Silicon Photonics for Next-Generation Optical Processing and Communications

by

Dusan Gostimirovic

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Carleton University
Ottawa, Ontario

Abstract

The rapid growth of transistor and fiber-optic technologies has brought unprecedented advancements in computing and telecommunications. While fiber-optics are poised to grow in speed this year, growth in computing power continues to slow down as the limits of transistor downscaling are approached. This has given rise to the growing field of silicon photonics, where the well-established microelectronic fabrication process is being adapted to create integrated devices that bridge the electronic and optical domains for large performance gains in data centers and high-performance computers. However, the inherent cost of switching between the two domains, and the fact that much of on-chip data transfer is still carried out by low-speed, high-power electronics, are problems that scale with the growing global demand for bandwidth.

In this thesis, a novel optoelectronic computing logic architecture based on the silicon photonics platform is presented. This architecture combines the best of optical data transfer and electronic control for the highest level of throughput presented in the literature—presenting a promising option for future, "Beyond Moore" computing and processing. The main driver of this architecture is the silicon microdisk modulator, which achieves the best combination of energy efficiency, operating speed/bandwidth, compactness, and cost of all previously demonstrated optoelectronic processing devices. New configurations of the microdisk modulator are introduced to further improve the performance, not just for optical logic, but for all optical processing and communications applications. One of these applications is a novel on-chip optical communications circuit, presented in this thesis, that achieves ultrahigh-bandwidth-density through efficient microdisk-based design.

As integrated photonic circuits like these become more complex, with hundreds or thousands of components on the chip, the design process becomes lengthier and more expensive. Even for small circuits, like those presented in this work, the time taken from design to characterization is delayed by having to design components secondary to the main devices (e.g., for coupling light into the chip). As the final part of this work, a machine learning based photonic device modeller is created to accelerate the photonic simulation and design process by multiple orders of magnitude, with minimal input from the designer.

Preface

This integrated thesis consists of work from the following journals, conference proceedings, and papers awaiting publication:

1. D. Gostimirovic and W. N. Ye, “A compact silicon-photonic mode-division (de)multiplexer using waveguide-wrapped microdisk resonators,” Opt. Lett. Submitted for publication, (October 2020).
2. © 2020 IEEE. Reprinted, with permission, from D. Gostimirovic and W. N. Ye, “Ultralow-power double vertical junction microdisk modulators,” IEEE J. Quantum Electron. Submitted for publication, (2020).
3. D. Gostimirovic, F. De Leonardis, R. Soref, V. M. N. Passaro, and W. N. Ye, “Ultrafast electro-optical disk modulators for logic, communications, optical repeaters, and wavelength converters,” Opt. Express 28(17), 24874–24888 (2020); https://doi.org/10.1364/OE.400716
4. © 2020 IEEE. Reprinted, with permission, from D. Gostimirovic and W. N. Ye, “An open-source artificial neural network model for polarization-insensitive silicon-on-insulator subwavelength grating couplers,” IEEE J. Quantum Electron. 25(3), (2018); https://doi.org/10.1109/JSTQE.2018.2885486
5. D. Gostimirovic and W. N. Ye, “Ultracompact CMOS-compatible optical logic using carrier depletion in microdisk resonators,” Sci. Rep. 7(1), (2017); https://doi.org/10.1038/s41598-017-12680-1
6. D. Gostimirovic and W. N. Ye, "A comparison of microresonator devices for WDM-compatible mode-division multiplexing," Proc. SPIE 11284, Smart

Photonic and Optoelectronic Integrated Circuits XXII, 112841E (26 February 2020); https://doi.org/10.1117/12.2550744

These works directly make up the majority of this thesis; however, certain sections were removed to avoid repeated information or added to provide further context and establish better transitions between the works. Chapter 2 consists of Work 5; Chapter 3 consists of Works 2 and 3; Chapter 4 consists of Works 1 and 6; and Chapter 5 consists of Work 4.

The student was the first author for each of these works, in which he was fully involved in setting up and conducting the research, obtaining data and analyzing results, and preparing and writing the material. Work 3 included multiple co-authors that contributed to each of these areas, but the student was still fully and equally involved in the entire process.

The information provided in this preface has been confirmed by the student's academic supervisor, Professor Winnie N. Ye.

Acknowledgements

Firstly, I would like to sincerely thank my thesis supervisor, Professor Winnie Ye, for her invaluable support throughout my graduate studies. Her knowledge has guided me in more ways than I can possibly count, and her motivation has pushed me to achieve more than I could have ever done without her. I am greatly appreciative of her allowing me to choose research topics based on my own interests, while also steering me towards realistic objectives along the path to success.

I would also like to express my gratitude towards the Computer Systems and Physical Facilities Manager, Nagui Mikhail, for his immense support throughout my studies. Getting access to software, lab/computer equipment, and facilities was always made easy by him, and his kindness and wisdom is something I will always cherish.

I am grateful to all of my fellow classmates, colleagues, and collaborators along the way. They are too many to name individually, but they have all shaped this thesis in different ways. Through lessons, conversations, and just leading by example, they have all also helped me become a better researcher and a better person.

I am thankful to have received financial support from Professor Ye, Carleton University, and the Ontario and Canadian Governments through funding, scholarships, and awards. With a special thank you to Ed Ireland for his generous support throughout my graduate studies. This thesis would not have been possible without their support.

Finally, I cannot thank my family and friends enough for their cheerful support throughout my many, many years in school. I am thankful to my parents for allowing me to focus solely on my studies and not having to worry about financial burdens; to my brother and my friends for keeping me motivated, grounded, and humble; and to my loving partner Tegan, for her incredible patience throughout the many long days spent in the lab or behind the computer screen working towards this point.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALU	Arithmetic Logic Unit	31
ANN	Artificial Neural Network	119
CMOS	Complementary Metal-Oxide-Semiconductor	1
CW	Continuous Wave	60
DEMUX	Demultiplexer	88
DWDM	Dense Wavelength-Division Multiplexing	92
EOL	Electro-Optic Logic	59
EOM	Electro-Optical Modulator	59
ER	Extinction Ratio	21
FCA	Free-Carrier Absorption	13
FCD	Free-Carrier Dispersion	12
FDTD	Finite-Difference Time-Domain	21
FEM	Finite Element Method	22
FF	Fill Factor	128
FSR	Free Spectral Range	90
IL	Insertion Loss	49
MDM	Mode-Division Multiplexing	8
MRR	Microring Resonator	13
MUX	Multiplexer	87
MZI	Mach-Zehnder Interferometer	13
OE	Optoelectronic	1
OEO	Optical–Electrical–Optical	60

OOL	Optical–Optical Logic	60
PD	Photodetector	60
PSO	Particle Swarm Optimization	120
ReLU	Rectified Linear Unit	123
ROADM	Reconfigurable Optical Add–Drop Multiplexer	60
SOI	Silicon on Insulator	18
SWG	Subwavelength Grating	119
TE	Transverse Electric	39
TM	Transverse Magnetic	128
TPA	Two-Photon Absorption	15
WDM	Wavelength-Division Multiplexing	8
WGM	Whispering Gallery Mode	59
XT	Crosstalk	89

Chapter 1: Introduction

In 2021, there will be 5.3 billion internet users [1]. The majority of the data they will send and receive (over 20 zettabytes per year) will interact with the thousands of servers in the hundreds of hyperscale data centers from the largest data companies in the world (e.g., Amazon, Google/Alphabet, Microsoft) [2]. Satisfying the growing demand for bandwidth is not cheap, however, as the cost of running datacenters worldwide has become environmentally significant, at 1% of the global electricity consumption [3]. The expected consumption would be much higher if significant efforts to reduce power consumption weren't already in place, from more efficient cooling and system-level designs, down to the optimization of individual chip-level components—though the latter is becoming more difficult to improve using conventional methods [4, 5].

A major bottleneck in fiber-optic, cloud-based systems is the interoperability of optical data transfer and electronic processing. Optical fiber connects the thousands of servers within a data center and connects the data center to the rest of the world; however, the components inside the servers are entirely electronic. The introduction of silicon photonics has reduced the cost of bringing low-power optical data links to the edge of the data center rack, with high-performance optoelectronic (OE) transceivers that efficiently convert optical data to electrical, and vice versa [6]. Over 24 million transceivers are projected to be in use by 2025. Silicon photonics brings optical processing and data communications to the chip with the same manufacturing processes already in place with the complementary metal-oxide-semiconductor (CMOS) microelectronics industry—meaning cheap and easy integration into existing processing and communications systems [7, 8, 9, 10]. Given the

massive scale of the deployment of these transceivers, it is clear that any improvements to their energy efficiency will make substantial reductions in system-level energy consumption. Furthermore, as the processing is still being handled in the electronic domain, the horizon is blue for the further integration of optics down to the most fundamental level.

1.1 The Future of Moore's Law

Although we are seeing an exponential increase in the global demand for bandwidth, which means faster, more-capable computing devices are of high desire, we are also seeing a reduction in performance improvement year over year. The underlying workhorse behind the digital revolution of the past 50 years has been the CMOS transistor, which has followed a ~2X reduction in size every two years, according to Moore's law [11]. The reduction in transistor size simultaneously reduces power consumption, increases speed, and allows for more of them to fit onto a single chip. Because of this, we can see an improvement from 3,500-transistor, 500 kHz chips in 1972 to multibillion-transistor, 3+ GHz chips today. Many have cited the end of Moore's law to be approaching or already here [4, 5, 12], as the transistor is starting to approach its atomic limits. Not only this, the network of metal interconnections between the transistors has become the major consumer of energy and producer of heat on the chip—becoming a major bottleneck to any improvements in transistor design. Estimations have shown that metal interconnects consume more than half of the chip's energy [13]. As battery consumption of personal devices and the cost of operating hyperscale data centers has become a primary concern for users, this places a large level of importance on chip-level interconnection.

To satisfy the performance demands that foundries are struggling to meet, chipmakers have turned to new, application specific circuit designs that use the same fundamental components but in different architectures for applications that general-purpose chips cannot handle as well. Graphical processing units (GPUs) have largely replaced CPUs for major applications like gaming, video processing, and machine learning. New circuits like the tensor processing unit (TPU) from Google have become popular for the ever-growing machine learning applications in almost every data science field today. Even the rise of quantum computing has come about to potentially solve tasks that even many extra years of Moore's law scaling would never be able to solve in a lifetime [14, 15]. While these new architectures are helping satisfy the demand currently, foundries are looking to new technologies in the push for "More Moore," where entirely different physical concepts like ferroelectrics, ferromagnetics, magnetoelectrics, and photonics have been heavily cited [16].

1.2 Silicon Photonics: A Feasible Next-Generation Technology

Silicon photonics has been presented as one of the few feasible next-generation technologies to satisfy the push for high-bandwidth, attojoule-scale signal processing and communications [13]. As optics are already used to transfer large amounts of data for cheap at the system level, with fiber optics, silicon photonics can introduce the same benefits up to and even on the chip itself. This technology is regarded as one of the most promising because of its low cost, ease of fabrication, and extraordinary light-guiding performance compared to other optical material platforms. With a high refractive index contrast to the surrounding cladding material (silicon dioxide), silicon waveguides are able to function

well at low microscale dimensions, which is critical for high-density, very-large scale integrated (VLSI) circuits. This high refractive index contrast also allows for efficient multiplexing of many data channels on a single waveguide by combining different wavelengths, spatial modes, and/or polarization states. Competing optical technologies that combine elements of groups III and V of the periodic table (e.g., GaAs and InP) are much more costly and difficult to fabricate due to the smaller scale of the infrastructure supporting it. Additionally, the Group III-V materials' refractive indices are lower, which leads to much larger waveguiding devices.

Silicon performs best in the telecommunications C-band (1515 to 1565 nm), as the optical absorption is lowest there. This fits seamlessly with the optics that are already used in long-haul fiber optic communications. For similar reasons, silicon is a poor material for detection and light amplification (lasing); thus, additional materials and/or fabrication steps add complexity and cost to the system. The guiding properties of silicon can be efficiently modified, however, in a variety of different ways to produce the optical switches in the millions of optoelectronic transceivers already in use worldwide [6]. The potential of silicon photonics for chip-level optical communications and data processing, however, remain as heavily researched topics with few current commercial applications.

Silicon-photonic optical modulators are the key drivers to the 100 Gb/s optoelectronic transceivers already in use in data centers today [17]. The same types of devices can be used to drive optical processing architectures and chip-level data communications. Novel optical computing architectures based on the use of these modulators in directed logic

circuits have been shown to perform each of the fundamental computing logic operations [18, 19, 20, 21, 22]. New architectures based on analog computing and deep neural networks have also been replicated with optical technologies and are expected to perform well above the current state-of-the-art GPUs and TPUs in modern machine learning tasks [23, 24]. Even memory has been demonstrated in silicon photonic devices [25]. Although optics can be controlled by optics through nonlinear processes, it is generally with very high powers; thus, most of the successful applications listed here combine optical data transfer with electrical control. Although it is currently unclear which architecture and combination of optics and electronics is best, it is clear that it will be some combination of the two, and that a purely optical chip may not see many realistic applications in at least the near future. The main limitations in the advancement of these designs in commercial applications is the relative size, power, and speed of the active (switching) photonic components, which are worse or not clearly better than current CMOS designs. Although a direct comparison of a transistor or logic circuit to a photonic circuit may not be feasible (a larger scope is necessary), it is clear that many improvements must still be made to bring performance to suitable levels.

Furthermore, the photonic design process is still relatively immature compared to CMOS. Although significant work has been done to create industry-standard EDA tools in combination with well-established foundry processes, the design-to-chip process still requires more manual work than that of CMOS. There is certainly much room for improvement in the automation and acceleration of photonic integrated circuit design.

1.3 Thesis Objectives and Organization

Although optical computing has been heavily cited as a potential successor to Moore's law, it currently falls behind the mature CMOS platform in almost every performance metric. Still, the capability of seamlessly integrating to ultrahigh-speed, low-power optical interconnects without costly optical–electrical–optical converters gives optical computing a large upside. The ideal chip combines the best of optical data transfer and electrical control to satisfy the growing demand for bandwidth without further contributing to the growing global energy usage. By improving on the compactness, speed, and energy efficiency of optical processing circuits, and by using only cost-effective, CMOS-compatible materials and designs, we can come closer to realizing the full potential of ultrahigh-speed optics on a computer chip.

The objective of this thesis is to explore novel silicon photonic integrated devices to significantly improve the performance of integrated optical processing and communications circuits. The main improvements come from novel optoelectronic device designs to create low-power, high-throughput optical logic architectures that fit seamlessly with high-speed optical data transfer interconnects. Secondly, to further increase the capacity of on-chip data transfer itself with low-power, high-bandwidth-density communications circuits based on similar devices used in the optical logic circuits. By removing the optical–electrical–optical bottleneck, improvements to optical processing and communications at the device level can make substantial improvements to the system level that may otherwise not be possible with the diminishing improvements in Moore's law and electrical data transfer. Lastly, new design methodologies that leverage the

growing field of machine learning are explored and adapted to silicon photonics to accelerate and automate the design process of new integrated processing and communications circuits.

The organization of this thesis are as follows:

Chapter 2.1 presents the current state of optical processing and computing, where the potential benefits and challenges of each method are summarized. Chapter 2.2 presents a novel optoelectronic silicon modulator (logic cell) designed for high-throughput optical logic. In Chapter 2.3, a novel optoelectronic computing architecture based on the logic cell is presented. This architecture works with proven silicon photonic devices that fit seamlessly with the existing CMOS microelectronic technology for the coexistence of high-performance optics and electronics on a single chip. Finally, a summary discussion on this topic is presented in Chapter 2.4.

In Chapter 3.1, new ultra-efficient optoelectronic modulator designs are presented, which offer large energy efficiency improvements over the logic cell modulator in Chapter 2.2 and the state-of-the-art designs in the literature. In Chapter 3.2, these high-efficiency designs are adapted to high-speed modulators, logic gates, optical repeaters, and wavelength converters to demonstrate their wide potential for high-speed, low-power applications. Finally, a summary discussion on this topic is presented in Chapter 3.3.

Chapter 4.1 introduces an ultrahigh-bandwidth-density on-chip optical communications circuit that combines wavelength-division multiplexing (WDM) and mode-division multiplexing (MDM). Chapter 4.2 presents the design of the waveguide-wrapped microdisk resonator that improves the bandwidth, compactness, and energy consumption of this circuit. A detailed comparison is made with the conventional, microring resonators to demonstrate the benefits of this design. Ultimately, a WDM–MDM circuit is designed to fit seamlessly with the optoelectronic computing architecture presented in the previous chapter by using the same materials and device structures (silicon microdisk resonators/modulators). Chapter 4.3 outlines the process of fabricating these designs, and Chapter 4.4 presents the experimental results. Finally, a summary discussion on this topic is presented in Chapter 4.5.

As the devices and circuits presented in Chapters 2 to 4 are only small parts of a full photonic integrated circuit, much work had to be done on their complete integration. In Chapter 5, a design methodology based on machine learning is introduced as a way to automate and accelerate the design of the photonic components needed at the full circuit level. This method significantly speeds up the time the design idea can make it to fabrication. This chapter presents the artificial neural network training of photonic device models, and the use of these models in the rapid design of components like the polarization-insensitive subwavelength grating coupler. Chapter 5.1 of this chapter presents the process of modelling photonic integrated devices with artificial neural networks. Chapter 5.2 applies this method to the automated and accelerated design of polarization-insensitive

subwavelength silicon grating couplers. Finally, a discussion on the future work of this topic is presented in Chapter 5.3.

In Chapter 6, a conclusion is made on the designs and results presented in this thesis. Chapter 6.1 discusses the significant impact of the new devices, circuits, and design methodologies presented, and Chapter 6.2 lists the achievements, publications, and awards earned as a result of them. Finally, Chapter 6.3 discusses multiple future works for the topics presented in this thesis.

Chapter 2: Design of Next-Generation Optical Processing and Computing Circuits

The promise of optical computing and processing stems from the benefits already demonstrated in optical-fiber communications: speed-of-light data transfer, the potential of transferring multiple data signals simultaneously, and the removal of significant performance bottlenecks created by the many optical–electrical–optical conversions. Although the minimum dimensions of silicon photonic devices may never reach that of modern CMOS transistors, the benefits listed above can help next-generation computing reach new levels of performance when properly leveraged. In particular, when more of the chip-level electrical interconnects are replaced with high-speed, low-power optics, to perform computing operations on the same optical data without making costly conversions to/from the electrical domain.

This chapter introduces a novel optoelectronic computing architecture based on a "logic cell" that improves computational throughput and energy efficiency over previous optoelectronic logic demonstrations in the literature.

The simulation and design work of the optoelectronic logic cell and architecture was documented and published in Scientific Reports in 2017 and was presented at SPIE Photonics West in 2017.

2.1 Current State of Integrated Optical Processing

Optical logic is just one of many possible technologies that have been presented for the push for "More Moore", and it itself has many possible implementations. Not only in the type of computing (e.g., traditional von Neumann, quantum, neuromorphic), but also in the hardware implementation of the architecture. For an optical implementation of a traditional, von Neumann computer (i.e., fundamental logic gates being used for general-purpose computing), a variety of fundamental switching devices analogous to the CMOS transistor can be used to drive the computing operations. The ultimate performance of the proposed optical logic architecture—and if it is, in fact, a suitable CMOS successor—depends significantly on the performance of the individual switching mechanism and how it fits into the overall computing architecture. The following sections outline and discuss the current methods for optical processing.

2.1.1 Passive Effects

In its simplest form, optical logic can operate at the speed of light, passively, with zero switching power consumption. Interferometric devices like the Y-branch splitter/combiner can been used as optical logic gates by working off the phase relationship of the inputted optical signals. An example of an all-optical passive XOR gate is shown in Figure 2.1, where two signals with a π phase shift from each other are combined at the Y-branch logic gate. A signal appears at the output if only one signal is present. If both signals are present, they destructively interfere, and the XOR logic condition is satisfied. This operation occurs with near zero delay with the only energy consumption coming from the laser itself (which is likely to be external to the chip).

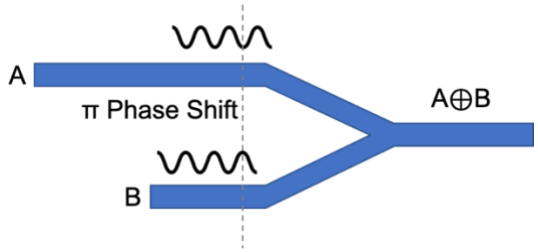


Figure 2.1 Schematic of a passive all-optical XOR logic gate based on Y-branch interference. If only one input is present, the output will be 1; if both inputs are present, they destructively interfere, and the output will be 0.

These logic gates are typically made into photonic crystals, where extremely small footprints are achieved [26, 27]. However, the interferometric (passive) approach is inherently limited by the strict coherence conditions required for combinational logic circuits. For the example of the XOR gate, the phase of the output signal cannot be predicted, and therefore it cannot be used for successive phase-dependent logic operations. For other operations, like the OR gate, which would combine two in-phase signals, it cannot be predicted if the constructive output will be 0, 1, or 2, which further complicates the operation of cascading logic gates.

2.1.2 Free-Carrier Effects

Free-carrier effects in doped silicon-photonic resonant and interferometric devices are predominantly used to drive the modulators in transceivers for data-center optical communications [28]. By integrating a pn diode into the silicon waveguide, the added dopants change the real and imaginary refractive indices. Applying a bias to the pn diode further changes the refractive index of the waveguide through free-carrier dispersion (FCD)

and absorption (FCA), which can produce big changes in the pass or drop propagative state of devices like the microring resonator or Mach-Zehnder interferometer (MZI). These effects have also been used to perform optoelectronic logic. The directed logic architecture [18, 19, 20, 21, 22, 29] has been used to perform single logic gates, combinational logic circuits, and FPGA-like switching arrays. An example of the directed logic architecture is shown in Figure 2.2. Unlike conventional Boolean logic, which accumulates delay at each sequential gate, directed logic switches each gate independently and simultaneously, so that the delay is simply the time of a single switch and the small delay taken for the optical signal to propagate through the circuit. Therefore, the relative throughput of this type of architecture grows with the size of the circuit.

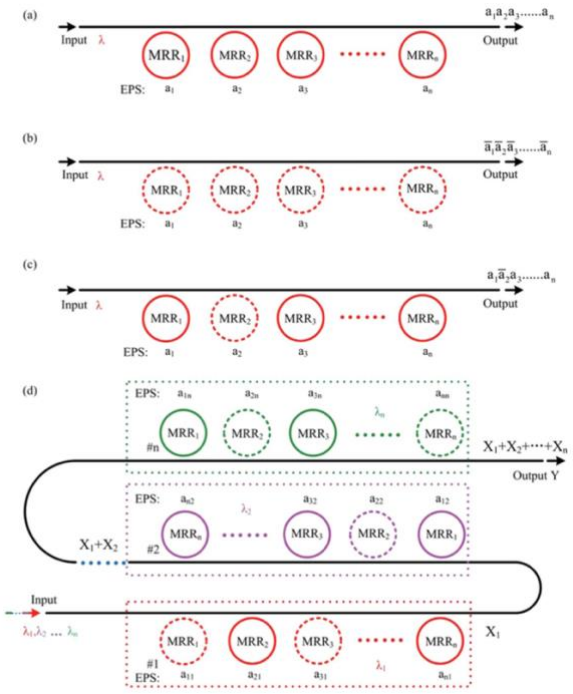


Figure 2.2 Optical logic based on a directed logic architecture [29]. Each microring resonator (MRR) is switched on or off, simultaneously, to create a different path for each wavelength of light inputted into the circuit. Each path would represent a different condition in the logic circuit.

Table 2.1 shows a comparison of a modern, 15 nm CMOS 2-bit NAND gate with a typical CMOS-compatible photonic version based on demonstrations in the literature. For every performance metric, the optical gate severely underperforms. Although a direct comparison is not suitable, as an optical logic architecture would result in the removal of many electrical interconnects (which consume most of the power and add significant delays in processing [13]), it is still clear that there is room for improvement.

Table 2.1 Comparison of a modern CMOS logic gate and the best-performing optical equivalent.

	CMOS[a] [16]	**Photonic**[b]
Size	100 μm^2	Hundreds of μm^2
Energy	50 aJ/bit	Tens of fJ/bit
Delay	0.5 ps	Hundreds of ps

2.1.3 Nonlinear Effects

Multiple nonlinear effects are present in silicon, which can be used to process signals at high speeds. These effects arise from the change in polarization caused by the interaction between the electric field of the optical signal and the electrons (or phonons) of the nonlinear material, as given by [30]

$$\mathbf{P}(t) = \varepsilon_0\left(\chi^{(1)}\mathbf{E}(t) + \chi^{(2)}\mathbf{E}^2(t) + \chi^{(3)}\mathbf{E}^3(t) + \cdots\right). \tag{2.1}$$

[a] In 2020, high-performance chips are using 7 nm transistors, which brings these numbers down further, but within the same general magnitude.
[b] Due to the wide variation in performance across the literature, general magnitudes are taken from a variety of sources cited in this thesis, rather than specific numbers. There are no photonic demonstrations that show high performance in each of the three metrics.

The first-order term $\chi^{(1)}$ is generally associated with the susceptibility of free carriers absorbed from a strong optical signal. Like with the optoelectronic free-carrier effect, carriers absorbed optically induce changes to the real and imaginary parts of the material's refractive index. At telecommunications wavelengths, this process (in silicon) arises from two-photon absorption (TPA). All-optical modulators and logic gates have been demonstrated with this effect [31, 32]; however, due to the long free-carrier lifetime in silicon, the modulation speed is limited to MHz rates.

The second-order term $\chi^{(2)}$ is generally associated with the optoelectronic Pockels effect, which is used in many high-performance optoelectronic modulators [33, 34]. However, $\chi^{(2)}$ is absent in centrosymmetric crystals like silicon. The symmetry can be broken by introducing strain into the material [35], but the process can be difficult, and the resulting devices do not necessarily perform better than carrier-based optoelectronic modulators [36].

The third-order term $\chi^{(3)}$ produces a plethora of nonlinear effects in silicon. This is perhaps best associated with the Kerr effect [30],

$$n = n_2 I, \tag{2.1}$$

where the refractive index changes by the product of the optical intensity and the nonlinear index of the material n_2. For low-power applications, n_2 must be as high as possible. The nonlinear index of silicon is low, but other CMOS-compatible materials like silicon nitride

and hydrogenated amorphous silicon have indices high enough for all-optical modulation [37, 38, 39, 40]. However, the best-performing devices still consume multiple orders of magnitude more energy than the best optoelectronic devices. The change in refractive index occurs instantaneously, but the operating speed is still limited to the photon lifetime within the resonant device the effect is hosted in.

Another commonly used third-order nonlinear effect, stimulated Raman scattering, induces a strong factor of absorption at a set frequency away from a strong inputted optical signal. This effect allows for wavelength-selective modulation in microring resonators [41], enabling WDM modulation with a single switch. Like with the Kerr effect, this process requires optical powers above what is acceptable for next-generation processing.

Perhaps the best-performing nonlinear effect for optical logic is four-wave mixing, where three photons of certain frequencies are combined to generate a fourth photon at a new frequency. The fundamental logic operations can be performed within a straight, non-resonant, waveguide, leading to GHz operational speeds [42, 43, 44]; however, the devices are in the millimeter scale and are thus unfeasible for high-density computing circuits.

2.1.4 Other Effects

Other modulation mechanisms that do not fit into the previous three categories have been used to perform optical logic. The heat-induced change in refractive index, which is commonly used to tune silicon photonic devices post-fabrication, has been used to demonstrate the proof-of-concept for many optical logic architectures [45, 46]. The slow

timescale of the thermodynamics limits the devices to kHz operating speeds and is thus unsuitable for high-performance computing. Semiconductor optical amplifiers were the choice for many early optical logic designs [47, 48]; however, the materials needed for them are not CMOS-compatible, and are thus unsuitable for low-cost, high-volume production. Plasmonic modulator devices, which utilize the surface plasma oscillations created through the interface of a metal and a dielectric have been shown to operate at very high speeds [49, 50], but the significant optical loss and fabrication concerns introduced by the tight dielectric–metal integration generally outweigh the performance improvements.

2.2 Ultracompact Microdisk Modulator for Logic Gate Operations

There are many factors to consider when proposing an optical replacement for CMOS logic. The primary four metrics are cost, speed, size, and energy consumption. To keep costs low, the proposed logic architecture must leverage CMOS through silicon photonics. Thus, many well-performing nonlinear materials are not considered (e.g., III-V Pockels modulators and semiconductor optical amplifiers). Effects that cannot reach GHz operating speeds (e.g., thermal effects, two-photon absorption, and free-carrier injection) are also not considered. In general, the all-optical nonlinear effects require unsuitably high operating powers or long device structures to operate. These outweigh the benefits of all-optical processing and instantaneous material property changes until significant advancements are demonstrated in silicon.

The switching method that offers the best balance of performance is free-carrier depletion in silicon microresonators with integrated pn diodes. By applying a reverse bias to the pn diode (junction), the refractive index of the ring is quickly changed, which shifts its resonance wavelength. By shifting the resonance wavelength of the ring, an optical signal will either pass by or pass through it—effectively working like a switch. Free carriers can be injected into or depleted out of the depletion region of an integrated pn diode. Due to the long free-carrier recombination lifetime in silicon (~700 ps [37]), free-carrier injection is limited to ~200 MHz modulation speeds and is thus unfeasible for high-performance logic applications. The free carriers can be actively swept away after injection, but with the expense of operating power and added driving circuitry. It is for these reasons that carrier-depletion modulators dominate in modulators for communications and why it is chosen as the driver for this proposed optical logic architecture. These modulators come in the form of microresonators or MZIs. Microresonators are chosen here simply because they can be made many times smaller than MZIs, which is a primary concern of a CMOS replacement. For the current state of optical modulation devices, the carrier-depletion microresonator is the only device with potential for large-scale integrated optical logic.

Free-carrier effects in silicon-on-insulator (SOI) microresonators have already been used to perform optical logic [18, 19, 21, 22, 29, 51, 52], but there is still much room for improvement in both the architecture of the logic and the optimization of the switching structure itself. The biggest improvement to the structure comes with the optimization of the orientation of the pn junction. Although laterally oriented pn junctions are the most common for these types of devices, a vertical alignment (p on top of n, and vice versa) is

more efficient, as the same change in the depletion width of the pn junction spans a larger percentage of the more vertically confined mode in standard, 220×500 nm SOI waveguides. The difference in length between the two pn junction orientations is visualized by the waveguide cross-sections in Figure 2.3. With a longer pn junction, the operating voltage can be reduced, which reduces the dynamic energy consumption. To integrate a vertical pn junction in a microring resonator, one contact electrode would have to be placed on top of the ring, as shown in Figure 2.4a, or an L-shaped junction would be required, as shown in Figure 2.4b. The latter requires multiple ion implantation steps and does not take advantage of the full width of the optical mode. And the former would introduce substantial optical losses from the metal being in close proximity to the optics. For vertical pn junctions, microdisks are used instead of microrings, as shown in Figure 2.4c, as it is much easier to integrate the contact electrodes.

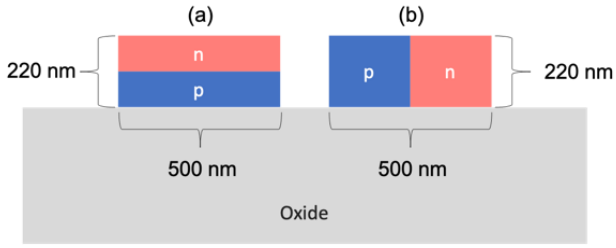


Figure 2.3 Cross-section views of 220×500 nm SOI waveguides with a (a) vertical pn junction and a (b) lateral pn junction. The vertical orientation allows for a longer pn junction, and therefore a greater modulation efficiency.

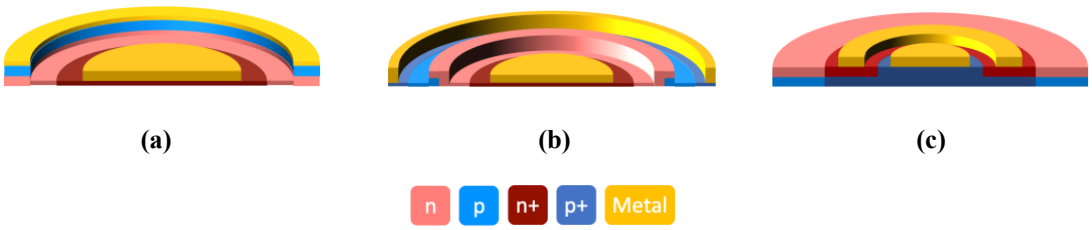
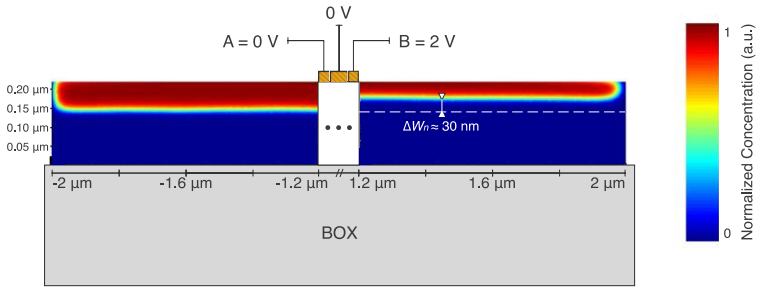


Figure 2.4 Cross-section examples of vertically oriented pn junctions in (a) a microring resonator with top-level contact electrodes, (b) a microring resonator with an L shape, and (c) a microdisk resonator. Note that the coupled waveguides are not shown in these views.

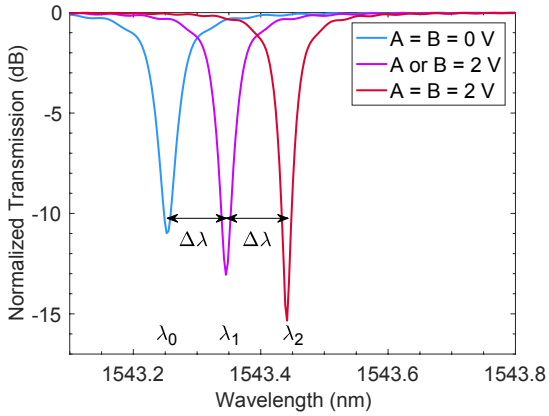
For the proposed logic device, which will be referred to as a logic cell, the integrated pn diode junction is split in half and electrically isolated to create two independently working diodes connected to the two control inputs of a basic logic operation. At the very least, this splits the size of the operation (doubling throughput), as all previous designs use one switch per control input. Furthermore, wavelength-division multiplexing enables parallel operations for enhanced functionality per unit area and time or multiple instruction, single data (MISD) processing. With three equally spaced wavelengths, one microdisk switch simultaneously performs OR, AND, and XOR operations—as well as their inversions. These six operations, and the built-in half adder via the XOR and AND operations, make up the fundamental logic cell of this architecture.

The vertical-junction microdisk modulator was simulated to characterize its performance benefits. A 3 μm thick layer of buried oxide sits beneath the 0.22 μm thick silicon disk and bus waveguides. A disk radius of 2 μm is chosen for high-density integration. The bus waveguides are 0.4 μm wide and are spaced 0.16 μm away from the disk. The n-doped region extends to the outer edge of the disk, has a concentration of 3×10^{17} cm^{-3}, and a depth

of 0.11 μm. Likewise, the p-doped region underneath has a concentration of 2×10^{17} cm^{-3} and a height of 0.11 μm. To isolate the two diodes, a 0.6 μm wide region in the middle of the disk is left undoped. Underneath the grounded metal contact is a highly doped p+ region, with a concentration of 2×10^{19} cm^{-3}, an upper-half radius of 0.6 μm, and a lower-half radius of 1.2 μm. Underneath the active metal contacts are highly doped n+ regions, with concentrations of 3×10^{19} cm^{-3} and outer radii of 1.2 μm. These highly doped regions are added to provide higher conductivity with the metal. In Figure 2.5a, a finite-element method drift–diffusion simulation from Lumerical (CHARGE) [53] shows the two diodes working independently of each other, where the concentration of free electrons (in this example) spreads by ΔW_n = 30 nm for the reverse-biased right diode (−2 V), while remaining confined in the center for the unbiased left diode. These simulations are then imported into Lumerical's finite-difference time domain (FDTD) [54] optical solver to demonstrate the resonance-shifting behavior of the switch, as shown in Figure 2.5b. Note that this microdisk resonator has an unbiased quality factor (Q) of approximately 31,000. One 2 V input shifts the resonance peak by $\Delta\lambda$ = 90 pm (45 pm/V), with a modulation depth of 12.6 dB, and an extinction ratio (ER) of 13 dB. With both inputs on, an additional 90 pm shift is observed, with a modulation depth of 15 dB, and an ER of 15.3 dB.



(a)



(b)

Figure 2.5 Finite element method (FEM) simulations visualizing the concentration of free electrons in the left- and right-side diodes operating at 0 V and –2 V biases, respectively (concentration of free holes not shown). The right-side diode shows a 30 nm larger electron depletion width ΔW_n, while the left-side diode remains undisturbed. (b) FEM–FDTD simulations showing the selective through-port transmission spectra of the disk under zero, one (A or B), and two (A and B) 2 V controls.

The switching energy per diode is defined as $E = CV^2$, where V is the applied bias, and C is the capacitance of the pn junction. Given that the probability of the 0-to-1 switching case is 1/4, the energy consumed per bit is defined as $E/bit = CV^2/4$. The FEM simulations show a capacitance of 12 fF per diode, resulting in 12 fJ/bit dynamic switching operation. This compares well to the experimentally verified microdisk modulators of 344 and 0.9 fJ/bit

operation [55, 56]. The total switching delay is defined as $t_{switch} = \tau_{RC} + \tau_{photon}$, where τ_{RC} is the RC constant of the electronics, and $\tau_{photon} = \lambda Q/2\pi c$ is the photon lifetime in the disk cavity. Given the Q of the disk in the simulations, τ_{photon} is expected to be 25 ps. Because this is generally much longer than the RC constant of the contact points, the switching delay is primarily limited by the photon lifetime, resulting in a 28 Gb/s maximum NRZ bit rate. Note that while reducing Q (increasing resonance peak width) reduces delay, a corresponding increase in voltage is required to maintain the same modulation depth. This introduces a significant speed–power tradeoff. A better path of innovation comes with a more efficient use of each resonator in the logic architecture (this work), by increasing switching efficiencies ($\Delta\lambda/V$) through better junction-to-mode overlapping, or by improved etching and lithography to further reduce the minimum radii of fabricated disks. Smaller devices lead to increased computational throughput per chip and decreased total capacitance and dynamic energy consumption.

Note that carrier-based switching transfers heat from the electrical contacts to the disk, adding a significant albeit slow redshift in the transmission spectrum. High-Q, high-sensitivity devices, such as this one, are especially prone to errors from internal and external sources of heating and cooling. Therefore, integrated thermal controllers may be needed for low-error operation. These have been reported to only add low femtojoule levels of energy consumption per bit [55].

2.3 Ultracompact Logic Architecture Based on Microdisk Modulators

One of the ultimate goals of this optical logic architecture is to increase the functionality of each component, as silicon photonic structures are relatively large compared to CMOS transistors. As introduced in the previous section, the proposed optical logic cell splits the pn diode of the microdisk to create two independent diodes. Using WDM, each fundamental logic operation can be performed with a single microdisk. This all-gate cell is given in comparison to previous demonstrations of CMOS-compatible optical logic that use one or (more often) more switches for each gate operation. A summary of different devices is shown in Table 2.2. Note that the area values are based on simple calculations of waveguide and resonator dimensions, and the space in between each structure is not accounted for. These values can potentially be reduced further by using lower-radii microresonators. Care must also be taken in comparing the ideal energy and speed figures of simulations with experimental works, as certain factors have not been accounted for (e.g., RC and propagation delays). With all factors considered, the proposed cell in this work still improves computational throughput and efficiency over previous demonstrations of optical logic through compact cell- and circuit-level designs and state-of-the-art modulation. Furthermore, the cell can easily expand to perform compact higher-order operations like N-bit AND/OR gates, adders, comparators, encoders, and decoders that are each the smallest reported of their kind in literature. The cells in these circuits follow the directed logic paradigm [57], like certain previously reported works [18, 19, 20, 21, 22], in that they each switch independently of each other, simultaneously. Unlike conventional logic, that accumulates delay at each sequential gate, the relative throughput of this type of architecture grows with the size of the circuit.

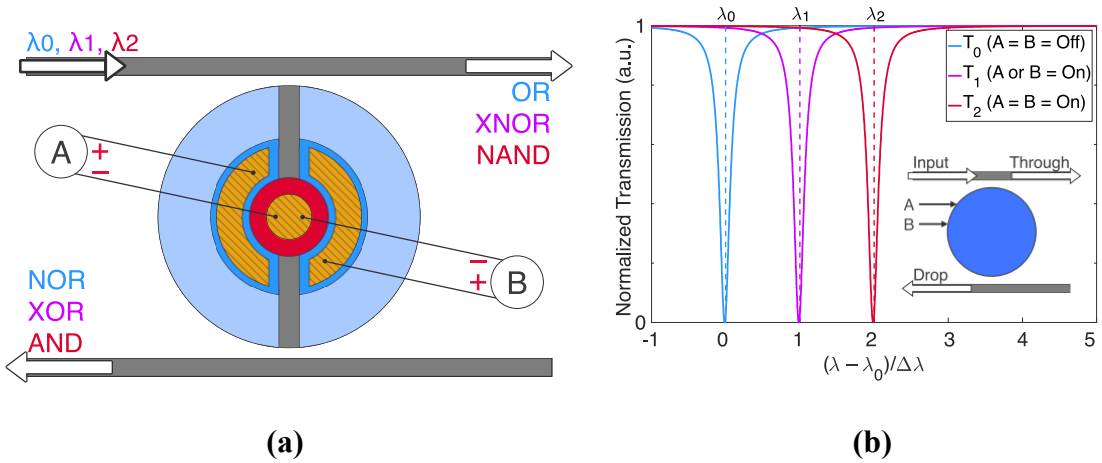
Table 2.2 Comparison of previously reported CMOS-compatible optical logic devices.

Structure	Energy (fJ/bit)	Speed (Gb/s)	Area (μm^2)	Demonstrated Logic Function(s)	Required # of Devices
TPA Waveguide [32]	2×10^3	10	4.8×10^3	NOR	1
Thermal Ring [20]	—	20×10^{-6}	0.9×10^6	X(N)OR	2
FCD Ring [58]	3.2×10^3	0.31	90	(N)AND	1
FCD Ring Depletion [19]	1.2×10^3	3	8×10^3	OR, AND, X(N)OR, A + B', AB'	8
FCD Ring Injection [21]	—	0.1	630	X(N)OR	2
FCD Ring Depletion [22]	—	12.5	630	X(N)OR	2
This Work (Simulation)	12	28	15	(N)OR, (N)AND, X(N)OR	1

This optoelectronic logic architecture is based on a fundamental logic cell that performs each of the fundamental logic operations depending on the wavelength of the signal sent through it. The design of this logic cell is critical to the overall performance of the logic circuit and was designed carefully because of this. The design of the passive, microdisk resonator structure was designed in conjunction with the active, integrated pn diode that modulates it. In this section, the design of the logic cell is outlined, and the system-level architecture is described with multiple examples of higher-order directed logic circuits. The fabrication of the logic cell, as well as the experimental results are presented, and a discussion is made on the outlook of this architecture—and optical logic in general—for next-generation information processing.

2.3.1 Optoelectronic Logic Cell: Simultaneous (N)AND, (N)OR, and X(N)OR Operations

The logic setup is as shown in Figure 2.6. Three different wavelengths are inputted into the cell; they operate at 0, 1, and 2 $\Delta\lambda$ away from the disk's unbiased resonance peak of interest and are used to perform the (N)OR, (N)AND, and X(N)OR operations, respectively. Note that $\Delta\lambda$ is the wavelength shift induced by a single 2 V control (on either diode). λ_0 is centered 0 $\Delta\lambda$ away from the unbiased resonance peak. If neither of controls A or B are on, the signal propagates to the drop port and sets NOR high. If either or both of the controls are on, the resonance peak shifts past the signal, which propagates to the through port, setting OR high. λ_1 is centered 1 $\Delta\lambda$ away from the unbiased resonance peak. If neither of the controls are on, the signal propagates to the through port and sets XNOR high. If only one control is on, the resonance peak shifts into alignment with the signal, which propagates to the drop port, setting XOR high. Should both controls be on, the peak overshoots the signal, which reverts back to its original, XNOR output. Lastly, λ_2 is centered 2 $\Delta\lambda$ away from the unbiased resonance peak. If neither of the controls are on, the signal propagates to the through port and sets NAND high. The same result occurs for the one-control case, as the peak has not shifted far enough to bring the signal onto resonance. When both controls are on, the resonance peak shifts into alignment with the signal, which propagates to the drop port, setting AND high. Note that with XOR and AND operations occurring simultaneously, this logic cell also produces the SUM and CARRY bits of a half adder—bringing 7X total logical functionality per cell.



(a) **(b)**

Figure 2.6 Setup of the fundamental logic cell. Three logic signals, with separated wavelengths, are inputted into the top bus and propagate to the through or drop ports as a function of the two electrical controls, A and B. (b) Given the electrical input combination, the through-port transmission spectrum of the cell shifts by 0, 1, or 2 Δλ to bring onto resonance the NOR, XOR, and AND operations, respectively. At any given time, one drop-port operation and two through-port operations are true. Inset shows a simplified symbol of the logic cell.

2.3.2 Higher-Order Logic Operations with Multiple Logic Cells

The optical logic cell serves as a building block to higher-order computing operations. These operations generally require the output of one stage of logic to directly drive the input of the next; however, this necessitates the use of costly OE converters and adds inherent delays from its one-after-another operational structure. Instead, a directed logic approach is used, in which multiple logic cells are concatenated, and a single or multiple optical logic signals propagate from input to output according to which cells are in on or off resonant modes. All cells are switched simultaneously in this approach, giving it a significant advantage over conventional logic, as switching delays do not scale with the size of the operation. To specify each cell's gate operation(s), their resonance peaks of

interest are designed (or perhaps tuned) to operate 0, 1, or 2 $\Delta\lambda$ away from the optical logic signal(s). As a proof of example, this approach is used to perform five important logic circuits in modern computers: N-bit AND/OR gates, adders, comparators, encoders, and decoders. Low-bit-count examples are shown here, with additional discussions on their bitwise expansion. With ultracompact design, simultaneous cell switching, and the use of high-performance silicon-photonic modulation schemes, these circuits show significant improvements over previous demonstrations of optical logic while introducing certain advantages over digital logic.

Performing optoelectronic operations with multiple microresonators/modulators can be made difficult when thermal fluctuations and fabrication variations are significant. As each MRR must operate at a precise resonance wavelength in relation to the other MRRs, keeping them aligned further necessitates the use of active tuning circuitry. The logic cell presented in this work features a special advantage over previous demonstrations of MRR-based logic, as it is the only device that performs the fundamental logic operations in just one resonator—alleviating this resonance alignment issue. However, multiple resonators are still required for higher order operations, and will therefore require active alignment.

For any optoelectronic operation that features multiple, cascading devices, the potential issue of accumulated optical loss must be considered. Unlike traditional, CMOS-based logic, where the level of the signal is restored after every gate, the directed logic architecture leaves the inputted optical signal(s) untouched throughout the circuit. For the higher order circuits presented in this work, the optical signals do not resonate with more

than two microresonators, and therefore would not accumulate too much loss. For much larger circuits, however, a simple optical signal restoration step may be necessary [52]. For the circuits presented here, which are based on the ultracompact logic cell, much fewer devices are required per operation, which therefore reduces optical loss over previous demonstrations.

Figure 2.7 shows a 4-bit AND/OR gate using two drop-port coupled logic cells. λ_0 and λ_2 ($\lambda_0 + 2\Delta\lambda$) are inputted at the top waveguide. When all four input controls (A, B, C, and D) are on, λ_2 propagates down to the drop port of Cell 1 as a function of AND4; otherwise, it exits at one of Cell 0's output ports, where it is to be filtered out or accepted as a function of NAND4. When no controls are on, λ_0 propagates to the drop port of Cell 1, where it is to be filtered out or accepted as a function of NOR4. If A or B are switched on, λ_0 propagates to the through port of Cell 0; if not, and C or D are switched on, it will propagate to the next drop port. Note that the waveguides shown in these figures are kept short for visual demonstration only; further routing may be needed in the fabrication layout. For larger bit counts, additional cells are coupled by their drop ports. N/2 disks are used for N-bit AND/OR operations, resulting in an ultracompact 1/4 disks per operation per bit (1/8 if NAND and NOR operations are accepted). The only previously demonstrated microresonator-based optical N-bit AND/OR gates are with reconfigurable directed logic circuits, which use one ring per operation per bit [18].

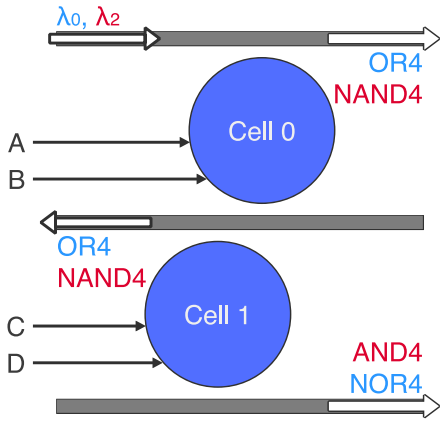


Figure 2.7 Schematic of the 4-bit AND/OR gate. Two drop-port coupled logic cells simultaneously perform AND4 and OR4 operations (and their inversions) on two separate optical logic signals.

Figure 2.8 shows a 4-bit comparator using four through-port coupled logic cells. A comparator is a fundamental logic circuit, used in various stages of a microprocessor's pipeline (decoding, arithmetic, addressing), that determines the equality of two binary numbers. If a greater/less than comparison is required, an additional four cells are drop-port coupled underneath. λ_1 is inputted into Cell 0 and propagates to the through port of Cell 3 as a function of

$$Z = (\overline{A_3 \oplus B_3})(\overline{A_2 \oplus B_2})(\overline{A_1 \oplus B_1})(\overline{A_0 \oplus B_0}). \quad (2.2)$$

Recalling the operation of a single cell, λ_1 operates at 1 $\Delta\lambda$ away from the resonance peak of interest and exits at the through port as a function of XNOR. Passing through multiple, serially coupled cells ANDs multiple XNOR operations together. If $A_n \neq B_n$, λ_1 resonates with Cell n and propagates to the corresponding cell underneath to check which input triggered the inequality. Only one input control is necessary for this check cell—say A_n—

because it is centered on λ_1 (red-detuned 1 $\Delta\lambda$ from the standard cell). If the logic signal exits at the through port, it is because A_n switched it, and therefore A > B. Otherwise, if the signal exits at the drop port, A < B. Note that, again, additional routing and coupling is required to combine the multiple output ports of the greater/less than operations. This circuit presents the most compact design for an optical bitwise comparator, simply using N disks for an N-bit equality comparison, and an additional N disks for the greater/less than functionality. Previously reported microresonator-based bitwise comparator designs use 59 rings (plus an OE conversion) for a 4-bit operation [18] and 17 rings for a 2-bit operation [59].

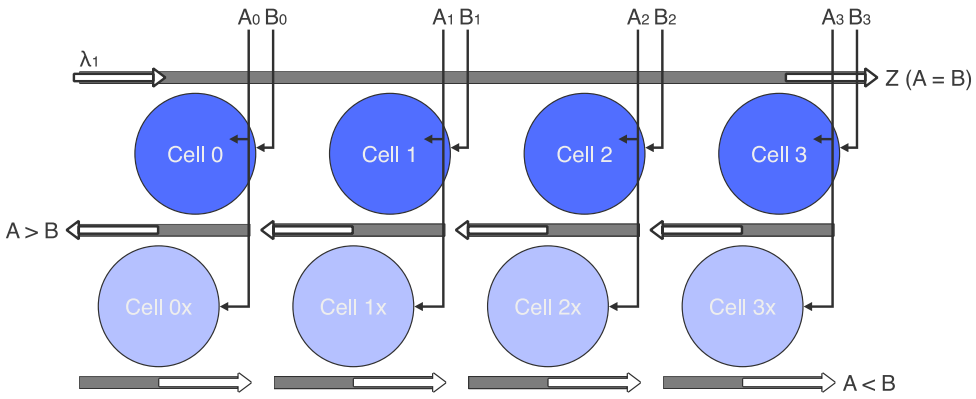


Figure 2.8 Schematic of the 4-bit comparator. For through-port coupled logic cells determine if two 4-bit numbers are equal. Cells 0x–3x are added to determine which number is larger; they are aligned with λ_0.

Figure 2.9 shows a 2-bit optical adder. An adder is a fundamental digital circuit of a microprocessor's arithmetic logic unit (ALU), in which every instruction of a computer is passed through. It is also used to generate memory addresses for loading and storing data. The proposed design combines a single-cell half adder stage and a multicell full adder

stage. Each stage has two optical inputs, $\lambda_{1,n}$ and $\lambda_{2,n}$, and two electrical controls, A_n and B_n, that work to create the SUM and CARRY outputs. $\lambda_{1,0}$ and $\lambda_{2,0}$ are inputted into Cell 0 to produce S_0 and C_{in} as functions of $A_0 \oplus B_0$ and $A_0 B_0$, respectively. The outputs of the full adder are defined as

$$S_1 = A_1 \oplus B_1 \oplus C_{in} \tag{2.3}$$

$$S_2 = C_{out} = A_1 B_1 + C_{in}(A_1 \oplus B_1). \tag{2.4}$$

S_1 is set high if only one of A_1, B_1, or C_{in} are high—or if all three are. Cell 1 produces S_1 from $\lambda_{1,1}$ as a function of $A_1 \oplus B_1$. Cell 2 is introduced underneath it, and blue-detuned 1 $\Delta\lambda$, to cancel S_1 if C_{in} is high (a function of $A_0 B_0$). Cell 3 (red-detuned 1 $\Delta\lambda$) cancels S_1 as a function of $A_1 \oplus B_1$ to round out the possible cases for (2.3), while simultaneously producing the second product of (2.4). Cell 4 works on $\lambda_{2,1}$ as a function of $A_1 B_1$ to produce the first product of (2.4). As a sum of products, Cells 3 and 4 can set S_2 high. This presents the most compact design for an optical adder, using $4N - 3$ disks for an N-bit operation. Previously reported ring-based adder designs use two rings for a half adder [51] and 15 rings for a 2-bit adder [60].

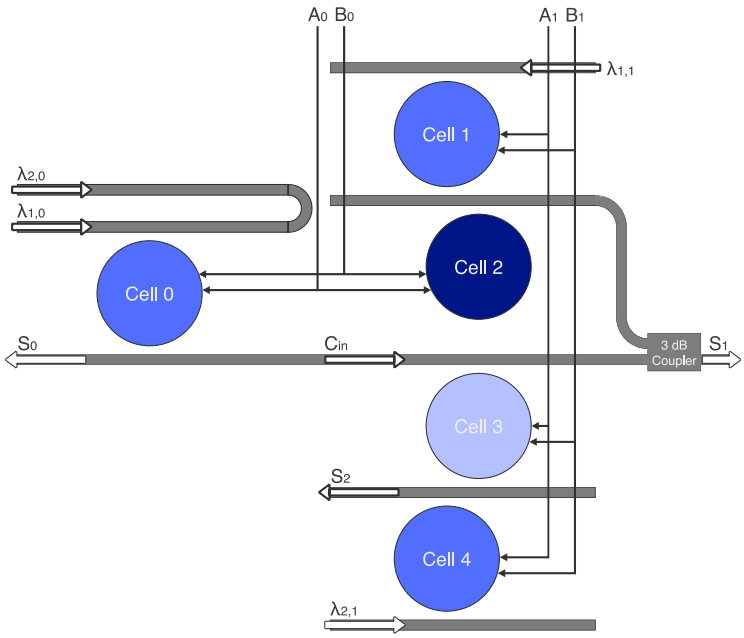


Figure 2.9 Schematic of the 2-bit optical adder. Cell 0 represents the half adder of the first stage of addition, and Cells 1–4 represent the full adder of the second stage. Cells 2 and 3 are blue- and red-detuned 1 Δλ away from the standard (blue) logic cells, respectively. For N-bit operations, N – 1 full adders are connected in series, for a total of 4N – 3 disks.

Figure 2.10a shows a 4-bit encoder. This circuit compresses a 4-bit number into a 2-bit binary representation of itself, as shown in Table 2.3, for reduced-bandwidth-use data transfer. This is done by two separate suboperations,

$$Q_0 = A_1 A_3 \tag{2.5}$$

$$Q_1 = \overline{A_0}\,\overline{A_1} = \overline{A_0 + A_1}, \tag{2.6}$$

Table 2.3 Truth table for a 4-to-2 binary encoder.

A_3	A_2	A_1	A_0	Q_1	Q_2
0	0	0	1	0	0
0	0	1	0	0	1
0	1	0	0	1	0
1	0	0	0	1	1
0	0	0	0	X	X

performed by two separate logic cells. Because these cells are not connected, they can be arranged differently for each application-specific circuit for a more optimized use of space. Q_1 is converted from an AND operation of two inverted control signals, by De Morgan's law, to a NOR operation of two noninverted controls. For higher-order encoding, the necessary AND and OR operations are added. For example, an 8-to-3 encoder can be reduced to three 4-bit OR gates.

$$Q_0 = A_1 + A_3 + A_5 + A_7 \tag{2.7}$$

$$Q_1 = A_2 + A_3 + A_6 + A_7 \tag{2.8}$$

$$Q_2 = A_4 + A_5 + A_6 + A_7, \tag{2.9}$$

which require six logic cells. This is reduced to five cells, given that the $A_5 + A_6$ operation is shared amongst two of the three operations. This is the most compact design for an optical encoder, using two and five disks for the 4-to-2 and 8-to-3 encoders, respectively.

The only previously reported design is an 8-to-3 priority encoder, which is inherently more complex, using 47 microresonators and an OE conversion [18].

Binary decoders are also more complex in design. Figure 2.10b shows a 2-to-4 decoder using five logic cells. This circuit produces the original 4 bits of the 2-bit encoded number by

$$D_0 = \overline{A_0}\,\overline{A_1} = \overline{A_0 + A_1} \tag{2.10}$$

$$D_1 = A_0\,\overline{A_1} \tag{2.11}$$

$$D_2 = \overline{A_0}\,A_1 \tag{2.12}$$

$$D_3 = A_0\,A_1. \tag{2.13}$$

D_0 and D_3 are combined into logic cell $D_{0/3}$, which performs the necessary AND and NOR operations. The remaining two operations require an inverted control signal, which cannot be done easily with our logic cell. Instead, two drop-port coupled double disk resonators are used, where each switch is controlled by a single input, A_0 or A_1. The double disk for the D_1 operation outputs a high bit when A_0 is high and A_1 is low. This is done by blue-detuning Cell D $1\Delta\lambda$ away from the location of the standard logic cell, while Cell D′ remains the same. The double disk for the D operation is configured in a similar way,

except that Cell D′ is blue-detuned, and Cell D_2 is not. This is the only microresonator-based optical binary decoder.

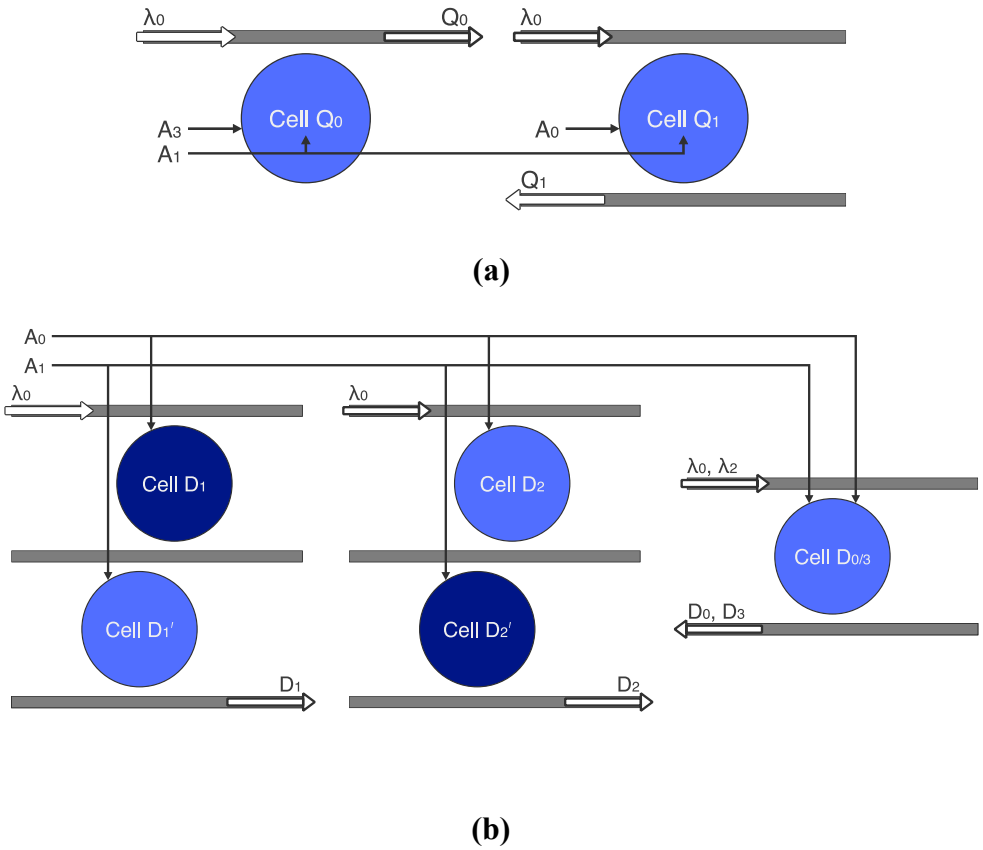


Figure 2.10 Schematics for the (a) 4-to-2 encoder and (b) the 2-to-4 decoder. Each circuit is a collection of AND and OR gates that can be arranged for a more optimized use of space in application-specific circuits. The dark blue cells are blue-detuned 1 Δλ away from the standard blue cells.

2.4 Summary

This chapter presents a high-throughput CMOS-compatible optoelectronic directed logic architecture that uses ultracompact vertical pn junction microdisk modulators to achieve a higher Δλ/V than conventional microring modulators; two electrical controls per switch to (at least) halve the size of logic operations; and wavelength-division multiplexing to

perform (N)OR, (N)AND, and X(N)OR operations simultaneously on a single switch. These basic gates are expanded on to create the most compact CMOS-compatible designs for N-bit logic gates, adders, comparators, encoders, and decoders. By using the fewest amount of (high-performance) switches per logic operation, this architecture can achieve the highest throughput and lowest energy consumption of all previously reported optical logic. Performance can be further optimized by reducing disk radii (resulting in reduced area and capacitance/energy) and enhancing junction-to-mode overlap for ultralow-voltage operation.

Chapter 3: Optimization of Ultrahigh-Efficiency Optoelectronic Modulators

To understand the true power-saving potential of optimizing the positioning and orientation of a single pn junction, a comprehensive comparison is made between the conventional, lateral pn junction and its equivalent vertical pn junction. Following that, two novel configurations of the vertical junction design are introduced to further improve the power budget: the double vertical and staggered double vertical pn junctions, which overlap even stronger with the resonant optical signal. These designs consume over 4 times less energy than the conventional, lateral design, providing a pathway to the attojoule-per-bit target set for next-generation information processing and communications [13].

Efficient pn junction designs are useful for a wide variety of optical computing and communications and devices—not just for low power, but for high operating speeds as well. This chapter also explores the high-bandwidth design of multiple optoelectronic modulators and switches using the efficient double vertical junction designs presented in this work.

The work on pn junction comparisons was presented in PIERS 2019, with a publication in the IEEE Journal of Selected Topics in Quantum Electronics in press. And the work done on ultrafast optoelectronic modulators for logic, communications, optical repeaters, and wavelength converters was published in Optics Express in 2020.

3.1 Comparing the Four pn Junction Designs

Four silicon microdisk/ring modulators, as shown in Figure 3.1, are designed with like parameters to establish a common base of comparison in each device. The main difference in each design is the orientation and placement of the integrated pn junction(s). These differences determine the amount of interaction between the free carriers of the doping regions and the optical mode of the resonator cavity; where, ultimately, more interaction means better power performance from the modulator. For the vertical-junction devices, microdisks are chosen rather than microrings due to the ease of electrical access to the doped regions of a vertically oriented pn junction. The passive silicon microdisk structures are all 0.22 μm thick, have 5 μm radii, and are hosted between upper and lower silicon dioxide cladding layers. A 50-nm-thick slab regions is added to each device to host the outer contact electrode. Although this slab is unnecessary for the microdisk modulators (both contacts can be integrated within the boundaries of the disk), it allows the n+ and p+ regions to be separated, which significantly reduces the device capacitance. The microring cavity, which is otherwise identical to the microdisk, is 500 nm wide. The dimensions of each device's doping regions are optimized to produce the largest change in effective refractive index (Δn_{eff}) of the fundamental transverse-electric mode (TE_0) between 0 and −1 V applied biases. For consistency, the peak carrier concentrations of all n and p regions are set to $N = N_A = N_D = 2\times10^{18}$ cm^{-3}, and all p+ and n+ regions are set to 10^{20} cm^{-3}.

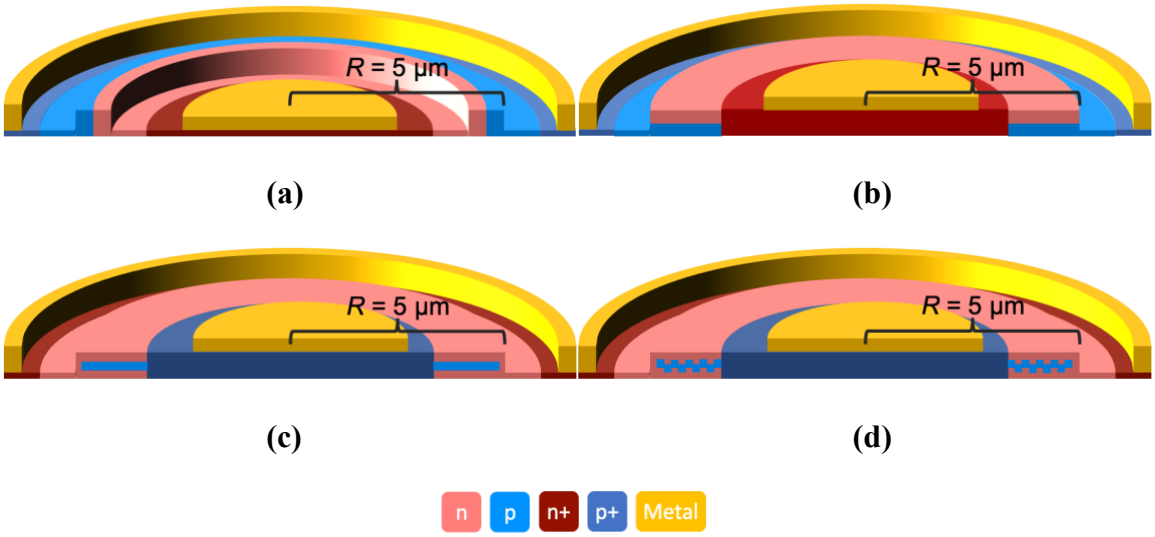


Figure 3.1 Cross-sectional views of the (a) lateral, (b) vertical, (c) double vertical, and (d) staggered double vertical pn junction microring/disk modulators. Note that the coupled waveguides are not shown in these views.

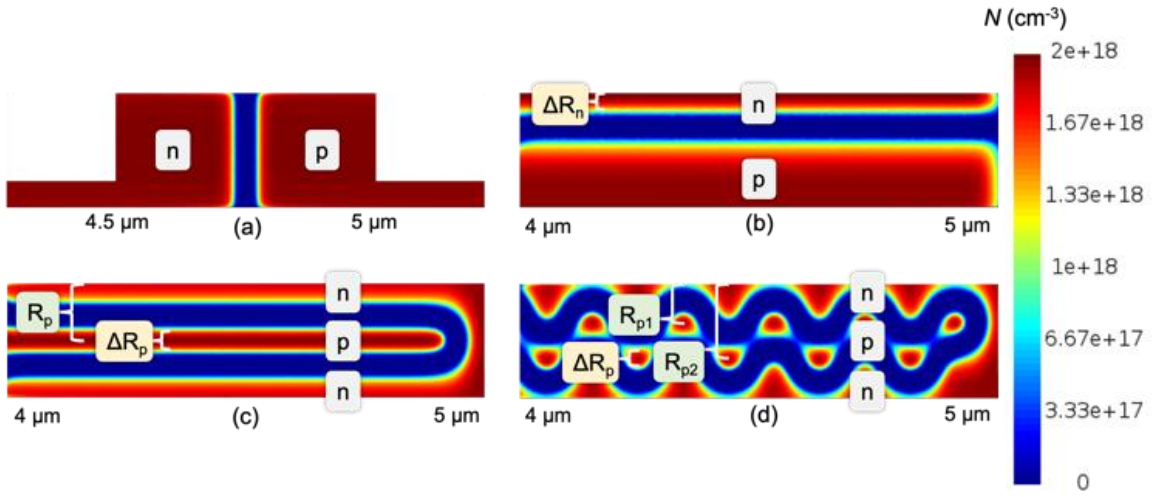


Figure 3.2 Cross-sectional views of the simulated doping profiles of the (a) lateral, (b) vertical, (c) double vertical, and (d) staggered double vertical pn junctions. Red regions are doped (n and p); blue regions are the depletion regions.

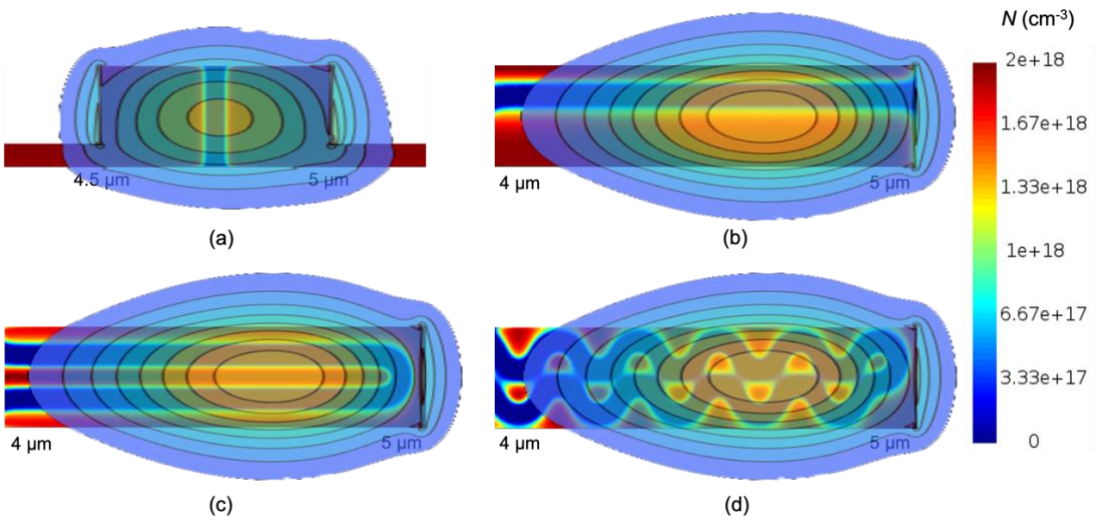


Figure 3.3 Simulated resonant mode profiles overlapped with the simulated free-carrier profiles of the (a) lateral, (b) vertical, (c) double vertical, and (d) staggered double vertical pn junctions.

The lateral pn junction modulator, shown in Figure 3.1a, features an n region in the interior half of the ring, and a p region around the exterior. A smaller-radius, high-concentration, n+ region is also located in the interior slab of the ring, with an anode electrode hosted on top of it. A high-concentration, p+ region is located around the outer slab region of the ring, with a cathode electrode hosted on top of it. In total, this device requires four ion implantation (doping) steps. The high concentration doping regions simply provide a better contact between the pn junction and the contact electrodes. These regions are placed just far enough away from the resonant mode to not introduce excessive carrier-induced optical loss. Each doping region extends from the top to the bottom of the silicon layer; the only optimization that is made is on the lateral positioning of the pn junction.

The optimized lateral positioning of the pn junction is found by running a parametric sweep of the junction location from the interior edge of the ring (4.5 µm) to the exterior (5 µm).

At each location, a simulation was run in a finite-element drift–diffusion solver (Lumerical CHARGE) to find the free-carrier concentration of a 2D slice around the pn junction: one simulation was run for a 0 V applied bias, and another for −1 V. The two free-carrier concentration profiles were then exported to a finite-difference eigenmode solver (Lumerical MODE) to find the effective refractive index of the resonant TE_0 mode. To apply the plasma dispersion effect to the simulation, the solver's Soref–Bennett index-perturbation material model is used. For the subsequent devices, the same optimization process is used, but with a different sweep of the junction parameters.

The vertical pn junction modulator, shown in Figure 3.1b, features an n region at the top of the disk, and a p region at the bottom. Because the optical mode of a microdisk resonator is more confined vertically, it is more sensitive to any changes in the depletion width of a vertical junction than that of a lateral junction. Like the microring modulator, the n+ region and cathode contact are positioned in the center of the disk, and the p+ region and anode contact are positioned around the disk. Although the device can be made more compact by integrating both electrodes within the boundaries of the disk, this approach removes the high-concentration, high-capacitance pn junctions—which don't contribute to modulation—by placing the n+ and p+ regions on opposite sides of the main pn junction. The microdisk still offers the benefits in ease of contact placement on the top (n) region of the modulator, as well as the elimination of the loss-inducing etched inner sidewall of a microring cavity.

The microdisk still offers the benefits of low resistance and ease of contact placement for the top (n) region of the modulator. In addition, the microdisk configuration eliminates the loss-inducing etched inner sidewall of a microring cavity, which can contribute over 15 dB/cm for waveguides of a similar size [61]. Furthermore, microdisks have been shown to be more tolerant to fabrication deviations and have higher thermal tuning efficiencies [62].

Similar to the case of a lateral-junction device, this device requires four ion implantation steps; however, it does require precise control of the range (depth) and straggle (spread) of the n and p implantation steps, which can vary after thermal annealing (implant activation). Although the tight vertical confinement of the resonant optical mode may make the performance of the device less tolerant to fabrication variations, multiple successful experimental demonstrations of this junction have been reported [55, 56].

The double vertical pn junction modulator, shown in Figure 3.1c, features a middle p region sandwiched by the two n regions at the top and bottom of the disk. This doping distribution creates two vertical pn junctions, effectively doubling the interaction with the optical mode. Note that the middle p region is kept shorter than the microdisk radius as this creates a connection point between the upper and lower n regions. The p+ region and anode contact are positioned in the center of the disk, and the n+ region and cathode contact are positioned around the disk. The implantation process is similar to that of a single vertical junction device, except that the p region is implanted deeper into the disk. Although this junction adds further fabrication complexity, similar implant sequences have been experimentally

demonstrated in the literature [63], indicating the fabrication feasibility of the proposed double vertical junction.

The staggered double vertical pn junction modulator, shown in Figure 3.1d, features similar doping regions to the double vertical pn junction modulator, but with a staggered profile for the p region. Adding a lateral component to a vertical pn junction simply extends its interaction length with the optical mode. The staggered profile is formed with two separate p implants. Each implant is masked with an on–off periodic pattern, but one has a half-pitch lateral shift and is implanted with a different range. This device is the most complex of the four to fabricate, but still only requires five implant steps.

3.1.1 Optimizing the Four pn Junction Designs with Common Parameters

After the optimization of each of the four modulator configurations, where the implant parameters are set to what produces the largest change in effective refractive index per volt ($\Delta n_{eff}/V$), the remaining optical and electrical properties of the optimized devices are found and compared. The optimized carrier concentration profiles are used to find the carrier-induced optical loss (α_{0V}), resonance wavelength shift per volt ($\Delta\lambda/V$), device capacitance (C), minimum operating bias (V), and dynamic energy consumption per bit (E/bit). As the devices increase in design complexity, the performance improves significantly.

In the simulations using Lumerical CHARGE, the change in refractive index per volt $\Delta n_{eff}/V$ is obtained by first running two steady-state finite element simulations (at biases of 0 V and −1 V). The applied electric field and free-carrier density profiles of each device

model were calculated by solving the Poisson and drift–diffusion equations. The built-in semiconductor material model for silicon was used (with the trap-assisted, Auger, band-to-band tunneling, and radiative recombination models), and the implant profiles were created by specifying the positioning, concentration, range (R), and straggle (ΔR) of each n- and p-type implant. The simulated free-carrier density profiles were then exported to the MODE solver to calculate the difference in n_{eff} for the two states. Because these were 2D cross-sectional simulations of relatively small devices, small automated meshes (< 0.01 μm) were used throughout the optimization for high accuracy results within a reasonable timeframe. The simulated free-carrier profiles of each device are shown in Figure 3.2, and the overlapping optical mode profiles are shown in Figure 3.3.

The lateral pn junction modulator was optimized by sweeping the lateral position of the pn junction across the microring cavity. At its optimal position in the center of the cavity, which is offset 10 nm inwards from the center of the optical mode, the change in effective refractive index $\Delta n_{eff}/V$ is found to be 1.0×10^{-4} V^{-1}. As understood by the work done on offset-junction modulators, a depleted p region produces a larger change in refractive index than a depleted n region. Junction offsets as much as 50 nm towards the n region have been shown to produce greater index changes than with no offset [64, 65, 66, 67]; however, for this 5 μm microring, where $N_A = N_D$, the effect is less pronounced. Because of the relatively loose lateral confinement of the optical mode, the tolerance to fabrication error is relatively high, with 10-nm junction positioning variations only reducing $\Delta n_{eff}/V$ by 0.7%. Under no applied bias, the carrier-induced optical loss α_{0V} of the TE_0 mode is found, with the same Soref–Bennett eigenmode solver, to be 51.6 dB/cm. By swapping the n and p regions, so

that the p region is now in the interior half of the ring, and the pn junction is reoptimized to a position 10 nm outwards from the center of the mode, $\Delta n_{eff}/V$ and α_{0V} do not significantly change; this is because the optical mode is effectively symmetrical around the depletion region for this microring.

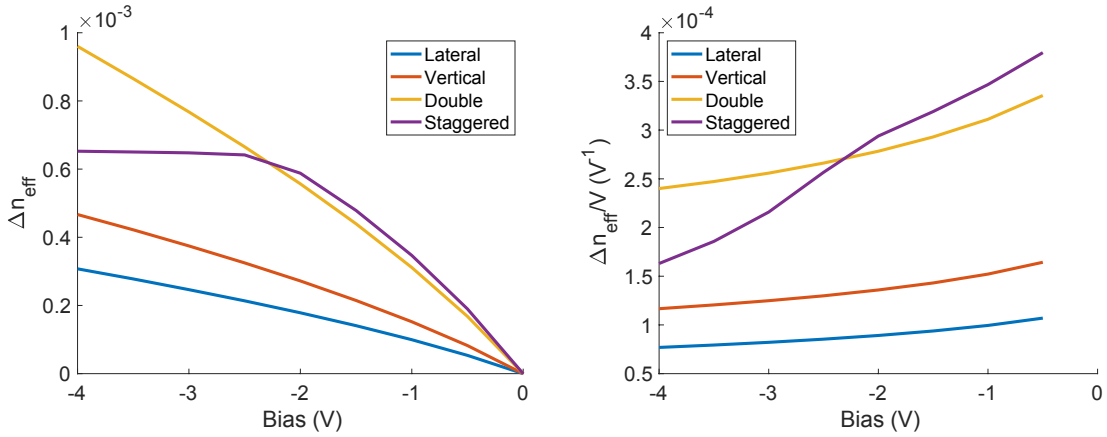
The vertical pn junction modulator was optimized by sweeping the straggle of the n region at the top of the disk. By sweeping the straggle of this implant, the vertical position of the pn junction is effectively being swept. At the optimal straggle of ΔR_n = 55 nm, which creates a pn junction 44 nm above the center of the disk, $\Delta n_{eff}/V$ of the resonant TE_0 mode is found to be 1.5×10^{-4} V^{-1}, which is 1.5 times larger than that of the lateral device. Because the vertical mode confinement is tighter than the lateral confinement, the effect of an offset pn junction is relatively high here; hence, the larger junction offset from the center of the mode. The tolerance to fabrication error is still relatively high, with a ±10% variation in ΔR_n resulting in only a 0.4% reduction of $\Delta n_{eff}/V$. α_{0V} is found to be 37.5 dB/cm, which is 1.4 times less than that of the lateral device. This reduction in carrier-induced loss is also due to the greater interaction between the optical mode and the (less lossy) depletion region. By simply changing the orientation of a single pn junction, the modulation efficiency and loss metrics are simultaneously improved, which both work to reduce energy consumption. By swapping the n and p regions, so that the n region is now at the bottom of the disk, and the pn junction is reoptimized to a position 44 nm below the center of the disk, there is no difference in performance, as the device is symmetrical along the lateral axis.

The design of the double vertical pn junction modulator is straightforward. Because holes produce a larger change in refractive index, an n-p-n vertical doping sequence is chosen. The range of the p implant is best set to the vertical center of the disk, as moving one junction closer to the center of the mode would move the other away from it. Ideally, the straggle of the p region is small, to keep the depletion regions closer to the center of the mode, but this is limited by the implant physics. As will be explained in Chapter 3.1.3, ΔR_p is limited to 40 nm, which creates junctions 48 nm above and below the center of the disk. $\Delta n_{eff}/V$ of the resonant TE_0 mode is found to be 3.0×10^{-4} V^{-1}, which is 3.0 and 2.0 times larger than the lateral and single-vertical devices, respectively. α_{0V} is found to be 30.8 dB/cm, which is 1.7 times less than the lateral device and 1.2 times less than the single-vertical device. The reduction in loss is attributed to the mode now overlapping even more with the depletion regions.

The staggered double vertical pn junction modulator was optimized by sweeping the relative ranges of the two p implants from the center of the disk. With both at the center of the disk, the modulator behaves like the double vertical pn junction device. The pitch of the implant mask is also swept, which determines the number of lateral components in each pn junction. With a mask pitch of 175 nm, and optimal range values of 75 nm for the first implant, and 145 nm for the second implant, $\Delta n_{eff}/V$ of the resonant TE_0 mode is found to be 3.4×10^{-4} V^{-1}, which is a slight increase from the double-vertical device. Although more of the perturbed depletion region overlaps with the optical mode now (from the combined lateral and vertical components of the staggered profile), the top-and-bottommost sections of the pn junctions are further away from the center of the mode; thus, the significantly

elongated pn junction does not significantly improve modulation efficiency. α_{0V}, however, is found to be 22.6 dB/cm, which is 1.4 times less than the double-vertical device and 2.3 times less than the lateral device. Again, the lower loss is attributed to more of the depletion region now being centered on the optical mode. For applications with strict loss budgets, this device is ideal.

Despite the double junction devices being significantly more efficient, they do show saturation at larger reverse biases. This is more evident for the staggered double junction device, as shown in Figure 3.4. The middle p region is very thin for the double-junction devices. At large reverse biases, they become fully depleted, resulting in no further increase in Δn_{eff}. This is evident by the flattening of the curve for the staggered device with large reverse biases in Figure 3.4a, and the dip in the curve in Figure 3.4b. This is not so much the case for the single-junction devices, where Δn_{eff} continues to increase for large reverse biases. However, for each device, a slight decrease in $\Delta n_{eff}/V$ is observed, as the depletion region slowly moves outward from the center of the optical mode, where the carrier depletion effect is the greatest. These issues are likely insignificant for the proposed double junction devices, as the operating biases are less than –2 V. For ultrahigh-speed, low-Q resonators, where larger applied biases would be necessary, this may be of significance.



(a) **(b)**

Figure 3.4 Saturation of Δn_{eff} (modulation efficiency) at large reverse biases (a) as is and (b) per volt for the four modulator configurations.

The change in resonance wavelength $\Delta\lambda$ is calculated by a 2.5D FDTD solver (Lumerical varFDTD) for the four optimized modulator designs. The calculated values are shown in Table 3.1, and the simulated transmission spectra of the four modulators are shown in Figure 3.5. The waveguide–ring gap size of the lateral-junction microring modulator is designed to achieve a minimum transmission of −10 dB. For this design, the quality factor $Q \approx 2{,}900$. Figure 3.5a shows the transmission spectra for the lateral-junction device under 0 and −1 V biases, as well as the bias required for shifting the resonance where the insertion loss (IL) and the ER are 2 dB and 7 dB, respectively. These target values are based on common values found in the literature [55, 56, 64, 65, 66, 67, 68, 69, 70, 71]. For microdisks with the same Q, the vertical-junction devices achieve progressively larger ER values, due to less carrier-induced loss, and require progressively lower biases to shift the resonance.

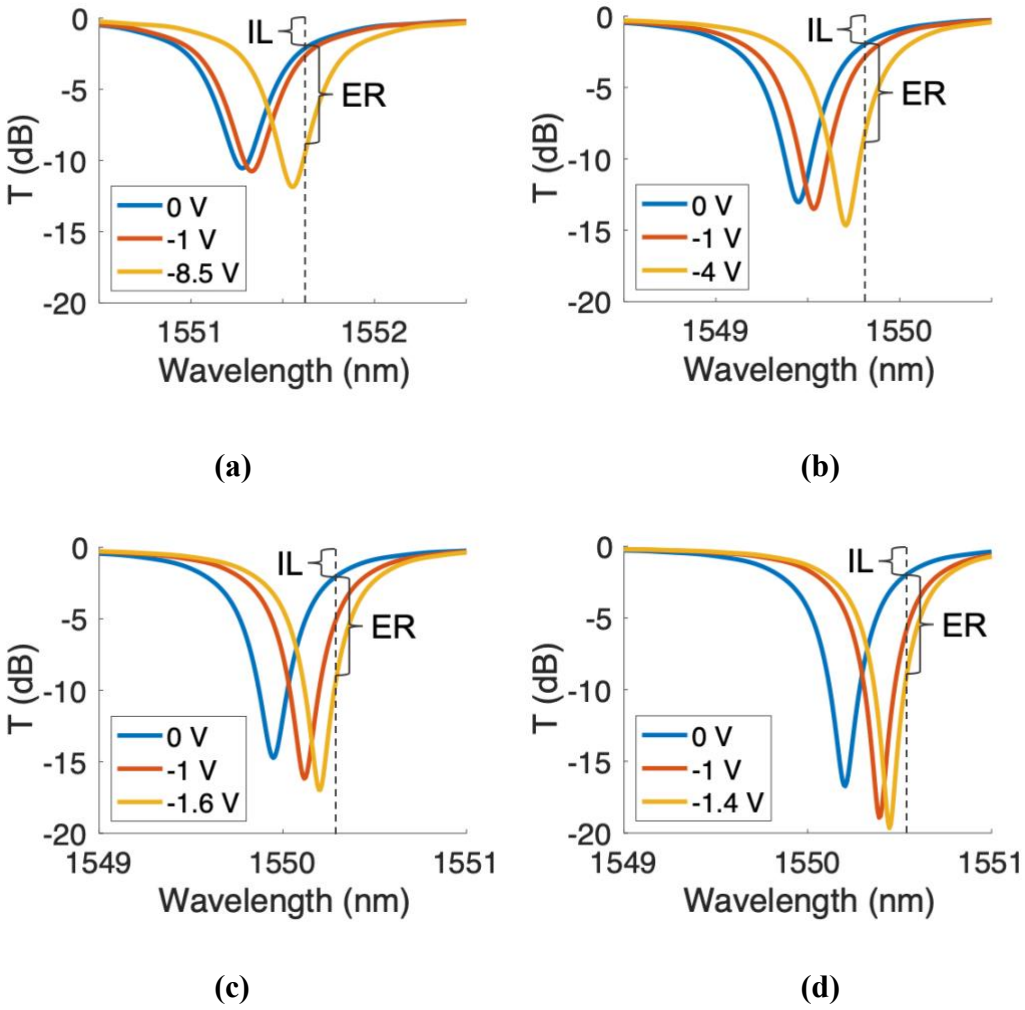


Figure 3.5 Through-port transmission spectra of the (a) lateral, (b) vertical, (c) double vertical, and (d) staggered double vertical junction modulator designs under 0 and −1 V biases, as well as the bias required for a full resonance shift (IL = 2 dB, ER = 7 dB).

Table 3.1 Performance properties of the four modulator designs for Q = 2,900, IL = 2 dB, ER = 7 dB

Device	$\Delta n_{eff}/V$ (V^{-1})	$\Delta\lambda/V$ (pm/V)	V (V)	C (fF)	E/bit (fJ/bit)	Improvement Factor
Lateral	1.0×10^{-4}	56	-8.5	7.40	134	1.0
Vertical	1.5×10^{-4}	88	-4.0	28.0	112	1.2
Double	3.0×10^{-4}	180	-1.6	71.0	45.5	2.9
Staggered	3.4×10^{-4}	201	-1.4	62.3	30.5	4.4

The quality factor of the resonator ultimately affects the modulation performance. A high-Q resonator has a narrower resonance peak, which requires less $\Delta\lambda$ to fully switch; thus, reducing the required applied bias and energy consumption. A high-Q resonator cavity also holds a photon for longer, which reduces the speed of modulation. The design optimizations in this work avoid this speed–power tradeoff, which enables the choice between same-speed, lower-power or same-power, higher-speed operation.

Of critical importance to the performance of these devices is the energy consumption per bit, where the ultimate goal is to reach the attojoule-per-bit mark for next-generation signal processing [13]. Energy consumption per bit is given by $E/bit = CV^2/4$, where C is the total device capacitance, and the factor of four is introduced by the fact that the 0-to-1 transition (of the four possible transitions) is the only one that draws power in the dynamic state. Leakage current is simulated to be on the picoamp scale for all four devices, so static power consumption is considered negligible. There is, however, the added cost of running the laser, which will not be included in these calculations, as it is a system cost. It is important to note though that the better-performing devices add less loss to the system and would potentially allow for lower optical operating powers. For the main, dynamic energy

consumption, the two factors create a natural tradeoff: A longer depletion region is more capacitive; however, it reduces the required operating bias by increasing the change in resonance wavelength. Because the dynamic energy consumption has a quadratic dependency on bias, the reduction of bias is prioritized over capacitance.

The device capacitance of each modulator was found using the charge simulations in Lumerical CHARGE and are listed in Table 3.1, where the longer pn junctions produce considerably higher capacitances. Although the staggered double junction is the longest, it has a lower capacitance than the simple double junction. This is because the carrier concentration profiles in the junction are not abrupt (i.e., the concentration gradually transitions from 0 to the maximum N_A or N_D), and the doping regions of the staggered junction that reach their peak concentration are much smaller than those of the simple double junction—resulting in a lower overall charge (capacitance). This also contributes to the low carrier-induced loss of the staggered junction.

Despite having a 1.5X larger $\Delta n_{eff}/V$, the single vertical junction only offers a moderate improvement over the lateral junction due to an increased device capacitance. However, significant energy efficiency improvements of 2.9X and 4.4X are found in the double vertical and staggered double vertical junction devices, respectively. As $\Delta n_{eff}/V$ is increased, the quadratic dependency V has on E/bit is reduced, allowing for more substantial energy efficiency improvements. For a more reasonable Q of ~9,000, the E/bit drops to 6.2 fJ/bit (at V = –0.5 V) for the same IL and ER. Above a Q of ~20,000, the device drops below 1 fJ/bit (at V < –0.2 V). The benefits gained from the large performance

improvements of the double-junction devices outweigh their slightly more difficult fabrication requirements.

The energy consumption of these devices can be further reduced with additional optimizations of the pn junctions. The peak doping concentrations were set to a reasonable range found from similar devices in the literature, but there is room for fine enhancements. As well, the radius of the microdisk can be reduced, as this reduces the device capacitance.

3.1.2 Characterizing the Speed Performance

Higher modulation efficiencies allow for higher-speed, lower-power or much-higher-speed, same-power operation. Figure 3.6 presents the speed–power characterization of the four devices compared in this work. The CHARGE solver was used again to calculate the capacitance, with an applied bias of between 0 and −9 V, as shown in Figure 3.6a. Note that as the reverse bias is increased, and the depletion width(s) of the junction(s) are widened, capacitance is reduced. Although a longer, more-efficient junction has more capacitance at smaller reverse biases, the middle region of the n-p-n junctions fully depletes at larger reverse biases; hence, the capacitance is significantly reduced. This carries large benefits in the speed and power performance of the device.

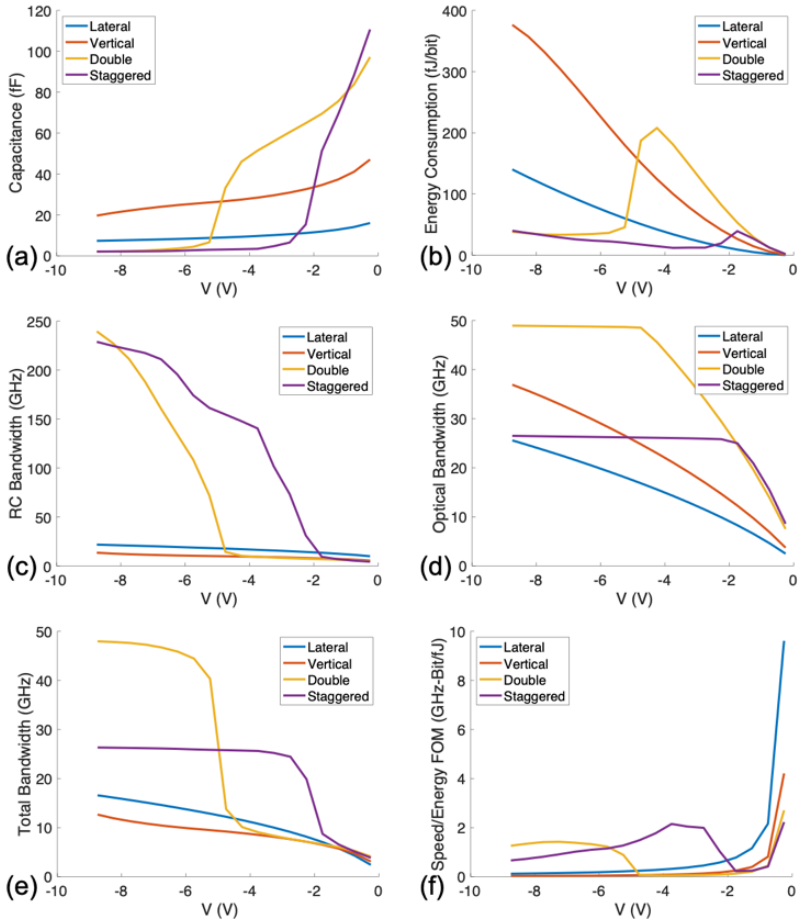


Figure 3.6 Simulated (a) capacitance, (b) energy consumption per bit, (c) electrical bandwidth, (d) optical bandwidth, (e) total bandwidth, and (f) bandwidth per energy figure of merit, with respect to reverse bias, for the four devices under comparison.

Figure 3.6b shows the energy consumption ($E/bit = CV^2/4$) of the devices, where a large drop in consumption is observed at large reverse biases. This large drop in capacitance promotes a large increase in the electrical bandwidth (B_{RC}) of the pn junction, as shown in Figure 3.6c. Here, $B_{RC} = 1/2\pi RC$, where the resistance R is calculated by CHARGE from the current between the two contact electrodes. The resistances of the lateral, vertical, double vertical, and staggered double vertical devices were calculated to be 999 Ω, 601 Ω, 334 Ω, and 334 Ω, respectively. As the microdisk devices allows for a thicker middle

electrode, the resistance of the single vertical junction device can be reduced by swapping the n and p regions, such that the low-mobility p region is in the middle, like it is for the double junction devices.

The optical bandwidth, which is shown in Figure 3.6d, is a function of $B_{opt} = \sqrt{\sqrt{2}-1}\, c_0 \delta\lambda/\lambda_0^2$, where $\delta\lambda$ is taken as the maximum $\Delta\lambda$ shift at a given reverse bias, and λ_0 = 1,550 nm. The higher-efficiency devices perform better here, though the aforementioned saturation of the double junction devices causes the bandwidth to saturate at larger reverse biases. The total bandwidth, as is shown in Figure 3.6e, is a function of electrical and optical bandwidth: $B = B_{RC}B_{opt}/\sqrt{B_{RC}^2 + B_{opt}^2}$.

A GHz-bit/fJ figure of merit is introduced and plotted in Figure 3.6f to visualize the speed-to-power ratio of each device. At reasonable operating voltages (less than −1 V is not feasible for the single junction devices, given the small resonance shift), the double junction devices perform better than the single junction devices. The most ideal device obtained is the staggered double vertical junction operating at –4 V (26 GHz, 12 fJ/bit).

Note that the plots in Figure 3.6 do not take IL and ER into account, and that Figure 3.6d, in particular, is based on a larger $\Delta\lambda$ than is necessary to maintain a large ER and low IL. However, the plots demonstrate the relative behavior of the devices. At the operating voltages presented in Table 3.1, the low-capacitance, high-energy devices generally perform faster; however, the double junction devices can be operated at higher voltages, as shown in Figure 3.6e, to achieve much higher speeds at reasonably low powers.

3.1.3 Fabrication Feasibility: Characterization of the Ion Implantation Process

To verify the feasibility of fabricating the proposed double vertical junction devices, and to obtain realistic range and straggle values for the CHARGE optimizations, a Taurus TSUPREM-4 [72] 2D process simulation is performed. Although similar junctions have been experimentally verified in the literature [63], the device dimensions presented here are different and must be characterized accordingly.

The double vertical junction can be created by ion implantation, first doping the entire disk n type, then implanting the middle p region. For a straightforward simulation, the doping is split into three different steps. The upper n region is formed by implanting phosphorus ions with an energy of 20 keV and a dosage of 5.1×10^{13} cm^{-2}; the lower n region is formed by implanting phosphorus ions with an energy of 120 keV and a dosage of 0.8×10^{13} cm^{-2}; and the middle p region is formed by implanting boron ions with an energy of 20 keV and a dosage of 5.1×10^{13} cm^{-2}. A rapid thermal annealing step of 1,000 °C for 5 s is used to activate the implants. With this implantation process, the straggle of the middle p region ΔR_p is found to be 40 nm, which is set as constant for the CHARGE optimizations. The n+ and p+ regions use a similar implant process, but with two orders of magnitude higher doses. The unbiased doping profile of this process is shown in Figure 3.7. The staggered double vertical junction follows a similar implant process. The implant is masked with a minimum dimension of 175 nm, which is of the order of the microdisk resonator gap size and can be resolved easily by modern e-beam and optical lithography processes. Based on the demonstrated process simulation results, and the chosen minimum feature sizes and

dimensions, although challenging, the double junction devices can be experimentally realized using currently available fabrication services.

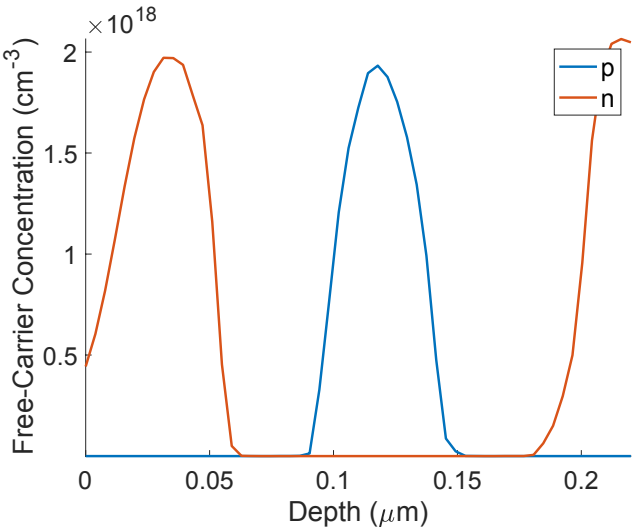


Figure 3.7 Simulated free-carrier concentration of the double vertical pn junction at 0 V bias.

The staggered double vertical junction follows a similar implantation process; however, it requires two p-type implants to form the staggered profile. Each p-type implant is masked by an on–off pattern. The pitch of the on–off pattern of the mask is 175 nm, which can be readily fabricated by available lithography processes. It is important to note that one mask must be offset radially by a half period to allow for the two different profiles: one with a shallower p region, and one with a deeper p region. The shallow-p profile is shown in Figure 3.8a and is formed by implanting BF2 with an energy of 40 keV and a dosage of 9.1×10^{13} cm^{-2}. In addition to using BF2 instead of boron, a 10 nm screening oxide is included in the simulation to create the shallow junction required for this device. The deep-p profile is shown in Figure 3.8b and is formed by implanting boron ions with an energy of 45 keV and a dosage of 3.9×10^{13} cm^{-2}. Because of the addition of the screening oxide,

and the different p implants, the n implants must also be adjusted from the double vertical junction device. The top and bottom n regions of the shallow stagger implant are set to 10 keV, 8.8×10^{13} cm^{-2} and 120 keV, 0.65×10^{13} cm^{-2}, respectively. The top and bottom n regions of the deep stagger implant are set to 10 keV, 1.3×10^{13} cm^{-2} and 120 keV, 0.9×10^{13} cm^{-2}, respectively. The same 1,000 °C, 5 s thermal annealing step is used for this junction. Note that the upper and lower n regions of the shallow and deep stagger implants in Figure 3.8, respectively, reach lower peak concentrations, as is also shown in the CHARGE simulation of Figure 3.2d. Based on the demonstrated process simulation results, and the chosen minimum feature sizes and dimensions, we conclude that, although potentially challenging, the double junction devices can be experimentally realized using currently available fabrication services.

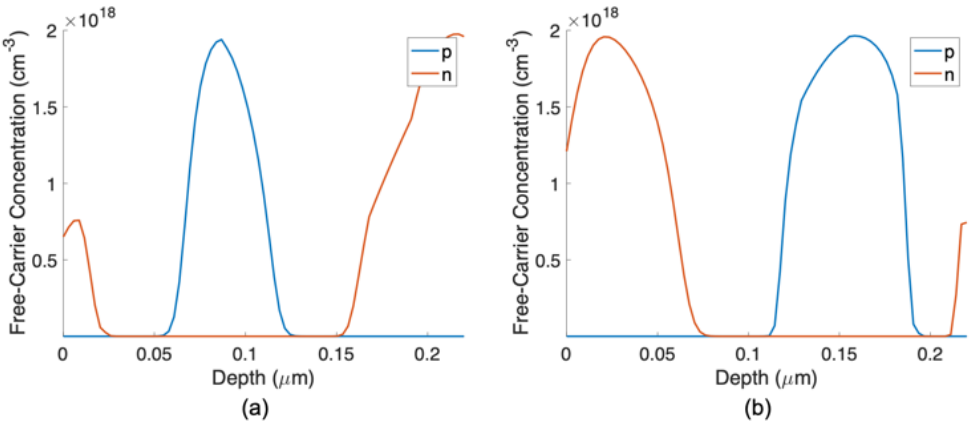


Figure 3.8 Simulated free-carrier concentration of the staggered double vertical pn junction at 0 V bias for (a) the shallow-p section and (b) the deep-p section.

3.2 Microdisk Modulators for Ultrahigh-Speed Applications

This theoretical, simulation-based study proposes a high-efficiency bus-coupled electro-optical microdisk modulator design for ultrahigh-speed and low-power communications and computing applications on the silicon-photonics integrated-chip platform. The proposed modulator features a depletion-mode U-shaped pn junction which maximizes the depleted junction overlap with the whispering gallery mode (WGM) of the microdisk. Consequently, a maximum change in the disk's effective refractive index can be induced, thereby leading to a large resonance wavelength shift per volt applied to the junction. The high modulation efficiency achieved by this design allows for a low-Q resonant operation at suitably low voltages to circumvent the speed–power trade-off in conventional resonator-based modulator designs [17, 64, 65, 66, 67] while maintaining an ultracompact footprint and high ER.

The modulator is designed on the SOI platform to enable low-cost and high-volume production of CMOS-compatible photonic and electro-optical (EO) circuits. The proposed modulator is optimized for use in optical communications, switching, and computing applications. The first device investigated is a stand-alone electro-optical modulator (EOM). A pulley-style coupler [52, 73, 74], in which the bus waveguide bends around the disk, is used to tailor the waveguide–disk coupling condition for a high ER with low Q factors required for operating at high speeds. A second bus waveguide, also in the pulley coupler configuration, is added in the second modulator device. This 2x2 device can be utilized in important communications and computing applications, such as electro-optical logic (EOL) gates [52, 75, 76, 77], WDM NxNxMλ spatial routing switches [78, 79], and

reconfigurable optical add-drop multiplexers (ROADMs) [80, 81]. For the two-bus configuration, the coupling angle of both pulleys is adjusted to maintain high ER and low Q factor values.

The implementation of the proposed modulator is also demonstrated in multifunctional optical-to-electrical-to-optical (OEO) devices for optical signal restoration-with-gain, wavelength conversion, and optical–optical logic (OOL). In these OEO devices, a CMOS-compatible, GeSn 1,550-nm photodetector (PD) is used to convert a premodulated optical input signal into a photocurrent, which drives the DC-biased circuitry of the EOM [52]. The EOM, in turn, modulates a separate, continuous wave (CW) optical signal injected from the disk's input bus. By using a similar or higher intensity for the CW signal, this process effectively restores the intensity of the original signal that is incident upon the PD, thus adding optical gain to that signal. The OEO device, in addition to working as an "amplifier-less" optical repeater, can work as a wavelength converter by using a different wavelength for the CW signal. The second application of the OEO configuration adds a second PD-driven electrical input and a second bus waveguide to the EOM to perform a full set of OOL operations. With the two (waveguide-PD) inputs required of a logic gate, this device simultaneously performs OR, AND, and XOR operations, as well as their inversions: NOR, NAND, and XNOR, employing only a single microresonator.

Given the high modulation efficiency of the depletion-mode U-junction modulator, the EOM, EOL, NxNxMλ cross-connect, ROADM, OEO repeater/converter, and OOL all operate, as optimized here, between 7.6 and 50 GHz in bandwidth, with suitably low levels

of energy consumption, between 0.4 and 9.8 fJ/bit, while maintaining more than 10 dB ER in an ultracompact, 2-μm-radius disk.

This work presents the design and optimization of the U-junction microdisk modulator and its implementation in several real and proposed applications in computing and communications. Primary design considerations are made to maximize operating bandwidth using CMOS-compatible features and materials. Although an experimental demonstration of the devices is not included here, special considerations on their fabrication feasibility are discussed and are supplemented with numerical analysis. This work presents a feasible design approach to enhance the performance of electro-optical devices in next-generation communications and computing circuits.

3.2.1 High-Bandwidth Resonant 1x1 and 2x2 Optoelectronic Devices

Figure 3.9 and Figure 3.10 present the four devices optimized for ultrahigh bandwidth. The real and potential applications of the devices are as follows. The modulator shown in Figure 3.9a could form the heart of a chip-scale data transmitter that feeds an optical fiber, or it could function as an EO logic element where many such 1x1s are deployed to create a complex on-chip logic photonic integrated circuit (PIC); while for the Figure 3.9b 2x2, the EO logic gates shall have complimentary outputs. The 2x2 switch in Figure 3.9b (often used as 1x2 in high-radix matrix switches [82]) also has some major applications in NxNxMλ cross-connect switches [83]. Interesting microring-based WDM cross-connect architectures have been proposed recently for waveguide-connected sets of the Figure 3.9b devices [78, 84, 85]. The performance predicted here for the Figure 3.9b disks suggests

that they are a better choice than EO microring resonators for the data-center matrix switching application. In Figure 3.10, each diode symbol represents a strip waveguide with its own waveguide PD at the strip end, while its output provides the electrical drive (which is demodulated from the original optical signal) for the new higher-intensity CW signal entering the disk. A required load resistor R_L is placed in shunt with the EOM input terminals because this resistance value is chosen to convert the PD photocurrent into the required reverse-bias voltage on the disk. The primary use of the structures in Figure 3.10 is the "amplifier-less" optical repeater; that is, a device that restores or increases the amplitude of light incident on the PD by transferring its data to a new higher-intensity CW signal. This OEO is equivalent in many ways to a semiconductor optical amplifier. The motivation for the OEOs is to deploy a group of such "gain" elements within a "large-scale" photonic integrated circuit to restore signals that become attenuated when traveling through the circuit. As a wavelength converter, this OEO design is attractive because it can change the operating wavelength while adding gain. The OOL application, while fascinating in principle, needs specialized computer architectures to fully exploit its potential in advancing the electronic–photonic digital computer.

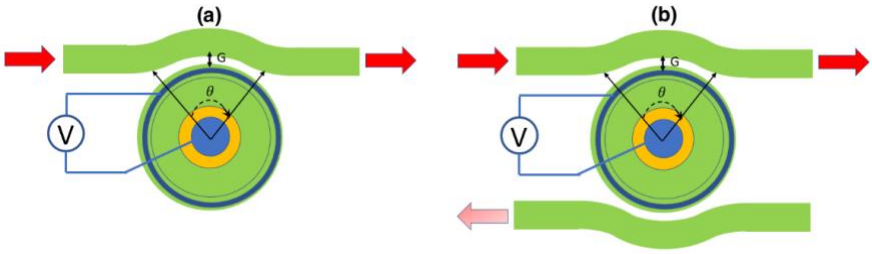


Figure 3.9 Top view of the EO microdisk standing alone for (a) EOM and (b) EO 2x2 "switch" for EO logic, WDM cross-connect switching networks, and ROADMs. Metal contact regions are shown in blue. The coupling angle and gap size between the disk and the bus waveguide are denoted by θ and G, respectively.

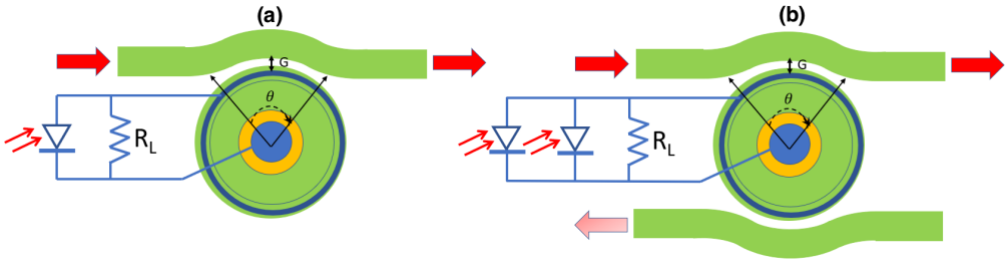


Figure 3.10 Top view of the photodetector-driven EO microdisk for OEO repeaters-with-gain and for OEO wavelength converters, and (b) the 2x2 device for multi-operational optical–optical logic gates. Each diode symbol represents a separate strip waveguide with its own waveguide PD at the strip end (two such waveguides for the OOL), as in Ref. [52].

The high bandwidth targeted in this paper for the proposed device structures serves three purposes: it provides an extremely high (i) "information bandwidth" for analog or digital modulation of light in an optical interconnect or communications link, (ii) bit/sec rate for digital modulation—a rate that is about twice the optical bandwidth, and (iii) switching rate, or rate of "reconfiguration" for 2x2 switching elements.

3.2.2 High-Bandwidth Microdisk Modulator

The simulations begin with an investigation of the device in Figure 3.9—a stand-alone EOM. Here, the U-junction depletion-mode microdisk modulator is a silicon rib-edge disk that is side coupled to one or two silicon strip bus waveguides. The rib-to-strip-bus side coupling is similar to that of a fully etched disk.

The design considerations of this device are made to maximize bandwidth for optical communications and computing applications. The standard, 220-nm-thick silicon core with a buried 3-μm silicon dioxide platform is used with an ultracompact disk radius of 2 μm to minimize the overall device footprint. The key feature of the microdisk modulator is the U-shaped pn junction, as shown in Figure 3.11, which is designed to maximize its depletion overlap with the WGM of the disk cavity. Vertically oriented pn junctions naturally overlap better with the wide aspect ratio of the SOI modes [55, 56, 63, 86]; the U-shaped junction is effectively a double vertical pn junction, with an n-p-n vertical sequence. A 50-nm-thick outer slab region (rib platform) is used to host the circular cathode contact to the U-shaped n region. The rib-shaped perimeter delivers good coupling to the bus waveguides—a coupling similar to that of fully etched disks. Previous demonstrations of microdisk modulators kept both electrodes within the WGM [55, 56], but this creates highly doped pn junctions that add to the total device capacitance (energy consumption) without contributing to the free-carrier plasma effect. This design only has two pn junctions: (i) the main, U-shaped junction that contributes to modulation and (ii) the small parasitic junction between the p+ region and the top of the n region. The depth of the p+ region is kept shorter than the depth of the silicon so as not to make a junction with the bottom n region. This is

a unique structure with rather complex fabrication requirements; however, a U-junction of similar feature sizes has been successfully demonstrated in other fabricated SOI devices [63].

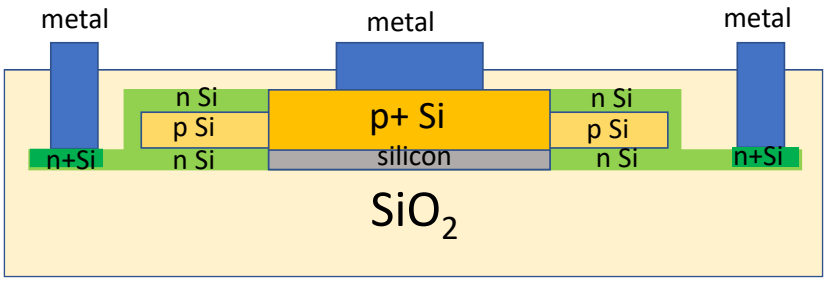


Figure 3.11 Cross-section view of the U-junction depletion-mode microdisk modulator. Note that the coupled waveguides are not shown in this view.

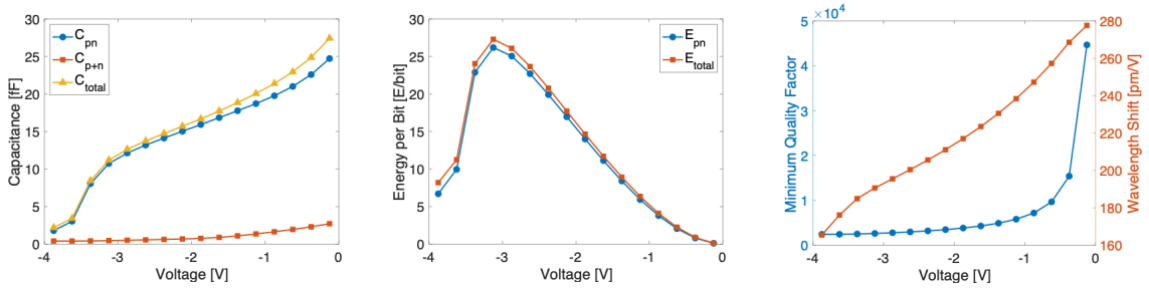
The microdisk resonator has several advantages over the microring resonator. The disk configuration is more efficient for hosting vertical pn junctions [55, 56, 77]. Furthermore, as detailed in Ref. [55], the plot of the internal quality factor as a function of radius for microdisk and microring resonators shows better performances for the microdisks as a result of the higher optical confinement. Moreover, the TE radial mode profiles indicate larger bending losses for microring resonators with radii around 2 μm. This disk is different from that in Ref. [55] as it has an outer rib platform; however, without the etched inner wall of a microring, the WGM of the disk still expects to experience less loss through less sidewall roughness. This is very important for highly compact devices such as this one. Furthermore, disks offer higher resistance to thermal and fabrication variations [62]. The higher resistance to thermal variations means less static energy consumption is required to keep the disk set to the desired wavelength.

The free-carrier plasma dispersion effect triggers the free carriers of the doped silicon regions to induce a change in the real and imaginary refractive indices of the silicon. By reverse biasing a pn junction, the depletion width is increased, and the spatial change in the free-carrier concentration produces a small but effective change in the refractive index. The concentration, orientation, and boundaries of the doped silicon regions are optimized to achieve the maximum change in effective refractive index Δn_{eff} of the WGM, while keeping the carrier-induced loss low. The boundaries of the highly doped, n+ and p+ regions are placed just far enough away from the WGM to minimize contact resistance without adding significant carrier-induced loss. The p region is placed in the vertical center of the disk because holes produce a greater change in refractive index than electrons, as indicated by the Soref–Bennett equations [87]. The commercial drift–diffusion solver from Lumerical (CHARGE) is used to calculate the charge profiles of the device for different reverse biases and doping parameters. These are exported to the mode solver from Lumerical (MODE) to calculate the effective refractive index and loss of the WGM for each bias condition.

For high-bandwidth operation, a microresonator-based modulator must operate with a low Q factor in order to minimize the photon lifetime delay within its cavity. Low-Q modulators, which have “wide” resonance transmission peaks, have in the past required increased reverse biases to maximize the modulation depths with inevitably high energy consumption. With the high $\Delta n_{eff}/V$ achieved by the present design, high-speed, low-Q resonators can be modulated at reasonably low reverse biases.

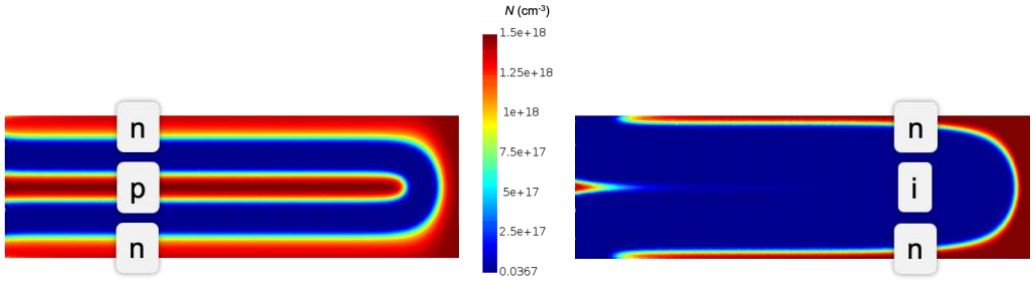
A parameter sweep of the peak doping concentration N and implant depth (range) R_p is performed to optimize the bandwidth within a reasonable voltage range of 0 to -4 V and with a realistic p-region straggle of ΔR_p = 40 nm (as determined by the simulations in Chapter 3.2.3). This device achieves its peak bandwidth with $N = N_A = N_D = 1.5\times10^{18}$ cm^{-3} and R_p = 110 nm, for a high Δn_{eff} of 3.5×10^{-4} V^{-1} (at a -1V bias). The critical performance metrics of the optimized U-junction microdisk modulator are plotted in Figure 3.12. Figure 3.12a plots the device capacitance as a function of reverse bias, which effectively demonstrates the behaviour of the U-junction. As the reverse bias is increased, the width of the depletion region is increased, which steadily decreases the capacitance. As the reverse bias is increased past -3.25 V, the capacitance drops off significantly. At this bias, the narrow middle p region becomes almost fully depleted, and the n-p-n vertical sequence effectively becomes an n-i-n. Operating in the fully depleted region, at -3.75 V, is beneficial for high-speed and low-power operation, as the U-junction capacitance C_{pn} is only 2.4 fF, and the total device capacitance C_{total} is 2.8 fF. Figure 3.12b uses the device capacitance and reverse voltage values to calculate $E/bit = CV^2/4$. This plot starts to follow the quadratic nature of the equation; however, it also drops off once the p region becomes depleted and capacitance is minimized. At the minimum capacitance point at -3.75 V, E/bit = 9.8 fJ/bit. The device can be operated in the low-bias region, for less than 5 fJ/bit of operation, but the minimum Q factor and maximum modulation speed are affected. Figure 3.12c presents the wavelength shift per volt produced at a given bias. The calculated wavelength shift is used to determine the minimum Q factor to design for, as any lower Q (i.e., with a wider resonance width) cannot be fully shifted by the given bias point. At the operating bias of -3.75 V, the device shifts 640 pm (170 pm/V), with an ultralow minimum

Q of 2,427. It is clear from these plots that operating in the fully depleted region permits simultaneous high speed and low power performance. It should be noted that, although the device performs better at high reverse biases, the efficiency drops off significantly past the full depletion point, as there is no more pn junction to modulate. The full depletion of the p region is demonstrated in Figure 3.13, which shows the free-carrier profile of the junction for 0 and -3.5 V biases.



(a) **(b)** **(c)**

Figure 3.12 (a) Capacitance, (b) energy consumption (in fJ/bit), and (c) resonance wavelength shift and minimum operating Q factor, with respect to reverse bias, for the stand-alone U-junction microdisk EOM devices presented in Figure 3.9.



(a) **(b)**

Figure 3.13 Free-carrier profiles of the U-junction for (a) 0 and (b) -3.5 V biases. For -3.5 V, the p region is almost fully depleted; however, a small concentration of holes still exists at the interface with the p+ region (not shown). The depletion region is shown in blue, and the free carriers (both p and n) are shown in red.

The EOM in Figure 3.9a couples a single bus waveguide to the microdisk, where a reverse bias is applied to the U-junction. A pulley-style coupler is used in all four of the proposed devices because the coupling condition (and the disk's Q factor) can be adjusted by simply changing the coupling angle (θ) around the disk. By increasing the degree of bending around the disk, the Q factor and photon lifetime of the resonator are lowered, and the potential bandwidth is increased. Figure 3.14 plots the maximum bandwidth of the device with respect to the applied reverse bias. The maximum bandwidth of the device is a function of the photon (optical) and RC (electrical) lifetimes of the microdisk and pn junction, respectively, which can be estimated by [52]

$$\Delta f_{3dB,opt} = \sqrt{\sqrt{2}-1}\frac{c_0 \delta\lambda}{\lambda_0^2} \tag{3.1}$$

$$\Delta f_{3dB,elec} = \frac{1}{2\pi RC} \tag{3.2}$$

$$\Delta f_{3dB,total} = \frac{\Delta f_{3dB,elec}\Delta f_{3dB,opt}}{\sqrt{\left(\Delta f_{3dB,elec}\right)^2 + \left(\Delta f_{3dB,opt}\right)^2}}, \tag{3.3}$$

where c_0 is the vacuum speed of light, λ_0 is the resonance wavelength, $\delta\lambda$ is the linewidth of the resonance, and R and C are the resistance and capacitance of the U-junction, respectively.

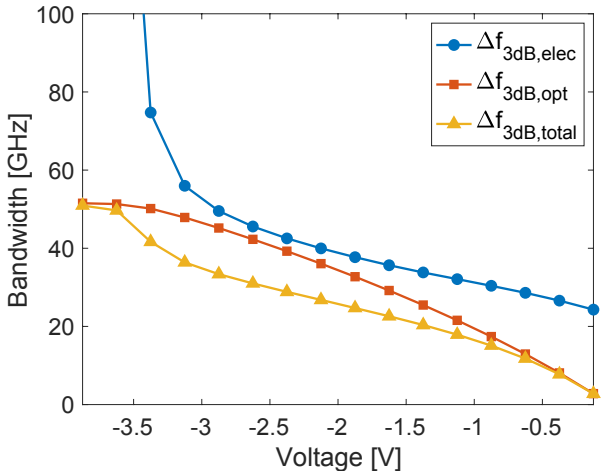


Figure 3.14 Optical, electrical, and total bandwidth, with respect to reverse bias, for the stand-alone U-junction microdisk EOM devices presented in Figure 3.9.

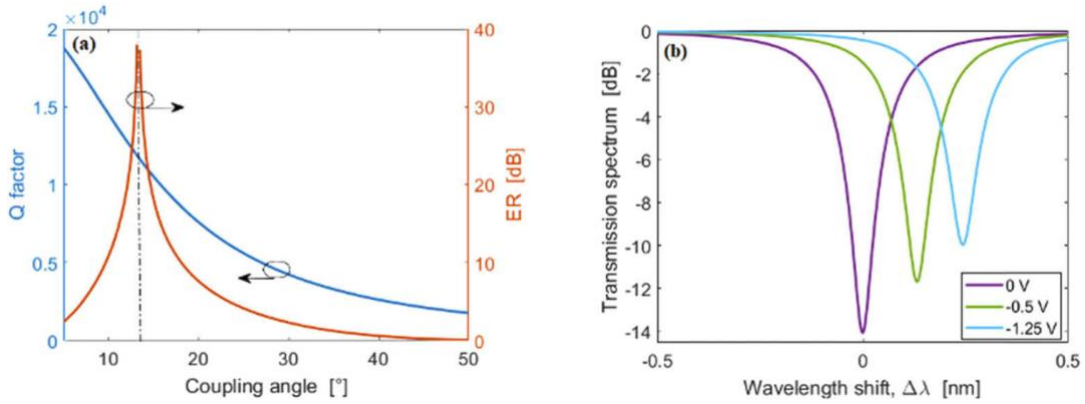
As the reverse bias increases, the pn junction induces a greater $\Delta\lambda$ in the resonance of the microdisk resonator, which enables the use of high-$\delta\lambda$, low-Q disks. At the same time, the depletion width expands, and the edges of the n region move slowly away from the center of the WGM, where the carrier dispersion effect peaks. This effect is evident through the slow saturation of the optical bandwidth until -3.75 V, where the middle p region is fully depleted, and the bandwidth curve flattens out completely. By increasing the straggle of the p region in the U-junction, the full bandwidth flattening would occur at a larger reverse bias, but the peak bandwidth would be lower, limited by the modulation efficiency with less junction–mode interaction.

Electrical bandwidth is a product of the pn junction capacitance and contact resistance in the electrodes. The resistance was calculated to be 265 Ω using the drift–diffusion solver from Lumerical. As the reverse bias is increased past the full junction depletion, capacitance is significantly reduced and the electrical bandwidth significantly increases;

thus, the total bandwidth is entirely limited by optical bandwidth, as illustrated in Figure 3.14. At -3.75 V, the EOM operates at 50.3 GHz and 9.8 fJ/bit.

The carrier-induced loss of the EOM is 27.6 and 10.8 dB/cm for 0 and -3.75 V, respectively. Because a maximum bandwidth is achieved at an ultralow Q factor of 2,427, this loss does not significantly affect the ultimate operation of the device. However, the pulley-coupled waveguide must still be optimized to achieve the desired Q factor with a maximum ER. The preliminary investigations, as shown in Figure 3.15a, indicate that the required Q induces an ER value of only ~3 dB. Note that in the optical calculations, the simulated carrier-induced loss and the standard, 2 dB/cm waveguide propagation loss are taken into account. The bending loss is considered to be negligible compared to these two. Because ER and Q both decrease with an increase in coupling angle, there is a trade-off between ER and bandwidth for this 1x1 EOM. This trade-off does not exist for low coupling angles, as ER increases with a reduced Q, but the Q values are too high for high-speed operation. To ensure high-performance operation, a requirement on all devices presented in this work is imposed: that they must offer an ER of at least 10 dB. To satisfy this condition, the first device must operate with a minimum Q of 8,382, which is achieved with a coupling angle of $\theta = 18.5°$, a gap size of $G = 100$ nm (as with each device presented here), and a waveguide width of 500 nm. At this Q, the bandwidth, bias, and energy consumption are all reduced to 13.4 GHz, -0.75 V, and 3.1 fJ/bit, respectively. Should a lower ER be acceptable, the coupling angle and applied bias can simply be increased to the desired higher bandwidth. The previous discussion is summarized in Figure 3.15a, where both the Q and the ER values are plotted as a function of the coupling angle, assuming N_A

= $N_D = 1.5\times10^{18}$ cm^{-3}, a bus waveguide width of 500 nm, and a reverse voltage of -0.75 V. Under the same conditions, the transmission spectrum as a function of the wavelength shift, $\Delta\lambda = \lambda - \lambda_0$, is shown in Figure 3.15b, for different values of the reverse bias and assuming a fixed coupling angle of 18.5°. The applied bias affects the critical coupling condition to the microdisks, reflecting the changes in the ER of the resonances.

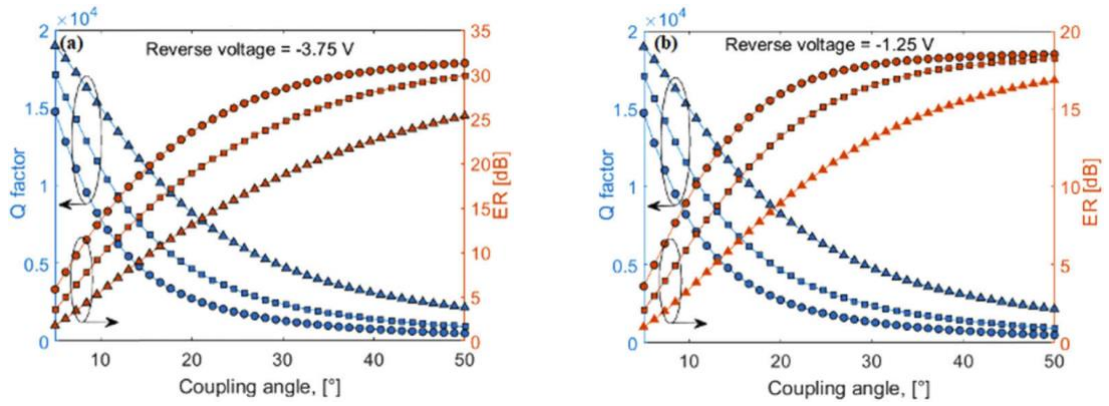


(a) **(b)**

Figure 3.15 (a) Q factor and ER at the Through port, as a function of the coupling angle, assuming a reverse voltage of -0.75 V; (b) Through-port transmission spectrum as a function of the resonance wavelength shift, for different values of the reverse bias. In the simulations, the waveguide width is 500 nm, $N_A = 1.5\times10^{18}$ cm^{-3}, $N_D = 1.5\times10^{18}$ cm^{-3}, and the coupling angle is 18.5°.

The 2x2 switch in Figure 3.9b adds a second pulley bus waveguide to the EOM for high-speed switching in WDM communications. Figure 3.12–Figure 3.14 hold true for this device, as the pn modulator itself remains the same. The addition of the second bus waveguide, however, changes the coupling condition, as the WGM now has two output accesses. Figure 3.16a and Figure 3.16b show the Q factor and the ER at the Through port, as a function of the coupling angle, for different values of the bus waveguide width W_{bus},

and assuming a reverse voltage of -3.75 and -1.25 V, respectively. The plots clearly indicate that for the greater coupling angle required for a low-Q operation, the ER now increases, which removes the trade-off with bandwidth; hence, the device can operate at its maximum bandwidth without dropping below the 10 dB ER target. With a waveguide width of 500 nm, and a coupling angle of 29.6°, the 2-bus ROADM and EOL element achieves the maximum, 50.3 GHz bandwidth with an ER of 24.4 dB. Operating in the low-bias region (e.g., -1.25 V) requires a larger Q factor of 5,310, which limits ER and bandwidth to 12.4 dB and 19.1 GHz, respectively, but with a reduced E/bit of 7.6 fJ/bit. For low-bias operation, the coupling angle and waveguide width were adjusted to 28.0° and 600 nm, respectively, to maximize ER and bandwidth.



(a) **(b)**

Figure 3.16 Q factor and ER at the Through port, as a function of the coupling angle, for different values of the bus waveguide width. In (a), the reverse bias is -3.75 V; and in (b), the reverse bias is -1.25 V. Circle: W_{bus} = 400 nm, Square: W_{bus} = 500 nm, Triangle: W_{bus} = 600 nm. The peak carrier concentrations of the doping regions are $N_A = N_D = 1.5\times10^{18}$ cm^{-3}.

The OEO repeater and wavelength converter presented in Figure 3.10a follows the same design as the EOM in Figure 3.9a, but with an added load resistor R_L connected across the pn junction, which is driven by the PDs photovoltage V_{ph}. V_{ph} is a function $\eta_{pd}P_{in}R_L$, where η_{pd} is the responsivity of the PD in photoconductive mode and P_{in} is the peak power of the PD's inputted optical signal. From Ref. [52], the responsivity of the GeSn PD is estimated to be 1.4 A/W, using a suitable optical input power of 0.5 mW. In these device architectures, the p-i-n PD is a p-GeSn/i-GeSn/n-GeSn homo-junction photodiode embedded in a ridge waveguide. In particular, the GeSn waveguide is a ridge structure with a height of H_{PD}, a width of W_{PD}, and a slab thickness of H_s. The ion implantation is performed at both lateral ends of the slab in order to fabricate the p and n regions, as well as the p++ and n++ contacts. Moreover, a Ge layer with a thickness d is necessary to facilitate the deposition of the GeSn alloy. As detailed in Ref. [52], the GeSn-on-Si PD presents a number of advantages compared to the standard Ge-on-Si PDs. In particular, (i) the GeSn-on-Si PD gives a better wavelength coverage due to its selectable bandgap; (ii) the 1,550-nm GeSn-on-Si PD has stronger absorbance (15,000 cm^{-1}), higher efficiency, and lower spatial volume; and (iii) hundreds or thousands of PDs could be closely integrated on-chip at a low cost. 3% of Sn content is used in order to optimize its absorption at 1,550 nm.

As with the EOM in Figure 3.9a, because this device uses one bus waveguide, it must operate with a high Q to achieve a suitably high ER. The added load resistor also significantly limits its electrical bandwidth. Here, the peak carrier concentration of the doping regions is reduced to $N_A = 0.8\times10^{18}$ cm^{-3} and $N_D = 1.5\times10^{18}$ cm^{-3} to deplete the p

region at a lower reverse voltage and to reduce the required load resistance. To produce the -0.75 V photovoltage required for the maximum bandwidth operation, the required load resistance is 1.1 kΩ. Given a fully depleted, low-C junction, the electrical bandwidth is larger than the optical bandwidth; thus, the total bandwidth is effectively limited by the Q factor (ER). Figure 3.17 presents the bandwidth plots with the added load resistor, where the total bandwidth maxes out at 11.1 GHz. It is clear that the overall bandwidth performance is dominated by the optical bandwidth. With a coupling angle and waveguide width of 15° and 500 nm, respectively, this device achieves the required Q of 11,100 with an ER of 17.0 dB and an E/bit of 0.67 fJ/bit. The Through-port transmission spectrum of this device is shown in Figure 3.18 for different reverse biases. Note that the smaller resonance shift between -0.5 and -0.875 V is due to the junction already being mostly depleted.

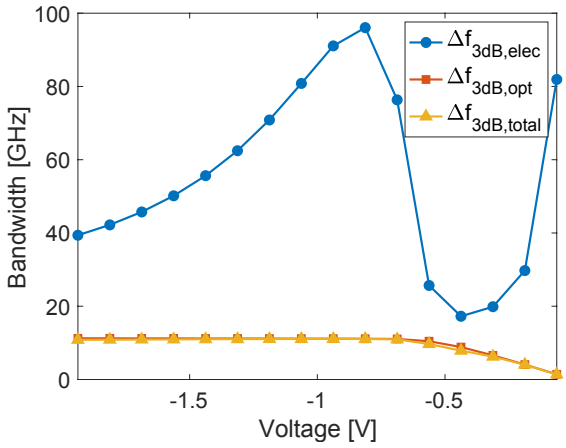


Figure 3.17 Optical, electrical, and total bandwidth, with respect to the applied reverse bias, for the PD-driven U-junction microdisk OEO and OOL devices presented in Figure 3.10. The sharp increase in electrical bandwidth is due to the sharp depletion of the middle p region; the decrease of electrical bandwidth past -1 V is due to R_L increasing with C already being at its minimum.

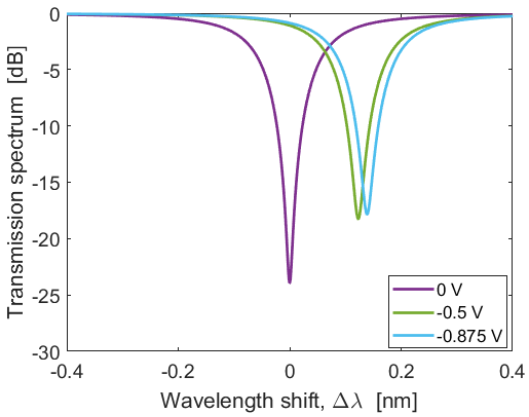


Figure 3.18 Through-port transmission spectra of the OEO device in Figure 3.10a, as a function of the resonance wavelength shift, for different values of the reverse bias, assuming a waveguide width of 500 nm, a coupling angle of 15°, $N_A = 0.8\times10^{18}$ cm^{-3}, and $N_D = 1.5\times10^{18}$ cm^{-3}.

The restorative, multi-operational optical–optical logic gate follows the same connected PD–EOM design as the OEO (as shown in Figure 3.10b); however, a second electrical input is added to perform 2-bit Boolean logic operations. The second electrical input is from an identical GeSn waveguide PD connected in parallel with the first PD. The total photovoltage is expressed by $V_{ph} = R_L I_1 + R_L I_2$, where I_1 and I_2 are the photocurrents generated by the two PDs. As demonstrated in Figure 7 of Ref. [52], the two electrical inputs are used to perform the OR, AND, or XOR operations by shifting the resonance wavelength by $\Delta\lambda$ (one input on) or $2\Delta\lambda$ (two inputs on), depending on the relative positions of the input and resonance wavelengths. With three input signals centered at λ_0, $\lambda_0 + \Delta\lambda$, and $\lambda_0 + 2\Delta\lambda$, this device performs the three fundamental logic operations simultaneously, for multiple instruction, single data (MISD) electro-optical computing. A second pulley coupler is added to the disk to capture the inverse of these operations as well, providing the complimentary NOR, NAND, and XNOR operations.

Adding the second bus waveguide to the OEO device configuration has the same effect the ROADM had, in that the ER increases with the coupling angle—eliminating the ER–bandwidth trade-off. Because of the negative impact the added load resistor has on the total bandwidth, the peak carrier concentration is reduced to $N = N_A = N_D = 0.8\times10^{18}$ cm^{-3} to increase bandwidth through a reduced capacitance. For this device, V_{ph} cannot be set to fully deplete the U-junction, as a second input would have no effect on modulation. Therefore, V_{ph} is set to half of the full depletion voltage. $1V_{ph}$ (one input on) is capable of shifting $1\Delta\lambda$ if the Q factor is doubled, and $2V_{ph}$ (two inputs on) will shift $2\Delta\lambda$. Because the U-junction is only partially depleted with $1V_{ph}$, the capacitance is higher, and a slightly reduced maximum bandwidth is expected compared to the OEO devices described. With a coupling angle of 20°, a waveguide width of 600 nm, and a V_{ph} of -0.75 V, the OOL device achieves 7.6 GHz of operation, 2.2 fJ/bit of energy consumption, and an ER of 12.5 dB.

3.2.3 Fabrication Feasibility: Characterization of the Ion Implantation Process

The ultimate implementation of the proposed junction design depends on the feasibility of its fabrication. Although the U-shaped design is uncommon, it has been experimentally demonstrated in other SOI devices with similar feature sizes [63]. These demonstrations indicate that the fabrication of a U-shaped junction is feasible. To further verify the feasibility of the optimized design, a TSUPREM-4 2D [72] process simulation is performed.

The doped regions of the U-junction are created by ion implantation. The device is modelled using implant parameters within the capabilities of commercially available ion implantation services. The n region is created by two separate implants. The upper n region of the 1x1 and 2x2 EOM devices can be formed by implanting phosphorus with an energy and dosage of 20 keV and 3.8×10^{13} cm^{-2}, respectively, and the lower n region can be formed by implanting phosphorus with an energy and dosage of 120 keV and 0.7×10^{13} cm^{-2}, respectively. The middle p region of the U-junction can be formed by implanting boron with an energy and dosage of 20 keV and 3.8×10^{13} cm^{-2}, respectively. The doses can be increased or decreased to meet the required concentrations of the other (OEO) devices, but the process remains the same. The p implant is the most demanding process step, as the straggle of the implant is low. After a thermal activation step of 1,000 °C for 5 s, the straggle of the p region ΔR_p is found to be 40 nm, which is used in the CHARGE optimizations. The n+ and p+ regions use a similar implant process, but with higher doses. The unbiased doping profile is shown in Figure 3.19, where the peak concentrations and junction spacing closely match those of the CHARGE optimizations.

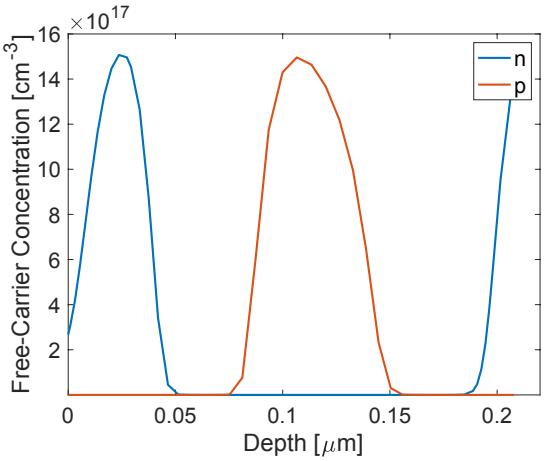


Figure 3.19 Simulated vertical free-carrier concentration of the npn-sequenced U-junction modulator for a 0 V bias.

The passive component of the device: a disk with a 2 μm radius and a 100 nm waveguide-to-disk gap size, is readily fabricable by modern e-beam or deep UV lithography tools. The widths of the doping regions are no smaller than 2.4 μm (p+ region) and can be masked by the same lithography tools. The 50-nm spacing between the middle p region and the edge of the disk may cause a disconnection between the upper and lower n regions if there are significant mask misalignments. This gap can be widened to increase tolerance to misalignments; however, the junction–mode overlap (modulation efficiency) will consequently decrease.

To demonstrate the fabrication tolerance of the U-junction EOM design, the bandwidth performance with respect to the range and straggle of the p region, the peak doping concentration, and the separation between the p region and the lateral edge of the disk (caused by variations in the size and lateral positioning of the p implant) are plotted in Figure 3.20. With a 10% increase in the separation between the p region and the edge of

the disk, the length of the pn junction is reduced, which reduces the modulation efficiency and maximum bandwidth by only 0.8%. For a 10% decrease in this separation, the length of the pn junction is increased, resulting in a 0.6% increase in the bandwidth. The efficiency will continue to increase (minimally) with the length of the pn junction until the separation is reduced to zero, where the two n regions become electrically isolated and the efficiency will be effectively halved. Extending or reducing the range of the p region by 10% causes one of the depletion zones to move further from the center of the WGM; hence, reducing the modulation efficiency. However, this only reduces the bandwidth by 2.6%, as the other depletion zone moves closer to the center of the WGM. By expanding or condensing the straggle of the p region by 10%, the U-junction will become fully depleted at a different reverse bias. Without adjusting the operating reverse bias, the bandwidth is reduced by as much as 16.1% for a larger p region straggle, and 13.1% for a smaller p region straggle. By adjusting the operating voltage to the new performance peak, the reduction is not as significant. Note that with a 15% increase in straggle, the outer n regions become very narrow; this reduces capacitance and increases electrical bandwidth. For 10% variations in peak doping concentration, bandwidth reduces by as much as 12.3%, but this reduction can also be less significant if the operating reverse bias is adjusted to match where the U-junction is fully depleted. The fabricated devices will perform reasonably close to the optimal performance by providing appropriate tuning to the operating voltages.

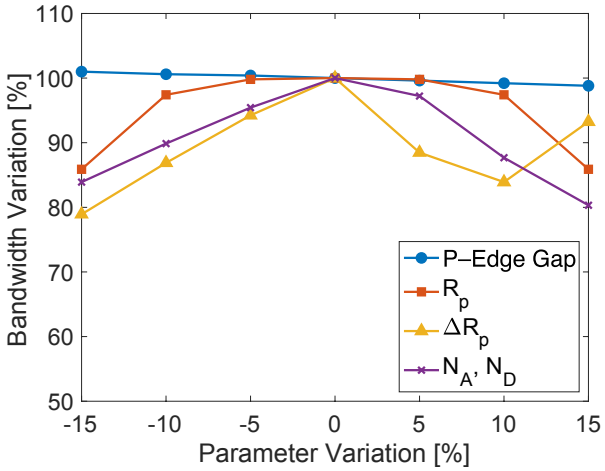


Figure 3.20 Change in EOM bandwidth performance with respect to fabrication variations in separation between the p region and the lateral edge of the disk, range of the p implant R_p, straggle of the p implant ΔRp, and peak doping concentration N.

3.2.4 Comparison of the Proposed Optoelectronic Devices

For all four of the devices depicted in Figure 3.9 and Figure 3.10, the simulation results are shown in Table 3.2. The 2x2 EOM, because of its low device resistance and low-Q operation, achieves the largest bandwidth among the four devices, at 50.3 GHz. Although it uses a relatively large reverse bias of -3.75 V, the low device capacitance keeps the energy consumption low. The 2x2 OOL, which adds a large PD-driven load resistor to the 2x2 EOM, uses a lower peak doping to reduce the RC time delay through less capacitance. The lower doping concentration reduces its modulation efficiency, increases the minimum Q factor, and limits the total bandwidth. The 1x1 devices (stand-alone EOM and OEO repeater and wavelength converter) must operate at high Q factors to achieve suitably high ERs, which significantly limits their optical bandwidths. For high-Q operation, these devices naturally operate best at low voltages. The peak concentration is reduced for the doping regions to deplete the U-junction at these lower voltages. These devices operate at

a relatively lower bandwidth with a low energy consumption. Each device is designed to achieve a minimum ER of 10 dB. Should a lower ER be acceptable, the Q factor can be reduced to increase the bandwidth.

Table 3.2 Comparison of the Four Proposed U-Junction Modulator Devices

Device	Bandwidth (GHz)	Bias (V)	E/bit (fJ/bit)	Q Factor	ER (dB)	Coupling Angle (°)	W_{bus} (nm)
1x1 OEO	11.1	-0.875	0.7	11,100	17.0	15.0	500
2x2 OOL	7.60	-0.75	2.2	10,019	12.5	20.0	600
1x1 EOM	13.4	-0.75	3.1	8,382	10.0	18.5	500
2x2 EOM	50.3	-3.75	9.8	2,427	24.4	29.6	500
2x2 EOM (Low V)	19.1	-1.25	7.6	5,310	12.4	28.0	600

It should be highlighted that the 1x1 OEO has a reasonably high bandwidth with an ultralow energy consumption. The low drive voltage will enable arrays of these repeaters to be deployed in photonic integrated circuits at which the optical signal levels are in the low mW range. The 2x2 EOM is particularly promising for high-radix matrix optical network switching with impressive bandwidths at both voltage levels. The 1x1 EOM is expected to advance the art of EO logic with its excellent combination of performance metrics.

In comparison to the previous work on OEO and OOL devices [52], which featured a vertical pn junction design, this work achieves a large energy efficiency improvement with a 2X bandwidth, while maintaining similar dimensions and ER performance, as shown in Table 3.3. A comparison to the state-of-the-art high-speed EOM devices is not straightforward, as this design direction—to reduce Q and increase bandwidth through high-efficiency pn junction designs—highlights the limitations of using 1x1 EOM

configurations, where operating at low Q factors coincides with a low ER. In this work, a minimum acceptable ER of 10 dB is chosen, which limits the bandwidth of the 1x1 EOM to 13 GHz. The state-of-the-art devices operate up to 50 GHz, but with ERs less than 10 dB [17, 88, 89, 90]. We have shown that our device can also operate at 50 GHz, but at the cost of an ER below 10 dB. While the performance is comparable, the modulation efficiencies of the devices reported in the literature are generally between 20 and 50 pm/V, making their energy consumption and device size to be larger than this. Devices that are also in the low fJ/bit energy consumption range are generally much slower, operating below 20 GHz and 10 dB ER [55, 56, 64, 65]. The improved 2x2 EOM configuration highlights the potential performance of this design, with an ER above 20 dB at the maximum bandwidth; however, a direct comparison is difficult as the prior art is lacking or deficient in this area. By simply comparing this 2x2 to the state-of-the-art 1x1 devices, however, comparable or better performance in bandwidth, energy efficiency, ER, and footprint is evident.

Table 3.3 Comparison to the State-of-the-Art High-Speed EOM, OOL, and OEO Devices

Device	Bandwidth (GHz)	E/bit (fJ/bit)	ER (dB)	Radius (μm)	Δλ/V (pm/V)
OEO (1x1)					
This Work	11.1	0.7	17.0	2	250
Previous Work [52]	6.5	5.8	16.6	2	62
OOL (2x2)					
This Work	7.6	2.2	12.5	2	250
Previous Work [52]	4.6	9.6	12.3	2	29
EOM					
This Work 1x1	13.4	3.1	10.0	2	250
This Work 1x1	50.3	9.8	3.0	2	170
This Work 2x2	50.3	9.8	24.4	2	170
1x1 [17]	56 (Gb/s)	70	6.0	10	32[a]
1x1 [88]	47	—	6.1	5	45
1x1 [89]	46	710	4.7	—	40
1x1 [90]	80 (Gb/s)	71[b]	—	8	22[a]

3.3 Summary

This chapter presents a comprehensive numerical comparison of four optoelectronic silicon microring/disk modulators. The key optimization in the design is the orientation and placement of the pn junction and how it overlaps with the resonant optical mode. A simple, vertically oriented pn junction achieves a 1.5X larger change in refractive index than a conventional, lateral pn junction; however, its larger capacitance only moderately reduces energy consumption. By adding a second pn junction to the device, and by combining vertical and lateral depletion components, an improvement in energy efficiency by more than 4X is demonstrated. With the same optimization, the carrier-induced optical loss is

[a] Value inferred from transmission spectrum.
[b] Value calculated from capacitance and voltage values using $CV^2/4$.

reduced by over 2X, allowing the device to operate with less optical power—further reducing the total energy consumption of the chip. A 4X increase in the energy efficiency of silicon-photonic modulator designs, like these, can make substantial reductions in the environmental cost taken by the growing data communications industry.

This chapter also presents high-efficiency depletion-mode U-junction microdisk modulators for ultrahigh-bandwidth operation in EO modulators, EO logic gates, high-radix EO matrix switches for WDM networks, ROADMs, OEO repeaters, OEO wavelength converters, and optical–optical multi-operational logic gates. The U-shaped junction produces a larger change in refractive index and resonance wavelength than conventional, lateral pn junctions, which allows for the use of low-Q, low photon lifetime resonators for high-speed operation of up to 50 GHz under suitably low voltages. The n-p-n sequence of the U-junction becomes fully depleted above the required operating voltage, where the capacitance is reduced below 3 fF. This behavior allows for a simultaneous high-speed and low-power operation that cannot be demonstrated in conventional, linear single-junction modulators. By using CMOS-compatible materials and processes for the devices presented in this work, low-cost, high-volume fabrication feasibility of the proposed designs is ensured. Given the impressive performance improvements over the existing devices reported in the literature, the proposed designs offer a promising potential in next-generation electro-optical communications and computing circuits.

Chapter 4: Design of Next-Generation Integrated Optical Communications Circuits

Energy consumption has become the main limiter to advancements in data processing and communications [13, 16]. Faster, more-functional circuits cannot be produced at scale if they cannot be powered economically. Improvements in efficiency can be made to computing architectures and CMOS transistors, but by and large, the main contributor to energy consumption is the metal interconnect network that transfers data within and between the processing chips. Data interconnects that span across the country or between racks in a data center have already been replaced by optical fiber, simultaneously improving the bandwidth, cost, and energy consumption of communication [91]. The next goal is to bring the performance improvements of optical data transfer to the chip itself [92, 93, 94, 95]. Energy consumption is expected to plummet simply by eliminating the charging of long metal interconnects and the abundant circuitry required to drive them (combined, they contribute to most of the chip's energy consumption [13]). Optical data transfer increases bandwidth simply by using photons to carry the signal, rather than electrons. But optics also allows for high bandwidth density, with relatively low crosstalk between neighboring signals. Perhaps most significantly, though, optical waveguides also increase bandwidth density through WDM and MDM. Although the former is a universally applied approach to combining signals in photonics, the latter has seen fewer applications because of the relatively high difficulty of selecting modes with low loss and low intermodal crosstalk. It has been shown, however, that combining WDM and MDM can increase bandwidth density by a factor of at least three, using simple and low-loss silicon photonic circuitry [96, 97, 98]. Silicon photonics provides the optimal platform for MDM

applications, as the high refractive index contrast between the core and cladding of the waveguide allows for more modes to propagate simultaneously for a fixed waveguide size. For the combined MDM–WDM architecture to be considered in high-volume optoelectronic circuits, however, improvements must still be made to its size, with a more compact design; bandwidth, with more channel spacing; and energy consumption, with attojoule-level optoelectronic modulation and detection.

This chapter introduces the design of novel high-bandwidth on-chip optical data communications circuits. These circuits combine wavelength- and mode-division multiplexing in high-efficiency microdisk resonators to send hundreds of channels down a single 500-nm-wide silicon waveguide. The design and fabrication process are detailed in this chapter and the experimental results validate the operation of the design.

This work was documented and presented at SPIE Photonics West 2020, with a submission for publication in Optics Letters.

4.1 Current State of High-Bandwidth Integrated Optical Communications Circuits

MDM multiplexers (MUX), which can be based on a variety of photonic devices, have been experimentally demonstrated in the literature [99, 100, 101, 102, 103, 104]. Only those based on the microring resonator have been shown to support both MDM and WDM—in addition to being compact and easy to design and fabricate. Although the microresonator-based circuits are WDM-compatible, they can be operated in a strictly MDM nature for applications that only have one laser source. In the proposed circuit, a

microresonator in a series of microresonators couples a specific mode and wavelength onto a multimode waveguide, as shown in Figure 4.1. The data-carrying signal can come externally or can be generated by modulating the microresonators of the MUX individually. For added WDM functionality, slightly detuned microresonators would be added to handle separate wavelength channels. For example, a three-mode, three-wavelength MUX would require nine microresonators to operate. After propagating across the chip on the multimode waveguide, the signals are demultiplexed (DEMUX) using a mirrored MUX circuit.

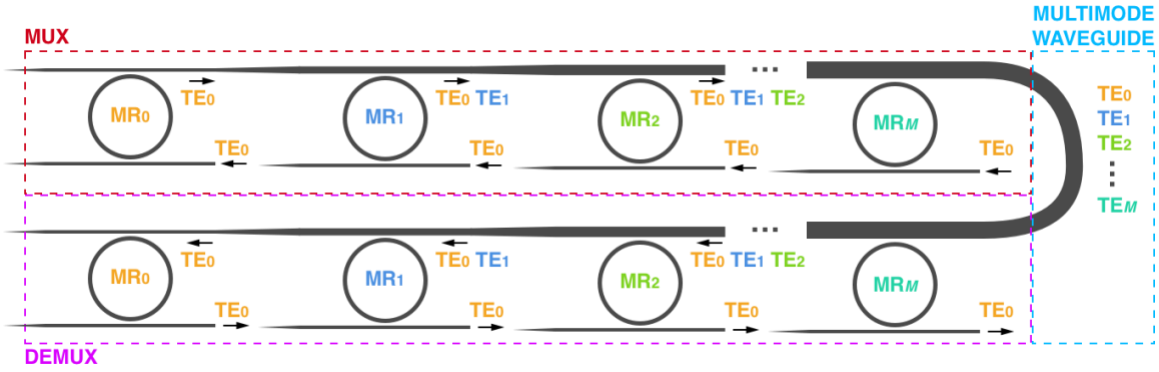


Figure 4.1 Visualization of the WDM-compatible MDM MUX/DEMUX circuit and the flow of data within it. Regular, point-coupled microring resonators are shown here, but they can be replaced by the other microresonator devices presented in this work. The multimode waveguide between the MUX and DEMUX is kept short for the purpose of visualization only.

Given that many microresonators are required for this operation, it is important to minimize their size. Of further importance, each microresonator-based modulator must satisfy the push for attojoule-level energy efficiency to be considered in next-generation optoelectronic communications circuits [13]. Perhaps the most promising optical modulation mechanism in silicon is the free-carrier-depletion vertical pn junction

microdisk modulator [55, 56, 105], as demonstrated in the previous chapter. The vertical pn junction produces a larger change in resonance wavelength, which has been demonstrated to reduce dynamic energy consumption by over an order of magnitude. The structure also eliminates the need for an electrode-supporting slab region, which further reduces the spatial footprint through tighter mode confinement and tighter integration of the contact electrodes. In this work, a variety of microresonator designs are explored for this application, with the microdisk resonator ultimately providing the best balance of performance.

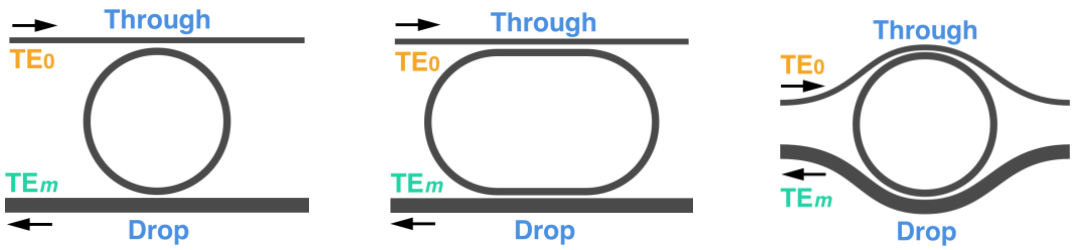
4.2 Design of Ultracompact, High-Bandwidth Optical Communications Circuits

The ultimate performance of the WDM–MDM MUX/DEMUX circuit depends heavily on the design of the microresonators that drive it, where a few performance tradeoffs must be considered. A comprehensive numerical characterization and comparison of three microring resonator types: point-coupled, racetrack, and waveguide-wrapped (which is novel to this application) are performed. Ultimately, the best-performing design will be extended to the microdisk resonator, where the aforementioned performance improvements of size and low-power modulation will be combined. For reasons that will become clear in this section, the microring resonator is much easier to simulate and optimize than the microdisk. Each microring design features a different coupling region which ultimately determines how well it transfers select modes between the multimode bus waveguide and the microring cavity. The IL and crosstalk (XT) achieved by the microring determines how many modes can be supported by the MUX/DEMUX circuit. The racetrack microring features a lengthened cavity section to increase the coupling strength, with the expense of

a larger spatial footprint and a smaller free spectral range (FSR). The waveguide-wrapped microring increases the length of the waveguide coupling section instead, which leads to better coupling strength without sacrificing space and FSR (bandwidth). Of course, an increase in bandwidth cannot be made without giving consideration to energy efficiency. The waveguide-wrapped design offers more flexibility in the types of pn modulators it supports, as the microring can be easily swapped with a microdisk, which naturally supports ultra-efficient vertical pn junction modulators.

4.2.1 Comparison of Microresonators for WDM–MDM Communications Circuits

The three WDM–MDM microring resonators compared in this work are shown in Figure 4.2. To establish a baseline level of performance comparison, the devices are designed on the same SOI platform. The silicon layer is 0.22 μm thick and is hosted between thick upper and lower silicon dioxide cladding layers. The radii of the microrings are all 5 μm, and the cavity widths of the microrings are all 0.44 μm (determined by matching the point-coupled microring to a 0.4 μm single-mode waveguide). To characterize and compare the optimal performance of the devices, the coupling interaction between the microring cavities and the drop waveguides of the first four TE modes are optimized. The coupling interaction is optimized based on the coupling length, the width of the drop waveguide, and the gap between the cavity and the drop waveguide. All other parameters remain constant.



(a) (b) (c)

Figure 4.2 The (a) point-coupled, (b) racetrack, and (c) waveguide-wrapped microring devices. The thick drop waveguide is the segment that is connected to the multimode bus waveguide and will vary in thickness depending on the TE$_m$ mode it handles. The thinner, through waveguide is an input/output port to/from the MUX/DEMUX circuit.

The most conventional device is the add–drop microring resonator with straight waveguides and a point coupler; it is shown in Figure 4.2a. To achieve optimal coupling of each mode, the width of the drop waveguide and its distance from the microring cavity are optimized. Given the short coupling length of this device, especially for a small ring radius, its performance is expected to be relatively poor. The racetrack microring resonator, shown in Figure 4.2b, improves on the previous device by extending the coupling length. This extension increases the coupling strength, with the expense of a larger microring cavity. In addition to increasing the total device footprint, increasing the size of the cavity reduces its FSR, which reduces the number of available wavelength channels (bandwidth). For this device, the gap size, width of the drop waveguide, and the length of the racetrack extension are optimized. The waveguide-wrapped microring, shown in Figure 4.2c, increases the coupling length without increasing the length of the microring cavity. For this device, the gap size, width of the drop waveguide, and the bend degree of the drop waveguide (which determines the coupling length) are optimized.

The commercial eigenmode solver from Lumerical (MODE) is used to calculate the effective refractive indices of the microring cavities and the different segments of the multimode bus waveguide that couple to it. The calculated effective refractive indices are used to match higher-order modes to the fundamental mode inside the microring cavity and find the width of the waveguide needed to support it. This index-matching process is shown in Figure 4.3, where the effective indices of the first seven TE modes are plotted with respect to waveguide width. The widths where the curves align with the constant index of the microring cavity are taken and set in the design. The optimal gap sizes between the waveguide and the microring cavity are found with the commercial finite-difference time domain solver from Lumerical. Because FDTD simulations require all electromagnetic fields to decay out of the simulation window to produce an accurate simulation [106], light-confining structures like the microring resonator are notoriously long to simulate. An equally accurate, but much quicker, simulation method is used for the optimization, where the S-parameters are calculated from half of a microring resonator, as shown in Figure 4.4. S-parameters are calculated for each device, for each mode order, and each gap size within a predefined sweep range to create compact models for use in the commercial system-level simulation tool from Lumerical (INTERCONNECT), which then calculates the transmission spectra of each device. The gap size of each device is optimized to obtain a resonance linewidth of 0.4 nm (50 GHz), which allows for dense WDM (DWDM) channel spacing.

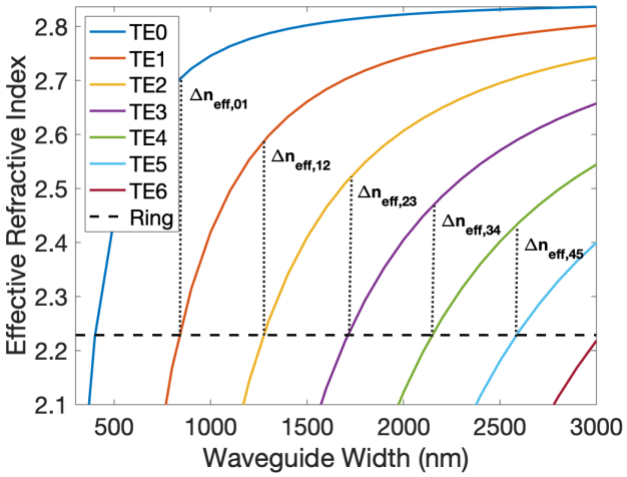


Figure 4.3 Simulated effective refractive indices of the first seven TE modes with respect to waveguide width. Dashed horizontal line shows the effective refractive index of the microring resonator, which has a constant width. All simulations taken at λ = 1.55 μm.

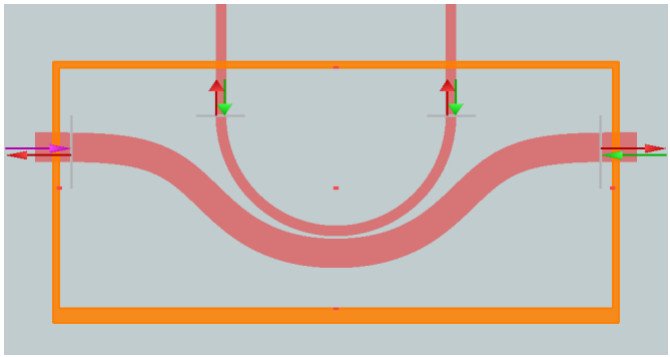


Figure 4.4 FDTD simulation window for the half-ring design and optimization method. For each possible source input, the S-parameters are extracted from each output. However, due to the symmetry of the device, only two inputs need to be considered—further reducing the simulation time.

The optimized device design parameters are listed in Table 4.1. For the racetrack and waveguide-wrapped microrings, the length of the coupling is also optimized. For the racetrack microring, the length of the TE_2 device is optimized to 5 μm to achieve the lowest possible IL for the worst-performing device. This length is applied to the rest of the

racetrack devices to maintain the same FSR across the board. For the waveguide-wrapped microring, the degree of the waveguide bend is increased to reduce IL and XT for each device. Because the circumference of the microring stays constant, the coupling length does not affect the FSR and can be different for each mode order.

Table 4.1 Optimized design parameters for the three WDM–MDM microring resonators.

Device	Mode	Waveguide Width (μm)	Gap Size (μm)
Point-Coupled	TE_0	0.40	0.16
	TE_1	0.84	0.12
	TE_2	1.27	0.11
	TE_3	1.71	0.10
Racetrack	TE_0	0.44	0.26
	TE_1	0.91	0.22
	TE_2	1.39	0.20
	TE_3	1.86	0.19
Waveguide-Wrapped	TE_0	0.40	0.10
	TE_1	0.84	0.10
	TE_2	1.22	0.10
	TE_3	1.64	0.10

The critical performance metrics are optimized for each WDM–MDM microring device and are listed in Table 4.2. These metrics are extracted from the simulated transmission spectra, which are shown in Figure 4.5–Figure 4.7. In terms of IL and XT, the regular, point-coupled microring performs the worst, as predicted, with the racetrack microring performing the best. However, because the racetrack microring has an extended circumference, its FSR is reduced, which limits the number of supported wavelength channels and potential bandwidth. The potential bandwidth is a product of the number of

supported wavelength channels, the number of supported modes, and the modulation speed of the microring device. The linewidths of the microrings' resonance peaks play a large role in this product. A narrow linewidth allows for more wavelength channels within the given FSR; however, IL and XT may suffer for higher-order modes as they may now be under coupled. A narrow linewidth also reduces the voltage (energy) required to modulate the microrings, but with the expense of a slower operating speed (long photon lifetime). Each of the 12 devices are normalized to a reasonable linewidth of 0.4 nm (50 GHz). Under this condition, each resonator type clearly supports at least four TE modes by the presented figures. With a reasonable operating speed of 10 Gb/s, the point-coupled and waveguide-wrapped microrings have the same potential bandwidth (due to having the same cavity circumference), which is 0.44 Tb/s more than the racetrack device (1.40 Tb/s). For higher mode counts, only the waveguide-wrapped and racetrack microrings will be able to scale in bandwidth, with the waveguide-wrapped microring also having the spectral advantage over the racetrack microring. For the greatest overall balance of performance, the waveguide-wrapped device is the most suitable for WDM–MDM applications.

Table 4.2 Optimized performance metrics for the three WDM–MDM microring resonators.

Device	Mode	IL (dB)	XT (dB)	FSR (nm)	Number of Channels	Bandwidth (Tb/s)
Point-Coupled	TE_0	0.39	N/A	18.65	46	1.84
	TE_1	0.54	-19.6			
	TE_2	1.26	-12.4			
	TE_3	3.10	-8.69			
Racetrack	TE_0	0.49	N/A	14.20	35	1.40
	TE_1	0.50	-33.8			
	TE_2	0.68	-25.4			
	TE_3	1.64	-36.7			
Waveguide-Wrapped	TE_0	0.74	N/A	18.65	46	1.84
	TE_1	0.62	-16.6			
	TE_2	0.66	-19.6			
	TE_3	1.74	-10.7			

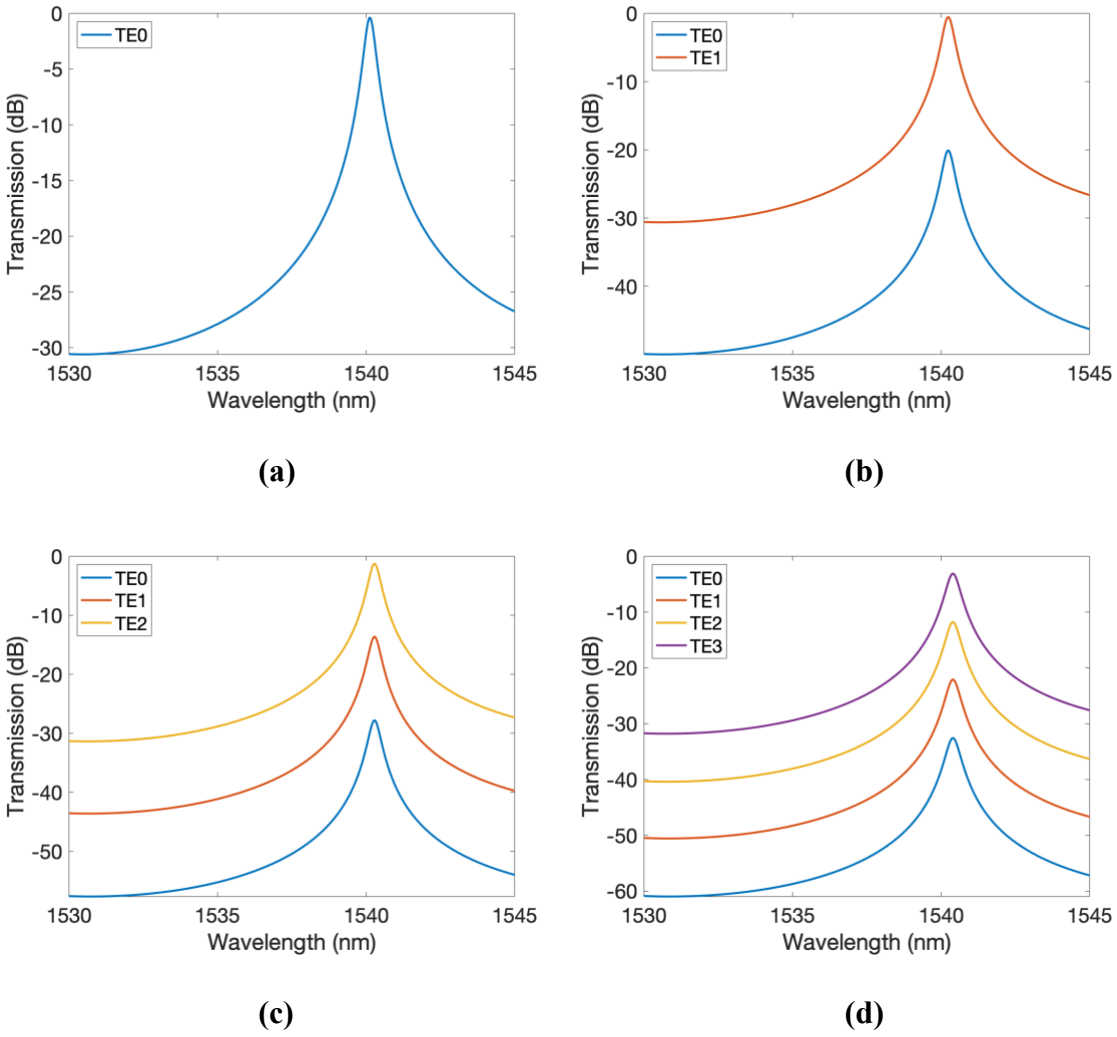


Figure 4.5 Drop-port transmission spectra for the (a) TE_0, (b) TE_1, (c) TE_2, and (d) TE_3 point-coupled microring WDM–MDM devices.

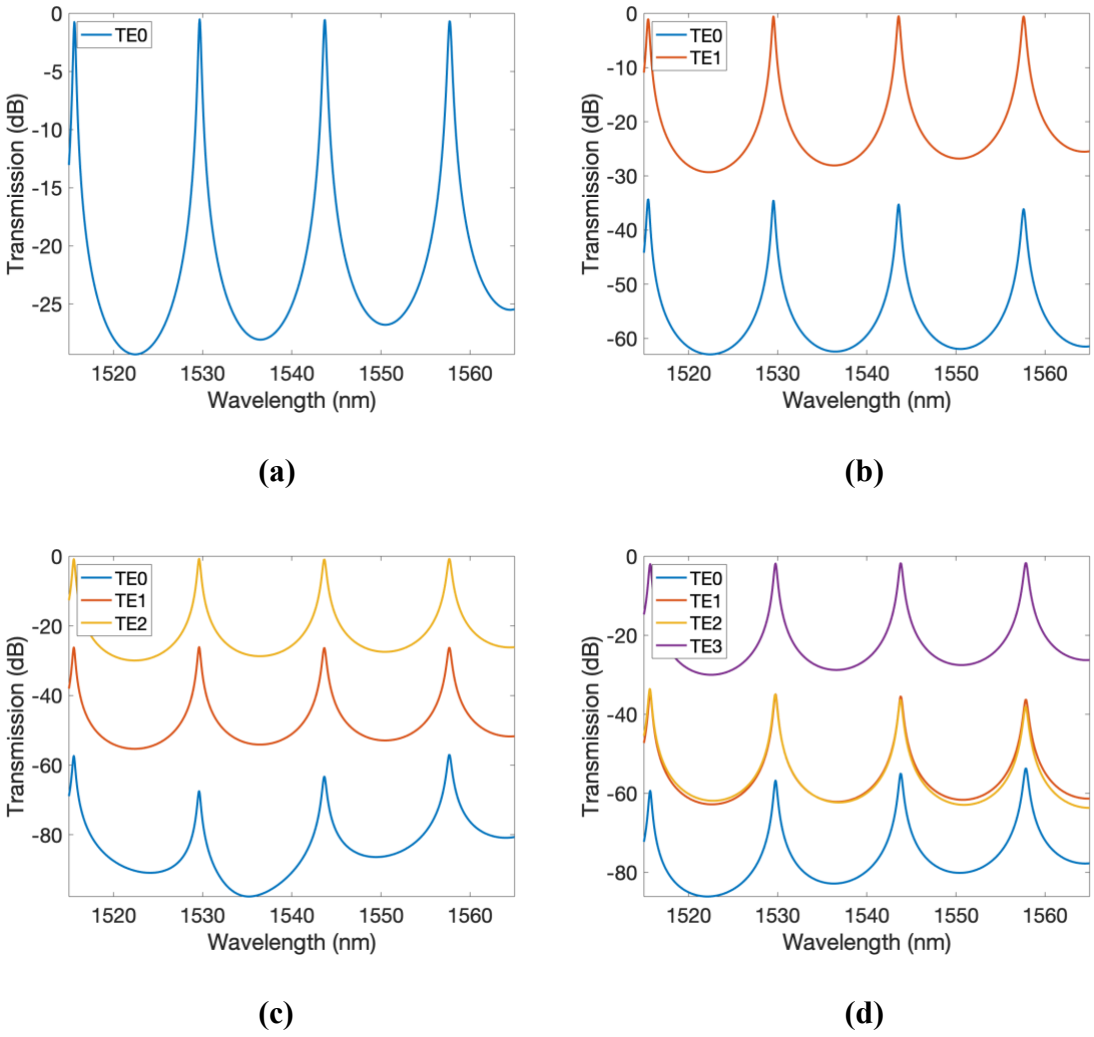


Figure 4.6 Drop-port transmission spectra for the (a) TE_0, (b) TE_1, (c) TE_2, and (d) TE_3 racetrack microring WDM–MDM devices.

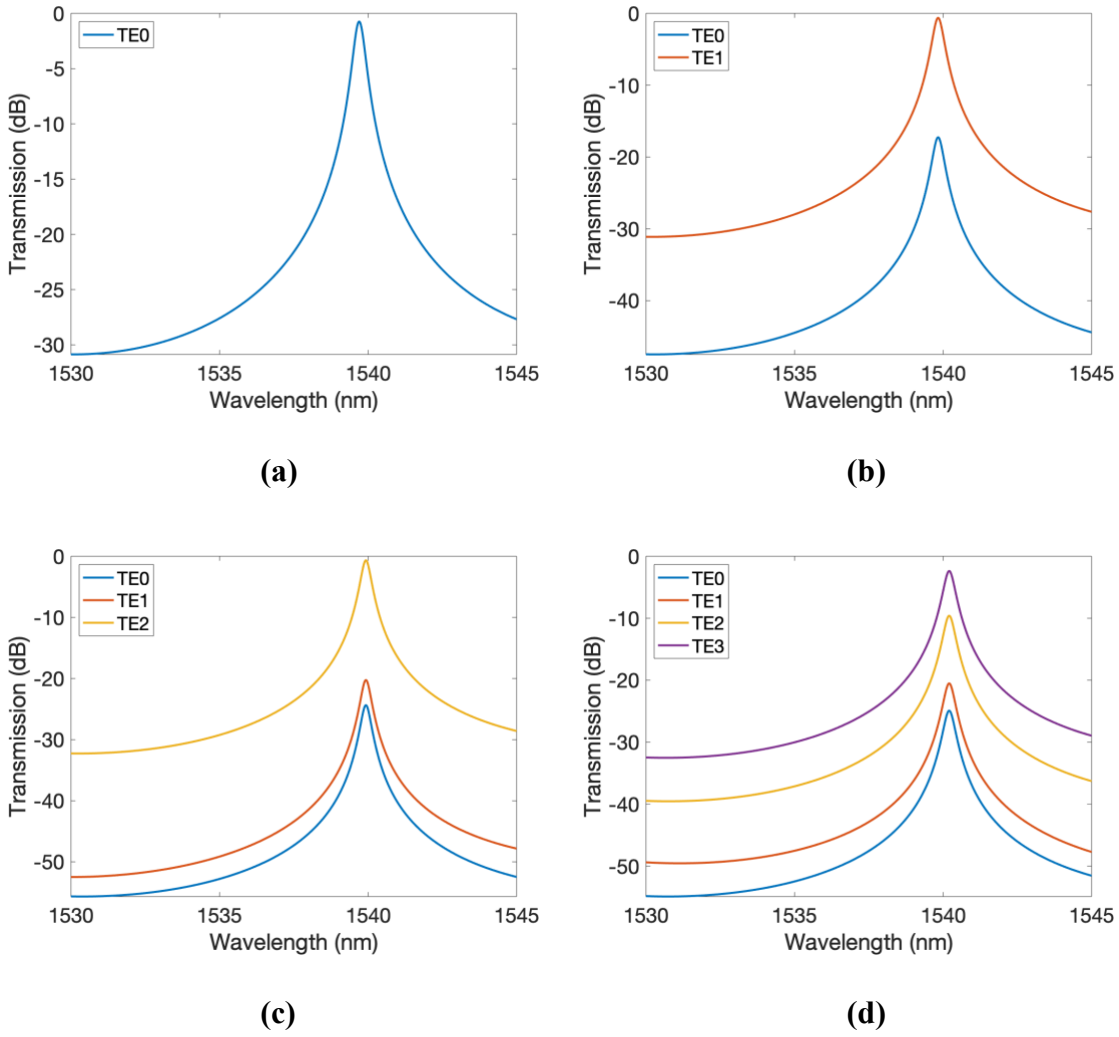
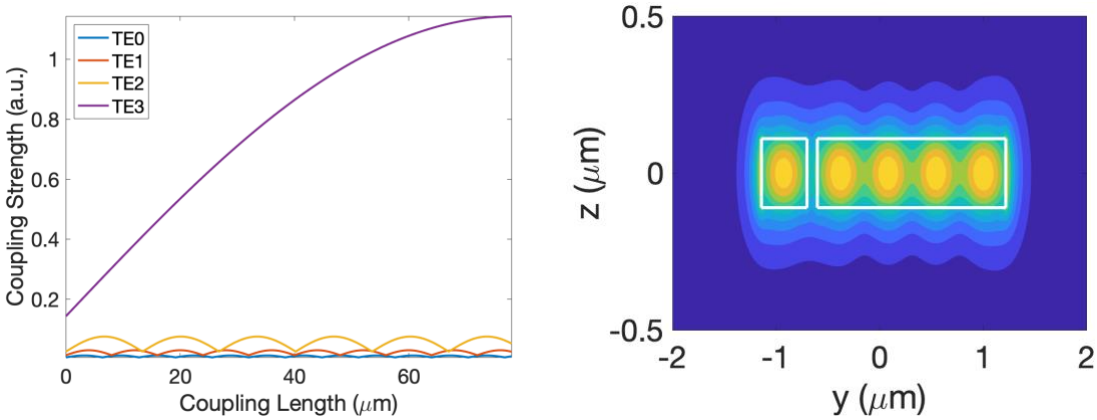


Figure 4.7 Drop-port transmission spectra for the (a) TE_0, (b) TE_1, (c) TE_2, and (d) TE_3 waveguide-wrapped microring WDM–MDM devices.

Figure 4.3 shows how the effective refractive indices of different TE mode orders scale with waveguide width and how each device can potentially support more than four mode orders. This is evident by the number of modes (waveguides) that can match the effective index of the microring cavity. However, the coupling strength diminishes for higher-order modes, which ultimately determines the IL, XT, and if the device can realistically support the mode. Figure 4.8a plots the coupling strength, with respect to coupling length, between

(for the example of Figure 4.8b) a TE_0-supporting waveguide and a TE_3-supporting waveguide. For increased coupling lengths, the coupling strength of the TE_0–TE_3 supermode follows a $\sin^2$ relationship. The length needed for maximum coupling strength would be too long to maintain an FSR suitable for DWDM operation, but a microring with even a low coupling strength can achieve critical coupling. Because the TE_3 waveguide does not match its lower-order modes to the microring cavity, their maximum strengths reach much lower values, which is a key component for achieving low XT. Ideally, a length is chosen where the undesired modes have the lowest coupling, and the desired mode has the strongest coupling—without using excessively long coupling sections.



(a) **(b)**

Figure 4.8 The (a) simulated coupling strength between a TE_0-supporting mode and a TE_3-supporting mode with respect to the coupling length. All simulations taken at λ = 1.55 μm. (b) An illustration of the TE supermode between the two waveguides.

For the higher-order devices, an increase in XT is observed. From Figure 4.3, the peak effective refractive indices are lower for higher-order modes. This is because of the

ultrawide aspect ratios of the waveguides that support them, which weakly confine the mode vertically. This reduction causes the difference in effective index Δn_{eff} between adjacent modes to become smaller for higher-order modes; thus, increasing XT. Because of the high index contrast of the SOI platform, many modes can still be supported without suffering from too much XT. The effect is present for each device type, but less so for the waveguide-wrapped and racetrack devices, as the difference in coupling strength between the desired and undesired modes is greater. The coupling length for the waveguide-wrapped device is not limited by FSR, like the racetrack device is, but it is limited by the circumference of the microring. A greater degree of wrapping also introduces more bending loss—especially for higher-order modes.

4.2.2 Design of Microdisk-Based Optical Communications Circuits

For a direct modulation approach, where the data signal is generated by the multiplexing microrings, considerations must be made on the embedded pn modulators. In the literature, the most promising modulator design for attojoule-scale operation is the vertical pn junction. This approach takes advantage of the wide aspect ratio of SOI waveguides to achieve orders of magnitude higher energy efficiencies than conventional, lateral pn junction designs. In order to contact to a vertical pn junction, a microdisk structure must be used, which would prohibit the use of a racetrack coupling section. Because the size of the resonator should be kept small, for the additional purpose of low device capacitance (voltage, energy), the use of the point-coupled microring should also be avoided, as it suffers from high IL and XT for small radii. Thus, the waveguide-wrapped design would fit best for this application. Although microdisks enable the use of these high-efficiency

modulators, its cavity is naturally multimode, which leads to multiple unwanted resonances within the free spectral range of its fundamental mode. This ultimately limits the bandwidth of the device. However, the microdisk structure has also been shown to achieve extremely small radii [107], due to the removal of the inner, loss-inducing wall of the microring cavity and the optical-confinement-reducing slab region needed to support a contact for a regular microring modulator. The further reduction of the radius will expand the working FSR, while also removing the higher-order modes within the cavity.

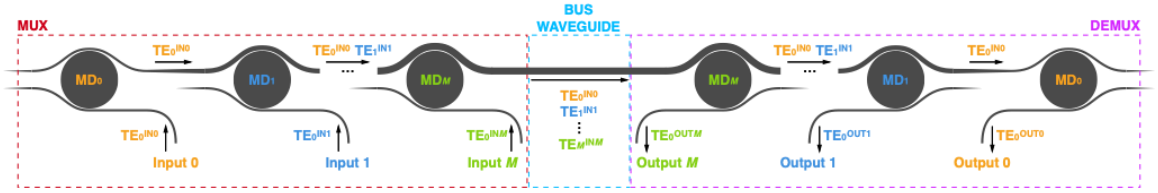


Figure 4.9 Diagram of the mode distribution within an M-mode microdisk-based MDM MUX/DEMUX circuit. Each microdisk resonator is connected to the main bus waveguide in series and has an input (MUX) or output (DEMUX) waveguide on its other side. The portion of the multimode bus waveguide between the MUX and DEMUX is kept short for visualization purposes. For added WDM functionality, detuned microdisks would be added in series to the existing microdisks.

Adapting the waveguide-wrapped design to the microdisk resonator follows a similar process, as shown in Figure 4.9. From a single TE_0 laser source, the first M microdisks multiplex M distinct modes, TE_0 to TE_M, onto a main multimode bus waveguide. For demonstration purposes, the continuous-wave operation of the circuit is characterized in this work. In a real-world application, the microdisks would be individually modulated to send on-off keyed data across the multimode bus waveguide to different parts of the chip. For an added WDM functionality (not shown in the figure), slightly detuned microdisks

would be added in series to the existing microdisks to handle separate wavelength channels. For example, a three-mode, three-wavelength MUX would require nine microdisks to operate (in the order of $MD_{0,\lambda 0}$, $MD_{0,\lambda 1}$, $MD_{0,\lambda 2}$, $MD_{1,\lambda 0}$, $MD_{1,\lambda 1}$, $MD_{1,\lambda 2}$, $MD_{2,\lambda 0}$, $MD_{2,\lambda 1}$, $MD_{2,\lambda 2}$). At the end of the circuit, the DEMUX is reciprocal to the MUX, where the signals travel in the same manner but in the reverse direction.

For the simulated characterization of the circuit, the design was separated into three different add-drop waveguide-wrapped microdisk resonators (denoted by MD_0, MD_1, and MD_M (MD_2) in Figure 4.9). Each microdisk works as a whispering gallery mode resonator. Wavelengths that are of integer multiples of the effective circumference of the disk are guided by total internal reflection and are dropped from the add waveguide to the drop waveguide. Each device features a 0.22 μm thick silicon disk, with a 5 μm radius, that is spaced 0.15 μm from each of its two waveguides. The input waveguide of each MUX resonator device (output waveguide for the DEMUX resonators) is 0.44 μm wide to match the effective refractive index of its TE_0 mode to that of the fundamental whispering gallery mode of the disk. MD_0 is symmetrical, but MD_1 and MD_2 feature wider drop waveguides to support the higher-order modes. To achieve the critical coupling condition, the widths of the MD_1 and MD_2 drop waveguides are optimized to match the effective refractive index of the fundamental whispering gallery mode inside the disk. Each device is optimized for a specific mode, while the other modes will only be weakly coupled, leading to low intermodal crosstalk.

As visualized in Figure 4.10, the commercial eigenmode solver from Lumerical (MODE) is used to simulate and visualize the profiles of the whispering gallery modes inside the microdisk resonator and the three supported TE modes of the bus waveguide. The wrapped waveguide works as a directional coupler to the microdisk. The bend degree is optimized to achieve the highest transmission for the desired mode and the lowest transmission for the undesired modes. Carefully designing the width of the drop waveguide and the length of the coupling portion reduces the insertion loss and intermodal crosstalk of the device. For the MD_1 and MD_2 devices, the optimized drop waveguides are 0.84 and 1.22 μm wide, respectively. Devices that support higher-order modes are created by the same process. For example, a TE_3-supporting device would simply require extending the waveguide to 1.54 μm; however, the insertion loss starts to become significant, as it becomes difficult to bend wider waveguides around a small disk without adding too much radiation loss. A larger disk would help to support more modes, with the expense of spatial footprint and free spectral range (reduced number of wavelength channels).

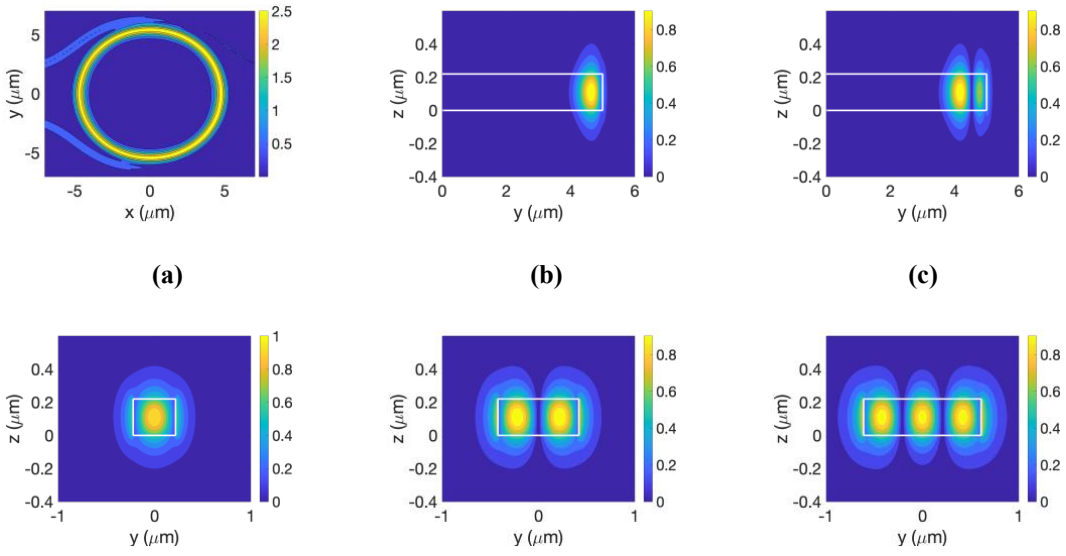


Figure 4.10 Simulated (a) top and (b) cross-sectional profiles of the microdisk's fundamental whispering gallery mode. (c) Cross-sectional profile of the microdisk's first higher order whispering gallery mode. Cross-sectional profiles of the (d) TE_0, (e) TE_1, and (f) TE_2 modes in the drop waveguides of MD_0, MD_1, and MD_2, respectively.

Because of the aforementioned issues with simulating the microdisk structures with the half-ring method, the full microdisk devices had to be simulated using the 3D finite-difference time-domain optical solver from Lumerical (FDTD). Figure 4.11 presents the simulated transmission spectra at the drop (output) ports of each of the first three DEMUX microdisk resonators (as shown in Figure 4.9), with the appropriate input modes. The select resonance peaks of the resonators are centered at 1,550.3 nm. Given the multimode nature of whispering gallery resonators, two higher order modes within the microdisk cavity produce two extra resonances on the spectra. Because the design of each microdisk was optimized for the fundamental whispering gallery mode, the higher order modes are not coupled as strongly. However, the first higher order mode (as shown in Figure 4.10c) is

strong enough to introduce significant WDM crosstalk. To avoid this resonance, the working FSR is reduced by 5.7 nm (i.e., the FSR is taken between the two highest order modes), which limits the number of available WDM–MDM channels. The radii of the microdisks can be reduced to eliminate the higher-order whispering gallery modes altogether (and further improve FSR, modulation efficiency, and compactness); however, insertion loss and intermodal crosstalk will suffer from the weaker coupling strength of the now lesser-wrapped waveguide. The insertion losses for the designed TE modes of the three microdisk resonators are found to be 0.06, 0.22, and 0.79 dB, respectively. And the worst-case intermodal crosstalk values for the MD_1 and MD_2 devices are found to be -22 and -12 dB, respectively. The MD_0 device's bus waveguide does not support higher order modes, thus crosstalk is assumed to be negligible.

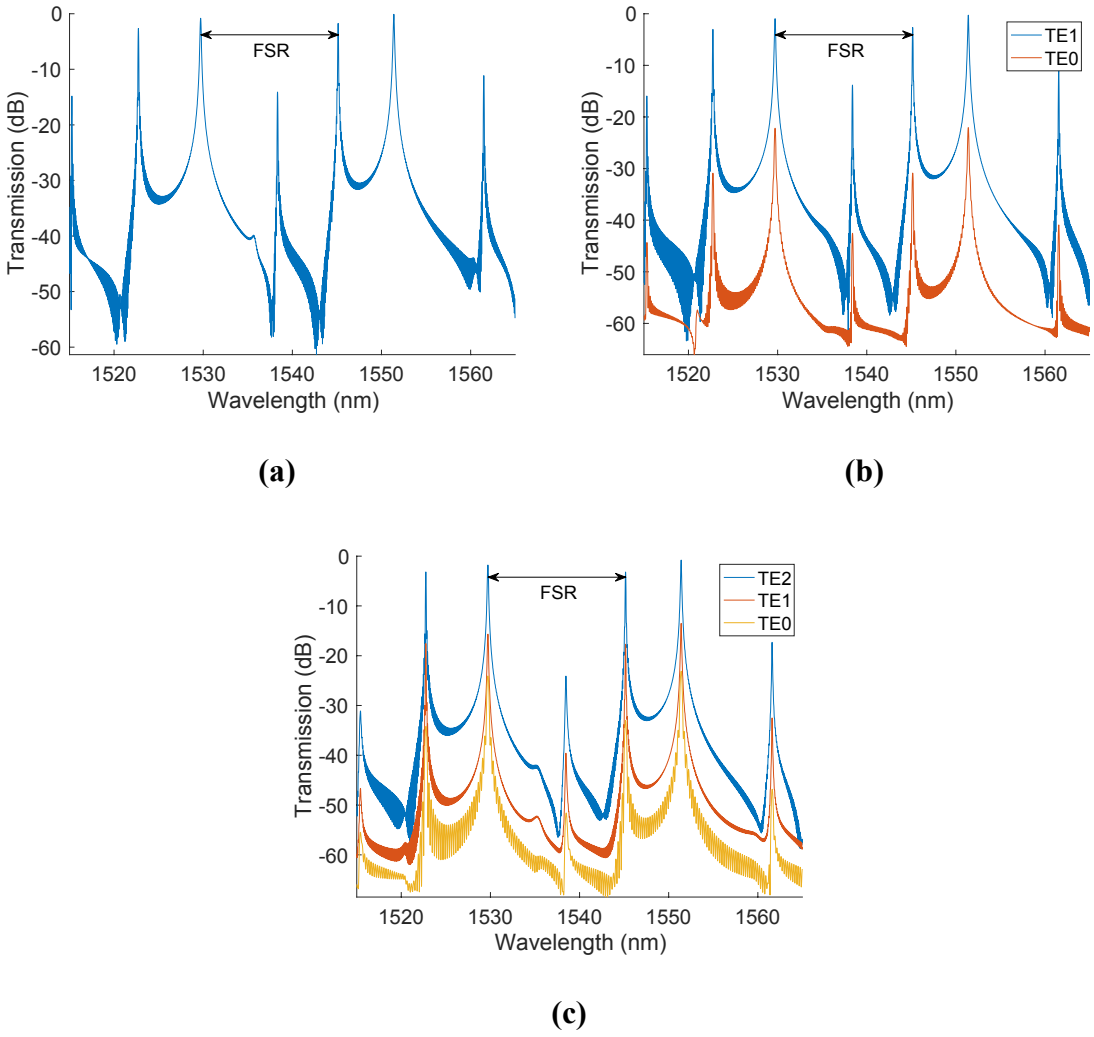


Figure 4.11 FDTD-simulated transmission spectra of the (a) MD_0, (b) MD_1, and (c) MD_2 DEMUX microdisk resonators. FSR is taken between the resonance peaks of the two highest order whispering gallery modes.

The microdisk devices are connected in series with adiabatic tapers to match the waveguides of various sizes. The bus waveguide that connects the MUX and DEMUX is kept short simply to match the spacing of the fiber array that is used for characterization. For the same reason, the microdisks are spaced by the same 127 μm pitch of the fiber array.

4.3 Fabrication of Microdisk-Based Optical Communications Circuits

The WDM–MDM MUX/DEMUX circuit was fabricated on an A*STAR Institute of Microelectronics (IME) multiproject wafer through CMC Microsystems and the Silicon Electronic-Photonic Integrated Circuits (Si-EPIC) Program from the University of British Columbia. The multiproject wafer features a 220 nm layer of silicon on top of a 2 μm layer of buried oxide and below a 3 μm layer of top oxide. The waveguides and resonators of the circuit are fully etched down to the buried oxide. The vertical grating couplers attached to each of the input/output ports are partially etched to achieve better coupling (by minimizing reflections) with the optical fiber. Integrated titanium nitride heaters are added to each microdisk to fix any misalignments of the transmission spectra introduced by fabrication variations. The layout of the chip, as shown in Figure 4.12, was created with KLayout [108], and the design rule checking was done with Calibre from Mentor Graphics [109] before submitting for fabrication. The layout includes the two-mode MDM MUX/DEMUX circuit, as well as three- and one-mode circuits for further testing. Basic waveguides and microdisk resonators are also included for baseline testing of the experimental setup. 16 chips were received: half were done with a 248 nm ultraviolet exposure, and the other half with a higher-resolution, 193 nm exposure. The chips that used the 248 nm exposure could not resolve the tight gap sizes between the disk and bus waveguides, so they were not used in this testing. Images of the full chip, the portion of the chip with this layout, the two-mode MUX/DEMUX circuits, and the three designed microdisk resonators are shown in Figure 4.13.

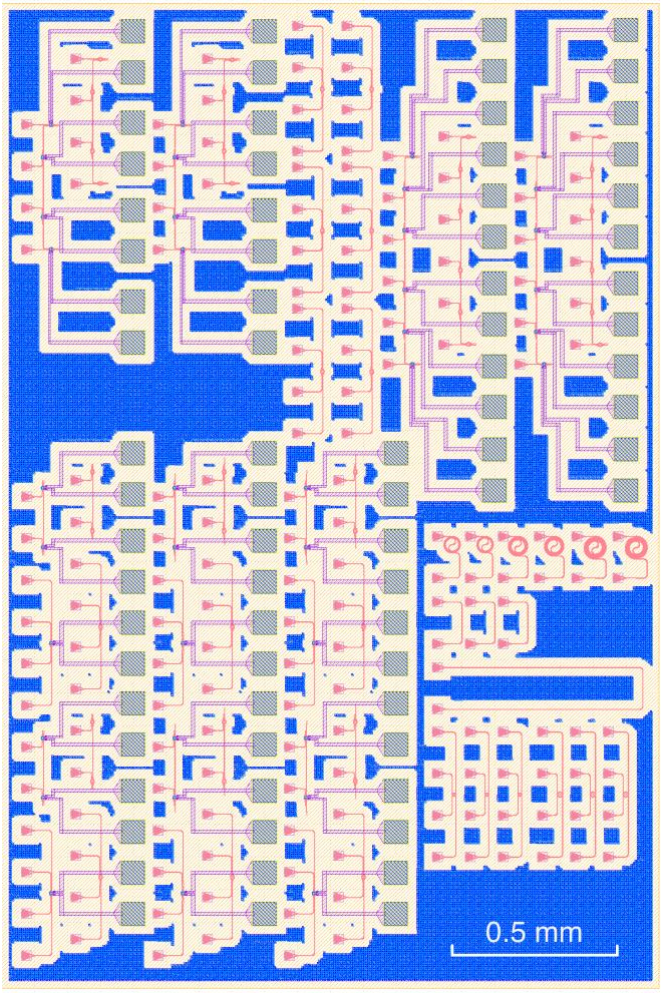


Figure 4.12 Layout of the chip received from the IME multiproject wafer.

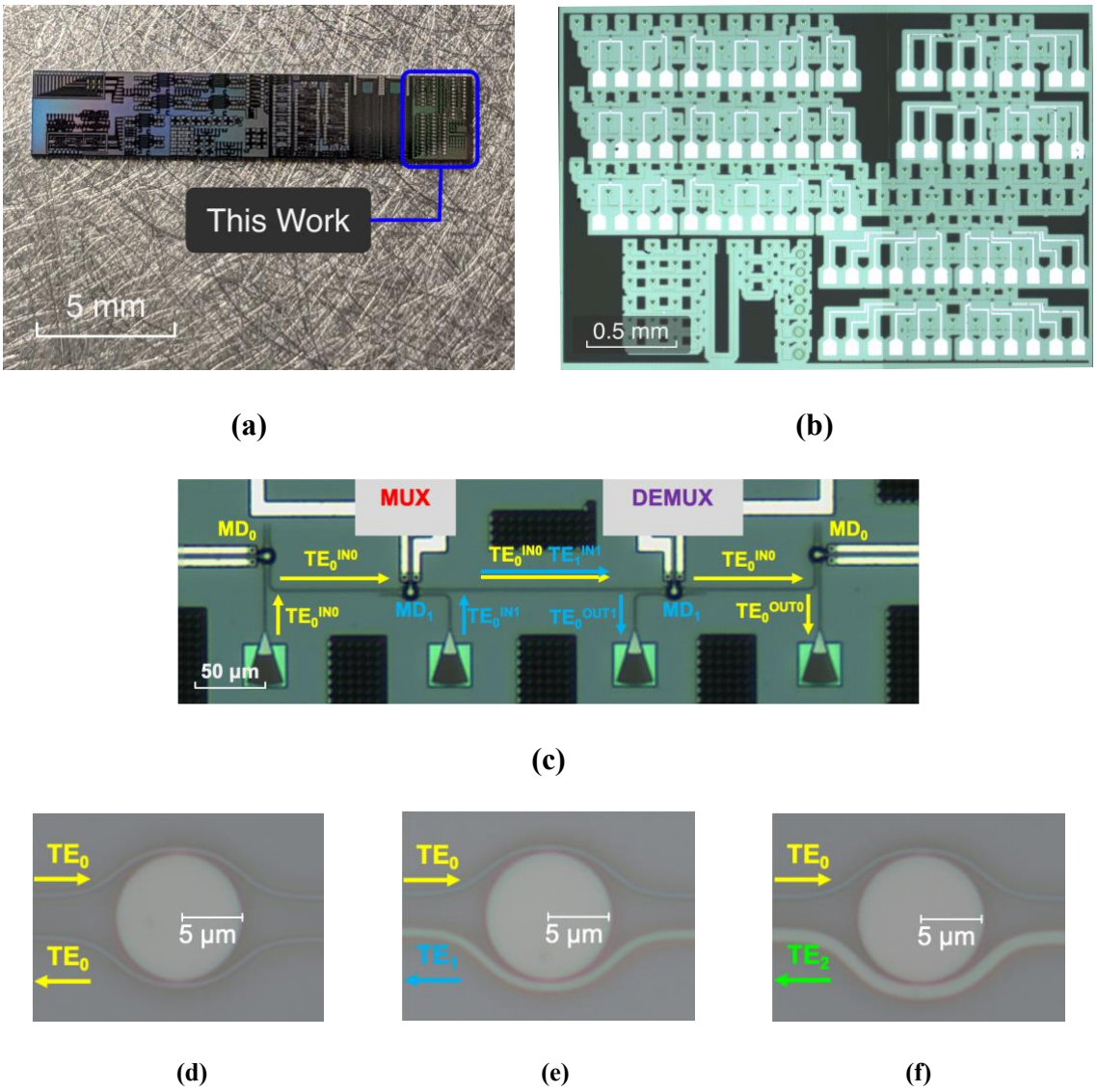
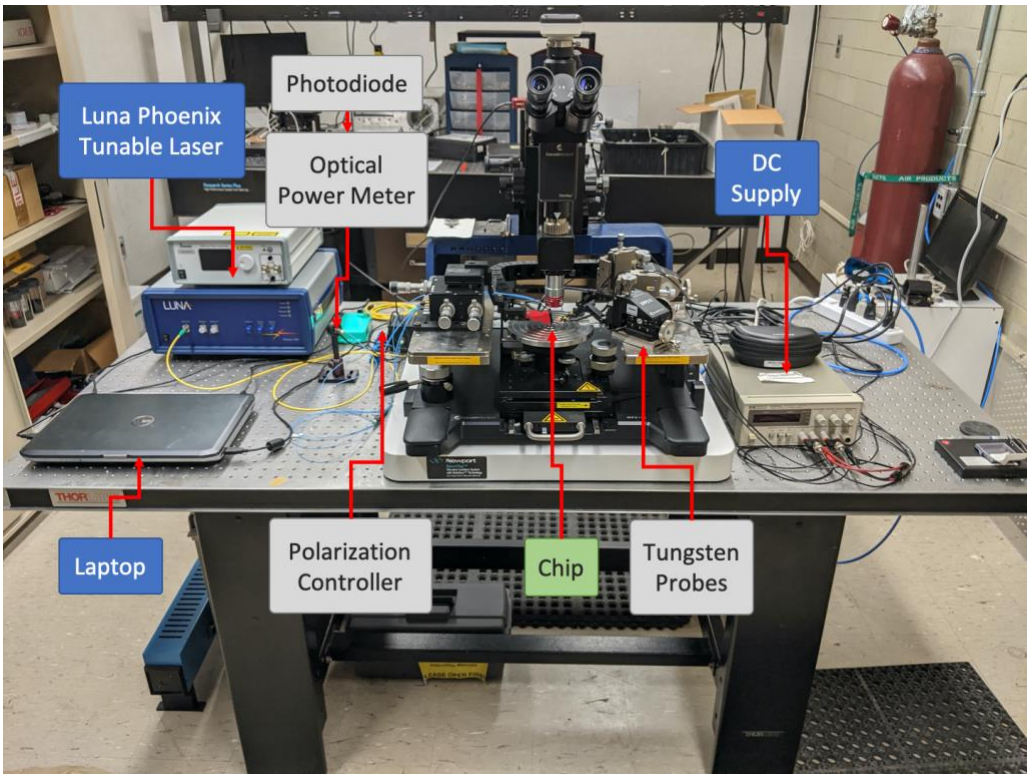


Figure 4.13 Image of (a) the chip diced from the IME multiproject wafer. Microscope images of (b) the layout of the WDM–MDM circuits, (c) the two-mode MUX/DEMUX circuit, (d) MD_0, (e) MD_1, and (f) MD_2.

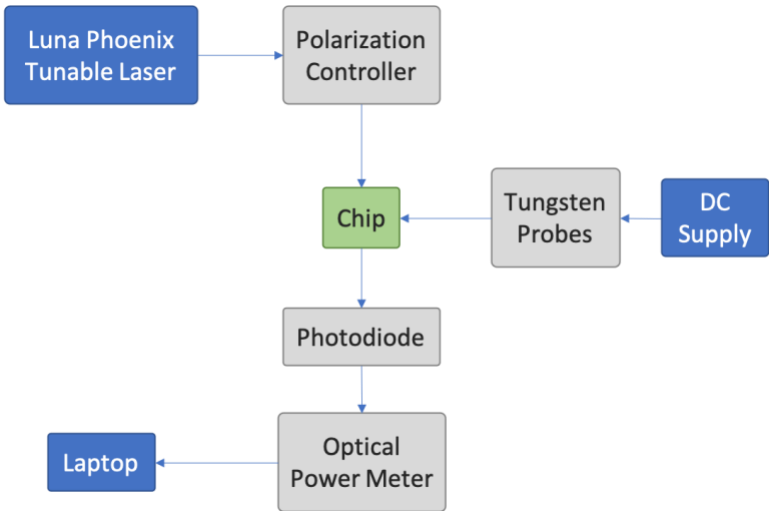
4.4 Experimental Characterization of Microdisk-Based Optical Communications Circuits

The experimental setup used to characterize these devices is shown in Figure 4.14. A tunable laser is the light source for the experiment. A fiber-based polarization controller

connects to the laser to establish a TE polarization state before coupling to a 4-channel polarization-maintaining TE fiber array. The spacing between the fibers (127 μm) matches that of the grating couplers in the layout so the other fibers can be used as additional inputs or outputs, depending on the device being measured. The fiber array is set to an angle that maximizes coupling at the selected resonance wavelength of the circuit. The output fiber array is connected to an optical power meter for real-time alignment of the fibers to the device. For the highest resolution transmission spectra measurements, the tunable laser is manually swept, and the transmission data is gathered at the optical power meter in wavelength increments of 0.01 nm (only where the transmission peaks actually occur). Four tungsten probes with variable DC supplies are contacted to the chip from the opposite side of the chip.



(a)



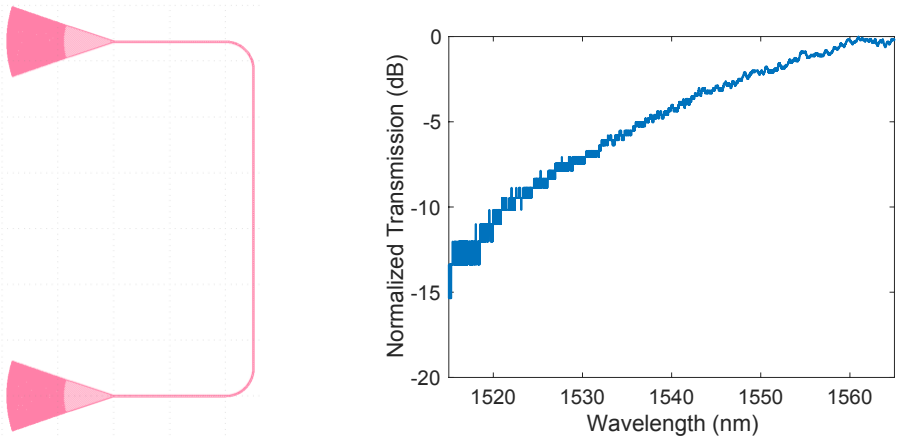
(b)

Figure 4.14 (a) Image and (b) diagram of the experimental setup used to characterize the WDM–MDM MUX/DEMUX circuits.

Two directly connected grating couplers, as shown in Figure 4.15a, are used to calibrate the angle of the fiber array and the polarization of the inputted light. These were set to maximize transmission at a wavelength of 1560 nm (near where the selected resonance peak of the circuit is). Prior to characterizing the (DE)MUX circuit, the standalone MD_0 and MD_1 devices, as shown in Figure 4.16a and Figure 4.16b, were measured to verify their mode-selective resonant behaviors. Figure 4.16c shows the through-port transmission spectrum from the input port for both devices. Because the two ports are connected to two TE_0 grating couplers, the TE_0 mode runs through both microdisk devices. Because MD_0 is designed to couple out TE_0, resonance peaks are observed on the spectrum; whereas for the MD_1 device, which is not designed to couple out TE_0, the resonance peaks are not observed. Figure 4.16d shows the drop-port transmission spectrum from the add port for both devices. Because both devices are designed to couple in TE_0, resonance peaks are observed for both. This verifies that the devices are working as expected. Note that the small wavelength shift between the two spectra is likely due to fabrication variations from the nominal design. Also note that the multiple resonance peaks per FSR is due to the multimode nature of the microdisk cavity.

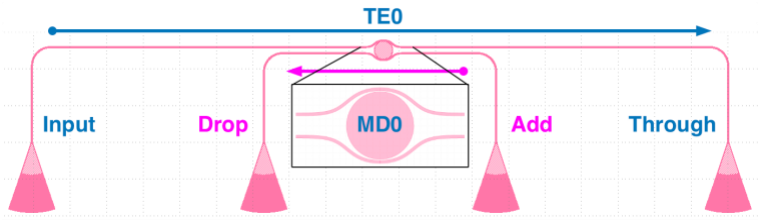
The experimentally measured transmission spectra of the 2-mode MUX/DEMUX are shown in Figure 4.17. Note that the grating coupler transmission spectrum in Figure 4.15b is subtracted from this to gather a more accurate result for the circuit itself. Each curve shows the transmission from Input *m* to Output *m*. The input TE_0^{IN0} is first multiplexed onto the multimode bus waveguide as TE_0 by the MUX MD_0, followed by passing through two MD_1 devices, before being demultiplexed out as TE_0^{OUT0} by the DEMUX MD_0. The

second input TE_0^{IN1} is injected and multiplexed by the MUX MD_1 onto the same multimode bus waveguide as TE_1, before being demultiplexed out as TE_0^{OUT1} by the DEMUX MD_1.

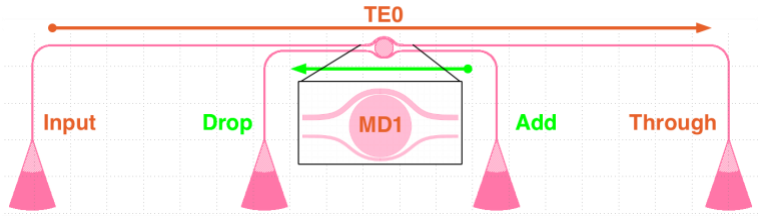


(a) **(b)**

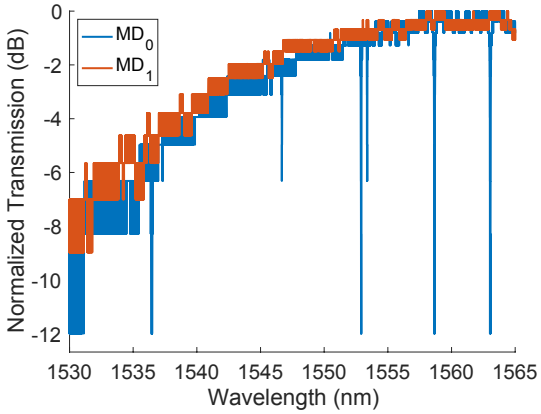
Figure 4.15 Layout of (a) two connected grating couplers used for measuring insertion loss attributed to fiber–chip coupling. The (b) transmission spectrum demonstrates the strong wavelength-dependency of the grating couplers.



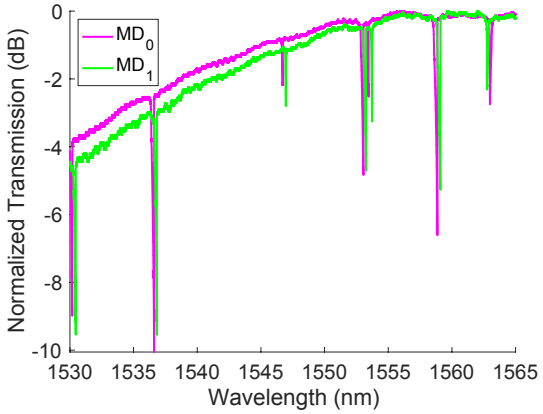
(a)



(b)



(c)



(d)

Figure 4.16 Standalone (a) MD_0 and (b) MD_1 devices used to demonstrate mode selectivity. Transmission spectra of the MD_0 and MD_1 devices for (c) the input to through port, and (d) the add to drop port.

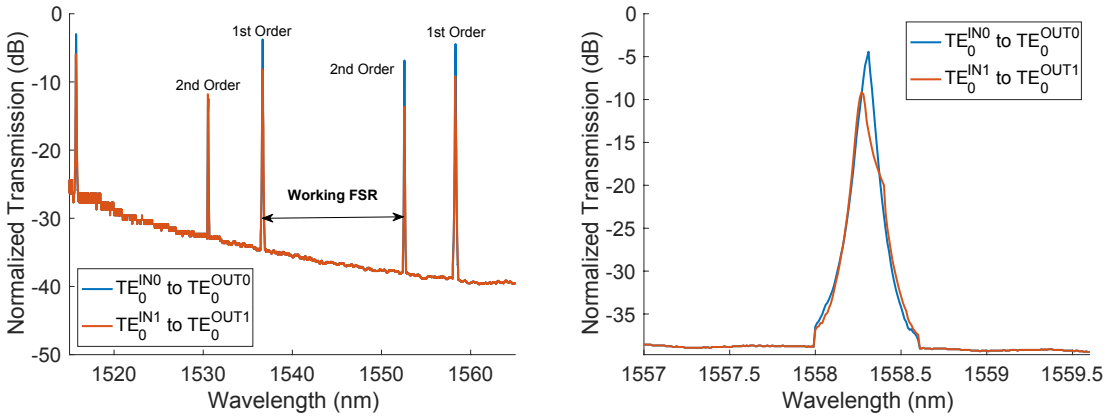


Figure 4.17 Experimental transmission spectra of a two-mode WDM–MDM MUX/DEMUX circuit, which includes a cascading link of MD_0, MD_1, MD_1, and MD_0 devices (as shown in Figure 4.13c). The transmissions spectra are measured for the TE_0^{IN0}–TE_0^{OUT0} and TE_0^{IN1}–TE_0^{OUT1} transmission links for (a) the entire C-band and (b) the resonance peak used for this circuit.

The selected resonance peaks of the outputs TE_0^{OUT0} and TE_0^{OUT1} are centered at 1558.31 and 1558.28 nm, respectively. These center wavelengths can be aligned together using the integrated thermal tuners. The quality factors of the resonators are ~26,000 (λ_{FWHM} = 0.06 nm). A resonator with this quality factor holds a photon with a lifetime of ~8.5 ps, which limits the single-channel modulation speed to 18 GHz. Reducing the quality factor of the resonators will increase the maximum speed of operation but will also increase the energy consumption and reduce the number of channels. This is a tradeoff that can be tuned to the desired specifications of the application. The FSR of all the microdisks (of 5 μm radii) are the same, 21.6 nm. Note that one other resonance was observed to appear within the FSR due to the higher order whispering gallery mode inside the disk cavity. This extra resonance reduces the working FSR to 15.9 nm, which still allows for 265 WDM channels and 530 combined WDM–MDM channels for this 2-mode MUX/DEMUX circuit. As shown in the

standalone transmission spectra in Figure 4.16d, higher order resonances likely exist but are too weak to be experimentally detected (due to slight misalignments of high-Q resonances). For every additional MDM mode added to the circuit, as supported by the simulations, an extra 265 channels would be added to the aggregate bandwidth. The insertion losses of the TE_0^{IN0}–TE_0^{OUT0} and TE_0^{IN1}–TE_0^{OUT1} transmission links are measured to be 4.0 and 13.1 dB, respectively. The intermodal crosstalk was found by measuring the TE_0^{IN1}–TE_0^{OUT0} and TE_0^{IN0}–TE_0^{OUT1} transmission links. Crosstalk is very low for both outputs, at -27.4 dB for TE_0^{OUT0} and -25.4 for TE_0^{OUT1}.

4.5 Summary

In this chapter, three microring resonators are numerically optimized and compared for the purposes of low-power, high-bandwidth WDM-compatible MDM: the point-coupled microring, the racetrack microring, and the waveguide- wrapped microring. For low-radii devices, which offer benefits of low spatial footprint, wide free-spectral range (bandwidth), and low capacitance (energy consumption), it is observed that regular, point-coupled microring resonators are not suitable, as they suffer from high insertion loss and crosstalk for higher-order modes. The racetrack microring improves on insertion loss and crosstalk by increasing the length of the coupling region (i.e., increasing coupling strength), but this has the inverse positive effect of low radii microrings. The waveguide-wrapped microring introduced in this work increases the coupling length/strength without increasing the size of the microring. While avoiding the downsides of the other two designs, the waveguide-wrapped microring also offers flexibility in the types of pn modulators it supports. For example, the microring can easily be replaced by a microdisk, which naturally supports

high-efficiency vertical-junction pn modulators (otherwise much more difficult to embed into the racetrack microring). Should there be a continued trend towards more optical communication and processing on the chip, combining WDM and MDM is a promising way of increasing aggregate bandwidth. The comparisons and discussions made here are important to consider when designing for low-loss, low-crosstalk, high-bandwidth, and low-power applications of the WDM–MDM MUX/DEMUX circuit.

Additionally, this chapter presents the design and experimental demonstration of the first WDM-compatible MDM MUX/DEMUX circuit based on waveguide-wrapped microdisk resonators. The introduction of waveguide-wrapped resonators allows for highly selective mode coupling without having to increase the size of the resonator cavities. Smaller resonator cavities reduce the spatial footprint of the entire circuit, reduce capacitance (energy consumption) of the potential modulation, and increase the number of supported wavelengths (bandwidth density) through their inherently larger FSRs. The microdisk structure provides a platform for attojoule-level energy consumption through vertical-junction depletion-type modulation. The experimentally demonstrated MUX/DEMUX supports the first two TE modes, offering a worst-case insertion loss of 13 dB, a worst-case intermodal crosstalk of -25.4 dB, and can support up to 530 channels of combined MDM–WDM data communication. The simulations show promising designs that can be scaled to support one or more additional modes, which would each add 265 channels to the circuit. For modest single-channel data rates, the proposed design demonstrates its high potential for low-energy, ultrahigh-bandwidth information processing and communications.

Chapter 5: Automated and Accelerated Design of Photonic Integrated Circuits with Machine Learning

Designing photonic circuits requires numerical electromagnetic solvers that can take minutes to hours to characterize a single device. Tens or hundreds of simulations are often needed to optimize the device for a specific figure of merit. Although this is unavoidable for most novel device-level designs, the development of standard secondary components of the photonic circuit further delays the time to fabrication and testing. Like the microelectronics (CMOS) industry, libraries of compact models have become more available for circuit-level simulations [110] but are not flexible to the specific needs of different applications (e.g., working at a different polarization or center wavelength). In this work, flexible compact models of essential secondary components are developed for the acceleration and automation of a wider range of future photonic circuit designs.

Artificial neural networks (ANNs) are universal approximators that can generalize the input–output relationship of any complex nonlinear function [111]. Although ANNs are most widely used in voice and image recognition applications, they have been used to accurately model microelectronic circuits as well [112, 113, 114]. Their use in photonics is limited, but recent works have presented models that accurately approximate light scattering spectra [115, 116]. Given a large dataset of inputs and outputs (i.e., the design properties and performance metrics of a photonic device), an ANN can learn the input–output relationship of any photonic device. In this work, the example of a polarization insensitive SOI subwavelength grating (SWG) coupler is used to demonstrate the capabilities of the proposed design method. With the trained ANN model of the coupler, a

set of simple linear algebraic and nonlinear activation calculations are made to predict the output of the device three orders of magnitude faster than the numerical simulations, without the need for empirical approximations. The ANN model uses a significant amount of computational resources up front in order to accelerate and automate future designs. Other frameworks have been developed for the automated design of photonic grating couplers [117]; however, every design requires the full, resource-heavy numerical process. The time-saving potential of the neural network method is demonstrated by running a brute-force, parametric sweep solver on the ANN grating coupler model, which finds an optimal design 61 times faster than a commercial, numerically solved particle swarm optimizer (PSO) [118] often used in the design of photonic devices [119, 120]. The trained model and the design framework are both made openly available on a public GitHub repository [121].

This work was documented and published in the IEEE Journal of Selected Topics in Quantum Electronics in 2018 and was presented at the 2018 IEEE International Conference on Group IV Photonics.

5.1 Artificial Neural Network Modelling of Photonic Devices

Designing photonic devices with the proposed machine learning framework is split into three main sections: data acquisition from numerical simulations, the construction and training of a deep ANN model, and device optimization using inference of the trained ANN model. This process does not change for different device types; thus, the process is generalized and automated for the accelerated development of future flexible, compact,

and open models of standard photonic components. The MATLAB numerical computing environment is used to handle this process, with Lumerical FDTD being used as a back-end finite-difference time-domain numerical optical solver. Lumerical's software suite integrates nicely into MATLAB, where scripts commands can be run from the MATLAB design environment, but any optical solver can in theory be used here.

The data needed to train and validate the model can be acquired experimentally, directly from real devices; however, training the model to a high accuracy requires hundreds of time- and resource-consuming measurements. Machine-learning-based design is thus more suitable when data are acquired numerically, as the measurements can be taken quickly and automatically, with a lower chance of external factors (e.g., measurement and fabrication error) being learned by the model. Given a reliable source of numerically solved data, the final accuracy of the trained model can be taken as the main source of device validation.

A simulation file is first created (manually) for the photonic device. Ideally, every parameter of the device is fully variable and learnable—to train a robust model useful for any design case—but compute resources are still limited. It is important to choose the parameters that are likely to cause a meaningful change in the device performance and to define a range of acceptable values for them. These may be continuous values, such as the height of a waveguide, or discrete, like specific materials or polarization states. The variable parameters are known as the features of the ANN model. Likewise, the outputs of interest are extracted into the labels of the model. Labels can be a single value, like the

center wavelength of a transmission spectrum, or vectorized into multiple labels, like for the full transmission spectrum at a device port.

The data acquisition algorithm generates a set of randomized feature values within the ranges specified, numerically solves for the specified labels, and stores them together in a data object as an example. This is repeated for a user-specified number of examples. A more suitable approach is also available, which uniformly sweeps through the range of features. Here, a specified resolution of the uniform sweep determines the number of examples that are acquired. A uniform order, however, introduces an external source of bias into the ANN model, negatively affecting its ability to generalize the photonic device. After the data are acquired, the order is automatically randomized to eliminate the bias. The example set is also automatically split into two separate groups: training and validation, with an 85:15 respective ratio set as the default. The model is trained with the training set and the performance of the model is evaluated with the validation set. Each feature and label are also normalized between zero and one, as the parameters of a photonic device may have largely different orders of magnitude, which otherwise negatively affect the training rate and accuracy.

Before doing any training, it is important to manually inspect the data to make sure the seeding simulations produced reasonable results, as the training algorithm does not have any knowledge of the underlying physics: only the relationship between the inputs and outputs. Because the data are acquired numerically, time is saved here by not having to inspect each example. If the simulation was set up correctly, every example can be expected

to be meaningful, even if the results are undesired. Unlike conventional, optimization-based design approaches, poor results are not discarded; rather, they are used to train the model to find optimized designs across a broad range of applications.

The proposed ANN structure (model) is the feedforward multilayer perceptron, as shown in Figure 5.1. The structure consists of at least three layers of fully interconnected neurons. The connections between the neurons are the synapses that each hold a value (weight) that ultimately represents a working part of the modelled photonic device. These weights are initially randomized with a mean of zero and a standard deviation of one but will be updated on every iteration of the training process. At the highest level of abstraction, a set of features are loaded onto the input layer and forward propagate through at least one hidden layer, until the output layer, where a set of predicted labels are produced. A deep learning model simply adds more hidden (middle) layers, which generally produce more accurate results for highly nonlinear functions [122]. At each layer k, we perform a multiply and accumulate operation:

$$\mathbf{y}^{(k)} = g\left(\mathbf{w}^{(k)}\mathbf{y}^{(k-1)} + \mathbf{b}^{(k)}\right), \tag{5.1}$$

where $\mathbf{w}^{(k)}$ is the weight matrix of the k^{th} layer, $\mathbf{b}^{(k)}$ is an added bias weight vector of the k^{th} layer, and $\mathbf{g(x)}$ is a rectified linear (ReLU) nonlinear activation function, given by

$$g(x) = \max(0, x). \tag{5.2}$$

The modeller has the choice of sigmoidal, hyperbolic tangent, and ReLU activation functions. In general, the ReLU function produces higher training accuracies, quicker. Because the model is predicting a set of continuous values at the output, the activation function is omitted at the output layer. For classifier models, which produce the probability of a relationship (e.g., image recognition), an activation function is applied to the output.

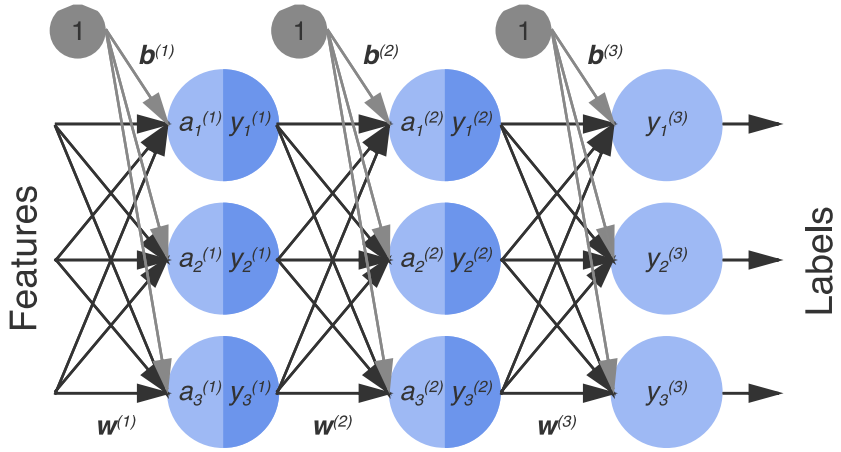


Figure 5.1 An example structure of a 3:3:3:3, fully interconnected multilayer perceptron artificial neural network. Blue circles are the neurons, grey circles are constant bias inputs, and the arrows are the weighted and trainable synapses.

After each forward propagation, the performance of the model is evaluated by calculating the squared error cost function, given by

$$\mathrm{E} = \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathbf{labels} - \mathbf{predictions})^2, \tag{5.3}$$

where n is the number of predictions made. The error is minimized with full-batch gradient descent, which adjusts the weights and biases of the model after every example is run through the network and evaluated by the cost function. Stochastic and mini-batch gradient descent, which update the weights after one or a few examples are evaluated, do not offer better performance for this model. To update the weights, the derivative of the cost function is found with respect to the weights, at every layer, in the process of backwards propagation. The chain rule is used here:

$$\frac{\partial \mathbf{E}}{\partial \mathbf{w}} = \frac{\partial \mathbf{E}}{\partial \mathbf{g}} \frac{\partial \mathbf{g}}{\partial \mathbf{a}} \frac{\partial \mathbf{a}}{\partial \mathbf{w}}, \tag{5.4}$$

where

$$\mathbf{a}^{(k)} = \mathbf{w}^{(k)} \mathbf{y}^{(k-1)} + \mathbf{b}^{(k)}. \tag{5.5}$$

Using (5.4), the weights and biases are updated by

$$\mathbf{w}^{(k)} = \mathbf{w}^{(k)} - \alpha \frac{\partial \mathbf{E}}{\partial \mathbf{w}^{(k)}} \tag{5.6}$$

$$\mathbf{b}^{(k)} = \mathbf{b}^{(k)} - \alpha \frac{\partial \mathbf{E}}{\partial \mathbf{b}^{(k)}}, \tag{5.7}$$

where α is the learning rate hyperparameter. Hyperparameters like this can cause major performance variations; it may take several iterations of trial and error to find the right

learning rate. Towards the end of training, the learning rate may be lowered to descend towards the minimum of the gradient function with fewer fluctuations in loss—ultimately finding a lower minimum. After the weights have been updated, a new full-batch forward propagation with prediction is made, which will produce a lower error value. This process is repeated until a suitably low error is achieved.

Inference is run on a well-trained ANN model to accurately predict the output of a device for any set of features within the range of simulated data. Inference is the process of forward propagation through the network to get a prediction, but with no backward propagation and weight adjustment to follow. Given the weights and activation functions used in the network, via the open-source model, inference can be run regardless of the simulation and training software used to develop the model.

To design for a specific device, any optimization algorithm can be used with the model, which can be considered as a black box that takes feature inputs and produces predictions very quickly. An evolutionary algorithm, such as particle swarm optimization, can be used; however, given the ultrafast prediction speed of an ANN model, a brute-force parametric sweep is more than capable of producing a rapid optimization. The open-sourced device development framework uses a simple parametric sweep based on user-defined conditions and constraints. A condition is what the framework solves for (e.g., minimize loss at λ_0 = 1.55 ± 0.01 μm), and a constant is a feature that is not included in the sweep (e.g., keep etch depth at 0.22 μm). Given a sweep resolution, the solver finds the best fits by running inference on each set of features in the sweep.

Given the difficulty of achieving 100% training accuracy, it may be necessary to cross reference the solved device in the original numerical solver and perform a quick, narrow optimization to find a truer optimum. If results vary significantly, it may be necessary to retrain the network or collect more training data.

5.2 Device Design Using Trained Artificial Neural Network Models

The generalized ANN modelling process is used to create a compact, flexible model of a polarization insensitive SOI SWG grating coupler, as shown in Figure 5.2. Due to the wide usage of this type of device for coupling light into and out of photonic circuits, each application may require slightly different performance metrics (e.g., center wavelength or polarization state). The trained ANN model accelerates the design process of the device, for an arbitrary set of metrics, in comparison with a conventional, numerically solved approach.

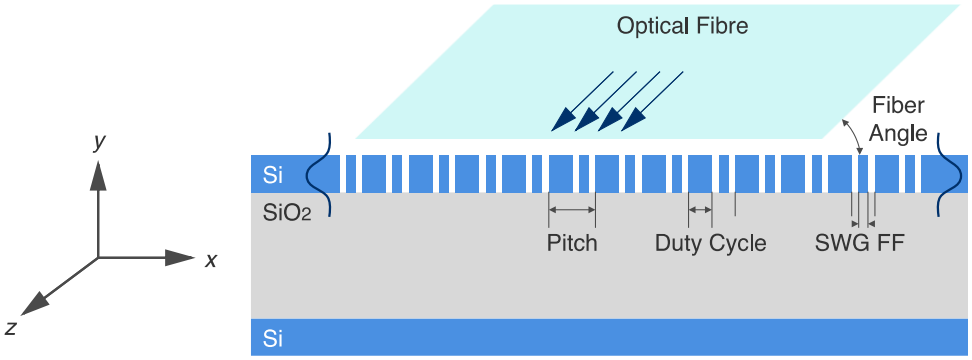
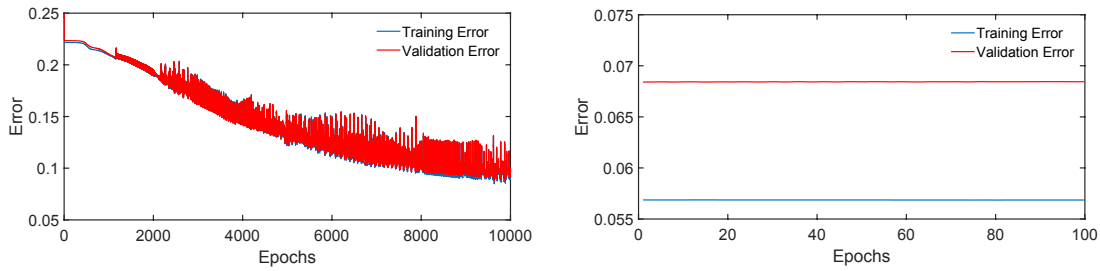


Figure 5.2 A 2D cross-section of a straight SOW subwavelength grating coupler. The labeled dimensions are the features (variable parameters) defined in our ANN model.

A 2D Lumerical FDTD simulation file is created for a straight (1D) grating coupler with alternating SWGs. Only a cross-sectional, 2D (x, y) FDTD simulation is needed to accurately characterize this device, as it extends uniformly in the z direction past the extent of the incident optical mode. For a focusing-style grating coupler [123], which reduces the required footprint, a longer-duration, 3D FDTD simulation would be needed. For simplicity, the modelling of the straight grating is demonstrated here, noting that the process would remain identical for the focused grating. The features (variable parameters) of the model are defined to be the fiber angle, polarization, grating pitch, the duty cycle of the large grating, and the fill factor (FF) of the subwavelength grating. The range of variable values are 5° to 20°, 0 (TE_0) and 1 (TM_0), 0.5 to 1.5 μm, 0.4 to 0.8, and 0.2 to 0.6, respectively. A uniform distribution of the variable parameters is created, producing 9,190 examples, and each set of features is simulated—each taking 23.8 seconds to complete on an 18-core, 3.4 GHz, Intel Core i9-7980XE CPU. For each simulation, the 200-point transmission spectrum at each end of the grating coupler is extracted for wavelengths between 1.3 and 1.7 μm. The spectra are then automatically discretized into their maximum transmission values, with the corresponding λ_0 values. Along with the automated data engineering steps the framework takes, the examples that have maxima on the edges of the spectrum are removed, as the true maximum may lie outside of it. 4,831 training examples and 852 validation examples are leftover.

Given the sizes of the feature and label vectors, the network starts as a 5-neuron input layer and a 4-neuron output layer—both without activation functions. Through trial and error, the highest model accuracy is achieved from a structure with three ultrawide (much larger

than the input/output layers) hidden layers of 100, 50, and 50 neurons. Each neuron is equipped with a ReLU activation function. The learning rate was originally set to $\alpha = 1\times10^{-4}$ and manually decreased when the error curve would flatten or start to experience many fluctuations. The final learning rate was $\alpha = 2\times10^{-6}$. At each iteration, a mean percentage error was calculated based on the training data and the validation data. The error plot for these two datasets is shown in Figure 5.3a for the first 10,000 iterations (epochs), which covered the steepest section of the gradient descent. The fluctuations are a result of the model bouncing around local minima as it descends down the error gradient. The training is stopped when the validation error reached a minimum value: 6.8%. Although further training will continue to reduce training error (past 5.7%), validation error will start to increase, implying an overfit function. Because the function is being generalized for any unseen set of features, basing the error on already seen data would be misleading. Figure 5.3b shows the last 100 iterations of training, where the error values have stabilized. Note that with more data, and further tuning of the structure of the ANN, the validation error can be further reduced.



(a) **(b)**

Figure 5.3 Training and validation error for the (a) first 10,000 iterations of the ANN training process, where the error begins at 0.31, and (b) the last 100 iterations, which shows a stable minimum of the trained function.

Inference of the trained model takes 0.013 seconds to complete, which is 1,830 times faster than the FDTD simulation. This is a relatively low speedup factor, as this particular simulation runs quickly. For many potential models (e.g., a ring resonator that simulates in 25 minutes [124]), the speedup factor would rise above 100,000.

To demonstrate and test the design process using the open-source ANN model and parametric-sweep solver, an SOI subwavelength grating coupler is designed and optimized at $\lambda_0 = 1.55$ μm for the fundamental TE mode. Using a sweep step resolution of nine, in 8.1 seconds, the solver finds the features of the device as: fiber angle = 5°, pitch = 0.875 μm, duty cycle = 0.65, and SWG fill factor = 0.6. The labels were found as: $T_{TE0,max} = -0.5$ and $\lambda_0 = 1.55$ μm. This device is run back into the FDTD solver to cross-check the results. Figure 5.4 shows the FDTD transmission spectrum ($T_{TE0,max} = -0.47$ at $\lambda_0 = 1.56$ μm), which is within the range of expected error. To compare, the same device is optimized with the particle swarm optimizer in Lumerical FDTD. A generation size of five is used, and the optimizer is stopped when the maximum transmission begins to converge (five

generations). In 492 seconds, the optimizer finds the device as: fiber angle = 8.5°, pitch = 0.886 μm, duty cycle = 0.8, and SWG fill factor = 0.6 to achieve: $T_{TE0,max} = -0.5$ at $\lambda_0 = 1.56$ μm. Despite using a less sophisticated optimization method, the ANN model arrives at a similar result and accelerates the design process by a factor of 61.

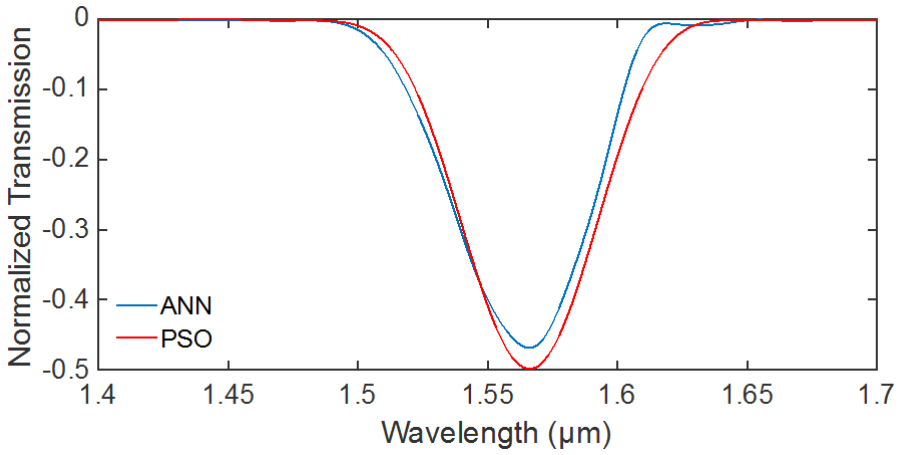


Figure 5.4 Normalized transmission spectra for the ANN- and PSO-designed single-polarization SOI grating couplers.

Because the ANN model was trained for both fundamental-order polarizations, at both ends of the grating coupler, a polarization-insensitive device can be easily designed as well. The maximum TM_0 transmission is found in the forward direction of the grating, and TE_0 in the backward direction. In 8.4 seconds, the solver finds the features of the device as: fiber angle = 14.375°, pitch = 0.75 μm, duty cycle = 0.65, and SWG fill factor = 0.25. The FDTD results were found as: $T_{TE0,max} = 0.34$ and $T_{TM0,max} = -0.32$ at $\lambda_{0,TE0} = 1.61$ μm and $\lambda_{0,TM0} = 1.591$ μm. Note that the opposite signs in transmission represent the counter-directional propagation of the two modes. The wavelength of polarization insensitivity is $\lambda_{0,TEM} = 1.596$ μm, where the matched transmission is $|T_{TEM,max}| = 0.31$ (−5.1 dB). This is within

the range of previously reported polarization-insensitive SOI grating couplers (−3 dB, −6.5 dB) [125]. Figure 5.5 shows the FDTD transmission spectra of the device.

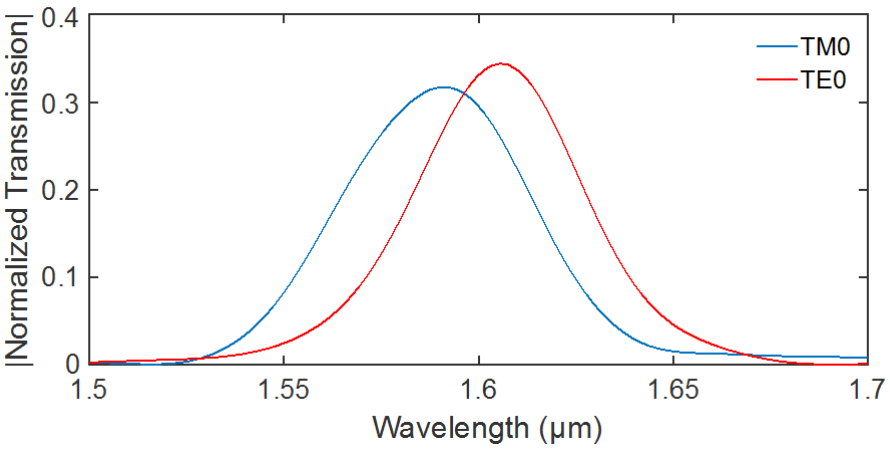


Figure 5.5 Absolute transmission spectra of the forward (TM_0) and backward (TE_0) directions of the polarization insensitive grating coupler.

5.3 Summary

This chapter presents an openly available compact, generalized model of a polarization insensitive SOI subwavelength grating coupler, featuring accelerated design of silicon photonic integrated circuits. The model is a deep artificial neural network trained by a dense set of finite-difference time-domain optical data in the proposed machine learning framework. The model predicts an output for the device 1,830 times faster than the corresponding numerical simulation, at 93.2% accuracy of the simulation. The open sourcing of standard photonic neural network models is beneficial for the rapid development of new photonic devices, given the speed and flexibility of the model, as well as the abstraction of the underlying physics and simulation into a simple, universally solvable package.

Chapter 6: Conclusion

This thesis presents multiple silicon photonic integrated circuit designs for the improved compactness, speed, and energy efficiency of optoelectronic processing (computing) and communications applications. By focusing on CMOS-compatible material platforms like silicon-on-insulator, these circuits can be cost-effectively produced at large scales using existing foundry processes.

Chapter 2 of this thesis presented the logic cell: a high-efficiency silicon microdisk optoelectronic modulator that simultaneously performs (N)AND, (N)OR, and X(N)OR logic operations in a single structure—a feature previously not found in the literature. The logic cell approach combines the best of electrical control and optical data transport by combining them at the logic device level. Although the system-level performance benefits of combining optical data transfer with electrical control were outside the scope of this work, this device presents the best mix of speed, compactness, and energy efficiency among all CMOS-compatible optical logic devices in the literature. Because the logic cell combines so many capabilities in a single structure, combining many of them to perform higher-order logic circuits is straightforward. In this work, N-bit optical AND/OR gates, adders, comparators, encoders, and decoders were demonstrated—each using many times fewer structures than previous demonstrations.

Chapter 3 explored novel pn junction designs to improve the efficiency of microresonator-based carrier-depletion modulators past the state of the art. Double-stacked vertical pn junctions, which may also be staggered to extend their length, increase the overlap with the

optical mode, which means the refractive index (resonance wavelength) of the device changes more per applied volt. This can result in direct reductions in the energy consumption of these devices, where this work shows a greater than 4X reduction over the conventional design. But given the direct tradeoff with speed and energy consumption in these devices, the improvement in efficiency means that they can operate at higher speeds as well. The final section in Chapter 3 presents the use of these novel pn junction modulator designs in logic, communications, optical repeaters, and wavelength converters, where multigigahertz-level speeds can easily be achieved. For each of these applications, these improved designs allow for the best mix of speed, compactness, and energy efficiency presented in the literature.

Chapter 4 applies the high-efficiency potential of microdisk resonators to ultrahigh-bandwidth on-chip data transfer. The combined wavelength- and mode-division multiplexing circuit has been shown to have the capability of sending hundreds of data channels down a single silicon waveguide. Previous demonstrations used microring resonators with long-straight bus waveguides, but this work shows that wrapping the bus waveguides around the resonator cavity improves the free spectral range (more wavelength channels) and compactness of the circuit, while also improving the mode selectivity. As the previous chapter presented the ultrahigh efficiency of microdisk modulators, they are used in this circuit as well. With microdisks, the novel waveguide-wrapped approach is absolutely necessary. The circuits were fabricated and experimentally verified. The experimental data shows that up to 530 combined channels are supported by a 2-mode circuit. The simulations show that more mode orders can be supported (at least three),

which would each add 265 channels of data for ultrahigh-bandwidth-density on-chip communications.

Chapter 5 adds a major improvement to the photonic integrated circuit design process used in Chapters 2 to 4 by leveraging machine learning. Artificial neural networks are used to model photonic integrated devices into compact flexible models that can be solved thousands of times faster than traditional electromagnetic solvers, and at high accuracies. This can automate and accelerate the design of common photonic components, which can significantly reduce the time to fabrication. This thesis work demonstrated the potential of this approach with the modelling and design of a polarization-insensitive subwavelength grating coupler. The trained neural network model predicts the outputs of the device 1,830 times faster than FDTD solvers, at an accuracy of 93.2%. Models like this can be made openly available to other research groups. Because these models are flexible, future design projects can adapt them to their own project specifications (e.g., different wavelength, minimum feature size, polarization) without having to make the many long simulations to fully design their own specific devices.

6.1 Significant Impact

The photonic integrated device and circuit designs presented in this thesis work offer a pathway towards ultraefficient, high-speed modulation and data transfer in next-generation processing and communications systems. With the exponential number of connected devices and bandwidth generated each year, improvements at the device and circuit level can make substantial improvements at the system level. Large performance gains are

already being made with the introduction of fiber optics and on-chip optical data transfer. With the compact, efficient optical computing architecture presented in this work, much of the electronic processing on the chip can start to be removed, along with any bottlenecking optical-electrical-optical converters. The ideal chip combines the best of optical data transfer and electrical control by integrating them at the logic level, rather than at added conversion points. This frees up space and energy to perform more operations on the chip. Furthermore, by using only CMOS-compatible designs, these devices can be seamlessly integrated into existing foundry processes to keep costs low for the high-volume production necessary to meet the growing bandwidth demands worldwide.

These same efficient microdisk modulators designed in Chapter 1 are easily adapted in this work to the application of ultrahigh-bandwidth-density on-chip data transfer. Optical data transfer already offers a much higher bandwidth potential than copper electronics; this work expands on the promising approach of combining wavelength- and mode-division multiplexing and further improves the compactness, bandwidth, and energy consumption over previous designs. Thus, more can be done on a single chip with a lesser energy impact.

This work demonstrates the high potential of high-speed photonics in optoelectronic integrated circuits. As these circuits become more complex, with hundreds or thousands of photonic components on the chip, the design process becomes more expensive and lengthier. The machine learning based design methodology presented in this thesis has the potential to accelerate the design process by many orders of magnitude, and with very little

input from the designer. This frees their time and energy to be spent on novel designs rather than optimizations of already-known designs.

The impact of this thesis work is also evident in the numerous academic awards received during the course of the program: NSERC Postgraduate Scholarship (Doctoral), Ontario Graduate Scholarship, Queen Elizabeth II Scholarship, Ed Ireland Award, Graduate Student Open Access Award, and CURIE Research Impact Endeavour.

6.2 Achievements

The following articles have been published in journals as a result of the work done for this thesis:

7. D. Gostimirovic and W. N. Ye, "A compact silicon-photonic mode-division (de)multiplexer using waveguide-wrapped microdisk resonators," Opt. Lett. Submitted for publication, (October 2020).
8. D. Gostimirovic and W. N. Ye, "Ultralow-power double vertical junction microdisk modulators," IEEE J. Quantum Electron. In press, (2020).
9. D. Gostimirovic, F. De Leonardis, R. Soref, V. M. N. Passaro, and W. N. Ye, "Ultrafast electro-optical disk modulators for logic, communications, optical repeaters, and wavelength converters," Opt. Express 28(17), 24874–24888 (2020).
10. D. Gostimirovic and W. N. Ye, "An open-source artificial neural network model for polarization-insensitive silicon-on-insulator subwavelength grating couplers," IEEE J. Quantum Electron. 25(3), (2018).

11. D. Gostimirovic and W. N. Ye, “Ultracompact CMOS-compatible optical logic using carrier depletion in microdisk resonators,” Sci. Rep. 7(1), (2017).

The following articles have been presented and published in conferences as a result of the work done for this thesis:

1. D. Gostimirovic and W. N. Ye, “A comparison of microresonator devices for WDM-compatible mode-division multiplexing,” Proc. SPIE 11284, 112841E (2020).
2. D. Gostimirovic and W. N. Ye, “Design and optimization of pn junctions in silicon microdisk modulators,” in Proceedings of PhotonIcs & Electromagnetics Research Symposium-Spring, (IEEE, 2019), pp. 3679–3682.
3. D. Gostimirovic and W. N. Ye, “Automating photonic design with machine learning,” in Proceedings of IEEE 15th International Conference on Group IV Photonics, (IEEE, 2018).
4. D. Gostimirovic and W. N. Ye, “CMOS-compatible optical AND, OR, and XOR gates using voltage-induced free-carrier dispersion and stimulated Raman scattering,” Proc. SPIE 10108, 101080Q (2017).
5. D. Gostimirovic and W. N. Ye, “Ultrafast all-optical arithmetic logic based on hydrogenated amorphous silicon microring resonators,” Proc. SPIE 9752, 97520X (2016).
6. D. Gostimirovic and W. N. Ye, “All-optical half-adder based on a hydrogenated amorphous silicon microring resonator,” in Proceedings of Photonics North, (IEEE, 2015).

6.3 Future Work

There are many directions to take when considering improvements to the works presented in this thesis. Chapter 6.3.1 presents the recommended future works for the design of active integrated photonic circuits for next-generation optical processing and computing, Chapter 6.3.2 presents the recommended future works for the design of photonic integrated circuits for next-generation on-chip optical communications, and Chapter 6.3.3 presents the recommended future works for the automated and accelerated design of next-generation photonic integrated circuits with machine learning.

6.3.1 Design of Next-Generation Optical Processing and Computing Circuits

Chapter 2 presented the fundamentals of the proposed optoelectronic logic architecture. The physics of the fundamental logic cell were studied and characterized, and the high-level design of logic gates and combinational logic circuits are designed. Many improvements can be made at the device and architectural levels. In Chapter 3, the double vertical and staggered double vertical modulator designs were introduced and shown to decrease energy consumption by over a factor of 4. Applying these designs to the logic cell would not only bring their performance levels closer to CMOS gates, but they open up potential structural advantages as well. At the simplest level, the demonstrated logic cell must be split into two pn diodes on either lateral side of the microdisk to account for both logic inputs. With the double junction designs, two independent junctions can be naturally formed without splitting, as shown in Figure 6.1. Although this would require tighter contact integration, with the potential for added capacitance, this has the benefit of using the entire circumference of the disk for pn modulation, as the nominal design has an

undoped region that is wasted in modulation. The lateral splitting of the pn diodes may still be useful, however, as the 2-disk 4-bit AND/OR gate may be replaced by a single double-junction disk split into four independent junctions. This design flexibility may be applied to other higher-order logic functions for the further optimization of chip space.

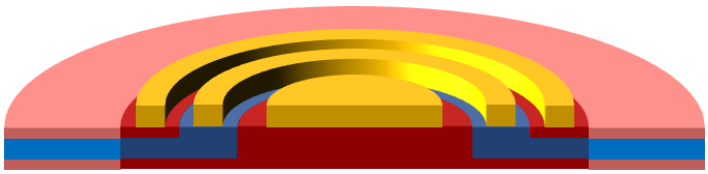


Figure 6.1 Optional contacting configuration for the double vertical junction microdisk modulator. Each n region (red) can be controlled by a separate logic input to naturally support logic operations without splitting the junctions laterally.

At the beginning of Chapter 2, simple, low-level comparisons were made to CMOS logic and other demonstrations of optical logic in the literature. Especially for CMOS, a direct comparison is difficult, as there are many external factors, such as the interconnection with the rest of the circuit, that would have to be considered. To gain a better picture of the potential of optical logic, the entire system must be considered in the comparison. To do so, a study can be made on which parts of the chip can be replaced by photonic components. The first and perhaps most important replacement would be the optical interconnects, which would reduce latency, power consumption, spatial footprint, and heat generation. From there, select circuit blocks like the arithmetic and logic unit (which many of the circuits presented here would fit into), would be replaced by an optical logic equivalent that would fit seamlessly into the optical interconnect network. Given similar performance

to CMOS, a significant performance increase is expected simply due to the benefits of the optical interconnects and the elimination of OEO conversions.

6.3.2 Design of Next-Generation Optical Communications Circuits

Despite the high performance achieved here, in simulation and experiment, there is still much room for device optimization and novel designs to further improve the total aggregate bandwidth of these circuits. The fabricated circuit only used two modes, but the simulations show that three and possibly more modes can realistically be used. Simply adding more mode orders (disks) would be the simplest path to increasing aggregate bandwidth.

As stated in the previous sections, the multimode whispering gallery nature of microdisk resonators creates multiple unwanted peaks within the FSR, which reduces the number of wavelength channels and, ultimately, bandwidth. These devices can be redesigned for disk radii as low as 1.5 μm [107], which would not only increase the FSR but eliminate higher-order modes altogether as it would be too small to support them.

For smaller microdisks, there is less circumference to wrap the waveguides around and create low IL and XT. Because only the higher-order modes require large degrees of wrapping, a design like what is shown in Figure 6.2 may perform better. Here, the TE_0-mode waveguide does not wrap around the disk, as it is easy to achieve critical coupling for it. This frees up more of the circumference of the disk to wrap the higher-order-mode waveguide around. Not only does this increase the degree of coupling, but it reduces the

loss from the modal mismatch between the straight and curved segments of the wrapped waveguide used in the nominal design.

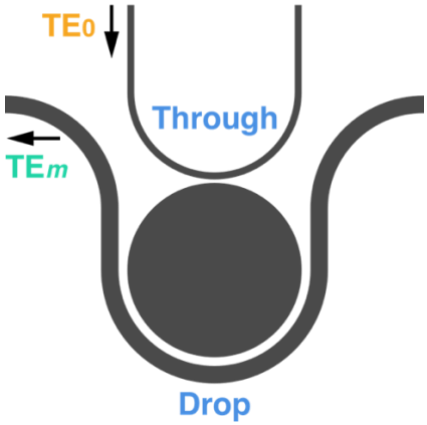


Figure 6.2 Possible configuration for the waveguide-wrapped microdisk resonator to increase the degree of bending (coupling strength) and reduce loss induced by the curved multimode waveguide.

6.3.3 Automated and Accelerated Design of Photonic Integrated Circuits with Machine Learning

This work presents an openly available framework that automates and accelerates the design of standard photonic devices using artificial neural networks. A compact, flexible model of a polarization insensitive subwavelength grating coupler is presented as an example, but any device can be modelled using the same process. Future work on this topic would aim to build a library of standard photonic components (e.g., microring/disk resonators, Y-junctions, Bragg gratings) that can be made openly available for future designers to access. Given the specific requirements of each design project, the flexible ANN models can rapidly produce the device designs that satisfy the design requirements of the chip. A large compute cost would be taken to acquire the initial training data, but this would be for the greater benefit of the acceleration of many future design projects.

Design acceleration would not only come in the form of reduced computational load, but in the initial cost needed for the designer to understand the underlying physics that describe the device's operation.

With a complete library of ANN models, it would not be difficult to create a system-level circuit solver, similar to Lumerical INTERCONNECT or PSpice. The compact models would not only allow for system-level simulations in reasonable timeframes, but the flexibility of the components would allow for rapid system-level optimizations that are otherwise unfeasible using current design approaches. Although significant research efforts have been made on optimizing individual photonic devices, much fewer have been made at the system level. As silicon photonics reaches larger levels of scale and implementation, singular device optimization may not necessarily produce overall performance improvements if they do not seamlessly fit with their surrounding components. One can imagine a system-level optimization that uses flexible ANN models to find the best performance when factors such as fabrication nonuniformities, chip-level thermal variances, and simple device-to-device interactions are in place.

For robust ANN models, which have many features and labels, it can take many thousands of simulation examples to accurately model. For devices like microring resonators, which can take hours to simulate, the data acquisition process may be unfeasibly long. Although one could take advantage of hyperscale cloud computing to distribute the load and acquire the data quicker, there may be some devices that still require too much data to accurately model. Despite the high sophistication of modern ANN training algorithms, they still rely

on large sets of data that can be difficult and costly to obtain. Work has been done on developing new training techniques that achieve high accuracies with less data [126, 127] but they have not been applied to the relatively new field of ANN-based photonic device design. Pre-training optimization algorithms can be used during data acquisition to only collect data that is meaningful to training. This can mean simultaneously training the model while acquiring data and steering the acquirer to the areas of the feature space that produce more meaningful results. This can be determined by how the training gradient responds to certain data. Other smart acquisition approaches can be made without the training entirely. Genetic algorithms like particle-swarm optimization can be used to acquire data from generally well-performing areas of the feature space based on one or two globally relevant figures of merit. For the example of a grating coupler, center wavelength, polarization, and bandwidth may change between applications, but coupling efficiency will always be important. The smart data acquisition algorithm can bias the scan towards devices that produce high coupling efficiencies, and the modeller can generalize based only off of that (mostly ignoring poorly performing devices).

The final device design (inferencing) stage may also take a long time for big, robust models. The example shown in this work inferences at the millisecond timescale, but much larger models can take longer. For large, system-level optimizations, these can add up to a significant amount. The example used in this work designs the device using a simple parameter scan that searches the entire feature space for a specific figure of merit. Smart genetic algorithms and gradient descent can be easily plugged in with the basic inference math to find designs with much fewer feed-forward iterations. Alternatively, because the

ANN is bidirectional, one could input the desired performance parameters at the output and feed backward to produce the corresponding design parameters. Work that has already been done here has shown that accuracy is low for this approach, as many designs can produce the same result, which the ANN struggles to map [115, 116]. However, tandem forward–backward networks have been able to increase accuracy of this approach mildly. Other demonstrations remove identical designs, but this adds bias to the model and reduces generalization. Although it is unclear if there is a specific ANN structure that could accurately perform this inverse modelling, it is clear that this would be the best approach to design, as the device is directly created from the desired performance parameters.

References

[1] Cisco Systems, Inc., "Cisco Annual Internet Report," [Online]. Available: https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html#~insights. [Accessed 15 September 2020].

[2] Cisco Systems, Inc., "Cisco Global Cloud Index: Forecast and Methodology, 2016–2021," 2018.

[3] E. Masanet, A. Shehabi, N. Lei, S. Smith and J. Koomey, "Recalibrating global data center energy-use estimates," *Science,* vol. 367, no. 6481, pp. 984-986, 2020.

[4] M. M. Waldrop, "More than Moore," *Nature,* vol. 530, no. 7589, pp. 144-148, 2016.

[5] P. Ye, T. Ernst and M. V. Khare, "The Last Silicon Transistor," *IEEE Spectrum,* pp. 30-35, August 2019.

[6] i-Micronews Media, "Silicon Photonics 2020," April 2020. [Online]. Available: https://www.i-micronews.com/products/silicon-photonics-2020/. [Accessed 15 September 2020].

[7] J. Bahram and S. Fathpour, "Silicon photonics," *IEEE J. Lightw. Technol.,* vol. 24, no. 12, pp. 4600-4615, 2006.

[8] G. T. Reed and A. P. Knights, Silicon photonics: an introduction, John Wiley & Sons, 2004.

[9] D. Thomson, A. Zilkie, J. E. Bowers, T. Komljenovic, G. T. Reed, L. Vivien, D. Marris-Morini, E. Cassan, L. Virot, J. M. Fédéli and J. M. Hartmann, "Roadmap on silicon photonics," *J. Opt.,* vol. 18, no. 7, 2016.

[10] R. Soref, "The past, present, and future of silicon photonics," *IEEE J. Sel. Topics Quantum Electron.,* vol. 12, no. 6, pp. 1678-1687, 2006.

[11] R. R. Schaller, "Moore's law: past, present and future," *IEEE Spectrum,* p. 1997, June 1997.

[12] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir and V. and Narayanan, "Leakage current: Moore's law meets static power," *Computer,* vol. 36, no. 12, pp. 68-75, 2003.

[13] D. A. B. Miller, "Attojoule optoelectronics for low-energy information processing and communications," *J. Light. Technol,* vol. 35, no. 3, pp. 346-396, 2017.

[14] A. Steane, "Quantum computing," *Rep. Prog. Phys.,* vol. 61, no. 2, 1998.

[15] L. Gyongyosi and S. Imre, "A survey on quantum computing technology," *Comp. Sci. Rev.,* vol. 31, pp. 51-71, 2019.

[16] D. E. Nikonov and I. A. Young, "Overview of beyond-CMOS devices and a uniform methodology for their benchmarking," *Proc. IEEE,* vol. 101, no. 12, pp. 2498-2533, 2013.

[17] J. Sun, R. Kumar, M. Sakib, J. B. Driscoll, H. Jayatilleka and H. Rong, "A 128 Gb/s PAM4 silicon microring modulator with integrated thermo-optic resonance tuning," *J. Lightw. Technol.,* vol. 37, no. 1, pp. 110-115, 2019.

[18] Q. Xu and R. Soref, "Reconfigurable optical directed-logic circuits using microresonator-based optical switches," *Opt. Express,* vol. 19, no. 6, pp. 5244-5259, 2011.

[19] C. Qiu, W. Gao, R. Soref, J. T. Robinson and Q. Xu, "Reconfigurable electro-optical directed-logic circuit using carrier-depletion micro-ring resonators," *Opt. Lett.,* vol. 39, no. 24, pp. 6767-6770, 2014.

[20] L. Zhang, R. Ji, L. Jia, L. Yang, P. Zhou, Y. Tian, P. Chen, Y. Lu, Z. Jiang, Y. Liu, Q. Fang and M. Yu, "Demonstration of directed XOR/XNOR logic gates using two cascaded microring resonators," *Opt. Lett.,* vol. 35, no. 10, pp. 1620-1622, 2010.

[21] L. Zhang, J. Ding, Y. Tian, R. Ji, L. Yang, H. Chen, P. Zhou, Y. Lu, W. Zhu and R. Min, "Electro-optic directed logic circuit based on microring resonators for XOR/XNOR operations," *Opt. Express,* vol. 20, no. 11, pp. 11605-11614, 2012.

[22] L. Yang, L. Zhang, C. Guo and J. Ding, "XOR and XNOR operations at 12.5 Gb/s using cascaded carrier-depletion microring resonators," *Opt. Express,* vol. 22, no. 3, pp. 2996-3012, 2014.

[23] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund and M. Soljačić, "Deep Learning with Coherent Nanophotonic Circuits," *Nat. Photonics,* vol. 11, pp. 441-446, 2017.

[24] A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.,* vol. 7, no. 7430, 2017.

[25] C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran and W. H. P. Pernice, "Integrated all-photonic non-volatile multi-level memory," *Nat. Photonics,* vol. 9, pp. 725-732, 2015.

[26] P. Rani, Y. Kalra and R. K. Sinha, "Realization of AND gate in Y shaped photonic crystal waveguide," *Opt. Commun.,* no. 298-299, pp. 227-231, 2013.

[27] H. Wang, X. Yu and X. Rong, "All-optical AND, XOR, and NOT logic gates based on Y-branch photonic crystal waveguide," in *SPIE Optical Engineering Conference*, 2015.

[28] G. T. Reed, G. Mashanovich, F. Y. Gardes and D. J. Thomson, "Silicon optical modulators," *Nat. Photonics,* vol. 4, pp. 518-526, 2010.

[29] Y. Tian, Z. Liu, H. Xiao, G. Zhao, G. Liu, J. Yang, J. Ding, L. Zhang and L. Yang, "Experimental demonstration of a reconfigurable electro-optic directed logic circuit using cascaded carrier-injection micro-ring resonators," *Sci. Rep.,* vol. 7, no. 6410, 2017.

[30] R. W. Boyd, Nonlinear Optics, Academic Press, 2019.

[31] C. Manolatou and M. Lipson, "All-optical silicon modulators based on carrier injection by two-photon absorption," *J. Light. Technol.,* vol. 24, no. 3, pp. 1433-1439, 2006.

[32] T. K. Liang, L. R. Nunes, M. Tsuchiya, K. S. Abedin, T. Miyazaki, D. Van Thourhout, W. Bogaerts, P. Dumon, R. Baets and H. K. Tsang, "High speed logic gate using two-photon absorption in silicon waveguides," *Opt. Commun.,* vol. 265, pp. 171-174, 2006.

[33] M. He, M. Xu, Y. Ren, J. Jian, Z. Ruan, Y. Xu, S. Gao, S. Sun, X. Wen, L. Zhou, L. Liu, C. Guo, H. Chen, S. Yu, L. Liu and X. Cai, "High-performance hybrid silicon

and lithium niobate Mach–Zehnder modulators for 100 Gbit s−1 and beyond," *Nat. Photonics,* vol. 13, pp. 359-364, 2019.

[34] C. Wang, M. Zhang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer and M. Lončar, "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature,* vol. 562, pp. 101-104, 2018.

[35] B. Chmielak, M. Waldow, C. Matheisen, C. Ripperda, J. Bolten, T. Wahlbrink, M. Nagel, F. Merget and H. Kurz, "Pockels effect based fully integrated, strained silicon electro-optic modulator," *Opt. Express,* vol. 19, no. 18, pp. 17212-17219, 2011.

[36] C. Lafforgue, M. Berciano, L. Deniel, G. Marcaud, X. Le Roux, C. Alonso-Ramos, D. Benedikovic, V. Vakarin, A. Ruiz-Caridad, P. Crozat, D. Marris-Morini, E. Cassan and L. Vivien, "High speed logic gate using two-photon absorption in siliconwaveguides," in *SPIE Silicon Photonics XV Conference*, 2020.

[37] J. S. Pelc, L. Rivoire, S. Vo, C. Santori, D. A. Fattal and R. G. Beausoleil, "Picosecond all-optical switching in hydrogenated amorphous silicon microring resonators," *Opt. Express,* vol. 22, no. 4, pp. 3797-3810, 2014.

[38] N. Vukovic, N. Healy, F. H. Suhailin, P. Mehta, T. D. Day, J. V. Badding and A. C. Peacock, "Ultrafast optical control using the Kerr nonlinearity in hydrogenated amorphous silicon microcylindrical resonators," *Sci. Rep.,* vol. 3, no. 2885, 2013.

[39] K. Ikeda, R. E. Saperstein, N. Alic and Y. Fainman, "Thermal and Kerr nonlinear properties of plasma-deposited silicon nitride/ silicon dioxide waveguides," *Opt. Express,* vol. 16, no. 17, pp. 12987-12994, 2008.

[40] W. Yoshiki and T. Tanabe, "All-optical switching using Kerr effect in a silica toroid microcavity," *Opt. Express,* vol. 22, no. 20, pp. 24332-24341, 2014.

[41] Y. H. Wen, O. Kuzucu, M. Fridman, A. L. Gaeta, L.-W. Luo and M. Lipson, "All-optical control of an individual resonance in a silicon microresonator," *Phys. Rev. Lett.,* vol. 108, no. 223907, 2012.

[42] J. Hou, L. Chen, W. Dong and X. Zhang, "40 Gb/s reconfigurable optical logic gates based on FWM in silicon waveguide," *Opt. Express,* vol. 24, no. 3, pp. 2701-2711, 2016.

[43] F. Li, T. D. Vo, C. Husko, M. Pelusi, D.-X. Xu, A. Densmore, R. Ma, S. Janz, B. J. Eggleton and D. J. Moss, "Silicon nanowire based exclusive-OR gate using nonlinear optics for 40Gb/s DPSK signals," *Opt. Express,* vol. 19, no. 21, pp. 20364-20371, 2011.

[44] C. Gui and J. Wang, "Simultaneous optical half-adder and half-subtracter sing a single-slot waveguide," *IEEE Photon. J.,* vol. 5, no. 5, 2013.

[45] L. Zhang, R. Ji, Y. Tian, L. Yang, P. Zhou, Y. Lu, W. Zhu, Y. Liu, L. Jia, Q. Fang and M. Yu, "Simultaneous implementation of XOR and XNOR operations using a directed logic circuit based on two microring resonators," *Opt. Express,* vol. 19, no. 7, pp. 6524-6540, 2011.

[46] Z. Ying, Z. Wang, Z. Zhao, S. Dhar, D. Z. Pan, R. Soref and R. T. Chen, "Silicon microdisk-based full adders for optical computing," *Opt. Lett.,* vol. 43, no. 5, pp. 983-986, 2018.

[47] T. Fjelde, D. Wolfson, A. Kloch, B. Dagens, A. Coquelin, A. Guillemot, F. Gaborit, F. Poingt and M. Renaud, "Demonstration of 20 Gbit/s all-optical logic XOR in integrated SOA-based interferometric wavelength converter," *Electron. Lett.,* vol. 36, no. 22, pp. 1863-1864, 2000.

[48] Z. Li, Y. Liu, S. Zhang, H. Ju, H. de Waardt, G. D. Khoe, H. J. S. Dorren and D. Lenstra, "All-optical logic gates using semiconductor optical amplifier assisted by optical filter," *Electron. Lett.,* vol. 41, no. 25, pp. 1397-1399, 2005.

[49] T. Birr, U. Zywietz, P. Chhantyal, B. N. Chichkov and C. Reinhardt, "Ultrafast surface plasmon-polariton logic gates and half-adder," *Opt. Express,* vol. 23, no. 25, pp. 31755-31765, 2015.

[50] W. Su and Z. Geng, "Terahertz all-optical logic gates based on a graphene nanoribbon rectangular ring resonator," *IEEE Photon. J.,* vol. 10, no. 6, 2018.

[51] Y. Tian, L. Zhang, J. Ding and L. Yang, "Demonstration of electro-optic half-adder using silicon photonic integrated circuits," *Opt. Express,* vol. 22, no. 6, pp. 6958-6965, 2014.

[52] R. Soref, F. De Leonardis, Z. Ying, V. M. N. Passaro and R. T. Chen, "Silicon-Based Group-IV O-E-O Devices for Gain, Logic, and Wavelength Conversion," *ACS Photonics,* vol. 7, pp. 800-811, 2020.

[53] Lumerical, Inc., "3D Charge Transport Simulator," [Online]. Available: https://www.lumerical.com/products/charge/. [Accessed 14 September 2020].

[54] Lumerical, Inc., "3D/2D Maxwell's Solver for Nanophotonic Devices," [Online]. Available: https://www.lumerical.com/products/fdtd/. [Accessed 14 September 2020].

[55] E. Timurdogan, C. M. Sorace-Agaskar, J. Sun, E. S. Hosseini, A. Biberman and M. R. Watts, "An ultralow power athermal silicon modulator," *Nat. Commun,* vol. 5, no. 4008, 2014.

[56] M. R. Watts, W. A. Zortman, D. C. Trotter, R. W. Young and A. L. Lentine, "Vertical junction silicon microdisk modulators and switches," *Opt. Express,* vol. 19, no. 22, pp. 21989-22003, 2011.

[57] J. Hardy and J. Shamir, "Optics inspired logic architecture," *Opt. Express,* vol. 15, no. 1, pp. 150-165, 2007.

[58] Q. Xu and M. Lipson, "All-optical logic based on silicon micro-ring resonators," *Opt. Express,* vol. 15, no. 3, pp. 924-929, 2007.

[59] P. Sethi and S. Roy, "All-optical ultrafast XOR/XNOR logic gates, binary counter, and double-bit comparator with silicon microring resonators," *Applied Optics,* vol. 53, no. 28, pp. 6527-6536, 2014.

[60] S. Lin, Y. Ishikawa and K. Wada, "Demonstration of optical computing logics based on binary decision diagram," *Opt. Express,* vol. 20, no. 2, pp. 1378-1384, 2012.

[61] E. Jaberamsary, T. M. B. Masaud, M. M. Milosevic, M. Nedeljkovic, G. Z. Mashanovic and H. M. H. Chong, "Scattering loss estimation using 2-D Fourier analysis and modeling of sidewall roughness on optical waveguides," *IEEE Photon. J.,* vol. 5, no. 3, 2013.

[62] Z. Ying, Z. Wang, Z. Zhao, S. Dhar, D. Z. Pan, R. Soref and R. T. Chen, "Comparison of microrings and microdisks for high-speed optical modulation in silicon photonics," *Appl. Phys. Lett.,* vol. 112, no. 11, 2018.

[63] Z. Yong, W. D. Sacher, Y. Huang, J. C. Mikkelsen, Y. Yang, X. Luo, P. Dumais, D. Goodwill, H. Bahrami, P. G.-Q. Lo, E. Bernier and J. K. S. Poon, "U-shaped PN junctions for efficient silicon Mach-Zehnder and microring modulators in the O-band," *Opt. Express,* vol. 25, no. 7, pp. 8425-8439, 2017.

[64] P. Dong, S. Liao, D. Feng, H. Liang, D. Zheng, R. Shafiiha, C.-C. Kung, W. Qian, G. Li, X. Zheng, A. V. Krishnamoorthy and M. Asghari, "Low Vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator," *Opt. Express,* vol. 17, no. 25, pp. 22484-22490, 2009.

[65] G. Li, X. Zheng, J. Yao, H. Thacker, I. Shubin, Y. Luo, K. Raj, J. E. Cunningham and A. V. Krishnamoorthy, "25Gb/s 1V-driving CMOS ring modulator with integrated thermal tuning," *Opt. Express,* vol. 19, no. 21, pp. 20435-20443, 2011.

[66] F. Y. Gardes, A. Brimont, P. Sanchis, G. Rasigade, D. Marris-Morini, L. O'Faolain, F. Dong, J. M. Fedeli, P. Dumon, L. Vivien, T. F. Krauss, G. T. Reed and J. Martí, "High-speed modulation of a compact silicon ring resonator based on a reverse-biased pn diode," *Opt. Express,* vol. 17, no. 24, pp. 21986-21991, 2009.

[67] J.-B. You, M. Park, J.-W. Park and G. Kim, "12.5 Gbps optical modulation of silicon racetrack resonator based on carrier-depletion in asymmetric p-n diode," *Opt. Express,* vol. 16, no. 22, pp. 18340-18344, 2008.

[68] X. Xiao, H. Xu, X. Li, Y. X. K. Hu, Z. Li, T. Chu, Y. Yu and J. Yu, "25 Gbit/s silicon microring modulator based on misalignment-tolerant interleaved PN junctions," *Opt. Express,* vol. 20, no. 3, pp. 2507-2515, 2012.

[69] M. Pantouvaki, H. Yu, M. Rakowski, P. Christie, P. Verheyen, G. Lepage, N. Van Hoovels, P. Absil and J. Van Campenhout, "Comparison of silicon ring modulators with interdigitated and lateral p-n junctions," *IEEE J. Sel. Topics Quantum Electron.,* vol. 19, no. 2, 2013.

[70] L. Alloatti, D. Cheian and R. J. Ram, "High-speed modulator with interleaved junctions in zero-change CMOS photonics," *Appl. Phys. Lett.,* vol. 108, no. 131101, 2016.

[71] J. C. Rosenberg, W. M. J. Green, S. Assefa, D. M. Gill, T. Barwicz, M. Yang, S. M. Shank and Y. A. Vlasov, "A 25 Gbps silicon microring modulator based on an interleaved junction," *Opt. Express,* vol. 20, no. 24, pp. 26411-26423, 2012.

[72] Synopsys, "Taurus TSUPREM-4: An advanced 1D and 2D process simulator," [Online]. Available: https://www.synopsys.com/silicon/tcad/process-simulation/taurus-tsuprem-4.html.

[73] E. S. Hosseini, S. Yegnanarayanan, A. H. Atabaki, M. Soltani and A. Adibi, "Systematic design and fabrication of high-Q single-mode pulley-coupled planar silicon nitride microdisk resonators at visible wavelengths," *Opt. Express,* vol. 18, no. 3, pp. 2127-2136, 2010.

[74] D. Gostimirovic and W. N. Ye, "A comparison of microresonator devices for WDM-compatible mode-division multiplexing," in *SPIE OPTO Smart Photonic and Optoelectronic Integrated Circuits XXII*, San Francisco, 2020.

[75] Z. Ying, C. Feng, Z. Zhao, R. Soref, D. Pan and R. T. Chen, "Integrated multi-operand electro-optic logic gates for optical computing," *Appl. Phys. Lett.,* vol. 115, no. 17, 2020.

[76] Z. Ying, C. Feng, Z. Zhao, S. Dhar, H. Dalir, J. Gu, Y. Cheng, R. Soref, D. Z. Pan and R. T. Chen, "Electronic- photonic arithmetic logic unit for high-speed computing," *Nat. Commun.,* vol. 11, no. 1, 2020.

[77] D. Gostimirovic and W. N. Ye, "Ultracompact CMOS-compatible optical logic using carrier depletion in microdisk resonators," *Sci. Rep.,* vol. 7, no. 1, 2017.

[78] L. Y. Dai, Y. H. Hung, Q. Cheng and K. Bergman, "Experimental demonstration of PAM-4 transmission through microring silicon photonic Clos switch fabric," in *Optical Fiber Communication Conference (OFC)*, 2020.

[79] Z. Wang, P. Yang, Y. S. Chang, J. Xu, X. Chen, Z. Wang and J. Feng, "A cross-layer optimization framework for integrated optical switches in data centers," *IEEE Trans. on Comput-Aided Des. Integr.,* vol. 39, no. 3, pp. 640-653, 2019.

[80] T. Barwicz, M. A. Popovic, F. Gan, M. S. Dahlem, C. W. Holzwarth, P. T. Rakich, E. P. Ippen, F. X. Kärtner and H. I. Smith, "Reconfigurable silicon photonic circuits for telecommunication applications," in *SPIE Laser Resonators and Beam Control*, 2008.

[81] D. Wu, Y. Wu, Y. Wang, J. An and X. Hu, "Reconfigurable optical add-drop multiplexer based on thermally tunable micro-ring resonators," *Opt. Commun.,* vol. 367, pp. 44-49, 2016.

[82] N. Parsons, A. Hughes and R. Jensen, "High radix all-optical switches for software-defined datacenter networks," in *42nd European Conference on Optical Communication*, 2016.

[83] N. Calabretta, W. Miao, K. Prifti and K. Williams, "System performance assessment of a monolithically integrated WDM cross-connect switch for optical data centre networks," in *42nd European Conference on Optical Communication*, 2016.

[84] Z. Wang, J. Xu, P. Yang, Z. Wang, L. H. K. Duong and X. Chen, "High-radix nonblocking integrated optical switching fabric for data center," *J. Lightw. Technol.,* vol. 35, no. 19, pp. 4268-4281, 2017.

[85] Y. Huang, Q. Cheng, Y. H. Hung, H. Guan, X. Meng, A. Novack, M. Streshinsky, M. Hochberg and K. Bergman, "Multi-stage 8x8 silicon photonic switch based on dual-microring switching elements," *J. Lightw. Technol.,* vol. 38, no. 2, pp. 194-201, 2020.

[86] T. Cao, Y. Fei, L. Zhang, Y. Cao and S. Chen, "Design of a silicon Mach–Zehnder modulator with a U-type PN junction," *Appl. Opt.,* vol. 52, no. 24, pp. 5941-5948, 2013.

[87] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.,* vol. 23, no. 1, pp. 123-129, 1987.

[88] M. Pantouvaki, S. A. Srinivasan, Y. Ban, P. De Heyn, P. Verheyen, G. epage, H. Chen, J. De Coster, N. Golshani, S. Balakrishnan, P. Absil and J. Van Campenhout, "Active components for 50 Gb/s NRZ-OOK Optical Interconnects in a solicon photonics platform," *J. Lightw. Technol.,* vol. 35, no. 4, pp. 631-638, 2017.

[89] H. Ramon, M. Vanhoecke, J. Verbist, W. Soenen, P. De Heyn, Y. Ban, M. Pantouvaki, J. Van Campenhout, P. Ossieur, X. Yin and J. Bauwelinck, "Low-power 56 Gb/s NRZ microring modulator driver in 28 nm FDSOI CMOS," *IEEE Photon. Technol. Lett.,* vol. 30, no. 5, pp. 467-470, 2019.

[90] R. Dubé-Demers, S. LaRochelle and W. Shi, "Ultrafast pulse-amplitude modulation with a femtojoule silicon photonic modulator," *Optica,* vol. 3, no. 6, pp. 622-627, 2016.

[91] Yole Développement, "Silicon photonics and photonics integrated circuits 2019," i-Micronews, 2019.

[92] C. Sun, M. T. Wade, Y. Lee, J. S. Orcutt, L. Alloatti, M. S. Georgas, A. S. Waterman, J. M. Shainline, R. R. Avizienis, S. Lin, B. R. Moss, R. Kumar, F. Pavanello, A. H. Atabaki, H. M. Cook, A. J. Ou, J. C. Leu, Y.-H. Chen, K. Asanović, R. J. Ram and M. A. Popović, "Single-chip microprocessor that communicates directly using light," *Nature,* vol. 528, no. 7583, pp. 534-538, 2015.

[93] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. W. Holzwarth, M. A. Popović, H. Li, H. I. Smith, J. L. Hoyt, F. X. Kärtner, R. J. Ram, V. Stojanović and K. Asanović, "Building many-core processor-to-DRAM with monolithic CMOS silicon photonics," *IEEE Micro,* vol. 29, no. 4, pp. 8-21, 2009.

[94] A. V. Krishnamoorthy, R. Ho, X. Zheng, H. Schwetman, J. Lexau, P. Koka, G. Li, I. Shubin and J. E. Cunningham, "Computer systems based on silicon photonic interconnects," *Proc. IEEE,* vol. 97, no. 7, pp. 1337-1361, 2009.

[95] C. Gunn, "CMOS photonics for high-speed interconnects," *IEEE Micro,* vol. 26, no. 2, pp. 58-66, 2006.

[96] L.-W. Luo, N. Ophir, C. P. Chen, L. H. Gabrielli, C. B. Poitras, K. Bergmen and M. Lipson, "WDM-compatible mode-division multiplexing on a silicon chip," *Nat. Commun.,* vol. 5, no. 3069, 2014.

[97] H. Jia, X. Fu, T. Zhou, L. Zhang, S. Yang and L. Yang, "Mode-selective modulation by silicon microring resonators and mode multiplexers for on-chip optical interconnect," *Opt. Express,* vol. 29, no. 4, pp. 2915-2925, 2019.

[98] M. Ye, Y. Yu, G. Chen, Y. Luo and X. Zhang, "On-chip WDM mode-division multiplexing interconnection with optional modulation function," *Opt. Express,* vol. 23, no. 25, pp. 32130-32138, 2015.

[99] B. Dorin and W. N. Ye, "Two-mode division multiplexing in a silicon-on-insulator ring resonator," *Opt. Express,* vol. 22, no. 4, pp. 4547-4558, 2014.

[100] Y. Ding, J. Xu, F. Da Ros, B. Huang, H. Ou and C. Peucheret, "On-chip two-mode division multiplexing using tapered directional coupler-based mode multiplexer and demultiplexer," *Opt. Express,* vol. 21, no. 8, pp. 10376-10382, 2018.

[101] J. B. Driscoll, R. R. Grote, B. Souhan, J. I. Dadap, M. Lu and R. M. Osgood Jr., "Asymmetric Y junctions in silicon waveguides for on-chip mode-division multiplexing," *Opt. Lett.,* vol. 38, no. 11, pp. 1854-1856, 2013.

[102] J. Wang, S. He and D. Dai, "On-chip silicon 8-channel hybrid (de)multiplexer enabling simultaneous mode- and polarization-division-multiplexing," *Laser Photon Rev.,* vol. 8, no. 2, pp. L18-L22, 2014.

[103] B. Stern, X. Zhu, C. P. Chen, L. D. Tzuang, J. Cardenas, K. Bergmen and M. Lipson, "On-chip mode-division multiplexing switch," *Optica,* vol. 2, no. 6, pp. 530-535, 2015.

[104] T. Uematsu, Y. Ishizaka, Y. Kawaguchi, K. Saitoh and M. Koshiba, "Design of a two-mode multi/demultiplexer consisting of multimode interference waveguides and a wavelength-insensitive phase shifter for mode-division multiplexing transmission," *J. Light. Technol.,* vol. 30, no. 15, pp. 2421-2426, 2012.

[105] D. Gostimirovic and W. N. Ye, "Design and optimization of pn junctions in silicon microdisk modulators," in *41st PIERS*, Rome, 2019.

[106] Lumerical, Inc., "Ring resonator getting started - Design and initial simulation," [Online]. Available: https://support.lumerical.com/hc/en-us/articles/360042800293-Ring-resonator-getting-started-Design-and-initial-simulation. [Accessed 2 April 2020].

[107] W. Shi, H. Yun, W. Zhang, C. Lin, T. K. Chang, Y. Wang, N. A. Jaeger and L. Chrostowski, "Ultra-compact, high-Q silicon microdisk reflectors," *Opt. Express,* vol. 20, no. 20, pp. 21840-21846, 2012.

[108] KLayout, [Online]. Available: https://klayout.de. [Accessed 30 September 2020].

[109] Mentor Graphics, "Calibre," [Online]. Available: https://www.mentor.com/training/course_categories/calibre. [Accessed 30 September 2020].

[110] Lumerical, Inc., "Photonic Model Development Kit," [Online]. Available: https://www.lumerical.com/products/cml-compiler/. [Accessed 2 April 2020].

[111] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.,* vol. 2, no. 5, pp. 359-366, 1989.

[112] Q.-J. Zhang, K. Gupta and V. Devabhaktuni, "Artificial neural networks for RF and microwave design—from theory to practice," *IEEE Trans. Microw. Theory Tech.,* vol. 51, no. 4, pp. 1339-1350, 2003.

[113] A. Hafid Zaabab, Q.-J. Zhang and M. Nakhla, "Device and circuit-level modeling using neural networks with faster training based on network sparsity," *IEEE Trans. Microw. Theory Tech.,* vol. 45, no. 10, pp. 1696-1704, 1997.

[114] H. Kabir, Y. Wang, M. Yu and Q.-J. Zhang, "Neural network inverse modeling and applications to microwave filter design," *IEEE Trans. Microw. Theory Tech.,* vol. 54, no. 4, pp. 867-879, 2008.

[115] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteira, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark and M. Soljačić, "Nanophotonic particle simulation and inverse design using artificial neural networks," *Sci. Adv.,* vol. 4, no. 6, 2018.

[116] D. Liu, Y. Tn, E. Khoram and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics,* vol. 5, no. 4, pp. 1365-1369, 2018.

[117] L. Su, R. Trivedi, N. V. Sapara, A. Y. Piggott, D. Vercruysse and J. Vučković, "Fully-automated optimization of grating couplers," *Opt. Express,* vol. 26, no. 4, pp. 4023-4034, 2018.

[118] Lumerical, Inc., "Optimization utility," [Online]. Available: https://support.lumerical.com/hc/en-us/articles/360034922953-Optimization-utility. [Accessed 2 April 2020].

[119] Y. Wang, W. Shi, X. Wang, Z. Lu, M. Caverley, R. Bojko, L. Chrostowski and N. A. F. Jaeger, "Design of broadband sub-wavelength grating couplers with low back reflection," *Opt. Lett.,* vol. 40, no. 20, pp. 4647-4650, 2015.

[120] J. G. Mutitu, S. Shi, C. Chen, T. Creazzo, A. Barnett, C. Honsberg and D. W. Prather, "Thin film silicon solar cell design based on photonic crystal and diffractive grating structures," *Opt. Express,* vol. 16, no. 19, pp. 15238-15248, 2008.

[121] D. Gostimirovic, "Photonmind," [Online]. Available: https://github.com/Dusandinho/photonmind. [Accessed 4 April 2020].

[122] H. Geoffrey, S. Osindero and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.,* vol. 18, no. 7, pp. 1527-1554, 2006.

[123] 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.

[124] Lumerical, Inc., "Lumerical's 2.5D FDTD Propagation Method," [Online]. Available: https://www.lumerical.com/learn/whitepapers/lumericals-2-5d-fdtd-propagation-method/. [Accessed 2 April 2020].

[125] Z. Cheng and H.-K. Tsang, "Experimentaldemonstrationofpolarization- insensitive air-cladding grating couplers for silicon-on-insulator waveguides," *Opt. Lett.,* vol. 39, no. 7, pp. 2206-2209, 2014.

[126] K. Konyushkova and R. Sznitman, "Learning active learning from data," *arXiv:1703.03365v3 [cs.LG],* 2017.

[127] P. Bachman, A. Sordoni and A. Trischler, "Learning algorithms for active learning," *arXiv:1708.00088v1 [cs.LG],* 2017.