Photonic Interconnection Network Architectures Using Wavelength-Selective Spatial Routing for Chip-Scale Communications

Johnnie Chan and Keren Bergman

***Abstract*—The overall performance of modern computing systems is increasingly determined by the characteristics of the interconnection network used to provide communication links between on-chip cores and off-chip memory. Photonic technology has been proposed as an alternative to traditional electronic interconnects because of its advantages in bandwidth density, latency, and power efficiency. Circuit-switched photonic interconnect topologies 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 induces high network resource contention, unfairness, and long latencies. This work improves upon the circuit-switching paradigm by introducing the use of on-chip wavelength-selective spatial routing to produce multiple logical communication layers on a single physical plane. This technique yields higher path diversity in photonic interconnection networks, demonstrating as much as 764% saturation bandwidth improvement with synthetic traffic and as much as 89% improvement in execution time and energy dissipation for traffic from scientific application traces.**

***Index Terms*—High-performance computing; Multiprocessor interconnection networks; Optical interconnects; WDM networks.**

I. Introduction

Modern computing platforms are leveraging the parallel computation capabilities of chip multiprocessors (CMPs) to achieve improved computational performance and energy efficiency. The transition away from single-core architectures to CMPs occurred due to fundamental limitations in the power dissipation capabilities of current packaging technology, which consequently hampered progress in on-chip clock-rate scaling. This subsequent shift toward CMP architectures creates a need for high-performance chip-scale interconnects for core-to-core and core-to-memory communications. Chip-scale interconnects have thus far been implemented with electronics, but they achieve limited on-chip and off-chip bandwidth scaling due to minimum achievable wire pitches and the same power dissipation restrictions that constrict the clock rate. These restrictions will prevent electronics from achieving the performance requirements of current and future applications that are communication bound. Additionally, there will be difficult challenges in meeting the connectivity requirements of computer systems with increasing number of cores. New technologies need to be explored to address these issues.

Integrated photonic technology is an attractive interconnect solution that can be used to mitigate the energy and bandwidth bottlenecks that are arising in CMP systems. Photonic interconnects can enable improved bandwidth density by leveraging wavelength-division multiplexing (WDM) to concurrently transmit multiple spectrally parallel streams of data through a single optical waveguide, which contrasts with electronic interconnects that require a unique metal wire per bit stream. As a result, photonics can alleviate the problems facing interconnect subsystems that are reaching limits in wire and input/output (I/O) pin density. Additionally, photonics also affords improved energy efficiency by eliminating the need for the electronic buffering and switching that are found in conventional electronic networks. The combined advantages of better bandwidth density and power efficiency make photonic interconnects a serious contender as a technological replacement for electronic interconnects.

This work describes a novel on-chip photonic interconnect architecture that leverages a new concept known as *wavelength-selective spatial routing* (WSSR) to increase the path diversity within previously proposed circuit-switched photonic networks for CMPs. 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

Manuscript received October 27, 2011; revised January 17, 2012; accepted January 20, 2012; published February 13, 2012 (Doc. ID 156326).

The authors are with the Department of Electrical Engineering, Columbia University, New York, New York, 10027, USA (e-mail: johnnie@ee.columbia.edu).

Digital Object Identifier 10.1364/JOCN.4.000189

1943-0620/12/030189-13/$15.00

superior bandwidth performance with a minimal increase in design complexity and latency.

The remainder of this paper describes the architectural details of the WSSR design and presents an analysis of its performance characteristics. First, Section II presents a generalized overview of how previously proposed photonic interconnection networks are designed and the optical switching mechanisms they use. Next, Section III discusses the concept of WSSR and the devices required to implement it. Section IV discusses the various network-level performance implications of the proposed technique. In Section V, synthetic and trace-driven simulation performance results of network topologies using WSSR are presented and discussed. Concluding remarks are made in Section VI.

II. Related Work

A. 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 complementary metal–oxide semiconductor (CMOS) compatibility [1–4]. Ring resonators selectively couple and reject lightwaves at periodic wavelengths in the optical spectrum. When a waveguide is properly positioned next to a ring resonator, lightwaves injected into the waveguide that are rejected by the ring (termed *off resonance*) will be transmitted (Fig. 1(a)). Lightwaves that couple into the ring (termed *on resonance*) will not be transmitted and will be dissipated by the ring (Fig. 1(b)). A ring can also be electrically manipulated to fluctuate between these two states to produce modulated light on the waveguide output. Alternatively, ring resonators can be designed to deliver on-resonance lightwaves onto a nearby secondary waveguide to enable filtering or switching functionality (Fig. 1(c)).

The free spectral range (FSR) of the ring resonator is inversely proportional to the circumference of the loop, and it 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. 1(d)). 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. 1(e)). In this 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, Figs. 1(d) and 1(e) 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 [3,4]. 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;

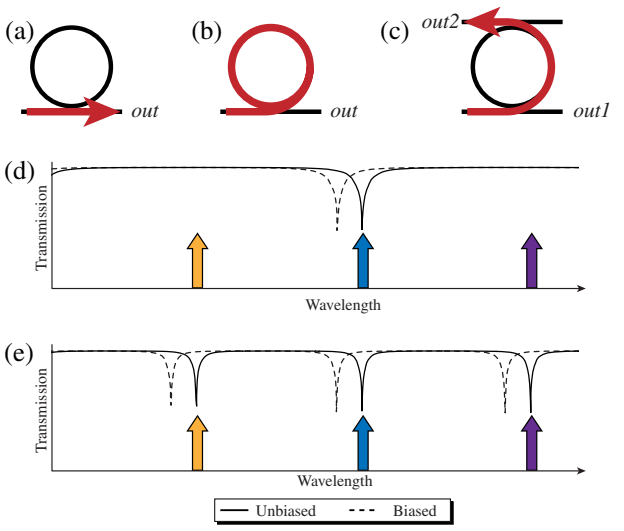


Fig. 1. (Color online) Ring resonator functional characteristics. (a) Off-resonance wavelength with a single waveguide. (b) On-resonance wavelength with a single waveguide. (c) On-resonance wavelength with secondary waveguide. (d) Transmission spectra of a long FSR ring resonator. (e) Transmission spectra 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.

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 a maximum wavelength channel count limitation of 62 when assuming 10 Gb/s data rates [5]. 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 [6], photonic bandgap structures [7], and the Vernier effect [8]. These techniques can be used to increase the FSR, and correspondingly increase the available spectrum and allowable number channels.

B. Photonic Interconnection Networks

Advancements in silicon photonic device technology have 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 [9,10], bends [9], crossings [11], filters [1], switches [2], modulators [3], and detectors [12]. 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 those of their electronic counterparts. In what can be considered as the first step toward 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 [13]. Network architects have also proposed a variety of

advanced novel interconnect designs in order to fully leverage the capabilities of photonics [14–22].

Wavelength-routed topologies are constructed using ring-resonator-based filters which accordingly route lightwaves based on their wavelengths [14–18]. 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 been 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 [19–21]. The ring resonators act as comb switches to simultaneously control the path of all incident wavelength channels (Fig. 1(e)). 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 [20].

The usage of time-division multiplexing (TDM) has also been previously proposed as a technique for improving optical on-chip network performance [22]. *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.

Incarnations of some of the aforementioned TDM routing and the WSSR concept presented in this work were previously proposed and analyzed for multiprocessor networks and wide-area networks [23,24]. 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 it showed that link multiplexing performed better in certain traffic configurations with a significant reduction in design complexity [23]. The alternative WDM technique was also described to have similar performance characteristics as the TDM case [24].

The various proposed photonic networks can be generally classified as leveraging a combination of three optical arbitration domains: time, wavelength, and space. The simplest case is the optical bus, which does not require any form of routing. Figure 2 illustrates the design space that is afforded by these arbitration methodologies and the relative placement of the aforementioned routing techniques. Included in the design space is WSSR, which will be discussed in the following section.

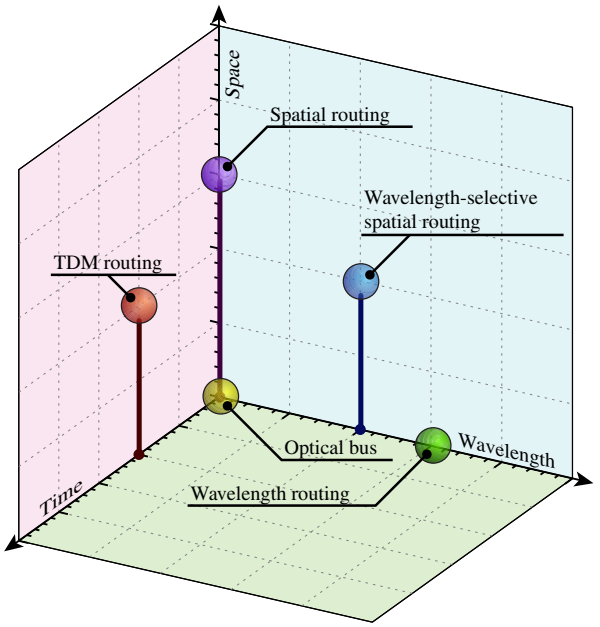


Fig. 2. (Color online) Routing technique design space based on three arbitration domains: time, wavelength, and space.

III. Wavelength-Selective Spatial Routing

WSSR is used to selectively manipulate wavelength channel subsets of a WDM signal as it propagates through a network [25]. WSSR can be qualified as a hybrid form of spatial routing and wavelength routing (Fig. 2). 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.

Figure 3(a) illustrates the inclusion of an additional set of three wavelength channels, interleaved amongst the original set of wavelengths that were shown in Fig. 1(e). 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 WSSR. Moreover, a second ring resonator can be cascaded and designed to align 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 [2]. Figures 3(b)–3(e) show the four possible routing configurations of the two-partition router, illustrating the independent controllability of each WDM partition.

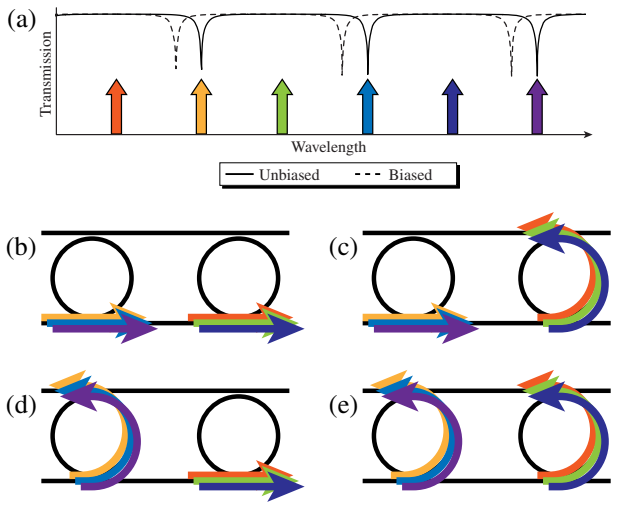


Fig. 3. (Color online) (a) Spectral placement of two WDM partitions (each containing three wavelength channels), with respect to the spectrum of an electro-optic broadband ring switch. (b)–(e) Four possible routing configurations for a two-partition router.

Previous work has demonstrated the transmission of six 10 Gb/s wavelength channels configured as two WDM partitions [25]. The experiment was limited to a demonstration of a slice of the two-partition WSSR switch due to availability of fabricated components; however, the transparency of the *auxiliary partition* in [25] shows the possibility of cascading a second copy of the device to fully construct the switch. Additionally, the experiment was conducted on a second-order ring resonator switch, which is in contrast with our current discussion which assumes a single-order ring device. Physical-layer characteristics such as resonance profile, switching energy, and insertion losses will differ from single-order versions of the device; nevertheless, the network-level functionality in its ability to selectively switch wavelength channels is identical.

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 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 Subsection IV.B.

We consider in this paper a range of ring diameters that have been experimentally verified to operate as switches. Preston *et al.* found that a minimum wavelength channel spacing of 0.8 nm was required for ring-modulated 10 Gb/s wavelength channels to maintain sufficiently low crosstalk levels (<−20 dB) [5]. This corresponds to a 200-μm-diameter ring switch, which has been demonstrated previously with an adequately wide passband for transmitting the high-speed data rate [2].

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 [26]. The highest number of WDM partitions we consider in our presented analysis is six, which requires a 4.8-nm FSR and a 33.3-μm ring. This falls outside the experimentally verified limit.

Independent routability of each WDM partition enhances the 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 [5]. 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 the 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 [19], 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 [18]. The 2D-HERT network uses 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.

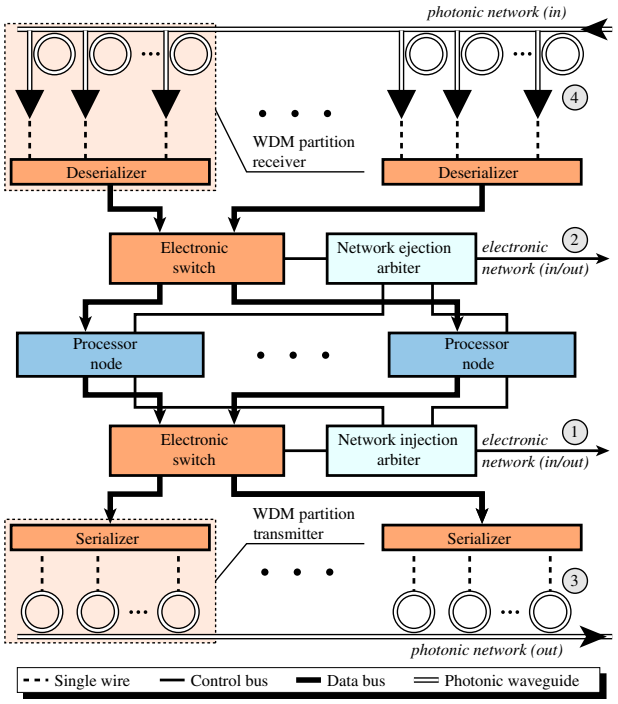


Fig. 4. (Color online) The WSSR gateway architecture with concentrating processing cores.

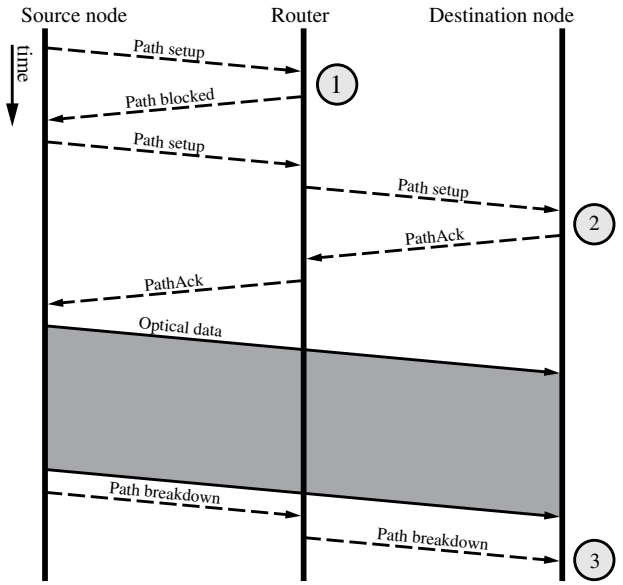


Fig. 5. Example timing diagram of the circuit-switching and WSSR allocation protocol. A path provisioning request is initially blocked, but is successful upon re-attempt.

An advantage of not utilizing wavelength for routing purposes is that we can exploit wavelength parallelism for enabling higher node-to-node data rates. 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 [19]. Processors interface with the network by communicating with a network gateway (Fig. 4). Resource allocation of photonic routers is accomplished on a separate lightweight 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. 4):

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 '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). Figures 5 and 6 illustrate the message transactions required in perform allocation and data transmission. In the example, the request is initially blocked at an 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 that particular source node and have not been deallocated 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 [24].

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

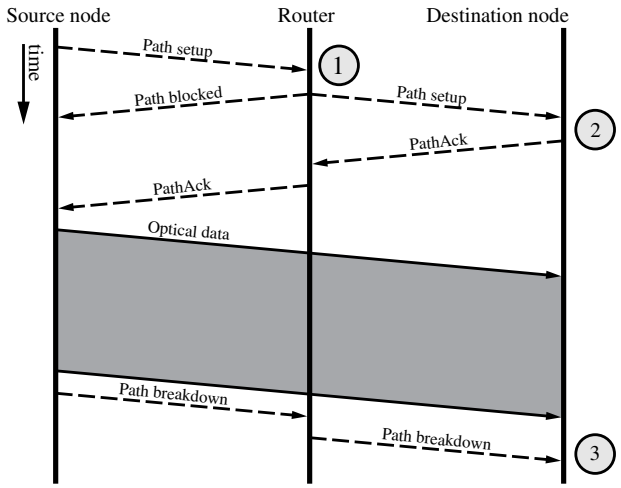


Fig. 6. Example timing diagram of the WSSR allocation protocol. If a single path provisioning request is attempted with multiple partitions, a path setup request can partially block on a particular partition while being successful on another partition.

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 it only continues propagation if at least a single WDM partition is still available. Figure 5 (marker '1') illustrates a situation in which the path is blocked for all partitions being considered. Figure 6 (marker '1') shows a sequence of events in which a subset of the available WDM partitions of the *PathSetup* message is 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 (Figs. 5 and 6 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, which is chosen at random from the pool of available channels (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 (Figs. 5 and 6 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.

A previously proposed circuit-switched topology is the TorusNX, which is designed with a reduced number of crossings and an optimized switching layout [21]. A 4×4 version of the TorusNX is illustrated in Fig. 7, consisting of

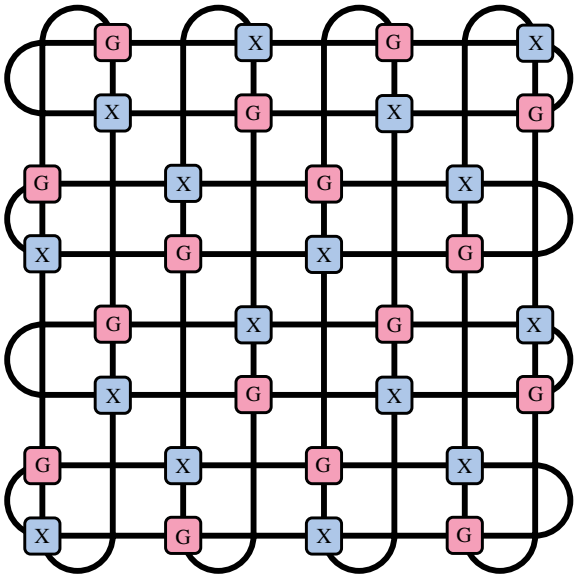


Fig. 7. (Color online) Schematic of the TorusNX topology. 'G' blocks represent gateway photonic switches and 'X' blocks represent 4×4 non-blocking photonic switches. Lines indicate bidirectional links which are composed of two waveguides that are used for counter-propagating lightwaves.

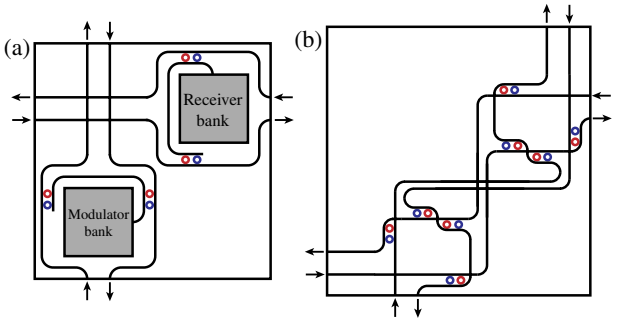


Fig. 8. (Color online) Schematic of the TorusNX photonic routers configured with two WDM partitions: (a) gateway switch and (b) 4×4 non-blocking photonic switch.

16 gateway switches and 16 4×4 non-blocking switches. The structure of each switch configured with two WDM partitions is given in Fig. 8. Each pair of rings (indicated by a red ring and a blue ring) composes the two cascaded rings that compose a two-partition router. Note that the original single-partition designs can be reconstructed by removing either the red or blue set of rings from the layout.

IV. Analysis

A. 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

TABLE I
INSERTION LOSS PARAMETERS

Parameter	Value	Ref.
Propagation loss (Silicon)	1.7 dB/cm	[9]
Waveguide crossing	0.16 dB	[11]
Waveguide bend	0.005 dB/90°	[9]
Drop into a ring	0.6 dB	[2]
Pass by a ring	0.005 dB	[2]

determine the requirements for the laser input power and for the receiver sensitivity. The insertion loss analysis assumes the parameters listed in Table I and is derived from experimentally validated published results.

The results of the analysis are shown in Fig. 9 for different levels of partitioning. 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 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 data rate (demonstrated in [12]) 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 [27] or lateral tapered waveguides [28].

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 CMOS-compatible platforms such as three-dimensional (3D) deposited technology can virtually eliminate these loss constraints and increase the feasibility of this type of network [29].

B. 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, the 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 the spectral density of the channel spacing, it 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.

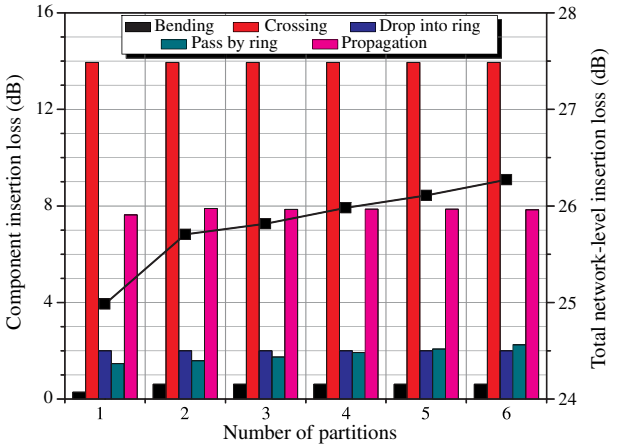


Fig. 9. (Color online) Insertion loss analysis of the TorusNX topology for varying levels of partitioning. Column plots correspond to worst-case insertion loss per component among all possible network paths (left-vertical axis). The line plot corresponds to the greatest total network-level insertion 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.

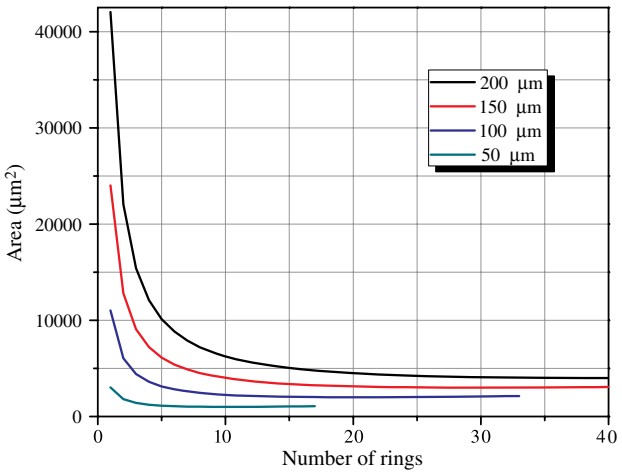


Fig. 10. (Color online) Photonic router footprint for varying number of rings (which corresponds to the number of WDM partitions enabled by the router). Legend indicates the ring diameter for the single-ring case.

The area footprint of a single WSSR router versus the number of WDM partitions (labeled as the number of rings) is analyzed in Fig. 10. The WSSR router footprint calculations assume a structure similar to those shown in Fig. 3. The only locations where waveguides are closely placed together are regions where optical coupling is 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 is 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 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 on 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 increases, 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 micro-ring [26]).

C. 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 the performance of on-chip networks [19]. 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 needed 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 [21]. 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 occur 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 an 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 that 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

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

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

$$L_C = \lim_{N\to\infty} P_C = \left(1 - \frac{1}{e}\right)^C. \tag{2}$$

Figure 11 plots Eq. (1) for $C = 1,\ldots,6$ and $3 \le N \le 100$, and appends the limit calculated from Eq. (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 such as WSSR can provide dramatic improvements to performance through a reduction in blocking probability.

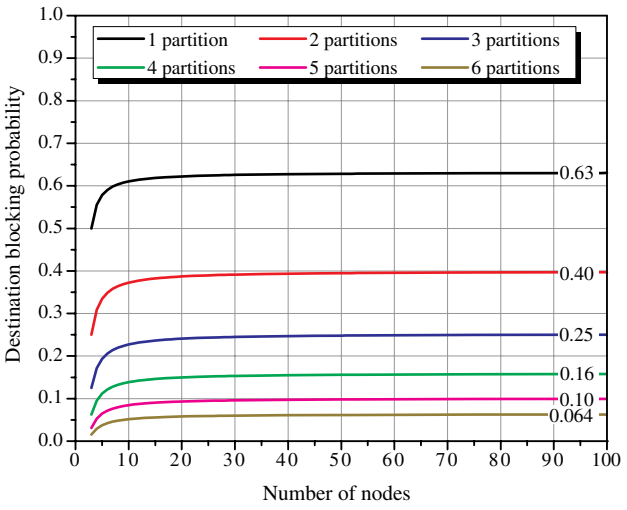


Fig. 11. (Color online) Destination blocking probability in a non-blocking network for varying number of interleaved channels. The limit of each blocking probability as $N \to \infty$ is superimposed on the right of the plot.

V. Simulation Results and Analysis

The partitioned-WDM network architecture is modeled and simulated in PhoenixSim, a chip-scale photonic network simulation environment [30]. All conducted simulations assumed a 2-cm × 2-cm 64-core CMP, which requires an 8 × 8 network.

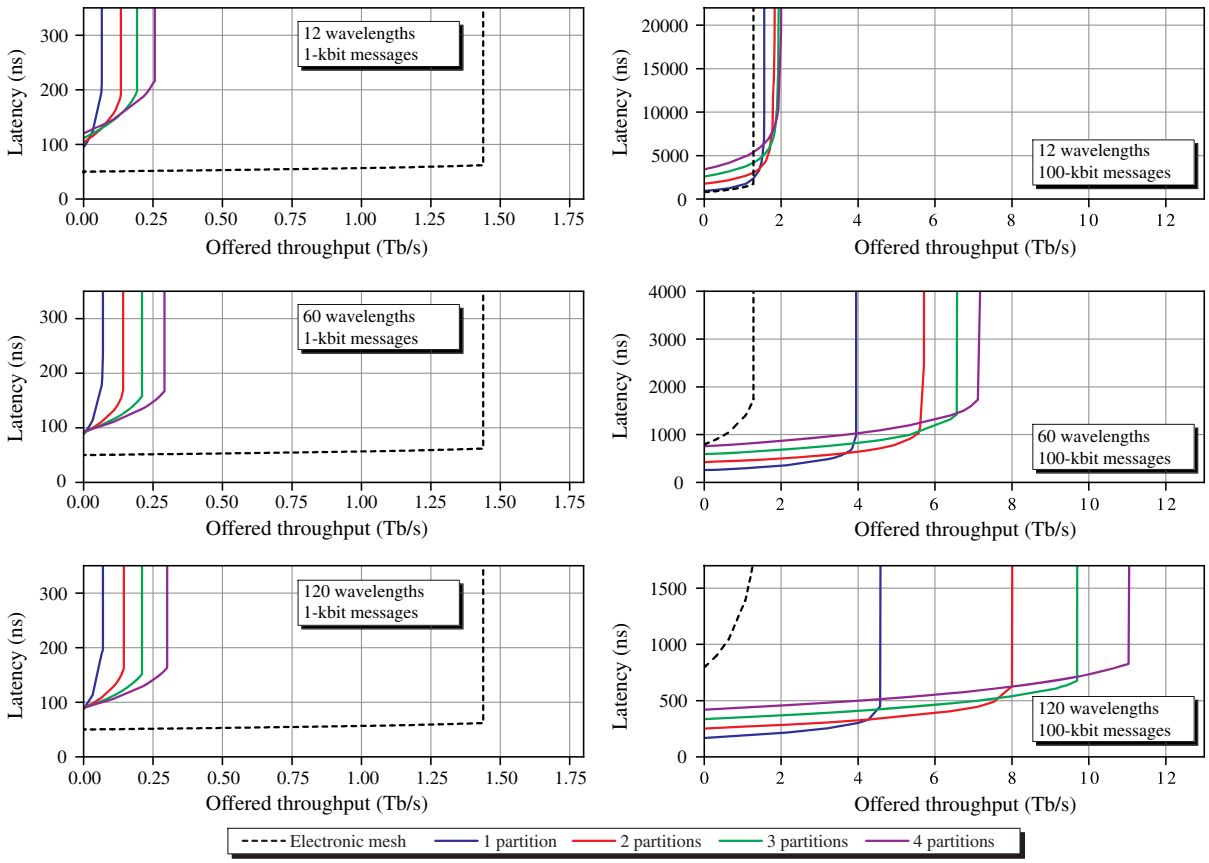


Fig. 12. (Color online) Average latency versus offered throughput for varying number of WDM partitions, message sizes, and number of wavelength channels. Electronic mesh performance is shown as a dotted line.

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 eight control messages. Path setup control messages have an assumed bit length of 32 bits. 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 WSSR, was used for this study [21].

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. 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.

A. 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 one to four to capture the performance effect that the WSSR technique provides. The total number of wavelengths was varied between 12 (low aggressiveness), 60 (medium aggressiveness), and 120 wavelength channels (high aggressiveness).

Figure 12 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. 12) 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 equal to the number of WDM partitions (e.g., four partitions results in a fourfold 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 four 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 [22].

The transmission of 100-kbit messages (right plots in Fig. 12) 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, having four partitions using a total of 120 wavelength channels achieves a saturation bandwidth improvement of 764% over the electronic mesh. The results show that modest gains are achievable using WSSR for even the low-aggressiveness version of the TorusNX network; however, greater gains can be expected as photonic device fabrication matures and the possible number of wavelength channels increases. 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 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.

B. 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 data rate 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 [31].
- *Cactus*: An astrophysics computation toolkit designed to solve coupled nonlinear hyperbolic and elliptic equations arising from general relativity [32].
- *GTC*: A 3D particle-in-cell application developed to study turbulent transport in magnetic confinement fusion [33].
- *MADbench*: A benchmark based on MADspec cosmology code, calculating the maximum likelihood angular power spectrum of the cosmic microwave background [34].

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 message size. This set of application traces forms a representative set of communication patterns that match a large class of scientific applications currently being investigated by the computational science research community. The characteristics of each application trace are summarized in Table II, and traffic pattern plots are given in Fig. 13.

TABLE II
APPLICATION TRACE CHARACTERISTICS

Application	Number of phases	Number of msgs.	Total data sent (B)	Avg. msg. size (B)
Paratec	34	126,059	5.4 M	43.3
Cactus	2	285	7.3 M	25,600
GTC	2	63	8.1 M	129,796
Madbench	195	15,414	86.5 M	5613

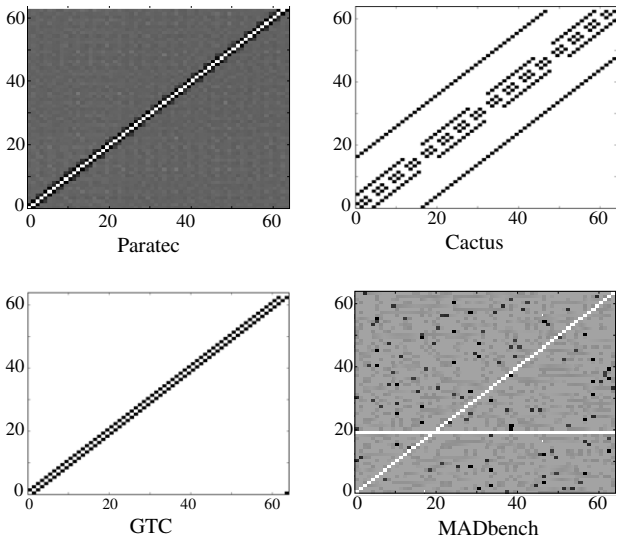


Fig. 13. Traffic pattern plots for the four scientific applications being considered. The left axis represents the source thread ID, and the bottom axis represents the destination thread ID. White blocks represent no communication load while darker shades of gray represent increased traffic load between the associated source–destination pair.

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 it 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 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. 14. The small messages found in the Paratec trace result 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.

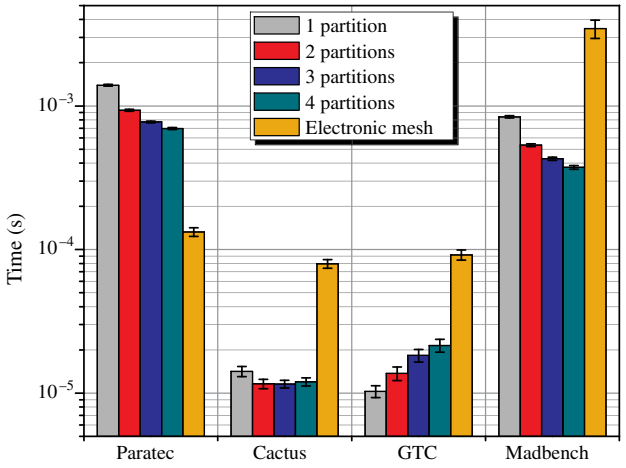


Fig. 14. (Color online) Total simulation time required to complete each application trace. Columns indicate the average resulting time, and error bars indicate one standard deviation of the sampled data.

The number of WDM partitions varies with each application with respect to the 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 the 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.

The lackluster performance of Cactus and GTC is also indicative of the limited traffic pattern as indicated in Fig. 13. 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 in agreement 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 more opportunities for the path parallelism to be exploited.

Power simulations in PhoenixSim utilize ORION 2.0 [35] for electronic router power performance and built-in static and dynamic power models for photonic components. The ORION 2.0 router model assumes the 32 nm technology node and normal threshold voltage (NVT). Power parameters for optical devices are summarized in Table III. Details of the power simulation methodology that is employed by PhoenixSim can be found in [30].

Figure 15 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, 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.

TABLE III
OPTICAL DEVICE POWER PARAMETERS

Parameter	Value
Ring switch dynamic energy	375 fJ[a]
Ring switch static energy	400 μW[b]
Modulation dynamic energy	85 fJ/bit[c]
Modulation static energy	30 W[c]
Detector energy	50 fJ/bit[d]
Thermal ring tuning	100 μW/ring[e]

Notes.
[a] Calculation based on carrier density, assuming 50-μm diameter, 320 nm × 250 nm micro-ring waveguide cross-section, 75% waveguide volume exposure, 1 V forward bias.
[b] Based on switching energy, including photon lifetime for re-injection.
[c] [36], static energy calculated for half a 10 GHz clock cycle, with 50% probability of a '1' bit.
[d] Conservative approximation assuming a femtofarad class receiver-less SiGe detector with C < 1fF.
[e] Assumes a 1-μW/degree tuning cost per ring, with a temperature deviation of 20 degrees.

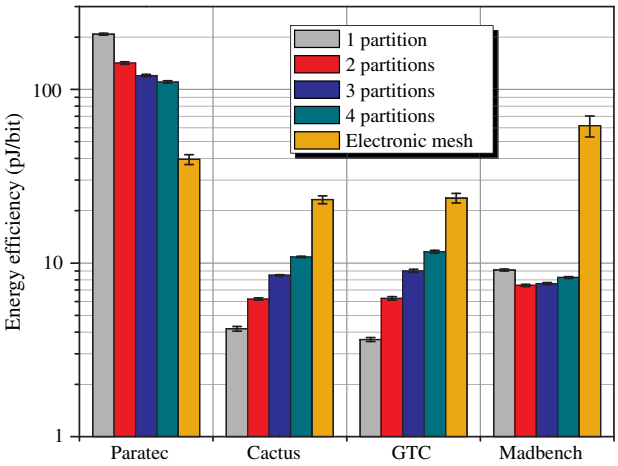


Fig. 15. (Color online) Network-level energy efficiency from each application trace. Columns indicate the average resulting energy efficiency, and error bars indicate one standard deviation of the sampled data.

The trace-driven results are summarized in Table IV, 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 single-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.

VI. CONCLUSIONS

This work has motivated and presented the use of WSSR, a novel interconnection network concept for reducing the path

TABLE IV
APPLICATION TRACE RESULTS SUMMARY

	Paratec	Cactus	GTC	Madbench
Execution time optimized				
Optimal number of partitions	4	3	1	4
Improvement	−425%	85%	89%	89%
Energy dissipation optimized				
Optimal number of partitions	4	1	1	2
Improvement	−205%	82%	85%	89%

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 the 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.

ACKNOWLEDGMENTS

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. We are also grateful for the support from the GRO program sponsored by Samsung Advanced Institute of Technology (SAIT).

The authors would like to thank Shoaib Kamil, Lenny Oliker, and John Shalf from CRD/NERSC at Lawrence Berkeley National Laboratory for their support in providing the application traces.

REFERENCES

[1] B. Little, J. Foresi, G. Steinmeyer, E. Thoen, S. Chu, H. Haus, E. Ippen, L. Kimerling, and W. Greene, "Ultra-compact Si-SiO_2 microring resonator optical channel dropping filters," *IEEE Photon. Technol. Lett.*, vol. 10, no. 4, pp. 549–551, Apr. 1998.

[2] B. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, "All-optical comb switch for multiwavelength message routing in silicon photonic networks," *IEEE Photon. Technol. Lett.*, vol. 20, no. 10, pp. 767–769, May 2008.

[3] S. Manipatruni, Q. Xu, B. Schmidt, J. Shakya, and M. Lipson, "High speed carrier injection 18 Gb/s silicon micro-ring electro-optic modulator," in *20th Annu. Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, Oct. 2007, pp. 537–538.

[4] H. L. R. Lira, S. Manipatruni, and M. Lipson, "Broadband hitless silicon electro-optic switch for on-chip optical networks," *Opt. Express*, vol. 17, no. 25, pp. 22271–22280, Dec. 2009.

[5] K. Preston, N. Sherwood-Droz, J. Levy, and M. Lipson, "Performance guidelines for WDM interconnects based on silicon microring resonators," in *Conf. on Lasers and Electro-Optics (CLEO)*, May 2011.

[6] J. Feng, Q. Li, and Z. Zhou, "Single ring interferometer configuration with doubled free-spectral range," *IEEE Photon. Technol. Lett.*, vol. 23, no. 2, pp. 79–81, Jan. 2011.

[7] 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 *5th IEEE Int. Conf. on Group IV Photonics*, Sept. 2008, pp. 82–84.

[8] Y. Yanagase, S. Suzuki, Y. Kokubun, and S. T. Chu, "Box-like filter response and expansion of FSR by a vertically triple coupled microring resonator filter," *J. Lightwave Technol.*, vol. 20, no. 8, pp. 1525–1529, Aug. 2002.

[9] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," *Nat. Photonics*, vol. 1, pp. 65–71, 2006.

[10] M. Gnan, S. Thorns, D. Macintyre, R. De La Rue, and M. Sorel, "Fabrication of low-loss photonic wires in silicon-on-insulator using hydrogen silsesquioxane electron-beam resist," *Electron. Lett.*, vol. 44, no. 2, pp. 115–116, Jan. 2008.

[11] W. Bogaerts, P. Dumon, D. V. Thourhout, and R. Baets, "Low-loss, low-cross-talk crossings for silicon-on-insulator nanophotonic waveguides," *Opt. Lett.*, vol. 32, no. 19, pp. 2801–2803, 2007.

[12] S. Assefa, B. G. Lee, C. Schow, W. M. Green, A. Rylyakov, R. A. John, and Y. A. Vlasov, "20 Gbps receiver based on germanium photodetector hybrid-integrated with 90 nm CMOS amplifier," in *CLEO 2011—Laser Applications to Photonic Applications*, 2011, PDPB11.

[13] 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 Conf.*, 2011, OWZ3.

[14] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. Holzwarth, M. Popovic, H. Li, H. Smith, J. Hoyt, F. Kartner, R. Ram, V. Stojanovic, and K. Asanovic, "Building many-core processor-to-DRAM networks with monolithic CMOS silicon photonics," *IEEE Micro*, vol. 29, no. 4, pp. 8–21, July–Aug. 2009.

[15] N. Kirman, M. Kirman, R. K. Dokania, J. F. Martinez, A. B. Apsel, M. A. Watkins, and D. H. Albonesi, "On-chip optical technology in future bus-based multicore designs," *IEEE Micro*, vol. 27, no. 1, pp. 56–66, 2007.

[16] M. J. Cianchetti, J. C. Kerekes, and D. H. Albonesi, "Phastlane: a rapid transit optical routing network," in *Proc. of the 36th Annu. Int. Symp. on Computer Architecture (ISCA)*, 2009, pp. 441–450.

[17] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn, "Corona: System implications of emerging nanophotonic technology," in *Proc. of the 35th Annu. Int. Symp. on Computer Architecture (ISCA)*, June 2008, pp. 153–164.

[18] S. Koohi, M. Abdollahi, and S. Hessabi, "All-optical wavelength-routed NoC based on a novel hierarchical topology," in *5th IEEE/ACM Int. Symp. on Networks on Chip (NoCS)*, May 2011, pp. 97–104.

[19] A. Shacham, K. Bergman, and L. Carloni, "Photonic networks-on-chip for future generations of chip multiprocessors," *IEEE Trans. Comput.*, vol. 57, no. 9, pp. 1246–1260, Sept. 2008.

[20] G. Hendry, S. Kamil, A. Biberman, J. Chan, B. Lee, M. Mohiyuddin, A. Jain, K. Bergman, L. Carloni, J. Kubiatowicz, L. Oliker, and J. Shalf, "Analysis of photonic networks for a chip multiprocessor using scientific applications," in *3rd ACM/IEEE Int. Symp. on Networks-on-Chip (NOCS)*, May 2009, pp. 104–113.

[21] J. Chan, G. Hendry, A. Biberman, and K. Bergman, "Architectural exploration of chip-scale photonic interconnection network designs using physical-layer analysis," *J. Lightwave Technol.*, vol. 28, no. 9, pp. 1305–1315, May 2010.

[22] G. Hendry, J. Chan, S. Kamil, L. Oliker, J. Shalf, L. Carloni, and K. Bergman, "Silicon nanophotonic network-on-chip using TDM arbitration," in *2010 IEEE 18th Annu. Symp. on High Performance Interconnects (HOTI)*, Aug. 2010, pp. 88–95.

[23] C. Qiao and R. Melhem, "Reducing communication latency with path multiplexing in optically interconnected multiprocessor systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 8, no. 2, pp. 97–108, Feb. 1997.

[24] X. Yuan, R. Melhem, and R. Gupta, "Distributed path reservation algorithms for multiplexed all-optical interconnection networks," *IEEE Trans. Comput.*, vol. 48, no. 12, pp. 1355–1363, Dec. 1999.

[25] 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," in *Optical Fiber Communication Conf.*, Mar. 2011.

[26] Q. Xu, D. Fattal, and R. G. Beausoleil, "Silicon microring resonators with 1.5-μm radius," *Opt. Express*, vol. 16, no. 6, pp. 4309–4315, Mar. 2008.

[27] G. Roelkens, D. V. Thourhout, and R. Baets, "High efficiency silicon-on-insulator grating coupler based on a poly-silicon overlay," *Opt. Express*, vol. 14, no. 24, pp. 11622–11630, Nov. 2006.

[28] V. R. Almeida, R. R. Panepucci, and M. Lipson, "Nanotaper for compact mode conversion," *Opt. Lett.*, vol. 28, no. 15, pp. 1302–1304, Aug. 2003.

[29] A. Biberman, K. Preston, G. Hendry, N. Sherwood-Droz, J. Chan, J. S. Levy, M. Lipson, and K. Bergman, "Photonic network-on-chip architectures using multilayer deposited silicon materials for high-performance chip multiprocessors," *ACM J. Emerging Technol. Comput. Syst.*, vol. 7, pp. 7:1–7:25, July 2011.

[30] J. Chan, G. Hendry, K. Bergman, and L. Carloni, "Physical-layer modeling and system-level design of chip-scale photonic interconnection networks," *IEEE Trans. Comput.-Aided Des.*, vol. 30, no. 10, pp. 1507–1520, Oct. 2011.

[31] A. Canning, L. Wang, A. Williamson, and A. Zunger, "Parallel empirical pseudopotential electronic structure calculations for million atom systems," *J. Comput. Phys.*, vol. 160, no. 1, pp. 29–41, 2000.

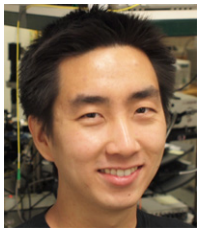
[32] Cactus Computational Toolkit [Online]. Available: http://www.cactuscode.org/.

[33] Z. Lin, S. Ethier, T. S. Hahm, and W. M. Tang, "Size scaling of turbulent transport in magnetically confined plasmas," *Phys. Rev. Lett.*, vol. 88, no. 19, 195004, Apr. 2002.

[34] J. Borrill, J. Carter, L. Oliker, and D. Skinner, "Integrated performance monitoring of a cosmology application on leading HEC platforms," in *Proc. of the 2005 Int. Conf. on Parallel Processing*, 2005, pp. 119–128.

[35] A. B. Kahng, B. Li, L.-S. Peh, and K. Samadi, "ORION 2.0: A power-area simulator for interconnection networks," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 1, pp. 191–196, Jan. 2012.

[36] M. Watts, D. Trotter, R. Young, and A. Lentine, "Ultralow power silicon microdisk modulators and switches," in *5th IEEE Int. Conf. on Group IV Photonics*, Sept. 2008, pp. 4–6.



Johnnie Chan (S'08) received his B.S. degree (with high distinction) in computer and electrical engineering and his M.S. degree in electrical engineering from the University of Virginia, Charlottesville, in 2005 and 2007, respectively. He is currently working toward a Ph.D. degree in the Department of Electrical Engineering, Columbia University, New York.

His current research interests include silicon photonic devices for chip-scale interconnection networks and optical networks for high-performance computing.

Keren Bergman (S'87–M'93–SM'07–F'09) received her B.S. degree from Bucknell University, Lewisburg, PA, in 1988, and her M.S. and Ph.D. degrees from Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively, all in electrical engineering.

She is currently a Professor and Chair of Electrical Engineering at Columbia University, New York, NY, where she also directs the Lightwave Research Laboratory (http://lightwave.ee.columbia.edu/). She leads multiple research programs on optical interconnection networks for advanced computing systems, data centers, optical packet-switched routers, and chip multiprocessor nanophotonic networks-on-chip.

Prof. Bergman is a Fellow of the IEEE and of the OSA, and she currently serves as the Co-Editor-in-Chief of the IEEE/OSA *Journal of Optical Communications and Networking*.