	virtual size_t storageSizeH() const = 0;
};

/*!
	\brief The interface for a factory class that creates and destroys
	    material plugins
*/
```

```
class IMaterialPluginFactory{
public:
       virtual IMaterialPlugin* createInstance()=0;
       virtual void destroyInstance(IMaterialPlugin* i)=0;
       virtual IMagneticMaterialPlugin* toMagneticMaterialPlugin(
          IMaterialPlugin* p)=0;
};

/*!
       \brief A templated implementation of the IMaterialPluginFactory
          class

       Plugin authors do not need to write a factory class, they can just
          use this class. It is written as a template so that it can be
          compiled into the plugin easily. This is done in the plugin code
           usign the MATERIAL_PLUGIN(T) macro.
*/
template<class T>
class MaterialPluginFactory : public IMaterialPluginFactory
{
       IMaterialPlugin* createInstance(){return new T();}
       void destroyInstance(IMaterialPlugin* i){delete i;}
       IMagneticMaterialPlugin* toMagneticMaterialPlugin(IMaterialPlugin* p
          ){return dynamic_cast<IMagneticMaterialPlugin*>(p);}
};

#ifdef WIN32
#define DLLEXPORT __declspec(dllexport)
#else
#define DLLEXPORT
#endif

//A macro to add the factory function to the plugin, instantiating the
   MaterialPluginFactory in the process
//All plugins should include this macro once in a source file. The argument
     T is the name of the user's plugin class
#define MATERIAL_PLUGIN(T) \
       extern "C" DLLEXPORT IMaterialPluginFactory* createFactoryV1(){
          return new MaterialPluginFactory<T>();} \
       extern "C" DLLEXPORT void destroyFactoryV1(IMaterialPluginFactory* f
          ){ delete f;}

#endif
```

Header File: NonlinearSi.h

```
#ifndef _NONLINEARSI_H
#define _NONLINEARSI_H

#include "imaterialplugin.h"

class NonlinearSiPlugin : public IMaterialPlugin
{
    public:
    NonlinearSiPlugin(){};
    virtual ~NonlinearSiPlugin(){};

    const char* name() const {return "Nonlinear Si";};
    const char* uniqueId() const {return "{4E1E0F3F-68C3-4ed2-88FC-
        EED2A42E8024}";};
    const char** parameterNames() const {return names;};
    float calculateEx( float U, float V, float Ex, float* storage);
    float calculateEy( float U, float V, float Ey, float* storage);
    float calculateEz( float U, float V, float Ez, float* storage);
    void initialize(const double** parameters, double dt);
    void initializeStorageEx(float* storage){};
    void initializeStorageEy(float* storage){};
    void initializeStorageEz(float* storage){};
    size_t storageSizeE() const {return 18;}; // this line returns the
        number of storage fields required; only the number required

    private:
    float calculate(int axis, float U, float V, float E, float* storage);

    //define the constants
    float c1[3];
    float c2[3];
    float c3[3];
    float c4[3];
    float c5[3];
    float c6[3];
    float c7[3];
    float c8[3];
    float c9[3];
    float c10[3];
    float c11[3];

    //define the initial and intrinsic free carrier concentrations and
        iteration
    float Ninit[3];
    float Nfi[3];
    float iter[3];
```

```
    //define other variables
    float dnplas_np1;
    float chiKerr_np1;
    float g1;
    float g2;

    //number of user specified properties +1
    static const char* names[17];

    static const double PI;
    static const double C0;
    static const double E0;
    static const double HBAR;
};

#endif
```

C++ Source File: NonlinearSi.cpp

```
#include "NonlinearSi.h"
#include <cmath>

/*!
    \class NonlinearSiPlugin

    \A material plugin for Lumerical FDTD solutions to model nonlinear
       effects in Si. This plugin is based on the algorithm in [1] below.
    \This plugin accounts for THG, the Kerr effect, TPA, FCA, and plasma
       dispersion. Raman and thermal effects are not included.
    \There is an assumption that the material response for the Kerr effect
       is instantaneous. This is acceptable providing the pulse length is
       not in the sub-10fs regime, comparable to the response time.
    \Although not hardcoded, this plugin is designed to be used around 1550
       nm, so the input n0 and lambda0 properties should be the refractive
       index of Si at 1550nm and 1550nm.
    \This plugin assumes that the base material in Lumerical was selected
       to be Si.
    \All units are SI.

    \Note that the way Lumerical's algorithm works, generated carriers are
       calculated for each component, based on the intensity of that field
       component alone.
    \This cannot be changed, as the field component calculations are
       independent, and a parameter cannot be passed between them (the
       storage fields are generated for each component, not passed between
       them).
```

```
    \If the fields are linearly polarized, or dominated by one component,
       the majority of the carriers will be generated by that component, so
        little error is introduced here.
    \However, this may cause issue when including plasma dispersion, as the
        different components will experience different refractive index
       changes, and hence the material will become artificially
       birefringent.
    \This will lead to different components travelling at different speeds
       within the material, and must be carefully checked.

    \Convergence testing through the iteration parameter is recommended, as
        the electric field at each timestep is calculated through an
       iterative loop

    \The code is written so that the user may generate third order
       nonlinear effects by inputting either chi3 or n2. Note that only one
        should be enabled at any given time, otherwise the nonlinearity
       will be effectively doubled.
    \Due to the convention taken here that P=e0[chi1 + chi3*E^2]*E, no
       factor of 3/4 is present. This is assumed to be absorbed into the
       value for chi3. As such, when converting between n2 and chi3, one
       should use the formula
    \n2 = chi3/(e0*c*n0^2). This convention aligns the results with the
       built in chi3 model present in Lumerical, and converting between
       chi3 and n2 in this way will produce equivalent results. The
       discrepancy in relations between
    \n2 and chi3 is discussed in detail in Appendix 5 of [2].

    [1] N. Suzuki, "FDTD analysis of two photon absorption and free-carrier
        absorption in Si high-index-contrast waveguides," IEEE J. Lightwave
        Technol. 25(9), 2495-2501 (2007).
    [2] P. N. Butcher and D. Cotter, The Elements of Nonlinear Optics (
       Cambridge University Press, 1990).
*/

//define constants
const double NonlinearSiPlugin::PI = 3.1415926535897931;
const double NonlinearSiPlugin::C0 = 2.99792458e8;
const double NonlinearSiPlugin::E0 = 8.854187817e-12;
const double NonlinearSiPlugin::HBAR = 1.05457148e-34;

//names of user specified parameters; last value must be 0
const char* NonlinearSiPlugin::names[17] = {"lambda0 (m)", "n0", "chi3 (m
   ^2/V^2)", "n2 (m^2/W)", "betaTPA (m/W)", "sigma0 (m^2)", "Ninitial (m
   ^-3)", "Nfintrinsic (m^-3)", "tauR (s)", "iteration", "Enable chi3 (1
   or 0)", "Enable n2 (1 or 0)", "Enable FCG (1 or 0)", "Enable FCA (1 or
   0)", "Enable TPA (1 or 0)", "Enable PD (1 or 0)",0};
```

```
void NonlinearSiPlugin::initialize(const double** parameters, double dt)
{
    for(int i=0; i<3; i++)
    {
        double lambda0 = parameters[0][i];
        double n0 = parameters[1][i];
        double chi3 = parameters[2][i];
        double n2 = parameters[3][i];
        double betaTPA = parameters[4][i];
        double sigma0 = parameters[5][i];
        Ninit[i] = float( parameters[6][i] );
        Nfi[i] = float( parameters[7][i] );
        double tauR = parameters[8][i];
        iter[i] = float( parameters[9][i] );

        double enableChi3 = parameters[10][i];
        double enableKerr = parameters[11][i];
        double enableFCG = parameters[12][i];
        double enableFCA = parameters[13][i];
        double enableTPA = parameters[14][i];
        double enablePD = parameters[15][i];

        double w0 = 2.*PI*C0/lambda0;
        double sigmaFCA = sigma0 * lambda0*lambda0/(1.55e-6)/(1.55e-6);

        c1[i] = float( enableFCG*(2.*tauR - dt) / (2.*tauR + dt) );
        c2[i] = float( enableFCG*(tauR*dt*C0*C0*E0*E0*n0*n0*betaTPA)/((2.*
            tauR+dt)*4.*HBAR*w0) );
        c3[i] = float( -8.8e-22 );
        c4[i] = float( -8.5e-18 );
        c5[i] = float( 0.8 );
        c6[i] = float( enablePD*lambda0 / (1.55e-6) );
        c7[i] = float( enableKerr*C0*E0*n0*n0*n2 );
        c8[i] = float( enableChi3*3.*chi3/2. );
        c9[i] = float( enablePD*2.*n0 );
        c10[i] = float( enableFCA*C0*n0*dt*sigmaFCA/2. );
        c11[i] = float( enableTPA*C0*C0*E0*n0*n0*betaTPA*dt/8. );

    }
}

float NonlinearSiPlugin::calculate(int axis, float U, float V, float E,
    float* storage)
{
    // pull values from storage
    float V_n = storage[0];
```

```
float U_n = storage[1];
float Nf = storage[2];
float dnplas_n = storage[3];
float chiKerr_n = storage[4];
float firstUpdate = storage[5];
//initially set both E_n and E_np1 to E at the previous time step (E_n)
//E_np1 will be updated and returned later. This is needed for the
    iteration
float E_n = E;
float E_np1 = E;
float V_np1 = V;
float U_np1 = U;

float ind = 1.;

//update Nf
Nf = c1[axis]*Nf+c2[axis]*abs(E_n)*abs(E_n)*abs(E_n)*abs(E_n);
storage[17] = c2[axis]*abs(E_n)*abs(E_n)*abs(E_n)*abs(E_n);
//on the first update, set Nf to the specified initial concentration
if(firstUpdate == 0.)
{
    Nf = Ninit[axis];
    firstUpdate = 1.;
}
//check that Nf is not less than the intrinsic free carrier density
// Nf cannot fall below the intrinsic value
if(Nf<Nfi[axis])
{
    Nf = Nfi[axis];
}

// Note that this calculation requires the carrier concentrations to be
     in cm^-3
// Thus, the carrier values are multiplied by 1e-6 to convert from m^-3
// The values of c3 and c4 cannot be scaled due to the N^0.8 dependence
// The N^0.8 dependence comes from experimenta values in Ref. 35 of
    Suzuki's paper.
// Also, c6 does not need to change, as the units cancel out.
float Nf_cm3 = Nf*0.000001;
dnplas_np1 = ( c3[axis]*Nf_cm3 + c4[axis]*pow(Nf_cm3,c5[axis]) ) * c6[
    axis]*c6[axis];

//iteratively calculate chiKerr, g1, g2, then E_np1
while(ind<=iter[axis])
{
    chiKerr_np1 = c7[axis] * abs(E_np1)*abs(E_np1);
    g1 = c8[axis]*(E_np1*E_np1 + E_n*E_n) + chiKerr_np1 + c9[axis]*
```

```
            dnplas_np1 + c10[axis]*Nf + c11[axis]*(abs(E_np1)*abs(E_np1) +
            abs(E_n)*abs(E_n));
        g2 = c8[axis]*(E_np1*E_np1 + E_n*E_n) + chiKerr_n + c9[axis]*
            dnplas_n - c10[axis]*Nf - c11[axis]*(abs(E_np1)*abs(E_np1) + abs
            (E_n)*abs(E_n));

        E_np1 = (V_np1-V_n)/(g1+U_np1) + E_n * (g2 + U_n)/(g1 + U_np1);

        ind = float( ind + 1. );
    }

    // put values into storage
    storage[0] = V_np1;
    storage[1] = U_np1;
    storage[2] = Nf;
    storage[3] = dnplas_np1;
    storage[4] = chiKerr_np1;
    storage[5] = firstUpdate;

    // storage fields for debugging
    storage[6] = c1[axis];
    storage[7] = c2[axis];
    storage[8] = c3[axis];
    storage[9] = c4[axis];
    storage[10] = c5[axis];
    storage[11] = c6[axis];
    storage[12] = c7[axis];
    storage[13] = c8[axis];
    storage[14] = c9[axis];
    storage[15] = c10[axis];
    storage[16] = c11[axis];

    return E_np1;
}

float NonlinearSiPlugin::calculateEx( float U, float V, float Ex, float*
    storage )
{
    return calculate(0, U, V, Ex, storage);
}

float NonlinearSiPlugin::calculateEy( float U, float V, float Ey, float*
    storage )
{
    return calculate(1, U, V, Ey, storage);
}
```

```
float NonlinearSiPlugin::calculateEz( float U, float V, float Ez, float*
    storage )
{
    return calculate(2, U, V, Ez, storage);
}

MATERIAL_PLUGIN(NonlinearSiPlugin);
```

The noblest pleasure is the joy of understanding.
—Leonardo da Vinci