Architectural Exploration and Design Methodologies of Photonic Interconnection Networks

Jong Wu Chan

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Architectural Exploration and Design Methodologies

of Photonic Interconnection Networks

Jong Wu Chan

Photonic technology is becoming an increasingly attractive solution to the problems facing today's electronic chip-scale interconnection networks. Recent progress in silicon photonics research has enabled the demonstration of all the necessary optical building blocks for creating extremely high-bandwidth density and energy-efficient links for on- and off-chip communications. From the feasibility and architecture perspective however, photonics represents a dramatic paradigm shift from traditional electronic network designs due to fundamental differences in how electronics and photonics function and behave. As a result of these differences, new modeling and analysis methods must be employed in order to properly realize a functional photonic chip-scale interconnect design.

In this work, we present a methodology for characterizing and modeling fundamental photonic building blocks which can subsequently be combined to form full photonic network architectures. We also describe a set of tools which can be utilized to assess the physical-layer and system-level performance properties of a photonic network. The models and tools

are integrated in a novel open-source design and simulation environment called PhoenixSim.

Next, we leverage PhoenixSim for the study of chip-scale photonic networks. We examine several photonic networks through the synergistic study of both physical-layer metrics and system-level metrics. This holistic analysis method enables us to provide deeper insight into architecture scalability since it considers insertion loss, crosstalk, and power dissipation. In addition to these novel physical-layer metrics, traditional system-level metrics of bandwidth and latency are also obtained.

Lastly, we propose a novel routing architecture known as wavelength-selective spatial routing. This routing architecture is analogous to electronic virtual channels since it enables the transmission of multiple logical optical channels through a single physical plane (*i.e.* the waveguides). The available wavelength channels are partitioned into separate groups, and each group is routed independently in the network. Each partition is spectrally multiplexed, as opposed to temporally multiplexed in the electronic case. The wavelength-selective spatial routing technique benefits network designers by provider lower contention and increased path diversity.

Contents

List of Figures vi

List of Tables xix

List of Abbreviations xxiii

1 Introduction 1

1.1 Photonics for Chip-Scale Computing 1

1.2 Photonics and Memory 6

1.3 Dissertation Overview 14

2 Literature Review 17

2.1 Silicon Photonic Devices for Communications 17

2.1.1 Waveguides 20

2.1.2 Couplers 22

2.1.3 Ring Resonators 24

2.1.4 Detectors 28

2.2 Photonic Interconnection Networks 29

2.3 Computer-Aided Design Tools 33

3 Design Methodology and Simulator for Chip-Scale Photonic Networks 36

3.1 Motivation for Photonic Simulation 37

3.2 Methodology and Design Flow Overview 38

3.3 Photonic Device Library 43

3.3.1 Static Elements 46

3.3.1.1 Straight Waveguides 46

3.3.1.2 Waveguide Bends 47

3.3.1.3 Waveguide Crossings 49

3.3.1.4 Couplers 51

3.3.2 Ring-Resonator Elements 53

3.3.2.1 Filters 55

3.3.2.2 Broadband Switches 55

3.3.2.3 Modulators 56

3.3.2.4 Receivers (Photo-Detectors) 57

3.3.3 Mach-Zehnder Elements 58

3.4 Physical-Layer Performance Analysis Tools 59

3.4.1 Optical Power Budget 59

3.4.2 Data Integrity 63

3.4.3 Power Dissipation 67

3.5 Integration With Other Simulators 67

3.6 Simulation Infrastructure 69

4 Physical-Layer Analysis of Photonic Interconnection Networks 71

4.1 Photonic Circuit Switching Primer 71

4.2 Insertion Loss Analysis of 4×4 Switch Designs for Photonic Networks 75

4.2.1 Simulation Results 78

4.3 Physical-Layer Analysis of Photonic Circuit Switching 81

4.3.1 Insertion Loss Analysis 83

4.3.1.1 Device Improvement 86

4.3.1.2 Topology Exploration 88

4.3.2 Crosstalk Analysis 96

4.3.3 Power Analysis 101

4.4 Comparative Analysis of Photonic Spatial Routing and Wavelength Routing 109

4.4.1 Optical Power Budget Analysis . . . 113
4.4.2 Network Performance . . . 115
4.4.3 Data Integrity Analysis . . . 117
4.4.4 Power Dissipation Analysis . . . 122

5 Wavelength-Selective Spatial Routing 125

5.1 Concept . . . 126
5.2 Experimental Validation . . . 138
5.2.1 Experimental Setup . . . 140
5.2.2 Experimental Results . . . 141
5.3 Analytical Analysis . . . 145
5.3.1 Optical Power Budget and Insertion Loss Analysis . . . 145
5.3.2 Photonic Footprint . . . 149
5.3.3 Contention Probability . . . 151
5.4 Simulation Results and Analysis . . . 155
5.4.1 Synthetic Traffic . . . 156
5.4.2 Trace Simulations of Scientific Applications . . . 159

6 Concluding Remarks 168

6.1 Contributions . . . 169

6.2 Future Work 171

6.3 Recommendations 171

References **174**

List of Figures

1.1 Illustrations of the current typical interconnect architecture. 6

1.2 Memory performance of commercial micro-processors in recent years and projections. 8

1.3 Illustration of an optically-attached memory compute system with a processor attached to a single memory bank (composed of multiple DIMMs) via an optical bus. 10

1.4 Illustration of an optical-network-attached memory compute system with a single processor attached to a single memory bank via a photonic interconnection network. 12

1.5 Processor I/O pin scaling of commercial micro-processors in recent years (estimated, red square markers). Plot also shows ITRS projections for targeted pin count in next decade (blue diamond markers), and the required pin count of a processor package in order to achieve a performance of 1 B/FLOP (green triangle markers). 13

1.6 Three engineering challenges towards the realization and commercialization of chip-scale photonic interconnection networks: devices, tools, and architectures. 14

2.1 High-level block diagram of all optical communication links. 19

2.2 Ring resonator functional characteristics. (a) Off-resonance wavelength with a single waveguide. (b) On-resonance wavelength with a single waveguide. (c) On-resonance wavelength with secondary waveguide. (d) Transmission spectra of a long FSR ring resonator. (e) Transmission spectra of a short FSR ring resonator. The solid and dotted spectra in (d) and (e) show the influence of electro-optic control on the resonances of the ring while in an electrically unbiased and biased state. 26

2.3 Routing technique design space based on three arbitration domains: time, wavelength, and space. 30

3.1 The design flow of modeling a network in the PhoenixSim environment. . . . 39

3.2 A subset of the photonic devices in the *Interconnect Building Block Library*. 40

3.3 (a) Schematic of a design for a 4×4 non-blocking photonic switch. (b) A screenshot of how PhoenixSim composes the switch by instancing basic photonic devices. (c) Microscope image of a 4×4 non-blocking switch fabricated at the Cornell Nanofabrication Facility. 41

3.4 Parameters for characterizing a photonic device using the *Basic Element Model*. 45

3.5 PhoenixSim representation of the straight waveguide geometry. 48

3.6 PhoenixSim representation of a 90° bending waveguide geometry. 49

3.7 PhoenixSim representation of the waveguide crossing geometry. 50

3.8 PhoenixSim representation of an example coupler geometry, connecting a silicon waveguide to a tapered fiber. In this example, the width along the lateral direction of the interface is dominated by the fiber diameter. The length accounts for the tapering at the fiber tip (right) and the inverse taper of the waveguide in the silicon substrate (right). 52

3.9 Organization of building block element classes within PhoenixSim. 54

3.10 Propagation through a ring-resonator device depends on the signal wavelength and the resonant modes of the device. (a) Small rings with larger mode spacings (shown as periodic peaks) can be designed to interact with a single wavelength channel from a WDM signal (indicated by arrows). (b) Broadband switch have tightly spaced modes, enabling many WDM channels to couple into the device cohesively. (c) The path of propagation depends on whether the wavelength of the message is on- or off-resonance with the ring. 54

3.11 Schematic of the conversion process between the spatially-parallel electronic domain and wavelength-parallel optical domain. 57

3.12 The relationship of various parameters affecting the optical power budget. The difference in power of the total WDM signal (large arrow on the left) and the individual wavelength channels (five smaller arrows on the right) constrains the scalability of the system. 60

3.13 Calculation of insertion loss for a small network segment. 62

3.14 Sources of noise and crosstalk within a chip-scale photonic system. 65

3.15 Organization of the PhoenixSim environment. 68

3.16 Simulation server rack located in the Lightwave Research Laboratory at Columbia University. Servers used for simulation are the first, second, fourth, and fifth from the top. 69

4.1 The envisioned chip stack for photonic circuit switching. Three logical primary layers consisting of a processing layer (bottom), electronic control plane layer (middle), and photonic data plane layer (top). 73

4.2 Structure of a 4×4 node torus topology. The waveguides that make up the torus network are shown as thick lines, and the gateway access network for injecting packets to and ejecting packets from the network shown as thin lines. The blocks represent the following: gateway switch (G), injection switch (I), ejection switch (E), and a 4×4 non-blocking switch (X). 76

4.3 Layout of a tile in the torus network. This includes the type (A) version of the 4×4 non-blocking switch shown in Fig. 4.4. 77

4.4 Three implementations of the 4×4 non-blocking switch. 79

4.5 Insertion loss distribution for folded torus topologies of size (a) 4×4, (b) 6×6, and (c) 8×8. Each graph contains plots of three differing switch designs. Inset within each graph is a table of minimum, mean and maximum insertion losses observed for each case. 80

4.6 4×4 Non-blocking Torus with 8 access points. X labels mark 4×4 non-blocking switching points. G labels mark access points. S labels indicate combined injection-ejection switching points. 82

4.7 Maximum possible network-level insertion loss by component for varying sizes the Torus and Non-blocking Torus using the parameters listed in Table 4.2. Labeled values represent the peak cumulative insertion loss (in dB) for the network. 85

4.8 Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Torus topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved). 87

4.9 Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Non-Blocking Torus topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved). 88

4.10 Light propagation in 1×2 PSE. (a) Off-resonance propagation with crossing. (b) On-resonance propagation with crossing. (c) Off-resonance propagation without crossing. (d) On-resonance propagation without crossing. 89

4.11 Light propagation in 2×2 PSE. (a) Off-resonance propagation with crossing. (b) On-resonance propagation with crossing. (c) Off-resonance propagation without crossing. (d) On-resonance propagation without crossing. 90

4.12 4×4 TorusNX network with 16 access points. 92

4.13 Design for a photonic gateway with an integrated bidirectional crossing. . . 93

4.14 (a) The basic unit of the Square Root topology, a 2×2 quad. (b) A 4×4 Square Root. 94

4.15 Maximum possible network-level insertion loss by component for varying sizes of TorusNX and Square Root using the parameters listed in Table 4.2. Labeled values represent the peak cumulative insertion loss in dB. 95

4.16 Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the TorusNX topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved). 96

4.17 Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Square Root topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved). 97

4.18 Optical SNR performance for varying message sizes assuming saturated network load, measured at the photodetectors. The line at OSNR=16.9 dB is where a bit-error-rate of 10^{-12} can be achieved, assuming an ideal binary receiver circuit and orthogonal signaling. 100

4.19 Power-dissipation breakdown of an 8×8 Torus topology over varying message sizes. 104

4.20 Power-dissipation breakdown of an 8×8 Non-blocking Torus topology over varying message sizes. 105

4.21 Power-dissipation breakdown of an 8×8 TorusNX topology over varying message sizes. 106

4.22 Power-dissipation breakdown of an 8×8 Square Root topology over varying message sizes. 106

4.23 Total network bandwidth of each network at saturation. 107

4.24 Total network bandwidth of each network at saturation. 108

4.25 The Photonic Mesh topology. (a) A high-level representation of a 4×4 Photonic Mesh. Parallel lines indicate two unidirectional waveguides, which are paired together to form bidirectional links. Boxes represent higher-order photonic components, which are labeled 'X' for 4×4 nonblocking crossbar switch, 'I' for injection gateway, and 'E' for ejection gateway. Also shown are detail schematics of the (b) 4×4 nonblocking crossbar switch, (c) injection gateway, and (d) ejection gateway. 110

4.26 The Photonic Crossbar topology. (a) A high-level representation of a 2×4 Photonic Mesh, connecting 16 cores. Boxes represent gateways with a concentration of two processing cores. (b) A detail schematic of the Photonic Crossbar gateway, showing 49 bypass waveguides and 7 waveguides with modulator and receiver banks used to communicate to the other 7 gateways. 111

4.27 Insertion loss results for the Photonic Mesh and Photonic Crossbar of varying sizes. Labeled values that overlay the columns indicate the worst-case total network-level loss values. Columns illustrate the worst-case loss associated with the individually labeled loss component which does not necessarily occur in the network path with the worst total loss. 115

4.28 Wavelength channel allotment in the Photonic Mesh and Photonic Crossbar for varying network sizes and optical power budgets. 116

4.29 Bandwidth and latency performance on the Electronic Mesh, Photonic Crossbar, and Photonic Mesh for 1-kbit and 100-kbit message sizes. 118

4.30 Average total noise power accumulated by each optical message in the Photonic Mesh and Photonic Crossbar for saturated network load. Laser noise, thermal noise, and shot noises are negligible quantities and are not listed. 120

4.31 Network-level power dissipation breakdown of the Electronic Mesh, Photonic Mesh and Photonic Crossbar for transmission of 1-kbit and 100-kbit messages. Values overlaying each column indicate the energy efficiency of the network in units of pJ/bit. 124

5.1 (a) Spectral placement of two WDM partitions (each containing three wavelength channels), with respect to the spectrum of an electro-optic broadband ring switch. (b)–(e) Four possible routing configurations for a two-partition router. 128

5.2 The WSSR gateway architecture with concentrating processing cores. 133

5.3 Example timing diagram of the circuit-switching and WSSR allocation protocol. A path provisioning request is initially blocked, but is successful upon re-attempt. 134

5.4 Example timing diagram of the WSSR allocation protocol. If a single path provisioning request is attempted with multiple partitions, a path-setup request can partially block on a particular partition while be successful on another partition. 135

5.5 Schematic of the TorusNX photonic routers configured with two WDM partitions: (a) gateway switch and (b) 4×4 non-blocking photonic switch. . . 139

5.6 Scanning electron microscope image of a second-order electro-optic microring switch. Blue and red coloring is provided to label slab regions with dopants. 140

5.7 Diagram of the experimental setup for demonstration of WSSR concept. The three major optical link components are represented in this setup: generation (top left block), manipulation (top right block), and reception (bottom block). 142

5.8 The through port and drop port spectra of the electro-optic microring switch in the passive state. 143

5.9 BER curves for each of the six wavelength channels and the back-to-back case which bypasses the chip. 144

5.10 10-Gb/s output eye diagrams at both output ports from the device in the active state. 145

5.11 10-Gb/s output optical packets at both output ports from the device in the active state. 146

5.12 Insertion loss analysis of the TorusNX topology for varying levels of partitioning. Column plots correspond to worst-case insertion loss per component among all possible network paths (left-vertical axis). The line plot corresponds to greatest total network-level insertion loss path among all possible network paths (right-vertical axis). The lossiest path does not necessarily correspond with the sum of the worst-case losses per component. 148

5.13 Photonic router footprint for varying number rings (which corresponds to the number of WDM partitions enabled by the router). Legend indicates the ring diameter for the single ring case. 150

5.14 Destination blocking probability in a non-blocking network for varying number of interleaved channels. The limit of each blocking probability as $N \rightarrow \infty$ is superimposed on the right of the plot. 154

5.15 Average latency versus offered throughput for varying number of WDM partitions, message sizes, and number of wavelength channels. Electronic mesh performance is shown as a dotted line. 157

5.16 Traffic pattern plots for the four scientific applications being considered. Left axis represents source thread ID, bottom axis represents destination thread ID. White blocks represent no communication load while darker shades of gray represent increased traffic load between the associated source-destination pair. 161

5.17 Total simulation time required to complete each application trace. Columns indicate the average resulting time, and error bars indicate one standard deviation of the sampled data. 163

5.18 Network-level energy efficiency from each application trace. Columns indicate the average resulting energy efficiency, and error bars indicate one standard deviation of the sampled data. 165

List of Tables

4.1 Insertion Loss Parameters - 4×4 Non-blocking Switch Study 78

4.2 Insertion Loss Parameters - Photonic Circuit-Switching Analysis 84

4.3 Crosstalk and Noise Parameters - Photonic Circuit-Switching Analysis . . . 98

4.4 Energy Dissipation Parameters - Photonic Circuit-Switching Analysis 102

4.5 Insertion Loss Parameters - PhoenixSim Case Study 114

4.6 Crosstalk and Noise Parameters - PhoenixSim Case Study 119

4.7 Power Dissipation Parameters - PhoenixSim Case Study 123

5.1 Insertion Loss Parameters - Wavelength-Selective Spatial Routing Analysis . 147

5.2 Application Trace Characteristics - Wavelength-Selective Spatial Routing Analysis 160

5.3 Optical Device Power Parameters - Wavelength-Selective Spatial Routing Analysis 164

5.4 Application Trace Results Summary - Wavelength-Selective Spatial Routing Analysis 167

Acknowledgements

First, I give my foremost thanks to my advisor, Professor Keren Bergman. Her mentorship and advise have been instrumental in my doctoral work and my development as a researcher.

Also, I acknowledge Professor Luca Carloni, with whom I have extensively worked with during my doctorate program. His enthusiasm and insights have been an inspiration to me.

Next, I'd like to acknowledge and thank my many fellow members of the Lightwave Research Laboratory that I've had the opportunity of interacting with over the several years that I've been apart of the group.

The most senior student member, Ben Lee, for his early advise and a frequent discussion in photonics and networking. To Gilbert Hendry, who I've worked with most closely during my doctoral program, our frequent discussions, collaborations, and camaraderie. To Sasha Biberman, Caroline Lai, Noam Ophir,

Kishore Padmaraju, Lin Xu, and Wenjia Zhang who I've had the privilege of working with. To Howard Wang, Ajay Garg, Dan Brunina, Michael Wang, and Bala Bathula for our many fruitful discussions. And last but not least, I'd like to acknowledge the newest members of the lab, Atiyah Ahsan, Cathy Chen, Robert Hendry, Qi Li, Dawei Liu, Lee Zhu, Christine Chen, and Gouri Dongaonkar, I look forward to seeing you all take our lab into newer, greater, more exciting research directions.

Finally, I'd like to give my great appreciation to the committee members which I have not name for their time and effort in participating in my doctoral defense: Richard Osgood, Gil Zussman, and Madeleine Glick.

List of Abbreviations

BER bit error rate

BERT bit-error-rate tester

CAD computer-aided design

CMOS complimentary metal-oxide semiconductor

CMP chip multi-processor

CPU central processing unit

CW continuous wave

DCA digital communication analyzer

DDR3 third generation double data-rate

DIMM dual in-line memory module

DRAM dynamic random-access memory

DTG data timing generator

DWDM dense wavelength-division multiplexer

EDFA erbium-doped fiber amplifier

FDTD finite-difference time-domain

FLOP floating-point operation

FSR free spectral range

HPC high-performance computing

I/O input-output

IC integrated circuit

ITRS International Technology Roadmap for Semiconductors

LA limiting amplifier

MOD modulator

MPI message-passing interface

MZI Mach-Zehnder interferometry

NoC network-on-chip

OSA optical spectrum analyzer

OSNR optical signal-to-noise ratio

PhoenixSim Photonic and Electronic Network Integration and Execution Simulator

PPG pulse pattern generator

PSE photonic switching element

RIN relative intensity noise

SMF single-mode fiber

SNR signal-to-noise ratio

SOI silicon-on-insulator

TDM time-division multiplexing

VOA variable optical attenuator

WDM wavelength-division multiplexing

WSSR wavelength-selective spatial routing

Chapter 1

Introduction

1.1 Photonics for Chip-Scale Computing

The performance improvement of microprocessors over the past four decades has been made possible by several technological developments and innovations. The primary catalyst for this progress has been the steady shrinking of transistor technology. Up till now, transistor scaling has remarkably been able to sustain the trend that is predicted by Moore's Law, which simply state, that the number of transistors that can fit onto a single integrated circuit will double every two years. The increased transistor density has consequentially lead to the development of faster and more complex processor technology. Evidence of this progression strategy could be seen with the marketing tactics utilized by the various micro-

processor manufacturers during the 90s and first few years around the turn of the century, where the processor clock rates were advertised as indication of their computational power (e.g. 500 MHz, 1.5 GHz, etc.). Furthermore, this law has become a self-fulfilling prophesy since processor manufacturers have utilized Moore's Law in their own strategies for targeted future product development. Indeed, eventually this became a race of companies trying to get the most megahertz, and eventually gigahertz, in their chips.

However, the race towards the fastest clock rate experienced a fundamental power-dissipation wall around the turn of the century. Power dissipated by a transistor scales linearly with the clock rate, therefore pushing on a processor's clock rate would eventually push the component into a regime where the packaging can no longer tolerate the thermal energy being produced. Performance improvement could no longer be achieved with a faster clock rate, and an alternative performance gaining technique needed to be employed. This roadblock has more or less forced the transition away from single-core central processing unit (CPU) design to chip-multiprocessor (CMP) design. This shifts performance progress away from faster single-threaded execution to slower execution of more instructions in parallel threads. A multitude of commercial and research chips have been released with high core counts in recent years such as the Sony/Toshiba/IBM's 9-core Cell Microprocessor [1], the Tilera Tile 64-core chip [2], and Intel's 80-core Teraflop Research Chip [3].

The shift from single-threaded execution to multi-threaded execution leads to fundamental changes in the way that computer programs are ran. The naive method for leveraging multi-core computers is to simultaneously run multiple independent programs, each on separate cores. The processor supplies more net computational 'force', but individual applications are still hampered by a performance ceiling limited by the clock speed. The challenging engineering task is to utilize multiple cores to accelerate the execution of a single program. This requires fundamental changes in the way application codes are structured and changes in methodologies utilized for software development. Parallel computing has had a long history of progress and innovation in the high-performance computing (HPC) field where it continues to make headway, however it is also beginning to emerge as a relevant and important research topic for computation at the node and embedded-system level [4].

Another consequence of this shift has been the requirement of a hardware interconnect subsystem to link the many cores together, as well as connecting the multiple cores to off-chip components such as main memory and input-output (I/O) signals. Thus far, electronic-enabled interconnects have been able to satisfy the communication requirements of current computing systems. However, as these systems continue to scale in performance and size, it becomes increasingly difficult to maintain a network that can both accommodate the communication demands and stay within power-dissipation limits of the system

package [5, 6]. Electronically-enabled interconnects in CMPs already account for over 50% of the dynamic power dissipated in some high-performance chips [7]. The portion of dissipated power that comes from the interconnect is expected to continue to grow with time and will become the limiting factor in performance scaling again.

For Moore's Law to continue, it has become clear that the ultimate solution to the performance and power problems will need to be realized through a paradigm shift in the way that computer architectures are built and designed. The shift can either be brought about through fundamental changes to the way that computation logic is devised, or alternatively, and more dramatically, through a a migration in underlying technology. One such solution that could potentially alleviate many of the problems facing CMPs is the usage of optics, or more specifically *photonics*.

Optical communications has steadily penetrated smaller and smaller scales of application domains as electronics wanes in its capabilities to support the needed amounts of data transmission [8]. Optics has long been established within the realm of long-haul and metro communications for its superior distance-bandwidth product in comparison to previously used long-haul electronic communication systems. Photonics in HPC is emerging as the solitary solution for data transport beyond several meters due to rising bandwidth requirements on the order of terabits per second. At the chip scale, designers continue to

struggle with the scaling of purely electronic systems which has led to photonics becoming an obvious candidate for supplying the necessary performance requirements of future CMP systems.

Photonics technology has emerged as a promising chip-scale interconnect solution to the various challenges facing CMP scaling. Photonic signaling using wavelength division multiplexing (WDM) can enable orders of magnitude higher bandwidth density than electronics which is becoming increasingly constrained by the wire and pin densities that can be achieved [9]. The power dissipation of photonic signaling can be designed to be practically independent of distance and data rate. This allows for high-speed data to flow seamlessly between the on- and off-chip domains. All the necessary optical devices for creating chip-scale photonic interconnection networks have been demonstrated using complementary metal-oxide semiconductor (CMOS)-compatible fabrication techniques [10, 11, 12]. This compatibility allows them to be economically produced in existing fabrication lines. Moreover, CMOS compatibility allows these optical devices to be directly integrated with electronic digital circuits, providing a flexible and powerful means to create a high-performance interconnect fabric.

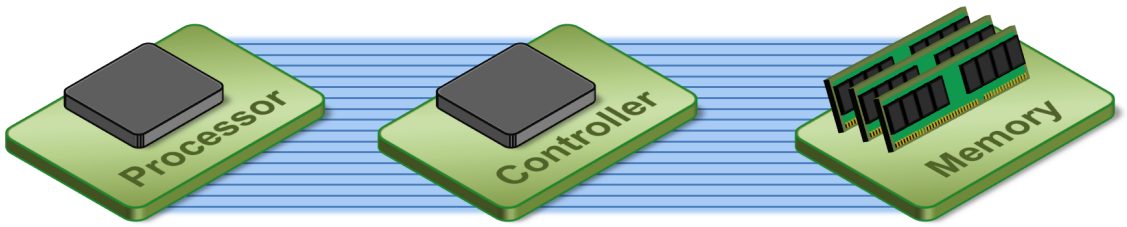

Figure 1.1: Illustrations of the current typical interconnect architecture.

1.2 Photonics and Memory

The distance independence and high datarate features of optics are remarkably well matched for a current major challenge facing computing systems: main memory interconnect architectures. The communication link between the CPU and main memory is a critical performance bottleneck for current fully electronic computing systems. Fig. 1.1 shows the current typical structure of a memory subsystem. The component which we might typically think of as the *processor* is the integrated circuit (IC) which handles the pipeline that performs arithmetic functions, logic functions, issues requests for data retrieval from memory, and issues requests for data transmission to memory. The memory *controller* translates processor memory requests into the logic signals necessary to access the requested memory elements. The *memory* itself in typically commercial systems is arranged as dual in-line memory modules (DIMMs) which are daughter cards mounted with several memory chips (typically dynamic random-access memory, also known as DRAM).

In regards to memory interactions, the processor only possesses the ability to directly retrieve from its on-chip cache. If a cache miss occurs, then the processor must communicate with main memory to interact with the addressed data. This action requires communication process to transpire off the chip, representing a domain boundary traversal and troublesome engineering challenge for system architects. Interactions occur as follows. First, the processor issues a request to the memory controller. Next, the memory controller must translate the request into the proper signaling to interact with the addressed memory cells. Lastly, main memory honors the commands from the memory controller and performs the requested action. In the event of a memory read operation, the data must be sent back through the controller and then to the processor. As is seen in the illustration (Fig. 1.1), the connection between the processor and the memory controller, and the memory controller and main memory requires a signaling bus that is composed of many wires in parallel. While many current commercial processors possess integrated memory controllers which combines the processor and memory controller into a single package, it does not preclude the need for a wide bus to interact with memory. Current third generation double data-rate (DDR3) DRAM requires 240 pins for proper electrical signaling. Memory systems have successfully been able to scale in capacity, however due to the need for complexing wiring have struggled in terms of bandwidth and latency improvement.

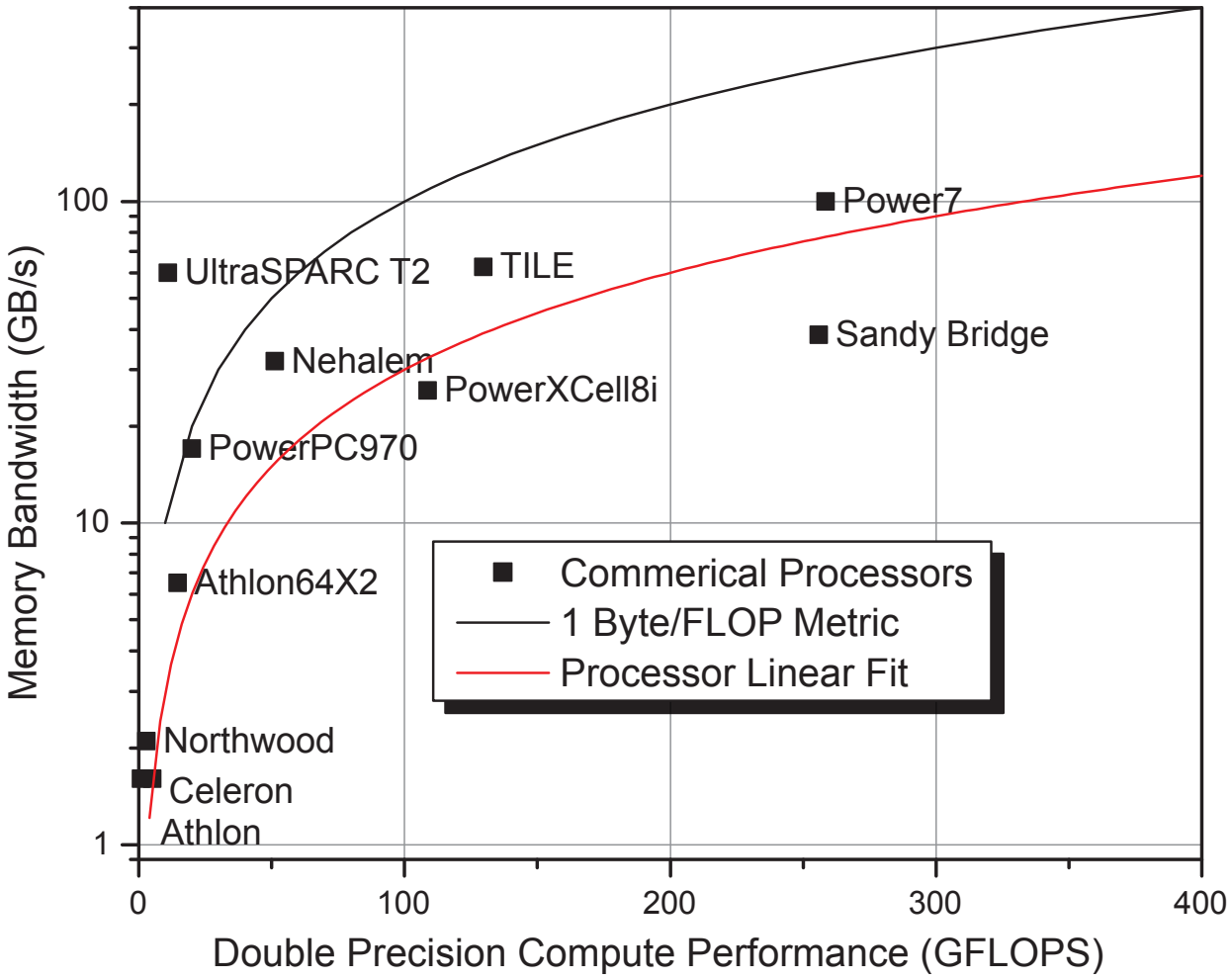

Figure 1.2: Memory performance of commercial micro-processors in recent years and projections.

A current metric that architects are increasingly specifying for a properly designed computer system is one byte I/O transferred per floating-point operations (FLOP). In other words, 1 B/FLOP specifies a balance between memory bandwidth and computation performance. Conventional computer architecture designs have been able to dodge this issue

by leveraged temporal and spatial locality of memory access. The presence of data locality enables the utilization of effective caching systems to hide access latencies. Fig. 1.2 shows the recent trend in computational performance versus the available memory bandwidth. The plot shows a trend in commercial processors that is half an order of magnitude below the 1 B/FLOP metric. However, new cluster computing application classes have risen in recent years which require a constant stream of data from main memory. This requirement for constant streams of large amounts of data effectively nullifies the performance that caching can bring.

The difficulties associated with scaling the latency and bandwidth performance of memory can be understood through a discussion on wire delays. The wire can be modeled as an RC circuit with time constant defined as $\tau = RC$, where R and C are the resistance and the capacitance of the wire. The time constant determines how quickly the wire will transfer a signal in response to an excitation by a driver. Qualitatively, a large τ corresponds to a large delay and lower frequency cutoff while a small τ corresponds to a small delay and higher frequency cutoff. It becomes apparent that shorter wires can produce a smaller τ from a lower resistance, however smaller and more densely packed wires will produce a larger τ.

Current memory sub-systems place the memory components (known as dual in-line memory modules, or DIMMs) near the CPU. The reason for the close proximity is to reduce

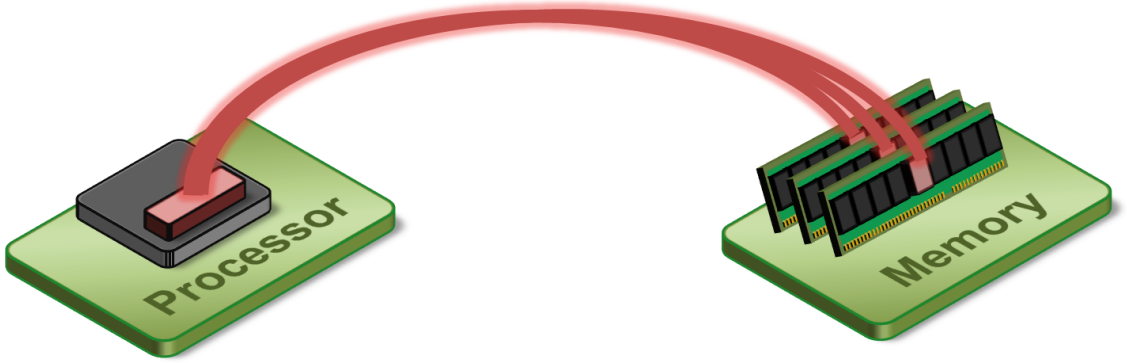

Figure 1.3: Illustration of an optically-attached memory compute system with a processor attached to a single memory bank (composed of multiple DIMMs) via an optical bus.

delay and increase frequency cutoff in the wire traces. In an attempt to optimize τ, system designers try to place the memory as close as possible to the CPU to reduce the lengths of wire. However, a design tradeoff arises from the need to meet capacity demands by including many DIMMs which conflicts with the available area when the traces are limited in length. Optics eliminates these RC-circuit properties and consequentially can eliminate distance-dependent performance. By enabling optical memory links, memory can be placed at farther distances while maintaining high data rates.

The advantages that optics can leverage naturally make it an ideal technological solution to the challenges facing memory for computing. The overarching vision for the application of photonics to memory systems in shown in Fig. 1.3. Research has shown that the enabling of optically-attached memory can provide significant performance advantages for typical

high-performance computational algorithms [13]. Fig. 1.3 shows a hypothetical optical link between a processor and memory DIMMs. The processor and DIMMs each have integrated photonic transceiver components. This close integration of electronic logic and photonic components is key to eliminating the need for board-level wire traces and consequently the delay characteristics of off-chip communications. IBM Research has experimentally demonstrated this tight integration of photonics with electronic drivers [14].

A potential extension of the optically-attached memory is the optical-network-attached memory which places an optical network between the processor and memory. This enables the possibility of utilizing multiple memory banks for a each processor chip. This is not practically feasible in the electrical domain due to the RC delay issues explained earlier. However, the bandwidth density offered by optics enables the creation of such a system. Fig. 1.4 illustrates this concept.

A final issue that the memory system of current computer systems face is in the available off-chip I/O bandwidth. While on-chip bus bandwidths can reach terabits-per-second scales, off-chip memory bandwidths are orders of magnitude less at 100's of gigabits-per-second. For example, the Tilera Tile processor is a 64-core chip arranged in an 8×8 mesh configuration with 2.56 Tb/s of bisection bandwidth and an off-chip memory bandwidth of 200 Gb/s [2]. This is primarily a limitation of the available pin count on chip packaging. Current state-of-

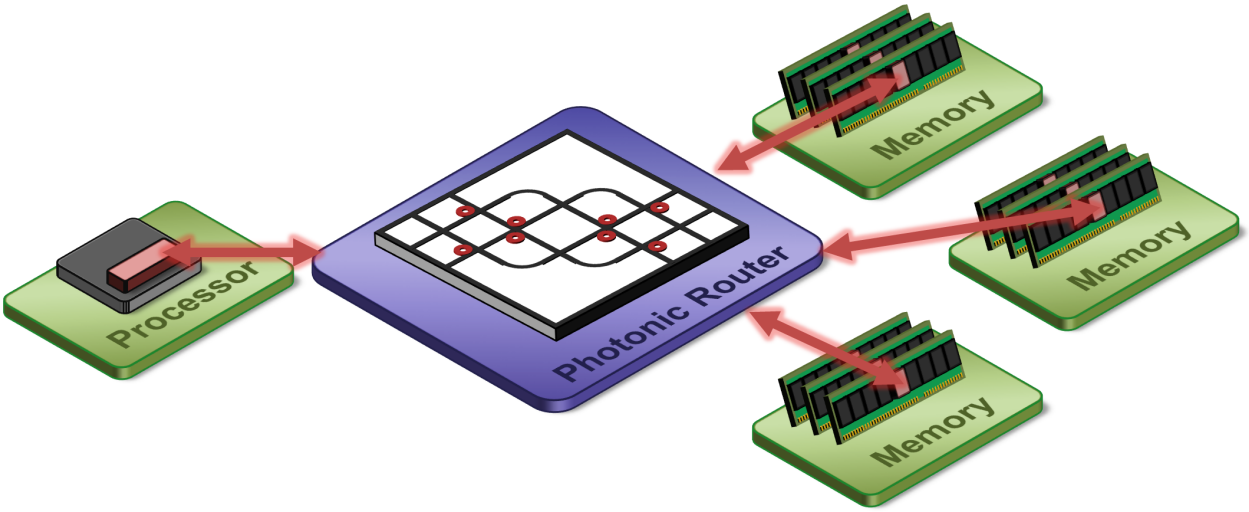

Figure 1.4: Illustration of an optical-network-attached memory compute system with a single processor attached to a single memory bank via a photonic interconnection network.

the-art chips contain a maximum of around 2000 pins, with a significant number of the pins being utilized for power delivery and grounding.

Fig. 1.5 plots rough estimates of the number of pins that are devoted to I/O for a sample set of processors (red squares) in the past decade. Fig. 1.5 also shows the targeted number of pins in the next decade which are values published by the International Technology Roadmap for Semiconductors (ITRS) in 2010 [15]. Lastly, the figure also shows the required pin count for each processor if it were to achieve the 1 B/FLOP metric, with estimated scaling of the clock frequency and improvements in processor performance. Notable is the fact that current commercial processors more or less closely flows the trend expected by the ITRS, however, this trend is almost an order of magnitude lower than the required pin count for 1 B/FLOP

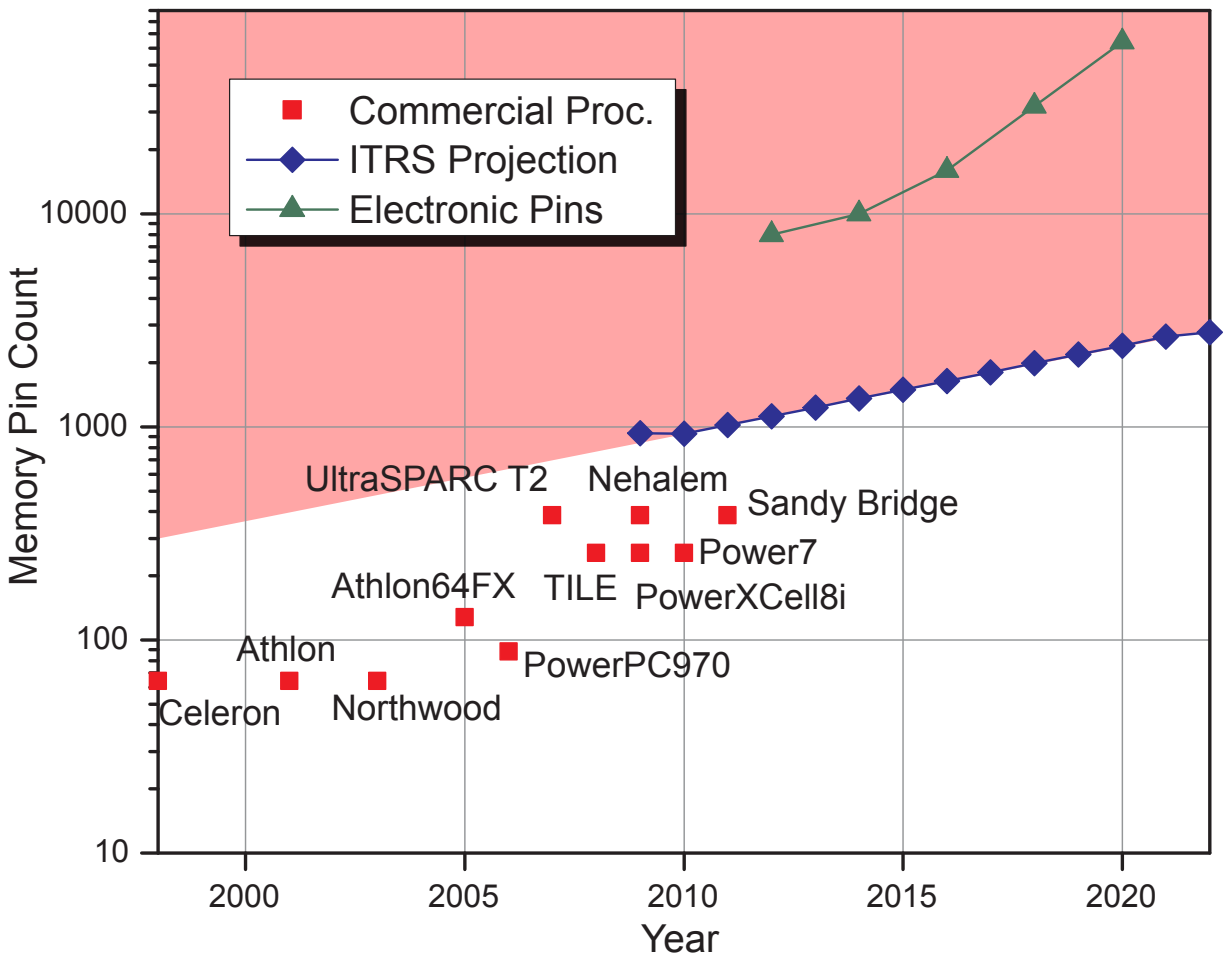

Figure 1.5: Processor I/O pin scaling of commercial micro-processors in recent years (estimated, red square markers). Plot also shows ITRS projections for targeted pin count in next decade (blue diamond markers), and the required pin count of a processor package in order to achieve a performance of 1 B/FLOP (green triangle markers).

performance. This electronic packaging problem is a potential area where photonics can bring about a paradigm shifting improvement to current architectures.

1.3 Dissertation Overview

The catalyst for the work presented here stems from a need to develop technological solutions for the performance scaling problems facing chip-scaling computing systems. The engineering challenges arise in three domains (Fig. 1.6), 1) devices, 2) tools, and 3) architectures. Within the device realm, physicists must design, create, and utilize new novel components for enabling the fundamental functions of an optical link. On the opposite side of the spectrum are the computer architects, who must create systems from the combination of fundamental devices to do useful things such as computation. Lastly, the domain that welds these two opposite but closely intertwined domains together are the tools, which must be designed and created in order to facilitate the collaborative and cohesive progress of the two areas. This work predominantly focuses the two later domains and emphasizes two particular topics: 1)

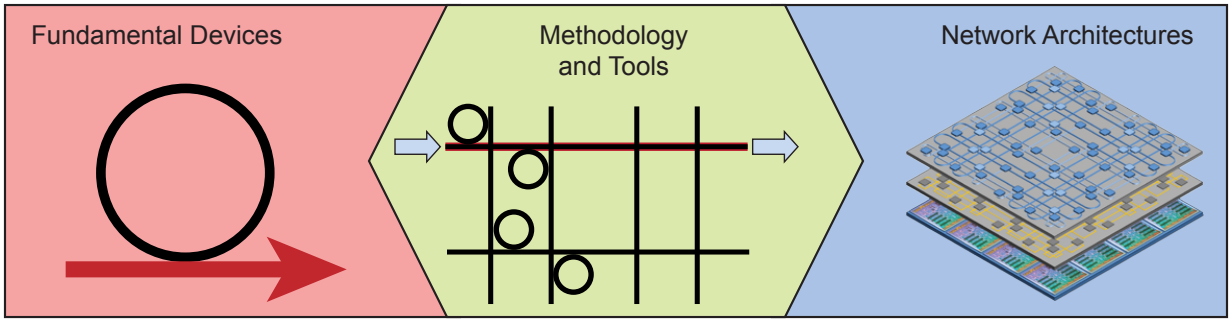

Figure 1.6: Three engineering challenges towards the realization and commercialization of chip-scale photonic interconnection networks: devices, tools, and architectures.

a methodology and associated tools for the designing and analyzing photonic interconnection networks, and 2) the utilization of this tool for designing new photonic architectures and studying their performance implications.

The organization of the remainder of this thesis is as follows:

Chapter 2 will provide a literature review of the current state of the art in photonics. The review will provide an introduction into three aspects of the field, 1) the *fundamental photonic devices* used to construct interconnection networks at the chip scale, 2) *photonic interconnection network architectures* used to link computation nodes, and 3) *software tools* used to design and understand photonic interconnection networks.

Chapter 3 discusses a photonic novel design methodology we developed for understanding photonic interconnection networks. We created PhoenixSim, a photonic network simulator, to implement this methodology. PhoenixSim is used for modeling and understanding photonic interconnection networks and is an integral in the research work presented in this thesis.

In Chapter 4, we review our research into the physical-layer performance of photonic interconnection networks. We highlight our study of physical-layer metrics (*e.g.* insertion loss, optical crosstalk) in the photonic networks which have no electronic equivalent. This fundamental understanding of photonic metrics enables photonic network architecture

designers to develop realistic architectures that obey the physical constraints of the photonic elements.

Chapter 5 discusses the development of wavelength-selective spatial routing which utilize new control techniques which have not been previously considered. This includes laboratory-based experimental validation of the concept, and simulated performance results.

Lastly, Chapter 6 provides concluding remarks. In particular, the major contributions of this work are summarized. In addition, areas of future work are also offered as crucial stepping stones in the realization and deployment of commercial chip-scale photonic systems.

Chapter 2

Literature Review

This chapter reviews several topics in the field of photonic interconnection networks that are relevant to the research presented in the later chapters of this dissertation. First, the fundamental set of devices that are used to construct photonic communication links are described. Then, a review of proposed photonic chip-scale architectures is presented. Lastly, an overview of research tools being developed for photonic interconnection networks is presented.

2.1 Silicon Photonic Devices for Communications

This section on devices reviews the components necessary in the creation of a optical communication channel. While there is a large variety of photonic devices being developed

and researched, the actual number of types of devices required for a optical network is fairly concise. The set of devices required for a photonic interconnection network are waveguides, couplers, modulators, detectors, switches, and filters. In this section, silicon photonic devices will be described, elucidated in terms of their usefulness towards chip-scale optical networks, and contrasted with electronic equivalents.

Fig. 2.1 illustrates the general structure of all optical communication channels, which comprises of the communicating nodes and the optical link itself. The optical link consists of three functional blocks: 1) generation, 2) transport and manipulation, and 3) reception. *Generation* occurs near a source node and involves the creation of a waveform in the optical domain for transporting useful information. *Transport and manipulation* is for controlling the movement of optical data so that the useful information can properly travel from source node to destination node. Lastly, *reception* enables the optical link to translate the useful information back into the electrical domain to be used by the computing resource at the destination node. In many cases, the transport and manipulation section of the link serves as the most important determiner of network performance since the generation and reception stages are generally very similar across all network architectures.These three components (generation, transport/manipulation, and reception) encompass everything needed for optical communications.

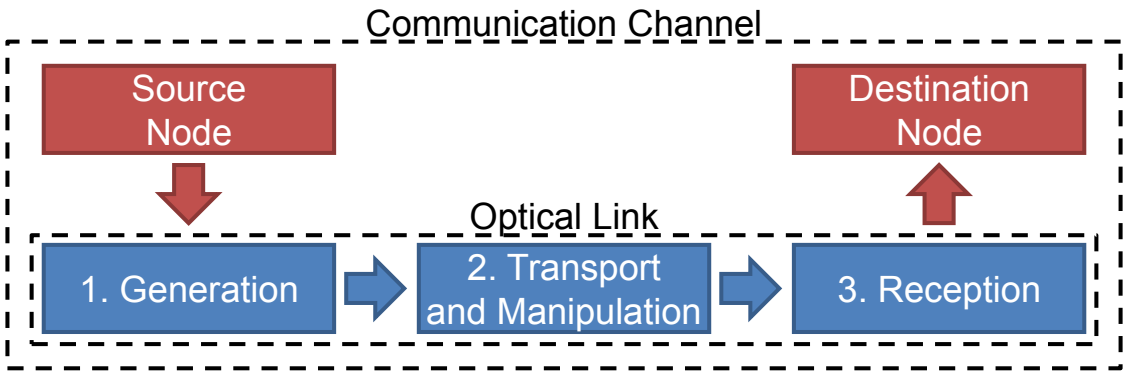

Figure 2.1: High-level block diagram of all optical communication links.

Although the high-level functional diagram of the canonical optical link (Fig. 2.1) is functionally similar to electronic interconnects, the two technology domains actually require fundamentally different design paradigms. In terms of generation and reception, the link requires a translation from electrons in the electrical domain to photons in the optical domain. Although all-optical computation and photonic logic (and in the same vein, quantum computing) are currently being proposed and researched [16, 17], the work presented here assumes the utilization of electronic-based logic and electronic-based compute nodes for the foreseeable future.

An additionally advantage of optical links is that they can uniquely leverage WDM, which is the ability to transmit multiple streams of data on a single physical waveguide by leveraging several optical carriers. Parallel data streams in an electronic network would require multiple spatially parallel wires. In contrast, a single waveguide (the photonic equivalent of a wire)

can transport several streams of optical data by utilizing a unique wavelength for each independent data stream. In non-high-power scenarios, each optical carrier with its unique wavelength will not interfere with any other signal flowing along the same waveguide. This WDM aspect of optical communications is a fundamental reason for why photonic networks-on-chip (NoCs) are attractive for providing high bandwidth links in future systems. The utilization of WDM in various architectures will be discussed in later chapters.

2.1.1 Waveguides

Waveguides can be regarded as the photonic equivalent of a wire. Waveguides are passive components which provide the physical links between all sources and destinations and enables connectivity between all photonic devices. Although they are simple devices, they are a fundamental elements in each of the three blocks of the canonical optical link. A photonic signal experiences insertion loss (*i.e.* attenuation) as it propagates through the waveguide due to free carrier absorption, light scattering at sidewall imperfections, and substrate leakage [18]. Most photonic devices are fabricated on silicon-on-insulator (SOI) technology, which limits waveguide placement to 2-D planar layouts. This means that the proper routing of data will require waveguide portions to be straight, as well as bend, and to cross. Each type of waveguide presents additional sources of loss which must be considered when determining

overall scalability of a photonic network.

A straight waveguide segment is simplest in design and will typically have the lowest loss when compared to any other waveguide variations (*i.e.* bends and crossings). A paper (Ref. [18]) published in 2005 surveyed waveguides published by research groups around the world, showing SOI-based waveguides with losses ranging from 2.4 dB/cm up to 110 dB/cm. More recently, silicon waveguides with cross sectional areas of approximately $500\,\text{nm} \times 250\,\text{nm}$ have been demonstrated improved losses of 1–2 dB/cm [19, 20]. Lower losses can be achieved using more exotic fabrication techniques such as with etchless silicon waveguides that have been shown to have losses of 0.3 dB/cm [21]. In terms of other current CMOS compatible materials besides crystalline silicon, silicon nitride is also a possible option due to its extremely low loss characteristics (losses of 0.1 dB/cm, [22]).

Just like an electronic wire, waveguides need to trace out paths with both straight sections and bending sections in order to divert signals correctly. Bends are necessary for the proper routing of optical paths, but also introduce an additional source of attenuation. This excess attenuation introduced by the bend is inversely related to the bending radius. Thus, a smaller bending radius produces a larger excess loss factor. The amount of loss has been experimentally measured to be 0.005dB per 90° with bending radius of 6.5 μm [19]. In general, bending radii longer than 5 μm produce negligible excess loss in comparison to the

propagation loss of the waveguide itself.

Waveguide crossings are inherently required in silicon-based on-chip topologies due to the 2-D planar nature of the technology platform. Crossings occur whenever two waveguides intersect and can exhibit both insertion loss and crosstalk which can have an impact on system scalability and performance. This is in distinct contrast with electronic interconnects, which do not allow arbitrary crossings of two wires since this would cause a short circuit. Since many topologies inevitably require a large number of waveguide crossings, it is important for these devices to exhibit both low insertion loss and low crosstalk. A $6\,\mu\text{m} \times 6\,\mu\text{m}$ double-etched crossing design has been fabricated and tested, and was shown to have fairly low insertion loss at 0.16 dB and high crosstalk suppression at about -40 dB [23].

2.1.2 Couplers

The cross-boundary interface that separates the on-chip and off-chip domain presents a distinct situation where photonics can break through performance bottlenecks that are typically experienced by electronics. The capacitive effects of metal wires cause limitations in both the distance and rate at which data can be transmitted electronically, consequently causing problems when trying to scale I/O performance which can potentially require long

wires that travel off-chip and across a board.

An optical coupler is a device that joins two waveguide segments together, including segments that might straddle an interface boundary. For example, this would occur at a chip I/O when transferring an optical signal from a on-chip silicon waveguide to an off-chip silica fiber. While traditional electronic system design is typically restrictive in cross-boundary data transmission (such as going from on-chip to off-chip), photonic interconnect-enabled systems possess the unique capability of crossing those boundaries with minimal impact on interconnect performance. Integrated optical I/O enables bandwidth transparency for off-chip signaling, and, unlike electrical I/O, the resulting signal integrity is much more resilient to propagation distance. Additionally, the power consumed in off-chip photonic communications is comparable to that of photonic on-chip message transfers, reducing the on- and off-chip bandwidth mismatch brought on by power limitations in current electronic systems.

A coupler of a signal on or off a chip can be accomplished through either a vertical coupler on the chip surface or a lateral coupler at the chip edge. The lateral coupling method transfers light at the chip edge which has demonstrated losses of less than 1 dB across a bandwidth of over 300 nm [24, 25, 26]. Vertical coupling utilizes Bragg gratings (periodic index changes) to allow the coupling of light into a waveguide. Vertical couplers produce losses below 1

dB which is comparable to lateral couplers, however suffer from much lower bandwidths at around 30 nm [27, 28]. Although vertical couplers exhibit smaller bandwidths, they can achieve much better alignment tolerances. Also vertical couplers can be placed anywhere on the surface of a chip which allows for flexibility in optical I/O placement and for chip-scale testing which is critical for achieving mass production of photonic chips.

2.1.3 Ring Resonators

The ring resonator is an instrumental device in the construction of photonic interconnection networks due to its versatility in implementing a variety of networking functions, compact footprint, and CMOS compatibility [29, 30, 31, 32, 33, 34, 35]. Ring resonators can be utilized to create modulators, filters, and switches.

Ring resonators are waveguides that form a closed loop which can be designed to manipulate the flow of light in a way that enables network functionality. Light interacts with the rings at specific periodically spaced wavelengths in the optical spectrum, called *resonant modes*. When a waveguide is properly positioned next to a ring resonator, lightwaves injected into waveguide that are rejected by the ring (termed *off resonance*) will be transmitted (Fig. 2.2a). Lightwaves that couple into the ring (termed *on resonance*) will not be transmitted and will be dissipated by the ring (Fig. 2.2b). A ring can also be

electrically manipulated to fluctuate between these two states to produce modulated light on the waveguide output. This modulated light provides the necessary mechanism to transfer an electrical signal into an optical signal for the generation block of the canonical optical link.

Alternatively, ring resonators can be designed to deliver on-resonance lightwaves onto a nearby secondary waveguide to enable filtering or switching functionality (Fig. 2.2c). Switching is critical components in the transport and manipulation stage of the optical link as it allows a system to divert and control the path the optical signal takes in the network. Filters can be utilized in both the generation and reception stages since it can be utilized for multiplexing or demultiplexing WDM signals. Filters can also be utilized in the transport/manipulation segment

The free spectral range (FSR) of the ring resonator is inversely proportional to the circumference of the loop, and quantifies the space between wavelengths that will couple and resonate with the ring. Modulators and filters which operate on a single wavelength will ideally have a small circumference and large FSR, thereby allowing only a single on-resonance wavelength and rejecting all other channels (Fig. 2.2d). When filtering or switching is required on more than a single wavelength, a smaller FSR is desirable, so that several wavelength channels can be concurrently on resonance with the ring (Fig. 2.2e). In this

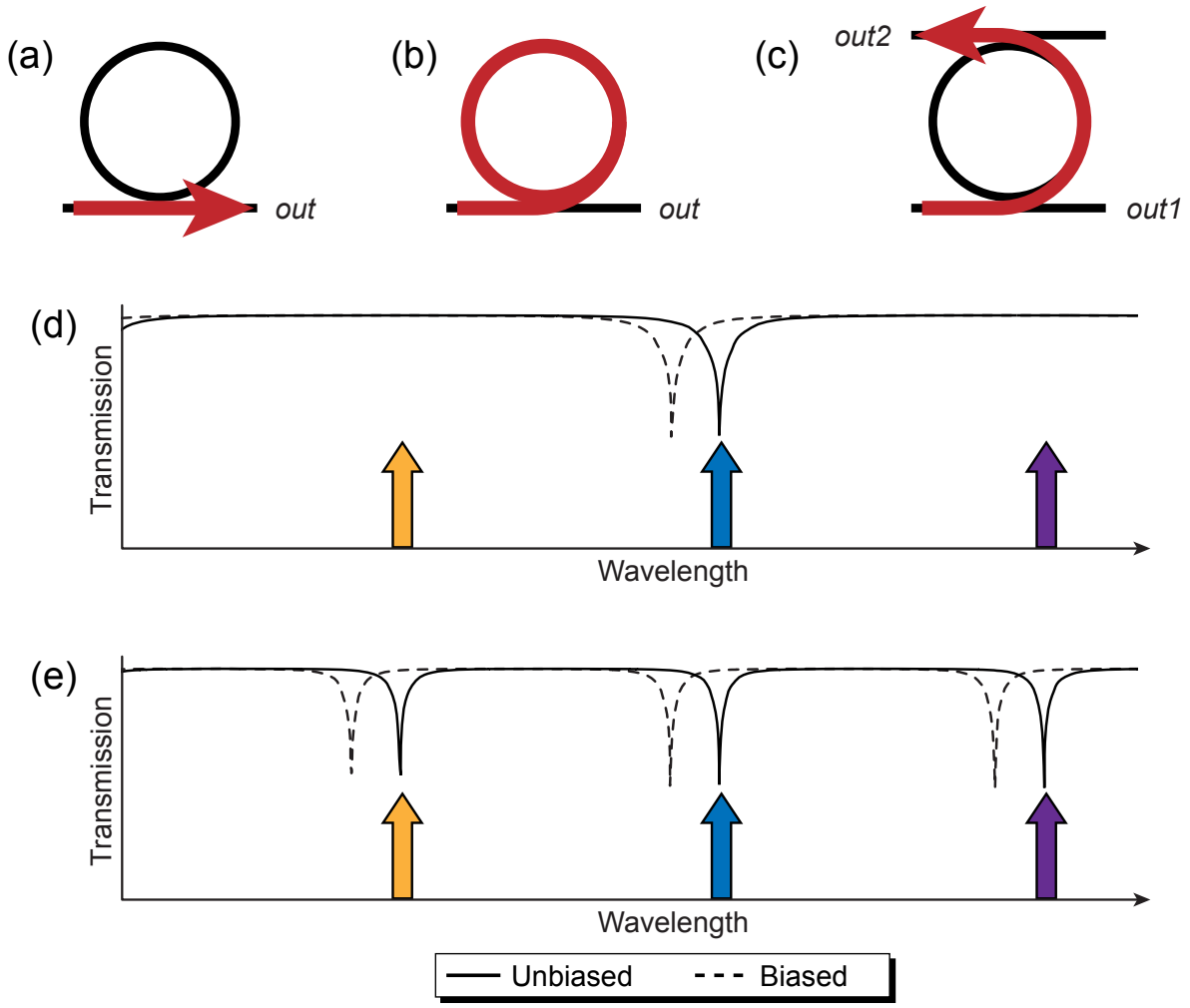

Figure 2.2: Ring resonator functional characteristics. (a) Off-resonance wavelength with a single waveguide. (b) On-resonance wavelength with a single waveguide. (c) On-resonance wavelength with secondary waveguide. (d) Transmission spectra of a long FSR ring resonator. (e) Transmission spectra of a short FSR ring resonator. The solid and dotted spectra in (d) and (e) show the influence of electro-optic control on the resonances of the ring while in an electrically unbiased and biased state.

manner, the single ring resonator can be used to simultaneously manipulate all channels in a WDM signal with no additional cost in complexity or footprint.

Moreover, Fig. 2.2d and Fig. 2.2e illustrate how electro-optic control through free carrier injection can be used to manipulate the resonant wavelengths of the ring for modulation or active switching [31, 32]. Electrical manipulation can be accomplished by creating a p-i-n structure on the ring with the waveguide acting as the intrinsic region. Electrically biasing the p-i-n structure will cause a shift in refractive index due to the free-carrier plasma dispersion effect in silicon [36]. This contrasts with thermal manipulation which uses the thermo-optic properties of the material for index changes [37]. The diverse range in functionality and the controllability offered by the ring resonator has been instrumental in the design of photonic interconnection networks.

The FSR imposes a limitation on the number of wavelength channels that can be utilized in a WDM system. Ring resonator modulators should affect only a single wavelength channel, therefore the periodic nature of the resonances imposes an inherent limitation on the number of channels possible. Preston, *et al.* showed that a WDM interconnect based on ring resonators will be able to maintain a satisfactorily low crosstalk level by having maximum wavelength channel count limitation of 62 when assuming 10-Gb/s datarates [38]. One cause of this limitation is that the minimum ring radius which can be fabricated also results

in a maximum FSR limit of 50 nm. This issue can be addressed by exploiting more exotic resonator designs which can significantly elongate the FSR such as interferometric combining [39], photonic bandgap structures [40], and the Vernier effect [41]. These techniques can be used to increase the FSR, and correspondingly increase the available spectrum and allowable number channels.

2.1.4 Detectors

Photo-detectors are used for converting optical messages back into the electrical domain and occurs in the reception end of the optical link. While the detection element itself is not a ring resonator, photo-detectors still require rings to properly filter individual wavelength channels from an entire WDM message. Each ring filter will only allow the light from a single wavelength channel to be incident on the photo-detector it precedes, thereby allowing the receiver to convert a single wavelength channel's worth of data back into the electrical domain. Similar to modulators, filtering should be accomplished without disturbing other adjacent wavelength channels by using as high an FSR as possible. Integrated high-speed germanium detectors have been demonstrated operating at speeds of 40 Gbps [42, 43].

2.2 Photonic Interconnection Networks

Advancements in silicon photonic device technology has brought about the development of all the functional components necessary in constructing chip-scale interconnection networks based on photonics. The set of fundamental devices include waveguides [19, 20], bends [19], crossings [23], filters [29], switches [30], modulators [31], and detectors [44]. Replicating the functionality of electronic interconnect designs with these photonic devices is possible, however the advantages that photonic technology offers will not be fully appreciated since their behavior and characteristics are fundamentally different from their electronic counterparts. Network architects have also proposed a variety of advanced novel interconnect designs in order to fully leverage the capabilities of photonics.

The various proposed photonic networks can be generally classified as leveraging a combination of three optical arbitration domains: time, wavelength, and space. Each arbitration domain provides a unique optical routing mechanism with different advantages and disadvantages. Fig. 2.3 provides a qualitative illustration of the design space that is afforded by these arbitration methodologies and the relative placement of the aforementioned routing techniques. A brief description of wavelength-selective spatial routing has been included in this literature review for completeness, nevertheless this architecture will be described in complete detail in the Chapter 5.

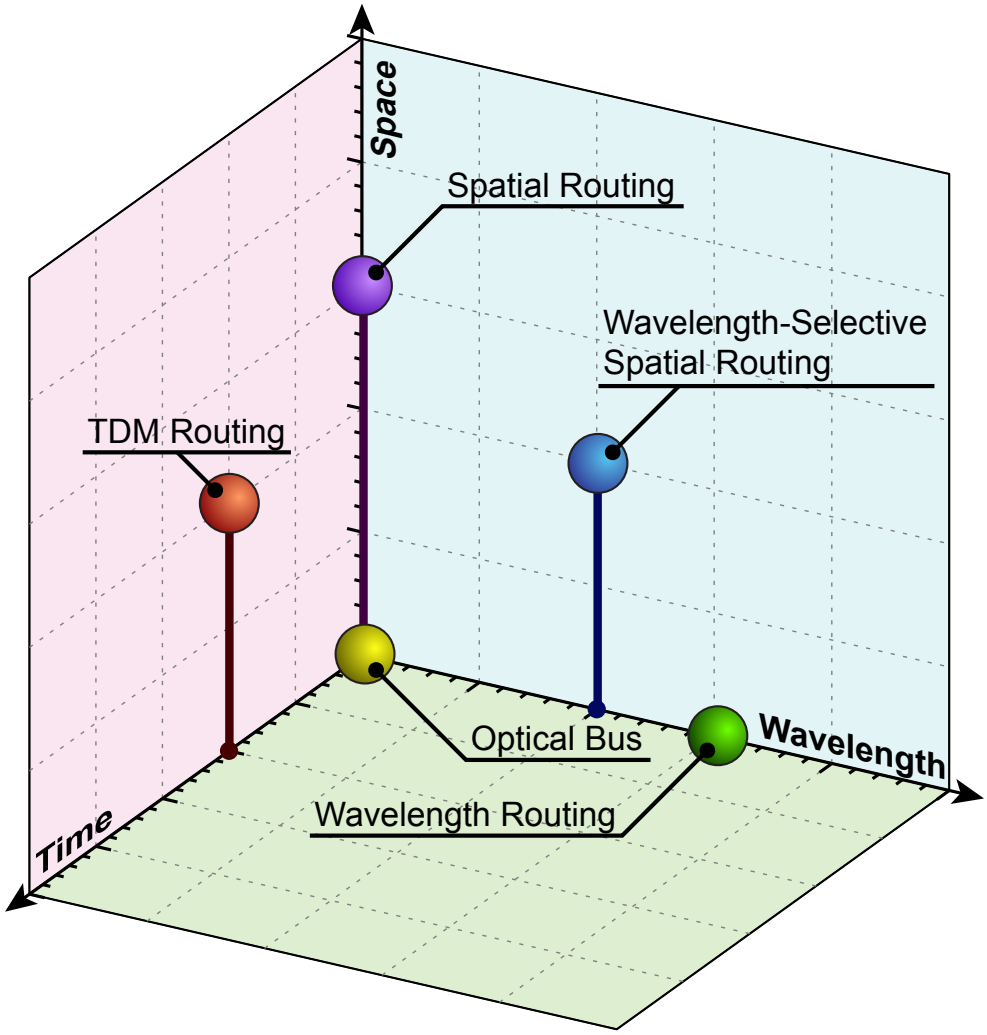

Figure 2.3: Routing technique design space based on three arbitration domains: time, wavelength, and space.

The simplest case is the optical bus which does not require any form of routing. Lack of an arbitration mechanism limits the network to a single source node and a single destination node. In what can be considered as the first step towards a full-scale photonic platform, Ophir

et al. demonstrated the operation of an *optical bus* (*i.e.* point-to-point link) operating at a data rate of 3 GHz [45].

Wavelength-routed topologies are constructed using ring-resonator-based filters which accordingly route lightwaves based on their wavelengths [46, 47, 48, 49, 50]. Any source node can address its intended destination through the selection of an appropriate transmission wavelength (i.e. source routing), which is then guided by the ring filters throughout the network. Transmission latencies can be designed to be extremely short when using wavelength routing since the propagation delay is simply the time of flight at the speed of light. However, spectral bandwidth is leveraged for routing purposes which could have otherwise be used to increase communication data rates.

Spatial routing uses electro-optic broadband ring resonators to guide a large set of parallel wavelength channels along an optical path [51, 52, 53]. The ring resonators act as comb switches to simultaneously control the path of all incident wavelength channels (Fig. 2.2e). Spatial routing requires a priori establishment of the entire optical path which is typically created using a circuit-switching style methodology. While spatial routing exhibits longer latencies than wavelength routing due to the overhead of the circuit-switching protocol, it is able to leverage the entirety of the available optical spectrum for data striping to create extremely high bandwidth links. A previous study showed that the circuit-switching

overhead can be amortized over large data messages, which is a characteristic in certain scientific applications typically executed on high-performance systems [52]. Section 4.1 provides a detailed description of the photonic circuit-switching design.

The usage of time-division multiplexing (TDM) has also been previously proposed as a technique for improving optical on-chip network performance [54]. *TDM routing* temporally divides the transmission medium into a continuous series of frames. Each frame is subdivided into several time slots which represents a different configuration of the entire optical network, and the set of all unique time slots completely connects all nodes in the network. The network is constructed using broadband ring switches, identical to the switches used for spatial routing, which are electro-optically reconfigured at the beginning of each time slot. A queued message at a source node will wait until an appropriate time slot arrives before it begins transmission, which contrasts with the spatial routing mechanism of immediately requesting the circuit allocation.

Wavelength-selective spatial routing (WSSR) is an extension of the spatial routing technique and is fully described in Chapter 5. This type of routing exploits the wavelength selectivity of ring resonators so that WDM signals can be partitioned into multiple logical network planes [55, 56]. Standard spatially routed networks requires a costly circuit switching protocol that can cause long delays when resources are over utilized. WSSR mitigates this

issue by distributing messages across several logical planes to reduce congestion.

Incarnations of some of the aforementioned TDM routing and the WSSR concept presented in this work were previously proposed and analyzed for multi-processor networks and wide area networks [57, 58]. The previous work showed that the use of WDM and TDM was effective for reducing network-level latency. With respect to TDM techniques, a comparison of link multiplexing and path multiplexing was conducted and showed that link multiplexing performed better in certain traffic configurations with a significant reduction in design complexity [57]. The alternative WDM technique was also described to have similar performance characteristics as the TDM case [58].

2.3 Computer-Aided Design Tools

As the interest for using photonic interconnects continues to grow, so does the need for computer-aided design (CAD) tools that can harness the potential of this new technology. In the realm of simulation, two levels exist which are of interest to photonic network designers: link-level and system-level. Simulation is an especially important predictive tool for gauging the performance of these photonic interconnect systems which are too complex for manufacturing in current fabrication technology. Beyond simulation, design tools will be needed to effectively and accurately design complex and efficient photonic interconnection

networks. Most conventional simulation and design tools are not ideally suited for capturing the physical and performance characteristics of chip-scale photonic interconnection devices and networks. Therefore the development of photonically-enabled tools is needed to fill the void.

As photonic interconnect topologies are becoming increasingly complex, layout tools and optimization techniques will be required for efficient and accurate design. Ding *et al.* have developed OIL (Optical Interconnect Library) a synthesis-like CAD tool for optimizing optical router designs in terms of insertion loss [59]. The methodology allows for constraint based optimization in terms of latency and insertion loss. Similarly, Minz *et al.* have devised a synthesis tool for timing-driven optimization of optical waveguide placement in an on-chip network [60]. VANDAL is a place-and-route tool for on-chip photonic architectures which uses a library of modeled and characterized components, and includes automation tools for rapid design and synthesis [61].

With link-level simulation, the primary concern is detailed physical modeling of all the end-to-end aspects of a photonic path to determine performance metrics such as signal integrity and link reliability. O'Connor *et al.* proposed a link-level simulation environment for heterogeneous photonic integrated circuits which leverages detailed synthesizable models of building-block components for the purpose of determining interconnect density, area,

link delay, and link power requirements [62]. Similarly, De Wilde *et al.* presented an approach for characterizing CMOS-to-CMOS links in terms of timing, error rates, and noise sensitivity [63]. The IBM optical link simulator was created to design and analyze telecom- and LAN-scale links through metrics such as failure rates, power penalties, and signal performance (*e.g.* eye diagrams) [64].

System-level simulation uses a higher-level of abstraction than link-level simulation and is primarily concerned with determining network performance metrics (e.g. bandwidth, application latency, and system power dissipation). Briere *et al.* have developed the ONoC SystemC model which focuses on the simulation of optical networks-on-chip using the SystemC framework and primarily addressing high-level system concerns including device timing and network-level power dissipation [65]. Their modeling is currently specific to topologies that leverage the *lambda router*, which routes optical traffic based on the wavelength of light that is being used by the source. OptiSim is a system-level simulator for modeling optical interconnects in board- and cluster-based computing [66].

Chapter 3

Design Methodology and Simulator for Chip-Scale Photonic Networks

In this chapter, we present a methodology for designing, modeling, and analyzing the performance of photonic interconnection networks [67, 68]. Furthermore, this chapter will highlight several techniques to synergistically study a photonic architecture's system-level properties through physical-layer analysis. We have developed the PhoenixSim environment which implements the described modeling and analysis aspects of our methodology and has been made publicly available [69]. PhoenixSim is implemented using OMNeT++, an open-source C++-based event-driven simulation environment [70, 71]. Our methodology and PhoenixSim represent a novel set of tools which system architects can use to see how

integrated photonics can potentially impact the performance of a particular computing system. PhoenixSim was initially planned and developed to specifically target silicon photonic architectures, however the simulation environment was designed to be generalized for photonic components of any material system (*e.g.* III-V materials) or scale (*e.g.* wide-area networks, telecom). The methodology and simulator are vital tools utilized for the architecture analysis in Chapter 4 and architecture design in Chapter 5.

PhoenixSim and the associated methodology are the successor to POINTS, which stands for Photonic On-Chip Interconnection Network Traffic Simulator [51]. POINTS was designed to look at the performance of photonic chip-scale architectures using synthetic-based traffic patterns.

3.1 Motivation for Photonic Simulation

While there are currently a large number of high-quality simulation environments available for studying networks architectures, none are capable of handling the unique architectures that are possible when considering chip-scale silicon photonics. Notable simulators of traditional systems include ns-2, NetSim, OPNET, and GloMoSim. These simulators typically support most standardized communication protocols (e.g. TCP) and are therefore well suited for traditional large scale networks. While these network simulators may support

optical components, the available library of elements are mostly relegated to commercially available equipment. For this reason, these simulators are unsuitable for the exotic network architectures that are available in the chip-scale domain.

An additional simulator characteristic that is needed is the ability to model the physical level of the optical components. Current fabrication technology is limited to simple device-level demonstrations (at most) for silicon photonics. A full scale network exceeds

PhoenixSim is primarily categorized as a system-level simulation environment that includes some aspects of link-level simulation. Our PhoenixSim environment closely resembles OptiSim (Ref. [66]) with respect to the use of a photonic building block library, and extractability of physical and system metrics. We differentiate our work from OptiSim through combination of our focus on chip-scale systems, support for spatial and temporal based photonic chip-scale architectures, and synergistic study of physical-layer and system-level performance metrics.

3.2 Methodology and Design Flow Overview

An overview of our design methodology is illustrated in Fig. 3.1. The sequence of design stages we employ for modeling photonic interconnection networks primarily consists of six design steps: 1) specification of the network building blocks, 2) specification of the target

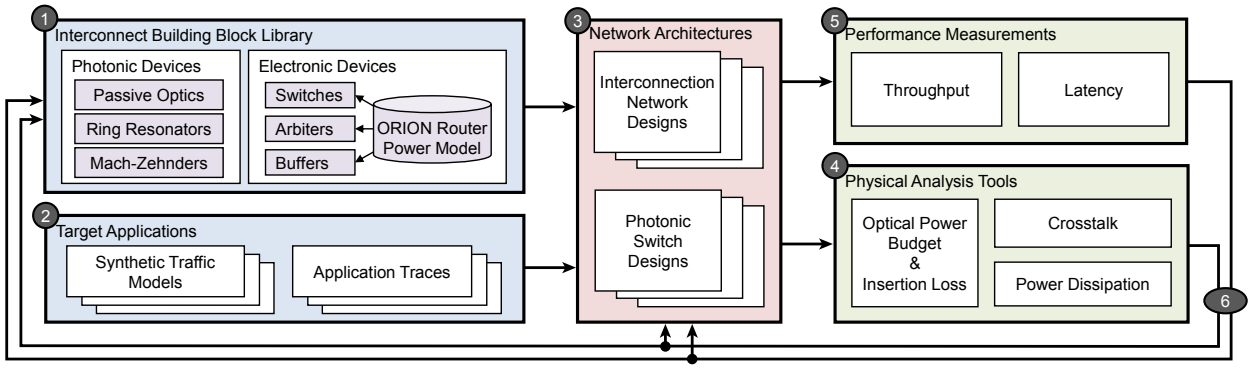

Figure 3.1: The design flow of modeling a network in the PhoenixSim environment.

application, 3) modeling of the network architecture, 4) system-level performance analysis, 5) physical-layer characterization, and 6) iterative refinement of parameters and design.

Step 1 (as labeled in Fig. 3.1) involves the specification of the fundamental network building blocks that will be used for creating the interconnection network. The collection of network building blocks is named the *Interconnect Building Block Library*. Within this library is a set of photonic devices that are characterized using the Basic Element Device Model (Fig. 3.2), described in further detail in Section 3.3. Users of this design methodology can choose to design a network based on the included library of devices, or extend the library themselves with other novel photonic building blocks.

The library for electronic building blocks consists of switch, arbitrator, and buffer blocks for creating standard pipelined routers. PhoenixSim leverages the ORION simulator [72] for deriving detailed values for electronic delay and energy dissipation. The electronic

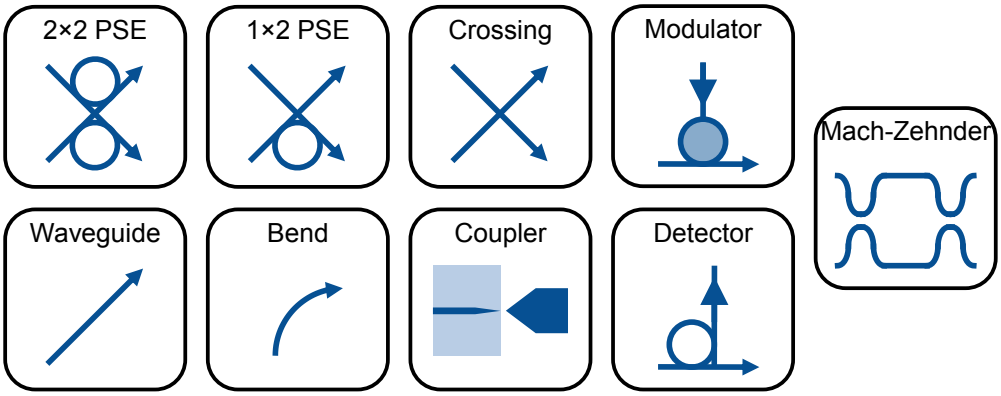

Figure 3.2: A subset of the photonic devices in the *Interconnect Building Block Library*.

router model is highly configurable and includes parameters for clock rate, buffer size, channel width, and number of virtual channels. In addition to the standard router design, the electronic router model also includes additional methods for interfacing with photonic devices. Electro-optic photonic devices can take an electronic input to influence its optical behavior and are essential components for enabling the active types of switching used in some proposed networks [51, 73].

Next, Step 2 consists of specifying the target application. PhoenixSim currently supports the use of both synthetically generated traffic patterns and communication traces, with eventual plans for integration with a cycle-accurate microarchitecture simulator. A variety of synthetic patterns have already been created within the environment (*e.g.* random, hotspot, nearest neighbor, and tornado) and is extensible to others. Communication traces can be generated by monitoring the network traffic during the execution of a real application and

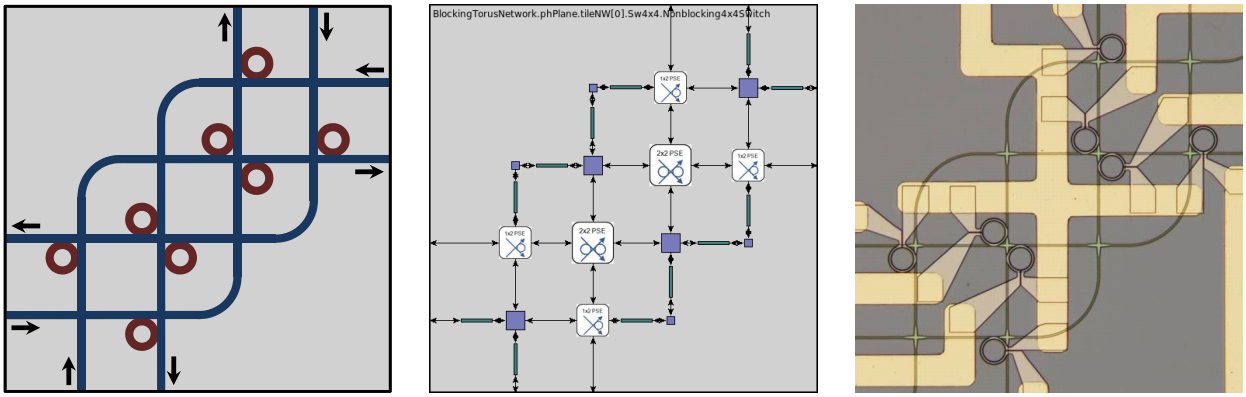

Figure 3.3: (a) Schematic of a design for a 4×4 non-blocking photonic switch. (b) A screenshot of how PhoenixSim composes the switch by instancing basic photonic devices. (c) Microscope image of a 4×4 non-blocking switch fabricated at the Cornell Nanofabrication Facility.

used as an input into PhoenixSim. Performance results gained by using communication traces are useful in assessing the application-specific performance gains of photonic networks [52].

The design and modeling of the network occurs in Step 3 of the design flow. The devices from the *Interconnect Building Block Library* can be combined to create higher-order networking components and entire interconnection network topologies. By accounting for the target applications, a network architect can optimize the topology design to target specific requirements such as message size, latency, and/or throughput. For instance, Fig. 3.3 illustrates how a 4×4 non-blocking switch can be derived within PhoenixSim by connecting various devices from the *Building Block Library*. Fig. 3.3a illustrates the schematic representation of the 4×4 non-blocking switch, while Fig. 3.3b depicts the

PhoenixSim representation as composed within the environment. In Fig. 3.3c, an image of an actual 4×4 non-blocking switch that was fabricated at the Cornell Nanofabrication Facility is shown [74].

Step 4 involves the characterization of the network architecture at the physical-layer, which involves metrics such as the optical power budget, crosstalk, and power dissipation. The overall physical-layer performance of a derived photonic component or topology can be determined from the aggregate performance of the individual photonic devices. Although this is not as rigorous as a true link-level simulator, this hierarchical building process enables an accurate first-order physical characterization of an entire network through the characterization of a small number of foundational components.

Step 5 measures the system-level performance characteristics of the network architecture in terms of data throughput and latency. Many of the physical properties that are identified in Step 4 have an impact on network functionality and scalability and play a crucial role in determining overall system performance.

Finally, Step 6 forms the basis for an iterative process, where the performance results and analysis of the modeled network can be used to refine the topology design and device parameters to further optimize the overall performance. Previous work has demonstrated the effectiveness of this iterative step. The initial physical-layer characterizations showed

the dramatic impact that waveguide crossing loss had on performance and a subsequent analysis of a system with improved crossings resulted in a dramatic improvement in overall performance [53].

3.3 Photonic Device Library

Our method for modeling photonic devices is designed to enable the assessment of the physical-layer performance at a first-order approximation while concurrently allowing for system-level analysis with a reasonable computational requirement. Many simulation packages use techniques such as finite-difference time-domain (FDTD) to accurately model an electromagnetic field according to Maxwell's equations. FDTD analysis, however, is usually limited to a single or small set of devices since it is computationally intensive and can have a large memory requirement. We use a more efficient level of abstraction by establishing a set of characteristic device parameters that are key to measuring the physical and system metrics which are important to our understanding of photonic interconnection networks. This simplified model enables PhoenixSim simulations to run on conventional computers in a period of minutes or hours. The device characteristics can be determined experimentally, through simulation, or projected. This set of modeled devices composes the *Photonic Device Library*. While the descriptions included in this paper mostly highlight silicon ring-based

topologies, the modeling methodology can easily be used to describe devices based on other technology domains such as Mach-Zehnders (also described in this section), photonic crystals, and MEMS.

The parameters used to describe basic photonic devices, called *Basic Elements*, are shown in Fig. 3.4. We refer to optical inputs and outputs as ports. Each port is physically bi-directional, therefore ports from which an optical signal can ingress into can also be used to egress from, and vice versa. Certain network topologies may still require uni-directional operation of the ports to facilitate simplicity or satisfy some other design requirement. Nonetheless, the bi-directional nature of each port is still represented for accuracy. The ports of the device are enumerated $0 \ldots N-1$ where N is the number of ports of a photonic device. The later figures in this section which show device geometry will have ports (represented by black dots) labeled with their assigned value. N also determines the size of additional parameter matrices used in defining the photonic device behavior and characteristics.

We use a logical routing table to determine the path a message takes through the device. Fig. 3.4 shows how the routing table can be represented as a length-N vector, where the index represents the ingression port of an optical signal and the value at the index represents the egression port.

Additionally, we use two tables to represent the latency and the optical insertion loss

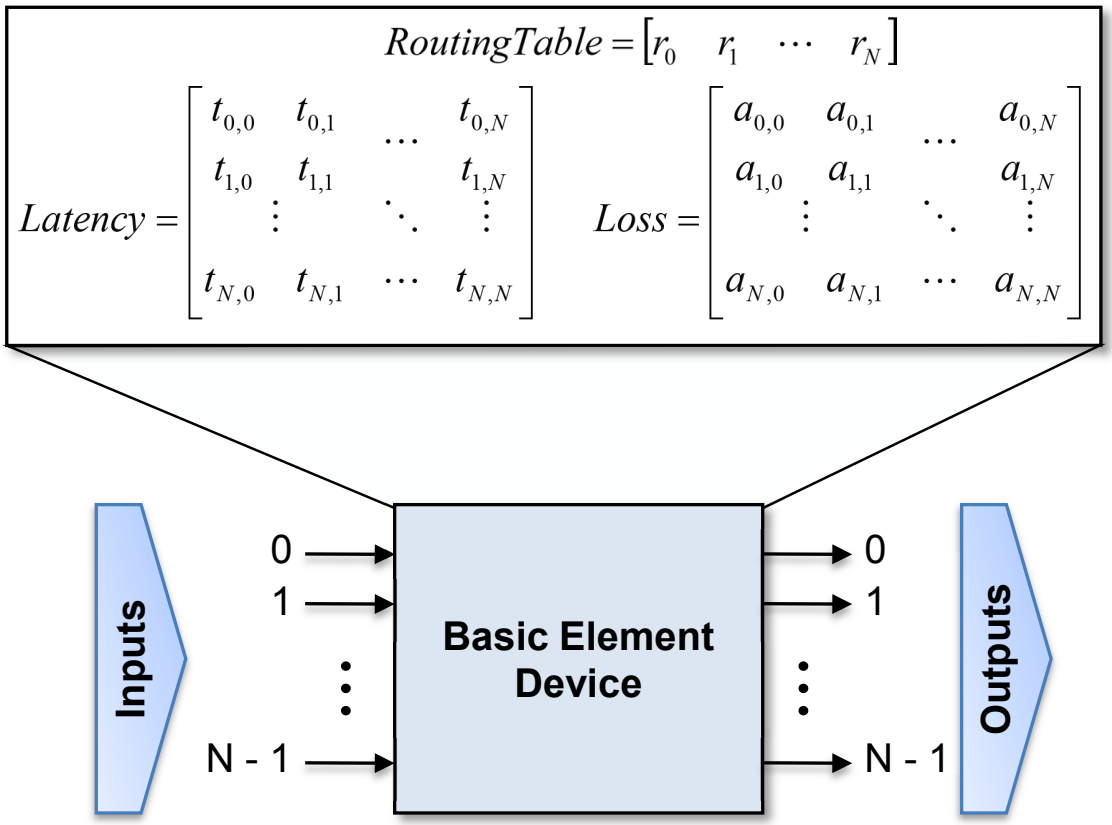

Figure 3.4: Parameters for characterizing a photonic device using the *Basic Element Model*.

properties of the device. Each property is represented as a $N \times N$ matrix where the row corresponds to the port through which the optical signal ingresses from (input) and the column represents the port from which the optical signal egresses from (output). Each entry in a matrix corresponds to the value used for the particular input/output combination. The latency for a particular input-output port combination is measured as the time between when optical signal enters the input port and when the same optical signal exits the output port. The insertion loss is a measure of the optical power attenuation an optical signal receives

when traveling through a device and is useful in characterizing network-level insertion loss and crosstalk.

3.3.1 Static Elements

The *Basic Element Model* is most suitable for describing static optical devices that have characteristics that do not change at runtime. The current library of devices focus on 2-D planar devices that are capable of being fabricated in a CMOS-compatible process. These static devices include waveguides, waveguide bends, waveguide crossings, and couplers.

3.3.1.1 Straight Waveguides

Straight waveguides can be characterized by its segment length and insertion loss. Propagation loss is affected by a variety of parameters including waveguide dimensions, fabrication technique, and material properties. Waveguides are modeled as 2-port devices with parameters for length, group velocity per unit length, and insertion loss per unit length.

A waveguide's routing table is $[1, 0]$; which indicates that an optical signal ingressing on either end will egress on the opposite side. For a waveguide of length L_{wg} and propagation delay t_{wg}, the latency matrix will be:

$$Latency_{wg} = \begin{bmatrix} - & L_{wg}t_{wg} \\ L_{wg}t_{wg} & - \end{bmatrix}$$

Note that the elements along the diagonal represent the latency of a reflection. Since reflections are nonexistent in waveguides, the elements of the matrix that represent the latency of the reflection are marked as don't-care values (–). Similarly, the same waveguide with propagation loss of α_{wg} will have a insertion loss matrix of:

$$Loss_{wg} = \begin{bmatrix} \infty & L_{wg}\alpha_{wg} \\ L_{wg}\alpha_{wg} & \infty \end{bmatrix}$$

While reflections do not occur in the waveguide, it is useful to assign infinite (∞) insertion loss to the reflection path for crosstalk calculation purposes.

The straight waveguide geometry is shown in Fig. 3.5. Although waveguide pitches are less than a micron in pitch, a large buffer needs to be enforced around the waveguide to prevent unintended evanescent coupling and crosstalk. PhoenixSim assumes a buffer of 2.5 μm for all devices, therefore the waveguide element utilizes an effective pitch of around 5 μm.

3.3.1.2 Waveguide Bends

Waveguide bends contribute additional insertion loss to the waveguide's existing propagation loss, which we refer to as bending loss. Bends are modeled as 2-port devices and take parameters for loss per degree and angle of the bend.

Similar to straight waveguides, waveguide bends also possess a routing table of $[1, 0]$. The

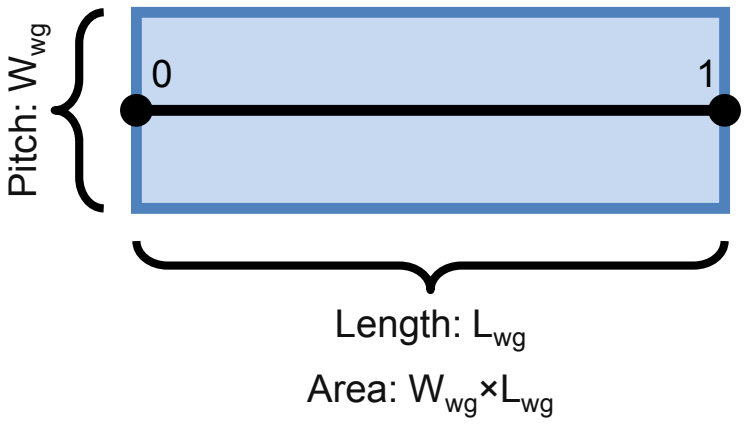

Figure 3.5: PhoenixSim representation of the straight waveguide geometry.

radius of the waveguide bend, L_{bend}, must be specified. For the purposes of simulation and layout, we assume a bending radius of 2.5 μm and parameterize the insertion loss according to the angle of the arc, θ_{bend}. The latency matrix for waveguide bends is:

$$Latency_{bend} = \begin{bmatrix} - & \theta_{bend} L_{bend} t_{wg} \\ \theta_{bend} L_{bend} t_{wg} & - \end{bmatrix}$$

We introduce an additional loss parameter for the bending loss, α_{bend}, which defines the total loss per 90° bend. Note that in the PhoenixSim definition, the bending loss parameter includes both the propagation loss of the waveguide as well as the excess loss due to the bend. This produces a loss matrix for bends as follows:

$$Loss_{bend} = \begin{bmatrix} \infty & 2\alpha_{bend}\theta_{bend}/\pi \\ 2\alpha_{bend}\theta_{bend}/\pi & \infty \end{bmatrix}$$

The bending waveguide element geometry is illustrated in Fig. 3.6. In the same fashion

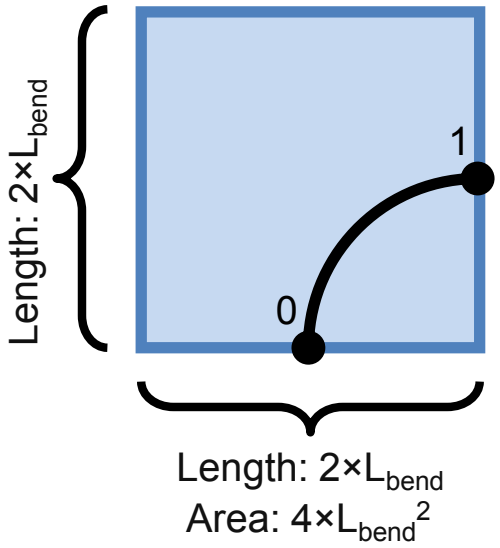

Figure 3.6: PhoenixSim representation of a 90° bending waveguide geometry.

as the straight waveguide, the design building block of the bending waveguide requires an area much larger than the waveguide itself to prevent coupling.

3.3.1.3 Waveguide Crossings

The model for crossings are configured as 4-port devices with parameters for the loss and crosstalk. Unlike the straight and bending waveguides previously described, the waveguide crossing is the first element thus far described that exhibits crosstalk, which is the act of inducing noise on a signal. The routing table is $[2, 3, 0, 1]$. The ordering of the indexes of the routing table are labeled in Fig. 3.7 and correspond with the cardinal directions in the following order: East, South, West, and North.

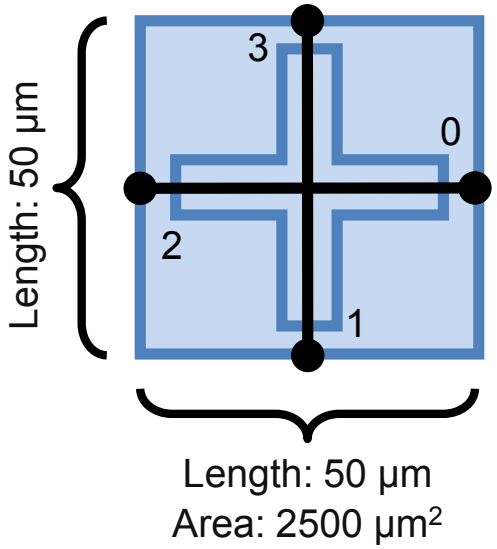

Figure 3.7: PhoenixSim representation of the waveguide crossing geometry.

The PhoenixSim waveguide crossing model assumes the design described by W. Bogaerts *et al.* [23]. The crossing is double etched at the intersection to create a mode expanding region to reduce the loss, crosstalk, and back reflection. PhoenixSim assumes a fixed crossing length of 50 μm for each waveguide, crossing at exactly the midpoint.

The latency matrix of the waveguide crossing is as follows:

$$Latency_{cross} = \begin{bmatrix} 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} \\ 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} \\ 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} \\ 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} & 50\mu m \cdot t_{wg} \end{bmatrix}$$

Observe that the latency takes the value $50\mu m \cdot t_{wg}$ for all possible input-output path combinations. Due to the symmetry that exists along each of the four arms of the crossing, any optical signal will always propagate along two arm segments (25 μm long, each).

For waveguide crossings, PhoenixSim will also request values for the following parameters: insertion loss, α_{cross}, crosstalk, α'_{cross}, and back reflection, R_{cross}.

$$Loss_{cross} = \begin{bmatrix} R_{cross} & \alpha'_{cross} & \alpha_{cross} & \alpha'_{cross} \\ \alpha'_{cross} & R_{cross} & \alpha'_{cross} & \alpha_{cross} \\ \alpha_{cross} & \alpha'_{cross} & R_{cross} & \alpha'_{cross} \\ \alpha'_{cross} & \alpha_{cross} & \alpha'_{cross} & R_{cross} \end{bmatrix}$$

Only the logically correct paths (north-to-south, south-to-north, east-to-west, and west-to-east) take the insertion loss value. Paths that must 'turn' at the intersection observe the crosstalk loss. Reflections (matrix diagonal) take on a non-negligible value unlike in the previous waveguide examples.

3.3.1.4 Couplers

Optical couplers are modeled as a 2-port device with a single parameter for insertion loss. Fig. 3.8 shows an example coupling interface between an on-chip silicon waveguide and an off-chip single-mode silica fiber. The routing table is $[1, 0]$.

The insertion loss of the coupler,$\alpha_{coupler}$, predominantly comes from the scattering and reflection, $R_{coupler}$, that occurs.

$$Latency_{bend} = \begin{bmatrix} R_{coupler} & \alpha_{coupler} \\ \alpha_{coupler} & R_{coupler} \end{bmatrix}$$

When considering the coupler dimensions, the device might require special conditioning of the waveguide and fiber on each side of the interface. Therefore the waveguide portion is

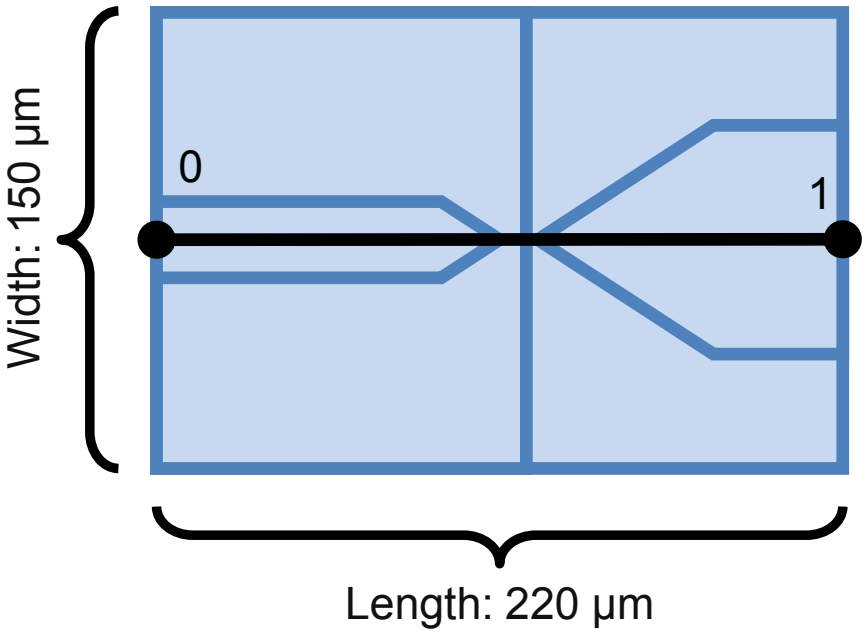

Figure 3.8: PhoenixSim representation of an example coupler geometry, connecting a silicon waveguide to a tapered fiber. In this example, the width along the lateral direction of the interface is dominated by the fiber diameter. The length accounts for the tapering at the fiber tip (right) and the inverse taper of the waveguide in the silicon substrate (right).

defined as having length, $L_{coupler.wg}$, and the fiber side is defined as having length, $L_{coupler.fiber}$. The coupling length is a summation of two waveguide segments, however the induced delay through the coupler is dependent on the effective index of both the silicon waveguide and the silica fiber. The high-confinement silicon waveguides can have an effective index of over 4 (highly dependent on waveguide dimensions) at 1550 nm [75] while standard single-mode fiber (SMF) possess an effective index of 1.47 at 1550 nm [76]. This large index contrast

requires the assumption of a different propagation delay for the fiber side, t_{fiber}.

The coupler latency is defined as follows:

$$Latency_{bend} = \begin{bmatrix} 2t_{wg}L_{coupler.wg} & t_{fiber}L_{coupler.fiber} + t_{wg}L_{coupler.wg} \\ t_{fiber}L_{coupler.fiber} + t_{wg}L_{coupler.wg} & 2t_{fiber}L_{coupler.fiber} \end{bmatrix}$$

3.3.2 Ring-Resonator Elements

As described in Section 2.1.3, ring resonators are extremely versatile structures that can be used to implement many network functions. To model the various ring resonator devices, we extend the *Basic Element Model* with subclasses for *Ring Elements* and *Dynamic Elements* (Fig. 3.9). The *Dynamic Element Model* is used to describe active devices which can exhibit changes in its routing table, latency matrix, and loss matrix during runtime. The properties of the active device during its operation is defined by *state* variables which can be changed and controlled. The *Ring Element Model* supports the definition of the resonant behavior of the devices. The behavior of ring-based devices is determined by the wavelength of the optical signal that interacts with the component. Also shown in Fig. 3.9 is how *Dynamic-Ring Elements* can be derived from the individual *Ring* and *Dynamic Element*. For instance, a ring-based broadband switch consists of a combination of ring resonators and electrical logic (described later) and can be electro-optically controlled to alter the optical flow of data.

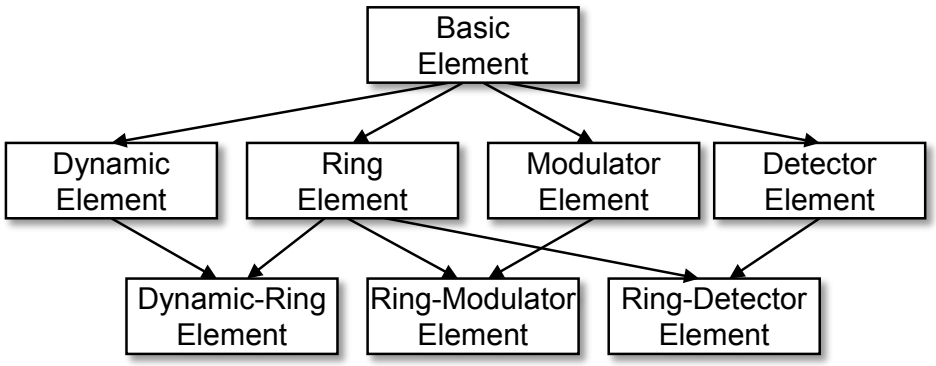

Figure 3.9: Organization of building block element classes within PhoenixSim.

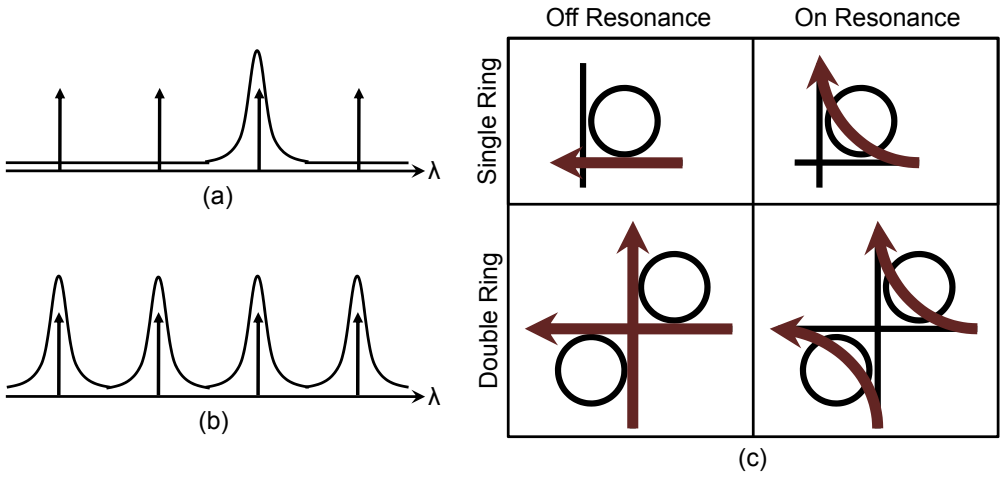

Figure 3.10: Propagation through a ring-resonator device depends on the signal wavelength and the resonant modes of the device. (a) Small rings with larger mode spacings (shown as periodic peaks) can be designed to interact with a single wavelength channel from a WDM signal (indicated by arrows). (b) Broadband switch have tightly spaced modes, enabling many WDM channels to couple into the device cohesively. (c) The path of propagation depends on whether the wavelength of the message is on- or off-resonance with the ring.

3.3.2.1 Filters

Optical filters are useful in selectively extracting a subset of wavelengths from a WDM message. In the limiting case, an extremely small ring will have a large FSR and allow the filtering of a single wavelength channel. Filtering is accomplished by aligning the spectral mode of the ring with the wavelength channel of interest (Fig. 3.10a). Light at wavelengths that align with the mode of the ring (on resonance) will couple from the ingression waveguide, into the ring structure, and out onto a secondary waveguide; wavelengths of light that are not aligned (off resonance) will be unperturbed by the ring and continue down the injection waveguide (Fig. 3.10c). We model ring filter devices as single-state 4-port *Ring Elements* with a parameter for the ring diameter (assuming a circle). Ring filters have been fabricated and demonstrated on SOI with 3-μm radius, corresponding to an FSR of 30 nm [29].

3.3.2.2 Broadband Switches

Ring resonators are also capable of controlling the flow of an entire WDM message by aligning each wavelength channel to a mode of the ring (Fig. 3.10b). This can be accomplished in a limited spectral range by using a large ring with a correspondingly small FSR. When all the wavelength channels are on resonance, the entire WDM message will couple into the ring and onto a second waveguide, similar to the case of the filter. Additionally, if the FSR is

manipulated electro-optically, all the modes can be shifted so that the wavelength channels are no longer on resonance, thus causing the entire WDM message to not couple into the ring. This functionality is illustrated in Fig. 3.10c for both a single-ring 1×2 photonic switching element (PSE) and a double-ring 2×2 PSE. These broadband switch elements are modeled as two-state 4-port devices. A 1×2 switch composed of a ring with a 100-μm radius and 0.8-nm FSR was shown to be capable of switching 20 wavelength channels simultaneously [30]. Elsewhere, a fifth-order switch was demonstrated being able to simultaneously route nine 40-Gbps wavelength channels for an aggregate data rate of 360 Gbps [33].

3.3.2.3 Modulators

Ring-based modulators are essentially high-speed switches. By electro-optically flipping the ring between an on- and off-resonance state, a series of 0's and 1's can be encoded onto an optical stream of light. Light that couples into the ring will not egress into another waveguide like the filters and switches, but will eventually dissipate within the ring. A modulator array can be formed with multiple ring modulators so that several wavelength channels can be encoded in parallel, creating a WDM signal (Fig. 3.11). Modulators should have a small ring diameter to create a large FSR to ensure that the modulation does not interfere with other spectrally adjacent wavelength channels. The modulator device is modeled as a single-state device with parameters for energy dissipated per modulated bit and ring diameter. Ring-

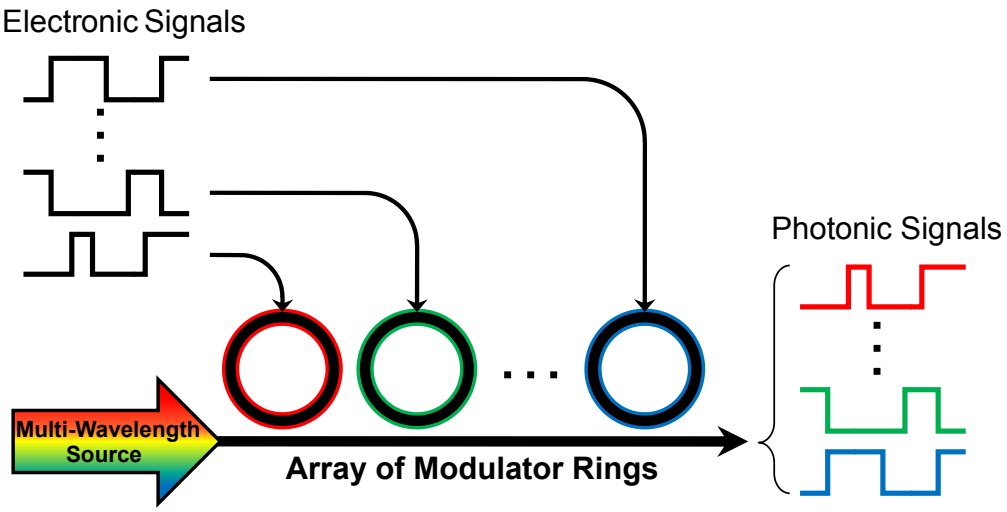

Figure 3.11: Schematic of the conversion process between the spatially-parallel electronic domain and wavelength-parallel optical domain.

based modulation has been demonstrated at rates of 12.5 Gbps in a 5-μm radius silicon ring resonator [34].

3.3.2.4 Receivers (Photo-Detectors)

PhoenixSim Detector Elements assume that a ring filter is placed before the photo-detector element for selecting specific wavelengths from a WDM signal. The *detector sensitivity* determines the minimum signal power that must be received at the photo-detector in order for data to be properly recovered from the optical domain and is an important parameter for determining the optical power budget (as discussed in Section 3.4). This ring-based detection device take parameters for energy dissipated per detected bit, sensitivity, and ring diameter.

Integrated high-speed germanium detectors have been demonstrated operating at speeds of 40 Gbps [42, 43].

3.3.3 Mach-Zehnder Elements

Switches and modulators can also be designed using the principle of Mach-Zehnder interferometry (MZI). Mach-Zehnder devices are designed to operate relatively uniformly over a large wavelength range and do not exhibit the sharp resonant peaks that ring resonators have. For instance, a MZI-based device can be used to modulate wavelengths of light that span a large continuous wavelength range while ring-resonator modulators are limited to specific resonance wavelengths. However, this operational difference between Mach-Zehnder devices and ring-resonator devices causes them to not be interchangeable. The ring-based network architectures analyzed in Section 4.4 are not compatible with these devices and would require significant changes in the designs. Models for 1×2 and 2×2 Mach-Zehnder switches are currently included in the *Photonic Device Library*. A modulator and switch based on MZI has been demonstrated operating at up to 10 Gbps [77].

3.4 Physical-Layer Performance Analysis Tools

The consideration of the photonic technology domain presents new design challenges that must be satisfied in order to produce feasible interconnect designs. Similar to electronics, it is important for photonic networks to consider power dissipation and system-level performance. Furthermore, photonic networks must also consider metrics that have no electronic equivalent such as insertion loss, the optical power budget, noise, and crosstalk. While a comprehensive analysis of a photonic interconnect design would involve the actual fabrication and operation of such a system, this is currently unrealistic since full-scale photonic on-chip networks are still in early stages of research. Therefore, the tools presented here can give important insight into the physical feasibility of the designs and the performance that is expected.

3.4.1 Optical Power Budget

The optical power budget of a photonic network assesses the amount of WDM parallelism and insertion loss that can be tolerated. Many currently proposed photonic interconnection networks assume off-chip lasers to provide the optical sources, which are then coupled into the chip where they are modulated, routed, and received. Optical amplification in an on-chip environment is not easily accomplished in the CMOS platform. For this reason, the power that is received at the photo-detectors must remain above a certain power threshold

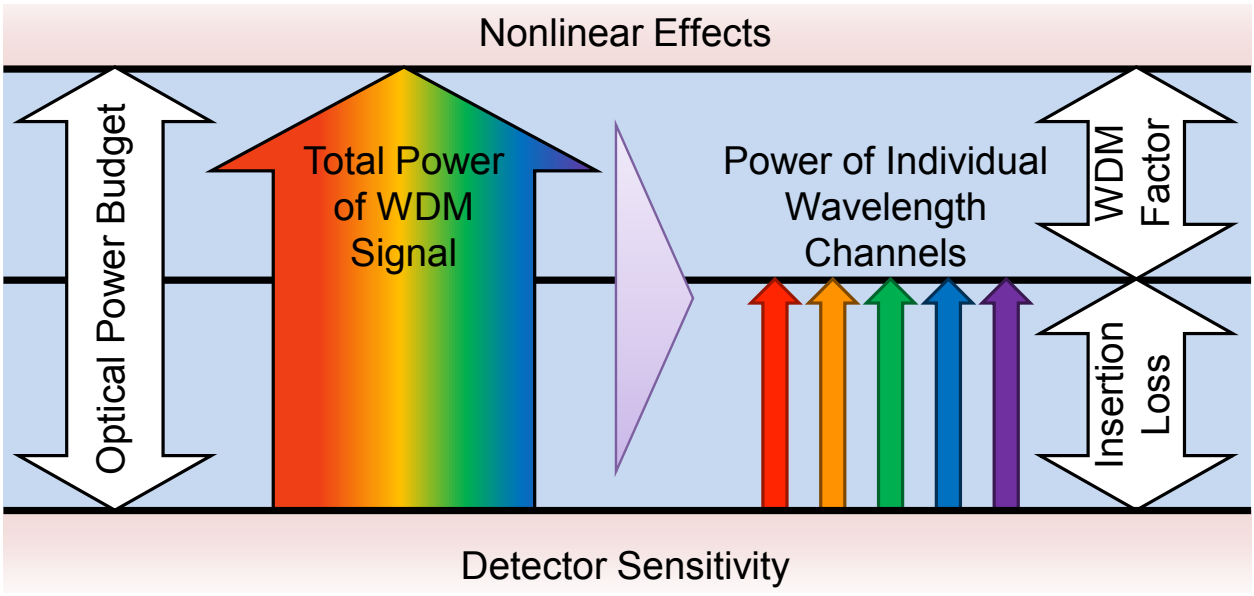

Figure 3.12: The relationship of various parameters affecting the optical power budget. The difference in power of the total WDM signal (large arrow on the left) and the individual wavelength channels (five smaller arrows on the right) constrains the scalability of the system.

(labeled the detector sensitivity in Fig. 3.12) to ensure proper detection of data bit streams. This limitation can be partially compensated for by increasing the optical power that is injected into the chip. However, this also exhibits an upper limitation due to nonlinearities of the silicon material which will potentially distort the signal. Distortions are caused by nonlinearities within silicon which contribute additional insertion losses and can also causes unwanted shifts in the resonances of ring resonators. This limit is labeled as *nonlinear effects* in Fig. 3.12. The difference in the two thresholds is called the *optical power budget*.

As shown in Fig. 3.12, the optical power budget affects the design choices of a given

network architecture by constraining the sum of the WDM factor and the network insertion loss. The *WDM factor* measures the power difference between an entire WDM signal and its constituent wavelength channels. This factor needs to be accounted for since the nonlinearity threshold is determined by the total power in the waveguide while the detector sensitivity depends on the power in the individual wavelengths. The remaining portion of the optical power budget must accommodate the worst-case insertion loss that an optical message could receive in the network. Fig. 3.13 shows an example of the calculation involved in determining the insertion loss for an optical signal being injected into a small network segment at 1 dBm. The signal is ejected at 0.24 dBm after propagating across a 0.1-cm distance, passing by two ring resonators, and entering four waveguide crossings. The total loss for this example is 0.76 dB. For a full-scale photonic network, all valid optical paths need to be examined to determine the highest-loss path.

The relationship between the various device limitations and system-level metrics is summarized in the inequality

$$P - S \geq IL_{max} + 10log_{10}n \tag{3.1}$$

where P is the power threshold we limit the optical power to and S is the detector sensitivity. The optical power budget is $P - S$. The worst-case optical path in terms of insertion loss is IL_{max} and n specifies the number of wavelength channels being used. P, S, and IL_{max} are

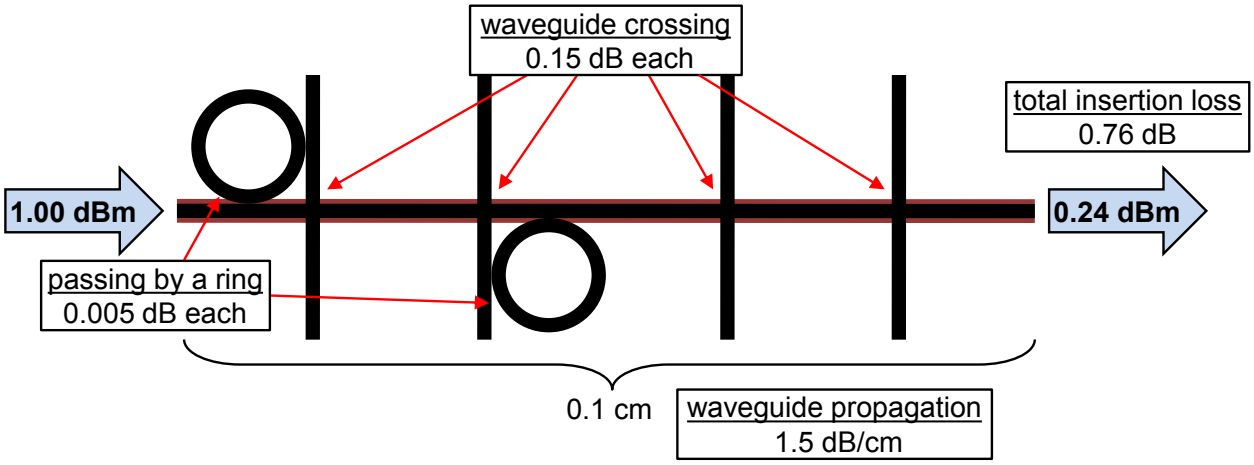

Figure 3.13: Calculation of insertion loss for a small network segment.

expressed in decibel units.

While it may be desirable to maximize the number of wavelength channels used to increase bandwidth through parallelism, and to create scalable photonic networks at the cost of higher insertion losses, Eq. (3.1) shows the inherent limitation to this. From an architectural standpoint, P and S are fundamental design constraints imposed by the photonic devices. Therefore, a designer must strike a balance between the desired link bandwidth and the desired complexity of the network. In Section 4.4, we illustrate the evaluation of these tradeoffs which are made possible by PhoenixSim.

3.4.2 Data Integrity

A variety of interactions in a photonic interconnection network will work to degrade the integrity of transmitted data. Our current noise modeling methodology accounts for intensity noise generated at the laser sources, inter-message crosstalk, intra-message crosstalk, and electrical noise generated by the optical receivers (Fig. 3.14). The standard figure of merit for measuring the quality of signal is the signal-to-noise ratio (SNR) which is defined as the ratio between signal power and noise power. More specifically, the optical SNR (OSNR) is the ratio of optical signal power to optical noise power at the point where the measurement is being made. From a system perspective, the SNR can be used to determine the statistical likelihood that each bit of data is transmitted erroneously (*e.g.* a transmitted 0 is detected as a 1), also called a *bit error rate* (BER). An understanding of the potential noise in any interconnection network is critical to determining the effective throughput of the system since error detection and correction will invariably cause performance overheads.

The first source of noise is from the laser sources which inherently cause random fluctuations in an optical signal, called intensity noise. This noise is quantified as relative intensity noise (RIN), which is the ratio of the power variance of the optical signal to the mean optical power squared. Quantum cascade lasers have a measured RIN on the order of -150 dB Hz^{-1} with an output of 10-dBm mean optical power [78]. To convert to a SNR, we

use the relation [79]:

$$SNR_{laser} = \frac{m^2}{2B \cdot RIN} \tag{3.2}$$

where B is the noise bandwidth, assumed equal to the modulation rate, and m is the modulation index, equal to $1 - E$, where E is the extinction ratio of the modulator.

A second source of noise is *inter-message crosstalk* which occurs when multiple photonic messages concurrently propagate through a photonic device. In a waveguide crossing for example, the ideal situation is for two orthogonally propagating messages to be completely isolated from each other with no interaction. However, in reality a small amount of optical power from each message will leak onto the other message. A similar situation occurs in ring-resonator filters and switches due to imperfect coupling of each wavelength channel.

For the N-port device, the crosstalk power that a message on a particular port receives is given by the sum of the power that is leaked by any existing messages on the other $N - 1$ ports. If M is the set of all signals present in the device and the power of a signal k is given by the variable P_k, then the crosstalk power seen by signal s is given by

$$\sum_{k \in M, k \neq s} \frac{P_k}{IL(portin_k, portout_s)} \tag{3.3}$$

which aggregates the unwanted signal power that leaks into the output port being used by s. Function IL refers to the *insertion-loss matrix* (that was described in Section 3.3) of the device model with arguments for the input and output port. In Eq. (3.3), $portin_k$ denotes

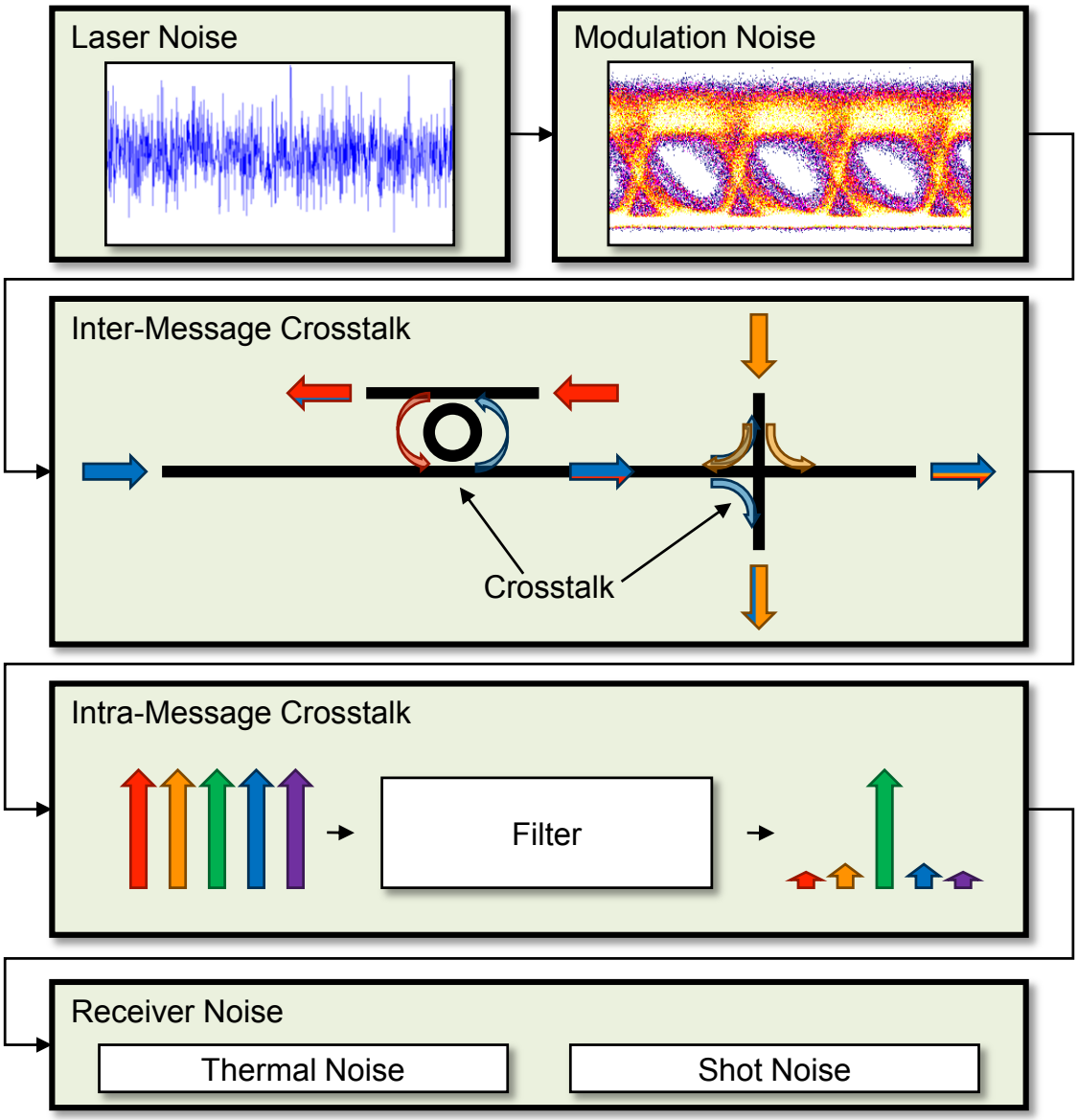

Figure 3.14: Sources of noise and crosstalk within a chip-scale photonic system.

the input port of a message k, and $portout_s$ denotes the output port of s. This calculation is a first-order approximation that only considers crosstalk for messages that coexist in a device and not from leaked power that propagates across multiple devices before interfering

with a foreign signal.

A third source of noise called *intra-message crosstalk* occurs due to imperfect filtering. For example, in order for a WDM message to be received and converted into an electrical signal, each wavelength channel must be individually filtered and fed into a photo-detector. Due to imperfect extinction, power from the adjacent wavelength channels will leak through causing an additional source of noise. Intra-message crosstalk will also occur in any other location in a photonic network where filtering functionality is involved. The spectral response of a ring resonator mimics a periodic Lorentzian function. For simplicity we assume a periodic flat passband and constant extinction ratio for the stop bands. Lastly, our receiver model includes thermal and shot noise.

The combined effect of these multiple sources of noise can be used to compute an SNR for the final detected signal with the following equation:

$$SNR = \frac{P}{N_{laser} + N_{inter} + N_{intra} + N_{therm} + N_{shot}} \tag{3.4}$$

where P is the signal power and N corresponds to the noise power associated with the noise or crosstalk source indicated by the subscript.

3.4.3 Power Dissipation

To compute the power dissipation of the modeled networks, we add up the energy dissipation events from all devices. Our photonic device library tracks the power dissipation according to the type of model that is used, and can include both static (over a duration of time) and dynamic (instantaneous) power dissipation. *Dynamic Element* devices can have static power dissipation, which is determined by the occupied state. *Dynamic Element* devices can also have dynamic power dissipation, which is accumulated whenever there is a state transition. An additional source of power dissipation are *Ring Element* devices, which require constant thermal tuning to compensate for fabrication uncertainty and ambient temperature shifts. *Modulator* and *Detector Elements* also dissipate power during the transmission and detection of data, respectively.

Electronic routers are modeled as standard three-stage pipelines. The power modeling of the electronic routers is accomplished by leveraging the ORION simulator, which is currently capable of modeling down to the 32 nm technology node [72].

3.5 Integration With Other Simulators

In addition to the PhoenixSim code base, we integrate and leverage a number of third party tools and simulators. This enables us to simulator our networks with a richer and

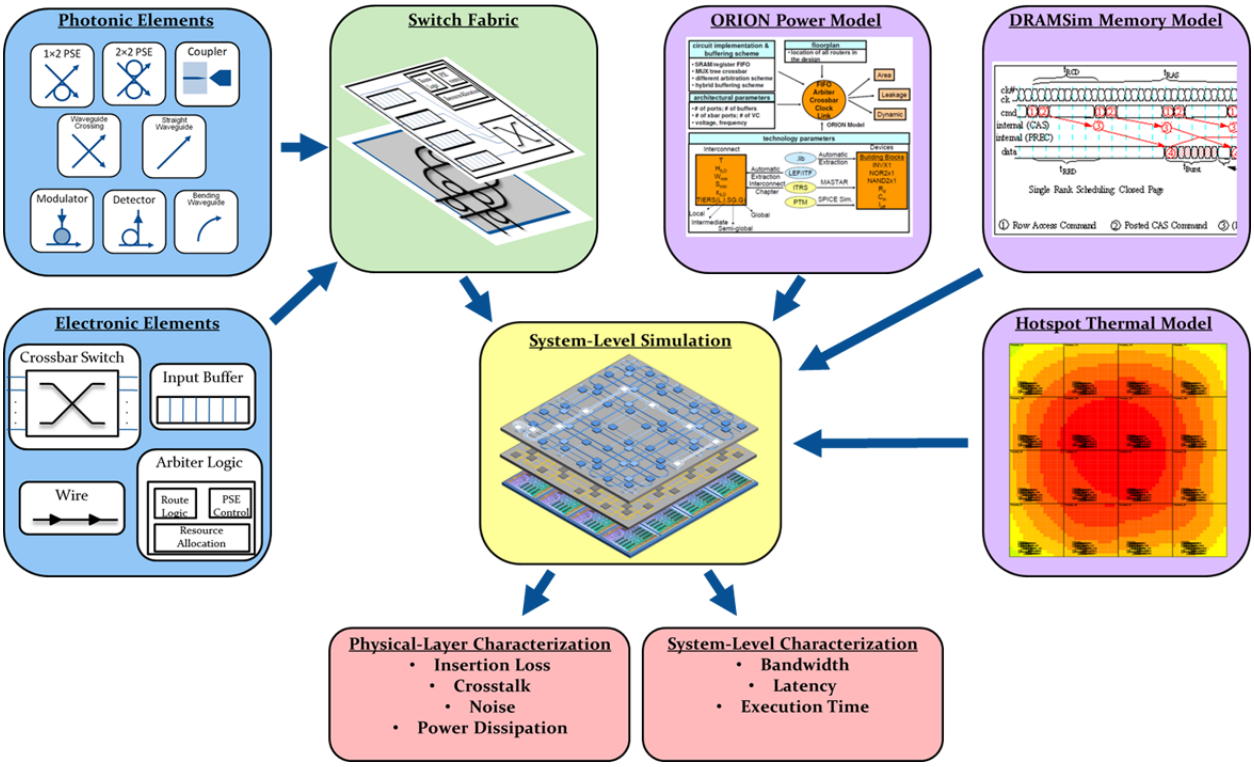

Figure 3.15: Organization of the PhoenixSim environment.

more detail level of precision. Fig. 3.15 shows the organization of PhoenixSim with these integrated third party tools. As mentioned before, we include ORION for its electronic router power model [72]. For modeling memory, we include DRAMSim which was developed at University of Maryland [80]. We also include the Hotspot thermal simulator which came from University of Virginia [81, 82].

Figure 3.16: Simulation server rack located in the Lightwave Research Laboratory at Columbia University. Servers used for simulation are the first, second, fourth, and fifth from the top.

3.6 Simulation Infrastructure

Research results often require the execution of 100's of simulation runs per plot. In order to condense simulation execution times, we built up a set of four multi-processor servers (Fig. 3.16) to allow us to execute multiple simulations in parallel. The four servers possess 64 processors and 288 GB of DRAM in aggregate. Ubuntu distributions were installed on

each machine. The large amount of DRAM was a relatively inexpensive upgrade added to the servers, which enabled us to perform some special case simulations which utilized an extremely large memory footprint (multiple gigabytes).

Chapter 4

Physical-Layer Analysis of Photonic Interconnection Networks

In this chapter, several synergistic physical-layer and system-level analyses of previously proposed network architectures are discussed.

4.1 Photonic Circuit Switching Primer

Since the spatial routing technique is a central component to most of the conducted research, a detailed description of this interconnect style is first provided here. Also, the new architecture discussed in Chapter 5 derives heavily from the circuit-switching protocol.

The high-level structure of the photonic circuit-switching technique is illustrated in

Fig. 4.1. A photonic circuit-switching enabled chip is composed of three logical layers: a processing layer, electronic control plane, and a photonic data plane. The processing plane is where the processing nodes sit and act as the sources and sinks for all communications. The top most layer, the photonic data plane, provides high-speed WDM-enabled optical links between any pair of communicating processors. However, because the photonic plane cannot be adjusted all-optically, the photonic devices need to be preconfigured before any optical data can be transmitted. For this reason, a electronic control plane is provided for the purposes of configuration.

Fig. 4.1 shows an example photonic plane topology (top layer) with lines representing waveguides and blocks representing photonic routers and gateways. Underlying the photonic plane is the electronic control plane composed of standard metal wires (yellow lines) and electronic routers (grey blocks). The electronic wires and routers are strategically placed so that the network exactly mirrors the photonic version. The reason for this placement is to facilitate the circuit-switching process. Each node of the processor has a connection to a gateway on the optical plane (for data generation and reception) and a connection to the control plane.

The steps for establishing an optical data path are as follows. A processor node with a request for sending data must first establish the photonic link using the electronic control

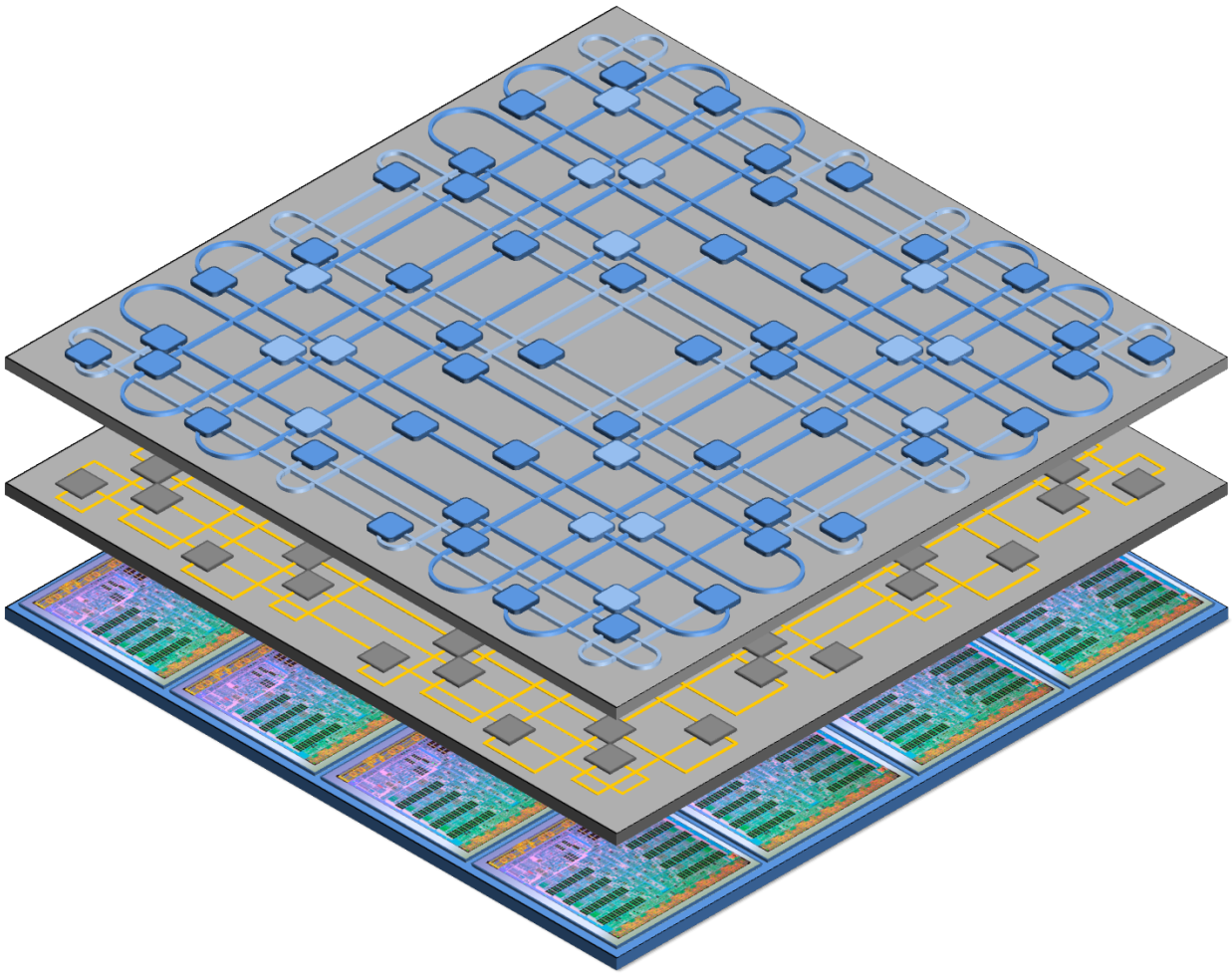

Figure 4.1: The envisioned chip stack for photonic circuit switching. Three logical primary layers consisting of a processing layer (bottom), electronic control plane layer (middle), and photonic data plane layer (top).

plane. The node inserts a *PathSetup* message which contains the destination address in the header. The *PathSetup* message travels through the electronic network and traces out a possible path for the optical message to take. At each router hop, the state of the associated

photonic router is checked (contained within the logic of the electronic router). If the path is available for use, then a reservation is set for the particular path and the *PathSetup* message proceeds towards the destination.

If any resource is unavailable (*e.g.* a ring switch has been electro-optically controlled and is currently controlling a signal) then a *PathBlocked* message must be returned. The *PathBlocked* message retraces the route of the *PathSetup* message so that all reservations can be canceled. Once the *PathBlocked* message reaches the source node, the node will be signaled to reattempt the transmission after some hold-off period.

If the *PathSetup* reaches the destination, the destination can deduce that a complete optical path on the photonic plane has been reserved, and returns a *PathAck* message. The *PathAck* retraces the exact same path all the way to the source. At each hop, the previously established reservation during the *PathSetup* traversal is exercised and the appropriate photonic devices are actuated. Before the *PathAck* proceeds, the newly activated devices are flagged so that other *PathSetup* messages cannot change them.

When the *PathAck* message reaches the source, the source knows that a complete optical path on the photonic data plane has been established and can begin to transmit data. Once the last data bit has been sent, a *PathBreakdown* message is immediately sent. Again, this message will trace out the same path as the original *PathSetup* message so that previously

allocated photonic devices are freed and able to be utilized by a future path request.

4.2 Insertion Loss Analysis of 4×4 Switch Designs for Photonic Networks

In this section, an analysis is conducted to compare the performance of varying 4×4 non-blocking switch designs. In mesh-style optical networks, the 4×4 switch provides the main routing mechanism for guiding lightwaves at each intersection of the network.

For this analysis, a circuit-switching folded torus topology is used [51]. The modulators and switches throughout the network (Fig. 4.2) are designed using ring resonator based electro-optic devices. Additional broadband switches are placed in the network to allow packets to enter (injection), route, and exit (ejection) the interconnection network topology. The separate electronic control plane provides the necessary functions to arbitrate a complete circuit switched optical path from source node to destination node. Fig. 4.2 shows the main folded torus in thick black lines. An additional gateway access network shown as thin red lines is required to enable entering and exiting the network.

An important issue not considered in previous work is the spatial layout of the optical components. The layout can significantly affect the available power budget of each optical signal. The simulation model assumes a tile size of 2.0 mm × 1.5 mm, which is the size

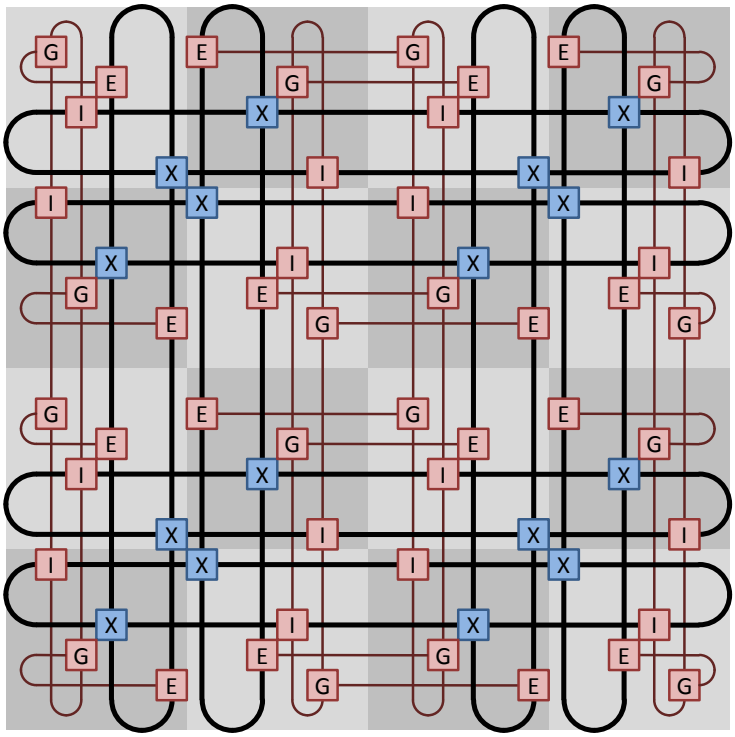

Figure 4.2: Structure of a 4×4 node torus topology. The waveguides that make up the torus network are shown as thick lines, and the gateway access network for injecting packets to and ejecting packets from the network shown as thin lines. The blocks represent the following: gateway switch (G), injection switch (I), ejection switch (E), and a 4×4 non-blocking switch (X).

of a single core in Intels 80-core chip [83]. Fig. 4.3 shows a typical layout of a tile in the photonic plane. Each tile consists of a gateway switch, injection switch, ejection switch, 4×4 non-blocking switch, and several optical paths to form the torus and gateway access network.

The network uses a folded-torus topology and X-Y dimensional ordered routing. The

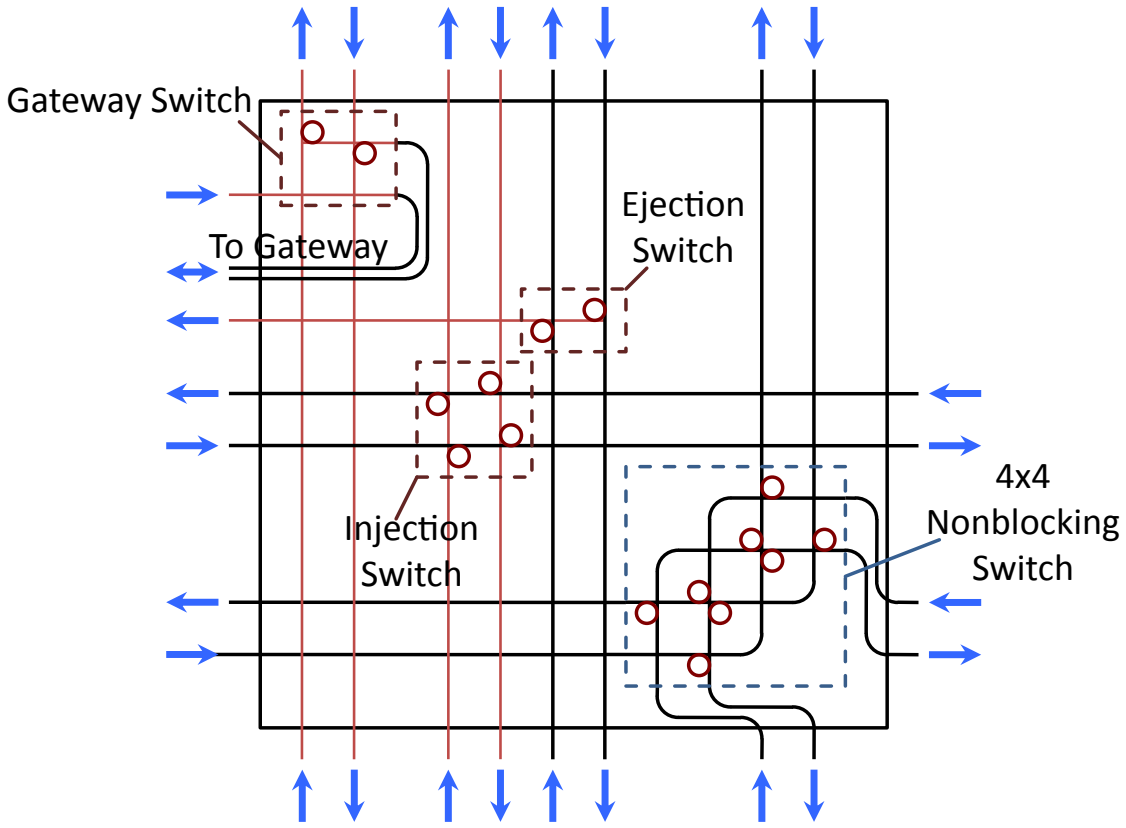

Figure 4.3: Layout of a tile in the torus network. This includes the type (A) version of the 4×4 non-blocking switch shown in Fig. 4.4.

simulation uses uniformly distributed generated transmission requests with exponentially distributed interpacket spacing. Although insertion loss is independent of network congestion, an arbitrary constant message length of 50 ns, equivalent to an 8 kb size packet at 160 Gb/s [30], is used. Layout differences and losses due to the on-chip routing of continuous-wave light and off-chip messages into each gateway are ignore for this simulation.

The insertion loss parameters are shown in Table 4.1. The values are obtained from

Table 4.1: Insertion Loss Parameters - 4×4 Non-blocking Switch Study

Parameter	Value	Ref.
Propagation Loss (Silicon)	1.5 dB/cm	[19]
Waveguide Crossing	0.05 dB	[23]
Waveguide Bend	0.005 dB/90°	[19]
Drop Into a Ring	0.5 dB	[30]
Pass By a Ring	0.005 dB	[30]

reported devices and predictions for future scaling.

4.2.1 Simulation Results

The analysis here investigates how insertion loss is affected by changes in topology size and different switch designs. Tori of size 4×4, 6×6, and 8×8 are considered. The different switch designs, described later, are labeled (A), (B), and (C) (Fig. 4.4).

Fig. 4.5 shows the distribution of insertion loss that a packet will experience when propagating from source to destination. Minimum losses for each switch layout remains constant for differing network sizes. For every additional two nodes in each dimension, 3.89 dB, 3.66 dB, and 3.36 dB of loss is added to the maximum loss for switch design A, B, and C, respectively. This is a result of the fact that the minimum length path from any two

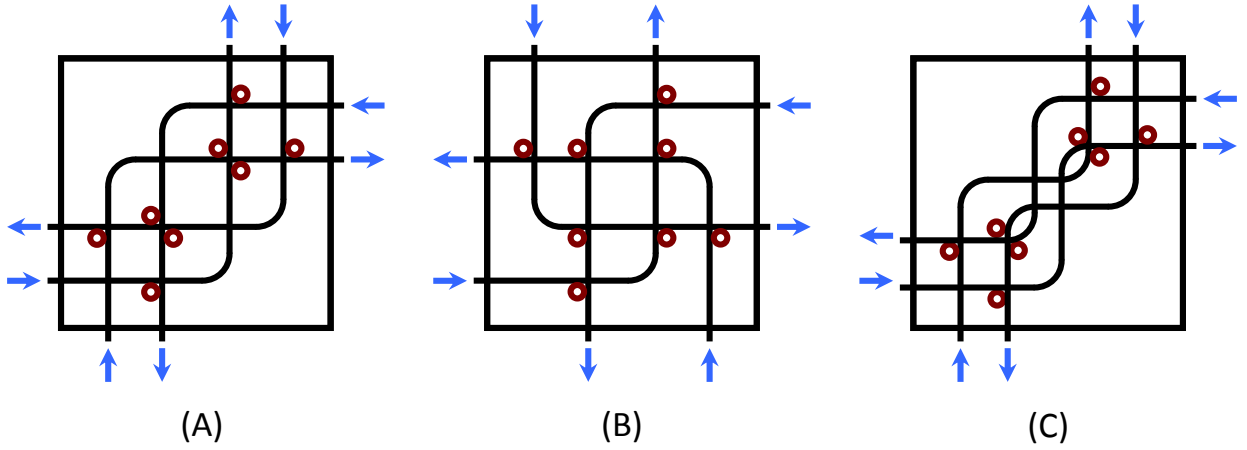

Figure 4.4: Three implementations of the 4×4 non-blocking switch.

nodes remains the same while the maximum length will change with number of nodes. The three, five, and seven peaks that appear in the distribution for the 4×4, 6×6, and 8×8 node torus networks, respectively, equates to the maximum number of 4×4 non-blocking switches an optical packet must travel through, which rises as the node count scales up.

Next, we explore the performance of three different 4×4 non-blocking switch designs (Fig. 4.4): (A) is a design first introduced in [84]. (B) contains a reduced number of waveguide crossings while keeping the number of ring resonator structures at eight. (C) differs from the previous two designs by allowing packets that require a straight path to propagate through without requiring a turn at a ring. Each design is non-blocking when no u-turns are allowed.

Each plot in Fig. 4.5 shows the general trend of the different switch designs. (B) has both a lower maximum and lower minimum loss, in comparison to (A), as expected. (C)

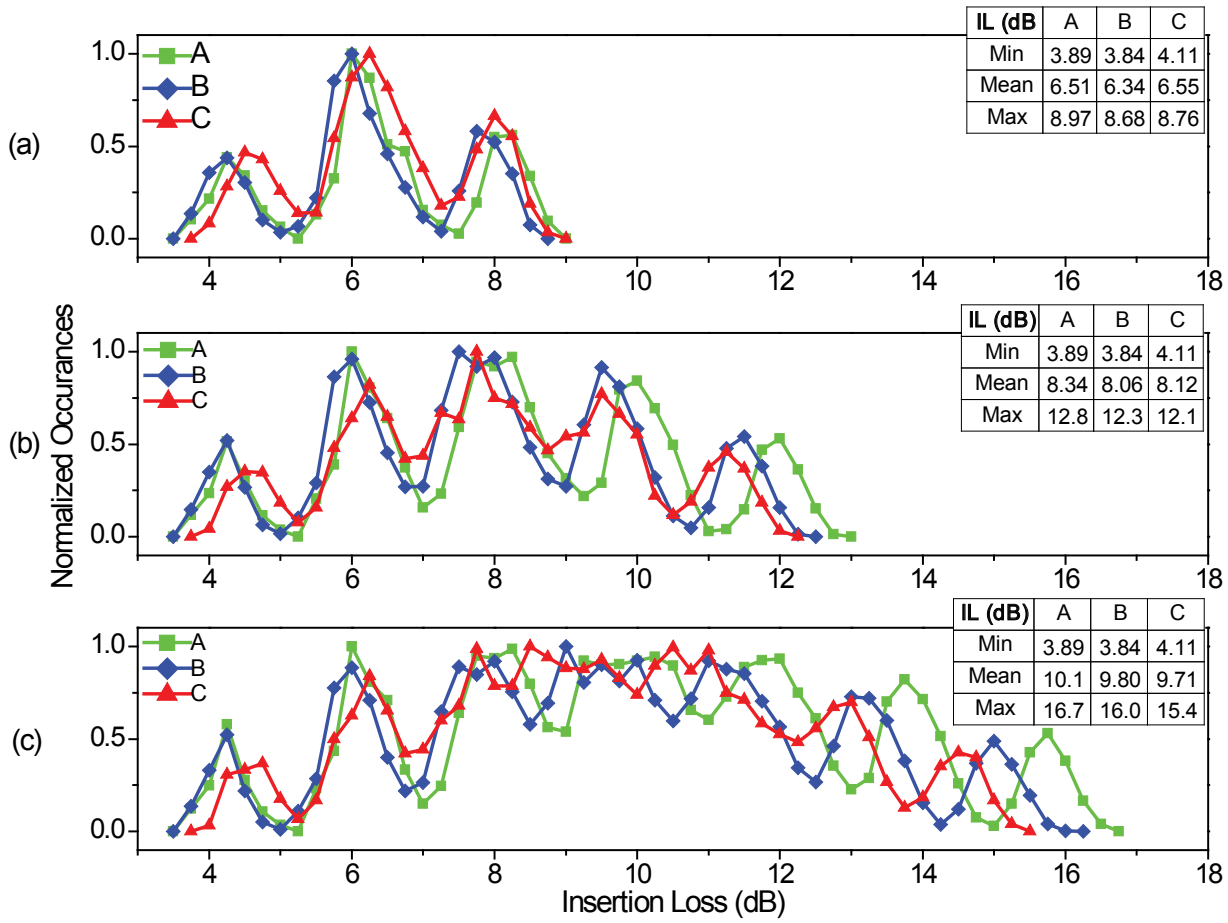

Figure 4.5: Insertion loss distribution for folded torus topologies of size (a) 4×4, (b) 6×6, and (c) 8×8. Each graph contains plots of three differing switch designs. Inset within each graph is a table of minimum, mean and maximum insertion losses observed for each case.

consistently has a higher loss for the lower bound of the distribution. Although (C) exhibits higher maximum loss in the 4×4 node network than (B), it shows lower loss at sizes of 6×6 nodes and higher. This is attributed to the fact that even though the minimum insertion loss for this 4×4 switch design is higher than the others, the straight path (from north to

south, east to west, or vice versa) has a lower loss because no rings are encountered. In contrast the paths in (A) and (B) that do not pass through any ring resonators implement a turn. The performance improvement noticed with switch (C) is a consequence of using dimensional ordered routing makes a single turn in any optical path, and mostly straight propagation through the switches.

4.3 Physical-Layer Analysis of Photonic Circuit Switching

This section focuses on the physical-layer analysis of space-switched photonic networks. Two previously proposed topologies are the Torus [51] and a Non-blocking Torus [85], shown in Fig. 4.2 and Fig. 4.6, respectively. We define a node (marked X) as the logical switching point on the network, whereas an access point (marked G) is a gateway where a network user (e.g. a processor node) can initiate or receive a transmission. The nodes are implemented with the non-blocking 4×4 switch. The primary folded-torus path in both networks is illustrated with thick lines to represent two waveguides forming a bi-directional link. The remaining thinner lines and blocks (I, E, and S) indicate the location of additional waveguides and switches that compose the access network, which is needed to enter and exit the tori.

The primary difference between the two topologies is the manner in which access points are mapped to nodes. The Torus has an access point mapped to every node, while the

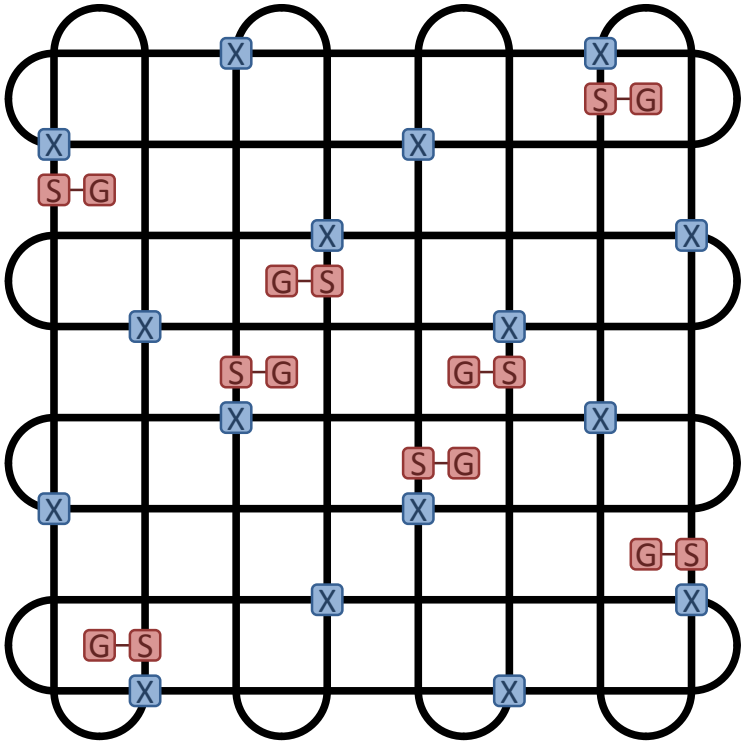

Figure 4.6: 4×4 Non-blocking Torus with 8 access points. X labels mark 4×4 non-blocking switching points. G labels mark access points. S labels indicate combined injection-ejection switching points.

Non-blocking Torus is limited to two access points on each row and column of nodes in the torus in order to achieve a strictly non-blocking network. For example, an 8×8 torus would allow 64 access points in a normal configuration, but would only allow 16 access points in a non-blocking configuration. Previous studies have shown that the non-blocking property can be advantageous in both throughput and latency compared to blocking networks [85], but performance improvements will be offset by the physical layer constraints that have not

previously been considered.

We simulate the networks using PhoenixSim, a physical-layer simulator that we have developed. The simulation topology models assume die sizes of 2.0 cm × 2.0 cm.

4.3.1 Insertion Loss Analysis

Our study assumes loss parameters close to currently realizable values and are summarized in Table 4.2. Note that ring resonators exhibit a strong thermal dependency which could potentially cause additional losses, increased crosstalk, and disruptions in the network. Thermal management of ring resonator devices is currently an active research topic with proposed solutions that include integrated heaters for thermal compensation [86] and athermal devices [87]. For this simulation work, we assume an adequate mechanism for managing this issue.

The maximum possible loss (across all paths) that a message will incur from each type of component in the Torus and Non-blocking Torus is shown in Fig. 4.7 for networks ranging from 4×4 to 18×18 nodes. Losses due to bending waveguides and passing a ring off resonance are negligible and are not shown. As the photonic network topology scales to support more access points, signals will incur higher losses due to more waveguide crossings and switching elements.

Table 4.2: Insertion Loss Parameters - Photonic Circuit-Switching Analysis

Parameter	Value	Ref.
Propagation Loss (Silicon)	1.5 dB/cm	[19]
Waveguide Crossing	0.15 dB	[23]
Waveguide Bend	0.005 dB/90°	[19]
Drop Into a Ring	0.5 dB	[30]
Pass By a Ring	0.005 dB	[30]

The waveguide crossings are shown to be the most significant component of optical losses reaching as high as 68% for the Torus and 61% for the Non-blocking Torus. The contribution of loss from dropping into a ring on resonance for the Torus and Non-blocking Torus regardless of topology size are approximately 17% and 20%, respectively, whereas propagation losses in the 4×4 configuration are as high as 43% and 49%, respectively, and gradually decrease in percentage as the topology size increases. The decreasing trend in percentage for propagation loss is due to the assumed fixed size of the die keeping the approximate maximum propagation distance equal while other components continue to scale in number as the topology size increases. Passing by rings off resonance and passing through waveguide bends induce relatively negligible losses in these topologies. Consequently, the most beneficial improvements to these networks can be achieved through either a reduction

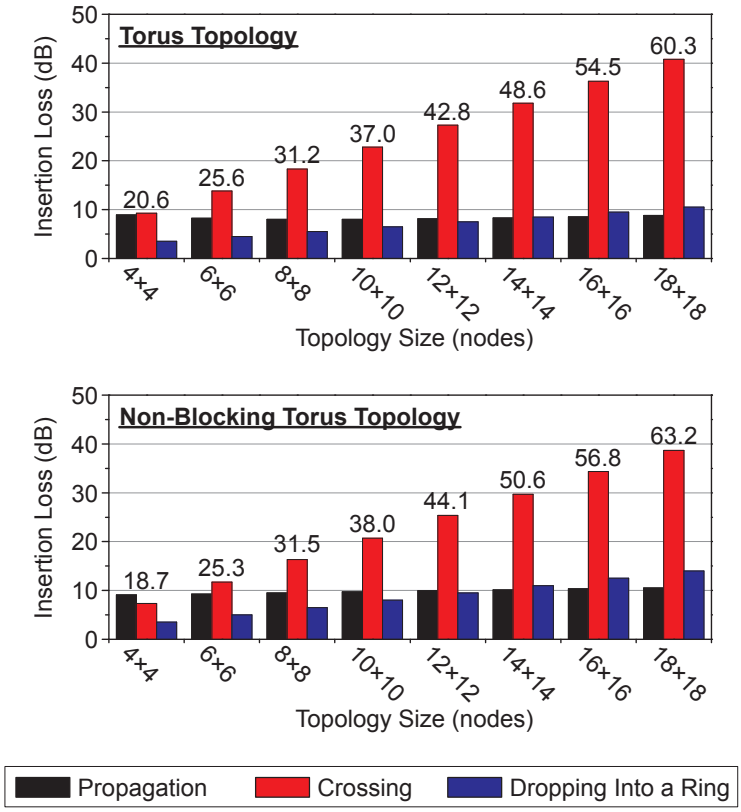

Figure 4.7: Maximum possible network-level insertion loss by component for varying sizes the Torus and Non-blocking Torus using the parameters listed in Table 4.2. Labeled values represent the peak cumulative insertion loss (in dB) for the network.

of waveguide crossing losses or through the redesign of the switching fabric layout to reduce the number of crossings.

4.3.1.1 Device Improvement

The previous analysis of network-level insertion loss of the Torus and Non-blocking Torus suggests that research advancements in lower-loss crossings will have the most impact in increasing system performance. In particular, two system parameters stand to gain with improvements in loss, the bandwidth available to each access point which is specified by the number of wavelengths, and the number of access points available in the network. We examine in simulation a hypothetical improvement in crossing loss, and use Equation 3.1 to determine the impact it will have on network scalability.

Fig. 4.8 and Fig. 4.9 shows the maximum number of wavelengths that are allowed for varying topology sizes and the change in performance when assuming a hypothetically better crossing loss of 0.05 dB (compared with 0.15 dB in the original case). The gains in system-level performance from the improved crossings are apparent from the networks support for more access points and greater numbers of wavelengths. For instance, assuming a 30-dB allowed network-level optical power budget, the maximum connectivity supported on the Torus scales from 36 access points when using the original crossings to 196 access points when using the improved crossings (a more than five-fold increase). Similarly, the Non-blocking Torus scales from 12 to 24 access points. On the other hand, we can fix the Torus topology to 36 access points and have a gain in the number of possible wavelength channels from 2

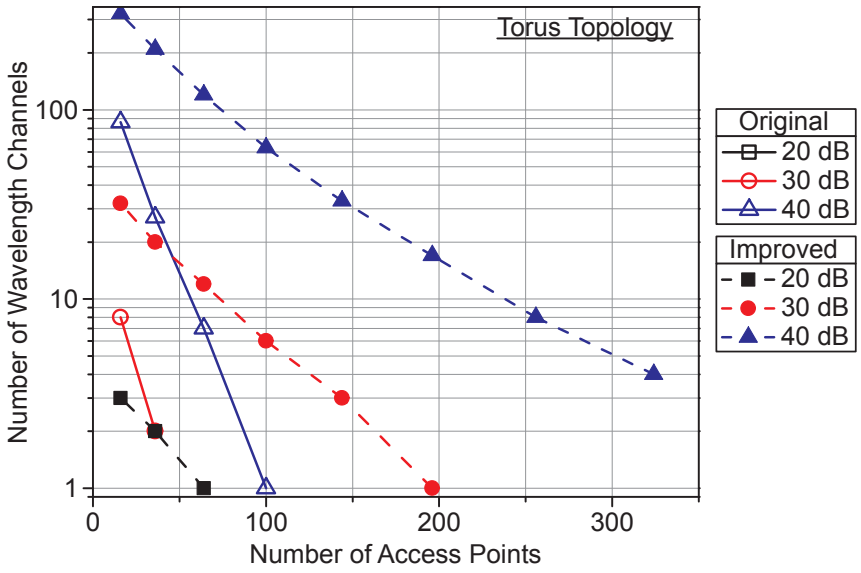

Figure 4.8: Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Torus topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved).

to 20 (ten-fold increase in bandwidth), while a Non-blocking Torus with 12 access points will increase from 2 to 15 wavelength channels. For the case of the Torus network operating with a 20-dB optical power budget and original parameter set, the network configuration is unable to produce any wavelengths since the worst-case insertion loss exceeds the optical budget.

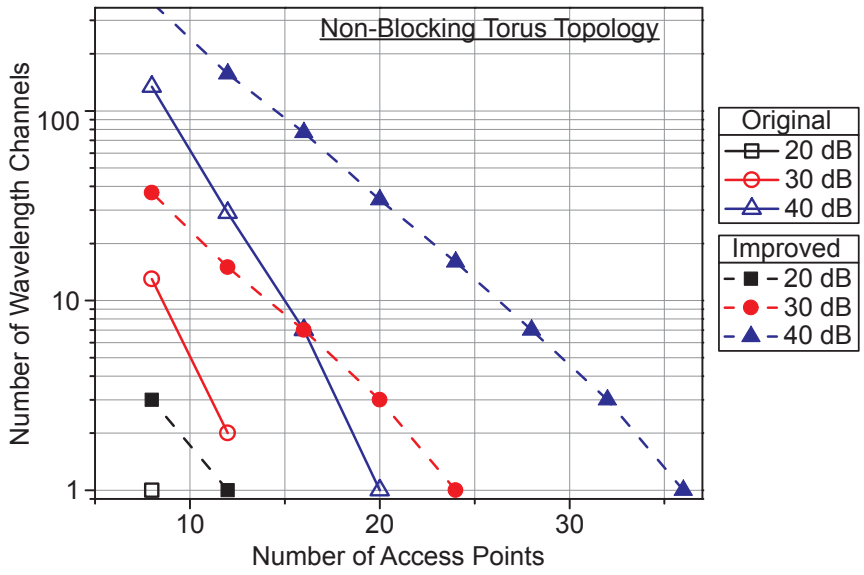

Figure 4.9: Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Non-Blocking Torus topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved).

4.3.1.2 Topology Exploration

Network performance improvement can also be achieved though design optimizations that decrease network-level insertion loss. As was shown, waveguide crossing losses are the dominant contribution to the total optical insertion loss. Therefore, designs that decrease the number of crossings will be advantageous. TorusNX and Square Root were designed with this objective in mind.

A significant amount of loss in the original Torus is attributed to two reasons. First,

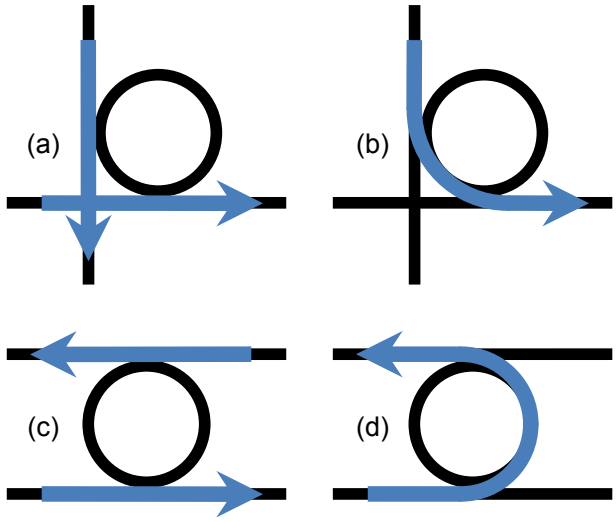

Figure 4.10: Light propagation in 1×2 PSE. (a) Off-resonance propagation with crossing. (b) On-resonance propagation with crossing. (c) Off-resonance propagation without crossing. (d) On-resonance propagation without crossing.

the usage of the access network introduces an additional set of waveguide crossings which produce a high insertion-loss overhead. Secondly, the Torus (and also Non-blocking Torus) is designed using only the 1×2 and 2×2 PSEs which both contain an embedded waveguide crossing (Fig. 4.10a and Fig. 4.10b shows the 1×2 case, Fig. 4.11a and Fig. 4.11b show the 2×2 case). These switch designs were suitable for prior investigations into photonic networks since the studies did not consider insertion loss, but our analysis shows that the overall system performance would be significantly impacted. In many circumstances, a designer can take advantage of an alternative 1×2 (Fig. 4.10c and Fig. 4.10d) and 2×2 (Fig. 4.11c and Fig. 4.11d) PSE design which eliminate the crossing and reduce the insertion loss impact on

off-resonance message traversal but keep similar switching functionality.

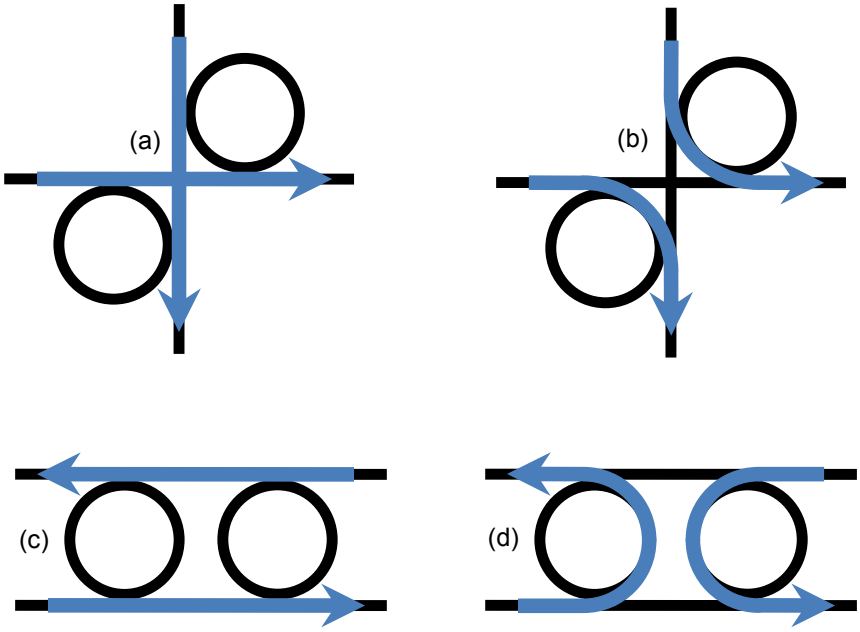

Figure 4.11: Light propagation in 2×2 PSE. (a) Off-resonance propagation with crossing. (b) On-resonance propagation with crossing. (c) Off-resonance propagation without crossing. (d) On-resonance propagation without crossing.

The TorusNX topology (Fig. 4.12) is designed to preserve the connectivity and scalability of the original Torus topology while lowering the overall insertion loss. The name of this topology means 'torus, no crossings' and alludes to the strategy used in the designing of this network. Many design decisions were made in order to significantly reduce waveguide crossings and to reduce the insertion loss overhead.

In contrast with the Torus which required a complex access network to facilitate injection

and ejection from the network, TorusNX uses a new gateway design (Fig. 4.13) which splits the access point into two blocks for modulation and detection and circumvents adding any additional crossings to the torus through the use of the 1×2 PSE variant. The modulation block enables a message to be injected north or south while the detection block can receive signals coming from the east or west direction. This scheme is well suited for dimension-ordered routing which is the assumed routing for this topology. TorusNX also uses an optimized version of the 4×4 non-blocking switch which was shown in Section 4.2 to perform better in dimension-order routed topologies.

The Square Root topology was also designed with fewer waveguide crossings and fewer switches in mind by simplifying the entire network into only using 4×4 non-blocking switches. In addition to the axioms used to reduce insertion loss in the physical layer, the Square Root also uses hierarchical organization to simplify routing, and path multiplicity between organizational units to increase performance.

The Square Root is constructed recursively beginning with a 2×2 *quad*, shown in Fig. 4.14a, which has no waveguide crossings outside the 4×4 switches. A 4×4 Square Root is composed of four sets of quads, and is shown in Fig. 4.14b, connecting quads through central switches and inter-quad express lanes. In a similar fashion, an 8×8 Square Root can be constructed from four 4×4 Square Roots. This recursive construction can be used to

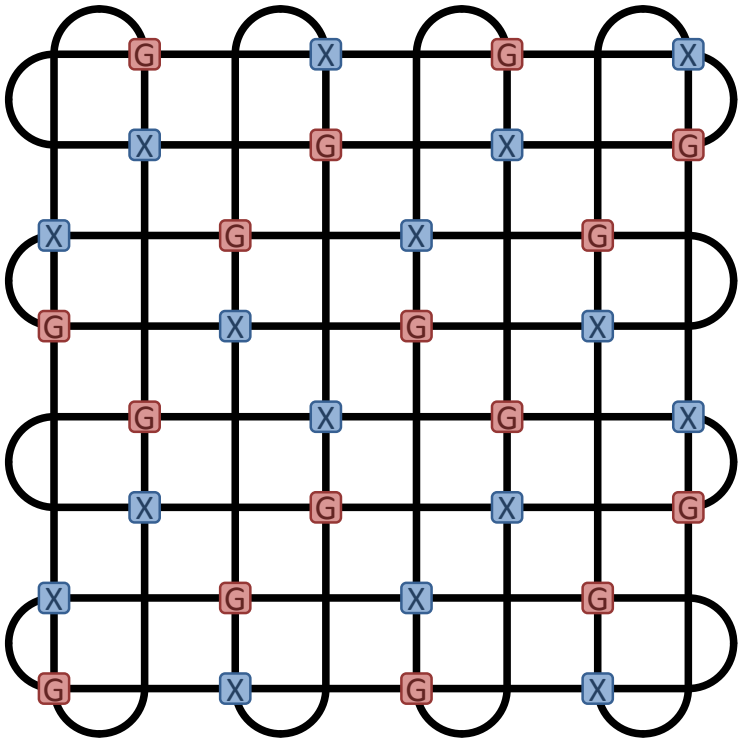

Figure 4.12: 4×4 TorusNX network with 16 access points.

build any size square topology with dimensions equal to any positive integer power of two.

The insertion loss performances of TorusNX and Square Root assuming realistic loss parameters are shown in Fig. 4.15. For the radixes examined, TorusNX has between 23% and 29% lower network-level insertion loss in comparison to the original Torus, while Square Root has between 31% and 46% lower loss. In the case of 8×8 topologies, the Torus contains 3200 waveguide crossings, while TorusNX reduces this number to 1796, and Square Root further reduces it to 1080. As before, improved crossing loss can also be applied to these designs to further improve the scalability and performance (Fig. 4.16 and Fig. 4.17). In

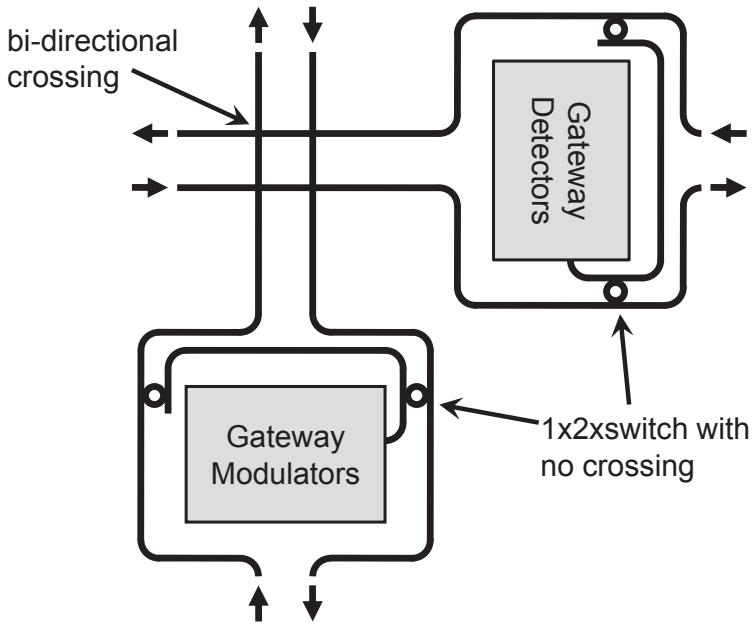

Figure 4.13: Design for a photonic gateway with an integrated bidirectional crossing.

both of the new networks, assuming the same 30-dB optical budget and improved crossing losses, both networks are able to achieve the maximum size network simulated in this study (324 access points for TorusNX, 256 access points for Square Root) and with the remaining optical budget transmit on seven wavelength channels.

The results of this insertion loss analysis clearly indicate that the newly developed networks are better in sustaining higher bandwidths and more access points for better overall system performance. However, for a fixed network design, optical power budget, and device performance, determining the optimal number of wavelengths and access points to use will largely depend on the specific system requirements being targeted. As an example, we can

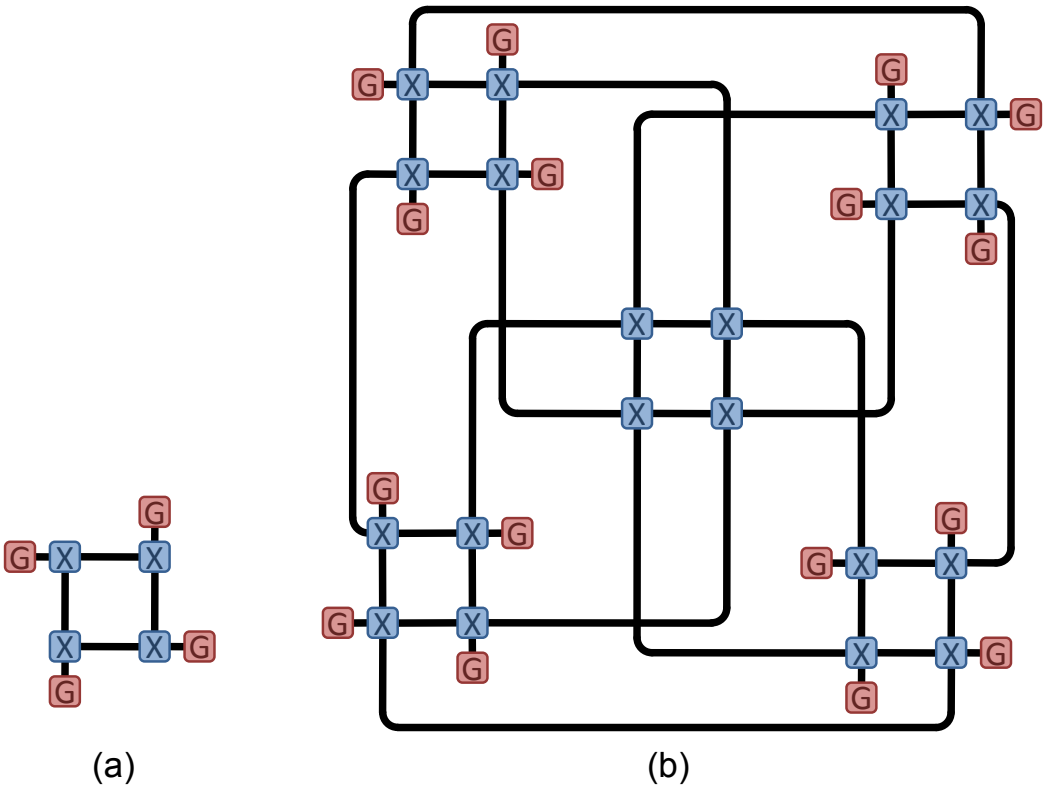

Figure 4.14: (a) The basic unit of the Square Root topology, a 2×2 quad. (b) A 4×4 Square Root.

choose to maximize the total ideal network throughput (number of access points × number of wavelengths per access point × data rate per wavelength) of the TorusNX topology. We assume a 30-dB optical budget, the improved device parameters, and a 10-Gbps modulation rate per wavelength. At one extreme, selecting the maximum number of access points (324) while using a single wavelength achieves a throughput of 22.6 Tbps. On the other hand, maximizing the number of wavelengths (70) would allow a total of 16 access points which

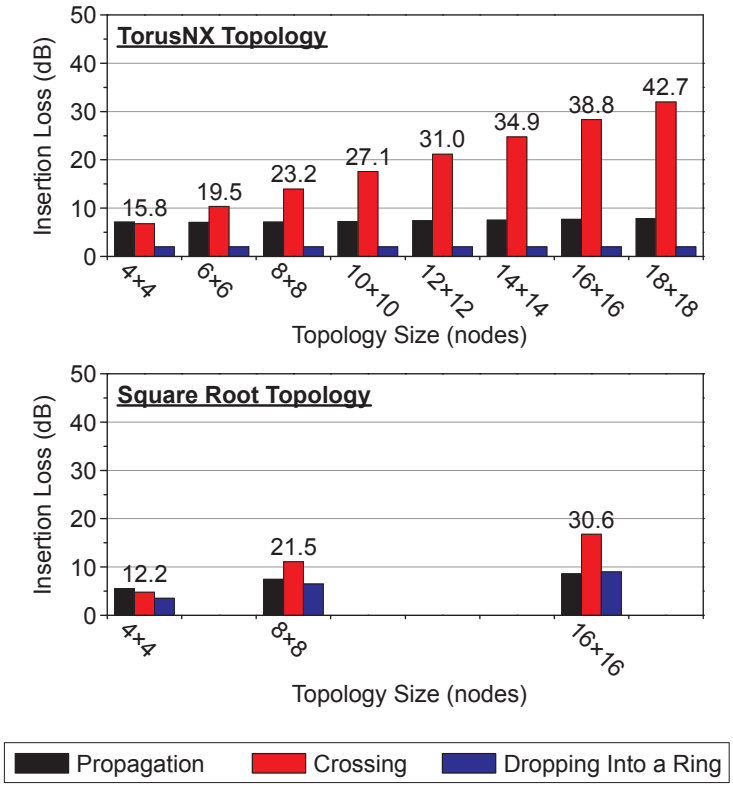

Figure 4.15: Maximum possible network-level insertion loss by component for varying sizes of TorusNX and Square Root using the parameters listed in Table 4.2. Labeled values represent the peak cumulative insertion loss in dB.

results in a throughput of 11.2 Tbps. A balance of the two parameters, in fact, achieves the best throughput performance at 27.4 Tbps when using 196 access points with 14 wavelengths.

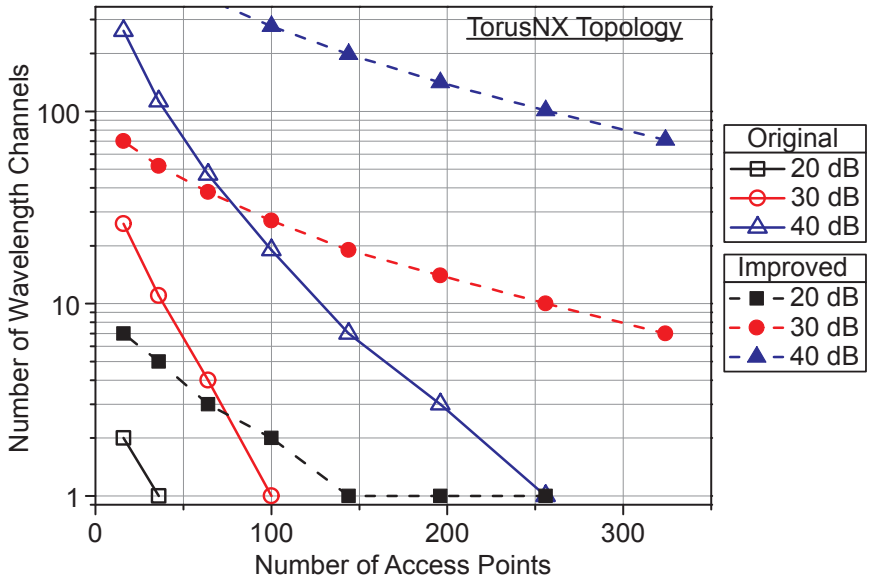

Figure 4.16: Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the TorusNX topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved).

4.3.2 Crosstalk Analysis

For system performance, it is useful to report the SNR, which is a measure of the integrity of the message being transmitted. The signal power is calculated based on the injected power and the network-level insertion loss, while the noise power is derived from the several sources outlined in Section 3.4.2. The crosstalk analysis we report in this analysis assumes non-WDM (single-wavelength) transmission, therefore we set N_{intra} equal to zero. This analysis only considers laser noise and inter-message crosstalk. For this reason the presented results can

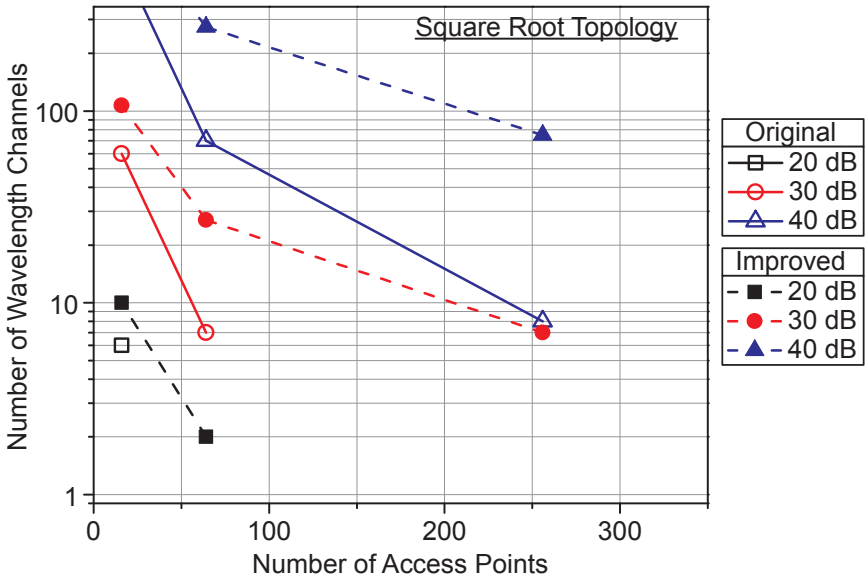

Figure 4.17: Upper limits on the number of wavelength channels allowed for a given number of access points assuming various network-level optical power budgets in the Square Root topology. Solid lines assume all realistic parameters (original) and dashed lines assume a hypothetical improvement in crossing loss (improved).

be thought of as an upper bound in OSNR performance.

Determination of laser noise is based on laser and modulator performance. For continuous-wave quantum cascade laser, RIN has been measured to be about -150 dB Hz^{-1} for a 10-mW output [78]. Silicon ring modulators have been demonstrated with extinction ratios of about 9 dB when modulated at 12.5 Gbps [34]. Poly-silicon ring modulators have also been demonstrated with extinction ratios of 16 dB during DC operation, and 10 dB with active signaling at 2.5 Gbps [88]. From Equation 3.2, we can solve for the laser noise

Table 4.3: Crosstalk and Noise Parameters - Photonic Circuit-Switching Analysis

Parameter	Value	Ref.
Lasers (RIN)	-150 dB/Hz	[78]
Modulation (Modulation Index)	16 dB	[88]
PSEs - Through Port Extinction Ratio	25 dB	[30]
PSEs - Drop Port Extinction Ratio	20 dB	[30]
Crossings (Crosstalk)	-40 dB	[23]

power since the signal power is known.

The crossings and ring switches are the main points for inter-message crosstalk. This leakage has been measured at -40 dB below the signal power [23]. Similarly, the ability of a ring to resonate or pass a particular optical wavelength channel is also non-ideal. A signal that is on resonance with the ring will mostly drop through the ring, however a small portion of the optical power will continue through in the off resonance direction. The same is true in the case of an off resonance signal, which will partially leak onto the on resonance direction. This small fraction of the optical signal can interfere with other propagating messages as more noise. This behavior is characterized by the extinction ratio, which has been measured experimentally to be 28.6 dB for the through port and 18.7 dB for the drop port [30]. All noise related parameters for the crosstalk analysis are listed in Table 4.3.

The OSNR measurements for the four networks are reported in Fig. 4.18 for varying message sizes. Communications on space-routed topologies have varying ratios of photonic activity to electronic activity due to the separate electronic control and photonic planes. Network activity exclusively takes place on the control plane during the provisioning and release stages of a photonic path, therefore no optical signal is injected during these periods. As the transmission message sizes increases, the ratio of photonic to electronic activity increases and is reflected by increased optical crosstalk and lower OSNR. We assume maximal loading of the network with uniform random traffic. Each network assumes an 8×8 topology.

For short messages, the message transmissions are dominated by the electronic control messages, therefore optical transmission is less frequent and crosstalk is less likely. In this limiting case, the OSNR is limited by the laser intensity noise. By solving for Equation 3.2 with the assumed parameters, we get an OSNR of about 47 dB, which corresponds well with the simulation results.

For large messages, the electronic path-setup time is amortized by long data transmissions, and the optical network becomes saturated with the long optical messages. In this case, inter-message crosstalk is highly likely to occur, causing more significant signal degradation. The Square Root topology performs the best for large messages with an OSNR of about 16.0 dB. Torus, Non-blocking Torus, and TorusNX achieve OSNRs of 11.3 dB, 13.2 dB, and 12.2 dB,

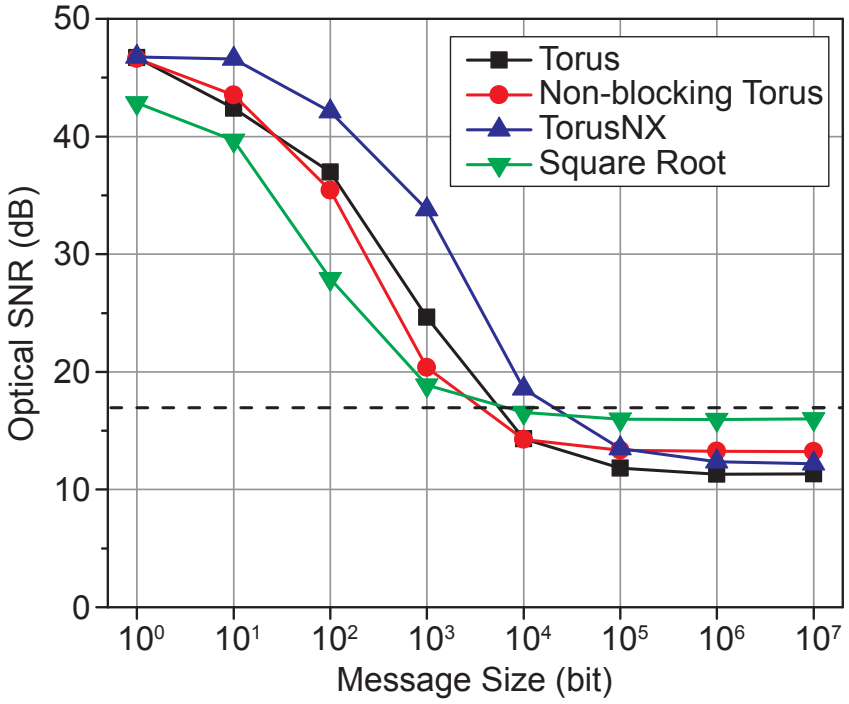

Figure 4.18: Optical SNR performance for varying message sizes assuming saturated network load, measured at the photodetectors. The line at OSNR=16.9 dB is where a bit-error-rate of 10^{-12} can be achieved, assuming an ideal binary receiver circuit and orthogonal signaling.

respectively.

Lack of signal integrity ultimately results in erroneous bits detected. If we assume orthogonal signaling, and an ideal optimal binary receiver, we can calculate the BER using the following Q function [89]:

$$BER = Q\left(\frac{E_b}{\mathcal{N}}\right) \tag{4.1}$$

E_b is the energy in each bit, and $\mathcal{N}$ is the power spectral density of the noise. The term inside

the radical is equivalent to the SNR of the signal. For a BER of 10^{-12}, the network requires a SNR of 16.9 dB (indicated in Fig. 4.18 by a horizontal line). This indicates that in the large-message cases, none of the networks are able to achieve this level of signal integrity. The achieved BERs for networks using 10^7-bit messages are 1.14×10^{-4} for the Torus, 2.20×10^{-6} for Non-blocking Torus, 2.36×10^{-5} for TorusNX, and 1.31×10^{-10} for Square Root. The high BERs can be lowered by using smaller messages, or mitigated through the use of a higher network-layer error correction scheme.

4.3.3 Power Analysis

The network-level power dissipation is a major component in limiting performance scaling of chip-scale systems. Photonic on-chip networks have been shown to drastically outperform electronic networks in both performance and energy, especially in the case of traffic patterns that require large data transmissions [52]. We conduct simulations to examine the dissipation of the four photonic networks.

Each network is assumed to use the maximum number of wavelengths allowed for the improved 8×8 topology assuming a 30-dB optical power budget according to the results in Fig. 4.8, Fig. 4.9, Fig. 4.16, and Fig. 4.17. The simulator uses the ORION model [72] for electronic router energy dissipation, which is configured for a 32-nm process with a normal

voltage threshold transistor type and a V_{dd} equal to 1.0 V. The electronic components in the network are clocked at 1.0 GHz. All electronic routers use a standard three-stage pipeline model with an 128-bit buffer on each input port and a flit-size of 32 bits. All control messages are 32 bits in size. The routers in the torus-like networks use dimension-ordered routing while Square Root uses a unique routing scheme that is optimized to equally distribute load and reduce propagation distance. All routers are modeled with credit-based flow control.

Table 4.4: Energy Dissipation Parameters - Photonic Circuit-Switching Analysis

Parameter	Value
Lasers (RIN)	-150 dB/Hz
Modulation (Modulation Index)	16 dB
PSEs - Through Port Extinction Ratio	25 dB
PSEs - Drop Port Extinction Ratio	20 dB
Crossings (Crosstalk)	-40 dB

The simulations assume integrated thermal tuners to manage thermal fluctuations in a chip, which will be strongly dependent on application activity. Thermal tuners integrated at each ring in the network assume approximately 1 μW/$^\circ$ K of power dissipation, while the system is assumed to have a mean temperature deviation of 20 degrees. Modulators assume a dynamic dissipation of 85 fJ for every bit transmitted (bit edges) and an additional 30 μW

of static power during periods when a constant signal is transmitted (hold periods). Switches exhibit higher dynamic and static dissipation than the ring modulators, at 375 fJ/bit and 400 μW, respectively, due to larger footprints. Photodetector energy is assumed to be 50 fJ/bit. The photonic power dissipation parameters used in this set of simulations are listed in Table 4.4.

The power performance is reported for each of the four networks, and assumes maximum loading with uniform random traffic on 8×8 topologies (Fig. 4.19, Fig. 4.20, Fig. 4.21, and Fig. 4.22). In all four network designs, the electronic buffers, crossbar circuit, and clock tree dissipate a clear majority of the network power. This is a clear indication that electronic power will remain as a relatively significant contributor to total network power dissipation even with photonic integration.

Additional notable trends can be reasoned by relating the power dissipated to the exhibited bandwidth performance of the networks. From Fig. 4.23 we can see the total network performance of the four networks. As the network assumes larger message sizes, the network throughput also rises due to the amortization of the circuit-switching overhead. Congestion of optical traffic on the photonic network plane causes the eventual saturation of the networks. TorusNX achieves the best network bandwidth at 7.80 Tbps, while Square Root, Torus, and Non-blocking Torus obtain throughputs of 3.75 Tbps, 2.45 Tbps, and 669

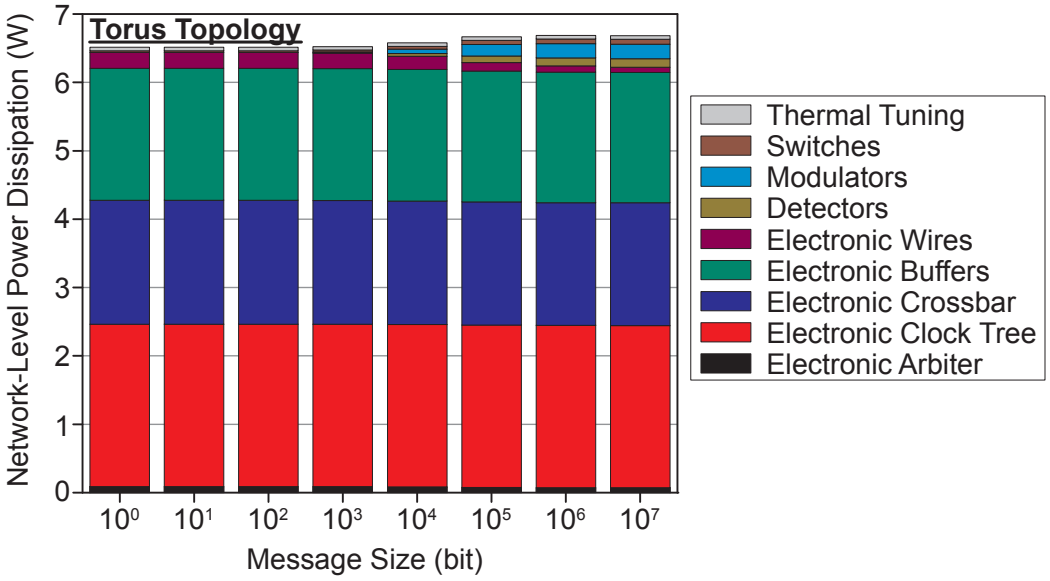

Figure 4.19: Power-dissipation breakdown of an 8×8 Torus topology over varying message sizes.

Gbps, respectively.

Relating back to the four power dissipation figures, we see that as the network achieves higher throughput with larger messages, the ratios in power dissipation shifts from high amounts of wire power dissipation and low photonic device power dissipation to low wire power dissipation and high photonic device power dissipation. This is evidence of the higher photonic network utilization and amortization of the electronic path-setup overhead. Furthermore, the total power dissipated by the electronic components in the network remains approximately constant regardless of network throughput since all the data is being sent optically.

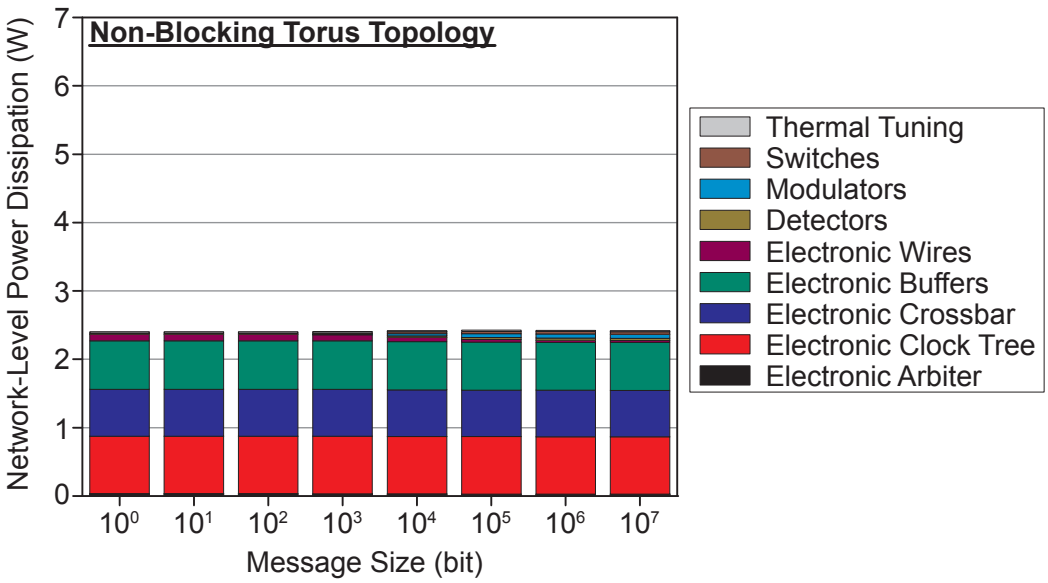

Figure 4.20: Power-dissipation breakdown of an 8×8 Non-blocking Torus topology over varying message sizes.

Fig. 4.24 combines the power and bandwidth results to plot the energy-per-bit efficiency of the networks. For the largest message size, TorusNX and Square Root achieve the best efficiencies at 585 fJ/bit and 681 fJ/bit. Torus achieves an efficiency of 2.73 pJ/bit, while Non-blocking Torus achieves an efficiency of 3.62 pJ/bit. The new network designs attain at least 75% better efficiency compared to the Torus, and at least 81% better efficiency than the Non-blocking Torus. This dramatic improvement is attributed to the lower-loss network designs which enable better bandwidth utilization and reductions in the number of required switches.

We see that although the Non-blocking Torus produces a comparatively reasonable

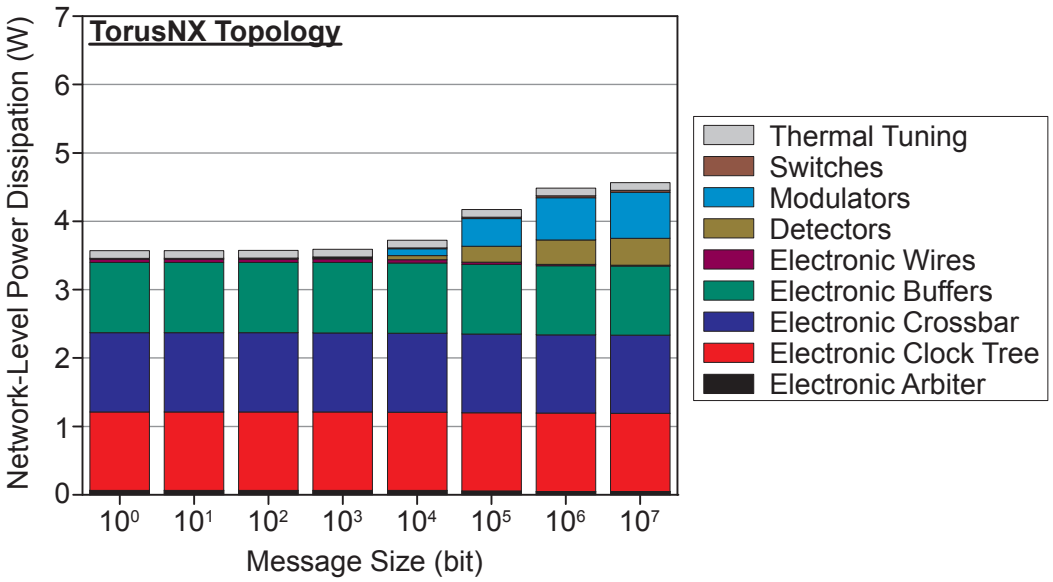

Figure 4.21: Power-dissipation breakdown of an 8×8 TorusNX topology over varying message sizes.

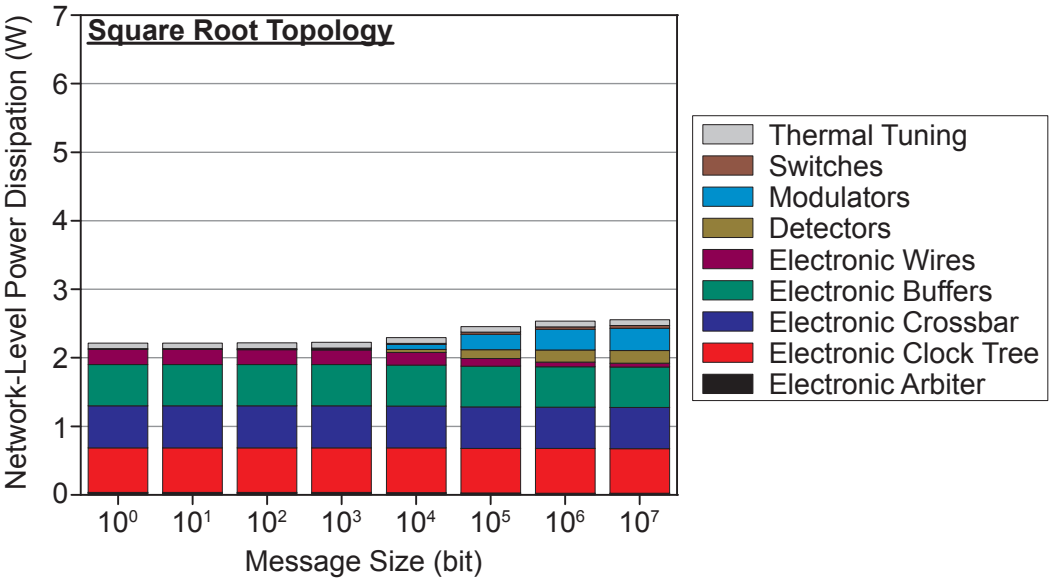

Figure 4.22: Power-dissipation breakdown of an 8×8 Square Root topology over varying message sizes.

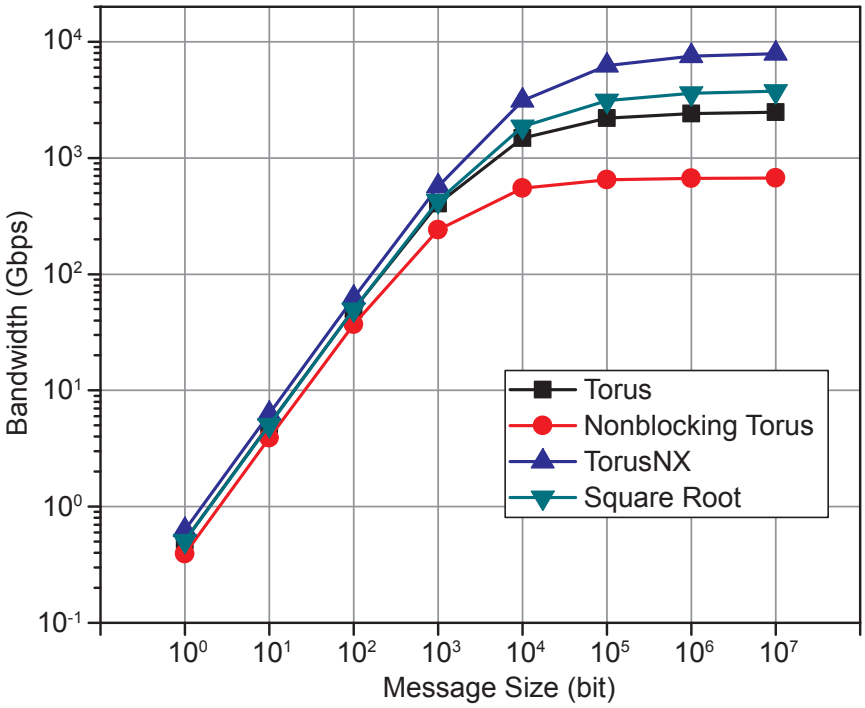

Figure 4.23: Total network bandwidth of each network at saturation.

absolute power dissipation measurement, the efficiency, for larger message sizes, is the worst of the four networks. Although the Non-blocking Torus has the advantage of being non-blocking, the fact that it supports fewer access points than the other three network designs results in a dramatic degradation in performance. Note that each network assumes the same topology size, however the Non-blocking Torus only uses 16 nodes due to the layout constraints. While it may seem reasonable to assume a 32×32 Non-blocking Torus so that each network can be normalized to the number of gateways, we can see from our original conclusions in Fig. 4 that a 64-gateway version is not possible. The insertion loss penalties

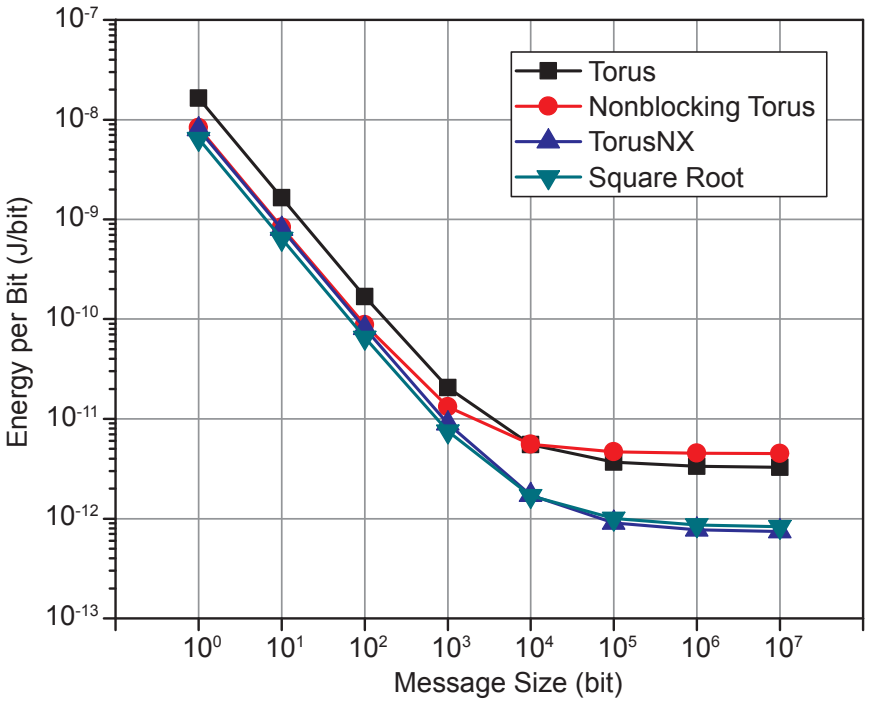

Figure 4.24: Total network bandwidth of each network at saturation.

usurp the benefits of the non-blocking property, resulting in bandwidth degradation.

While from an efficiency standpoint, larger message transmissions clearly perform better, the prior crosstalk simulations indicate that the OSNR also decreases with increased message size. This indicates that in order to maintain the high energy efficiency that these photonic topologies can provide, a scheme must be in place to either correct or mitigate these errors.

4.4 Comparative Analysis of Photonic Spatial Routing and Wavelength Routing

In this section, we model and compare two different photonic routing formats to demonstrate our design methodology and the versatility of the physical-layer analysis capabilities of PhoenixSim. The comparison will be between a spatial-routed network and a wavelength-routed network. We will show that the two networks offer different advantages depending on the considered metric and traffic pattern. Therefore this analysis serves to give system architects recommendations based on their design objectives.

The first photonic network we model for this case study is the Photonic Mesh which uses the circuit switching protocol described in Section 4.1. The Photonic Mesh (Fig. 4.25a) is similar to a typical electronic mesh since it is laid out in a matrix-like configuration of nodes, and has mechanisms for switching, entering the network, and exiting the network at each node. Although the mesh-based design presented here exhibits lower path diversity than other previously proposed circuit-switching topologies ([53, 90]), the simpler architecture is beneficial to overall performance by lowering total insertion loss.

A 4×4 nonblocking crossbar switch (Fig. 4.25b) is found at each node of the network and is optimized for dimension-ordered routing (which is the case for the Photonic Mesh) by

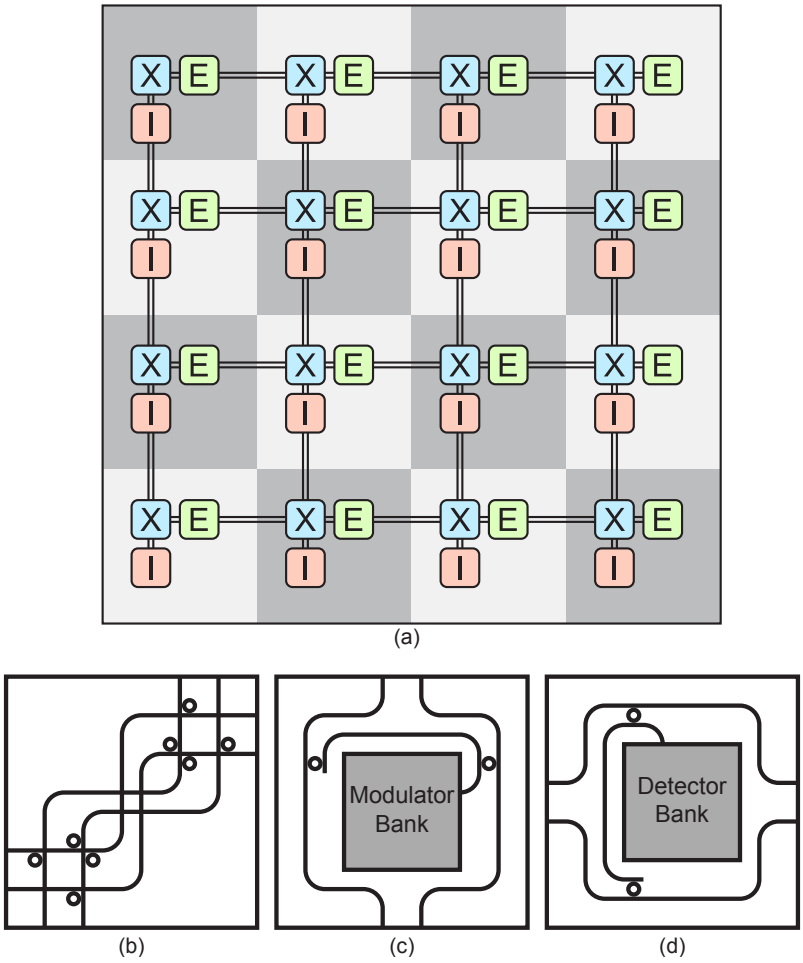

Figure 4.25: The Photonic Mesh topology. (a) A high-level representation of a 4×4 Photonic Mesh. Parallel lines indicate two unidirectional waveguides, which are paired together to form bidirectional links. Boxes represent higher-order photonic components, which are labeled 'X' for 4×4 nonblocking crossbar switch, 'I' for injection gateway, and 'E' for ejection gateway. Also shown are detail schematics of the (b) 4×4 nonblocking crossbar switch, (c) injection gateway, and (d) ejection gateway.

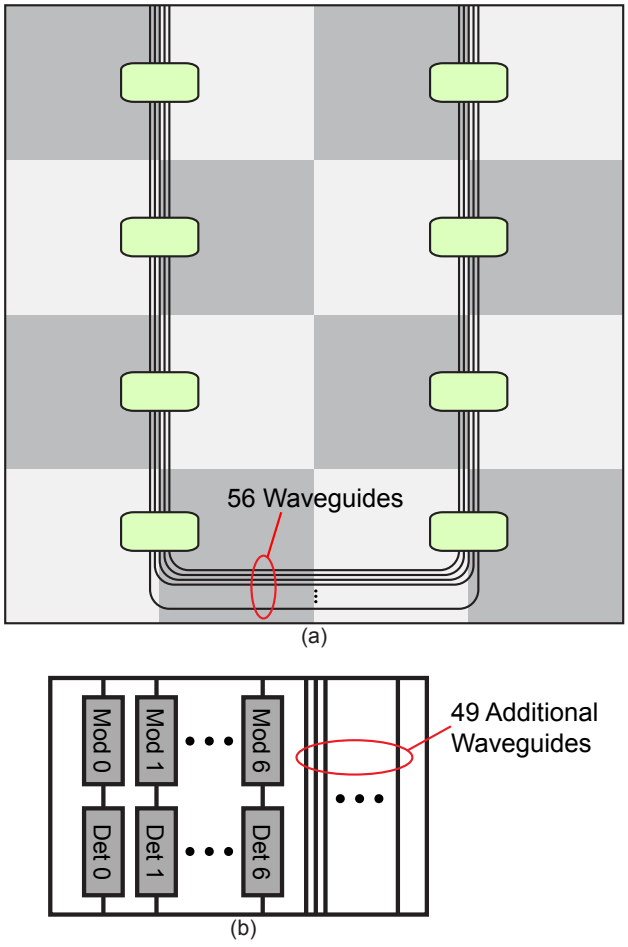

Figure 4.26: The Photonic Crossbar topology. (a) A high-level representation of a 2×4 Photonic Mesh, connecting 16 cores. Boxes represent gateways with a concentration of two processing cores. (b) A detail schematic of the Photonic Crossbar gateway, showing 49 bypass waveguides and 7 waveguides with modulator and receiver banks used to communicate to the other 7 gateways.

minimizing insertion losses along propagation paths that do not turn through the switch [91]. The injection gateway (Fig. 4.25c) and ejection gateway (Fig. 4.25d) designs, which are used by the underlying processing cores to transmit and receive optical data, are adapted from the TorusNX topology to help further reduce insertion loss overhead caused by more complex injection/ejection schemes [53]. Each switch and gateway is constructed using the devices previously described in Section 3.3.

The second photonic network we model for this case study is the Photonic Crossbar (Fig. 4.26). This design uses the crossbar concepts used previously in the Photonic Clos topology [92]. A set of waveguides are routed in a serpentine manner so that they intersect with all gateways in the network. Each individual waveguide is configured with two modulator banks and two receiver banks to connect a unique pair of gateways. For a topology with G gateways, a set of $G \cdot (G-1)/2$ waveguides is required to fully connect the network. Since the required number of waveguides grows quadratically with G, it can be advantageous to concentrate the traffic of a set of processing cores through a single photonic gateway. Each gateway exploits the bidirectionality of the waveguides and avoids receiving its own modulated signal by transmitting and receiving on different sets of wavelengths.

Fig. 4.26a shows an 8-gateway network with two processing cores connected to each gateway. The gateway design is illustrated in Fig. 4.26b. The gateway contains 49 bypass

waveguides which are ignored, and is connected to the remaining seven waveguides through a set of seven modulator banks and seven receiver banks. Each connected waveguide will transmit to and receive from one of the other seven gateways in the network. Attached to each photonic gateway is a nine-port electronic router which must transport messages to and from the group of cores to the appropriate photonic transmitter or receiver bank.

4.4.1 Optical Power Budget Analysis

First, we used PhoenixSim to model both photonic topologies and analyze the worst-case insertion loss for network radixes from 2×2 (4 nodes) to 10×10 (100 nodes). The insertion loss parameters used in this study are derived from experimentally demonstrated results and are listed in Table 4.5. Fig. 4.27 shows the maximum total loss exhibited within each network and the breakdown according to type of loss. All network sizes assumed total chip dimensions of 2 cm × 2 cm and the size of the network is designed to span the entire chip. Hence the spacing between nodes will decrease with larger radixes. Crossing loss and propagation loss are the most significant contributors to total loss in the Photonic Mesh and Photonic Crossbar, respectively. The 10×10 Photonic Mesh has 18.1 dB of crossing loss caused by the existence of a network path with 113 waveguide crossings, accounting for approximately 63% of the total network-level insertion loss. The serpentine waveguide

Table 4.5: Insertion Loss Parameters - PhoenixSim Case Study

Parameter	Value	Ref.
Propagation Loss (Silicon)	1.7 dB/cm	[19]
Waveguide Crossing	0.16 dB	[23]
Waveguide Bend	0.005 dB/90°	[19]
Drop Into a Ring	0.6 dB	[30]
Pass By a Ring	0.005 dB	[30]

design of the Photonic Crossbar causes repeated traversals of the chip, therefore causing high propagation loss. This analysis is an important indicator for device researchers who may seek to focus on improving the performance of a specific type of network architecture.

By taking the insertion loss results and applying Eq. (3.1), we can derive the allowed number of wavelength channels for varying radixes and optical power budgets (Fig. 4.28). For the specified optical power budgets, points below and to the left of the plotted curve indicate physically realizable combinations of network size and number of wavelength channels. For example, both networks are realizable as a 4×4 network using 32 wavelength channels with devices that stay above a 30-dB optical budget, however the fabrication of an 8×8 network with 32 wavelength channels and a more aggressive 40-dB budget will only be possible for the Photonic Mesh. Furthermore, the plot indicates that the Photonic Crossbar is not capable

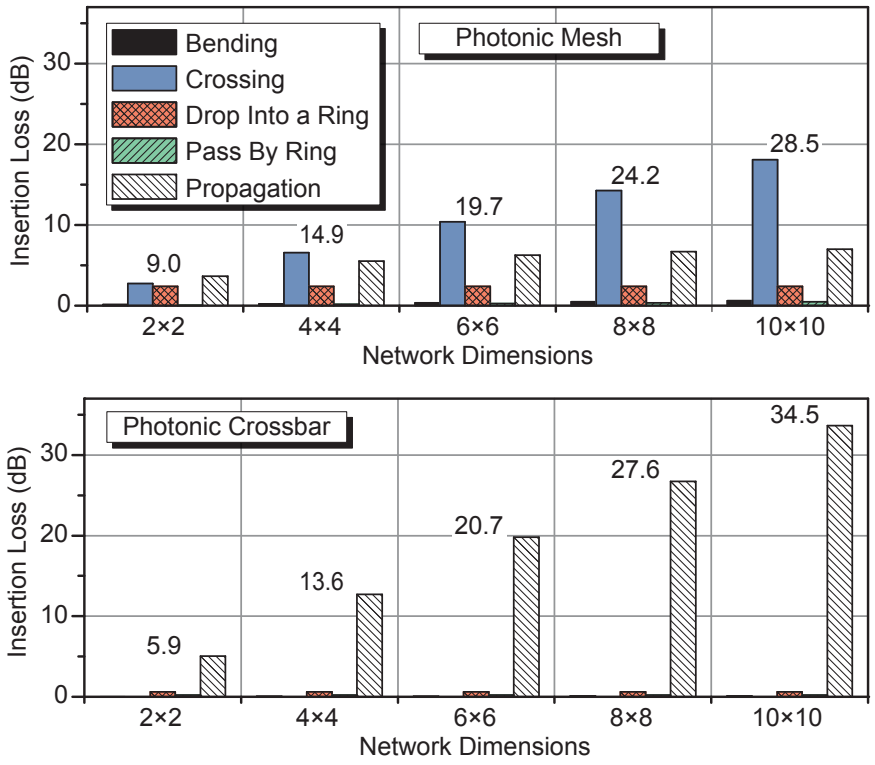

Figure 4.27: Insertion loss results for the Photonic Mesh and Photonic Crossbar of varying sizes. Labeled values that overlay the columns indicate the worst-case total network-level loss values. Columns illustrate the worst-case loss associated with the individually labeled loss component which does not necessarily occur in the network path with the worst total loss.

of operating at sizes of 10×10 or greater with a 30-dB optical power budget.

4.4.2 Network Performance

The performance and power dissipation of the on-chip network are both important considerations for future scaling of CMPs. For this analysis, we assume a 64-core processor

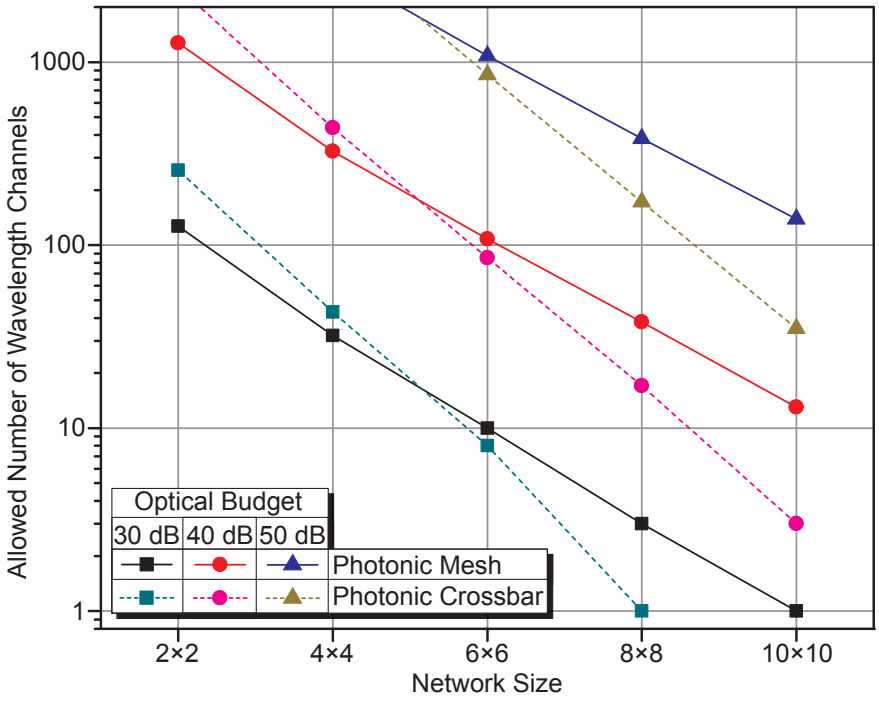

Figure 4.28: Wavelength channel allotment in the Photonic Mesh and Photonic Crossbar for varying network sizes and optical power budgets.

and compare the performance of the Photonic Mesh, the Photonic Crossbar, and a traditional electronic mesh. In each of the three networks, we assume a 2.5-GHz operating frequency for both electrical and optical signaling. Both photonic networks assume the use of 128-wavelength channels (the Photonic Crossbar will have two bi-directional 64-wavelength channels). Electronic routers for the Photonic Mesh are modeled with a 32-bit channel width and buffer size of 128 bit, which equates to a buffer depth of four control messages. Electronic routers for the Photonic Crossbar and electronic mesh have a 64-bit channel width

and a 1024-bit buffer size. Additionally, for the Photonic Crossbar we assume a concentration of eight cores per gateway. All simulations are based on uniform random traffic.

Fig. 4.29 plots the network-level bandwidth and latency of the three networks under consideration. For 1-kbit messages, we see that the Photonic Crossbar exhibits the highest throughput. The Photonic Mesh performs the worst as a result of the costly overhead associated with circuit switching. In the case of 100-kbit messages, the Photonic Mesh now achieves the best performance since the latency overhead of circuit switching is now amortized over the duration of the message transmission. This indicates that the most suitable network design will be dependent on the type of traffic exhibited by the system.

4.4.3 Data Integrity Analysis

Whereas the insertion loss has an impact on physical size and bandwidth of the network, the noise has an impact on the quality of the data stream. Given the same network configuration used in the Network Performance results, the average noise power for each wavelength channel for all WDM transmissions under saturated random-traffic load is plotted in Fig. 4.30. These noise power results are based on the crosstalk and noise parameters listed in Table 4.6. In this network, laser intensity noise, thermal noise, and shot noise are negligible quantities in comparison to inter-message and intra-message crosstalk.

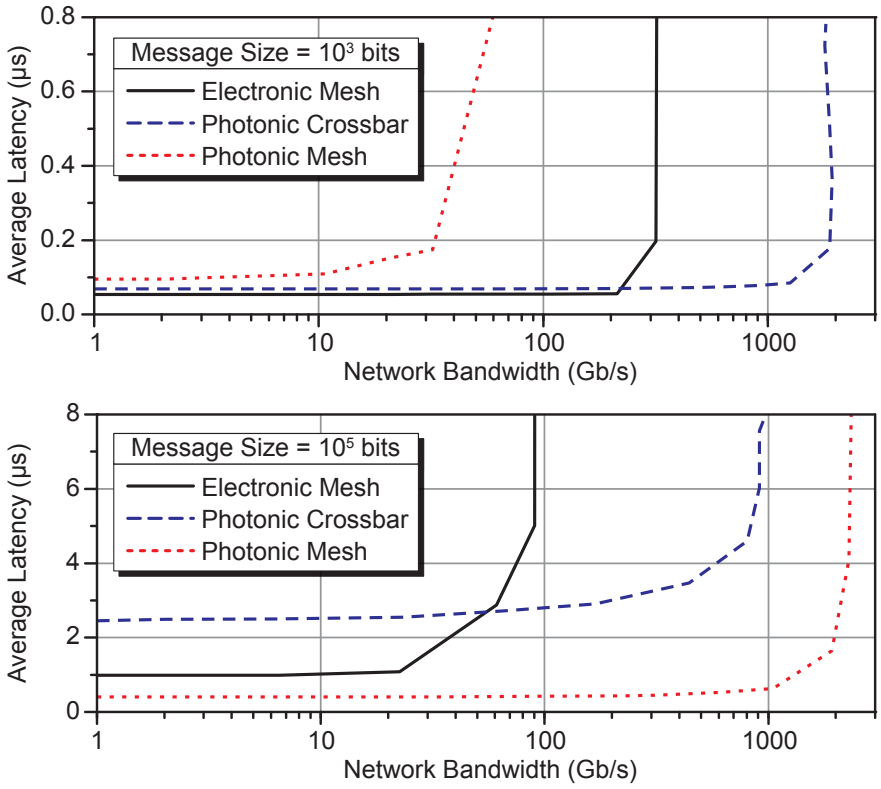

Figure 4.29: Bandwidth and latency performance on the Electronic Mesh, Photonic Crossbar, and Photonic Mesh for 1-kbit and 100-kbit message sizes.

In both networks intra-message crosstalk predominately occurs at the ejection gateway where filters are used to select individual wavelength channels. The amount of intra-message crosstalk power exhibited by each optical message is predominately dependent on the number of co-propagating wavelengths. Therefore it is practically independent of both network load and message size. We see that across the two different message sizes, the amount of intra-message crosstalk power remains approximately constant.

Table 4.6: Crosstalk and Noise Parameters - PhoenixSim Case Study

Parameter	Value	Ref.
Laser (Relative Intensity Noise)	-150 dB/Hz	[78]
Modulator (Extinction Ratio)	16 dB	[88]
PSE Through-Port (Extinction Ratio)	25 dB	[30]
PSE Drop-Port (Extinction Ratio)	20 dB	[30]
Waveguide Crossing (Crosstalk)	-40 dB	[23]

The trend in inter-message crosstalk reflects the probability that two WDM messages will intersect in the network. The Photonic Crossbar exhibits zero inter-message crosstalk since it contains no crossings or switches where a message intersection could occur. A longer duration optical packet from using fewer wavelength channels or large message sizes will create a scenario where the photonic message will occupy the network for a longer period of time, thereby increasing the likelihood that another message will be instanced in the network and interfere. In Fig. 4.30, we can see that indeed larger messages in the Photonic Mesh do produce a non-negligible amount of inter-message crosstalk.

Lastly, PhoenixSim also determines the signals SNR when the message is finally received. For 1-kbit message sizes, the average electrical SNR of the Photonic Mesh and Photonic Crossbar optical link is 6.4 dB and 3.5 dB, respectively. For 100-kbit message sizes, the

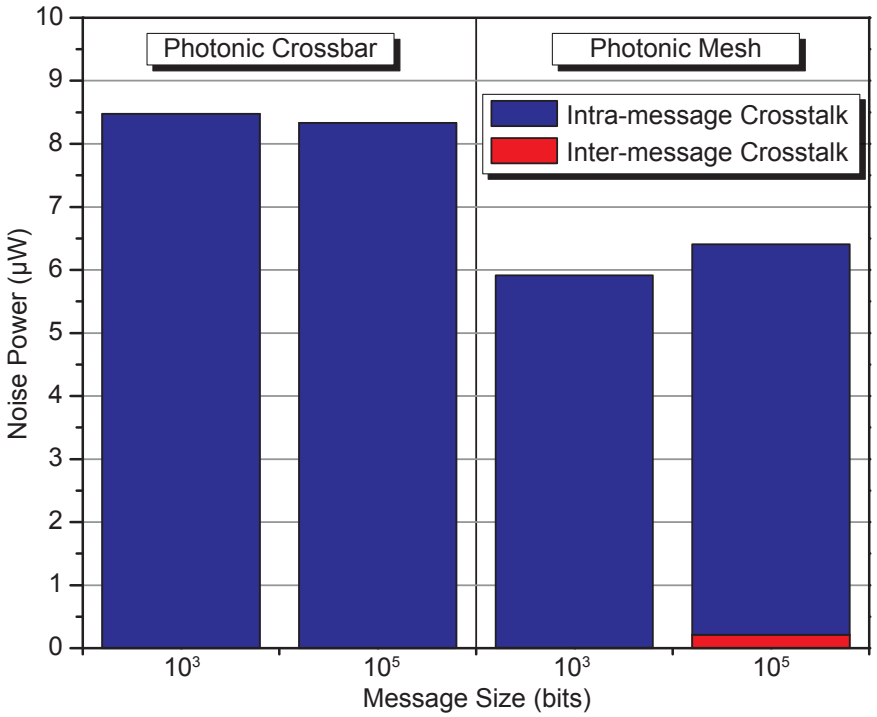

Figure 4.30: Average total noise power accumulated by each optical message in the Photonic Mesh and Photonic Crossbar for saturated network load. Laser noise, thermal noise, and shot noises are negligible quantities and are not listed.

average SNR for the Photonic Mesh and Photonic Crossbar is 6.5 dB and 2.9 dB, respectively. These results indicate that the Photonic Crossbar relatively outperforms the Photonic Mesh with respect to signal integrity. However these values also conclude that the optical link integrity of both networks will be detrimentally compromised. This performance penalty can be rectified by improved filter performance or through the use of fewer wavelength channels.

4.4 Comparative Analysis of Photonic Spatial Routing and Wavelength Routing

The zero-load plot in Fig. 4.30 shows the inter-message crosstalk noise power trend for varying number of wavelength channels used. The noise power exhibits a slight dependence on the number of wavelength channels. Up to about 20 wavelengths, there is an increase in intra-message crosstalk power due to the increasing number of adjacent wavelength channels that can leak through each filter. Above 20 wavelength channels, the intra-message crosstalk power decreases due to each wavelength channel only using a smaller fraction of the total allowed power according to the optical power budget.

Inter-message crosstalk does not occur in the zero-load case since the probability of two messages intersecting in the network is relatively low. In contrast, the noise power of a saturated network (Fig. 4.30) shows significant inter-message crosstalk power indicating that it is dependent on network load, as well as message size and the number of wavelength channels being used. The trend in inter-message crosstalk reflects the probability that two WDM messages will intersect in the network. With a high network load, there will be more messages in the network at any single point in time. The duration of the optical packet will be shorter when more wavelength channels are used or shorter messages are transmitted, thereby creating a smaller temporal opportunity for another message to interfere since the transmission will occur relatively quickly. A longer duration optical packet from using fewer wavelength channels or large message sizes will create a scenario where the photonic message

will occupy the network for a longer period of time, thereby increasing the likelihood that another message will be instanced in the network and interfere.

4.4.4 Power Dissipation Analysis

Lastly, we compare the power dissipation of the Electronic Mesh, Photonic Mesh, and Photonic Crossbar, assuming the same system configuration as before and the power parameters listed in Table 4.7. The total power dissipation of each network, operating with maximum load, is plotted in Fig. 4.31. Each column is broken down into categories of photonic-related dissipation from ring modulators, photo-detectors, optical switches, and thermal feedback tuning, and electronic-related dissipation from router logic, router buffers, and wires. While SerDes would be required in many proposed photonic interconnect architectures for every ring modulator and photo-detector to up and down convert to the photonic transmission clock, in this case study we assume the same 2.5-GHz clock for both electronic and photonic domains.

Regardless of the message size the Electronic Mesh dissipates approximately 8 W of power and the Photonic Mesh dissipates approximately 5 W. This is a result of both networks relying on some electronic routers to route data. Data on the Photonic Mesh is only transmitted optically, which provides a significant savings in power when the circuit-

Table 4.7: Power Dissipation Parameters - PhoenixSim Case Study

Parameter	Value
Modulators (Dynamic Energy)	85 fJ/bit
Modulators (Static Energy)	30 μW
Photodetectors	50 fJ/bit
PSEs (Dynamic Energy)	375 fJ/bit
PSEs (Static Energy)	400 μW
Thermal Ring Tuning	100 μW/ring

switching overhead can be amortized. In terms of energy efficiency when transmitting 1-kbit messages, the Photonic Crossbar outperforms the other networks at 2.9 pJ/bit, while the Photonic Mesh performs the worst at 55.9 fJ/bit. Nonetheless, with the larger 100-kbit messages, the Photonic Mesh achieves the highest efficiency with 3.2 pJ/bit as a result of the efficient optical transmission. This message-size/efficiency relationship of the circuit-switched Photonic Mesh design is a useful indicator as to which photonic design may be ideally suited for various application traffic patterns. For instance, photonic circuits have been shown to be ideally suited for many classes of scientific applications that require long data messages [52].

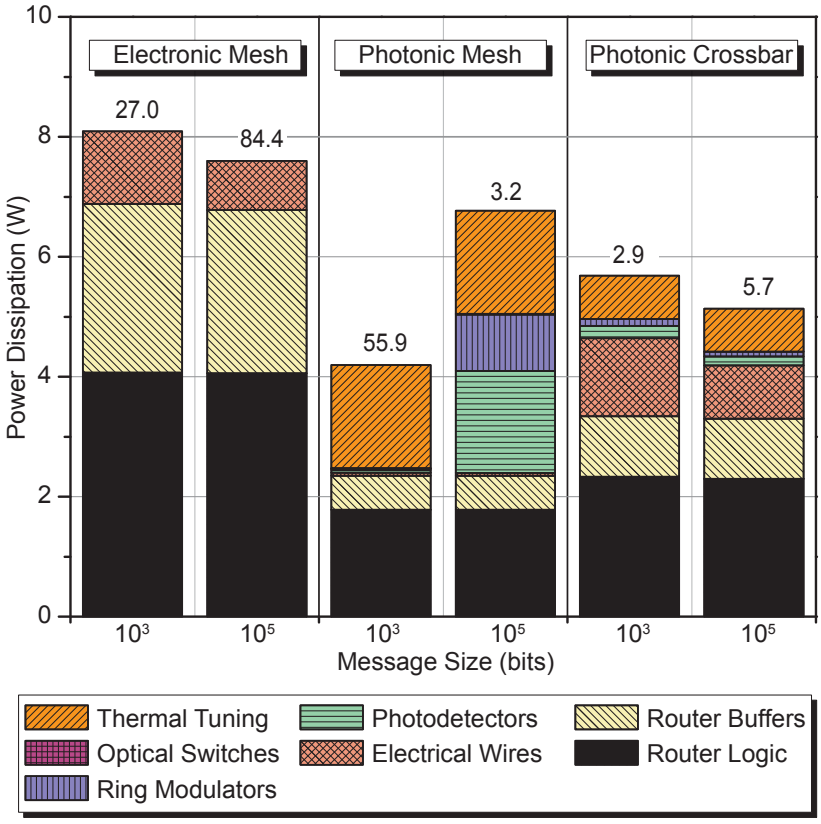

Figure 4.31: Network-level power dissipation breakdown of the Electronic Mesh, Photonic Mesh and Photonic Crossbar for transmission of 1-kbit and 100-kbit messages. Values overlaying each column indicate the energy efficiency of the network in units of pJ/bit.

Chapter 5

Wavelength-Selective Spatial Routing

The photonic interconnection networks that have been discussed thus far have exclusively looked at wavelength-routed architectures and spatial circuit-switching-style architectures. Previous research into photonic chip-scale networks have also exclusively focused on these two domains. Instead of developing alternative topologies for already proposed routing architectures, this chapter discusses the development of new switching methodology.

In this chapter, we describe a novel on-chip photonic interconnect architecture that leverages a new concept known as *wavelength-selective spatial routing* (WSSR) to increase path diversity within previously proposed circuit-switched photonic networks for CMPs [56]. Previous circuit-switched photonic network designs can suffer from longer latencies and degraded bandwidth performance due to low path diversity and high contention probability

caused by a fundamental architectural constraint that limits each physical optical path to a single communication link at any one point in time. Traditional networks can leverage electronic virtual channels to statistically multiplex several logical links through a single physical electronic bus. However, the standard virtual channel technique requires buffering and processing which are not economically feasible in the photonic domain. We alternatively propose the use of WSSR which uses *spectral multiplexing* to create several concurrent communication links with a single waveguide. Compared to traditional circuit-switched photonic networks, WSSR-enabled networks can achieve superior bandwidth performance with a minimal increase in design complexity and latency.

5.1 Concept

WSSR is used to selectively manipulate wavelength-channel subsets of a WDM signal as it propagates through a network [55]. WSSR can be qualified as a hybrid form of spatial routing and wavelength routing (Fig. 2.3). The WSSR scheme takes advantage of the unused spectrum that exists between the resonances of a broadband ring switch by interleaving additional wavelength channels in the unused spectral space. The newly interleaved channels can then be used to provide additional paths of communication in the network to increase overall network performance.

Fig. 5.1a illustrates the inclusion of an additional set of three wavelength channels, interleaved amongst the original set of wavelengths that were shown in Fig. 2.2e. Each grouping of three wavelengths, which composes a subset of the total set of wavelengths in the WDM system, is referred to as a *WDM partition*. The newly included partition will propagate past the ring resonator undisturbed, regardless of whether a voltage bias is being applied or not. The technique of isolating a single WDM partition for switching while ignoring the other remaining wavelengths is referred to as wavelength-selective spatial routing. Moreover, a second ring resonator can be cascaded and designed to aligned to the new set of wavelength channels forming a two-partition router. Introduction of the additional cascaded ring will in the worst case increase the insertion loss by only the through-port loss of a ring switch which has been measured to be negligible [30]. Fig. 5.1b–5.1e show the four possible routing configurations of the 2-partition router, illustrating the independent controllability of each WDM partition.

Notice that the previous example augmented the original ring with a second ring resonator of the same diameter. This produces a wavelength channel spectrum that is effectively twice as dense as that of the original case, however it ignores possible crosstalk consequences from placing wavelength channels closer together. This example also produces a more complex gateway since it requires a doubling of the number of modulator and detector elements at

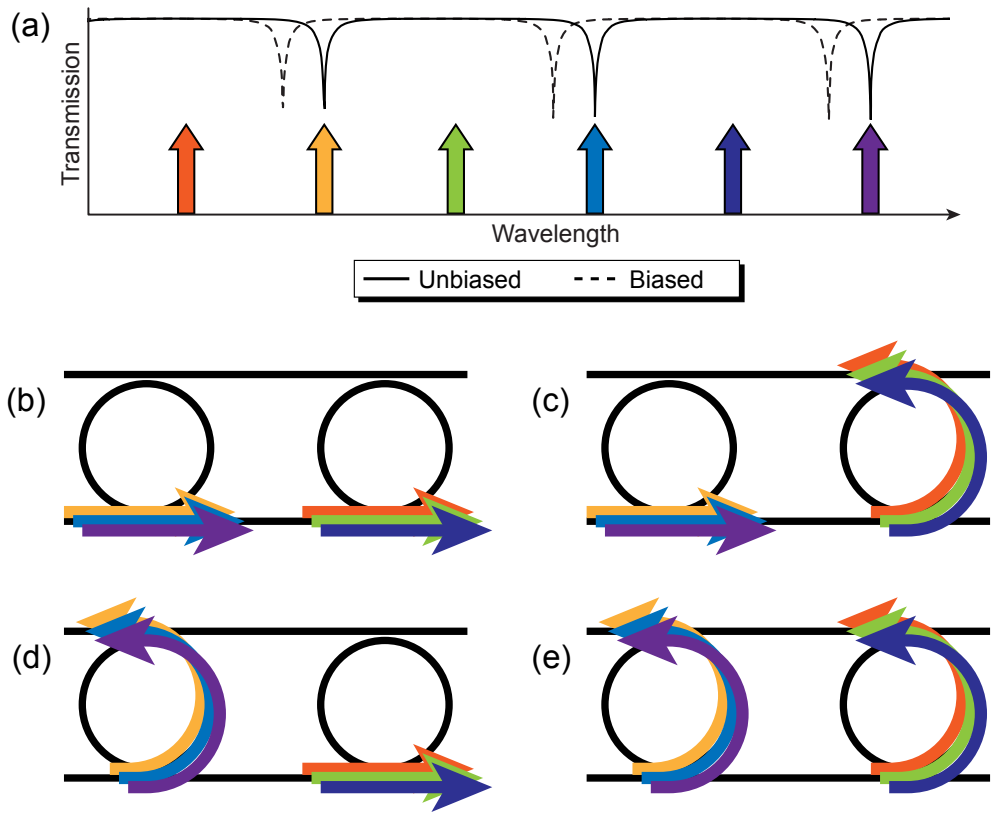

Figure 5.1: (a) Spectral placement of two WDM partitions (each containing three wavelength channels), with respect to the spectrum of an electro-optic broadband ring switch. (b)–(e) Four possible routing configurations for a two-partition router.

each node. Alternatively, the number of wavelength channels can be fixed to preserve the wavelength channel density and the rings can be designed to operate on a subset of the original wavelength channel set through an alteration of the FSR of the ring. A ring with half the diameter of the original will exhibit an FSR that is twice as wide and allow it to operate on half the original set of wavelength channels. This relationship between the

number of partitions (and thus the number of rings) and the area footprint of the router is explored in Section 5.3.2.

We consider in this paper a range of ring diameters that have been experimentally verified to operate as switches. Preston, *et al.* found that the a minimum wavelength channel spacing of 0.8 nm was required for ring modulated 10-Gb/s wavelength channels to maintaining sufficiently low crosstalk levels (< -20 dB) [38]. This corresponds to a 200-μm-diameter ring switch, which has been demonstrated previously with an adequately wide passband for transmitting the high-speed datarate [30].

We can reasonably assume that smaller diameter ring switches can also be produced due to the fact that reductions in ring circumference will only reduce the circulating loss in the ring. It is possible that the reduced loss will increase the Q factor to a point where the drop port resonance becomes too narrow to pass the high speed data signal. This can be remedied by inducing additional insertion loss with fabricated defects or additional doping. The smallest demonstrated ring resonator device we consider has a diameter of 3 μm due to the dominance of bending losses [93]. The most number of WDM partitions we consider in our presented analysis is six, which requires an 4.8-nm FSR and a 33.3-μm ring. This falls without the experimentally verified limit.

Independent routability of each WDM partition enhances path diversity and forms the

basis for WSSR. The number of WDM partitions is increased by interleaving additional sets of wavelengths, being only limited by the achievable wavelength channel density which must adhere to the aforementioned crosstalk constraints [38]. Single partition routers produce a degenerate case where the wavelength selectivity is eliminated, forming a purely spatially-routed design. Additionally, since the input-output port connectivity for all wavelength channels remains the same regardless of the number of WDM partitions, the entire router can be treated as a parameterized building block. These traits enable two features: 1) all previous spatially-routed topologies can be augmented with WSSR, and 2) the number of partitions and the network topology are independent design decisions that can be determined separately.

In a WSSR interconnect topology, each WDM partition can be regarded as an independent communication plane. This is conceptually analogous to electronic network multiplexing techniques such as traditional electronic virtual channels or the use of multiple physical networks. As stated previously, implementation of traditional virtual channels is difficult in the photonic domain due to impracticality of optical buffering and processing. The use of multiple physical planes is also detrimental since it will generate high insertion loss due to increased network complexity. Although increasing path multiplicity by adding extra paths in the network has been previously suggested [51], the previous analysis did

not consider the fundamental physical-layer constraints of the network. These issues are circumvented with WSSR since the network planes are multiplexed in the wavelength domain.

Koohi, *et al.* have proposed 2D-HERT, a wavelength-routed network, which uses a similar partitioned wavelength space for directing wavelength channels [50]. The 2D-HERT network use passive ring filters for guiding a subset of wavelengths. A source node employs source-routing through selection of an appropriate wavelength to establish the complete optical path since the wavelength will determine whether the lightwave will pass through or drop into each passive ring filter. Our WSSR technique is differentiated by the fact that we utilize active electro-optic ring switches for generating several WDM partitions that act like independent network planes. The selection of wavelength only determines which network plane is traversed, but has no role in determining the optical path. An advantage of not utilizing wavelength for routing purposes is that we can exploit wavelength parallelism for enabling higher node-to-node datarates. This type of network behavior is ideally suited for traffic with long-lived and large-message transmissions.

Allocation of WSSR network resources is accomplished using a circuit-switching methodology similar to the one used for spatial routing [51]. Processors interface with the network by communicating with a network gateway (Fig 5.2). Resource allocation of photonic routers is accomplished on a separate light-weight electronic packet-switched control plane, which has

a topology that replicates the photonic layout.

The gateway has the following principle network roles (enumerated in Fig 5.2):

1. **Electronic/Transmission**: Processing cores first send transmission requests to the *Network Injection Arbiter* logic which handles allocation of a *WDM Partition Transmitter* and the circuit-switching network protocol required to provision a photonic path.

2. **Electronic/Reception**: Requests from remote processing cores are sent to the *Network Ejection Arbiter* which will handle allocation of a *WDM Partition Receiver* and the circuit-switching network protocol for the reception end of the photonic link.

3. **Photonic/Transmission**: Each WDM Partition Transmitter is tuned to transmit on a different set of wavelengths, corresponding to a particular WDM partition.

4. **Photonic/Reception**: Each WDM Partition Receiver is tuned to receive on a different set of wavelengths, corresponding to a particular WDM partition.

The WSSR path-allocation protocol occurs through the transmission of a series of control messages on the control plane. All control messages contain fields for message type, source ID, destination ID, and WDM partition selection data. The WDM partition selection field contains two flags (bits) per WDM partition that exists in the system. The first bit labels

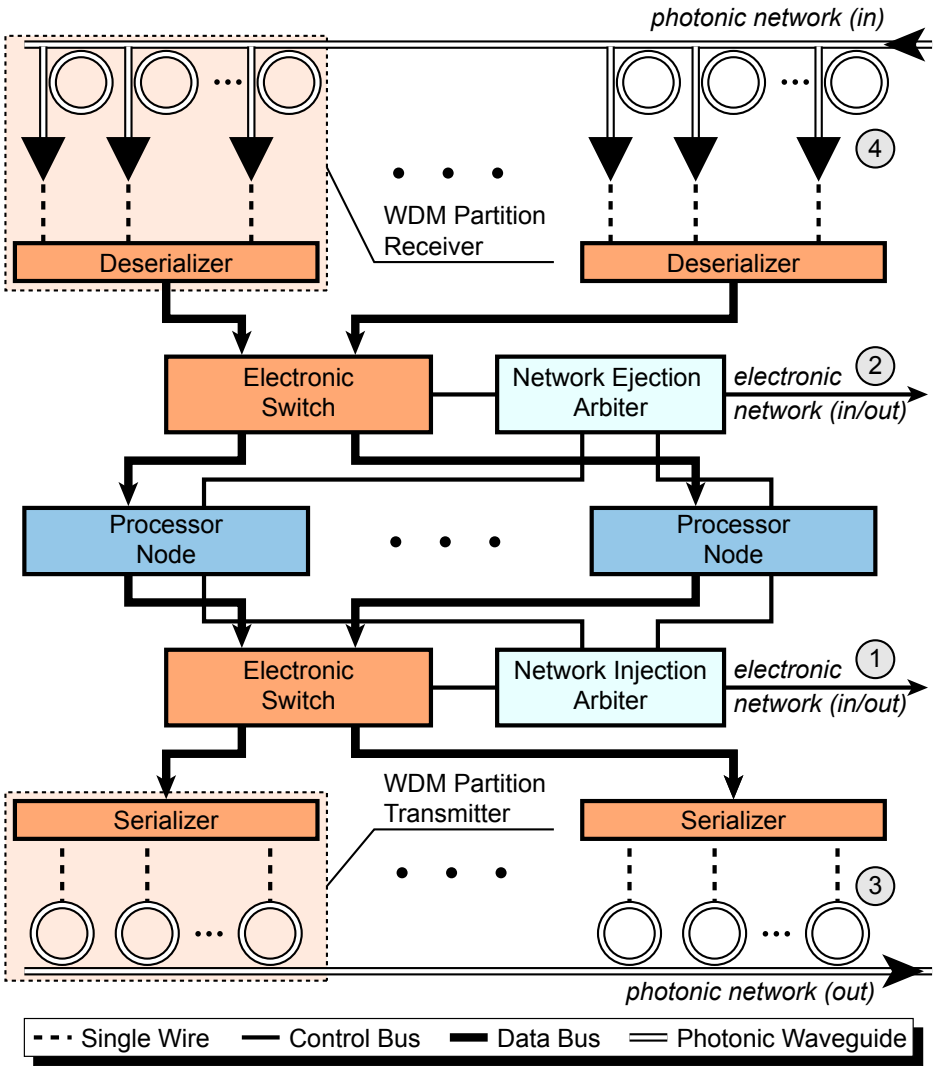

Figure 5.2: The WSSR gateway architecture with concentrating processing cores.

'check' (indicating a partition that is being considered for allocation) and the second bit labels 'available' (indicating the current assumed resource availability for the corresponding partition). Fig. 5.3 and Fig. 5.4 illustrate the message transactions required in perform allocation and data transmission. In the example, the request is initially blocked at an

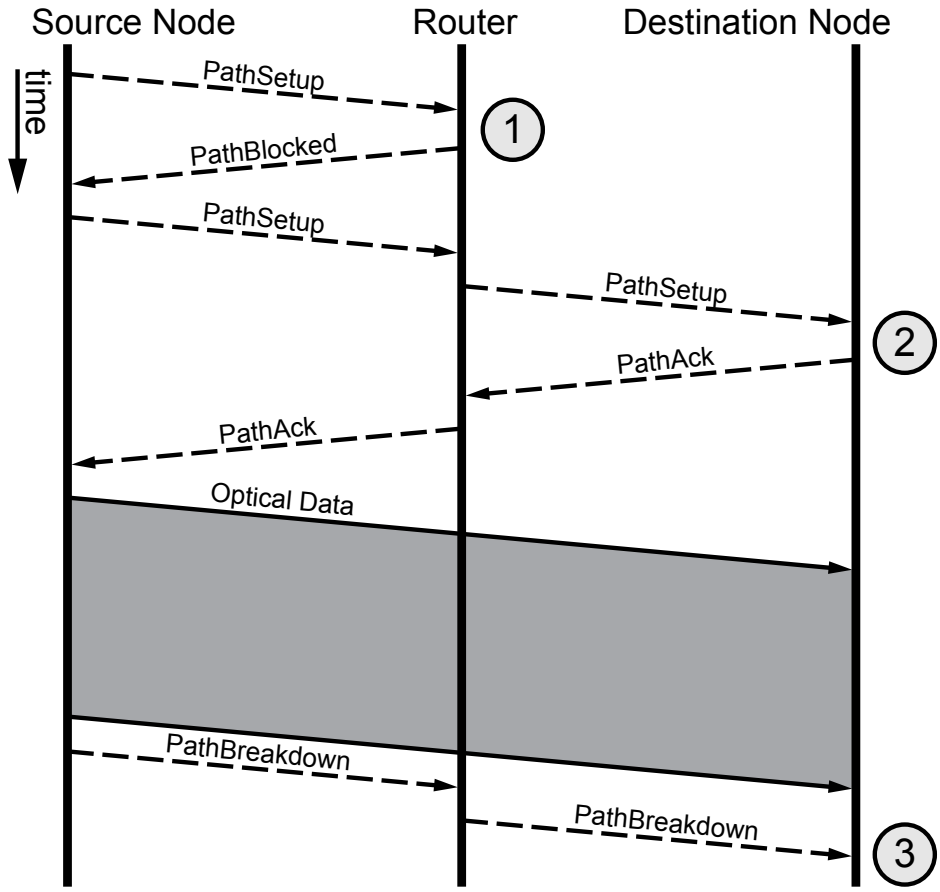

Figure 5.3: Example timing diagram of the circuit-switching and WSSR allocation protocol. A path provisioning request is initially blocked, but is successful upon re-attempt.

intermediate router, retries, and is subsequently successful in resource allocation and data transmission.

The allocation of a path begins with the transmission of a message of type *PathSetup* from a source node. The 'check' bit is set to 'true' on each partition for which allocation will be attempted. This automatically precludes partitions that have already been allocated from

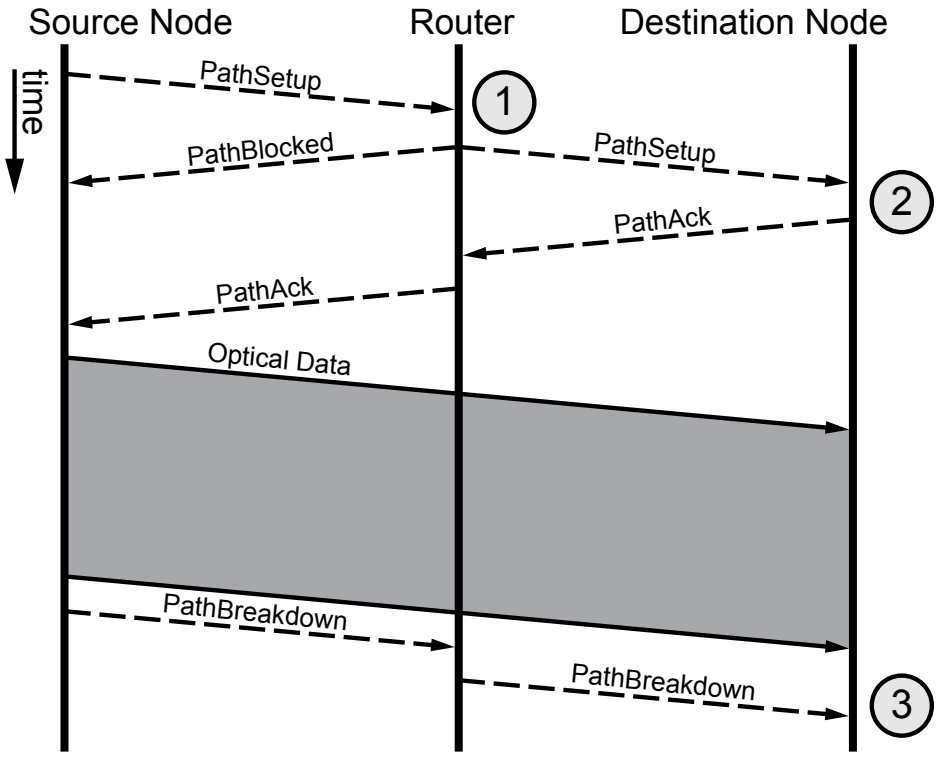

Figure 5.4: Example timing diagram of the WSSR allocation protocol. If a single path provisioning request is attempted with multiple partitions, a path-setup request can partially block on a particular partition while be successful on another partition.

that particular source node and have not been de-allocated yet, or on partitions where an allocation attempt is concurrently being made by another *PathSetup* message. Initially, the 'available' bits are all equal to the 'check' bits since an attempt at allocation is only performed if the partition is available at the gateway. For the simulation analysis presented in this paper, the gateway only attempts to allocate a single WDM partition per *PathSetup* message. Alternatively, the *PathSetup* message could set a 'true' value for all 'check' bits which are

free for allocation to increase the likelihood of finding a partition that is available. Previous work referred to the number of partitions used during each *PathSetup* as the reservation aggressiveness [58].

The *PathSetup* message travels on the control plane, attempting to provision each WDM partition which still has the 'check' and 'available' bit set as 'true'. Each photonic router in the network maintains its own reservation table, which is used to track circuits and WDM partitions that have been allocated or are in the process of being allocated. If any of the partitions are blocked, then a *PathBlocked* message is created and returned to the originating node, with the 'check' bit set to 'true' and 'available' bit set to 'false' for the blocked channels. The 'check' and 'available' bits that correspond to the blocked circuits are set to 'false' in the *PathSetup* message, and is only continues propagation if at least a single WDM partition is still available. Fig. 5.3 (marker '1') illustrates a situation where the path is blocked for all partitions being considered. Fig. 5.4 (marker '1') shows a sequence of events where a subset of the available WDM partitions of the *PathSetup* message are blocked. The alternative WDM partition enables the *PathSetup* message to proceed and complete the provisioning process.

A *PathSetup* message that reaches the destination gateway indicates that at least one source-to-destination circuit is available for photonic transmission (Fig. 5.3 and Fig. 5.4 at

marker '2'). The message is converted to a *PathAck* message, the source and destination ID are swapped, and the 'check' bits are preserved while the 'available' bit is set based on how many channels will be used for the transmission. The simulation studies in this paper limit the allocation to a single WDM partition, and is chosen at random from the pool of available channels as indicated by the available bits. However, alternative configurations could enable some or all of the available partitions to be aggregated together to allow for dynamic throughput allocation.

Upon completion of the photonic transmission, a *PathBreakdown* message is sent into the network from the source node (Fig. 5.3 and Fig. 5.4 at marker '3'). The 'check' bit is set for each partition that was allocated and is used to indicate to each photonic router along the path that the resources should be released and reservation table updated appropriately.

Koohi, *et al.* have proposed 2D-HERT, a wavelength-routed network, which uses a similar partitioned wavelength space for directing wavelength channels [50]. The 2D-HERT network use passive ring filters for guiding a subset of wavelengths. A source node uses source-routing through selection of an appropriate wavelength to establish the complete optical path since the wavelength will determine whether the lightwave will pass through or drop into each passive ring filter. Our WSSR technique is differentiated by the fact that we utilize active electro-optic ring switches for generating several independent network planes.

A previously proposed circuit-switched topology is the TorusNX which is designed with a reduced number of crossings and an optimized switching layout [53]. A 4×4 version of the TorusNX is illustrated in Fig. 4.12, consisting of 16 gateway switches and 16 4×4 non-blocking switches. The structure of each switch configured with two WDM partitions is diagrammed in Fig. 5.5. Each pair of rings (indicated by a red and blue ring) composes the two cascaded rings that compose a two-partition router. Note that the original single-partition design of the gateway appears in Fig. 4.13 and can be reconstructed by removing either the red or blue set of rings from the layout. Similarly, the single-partition 4×4 non-blocking switch design in Fig. 4.4c can be reconstructed by removing one set of rings.

5.2 Experimental Validation

We experimentally demonstrate the WSSR concept and report performance measurements of the active transmission of six 10-Gb/s WDM channels through a silicon electro-optic microring switch; the demonstration shows the active routing of a partition of three channels, while leaving the remaining three channels unperturbed [55]. We use a second-order electro-optic microring switch (Fig. 5.6) fabricated at the Cornell Nanofabrication Facility [32]. Previous work has shown active switching of 40-Gb/s data through this device [94].

For the purposes of this experimental validation, we define two WDM partitions: the

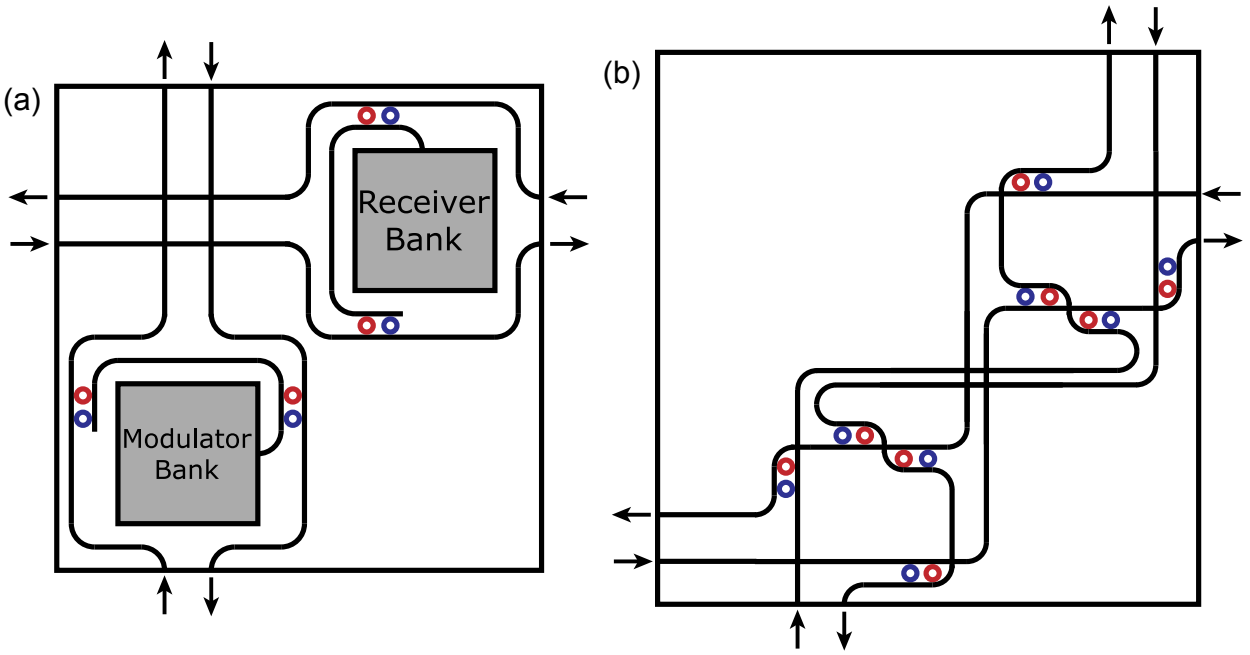

Figure 5.5: Schematic of the TorusNX photonic routers configured with two WDM partitions: (a) gateway switch and (b) 4×4 non-blocking photonic switch.

primary partition and the *auxiliary partition*. Each partition will consist of three wavelength channels. The auxiliary wavelength channels will propagate past the microring undisturbed, regardless of whether a bias is being applied to the ring or not. Two ring resonators can be cascaded and each aligned to a different WDM partition, forming a 2 partition 1×2 WSSR router (as depicted in Fig. 5.1.

To the best of our knowledge, this experiment also represents the first demonstration of a WDM data signal being switched through an electro-optic microring resonator. Note that the prior discussion on the proposed routing scheme assumes first-order microring devices;

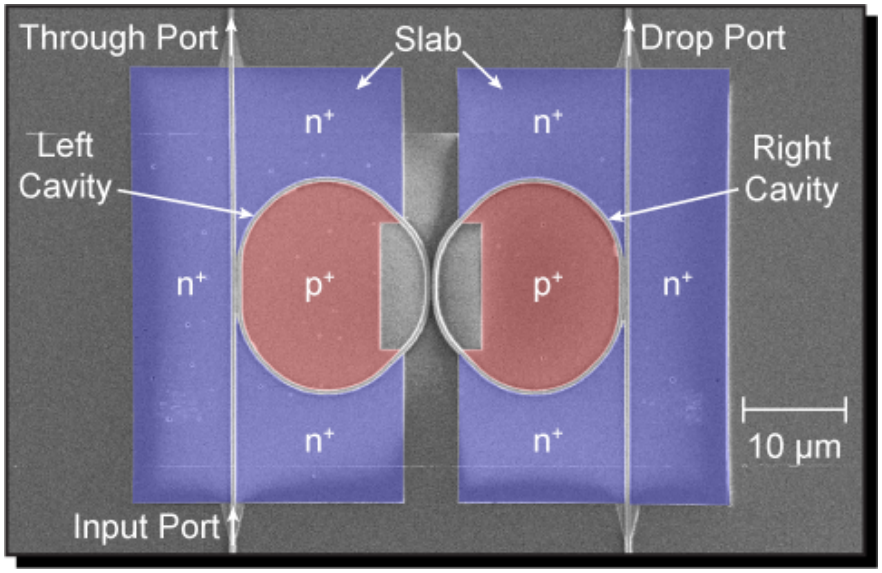

Figure 5.6: Scanning electron microscope image of a second-order electro-optic microring switch. Blue and red coloring is provided to label slab regions with dopants.

however, this concept can be readily applied to higher order devices. An advantage of the second-order device is that it exhibits hitless characteristics when the modes are shifted off resonance, producing suppressed resonances due to the Vernier effect, and thereby reducing the impact on adjacent wavelength channels.

5.2.1 Experimental Setup

The experimental setup (Fig. 5.7) consists of six continuous-wave (CW) laser sources combined using a dense wavelength-division multiplexer (DWDM). The six channels are simultaneously amplitude modulated using a LiNbO3 modulator (MOD) with a $2^7 - 1$

pseudo-random bit sequence generated by a pulse pattern generator (PPG) at 10 Gb/s. The six wavelength channels are decorrelated at the output of the modulator, amplified (EDFA), aligned for quasi-TM propagation, and injected into the chip using a tapered fiber. The microring switch is electrically contacted using high-speed electrical probes and driven using a data timing generator (DTG). At the output of the chip, a filter () is used to select a single WDM channel. The single channel is then amplified, filtered, and sent through a variable optical attenuator (VOA). Lastly, the optical channel is fed into a PIN photodiode with a transimpedence amplifier (PIN-TIA), followed by a limiting amplifier (LA). The received data is assessed using a bit-error-rate tester (BERT). A common 10-GHz clock is used to synchronize the PPG, DTG, and BERT. A digital communication analyzer (DCA) and optical spectrum analyzer (OSA) are positioned throughout the setup for capturing eye diagrams and optical spectra.

5.2.2 Experimental Results

We first measure the passive optical spectrum of the device, showing three resonant modes which will be used for switching the primary WDM partition (Fig. 5.8). To determine the optimal wavelength positioning of the primary WDM partition, we first scanned across the resonant mode centered at 1551 nm using a tunable CW laser on the drop port while

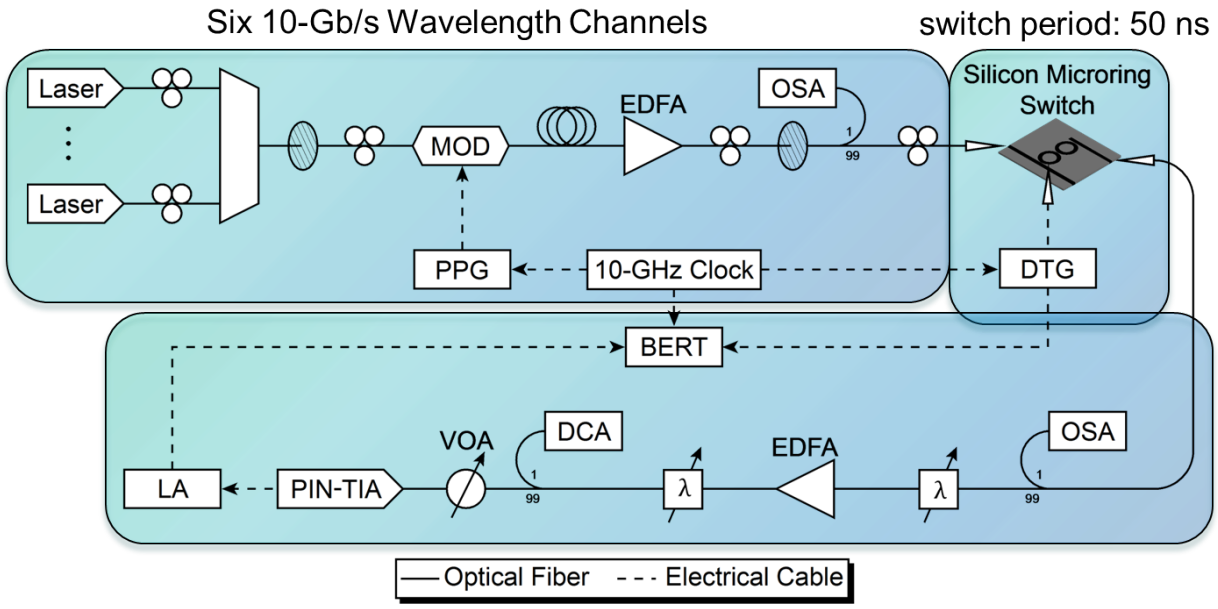

Figure 5.7: Diagram of the experimental setup for demonstration of WSSR concept. The three major optical link components are represented in this setup: generation (top left block), manipulation (top right block), and reception (bottom block).

optimizing the DTG voltage bias for maximum extinction ratio. With the DTG bias fixed, we then scanned across every mode of interest to measure the extinction ratio on both the through and drop ports. Channels positioned at 1541.25 nm, 1550.05 nm, and 1559.01 nm were selected for having the most balanced extinction between the two output ports (approximately 9-10 dB). The auxiliary partition wavelengths are positioned at 1536.88 nm, 1545.65 nm, and 1554.53 nm, and were arbitrarily selected to lie approximately between adjacent modes. In a fully-implemented version of the router, the auxiliary channels should

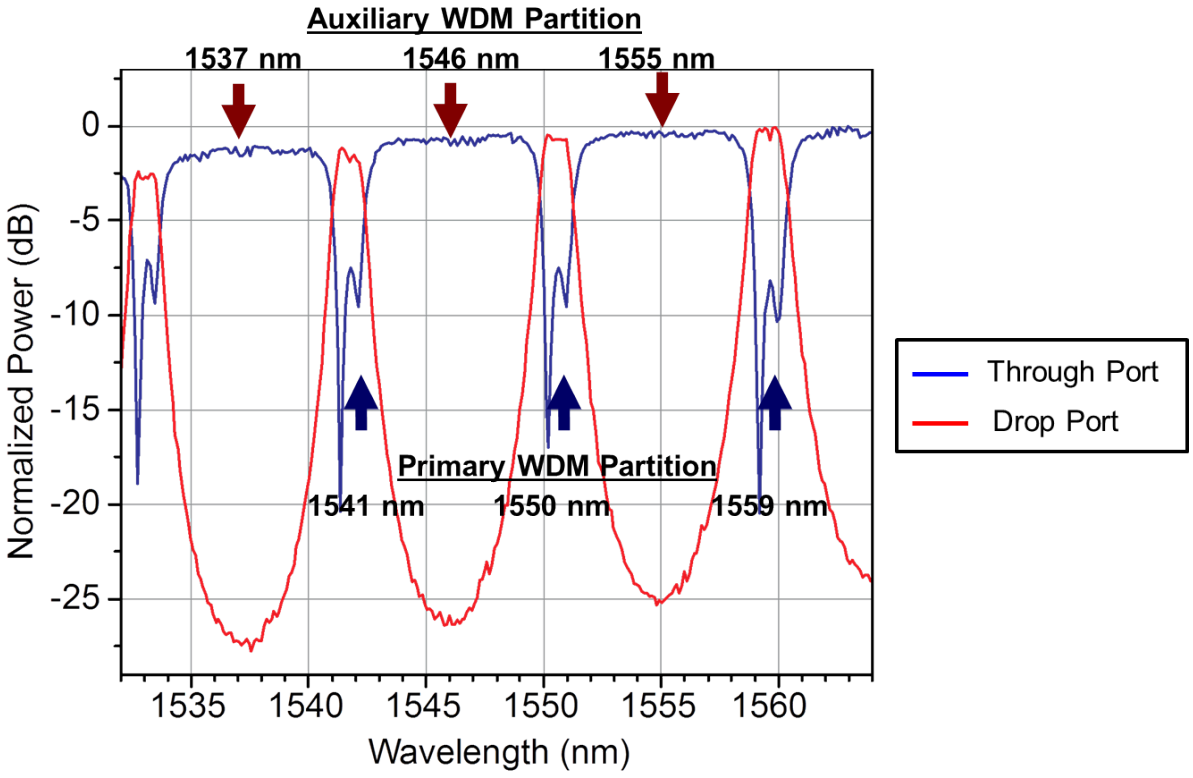

Figure 5.8: The through port and drop port spectra of the electro-optic microring switch in the passive state.

correspond to the resonances of the secondary microring. The power of each channel at the chip input is set to approximately 0 dBm and the switch is operated with a 100-ns period and 50% duty cycle, resulting in 50-ns long optical packets.

The BER curves are reported in Fig. 5.9 showing minimal data degradation. The BER curves of the primary WDM partition were shifted by 2.5 dB in order to account for differences in average measured power at the receiver due to the 50% duty cycle at the

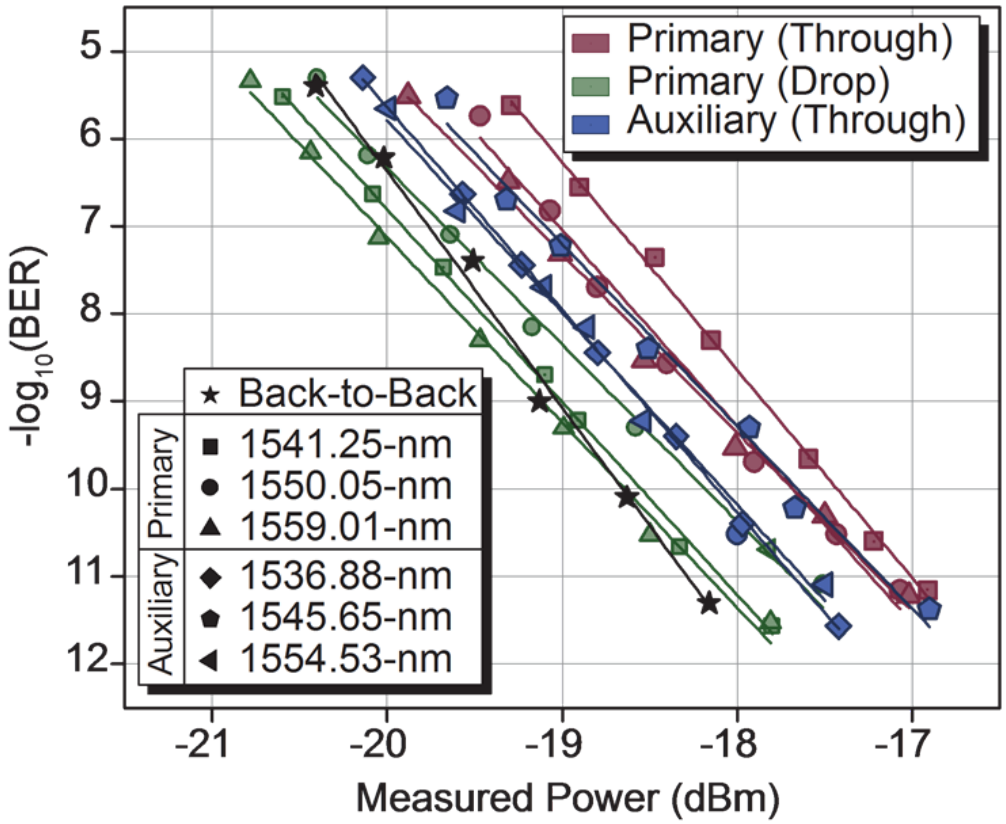

Figure 5.9: BER curves for each of the six wavelength channels and the back-to-back case which bypasses the chip.

device output. The shift was analytically calculated based on a 9-dB extinction ratio.

In Fig. 5.10, we record each 10-Gb/s eye diagrams, showing open eyes. Fig. 5.11 shows 50-ns long optical packets at the input and output ports for both primary and auxiliary partitions. Notice that the auxiliary partition packets transmit through the actuating ring switch undisturbed.

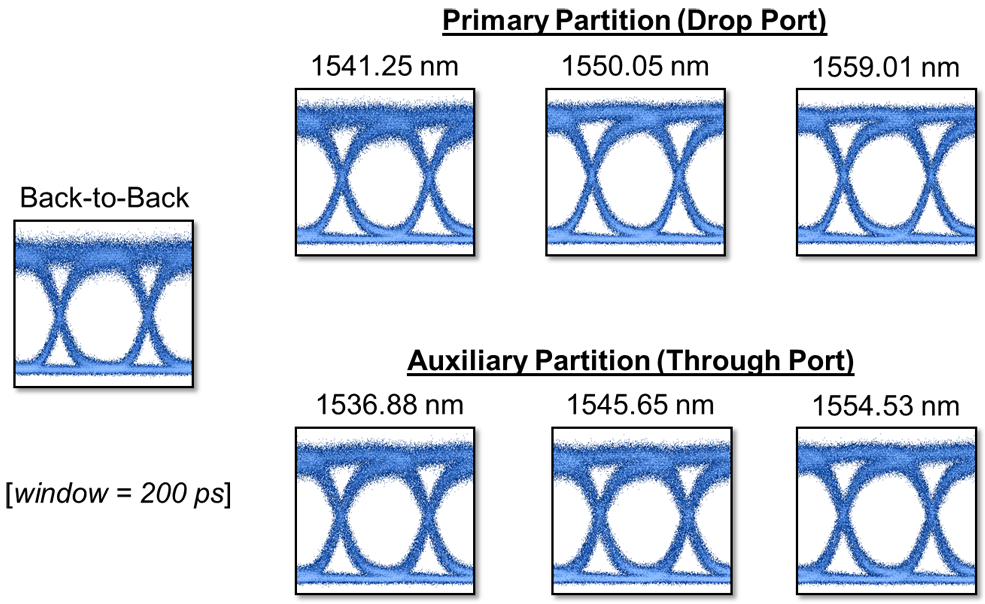

Figure 5.10: 10-Gb/s output eye diagrams at both output ports from the device in the active state.

5.3 Analytical Analysis

5.3.1 Optical Power Budget and Insertion Loss Analysis

The consideration of the physical-layer properties of the photonic network plays a critical role in determining the feasibility of implementing the network. Specifically, the optical power budget and network-level insertion loss will determine the requirements for the laser input power and for the receiver sensitivity. The insertion loss analysis assumes the parameters listed in Table 5.1 and are derived from experimentally-validated published results.

The results of the analysis are shown in Fig. 5.12 for different levels of partitioning.

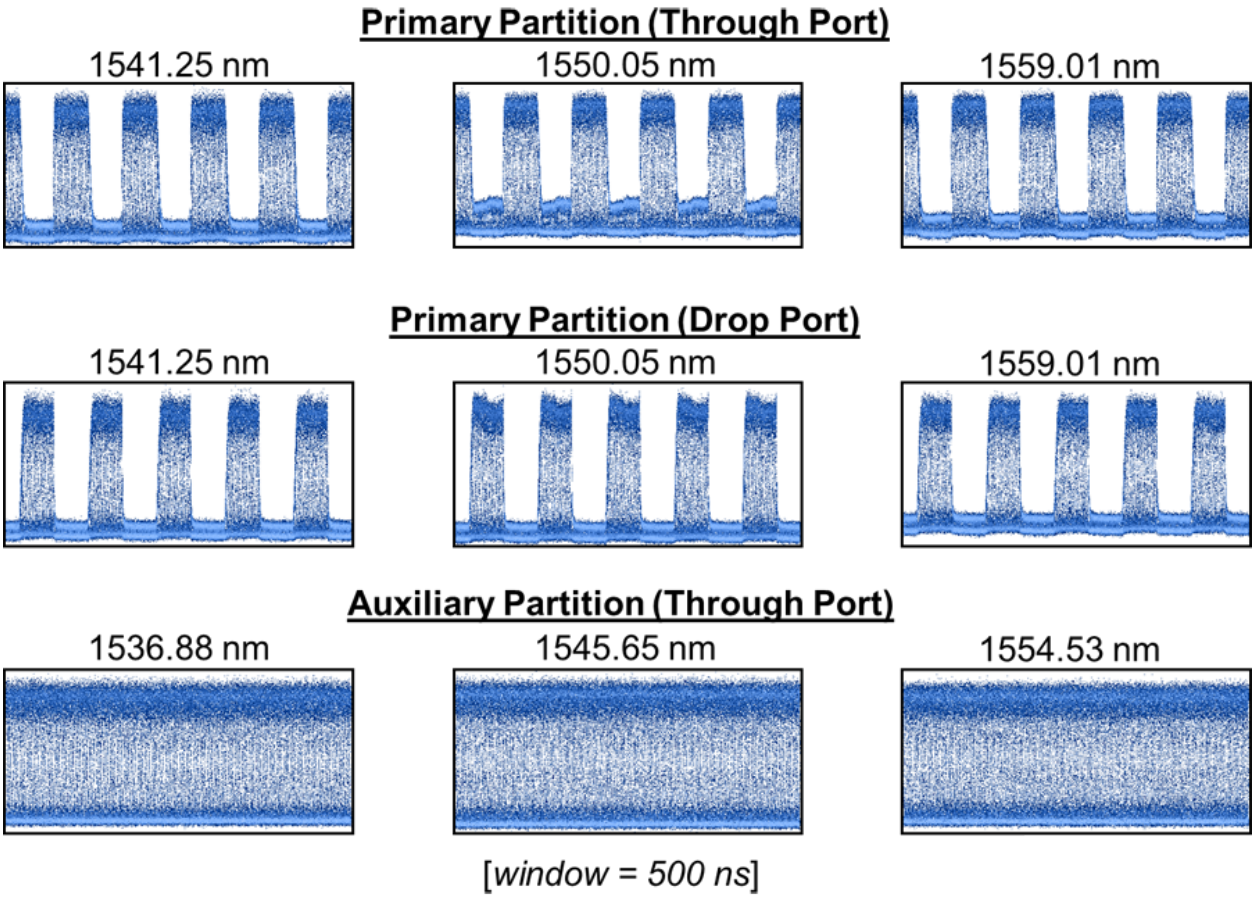

Figure 5.11: 10-Gb/s output optical packets at both output ports from the device in the active state.

An initial cost of 0.72-dB insertion loss is observed when transitioning from one to two partitions; this jump in loss is attributed to additional waveguides and bends required to accommodate the additional ring resonators. Scaling beyond two partitions requires an increase in waveguide propagation and in the number of times ring resonators are passed, nonetheless a minor 0.56-dB loss increase is observed when transitioning from two to six

Table 5.1: Insertion Loss Parameters - Wavelength-Selective Spatial Routing Analysis

Parameter	Value	Ref.
Propagation Loss (Silicon)	1.7 dB/cm	[19]
Waveguide Crossing	0.16 dB	[23]
Waveguide Bend	0.005 dB/90°	[19]
Drop Into a Ring	0.6 dB	[30]
Pass By a Ring	0.005 dB	[30]

partitions (0.14 dB per added partition).

The required laser power can be computed by adding the expected network loss to the receiver sensitivity. A receiver with a -17-dB sensitivity and operating at a 10-Gb/s datarate (demonstrated in [44]) would require a minimum injected laser power at the modulator of 8.0 dBm, 8.7 dBm, 8.8 dBm, and 9.0 dBm for one through four WDM partitions, respectively. We envision the optical-power delivery to the chip to either leverage vertical grating couplers [28] or lateral tapered waveguides [25].

We can also observe that the largest loss components arise from the waveguide crossings and the propagation. This shows that the introduction of WSSR into the photonic circuit-switching network topology only adds a small amount of loss to the network. Our presented analysis assumes a planar single-crystalline silicon fabrication platform, but alternative

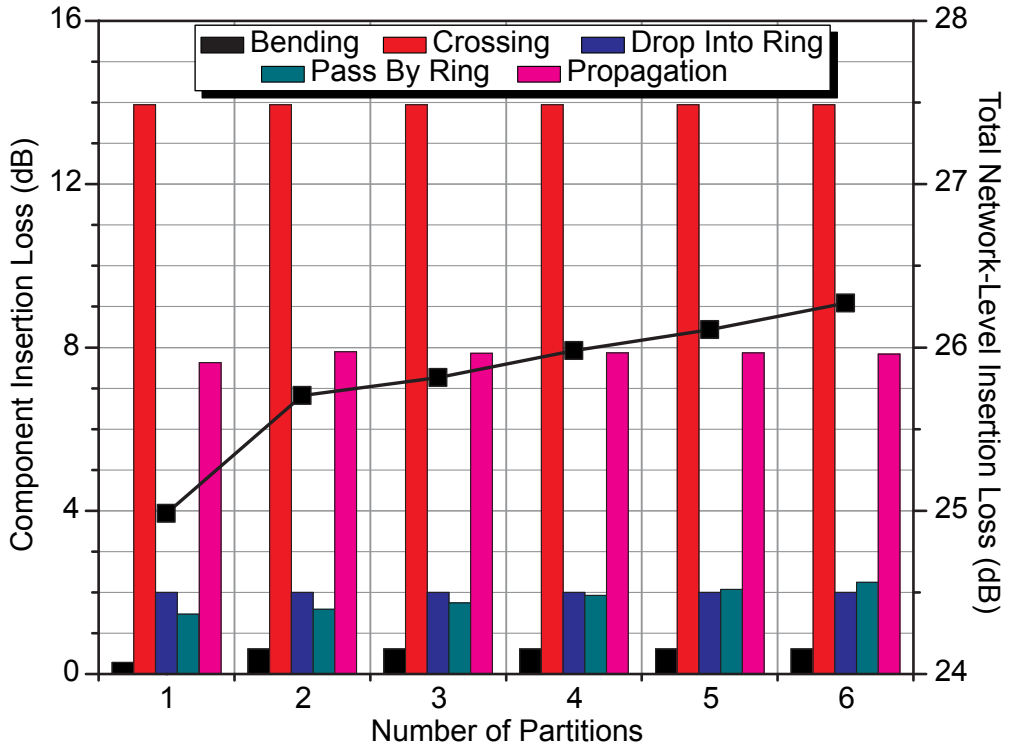

Figure 5.12: Insertion loss analysis of the TorusNX topology for varying levels of partitioning. Column plots correspond to worst-case insertion loss per component among all possible network paths (left-vertical axis). The line plot corresponds to greatest total network-level insertion loss path among all possible network paths (right-vertical axis). The lossiest path does not necessarily correspond with the sum of the worst-case losses per component.

CMOS-compatible platforms such as 3D deposited technology can virtually eliminate these loss constraints and increase the feasibility of this type of network [95].

5.3.2 Photonic Footprint

The nature of the WSSR mechanism requires multiple rings to enable the individual controllability of each WDM partition. As the number of WDM partitions increases one of two ring design changes can be employed. In the first case, ring diameters are fixed regardless of the number of WDM partitions which will produce a system with higher channel density and consequentially higher wavelength channel crosstalk. Alternatively, the wavelength channel density can be fixed by scaling the ring diameter inversely proportional to the number of WDM partitions. While this has the benefit of not increasing spectral density of the channel spacing, this also enables a reduced footprint of the photonic routing element. Our proceeding area analysis assumes the scaling of the ring diameters with a maximum considered diameter of 200 μm.

The area footprint of a single WSSR router versus the number of WDM partitions (labeled as the number of rings) is analyzed in Fig. 5.13. The WSSR router footprint calculations assume a structure similiar to those shown in Fig. 5.1. The only locations where waveguides are closely placed together are regions where optical coupling are required (*i.e.* where the optical signal enters and exits the ring resonator). To prevent optical coupling across waveguides that are meant to be isolated, a 5-μm gap are used (*e.g.* between adjacent rings). An additional 2.5-μm gap is assumed to be on the outside edge of the two straight waveguides

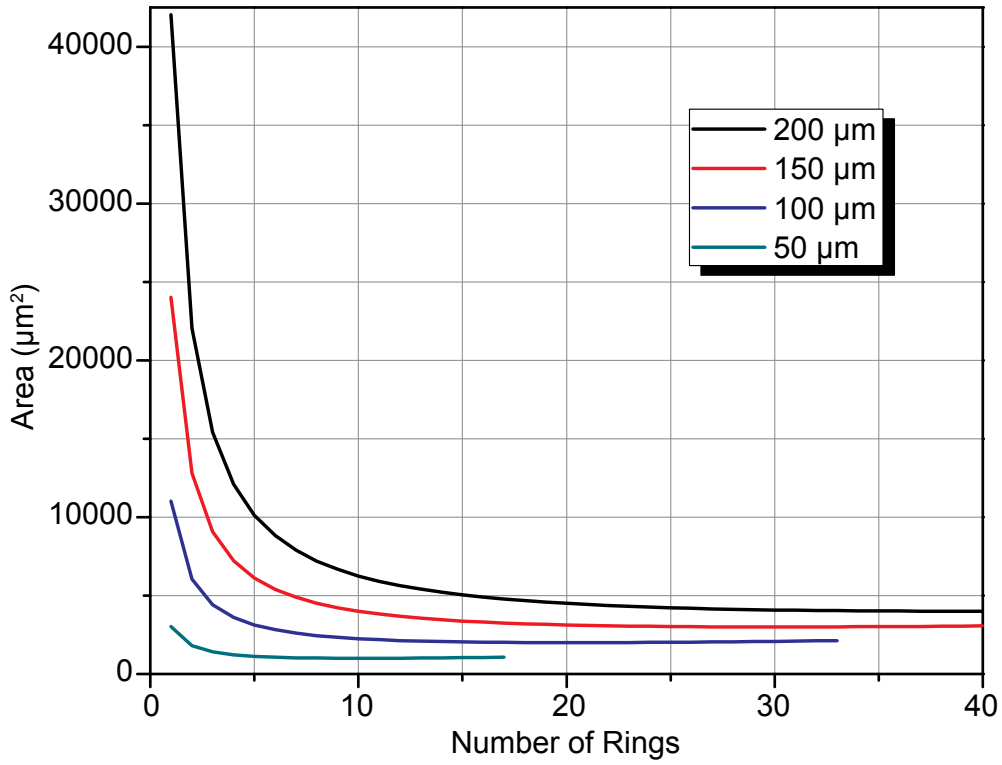

Figure 5.13: Photonic router footprint for varying number rings (which corresponds to the number of WDM partitions enabled by the router). Legend indicates the ring diameter for the single ring case.

to account for space required with other optical components (*e.g.* another photonic router) outside of the immediate WSSR router of interest.

The plot shows the area scaling for varying initial single-partition ring diameters (as indicated in the legend). The curves show an immediate area benefit for increasing the number of partitions. Not only is this beneficial for the WSSR technique, but this scaling can also be used to benefit standard circuit-switching architectures through a reduction of the

photonic footprint. The operational difference between WSSR and circuit switching is that the cascaded rings are used cohesively instead of independently. As the number of partitions increase, the area reduction diminishes and eventually an area increase is observed. The inflection point occurs at 40, 30, 20, and 10 rings for the 200, 150, 100, and 50 μm cases, respectively. Each curve ends at the point where the individual ring diameters would become less than 3 μm which is our assumed minimum size limit of the ring resonators (corresponding to the smallest known fabricated microring [93]).

5.3.3 Contention Probability

From a performance perspective, the added path diversity by WSSR allows multiple communication links to occupy the same waveguides and photonic routers, resulting in reduced network-level contention. Decreased contentions will reduce latencies caused by network resource unavailability, and increase network-level bandwidth due to the higher availability. Fundamentally, the use of multiple WDM partitions is equivalent to the concept of path multiplicity previously proposed and shown to improve performance of on-chip networks [51]. The primary difference in the two architectural concepts is in the usage of cascaded wavelength-selective spatially routed rings for WSSR and the overlay of additional waveguides and routing elements for path multiplicity. Previous work has

shown that waveguide crossings (which would be need for added path multiplicity) are the largest contributor to insertion loss while the through port ring switch losses (used in WSSR) contribute a negligible amount [53]. This trend is agreeable for the proposed WSSR routing design since we can observe that the number of through port traversals will increase, but no additional crossing traversals will be created.

Destination blocking occurs in the scenario when multiple source nodes request to transmit to a common destination node at the same time. This condition can occurs within many traffic patterns where transmission requests experience hotspots. In the context of traditional circuit switching, a destination can only receive from a single source at any period in time. WSSR can alleviate this issue by providing multiple receiver connections for each destination.

We assume a non-blocking network for the purpose of analyzing the contention characteristics of destination blocking. In a traditional circuit-switched non-blocking network, any idle source node can immediately transmit to its intended destination with the condition that the destination is not already receiving a message (*i.e.* no contention due to source blocking or circuit-path blocking). Consider a N node network, with a transmission being requested from source node A to destination node B. If all nodes aside from A have either established a connection or have been blocked (*i.e.* a saturated network), then there

are $N-2$ nodes that could block this new connection. Assuming nodes do not require the optical network to communicate with itself, then the probability that the connection from A to B will not be destination blocked is

$$P_a = \left(\frac{N-2}{N-1}\right)^{N-2}$$

Now we consider a non-blocking WSSR network with C WDM partitions. If each node is restricted to a single message transmission at a time (*i.e.* single transmitter per gateway), then the destination blocking probability of A is

$$P_C = 1 - \sum_{i=0}^{C-1} \binom{C}{i} P_a^{C-i} \cdot (1-P_a)^i \tag{5.1}$$

It can be easily shown that P_C will converge as $N \to \infty$, where the limits can be expressed exactly as

$$L_C = \lim_{N\to\infty} P_C = \left(1 - \frac{1}{e}\right)^C \tag{5.2}$$

Fig. 5.14 plots Eq. 5.1 for $C = 1 \ldots 6$ and $3 \leq N \leq 100$, and appends the limit calculated from Eq. 5.2. We can observe a dramatic destination blocking probability improvement from $L_1 = 0.63$ to $L_6 = 0.064$. Furthermore, networks containing more than ~25 nodes vary minimally in terms of destination blocking probability. This indicates that techniques

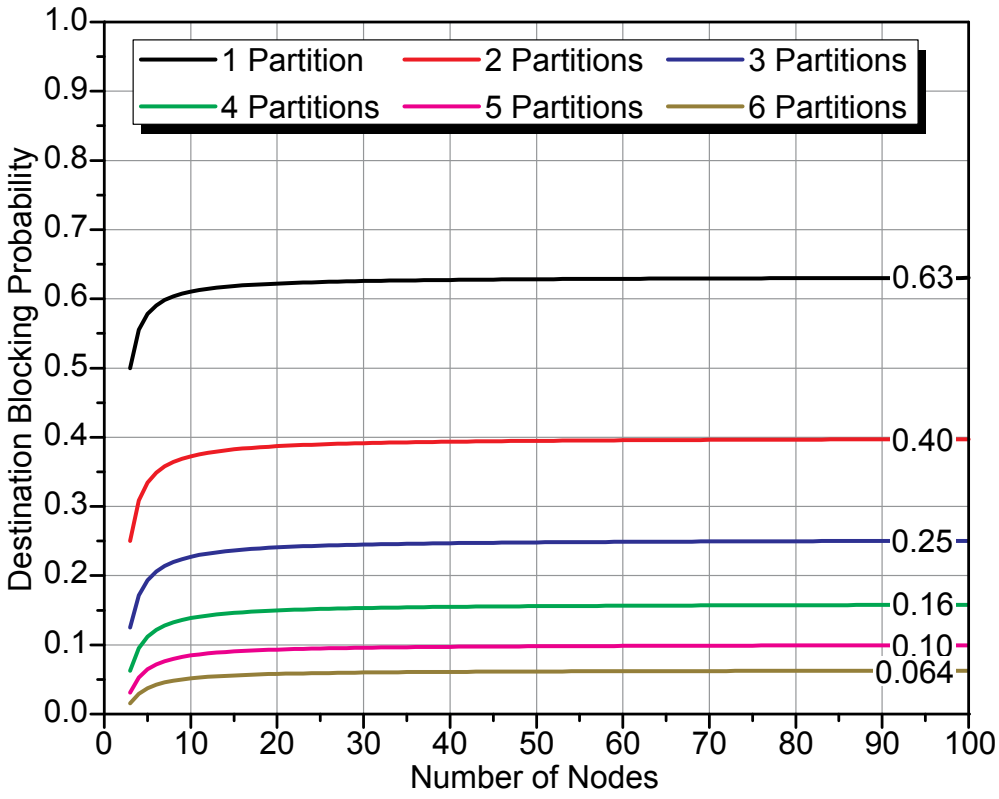

Figure 5.14: Destination blocking probability in a non-blocking network for varying number of interleaved channels. The limit of each blocking probability as $N \to \infty$ is superimposed on the right of the plot.

such as WSSR can provide dramatic improvements to performance through a reduction in blocking probability.

5.4 Simulation Results and Analysis

The partitioned-WDM network architecture is next modeled and simulated in PhoenixSim [68]. All conducted simulations assumed a 2-cm×2-cm 64-core CMP, which requires an 8×8 network.

The photonic architectures assume a 2.5-GHz clock for the electronic control plane. The control-plane routers utilize channel widths of 32 bit and 256-bit input buffers, corresponding to a buffer depth of 8 control messages. Path-setup control messages have an assumed bit length of 32 bit. Photonic networks are normalized by their total number of transmission wavelengths used, and wavelengths are evenly allocated among the WDM partitions. Each wavelength channel provides a 10-Gb/s serial data rate. The TorusNX, described in Section 4.3.1.2, photonic circuit-switching topology design was used for this study.

We also simulate a traditional electronic mesh network to serve as a baseline comparison for the proposed photonic architectures. Each electronic router assumes a channel width of 128 bit and utilizes a 2048-bit buffer on each input port. This electronic network model employs bubble flow control to prevent deadlocks. The electronic mesh network also operates on a 2.5-GHz clock, producing a link-level bandwidth of 320 Gb/s, and a network-level bisection bandwidth of 5.12 Tb/s for the 8×8 network.

5.4.1 Synthetic Traffic

Performance measurements were recorded for varying degrees of message size, total number of wavelength channels, and number of WDM partitions. Simulations were conducted with either a small (1-kbit) or large (100-kbit) message size. All synthetic traffic simulations utilized the standard uniform random traffic pattern. The number of WDM partitions ranged from 1 to 4 to capture the performance effect that the wavelength-selective spatial routing technique provides. The total number of wavelengths was varied between 12 (low aggressiveness), 60 (medium aggressiveness), and 120 wavelength channels (high aggressiveness).

Fig. 5.15 contains plots for each combination of message size and total number of wavelength channels specified. The dotted-line curves depict the performance of the standard electronic mesh which is only influenced by the message size.

Photonic network configurations using small 1-kbit messages (left plots in Fig. 5.15) achieve saturation bandwidth gains that scale proportionally with the number of WDM partitions used. In the case of 60 and 120 wavelength channels, the small message sizes result in negligible differences in serialization delay when scaling the number WDM partitions. Consequently, this results in a fixed zero-load latency (approximately 90 ns) regardless of the number of WDM partitions, and saturation bandwidth gains that are approximately

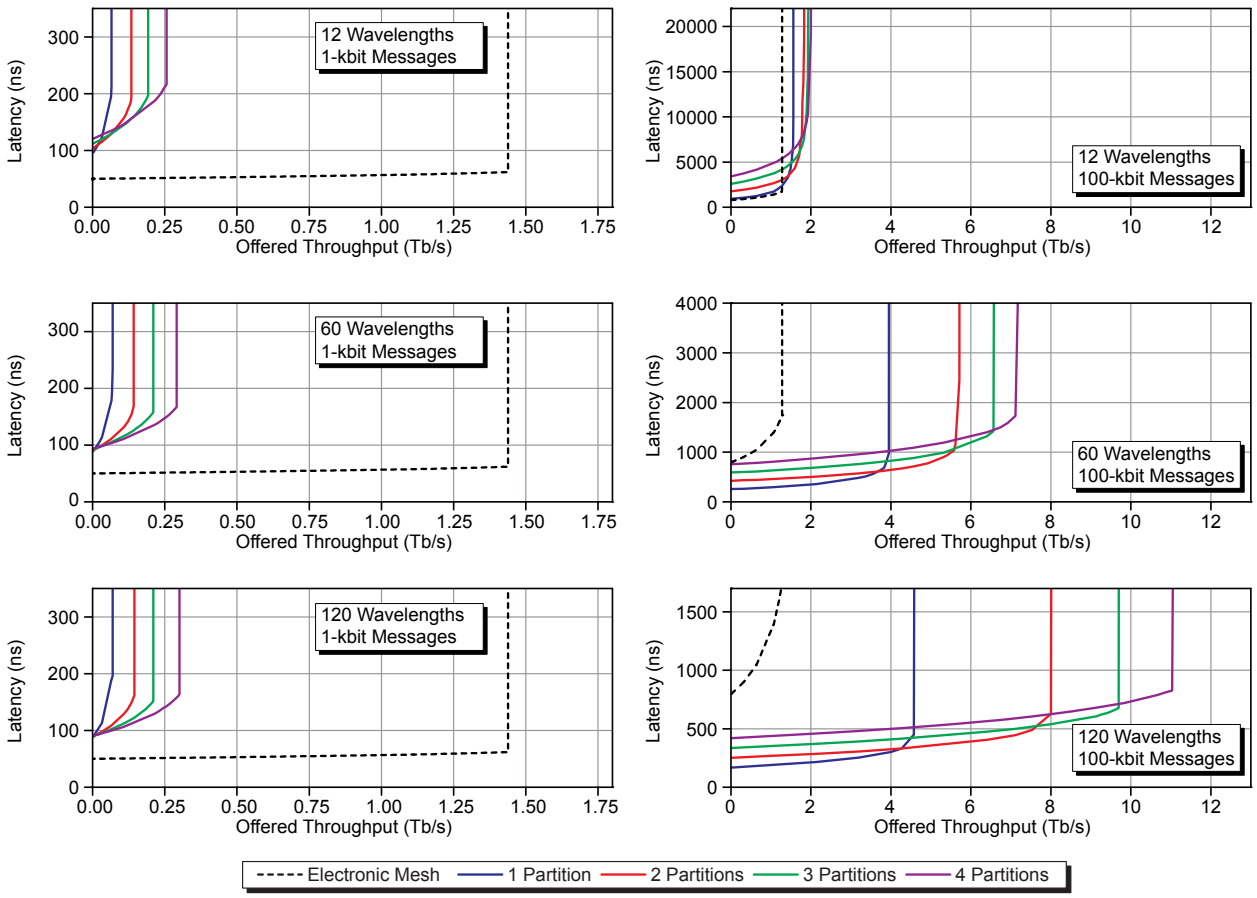

Figure 5.15: Average latency versus offered throughput for varying number of WDM partitions, message sizes, and number of wavelength channels. Electronic mesh performance is shown as a dotted line.

equal to the number of WDM partitions (*e.g.* 4 partitions results in a 4× improvement). Only in the case of 12 wavelength channels is there a perceivable difference in serialization delay which results in a slightly degraded zero-load latency (120 ns for 4 partitions) and lower gain in saturation bandwidth (approximately 90% gain per partition). The WDM

partition technique provides significant performance gains relative to the degenerate case, however the photonic network variants still underperform in comparison to the electronic mesh, a disadvantage that has been previously concluded for circuit-switched networks [54].

The transmission of 100-kbit messages (right plots in Fig. 5.15) on all the photonic network variants produce better performance values compared to the electronic mesh baseline. When compared to the degenerate case, the 12-wavelength system produces saturation-bandwidth gains of 14%, 21%, and 24% when utilizing two, three, and four WDM partitions, respectively. In the 120-wavelength channel case, the saturation bandwidth gain is 97%, 140%, and 169%, for the two, three, and four partition cases, respectively. In the best case, four partitions using a total of 120 wavelength channels achieves a saturation bandwidth improvement of 764% over the electronic mesh. This shows that modest gains are achievable using WSSR for nearer term photonic networks, however greater gains can be expected as photonic device fabrication matures. Due to the large message sizes, the serialization delay is significantly longer and has a greater impact on the zero-load latency. For each each set of plots with a common total wavelength count, the division of wavelength channels among WDM partitions produces noticeable differences in delay. This produces a noticeable trade-off when determining whether a system design should minimize latency or maximize bandwidth.

5.4.2 Trace Simulations of Scientific Applications

Presented next is an analysis of the performance of scientific applications on the proposed WSSR architecture. The photonic architectures assume the use of 120 wavelength channels, each transmitting a serial datarate of 10 Gb/s. The performance evaluation of the proposed architecture uses trace information extracted from four different message-passing interface (MPI) based scientific applications, summarized here:

- *Paratec* - a materials science application using the density functional theory method [96]
- *Cactus* - an astrophysics computation toolkit designed to solve coupled nonlinear hyperbolic and elliptic equations arising from general relativity [97]
- *GTC*- a 3D particle-in-cell application developed to study turbulent transport in magnetic confinement fusion [98]
- *MADbench* - a benchmark based on MADspec cosmology code, calculating the maximum likelihood angular power spectrum of the cosmic microwave background [99]

Each application trace contains a listing of all core-to-core communications that occurred during a complete execution of the algorithm on a 64 node system. Each trace entry lists the phase, source thread ID, destination thread ID, and the message size. This set of application traces form a representative set communication patterns that match a large class of scientific

Table 5.2: Application Trace Characteristics - Wavelength-Selective Spatial Routing Analysis

Application	Number of Phases	Number of Msgs.	Total Data Sent (B)	Avg. Msg. Size (B)
Paratec	34	126059	5.4M	43.3
Cactus	2	285	7.3M	25600
GTC	2	63	8.1M	129796
Madbench	195	15414	86.5M	5613

applications currently being investigated by the computational science research community. The characteristics of each application trace are summarized in Table 5.2 and traffic pattern plots are shown in Fig. 5.16.

The phase value is used to indicate MPI barriers during the execution of the code. A single predetermined core acts as a *master* node, and collects phase completion messages from all other *slave* nodes. Upon reception of completion messages from all nodes, the master node will broadcast commands to begin the following phase of execution. For this study, the described synchronization process occurs using the electronic control plane.

Source and destination thread IDs label the transmitting and receiving threads of the application. This is differentiated from the source and destination core of the micro-architecture. This distinction occurs due to the fact that the optimal thread-to-core mapping

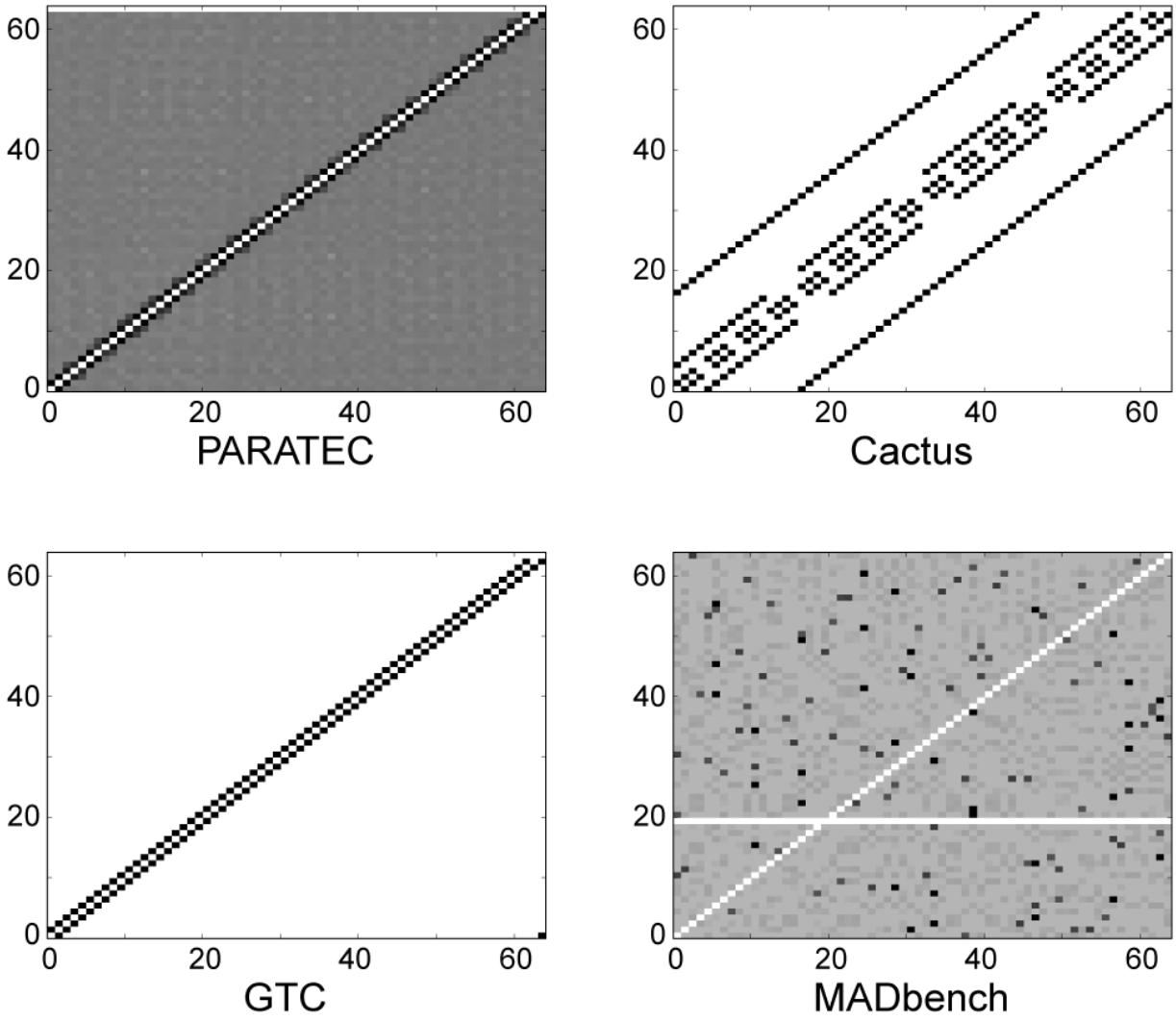

Figure 5.16: Traffic pattern plots for the four scientific applications being considered. Left axis represents source thread ID, bottom axis represents destination thread ID. White blocks represent no communication load while darker shades of gray represent increased traffic load between the associated source-destination pair.

is not necessarily known. For this reason, random thread mappings are used for this simulation work and the mean and standard deviation statistics are reported in the results.

The execution time statistics for the photonic networks and the electronic mesh are shown in Fig. 5.17. The small messages found in the Paratec trace results in the photonic networks having lower performance than the electronic mesh, which is in agreement with the results of the synthetic traffic. However, the photonic networks perform better than the electronic mesh in the remaining three applications as a result of larger message sizes.

The number of WDM partitions varies with each application with respect to shortest execution time. The relatively small message sizes of Paratec and Madbench receive the greatest benefit from using four partitions, achieving 50% and 56% improvements over the single partition case, respectively. Paratec underperforms the electronic mesh due to the circuit-setup overhead associated with the WSSR technique. In the case of Madbench, the execution time reduction compared to the electronic mesh is 89%. GTC uses the largest messages and receives the greatest advantage with a single partition, which results in a time improvement of 89% over electronic mesh. Cactus, which contains messages of an intermediate size, optimally performs with two or three partitions resulting in an improvement of 85% compared to the electronic mesh.

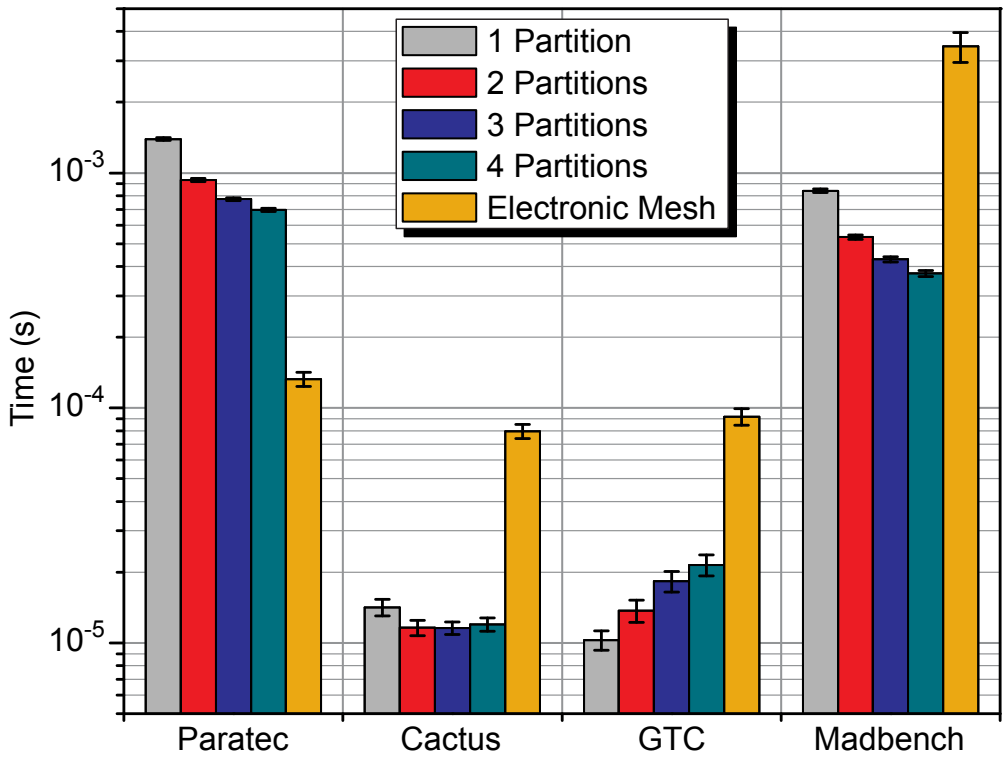

Figure 5.17: Total simulation time required to complete each application trace. Columns indicate the average resulting time, and error bars indicate one standard deviation of the sampled data.

The lackluster performance of Cactus and GTC is also indicative of the limited traffic pattern as indicated in Fig. 5.16. Each source only transmits to a limited number of destinations, therefore the network is less able to exploit the path diversity that is provided with each additional WDM partition. This observation is agreeable with the results from the uniform random traffic where the network was able to achieve better performance with a greater number of partitions since all possible source-destination pairs were utilized creating

Table 5.3: Optical Device Power Parameters - Wavelength-Selective Spatial Routing Analysis

Parameter	Value
Ring Switch Dynamic Energy	375 fJ[1]
Ring Switch Static Energy	400 μW[2]
Modulation Dynamic Energy	85 fJ/bit[3]
Modulation Static Energy	30 W[3]
Detector Energy	50 fJ/bit[4]
Thermal Ring Tuning	100 μW/ring[5]

[1]Calculation based on carrier density, assuming 50-μm diameter, 320-nm $\times$ 250-nm micro-ring waveguide cross-section, 75% waveguide volume exposure, 1-V forward bias.

[2]Based on switching energy, including photon lifetime for re-injection.

[3][35], static energy calculated for half a 10GHz clock cycle, with 50% probability of a 1 bit.

[4]Conservative approximation assuming femto-farad class receiverless SiGe detector with C $<$ 1fF.

[5]Assumes a 1-μW/degree tuning cost per ring, with a temperature deviation of 20 degrees.

more opportunities for the path parallelism to be exploited.

Power parameters for optical devices are summarized in Table 5.3.

Fig. 5.18 depicts the network-level energy efficiency during the runtime of the application traces. Static power dissipation is a major component of the total energy expended, therefore

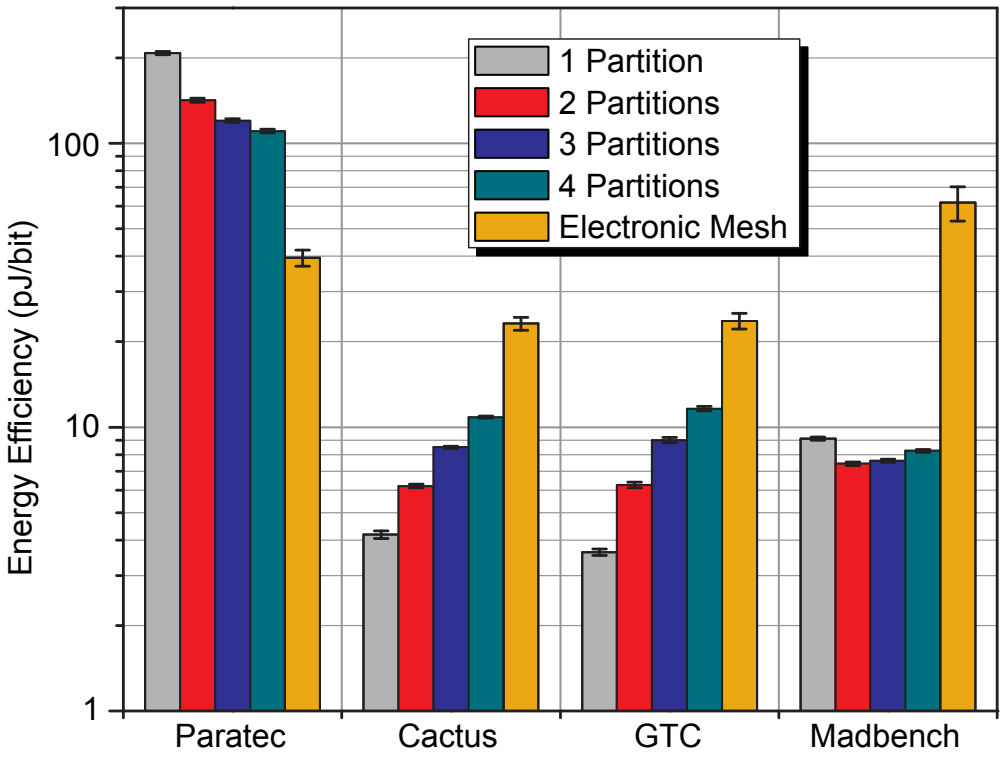

Figure 5.18: Network-level energy efficiency from each application trace. Columns indicate the average resulting energy efficiency, and error bars indicate one standard deviation of the sampled data.

we see a positive correlation between the time and energy results. The photonic networks are able to outperform the electronic mesh in each application except for Paratec. Despite this disadvantage, the photonic network achieves the best energy performance in Paratec using four WDM partitions. Cactus and GTC achieve the best energy performance when using a single partition.

The trace-driven results are summarized in Table 5.4, indicating the best performing number of channels and the percentage improvement. Among the photonic network variants considered, only in Paratec does a four-partition WSSR network perform the best for both execution time and energy. For GTC, the large message sizes benefited the 1-partition network the most by taking advantage of the largest link-level bandwidths and low network-level congestion. This performance dependency on the message size elucidates an opportunity to create a WSSR network design that can dynamically allocate a specific number of channels to optimize network performance.

Table 5.4: Application Trace Results Summary - Wavelength-Selective Spatial Routing Analysis

Execution Time Optimized				
	Paratec	Cactus	GTC	Madbench
Optimal Number of Partitions	4	3	1	4
Improvement	-425%	85%	89%	89%

Energy Dissipation Optimized				
	Paratec	Cactus	GTC	Madbench
Optimal Number of Partitions	4	1	1	2
Improvement	-205%	82%	85%	89%

Chapter 6

Concluding Remarks

The inevitable abandonment of electronic "long distance" wiring has been endlessly predicted, however an effective substitute technology for electronic wires at has yet to be fully developed and commercialized. This is a surprising conclusion since optics now dominates true long distance communications such as across regions or cities. However, "long distance" is no longer across kilometer-scale distances, but is now considered the domain at the computer cluster and computer rack scale. The demand for higher bandwidth, lower latency, and better energy efficiency over distances of meters and centimeters is quickly out pacing what electronics can do at these distances as well. Even distances of millimeters across a chip microprocessor are now being considered too energy intensive for future scalability. We are quickly approaching an era when electronics simply cannot scale, and a true alternative

technology needs to appear.

The adoption of silicon photonics is a logical approach to meeting these performance demands. As was described in this work, as well as countless other publications from many other research groups around the world, photonics has the proven capability of providing huge benefits to computing systems. However, performance is not the reason for this logical outcome. What silicon photonics has is the huge amount of momentum of CMOS electronics due to the amount of invested infrastructure that the computing industry has placed towards it. Unlike alternative exotic material systems for creating computer systems (*e.g.* carbon nanotubes, diamonds, graphene), silicon photonics is completely CMOS compatible, enabling it to be produced in the same multi-billion dollar foundries that exist for regular electronics. This is an extremely attractive approach since the heavily invested infrastructure can continue to be utilized. *Momentum*: it is for this very reason that silicon photonics will likely be the next stepping stone in the future progression of computing.

6.1 Contributions

The main contributions of this work are summarized as follows:

We have described a methodology for modeling, designing, and analyzing photonic interconnection networks at both the physical-layer and system-level. A *Photonic Device*

Library has been devised to describe any type of fundamental photonic elements, which can then be combined and used to model large-scale photonic components and network topologies. We developed a set of physical-layer tools to accurately determine physical properties of the photonic networks and examine how they impact the network architectures in terms of system performance. Our PhoenixSim environment implements this methodology, which we have made open source and publicly available. The device library, analysis tools, and simulation environment form a comprehensive design flow for understanding and designing photonically-enabled computing systems.

Next, we analyzed the physical-layer performance of photonic networks on chip. Previous proposed designs did not consider physical properties such as insertion loss, cross talk, and power dissipation. In fact, many previously proposed network topologies possess such high insertion loss, that it would be impossible to successfully deliver any optical data. With the proposed methodology, we are able to re-examine circuit-switching and wavelength-routed topologies and perform synergistic studies of physical-layer and system-level metrics. We are able to determine realistic design space parameters such as network size scalability and wavelength parallelism. These results bring silicon photonics a step close to reality since designs conform to verified device performances.

This work presented the use of wavelength-selective spatial routing, a novel interconnection

network concept for reducing path diversity and increasing the performance of interconnection networks for CMPs. This design is extensible to previous circuit-switching photonic topologies and is shown to improve network performance for both synthetic and trace traffic in specific cases. We observed that the WSSR architecture is ideally suited for applications with communication patterns that are scattered, enabling the traffic to exploit the path diversity and transmission parallelism that is provided by the spectrally-multiplexed WDM partitions. This work contrasts with almost all other published works on chip-scale photonic networks since they solely rely on redesigning the topology of circuit-switching or wavelength-routed architectures. Instead, this work focuses on an completely new routing architecture. It can be argued that new topologies not useful since it is difficult to predict long-term computing requirements and constraints, but a new architecture provides another conceptual 'tool' which can be utilized in a design in the future.

6.2 Future Work

6.3 Recommendations

Fundamentally, silicon photonics can clearly provide scalable performance improvements for computing systems. However, there are still many realistic engineering challenges that must

be overcome in order for photonics to be fully realized.

First, which were vaguely mentioned in this work, but have become a central discussion point for manufacturing resonator-based photonic devices in temperature tolerance. As photonic device fabrication becomes mature and components increase in complexity, sensitivity to temperature fluctuation becomes an dominant concern. Due to the characteristics of a resonator, small perturbations in temperature results in large shifts in spectral response. Recent work has shown the utilizing of a PID controller for on-line temperature stablization [100]. Alternatively, novel athermal resonators have also been proposed which reduces thermal dependence of the device [101]. Although there is excellent progress, significantly better techniques need to be devised so that chip-scale photonic systems can be reliably deployed.

In addition to thermal perturbation mitigation, a method for post-fabrication tuning must also be utilized to compensate for fabrication variances. The behavior of optical components is extremely sensitive to variations in geometry dimensions which can to easily be avoided or corrected. In many of the networks considered in this work, a single optical link could have upwards of over a hundred optical components along the path. Each device would need to be exactly tuned correctly so that the link can behave as expected. A possible solution is to use the effects of thermal fluctuation to counteract manufacturing variation.

The final, but possibly the most important challenge towards commercialization is packaging. While packaging exists for discrete optical components such as modulators, detectors, lasers, amplifiers, etc., none yet exist for chip-scale silicon photonics. The packaging solution must meet two objectives: to be able to couple a large number of waveguides, and to be stable in the field. State-of-the-art fiber coupling solutions include multi-core fiber and fiber array connectors [102]. Future iterations of these multi-fiber-core technologies are expected to scale much farther. Alignment of fiber coupling to connect the chip will be critical, which must deal with a plethora of environmental influences such as mechanical motion, thermal fluctuation, and material deformation. Additionally, the low loss behavior of silicon photonics requires single-mode optics, which requires alignment tolerances of only a few microns.

None of these challenges represent fundamental physical limitations, therefore each is a surmountable research barriers that must first be conquered. It is only a matter of time and investment before silicon photonics technology becomes a commercial reality.

References

[1] J. Kahle. **The Cell Processor Architecture**. In *Microarchitecture, 2005. MICRO-38. Proceedings. 38th Annual IEEE/ACM International Symposium on*, page 3, nov. 2005. 2

[2] D. Wentzlaff, P. Griffin, H. Hoffmann, Liewei Bao, B. Edwards, C. Ramey, M. Mattina, Chyi-Chang Miao, J.F. Brown, and A. Agarwal. **On-Chip Interconnection Architecture of the Tile Processor**. *Micro, IEEE*, **27**(5):15–31, sept.-oct. 2007. 2, 11

[3] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, P. Iyer, A. Singh, T. Jacob, S. Jain, S. Venkataraman, Y. Hoskote, and N. Borkar. **An 80-Tile 1.28TFLOPS Network-on-Chip in 65nm CMOS**. In *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pages 98 –589, feb. 2007. 2

[4] Krste Asanovic, Ras Bodik, Bryan Christopher Catanzaro, Joseph James Gebis, Parry Husbands, Kurt Keutzer, David A. Patterson, William Lester Plishker, John Shalf, Samuel Webb Williams, and Katherine A. Yelick. **The Landscape of Parallel Computing Research: A View from Berkeley**. *EECS Department, University of California, Berkeley*, (UCB/EECS-2006-183), Dec. 2006. [Online]: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html`. 3

[5] J.D. Meindl. **Interconnect opportunities for gigascale integration**. *Micro, IEEE*, **23**(3):28–35, May-June 2003. 4

[6] R. Ho, K.W. Mai, and M.A. Horowitz. **The future of wires**. *Proceedings of the IEEE*, **89**(4):490 –504, Apr. 2001. 4

[7] Nir Magen, Avinoam Kolodny, Uri Weiser, and Nachum Shamir. **Interconnect-power dissipation in a microprocessor**. In *Proceedings of the 2004 international workshop on System level interconnect prediction*, SLIP '04, pages 7–13, 2004. 4

[8] A. F. Benner, M. Ignatowski, J. A. Kash, D. M. Kuchta, and M. B. Ritter. **Exploitation of optical interconnects in future server architectures**. *IBM Journal of Research and Development*, **49**(4.5):755 –775, july 2005. 4

[9] D. Miller. **Device Requirements for Optical Interconnects to Silicon Chips**. *Proceedings of the IEEE*, **97**(7):1166–1185, July 2009. 5

[10] M. Salib, L. Liao, R. Jones, M. Morse, A. Liu, D. Samara-Rubio, D. Alduino, and M. Paniccia. **Silicon photonics**. *Intel Technology Journal*, **8**(2):1442, 2004. 5

[11] M. Lipson. **Guiding, modulating, and emitting light on Silicon-challenges and opportunities**. *Lightwave Technology, Journal of*, **23**(12):4222 – 4238, Dec. 2005. 5

[12] C. Gunn. **CMOS photonics trade; - SOI learns a new trick**. In *SOI Conference, 2005. Proceedings. 2005 IEEE International*, pages 7 – 13, oct. 2005. 5

[13] G. Hendry, E. Robinson, V. Gleyzer, J. Chan, L. Carloni, N. Bliss, and K. Bergman. **Circuit-Switched Memory Access in Photonic Interconnection Networks for High-Performance Embedded Computing**. In *High Performance Computing, Networking, Storage and Analysis (SC), 2010 International Conference for*, pages 1–12, nov. 2010. 11

[14] B.G. Lee, C.L. Schow, A.V. Rylyakov, J.V. Van Campenhout, W.M.J. Green, S. Assefa, F.E. Doany, Min Yang, R.A. John, C.V. Jahnes, J.A. Kash, and Y.A. Vlasov. **Demonstration of a Digital CMOS Driver Codesigned and Integrated With a Broadband Silicon Photonic Switch**. *Lightwave Technology, Journal of*, **29**(8):1136 –1142, april15, 2011. 11

[15] International Technology Roadmap for Semiconductors. **2010 Report**. [Online]: `http://www.itrs.net`. 12

[16] A Demircan, Sh Amiranashvili, and G Steinmeyer. **Controlling light by light with an optical event horizon**. *Physical Review Letters*, **106**(16):163901, 2011. 19

[17] J. Hwang, M. Pototschnig, R. Lettow, G. Zumofen, A. Ren, S. Gotzinger, and V. Sandoghdar. **A single-molecule optical transistor**. *Nature*, **460**:76–80, 2009. 19

[18] D. Van Thourhout, P. Dumon, W. Bogaerts, G. Roelkens, D. Taillaert, G. Priem, and R. Baets. **Recent progress in SOI nanophotonic waveguides**. In *Optical Communication, 2005. ECOC 2005. 31st European Conference on*, **2**, pages 241–244, Sept. 2005. 20, 21

[19] Fengnian Xia, Lidija Sekaric, and Yurii Vlasov. **Ultracompact optical buffers on a silicon chip**. *Nature Photonics*, **1**:65–71, 2006. 21, 29, 78, 84, 114, 147

[20] M. Gnan, S. Thorns, D.S. Macintyre, R.M. De La Rue, and M. Sorel. **Fabrication of low-loss photonic wires in silicon-on-insulator using hydrogen silsesquioxane electron-beam resist**. *Electronics Letters*, **44**(2):115–116, Jan. 2008. 21, 29

[21] Jaime Cardenas, Carl B. Poitras, Jacob T. Robinson, Kyle Preston, Long Chen, and Michal Lipson. **Low loss etchless silicon photonic waveguides**. *OSA Optics Express*, **17**(6):4752–4757, 2009. 21

[22] Michael J. Shaw, Junpeng Guo, Gregory A. Vawter, Scott Habermehl, and Charles T. Sullivan. **Fabrication techniques for low-loss silicon nitride waveguides**. *Micromachining Technology for Micro-Optics and Nano-Optics III*, **5720**(1):109–118, 2005. 21

[23] Wim Bogaerts, Pieter Dumon, Dries Van Thourhout, and Roel Baets. **Low-loss, low-cross-talk crossings for silicon-on-insulator nanophotonic waveguides**. *OSA Optics Letters*, **32**(19):2801–2803, 2007. 22, 29, 50, 78, 84, 98, 114, 119, 147

[24] Sharee McNab, Nikolaj Moll, and Yurii Vlasov. **Ultra-low loss photonic integrated circuit with membrane-type photonic crystal waveguides**. *Opt. Express*, **11**(22):2927–2939, Nov 2003. 23

[25] Vilson R. Almeida, Roberto R. Panepucci, and Michal Lipson. **Nanotaper for compact mode conversion**. *OSA Optics Letters*, **28**(15):1302–1304, 2003. 23, 147

[26] T. Tsuchizawa, K. Yamada, H. Fukuda, T. Watanabe, Jun ichi Takahashi, M. Takahashi, T. Shoji, E. Tamechika, S. Itabashi, and H. Morita. **Microphotonics devices based on silicon microfabrication technology**. *Selected Topics in Quantum Electronics, IEEE Journal of*, **11**(1):232 – 240, jan.-feb. 2005. 23

[27] Dirk Taillaert, Peter Bienstman, and Roel Baets. **Compact efficient broadband grating coupler for silicon-on-insulator waveguides**. *OSA Optics Letters*, **29**(23):2749–2751, Dec 2004. 24

[28] Günther Roelkens, Dries Van Thourhout, and Roel Baets. **High efficiency grating coupler between silicon-on-insulator waveguides and perfectly vertical optical fibers**. *OSA Optics Letters*, **32**(11):1495–1497, 2007. 24, 147

[29] B.E. Little, J.S. Foresi, G. Steinmeyer, E.R. Thoen, S.T. Chu, H.A. Haus, E.P. Ippen, L.C. Kimerling, and W. Greene. **Ultra-compact Si-SiO_2 microring resonator optical channel dropping filters**. *IEEE Photonics Technology Letters*, **10**(4):549–551, Apr. 1998. 24, 29, 55

[30] B.G. Lee, A. Biberman, Po Dong, M. Lipson, and K. Bergman. **All-Optical Comb Switch for Multiwavelength Message Routing in Silicon Photonic Networks**. *IEEE Photonics Technology Letters*, **20**(10):767–769, May 2008. 24, 29, 56, 77, 78, 84, 98, 114, 119, 127, 129, 147

[31] S. Manipatruni, Qianfan Xu, B. Schmidt, J. Shakya, and M. Lipson. **High Speed Carrier Injection 18 Gb/s Silicon Micro-ring Electro-optic Modulator**. In *The 20th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, pages 537–538, Oct. 2007. 24, 27, 29

[32] H.L.R. Lira, S. Manipatruni, and M. Lipson. **Broadband hitless silicon electro-optic switch for optical networks on-chip**. In *Group IV Photonics, 2009. GFP '09. 6th IEEE International Conference on*, pages 253–255, sept. 2009. 24, 27, 138

[33] Yurii Vlasov, William M. J. Green, and Fengnian Xia. **High-throughput silicon nanophotonic wavelength-insensitive switch for on-chip optical networks**. *Nature Photonics*, **2**:242–246, Apr. 2008. 24, 56

[34] Qianfan Xu, Sasikanth Manipatruni, Brad Schmidt, Jagat Shakya, and Michal Lipson. **12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators**. *OSA Optics Express*, **15**(2):430–436, 2007. 24, 57, 97

[35] M.R. Watts, D.C. Trotter, R.W. Young, and A.L. Lentine. **Ultralow power silicon microdisk modulators and switches**. In *5th IEEE International Conference on Group IV Photonics*, pages 4–6, Sept. 2008. 24, 164

[36] M. Lipson. **Compact Electro-Optic Modulators on a Silicon Chip**. *IEEE Journal of Selected Topics in Quantum Electronics (JSTQE)*, **12**(6):1520–1526, Nov.–Dec. 2006. 27

[37] W.M.J. Green, H.F. Hamann, L. Sekaric, M.J. Rooks, and Y.A. Vlasov. **Ultra-compact reconfigurable silicon optical devices using micron-scale localized thermal heating**. In *Proceedings of Optical Fiber Communication Conference (OFC)*, Mar. 2007. 27

[38] K. Preston, N. Sherwood-Droz, J.S. Levy, and M. Lipson. **Performance guidelines for WDM interconnects based on silicon microring resonators**. In *Lasers and Electro-Optics (CLEO), 2011 Conference on*, May 2011. 27, 129, 130

[39] Junbo Feng, Qunqing Li, and Zhiping Zhou. **Single Ring Interferometer Configuration With Doubled Free-Spectral Range**. *Photonics Technology Letters, IEEE*, **23**(2):79–81, Jan. 2011. 28

[40] J. Garcia, A. Martinez, and J. Marti. **Optical add-drop multiplexer with FSR higher than 140 nm using ring resonators and photonic bandgap structures**. In *Group IV Photonics, 2008 5th IEEE International Conference on*, pages 82–84, Sept. 2008. 28

[41] Y. Yanagase, S. Suzuki, Y. Kokubun, and Sai Tak Chu. **Box-like filter response and expansion of FSR by a vertically triple coupled microring resonator filter**. *Lightwave Technology, Journal of*, **20**(8):1525–1529, Aug. 2002. 28

[42] Solomon Assefa, Fengnian Xia, Stephen W. Bedell, Ying Zhang, Teya Topuria, Philip M. Rice, and Yurii A. Vlasov. **CMOS-Integrated 40GHz Germanium Waveguide Photodetector for On-Chip Optical Interconnects**. In *Proceedings of Optical Fiber Communication Conference (OFC)*, page OMR4, 2009. 28, 58

[43] Laurent Vivien, Johann Osmond, Jean-Marc Fédéli, Delphine Marris-Morini, Paul Crozat, Jean-François Damlencourt, Eric Cassan, Y. Lecunff, and Suzanne Laval. **42 GHz p.i.n Germanium photodetector integrated in a silicon-on-insulator waveguide**. *OSA Optics Express*, **17**(8):6252–6257, 2009. 28, 58

[44] Solomon Assefa, Benjamin G. Lee, Clint Schow, William M. Green, Alexander Rylyakov, Richard A. John, and Yurii A. Vlasov. **20Gbps Receiver Based on Germanium Photodetector Hybrid-Integrated with 90nm CMOS Amplifier**. In *CLEO:2011 - Laser Applications to Photonic Applications*, page PDPB11. Optical Society of America, 2011. 29, 147

[45] N. Ophir, K. Padmaraju, A. Biberman, L. Chen, K. Preston, M. Lipson, and K. Bergman. **First Demonstration of Error-Free Operation of a Full Silicon On-Chip Photonic Link**. In *Optical Fiber Communication Conference*, page OWZ3. Optical Society of America, 2011. 31

[46] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C.W. Holzwarth, M.A. Popovic, Hanqing Li, H.I. Smith, J.L. Hoyt, F.X. Kartner, R.J. Ram, V. Stojanovic, and K. Asanovic. **Building Many-Core Processor-to-DRAM Networks with Monolithic CMOS Silicon Photonics**. *IEEE Micro*, **29**(4):8–21, July-Aug. 2009. 31

[47] Nevin Kirman, Meyrem Kirman, Rajeev K. Dokania, Jose F. Martinez, Alyssa B. Apsel, Matthew A. Watkins, and David H. Albonesi. **On-Chip Optical Technology in Future Bus-Based Multicore Designs**. *IEEE Micro*, **27**(1):56–66, 2007. 31

[48] Mark J. Cianchetti, Joseph C. Kerekes, and David H. Albonesi. **Phastlane: a rapid transit optical routing network**. In *Proceedings of the 36th Annual International Symposium on Computer Architecture (ISCA)*, pages 441–450, 2009. 31

[49] Dana Vantrease, Robert Schreiber, Matteo Monchiero, Moray McLaren, Norman P. Jouppi, Marco Fiorentino, Al Davis, Nathan Binkert, Raymond G. Beausoleil, and Jung Ho Ahn. **Corona: System Implications of Emerging Nanophotonic Technology**. *Proceedings of the 35th Annual International Symposium on Computer Architecture (ISCA)*, **0**:153–164, June 2008. 31

[50] S. Koohi, M. Abdollahi, and S. Hessabi. **All-optical wavelength-routed NoC based on a novel hierarchical topology**. In *Networks on Chip (NoCS), 2011 Fifth IEEE/ACM International Symposium on*, pages 97–104, May 2011. 31, 131, 137

[51] A. Shacham, K. Bergman, and L.P. Carloni. **Photonic Networks-on-Chip for Future Generations of Chip Multiprocessors**. *IEEE Transactions on Computers*, **57**(9):1246–1260, Sept. 2008. 31, 37, 40, 75, 81, 130, 131, 151

[52] G. Hendry, S. Kamil, A. Biberman, J. Chan, B.G. Lee, M. Mohiyuddin, A. Jain, K. Bergman, L.P. Carloni, J. Kubiatowicz, L. Oliker, and J. Shalf. **Analysis of photonic networks for a chip multiprocessor using scientific applications**. In *Proceedings of the 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, pages 104–113, May 2009. 31, 32, 41, 101, 123

[53] J. Chan, G. Hendry, A. Biberman, and K. Bergman. **Architectural Exploration of Chip-Scale Photonic Interconnection Network Designs Using Physical-Layer Analysis**. *IEEE/OSA Journal of Lightwave Technology*, **28**(9):1305–1315, May 2010. 31, 43, 109, 112, 138, 152

[54] G. Hendry, J. Chan, S. Kamil, L. Oliker, J. Shalf, L.P. Carloni, and K. Bergman. **Silicon Nanophotonic Network-on-Chip Using TDM Arbitration**. In *2010 IEEE 18th Annual Symposium on High Performance Interconnects (HOTI)*, pages 88–95, Aug. 2010. 32, 158

[55] J. Chan, N Ophir, C. P. Lai, A. Biberman, H. L. R. Lira, M Lipson, and K. Bergman. **Data Transmission Using Wavelength-Selective Spatial Routing for Photonic Interconnection Networks**. *Optical Fiber Communication Conference*, Mar. 2011. 32, 126, 138

[56] Johnnie Chan and Keren Bergman. **Photonic Interconnection Network Architectures Using Wavelength-Selective Spatial Routing for Chip-Scale Communications**. *J. Opt. Commun. Netw.*, **4**(3):189–201, Mar. 2012. 32, 125

[57] Chunming Qiao and R. Melhem. **Reducing communication latency with path multiplexing in optically interconnected multiprocessor systems**. *Parallel and Distributed Systems, IEEE Transactions on*, **8**(2):97–108, Feb. 1997. 33

[58] X. Yuan, R. Melhem, and R. Gupta. **Distributed path reservation algorithms for multiplexed all-optical interconnection networks**. *Computers, IEEE Transactions on*, **48**(12):1355–1363, Dec. 1999. 33, 136

[59] Duo Ding and David Z. Pan. **OIL: a nano-photonics optical interconnect library for a new photonic networks-on-chip architecture**. In *Proceedings of the 11th International Workshop on System Level Interconnect Prediction (SLIP)*, pages 11–18, July 2009. 34

[60] J.R. Minz, S. Thyagaraja, and Sung Kyu Lim. **Optical Routing for 3D System-On-Package**. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, **1**, pages 1–2, Mar. 2006. 34

[61] Gilbert Hendry, Johnnie Chan, Luca P. Carloni, and Keren Bergman. **VANDAL: A Tool for the Design Specification of Nanophotonic Networks**. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, Mar. 2011. 34

[62] Ian O'Connor, Faress Tissafi-Drissi, Frédéric Gaffiot, Joni Dambre, Michiel De Wilde, Joris Van Campenhout, Dries Van Thourhout, Jan Van Campenhout, and Dirk Stroobandt. **Systematic simulation-based predictive synthesis of integrated optical interconnect**. *IEEE Trans. Very Large Scale Integr. Syst.*, **15**:927–940, Aug. 2007. 35

[63] Michiel De Wilde, Olivier Rits, Wim Meeus, Hannes Lambrecht, and Jan Van Campenhout. **Integration of Modeling Tools for Parallel Optical Interconnects in a Standard EDA Design Environment**. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, Feb. 2004. 35

[64] P. K. Pepeljugoski and D. M. Kuchta. **Design of optical communications data links**. *IBM Journal of Research and Development*, **47**(2.3):223–237, Mar. 2003. 35

[65] Matthieu Briere, Emmanuel Drouard, Fabien Mieyeville, David Navarro, Ian O'Connor, and Frederic Gaffiot. **Heterogeneous Modelling of an Optical Network-on-Chip with SystemC**. In *Proceedings of the 16th IEEE International Workshop on Rapid System Prototyping (RSP)*, pages 10–16, June 2005. 35

[66] A.K. Kodi and A. Louri. **Optisim: A System Simulation Methodology for Optically Interconnected HPC Systems**. *IEEE Micro*, **28**(5):22–36, Sept.–Oct. 2008. 35, 38

[67] J. Chan, G. Hendry, A. Biberman, K. Bergman, and L.P. Carloni. **PhoenixSim: A simulator for physical-layer analysis of chip-scale photonic interconnection networks**. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pages 691 –696, March 2010. 36

[68] J. Chan, G. Hendry, K. Bergman, and L.P. Carloni. **Physical-Layer Modeling and System-Level Design of Chip-Scale Photonic Interconnection Networks**. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, **30**(10):1507–1520, Oct. 2011. 36, 155

[69] Photonic and Electronic Network Integration and Execution Simulator (PhoenixSim). [Online]: `http://lightwave.ee.columbia.edu/phoenixsim`. 36

[70] András Varga and Rudolf Hornig. **An overview of the OMNeT++ simulation environment**. In *Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, pages 1–10, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 36

[71] Andras Varga. **OMNeT++ Discrete Event Simulation System**. [Online]: `http://www.omnetpp.org`. 36

[72] Hang-Sheng Wang, Xinping Zhu, Li-Shiuan Peh, and Sharad Malik. **Orion: a power-performance simulator for interconnection networks**. In *Proceedings of the 35th Annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, pages 294–305, Nov. 2002. 39, 67, 68, 101

[73] Dana Vantrease, Robert Schreiber, Matteo Monchiero, Moray McLaren, Norman P. Jouppi, Marco Fiorentino, Al Davis, Nathan Binkert, Raymond G. Beausoleil, and Jung Ho Ahn. **Corona: System Implications of Emerging Nanophotonic Technology**. *Proceedings of the 35th Annual International Symposium on Computer Architecture (ISCA)*, pages 153–164, June 2008. 40

[74] Po Dong, Stefan F. Preble, and Michal Lipson. **All-optical compact silicon comb switch**. *Opt. Express*, **15**(15):9600–9605, Jul 2007. 42

[75] A. Sakai, G. Hara, and T. Baba. **Large effective index and low bend loss in SOI optical waveguides**. In *Lasers and Electro-Optics, 2001. CLEO/Pacific Rim 2001. The 4th Pacific Rim Conference on*, **1**, pages I–4 –I–5 vol.1, 2001. 52

[76] Corning Incorporated. **Corning SMF-28e+ Optical Fiber Product Information**, July 2011. [Online]: `http://www.corning.com/WorkArea/showcontent.aspx?id=41261`. 52

[77] Hui-Wen Chen, Ying-Hao Kuo, and John E. Bowers. **High speed hybrid silicon evanescent Mach-Zehnder modulator and switch**. *OSA Optics Express*, **16**(25):20571–20576, 2008. 58

[78] Tobias Gensty, Wolfgang Elsässer, and Christian Mann. **Intensity noise properties of quantum cascade lasers**. *OSA Optics Express*, **13**(6):2032–2039, 2005. 63, 97, 98, 119

[79] Christopher Miller. *Fiber Optic Test and Measurement*. Prentice Hall, 1998. 64

[80] David Wang, Brinda Ganesh, Nuengwong Tuaycharoen, Kathleen Baynes, Aamer Jaleel, and Bruce Jacob. **DRAMsim: a memory system simulator**. *SIGARCH Comput. Archit. News*, **33**(4):100–107, November 2005. 68

[81] Wei Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M.R. Stan. **HotSpot: a compact thermal modeling methodology for early-stage VLSI design**. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, **14**(5):501 –513, may 2006. 68

[82] Kevin Skadron, Mircea R. Stan, Karthik Sankaranarayanan, Wei Huang, Sivakumar Velusamy, and David Tarjan. **Temperature-aware microarchitecture: Modeling and implementation**. *ACM Trans. Archit. Code Optim.*, **1**(1):94–125, March 2004. 68

[83] S.R. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, A. Singh, T. Jacob, S. Jain, V. Erraguntla, C. Roberts, Y. Hoskote, N. Borkar, and S. Borkar. **An 80-Tile Sub-100-W TeraFLOPS Processor in 65-nm CMOS**. *Solid-State Circuits, IEEE Journal of*, **43**(1):29 –41, Jan. 2008. 76

[84] A. Shacham, B.G. Lee, A. Biberman, K. Bergman, and L.P. Carloni. **Photonic NoC for DMA Communications in Chip Multiprocessors**. In *High-Performance Interconnects, 2007. HOTI 2007. 15th Annual IEEE Symposium on*, pages 29–38, Aug. 2007. 79

[85] Howard Wang, Michele Petracca, Aleksandr Biberman, Benjamin G. Lee, Luca P. Carloni, and Keren Bergman. **Nanophotonic Optical Interconnection Network Architecture for On-Chip and Off-Chip Communications**. In *Optical Fiber Communication Conference and Exposition and The National Fiber Optic Engineers Conference*, page JThA92. Optical Society of America, 2008. 81, 82

[86] Nicolás Sherwood-Droz, Howard Wang, Long Chen, Benjamin G. Lee, Aleksandr Biberman, Keren Bergman, and Michal Lipson. **Optical 4x4 hitless slicon router for optical networks-on-chip (NoC)**. *Opt. Express*, **16**(20):15915–15922, Sep 2008. 83

[87] Mutsunori Uenuma and Teruaki Motooka. **Temperature-independent silicon waveguide optical filter**. *Opt. Lett.*, **34**(5):599–601, Mar 2009. 83

[88] Kyle Preston, Sasikanth Manipatruni, Alexander Gondarenko, Carl B. Poitras, and Michal Lipson. **Deposited silicon high-speed integrated electro-optic modulator**. *OSA Optics Express*, **17**(7):5118–5124, 2009. 97, 98, 119

[89] P. Lathi. *Modern Digital and Analog Communication Systems*. Oxford University Press, third edition, 1998. 100

[90] Michele Petracca, Benjamin G. Lee, Keren Bergman, and Luca P. Carloni. **Photonic NoCs: System-Level Design Exploration**. *IEEE Micro*, **29**:74–85, 2009. 109

[91] J. Chan, A. Biberman, B.G. Lee, and K. Bergman. **Insertion loss analysis in a photonic interconnection network for on-chip and off-chip communications**. In *21st Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, pages 300–301, Nov. 2008. 112

[92] Ajay Joshi, Christopher Batten, Yong-Jin Kwon, Scott Beamer, Imran Shamim, Krste Asanovic, and Vladimir Stojanovic. **Silicon-photonic clos networks for global on-chip communication**. In *Proceedings of the 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, pages 124–133, May 2009. 112

[93] Qianfan Xu, David Fattal, and Raymond G. Beausoleil. **Silicon microring resonators with 1.5-μm radius**. *Opt. Express*, **16**(6):4309–4315, Mar. 2008. 129, 151

[94] Aleksandr Biberman, Hugo L. Lira, Kishore Padmaraju, Noam Ophir, Michal Lipson, and Keren Bergman. **Broadband CMOS-Compatible Silicon Photonic Electro-Optic Switch for Photonic Networks-on-Chip**. In *Conference on Lasers and Electro-Optics*, page CPDA11. Optical Society of America, 2010. 138

[95] Aleksandr Biberman, Kyle Preston, Gilbert Hendry, Nicolás Sherwood-Droz, Johnnie Chan, Jacob S. Levy, Michal Lipson, and Keren Bergman. **Photonic network-on-chip architectures using multilayer deposited silicon materials for high-performance chip multiprocessors**. *J. Emerg. Technol. Comput. Syst.*, **7**:7:1–7:25, July 2011. 148

[96] A. Canning, L.W. Wang, A. Williamson, and A. Zunger. **Parallel Empirical Pseudopotential Electronic Structure Calculations for Million Atom Systems**. *Journal of Computational Physics*, **160**(1):29 – 41, 2000. 159

[97] Cactus Computational Toolkit. [Online]:http://www.cactuscode.org/. 159

[98] Z. Lin, S. Ethier, T. S. Hahm, and W. M. Tang. **Size Scaling of Turbulent Transport in Magnetically Confined Plasmas**. *Physical Review Letters*, **88**(19):195004, Apr 2002. 159

[99] J. Borrill, J. Carter, L. Oliker, and D. Skinner. **Integrated Performance Monitoring of a Cosmology Application on Leading HEC Platforms**. In *Proceedings of the 2005 International Conference on Parallel Processing*, pages 119–128, 2005. 159

[100] Kishore Padmaraju, Johnnie Chan, Long Chen, Michal Lipson, and Keren Bergman. **Dynamic Stabilization of a Microring Modulator Under Thermal Perturbation**. In *Optical Fiber Communication Conference*, page OW4F.2. Optical Society of America, 2012. 172

[101] Biswajeet Guha, Bernardo B. C. Kyotoku, and Michal Lipson. **CMOS-compatible athermal silicon microring resonators**. *Opt. Express*, **18**(4):3487–3493, Feb 2010. 172

[102] B.G. Lee, C. Baks, F.E. Doany, C. Jahnes, R. John, D.M. Kuchta, P. Pepeljugoski, A.V. Rylyakov, C.L. Schow, S. Assefa, W.M.J. Green, Y.A. Vlasov, and J.A. Kash. **Increasing bandwidth density in future optical interconnects**. In *Photonics Conference (PHO), 2011 IEEE*, pages 670 –671, oct. 2011. 173