Chip-Scale Photonic Architectures using Wavelength-Selective Spatial Routing for High-Performance Interconnection Networks

Johnnie Chan and Keren Bergman
Department of Electrical Engineering
Columbia University
New York, NY 10027, USA
johnnie@ee.columbia.edu

***Abstract*— Modern chip-scale computing system performance is increasingly becoming determined by the characteristics of the interconnection network. Photonic technology has been proposed as an alternative to traditional electronic interconnects for its advantages in bandwidth density, latency, and power efficiency. Circuit-switched photonic network architectures take advantage of the optical spectrum to create high-bandwidth transmission links through the transmission of data channels on multiple parallel wavelengths. However, this technique suffers from low path diversity and high setup-time overhead, which consequentially induces high network resource contention and long latencies. This work improves upon the circuit-switching paradigm by introducing the use of wavelength-selective spatial routing to produce multiple logical communication layers on a single physical photonic plane. With a photonic system leveraging 120 wavelength channels, this technique is shown in simulation to achieve as much as 169% and 764% saturation bandwidth improvement over the previous photonic circuit-switching design and the electronic mesh, respectively.**

I. INTRODUCTION

Future high-performance computing (HPC) systems will be increasingly met with more stringent physical demands and growing performance requirements to meet the needs of emerging HPC applications [1]. Performance requirements have thus far been obtained through the increased parallelism of chip multiprocessors (CMPs) and greater memory capacity; however, the current trend in processor and memory scaling will ultimately be met with fundamental roadblocks. Interconnecting the cores on a CMP is becoming increasingly difficult due to physical limitations in wire densities and packaging power dissipation. Moreover, memory and IO communications are also being bottlenecked by pin density and limited surface area on the chip package. These challenges have led system architects to investigate alternative technologies for handling the communication infrastructure required by HPC systems.

Integrated silicon photonics is currently seen as an attractive interconnect solution for mitigating the bandwidth and energy bottlenecks that are facing HPC systems. Photonic interconnects offers orders-of-magnitude improved bandwidth density over electronics by leveraging wavelength-division multiplexing (WDM) to concurrently transmit multiple spectrally-parallel streams of data through a single optical waveguide. Photonics also affords improved energy efficiency by eliminating the need for electronic buffers and switches that are required within typical electronic networks. These notable advantages have prompted the proposal of many novel photonics-enabled interconnect architectures [2]–[4].

This work builds upon the previously proposed circuit-switched photonic interconnection network architectures [4]. This type of architecture has been shown to improve network efficiency by at least an order-of-magnitude in a variety of signal processing and HPC scientific applications [5], [6]. This work further improves the performance of photonic circuit-switching architectures by introducing a new photonic chip-scale routing concept called *wavelength-selective spatial routing*. This new routing concept allows multiple communication circuits to be multiplexed onto a single physical link. The remainder of this paper is used to describe the wavelength-selective spatial routing concept, and present simulation results showing upwards of 169% gain in saturation-bandwidth performance over previous photonic circuit-switching interconnect designs, and 764% improvement over a standard electronic mesh topology.

II. WAVELENGTH-SELECTIVE SPATIAL ROUTING

Photonic circuit-switching architectures leverage electro-optically controlled ring-resonator switches to direct the flow of optical signals. The ring acts as an active comb filter, which allows multiple unique wavelengths of light to be simultaneously switched. The through state (Fig. 1a) allows optical signals to propagate past a ring unperturbed, while the drop state (Fig. 1b) will direct all wavelength channels onto a secondary waveguide. Each wavelength of light is tuned to a unique resonance peak of the ring resonator such that they are

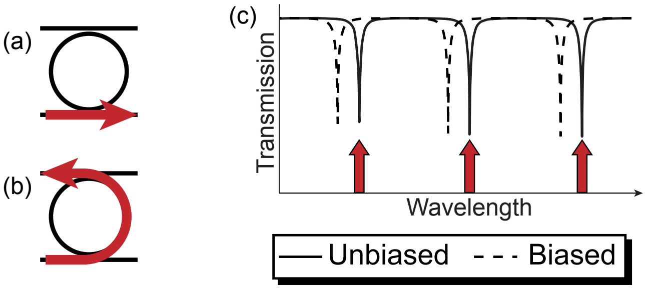

Figure 1. Previously proposed photonic circuit-switching technique. The (a) off-resonance through state and (b) on-resonance drop state of a ring resonator. (c) The positioning of transmission wavelengths relative to the through-port spectrum of the ring switch, with and without voltage bias.

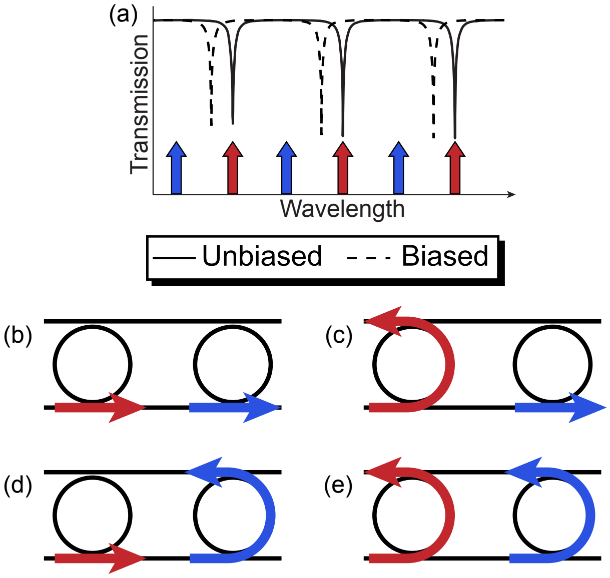

Figure 2. Two-WDM-partition router configuration. (a) Positioning of transmission wavelengths of two sets of wavelengths (each composing a WDM partition) relative to one ring of the router. One partition is colored in red, the other colored in blue. This router requires two cascaded rings and exhibits four states: (b) through-through, (c) drop-through, (d) through-drop, and (e) drop-drop.

simultaneously switched between the two states when the applied voltage bias is changed (Fig. 1c). Since the intended destination of an optical data signal cannot be readily interpreted optically, the complete source-destination optical path must be established before the signal can be transmitted. This work examines photonic circuit-switching-style networks, which require provisioning and de-provisioning of lightpaths through the transmission of control messages on a separate electronic control plane, a technique which has been previously proposed [4].

The fundamental nature of circuit-switched networks (both photonic and electronic) require the reservation of resources for the duration of a communication flow. In a simple network architecture, resource reservation will cause contention if multiple network nodes require common resources. To mitigate this issue, electronic circuit-switched networks can leverage virtual channels to *time multiplex* several communication flows through a single physical electronic bus. However, time multiplexing of optical data would require an uneconomical conversion to the electronic domain and additional logic. We alternatively propose the use of wavelength-selective spatial routing which uses *spectral multiplexing* to create several concurrent communication links within a single waveguide.

The presented architecture utilizes the wavelength-selective spatial routing technique to accomplish spectral multiplexing of communication links [7]. This transmission scheme leverages the unused spectrum that exists between the resonances of a broadband ring. This unallocated spectrum is depicted in Fig. 1c as high transmission regions that remain unperturbed, regardless of the voltage bias. Given this behavior, transmission wavelengths are chosen in such a way that the ring resonator manipulates a partial set of wavelength channels, while others completely ignored (Fig. 2a). A second ring resonator can then be cascaded and aligned to the ignored transmission wavelengths. This enables two sets of transmission wavelengths to be routed independently of each other (Fig. 2b-2e). Each of the two sets of wavelength channels in the example are labeled as *WDM partitions*, alluding to the fact that the switched channels represent just a subset of all the channels used in the system. Fig. 2 illustrates the routing configuration for two WDM partitions, however the concept can be extended to higher partition counts. The degenerate case of a single WDM partition represents the original circuit-switching concept. Additionally, since the number and relative location of the input and output ports remain the same with any number of partitions, previous circuit switch designs can be easily altered to take advantage of this novel concept.

The previously proposed technique for provisioning and de-provisioning of lightpaths has been augmented to support the partitioned WDM architecture. Each WDM partition represents an independent photonic transmission plane, therefore each node is capable of simultaneously provisioning and transmitting multiple independent data streams. For a WDM-partitioned architecture with *N* partitions, a single source node will assign a message that is awaiting transmission to a single partition and continuously attempt to allocate on that path until it is successful. This provisioning policy allows for *N* different path-setup attempts simultaneously, assuming that the node is not already transmitting on any partition. An alternative technique is to leverage all *N* WDM partition concurrently for the path-setup of a single message to increase the likelihood of a successful path allocation. In the case that multiple partitions are simultaneously successful, a single partition is chosen and the remainder are released for a subsequent message to use in its path setup.

III. SIMULATION SETUP AND RESULTS

The partitioned-WDM network architecture is modeled and simulated in PhoenixSim, a chip-scale photonic network simulation environment [8]. Simulations assumed an 8×8 network used to interconnect a 2-cm×2-cm 64-core CMP.

The photonic architectures assume a 2.5-GHz clock for the electronic control plane. The control-plane electronic 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 length of 32 bits. To fairly compare the use of WDM partitions, the 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 photonic circuit-switching topology design, augmented with WDM partitions, was used for this study [9].

The TorusNX is a previously proposed circuit-switched topology designed with a reduced number of crossings and an optimized switching layout [9]. A 4×4 version of the TorusNX is illustrated in Fig. 3, consisting of 16 gateway switches and 16 4×4 non-blocking switches. The structure of each switch configured with two photonic WDM partitions is diagrammed in Fig. 4. Each pair of rings (indicated by a red and blue ring) composes the two cascaded rings that compose a two WDM-partition router. Note that the original single-partition designs

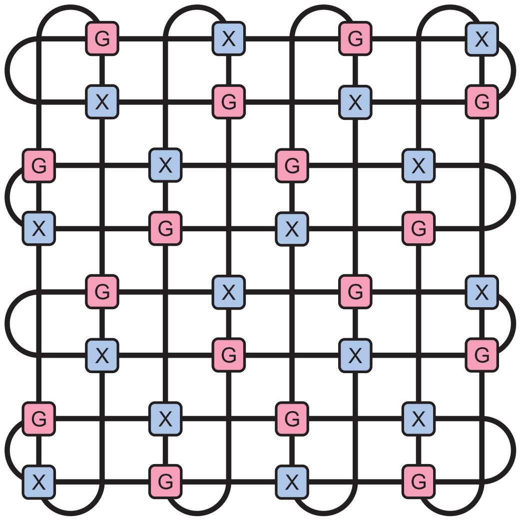

Figure 3. Schematic of the TorusNX topology. 'G' blocks represent gateway photonic switches and 'X' blocks represent 4×4 non-blocking photonic switches. Lines indicate bi-directional links which are composed of two waveguides that are used for counter-propagating lightwaves.

can be reconstructed by removing either the red or blue rings from the layout.

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 bits and utilizes a 2048-bit buffer on each input port. The electronic mesh network also operates on a 2.5-GHz clock.

Performance measurements were recorded for varying degrees of message size, total number of wavelength channels, and number of WDM partitions (Fig. 5). Simulations were conducted with either a small (1-kbit) or large (100-kbit) message size. The smaller messages are representative of general purpose processing which typically use smaller message transmissions (e.g. cache line), while the larger messages are characteristic of many scientific applications. The total number of wavelengths was varied between 12 (low aggressiveness), 60 (moderate aggressiveness), and 120 wavelength channels (high aggressiveness). The number of WDM partitions ranged from 1 to 4 to capture the performance effect that the partitioning technique provides. The dotted-line curves in Fig. 5 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) 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 of 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 equal to the number of WDM partitions (e.g. 4 WDM partitions results in a 4× improvement). Only in the case of 12 wavelength channels is there a perceivable difference in serialization delay that results in a slightly degraded zero-load latency (120 ns for four WDM partitions) and lower gain in saturation bandwidth (approximately 90% gain per WDM partition). The partitioning technique provides significant performance gains relative to single channel case, however the photonic network variants still underperform in comparison to the electronic mesh, a disadvantage that has been previously shown for circuit-switched networks [10].

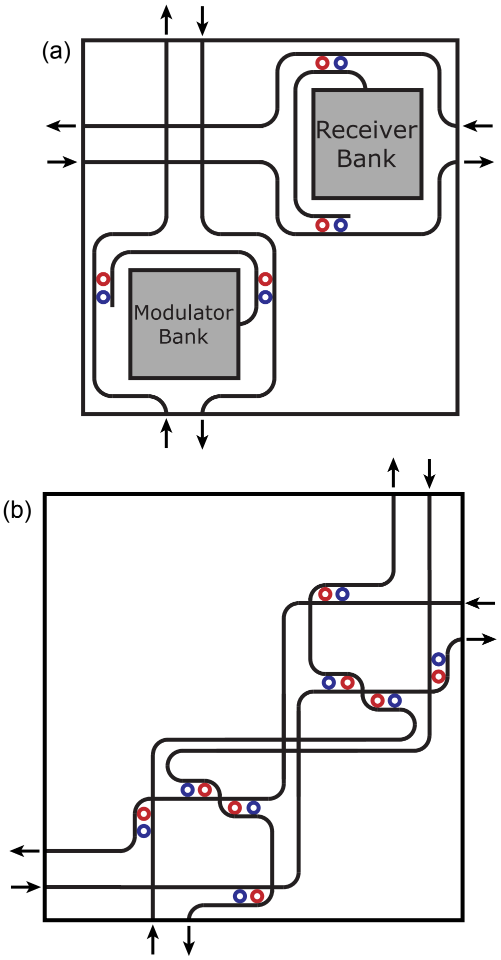

Figure 4. Schematic of the TorusNX photonic routers configured with two WDM partitions: (a) gateway switch and (b) 4×4 non-blocking photonic switch.

The transmission of 100-kbit messages (right plots in Fig. 5) 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 partitions cases, respectively. In the best case, four partitions using a total of 120 wavelength channels achieve a saturation bandwidth improvement of 764% over the electronic mesh. This shows that modest gains are achievable using the WDM partitioning technique 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 set of plots with a

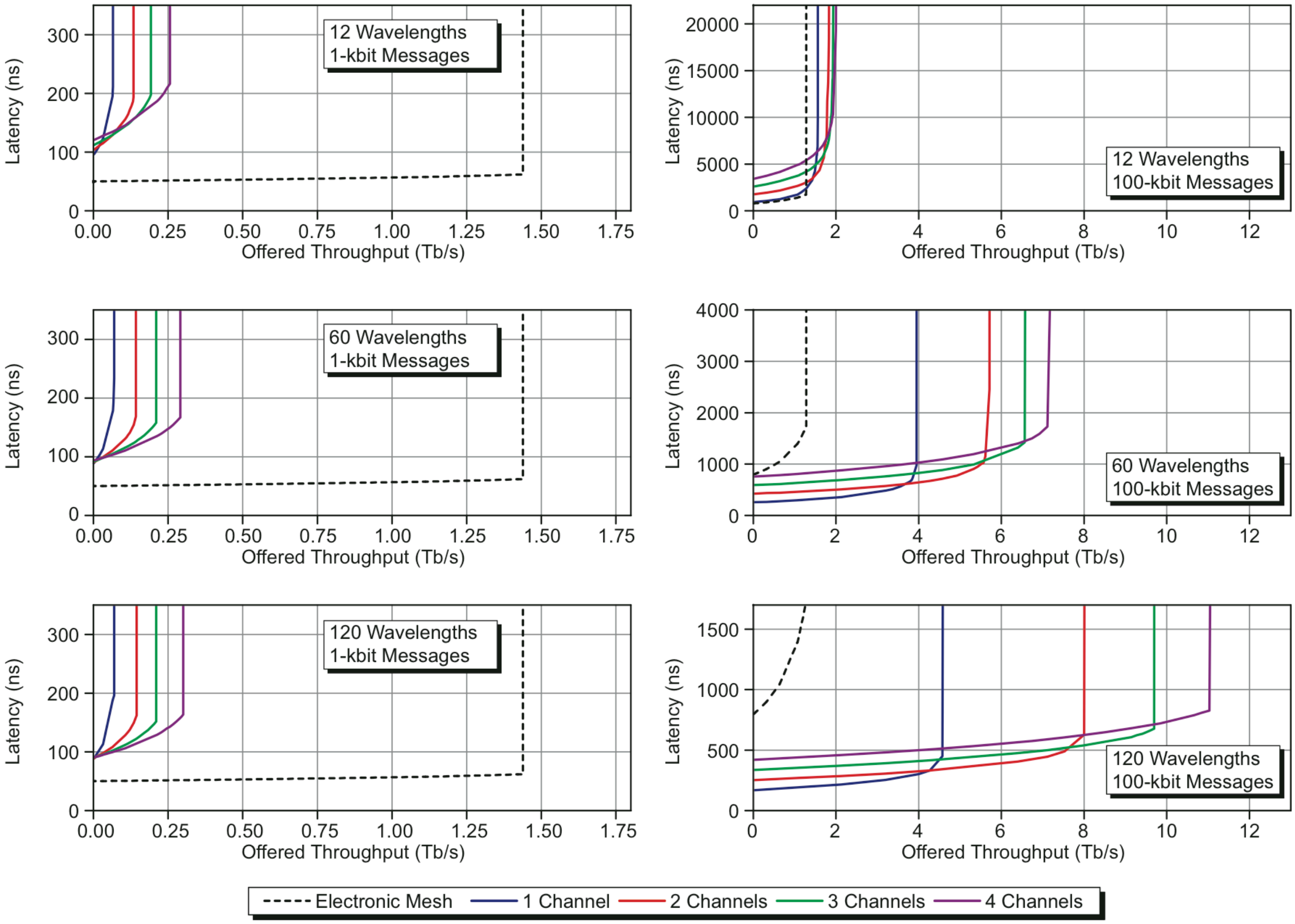

Figure 3. Average latency versus offered throughput for varying number of WDM partitions, message sizes, and number of wavelength channels. Electronic mesh performance is shown as a dotted line.

common total wavelength count, the division of wavelength channels among partitions produces noticeable differences in delay. This produces a noticeable trade-off when determining of a system design should minimize latency or maximize bandwidth.

IV. CONCLUSIONS

This work has motivated and presented the use of wavelength-selective spatial routing for producing WDM partitions, a novel interconnection network concept for increasing path diversity and increasing the performance of CMPs. This design is extensible to previous circuit-switching photonic topologies and is shown to improve network bandwidth performance.

ACKNOWLEDGMENT

This work was supported by the Interconnect Focus Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation and DARPA program.

REFERENCES

[1] E. Robinson, *et al.*. “Photonics for HPEC: a low-powered solution for high bandwidth applications,” in OFC, OWH2 (2011).

[2] C. Batten, *et al.*, "Building many-core processor-to-DRAM networks with monolithic CMOS silicon photonics," IEEE Micro **29** (4) pp. 8–21 (2009).

[3] D. Vantrease, *et al.*, "Corona: System implications of emerging nanophotonic technology," in ISCA, pp. 153–164 (2008).

[4] A. Shacham, *et al.*, "Photonic networks-on-chip for future generations of chip multiprocessors," IEEE Trans. Comput. **57** (9) pp. 1246–1260 (2008).

[5] G. Hendry, *et al.*, "Circuit-switched memory access in photonic interconnection networks for high-performance embedded computing," in SC, (2010).

[6] G. Hendry, *et al.*, “Analysis of photonic networks for a chip multiprocessor using scientific applications,” in NOCS, pp. 104–113 (2009).

[7] J. Chan, *et al.*, “Data transmission using wavelength-selective spatial routing for photonic interconnection networks,” in OFC, (2011).

[8] J. Chan, *et al.*, “Phoenixsim: A simulator for physical-layer analysis of chip-scale photonic interconnection networks,” in DATE, pp. 691–696 (2010).

[9] J. Chan, *et al.*, “Architectural exploration of chip-scale photonic interconnection network designs using physical-layer analysis,” IEEE/OSA JLT, **28** (9) pp. 1305–1315, (2010).

[10] G. Hendry, *et al.*, “Silicon nanophotonic network-on-chip using TDM arbitration,” in HOTI, pp. 88–95, (2010).