Photonic Integrated Circuit Packaging Using Silicon Based Optical Interconnects

by

ENS Drew Michael Weninger

Bachelor of Science (B.Sc.) in Honors Physics
United States Naval Academy, Annapolis, Maryland (2019)

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Materials Science and Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2021

Author . . . .
Department of Materials Science and Engineering
May 24, 2021

Certified by . . . .
Lionel C Kimerling
Advisor, Thomas Lord Professor of Materials Science and Engineering, MIT

Certified by . . . .
Anuradha Murthy Agarwal
Co-advisor, Principal Research Scientist, DMSE, MIT

Certified by . . . .
Samuel Serna Otalvaro
Co-advisor, Assistant Professor of Physics, Bridgewater State University

Certified by . . . .
Cort Johnson
Co-advisor, Senior Technical Staff Member, The Charles Stark Draper Laboratory

Accepted by . . . .
Frances M. Ross
Chair, Departmental Committee on Graduate Studies, DMSE, MIT

Photonic Integrated Circuit Packaging Using Silicon Based Optical Interconnects

by

ENS Drew Michael Weninger

Submitted to the Department of Materials Science and Engineering
on May 24, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Materials Science and Engineering

Abstract

As electrical interconnections reach their physical limit, optoelectronic packaging solutions are a critical bottleneck in the effort to seamlessly integrate silicon photonic devices - a technology capable of handling high data rates for next generation data center telecom applications. In this thesis, novel silicon based optical couplers capable of low loss, robust connections from an optical fiber to a photonic integrated circuit (PIC) and a PIC to another PIC are presented. The first of these, the fiber-to-chip coupler, utilizes a graded index material stack in the vertical direction and a non-adiabatic taper in the horizontal direction to focus light into a single mode waveguide, all while maintaining planarity and a monolithic design. Distinguishing features from prior designs are made in the form of coupler rotation, added structural elements, and removal of all curved facets. Notable advantages include customization of the output PIC waveguide, high intrinsic coupling efficiency, wide alignment tolerances, CMOS compatibility, and scalability to mass manufacturing.

Following this will be the description of the adiabatic, inverse $Si_xO_yN_z$ cross tapers which provide vertical evanescent coupling of light from one PIC to another PIC. A final beam expansion design is presented for chip-to-cip coupling, one capable of maintaining silicon input and output waveguides while increasing tolerances to within the accuracy capabilities of high speed pick and place die bonders. Simulations using 3D finite-difference time-domain (FDTD) and eigenmode expansion (EME) methods were utilized to determine quantitative performance metrics including coupling efficiency, packaging misalignment tolerance, and wavelength and polarization dependence. The comparison of these coupling designs to other state of the art fiber-to-chip and chip-to-chip coupling designs was also done to evaluate the their performance in the context of their peers.

Thesis Supervisor: Lionel C. Kimerling
Title: Advisor, Thomas Lord Professor of Materials Science and Engineering, DMSE, MIT

Thesis Supervisor: Anuradha Murthy Agarwal
Title: Co-advisor, Principal Research Scientist, DMSE, MIT

Thesis Supervisor: Samuel Serna Otalvaro
Title: Co-advisor, Assistant Professor of Physics, Bridgewater State University

Thesis Supervisor: Cort Johnson
Title: Co-advisor, Senior Technical Staff Member, The Charles Stark Draper Laboratory

Acknowledgements

Coming to MIT was a dream turned into a reality for me. The fact that I was able to come here, and that I am still here now, is less a reflection of me than it is of all the people who offered me help, support, encouragement, and advice along the way. There are far too many to name, but I'll do my best here. First, I would like to acknowledge and express my deep gratitude for the great mentorship of my research advisors, Professor Kimerling and Dr. Agarwal, throughout my time spent at MIT thus far. Kim and Anu's encouragement, confidence in my abilities, and extraordinary wealth of experience made carrying out this research significantly easier, more productive, and most importantly fun. In the midst of a pandemic, their strong leadership helped me stay on track and level headed when there was so much to be concerned about - I am beyond thankful I get to work with them. I want to thank Dr. Samuel Serna, a former postdoc in the EMAT Group and now Assistant Professor of Physics Department at Bridgewater State University. Samuel continues to offer his time and wisdom to deepen my learning and make this project a success, even in the midst of starting his own laboratory at BSU - he's truly brilliant and an even better person. I also want to acknowledge my classmates in DMSE and group members in the EMAT/PMAT Groups, especially Katherine Stoll and Eveline Postelnicu, for cultivating an outstanding research environment and, above all else, offering their friendship. I am also grateful to The Charles Stark Draper Laboratory for funding my education over the past two years through their fellowship program, and especially to my Draper advisor, Cort Johnson, for allowing me the freedom and flexibility to pursue a wide latitude of research interests and offering valuable, precise insight throughout the course of the last two years.

Lastly, it simply wouldn't be right if I did not thank my family and friends for understanding my mysterious, busy schedule and providing me with whatever I needed, whenever I needed it. To Wolf, it can't be put into words how much we all continue to miss you each and every day. Without knowing it, you pushed me to be the best version of myself I

could be, whether that meant trying to follow in your footsteps as an athlete or match your passion for military service by going to USNA. I can only hope to live with the same level of character and purpose as you did in the 24 years we shared together. To all the people I named and all the people I didn't, but should've - thank you all. It goes without saying that I am ecstatic I chose to stay and for what is in store over the remainder of my time at MIT, and I look forward to the challenges it may bring.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

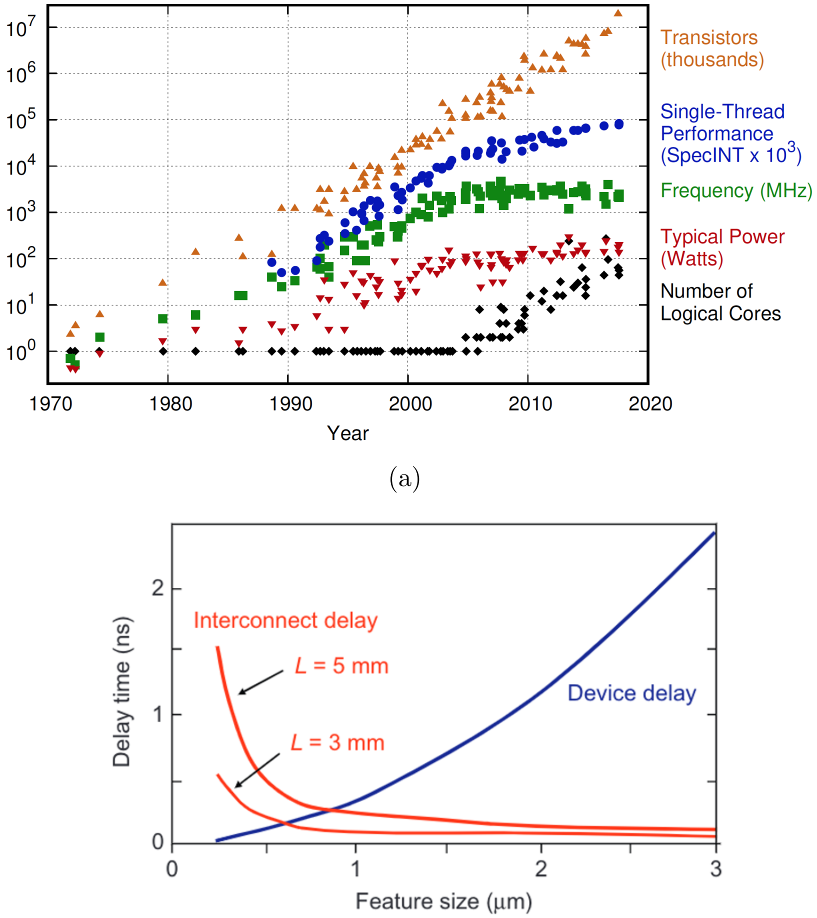
Part I

Introduction, Motivation, and Background

Chapter 1

Motivation for Optical Interconnect Research

The advance of the modem electronic industry has totally revolutionized the way human beings are living and thinking. Moore's law has been governing the scaling of electronic integrated circuits (ICs) for more than 50 years since Dr. Gordon E. Moore introduced it in 1965 [1]. It states that the number of transistors on a silicon (Si) chip doubles every 18 months, and is shown in Figure 1.1(a) from [2]. The drive to continue Moore's Law has led the minimum feature size on ICs to advance from 10 μm down to 7-10 nm, the cost per transistor to decrease by seven orders of magnitude, and the maximum number of transistors per chip to increase by at least 10 orders of magnitude [3]. By increasing density, the performance of the semiconductor microprocessors is improved. Specifically, this is accomplished by increasing the frequency at which transistors can be turned on and off, governed by the gate delay. By decreasing size, this gate delay continues to be reduced. However, as the minimum feature size shrinks, the thinner line width and pitch results in an increase in resistance of the metal lines interconnecting transistors and an increase in capacitance between neighboring metal lines. The delay in the metal interconnects is governed by RC delay, which is equal to resistance times capacitance. Therefore, with increased complexity and total length, the copper interconnects become the bottleneck in the improvement of IC performance as their RC delay surpasses the gate delay [4]. This includes copper interconnects not on the IC, but routing signals from external networks onto the microprocessor. This relationship is shown by Figure 1.1(b) [4].



(a)

(b)

Figure 1.1: An illustration of the limits facing copper interconnects in microelectronic devices. In (a) Moore's Law demonstrates the increase in the number of transistors on an electronic IC, thus decreasing the size of the interconnects and transistors. Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten. New plot and data collected for 2010-2019 by K. Rupp. Image from [2]. In (b) the gate delay vs RC delay is shown for Cu interconnects. Note that as minimum feature size decreases, RC delay dominates, image from [4].

At the same time as RC delay becomes a roadblock to continued IC performance, data centers continue to see a skyrocketing of information transfer and storage. In addition, the nature of modern applications requires that at least three quarters of the traffic stays internal to the data center through communication between racks as opposed to traffic that enters and exits the data center [5]. This equates to significantly larger amounts of information transferred over the same copper interconnects. Meanwhile, the energy consumed during data transfer becomes an issue whether it's by servers processing the data, cooling devices to dissipate all the heat being generated by the servers, or in powering the storage devices. Data centers currently account for approximately 2.5% of electricity consumption and 1.5% of the total energy consumed in the US as of 2018, at a cost of $4.5 billion according to [6, 7]. In [5] it was put in more practical terms - Google reports that for a month's worth of the typical active Google user activity (25 searches, one hour of YouTube, Gmail and other services per day), Google emits the equivalent amount of CO_2 as driving a car for one mile [5].

Even as smaller node sizes continue to allow the number of the transistors to increase, diminishing returns on performance are struggling to keep pace with increased data center traffic. It has now become cost effective to turn to alternative solutions for interconnects that can handle higher bandwidth. The most promising option to accomplish this complex task is in silicon photonics. Optical interconnects offer high data bandwidth, no heat dissipation (less power consumption), and no electromagnetic interference, allowing for a tremendous improvement. On the micro- and nano-scale, silicon provides an ideal high index contrast material platform, allowing for chip level photonics integration by leveraging of current CMOS technologies. This compatibility with current fabrication techniques significantly reduces the potential cost and provides a scalable solution for mass production. At the same time, functionalities such as modulators and transceivers are being increasing switched over to photonic devices on photonic integrated circuits (PICs). In other words, silicon photonics allows us to achieve the next generation of performance through devices which can be integrated alongside current electronics technology, without a total overhaul of the manufacturing system. As functionality and routing is converted more to photonic devices, those devices are getting closer and closer to the electronic microprocessor. This is shown in Figure 1.2. To this end, advanced packaging solutions are needed to provide simultaneous connection between optical fibers, electrical traces, PICs, ICs, and all combinations thereof.

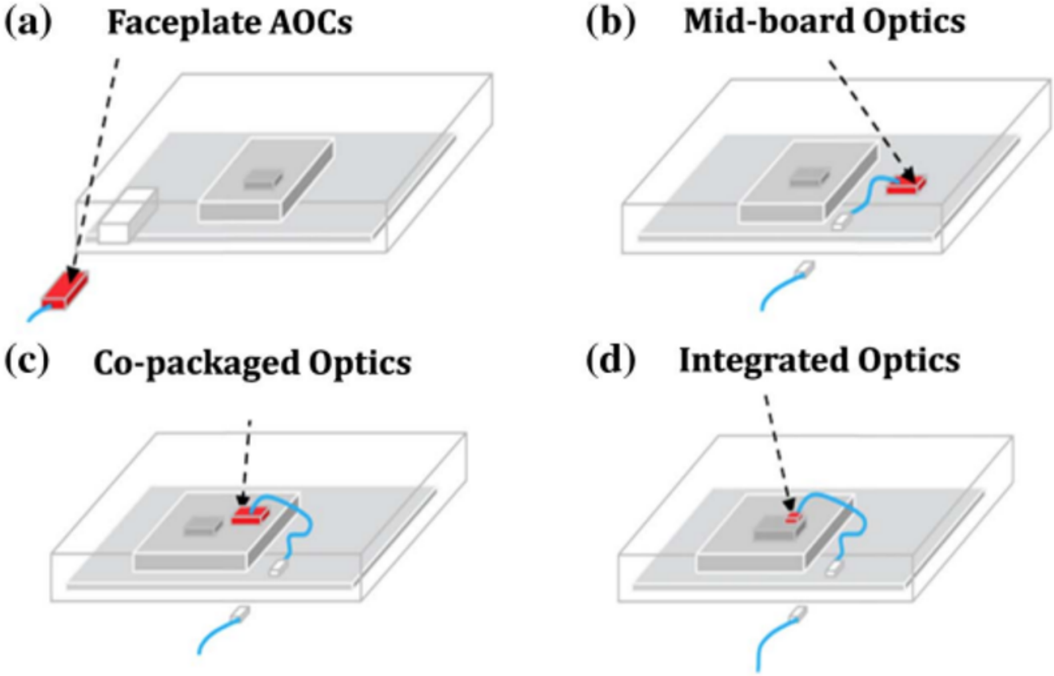


Figure 1.2: PICs are being brought closer to the electronic IC by a combination of interconnection and functionality transfer to optical devices, image originally from [8] ©2015 IEEE.

There have been considerable advances in high-bandwidth optical interconnects for the data center. Large-scale data centers adopted optical transmission technologies during the transition from the 1 to 10 Gb/s link data rate between 2007 and 2010 [9]. In 2007, Google introduced optical communication in its data centers in the form of 10 Gb/s vertical-cavity surface-emitting lasers (VCSELs) and multimode-fiber-based small form-factor pluggable (SFP) transceivers for reaches up to 200 m [10]. This was followed in 2017, when 40 Gb/s Ethernet-based DCIs were deployed in production data centers [11]. Likewise, 100 Gb/s links have been commercially available since 2014, and are currently installed in production data centers, with 400 Gb/s equipment is expected to emerge in the near future [10].

Due to this rapid uptick in optical interconnect technology, the Integrated Photonics System Roadmap-International (IPSR-I) is developing a common roadmap for low-cost, high-volume manufacturing of photonics for data and telecommunications systems [12]. Among the key findings was the emphasis on early deployment of "2.5-D" integrated photonic technologies, packages in which chips are placed side by side and interconnected through an interposer or substrate. By 2025, there should be pervasive deployment of wavelength-division multiplexed (WDM) interconnects and the beginnings of commercial chip-to-chip intra-package photonic interconnects. This would allow for hyperscale datacenters to not only meet current data traffic requirements, but be prepared for future energy and bandwidth needs as exponential signal transfer trends continue.

This technology is by no means limited to datacom and telecom application in hyperscale datacenters. The importance of photonic packaging extends to several other core market areas as well, including radio-frequency (RF) communication in drone technology, the Internet of Things (IoT) and biochemical sensing devices, and the autonomous vehicle sector where LiDAR usage is crucial. Each of these markets displays a situation where high bandwidth signal routing is essential. For example, in a self-driving car the sensors responsible for relaying data regarding the environment such as road signs, other vehicles, moving objects or pedestrians, weather, terrain information, etc. must be constantly providing feedback to a microprocessor which can make decisions on the next action, even if that action is to do nothing, resulting in data traffic stacking up very, very quickly. Therefore, ubiquitous in these future technologies is the need for low loss, robust optical interconnections capable of handling these exceptional future data rates and manufacturable on a large scale at low cost. The goal of this work is to provide insight into a potential solution for some optical interconnection regimes - specifically the fiber-to-chip and chip-to-chip coupling regimes.

1.1 Thesis Structure

This thesis will discuss optical interconnection solutions in the form of a fiber-to-chip coupler and a chip-to-chip coupler for high efficiency, high tolerance signal transfer. First, necessary background information will be provided on the areas needed to fully understand the design of both couplers, including waveguides, optical fibers, refractive index, sources of loss, the simulation methods used, and the methods of optical coupling used. Following this, the research done on fiber-to-chip and chip-to-chip coupling design will be presented in the following order: past designs of similar structures, novel additions and modifications made to those designs, results of the performance after the changes are implemented, and a comparison to other state of the art couplers in terms of a specific number of performance metrics. The hope is that the reader will be presented the background to understand the final results, the reasoning behind certain design considerations, and then the context for how this coupler fits into the plethora of ongoing study into optical coupling techniques.

Chapter 2

Waveguides

Waveguides are optical components that direct the propagation of light on chip similar to how metallic wiring, such as Cu or Al, guides electrical signals. In the context of this work, waveguides will reference on-chip structures. They can be fabricated from a multitude of materials including pure silicon, silicon compounds such as silicon nitride or silicon germanium, or other materials such as polymers. There are multiple types of waveguides, but this study will focus on channel or strip waveguides. Channel or strip waveguides consist of a dielectric core and a lower refractive index cladding, where light is confined to the core by total internal reflection similar to optical fibers (which are discussed in Section 2.4).

2.1 Optical Modes

When light enters a waveguide from another structure, such as from an optical fiber or a laser, it can excite a particular transverse electromagnetic field distribution. When the normalized transverse distribution does not change along the propagation direction in the structure, the field distribution is called a mode. Depending on waveguide's cross-sectional dimensions, the core to cladding refractive index contrast, and the wavelength of the propagating light, a waveguide can support anywhere from one to ten or more modes. In the latter case, the waveguide is said to be "multi-moded" and when light enters the core, a linear combination of its modes will be excited. However, each mode in a multimode waveguide has its own separate phase and group velocity. The group velocity of light dictates how fast a packet of light can travel, i.e. how fast the mode can travel in the waveguide. As a result, the multi-moded field will change as it propagates, or otherwise known as attenuation, due to the cumulative difference between the phases of all the modes and due to modal dispersion (the difference in group velocities of the modes). In order to avoid these effects, it is desirable to design the waveguide such that only the first or fundamental mode can propagate - this removes modal dispersion from the picture so the normalized transverse field is unchanging in space and time. These waveguides are termed single mode (SM) waveguides.

Mathematically, modes are governed by the Helmholtz equation. The Helmholtz equation stems from Maxwell equations without sources in an isotropic media. Maxwell's equations are a set of four equations that describe electromagnetic radiation waves. These equations were first presented by James Maxwell in 1864 [13] and today are commonly written in differential form as:

$$\nabla \times H = \frac{\partial}{\partial t} D + J \tag{2.1}$$

$$\nabla \times E = -\frac{\partial}{\partial t} B \tag{2.2}$$

$$\nabla \cdot D = \rho \tag{2.3}$$

$$\nabla \cdot B = 0 \tag{2.4}$$

with the so called constitutive relations as follows:

$$J = \sigma E \tag{2.5}$$

$$D = \varepsilon E \tag{2.6}$$

$$H = \frac{B}{\mu} \tag{2.7}$$

where E and H are the electric and magnetic field strengths respectively, D is the electric displacement, B is the magnetic flux density, J and ρ are the current and charge density, ε is the material permittivity, and μ is the permeability [14]. In general, E,B,H,D, and J are vectors while ε and μ are tensors. However, since we are working in photonics, we are only interested in source-free situations where $\rho = 0$ and $J = 0$. The equations simplify further if our material system is isotropic, meaning ε and μ are scalars in Equations 2.6 and 2.7. In the end, we are left with the source-free Maxwell equations in isotropic media:

$$\nabla \times H = \frac{\partial}{\partial t}(\varepsilon E) \tag{2.8}$$

$$\nabla \times E = \frac{\partial}{\partial t}(\mu H) \tag{2.9}$$

$$\nabla \cdot E = 0 \tag{2.10}$$

$$\nabla \cdot H = 0 \tag{2.11}$$

By separating the fields into their components to obtain scalar partial differential equations, we can solve for E_x through elimination of H_x, H_y, and H_z:

$$0 = \left[\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) - \mu\varepsilon\frac{\partial^2}{\partial t^2}\right] E_x = \left(\nabla^2 - \mu\varepsilon\frac{\partial^2}{\partial t^2}\right) E_x \tag{2.12}$$

Doing likewise with E_y and E_z leaves us with three equations of the same form as Equation 2.12, which can be combined to give:

$$0 = \left(\nabla^2 - \mu\varepsilon\frac{\partial^2}{\partial t^2}\right) E \tag{2.13}$$

The resulting wave equation signifies that E behaves as a wave with velocity $v = (\mu\varepsilon)^{-\frac{1}{2}} = \frac{c_0}{n}$. The same equation can be derived for H and shows the quantity to be a wave with the same properties as E. Assuming z as the propagation direction, we can choose E such that $E_{xy}(z,t) = \tilde{A}(x,y)e^{\beta z}e^{-\omega t}$ where $\tilde{A}(x,y)$ is the transverse electric field amplitude. In this case, then Equation 2.13 simplifies to:

$$\left[\nabla^2 + k_0^2 n^2 - \beta^2\right] \tilde{A} = 0 \tag{2.14}$$

where

$$\nabla_{x,y}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2.15}$$

$$k_0 = \frac{2\pi}{\lambda_0} \tag{2.16}$$

$$\beta = n_{eff} k_0 \tag{2.17}$$

and n is the refractive index as a function of position, β is the propagation constant, n_{eff} is the effective refractive index, k_0 is the free space wavenumber, λ_0 is the free space wavelength, and $\tilde{A}$ is the electric field transverse spatial amplitude. The eigenvalues of this equation are the values of β allowed for a specific spatially distribution of n, while the eigenfunctions of this equation are the amplitudes $\tilde{A}$ for each β. This is analogous to quantum mechanical particle governed by a Schrödinger equation, where the β is analogous to the allowed energies and $\tilde{A}$ is analogous to the wavefunction. Each value of β corresponds to a specific n_{eff}, an effective refractive index, which describes the refractive index actually dictating the propagation of the guided mode. The effective refractive index occurs because the mode propagates in both the core and the adjacent cladding, thus there is optical power in both the core and the cladding. The effective refractive index is not a simple weighted average of the core and the cladding indices; however, an approximation is possible by looking at the energy of the propagating mode. The energy of a mode is determined by a term in the Poynting theorem, which is conservation of energy for light. From the Poynting theorem terms, the time averaged energy of the mode is:

$$\langle U_{em} \rangle = \int_T \oint_S (E \times H)\, dS dt$$

where S is an area on a plane perpendicular to the propagation axis, E is the electric field, H is the magnetic field strength, T is one period of the light wave, and U_{em} is the energy of the electromagnetic mode. Higher order modes are more energetic. By limiting the integration to the core area, we can calculate P_{core}, the power in the core; likewise, we can also find the power in the cladding, P_{clad} by restricting the integration to the cladding region. P_{core} + P_{clad} should yield the total power (or energy) of the mode. The power distribution of the mode can be used to calculate its effective index, n_{eff} , approximated by

$$n_{eff} = \frac{\mathrm{P}_{core} n_{core} + \mathrm{P}_{clad} n_{clad}}{\mathrm{P}_{total}}$$

where the power in the core and power in the cladding for a cross sectional waveguide can be quickly calculated using numerical software. Moreover, the effective index of a guided mode is extremely important in photonics. For one, they can provide a qualitative understanding of whether or not a given mode will propagate in a waveguide of a given material and dimension. For instance, if the effective index of the mode is greater than the cladding index, then the mode will propagate as a guided mode in the waveguide. If the effective index of the mode is less than the cladding index, then the mode will become a radiation mode and attenuate out, not being able to propagate through the waveguide for any appreciable distance. This will be one type of method used to determine the maximum width allowed for a given material and waveguide thickness since it is easy to compute effective indices using Lumerical's finite domain eigensolver (FDE) tool. One last note on Maxwell's equations and optical modes is that boundary conditions dictate that E and H in the cladding decay exponentially as we move away from the core, perpendicular to the propagation axis. These are the evanescent tails of the mode, and will be important for our evanescent chip-to-chip coupler depicted in Chapter 6.

2.2 Refractive Index

In order to properly design an optical coupler, we must be able to use all of the potential tools (i.e. design parameters) at our disposal. One of those crucial tools is the refractive index of the material, which in a basic sense is a measure of the retardation of the light for a given material. The background presented here follows the notation of Feynman in [15]. The fundamental basis for the refractive index in a material is the interaction of the electromagnetic fields of the incoming wave (the light) and the electric charges in the material (mainly electrons). On an atomic scale, the electrons act like oscillators on springs, and when an external electromagnetic field is applied, the system acts like a damped harmonic oscillator with a driving force proportional to the electric field of the light:

$$m\left(\frac{d^2x}{dt^2}+\omega_0^2 x\right)=F \tag{2.18}$$

$$F=q_eE_s=q_eE_0e^{i\omega t} \tag{2.19}$$

This differential equation can be solved, and the solutions for position can be used to determine the electric field response from the electrons in the material from the known relationship for the electric field at a point due to an oscillating plane of charge.

$$x=\frac{q_eE_0}{m(\omega_0^2-\omega^2)}e^{i\omega t} \tag{2.20}$$

$$E_a=\frac{-\eta q_e}{2\epsilon_0 c}\left[\frac{dx}{dt}\right]_{t=\ t-z/c} \tag{2.21}$$

$$E_a=-\frac{\eta q_e}{2\epsilon_0 c}\left(\frac{i\omega q_eE_0}{m(\omega_0^2-\omega^2)}e^{i\omega(t-\frac{z}{c})}\right) \tag{2.22}$$

From experiment, we know that light propagating in a material travels slower than the speed of light in a vacuum. Therefore, we can assume traveling through a material will cause some additional time delay, resulting in a phase shift in the electric field. By Taylor expanding the phase shift for a small material thickness, we can obtain a term for the electric field contribution from all the electrons in the material.

$$E_{\text{after material}} = e^{-i\omega(n-1)\frac{\Delta z}{c}} E_0 e^{i\omega(t-\frac{z}{c})} \tag{2.23}$$

$$e^{-i\omega(n-1)\frac{\Delta z}{c}} \approx 1 - i\omega(n-1)\frac{\Delta z}{c} \tag{2.24}$$

$$E_{\text{after material}} = \underbrace{E_0 e^{i\omega(t-\frac{z}{c})}}_{E_s} - \underbrace{i\omega(n-1)\frac{\Delta z}{c} E_0 e^{i\omega(t-\frac{z}{c})}}_{E_a} \tag{2.25}$$

Comparing this result to the result from the driven oscillator differential equation, we see that the refractive index is a function of the number of atoms per unit volume of the material (the atomic concentration) N and the frequency of the light ω.

$$n = 1 + \frac{Nq_e^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)}$$

Note that empirical relationships can also be determined relating the refractive index to wavelength for different compounds, called Sellmeier equatons, and these equations are typically used to determine the refractive index of a complex material for use in a device (such a coupler). Also note that another way of writing this result is in terms of the materials susceptibility, χ. Thus, by changing the materials atomic concentration, or susceptibility, it is possible to tune the refractive index for a given frequency. In the case of our couplers, we will be focusing tuning the refractive index of $\text{Si}_x\text{O}_y\text{N}_z$ by altering the oxygen content of the layer. A chart showing how varying the oxygen, nitrogen, and silicon content in $\text{Si}_x\text{O}_y\text{N}_z$ can vary properties such as thermoelectric insulation, refractive index, and band gap is shown in Figure 2.1 [16].

The result that refractive index is a function of material parameters such as is important for our fiber-to-chip and chip-to-chip couplers because we will tune the refractive index of the material by changing the material's composition, as will be described in Section 2.2.1.

In Lumerical's simulation software, dispersive materials with tabulated refractive index (n, k) data as a function of wavelength are incorporated by using a multi-coefficient material

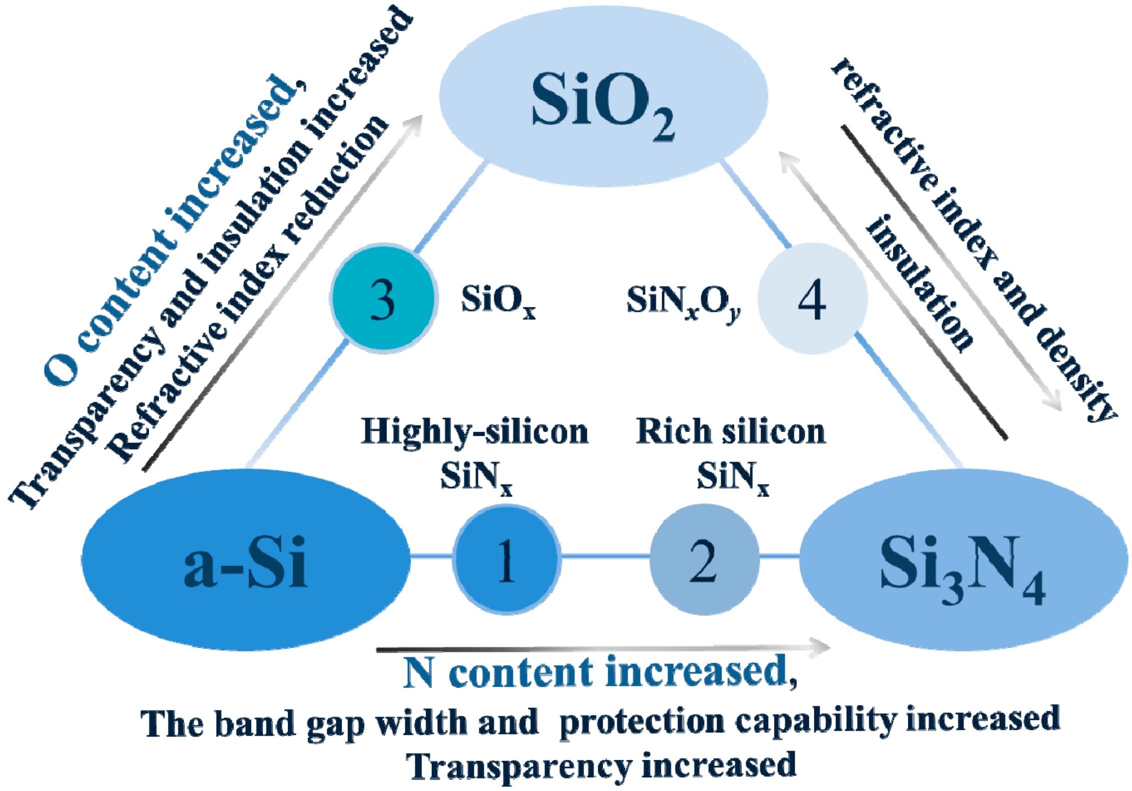


Figure 2.1: A chart showing how changing oxygen, nitrogen, or silicon content alters properties such as the refractive index, thermoelectric insulation, and band gap. Image reused from [16] without changes. This is qualitative to yield a general idea for how we will tune the refractive index of $Si_xO_yN_z$ by controlling its oxygen content and develop a highly efficient coupler.

model. Alternatively, specific models such as the plasma (Drude), Debye or Lorentz can be used, although we will not make use of those types of models in this report.

2.2.1 GRaded INdex (GRIN) Structures

A graded index (GRIN) material has a refractive index that varies along an axis. In the case of ray optics, this position dependent refractive index is continuous, often fabricated by adding impurities (dopants) of controlled concentrations. An example is in continuously graded optical fibers, where the core of the fiber is doped with an index raising agent such as germania (GeO_2), phosphorus pentoxide (P_2O_5), or alumina (Al_2O_3) or the cladding is doped with index lowering agents such as fluorine or boron oxide (B_2O_3). Similarly, because continuously graded index profiles require a higher fabrication complexity, step graded index profiles are often used as well. In a step GRIN system, regions of uniform, but different, refractive index are placed in contact with one another to achieve the desired GRIN profile. For our fiber-to-chip coupler, we use a step GRIN system with the number of layers and the refractive index of each layer tailored to fit a specific refractive index profile. The GRIN is achieved in our case by depositing $\mathrm{Si}_x\mathrm{O}_y\mathrm{N}_z$ layers with varying levels of oxygen. Thus, the refractive index is engineered to near near SiO_2 on the bottom with a maximum of SiN ($n = 2.1$) on the top (that is, if the film is deposited with no or very little oxygen content). This is accomplished through close control of the oxygen content during deposition using low temperature plasma enhanced chemical vapor deposition (PECVD) by varying the flow of SiH_4 and $\mathrm{N}_2\mathrm{O}$ in the chamber [17]. By appropriate choice of the refractive index profile, a GRIN system can have the same effect on light as a conventional optical component, such as a prism or a lens. The specific refractive index profile we are interested in is the case of a index profile which is described by the following equation:

$$n^2(y) = n_0^2(1 - m^2y^2)$$

where n_0 is the refractive index at $y = 0$, m is a constant which describes the stretching of the index profile, and y is the vertical distance from the propagation axis. Usually, m is chosen to be sufficiently small such that $m^2y^2 << 1$. If this assumption is true, we can do a Taylor expansion about $y = 0$ and the following approximation holds:

$$n^2(y) = \sqrt{n_0^2(1 - m^2y^2)} \approx n_0(1 - \frac{1}{2}m^2y^2) \tag{2.26}$$

which describes $n(y)$ as a parabolic distribution. To describe how this index distribution might affect propagation of light, let's consider ray optics and the paraxial ray equation for a GRIN slab which governs the trajectories of light rays inside the slab [18]:

$$\frac{d^2}{dz^2} = \frac{1}{n(y)}\frac{dn(y)}{dy} \tag{2.27}$$

In Equation 2.26, we assume that $n(y) - n_0 << n_0$ so that the fractional change of the refractive index is very small, take the derivative with respect to y and then solve Equation 2.27 to obtain the following solutions:

$$y(z) = y_0 \cos mz + \frac{\theta_0}{m}\sin mz \tag{2.28}$$

where $y_0 = y(0)$ and $\theta_0 = \frac{dy}{dz}\Big|_{z=0}$. The result are oscillatory functions with a period of $\frac{2\pi}{m}$. If we focus our attention to a single period, we can see how the GRIN slab can be used as a lens. This concept was reviewed in the context of ray optics, but a similar argument can be made in terms of wave optics as well. In wave optics, as described above, we consider an optical mode with a transverse electromagnetic field distribution governed by Equation 2.12. By inserting the parabolic index profile of Equation 2.26 into Equation 2.12 and solving for the mode profiles, we can obtain solutions which are Hermite-Gaussian functions of the form [19]:

$$M_\nu(y) = H_\nu\left(y\sqrt{k_0 n_0 m}\right)\exp\left(-\frac{y^2 k_0 n_0 m}{2}\right)$$

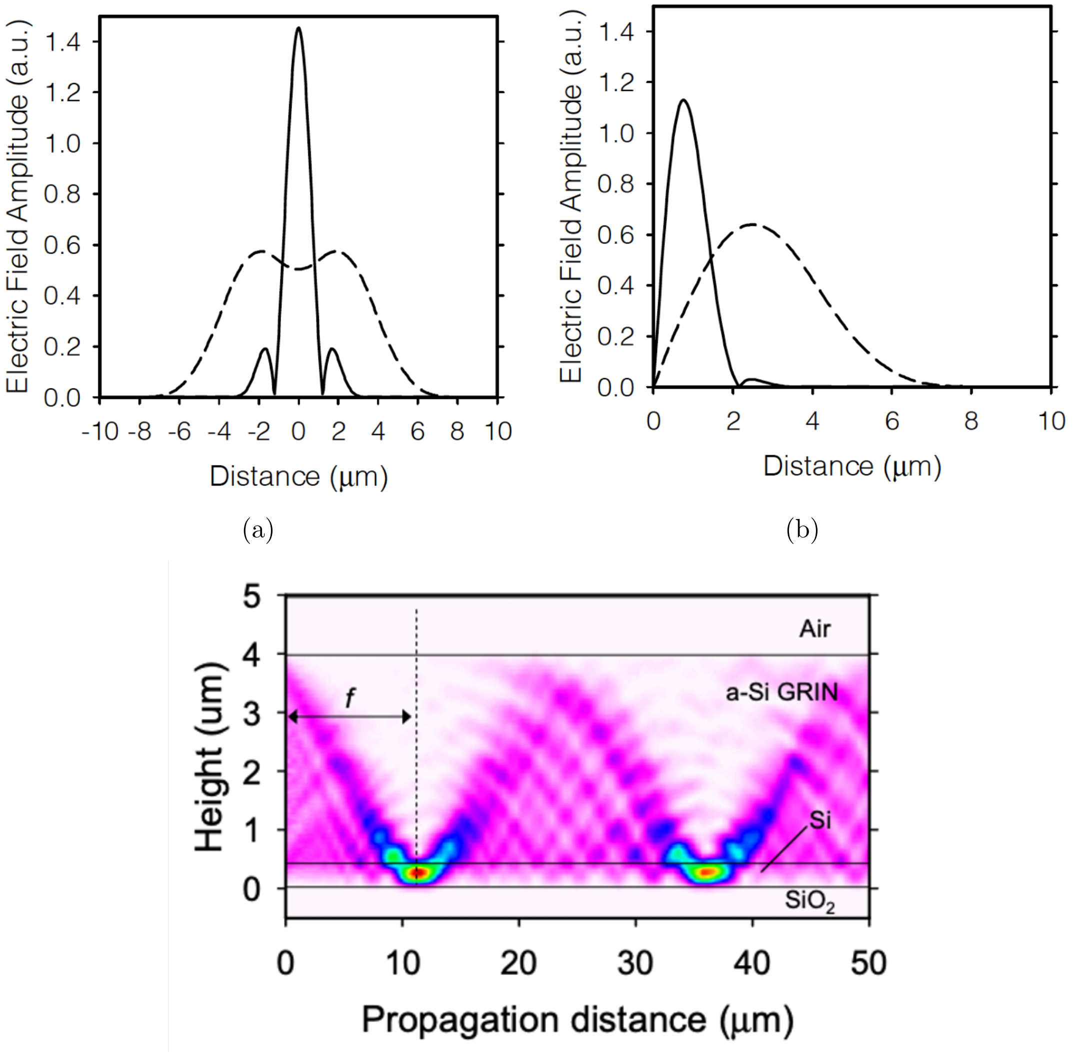
with propagation constants given by:

$$\beta_\nu = k_0 n_0\sqrt{1 - \frac{m(2\nu+1)}{k_0 n_0}}$$

and a total guided field governed by a superposition of these mode profiles

$$E(y,z) = \sum_\nu c_\nu M_\nu(y)\text{exp}\left(-i\beta_\nu z\right) \tag{2.29}$$

where c_ν are the expansion coefficients determined by initial conditions (i.e. the input field launched into the structure), while the term $\exp(-i\beta_\nu z)$ gives the phase evolution of the field along the propagation direction. For a symmetric GRIN structure excited by a symmetric input field, the even (symmetric) modes in Equation 2.29 are excited and $c_\nu = 0$ for odd values of ν. If the combined expansion coefficients and phase term, $c_\nu \exp(-i\beta_\nu z)$, in Equation 2.29 are all real and positive, then the mode superposition generates a narrow field distribution. On the other hand, when $c_\nu \exp(-i\beta_\nu z)$ for each successive mode differs by π (i.e. successive terms alternate in sign) a broad field distribution is generated. For the lower order modes where $\nu << \frac{k_0 n_0}{m}$, the total field profile will periodically oscillate between the these two distributions with a period $\frac{\pi}{m}$, yielding an effective lens focal length of $\frac{\pi}{2m}$. If we assume that $\theta_0 \approx 0$ in Equation 2.28 and the first term dominates (which is reasonable), then the effective focal length according to the ray model and wave model are equivalent [20]. The same logic can be applied to the case of an asymmetric GRIN profile as well, resulting in an oscillation between a narrow and broad electric field distribution. This oscillation can be visualized using Figure 2.2 (all plots were reused from [19] which shows the oscillation for a symmetric GRIN, an asymmetric GRIN, and a cross sectional view of a GRIN slab to see how this might be applied for optical coupling.



(a)

(b)

(c)

Figure 2.2: The oscillation of the electric field distribution in a rectangular block of graded index material. In (a) the field oscillation is shown for a symmetric GRIN material while in (b) it is shown for an asymmetric GRIN material. Note that the field oscillates between a narrow distribution shown in the solid line for both cases and a broad distribution shown in the dotted line for both cases. As will be seen in later sections, the narrow distribution will be essential for coupling light into the narrow waveguides on the chip while the broad distribution is important for mode size and shape matching the large 10.4 μm MFD of the SMF at the input for our fiber-to-chip coupler. All images are from [19].

2.3 Adiabatic Tapers

Thus far, it has been assumed that optical modes travel in strip waveguides and the optical modes are normal modes, meaning they are orthogonal to one another. Mathematically, this means that the transverse electric field amplitude described by Equation 2.14 follows the constraint below:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{A}_i(x,y)\tilde{A}_j^*(x,y)dxdy = 0$$

However, in waveguides which vary their geometry in the direction of propagation, this relationship no longer holds and the normal modes we are accustomed to no longer exist. Instead, we need to discuss modes as “locally” normal modes. In essence, this means that if the waveguide is tapered, then the dimensions at some point z_0 along the propagation direction are used to solve for the normal modes of the structure at that point as if the structure did not vary with z. The solutions are then the local normal modes of the waveguide structure which is varying in reality. At a waveguide position z_0, these normal mode solutions are then said to be the local normal modes of the actual varying waveguide structure. By solving for the local normal modes along the propagation direction, the local normal mode representation can become a function of z and they can be used to determine modal evolution in the taper.

However, analytically determining modal evolution in the taper is complex, so a qualitative description will be presented. The complexity is because, in contrast to the normal modes of a fixed structure, the amplitudes of the local normal modes in a varying waveguide structure are not necessarily constant. This means that power transfer between the local normal modes of a varying structure occurs, and the magnitude of that power transfer depends on the rate of change of the geometry of the structure [21]. With that being said, by defining a transition between two waveguide structures which is extremely slow, i.e. it takes place gradually with propagation distance z, the power transfer between local normal modes can be made negligible. This is highly desirable because power transfer away from the fundamental local normal mode will result in loss during the transition process once the final output is reached and propagation returns to single mode conditions. This type of transition is referred to as an adiabatic transition or an adiabatic tapering. In other words, in an adiabatic taper, power injected initially in a given local normal mode will stay in that mode throughout the transition (i.e. the tapering process). During the transition, the local normal mode may change its shape, but coupling to the other local normal modes is assumed not to occur, with power put into the first-order local normal mode ending up in the first-order local normal mode, for example. The discussion here uses the phrase transition liberally to reference a taper connecting two waveguides and converting a mode in size and shape from one waveguide to another, or a set of tapers near one another to couple light from one taper to another taper. The latter of these situations is of interest for our chip-to-chip coupler and will be fleshed out in later chapters.

Guidelines for best practices in design of high efficiency tapers were provided in [22] and these guidelines have been used to develop adiabatic tapers for optical fiber-to-chip coupling and chip-to-chip coupling in [23], showing their validity experimentally. The guidelines relied on a length-scale and a weak power transfer criterion to compute the minimum taper length required to avoid power transfer from the fundamental order mode to the next, unwanted local normal mode. The rule of thumb for the length-scale of the transition was that it must be much larger than the coupling length between the mode of interest and the closest mode in terms of the propagation constant. Mathematically, this adiabatic criterion equates to the following relation (assuming a smooth and uniform taper of length L):

$$L >> \frac{2\pi}{\beta_0 - \beta_1} \tag{2.30}$$

where β_0 and β_1 are the respective propagation constants of the mode of interest (the fundamental mode for this report) and the closest mode the system could couple to. Equation 2.30 represents an important subtlety - it says that geometrical changes of the structure must be gradual to ensure no power transfer but compared to other approaches, such as conventional directional couplers without tapers, does not require a critical length for 100% transfer. One of the results of this is that adiabatic devices generally are not sensitive to changes in wavelength and do not exhibit large polarization dependent losses. In addition, this comes with the added benefit that they typically do not display significant mode beating characteristics like directional couplers. Adiabatic tapers will be essential to the design of a highly efficient chip-to-chip coupler to be discussed in the following chapters.

2.4 Optical Fibers

Now that we have covered basic photonic notions like waveguides, optical modes, and refractive index, we can discuss a crucial component for large scale data networking - the optical fiber. An optical fiber guides light via total internal reflection, similar to the waveguide, but is responsible for the majority of off chip light routing over longer distances. Similar to waveguides, there are both single mode and multi mode fibers, with the standard fiber for telecom and datacom applications being single moded in order to lower attenuation and avoid modal dispersion. There are also polarization maintaining fibers (PMFs), a subset of SMFs which keep linear polarized input light as linear polarized light upon output from the waveguide after propagation with little cross-coupling of optical power between the two modes. This is important because in a typical SMF, the polarization state is randomly oriented upon output; however, in the SM waveguide the polarization of each mode causes the mode to take a defined orientation. Compared to multi-mode fibers, SMFs allow communication distances to be significantly longer. For example, the attenuation coefficient for a SMF28 ultra optical fiber, a measure of loss per unit length for a fiber system, is lower than 0.18 and 0.32 dB/km at wavelengths of 1.55 and 1.31 μm, respectively. This is compared to a multi mode fiber at 1.31 μm (multimode fibers are not designed to operate at 1.55 μm due to high attenuation) which has an attenuation coefficient between 0.8 and 1.4 dB/km. SMFs typically contain a 125 μm cladding layer encapsulating a core 6-10 μm in diameter [24]. The SiO_2 cladding layer has an index of refraction of $n_{\mathrm{clad}} = 1.45$ while the fiber core has a refractive index $n_{\mathrm{core}} = n_{\mathrm{clad}} + 0.03n_{\mathrm{clad}}$, a value just 0.3% higher than SiO_2 and a very low contrast configuration. For example, for a SMF with an 8.2 μm core diameter, the resulting MFDs are 10.4 and 9.2 μm at wavelengths of 1.55 and 1.31 μm respectively. For our fiber-to-chip coupler, we will always use SMFs because of their fundamental role in the future of datacom and telecom applications.

2.5 Sources of Loss

In order to master the transmission of light, we must understand why light attenuates in a medium or during a coupling transition. There are three main sources of loss when light travels in waveguides or in optical fibers: roughness scattering, material attenuation and optical leakage (as a result of imperfect waveguiding). Roughness scattering includes line edge roughness due to imperfect lithography and pattern transfer for waveguides and frozen-in "surface capillary waves" for optical fibers (which will not be described here). Note that high index contrast systems suffer from high scattering losses, seen by the approximation $\alpha_{wg} \propto (n_{core}^2 - n_{clad}^2)\,\sigma^2$ where σ is the RMS roughness [25]. Therefore, our material system incorporates lower index contrast configurations for the coupling elements which do not exhibit any bending and are extremely planar. This will reduce roughness scattering losses within the coupling element, while allowing for higher index materials to be used for signal routing where bending and higher confinement is necessary. Material attenuation is a result of different types of absorption - electronic, vibrational, impurities, free carriers - or, for glasses, Rayleigh scattering. To avoid material attenuation, 1.55 μm and 1.31 μm wavelengths are chosen for telecom and datacom, respectively, as these wavelengths represent loss minimums for light propagating in silica optical fibers. As a result, this work will focus on developing optical coupling devices for 1.55 μm. Optical leakage refers to bending loss or leakage to the substrate such as bulk silicon in the case of SOI waveguides. The losses associated with bending present an interesting challenge for chip-to-chip optical coupling, where light must change planes vertically without using bends of high curvature. Our coupler designs do not exhibit bending and, from the instant light enters the chip from the optical fiber, take care to vertically guide the mode towards the surface and confine it there (as opposed to vertically down towards the substrate as in prior fiber-to-chip coupling designs. This will be discussed more in Section 5.1.2).

Note that these sources of loss occur during propagation of the light. When light is being coupled between two components, such as from an optical fiber or to another chip, additional sources of loss appear due to reflections, polarization dependent losses, and mode size and shape mismatch, and. Reflections appear at interfaces where abrupt changes in refractive index cause the light to reflect or scatter, inducing loss into the system. This is especially important for our fiber-to-chip coupler, where the optical fiber's refractive index close to that of silica often is significantly different from waveguides on the chip, such as silicon with an index of 3.48 at $\lambda = 1.55\mu$m. As for polarization dependent loss, this occurs when the coupler has a decrease in coupling efficiency for TE polarized light versus TM polarized light (assuming all components only support single mode propagation, as is the case for the optical fibers and waveguides in this work). This is because the two polarizations have different spatial distributions for the electric field, which may cause the coupling to be less efficient depending on the coupling method. Coupling via surface gratings suffers from this type of loss in particular because the grating period is fundamentally related to the wavelength of light, and the polarization is coupled to wavelength (this connection is discussed further in Section 3.1. This is particularly important in the context of datacom/ telecom applications, because at low data rates, PDL is a minor contributor to loss; however, at 10 Gbps and above, PDL becomes comparable to insertion loss. Therefore, showing low PDL at 1.55 μm will be necessary for our couplers.

The next source of coupling loss, mode size and shape mismatch, is very important and represents the key challenge to optical fiber-to-chip coupling. This challenge is encapsulated in Figure 2.3, which is modified from [26].

Using this diagram, it is clear the size and shape of the SMF mode is around 10 μm and circular while the size and shape of a SM SOI waveguide mode is 220 nm by 500 nm and

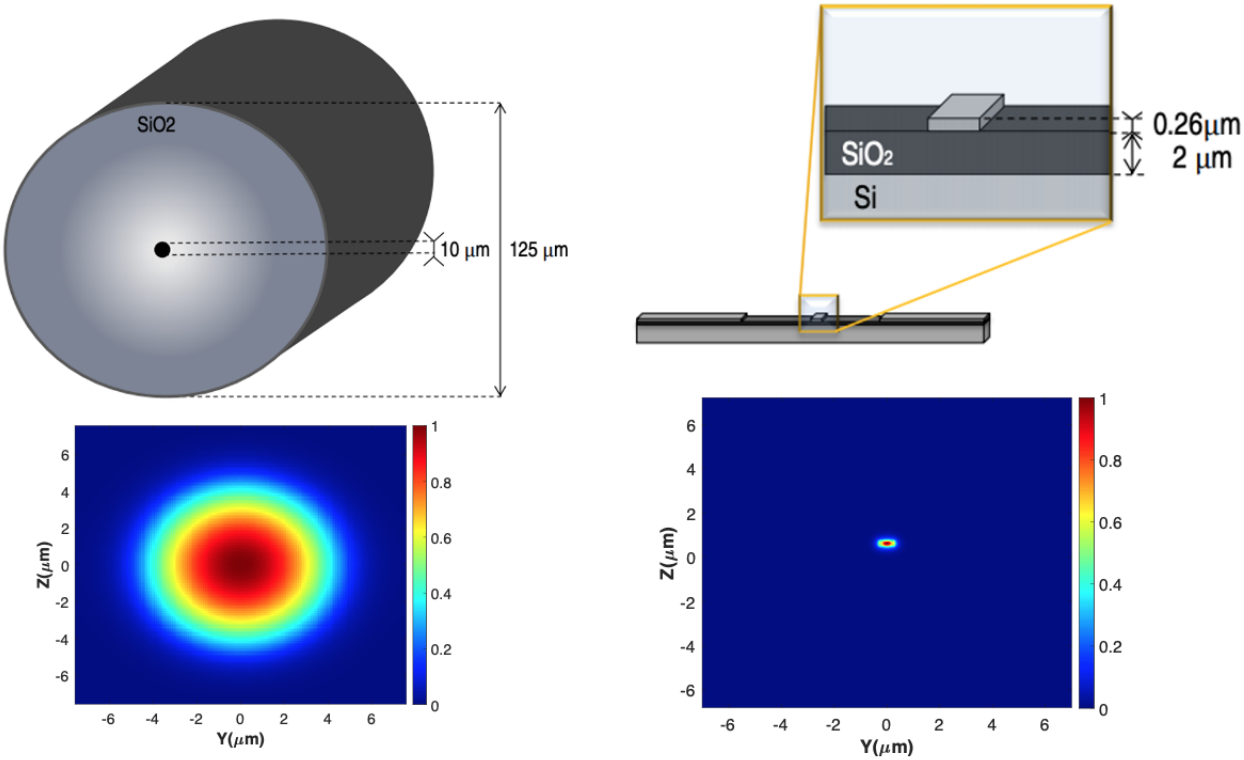


Figure 2.3: Modal size and shape mismatch between on chip SOI waveguides and a SM fiber with core diameter of 10 μm, modified from [26]

elliptical. The basis of optical coupling between two structures is modal overlap - the higher the amount of overlap between the size, shape, and polarization of the allowed modes in the SM waveguide and the SMF, the lower the losses due to modal transformation between SMF and SM waveguide. Modal overlap is characterized by Equation 2.31, which represents the similarity in modal expansion coefficients and thus coupling efficiency.

$$\eta = \frac{\int E_1 E_2 dS}{\int |E_1|^2 dS \int |E_2|^2 dS} \tag{2.31}$$

where η represents modal overlap, and E_1 and E_2 are the transverse electric field profiles in the SMF and SM waveguide, respectively. Note that E_2 is the modal distribution at the input of the light to the waveguide. Due to the significant size and shape modal mismatch, the result is heavy coupling losses if no optical coupling device is employed. The loss can be approximated without doing the overlap integral by using Equation 2.32 from [27]. A simple calculation for a SMF with a MFD of 10 μm butt coupled directly to square a 0.9 μm x 0.9 μm $Si_xO_yN_z$ SM waveguide shows calculated losses on the order of 15 dB, which is 3% coupling efficiency.

$$\text{loss (dB)} = -10 \log \left[\frac{4}{\left(\frac{MFD_1}{MFD_2} + \frac{MFD_2}{MFD_1} \right)^2} \right] \tag{2.32}$$

In order to mode match, and reduce reflection, vertical and lateral confinement of the mode into the waveguide is necessary. This requires a separate device between the SM waveguide and the SMF which can transform the mode - the optical coupler.

Chapter 3

Optical Couplers

3.1 Coupler Types

In general, there are four types of coupling schemes in widespread use today for optical fiber-to-chip coupling and chip-to-chip coupling, they are: edge coupling, grating coupling, evanescent coupling, and beam expander coupling designs. Note that in addition to these, other coupling schemes are under development and show promise, including photonic wire bonding, where two-photon polymerization is used to deposit a waveguide material connecting SMF to SM waveguide, or SM waveguide to SM waveguide on separate chips [28, 29]. This deposited material guides light between the two structures analogous to how metallic wire bonds conducts electric current between two separate chips. However, this method will not be discussed further here due to its relative fabrication complexity - our focus is on developing couplers which are mass manufacturable utilizing current CMOS microelectronics fabrication facilities. Descriptions of each of the four coupling schemes are presented in order to understand why edge and evanescent coupling schemes were chosen for our design and also for future comparison in order to judge the performance of our design.

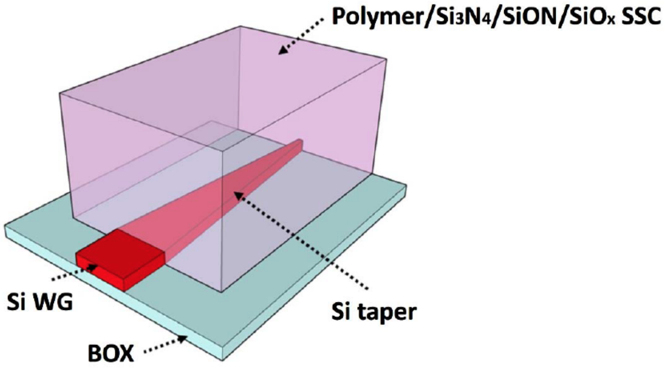
3.1.1 Edge Couplers

The first common coupling technique is edge or butt coupling. Edge couplers are located on the perimeter of the PIC, coming into contact or extremely close proximity to a SMF and delivering light in the same plane. They are able to do so by through minimizing reflection at the chip-fiber interface and matching the size and shape of the mode from the SMF to that of the SM waveguide at the interface. Numerous techniques have been employed to achieve this, including tapers with linear and nonlinear profiles, multi-tip tapers and trident taper structures, waveguide assists for vertical confinement, cascaded multi-stage tapers, 3D tapers, cantilever structures, and sub-wavelength gratings (which are not grating couplers, but a type of edge coupling scheme) [30, 31, 32, 33, 34, 35, 36]. These different edge coupler variants are summarized in Table 5.4, which was taken and modified from [37].

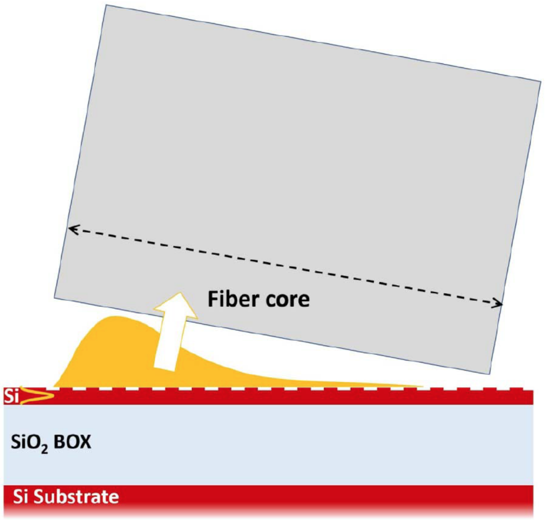
Edge couplers can achieve rather high coupling efficiency, broad bandwidth, and polarization independence. Their drawbacks include relatively larger footprints than grating couplers (adiabatic tapers can range from 500 μm to 1 mm), a fixed coupling position at the edge of the chip, and the need for high quality polishing of the facet (since they are located on the edge of the chip) which adds costs and hampers mass manufacturing. In addition, they tend to suffer from poorer packaging misalignment tolerances compared to grating couplers, typically requiring sub-micron alignment in order achieve less than 1 dB of excess loss. Packaging misalignment is a crucial performance metric when considering practical coupler performance and scalability, and is defined in Section 3.3.

3.1.2 Grating Couplers

The second type of fiber-to-chip coupler is the grating coupler. A grating coupler operates by a surface grating etched into the surface of the chip with a period on the order of the wavelength of light being coupled. This results in diffraction of light entering vertically from above, such as when a SMF is placed with the facet aiming downward or when the SMF is horizontally, but the facet of the SMF is cleaved such that internal reflection deflects the beam downwards. In either case, the resulting diffraction alters the wave-vector of the incoming light, or propagation direction of light, into the plane of the chip and vice versa. The light then undergoes mode spot size conversion to match the size of the coupled mode in the waveguide. Simple examples of an edge coupler, in the form an inverse taper, and a grating coupler are shown in Figure 3.1 for context. The performance of common grating couplers is shown in Table 3.1 taken from [37] with modifications.



(a)



(b)

Figure 3.1: Examples of fiber-to-chip couplers patterned on SOI substrates. In (a) an edge coupler and in (b) a grating coupler (both from [38]).

Grating couplers have some major advantages including compact size, wafer-level testing capability, and flexible coupling position (because they are not constrained to the edge of the chip or wafer); however, there are also drawbacks. These include a relatively low coupling efficiencies - the most important figure of merit for optical couplers - typically below 50%. In addition, they suffer from narrow bandwidth and high wavelength sensitivity. This stems from the fact that in order to ensure coupling into the guided modes of the waveguide, the grating period, duty cycle, fill factor, etc. must be directly tied to wavelength via the Bragg condition:

$$n_{\text{eff}} - n_1 \sin\theta_1 = \frac{\lambda}{\Lambda}$$

Grating Coupler Design	**Coupling Loss (dB)**	**Bandwidth**	**Polarization**	**Comments**	**Year and Reference**
Shallow etched grating	3.1	Approximately 40 nm @ 1 dB	TE	Standard 248 nm DUV[1] used with a shallow etch, low loss.	2013 [39]
Subwavelength grating	3.7	60 nm @ 1 dB	TM	Standard 193 nm DUV used, low loss, wide bandwidth, sensitive to fabrication.	2010 [40]
Grating with silicon overlay	1.6	80 nm @ 3 dB	TE	Standard 193 nm DUV plus amorphous silicon overlay, low loss.	2010 [41]
Polarization-splitting grating	4.2	\	TE and TM	EBL[2] with a shallow etch, low PDL[3], low loss.	2019 [42]
Bi-wavelength grating	7.1	35 nm @ 3 dB	TE and TM	Standard 248 nm DUV with a shallow etch, high loss.	2013 [43]

Table 3.1: Overview of Grating Coupler Configurations [37]

[1] DUV = deep ultraviolet patterning and lithography
[2] EBL = electron beam lithography
[3] PDL = polarization dependent loss

3.1.3 Free Form Couplers

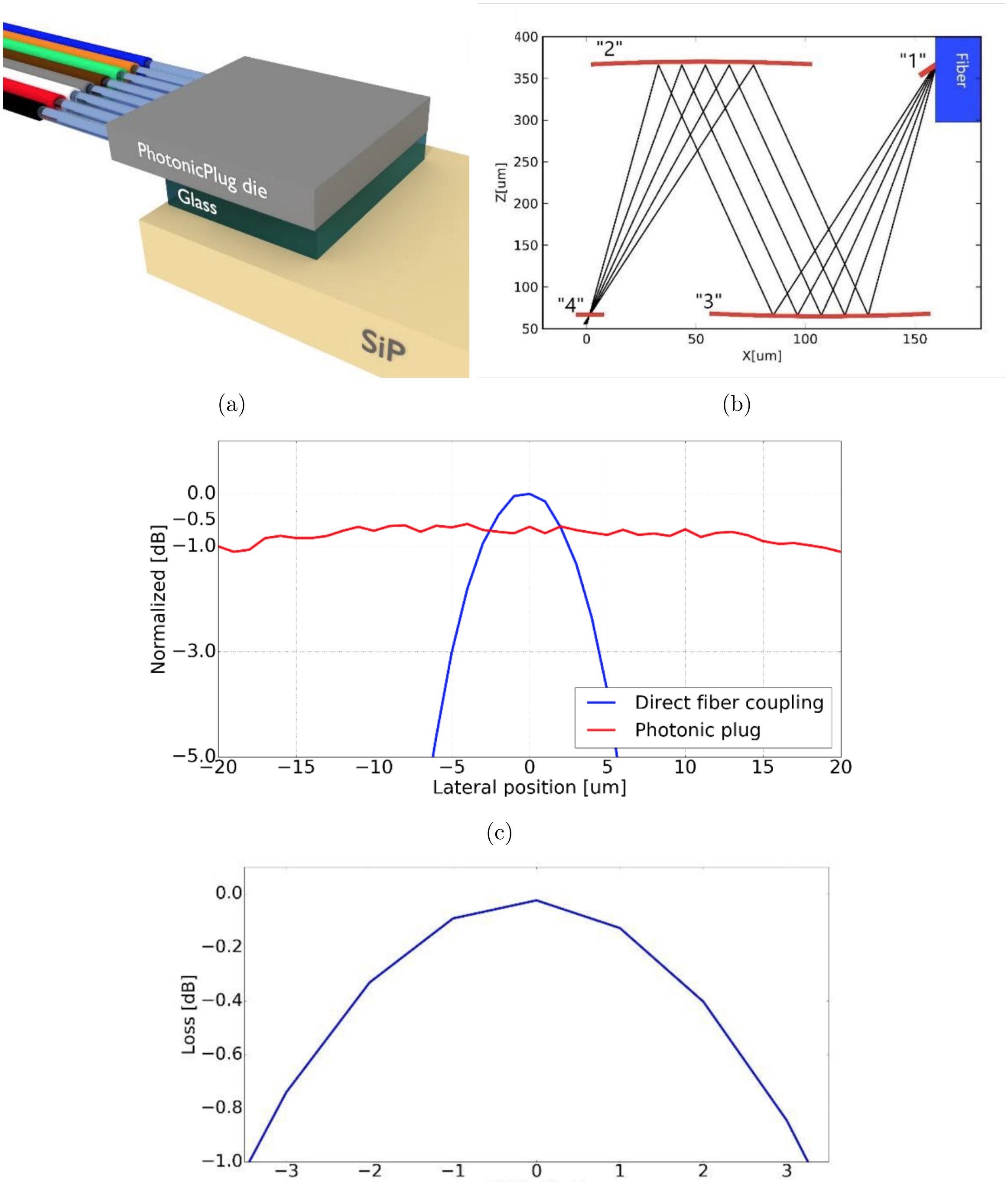
The third type of fiber-to-chip coupler is the free form coupler. In this work, "free form" means the use of micro-lenses and mirrors in order to expand the input beam and direct light through free space to couple into the SM waveguide. The basic premise of this beam expansion method is that by expanding the MFD of from the facet of the SMF by a low numerical aperture (NA) lens and then coupling into the SM waveguide using a high NA lens, alignment tolerances can be expanded significantly, as can footprint. The alignment tolerance is expanded by increasing the size of the mode significantly between the SMF and the SM waveguide, and then designing the coupler to achieve modal overlap for this increased MFD instead of the 10 μm MFD at the input facet which puts an upper limit on alignment tolerances of around $\pm 2.5\mu$m. The widened alignment tolerances of designs using free form couplers can be found in Figure 3.2, 3.3, or 3.4 from [44, 45, 46].

The drawbacks of these free form coupler designs include the use of curved structures in micro-lenses and mirrors, which are difficult to implement in mass manufacturing, and the use of diffraction as the basis for coupling which induces chromatic aberration blurring, an effect limiting the broadband capabilities of such a coupler. Broadband capabilities become important when discussing the use of such a structure for applications such as WDM which relies on multiple wavelengths being able to be coupled into a SM waveguide simultaneously. In addition, the free form examples quoted here also still include a grating coupler as the final step to couple light into the SM waveguide. In other words, the free form coupler is able to maintain a focus on the grating coupler for wider alignment tolerances, but still relies on a grating or micro-mirror to couple light into the SM waveguide, thereby acting like an additional tolerance expanding connector. However, this means any disadvantages of the grating coupler are present in the free form coupler, plus the additional fabrication complexity of the micro-lenses and micro-mirrors. Finally, the footprint advantage the grating coupler maintained no longer applies either, since the free space optics requires larger spaces to beam expand and focus.

3.1.4 Evanescent Couplers

The last type of coupler is the evanescent coupler. The basis of this style is the use of the evanescent tail of the modal distribution to transfer optical power from one optical waveguiding structure to another when the structures are placed side by side with one another (or overlaid on one another vertically - this is different from edge couplers which are adjoined end to end at a facet). For example, a conventional directional coupler with two adjacent, parallel waveguides is a well understood instance of evanescent coupling [47, 18]. An image of a directional coupler can be found in Figure 3.5 from [48]. According to [18], 100% power transfer is possible between the two waveguides, but requires zero phase mismatch between the light propagating in each channel (meaning the effective indices of the two waveguides are identical throughout the length of the waveguide) and the coupling length must be exactly equal to the specific transfer distance at which 100% power transfer is complete. Modification of this conventional directional coupler can allow for tailoring of the transfer distance as well as the gap between the two waveguides. The performance of a few current evanescent coupling designs or evanescent-edge coupler hybrids can be found in Table 6.2 and Table 5.5 later on with their respective sources.

Evanescent couplers are typically polarization dependent since the TE and TM modes have different distributions for the evanescent tail in the waveguide, meaning modal overlap will be different between modes in the two waveguides. Another important aspect of evanescent couplers are their alignment tolerances, which are often quite narrow as coupling occurs between waveguides (which have dimensions on the order of hundreds of nanometers, or single microns) and because the evanescent tail falls off exponentially from the edge of the waveguide core, meaning they typically must be in close proximity to one another. However, evanescent couplers have a relatively robust broadband nature similar to edge couplers, limited only by the material used (i.e. if a single mode condition is required in the waveguide, and the waveguide dimensions are a function of the wavelength of light via the refractive index of the specific material and Sellmeier equations, then the waveguide is the limiting feature of how far from 1.55 μm or 1.31 μm the wavelength can be while maintaining sufficient coupling efficiency). Traditionally, using a conventional directional coupler as a basis, evanescent couplers can have large footprints on the chip as large lengths are necessary for 100% power transfer between waveguides. Lastly, evanescent couplers have the ability to be located anywhere on the surface of the chip, meaning no fine polishing back end processing steps are necessary in fabrication (similar to the advantage of a grating coupler).

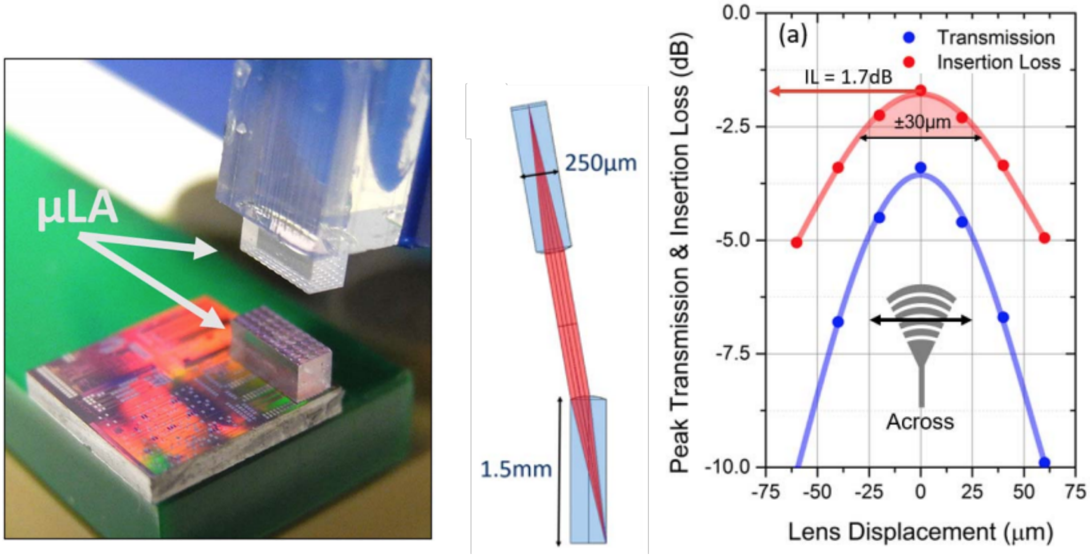


(a) (b)

(c)

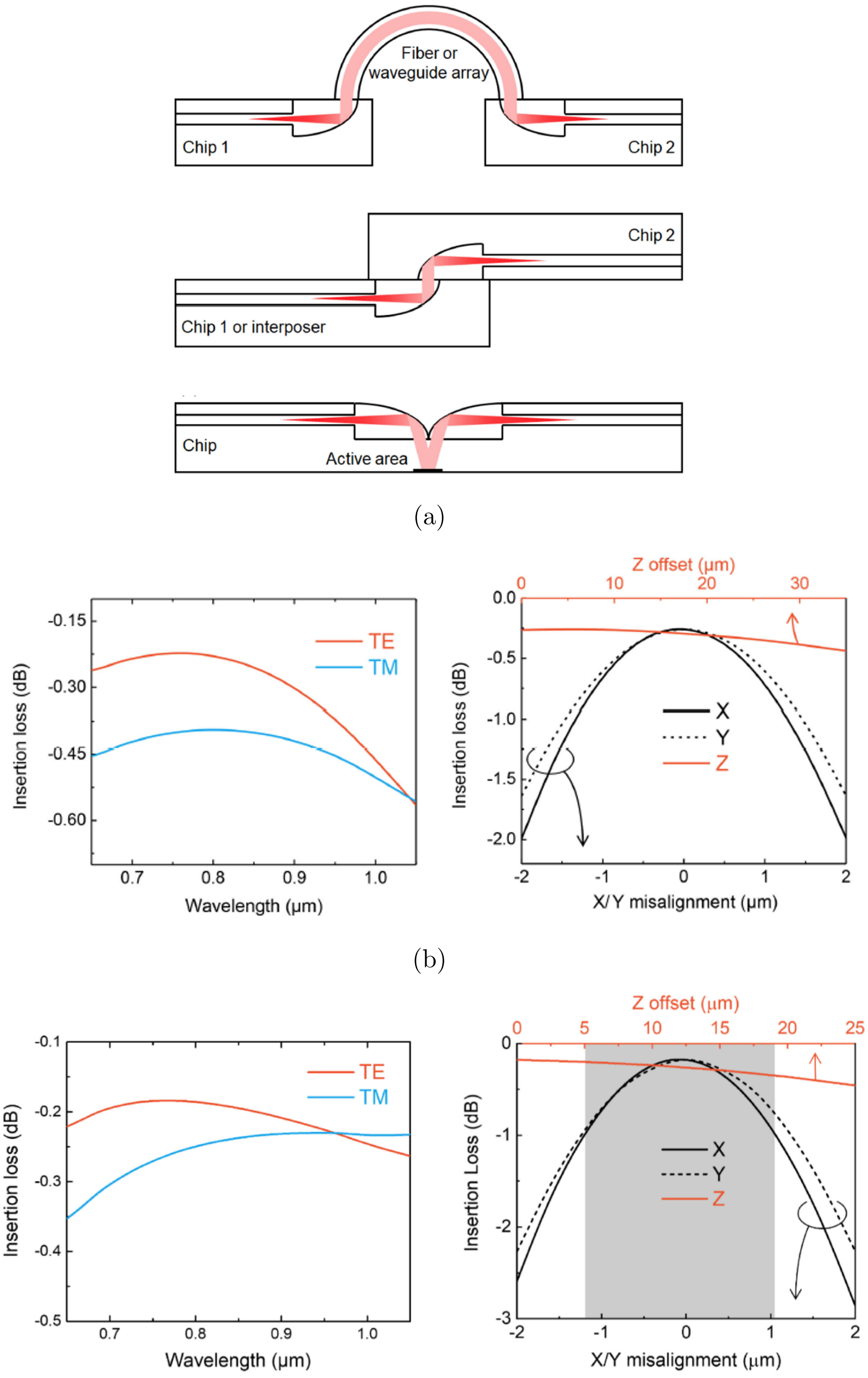
(d)

Figure 3.2: The Photonic-plug design. In (a), a schematic of the SMF array connected to the Photonic-plug which couples light to a Si-PIC without free space interaction through the use of a glass interposer piece. In (b), the method of coupling using mirror elements in "1"-"3" and either a grating coupler in "4" or another mirror element to coupling light into the SM waveguide. In (c) and (d) the lateral and rotational alignment tolerances of butt coupling versus free form coupling for Photonic-plug. All images are from [44].



(a) (b) (c)

Figure 3.3: Tyndall's free form pluggable coupler design. In (a), a picture of the experimental device with a SMF array connected micro-lenses expand the beam of light to then be focused by micro-lenses in an array on the Si-PIC. In (b), a schematic of the beam expansion during the free form coupling. In (c) the lateral alignment tolerances of the pluggable free form coupler - note that the alignment tolerances are wide enough to employ molded plastic to fit the fiber array and mechanically connect it to the Si-PIC using a LEGO piece. All images are from [45] ©2017 IEEE.



(a)

(b)

(c)

Figure 3.4: The optical free form couplers for high density integrated photonics (OFFCHIP) design. In (a), a schematic of the potential applications and uses of the coupler showing the potential for fiber-to-chip coupling, chip-to-chip coupling, or both. In (b), the misalignment tolerances and the wavelength dependency when the coupler is used to couple light chip-to-chp. In (c) the misalignment tolerances and wavelength dependency for when the coupler is used to couple light vertically from a SMF. All images are from [46].

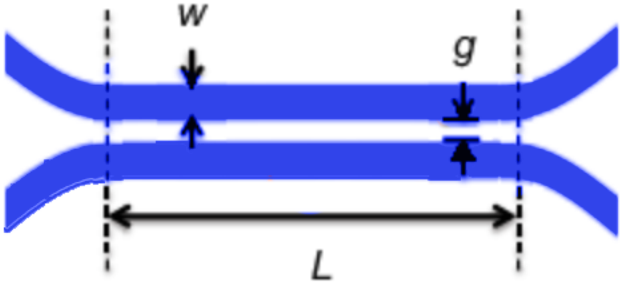
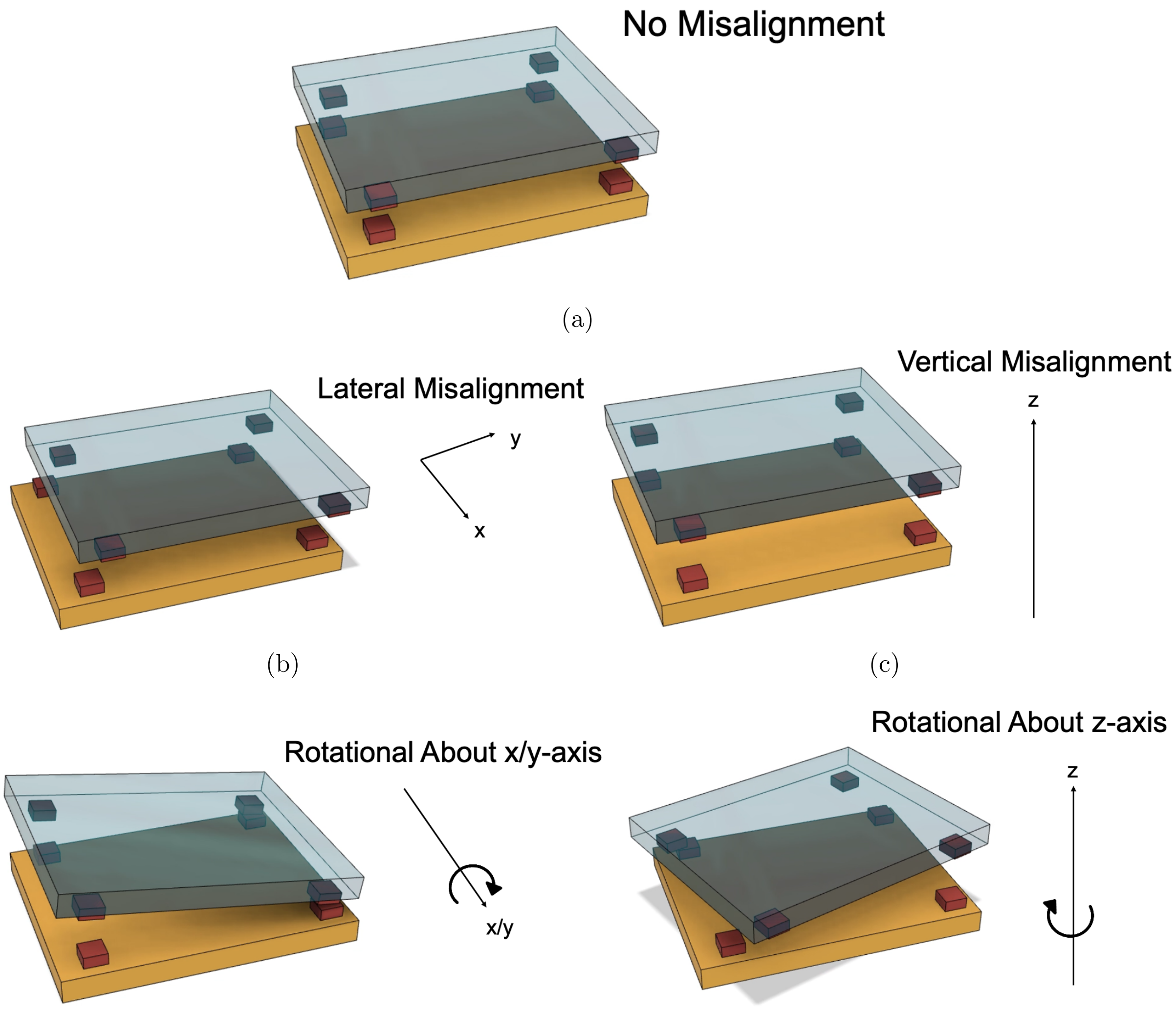


Figure 3.5: Standard directional coupler design, from [48].

3.2 Chip-to-Chip Coupling

Note that all of the above categories of couplers were quoted in terms of fiber-to-chip optical coupling, but just as easily could be used for chip-to-chip coupling (edge couplers included). In the case of chip-to-chip coupling, mode mismatch between the SMF and the SM waveguide is no longer the critical issue - it is instead replaced with the need for out of plane vertical coupling while overcoming a vertical spacing and now dealing with the full effects of all 6 degrees of freedom. Let's begin with the first of these issues, out of plane vertical coupling. The need for out of plane vertical coupling arises because in the majority of situations a flip chip bonding scheme is used. In flip chip bonding the "active" layer side of one chip (the layer or side where all structures have been fabricated on) will be rotated 180°, placed and aligned on top of a second chip which has its active layer facing in the original, upward direction, and then the two chip are adhered together through some type of assembly method such as ultraviolet (UV) epoxy, wirebonding, etc. With this type of assembly scheme, the light, traveling horizontally in a SM waveguide on the lower chip, must change directions to travel vertically in order to eventually couple into the SM waveguides on the upper chip and continue traveling in the horizontal direction again. To add to this, the adhesive material joining the two chips most likely has some non-zero thickness, meaning the mode needs to overcome a vertical gap which may be significant, even on the order of hundreds of microns if an adhesion mechanism such as a solder joint is used. Note that this is assuming a method such as direct bonding was not used, which occurs when two wafers are placed in direct contact with one another and bond at room temperature, resulting in no appreciable vertical gap. Of course, this means the surface would need to be clean, smooth, and flat in this case to avoid the formation of voids, which could hamper chip-to-chip optical performance.

Lastly, in terms of chip-to-chip coupling one of the most important challenges is the presence of 6 degrees of freedom. In the case of fiber-to-chip coupling, fiber alignment to the chip is also an issue, but some of the modes of misalignment are suppressed such as twisting



(a)

(b)

(c)

(d)

(e)

Figure 3.6: Translational and rotational degrees of freedom when aligning two chips together. In all of the images, there are red blocks on each chip, and the goal is to align those red blocks while bringing the chips in contact with one another. The red blocks could represent components of an optical coupler, for example. In (a) perfect alignment is shown while in (b) a lateral shift and in (c) a vertical shift is shown. In (d) and (e), the rotational modes of alignment are shown. As will be discussed in subsequent sections, rotation about the z axis has the lowest alignment tolerance for our chip-to-chip coupler.

the fiber (there is cylindrical symmetry and if the coupler is polarization invariant than performance is unchanged by this action) or, in the case of edge couplers, the gap between the chip facet and fiber facet has relatively little effect on coupling efficiency. Likewise, one can imagine that fiber tilt misalignment (either around the z axis or y axis according to Figure 3.6), although still affecting coupling efficiency, are relatively suppressed in fiber-to-chip edge coupling since the size of the fiber's MFD is very large in comparison to the size of the coupler and mode in the SM waveguide. However, in the case of chip-to-chip coupling, these modes are no longer approximately negligible - all three translational degrees of freedom affect coupling performance, and rotation about all three axes can significantly affect coupling. For example, take the case of the Photonic-plug developed in [44] for fiber-to-chip coupling - the rotational alignment tolerance shown in Figure 3.2(d) is on the order of 0.1 milliradians, or approximately 0.006 degrees, meaning just a minor misalignment between chips during assembly could vastly alter coupling performance.

To overcome these challenges to optical coupling, our proposed fiber-to-chip and chip-to-chip coupling schemes attempt to improve upon the negatives of edge couplers and evanescent couplers by expanding misalignment tolerances beyond 1 μm, reducing footprint, increasing planarity, and decrease loss to develop a set of robust, highly efficient, polarization independent, broadband coupling solutions.

3.3 Performance Metrics and Design Recommendations for Optical Couplers

For chip-to-chip and fiber-to-chip coupling, there are several quantitative performance parameters of note: coupling efficiency (or coupling loss in dB), device footprint, operating bandwidth, polarization dependent losses, fabrication deviation tolerance, and packaging misalignment tolerance [37]. Qualitatively, the important items to consider include the"adiabaticity" of the system, usually referring to modal conversion in tapered structures, CMOS fabrication compatibility and material selection, and the fabrication complexity of the design. Of the quantitative metrics, intrinsic coupling efficiency is the most critical and fundamental parameter for an edge coupler. While it can be related to Equation 2.31, we will directly define coupling efficiency as the ratio of output power over the input power after light transmission and mode conversion inside a coupler as done in [37]. Packaging misalignment refers to the displacement in $x/y/z$ between the edge coupler and the fiber's center axis. Packaging misalignment tolerance, a measure of the allowable error for a coupler design, is typically quoted in terms of 1 dB excess loss, defined as the maximum misalignment in $x/y/z$ allowable before 1 dB of additional loss is reached (with the instrinsic input coupling loss used as the reference). Adiabaticity, as discussed in Section 2.3, refers to the desire for tapers in couplers that are long such that losses from modal conversion between local normal modes within the waveguide are negligible. Device footprint indicates the area or length the coupler occupies on the chip - for our couplers, we first attempt to maximize the coupling efficiency and packaging misalignment tolerance, and then return to the challenge of device footprint. The space occupied by the coupler can become especially important for adiabatic couplers, such as our chip-to-chip tapers, but ultimately depends on the space available on the chip and the purpose of the chip (is it an interposer, a PIC, etc.).

Next is operation bandwidth - or the wavelength range the coupler can operate within before incurring an additional 1 dB loss penalty. This metric is important because it is often desirable to 1) use sources of light with varying wavelength such as tunable lasers (to conduct biochemical spectroscopy sensing for example), without needing to re-design the coupling device or incur additional loss penalties and 2) using multiple wavelengths has multiple important applications both in datacom/telecom architecture, such as the future of WDM, and outside of it, for instance in triplexors for fiber-to-the-home applications, whose devices rely on 1310, 1540, and 1560 nm light [49]. Polarization dependent losses were depicted in Section 2.5, referring to loss penalties incurred when the coupler converts the randomized input polarization from a SMF to the preferred polarization orientation in a SM waveguide. Fabrication deviation tolerance refers to errors occurring in the processing of the coupler, prior to packaging, such as during mask alignment during lithography. The type of method used heavily influences fabrication misalignment, but this source of error is less impactful than the packaging misalignment errors introduced if passive assembly with a minimum number of active alignment steps is desired.

Moreover, on top of the need to pursue excellent performance metrics, the optical coupler also must follow a set of design guidelines which will assist in allowing the coupler to be mass manufacturable. Along those lines, developing a coupler capable of being mass manufactured using current microelectronics fabrication facilities means an understanding of what those facilities excel at is also necessary. This includes following design guidelines and recommendations for optical couplers outlined within the IPSR-I 2020 in the "Interconnects", "Assembly", "Packaging", and "Silicon Photonics" chapters. Examples of these recommendations include the following considerations:

- Use of planar structures
- Increased packaging alignment tolerances for fiber-to-waveguide coupling
- A passive fiber-to-chip or chip-to-chip alignment mechanism OR an active alignment mechanism with minimized active steps
- Little back end processing (this includes high quality polishing for facets, additional coatings or layers, etc.)
- Integrated electronics and photonics while managing thermal budget requirements. This can be challenging when considering the sheer number of different materials in an optoelectronic package, especially when photonics is currently heterogeneously processed and packaged.
- Must pass all reliability tests
- Minimization of the number of parts

By loosely using these guidelines to develop our edge and evanescent couplers, we will have an inherent upside of mass manufacturability, a central theme of this work. In other words, this is yet another qualitative set of performance metrics which can help distinguish our coupler from other similar designs in the field. By using these qualitative and quantitative performance metrics, we can objectively evaluate and compare our coupler to those developed elsewhere, devices which are setting the current standard for photonics packaging. This will be done following the descriptions of our modified fiber-to-chip coupler and chip-to-chip coupler in the following part of this thesis.

Chapter 4

Simulation Methods

Simulations of the interaction of light with matter are advantageous for several reasons. In general, they offer a cheap (paying only for the software license and computational energy) and rapid method to gain an intuitive understanding of how various structures affect the transmission of light. Some may argue simulations can take long periods of time depending on the type of method used, the software implementing that method, and the quality of the computer the software is run on. This is absolutely true; however, it is normally faster and easier to test a new design or design modifications via simulation compared to trial and error in fabrication which will likely take longer and needlessly burn through resources. Additionally, with simulations one has the ability to change parameters quickly and easily in an idealized environment and then evaluate the performance of a device. In other words, they provide a starting point for fabrication. This safe, regulated environment also allows us to test the boundaries of physical laws. To go up against the physical boundaries of a device, the user must have an understanding of the limits of the simulation as well.

First, regardless of the simulation method, in an idealized environment it is nearly impossible to account for everything that will affect the performance of an actual device ranging from fabrication imperfections to dynamic environments. Second, simulations frequently contain inherent assumptions and approximations which can have significant roles in how certain performance metrics are computed, thus affecting design significantly. Therefore, despite their usefulness, simulations provide only a partial insight into device's performance. Prototyping and testing are the best method for proving that the design indeed works. It is a best practice to work in combination with experiments, simulating the initial design and subsequently refine that design iteratively, using feedback from device fabrication and simulations together to drive improvement.

In this chapter, we review the simulation methods we use to evaluate device behavior. Both finite difference time domain (FDTD) method and eigenmode expansion method (EME) allow us to study the propagation of light through our coupler and into SM waveguides, so they will be discussed thoroughly. From this, we can determine the optimal device parameters to be used later as the starting point for mask layout and fabrication. Because we will use a particle swarm optimization (PSO) method to obtain the combination of parameters which results in the best performance, this concept will be discussed briefly as well.

4.1 Finite Difference Time Domain Method

The explanation given here follows the same notation and presentation used by Schneider in [50]. The finite difference time domain (FDTD) method solves Maxwell's equations, specifically those for a source-free, isotropic media found in Equations 2.8-2.11, using the Yee algorithm. The Yee algorithm determines the behavior of E and H in a region of space over time, to be discussed further in the following section.

4.1.1 The Yee Algorithm

The Yee algorithm, first proposed in 1966 by Kane Yee and the basis of FDTD methods, provides a means to numerically solve for electric and magnetic fields in time over some region of space using Equations 2.8 and 2.9 (Faraday's Law and the Ampere-Maxwell Law in a source-free, isotropic media) [51, 50]. The algorithm is remarkably exact - the replacement of derivatives by finite differences is the only approximation. Specifically, this approximation replaces temporal and spatial derivatives using the second-order central difference approximation:

$$\left.\frac{df(x)}{dx}\right|_{x=x_0} \approx \frac{f(\delta+\frac{1}{2})-f(\delta-\frac{1}{2})}{\delta} \tag{4.1}$$

This approximation was obtained by writing the definition of a Taylor expansion about $x = x_0$ with an offset $\pm\frac{\delta}{2}$, subtracting the positive and negative offset equations, and ignoring terms of δ^2 order or higher. It is important to also describe the mesh overlying the spatial domain on which the numerical calculations take place. One might expect to simply calculate E and H across each point in a domain at a certain time, then advance the fields at each point in time to determine the evolution. However, because the primary equations used to solve for E and H are coupled, the two fields cannot be calculated at the same coordinate points. To resolve this issue, E is calculated on the "original" lattice (the standard laboratory reference frame spatial domain) and H on a second lattice displaced from the original by half a unit step in each direction (this can be visualized in Figure 4.2 which was reused from [52]). In this configuration, each E point is surrounded by four points at which H is calculated and vice versa. A unit cell in this space is referred to as the Yee lattice or Yee cell. The fields are then alternately computed in time across the entire mesh. The E field (and analogously the H field) at any given point depends on its value at that point at an earlier time, and on the values of H (and analogously the E field) at the surrounding points [51]. Summarizing these steps, the Yee algorithm operates as follows:

1. Replace all the derivatives in Ampere-Maxwell's and Faraday's laws (Equations 2.8 and 2.9) with finite differences described by the second-order central difference approximation (Equation 4.1). Discretize the spatial and and temporal lattices so that the E and H fields are staggered in both space and time (i.e. form the Yee lattice).

2. Solve the resulting difference equations to obtain "update equations". These are used

to determine future fields based on the already known past fields

3. Go forward one time-step and evaluate the H fields. Now these H fields are known and they then operate as a past field for the next time-step evolution.
4. Go forward one time-step and evaluate the E fields. Now these E fields are known and they then operate as a past field for the next time-step evolution.
5. Repeat the previous two steps until the fields are calculated over the entire simulations region and for the desired amount of time.

While we will not discuss it in depth here, Figure 4.1 (reused from [50]) shows a one dimensional example of a Yee lattice both before and after a single time step has evolved as an example of how the Yee lattice and FDTD method is used to determine the electric and magnetic field over a region of space from the past evolution of the field.

Another important item to describe is mesh stability - in order to accurately simulate electromagnetic fields with FDTD on the Yee lattice, the grid spacing has to be sufficiently small so that field strength does not appreciably change between adjacent grid points. If this condition is true, the finite difference approximation (i.e. the second-order central difference approximation in this case) holds. In other words, this implies that for computational stability the distance traveled by electromagnetic energy in a temporal step $c_0\Delta t$ should not exceed the total spatial step. This results in the following relationship:

$$c_0\Delta t \leq \left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}\right)^{-\frac{1}{2}}$$

This equation has to be satisfied for spatial steps Δx, Δy, and Δz, and temporal step Δt. Along the same lines, appropriate boundary conditions are needed or else extraneous reflections will be introduced at the edges of the computational domain, which are undesirable. To adjust for this type of error, so called “perfectly matched layers” (PMLs) are used - they absorb radiation without creating reflections at the simulation boundaries [53]. The

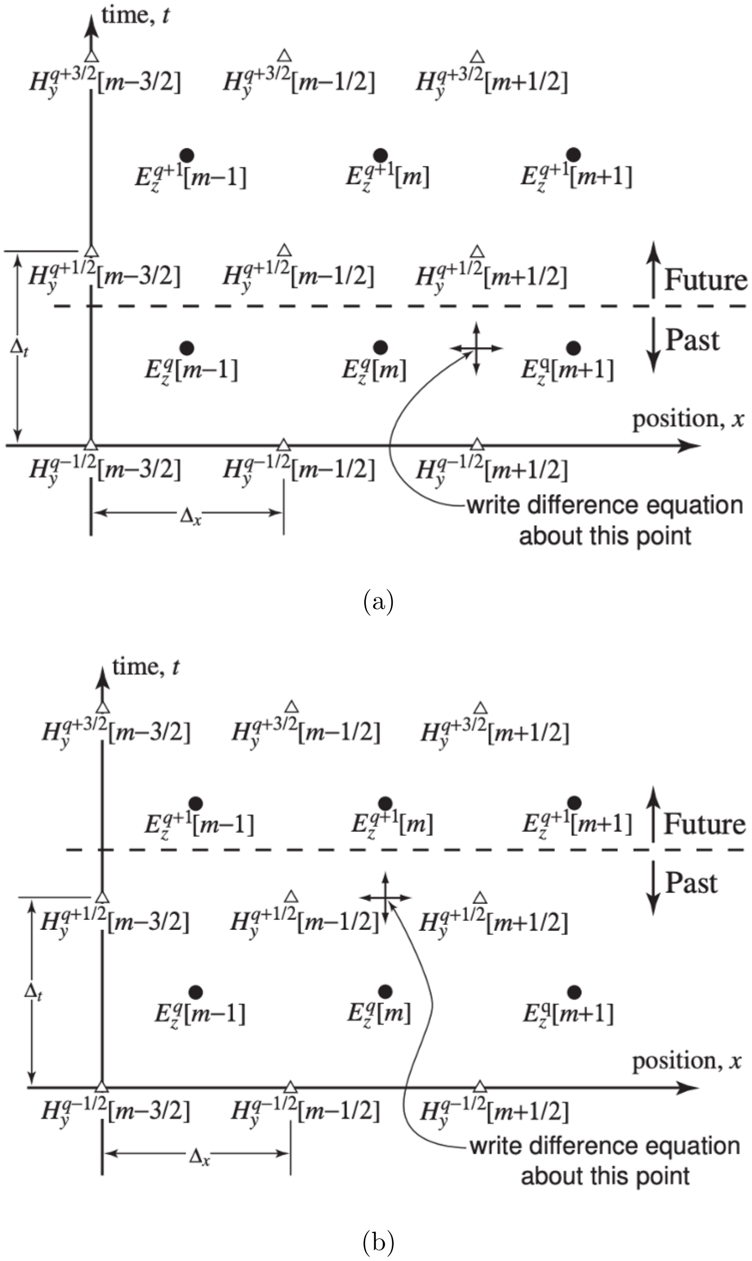


Figure 4.1: A one dimensional example of the Yee lattice or Yee cell. In (a) the simulation status is shown at some time with the dotted line representing the time t. Following execution of one time step, the system evolves to (b). This cycle continues until the entire region of interest has calculated electric and magnetic field values. All images here are taken and reused from [50].

3D FDTD solver in Lumerical supports a range of boundary conditions in addition to PML, such as periodic and Bloch, but we will use PML for the entirety of our simulations.

The last item to note on the FDTD method are the specific form of Maxwell's equa-

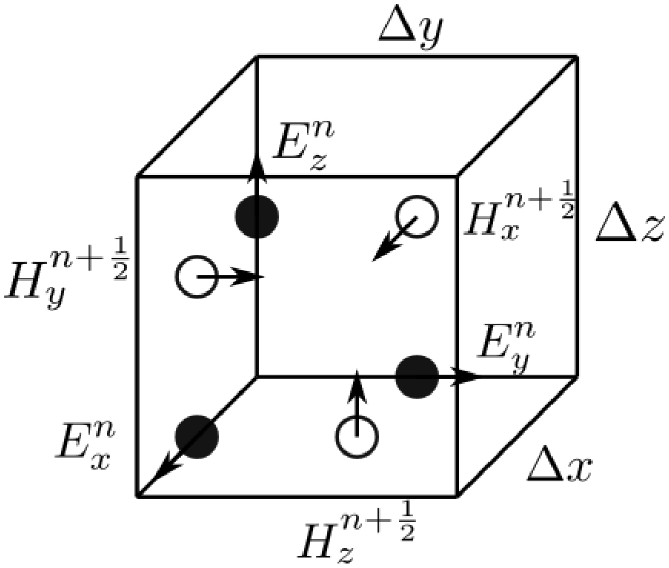


Figure 4.2: The full three dimensional Yee lattice showing the offset electric and magnetic fields for completeness, image reused from [52]. This is what Lumerical software will use in our 3D FDTD simulations to propagate the fields.

tions to be used. If specifically discussing Lumerical's 3D FDTD software suite, Maxwell's equations are further simplified by assuming that the structure is infinite in the propagation dimension and that the fields are independent of propagation direction (taken to be z here). Mathematically, this corresponds to:

$$\varepsilon(\omega, x, y, z) \approx \varepsilon(\omega, x, y) \tag{4.2}$$

$$\frac{\partial E}{\partial z} = \frac{\partial H}{\partial z} \approx 0 \tag{4.3}$$

This is essentially saying that the mesh size (the Yee lattice) is fine enough that the structure appears infinite in the propagation direction. Following this approximation, Maxwell's equations are split into two independent sets of equations composed of three vector quantities. These are termed the TE (transverse electric) equation, corresponding to components E_x, E_y, and E_z components, and TM (transverse magnetic) equation, corresponding to the

H_x, H_y, and H_z components. For example, in the TM case, Maxwell's equations reduce to:

$$\frac{\partial D_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \tag{4.4}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu_0}\frac{\partial E_z}{\partial x} \tag{4.5}$$

$$\frac{\partial H_x}{\partial t} = -\frac{1}{\mu_0}\frac{\partial E_z}{\partial y} \tag{4.6}$$

$$D_z(\omega) = \varepsilon(\omega)E_z(\omega) \tag{4.7}$$

The FDTD method on the Yee lattice can then be used to solve these relationships in the x-y plane only, eliminating the complexity of derivatives in z during calculations. The result is a 3D FDTD method very capable of simulating micro and nanophotonic devices such as the couplers we present in later chapters.

4.2 Eigenmode Expansion Method

An alternative to the FDTD method, and to other methods such as beam propagation method (BPM) or finite element method (FEM), is the eigenmode expansion method (EME). The idea behind EME is relatively simple - take a cross section of the structure along the propagation direction and then solve the Helmholtz equation (i.e. Equation 2.13) to determine the eigenvalues (β's in Equation 2.13) and the eigenfunctions ($\tilde{A}$'s) which characterize the mode. This is accomplished in two separate steps for Lumerical's EME solver:

1. The modal decomposition of electromagnetic fields into a basis set of eigenmodes. As described above, this involves taking slices (called cells) of the structure's geometry (with the number of cells determined by the user) along the propagation direction, and then solving for the modes at the interface between adjacent cells using the Helmholtz equation. The user allocates two port regions with four possible paths for light (input and output of each port). Once the modes are determined, the scattering matrices (called the S-matrix) are calculated by computing the modal overlap integral (i.e.

Equation 2.31) between the tangential E and H fields at the cell boundaries. The S-matrix represents transmission for each input and output at each port (i.e. input, output, and reflections at the input and output interfaces). The calculation of the modes (and thus the S-matrix) is the most time consuming portion of the EME calculation.

2. The solutions obtained for each cell are then used to determine light propagation in the forward and backward propagation direction to calculate the S-matrix of the entire device (not simply the cell interface). This step can be carried out very quickly compared to the first step.

Using an EME method which operates this way has several strengths. First, it is ideal for the modeling of light propagation over long distances in waveguides or optical fibers. This is because the computational cost of the method scales exceptionally well with the device length, making it much more efficient for the design and optimization of long tapers (such as the adiabatic, inverse cross tapers we will present in later chapters) and periodic devices. This is in comparison to 2D or 3D FDTD methods requiring fine discretization (typically on the scale of the wavelength of the light being simulated) along the direction of propagation, yielding immediate benefits to EME in terms of time. Second, the mode calculation can take advantage of symmetries of the structure, is fully vectorial (provided that it relies on a fully vectorial mode solver), and fully bidirectional. In addition, using an S-matrix approach provides a flexible calculation framework, potentially allowing users to only re-calculate modified parts of the structure when performing parameter scan studies while ensuring all reflections are taken into account.

Lastly, unlike other methods such as the beam propagation method (BPM), which is only valid under the slowly varying envelope approximation, EME provides a rigorous solution to Maxwell's equations in the cross sectional analysis. The accuracy of BPM also becomes compromised in components with high refractive-index contrast, making it unsuitable for photonic components manufactured from silicon or other high index contrast material systems (which our couplers will be developed from). Along the same lines, in FDTD based simulations, it's important to use a smaller mesh in high index materials, and to maintain a minimum number of mesh points per wavelength; however, this constraint does not exist for the EME solver. Of course, these advantages come with a few drawbacks, the most significant being that EME is limited to linear problems; nonlinear problems may be modeled using iterative techniques. Also, EME is a frequency-domain solver, so you will only get results for a single frequency. Finally, EME may be inefficient for modeling structures requiring a very large number of modes, which limits the size of the cross-section for 3D problems - note that the goal of our edge and evanescent couplers is to develop single mode coupling devices, so this is not necessarily as relevant.

In the past, researchers' personal computers could only handle 2D simulations, so the simulation itself was unable to account for the effect of the graded index stack on the mode shape and effective index. To get around this, past researchers used the effective index method (EIM) to collapse the 3D coupler into a 2D structure which FDTD can handle. The effective index method translates the index a mode in a 3D structure sees to the core index of a slab waveguide in a 2D structure. The index depends on the mode number, wavelength, and device cross-section. This limits the analysis to only one mode at a time and that mode will vary with propagation distance because of the taper's changing cross-section.

4.3 Particle Swarm Optimization Method

In this research, simultaneous optimization of large numbers of parameters to maximize coupling efficiency is highly desirable. Previously, executing sweeps over individual parameters such as coupler lengths or taper input and tip widths provided an optimization technique by determining the parameter resulting in a peak coupling efficiency. However, by using particle swarm optimization (PSO), simultaneous optimization is possible. The PSO analysis is a "population based, stochastic" optimization technique, inspired by the social behavior of flocks of birds or schools of fish [54, 55], and actually has an easy classical mechanics analogy which can be used to describe it. Population based indicates that the method maintains and improves multiple potential solutions instead of focusing on a single solution. Stochastic means that the method generates and uses random variables. This type of method has been widely used for various kinds of design optimization problems [55] including nanophotonic design [56, 57, 58, 59, 60]. In PSO, the potential solutions, called "particles", begin at random "positions" (each position corresponds to a randomized set of the parameters being investigated), and then move within the parameter search space. In other words, different sets of values of parameters within the search space boundaries can be considered as different positions for the particles. As the particles (the solutions) move in the parameter search space, they are subject to the following "forces" in a classical mechanics analogy :

- A spring, or restorative, force towards a single, individual particle's best position ever attained, labeled p
- A spring force towards the best overall position, labeled g, achieved by any of the available particles moving through the search space.
- A frictional force proportional to the velocity v of the particle.

With these three forces in mind, the algorithm used by Lumerical's PSO follows the following steps in order to arrive at an optimized solution:

1. Set the number of particles N (i.e. mesh the total number of possible solutions by indicating how many are to be tested) and initialize the positions of each particle
2. Execute the simulation (FDTD, EME, etc.) to determine the figures of merit (i.e. transmission or coupling efficiency for the case of our couplers) and then calculate restoring forces p and g for each particle.

3. Calculate the new velocities for each particle based on the forces applied to the particle. This is done using the following relationship where t is the iteration counter:

$$v_t = v_{t-1} + c_1\eta_1(p_{t-1} - x_{t-1}) + c_2\eta_2(g_{t-1} - x_{t-1}) + (\omega - 1)v_{t-1} \tag{4.8}$$

4. Update the positions x of each particle based on the velocity using the following equation where the time step is 1 (so v has the same units as x):

$$x_t = x_{t-1} + v_t$$

5. Use the updated positions to return to step 2 and repeat iteratively until convergence is achieved using a tolerance set by the user. Convergence indicates the velocity of the particle with the best overall position has essentially slowed to zero in the classical mechanics analogy. For example, if the difference between consecutive maximum figure's of merit is less than 0.00001, the optimization is complete.

In Equation 4.8, c_1 and c_2 are the "cognitive" and "social" factors, respectively; ω is called the inertial weight; and η_1 and η_2 are random numbers generated between 0 and 1. In other words, Equation 4.8 is an update equation for velocity using weighted terms for how close the particle is to its personal best and the global best while the $\omega vt - 1$ term provides a frictional force. Lumerical's PSO implementation uses default values for the weighting coefficiencts c_1, c_2 and ω that have shown to converge well in many test optimization problems for photonic design. These values will not be edited for the purposes of this thesis. If more information is desired, a detailed description of the algorithm and the difference coefficients can be found in Refs. [54] or [55].

4.4 Summary

For our simulations, we primarily rely on the 3D FDTD and EME methods to help determine the optimal dimensions and material composition for our edge and evanescent coupler design as well as an intuitive picture of how light interacts with the structures. This work is an improvement over previous analyses whose calculations were limited to 2D FDTD simulations, the use of beam propagation method (BPM), and ray matrices. The use of 2D FDTD simulations is limited by the approximations of the EIM and the fact that only one mode can be considered at a time even in multimoded structures, while BPM is limited by assumptions that may or may not be valid given the nature of the coupler's design. While ray matrices are useful to focus on how changing a single parameter affects a second parameter and eliminates the coupling of other parameters in the overall structure, they also have their restrictions since the physics they are based in is not as accurate as electromagnetic wave optics. That being said, all simulations are of course performed in an idealized environment and not privy to the imperfections based in small fabrication condition variations. Thus, the ultimate check will be whether experiment matches with the design simulations.

Part II

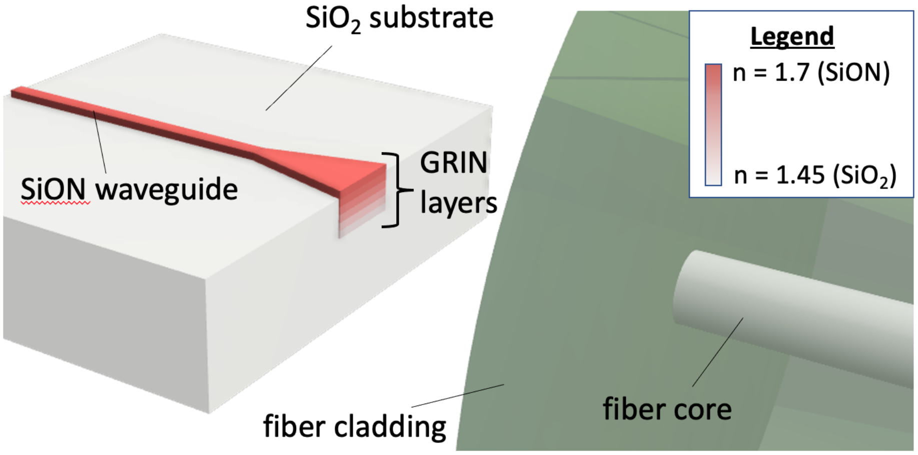
Optical Coupler Design

Chapter 5

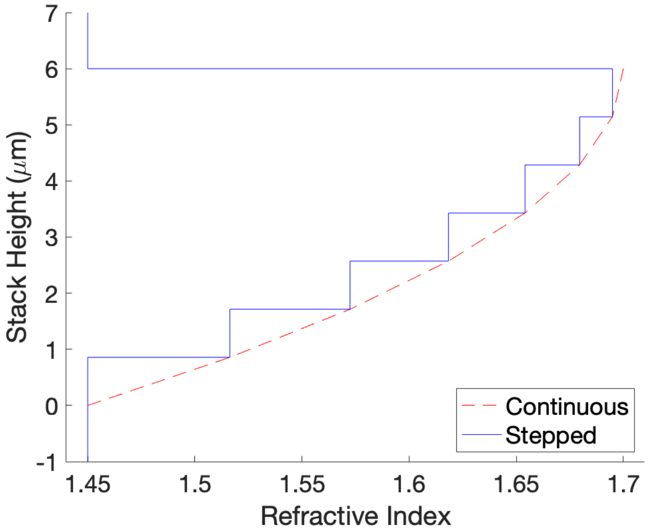
Fiber-to-Chip Coupler

Our proposed technique attempts to achieve three dimensional mode confinement through the combination of a GRIN lens with a horizontal taper. As discussed in Section 2.2.1, grading the index results in the structure focusing (and defocusing) the light similar to that of a micro-lens, allowing us to focus the input mode to a narrower MFD, thus achieving higher modal overlap upon entry to the SM waveguide. In our design, the graded index feature is achieved by depositing $\mathrm{Si}_x\mathrm{O}_y\mathrm{N}_z$ (abbreviated SiON for the remainder of this chapter) layers with consecutively higher indices of refraction on top of one another, such that the final stack has an asymmetric, parabolic index profile. The tuning of the index to match a parabolic profile is accomplished through close control of the oxygen content during deposition using low temperature plasma enhanced chemical vapor deposition (PECVD). This is done by varying the flow of SiH_4 and $\mathrm{N}_2\mathrm{O}$ in the deposition chamber, a process which was optimized in [17] in order to achieve tunable SiON in terms of refractive index, Si-H to N-H ratio, etc. For our design, the refractive index can be engineered to be near that of SiO_2 on the bottom and can have a maximum near that of $\mathrm{Si}_x\mathrm{N}_y$ ($n = 2.03$ for stoichiometric silicon nitride, abbreviated SiN for the remainder of this chapter) on the top (this would result if the film is deposited with no or very little oxygen content). The specific refractive index range used in this study and its effects on the design will be discussed in subsequent paragraphs.

Horizontal confinement is achieved by geometry manipulation, namely through a non-adiabatic taper. Each dimension - the taper length, the input facet width, and the output facet width - contributes towards mode shape and size matching between output waveguide and input fiber in order to maximize modal overlap. We know from the background discussed in Section 2.5 that modal overlap during light entry is important; therefore, the input width is primarily dictated by the MFD of the SMF for mode size matching. The length, however, is determined by distance from the input where, in terms of wave optics, the mode's first minimum oscillation amplitude occurs, i.e. the period of the GRIN lens. In terms of ray optics, this is the effective focal length for ray convergence. Recall from Section 2.2.1 that for a structure, such as that shown in Figure 5.1 with a parabolic, symmetric (or asymmetric) graded index profile governed by Equation 2.26, the light ray (or the modal distribution, if you like) will self focus and oscillate around the center axis of symmetry with a period of oscillation given by π/m [18, 61].



(a)



(b)

Figure 5.1: Asymmetric, tapered GRIN edge coupler index profile. In (a) a practical schematic of the GRIN coupler where it is buried in an SiO$_2$ cladding. In (b) the asymmetry, parabolic index profile of the structure is detailed, highlighting how the index of each step in the stack is carefully selected to achieve the profile outline with the dotted line.

5.1 Asymmetric, Tapered $Si_xO_yN_z$ GRIN Coupler

5.1.1 Design Overview

In this section we will begin with a design overview, using prior versions to detail the evolution of the GRIN edge coupler. This will be followed by the detailing of necessary changes for integration with our chip-to-chip coupler, and how those changes can be optimized to maintain high coupling efficiency and practical alignment tolerances. The initial design was developed in [61] and is shown in Figure 5.2. The device used a symmetric, parabolic index profile instead of an asymmetric profile, and used a lens on the input facet with a very small radius of curvature of 2.5 μm compared to 9 μm used in following iterations [61, 62]. In addition, the length of the initial design was very short - on the order of 5 μm versus the 20 μm lengths used later on. Maximum coupling efficiencies for this design were calculated using FDTD simulations to be 82% [61]. This high coupling efficiency can be attributed partially to the focusing by the highly curved input lens and also to the use of lens tipped fibers which contributed to mode size and shape matching prior to interaction with the coupler. Although the use of high curvatures lenses aid in focusing light rays and shortening the required chip footprint, they diverge from design recommendations for low cost, low fabrication complexity designs for mass manufacturing, such as those outlined earlier in the IPSR-I 2020 “Assembly” chapter highlighting the need for planar structures [12] .

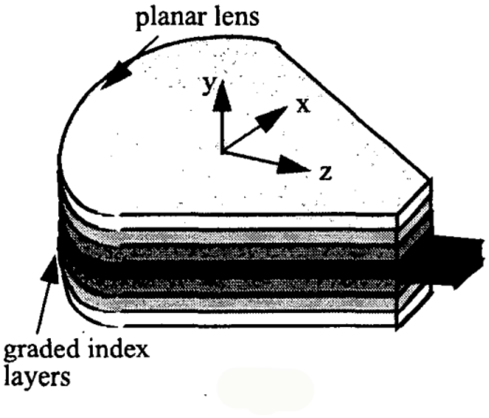


Figure 5.2: Initial design for symmetric, parabolic GRIN edge coupler proposed and simulated in [61] with a peak coupling efficiency of 82%.

The next model of this design removed the lens on the input facet, used non-lens tipped NuFern980$^{\mathrm{TM}}$ fibers for coupling, and switched to an asymmetric, parabolic index profile. This iteration was developed in [63] and is shown in Figure 5.3.

The maximum coupling efficiencies achieved experimentally were 65% (1.9 dB loss) at

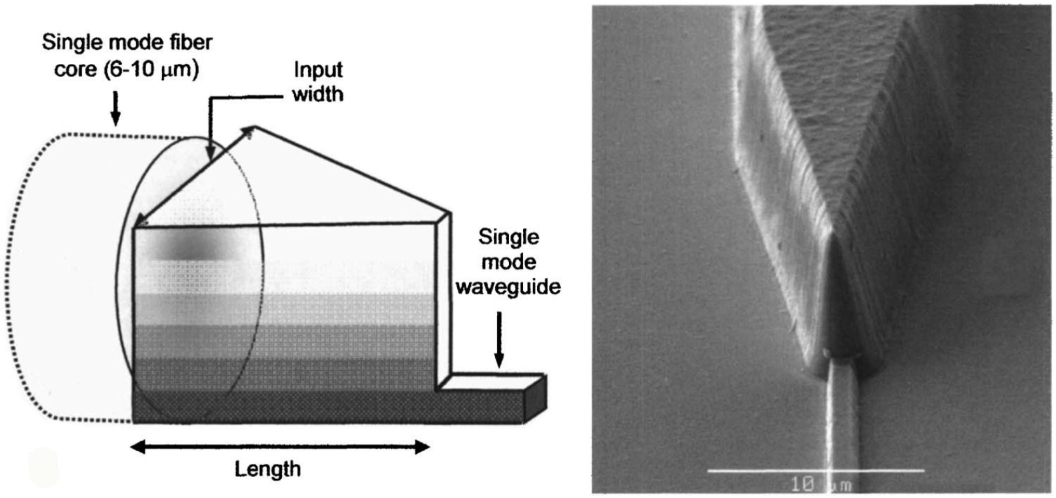
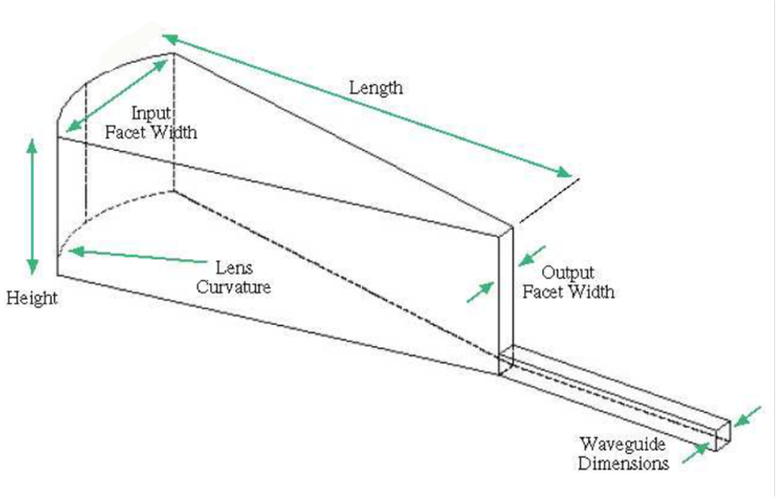


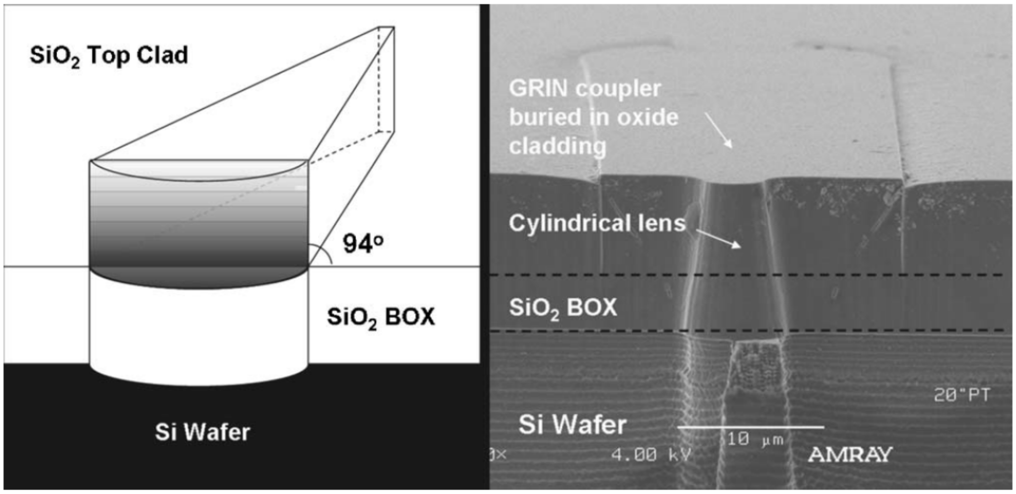
Figure 5.3: Next generation design for an asymmetric, parabolic GRIN edge coupler, simulated and fabricated in [63] with a peak coupling efficiency of 65% (1.9 dB loss) at 1540 nm and 60% (2.2 dB loss) at 1550 nm.

1540 nm and 60% (2.2 dB loss) at 1550 nm. The fabrication process for this coupler was then optimized in [62] through the use of an improved sidewall etching process using a plasma etch with trifluoromethane, leading to a maximum coupling efficiency of 80% for a flat faceted input and 90% coupling efficiency for lensed input facet [62]. The optimized structures were 20 μm in length, 6 μm thick, and 7 μm wide at the input facet, and a schematic as well as cross-sectional SEM images are shown in Figure 5.4.

Previous GRIN models had structures optimized for 1550 nm light; however, BPM simulations in these studies indicated that vertical focal length depended weakly on wavelength (20.5 μm for 1310 nm compared to 20.6 μm for 1550 nm). The 2D FDTD simulation results in Figure 5.5(a) from [49] showed coupling efficiency does not change either. These results are shown alongside experimental results demonstrating the broadband nature of the device. Based on this data it is reasonable to expect the coupler will perform well at 1310 nm even though it was not optimized for that wavelength.



(a)

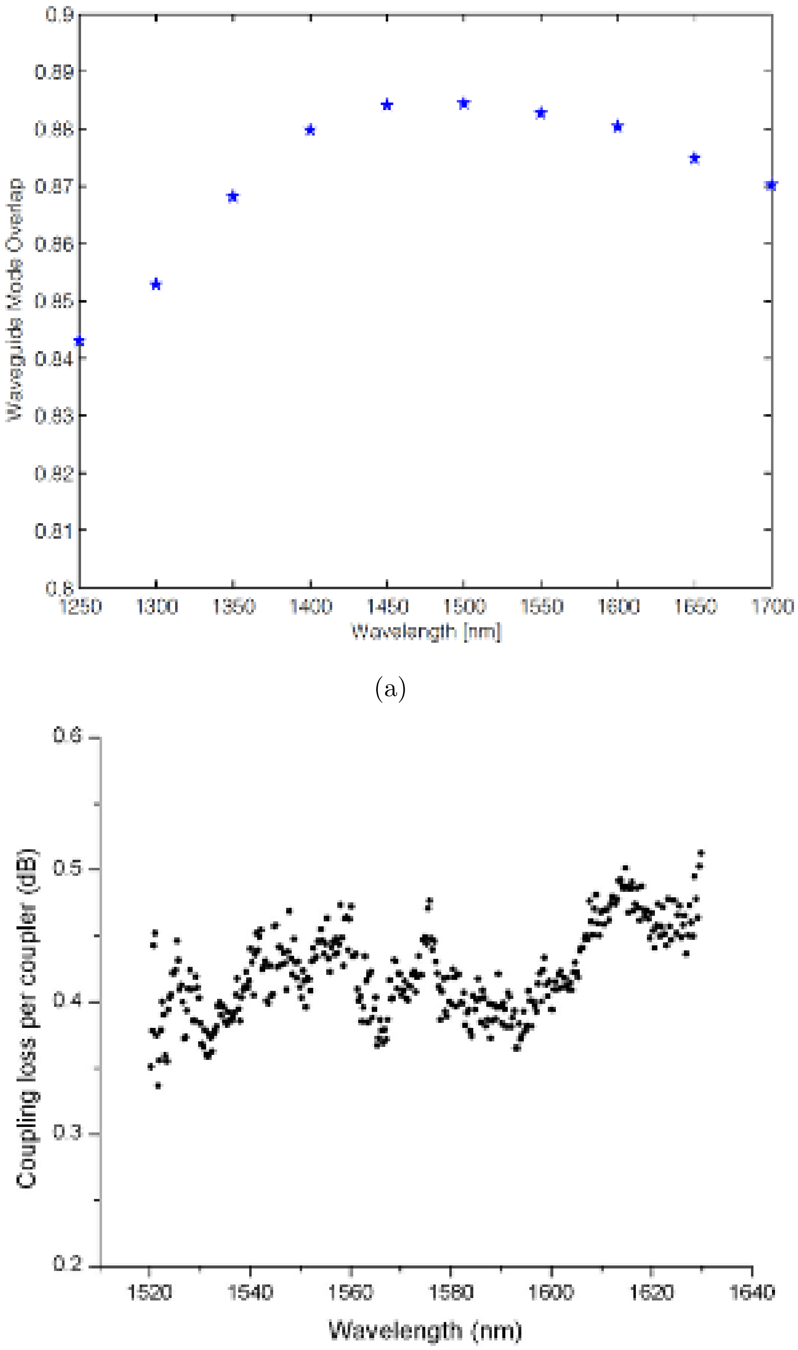


(b)

Figure 5.4: Design modification of the asymmetric, parabolic GRIN edge coupler. In (a) a schematic image showing the design parameters explored [49]. This design was then simulated and fabricated in [62] with a peak coupling efficiency of 80% for a flat faceted input and 90% coupling efficiency for lensed input facet at 1550 nm. This was due to an improved plasma etch process using trifluoromethane increased the sidewall angle to 94°. In (b) a cross-sectional schematic and SEM image of the fabricated coupler showing the sidewall improvement, again from [62].

5.1.2 Design Modifications and Optimization

Here, we will discuss deviations made from past designs, the reasoning behind them, and how these modification can be optimized to maintain coupling efficiency. In the prior design studies conducted in [49], the design parameters allowed to vary were GRIN coupler length,



(a)

(b)

Figure 5.5: Coupling efficiency versus wavelength results for an asymmetric GRIN edge coupler. In (a) 2D FDTD simulations were carried out for a flat faceted coupler [49]. In (b) experimental results are reported for a fabricated lensed coupler [62]. Note that although (b) is for a lensed GRIN edge coupler, [63] showed the same effect for a fabricated flat faceted coupler.

the input facet width, the output facet width, and the lens radius, while fixed design parameters included wavelength, input fiber parameters, graded index profile, GRIN thickness, taper profile final waveguide index, and final waveguide dimensions. Our design modifications add new elements to the coupler, change the orientation of the coupler in space, and allow design parameters which were constrained previously to vary. On top of these changes, we optimize our new design for coupling from a SMF with a core diameter of 8.2 μm which is more in line with widely used SMFs today.

The first modification is the 180 degree rotation along the x axis, so the highest index layer is on top. The primary reason for this is to allow the mode to travel vertically upward. This is important because the light must be near the surface of the lower chip in order for our evanescent chip-to-chip coupler to operate (which will occur in a different region of the chip). Note that this does not mean our PIC-to-PIC coupler cannot operate without the GRIN coupler and vice versa, we are simply assuming here that they are used in series with one another for completion. Because of the symmetry of the system, this rotation should result in zero change to coupling efficiency, just a change in the output location. If the output location needed to remain as the 220 nm silicon layer on 3 μm oxide (assuming standard SOI), this rotation would not be necessary; however, the modifications discussed in this section could still be applied to maintain coupling efficiency and alignment tolerance while avoiding substrate leakage.

The second modification is the addition of a crossed taper above the output waveguide of the GRIN coupler. Again, this change was made for two reasons. First, this allows for customization of the final waveguide output. For example, it might be desirable for the on chip waveguides to be silicon or silicon nitride with smaller dimensions and tighter mode confinement instead of lower index SiON with a large mode and large required bending radii. For our situation in particular, we want to demonstrate how light can enter through the fiber onto a lower chip and then onto a separate upper chip, so the final output waveguide will be customized to be the SM waveguide used in the chip-to-chip coupler. Basically, the overlayed taper in our case allows for separation of the GRIN design from the evanescent coupler design and the performance of each can be optimized without coupling the parameters. An added advantage of this is that this allows for the output waveguide attached directly to the GRIN lens stack to vary more readily in geometry without being constrained to the 220 nm thickness used for other components of the chip or the $n = 1.7$ index and square dimensions of the original design. Essentially, it creates a hybrid edge-evanescent coupler, where light can enter efficiently from the fiber to a more confined mode size (but not the exact size of the final SM waveguide), and then the final extraction of the mode happens very efficiently in an evanescent fashion.

The third structural change to the coupler is the removal of the lensed facet at the input of the GRIN structure. The removal of the lens was done in order to more rigorously following design guidelines set out by institutions such as the IPSR-I dictating increased use of planar layers is desirable if the coupler is to be adopted for mass manufacturing. Moreover, 3D FDTD simulations show the lens adds marginal benefits to coupling efficiency - on the order of 0.5 added %; however, in terms of additional assembly steps the impact is much larger. Companies responsible for assembly of SMFs in V-groove arrays to chip facets need to add a sub-mount in the case of a not perfectly polished, flat facet which is the case for an input lens. This sub-mount is an added processing step which only introduces more error in potential alignment and implies higher cost and energy to do so. These additions, subtractions, and orientation changes can be seen by reviewing Figure 5.8 showing explicitly a schematic of our modified edge coupler.

The next series of changes result from allowing previously constrained design parameters to vary, and then optimizing their values based on the particle swarm optimization (PSO) method. In prior studies, the effect of each parameter was investigated separately using parameter sweeps in 2D FDTD and BPM (note that the discussion of those parameters, i.e. length, input and output width, and lens radius, will not be repeated here and the reader is directed to Chapter 4 of [49] which details these effects and provides design rules for future systems). We will allow the input width, output width, and taper length to vary as in the previous studies; however, we will also allow the index of the uppermost layer, the GRIN output waveguide index and the GRIN output waveguide dimensions to vary also. Because all of these parameters are coupled together, investigating their individual affects can be extremely challenging. Instead, we will use the PSO method in Lumerical's 3D FDTD simulations to obtain a computer aided design which has optimized these parameters to achieve the highest possible coupling efficiency and misalignment tolerances. Note that the parameters named above are the ones which the PSO method will optimize, but not the only parameters varying between two designs - values for parameters such as taper tip width and final waveguide output width are different, but only because they were structurally absent from the reference design (they are not optimized via PSO method). The graded index profile will still be constrained to be parabolic following Equation 2.26 and the number of layers will be constrained to be 10.

Prior to allowing these parameters to vary, we will use the system to optimize an identically constrained system to that found in [49]. Following this, we will alter the SMF core diameter (priorly constrained) to 8.2 μm and re-optimize in order to provide a solution more compatible with current, widely used SMFs. The first of these simulations was done because we are using a more advanced simulation method and software with Lumerical's 3D FDTD, but we can show that the PSO method provides similar results to those previously calculated, proving its accuracy and utility. The second of these simulations (the changing of the SMF to have a core diameter of 8.2 μm) was done to provide a reference for coupling efficiency and misalignment tolerance once other design parameters are allowed to vary. A table showing parameter constraints, optimized varying parameters using PSO, and previously determined parameters from [49] can be found in Table 5.1.

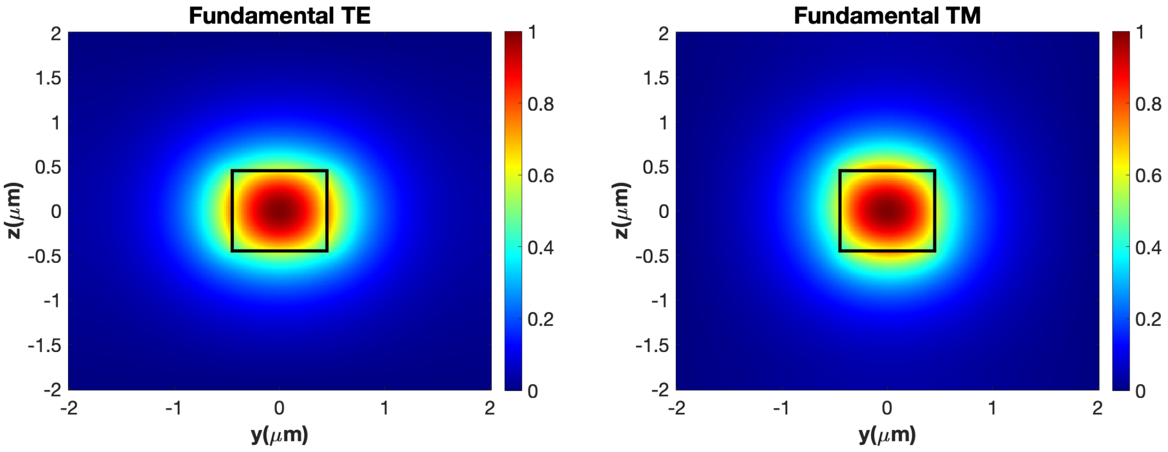
Notice that the optimized parameters vary by less than 2.6% for coupler length, 0% for input facet width, and 8.7% for output facet width. Note that the lens radius is significantly different from prior simulations; however, the experimental measurements in [62] show a

Parameter	**2D FDTD and BPM**	**3D FDTD PSO**
Variable Parameters		
Input width (μm)	7.7	7.2
Output width (μm)	2.3	2.5
Taper length (μm)	19.5	20
Lens radius (μm)	4	9
Total thickness[1] (μm)	6	6.5
Constrained Parameters		
Bottom layer index	1.52	
Top layer index	1.7	
Number of layers	7	
Waveguide width (μm)	0.9	
Total thickness (μm)	6	
Output waveguide index	1.7	
Fiber core diameter (μm)	6.8	
Fiber core index	1.7	
Fiber cladding index	1.45	

Table 5.1: Optimized parameters for asymmetric, tapered GRIN edge coupler following PSO using 3D FDTD simulations in Lumerical. Note that the reference design is that found in Figure 5.4 from [49, 62].

[1] Note that GRIN thickness was previously constrained to 6 μm, but was allowed to vary during the this optimization.

maximum coupling efficiency for a lens radius between 7-8 μm which is closer to our determined value. Images of the output modes are provided in Figure 5.6 as well, for context when discussing design changes. Having shown the similarity between previous optimizations and using PSO, we now change the SMF characteristics and run the optimization. A table showing the optimized parameters and coupling efficiency for a SMF with a core diameter of 8.2 μm (thus a MFD of 10.4 μm from Section 2.4) can be found in Table 5.2 while misalignment tolerances can be found in Figure 5.10(b). Notice that from Table 5.2 and Figure 5.10(d), the structure's input and output width, and length, only needed mild adjustments (from 7.2 μm, 2.5 μm, and 20 μm to 7.2 μm, 1.8 μm, and 30 μm, respectively) to maintain the same coupling efficiency, polarization independence, and broadband nature for input from a SMF with a different core diameter. Note that the thickness of the GRIN was allowed to vary across these two simulations from 6 μm to 10 μm, but was user constrained and not allowed to vary during optimization. The number of layers was also changed from 7 to 9 in order to maintain an approximately parabolic graded index profile. The increase in GRIN thickness was allowed in order to achieve more suitable mode matching conditions, even for a relatively low index contrast system. It is the thickness change in the GRIN which causes the shift in necessary coupling length, because this alters m in Equation 2.26 and thus the effective focal length of the GRIN lens. In terms of fabrication, this can have consequences - the previously optimized process in [17] was developed for the 6 μm stack. A thicker stack will likely induce additional stresses and sloped sidewalls. To provide more context, the added stresses can cause wafer bowing during deposition, an issue mentioned in detail in [17]. Basically, because the SiON stack has a high compressive stress associated with it due to the oxygen content, and because it is only deposited on one side of the SOI wafer, this causes the wafer to bow on the order of 125 μm. This in turn requires oxide deposition on the backside to compensate [17]. With that being said, the same techniques used in [17] to overcome additional stresses and in [62] to overcome sidewall sloping are scalable, and could be used to re-optimize the process for a 10 μm stack compared to the original 6μm stack.



(a) (b)

Figure 5.6: The fundamental TE and TM modes for a single moded, square SiON waveguide of 0.9 by 0.9 μm^2 dimensions and $n = 1.7$ index. This is the output waveguide of the reference design GRIN coupler shown for reference compared to the output mode of the modified coupler shown later in this chapter.

Parameter	**6.8 μm MFD**	**10.4 μm MFD**
Variable Parameters		
Input width (μm)	7.2	7.2
Output width (μm)	2.5	1.8
Taper length (μm)	20	30
Fiber core diameter (μm)	6.8	8.2
Constrained Parameters		
Bottom layer index	1.52	
Top layer index	1.7	
Number of layers	7	9
Waveguide width (μm)	0.9	
Total thickness (μm)	6	10
Output waveguide index	1.7	
Fiber core index	1.7	
Fiber cladding index	1.45	

Table 5.2: Optimized parameters for asymmetric, tapered GRIN edge coupler following PSO using 3D FDTD simulations in Lumerical. Note that the reference design is that found in Figure 5.4 from [49, 62].

Now, we allow the input width, taper length, output width, GRIN output waveguide dimensions, GRIN output waveguide index, and uppermost layer index to vary while adding in the final cross taper for modal extraction, thus creating the hybrid edge-evanescent coupler (later it will be shown that our design is optimized for 1550 nm light, but has a broad 1 dB bandwidth, so wavelength can also be varied for WDM applications). A labeled cross sectional view of the simulation can be found in Figure 5.7 showing the structure with the output monitor. Additionally, a cross sectional motion monitor was used to visualize the evolution of the mode in the x-z plane as it transits from the fiber to the output waveguide. Snapshots of that motion monitor can be found in Figure 5.9, showing the size conversion from a MFD of 10.2 μm and effective area of 81.7 μm^2 to the SM waveguide with an effective area of 2.9 μm^2. Notice from the images that the location of highest loss is at the interface of the GRIN stack to the output GRIN waveguide. This is due to multiple reasons - first, the GRIN stack, being especially thick, excites higher order modes which carry optical power; however, the output waveguide of the GRIN is single moded and this transition is abrupt, resulting in mode conversion loss. A potential solution might seem to make the output GRIN waveguide larger, even approaching multimodal dimensions, so that this modal conversion loss does not occur and because this would better mode match the fundamental mode in the waveguide and the fundamental mode in the GRIN stack at the designed taper length. This actually would worsen the situation, though, due to the presence of the cross taper. The cross taper is intentionally single moded for customization, and so if the GRIN's output waveguide supported multiple modes, the cross taper would efficiently couple only the fundamental mode. Due to the modal dispersion (difference in group velocities), the higher order modes would continue to travel in the larger waveguide at a different speed until the waveguide reached an abrupt end, decreasing coupling efficiency below 20% or even below 10% depending on the dimensions used. The end conclusion is a design rule - the output GRIN waveguide should be single moded, even if a cross taper is used to extract the mode and as the final output waveguide of the entire coupler.

A table showing parameter constraints and optimized varying parameters using PSO can be found in Table 5.3. There are several items to note based on Table 5.3. First is the change made to the output waveguide dimensions. By increasing the width and the thickness of the output waveguide to its maximum value, constrained by the need to only support a single mode for the reasons described in the prior paragraph, we can more effectively mode match the mode at the edge of the GRIN stack and the intermediate waveguide. Note that we also tapered the GRIN output waveguide located under the additional taper containing the final coupler output waveguide (so they are overlaying cross tapers, similar to our evanescent coupler with a significantly thicker lower taper). The final output waveguide of the coupler is still single moded and constrained to a 0.2 by 3.2 μm^2 waveguide at an index of $n = 1.7$ on the lower chip for ease of connection to our chip-to-chip coupler. Recall this taper is customizable - it can be whatever material is desirable for the PIC and have a refractive index as high as silicon (3.48) as long as the taper length and tip width are properly adjusted (within reason - the lowest tip width necessary is for silicon in order to lower the effective index below that of the GRIN output waveguide. That being said, the tip width is still above 100 nm, as this results in an effective index for the mode below 1.44 which is low enough for effective index matching to occur). The second is the lengthening of the GRIN stack taper to have a shallower, but still linear, taper angle. This lengthening results in a lower modal conversion loss as the light is horizontally confined from the input fiber. In essence, the GRIN stack taper is becoming more adiabatic, although still non-adiabatic on the whole. The change in GRIN length this time is most likely caused by two factors - the increase in the number of layers from 9 to 10 (due to the method of coding the GRIN structure this actually slightly increases the GRIN thickness) and the increase in the input width from 7.2 μm to 7.9 μm for modal matching to the fiber. Both of these factors changing the GRIN length makes sense when we review Equation 2.26 where the focal length was $\pi/2m$, where m has now been changed to 0.0460 μm^{-1} from 0.0511μm^{-1}, so coupler length must increase to match the focal length.

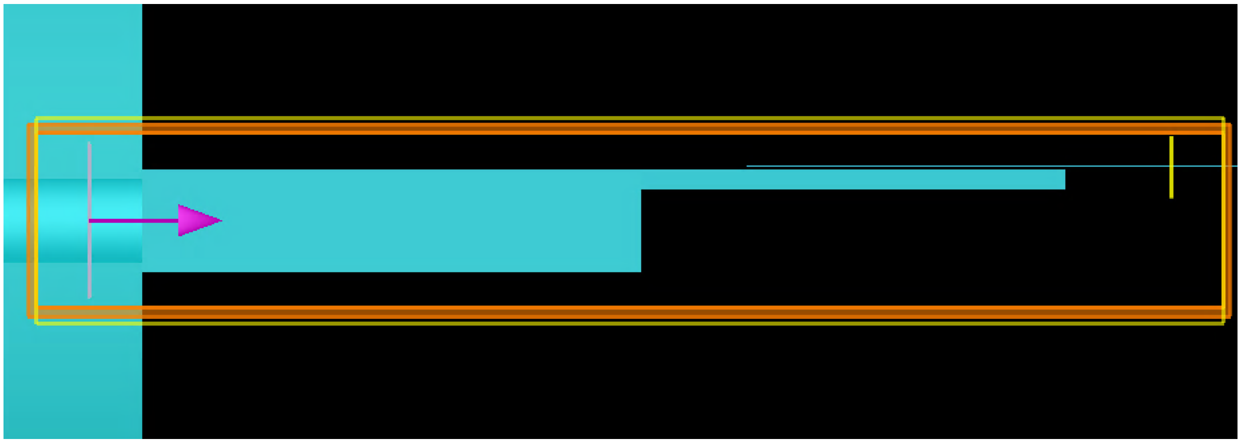


Figure 5.7: A cross section of the 3D FDTD setup showing the input mode, the structure, the output monitor, the FDTD simulation region, and the region of movie capture. Note that the layers of the GRIN cannot be noticed visually because in Lumerical similar refractive index materials are made a similar color, and the difference between the top and bottom layers of the GRIN is approximately 0.18.

With the parameters of our modified design optimized, we can determine the performance metrics as outlined in Section 3.3, and then compare these performance metrics against competing designs in the photonics field. First, we can compare the metrics between our reference designs and our modified designs. The coupling efficiency in the x,y, and z directions for the

Parameter	Reference Design	Modified Design
Variable Parameters		
Input width (μm)	7.2	7.9
Output width (μm)	1.8	1.8
Taper length (μm)	30	47
GRIN output waveguide width (μm)	0.9	1.5
GRIN output waveguide thickness (μm)	0.9	1.5
Top layer index	1.7	1.6
GRIN output waveguide index	1.7	1.6
Constrained Parameters		
Final output waveguide index	\	1.7
Cross Taper Gap (μm)	\	0.2
Final output waveguide thickness (μm)	\	0.2
Taper tip width[1] (μm)	\	0.1
Final output waveguide width (μm)	\	3.2
Total thickness (μm)	10	
Number of layers	10	
Bottom layer index	1.52	
Fiber core diameter (μm)	8.2	
Fiber core index	1.465	
Fiber cladding index	1.45	

Table 5.3: Optimized parameters for the GRIN coupler comparing the reference design to the modified design using PSO and 3D FDTD simulations. Note that the reference design is that found in Figure 5.4 from [49, 62].

optimized reference design with a 6.8 μm MFD fiber versus a 10.4 μm MFD fiber are shown in Figure 5.10. Also, the wavelength and polarization dependence of the reference design is shown to contain a high 1 dB bandwidth and low PDLs based on the results in Figure 5.10(d). Next, the coupling efficiency of our proposed design has a peak at 94% as shown in Figure 5.10(b). The 1dB misalignment tolerance was simulated to be ± 2.2 μm and ± 2.5 μm in the y and z directions respectively (the coordinate system in Figure 5.10(a) is assumed as reference). Again note that the x misalignment tolerance is relatively large (greater than 5 μm at least, although not simulated due to the length of time required for the simulation to run) even in the presence of an air gap. If the gap between the chip facet and fiber facet

---

[1] Note that the same tip width was used for the GRIN output waveguide taper as the final coupler output taper, for convenience.

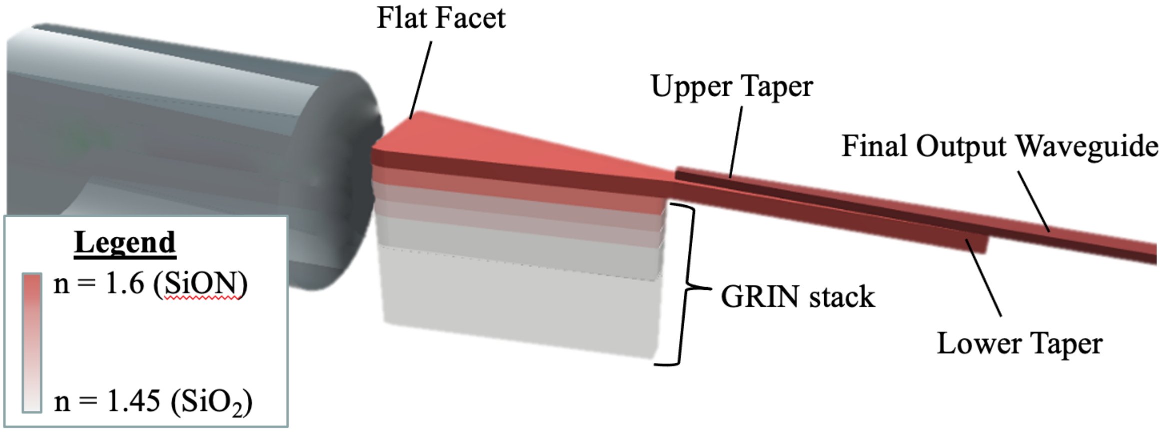
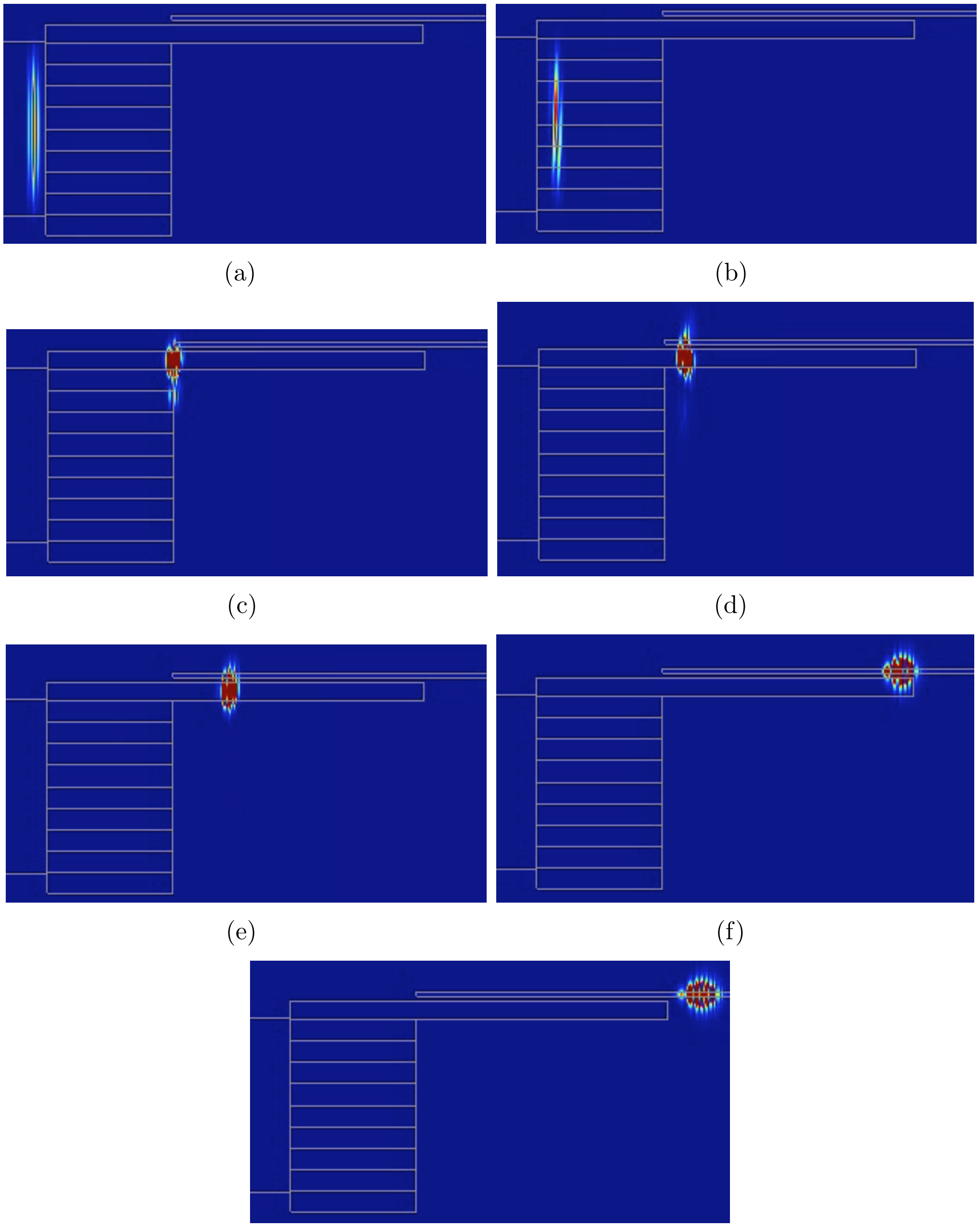


Figure 5.8: Our modified design for the asymmetric, parabolic GRIN edge coupler. Note the key differences between this and prior designs including a 180° rotation of the coupler to couple light vertically into the upper layers of the chip and the addition of a cross taper to take light into a SM waveguide and allow for enhanced flexibility in the GRIN output waveguide dimensions.



(a) (b)

(c) (d)

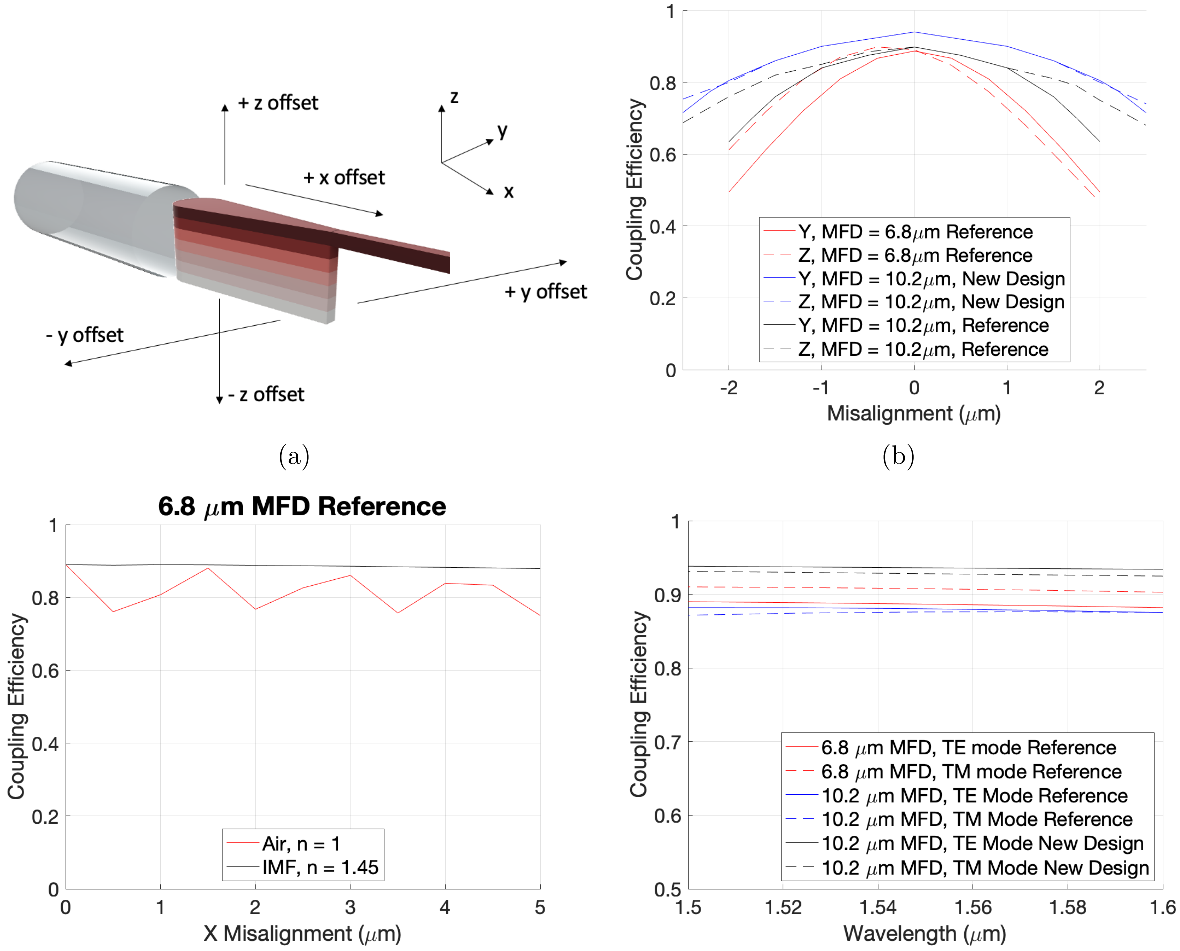
(e) (f)

(g)

Figure 5.9: Snapshots of the motion monitor tracking the light entering from the fiber (left in each image) to entry into the final SM waveguide (upper right of each image). Notice that the location of the highest loss is at the interface of the GRIN stack and the GRIN output waveguide, and that the addition of the cross taper induces little loss, especially if there is space to make the taper adiabatic (increase the length).

is filled with air with a refractive index of 1, the gap essentially forms a cavity instead of a continuous material as in the case of index matching fluid (IMF). The cavity is expected to cause oscillation in coupling efficiency due to the imperfect reflections at the interfaces of the fiber facet and the chip facet (forming small resonances). This is reflected in Figure 5.10(c), while the IMF plot shows the expected result for practically zero reflection at the interface. The design now has a larger longitudinal footprint compared to former models due to the lengthening of the GRIN and the addition of cross tapers at the output. The exact length of the cross tapers depends on if they are adiabatic (lowest loss) and what material is the final output waveguide. For our system, where if we assume the final output waveguide leads directly into the evanescent coupler and thus must be SiON with $n = 1.7$ and width 3.2 μm, the tapers are 100 μm long to ensure they are adiabatic. However, this design is still less than 150 μm so in comparison to other designs in service such as Si nano-tapers it is still reasonable.

Another important item to note aside from the performance metrics is that this design is bidirectional, meaning that the same design can be used for light traveling from chip-to-fiber as fiber-to-chip. This significantly simplifies fabrication by implying that the same optimized process flow (including materials, dimensions, tools, temperatures, etc.) can be used to create the input couplers as the output couplers; in fact, because of this they could be done in the same step, just in a different location along the chip. This idea is demonstrated in Figure 5.12, showing the path taken by light in the case where it is going from chip-to-fiber and the fiber-to-chip case as well. This idea can be verified by calculating coupling efficiency and misalignment tolerances in the reverse propagation direction, using the 3D FDTD simulation cross sectional setup shown Figure 5.11. Note that this makes sense, since the GRIN section of the coupler acts like a lens in either propagation direction - once the light enters the GRIN it is controlled by the index profile. Another way of explaining it is to through considering mode as a transverse electromagnetic wave amplitude instead of ray optics. The field distribution of the mode within the GRIN oscillates between a broad field



(a) (b)

(c) (d)

Figure 5.10: The asymmetric, tapered GRIN edge coupler's 3D FDTD simulated alignment tolerances. In (a) a diagram showing the coordinate system and definitions of ± x, y, and z offsets. In (b) the alignment tolerance when the GRIN coupler is moved relative to the fiber in the y and z direction. Note that in the misalignment tolerance is asymmetric due to the asymmetric index profile of the GRIN. This means that if the fiber is misalignment too low on the GRIN there will be slightly higher losses than if the center of the fiber is accidentally aligned with the top layer (the highest index layer) of the GRIN. In (c) the x misalignment tolerance is shown for an air gap between the fiber facet and the GRIN coupler and for an index matching fluid (IMF) of $n = 1.4587$ (the same as that quoted in [63, 62]. In (d) the coupling efficiency versus wavelength shows little appreciable change over a broad span, indicating applicability in broadband applications such as WDM.

and a narrow field regardless of propagation direction, as was described in Section 2.2.1. This means that, since the GRIN length is determined such that the wide distribution is achieved near the SMF and the narrow distribution is achieved near the output waveguide, that a mode propagating in reverse will expand from the narrow distribution to the wide distribution and efficiently couple into the fiber. The result is a high coupling efficiency in the reverse direction of 99.5%.

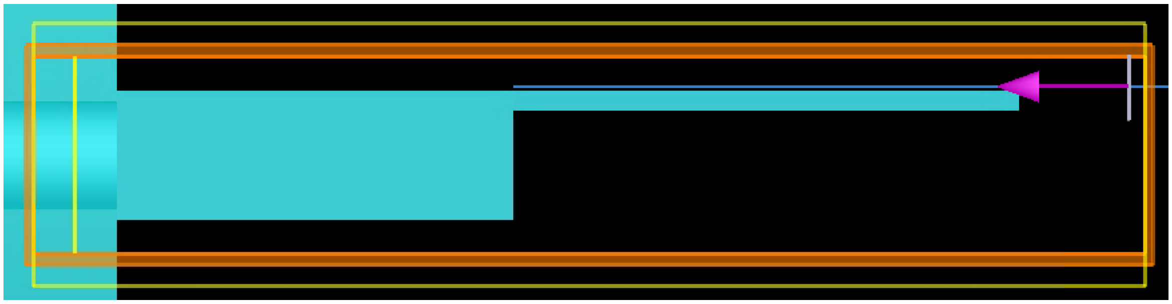


Figure 5.11: A cross section of the 3D FDTD setup showing the input mode, the structure, the output monitor, the FDTD simulation region, and the region of movie capture for the reverse propagation direction. The high coupling efficiency of 99.5% in the reverse direction demonstrates the bidirectional nature of the coupler.

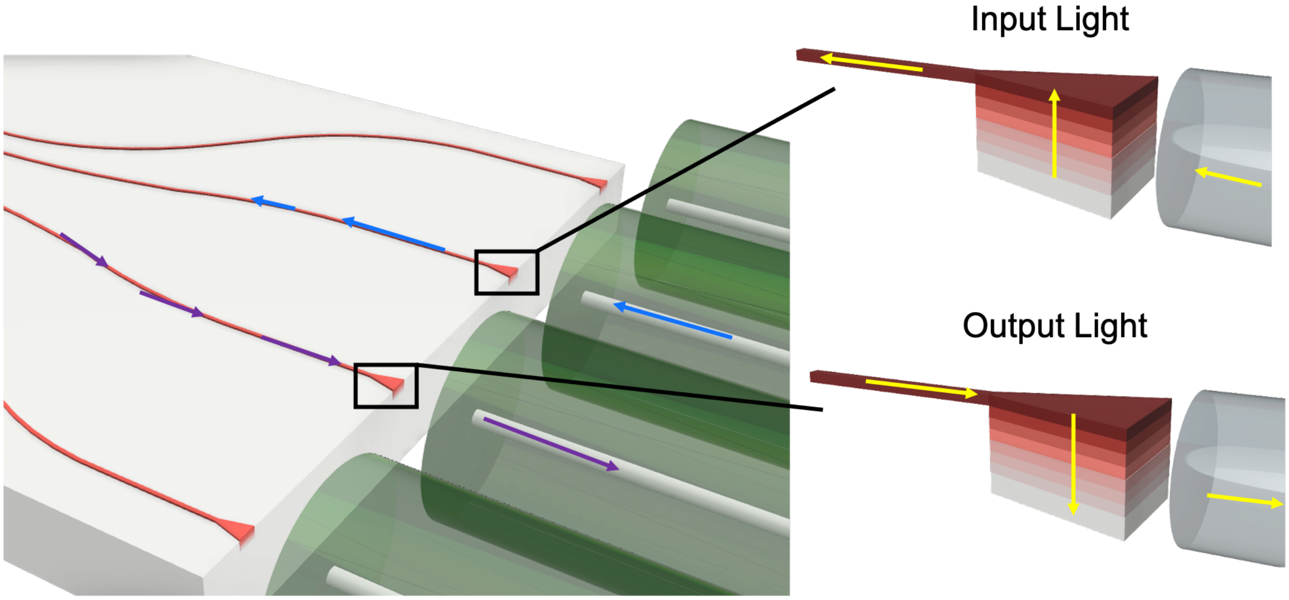


Figure 5.12: The bidirectional nature of the GRIN coupler. The blue lines on the chip waveguides indicate light coming onto the chip from the fiber, while the purple lines indicate light going from the chip into the fiber. Zoomed images are shown on the right for clarification, using yellow lines to indicate light (the colors were arbitrarily chosen for better visibility in the image).

Coupler	IL[1] (dB)	Mode	Size (μm)	1 dB TOL[2] (μm)	1 dB BW[3] (nm)	Comments	Year and Ref.
Nonlinear inverse tapers	1.37	TE	250	< ± 1	> 100	Smaller footprint vs normal taper, simple fabrication process.	2011 [64]
	2.13	TM	200				
	1.37	TE	170				
	2.12	TM	140				
	1.39	TE	170				
	1.87	TM	150				
Double-tip inverse taper	1.10	TE	40	± 0.75 (x) ± 0.5 (y)	>150	Lower IL & higher tolerance than single-tip, ultra-compact.	2016 [30]
	1.52	TM					
Trident-tip	<1.5	TE	300	±1.1	>35	Lower PDL than normal taper, feasible lithography process.	2015 [31]
	<1.7	TM					
SWG with SiN assists	0.75	TE	1500	±1.3	>150	High efficiency, complicated fabrication process, large footprint.	2016 [65]
Si taper under SiN rods & SiON cladding	0.5	TE	~500	±2.2 (x/y) ±2.4 (z)	\	Low IL, broadband, low PDL, but complex & difficult for mass production.	2016 [32]
	0.9	TM					
Cascaded claddings & cantilever	1.5	TE	240	±2	> 100	Good performance, but infeasible with overlapped tapers & cantilever.	2016 [66]
	2.1	TM					
Polymer cladded nano-taper	0.66	TE	300	< ± 1	> 100	Single taper and SU-8 polymer with EBL, simple design, high cost.	2010 [67]
	0.36	TM					
3D knife taper using double-patterning	0.35	TE	~180	\	\	Low IL with i-line stepper & angled sidewall dry-etching, complex fabrication.	2013 [34]
	0.21	TM					
Cantilever	1.0	TE	250	±2.75 (x/y) ± 2.14 (z)	>120	Complex fabrication, low feasibility & integrability.	2011 [68]
	1.5	TM					

Table 5.4: Overview of edge coupler configurations. Data and comments adapted from [37].

[1] IL = Insertion Loss
[2] TOL = misalignmen tolerance
[3] BW = bandwidth

5.1.3 Device Comparison and Discussion

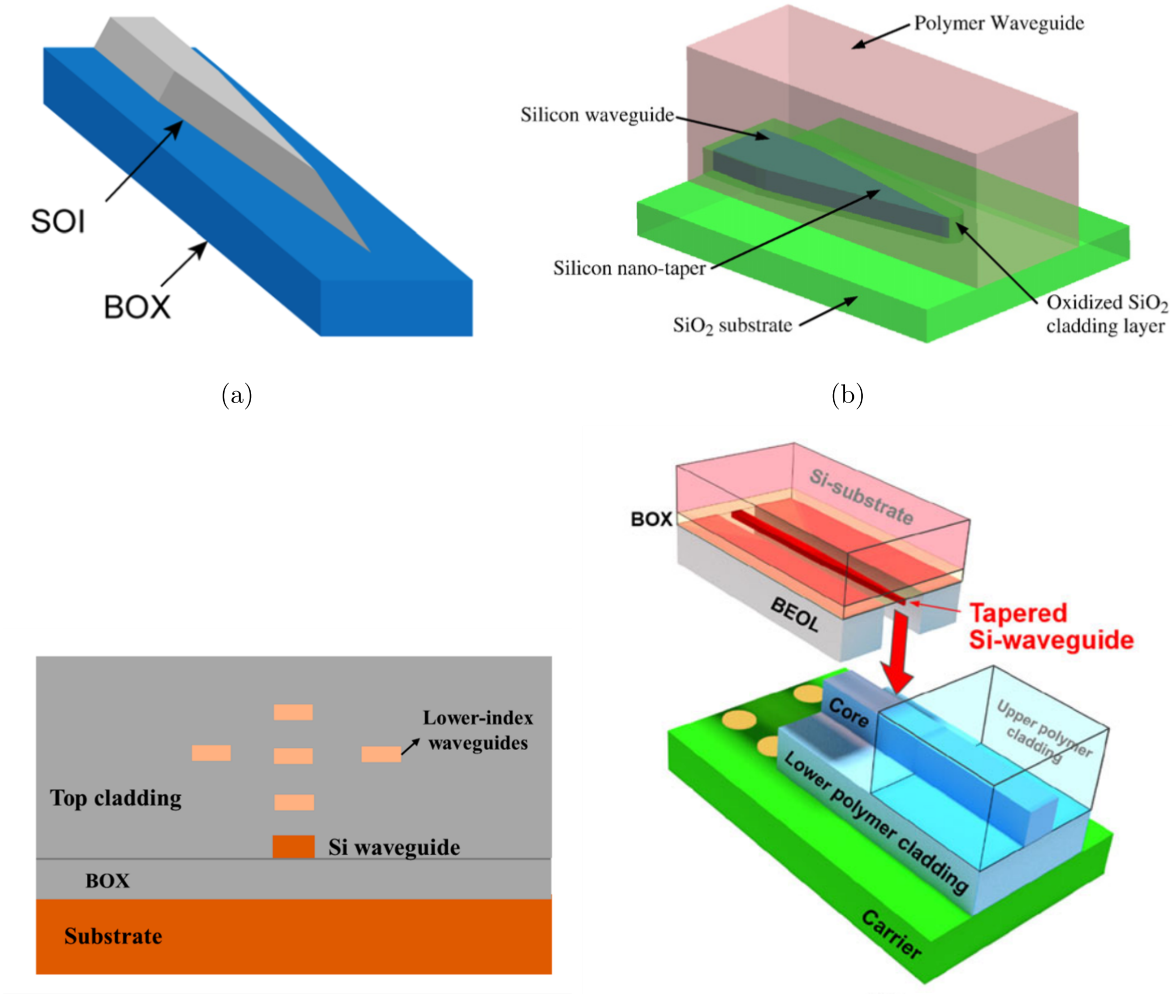
Through a review of edge couplers in Table 5.4 and the grating, evanescent, and free form couplers in Table 5.5, our simulated fiber-to-chip coupler stacks up well according to

Coupler	**IL (dB)**	**Mode**	**Size (μm)**	**1 dB TOL (μm)**	**1 dB BW (nm)**	**Comments**	**Year and Ref.**
OFFCHIP free form coupler	-0.18	TE	30 (x) 10 (y)	±1.1 (x/y) $>$ 25 (z)	>350	Excellent z tolerance, small footprint, not monolithic.	2020 [46]
	-0.25	TM					
Photonic-plug free form coupler	-0.6	TE	3000 by 2000	± 20 (x/y) ± 15 (z)	$>$ 60	Ultra-wide translational tolerance, mrad scale angular leveling necessary, not monolithic, large footprint	2020 [44]
Micro-lens free form coupler	-1.7	TE	250 (r) 1500 (l)	±25 (x/y)	$<$ 20	Ultra-wide alignment tolerance, pluggable fiber array with Legos$^{\text{TM}}$, very large footprint, non-monolithic	2017 [45]
Hybrid polyemer evanescent-edge coupler	-1.3	TE	37.5-250	±2	$>$ 100	Novel passive assembly, III-V chip integration, fine patterning of Si nanotaper, polymer reliability concerns.	2016 [69]
	-1.5	TM					
Hybrid polymer interposer coupler	-2	TE	>1000	$<$ ± 2	>300	Simple design, low fabrication complexity, thick polymer interposer may prevent mass production feasibility.	2017 [23]

Table 5.5: Overview of various grating, evanescent, and free form coupler configurations for coupling from SMF.

the performance metrics outlined in Section 3.3. First, in terms of quantitative performance metrics, our GRIN coupler display a coupling efficiency of 94% (0.27 dB of loss), 1 dB misalignment tolerance of 2.2 μm in y and 2.5 in z, and 1 dB bandwidth of greater than 100 nm at least. These are competitive compared to many of the listed couplers which have coupling losses greater than 1 dB, and 1 dB misalignment tolerances below 1.5 μm. Notable exceptions, or designs we are nearly equal to, include the 3D knife taper [34], the nano-taper with a polymer cladding [67], the Si nano-taper under multiple SiN assisting waveguides [32], and the polymer waveguide interposer concept from [23]. Images of these designs are shown in Figure 5.13 for reference and to put the following remarks in context. Here is where the qualitative performance metrics become important. The main remark is that the fabrication complexity of these designs is high due to lack of planarity or use of materials which either are not CMOS compatible or contain reliability challenges. For instance, the 3D knife taper requires fine control of sidewall angles during the etching steps in all three dimensions. In the case of the SiN assisted Si taper there is the need for multiple, different lithographic masks to be finely aligned to one another as the SiN assisting waveguides are deposited. While the alignment tolerances of aligning multiple masks are better than the alignment tolerances of fiber arrays in V-grooves attached to the photonic chip facet, making this design more practical, the mask alignment is not negligible. For a DUV system, mask alignment tolerance is typically on the order of 0.2 μm for projection aligners and 1μm for contact aligners [70]. In terms of the Si nano-taper with a polymer cladding, the presence of a narrow nano-taper tip width of around 40 nm will require more advanced patterning methods such as electron beam lithography. Plus this design relies on polymers which, generally speaking, contain reliability challenges. The polymers, included to reduce the index contrast between SMF ($n = 1.46$) and silicon waveguides ($n = 3.45$), suffer from unreliability under high humidity and high temperatures. This is a challenge in general, but is especially undesirable in photonic or optoelectronic packaging where flip chip bonding processes will eventually be used, induces thermal cycling. Likewise, if a flip chip bonding method is used such as thermo-compression bonding where a vertical force is used on the top chip during bonding, the lower polymer based chip would need to be mechanically stable to avoid deformation which might present challenges. Additionally, if the coupler's primary usage is outside the datacom/telecom market, such as in automated vehicles or the IoT, environmental parameters such as moisture may be present challenges for the use of polymers. Lastly, some polymers such are opaque in the telecom C band (λ=1.53-1.565 nm) due to absorption in their vibrational energy levels associated with C-H bonds [38, 71], making their inclusion detrimental for telecom applications. Note that the polymer used in the evanescent coupler developed in [23] is CMOS compatible and UV cured and thermally baked at 200° to provide better long term stability, also demonstrating the thermal budget can handle solder reflow temperatures which typically are close to 217°, the eutectic temperature of Sn-3.5Ag. Also, the specific polymer used in [23] has low absorption near 1310 and 1550 nm, showing that this generalization regarding the reliability issues of polymers is not always the case for optical couplers.

On the other hand, due to the material choice of SiON for the stack layers and an SOI based chip, this design can be processed using a standard CMOS compatible process flow for Si, while maintaining the reliability of a silicon materials platform. In addition, the GRIN coupler utilizes large feature sizes (the minimum feature size is the taper tip width, on the order of 0.2 μm or greater), making our fabrication complexity relatively low. Note that the primary fabrication roadblocks in our design are due to the GRIN stack thickness, material composition, and the sidewall angle. In terms of material composition, the choice of SiON or SiN layers means the potential for increased absorption loss due to N-H bonds. However, the processing optimization conducted in [17] shows the percentage of N-H bonds for low temperature PECVD can be reduced to below 5%, a significant improvement over previous fabrication methods. This is also taken into account in our design modifications by switching to a lower index than was previously used for the upper index limit of the GRIN, providing a lower nitrogen ratio and, in theory, fewer numbers of N-H bonds. The other two factors, GRIN stack thickness and the sidewall angle, were discussed in Section 5.1. The conclusion was that improved processing (mainly the switch in etching precursors through the use of trifluoromethane and oxide deposition to compensate for wafer bowing) resulted in an improved sidewall angle of 94° and the ability to deposit large thicknesses (greater than 6 μm), improving measured coupling efficiency to be nearly that of simulated values. This claim is supported by evidence, mainly that 3D FDTD simulations show a theoretical maximum coupling efficiency for the optimized design presented in [49, 63] of approximately 89% (both the lens and the flat facet versions are close to this number because recall that the



(a) (b)

(c) (d)

Figure 5.13: Relavent fiber-to-chip coupling alternatives close in terms of performance metrics to our GRIN coupler. In (a) a 3D knife taper from [34], in (b) a polymer cladded silicon nanotaper from [67], in (c) a silicon nanotaper assisted by mode expanding SiN waveguides [32], and in (d) a polymer waveguide on a polymer interposer coupling light from a single mode fiber [23]. Note that in (d) the silicon waveguide is on a separate chip, so the image actually shows two separate couplers - the evanescent coupler from polymer interposer to PIC and the fiber-to-chip polymer cladded edge coupler.

lens only adds around 0.5% efficiency). Meanwhile, the experimental coupling efficiency was measured to be 81% for the flat facet and 90% for the lensed facet in [62] following process optimization. Thus, these concerns are mitigated through modest design changes and the process optimization shown in prior work.

Returning to comparison of other performance metrics, our widened alignment tolerance of 2.2 μm in y and 2.5 in z is competitive as well, although the SiN assisted Si taper and cascaded cladding mode converters both showing alignment tolerances outside of 2 μm as well. However, on this point it should be noted that the 1 dB alignment tolerance performance marker can be somewhat misleading, because it details loss relative to the coupler's zero misalignment insertion loss rather than a standard value. Therefore couplers may have a wide 1 dB alignment tolerance, but could also start with a lower insertion loss also, as is the case with the cascaded cladding mode converter which begins with an insertion loss of 1.5-2.1 dB. It is also important to note when referring to alignment tolerance the importance again of fabrication complexity, specifically when comparing to free form couplers. Free form couplers, such as those developed in [44, 45] and [46] (shown in Figure 3.2, 3.3, and 3.4 for reference and context) rely upon micro-lenses and micro-mirrors with curved shapes and often requiring different materials. This is in contrast to our monolithic, planar design which is more ready for immediate mass production.

5.1.4 Proposed Future Work

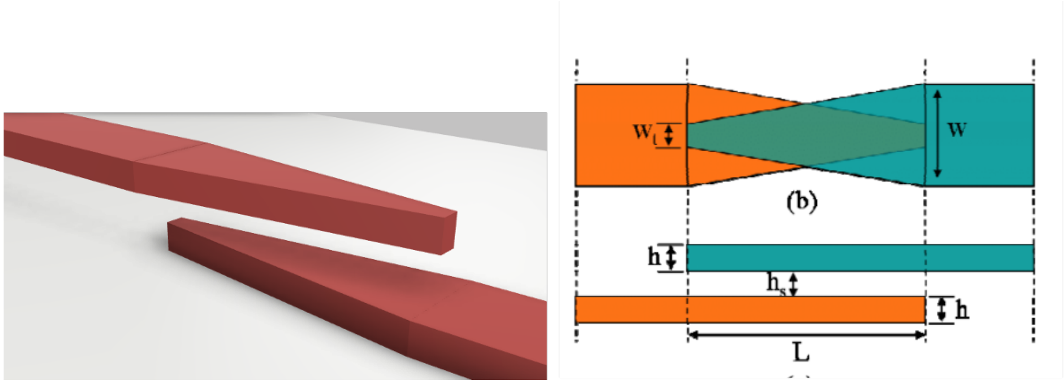
Immediate next steps for the GRIN fiber-to-chip coupler include design of the mask and the material characterization for SiON of varying oxygen concentration. The mask will be relatively simple considering the planarity of the system, the only complexity stemming from the need for multiple layers. Material characterization and optimization of the process for SiON GRIN structures was accomplished in [17]; however, part of the design modifications involved altering previously constrained parameters in order to improve efficiency, tolerances, and be more integrable with standard SMFs. Because parameters such as the GRIN output waveguide dimensions and index, GRIN thickness and index profile, number of layers, and final output waveguide material take different values, process changes will be necessary. Additionally, the process developed in [17] was over ten years ago, meaning a similar, but modestly modified process flow will be necessary to create. The past decade has come with improvements in deposition tools and PECVD techniques which will hopefully aid in the endeavor to avoid stress buildup, sidewall sloping, high N-H bond ratios, and other material parameters of interest measured during the characterization period. Characterization will also include propagation loss measurements of the final SiON waveguides, using waveguides deposited with different refractive indices and dimensions and measuring them via the cutback method, spirals, paperclips, and ring resonators. Of course, there are also a plethora of design aspects which could be explored as well such as how a nonlinear GRIN taper affects efficiency, how the addition of an anti reflective coating could assist in reducing back scattering at the fiber-chip interface, or how a different substrate such as glass could provide stress relief and reduced wafer bowing during deposition of thicker GRIN stacks. There is also the idea that Bragg mirrors could be added on either side of the GRIN to reflect uncoupled light during misalignment and improve efficiency, which could be explored in future work.

Chapter 6

Chip-to-Chip Coupling

6.1 Design Overview

Once light is on the lower chip, we will use evanescent coupling to transfer the signal onto the upper chip. Before discussing our novel design for chip-to-chip coupling, we will give a general overview of the solution used to vertically couple light between different planes on the same chip, including prior designs and how they work, followed by how we will modify them for robust chip-to-chip coupling. In general, co-planar designs with SOI substrates are typically used to interconnect components in the same plane. However, if all integration occurred in the same plane, the pattern density would be limited and it would create restrictions for scaling. Connection in the vertical dimension, allowing for 3D photonic integration, can significantly increase the device density on a single wafer or chip. The simplest initial design for a vertical coupler is based on a directional coupler for two waveguides, but with the waveguides stacked vertically instead of horizontally side by side. From Section 3.1.4, we know that for a conventional directional coupler with two adjacent, parallel waveguides 100% power transfer requires zero phase mismatch, meaning the effective indices of the two waveguides are identical throughout the length of the waveguide, and the coupling length is exactly equal to the transfer distance at which 100% power transfer is complete. This design is not robust though because achieving identical effective refractive indices throughout the length of the waveguide requires identical intrinsic refractive indices if the waveguides have identical cross sections along their length. This is challenging to do in terms of fabrication due to the waveguides potentially being composed of two separate materials, for example amorphous Si (a-Si) and SOI, and due to slight variation in waveguide geometry that comes with fabrication error. In addition, vertical misalignment, an important factor in chip-to-chip alignment, induces a different coupling gap, meaning the coupling efficiency is no longer at a maximum for a given coupling length. A more robust design includes the use of tapers on both the upper and lower waveguides to ensure effective index matching occurs. Both a perspective view and a side view the vertical cross tapers are shown in Figure 6.1(a) and 6.1(b), respectively [72].



(a) (b)

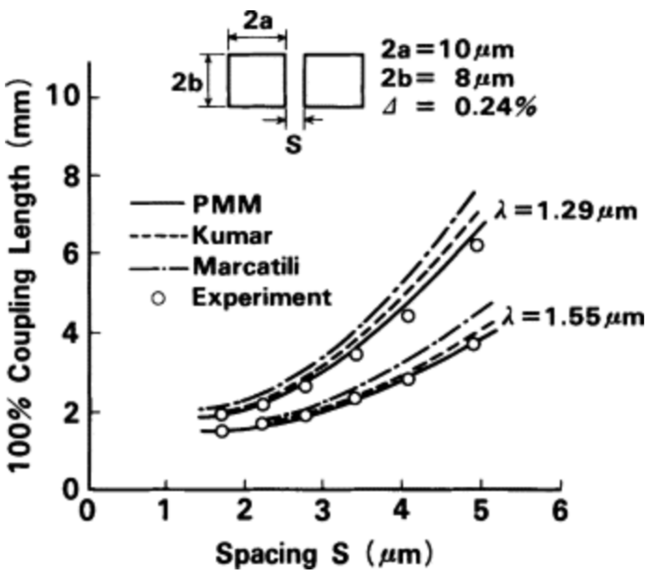
Figure 6.1: Vertical cross taper coupler design. In (a) our vertical cross taper, with (b) being the same structure but a top view [72].

This type of design also reduces oscillation of power between waveguides or mode beating, showing very little coupling of power back into the input waveguide. Another advantage is that the design is its bidirectionalty due to its symmetric nature, meaning coupling of light from the bottom chip to the top chip and vice versa does not require an alternate design. Different variations of this design have been developed [72, 73, 74, 75, 76], and our proposed design is itself a variant of the couplers fabricated in [72]. The driving force for these alternative designs was to increase the vertical coupling gap. For example, modifications of [72] were made achieve high coupling efficiency under the constraint that a waveguide to waveguide spacing greater than 1 μm is necessary, such as when cross talk is to be avoided between two chips or layers on the same chip These modifications were necessary because $L_c \propto \exp d/\alpha$ where L_c is the length for maximum coupling, d is coupling gap, and α is the extinction coefficient in the gap region. Thus, for larger vertical coupling gaps an exponentially longer taper length is necessary in order to achieve adiabatic coupling saturation. This general idea is demonstrated by Figure 6.2 from [77]. Note that this is not to be confused with expanding the Z misalignment tolerance, because an optimized taper for 1 μm coupling gap will still experience decreased coupling efficiency when the coupling gap shifts from its optimized value.

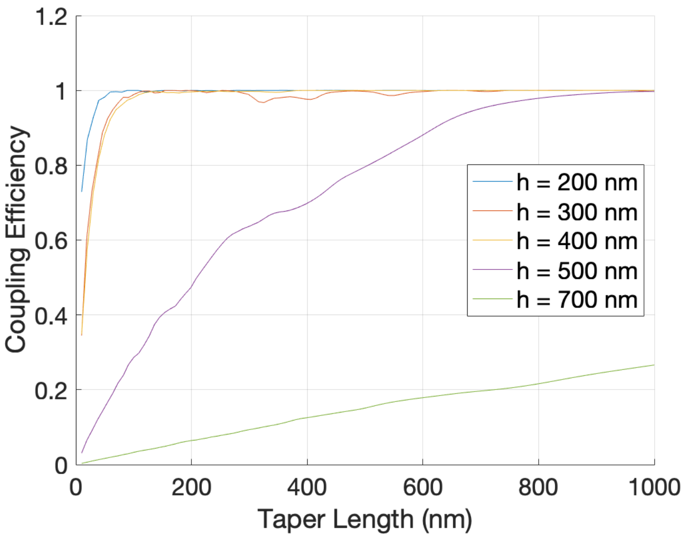
The prior fabrication of the adiabatic, inverse cross tapers demonstrated consistent coupling losses of 0.20 ± 0.05 dB (95% coupling efficiency) [72]. The experimental data also demonstrated at least 100 nm broadband tolerance, shown by the responsivity results of Figure 6.3(b). To define the SOI and a-Si tapers which had varying tip widths of 200, 250, and 300 nm, deep UV (DUV) photolithography was used in order to achieve the accuracy necessary for resolution. However, even with the enhanced resolution, tip widths less than 200 nm were not created, despite simulations indicating going below this limit would provide further increase in coupling efficiency for the same coupling length.

6.1.1 Novel Chip-to-Chip Evanescent Coupler

As stated in Section 6.1, our design begins with that of [72], but quickly implements significant design changes. The design parameters of the vertical cross tapers include the refractive indices, material composition, tip width, input waveguide width, taper length, waveguide and taper thickness, cladding material, and spacing between the upper and lower tapers. Note that for all of the design parameters aside from taper spacing, the parameters can be asymmetric, meaning the lower and upper tapers can have different values for each parameter should this show improvement in coupling efficiency, alignment tolerance, band-



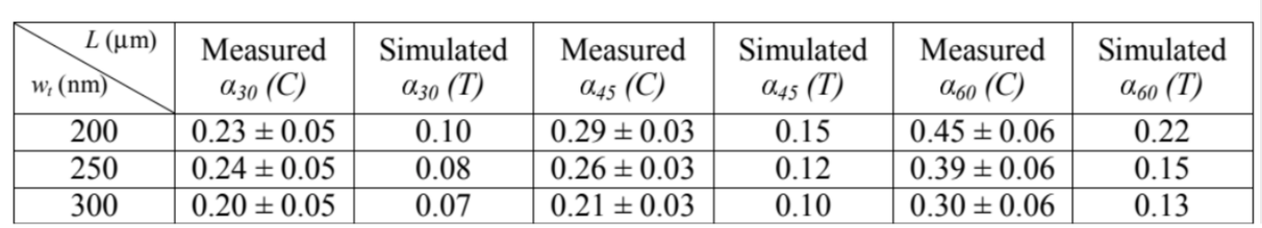
(a)



(b)

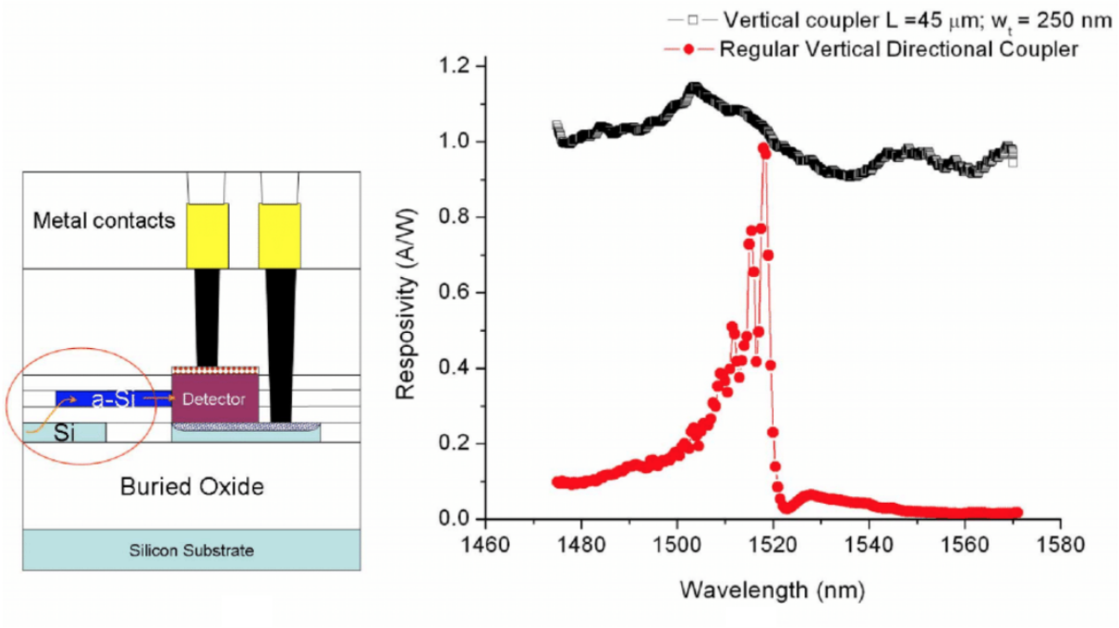
Figure 6.2: Coupling gap relationship with maximum efficiency coupling length. In (a) the relationship is shown for a standard directional coupler with no tapers, with the image taken from [77]. In (b) the same relationship is shown for our cross taper structure. This problem is mitigated in the $\mathrm{Si}_x\mathrm{O}_y\mathrm{N}_z$ configuration, which has a wider coupling gap tolerance.

width, etc. While fiber-to-chip coupling involves standardized fiber sizes to adjust coupler parameters for, PICs, even in silicon photonics, can vary from one to another in terms of their waveguide dimensions, cladding and passivation layers, materials used, etc. This means



L (μm) / w_t (nm)	Measured α_{30} (C)	Simulated α_{30} (T)	Measured α_{45} (C)	Simulated α_{45} (T)	Measured α_{60} (C)	Simulated α_{60} (T)
200	0.23 ± 0.05	0.10	0.29 ± 0.03	0.15	0.45 ± 0.06	0.22
250	0.24 ± 0.05	0.08	0.26 ± 0.03	0.12	0.39 ± 0.06	0.15
300	0.20 ± 0.05	0.07	0.21 ± 0.03	0.10	0.30 ± 0.06	0.13

(a)



(b)

Figure 6.3: Performance and experimental setup for testing cross tapers in [72]. In (a), a table showing simulated loss versus measured loss for varying tip widths and taper lengths. Note that the measured loss is the total loss, and is primarily due to scattering at the finite tip as opposed to intrinsic coupling loss, which was simulated to be above 99%. The setup in (b) was used to measure a broadband response, where the a-Si/c-Si cross tapers were used alongside a standard vertical directional coupler in identical setups to couple light into a GeSi photodetector. The GeSi photodetector's measured responsivity is proportional to the coupling efficiency of the system, demonstrating a distinct peak at 1520 nm for the taperless directional coupler. However, the adiabatic, inverse cross tapers exhibited broadband characteristics between 1470 - 1570 nm.

it may also be important to ensure the design is robust enough to handle the situation where one of the tapers is constrained due to other components of the photonic chip. For example, one of the chips may have the necessity that the outermost waveguiding layer be crystalline Si (c-Si) or amorphous silicon (a-Si), meaning that that chip's tapers would have constraints on all parameters due to silicon's HIC. We will demonstrate the effect of symmetric and asymmetric parameters here in an effort to cover different situations which may arise in implementing this type of evanescent coupler. We will also present a solution to the scenario where the primary waveguiding layer must be constrained to HIC silicon waveguides and light is to be coupled chip to chip. Note that in [72], the details and explanation of how taper length, tip width, and very slight variation of refractive index asymmetrically (using a-Si versus c-Si, with $\Delta n \approx$ 0.00-0.22) affect coupling efficiency are presented and those arguments will not be re-printed here, just mentioned as fact with the reader encouraged to refer to the reference for more details.

Out of all the listed design parameters, we will only constrain the cladding material to be that of SiO_2 with $n = 1.45$ for our simulations. This means the background index for all simulations is set to 1.45. In terms of substrate leakage, this is reasonable because our materials are silicon based and can be fabricated on 3 μm SOI wafers which use silica as a cladding layer from the Si bulk substrate. The only time this may need to be revisited is in the case when the taper material and geometry being used expand the beam to such a point that it may overcome the approximately 3 μm gap and leak into the Si. This case arises for our LIC evanescent coupling, and in these cases a bulk substrate will need to be included despite the increase to simulation time.

Aside from substrate leakage concerns, using a background index of 1.45 may also seem incorrect since there are two separate chips and an air gap may arise between the surfaces which should need to be incorporated into the simulations. An air gap is expected to introduce loss into the system in two ways: via reflections off of the upper and lower chips, and through the creation of an optical cavity which would cause coupling efficiency to oscillate as in the case of an air gap between fiber and facet for the fiber-to-chip coupler (see Figure 5.10(c) and associated discussion). However, it is very common to implement curable UV epoxies and other underfill materials between chips in a flip chip bonded system, and these materials are wide available with tunable indices between 1.4-1.6. In our case we would select an epoxy near that of silica as done in [62] where $n = 1.4587$. In other words, it is assumed the chips are either in direct contact with one another (accomplished, for example, by etching a ledge into the lower chip or substrate as done in [78, 69, 79, 80]) or an intermediate layer provides index matching capabilities, meaning there is no air gap between the upper and lower chips.

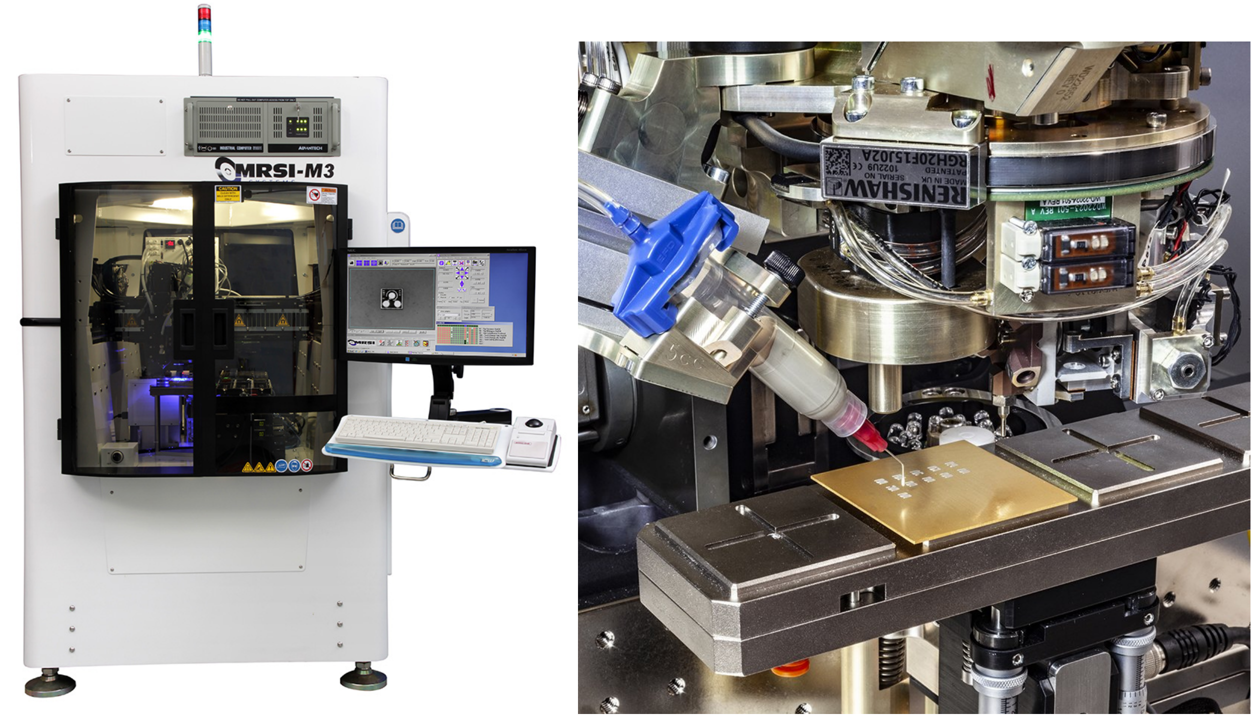
Other self-imposed constraints on the design of the taper include the compatibility with high speed lithography tools and lower exposure resolutions. This constraint will yield a scalable solution - the final chip-to-chip coupling design should be able to be patterned with tools used in mass manufacturing compared to slower, yet highly accurate lithography tools used in academic research spaces such as electron beam lithography (EBL) equipment. Per the IPSR-I 2020 “Silicon Photonics” chapter, most silicon photonics chips with thin/thick SOI and bulk Si, use 248 nm DUV, 193 nm DUV and 193 nm immersion lithography technologies. These lithography options enable feature sizes of sub-100nm, with the critical dimension (CD) outlined by the following equation:

$$\mathrm{CD} = k_1 \frac{\lambda}{\mathrm{NA}} \tag{6.1}$$

where k_1 is a coefficient that encapsulates process-related factors (typically equals 0.4 for production with a theoretical minimum of 0.25), λ is the wavelength of exposure, and NA is the numerical aperture of the lens as seen by the wafer. Plugging in reasonable NA’s into Equation 6.1 yield typical resolutions for 248 and 193 nm DUV lithography on the order of approximately 50 nm. While these types of resolutions are possible using a method employed in mass manufacturing, it is desirable to increase the minimum feature size of our coupler to even larger values in order to utilize even faster lithographic tools, such as the MLA150 direct write tool located at MIT’s MIT.nano facility. The MLA150 uses a 375-405 nm laser for exposure and, depending on the installed “write mode”, allows for a minimum feature size between 0.6 and 1 μm. The tool is significantly faster, at least for research and development purposes, compared to DUV lithography because it does not require a hard mask for exposure and operates based on projection lithography, writing the mask directly based on the uploaded .gds file.

In terms of novelties compared to past designs, there are several. First and foremost, the fact that our new design will demonstrate chip-to-chip evanescent coupling is a critical difference which presents unique challenges. As mentioned in Section 3.2, the most imposing challenge to face is misalignment between the two tapers in all 6 degrees of freedom (translational and rotational) due to flip-chip assembly. Since alignment previously was between tapers in different layers but on the same chip, the alignment errors were governed entirely by mask misalignment during lithographic process steps which are typically less than 0.1 nm; in fact, there was no investigation of the lateral, vertical, or rotational misalignment tolerances at all. However, for flip-chip assembly, high speed pick-and-place tools are typically used for alignment, positioning, and bonding. Many commercially available tools have alignment accuracies outside of ± 1 μm, meaning that the tool cannot guarantee sub-micron passive alignment. Examples including the MRSI M-3 in MIT.nano's LEAP packaging facility, shown in Figure 6.4(a), or Palomar's 6532 Die Bonder, shown in Figure 6.4(b), coming in at ± 3 μm and ± 1.5 μm quoted alignment accuracies, respectively [81, 82]. The goal of the chip-to-chip coupler design changes is to expand the alignment tolerance so that it can fall within the alignment accuracy of the die bonder, again resulting in a solution that is scalable to mass manufacturing.

With these constraints, novelties, goals, and challenges in mind, we consider the system we will use as a reference before we test the effects of modifications. As a control and reference system, we will use the Si/Si (read "silicon on silicon") system described in [72]. In this system, the Si/Si waveguides have input widths of 500 nm in order to ensure single mode transmission; therefore, the misalignment tolerance for 1 db excess loss has been simulated to be around ± 0.2 μm in the y/z direction and greater than ± 60 μm in the x direction. For the y direction, as will become clearer once the HIC silicon system is switched to a LIC system, the misalignment tolerance is governed primarily by the expansion of the mode



(a) (b)

Figure 6.4: State-of-the-art die bonders. In (a) the MRSI-M3 tool [81] and in (b) the Palomar 6532H tool [82]

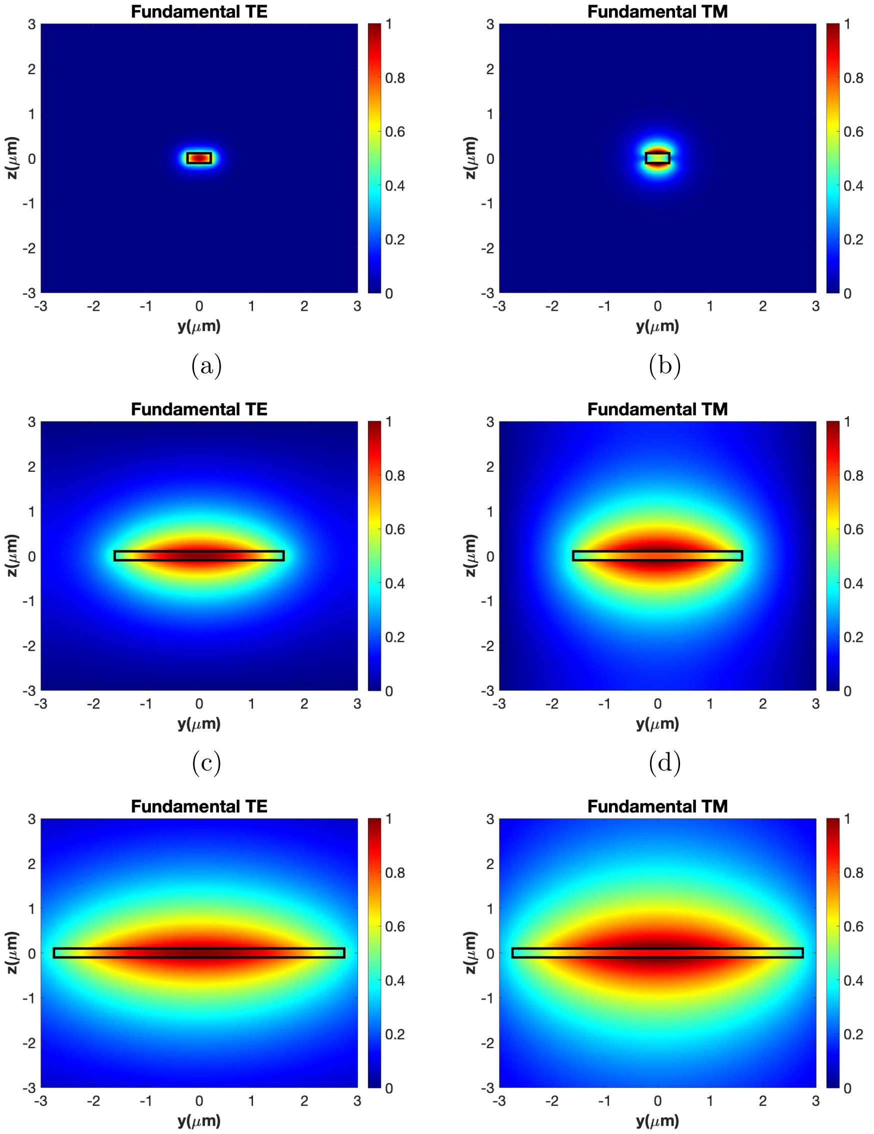
from the waveguide as opposed to the percentage of overlap for the two tapers. For the Si/Si system with a tip width of 250 nm and input waveguide width of 500 nm, even at ± 200 nm y offset the two tapers are still overlapped by 33% in terms of area (calculated by the taking the area of overlap divided by the area of the trapezoidal taper). However, because silicon has HIC in comparison to the silica cladding with a large confinement factor of 77%, the modal overlap is small for small relative offsets, leading to a small misalignment tolerance. For the z direction, the misalignment tolerance is governed by the exponential tail of the evanescent mode, as described in Section 6.1. Recall this type of tolerance can be overcome through longer tapers, but the necessary taper length exponentially increases with the increasing coupling gap. This again is exacerbated by the HIC nature of silicon - the mode has high confinement and the evanescent tail barely extends past the surface into the coupling gap. This is yet another tolerance which can benefit from modal expansion and using a LIC material. The x misalignment, on the other hand, is very robust due to the basis of coupling being effective index matching. As long as the point where the effective indices of refraction match remains overlapped by the other taper, efficient coupling will still occur. For symmetric tapers, this means that the tapers just need their tips overlapping to have efficient coupling, yielding well greater than 60 μm x tolerance. For asymmetric tapers, the point of coupling will depend on the index contrast between the upper and lower tapers, meaning for tapers with similar refractive indices there is still a large x tolerance, but for systems with a large Δn this tolerance is no longer negligible.

Finally, rotational misalignment tolerances are also shown in Figure 6.8, where the axis in the legend corresponds to the axis of rotation and a right hand coordinate system is assumed. Note that rotational misalignment tolerance for rotation around the x axis is not shown because this alignment tolerance is essentially negligible (while the taper does not have symmetry on under this rotation, its form is very similar when rotated even by large angle about the x axis. This means the effective index is approximately the same and coupling efficiency varies very little under this rotation). The rotational alignment tolerances can be understand by thinking about how rotation is similar to each point along the taper under going a shift in x, y, and z. For example, a rotation about the z axis of the upper taper will result the upper taper tip shifting away from the lower taper in x and y and upper taper input and waveguide shifting away from the lower taper tip in x and y. This means light is now misaligned, and must undergo a directional change in the x-y plane as well depending on the angle of misalignment. While rotation about the y axis also results in shifts away from the other taper, this time in z and x, there is no additional in plane directional change needed (light needs escalate vertically, not horizontally). This makes rotation about the z axis the primary rotational alignment tolerance limiter. Considering both the translational and rotational alignment tolerances together, it's clear modifications of design parameters are necessary in order to obtain a robust coupling solution - these modifications will be presented in the following sections.

6.2 $\mathrm{Si}_x\mathrm{O}_y\mathrm{N}_z$ Adiabatic, Inverse Cross Tapers

The narrow ± 0.2 μm range for the Si/Si system can be increased by switching to a $\mathrm{Si}_x\mathrm{O}_y\mathrm{N}_z$ (abbreviated SiON for the remainder of this work) material system with a lower refractive index. Specifically, the refractive index typically ranges from 1.5-2.2 depending on the ratio of oxygen and nitrogen precursor gases during deposition. This change in material offers two key advantages. First, because of the low index contrast present in the system if SiO_2 is the cladding (Δn can be as low as approximately 0.05), the fundamental mode is significantly expanded to a larger MFD. In turn, this leads to larger misalignment tolerances, because a larger MFD means there is a larger physical region where the optical modes of the upper and lower waveguides can overlap and couple. This concept can also be visualized in Figure 6.5 by looking at the relative size of the MFD of the silicon waveguide compared to different silicon oxynitride waveguides. Second, SiON waveguides allow for a relaxing of the dimensional constraints originally placed on the tapers when crystalline or amorphous silicon was used. For the lower index system, larger possible input waveguide widths are possible while maintaining single mode propagation. If numerical software is available, a fast way to determine the the largest waveguide width constrained by single mode condition is to calculate the effective indices of the modes and determine when the first higher order mode's effective refractive index is greater than the claddings index - this is the cutoff for guided propagation of that higher order mode. This idea of widening the geometry can be verified by Figure 6.5 which shows the mode in a Si waveguide of 440 nm width and 220 nm thickness, compared to a stoichiometric silicon oxynitride waveguide of refractive index 1.7 which has a thickness of 200 nm and width of 3.2 μm as well as a SiON waveguide with $n = 1.6$ and width 5.5 μm. The combination of the geometry with the natural mode expansion due to material change, results in a confinement factor which is significantly reduced for the SiON waveguide, with values of 19% and 13% for the $n = 1.7$, 3.2 μm wide and $n = 1.6$, 5.5 μm cases respectively.

However, if the refractive index, input waveguide width, and taper tip width are allowed



(a) (b)

(c) (d)

(e) (f)

Figure 6.5: Cross sectional mode profiles for Si and SiON strip waveguides calculated using finite domain eigensolver (FDE) in Lumerical MODE. In (a) and (b) the fundamental TE and TM mode profiles of a 440 by 220 nm single mode Si waveguide. In (c) and (d) the same two modes are shown for a 3.2 μm by 200 nm stoichiometric SiON strip waveguide with $n = 1.7$. In (e) and (f) the modes are shown for a 5.5 μm by 200 nm SiON waveguide with $n = 1.6$. These widths are the largest widths possible for a 200 nm SiON strip waveguide in a SiO_2 cladding while maintaining the single mode condition.

to vary as described above, then the taper length must be adjusted in order to meet adiabatic requirements which will allow for low loss from mode conversion, lower wavelength dependency, and polarization insensitivity. In order to determine this length, one could determine the propagation constants for the silicon oxynitride waveguide dimensions and use Equation 2.30 to determine a suitable length. However, a more practical method is to perfectly align the tapers in Lumerical and use EME to calculate coupling efficiency as a function of taper length for the overlaying tapers. Coupling efficiency is expected to increase rapidly as taper length increases and the mode is allowed to convert shape and size more gradually; however, there will be a length where the coupling efficiency saturates and this will correspond to the adiabatic length of the taper. Because the EME method is suited extremely well for determining how changes in geometry in the propagation direction affect performance parameters, this simulation can be done very quickly and efficiently. The results are shown in Figure 6.6, demonstrating that around 85 μm for $n = 1.7$ and 255 μm for $n = 1.6$ for the taper length the coupling efficiency saturates and this length can be used and assumed to be adiabatic. While there are still mild oscillations occurring due to increased mode beating for wider tapers, this oscillation only causes the coupling efficiency to drift less than 0.5% from its average value of 99.3% efficiency beyond these lengths. In fact, if the gain in misalignment tolerance is considered to be a greater benefit than this smaller change in coupling efficiency, the length can be reduced to approximately 45 μm and 185 μm if a fluctuation of around 1.5% is allowed from an average efficiency of 98.3%. Also, notice that increasing the lateral dimensions of the input waveguide width results in an increase in the necessary adiabatic length. This makes sense - a longer taper length is needed to ensure that mode conversion occurs slowly in the case where a large input waveguide width exists and the taper width is relatively unchanged. The final design parameters of the system are reported in Table 6.1.

To study the improvement in misalignment tolerance, we will first consider the scenario where both tapers are identical (i.e. both tapers have been shifted to silicon oxynitride), followed by the scenario where the upper taper is constrained to be silicon while the lower taper is varied to that of silicon oxynitride. Lastly, we will consider the situation where both the input and output waveguides are silicon, but an intermediate structure is used to mode expand and improve tolerances. For the SiON/SiON system, the translational misalignment

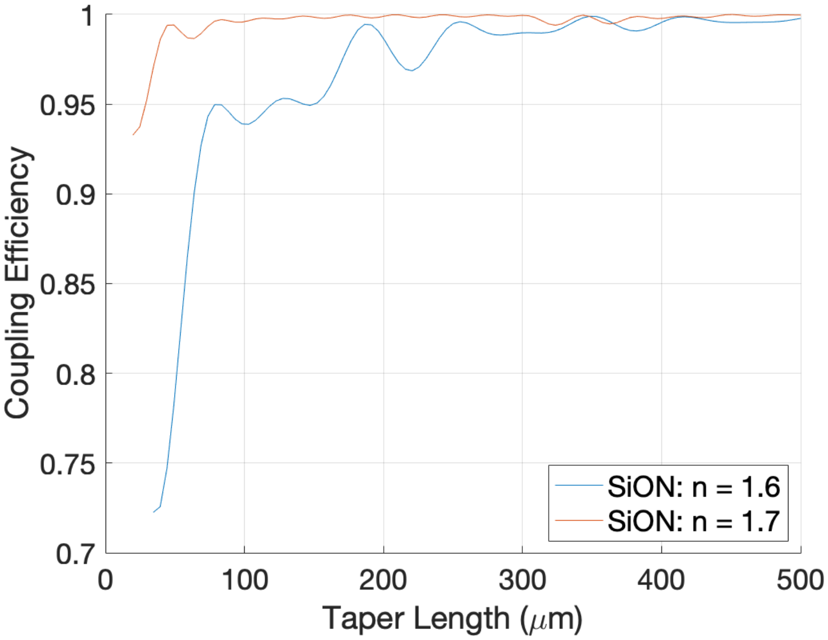


Figure 6.6: The adiabatic length for SiON tapers of refractive index 1.7 and 1.6. The thickness of the waveguide is held at 0.2 μm while the width is increased to the maximum allowed while only allowing one guided mode to propagate. For $n = 1.7$ this is 3.2 μm and for $n = 1.6$ this is 5.5μm; thus, it makes sense that the adiabatic length for the lower index taper is much longer than that of the higher index taper, since the input width is wider.

tolerance in the lateral and vertical direction is shown in Figure 6.8(b) and 6.8(c). In these plots, 3D FDTD simulations demonstrated that shifting from Si to SiON for both tapers has a minimal effect on intrinsic coupling efficiency at zero misalignment, shifting from 99.95% to 99.5% (for both $n = 1.6$ and $n = 1.7$). This is countered by a coupling efficiency above 80% beyond ± 2-3.5 μm in y and ± 1.5-3.2μm in z for $n = 1.7 - 1.6$ (this is 1 dB excess loss). In terms of bandwidth, this device has low wavelength dependance as evidenced by the simulations in Figure 6.8(a) which show a 1 dB bandwidth well above 300 nm. For polarization dependent loss, the coupling efficiency of the TE mode is 99.9% while the coupling efficiency of the TM mode is 99.1% for the $n = 1.7$ case, for example, showing less than 1% loss between states.

For the Si/SiON system, the translational misalignment tolerance in the lateral and vertical direction is shown in Figure 6.8(b) and 6.8(c). In these plots, 3D FDTD simulations

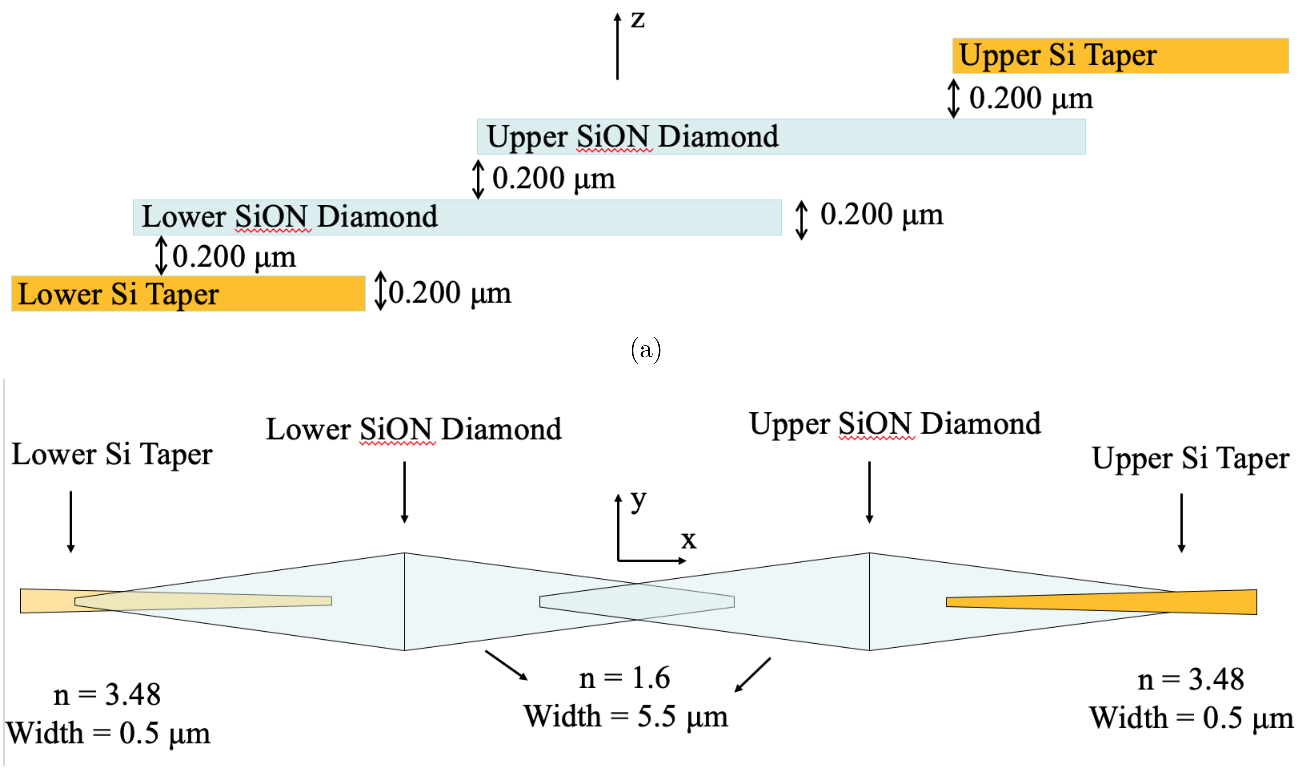
Parameter	**Si/Si**	**SiON/SiON**	**SiON/SiON**	**Si/SiON**
Bottom waveguide width (μm)	0.5	3.2	5.5	5.5
Top waveguide width (μm)	0.5	3.2	5.5	0.5
Taper length (μm)	60	100	300	300
Taper thickness (μm)	0.2			
Bottom taper index	3.48	1.7	1.6	1.6
Top taper index	3.48	1.7	1.6	3.48
Spacing (μm)	0.2			
Taper tip width (μm)	0.2			0.1 (upper) 0.2 (lower)
Total thickness (μm)	0.6			

Table 6.1: Final parameters for adiabatic, inverse tapered evanescent coupler using 3D FDTD simulations in Lumerical. Note that the material on the left indicates the upper taper material and the material on the right indicates the lower taper material (i.e. SOI/SiON means a c-SI upper taper and SiON lower taper).

demonstrated that shifting from SiON to Si for the upper taper barely affects intrinsic coupling efficiency at zero misalignment by shifting from greater than 99.95% to 99.7%. This system is still able to maintain an impressive coupling efficiency above 80% beyond ± 1.5 μm in y and ± 1 μm in Z (this is 1 dB excess loss). In terms of bandwidth, this device has low wavelength dependance as evidenced by the simulations in Figure 6.8(a) which show a 1 dB bandwidth well above 300 nm. For polarization dependent loss, the coupling efficiency of the TE mode is 96.4% while the coupling efficiency of the TM mode is 95.1%.

Finally, the last system we will present is the configuration where the input and output waveguides are constrained to be HIC materials with narrow dimensions, such as a SM silicon waveguide with 0.2 x 0.5 μm^2 thickness and width. In effect, the solution proposed here is a combination of the systems already described, mainly the $n = 1.6$ SiON/SiON system. The structure will have four elements: the input waveguide, the output waveguide, and a beam expansion component on each chip. The beam expansion component is a diamond shape meant to widen the alignment tolerances by expanding the mode in the region where packaging misalignment is likely to occur. While the overall shape is that of a diamond, it can actually be considered two adiabatic tapers back-to-back, such that the wide side of the first taper is butt coupled to the wide side of the second taper. The general idea is that light travels into the SM silicon waveguide, then couples into the first taper of the first diamond with high efficiency as has already been detailed (this is just the Si/SiON system). The light then immediately travels into the next taper in the same plane while undergoing mode expansion. The now significantly larger mode then evanescently couples to the upper chip via another set of overlaying cross tapers, where it then travels to the final coupling section - a set of overlaying cross tapers with the final upper taper being that of silicon. Through this design, the beam expansion component of either chip is aligned to the silicon component with the only source of misalignment being fabrication misalignment. Then, once the two chips are packaged, the components actually doing the coupling are the diamond structures on either chip, which have significantly larger modes, and thus wider tolerances in all three translational directions. A schematic of the design is shown in Figure 6.7 with a top and side view for reference. The y misalignment tolerance of the structure is shown alongside the other systems in Figure 6.8(b), and the z alignment is taken to be comparable as well while these simulations are ongoing.

For all of these systems, even the asymmetric case where Si sits on SiON, due to the fact that the evanescent coupling of the mode relies on effective index matching (as depicted in Section 3.1.4) the coupler is inherently bidirectional, operating identically from light traveling from SiON to Si as from Si to SiON. This is notable and not necessarily intuitive, because it seems slightly contradictory that the mode would tend towards a structure with a lower refractive index in the case of light coupling from Si down to SiON. However, because the structure of the SOI waveguide tapers to appropriately thin tip width where the effective index is significantly reduced and close to that of the SiO_2 cladding, meaning the effective indices will still match at a point along the tapers and efficient coupling will still occur. In other words, this means that 1) the tip width has been chosen thin enough such that the effective index of the Si taper is able to fall below the refractive index of the SiON structure and 2) the point of coupling has been shifted closer to the tip of the Si taper where the effective index is lower and matches the SiON taper. Of course, as in the case of the fiber-

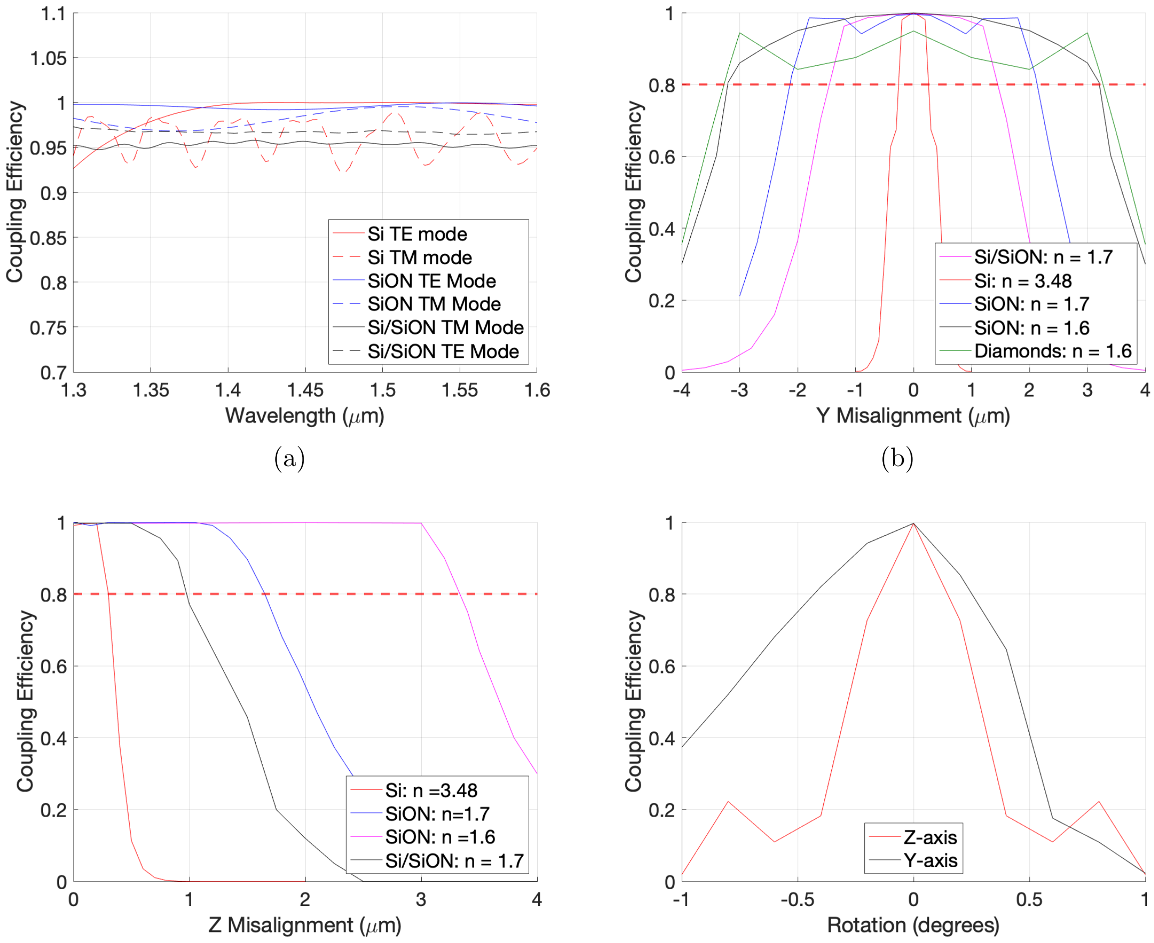


(a)

(b)

Figure 6.7: The diamond structure capable of coupling light between two chips when the input and output waveguides are constrained to be silicon, or some other HIC system. In (a) a side view of the structure while (b) shows a top down view.

to-chip coupler, this is important considering that one device design can be used for vertical escalation of the mode in either direction, significantly simplifying fabrication requirements.



(a) (b) (c) (d)

Figure 6.8: Characterization of misalignment tolerance, wavelength, and polarization dependence for Si and SiON adiabatic cross tapers. In all figures, “Si” indicates the Si/Si system and “SiON” indicates the SiON/SiON system. Where unmarked, “SiON” alone indicates the $n = 1.7$ system. In (a), low wavelength dependance shows that the 1 dB bandwidth is greater than 300 nm. The polarization does not have a large influence on coupling efficiency due to the adiabaticity of the tapers. In (b) the lateral misalignment tolerance is expanded by switching to SiON taper configurations, with a 1 dB tolerance of $\pm$ 2μm for $n = 1.7$ and $\pm$ 3 μm for $n = 1.6$. In (d) the rotational misalignment tolerance is shown only for the SiON system to yield an understanding of the order of magnitude of rotational tolerances, with rotation around the z axis clearly being the limiting factor. Note that rotational around the x axis is not shown because this is negligible.

6.3 Si/Si System Mask

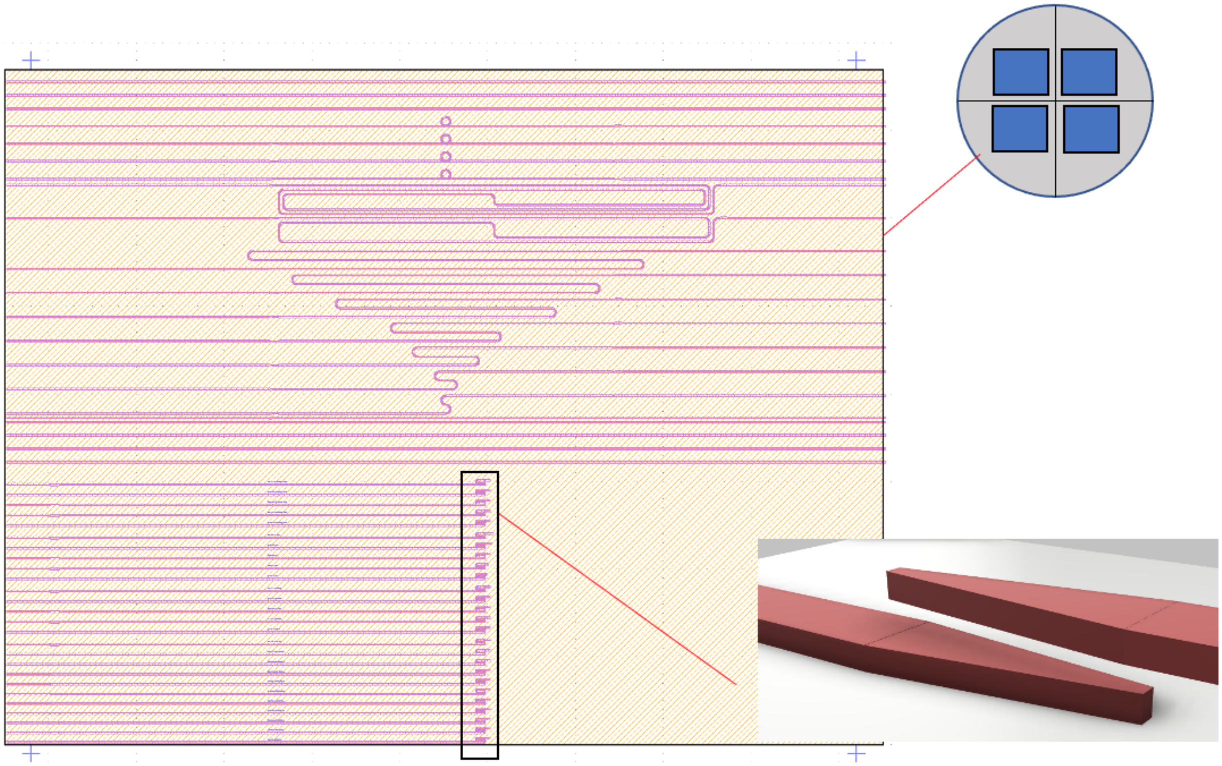
Part of this design development was creation of an appropriate mask for fabrication of the vertical cross tapers and it is presented here for completeness. The mask was made with the intention that it would be utilized to fabricate the Si/Si system. This system was chosen for the initial samples to establish a reference or benchmark for chip-to-chip coupling efficiency and misalignment tolerances before parameters such as material composition or geometry are varied. On this note, lithography of SOI structures has been extensively studied which reduces the process variation and allows a more directed study of the efficiency and tolerances of the cross tapers.

The mask design for the SOI cross tapers is shown in Figure 6.9. The mask is designed to simultaneously provide feedback on $x/y/z$ (translational) misalignment tolerances as well as verify tip width and coupler length dependencies. The circle split into four quadrants represents a wafer cleaved into four chips, with the mask patterning each of the chips identically (note that the diagram is not to scale, and the total footprint of the mask is 1 cm by 0.75 cm). This will be for measuring dependency on the z misalignment tolerance of our design - each of those four (or more) chips will have identical layouts, but each will have an incrementally thicker oxide deposited on top, thus making the spacing incrementally larger for the flip chip system. To measure the x alignment tolerance, tapers on the mask were intentionally offset in the x direction by decreasing the horizontal distance from the tip of the upper taper to the input waveguide width of the lower taper. When the two chips are flip chip bonded together, this will result in perfect overlap in y, but impartial overlap in x. The mask design could not inherently account for rotational misalignment tolerances without incurring undue complexity. Thus, this will need to be determined by active tilting and twisting of the upper chip relative to the lower chip; however, since tilting of the upper chip by the amounts necessary to measure misalignment tolerance (i.e. less than 1 degree), twisting will likely be the only way to fully characterize this type of loss, and even this will be difficult due to the narrow rotational tolerances.

In order to properly measure the y misalignment tolerance, very small offsets done in increments of 100 nm offsets in y are necessary in order to attain enough data points for trustworthy results. Due to the structure's minimum feature size (a tip width of 200 nm) and the need for a high resolution pattern, electron beam lithography (EBL) is proposed for initial fabrication and proof of concept as it can achieve resolutions down to 1 nm. The addition of ring resonators, paperclips, spirals, and varying waveguide widths were added as well to verify waveguide loss characterization. By subtracting out other sources of loss, we can properly determine the intrinsic coupling efficiency of the cross tapers. These loss measurements would need to be done prior to any flip-chip bonding occurring.

It is also important to discuss again the lithography changes that can be made due to the shift in taper material and geometric parameters. For example, for the SiON/SiON system which has geometric parameters above 600 nm, the structures could be patterned using quicker, cheaper direct write methods and tools mentioned in Section 6.1.1 such as 373 or 405 nm optical lithography and the MLA150 tool. This, combined with the option to pattern using industry standards such as 248 or 193 nm DUV, makes this design an ideal candidate for mass production according to 3D FDTD simulation, one of our central design goals. In terms of future research and development, the ability to use the MLA150 tool is important, because not only can be the tapers themselves be defined more easily, but so can the pattern they are arranged in. For example, in the case of the Si/Si system, less than 100 nm resolution is required not only because the tapers have small features, but because the patterning necessary to measure the effect of y misalignment on coupling efficiency requires tapers to be distanced by 100 nm steps or less. This is of course because the Si/Si system has a 1 dB misalignment tolerance of 250 nm, meaning any jumps made higher than 100 nm will result in exponential loss increases and the inability to experimentally characterize the y misalignment tolerance as a reference for future modifications. With the SiON/SiON and Si/SiON system with a 1 dB misalignment tolerance on the order of 1.5-3.5μm and 1.5 μm, respectively, steps can be made in 500 nm or 750 nm jumps - resolutions attainable on more economic machinery (this assumes that the SOI structures were patterned using a technique such as DUV lithography due to the tip width being below 600 nm, but above 100 nm).

Returning to the discussion of the mask layout, it is assumed that active chip-to-chip



(a)

Figure 6.9: Mask layout and .gds file design for a-Si/SOI cross taper system. The upper half of the mask is for loss characterization, while the lower half includes groups of cross tapers for X/Y misalignment, tip width variation, and coupler length variation measurements. The circle split into four quadrants represents cleaved chips which will have different deposited oxide thicknesses, allowing for characterization in all three translational DOF.

alignment would be used in order to accurately characterize the misalignment tolerance of the cross tapers and troubleshoot unforeseen mask design flaws. By not including any type of passive alignment schemes in the first trial, such as an automated pick-and-place tool, we eliminate a potential degree of freedom and streamline the characterization and optimization process for the chip-to-chip coupler. A potential active alignment method which could be used with the designed mask might include using a Maple Leaf photonic measurement system. This system allows for piezoelectric motor integration which can be used to control an arm with an adhesive tip with extreme accuracy and precision. The use of the arm in conjunction with an IR microscope and vacuum holder can be used to actively align two chips to less than 100 nm with a high degree of reliability, thereby accurately test the Si/Si cross taper configuration. This method of active alignment, although not yet completed, is important to mention here because it affects mask design. First, metallic alignment marks must be patterned using the above mask so that they can be seen while looking through an IR microscope, remembering that silicon is transparent in the IR. The mask also needed to contain lithographically defined Vernier calipers for each chip. The calipers are used as alignment position measurement redundancy because they provide a second highly accurate measurement of the alignment of the system in addition to cross and ring alignment markers which the user can visually ensure are aligned. However, in general the idea is that once these narrow misalignment tolerances are increased to beyond 1 μm with the use of SiON, this active alignment scheme would not be utilized as the transition to high speed pick and place tools would be possible.

6.4 Device Comparison and Discussion

Having depicted previous work, design modifications, new device performance, and the status of mask development, we can begin to discuss how this evanescent coupler compares to chip-to-chip coupling devices developed elsewhere. In Table 6.2, relatively recent examples of chip-to-chip optical coupling are listed (note that this table is by no means all encompassing. For example, the chip-to-chip optical coupling package developed in [83, 84, 85] reports coupling light using grating couplers combined with a self-alignment scheme where large photoresist balls sit in pyramidal pits, etched to provide sub-micron alignment. The misalignment tolerance of the coupler is assumed to be sub-micron as well, but no other information about the coupler is explicitly stated. Studies such as these, which have accomplished chip-to-chip coupling, but did not report enough information on the optical components for fair comparison to our design, were left out from the table. The table was simply intended to yield a general idea of other state of the art designs and their performance metrics in order to properly critique our design).

Package	**Coupler**	**Loss (dB)**	**1dB TOL (μm)**	**1dB BW (nm)**	**Year and Ref.**
VCSEL to Si-PIC	Grating Coupler	-11.8	± 1.6	< 50	2016 [86]
III-V die to Si-PIC	Edge Coupler	-1.1	$< \pm 1$	> 100	2016 [69, 79, 80]
Si-PIC to Si-PIC	Grating Coupler	-4.0	± 1.25	< 10	2017 [87]
Si-PIC to Polymer-interposer	Evanescent Coupler	-1.4 (TE) -0.6 (TM)	± 2	≥ 300	2017 [23]
InP die to Si-interposer	Edge Coupler	-1.4	0.45 (vertical) ± 0.6 (lateral)	\	2019 [78]
Si-PIC to Si-PIC	Free Form Coupler	-0.22 (TE) -0.25(TM)	> 35 (vertical) ± 1.3 (lateral)	> 300	2020 [46]

Table 6.2: Summary of a few relatively recent examples of chip-to-chip optical coupling schemes. Note that this list is not all encompassing and is meant to yield a general idea of competing designs and performance.

Through a review of the packages, chip-to-chip couplers, and alignment tolerances in Table 6.2, it can be seen that our simulated chip-to-chip coupler stacks up well according to the performance metrics outlined in Section 3.3. First, in terms of intrinsic coupling efficiency or loss, our vertical cross tapers display a coupling efficiency of greater than 99.5 %, or 0.02 dB of loss, across the board. This is competitive compared to the listed couplers, nearly all of which are north of 1 dB of loss. Second, our widened alignment tolerances from 1.5-3.5 μm (depending on the configurations) are very competitive as well, even compared to grating couplers, all while offering the benefits of low wavelength dependency and polarization sensitivity due to the adiabatic nature of the tapers. In addition, our design is CMOS compatible, meaning the materials can easily be integrated into current microelectronics fabrication facilities, and entirely planar, thus meeting design recommendations set out by standard setting entities such as IPSR-I and removing hurdles to the mass production of the coupler. This is in contrast to other types such as free form couplers, which offer vastly wider alignment tolerances, but come at the cost of non-monolithic fabrication occurring outside of the process flow of typical optoelectronic packages - a detriment to its case for immediate mass manufacturing. An important point to note is that several of the entries in Table 6.2 did not provide simulation results for performance metrics, only experimental results. Experimental measurements are more valuable to determine how the coupler will truly perform, but makes design comparison difficult here since fabrication has not occurred for our coupler yet. For example, the evanescent coupler fabricated in [23] used simulation to aid in design, but did not report the simulated performance metrics themselves. While loss in both their polymer waveguide and Si waveguide systems can be characterized and subtracted out to determine intrinsic efficiency, this procedure has some randomized error. In addition, that specific study optimized design, materials, and processes for 1310 nm light as opposed to 1550 nm light (although the polymer waveguide technology used there has been developed for C-band operation with relatively low absorption) making comparison more challenging. This applies to other couplers as well, such as the OFFCHIP free form coupler in [46], because the wavelength range used was in the visible and very near IR. Therefore, even in cases where the couplers seem extremely similar (both evanescent couplers reliant upon adiabatic tapers), the combination of these two factors makes comparison between the two indirect. Nevertheless, a qualitative metric which is worth noting is that our coupler is silicon based in both the upper and lower taper compared to a polymer based lower waveguide as is the case for [23], meaning there is the potential for enhanced reliability mechanically, thermally, and in terms of humidity control.

6.5 Proposed Future Work

The way forward on this coupler following simulations, development of the mask for the Si/Si system, and finding a suitable active alignment system for the first batch of samples, is to fabricate the initial test units using primarily EBL and reactive ion etching (RIE). The active alignment procedure outlined above has yet to be proven viable as well, which will be key to establishing reference samples. Following the fabrication, alignment, and measurement of the coupling efficiency of the Si/Si system, the clear follow on steps are to then repeat the same steps (mask development, fabrication and loss characterization, alignment) except now using the SiON system whether on a single chips or both chips. Of course, due to the change in material and dimensions, these steps will be entirely different - the dimensions of the mask will be vastly changed, the process flow (including patterning methods and tools) will be altered to accommodate lithographic methods used in mass manufacturing, and the alignment will shift from active alignment to passive alignment using a high speed pick and place tool. Based on the measurements of alignment tolerance and coupling efficiency, our design can then be iteratively improved through further simulations using 3D FDTD, EME, and other means guided by experimental results. In totality, the 3D FDTD and EME simulations of our evanescent coupler show promising results that could prove to be a notable advance in chip-to-chip couplers following optimization of the fabrication and assembly processes.

6.6 Conclusion

As we continue on our current upward trajectory of data center traffic, we are supported by a copper interconnect foundation reaching the limits of its possible performance. Counter to this are the advancements being made in the field of silicon photonics, showing a potential avenue for higher bandwidth and CMOS compatible mass production. As this transition occurs, novel interfaces will be needed to help facilitate the integration of optical components both closer to electronics and, eventually, in replacement of them through all optical signal processing. These interfaces need to be highly efficient, robust, and scalable, and our proposed optical couplers provide insight into how those interfaces might be achieved. It does so through an integrated set of asymmetric, tapered GRIN edge couplers and adiabatic, inverse cross tapers which couple light from fiber-to-chip and chip-to-chip with greater than 95% and 99% predicted efficiency, respectively. This will be accomplished by expanding tolerances of optical components beyond the minimum thresholds for passive high speed pick and place tools and known methods of fiber-to-chip adhesion. The knowledge gained from the design, simulation, and, in the future, fabrication of these optical couplers will hopefully help advance not only the development of optoelectronic packaging for applications in hyperscale datacenters, but also adjacent domains such as biochemical sensing in challenging environments. I hope that through this research and future work in this area, we are able to contribute to the next generation of optical interconnect technology.

Bibliography

[1] G. E. Moore *et al.*, "Cramming more components onto integrated circuits," 1965.

[2] "Microprocessor trend data," 7 2020.

[3] Y.-L. Cheng, C.-Y. Lee, and Y.-L. Huang, "Copper metal for semiconductor interconnects," 2016.

[4] R. H. Havemann and J. A. Hutchby, "High-performance interconnects: an integration overview," *Proceedings of the IEEE*, vol. 89, no. 5, pp. 586–601, 2001.

[5] M. Glick, "The role of integrated photonics in datacenter networks," *Next-Generation of Optical Networks for Data Centers and Short Reach LinksIV*, 2017.

[6] R. E. Brown, R. Brown, E. Masanet, B. Nordman, B. Tschudi, A. Shehabi, J. Stanley, J. Koomey, D. Sartor, P. Chan, J. Loper, S. Capana, B. Hedman, R. Duff, E. Haines, D. Sass, and A. Fanara, "Report to congress on server and data center energy efficiency," *Public Law 109-431: Appendices*, 8 2007.

[7] U. S. Energy Information Administration, "Annual energy outlook 2021 (aeo2021)," 2 2021. https://www.eia.gov/outlooks/aeo/.

[8] C. Kachris and I. Tomkos, "A roadmap on optical interconnects in data centre networks," in *2015 17th International Conference on Transparent Optical Networks (ICTON)*, pp. 1–3, 2015.

[9] Q. Cheng, M. Bahadori, M. Glick, S. Rumley, and K. Bergman, "Recent advances in optical technologies for data centers: a review," *Optica*, vol. 5, pp. 1354–1370, Nov 2018.

[10] X. Zhou, H. Liu, and R. Urata, "Datacenter optics: requirements, technologies, and trends (invited paper)," *Chin. Opt. Lett.*, vol. 15, p. 120008, May 2017.

[11] A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead, R. Bannon, S. Boving, G. Desai, B. Felderman, P. Germano, A. Kanagala, J. Provost, J. Simmons, E. Tanda, J. Wanderer, U. Hölzle, S. Stuart, and A. Vahdat, "Jupiter rising: A decade of clos topologies and centralized control in google's datacenter network," *SIGCOMM Comput. Commun. Rev.*, vol. 45, pp. 183–197, 8 2015.

[12] M. Lebby, R. Pitwon, and M. e. a. Robertson, "Datacenter telecom," *Integrated Photonic Systems Roadmap-International (IPSR-I) 2020*, 2020.

[13] J. C. Maxwell, "A dynamical theory of electromagnetic field," *Royal society transactions*, vol. 155, pp. 526–597, 1864.

[14] J. A. Kong, *Electromagnetic wave theory.* Cambridge, MA: EMW Pub., 2005.

[15] R. P. Feynman, R. B. Leighton, and M. L. Sands, *The Feynman lectures on physics.* Reading, Mass.: Addison-Wesley Pub. Co., 1963.

[16] Y. Shi, L. He, F. Guang, L. Li, Z. Xin, and R. Liu, "A review: Preparation, performance, and applications of silicon oxynitride film," *Micromachines*, vol. 10, no. 8, 2019.

[17] V. Nguyen, *Efficient Power Coupling to Waveguides in High Index Contrast Systems.* PhD thesis, Massachusetts Institute of Technology, 2 2006.

[18] B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics*, pp. 1–40. Hoboken, NJ: John Wiley & Sons Inc., 2nd ed., 2007.

[19] A. Delâge, S. Janz, B. Lamontagne, A. Bogdanov, D. Dalacu, D.-X. Xu, and K. Yap, "Monolithically integrated asymmetric graded and step-index couplers for microphotonic waveguides," *Opt. Express*, vol. 14, pp. 148–161, Jan 2006.

[20] A. Delage, S. Janz, D. X. Xu, D. Dalacu, B. Lamontagne, and A. L. Bogdanov, "Graded-index coupler for microphotonic SOI waveguides," in *Photonics North 2004: Optical Components and Devices* (J. C. Armitage, S. Fafard, R. A. Lessard, and G. A. Lampropoulos, eds.), vol. 5577, pp. 204 – 212, International Society for Optics and Photonics, SPIE, 2004.

[21] T. Tamir, G. Griffel, and H. L. Bertoni, *Guided-Wave Optoelectronics: Device Characterization, Analysis, and Design.* Springer Publishing Company, Incorporated, 1st ed., 2013.

[22] F. Gonthier, "Tapered single-mode fibres and devices. part 1: Adiabaticity criteria," *IEE Proceedings J (Optoelectronics)*, vol. 138, pp. 343–354(11), October 1991.

[23] R. Dangel, A. La Porta, D. Jubin, F. Horst, N. Meier, M. Seifried, and B. J. Offrein, "Polymer waveguides enabling scalable low-loss adiabatic optical coupling for silicon photonics," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 4, pp. 1–11, 2018.

[24] Corning Incorporated, *Corning ® SMF28 ® Ultra Optical Fiber Product Specifications*, 9 2019.

[25] F. P. Payne and J. P. R. Lacey, "A theoretical analysis of scattering loss from planar optical waveguides," *Optical and Quantum Electronics*, vol. 26, no. 10, pp. 977–986, 1994.

[26] S. F. Serna Otálvaro, *Design and characterization of Silicon Photonic structures for third order nonlinear effects.* PhD thesis, Université Paris Saclay, 2016. 2016SACLS409.

[27] W. B. Joyce and B. C. DeLoach, "Alignment of gaussian beams," *Appl. Opt.*, vol. 23, pp. 4187–4196, 12 1984.

[28] N. Lindenmann, G. Balthasar, D. Hillerkuss, R. Schmogrow, M. Jordan, J. Leuthold, W. Freude, and C. Koos, "Photonic wire bonding: a novel concept for chip-scale interconnects," *Opt. Express*, vol. 20, pp. 17667–17677, 7 2012.

[29] N. Lindenmann, S. Dottermusch, M. L. Goedecke, T. Hoose, M. R. Billah, T. P. Onanuga, A. Hofmann, W. Freude, and C. Koos, "Connecting silicon photonic circuits to multicore fibers by photonic wire bonding," *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 755–760, 2015.

[30] J. Wang, Y. Xuan, C. Lee, B. Niu, L. Liu, G. N. Liu, and M. Qi, "Low-loss and misalignment-tolerant fiber-to-chip edge coupler based on double-tip inverse tapers," in *Optical Fiber Communication Conference*, p. M2I.6, Optical Society of America, 2016.

[31] X. Tu, P. Dumais, M. Li, D. Goodwill, H. Fu, D. Geng, and E. Bernier, "Low polarization-dependent-loss silicon photonic trident edge coupler fabricated by 248 nm optical lithography," in *Asia Communications and Photonics Conference 2015*, p. AS4B.3, Optical Society of America, 2015.

[32] M. M. Sisto, B. Fisette, J.-E. Paultre, A. Paquet, and Y. Desroches, "Novel spot size converter for coupling standard single mode fibers to SOI waveguides," in *Silicon Photonics XI* (G. T. Reed and A. P. Knights, eds.), vol. 9752, pp. 199 – 213, International Society for Optics and Photonics, SPIE, 2016.

[33] K. Kruse and C. T. Middlebrook, "Polymer taper bridge for silicon waveguide to single mode waveguide coupling," *Optics Communications*, vol. 362, pp. 87–95, 2016. Polymer Photonics and Its Applications.

[34] R. Takei, M. Suzuki, E. Omoda, S. Manako, T. Kamei, M. Mori, and Y. Sakakibara, "Silicon knife-edge taper waveguide for ultralow-loss spot-size converter fabricated by photolithography," *Applied Physics Letters*, vol. 102, no. 10, p. 101108, 2013.

[35] J. V. Galán, P. Sanchis, G. Sánchez, and J. Martí, "Polarization insensitive low-loss coupling technique between soi waveguides and high mode field diameter single-mode fibers," *Opt. Express*, vol. 15, pp. 7058–7065, 5 2007.

[36] V. Donzella, A. Sherwali, J. Flueckiger, S. T. Fard, S. M. Grist, and L. Chrostowski, "Sub-wavelength grating components for integrated optics applications on soi chips," *Opt. Express*, vol. 22, pp. 21037–21050, 5 2014.

[37] X. Mu, S. Wu, L. Cheng, and H. Fu, "Edge couplers in silicon photonic integrated circuits: A review," *Applied Sciences*, vol. 10, no. 4, 2020.

[38] R. Marchetti, C. Lacava, L. Carroll, K. Gradkowski, and P. Minzioni, "Coupling strategies for silicon photonics integrated chips [invited]," *Photon. Res.*, vol. 7, pp. 201–239, 2 2019.

[39] L. He, Y. Liu, C. Galland, A. E.-J. Lim, G.-Q. Lo, T. Baehr-Jones, and M. Hochberg, "A high-efficiency nonuniform grating coupler realized with 248-nm optical lithography," *IEEE Photonics Technology Letters*, vol. 25, no. 14, pp. 1358–1361, 2013.

[40] R. Halir, P. Cheben, J. H. Schmid, R. Ma, D. Bedard, S. Janz, D.-X. Xu, A. Densmore, J. Lapointe, and I. Molina-Fernández, "Continuously apodized fiber-to-chip surface grating coupler with refractive index engineered subwavelength structure," *Opt. Lett.*, vol. 35, pp. 3243–3245, Oct 2010.

[41] D. Vermeulen, S. Selvaraja, P. Verheyen, G. Lepage, W. Bogaerts, P. Absil, D. V. Thourhout, and G. Roelkens, "High-efficiency fiber-to-chip grating couplers realized using an advanced cmos-compatible silicon-on-insulator platform," *Opt. Express*, vol. 18, pp. 18278–18283, Aug 2010.

[42] Y. Xue, H. Chen, Y. Bao, J. Dong, and X. Zhang, “Two-dimensional silicon photonic grating coupler with low polarization-dependent loss and high tolerance,” *Opt. Express*, vol. 27, pp. 22268–22274, Aug 2019.

[43] M. Streshinsky, R. Shi, A. Novack, R. T. P. Cher, A. E.-J. Lim, P. G.-Q. Lo, T. Baehr-Jones, and M. Hochberg, “A compact bi-wavelength polarization splitting grating coupler fabricated in a 220 nm soi platform,” *Opt. Express*, vol. 21, pp. 31019–31028, Dec 2013.

[44] A. Israel, F. Ulfan, L. Pascar, S. Menezo, Q. Wilmart, S. Malhouitre, Y. Wang, C. Li, P. G. Q. Lo, and H. Taha, “Photonic plug for scalable silicon photonics packaging,” in *Optical Interconnects XX* (H. Schröder and R. T. Chen, eds.), vol. 11286, pp. 52 – 61, International Society for Optics and Photonics, SPIE, 2020.

[45] C. Scarcella, K. Gradkowski, L. Carroll, J.-S. Lee, D. Fowler, and P. O’Brien, “Pluggable single-mode fiber-array-to-pic coupling using micro-lenses,” *IEEE Photonics Technology Letters*, pp. 1–1, 10 2017.

[46] S. Yu, H. Zuo, X. Sun, J. Liu, T. Gu, and J. Hu, “Optical free-form couplers for high-density integrated photonics (offchip): A universal optical interface,” *J. Lightwave Technol.*, vol. 38, pp. 3358–3365, Jul 2020.

[47] H. Haus and K. Šipilov, *Waves and Fields in Optoelectronics*. Prentice-Hall series in solid state physical electronics, Prentice-Hall, 1984.

[48] Z. Lu, D. Celo, P. Dumais, E. Bernier, and L. Chrostowski, “Comparison of photonic 22 3-db couplers for 220 nm silicon-on-insulator platforms,” *2015 IEEE 12th International Conference on Group IV Photonics (GFP)*, pp. 57–58, 2015.

[49] T. Montalbo, “Fiber to waveguide couplers for silicon photonics,” Master’s thesis, Massachusetts Institute of Technology, 9 2004.

[50] J. B. Schneider, “Understanding the finite-difference time-domain method.” `www.eecs.wsu.edu/~schneidj/ufdtd`, 2010. Accessed:2021-05-03.

[51] K. Yee, “Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media,” *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[52] “Finite-difference time-domain method.” `https://en.wikipedia.org/wiki/Finite-difference_time-domain_method`. Accessed: 2021-05-03.

[53] J.-P. Berenger, “A perfectly matched layer for the absorption of electromagnetic waves,” *Journal of Computational Physics*, vol. 114, no. 2, pp. 185–200, 1994.

[54] J. Robinson and Y. Rahmat-Samii, “Particle swarm optimization in electromagnetics,” *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 2, pp. 397–407, 2004.

[55] K. E. Parsopoulos and M. N. Vrahatis, “Particle swarm optimization and intelligence: Advances and applications,” 2010.

[56] J. Pond and M. Kawano, “Virtual prototyping and optimization of novel solar cell designs,” in *Photonics North*, 2010.

[57] J. Pond and M. Kawano, “Virtual prototyping and optimization of novel solar cell designs,” in *Photonics North 2010* (H. P. Schriemer and R. N. Kleiman, eds.), vol. 7750, pp. 626 – 631, International Society for Optics and Photonics, SPIE, 2010.

[58] J. G. Mutitu, S. Shi, C. Chen, T. Creazzo, A. Barnett, C. Honsberg, and D. W. Prather, “Thin film silicon solar cell design based on photonic crystal and diffractive grating structures,” *Opt. Express*, vol. 16, pp. 15238–15248, Sep 2008.

[59] M. Shokooh-Saremi and R. Magnusson, “Leaky-mode resonant reflectors with extreme bandwidths,” *Opt. Lett.*, vol. 35, pp. 1121–1123, Apr 2010.

[60] R. Magnusson, M. Shokooh-Saremi, and E. G. Johnson, “Guided-mode resonant wave plates,” *Opt. Lett.*, vol. 35, pp. 2472–2474, Jul 2010.

[61] C. Manolatou, *Passive Components for Dense Optical Integration Based on High Index-Contrast Systems*. PhD thesis, Massachusetts Institute of Technology, 6 2001.

[62] R. Sun, V. Nguyen, A. Agarwal, C.-y. Hong, J. Yasaitis, L. Kimerling, and J. Michel, “High performance asymmetric graded index coupler with integrated lens for high index waveguides,” *Applied Physics Letters*, vol. 90, no. 20, p. 201116, 2007.

[63] V. Nguyen, T. Montalbo, C. Manolatou, A. Agarwal, C.-y. Hong, J. Yasaitis, L. C. Kimerling, and J. Michel, “Silicon-based highly-efficient fiber-to-waveguide coupler for high index contrast systems,” *Applied Physics Letters*, vol. 88, no. 8, p. 081112, 2006.

[64] G. Ren, S. Chen, Y. Cheng, and Y. Zhai, “Study on inverse taper based mode transformer for low loss coupling between silicon wire waveguide and lensed fiber,” *Optics Communications*, vol. 284, no. 19, pp. 4782–4788, 2011.

[65] M. Papes, P. Cheben, D. Benedikovic, J. H. Schmid, J. Pond, R. Halir, A. O.-M. nux, G. Wangüemert-Pérez, W. N. Ye, D.-X. Xu, S. Janz, M. Dado, and V. Vašinek, “Fiber-chip edge coupler with large mode size for silicon photonic wire waveguides,” *Opt. Express*, vol. 24, pp. 5026–5038, Mar 2016.

[66] Q. Fang, J. Song, X. Luo, X. Tu, L. Jia, M. Yu, and G. Lo, “Low loss fiber-to-waveguide converter with a 3-d functional taper for silicon photonics,” *IEEE Photonics Technology Letters*, vol. 28, no. 22, pp. 2533–2536, 2016.

[67] M. Pu, L. Liu, H. Ou, K. Yvind, and J. M. Hvam, “Ultra-low-loss inverted taper coupler for silicon-on-insulator ridge waveguide,” *Optics Communications*, vol. 283, no. 19, pp. 3678–3682, 2010.

[68] Q. Fang, J. Song, X. Luo, M. Yu, G. Lo, and Y. Liu, “Mode-size converter with high coupling efficiency and broad bandwidth,” *Opt. Express*, vol. 19, pp. 21588–21594, Oct 2011.

[69] T. Barwicz, Y. Taira, T. W. Lichoulas, N. Boyer, Y. Martin, H. Numata, J. Nah, S. Takenobu, A. Janta-Polczynski, E. L. Kimbrell, R. Leidy, M. H. Khater, S. Kamlapurkar, S. Engelmann, Y. A. Vlasov, and P. Fortier, “A novel approach to photonic packaging leveraging existing high-throughput microelectronic facilities,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 6, pp. 455–466, 2016.

[70] “Manual high precision mask and bond aligner.” `https://www.suss.com/en/products-solutions/mask-aligner/ma300-gen2`. Accessed: 2021-05-03.

[71] M.-C. Oh, K.-J. Kim, W.-S. Chu, J.-W. Kim, J.-K. Seo, Y.-O. Noh, and H.-J. Lee, “Integrated photonic devices incorporating low-loss fluorinated polymer materials,” *Polymers*, vol. 3, no. 3, pp. 975–997, 2011.

[72] R. Sun, M. Beals, A. Pomerene, J. Cheng, C. yin Hong, L. Kimerling, and J. Michel, “Impedance matching vertical optical waveguide couplers for dense high index contrast circuits,” *Opt. Express*, vol. 16, pp. 11682–11690, 8 2008.

[73] K. Itoh, Y. Hayashi, J. Suzuki, T. Amemiya, N. Nishiyama, and S. Arai, “Double taper interlayer transition coupler for 3d optical interconnection with heterogeneous material stacking,” in *2016 International Conference on Solid State Devices and Materials*, 9 2016.

[74] K. Itoh, N. Nishiyama, Y. Hayashi, J. Suzuki, T. Amemiya, and S. Arai, “Design of curved taper interlayer coupler for multilayered photonic integrated circuits,” in *2017 JSAP Autumn Meeting*, 9 2017. (in Japanese).

[75] N. MacFarlane, M. R. Kossey, J. R. Stroud, M. A. Foster, and A. C. Foster, “A multi-layer platform for low-loss nonlinear silicon photonics,” *APL Photonics*, vol. 4, no. 11, p. 110809, 2019.

[76] S. Gao, Y. Wang, K. Wang, and E. Skafidas, “Polarization insensitive vertical coupler for multi-layer silicon photonic integrated circuits,” in *2015 IEEE Optical Interconnects Conference (OI)*, pp. 70–71, 2015.

[77] N. Takato, K. Jinguji, M. Yasu, H. Toba, and M. Kawachi, “Silica-based single-mode waveguides on silicon and their application to guided-wave optical interferometers,” *Journal of Lightwave Technology*, vol. 6, no. 6, pp. 1003–1010, 1988.

[78] Y. Martin, J. S. Orcutt, C. Xiong, L. Schares, T. Barwicz, M. Glodde, S. Kamlapurkar, E. J. Zhang, W. M. J. Green, V. Dolores-Calzadilla, A. Sigmund, and M. Moehrle, “Flip-chip iii-v-to-silicon photonics interfaces for optical sensor,” in *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*, pp. 1060–1066, 2019.

[79] J. Nah, Y. Martin, S. Kamlapurkar, S. Engelmann, R. L. Bruce, and T. Barwicz, “Flip chip assembly with sub-micron 3d re-alignment via solder surface tension,” in *2015 IEEE 65th Electronic Components and Technology Conference (ECTC)*, pp. 35–40, 2015.

[80] Y. Martin, J. Nah, S. Kamlapurkar, S. Engelmann, and T. Barwicz, “Toward high-yield 3d self-alignment of flip-chip assemblies via solder surface tension,” in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, pp. 588–594, 2016.

[81] “Mrsi-m3: 3- micron die bonder.” `https://mrsisystems.com/mrsi-m3/`. Accessed: 2021-03-22.

[82] “Palomar 6532hp die bonder.” `https://www.palomartechnologies.com/6532hp-die-bonder-data-sheet-lfw`. Accessed: 2021-03-22.

[83] M. Zia, C. Zhang, H. Yang, L. Zheng, and M. Bakir, “Chip-to-chip interconnect integration technologies,” *IEICE Electron. Express*, vol. 13, p. 20162001, 2016.

[84] H. S. Yang, C. Zhang, and M. S. Bakir, “Self-alignment structures for heterogeneous 3d integration,” in *2013 IEEE 63rd Electronic Components and Technology Conference*, pp. 232–239, 2013.

[85] H. S. Yang and M. S. Bakir, “Design, fabrication, and characterization of freestanding mechanically flexible interconnects using curved sacrificial layer,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 4, pp. 561–568, 2012.

[86] H. Lu, J. S. Lee, Y. Zhao, C. Scarcella, P. Cardile, A. Daly, M. Ortsiefer, L. Carroll, and P. O’Brien, “Flip-chip integration of tilted vcsels onto a silicon photonic integrated circuit,” *Opt. Express*, vol. 24, pp. 16258–16266, Jul 2016.

[87] Y. D. Zonou, S. Bernabe, D. Fowler, M. Francou, O. Castany, P. Arguel, and O. Castany, “Self-alignment with copper pillars micro-bumps for positioning optical devices at submicronic accuracy,” in *2017 IEEE 67th Electronic Components and Technology Conference (ECTC)*, pp. 557–562, 2017.