LOW NOISE ALL OPTICAL SWITCH AND GeSn LASER FOR SILICON PHOTONICS

Thesis

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Electro-Optics

By

Yun Zhao

UNIVERSITY OF DAYTON

Dayton, Ohio

May, 2016
LOW NOISE ALL OPTICAL SWITCH AND GeSn LASER FOR SILICON

PHOTONICS

Name: Zhao, Yun

APPROVED BY:

______________________________  ________________________________
Imad Agha, Ph.D.                Jay Mathews, Ph.D.
Advisory Committee Chairman     Committee Member
Assistant Professor, Department of
Physics                          Assistant Professor, Department of
                                 Physics

______________________________
Andrew Sarangan, Ph.D.
Committee Member
Professor, Electro-Optics program

______________________________  ________________________________
John G. Weber, Ph.D.            Eddy M. Rojas, Ph.D., M.A., P.E.
Associate Dean                 Dean
School of Engineering           School of Engineering

ii
ABSTRACT

LOW NOISE ALL OPTICAL SWITCH AND GeSn LASER FOR SILICON PHOTONICS

Name: Zhao, Yun
University of Dayton
Advisor: Dr. Imad Agha

Silicon based electronics revolutionized the second half of the 20th century. Moore’s law has been successfully predicting the growth of integrated circuits for more than 40 years, but the trend has slowed down. To keep meeting with the ever so strong demands, new breakthroughs are keenly needed. With the inspiring success of optical communication technology, many researchers believe this breakthrough relies on bringing optics onto silicon chips. This leads to the study of silicon photonics.

All optical switches are one of the versatile components in silicon photonics. In optical interconnections, where on chip metal wires are replaced by waveguides, it can route the signals to the right destination. Compared with traditional optical switches, an all optical switch requires less power and generates less heat, which qualities are highly valued for large data center and high performance computing applications. In the future optical computing systems, it can serve as the logic gates. It is also likely to find its place in the more advanced quantum communication networks. An electrically pumped on silicon chip laser is another fundamental component for silicon photonics. A semiconductor laser that can be reliably grown on silicon is not satisfactorily found.

In this thesis, we demonstrate two promising devices that is useful for silicon photonics: A low
noise all optical switch based on a third order nonlinear effect and a GeSn laser material that can be monolithically deposited onto silicon.
Dedicated to my parents and my uncle who encouraged the boy to chase his dream.
ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Imad Agha for his financial and mental support of my study in the University of Dayton. I cannot achieve this without his trust. I would also like to give my special thanks to Dr Jay Mathews for his valuable suggestions to my research and my career. I would then like to thank Dr. Andrew Sarangan for providing his clean room for part of the experiment and for the enlightening conversations that greatly broaden my vision. Thank Aimee Price at Ohio State University Nanotech West lab for doing the electron beam lithography. A large part of the fabrication is finished in the Nanotech West public cleanroom. Thank my colleagues at Arizona state university, Dr. James Gallagher, Dr. Jose Menendez and Dr. John Kouvetakis for providing the GeSn material for the laser experiment. Thank Dr. Andy Chong, Dr. Cong Deng and Dr. Joesph Haus at University of Dayton for generously sharing their equipments.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>I. INTRODUCTION AND MOTIVATION</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Introduction to Silicon Photonics</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Introduction to Nonlinear Optics</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Motivation of Research</td>
<td>6</td>
</tr>
<tr>
<td>II. THEORY</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>2.1</td>
<td>Theory of Guided-wave Optics</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Guided Modes</td>
<td>8</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Dispersion</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Four Wave Mixing And Nonlinear Optics</td>
<td>11</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Nonlinear Susceptibilities</td>
<td>11</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Nonlinear Maxwell’s Wave Equation</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Semiconductor Lasers</td>
<td>23</td>
</tr>
<tr>
<td>III. A NOVEL ALL-OPTICAL SWITCH</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Waveguide Design</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>Waveguide Fabrication</td>
<td>31</td>
</tr>
<tr>
<td>3.3</td>
<td>Loss Measurement</td>
<td>35</td>
</tr>
<tr>
<td>3.4</td>
<td>Four Wave Mixing Test</td>
<td>38</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Illustration of three FWM processes. The longer rods represent strong pumps. The shorter rods represent the signal and the idler. The arrows indicate photon number conservation. Each arrow stands for one photon. .................................................. 5

1.2 An analogy between electronic transistor and FWM-BS all-optical switch. ............. 6

2.1 (a) An example of rectangular waveguide. The core is composed of silicon \( n \approx 3.5 \) and the cladding is silicon dioxide \( n \approx 1.5 \). (b) The fundamental TE mode of the waveguide in (a). (c) The fundamental TM mode of the waveguide in (a) .................. 9

2.2 (a) Conversion efficiency vs. propagation distance. Parameters are \( P_1(0) = 1 \text{W}, P_2(0) = 0.05 \text{W}, P_3(0) = 0.03 \text{W}, \gamma = 125 \text{W}^{-1}\text{m}^{-1}, \Delta \beta = 30\text{m}^{-1} \). Curve (a) is analytical solution from (2.40). (b) is numerical solution for (2.26) by 4th order Runge-Kutta method. (c) is conversion efficiency upper bound estimation by letting \( P_2(z) = P_2(0) \) in (2.35). (b) Unit ellipse for the definition of elliptical function. 20

2.3 (a) Illustration of an excited indirect band gap material. (b) Excited quasi direct band gap material with n type doping. .................................................. 25

2.4 Laser output power vs. input power curves. Source: Lasers (Mill Valley, CA: University Science Books, ©1986 [1]). Using with permission. ................................. 27

3.1 (a) Group velocity dispersion curve of SOI waveguides passivated by SiO\(_2\). (b) Group velocity dispersion curve of SOI waveguides passivated by Si\(_3\)N\(_4\). .............. 29

3.2 (a) Group velocity dispersion sidewall angle dependence. (b) Structure used in simulation. [2] ................................................................. 30

3.3 (a) Effective area vs. width. The height is 500 nm for all data points. (b) Field expansion at small core size. .................................................. 30
3.4 (a) Finite difference beam propagation method simulation of mode evolution during transition. An 150 µm transition length is enough to preserve mode quality. 500 µm transition length is chosen in real devices to increase tolerance. (b) Optical microscope image of waveguide to lens fiber light coupling. 

3.5 SEM images of different waveguides. (a) Stepper photolithography waveguide. (b) Electron beam lithography waveguide. (c) Electron beam lithography waveguide with resist reflow. (d) overview of (c) at the 4µm taper region. 

3.6 Setup for waveguide loss measurement. TL, turnable laser. PC, polarization controller. LF, lens fiber. TS, 3D translation stage. RS, 3D rotation stage. Si, tested sample. CM, compound microscope. CCD, camera. PD, calibrated power detector. 

3.7 Theory and experiment of transmission curve. 

3.8 Setup for FWM-BS experiment. TL, tunable laser. PC, polarization controller. FC, fiber coupler. AM, amplitude modulator. PPG, pulse pattern generator. EDFA, erbium doped fiber amplifier. DWDMs, dense wavelength division multiplexers. FD, fiber delay line. LF, lens fiber. TS, translation stage. RS, rotation stage. CM, compound microscope. CCD, camera. Si, tested silicon sample. AA, adjustable attenuator. BD, balanced detector. ESA, electrical spectrum analyzer. OSA, optical spectrum analyzer. 

3.9 (a) OSA measurement of the fiber coupled output spectrum. (b) Split-step Fourier method simulation of similar condition. 

3.10 (a) Measurement of balanced detector. (b) Blow up of the selected area in (a). 

3.11 (a) Conversion efficiency vs. input bias power. The control is kept at 88 mW. (b) Conversion efficiency vs. input control power. The bias is kept at 775 mW. 

4.1 (a) Energy gap of GeSn alloy vs. Sn concentration. [3]. (b) Spontaneous emission spectrum for various GeSn alloys. [3] (c) Illustration of Optical pump. (d) Illustration of material structure. 

4.2 Test setup for Ge-Sn laser. Lp, laser pump. OF, optical fiber. SL1 spherical lens, \( f = 60 \text{ mm}, D = 2'' \). CL, cylindrical lens, \( f = 200 \text{ mm}, D = 2'' \). FM, flat mirror. SL2, spherical lens, \( f = 80\text{ mm}, D = 1'' \). TS, translation stage. CM, compound microscope. CCD, camera. FC, fiber collimator. LPF, long-pass filter. PD, photodetector. TIA, transimpedance amplifier. DAQ, data acquisition card.
4.3 (a) Laser efficiency measurement without pump contamination. (b) Laser efficiency measurement with pump contamination. (c) Pump power collected by optical fiber. (d) Microscope image of light coupling configuration.
LIST OF TABLES

3.1   Propagation loss measurement with Fabry-Perot method. . . . . . . . . . . . . 37
1.1 Introduction to Silicon Photonics

Silicon based electronics made its debut by Kilby’s demonstration of the first integrated circuit in 1958. Soon after, the processing capability of electronic processors has been on a linear increasing trajectory. As Gordon E. Moore, co-founder of Intel, observed, the number of components per integrated circuit doubles every two years, which is know as Moore’s law. This has been a beacon for the semiconductor industry for the past 4 and a half decades. In 2015, IBM demonstrated 7 nm node chip with silicon-germanium transistors [4]. This size is smaller than that of 30 silicon atoms. With the decreasing of transistor size and increasing of operational frequency, problems such as heat dissipation and electron leakage become significant and hard to solve. It is agreed that electronic devices will fall behind Moore’s law in the near future.

Apart from the computational power of each integrated processors, information processing speed is also limited by the information transmission speed. This problem became significant first for long distance communications. A 2.5 km long copper wire can transmit only 1.5 Mb of information per second. As a comparison, the Intel Core i7 980XE PC processor can perform 109 giga float-point operations per second (GFLOPS). Fortunately, the solution is found – the optical fibers. Different from electrons (fermions), photons (bosons) of disparate frequency can co-propagate without interference. Thus, one optical fiber can carry multiple pulse trains encoded in different optical
frequencies. The corresponding technology to combine and separated these pulse trains is called dense wavelength division multiplexing (DWDM). As a result, one single optical fiber, even if 100 times longer than copper wire, has 1000 times its information capacity. It is one of the fundamental hardware for today’s Internet system. Moreover, the role of optical fiber represented optical communication technology becomes bigger during recent years. Nowadays, it does not only appear in metropolitan area networks (MANs) but also in local area networks (LANs). Most recently, Google has initiated Google Fiber service that brings optical signal directly into end user devices. This allowed 1000 Mb/s Internet connection speed for a common family.

Inspired by the optics success in telecommunications, both academic and industrial researchers started to implement optics to solve intra silicon chip communication [5, 6, 7] and even computation problems[8]. This field is called silicon photonics. Conventionally, silicon is not considered as a high quality material for optics. It is known as an indirect band gap material which results in very low light emitting efficiency. It also lacks of second order nonlinearity which is the basis of commercial electro-optical (EO) modulators for telecommunications. However, silicon is meanwhile one of the most practical materials for integrated optics based on nowadays industrial infrastructure. Firstly, silicon is transparent for wavelength from 1.1 $\mu$m up to 7 $\mu$m [9], which is compatible with the signal used in optical telecommunications (1.3 $\mu$m to 1.6 $\mu$m). Secondly, silicon has large refractive index contrast to its oxide. Compared with single mode fiber’s 10 $\mu$m mode field diameter (@1.55 $\mu$m wavelength), silicon can confine the light passing through it within 300 nm. This is undoubtedly a big advantage for integration purpose. Also, even though without second order nonlinearity, silicon has two orders larger third order nonlinearity than optical fibers [10]. If used properly, this would allow direct information exchange between different communication channels. Finally, the manufacturing process established for silicon electronics can be easily migrated to silicon optics. Ensuring low cost for accurate and complicated fabrications. The silicon on insulator
(SOI) platform was developed to reduce parasitic device capacitance for microelectronics and is now a low cost standard material for integrated circuits. Incidentally, the buried oxide structure is also very suitable for confining light on the top silicon layer. In 2015, researchers in University of California Berkly, Massachusetts Institute of Technology and University of Colorado Boulder together demonstrated a electronic-optical hybrid microprocessor that is manufactured by a standard microelectronics foundry process [11].

A typical silicon photonics circuit consists of the following components: lasers/optical amplifiers, waveguides, modulators/switches, multiplexers/demultiplexers and detectors. In this thesis, we focus on lasers and switches.

An all silicon laser via Raman scattering was demonstrated by Intel in 2005 [12]. But its requirement of an optical pump limited its application. Nowadays, almost all of the commercial semiconductor lasers are based on III-V materials such as GaAs and InP. However, GaAs and InP have 4.1% and 8.1% of lattice mismatch compared with silicon respectively [9]. In order to integrate such lasers on a silicon chip, one of the bonding techniques is needed [13]. This, however, dramatically slows down the speed for large scale manufacturing. Recently, researchers have reported monolithic growth of GaAs quantum dot laser on silicon with a total of 2.5 µm buffer layers. But for integration purpose, a well defined positioning and a good optical mode is preferred. Another approach is through band structure engineering of Ge. Although Ge is also an indirect band gap material, the direct energy gap (Γ valley) is only 0.15 eV higher than the indirect energy gap (L valley). This difference can be further mitigated by applying tensile strain to the crystal. In 2012, a group at MIT demonstrated the first electrically pumped Ge-on-Si laser [14]. However, this laser is still fabricated from indirect band gap Ge. Doping was applied to increase direct gap electron density. Ge band engineering can also be achieved by incorporating Sn. Direct band gap GeSn-on-Si lasing has been observed under very low temperature in 2015 [15].
Optical switching is another crucial component for both optical interconnection and computation. In today’s telecommunication networks, ‘optical-to-electronic-to-optical’ (OEO) switching is commonly used. OEO provides the greatest flexibility. The pulse is regenerated during OEO thus reshaping, time jitter correction and wavelength switching is easy to implement. It also provides bit flow buffering which has not been satisfactorily achieved by all-optical switching techniques. However, OEO scheme is protocol dependent and requires high speed detectors on each node. It also needs to switch with every bit of data. This leads to a large amount of power dissipation that scales with the bit rate. All-optical switches, on the other hand, are bit rate insensitive. They switch per message rather than per bit thus is more energy economic. A survey conducted by several Internet companies and U.S. Department of Energy showed that all-optical networks could save up to 75% of the energy for large data centers [16]. Other than data centers, all-optical switches can also boost high performance computing (HPC) [17]. In short term, all-optical interconnection could increase the communication speed between different computation parts. In long term, all-optical switch based optical logic gates could help realize optical computers and even quantum computers. The study of all-optical switching has started as early as 1980s [18]. However, the first relatively low power silicon based switching was only demonstrated in 2004 by groups in Cornell University [19].

1.2 Introduction to Nonlinear Optics

Nonlinear optics studies the light matter interaction that is not proportional to input electric or magnetic field. Event though this nonlinear response is an inherent property of all matters, it is so weak that very high optical intensity is needed to make it observable. In fact, the first harmonic generation could not be demonstrated until the invention of laser [20]. Applications of nonlinear optics thrived since then. Today, even some laser pointers has nonlinear crystal inside it.

One of the most common applications of nonlinear optics is optical wavelength conversion
(OWC). It has been proven useful in today's communication networks. Implementations such as pulse regeneration [21, 22], signal demultiplexing [23] and all-optical gates [24] have been demonstrated. Furthermore, if OWC can be achieved without added noise, it is possible to realize quantum wavelength conversion (QWC) where the coherence of a photon is maintained after the photon is switched to another wavelength [25]. This would be very useful if quantum communication network is to be established where detecting then regenerating a quantum state is fundamentally impossible.

A common way to realize wavelength conversion is through four wave mixing (FWM). It is a third order nonlinear phenomenon that exists in both optical fibers and other silicon-based materials. In FWM, four different wavelengths (three if degenerate condition is satisfied) are involved. Conventionally, the strong laser(s) is called pump. The weak input laser(s) is called signal and the newly generated component(s) is called idler. Following the terminologies in [26], we categorize FWM
into 3 processes – modulation instability (MI), phase conjugation (PC) and Bragg scattering (BS) based on the location of pump(s) and signal in frequency domain (Figure 1.1). Both MI and PC processes involve energy flowing from pump to signal and idler, thus amplify vacuum fluctuations and introduce noise. BS, however, keeps the total energy of signal and idler constant hence is free from excessive noise [27, 28].

1.3 Motivation of Research

In this thesis, we experimentally demonstrate two devices for silicon photonic circuits – an low noise all-optical switch based on four wave mixing Bragg scattering (FWM-BS) and a silicon compatible laser based on germanium tin alloy.

One of the biggest concern for all-optical switch is the switching power. Even though silicon itself has large optical damage threshold, most of the EO modulators do not. Thus low power low noise switching is preferable. To achieve this, we use asymmetrically pumped FWM-BS, where one pump is kept sufficiently strong yet is not encoded with message and the other pump weak with encoded message. The existence of the strong pump is to ensure efficient nonlinear reaction. It can be a amplified pulse laser. The switching of the signal is solely controlled by the weak pump. An analogy to the electronic transistor can be made. As shown in Figure 1.2, the information free

Figure 1.2: An analogy between electronic transistor and FWM-BS all-optical switch.
strong pump is similar to the bias voltage in transistors. They provide the environment for the device to work properly. The weak pump is analogous to the control voltage who decides when the input goes to the output signal. Hence we will call the strong pump as bias and the weak pump as control in this thesis. Just like transistors have multiple working regions, the FWM-BS all-optical switch can also be configured to fulfill different tasks. If we detect at the signal wavelength, the device performs as an inverse switch, where the turned on control port converts energy from input signal to idler thus turns signal off. If the detection is performed at the idler wavelength, it acts as a direct switch where the output (idler) is turned on with the control. Due to the tensor nature of third order susceptibility, this switch can also work as a polarization switch where the polarization state of the idler is switched by the polarization of the control. With the availability of the passive on chip DWDM, this switch can also work as a router since different wavelengths can be easily separated to go to different destinations.

The challenge for silicon-compatible lasers arise from lattice matching. Excessive lattice mismatch would cause large internal strain and cause structure defects or even break down the crystal. Si, Ge and Sn are all group IV elements thus Ge and α-Sn have the same diamond lattice structure as Si. Even though lattice constant mismatch still exists between Ge, Sn and Si, GeSn alloy has been successfully grown onto silicon and demonstrated as a high efficiency on silicon detector [29]. In this work, we experimentally demonstrate the potential of lasing of GeSn alloy at room temperature without strain engineering. The GeSn active layer is deposited on silicon as thin film and can be fabricated into arbitrary shapes.
CHAPTER II

THEORY

2.1 Theory of Guided-wave Optics

2.1.1 Guided Modes

As metal wires are used to transmit electronic signals, proper optical waveguides can transmit optical signals with very low loss. When the waveguide size is comparable to or even smaller than the signal wavelength (1.3\(\mu\)m to 1.6\(\mu\)m in free space for telecommunication bands), the wave nature of photon results in that only certain distribution of optical field can propagate without diffraction. These specific distributions are called *modes*. For rectangular dielectric waveguides, as shown in Figure 2.1, the eigen modes can be solved from source free Maxwell’s wave equation

\[
\nabla^2 \vec{E} - \mu_0 \epsilon \frac{\partial^2 \vec{E}}{\partial t^2} = -\nabla (\frac{\vec{E} \cdot \nabla \epsilon}{\epsilon}) \tag{2.1}
\]

where \(\vec{E}\) is the electric field vector. \(\mu_0\) is the free space permeability because most materials used for waveguides are non magnetic. \(\epsilon\) is the scalar material permittivity. The scalar form is valid when the material does not possess birefringence or when the electric field is along one of the crystal axes. \(\epsilon\) varies in space for waveguide problems. It is also likely a function of frequency for many cases. To search for non diffracting modes, we assume the solutions have the form of

\[
\vec{E}(x, y, z) = \hat{E} E(x, y) e^{i\beta z} \tag{2.2}
\]
where \( \hat{i} (= \hat{x} \text{ or } \hat{y}) \) stands for the unit vector of \( x \) or \( y \) direction. \( \beta \) is the longitudinal propagation wavevector which is to be decided. \( E(x, y) \) is the transverse wave distribution that is also unknown for now. Substituting (2.2) into (2.1), we get

\[
\nabla_\perp^2 E + (n^2 k_0^2 - \beta^2) E = -\frac{\partial}{\partial \hat{i}} (\hat{i} \cdot \nabla n^2 E)
\]

(2.3)

Here, \( \nabla_\perp = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \). \( n^2 = \frac{\epsilon \epsilon_0}{\epsilon_0} \). \( k_0 \) is the wavevector in free space. One extra requirement is that the field must be localized on the transverse plane, i.e. \( 0 < \int_{-\infty}^{+\infty} E^*(x, y)E(x, y)dx dy < \infty \). This localization constrain leads to the discretization of wavevector \( \beta \). It also set up the requirement that core refractive index needs to be larger than that of cladding. For convenience, a dimensionless effective refractive index is used in place of \( \beta \). It is defined as

\[
n_{\text{eff}} = \frac{\beta}{k_0}
\]

(2.4)

Normally, \( n_{\text{eff}} \) has a value between the core and cladding index.

In this thesis, we address the solutions for \( \hat{i} = \hat{x} \) as TE modes and those for \( \hat{i} = \hat{y} \) as TM modes. For rectangular waveguides, approximated analytical solutions can be found following the method in chapter 5 of book [30]. For most other cases, analytical solutions are not available and numerical methods are needed.

Figure 2.1: (a) An example of rectangular waveguide. The core is composed of silicon (\( n \approx 3.5 \)) and the cladding is silicon dioxide (\( n \approx 1.5 \)). (b) The fundamental TE mode of the waveguide in (a). (c) The fundamental TM mode of the waveguide in (a)
2.1.2 Dispersion

In optics and quantum physics, the dispersion relation is the relation between frequency and wavevector or their equivalents such as angular frequency, wavelength, photon energy, photon momentum and refractive index. In waveguides, this wavevector is the forward propagation wavevector, namely $\beta$ in (2.2). Conventionally, we say dispersion exists when frequency is not proportional to wavevector. This is equivalent to say that different frequency propagate at disparate speed.

In waveguides, there are three major source of dispersion – material dispersion, waveguide dispersion and modal dispersion. Material dispersion refers to the non constant refractive index against frequency in the bulk material. It is independent of waveguide structure. Waveguide dispersion arises from the eigen-mode problem described by (2.3). (2.3) solves for $\beta$ for each angular frequency $\omega$. This usually leads to a non analytical equation and $\beta$ is normally not proportional to $\omega$. Modal dispersion is defined to account for the fact that, often, more than one solution (mode) can be found in (2.3) for one given frequency. It is also explained as a different propagation speed between modes. The combination of material dispersion and waveguide dispersion is referred to as chromatic dispersion. It can be characterized by different orders of Taylor expansions of longitudinal propagation constant $\beta$

$$\beta_i(\omega_0) = \frac{\partial^i \beta}{\partial \omega^i} \bigg|_{\omega=\omega_0} \quad (2.5)$$

$\beta_0$ is always 0. $\beta_1$ is the reciprocal of group velocity. $\beta_2$ is often referred to as the first dispersion term. It is called group velocity dispersion (GVD). It has an alternative form which is more often used in industry as

$$D(\lambda_0) = -\frac{\lambda}{c} \frac{\partial^2 n_{eff}}{\partial \lambda^2} \bigg|_{\lambda=\lambda_0} \quad (2.6)$$

where $c$ is the speed of light in vacuum. It worth mentioning that $\beta_2 = -\frac{\lambda^2}{2\pi c} D$. $D$ is usually expressed in the unit of ps/(nm · km). The total dispersion is called normal when $D$ is negative and anomalous when $D$ is positive.
Dispersions in waveguides are well studied because it is a main limitation of information transmission capacity. For this thesis, we will focus more on its influence on nonlinear effects (nonlinear phase matching). This will be introduced in section 2.2.2.

2.2 Four Wave Mixing And Nonlinear Optics

2.2.1 Nonlinear Susceptibilities

When electromagnetic (EM) waves propagate through matter, the electric field cause atoms or molecules to polarize, inducing dipole moments. Here the magnetic effect is not discussed because it is significantly smaller in majority of dielectric and semiconductor materials, including silicon. The internal electric field, then, is the superposition of the source field and the dipole field. However, it is not guaranteed that this dipole field strictly follow the oscillation of the source field. In fact, even though weak, this dipole oscillation always contains frequency that is the harmonics of the source frequency. These harmonic frequencies give rise to different orders of nonlinear effects.

Next, we use a semi-classical model to introduce different orders of susceptibilities which characterize the strength of nonlinear effects. For simplicity, we assume the material is a two level system, but this model can be applied to any amount of energy levels.

Suppose $\hat{H}_0$ is the steady state Hamiltonian and $u_0(\vec{r}), u_1(\vec{r})$ are the two eigen states corresponding to eigen energies $\mathcal{E}_0 = \hbar\omega_0, \mathcal{E}_1 = \hbar\omega_1$. The electric field has the form of $\vec{E} = \frac{1}{2} \sum_p \vec{E}_p e^{-i\omega_p t} + c.c$. The spatial variation of electric field is much longer than the electron orbits thus we assume $\vec{E}$ is only a function of time. The Schrödinger equation for the new system is

$$ (\hat{H}_0 - \hat{\mu} \cdot \vec{E})\psi = i\hbar \frac{\partial}{\partial t}\psi \quad (2.7) $$

where $\hat{\mu} = -e\vec{r}$ is the electric dipole moment operator. Applying perturbation method, we assume

$$ (\hat{H}_0 - \lambda\hat{\mu} \cdot \vec{E})\psi = i\hbar \frac{\partial}{\partial t}\psi \quad (2.8) $$
\[
\psi(\vec{r}, t) = \sum_{N=0}^{+\infty} \chi^N \psi^{(N)}(\vec{r}, t) \\
= \sum_{N=0}^{+\infty} \sum_{m=0,1} \chi^N a_m^{(N)}(t) u_m(\vec{r}) e^{-i\omega_m t}
\]  
(2.9)

Comparing orders larger than 0\textsuperscript{th} \((N \geq 1)\) yields

\[
i\hbar \dot{a}_n^{(N)} = - \sum_{m=0,1} a_m^{(N-1)} <u_n|\hat{\mu} \cdot \vec{E}|u_m> e^{i(\omega_n - \omega_m)t}
\]  
(2.11)

where \(n = 0\) or 1. We assume the system stays in ground state initially, meaning \(a_0^{(0)} = 1, a_1^{(0)} = 0\).

Then

\[
a_{n}^{(1)}(t) = \sum_p \frac{1}{2\hbar} <u_n|\hat{\mu}|u_0> \vec{E}_p \left[e^{i(\omega_n - \omega_0 + \omega_p)t} + e^{i(\omega_n - \omega_0 - \omega_p)t} \right], n = 1, 2
\]  
(2.12)

where \(\omega_p\) are the frequencies contained in the input electric field. We denote \(<u_n|\hat{\mu}|u_m>\) as \(\vec{\mu}_{n,m}\), \(\omega_1 - \omega_0\) as \(\Delta \omega\).

Macroscopically, we define \(\vec{P}^{(1)} = N <\vec{p}^{(1)} >\), where \(N\) is the number density of atoms. We would like to find first order susceptibility tensor \(\bar{\chi}\) having the form \(\vec{P}^{(1)} = \epsilon_0 \bar{\chi} \vec{E}\). If we write \(\vec{\mu}_{n,m}\) as \((\vec{\mu}_{n,m}^1, \vec{\mu}_{n,m}^2, \vec{\mu}_{n,m}^3)\), Then

\[
\bar{\chi}_{ij}^{(1)}(\omega_p) = \frac{N}{\epsilon_0 \hbar} \left( \frac{\mu_{1,0} \mu_{0,1}^*}{\Delta \omega - \omega_p} + \frac{\mu_{0,1} \mu_{1,0}^*}{\Delta \omega + \omega_p} \right)
\]  
(2.15)

It worth mentioning that (2.15) quantitatively explains the appearance of material dispersion introduced in section 2.1.2.

We now look at the 2nd order susceptibility. (2.11) shows that the oscillation in \(a_n^{(1)}\) is carried to \(a_n^{(2)}\), superposing with the oscillation of \(\vec{E}\). \(a_n^{(2)}\) will also be a summation of oscillating terms.

Here we only look at terms related with \(\omega_p + \omega_q\). The whole solution can be found by sum \(p\) and \(q\).
We define the second order susceptibility \( \chi^{(2)} \). Here, we only present the result.

\[
a^{(2)}_n(t) = \frac{1}{4\hbar^2} \sum_{m=0,1} \left[ \frac{\langle \vec{\mu}_{m,0} \cdot \vec{E}_P \rangle (\vec{\mu}_{m,m} \cdot \vec{E}_q)}{(\omega_m - \omega_0 + \omega_p)(\omega_n - \omega_0 + \omega_p + \omega_q)} e^{i(\omega_n - \omega_0 + \omega_p + \omega_q)t} 
+ \frac{\langle \vec{\mu}_{m,0} \cdot \vec{E}_P \rangle (\vec{\mu}_{m,m} \cdot \vec{E}_q)}{(\omega_m - \omega_0 - \omega_p)(\omega_n - \omega_0 - \omega_p - \omega_q)} e^{i(\omega_n - \omega_0 - \omega_p - \omega_q)t} \right], n = 1, 2
\]  

(2.16)

Notice that when \( p \neq q \), switching \( p \) and \( q \) gives us another term which contributes to the same frequency. Similar to 1\(^{st}\) order, we calculate \( < \hat{\rho}^{(2)} > \) as

\[
< \hat{\rho}^{(2)}(\omega_p + \omega_q; \omega_p, \omega_q) >= < |\psi|^2 > + < |\psi|^1 |\psi|^1 > + < |\psi|^2 |\psi|^0 >
\]

(2.17)

We define the second order susceptibility \( \chi^{(2)}_{ijk} \) satisfying the condition

\[
P^{(2)}_{i}(\omega_p + \omega_q) = N < p_i(\omega_p + \omega_q) >= \epsilon_0 \chi^{(2)}_{ijk}(\omega_p + \omega_q; \omega_p, \omega_q) E_j(\omega_p) E_k(\omega_q)
\]

(2.18)

where Einstein’s summation over \( j, k \) is assumed. The detailed derivation can be found in [31]. Here, we only present the result.

\[
\chi^{(2)}_{ijk}(\omega_p + \omega_q; \omega_p, \omega_q) = \frac{N}{\epsilon \hbar^2} P_I \sum_{m,n=0,1} \left[ \frac{\mu_{i0n}^i \mu_{nm}^i \mu_{m0}^k}{(\omega_n - \omega_p - \omega_q)(\omega_m - \omega_p)} + \frac{\mu_{i0n}^j \mu_{nm}^j \mu_{m0}^k}{(\omega_n + \omega_p)(\omega_m - \omega_p)} \right]
\]

(2.19)

where \( P_I \) is the intrinsic permutation operator. As we just explained, \( \chi^{(2)}_{ijk}(\omega_p + \omega_q; \omega_p, \omega_q) \) and \( \chi^{(2)}_{ijk}(\omega_p + \omega_q; \omega_q, \omega_p) \) contribute to the same term when \( \omega_p \neq \omega_q \). It is not practical to separate the contribution of the two terms. Thus we use operation \( P_I \) to indicate the average of all possible permutations of \( \omega_p \) and \( \omega_q \).

Equation (2.19) shows that if the material is unpolarized, namely \( \vec{\mu}_{nn} = \vec{0}, n = 0, 1 \), \( \chi^{(2)} \) vanishes. As a consequence, silicon crystal, formed by the same atoms, does not show second order nonlinear effects.

The same procedure can be carried on to find all orders of susceptibilities. However, each order gets more complicated. In fact, even for a two level system, the full expansion of \( \chi^{(3)}_{kijh} \) contains
192 terms, as can be found in [31]. For cases where the interested frequencies are far from atomic resonance or semiconductor band gap, it is common to only use a constant $\chi^{(3)}_{kjih}$. Practically, third order nonlinearity is measured in the form of nonlinear refractive index $n_2$. It arises from the fact that

$$D_i(\omega) = \epsilon_0[\epsilon_i E_i(\omega) + \chi^{(3)}_{iiii} E_i(\omega) E_i(-\omega) E_i(\omega)]$$  \hspace{1cm} (2.20)

where the electric field is assumed to be aligned with axis $i = x, y, \text{or} z$. $D_i$ is the displacement vector. If the conventional form $D = \epsilon_0 n_2 E$ is to still be used, we need to modify $n$ as

$$n = n_0 + n_2 I$$  \hspace{1cm} (2.21)

where $n_0$ is the linear refractive index. $I$ is the irradiance of the light. A link between $n_2$ and $\chi^{(3)}_{iiii}$ can be established as

$$\chi^{(3)}_{iiii} = \frac{4}{3} n_0^2 \epsilon_0 n_2$$  \hspace{1cm} (2.22)

Also, if the material is silicon, due to the rotational symmetries of its lattice structure, all $\chi^{(3)}_{iiii}, i = x, y, z$ are equal [32].

### 2.2.2 Nonlinear Maxwell’s Wave Equation

The previous section explained how new frequencies are created through light matter interaction. In this section, we talk about how to accumulate these new frequencies radiations. For most applications, the nonlinear material does not contain only one atom. To accumulate new frequencies efficiently, as the wave propagates, the later generated wave is best in phase with the previously generated one. This imposes a constraint to the speed-frequency relation, i.e. dispersion relation. Following, we will focus on $\chi^{(3)}$ phenomena and quantitatively show the dispersion requirements.

As we see from the previous section, the nonlinearity is introduced by the polarization term $\vec{P}$. We start from source free Maxwell’s wave equation

$$\nabla^2 \vec{E} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \vec{P}}{\partial t^2}$$  \hspace{1cm} (2.23)
Instead of writing $\vec{P}$ as $\epsilon_0 \chi \vec{E}$ as in linear problems, we write it as $\epsilon_0 (\chi^{(1)}(\omega) E + \chi^{(3)} |E|^2 E)$. Notice that we also assumed $\chi^{(1)}, \chi^{(3)}, E$ being scalar and $\chi^{(3)}$ being constant. The scalar form is valid when the material is isotropic or all the electric fields lie on the crystal axes. Generally, the exact value of $\chi^{(3)}$ also depend on the polarization direction of the involved fields. For simplicity, we assume all fields are co-aligned. It can also be justified that the wavelength dependence of $\chi^{(3)}$ only play a minor part in the final results thus we assume it constant for simplicity. The above equation becomes

$$\nabla^2 E - \mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2} = \mu_0 \epsilon_0 \frac{\partial^2}{\partial t^2} (\chi^{(1)}(\omega) E + \chi^{(3)} |E|^2 E) \tag{2.24}$$

For FWM-BS, we consider four frequency components $\omega_1 + \omega_i = \omega_2 + \omega_s$. If we see (2.23) as the summation of several temporal oscillating terms being 0, we can say that each oscillating frequency has a total amplitude of 0. Thus we can write (2.23) into four separate equations for the four interested frequencies as

$$\nabla^2 E_1 - \mu_0 \epsilon_0 \frac{\partial^2 E_1}{\partial t^2} = \frac{3}{4} \mu_0 \epsilon_0 \chi^{(3)} \frac{\partial^2}{\partial t^2} (|E_1|^2 E_1 + 2 |E_2|^2 E_1 + 2 |E_s|^2 E_1 + 2 |E_i|^2 E_1 + 2 E_2 E_s E_i^*) \tag{2.25a}$$

$$\nabla^2 E_2 - \mu_0 \epsilon_0 \frac{\partial^2 E_2}{\partial t^2} = \frac{3}{4} \mu_0 \epsilon_0 \chi^{(3)} \frac{\partial^2}{\partial t^2} (2 |E_1|^2 E_2 + |E_2|^2 E_2 + 2 |E_s|^2 E_2 + 2 |E_i|^2 E_2 + 2 E_1 E_s^* E_i) \tag{2.25b}$$

$$\nabla^2 E_s - \mu_0 \epsilon_0 \frac{\partial^2 E_s}{\partial t^2} = \frac{3}{4} \mu_0 \epsilon_0 \chi^{(3)} \frac{\partial^2}{\partial t^2} (2 |E_1|^2 E_s + 2 |E_2|^2 E_s + |E_s|^2 E_s + 2 |E_i|^2 E_s + 2 E_1 E_2^* E_i) \tag{2.25c}$$

$$\nabla^2 E_i - \mu_0 \epsilon_0 \frac{\partial^2 E_i}{\partial t^2} = \frac{3}{4} \mu_0 \epsilon_0 \chi^{(3)} \frac{\partial^2}{\partial t^2} (2 |E_1|^2 E_i + 2 |E_2|^2 E_i + 2 |E_s|^2 E_i + |E_i|^2 E_i + 2 E_1^* E_2 E_s) \tag{2.25d}$$

To describe this process occurring in waveguides, we assume the solution has the form of $E_n(x, y, z) = F(x, y) A_n(z) \exp[i(\beta_n z - \omega_n t)], n = 1, 2, s, i$. Notice that we assume transverse field of guided modes
for different frequency to be the same. Then (2.25) can be simplified into

$$\frac{dA_1}{dz} = i\gamma_1(|A_1|^2 + 2|A_2|^2 + 2|A_s|^2)A_1 + i2\gamma A_2 A_s A^*_s e^{-i\Delta \beta z}$$ (2.26a)

$$\frac{dA_2}{dz} = i\gamma_2(2|A_1|^2 + |A_2|^2 + 2|A_s|^2)A_2 + i2\gamma A_1 A^*_s A_s e^{i\Delta \beta z}$$ (2.26b)

$$\frac{dA_s}{dz} = i\gamma_s(2|A_1|^2 + 2|A_2|^2 + |A_s|^2 + 2|A_i|^2)A_s + i2\gamma A_1 A^*_s A_s e^{i\Delta \beta z}$$ (2.26c)

$$\frac{dA_i}{dz} = i\gamma_i(2|A_1|^2 + 2|A_2|^2 + 2|A_s|^2 + |A_i|^2)A_i + i2\gamma A^*_1 A_2 A_s e^{-i\Delta \beta z}$$ (2.26d)

where,

$$\gamma_n = \frac{n_2\omega_n}{cA_{eff}}$$ (2.26e)

$$A_{eff} = \frac{\left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(x,y)|^2 dxdy\right)^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(x,y)|^4 dxdy}$$ (2.26f)

$$\Delta \beta = \beta_1 + \beta_i - \beta_2 - \beta_s$$ (2.26g)

$A_{eff}$ is known as the effective area. $A_n$ is chosen so that $|A_n|^2$ stands for total power of the mode.

For the SOI waveguides that this work focuses on, the nonlinearity of cladding (SiO$_2$) is negligible compared with core (Si), thus the integration in the denominator is only over the core area. $n_2$ is the nonlinear refractive index and $\gamma_n$ the nonlinear coefficient. For our purpose, the $\omega_n$ are close thus we replace $\gamma_n$ with a single $\gamma$. The terms in the first parentheses on the right hand side of the equations corresponds to self and cross phase modulation and the second terms corresponds to frequency mixing. (2.26) can be solved analytically, as shown in [33]. Here, we show that if we apply the asymmetric pump condition, i.e. one pump is nearly undepleted and the other is comparable with the signal, a simpler solution can be yielded.
If we assume $|A_1(0)|^2$ is much larger than any other fields, (2.26) can be re-expressed as

\[
\frac{dA_1}{dz} = i\gamma |A_1|^2 A_1 \tag{2.27a}
\]
\[
\frac{dA_2}{dz} = i2\gamma |A_1|^2 A_2 + i2\gamma A_1 A_2^* A_1 e^{i\Delta \beta z} \tag{2.27b}
\]
\[
\frac{dA_s}{dz} = i2\gamma |A_1|^2 A_s + i2\gamma A_1 A_2^* A_1 e^{i\Delta \beta z} \tag{2.27c}
\]
\[
\frac{dA_i}{dz} = i2\gamma |A_1|^2 A_i + i2\gamma A_1 A_2^* A_s e^{-i\Delta \beta z} \tag{2.27d}
\]

The justification goes as follows. First, the summation of $|A_n|^2$ collapse to $|A_1|^2$ under the assumption that $|A_1|^2$ is significantly stronger than others. Second, each complex equation is essentially two equations, one of the real part and another of the imaginary part. The $e^{-i\Delta \beta z}$ dependent terms contribute to a different part than the $|A_n|^2$ dependent terms thus they should not be neglected. Finally, $|A_1|$ is large so that its amplitude change is insignificant during frequency mixing process based on energy conservation. Thus we ignore the $e^{-i\Delta \beta z}$ dependent term in (2.27a). The solution of (2.27a) can be expressed as

\[
A_1 = \sqrt{P_1} e^{i\gamma P_1 z} \tag{2.28}
\]

where $P_1 = |A_1(z)|^2 = |A_1(0)|^2$. Then we assume $A_n(z) = \xi_n(z) e^{i\phi_n(z)}$, $n = 2, s, i$, where both $\xi_n(z)$ and $\phi_n(z)$ are real. $\xi_n(z)$ corresponds to amplitude change and $\phi_n(z)$ corresponds to initial and nonlinear phase. By Substituting them into (2.27b) – (2.27d) and comparing the real and imaginary parts, we have

\[
\frac{d\xi_2}{dz} = -2\gamma \sqrt{P_1} \xi_2 \xi_i \sin \theta \tag{2.29a}
\]
\[
\frac{d\xi_s}{dz} = -2\gamma \sqrt{P_1} \xi_2 \xi_i \sin \theta \tag{2.29b}
\]
\[
\frac{d\xi_i}{dz} = 2\gamma \sqrt{P_1} \xi_2 \xi_s \sin \theta \tag{2.29c}
\]
\[
\frac{d\theta}{dz} = \Delta \beta - \gamma P_1 + \frac{\cos \theta}{\sin \theta} \left( \frac{1}{\xi_2} \frac{d\xi_2}{dz} + \frac{1}{\xi_s} \frac{d\xi_s}{dz} + \frac{1}{\xi_i} \frac{d\xi_i}{dz} \right) \tag{2.29d}
\]
where \( \theta = \gamma P_1 z - \phi_2 - \phi_s + \phi_i + \Delta \beta z \). From (2.29) we have

\[
\xi_2^2 + \xi_i^2 = C_1 \tag{2.30a}
\]
\[
\xi_3^2 + \xi_s^2 = C_2 \tag{2.30b}
\]
\[
(\xi_2 \xi_s \cos \theta + q \xi_i) \xi_i = C_3 \tag{2.30c}
\]

where \( q = \frac{\gamma P_1 - \Delta \beta}{4 \gamma \sqrt{P_1}} \). \( C_1, C_2, C_3 \) are invariants. (2.30a) comes from \( \xi_2 \cdot (2.29a) + \xi_i \cdot (2.29c) \) and (2.30b) comes from \( \xi_s \cdot (2.29b) + \xi_i \cdot (2.29c) \). (2.30) requires more explanation. We first multiply (2.29d) with \( \sin(\theta) \xi_2 \xi_s \xi_i \). rearrange it so it becomes

\[
\frac{d(\xi_2 \xi_s \xi_i)}{dz} \cos(\theta) - \xi_2 \xi_s \xi_i \sin(\theta) \frac{d\theta}{dz} + (\Delta \beta - \gamma P_1) \xi_2 \xi_s \xi_i \sin(\theta) = 0 \tag{2.31}
\]

The first two terms can be combined into one derivation. We then substitute (2.29c) into the third term.

\[
\frac{d(\xi_2 \xi_s \xi_i \cos(\theta))}{dz} + \frac{\Delta \beta - \gamma P_1}{2 \gamma \sqrt{P_1}} \xi_i \frac{d\xi_i}{dz} = 0 \tag{2.32}
\]

Subsequently,

\[
\frac{d}{dz} (\xi_2 \xi_s \xi_i \cos(\theta) + \frac{\Delta \beta - \gamma P_1}{4 \gamma \sqrt{P_1}} \xi_i^2) = 0 \tag{2.33}
\]

This is equivalent to (2.30c)

Let \( \rho_2 = |A_2(0)|^2, \rho_s = |A_4(0)|^2, A_i(0) = 0 \). Thus \( C_1 = \rho_2, C_2 = \rho_s, C_3 = 0 \). Since \( \xi_i \) can not be 0 for the whole process, \( C_3 = 0 \) implies

\[
\xi_2 \xi_s \cos \theta = -q \xi_i \tag{2.34}
\]

As a consequence, we can quickly estimate the possible conversion efficiency as

\[
\frac{P_i(z)}{P_s(z) + P_i(z)} \leq \frac{16 P_2(z) P_1}{(P_1 - \frac{\Delta \beta}{\gamma})^2 + 16 P_2(z) P_1} \tag{2.35}
\]

To actually solve for \( \xi_i \), we substitute (2.30) into (2.29c).

\[
\frac{d z}{d \xi_i} = \pm \frac{1}{2 \gamma \sqrt{P_1}} \frac{1}{\sqrt{\xi_i^4 - (\rho_2 + \rho_s + q^2) \xi_i^2 + \rho_2 \rho_s}} \tag{2.36}
\]

\[\text{Page} 18\]
With some substitution, we can write this into standard elliptical integration

\[
\frac{dt}{du} = \pm \frac{1}{\sqrt{(1 - u^2)(1 - k^2u^2)}}
\]  \hspace{1cm} (2.37)

where,

\[
u = \frac{\xi_i}{\sqrt{r_1}}
\]  \hspace{1cm} (2.38a)

\[
r_1 = \frac{(\rho_2 + \rho_s + q^2) - \sqrt{(\rho_2 + \rho_s + q^2)^2 - 4\rho_2\rho_s}}{2}
\]  \hspace{1cm} (2.38b)

\[
r_2 = \frac{(\rho_2 + \rho_s + q^2) + \sqrt{(\rho_2 + \rho_s + q^2)^2 - 4\rho_2\rho_s}}{2}
\]  \hspace{1cm} (2.38c)

\[
t = 2\gamma \sqrt{P_1r_2z}
\]  \hspace{1cm} (2.38d)

\[
k = \sqrt{\frac{r_1}{r_2}}
\]  \hspace{1cm} (2.38e)

The solution for (2.37) is Jacobi elliptic sn function

\[
u = \pm \text{sn}(t, k)
\]  \hspace{1cm} (2.39)

Finally the power flow can be expressed as

\[
P_1(z) = r_1\text{sn}^2(2\gamma \sqrt{P_1r_2z}, k)
\]  \hspace{1cm} (2.40)

where

\[
r_1, r_2 = \frac{(P_2(0) + P_s(0) + q^2) \pm \sqrt{(P_2(0) + P_s(0) + q^2)^2 - 4P_2(0)P_s(0)}}{2}
\]  \hspace{1cm} (2.41a)

\[
k = \sqrt{\frac{r_1}{r_2}}
\]  \hspace{1cm} (2.41b)

\[
q = \frac{\gamma P_1 - \Delta \beta}{4\gamma \sqrt{P_1}}
\]  \hspace{1cm} (2.41c)

\[
\Delta \beta = \beta_1 + \beta_i - \beta_2 - \beta_s
\]  \hspace{1cm} (2.41d)

To understand the final expression, we first briefly introduce the Jacobi elliptic functions. As
Figure 2.2: (a) Conversion efficiency vs. propagation distance. Parameters are \( P_1(0) = 1 \text{W}, P_2(0) = 0.05 \text{W}, P_3(0) = 0.03 \text{W}, \gamma = 125 \text{W}^{-1} \text{m}^{-1}, \Delta \beta = 30 \text{m}^{-1} \). Curve (a) is analytical solution from (2.40). (b) is numerical solution for (2.26) by 4th order Runge-Kutta method. (c) is conversion efficiency upper bound estimation by letting \( P_2(z) = P_2(0) \) in (2.35). (b) Unit ellipse for the definition of elliptical function.

Shown in Figure 2.2(b), we first normalize the short axis of an ellipse to 1. We use \( k \) to represent eccentricity as

\[
k = \sqrt{1 - \frac{1}{a^2}} \quad \text{(2.42)}
\]

where \( a \geq 1 \). A counterpart of radius in trigonometry is defined as

\[
u = \int_0^\theta r \, d\theta \quad \text{(2.43)}
\]

Notice that \( r \) is a dimensionless value since the ellipse is normalized. Then several Jacobi elliptic functions can be defined.

\[
\begin{align*}
sn(u, k) &= y_0 \quad \text{(2.44)} \\
cn(u, k) &= \frac{x_0}{a} \quad \text{(2.45)} \\
dn(u, k) &= \frac{r}{a} \quad \text{(2.46)}
\end{align*}
\]

We see that \( sn(u, k) \) is a periodic function with period depending on \( k \). Its value vary from -1 to 1. It reduces to sine function when \( k = 0 \).
Now we look back at (2.40). The highest conversion efficiency depends only on $r_1$. The maximum value of $r_1$ is achieved when $q = 0$, leading to the phase matching condition that

\[ \Delta \beta = \beta_1 + \beta_i - \beta_2 - \beta_s = \gamma P_1 \]  

(2.47)

When one strong pump exists, FWM-MI will happen. The analysis is similar to FWM-BS but is in fact simpler. The coupled amplitude equations can be written as

\[ \frac{dA_p}{dz} = i\gamma \left[ (|A_p|^2 + 2|A_s|^2 + 2|A_i|^2)A_p + 2A_p^*A_sA_i e^{-i\Delta \beta z} \right] \]

\[ \frac{dA_s}{dz} = i\gamma \left[ (2|A_p|^2 + |A_s|^2 + 2|A_i|^2)A_s + A_p^*A_sA_i e^{i\Delta \beta z} \right] \]

\[ \frac{dA_i}{dz} = i\gamma \left[ (2|A_p|^2 + 2|A_s|^2 + |A_i|^2)A_i + A_p^*A_sA_i e^{i\Delta \beta z} \right] \]

(2.48)

where $\Delta \beta = 2\beta_p - \beta_s - \beta_i$. Assuming $A_p$ is undepleted, the solution can be found as [34]

\[ P_i = \frac{\gamma^2 P_p^2}{g^2} P_{s0} \sinh^2(gz) \]  

(2.49)

where

\[ g = \sqrt{\gamma^2 P_p^2 - \left( \frac{\kappa}{2} \right)^2} \]  

(2.50)

\[ \kappa = \Delta \beta - 2\gamma P_p \]  

(2.51)

$P_{s0}$ is the initial power of signal. It can be concluded that $P_i$ grows exponentially when $\Delta \beta < 0$ and is sinusoidal when $\Delta \beta > 0$. Since $2\omega_p = \omega_s + \omega_i$, we can expand $\Delta \beta$ in terms of $\omega$ as

\[ \Delta \beta \approx -\frac{\beta_2}{2} \left[ (\omega_p - \omega_i)^2 + (\omega_s - \omega_p)^2 \right] \]  

(2.52)

As a comparison, same argument can be applied to FWM-BS phase matching condition

\[ \Delta \beta_{BS} \approx \beta_2 \left[ \frac{1}{2} (\omega_1 + \omega_2) (\omega_1 - \omega_2) (\omega_1 - \omega_s) \right] \]  

(2.53)

where $\beta_2$ is defined as (2.5). The sign of the phase mismatch in FWM-MI is solely defined by the sign of dispersion while such sign of FWM-BS can be adjusted by frequency configuration.
the $\beta_2$ value is positive, correspondingly $D$ (defined in (2.6)) is negative, FWM-MI process can be significantly suppressed. FWM-BS can still have high efficiency as long as the strong pump is the highest or lowest frequency.

Due to the complexity of the nonlinear interaction, an analytical solution is not available for most cases. The split-step Fourier method (SSFM) is often used to find numerical solutions for various nonlinear problems. At the end of this part, we introduce a simple 1D SSFM which considers dispersion and third order nonlinearity simultaneously. Rather than looking at each wavelength, SSFM applies slowly varying envelope approximation on (2.24) directly. The starting equation of SSFM is

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2} A - \frac{i\beta_2}{2} \frac{\partial^2 A}{\partial T^2} + i\gamma|A|^2 A - \frac{\beta_T}{2A_{\text{eff}}}|A|^2 A$$

(2.54)

where $A$ is the total wave packet. $|A|^2$ corresponds to power. $T = t - \beta_1 z$ is the equivalent time in moving frame. $\beta_T$ is the two photon absorption(TPA) coefficient. $A_{\text{eff}}$ again is the effective area. Higher order dispersions, Raman scattering and self steepening can also be implemented in the equation. (2.54) has the symbolic solution of

$$A(z,T) = e^{z(\hat{L} + \hat{N})} A(z_0, T)$$

(2.55)

where

$$\hat{L} = -\frac{\alpha}{2} - \frac{i\beta_2}{2} \frac{\partial^2}{\partial T^2}$$

(2.56)

$$\hat{N} = i\gamma|A|^2 - \frac{\beta_T}{2A_{\text{eff}}}|A|^2 A$$

(2.57)

As the name implies, $\hat{L}$ is the operator for linear processes and $\hat{N}$ is the operator for nonlinear processes. We assume for a small step size $\Delta z$, The commutator of $\hat{L}$ and $\hat{N}$ is insignificant. Thus

$$A(z_0 + \Delta z) \approx e^{\Delta z \hat{L}} e^{\Delta z \hat{N}} A(z_0, T)$$

(2.58)
$e^{\Delta z \hat{N}}$ can be implemented directly in time domain by assuming $i\gamma |A|^2 = i\gamma |A(z_0, T)|^2$. $e^{\Delta z \hat{L}}$ is better implemented in frequency domain, with the form $e^{-\frac{\alpha}{2} + \frac{i\beta}{2} \omega^2}$. The full solution can be found by repeating (2.58).

2.3 Semiconductor Lasers

Optical gain relies on electron population inversion and stimulated emission. However, the population inversion behavior in semiconductor lasers is very distinct from that in atomic energy level based lasers (gas lasers and ion doped lasers). In semiconductors, the excited electrons will immediately relax to one of the lowest available energy levels inside the conduction band, following Fermi-Dirac distribution. Two criteria need to be satisfied in order to let the electron relax back to valence band through radiative recombination. Energy conservation

$$E_e - E_h = h\nu_p$$  (2.59)

and momentum conservation

$$\hbar \vec{k}_e - \hbar \vec{k}_p = \hbar \vec{k}_h$$  (2.60)

where $E_e$ and $\hbar \vec{k}_e$ are the energy and momentum of the electron respectively. $E_h$ and $\hbar \vec{k}_h$ are the energy and momentum of an unoccupied state (called a hole) in the valence band. $h\nu_p$ and $\hbar \vec{k}_p$ are energy and momentum of the radiated photon. The momentum of a photon is generally too small to compensate for any momentum mismatch of electron states. So radiative recombination (RR) would require the electron and the hole states to correspond to almost the same momentum. The existence of phonons (lattice vibration) can also possibly compensate for momentum mismatch, but usually with significantly lower recombination rate. Unfortunately, RR is not the only recombination process that happens in semiconductors. Other non-radiative recombination (NRR) processes such as Shockley-Reed-Hall recombination happens simultaneously. A low RR rate will cause the majority part of the excited electrons relax back to valence band through NRR processes. The internal
quantum efficiency of radiation is defined as

\[ \eta_i = \frac{\tau_n}{\tau_r + \tau_n} \]  

(2.61)

where \( \tau_r \) is the radiative lifetime and \( \tau_n \) is the non radiative lifetime. The optical gain for photon at frequency \( \nu \) can be expressed as [35]

\[ \gamma_0(\nu) = \frac{\eta_i \lambda^2}{8\pi\tau_r} \rho(\nu) f_g(\nu) \]  

(2.62)

where

\[ \rho(\nu) = \frac{h m_r}{m_e} \rho_c(E_2) = \frac{(2m_r)^{3/2}}{\pi h^2} \sqrt{\hbar \nu - E_g} \]  

(2.63)

\[ f_g(\nu) = e^{-\frac{E_2-E_1}{k_B T}} + e^{-\frac{E_1-E_1}{k_B T}} - 1 \]  

(2.64)

\( \frac{1}{m_r} = \frac{1}{m_e} + \frac{1}{m_h} \) is the effective mass. \( E_2 \) and \( E_1 \) are the corresponding electron energy states that emit the photon. \( h \nu = E_2 - E_1 \). \( E_g \) is the band gap and \( k_B \) is the Boltzmann constant. \( \rho(\nu) \) is the optical joint density of states describing the number of matched electron-hole state pairs and \( f_g(\nu) \) is the Fermi inversion factor describing the probability of photon emission minus the probability of photon absorption. For the material in Figure 2.3(a), if the emission is from point \( a \) to point \( c \), the electron occupation probability is close to 0 so that gain is close to 0 or even negative if \( c \) is not fully occupied by holes. Similarly, if the emission is from \( b \) to \( d \), the gain is smaller than 0 due to lack of holes. If the emission is to happen from \( b \) to \( c \), \( \eta_i \) would be small and \( \tau_r \) would be large as phonon is required. Gain is possible but would be too small to overcome any external loss. We call the energy difference between \( a \) and \( c \) direct energy gap (DEG) to distinguish from band gap (BG) which is defined by the energy gap between lowest conduction band energy and highest valence band energy. We also call indirect energy gap (IEG) as the smallest gap other than DEG. It is the gap between \( b \) and \( c \) in 2.3(a). The material is direct band gap if \( \text{DEG} = \text{BG} \). It is indirect band gap if \( \text{IEG} = \text{BG} \). We call the material quasi direct band gap when \( \text{DEG} \) is not very far from \( \text{BG} \). In such case, it is possible to have excited electrons occupy the direct gap states. This probability
could be further enhanced by doping with electron donors through which the bottom of conduction band would be preoccupied by doped electrons and more excited electrons would fall on the direct gap states.

![Figure 2.3:](a) Illustration of an excited indirect band gap material. (b) Excited quasi direct band gap material with n type doping.)

The emission efficiency curve is one of the common ways to distinguish stimulated emission (lasers) from spontaneous emission (LEDs). For spontaneous emission based devices, the emission rate is proportional to excited electrons which is generally proportional to pump power. Lasers, however, has piecewise defined emission rates. When the gain is lower than the cavity loss, it follows spontaneous emission with possible amplified stimulated emission (ASE). When the gain exceeds the cavity loss, feedback gain is established and the output power grows in a dramatically faster rate. To quantitatively describe this, we make two plausible assumptions. First, the material gain is proportional to pump power. Second, the saturation power is independent of the pump power. Assumption two is based on the fact that the saturation intensity depends only on gain spectrum and electron lifetime. We will also ignore the interference between forward and backward traveling waves since they happen in a considerably small scope. The laser cavity is assumed to be a Fabry-Perot type.

For high reflection mirror cavity lasers, the intracavity field is nearly constant. The stability
condition is
\[
\frac{kP_p}{1 + \frac{P_i}{P_s}} = \alpha
\]  
(2.65)

where \(kP_p\) is the small signal gain under pump power \(P_p\). \(P_i\) is the intracavity power. \(P_s\) is the saturation power and \(\alpha\) is the cavity loss. We also define threshold power as
\[
P_{th} = \frac{\alpha}{k}
\]  
(2.66)

It is the power at which small signal gain overcomes cavity loss. Notice that only half of the intracavity power is contributed by forward traveling wave, so the emission efficiency can be written as
\[
P_{out} = \frac{P_i}{2} = TP_{sat} \frac{P_p}{P_{th}} - 1
\]  
(2.67)

where \(T\) is the transmission coefficient for one facet.

If mirror loss is significant and propagation loss is small, we have to consider the power variance inside the cavity. If we assume \(P_+(z)\) and \(P_-(z)\) are the power of the forward and backward traveling wave respectively, the stability condition can be expressed as
\[
\gamma(z) = \frac{kP_p}{1 + \frac{P_+(z) + P_-(z)}{P_s}} = \frac{1}{P_+(z)} \frac{dP_+(z)}{dz} = -\frac{1}{P_-(z)} \frac{dP_-(z)}{dz}
\]  
(2.68)

with the boundary conditions
\[
R_1 P_-(0) = P_+(0)
\]  
(2.69)

\[
R_2 P_+(l) = P_-(l)
\]  
(2.70)

where \(R_1\) and \(R_2\) are the reflectivity of the two mirrors. \(l\) is both the cavity length and the gain material length. For simplicity, we take \(R_1 = R_2 = R\). Then the solution of (2.68) is
\[
P_{out} = P_+(l)T = \frac{P_s}{2} (kP_pl + \ln R)
\]  
(2.71)

Since mirror loss is the dominant factor of cavity loss, we have
\[
kP_{th} = -\frac{1}{l} \ln R
\]  
(2.72)
Replacing $kl$ by $-\frac{1}{P_{th} \ln R}$, the emission efficiency becomes

$$P_{out} = P_{+}(l)T = \frac{P_s}{2} \ln \left( \frac{1}{R} \right) \left( \frac{P_p}{P_{th}} - 1 \right)$$

(2.73)

In both cases, the laser emission does not occur until pump power exceeding threshold power. This is justified by most of the well studied laser systems.

Figure 2.4: Laser output power vs. input power curves. Source: Lasers (Mill Valley, CA : University Science Books, ©1986 [1]). Using with permission.
CHAPTER III

A NOVEL ALL-OPTICAL SWITCH

3.1 Waveguide Design

In the previous chapter, we derived an analytical solution for Asymmetrically pumped FWM-BS (2.40). We discovered that a configuration to suppress noisy FWM-MI and enhance noise free FWM-BS simultaneously is possible. The requirement is setting the weak pump wavelength in the middle of the undepleted pump and the signal meanwhile setting the GVD coefficient $D$ negative. We use silicon waveguides to perform the experiment for two major advantages. First, silicon waveguide can confine the field in a small area for a long distance. Second, the dispersion can be well controlled by the structure. Silicon on insulator (SOI) is chosen as the platform. GVD coefficient is an inherent property of waveguide material and structure thus needs to be considered during design the waveguides. Another designing consideration is field intensity inside waveguides which is directly linked to the efficiency of FWM-BS. This can be characterized by input power and effective mode area. The former is directly controlled by the input sources. The latter is defined in (2.26f). Like GVD, effective mode area is also related with waveguide material and structure. Finally, silicon waveguide modes are dramatically different from optical fiber modes, special treatment is needed to improve light coupling between fibers and waveguides. In this section, we use numerical tools to find out the proper waveguide structure for the FWM-BS based optical switching.
As introduced in Chapter II, GVD is a combination of material dispersion (MD) and waveguide dispersion (WD). It has been shown that in small structure SOI waveguides WD predominates GVD and can be tailored with large flexibility [36]. We used a Finite Element numerical solver to calculate the GVD coefficient for TE\textsubscript{11} modes in silicon waveguides with 500\,nm height and various widths. Two cladding materials, SiO\textsubscript{2} and Si\textsubscript{3}N\textsubscript{4} are considered. As shown in Figure 3.1, Si\textsubscript{3}N\textsubscript{4} cladding gives a more negative $D$ coefficient, but SiO\textsubscript{2} cladding also satisfies our purpose for width larger than 0.8\,$\mu$m.

Due to the complexity of fabrication, perfect vertical sidewalls are not always achievable. Our further simulation shows that while sidewall angle does affect the GVD coefficient, the influence is similar to changing of widths and is minimal. The result is shown in Figure 3.2.

Next we consider the effective area. It is known that the field size does not decrease monotonically with the decreasing of core size. Again, we simulated silicon waveguides with 500 nm height and various width. Here we show the result of SiO\textsubscript{2} cladding but Si\textsubscript{3}N\textsubscript{4} have vary similar numbers. From Figure 3.3, we see that effective area reaches its minimum at width between 400 nm to 600 nm. Considering smaller waveguides are also more susceptible to scattering loss, we choose to fabricate waveguides with size from 600 nm $\times$ 500 nm to 1 $\mu$m$\times$500 nm.

Figure 3.1: (a) Group velocity dispersion curve of SOI waveguides passivated by SiO\textsubscript{2}. (b) Group velocity dispersion curve of SOI waveguides passivated by Si\textsubscript{3}N\textsubscript{4}.
To have a good light coupling efficiency, it is preferable to match the input field with the waveguide mode. The power coupling efficiency between single optical modes can be estimated by

\[
\eta = \left[ \frac{4\beta_i\beta_t}{(\beta_i + \beta_t)^2} \right] \frac{\int\int_{-\infty}^{\infty} E_t(x,y)E_i^*(x,y)dxdy}{\int_{-\infty}^{\infty} |E_t(x,y)|^2dxdy \int_{-\infty}^{\infty} |E_i(x,y)|^2dxdy} \tag{3.1}
\]

where \( E_i/E_t \) is the incident/transmitted field and \( \beta_i/\beta_t \) its normal propagation constant. To achieve good mode matching, we designed tapers on both sides of waveguides with 500 \( \mu m \) transition length. The taper size is chosen to be 2,3 and 4 \( \mu m \). At the input side, we choose to use tapered
lens fiber which produces an approximate Gaussian spot with 2 µm waist. We also used finite difference beam propagation method (BPM) to verify that the optical mode order is not perturbed during transition region, as shown in Figure 3.4.

Figure 3.4: (a) Finite difference beam propagation method simulation of mode evolution during transition. An 150 µm transition length is enough to preserve mode quality. 500 µm transition length is chosen in real devices to increase tolerance. (b) Optical microscope image of waveguide to lens fiber light coupling.

### 3.2 Waveguide Fabrication

In this work, two types of lithography techniques, stepper photolithography and electron beam lithography (EBL), have been investigated for waveguide fabrication. For both lithography techniques, the silicon substrate is coated with ultraviolet light (UV) sensitive or electron beam (E-beam) sensitive polymer, called photoresist or E-beam resist. The chemical property of these resists is altered after its absorbing of UV light or electrons. If the resist becomes more reactive (inactive) after exposing, it is called positive (negative) resist. The comparably reactive part of the resist can be dissolved in proper solvents, called developer, while leaving the inactive part intact. This allows an optical (or electron beam) pattern be transmitted onto a physical mask on top of the silicon wafer.
in a sub-micrometer to nanometer scale precision. Followed by etching procedures, the pattern can eventually be transmitted onto silicon itself.

For waveguide fabrication, the wall smoothness is of special concern. The feature roughness, usually much smaller than the wavelength, can be regarded as an extra dielectric particle. The time varying electric field induces a time varying dipole moment on the particle, which can radiate energy to directions other than the incident direction. This phenomenon is called Rayleigh scattering. It is much severe in integrated waveguides than in optical fibers. The reason is as follows. First, integrated waveguides usually have higher index contrast between core and cladding. This cause a much larger permittivity difference between roughness and environment hence more significant power scattering. As an example, a typical silicon–silicon dioxide waveguide has more than 300 times refractive index contrast (100000 times permittivity contrast) than Corning SMF-28 optical fiber. Second, due to strong light confinement, silicon waveguide modes have large portions of power concentrated near the boundaries, i.e. large portions of power that is susceptible to scattering loss. Lots of works has been done to reduce sidewall roughness of silicon waveguides, including surface passivation [37], thermal oxidation [38], annealing with hydrogen ambient [39], and resist reflow. Thermal oxidation and annealing both require extreme conditions (above 1100°C) and have poor structure selectivity. Thus these two methods are not compatible to complementary metal-oxide semiconductor (CMOS) processes. In this work, we choose to use resist reflow and passivation with silicon dioxide or silicon nitride. The maximum temperature required is around 350°C during plasma enhanced chemical vapor deposition (PECVD). The whole process is CMOS compatible.

Stepper photolithography is the most widely used lithography technique in semiconductor industry. It projects the pattern from a otherwise manufactured mask onto the photoresist, similar to a slide projector but usually with reduced image size. The silicon wafer is loaded on a stage beneath
the projection lens. After one exposure, the stage "steps" and move the wafer to its next exposure location. Stepper lithography has the advantage of fast processing speed (one exposure needs only a few seconds) and low marginal cost. In this work, we use an i-line (365 nm wavelength) GCA 6100 stepper. It provides a 5X pattern size reduction. The depth of focus is 2 µm and the resolution is 550 nm. The field size for each exposure is 17.5 mm. The procedure used is as follow.

1. Clean the sample with acetone followed by isopropanol (IPA). Dry with nitrogen.

2. Bake the sample in hexamethyldisilizane (HMDS) prime oven for 30 minutes.

3. Spin coat with SPR-955CM photoresist. Keeping the velocity at 3000 rps for 30 s yields photoresist thickness of 0.7 µm.

4. Soft bake sample on hotplate at 105°C for 60 s.

5. Expose on stepper for 1.8 s.

6. Post exposure bake at 110°C for 60 s.

7. Develop in Shipley MF®CD-26 developer for 60 s.

8. Rinse in flowing deionized (DI) water for 120 s.

Electron beam lithography takes advantage of the fact that electrons has much shorter wavelength (1.23 nm) than UV light (from 13.5 nm for extreme UV to 436 nm for g-line). A typical resolution for EBL is 25 nm. The limitation arise from the scattering of electrons inside resists rather than diffraction. Due to source brightness problem, EBL use point scan method to construct the whole pattern rather than single exposure. On the one hand, this means a physical mask is not necessary for EBL process. On the other hand, the exposure time needed for EBL is significantly longer than photolithography. In industry, EBL is usually used to fabricate stepper lithography mask, but
not used for large scale manufacturing. In this work, the EBL tool used is Vistec EBPG5000. The procedure is as follow

1. Clean the sample with acetone followed by IPA. Dry with nitrogen.

2. Bake the sample on hotplate at 115°C for 5 minutes.


4. Expose. Use 600µC/cm² dose and 100 kV acceleration voltage. The spot diameter is 60 nm and current is 40 nA. Both resolution and step size are 25 nm.

5. Develop in MF™-319 for 70 s.

6. Rinse in DI water for 5 min.

7. To mitigate resist roughness, reflow on hotplate at 120°C for 5 minutes.

To transfer the resist pattern onto silicon itself, inductively coupled plasma reactive ion etching (ICP-RIE) is used. ICP-RIE is one of the chemical-mechanical dry etching techniques, meaning the ions have both high momentum to bombard target atoms out of their location and proper chemical property to form gaseous reaction products with the target atoms. The recipe used for silicon etching is

<table>
<thead>
<tr>
<th>Gas</th>
<th>Flow Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl₂</td>
<td>20 sccm</td>
</tr>
<tr>
<td>RIE power</td>
<td>28 W</td>
</tr>
<tr>
<td>ICP power</td>
<td>500 W</td>
</tr>
<tr>
<td>He flow</td>
<td>5 sccm</td>
</tr>
<tr>
<td>Cathode temperature</td>
<td>19°C</td>
</tr>
<tr>
<td>Pressure</td>
<td>5 mTorr</td>
</tr>
</tbody>
</table>

The etch rate for silicon is 3.5 nm/s. The selectivity for SPR-955CM is 1:1 (silicon vs. resist), for maN-2403 is 2:1.

The fabrication results are shown in Figure 3.5. The scan electron microscope (SEM) images
show that EBL waveguides have smaller overall roughness compared with stepper waveguides. Resist reflow further reduces the roughness of EBL significantly. As expected, resist reflow generates rounded sidewalls. But this does not have significant influence for our application. The final samples consists of 21 mm waveguides with height of 500 nm and width varying from 600 nm to 1 µm in 100 nm increments. Taper sizes include 2 µm, 3 µm and 4 µm. Either SiO₂ or Si₃N₄ claddings is used.

### 3.3 Loss Measurement

The propagation loss of the waveguide is estimated from the Fabry-Perot effect shown by the waveguide. The transmission coefficient of a normal incident Fabry-Perot cavity can be expressed
as

\[ T_{FP} = \kappa^2 \frac{(1 - R)^2 e^{-\alpha L}}{1 + R^2 e^{-2\alpha L} + 2Re^{-\alpha L} \cos \left( \frac{4\pi n L}{\lambda} \right)} \]  (3.2)

where \( \kappa \) is the mode coupling efficiency for each facet. \( R \) is the reflectivity of the silicon-air facet. \( L \) is the length of the waveguide. \( n \) is the effective refractive index. \( \alpha \) is the propagation loss. With fine turning of the wavelength and the laser induced thermal expansion, the \( \cos \left( \frac{4\pi n L}{\lambda} \right) \) term can scan more than a whole period. If we keep the input power \( P_{in} \) constant and define

\[ \eta = \frac{T_{FP_{max}}}{T_{FP_{min}}} = \frac{P_{max}}{P_{min}} \]  (3.3)

where \( P_{max}/P_{min} \) is the maximum/minimum output power. Then the propagation loss can be found as

\[ \alpha = \frac{1}{L} \ln \left( R \sqrt{\eta + 1} \right) \]  (3.4)

Alternatively

\[ \alpha_{dB} = 10 \frac{\log_{10} \left( R \sqrt{\eta + 1} \right)}{L} \]  (3.5)

If we define

\[ T_{max} = \frac{P_{max}}{P_{in}} \]  (3.6a)
\[ T_{min} = \frac{P_{min}}{P_{in}} \]  (3.6b)

Then we can express \( \kappa \) as

\[ \kappa = \frac{2}{1 - R} \sqrt{\frac{T_{max} T_{min}}{T_{max} - T_{min}}} R \]  (3.7)

The test setup is shown in Figure 3.6. A 0.001 nm finely turnable laser at telecommunications band is used as the light source. Two single mode lens fibers are used for input and output coupling to achieve a good selectivity to fundamental waveguide modes. The light coupling is assisted by peizo controlled stages and an optical microscope. The input and output powers are recorded by a calibrated InGaAs power detector.
Figure 3.6: Setup for waveguide loss measurement. TL, turnable laser. PC, polarization controller. LF, lens fiber. TS, 3D translation stage. RS, 3D rotation stage. Si, tested sample. CM, compound microscope. CCD, camera. PD, calibrated power detector.

Table 3.1: Propagation loss measurement with Fabry-Perot method.

<table>
<thead>
<tr>
<th>Width (µm)</th>
<th>Propagation Loss (dB/cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.4</td>
</tr>
<tr>
<td>0.9</td>
<td>3.7</td>
</tr>
<tr>
<td>0.7</td>
<td>4.3</td>
</tr>
<tr>
<td>0.6</td>
<td>6.1</td>
</tr>
</tbody>
</table>

The measured propagation loss for TE\textsubscript{11} mode is listed in Table 3.1. The coupling efficiency for 2 µm, 3 µm, 4 µm tapers are 10 dB/facet, 5.5 dB/facet, 4.9 dB/facet respectively. The 4 µm taper, however, has a large coupling efficiency variation between waveguides. This is suspected to be mode hoping because silicon waveguide is heavily multimoded at this size.
3.4 Four Wave Mixing Test

In this part of the experiment, we use modulated laser pulses to characterize FWM-BS inside the fabricated waveguides. The setup is shown in Figure 3.8. Three fiber coupled tunable lasers at telecommunication band are used to generate the bias, control and input signal respectively. All three laser signals are combined into one optical fiber by fused fiber couplers. The combined beam is then modulated by an electro-optical amplitude modulator. Three polarization controllers are used to ensure good extinction ratio. The modulated light is then amplified by an erbium doped fiber amplifier (EDFA) to provide the necessary power for nonlinear interaction. The power ratio of the amplified signals can be controlled by the input laser powers and the wavelength locations. The amplified laser signals are split and recombined by dense wavelength division multiplexers.
(DWDM) to filter out the majority part of the EDFA amplified spontaneous emission (ASE) noise. Polarization controllers are also used on each channel to ensure well controlled fiber to waveguide coupling efficiency. Fiber delay lines are used to ensure the synchronization of the pulses. The recombined lasers are coupled into the silicon waveguide by microscope assisted butt coupling. Due to the big coupling efficiency difference between waveguide TE$_{11}$ and TM$_{11}$ modes, the coupling efficiency of each wavelength can be tuned by its polarization controller respectively. This together with laser source power tuning allows a flexible power control for each input channel. An optical spectrum analyzer (OSA) is used to analyze power conversion efficiency which is defined as

$$\text{Conversion efficiency} = \frac{\text{Power in one idler}}{\text{Power in signal} + \text{Power in two idlers}}$$

(3.8)

Noise performance is evaluated by balanced detection, where a DWDM is used to filter out the interested channel and a 40/60 fused fiber coupler and an adjustable attenuator are used to get two signals with the same power. The lengths of the two optical fibers are also carefully controlled to minimize the time jitter. An electrical spectrum analyzer (ESA) is used to analyze the radio frequency (RF) power spectrum.

![Figure 3.9](image)

Figure 3.9: (a) OSA measurement of the fiber coupled output spectrum. (b) Split-step Fourier method simulation of similar condition.

30 ns long pulse is used in this experiment. The 0.9 µm wide waveguides are chosen to perform
the experiment out of a balance between propagation loss and effective area. The OSA measurement is shown in Figure 3.9. The calibration is done as follows. The total input average power is directly measured by a calibrated InGaAs power meter. The power ratio between each input is estimated by the output spectrum measurement at low power. The output spectrum is used because the input power is partially controlled by the polarization which cannot be measured directly by the OSA. The average power for each input is found according to the ratio. Finally, the peak power is found by dividing the average power by the duty cycle and extinction ration. The input bias pulse peak power is found to be 775 mW and the control pulse 60 mW at the input of the waveguide. A -16 dB conversion from signal to each idler is observed. Cascaded FWM-BS idlers are also observed. As expected, FWM-MI idlers also appears, however, the highest idler is still more than 15 dB lower than its corresponding signal so no significant gain is provided by this process. To evaluate the performance of the device, we simulated the same condition with Split-step Fourier method (SSFM). Extra two photon absorption loss is added to the simulation. The nonlinear refractive index $n_2$ and the two photon absorption coefficient $\beta_T$ is found according to the measurement by [40]. As shown in (b), the simulation spectrum well matches the experiment spectrum. The simulation predicted a -15.5 dB conversion efficiency for each idler. Error sources could be as follows. First, the $n_2$ and $\beta_T$ may not exactly match our sample. As the author of [40] pointed out, the measurement done by different groups or with different methods does not alway agree with each other. Secondly, the simulation only considered dispersion, and both real(FWM) and imaginary (TPA) part of $\chi^{(3)}$ process. Other processes such as Raman scattering, self steepening and free carrier absorption are not included. Thirdly, the effective area is taken from the simulation for rectangular waveguide. The real shape is likely smaller since the sidewall of the waveguides curved in.

The Raman gain spectrum in silicon is around 105 GHz thus is not a major source of noise for the signal and idlers. However, free carrier absorption inside silicon can be strong due to the
existence of the strong pump and TPA. Even though electron momentum relaxation is relatively fast (subpicoseconds), it is still slow compared with optical frequency. The excited electrons will instantaneously influence the optical field after their creation [32], causing extra noise. Also, a maximum gain of 3dB can still be introduced by non phase matched FWM-MI. We use balanced detector followed by ESA analysis to measure the added noise. When the two input ports of the balanced detector have the same power, common mode noise such as relative intensity noise caused by spontaneous emission is rejected. The 0 frequency spectrum of ESA should be 0 and other frequencies corresponds to shot noise level. When one input port is attenuated to 0, the 0 frequency of the ESA spectrum corresponds to average power and the other frequencies corresponds to intensity variance at that frequency. Because the total input power is now decreased by a factor of 2, another 3dB power should be added to the second measurement to make it consistent with the first. In practice, common mode noise rejection is finite. As shown in Figure 3.10, All noise measurements are calibrated to the same power level. The shot noise is supposed to be only dependent on photon number, namely power, so only one balanced measurement curve is plotted. Due to the amplification through EDFA, the initial signal has 15 dB added noise. The processes occurring inside silicon
waveguides add another 3 dB of noise to both signal and idler. Other than the non avoidable noise sources discussed above, this is also very likely caused by the excessive noise carried by the bias and control lasers.

The independent influence of the control and bias power to the conversion efficiency is also experimentally measured (Figure 3.11). We notice that both curve showed close to linear behavior. However, in the conversion efficiency vs. bias curve, the slope is not 1. This is because the bias is the main contribution to TPA (include cross two photon absorption). Though analytical relation is not derived, we guess this is because TPA has the same intensity dependence as $\chi^{(3)}$ wave mixing. The combined effect causes a slowed down conversion, but it still follows a linear power dependence even though TPA is quite significant. The conversion efficiency vs. control line showed a close to 1 slope. This mean aside from binary switching, continuous modulation is also achievable.

Several potential improvements can be applied to increase the conversion efficiency. Firstly, a low propagation loss will be helpful. At the current stage, smaller core waveguides does not perform well because of the propagation loss. Conversion efficiency can be significantly improved if propagation loss can be reduced. Surface melting and re-consolidation processes are possible.
because of the good geometric shape tolerance of the nonlinear process. Secondly, for intra chip interconnections, telecommunication band wavelength is not necessary. Two photon absorption decrease to nearly 0 when the wavelength exceeds 1.7 $\mu$m [32]. Silica optical fiber is still suitable for short distance communication at this wavelength. We will introduce the study of an on silicon laser emitting at 2 $\mu$m in the next chapter.
As we have introduced in the previous chapters, the band gap of Ge can be altered by incorporating with Sn. More specifically, both direct energy gap (DEG) and indirect energy gap (IEG) decreased with the increasing of Sn concentration, but the DEG decreases faster. As shown in Figure 4.1(a), the DEG becomes material band gap at Sn concentration larger than 7.4%. However, higher Sn concentration is more likely to result in defects of the crystal structure due to the lattice constant mismatch. These defects lead to higher non-radiative recombination rate and lower emission quantum efficiency. In this work, we demonstrate observed stimulated emission at room temperature using 4.4% Sn concentration Ge with phosphorus doping.

The material fabrication is done in Arizona State University. First a 850 nm thick pure Ge buffer layer is deposited on p type silicon by low pressure chemical vapor deposition (LPCVD). This layer helps reduce the lattice mismatch between the top Ge and bottom Si. Then 625 nm of GeSn layer is deposited by the same process. Phosphorous doping is done in situ simultaneously. The doping level is $3.7 \times 10^{19}$ cm$^{-3}$. 4 mm long ridge waveguides are then fabricated by photolithography and reactive ion etching. The waveguide width vary from 1.5 $\mu$m to 100 $\mu$m. The surface is then protected by around 1 $\mu$m thick SiO$_2$ through plasma enhanced chemical vapor deposition (PECVD). Finally, the facets are polished by mechanical polishing. Due to the high refractive index of Ge ($\approx 4.0$), each polished facet has 37% reflectivity. With the strong mode confinement, these waveguides can be
regarded as Fabry-Perot resonators with 3.7 finesse. These resonators behave as cavities for laser operation. The expected direct energy gap is 6.6 eV [3], corresponding to a emitting wavelength of 1.88 \( \mu m \).

![Figure 4.1: (a) Energy gap of GeSn alloy vs. Sn concentration. [3]. (b) Spontaneous emission spectrum for various GeSn alloys. [3] (c) Illustration of Optical pump. (d) Illustration of material structure.](image)

The stimulated emission testing setup is shown in Figure 4.2. Optical pump is chosen to increase gain area and avoid unexcited material absorption loss that is commonly seen in homojunction diode lasers. Both Ge buffer layer and silicon layer have larger band gap than the emission wavelength.
so direct absorption is avoided. Free carrier absorption and two photon absorption are negligible at the targeted power. A high power fiber laser at 974 nm wavelength is used as the pump source. The output beam has a diameter of 0.4 mm inside the optical fiber. This beam is collimated by a 60 mm focal length lens after leaving the fiber. The collimated beam is 1” in diameter. It then passes through a cylindrical lens to purposely introduce astigmatism. This eventually results in an elliptical spot on the sample surface. The spot length is controlled by the focal length of this lens. The spot width is independent of it. Two gold coated mirrors are used to redirect the beam for top down projection. A second spherical lens is used to compress the beam into 5 mm × 0.9 mm elliptical spot at its focal plane. The projection direction forms a 30° angle with the vertical direction, leaving room for the microscope to monitor the light output coupling process. The tested sample is set at the focal plane by a translation stage. The optical output is coupled into a multimode fiber with the help of another translation stage and the microscope. The collected emission is then collimated and passed through a long pass filter cutting off at 1560 nm. This filter avoids the detection being polluted.
by the pumping light. The detection is down by an InGaAs photodetector and a transimpedance amplifier.

![Graphs](image)

Figure 4.3: (a) Laser efficiency measurement without pump contamination. (b) Laser efficiency measurement with pump contamination. (c) Pump power collected by optical fiber. (d) Microscope image of light coupling configuration.

The emission measurement is taken from the 3\(\mu\)m wide GeSn waveguides. Emission from multiple waveguide emitters is collected simultaneously, as shown in Figure 4.3(d). Figure 4.3(a) shows the output power vs. pump power curve. Strong nonlinear increasing is observed after pump power exceeding 40 mW per waveguide. To completely rule out the possibility of pumping light contamination, the pump filter LPF is removed and the measurement is repeated, shown in (b). Finally, the fiber is intentionally misaligned with the waveguides to get the pump power behavior measurement.
(c). We observed a considerable amount of pump light inside the collecting fiber. This is due to the large core (62.5 µm) and large numerical aperture of gif625 fiber. However, this part of the power increases linearly with the total pump power and the LPF filtering is proven efficient. As spontaneous emission also increase nearly linearly with the pump power, we conclude that the nonlinear behavior in (a) is resulted from stimulated emission inside the GeSn waveguides. Even though Figure 4.3 resembles the laser efficiency curves shown in Figure 2.4 and equation (2.73), to this stage, it is not conclusive to distinguish between ASE and lasing. But a material capable of producing ASE is generally also capable of producing lasing given the right cavity. So we claim that GeSn alloy is a room temperature laser material.
CHAPTER V

CONCLUSIONS AND FUTURE WORKS

We have theoretically analyzed and experimentally demonstrated an on silicon low noise all-optical switch. The extinction ratio can be further improved by increasing the smoothness of the waveguide structures, applying resonance structures or using longer wavelength as the bias. Because silicon waveguides have good bending loss tolerance, the footprint of the switches can also be significantly reduced by bending the waveguides. Polarization switching can be tested in the near future.

We have also experimentally demonstrated the lasing potential of the GeSn alloy material. Obvious stimulated emission is observed. Further spectrum measurement is needed to better understand the coherence of the emission.


