OPTICAL COUPLER DESIGN AND EXPERIMENTAL DEMONSTRATION FOR 2.5D/3D HETEROGENEOUS INTEGRATED ELECTRONICS

A Thesis Dissertation
Presented to
The Academic Faculty

By

Congshan Wan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

May 2019

OPTICAL COUPLER DESIGN AND EXPERIMENTAL DEMONSTRATION FOR 2.5D/3D HETEROGENEOUS INTEGRATED ELECTRONICS

Approved by:

Dr. Muhannad S. Bakir, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Thomas K. Gaylord, Co-advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Benjamin D. B. Klein
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Stephen E. Ralph
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Oliver Brand
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Suresh K. Sitaraman
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: March 29, 2019

Nothing in life is to be feared. It is only to be understood.

Marie Curie

To my parents,

Renjun Wan and Huaying Zhang,

and my husband,

Li Wang

ACKNOWLEDGEMENTS

"The way of progress is neither swift nor easy." Yet every step forward, no matter big or small, brings hope, confidence and fulfillment. In my journey through the Ph.D. study, I have experienced thousands of such moments, confronting the setbacks and tackling the problems. I don't give up because of the encouragement and consolation bestowed on me from my advisors, family, and friends. Every failure makes me stronger, and I feel more grateful to those around me.

I would like to express my sincere gratitude to my research advisors, Dr. Bakir and Dr. Gaylord. They gave me an opportunity when I was misoriented and guided me to the field of optics and photonics. They helped me identify useful research topics and develop good research habits. I appreciate them for giving me the freedom to explore the field on my own, acknowledging my works and contributions, and considering my personal needs.

I also wish to thank previous group members, Yue Zhang, Chaoqi Zhang, Xuchen Zhang, Li Zheng and Yang Zhang, for welcoming me into the group and helping me get familiar with the lab and the group culture. Even though I work on a completely different field, I have learnt a lot from the discussions with my current labmates, Muneeb Zia, Hanju Oh, Thomas Sarvey, William Wahby, Reza Abbaspour, Joe Gonzalez, Paul Jo, and Md Obaidul (Opu) Hossen, on heterogeneous integrations, microfabrications, flexible interconnects, through-silicon vias, fluidic coolings and power managements. These discussions broaden my horizon and enrich my experience, and they allow me to see a broad picture in which optics and photonics can be integrated. It is also rewarding to work with the smart and energetic new members, Sreejith Rajan, Ting Zheng, Ankit Kaul and Ming Jui (Carl) Li. Talking with them brings back my memory when I first joined the group, full of vigor. I wish them all success in their future career.

Many of the fabrication and testing results presented in this dissertation were not possible without the collaboration with Dr. Adibi's group. I wish to thank Dr. Adibi for offering

me full access to his labs and Tianren Fan for helping me on the experiments.

Ph.Ds' lives are not only about research, and thankfully I have a few best friends to enjoy the days. I'd like to thank my friends Ruxiu Liu and Sicheng Wang for all the fun we shared together. Those exotic restaurants, board game nights, and strenuous hiking are the best treatment for fighting fatigue and depression.

Lastly, I would like to thank my parents and my husband. My parents shaped who I am when I was young, and I carry those identities with me until today. They make me persistent in front of tasks, curious about the unknown world, and resistant to negative influences. My parents are strong believers in education and they invest all their time and effort to give me a better chance. Studying in the United States separates us physically, but I never feel alone because I know my parents are always there thinking of me. Whenever I need them, they are there. Their support, encouragement, and tiny little jokes are always there. I am fortunate to be born in this caring and supportive family. I am also lucky to meet my husband, Li Wang, at Georgia Tech. He is my best friend and companion. He is so smart that I could discuss my research, which is not his field, with him and get some inspirations. In front of him, I don't have to be strong or reasonable all the time. I'd like to thank him for sharing every moments with me and brightening up my days.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Digital computing systems, which consist of devices for logical functions and storage as well as interconnects for information transmission, have benefited from the device scaling. The exponential reduction in feature sizes of CMOS microprocessors, known as Moore's Law [1], leads to gigascale integration (GSI) which places immense bandwidth demands on chip-to-chip and chip-to-package interconnects. This evolution is breaking the balance between devices and interconnects, since present-day electrical interconnect performance does not scale as the devices. As such, optical interconnects have been proposed as a potential solution which can overcome design challenges of electrical interconnects due to their completely different physics. Optical fibers are one successful example of leveraging optical signals for long-distance communication; furthermore, chip-level or package-level optical signalings have demonstrated their advantages in heterogeneous integration platforms consisting of both electronics and photonics. A critical factor for enabling chip-to-chip and chip-to-package optical interconnection is efficient optical coupling, which involves the design of high-efficiency and large-bandwidth couplers as well as the development of optical alignment techniques. This research addresses challenges in optical coupling through the design, modeling, analysis, fabrication and testing of optical couplers and the demonstration of a passive self-aligning integration platform.

CHAPTER 1
INTRODUCTION

The rapid growth in data centers, 5G communication, augmented reality, autonomous vehicles, and artificial intelligence, etc., creates increasing demands for higher data rate and larger processing throughput. Even though electronics have pushed their limits to satisfy these needs, heterogeneous integration with photonics is inevitable, as predicted by Integrated Photonic Systems Roadmap International (IPSR-I) [2]. This is because optical signals have many benefits such as large bandwidth, low latency, reduced energy dissipation, dense interconnect design, and negligible dispersion, etc. Optical fibers have successfully demonstrated their superiority in long-haul signal traffic over electrical cables because the loss at high frequencies in electrical wires is significant; the energy efficiency of electrical links drops notably when transmitting large throughput over a large distance ($>$ 1 mm). For shorter-distance signal communication inside computing systems, the buses that carry information on the chip-level or the package-level run at rates much slower than the clock rate on the chips due to various limitations with electrical interconnects, such as wave reflection, crosstalk and inductance, etc. In addition, the scaling of electrical interconnects becomes a critical limiting factor to the performance of the overall computing system. On the contrary, an optical system can address most of the problems encountered in electrical interconnects. As technology advances, the emergence of 2.5D/3D integrated photonics [2] offers opportunities for denser and more complex network designs without the challenges of waveguide crossings and crosstalk. Optical couplers are critical components to achieve efficient integrated photonics, which enable chip-to-chip, chip-to-package, and package-to-package optical connections. Therefore, designing high-performance optical couplers, e.g. fiber couplers and interlayer couplers, and developing accurate and robust integration approaches are important to the field of packaging and integration.

Here in the Introduction, we review recent developments in high performance computing (HPC). Next we discuss the increasing difficulties in continuing HPC performance enhancement due to the problems of electrical interconnects. Subsequently, the potentials of optical interconnects are evaluated and some commercial optical products are introduced. Further, we discuss challenges in current optical interconnect technologies and the necessity of heterogeneous integration. Finally, the research presented in this thesis will be briefly summarized.

1.1 Commercial Processors

Individual processors, such as the central processing units (CPUs) and specialized graphic processing units (GPUs), comprise of electronic circuits that implement the basic arithmetic, logic, control, input/output (I/O), and memory operations specified by the instructions. Most modern CPUs are contained on a single integrated circuit (IC), on which other components, such as memory, and peripheral interfaces, etc., are also integrated. Such integrated "microprocessor" chip may also be called microcontroller or System-on-Chip (SoC). An IC with multiple CPUs, or "cores", is refered to as a "socket" [3]. Some specialized integrated circuits can be configured by a user after manufacturing, such as the Field-Programmable Gate Arrays (FPGAs). FPGAs contain an array of programmable logic blocks, memory elements, and a hierarchy of reconfigurable interconnects. Among all the components on an IC, memory elements limit the overall bandwidth of the microprocessor. The critical requirement for any system's performance is the ability to read/write large amounts of data from/to memory efficiently, and the memory bandwidth creates a bottleneck for next-generation platforms, including data centers, high-performance computing systems, 8K broadcasting, wireline networking, and Internet of Things (IoT). The memory bandwidth limitation makes it harder to achieve the growing data traffic rate, which is projected to be 4.8 ZB per year (approximately 150 TB/s) globally by 2022 [4]. Actually,

the memory bandwidth is largely limited by the I/O bandwidth, since it is not physically possible to integrate enough I/O pins to support a wide enough memory bus delivering the required bandwidth. Adding more components cannot solve the problem due to the increased power and form factor impact. To mitigate this issue, Intel launched the Stratix 10 MX FPGA (Fig. 1.1(a)) in 2017, whose components are schematically shown in Fig. 1.2. This FPGA consists of two HBM2 (second-generation High-Bandwidth Memory) memory tiles (up to four tiles per package) and four transceivers which are interconnected to the central FPGA core fabric via the Embedded Multi-die Interconnect Bridge (EMIB) technology [5]. Each HBM2 memory tile contains either 4 or 8 memory layers and supports up to 16 independent channels (64 bits each). Each channel can run at data rates of up to 2 GB/s (16 Gb/s) and provide up to 16 GB/s of aggregate bandwidth per channel [6]. Xilinx's Virtex UltraScale+ FPGA (Fig. 1.1(b)) demonstrates 8 GB memory and 8384 Gb/s bandwidth [7]. Intel also launched the Xeon Scalable CPUs to achieve a high aggregate bandwidth of 1023.7 Gb/s. To compete with the Intel Xeon Scalable CPUs, AMD launched the EPYC SoCs, each of which features up to 32 cores, up to 8 channels of DDR-4 memory, large amount of I/O (128 PCIe lanes per CPU), and 2 TB of high-speed memory per socket [8]. Both Intel processor and AMD processor are shown in Fig. 1.3. NVIDIA launched the Tesla V100 GPUs (Fig. 1.4) for high performance computing in 2017. Tesla V100 offers the performance of up to 100 CPUs in a single GPU with a memory bandwidth of 900 GB/s (7200 Gb/s). Nevertheless, its interconnect bandwidth is relatively low (300 GB/s) [9]. Recently, AMD unveiled Radeon VII (Fig. 1.5), the worlds first 7 nm gaming graphics card that breaks the terabyte memory barrier. It is designed to deliver exceptional performance for the latest gaming, e-sports, Virtual Reality (VR) and 3D rendering applications. It features 60 computing units/3840 stream processors running at up to 1.8 GHz, 16 GB of ultra-fast HBM2 memory, and 1 TB/s (8000 Gb/s) memory bandwidth [10]. Some technical specifications of these products are summarized in Table 1.1.

(a) Intel Stratix 10MX FPGA

(b) Xilinx Virtex UltraScale+ FPGA

Figure 1.1: Intel Stratix 10 MX FPGA [11] and Xilinx's Virtex UltraScale+ FPGA [12].

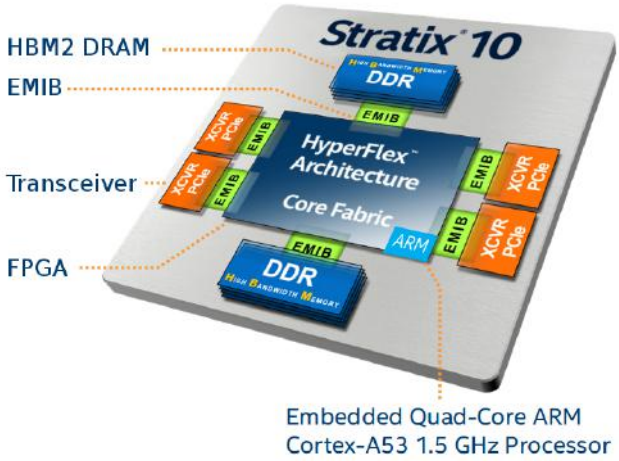
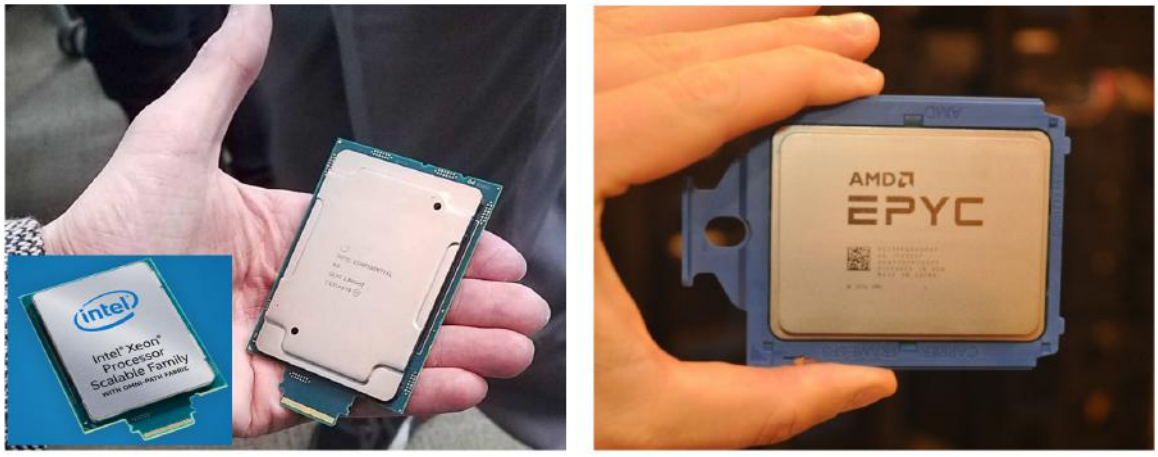
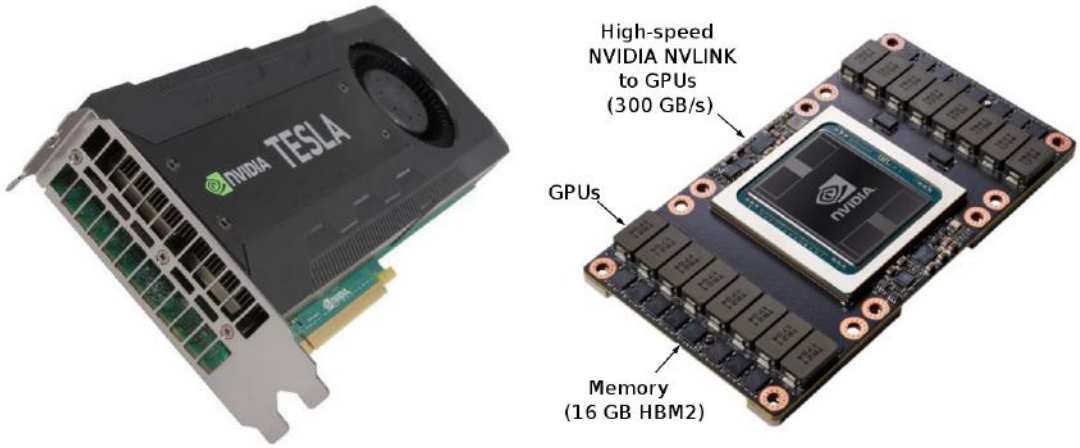


Figure 1.2: Schematic showing the components on the Intel Stratix 10 MX FPGA. Figure is modified based on [13].



(a) Intel Xeon processor (b) AMD EPYC processor

Figure 1.3: Intel Xeon Scalable processor [14] and AMD EPYC processor [15].



(a) NVIDIA TESLA V100 GPU (b) Inside of NVIDIA TESLA V100

Figure 1.4: NVIDIA Tesla V100 GPU and its components. Figure is modified based on [16, 17].

Figure 1.5: Recently launched AMD Radeon VII GPU [18].

Table 1.1: Some commercial processors for high performance computing.

Product Type	Product Name	Application	Aggregate Bandwidth [Gb/s]	Memory [GB]	Power Consumption [W]	Max I/O	Launch Year	Reference
Intel CPU	Xeon Scalable Platinum 8176	High Performance Computing	1023.7	768	165 [a]	–	2017	[19]
Intel CPU	Xeon E5 2699A v4	High Performance Computing	614.4	1540	145 [a]	–	2016	[20, 21]
Intel FPGA	Stratix 10 MX	High Performance Computing	4096	–	125 [a], 182 [b]	–	2017	[6, 22, 23]
AMD GPU	Radeon VII	Gaming and Virtual Reality	8000	16	295	–	2019	[24, 25]
AMD CPU	EPYC 7601	Server and Data Center	1365.6	2048	180 [a]	4096	2017	[20]
NVIDIA GPU	Tesla V100	High Performance Computing	7200	32	300 [c]	–	2017	[9]
Xilinx FPGA	Virtex UltraScale+ XCVU37P	Cloud Computing	8384	8	–	624	2016	[26]

[a] Thermal design power: average power the processor dissipates when operating at base frequency with all cores active
[b] Power dissipation of the whole package measured in the lab [c] Maximum power consumption

1.2 Problems of Electrical Interconnects

As we can see from the previous section, the most advanced processor, AMD Radeon, already uses the 7 nm technology. Even though the technology may keep advancing to even higher resolution, e.g. 5 nm node, it poses increasing challenge on the lithography and fabrication process. In addition, the 1 TB/s bandwidth achieved by the latest processor is far less than the projected hundreds of terabyte per second bandwidth demands [4]. This may largely be due to the limited interconnect bandwidth, as shown in the case of NVIDIA Tesla V100.

Undoubtedly, electrical wiring has been successful in interconnecting semiconductor chips at low cost for decades. Nevertheless, as the semiconductor industry keeps delivering chips with greater transistor densities, larger bandwidth and higher clock speeds due to the scaling-down of feature sizes, electrical interconnects meet a bottleneck to keep up with the scaling. It is increasingly difficult to keep the balance between on-chip logic operations and the off-chip read/write operations.

1.2.1 Interconnect Density

One of the most obvious problems of electrical wires is resistance. Resistance, especially at high frequencies, gives rise to signal attenuation, delay, and distortion. In order to reduce resistance, cross-sectional areas of the electric wires need to be relatively large, resulting in low interconnect density and the so-called scaling problem in a limited space. Such thick wires also increase cost for long lines. We can refer to the underlying physics to better understand the scaling limitation. There are two types of electrical wires, namely resistive-capacitive (RC) lines and inductive-capacitive (LC) lines. For modern digital systems, information is processed at the gigahertz frequency level, and on-chip interconnects are primarily bulk-resistance-limited RC lines, while the off-chip links are primarily skin-effect-limited LC transmission lines. Now take a RC line, in which the capacitance of the

line is charged through the bulk resistance, as a simple example of the on-chip electrical wire. The wire has an effective cross-sectional area A, a capacitance per unit length C_l, and a resistance per unit length R_l. The total RC time constant of the wire is $R_l C_l$. Suppose the line is shrunk in all three dimensions by a factor of s (s <1). The cross-sectional area would shrink by s^2, increasing the resistance per unit length to R_l/s^2. The length of the line would be shrunk to sl. However, the capacitance per unit length remains unchanged because it only depends on the geometry of the line. As a result, the total RC time constant would be $(R_l/s^2)C_l(sl)^2 = R_l C_l l^2$, so the shrinkage of the wire doesn't contribute to any change in RC time constant in this simplified analysis [27, 28]. This means that the time delay caused by the interconnect will not be reduced by scaling-down the wires. As transistors get faster due to their dimension shrinkage, the interconnects will be even harder to keep up with the fast switching of the devices. Using LC lines is less likely to solve this dilemma, because they follow a similar scaling law. In fact, the bit-rate capacity of both RC and LC lines can be modeled as

$$B \leq B_0 \frac{A}{l^2}, \tag{1.1}$$

where A is the cross-sectional area, l is the total length, B is the total number of bits per second, and the constant B_0 is approximately 10^{15} bits/s for high-performance strip lines and cables, 10^{16} bits/s for small on-chip interconnects (RC lines), and 10^{17}-10^{18} bits/s for off-chip equalized lines (LC lines with resistive loss, or RLC lines) [29]. From Eq. (1.1), the dimensionless ratio A/l^2 is what makes the bit-rate capacity of the electrical wires independent of the size.

1.2.2 Interconnect Bandwidth

The continuous scaling of semiconductor devices, which allows more processing power and integrated functionalities, poses increasing challenges to inter-chip and intra-chip commu-

nications. The challenges are reflected by two important aspects, namely bandwidth density and energy efficiency. Bandwidth density, defined as gigabits per second per square millimeter (Gb/s/mm^2), determines the aggregate throughput; and energy efficiency, defined as picojoules per bit (pJ/bit), indicates the overall power consumption. Bandwidth demands in servers and other high performance computing environments have been increasing at least as fast as Moore's law with aggregate chip bandwidth exceeding terabits per second, which significantly exceed the growth rate of the number of I/O pins on the package. The major limitation of electrical interconnects is the low bandwidth-distance product (Bd in the unit of GHz$\times$m, where B and d represent bandwidth and distance, respectively) of the metallic medium. This metric, Bd, indicates that the distance over which an electrical link can reliably transmit a signal is limited by the frequency-dependent attenuation of the transmission medium at a given signaling rate [30]. Therefore, the energy efficiency of electrical links drops significantly when transmitting large data rates over a large distance ($>$ 1 mm), resulting in narrow frequency bandwidth.

1.2.3 Interconnect Energy

From the discussion of the previous section, it is obvious that the bandwidth density is closely related to the energy efficiency. Energy is dissipated via many sources in information processing, e.g. logic operation, read/write operation, thermal dissipation, information transmission through interconnects, and leakage and sub-threshold current, etc., but surprisingly most of the energy loss is resulted from the interconnects [31]. For example, a simple crosspoint switch consumes about 20$\sim$40 pJ/bit. However, the energy consumed by the switching circuits is less than 1 pJ/bit, while the dominant energy consumers are the electrical interconnects of the chip [31]. This phenomenon is aggravated in a switching system consisting of multiple concatenated switching chips. Commercial internet routers consume several nanojoules per bit [32], because their routing functions are far more complex than a simple crosspoint switch and involve much more extensive processing and data

movement. Chip communication in complicated system relies heavily on data transmission through the interconnects on chip, package, board, connector, and backplane, and thus much energy is lost due to the signal propagation on the interconnects associated with charging and discharging the capacitance of signal lines. In addition, electrical signals may be amplified by repeaters to improve the bandwidth of longer interconnects, however, at the expense of significant power consumption.

1.3 Potentials of Optical Interconnects

It is not difficult to see that optics could solve many physical problems of electrical interconnects, including precise clock distribution, system synchronization, signal integrity and timing, power dissipation, and bandwidth of long interconnections. Optics may also relieve stringent design problems, such as crosstalk, voltage isolation, wave reflection, impedance matching, and pin inductance. Optical interconnects can deliver precise timing because optical waveguides/fibers have low dispersion and negligible temperature influences. As carrier frequency of the signal (petahertz, 10^{15} Hz) is much higher than the modulation frequency (bit rate), light propagation is essentially not affected by the frequency modulation. Even though optics also exhibits crosstalk between channels, such crosstalk doesn't depend on bit rate. The on-chip interconnect density can be potentially increased since optical interconnects do not suffer from the aspect ratio limit. The physics of loss and signal distortion in optical interconnects is completely different, and the major loss in optical systems is usually not associated with propagation distance but rather with components and connections. This is why optical fibers have been widely used for long-distance communication, such as wide area networks (WANs) and metropolitan area networks (MANs). Each individual section of fibers can be 100 km in length between repeater stations due to their very low loss and dispersion. Bandwidth requirement can be achieved by the use of wavelength-division multiplexing, and there is negligible power loss along the optical channels (waveguides and fibers).

1.4 Commercial Optical Transceivers

In order to eliminate the networking bottlenecks that result in stranded computing capacity and gain much faster data transfer over longer distances, many traditional electronics manufacturers and emerging technology startups have begun to focus on integrated photonics which enables high-bandwidth and low-consumption networking. Optical transceivers are the most common type of devices that utilize photonics technology. An optical transceiver is a device that uses optical fibers to send and receive data. The transceiver has electronic components, electrical-to-optical (E-O/O-E) converters such as lasers and photodetectors, to encode/decode data into light pulses and then send them to the other end as electrical signals. In the second quarter of 2019, Intel announced the silicon photonics optical transceiver 100G CWDM4 QSFP28 compatible with single-mode fiber connectors and cable infrastructures for uncooled operations under CWDM (Coarse Wavelength Division Multiplexing) wavelengths (1271 nm, 1291 nm, 1311 nm, and 1331 nm). The transceiver is expected to support 100 GbE (transmitting Ethernet frames at 1 Gb/s) over 10 km, and it is targeted at large scale cloud and enterprise data centers as well as Ethernet switch, router, and client-side telecom interfaces [33]. A privately-held startup, Luxtera, has successfully demonstrated LUX42604 QSFP28 optical modules, which are pluggable optical transceivers containing four parallel fiber optic transceivers with full duplex operation, each operating at data rates from 1 Gb/s up to 25.78 Gb/s and supporting a reach up to 2 km over standard single-mode fiber (SMF). Luxtera's technology motivated Cisco to acquire Luxtera in the third quarter of 2019. The acquisition will help Cisco in achieving 100GbE/400GbE optics, silicon, and process technology [34]. NeoPhotonics' 100G CFP2 LR4 transceiver supports 103.1 Gb/s and 111.8 Gb/s aggregate bit rates up to 10 km single-mode fiber transmission for data center applications. NeoPhotonics also provides coherent modules, optical switches, drivers and amplifiers, network monitors and passive components. Inphi's COLORZ Silicon photonics technology is a low power, cost effective 100

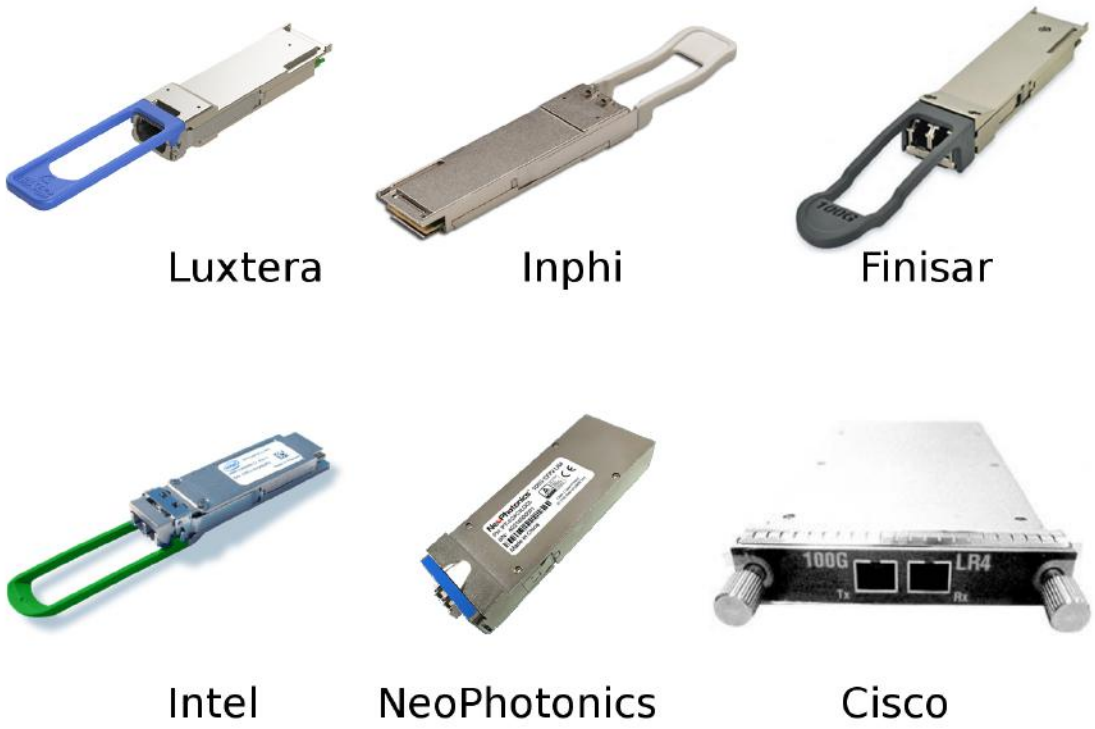


Figure 1.6: Commercial optical transceivers by Luxtera, Inphi, Finisar, Intel, NeoPhotonics, and Cisco.

Gb Ethernet DWDM (Dense Wavelength Division Multiplexing) platform in QSFP28 form factor for data center interconnect (DCI) within an 80 km distance. It can deliver 4 Tb/s capacity over a single fiber. Finisar produced the 100G 100m QSFP28 SWDM4 Optical Transceiver designed for use in 100G Ethernet links over duplex multimode fiber (MMF) with a maximum link length of 150 m. Figure 1.6 shows the discussed commercial optical transceivers, and some technical specifications of these products are summarized in Table 1.2. Ayar Labs, a silicon photonics startup, is developing its electro-optical I/O chips, TeraPHY silicon chip, which will become the basis of its first commercial product. The initial chip is envisioned to deliver 3.2 Tb/s achieved by 8 single-mode fibers transmitting data at 400 Gb/s. These optical transceivers are now widely used in high-speed datacom and telecom, Local Area Networks (LANs), high performance computing (HPC), mobile infrastructure, and Storage Area Networks (SANs).

Table 1.2: Some commercial optical transceivers.

Company Name	Bandwidth [GbE]	Power [W]	Form Factor	Fiber Type	Link Length [km]	Operation Wavelength [nm]	Electrical Interface
Intel	100	< 3.5	QSFP28	SMF	< 10	CWDM (1271, 1291, 1311, and 1331)	IEEE 802.3bm CAUI-4
Luxtera	100	< 3.5	QSFP28	SMF	< 2	1310	IEEE 802.3bm CAUI-4
Inphi	100	–	QSFP28	–	< 80	DWDM	IEEE 802.3bm CAUI-4
Finisar	100	< 3.5	QSFP28	MMF	< 0.15	–	IEEE 802.3bm CAUI-4
NeoPhotonics	100	–	CFP2	SMF	< 10	–	IEEE 802.3ba 100GBASE-LR4
Cisco	100	< 24	CFP	SMF	< 25	DWDM (1295.6, 1300.1, 1304.6, and 1309.1)	IEEE 802.3ba

1.5 Challenges of Optical Interconnects

Nevertheless, on-chip optical wires must be dense enough, at least on the order of hundreds or more likely thousands per chip, to show energy and cost advantages [27, 28]. Without such number, the on-chip interconnects would have to remain electrical. This is because the optical transmitters and receivers (the E-O/O-E converters) typically consume significant amount of power. Even though the waveguiding channels of optical links dissipate minimal energy, it is the total energy per bit, including the power of both the transmitter and receiver, that becomes a challenge for reducing interconnect energy. Optical interconnects also face system integration complexity, immature technology, environment sensitivity, and high cost. Therefore, heterogeneous integration, which complements electrical interconnects and optical interconnects by leveraging the advantages of both electronics and optics, is inevitable to satisfy present-day digital processing requirements. As such, a promising solution at present is to bring off-chip signals from optical fibers and convert them to on-chip electrical signals using photodetectors, which is proved by the popularity of optical transceivers. In parallel with the technology evolution for optical interconnects, there is a growing trend for 2.5D/3D integrated photonics, which emphasizes the need for high-efficiency interlayer optical routing. As most of the power loss comes from the optical connectors, high-efficiency optical couplers play an important role in the overall heterogeneous integration.

1.6 Heterogeneous Integration Platforms

Many heterogeneous integration platforms have been proposed to interconnect multiple dice of various functionalities, including ASICs, CPUs, GPUs, FPGAs, microsensors, MEMS, RF components, and photonics, into a single package [38, 39]. Silicon interposer, as shown in Fig. 1.7, is demonstrated to provide high-density (more than 10,000 connections between dice) and low-latency ($\sim$ 1 ns) connections [35, 40]. However, the size of

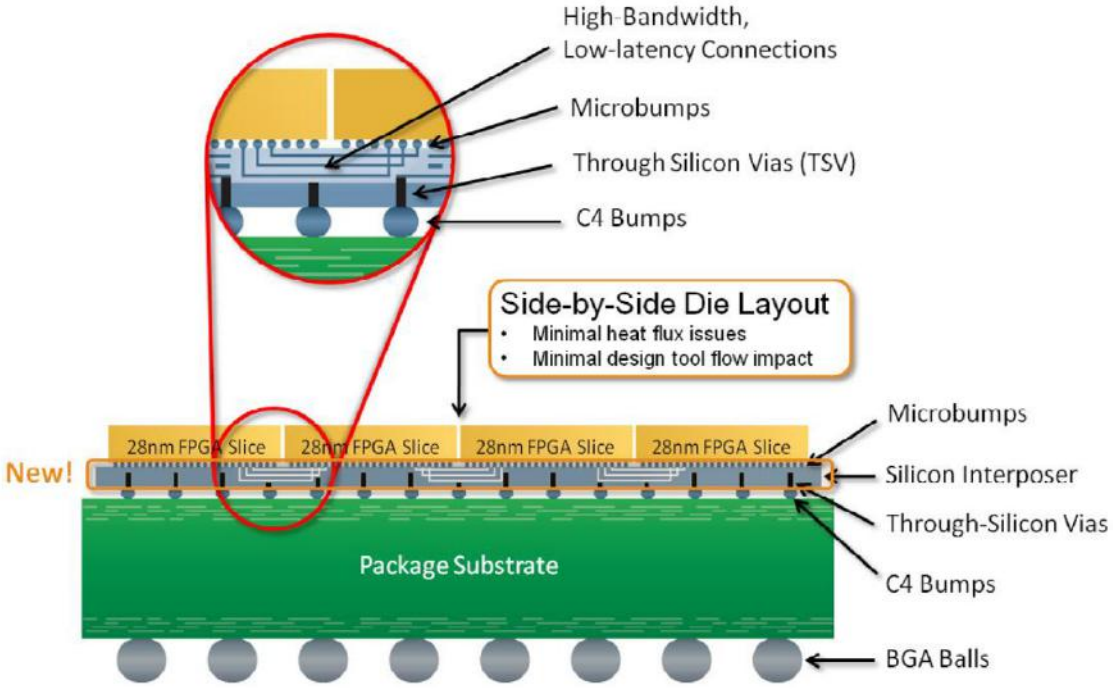


Figure 1.7: FPGA dice (or other dice) are bonded to a silicon interposer that provides relatively high-bandwidth and low-latency electrical interconnections; through-silicon vias and C4 bumps create connections to system I/O, power, clocks and other signals through the package substrate. Figure is obtained from [35].

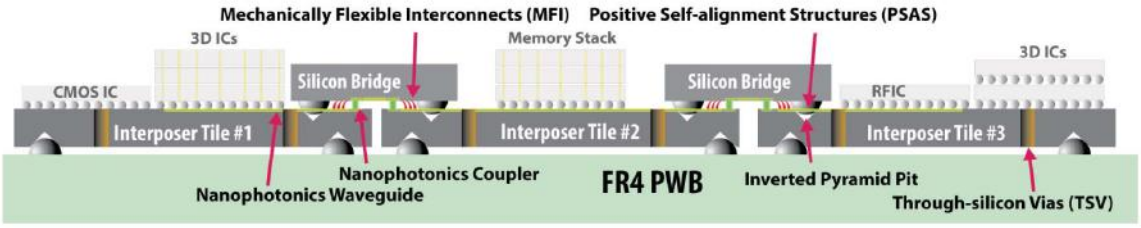
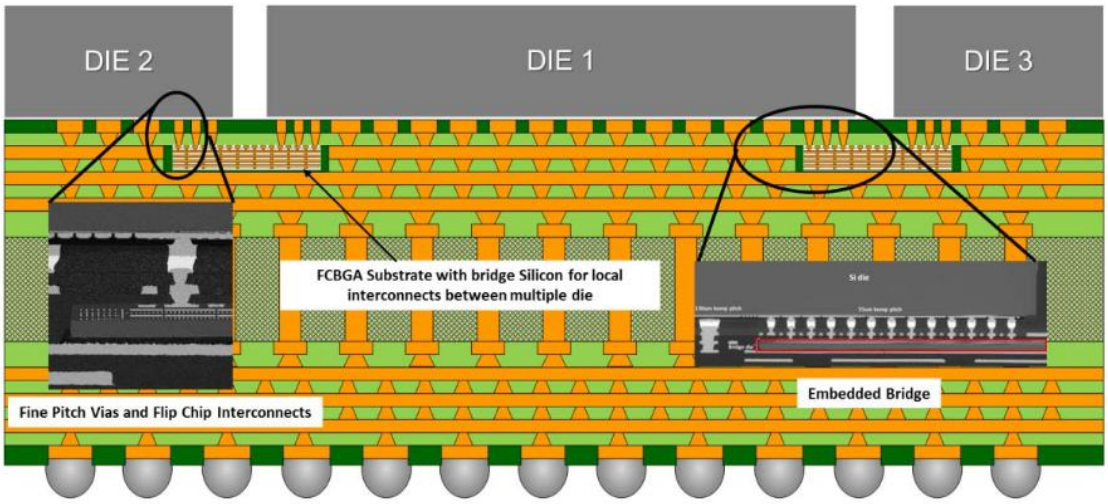
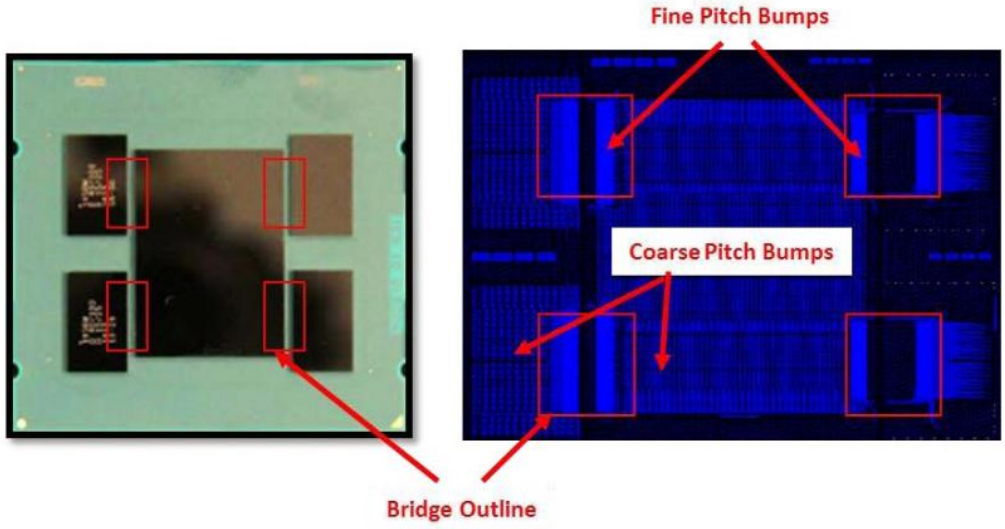


Figure 1.8: Silicon interposers are directed mounted on the FR4 printed circuit board and the interposers are connected by Silicon bridges. Figure is obtained from [36].



(a)



(b)

Figure 1.9: Schematic and sample images showing the EMIB architecture: (a) cross-section view; (b) top view of a test vehicle and its design layout to highlight localized high density interconnects and pitches. Figure is obtained from [5].

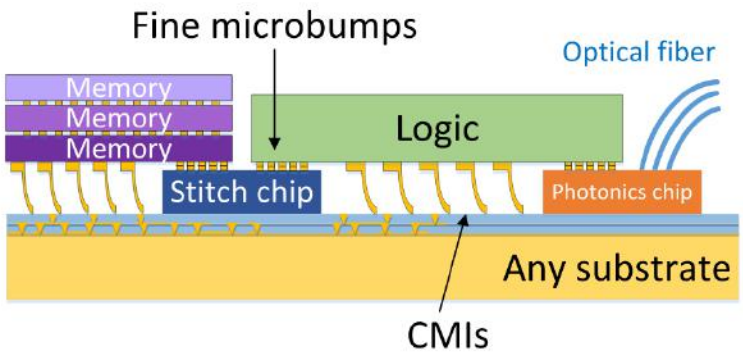


Figure 1.10: Schematic showing the HIST architecture: a stitch chip with high-density fine-pitch wires is placed between the substrate and the chiplets. Fine-pitch microbumps are used to bond chiplets to the stitch chip to provide high-bandwidth and low-energy signaling. Compressible microinterconnects (CMIs) are used to compensate for package nonplanarity and enable chiplet-package interconnection. Figure is obtained from [37].

an interposer is limited by both mechanical and economical constraints, which ultimately restrict the number of chips that can be assembled on top of the interposer. As a result, multiple interposers may be necessary to build a contiguous system enabled by the use of silicon bridges [36] (Fig. 1.8). Embedded Multi-die Interconnect Bridge (EMIB) technology [5] provides localized high-density fine-pitch interconnects between two or more dice on an organic package substrate, opening up new opportunities for heterogeneous on-package integration. As shown in Fig. 1.9, a thin silicon bridge is embedded within the top two layers of an organic package and connected to flip-chip pads on the package substrate through package vias. The on-package interconnect is not affected by the presence of the bridge which offers more localized high-density wiring. EMIB also has the advantages such that there is no practical limits to die size, no through-silicon-via (TSV) integration, and it can leverage existing organic substrate manufacturing. Recently, a Heterogeneous Interconnect Stitching Technology (HIST) [37] is demonstrated to establish the interconnection of multiple dice or chiplets of various functionalities, as shown in Fig. 1.10. Different from EMIB which requires embedded bridges, HIST is based on die-to-die face-to-face bonding, and thus there are no intermediate package levels, which enables higher signal I/O pitch and lower capacitance. HIST platform achieves a similar signal bandwidth density

as the silicon interposer technology [40] and exhibits many advantages similar to those of EMIB, including flexible designs, improved scalability, improved signaling, reduced cost, and versatile substrate choices.

1.7 Optical Coupling Structures

The co-integration of photonic chips on electronic packages containing logic digital integrated circuits presents a number critical challenges, one of which is efficient optical coupling on the package-level and the chip-level. Off-chip optical signals are usually introduced to the chip/package via optical fibers, which requires fiber-to-chip couplers. In addition, interlayer couplers are necessary to achieve optical coupling between two separate overlaid chips or between the layers of on-chip interconnect stack.

Common fiber-to-chip coupling schemes include vertical coupling and lateral coupling to on-chip waveguides. One benefit of vertical coupling is that fibers can be arranged in a 2D array, increasing the number of on-chip I/O ports. Vertical coupling is essentially achieved using diffraction gratings (Fig. 1.11(a)), which diffract light and change its propagation direction. 1D or 2D rectangular waveguide gratings have been comprehensively studied to date because their planar geometries make them compatible with IC wafer-scale fabrication and testing. Extensive research has been conducted to increase the grating coupling efficiencies. The most effective approaches have been choosing materials with large refractive index contrast (e.g. Silicon-On-Insulator (SOI) platform) and optimizing the grating period, etch depth, and fill factor [41], which fundamentally change the diffraction behavior. However, further efficiency improvement into the desired diffraction direction continues to be a challenge due to the presence of substrate leakage loss, back-reflection, and forward-transmission (Fig. 1.12). These problems have been mitigated by applying poly-Si overlayers (to enhance directionality) [42], asymmetric profiles (to enhance directionality) [43], distributed Bragg reflectors (DBR) or metal reflectors (to reduce

substrate leakage) [44], apodized or chirped gratings (to reduce back-reflection or to produce Gaussian profile) [45], in-plane grating reflectors (to reduce forward-transmission) [46], dual-grating layers [47], and novel subwavelength structures [48, 49]. Focusing gratings [50], whose grating rulings are curved, are also proposed to counteract the spreading of the diffracted fields. Some reported SOI gratings are summarized in Table 1.3. It is observed that narrow grating ridges ($\sim$100 nm) are required to achieve relatively high efficiencies ($>$70%), and such gratings need to be patterned using high-resolution fabrication techniques, e.g. e-beam lithography. The methods used to increase the directionality, e.g. depositing overlayers or applying reflectors, also add complexity to the fabrication process. Apart from the fabrication difficulties, grating assembly poses additional challenges to photonics integration and packaging processes because rectangular grating couplers are very sensitive to misalignments [51]. In addition, all the grating designs have relatively narrow spectral bandwidth, and those with 100 nm 1dB bandwidth have very low coupling efficiency. All these issues regarding resolution, directionality, misalignment and bandwidth limitation originate from the Floquet and Bragg conditions, which rectangular gratings must inherently follow. As a result, grating coupler may not simultaneously satisfy both broad bandwidth and high efficiency requirements for on-chip signaling.

Fiber lateral coupling includes butt coupling (or edge coupling, Fig. 1.11(b)) and evanescent coupling. Butt coupling of fibers to waveguides usually involves structures like v-grooves [65, 66] for holding the fibers parallel to the waveguides and inverse tapers [67] on top of the waveguides for mode conversion (spot size conversion). To further decrease the mode field diameter, some form of lenses, e.g. conventional lens [68], axicon lens [69], surface-relief lens [70], and graded index lens [71], etc., are fabricated on the tips of the fibers. In the evanescent coupling scheme, fibers are stripped off the cladding and placed in adjacent to waveguides [72] or nanoribbons [73]. The stripped fibers need to be side-polished and the coupling length is relatively large (on the order of 100 μm), which is generally difficult to fabricate and less applicable to compact chip-level integration. Thus,

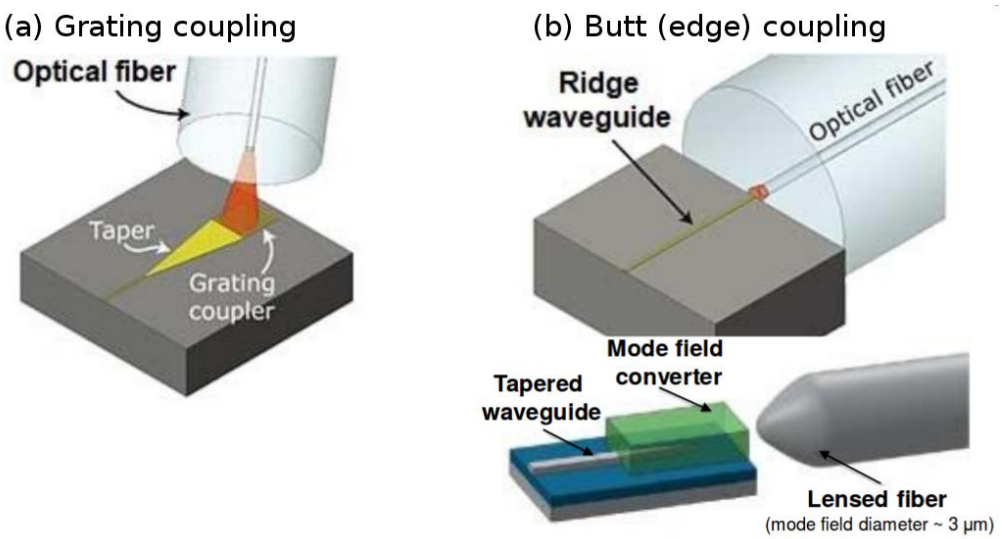


Figure 1.11: Common fiber coupling scheme including (a) vertical grating coupling and (b) lateral butt (edge) coupling.

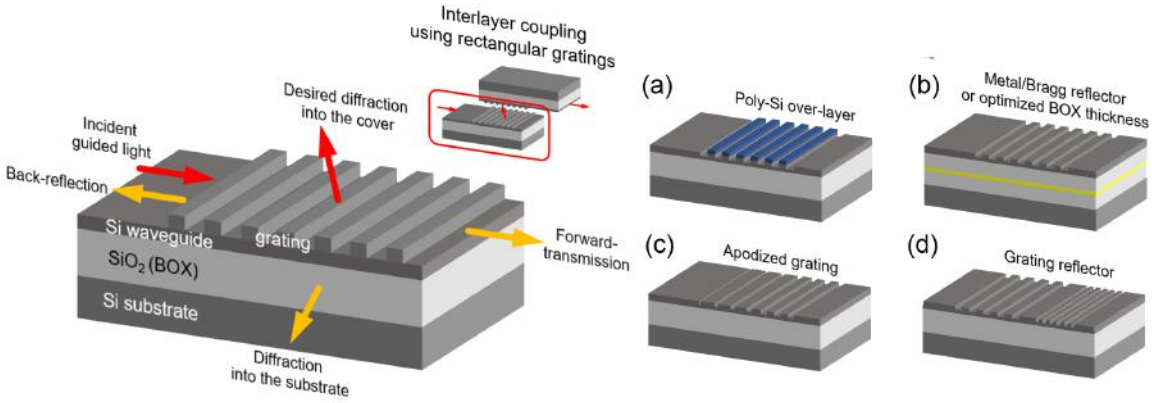


Figure 1.12: Optical losses present in rectangular gratings used for interlayer coupling. Possible solutions to reduce losses in a binary SOI grating out-coupler are (a)(b) increasing directionality to cover, (c) reducing back-reflection, and (d) reducing forward-transmission.

Table 1.3: Some reported SOI gratings for TE polarization. The column labels represent Si waveguide thickness (t_{Si}), grating period (Λ), grating etch depth (t_g), grating fill factor (f), grating line resolution (δ), grating length or dimension (L), single-grating efficiency (η_s), working wavelength (λ), and 1dB or 3dB bandwidth (1dB/3dB). CP, AP and DE represent chirped grating period, apodized fill factor, and double etch depths/widths, respectively.

SOI Structure						Measurement			Simulation			Reference
t_{Si} [nm]	Λ [nm]	t_g [nm]	f	δ [nm]	L [μm]	η_s [% / dB]	λ [nm]	1dB / 3dB [nm]	η_s [% / dB]	λ [nm]	1dB / 3dB [nm]	and Comments
220	CP	220	AP	120	11	35 / -4.6	1536	47 / –	72 / -1.4	1550	38 / –	[52] [a]
250	CP	–	AP	60	14	64 / -1.9	1524	42 / 70	74 / -1.3	1520	50 / 80	[53] [a]
220	640	70	0.5	320	12.8	91 / -0.41 [b]	1590	–	94 / -0.28 [b]	1550	–	[54] [c]
220	630	70	0.5	315	15.75	69.5 / -1.6	1520	36 / 63	76 / -1.2	1550	40 / 60	[55] [d]
220	CP	100	AP	30	13	–	–	–	88 / -0.6	1550	– / 40	[56] [d]
220	630	70	0.5	315	–	–	–	–	79 / -1.0	1550	–	[57] [d]
220	610	70	0.5	305	12.2	55 / -2.6	1530	50 / –	66 / -1.8	1550	55 / –	[42] [e]
340	630	340	0.85	95	10	44 / -3.6	1560	– / 50	–	–	–	[58]
340	CP	200	AP	43	15	75.8 / -1.2	1533	– / 45	79 / -1.0	1550	– / 48	[45]
220	701	220	0.5	350	14	–	–	–	49 / -3.1	1550	20 / 35	[59]
220	568	DE	AP	100	–	65 / -1.9	1310	23 / –	85 / -0.7	1310	25 / –	[60] [f]
220	780	DE	DE	100	30	74 / -1.3	1550	– / 52	78 / -1.1	1550	– / 50	[48] [f]
220	663	DE	AP	87	15	71 / -1.5	1533	– / 49	87 / -0.6	1550	– / 50	[49] [f]
510	545	350	AP	–	–	–	–	–	81 / -0.9	1560	42 / –	[61] [f]
380	430	70	0.5	215	14	43 / -3.7	1310	–	–	–	–	[62]
300	565	150	0.77	130	12.4	–	–	–	78 / -1.1	1550	– / 65	[47] [g]
120	–	–	–	–	–	92 / -0.36	1197	–	95 / -0.2	1200	– / 100	[63] [g]
250	420	250	0.54	227	10	–	–	–	85.3 / 0.69	1310	– / 20	[64] [h]

[a] Optimized BOX layer thickness [b] η_s approximated by $\sqrt{\eta_c}$ [c] Gold reflector [d] Bragg reflector [e] Poly-Si overlay
[f] Subwavelength structure [g] Dual-grating layer [h] In-plane reflector

the fiber butt coupling scheme is more commonly used, and some representative research results are summarized in Table 1.4. It is observed that the coupling efficiency is relatively high compared with the vertical grating coupling approach. Another advantage of lateral coupling is that the spectral bandwidth is much wider because there is no wavelength selectivity involved in the physics of this coupling scheme. Nevertheless, the I/O density achieved using lateral coupling is limited due to the much larger diameter of optical fibers and limited space along each edge of the chip.

As technology moves toward heterogeneous multi-die integration within a single package using 2.5D and 3D integrated circuits [80], out-of-plane interlayer optical connectors are indispensable for within- and off-package optical connectivity. Potential interlayer couplers include 45° mirrors, evanescent couplers, and diffraction gratings. Reflection mirrors are commonly used in the early stage of photonic integration due to their simple physics and relatively easy fabrication process. They are usually imprinted into polymer waveguides or etched along crystallographic planes of Si waveguides (resulting in 54.7° slant angle). However, the coupling efficiency of reflection mirrors is limited by the size of the mirror and material interfaces. Further efficiency reduction may occur due to guided mode mismatch when the input waveguide and the receiving waveguide are made of different materials. In addition, their bulky shapes also make reflection mirrors unsuitable for compact integration. Evanescent waveguide couplers are commonly used in in-plane optical modulators, interferometers, amplifiers [81, 82, 83], etc. When the two adjacent waveguides are stacked vertically, the evanescent couplers can be used for vertical interlayer coupling. Nevertheless, the interlayer gap needs to be relatively small in order to achieve high coupling efficiency, which is likely to induce interference between other circuits components on the two layers. Diffraction gratings are promising because they can achieve relatively large interlayer distances and their planar geometries make them compatible with compact photonic integration. Since diffraction gratings are useful for both fiber coupling and interlayer coupling, they will be the focus of the next sections.

Table 1.4: Some reported fiber-butt/edge coupling for TE polarization.

Waveguide Geometry			Lensed Fiber Type	Coupling Structure			Wavelength [nm]	Coupling Efficiency [% / dB]	Reference
Material	Height [nm]	Width [nm]		Spot Size Converter Cross-Section [μm $\times$ μm]	Waveguide Taper Tip Width [nm]	Waveguide Taper Length [μm]			
SOI	220	590	Spherical	3 $\times$ 3 (Polymer)	175	175	1550	64 / -1.9	[68]
SOI	300	300	No lens	3 $\times$ 3 (Polymer)	60	200	1560	83 / -0.8	[67]
SOI	220	450	No lens	2 $\times$ 2 (Polymer)	75	150	1550	79 / -1.4	[74]
SOI	250	500	Spherical	No converter	160	100	1550	85 / -0.7	[75]
SOI	240	500	No lens	0.9 $\times$ 4.5 (Si_3N_4)	Fork shape		1550	78 / -1.1	[76]
SOI	150	–	No lens	Metamaterial (SiO_2)	175	875	1310	74 / -1.3	[77]
Index 2.05	400	700	No lens	4.5 $\times$ 4.5 (index 1.45)	110	350	–	79 / -1	[78]
InP	–	–	3D printed	3D printed converter			–	87 / -0.6	[79]

1.8 Diffraction Grating Couplers

Diffraction gratings, which can be categorized as surface-relief gratings and volume holographic gratings, have many applications besides optical couplers. Surface-relief gratings, with variations in the surface profile, have been widely used in the field of photonics, (e.g. couplers [84, 85], polarizers [86], modulators [87], and switches [88], etc.) as well as other fields such as optoelectronics [89, 90], sensors [91, 92], and antennas [93]. With reduced thicknesses, planar geometries, and CMOS compatible materials, surface-relief gratings have become a popular solution in compact interconnect technologies. Volume holographic gratings, with variations in the refractive index of the material which are mostly polymer-based and glass-based, have been extensively explored in the field of data storage [94], optical correlation [95], optical information encryption [96], fiber communication [97], and spectroscopy [98], etc., due to their large bandwidth storage capability and high sensitivity in wavelength and angle. In order to fully exploit the benefits of diffraction gratings, it is important to study their diffraction behaviors using grating theories.

1.9 Diffraction Grating Theory

A number of theories have been proposed to study the grating diffraction phenomenon, including rigorous approaches and approximation methods. The rigorous approaches can be classified into two categories, namely the integral methods and the differential methods. The rigorous approaches usually give numerically accurate results but can potentially be computationally expensive, while approximation methods yield simpler but less accurate solutions. Nevertheless, all theories begin with the scalar wave equation $\nabla^2 F + k^2 n^2 F = 0$ and represent the diffracted fields ($F = E_y$ for transverse electric (TE) polarization or H_y for transverse magnetic (TM) polarization) in a plane wave expansion (also called Floquet modal expansion or Rayleigh expansion) as $F(x,z) = \sum_{i=-\infty}^{\infty} C_i \exp(jk_{x,i}x)\exp(jk_{z,i}z)$.

1.9.1 Integral and Differential Method

The integral methods involve the calculation of an integral equation (and sometimes of coupled integral equations), while the differential methods require the construction of an infinite system of coupled differential equations. The integral methods were first developed to analyze the diffraction problem by perfectly conducting (metallic) gratings in the visible spectrum. As more advances in dielectric gratings in the UV and IR ranges took place, the integral methods have become inadequate. Differential methods are developed to study dielectric gratings and more complex objects in the spectral region where integral methods are less efficient. Differential methods can be classified into two types: 1) representing fields using a suitable basis (for gratings, the periodicity of the fields leads to exponential basis) and casting the field expressions into a set of ordinary differential coupled equations (with one variable), and 2) discretizing the problem space into grids and directly solving partial differential equations (multiple variables) derived from Maxwell's equations at each grid point using a finite difference scheme [99]. The first type usually results in matrix formulations, and the numerical determination of these matrices usually governs the feasibility of the method. For the second type, the formulation is simple but requires long computing time and is limited by the numerical instabilities of the algorithm.

1.9.2 Modal Approach

Modal approach, a differential method based on eigenmode expansions, analyzes the diffraction problem in which a plane wave is incident on a grating bounded by two different media on two sides. In this approach, the grating is divided into layers featuring parallel interfaces. The permittivity varies only in the plane of the layer and is assumed to be constant along the perpendicular direction allowing the separation of spatial variables. Within each layer the eigenmodes of the electromagnetic field are calculated and the general solution is then expressed by means of an eigenmode expansion [100]. The expansion coefficients can be found by solving the matrix equation generated by proper boundary conditions. In practice,

the expansion has to be truncated to obtain a finite dimensional matrix. The computation time of this method is generally low because all the calculations are explicit. The only time limiting step is the matrix inversion, which is based on the number of diffraction orders used. This method is only suitable for certain profiles, e.g. 1D gratings, while it is insufficient to treat profiles involving complicated 2D mathematics.

1.9.3 Rigorous Coupled-Wave Analysis

Rigorous Coupled-Wave Analysis (RCWA) [101], also called Floquet modal method, has been used to analyze diffraction of both surface-relief gratings and volume gratings. It is based on the Modal approach but uses a different set of predetermined basis functions to express the fields in the grating. An advantage of RCWA is that additional layers, e.g. a waveguide or another grating, can be added into the analysis. Due to its efficient computation inherited from the modal approach and its capability of analyzing waveguide grating structures, RCWA has become a popular theory to study diffraction gratings and it is also the basis theory used in this work. However, RCWA is typically applied to solve reflected and transmitted diffraction in the plane normal direction (perpendicular to the plane of grating) and it is not applicable to surface waves (also referred as leaky wave due to attenuations) parallel to the interfaces. Therefore, Leaky Wave (RCWA-LW) approach [102] and the proposed "Equivalent Index Slab" method (RCWA-EIS) are suggested to analyze the waveguide grating in-coupling or out-diffraction where surface waves are involved. The comparison between RCWA-LW method and RCWA-EIS method will be discussed in Chapter 2.

1.9.4 Transmission Line Approach

Transmission line approach is similar to RCWA except it uses voltages V_{mn} and currents I_{mn} to represent the coupled space harmonics in the grating layer. The parameter V_{mn} is related to I_{mn} via $V_{mn} = \sum_r Z^{(g)}_{m,n-r} I_{mr}$, where $Z^{(g)}_{m,n-r}$ is an impedance that represents the

coupling of the rth harmonic of the magnetic field to the nth harmonic of the electric field. The value of $Z^{(g)}_{m,n-r}$ is determined by the Maxwell's equations. Similarly, the parameters I_{mn} is related to V_{mn} via $I_{mn} = \sum_r Y^{(g)}_{m,n-r} V_{mr}$ by admittance $Y^{(g)}_{m,n-r}$. Transmission line approach can solve the surface wave problem in a similar way as the RCWA-LW approach. From a different perspective, it treats the grating structure as a transmission line equivalent network, and it is easier to calculate the reflection coefficient at each interface, e.g. $\rho_n = (Y^g_n - Y^s_n)/(Y^g_n + Y^s_n)$ at the grating-substrate interface [103]. The downside of the transmission line approach is that it is not intuitive to get the electric and magnetic field expressions which require additional conversions from the voltages and currents.

1.9.5 Coupled Wave Theory

Coupled wave theory, proposed by Kogelnik [104], can be categorized as an approximation method because only the incident reference wave R and the outgoing signal wave S that obey the Bragg condition are considered while the other diffraction orders are neglected. This assumption limits the coupled wave theory to thick hologram gratings. The model is based on a set of two equations (the coupled wave equations) regarding S and R as $C_R R' + \alpha R = -j\kappa S$ and $C_S S'(\alpha + j\theta)S = -j\kappa R$, where κ is the coupling constant, which indicates the energy interchange between S and R. The solution to the coupled wave equations is in the general form $R = r_1 \exp(\gamma z) + r_2 \exp(-\gamma z)$ and $S = s_1 r_1 \exp(\gamma z) + s_2 r_2 \exp(-\gamma z)$. The diffraction efficiency of the volume grating can be determined as $\eta = (|C_S| S S^*)/C_R$.

1.9.6 Perturbation Method

As a type of approximation method, perturbation method is commonly used to analyze surface-relief gratings with small perturbations whose relative permittivity can be written analytically. The field in the grating is first expressed in a plane wave expansion $E_y(x, z) = \sum_m \tilde{E}_m \exp(j\beta_m z)$. The grating permittivity, expressed in a Fourier series, and the field

expansion are substituted into the wave equation, resulting in a second-order differential equation in $\tilde{E}_m$ whose solution is subject to the continuity of E_y and H_z at the boundaries. The unperturbed case, e.g. uniform waveguide, is first determined by considering only the 0th order term in the second-order differential equation, and then the calculated $\tilde{E}_m(0)$ and β_0 are used to approximate higher order terms. This process provides a first-order perturbation solution for $\tilde{E}_m$ [105].

1.9.7 Coupled Mode Theory

Coupled mode theory, an essential perturbation approach, is useful for treating problems involving energy exchange between modes, e.g. grating diffraction, evanescent coupling, electrooptic modulation, photoelastic and magnetooptic modulation, and optical filtering, etc. This theory essentially solves the coupled equations for two electromagnetic modes $a(z,x,t) = A\exp[j(\omega_a t \pm \beta_a z)]f_a(x)$ and $b(z,x,t) = B\exp[j(\omega_b t \pm \beta_b z)]f_b(x)$ under perturbation, where the complex amplitudes A and B vary according to the coupled mode equations $\frac{dA}{dz} = \kappa_{ab}B\exp(-j\Delta z)$, and $\frac{dB}{dz} = \kappa_{ba}A\exp(+j\Delta z)$, where Δ is the phase-mismatch constant, and κ is the coupling coefficient [106]. The solutions to the coupled mode equations depend on the power conservation condition, mode propagation directions, and boundary conditions.

1.9.8 Commercial PDE Solver

With simple user interfaces, commercial software packages such as Lumerical FDTD (finite-difference time-domain), COMSOL multiphysics (finite element method, FEM), or CAMFR (CAvity Modelling FRamework, eigenmode expansion) are popular tools to simulate grating structures. FDTD and FEM are based on the second type differential method, while CAMFR is based on the first type. However, it is typically time-consuming to use these tools for diffraction problem analysis. FDTD discretizes the problem in space and time, and it generally requires more computation points (higher resolution) to properly simulate

a complicated structure in 3D, which greatly increases computation time. FEM requires the generation of connected meshes and nodal matrices which will be problematic when treating and updating large and complex geometries. What's more, FEM heavily relies on numerical integration which is a typically slow calculation process. CAMFR treats the field as a sum of local eigenmodes in each z-invariant layer. It approximates the grating layer as a uniform dielectric layer, and thus it is challenging to simulate complex structures which require large number of modes.

1.9.9 Optimization Solver

Optimization is a critical step to design efficient waveguide grating couplers. Optimization algorithms, such as genetic algorithm [107, 108], partical swarm [108, 109], simulated annealing [110, 111], trust regions method [112], interior point method [113], and gradient descent method [114], etc., need to be used in addition to the grating theories. The optimization algorithms take grating parameters (e.g. period, grating depth, fill factor and diffraction angle, etc.) as variables to optimize the target parameter (diffraction efficiency calculated by grating theories). This process involves multiple iterations and comparisons which greatly increase computation loads. As a result, an accurate and efficient grating theory, such as RCWA, is helpful in speeding up the optimization.

1.10 Research Overview

The objective of the dissertation is to theoretically design and experimentally demonstrate optical couplers for 2.5D/3D heterogeneous integrated electronics. Specifically, this thesis focuses on the design, simulation, fabrication, integration and testing of optical couplers for fiber coupling and interlayer coupling applications. In Chapter 2, a new concept, the "Equivalent Index Slab (EIS)" method, which extends the Rigorous Coupled-Wave Analysis (RCWA) to rectangular grating diffraction involving surface waves, is proposed to offer a comprehensive study of rectangular grating couplers with various periodic profiles in

terms of efficiency optimization and angular misalignment analysis. Based on the understanding of rectangular grating performance and limitations, Chapter 3 introduces a fundamentally new coupling structure, Grating-Assisted-cylindrical-Resonant-Cavities (GARC) coupler, to achieve efficient and broadband interlayer coupling in 3D integrated photonics. GARC coupler is based on evanescent field coupling between waveguides and the interconnecting via, and the via serves as a cylindrical resonant cavity which is further assisted by the circular gratings to enhance the field. Chapter 4 switches focus from interlayer coupling to fiber coupling, and a passive self-aligning fiber-to-chip coupling method, Fiber-Interconnect Silicon Chiplet Technology (FISCT), is demonstrated using a combination of silicon micromachining and 3D microprinting to achieve efficient and convenient fiber assembly. Lastly in Chapter 5, the thesis is concluded by potential future work.

CHAPTER 2

RECTANGULAR WAVEGUIDE GRATING COUPLER: EFFICIENCY OPTIMIZATION AND ANGULAR MISALIGNMENT ANALYSIS

Rigorous Coupled-Wave Analysis (RCWA) is an accurate and efficient method to calculate grating reflection and transmission efficiencies of various diffraction orders in the plane normal direction (z direction). RCWA assumes that the grating is infinitely long in the x direction, and its formulation relies on energy conservation in the z direction. As a result, there is no direct relation between the in-coupling power and the guided power (propagating in the x direction), and thus RCWA doesn't provide the in-coupling/out-diffraction efficiencies of waveguide gratings in which surface waves (or leaky waves) are involved. RCWA-Leaky Wave (LW) approach is proposed to investigate the out-diffraction process of thick volume gratings made of polymer [102], but it is tedious to implement and less practical for thin SOI gratings. RCWA-LW approach relies on the determination of a complex propagation constant $\tilde{\gamma} = \beta - j\alpha$, which is necessary for the calculation of out-diffraction efficiencies. The parameter $\tilde{\gamma}$ is obtained such that the determinant of the boundary condition matrix $\tilde{\mathbf{M}}$ is zero, e.g., $\det(\tilde{\mathbf{M}}) = 0$, and in RCWA-LW approach, this calculation is realized by the Muller method, which requires careful choices of initial guesses in order to find meaningful converged results. The difficulty in finding $\tilde{\gamma}$ is more apparent in the analysis of SOI gratings. This is because the calculation of $\tilde{\gamma}$ is a numerical sensitive step limited by both the condition of the matrix and the precision of the computer. The boundary condition matrix $\tilde{\mathbf{M}}$, whose dimension is based on the total number of grating interfaces and the number of truncated diffraction orders, is usually sparse and big. The matrix $\tilde{\mathbf{M}}$ is ill-conditioned when a grating with large refractive index contrasts, e.g. SOI grating, is under evaluation. The limited floating point arithmetic of the computer also gives additional round-off errors. In addition, RCWA-LW approach is insufficient to study the in-coupling

process, so it can't adopt the useful 3D RCWA formulations of conical incidence.

In this chapter, a new concept, "Equivalent Index Slab (EIS)", is proposed to determine the complex propagation constant $\tilde{\gamma}$, which circumvents the problem of solving the determinant of a large-dimension matrix [41]. This extends the applicability of RCWA-EIS method to arbitrary material systems. In conjunction with optimization algorithms in Matlab or Python, RCWA-EIS method can be applied to optimize grating in-couplers as well as out-couplers due to light reciprocity. Since RCWA-EIS formulations take into account the in-coupling process, conical incidence on the waveguide grating can be considered, which facilitates the analysis of angular misalignment effects on the interlayer grating coupling.

2.1 RCWA-EIS Method

RCWA-EIS method starts with the formulations of the conventional RCWA [115]. A surface-relief binary waveguide grating in the transverse electric (TE) polarization (E_y, H_x and H_z) is considered as an example. As shown in Fig. 2.1, the waveguide grating structure can be divided into four layers, namely cover, grating, waveguide, and substrate. The grating layer is composed of a periodic distribution of grating ridges (with refractive index n_{rd}) and grating grooves (with refractive index n_{gr}), and the fraction of the grating ridge with respect to the whole period Λ is indicated by the fill factor f. For non-binary grating profiles (arbitrary periodic shapes), the entire grating structure is horizontally sliced into sublayers which can be approximated as a series of binary gratings with varying fill factors or ridge locations [116, 117]. A plane wave with free-space wavelength λ_0 is obliquely incident onto the grating at an angle θ.

2.1.1 Step 1: Field Calculation by Conventional RCWA

For simplicity, the formulations introduced here focus on the binary grating. The relative permittivity of the binary grating can be expanded in a Fourier series along the x direction

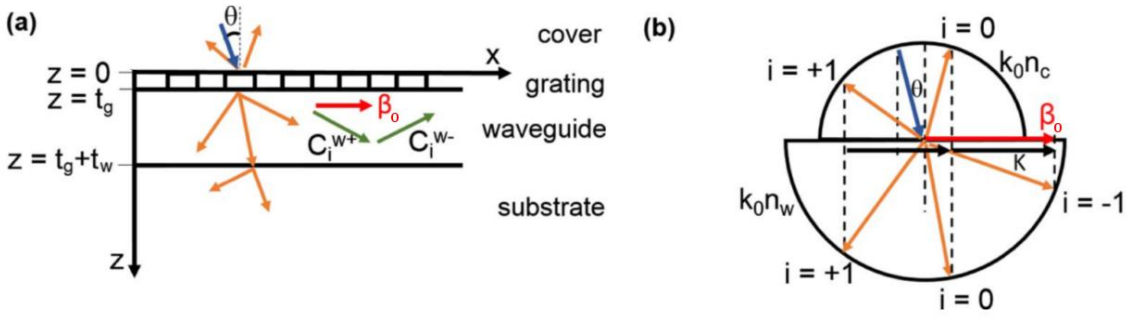


Figure 2.1: Schematic representation of (a) the grating in-coupling process and (b) the phase diagram based on the Floquet condition. This example shows multiple diffraction orders in the cover and substrate (not optimized for high efficiency), and the $i = -1$ order is a possible guided order whose propagation constant in the x direction is approximately equal to the guided mode propagation constant β_0 in the waveguide.

in the form

$$\epsilon_g(x) = \epsilon_{g0} + (\epsilon_{rd} - \epsilon_{gr}) \sum_{h=-\infty, h\neq 0}^{\infty} \frac{\sin(\pi h f)}{\pi h} \exp(jhKx), \tag{2.1}$$

where ϵ_{g0} is the average grating permittivity defined as $\epsilon_{g0} = {n_{rd}}^2 f + {n_{gr}}^2(1-f)$, and K is the grating vector magnitude ($K = 2\pi/\Lambda$). For the following analysis, $\tilde{\epsilon}_h$ is defined as

$$\tilde{\epsilon}_0 = \epsilon_{g0}, \tag{2.2}$$

$$\tilde{\epsilon}_h = (\epsilon_{rd} - \epsilon_{gr}) \frac{\sin(\pi h f)}{\pi h}. \tag{2.3}$$

The grating permittivity definitions of asymmetric surface-relief gratings, such as parallelogramic gratings and sawtooth gratings, and volume holographic gratings are introduced in Appendix A.

The periodic structure perturbs the incident plane wave into multiple discrete directions indicated by the diffraction orders, and it produces both forward-diffracted (transmitted) waves ($+z$ direction) and backward-diffracted (reflected) waves ($-z$ direction). Since the incident wave is TE polarized, the electric field has only a y component while magnetic field has both x and z components. The electric fields in all layers are expressed as plane

wave expansions $E_y = \sum_i Z(z)\exp(-jk_{x,i}x)$, where i indicates the ith diffraction order, $Z(z)$ is a function in the z direction containing unknown field amplitudes, and $k_{x,i}$ is the propagation constant in the x direction.

Specifically, the electric field in the cover is the sum of the incident and backward-diffracted waves. The normalized total electric field in the cover is expressed as

$$\begin{aligned} E_{cy}(x,z) = &\exp[-jk_0 n_c(\sin\theta x + \cos\theta z)] \\ &+ \sum_i R_i \exp(-jk_{x,i}x + jk_{cz,i}z), \end{aligned} \tag{2.4}$$

and the normalized electric field in the substrate is expressed as

$$E_{sy}(x,z) = \sum_i T_i \exp[-jk_{x,i}x - jk_{sz,i}(z - t_g - t_w)]. \tag{2.5}$$

The electric field inside the slab waveguide is a superposition of forward-propagating ($+z$) waves and backward-propagating ($-z$) waves expressed as

$$\begin{aligned} E_{wy}(x,z) = \sum_i \{&C_i^{w+} \exp[-jk_{wz,i}(z - t_g)] \\ &+ C_i^{w-} \exp[+jk_{wz,i}(z - t_g - t_w)]\} \exp(-jk_{x,i}x), \end{aligned} \tag{2.6}$$

where k_0 is the wave vector magnitude in free space ($k_0 = 2\pi/\lambda_0$), the summation is from $i = -(s-1)/2$ to $(s-1)/2$, s is the total number of diffraction orders (an odd number for calculation convenience), R_i and T_i are normalized *i*th backward-diffracted (reflected) amplitude and forward-diffracted (transmitted) amplitude, respectively, t_g and t_w are grating layer thickness and waveguide thickness, respectively, $k_{x,i}$ is the *i*th propagation constant in the x direction defined by the Floquet condition (Fig. 2.1(b))

$$k_{x,i} = k_0 n_c \sin\theta - iK, \tag{2.7}$$

and $k_{rz,i}$ is the propagation constant in the z direction defined as

$$k_{rz,i} = \begin{cases} \sqrt{{n_r}^2{k_0}^2 - {k_{x,i}}^2} & , n_r k_0 > |k_{x,i}| \\ -j\sqrt{{k_{x,i}}^2 - {n_r}^2{k_0}^2} & , n_r k_0 < |k_{x,i}| \end{cases} \quad r = c, w, s. \tag{2.8}$$

According to Maxwell's equation $\bar{H} = \frac{j}{\omega\mu}\nabla \times \bar{E}$, the tangential magnetic fields in the cover, substrate, and waveguide are expressed as

$$\begin{aligned} H_{cx}(x,z) =& \frac{1}{j\omega\mu_0}\{-jk_0 n_c \cos\theta \exp[-jk_0 n_c(\sin\theta x + \cos\theta z)] \\ &+ \sum_i jk_{cz,i}R_i \exp(-jk_{x,i}x + jk_{cz,i}z)\}, \end{aligned} \tag{2.9}$$

$$H_{sx}(x,z) = \frac{1}{j\omega\mu_0}\sum_i (-jk_{sz,i})T_i \exp[-jk_{x,i}x - jk_{sz,i}(z - t_g - t_w)], \tag{2.10}$$

$$\begin{aligned} H_{wx}(x,z) =& \frac{1}{j\omega\mu_0}\sum_i \{C_i^{w+}(-jk_{wz,i})\exp[-jk_{wz,i}(z - t_g)] \\ &+ C_i^{w-}(+jk_{wz,i})\exp[+jk_{wz,i}(z - t_g - t_w)]\}\exp(-jk_{x,i}x). \end{aligned} \tag{2.11}$$

The electric field in the grating region ($0 < z < t_g$) is expressed by a Fourier expansion in spatial harmonics as

$$E_{gy}(x,z) = \sum_i S_{gy,i}\exp(-jk_{x,i}x), \tag{2.12}$$

and the tangential magnetic field in the grating region is derived from Maxwell's equation as

$$H_{gx}(x,z) = -j\sqrt{\frac{\epsilon_0}{\mu_0}}\sum_i U_{gx,i}\exp(-jk_{x,i}x), \tag{2.13}$$

where $U_{gx,i} = \frac{1}{k_0}\frac{\partial S_{gy,i}}{\partial z}$. Equation (2.12) satisfies the wave equation for TE polarization in the grating region

$$\nabla^2 E_{gy} + k_0^2 \epsilon_g(x) E_{gy} = 0. \tag{2.14}$$

Equation (2.12) is substituted into Eq. (2.14) and the following equation can be derived

$$\frac{\partial^2 S_{gy,i}}{\partial z^2} + (k_0^2 \epsilon_g - k_{x,i}^2) S_{gy,i} = 0, \tag{2.15}$$

which can be written in a matrix

$$\left[\frac{\partial^2 \mathbf{S}_{gy}}{\partial (z')^2}\right] = \left[\mathbf{A}_g\right] \left[\mathbf{S}_{gy}\right], \tag{2.16}$$

where $z' = k_0 z$, $\mathbf{A}_g = \mathbf{K}_x{}^2 - \mathbf{E}_g$, $\mathbf{K}_x$ is a diagonal matrix with (i,i)th element being k_{xi}/k_0, and $\mathbf{E}_g$ is the matrix of permittivity coefficient $\tilde{\epsilon}_h$ shown in Eq. (2.2) and (2.3)

$$\mathbf{E}_g = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \dots & \tilde{\epsilon}_0 & \tilde{\epsilon}_{-1} & \tilde{\epsilon}_{-2} & \tilde{\epsilon}_{-3} & \tilde{\epsilon}_{-4} & \dots \\ \dots & \tilde{\epsilon}_1 & \tilde{\epsilon}_0 & \tilde{\epsilon}_{-1} & \tilde{\epsilon}_{-2} & \tilde{\epsilon}_{-3} & \dots \\ \dots & \tilde{\epsilon}_2 & \tilde{\epsilon}_1 & \tilde{\epsilon}_0 & \tilde{\epsilon}_{-1} & \tilde{\epsilon}_{-2} & \dots \\ \dots & \tilde{\epsilon}_3 & \tilde{\epsilon}_2 & \tilde{\epsilon}_1 & \tilde{\epsilon}_0 & \tilde{\epsilon}_{-1} & \dots \\ \dots & \tilde{\epsilon}_4 & \tilde{\epsilon}_3 & \tilde{\epsilon}_2 & \tilde{\epsilon}_1 & \tilde{\epsilon}_0 & \dots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{2.17}$$

For a binary grating whose grating profile is symmetric, $\mathbf{E}_g$ is a symmetric matrix in which $\tilde{\epsilon}_{-h} = \tilde{\epsilon}_h$.

Equation (2.16) is a typical second-order homogeneous system of differential equations and its solution can be expressed by eigenvalues and eigenvectors of matrix $\mathbf{A}_g$. As a result, the space harmonic expansions of tangential electric and magnetic fields in the grating are

expressed as

$$S_{gy,i}(z) = \sum_{p=1}^{s} w_{i,p}^{g} \{ C_p^{g+} \exp(-k_0 q_p^g z) + C_p^{g-} \exp[k_0 q_p^g (z - t_g)] \}, \tag{2.18}$$

$$U_{gx,i}(z) = \sum_{p=1}^{s} v_{i,p}^{g} \{ -C_p^{g+} \exp(-k_0 q_p^g z) + C_p^{g-} \exp[k_0 q_p^g (z - t_g)] \}, \tag{2.19}$$

where $w_{i,p}^g$ is the (i,p)th element of the eigenvector matrix $\mathbf{W}_g$ and q_p^g is the positive square root of the (p,p)th element of the eigenvalue matrix $\mathbf{Q}_g$. According to Eq. (2.13), it can be found that $v_{i,p}^g = w_{i,p}^g q_p^g$ and therefore $\mathbf{V}_g = \mathbf{W}_g \mathbf{Q}_g$. C_p^{g+} and C_p^{g-} are unknown coefficients which will be determined from boundary conditions. The term $\exp(-k_0 q_p^g z)$ represents forward-propagating $(+z)$ waves and the term $\exp[k_0 q_p^g (z - t_g)]$ represents backward-propagating $(-z)$ waves in the grating region.

The coefficients for the cover, grating, waveguide, and substrate R_i, C_p^{g+}, C_p^{g-}, C_i^{w+}, C_i^{w-}, and T_i can be determined by matching the tangential electric field E_y [Eq. (2.4), (2.12), (2.6), (2.5)] and the tangential magnetic field H_x [Eq. (2.9), (2.13), (2.11), (2.10)] at all boundaries. At $z = 0$ (boundary between cover and grating):

$$\begin{bmatrix} \delta_{i0} \\ j n_c \cos\theta \delta_{i0} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -j\mathbf{Y}_c \end{bmatrix} \begin{bmatrix} \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_g & \mathbf{W}_g \mathbf{X}_g \\ \mathbf{V}_g & -\mathbf{V}_g \mathbf{X}_g \end{bmatrix} \begin{bmatrix} \mathbf{C}_g^+ \\ \mathbf{C}_g^- \end{bmatrix}; \tag{2.20}$$

at $z = t_g$ (boundary between grating and waveguide):

$$\begin{bmatrix} \mathbf{W}_g \mathbf{X}_g & \mathbf{W}_g \\ \mathbf{V}_g \mathbf{X}_g & -\mathbf{V}_g \end{bmatrix} \begin{bmatrix} \mathbf{C}_g^+ \\ \mathbf{C}_g^- \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{X}_w \\ j\mathbf{Y}_w & -j\mathbf{Y}_w \mathbf{X}_w \end{bmatrix} \begin{bmatrix} \mathbf{C}_w^+ \\ \mathbf{C}_w^- \end{bmatrix}; \tag{2.21}$$

at $z = t_g + t_w$ (boundary between waveguide and substrate):

$$\begin{bmatrix} \mathbf{X}_w & \mathbf{I} \\ \mathbf{X}_w & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_w^+ \\ \mathbf{C}_w^- \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{Y}_{s/w} \end{bmatrix} \begin{bmatrix} \mathbf{T} \end{bmatrix}, \tag{2.22}$$

where $\mathbf{X}_g$, $\mathbf{X}_w$, $\mathbf{Y}_c$, $\mathbf{Y}_w$ and $\mathbf{Y}_{s/w}$ are diagonal matrices with diagonal elements $\exp(-k_0 q_p^g t_g)$, $\exp(-jk_{wz,i}t_w)$, $k_{cz,i}/k_0$, $k_{wz,i}/k_0$, and $k_{sz,i}/k_{wz,i}$, respectively.

The matrix equations can be merged into a nonhomogeneous system of equations as

$$\begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{W}_g & \mathbf{W}_g\mathbf{X}_g & \mathbf{0} & \mathbf{0} \\ j\mathbf{Y}_c & \mathbf{0} & \mathbf{V}_g & -\mathbf{V}_g\mathbf{X}_g & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_g\mathbf{X}_g & \mathbf{W}_g & -\mathbf{I} & -\mathbf{X}_w \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_g\mathbf{X}_g & -\mathbf{V}_g & -j\mathbf{Y}_w & j\mathbf{Y}_w\mathbf{X}_w \\ \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{X}_w & \mathbf{I} \\ \mathbf{0} & -\mathbf{Y}_{s/w} & \mathbf{0} & \mathbf{0} & \mathbf{X}_w & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ \mathbf{T} \\ \mathbf{C}_g^+ \\ \mathbf{C}_g^- \\ \mathbf{C}_w^+ \\ \mathbf{C}_w^- \end{bmatrix} = \begin{bmatrix} \delta_{i0} \\ jn_c \cos\theta\delta_{i0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{2.23}$$

represented by $\mathbf{Mx} = \mathbf{b}$, and the field amplitudes $\mathbf{x}$ can be calculated by $\mathbf{x} = \mathrm{inv}(\mathbf{M})\mathbf{b}$. The matrix inversion can be calculated by $\mathrm{inv}(\mathbf{M}) = \mathbf{V}(1/\mathbf{S})\mathbf{U}^T$, where $\mathbf{U}$, $\mathbf{S}$, and $\mathbf{V}$ are matrices obtained from the singular value decomposition $\mathbf{M} = \mathbf{USV}^T$. The amplitudes can also be determined by a transfer matrix approach regarding Eq. (2.20) to (2.22) [116], which is the preferred method to ensure numerical stability for asymmetrical gratings involving a large number of sublayers.

Since RCWA describes the response due to a plane wave incident upon a multilayer structure with infinite boundaries along the x direction, power conservation is only fulfilled in the z direction which is perpendicular to the infinite boundaries, while power flow in the x direction is not involved in the power conservation. There is no direct comparison between the incident light power and the guided power in the waveguide. This poses a difficulty in determining the amount of power coupled into the guided mode in the x direction, and the in-coupling efficiency can't be directly calculated. As a result, we need to

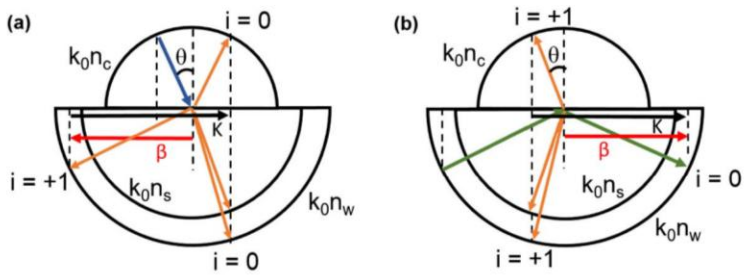


Figure 2.2: Phase diagram of (a) grating in-coupling and (b) out-diffraction process with only the $i = 0$ and $i = +1$ orders. The light in-coupling into the $i = +1$ order in (a) is the reciprocal process of the guided wave out-diffraction into the $i = +1$ order in (b).

take advantage of the light reciprocity and use the out-diffraction process to indirectly calculate the in-coupling diffraction efficiency. For example, the in-coupling efficiency to the $i = +1$ order into the waveguide in Fig. 2.2(a) is the same as the out-diffraction efficiency to $i = +1$ order into the cover in Fig. 2.2(b).

Previous work, RCWA-LW approach [102, 118, 119], attempted to solve the out-diffraction efficiency by removing the incident light contribution (the terms involving δ_{i0} in Eq. (2.23)) and casting the problem into a homogeneous system of equations in the form

$$\begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{W}_g & \mathbf{W}_g\mathbf{X}_g & \mathbf{0} & \mathbf{0} \\ j\tilde{\mathbf{Y}}_c & \mathbf{0} & \mathbf{V}_g & -\mathbf{V}_g\mathbf{X}_g & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_g\mathbf{X}_g & \mathbf{W}_g & -\mathbf{I} & -\mathbf{X}_w \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_g\mathbf{X}_g & -\mathbf{V}_g & -j\tilde{\mathbf{Y}}_w & j\tilde{\mathbf{Y}}_w\mathbf{X}_w \\ \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{X}_w & \mathbf{I} \\ \mathbf{0} & -\tilde{\mathbf{Y}}_{s/w} & \mathbf{0} & \mathbf{0} & \mathbf{X}_w & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ \mathbf{T} \\ \mathbf{C}_g^+ \\ \mathbf{C}_g^- \\ \mathbf{C}_w^+ \\ \mathbf{C}_w^- \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{2.24}$$

which can be represented by $\tilde{\mathbf{M}}\mathbf{x} = \mathbf{0}$. The problem becomes finding a complex propagation constant $\tilde{\gamma} = \beta - j\alpha$ with unknown positive real number β and α such that the determinant of the boundary condition matrix $\tilde{\mathbf{M}}$ is minimized (close to zero). As a complex propagation constant is involved in the calculation, Eq. (2.7) is modified to

$$\tilde{k}_{x,i} = \tilde{\gamma} - iK = (\beta - iK) - j\alpha, \tag{2.25}$$

and Eq. (2.8) becomes [120]

$$\tilde{k}_{rz,i} = \begin{cases} \sqrt{{n_r}^2 {k_0}^2 - \tilde{k}_{x,i}^2} & \mathrm{Re}(\tilde{k}_{rz,i}) > \mathrm{Im}(\tilde{k}_{rz,i}) \\ -\sqrt{{n_r}^2 {k_0}^2 - \tilde{k}_{x,i}^2} & \mathrm{Re}(\tilde{k}_{rz,i}) < \mathrm{Im}(\tilde{k}_{rz,i}) \end{cases} \quad r = c, s. \tag{2.26}$$

Then, $\tilde{\mathbf{Y}}_c$,$\tilde{\mathbf{Y}}_w$, and $\tilde{\mathbf{Y}}_{s/w}$ are defined as $\tilde{k}_{cz,i}/k_0$, $\tilde{k}_{wz,i}/k_0$ and $\tilde{k}_{sz,i}/k_{wz,i}$, respectively. The application of this method is limited to grating structures with small index modulations ($\Delta n \sim 0.01 - 0.5$, e.g. polymer-based or glass-based volume gratings); however, the algorithm is difficult to find a $\tilde{\gamma}$ within reasonable bounds when the index difference is large ($\Delta n > 1$, e.g. SOI gratings), even though Δn is successively increased from small values to the desired value and the converged results obtained from the Muller method are used as the initial guesses for the next steps.

Therefore, an alternative method has to be applied to find the complex propagation constant $\tilde{\gamma}$ for gratings with large index contrasts in order to calculate the out-diffraction efficiencies of the waveguide gratings, which leads to the introduction of RCWA-EIS method. In RCWA-EIS method, the grating is designed such that only $i = 0$ and $i = +1$ orders are propagating whereas the other orders are all evanescent, which ensures the minimum number of propagating orders and thus maximizes diffraction efficiencies. This assumption is valid in grating optimization where high efficiency per order is needed, and it also reduces the dimension of matrices involved in subsequent steps and minimizes numerical instability. The $i = +1$ diffracted order is coupled into the waveguide, as shown in Fig. 2.2. All fields in the structure, as well as the propagation constant β, can be determined from the in-coupling process analyzed by the conventional RCWA. By reciprocity, in-coupled light and out-diffracted light should have the same real propagation constant β in the x direction. The problem becomes finding the radiation factor α involved in Eq. (2.25) that will be used to calculate the diffraction efficiencies.

2.1.2 Step 2: Equivalent Index Slab Definition

The grating layer, whose field expansion involves a sum of exponential terms, is replaced by L layers of uniform equivalent slabs with unknown refractive indices $\tilde{n}_l$ and the electric field expansion of the lth slab in the z direction is

$$\begin{aligned} E_{ly,i}(z) =& C_i^{l+} \exp\left\{-j\tilde{k}_{lz,i}[z-(l-1)t_l]\right\} \\ &+ C_i^{l-} \exp\left[+j\tilde{k}_{lz,i}(z-lt_l)\right], \end{aligned} \tag{2.27}$$

where $t_l = t_g/L$, t_g is the grating height, and the parameter L is chosen to be 4 which will be explained later. With only two diffraction orders (0 and $+1$) involved, the only requirement for the definition of the equivalent slabs is matching the field amplitudes and phases of these two orders at the cover-grating interface and the grating-waveguide interface. Note that the field coefficients R_0, R_1, C_0^{w+}, C_0^{w-}, C_1^{w+}, and C_1^{w-} are all obtained from the conventional RCWA in Step 1. As shown in Fig. 2.3, the equivalent slabs generate the same fields outside grating region, even though the field distributions inside the grating region may vary. Retaining the field profiles in the cover and substrate is important because the radiation losses are mainly due to the radiation in the cover and substrate. Similar to a "black box", the complicated field expansions inside of the "box" are replaced by simple expressions while the outside fields remain unchanged. The EIS concept is similar to the method introduced in [121] in which the grating layer is replaced by a homogeneous dielectric slab with pre-defined index and then simulated by the transmission line approach. But here, the equivalent slab indices are varied based on the fields outside the gratings. Note that the contribution of the evanescent orders still exist in defining the equivalent index slabs because the field coefficients used in the EIS definition are determined by the conventional RCWA of the in-coupling process which considers all possible orders (s truncated orders implemented in the computer).

RCWA-EIS approach works well when only two propagating diffraction orders are

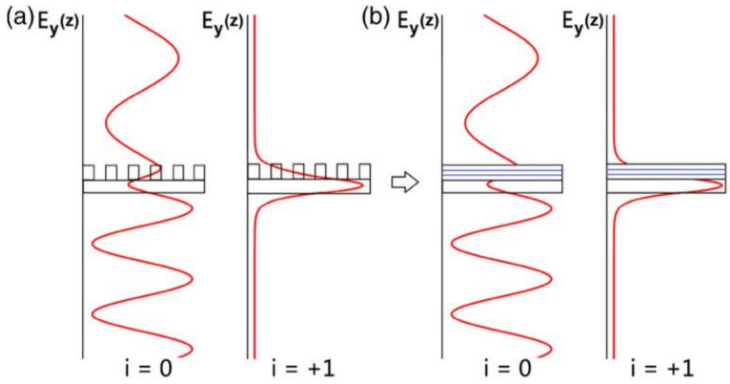


Figure 2.3: Schematic representation of field repetition outside of the grating by equivalent index slabs. (a) Electric field amplitude along the z direction of the multilayer grating structure, and (b) equivalent index slabs are used to replace the grating layer.

considered. If more diffraction orders are involved, a larger boundary condition matrix in the form of Eq. (2.24) has to be constructed and the formulation is then no different from RCWA-LW approach. In that case, there is no point in finding equivalent index slabs to replace the grating. In other words, RCWA-EIS approach is a simplified version of RCWA-LW approach relying on multiple equivalent index slabs to represent the grating and it can only efficiently treat a small number of propagating diffraction orders.

Specifically, by imposing boundary conditions on the electric and magnetic fields, the following transfer matrix formulation can be obtained for the $i = 0$ order:

$$\begin{bmatrix} 1 \\ -k_0 n_c \cos\theta \end{bmatrix} + \begin{bmatrix} 1 \\ k_{cz,0} \end{bmatrix} R_0 = \prod_{l=1}^{L} \begin{bmatrix} 1 & X_{l,0} \\ -\tilde{k}_{lz,0} & \tilde{k}_{lz,0} X_{l,0} \end{bmatrix} \begin{bmatrix} X_{l,0} & 1 \\ -\tilde{k}_{lz,0} X_{l,0} & \tilde{k}_{lz,0} \end{bmatrix}^{-1} \begin{bmatrix} 1 & X_{w,0} \\ -k_{wz,0} & k_{wz,0} X_{w,0} \end{bmatrix} \begin{bmatrix} C_0^{w+} \\ C_0^{w-} \end{bmatrix} = \begin{bmatrix} P_{0,11} & P_{0,12} \\ P_{0,21} & P_{0,22} \end{bmatrix} \begin{bmatrix} C_0^{w+} \\ C_0^{w-} \end{bmatrix}, \tag{2.28}$$

where $X_{l,0} = \exp(-j\tilde{k}_{lz,0}t_l)$, $X_{w,0} = \exp(-jk_{wz,0}t_l)$, and $\tilde{k}_{lz,0}$ is defined as

$$\tilde{k}_{lz,i} = \begin{cases} \sqrt{\tilde{n}_l^2 {k_0}^2 - {k_{x,i}}^2} & , \mathrm{Re}(\tilde{k}_{lz,i}) > \mathrm{Im}(\tilde{k}_{lz,i}) \\ -\sqrt{\tilde{n}_l^2 {k_0}^2 - {k_{x,i}}^2} & , \mathrm{Re}(\tilde{k}_{lz,i}) < \mathrm{Im}(\tilde{k}_{lz,i}) \end{cases} \tag{2.29}$$

because the unknown equivalent slab indices $\tilde{n}_l$ are complex numbers with no physical meanings (pure mathematical representation). Equation (2.28) can be transformed into

$$P_{0,11}C_0^{w+} + P_{0,12}C_0^{w-} - 1 - R_0 = 0, \tag{2.30}$$

$$P_{0,21}C_0^{w+} + P_{0,22}C_0^{w-} - k_{cz,0}R_0 + k_0 n_c \cos\theta = 0. \tag{2.31}$$

The transfer matrix formulation for the $i = +1$ order has a similar form as Eq. (2.28) but excludes the incident light contribution (1 and $k_0 n_c \cos\theta$), and it can be transformed into

$$P_{1,11}C_1^{w+} + P_{1,12}C_1^{w-} - R_1 = 0, \tag{2.32}$$

$$P_{1,21}C_1^{w+} + P_{1,22}C_1^{w-} - k_{cz,1}R_1 = 0, \tag{2.33}$$

where $k_{cz,0}$, $k_{wz,0}$, R_0, R_1, C_0^{w+}, C_0^{w-}, C_1^{w+} and C_1^{w-} are known from the in-coupling calculation.

The problem then becomes finding L equivalent refractive indices such that Eq. (2.30) to (2.33) are satisfied simultaneously. A total of 4 equations related to the E fields and H fields of each of the $i = 0$ order and $i = +1$ order are involved. The Matlab function *fsolve* with the Trust-Region-Dogleg (TRD) algorithm are used to find the equivalent indices [43]. The TRD algorithm is specially designed to solve nonlinear equations, and it requires the number of equations be the same as the number of unknowns, which gives a unique set of solutions. This is the reason why the number of equivalent index slabs L is chosen to be 4. The TRD algorithm usually gives small function values, e.g. less than 1×10^{-8}. The resulting 4 equivalent indices are verified further by calculating the propagation constant β_{eq}

in the multilayer structure consisting of the cover, the 4 equivalent index slabs, the original waveguide, and the substrate, using the formulation reported in [122]. The calculated propagation constant β_{eq} is then compared with $k_{x,1}$ (the propagation constant of the coupled order along the x direction).

2.1.3 Step 3: Radiation Factor Calculation

After the equivalent indices are determined, the radiation factor α is calculated using a similar process as discussed in [122]. But here, we will consider two out-diffracted orders with $\tilde{k}_{x,0} = \beta - j\alpha$ and $\tilde{k}_{x,1} = K - \beta - j\alpha$ as propagation constants in the x direction. The sign of the real part of $\tilde{k}_{x,1}$ is not of consequence since $\tilde{k}_{x,1}$ is only used to calculate $\tilde{k}_{cz,1}$, $\tilde{k}_{sz,1}$ and $\tilde{k}_{wz,1}$ based on Eq. (2.26). All layers, including cover, equivalent index slabs, waveguide and substrate, are considered in the transfer matrix formulations. Specifically, the transfer matrix formulations for both orders are in the form of Eq. (2.28) (without incident light) with the replacement

$$\begin{bmatrix} C_{0/1}^{w+} \\ C_{0/1}^{w-} \end{bmatrix} = \begin{bmatrix} X_{w,0/1} & 1 \\ -\tilde{k}_{wz,0/1} X_{w,0/1} & \tilde{k}_{wz,0/1} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -\tilde{k}_{sz,0/1} \end{bmatrix} T_{0/1}, \tag{2.34}$$

and can be written in a simplified form as

$$\begin{bmatrix} 1 \\ \tilde{k}_{cz,0/1} \end{bmatrix} R_{0/1} = \begin{bmatrix} Q_{0/1,1} \\ Q_{0/1,2} \end{bmatrix} T_{0/1}, \tag{2.35}$$

which results in $R_{0/1} = Q_{0/1,1} T_{0/1}$, where $0/1$ indicates 0 or $+1$ order. Therefore, we need to find a radiation factor α such that $R_0 - Q_0 T_0$ and $R_1 - Q_1 T_1$ are simultaneously close to zero for the existence of an out-diffracted order. Since this step involves two equations (Eq. 2.35) and one unknown (α), the resulting least-squares problem is solved using *fsolve* with the Levenberg-Marquardt (LM) algorithm. The LM algorithm minimizes the sum of

squares of the functions, and the nonlinear system it deals with could be underdetermined, critical-sized, or overdetermined; that is, the number of unknowns could be greater than, equal to, or less than the number of equations. However, the lower and upper bound of α cannot be defined in *fsolve* and this may lead to nonphysically small or negative values of α. In many cases, these nonphysical values may be avoided by slightly changing the values ($\pm 10^{-4}$) of input parameters, e.g. coupling angle, grating period, grating height, etc. Note that the field amplitudes inside the grating are assumed to be the same for the in-coupling and out-diffracted situations, similar to the assumption in the RCWA-LW approach [118, 119, 123].

The matrix dimension treated here (2×2) is much smaller than that used in RCWA-LW approach ($6s\times6s$), where s is the total number of diffraction orders, which avoids the problematic step of solving the determinant in RCWA-LW approach. Step 2 and 3, which manifest the EIS concept, eliminate the less important evanescent orders considered in RCWA-LW approach when calculating the radiation factor, thus greatly improving the calculation efficiency and numerical stability.

2.1.4 Step 4: Out-Diffraction Efficiency Calculation

After α is determined, the out-diffraction efficiency, equivalently in-coupling efficiency, can be calculated by substituting the complex propagation constant $\tilde{\gamma} = \beta - j\alpha$ into Eq. (2.24), and conducting the singular value decomposition of the matrix $\tilde{\mathbf{M}}$, which is $\tilde{\mathbf{M}} = \mathbf{USV}^T$. The solution to the homogeneous system $\tilde{\mathbf{M}}\mathbf{x} = \mathbf{0}$ is the the column vector of $\mathbf{V}$ corresponding to the smallest singular value. The values of R_i and T_i can be calculated correspondingly. The out-diffraction efficiency can be determined by first calculating the power flow in the z direction. The energy flux density is represented by the time averaged Poynting vector $\bar{S} = \frac{1}{2}\,\mathrm{Re}(\bar{E} \times \bar{H}^*)$, which can be reduced to $S_z = -\frac{1}{2}\,\mathrm{Re}(E_y H_x^*)$ for power flow in the z direction. The reflected power density and transmitted power density

are calculated as:

$$S^r_{z,i} = \frac{1}{2} R_i R_i^* \frac{\mathrm{Re}(\tilde{k}_{cz,i})}{\omega\mu_0}, \tag{2.36}$$

$$S^t_{z,i} = \frac{1}{2} T_i T_i^* \frac{\mathrm{Re}(\tilde{k}_{sz,i})}{\omega\mu_0}, \tag{2.37}$$

for $\mathrm{Re}(\tilde{k}_{cz,i}) < k_0 n_c$ and $\mathrm{Re}(\tilde{k}_{sz,i}) < k_0 n_s$, respectively. Otherwise, set $S^r_{z,i}$ and $S^t_{z,i}$ to zero. The preferential coupling ratios $PC_{c,i}$ and $PC_{s,i}$ of the ith order are defined as

$$PC_{c,i} = \frac{S^r_{z,i}}{\sum_i (S^r_{z,i} + S^t_{z,i})}, \tag{2.38}$$

$$PC_{s,i} = \frac{S^t_{z,i}}{\sum_i (S^r_{z,i} + S^t_{z,i})}. \tag{2.39}$$

Finally, the diffraction efficiencies $DE_{c,i}$ and $DE_{s,i}$ of the ith order at a given grating length ℓ are estimated by an exponential decaying distribution as

$$DE_{c,i} = PC_{c,i}[1 - \exp(-2\alpha\ell)], \tag{2.40}$$

$$DE_{s,i} = PC_{s,i}[1 - \exp(-2\alpha\ell)]. \tag{2.41}$$

2.2 Efficiency Optimization

2.2.1 Optimization Procedures

The grating coupler is optimized using the Matlab optimization solver *fmincon* together with RCWA-EIS method. RCWA-EIS method determines the in-coupling/out-diffraction efficiencies DE of the grating coupler in a specific configuration (e.g. given grating profile, diffraction angle θ, grating height t_g, grating period Λ, fill factor f, etc.), and the solver *fmincon* minimizes the target function $(1 - DE)$ by varying the grating parameters according to a set of constraints. The $i = +1$ order out-diffraction efficiency $DE_{c,1}$, equivalently the $i = +1$ order in-coupling efficiency (Fig. 2.2), determined by RCWA-EIS method is

the target to optimize in the solver *fmincon*. Since the grating is designed to couple the $i = +1$ order diffracted wave into the waveguide, $k_{x,1}$ should be comparable to the fundamental propagation constant of the slab waveguide β_0, which can be calculated using the method introduced in [122]. In the optimization, $k_{x,1}$ is restricted to be in the range $|\beta_0 - k_{x,1}| < k_0 n_w - \beta_0$. The inputs to the optimization solver *fmincon* are as follows: (1) the initial values of a set of undecided parameters, e.g. incident angle (or coupling angle) θ, fill factor f, grating period Λ, grating thickness t_g, etc., (2) the lower and upper bounds of each parameter, (3) the function tolerance, and (4) the constraint $|\beta_0 - k_{x,1}| < k_0 n_w - \beta_0$. The model outputs the final values of those parameters that give the minimum value of $1 - DE_{c,1}$, which corresponds to maximizing $DE_{c,1}$. The optimized structure can be validated by 2D FDTD simulation (e.g. MEEP program developed at Massachusetts Institute of Technology). Multiple runs with different sets of initial values can be conducted in order for the solver to find the global minimum.

2.2.2 Results and Discussion

The interlayer grating coupling problem is schematically depicted in Fig. 2.4. Surface-relief gratings (binary, parallelogramic, and sawtooth gratings) as well as volume gratings with sinusoidally varying refractive indices are considered here. Some basic parameters involved in the models are as follows: free-space wavelength of incident light λ_0, coupling angle in the cover θ, grating thickness t_g, waveguide thickness t_w, grating fill factor f, grating period Λ, number of grating period N, and grating slant angle φ (for parallelogramic and volume gratings). TE$_z$ (E_y, H_z and H_x) polarized planar incidence is considered in all the models. The formulations of the grating models follow the previous RCWA work [101, 124, 115, 41]. The permittivity formulations of parallelogramic, sawtooth and volume gratings are introduced in the Appendix A.

For the analysis of surface-relief gratings, three grating parameters, namely coupling angle (θ), grating period (Λ), and grating thickness (t_g), are variables to be optimized. The

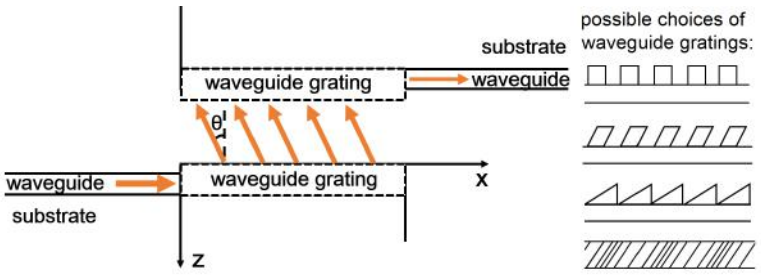


Figure 2.4: Schematic representation (not to scale) of the grating-to-grating coupling process. The waveguide grating region indicated by the dashed box represents one of the grating structures at the right side of the figure.

known parameters are as follows: free-space wavelength $\lambda_0 = 1.55$ μm, cover refractive index $n_c = 1$, substrate refractive index $n_s = 1.45$, grating groove refractive index $n_{gr} = 1$, grating ridge refractive index $n_{rd} = 2.46$, waveguide refractive index $n_w = 3.45$, grating fill factor $f = 0.5$ (for binary and parallelogramic gratings), and waveguide thickness $t_w =$ 0.22 μm. The fundamental mode propagation constant of the 0.22 μm thick air/Si/SiO_2 slab waveguide is calculated to be $\beta_0 = 11.3710$ μm^{-1}. The total number of space harmonics is set to be $s = 7$.

For the analysis of volume gratings, three grating parameters, namely coupling angle (θ), grating period (Λ), and slant angle (φ), are variables to be optimized. The known parameters are as follows: free-space wavelength $\lambda_0 = 1.55$ μm, cover refractive index $n_c = 1$, substrate refractive index $n_s = 1.45$, waveguide refractive index $n_w = 1.8$, average grating refractive index $n_g = 1.8$, grating refractive index modulation $\Delta n_g = 0.1$ ($\Delta\epsilon \approx 2n_g\Delta n_g$) , waveguide thickness $t_w = 0.4$ μm, and grating thickness $t_g = 0.4$ μm. The fundamental mode propagation constant of the 0.4 μm thick waveguide is calculated to be $\beta_0 = 6.3008$ μm^{-1}. The total number of space harmonics is again set to be $s = 7$.

The optimization of the grating parameters is carried out using the Matlab function *fmincon*. The grating is designed to couple the $i = +1$ order diffracted light into/out of the waveguide, and thus the target of optimization is the diffraction efficiency of the $i = +1$ order ($DE_{c,1}$), and $k_{x,1}$ should be comparable to β_0. Optimizing the single grating diffraction efficiency $DE_{c,1}$ is the same as optimizing the grating-to-grating coupling efficiency which

is approximated by $DE_{c,1}^2$. To ensure the maximum diffraction efficiency of the $i = +1$ order, diffraction orders other than $i = 0$ and $i = +1$ should not propagate. Furthermore, to ensure the correctness of the 4 equivalent indices, the propagation constant calculated using the equivalent indices, β_{eq} (taking the real part), should be comparable to $k_{x,1}$, and the difference between these two values is restricted to be less than 2%. As a result, the nonlinear constraints in the *fmincon* function is thus set as follows: (1) $|\beta_0 - k_{x,1}| < k_0 n_w - \beta_0$, (2) $k_{x,-1} > k_0 n_w$, (3) $|k_{x,1} - \beta_{eq}|/|k_{x,1}| < 0.02$, and (4) $\alpha > 0$. The lower and upper bounds of the three variables [θ, Λ, t_g] are set to be [0.1 rad (5.73°), 0.3 μm, 0.05 μm] and [$\pi/4$ rad (45°), 1.55 μm, 0.4 μm], respectively. Taking into account fabrication limits, the thickness of the grating layer should be less than 0.4 μm [125], and the minimum coupling angle should be set to 0.1 rad (5.73°) to reduce the possibility of coupling into the $i = -1$ diffraction order. For the volume grating, the lower and upper bounds of [θ, Λ, φ] are set to be [0.1 rad (5.73°), 0.2 μm, 0.3 rad (17.19°)] and [0.4 rad (22.91°), 1 μm, 1.3 rad (74.48°)], respectively.

For all the grating structures presented in this paper, the Matlab function *fsolve* with the TRD algorithm is used to find the equivalent indices. The function tolerance of *fsolve* is set to be 1×10^{-8}. The Matlab function *fsolve* with the LM algorithm is used to find the radiation factor of the multilayer structure, and its function tolerance is set to be 1×10^{-8}. The optimization is carried out using the Matlab function *fmincon* with nonlinear constraints specified above. Parameter sweep of initial values, e.g. [θ, Λ, t_g] for the surface-relief gratings and [θ, Λ, φ] for the volume gratings, can be conducted to avoid finding of unreasonable values of the radiation factor α. Normally, α for parallelogramic gratings, sawtooth gratings, and volume gratings are in the range of $[0.01,\ 0.1]$, $[0.001,\ 0.02]$, and $[0.001,\ 0.02]$, respectively. In general, the search algorithms work well for grating structures with relatively large α, e.g. parallelogramic gratings and binary gratings; for grating structures with relatively small α, e.g. sawtooth gratings and volume gratings, parameter sweep of initial values is highly recommended.

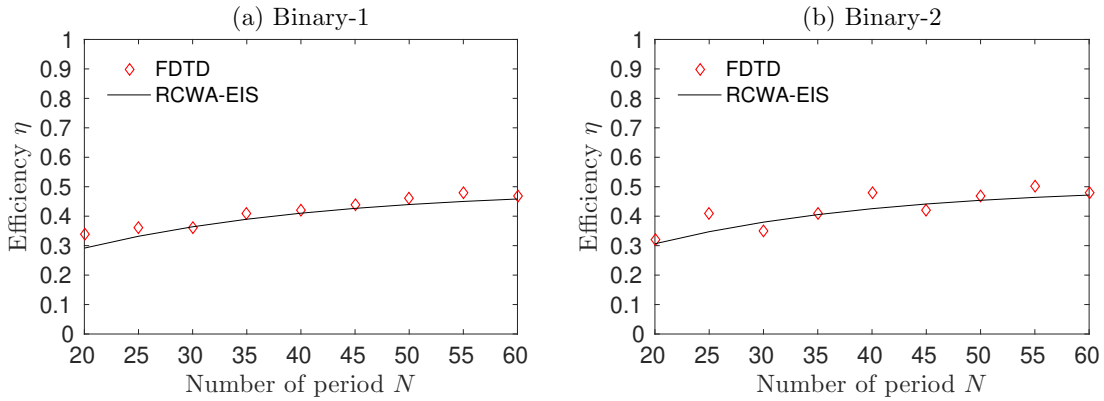


Figure 2.5: Single grating diffraction efficiency ($DE_{c,1}$) as a function of number of periods (N) or grating length ($\ell = N\Lambda$) for binary gratings (binary-1 and binary-2) summarized in Table 2.1.

Binary Grating Coupler

Two examples of optimized binary gratings are summarized in Table 2.1. Each optimization starts with a set of initial values. If the initial values are far from the ideal values corresponding to the optimized structure (relatively large diffraction efficiencies), the optimization function will still find the final values that are close to the ideal values, even though the diffraction efficiencies may not be maximized or accurate. Those final values can serve as rough estimates of the ideal values. If the initial values are in the vicinity of the ideal values, the diffraction efficiencies found by the optimization function are maximized and relatively accurate. More examples are shown in [41]. Multiple optimizations should be carried out to ensure finding the global maximum, which is still computationally efficient due to the fast calculation of each optimization.

From Fig. 2.5, it is observed that the optimized binary gratings have diffraction efficiencies approaching 50%, which is limited by the preferential coupling ratio PC_c. Since binary gratings have symmetric profiles and roughly equal indices of the waveguide claddings (1 for air and 1.45 for SiO_2), the guided power is diffracted approximately equally into the cover and substrate. The curves corresponding to the FDTD results oscillate at small numbers of grating periods. This may due to two reasons: (1) at small grating lengths, the fields in the structures have not yet reached steady state, and thus the contribution of

Table 2.1: Optimized parameters and calculated in-coupling/out-diffraction efficiencies for selected cases of binary gratings, parallelogramic gratings, sawtooth gratings, and the volume gratings. $PC_{c,1}$ is the preferential coupling ratio, defined as the fraction of the total radiation power that is diffracted into the $i = +1$ order in the cover.

Case No.	Initial Values				Optimized Values						
	θ [rad]	Λ [μm]	t_g [μm]	ϕ [rad]	θ [rad]	Λ [μm]	t_g [μm]	ϕ [rad]	$PC_{c,1}$	β [μm^{-1}]	α [μm^{-1}]
binary-1	0.30	0.60	0.20	–	0.3001	0.6000	0.2340	–	0.4914	9.2736	0.0375
binary-2	0.10	0.55	0.30	–	0.2801	0.5311	0.2432	–	0.5011	10.7099	0.0445
f-pa-1	0.30	0.55	0.30	–	0.3005	0.5373	0.3124	0.9942	0.8175	10.4950	0.0279
f-pa-2	0.10	0.52	0.39	–	0.1956	0.5589	0.3087	0.9757	0.7923	10.4542	0.0238
b-pa-1	0.10	0.53	0.34	–	0.2242	0.5407	0.3095	0.9911	0.8038	10.7192	0.0139
b-pa-2	0.11	0.48	0.12	–	0.1507	0.4912	0.1575	1.0418	0.8006	12.1829	0.0122
f-saw-1	0.10	0.47	0.37	–	0.2345	0.4793	0.3667	–	0.7931	12.1672	0.0096
f-saw-2	0.10	0.52	0.31	–	0.2057	0.5434	0.2872	–	0.8033	10.7348	0.0010
b-saw-1	0.10	0.60	0.1	–	0.2317	0.5356	0.2079	–	0.5840	10.7997	0.0097
b-saw-2	0.25	0.50	0.25	–	0.2376	0.5088	0.2409	–	0.4864	11.3951	0.0099
vol-1	0.30	0.70	–	1.2	0.1041	0.7019	–	0.9255	0.5968	6.7298	0.0098
vol-2	0.30	0.70	–	1.1	0.2933	0.6974	–	0.9630	0.6274	6.2235	0.0079

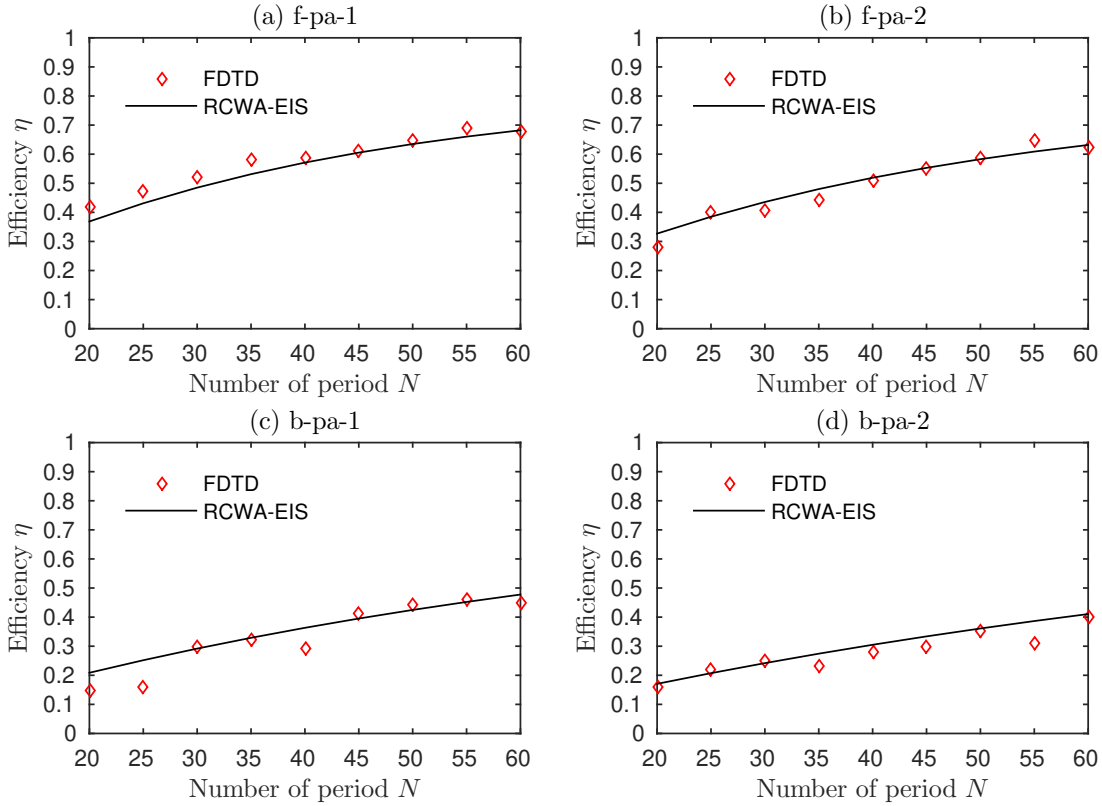


Figure 2.6: Single grating diffraction efficiency ($DE_{c,1}$) as a function of number of periods (N) or grating length ($\ell = N\Lambda$) for parallelogramic gratings, where Λ is given in Table 2.1.

scattering at the waveguide ends becomes much more significant; (2) there is a stability issue in the FDTD calculation when a transient source, e.g. a Gaussian source, is used, but as time increases, the nonphysical transients will decay. On the other hand, neither the scattering effect nor the transient field is considered in the RCWA, and RCWA-EIS method generates smooth curves as a result of the exponential mathematical model Eq. (2.40).

Parallelogramic Grating Coupler

According to Li *et al.* [126], the optimized slant angle for a forward-slanted paralellogramic grating (geometry defined in Appendix A) is defined as

$$\varphi = \arcsin\left(\frac{\lambda_0}{\Lambda\sqrt{n_{gr}^2 + f(n_{rd}^2 - n_{gr}^2) - \beta_0^2 + 2\beta_0\lambda_0/\Lambda}}\right). \tag{2.42}$$

Therefore, the slant angle φ is determined once the grating period Λ is found for a given material system.

Four optimized paralellogramic gratings, two forward-slanted cases (case f-pa-1 and case f-pa-2) and two backward-slanted cases (case b-pa-1 and case b-pa-2), are summarized in Table 2.1 which includes the optimized parameters θ, Λ, and φ as well as the resulting parameters such as the preferential coupling ratio in the $i = +1$ order ($PC_{c,1}$), coupled mode propagation constant ($\beta = k_{x,1}$) and the radiation factor (α). The single grating diffraction efficiencies ($DE_{c,1}$) as a function of number of grating periods (N) for the parallelogramic gratings are shown in Fig. 2.6. The results obtained from RCWA-EIS method demonstrate good agreement with those calculated by FDTD simulations. Some of the data plots corresponding to the FDTD calculations exhibit irregularity due to the numerical instability induced by the slanted grating structure (staircase approximation of the FDTD grids). The error may be reduced by increasing the resolution of the FDTD simulation, but it requires more computation time. It is observed that the optimized forward-slanted gratings have much larger diffraction efficiencies than the optimized backward-slanted gratings, which agrees with the conclusion that forward-slanted parallelogramic gratings give higher diffraction efficiencies as stated in Li *et al.* [126]. It is also observed that the preferential coupling ratio ($PC_{c,1}$) of the forward-slanted parallelogramic gratings is much larger than 0.5, which is a good demonstration of the benefit of using asymmetric grating profiles. It is known that gratings with symmetric profiles, e.g. sinusoidal and rectangular, radiate the incident power (or guided power) almost equally into the cover and the substrate, resulting in a preferential coupling ratio (or radiation directionality) close to 50% provided the refractive index of the cover is comparable to that of the substrate, as shown in the binary grating section. On the contrary, gratings with asymmetric profiles, e.g. blazed (trapezoidal, sawtooth or triangular) and parallelogramic, emit more than 90% of the radiated power into the cover [127, 105, 128, 129]. Nevertheless, the radiation factor α, defined as the radiated power per unit grating length, of the blazed grating is very small, and thus a longer

grating is required to radiate the same amount of incident (or guided) power. Researchers have already found that parallelogramic gratings provide both larger radiation factor and higher radiation directionality than gratings with other tooth profiles [130, 126]. Therefore, parallelogramic gratings offer the benefits of simultaneously obtaining high diffraction efficiency and high device compactness. From Table 2.1, it is found that the radiation factors α of the forward-slanted parallelogramic gratings are relatively large compared to those of other grating profiles. In spite of their relatively complex profiles, parallelogramic gratings can be fabricated using the methods such as those presented in [130, 131].

Sawtooth Grating Coupler

Four optimized sawtooth gratings, two forward-slanted cases (case f-saw-1 and case f-saw-2) and two backward-slanted cases (case b-saw-1 and case b-saw-2), are summarized in Table 2.1. The single grating diffraction efficiencies ($DE_{c,1}$) as a function of number of grating periods (N) for the sawtooth gratings are shown in Fig. 2.7. The results obtained from RCWA-EIS method demonstrate good agreement with those from FDTD calculations. The optimized forward-slanted gratings have larger diffraction efficiencies than the optimized backward-slanted gratings, though the improvement of the forward-slanted profile is not as significant as in the case of parallelogramic gratings. The preferential coupling ratios of the forward-slanted sawtooth gratings are also much larger than 50%. Aoyagi *et al.* [132] reported that the blazed grating can direct 97% of the total radiated power into the desired angle. Although the preferential coupling ratios of sawtooth gratings shown in Table 2.1 are approximately 80%, they can be optimized if the target function is $PC_{c,1}$ rather than $DE_{c,1}$. Nevertheless, the resulting diffraction efficiencies are small due to the small radiation factor common in blazed gratings.

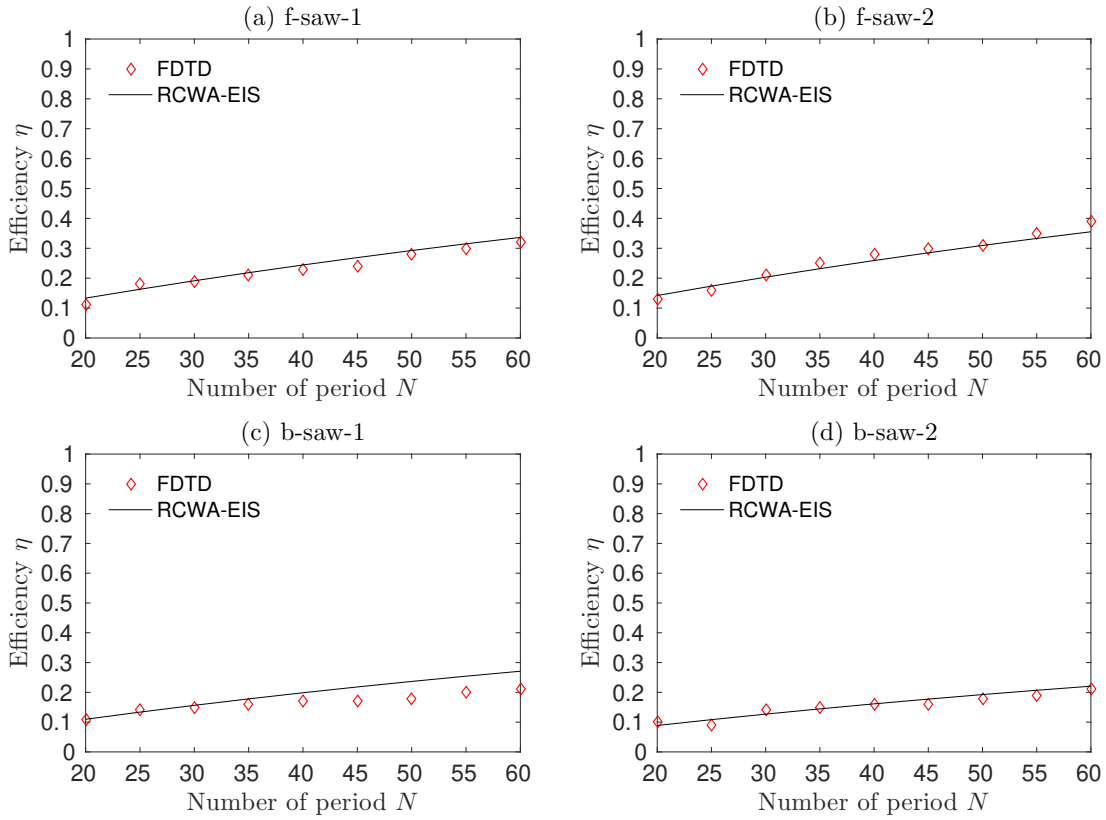


Figure 2.7: Single grating diffraction efficiency ($DE_{c,1}$) as a function of number of periods (N) or grating length ($\ell = N\Lambda$) for sawtooth gratings, where Λ is given in Table 2.1.

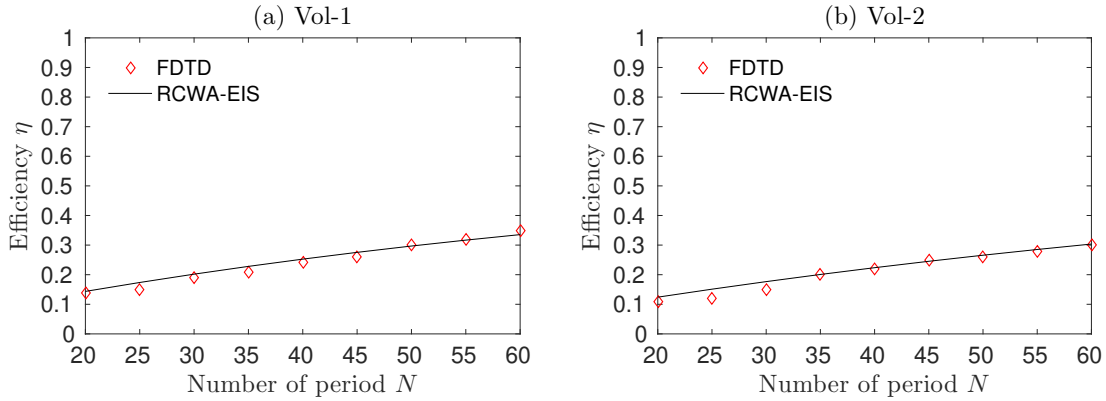


Figure 2.8: Single grating diffraction efficiency ($DE_{c,1}$) as a function of number of periods (N) or grating length ($\ell = N\Lambda$) for volume gratings, where Λ is given in Table 2.1.

Volume Grating Coupler

The volume grating is taken to be a section of the waveguide with sinusoidal index modulation; that is, the volume grating is in the waveguide instead of on top of the waveguide as in the previous examples. Two optimized volume gratings are summarized in Table 2.1. The single grating diffraction efficiencies ($DE_{c,1}$) as a function of number of grating periods (N) for the volume gratings are shown in Fig. 2.8. The results obtained from RCWA-EIS method demonstrate good agreement with those calculated by FDTD. As the grating is a part of the waveguide, the guided field is largely affected by the index-varying region. Nevertheless, the diffraction efficiency is relatively small due to the small index modulation Δn_g used in this model. Since the majority of the volume gratings are made of polymers with typical indices from 1.3 to 1.8, and the index modulation Δn_g is usually less than 0.1, the diffraction efficiency of the volume gratings should not be compared with those of surface-relief gratings with large index differences. Here, the results are used to demonstrate the feasibility of RCWA-EIS method in simulating arbitrary volume gratings.

Binary Grating Coupler with Reflector

Surface-relief gratings with bottom reflectors can also be optimized using RCWA-EIS method. Gratings with symmetric profiles have approximately half of the out-diffracted power travelling into the substrate, and thus it is beneficial to recycle the downward-diffracted light using an adjacent-layer reflector. Two types of reflectors are considered, namely grating reflectors and metal reflectors. RCWA-EIS formulation remains the same except for the addition of more layers corresponding to the reflectors. The grating reflector is replaced by equivalent index slabs, while the metal layer is treated as a uniform layer with negative real and positive imaginary relative permittivity, e.g. for gold, ϵ_r = 112.68+j6.852 at 1.55 μm wavelength. The reflectors are added to several optimized binary gratings. To optimize the reflector structure, three parameters, namely the buried oxide layer (BOX)

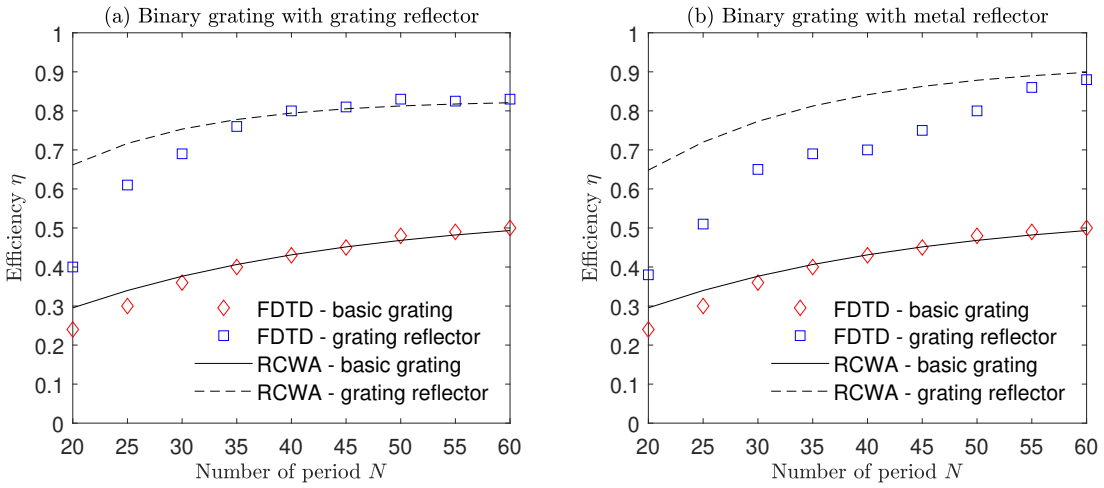


Figure 2.9: In-coupling/out-diffraction efficiencies ($DE_{c,1}$) as a function of number of grating periods (N) for optimized binary gratings with bottom grating reflector and metal reflector. Applying reflector results in twofold increase in diffraction efficiencies.

thickness (t_{box}), grating reflector thickness (t_{bg}), and the horizontal shift of the grating reflector relative to the optimized binary grating (Δ_x), are set as unknowns for the case of grating reflector, whereas two parameters, namely the BOX thickness (t_{box}) and metal layer thickness (t_m), are set as unknowns for the case of metal reflector.

An optimized binary grating with parameters t_g = 0.27 μm, Λ = 0.67 μm, and θ = 0.10 rad (5.73°) is used as the basic structure to which reflectors can be applied. Optimized parameters found for the grating reflector are $\Lambda_{bg} = \Lambda = 0.67$ μm, $t_{bg} = 0.31$ μm, $t_{box} = 0.22$ μm, and $\Delta x = 0.30$ μm, while those for the metal reflector are $t_m = 0.51$ μm, and $t_{box} = 1.02$ μm. Figure 2.9 shows the in-coupling/out-diffraction efficiency ($DE_{c,1}$) as a function of the number of grating periods (N) for the basic grating with/without the reflectors. From RCWA-EIS results, applying the grating reflector results in twofold increase in diffraction efficiencies. Mismatch in FDTD and RCWA-EIS data is due to the geometric issue that RCWA assumes infinite long gratings, while FDTD simulates gratings with real lengths.

In conclusion, RCWA-EIS method together with the Matlab optimization solver *fmincon* can be used to optimize grating couplers with various profiles. The advantages of this optimization method are arbitrary choices of grating profiles, wide parameter search spaces, easy implementation, fast calculation, and accurate results.

2.3 Angular Misalignment Analysis

Alignment of diffraction gratings has been a major bottleneck that limits the interlayer grating coupling efficiency. The need to align the grating couplers and other optical components greatly reduces the possibility of cost-effective manufacturing and automation. Misalignment may result from mechanical and thermal effects which are unavoidable during device handling and operation. For example, thermal warpage may be induced in the package due to strains caused by different coefficients of thermal expansion of package materials, and gratings defined on such package are most likely to be affected. As a result, understanding the underlying mechanism and designing misalignment-tolerant grating couplers are important to the fields of packaging and testing.

There are six degrees of freedom to be considered for misalignment analysis, including three lateral (x, y, z) and three angular (roll, pitch, yaw), where roll, pitch, and yaw are the rotational angles about the x, y, and z axes, respectively. For chip-to-chip grating couplers, the lateral displacements of volume holographic gratings [133, 134] and surface-relief gratings [135, 136] have been experimentally studied. There is limited effort analyzing rotational misalignment, and such efforts all focus on rotations about the y axis, or equivalently, changes in the incident angle [137, 138]. Wu *et al.* [139] analyzed the yaw, pitch, and roll rotation of a volume grating under vertical incidence (zero incident angle), but they only considered transmission through the grating instead of coupling into a waveguide. Other efforts related to misalignment analysis in the field of optical communication mainly focus on fiber-to-grating coupling [57, 140, 141, 142] and laser-to-grating coupling [143]. Misalignment-tolerant structures have been proposed to provide solutions to misalignment problems but those designs did not give general physical insight into the underlying mechanisms [144, 145, 146]. Consequently, it is necessary to build a comprehensive simulation model for interlayer grating coupling under angular misalignment.

This work, for the first time, calculates the interlayer coupling efficiencies of waveg-

uide gratings under arbitrary rotations, using 3D RCWA for conical diffraction [115] together with RCWA-EIS method [41, 43]. The proposed model can handle arbitrary rotations and calculate interlayer grating coupling efficiency under the pre-defined rotations for optimized gratings (with two major propagating diffraction orders). The model can also produce the change of single grating diffraction efficiency (or equivalently, the interlayer grating coupling efficiency) as a function of rotation angle about a particular direction. Implemented in Matlab and providing a simple software scheme, the model is computationally efficient and numerically accurate, compared with the hefty simulation time required for 3D FDTD simulations. For example, it generally requires 3 days to finish one 3D FDTD simulation at a resolution of 60 pixels per unit distance with 20-core parallel computing, and it would take even longer when the resolution is set to a higher value. In comparison, it takes only 0.3 second for RCWA-EIS method to calculate the efficiency of a rotated grating. Given a specific requirement, such as the maximum angular rotation of an interconnect system, the model is capable of optimizing the grating parameters by using the Matlab optimization toolbox.

2.3.1 Theory and Formulation

The general 3D interlayer grating coupling configuration is depicted in Fig. 2.10. A binary rectangular-groove grating consisting of cover, grating, waveguide, and substrate layers is considered as an example. The bottom grating is rotated relative to the top grating. The coordinate axes of the top and bottom waveguide gratings are labeled as (x_t, y_t, z_t) and (x_b, y_b, z_b), respectively. A TE-polarized guided wave (with field components E_y, H_x, H_z) of free-space wavelength λ_0 is launched into the top waveguide along the $-x_t$ direction and out-diffracted from the top grating, and the out-diffracted light is then incident conically onto the bottom grating. Assume the gratings are optimized for high diffraction efficiencies and only two orders ($i = 0$ and $i = +1$) are propagating in the cover. The analysis begins with determining the conical incidence configuration (proper Euler angles) from the

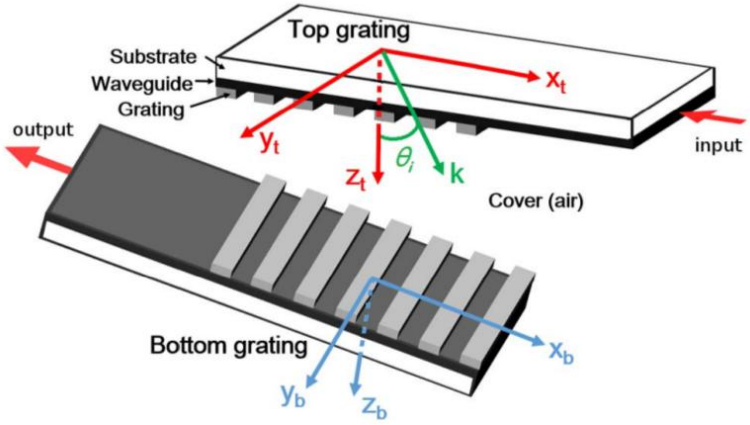


Figure 2.10: Schematic representation (not to scale) of the 3D interlayer grating coupling configuration under angular misalignment.

relative rotation of the bottom grating. Then the conical incidence can be analyzed using the traditional 3D RCWA formulations, and the field amplitudes and phases at all interfaces can be determined. The equivalent index slabs can be subsequently defined and coupling efficiency into the waveguide can be calculated.

Step 1: Determination of Rotated Grating Coordinate System

The first step of the misalignment analysis is to determine the coordinate axes of the bottom grating structure, designated as $\hat{x}_b$, $\hat{y}_b$, and $\hat{z}_b$. The coordinate system of the top grating structure, defined by the axes $\hat{x}_t$, $\hat{y}_t$, and $\hat{z}_t$, is set as the reference coordinate system. Since the rotational misalignment is the target of the analysis, the separation between the top and bottom gratings is not emphasized. For simple calculations, both the top and bottom coordinate systems are placed at a common origin and thus no translation is involved. The bottom grating structure is rotated about an arbitrary axis defined by the unit vector $\hat{a}$. As shown in Fig. 2.11, the rotation of a vector by an angle δ about the axis $\hat{a}$ can be achieved by the following steps:

(1) rotate the coordinate system about the $\hat{z}_t$ axis such that $\hat{a}$ lies in the $x_t z_t$ plane: $R_z(\theta_1)$;

(2) rotate the coordinate system about the $\hat{y}_t$ axis such that $\hat{a}$ lies along the z_t axis: $R_y(\theta_2)$;

(3) rotate the coordinate system by the desired angle δ about the z_t axis: $R_z(\delta)$;

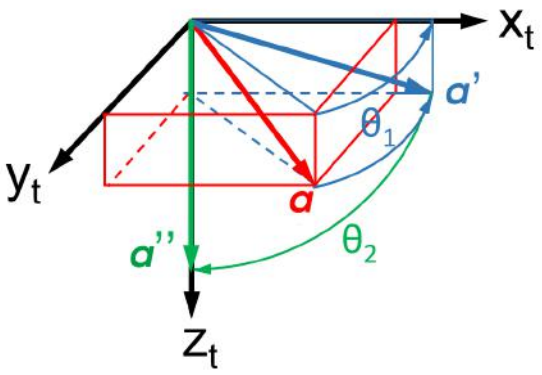


Figure 2.11: Rotation about an arbitrary axis $\hat{a}$ can be decomposed into a series of rotation operations. The first step is rotating the vector a about the z_t axis by an angle θ_1 such that a' is in the $x_t z_t$ plane, and the second step is rotating the vector a' about the y_t axis by an angle θ_2 such that a'' is along the z axis. In this figure, both θ_1 and θ_2 are negative.

(4) perform the inverse of Step (2): $R_y^{-1}(\theta_2) = R_y(-\theta_2)$;

(5) perform the inverse of Step (1): $R_z^{-1}(\theta_1) = R_z(-\theta_1)$,

where the rotation matrix $R_z(\gamma)$ is

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.43}$$

if the coordinate system is rotated clockwise (or equivalently, the vector is rotated counter-clockwise) by an angle γ about the z axis, and the rotation matrix $R_y(\beta)$ is

$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \tag{2.44}$$

if the coordinate system is rotated clockwise by an angle β about the y axis. As a result, the total operation of rotating about the axis $\hat{a}$ is $R_a = R_z^{-1}(\theta_1)R_y^{-1}(\theta_2)R_z(\delta)R_y(\theta_2)R_z(\theta_1)$, and the bottom coordinate system can be determined as $\hat{x}_b = R_a\hat{x}_t$, $\hat{y}_b = R_a\hat{y}_t$, and $\hat{z}_b = R_a\hat{z}_t$.

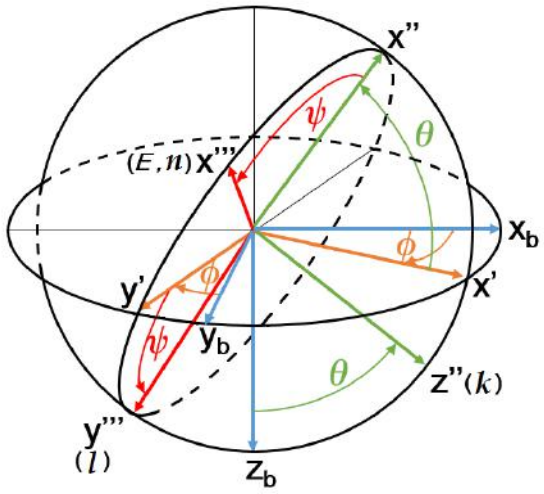


Figure 2.12: Proper Euler angles (ϕ, θ, ψ) for the z-y'-z'' type of intrinsic rotation. The coordinate system first rotates about the z axis (same as z_b and z') by ϕ, then rotates about the y' axis (same as y'') by θ, and finally rotates about the z'' axis (same as k and z''') by ψ.

Step 2: Determination of Proper Euler Angles

The second step of the misalignment analysis is to determine the proper Euler angles (ϕ, θ, ψ), as shown in Fig. 2.12. For a given rotational misalignment configuration, the coordinate axes $(x_t,\ y_t,\ z_t)$ and $(x_b,\ y_b,\ z_b)$ are known. The bottom coordinate axes are expressed in the top coordinate system. From the system $(x_t,\ y_t,\ z_t)$, the system (n, l, k) can be determined as the following unit vectors: $\hat{n} = [0, 1, 0]$, $\hat{l} = [-\cos\theta_{inc}, 0, \sin\theta_{inc}]$, and $\hat{k} = [\sin\theta_{inc}, 0, \cos\theta_{inc}]$, where $\hat{n} = \hat{y}_t = \hat{E}$, $\hat{l} = \hat{k} \times \hat{n}$, and θ_{inc} is the angle between $\hat{k}$ and $\hat{z}_t$ or the incident angle. The direction of the line of nodes, or $\bar{y}'$, is determined by $\bar{y}' = \hat{z}_b \times \hat{k}$. Then ϕ is the angle between $\bar{y}_b$ and $\bar{y}'$, defined as $\phi = \cos^{-1}[\bar{y}_b \cdot \bar{y}'/(|\bar{y}_b||\bar{y}'|)]$; θ is the angle between $\bar{z}_b$ and $\bar{k}$, defined as $\theta = \cos^{-1}[\bar{z}_b \cdot \bar{k}/(|\bar{z}_b||\bar{k}|)]$ (note θ and θ_{inc} are completely different variables; θ_{inc} is the incident angle); and ψ is the angle between $\bar{l}$ and $\bar{y}'$, defined as $\psi = \cos^{-1}[\bar{l} \cdot \bar{y}'/(|\bar{l}||\bar{y}'|)]$. The above definitions only give the values of the angles. The sign of the angle is defined using the right-hand rule: curl the fingers along the rotation direction, and if the thumb points to the positive direction of the rotation axis, the rotation angle is positive; otherwise, the rotation angle is negative. According to this

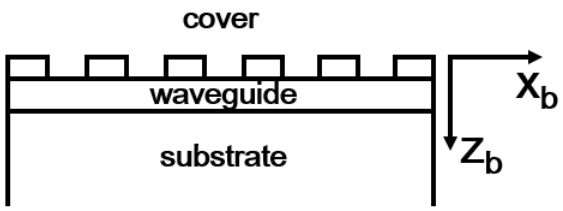


Figure 2.13: Coordinate axes of the bottom grating; $z = 0$ is located at the top surface of the grating.

statement, both θ and ψ are always positive. The sign of ϕ is determined as follows: if the x component of $\hat{y}'$ is larger than the x component of $\hat{y}_b$, ϕ is negative; otherwise, ϕ is positive.

Step 3: Formulations in the system (x, y, z)

After the proper Euler angles (ϕ, θ, ψ) are determined, the conical diffraction problem of the bottom waveguide grating can be formulated according to [115] with some modifications. The coordinate system (x_b, y_b, z_b) is indicated by (x, y, z) for simplicity, and $z = 0$ is set at the top surface of the grating, as indicated in Fig. 2.13. The electric and magnetic fields of the cover, grating, waveguide and substrate layers are expressed as follows:

$\diamond$ Electric field in the cover:

$$E_c = E_{inc} + \sum_i (R_{xi}\,\hat{x} + R_{yi}\,\hat{y} + R_{zi}\,\hat{z}) \exp(-jk_{x,i}x - jk_y y + jk_{cz,i}z), \tag{2.45}$$

where

$$\begin{aligned} E_{inc} =& [(\cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi)\,\hat{x} \\ &+ (\cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi)\,\hat{y} \\ &- \cos\psi\sin\theta\,\hat{z}] \\ &\cdot \exp[-jk_0 n_c(\sin\theta\cos\phi\,x + \sin\theta\sin\phi\,y + \cos\theta\,z)]; \end{aligned} \tag{2.46}$$

◇ Magnetic field in the cover:

$$H_c = H_{inc} + (-j\sqrt{\frac{\epsilon_0}{\mu_0}}) \sum_i (A_{xi}\ \hat{x} + A_{yi}\ \hat{y} + A_{zi}\ \hat{z}) \cdot \exp(-jk_{x,i}x - jk_y y + jk_{cz,i}z), \tag{2.47}$$

where

$$\begin{aligned} H_{inc} = n_c\sqrt{\frac{\epsilon_0}{\mu_0}}[&(-\cos\psi\sin\phi - \sin\psi\cos\phi\cos\theta)\ \hat{x} \\ &+ (\cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta)\ \hat{y} \\ &+ H_{inc,z}\ \hat{z}] \\ \cdot \exp[&-jk_0 n_c(\sin\theta\cos\phi\ x + \sin\theta\sin\phi\ y + \cos\theta\ z)]; \end{aligned} \tag{2.48}$$

◇ Electric field in the grating:

$$E_g = \sum_i [S_{xi}(z)\ \hat{x} + S_{yi}(z)\ \hat{y} + S_{zi}(z)\ \hat{z}] \exp(-jk_{x,i}x - jk_y y); \tag{2.49}$$

◇ Magnetic field in the grating:

$$H_g = -j\sqrt{\frac{\epsilon_0}{\mu_0}} \sum_i [U_{xi}(z)\ \hat{x} + U_{yi}(z)\ \hat{y} + U_{zi}(z)\ \hat{z}] \cdot \exp(-jk_{x,i}x - jk_y y); \tag{2.50}$$

◇ Electric field in the waveguide:

$$E_w = \sum_i [P_{xi}(z)\ \hat{x} + P_{yi}(z)\ \hat{y} + P_{zi}(z)\ \hat{z}] \exp(-jk_{x,i}x - jk_y y); \tag{2.51}$$

◇ Magnetic field in the waveguide:

$$\begin{aligned} H_w =&(-j\sqrt{\frac{\epsilon_0}{\mu_0}})\sum_i[Q_{xi}(z)\ \hat{x} + Q_{yi}(z)\ \hat{y} + Q_{zi}(z)\ \hat{z}] \\ &\cdot \exp(-jk_{x,i}x - jk_y y); \end{aligned} \tag{2.52}$$

◇ Electric field in the substrate:

$$\begin{aligned} E_s =&\sum_i(T_{xi}\ \hat{x} + T_{yi}\ \hat{y} + T_{zi}\ \hat{z}) \\ &\cdot \exp[-jk_{x,i}x - jk_y y - jk_{sz,i}(z - t_g - t_w)]; \end{aligned} \tag{2.53}$$

◇ Magnetic field in the substrate:

$$\begin{aligned} H_s =&(-j\sqrt{\frac{\epsilon_0}{\mu_0}})\sum_i(B_{xi}\ \hat{x} + B_{yi}\ \hat{y} + B_{zi}\ \hat{z}) \\ &\cdot \exp[-jk_{x,i}x - jk_y y - jk_{sz,i}(z - t_g - t_w)], \end{aligned} \tag{2.54}$$

where the summation is from $i = -(s-1)/2$ to $(s-1)/2$ and s is the total number of diffraction orders (an odd number for calculation convenience), and the propagation constants $k_{x,i}$, k_y, $k_{cz,i}$ and $k_{sz,i}$ are defined as follows:

◇ Propagation constant in the x direction:

$$k_{x,i} = k_0 n_c \sin\theta\cos\phi - iK; \tag{2.55}$$

◇ Propagation constant in the y direction:

$$k_y = k_0 n_c \sin\theta\sin\phi; \tag{2.56}$$

◇ Propagation constant in the z direction:

$$k_{rz,i} = \begin{cases} \sqrt{{n_r}^2{k_0}^2 - {k_{x,i}}^2 - {k_y}^2} & , {n_r}^2{k_0}^2 > {k_{x,i}}^2 + {k_y}^2 \\ -j\sqrt{{k_{x,i}}^2 + {k_y}^2 - {n_r}^2{k_0}^2} & , {n_r}^2{k_0}^2 < {k_{x,i}}^2 + {k_y}^2 \end{cases} \quad r = c,\ s, \tag{2.57}$$

where $k_0 = 2\pi/\lambda_0$.

In the grating region, the electric field and magnetic field vectors satisfy the Maxwell's equations

$$\nabla \times \bar{E}_g = -j\omega\mu_0\bar{H}_g, \tag{2.58}$$

$$\nabla \times \bar{H}_g = j\omega\epsilon_0\epsilon(x)\bar{E}_g. \tag{2.59}$$

Substituting Eq. (2.49) and (2.50) into Eq. (2.58) and (2.59) and eliminating the z component of the fields (E_{gz} and H_{gz}) result in the coupled-wave equations in a matrix form

$$\begin{bmatrix} \partial\mathbf{S}_y/\partial(z') \\ \partial\mathbf{S}_x/\partial(z') \\ \partial\mathbf{U}_y/\partial(z') \\ \partial\mathbf{U}_x/\partial(z') \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{K}_y\mathbf{E}_{g,n}^{-1}\mathbf{K}_x & \mathbf{I} - \mathbf{K}_y\mathbf{E}_{g,n}^{-1}\mathbf{K}_y \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_x\mathbf{E}_{g,n}^{-1}\mathbf{K}_x - \mathbf{I} & -\mathbf{K}_x\mathbf{E}_{g,n}^{-1}\mathbf{K}_y \\ \mathbf{K}_x\mathbf{K}_y & \mathbf{E}_{g,r}^{-1} - \mathbf{K}_y^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{K}_x^2 - \mathbf{E}_{g,n} & -\mathbf{K}_x\mathbf{K}_y & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \\ \mathbf{U}_y \\ \mathbf{U}_x \end{bmatrix}, \tag{2.60}$$

where $\mathbf{K}_x$ is a diagonal matrix with the (i,i)th element equal to $k_{x,i}/k_0$, $\mathbf{K}_y$ is a diagonal matrix with the (i,i)th element equal to k_y/k_0, z' has a normalized value z/k_0, and the permittivity matrices $\mathbf{E}_{g,n}$ and $\mathbf{E}_{g,r}$ consist of elements $(\epsilon_{rd} - \epsilon_{gv})\sin(i\pi f)/(i\pi)$ and $(1/\epsilon_{rd} - 1/\epsilon_{gv})\sin(i\pi f)/(i\pi)$, respectively, based on Eq. (2.3). The reciprocal permittivity matrix $\mathbf{E}_{g,r}$ is introduced to reduce numerical error in the product of discontinuous functions $\epsilon(x)S_{x,i}$ involved in Eq. (2.59) [147]. Equation (2.60), a $(4s\times4s)$ matrix, is reduced

to either of the following two $(2s \times 2s)$ matrices

$$\begin{bmatrix} \partial^2 \mathbf{S}_y/\partial(z')^2 \\ \partial^2 \mathbf{S}_x/\partial(z')^2 \end{bmatrix} = \begin{bmatrix} \mathbf{K}_x^2 + [\mathbf{K}_y \mathbf{E}_{g,n}^{-1} \mathbf{K}_y - \mathbf{I}]\mathbf{E}_{g,n} & \mathbf{K}_y[\mathbf{E}_{g,n}^{-1} \mathbf{K}_x \mathbf{E}_{g,r}^{-1} - \mathbf{K}_x] \\ \mathbf{K}_x[\mathbf{E}_{g,n}^{-1} \mathbf{K}_y \mathbf{E}_{g,n} - \mathbf{K}_y] & \mathbf{K}_y^2 + [\mathbf{K}_x \mathbf{E}_{g,n}^{-1} \mathbf{K}_x - \mathbf{I}]\mathbf{E}_{g,r}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \end{bmatrix}, \tag{2.61}$$

$$\begin{bmatrix} \partial^2 \mathbf{U}_y/\partial(z')^2 \\ \partial^2 \mathbf{U}_x/\partial(z')^2 \end{bmatrix} = \begin{bmatrix} \mathbf{K}_y^2 + \mathbf{E}_{g,r}^{-1}[\mathbf{K}_x \mathbf{E}_{g,n}^{-1} \mathbf{K}_x - \mathbf{I}] & \mathbf{K}_y[\mathbf{K}_x - \mathbf{E}_{g,r}^{-1} \mathbf{K}_x \mathbf{E}_{g,n}^{-1}] \\ [\mathbf{K}_y - \mathbf{E}_{g,n} \mathbf{K}_y \mathbf{E}_{g,n}^{-1}]\mathbf{K}_x & \mathbf{K}_x^2 + \mathbf{E}_{g,n}[\mathbf{K}_y \mathbf{E}_{g,n}^{-1} \mathbf{K}_y - \mathbf{I}] \end{bmatrix} \begin{bmatrix} \mathbf{U}_y \\ \mathbf{U}_x \end{bmatrix}. \tag{2.62}$$

The diagonal matrix $\mathbf{K}_y$ can be simplified as $k_y \mathbf{I}$ because the value of k_y doesn't depend on the diffraction order. Let $\mathbf{A} = \mathbf{K}_x^2 - \mathbf{E}_{g,n}$ and $\mathbf{B} = \mathbf{K}_x \mathbf{E}_{g,n}^{-1} \mathbf{K}_x - \mathbf{I}$. Equation (2.61) and (2.62) can be further simplified as

$$\left[\partial^2 \mathbf{S}_x/\partial(z')^2\right] = \left[k_y^2 \mathbf{I} + \mathbf{B}\mathbf{E}_{g,r}^{-1}\right] \left[\mathbf{S}_x\right] = \left[\mathbf{M}_s\right] \left[\mathbf{S}_x\right], \tag{2.63}$$

$$\left[\partial^2 \mathbf{U}_y/\partial(z')^2\right] = \left[k_y^2 \mathbf{I} + \mathbf{A}\right] \left[\mathbf{U}_x\right] = \left[\mathbf{M}_u\right] \left[\mathbf{U}_x\right]. \tag{2.64}$$

The simplified coupled-wave equations Eq. (2.63) and (2.64) can be solved by calculating the eigenvalues and the eigenvectors associated with the two $(s \times s)$ matrices $\mathbf{M}_s$ and $\mathbf{M}_u$. As a result, the space harmonics of the tangential electric and magnetic fields in the grating are given as follows:

$\diamond$ x-component of electric field of ith order in the grating:

$$S_{xi}(z) = \sum_{p=1}^{s} w_{2,i,p}^{g} \{ C_{2,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] + C_{2,p}^{g-} \exp[k_0 q_{2,p}^{g}(z - z_2)] \}; \tag{2.65}$$

$\diamond$ y-component of electric field of ith order in the grating:

$$\begin{aligned} S_{yi}(z) = & \sum_{p=1}^{s} v_{11,i,p}^{g} \{ C_{1,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] \\ & + C_{1,p}^{g-} \exp[k_0 q_{2,p}^{g}(z - z_2)] \} \\ & + \sum_{p=1}^{s} v_{12,i,p}^{g} \{ C_{2,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] \\ & + C_{2,p}^{g-} \exp[k_0 q_{2,p}^{g}(z - z_2)] \}; \end{aligned} \tag{2.66}$$

$\diamond$ x-component of magnetic field of ith order in the grating:

$$U_{xi}(z) = \sum_{p=1}^{s} w_{1,i,p}^{g} \{ - C_{1,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] + C_{2,p}^{g-} \exp[k_0 q_{1,p}^{g}(z - z_2)] \}; \tag{2.67}$$

$\diamond$ y-component of magnetic field of ith order in the grating:

$$\begin{aligned} U_{yi}(z) = & \sum_{p=1}^{s} v_{21,i,p}^{g} \{ - C_{1,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] \\ & + C_{1,p}^{g-} \exp[k_0 q_{2,p}^{g}(z - z_2)] \} \\ & + \sum_{p=1}^{s} v_{22,i,p}^{g} \{ - C_{2,p}^{g+} \exp[-k_0 q_{2,p}^{g}(z - z_1)] \\ & + C_{2,p}^{g-} \exp[k_0 q_{2,p}^{g}(z - z_2)] \}, \end{aligned} \tag{2.68}$$

where the superscript g indicates the grating layer, $z_1 = 0$, $z_2 = t_g$, $w_{1,i,p}$ and $q_{1,p}$ are the eigenvectors and the positive square root of the eigenvalues of the matrix $\mathbf{M}_u$ in Eq.

(2.64), respectively, $w_{2,i,p}$ and $q_{2,p}$ are the eigenvectors and the positive square root of the eigenvalues of the matrix $\mathbf{M}_s$ in Eq. (2.63), respectively. The vectors $w_{1,i,p}$ and $w_{2,i,p}$ are arranged as the columns of the matrix $\mathbf{W}_1$ and $\mathbf{W}_2$, respectively, and the matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are diagonal matrices with the diagonal elements $q_{1,p}$ and $q_{2,p}$, respectively. The quantities $v_{11,i,p}$, $v_{12,i,p}$, $v_{21,i,p}$,and $v_{22,i,p}$ are the elements of the matrices $\mathbf{V}_{11}$, $\mathbf{V}_{12}$, $\mathbf{V}_{21}$, and $\mathbf{V}_{22}$ defined by

$$\mathbf{V}_{11} = \mathbf{A}^{-1}\mathbf{W}_1\mathbf{Q}_1, \tag{2.69}$$

$$\mathbf{V}_{12} = (k_y/k_0)\mathbf{A}^{-1}\mathbf{K}_x\mathbf{W}_2, \tag{2.70}$$

$$\mathbf{V}_{21} = (k_y/k_0)\mathbf{B}^{-1}\mathbf{K}_x\mathbf{E}^{-1}\mathbf{W}_1, \tag{2.71}$$

$$\mathbf{V}_{22} = \mathbf{B}^{-1}\mathbf{W}_2\mathbf{Q}_2. \tag{2.72}$$

The space harmonics of the tangential electric and magnetic fields in the waveguide, namely P_{xi}, P_{yi}, Q_{xi}, and Q_{yi}, are of the same forms as Eq. (2.65) to (2.68) except changing the superscript g to w and setting $z_1 = t_g$ and $z_2 = t_g + t_w$. The parameters w, v, and q for the waveguide layer have the same forms as in the grating but can be simplified due to the constant permittivity of the waveguide. In fact, $w^w_{1,i,p} = w^w_{2,i,p} = w^w_{i,p}$ and $q^w_{1,p} = q^w_{2,p} = q^w_p$. Furthermore, given a particular diffraction order i, there exists one and only one $p = \gamma$ such that $w^w_{i,\gamma} = 1$ and $w^w_{i,p} = 0$ for $p \neq \gamma$. As a result, the space harmonics of the tangential electric and magnetic fields in the waveguide can be simplified as follows:

$\diamond$ x-component of electric field of ith order in the waveguide:

$$\begin{aligned} P_{xi}(z) =& C^{w+}_{2,\gamma} \exp[-k_0 q^w_\gamma (z - t_g)] \\ &+ C^{w-}_{2,\gamma} \exp[k_0 q^w_\gamma (z - t_g - t_w)]; \end{aligned} \tag{2.73}$$

$\diamond$ y-component of electric field of ith order in the waveguide:

$$\begin{aligned} P_{yi}(z) =&(v^w_{11,i,\gamma}C^{w+}_{1,\gamma} + v^w_{12,i,\gamma}C^{w+}_{2,\gamma})\exp[-k_0q^w_\gamma(z-t_g)] \\ &+ (v^w_{11,i,\gamma}C^{w-}_{1,\gamma} + v^w_{12,i,\gamma}C^{w-}_{2,\gamma})\exp[k_0q^w_\gamma(z-t_g-t_w)] \\ =&I^{w+}_i\exp[-k_0q^w_\gamma(z-t_g)] + I^{w-}_i\exp[k_0q^w_\gamma(z-t_g-t_w)]; \end{aligned} \tag{2.74}$$

$\diamond$ x-component of magnetic field of ith order in the waveguide:

$$\begin{aligned} Q_{xi}(z) =& -C^{w+}_{1,\gamma}\exp[-k_0q^w_\gamma(z-t_g)] \\ &+ C^{w-}_{1,\gamma}\exp[k_0q^w_\gamma(z-t_g-t_w)]; \end{aligned} \tag{2.75}$$

$\diamond$ y-component of magnetic field of ith order in the waveguide:

$$\begin{aligned} Q_{yi}(z) =&(-v^w_{21,i,\gamma}C^{w+}_{1,\gamma} - v^w_{22,i,\gamma}C^{w+}_{2,\gamma})\exp[-k_0q^w_\gamma(z-t_g)] \\ &+ (v^w_{21,i,\gamma}C^{w-}_{1,\gamma} + v^w_{22,i,\gamma}C^{w-}_{2,\gamma})\exp[k_0q^w_\gamma(z-t_g-t_w)] \\ =&J^{w+}_i\exp[-k_0q^w_\gamma(z-t_g)] + J^{w-}_i\exp[k_0q^w_\gamma(z-t_g-t_w)]. \end{aligned} \tag{2.76}$$

Actually, $k_0q^w_\gamma = jk_{wz,i}$, where $k_{wz,i}$ is defined in the same way as in Eq. (2.57).

Step 4: Formulations in the system (x', y', z')

Now the coordinate system is switched from (x, y, z) to (x', y', z'), as indicated in Fig. 2.12. This step simplifies the E field and H field expressions of the incident wave, as can be seen by comparing Eq. (2.46) and Eq. (2.48) with Eq. (2.78) and Eq. (2.80), respectively. The main goal of switching the coordinate system is to follow the formulation in [115]. The final results will not be affected if the calculations are done in (x, y, z). For a particular diffraction order i, the rotation angle φ_i about the z_b axis is determined by $\varphi_i = \arctan(k_y/k_{xi})$, and this rotation angle can be visualized in Fig. 2.14. For the $i = 0$ order, $\varphi_0 = \phi$. The new coordinate system is related to the original system by $X' =$

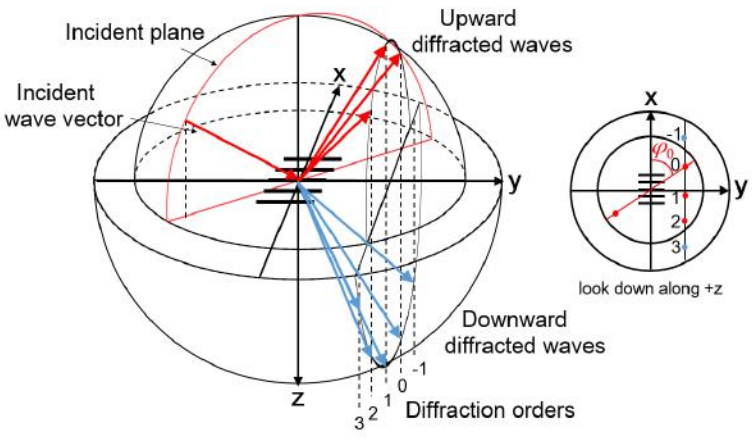


Figure 2.14: 3D wave vector diagram of conical diffraction. The top half sphere represents the $k_0 n_c$ surface, and the bottom half sphere represents the $k_0 n_s$ surface.

$\cos\varphi_i\, X + \sin\varphi_i\, Y$ and $Y' = -\sin\varphi_i\, X + \cos\varphi_i\, Y$, where X and Y represent variables in the original coordinate system, and X' and Y' represent variables in the new coordinate system. The formulations follow [115] except for those representing the tangential fields in the cover and substrate. Only the modifications are presented here. In the new coordinate system:

◇ Electric field in the cover:

$$E'_c = E'_{inc} + \sum_i (R'_{xi}\,\hat{x}' + R'_{yi}\,\hat{y}' + R_{zi}\,\hat{z}) \cdot \exp(-jk'_{x,i}x' - jk'_y y' + jk_{cz,i}z), \tag{2.77}$$

where

$$E'_{inc} = (\cos\psi\cos\theta\,\hat{x}' + \sin\psi\,\hat{y}' - \cos\psi\sin\theta\,\hat{z}) \cdot \exp[-jk_0 n_c(\sin\theta\,x' + \cos\theta\,z)]; \tag{2.78}$$

◇ Magnetic field in the cover:

$$H'_c = H'_{inc} + (-j\sqrt{\frac{\epsilon_0}{\mu_0}}) \sum_i (A'_{xi}\,\hat{x}' + A'_{yi}\,\hat{y}' + A_{zi}\,\hat{z}) \cdot \exp(-jk'_{x,i}x' - jk'_y y' + jk_{cz,i}z), \tag{2.79}$$

where

$$H'_{inc} = n_c\sqrt{\frac{\epsilon_0}{\mu_0}}[-\sin\psi\cos\theta\ \hat{x}' + \cos\psi\ \hat{y}' + \sin\psi\sin\theta\ \hat{z}] \cdot \exp[-jk_0 n_c(\sin\theta\ x' + \cos\theta\ z)]; \tag{2.80}$$

◇ Electric field in the substrate:

$$E'_s = \sum_i (T'_{xi}\ \hat{x}' + T'_{yi}\ \hat{y}' + T_{zi}\ \hat{z}) \cdot \exp[-jk'_{x,i}x' - jk'_y y' - jk_{sz,i}(z - t_g - t_w)]; \tag{2.81}$$

◇ Magnetic field in the substrate:

$$H_s = (-j\sqrt{\frac{\epsilon_0}{\mu_0}})\sum_i (B'_{xi}\ \hat{x}' + B'_{yi}\ \hat{y}' + B_{zi}\ \hat{z}) \cdot \exp[-jk'_{x,i}x' - jk'_y y - jk_{sz,i}(z - t_g - t_w)], \tag{2.82}$$

From Maxwell's equations and the characteristics of plane waves $\bar{E}\cdot\bar{k} = 0$, the magnetic field components A'_{xi}, A'_{yi}, B'_{xi}, and B'_{yi} can each be expressed as a function of the electric field components:

$$A'_{xi} = \frac{j}{k_0 k_{c,zi}}[k'_{xi}k'_y R'_{xi} + ({k_0}^2{n_c}^2 - {k'_{xi}}^2)R'_{yi}], \tag{2.83}$$

$$A'_{yi} = \frac{-j}{k_0 k_{c,zi}}[({k_0}^2{n_c}^2 - {k'_y}^2)R'_{xi} + k'_{xi}k'_y R'_{yi}], \tag{2.84}$$

$$B'_{xi} = \frac{-j}{k_0 k_{s,zi}}[k'_{xi}k'_y T'_{xi} + ({k_0}^2{n_s}^2 - {k'_{xi}}^2)T'_{yi}], \tag{2.85}$$

$$B'_{yi} = \frac{j}{k_0 k_{s,zi}}[({k_0}^2{n_s}^2 - {k'_y}^2)T'_{xi} + k'_{xi}k'_y T'_{yi}]. \tag{2.86}$$

Then, all the tangential fields in the new coordinate system are matched at boundaries $z = 0$ (cover-grating interface), $z = t_g$ (grating-waveguide interface) and $z = t_g + t_w$ (waveguide-substrate interface), resulting in the following matrices:

$\diamond$ at $z = 0$:

$$\begin{bmatrix} \sin\psi\ \delta_{i0} \\ jn_c \sin\psi\cos\theta\ \delta_{i0} \\ \cos\psi\cos\theta\ \delta_{i0} \\ -jn_c\cos\psi\ \delta_{i0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -j\mathbf{RA} & -j\mathbf{RB} \\ \mathbf{I} & \mathbf{0} \\ j\mathbf{RC} & j\mathbf{RA} \end{bmatrix} \begin{bmatrix} \mathbf{R}'_x \\ \mathbf{R}'_y \end{bmatrix} = \begin{bmatrix} \mathbf{V}^g_{ss} & \mathbf{V}^g_{sp} & \mathbf{V}^g_{ss}\mathbf{X}^g_1 & \mathbf{V}^g_{sp}\mathbf{X}^g_2 \\ \mathbf{W}^g_{ss} & \mathbf{W}^g_{sp} & -\mathbf{W}^g_{ss}\mathbf{X}^g_1 & -\mathbf{W}^g_{sp}\mathbf{X}^g_2 \\ \mathbf{V}^g_{ps} & \mathbf{V}^g_{pp} & \mathbf{V}^g_{ps}\mathbf{X}^g_1 & \mathbf{V}^g_{pp}\mathbf{X}^g_2 \\ \mathbf{W}^g_{ps} & \mathbf{W}^g_{pp} & -\mathbf{W}^g_{ps}\mathbf{X}^g_1 & -\mathbf{W}^g_{pp}\mathbf{X}^g_2 \end{bmatrix} \begin{bmatrix} \mathbf{C}^{g+}_1 \\ \mathbf{C}^{g+}_2 \\ \mathbf{C}^{g-}_1 \\ \mathbf{C}^{g-}_2 \end{bmatrix}; \tag{2.87}$$

$\diamond$ at $z = t_g$:

$$\begin{bmatrix} \mathbf{V}^g_{ss}\mathbf{X}^g_1 & \mathbf{V}^g_{sp}\mathbf{X}^g_2 & \mathbf{V}^g_{ss} & \mathbf{V}^g_{sp} \\ \mathbf{W}^g_{ss}\mathbf{X}^g_1 & \mathbf{W}^g_{sp}\mathbf{X}^g_2 & -\mathbf{W}^g_{ss} & -\mathbf{W}^g_{sp} \\ \mathbf{V}^g_{ps}\mathbf{X}^g_1 & \mathbf{V}^g_{pp}\mathbf{X}^g_2 & \mathbf{V}^g_{ps} & \mathbf{V}^g_{pp} \\ \mathbf{W}^g_{ps}\mathbf{X}^g_1 & \mathbf{W}^g_{pp}\mathbf{X}^g_2 & -\mathbf{W}^g_{ps} & -\mathbf{W}^g_{pp} \end{bmatrix} \begin{bmatrix} \mathbf{C}^{g+}_1 \\ \mathbf{C}^{g+}_2 \\ \mathbf{C}^{g-}_1 \\ \mathbf{C}^{g-}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}^w_{ss} & \mathbf{V}^w_{sp} & \mathbf{V}^w_{ss}\mathbf{X}^w_1 & \mathbf{V}^w_{sp}\mathbf{X}^w_2 \\ \mathbf{W}^w_{ss} & \mathbf{W}^w_{sp} & -\mathbf{W}^w_{ss}\mathbf{X}^w_1 & -\mathbf{W}^w_{sp}\mathbf{X}^w_2 \\ \mathbf{V}^w_{ps} & \mathbf{V}^w_{pp} & \mathbf{V}^w_{ps}\mathbf{X}^w_1 & \mathbf{V}^w_{pp}\mathbf{X}^w_2 \\ \mathbf{W}^w_{ps} & \mathbf{W}^w_{pp} & -\mathbf{W}^w_{ps}\mathbf{X}^w_1 & -\mathbf{W}^w_{pp}\mathbf{X}^w_2 \end{bmatrix} \begin{bmatrix} \mathbf{C}^{w+}_1 \\ \mathbf{C}^{w+}_2 \\ \mathbf{C}^{w-}_1 \\ \mathbf{C}^{w-}_2 \end{bmatrix}; \tag{2.88}$$

$\diamond$ at $z = t_g + t_w$:

$$\begin{bmatrix} \mathbf{V}_{ss}^{w}\mathbf{X}_1^{w} & \mathbf{V}_{sp}^{w}\mathbf{X}_2^{w} & \mathbf{V}_{ss}^{w} & \mathbf{V}_{sp}^{w} \\ \mathbf{W}_{ss}^{w}\mathbf{X}_1^{w} & \mathbf{W}_{sp}^{w}\mathbf{X}_2^{w} & -\mathbf{W}_{ss}^{w} & -\mathbf{W}_{sp}^{w} \\ \mathbf{V}_{ps}^{w}\mathbf{X}_1^{w} & \mathbf{V}_{pp}^{w}\mathbf{X}_2^{w} & \mathbf{V}_{ps}^{w} & \mathbf{V}_{pp}^{w} \\ \mathbf{W}_{ps}^{w}\mathbf{X}_1^{w} & \mathbf{W}_{pp}^{w}\mathbf{X}_2^{w} & -\mathbf{W}_{ps}^{w} & -\mathbf{W}_{pp}^{w} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^{w+} \\ \mathbf{C}_2^{w+} \\ \mathbf{C}_1^{w-} \\ \mathbf{C}_2^{w-} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ j\mathbf{TA} & j\mathbf{TB} \\ \mathbf{I} & \mathbf{0} \\ -j\mathbf{TC} & -j\mathbf{TA} \end{bmatrix} \begin{bmatrix} \mathbf{T}_x' \\ \mathbf{T}_y' \end{bmatrix}, \tag{2.89}$$

where

$$\begin{aligned} \mathbf{V}_{ss}^{r} &= \mathbf{F}_c\mathbf{V}_{11}^{r}, & \mathbf{V}_{sp}^{r} &= -\mathbf{F}_s\mathbf{W}_2^{r} + \mathbf{F}_c\mathbf{V}_{12}^{r}, \\ \mathbf{W}_{ss}^{r} &= \mathbf{F}_c\mathbf{W}_1^{r} + \mathbf{F}_s\mathbf{V}_{21}^{r}, & \mathbf{W}_{sp}^{r} &= \mathbf{F}_s\mathbf{V}_{22}^{r}, \\ \mathbf{V}_{ps}^{r} &= \mathbf{F}_s\mathbf{V}_{11}^{r}, & \mathbf{V}_{pp}^{r} &= \mathbf{F}_c\mathbf{W}_2^{r} + \mathbf{F}_s\mathbf{V}_{12}^{r}, \\ \mathbf{W}_{ps}^{r} &= -\mathbf{F}_s\mathbf{W}_1^{r} + \mathbf{F}_c\mathbf{V}_{21}^{r}, & \mathbf{W}_{pp}^{r} &= \mathbf{F}_c\mathbf{V}_{22}^{r}, \end{aligned} \tag{2.90}$$

where $r = g$ or w, $\mathbf{F}_c$, $\mathbf{F}_s$, **RA**, **RB**, **RC**, **TA**, **TB**, and **TC** are diagonal matrices with diagonal elements $\cos\varphi_i$, $\sin\varphi_i$, $k'_{xi}k'_y/(k_0k_{c,zi})$, $[k_0^2n_c^2-(k'_{xi})^2]/(k_0k_{c,zi})$, $[k_0^2n_c^2-(k'_y)^2]/(k_0k_{c,zi})$, $k'_{xi}k'_y/(k_0k_{s,zi})$, $[k_0^2n_s^2-(k'_{xi})^2]/(k_0k_{s,zi})$, $[k_0^2n_s^2-(k'_y)^2]/(k_0k_{s,zi})$, respectively, and the matrices $\mathbf{V}_{11}^{r}$, $\mathbf{V}_{12}^{r}$, $\mathbf{V}_{21}^{r}$, and $\mathbf{V}_{22}^{r}$ are defined by Eq. (2.69) to (2.72). Note the field amplitudes in the grating region correspond to the (x_t, y_t, z_t) coordinate system.

After solving the above matrix equations using the transfer matrix approach [116] or singular value decomposition [41], the field amplitudes R, T and C can be determined. The diffraction efficiencies in the cover (DE_c) and the substrate (DE_s) can be subsequently cal-

culated:

$$DE_{c,i} = \frac{\mathrm{Re}[R'_{xi}(-jA'_{yi})^* - R'_{yi}(-jA'_{xi})^*]}{n_c \cos\theta}, \tag{2.91}$$

$$DE_{s,i} = \frac{\mathrm{Re}[T'_{xi}(-jB'_{yi})^* - T'_{yi}(-jB'_{xi})^*]}{n_c \cos\theta}. \tag{2.92}$$

The above formulations produce the same results as those presented in [115]. The sum of the diffraction efficiencies is unity as a result of energy conservation in the z direction. Compared with those presented in [115], the present formulations have the advantages that the x and y component of electric and magnetic fields are clearly presented and calculated, and they can be easily referenced in the following analysis.

Step 5: Determination of Coupling Efficiency into the Waveguide

As explained in Section 2.1 and demonstrated in Section 2.2, RCWA-EIS method is capable of calculating grating in-coupling efficiencies into the waveguide for various grating profiles under the planar diffraction. RCWA-EIS method can also be extended to analyze the coupling efficiency under conical mounting, provided the propagating $i = 0$ and $i = +1$ orders are dominant.

Under perfect interlayer grating alignment, a TE-polarized guided wave (with field components E_y, H_z, H_x) with propagation constant β is diffracted out of the top grating in the direction of $i = +1$ order and incident onto the bottom grating. As a result, the propagation constant of the coupled wave in the bottom waveguide should also be β. When the bottom grating is rotated relative to the top grating, the out-diffracted light from the top grating is conically incident on the bottom grating. For demonstration purposes, only coupling into the TE mode of the bottom grating is considered.

The first step in determining the coupling efficiency into the bottom waveguide is to find the equivalent index slabs with indices $\tilde{n}_l$ that reproduce the fields outside of the grat-

ing layer. All fields in the structures as well as the propagation constants can be readily obtained from the in-coupling process discussed in the previous sections. Four equivalent index layers with uniform refractive indices are used to replace the grating layer, and the ith order electric field of the lth slab in the y direction is represented as in Eq. (2.27), where $t_l = t_g/4$. With small rotation angles, higher diffraction orders may propagate but they are not dominant. Therefore, only two diffraction orders ($i = 0$ and $i = +1$) will be considered in determining the equivalent indices. By imposing boundary conditions on the tangential electric (E_y) and magnetic (H_x) fields in the coordinate system (x_b, y_b, z_b), the following transfer matrix formulation can be obtained for the $i = 0$ and $+1$ orders:

$$\begin{bmatrix} a\,\delta_{i0} + R_{yi} \\ b\,\delta_{i0} - jk_0 A_{xi} \end{bmatrix} = \prod_{l=1}^{4} \begin{bmatrix} 1 & X_{l,i} \\ -\tilde{k}_{lz,i} & \tilde{k}_{lz,i} X_{l,i} \end{bmatrix} \begin{bmatrix} X_{l,i} & 1 \\ -\tilde{k}_{lz,i} X_{l,i} & \tilde{k}_{lz,i} \end{bmatrix}^{-1} \begin{bmatrix} I_i^{w+} + I_i^{w-} X_{w,i} \\ -jk_0(-C_{i,\gamma}^{w+} + C_{i,\gamma}^{w-} X_{w,i}) \end{bmatrix}, \tag{2.93}$$

where $X_{l,i} = \exp(-j\tilde{k}_{lz,i} t_l)$, $X_{w,i} = \exp(-jk_{wz,i} t_w)$, $a = \cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi$, $b = -k_0 n_c(\cos\psi\sin\phi + \sin\psi\cos\phi\cos\theta)$, $R_{yi} = \sin\varphi_i\, R'_{xi} + \cos\varphi_i\, R'_{yi}$, $A_{xi} = \cos\varphi_i\, A'_{xi} - \sin\varphi_i\, A'_{yi}$, and $\tilde{k}_{lz,i}$ is defined by Eq. (2.29) in which $k_{x,i}$ is defined by Eq. (2.55).

Equation (2.93) generates 2 equations for each of the $i = 0$ and $i = +1$ orders, and thus there will be a total of 4 equations. The problem then becomes finding 4 equivalent refractive indices $\tilde{n}_l$ such that the 4 equations are satisfied simultaneously. This is achieved with the Matlab function *fsolve* with the Trust-Region-Dogleg algorithm [43].

After the equivalent indices are determined, the radiation factor α is the single unknown

which is incorporated in the following transfer matrix formulation

$$\begin{bmatrix} 1 \\ \tilde{k}_{cz,i} \end{bmatrix} R_i = \prod_{l=1}^{4} \begin{bmatrix} 1 & X_{l,i} \\ -\tilde{k}_{lz,i} & \tilde{k}_{lz,i}X_{l,i} \end{bmatrix} \begin{bmatrix} X_{l,i} & 1 \\ -\tilde{k}_{lz,i}X_{l,i} & \tilde{k}_{lz,i} \end{bmatrix}^{-1} \begin{bmatrix} 1 & X_{w,i} \\ -\tilde{k}_{wz,i} & \tilde{k}_{wz,i}X_{w,i} \end{bmatrix} \begin{bmatrix} X_{w,i} & 1 \\ -\tilde{k}_{wz,i}X_{w,i} & \tilde{k}_{wz,i} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -\tilde{k}_{sz,i} \end{bmatrix} T_i = \begin{bmatrix} Q_{i,1} \\ Q_{i,2} \end{bmatrix} T_i \tag{2.94}$$

valid for both $i = 0$ and $i = +1$ orders, where $\tilde{k}_{cz,i}$, $\tilde{k}_{wz,i}$, $\tilde{k}_{sz,i}$ are calculated based on Eq. (2.26), in which $\tilde{k}_{x,0} = \beta - j\alpha$ and $\tilde{k}_{x,1} = K - \beta - j\alpha$, where K is the grating vector ($K = 2\pi/\Lambda$, Λ is the grating period). The sign of the real part of $\tilde{k}_{x,1}$ is not of consequence since $\tilde{k}_{x,1}$ is squared when calculating propagation constants in z direction. Equation (2.94) results in 2 equations $\tilde{k}_{cz,i}Q_{i,1} = Q_{i,2}$ ($i = 0$ or 1) with one unknown α, which can be solved by Matlab function *fsolve* with the Levenberg-Marquardt algorithm.

After α is determined, the diffraction efficiency can be calculated using the method stated in Section 2.1, except changing the expression for $k_{x,i}$ from Eq. (2.7) to Eq. (2.55) when calculating the matrices involved in Eq. (2.24).

2.3.2 Results and Discussion

For demonstration purposes, two optimized binary grating designs obtained by RCWA-EIS method are used to investigate the rotation effects on interlayer grating coupling efficiency. The given parameters are as follows: free-space wavelength $\lambda_0 = 1.55$ μm, cover refractive index $n_c = 1$, substrate refractive index $n_s = 1.45$, grating groove refractive index $n_{gr} = 1$, grating ridge refractive index $n_{rd} = 2.46$, waveguide refractive index $n_w = 3.45$, grating fill factor $f = 0.5$, number of grating periods $N = 30$ or 50 (depending on the rotation configuration), waveguide thickness $t_w = 0.22$ μm, and waveguide separation in z direction

Table 2.2: Optimized parameters and calculated single grating diffraction efficiencies for binary grating case 1 and 2.

Case Number	θ_i [rad / °]	Λ [μm]	t_g [μm]	PC	α [μm^{-1}]	DE @ N=50	DE @ N=30
1	0.2665 rad (15.27°)	0.6316	0.3951	0.7734	0.0138	0.4499	0.3150
2	0.2021 rad (11.58°)	0.6477	0.3249	0.7163	0.0167	0.4735	0.3420

$d = 2$ μm. The fundamental mode propagation constant of the 0.22 μm thick waveguide is calculated to be $\beta_0 = 11.3710$ μm^{-1}. The total number of space harmonics is set to be $s = 7$. Under the perfectly aligned conditions, the coupling angle θ, grating period Λ, grating thickness t_g, preferential coupling ratio (PC), radiation factor (α), single grating diffraction efficiency (DE) at $N = 30$ and $N = 50$ of two optimized cases are summarized in Table 2.2. Randomly selecting grating parameters may cause irregular and unstable results, which cannot correctly represent changes and trends in diffraction efficiencies. Each case will be investigated in the situation for which the bottom substrate is rotated about the x_t, z_t, and the vector [2 2 1] defined in the top coordinates, respectively. The coupling efficiency into the TE mode (with field components E_y, H_x, H_z) is calculated.

The interlayer grating coupling efficiency η can be approximated by $\eta = DE_t \cdot DE_b$, where DE_t and DE_b are the single grating diffraction efficiencies of the top and the bottom grating, respectively. Since the misalignment is treated as the relative rotation of the bottom grating with respect to the top grating, the top grating diffraction efficiency is the same as that in the perfectly aligned situation; that is, $DE_t = \sqrt{\eta_0}$, where η_0 is the interlayer grating coupling efficiency under perfect alignment. Thus, the bottom grating diffraction efficiency can be approximated as $DE_b = \eta/\sqrt{\eta_0}$. This approximation is used to obtain bottom grating diffraction efficiency from the interlayer grating coupling efficiency calculated by 3D FDTD. RCWA-EIS method considers coupling the out-diffracted light from the top grating into the TE mode of the bottom waveguide. 3D FDTD calculates the total guided flux (both TE and TM contributions) in the bottom grating, but the TM contributions are smaller than TE ones, as discussed in the following section.

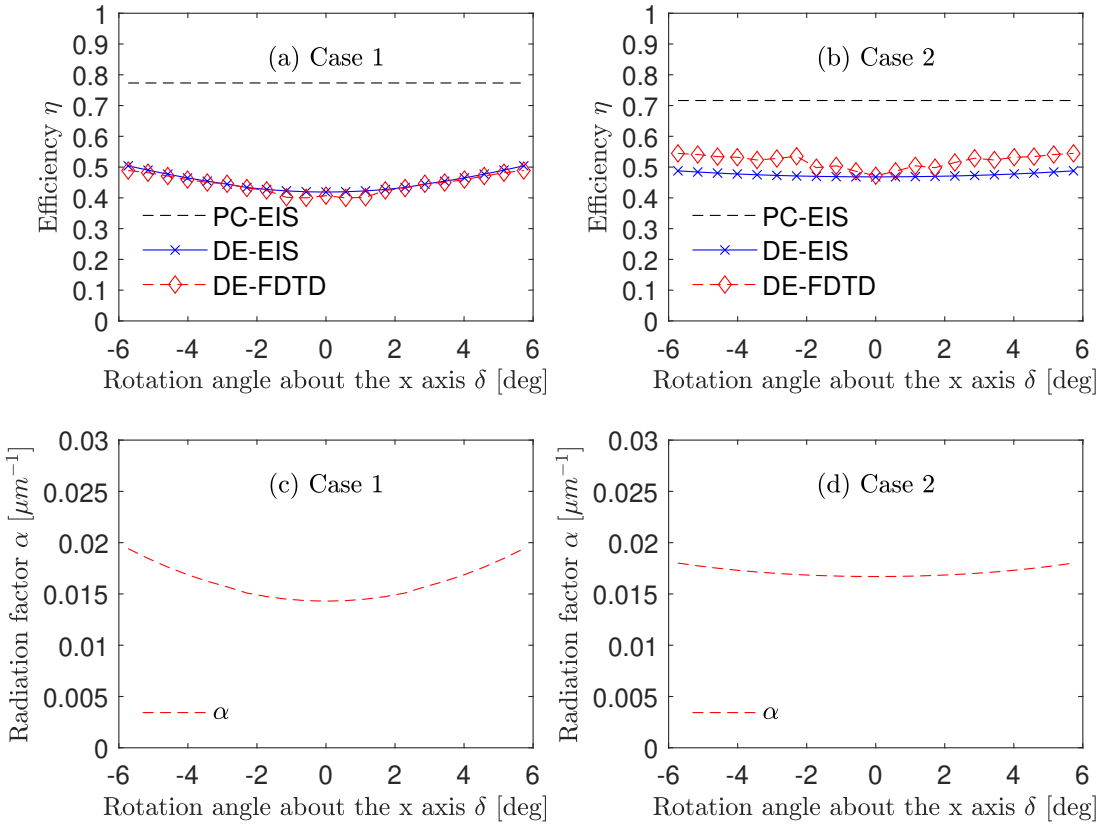


Figure 2.15: Plots of PC-EIS, DE-EIS, DE-FDTD, and α for case 1 and 2, respectively, when the bottom system is rotated about the x_t axis. The number of grating period N is set as 50.

Rotation about the x_t axis

Figures 2.15(a) and (b) show the preferential coupling ratios (PC-EIS) or the branching ratio calculated by RCWA-EIS method, the bottom grating diffraction efficiencies calculated by RCWA-EIS method (DE-EIS) and 3D FDTD simulation (DE-FDTD) for the two cases when the bottom grating is rotated about the x_t axis. Figures 2.15(c) and (d) show the radiation factor α for the two cases. It can be observed that PC exhibits negligible changes as the rotation angle about the x_t changes from -5.73° (-0.1 rad) to 5.73° (0.1 rad). This indicates that for an infinitely long and infinitely wide grating, the in-coupled light of TE polarization is unaffected when the rotation angle is small ($\pm$ 5.73°). This is due to the fact that the Bragg condition is not significantly disturbed when the grating is rotated

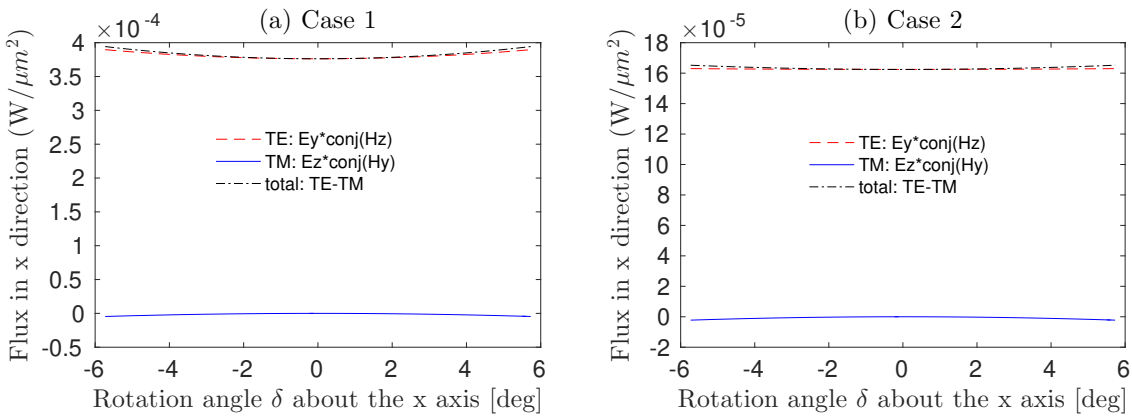


Figure 2.16: Guided power along x_b direction in the bottom waveguide, including TE power, TM power and total power, calculated by RCWA. Power is integrated from $z = t_g$ to $z = t_g + t_w$ for a unit length in y_b direction.

about the x_t (x_b equivalently) axis, which is perpendicular to the grating grooves. It can also be observed that the shape of the DE-EIS curves is largely determined by the shape of the α curves, which means that the radiation factor is critical in determining the bottom grating diffraction efficiency when the grating length $\ell = N\Lambda$ is limited in this rotation configuration. The radiation factors smoothly increase as the rotation angle increases from the perfectly aligned situation ($\alpha = 0.0138$ for case 1 and $\alpha = 0.0167$ for case 2 at rotation angle = 0°), giving rise to valley-shaped DE-EIS curves. At rotation angles $\pm 5.73°$ ($\pm$ 0.1 rad), the DE-EIS curve increases about 10% for case 1 and 5% for case 2. The increase in diffraction efficiency is well known in over-modulated gratings under off-Bragg conditions [104]. This can be understood as follows: as rotation angle increases, the efficiency of the $i = +1$ order may increase to such a point when the $i = +2$ order starts to propagate, from which the efficiency of the $i = +1$ order will drop and power is allocated to the $i = +2$ order. The FDTD results also demonstrate the trend in which diffraction efficiency increases as rotation angle about the x_t axis increases. The 3D FDTD calculates the total guided flux in the bottom waveguide along the x_b direction defined by the following equation:

$$\bar{S} \cdot \hat{x}_b = \frac{1}{2} \mathrm{Re}(E_y \cdot H_z^* - E_z \cdot H_y^*), \tag{2.95}$$

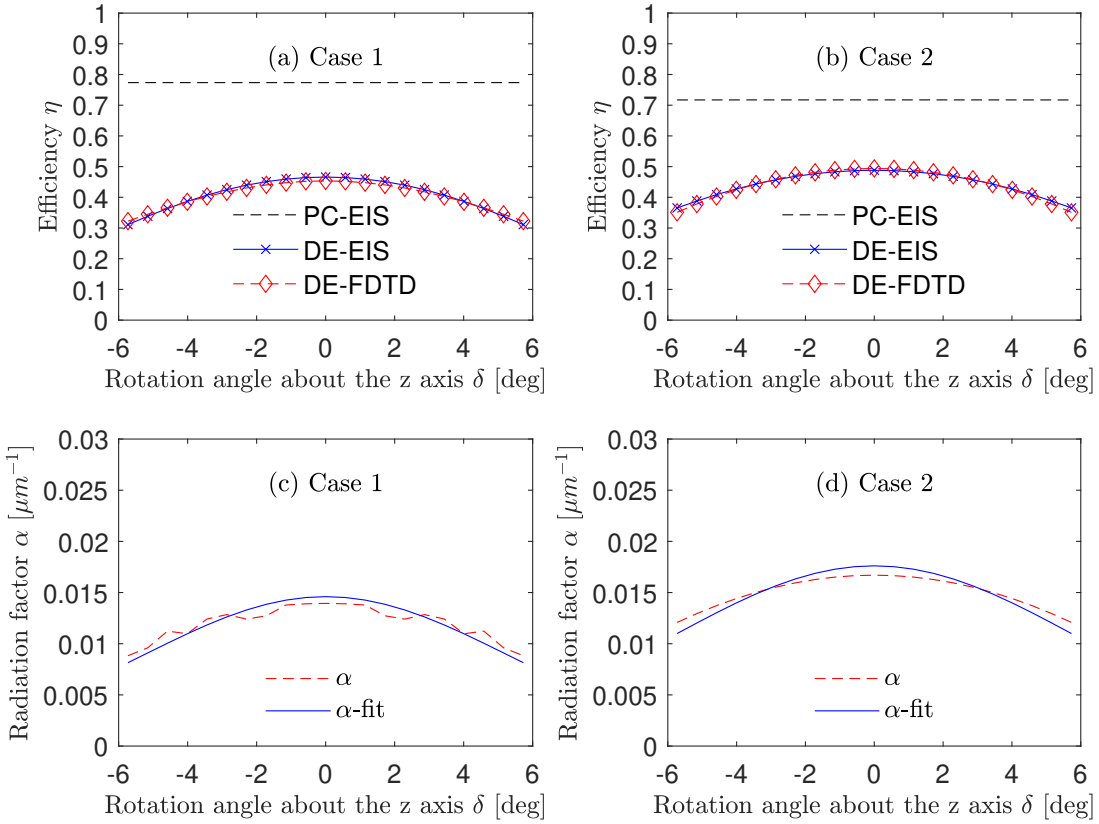


Figure 2.17: Plots of PC-EIS, DE-EIS, DE-FDTD, and α for case 1 and 2, respectively, when the bottom system is rotated about the z_t axis. The number of grating periods N is set as 50.

where the first part is TE contribution and the second part is TM contribution. However, the TM power is much smaller, as shown in Fig. 2.16. As a result, the flux calculated by 3D FDTD can be treated as TE type. An important observation should be made regarding the numerical instability of the 3D FDTD calculation, and this is discussed in Appendix B.

Rotation about the z_t axis

Figures 2.17(a) and (b) show the PC-EIS, DE-EIS and DE-FDTD plots for the two cases when the bottom grating is rotated about the z_t axis. Figures 2.17(c) and (d) show the α plots for the two cases. The α curve for case 1 exhibits a noticeably wavy shape due to numerical sensitivity problems. In order to give a better presentation, the curve is Gaussian-fitted. The same fitting procedure is also done on the α curve for case 2. Like in the

x-rotation cases, the PC curves exhibit negligible changes as the bottom grating rotates about the z_t axis, and the shapes of the DE-EIS curves are determined by those of the radiation factors. At rotation angles $\pm 5.73^\circ$ (± 0.1 rad), the DE-EIS curve decreases about 25% for both cases. The FDTD results are in good agreement with the RCWA-EIS ones because rotating about the z_t axis, or equivalently, the surface normal of the waveguide, does not suffer from the staircase approximation problem (discussed in Appendix B). In other words, the waveguide surface remains flat, and the only place that requires the staircase approximation is on the edges. Since the width of the waveguide ($W = 10$ μm) is set much larger than the waveguide thickness ($t_g = 0.22$ μm) to model a slab waveguide, the rough edges would not affect the field distribution in the slab structure.

Rotation about the vector [2 2 1]

Figures 2.18(a) and (b) show the PC-EIS, DE-EIS and DE-FDTD plots for the two cases when the bottom grating is rotated about an axis defined by the vector [2 2 1] in the top coordinate system. Figures 2.18(c) and (d) show the α plots for the two cases. Since the bottom waveguide may be in contact with the top waveguide after rotations in FDTD models, a shorter grating length with $N = 30$ is used instead of $N = 50$. The locations of the corners of the waveguides and grating ridges are calculated to ensure the top and bottom structures are not in contact. In this rotation configuration, α changes significantly and it influences the definition of DE-EIS. The resulting shapes of the DE-EIS curves are not symmetric. The rotation axis has a y_t component, and rotating about the y_t axis has an equivalent effect as changing the incident angle of the incoming light to the bottom grating, which greatly disturbs the Bragg condition and induces a large change in the diffraction efficiency. Both the DE-EIS and DE-FDTD curves first increase and then decrease when the rotation angle changes from 0° to -5.73° (-0.1 rad), while the curves monotonically decrease when the rotation angle changes from 0° to 5.73° (0.1 rad). The efficiency hill in the $\delta < 0^\circ$ region may also due to the over-modulation of gratings under off-Bragg conditions.

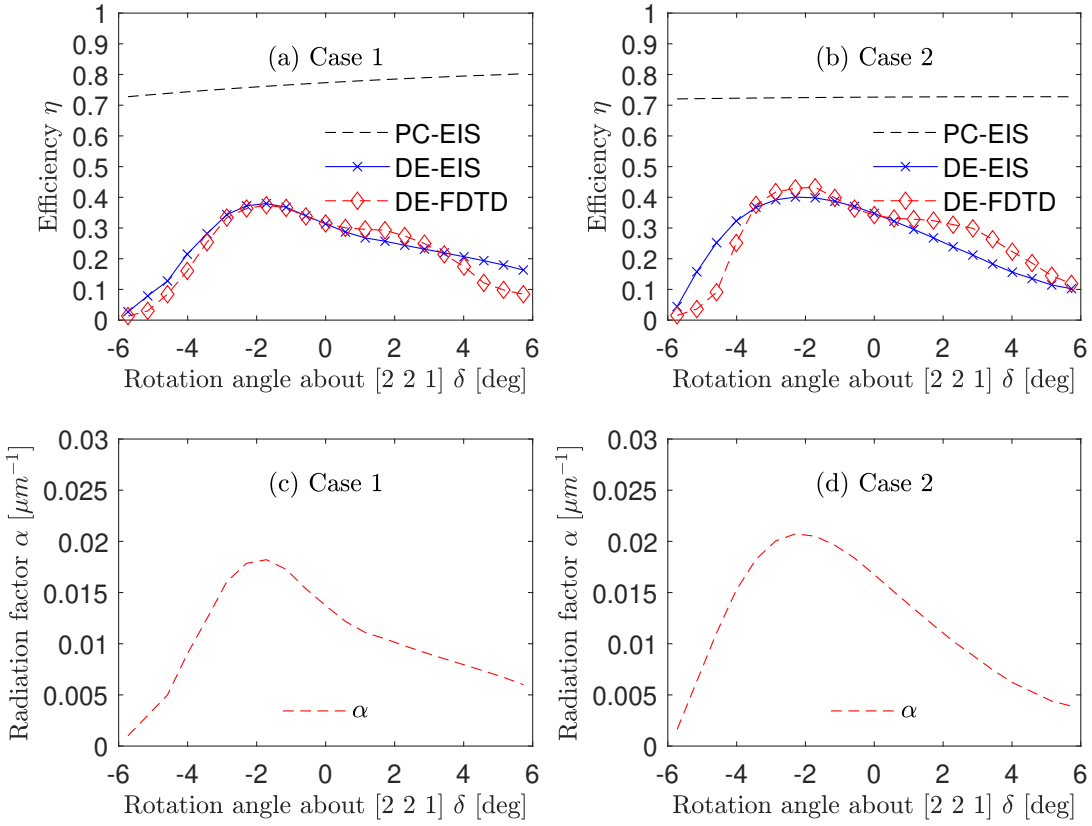


Figure 2.18: Plots of PC-EIS, DE-EIS, DE-FDTD, and α for case 1 and 2, respectively, when the bottom system is rotated about the vector [2 2 1]. The number of grating period N is set as 30.

While the data points of DE-EIS and DE-FDTD at 0° rotation match well, the general shapes of the curves at the two ends exhibit noticeable discrepancies. This is because an infinitely long slab waveguide grating is assumed in RCWA-EIS method, while structures with limited dimensions are modeled in the FDTD simulation. The influence of this geometric factor is explained using Figure 2.19. The input light is launched into the top waveguide along the $-x$ direction and out-diffracted when it is incident onto the top grating. Both the out-diffracted power and the guided power in the top waveguide exhibit an exponential decay along the $-x$ direction. When the bottom waveguide is rotated by $+\delta$ about the y axis, the right end of the bottom waveguide is closer to the top waveguide, and the high power (indicated by the thickest arrow) confined in the top waveguide may be evanescently coupled to the bottom waveguide, resulting in larger diffraction efficien-

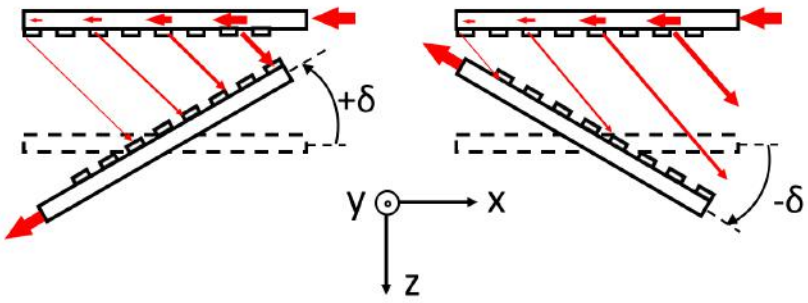


Figure 2.19: Schematic diagram show positive ($+\delta$) and negative ($-\delta$) rotation of the bottom grating about the y axis. Rotation angles are exaggerated.

cies from the FDTD simulations than those from the RCWA-EIS calculations at positive rotation angles. On the contrary, the left end of the bottom waveguide is closer to the top grating when the bottom grating is rotated by $-\delta$. Since the power at the left end is much smaller, the coupled power in the bottom waveguide is negligible. The same reasoning can be made in explaining the offsets of the DE curves obtained by RCWA and FDTD in Fig. 2.18. The rotation configuration about the vector [2 2 1] is shown in Fig. 2.20. The positively rotated grating is closer to the high power end of the top grating, and thus FDTD results show a moderate efficiency increase compared with the RCWA-EIS results in the $\delta > 0^\circ$ region, while the negatively rotated bottom grating is closer to the low power end of the top grating, resulting in smaller FDTD values than the RCWA-EIS results in the $\delta < 0^\circ$ region. As a result, system designers should try to align the grating such that there is minimal rotation about the y axis, e.g. rotation should be limited to $\pm 3^\circ$ to achieve less than 10% change in DE_b.

2.4 Conclusion

RCWA-EIS method is introduced to calculate waveguide grating coupling efficiencies. The "Equivalent Index Slab" (EIS) concept is proposed to extend the traditional RCWA method to the analysis of waveguide gratings involving surface waves. RCWA-EIS method overcomes the numerical instability issue in RCWA-LW approach regarding grating structures

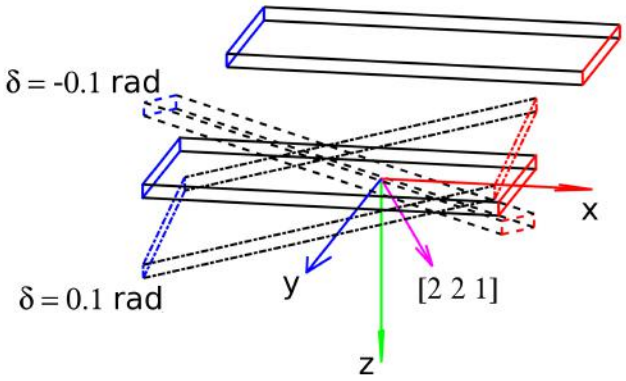


Figure 2.20: The Matlab-generated figure shows the rotation of the bottom grating about the vector [2 2 1]. The geometries defined by the solid lines represent the unrotated top and bottom gratings. The geometries defined by the dash-dot lines and dashed lines represent the bottom grating rotated 0.1 rad and -0.1 rad about [2 2 1], respectively.

with large index contrasts, and it is more computational efficient than FDTD simulations. RCWA-EIS method approximates the interlayer grating coupling efficiency by the product of single grating coupling efficiencies in either the in-coupling or the out-diffraction process due to light reciprocity. The grating optimization is performed by incorporating optimization algorithm with RCWA-EIS method. By using 3D RCWA formulation for conical diffraction, RCWA-EIS method can also be applied to analyze angular (rotational) misalignment of interlayer waveguide gratings. The advantages of RCWA-EIS method include arbitrary choices of periodic grating profiles, wide parameter space search, fast calculation, easy implementation, and accurate results. In summary, this work offers a feasible simulation tool for the design and analysis of basic periodic waveguide grating couplers for optical interconnects in high-density electronics.

CHAPTER 3
GRATING-ASSISTED-CYLINDRICAL-RESONANT-CAVITIES (GARC) INTERLAYER COUPLER: THEORETICAL DESIGN AND SENSITIVITY ANALYSIS

Rectangular diffraction gratings are widely used to achieve interlayer optical coupling. However, various methods, including increasing directionality, reducing substrate leakage, and suppressing backward-reflection/forward-transmission, have been explored, and further improvements in grating performance have been slowing down. As a result, many challenges still exist which obstruct rectangular gratings from serving as efficient and broadband couplers. The basic binary gratings have relatively low efficiencies, e.g. 30% or -5.2 dB [148], without additional features such as overlayers, bottom reflectors, and slanted grating ridges, etc., as can be seen from Table 1.3. Applying these additional features increases the efficiency at the expense of fabrication complexity and cost. For SOI gratings, narrow grating ridges ($\sim$100 nm) are required to achieve relatively high efficiency ($>$70% or -1.5 dB, see Table 1.3), and such gratings can only be patterned using high-cost fabrication techniques, e.g. deep UV lithography or e-beam lithography. In addition, rectangular gratings exhibit narrow spectral bandwidth (e.g. λ_{1dB} = 60 nm and λ_{3dB} = 90 nm), making them less efficient for broadband applications. Apart from the fabrication difficulties and bandwidth limitations, grating assembly poses additional challenges to photonic integrations because rectangular grating couplers are sensitive to both translational [133] and rotational misalignments [51]. All of these issues originate from the Floquet and Bragg conditions which rectangular gratings must inherently follow.

In this chapter, a fundamentally new approach, Grating-Assisted-cylindrical-Resonant-Cavities (GARC) coupler [149, 150], is proposed to achieve efficient and broadband interlayer optical coupling. The coupler consists of a pair of circular gratings with a high-index

via. The via plays a significant role as a cylindrical resonant cavity in the vertical direction, and the GARC coupler thus benefits from the large increase of the resonant optical field introduced by the cavity. In addition, the outer circular gratings function as distributed Bragg reflectors, forming another set of resonant cavities in the horizontal direction and enhancing the fields which are modulated by the inner circular gratings. Since the GARC couplers exploit constructive interference and evanescent coupling instead of diffraction (following the Floquet condition) exhibited in rectangular gratings, grating ridges of GARC can be relatively wide (defined by the in-plane interference) and adjacent-layer reflectors are not necessary (no substrate leakage loss). At 1.55 μm wavelength, the simulated coupling efficiency obtained from an optimized Si/SiO_2 GARC coupler (68% or -1.7 dB interlayer coupling efficiency, see Appendix B for the determination of efficiencies based on FDTD simulation resolutions) are relatively high compared with those obtained from the conventional SOI rectangular gratings (Table 1.3). The most significant benefit of the GARC coupler is the wide spectral bandwidth (λ_{1dB} = 170 nm and λ_{3dB} = 270 nm for an optimized Si/SiO_2 GARC), which can't be achieved by conventional rectangular gratings. The GARC coupler also demonstrates many advantages such as CMOS compatible fabrication, flexible choices of interlayer separation, and high misalignment tolerances.

3.1 Theoretical Design

Figure 3.1 shows two optimized designs based on different materials, Si/SiO_2 and Si_3N_4/SiO_2, in which waveguides and a via (with relatively high refractive indices) are configured to couple optical signals between layers separated by an air gap. This structure can be implemented between two separate overlaid chips or embedded within an on-chip interconnect stack. In the second case, the air gap can be filled with SiO_2, which will cause only small changes to the overall design. Transverse electric (TE, H_z, H_r, H_θ, E_r, E_θ) polarization is considered for demonstration.

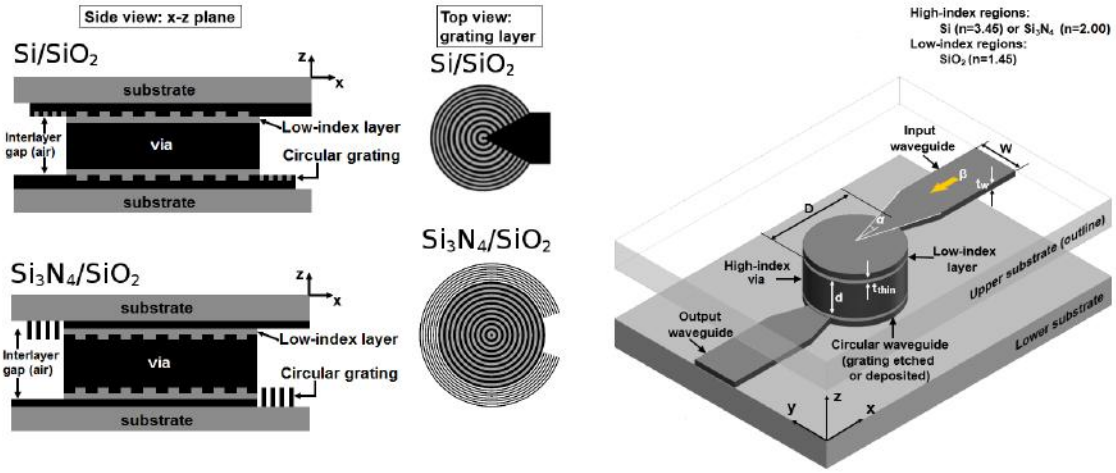


Figure 3.1: Schematic representation of the GARC interlayer coupler. Two types of GARC structures determined by different material systems (Si/SiO_2 and Si_3N_4/SiO_2) are shown.

3.1.1 GARC Coupler Model

The intuition of the GARC coupler design is as follows. The GARC structure has two types of resonant cavities essentially; the cylindrical via is more critical which provides the vertical coupling path, and the in-plane circular gratings assist the cylindrical via by enhancing field in the horizontal direction. Thus the circular grating definition should follow the interference condition of the cylindrical wavefront inside of the circular grating.

The analysis begins by considering a uniform circular slab waveguide in which the TE_z (E_z = 0) polarization is assumed to be supported. According to Maxwell's equations in cylindrical coordinates, the field components E_r, E_θ, H_r, and H_θ can be expressed based on H_z as follows:

$$E_r = -\frac{j\omega\mu_0}{\beta^2}\frac{1}{r}\frac{\partial H_z}{\partial \theta}, \tag{3.1}$$

$$E_\theta = \frac{j\omega\mu_0}{\beta^2}\frac{\partial H_z}{\partial r}, \tag{3.2}$$

$$H_r = \frac{1}{\beta^2}\frac{\partial^2 H_z}{\partial r \partial z}, \tag{3.3}$$

$$H_\theta = \frac{1}{\beta^2}\frac{1}{r}\frac{\partial^2 H_z}{\partial \theta \partial z} \tag{3.4}$$

for TE_z waves, where β, as can be seen later, is the transverse (perpendicular to z) propagation constant of a slab mode, and μ_0 is the permeability of free space (assume the layer is non-magnetic).

In each layer, the longitudinal component H_z, denoted as ψ, satisfies the wave equation

$$(\nabla^2 + \omega^2 \mu_0 \epsilon_0 \epsilon)\psi = 0, \tag{3.5}$$

where the Laplacian operator ∇^2 in cylindrical coordinates is given by

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}. \tag{3.6}$$

Assume the solutions to Eq. (3.5) are separable as follows

$$\psi(r, \theta, z) = H_z = R(r)Z(z)\Theta(\theta), \tag{3.7}$$

where $R(r)$, $Z(z)$, and $\Theta(\theta)$ describe the field distribution in the r, θ, and z directions, respectively.

Substituting Eq. (3.7) into Eq. (3.5) results in the following equations

$$\frac{d^2 Z}{dz^2} + (\omega^2 \mu_0 \epsilon_0 \epsilon - \beta^2)Z = 0, \tag{3.8}$$

$$\frac{d^2 \Theta}{d\theta^2} + \nu^2 \Theta = 0, \tag{3.9}$$

$$r\frac{d}{dr}\left(r\frac{dR}{dr}\right) + [(\beta r)^2 - \nu^2]R = 0, \tag{3.10}$$

where ν and β are constants of separation.

Equation (3.8) is the one-dimensional scalar wave equation describing slab modes in dielectric waveguides (same as Eq. (2.15) in Chapter 2), and the term $\omega^2 \mu_0 \epsilon_0 \epsilon - \beta^2 = \epsilon k_0^2 - \beta^2 = \gamma^2$, where γ is the propagation constant in the z direction of the interested layer. As a result, the function $Z(z)$ should be in the form of $\exp(j\gamma z)$ or $\cos(\gamma z)$. Periodic

condition is satisfied in Θ described by Eq. (3.9) such that $\Theta(\theta + 2\pi) = \Theta(\theta)$. Therefore, the constant ν, called the cylindrical mode order, has to be an integer (e.g. 0, ±1, ±2,...etc). The function $\Theta(\theta)$ should be in the form of $\exp(j\nu\theta)$, and it describes the orbital angular momentum. Equation (3.10) characterizes Bessel's equation of integer order ν, which describes the radial propagation of cylindrical waves. The solution can be Bessel function of first kind $J_\nu(\beta r)$ or Hankel functions $H_\nu^{(1)}(\beta r)$ and $H_\nu^{(2)}(\beta r)$, depending on whether the radial behavior of the field is standing or radially outward-/inward-propagating. If there exists a circular Bragg reflector surrounding the inner circular waveguide, the waveguide field will be standing, expressed by $J_\nu(\beta r)$; if there is no Bragg reflector, the waveguide field will be outward-propagating, expressed by $H_\nu^{(2)}(\beta r)$. The field expression $J_\nu(\beta r)$ is chosen because the GARC coupler has an outer circular grating functioning as a distributed Bragg reflector. As a result, the field expression for the H_z component is as follows:

$$H_z = \sum_{\nu=-\infty}^{\infty} C_\nu J_\nu(\beta r)\cos(\gamma z)\exp(j\nu\theta). \tag{3.11}$$

For the TE polarization, the E_θ field in the circular slab waveguide is of primary importance because it is tangent to the cylindrical wavefront. In the limiting case where the radius of the circular slab is infinite, the cylindrical wavefront becomes planar and the E_θ field can be treated as the E_y field of a typical rectangular slab waveguide that supports TE modes (E_y, H_x, and H_z). The E_θ field is expressed according to Eq. (3.2) and (3.11) as

$$E_\theta = \frac{j\omega\mu_0}{\beta^2}\sum_{\nu=-\infty}^{\infty} C_\nu \frac{d[J_\nu(\beta r)]}{dr}\cos(\gamma z)\exp(j\nu\theta). \tag{3.12}$$

The guided-mode wavelength λ_ν of the νth mode is determined as the difference between two consecutive zeros of the function $d[J_\nu(\beta r)]/dr$, denoted as $J'_\nu(\beta r)$, which describes the radial variation of the E_θ field [151]. As a result, λ_ν is larger at the center and approaches $\lambda_{eff} = 2\pi/\beta_0$ at greater radial distances r, where λ_{eff} and β_0 are the effective wavelength and the transverse propagation constant of a guided mode with a planar wave-

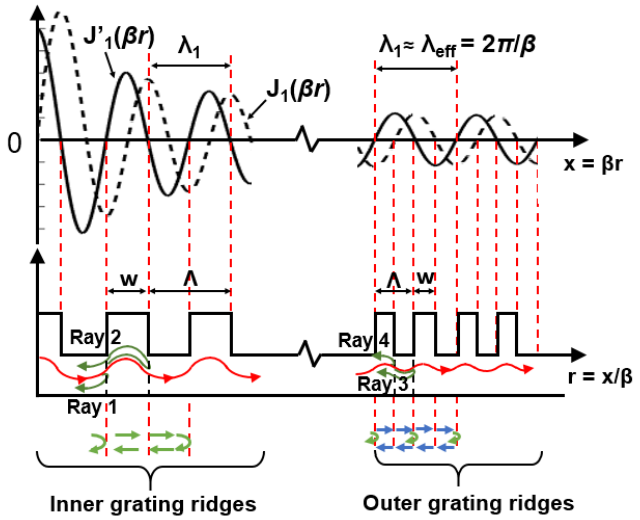


Figure 3.2: Circular grating ridge definition and in-plane ray reflection scheme in a circular grating. The green arrows below the grating indicate a phase change of π, and the blue arrows indicate a phase change of $\pi/2$.

front, respectively.

When the circular grating is present, the guided mode in the circular waveguide will be perturbed but β will not be affected, which differs from the case of a rectangular grating whose ith order horizontal propagation constants β_i are modified by the grating vector magnitude K through the Floquet condition $\beta_i = \beta_0 - iK$, which creates multiple and possibly propagating diffraction orders. This is because the radial fields in the circular region, represented by $J_\nu(\beta r)$ or $J'_\nu(\beta r)$, are not periodic. Since β is invariant and larger than the propagation constant k_{clad} in the low-index cladding, the longitudinal field in the cladding is evanescent, preventing circular gratings from emitting propagating waves and behaving as efficient interlayer couplers. Therefore, a high-index via needs to be incorporated between the two circular gratings to provide an optical path for vertical coupling. The evanescent behavior in the cladding makes it unnecessary to incorporate reflectors because all the propagating light will be concentrated in the high-index via, and thus substrate leakage does not exist.

The circular grating can be divided into two regions. The inner region is designed

for field perturbation, and the outer region serves as an in-plane circular distributed Bragg reflector. The arrangement of grating ridges is based on the ray optics concept depicted in Fig. 3.2, and the phase changes are indicated by the green (phase change of π) and blue (phase change of $\pi/2$) arrows below the grating. Note that the reflected light will experience a phase change of π when it reflects from a medium of higher index and no phase change when it reflects from a medium of lower index. In the inner grating, the radially propagating waves (red rays) constructively interfere if the inner grating periods Λ_i satisfy

$$\Lambda_i = p\lambda_\nu, \tag{3.13}$$

and the reflected rays 1 and 2 will be cancelled if the inner grating ridge widths w_i are

$$w_i = \frac{q}{2}\lambda_\nu, \tag{3.14}$$

where p and q are arbitrary integers. For convenience, q is set equal to p and thus $w_i = \Lambda_i/2$. The outer grating is designed in such a way that the reflected waves (ray 3 and 4) are enhanced, so the outer grating periods Λ_o are defined as

$$\Lambda_o = \frac{s}{2}\lambda_\nu, \tag{3.15}$$

and the outer grating ridge widths w_o need to satisfy

$$w_o = \frac{2t-1}{4}\lambda_\nu, \tag{3.16}$$

where s and t are arbitrary integers. According to the analysis above, the grating periods Λ and the grating ridge widths w need to be defined locally since λ_ν depends on the radial location. As p, q, s and t can be arbitrary integers, the grating ridges can be relatively wide provided the interference conditions are satisfied.

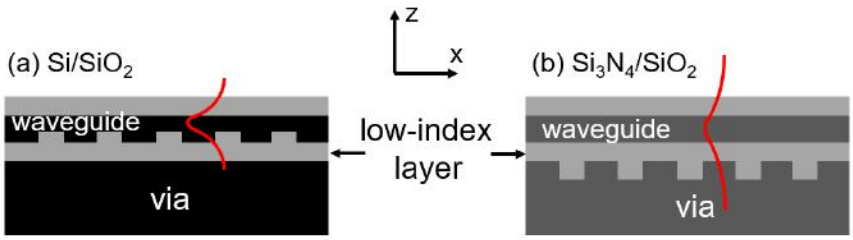


Figure 3.3: Layer stacking sequence and longitudinal field distribution in two types of GARC couplers: (a) Si/SiO_2 and (b) Si_3N_4/SiO_2. Darker color indicates higher index.

The cylindrical mode order $\nu = 1$ is chosen to define λ_ν because of its 180° rotational symmetry. According to Eq. (3.12), the field component $\exp(-j\nu\theta)$ with $\nu = 1$ has one nodal line. This field pattern will promote the field coupling from the input waveguide to the output waveguide with minimal loss along the nodal line. By defining the grating ridges based on λ_1, the other mode orders will be suppressed since only the $\nu = 1$ order satisfies the interference conditions.

A thin layer of a low-index material is sandwiched between the grating and the via to enhance the in-plane resonance. The field in the top input grating can be evanescently coupled to the via and transported to the output grating on the bottom, or vice versa. The presence of the thin layer impedes the direct propagation from the high-index grating ridges to the via and forms another resonant cavity along the z direction. This effectively allows more time for the light to interact with the grating and further suppresses modes other than the $\nu = 1$ order. Since the light is initially launched from a rectangular slab waveguide, a waveguide taper is used to gradually convert the plane wavefront in the rectangular waveguide to the cylindrical wavefront in the circular waveguide. The structure is less prone to misalignment because it doesn't follow the "sensitive" Floquet condition.

The stacking sequence of the GARC coupler is determined by the materials used. When the high-index regions (darker regions in Fig. 3.1) are made of Si, the grating is etched into the Si circular waveguide and filled with SiO_2. The low-index layer, made of SiO_2, is sandwiched between the grating and the via. Whereas in the Si_3N_4/SiO_2 GARC

coupler, the grating is etched into the via and the low-index layer is located between the waveguide and the grating. This is because that the longitudinal field of the Si/SiO_2 waveguide is more confined in the waveguide core due to a larger refractive index difference (Fig. 3.3). In order to effectively perturb the field in the circular waveguide, the grating needs to be etched into the waveguide. The etch depth should be shallow to reduce back-reflection into the input taper. By contrast, the longitudinal field extends further out of the waveguide core of the Si_3N_4/SiO_2 waveguide, and thus a grating layer adjacent to the waveguide is sufficient to perturb the field, while an etched grating may induce too much back-reflection.

GARC couplers for the transverse magnetic (TM_z, $H_z = 0$) polarization follow the same design process except that the TM guided-mode propagation constant β_{TM} is used (e.g.for a 0.22 μm thick air/Si/SiO_2 waveguide, β_{TE} = 11.371 μm^{-1} for the TE polarization whereas β_{TM} = 7.5996 μm^{-1} for the TM polarization), and the relative permittivity is introduced to determine the grating ridges. For TM waves, the field components E_r, E_θ, H_r, and H_θ is related to E_z as

$$E_r = \frac{1}{\beta^2}\frac{\partial^2 E_z}{\partial z \partial r}, \tag{3.17}$$

$$E_\theta = \frac{1}{\beta^2}\frac{1}{r}\frac{\partial^2 E_z}{\partial \theta \partial z}, \tag{3.18}$$

$$H_r = \frac{j\omega\epsilon_0\epsilon}{\beta^2}\frac{1}{r}\frac{\partial E_z}{\partial \theta}, \tag{3.19}$$

$$H_\theta = -\frac{j\omega\epsilon_0\epsilon}{\beta^2}\frac{\partial E_z}{\partial r}. \tag{3.20}$$

The component H_θ should be used to determine the grating ridges, and its formulation contains the permittivity, which is a location dependent variable.

3.1.2 Optimized Configurations

Figure 3.4 shows the cross-sectional views of the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers. The grating ridge definitions for the Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers are

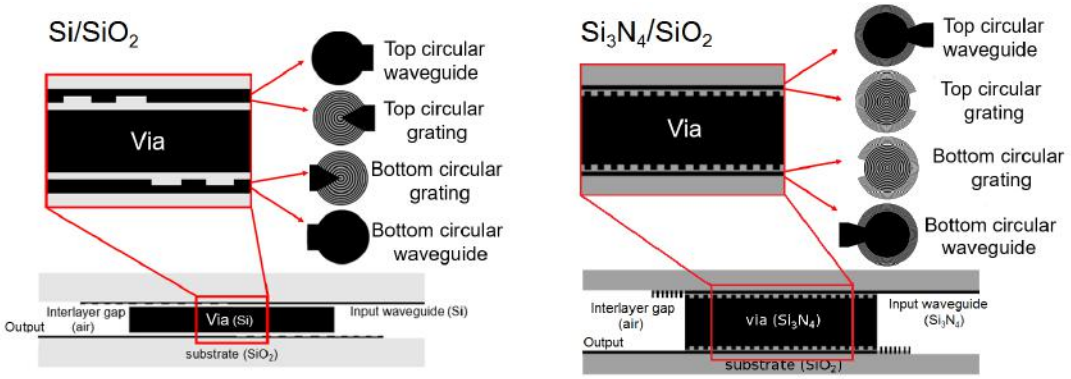


Figure 3.4: Cross-sectional views of the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers.

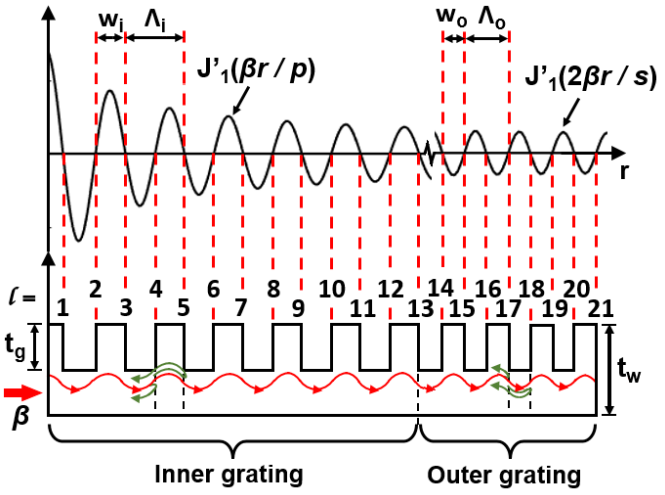


Figure 3.5: Circular grating ridge definition for the optimized Si/SiO_2 GARC coupler with parameters $p = 2$, $s = 3$, and $t = 2$.

schematically demonstrated in Fig. 3.5 and Fig. 3.2, respectively. Some key parameters are summarized and compared in Table 3.1. Details of the Si/SiO_2 and Si_3N_4/SiO_2 GARC designs are shown in Table 3.2 and Table 3.3, respectively.

Full-wave simulations were performed using MEEP 3D FDTD (see Appendix B for general FDTD formulations). The field patterns of the H_z component at different planes (waveguide, grating, and vertical cross-section) of the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers are shown in Fig. 3.6. For both cases, there is a noticeable amount of desctructive interference in the top waveguide (input taper) due to the back-reflection, and multimodes are observed in both the rectangular and circular slab waveguides. If needed, higher-order modes in the GARC output slab waveguide can be suppressed by tapering the slab waveguide to a ridge waveguide, as illustrated in Fig. 3.7. This is particularly important in photonic circuits that use ridge waveguides which are evanescently coupled to ring resonators or interferometers, etc. As shown in the vertical cross-sections, there is no propagating field in the substrate near the inner grating regions, but diffracted light is observed near the outer grating regions, which accounts for the power loss. Another source of the power loss is the field reflected at possible boundaries, e.g. wall of the cylindrical via and distributed Bragg gratings, and field confined in the resonant cavities.

In some cases, diffraction occurs in the outer gratings of GARC couplers because the outer grating periods are approximately equal, as shown in Table 3.2 and Table 3.3. The Bessel function $J_1(\beta r)$, on which the circular gratings are based, has an oscillation period larger near the origin and smaller at larger radial distances r. At larger argument r, the Bessel function $J_1(\beta r)$ can be approximated as the periodic function $\cos(\beta r)$, and circular grating in this region is diffractive. This phenomenon will become more obvious when the inner grating is small and the majority of the input field is forward transmitted to the

Table 3.1: Parameter values for the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers.

Case	β [μm^{-1}]	n_{high}	t_w [μm]	t_g [μm]	t_{thin} [μm]	D_{in} [μm]	D_{out} [μm]	W [μm]	α [$^\circ$]	d [μm]	p	s	t	η_c [%]
Si/SiO_2	11.37	3.45	0.22	0.04	0.1	14.08	20.72	10	28.65	2	2	3	2	68
Si_3N_4/SiO_2	6.76	2	0.34	0.34	0.1	19.28	25.79	10	20	4.7	1	1	1	41

Table 3.2: Parameters for the optimized Si/SiO_2 GARC coupler.

Inner and Outer Circular Grating Ridge Edges					
l	ω_l [a]	r [b] [μm]	w_r [c] [μm]	w_v [d] [μm]	Λ [e] [μm]
1	1.8412	0.3238	0.3238	–	–
2	5.3314	0.9377	–	0.6139	–
3	8.5363	1.5014	0.5637	–	1.1776
4	11.7060	2.0589	–	0.5575	–
5	14.8636	2.6143	0.5554	–	1.1129
6	18.0155	3.1687	–	0.5544	–
7	21.1644	3.7225	0.5538	–	1.1082
8	24.3113	4.2760	–	0.5535	–
9	27.4571	4.8293	0.5533	–	1.1068
10	30.6019	5.3824	–	0.5531	–
11	33.7462	5.9355	0.5530	–	1.1061
12	36.8900	6.4884	–	0.5530	–
13	40.0334	7.0413	0.5529	–	1.1059
14	43.1766	7.4559	–	0.4146	–
15	46.3196	7.8706	0.4146	–	0.8292
16	49.4624	8.2851	–	0.4146	–
17	52.6050	8.6997	0.4146	–	0.8292
18	55.7476	9.1142	–	0.4145	–
19	58.8900	9.5288	0.4145	–	0.8290
20	62.0323	9.9433	–	0.4145	–
21	65.1746	10.3578	0.4145	–	0.8290
Other Parameters					
β [f] [μm^{-1}]		11.37	λ_0 [g] [μm]		1.55
k_0 [μm^{-1}]		4.05	n_{SiO_2} [h]		1.45
n_{Si} [h]		3.45	γ_{Si} [i] [μm^{-1}]		8.12
t_w [μm]		0.22	t_g [μm]		0.04
t_{thin} [μm]		0.1	W [μm]		10
ϕ [°]		35.5	α [j] [°]		28.65
D_{inner} [k] [μm]		14.08	D_{outer} [k] [μm]		20.72
p, s, t [l]		2, 3, 2	d [μm]		2
N_{inner} [m]		6	N_{outer} [m]		4

[a] lth zero of $J_1'(x) = d[J_1(x)]/dx$
[b] Location of grating ridge edges as shown in Fig. 3.5
[c] Grating ridge width [d] Grating groove width
[e] Local grating period [f] Radial (r) propagation constant
[g] Free-space wavelength [h] n represents refractive index
[i] Longitudinal (z) propagation constant
[j] Waveguide taper half-angle (within circular gratings)
[k] Inner and outer grating diameters
[l] Parameters used to define circular grating ridges [149]
[m] Number of grating periods of the inner and outer gratings

Table 3.3: Parameters for the optimized Si_3N_4/SiO_2 GARC coupler.

Inner and Outer Circular Grating Ridge Edges					
l	ω_l [a]	r [b] [μm]	w_r [c] [μm]	w_v [d] [μm]	Λ [e] [μm]
1	1.8412	0.2723	0.2723	–	–
2	5.3314	0.7886	–	0.5162	–
3	8.5363	1.2626	0.4740	–	0.9902
4	11.7060	1.7314	–	0.4688	–
5	14.8636	2.1985	0.4670	–	0.9358
6	18.0155	2.6647	–	0.4662	–
7	21.1644	3.1304	0.4658	–	0.9320
8	24.3113	3.5959	–	0.4655	–
9	27.4571	4.0612	0.4653	–	0.9308
10	30.6019	4.5263	–	0.4651	–
11	33.7462	4.9914	0.4650	–	0.9301
12	36.8900	5.4564	–	0.4650	–
13	40.0334	5.9213	0.4649	–	0.9299
14	43.1766	6.3862	–	0.4649	–
15	46.3196	6.8511	0.4649	–	0.9298
16	49.4624	7.3159	–	0.4648	–
17	52.6050	7.7808	0.4648	–	0.9296
18	55.7476	8.2456	–	0.4648	–
19	58.8900	8.7104	0.4648	–	0.9296
20	62.0323	9.1752	–	0.4648	–
21	65.1746	9.6399	0.4648	–	0.9296
22	68.3168	9.8723	–	0.2324	–
23	71.4590	10.1047	0.2324	–	0.4648
24	74.6011	10.3371	–	0.2324	–
25	77.7432	10.5695	0.2324	–	0.4648
26	80.8852	10.8019	–	0.2324	–
27	84.0272	11.0343	0.2324	–	0.4648
28	87.1692	11.2667	–	0.2324	–
29	90.3111	11.4991	0.2324	–	0.4648
30	93.4530	11.7315	–	0.2324	–
31	96.5949	11.9639	0.2324	–	0.4648
32	99.7368	12.1963	–	0.2324	–
33	102.8787	12.4287	0.2324	–	0.4648
34	106.0205	12.6610	–	0.2324	–
35	109.1623	12.8934	0.2324	–	0.4648

Other Parameters			
β [μm^{-1}]	6.76	λ_0 [μm]	1.55
k_0 [μm^{-1}]	4.05	n_{SiO_2}	1.45
$n_{Si_3N_4}$	2	$\gamma_{Si_3N_4}$ [μm^{-1}]	4.47
t_w [μm]	0.34	t_g [μm]	0.34
t_{thin} [μm]	0.1	W [μm]	10
ϕ [°]	33.5	α [°]	20
D_{inner} [μm]	19.28	D_{outer} [μm]	25.79
p, s, t	1, 1, 1	d [μm]	4.7
N_{inner}	10	N_{outer}	7

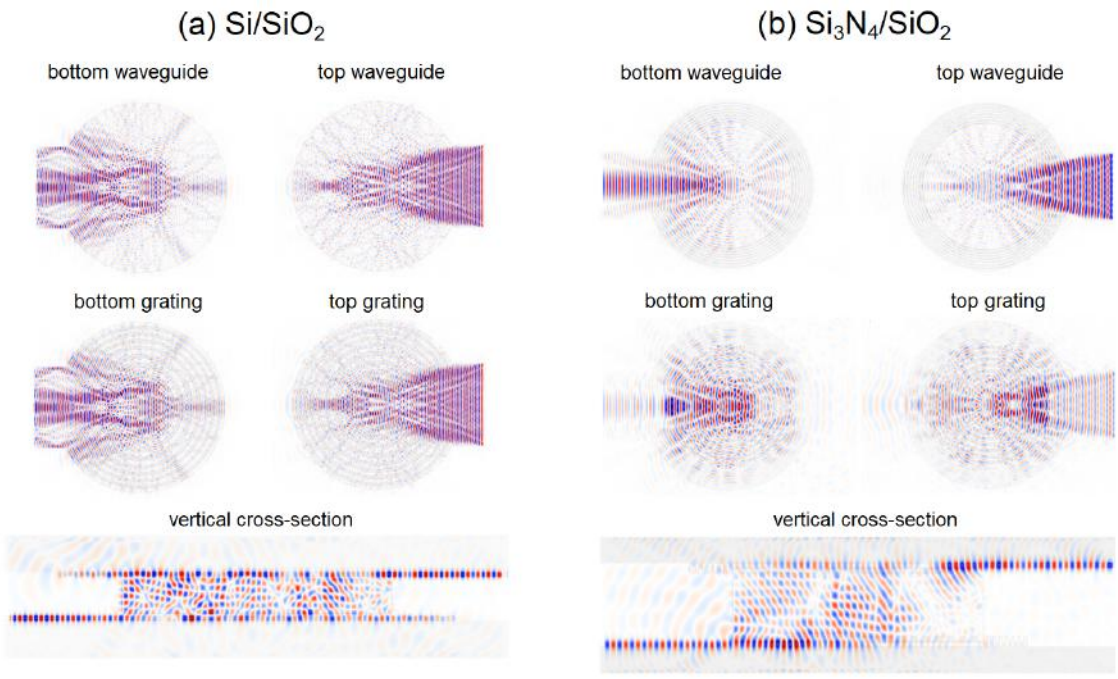


Figure 3.6: H_z field patterns of the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers simulated by MEEP 3D FDTD.

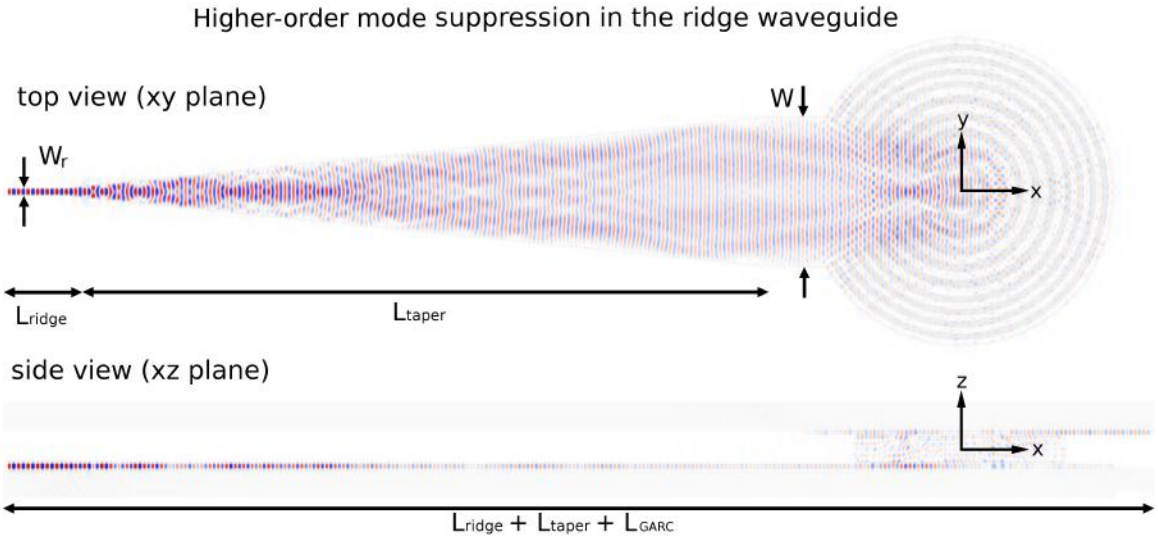


Figure 3.7: Waveguide taper is added to coupler the GARC output slab waveguide to a ridge waveguide. Color is scaled based on the maximum power in the simulation.

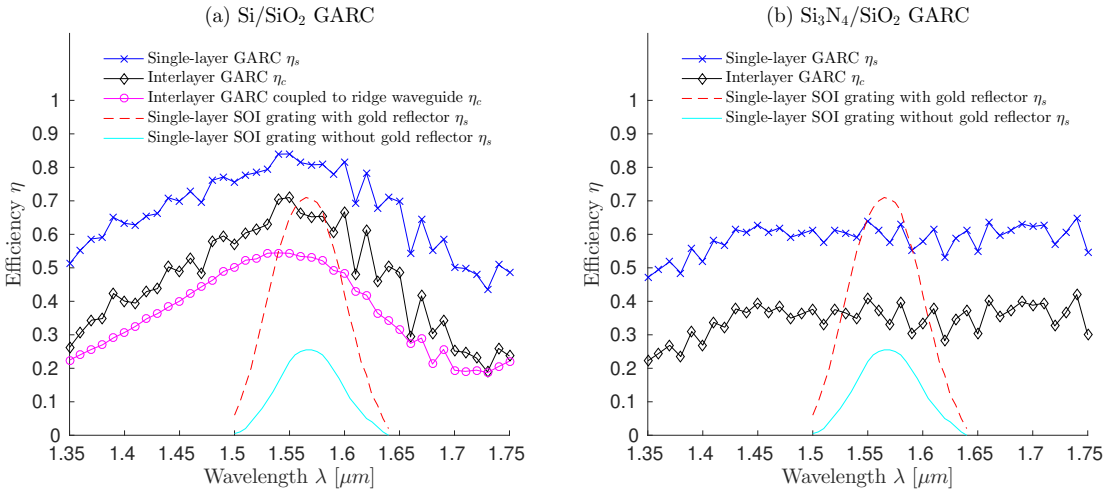


Figure 3.8: Spectral response for the optimized Si/SiO_2 and Si_3N_4/SiO_2 GARC couplers as well as that for the rectangular grating reported in [148].

outer grating. If the inner grating is properly designed such that most of the input field is allowed to be coupled to the high-index interconnecting via, there will be minimal amount of forward-transmitted field reaching the diffractive outer grating. The method to define circular gratings based on Bessel functions does not contradict common circular grating designs used as laser cavities, e.g. [152, 153, 154], etc., in which strictly periodic circular gratings are used. For example, in [152], the laser cavity consists of a uniform disk in the center (radius 5 μm) surrounded by circular Bragg reflector (from r = 5 μm to 60 μm). The central disk, where the pump medium is located, circumvents the necessity of grating definition, while the radial locations for the circular Bragg gratings are large enough such that the gratings can be treated as periodic.

The Si/SiO_2 GARC coupler has a higher interlayer coupling efficiency (η_c = 68% or -1.7 dB) than the Si_3N_4/SiO_2 GARC coupler (η_c = 41% or -3.9 dB) due to a larger refractive index contrast and a stronger field confinement in the cavity. The circular grating ridges are defined according to the interference conditions Eq.(3.13) to (3.16) using different sets of parameters p, s and t, such that the minimum w_i of the Si/SiO_2 GARC coupler is 553 nm and that of the Si_3N_4/SiO_2 is 465 nm. Both GARC couplers exhibit wide spectral responses, e.g. λ_{3dB} = 270 nm and 400 nm for the interlayer coupling efficiency (η_c) of the

Si/SiO_2 and Si_3N_4/SiO_2 GARC coupler respectively, as shown in Fig. 3.8. The single-layer efficiencies η_s are approximated as the square root of the interlayer coupling efficiencies due to the reciprocity nature of the GARC couplers, and the curves of the single-layer coupling efficiencies are shown to compare the performance of the GARC couplers with that of the rectangular grating couplers reported in [148]. Strictly speaking, the GARC structure can't be separated into functional top and bottom gratings because it requires the presence of the cylindrical via to provide an optical path, and thus the notion of the single-layer efficiency is meaningless and it is only used for comparisons. Higher-order modes in the transverse direction (xy plane) will typically be excited in the slab waveguides. To suppress the higher-order modes in the transverse direction, a waveguide taper (not to be confused with the taper within the circular grating) can be added to gradually narrow the width (W) of the slab waveguide down to that of the ridge waveguide (W_r). For the Si/SiO_2 GARC coupler, a waveguide taper with a taper half-angle 0.1 rad can be used to couple the slab modes into a ridge waveguide with a width of W_r = 0.5 μm (Fig. 3.7), and the resulting spectral response shown in Fig. 3.8 has correspondingly fewer oscillations and reduced efficiency due to the removal of higher-order modes. The coupling efficiency could be further improved by optimizing the taper angle and length.

The diameter of the inner circular grating D_{inner}, the width of the rectangular slab waveguide W, and the taper half-angle α (the taper within the circular gratings) affect the determination of the via height d or equivalently, the interlayer separation. The via diameter is the same as D_{inner}. Figure 3.9 shows the H_z field of the Si/SiO_2 GARC coupler with five different taper half-angles. The angle 0.5 rad is the limiting case in which the taper side length equals $R_{outer} = D_{outer}/2$. The parameters d = 2 μm, D_{inner} = 14.08 μm, and W = 10 μm are the same for all cases. The smaller the taper half-angle, the longer the input taper (the entire taper is not shown for α = 0.1 rad to 0.3 rad). It is observed that the "destructive interference region (DIR)" (indicated by the dashed arrow), the null field in the input taper due to destructive interference of input field and back-reflected field, moves in-

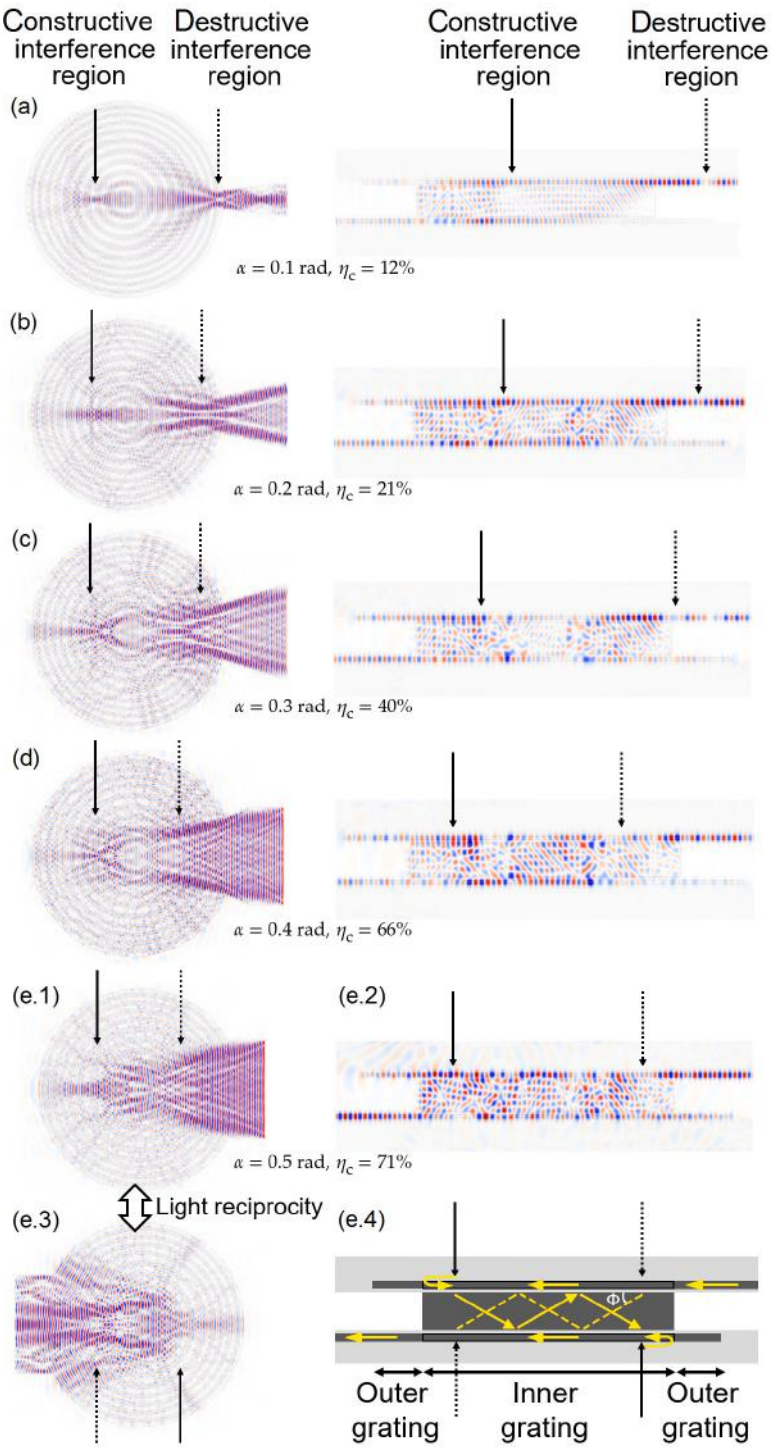


Figure 3.9: The H_z field distribution in the top grating and in the vertical cross-section of the Si/SiO_2 GARC coupler with five different taper half-angles: (a) 0.1 rad = 5.73°, (b) 0.2 rad = 11.46°, (c) 0.3 rad = 17.19°, (d) 0.4 rad = 22.92°, and (e) 0.5 rad = 28.65°. For case (e), the H_z field distribution in the bottom grating is shown in (e.3), and the ray representation of the vertical resonator in the via is shown in (e.4).

wards from the outskirts of the circular grating as taper half-angle increases. Once the input field reaches the inner grating, the field will be evanescently coupled to the high-index via and propagate at the angle $\phi = \arctan(\gamma_{Si}/\beta) = 35.5°$. Note that not all of the field is coupled into the via; instead, a portion of the input field still propagates in the slab waveguide with an approximately exponential-decaying intensity due to the presence of the low-index layer. The obliquely propagating field in the via will be reflected from the bottom circular grating and then obliquely propagate upward with the angle $\pi - \phi$. Eventually, most of the reflected field will constructively interfere at the top grating, forming a "constructive interference region (CIR)" indicated by the solid arrow. It is desired that the CIR and DIR be located symmetrically about the grating center and both regions be located within the inner grating, e.g. Fig. 3.9 (e.1). This symmetry ensures that another set of CIR and DIR, whose locations are interchanged, are formed at the bottom grating, e.g. Fig. 3.9 (e.3). In other words, inversion symmetry about the center of the via has been achieved and the structure exhibits reciprocal behavior. Thus, α and W determine the location of the DIR; D_{inner} restricts the location of the CIR and DIR; and d controls the location of the CIR and the effectiveness of the vertical resonator.

3.2 Sensitivity Analysis

The optimized couplers, whose components are assumed to be exactly shaped and aligned in the simulation, may not be fabricated and assembled precisely as designed. Here the effects of variations in both vertical and horizontal dimensions are analyzed using 3D FDTD. The vertical variations include changes in via height, SiO_2 layer thickness, gap between the SiO_2 layer and the via, and grating etch depth; the horizontal variations include changes in outer and inner grating ridge widths, horizontal shifts, and via tapering. This sensitivity analysis will offer insights on experimental designs and possible outcomes for GARC coupler fabrications.

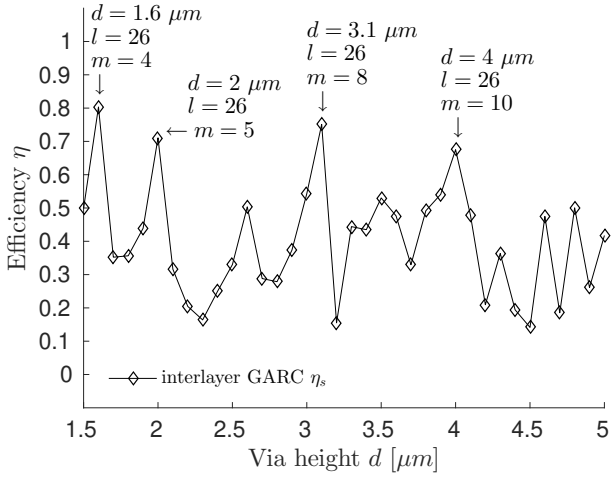


Figure 3.10: Plot of GARC interlayer coupling efficiency as a function of via height.

Table 3.4: Resulting via height d in μm for the (l,m)th resonance at 1.55 μm free-space wavelength (R = 7.0413 μm).

m \ l	24	25	**26**	27	28
4	1.3801	1.4682	**1.5807**	1.7302	1.9413
5	1.7251	1.8353	**1.9758**	2.1627	2.4266
6	2.0701	2.2023	2.3710	2.5952	2.9119
7	2.4151	2.5694	2.7662	3.0278	3.3972
8	2.7602	2.9364	**3.1613**	3.4603	3.8825
9	3.1052	3.3035	3.5565	3.8929	4.3678
10	3.4502	3.6705	**3.9516**	4.3254	4.8531

The effect of varying via height d on the GARC coupler efficiency is shown in Fig. 3.10. It is observed that the efficiency is sensitive to via height changes. Nevertheless, the layer thickness could be closely controlled during the deposition process and the optimal via height could be achieved without significant difficulty. Since the via functions as a cylindrical resonant cavity, the resonant wavelengths in the via are [155]

$$\lambda_{lm} = \frac{2\pi}{\sqrt{\left(\frac{\omega_l}{R}\right)^2 + \left(\frac{m\pi}{d}\right)^2}}, \tag{3.21}$$

where l and m indicate the resonant mode order in the radial (r) and longitudinal (z) directions, respectively, ω_l is the lth zero of the function $J_1'(x)$, R = 7.0413 μm is the radius

of the via, and d is the via height. Since the GARC coupler is designed for the free-space wavelength λ_0 = 1.55 μm, the resonance wavelength in the Si via is $\lambda_{lm} = \lambda_0/n_{Si}$, which can be achieved using various sets of l, m, and d. The via height d at the (l,m)th resonance can be found from Eq. (3.21), and some results are shown in Table 3.4. Since the radius of the cylindrical cavity is fixed at R = 7.0413 μm, the resonance mode order l in the radial direction can be approximated as

$$R = \frac{l}{2}\lambda_r, \tag{3.22}$$

where λ_r is the radial component of λ_{lm} ($\lambda_r = 2\pi/\beta$ = 0.5526 μm). Thus l is approximated as 26, resulting in the most prominent resonances in the cavity. The highlighted d values in Table 3.4, which are calculated using l = 26, correspond to the peaks in Fig. 3.10. Even though the via heights 1.6 μm and 3.1 μm lead to higher coupling efficiency (80% and 75%, respectively), the via height 2 μm is chosen in consideration of fabrication feasibility and device performance. On one hand, it may be challenging to deposit a Si via of more than 2 μm thick using conventional LPCVD without sacrificing deposition speed and layer uniformity. On the other hand, interlayer separations less than 2 μm may induce optical crosstalk between two layers, thus degrading the overall 3D interconnection. The case d = 2 μm has a longer coupling time (1.32 ps, temporal bandwidth δ_t = 0.76 THz) than d = 3.1 μm (0.83 ps, δ_t = 1.21 THz) due to a stronger resonance. Nevertheless, the GARC couplers would be fast enough to satisfy the terahertz communication requirements.

Keeping the interlayer separation d at 2 μm, the effect of changing the thickness of the SiO_2 thin layer, t_{thin}, is shown in Fig. 3.11. The resonance in the via is weaker for smaller t_{thin}, while evanescent coupling is prohibited for larger t_{thin}, both resulting in lower coupling efficiencies. The plateau regions may due to the limited resolution used in the FDTD simulation which can't take into account small distance adjustments.

In the case of coupling between layers separated by an air gap, the GARC coupler may be divided into two parts for easier fabrication, one being the top waveguide and the via, the other being the bottom waveguide. Here the effect of vertical gaps between the two

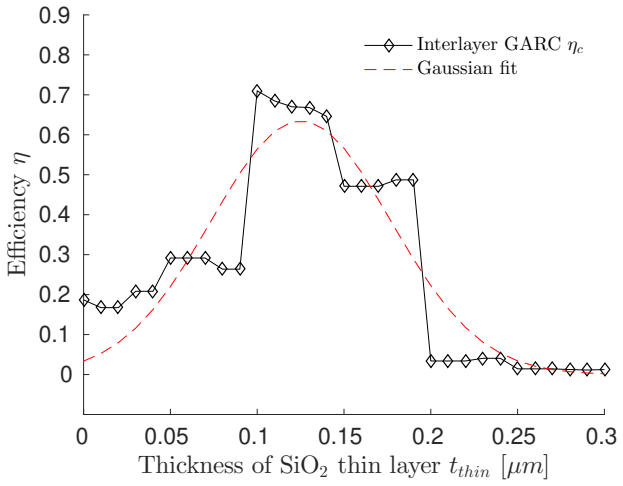


Figure 3.11: Effect of varying SiO_2 thin layer thickness t_{thin} on the optimized Si/SiO_2 GARC coupler with $d = 2$ μm.

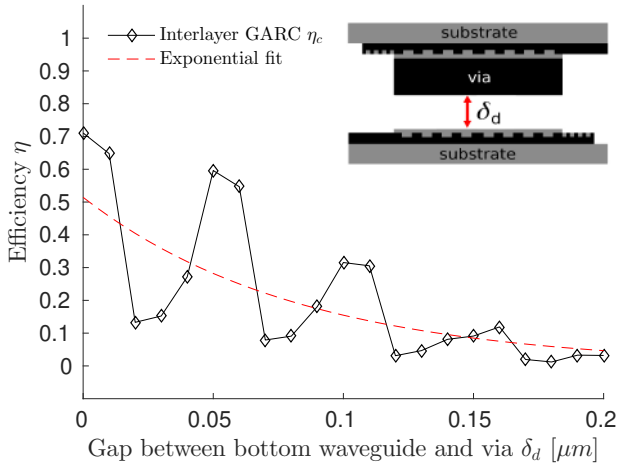


Figure 3.12: Effect of varying vertical gap between via and bottom waveguide δ_d on the optimized Si/SiO_2 GARC coupler with target interlayer separation $d = 2$ μm.

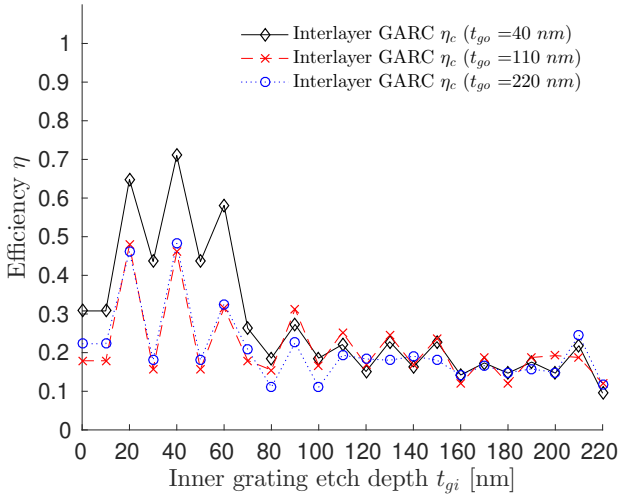


Figure 3.13: Effect of varying the inner grating etch depth t_{gi} and the outer grating etch depth t_{go} on the optimized Si/SiO_2 GARC coupler with d = 2 μm.

parts is analyzed, assuming the via height is exactly 2 μm as designed. As shown in Fig. 3.12, the efficiency drops in an oscillating and exponentially decaying fashion. The irregular efficiency drops may due to the resolution problem discussed in Appendix B. Since the power in the via is evanescently coupled to the bottom waveguide, it is expected that a large gap, e.g. δ_d >200 nm, will prevent the coupling. From this analysis, we can see that the gap between layers of GARC coupler is detrimental to the coupling and it should be avoided. Therefore, ideally, the GARC structure should be grown layer by layer in a bottom-up approach.

To reduce fabrication difficulty, the inner and outer grating etch depths are kept the same; that is, $t_{gi} = t_{go} = t_g$. For analysis purposes, varying the etch depths of the inner and outer grating separately offers a better understanding of the effects of circular gratings and the resonant cavity. Figure 3.13 shows the efficiency change with respect to the variation of inner grating etch depth t_{gi} for three outer grating etch depths t_{go} = 40 nm, 110 nm, and 220 nm. At t_{gi} = 0, the inner circular grating is absent and the interlayer coupling efficiency is 20% to 30%. This is because the field in the inner circular slab waveguide, which is launched from the input taper, is not modulated and the field pattern is not sym-

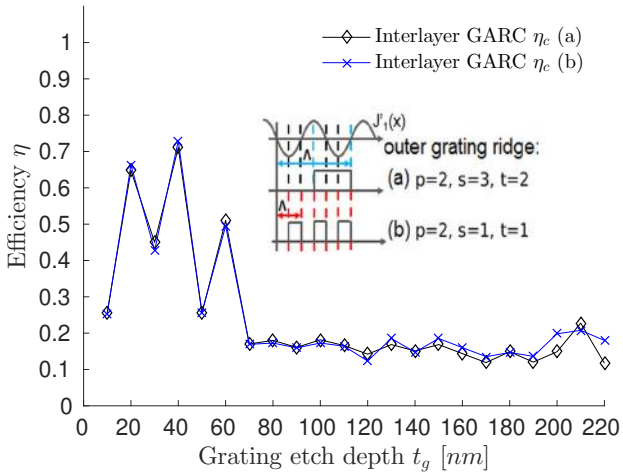


Figure 3.14: Effect of varying grating etch depth t_g on the optimized Si/SiO_2 GARC coupler with d = 2 μm. t_g indicates the etch depth for both the inner and outer circular gratings.

metric about the circular waveguide center. As a result of the asymmetry, the field does not constructively interfere, even though the outer circular grating forms a resonant cavity. As the inner grating depth increases, the index contrast becomes larger, and so does the field modulation. The curves for the three outer grating depths have similar trends, but coupling losses are higher for larger t_{go} possibly due to stronger back-reflection. The highest coupling efficiency for the three cases t_{go} = 40 nm, 110 nm, and 220 nm are all obtained at t_{gi} = 40 nm. It can be concluded that, first, the shallow etch depth is necessary to achieve high coupling efficiency, and second, there is no benefit in defining different etch depths for the inner and outer circular gratings.

The effect of changing the grating etch depth t_g and changing the outer grating ridge definition (values of s and t) are also explored. As shown in Fig. 3.14, the efficiencies drop rapidly for $t_g > 80$ nm due to significant back-reflection from the circular grating into the input waveguide taper. The oscillations in the efficiencies are due to the effect of interference and simulation resolution. The efficiency plots of the GARC coupler with s = 3 and t = 2 coincide with those with s = 1 and t = 1, which means that the size of the inner grating is properly designed such that most of the input field propagates into the via and

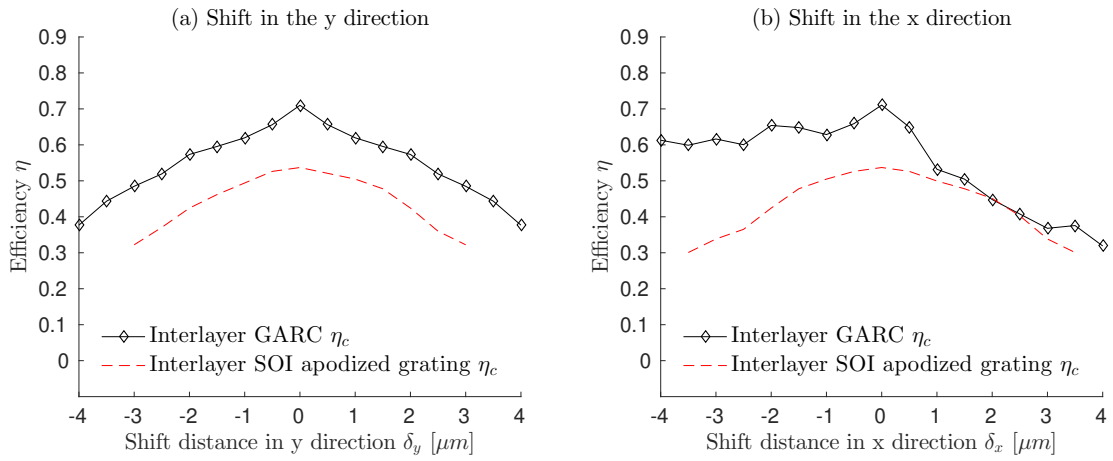


Figure 3.15: Effect of translational shift in the y direction (a) and the x direction (b) on the optimized Si/SiO_2 GARC coupler specified in Table 3.2 and that of an apodized SOI grating coupler reported in [136].

there are minimal forward-transmitted field reaching the outer grating region. Thus, for this case, it is possible to remove the outer grating region to reduce the size of the GARC coupler without severe degradation of the coupling efficiency. Nevertheless, the current design includes the outer grating, and the values s = 3 and t = 2 are chosen to define the outer circular grating because they result in wider grating ridges.

Figure 3.15(a) shows the effect of translational shifts of the bottom waveguide with respect to the via in the y direction (the axes are shown in Fig. 3.1), assuming the circular gratings are exactly defined in the circular waveguides on both layers and the top circular waveguide is exactly aligned to the via. Since the GARC coupler is symmetric about the xz plane, the efficiency changes for the positive and negative shifts are identical. A misalignment tolerance of ± 2 μm causes about 1 dB excess loss for the the GARC interlayer coupling efficiency, which is about the same amount of tolerance as for an SOI apodized grating coupler reported in [136]. However, the layers of the GARC coupler are patterned lithographically, which can be aligned more accurately than the post assembly of rectangular gratings using flip-chip bonder or active alignment, etc. The effect of shift in x direction is shown in Fig. 3.15(b). Shifting in the positive x direction has a larger effect on the efficiency than shifting in the negative x direction, which can be explained with the illustration

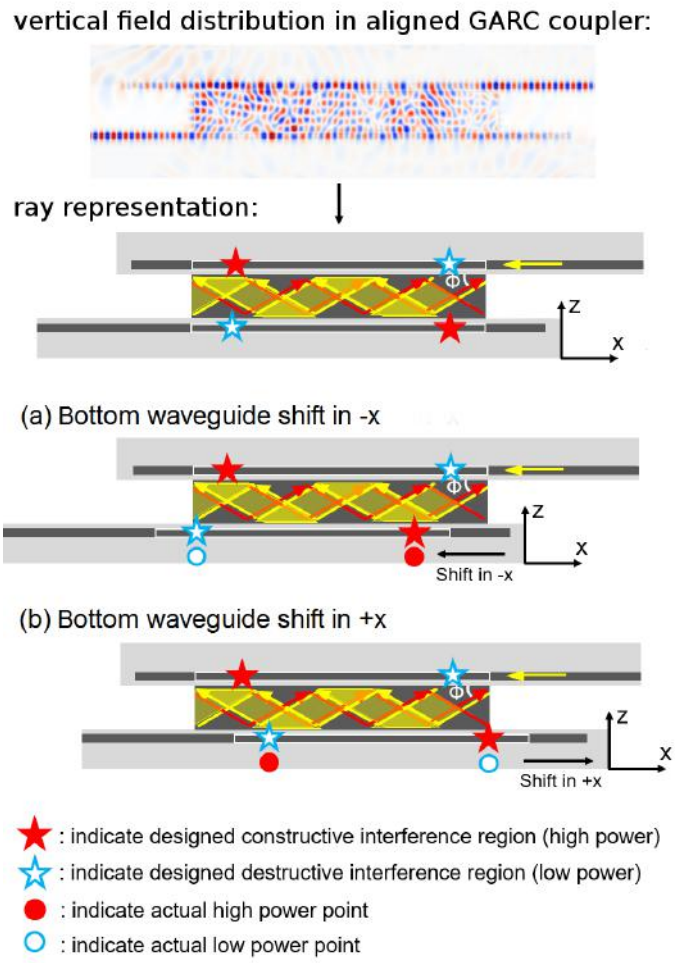


Figure 3.16: Effect of translational shift in the x direction on coupler performance.

in Fig. 3.16. The shaded yellow region indicates a relatively high power region in the via. When the bottom waveguide is shifted in the $-x$ direction, high-power region and low-power region on the bottom waveguide are approximately located at the pre-designed CIR and DIR, respectively. By contrast, the high-power region and low-power region interchange their locations when the bottom waveguide is shifted in the $+x$ direction, which violates the light reciprocity and thus induces a relatively larger power loss. Rotational misalignment, which is particularly detrimental to the rectangular gratings, is not likely to occur in the GARC structure due to the presence of the via which offers a mechanical support between layers.

Furthermore, varying the grating ridge width is another factor to consider. In the simulation, ridge locations with 4 decimal place accuracy are employed even though the simulation resolution may not handle such level of accuracy. The high level accuracy is neither realistic in fabrication. Keeping the other parameters unchanged, a Si/SiO_2 GARC coupler with a strictly periodic 0.55 μm inner grating ridge/groove width and 0.41 μm outer grating ridge/groove width is simulated, and the interlayer coupling efficiency is simulated as 65%, which is 4% smaller than the optimized GARC coupler specified in Table 3.2. This is due to the fact that the uniform inner grating doesn't follow the radial field distribution described by the 1st order Bessel function, and the resulting via radius doesn't fully satisfy the resonant condition Eq. 3.21. Although it is common practice to use a strictly periodic grating, the Bessel-function-defined periods are shown to provide better performance.

Tapering effect of the via is also investigated. During the etching process, the side wall of the cylindrical via may not be exactly vertical. The effect of via radius change δ_r is shown in Fig. 3.17. Two scenarios are investigated: the via may be either tapering up or tapering down depending on the fabrication process. In either case, the narrow end of the via has a change of δ_r (a negative value) compared with the wide end whose radius is set as a fixed value R = 7.0413 μm. The plots are symmetric about the axis at δ_r = 0, which indicates a perfectly vertical side wall. This is reasonable because of the 180° rotational

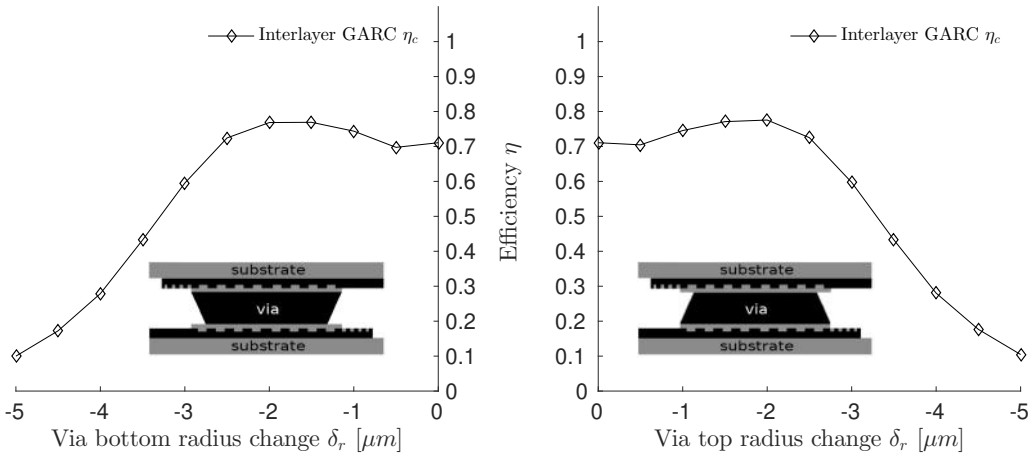


Figure 3.17: Effect of via tapering on the optimized Si/SiO_2 GARC coupler with $d = 2$ μm.

symmetry of the GARC structure and the light reciprocity. The efficiencies exhibit moderate increase for $|\delta_r| < 2$ μm (28% change relative to R), which may due to a stronger field confinement at the narrow taper end. However, further radius deductions will violate the radial resonance condition in the via, thus inducing large coupling loss. As a result, tapering effect should be controlled within 2 μm radius deviations. Nevertheless, a tapering effect with more than 28% radius decrease is unlikely to occur.

Based on these analysis, it is found that the vertical variations have larger effects on GARC coupler performance. This is because the most critical resonance is within the cylindrical via along the vertical direction, and the horizontal resonant cavities assist the vertical coupling by enhancing the field with 180° symmetry. In addition, the vertical dimension (2 μm) is much smaller than the horizontal dimension (20 μm), and thus the same amount of variation will have larger effect on the vertical dimension.

Lastly, the GARC coupler is polarization dependent. The optimized GARC coupler for the TE polarization is not effective for coupling a TM polarized guided mode, and the coupling efficiency for TM polarization is $\eta_c = 20\%$. This is because the interference condition designed for the TE polarization doesn't satisfy the TM guided mode. On one hand, the radial propagation constants for the two polarizations are not the same, resulting different

interference condition in the radial direction. On the other hand, the longitudinal propagation constants $\gamma_{Si} = \sqrt{k_0^2 n_{Si}^2 - \beta^2}$ are also different, causing different resonant conditions in the vertical cavity.

3.3 GARC Coupler Design Flow

It is noted that high-efficiency GARC couplers are not only limited to the presented designs. By properly choosing the sizes of the resonant cavities, various interlayer distances d and coupler radii R can be achieved. A design flowchart is provided in Fig. 3.18 to accommodate various potential structure requirements. The zeros of Bessel function $J_1'(x)$ and the fundamental propagation constant β can be pre-calculated. Given the minimum grating period Λ_{min}, the parameters p, s, and t, which correspond to grating ridge widths, can be determined. The targeted interlayer distance and via radius are set as the initial values d_i and R_i, respectively. The updated values, d_f and R_f, will be determined by the initial values as well as by Eq. (3.21). The loop will be terminated if $|d_f - d_i| < tol$, where tol is the tolerance. If the via radius is more critical, the decision point should be $|R_f - R_i| < tol$, and the process can be accordingly modified. After d_f and R_f are obtained, FDTD can be used to slightly adjust the parameters involved in the GARC model due to the complexity of the structure.

Lastly, the differences between circular gratings used in GARC and the focusing grating couplers need to be clarified. Focusing grating couplers, whose rulings are curved, have been reported to achieve efficient fiber-to-grating coupling [50, 156, 157]. These curved gratings, even though similar in shape to the circular gratings used in the GARC coupler, are based on a fundamentally different concept. The curved gratings are defined by the interference of two focusing beams (Beam 1 and 2) [156]. By directing Beam 1 toward the defined grating, Beam 2 can be generated, and vice versa. Similarly, the curved gratings used in [50, 156] are formed by the interference between an incident beam in the air, either a plane wave or a focusing beam, and a focusing wave in the waveguide. As a result, an

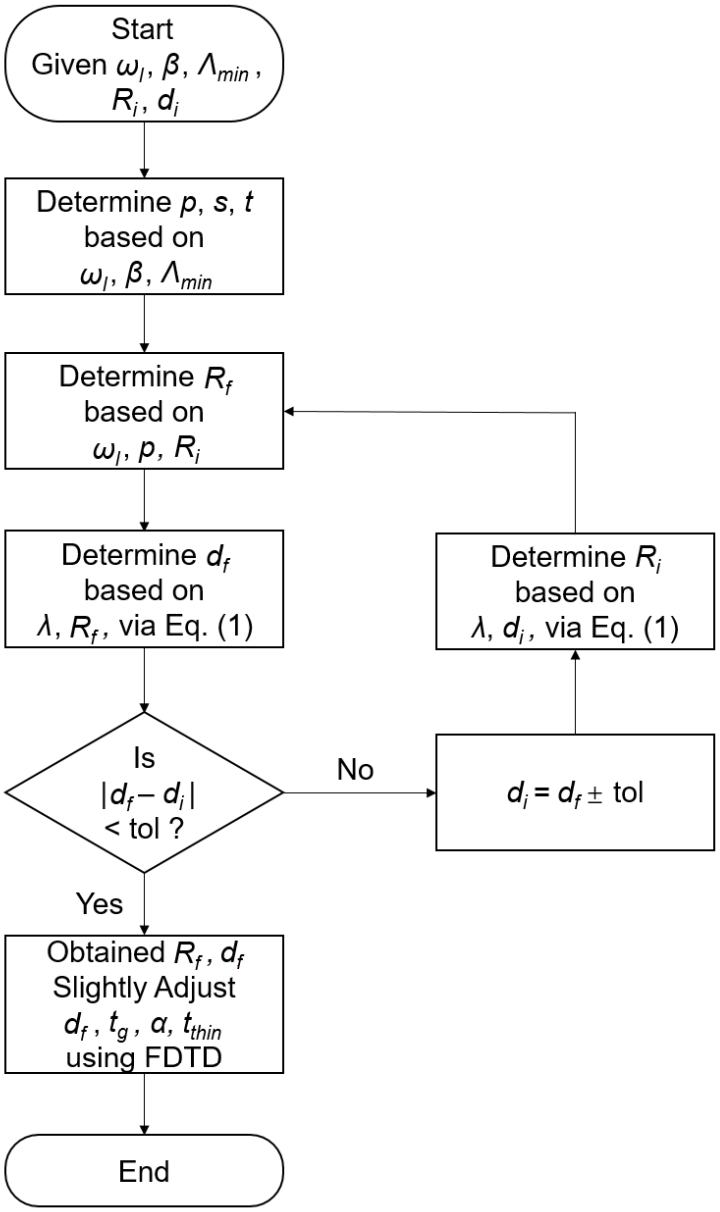


Figure 3.18: Process flowchart for the design of GARC couplers that satisfies targeted interlayer distance d.

externally incident wave, e.g., beam from the fiber, can excite the guided focusing wave, thus achieving the fiber-to-waveguide coupling. The focusing grating couplers defined by external and guided wave interference realize fiber coupling to ridge waveguides. On the other hand, the circular gratings used in a GARC coupler are defined by internal cylindrical wavefronts in the cylindrical cavity, and no external beams are involved. Since the inner circular gratings are not diffractive, GARC couplers can't serve as fiber-to-chip couplers.

3.4 Conclusion

A new type of interlayer coupling structure, Grating-Assisted-cylindrical-Resonant-Cavities (GARC) coupler, consisting of circular gratings and a high-index via has been designed. Two GARC couplers, made of Si/SiO_2 and Si_3N_4/SiO_2, respectively, have been introduced. The Si/SiO_2 GARC coupler is optimized to achieve 68% interlayer coupling efficiency (approximately 82% single-layer efficiency for comparison with rectangular diffraction gratings). The coupling structure is compact (20.72 μm), broadband ($\delta_{\lambda,1dB}$ = 170 nm or $\delta_{\lambda,3dB}$ = 270 nm), and relatively straightforward to fabricate due to the wide grating ridges and the absence of adjacent-layer reflectors. The GARC coupler performance is less sensitive to variations in grating ridge width, but it is more affected by the vertical layer thickness, e.g. grating etch depth and via height. However, compared to the horizontal feature definition which is limited by the lithography resolutions, vertical layer thickness can be more easily controlled (within nanometer range) by regulating the fabrication process. The GARC coupler is similar in principle to a FabryPerot resonator. Thus, it is conceivable for it to operate as an amplifier if it is doped with a rare-earth element or to operate as a modulator together with electro-optic materials. Overall, the GARC structure represents a promising candidate to achieve efficient and broadband interlayer coupling for 2.5D and 3D IC technologies.

CHAPTER 4

FIBER-INTERCONNECT SILICON CHIPLET TECHNOLOGY (FISCT): FIBER-TO-CHIP ASSEMBLY AND OPTICAL TESTING

Heterogeneous integration of electronics with photonics is a promising solution to meet the high-bandwidth, low-latency, and low-energy consumption needs of modern computing systems. Optical fibers, which offer extremely low loss, are especially critical in long-haul applications as well as in shorter distances found in data-centers and in high-performance computing (HPC). In the latter applications, photonic packaging, assembly, and interfacing to silicon electronics play a critical role in determining overall module performance, energy, and cost. In order to create seamless polylithic integration of photonics and electronics, advances in packaging and assembly are critical. In particular, the accurate alignment and assembly of fibers or fiber arrays to a photonic integrated circuit (PIC) are crucial steps to realize high-efficiency optical packaging and integration [158]. Common fiber alignment techniques include active alignments and passive alignments. With a typical alignment accuracy of approximately 100 nm, active alignment schemes require a complex setup consisting of a laser light source and a photodetector, with alignment adjustments realized through the use of a microscope and rotational stages [159]. The fibers or fiber arrays are aligned and assembled serially, which is not time efficient. On the other hand, passive alignment relies on the placement of fibers onto properly designed mechanical structures with a typical alignment accuracy of approximately 1 μm. While passive alignment has a lower alignment accuracy than active alignment, it is more time efficient and amenable to scale-up. Some passive alignment techniques include v-grooves [65, 66], plugs [160], ferrules [65], and light-splitting techniques [161] that only work for lateral couplings, e.g. edge coupling of fibers to waveguides, evanescent coupling of stripped fibers to waveguides, or coupling fibers mounted parallel to grating surfaces with the assistance of mirrors [161,

162], and thus they do not accommodate all fiber integration schemes. Passive alignment structures for vertical coupling have also been proposed [163], but they typically only work for coupling at a single location and thus it is not scalable. Figure 4.1 summarizes some of the passive alignment structures.

In this chapter, a passive fiber-array alignment and assembly approach using Fiber-Interconnect Silicon Chiplet Technology (FISCT) is proposed. FISCT consists of two key elements: 1) a silicon carrier with integrated lithographically defined mechanical self-aligning features and through-vias, and 2) 3D microprinted fiber ferrules that reside within the vias of the silicon carrier. FISCT enables massive precision self-alignment and assembly of optical fibers onto a package/substrate or a photonic chip and it is compatible with flip-chip bonding technology. The design and fabrication of the FISCTs are versatile and scalable to accommodate various photonic integrated circuit topologies and packages. The proposed FISCT has the potential in applications that include permanent or temporary fiber attachment. Note that the fiber chiplet concept aims to co-exist with the recent trends in digital electronics in which dice of heterogeneous functionality and materials (e.g. chiplets) are densely interconnected using 2.5D or 3D heterogeneous integration [164, 165]. Here the FISCT extends the chiplet concept to photonic devices and interconnections.

4.1 FISCT Structure

The general concept of FISCT is illustrated in Fig. 4.2(a). A polylithic integration platform (Heterogeneous Interconnect Stitching Technology (HIST) [164]) is formed using electrical and optical stitch chips. The electrical stitch chips, in the simplest form, provide high-density and low-energy connectivity between neighboring dice, while photonic stitch chips provide near-logic E-O/O-E conversion and interfacing to optical fibers. Here, FISCT addresses the technology challenges in interfacing a massive number of optical fibers to a photonic stitch chip with integrated waveguides and diffractive optical couplers [164]. Figure 4.2(b) and (c) illustrate details and assembly of the FISCT platform.

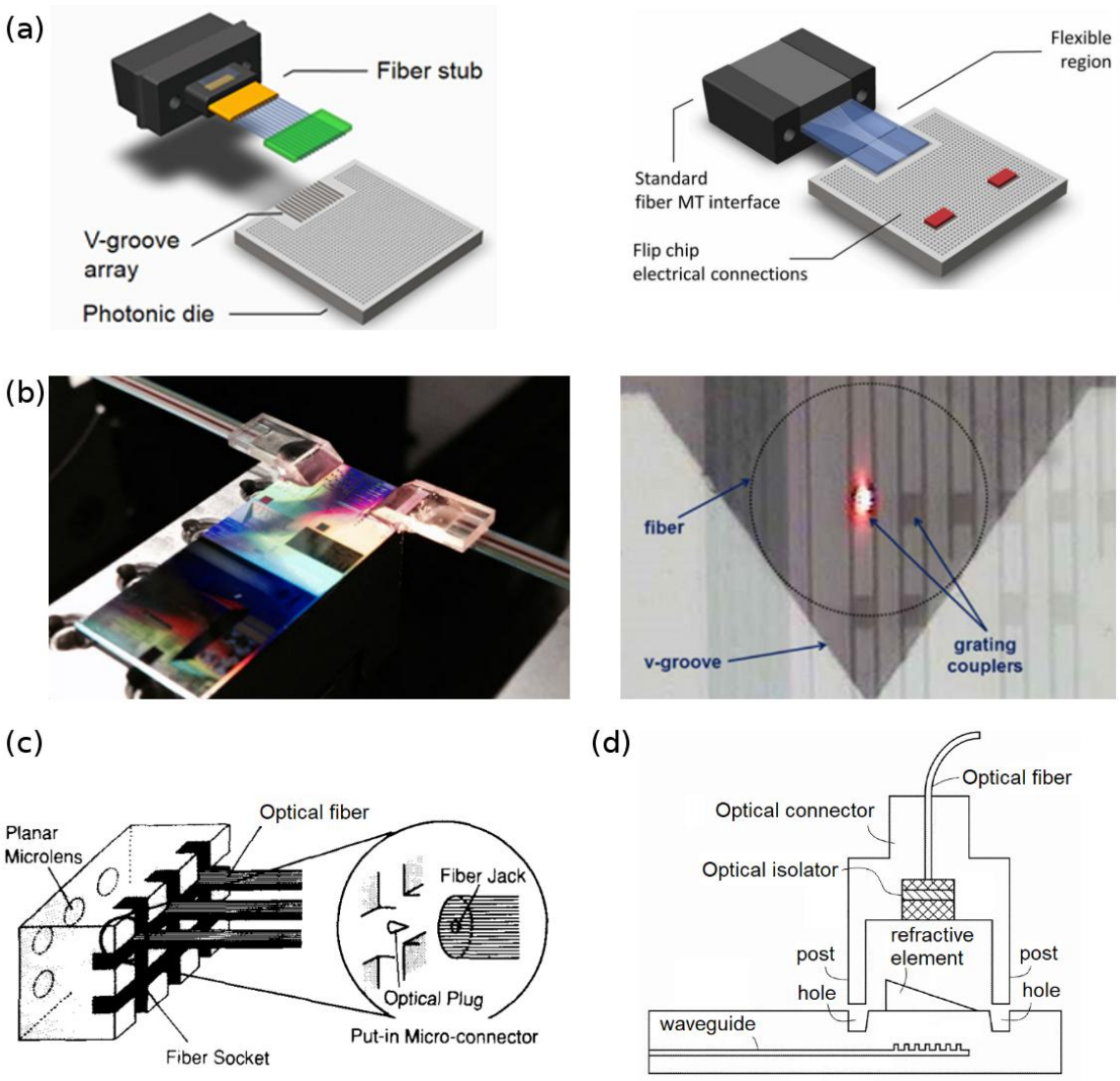


Figure 4.1: Some reported passive alignment methods: (a) fiber lateral alignment using fiber ferrule and v-grooves [65]; (b) horizontal fiber-to-grating coupling using light splitting techniques [161]; (c) fiber-to-lens array coupling using plugs [160]; and (d) fiber vertical coupling to diffraction elements using alignment structures [163].

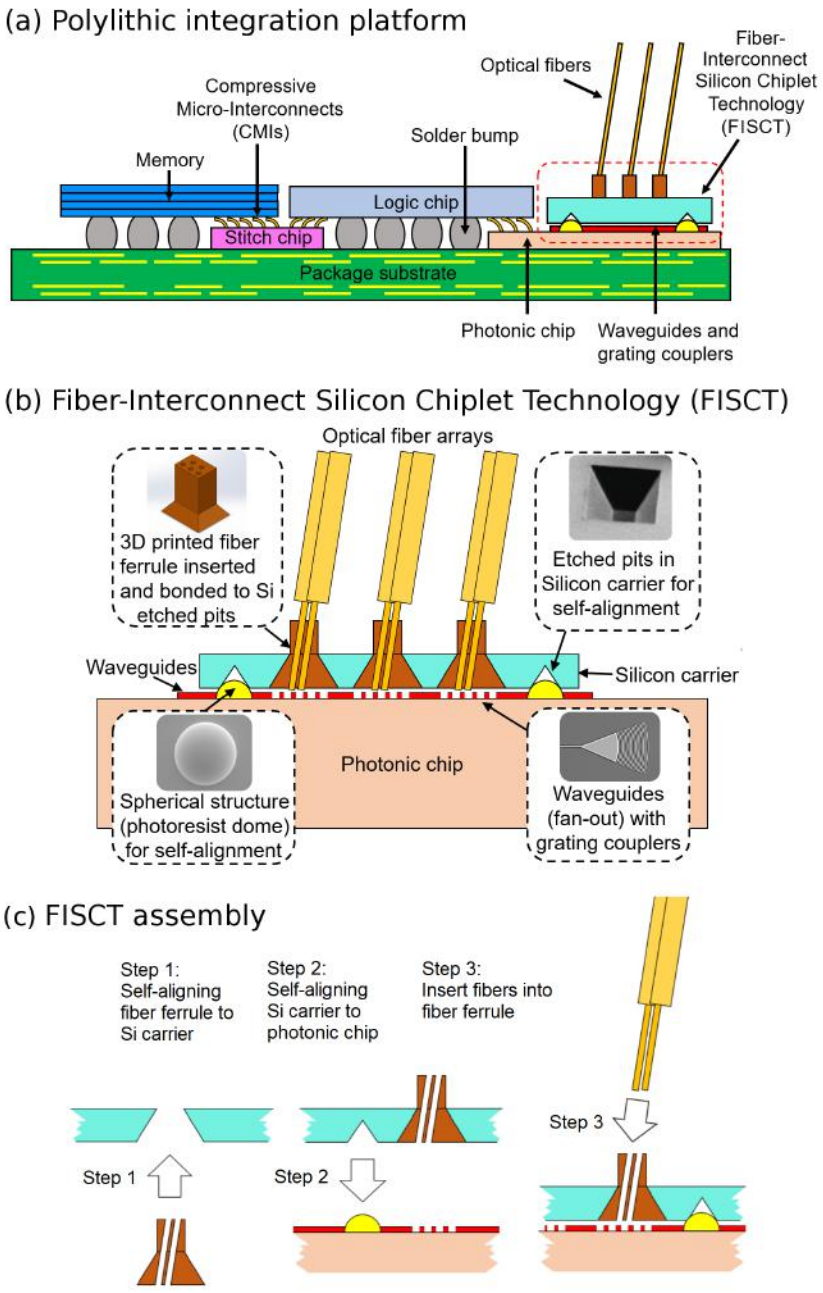


Figure 4.2: A conceptual view of the proposed Fiber-Interconnect Silicon Chiplet Technology (FISCT) in the polylithic integration: (a) fiber arrays are aligned to the polylithic HIST platform [164]; (b) an enlarged figure shows the details of the FISCT cross-section; and (c) the FISCT assembly process.

The design of FISCT is flexible and can be easily modified based on photonic integrated circuits (PIC) design. The structure is composed of a Si carrier substrate, self-alignment structures (etched pits and photoresist domes) [36, 166], and fiber ferrules microprinted by 3D laser lithography (two photon polymerization). The Si carrier is lithographically patterned and wet etched, resulting in etched pits of two sizes. The small pits are self-aligned to the reflowed photoresist domes deposited on the PIC, forming the positive self-alignment structure (PSAS) [36, 166]. The larger pits are fully etched through the Si carrier substrate, and the laser-printed fiber ferrules are inserted within the through-Si pits and self-aligned with the pits due to the complementary geometries. The fiber ferrules can be secured using epoxy, polymer, or mechanical latches (as shown in the experiment). Fiber placement aid, which is also microprinted by 3D laser lithography, is placed on top of the fiber ferrules for easier fiber insertion. Segments of the cleaved optical fibers without their coating are inserted into the channels within the fiber ferrules, which are fabricated with a pre-defined incline angle with respect to the normal, until they reach the surface of the PIC where they are self-aligned with the grating couplers.

4.2 FISCT Fabrication

The fabrication of the current FISCT approach (a test structure is shown in Fig. 4.3) can be divided into two parts: the chemically etched Si carrier and the 3D microprinted fiber ferrules (which are then inserted into the silicon carrier). Once formed, FISCT is self-aligned and assembled on a Silicon-On-Insulator (SOI) substrate with integrated ridge waveguides and surface-relief grating couplers. Specifically, the reflowed photoresist domes on the SOI substrate self-align with the smaller pits on the silicon carrier. Since the smaller pits are also lithographically aligned with the pits that hold the fiber ferrules, the fibers become aligned with the grating couplers on the SOI substrate. The details of the fabrication process are described in the following sections.

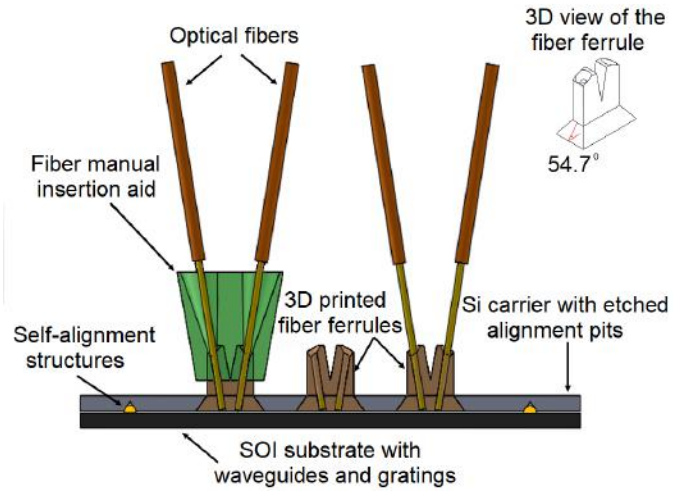


Figure 4.3: Schematic showing the preliminary FISCT platform used for demonstration and testing. The fiber ferrules hold two fibers for each of the grating-waveguide-grating photonic circuits on the SOI substrate.

4.2.1 Si Carrier Substrate

The process starts with a 300 μm thick, $\langle 100 \rangle$-oriented, and double side polished Si wafer. First, a 200 nm thick Si_3N_4 layer is deposited on both sides of the Si carrier substrate using LPCVD followed by pit lithographic openings on one side of the wafer. Next, the patterned wafer is placed into an RIE chamber to etch away the exposed Si_3N_4 layer. After photoresist removal, a patterned Si_3N_4 layer is revealed, which serves as the wet etching mask for the exposed Si. The wafer is next immersed in a 45% KOH bath at 90°C for 4 hours resulting in etched pits in the Si substrate. Note, during this process step, two different sizes of pits are formed simultaneously, and thus, all pits are lithographically self-aligned. The resulting pit size simply depends on the size of the window in the Si_3N_4 hard mask.

4.2.2 3D Microprinted Fiber Ferrules

The next step is to fabricate the 3D microprinted fiber ferrules. The ferrule structure is first designed using Solidworks. The bottom portion has a tapering angle of 54.7° corresponding to the angle between $\langle 100 \rangle$ and $\langle 111 \rangle$ crystallographic planes of the Si carrier. A channel through which the fiber is inserted is formed within the fiber ferrule; the chan-

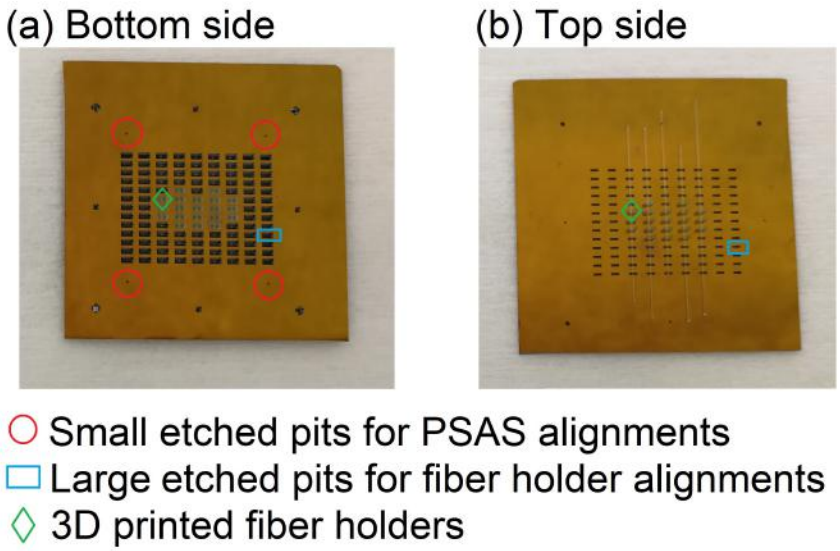


Figure 4.4: Bottom side and top side of the assembled passive alignment structure.

nel tilt angle (9°) matches the optimal fiber coupling angle and is pre-determined during the design phase. The fiber locations relative to the grating couplers should be properly designed based on the interlayer gap between the Si carrier and the SOI substrate. The design is then transferred to a Nanoscribe laser lithography system, which is based on two photon polymerization. The fiber ferrules are microprinted using in-house IP-S resist on a conductive substrate (ITO or Si) and the exposed structures are developed in SU8 developer for 20 min. The fiber ferrules can be printed in a 2D array with the same pitch as the through-Si etched pits in the Si carrier substrate, which makes the final assembly relatively simple. The assembled FISCT can be secured using fibers with coatings as latches (for proof of concept) or using photoresist/epoxy as glues. The Si carrier has a yellow sheen due to the presence of the thin Si_3N_4 layer. The bottom side (facing the SOI substrate) and top side (from which the fibers are inserted) of the assembbed FISCT structure are shown in Fig. 4.4. On the bottom side, the small etched pits labeled in red circles are used for self-alignment with the photoresist domes; the green diamond shape circles the location where the 3D printed fiber ferrule is mated with the large etched pits. On the top side, it is observed that the top side of the 3D printed fiber ferrule stands out of the Si carrier.

4.2.3 Photonic Circuit Substrate

The SOI substrate consists of a 2D array of passive photonic circuits. As a test circuit, two focusing grating couplers are connected by a ridge waveguide. Optical signal from the input fiber is coupled into the ridge waveguide via one grating coupler. The signal would then propagate along the ridge waveguide and diffract out once it reaches the second grating coupler. The out-diffracted signal is then picked up by the output fiber. The circuits with alignment marks are fabricated using an SOI wafer with a 250 nm thick device layer and a 3 μm thick buried oxide layer. The gratings, ridge waveguides and alignment marks are patterned in HSQ negative resist using e-beam lithography and fully etched in an ICP chamber. An SEM image of the fabricated grating is shown in Fig. 4.5. The resulting grating couplers have a theoretical fiber coupling efficiency of 30% (-5 dB), which can be improved by shallow etching the grating grooves but requires two step e-beam lithography. However, the fabrication of high-efficiency grating couplers is not the focus of this research. Some high-efficiency grating coupler designs can be found in Table 1.3.

Once the gratings and waveguides are fabricated, photoresist domes (i.e., PSAS) are formed by first spin-coating a positive photoresist (AZ40XT-11D) and then patterning photoresist cylinders using photolithography. The photoresist cylinders are reflowed yielding domes with desired height and radius. The size of the photoresist cylinder, e.g. radius and film thickness, should be co-designed with the Si etched pit openings [36].

4.3 Optical Testing

The gratings are first characterized using the fiber active stage alignment. The light source is swept from 1560 nm to 1630 nm. The fabricated grating coupler was optimized for the wavelength 1550 nm. The measured peak efficiency is approximately located at 1610 nm due to fabrication variations.

After obtaining the reference measurements using the active stage alignment, the fabri-

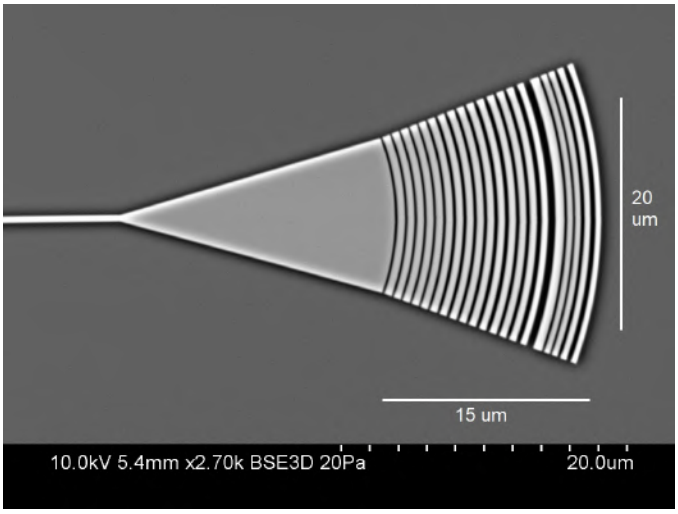


Figure 4.5: SEM image of the fabricated focusing grating coupler.

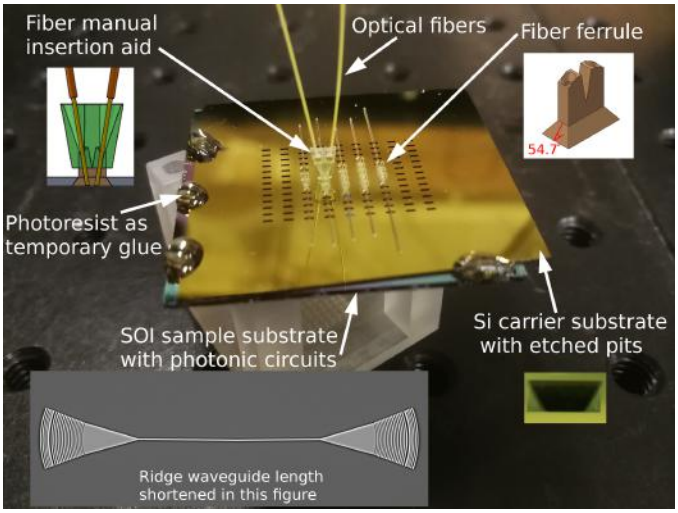


Figure 4.6: Testing setup after aligning the FISCT to the photonic circuit substrate.

cated FISCT is picked up using a flip-chip bonder and "dropped" above the SOI substrate. The PSAS-pits combination will ensure that FISCT and the SOI substrate will be precisely aligned and separated by a pre-defined interlayer gap. A small force is applied on top of the fiber ferrules while photoresist is dispensed on the edges of the aligned substrates to temporarily adhere the two parts together. The force is maintained until the photoresist is solidified. During this process, the PSAS also prevents the relative sliding of the substrates. A fiber insertion aid is placed on top of the fiber ferrule to assist fiber insertion by hand with the unaided eye, although, fibers (or fiber ribbons) are envisioned to be inserted with a fiber placement tool. The input and output fibers are subsequently inserted into the corresponding channels of the fiber ferrule.

In order to counterbalance possible fabrication variations and predict misalignment tolerances, the locations of the photonic circuits on the SOI substrate are shifted relative to the pit locations on the Si carrier in 1 μm increments along each of the x and y directions, and the amount of the shift is based on the coordinate of the pit-circuit pair. For example, center location g5 with coordinate (0,0) is perfectly aligned by design, while the circuit at location f4 with coordinate (-1, 1) is shifted by 1 μm in both of the $-x$ and $+y$ direction. The test locations are shown in Fig. 4.7.

Figure 4.8 shows the measurement results at three locations, namely g5 (no shift), g6 (circuit shifted 1 μm in the $+x$ direction), and h5 (circuit shifted 1 μm in the $-y$ direction). The black solid curves correspond to the active stage alignment while the blue dashed curves are obtained using the proposed passive FISCT. The left set of figures show the data measured in the unit of volts, which is the output of the photodetector, and the right column converts the relative intensity (the ratio of the passive FISCT measurements to the maximum measurement data by active alignment) into dB. At location g5, FISCT and SOI layers are designed to be perfectly aligned, and the peak signal (filtered data) measured by the passive alignment structure is approximately 86% (-0.67 dB) of that measured by the active stage alignment. The slightly lower efficiency measured by FISCT is probably due

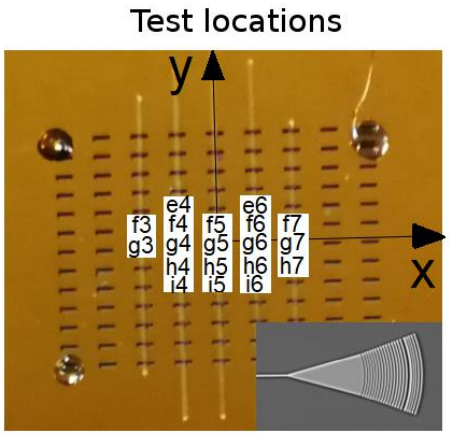


Figure 4.7: Testing locations of the photonic circuits.

to two reasons: on one hand, fabrication variations may cause the optimal fiber alignment location to shift relative to the grating, which can't be compensated by the fabricated fiber ferrules; on the other hand, the z location of the fiber inside of the ferrule is not controlled in the current design of demonstration, and thus the x location of the fiber tip may shift from the ideal aligned position. A change of 1 μm in the z direction results in a 158 nm lateral shift in the x direction, given the 9° tilt angle. Since the fiber is only allowed to move along a line in the fiber ferrule while it can move freely in space under active stage alignment, the extra degree of restriction causes the intensity of the passive method to be smaller. The peak signals at locations g6 and h5 are approximately 75% (-1.2 dB) and 50% (-3 dB) of the reference peaks at the corresponding locations, respectively. This implies that the misalignment along the x direction has a smaller effect than the misalignment along the y direction. This is because shifting the fiber along the y direction will break the symmetry and affect the coupling mode. Since fiber mode is approximately 10 μm in diameter and the focusing grating has a length of roughly 15 μm, shifting along the x direction is tolerable. In addition, the relative shift in the y direction doesn't depend on the z dimension if the fiber is controlled in the xz plane; the major y misalignment comes from the assembly of the Si carrier and the photonic chip.

It is noticed that the peak locations of the passive FISCT alignment and the active

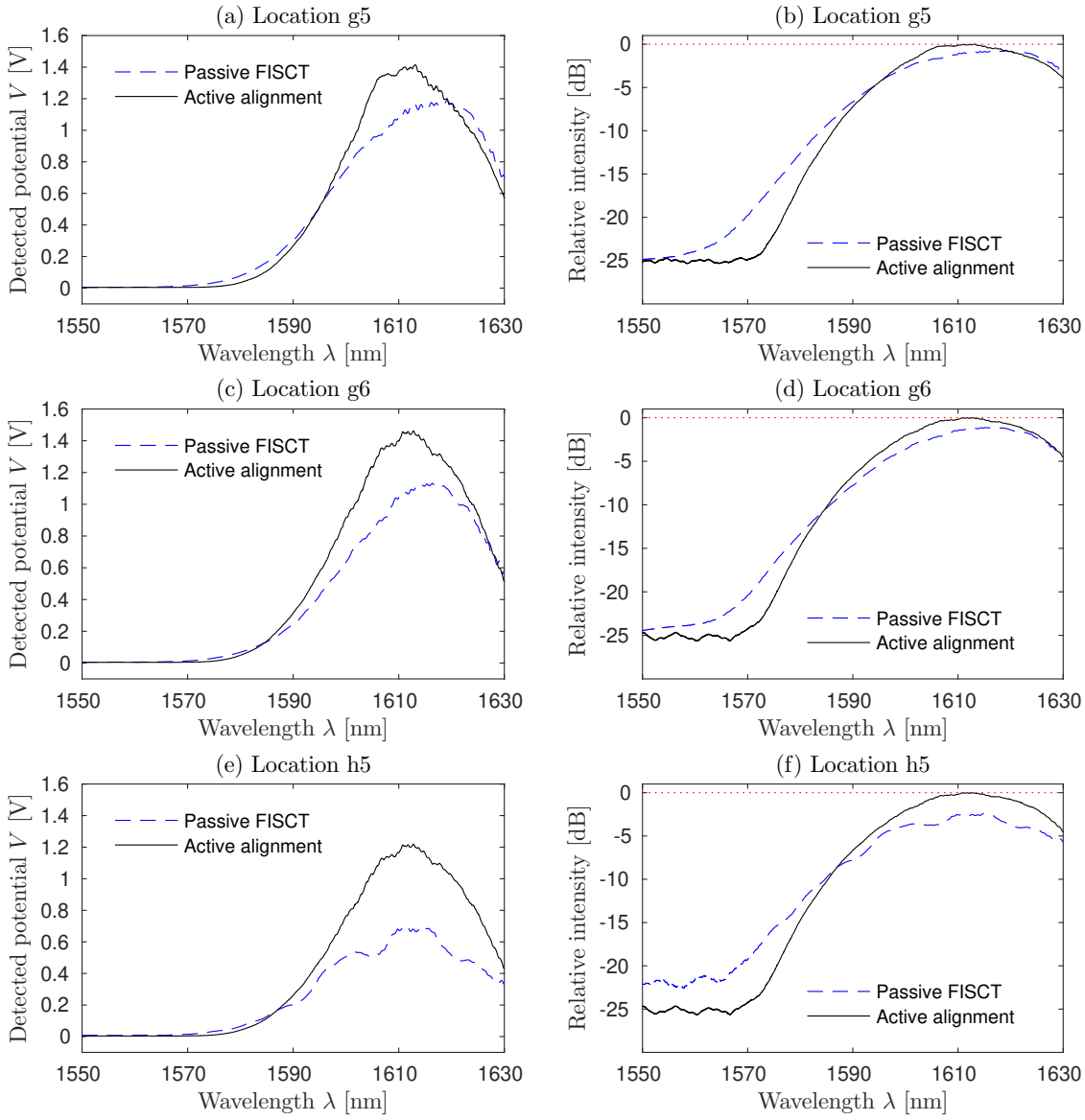


Figure 4.8: Comparison of the measurement results obtained from the passive FISCT alignments and the active stage alignments at three testing locations g5, g6, and h5. The left set of the figures are plotted in actual detected voltage, while the right set of the figures are plotted in dB with reference to the maximum signal at each location.

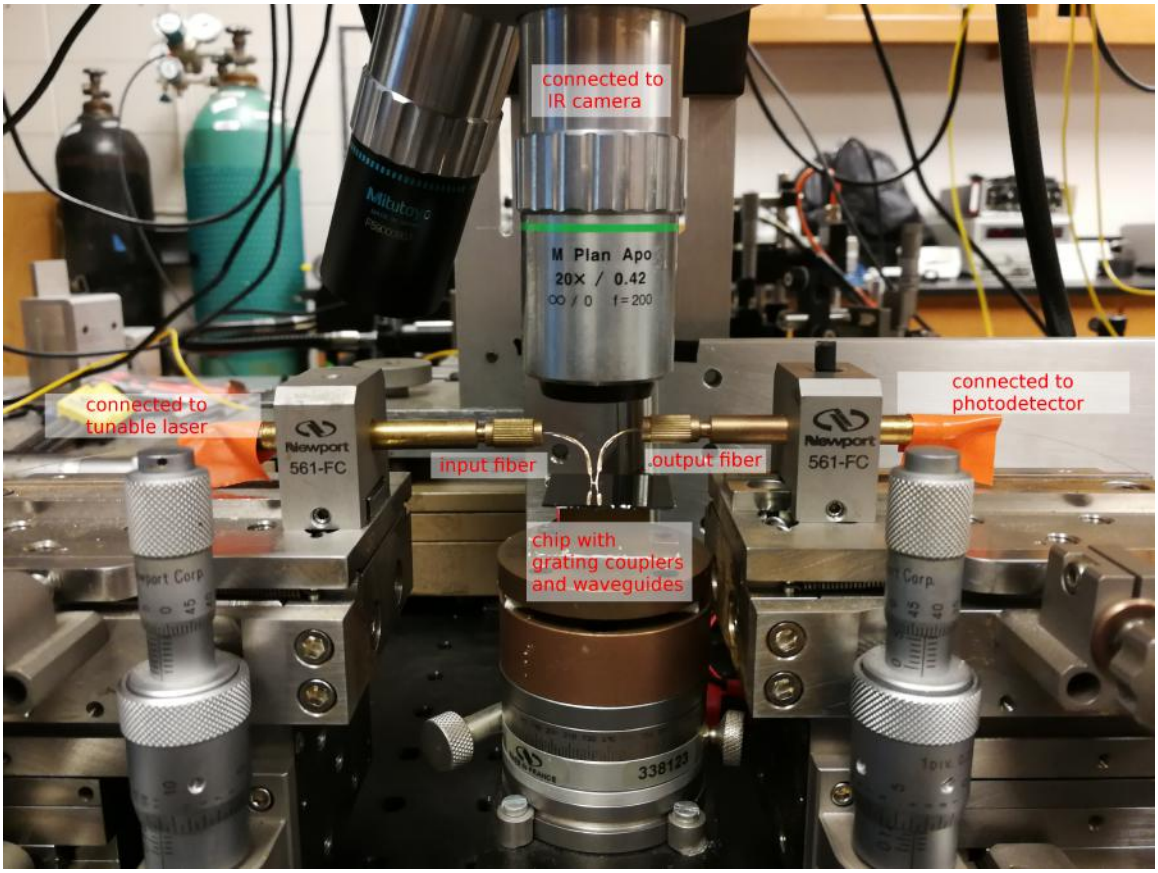


Figure 4.9: Active stage alignment setup whose measurements serve as the reference for the FISCT measurements.

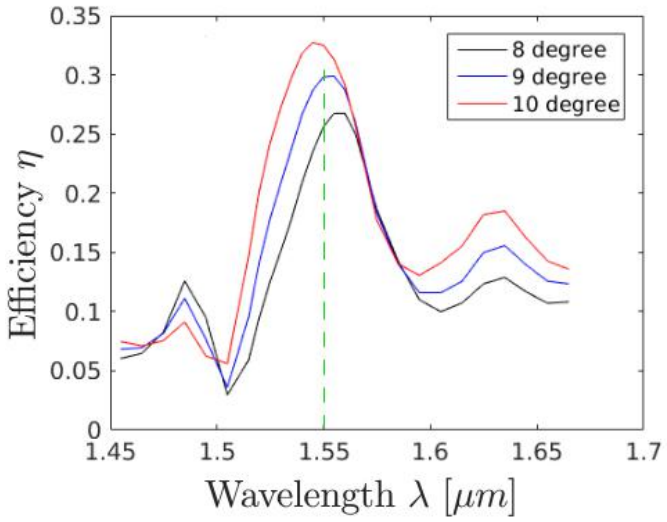


Figure 4.10: FDTD simulation of the effect of changing fiber coupling angle .

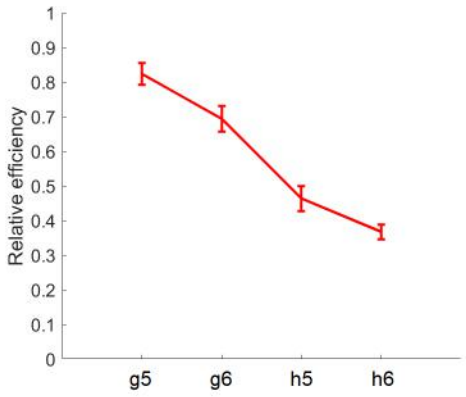


Figure 4.11: Multiple measurement data at location g5, g6, h5, and h6 (including their equivalent positions). Red line indicates the mean of the data at each location; error bar shows that 95% of the measured data lies within the range at each location.

stage alignment are different in Fig. 4.8. This is because the fiber insertion angles may be different in the two alignment methods. In the FISCT alignment, the fiber coupling angle is controlled to be 9° by the 3D microprinted channel within the ferrule, which may not be the same as in the active stage alignment. Figure 4.9 is a picture of the active stage alignment setup, from which it can be observed that the fiber coupling angle is controlled by a manually bent metal tube on the fiber chuck, and the two fibers are not arranged symmetrically. The fiber coupling angle affects the coupling efficiency and induces peak efficiency shifts, as can be seen from the FDTD simulation shown in Fig. 4.10. This manifests the feasibility of FISCT in controlling the fiber angle and its importance in efficient fiber integration.

The measurement of the passive FISCT is repeatable after retracting and re-inserting the fibers. For example, Fig. 4.11 shows multiple testing data at location g5, g6 (and its equivalent location g4), h5 (and its equivalent location f5), and h6 (and its equivalent locations f4, h4, and f6). The red line traces the means of the repeated passive measurement data, and 95% of the measurements locate within the range indicated by the error bar.

High intensity can be obtained by simply inserting the fibers into the fiber ferrules and slightly adjusting the fiber polarizations, which is more time-efficient compared with the active alignment schemes. The fiber is inserted until it reaches the SOI substrate; the fiber

end facet and the gratings were found to be undamaged. By properly designing the fiber ferrules and the Si carrier substrate, fiber arrays can be aligned using the proposed FISCT.

4.4 Fiber Array Integration

An example of fiber array integration by the FISCT platform is shown in Fig. 4.12. The waveguides of the photonic circuits can be short or long; for the short waveguides, fibers are assembled using a single ferrule, designated as "short-length ferrule", and for the long waveguides, fibers are assembled using a pair of ferrules separated by a designed distance, designated as "long-length ferrule". The channels inside of the ferrules are arranged in the close-packed formation in order to achieve the most compact fiber integration. An optical image of the fiber array integration is shown in Fig. 4.13. In the example illustrated in Fig. 4.12, the ferrules are secured on top of the Si carrier using mechanical structures, such as horizontally inserted fibers (for short-length ferrules) and 3D printed latch structures (for long-length ferrules). Optical images of the ferrules are shown in Fig. 4.14 and Fig. 4.15.

4.5 Conclusion

A passive Fiber-Interconnect Silicon Chiplet Technology (FISCT) is proposed and demonstrated to achieve accurate and reliable fiber-to-chip alignment, assembly, and packaging. It is compatible with current chiplet and flip-chip bonding technology, and its design is versatile and scalable to satisfy various integration requirements. FISCT enables time-efficient alignment, assembly, and packaging of fibers or fiber arrays and it is suitable for fiber-to-chip, fiber-to-package, prototyping, and testing applications.

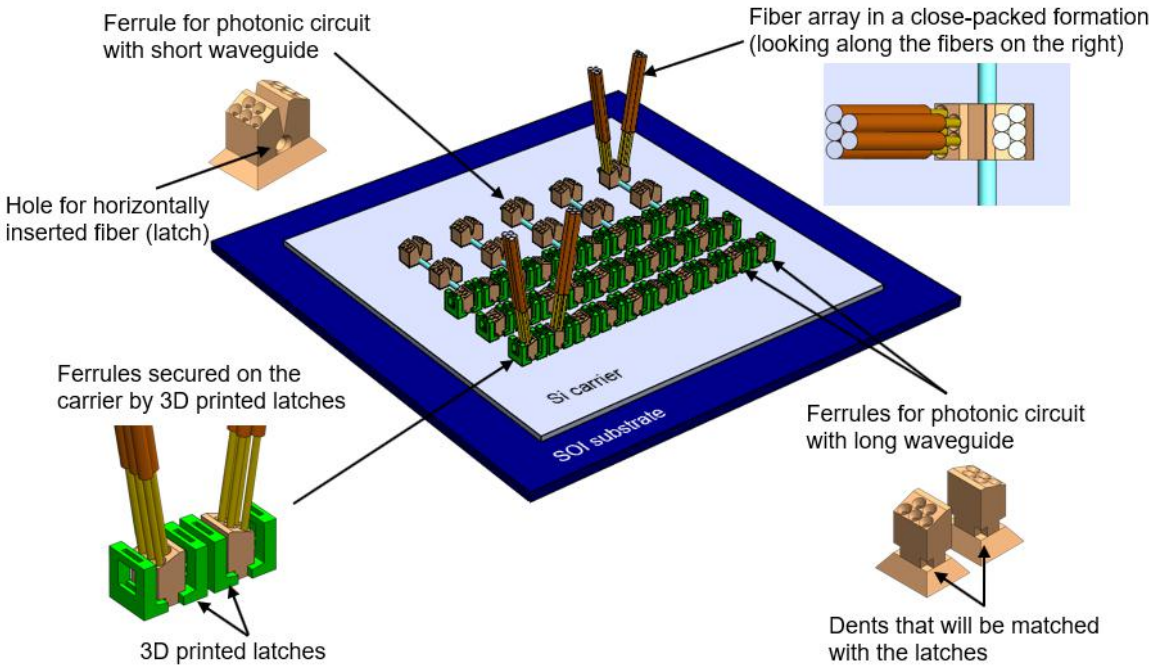


Figure 4.12: Fiber array integration using FISCT.

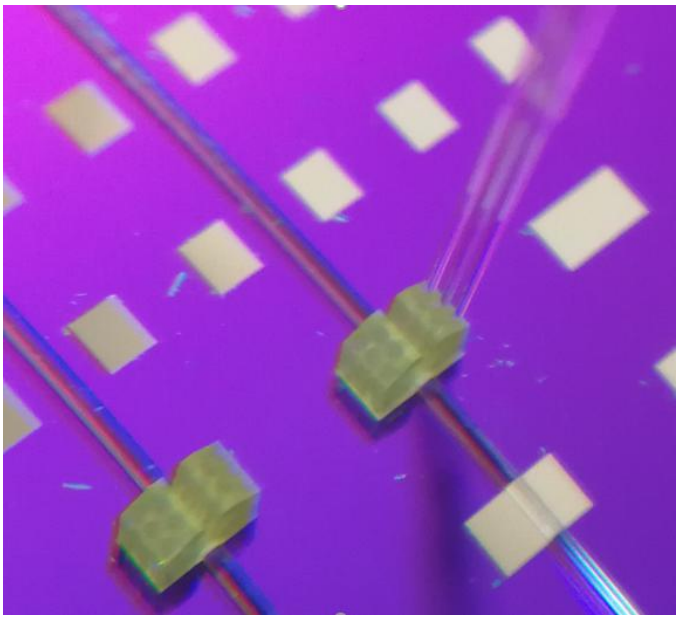


Figure 4.13: Close-packed fiber array inserted to a 3D microprinted fiber ferrule. The purple color of the substrate is due to the presence of a thin Si_3N_4 layer, whose thickness is thicker than the Si_3N_4 layer on the Si carrier shown in Fig. 4.6. The scattered blue debris on the Si carrier are the broken Si_3N_4 pieces.

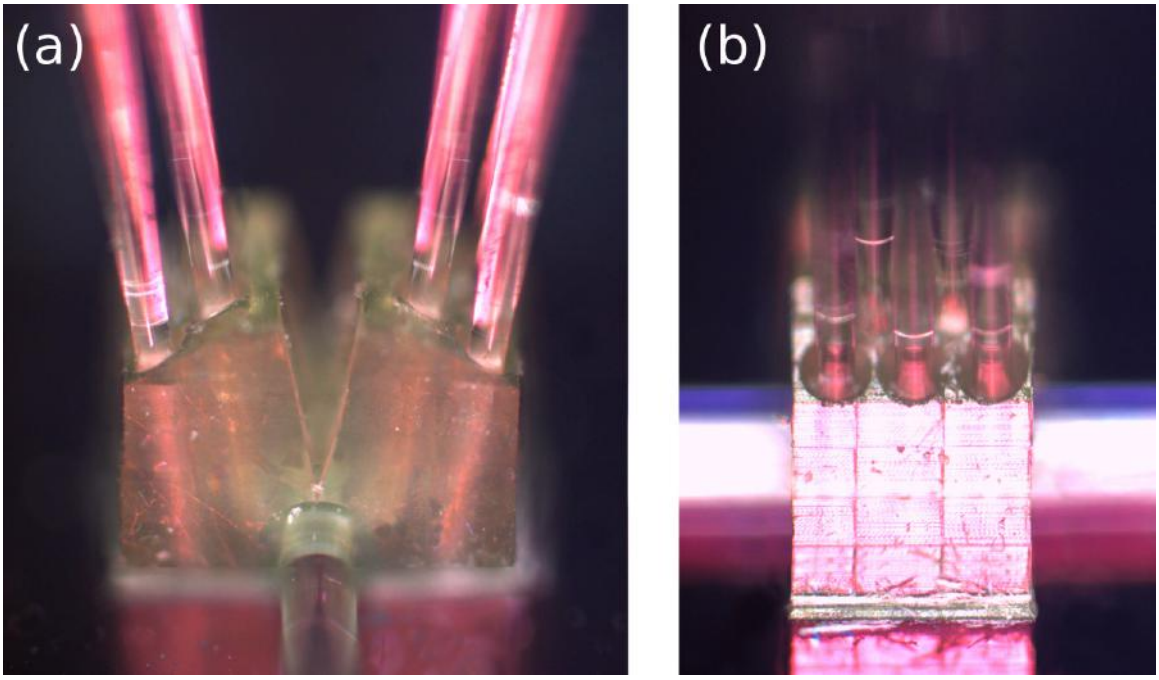


Figure 4.14: Optical images of fiber ferrules for short photonic circuits with fiber array inserted into the channels. The ferrule is secured on top of the Si carrier by a fiber horizontally inserted through the middle of the ferrule.

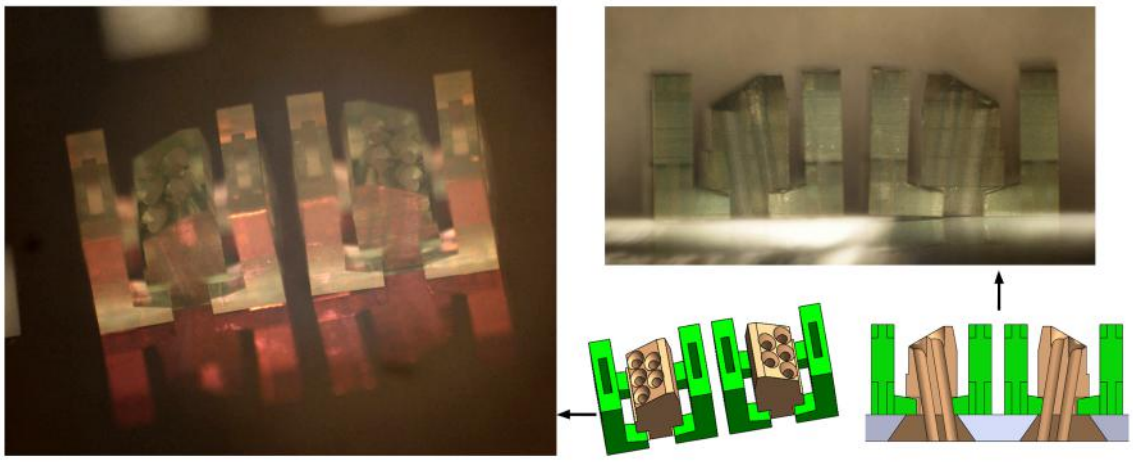


Figure 4.15: Optical images of a pair of 3D microprinted fiber ferrules for long photonic circuits. The ferrules are secured on top of the Si carrier by 3D microprinted latches.

CHAPTER 5
FUTURE WORK

Optical interconnects have shown their potential to satisfy the ever-growing need for large bandwidth, low energy, and high data rate. The heterogeneous multi-die integration within a single package using both off-chip and on-chip optical signaling increases the need for within- and off-package optical connectivity. Innovations in optical coupler design and system integration are therefore ones of the highly sought-after topics. The proposed GARC coupler and the FISCT integration platform are promising solutions to meet this demand. The GARC couplers are new interlayer coupling structures that have never been exploited before. The benefits of the GARC couplers, including but not limited to wide spectral bandwidth and high coupling efficiency, are only demonstrated by simulation, and thus the structures require experimental validation and performance evaluation. Efficient fiber-to-chip coupling and assembly are also important to optical system integration, and the FISCT platform offers more possibilities to integrate fiber arrays since it is compatible with various photonic circuit designs and heterogeneous integration.

5.1 Si/SiO_2 GARC Coupler Demonstration and Application

5.1.1 GARC Coupler Fabrication

The Si/SiO_2 GARC coupler can be fabricated using the process shown in Fig. 5.1. An SOI wafer with a 220 nm thick Si device layer and a 3 μm thick SiO_2 buried oxide layer can be used as the substrate. The first step is to etch alignment marks into the SOI wafer. The alignment marks can be patterned using Heidelberg Maskless aligner MLA150 or Elionix ELS-G100 e-beam lithography (for better alignment accuracy). The etch mask can be either a lifted-off chromium layer (step (a.1) to (a.4)) or a positive resist layer (e.g. SC1827 for

photolithography or PMMA for e-beam lithography). The exposed regions in the mask are etched sequentially using Cl_2 (Si etch) and using CF_4, O_2, and H_2 (SiO_2 etch) in Plasma-Therm ICP, resulting in 2 μm deep trenches. After the etch mask is removed and substrate is cleaned, a 120 nm thick Si_3N_4 layer is deposited onto the substrate using LPCVD process in Tystar nitride furnace. This step will define a Si_3N_4 overlayer on top of the rectangular grating to enhance the coupling efficiency, but it is optional. Subsequent step involves circular and rectangular grating patterning using Elionix ELS-G100 EBL system. The Si_3N_4 layer is fully etched (same etching gas as for SiO_2) followed by 40 nm etch into the Si device layer to create the grating ridges, which is also carried out in Plasma-Therm ICP. Next the rectangular grating regions are patterned and covered by negative ebeam resist (e.g. HSQ) and the uncovered Si_3N_4 regions are etched. A similar process is used to isolate the whole waveguide structures by fully etching the Si device layer. The following SiO_2 thin layer can be CVD-deposited or sputtered, depending on further investigations of the film quality. Subsequently, a 2 μm thick Si layer is LPCVD-deposited using Tystar poly furnace, followed by ICP etching to define the Si via.

A preliminary verification can be carried out half way through the process after the Si via is defined. Since the via functions as a resonant cavity and only a portion of its power leaks out though the SiO_2 thin layer, the field pattern at the top of the via should be similar to the simulated field pattern as shown in Fig. 5.2. After introducing optical signal into the bottom input waveguide, the field pattern on the top of the via can be imaged by an IR camera or a visible light camera by the assistant of upconversion nanoparticles, e.g. Er doped $NaYF_4$, which convert IR light into visible spectrum. The image can be compared with the simulated field pattern to approximate the resonance behavior of the via. Complete coupling efficiency and spectral response measurements can be conducted using a tunable laser and a power meter after the entire GARC coupler is fabricated. The verified GARC coupler can be further integrated into the photonic chips or heterogeneous integration platforms.

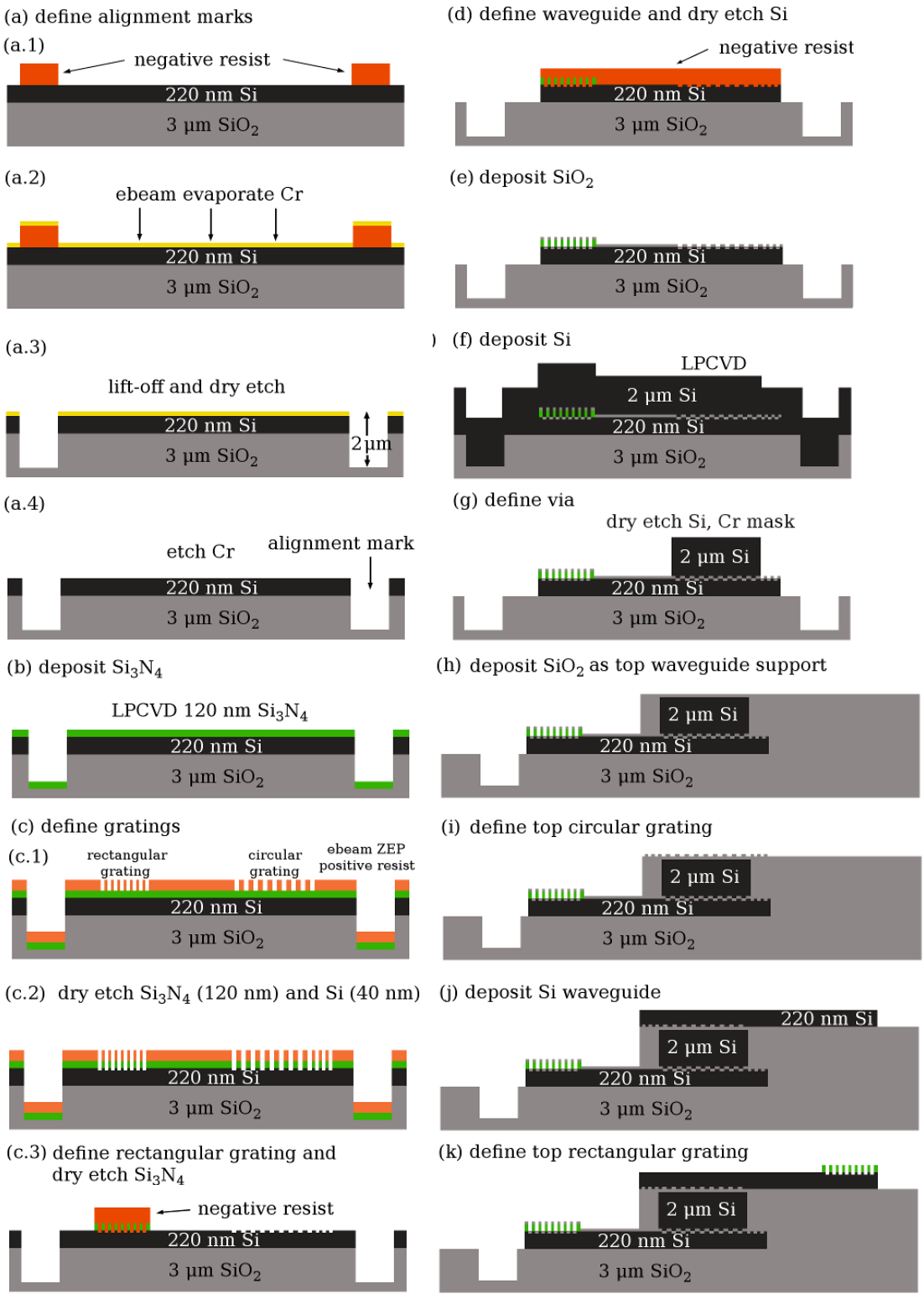


Figure 5.1: Fabrication steps for the Si/SiO_2 GARC coupler.

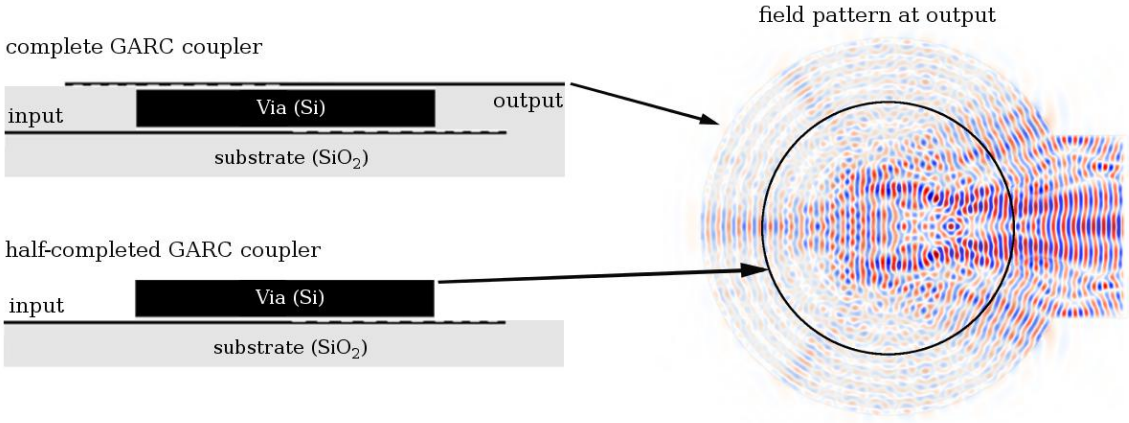


Figure 5.2: H_z field pattern at top of the via, assuming optical signal is launched from the bottom left waveguide. The field patterns at the top of the via are similar for the complete and half-completed structures.

5.1.2 GARC Coupler for Multilayer Photonic LIDAR

The GARC couplers can be used in the field of LIDAR, which is a popular tool for autonomous driving nowadays [167, 168]. Beam steering in LIDAR is usually achieved by MEMS [169], digital micromirrors [170], and liquid crystal devices [171]. There is a growing interest in photonic LIDAR due to its compact size and beam steering capability. For example, Fig. 5.3 shows an example of a photonic LIDAR with phase shifter circuits and beam steering apertures consisting of ridge waveguides with gratings etched on the edges (Fig. 5.4). The phase shifter circuits in photonic LIDAR [172] occupy a lot of space; in this case, if a multilayer structure is used by defining another layer of beam steering apertures [173] on top of the phase shifter circuits, the device size can be further reduced. GARC couplers can be applied in such multilayer structure to achieve efficient and broadband interlayer coupling.

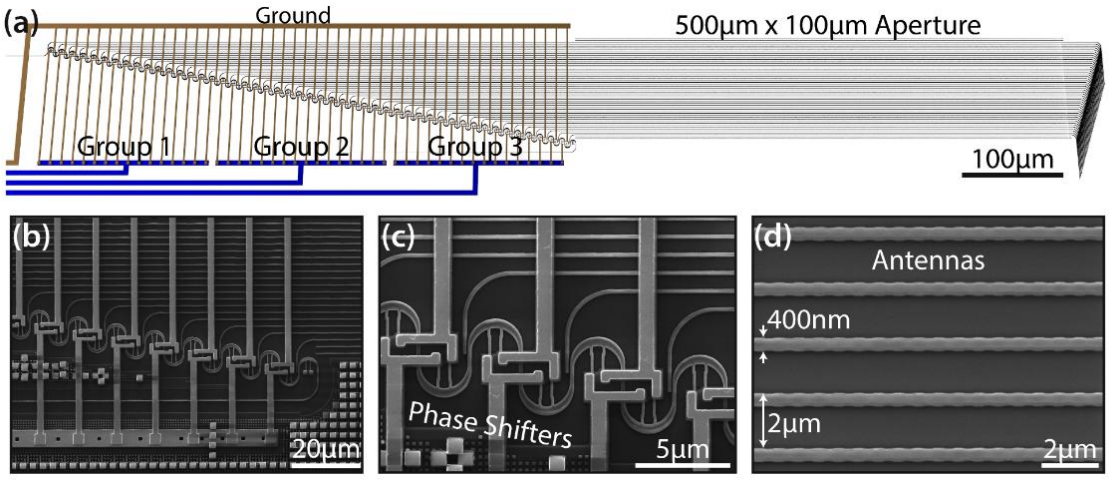


Figure 5.3: (a) An example of a photonic LIDAR consisting of an optical phased array; (b) SEM image of the cascaded phase shifter architecture; (c) close-up view of the thermal phase shifters; and (d) the fully-etched silicon grating-based antennas with a waveguide width of 400 nm and a pitch of 2 μm. Figure is obtained from [172]. The antennas shown in (d) have a similar structure as shown in Fig. 5.4.

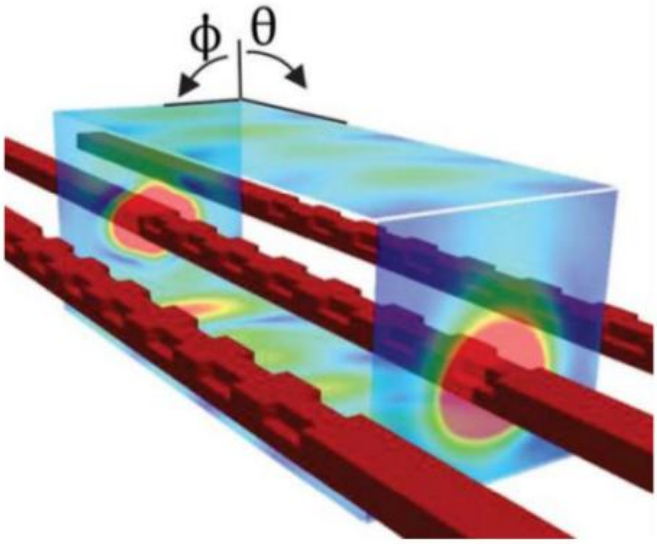


Figure 5.4: Beam steering angles realized by ridge waveguides with edge-defined gratings. Figure is obtained from [173].

5.2 FISCT Platform Modification and Integration

5.2.1 Vertical Control of Fiber Insertion

The FISCT demonstration in Chapter 4 can be further improved by designing microstuctures to control the vertical (z) location of the fiber tips. One one hand, the z location is correlated with the x location due to the existence of the fiber tilt angle, and thus the uncontrollable vertical location may cause deviations from the ideal x alignment position. One the other hand, the emitted field out of the fiber tip may have different radiation patterns, and the emitted power many not be efficiently collected by the grating if the distance between fiber facet and the substrate is too large.

5.2.2 3D Microprinted Out-of-Plane Coupler

In Chapter 4, a preliminary FISCT structure is demonstrated to couple optical fibers to on-chip waveguides through diffraction grating couplers. The reason for using diffraction gratings as couplers is that gratings can diffract the incoming fiber signal into the guided wave propagation direction in the planar waveguide, thus achieving vertical-to-lateral beam transition. Since diffraction gratings are sensitive to wavelength and have limited bandwidth due to the Floquet and Bragg condition, they may not simultaneously satisfy the broad bandwidth and high efficiency requirements. Other types of couplers that achieve vertical-to-lateral transition can also be used in the FISCT platform. For example, Blaicher *et al*. [174] experimentally demonstrated an out-of-plane fiber-to-waveguide coupler using a 3D microprinted structure consisting of free-form lens, polymer waveguide bend, and double tapers, as shown in Fig. 5.5. This type of coupler is a bent version of the fiber-butt (lateral) coupling structure, which has the advantage of broad bandwidth. It was reported that the 3D microprinted out-of-plane coupler has a coupling efficiency of -0.8 dB (83%) at the wavelength 1550 nm with a 1dB bandwidth exceeding 100 nm [174]. To make the fiber-integration more compact, the 3D microprinted out-of-plane couplers can be arranged

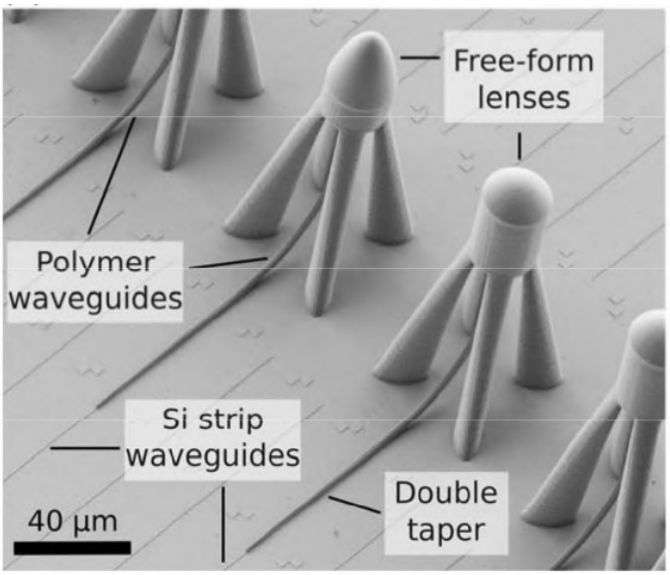


Figure 5.5: Scanning electron micrograph of 3D microprinted out-of-plane couplers, comprising a dielectric polymer waveguide connected to a silicon nanowire on one side, and a microlens on the other side. Picture is obtained from [174].

in an array with a shared supporting structure, and an example structure is fabricated using the Nanoscribe laser lithography system at Georgia Tech, as shown in Fig. 5.6. The arrangement of the couplers can be co-designed with the FISCT platform.

Further efficiency enhancement of the 3D microprinted coupler can be achieved by optimizing the lens, taper, and waveguide bend. Since lens structure is essentially a spot size converter, the first element that affects the fiber coupling is the lens profile, which determines the beam convergence behavior, such as beam waist and its location. Beam waist is an important parameter because it determines the width of the connecting polymer light guide and the degree of mode matching with the on-chip waveguide. In theory, the polymer taper, which connects the polymer lens and the transition waveguide bend, should be designed with the same convergence angle as the beam, and the connecting polymer transition waveguide should begin at the beam waist. This can be demonstrated using Finite-Difference Beam Propagation Method (FD-BPM, see Appendix C for derivation). For example, Fig. 5.7 shows the BPM-simulated convergence behavior of fiber input light for three parabolic lenses whose profiles are defined by $y^2 = 2px$, where x is the horizontal

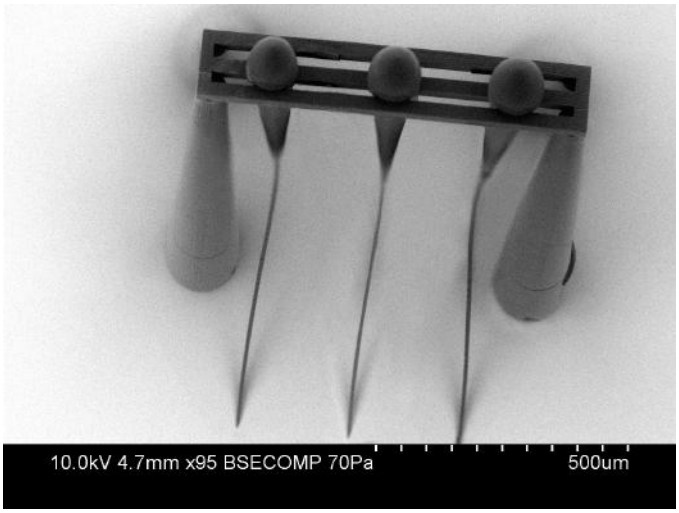


Figure 5.6: Scanning electron micrograph of 3D microprinted out-of-plane couplers secured on one supporting structure. The spacing between each coupler is 250 μm which corresponding to the diameter of an optical fiber with coating.

direction and y is the vertical direction. The light blue shaded region indicates the polymer lens tip and the connecting circular light guide (a very wide light guide is used to show the convergence behavior of the beams). The beam converges at a shorter distance for lens with sharper tip (smaller p value), as shown in Fig. 5.7. The field profiles at the beam waist for the three cases are plotted in Fig. 5.8, which reveals that the FWHM beam waist diameters are 3.46 μm, 4.42 μm, and 5.72 μm for p = 10, 20, and 30, respectively. Nevertheless, the $|E_z|^2$ integration over the peak region, which is proportional to the field intensity, shows that the case p = 30 contains more power confined in the waist region (integrated values are 25.6679, 42.3958, and 61.9373 for p = 10, 20, and 30, respectively). This can be seen from Fig. 5.7(a) that there are more stray rays in the p = 10 case which carry a lot of energy away from the beam waist. This phenomenon implies that narrowing down the lens tip may reduce the converged beam waist at the expense of energy loss. Further optimization needs to be conducted to evaluate this trade-off. For example, lens surface profile defined by other mathematical functions or random spline control points can be used. To assist the BPM simulation method, ray tracing software can also be exploited.

In addition, it is observed that the convergence length is very large, e.g. 40.3 μm,

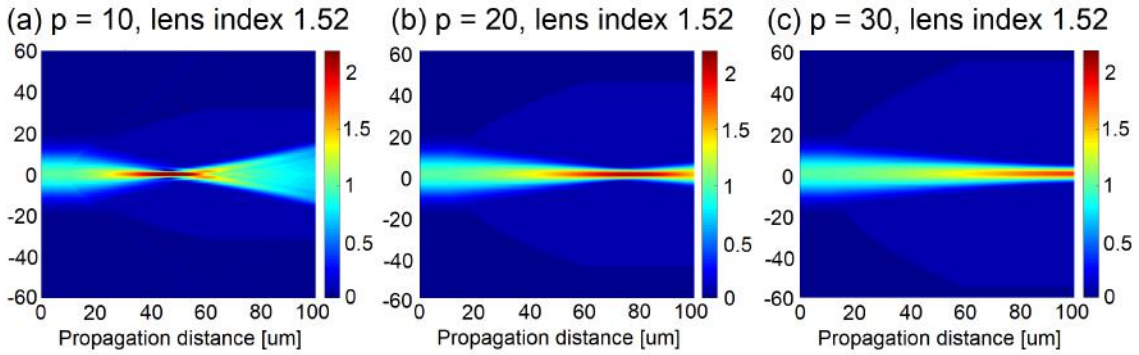


Figure 5.7: BPM-simulated beam convergence by the presence of parabolic lens with surface profile defined as $y^2 2 = 2px$: (a) p = 10, (b) p = 20, and (c) p = 30.

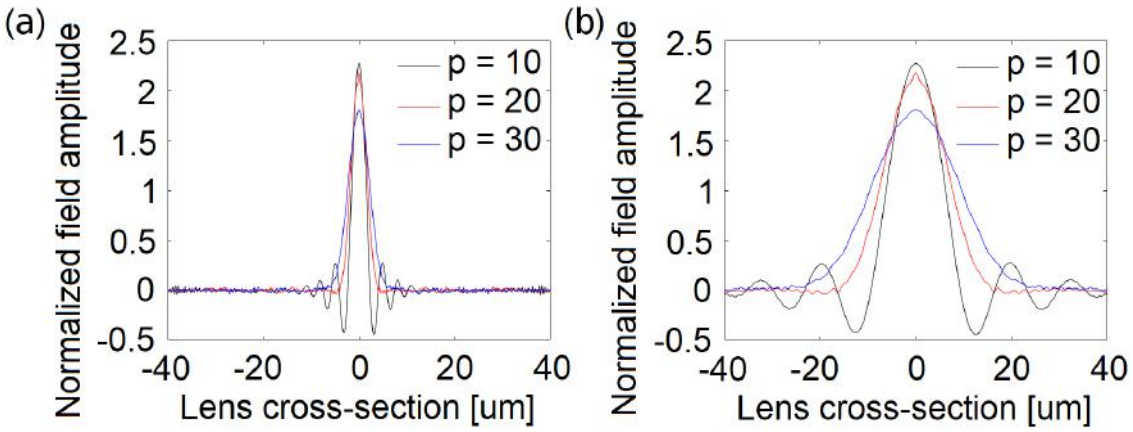


Figure 5.8: Field profile of the converged beam at the beam waist for the three cases. The right figure (b) is an enlarged view of the sharp peaks in the left figure (a).

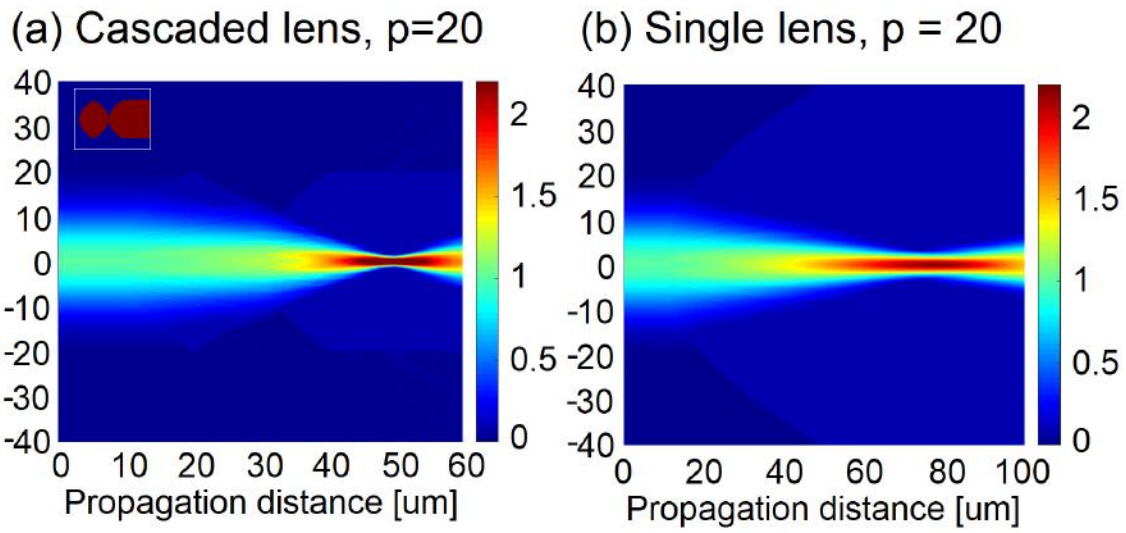


Figure 5.9: Comparison of convergence behavior of cascaded lenses (a) and a single lens (b) with same parabolic p value. Cascaded lenses result in shorter convergence distance.

81.6 μm, 99.8 μm for p = 10, 20, 30, respectively, even though this demonstration is not an optimized case. Such large convergence length is unfavorable to achieve efficient 3D microprinting and compact fiber assembly. Therefore, it may be beneficial to include multiple cascaded lens or converging surfaces to decrease the convergence length. Figure 5.9 shows the advantage of cascaded lens in reducing the beam convergence distance. The cascaded lens surfaces in Fig.5.9(a) are defined by the same parabolic function $y^2 = 2px$ with p = 20 as that of the single lens in Fig. 5.9(b). The same Gaussian input field is incident on both structures, and it is shown that the cascaded lens results in a shorter convergence distance (49.7 μm), compared with that obtained from the single lens (81.6 μm). Also note that the converged beam waist is smaller for the cascaded lens. This trend shows that incorporating cascaded lens or curved surfaces offers a promising solution in reducing beam convergence length and spot size.

The lens structure can be further optimized by selecting materials with higher refractive indices. Since the lens structure is microprinted using two photon polymerizations, any photosensitive materials with potential low loss can be used. For example, Fig. 5.10 compares two single parabolic lens with different refractive indices. Lens with higher in-

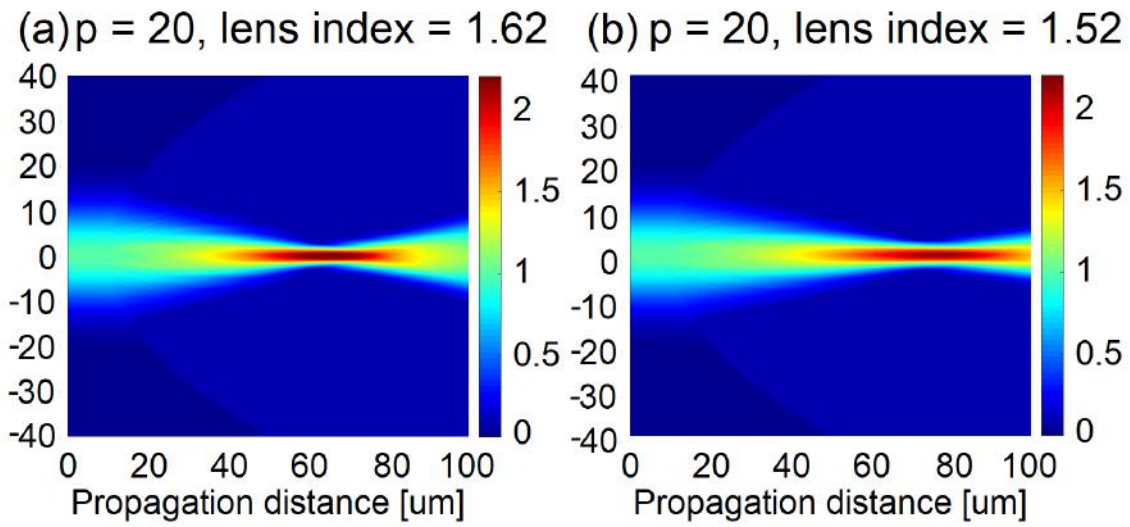


Figure 5.10: Comparison of beam convergence of lenses with different refractive indices. Both lenses have the same parabolic surface profile.

dex (Fig. 5.10(a), n = 1.62) converges beam to a narrower spot size at a shorter distance than the one with lower index (Fig. 5.10(b), n = 1.52).

Another important component of the coupling structure is the double tapers used to achieve evanescent coupling between the polymer transition waveguide and the planar waveguide. The optimization of the double tapers involves many parameters, such as taper tip width, taper length, waveguide thickness, waveguide width, materials of both the on-chip waveguide and the polymer light guide. The optimization can be carried out using COMSOL (FEM) or Lumerical (FDTD) with optimization algorithms, which require relatively intensive computation. Theoretically, a longer taper will increase the coupled power because the mode field change is gradual and back-reflection is small. As shown in Table 1.4, high efficiency tapers usually have lengths larger than 100 μm. Nevertheless, by optimizing other parameters of the double tapers, it may be possible to reduce the coupling length and thus reduce the coupler footprint. Note that the waveguide coupling structure is not limited to the double taper design, and other efficient evanescent coupling structures reported in the literature can also be explored. This part of the 3D microprinted coupling structure requires extensive research in order to find optimal combination of the vast pa-

rameter space in the design, but it also gives more opportunity to increase the coupling efficiency.

The third key component of the coupling structure is the polymer guide bend. Theoretically, the larger the curvature (smaller bend radius), the larger the bending loss. A larger bend radius would reduce the bending loss, but the light guide length required to achieve specific bend angle would increase, which may not favor compact integration designs. Change in material index will also affect the bending loss. It is expected that light guide with higher refractive index (higher index contrast with cladding material) gives smaller bending loss at a given bending curvature, due to the more confined mode field in the light guide. Thus using polymers with higher refractive index will help in reducing the dimension of the light guide bend. However, the structure with reduced dimension may require higher fabrication accuracy since it is more sensitive to surface roughness which gives rise to considerable propagation losses.

Other design considerations may involve reflection reduction and structure protection. For example, the 3D printed coupling structure can be embedded in a low-index cladding polymer, oil, or fluid to reduce reflection at all possible surfaces and protect the 3D structure. Using this approach may reduce the index contrast of the polymer light guide and affect the bending loss. In summary, there are many trade-offs in the design and optimization of the 3D printed fiber coupling structure, and all the three key components, lens, guide bend and evanescent coupler, are related and should be considered at the same time.

5.2.3 Heterogeneous Integration of FISCT

As heterogeneous integration is the trend of current digital systems, FISCT should be eventually integrated to electronic packaging, such as the Heterogeneous Interconnect Stitching Technology (HIST) platform. Even though the research focus of this thesis is passive optical interconnects, photonic integrated circuits (PICs) also consist of active components such as lasers, detectors, modulators, etc. As more electrical routings and components are

incorporated, either on the photonic chip/package or on the Si carrier substrate of FISCT, proximity effects such as thermal, optical, and RF crosstalk may occur and affect the performance of the photonic circuits.

Previous research has been conducted on the analysis of thermal crosstalk on modulators [175], RF crosstalk on modulators [176], optical radiative crosstalk between waveguides due to sidewall roughness [177], electrical crosstalk (leakage and induction currents, parasitic capacitance and inductance) between detectors and lasers [178, 179], and RF effects on electro-optical Mach-Zehnder modulators (MZMs) [180], etc. For example, in the optical radiative crosstalk study [177], it is found that the rough waveguide sidewalls induce radiative coupling region between two passive waveguides separated by a gap (g) much larger than the evanescent field coupling distance ($g < 3$ μm). Field propagating in one waveguide also excites higher order modes in the adjacent waveguide, as shown in Fig. 5.11. The electrical crosstalk study [180] demonstrates that the overall crosstalk between the RF interconnects and the electro-optical phase-shifters originated from radiative and substrate coupling between electrical lines and shared ground connection. Noise in the drive voltage in one of the MZM will convert to optical phase noise, which consequently gives rise to intensity noise at the interferometer output. Therefore, electrical crosstalk significantly affects the optical performance of modulators. Thermal noise contributes to refractive index variations for thermal-tuning modulators, which leads to drift of DC switching voltage and degradation of the dynamic optical extinction ratio (ER, which is a key parameter in optical communications since the quality of the transmissions, the bit error rate (BER), is strictly related to the ER value) [175]. In addition, thermal-mechanical analysis is also critical because the thermal expansion of substrate material will cause misalignment and deteriorate proper current injection and optical coupling. The mismatch of thermal expansion coefficients of different substrates will induce warpage and high-stress points which will distort the overall integration. In addition, the thermal expansion response of the 3D microprinted fiber ferrule also needs further investigation. The expansion of the

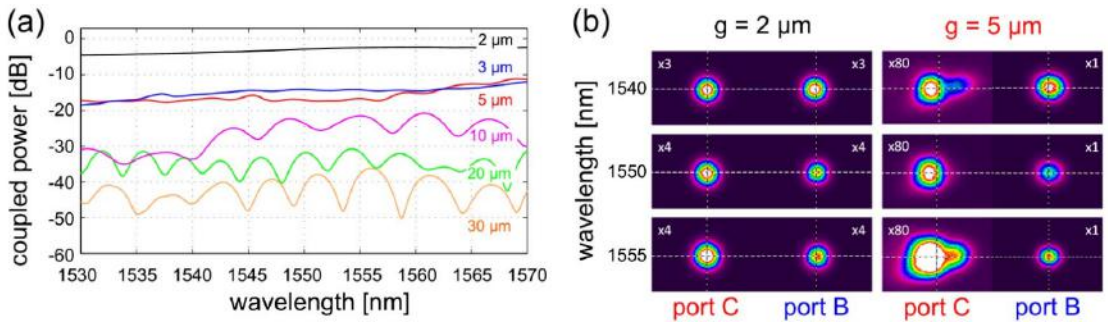


Figure 5.11: (a) Normalized coupled power in the adjacent waveguide measured at different values of the gap g. For $g > 5$ μm, the spectral ripples suggest the emergence of the first higher order mode and the "beating" (interference between two slightly different frequencies) between the fundamental mode and the higher order mode. (b) Near-field measurements at the output ports of two waveguides for three wavelengths. For $g > 5$ μm, multimode propagation appears. Figure is obtained from [177].

fiber ferrule may be smaller or larger than that of the Si carrier substrate, which may cause the sliding of ferrule in the etched pit or the breaking of the Si carrier. Materials with better thermal compliance with Si may be chosen to reduce possible misalignment and damage. As a result, the successful integration of FISCT on electronic packaging requires extensive optical, electrical, thermal and mechanical analysis.

Appendices

APPENDIX A

GRATING PERMITTIVITY DEFINITION

A.1 Parallelogramic Grating

For an arbitrary surface-relief grating, the relative permittivity of the grating region can be expressed as the Fourier series along the x direction

$$\epsilon_g(x) = \epsilon_{gr} + (\epsilon_{rd} - \epsilon_{gr}) \sum_h \tilde{\epsilon}_h \exp(jhKx), \tag{A.1}$$

where $\tilde{\epsilon}_h$ is the hth Fourier coefficient, K is the grating vector magnitude ($K = \frac{2\pi}{\Lambda}$, Λ is the grating period), and ϵ_{gr} and ϵ_{rd} are the permittivity of the grating groove and grating ridge, respectively. A non-binary grating can be horizontally sliced into a total of L sublayers, and each sublayer can be represented by a binary grating. For each sublayer, the matrix of permittivity coefficient $\tilde{\epsilon}_h$ as defined in Eq. (2.17) is no longer symmetric.

Figure A.1 shows two possible parallelogramic gratings, one with slant angle $\varphi < 90^\circ$ (designated as "forward-slanted") and the other with $\varphi > 90^\circ$ (designated as "backward-slanted"). The guided wave is incident on the grating in the $+x$ direction from the left. If $\Delta < W$, where $\Delta = t_g / \tan\varphi$ and $W = (1 - f)\Lambda$ as shown in Fig. A.2, the hth Fourier coefficient of the lth sublayer of the forward-slanted parallelogramic grating can be expressed as

$$\tilde{\epsilon}_{l,h}(x) = \frac{1}{\Lambda} \int_{\Delta - l\delta}^{\Delta - l\delta + f\Lambda} \exp(-jhKx)dx, \tag{A.2}$$

where $\delta = \Delta / L$. If $\Delta > W$, the parallelogramic grating will be separated into two sections, and the top and the bottom sections are divided into L_1 and L_2 slices, respectively. The hth Fourier coefficient of the lth sublayer of the forward-slanted parallelogramic grating is expressed as

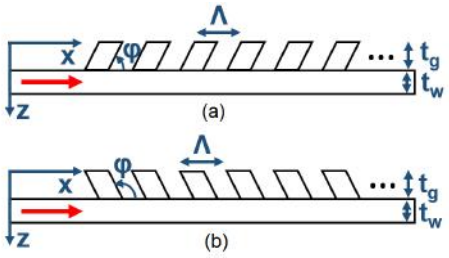


Figure A.1: Schematic representation (not to scale) of a guided wave incident on a parallelogramic grating with (a) slant angle $\varphi < 90°$ or (b) $\varphi > 90°$.

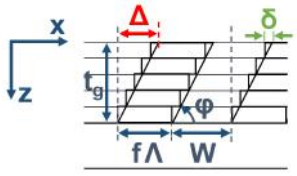


Figure A.2: Sublayers of a forward-slanted parallelogramic grating with $\Delta < W$.

$$\tilde{\epsilon}_{l,h}(x) = \begin{cases} \frac{1}{\Lambda}[\int_0^{\rho - l\delta_1} \exp(-jhKx)dx \\ + \int_{\rho - l\delta_1 + W}^{\Lambda} \exp(-jhKx)dx], & l \in [1, L_1] \\ \frac{1}{\Lambda}\int_{W - l\delta_2}^{\Lambda - l\delta_2} \exp(-jhKx)dx, & l \in [L_1 + 1, L_1 + L_2] \end{cases} \tag{A.3}$$

where $\rho = \Delta - W$, $t_2 = W \tan\varphi$, $t_1 = t_g - t_2$, $\delta_1 = t_1/(L_1 \tan\varphi)$, and $\delta_2 = t_2/(L_2 \tan\varphi)$, and t_1 and t_2 are the thickness of the top and bottom layers, respectively.

Similarly, the $\tilde{\epsilon}_{l,h}(x)$ of the lth sublayer of the backward-slanted parallelogramic grating can be expressed as

$$\tilde{\epsilon}_{l,h}(x) = \frac{1}{\Lambda}\int_{W - \Delta + (l-1)\delta}^{\Lambda - \Delta + (l-1)\delta} \exp(-jhKx)dx \tag{A.4}$$

for $\Delta < W$, and

$$\tilde{\epsilon}_{l,h}(x) = \begin{cases} \frac{1}{\Lambda}[\int_0^{\Lambda - \Delta + (l-1)\delta_1} \exp(-jhKx)dx \\ + \int_{\Lambda - \rho + (l-1)\delta_1}^{\Lambda} \exp(-jhKx)dx], & l \in [1, L_1] \\ \frac{1}{\Lambda}\int_{(l-1)\delta_2}^{(l-1)\delta_2 + f\Lambda} \exp(-jhKx)dx, & l \in [L_1 + 1, L_1 + L_2] \end{cases} \tag{A.5}$$

for $\Delta > W$.

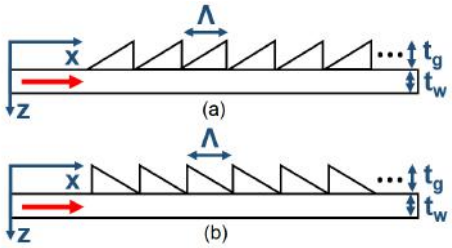


Figure A.3: Schematic representation (not to scale) of a guided wave incident on a sawtooth grating with (a) forward-slanted ridges or (b) backward-slanted ridges.

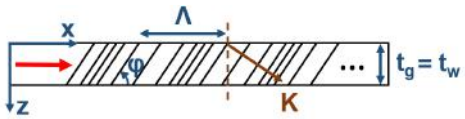


Figure A.4: Schematic representation (not to scale) of a guided wave incident on a volume grating.

A.2 Sawtooth Grating

Figure A.3 shows two types of sawtooth gratings, namely (a) "forward-slanted" and (b) "backward-slanted". The hth Fourier coefficient of the lth sublayer is expressed as

$$\tilde{\epsilon}_{l,h}(x) = \frac{1}{\Lambda}\int_{\Lambda-l\delta}^{\Lambda} \exp(-jhKx)dx \tag{A.6}$$

for the forward-slanted sawtooth grating, and

$$\tilde{\epsilon}_{l,h}(x) = \frac{1}{\Lambda}\int_{0}^{l\delta} \exp(-jhKx)dx \tag{A.7}$$

for the backward-slanted sawtooth grating, where $\delta = \Lambda/L$.

A.3 Volume Grating

Figure A.4 shows the structure of a volume grating with slant angle φ. The permittivity of the volume grating varies sinusoidally, which is expressed as

$$\epsilon_g = \epsilon_{g0} + \Delta\epsilon\cos(K\sin\varphi\, x + K\cos\varphi\, z), \tag{A.8}$$

where ϵ_{g0} is the average permittivity and $\Delta\epsilon$ is the amplitude of the sinusoidal variation. The ith propagation constant in the x direction is defined as

$$k_{x,i} = k_0 n_g \sin\theta' - iK\sin\varphi, \tag{A.9}$$

and the ith propagation constant in the z direction in the volume grating is defined as

$$k_{gz,i} = k_0 n_g \cos\theta' - iK\cos\varphi, \tag{A.10}$$

where $n_g = \sqrt{\epsilon_{g0}}$, and θ' is the 0th order refraction angle inside of the grating ($k_0 n_c \sin\theta = k_0 n_g \sin\theta'$ where $k_0 = \frac{2\pi}{\lambda_0}$ and n_c is the refractive index of the cover). The electric field in the volume grating is expressed by Fourier expansions in spatial harmonics as

$$E_{gy} = \sum_i S_{gy,i} \exp(-jk_{x,i}x - jk_{gz,i}z), \tag{A.11}$$

and it satisfies the wave equation

$$\bigtriangledown^2 E_{gy} + k_0^2 \epsilon_g E_{gy} = 0. \tag{A.12}$$

Substituting Eq. (A.11) into Eq. (A.12) gives

$$\begin{aligned} &\tfrac{\partial^2 S_{gy,i}}{\partial z^2} - j4\pi\left(\tfrac{\sqrt{\epsilon_{g0}}\cos\theta'}{\lambda} - \tfrac{i\cos\varphi}{\Lambda}\right)\tfrac{\partial S_{gy,i}}{\partial z} \\ &+\tfrac{4\pi^2}{\Lambda^2} i(m-i)S_{gy,i} + \tfrac{2\pi^2\Delta\epsilon}{\lambda^2}(S_{gy,i+1} + S_{gy,i-1}) = 0, \end{aligned} \tag{A.13}$$

where $m = \frac{2\Lambda}{\lambda}\sqrt{\epsilon_{g0}}\cos(\theta' - \varphi)$. The permittivity matrix of the volume grating can be expressed in the form of Eq. (A5) in Moharam *et al.* [124], and the eigenvector matrix **W** and eigenvalue matrix **Q** of the permittivity matrix can be calculated. The electric field

expression in the volume grating is expressed as

$$S_{gy,i}(z) = \sum_{p=1}^{2s} C_p w_{i,p} \exp(q_p z), \tag{A.14}$$

where $2s$ is the total number of diffraction orders, $w_{i,p}$ is the (i,p)th element of $\mathbf{W}$ (size $2s \times 2s$), and q_p is the (p,p)th element of $\mathbf{Q}$ (size $2s \times 2s$).

If the volume grating is fabricated in the waveguide, the nonhomogenous system of equations related to the boundary conditions is in the form of

$$\begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{M}_{g1} \\ -j\mathbf{Y}_c & \mathbf{0} & \mathbf{M}_{g2} \\ \mathbf{0} & -\mathbf{I} & \mathbf{M}_{g3} \\ \mathbf{0} & j\mathbf{Y}_s & \mathbf{M}_{g4} \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ \mathbf{T} \\ \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} = \begin{bmatrix} \delta_{i0} \\ a\delta_{i0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{A.15}$$

where $a = -jk_0 n_c \cos\theta$, $\mathbf{Y}_c$ and $\mathbf{Y}_s$ are diagonal matrices with diagonal elements $k_{cz,i}$ and $k_{sz,i}$ defined in Eq. (2.8), respectively, $\mathbf{C}_1$ and $\mathbf{C}_2$ are vectors consisting of C_1 to C_s and C_{s+1} to C_{2s}, respectively, $\mathbf{M}_{g1}$ is the matrix consisting of the first s rows of $\mathbf{W}$, the (i,p)th element of $\mathbf{M}_{g2}$ is $M^{g1}_{i,p}(q_p - jk_{gz,i})$, the (i,p)th element of $\mathbf{M}_{g3}$ is $M^{g1}_{i,p}\exp(q_p - jk_{gz,i})$, and the (i,p)th element of $\mathbf{M}_{g4}$ is $M^{g1}_{i,p}(q_p - jk_{gz,i})\exp(q_p - jk_{gz,i})$. The size of $\mathbf{M}_{g1}$, $\mathbf{M}_{g2}$, $\mathbf{M}_{g3}$ and $\mathbf{M}_{g4}$ are $s \times 2s$, while the others are $s \times s$. If the volume grating is configured above the waveguide, the boundary condition matrix is in the form of Eq. (2.23) with appropriate modifications.

APPENDIX B

3D FINITE-DIFFERENCE TIME-DOMAIN METHOD

Finite-different time-domain (FDTD) method is a time-domain numerical analysis technique for modeling computational electromagnetics and finding approximate solutions based on Maxwell's differential equations. It has been widely used to model complex electromagnetic problems in the field of ultralow-frequency geophysics [181], radar [182], antenna [183], wireless communications [184], biomedical imaging [185], photonic crystals [186], and diffraction optics [187], etc. Kane Yee first introduced the FDTD scheme to electromagnetic problems in 1996 [188]. In Yee's FDTD formulation, Maxwell's curl equations are discretized in time and space by using centered finite difference operators. The space discretization make it flexible to model radiation or scattering from objects with arbitrary shapes.

B.1 Formulation in the Central Region

IN Yee's FDTD formulation, the computational space is discretized into a 3D domain in the size of $N_x \times N_y \times N_z$, where N_x, N_y, and N_z are the total number of cells in the x, y, and z direction, respectively. The domain is populated by staggered unit cells of E and H fields as shown in Fig. B.1. The unit step in the x, y, and z direction are set as Δx, Δy, and Δz, respectively. The unit time step is set as Δt based on the stability condition

$$\Delta t \leq \frac{1}{c\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}}. \tag{B.1}$$

In order to reduce the effect of scattering at domain boundaries, the whole structure is surrounded by a perfectly matched layer (PML).

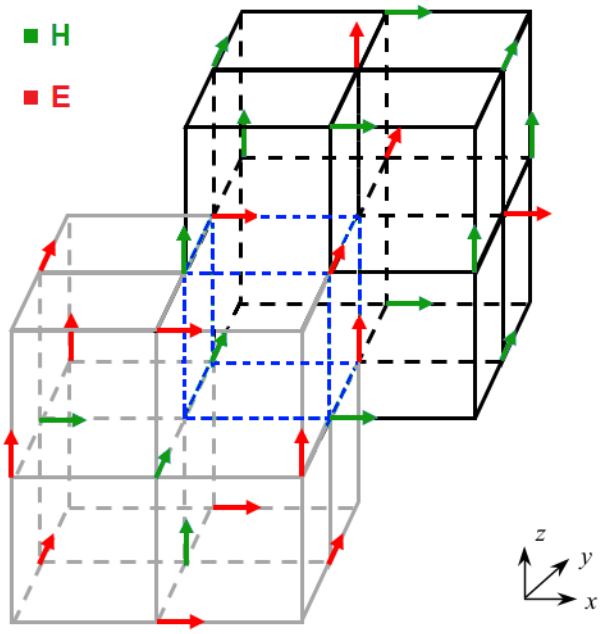


Figure B.1: Staggered FDTD unit cells in which green arrows indicate H fields and red arrows indicate E fields.

Ampere's equation

$$\epsilon_0 \epsilon_r \frac{\partial \bar{E}}{\partial t} = \nabla \times \bar{H} - \bar{J} \tag{B.2}$$

is first separated into the x, y, and z component as follows:

$$\epsilon_0 \epsilon_r \frac{\partial E_x}{\partial t} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - J_x, \tag{B.3}$$

$$\epsilon_0 \epsilon_r \frac{\partial E_y}{\partial t} = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - J_y, \tag{B.4}$$

$$\epsilon_0 \epsilon_r \frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - J_z. \tag{B.5}$$

If input field is polarized in the z direction (assumed for the subsequently derivations), the component J_x and J_y will be zero. Similarly, Faraday's equation

$$\mu_0 \frac{\partial \bar{H}}{\partial t} = -\nabla \times \bar{E} - \bar{M} \tag{B.6}$$

is separated into the x, y, and z component as follows:

$$\mu_0 \frac{\partial H_x}{\partial t} = -\frac{\partial E_z}{\partial y} + \frac{\partial E_y}{\partial z} - M_x, \tag{B.7}$$

$$\mu_0 \frac{\partial H_y}{\partial t} = -\frac{\partial E_x}{\partial z} + \frac{\partial E_z}{\partial x} - M_y, \tag{B.8}$$

$$\mu_0 \frac{\partial H_z}{\partial t} = -\frac{\partial E_y}{\partial x} + \frac{\partial E_x}{\partial y} - M_z, \tag{B.9}$$

where $M_x = M_y = M_z = 0$ for a J_z excited input field.

The time stepping formulas for the field E_x, E_y, E_z, H_x, H_y, and H_z are found by expanding the spatial and time derivatives in Eq. (B.3) to (B.5) and Eq. (B.7) to (B.9) using the central difference and forward difference formula, respectively:

$$\begin{aligned} &E_x^{n+1}(i+\frac{1}{2},j,k) = E_x^n(i+\frac{1}{2},j,k) + \frac{\Delta t}{\epsilon_0 \epsilon_r} \times \\ &\{\frac{1}{\Delta y}[H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) - H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j-\frac{1}{2},k)] \\ &-\frac{1}{\Delta z}[H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) - H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k-\frac{1}{2})]\}, \end{aligned} \tag{B.10}$$

$$\begin{aligned} &E_y^{n+1}(i,j+\frac{1}{2},k) = E_y^n(i,j+\frac{1}{2},k) + \frac{\Delta t}{\epsilon_0 \epsilon_r} \times \\ &\{\frac{1}{\Delta z}[H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) - H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k-\frac{1}{2})] \\ &-\frac{1}{\Delta x}[H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) - H_z^{n+\frac{1}{2}}(i-\frac{1}{2},j+\frac{1}{2},k)]\}, \end{aligned} \tag{B.11}$$

$$
\begin{aligned}
&E_z^{n+1}(i,j,k+\frac{1}{2}) = E_z^n(i,j,k+\frac{1}{2}) + \frac{\Delta t}{\epsilon_0\epsilon_r}\times \\
&\{\frac{1}{\Delta x}[H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) - H_y^{n+\frac{1}{2}}(i-\frac{1}{2},j,k+\frac{1}{2})] \\
&-\frac{1}{\Delta y}[H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) - H_x^{n+\frac{1}{2}}(i,j-\frac{1}{2},k+\frac{1}{2})] \\
&-J_z^{n+1}(i,j,k+\frac{1}{2})\},
\end{aligned}
\tag{B.12}
$$

$$
\begin{aligned}
&H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) = H_x^{n-\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) - \frac{\Delta t}{\mu_0}\times \\
&\{\frac{1}{\Delta y}[E_z^n(i,j+1,k+\frac{1}{2}) - E_z^n(i,j,k+\frac{1}{2})] \\
&-\frac{1}{\Delta z}[E_y^n(i,j+\frac{1}{2},k+1) - E_y^n(i,j+\frac{1}{2},k)]\},
\end{aligned}
\tag{B.13}
$$

$$
\begin{aligned}
&H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) = H_y^{n-\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) - \frac{\Delta t}{\mu_0}\times \\
&\{\frac{1}{\Delta z}[E_x^n(i+\frac{1}{2},j,k+1) - E_x^n(i+\frac{1}{2},j,k)] \\
&-\frac{1}{\Delta x}[E_z^n(i+1,j,k+\frac{1}{2}) - E_z^n(i,j,k+\frac{1}{2})]\},
\end{aligned}
\tag{B.14}
$$

$$
\begin{aligned}
&H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) = H_z^{n-\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) - \frac{\Delta t}{\mu_0}\times \\
&\{\frac{1}{\Delta x}[E_y^n(i+1,j+\frac{1}{2},k) - E_y^n(i,j+\frac{1}{2},k)] \\
&-\frac{1}{\Delta y}[E_x^n(i+\frac{1}{2},j+1,k) - E_x^n(i+\frac{1}{2},j,k)]\}.
\end{aligned}
\tag{B.15}
$$

B.2 Formulation in the PML Region

The time-stepping formulation in the PML region is derived by first defining the relative permittivity or permeability matrix of the artificial PML region as

$$M = \bar{\bar{\epsilon}}_r = \bar{\bar{\mu}}_r = \begin{bmatrix} \frac{s_y s_z}{s_x} & 0 & 0 \\ 0 & \frac{s_x s_z}{s_y} & 0 \\ 0 & 0 & \frac{s_y s_x}{s_z} \end{bmatrix}, \tag{B.16}$$

where $s_q = 1 - j\frac{\sigma_q}{\omega\epsilon_0}$, q = x, y, or z, such that $\bar{\bar{\epsilon}} = \epsilon_0\bar{\bar{\epsilon}}_r$ and $\bar{\bar{\mu}} = \mu_0\bar{\bar{\mu}}_r$. The parameter σ_q is calculated from the following equation

$$\sigma_q(q_i) = -\left(\frac{q_i}{d_q}\right)^m \frac{(m+1)\ln(R)}{2\eta_0 d_q}, \tag{B.17}$$

where q_i indicates the grid location in the q direction, d_q is the thickness of PML in the q direction, R is the desired reflectance, and m is an arbitrary number in the range of [3,4]. The matrix M in Eq. (B.16) can be separated into two matrices as follows:

$$M = M_1 M_2 = \begin{bmatrix} s_y & 0 & 0 \\ 0 & s_z & 0 \\ 0 & 0 & s_x \end{bmatrix} \begin{bmatrix} \frac{s_z}{s_x} & 0 & 0 \\ 0 & \frac{s_x}{s_y} & 0 \\ 0 & 0 & \frac{s_y}{s_z} \end{bmatrix}. \tag{B.18}$$

As a result, the Maxwell's equation $\nabla \times \bar{E} = -j\omega\bar{\bar{\mu}}\bar{H}$ can be separated as

$$\nabla \times \bar{E} = -j\omega M_1 \bar{B}, \tag{B.19}$$

$$\bar{B} = \mu_0 M_2 \bar{H}; \tag{B.20}$$

the Maxwell's equation $\nabla \times \bar{H} = j\omega\bar{\bar{\epsilon}}\bar{E}$ can be separated as

$$\nabla \times \bar{H} = j\omega M_1 \bar{D}, \tag{B.21}$$

$$\bar{D} = \epsilon_0 M_2 \bar{E}. \tag{B.22}$$

Then, the time-stepping formulas in the PML region can be expressed as follows:

$$\begin{aligned} D_x^{n+1}(i+\frac{1}{2},j,k) &= \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0}}\{(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})D_x^n(i+\frac{1}{2},j,k) \\ &+ \frac{1}{\Delta y}[H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) - H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j-\frac{1}{2},k)] \\ &- \frac{1}{\Delta z}[H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) - H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k-\frac{1}{2})]\}, \end{aligned} \tag{B.23}$$

$$\begin{aligned} E_x^{n+1}(i+\frac{1}{2},j,k) &= \frac{1}{\epsilon_0\epsilon_s(\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0})}\times \\ &\{\epsilon_0\epsilon_s(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})E_x^n(i+\frac{1}{2},j,k) \\ &+ (\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0})D_x^{n+1}(i+\frac{1}{2},j,k) \\ &- (\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})D_x^n(i+\frac{1}{2},j,k)\}; \end{aligned} \tag{B.24}$$

$$\begin{aligned} D_y^{n+1}(i,j+\frac{1}{2},k) &= \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0}}\{(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})D_y^n(i,j+\frac{1}{2},k) \\ &+ \frac{1}{\Delta z}[H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) - H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k-\frac{1}{2})] \\ &- \frac{1}{\Delta x}[H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) - H_z^{n+\frac{1}{2}}(i-\frac{1}{2},j+\frac{1}{2},k)]\}, \end{aligned} \tag{B.25}$$

$$
\begin{aligned}
&E_y^{n+1}(i,j+\frac{1}{2},k) = \frac{1}{\epsilon_0\epsilon_s(\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0})}\times \\
&\{\epsilon_0\epsilon_s(\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})E_y^n(i,j+\frac{1}{2},k) \\
&+(\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0})D_y^{n+1}(i,j+\frac{1}{2},k) \\
&-(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})D_y^n(i,j+\frac{1}{2},k)\};
\end{aligned}
\tag{B.26}
$$

$$
\begin{aligned}
&D_z^{n+1}(i,j,k+\frac{1}{2}) = \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0}}\{(\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})D_z^n(i,j,k+\frac{1}{2}) \\
&+\frac{1}{\Delta x}[H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2})-H_y^{n+\frac{1}{2}}(i-\frac{1}{2},j,k+\frac{1}{2})] \\
&-\frac{1}{\Delta y}[H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2})-H_x^{n+\frac{1}{2}}(i,j-\frac{1}{2},k+\frac{1}{2})]\},
\end{aligned}
\tag{B.27}
$$

$$
\begin{aligned}
&E_z^{n+1}(i,j,k+\frac{1}{2}) = \frac{1}{\epsilon_0\epsilon_s(\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0})}\times \\
&\{\epsilon_0\epsilon_s(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})E_z^n(i,j,k+\frac{1}{2}) \\
&+(\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0})D_z^{n+1}(i,j,k+\frac{1}{2}) \\
&-(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})D_z^n(i,j,k+\frac{1}{2})\},
\end{aligned}
\tag{B.28}
$$

where ϵ_s is the permittivity of materials inside of the PML;

$$
\begin{aligned}
&B_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) = \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0}}\times \\
&\{(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})B_x^{n-\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) \\
&-\frac{1}{\Delta y}[E_z^n(i,j+1,k+\frac{1}{2})-E_z^n(i,j,k+\frac{1}{2})] \\
&+\frac{1}{\Delta z}[E_y^n(i,j+\frac{1}{2},k+1)-E_y^n(i,j+\frac{1}{2},k)]\},
\end{aligned}
\tag{B.29}
$$

$$
\begin{aligned}
&H_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) = \frac{1}{\mu_0(\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0})}\times \\
&\{\mu_0(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})H_x^{n-\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) \\
&+(\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0})B_x^{n+\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2}) \\
&-(\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})B_x^{n-\frac{1}{2}}(i,j+\frac{1}{2},k+\frac{1}{2})\};
\end{aligned}
\tag{B.30}
$$

$$
\begin{aligned}
&B_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) = \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0}}\times \\
&\{(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})B_y^{n-\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) \\
&-\frac{1}{\Delta z}[E_x^n(i+\frac{1}{2},j,k+1)-E_x^n(i+\frac{1}{2},j,k)] \\
&+\frac{1}{\Delta x}[E_z^n(i+1,j,k+\frac{1}{2})-E_z^n(i,j,k+\frac{1}{2})]\},
\end{aligned}
\tag{B.31}
$$

$$
\begin{aligned}
&H_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) = \frac{1}{\mu_0(\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0})}\times \\
&\{\mu_0(\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})H_y^{n-\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) \\
&+(\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0})B_y^{n+\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2}) \\
&-(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})B_y^{n-\frac{1}{2}}(i+\frac{1}{2},j,k+\frac{1}{2})\};
\end{aligned}
\tag{B.32}
$$

$$
\begin{aligned}
&B_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) = \frac{1}{\frac{1}{\Delta t}+\frac{\sigma_x}{2\epsilon_0}}\times \\
&\{(\frac{1}{\Delta t}-\frac{\sigma_x}{2\epsilon_0})B_z^{n-\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) \\
&-\frac{1}{\Delta x}[E_y^n(i+1,j+\frac{1}{2},k)-E_y^n(i,j+\frac{1}{2},k)] \\
&+\frac{1}{\Delta y}[E_x^n(i+\frac{1}{2},j+1,k)-E_x^n(i+\frac{1}{2},j,k)]\},
\end{aligned}
\tag{B.33}
$$

$$
\begin{aligned}
H_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) &= \frac{1}{\mu_0(\frac{1}{\Delta t}+\frac{\sigma_y}{2\epsilon_0})}\times \\
&\{\mu_0(\frac{1}{\Delta t}-\frac{\sigma_y}{2\epsilon_0})H_z^{n-\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) \\
&+(\frac{1}{\Delta t}+\frac{\sigma_z}{2\epsilon_0})B_z^{n+\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k) \\
&-(\frac{1}{\Delta t}-\frac{\sigma_z}{2\epsilon_0})B_z^{n-\frac{1}{2}}(i+\frac{1}{2},j+\frac{1}{2},k)\}.
\end{aligned}
\tag{B.34}
$$

Finally, the entire model is truncated using the perfect electric conductor (PEC) boundary condition at all outer surfaces; that is, $E_x = E_y = 0$ at the top and bottom surfaces (xy plane), $E_x = E_z = 0$ at the front and back surfaces (xz plane), and $E_y = E_z = 0$ at the right and left surfaces (yz plane).

Source excitation can be achieved using an electric current source (J_z as an example) at desired locations. The source should be gradually pumped up to avoid the excitation of transient fields and undesirable frequencies.

B.3 Post-Processing of Simulated Fields

The most common post-processing of FDTD simulation is the computation of transmission and reflection spectra. The power P (the amount of energy transferred per unit time) through a certain area (for 3D simulation) or line (for 2D simulation) at a given frequency ω can be calculated by integrating the Poynting vector (the directional energy transfer per unit area per unit time or, equivalently, the directional flux) in the plane normal direction $\hat{n}$ over the area/line as

$$
P(\omega) = \mathrm{Re}\left\{\int \bar{E}_\omega^*(\bar{x}) \times \bar{H}_\omega(\bar{x}) \cdot \hat{n} d^2\bar{x}\right\}, \tag{B.35}
$$

where $\bar{x}$ is a vector specifying the locations in the flux region. Note that the power calculated by Eq. (B.35) has a time dependence. If we are interested in the flux of an input pulse or a guided mode with multiple frequencies, we need to accumulate the Fourier Transforms of $\bar{E}(\bar{x})$ and $\bar{H}(\bar{x})$ at every location in the flux region. Let $f(t) = \bar{E}(\bar{x})$ or $\bar{H}(\bar{x})$. For each point in the flux region, the discrete Fourier transform in the time domain of $f(t)$ is the frequency domain function $F(\omega) = \bar{E}_\omega(\bar{x})$ or $\bar{H}_\omega(\bar{x})$:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \sum_n \exp(j\omega n\Delta t) f(n\Delta t)\Delta t \approx \frac{1}{\sqrt{2\pi}} \int \exp(j\omega t) f(t) dt. \tag{B.36}$$

Then the resulting power at the desired frequency is calculated according to Eq. (B.35) using the Fourier transformed fields $\bar{E}_\omega(\bar{x})$ and $\bar{H}_\omega(\bar{x})$. The power $P(\omega)$ needs to be normalized by relating it with the incident power at each frequency in order to get the meaningful transmission/reflection spectrum. Sometimes, a reference case with no scattering centers needs to be used to determine the reflected flux because the total field calculated at the input port is a superposition of the input field and the reflected field.

Another useful post-processing is finding the modes of waveguiding structures. Eigenmodes can be calculated for a specific cross-section at a given frequency according to the modal methods. To identify the modes simulated by the FDTD, an overlap integral between the simulated field distribution ($\bar{E}_s(\bar{x})$ and $\bar{H}_s(\bar{x})$) and the calculated eigenmodes ($\bar{E}_e(\bar{x})$ and $\bar{H}_e(\bar{x})$) is performed as

$$OI = \mathrm{Re}\left\{ \frac{\left(\int \bar{E}_s(\bar{x}) \times \bar{H}_e^*(\bar{x}) \cdot \hat{n} d^2\bar{x}\right)\left(\int \bar{E}_e(\bar{x}) \times \bar{H}_s^*(\bar{x}) \cdot \hat{n} d^2\bar{x}\right)}{\left(\int \bar{E}_s(\bar{x}) \times \bar{H}_s^*(\bar{x}) \cdot \hat{n} d^2\bar{x}\right)\left(\int \bar{E}_e(\bar{x}) \times \bar{H}_e^*(\bar{x}) \cdot \hat{n} d^2\bar{x}\right)} \right\}. \tag{B.37}$$

Many post-processing operations, including the two discussed above, are available in the open source software MEEP and other commercial software, such as Lumerical FDTD, and Synopsys FullWAVE, etc.

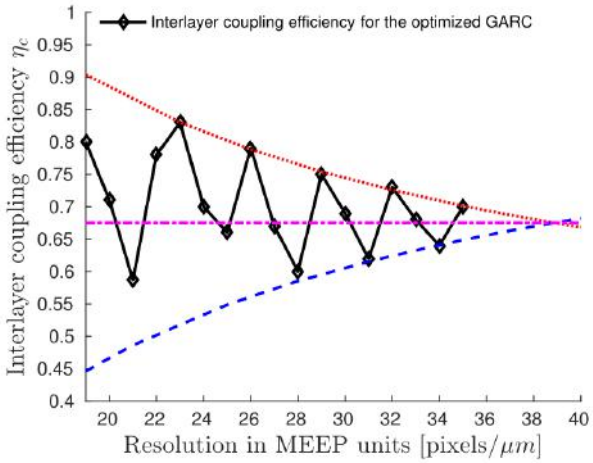


Figure B.2: Interlayer efficiency of GARC as a function of 3D FDTD resolution. The approximately converged value is 68%.

B.4 Effect of Resolution on FDTD Results

The accuracy of the 3D FDTD calculation is largely determined by the resolution, or number of pixels (grids) per unit distance. The 3D FDTD is based on a set of cubic grids, which cannot properly represent a slanted or curved surface; instead, a staircase substitution is necessary to model the rotated/curved 3D geometry. This requires relatively high resolution to correctly calculate the fields and fluxes. Typically, the results display an oscillating and converging trend, as the resolution of FDTD simulation increases. The oscillatory behavior is due largely to the effects of the finite pixel size and the slight mismatches between the computational grid and the device structure. Subpixel averaging, a process to assign carefully designed average values of the permittivity to pixels at the structural boundaries in MEEP [189], may be incorporated to give more stable results. Usually at least 10 pixels/wavelength in the highest dielectric should be used. In Chapter 3, interlayer coupling efficiency of the GARC coupler was simulated using the MEEP 3D FDTD software package. This was done for a range of resolutions from 19 to 35 pixels/μm. From these calculations, the efficiency at convergence is approximately 68%, as shown in Fig. B.2. The simulation results in the sensitivity analysis are obtained using the resolution 20 pixels/μm considering the time-accuracy trade-off.

APPENDIX C

2D FINITE-DIFFERENCE BEAM PROPAGATION METHOD

Beam propagation method (BPM) has been successfully used to analyze guided wave structures, such as waveguide branches, waveguide couplers, optical fibers, etc., in which the beam propagation direction doesn't deviate too much from the optical axis. In the BPM formulation, the wave guiding structure is reduced to a 1D cross-sectional index profile by defining effective indices to various parts of the structure. Then the paraxial wave equation is solved with respect to those effective indices at different locations along the optical axis. Even though the effective index and paraxial approximations are not adequate for certain cases, BPM can still be applied to get qualitative predictions about the behavior of a specific guided-wave structure. For more accurate results, a full-wave simulation such as FDTD should be used, but this requires extensive computational load. Both fast Fourier transform (FFT) and finite difference (FD) method have been proposed to solve the partial differential equation (paraxial wave equation) [190]. Here we will briefly introduce and compare these two methods.

C.1 Derivation of Paraxial Wave Equation

Prior to the introduction of BPM, the scalar wave equation is derived first. Starting from the Maxwell's equations (Eq. (B.2) and Eq. (B.6)) and auxiliary functions, the vector wave equation

$$\nabla \times \nabla \times \bar{E} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\bar{\bar{\epsilon}}_r \cdot \bar{E} = 0 \tag{C.1}$$

can be derived. By using the vector identity

$$\nabla \times \left(\nabla \times \bar{A}\right) = \nabla\left(\nabla \cdot \bar{A}\right) - \nabla^2 \bar{A}, \tag{C.2}$$

the scalar wave equation (Helmholtz equation)

$$\nabla^2 \xi + k_0^2 \epsilon_r \xi = 0 \tag{C.3}$$

is obtained for each component of $\bar{E}$. In the paraxial limit, the index modulation δn is assumed to be small, so Eq. (C.3) can be rewritten as

$$\nabla^2 \xi + k_0^2 \left(n_0 + \delta n\right)^2 \xi = 0. \tag{C.4}$$

Now assume the waveguide structure, whose index profile is represented by $n(x,z) = n_0 + \delta n$, carries TE-polarized field. The only electric field component ξ_y can be expressed as follows:

$$\xi_y(x,y,z) = E_y(x,y,z)\exp(-jk_0 n_0 z), \tag{C.5}$$

where $E_y(x,y,z)$ is the field amplitude distribution (beam envelope). Assume the field is uniform across the y dimension ($\partial/\partial y = 0$). Substitute Eq. (C.5) into Eq. (C.4), resulting in

$$\left(\frac{\partial^2}{\partial z^2} - 2jk_0 n_0 \frac{\partial}{\partial z} - (k_0 n_0)^2 + \frac{\partial^2}{\partial x^2}\right) E_y + k_0^2 \left[n_0^2 + 2n_0 \delta n + (\delta n)^2\right] E_y = 0, \tag{C.6}$$

where the $\partial^2/\partial z^2$ term can be treated as zero by the slow-varying envelope approximation, and $(\delta n)^2 = 0$ due to small index perturbation (so $2n_0 \delta n \approx n^2(x,z) - n_0^2$). As a result, Helmholtz equation can be reduced to the paraxial wave equation:

$$2jk_0 n_0 \frac{\partial E_y}{\partial z} = \frac{\partial^2 E_y}{\partial x^2} + k_0^2 [n^2(x,z) - n_0^2] E_y. \tag{C.7}$$

C.2 Finite-Difference BPM

In the FD-BPM approach, the 2D simulation space is divided into grids with sides Δx and Δz. The partial differential equation Eq. (C.7) is represented by the finite difference approximation

$$2jk_0n_0\frac{\partial E_i}{\partial z} = \frac{E_{i-1} - 2E_i + E_{i+1}}{\Delta x^2} + k_0^2[n_i^2(z) - n_0^2]E_i, \tag{C.8}$$

where E_i is the electric field E_y at ($i\Delta x$, z) location, and i is the location index along the transverse direction x. Integrate Eq. (C.8) in the interval [z, $z + \Delta z$] and approximate the integration by trapezoidal rule, the electric field at location $z + \Delta z$ can be related to the electric field at location z by

$$-aE_{i-1}(z+\Delta z)+bE_i(z+\Delta z)-aE_{i+1}(z+\Delta z) = aE_{i-1}(z)+cE_i(z)+aE_{i+1}(z), \tag{C.9}$$

where

$$a = \frac{\Delta z}{2\Delta x^2}, \tag{C.10}$$

$$b = \frac{\Delta z}{\Delta x^2} - \frac{\Delta z}{2}\left[n_i^2(z+\Delta z) - n_0^2\right] + 2jk_0n_0, \tag{C.11}$$

$$c = -\frac{\Delta z}{\Delta x^2} + \frac{\Delta z}{2}\left[n_i^2(z) - n_0^2\right] + 2jk_0n_0. \tag{C.12}$$

C.3 Fast Fourier Transform BPM

In the FFT-BPM approach, Eq. (C.7) can be separated into the diffraction contribution

$$\frac{\partial E_y(x,z)}{\partial z} = \frac{-j}{k_0n_0}\frac{\partial^2 E_y(x,z)}{\partial x^2}, \tag{C.13}$$

and the refraction contribution

$$\begin{aligned}\frac{\partial E_y(x,z)}{\partial z} &= \frac{-j}{k_0 n_0} k_0^2 (n^2 - n_0^2) E_y(x,z) \\ &= \frac{-j}{n_0} k_0 (n^2 - n_0^2) E_y(x,z). \end{aligned} \tag{C.14}$$

Fourier transforming Eq. (C.13) results in

$$\frac{\partial \tilde{E}_y(k_x,z)}{\partial z} = \frac{-j}{k_0 n_0}(-k_x^2)\tilde{E}_y(k_x,z), \tag{C.15}$$

where k_x indicates reciprocal space of x and $\tilde{E}_y$ represents the Fourier transform of E_y. Then a solution to Eq. (C.15) is

$$\tilde{E}_y(k_x,z) = \exp\left[\frac{-j}{k_0 n_0}(-k_x^2)z\right]. \tag{C.16}$$

The Fourier-transformed field at the next location $\tilde{E}_y(k_x, z+\Delta z)$ is then related to the transformed field $\tilde{E}_y(k_x,z)$ at the current location as

$$\tilde{E}_y(k_x, z+\Delta z) = \tilde{E}_y(k_x,z)\exp\left[\frac{-j}{k_0 n_0}(-k_x^2)\Delta z\right]. \tag{C.17}$$

Similarly, the solution to Eq.(C.14) is easy to obtain as

$$E_y(x,z) = \exp\left[\frac{-j}{n_0} k_0 (n^2 - n_0^2) z\right], \tag{C.18}$$

and the field at the next z location $E_y(x, z+\Delta z)$ is

$$E_y(x, z+\Delta z) = E_y(x,z)\exp\left[\frac{-j}{n_0} k_0 (n^2 - n_0^2)\Delta z\right]. \tag{C.19}$$

For the diffraction contribution along, the iterative formulation for $E_y(x, z + \Delta z)$ is

$$\begin{aligned} E_y(x, z + \Delta z) &= FT^{-1}\left\{\tilde{E}_y(k_x, z + \Delta z)\right\} \\ &= FT^{-1}\left\{\tilde{E}_y(k_x, z)\exp\left[\frac{-j}{k_0 n_0}(-k_x^2)\Delta z\right]\right\} \\ &= FT^{-1}\left\{FT\left\{E_y(x, z)\right\}\exp\left[\frac{jk_x^2}{k_0 n_0}\Delta z\right]\right\}. \end{aligned} \tag{C.20}$$

Now add the refraction contribution to Eq. (C.20):

$$E_y(x, z + \Delta z) = FT^{-1}\left\{FT\left\{E_y(x, z)\right\}\exp\left[\frac{jk_x^2}{k_0 n_0}\Delta z\right]\right\}\exp\left[\frac{-j}{n_0}k_0(n^2 - n_0^2)\Delta z\right]. \tag{C.21}$$

Equation (C.21) is the final FFT-BPM formulation which can be solved by fast Fourier transform.

For both FD-BPM and FFT-BPM, the accuracy depends on the resolution in the transverse direction Δx and the longitudinal direction Δz. It is found that FFT-BPM degrades faster at smaller Δz, while the degradation in FD-BPM is slower and well behaved. In addition, the numerical error of FFT-BPM exibits an oscillatory behavior as a function of Δz, while the FD-BPM result is smooth [190].

REFERENCES

[1] *The National Technology Roadmap for Semiconductors – Technology Needs*, Semiconductor Industry Association, 1997.

[2] M. Glick, L. C. Kimmerling, and R. C. Pfahl, "A roadmap for integrated photonics," *Opt. Photon. News*, vol. 29, no. 3, pp. 36–41, 2018.

[3] *AMD EPYC empowers signle-socket servers*, https://www.amd.com/system/files/2017-05/TIRIAS-AMD-Single-Socket-Server.pdf, May 16, 2017.

[4] *Cisco visual networking index: forecast and trends, 20172022*, Document ID:1543-280537836565, 2018.

[5] R. Mahajan, R. Sankman, N. Patel, D.-W. Kim, K. Aygun, Z. Qian, Y. Mekonnen, I. Salama, S. Sharan, D. Iyengar, *et al.*, "Embedded multi-die interconnect bridge (EMIB)–a high density, high bandwidth packaging interconnect," in *66th IEEE Electron. Compon. Technol. Conf.*, 2016, pp. 557–565.

[6] *Intel Stratix 10 MX devices solve the memory bandwidth challenge*, https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/wp/wp-01264-stratix10mx-devices-solve-memory-bandwidth-challenge.pdf.

[7] *UltraScale architecture and product data sheet: overview*, https://www.xilinx.com/support/documentation/data_sheets/ds890-ultrascale-overview.pdf.

[8] *MongoDB and AMD EPYC for the intelligent operational data platform*, https://www.amd.com/system/files/documents/mongodb-and-amd-epyc-for-the-intelligent-operational-data-platform.pdf, Oct. 2018.

[9] *NVIDIA TESLA V100 GPU accelerator*, https://images.nvidia.com/content/technologies/volta/pdf/tesla-volta-v100-datasheet-letter-fnl-web.pdf.

[10] *AMD unveils worlds first 7 nm gaming GPU delivering exceptional performance and incredible experiences for gamers, creators and enthusiasts*, https://www.amd.com/en/press-releases/2019-01-09-amd-unveils-

world-s-first-7nm-gaming-gpu-delivering-exceptional, Jan. 9, 2019.

[11] *Stratix 10 MX FPGA development kit*, https://www.intel.com/content/www/us/en/programmable/products/boards_and_kits/dev-kits/altera/s10_fpga_dev_kit.html.

[12] *Want to get on board the Xilinx UltraScale+ FPGA express now? BittWare ships two boards to help you do just that*, https://forums.xilinx.com/t5/Xcell-Daily-Blog-Archived/Want-to-get-on-board-the-Xilinx-UltraScale-FPGA-express-now/ba-p/727882, Oct. 11, 2016.

[13] *Intel Stratix 10 MX FPGA highlights*, https://www.semiwiki.com/forum/content/6194-intel-stratix-10-mx-fpga-highlights.html, Sep. 15, 2016.

[14] *Intel Xeon Platinum 8176 Scalable processor review*, https://www.tomshardware.com/reviews/intel-xeon-platinum-8176-scalable-cpu,5120.html, Jul. 17, 2017.

[15] *AMD EPYC 7351P linux performance: 16 core / 32 thread server CPU for $750*, https://www.phoronix.com/scan.php?page=article&item=amd-epyc-7351p&num=1, Oct. 19, 2017.

[16] *NVIDIA bumps all TESLA V100 models to 32 GB, effective immediately*, https://www.anandtech.com/show/12576/nvidia-bumps-all-tesla-v100-models-to-32gb, Mar. 27, 2018.

[17] *NVIDIA TESLA V100 (with specs)*, https://nvidianews.nvidia.com/file?fid=5b19b7a02cfac270e487b566.

[18] *AMD Radeon VII graphics card will support linux on launch day*, https://www.techradar.com/news/amd-radeon-vii-graphics-card-will-support-linux-on-launch-day, Jan. 22, 2019.

[19] *Intel Xeon Platinum 8176 processor*, https://ark.intel.com/content/www/us/en/ark/products/120508/intel-xeon-platinum-8176-processor-38-5m-cache-2-10-ghz.html.

[20] *AMD EPYC SoC delivers exceptional results on the STREAM benchmark on 2P servers*, https://www.amd.com/system/files/2017-06/AMD-EPYC-SoC-Delivers-Exceptional-Results.pdf, Jun, 2017.

[21] *Intel Xeon processor E5-2699A v4*, https://ark.intel.com/content/www/us/en/ark/products/96899/intel-xeon-processor-e5-2699a-v4-55m-cache-2-40-ghz.html.

[22] *Intel Stratix 10 MX FPGAs*, https://www.intel.com/content/www/us/en/products/programmable/sip/stratix-10-mx.html.

[23] J. Fowers, K. Ovtcharov, M. Papamichael, T. Massengill, M. Liu, D. Lo, S. Alkalay, M. Haselman, L. Adams, M. Ghandi, *et al.*, "A configurable cloud-scale DNN processor for real-time AI," in *Proc. 45th Annu. Int. Symp. Comput. Archit.*, 2018, pp. 1–14.

[24] *AMD Radeon VII graphics card*, https://www.amd.com/en/products/graphics/amd-radeon-vii.

[25] *AMD Radeon VII*, https://www.techpowerup.com/gpu-specs/radeon-vii.c3358, Mar 14, 2019.

[26] *Virtex UltraScale+ HBM VCU128-ES1 FPGA evaluation kit*, https://www.xilinx.com/products/boards-and-kits/vcu128-es1.html.

[27] D. A. Miller, "Optical interconnects to electronic chips," *Appl. Opt.*, vol. 49, no. 25, F59–F70, 2010.

[28] ——, "Rationale and challenges for optical interconnects to electronic chips," *Proc. IEEE*, vol. 88, no. 6, pp. 728–749, 2000.

[29] D. A. Miller and H. M. Ozaktas, "Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of the system architecture," *J Parallel Distrib. Comput.*, vol. 41, no. 1, pp. 42–52, 1997.

[30] W. J. Dally, M.-J. E. Lee, F.-T. An, J. Poulton, and S. Tell, "High-performance electrical signaling," in *IEEE Proc. 5th Int. Conf. Massively Parallel Process.*, 1998, pp. 11–16.

[31] A. V. Krishnamoorthy, K. W. Goossen, W. Jan, X. Zheng, R. Ho, G. Li, R. Rozier, F. Liu, D. Patil, J. Lexau, *et al.*, "Progress in low-power switched optical interconnects," *IEEE J. Sel. Top. Quantum Electron.*, vol. 17, no. 2, pp. 357–376, 2011.

[32] D. T. Neilson, "Photonics for switching and routing," *IEEE J. Sel. Top. Quantum Electron.*, vol. 12, no. 4, pp. 669–678, 2006.

[33] *Intel Silicon Photonics 100G CWDM4 optical transceiver brief*, https://www.intel.com/content/www/us/en/architecture-and-technology/

silicon-photonics/optical-transceiver-100g-cwdm4-qsfp28-brief.html.

[34] *Cisco announces intent to acquire silicon photonics leader, Luxtera*, https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1959037, Dec. 18, 2018.

[35] N. Kim, D. Wu, D. Kim, A. Rahman, and P. Wu, "Interposer design optimization for high frequency signal transmission in passive and active interposer using through silicon via (TSV)," in *61st IEEE Electron. Compon. Technol. Conf.*, 2011, pp. 1160–1167.

[36] H. S. Yang, C. Zhang, and M. S. Bakir, "Self-aligned silicon interposer tiles and silicon bridges using positive self-alignment structures and rematable mechanically flexible interconnects," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 4, no. 11, pp. 1760–1768, 2014.

[37] X. Zhang, P. K. Jo, M. Zia, G. S. May, and M. S. Bakir, "Heterogeneous interconnect stitching technology with compressible microinterconnects for dense multi-die integration," *IEEE Electron Device Lett.*, vol. 38, no. 2, pp. 255–257, 2017.

[38] J. Q. Lu, "3D hyperintegration and packaging technologies for micro-nano systems," *Proc. IEEE*, vol. 97, no. 1, pp. 18–30, 2009.

[39] K. W. Lee, A. Noriki, K. Kiyoyama, T. Fukushima, T. Tanaka, and M. Koyanagi, "Three-dimensional hybrid integration technology of CMOS, MEMS, and photonics circuits for optoelectronic heterogeneous integrated systems," *IEEE Trans. Electron Devices*, vol. 58, no. 3, pp. 748–757, 2011.

[40] C. Erdmann, D. Lowney, A. Lynam, A. Keady, J. McGrath, E. Cullen, D. Breathnach, D. Keane, P. Lynch, M. De La Torre, *et al.*, "A heterogeneous 3D-IC consisting of two 28 nm FPGA die and 32 reconfigurable high-performance data converters," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 258–269, 2015.

[41] C. Wan, T. K. Gaylord, and M. S. Bakir, "Grating design for interlayer optical interconnection of in-plane waveguides," *Appl. Opt.*, vol. 55, no. 10, pp. 2601–2610, 2016.

[42] G. Roelkens, D. Vermeulen, D. Van Thourhout, R. Baets, S. Brision, P. Lyan, P. Gautier, and J. M. Fedeli, "High efficiency diffractive grating couplers for interfacing a single mode optical fiber with a nanophotonic silicon-on-insulator waveguide circuit," *Appl. Phys. Lett.*, vol. 92, no. 13, p. 131 101, 2008.

[43] C. Wan, T. K. Gaylord, and M. S. Bakir, "RCWA-EIS method for interlayer grating coupling," *Appl. Opt.*, vol. 55, no. 22, pp. 5900–5908, 2016.

[44] G. Roelkens, D. Vermeulen, S. Selvaraja, R. Halir, W. Bogaerts, and D. Van Thourhout, "Grating-based optical fiber interfaces for silicon-on-insulator photonic integrated circuits," *IEEE J. Sel. Top. Quantum Electron.*, vol. 17, no. 3, pp. 571–580, 2011.

[45] X. Chen, C. Li, C. K. Fung, S. M. Lo, and H. K. Tsang, "Apodized waveguide grating couplers for efficient coupling to optical fibers," *IEEE Photon. Technol. Lett.*, vol. 22, no. 15, pp. 1156–1158, 2010.

[46] D. Taillaert, W. Bogaerts, P. Bienstman, T. F. Krauss, P. V. Daele, I. Moerman, S. Verstuyft, K. D. Mesel, and R. Baets, "An out-of-plane grating coupler for efficient butt-coupling between compact planar waveguides and single-mode fibers," *IEEE J. Quantum Electron.*, vol. 38, no. 7, pp. 949–955, 2002.

[47] M. Dai, L. Ma, Y. Xu, M. Lu, X. Liu, and Y. Chen, "Highly efficient and perfectly vertical chip-to-fiber dual-layer grating coupler," *Opt. Express*, vol. 23, no. 2, pp. 1691–1698, 2015.

[48] D. Benedikovic, C. Alonso-Ramos, P. Cheben, J. H. Schmid, S. Wang, D. X. Xu, J. Lapointe, S. Janz, R. Halir, A. Ortega-Moñux, J. M. Fedeli, S. Janz, I. Molina-Fernandez, and M. Dado, "High-directionality fiber-chip grating coupler with interleaved trenches and subwavelength index-matching structure," *Opt. Lett.*, vol. 40, no. 18, pp. 4190–4193, 2015.

[49] T. Watanabe, M. Ayata, U. Koch, Y. Fedoryshyn, and J. Leuthold, "Perpendicular grating coupler based on a blazed antiback-reflection structure," *IEEE J. Lightwave Technol.*, vol. 35, no. 21, pp. 4663–4669, 2017.

[50] F. Van Laere, T. Claes, J. Schrauwen, S. Scheerlinck, W. Bogaerts, D. Taillaert, L. O'Faolain, D. Van Thourhout, and R. Baets, "Compact focusing grating couplers for silicon-on-insulator integrated circuits," *IEEE Photon. Technol. Lett.*, vol. 19, no. 23, pp. 1919–1921, 2007.

[51] C. Wan, T. K. Gaylord, and M. S. Bakir, "Rigorous coupled-wave analysis equivalent-index-slab method for analyzing 3D angular misalignment in interlayer grating couplers," *Appl. Opt.*, vol. 55, no. 35, pp. 10 006–10 015, 2016.

[52] M. Antelius, K. B. Gylfason, and H. Sohlström, "An apodized SOI waveguide-to-fiber surface grating coupler for single lithography silicon photonics," *Opt. Express*, vol. 19, no. 4, pp. 3592–3598, 2011.

[53] Y. Tang, Z. Wang, L. Wosinski, U. Westergren, and S. He, "Highly efficient nonuniform grating coupler for silicon-on-insulator nanophotonic circuits," *Opt. Lett.*, vol. 35, no. 8, pp. 1290–1292, 2010.

[54] J. H. Kang, Y. Atsumi, Y. Hayashi, J. Suzuki, Y. Kuno, T. Amemiya, N. Nishiyama, and S. Arai, "50 Gbps data transmission through amorphous silicon interlayer grating couplers with metal mirrors," *Appl. Phys. Express*, vol. 7, no. 3, p. 032 202, 2014.

[55] S. K. Selvaraja, D. Vermeulen, M. Schaekers, E. Sleeckx, W. Bogaerts, G. Roelkens, P. Dumon, D. Van Thourhout, and R. Baets, "Highly efficient grating coupler between optical fiber and silicon photonic circuit," in *Conf. Lasers Electro-Optics*, Optical Society of America, 2009, CTuC6.

[56] D. Taillaert, P. Bienstman, and R. Baets, "Compact efficient broadband grating coupler for silicon-on-insulator waveguides," *Opt. Lett.*, vol. 29, no. 23, pp. 2749–2751, 2004.

[57] D. Taillaert, F. Van Laere, M. Ayre, W. Bogaerts, D. Van Thourhout, P. Bienstman, and R. Baets, "Grating couplers for coupling between optical fibers and nanophotonic waveguides," *Jpn. J. Appl. Phys.*, vol. 45, no. 8R, pp. 6071–6077, 2006.

[58] J. Bolten, J. Hofrichter, N. Moll, S. Schönenberger, F. Horst, B. J. Offrein, T. Wahlbrink, T. Mollenhauer, and H. Kurz, "CMOS compatible cost-efficient fabrication of SOI grating couplers," *Microelectron. Eng.*, vol. 86, no. 4, pp. 1114–1116, 2009.

[59] B. Schmid, A. Petrov, and M. Eich, "Optimized grating coupler with fully etched slots," *Opt. Express*, vol. 17, no. 13, pp. 11 066–11 076, 2009.

[60] X. Chen, D. J. Thomson, L. Crudginton, A. Z. Khokhar, and G. T. Reed, "Dual-etch apodised grating couplers for efficient fibre-chip coupling near 1310 nm wavelength," *Opt. Express*, vol. 25, no. 15, pp. 17 864–17 871, 2017.

[61] L. Yu, L. Liu, Z. Zhou, and X. Wang, "High efficiency binary blazed grating coupler for perfectly-vertical and near-vertical coupling in chip level optical interconnections," *Opt. Commun.*, vol. 355, pp. 161–166, 2015.

[62] S Lardenois, D Pascal, L Vivien, E Cassan, S Laval, R Orobtchouk, M Heitzmann, N Bouzaida, and L Mollard, "Low-loss submicrometer silicon-on-insulator rib waveguides and corner mirrors," *Opt. Lett.*, vol. 28, no. 13, pp. 1150–1152, 2003.

[63] J. Notaros, F. Pavanello, M. T. Wade, C. M. Gentry, A. Atabaki, L. Alloatti, R. J. Ram, and M. A. Popović, "Ultra-efficient CMOS fiber-to-chip grating couplers," in *Optical Fiber Commun. Conf.*, 2016, pp. 1–3.

[64] H. L. Tseng, E. Chen, H. Rong, and N. Na, "High-performance silicon-on-insulator grating coupler with completely vertical emission," *Opt. Express*, vol. 23, no. 19, pp. 24 433–24 439, 2015.

[65] T. Barwicz, Y. Taira, T. W. Lichoulas, N. Boyer, Y. Martin, H. Numata, J. W. Nah, S. Takenobu, A. Janta-Polczynski, E. L. Kimbrell, *et al.*, "A novel approach to photonic packaging leveraging existing high-throughput microelectronic facilities," *IEEE J. Sel. Top. Quantum Electron.*, vol. 22, no. 6, pp. 455–466, 2016.

[66] R. Hauffe, U Siebel, K Petermann, R Moosburger, J. R. Kropp, and F Arndt, "Methods for passive fiber chip coupling of integrated optical devices," *IEEE Trans. Adv. Packag.*, vol. 24, no. 4, pp. 450–455, 2001.

[67] T Shoji, T Tsuchizawa, T Watanabe, K Yamada, and H Morita, "Spot-size converter for low-loss coupling between 0.3-um-square Si wire waveguides and single-mode fibers," in *15th Ann. Mtg. IEEE Lasers Electro-Opt. Soc.*, vol. 1, 2002, pp. 289–290.

[68] G. Roelkens, P. Dumon, W. Bogaerts, D. Van Thourhout, and R. Baets, "Efficient silicon-on-insulator fiber coupler fabricated using 248-nm-deep UV lithography," *IEEE Photon. Technol. Lett.*, vol. 17, no. 12, pp. 2613–2615, 2005.

[69] H. Melkonyan, K. Sloyan, K. Twayana, P. Moreira, and M. S. Dahlem, "Efficient fiber-to-waveguide edge coupling using an optical fiber axicon lens fabricated by focused ion beam," *IEEE Photon. J.*, vol. 9, no. 4, pp. 1–9, 2017.

[70] F Schiappelli, R Kumar, M Prasciolu, D Cojoc, S Cabrini, M De Vittorio, G Visimberga, A Gerardino, V Degiorgio, and E Di Fabrizio, "Efficient fiber-to-waveguide coupling by a lens on the end of the optical fiber fabricated by focused ion beam milling," *Microelectron. Eng.*, vol. 73, pp. 397–404, 2004.

[71] H. Melkonyan, K. Al Qubaisi, K. Sloyan, A. Khilo, and M. S. Dahlem, "Gradient-index optical fiber lens for efficient fiber-to-chip coupling," *Opt. Express*, vol. 25, no. 12, pp. 13 035–13 045, 2017.

[72] P. H. Howerton and D. Cordray, "Diode pumping of a solid state laser using evanescent field fiber optic coupling: a proposed technique," *Appl. Opt.*, vol. 30, no. 15, pp. 1911–1915, 1991.

[73] J. Yu, S. Jin, Q. Wei, Z. Zang, H. Lu, X. He, Y. Luo, J. Tang, J. Zhang, and Z. Chen, "Hybrid optical fiber add-drop filter based on wavelength dependent light coupling between micro/nano fiber ring and side-polished fiber," *Sci. Rep.*, vol. 5, p. 7710, 2015.

[74] S. J. McNab, N. Moll, and Y. A. Vlasov, "Ultra-low loss photonic integrated circuit with membrane-type photonic crystal waveguides," *Opt. Express*, vol. 11, no. 22, pp. 2927–2939, 2003.

[75] J. Cardenas, C. B. Poitras, K. Luke, L.-W. Luo, P. A. Morton, and M. Lipson, "High coupling efficiency etched facet tapers in silicon waveguides," *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2380–2382, 2014.

[76] Z. Lu, P. Yin, and K. Shi, "Fiber-to-waveguide and 3D chip-to-chip light coupling based on bent metal-clad waveguides," *arXiv preprint arXiv:1606.00417*, 2016.

[77] T. Barwicz, A. Janta-Polczynski, M. Khater, Y. Thibodeau, R. Leidy, J. Maling, S. Martel, S. Engelmann, J. S. Orcutt, P. Fortier, *et al.*, "An O-band metamaterial converter interfacing standard optical fibers to silicon nanophotonic waveguides," in *Opt. Fiber Commun. Conf.*, 2015, Th3F–3.

[78] K. K. Lee, D. R. Lim, D. Pan, C. Hoepfner, W.-Y. Oh, K. Wada, L. C. Kimerling, K. P. Yap, *et al.*, "Mode transformer for miniaturized optical circuits," *Opt. Lett.*, vol. 30, no. 5, pp. 498–500, 2005.

[79] P. I. Dietrich, M Blaicher, I Reuter, M Billah, T Hoose, A Hofmann, C Caer, R Dangel, B Offrein, U Troppenz, *et al.*, "In situ 3D nanoprinting of free-form coupling elements for hybrid photonic integration," *Nat. Photonics*, p. 1, 2018.

[80] A. V. Krishnamoorthy, H. D. Thacker, O. Torudbakken, S. Müller, A. Srinivasan, P. J. Decker, H. Opheim, J. E. Cunningham, I. Shubin, X. Zheng, *et al.*, "From chip to cloud: optical interconnects in engineered systems," *J. Light. Technol.*, vol. 35, no. 15, pp. 3103–3115, 2017.

[81] S. Feng, T. Lei, H. Chen, H. Cai, X. Luo, and A. W. Poon, "Silicon photonics: from a microresonator perspective," *Laser Photonics Rev.*, vol. 6, no. 2, pp. 145–177, 2012.

[82] H. Park, A. W. Fang, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "A hybrid AlGaInAs–silicon evanescent amplifier," *IEEE Photon. Technol. Lett.*, vol. 19, no. 4, pp. 230–232, 2007.

[83] A. W. Fang, H. Park, Y. H. Kuo, R. Jones, O. Cohen, D. Liang, O. Raday, M. N. Sysak, M. J. Paniccia, and J. E. Bowers, "Hybrid silicon evanescent devices," *Mater. Today*, vol. 10, no. 7-8, pp. 28–35, 2007.

[84] R. Halir, A. Ortega-Monux, J. H. Schmid, C. Alonso-Ramos, J. Lapointe, D. X. Xu, J. G. Wanguemert-Perez, I. Molina-Fernandez, and S. Janz, "Recent advances in silicon waveguide devices using sub-wavelength gratings," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 4, pp. 279–291, 2014.

[85] J. Kang, Y. Atsumi, M. Oda, T. Amemiya, N. Nishiyama, and S. Arai, "Layer-to-layer grating coupler based on hydrogenated amorphous silicon for three-dimensional optical circuits," *Jpn. J. Appl. Phys.*, vol. 51, no. 12R, p. 120 203, 2012.

[86] Y. Tang, D. Dai, and S. He, "Proposal for a grating waveguide serving as both a polarization splitter and an efficient coupler for silicon-on-insulator nanophotonic circuits," *IEEE Photon. Technol. Lett.*, vol. 21, no. 4, pp. 242–244, 2009.

[87] O. Solgaard, F. S. A. Sandejas, and D. M. Bloom, "Deformable grating optical modulator," *Opt. Lett.*, vol. 17, no. 9, pp. 688–690, 1992.

[88] Y. Shibata, S. Oku, Y. Kondo, T. Tamamura, and M. Naganuma, "Semiconductor monolithic wavelength selective router using a grating switch integrated with a directional coupler," *J. Lightwave Technol.*, vol. 14, no. 6, pp. 1027–1032, 1996.

[89] G. Li, Y. Hashimoto, T. Maruyama, and K. Iiyama, "High-efficiency optical coupling to planar photodiode using metal reflector loaded waveguide grating coupler," *Opt. Quant. Electron.*, vol. 45, no. 7, pp. 657–663, 2013.

[90] S. R. Park, O. J. Kwon, D. Shin, S. H. Song, H. S. Lee, and H. Y. Choi, "Grating micro-dot patterned light guide plates for LED backlights," *Opt. Express*, vol. 15, no. 6, pp. 2888–2899, 2007.

[91] K. M. Byun, S. J. Kim, and D. Kim, "Grating-coupled transmission-type surface plasmon resonance sensors based on dielectric and metallic gratings," *Appl. Opt.*, vol. 46, no. 23, pp. 5703–5708, 2007.

[92] M.-S. Kwon, "Disposable and compact integrated plasmonic sensor using a long-period grating," *Opt. Lett.*, vol. 35, no. 22, pp. 3835–3837, 2010.

[93] N. Eriksson, M. Hagberg, and A. Larsson, "Highly directional grating outcouplers with tailorable radiation characteristics," *IEEE J. Quantum Electron.*, vol. 32, no. 6, pp. 1038–1047, 1996.

[94] S. H. Lin, K. Y. Hsu, W. Z. Chen, and W. T. Whang, "Phenanthrenequinone-doped poly (methyl methacrylate) photopolymer bulk for volume holographic data storage," *Opt. Lett.*, vol. 25, no. 7, pp. 451–453, 2000.

[95] C. Gu, H. Fu, and J. R. Lien, "Correlation patterns and cross-talk noise in volume holographic optical correlators," *J. Opt. Soc. Am. A*, vol. 12, no. 5, pp. 861–868, 1995.

[96] C. C. Sun and W. C. Su, "Three-dimensional shifting selectivity of random phase encoding in volume holograms," *Appl. Opt.*, vol. 40, no. 8, pp. 1253–1260, 2001.

[97] K. O. Hill and G. Meltz, "Fiber Bragg grating technology fundamentals and overview," *J. Lightwave Technol.*, vol. 15, no. 8, pp. 1263–1276, 1997.

[98] J. M. Tedesco, H. Owen, D. M. Pallister, and M. D. Morris, "Principles and spectroscopic applications of volume holographic optics," *Anal. Chem.*, vol. 65, no. 9, 441A–449A, 1993.

[99] R Petit, *Electromagnetics Theory of Gratings*. Springer-Verlag, 1980.

[100] R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," *IEEE Trans. Microwave Theory Tech.*, vol. 25, no. 1, pp. 18–24, 1977.

[101] T. K. Gaylord and M. G. Moharam, "Analysis and applications of optical diffraction by gratings," *Proc. IEEE*, vol. 73, no. 5, pp. 894–937, 1985.

[102] S. M. Schultz, E. N. Glytsis, and T. K. Gaylord, "Design of a high-efficiency volume grating coupler for line focusing," *Appl. Opt.*, vol. 37, no. 12, pp. 2278–2287, 1998.

[103] S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielect waveguides," *IEEE Trans. Microwave Theory Tech.*, vol. 23, no. 1, pp. 123–133, 1975.

[104] H. Kogelnik, "Coupled wave theory for thick hologram gratings," *Bell Syst. Tech. J.*, vol. 48, no. 9, pp. 2909–2947, 1969.

[105] W. Streifer, D. Scifres, and R. Burnham, "Analysis of grating-coupled radiation in GaAs: GaAlAs lasers and waveguides-I," *IEEE J. Quantum Electron.*, vol. 12, no. 7, pp. 422–428, 1976.

[106] A. Yariv, "Coupled-mode theory for guided-wave optics," *IEEE J. Quantum Electron.*, vol. 9, no. 9, pp. 919–933, 1973.

[107] R. E. Dorsey and W. J. Mayer, "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features," *J. Bus. Econ. Stat.*, vol. 13, no. 1, pp. 53–66, 1995.

[108] K. O. Jones, "Comparison of genetic algorithm and particle swarm optimization," in *Proc. Int. Conf. Comput. Sys. Technol.*, 2005, pp. 1–6.

[109] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, 2002.

[110] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Sci.*, vol. 220, no. 4598, pp. 671–680, 1983.

[111] R. A. Rutenbar, "Simulated annealing algorithms: an overview," *IEEE Circuits Devices Mag.*, vol. 5, no. 1, pp. 19–26, 1989.

[112] N. M. Alexandrov, J. Dennis, R. M. Lewis, and V. Torczon, "A trust-region framework for managing the use of approximation models in optimization," *Struct. Optim.*, vol. 15, no. 1, pp. 16–23, 1998.

[113] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM J. Optim.*, vol. 5, no. 1, pp. 13–51, 1995.

[114] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[115] M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A*, vol. 12, pp. 1068–1076, 1995.

[116] M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," *J. Opt. Soc. Am. A*, vol. 12, no. 5, pp. 1077–1086, 1995.

[117] M. G. Moharam and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. Am.*, vol. 72, no. 10, pp. 1385–1392, 1982.

[118] R. A. Villalaz, E. N. Glytsis, and T. K. Gaylord, "Volume grating couplers: polarization and loss effects," *Appl. Opt.*, vol. 41, no. 25, pp. 5223–5229, 2002.

[119] A. D. Papadopoulos and E. N. Glytsis, "Preferential-order waveguide grating couplers: a comparative rigorous analysis using the finite-difference time-domain method," *Appl. Opt.*, vol. 49, no. 30, pp. 5787–5798, 2010.

[120] M Neviere, "The homogeneous problem," in *Electromagnetic theory of gratings*, Springer, 1980, pp. 123–157.

[121] S. Zhang and T. Tamir, "Analysis and design of broadband grating couplers," *IEEE J. Quantum Electron.*, vol. 29, no. 11, pp. 2813–2824, 1993.

[122] E. Anemogiannis and E. N. Glytsis, "Multilayer waveguides: efficient numerical analysis of general structures," *J. Lightwave Technol.*, vol. 10, no. 10, pp. 1344–1351, 1992.

[123] S. M. Schultz, E. N. Glytsis, and T. K. Gaylord, "Design of a high-efficiency volume grating coupler for line focusing," *Appl. Opt.*, vol. 37, no. 12, pp. 2278–2287, 1998.

[124] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Am.*, vol. 71, no. 7, pp. 811–818, 1981.

[125] M. Sodagar, R. Pourabolghasem, A. A. Eftekhar, and A. Adibi, "High-efficiency and wideband interlayer grating couplers in multilayer Si/SiO_2/SiN platform for 3D integration of optical functionalities," *Opt. Express*, vol. 22, no. 14, pp. 16 767–16 777, 2014.

[126] M. Li and S. J. Sheard, "Waveguide couplers using parallelogramic-shaped blazed gratings," *Opt. Commun.*, vol. 109, no. 3, pp. 239–245, 1994.

[127] D Marcuse, "Exact theory of TE-wave scattering from blazed dielectric gratings," *Bell Syst. Tech. J.*, vol. 55, no. 9, pp. 1295–1317, 1976.

[128] K. C. Chang and T Tamir, "Simplified approach to surface-wave scattering by blazed dielectric gratings," *Appl. Opt.*, vol. 19, no. 2, pp. 282–288, 1980.

[129] S. Peng and T Tamir, "Directional blazing of waves guided by asymmetrical dielectric gratings," *Opt. Commun.*, vol. 11, no. 4, pp. 405–409, 1974.

[130] M. Li and S. J. Sheard, "Experimental study of waveguide grating couplers with parallelogramic tooth profiles," *Opt. Eng.*, vol. 35, no. 11, pp. 3101–3106, 1996.

[131] T Liao, S. Sheard, M. Li, J. Zhu, and P. Prewett, "High-efficiency focusing waveguide grating coupler with parallelogramic groove profiles," *J. Lightwave Technol.*, vol. 15, no. 7, pp. 1142–1148, 1997.

[132] T. Aoyagi, Y. Aoyagi, and S. Namba, "High-efficiency blazed grating couplers," *Appl. Phys. Lett.*, vol. 29, no. 5, pp. 303–304, 1976.

[133] M. Cabezón, I. Garcés, A. Villafranca, J. Pozo, P. Kumar, and A. Kaźmierczak, "Silicon-on-insulator chip-to-chip coupling via out-of-plane or vertical grating couplers," *Appl. Opt.*, vol. 51, no. 34, pp. 8090–8094, 2012.

[134] S. K. Patra, J. Ma, V. H. Ozguz, and S. H. Lee, "Alignment issues in packaging for free-space optical interconnects," *Opt. Eng.*, vol. 33, no. 5, pp. 1561–1570, 1994.

[135] S. Bernabé, C. Kopp, M. Volpert, J. Harduin, J. M. Fédéli, and H. Ribot, "Chip-to-chip optical interconnections between stacked self-aligned SOI photonic chips," *Opt. Express*, vol. 20, no. 7, pp. 7886–7894, 2012.

[136] J. Yao, X. Zheng, G. Li, I. Shubin, H. Thacker, Y. Luo, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, "Grating-coupler based low-loss optical interlayer coupling," in *8th IEEE Intl. Conf. Group IV Photon.*, 2011, pp. 383–385.

[137] A. V. Mulé, R. Villalaz, T. K. Gaylord, and J. D. Meindl, "Quasi-free-space optical coupling between diffraction grating couplers fabricated on independent substrates," *Appl. Opt.*, vol. 43, no. 29, pp. 5468–5475, 2004.

[138] G. S. Chan, M. Li, and S. J. Sheard, "Grating-assisted waveguide chip-to-chip interconnects," *Proc. SPIE*, vol. 2213, pp. 267–277, 1994.

[139] S. D. Wu, T. K. Gaylord, E. N. Glytsis, and Y.-M. Wu, "Angular sensitivities of volume gratings for substrate-mode optical interconnects," *Appl. Opt.*, vol. 44, no. 21, pp. 4447–4453, 2005.

[140] D. T. Neilson, "Tolerance of optical interconnections to misalignment," *Appl. Opt.*, vol. 38, no. 11, pp. 2282–2290, 1999.

[141] R. Zhang and F. G. Shi, "Novel fiber optic alignment strategy using Hamiltonian algorithm and Matlab/Simulink," *Opt. Eng.*, vol. 42, no. 8, pp. 2240–2245, 2003.

[142] Z Tang, R Zhang, S. Mondal, and F. Shi, "Fiber-optic alignment automation: tolerance analysis in coupling with a wedged single mode fiber," *Opt. Commun.*, vol. 199, pp. 1–4, 2001.

[143] S Romero-García, B Marzban, S. S. Azadeh, F Merget, B Shen, and J Witzens, "Misalignment tolerant couplers for hybrid integration of semiconductor lasers with silicon photonics parallel transmitters," *Proc. SPIE*, vol. 9133, 91331A, 2014.

[144] S. Romero-García, B. Shen, F. Merget, B. Marzban, and J. Witzens, "Alignment tolerant couplers for silicon photonics," *IEEE J. Sel. Top. Quantum Electron.*, vol. 21, no. 6, pp. 765–778, 2015.

[145] Z. Zhang, B. Huang, Z. Zhang, H. Liu, H. Li, and H. Chen, "Misalignment-tolerant silicon optical modulator with surface-normal optical interface," *IEEE Photon. Technol. Lett.*, vol. 27, no. 10, pp. 1052–1055, 2015.

[146] H. Yamada, M. Nozawa, M. Kinoshita, and K. Ohashi, "Vertical-coupling optical interface for on-chip optical interconnection," *Opt. Express*, vol. 19, no. 2, pp. 698–703, 2011.

[147] L. Li, "Use of fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A*, vol. 13, no. 9, pp. 1870–1876, 1996.

[148] F. Van Laere, G. Roelkens, M. Ayre, J. Schrauwen, D. Taillaert, D. Van Thourhout, T. F. Krauss, and R Baets, "Compact and highly efficient grating couplers between optical fiber and nanophotonic waveguides," *IEEE J. Lightwave Technol.*, vol. 25, no. 1, pp. 151–156, 2007.

[149] C. Wan, T. K. Gaylord, and M. S. Bakir, "Circular waveguide grating-via-grating for interlayer coupling," *IEEE Photon. Technol. Lett.*, vol. 29, no. 21, pp. 1776–1779, 2017.

[150] ——, "Grating-assisted-cylindrical-resonant-cavities interlayer coupler," *Appl. Opt.*, vol. 57, no. 18, pp. 5079–5089, 2018.

[151] R. M. Schimpe, *Cylindrical diffraction grating couplers and distributed feedback resonators for guided wave devices*, US Patent 4,743,083, 1988.

[152] E. Ben-Bassat and J. Scheuer, "Optimal design of radial Bragg cavities and lasers," *Opt. Lett.*, vol. 40, no. 13, pp. 3069–3072, 2015.

[153] X. Li and S. Yu, "Static and dynamic modeling of circular grating-coupled distributed feedback lasers," *IEEE J. Quantum Electron.*, vol. 44, no. 8, pp. 770–776, 2008.

[154] X. Gong, A. Chan, and H. Taylor, "Lateral mode discrimination in surface emitting DBR lasers with cylindrical symmetry," *IEEE J. Quantum Electron.*, vol. 30, no. 5, pp. 1212–1218, 1994.

[155] D. M. Pozar, "Circular waveguide cavity resonators," in *Microwave Engineering*, Wiley, 2011, ch. 6.4, pp. 288–292.

[156] D. Heitmann and C. Ortiz, "Calculation and experimental verification of two-dimensional focusing grating couplers," *IEEE J. Quantum Electron.*, vol. 17, no. 7, pp. 1257–1263, 1981.

[157] V. Kiselev and S. Shaposhnikov, "Wide-aperture focusing in the case of excitation of an optical waveguide through an annular grating," *Sov. Phys. Tech. Phys.*, vol. 33, pp. 987–989, 1988.

[158] A. Mekis, S. Gloeckner, G. Masini, A. Narasimha, T. Pinguet, S. Sahni, and P. De Dobbelaere, "A grating-coupler-enabled CMOS photonics platform," *IEEE J. Sel. Top. Quantum Electron.*, vol. 17, no. 3, pp. 597–608, 2011.

[159] M. Carminati, S. Grillanda, P. Ciccarella, G. Ferrari, M. J. Strain, M. Sampietro, A. Melloni, and F. Morichetti, "Fiber-to-waveguide alignment assisted by a transparent integrated light monitor," *IEEE Photon. Technol. Lett.*, vol. 27, no. 5, pp. 510–513, 2015.

[160] A. Sasaki, T. Baba, and K. Iga, "Put-in microconnectors for alignment-free coupling of optical fiber arrays," *IEEE Photon. Technol. Lett.*, vol. 4, no. 8, pp. 908–911, 1992.

[161] N. Pavarelli, J. S. Lee, M. Rensing, C. Scarcella, S. Zhou, P. Ossieur, and P. A. OBrien, "Optical and electronic packaging processes for silicon photonic systems," *J. Light. Technol.*, vol. 33, no. 5, pp. 991–997, 2015.

[162] R. R. Vallance, L. Shuhe, R. D. Dannenberg, and M. K. Barnoski, *Optical connection of optical fibers to grating couplers*, US Patent App. 14/714,247, 2016.

[163] S. P. Anderson, V. Patel, and D. Piede, *Passively placed vertical optical connector*, US Patent App. 14/070,962, 2015.

[164] X. Zhang, P. K. Jo, M. Zia, G. S. May, and M. S. Bakir, "Heterogeneous interconnect stitching technology with compressible microinterconnects for dense multi-die integration," *IEEE Electron Device Lett.*, vol. 38, no. 2, pp. 255–257, 2017.

[165] P. K. Jo, X. Zhang, J. L. Gonzalez, G. S. May, and M. S. Bakir, "Heterogeneous multi-die stitching enabled by fine-pitch and multi-height compressible mircointerconnects (CMIs)," *IEEE Trans. Electron Devices*, vol. 65, no. 7, pp. 2957–2963, 2018.

[166] H. S. Yang, C. Zhang, and M. S. Bakir, "A self-aligning flip-chip assembly method using sacrificial positive self-alignment structures," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 6, no. 3, pp. 471–477, 2016.

[167] L. Wang, "Multi-robot coordination and safe learning using barrier certificates," PhD thesis, Georgia Institute of Technology, 2018.

[168] L. Wang, A. D. Ames, and M. Egerstedt, "Safety barrier certificates for collisions-free multirobot systems," *IEEE Trans. Robot.*, vol. 33, no. 3, pp. 661–674, 2017.

[169] A. Tuantranont, V. Bright, J Zhang, W Zhang, J. Neff, and Y. Lee, "Optical beam steering using MEMS-controllable microlens array," *Sens Actuators A Phys.*, vol. 91, no. 3, pp. 363–372, 2001.

[170] B. Smith, B. Hellman, A. Gin, A. Espinoza, and Y. Takashima, "Single chip lidar with discrete beam steering by digital micromirror device," *Opt. Express*, vol. 25, no. 13, pp. 14 732–14 745, 2017.

[171] J. Kim, C. Oh, M. J. Escuti, and S. Serati, "Wide-angle nonmechanical beam steering using thin liquid crystal polarization gratings," in *Adv. Wavefront Control: Methods, Dev. Appl. VI*, vol. 7093, 2008, p. 709 302.

[172] C. V. Poulton, A. Yaacobi, Z. Su, M. J. Byrd, and M. R. Watts, "Optical phased array with small spot size, high steering range and grouped cascaded phase shifters," in *Integr. Photon. Res. Silicon NanoPhoton.*, 2016, IW1B–2.

[173] A. Yaacobi, J. Sun, M. Moresco, G. Leake, D. Coolbaugh, and M. R. Watts, "Integrated phased array for wide-angle beam steering," *Opt. Lett.*, vol. 39, no. 15, pp. 4575–4578, 2014.

[174] M. Blaicher, M. R. Billah, T. Hoose, P.-I. Dietrich, A. Hofmann, S. Randel, W. Freude, and C. Koos, "3D-printed ultra-broadband highly efficient out-of-plane coupler for photonic integrated circuits," in *Conf. Lasers Electro-Opt.*, IEEE, 2018, pp. 1–2.

[175] G. Gilardi, W. Yao, H. R. Haghighi, M. K. Smit, and M. J. Wale, "Substrate thickness effects on thermal crosstalk in InP-based photonic integrated circuits," *J. Light. Technol.*, vol. 32, no. 17, pp. 3061–3066, 2014.

[176] W. Yao, G. Gilardi, M. Smit, and M. J. Wale, "Performance degradation of integrated modulator arrays due to electrical crosstalk," in *Adv. Photon. Commun.*, 2014, IW3A–3.

[177] D. Melati, F. Morichetti, G. G. Gentili, and A. Melloni, "Optical radiative crosstalk in integrated photonic waveguides," *Opt. Lett.*, vol. 39, no. 13, pp. 3982–3985, 2014.

[178] D. Varoutas, A. Arapoyianni, and T. Sphicopoulos, "Modeling of electrical crosstalk in OEIC modules," *Fiber Integrated Opt.*, vol. 24, no. 2, pp. 91–111, 2005.

[179] S Chandrasekhar, L. Garrett, L. Lunardi, A. Dentai, C. Burrus, and E. Burrows, "Investigation of crosstalk performance of eight-channel pin/HBT OEIC photoreceiver array modules," *IEEE Photon. Technol. Lett.*, vol. 8, no. 5, pp. 682–684, 1996.

[180] W. Yao, G. Gilardi, N. Calabretta, M. K. Smit, and M. J. Wale, "Experimental and numerical study of electrical crosstalk in photonic-integrated circuits," *J. Light. Technol.*, vol. 33, no. 4, pp. 934–942, 2015.

[181] J. J. Simpson and A. Taflove, "A review of progress in FDTD Maxwell's equations modeling of impulsive subionospheric propagation below 300 kHz," *IEEE Trans. Antennas Propag.*, vol. 55, no. 6, pp. 1582–1590, 2007.

[182] J. M. Bourgeois and G. S. Smith, "A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with experiment," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 36–44, 1996.

[183] R. Luebbers and H. Langdon, "A simple feed model that reduces time steps needed for FDTD antenna and microstrip calculations," *IEEE Trans. Antennas Propag.*, vol. 44, no. 7, pp. 1000–1005, 1996.

[184] H.-R. Chuang and L.-C. Kuo, "3-D FDTD design analysis of a 2.4-GHz polarization-diversity printed dipole antenna with integrated balun and polarization-switching circuit for wlan and wireless communication applications," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 2, pp. 374–381, 2003.

[185] S. C. Hagness, A. Taflove, and J. E. Bridges, "Two-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: fixed-focus and antenna-array sensors," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 12, pp. 1470–1479, 1998.

[186] M. Fujita, S. Takahashi, Y. Tanaka, T. Asano, and S. Noda, "Simultaneous inhibition and redistribution of spontaneous light emission in photonic crystals," *Sci.*, vol. 308, no. 5726, pp. 1296–1298, 2005.

[187] Z. Sun and H. K. Kim, "Refractive transmission of light and beam shaping with metallic nano-optic lenses," *Appl. Phys. Lett.*, vol. 85, no. 4, pp. 642–644, 2004.

[188] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.*, vol. 14, no. 3, pp. 302–307, 1966.

[189] A. F. Oskooi, D Roundy, M Ibanescu, P Bermel, J. D. Joannopoulos, and S. G Johnson, "MEEP: a flexible free-software package for electromagnetic simulations by the FDTD method," *Comput. Phys. Commun.*, vol. 181, pp. 687–702, 2010.

[190] Y. Chung and N. Dagli, "An assessment of finite difference beam propagation method," *IEEE J. Quantum Electron.*, vol. 26, no. 8, pp. 1335–1339, 1990.