NANOPHOTONIC INTERCONNECT ARCHITECTURES FOR MANY-CORE MICROPROCESSORS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mark J. Cianchetti

January 2012

NANOPHOTONIC INTERCONNECT ARCHITECTURES FOR MANY-CORE MICROPROCESSORS

Mark J. Cianchetti, Ph.D.

Cornell University 2012

Nanophotonics is an emerging technology that has the potential to improve the performance and energy consumption of inter- and intra-die communication in future chip multiprocessors. To date, the successful demonstration of a working large-scale system has been hampered by integration challenges and temperature sensitivity of the optical building blocks. Moreover, current approaches to interfacing with these devices are either CMOS incompatible or degrade the potential Tb/s modulation capability to only tens of Gb/s. At first glance it may seem like all of these challenges hint at today's nanophotonic devices being too impractical. However, using a combination of proposed solutions at the device and architectural level, a rich tradeoff space begins to emerge that is still largely untouched due to the knowledge gap between nanophotonic researchers on both sides of the spectrum. To this end, this dissertation attempts to fill this gap by targeting both device and system level research in an integrated fashion.

We begin with an extended background and related work section that presents the relevant parameters and functionality of key optical devices for designing interconnection networks at the architecture level. Following this, we give a detailed discussion on the system level implications of optics including communication methods and summaries of recent network architectures for both on-chip and off-chip signaling with important takeaways for designing future systems.

The lack of a comprehensive and accurate modeling strategy for optical com-

ponents in the architecture community has lead to potentially inaccurate, and inflated, power and performance estimates. Since better representation of optical devices in architectural level simulations is essential to producing trustworthy results, we present a comprehensive, mathematical model for all of the major optical building blocks. To our knowledge, this is the first comprehensive model of all relevant optical devices specifically tailored to system level design for architects.

An interesting aspect of architectural research in the field of optics is that there is not a natural progression of scaling parameters that will necessarily dictate future designs as is the case in CMOS. Because nanophotonics is an emerging technology, the potential is limitless for creating new devices that solve previous challenges. Optical packet switching is a promising approach for overcoming the performance and power limitations of bus-based on-chip networks. We present two variations of *Phastlane*, the first proposed nanophotonic packet switched architecture. In our evaluation, we demonstrate the potential improvements in system performance and power consumption across a range of modulator and receiver parameters. We also augment this analysis with projections for current optical devices using our mathematical device model.

Finally, we propose alternatives for overcoming some of the limitations of both Phastlane architectures in the event that future optical components stagnate at current performance and power consumption. Also, we use our device model to explore a less aggressive approach to nanophotonics that judiciously combines electrical and optical interconnect.

BIOGRAPHICAL SKETCH

Mark James Cianchetti graduated from the University at Buffalo in 2006 with two B.S. degrees in Computer and Electrical Engineering. He worked as a student researcher at the Center for Computational Research (CCR) throughout his four undergraduate years in the area of computational biology. He also joined the Research Experience for Undegraduates (REU) program in nanostructured devices where he worked on the fabrication of single electron transistors in the summer of 2005. Mark was inducted into Tau Beta Pi, Eta Kappa Nu, Phi Eta Sigma, Phi Beta Kappa and Golden Key honor societies and joined the advanced honors program in his freshman year. In 2006 he graduated summa cum laude.

Following graduation from the University at Buffalo, in June of 2006 Mark joined the Computer Systems Laboratory at Cornell University as an M.S./Ph.D. student. There he worked on nanophotonic interconnect for future chip multiprocessors under the supervision of Professor David H. Albonesi, publishing novel optical packet switched architectures in the International Symposium on Computer Architecture and a special issue of the Journal of Emerging Technologies. In 2008 Mark won an Intel Fellowship and defended his dissertation research in August of 2011.

In September of 2011 Mark joined Intel in Portland Oregon as a computer architect working in the Visual and Parallel Computing Group.

This dissertation is dedicated to my wife Flor, Mom, Dad, Grandpa and Grandma, Bryan and Sarbear. Thank you all so much for making me smile, and for pulling me through the tough times.

ACKNOWLEDGEMENTS

I would first like to thank my wife Flor for all her love and support. You helped me to see the good in bad and to always trust that God will work things out. I would not have survived these past five years without you by my side. Thank you, thank you and thank you!

To my Mom, thank you for making me smile, and for all your nightly letters of wisdom. Your consistent support, understanding, and love really lifted my spirit when I needed it. You helped me to forget about all the stress and taught me to just be happy.

To my Dad, your insistence on getting it done, and pushing through till the end was probably the only reason I survived all those sleepless nights in front of my computer. Thank you for always giving me good advice and for motivating me to keep going.

I'd like to thank my Ph.D. advisor, David Albonesi, for taking the tremendous amount of time to mold me into the researcher I am today. Thank you for spending hours and hours correcting my papers and slides, and for being patient all the while! You taught me what it really means to be a hard worker!

Christopher Batten and Edward Suh were also on my Ph.D. committee and provided invaluable feedback during my A and B exams. Chris, thank you for really inspiring me and for all your useful suggestions and help. Nanophotonics at Cornell, including myself, really benefited from your arrival!

Michal's Nanophotonics Group was very generous in answering my many emails about device parameters and providing thorough feedback on our system level nanophotonic designs. I'd especially like to thank Nicolás Sherwood Droz, Kyle Preston, Biswajeet Guha, Sasikanth Manipatruni, and Yoon Ho Daniel Lee for always being very helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Integrated nanophotonics is an emerging technology that has recently gained research momentum as a potential replacement for electrical interconnect in future chip multiprocessors. Previous work at the device and architectural levels have demonstrated the low power consumption and high bandwidth density that optical communication enables for both inter- and intra-die applications [43] [56] [65] [72]. However, large challenges still exist in forming a successful marriage between optical devices and conventional CMOS transistors to demonstrate a functional system. Finding a suitable method for integrating optical components and CMOS transistors that minimizes fabrication costs, utilizes standard processing techniques, and does not impact the functionality of both technologies is still an active area of research. The extreme temperature sensitivity of nanophotonic building blocks, which cease to operate correctly with temperature fluctuations as low as one degree celsius, makes integration even more difficult. Lastly, electrical interfacing circuits for performing modulation, switching and receipt severely limit the fundamental communication bandwidth that could theoretically be achieved from Tb/s to 10's of Gb/s.

These daunting challenges hint at the conclusion that today's nanophotonic devices are too impractical for constructing a network to facilitate communication in future computing systems. Although it is entirely possible that device researchers may develop solutions to the above problems using completely different components with the same functionality, using today's fundamental building block for optical networks, the ring resonator, key hurdles still exist in creating a full, functioning system. Since all of the listed challenges can be mitigated through a combination of system and device level innovation, it is essential that nanopho-

tonic researchers on both sides of the spectrum fill the current knowledge gap that exists between them. Following the first system paper in 2006 that proposed an optical bus for global processor communication [33], the lack of a comprehensive and accurate modeling strategy for optical components in the architecture community has led to potentially inaccurate, and inflated, power and performance estimates. This was shown by device level research that examined well-known nanophotonic architectures, pointing out key modeling errors [8]. However, the gap in knowledge between systems and devices also exists in the latter, where designing more efficient devices is only possible with an intimate knowledge of chip multiprocessors and interconnection networks.

In this dissertation, we attempt to address this knowledge gap and fill the spectrum between devices and systems. In Chapter 2, we present an extended background and related research on nanophotonic interconnects from both device and architectural perspectives. This includes examining the architectural parameters relevant to waveguides, ring resonators and optical receivers and how to use these components to form a basic optical link and an optically switched variation. We also present the challenges and tradeoffs associated with the aforementioned integration strategies. For device researchers and architects new to the field of nanophotonics, we explain the primary bus and switched based communication protocols that have been proposed. We then provide a detailed discussion of recent work in on-chip optical networks in chip multiprocessors, concluding with work that's been done in inter-die optical and high-performance electrical alternatives.

Since better modeling of optical devices in architectural level simulations is essential to producing trustworthy results, we present a comprehensive, mathematical model for all of the components in Chapter 3. To our knowledge, this chapter is the first fully encompassing piece of literature that combines the mod-

eling strategies of all relevant optical devices specifically tailored to system level design for architects.

One attraction of being an architectural researcher in the field of optics is that there is not a natural progression of scaling parameters that clearly dictate future designs as is the case in CMOS. Because nanophotonics is an emerging technology, the potential is limitless for creating new devices that solve the challenges listed earlier and we don't necessarily know what the future will bring. In Chapters 4 and 5 we present two variations of Phastlane, the first proposed optical packet switched network architecture. We demonstrate the potential improvements in system performance and power consumption across a range of assumed parameters for modulators and receivers. We also augment this analysis with projections for current optical devices using our device model from Chapter 3. Along with presenting a novel packet switched approach, another contribution of our Phastlane work at the device end comes form demonstrating required system parameters (i.e., modulator and receiver energy consumption) for producing an interconnect that's competitive with highly-aggressive electrical alternatives.

In Chapter 6 we conclude with important strategies for reducing the knowledge gap between optical devices and systems and discuss how architects can benefit from improved modeling strategies to guide the design of future nanophotonic networks. In Chapter 7 we present detailed proposals for future work that overcome some of the limitations in both Phastlane architectures in the event that optical devices stagnate at current performance and power consumption. This includes combining time-division-multiplexing and wavelength-division-multiplexing with a modified flow control to improve latency characteristics. We also examine varying router radices to enable low diameter network topologies such as a flattened butterfly. Also, to explore a less aggressive approach to nanophotonics, we use

our device model from Chapter 3 and the projections that it shows to devise a blueprint for mutual collaboration between electrical and optical interconnects.

CHAPTER 2

BACKGROUND AND RELATED WORK

We divide this chapter into two broad categories: An introduction to integrated nanophotonic devices and the relevant design parameters for system level architects, and a comprehensive overview of key design strategies for architecting an optical interconnect in a future chip multiprocessor. We begin the first half with a discussion on the emerging field of nanophotonics and the potential it has to revolutionize communication between processors within a die and off-chip to other system components. Next, we include a high-level description of recent work that's been done in the field of silicon based optical devices that pertains to network design. Finally, we combine all of the devices we introduced to show a full optical communication link with and without active broadband switching.

In the second half of this chapter we switch directions and discuss how to use the previously discussed optical devices to form an interconnection network. We begin by presenting communication protocols used in arbitrating for, and transmitting on, shared network resources. Many architectural level network proposals using optics have appeared since the first ring based approach in 2006 [33]. To conclude this chapter, we choose some of the unique approaches from this prior work and present the key design takeaways.

2.1 Enabling Device Technology

In this section we begin with an overview of emerging nanophotonic architectures, focusing on current and projected performance, energy and area overheads of relevant devices. We then present key design parameters and tradeoffs of the fundamental building blocks of an optical communication link for architectural level design. We include recent work that examines techniques for pushing com-

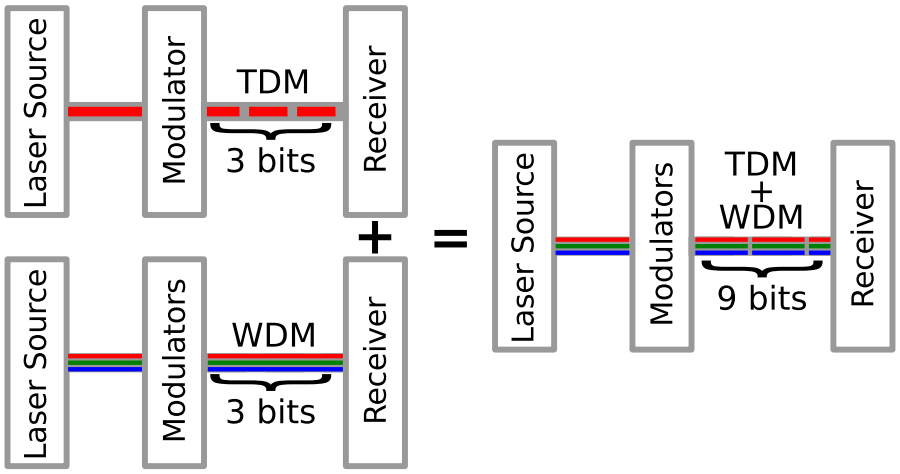


Figure 2.1: A laser source supplies light to modulators that turn the light on or off depending on an electrical control signal. In a TDM-only system, the entire data packet is transmitted in time such that each bit is a small slice of light in the link. In the WDM-only variation, the entire packet is transmitted on multiple wavelengths, each of which represents a bit of data. To achieve very high bandwidth communication, WDM and TDM can be combined.

munication performance into the 10's of Gb/s. Then we present an overview of a photodetector and accompanying amplifier stages for converting light pulses into electrical bits. Finally, we use all of the presented optical devices to form two variations of a full optical link and summarize fabrication strategies for integrating optics with a CMOS chip multiprocessor.

2.1.1 Nanophotonics Overview

Integrated optical communication has the potential to offer advantages over traditional electrical wires in four main categories: bandwidth density, device latency, energy consumption and area overhead. Communication in an optical link utilizes time-division-multiplexing (TDM) and wavelength-division-multiplexing (WDM) to transmit information between source and destination. TDM decomposes data into multiple bits that trail one another in an optical link. The rate at which these serialized bits are transmitted is dictated by the smallest of the maximum rate of

modulation at the front end and maximum receive rate at the back end. WDM further increases total communication bandwidth by using multiple wavelengths of light to simultaneously transmit data in parallel. In Figure 2.1 we demonstrate a TDM-only optical link, a WDM-only optical link and a combined version. In the TDM-only variation, the data packet being transmitted is modulated such that bits are serially transmitted in time to the receiver. In WDM-only, the same three bits are instead encoded using wavelengths of light. This is beneficial in low latency applications where the time for a data packet to reach a destination is only dictated by the time it takes for the front of the signal to reach the end receiver. We demonstrate a WDM-only system implementation in Chapters 4 and 5 when we introduce Phastlane and Phastlane 2.0. Lastly, to get a very high degree of communication bandwidth WDM and TDM can be combined. Exceptional bandwidth density comes from the simultaneous use of TDM and WDM and the nanometer sized width of a single link ($\sim$450nm [57]).

Latency characteristics fall into two categories: switching speeds of the optical devices and the velocity of the optical signal in a link. Depending on the material used to construct the optical network, the latter has been shown to be about 10.45ps/mm in single crystalline silicon links [11]. This equates to a speedup of about 2x over a highly-optimized electrical wire [20]. Using optical links made of silicon nitride (Si_3N_4), the latency can be reduced to approximately 6ps/mm [16] at a cost of increased link width and spacing. We discuss structure, performance and power characteristics of optical links in more detail in Section 2.1.2.

The latency of the optical devices are related to the maximum achievable communication rate that the bits in TDM modulated data can be transmitted and received. We show in Chapter 3 that using a CMOS process it is possible to make these devices very fast ($\sim$10's of ps). In Chapters 4 and 5 we present the Phast-

lane architectures, which utilize the optical devices in a combinational manner, requiring each optical component latency to be as small as possible. However, we also demonstrate that designing for very low latency potentially leads to degraded WDM level and high laser power requirements. Depending on the functionality of a network architecture, the latency of the devices may not be as important as high bandwidth transmission. In addition to increasing the switching rate, bandwidth can also be increased by adding more wavelengths or optical links.

The energy consumption of an optical link can be divided into two main components: electrical energy spent in transmitting and receiving the optical data, and the energy required to power a laser for supplying the wavelengths of light to the modulators. When an optical signal travels between a source and destination node, it encounters multiple points of power loss. The primary reasons for signal attenuation are due to roughness in the optical link (due to fabrication imperfections), absorption of light by the link's material, and insertion loss as the light passes other devices while traveling to its destination node. We discuss each of these loss mechanisms in more detail and the relationship between laser power requirements and optical link loss in Chapter 3.

At the end of a communication link the receiver's photodetector requires enough optical power to mitigate the potential for bit errors. This power level dictates the characteristics of the laser at the front end, which must supply enough power to potentially multiple wavelengths in a waveguide to account for all of the insertion loss it will experience prior to reaching the detector. Depending on the network architecture and number of communication nodes, the laser power in future chip multiprocessors may be in the 10's of watts [16] [65].

The second component of energy consumption is from the electrical power dissipation in the transmitter and receive components. We show in Chapter 3

that a wide range of tradeoffs exist in projecting the power consumption of these devices; however, typically this number can be projected to fall into the 1's to 10's of pJ per bit per device.

The area requirements of a basic optical link (i.e., a link consisting of modulators at the front-end, links for the data to travel though, and receivers at the back-end) are dictated by the size of the modulators, width and spacing requirements of the link and size of the receiver components. A modulator's size is dependent on the amount of optical insertion loss it adds to the system (due to bending losses) and the material from which it is fabricated, that typically amounts to a minimum of $28um^2$ [57]. The electrical portion of the ring is a driver circuit and depending on the required drive strength and technology node, should fit within the dimensions of the ring portion of the modulator. The dimensions of a link depend on its material composition. For example, a link fabricated in single crystalline silicon has a width of approximately 450nm and a similar spacing requirement [57], whereas in silicon nitride this increases to 1μm and 10μm, respectively [16]. Lastly, the receiver is composed of two components, the photodetector and a series of amplifying stages for converting the optical signal into a digital level voltage. The size of a germanium based photodetector is limited by the width of a single optical link in one dimension ($\sim$450nm) and a required length for absorbing the light in the second dimension ($\sim 10\mu$m [12] [13]). The amplifying stages that we examine in this dissertation (see Chapter 3) are a series of three to four CMOS inverters.

Initial work at the device and system level has shown that the emerging field of nanophotonics has the potential to revolutionize the way that processors communicate within and between dies. The bandwidth density, latency, energy and area characteristics of current fabricated devices and projections into the future suggest it may be beneficial to replace traditional electrical interconnect with opti-

cal links in future chip multiprocessors. We believe that these projections warrant research into the examination of interconnects at the architectural level for inter- and intra-die communication. In the following sections we provide an overview of the basic building blocks of an optical link, providing the details relevant for an understanding of how to use these components in system level design.

2.1.2 Optical Waveguides

Optical communication links are built using a structure known as a waveguide, which guides multiple wavelengths of light simultaneously between two points. When a material with a high index of refraction (guiding medium) is surrounded by a material with a lower index (cladding medium), light supplied by a laser source bounces off the sides of the waveguide and propagates in a forward direction. Within the waveguide, a signal propagates at a certain speed typically dependent on its wavelength and material properties, but can be estimated for the purpose of system design to be 10.45ps/mm [11] in silicon cladded with silicon dioxide (SiO_2). Other materials for fabricating the waveguide have also been examined. For instance, a silicon nitride guiding medium reduces this latency to approximately 6ps/mm [26].

A single waveguide is generally on the order of 100's of nm to μm's in width (450nm for single crystalline silicon and 1μm for silicon nitride) with a spacing dependent on the index contrast between the guiding and cladding material. In a silicon nitride waveguide cladded with dioxide, a low resulting index contrast requires a waveguide spacing of approximately 10μm [16]. This is reduced to around 450nm for the silicon based waveguides [57]. Thus, while the nitride material offers better latency, depending on the network architecture it may not provide enough bandwidth density due to its larger waveguide width and spacing requirements.

The power requirements of the laser that supplies wavelengths of light in a WDM waveguide is dependent on the points of loss in the network architecture. As light propagates down a waveguide it attenuates very slightly due to side-wall roughness (from fabrication imperfections) and absorption of light (since the wavelengths of light are typically large enough to overcome the bandgap of the waveguide material). These loss mechanisms increase as more optical power is put into the waveguide due to nonlinear affects, which are described in more detail in Chapter 3. Assuming that the total power in the waveguide is small enough to neglect nonlinear attenuation, the propagation loss of a silicon waveguide is approximately 1dB/cm [21] and the nitride waveguide less at 0.1dB/cm due to the smaller index contrast [16].

A nanophotonic communication network among tens and eventually hundreds of processors on a die requires a very complex array of optical waveguides. Depending on the topology of the network architecture, waveguide crossings or multiple waveguide layers may be required. The latter is similar to the different metal layers in a CMOS metal stack. Single crystalline waveguides cannot be deposited and thus multiple layers are infeasible, resulting in a loss of 0.045dB ($\sim$1%) per crossing [8]. Silicon nitride waveguides have the benefit of being fabricated with back-end-of-line (BEOL) techniques with multiple deposited layers that eliminate crossings [8]. We discuss fabrication techniques further in Section 2.1.6.

2.1.3 Optical Ring Resonator

The ring resonator is the fundamental building block of a nanophotonic interconnect. It's use as a modulator, switching element and demultiplexer covers all of the necessary functionality for implementing a network that uses light for communicating data. In this section, we provide a brief overview of this device including

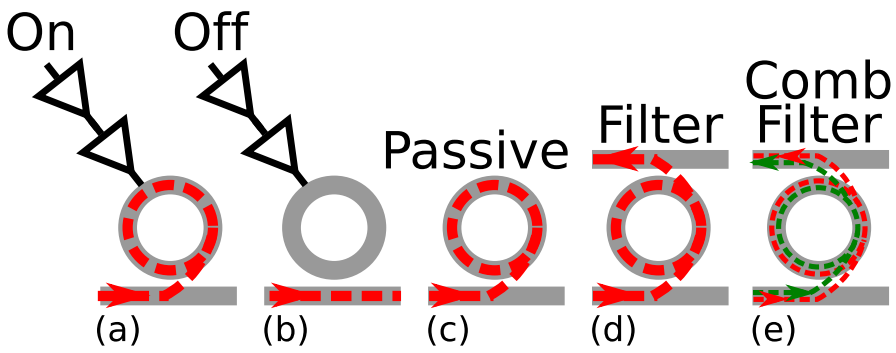


Figure 2.2: An optical ring resonator can be actively tuned to a particular wavelength passing in the waveguide. When the ring is turned on in (a), the wavelength leaves the waveguide and enters the ring. Similarly when the ring is off in (b), the wavelength continues in the waveguide. It is also possible to passively tune a ring resonator at fabrication time to always remove a particular wavelength from the neighboring waveguide as in (c). A filter is necessary for implementing a switching element or at the end of an optical link for demultiplexing individual wavelengths so that they can be routed to separate photodetectors. In (d) we show a filter that can be actively or passively tuned to a wavelength in the waveguide. Lastly, a comb filter has the same functionality as the filter in (d) except that it removes all of the wavelengths from the waveguide when it is turned on in (e).

recent work that has looked at improving its switching performance in fabricated implementations, with the potential to enable high bandwidth, low energy data transmission using a small area footprint. A detailed mathematical analysis of the ring's operation is presented in Chapter 3.

The basic functionalities of a resonator are shown in Figure 2.2. Only specific frequencies of light will enter the ring without continuing past in the waveguide. These frequencies acquire enough phase shift in the ring to cause destructive interference with the light in the coupled waveguide. The resonant frequencies can be dynamically tuned by injecting or removing charge carriers from the ring to change its index of refraction.

A modulator is implemented using the device configurations in (a) and (b) where an electrical driver turns on and off the ring, forcing light into the ring and allowing it to pass by, respectively. In the case of the modulator, only a single

wavelength of light in the WDM waveguide will enter into the ring. To modulate multiple wavelengths of light, a separate modulator per wavelength is required. Passive operation of a ring resonator is shown in (c) where light of a specific wavelength always enters the ring, which is not powered by a driving circuit.

Filtering a specific wavelength from the waveguide and transferring it to another waveguide is important for optical data switching and at the end of a basic optical link to demultiplex wavelengths of light off the WDM waveguide for delivery to the photodetectors. This type of ring, known as an add/drop filter, is shown in (d) and can be operated actively or passively. Extending this functionality for the purpose of broadband switching (i.e., simultaneous removal of all wavelengths in a WDM waveguide) is also possible as shown in (e). As with the previous filter, this ring can also be actively or passively operated.

Active tuning of a ring for modulation or optical routing of data packets through ring switches needs to be very fast to enable high performance packet communication. As a result, previous work has examined techniques for achieving ultra low switching speeds of these devices. Pre-emphasis uses an initial voltage spike to quickly inject carriers into a PIN diode built around the ring, thus turning on the device, and then an immediate decrease in applied voltage to avoid injecting an excess of carriers [70]. This latter part is important as the time to turn off the ring is directly related to the amount of injected carriers. Rapid turn-off is accomplished by applying a negative voltage across the ring. A data rate of 12.5Gb/s is demonstrated at an initial turn-on voltage of 8V, which is then reduced to 4V to keep the resonator on. A reduced voltage of -4V is used to turn off the device. While this method allows high-speed operation of a ring resonator, scaled CMOS supply voltages are typically 1V or less, and thus applying such high and negative drive voltages may be challenging.

One method for reducing the required voltage to turn on a resonator is to bias it at a pre-determined voltage chosen so that a small signal swing around it will produce a large shift in the wavelength that it filters [44]. While this mitigates potentially high driving voltages, this device still suffers from slow performance at a maximum operation rate of only 1 Gb/s. Also, because a bias of 0.96V with a signal swing of 150mV are required, depending on the CMOS driving technology, this still may not be achievable at sub-volt supply voltages in future scaled technologies.

One promising method for increasing the switching speed of an optical ring resonator is through the use of ion implantation [66] [68]. Various work has examined how device speed characteristics can be improved by injecting ions, such as oxygen, into the ring to improve its carrier recombination lifetime. We show in Chapter 3 that the speed of a PIN diode switched ring resonator is estimated as $2.3\tau_c$, where τ_c is the carrier recombination lifetime of the diode. However, injecting ions into the ring does not provide free performance gains. The newly implanted ions act as absorption centers for light propagating in the ring, increasing the optical power attenuation. The use of ion implants is a very promising avenue for reducing device latency and keeping drive voltages at a low enough level to be compatible with scaled CMOS technologies.

The use of a PN diode to turn on and off a ring resonator is also possible through reverse bias carrier removal [73]. This is accomplished by splitting the ring into N and P doped regions, which results in switching speeds as high as 27 Gb/s. However, as with many of the previous methods for improving performance, this comes at a cost of having to apply a 10V reverse bias across the diode.

In terms of both of these alternatives, the fundamental speed limitation to turn on and off a ring is dictated by the photon lifetime in the device [42]. For a typical micrometer sized ring resonator this is on the order of a few pico seconds

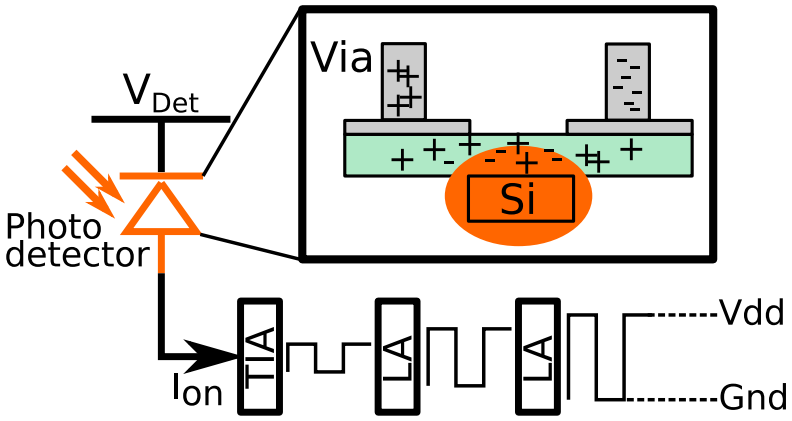


Figure 2.3: Building blocks of an optical receiver for converting light pulses into electrical bits of data. Light traveling in a silicon based waveguide strikes the photodetector and produces electrons and holes. These charges are swept across the detector and into the terminals where they are used to form the input to the amplifying stages. Here a transimpedance amplifier and a number of limiting amplifiers build the signal up to a digital voltage level.

and is determined by the amount of time required for the light of a particular wavelength to enter the ring and destructively cancel itself from continuing past in the waveguide. The critical path for turning it on and off is thus formed by the slower time to inject and remove charge carriers.

2.1.4 Optical Receiver

The optical receiver sits at the back end of an optical waveguide and converts light into a digital voltage level signal. Demultiplexing ring resonators separate the wavelengths of light in the WDM waveguide and route each one separately to a different receiver, which is shown in Figure 2.3. At the front of the receiver sits the photodetector, which is based on a metal-semiconductor-metal (MSM) design [12] [13] and converts light into holes and electrons. As light traveling in the silicon based waveguide strikes the detector, its frequency is high enough to overcome the detector material's bandgap energy. As a result, free charge carriers

are generated. Since the detector is biased at V_{Det}, the generated charge is quickly swept to the terminals of the detector where they form an input voltage to the following amplifier stages.

One way to implement the amplifier is to divide it into multiple stages. In Figure 2.3, for example, we show a high-gain transimpedance amplifier followed by multiple limiting amplifiers to achieve a digital level voltage signal at the end of the chain [30]. However, noise sources in the transistors of the amplifier may cause erroneous behavior, since a digital zero (one) may be latched at the sampling point of the signal, when the actual intended bit was meant to be a digital one (zero). To quantify this problem, the receiver circuitry has an associated bit-error-rate (BER) metric which dictates the probability that a single bit will be erroneously mistaken for its inverted form.

We describe the BER of a receiver in Figure 2.4 where an output voltage from the amplifiers is ready to be latched at time *Sample* by the network clock circuitry. If the voltage is above the *Threshold* point, it is considered a digital one, and if it is below, a digital zero. One way to determine the BER is to estimate the probability of erroneously sampling the data at a point when the noise fluctuation causes a mistaken bit by two gaussian probability density functions. Each one represents the probability that the noise sources in an intended digital level signal will cause erroneous sampling. *P(0|1)*, for example, is the probability that a digital zero will be sampled when the bit was actually a digital one. The level of noise that occurs impacts the variance of the gaussians, which we assume to be the same when we discuss BER in more detail in Chapter 3.

There are five primary figures of merit for detectors that pertain to architectural level network design. These are the size of the detector, its maximum operating bandwidth, responsivity, required bias voltage, and fabrication compatibility with

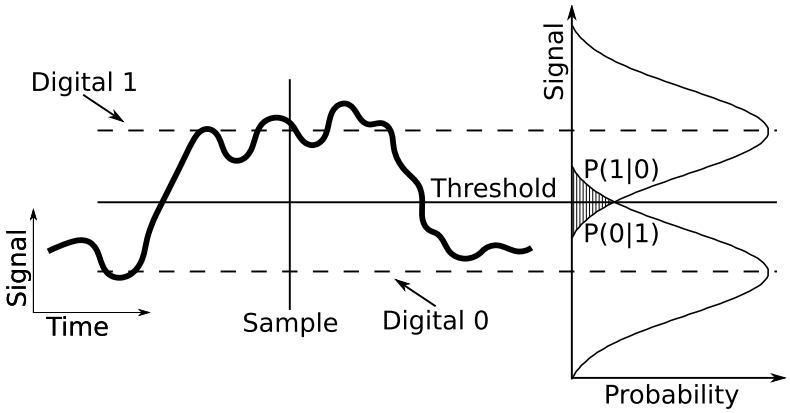


Figure 2.4: Bit-error-rate (BER) dictates the probability that a single bit will be received as a digital one, when it was actually intended to be a zero or vice-versa. This is due to signal noise generated by thermal fluctuations, dark current from the detector and leakage currents in the transistors. *Threshold* represents the voltage level that distinguishes a digital one from a zero. Typically a gaussian is used to represent the probability density function of noise generating an erroneous bit at the *Sample* point. These erroneous probabilities are denoted as P(0|1) and P(1|0), or the probability that the receiver sees a digital zero given that a one was actually present and the reverse, respectively.

current CMOS processes. The size of the detector is related to its ability to efficiently absorb all of the light in a passing waveguide in the shortest distance possible. Typically this is dependent on the type of material used to fabricate the detector and is why a vast majority of research in this area has examined the use of germanium. Germanium is a direct bandgap material that can be overcome by single photon absorption of light in the regime of wavelengths that are used in nanophotonic networks. Photon absorption is more efficient, therefore, than in a silicon based material, which has an indirect bandgap. The maximum bandwidth and responsivity of the detector determine, respectively, its maximum speed and the amount of optical power that needs to be absorbed to obtain a certain BER. A few of the physical properties that limit the speed of detection are the type of photodetector (i.e., metal-semiconductor-metal or PIN), its geometry, bias voltage, and presence of parasitics. Lastly, the fabrication compatibility with CMOS is vi-

tal to the successful integration of an optical link in future chip multiprocessors. In the rest of this section we examine some recent device research that has examined different photodetectors and the tradeoffs associated with each of the five primary figures of merit just discussed.

Germanium based photodetectors are attractive for future nanophotonic interconnect because of their compact size ($\sim$1μm by 10's of μm), high responsivity ($\sim$.44A/W) and low dark current [12] [13]. Dark current is the amount of current that leaves the detector when no light appears in its neighboring waveguide. It is important to minimize this since the BER of the amplifier stages may erroneously produce a digital one when no signal is actually present. In this dissertation we assume a germanium based detector using an MSM configuration as shown in Figure 2.3. Here the charge that is generated by the light is quickly swept to one of the terminals for current generation requiring a voltage less than $\sim$1V. The speed of the device is carrier time limited, and thus the achievable bandwidth is based on the time for the charge carriers to be swept across the germanium region into one of the terminals. Methods for estimating this latency and calculating the resulting bandwidth of the detector are presented in Chapter 3.

Silicon based photodetectors might provide a more economical alternative to germanium detectors since the latter requires the bonding of two wafers. A silicon based detector can be fabricated in current CMOS processes. Previous work has examined the use of a PIN based photodiode using silicon implanted with ions to increase linear absorption of light [25]. With increased linear absorption, it is possible to get around the indirect bandgap problem of silicon. However, one problem with these approaches is the length of the detector, which is on the order of millimeters to absorb a waveguide's optical signal. Increasing the number of ions in the diode reduces this distance, but at the cost of a higher required bias

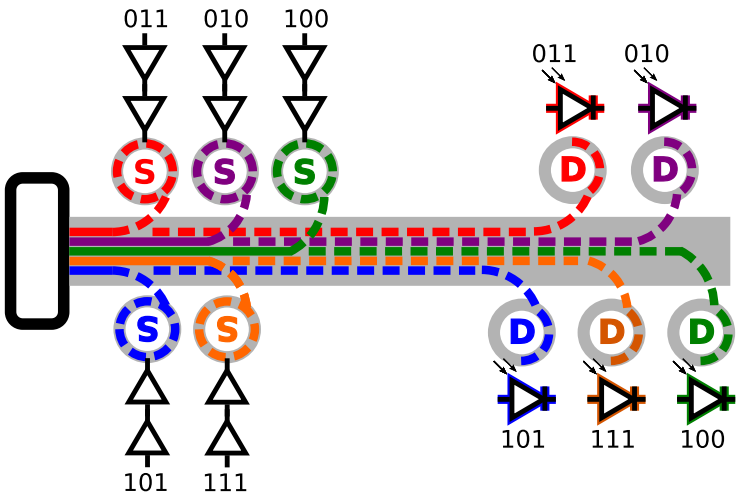


Figure 2.5: A full WDM link that uses multiple wavelengths and TDM to communicate data to a downstream node. A ring modulator per wavelength converts electrical bits of data into the optical domain where the light travels at high-speed to the end of the link. There, passive ring resonators demultiplex each wavelength and deliver them to detectors for conversion to electrical voltages. Here *S* denotes the ring modulators belonging to the source node, and *D* the demultiplexing resonators at the destination node.

voltage ($\sim$>5V). Other previous work in this area uses defect generation caused by protons to generate mid-level bandgap energy states to increase the optical absorption of silicon [10].

Another interesting approach is to use polycrystalline silicon to form a detector, which has the advantage that it can be deposited on top of a processor die along with silicon nitride waveguides [58]. Polycrystalline silicon shows increased absorption over the single crystalline version. The device fabricated in this work showed a responsivity as high as 0.15A/W with a required reverse bias voltage of -13V. Another advantage to this approach is that the detector *is* the demultiplexing ring resonator. This is beneficial from an area standpoint because it saves the total area that would have been occupied by all of the detectors in the interconnect.

2.1.5 Combining Devices to Form an Optical Link

A complete optical link comprising a single waveguide with five wavelengths, each time-division-multiplexed to facilitate high bandwidth data transmission between two points is shown in Figure 2.5. The total bandwidth of the link is decided by the allowed WDM level and individual data rates of each transmitted wavelength. The latter is dictated by one of three system design parameters: the maximum data rate of the modulator, the maximum receiver bandwidth or by the demultiplexing resonators at the end of the link. In Section 2.1.3 we described how previous research has pushed the maximum achievable data rate in modulators to the GHz range with different techniques like ion implantation and voltage pre-emphasis. We examined the receiver in Section 2.1.4 and further examine its latency characteristics in Chapter 3. However, another design parameter that can potentially hinder the total system data rate is the bandwidth of the demultiplexing resonators. Modulation of a data signal using on/off switching (the optical ring modulators turn light on for a digital one, and off for a digital zero) produces high and low frequency sidebands around the wavelengths of light provided by the laser. The demultiplexing resonators are tuned to a specific wavelength, but attenuate other frequency components that might exist around that wavelength. We examine this in more detail in Chapter 3.

We show later in Section 2.2 that through the use of multiple optical links similar to the one in Figure 2.5, it's possible to form a wide variety of network communication topologies. However, every communication event between two points requires a full transmit and receive of all of the data in a packet. Thus, to get to a final destination point, a source node must either transmit its data directly to the destination (in a point-to-point fashion), or undergo multiple full data receives and transmits prior to reaching it. One way to eliminate these additional transmits

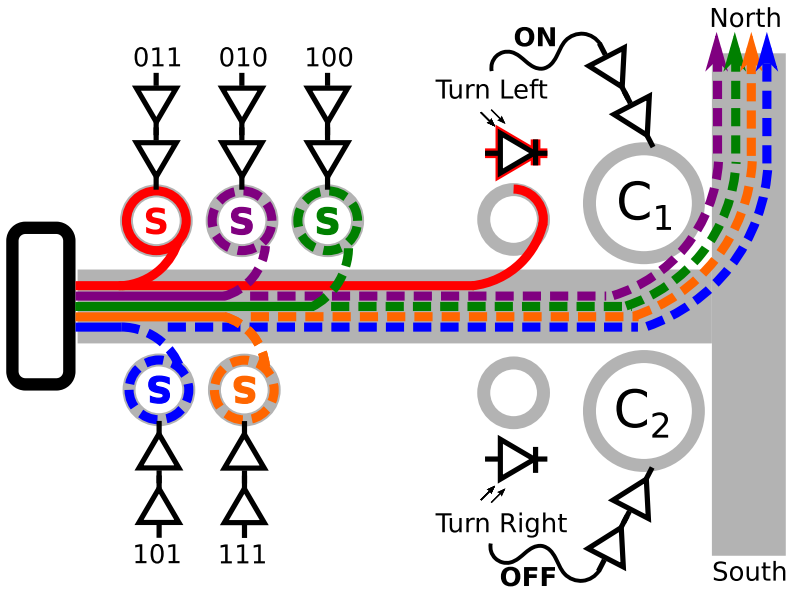


Figure 2.6: Optical data switching avoids the need to transmit and receive an entire data packet potentially multiple times between a source and destination. In this example the red wavelength encodes whether the packet desires the North output depending on whether its light is on or off. This control signal passively couples into the ring resonator prior to the two comb switches. Light that is received is used to turn on the first comb filter (C1) to route the entire data packet out the North port. The wavelength encoding the South output is not present in this example, causing the comb filter C2 to remain off.

and receives is through the use of an *optically switched link* shown in Figure 2.6. Here an incoming data packet has the opportunity to leave out the North port or South port. The only bit that is optically received is the red wavelength, which is the control signal for the comb switch, C_1, leading to the North output. In this example, since this bit is present (i.e., light is on), it is used to form the driving signal across C_1, routing the entire data packet out the North.

In Chapters 4 and 5 we present two chip multiprocessor network architectures that utilize optical packet switching to transmit a packet through multiple hops in a mesh topology. We propose two optical router architectures that utilize pre-computed routing bits for turning on comb switches in each crossbar along the destination path. Because the optical devices are used for control signals, they form

the critical delay of the packet through the network. Therefore, it is important to optimize the latencies of these devices to maximize the distance that a packet can reach in a network clock cycle.

2.1.6 Fabrication Techniques

Various companies and academic institutions have developed novel methods for accommodating the cumbersome requirements of optical devices in an attempt to integrate them with conventional CMOS fabrication processes. These methods fall into three broad categories as shown in Figure 2.7: monolithic, deposited and 3D integration of the optical components with CMOS transistors. Each one of these techniques has advantages and disadvantages, and in this section we briefly overview the tradeoffs associated with each approach. The goal is to integrate nanophotonics with a CMOS circuit without impacting the performance of either fabricated separately, and to do so as cheaply as possible and without having to perform unconventional processing.

Monolithic integration is an attractive way to use a current CMOS design flow and still be able to obtain the benefits of optical communication within a die. Previous work has examined a bulk CMOS 28nm technology with many of the optical components necessary for building a complete link [47]. Light is coupled into the chip using vertical gratings for guiding it into waveguides fabricated in the polysilicon metal layer. This layer is also used to form ring resonators, including a second order filter bank (i.e., two ring resonators cascaded to widen and/or create a box-like resonance response). Some of the challenges associated with this approach is the loss inherent in the polysilicon, which could reach 1000dB/cm. One of the reasons for this high loss is due to poor confinement of the mode in the waveguide, since the oxide layer surrounding the polysilicon is not thick enough. To mitigate

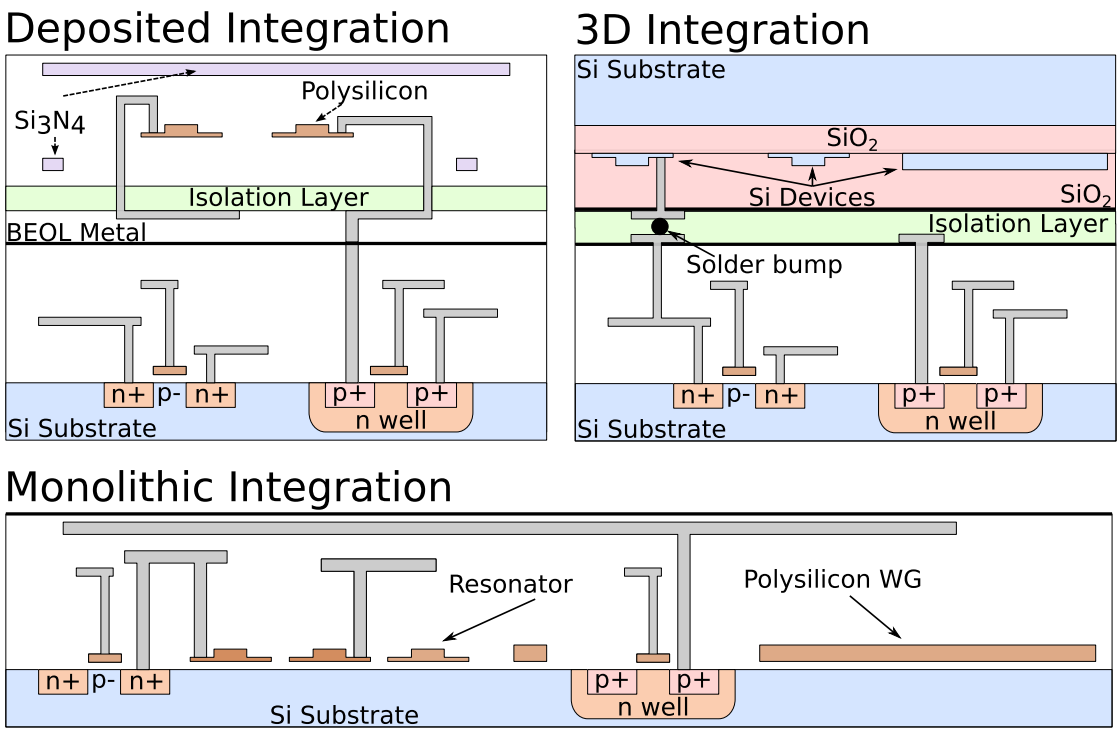


Figure 2.7: Three primary methods for integrating optical interconnects with a conventional CMOS process technology. The first method uses standard CMOS techniques to deposit optical devices above the processor metal layer post-fabrication. One advantage of this approach is that it enables multiple waveguide layers, which eliminates optical power loss due to waveguide crossings in a complex network topology. One of many 3D approaches uses die bonding facilitated by micro solder bumps that join two separate dies, each optimized for either the optical or CMOS devices. Monolithic fabrication uses a conventional CMOS process to integrate the optical components alongside the transistors. This has the benefit of low cost, but uses potentially precious real estate in the active layer.

this, post fabrication etching is used to remove silicon below the polysilicon and fill it with silicon dioxide. This reduces the propagation losses to a more manageable, although still high, 55dB/cm. One problem that is still being researched is process variation that exists across a CMOS die and the resulting resonance shift of ring resonators, which could potentially be mitigated using ring heaters. Another area that's being examined is the use of the silicon-germanium layer present for stress engineering the p-type transistor for fabricating germanium based photodetectors.

To accommodate all of the requirements of optical devices, the 3D approach

allows them to be fabricated in a process highly optimized for nanophotonics. For example, instead of fabricating the waveguides using a high loss polycrystalline silicon, single crystalline silicon would dramatically reduce optical signal attenuation. Using this approach, both the optical devices and the CMOS transistors are completely separate from one another at fabrication time until they are bonded to one another post-fabrication. Recently, an optical ring resonator was manufactured in a Luxtera-Freescale 130nm SOI CMOS optimized specifically for nanophotonics [73]. A cascode driver circuit was separately fabricated in a 90nm bulk CMOS and subsequently attached to the optical die using microbumps. These bumps are attached to the bonding pads at the top of each of the two dies to join them and create a channel for communication. In this way, the underlying electrical circuit is able to provide a driving voltage to the ring resonator above. Some of the disadvantages of this approach are the potential thermal problems that can result from stacked layers and the added complexity of bonding two dies together. Other work has examined epitaxial growth of silicon islands [48], oxygen ion implantation [36] and wafer bonding [23] to form a vertical optical layer, none of which are currently compatible with standard CMOS processing.

An alternative to 3D chip stacking is to use materials that can be deposited using a back-end-of-line (BEOL) approach following the fabrication of underlying CMOS circuits [56]. In this approach, low loss, low latency silicon nitride waveguides transmit light between two points using polycrystalline silicon ring resonators, both of which can be deposited. Some of the advantages of this method over pure 3D integration are the introduction of multiple waveguides layers and higher communication bandwidth between layers due to the use of vias instead of micro bumps. The former is especially important in potentially complex network topologies used in many core chip multiprocessors. We showed in Section 2.1.2

that the optical power loss per waveguide crossing is approximately 1%, which may compound to a large number if the interconnect is not carefully designed. High communication bandwidth between the optical devices and electronics is also important as some architectural proposals for using nanophotonics integrate as many as one million ring modulators [65].

2.2 On-chip Optical Interconnect Architectures

In this section, we present an architectural level analysis of nanophotonic interconnect in future chip multiprocessors. We begin with communication methodologies for facilitating optical data transmission between nodes in many core network architectures, derived from previous optical network proposals. Photons cannot be buffered and to date no suitable logic exists for manipulating light; thus, a network design must carefully choose a network topology, flow control and routing algorithm that avoids these drawbacks while still benefiting from the low power, high bandwidth data transmission that optics offers. In the second portion of this section, we provide a literature review of recent on-chip nanophotonic interconnect proposals and the key takeaways from each of these studies that can help system level designers create more powerful and efficient networks.

2.2.1 Communication Methodologies

Point-to-point

Point-to-point network (P2P) topologies require a dedicated communication path between every source/destination pair in the system as shown in Figure 2.8. In this small example, two sources, *S1* and *S2*, use different wavelengths of light to communicate with downstream destinations. *S1* has exclusive use of the red and

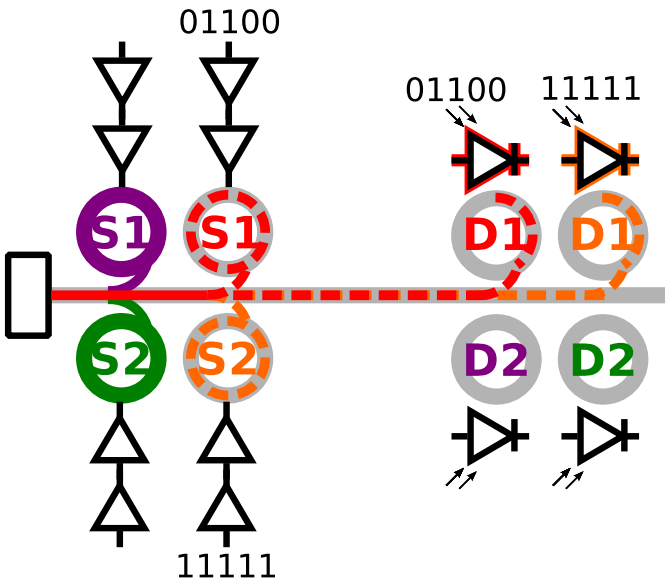


Figure 2.8: In point-to-point communication, both sources communicate with the two destinations using a unique wavelength of light. In this example *S1* transmits data to *D1* and simultaneously *S2* to *D1* as well. Notice that the purple and green wavelengths are not being used since *S1* and *S2* are not communicating with *D2* and *D1* respectively.

purple wavelengths for destinations *D1* and *D2*, respectively, and *S2* has exclusive use of the orange and green wavelengths for transmitting data to *D1* and *D2*, respectively. In the example in the figure, *S1* is sending a packet to *D1* and *S2* is also transmitting to the same location. Notice that neither the purple nor the green light are being used since transmission to the corresponding destinations is not occurring.

Although in this small example we provide distinct communication paths between every source/destination pair with single wavelengths, increasing the communication bandwidth is possible by increasing the number of wavelengths (i.e., the level of WDM) and also by increasing the number of waveguides in the system. One disadvantage of using P2P communication in a nanophotonic interconnect is the potentially high bisection bandwidth required for a large number of nodes. This problem is not unique to optics and also exists in an electrical implementation.

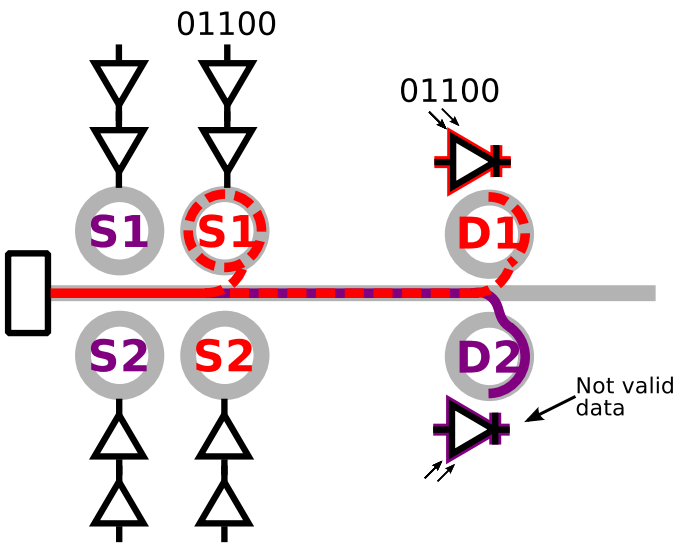


Figure 2.9: Multiple-writer-single-reader (MWSR) requires global arbitration for the red wavelength and purple wavelength corresponding to *D1* and *D2*, respectively. Both sources are able to modulate light on the purple and red wavelengths depending on the intended destination of their packet. In this example, because neither *S1* nor *S2* are communicating with *D2*, this wavelength of light is unmodulated and thus invalid data enters *D2*.

Multiple-writer-single-reader

Electrical packet switched routers typically utilize multiple-writer-single-reader (MWSR) communication to transmit packets between input ports and output ports. In the optical domain, MWSR has been used to facilitate communication between different processing nodes in a network architecture. To transmit to a destination, the source must arbitrate for exclusive use of the destination's communication path. In a conventional electrical router, for example, switch arbitration occurs for deciding which input port can exclusively access an output port. Figure 2.9 shows a small example that uses MWSR between two source and two destination nodes in an optical interconnect. The red wavelength is exclusive to destination *D1* and the purple to *D2*. If either *S1* or *S2* want to simultaneously transmit to the same destination, they must arbitrate for the exclusive use of the corresponding wavelength. In this example *S1* is sending a data packet to *D1*.

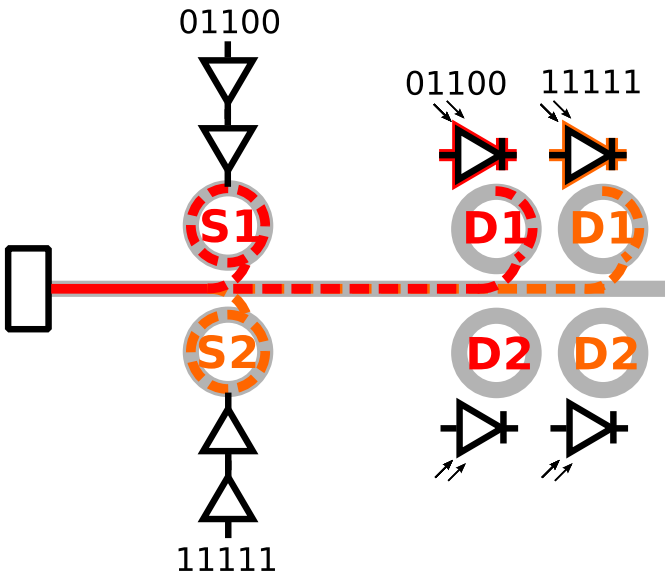


Figure 2.10: Single-writer-multiple-reader (SWMR) assigns the red wavelength to *S1* and the orange to *S2*. Any communication that occurs out of a source regardless of the destination will modulate the data on its assigned wavelength of light. In this example both *S1* and *S2* are transmitting data to *D1*. Both destination nodes are able to read all of the wavelengths in the system, in this case orange and red.

Notice that the purple wavelength of light is not being modulated by a source node, and thus holds no useful data. Destination *D2* still receives this light since it belongs to this node, but does not actually use the data. One disadvantage of MWSR in large network architectures is the potentially long latency in performing global arbitration required to gain exclusive use of a destination's set of wavelengths. Therefore, the system architect must strike a careful balance between the latency to transmit a packet, and the overheads associated with performing arbitration. One way to mitigate the arbitration latency at the expense of increased network diameter is to use multiple sub-networks, each with a more localized arbitration scheme.

Single-writer-multiple-reader

Single-writer-multiple-reader is the opposite of MWSR in that every source node only writes to a particular set of wavelengths and waveguides, but every destination has access to all of the wavelengths and waveguides in the system. A small example showing the concept of SWMR is shown in Figure 2.10, where each source node transmits data on a unique wavelength of light. Source *S1* uses the red wavelength and *S2* the orange, and both sources are simultaneously communicating with destination *D1*. As with the previous communication methodologies, transmission bandwidth can be increased by adding more wavelengths and/or waveguides to the system.

Since every destination node can read the transmission contents of every source in the system, two variations of SWMR exist depending on the power and performance requirements of the network. In the first version, every node in the system receives a portion of the optical power in every transmitted packet. Following receive and translation to an electrical signal, the nodes will determine whether they were the intended destination of the packet. If not, the contents are simply discarded. While this method offers good performance since arbitration and control signaling are eliminated, it requires high power dissipation because every node reads the packet's contents. To overcome this problem, the second variation of SWMR uses reservation assisted tuning for data transmission. When a source node want to transmit a packet, it will first send a small reservation packet that is received by all the destinations in the system. Following this, only the intended destination will turn on its receivers corresponding to the wavelengths and waveguides of the source node. Thus, unnecessary power loss is avoided since only the true destination reads the data packet, but this also comes at a cost of increased latency to set up a communication path prior to sending the actual data packet.

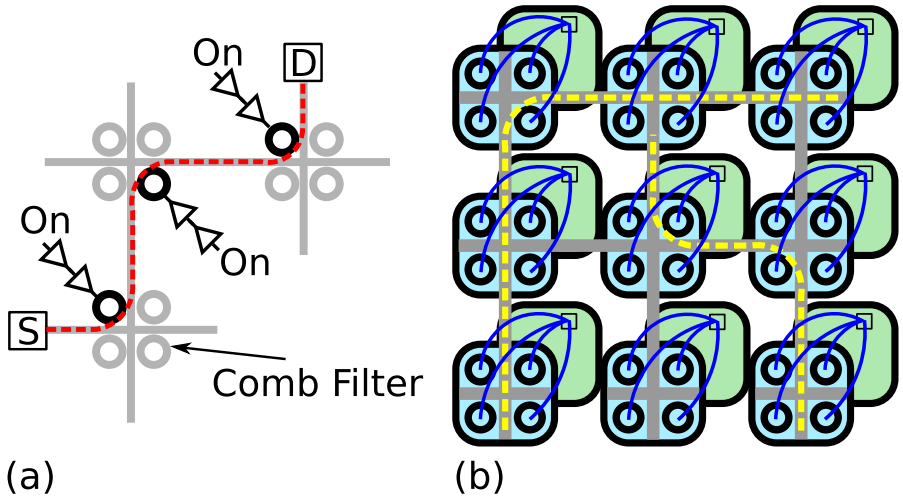


Figure 2.11: Circuit switched communication configures ring resonator comb filters ahead of data transmission. When all of the rings have been properly configured to form the path between a source/destination pair, optical signals are transmitted from source to destination as shown in (a). When the entire data packet has been transmitted, the path is torn down and parts of it can be reused to form different network paths. Using this functionality, it's possible to form different network topologies including the mesh shown in (b), where a control network configures the optical comb filters.

It may also be the case that a destination cannot simultaneously receive packets from multiple source nodes. In this case, arbitration will occur at the receiver, which notifies losing source transmitters when they can send their packets, or to retransmit their data to participate in another round of arbitration. If a broadcast based scheme is used, wasted transmission and receipt of potentially large data packets may occur. Thus, it may be beneficial from a power standpoint to use reservation packets for arbitration prior to optically modulating the source's payload.

Circuit switched

Circuit switched operation of a nanophotonic interconnect uses a separate control network (previous work has used a light weight packet switched electrical network [62]) to set up optical comb filters between a source/destination pair ahead

of packet transmission as shown in Figure 2.11(a). Once the setup is complete, the data can be transmitted unimpeded to the destination node at the endpoint of the path. As in electrical circuit switched networks, while the path is being used, other packets requiring overlapping resources must wait for the completion of transmission. The electrical control is also used to tear down the optical path after the destination receives the entire packet. Figure 2.11(b) shows an example of a mesh network topology that consists of multiple optical router banks of comb filters which are pre-configured by an electrical set up network.

One of the benefits of the circuit switched approach is the simplicity of the optical data path, which obliviously travels along a pre-configured waveguide route until it reaches a receiver. Unlike the previous methods that use MWSR, SWMR or P2P communication, all waveguides and wavelengths along the path between source and destination are exploited. This is particularly beneficial for very large packet sizes that might potentially face significant serialization penalties using a subset of a source's total communication bandwidth, which may even be unused in the case of MWSR, for example, when no nodes are communicating with a certain destination. One disadvantage of a circuit approach is in the context of a chip multiprocessor running a shared memory program. The amount of data in a cache line may not be enough to amortize the latency overhead of setting up the optical data path prior to sending a packet. Thus the designer should be aware of the communication characteristics of the underlying processing architecture prior to choosing one of the methods described in this section.

Optical packet switched

Optical packet switching permits data to traverse through multiple points in a network without requiring translation between the electrical and optical domains. Figure 2.6 demonstrates how an optical control signal is received into the electrical

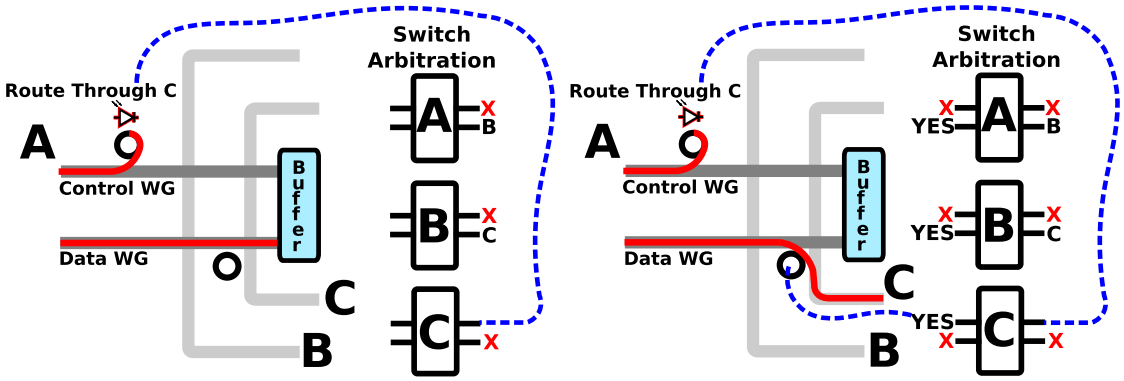


Figure 2.12: An optical control signal travels in parallel with its payload data and upon entering the input port of an optical router, translates to the electrical domain for participating in switch arbitration. Assuming that it wins, the electrical grant signal is used to drive the appropriate comb filters in the optical switch for routing the payload portion of the packet. A packet is electrically buffered at the end of a network clock cycle, or if it loses arbitration, in which case it is optically retransmitted into the network in a future network cycle.

domain for controlling a packet's route by turning on an appropriate comb filter. The first variation of optical packet switching is known as burst transmission. This approach is similar to circuit switching in that it uses an optical control signal that travels just enough ahead of an optical data packet to turn on the proper comb filters for routing the payload [4] [14]. Previous work in burst switching requires dropping packets or deflective routing if a packet is unable to leave out its desired destination port.

The second variation of optical packet switching, shown in Figure 2.12, eliminates the timing between the optical control signal and payload, since in some cases there may be uncertainty associated with the delay through an optical router. In this figure, an incoming packet on input port A desires to leave through output port C and uses its translated routing control signal to participate in switch arbitration. The packet's payload is sent simultaneously with the control and is either routed through the switch or electrically buffered if the packet loses switch arbitration. In this example, arbitration is won and an electrical signal turns on the

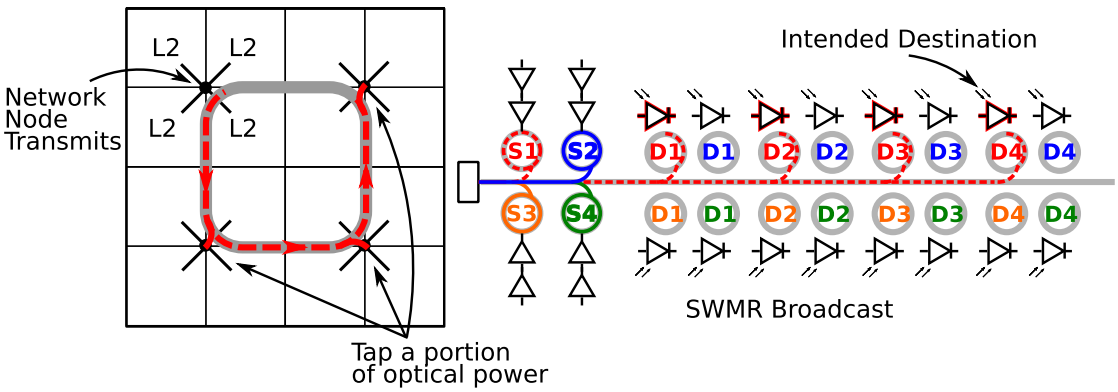


Figure 2.13: The Cornell ring architecture uses a single-writer-multiple-reader broadcast based bus to transmit data between four network nodes. Each network node is composed of four L2 caches, each of which belongs to a group of four processors. In this example, *S1* is transmitting data to *D4* using the red wavelength, which is broadcast to each destination in the system. Upon reading the packet's intended target, only *D4* will use its contents. In the actual paper, the communication bandwidth is multiplied by utilizing multiple wavelengths and waveguides.

appropriate comb filters in the optical crossbar so that the packet's payload and remaining control bits continue to the downstream router. In Chapters 4 and 5 we present two nanophotonic architectures that use this version of optical packet switching to route packets between source and destination.

2.2.2 Nanophotonic System Proposals

Previous research in nanophotonic networks for on and off-chip communication has produced many creative and unique ideas for overcoming the limitations of photonics (i.e., lack of buffering and logic) while exploiting its benefits. In this section, we present recent architectural level proposals for high bandwidth communication between processors, processors and DRAM, and multiple dies in a high performance server environment. We discuss the key features of each proposal and how the use of optics in place of an electrical network benefitted power consumption and performance.

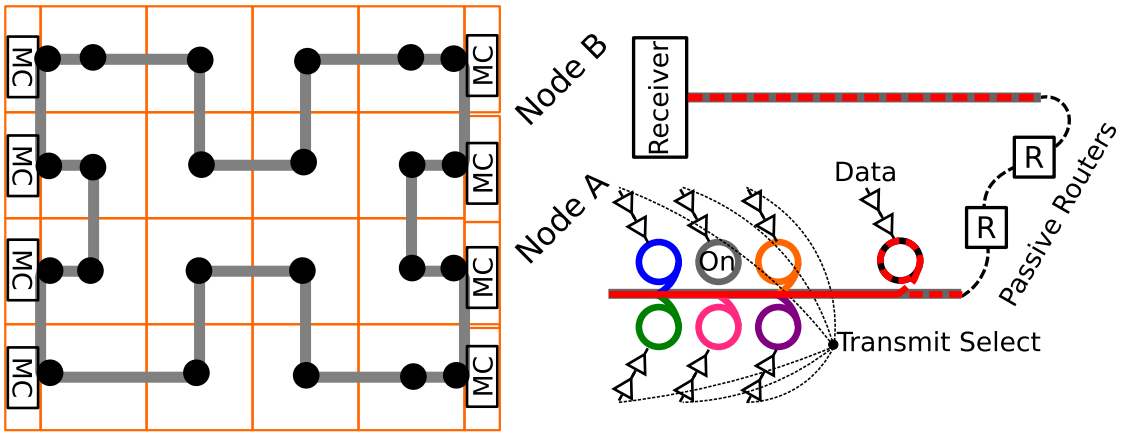


Figure 2.14: Prior to transmitting into the network, a source node arbitrates for the use of its intended destination's output port. Assuming that it wins, it optically transmits its packet on a pre-assigned set of wavelengths that passively traverse over a torus topology (layed out in a bus fashion) in an oblivious route which guarantees its successful delivery to the end node. Using a combination of wavelengths and packet routing, transmitted packets never encounter contention once sent into the network. Every node is only capable of transmitting and receiving to and from a single destination and source. In this example, Node A transmits to Node B and thus tunes its transmission resonators to use the red wavelength. Similarly, the destination node will tune its resonators to only allow the red wavelength to reach its receiver.

Kirman et al. [33] propose a hierarchical interconnect for communication among 64 cores in 32nm technology. A group of four cores sharing an L2 cache communicate with four other groups through an electrical switch as shown in Figure 2.13. The four 16-processor nodes in turn perform packet transmission using an optical ring that implements a single-writer, multiple-reader bus broadcast protocol. Each node writes to the bus using its own unique wavelengths, which obviates the need for arbitration, and information is read by coupling a percentage of the power from each signal.

Kirman et al. [34] propose a passively routed torus network that optically routes transmissions through statically configured switches. The switch configurations are fixed at design time to route wavelengths between input and output ports. When a node submits a packet into the network, it transmits on particular wavelengths

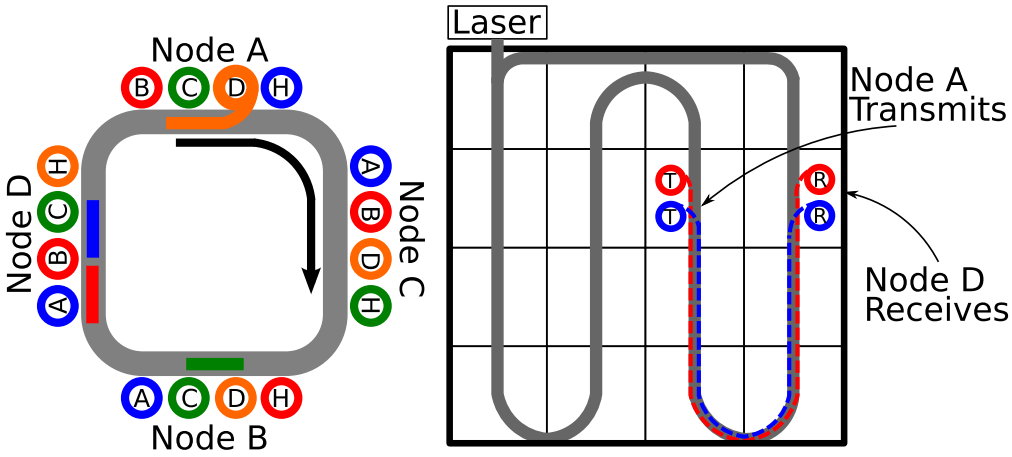


Figure 2.15: The Corona architecture is a global crossbar implemented using optical busses that use a multiple-writer-single-reader communication protocol. Because MWSR requires global arbitration for transmitting to end nodes, a global token bus is used for competing source nodes. Here a different wavelength of light represents the right to transmit to a particular node. In this example Node A wants to transmit to Node D and attempts to remove the orange wavelength, successfully doing so. The crossbar is layed out in a serpentine format and since Node A has the proper arbitration token, it transmits to downstream node D.

corresponding to its desired destination. These wavelengths route through the passive resonators in the interconnect towards the destination. The network is laid out in a bus configuration to avoid waveguide crossings and increase bisection bandwidth as shown in Figure 2.14. Here black dots represents nodes that are connected in the network, where each node is either an L2 cache or a memory controller (denoted as *MC*). When a source node needs to send data to a destination, it partakes in global arbitration via a separate optical network that implements a point-to-point communication protocol. Following arbitration, winners are able to transmit their packets into the network by tuning the appropriate transmit resonator via *Transmit Select*, traversing an oblivious routed set of switches prior to finally reaching the destination's receiver at the other end.

Vantrease et al. [65] propose optical buses for communication among 256 cores in 16nm technology. Similar to [33], multiple cores are grouped as a node and

communicate through an electrical sub-network. Inter-node communication occurs through a set of multiple-writer, single-reader buses (one for each node) that together form a crossbar as shown in Figure 2.15. Optical token arbitration resolves conflicts for writing a given bus. An optical token travels around a special arbitration waveguide, and a node reads and removes the token before communicating with its intended target. Following transmission, the node reinjects the token into the waveguide for use by other requestors. Chip-to-chip serial optical links communicate with main memory modules that are divided among the network nodes.

Vantrease et al. [64] build on their Corona work by examining two schemes for implementing optical switch arbitration in a global crossbar. The first scheme uses a single optical token per destination node that continually circulates around an arbitration waveguide. Any source node needing to send a packet will attempt to sink the token corresponding to the desired destination. Following this action, it may keep the token for a pre-specified length of time, sending up to N packets prior to retransmitting it onto the arbitration waveguide for use by other nodes. Credit flow control is enabled by encoding the number of free entries corresponding to the downstream buffer into the token. A token slot scheme is also proposed where instead of arbitrating for the exclusive use of a channel across multiple cycles, source nodes arbitrate for transmission slots in the channel, which corresponds to a much finer time granularity. Node starvation is handled in both schemes by explicitly notifying the destination node, which in turn takes appropriate action by allowing starved nodes to transmit into its buffers.

Shacham et al. [62] propose a 2D optical Torus topology similar to the architecture shown in Figure 2.11. Data transfer occurs through a grid of waveguides with resonators at crosspoints for turns. Control is handled by an electrical set-

up/tear-down network. To enable data transfer, a packet sent on the electrical network moves toward the destination and reserves the optical switches along its route. When this path is established, the source transfers data at high bandwidth using the optical network. Finally, a packet is sent on the electrical network to tear-down the established path.

Pan et al. [51] propose a system with 256 processors interconnected in clusters of eight using a concentrated, electrical mesh topology. Global communication between the different clusters occurs over an optical bus using a reservation-based single-writer, multiple-reader configuration. Upon transmitting a packet to a destination, the source node globally broadcasts a reservation signal to all downstream intra-cluster nodes. These nodes tune into this signal, but only the intended destination will receive the data packet.

Pan et al. [50] propose a multiple-writer, multiple-reader bus for mitigating static laser power through globalized sharing of network channels. A token slot arbitration scheme is also presented that differs from [64] by preventing node starvation through a two-pass technique. The arbitration waveguide wraps around all nodes twice, where in the first pass every node has a guaranteed slot for transmission. In the second pass any node may use any available slot. In [64], credit flow control is encoded in the number of available free slots, where if no available buffers exist in a downstream node, no transmission slots are visible to source nodes. In [50], credits are encoded in separate tokens that also circulate around the arbitration waveguide and must be captured by a source node prior to transmission.

Joshi et al. [29] describe an optically implemented version of a reconfigurable non-blocking Clos network scaled up from the two simplified variations shown in Figure 2.16. The first version uses three stages of electrical routers that are at-

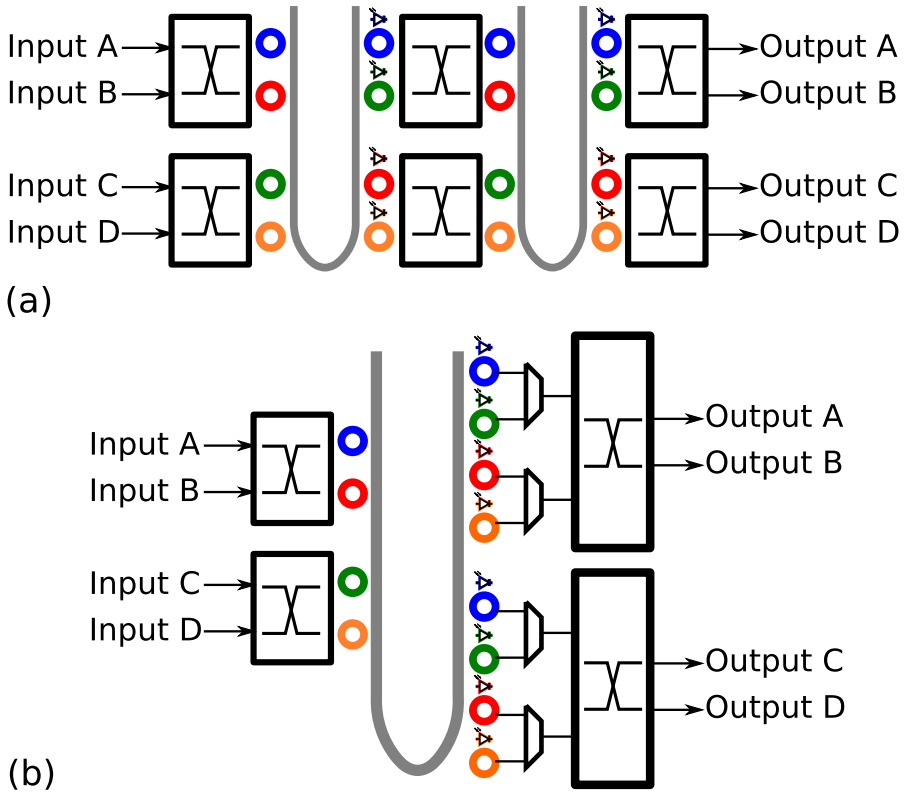


Figure 2.16: The Clos architecture is reconfigurably nonblocking and has the potential for better performance than other optical network topologies. For simplicity, we show a scaled down version of the network used by the authors. Two variations of the Clos are shown, one with an electrically routed middle stage (a), and the other using a SWMR photonic replacement (b). One of the advantages of the photonic replacement is that the electrical packet has to undergo fewer optical-to-electrical and electrical-to-optical conversions before reaching its destination, potentially reducing power consumption.

tached to one another via point-to-point connections using multiple wavelengths and waveguides for high bandwidth data transfer. In the second implementation, the middle set of electrical routers are removed and replaced with a single-writer-multiple-reader optical bus. This could be beneficial from a power standpoint since data packets undergo fewer optical to electrical conversions and vice versa. The authors examine the point-to-point variation in Figure 2.16(a) and demonstrate significantly less optical power, thermal tuning power and area overhead compared to a global optical crossbar network. Additionally, they compare against electri-

cal Clos and mesh networks, demonstrating improved energy efficiency at similar performance.

Optical burst switching operates by transmitting variable sized data bursts behind a path setup signal that configures every switch ahead of time according to the packet's desired destination as described earlier in Section 2.2.1. Traditionally, if a burst's control signal is unable to obtain a switch, it is dropped. Other work has examined deflection routing and delay lines to partially remedy this problem [4, 14].

Our Phastlane architectures leverage elements of each of these prior proposals. Like Shacham et al., we use a grid of waveguides with turn resonators, but there are several important distinctions between our proposals, some of which are due to differences in data payload size. We rely on only WDM to pack a narrow packet into one cycle, while they use WDM and TDM to achieve very high bandwidth transfer of a much greater amount of data. We optically send control along with the data to set up the router switches on the fly rather than use a slower electrical control network.

Lastly, recent work has examined the extreme temperature sensitivity of optical ring resonators in an on-chip context. Nitta et al. [46] raise the issue of thermal runaway when using a combination of carrier injection and heating to correct resonance shift in a ring. They also showed that using only resistive heating results in high power consumption in excess of 100W for a die area of approximately 400mm^2 and $\sim$500K resonators used in a global crossbar topology. To combat some of these issues they propose a sliding window technique that inserts rings into the spectral ends of a resonator bank. Rings are grouped together according to location and using only current injection, proper operation of the system resonances is enabled.

2.3 Inter-die Optical Interconnect

Off-chip optical links provide high bandwidth, power efficient communication in a system composed of multiple dies. This enables a) cost effective systems by decomposing a system-on-chip (SoC) into smaller pieces, increasing yield and potentially mitigating non-recurring engineering (NRE) costs, b) macrochips and mixed technology systems that are not possible to monolithically integrate, and c) energy efficient, high bandwidth communication between processors and DRAM, and processors across different server components. The following work proposes optical network solutions for inter-die interconnection.

Beamer et al. [7] propose to optically guide data and commands signals from a processor memory controller to an off-chip DRAM module. The photonic links extend deep into the DRAM all the way to individual banks, providing a high degree of energy efficiency compared to electrical alternatives. Additionally, the optical links enable high aggregate pin-bandwidth density through dense wavelength-division-multiplexing.

Udipi et al. [63] also examine the use of nanophotonics for improving the latency, bandwidth and energy characteristics of off-chip DRAM accesses. Optics is used to overcome the pin bandwidth and energy limitations of conventional inter-die communication using electrical wires. In combination with 3D chip stacking and offloading much of the functionality of the memory controller to a localized spot on the DRAM chip, the authors achieve better performance at reduced power consumption.

Beamer et al. [5, 6] examine the use of point-to-point optical fibers for connecting processors in a multi-socket system to memory controllers using a star optical coupler. Processors and caches are organized into clusters where each die consists of multiple clusters and memory controllers. Every cluster has a direct connection

to every other memory controller in the system using the high bandwidth off-chip optical links.

Koka et al. [35] propose a multi-chip substrate for building macrochips, using optics to interconnect separate processor and memory dies. A processor interfaces with the substrate and other processor dies through vertical optical couplers. Every processor die has an associated memory die that it communicates with via electrical proximity coupling [15]. This study explores different optical network topologies ranging from fully connected point-to-point networks to a circuit switched network using a torus topology.

Pan et al. [49] compose macrochips using optically interconnected processor dies. They propose off-chip optics to overcome the power density of very large SoCs. By breaking an SoC into smaller components and connecting them using optical fibers, cooling costs are mitigated without impacting network performance.

Cianchetti et al. [17] monolithically disintegrate an SoC into smaller optically interconnected chiplets (dies) to reduce development and fabrication costs. They propose a passive optical hub chip for connecting the chiplets in a flattened butterfly topology to minimize inter-die signal attenuation. Macrochips are also enabled by using the hub chip to interconnect multiple dies with total system area larger than the reticle limit.

Binkert et al. [9] extend the on-chip Corona topology to a full chip optical router architecture. Nanophotonic signals couple into the die via fibers and are immediately translated to the electrical domain for buffering. The authors found no difference between a centralized electrical arbiter and the use of optical arbitration, and thus opted for the former. Like the Corona crossbar, nodes communicate using a multiple-writer-single-reader protocol, but unlike Corona the larger die area enables a switch speedup of 2X. This is also partially due to the use of con-

centration, where four nodes share an input into the crossbar. Similar to [46], the authors propose the use of additional ring resonators with resonances in between the channel spacings of the system. This reduces power consumption in the resistive heaters responsible for maintaining the wavelength resonances of the system for proper operation.

2.4 High Performance Electrical Interconnects

Packet switched networks can adversely impact the latency of a packet by incurring many per hop router delays. This can become detrimental to network performance especially in high diameter topologies. Phastlane is able to achieve low average packet latencies over a wide variety of traffic patterns without sacrificing throughput. The following work attempts to achieve the same result in the electrical domain.

Kim [31] simplifies an electrical router microarchitecture by eliminating separable switch allocators in favor of fixed priority arbitration. Additionally, a dimension-sliced switch allows a packet to traverse the router and inter-router links in a single clock cycle. Starvation can result from the fixed priority arbitration but is resolved through delayed flow control credits, which prevent upstream routers from sending additional packets to a processor's local router, allowing the starved processor to inject into the network.

Peh and Dally [53] propose speculative router pipeline execution. To reduce the router pipeline of virtual channel routers, switch arbitration is performed in parallel with virtual channel arbitration. To decrease the performance impact of using speculation, non-speculative requests are given priority over speculative ones. Overall they demonstrate that a virtual channel router can have similar zero-load latency as a wormhole router through speculative pipeline execution and

still provide high throughput. Look-ahead routing also reduces the per-hop pipeline depth in a router by precomputing a packet's desired path in the previous upstream router [24].

Park et al. [52] decompose an electrical router into multiple stacked dies to decrease network packet latencies and energy consumption. A router's buffers, crossbar and control logic are spread across multiple layers, decreasing its area footprint and allowing a packet to traverse the switch and inter-router link in a single cycle. Average packet latencies are further decreased through the addition of express channels, which are enabled because of the area savings. Other work has also used 3D integration to increase network performance [32] [71].

Dally [18] attempts to mitigate a packet's per hop-router delay through Express Cubes, which is a k-ary n-cube topology augmented by one or more levels of express channels that allow non-local messages to bypass routers. This is accomplished with Interchanger nodes, which are equivalent to a router architecture except that they are connected to one another across large distances. When a packet reaches an Interchanger, it may either continue to the next downstream router, or it may enter an express channel where it is bypassed to the next downstream Interchanger.

Kumar et al. [38] propose Express Virtual Channels to reduce packet latency in an electrical router beyond techniques such as lookahead routing and speculation [53] and without needing additional physical channels as in [18]. Packets traveling in an express virtual channel that passes through a router are given priority over all other packets requiring the same output port, allowing a packet on an express lane to have a reduced latency path to its destination. Starvation at a router is eliminated by explicit upstream signaling, which disables express channels passing through the router.

Our goal in Phastlane is to reduce a packet's latency path through the network.

However, unlike Kumar et al. [38] and Dally [18], we do not accomplish this by allowing a packet to bypass router pipeline stages. Perhaps Kim [31] is most similar to our work in spirit in that it attempts to simplify the architecture in order to gain the benefits of reduced latency. In Phastlane we use predecoded source routing and rotating switch priority in the optical domain to minimize router delay. Similar to Peh and Dally [53], delay is further reduced in Phastlane by using a form of speculation (switch pre-configuration), whereby ports are configured ahead of packet traversal to commonly-used straight path outputs.

CHAPTER 3

NANOPHOTONIC DEVICE MODEL

In this chapter, we expand on the nanophotonic building block overview given in Chapter 2 to include detailed equations for modeling performance and power. In Section 3.1, we begin with an in depth introduction to the optical ring resonator and describe the fundamental design parameters for architecting a high bandwidth, optical communication link. Following this section, we analyze each of the components in the link separately, continually building on the analysis to conclude with projected nanophotonic device performance and power consumption estimates for scaled CMOS technology nodes at the end of the chapter. The tradeoff between wavelength-division-multiplexing (WDM) and enabled optical data rate in a waveguide is examined in Section 3.2. We then provide a model for carrier injection into an optical ring modulator and show how ion implantation increases achievable signal data rate at the expense of increased optical propagation loss due to absorption. Section 3.4 describes the tradeoffs in an optical receiver and examines the bit-error-rate (BER) as a function of data rate and power consumption. Optical insertion losses are an important design parameter in a nanophotonic interconnect since they directly impact the required level of laser power. In Section 3.5 we provide results for optical loss in the modulator and demultiplexing resonator banks at the front and end of an optical link, respectively. We discuss nonlinear signal attenuation in a waveguide and the loss of ring resonator functionality from thermal fluctuations and free charge carriers in Section 3.6. Finally, Section 3.7 culminates in an optical device tradeoff analysis.

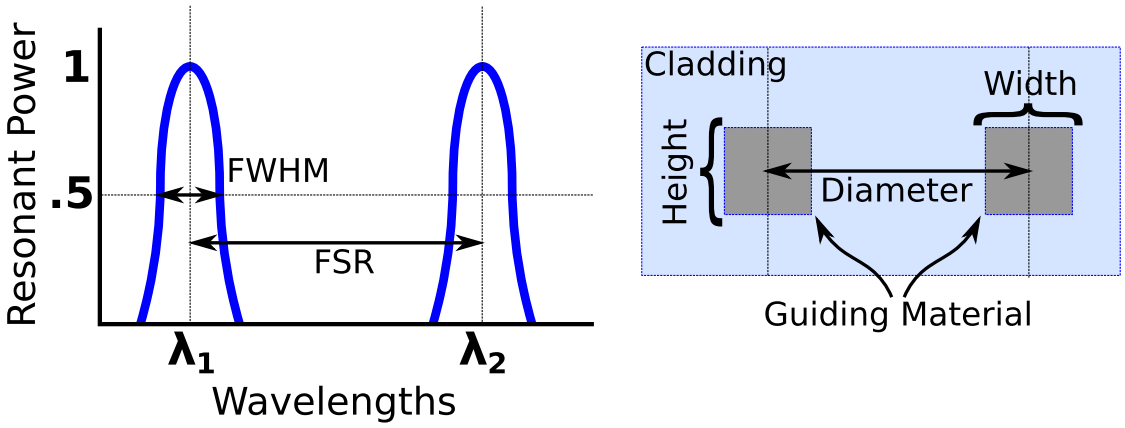


Figure 3.1: Defining characteristics of an optical ring resonator. The Free-Spectral-Range (FSR) dictates the spacing between cyclical resonant peaks. The Full-Width-Half-Maximum (FWHM) is the width of a resonant peak at half maximum. The resonators that we examine in this dissertation are rectangular waveguides with the optical signal confined in the guiding material buried in a cladding material. Evanescent tails are used to couple light between waveguide and ring resonator. The diameter of the resonator is defined as the center-to-center waveguide distance when looking at the cross-section of the ring.

3.1 Fundamentals of Nanophotonic Links

In this section, we begin with an introduction to the ring resonator and describe the characteristics that are most pertinent to low power and high performance data transfer. The versatility of the ring allows it to be simultaneously used as a transmitter, switching element and demultiplexer at the receiving end of a waveguide. Using this device as a foundation, we show how to build a high performance communication link. We demonstrate how the FSR of the system can be calculated and the resulting system WDM level based on a set channel spacing between different wavelengths. By utilizing multiple rings, each capable of transmitting data at GHz frequencies, total communication bandwidth in the Tb/s can be achieved.

3.1.1 Optical Ring Resonator

Ring resonators are the fundamental building block of integrated nanophotonic interconnect. Their compact size (microns in diameter), low power consumption (pJ) and high speed operation (GHz) has contributed to extensive studies in on-chip and off-chip communication in future computing systems. When coupled next to an optical waveguide, the ring will sink multiple wavelengths corresponding to its resonant peaks, each of which is spaced a set distance from the other known as the Free-Spectral-Range (FSR) as illustrated in Figure 3.1. The width of each resonant peak is characterized by the Full-Width-Half-Maximum (FWHM) parameter. Both the FSR and FWHM dictate the level of WDM that can be used in a waveguide, which is examined further in Section 3.1.2. The last parameter that characterizes the performance of a ring resonator is its Quality Factor, defined as:

$$\text{Quality Factor} = \frac{\text{Total Energy in Ring}}{\text{Energy loss per round trip}} \tag{3.1}$$

This is also written as:

$$\text{Quality Factor} = \frac{\lambda_{center}}{\text{FWHM}} \tag{3.2}$$

In a ring resonator the quality factor is negatively impacted by increasing optical loss or by adding more coupling sources. The latter is the case for a ring resonator coupled to two waveguides. A large quality factor is important because it enables a high degree of WDM and low switching power when used as a modulator. However, there is a tradeoff that is further examined in Section 3.2 where if the quality factor becomes too high, the maximum per wavelength data rate a ring can support is reduced accordingly.

The switching properties of the ring resonator are important for electrical signal modulation and optical packet switching in a chip multiprocessor's network-on-

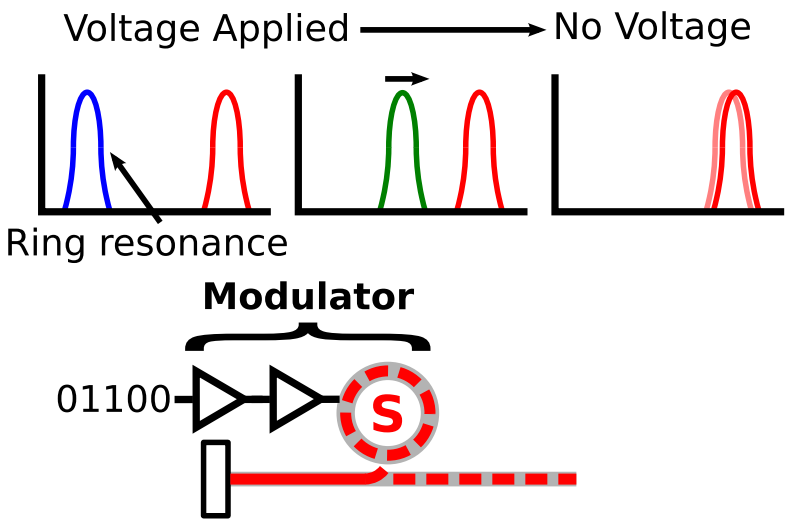


Figure 3.2: Electrical carrier injection into a ring resonator shifts its resonant peaks. In this example when a voltage is applied across the resonator by a driver, the resonator allows the light to pass by. When the voltage is removed, its resonant peaks are shifted such that one of them matches the wavelength in the waveguide, thus removing it. This mechanism enables high-speed signal modulation from the electrical to optical domain.

chip (NoC) and off-chip interconnect. The ring is turned on and off by carrier injection into a PIN diode surrounding the device waveguide. A model for this injection is described in more detail in Section 3.3. Figure 3.2 demonstrates how optical modulation occurs by applying and removing a voltage across the ring. When the applied voltage is eliminated, carriers are removed from the device and its resonant frequency peaks shift. When they shift enough to match one of the wavelengths in the neighboring waveguide, that wavelength is captured from the waveguide and routed into the ring. Similarly, when the driving voltage is applied, carriers are injected and the majority of the wavelength's power passes by the ring untouched. Besides data modulation, this functionality is also important in switching applications where light is diverted or routed into a particular waveguide.

Optical ring resonators serve the following important functions in a nanophotonic interconnect as shown in Figure 3.3:

1. Modulators - Carrier injection from an electrical control signal is used to

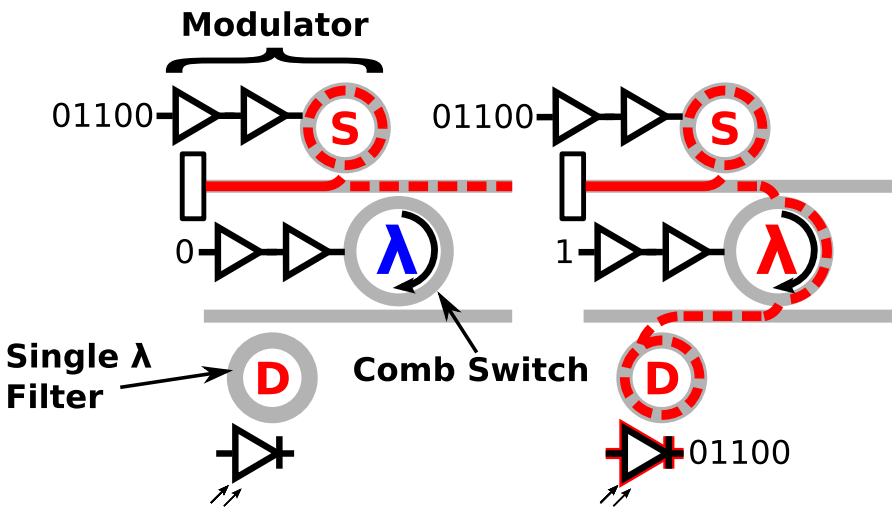


Figure 3.3: Optical ring resonators can be used as modulators, switches and filters. The data flows through a waveguide where it can be switched to a different direction and subsequently filtered and then received by a photodetector. The different operation modes of the ring makes it the fundamental building block of an optical network.

change the resonant frequencies of a ring. This enables the conversion of electrical input data to optical pulses of light. An electrical driver circuit turns the ring on and off, which causes it to sink or ignore light passing in the waveguide thereby forming the digital ones and zeros. Because every ring can be designed to modulate different wavelengths of light, the simultaneous use of multiple modulators enables a high degree of WDM. Electrical carrier injection for the purposes of modulation is discussed further in Section 3.3.

2. Comb switches - The ring resonator sandwiched in between two waveguides in Figure 3.3 transfers all of the wavelengths from one waveguide to the other. Similar to the modulator operation, an electrical signal can be applied across the ring to turn it on and off, corresponding to either switching light from the top waveguide to the bottom or allowing it to pass through, respectively.

3. Wavelength dependent filters - These filters are necessary to demodulate wavelengths of light from the waveguide to feed into optical photodetectors. Each filter only sinks one of potentially many wavelengths (unlike comb switches) passing

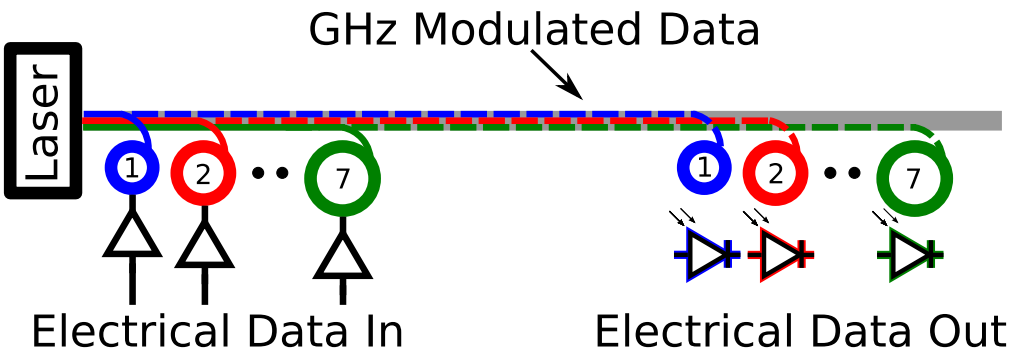


Figure 3.4: Multiple ring modulators and downstream receivers operate on a distinct wavelength that simultaneously travels with other modulated wavelengths in the same waveguide. These wavelengths are separated from their neighbors by a spectral distance known as the channel spacing.

through in the neighboring waveguide. Like the previous resonator functionalities, this filter can also be actively switched.

3.1.2 Wavelength-Division-Multiplexing

Wavelength-division-multiplexing allows multiple distinct wavelengths to be packed into a single waveguide for achieving high bandwidth density. Figure 3.4 shows a nanophotonic communication link that builds on Figure 3.3 through the addition of more ring modulators and receivers to take advantage of WDM. We also removed the comb filter since it is not related to our discussion on high data rate communication. Additionally, to ease explanation, we only show a WDM level of seven wavelengths; however, in an actual interconnect this number would probably be in the tens of wavelengths. Since each modulator simultaneously transmits data, it is important that they each operate using a distinct wavelength of light. The WDM structure of the optical link in the figure uses seven separate modulator and receiver pairs to transmit each wavelength. The total number of wavelengths that can be simultaneously transmitted in a single waveguide is limited by the FSR of the system and the channel spacing of each resonant peak as shown in Figure 3.5.

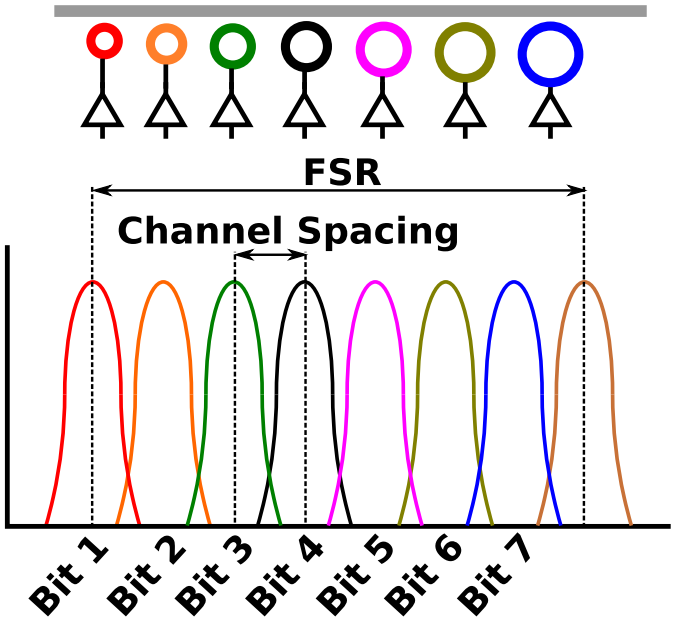


Figure 3.5: The FSR spacing between resonant peaks can be used to determine the amount of available WDM, which is influenced by three parameters: the FSR, FWHM and channel spacing between adjacent rings. Equation 3.4 describes how the level of achievable WDM is calculated.

Notice that to modulate more wavelengths the rings gradually become larger to change their resonant frequencies to be unique from the others. The system's FSR is dictated by the spacing between resonant peaks of the largest ring resonator that can't be used because of wavelength overlap. The channel spacing is chosen based on the desired level of optical power loss and is described further in Section 3.5.

The resonant frequencies of a ring with radius r and effective index of refraction n_{eff} are modeled using [57]:

$$\mathrm{m} * \lambda_{\mathrm{m}} = 2 * \pi * \mathrm{r} * \mathrm{n}_{\mathrm{eff}} \tag{3.3}$$

Here m is an integer representing the mode order of the resonant wavelength, and λ_m is the mode order's resonant wavelength. In Figure 3.6 we use this equation to obtain the system FSR as dictated by the largest ring modulator that will cause resonance overlap in the WDM link. To find n_{eff} we assume the ring resonator is

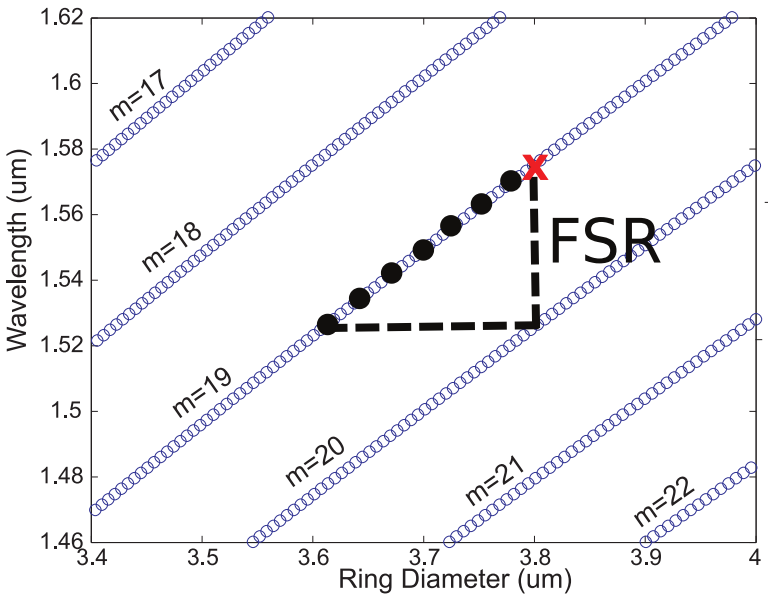


Figure 3.6: Equation 3.3 plotted across different sized ring resonators guided in single crystalline silicon. We show the range of wavelengths used in our WDM link and the system FSR, which is limited by the overlap of the m+1th mode of the largest (unused) ring on the mth mode of the smallest (used) ring.

fabricated in single crystalline silicon (Si) and cladded with silicon dioxide (SiO_2). We utilize the effective index method with a waveguide width of 450nm and height of 250nm to calculate the propagation coefficient as a function of wavelength, $\beta(\lambda)$. We model the wavelength dependent refractive indices of the guiding and cladding materials and calculate the effective index of refraction using $n_{eff} = \beta(\lambda)/k_o$. Here k_o is the vacuum wavevector and can be written as $2^*\pi/\lambda$ [54].

The diagram markings in Figure 3.6 show the wavelengths utilized in our WDM communication link and the FSR of the system, which is limited by the largest (unused) ring to avoid overlapping its m+1th mode on the mth mode of the smallest (used) ring. For the rest of the results in this section we assume our rings operate at the m=19 mode [57] with a resulting FSR of approximately 50nm. This FSR value dictates the amount of WDM that is achievable in the link using the following equation:

$$\text{WDM} = \frac{\text{FSR}}{\text{Multiplier} * \text{FWHM}} \tag{3.4}$$

ChannelSpacing describes the number of wavelengths between consecutive resonant peaks in Figure 3.5 and is given by the $Multiplier \times FWHM$ term. As the wavelength channels are brought closer together, the achievable level of WDM, and thus link bandwidth, increases. However, this comes at a cost of increased optical insertion loss and potentially nonlinearity induced loss. The latter is further described in Section 3.6 and is detrimental to the proper functionality of the system's ring resonators. The rest of this chapter is devoted to exploring these tradeoffs.

3.2 Tradeoffs in WDM and Optical Data Rate

Three nanophotonic device parameters dictate the total communication bandwidth through a link assuming a fixed FSR: the quality factor of the resonators, the channel spacing between resonant peaks and the modulation rate of each wavelength. In this section, we begin by examining the tradeoff associated with WDM and per channel data rate. In these results various channel spacing assumptions are also included to determine how they impact total link bandwidth. This section serves as a foundation for following sections by presenting a range of parameters. Each device assumption is separately analyzed to examine reasonable modulator and receiver data rates in scaled CMOS technologies, and also resulting optical power loss associated with channel spacing and thus WDM level.

Previous work [40] has shown that higher levels of signal attenuation occur in a ring resonator as the per channel data rate that passes through it grows. This property is shown in Figure 3.7 where high frequency sidebands in a 10Gb/s non-return-to-zero (NRZ) signal are attenuated by the resonator. Here the ring has a quality factor of 20,000 and a bandwidth of 9.6GHz [40]. The 3dB attenuated

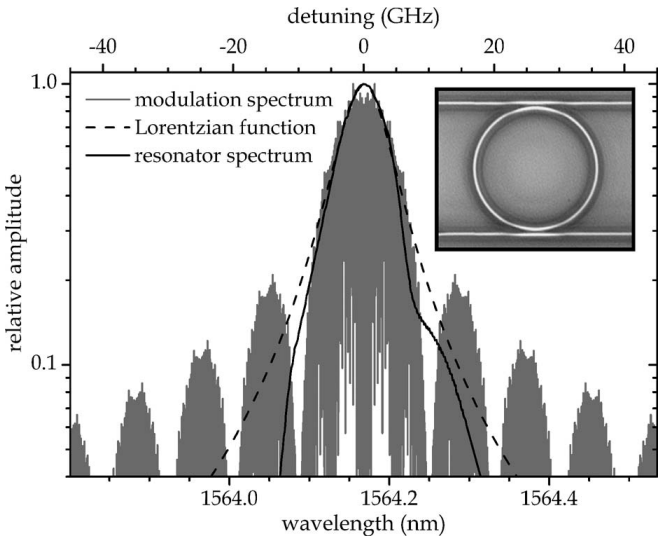


Figure 3.7: A fabricated ring resonator operating at a quality of 20,000 (9.6GHz bandwidth) with a 10Gb/s data rate signal being passed through it at one of its resonant wavelengths [40].

data rate of the incoming signal occurs at the bandwidth of the ring resonator divided by 0.75 [57]. For a bandwidth of 9.6GHz, this corresponds to a data rate of 12.8Gb/s. As the system's per wavelength data rate grows, the quality factor of the ring resonators must shrink to avoid excessive optical loss. The bandwidth of a ring, in hertz, can be found using its FWHM as:

$$\mathrm{BW}_{\mathrm{Ring}} = \frac{3 * 10^8}{\lambda_{\mathrm{o}}} * \frac{FWHM}{\lambda_{\mathrm{o}}} \tag{3.5}$$

The high frequency sidebands in Figure 3.7 extend further away from the resonator's operating wavelength as the data rate is increased. The reason for this is due to the Fourier components of the modulating square wave, which create the per wavelength on/off optical bits in the waveguide. The square modulation wave is composed of many high frequency components that are combined with the original (pre-modulated) optical data frequency, thus broadening its spectrum [40]. As the frequency of the square modulation wave grows, so do its number of higher

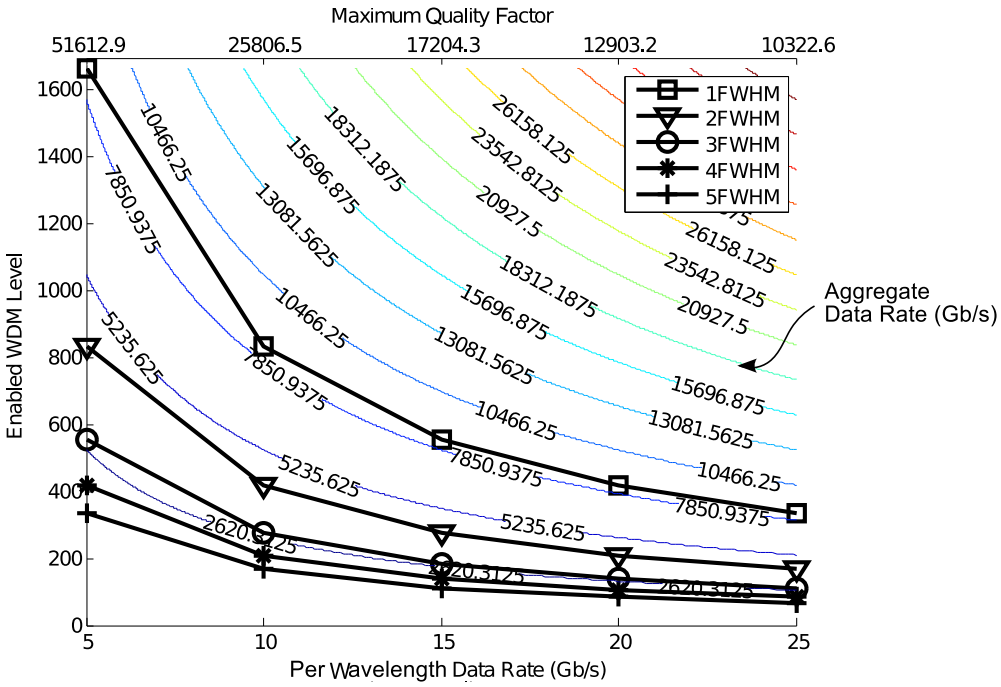


Figure 3.8: Tradeoffs in data rate versus required minimum ring resonator bandwidth. As the data rate is increased, the quality factor of a ring resonator must be lowered to avoid excessive attenuation of the signal. However, this also reduces the enabled level of WDM in the link. In the diagram we also show different channel spacing assumptions ranging from one to five FWHM lengths.

frequency components.

We demonstrate the tradeoffs associated with channel spacing, per wavelength data rate and ring quality factor in Figure 3.8. Here we choose a ring resonator diameter to obtain a center wavelength, λ_o, of 1550nm and assume the system FSR of 50nm found previously in Section 3.1.2. Along the x-axis we show various system data rates ranging from five to twenty-five Gb/s and along the top axis the maximum ring resonator quality factor to achieve each rate.

As the data rate is increased, the FWHM of the rings also increase by the same amount. Because the WDM of the system is inversely proportional to the FWHM from Equation 3.4, at a fixed channel spacing the total link bandwidth remains fixed. We calculate the total link bandwidth to be: $Data\ Rate \times WDM$

Level. Channel spacing assumptions ranging from one to five FWHM distances are shown [57]. As expected, a small spacing and thus more tightly packed channels provides a higher level of WDM. We explore the optical power tradeoff associated with tight channel spacings in Section 3.5.

3.3 Optical Ring Modulator

In this section, we present a performance and power model for carrier injection into a ring resonator used as a modulator. We assume that the ring is driven by a scaled CMOS inverter that is limited to providing a supply voltage less than the technology's Vdd. The data rate of the modulator (without the driver) is dictated by its carrier injection characteristics and is limited by the device carrier recombination lifetime. To improve the data rate, we present recent work in ion implantation for creating carrier recombination centers in the material, thus lowering the lifetime but at a cost of increased waveguide propagation losses due to absorption of light by the implants. However, we demonstrate that the upper bound on modulation rate is hampered by the scaled CMOS driver, which is unable to provide enough drive strength to the ring. Finally, we show power consumption projections using scaled transistor technologies across a range of ion implantation dosages.

3.3.1 Carrier Injection Model

Optical signal modulation and packet switching require a device that can be tuned and detuned to wavelengths traveling in a neighboring waveguide. When the ring is tuned, wavelengths corresponding to one of the resonant peaks are removed from the waveguide. Similarly, in detuned operation, the wavelengths are free to continue past the ring. Charge injection through electrical driving circuitry

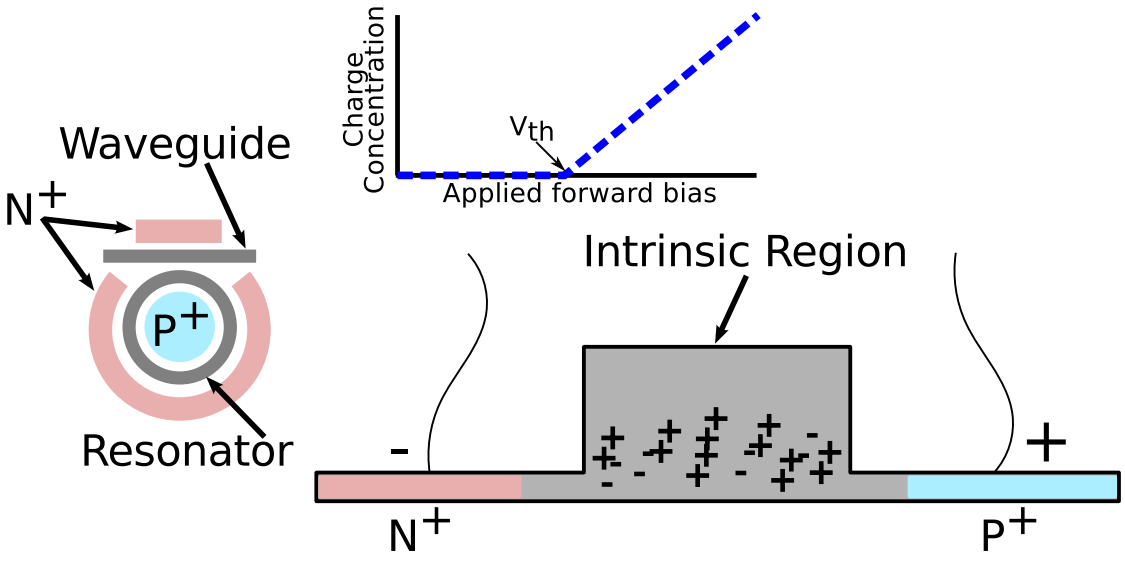


Figure 3.9: Charge injection into the ring resonator is accomplished by placing a PIN diode across the ring waveguide. The top view shows the P+ and N+ doped regions, where the ring corresponds to the intrinsic region. The diode is formed across a slab portion of the waveguide, which is shown in the lateral view. The silicon portion of the waveguide is extended outwards for doping. The diode can be modeled as a series resistor, where the amount of steady state charge after a forward driving voltage of V_{th} in the ring rises linearly and is equal to $I_{diode} \times \tau_c$.

performs the tuning. A PIN diode is fabricated around the resonator such that the ring acts as the intrinsic region. As a forward driving voltage is placed across the contacts of the diode, charge carriers are injected into the ring. This causes a blueshift in the resonant peaks (i.e., shift to lower wavelengths). When the forward driving voltage is removed, the ring relaxes back to its original state.

Figure 3.9 shows a top and lateral view of a resonator with a PIN diode placed across its waveguide. To form the diode regions both highly doped P and N type silicon are placed on a slab close to the ring. The slab can be seen in the lateral view, where the silicon forming the ring is partially extended outwards.

The current in the diode can be approximated by the following equation [70]:

$$\mathrm{I_{diode}} = \frac{\mathrm{V_{drive}} - \mathrm{V_{th}}}{R} \tag{3.6}$$

Where R is the series resistance of the diode which is largely dominated by the

contact resistance (5-100Ω) [42] [70]. V_{drive} is the forward bias placed across the diode, and V_{th} is the threshold of the diode (0.5-0.7V) [44] [70]. The level of steady state charge injected into the ring is described by [70]:

$$Q_{injected} = I_{diode} * \tau_c \tag{3.7}$$

Here τ_c is the carrier recombination lifetime of the device, which depends on its material composition. This is an important equation as it dictates the amount of drive voltage that must be applied across the ring to build up a specific amount of charge, $Q_{injected}$, in the intrinsic region. As more charge is injected into the ring (i.e., $Q_{injected}$ grows), its resonant peaks continue to shift. The optical transmission of a passing wavelength out the Through Port (i.e., the percentage of power that does not couple into the ring resonator) dictates the required quantity of charge injection. If, for example, a wavelength couples into the ring when no driving voltage is applied, the power out the Through Port is close to zero. However, as a voltage is applied and charge is injected, more power from the wavelength leaves out the Through Port instead of coupling into the ring. $Q_{injected}$ in this case must be high enough so that this power is close to 100% of the power held in the wavelength. The transmission characteristics of a single crystalline silicon ring resonator are altered due to the following change in refractive index caused by $Q_{injected}$ [42]:

$$\Delta n = -[8.8 * 10^{-22} * \Delta N + 8.5 * 10^{-18}(\Delta P)^{.8}] \tag{3.8}$$

Where ΔN (cm^{-3}) is the electron concentration change in the ring resonator and ΔP (cm^{-3}) is the hole concentration change. From this equation the total quantity of charge injected can be written as a function of ΔN and ΔP:

$$\mathrm{Q_{injected}} = (\Delta \mathrm{N} + \Delta \mathrm{P}) * \mathrm{q} * \mathrm{Volume}_{Ring} \tag{3.9}$$

Here q is the elementary charge of $1.60217646 \times 10^{-19}$ coulombs and Volume_{Ring} is the volume of the optical ring resonator, minus the doped regions forming the diode. Based on these equations it's evident that as a ring resonator's volume increases the required $\mathrm{Q}_{injected}$ also increases. Similarly, if the FWHM of the ring is large, more charge injection will also be required than for a ring with a smaller FWHM. This is because the resonance of the former ring will have to be wavelength shifted more to avoid excessive optical power loss. A large value of $\mathrm{Q}_{injected}$ forces the drive voltage, V_{drive}, across the resonator to grow over the required voltage for a smaller $\mathrm{Q}_{injected}$. Using Equations 3.6 and 3.7, V_{drive} can be written as:

$$\mathrm{V_{drive}} = \mathrm{Q_{injected}} * \frac{R}{\tau_c} + \mathrm{V_{th}} \tag{3.10}$$

Depending on the calculated value of V_{drive}, a CMOS process may not be able to supply enough voltage. In this case, resonator insertion losses will grow as the value of $\mathrm{Q}_{injected}$ falls below what is required. This is because the ring resonator will not be able to obtain enough modulator depth or shift to either completely extinguish the light from the waveguide or allow it to travel past. Ring resonator insertion loss are further discussed in Section 3.5.

When a forward bias voltage is applied across the diode, charge builds up in the intrinsic region over time. Similarly, when the bias is removed the charge will quickly recombine and disappear. Charge build up and discharge are respectively modeled using [70]:

$$\frac{dQ}{dt} = \frac{\mathrm{V_{drive}} - \mathrm{V_{th}}}{\mathrm{R}} - \frac{\mathrm{Q}}{\tau_c} \tag{3.11}$$

$$\frac{dQ}{dt} = \frac{-\mathrm{V_{th}}}{\mathrm{R}} - \frac{\mathrm{Q}}{\tau_{\mathrm{c}}} \tag{3.12}$$

From these equations its possible to derive the time required to charge and discharge the diode. In the system space this is the operational latency to turn on and off the device and is directly related to the carrier recombination life, τ_c, of the underlying material of the device. The photon lifetime of the modulator can be written as $\mathrm{Q}\times\lambda/(2\times\pi\times\mathrm{c})$ where Q is the quality factor of the ring, and c is the speed of light in a vacuum. τ_{ph} represents the fundamental latency of the ring modulator for the optical field inside of it to build up or decay down. The value of τ_{ph} has been shown to be on the order of only a few ps [42] and thus the data rate of the device is dominated by the carrier injection properties. We assume that when the resonator is turned off, its voltage is simply removed, rather than applying a negative bias. As a result, the latency to inject carriers into the ring is larger than the time required to remove them. For simplicity, however, we assume the longer of the two to be the modeled turn-on and turn-off delay of the device:

$$\mathrm{Latency_{on,off}} = \frac{2.3 * \tau_{\mathrm{c}}}{2} \tag{3.13}$$

Various work has examined how to improve the latency of turning on and off the ring through pre-emphasis techniques [70], smart biasing [44] and applying negative biases to more quickly remove the charge from the intrinsic region [70]. In this work we do not adopt these techniques since they may require negative or higher voltages than a technology's Vdd and potentially complex timing schemes. These directly impact the driving circuitry, which are examined in Section 3.3.3.

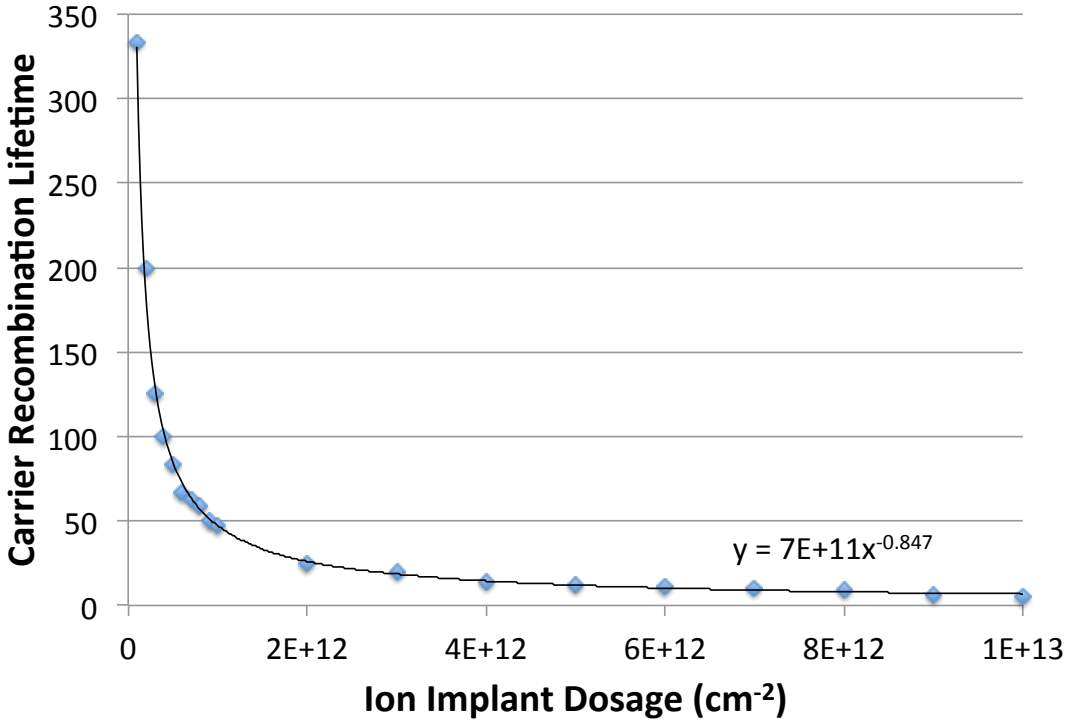


Figure 3.10: Carrier recombination lifetime reduction in single crystalline silicon from implanting oxygen ions [66]. As more ions are implanted the carrier lifetime reduces to below 10ps. However, this comes at a cost of increased propagation loss in the waveguide due to added optical absorption by the oxygen ions.

3.3.2 Reducing τ_c with Ion Implantation

The free carrier lifetime of the ring resonator can be significantly reduced by oxygen ion implantation [66]. The drawback of this technique is the resulting propagation losses generated (due to increased absorption of light) in the waveguide from the implants, which can be minimized if the irradiation energy and dosage are correctly chosen. Single crystalline silicon has a carrier recombination lifetime, τ_c, of approximately 450 ps [55]. Using experimentally demonstrated reductions in τ_c and increases in waveguide propagation loss from [66], we assume the parameters in Figures 3.10 and 3.11 for our results.

Although a substantial reduction in τ_c is possible with ion implantation, large propagation losses begin to accrue at higher implant dosages. Eventually the im-

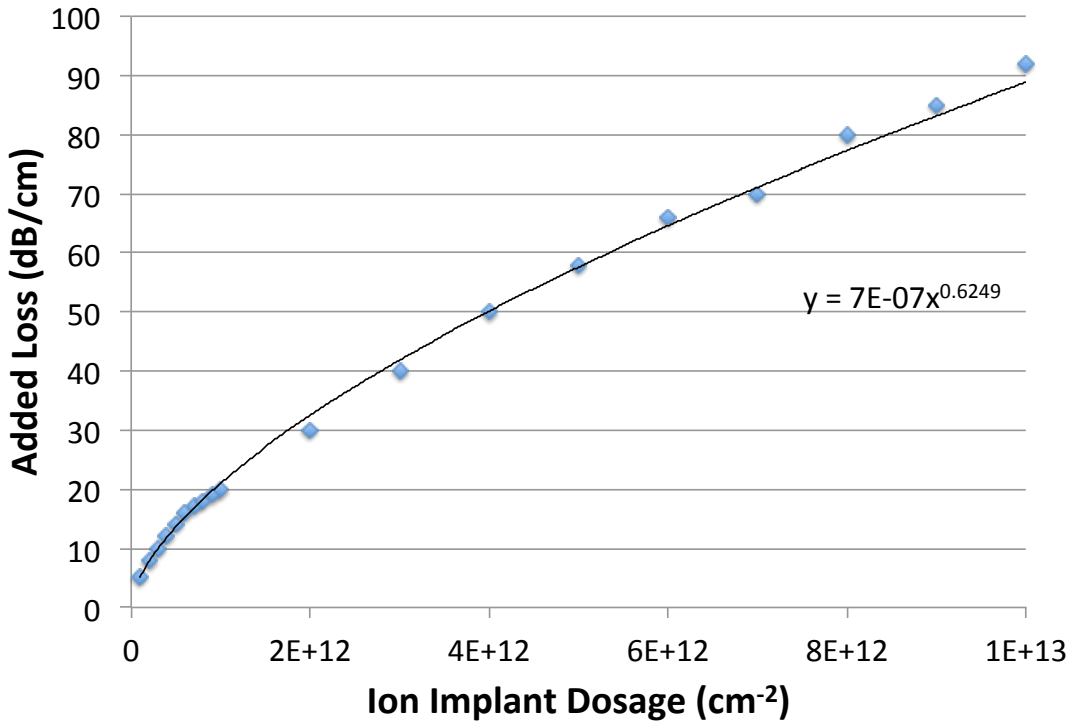


Figure 3.11: Increasing the oxygen ion dosage in silicon decreases its free carrier lifetime at the cost of increased propagation loss. This loss arises from increased absorption of the optical signal by the oxygen ions.

provements in τ_c saturate. To model this behavior, we augment the ring resonator and carrier injection equations shown previously to demonstrate how performance and power consumption are influenced by this technique.

3.3.3 Driver Model

In this section we assume carrier injection into a ring resonator is performed with an inverting driver circuit, which is shown in Figure 3.12. The inverter drives the voltage that turns the resonator on and off. This serves to either modulate electrical data into the optical domain, or as an optical switching element controlled electrically. Also shown in the figure is the equivalent RC circuit model used for analyzing delay and power consumption. Here the parameter R_{on} is the minimal

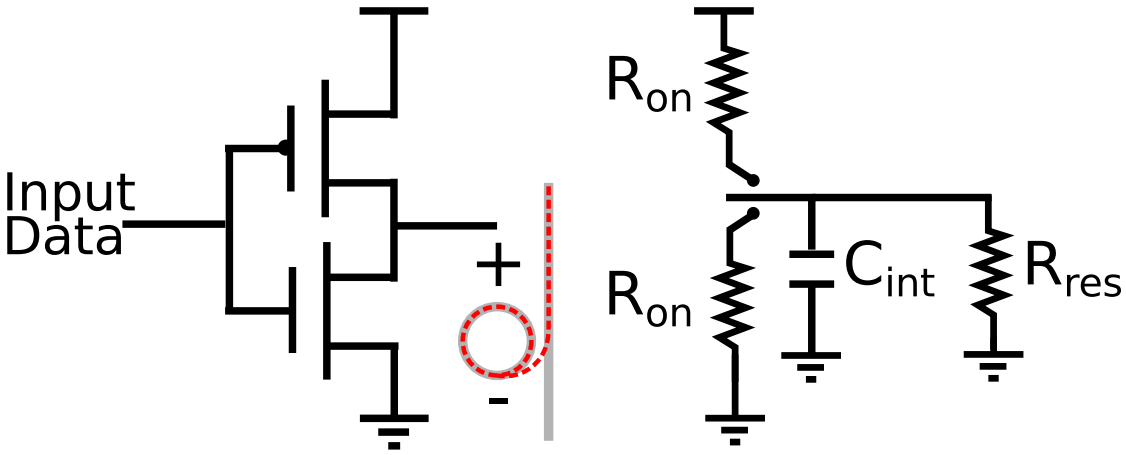


Figure 3.12: The ring resonator driver consists of a properly sized CMOS inverter with the ring resonator load. The voltage required by the ring resonator is based on its size and FWHM characteristics. The driver can be modeled using RC analysis with the assumption that each transistor has a specific on resistance, denoted as R_{on}. Under GHz frequencies the PIN diode across the ring is modeled as a resistance [70]. Thus, the capacitive load is the driver's intrinsic capacitance. The resistance of the resonator, R_{res}, is dominated by its contact resistance.

sized resistance of a CMOS transistor in saturation and C_{int} is the corresponding intrinsic capacitance of the minimally sized inverter. Under GHz frequencies, the PIN diode across the ring resonator can be treated as a resistive load, R_{res} [70]. The value of R_{res} is largely dominated by the contact resistance connecting to the diode (i.e., the electrical vias and metal wiring).

In this section, we examine the power and performance of the driver across multiple scaled CMOS technology nodes. The key parameters used in our results are the scaled technology size, the supply voltage, saturation current of a minimally sized transistor, gate oxide thickness and overlap capacitances. Table 3.1 shows the values of these parameters for 29, 20, 15.3 and 10.7nm technologies which were obtained from the 2009 ITRS [27].

The required drive voltage of the resonator, denoted as V_{drive} from Equation 3.10, may not necessarily be feasible depending on the available supply voltage for a technology node. Thus, we define V_{res} to be the actual voltage that the in-

	Tech (nm)	Vdd	Idsat (uA/nm)
$\bowtie$	29	1	0.83
$\prec$	20	0.87	1.45
ω	15.3	0.78	1.78
✓	10.7	0.68	2.1
	T_{ox} (nm)	$Coverlap_{nmos}$ (fF)	$Coverlap_{pmos}$ (fF)
$\bowtie$	1.32	0.041	0.036
$\prec$	0.95	0.039	0.034
ω	0.753	0.038	0.033
✓	0.551	0.036	0.031

Table 3.1: CMOS transistor scaling parameters [27].

verter can deliver to the resonator as:

$$\mathrm{V_{res}} = \min(\mathrm{P_{supply}} * \mathrm{Vdd}, \mathrm{V_{drive}}) \tag{3.14}$$

The parameter P_{supply} specifies the percentage of supply voltage, Vdd, that is placed across the ring resonator. In this work, we choose P_{supply} to be 0.9. This strikes a good balance between driver delay and power consumption, and the resulting optical power loss from the inability to supply the full V_{drive}. To achieve a voltage of V_{res}, the transistors in the inverter must be sized appropriately. We calculate the required transistor resistance, R_{on}, using the following equation:

$$\frac{\mathrm{Vdd}}{\mathrm{R_{on}} + \mathrm{R_{res}}} = \frac{\mathrm{V_{res}}}{\mathrm{R_{res}}} \tag{3.15}$$

Once R_{on} has been determined, calculating the proper transistor sizing is straightforward. The *Size* parameter is the absolute required width of the transistor (in nm).

$$\frac{\mathrm{Vdd}}{\mathrm{Size} * Idsat} = \mathrm{R_{res}} \tag{3.16}$$

Following the calculation of the required transistor resistance and corresponding sizing factor, the intrinsic capacitive load of the driver can be determined. In this

work, we assume that the intrinsic load of the inverter is equal to half the total gate capacitance [30]. We can calculate the total inverter gate capacitance using the parameters from Table 3.1:

$$\mathrm{C_{ox}} = \frac{\varepsilon}{\mathrm{T_{ox}}}(\mathrm{F/nm^2}) \tag{3.17}$$

C_{ox} is the capacitance per nm^2 due to the gate oxide thickness and dielectric. The SiO_2 dielectric has a permittivity of $\epsilon = 3.5{\times}10^{-20}$ F/nm. The gate to channel capacitance, C_{gc}, of a minimum sized transistor is approximated from the value of C_{ox} through:

$$\mathrm{C_{gc}} = \mathrm{Tech}^2 * \mathrm{C_{ox}} \tag{3.18}$$

Where *Tech* is the scaled CMOS processor technology (in nm).

Finally the total gate capacitance of the NMOS and PMOS transistors is a combination of the gate to channel capacitance, C_{gc}, and the overlap capacitances of the gate to source and gate to drain:

$$\mathrm{Cg_{nmos}} = 2 * \mathrm{Coverlap_{nmos}} + \mathrm{C_{gc}} \tag{3.19}$$

$$\mathrm{Cg_{pmos}} = 2 * \mathrm{Coverlap_{pmos}} + \mathrm{C_{gc}} \tag{3.20}$$

The total gate capacitance looking into the minimum sized inverter is estimated based on the gate capacitances of the NMOS and PMOS:

$$\mathrm{Cg_{inverter}} = \mathrm{Cg_{nmos}} + \mathrm{Cg_{pmos}} \tag{3.21}$$

Sizing the transistors in the driver changes the value of the intrinsic and gate capacitances through a multiplication by the relative sizing value. Thus, we cal-

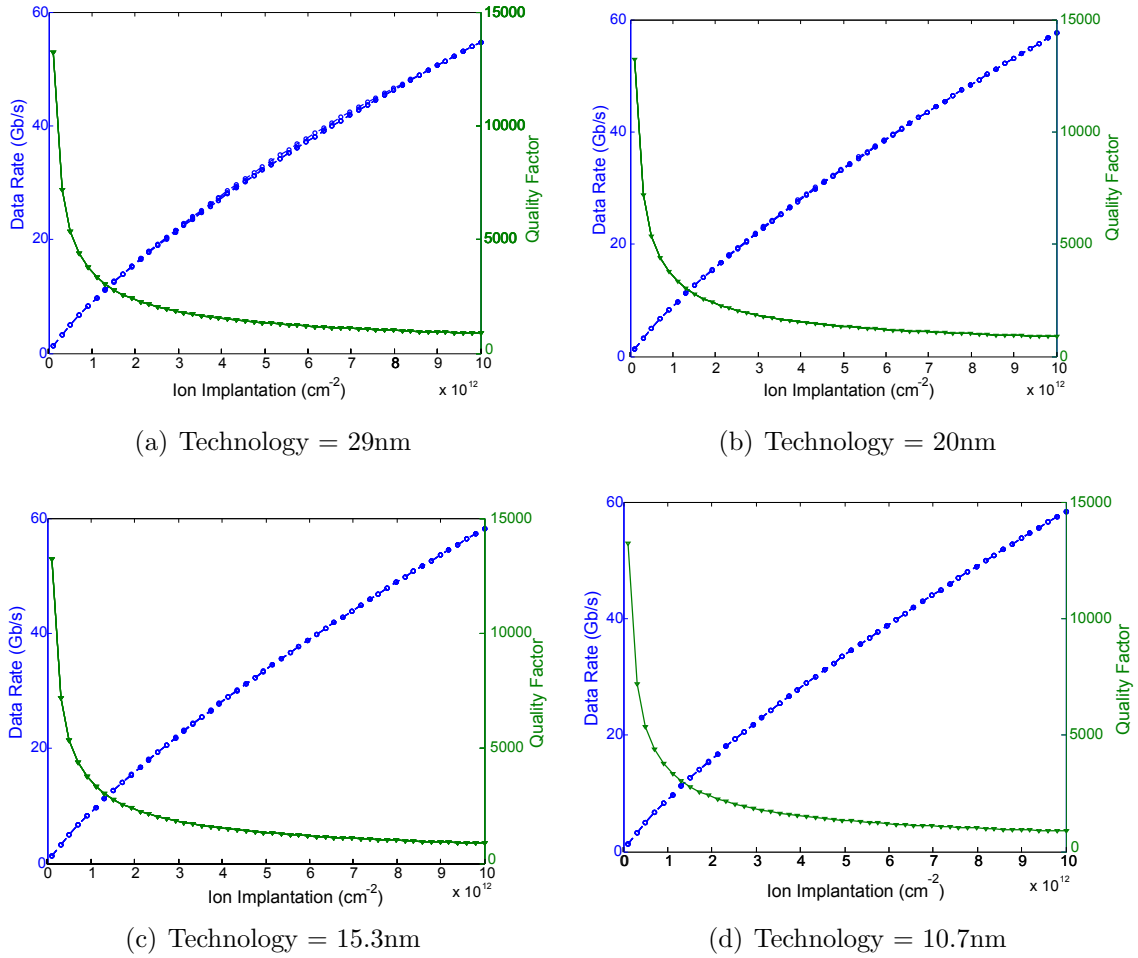
culate the model parameter, C_{int}, from Figure 3.12 to be:

$$C_{\text{int}} = \frac{1}{2} * Cg_{\text{inverter}} * \frac{\text{Size}}{\text{Tech}} \tag{3.22}$$

Where the *Size/Tech* calculation is the relative transistor sizing factor in the driver.

Using the projected carrier recombination lifetime improvements as a function of the ion implantation dosage and associated propagation loss from Figures 3.10 and 3.11, we generate performance results for 29, 20, 15.3 and 10.7nm CMOS technologies. These results are shown in Figure 3.13. For each technology, resonance shift amounts, $\Delta\lambda_o$, ranging from one to five FWHM are plotted. These shift amounts represent how far the ring's resonance peaks are moved to lower wavelengths of light (i.e., blue shifted). The difference in achieved data rate between the values of $\Delta\lambda_o$ within the same technology are negligible. As more ion implants are added to the ring resonator, the combined driver + resonator delay falls, reaching close to 60Gb/s. However, this comes at a cost of reduced ring quality factor, which is only dependent on the ion implantation dosage. Between the different technologies, the driver latency is primarily dominated by the resonator latency, and only slight improvements in data rate can be seen in going from 29 to 10.7nm at the highest dosage.

The voltage supply reduces as the technology node shrinks from 29 to 10.7nm according to Table 3.1. Because of this, it may be impossible for a driver to power a ring resonator with high ion implantation dosage, since the resulting quality factors of the ring will be very small. A smaller quality factor means that a larger voltage needs to be applied across the ring to shift it. This problem is exacerbated by the resonance shift amount, which is varied from one to five FWHM. In Figure 3.15 we plot the saturation points (i.e., when the required voltages becomes too large for the driver) of each technology node as a function of the total resonance shift amount, $\Delta\lambda_o$. As the shift amount grows, the highest ion implantation dosage that



(a) Technology = 29nm

(b) Technology = 20nm

(c) Technology = 15.3nm

(d) Technology = 10.7nm

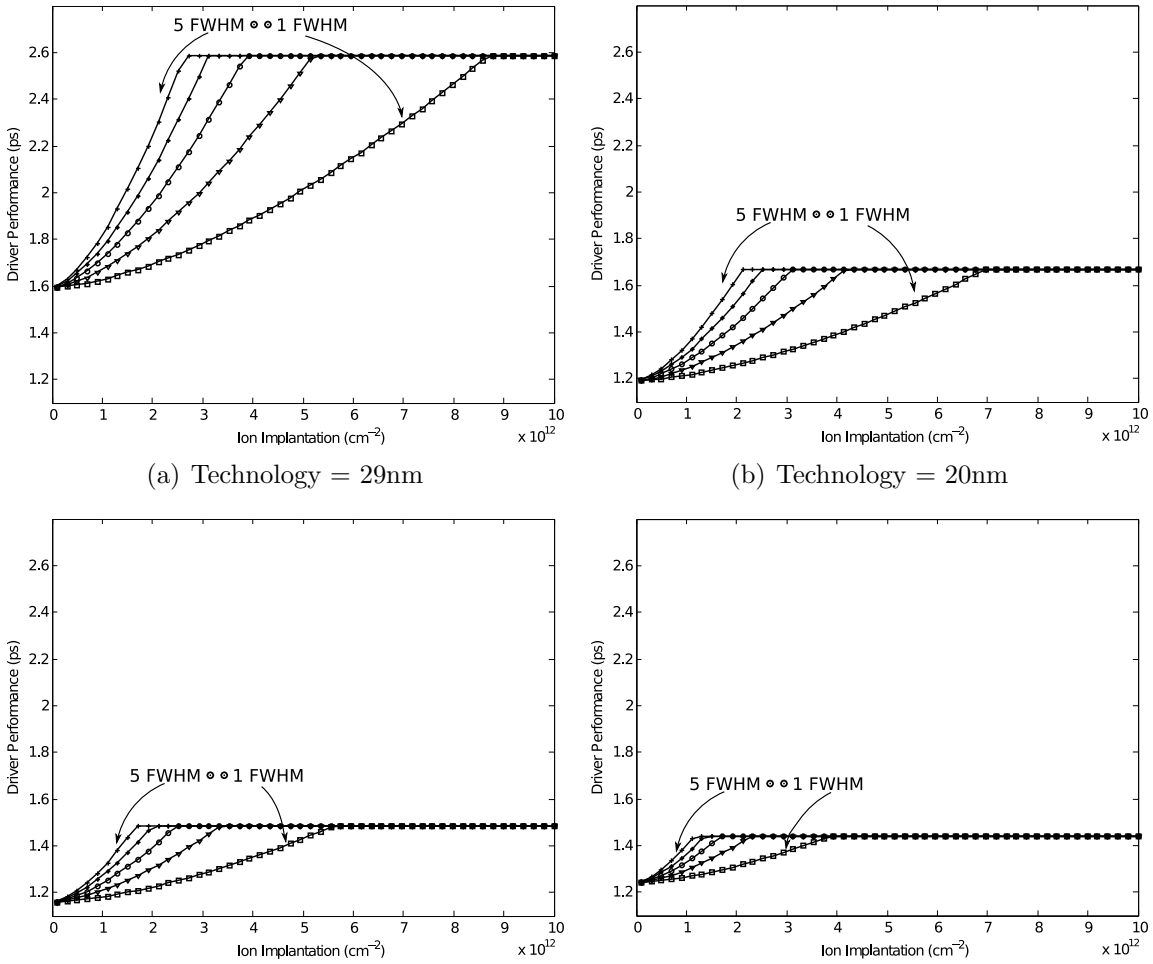
Figure 3.13: Ring modulator performance results for 29, 20, 15.3 and 10.7nm technology. Adding more ions to the ring resonator causes its quality factor to degrade due to increasing propagation losses. This is shown by the green triangle line, where implants above $1{\times}10^{12}$ cm^{-2} reduce the ring modulator quality factor to less than 5,000. The other blue line indicates the total modulator performance (driver circuitry + resonator activation/deactivation). This line is actually composed of multiple lines showing the difference in modulator bandwidth at different resonance shift amounts ranging from one to five FWHM. However, the difference in driver latency across these design points is negligible.

can be used shrinks. As technology scales, the inability to drive a ring with a high concentration of ions gets worse, since the technology's supply voltage decreases.

The driver delays without the ring response times from Equation 3.13 are shown in Figure 3.14. Depending on the value of $\Delta\lambda_o$ and technology node, the driver eventually saturates at a particular ion implantation dosage. As technology scales and the Vdd supply voltage continues to reduce, the saturation dosage comes increasingly earlier to the point where at 10.7nm, the driver latencies are degraded over 15.3nm. This is due to the large relative sizing factor of the transistors needed to provide enough current to the ring (even without ion implants) and the resulting increase in intrinsic capacitance, which causes an increase in delay.

Using the maximum ion implantation dosages from Figure 3.15 that can be successfully driven by a technology node at a particular $\Delta\lambda_o$, the data rate results from Figure 3.13 can be used to extract the resulting maximum supported data rate of the driver + ring transmitter as shown in Figure 3.16. It's evident from the results that larger technology nodes are able to achieve a higher data rate because of their greater voltage supply. Similarly, as the amount of resonance shift increases, the required drive voltage across the resonator also increases. A high quality factor, and thus less ion implantation, is desirable in smaller technology nodes where the limited voltage supply sometimes makes it difficult to provide the required V_{drive}.

Overcoming this problem requires a separate voltage supply or charge pump and transistors designed to operate at voltages larger than scaled Vdd's. Assuming this is possible, more ion implantation can be used for higher data rate. Additionally, this also enables pre-emphasis switching that avoids the added propagation loss of ion implants, and resulting reduction in WDM, through fast carrier injection using a voltage spike. However, this comes at a cost of increased driver complexity.



(a) Technology = 29nm

(b) Technology = 20nm

(c) Technology = 15.3nm

(d) Technology = 10.7nm

Figure 3.14: The inverting driver performance across the scaled technology nodes from Figure 3.13. Notice that the ring resonator response times dominate the small driver latencies. Depending on the technology, the driver performance saturates at different ion implantation dosages when it can no longer deliver enough supply voltage to the ring.

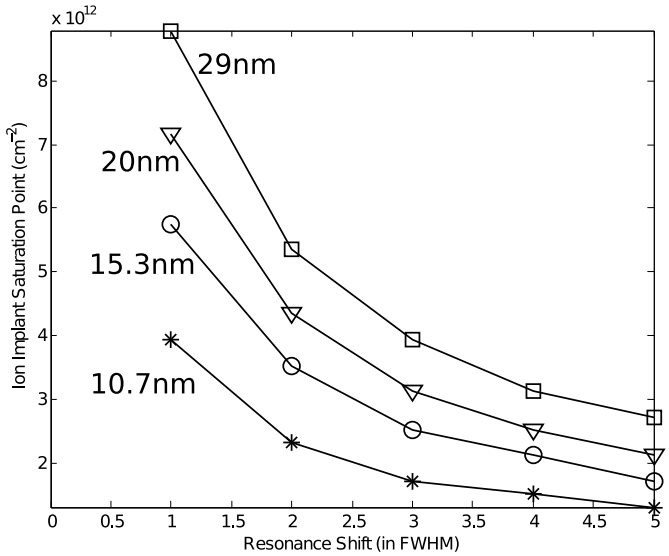


Figure 3.15: More charge injection is required as a ring's FWHM grows or the distance at which it has to shift increases. As the required $Q_{injected}$ increases, the voltage which must be applied across the ring to obtain that charge must also increase. In this graph, we show four scaled CMOS technology nodes and the first ion implantation dosage that requires a drive voltage higher than the supply voltage of the driver. As the shift distance increases from one to five FWHM, the maximum ion dosage that can be driven degrades since more charge injection is required.

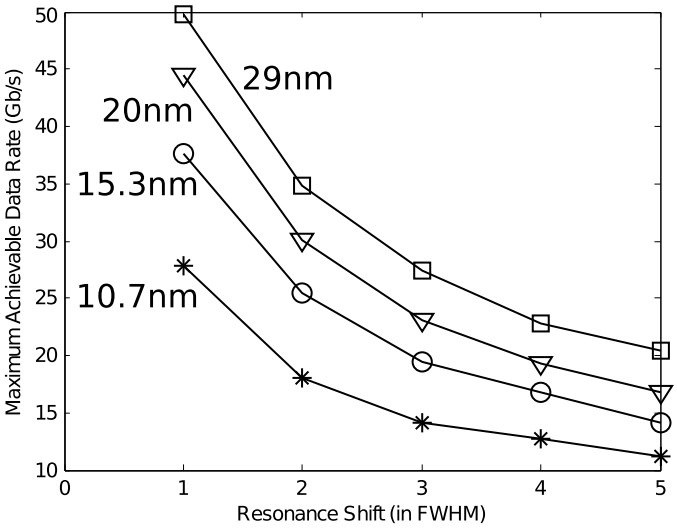


Figure 3.16: Using the maximum achievable ion implantation dosages across scaled technologies and resonance shifts in Figure 3.15, we extract maximum enabled data rates from Figure 3.13. Older technology nodes are able to provide better data rates because of their larger voltage supply and thus larger ion implantation dosages.

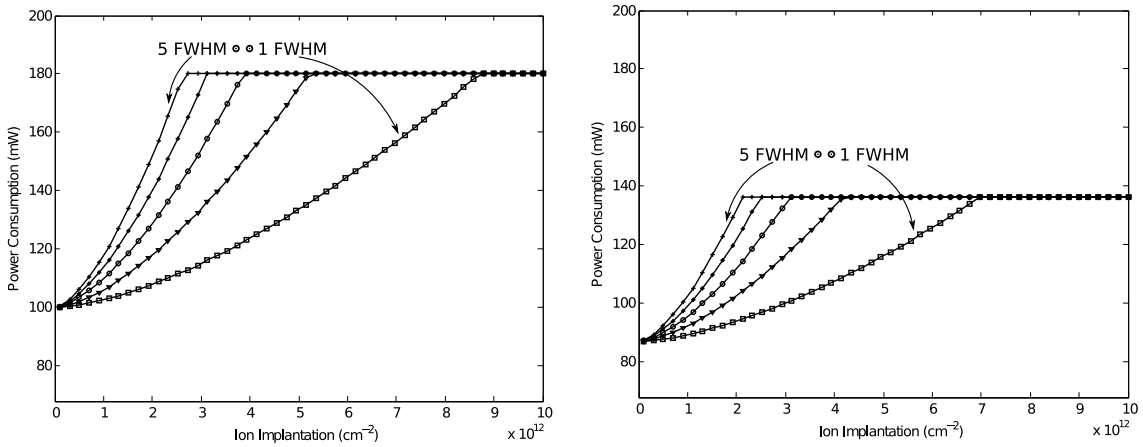
Alternative methods to ion implantation for turning on and off a ring resonator are examined in Chapter 2.

Power consumption results for the different technology nodes are shown in Figure 3.17. Across all ion implantation dosages, the power consumption of smaller technology nodes is lower. Prior to saturation this is due to smaller nodes being able to offer the ring modulator the same current as the older technologies at a reduced voltage supply. However, the figures also demonstrate that the smaller technologies for a particular $\Delta\lambda_o$ saturate at lower ion implantation dosages. This is again due to their lower power supply which at higher implantation dosages, and thus lower quality factors, is unable to provide enough voltage to fully shift the ring's resonances.

As the ion implantation in the ring is reduced, the switching data rate is also approximately reduced by the same factor as shown in Figure 3.13. However, due to the nonlinear dependence of resonance shift on injected carrier concentration from Equation 3.8, the resulting power reduction factor is less. Another tradeoff is the WDM level, which increases as ion implantation reduces because of the resulting jump in ring quality factor. Therefore, based on the requirements of a nanophotonic interconnect, the proper level of ion implantation should be carefully chosen to balance these tradeoffs.

3.4 Optical Receiver

In this section, we begin with a performance and power consumption model of the photodetector for converting light in a waveguide to an electrical signal. This device uses photons to generate free charge carriers that serve as the input to front-end amplifying stages. The amplifiers translate the photodetector input signal to a digital voltage level which can be subsequently used by the destination node. In



(a) Technology = 29nm

(b) Technology = 20nm

(c) Technology = 15.3nm

(d) Technology = 10.7nm

Figure 3.17: Ring modulator power results for 29, 20, 15.3 and 10.7nm technology. As ion implantation dosage increases, more power is expended by the resonator driver. Similarly, as a larger resonance shift is required, a greater V_{drive} must be supplied. Depending on the resonance shift amount, the driver will be unable to provide enough voltage to the ring, thus saturating its power consumption.

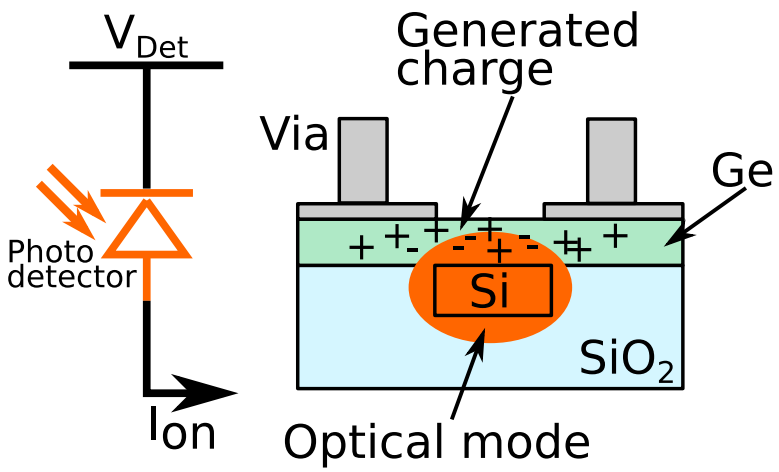


Figure 3.18: Single crystalline germanium detector based on [12] [13]. The detector is biased at a voltage high enough to cause velocity saturation in the electron and hole charge carriers (0.6V). A single crystalline silicon waveguide is fabricated below the germanium detector. The power from the optical mode in the waveguide excites charge carriers in the germanium, which are swept across the electrical field created by the bias voltage. The waveguide is assumed to be surrounded by a silicon dioxide cladding material. The photocurrent, denoted as I_{on}, supplies a series of amplifier stages that inflate the signal to a digital-level output voltage.

this section, we examine the performance, power consumption and bit-error-rate (BER) of an inverter based receiver across scaled CMOS technologies. We conclude with a BER analysis that examines the probability of encountering undetectable errors in a cache line sized packet using bit parity.

3.4.1 Photodetector

An optical photodetector converts photons traveling in the waveguide to an electrical current, which is further amplified to a digital voltage level. In this dissertation, we assume single crystalline germanium based detectors since they can be easily bonded above silicon based waveguides [12] [13]. The detector design is shown in Figure 3.18. In a WDM system each wavelength will have its own photodetector that transforms the light into a current, denoted as I_{on}. Silicon waveguides surrounded by a silicon dioxide cladding carry the optical signal to the germa-

Detector Parameter	Value
Velocity Saturation Bias (V_{det})	0.6V
Electron Velocity Saturation	6×10^6 cm/s
Hole Velocity Saturation	6×10^6 cm/s
Length (L)	$30u$m
Inter Contact Gap (t)	450nm
Per Contact Width (D)	350nm
Detector Responsivity	0.44A/W
Detector Dark Current	10^{-7} Amps

Table 3.2: Germanium photodetector parameters.

nium portion of the detector, which sits above the waveguide. In this region, light surrounding a center wavelength of 1550nm is energetic enough to overcome the bandgap energy of germanium, exciting charge carriers. These electrons and holes are quickly swept to the contact vias through an applied detector bias, denoted as V_{det}.

The detector bias is chosen such that the electron and hole velocities are saturated. This enables maximum performance by minimizing the amount of time required for the electrons and holes to drift through the germanium to one of the contact terminals. Based on [13] we assume a saturation V_{det} of 0.6V; however, this could be further improved by optimizing the detector geometry or doping the contact regions to form PIN diodes.

In this chapter, we examine single crystalline silicon waveguides for light propagation, which are approximately 450nm wide for single mode operation [26]. This width, denoted as t, forms the inter contact gap of the detector (i.e., the distance between the two metal terminals). The latency response of the detector can be calculated as a function of t using the following equation:

$$\text{Risetime}_{detector} = \frac{t * \chi}{2 * V} \tag{3.23}$$

Here V is the velocity saturation of holes and electrons, and χ is referred to as

the carrier drift distance corrective coefficient [2]. We find the value of χ to be 2.4 based on comparison with [12].

The capacitance of the detector, denoted as C_{det}, impacts the total performance of the receiver including the amplifiers since it adds additional input capacitance to the transimpedance stage. This capacitance is a function of the total length of the device, L, the number of contact terminal pairs (here we only use one), the permittivity of germanium and an experimentally determined parameter η, defined below. Using these values C_{det} is calculated as [2]:

$$C_{\text{det}} = .226 * N * L * \epsilon_{\text{o}} * (\epsilon_{\text{s}} + 1) * (6.5 * \eta^2 + 1.08 * \eta + 2.37) \tag{3.24}$$

where the η parameter is a function of the width of each contact, D, and the distance between them, t:

$$\eta = \frac{\text{D}}{\text{t}} + \text{D} \tag{3.25}$$

Using Equation 3.23 and assuming a silicon based waveguide, the latency of the detector is calculated using [11]:

$$\text{Latency} = 0.315 * \text{Risetime}_{\text{detector}} \tag{3.26}$$

The bandwidth of the detector is related to $\text{Risetime}_{detector}$ as [11]:

$$\text{BW}_{\text{detector}} = \frac{0.35}{\text{Risetime}_{\text{detector}}} \tag{3.27}$$

Previous work has determined the data rate of a non-return-to-zero (NRZ) signal from a detector's bandwidth to be $0.7 \times \text{BW}_{detector}$ [30]. Assuming a silicon based waveguide width of 450nm, the detector rise time is calculated from Equation 3.23 to be approximately 9ps. This equates to a latency of 2.84ps and a

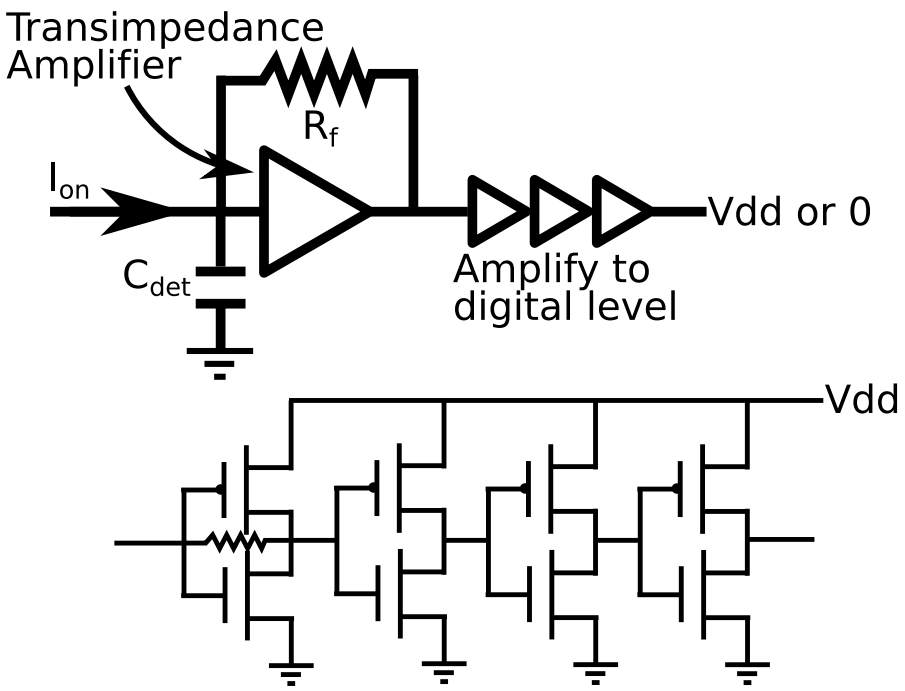


Figure 3.19: The optical receiver uses the photodetector current, I_{on}, as input into a transimpedance amplifier. The feedback resistance, R_f, self-biases the transimpedance stage at Vdd/2, and as a result, the amplifier stages following it. The detector capacitance is denoted as C_{det}. The amplifiers following the first stage further inflate the signal to a digital voltage level. Each amplifier is implemented using an inverter, where the first differs from the rest because of the feedback resistance.

maximum data rate through the detector of 39Ghz. The data rate of the full receiver is determined by the minimum of the detector and amplifying stages. While the detector does consume static power due to dark current, this is negligible compared to the power consumption of the following transimpedance and digital amplifying stages.

3.4.2 Front-End Receiver Components

In this section, we present a receiver model for analyzing power and performance tradeoffs based on [30]. Following the detector, the photocurrent I_{on} is fed into a transimpedance amplifier. Here the current is converted to an amplified voltage and further inflated by a series of inverting stages to a digital level. The complete

receiver architecture is shown in Figure 3.19. Each amplifier is implemented as an inverter with the transimpedance stage using a feedback resistance, denoted as R_f. This feedback effectively biases the inverters at Vdd/2, which maximizes the gain of each stage. The capacitance from the detector is added to the model and is denoted in the figure as C_{det}.

3.4.3 Spectral Bandwidth

The gain from the transimpedance amplifier is a function of the feedback resistance, R_f, the total output resistance, R_o, and transimpedance, g_m. The output resistance is the parallel combination of the NMOS and PMOS R_{on} values, and the total transimpedance is the sum of the two transimpedances of each transistor:

$$\text{Transimpedance}_{\text{gain}} = \text{R}_\text{o} * \frac{(\text{g}_\text{m} * \text{R}_\text{f} - 1)}{\text{R}_\text{o} + \text{R}_\text{f}} \tag{3.28}$$

Under the assumption that the bandwidth constraint of the receiver is dominated by the input pole of the transimpedance stage [30], the total spectral bandwidth of the receiver can be estimated using:

$$\text{BW}_{\text{Receiver}}\ (\text{Hz}) = \frac{\text{Transimpedance}_{\text{gain}} + 1}{2 * \pi * \text{R}_\text{f} * \text{C}_\text{t}} \tag{3.29}$$

Here C_t is the total capacitance looking into the front-end of the receiver. This includes the parallel combination of C_{det}, the gate capacitance looking into the transimpedance stage, and the gate to drain overlap capacitances of the two transistors in that amplifier. We denote the total gate to drain capacitance in this first stage as C_f. Similar to the previous detector analysis from Section 3.4.1, the maximum data rate of the receiver is estimated using its bandwidth as: $0.7\times\text{BW}_{Receiver}$ [30]. The latency of the receiver can also be calculated using $\text{BW}_{Receiver}$ [11]:

$$\text{Latency}_{Receiver} = \frac{0.7}{2 * \pi * \text{BW}_{\text{Receiver}}} \tag{3.30}$$

Using this equation, a receiver with a spectral bandwidth of 25GHz achieves a latency of 4.5ps and at 50GHz this reduces to 2.2ps. Previously we calculated the latency of the germanium based detector using Equation 3.26 to be 2.84ps assuming a silicon based waveguide with width 450nm. Thus, the total delays of the receiver circuitry assuming 25 and 50Ghz spectral bandwidth are 7.34ps and 5.04ps, respectively.

3.4.4 Noise Model and BER

The second important characteristic of the amplifier circuit is its bit-error-rate (BER), which denotes the number of detection errors per bit. Obviously, a lower BER is better, and receivers in the literature target anywhere from 10^{-15} to 10^{-18} [1]. In this section, we adopt the error model proposed in [30] that takes into account the following noise sources in the transimpedance amplifier: thermal noise from the feedback resistor, dark current from the detector, leakage current in the amplifier and thermal noise in the transistor channels. The Q parameter of the receiver is shown below and is a function of the detector current, I_{on}, and σ_{on}. The variable σ_{on} is the square root of the variance assuming a gaussian distribution of noise around I_{on} as discussed previously in Section 2.1:

$$\text{Q} = \text{I}_{\text{on}} \div 2/(2 * \sigma_{\text{on}}) \tag{3.31}$$

where σ_{on} can be written as follows:

$$\sigma_{\text{on}} = \sqrt{4 * \kappa * \frac{\text{Temp}}{Rf} + 2 * \text{q} * \gamma * \theta_1 + 4 * \kappa * \text{Temp} * \text{ecnf} * (2 * \pi * \text{C}_\text{t})^2 \div \text{g}_\text{m} * \theta_2} \tag{3.32}$$

In this equation, κ is the Boltzmann constant, or $1.38{*}10^{-23}{*}\mathrm{m}^2{*}\mathrm{kg}{*}\mathrm{s}^{-2}{*}\mathrm{K}^{-2}$, *Temp* is the operating temperature of the device in Kelvin, and q is the electron charge, or $1.6{\times}10^{-19}$ C. The parameters γ and *ecnf* are defined as follows, where *Tech* is the CMOS technology node of the receiver in nm [30] [37]:

$$\gamma = \mathrm{I_{dark}} + 2 * \mathrm{I_{Leakage}} \tag{3.33}$$

$$\mathrm{ecnf} = 3 - .002 * (\mathrm{Tech} - 100) \tag{3.34}$$

Lastly, the two θ parameters are defined as:

$$\theta_1 = \frac{1 + \mathrm{g_m} * \mathrm{R_o}}{4 * (\mathrm{R_o} * (\mathrm{C_{inter}} + \mathrm{C_{outer}}) + \mathrm{R_f} * (\mathrm{C_f} + \mathrm{C_{inter}}) + \mathrm{g_m} * \mathrm{R_o} * \mathrm{R_f} * \mathrm{C_f})} \tag{3.35}$$

$$\theta_2 = \frac{(1 + \mathrm{g_m} * \mathrm{R_o})^2}{16 * \pi^2 * (\mathrm{R_o} * (\mathrm{C_{inter}} + \mathrm{C_{outer}}) + \mathrm{R_f} * (\mathrm{C_f} + \mathrm{C_{inter}}) + \mathrm{g_m} * \mathrm{R_o} * \mathrm{R_f} * \mathrm{C_f})} * \frac{1}{(\mathrm{R_o} * \mathrm{R_f} * (\mathrm{C_f} * (\mathrm{C_{inter}} + \mathrm{C_{outer}}) + \mathrm{C_{inter}} * \mathrm{C_{outer}}))} \tag{3.36}$$

The first term under the square root in Equation 3.32 is the thermal current noise in the transimpedance amplifier due to the feedback resistance R_f. The second term is the current noise from dark and leakage sources. The third term is the thermal (Johnson) noise in the transistor channels. Here C_{inter} is the sum of the detector capacitance, C_{det}, and the input gate capacitance to the transimpedance stage. C_{outer} is the sum of the total output diffusion capacitance of the transimpedance stage and gate capacitance of the next amplifying stage. Following the calculation of σ_{on} and the receiver Q, the BER can be calculated using the *complementary error function* (erfc) as follows [37]:

$$\mathrm{BER} = 0.5 * \mathrm{erfc} * (\mathrm{Q} \div \sqrt{2}); \tag{3.37}$$

3.4.5 Power Modeling

Static power dominates the total energy consumption of the receiver circuitry [30]. The amount of static power consumed depends on the size of the transistors in the receiver and the number of inverting amplifiers following the transimpedance stage necessary to obtain a digital level voltage. The input voltage to the receiver formed by the input current I_{on} is:

$$V_{\text{front}} = \frac{I_{\text{on}}}{2 * R_{\text{f}} \div (\text{Transimpedance}_{\text{gain}} + 1)} \tag{3.38}$$

The gain of each inverting stage following the transimpedance amplifier is also a function of the combined transconductance of its two transistors, g_m, and their total output resistance R_o:

$$\text{Amplifiergain} = g_{\text{m}} * R_{\text{o}} \tag{3.39}$$

With the total gain equations of the receiver, the voltage at the input of the receiver and the required digital voltage at the output, the total number of inverting stages, *N*, following the transimpedance amplifier can be calculated using:

$$V_{\text{front}} * \text{Transimpedance}_{\text{gain}} * \text{Amplifiergain}^{\text{N}} = \text{Vdd} \tag{3.40}$$

The total static power consumption of the receiver is simply the saturation currents of the inverters multiplied by the supply voltage:

$$\text{Power}_{\text{static}} = \text{Idsat} * \text{Vdd} * (\text{N} + 1) \tag{3.41}$$

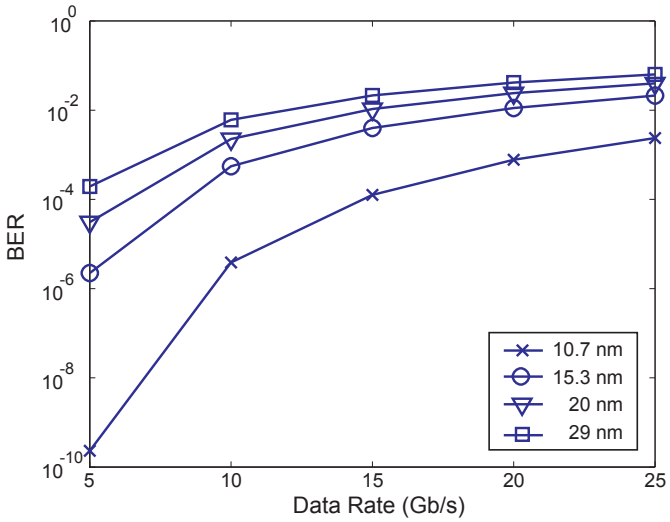


Figure 3.20: Bit-error-rates as a function of CMOS technology node and target receiver data rate with optical input power = 10μW. Smaller transistor technologies achieve a better BER for a fixed data rate due to reductions in thermal channel noise. This is also the case when the data rate within the same technology is reduced through increasing the size of the receiver transistors.

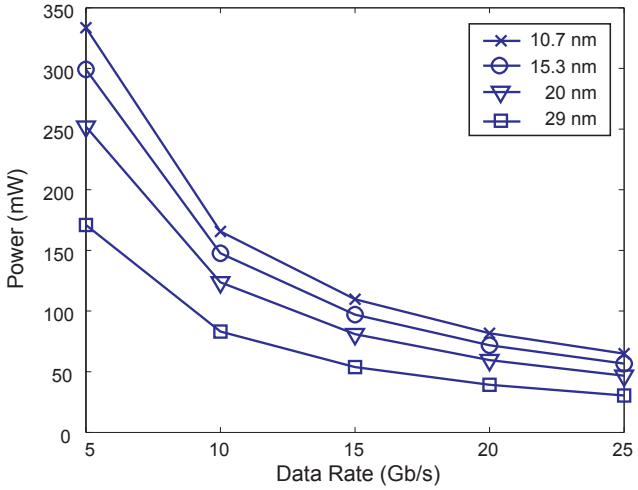


Figure 3.21: Receive static power consumption as a function of CMOS technology node and target receive data rate with optical input power = 10μW. Within a technology node, increasing data rate reduces static power consumption since resulting transistor sizes are made smaller, thus drawing less current. As technology scales, power consumption worsens due to increased relative sizing parameters and drive currents to achieve a fixed data rate.

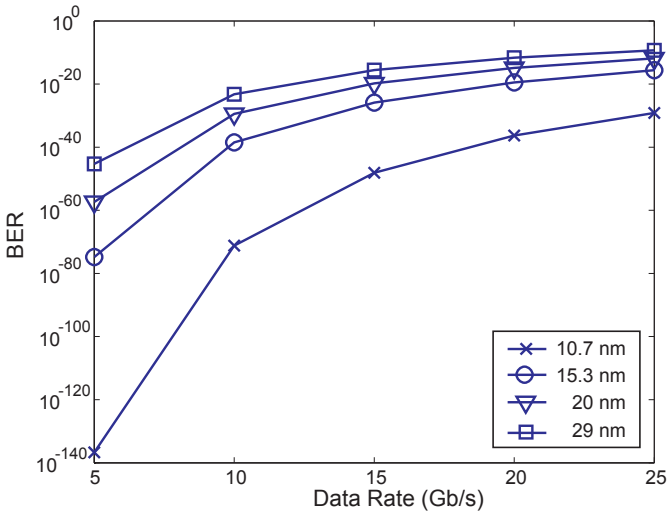


Figure 3.22: Bit-error-rates as a function of CMOS technology node and target receiver data rate with optical input power = 40μW.

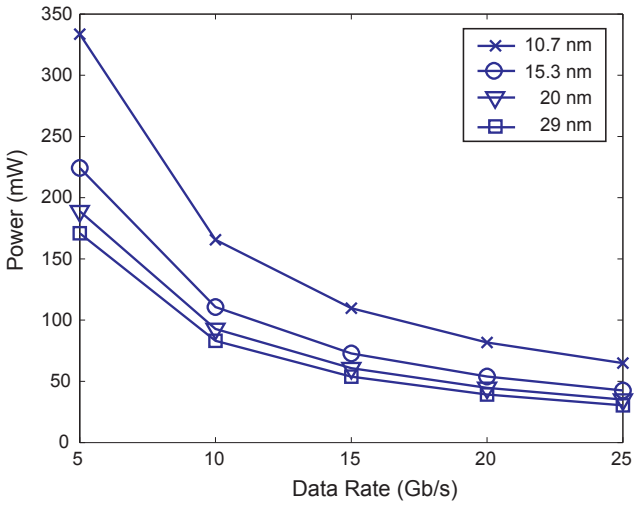


Figure 3.23: Receive static power consumption as a function of CMOS technology node and target receive data rate with optical input power = 40μW.

3.4.6 Power, Performance and BER Results

We provide an optical receiver data rate analysis using two assumed optical input powers at the detector: 10μW and 40μW. We vary the achieved data rate of the receiver (which is dominated by the amplifying stages following the detector based on the analysis in Section 3.4.1) and report the resulting BER and static power consumption using 29, 20, 15.3 and 10.7nm CMOS technologies. These results are shown in Figures 3.20 and 3.21 for 10μW optical input power and Figures 3.22 and 3.23 for 40μW optical input power. We do not report dynamic energy consumption since this has been shown to be dominated by the static power [30].

For a given technology, the BER gets consistently worse as the data rate of the receiver is increased. The reason for this is evident from Equations 3.31 and 3.29 that show the receiver Q parameter and bandwidth of the receiver, respectively. For the results presented, we assume a fixed feedback resistance of $R_f = 1k\Omega$. As the transistor sizes in the receiver are reduced, the spectral bandwidth response, $BW_{receiver}$, increases. However, this comes at a cost, namely the reduction in the receiver's Q parameter due to increased thermal noise from the feedback resistance, increased shot noise from gate and subthreshold leakage currents, and also increased thermal (Johnson) noise in the transistor channels.

As the transistor technology scales, the BER at a given data rate improves. We find that the thermal channel noise component from Equation 3.32 dominates the other two noise sources [37] in calculating the receiver Q. Because this noise source is inversely proportional to the transconductance of the channel, g_m, larger transistor sizes and also scaling improves the Q parameter and thus the BER of the receiver.

At a given data rate, the power consumption of the receiver increases as tech-

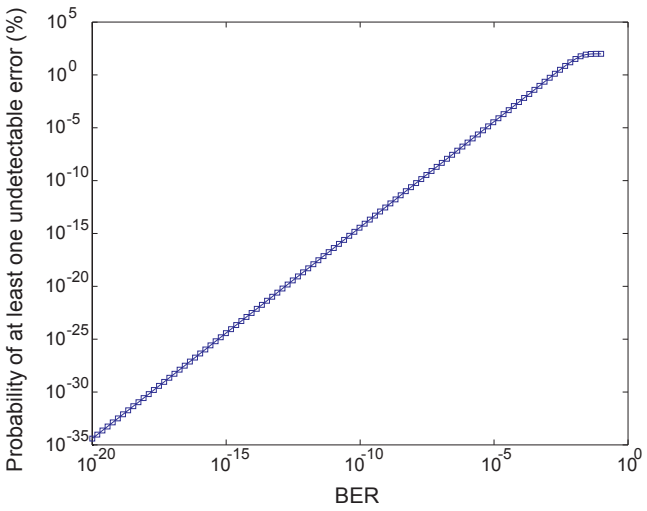


Figure 3.24: A parity bit is used to protect a group of 16 bits in a 64 byte packet. Within the 16 protected bits it's possible to encounter an undetectable error if an even number of bits are erroneously flipped. In this plot we show the probability of at least one undetectable error occurring in the packet as a function of the assumed system BER. As the BER rises, the probability quickly approaches 100% but also falls very rapidly as the BER improves.

nology scales. Increasing drive strength per absolute width as transistors shrink allows them to achieve a larger data rate at the same relative transistor sizing factor as earlier generations. As a result, because we fix the data rates across the technologies, the smaller nodes require larger relative sizing factors. Drive currents are scaling faster than supply voltages, and thus the static power consumption of scaled technologies is larger than previous generations in this analysis. It is possible to reduce this power consumption by decreasing transistor sizes, thus reducing BER at the same time. However, this also increases the receiver data rate, which we want to fix for comparison purposes across the different technologies.

It might seem counterintuitive that as the data rate of the receiver is increased the static power consumption reduces. However, because we fix the feedback resistance, we are effectively trading off power for increased BER at higher receiver data rates. The reason we fix R_f is to minimize the static power consumption at

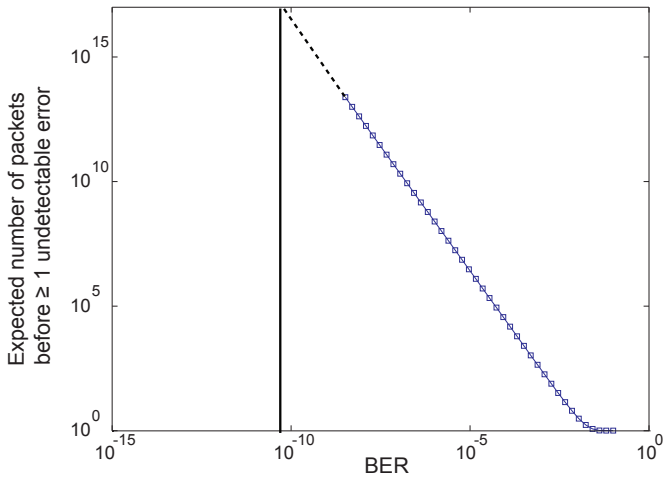


Figure 3.25: To put the data in Figure 3.24 in context, we calculate the expected number of packets that must be received prior to encountering a packet with at least a single undetectable error in one of its parity groups. Here we assume a 64 byte packet with 16 bit groups protected by a single parity bit. With a BER above 10^{-2} every packet that is received will probably have at least a single undetectable error. This number quickly improves beyond 10^{-4}.

a particular data rate. It's possible to save even more power if R_f is increased, but this comes at a cost of greater design complexity to form a high resistance in a scaled CMOS process. We chose a 1kΩ R_f based on the largest obtainable resistance from the data sheets of a scaled IBM process. Fixing BER and obtaining increased data rate at the cost of power consumption might be possible by allowing a decrease in the feedback resistance.

We conclude this section with insight into the meaning of BER and how this number relates to the probability of receiving a data packet with undetectable errors. For these results we assume a packet is a 64 byte cache line being transmitted between processors in a shared memory architecture. Additionally, parity bits are used to protect groups of 16 bits, requiring an additional four bytes to protect the entire packet. This type of protection is effective if the number of bit errors in the data is odd but is ineffective if an even number of flips occur. We calculate the probability of an even number of bit flips occurring within the 16 bit protected

data as:

$$\mathrm{P(undetectable\ error)} = \sum_{\mathrm{X}=1}^{8} \binom{16}{2\mathrm{X}} * \mathrm{BER}^{2\mathrm{X}} * (1 - \mathrm{BER})^{16-2\mathrm{X}} \tag{3.42}$$

This is the probability that an undetectable error will still occur in the presence of parity in a 16 bit block of data within the packet. To calculate the probability of encountering at least one undetectable error in the entire packet, we use the following:

$$\mathrm{P}(\geq 1\ \text{undetectable error in packet}) = \sum_{\mathrm{X}=1}^{\mathrm{PSize}} \binom{\mathrm{PSize}}{\mathrm{X}} * \mathrm{P(undetectable\ error)}^{\mathrm{X}} * (1 - \mathrm{P(undetectable\ error)})^{\mathrm{PSize}-\mathrm{X}} \tag{3.43}$$

Here *PSize* is equal to $64/2 = 32$ sets of protected data in the entire packet. We show the probability of encountering at least one undetectable error in the packet in Figure 3.24 as a function of the system's BER. At high bit error rates, the probability reaches 100% but falls rapidly as the BER shrinks to 10^{-20}. To put these numbers into context, we calculate the expected number of packets that have to be received prior to encountering at least a single undetectable error in Figure 3.25. This expectation is calculated as follows:

$$\mathrm{Expectation} = \sum_{\mathrm{n}=1}^{\infty} \mathrm{n} * \mathrm{P}(\geq 1\ \text{undetectable error in packet}) * \mathrm{P}(errorfree)^{n-1} \tag{3.44}$$

At BER assumptions above 10^{-2}, just about every packet will have an undetectable error but this number quickly falls to hundreds and eventually thousands below a BER of 10^{-4}.

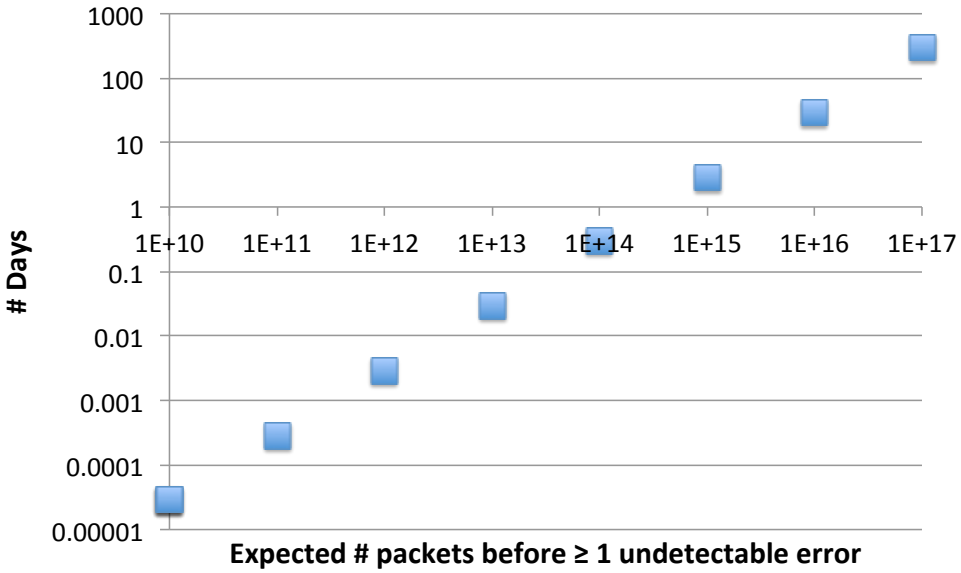


Figure 3.26: Assuming a network node operates at a 4GHz clock rate and receives a packet per cycle, we show the number of days to accumulate different numbers of packets. This data can be correlated with Figure 3.25 to approximate the required BER.

If we assume a processor node operates at a 4GHz clock rate and receives a packet every cycle, the number of days that it takes for an error to occur are shown in Figure 3.26. Approximately 25 years is achieved if the number of packets received prior to an error from Figure 3.25 is 3.2×10^{18}. According to that figure, this corresponds to a BER of 10^{-13}.

One of the key values of using error protection schemes is the ability to operate under reduced BER. In this example, we show parity at a granularity of two bytes. Further improvements in BER will occur if this granularity is reduced to a single byte, but at a cost of requiring more parity bits. As shown previously in this section, the required optical input power at the detector is one knob that can be turned for improving BER without diminishing receiver performance. If the bits are properly protected, a very small input power at the detector is required and could approach approximately 10uW or lower as shown previously in Figure 3.20.

3.5 Optical Insertion Loss

In this section, we examine the insertion loss at the front-end modulator and back-end demultiplexing portions of an optical link. At the modulator side, a wavelength which does not couple into its ring will pass by it to represent a digital one in a non-return-to-zero (NRZ) signaling scheme. However, as it passes by the modulator a portion of the signal still couples into the ring. The two nearest resonators that modulate neighboring wavelengths will also couple some of its power. The total power loss due to this unintended coupling is referred to as insertion loss. In this section, we explore the worst-case behavior for insertion loss at the modulator end and determine how it varies as a function of the system channel spacing and resonance shift amount. Finally, we also show insertion loss at the back-end of the link in the demultiplexing ring resonator array. Here a wavelength's power loss is due to nearest neighbor crosstalk and going through an add/drop filter. Furthermore, since these devices are passive, there is no resonance shift and the insertion loss is only dependent on the system channel spacing.

3.5.1 Ring Resonance Model

To accurately estimate the insertion loss in the modulator and demultiplexer arrays, we present a model based on [59] to mathematically describe the transfer of optical power into a ring resonator. We begin with a resonator coupled to a single waveguide, which is representative of a modulator. Following this, we present a model for a resonator coupled to two waveguides, which describes comb switch and wavelength specific filter functionality at the demultiplexing array. The model parameters and analyzed configurations for both the single waveguide and double waveguide variations are shown in Figure 3.27. Here $k_{1,2}$ and $t_{1,2}$ are the complex coupling coefficients of the system. In this chapter, we assume lossless coupling;

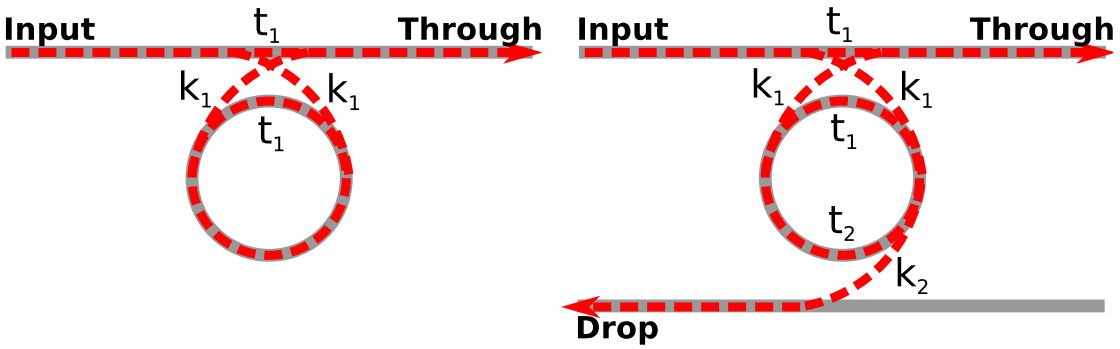


Figure 3.27: Two ring resonator models are shown for describing the behavior of a single ring resonator coupled to one neighboring waveguide and a single ring resonator asymmetrically coupled to two neighboring waveguides. In the former case, light enters the Input port and may be absorbed in the ring or leave out the Through port. In the latter case, light that enters the ring leaves out the Drop port. Variables $t_{1,2}$ and $k_{1,2}$ represent the coupling coefficients of the system and are based on [59].

that is, $\left|k^2\right| + |t^2| = 1$ at each coupling region. The $k_{1,2}$ coefficients present a $\pi/2$ complex phase change to the optical signal, and the $t_{1,2}$ coefficients have no phase shift [57].

Single Resonator Coupled to a Single Waveguide

In this scenario light enters the device through the Input port of the waveguide. Depending on the resonant frequency of the ring, it may traverse past or be diverted into it. In the former case, it will leave out the Through port of the device. The percentage of power that leaves out the Through port is described by the following equation:

$$\frac{t_1 - \alpha * (1 - t_1^2) * e^{j*\theta}}{1 - \alpha * t_1 * e^{j*\theta}} \tag{3.45}$$

where α is the power transfer in the ring after one full round trip ($e^{-PxCircumference/2}$, where P is the propagation loss in the ring per distance), t_1 is the complex coupling coefficient shown in Figure 3.27 and θ is the propagation coefficient in the forward direction multiplied by -1 and the circumference of the ring. The prop-

agation coefficient is dependent on the guiding film, cladding, data wavelength, and thickness and height of the waveguide. Lastly, t_1 is chosen to be equal to α for critical coupling. In critical coupling all of the light is extinguished from the waveguide when its wavelength matches the resonant wavelength of the ring.

Single Resonator Coupled to Two Waveguides

The equations derived for power transfer in the single waveguide case above are changed when another waveguide is added to the system. This is due to the addition of another coupling region to the resonator. In the following analysis we assume that the coupling coefficients between the ring and both waveguides are asymmetric [57]. Because the quality factor of a similarly designed ring with two coupled waveguides will be worse than a ring with only one coupled waveguide, we must be careful to minimize propagation loss in the ring. This is explained by Equation 3.1 and the larger amount of loss in the ring due to the added coupling region. Carefully choosing the coupling coefficients allows a tradeoff between desired insertion loss and ring quality factor.

The following equations can be used to describe the percentage power transfer in the Through and Drop ports of the ring on the right side of Figure 3.27:

$$\mathrm{Power_{Through}} = \frac{\mathrm{t_1} - (1 - \mathrm{t_1^2}) * \alpha * \mathrm{t_2} * \mathrm{e^{j*\theta}}}{(1 - \alpha * \mathrm{t_1} * \mathrm{t_2} * \mathrm{e^{j*\theta}}} \tag{3.46}$$

$$\mathrm{Power_{Drop}} = \frac{-\sqrt{\alpha} * \sqrt{1 - t_1^2} * \sqrt{1 - t_2^2} * e^{j*\theta}}{1 - \alpha * t_1 * t_2 * e^{j*\theta}} \tag{3.47}$$

where the parameter definitions for α, $t_{1,2}$ and θ are the same as in Equation 3.45. For critical coupling $\mathrm{t_1}$ is set equal to α*$\mathrm{t_2}$ [57].

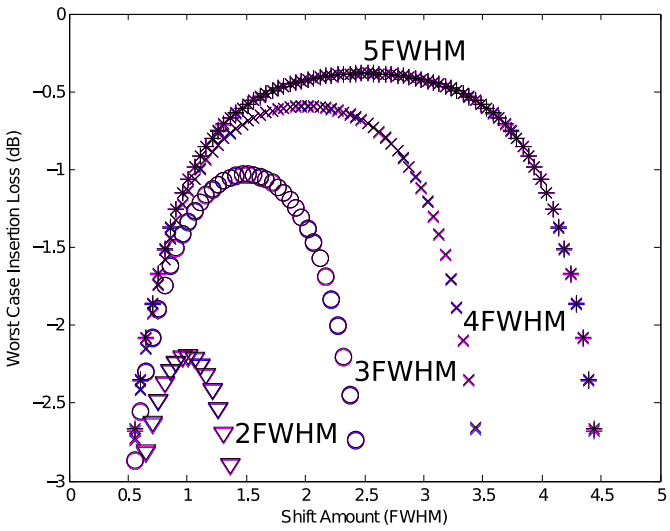


Figure 3.28: Worst case modulator insertion loss is calculated using nearest neighbor crosstalk and self insertion loss. Results are shown for different assumed channel spacings and resonance shift amounts. Depending on the desired level of insertion loss, reasonable laser power requirements are achievable at channel spacings ranging from three to five FWHM. If the peaks are spaced closer, insertion loss becomes excessive. The optimum resonance shift is found to be the channel spacing divided in half.

3.5.2 Power Results

The first set of results that we present are for the insertion loss in the modulator array. We assume a single crystalline silicon ring resonator centered at $\lambda_o = 1550$nm and a waveguide propagation loss of 1dB/cm [21]. We vary the channel spacing and resonance shift amount from one to five FWHM and show the worst-case insertion loss of a wavelength traveling through the modulator array. This worst-case loss occurs when a wavelength passes its shifted modulator (i.e., the wavelength is off resonance), and also by the non-shifted lower wavelength neighbor and shifted upper wavelength neighbor.

The results have an arch pattern due to the worst-case shifting behavior that we model. If the resonance shift is too small, the wavelength will still mostly couple into its parent modulator as it passes by. As the resonance shift grows, it will increasingly couple into the upper wavelength modulator of its neighbor.

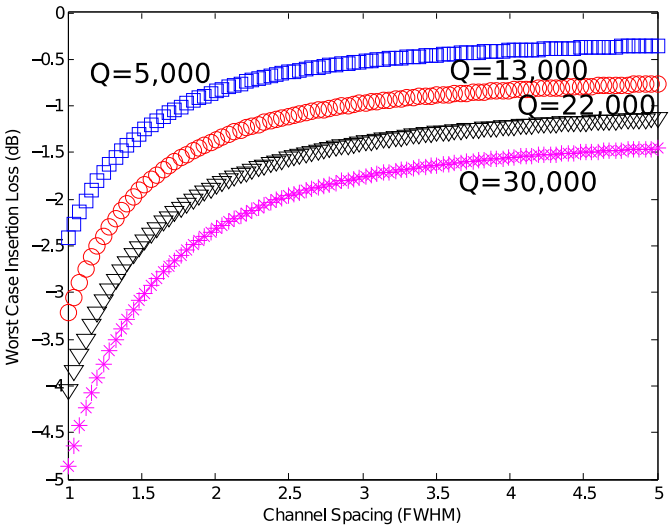


Figure 3.29: Demultiplexer array insertion loss due to nearest neighbor crosstalk and self insertion loss through a ring resonator. We show results for different assumed ring quality factors since an add/drop filter's self insertion loss will change depending upon its FWHM.

Across all the channel spacings, the least amount of insertion loss occurs when the resonance is shifted by the channel spacing divided by two. Results for a single FWHM channel spacing are not shown since the insertion loss is beyond 3dB (50% loss). Based on these results, we believe a channel spacing ranging from three to five FWHM depending on system power requirements will yield reasonable laser requirements. These channel spacings allow insertion loss per wavelength in passing the modulator array to be approximately 1dB (20%) or less. The modulator results are quality factor independent since all units are normalized to the FWHM parameter.

Lastly, we show results for insertion loss at the demultiplexing array at the end of an optical link. Here, nearest neighbor crosstalk from the higher and lower nearest wavelength rings and transmission through the Drop Port of an add/drop ring are responsible for optical power loss. These results are shown in Figure 3.29 for quality factors ranging from 5,000 to 30,000 in four steps. Unlike in the mod-

ulator case, as the quality factor of an add/drop resonator is increased, the loss out the drop port also increases. Thus, for a particular channel spacing, a higher quality factor ring will have greater insertion loss out its Drop Port. Similarly, as the wavelength distance between the two nearest resonance neighbors narrows, the crosstalk loss increases.

3.6 Nonlinear Device Behavior

When light propagates down a waveguide it experiences power attenuation from scattering due to sidewall roughness and linear absorption. The former occurs as a result of fabrication, which may generate roughness along the sides of the waveguide, causing light to scatter and attenuate the signal. Linear absorption uses a signal photon to excite an electron from the valence to the conduction band, and is linearly proportional to the amount of power in the waveguide. These generated free carriers add extra loss to the signal propagation if the total signal power flowing through the waveguide becomes large. Mitigating these loss components is important because they directly influence the amount of power that a laser must supply to the interconnect.

At first thought it might make sense to add more power to a waveguide to combat increasing insertion losses through an optical link. However, as the amount of power in the waveguide is further increased, nonlinearity affects begin to dominate propagation loss and optical device behavior. Two photon absorption (TPA) utilizes two photons which simultaneously strike an electron in the valence band, causing it to rise to the conduction band as shown in Figure 3.30. Absorption of photons does not impact the propagation loss through the waveguide as much as free carrier absorption (FCA). Because at high optical signal powers more free carriers are generated, they begin to absorb light causing potentially large signal

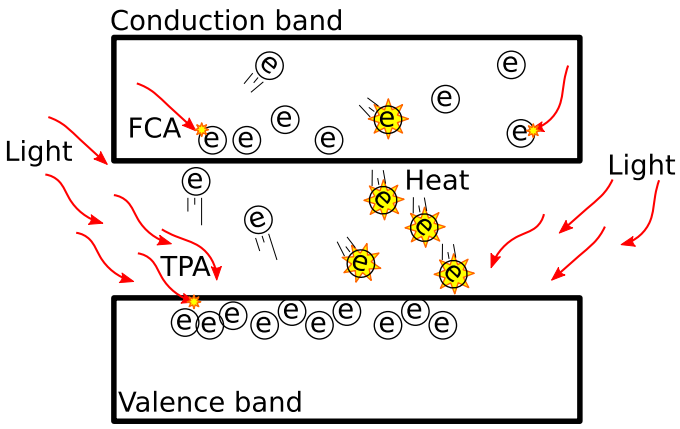


Figure 3.30: As the amount of optical power contained in a waveguide grows, nonlinearities create additional propagation loss and change the designed resonance behavior of system rings. Two photon absorption grows nonlinearly with the intensity of light, and thus becomes the dominant mechanism for generation of free charge carriers at high optical powers. These free charge carriers in the conduction band absorb more light, adding to signal propagation loss. Some of these carriers fall to a lower energy level, releasing a phonon in the process. These phonons cause heat to build up in the device. In the case of a ring resonator, the added free charge carriers cause a blueshift from the designed ring resonator, and the greater temperature causes a dominating red shift. Thus, along with adding propagation loss to a waveguide, nonlinearities cause ring resonators to function improperly.

attenuation. Previous work has shown a one cm long silicon waveguide with 60 wavelengths, each 0.8mW, will generate a total nonlinear loss per wavelength of 0.49dB [57]. This can be improved by reducing the free carrier recombination lifetime below the 500ps used in that work since this causes the generated free carriers to more quickly disappear.

The ring resonance shift caused by nonlinearities degrades the operation of an optical link. The creation of free charge carriers via FCA is the means for modulating an electrical input signal. However, in this context, the unwanted resonance shift causes erroneous behavior. As more charge carriers are created, more phonons (vibrations) are released as a result of the electrons in the conduction band consistently dropping energy states. When an electron falls from a higher energy to a lower energy, it releases a phonon that causes a temperature rise

inside the device. The injection of free carriers into the ring forces the resonance wavelength to blue shift (i.e., move to lower wavelengths of light), whereas the temperature shift will gradually begin to dominate and force the resonance to red shift (i.e., move to higher wavelengths of light). Previous work has shown the total optical power limit in a ring resonator to be approximately 0.8mW [57].

3.7 Putting it All Together

In this section, we build on all of the previous analysis in this chapter to derive estimated optical device parameters for system architects. We first examine the transmitter and receive device components individually and draw high-level conclusions about projected power and performance expectations. Then, we tie all of these conclusions together to form projections for the full optical communication link based on the design in Figure 3.4. This methodology can aid system architects looking to design and simulate realistic nanophotonic interconnects for future chip multiprocessors. One drawback of previous architectural level networks is the lack of consistency and accuracy in assumed optical device parameters. Our goal in this chapter is to form a coherent source of information for architects to use for learning about the relevant parameters of emerging nanophotonics. To date, no work has attempted to create an all encompassing model and accompanying literature describing in mathematical detail the operation principles of optical transmitter and receivers tailored to system architects.

3.7.1 Ring Modulator

In Section 3.3 we showed how implanting oxygen ions into a ring resonator improves the data rate of the device at the expense of increased propagation loss.

Two parameters impact the required voltage across the ring: its quality factor and the desired resonance shift amount. The quality factor is directly related to the propagation loss. The desired resonance shift amount is dictated by the required insertion losses in the system. We presented a model for CMOS inverter drivers across scaled technologies and demonstrated how the supply voltages are unable to provide the required V_{drive} across a ring as the ion implantation and/or resonance shift amount become too large. Figure 3.16 presents the achievable data rates across the different technologies and resonance shift assumptions. Larger transistor nodes achieve higher data rates due to their increased voltage supply, thus providing V_{drive} to a resonator with high ion implantation, and resulting low carrier recombination lifetime. In Section 3.5, we found that the worst case modulator array insertion gives reasonable laser power requirements if the channel spacing of the system is from three to five FWHM. The corresponding optimized resonance shift amounts for these channel spacings are 1.5, 2 and 2.5 FWHM, respectively. Using these projections its possible to observe the achievable modulator data rate and associated power consumption using Figures 3.16 and 3.17.

3.7.2 Optical Receiver

The optical receiver data rate is dependent on the desired BER, where higher data rates are possible but at the cost of more transmission errors. Increasing the amount of optical power at the receiver's detector improves the BER at a set data rate, but results in increased optical power consumption. In Section 3.4 we analyzed two assumed optical input powers of 10μW and 40μW where the former represents the typical assumption in architectural level papers and the latter an upper bound to show how the BER improves. Based on the results in Figure 3.20, 3.21, 3.22 and 3.23 we project that scaled technology nodes will achieve

a data rate of 25Gb/s with an optical input power between 10μW and 40μW to obtain the required BER.

3.7.3 Full Optical Communication Link

We conclude the optical communication link analysis presented in this chapter by giving performance projections for achievable data rates across scaled CMOS technologies. In Figure 3.8 we showed how the required communication data rate fixes the maximum bandwidth of the system ring resonators and the resulting levels of WDM at different channel spacing assumptions. In Figure 3.13 we showed how the data rate of the ring modulator increases as more ion implantation is used across different CMOS technologies. We augment the data in Figure 3.8 with the new data from Figure 3.13 for each technology in Figures 3.31, 3.32, 3.33 and 3.34. The reason for the difference between each technology data and the original data in Figure 3.8 is due to reductions in quality factors because of ion implantation from the maximum value calculated in Section 3.2 (i.e., Max Data Rate = $\mathrm{BW}_{ring}/.75$). Additionally, we use the maximum data rate data from Figure 3.16 and a maximum receiver data rate of 25Gb/s to further narrow the space in Figures 3.31, 3.32, 3.33 and 3.34. For each line (which represents a different channel spacing assumption from three to five FWHM), we draw a star denoting the maximum data rate that could be achieved by that technology as limited by either the receiver or the ring modulator data from 3.16. Although based on the data in Figure 3.13 adding more ions seems to improve performance of the ring modulator (and thus the full optical link), *ultimately the CMOS driver and receiver limit the total achievable data rate* and not the ring resonator.

The WDM level and per wavelength data rate are tuning knobs that can be adjusted to design an optical link for a target aggregate data rate. Two design

points with equal targets are circled in Figure 3.31. The first design point minimizes the per wavelength data rate by reducing the ring resonator ion implantation dosage. The resulting increase in quality factor, combined with a reduction in channel spacing, raises the WDM level to achieve the performance target. The second design point increases the per wavelength data rate and can reach the target at a wider channel spacing because of the slower reduction in WDM level.

The design point with higher data rate provides lower total power consumption in the modulators and receivers. The modulators have a sub linear reduction in power consumption as the per wavelength data rate is decreased based on the analysis in Section 3.3. Static power consumption is dominant in the receivers and reduces as the per wavelength data rate is increased, at a cost of BER. This design point also enables lower external laser requirements because of the increase in channel spacing.

One advantage of the second design point that reduces the per wavelength data rate is the resulting increase in WDM level. In both Phastlane architectures, high-speed packet transmission is possible by eliminating serialization latency. Thus, an entire packet is encoded in the WDM wavelengths, and the critical delay through the network routers is determined only by the time it takes for the head of the signal to reach the end receivers.

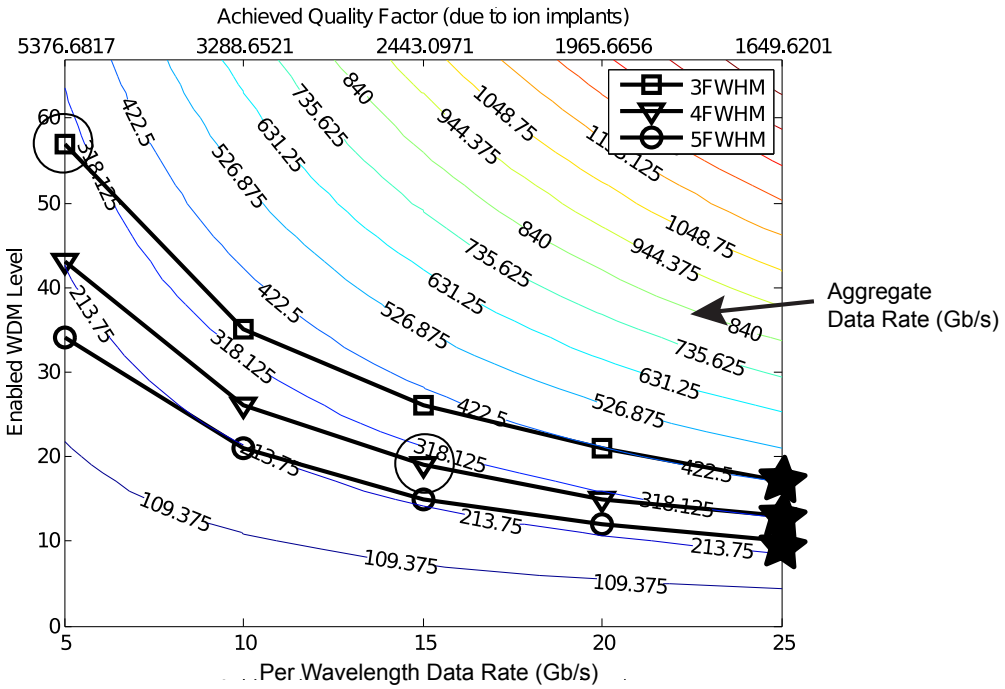


Figure 3.31: Performance results for the maximum data rate and total transmission bandwidth through an optical link at 29nm technology. The lines represent channel spacing assumptions from three to five FWHM and the achievable WDM using ion implantation from Figure 3.13 at 29nm. The dots represent a voltage limited modulator (i.e., the driver circuitry cannot provide enough V_{drive} across the ring to shift resonance) or a receiver limitation (we concluded in Section 3.4 that the maximum data rate is approximately 25Gb/s). Although based on Figure 3.13 adding more ions seems to improve performance of the ring modulator, ultimately the CMOS driver and receiver limit the total achievable data rate. The circles show two design points that tradeoff per wavelength data rate and WDM level to achieve the same aggregative data rate. These tradeoffs are discussed in Section 3.7.3.

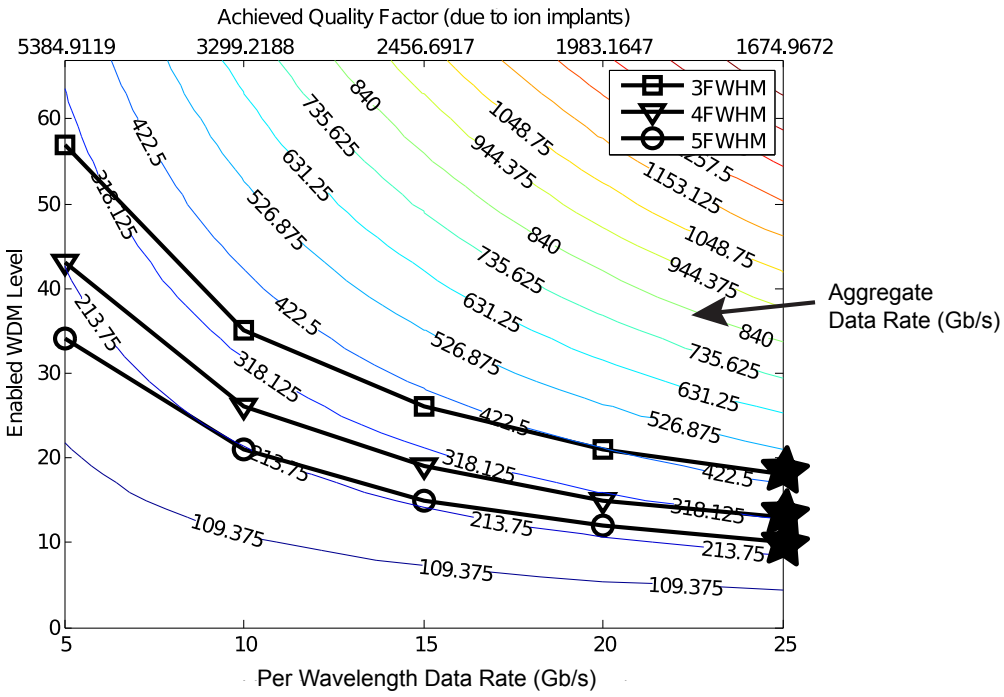


Figure 3.32: Performance results for the maximum data rate and total transmission bandwidth through an optical link at 20nm technology. The lines represent channel spacing assumptions from three to five FWHM and the achievable WDM using ion implantation from Figure 3.13 at 20nm. The dots represent a voltage limited modulator (i.e., the driver circuitry cannot provide enough V_{drive} across the ring to shift resonance) or a receiver limitation (we concluded in Section 3.4 that the maximum data rate is approximately 25Gb/s). Although based on Figure 3.13 adding more ions seems to improve performance of the ring modulator, ultimately the CMOS driver and receiver limit the total achievable data rate.

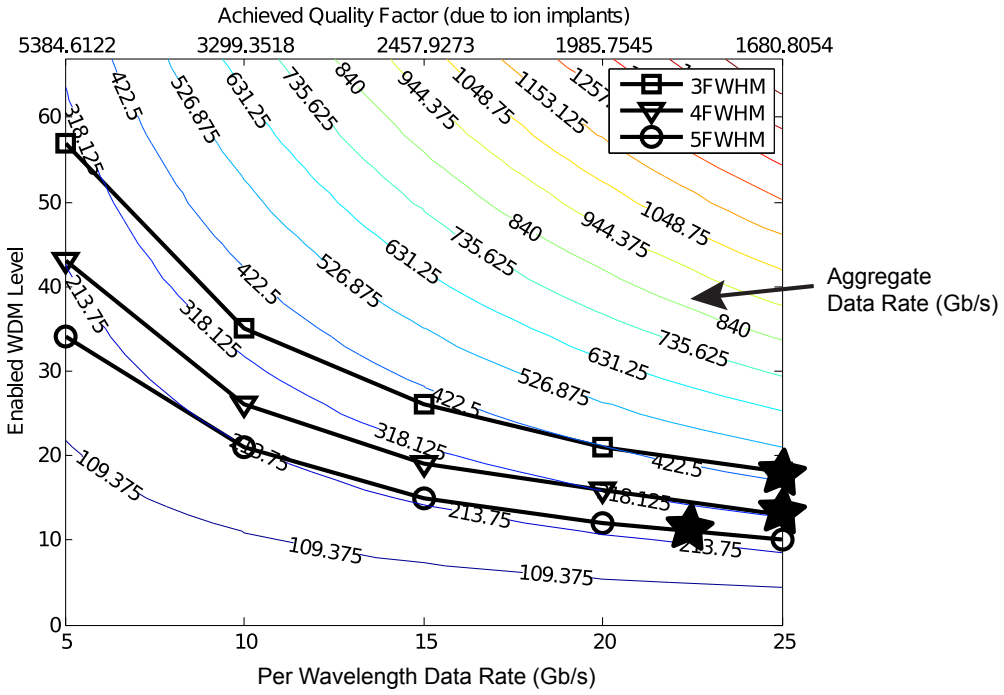


Figure 3.33: Performance results for the maximum data rate and total transmission bandwidth through an optical link at 15.3nm technology. The lines represent channel spacing assumptions from three to five FWHM and the achievable WDM using ion implantation from Figure 3.13 at 15.3nm. The dots represent a voltage limited modulator (i.e., the driver circuitry cannot provide enough V_{drive} across the ring to shift resonance) or a receiver limitation (we concluded in Section 3.4 that the maximum data rate is approximately 25Gb/s). Although based on Figure 3.13 adding more ions seems to improve performance of the ring modulator, ultimately the CMOS driver and receiver limit the total achievable data rate.

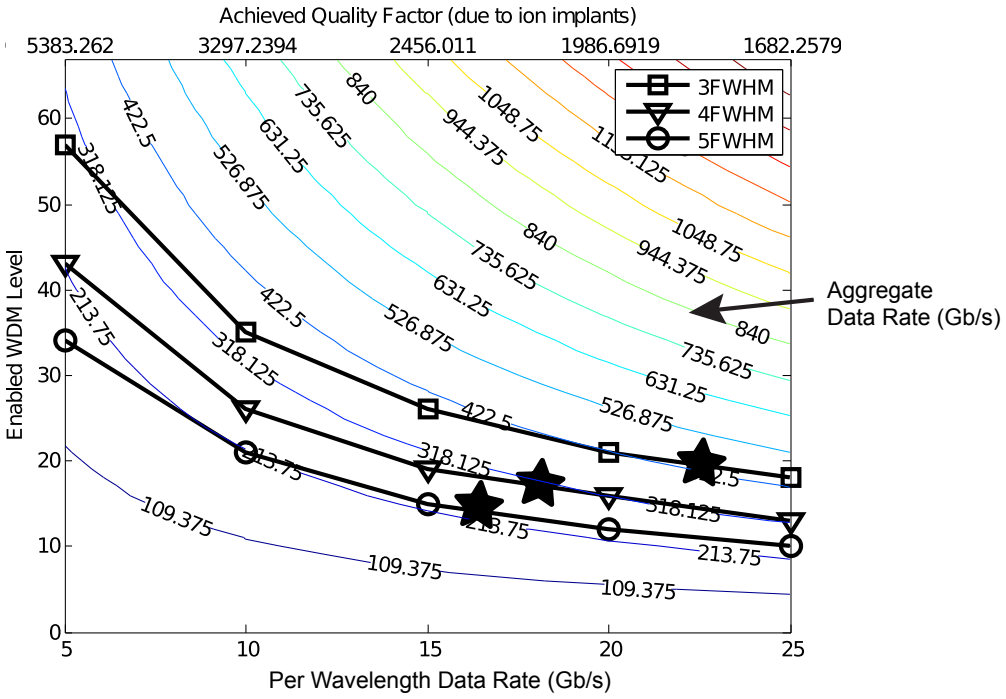


Figure 3.34: Performance results for the maximum data rate and total transmission bandwidth through an optical link at 10.7nm technology. The lines represent channel spacing assumptions from three to five FWHM and the achievable WDM using ion implantation from Figure 3.13 at 10.7nm. The dots represent a voltage limited modulator (i.e., the driver circuitry cannot provide enough V_{drive} across the ring to shift resonance) or a receiver limitation (we concluded in Section 3.4 that the maximum data rate is approximately 25Gb/s). Although based on Figure 3.13 adding more ions seems to improve performance of the ring modulator, ultimately the CMOS driver and receiver limit the total achievable data rate.

CHAPTER 4

PHASTLANE NANOPHOTONIC INTERCONNECT

In this chapter, we present Phastlane, the first optical packet switched network for future chip multiprocessors. We begin with a detailed overview of the proposed network architecture, examining Phastlane's unique switch design, implementation of fixed priority output port allocation, drop signaling flow control, and source based routing to enable high speed packet transmission without sacrificing network bandwidth. We present results that utilize our scaled optical device projections from Chapter 3 to compare Phastlane against an aggressive electrical baseline across both synthetic and Splash workloads. We demonstrate the feasibility of our approach using these projections, but also show that further device innovation is required to make Phastlane's on-chip electrical and laser power consumption competitive with the electrical baseline.

4.1 Network Architecture

One advantage of on-chip silicon photonics is its low latency transmission over distances long enough to amortize the costs of modulation, detection, and conversion. In 16nm technology, the distance beyond which optics achieves lower delay than optimally repeatered wires is expected to be 1-2mm [11], making optical transmission profitable for even single hop network traversals. Our goal, therefore, is to architect an optical switch network that matches the latency and bandwidth of a state-of-the-art electrical network at short distances, that exploits the ability of optics to traverse multiple hops in a single cycle in the case of no contention, and that uses a cache line as the unit of transfer. Meeting these goals requires simplicity in the control path. In particular, we opt for dimension-order routing, fixed-priority arbitration, and simply dropping a packet when buffer space is un-

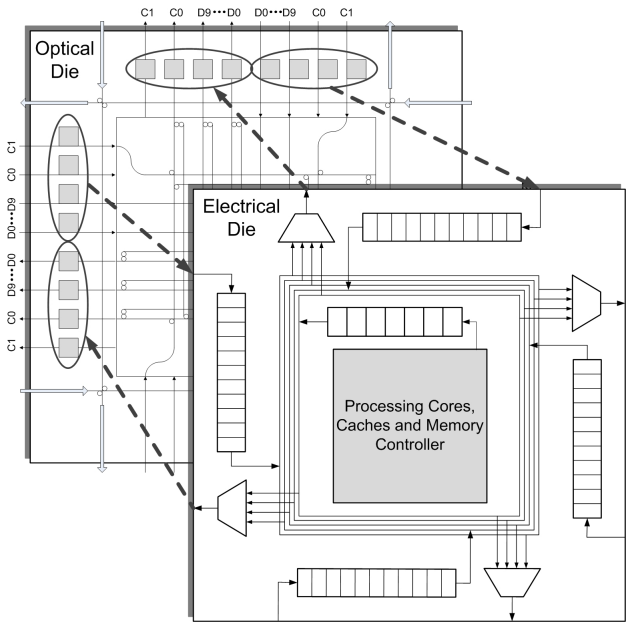


Figure 4.1: Overall diagram of a Phastlane router showing the optical and electrical dies, including optical receiver and driver connections to the electrical input buffers and output multiplexers. The input buffers capture incoming packets only when they are blocked from an optical output port.

available. Although these choices impact network efficiency, they permit optical data transmission over long distances to be minimally impeded by control circuitry.

Our design targets cache coherent multicore processors in the 16nm generation with tens to hundreds of cores and a highly-interleaved, main memory using multiple on-chip memory controllers. High bandwidth density and low latency are simultaneously met using WDM to pack many bits into each waveguide and simple predecoded source routing and fixed priority arbitration.

The optical components of the Phastlane 8x8 mesh network are located on a separate chip integrated into a 3D structure with the processor die. Figure 4.1 shows one of the 64 nodes of the Phastlane network. The node includes one or more processing cores, a two-level cache hierarchy, a memory controller (MC), and the electrical components of the router. The 64 MCs are interleaved on a cache line basis with high bandwidth serial optical links – like those proposed for Corona [65]

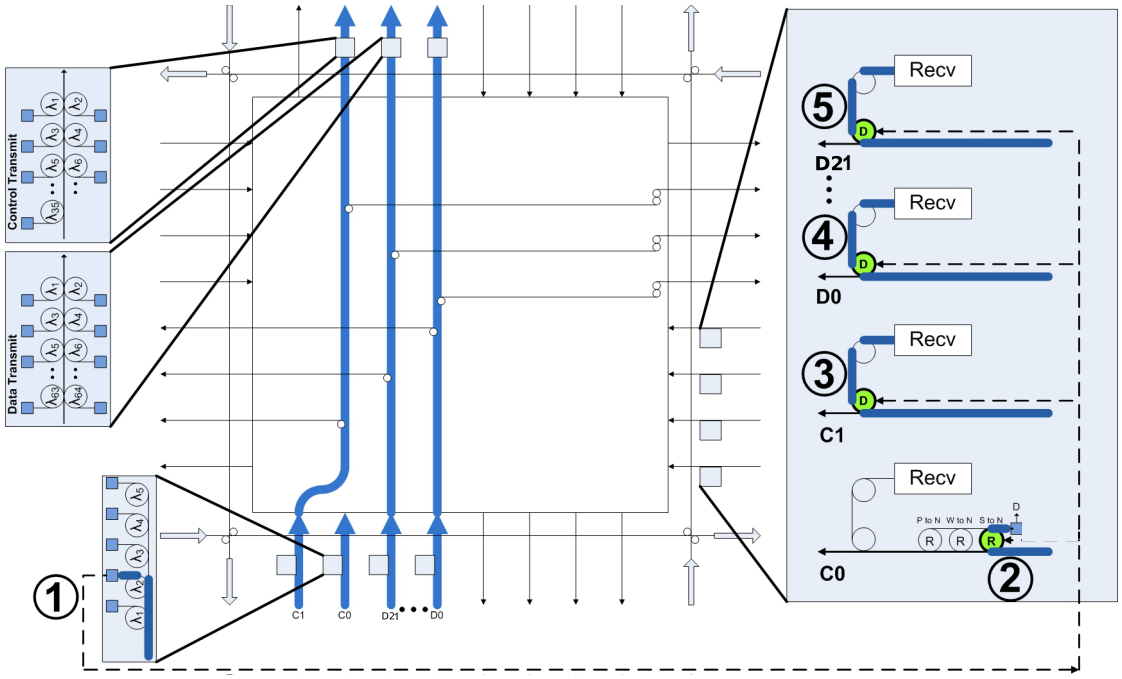


Figure 4.2: Phastlane optical switch, showing a subset of the signal paths for an incoming packet on the S port and the process of receiving an incoming blocked packet on the E input port.

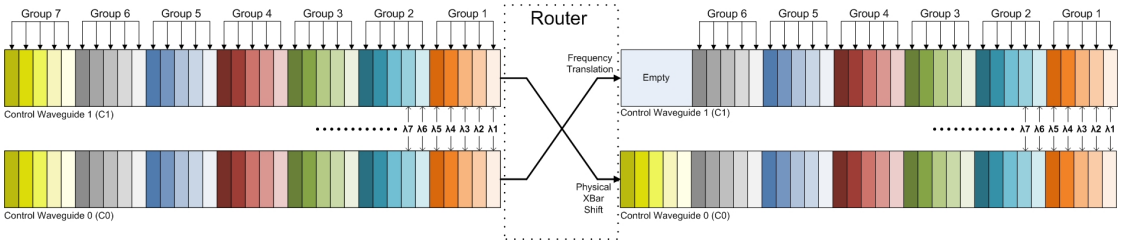


Figure 4.3: C0 and C1 control waveguides. As inputs, they together hold up to 14 groups of five control bits for each router. The Group 1 bits in the C0 waveguide are used to route the packet through the current router. On exiting the router, the Group 2-7 bits are frequency translated to the Group 1-6 positions and output on the C1 waveguide, while the C1 waveguide is physically shifted to the C0 position at the output port.

– connecting each MC to off-chip DRAM.

4.1.1 Router Microarchitecture

Figure 4.2 shows a portion of the optical components of a single Phastlane router. Only a fraction of the input and output waveguides and circuitry are shown for clarity. Resonator/receiver pairs at each of the four (N, S, E, and W) input ports receive packets that are either destined for this node or that are blocked. Transmitter/modulator pairs at each output port drive packets from the local node buffer or from one of the input port buffers. Incoming packets that turn left or right pass through the resonators located inside the router to the coupled perpendicular waveguides.

Unlike the Columbia approach [62], Phastlane has no electrical setup/teardown network. Rather, precomputed control bits for each router are optically transmitted in separate waveguides *in parallel* with the data, and these bits are used to implement simple dimension-order routing and fixed priority arbitration. Each packet consists of a single flit, which contains a full cache line (64 bytes) of Data, the Address, Operation Type and Source ID bits, Error Detection/Correction and miscellaneous bits, and Router Control bits for each of the intermediate routers as well as the destination router. Twenty two waveguides (D0-D21 in Figure 4.2) assuming 35-way WDM transmit the entire packet with the exception of the Router Control, which is evenly divided between two additional waveguides (C0 and C1) as shown in Figure 4.3. The Router Control consists of Straight, Left, Right, Local, and Multicast routing control bits for each of the up to 14 routers that may be traversed in the 8x8 network. The first three bits map to the three possible output ports. The Local bit indicates whether the router should accept the packet for its local node. The Multicast bit indicates a multicast operation as discussed in Section 4.1.1.

Returning to Figure 4.2, consider a packet arriving at the S input port. The

C0 waveguide contains the five control bits for this router on wavelengths $\lambda_1 - \lambda_5$ (Group 1), and up to six other sets of control bits on $\lambda_6 - \lambda_{35}$ (Groups 2-7). All of the C0 bits are received by the resonator/receiver pairs shown on the C0 S input port. The Group 1 control bits are used to route the packet through the switch while the remaining control bits are frequency translated as described below. If the Group 1 Local bit is set, resonator/receiver pairs on D0-D21, C0 and C1 are activated to receive the packet. Otherwise, the packet enters the router and continues on the straightline path (its desired route) towards the N output port. The first set of resonators in the crossbar are activated by the Left bit while the last set are activated by the Right bit. If neither of these are set, the Straight bit is set and the packet exits through the N port. As shown in Figure 4.3, the C1 waveguide is physically shifted to assume the C0 position at the corresponding output port. The remaining $\lambda_6 - \lambda_{35}$ control bits in C0 are frequency translated to $\lambda_1 - \lambda_{30}$ and are transmitted on the C1 waveguide of the selected output port. This physical shift and frequency translation lines up the control fields for subsequent routers.

Since the straightline paths through the router have priority over turns, the C0 Group 1 Straight bit from the S port, when set, blocks incoming packets from the E and W ports from exiting through the N port. For example, if the Right bit for the E input port is set, then this packet must be received or dropped – depending on the available buffer space – to avoid contention with the packet traveling from the S input to the N output. The resonator/receiver pair labelled ① in Figure 4.2 detects this situation causing the packet on the E input to buffer. The Group 1 Straight bit from the S input port (①) activates ② which receives the set Group 1 Right bit off the C0 waveguide on the E input port, forming a drop signal used to buffer the packet. Additionally, resonator/receiver pairs ③-⑤ are activated to

receive the packet on the E input port, preventing it from contending with the packet traveling from S to N. By using predecoded fields to directly control turn resonators and to receive lower priority packets, data transmission through the router crossbar is minimally disrupted by control complexity. This characteristic permits low latency transmission through the switch.

Electrical Buffers and Arbitration

Each router has five sets of buffers in the electrical domain, four corresponding to the N, S, E, and W input ports and one for the local node (Figure 4.1). A newly arriving blocked packet is received, translated, and placed in the corresponding buffer if there is space. Buffered packets have priority for output ports over newly arriving packets. A rotating priority arbiter selects up to four packets from these queues to transmit to the four output ports. Any incoming packets that conflict with a buffered packet for an output port are received and buffered if there is space. When no buffered packet competes for an output port then the aforementioned fixed-priority scheme determines the winner among the newly arriving packets.

Drop Signal Return Path

Phastlane's simplified optical-based control approach leads to dropping packets if an output port is blocked and an input buffer is full. In order to rapidly signal a dropped packet condition, depending on the situation, one of four actions are taken when a packet arrives at an intermediate router:

- The packet is not blocked; in this case, the router registers the received and translated Straight, Left, and Right bits in order to set up a drop signal return path in the next cycle in case the packet is eventually dropped;
- The packet is blocked but the input port buffer is not full; in this case, the

router receives, translates, and buffers the packet and assumes responsibility for its delivery;

- The packet is blocked and the input port buffer is full; in this case, the packet is dropped and the router transmits an asserted Packet Dropped signal and the router's Node ID on the return path output port in the next cycle.
- The Local bit is set, either because the current node is the destination or an interim stopping point, causing the packet to buffer at the end of the clock cycle.

The network includes return paths for signaling the source that its packet was dropped by a particular node[1]. The source may be the original sender of the packet or an intermediate router that buffered the packet (second scenario above). As a packet moves through the network, each router registers the C0 Group 1 Straight, Left, and Right bits. In the next cycle, each router uses these signals to activate the correct return path in case a drop condition needs to be communicated to the source. The router that drops the packet transmits an asserted Packet Dropped signal and its six-bit Node ID on the return path waveguide. These signals propagate through the return path constructed by each router back to the source. The source takes appropriate action (e.g., backoff and resend) upon receiving the Packet Dropped signal. If a source does not receive a Packet Dropped signal in the cycle immediately following transmission, then either the packet arrived at its destination or an interim node has assumed responsibility for its delivery.

The circuitry for constructing this path is straightforward given the predecoded control fields. Referring again to Figure 4.2, the large arrows show the return path input and output ports. Return paths flow in the opposite direction that packets

[1]By definition, each return path is unique and cannot overlap with the return path of any other packet in the same cycle.

travel through the router. For example, a packet that entered the N port and exited the E port would have the return path shown in the upper right corner of the router activated in the following cycle. The latched value of the Group 1 Left input from the N port controls the resonator shown in that corner, which makes a return path connection between the E and N ports. If the packet was dropped at this router, then transmitter/modulator pairs connected to the N return path output transmit the seven-bit optical signal in the following cycle.

Pipelined Transmission in Large Networks

For large networks, such as the 8x8 mesh that we investigate, single cycle corner-to-corner transmission is infeasible at high network clock rates. For these networks, the transmission is completed in multiple cycles, using interim nodes to buffer the packet. In our network at 16nm and using our optical device projections from Chapter 3, three hops can be traversed in one cycle when taking into account the worst-case situation of contention at every router and late arrival of the packet compared to competing packets. For transmissions requiring more than three hops, the source picks the nodes three, six, nine and twelve hops away along dimension order as interim destinations. The Local bits for the interim nodes and the final destination are set. Each interim node detects that their Local bit is set and places the packet in the input buffer if there is room, and otherwise drops the packet. For the former case, upon detecting that another Local bit is set, it assumes responsibility for sending the packet to either the next interim node or the final destination. If the packet is blocked and buffered in an intermediate node before reaching an interim node, the intermediate node may choose to bypass the original interim node and send the packet further (perhaps directly to its destination). It does so by modifying the Local bits of the packet.

Multicast Operations

In a snoopy cache-coherent system, L2 miss requests and coherence messages such as invalidates are broadcast to every node. In Phastlane, a broadcast consists of multiple multicast packets. Multicast packets have a set Multicast bit in the 5-bit router control field. The broadcasting node sends up to 16 multicast messages (eight if it is located on the top or bottom rows of the network).

For a given router, if the Group 1 Multicast bit is set but the Local bit is not, the router receives a portion of the power transmitted on the input lines through separate broadcast resonator/receivers. Since only a portion of the power is extracted, the packet continues through the selected output port to the next router in the absence of contention. If the Group 1 Local bit is also set, the packet is received through the local receive resonator making this router merely an interim node for a multicast packet. In this case, it either drops the packet if it has no buffer space available, or buffers the packet and assumes responsibility for completing the multicast. If neither bit is set, it simply routes the packet without receiving it.

If a multicast packet is dropped, the source examines the Node ID of the dropped packet return path and determines which nodes already received the multicast message. It clears the Multicast bits for these nodes for the resent packet.

4.2 Evaluation Methodology

To evaluate our proposed optical network, we developed a cycle-accurate network packet simulator that models components down to the flit-level. The simulator generates traffic based on a set of input traces that designate per node packet injections. All network components and functionality described in Section 4.1

Flits per Packet	1 (80 Bytes)
Routing Function	Dimension-Order
Number of VCs per Port	10
Number of Entries per VC	1
Wait for Tail Credit	YES
VC_Allocator	ISLIP [45]
SW_Allocator	ISLIP [45]
Total Router Delay	2 or 3 cycles
Input Speedup	4
Output Speedup	1
Buffer Entries in NIC	50

Table 4.1: Baseline electrical router parameters.

Benchmark	Experimental Data Set
Barnes	64 K particles
Cholesky	tk29.O
FFT	4 M particles
LU	2048x2048 matrix
Ocean	2050x2050 grid
Radix	64 M integers
Raytrace	balls4
Water-NSquared	512 molecules
Water-Spatial	512 molecules
FMM	512 K particles

Table 4.2: Splash benchmarks and input data sets.

are fully modeled, including finite buffering in the network-interface controller. In order to do a power comparison with the electrical baseline, we also model dynamic power consumption and static leakage power in a manner similar to [33].

We evaluate the electrical baseline network using a modified version of Booksim [19] augmented with dynamic and static leakage power models. The models use CACTI for buffers, and [3] for all other components. We also integrated finite NIC buffering as well as Virtual Circuit Tree Multicasting [28] to perform packet broadcasts. Finally, we changed Booksim to input the same trace files used for our optical simulator.

Simulated Cache Sizes	32KB L1I, 32KB L1D, 256KB L2
Actual Cache Sizes	64KB L1I, 64KB L1D, 2MB L2
Cache Associativity	4 Way L1, 16 Way L2
Block Size	32B L1, 64B L2
Memory Latency	80 cycles

Table 4.3: Cache and memory controller parameters.

The electrical baseline is an aggressive router optimized for both latency and bandwidth. The router assumes a virtual-channel architecture with the parameters shown in Table 4.1. In order to perform a fair performance comparison with our optical configurations, we assume both low latency and high saturation bandwidth for the electrical network. We reduce serialization latency by using a packet size of one flit, the same as in Phastlane. Doing so also gives no bandwidth density advantage to the optical network since the bisection bandwidth is fixed. We further assume that pipeline speculation and route-lookahead [53] reduce the per hop router latency of the baseline electrical router to 2-3 cycles for every flit. Finally, we assume that the electrical baseline can accept an input flit from each input port every cycle. These flits do not require the cross-bar and instead can be directly accepted by the processor one cycle after the flit enters the router, which is also assumed in our Phastlane architecture.

We evaluate Splash benchmarks and synthetic traffic workloads. By varying the injection rates of the synthetic benchmarks, we obtain saturation bandwidth and average packet latencies. We created Splash traces using the SESC simulator [60]. Each benchmark was run to completion with the input sets shown in Table 4.2. The modeled system consists of 64 cores with private L1 and L2 caches. Each core is 4-way out-of-order and has the cache and memory parameters shown in Table 4.3. As is typical when using Splash for network studies, the cache sizes are reduced to obtain sufficient network traffic.

Total execution times of the Splash benchmarks are found using the average packet latencies from the network trace simulations. These results form a static network latency in SESC on top of which each Splash benchmark is run to completion. Finally, we assume a 16nm technology node operating at a 4GHz processor and network clock with a supply voltage of 1.0V.

4.3 Results

In this section, we present power and performance results for our Phastlane network architecture against an aggressive electrical baseline. We start with performance results for Splash and synthetic benchmarks, and conclude with power results including the external laser component. Across all the performance and power results, we utilize our scaled optical device projections from Chapter 3.

4.3.1 Performance Results

We begin with a synthetic benchmark analysis that compares Phastlane against the baseline electrical network with a three cycle router latency, denoted as *Electrical3*. We show four different variations of Phastlane where *Optical3* represents an achievable packet hop count of three routers per cycle based on the scaled optical device projections from Chapter 3. These scaled parameters are shown in Table 4.4 along with other physical design parameters such as the number of waveguides utilized in the network data path. Three other optical configurations are also shown, *Optical4*, *Optical5* and *Optical8* (representing configurations where a packet can make four, five and eight hops per cycle), to examine whether better optical device performance yields increased network performance. Lastly, we also show *Electrical2*, an aggressive version of the baseline network that has only a two

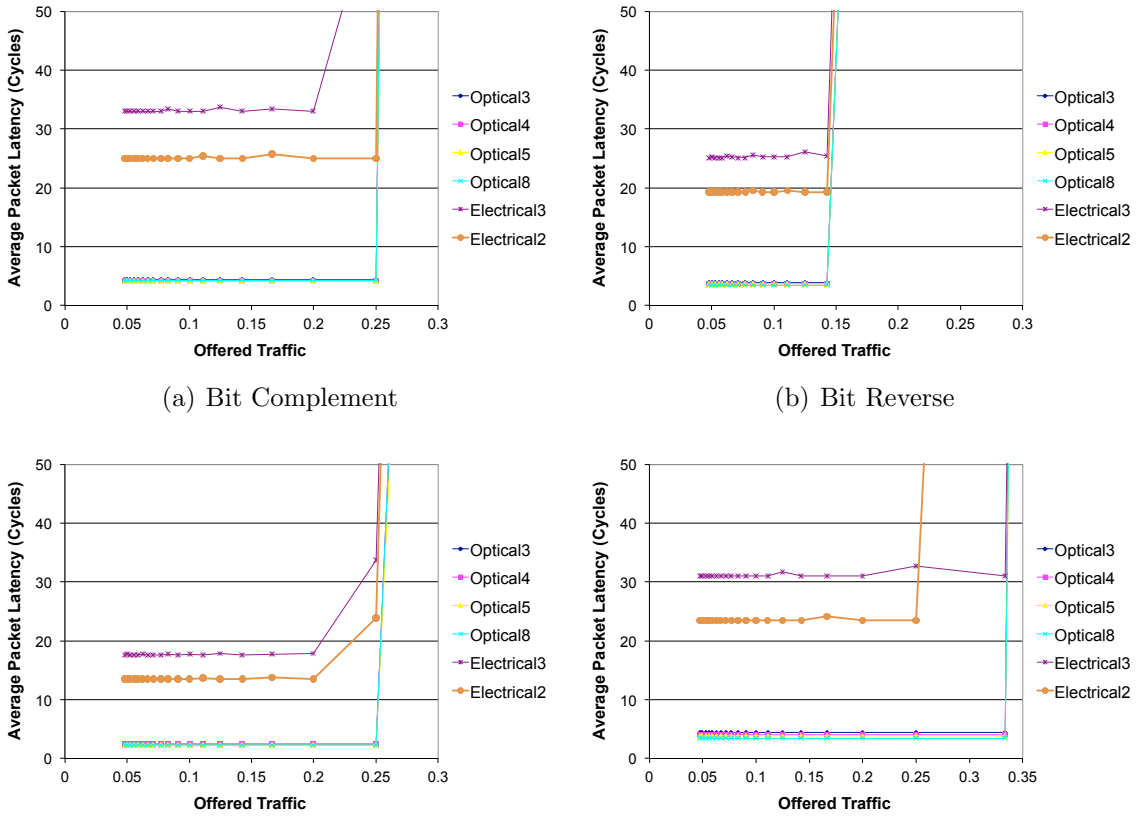
Level of WDM	35
Receiver Latency	7ps
Optical Transmitter Latency	33ps
Comb Filter Latency	29ps
Optical Signal Propagation	6ps/mm
Data Path Width (WGs)	24
# flits per packet	1
# Hops	3
Total Node Area	$2mm^2$
Channel Spacing (units of FWHM)	3
Number of Optical Layers	2

Table 4.4: Phastlane device parameters.

cycle latency per router.

We first evaluate average packet latency and saturation bandwidth using the synthetic workloads shown in Figure 4.4 for Bit Complement, Bit Reverse, Shuffle and Tornado. Across the four benchmarks, the different optical configurations see a small improvement as more hops can be traversed per cycle, achieving approximately 5-10X lower latency than the electrical networks. This is due to the behavior of the traffic patterns, which have many source, destination pairs that are close enough to not need the more aggressive configurations. Also, due to congestion in the network, most packets are blocked in switch arbitration before they reach the maximum hop count.

Figure 4.5 shows network speedup for the Splash benchmarks. For eight of the benchmarks, the optical three-hop network achieves a network speedup of over 1.9X (and over 2.5X for three benchmarks) compared to the electrical network. For most of the benchmarks the four, five and eight hop networks perform marginally better than the three-hop network; this result indicates that our projected scaling of the optical components will not dramatically impact performance. While overall, the optical configurations far outperform the baseline electrical network, the performance of Barnes, Cholesky, Ocean and FMM are highly sensitive to the



(a) Bit Complement

(b) Bit Reverse

(c) Shuffle

(d) Tornado

Figure 4.4: Average packet latency as a function of injection rate for four synthetic traffic patterns. We show results for two electrical packet switched networks, *Electrical3* and *Electrical2*, representing three and two pipeline stages per router, respectively. Four optical configurations are shown, *Optical3*, *Optical4*, *Optical5* and *Optical8*, where the number of router hops a packet can traverse per cycle is denoted by the trailing number.

amount of buffering at every router input port, and thus the number of dropped packets. These dropped packets steal resources from other packets, and also must be retransmitted, which impacts network performance. This result highlights a weakness of our simplified network control: with insufficient buffering, some traffic patterns may lead to many dropped packets that saturate the network. We address this issue in Chapter 5.

Lastly, we show system performance (i.e., improvement in execution time) across the Splash benchmarks for the *Optical3* configuration relative to the elec-

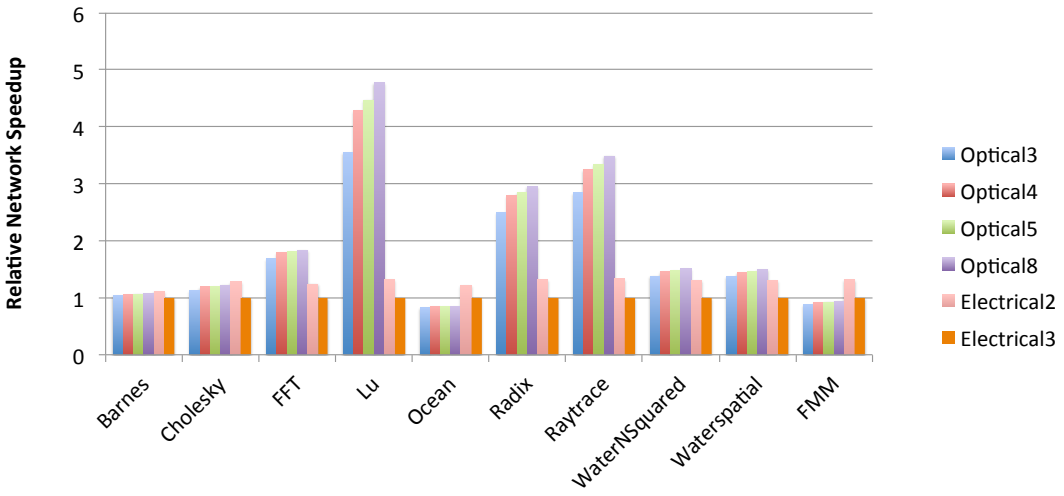


Figure 4.5: Network performance results for Splash benchmarks. We show results for two electrical packet switched networks, *Electrical3* and *Electrical2*, representing three and two pipeline stages per router, respectively. Four optical configurations are shown, *Optical3*, *Optical4*, *Optical5* and *Optical8*, where the number of router hops a packet can traverse per cycle is denoted by the trailing number.

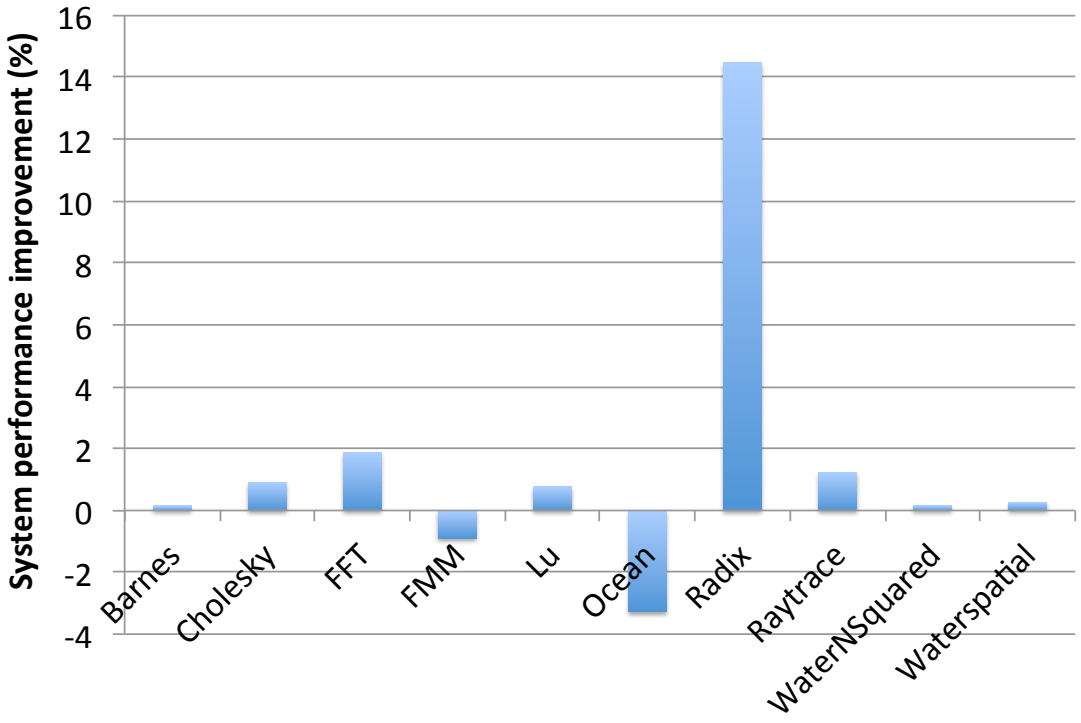


Figure 4.6: Relative system performance for the Splash benchmarks using the *Optical3* configuration and the *Electrical3* electrical baseline network.

Component	Power	Energy/bit
Receiver	42.5mW	5pJ/bit
Optical Transmitter	85mW	10pJ/bit
Optical Comb Filter	400mW	50pJ/bit

Table 4.5: Phastlane optical device energy consumption.

trical baseline *Electrical3* in Figure 4.6. Across all benchmarks, Phastlane has a 1.6% speedup.

4.3.2 Power Results

We use our device model to project the power consumption requirements of the optical building blocks in Phastlane. These parameters are shown in Table 4.5 for the optical receiver, transmitters and also comb filters in the optical crossbar. The resulting network power consumption using these projections is shown in Figure 4.7 for the Splash benchmarks. Phastlane's power consumption is well above the electrical networks due to our energy projections, which must be lowered through further innovation at the device level from pJ's/bit to hundreds of fJ's/bit in order to show improvement.

In Figure 4.8 we show Splash power results assuming aggressive optical device scaling [7]. Here, the optical modulator consumes 120fJ/bit and the receiver 80fJ/bit. Across all of the benchmarks, the average improvement in power consumption for *Optical3* is 31.8%. These results demonstrate the importance of continued device innovation and the resulting improvements in power consumption that could follow.

The second component of power consumption in an optical network is from the external laser source, which supplies the light to the chip for forming the optical packets. To calculate the required laser power we use the optical loss components shown in Table 4.6. These values represent the power loss associated with being

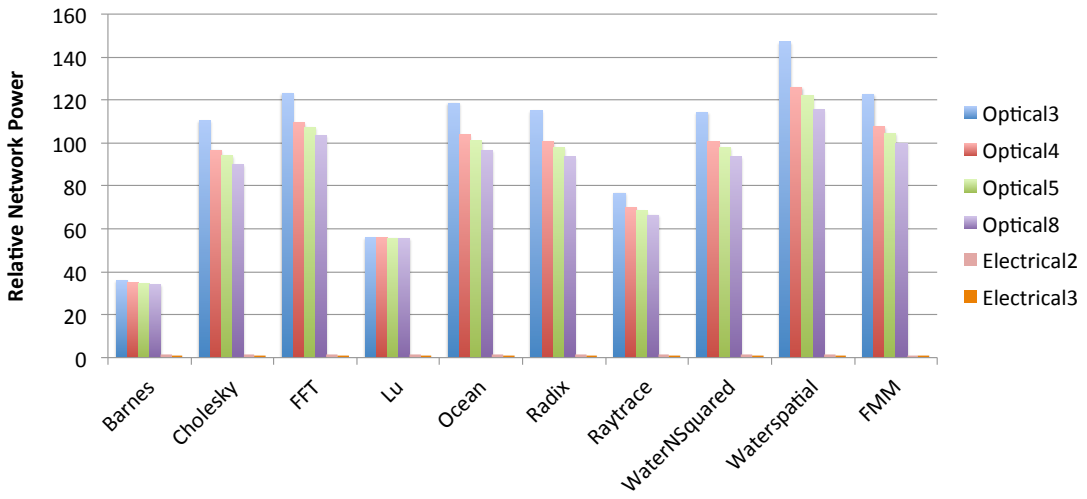


Figure 4.7: Network power consumption results for Splash benchmarks. We show results for two electrical packet switched networks, *Electrical3* and *Electrical2*, representing three and two pipeline stages per router, respectively. Four optical configurations are shown, *Optical3*, *Optical4*, *Optical5* and *Optical8*, where the number of router hops a packet can traverse per cycle is denoted by the trailing number.

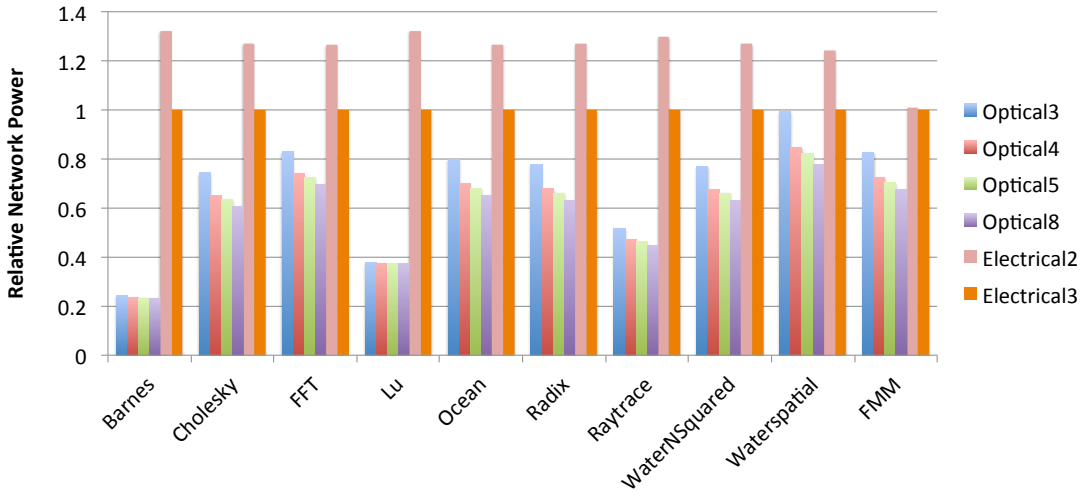


Figure 4.8: Network power consumption results for Splash benchmarks. Optical receiver and transmitter energy consumption is optimistically scaled to 80fJ/bit and 120fJ/bit, respectively [7].

Transmitter Through Loss	1.1dB
Demux Insertion Loss	0.6dB
Comb Filter Insertion Loss	0.1dB
Comb Filter Through Loss	0.1dB
Waveguide Propagation Loss	0.1dB/cm
Laser Chip Coupling Loss	3dB
Required Laser Power	109W

Table 4.6: Phastlane optical loss projections.

transmitted into the network and received, passing by or through a comb filter in the optical crossbar, propagation inside of a waveguide, and coupling from the laser into the chip, respectively. Adding all of these components up, and assuming that 40μW watts of optical laser power is necessary at the detector (based on our BER projections from Chapter 3), we calculate the laser power requirements to be 109W. We discuss the Phastlane power problem more in Chapter 7 where we propose potential solutions.

CHAPTER 5

PHASTLANE 2.0 NANOPHOTONIC INTERCONNECT

In this chapter, we present Phastlane 2.0, a hybrid electrical/optical router design that builds on the Phastlane architecture described in Chapter 4 through the complete redesign of the crossbar, flow control scheme, output port arbitration and source routed control encoding. We begin with a detailed description of the optical architecture, describing the circular waveguide switch design for localizing all routing logic to to an input port, which removes delays associated with control signaling. Phastlane 2.0 uses a novel optical arbitration scheme that implements rotating priority and lends itself to the use of on/off flow control to avoid dropping packets. Next, switch pre-configuration is described for statically setting the switch to join straight path ports prior to packet traversal at the beginning of every clock cycle. Through pre-configuration packets can traverse up to four router hops in a network clock cycle in the absence of contention. Lastly, we present results for power and performance relative to an aggressive electrical baseline using scaled optical device projections from Chapter 3.

5.1 Network Architecture

5.1.1 Router Microarchitecture

One advantage of on-chip silicon photonics is its low latency transmission over distances long enough to amortize the costs of modulation, detection, and conversion. In 16nm technology, the distance beyond which optics achieves lower delay than optimally repeatered wires is expected to be 1-2mm [11], making optical transmission profitable for even single hop network traversals. Similar to the original Phastlane design, our goal is to architect an optical router that matches the performance

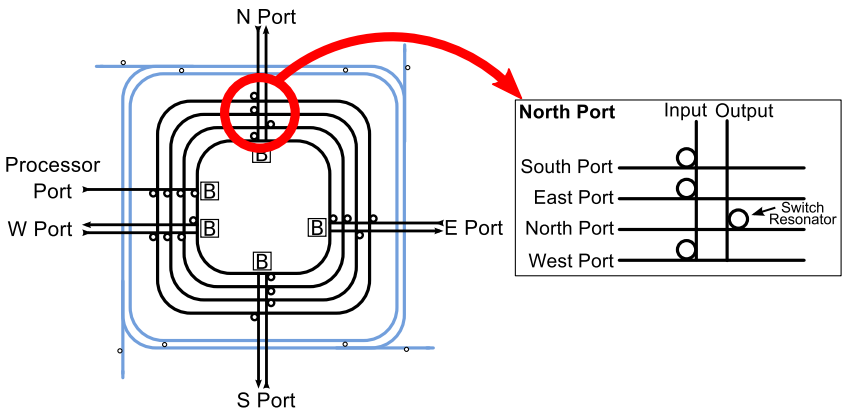


Figure 5.1: Proposed optical switch architecture. The four innermost circular waveguides correspond to each of the output ports of the switch. Switch Resonators allow a packet on an input port to be routed to any of the other output ports.

of a state-of-the-art electrical switch under high load, but enables multiple hops to be traversed in a network cycle under reduced load. This is possible through simplicity in the router control path and switch pre-configuration, which allows an incoming packet to travel through a switch with minimal delay.

Our design targets cache coherent multicore processors in the 16nm generation with tens to hundreds of cores and a highly-interleaved, main memory using multiple on-chip memory controllers. Each node includes one or more processing cores, a two-level cache hierarchy, a memory controller (MC), and a network switch. The MC's are interleaved on a cache line basis with high bandwidth serial optical links like those proposed for Corona [65] connecting each MC to off-chip DRAM.

5.1.2 Switch Design

Figure 5.1 shows a portion of the optical components in our proposed radix five optical switch. Two of the five ports, one being the port to the local processor, are located on the west side of each router. Only two waveguides per port are shown for simplicity, a single input and a single output waveguide. A data path width

of twenty four waveguides is actually implemented to achieve high bandwidth, low latency network communication. The local processor only has an input port because it receives packets via the buffers located at the other input ports. Each of the four circular waveguides in the centermost portion of the switch correspond to one of the four output ports. The North, South, East and West input port waveguides connect to three of these output port circular waveguides through coupling resonators, and the Processor input port connects to all four.

The blow-up shows the Switch Resonators in the North Port and illustrates these connections where resonators enable the input waveguide to couple to the South, East and West ports. Similarly, its output waveguide couples to the portion of the switch corresponding to the North Port. Port buffers are located at the center of the router at each input port. A packet is buffered when it reaches its destination (final or interim; see Section 5.1.6) or if it is unable to win arbitration for its desired switch output port, causing it to block. In the latter case, no Switch Resonators will be set and the packet will be forced to enter the buffer.

The switch design eliminates the optical power loss associated with waveguide crossings through the use of waveguide layers [20]. Waveguide links connecting one router to another are implemented on a different layer than the circular waveguides in the switch. Light couples between the layers through the ring resonators in the switch and router input ports.

Unlike the Columbia approach [62], our proposed optical switch has no electrical setup/teardown network. Rather, precomputed control bits for each router are optically transmitted in separate waveguides *in parallel* with the data, and these bits are used to implement simple dimension-order routing. Each packet consists of a single flit, which contains a full cache line (64 bytes) of Data, the Address, Operation Type, Error Detection/Correction and miscellaneous bits. Router Control

bits are also contained in the packet which are used at the source, intermediate and destination routers.

Theoretically, the Router Control could consist of 64 distinct routing groups, each of which corresponds to an individual node in the network. Prior to entering the network, a packet sets its Router Control by configuring only the routing groups corresponding to the switches it will traverse. All 64 possible routing groups each have six different wavelengths corresponding to the four possible outputs plus Valid and Multicast bits. In the simplest case, all routing groups will be placed on a single waveguide such that every bit is implemented with a different wavelength. However, it is also possible to spread the groups across different waveguides to decrease wavelength usage. This is feasible because each router is statically configured to read its own routing group (i.e., proper wavelengths and waveguide) when a packet enters one of its input ports. The first four Router Control bits in a routing group map to the four possible outputs a packet can leave through in a router. If a packet enters the North, East, South or West ports, one of these bits represents the Local bit (also called an Interim bit when the current router is not the packet's final destination), which dictates whether the router should accept the packet for its local node. The Multicast and Interim bits are discussed in Sections 5.1.5 and 5.1.6. The Valid bit is utilized in switch pre-configuration, which is introduced in Section 5.1.7

In this work, we implement an improvement over the use of routing groups corresponding to every router in the network in a packet's Router Control. We utilize routing groups that correspond to every input port in the network. Furthermore, if packets are routed deterministically, certain sets of input ports can share the same routing group since it is never the case that more than one of them can be used by a packet traveling to its destination. Compared to using per-hop routing

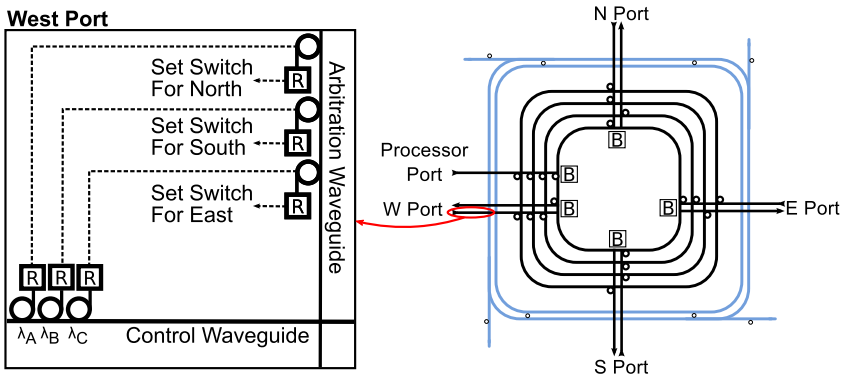


Figure 5.2: Switch input ports receive control bits to set up the switch for proper routing. Three of the six control bits are used for routing the packet to the proper output port. These control bits are received and used in switch arbitration.

groups, this permits reducing the number of required routing groups from 64 to 15, allowing us to drastically reduce the required number of wavelengths to implement the control.

Consider a packet arriving at the West input port as shown in Figure 5.2. The Control waveguide contains the six control bits for this router in its associated control group, and up to 29 more control bits depending on the distance between the packet's source and destination (we show in Section 5.4 that a waveguide can have a WDM level of 35). More control bits can be added to support larger hop counters by adding more waveguides. Three of the control bits–East, North and South–represent the desired route of the packet through the router. These bits are received and used to drive resonators connected to the Arbitration Waveguide where each of its resonators represent a different output port request. When a resonator is turned on, it generates a request for that output port. When arbitration has finished, the results are used to set the appropriate resonators in the switch.

It is important to note that a packet's payload data arrives *in parallel* with its control bits. Within each router, the control signals arbitrate for and set resonators to route the payload data through the switch. While this occurs, the payload data

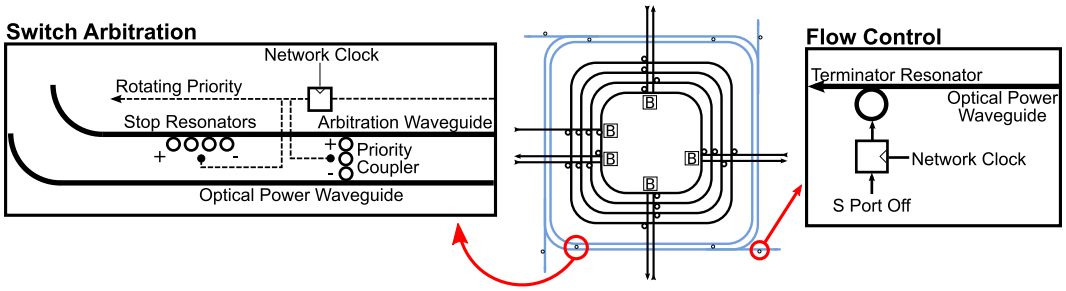


Figure 5.3: Switch arbitration is achieved using the two outermost circular waveguides in the optical router. An external laser source couples tokens into the Optical Power Waveguide at the four corners of the switch. Depending upon which priority coupler is activated, these tokens will couple into the Arbitration Waveguide at different points for use in switch arbitration. Stop Resonators absorb the arbitration wavelengths that haven't been sinked by an input port. The Rotating Priority signal is passed in a rotating fashion to turn on a different Priority Coupler each cycle. Optical flow control utilizes the Optical Power Waveguide. If any of the token off signals are activated, Terminator Resonators prevent these tokens from being available for switch arbitration.

travels to the optical receiver just prior to the electrical buffering. If output port arbitration is won, the crossbar resonators are properly set and the payload data is routed around the circular waveguide and out the corresponding output port. This occurs through multiple switches within a given clock cycle. Thus, the control signals are on the critical path timing-wise. If the packet doesn't win output port arbitration, none of the crossbar resonators are turned on and the packet is latched into the electrical buffer.

All of a packet's routing, switch arbitration and switch setup operations are performed locally at each input port, which eliminates potentially high latency electrical operations associated with lengthy control signaling paths.

5.1.3 Switch Arbitration

Switch arbitration is enabled by the two outermost circular waveguides in the switch shown in Figure 5.3. Ring resonators (Priority Couplers) join the two waveguides at particular points in the loop, shown in the left blowup image. At each of

the four corners of the switch light is coupled into the Optical Power Waveguide, which consists of four wavelengths (referred to as *tokens*), each corresponding to an output port in the switch. Every cycle only one Priority Coupler is activated by the Rotating Priority signal, allowing the light from the Optical Power Waveguide to couple into the inner Arbitration Waveguide. The switch arbitration priority changes every cycle as the Rotating Priority signal moves around the ring. The Stop Resonators prevent light from circulating around the Arbitration Waveguide more than once.

After a packet's control bits are translated to the electrical domain at a router's input port, they are used in switch arbitration. If a packet requests a particular output port, it will attempt to sink the wavelength associated with that output port's token. Light propagates in the counter clockwise direction in the Arbitration Waveguide, and input ports closest to the activated Priority Coupler in this direction have higher priority than others that are further away. When an input port arbitrates for an output port, it will sink the corresponding token by turning on the appropriate ring resonator along the Arbitration Waveguide such that any lower priority input ports no longer see that token wavelength. A packet on an input port may only exit an output port through the switch if it has its token. For example consider the input port highlighted in Figure 5.2. If a packet enters this port, the control wavelengths used for routing are received and used to drive an appropriate ring resonator on the Arbitration Waveguide. If the packet desires to be routed out the East Port, the third resonator from the top will be turned on. If the token for the East Port is available on the Arbitration Waveguide, it will be sinked off such that any lower priority input ports can no longer see it. Then it will be used to locally set the input port's Switch Resonators (see Figure 5.1) so that the packet can be properly routed to the East Port.

5.1.4 Electrical Buffering and Flow Control

Packets that do not couple into the switch waveguides continue to the buffers at the center of the switch. Here they are electrically received and latched into the input port's queue. In this study, we implement on/off router flow control because it requires very little additional hardware complexity over what we have already discussed. The router flow control utilizes the switch Arbitration Waveguide through the Terminator Resonators, one of which is shown in the right blowup in Figure 5.3, which are located where light couples into the Optical Power Waveguide used for output port arbitration. Each Terminator Resonator corresponds to the wavelength of one of the output port tokens on the Arbitration Waveguide. If there are no free downstream buffer entries through an output port, the input ports should be prevented from sourcing a token for that output. A X Port Off signal, where X is North, South, East, or West, achieves this purpose. If a X Port Off signal is set, there will be no token corresponding to that output port available on the Arbitration Waveguide, forcing an incoming packet requiring that output to be buffered. Assuming that a downstream router can send an On or Off signal to an adjacent upstream router electrically in a single cycle, three buffers per input port are required to cover the round trip delay enabling full throughput. While a packet requires only a fraction of a cycle to travel across a network hop, a new packet will not utilize that same channel until the following network clock cycle.

At the beginning of each cycle, every input port buffer counts its number of free entries. If this number is one, an Off signal is sent upstream. On the following cycle, this signal latches into a register as shown in Figure 5.3 and turns off the appropriate token by turning on the corresponding Terminator Resonator. Similarly, when the number of free entries is two, no signal is sent and the Terminator Resonator is turned off the following cycle, allowing the token to flow.

Because of the delayed flow control signaling, it is possible for an input port to be transmitting a previously buffered packet and receiving a new packet in the same cycle. In order to avoid collisions between the two packets, the latter is bypassed to the center switch buffers via the Bypass Path shown in Figure 5.4. The Block Resonators, denoted in the diagram by resonators with a *B*, prevent a failed packet transmission at the local input port from interfering with an incoming packet on the Bypass Path. When a packet is transmitted from an input port's queue, the Ongoing Transmission signal is activated, turning on the Block Resonators and Bypass Path. The Transmit Resonators are utilized to insert the packet transmission into the router just prior to the control logic used for routing, switch setup and arbitration. If the packet does not win its desired switch output port, it should not be re-buffered in the input port queue since it already exists at the head. Transmission failure is detected by noting the existence of a packet's Valid bit coupling through the Block Resonators. When this occurs, the input port knows to retransmit the packet at the head of the queue. Similarly, if the Valid bit is clear, it knows to pop off the packet at the head of its input port queue.

5.1.5 Multicast Operations

In a cache-coherent system, particular requests may be broadcast to every node. In the optical mesh topology that we evaluate in this study, a broadcast consists of multiple multicast packets. Multicast packets have the Multicast control bit set in the six bit router control group. For a 64 node system, the broadcasting node sends up to 16 multicast messages (eight if it is located on the top or bottom rows of the network).

For a given router, if the Multicast control bit is set, the Multicast Resonators are turned on as shown in Figure 5.4. The router then receives a por-

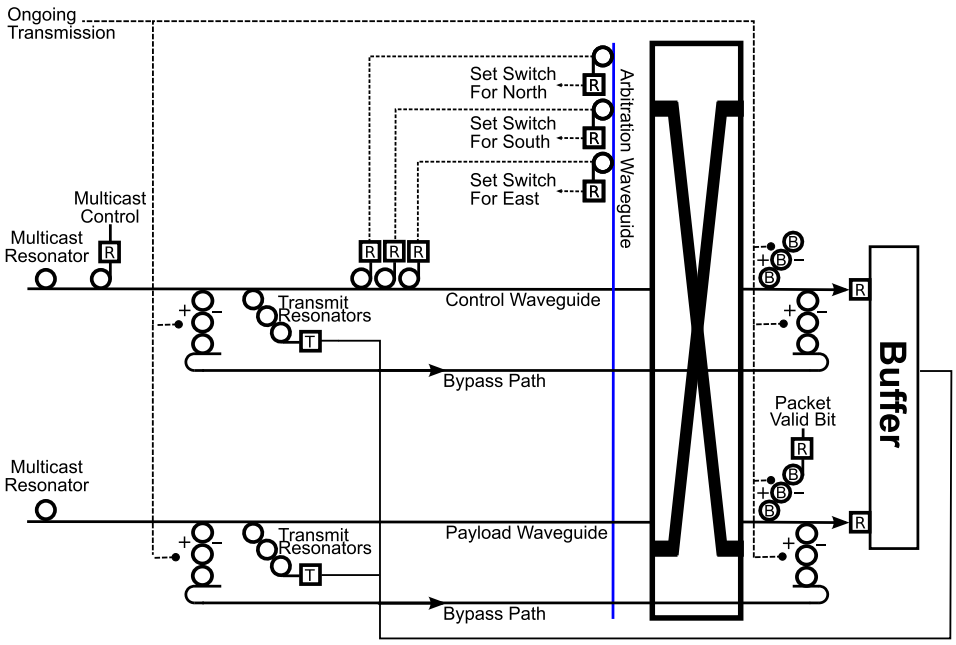


Figure 5.4: Upon transmission in the network, a packet will utilize the Transmit Resonators to enter the router prior to the control logic. Any upstream packet that arrives on the same input port during a packet transmission must be buffered in order to avoid packet collisions. We do this through the Bypass Path and Block Resonators (designated by 'B').

tion of the power transmitted on the input lines through separate broadcast resonator/receivers via the Multicast Resonators. Since only a portion of the power is extracted, the packet continues through the selected output port to the next router in the absence of contention. The Multicast Resonators are placed prior to the Bypass and Transmit Resonators so that a packet does not perform unnecessary multicasts when blocked, buffered and retransmitted. One way to implement a Multicast Resonator is to vary its size such that its resonant frequencies are slightly shifted from the frequencies used to carry the network packets. This allows it to couple only a small percentage of the packet's power.

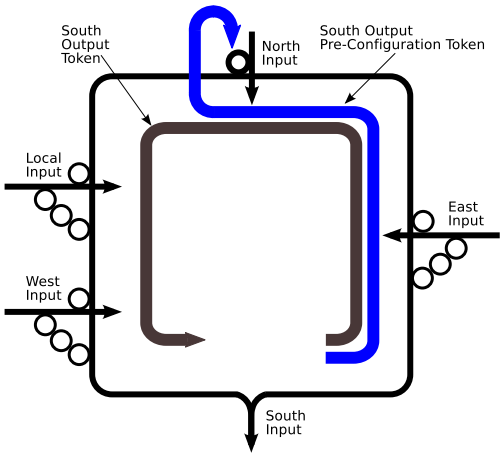
5.1.6 Interim Buffering

For large networks, all possible destinations may not be reachable in a single cycle. In these cases, the packet needs to be buffered at one or more interim nodes on its way to its final destination. We accomplish this using an Interim bit in every router's control group. When this bit it set, we force the packet to be buffered at that node. In the case that a packet can traverse four hops in a network clock cycle, one way of implementing this is to set the Interim bit in every fourth router control group along its network path at the source node prior to its transmission. If a packet is prematurely buffered due to losing switch arbitration at a node, that node recalculates the Interim bits.

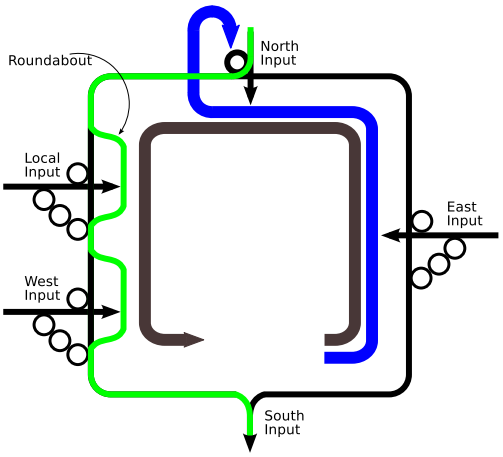
5.1.7 Switch Pre-Configuration

Because we implement dimension-order routing, a packet will spend most of its time traversing a router from the North port to the South port, East port to West port or vice versa. To minimize the per router hop delay, we implement a switch pre-configuration technique that statically joins the East/West and North/South router ports at the beginning of each network cycle prior to packet transmission. If an incoming packet enters an input port with a correctly configured output port, it continues through to the downstream router using a reduced latency path. Only when a packet desires an output port that differs from the straight output must it resort back to waiting for the control bits to properly set the switch as discussed in Section 5.1.3.

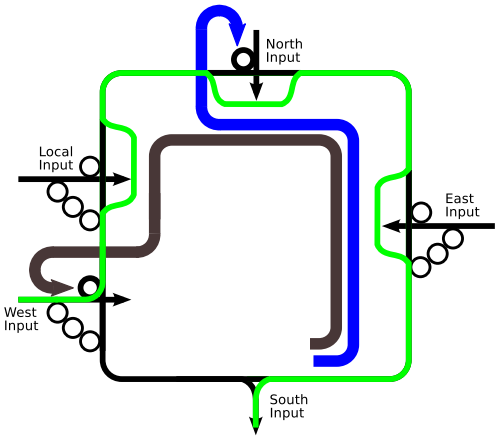
Input ports are statically configured to connect to the straight output ports through four additional tokens on the Arbitration Waveguide. We refer to these tokens as Pre-Configuration Tokens and they correspond to the North, East, South and West output ports. Thus the Arbitration Waveguide has four Pre-



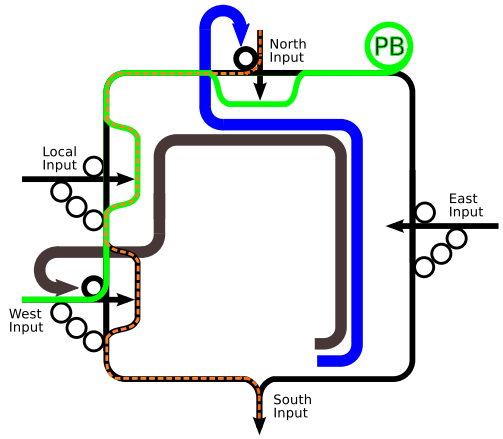
(a) North input is pre-configured for South output.



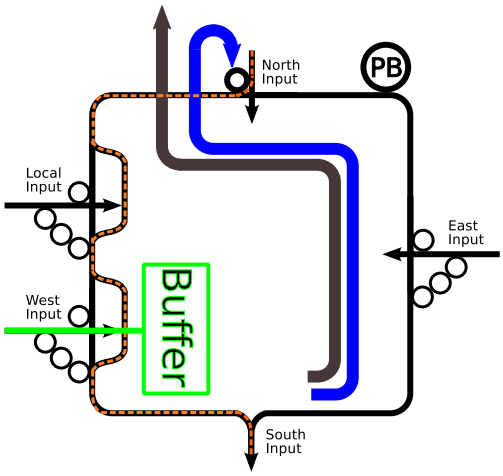
(b) North packet uses pre-configured route.



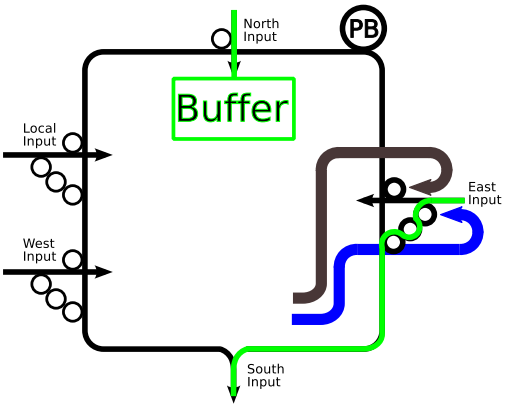
(c) Lower priority West packet uses South output.



(d) West packet blocked by North packet.



(e) West packet loses Output Token and buffers.



(f) High priority East packet uses South output.

Figure 5.5: East, West, North and South inputs are statically pre-configured to connect to straight path output ports. For clarity, only the ports connecting to the South output are shown.

Configuration Tokens and the four Output Tokens described in Section 5.1.3. In Figure 5.5a the North port statically pre-configures itself at the beginning of the clock cycle by taking the Pre-Configuration Token for the South output from the Arbitration Waveguide. It uses this token to turn on the Switch Resonators for connecting to the South output. The South Output Token remains on the Arbitration Waveguide. Following switch pre-configuration, a packet may enter the North input requiring the South output as shown in Figure 5.5b. Because the switch was previously set up to make this connection, the packet can traverse the router with a reduced delay path. Roundabout waveguides allow the packet to bypass the Switch Resonators of the other input ports. A roundabout waveguide in combination with the switch waveguide that it is attached to functions as an asymmetric y-branch for variable power splitting [41] [61]. Any light that is traveling through the switch that does not belong to the input port corresponding to the roundabout will couple into it entirely, allowing it to bypass that input port's switch resonators. Additionally, light from an input port will not couple into its own roundabout because of the y-branch functionality. For simplicity, the following discussion on switch pre-configuration will only refer to the input ports connected to the South output.

Switch pre-configuration still respects the rotating priority arbitration scheme introduced in Section 5.1.3. Any packet that enters the switch on the Local, West or East inputs and requires the South output will attempt to take both the South Pre-Configuration Token and the South Output Token. Packets on input ports with lower arbitration priority than the North can only access the South Output Token. This is shown in Figure 5.5c where the lower priority West input routes to the South using the Output Token to turn on one of its two sets of Switch Resonators. One set is turned on by the South Pre-Configuration Token and

the other by the South Output Token, where the former takes precedence. Thus because the West input was only able to take the Output Token for the South, it traverses the switch in the direction that forces it to pass by the North input. If in the same cycle a packet on the North enters the router and simultaneously requires the South output, it will take the South Output Token away from the West input and turn on the Pre-configuration Block Resonator. This is denoted in Figure 5.5d by the resonator with *PB*. When this resonator is turned on, the packet from the West input is blocked from leaving through the South, allowing the North packet to traverse the switch without having to wait for it to buffer. Shortly thereafter, the West input will be forced to turn off its Switch Resonators (since it no longer has the Output Token) and buffer as shown in Figure 5.5e. However, the incoming packet on the North does not have to wait for this to occur before leaving the router.

In Figure 5.5f the East input has higher arbitration priority than the North input, allowing an incoming packet there to take both tokens for the South output. In this case both sets of Switch Resonators are turned on, but precedence is given to the set turned on by the Pre-Configuration Token, which routes the packet in the crossbar away from the North input. This occurs because, if in the same cycle a packet on the North port enters the router and also wants to leave through the South, it will turn on its Pre-Configuration Block Resonator regardless of whether it still has access to the switch. However, because the North packet no longer has its Switch Resonators turned on, it will be buffered as shown in the diagram.

If an incoming packet on the North port requires other than the pre-configured South port, the packet must turn on the appropriate resonators in the Arbitration Waveguide, attempting to take both the Pre-Configuration and Output Token for the desired output. Thus the router delay for a packet that enters a port with an

incorrectly pre-configured path, or a packet that enters through the Local input port (which does not have a pre-configured route) is the same as when Switch Pre-Configuration is not supported.

The addition of Switch Pre-Configuration requires the flow control implementation to be slightly modified. When a downstream buffer is full, the flow control signal that propagates back must turn on Terminator Resonators for both the Output Token and the Pre-Configuration Token.

5.2 Optical Router Design Analysis

5.2.1 Critical Delay

The critical delay timing components of our proposed optical switch architecture can be divided into three broad categories. The first category is Router Setup, which is composed of switch tasks that are completed prior to packets transmitting into the network. These tasks consist of setting up the optical switch arbitration including turning on the appropriate Priority Coupler, Stop Resonators and propagating the output tokens around the Arbitration Waveguide. This step also involves turning on the Transmit Resonators and associated Bypass Resonators and Block Resonators. Lastly, flow control signals from downstream routers transferred during the previous cycle are used to set the proper Terminator Resonators on the Optical Power Waveguide. In parallel with router setup, if supported, switch pre-configuration statically configures the network switches.

The second timing category is Router Traversal and consists of two possible delay paths through a network router. The first path requires the packet to wait for control bit translation, switch arbitration and setup prior to entering the crossbar. This occurs when a) a packet enters through the Local input, which has no pre-

Component	Experimental delay (ps)
Optical Transmitter	33 ps
Optical Receiver	7 ps
Optical Comb Filter	29 ps
Optical Signal Propagation	6 ps/mm [26]

Table 5.1: Predicted optical component delay values for 16nm.

configured route, b) a packet makes a turn, or c) pre-configuration is not supported. The second type of path occurs when switch pre-configuration is supported and matches the desired output of a packet. Here, the optical packet continues through the router with a minimally impeded delay.

The last timing category, Cycle Termination, occurs at the end of a clock cycle when a packet enters an input port and uses its Interim control bit to buffer. This consists of receiving the Interim control signal, turning on the Bypass Path and buffering the packet.

In parallel with network packet transmission, each input port buffer performs an appropriate flow control action. This involves determining the number of free buffer entries and sending an off signal to an upstream router if necessary.

The individual delay parameters for the optical components used in our critical delay analysis are found in Table 5.1 and are based on our optical device projections from Chapter 3. We determine that our switch pre-configuration scheme allows a packet to traverse four hops in a single 4GHz network cycle, versus only two hops with no pre-configuration.

5.2.2 Area

The area of the optical components in our proposed router should not exceed the area of the electrical components in a network node, otherwise the latter will need to artificially increase in size to line up the related components. Moreover, the

Transmitter Through Loss	1.1dB
Demux Insertion Loss	0.6dB
Comb Filter Insertion Loss	0.1dB
Comb Filter Through Loss	0.1dB
Waveguide Propagation Loss	0.1dB/cm
Laser Chip Coupling Loss	3dB
Required Laser Power	161W

Table 5.2: Phastlane 2.0 optical loss projections.

electrical components of the router, such as the resonator drivers and receiver amplifiers, should only marginally increase the area of the processor die.

To estimate the area of the processor die, we adopted the methodology of Kumar et al. [39] for 16nm technology. For a single processor core with 64 KB L1 caches, a 2MB L2 cache, and Memory Controller the total area is approximately 3.5mm^2. For two cores and four cores sharing an L2 cache, the area is approximately 4.5mm^2 and 6.5mm^2, respectively. In this study we assume a concentration factor of two per router input (i.e., a network node consists of four processors and pairwise sharing of an L2 cache, two of which share a router input port). The area of the optical components of our proposed router consume approximately 8mm^2 under the assumption that a router's datapath uses 24 waveguides to route a single flit packet, allowing it to be deposited above the processor without the need to grow its area. The electrical components of the optical network which facilitate the communication between the electrical and optical domains (i.e., receiver amplifiers and transmitter driver circuitry), consume approximately 0.12mm^2 per router on the electrical die. This represents a 3% area overhead over a single processing core.

5.2.3 Optical Power

In this work, we assume that a laser externally supplies light to the on-chip interconnect through vertical coupling and incurs signal attenuation through the

Flits per Packet	1 (80 bytes)
Routing Function	Dimension-Order
Number of VCs per Port	4
Number of Entries per VC	1
Wait for Tail Credit	YES
VC_Allocator	ISLIP [45]
SW_Allocator	ISLIP [45]
Total Router Delay	2 cycles

Table 5.3: Baseline electrical router parameters.

previously calculated loss components shown in Table 5.2. When the laser couples into the chip, it incurs a 3 dB loss [33]. Traversing through the modulator ring array and also at the end of the optical link to demultiplex wavelengths, losses of 1.1dB and 0.6dB, respectively, occur. We also model the loss traveling past and into the comb filters in the optical crossbar, and propagation losses in the silicon nitride waveguides [26].

We calculate the laser requirements based on the optical power that each node requires to be able to transmit up to four packets through its output ports, broadcasting a portion of their power to every subsequent node they traverse. The laser is always on and thus contributes to the static power consumption of the network, albeit externally to the chip. Based on our analysis in Chapter 3, we found that a detector has a responsivity of 0.44A/W and requires 40μW of optical power to achieve a reasonable BER. Using these parameters, we estimate that the chip will require 161W of optical power to handle the requirements of the network. In Chapter 7, we discuss and propose potential methods for mitigating this large power requirement.

Simulated Cache Sizes	32KB L1I&L1D, 256KB L2
Actual Cache Sizes	64KB L1I&L1D, 2MB L2
Cache Associativity	4 Way L1,16 Way L2
Block Size	32B L1, 64B L2
Memory Latency	80 Cycles

Table 5.4: Memory parameters.

5.3 Evaluation Methodology

To evaluate our proposed optical network, we developed a cycle-accurate network packet simulator that models components down to the flit-level. The simulator generates traffic based on a set of input traces that designate per node packet injections. In order to do a power comparison with the electrical baseline, we model dynamic power consumption and static leakage power in a manner similar to [33].

We evaluate the electrical baseline network using a modified version of Booksim [19] augmented with dynamic and static leakage power models. The models use CACTI for buffers, and the methodology of [3] for all other components. We also implemented Virtual Circuit Tree Multicasting [28] to perform packet broadcasts. Finally, we changed Booksim to input the same trace files used for our optical simulator.

The electrical baseline is an aggressive router optimized for both latency and bandwidth. The router assumes a virtual-channel architecture with the parameters shown in Table 5.3. In order to perform a fair performance comparison with our optical configurations, we assume both low latency and high saturation bandwidth for the electrical network. We reduce serialization latency by using a packet size of one flit, the same as in our proposed architecture. Doing so also gives no bandwidth density advantage to the optical network. We further assume that pipeline speculation and route-lookahead [53] reduce the per hop router latency of

the baseline electrical router to 2 cycles for every flit. The bisection bandwidth of the electrical and optical systems are matched at 4 TB per second.

We evaluate SPLASH2 benchmarks and synthetic traffic workloads. By varying the injection rates of the synthetic benchmarks, we obtain saturation bandwidth and average packet latencies. We created SPLASH2 traces using the SESC simulator [60]. The modeled system consists of 64 cores with private L1 and L2 caches. Each core is 4-way out-of-order and has the cache and memory parameters shown in Table 5.4. As is typical when using SPLASH2 for network studies, the cache sizes are reduced to obtain sufficient network traffic.

Total execution times of the Splash benchmarks are found using the average packet latencies from the network trace simulations. These results form a static network latency in SESC on top of which each Splash benchmark is run to completion. Finally, we assume a 16nm technology node operating at a 4GHz processor and network clock with a supply voltage of 1.0V.

5.4 Results

In this section, we examine the latencies of the optical transmitter and receiver in the context of the Phastlane 2.0 nanophotonic router architecture. We first provide a basic background of the key devices that form the critical path through the network. Using the device modeling analysis from Chapter 3, we provide realistic performance and power consumption estimates tailored towards the requirements of the network architecture.

Using these parameters, we present power and performance results for our network architecture against an aggressive electrical baseline. We start with performance results for Splash and synthetic benchmarks, and conclude with power results.

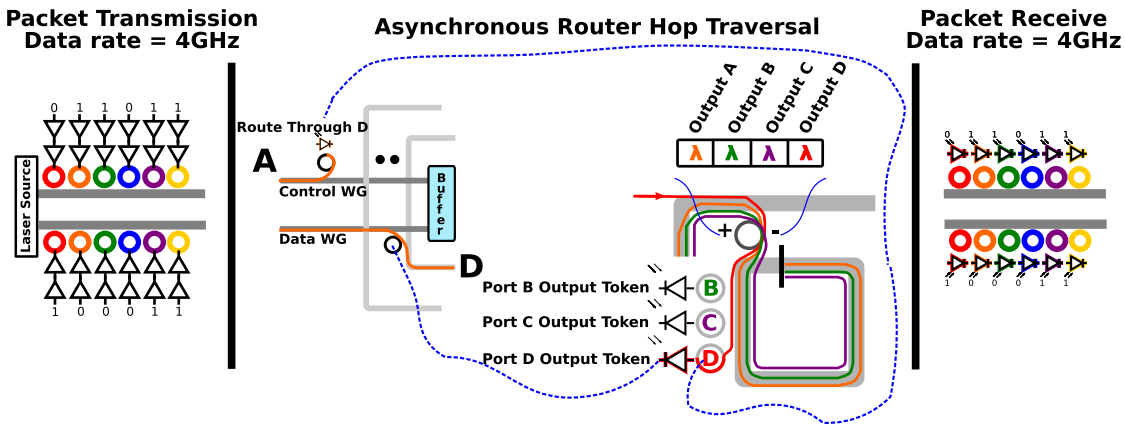


Figure 5.6: At the beginning of every network clock cycle packets are transmitted into the Phastlane 2.0 network using only WDM to encode the packet's data. Packets traverse multiple asynchronous hops between source and destination. Upon entering an input port, a portion of the packet's pre-computed control bits are electrically translated to participate in switch arbitration. An optical arbitration bus implements a high-speed, rotating priority token scheme that utilizes ring resonators on an Arbitration Waveguide to compete for output ports. Assuming that an input port wins arbitration and is able to sink the token corresponding to its desired output port, this signal will form a driving voltage across the appropriate comb filters in the crossbar. The optical packet is then routed through the crossbar and to a downstream switch. Packets are electrically buffered at the end of a clock cycle, or in the event that switch arbitration is lost.

5.4.1 Critical Network Components

The data path a packet takes between source and destination in Phastlane 2.0 is shown in Figure 5.6. At the beginning of every network clock cycle the ring modulators transmit the signal into the network. Following this, the packet traverses potentially many asynchronous router hops before it reaches its final destination node. Within each router, the packet uses precomputed routing bits to arbitrate for an output port and traverse the crossbar. If arbitration is lost, the packet is buffered at the end of the cycle. Otherwise it continues from router to router until it reaches its destination, or it is forced to buffer at the end of the network cycle.

Both of the Phastlane architectures presented in this dissertation are unique in that they are highly dependent on the latency of the optical transmitters and receivers. In Figure 5.7 we show the key devices that form the critical path through

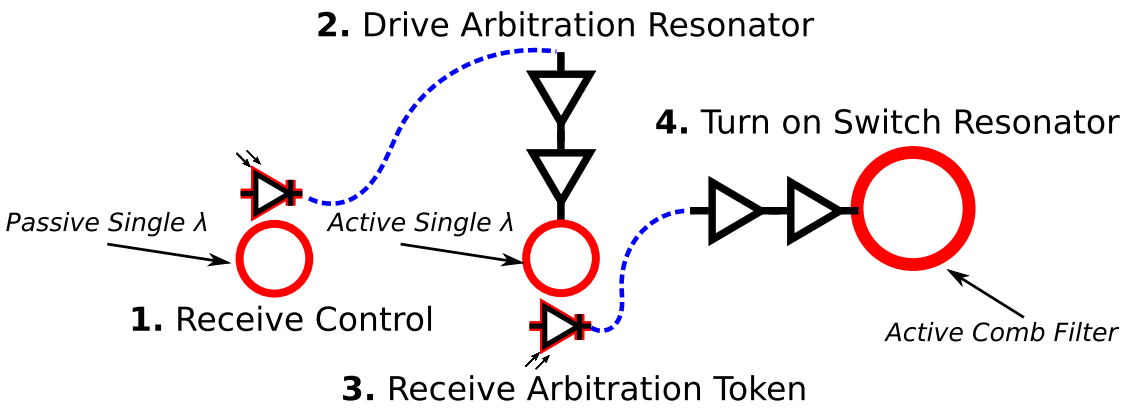


Figure 5.7: The critical components of an asynchronous optical router in Phastlane 2.0 without switch pre-configuration. Upon entering an input port, a portion of a packet's control bits are electrically translated and used to drive a ring resonator on the Arbitration Waveguide to compete in switch arbitration. Assuming that it wins arbitration, the optical token is electrically received and used to form the driving voltage across a comb filter in the crossbar. Once this filter is turned on, the packet is free to traverse the crossbar.

a Phastlane 2.0 optical router without switch pre-configuration. When a packet first enters the router, a portion of its pre-computed control bits (contained within the packet) are optically received to form a driving signal across the appropriate ring resonator on the Arbitration Waveguide. This allows the packet to arbitrate for an output token associated with its desired output port. If the token is available, it will be optically received and used to drive a comb filter in the switch. Following this, the packet is free to traverse in the crossbar and leave to the next downstream router. To optimize the number of hops that a packet can take in a single cycle, it is important that these basic components offer ultra low latency.

Optical Receiver Latency

Based on the optical modeling presented in Section 3.4, we concluded that a receiver data rate of 25 GHz would be possible by the 16nm technology node. Using Equation 3.30 from Section 3.4 we found the latency of the front-end portion of the receiver to be 4.5ps (without the detector). In Section 3.4.1, we calculated

the latency of the photodetector using Equation 3.26 to be 2.84ps. Thus the total latency of the full receiver is approximately 7ps.

Arbitration Waveguide Resonator

A packet activates the appropriate resonator on the Arbitration Waveguide to receive a token corresponding to its desired output port. Since the output token is immediately received and used to drive the signal across the comb filter, the loss leaving out the Arbitration Waveguide resonator is not vital to its operation. Using the data from Figure 3.16, and a resonance shift amount λ_o = 1FWHM, we find that the ring can operate at a data rate of approximately 37Gb/s. The resulting latency of the device can be found using:

$$\text{Latency}_{\text{Ring}} = \frac{0.5}{\text{Ring}_{\text{BW}}} \tag{5.1}$$

Using this equation we found the latency of the resonator to be 13.5ps.

Transmitting Into The Network

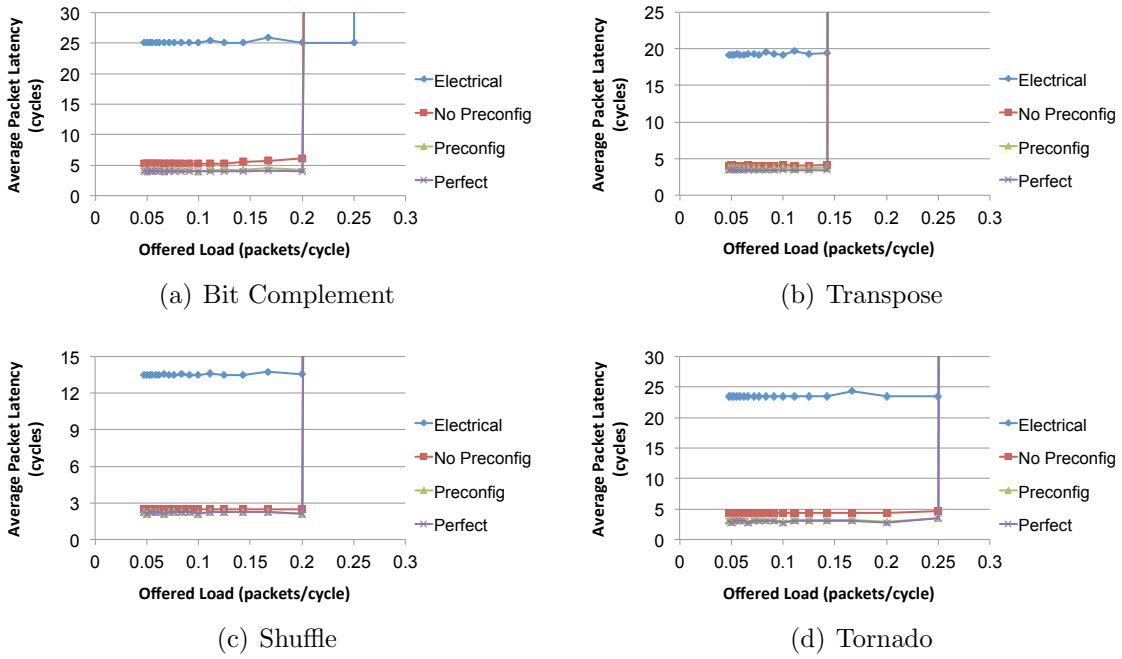
The transmission of a packet into the network at the beginning of the clock cycle trades off latency for WDM as shown by the data in Figure 3.33. If we assume a channel spacing of 3 FWHM, it's possible to achieve a high data rate (and thus low latency) at the cost of WDM, or more WDM at the cost of data rate. To balance between the two, we choose a data rate of 15Gb/s, which allows us to use a WDM level of 35 wavelengths. It is important not to reduce the level of WDM too much as this forces us to serialize packets through the network since we do not utilize time-division-multiplexing (TDM) for transmitting packets. Using Equation 5.1 we obtain an initial packet transmit latency of 33ps.

Level of WDM	35
Receiver Latency	7ps
Optical Transmitter Latency	33ps
Comb Filter Latency	29ps
Arbitration Ring Latency	13.5ps
Data Path Width (WGs)	24
# flits per packet	1
# Hops with Pre-Configuration	4
# Hops without Pre-Configuration	2
Total Node Area	$8mm^2$
Channel Spacing (units of FWHM)	3
Number of Optical Layers	2

Table 5.5: Phastlane 2.0 device parameters.

Optical Comb Filter

The optical comb filter is the fundamental component of the crossbar in both Phastlane architectures. It allows all of the wavelengths in a packet to be simultaneously switched to a router output port. One of the challenges associated with this ring resonator is its large size ($\sim$100μm in diameter for an assumed channel spacing of three FWHM and WDM = 35λ to match the parameters of the ring transmitters used to inject packets at the beginning of every cycle). Using an archimedean configuration, it's possible to achieve a 70 fold reduction in size of the total footprint of the ring, allowing us to fit the comb filter in an area of approximately 1×10^{-4} mm^2 [69]. Although the footprint of the ring is reduced using archimedean folding, the amount of charge that must be injected for a particular resonance shift is the same as for the ring prior to folding. To mitigate the potentially high driving voltage of the comb filter (i.e., above the Vdd supply of 16nm), we assume a 3dB loss going through the ring. Instead of shifting it 1.5FWHM according to Figure 3.28 for a channel spacing of three, we only shift it .75FWHM. Additionally, we assume that four inverting drivers inject charge into four separate regions of the ring. This allows the drivers to utilize a high level of



(a) Bit Complement

(b) Transpose

(c) Shuffle

(d) Tornado

Figure 5.8: Average packet latency as a function of injection rate for four synthetic traffic patterns. We show results for the two cycle electrical baseline, denoted as *Electrical*, and our optical configurations, *No Preconfig* (2 hops), *Preconfig* (4 hops) and *Perfect* (full network diameter).

ion implantation, and thus a low device latency, without the ring requiring a drive voltage higher than the inverters are able to supply. Using these assumptions, we calculate the latency of the comb filter to be 29ps.

5.4.2 Performance Results

We begin with performance results for four synthetic benchmarks shown in Figure 5.8. Three optical configurations are presented, *No Preconfig*, *Preconfig* and *Perfect*, representing Phastlane 2.0 routers without pre-configuration, with pre-configuration, and that can reach the entire extent of the network in a single cycle, respectively. The baseline electrical network is denoted as *Electrical* in our results. Based on our previous analysis in this section, we use the device parameters from Table 5.5 for the rest of our analysis in this chapter. As with the results from

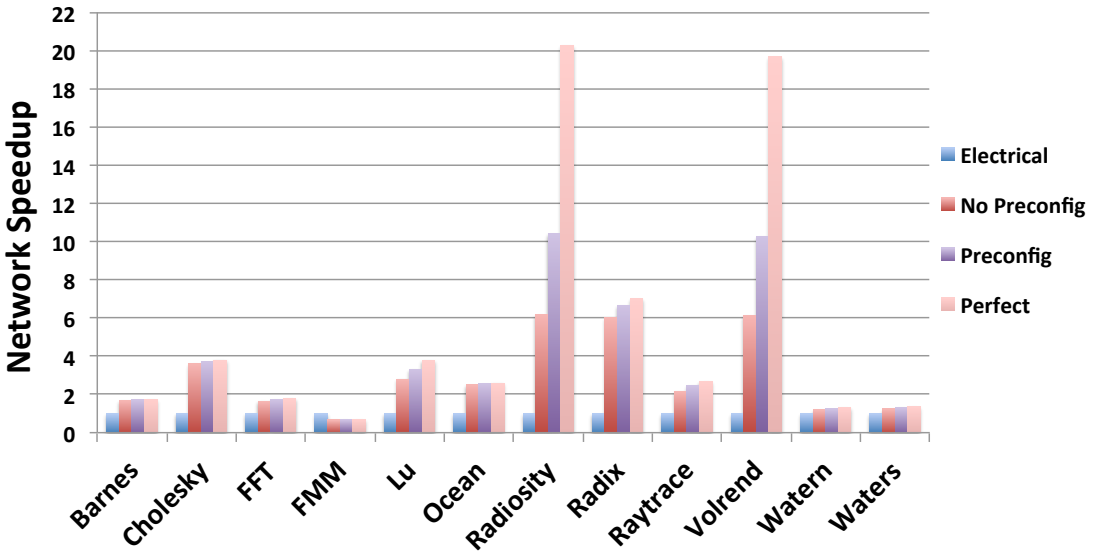


Figure 5.9: Network performance results for Splash benchmarks. We show results for the two cycle electrical baseline, denoted as *Electrical*, and our optical configurations, *No Preconfig* (2 hops), *Preconfig* (4 hops) and *Perfect* (full network diameter).

Chapter 4, the synthetic benchmarks are insensitive to the different optical network configurations. This is due to many of the source/destination pairs being close enough to not require more aggressive devices, and also because of switch arbitration, which forces losing packets to buffer.

Network performance results for Splash are shown in Figure 5.9 relative to the electrical baseline over the range of different Phastlane 2.0 network configurations. Across all of the benchmarks, the *No Preconfig* configuration achieves a 2X speedup, the *Preconfig* a 4X speedup, and the *Perfect* configuration that can reach the entire extent of the network a 6X speedup. Phastlane performs worse on FMM because the virtual channels in the electrical baseline, which are not present in our optical router, enable "turning lanes" which help increase the network saturation point. This is also the case in the synthetic benchmark Bit Complement.

Lastly, we demonstrate how the optical router architecture that uses switch preconfiguration impacts total system performance (i.e., execution time) of the shared memory architecture. These results are shown in Figure 5.10, demonstrating a

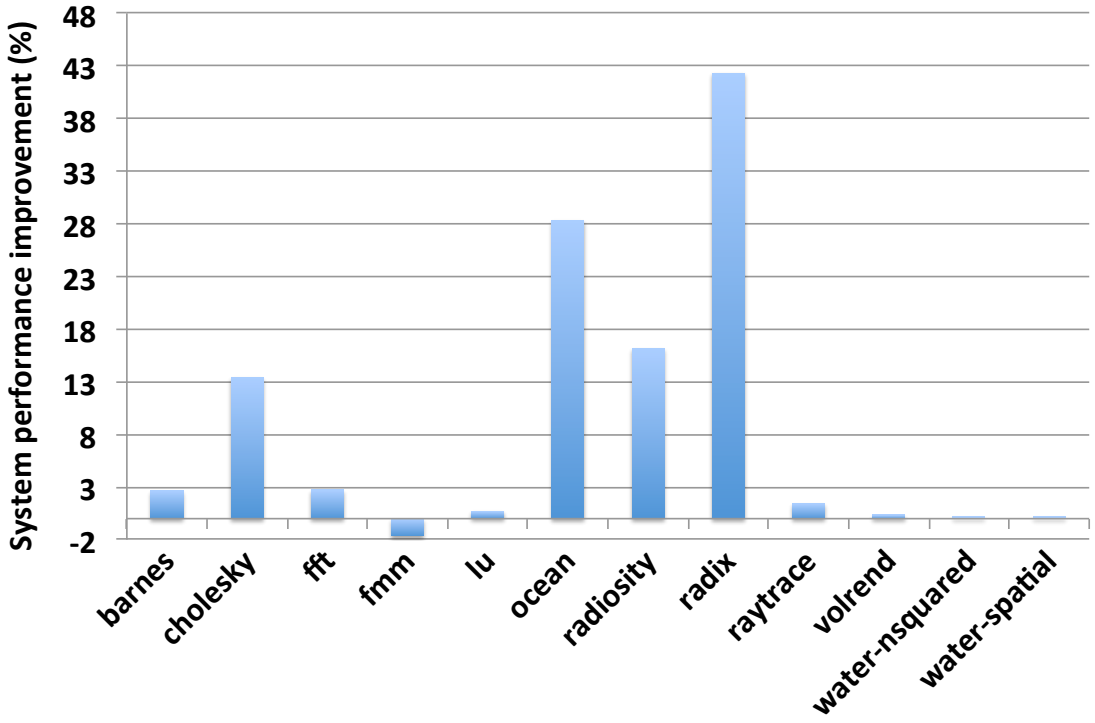


Figure 5.10: Relative system performance for the Splash benchmarks using the *Pre-config* configuration against the electrical baseline network. Across all the benchmarks, Phastlane 2.0 achieves an 8.9% speedup.

9% improvement in system performance over the baseline architecture using the aggressive electrical network.

5.4.3 Power Results

In this section, we present the on-chip power consumption (i.e., excluding external laser requirements) of our optical architecture against the electrical baseline. The device parameters that we assume for each building block are derived from our model and shown in Table 5.6. These building blocks are the optical transmitters for inserting optical packets into the network, the comb filter and arbitration ring resonators and the receivers for buffering optical packets at the end of every network clock cycle.

Component	Power	Energy/bit
Receiver	42.5mW	5pJ/bit
Optical Transmitter	85mW	10pJ/bit
Optical Comb Filter	400mW	50pJ/bit
Arbitration Ring	100mW	12.5pJ/bit

Table 5.6: Phastlane 2.0 optical device energy consumption.

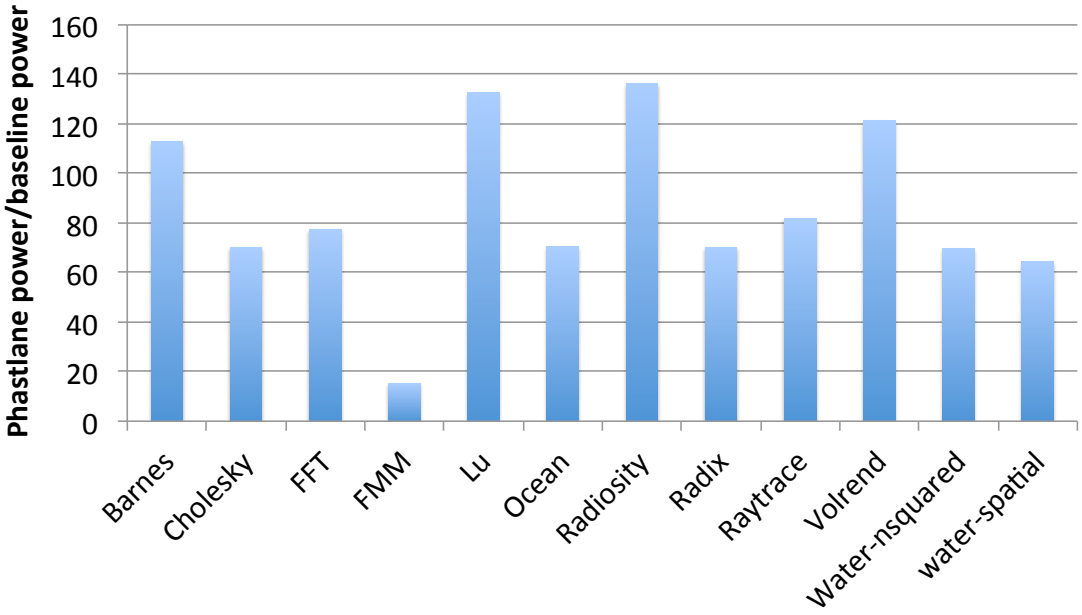


Figure 5.11: Relative network power consumption results for Splash benchmarks using the *Preconfig* configuration against the electrical baseline network. We examine potential ways to mitigate the high power consumption of our optical architecture in Chapter 7.

We show power consumption results for the optical architecture with preconfiguration in Figure 5.11. As with the Phastlane architecture first presented in Chapter 4, the power consumption is well above the electrical network due to our device energy projections, which must be lowered from pJ's/bit to hundreds of fJ's/bit in order to show improvement. In Chapter 7 we examine some techniques that could be used to mitigate this steep energy requirement.

In Figure 5.12 we show Splash power results assuming aggressive optical device scaling [7]. Here, the optical modulator consumes 120fJ/bit and the receiver 80fJ/bit. Across all of the benchmarks, the average improvement in power con-

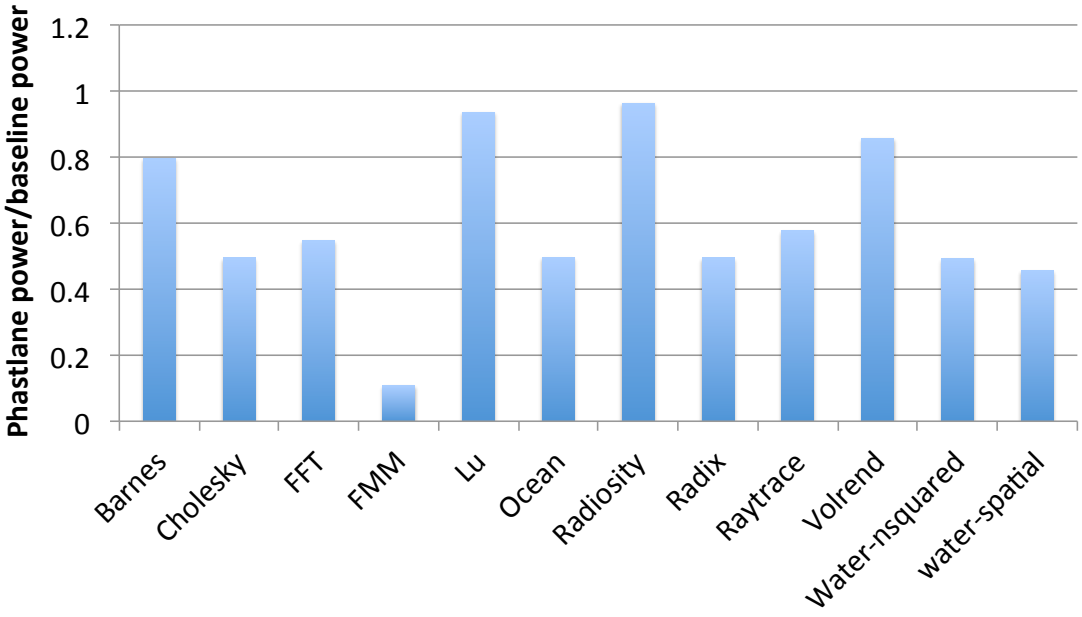


Figure 5.12: Relative network power consumption results for Splash benchmarks using the *Preconfig* configuration against the electrical baseline network. Optical receiver and transmitter energy consumption is optimistically scaled to 80fJ/bit and 120fJ/bit, respectively [7]. The average power reduction across all of the benchmarks is 40%.

sumption is 40%. These results demonstrate the importance of continued device innovation and the resulting improvements in power consumption that could follow.

CHAPTER 6

CONCLUSIONS

In Chapter 4 we present Phastlane, a hybrid electrical/optical routing network for future large scale, cache coherent multicore microprocessors. The heart of the Phastlane network is a low-latency optical crossbar that uses simple predecoded source routing and fixed priority switch arbitration to transmit cache-line-sized packets several hops in a single clock cycle under contentionless conditions. When contention exists, the router makes use of electrical buffers and, if necessary, a high speed drop signaling network. We examine performance and power consumption against an electrical baseline using the scaled optical device projections from our model in Chapter 3. On a set of ten SPLASH2 benchmarks, Phastlane achieves 1.7X better network performance, but at a cost of increased power consumption, a problem that we address in Chapter 7. However, if further innovation reduces modulator and receiver energy consumption from pJ's/bit to 100's of fJ/bit, the on-chip power consumption reduces to 30% below the baseline.

In Chapter 5 we introduce Phastlane 2.0, a novel optical router architecture that builds on Phastlane through the complete redesign of the optical router. We present a new switch architecture that localizes all router control within each input port, removing any delays associated with propagation of electrical control signals. We also incorporate an optical implementation of rotating priority switch arbitration, guaranteeing fairness to all packets. On/off flow control is introduced to remedy the potential problem of dropped packets under periods of high contention. Lastly, we present a mechanism for pre-configuring switch state by joining straight path ports at the beginning of every clock cycle, allowing a packet to achieve ultra-low network latency. On a set of twelve SPLASH2 benchmarks, Phastlane 2.0 achieves 4X better network performance. However, as with the original Phast-

lane architecture, the network consumes more power than the electrical baseline using our optical device projections. Further innovation in receiver and modulator energy consumption from pJ's/bit to 100's of fJ/bit will reduce the on-chip power consumption to achieve a 40% savings over the baseline.

Lastly, we present some key design strategies for implementing nanophotonic interconnection networks that demonstrate how to exploit its benefits while avoiding its weaknesses. We develop these rules based on our Phastlane architectures and detailed device level model in Chapter 3.

- The main benefit of nanophotonics over electrical wires is high bandwidth density using wavelength-division-multiplexing (WDM) and time-division-multiplexing (TDM). We demonstrate that network packets can be modulated at a maximum data rate of 25 Gb/s in scaled technology nodes enabling an aggregate bandwidth per link of over 200 Gb/s.

- Total energy consumption from an external laser source, modulators and receivers is the primary design constraint in a nanophotonic interconnect. In Chapter 7 we identify some solutions to mitigate the power consumption of these components, which can quickly become large if the interconnect is not properly designed.

- The maximum data rate of a ring modulator can be improved over the limits imposed by the carrier recombination lifetime using pre-emphasis, reverse bias depletion or ion implants. We focus on the latter because it enables compatibility with scaled CMOS voltage supplies through a simple inverting driver circuit, achieving as high as 30 Gb/s at 16nm.

- The maximum data rate of an optical receiver is set by the required BER and available optical power at the photodetector, which directly impacts the amount of external laser power that needs to be supplied to the interconnect.

Packets should have error correction/detection bits embedded in them so that the required power at the detector can be lowered while still achieving an acceptable BER. In our results, we show that using parity bits to protect every two bytes of data in a cache line sized packet, a BER of 10^{-13}, yields an undetectable error in the system every twenty five years. However, this could be improved by using more parity bits, or more complex error detection schemes. We found that the receiver in scaled technologies has a maximum data rate of 25 Gb/s, and that the optical power at the detector should be adjusted between 10μW and 40μW to obtain the required BER.

- Attention should be given to the integration strategy for fabricating the nanophotonic devices and the various tradeoffs. In Section 2.1 we present two different materials for constructing the optical components. For example, the waveguides can be fabricated in single crystalline silicon with a propagation loss of around 1dB/cm and signal latency of 10.45 ps/mm, whereas silicon nitride has a lower propagation loss of .1dB/cm and signal latency of 6 ps/mm, but larger area requirements. The latter material also enables multiple waveguide layers, which could avoid excessive power loss in complex network topologies that require many waveguide crossings. Similar design consideration should be given to the ring resonators, where polysilicon can be deposited with the silicon nitride waveguides, but has approximately 10X the propagation loss of single crystalline silicon.

Overall, electrical wires cannot match the superior bandwidth density of optical waveguides. Energy consumption should be the primary constraint when designing a nanophotonic interconnect, but could be mitigated through careful choice of the network topology and flow control. The utilization of the external laser supply should be maximized since it adds to the total power consumption.

CHAPTER 7

FUTURE WORK

In this chapter, we first examine fundamental challenges to the integration of nanophotonics in future chip multiprocessors. We then show methods for reducing the power consumption of the Phastlane architectures, a problem that was presented in Chapters 4 and 5. There we found the laser power requirements and electrical energy of the optical components to be higher than the total power consumption of the electrical baseline. Other improvements to the Phastlane architectures are also proposed, including exploiting time-division-multiplexing, increasing the router radix for use in other network topologies and mitigating the overheads associated with switch arbitration. Utilizing our optical device model from Chapter 3, we conclude this chapter with a novel architecture that follows the design guidelines from Chapter 6 and combines the advantages of electrical wires and optical waveguides to form a hybrid interconnect. We present a basic blueprint for this design, proposing future work to further examine its potential for improving performance in a chip multiprocessor without requiring excessive energy consumption.

7.1 Fundamental Challenges

In this dissertation, we present an extensive overview of the basic building blocks of a nanophotonic interconnect for future chip multiprocessors. We provide a background of the optical devices and recent architectural level research that has examined how to use these components to benefit the performance and power consumption in communication networks. However, large challenges still exist in forming a successful union between optical devices and conventional CMOS transistors to demonstrate a functional system.

The first challenge is to successfully integrate optical components that interface with controlling transistors such that neither has to sacrifice power, performance or density. One recent proposal uses a standard bulk CMOS process and its polycrystalline silicon layer to form waveguides and ring resonators, the fundamental building blocks of an optical link [47]. Although this method is appealing from the point of view of design cost, challenges still exist in reducing resulting waveguide propagation loss below the achievable 55dB/cm, and in determining how to efficiently detect light to perform optical to electrical conversion. Additionally, monolithically integrating the optical components uses potentially valuable transistor real estate. Other work that may address these problems separates the nanophotonic and CMOS components from one another using dual 3D integrated layers. Previous research in this area has examined flip chip bonding to join two dies, one optimized using a Luxtera-Freescale 130nm non-standard SOI process specifically targeted for optics, and the other using a standard 90nm bulk CMOS technology [73]. However, only a single modulator was fabricated in this work. Others have examined epitaxial growth of silicon islands [48], oxygen ion implantation [36] and wafer bonding [23] to form a vertical optical layer, none of which are compatible with a standard CMOS technology. Back-end-of-line deposition of polycrystalline silicon and silicon nitride above a pre-fabricated electrical die has the benefits of enabling multiple waveguide routing layers and uses standard CMOS fabrication techniques [56]. However, still in its nascent stages, it is unclear whether this technology will come to fruition.

Another problem is the extreme temperature sensitivity of current optical devices, which cease to function as designed with changes in temperature as small as 1° Celsius [47]. This extreme temperature sensitivity makes their practical use in an uncontrolled environment impossible. To combat this problem, heaters can be inte-

grated next to resonators, red shifting their wavelength response (i.e., moving them to larger values) as die temperatures fall below a ring's design point [22] [47] [67]. However, these circuits have a nontrivial power cost that's compounded by the use of over a million rings in some recently proposed nanophotonic architectures [65]. Research in this area has demonstrated that for a crossbar network with ∼500K resonators on a 484mm^2 die, the trimming power due to heaters to correct for a temperature range of 20° would require a maximum of 100W [46]. This work also shows that the use of ring carrier injection to blue shift (i.e., move to lower wavelengths) ring resonances in combination with heaters leads to thermal runaway. Various work has addressed these issues by including *spare* rings that are not used under normal operation, but allow heater power to be mitigated by inserting additional resonances at the front and back of the WDM spectral window [9] [46] [63]. However, this adds a nontrivial area cost that depends on the granularity of temperature fluctuations in the system and thus the number of rings that can be grouped into banks.

Assuming that successful integration of nanophotonics is achieved, arguably the greatest limitation in exploiting its potential bandwidth and energy benefits comes from the electrical interfacing circuitry. Recently proposed modulators are ring based and are turned on and off through PIN diode carrier injection, or PN diode depletion [55] [73]. The idea behind these approaches is to inject or remove charge carriers from the ring resonator to shift its effective index of refraction, causing its resonance peaks to blue or red shift, respectively. However, there are two primary challenges with these approaches that hinder the rate at which a ring can be switched using a conventional CMOS driver. The first is the latency required to turn the ring on and off, which is dominated by slow carrier injection or depletion characteristics. Previous work has examined how to overcome these

limitations using PIN diode carrier injection and pre-emphasis [70], but at the cost of requiring driving voltages beyond the reach of scaled CMOS technologies. The fundamental switching speed of the resonator is set by its photon lifetime, which is typically on the order of a few pico-seconds [42], resulting in bandwidth close to Tb/s. Recently proposed modulators are still well below this limit, operating in the low GHz range [55] [70]. The second challenge is the driving voltage across the ring to obtain reasonable extinction and low optical insertion loss. One technique uses ion implantation [66] [68] for reducing the carrier recombination lifetime of silicon and thus the latency to switch the ring, but at the cost of driving voltage, which must grow to offset the increased optical absorption by the implants. We explore the use of ions and the tradeoffs associated with driving voltage, latency and propagation loss in Chapter 3 as a means of achieving high data rate in a scaled CMOS technology.

7.2 Phastlane Architectures

A challenging problem in designing optical interconnects is overcoming the high power requirements of a statically tuned, external laser source. Since this laser cannot be dynamically modified to suit the actual power requirements of the network architecture, it must be provisioned for worst case behavior. This results in many optical components in the network that are supplied with light, wasting energy as they idly wait to deliver requests and responses from the underlying processors and memories. Future work must look at the laser as a shared resource that can be distributed to requesting optical transmitters.

In our Phastlane architectures every input port of every node must have laser power to be able to transmit packets in the worst case situation (i.e., every input port sends a packet to an output port that travels the furthest hop count the

optical components can support each cycle). However, this worst case situation is impossible to achieve for any variation of Phastlane where the packet can traverse at least two hops per cycle. This is because the laser is statically distributed via fibers following chip fabrication, removing any possibility of dynamic tuning. Future work should examine how to distribute, share and arbitrate for laser power.

The other component of power consumption in a nanophotonic interconnect is from the electrical interfacing circuits that modulate, switch and receive optical packets. This component results in both Phastlane architectures consuming considerable power because of the large amount of switching that occurs between a packet leaving its source node and reaching its destination. Broadband comb filters have considerably more waveguide area than the transmitting and demultiplexing rings, forming a large portion of the total power dissipation. One way to mitigate the amount of switching is to utilize a different network topology. A mesh network has a large diameter relative to a flattened butterfly, which utilizes cross chip links to interconnect routers in the same rows and columns. However, implementing this topology requires research in scaling the radix of the Phastlane router architectures. Another way to mitigate the amount of switching, and thus energy consumption, would be to guarantee that once a packet is injected into the network, it reaches the maximum number of hops that it can traverse per cycle (i.e., based on the critical delay of the optical devices) to minimize the amount of switching between source and destination. This injection technique seeks to avoid transmitting an optical packet unless it can encounter very little contention that could cause it to prematurely buffer.

Switch arbitration is the largest hindrance to achieving more performance in both Phastlane routers. In this dissertation, we utilize two variations of optical arbitration: a fixed priority scheme that turns certain input ports off by detuning

their ring resonators, and rotating priority through an optical token bus. However, it is unclear whether there are opportunities to improve performance and power dissipation with electrical alternatives that are specifically targeted for ultra low latency. Additionally, another option that could be implemented electrically or optically uses a token migration policy that takes advantage of traffic phases in the network to reduce latency. In this policy, a token corresponds to winning the use of an output port. As an input port consistently uses and wins the token, it is gradually migrated closer to that input port for faster use. As other input ports also require the same output port, the token gradually migrates back towards the center of the switch where it can be equally shared.

Lastly, we saw in Chapter 3 that time-division-multiplexing (TDM) can be used to transmit bits of data at a very high data rate (approximately 25Gb/s). Phastlane could potentially exploit from this capability but must be redesigned since data transmission is currently accomplished using only wavelength-division-multiplexing (WDM). Thus, each bit of the transmitted packet is encoded using a separate wavelength, and the critical delay path between a source and destination is only dictated by the time it takes the front of the light composing the packet to reach the final receiver. One way to accomplish this change would be to pipeline data and control, such that as the control is setting up the path between a source and destination node, data propagates along the paths that were set up the previous cycle. A reason to use this revised approach would be to reduce the number of required waveguides in the data path, potentially leading to reductions in area, router latency and power.

7.3 Hybrid Network Architectures

In this section, we utilize the key takeaways from our optical device model in Chapter 3 to demonstrate a network architecture that exploits the bandwidth density of nanophotonics, to complement an on-chip electrical communication network. We design this system according to the following observations taken from Chapter 6:

- Power consumption of the optical components due to the laser and on-chip transmission and receipt is a primary design constraint. We utilize point-to-point (P2P) links to avoid optical insertion loss caused by additional rings coupled to the waveguide. Additionally, the laser is a shared resource, where sources arbitrate for use of a portion of the power in available wavelengths based on an *a priori* knowledge of how far the packet has to travel (since all links are P2P). The electrical power consumption of the optical devices is minimized through the use of point-to-point (P2P) links, which eliminates multiple receipts and transmits between a packet's source and destination.
- The key advantage of optical waveguides over electrical wires is improved bandwidth density. Based on ITRS data [27], for a global electrical wire and optical waveguide both transmitting data at a rate of 5Gb/s, nanophotonics enables approximately a 10X improvement in bandwidth density. Therefore, optical links should only be used in parts of a computing system that suffer from the lack of bandwidth. Excessive use of optics may lead to more complex network topologies and resulting inefficiencies in energy consumption.
- An important means of mitigating laser power requirements is the use of error detection/correction codes embedded within a transmitted packet. We saw in Chapter 3 that a detector power requirement of 10μW makes a huge difference in total network energy consumption over a value of 40μW, but

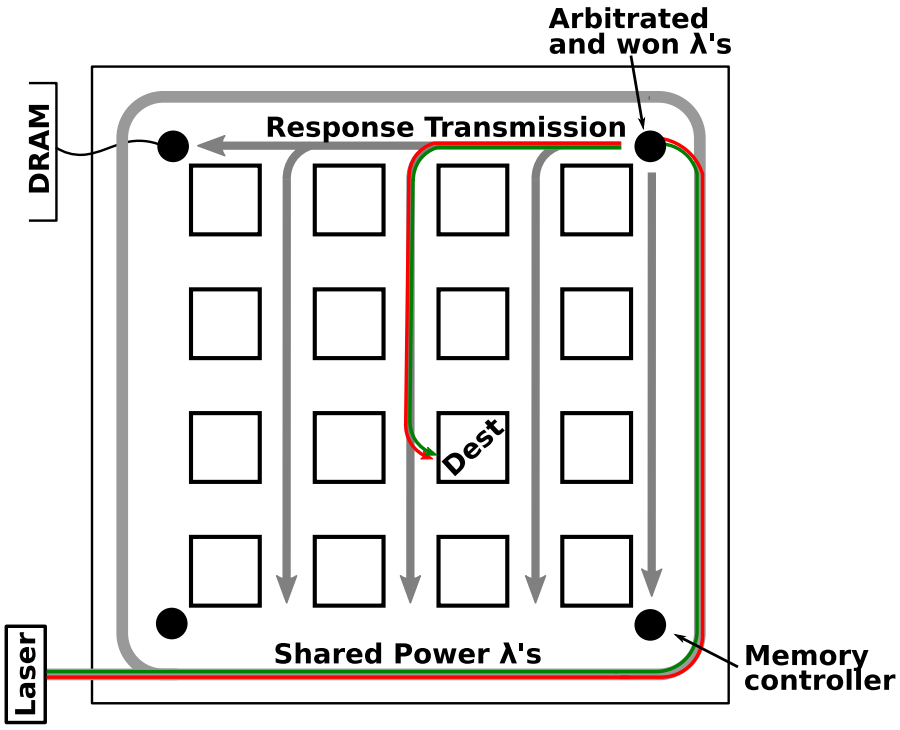


Figure 7.1: High level design of a hybrid electrical, optical interconnection network for future chip multiprocessors. Four memory controllers are situated at the corners of the network, which utilizes physically separate electrical, flattened butterfly topologies for shared memory requests and responses. Each node consists of multiple processors and cache memories and connects to the rest of the system using concentrated routers (i.e., multiple processors share the same input port). The optical interconnect is a P2P network that delivers responses from the memory controllers to different nodes. These P2P links utilize a shared laser resource using a smart arbitration scheme for obtaining power from the wavelengths on the surrounding distribution waveguide.

at the cost of increased bit-error-rate (BER). Another way to accomplish this would be to include additional ring resonators in the system channel spacings where temperature fluctuations impact device functionality. This could allow for the optical bits to enter through a neighboring ring resonator for modulation or demultiplexing without incurring an increase in BER.

Using these design guidelines generated from our optical device model, we show a high level blueprint of a hybrid network architecture that benefits from the use of nanophotonics and traditional electrical interconnect in Figure 7.1. The system

consists of multiple processors and memories grouped together to form concentrated inputs into an electrical, flattened butterfly router. Physically separate networks are tailored to the requests and responses of a shared memory system, allocating more bandwidth to the large response packets. As one example, four memory controllers are situated at the corners of the system, and utilize the optical interconnect to deliver responses from the DRAM to any node. These connections are implemented using P2P channels and flexible bandwidth transmission based on the power envelope of the laser.

Smart arbitration allows the memory controllers to quickly arbitrate for pieces of the available wavelengths of light (i.e., a portion of the light's power) using the token migration policy discussed earlier. Depending on the amount of power that is obtained for transmission, multiple response packets could be simultaneously sent from a memory controller to different nodes in the network. Communication bandwidth can be increased by raising the power envelope of the external laser source such that more shared energy is available to the transmitters for sending packets into the network.

As was done in the Phastlane architectures, communication encodes bits using different wavelengths of light in WDM, and can either be pipelined if the destination node is far enough that it can't be reached in a single network clock cycle, or bandwidth can be sacrificed by asserting the signal at the source for more than a single cycle. In the former case, more energy is required to do multiple transmits and receives prior to a packet arriving at its destination.

In the example shown in the diagram, the upper right memory controller successfully arbitrates for all of the power in two wavelengths on the waveguide surrounding the system that is supplied by the external laser source. It uses these two wavelengths to transmit a DRAM memory response to an inner node in the

network using the P2P channel link that joins both of them together.

Hybrid architectures are an interesting way to exploit the benefits of optics while trying to mitigate its weaknesses, namely potentially large power requirements if not designed carefully. Other ways to benefit an underlying electrical interconnect could include using optics to enable globally adaptive feedback for better packet routing. This feedback would be very beneficial to overcome the weaknesses of adaptive routing in electrical networks, which utilizes local feedback from neighboring switches to make routing decisions. Due to the lack of a global view, these routing algorithms suffer from accidentally routing a packet into a network hotspot.

Continued research in device modeling and design of nanophotonic networks for future chip multiprocessors will bring the power and performance of these architectures to a level unachievable with traditional electrical wires.

BIBLIOGRAPHY

[1] G. Agrawal. Fiber-Optic Communication Systems. Wiley-Interscience, 2002.

[2] S. Averine, Y. Chan, and Y. Lam. Geometry optimization of interdigitated Schottky-barrier metal-semiconductor-metal photodiode structures. *Journal of Solid-State Electronics*, 45(3), 2001.

[3] J. Balfour and W. Dally. Design Tradeoffs for Tiled CMP On-Chip Networks. In *International Conference on Supercomputing*, 2008.

[4] T. Battestilli and H. Perros. An Introduction To Optical Burst Switching. *IEEE Communications Magazine*, 41(8), 2003.

[5] S. Beamer, K. Asanovic, C. Batten, A. Joshi, and V. Stojanovic. Designing Multi-socket Systems Using Silicon Photonics. In *International Symposium on Super Computing*, 2009.

[6] S. Beamer, K. Asanovic, C. Batten, A. Joshi, and V. Stojanovic. Designing Multi-socket Systems Using Silicon Photonics. In *University of California at Berkeley Technical Report*, 2009.

[7] S. Beamer, C. Sun, Y. Kwon, A. Joshi, C. Batten, V. Stojanovic, and K. Asanovic. Re-Architecting DRAM Memory Systems with Monolithically Integrated Silicon Photonics. In *International Symposium on Computer Architecture*, 2010.

[8] A. Biberman, K. Preston, G. Hendry, N. Sherwood-Droz, J. Chan, J. Levy, and K. Bergman. Photonic Network-on-Chip Architectures Using Multilayer Deposited Silicon Materials for High-Performance Chip Multiprocessors. *ACM Journal on Emerging Technologies in Computing Systems*, 7(2), 2011.

[9] N. Binker, A. Davis, N. Jouppi, M. McLaren, N. Muralimanohar, R. Schreiber, and J. Ahn. The Role of Optics in Future High Radix Switch Design. In *International Symposium on Computer Architecture*, 2011.

[10] J. Bradley, P. Jessop, and A. Knights. Silicon waveguide-integrated optical power monitor with enhanced sensitivity at 1550 nm. *Applied Physics Letters*, 86(24), 2005.

[11] G. Chen, H. Chen, M. Haurylau, N. Nelson, P. Fauchet, E. Friedman, and D. H. Albonesi. Predictions of CMOS Compatible On-Chip Optical Interconnect. In *International Workshop on System Level Interconnect*, 2005.

[12] L. Chen, P. Dong, and M. Lipson. High performance germanium photodetectors integrated on submicron silicon waveguides by low temperature wafer bonding. *Optics Express*, 16(15), 2008.

[13] L. Chen and M. Lipson. Ultra-low capacitance and high speed germanium photodetectors on silicon. *Optics Express*, 17(10), 2009.

[14] Y. Chen, C. Qiao, and X. Yu. Optical Burst Switching: A New Area in Optical Networking Research. *IEEE Network Magazine*, 18(3), 2004.

[15] A. Chow, D. Hopkins, R. Drost, and R. Ho. Enabling technologies for multi-chip integration using Proximity Communication. In *International Symposium on VLSI Design, Automation and Test*, 2009.

[16] M. Cianchetti and D. H. Albonesi. A Low-Latency, High-Throughput On-Chip Optical Router Architecture for Future Chip Multiprocessors. *ACM Journal on Emerging Technologies in Computing Systems*, 7(2), 2011.

[17] M. Cianchetti, N. Sherwood-Droz, and C. Batten. Implementing System-in-Package with Nanophotonic Interconnect. In *Workshop on Interaction between Nanophotonic Devices and Systems (in conj. with MICRO-43)*, 2010.

[18] W. Dally. Express Cube: Improving the Performance of k-ary n-cube Interconnection Networks. *IEEE Transactions on Computers*, 40(9), 1991.

[19] W. Dally and B. Towles. Principles and Practices of Interconnection Networks. Morgan Kaufmann, 2007.

[20] R.K. Dokania and A.B. Apsel. Analysis of Challenges for On-Chip Optical Interconnects. In *Great Lakes Symposium on VLSI*, 2009.

[21] P. Dong, S. F. Preble, and M. Lipson. All-optical compact silicon comb switch. *Optics Express*, 15(15), 2007.

[22] P. Dong, W. Qian, H. Liang, R. Shafiiha, N-N. Feng, D. Feng, X. Zheng, A. V. Krishnamoorthy, and M. Asghari. Low power and compact reconfigurable multiplexing devices based on silicon microring resonators. *Optics Express*, 18(10), 2010.

[23] J. Fedeli, M Migette, L. Cioccio, L. Melhaoui, R. Orobtchouk, C. Seassal, P. Rojo-Romeo, F. Mandorlo, D. Morini, and L. Vivien. Incorporation of

a photonic layer at the metallization levels of a CMOS circuit. In *IEEE International Conference on Group IV Photonics*, 2006.

[24] M. Galles. Scalable Pipelined Interconnect for Distributed Endpoint Routing: The SGI SPIDER Chip. In *International Symposium on Hot Interconnects*, 1996.

[25] M. Geis, S. Spector, M. Grein, J. Yoon, D. Lennon, and T. Lyszczarz. Silicon waveguide infrared photodiodes with greater than 35 GHz bandwidth and phototransistors with 50 A/W response. *Optics Express*, 17(7), 2009.

[26] A. Gondarenko, J. Levy, and M. Lipson. High confinement micron-scale silicon nitride high Q ring resonator. *Optics Express*, 17(14), 2009.

[27] ITRS. International Technology Roadmap for Semiconductors (ITRS) 2009 edition, http://public.itrs.net. 2009.

[28] N. E. Jerger, L-S. Peh, and M. Lipasti. Virtual CIrcuit Tree Multicasting: A Case for On-Chip Hardware Multicast Support. In *International Symposium on Computer Architecture*, 2008.

[29] A. Joshi, C. Batten, Y. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic. Silicon-Photonic Clos Networks for Global On-Chip Communication. *Optics Letters*, 29(24), 2009.

[30] P. Kapur. Scaling Induced Performance Challenges/Limitations of On-Chip Metal Interconnects and Comparison with Optical Interconnects. In *Dissertation, Stanford University*, 2002.

[31] J. Kim. Low-Cost Router Microarchitecture for On-Chip Networks. In *International Symposium on Microarchitecture*, 2009.

[32] J. Kim, C. Nicopoulous, D. Park, R. Das, Y. Xie, N. Narayanan, M. Yousif, and C. Das. A Novel Dimensionally-Decomposed Router for On-Chip Communication in 3D Architecture. In *International Symposium on High Performance Computer Architecture*, 2007.

[33] N. Kirman, M. Kirman, R. Dokania, J. Martinez, A. Apsel, M. Watkins, and D. H. Albonesi. Leveraging Optical Technology in Future Bus-based Chip Multiprocessors. In *International Symposium on Microarchitecture*, 2006.

[34] N. Kirman and J. Martinez. An Efficient All-Optical On-Chip Interconnect

Based on Oblivious Routing. In *Architectural Support for Programming Languages and Operating Systems*, 2010.

[35] P. Koka, M. McCracken, H. Schwetman, X. Zheng, R. Ho, and A. Krishnamoorthy. Silicon-Photonic Network Architectures for Scalable, Power-Efficient Multi-Chip Systems. In *International Symposium on Computer Architecture*, 2010.

[36] P. Koonath, T. Indukuri, and B. Jalali. Monolithic 3-D Silicon Photonics. *Journal of Lightwave Technology*, 24(4), 2006.

[37] A. Krishnamoorthy and D. Miller. Scaling Optoelectronic-VLSI Circuits into the 21st Century: A Technology Roadmap. *IEEE Journal of Selected Topics in Quantum Electronics*, 2(1), 1996.

[38] A. Kumar, L-S. Peh, P. Kundu, and N. Jha. Express Virtual Channels: Towards the Ideal Interconnection Fabric. In *International Symposium on Computer Architecture*, 2007.

[39] R. Kumar, D. Tullsen, and N. Jouppi. Core Architecture Optimization for Heterogeneous Chip Multiprocessors. In *International Symposium on Parallel Architectures and Compilation Techniques*, 2006.

[40] B. Lee, B. Small, K. Bergman, Q. Xu, and M. Lipson. Transmission of high-data-rate optical signals through a micrometer-scale silicon ring resonator. *Optics Letters*, 31(18), 2006.

[41] H. Lin, J. Su, R. Cheng, and W. Wang. Novel Optical Single-Mode Asymmetric Y-Branches for Variable Power Splitting. *IEEE Journal of Quantum Electronics*, 35(7), 1999.

[42] M. Lipson. Compact Electro-Optic Modulators on a Silicon Chip. *IEEE Journal of Slected Topics in Quantum Electronics*, 12(6), 2006.

[43] H. L. R. Lira, S. Manipatruni, and M. Lipson. Broadband hitless silicon electro-optic switch for on-chip optical networks. *Optics Express*, 17(25), 2009.

[44] S. Manipatruni, K. Preston, L. Chen, and M. Lipson. Ultra-low voltage, ultra-small mode volume silicon microring modulator. *Optics Express*, 18(17), 2010.

[45] N. McKeown. The iSLIP Scheduling Algorithm for Input-Queued Switches. *ACM Transactions on Networking*, 7(2), 1999.

[46] C. Nitta, M. Farrens, and V. Akella. Addressing System-Level Trimming Issues in On-Chip Nanophotonic Networks. In *International Symposium on High Performance Computer Architecture*, 2011.

[47] J Orcutt, A. Khilo, C. Holzwarth, M. Popovic, H. Li, J. Sun, T. Bonifield, R. Hollingsworth, F. Kartner, H. Smith, V. Stokanovic, and R. Ram. Nanophotonic Integration in State-of-the-art CMOS foundries. *Optics Express*, 19(3), 2011.

[48] S. Pae, T. Su, J. Denton, and G. Neudeck. Multiple Layers of Silicon-on-Insulator Islands Fabrication by Selective Epitaxial Growth. *IEEE Electronic Device Letters*, 20(5), 1999.

[49] Y. Pan, Y. Demir, N. Hardavellas, J. Kim, and G. Memik. Exploring Benefits and Designs of Optically Connected Disintegrated Processor Architecture. In *Workshop on Interaction between Nanophotonic Devices and Systems (in conj. with MICRO-43)*, 2010.

[50] Y. Pan, J. Kim, and G. Memik. FlexiShare: Channel Sharing for an Energy-Efficient Nanophotonic Crossbar. In *International Symposium on High Performance Computer Architecture*, 2010.

[51] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary. Firefly: Illuminating Future Network-on-Chip with Nanophotonics. In *International Symposium on Computer Architecture*, 2009.

[52] D. Park, S. Eachempati, R. Das, A. Mishra, Y. Xie, N. Vijaykrishnan, and C. Das. MIRA: A Multi-Layered On-Chip Interconnect Router Architecture. In *International Symposium on High Performance Computer Architecture*, 2008.

[53] L. Peh and W. Dally. A Delay Model and Speculative Architecture for Pipelined Routers. In *International Symposium High Performance Computer Architecture*, 2001.

[54] C. Pollock and M. Lipson. Integrated Photonics. Kluwer Academic Publishing, 2003.

[55] K. Preston, S. Manipatruni, A. Gondarenko, C. Poitras, and M. Lipson. Deposited Silicon High-Speed Integrated Electro-Optic Modulator. *Optics Express*, 17(7), 2009.

[56] K. Preston, B. Schmidt, and M. Lipson. Polysilicon photonic resonators for large-scale 3D integration of optical networks. *Optics Express*, 15(25), 2007.

[57] K. Preston, N. Sherwood-Droz, J. Levy, H. Lira, and M. Lipson. Design rules for WDM optical interconnects using silicon microring resonators. In *submission*.

[58] K. Preston, M. Zhang, and M. Lipson. Waveguide-Integrated Photodiode in Deposited Silicon. *Optics Express*, 36(1), 2010.

[59] D. Rabus. Integrated Ring Resonators: The Compendium. Springer, 2007.

[60] J. Renau, B. Fraguela, J. Tuck, W. Liu, M. Prvulovic, L. Ceze, S. Sarangi, P. Sack, K. Strauss, and P. Montesinos. SESC Simulator. http://sesc.sourceforge.net, 2005.

[61] A. Sakat, T. Fukazawa, and T. Baba. Low Loss Ultra-Small Branches in a Silicon Photonic Wire Waveguide. *IECE Transactions on Electronics*, E85C(4), 2002.

[62] A. Shacham, K. Bergman, and L. Carloni. On the Design of a Photonic Network-on-Chip. In *International Symposium on Networks-on-Chip*, 2007.

[63] A. Udipi, N. Muralimanohar, R. Balsubramonian, A. David, and N. Jouppi. Combining Memory and a Controller with Photonics through 3D-Stacking to Enable Scalable and Energy-Efficient Systems. In *International Symposium on Computer Architecture*, 2011.

[64] D. Vantrease, N. Binkert, R. Schreiber, and M. Lipasti. Light Speed Arbitration and Flow Control for Nanophotonic Interconnects. In *International Symposium on Microarchitecture*, 2009.

[65] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. Jouppi, M. Fiorentino, A. David, N. Binkert, R. Beausoleil, and J. Ahn. Corona: System Implications of Emerging Nanophotonic Technology. In *International Symposium on Computer Architecture*, 2008.

[66] M. Waldow, T. Pltzing, M. Gottheil, M. Frst1, J. Bolten2, T. Wahlbrink2, and H. Kurz. 25ps all-optical switching in oxygen implanted silicon-on-insulator microring resonator. *Optics Express*, 16(11), 2008.

[67] M. Watts, W. Zortman, D. Trotter, G. Nielson, D. L. Luck, and R. W. Young.

Adiabatic Resonant Microrings (ARMs) with Directly Integrated Thermal Microphotonics. In *Conference on Lasers and Electroopics (CLEO)*, 2009.

[68] N. Wright, D. Thomson, K. Litvinenko, W. Headley, A. Smith, A. Knights, J. Deane, F. Gardes, G. Mashanovich, R. William, and G. Reed. Free carrier lifetime modification for silicon waveguide based devices. *Optics Express*, 16(24), 2008.

[69] D. Xu, A. Delage, R. McKinnon, M. Vachon, R. Ma, J. Lapointe, A Densmore, P. Cheben, S. Janz, and J. Schmid. Archimedean spiral cavity ring resonators in silicon as ultra-compact optical comb filters. *Optics Express*, 18(3), 2010.

[70] Q. Xu, S. Manipatruni, B. Schmidt, K. Shakya, and M. Lipson. 12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators. *Optics Express*, 15(2), 2007.

[71] Y. Xu, D. Du, B. Zhao, X. Yhou, Y. Zhang, and J. Yang. A Low-Radix and Low-Diameter 3D Interconnection Network Design. In *International Symposium on High Performance Computer Architecture*, 2009.

[72] I. Young, E. Mohammed, J. Liao, A. Kern, S. Palermo, B. Block, M. Reshotko, and P. Chang. Optical I/O Technology for Tera-Scale Computing. *IEEE Journal of Solid-State Circuits*, 45(1), 2010.

[73] X. Zheng, J. Lexau, Y. Luo, H. Thacker, T. Pinguet, A. Mekis, G. Li, J. Shi, P. Amberg, N. Pinckney, K. Raj, R. Ho, J. Cunningham, and A. Krishnamoorthy. Ultra-low-energy all-CMOS modulator integrated with driver. *Optics Express*, 18(3), 2010.