

Silicon Photonic Switch Fabrics in Computer Communications Systems

Benjamin G. Lee, *Senior Member, IEEE, Member, OSA*, Nicolas Dupuis, Petar Pepeljugoski, Laurent Schares, Russell Budd, Justin R. Bickford, *Member, IEEE*, and Clint L. Schow, *Senior Member, IEEE, Senior Member, OSA*

(Invited Paper)

Abstract—We discuss silicon photonic switch fabric designs that target data-intensive computing networks, reviewing recent results, and projecting future performance goals. We analyze the achievements of demonstrated hardware in terms of switching time, footprint, crosstalk, and power consumption, concluding that the most crucial metric to improve upon is net loss. We propose integrating semiconductor optical amplifiers into the switch fabric using either flip-chip or wafer-bonding technology, and investigate its potential merits alongside several challenges in implementation. Furthermore, we explore the dominant causes of crosstalk, and discuss manners for reducing it. We perform switch simulations that project a 7-dB reduction in crosstalk, when using a push-pull, rather than a single-ended phase shifter drive scheme. We also evaluate crosstalk effects on transmission performance using a full-link model that incorporates multiple crosstalk-accumulating photonic switch hops. The study demonstrates the degree to which crosstalk may degrade signal integrity after just a few occurrences. Finally, a comparison of four topologies highlights tradeoffs in physical-layer design and scheduling complexity, illustrating the scales that may be accomplished with the simplest topologies, and the device improvements required to achieve the more robust architectures.

Index Terms—Multiprocessor interconnection networks, optical switch, silicon photonics.

I. INTRODUCTION

POINT-TO-POINT optical interconnects are widely deployed in the current generation of high-performance computers (HPC) targeted toward scientific applications, with some machines comprising hundreds of thousands of optical links [1], [2]. Their bandwidth, power, and reach advantages over electrical cabling have enabled record-setting super-computing performance [3]. To scale these systems from one generation to the next while keeping costs within reason, designers commonly maintain low byte-to-flop ratios and rely on ever-improving optical devices and link architectures with increased bandwidths

Manuscript received September 26, 2014; revised November 7, 2014; accepted November 10, 2014. Date of publication December 18, 2014; date of current version February 20, 2015. This work was supported by the Defense Advanced Research Projects Agency and the Army Research Laboratory under Contract W911NF-12-2-0051. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of DARPA or the Department of Defense. Approved for public release, distribution unlimited.

B. G. Lee, N. Dupuis, P. Pepeljugoski, L. Schares, R. Budd, and C. L. Schow are with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: bglee@us.ibm.com; ndupuis@us.ibm.com; petarp@us.ibm.com; schares@us.ibm.com; rbudd@us.ibm.com; cshaw@us.ibm.com).

J. R. Bickford is with the U.S. Army Research Laboratory, Adelphi Laboratory Center, Adelphi, MD 20783 USA (e-mail: justin.r.bickford.civ@mail.mil).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>

Digital Object Identifier 10.1109/JLT.2014.2371616

and decreased footprints and power consumptions per unit cost. As a result, HPC machines are often tuned to maximize floating-point performance on easily parallelizable workloads that do not require significant network throughput.

In contrast, the concept for a new class of machines is gaining recognition as a vital resource to a broad array of users spanning the fields of health informatics, cyber-security, marketing, finance, and national defense. These machines will be required to process massive volumes of data, often in real-time. Many data-intensive computing applications are geared toward exploring large unstructured data sets while requiring mostly lightweight computational operations. Therefore, they place a much bigger demand on computer communications systems by requiring the frequent and fast exchange of many short messages, on the order of 1 kB or less. In these systems, it will be imperative to employ optical technologies that offer more than copper cable replacement. Electronically controlled photonic switches provide the capability to avoid pin- and power-limited electronic switch chips and the costly optoelectronic and electro-optic (EO) conversions that accompany them. Re-architecting the system around this premise may afford significant improvements to the bandwidth and latency characteristics of the interconnection network [4]. For this to occur, technological advancements in optical switching hardware are essential.

Three-dimensional (3-D) micro-electro-mechanical systems (MEMS) provide a mature optical switching technology that has demonstrated unparalleled port counts and successful commercialization [5]. Although their millisecond-scale reconfiguration times restrict the set of applications to which they may offer significant performance enhancements, certain computing applications that are characterized by long-lived communication patterns can benefit greatly by running over networks with 3-D MEMS switches [6], [7]. Moreover, using hybrid networks that employ MEMS-based optical circuit switching in parallel with electrical packet switching has been shown to deliver performance gains in datacenter environments [8], [9]. However, for data-intensive applications the communication patterns associated with exploring random data sets typically yield short-lived flows, making MEMS an ill-suited choice for these applications.

Various optical switching technologies have also been reported demonstrating microsecond-scale reconfiguration times [10], [11]. Using switches like these, researchers have extended the hybrid network concept, allowing the optical circuit switches to handle more diverse workloads and exist closer to the end hosts—even within the top-of-rack switch boxes [12], [13]. Nevertheless, microsecond reconfiguration times are still prohibitive

for the short-lived flows that exist in data-intensive workloads. Considering data rates on the order of 400-Gb/s per link (16 wavelength channels modulated at 25 Gb/s), typical packet durations could be less than 20 ns. Thus, if a switch is to operate at the packet granularity, reconfiguration times of a few nanoseconds or less are best. End-to-end reconfiguration times on the order of tens of nanoseconds could also be tolerated in multi-plane or multi-path networks or for applications that allow somewhat longer messages.

Certain planar photonic switching technologies have been shown to deliver high switch density and low power consumption with nanosecond-scale reconfiguration times. To date, research demonstrations have involved relatively low port counts, limited by high optical losses and crosstalk. In this article, we review recent photonic switch demonstrations and summarize our latest work. We provide insight into our approach for achieving scaled photonic switch hardware, discuss tradeoffs in design choices, and highlight areas that require improvement. Subsequently, we discuss concepts to break through the primary scaling limitations that currently impede the utility of photonic switch fabrics. Finally, we investigate which topologies lend themselves to photonic integration most appropriately, considering the loss- and crosstalk-limited design space. We conclude by offering an outlook for the future of photonic switching in computing systems.

II. NANOSECOND-SCALE PLANAR PHOTONIC SWITCH FABRICS

We are interested in technologies that can provide optical switching with reconfiguration times at the nanosecond scale and can be integrated in a planar high-density footprint. Some candidates include broadcast-and-select topologies based on III–V gate arrays [14]–[17], space/wavelength-selective cross-connects [18], and multi-stage topologies using interferometric switches with LiNbO₃ phase shifters [19]. We focus our discussion, however, on switch fabrics implemented on a silicon substrate due to potential advantages in cost, power, or area. Furthermore, although we are interested in switching technologies with nanosecond-scale reconfiguration times, we cite some works demonstrated with thermo-optic phase shifters that operate at the microsecond scale or longer. We include them because they may be straight-forwardly redesigned to employ faster electronic switch actuators, albeit with some performance penalty.

There have recently been numerous reports of four-to-eight port switch fabrics that meet the above criteria [20]–[27], all of which employ either ring resonators (RRs) or Mach–Zehnder (MZ) interferometers as the basic switching elements. MZs provide broad, continuous spectral coverage with a simple design, making them well-suited to switch wavelength-division multiplexed (WDM) signals aggregately, regardless of channel spacing—even in systems with loose specifications for the alignment of the wavelength channels to the defined grid. RRs potentially operate at much lower power and area than MZs, and likewise may be used to route WDM messages if the resonator’s free-spectral range equals the channel spacing [28]. Unfortunately, ensuring the wavelength alignment of these RRs typically requires control loops that complicate the design and add to the system’s switching energy. Furthermore, achieving

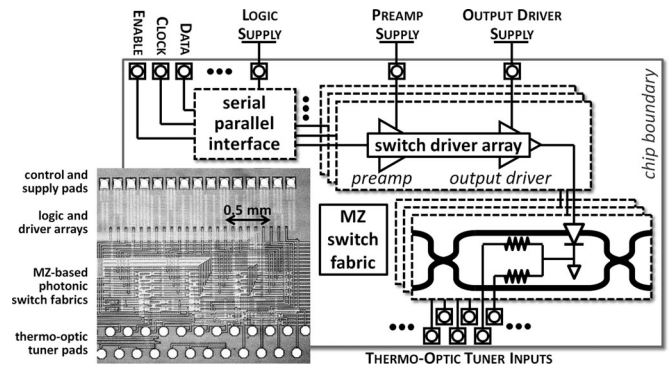


Fig. 1. Schematic diagram and inset die image of a test site that demonstrated monolithically integrated CMOS-driven scaled photonic switch fabrics [30].

a flat-top filter response with low group-delay ripple is imperative for preserving high-speed signal integrity, and generally requires a more complex design involving higher-order filters that add insertion loss [29].

These reported switch fabrics have been arranged in various topologies, including cross-point switch matrices [20], [21], combinations of cross-point switch matrices [22], tree-multiplexer switch matrices [23], Spanke-Beneš networks [24], and several custom topologies [25]–[27]. A comparison of the performance tradeoffs of a selection of topologies is saved for Section IV.

In our own work, we realized up to eight-port switch fabrics comprised of 2×2 MZ switching elements [30]. The demonstration was unique in that it was implemented with monolithically integrated digital CMOS logic and device drivers in IBM’s 90 nm Silicon Integrated Nanophotonics Technology [31]. The electronics included standard logic cells forming a serial-to-parallel interface for addressing each MZ in the fabric attached to inverter-based drivers for directly driving the switch electrodes. The photonics, which were designed to work with the transverse-electric polarization, included passive directional couplers, waveguide crossings, resistive thermo-optic phase tuners, and EO phase modulators implemented using forward-biased p-i-n diodes. Fig. 1 shows a schematic with an inset die image of a test site including the 4×4 and 8×8 photonic switch fabrics. In our past work, the topology has not been our primary design focus but will gain importance now that the technology has been demonstrated. For the building blocks of these fabrics—the MZ switches—we have shown devices that deliver reconfiguration times below 5 ns, a footprint of about 0.02 mm^2 , power dissipation under 2 mW, and crosstalk less than -15 dB over more than 75 nm of spectral bandwidth [30]. In the following paragraphs, we discuss the achieved switching time, crosstalk, footprint, and power dissipation in the context of scaling toward fabrics with higher port counts.

A. Reconfiguration Times

The reconfiguration times for our reported switch fabrics have consistently been below 10 ns [30]. Minimizing path setup time in circuit-switched networks with short-lived communication flows is a very important consideration, and reducing reconfiguration times to ~ 1 ns would have a positive impact on

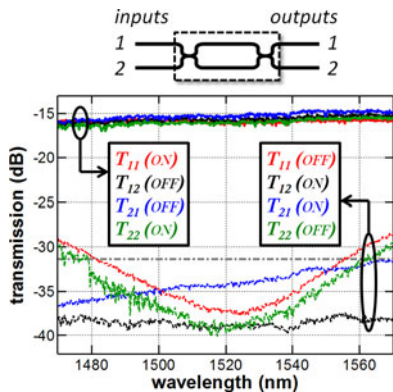


Fig. 2. Spectral response of a CMOS-driven broadband MZ 2×2 switching element, plotted for the four input/output port configurations (T_{11} , T_{12} , T_{21} , and T_{22}) in both the ON and OFF states. The gray dash-dot line denotes the -15 dB crosstalk threshold. Results originally published in [30].

throughput and latency. Certainly, techniques have been reported that quicken the transitions, both at the device- and system-level [32], [33], but this is not the most critical performance improvement needed for switch hardware, as 10-ns reconfiguration times can deliver good network performance with appropriate message sizes and scheduling algorithms [4].

B. Crosstalk

Fig. 2 displays the spectral characteristics of our 2×2 switching element [30]. In the figure, we use the notation T_{ij} to refer to the configuration where light is injected on input port i and collected from output port j . Eight curves are shown representing all four T_{ij} configurations in both the ON and OFF states. Each of the crosstalk curves remains below -15 dB over a 75-nm spectral bandwidth. At band center, near a wavelength of 1520 nm, the maximum crosstalk is -18 dB. The 2×2 switch lacks low-speed thermo-optic tuners, which were included in the larger fabrics to cancel phase errors that are generated by process imperfections. As a result, improvements in crosstalk can be expected in the 2×2 switch with the inclusion of phase tuners, ensuring the MZ can operate at its ideal bias point. Crosstalk can be further improved in devices operated under a push-pull drive scheme (discussed further in Section III-C). Overall, the level of crosstalk suppression demonstrated here is insufficient for achieving large-scale fabrics that maintain the required signal integrity. The values of crosstalk required of the switching elements are heavily design-dependent, but are discussed in more detail in Section III-C.

C. Footprint

For the multiport fabrics, the area occupied by the interface logic and switch drivers was 0.007 and 0.015 mm^2 per driver, respectively [30]. For example, a 16×16 fabric configured in a cross-point switch matrix would require 256 MZs. The total area of the photonics would be about 6.4 mm^2 using the area for the 2×2 switch described above and adding 25% overhead for routing wires and waveguides. The driver footprint would

be under 4 mm^2 , even using the over-sized drivers reported in previous demonstrations [34].

If monolithic integration of the drivers and switch fabric is not available, the case will be quite different. Assuming two thermo-optic tuners and two (push-pull) diode phase shifters are employed in each MZ, about 1000 bias and control signals plus at least as many ground connections will be needed in order to operate the fabric. For each of these pads to be connected to an external driver chip using 150- μm -pitch flip-chip solder bumps, an area of about 45 mm^2 will be required. This is not infeasible, but is substantially larger than the 10.4 mm^2 required for the photonics and electronics alone. Consequently, the digital control interface that is made possible through monolithic integration becomes increasingly advantageous as the fabric's scale increases since pad-limited footprints counter otherwise compelling densities. For our example here, an area penalty of $\sim 4.5\times$ is incurred when monolithic integration is not available. Furthermore, having logic very close to the switching devices may prove beneficial in implementing additional intelligence related to scheduling, controlling, or regulating the switch hardware.

D. Power Consumption

The 4×4 switch shown in our previous demonstrations had a total electrical power under 50 mW, about three-fourths of which resulted from the thermo-optic phase tuners. If four WDM channels are injected into each port, this power overhead can be shared among the 16 independent bit streams which are passing through the fabric, resulting in almost negligible added power to even the most efficient silicon photonic links reported [35], [36]. Nevertheless, accounting for the power forfeited due to optical losses is altogether different. Assuming optimistically an insertion loss of 1.5 dB in each MZ switch with another 1.5 dB lost in each coupling facet, and neglecting all other losses, the fabric insertion loss would be only 6 dB. If the switch is inserted into a link with a laser operating with 5 mW of output power and 20% wall-plug efficiency, and the laser power is increased to compensate the switch losses, an additional 75 mW of power will have to be added to each link passing through the switch. As a result, insertion losses may be the most beneficial area upon which to improve.

III. IMPROVING SCALABILITY

The degree to which photonic switch fabrics may be scaled to realize larger port counts or cascaded to achieve multiple network hops is limited chiefly by physical-layer impairments. This section addresses the limitations imposed by optical losses, the degradation imposed by amplification, and the inter-port crosstalk incurred in the switch elements. Other limitations resulting from amplifier and waveguide nonlinearities or dispersion may contribute as well but are not considered here.

A. Limitations of Optical Loss

Fig. 3(a) depicts an analysis of an optical link margin for a 16-channel WDM link passing through a future 16×16 photonic switch implemented with a simple butterfly network

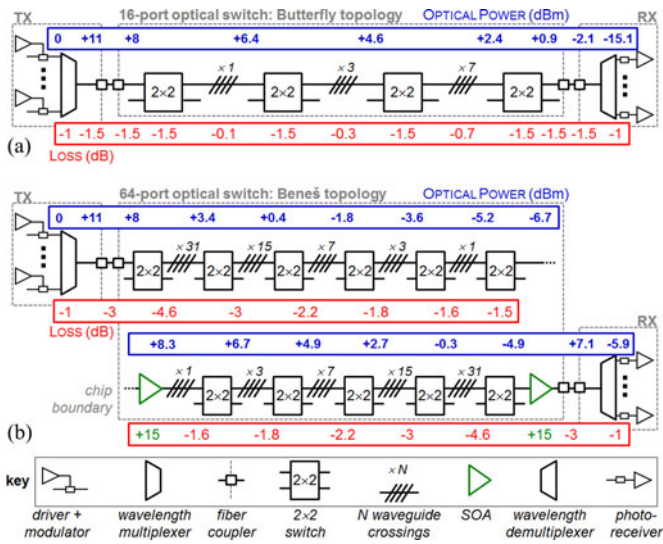


Fig. 3. Optical link budgets for (a) a 16-port butterfly network topology (see Fig. 7) without amplification and (b) a 64-port Beneš network topology (see Fig. 7) with two stages of SOAs. Number lines above the diagram track the aggregated average optical power, while those below the diagram denote the optical losses of a component or a few components.

topology. Our assumptions about the insertion loss of the various components encountered in the link are forward-looking—especially when accounting for the worst-case losses across the wavelength band, the switching states, and the manufacturing process tolerances. A MZ switch is assumed to incur 1.5 dB of loss, a waveguide crossing 0.1 dB, each waveguide/fiber transition 1.5 dB, and wavelength multiplexers and demultiplexers 1 dB. Moreover, the average optical power egressing from the modulator is taken to be 0 dBm, corresponding for instance to 2 mW of peak power with high extinction ratio. Even so, the link budget is impractical to close with < -15 dBm arriving at the photodetector. The most sensitive reported datacom-grade optical receivers at speeds of 25 Gb/s and above have shown sensitivities > -16.5 dBm for bipolar circuit implementations [37] and > -11 dBm for low-power CMOS implementations [38]. This leaves at best < 1.5 dB of margin before including additional penalties due to crosstalk in the switch fabric, optical and electrical inter-channel WDM crosstalk, relative intensity noise (RIN), interferometric noise, and other effects.

The diminished power efficiencies associated with increasing laser output power to facilitate larger link budgets has previously been discussed. In addition, there can be consequences from waveguide nonlinearities and two-photon absorption when optical power levels are too high. Therefore, optical amplification is necessary for scaling or cascading switch fabrics to achieve a network size that is suitable for addressing relevant data-intensive problem sizes. Semiconductor optical amplifiers (SOAs) are a natural choice given their ability to be integrated at the chip level, preserving the density advantages of the photonic switches. Although the power consumption of an SOA is large compared to other elements within the switch fabric, it is not impractical because its power will be amortized over multiple wavelength channels. Typical power consumption for an uncooled SOA is below 1 W, and often significantly less.

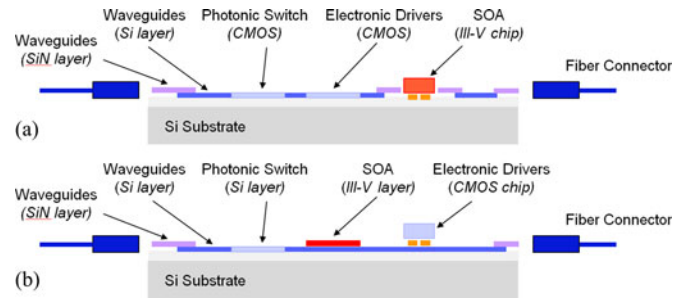


Fig. 4. Diagrams of two packaging platforms. (a) Monolithic integration of electronic driver and photonic switch in CMOS platform with flip-chip integration of SOA. (b) Heterogeneous integration of photonic switch and SOA with flip-chip integration of electronic driver IC.

For a 16-channel WDM link at speeds beyond 25 Gb/s, this corresponds to an added energy of less than 2.5 pJ/b per gain stage.

Fig. 3(b) depicts an analysis of an optical link margin for a 16-channel WDM optical link passing through a 64×64 photonic switch with integrated SOAs, implemented in a rearrangeably non-blocking Beneš network topology. Assuming 15-dB of net gain per SOA, two SOA stages provide more than enough amplification to compensate the optical losses. Given the well-known output power and dynamic range limitations of SOAs, their placement within the fabric and the wavelength grid assignment are important considerations. These are not wholly addressed here since they depend on the specifications of the particular SOAs employed. Nevertheless, it is apparent that a fully integrated approach with the ability to insert SOAs into arbitrary locations within the fabric—rather than at the inputs or outputs only—will lead to a more optimized system. Therefore, developing a novel optical and electrical packaging scheme is an important task. Previously, SOAs have been successfully integrated into planar waveguide platforms with silica-on-silicon waveguides by using passively aligned flip-chip bonding [39]. A few researchers have demonstrated flip-chip attachment of SOAs to silicon photonic platforms as well, where the smaller mode size requires tighter alignment tolerances even in the presence of spot-size converters [40]–[42]. However, only [42] has demonstrated the ability to couple both input and output facets to silicon waveguides, and this was done using a $4\text{-}\mu\text{m}$ thick silicon waveguide layer which is incompatible with traditional active silicon processing. Achieving the sub-micrometer alignment precision required to enable efficient optical coupling is a challenging task.

Fig. 4 highlights two plausible integration approaches. The first [see Fig. 4(a)] employs monolithic integration to combine the electronic drivers with the photonic switch elements, while using flip-chip integration to attach SOA arrays. This approach requires a photonics-enabled CMOS process—such as the one described in the previous section—that includes passive waveguides with enlarged mode-field diameter for efficient coupling to fibers and SOA devices. In this approach, the flip-chip integration is complicated by the alignment of the optical waveguides on the two surfaces. In addition to more stringent alignment tolerances, index-matching and facet preparation are essential

considerations for ensuring low coupling losses and reducing reflectivity. Here, even small reflections can introduce ripple in the gain spectrum or cause the SOA to act as a laser. Nevertheless, the benefits of monolithic electronic/photonic integration are many, and have already been discussed.

The second approach [see Fig. 4(b)] employs heterogeneous integration (e.g., wafer bonding [43]) to combine the photonic switch and optical amplifier functions, while using flip-chip integration to attach electronic drivers. Here, the flip-chip assembly is standard, not requiring sub-micrometer alignments or optical facet preparations. However, this approach requires the development of a silicon/III–V photonic platform with silicon p-n junction devices, which to the authors' knowledge has not been reported. This platform necessarily combines both silicon EO phase modulators and III–V optical amplifiers, while preferably offering similar performance to the bulk-optimized components. For highly-scaled fabrics, drawbacks in pin-limited density may be observed as well.

B. Limitations of Amplification Noise

With optical amplification integrated into the switch fabric, the design point moves from loss limited to noise limited. With each SOA encountered, the signal is degraded by the SOA's noise figure as amplified spontaneous emission is added to the signal spectrum. Furthermore, in saturation the SOA's fast gain dynamics result in pattern-dependent effects that distort the shapes of the waveforms and introduce inter-channel crosstalk. Therefore, with every SOA hop the window of acceptable input power levels (dynamic range) shrinks. Furthermore, since the gain spectrum is not flat, the spectral power distribution becomes less uniform with each hop. Channels experiencing too much gain or too little may be pushed outside of the dynamic range at the next hop, thus effectively shrinking the bandwidth.

Despite these challenges, cascaded SOA studies have demonstrated successful multichannel multi-hop transmissions [44]–[46]. Sun *et al.* demonstrate the transmission of a 32-channel WDM signal through a cascade of three inline SOAs with a saturated gain of 15 dB per stage [44]. Here, two methods are employed to improve signal integrity in the saturated-gain regime. The first is the use of many WDM channels which on average provides a steadier aggregate optical power level within the SOA and reduces gain fluctuations. The second is the addition of a reservoir channel, encoded in an analog fashion to counter variations in transmitted optical power among the other WDM channels. Spiekman *et al.* demonstrate the transmission of an eight-channel WDM signal through a cascade of five inline SOAs with 12–14 dB of gain per stage [45]. Moreover, SOAs were used as the exclusive amplifier technology in the link, which included three booster SOAs in the transmitter and one preamplifier SOA in the receiver, totaling nine cascaded SOAs. The SOAs were operated in the unsaturated regime yielding low distortion and crosstalk. Even much larger numbers of cascaded SOAs have been demonstrated when gain per stage has been kept very low and input power precisely controlled [46]. SOAs with large saturation output power can also improve performance by increasing the dynamic range [47]. Certainly, the design of

links with cascades of SOAs must be carefully considered, but researchers have demonstrated that it is possible.

C. Limitations of Crosstalk

Achieving acceptable crosstalk performance is another challenge for planar photonic switch fabrics, and must be addressed before large-scale implementations can be realized. Here, optical crosstalk is defined as the ratio of power present on the destructive port to that present on the constructive port of the MZ output coupler. There are two primary effects that contribute to crosstalk which we consider here: 1) phase errors and 2) power imbalance inside the MZ. Many factors can lead to one or both of these effects such as fabrication imprecisions, imperfect drive voltage, and temperature or wavelength sensitivity. Phase errors can be corrected during operation using thermo-optic phase tuners, as were utilized in the previous scaled switch fabric demonstrations [30]. Power imbalance is however a significant concern when using free-carrier plasma dispersion effect devices [48]. The injection of free-carriers in p-i-n diodes enables the realization of power- and area-efficient phase shifters that achieve nanosecond-scale switching times, but the inherent optical loss that results from free-carrier absorption leads to crosstalk by creating a power imbalance between the two arms of the MZ. For a given phase shifter design, there is therefore a limit to the minimum achievable crosstalk.

Assuming input and output couplers with an exact 3-dB split ratio and a relative phase shift between the arms of exactly π , the optical crosstalk at the destructive port can be described as:

$$10 \log_{10} \left[\tanh \left(\frac{1}{4} \Gamma L \alpha \right) \right]^2 \quad (1)$$

with Γ the optical confinement factor, L the length of the phase shifter, and α the free carrier absorption losses, which can be calculated using Soref and Bennett's empirical formula expressed as a function of free carrier density [49]. Assuming a phase shifter with a length of 250 μm and an optical confinement factor of 0.7 at a wavelength of 1550 nm, the free-carrier density required to achieve a π phase shift is $1.6 \times 10^{18} \text{ cm}^{-3}$, and the resulting free-carrier absorption loss is 23 cm^{-1} . This amount of imbalance generates about -20 dB of crosstalk.

In these calculations we assumed a simple structure consisting of a MZ with only one EO phase shifter. If we instead use a phase shifter in each arm and drive them in a push-pull manner, a phase shift of only $\pi/2$ per arm is needed. The carrier density is relaxed, and the free carrier absorption is reduced. Using the same phase shifter, we find that the required carrier density is $7.0 \times 10^{17} \text{ cm}^{-3}$, and the free-carrier absorption loss is 10 cm^{-1} , which yields -27 dB of crosstalk.

Fig. 5 presents the simulated transmission spectra, comparing single-ended and push-pull drive. The spectra were obtained using the transfer matrix approach described in [50]. We investigate a switch that employs the same phase shifter described above and simple directional couplers with relatively narrow bandwidth optimized at a wavelength of 1550 nm. For each case, we plot the transmission of both outputs in the ON-state (π phase difference) and OFF-state (0 phase difference), the input

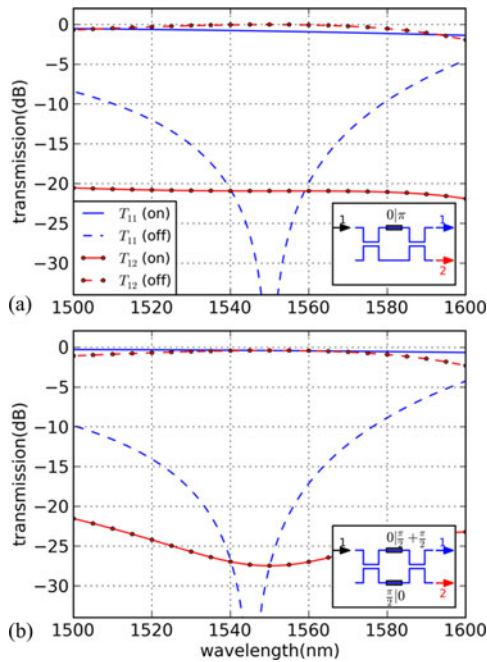


Fig. 5. Simulation of the transmission spectra of a 2×2 MZ switch under (a) single-ended and (b) push-pull drive. Insets show the MZ driving scheme.

port being fixed. In the following, we conform to the T_{ij} notation described in Section II. The insets in Fig. 5 show single-ended and push-pull drive schematics. In Fig. 5(a), only one phase shifter is used and is modulated from 0 to π (OFF to ON). In Fig. 5(b), the first arm's phase shifter is modulated from 0 to $\pi/2$ while the second arm's is modulated from $\pi/2$ to 0. Push-pull drive requires a fixed extra path length of $\pi/2$ in the first arm.

Concerning insertion losses, in the single-ended case the loss at band center of T_{11} (ON) is 0.9 dB while T_{12} (OFF) experiences no loss in our idealized scenario. When operating push-pull, this asymmetry disappears because there is always one phase shifter being operated. In that case, the insertion loss is 0.4 dB for both T_{11} (ON) and T_{12} (OFF).

Concerning crosstalk, we find at the center wavelength of the directional coupler (1550 nm) the same crosstalk limits as calculated above. When moving away from the optimal wavelength of the directional coupler, the coupling coefficient decreases (to shorter wavelengths) or increases (to longer wavelengths), which is an additional mechanism generating power imbalance in the MZ. Thus, the limit of crosstalk is given in equation (1), while the bandwidth of the directional coupler establishes the spectral shape of the crosstalk curve for a given port configuration as shown in Fig. 5.

The existence of crosstalk in the switch has significant implications on scalability, since each MZ can add its crosstalk to the signal's noise. The crosstalk impacts both the amplitude margin (vertical eye opening) and timing margin (horizontal eye opening), and limits the ultimate scale that can be achieved. The wavelength dependence of the switch, as noted above, further aggravates the crosstalk effect so that scalability and utilized bandwidth are correlated for a given switch design.

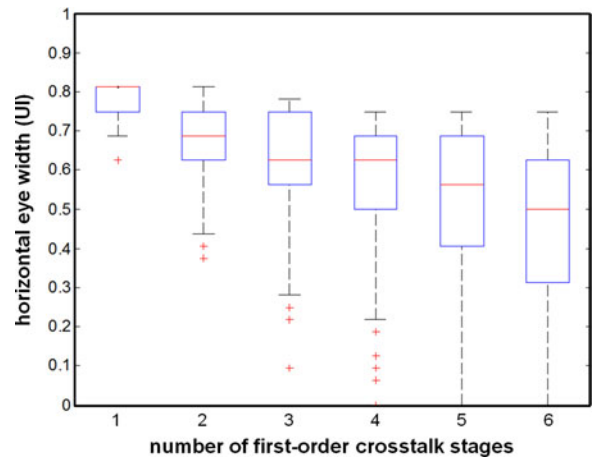


Fig. 6. Statistical results plotting the horizontal eye width, measured in unit intervals (UI) and evaluated at a BER of 10^{-12} , as a function of the number of first-order crosstalk stages encountered.

Using an analytical model for the simulated switch with push-pull drive, we conducted an analysis of the impact of crosstalk on the achievable switch size. We built a full link model which incorporates the switch model, using typical numbers for the transmitter optical modulation amplitude, receiver sensitivity, laser RIN, connector reflection, and loss. At the receiver output we calculated the horizontal eye opening at a bit error rate (BER) of 10^{-12} . Other link parameters, like the wavelengths of the two interfering signals and the receiver bandwidth, also affect the magnitude of the crosstalk. Due to the random nature of the laser wavelengths of the signals propagating through the switch, and random phase at each first-order crosstalk stage, the simulation was repeated 1000 times to probe the possible parameter space. The link power budget was modeled at 13 dB with 5 dB of loss present in the link at various connectors and couplers.

Using boxplots (see Fig. 6), we show representative results illustrating the dependence of the horizontal eye opening on the number of first-order crosstalk stages encountered. We investigate the dependence only considering first-order crosstalk, ignoring all others. (A first-order crosstalk stage is one in which signals from two input ports simultaneously pass through to two output ports without contention [51].) The top and bottom solid lines (blue) in the box represent the 25th and 75th percentile of the data, and the line in the middle of the box (red) denotes the median. The data extends one and a half times the interquartile range. The whiskers (dashed lines) cover $\sim 99.3\%$ of the data, while the crosses (red) outside the whiskers represent outliers, but are included for thoroughness.

These results indicate that the horizontal eye window significantly narrows after only the fourth first-order crosstalk stage, and some links will have closed eyes after only five hops. While the model relies on many estimated parameters (such as the distribution of the lasing wavelengths of interfering channels), we feel that they are all based on sound projections. Nevertheless, the results are intended to be qualitative—indicating that even just a few stages of crosstalk can significantly degrade the link. On the other hand, by selecting appropriate topologies, significantly sized fabrics can be realized with limited occurrences of first-order crosstalk.

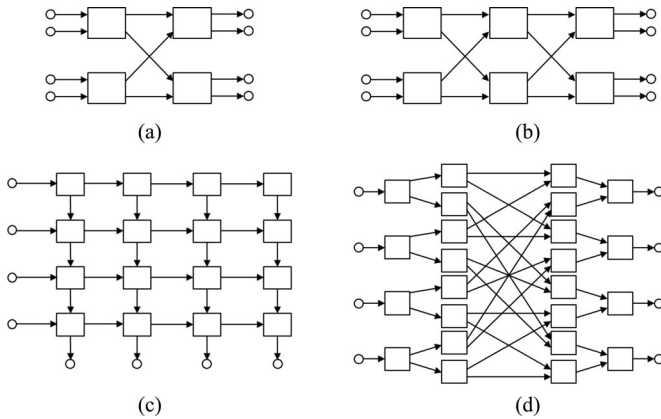


Fig. 7. Schematics of four 4×4 switch topologies constructed from input/output ports (\circ), 2×2 switch elements (\square), and internal connections (\rightarrow). The four topologies are: (a) butterfly network, (b) Beneš network, (c) cross-point switch matrix, and (d) tree-multiplexer switch matrix.

IV. FABRIC TOPOLOGIES

In this section, we investigate the loss and crosstalk limitations within four fabric topologies. We consider the butterfly network, the Beneš network, a cross-point switch matrix, and a tree-multiplexer switch matrix (also called a switch-and-select topology [23]) for port counts scaling up to 128 ports. These four networks shown in Fig. 7 give a relevant slice of the design tradeoffs that exist between the physical-layer impairments and the routing and scheduling complexity. The butterfly achieves a given scale with a minimum number of switch hops, representing the best case possible from the physical layer. As a result, it provides no path diversity, thus suffering from internal blocking characteristics. The Beneš is rearrangeably non-blocking, while the cross-point and tree-multiplexer switch matrices are strictly non-blocking.

For our investigation, we are concerned with the optical losses incurred by switch elements and waveguide crossings. Optical coupling and propagation losses are approximately equal for all of the topologies at a given scale. Fig. 8(a) and (b) plots the maximum number of switch elements and waveguide crossings, respectively, encountered by a signal traversing these four networks. Note that the cross-point has zero waveguide crossings, and thus does not appear in Fig. 8(b). Borrowing the assumptions of optical losses from Section III-A, the 32-port cross-point switch would incur up to 95 dB of loss for the worst path and as little as 1.5 dB for the best path. This severe non-uniformity renders the topology impractical for photonic networks of more than a few ports. In the other three topologies all paths through the network undergo the same number of MZ hops, with just the variation in waveguide crossings encountered contributing to non-uniformity.

The tree-multiplexer switch, aside from being strictly non-blocking, has several favorable qualities. Firstly, it employs only 1×2 and 2×1 switching elements. Thus, crosstalk is reduced to only second- and higher-order effects. Even more, algorithms can be devised to limit the second-order crosstalk, as has been shown in a related topology: the dilated banyan [51]. Addition-

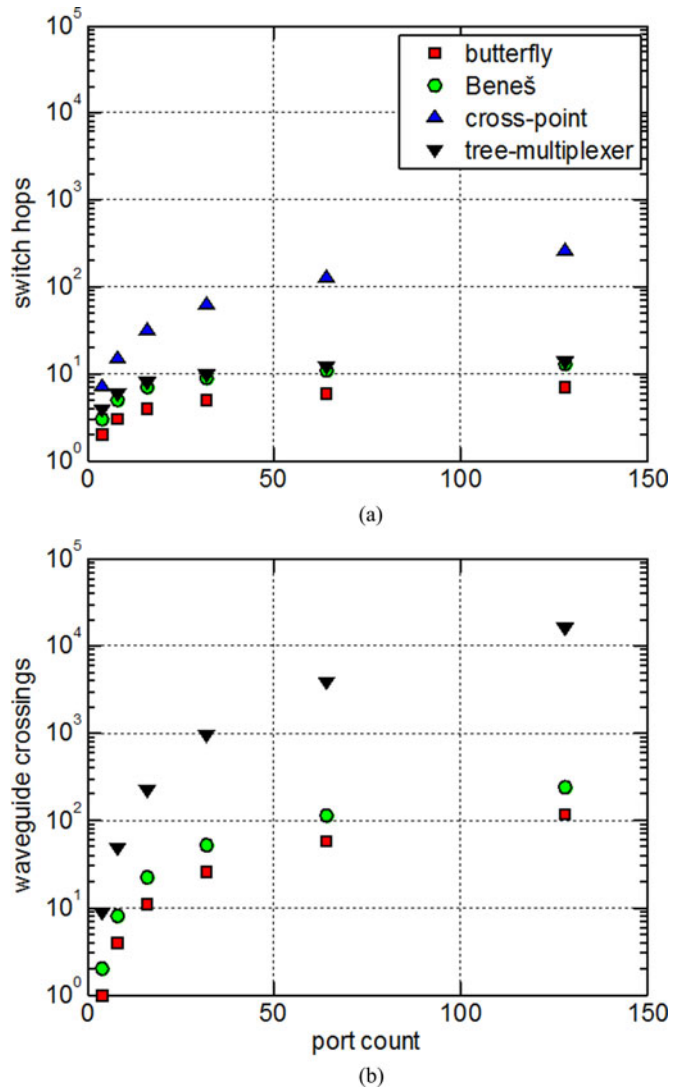


Fig. 8. Plots showing the number of (a) MZ hops and (b) waveguide crossings encountered in the worst-case path through the four representative topologies as a function of port count.

ally, the tree-multiplexer switch contains only a single additional MZ hop for a given scale compared to the Beneš network. Its most significant drawback, however, is its high crossing count. A signal propagating through a 32-port tree-multiplexer switch undergoes nearly one thousand waveguide crossings in the worst case, while encountering zero in the best case. Here, an ultra-low loss and low crosstalk waveguide crossing solution (e.g., multi-planar waveguide routing [52]) is needed in order for the tree-multiplexer switch to be a practical choice.

The butterfly network and Beneš network provide the best scaling performance, and should be chosen if (1) a blocking or a rearrangeably non-blocking topology can be tolerated by the routing and scheduling controllers and (2) the crosstalk is sufficiently low for the required scale that link integrity may be maintained through the fabric. In general, more work will be needed. A hybrid topology that draws from two or more traditional architectures may find an acceptable compromise.

V. OUTLOOK

In this work we have reviewed switch fabrics that are capable of achieving nanosecond-scale reconfigurability, low electrical power consumption, and high integration density. To date, these fabrics have been restricted to low-port-count implementations, limited by the physical-layer constraints of loss and crosstalk. If silicon photonic switch fabrics are to make an impact in computer communications systems, it is likely that a first-insertion point would be in a highly interconnected mid-size system with hundreds of nodes rather than in extreme-scale high-performance computing systems or datacenters with hundreds of thousands of nodes. Consequently, we have highlighted a target application of data-intensive computing, where the benefits of optical switching could potentially translate into computational performance gains, and where system sizes are not expected to soon reach extreme scales.

Nevertheless, exceeding the port counts of current implementations will certainly be required. We have outlined an approach to scale port counts using integrated optical amplification to overcome losses. Two manufacturing and packaging platforms are proposed as realistic choices for integrating arrays of SOAs with planar switch fabrics. Both approaches have drawbacks that require advancements in packaging technologies. New approaches to packaging or novel integration platforms may deliver improved performances beyond what is envisioned here.

We have further attempted to lay out the chief implementation challenges associated with delivering a scaled silicon photonic switch fabric consistent with the above-referenced packaging schemes. Amplification noise and crosstalk, as well as switch crosstalk, must be carefully controlled. In addition, reflections from various interfaces must be managed with high precision. However, with proper design we have shown a path toward reaching the desired goals. We have shown that this path hinges on device improvements as well as on novel topology designs that impact these physical layer constraints.

Apart from the issues described here, there are many other uncertainties as well. Most of the switches reported today work for a single polarization. A dual-polarization switch or a polarization-diverse implementation adds complexity, loss, and power consumption to any design, but will be required in robust implementations. Temperature tolerance, generally less of a concern with MZ switches than with RR switches, cannot be overlooked, especially if uncooled SOA operation is required for low-power solutions. Finally, nanosecond-reconfigurable switch fabrics are not beneficial if link retiming and retraining sequences take microseconds or longer to synchronize. Therefore, burst-mode receivers that operate over a large range of input power levels and have locking times similar to switch reconfiguration times are critical. If these and other challenges can be successfully resolved, photonic switch fabrics may provide a major performance enhancement to data-intensive computer communications systems.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Shah and Dr. N. DiLello of the Defense Advanced Research Projects Agency, Microsys-

tems Technology Office, and Dr. W. Chang of the Army Research Laboratory for inspiration and support. The IBM authors would also like to thank their many colleagues at IBM Research that provided helpful discussions and technical expertise.

REFERENCES

- [1] D. Chen, N. A. Easley, P. Heidelberger, R. M. Senger, Y. Sugawara, S. Kumar, V. Salapura, D. Satterfield, B. Steinmacher-Burow, and J. Parker, "The IBM blue gene/Q interconnection fabric," *IEEE Micro*, vol. 32, no. 1, pp. 32–43, Jan./Feb. 2012.
- [2] A. Benner, D. M. Kuchta, P. K. Pepeljugoski, R. A. Budd, G. Hougham, B. V. Fasano, K. Marston, H. Bagheri, E. J. Seminario, H. Xu, D. Meadowcroft, M. H. Fields, L. McColloch, M. Robinson, F. W. Miller, R. Kaneshiro, R. Granger, D. Childers, and E. Childers, "Optics for high-performance servers and supercomputers," presented at the *Opt. Fiber Commun. Conf.*, San Diego, CA, USA, Mar. 2010, Paper OTuH1.
- [3] TOP500 Supercomputer Site. (2014). [Online]. Available: <http://top500.org/>
- [4] L. Schares, B. G. Lee, F. Checcoli, R. Budd, A. Ryljakov, N. Dupuis, F. Petriani, C. L. Schow, P. Fuentes, O. Mattes, and C. Minkenberg, "A throughput-optimized optical network for data-intensive computing," *IEEE Micro*, vol. 34, no. 5, pp. 52–63, Sep./Oct. 2014.
- [5] M. Wu, O. Solgaard, and J. E. Ford, "Optical MEMS for lightwave communication," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4433–4454, Dec. 2006.
- [6] L. Schares, X. J. Zhang, R. Wagle, D. Rajan, P. Selo, S. P. Chang, J. Giles, K. Hildrum, D. Kuchta, J. Wolf, and E. Schenfeld, "A reconfigurable interconnect fabric with optical circuit switch and software optimizer for stream computing systems," presented at the *Opt. Fiber Commun. Conf.*, San Diego, CA, USA, Mar. 2009, Paper OTuA1.
- [7] S. Kamil, L. Oliker, A. Pinar, and J. Shalf, "Communication requirements and interconnect optimization for high-end scientific applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 2, pp. 188–202, Feb. 2010.
- [8] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: A hybrid electrical/optical switch architecture for modular data centers," in *Proc. ACM SIGCOMM Conf.*, Aug. 2010, pp. 339–350.
- [9] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. E. Ng, M. Kozuch, and M. Ryan, "c-through: Part-time optics in data centers," in *Proc. ACM SIGCOMM Conf.*, Aug. 2010, pp. 327–338.
- [10] T. Strasser and J. Wagener, "Wavelength-selective switches for ROADM applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1150–1157, Sep./Oct. 2010.
- [11] S. Han, T. J. Seok, N. Quack, B.-W. Yoo, and M. C. Wu, "Monolithic 50 × 50 MEMS silicon photonic switches with microsecond response time," presented at the *Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, Paper M2K.2.
- [12] N. Farrington, A. Forencich, P.-C. Sun, S. Fainman, J. Ford, A. Vahdat, G. Porter, and G. C. Papen, "A 10 μs hybrid optical-circuit/electrical-packet network for datacenters," presented at the *Opt. Fiber Commun. Conf.*, Anaheim, CA, USA, Mar. 2013, Paper OW3H.3.
- [13] H. Liu, F. Lu, R. Kapoor, A. Forencich, G. M. Voelker, G. Papen, A. C. Snoeren, and G. Porter, "REACToR: A reconfigurable packet and circuit ToR switch," presented at the *Photon. Soc. Summer Top. Meet.*, Waikoloa, Hawaii, Jul. 2013, Paper TuE3.1.
- [14] M. Gustavsson, B. Lagerström, L. Thylén, M. Janson, L. Lundgren, A.-C. Mörner, M. Rask, and B. Stoltz, "Monolithically integrated 4 × 4 InGaAsP/InP laser amplifier gate switch arrays," *Electron. Lett.*, vol. 28, no. 24, pp. 2223–2225, Nov. 19, 1992.
- [15] T. Segawa, M. Nada, M. Nakamura, Y. Suzuki, and R. Takahashi, "An 8 × 8 broadcast-and-select optical switch based on monolithically integrated EAM-gate array," presented at the *Eur. Conf. Opt. Commun.*, London, U.K., Sep. 2013, Paper We.4.B.1.
- [16] A. Wonfor, H. Wang, R. V. Pentyl, and I. H. White, "Large port count high-speed optical switch fabric for use within datacenters," *J. Opt. Commun. Netw.*, vol. 3, no. 8, pp. A32–A39, Aug. 2011.
- [17] R. Stabile, A. Alboreo-Mejia, and K. Williams, "Monolithic active-passive 16 × 16 optoelectronic switch," *Opt. Lett.*, vol. 37, no. 22, pp. 4666–4668, Nov. 15, 2012.
- [18] R. Stabile, A. Rohit, and K. A. Williams, "Monolithically integrated 8 × 8 space and wavelength selective cross-connect," *J. Lightw. Technol.*, vol. 32, no. 2, pp. 201–207, Jan. 15, 2014.

- [19] J. Watson, M. A. Milbrodt, K. Bahadori, M. F. Dautartas, C. T. Kemmerer, D. T. Moser, A. Schelling, T. O. Murphy, J. Veselka, and D. A. Herr, "A low-voltage 8×8 Ti:LiNbO₃ switch with a dilated-Beneš architecture," *J. Lightw. Technol.*, vol. 8, no. 5, pp. 794–801, May 1990.
- [20] A. W. Poon, X. Luo, F. Xu, and H. Chen, "Cascaded microresonator-based matrix switch for silicon on-chip optical interconnection," *Proc. IEEE*, vol. 97, no. 7, pp. 1216–1238, Jul. 2009.
- [21] P. DasMahapatra, R. Stabile, A. Rohit, and K. A. Williams, "Optical crosspoint matrix using broadband resonant switches," *J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, 5900410, pp. 1–10, Jul./Aug. 2014.
- [22] K. Suzuki, K. Tanizawa, T. Matsukawa, G. Cong, S.-H. Kim, S. Suda, M. Ohno, T. Chiba, H. Tadokoro, M. Yanagihara, Y. Igarashi, M. Masahara, S. Namiki, and H. Kawashima, "Ultra-compact 8×8 strictly-non-blocking Si-wire PILOSS switch," *Opt. Exp.*, vol. 22, no. 4, pp. 3887–3894, Feb. 24, 2014.
- [23] L. Chen and Y.-K. Chen, "Compact, low-loss and low-power 8×8 broadband silicon optical switch," *Opt. Exp.*, vol. 20, no. 17, pp. 18977–18985, Aug. 13, 2012.
- [24] J. Xing, Z. Li, P. Zhou, X. Xiao, J. Yu, and Y. Yu, "Nonblocking 4×4 silicon electro-optic switch matrix with push-pull drive," *Opt. Lett.*, vol. 38, no. 19, pp. 3926–3929, Oct. 1, 2013.
- [25] N. Sherwood-Droz, H. Wang, L. Chen, B. G. Lee, A. Biberman, K. Bergman, and M. Lipson, "Optical 4×4 hitless silicon router for optical networks-on-chip (NoC)," *Opt. Exp.*, vol. 16, no. 20, pp. 15915–15922, Sep. 29, 2008.
- [26] R. Ji, J. Xu, and L. Yang, "Five-port optical router based on microring switches for photonic networks-on-chip," *Photon. Technol. Lett.*, vol. 25, no. 5, pp. 492–495, Mar. 1, 2013.
- [27] X. Li, X. Xiao, H. Xu, Z. Li, T. Chu, J. Yu, and Y. Yu, "Mach-Zehnder-based five-port silicon router for optical interconnects," *Opt. Lett.*, vol. 38, no. 10, pp. 1703–1705, May 15, 2013.
- [28] B. G. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, "All-optical comb switch for multiwavelength message routing in silicon photonic networks," *Photon. Technol. Lett.*, vol. 20, no. 10, pp. 767–769, May 15, 2008.
- [29] F. Xia, L. Sekaric, M. O. Boyle, and Y. Vlasov, "Coupled resonator optical waveguides based on silicon-on-insulator photonic wires," *Appl. Phys. Lett.*, vol. 89, pp. 041122-1–041122-3, Jul. 26, 2006.
- [30] B. G. Lee, A. V. Rylyakov, W. M. J. Green, S. Assefa, Christian W. Baks, R. Rimolo-Donadio, D. M. Kuchta, M. H. Khater, T. Barwicz, C. Reinholm, E. Kiewra, S. M. Shank, C. L. Schow, and Y. A. Vlasov, "Monolithic silicon integration of scaled photonic switch fabrics, CMOS logic, and device driver circuits," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 743–751, Feb. 15, 2014.
- [31] S. Assefa, S. Shank, W. Green, M. Khater, E. Kiewra, C. Reinholm, S. Kamalpurkar, A. Rylyakov, C. Schow, F. Horst, H. Pan, T. Topuria, P. Rice, D. M. Gill, J. Rosenberg, T. Barwicz, M. Yang, J. Proesel, J. Hofrichter, B. Offrein, X. Gu, W. Haensch, J. E.-Monaghan, and Y. Vlasov, "A 90 nm CMOS integrated nano-photonics technology for 25 Gbps WDM optical communications applications," presented at the *Int. Electron Devices Meet.*, San Francisco, CA, USA, Dec. 2012, Paper 33.8.
- [32] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators," *Opt. Exp.*, vol. 15, no. 2, pp. 430–436, Jan. 22, 2007.
- [33] W. Green, M. J. Rooks, L. Sekaric, and Y. A. Vlasov, "Ultra-compact, low RF power, 10 Gb/s silicon Mach-Zehnder modulator," *Opt. Exp.*, vol. 15, no. 25, pp. 17106–17113, Dec. 10, 2007.
- [34] B. Lee, C. L. Schow, A. V. Rylyakov, J. V. Campenhout, W. M. J. Green, S. Assefa, F. E. Doany, M. 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," *J. Lightw. Technol.*, vol. 29, no. 8, pp. 1136–1142, Apr. 15, 2011.
- [35] X. Zheng, Y. Luo, J. Lexau, F. Liu, G. Li, H. D. Thacker, I. Shubin, J. Yao, R. Ho, J. E. Cunningham, and A. V. Krishnamoorthy, "2-pJ/bit (on-chip) 10-Gb/s digital CMOS silicon photonic link," *Photon. Technol. Lett.*, vol. 24, no. 14, pp. 1260–1262, Jul. 15, 2012.
- [36] N. Dupuis, B. Lee, J. Proesel, A. V. Rylyakov, R. Rimolo-Donadio, C. Baks, C. L. Schow, A. Ramaswamy, J. E. Roth, R. Guzzon, B. Koch, D. K. Sparacin, and G. Fish, "30 Gbps optical link utilizing heterogeneously integrated III-V/Si photonics and CMOS circuits," presented at the *Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, Paper Th5A.6.
- [37] A. Rylyakov, C. Schow, J. Proesel, D. M. Kuchta, C. Baks, N. Y. Li, C. Xie, and K. Jackson, "A 40-Gb/s, 850-nm, VCSEL-based full optical link," presented at the *Opt. Fiber Commun. Conf.*, Los Angeles, CA, USA, Mar. 2012, Paper OThE.1.
- [38] J. Proesel, B. G. Lee, C. W. Baks, and C. Schow, "35-Gb/s VCSEL-based optical link using 32-nm SOI CMOS circuits," presented at the *Opt. Fiber Commun. Conf.*, Anaheim, CA, USA, Mar. 2013, Paper OM2H.2.
- [39] G. Maxwell, A. Poustie, C. Ford, M. Harlow, P. Townley, M. Nield, I. Lealman, S. Oliver, L. Rivers, and R. Waller, "Hybrid integration of monolithic semiconductor optical amplifier arrays using passive assembly," in *Proc. Electron. Compon. Technol. Conf.*, 2005, pp. 1349–1352.
- [40] S. Tanaka, S.-H. Jeong, S. Sekiguchi, T. Kurahashi, Y. Tanaka, and K. Morito, "High-output-power, single-wavelength silicon hybrid laser using precise flip-chip bonding technology," *Opt. Exp.*, vol. 20, no. 27, pp. 28057–28069, Dec. 17, 2012.
- [41] C. Stamatiadis, L. Stampoulidis, K. Vysokinos, I. Lazarou, D. Kalavrouziotis, L. Zimmermann, K. Voigt, G. Preve, L. Moerl, J. Kreissl, and H. Avramopoulos, "A hybrid photonic integrated wavelength converter on a silicon-on-insulator substrate," presented at the *Proc. Opt. Fiber Commun. Conf.*, Los Angeles, CA, USA, Mar. 2012, Paper OM3E.1.
- [42] D. Fitsios, T. Alexoudi, G. T. Kanellos, K. Vysokinos, N. Pleros, T. Tekin, M. Cherchi, S. Ylino, M. Harjanne, M. Kapulainen, and T. Aalto, "Dual SOA-MZI wavelength converters based on III-V hybrid integration on a μm -scale Si platform," *IEEE Photon. Technol. Lett.*, vol. 26, no. 6, pp. 560–563, Mar. 15, 2014.
- [43] A. W. Fang, H. Park, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "Electrically pumped hybrid AlGaInAs-silicon evanescent laser," *Opt. Exp.*, vol. 14, no. 20, pp. 9203–9210, Oct. 2, 2006.
- [44] Y. Sun, A. K. Srivastava, S. Banerjee, J. W. Sulhoff, R. Pan, K. Kantor, R. M. Jopson, and A. R. Chraplyvy, "Error-free transmission of 32×2.5 Gbit/s DWDM channels over 125 km using cascaded in-line semiconductor optical amplifiers," *Electron. Lett.*, vol. 35, no. 21, pp. 1863–1865, Oct. 14, 1999.
- [45] L. H. Spiekman, J. M. Wiesenfeld, A. H. Gnauck, L. D. Garrett, G. N. van den Hoven, T. van Dongen, M. J. H. Sander-Jochem, and J. J. M. Binsma, "8 \times 10 Gb/s DWDM transmission over 240 km of standard fiber using a cascade of semiconductor optical amplifiers," *Photon. Technol. Lett.*, vol. 12, no. 8, pp. 1082–1084, Aug. 2000.
- [46] O. Liboiron-Ladouceur, B. Small, and K. Bergman, "Physical layer scalability of WDM optical packet interconnection networks," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 262–270, Jan. 2006.
- [47] K. Morito, S. Tanaka, S. Tomabechei, and A. Kuramata, "A broad-band MQW semiconductor optical amplifier with high saturation output power and low noise figure," *Photon. Technol. Lett.*, vol. 17, no. 5, pp. 974–976, May 2005.
- [48] J. Van Campenhout, W. M. J. Green, S. Assefa, and Y. A. Vlasov, "Low-power, 2×2 silicon electro-optic switch with 110-nm bandwidth for broadband reconfigurable optical networks," *Opt. Exp.*, vol. 17, no. 26, pp. 24020–24029, Dec. 21, 2009.
- [49] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.*, vol. QE-23, no. 1, pp. 123–129, Jan. 1987.
- [50] J. Van Campenhout, W. M. J. Green, and Y. A. Vlasov, "Design of a digital, ultra-broadband electro-optic switch for reconfigurable optical networks-on-chip," *Opt. Exp.*, vol. 17, no. 26, pp. 23793–23808, Dec. 11, 2009.
- [51] Y. Qian, H. Mehrvar, H. Ma, X. Yang, H. y. Fu, K. Zhu, D. Geng, D. Goodwill, P. Dumais, and E. Bernier, "Crosstalk optimization in low extinction-ratio switch fabrics," presented at the *Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, Mar. 2014, Paper th1i.4.
- [52] N. Sherwood-Droz and M. Lipson, "Scalable 3D dense integration of photonics on bulk silicon," *Opt. Exp.*, vol. 19, no. 18, pp. 17758–17765, Aug. 29, 2011.

Benjamin G. Lee (M'04–SM'14) received the B.S. degree from Oklahoma State University, Stillwater, OK, USA, in 2004, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, USA, in 2006 and 2009, respectively, all in electrical engineering.

In 2009, he became a Postdoctoral Researcher at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, where he is currently a Research Staff Member. He is also an Assistant Adjunct Professor of electrical engineering with Columbia University. His research interests include silicon photonic devices, integrated optical switches and networks for high-performance computing systems and datacenters, and highly parallel multimode transceivers.

Dr. Lee is a Member of the Optical Society and the IEEE Photonics Society, where he serves as an Associate Vice-President of Membership. He currently serves on the Technical Program Committees for the Optical Fiber Communications Conference and the Optical Interconnects Conference.

Nicolas Dupuis received the B.S. and M.S. degrees from Universite Blaise Pascal, Clermont-Ferrand, France, and the Ph.D. degree from Universite de Lorraine, Metz, France, in 2009, all in physics. He is currently a Postdoctoral Researcher at the IBM T.J. Watson Research Center, Yorktown Heights, NY, USA. His research interests include silicon photonics, optical switching, and optical link modeling. Before joining IBM, he was with Bell Laboratories, Crawford Hill, NJ, USA, working on high-speed InP-based photonic circuits.

Justin R. Bickford (M'97) received the B.S. degree from the Rochester Institute of Technology (RIT), Rochester, NY, USA, in 2002, and the M.S. and Ph.D. degrees from the University of California, San Diego (UCSD), CA, USA, in 2004 and 2008, respectively, all in electrical engineering.

In 2000, he was an Assistant Process Engineer with Texas Instruments. From 2001 to 2002, he was an undergraduate Research Assistant and Teaching Assistant with RIT. From 2004 to 2008, he was a graduate Research Assistant and Teaching Assistant with UCSD. Since 2008, he has been an Optoelectronics Researcher with the U.S. Army Research Laboratory, Adelphi, MD, USA. His research interests include MOCVD growth, heterogeneous monolithic integration, III-V, II-VI, and Si-photonic materials characterization, micro- and nano-device fabrication, RF photonics, and computational imaging.

Dr. Bickford is a Member of the American Vacuum Society.

Clint L. Schow (SM'10) received the Ph.D. degree in electrical engineering from the University of Texas at Austin, Austin, TX, USA, in 1999. He joined IBM in Rochester, MN, USA, assuming responsibility for the optical receivers used in IBM's optical transceiver business. From 2001 to 2004, he was with Agility Communications in Santa Barbara, CA, USA, developing high-speed optoelectronic modulators and tunable laser sources. In 2004, he joined the IBM T.J. Watson Research Center, Yorktown Heights, NY, USA, as a Research Staff Member and currently manages the Optical Link and System Design group responsible for optics in future generations of servers and supercomputers. He has directed multiple DARPA-sponsored programs investigating chip-to-chip optical links, nanophotonic switches, and future systems utilizing photonic switching fabrics. He is a Senior Member of the OSA, has published more than 125 journal and conference articles, and has ten issued and more than 20 pending patents.

Petar Pepeljugoski, biography not available at the time of publication.

Laurent Schares, biography not available at the time of publication.

Russell Budd, biography not available at the time of publication.