Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects

Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects

Xian Xiao, Roberto Proietti, Gengchen Liu, Hongbo Lu, Pouya Fotouhi, Sebastian Werner, Yu Zhang, and S. J. Ben Yoo, *Fellow, IEEE*

***Abstract*—This paper reports the first experimental demonstration of silicon photonic (SiPh) Flex-LIONS, a bandwidth-reconfigurable SiPh switching fabric based on wavelength routing in arrayed waveguide grating routers (AWGRs) and space switching. Compared with the state-of-the-art bandwidth-reconfigurable switching fabrics, Flex-LIONS architecture exhibits 21× less number of switching elements and 2.9× lower on-chip loss for 64 ports, which indicates significant improvements in scalability and energy efficiency. System experimental results carried out with an 8-port SiPh Flex-LIONS prototype demonstrate error-free one-to-eight multicast interconnection at 25 Gb/s and bandwidth reconfiguration from 25 Gb/s to 100 Gb/s between selected input and output ports. Besides, benchmarking simulation results show that Flex-LIONS can provide a 1.33× reduction in packet latency and > 1.5× improvements in energy efficiency when replacing the core layer switches of Fat-Tree topologies with Flex-LIONS. Finally, we discuss the possibility of scaling Flex-LIONS up to N=1024 ports (N=M×W) by arranging M^2 W-port Flex-LIONS in a Thin-CLOS architecture using W wavelengths.**

***Index Terms*—Arrayed waveguide grating router, optical interconnections, optical switches, photonic integrated circuits, silicon photonics.**

I. Introduction

Modern high performance computing (HPC) and datacenter systems are growingly adopting heterogeneous memory and processor nodes (Fig. 1 (a)) to better utilize resources for various tasks [1,2]. The communication patterns in such systems driven by modern workloads tend to be temporally bursty and spatially nonuniform [3-5]. The hotspots and coldspots simultaneously created in different locations in the network can lead to heavy congestions in some links, while others are poorly utilized, negatively affecting the overall throughput and energy efficiency performance. However, today's interconnection networks based on electronic switches and optical fibers are inherently rigid, incapable of changing the network topology or link bandwidth, while adaptive routing techniques cannot adequately cope with the significant variations of traffic patterns. On the other hand, all-to-all interconnections are essential for many applications, including map-reduce based applications, parallel sorting applications, and deep neural network (DNN) applications. It would then be desirable to design a bandwidth-reconfigurable interconnection network that can support all-to-all connectivity and adapt its connectivity to the traffic demand of hotspots when necessary.

In the past few years, there has been significant attention to the application and development of optical switching fabric for bandwidth reconfiguration between computing nodes or Top-of-Rack switches [6,7]. At the physical layer, silicon photonics (SiPh) offers a variety of integrated devices with the capability of wavelength routing and space switching, thereby support dynamic configuration and reconfiguration in both spectral and spatial domains. Indeed, wavelength-and-space selective switching fabrics that can reconfigure the bandwidth between selected pair of input and output ports have been demonstrated with InGaAsP/InP arrayed waveguide grating routers (AWGR) + semiconductor optical amplifiers (SOAs) [8], SiPh echelle gratings + MEMS arrays [9], and SiPh multi-wavelength selective crossbar [10]. However, all these reported switching fabrics exhibit poor scalability and low energy efficiency due to either high insertion losses induced by power splitters [8], O(N^2) waveguide crossings in the worst-case path [9], or large number (O(N^3)) of required switching elements [9][10].

In [11,12], we proposed a bandwidth-reconfigurable all-to-all interconnection switch, 'Flexible Low-Latency Interconnect Optical Network Switch (Flex-LIONS),' enabled by combining an AWGR-based all-to-all interconnection, microring resonator

Manuscript received June 7, 2019. This work was supported in part by DoD contract H98230-16-C-0820 and NSF grant 1611560.

X. Xiao, R. Proietti, G. Liu, H. Lu, P. Fotouhi, S. Werner, Y. Zhang, and S. J. B. Yoo are with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: xxxiao@ucdavis.edu; rproietti@ucdavis.edu; genliu@ucdavis.edu; hlu@ucdavis.edu; pfotouhi@ucdavis.edu; swerner@ucdavis.edu; dyuzhang@ucdavis.edu; sbyoo@ucdavis.edu).

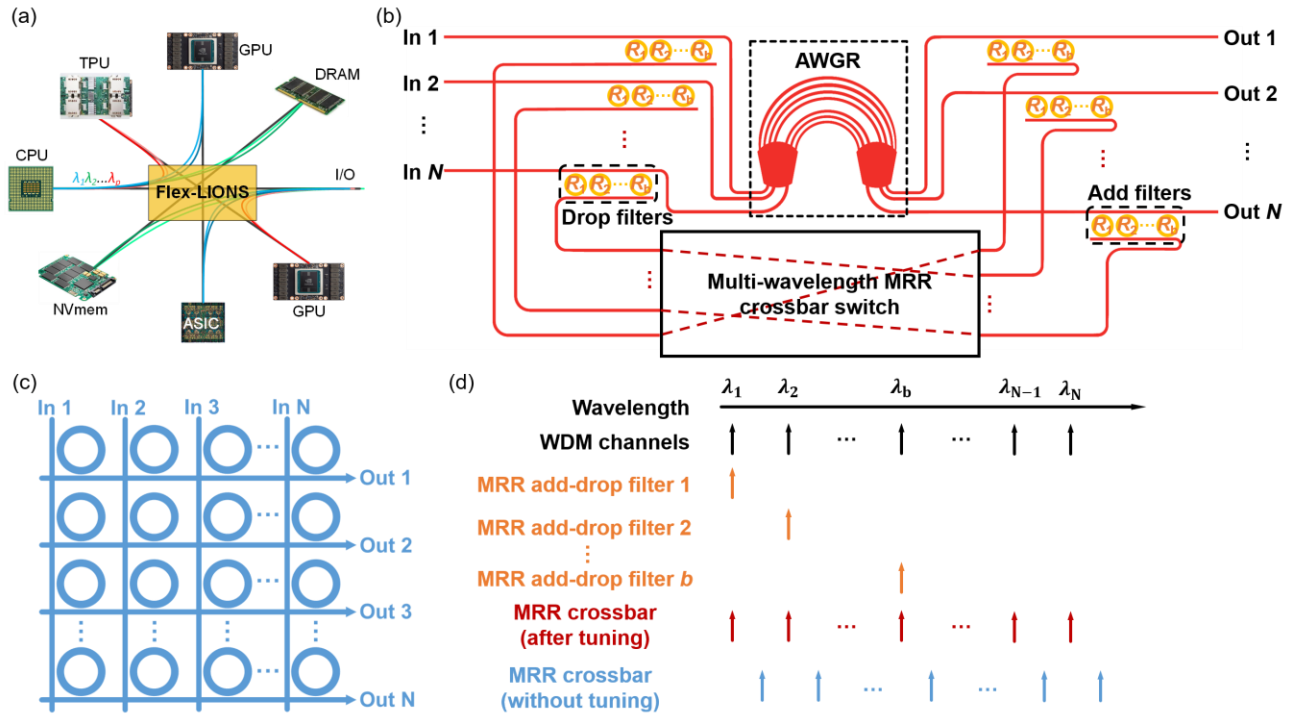

Fig. 1. (a) Modern HPC systems with heterogeneous processor and memory nodes. (b) *N*×*N* Flex-LIONS architecture with *N*×*N* AWGR, *b* MRR add-drop filters at each input and output ports, and *N*×*N* multi-wavelength MRR crossbar switch. (c) Schematic of multi-wavelength MRR crossbar switch. (d) Schematic of the wavelength relation between the WDM channels and the resonances of MRR add-drop filters and multi-wavelength MRR crossbar switch.

(MRR) add-drop filters, and multi-wavelength spatial switches. The multi-wavelength spatial switches which can be wide-band MEMS switches [13-15] or wide-band Benes MZS networks [16,17] are required to switch all the wavelength division multiplexing (WDM) signals simultaneously. In this paper, we choose to utilize a multi-wavelength MRR crossbar switch with MRR-based comb switch as building blocks [18]. The detailed principle will be discussed in Section II.

This paper reports the first experimental demonstration of SiPh Flex-LIONS that allows reconfiguration from 'fully all-to-all' to 'partially all-to-all' interconnect topology with bandwidth enhancements between selected port pairs. We designed and fabricated the SiPh Flex-LIONS chip on a multi-layer platform which monolithically integrates SiPh MRR add-drop filters, multi-wavelength MRR crossbar switch, and a low-crosstalk 8×8 200-GHz cyclic silicon nitride (SiN) AWGR. System measurement results show error-free one-to-eight multicast interconnection at 25 Gb/s, and bandwidth reconfiguration from 25 Gb/s to 100 Gb/s. Benchmarking simulation demonstrated a 1.33× reduction in packet latency and 1.5× improvements in energy efficiency when replacing the core layer switches of Fat-Tree topologies with Flex-LIONS.

The remainder of the paper is organized as follows. Section II introduces the architecture and principle of Flex-LIONS. Section III compares Flex-LIONS with the state-of-the-art bandwidth-reconfigurable switching fabrics. Section IV details the design, fabrication, packaging, and system testing of an 8×8 prototype SiPh Flex-LIONS. Section V shows the benchmarking simulations that evaluate the potential benefits of reconfiguration. Section VI discusses the scalability towards high-radix bandwidth-reconfigurable optical interconnections. Section VII concludes the paper.

II. Flex-LIONS Architecture and Principle

Fig. 1(b) illustrates the architecture of SiPh Flex-LIONS. It has an *N*-port cyclic AWGR at the core and includes *b* MRR add-drop filters at each AWGR input/output port. For uniform-random traffic, all MRR add-drop filters can be set off-resonance so that each input port provides *N* WDM signals to interconnect with all the *N* output ports according to the all-to-all wavelength routing property of the AWGR [19-22]. For non-uniform traffic or for resolving hot-spots, the MRR add-drop filters can be tuned in resonance to selected wavelengths channels so that those channels can be spatially switched by the multi-wavelength MRR crossbar switch shown in Fig 1(c). For instance, *b* of the *N* wavelengths from input port *i* can be dropped and then routed to the desired output port *j* by the *N*×*N* multi-wavelength MRR crossbar switch (strictly non-blocking). Here, the free spectral range (FSR) of the multi-wavelength MRR is designed to match with the AWGR channel spacing (i.e., WDM channel spacing) so that all the *b* wavelength can be simultaneously routed by tuning the desired multi-wavelength MRR in the crossbar (Fig. 1 (d)). In this way, the bandwidth between input port *i* and output port *j* is effectively increased by a factor of *b*.

III. State-of-the-Art Reconfigurable Switching Fabrics

There has been a significant amount of architectural and experimental works on optical switching fabrics for HPC and datacenter systems, including SOA-gate based optical switches [23], SiPh MZI-based optical switches [16,17], SiPh MRR-based optical switches [24,25], SiPh MEMS switches [13-15],

TABLE I
COMPARISON WITH THE STATE-OF-THE-ART BANDWIDTH-RECONFIGURABLE SWITCHING FABRICS

Architecture	Port Count	On-Chip Loss (dB)	Crosstalk (dB)	Footprint (mm^2)	Number of Switching Elements*	On-Chip Loss for $N\times N$ Scale (dB)	Reference
InP AWGRs +SOA gates	4×4	23.7	-12	4.2×3.6	$2N^2$	$(N\text{-}1)\times0.5$ $+\log_2 N\times7+8.5$	[8]
Si echelle gratings +MEMS arrays	8×8	16	Adjacent ch.: -17 Non-adjacent ch.: -30	9.7×6.7	N^3	$N\times0.18+N(N\text{-}1)\times0.034+12.6$	[9]
Multi-Wavelength Selective Crossbar	8×4	14	-32**	1.92×4.15	N^3	$(N\text{-}1)\times1.2+4.7$	[10]
Flex-LIONS with Conventional Multi-wavelength Crossbar	8×8	6	Adjacent ch.: -15 Non-adjacent ch.: -28	10×5.6	$3N^2$	$(2N+5)\times0.1$ $+(2N\text{-}2)\times0.09+3.5$	This work

* For $N\times N$ scale
** By using dual MRR switch

and SiPh AWGR-based switches [19]. However, most of these switching fabrics only have single wavelength connectivity between I/O ports and the bandwidth cannot be steered to match the interconnections with specific application and traffic patterns.

On the other hand, multi-wavelength wavelength-and-space selective bandwidth-reconfigurable switching fabrics exhibit better flexibility in interconnection patterns thanks to the ability to reconfigure the connectivity between I/O ports using any combination of the input wavelengths. Table I compares Flex-LIONS with various state-of-the-art approaches including InP AWGRs + SOA gates [8], SiPh echelle gratings + MEMS arrays [9], and SiPh multi-wavelength selective crossbar [10]. In particular, the comparison study takes into account worst-case on-chip loss, crosstalk, footprint, and the number of switching elements (SOA gates, MEMS switches, MRRs). Here we assume all the wavelength channels can be reconfigured (b=N for Flex-LIONS).

To evaluate the scalability to high radix, the number of switching elements and the on-chip loss as a function of the number of ports (N) are the primary metrics. It can be seen that [8] architecture has the problem of high on-chip insertion loss due to a large number of power splitters. Although the SOA gates can be used to compensate such high loss, the low energy efficiency prevents [8] architecture from scaling up to high radix. Reference [9] architecture suffers not only from the high number of switching elements (N^3) but also from high on-chip insertion loss since the number of waveguide crossings increases by ~N^2, while the number of waveguide crossings in Flex-LIONS increases only by ~N. Reference [10] architecture also has the issue of a high number of switching elements which limits the scalability. Taking N=64 for example, our proposed Flex-LIONS can save 21× in number of switching elements compared with [9] and [10] architectures, with 2.9×, 5.7×, and 2.8× lower on-chip loss compared with [8], [9], and [10] architectures, respectively.

IV. SILICON PHOTONIC 8×8 FLEX-LIONS

This section presents a detailed description of the design, fabrication, packaging, and system testing of the SiPh 8×8 Flex-LIONS (N=8, b=3).

A. Design

We designed our SiPh Flex-LIONS device on a multi-layer platform with silicon-dioxide cladding on silicon-on-insulator (SOI) wafers, as shown in Fig. 2(a) [26]. The bottom layer is the silicon (Si) waveguide layer, which contains the MRR add-drop filters and the multi-wavelength MRR crossbar switch. Ridge Si waveguides with 220-nm height and 500-nm width are used for low propagation loss. 600 nm above the Si waveguide layer is the SiN waveguide layer which contains the 200-GHz-spacing 8×8 cyclic AWGR (the detailed design procedures can be found in [27]). Ridge SiN waveguides with 200-nm height and 2-μm width are used for low propagation loss and relatively large bending radius. On top of the 2-μm-thick silicon dioxide cladding are the 100-nm-thick titanium (Ti) heater layer and 800-nm-thick contact metal layer for thermo-optical (TO) tuning of the MRRs.

Fig. 2(b) shows the schematic of the 8×8 SiPh Flex-LIONS layout. Edge coupler arrays with 127-μm-pitch are used to reduce the coupling loss from the fiber to the chip, as shown in Fig. 2(c). The designed radii of the two MRRs are 4.75 μm and 63 μm corresponding to FSRs of 19 nm and 1.6 nm, respectively. The gap between the bus waveguides and the MRRs are fabrication-calibrated to be 300 nm and 450 nm to minimize the insertion loss for dropping.

SiPh low-loss and low-crosstalk multimode interference (MMI) crossings are essential components to keep the overall insertion loss low [28]. Fig. 2(d) presents the physical dimensions of our crossing design. Fig. 2(e) shows the FDTD simulations of insertion loss with various taper length (L_T) and multimode region lengths (L_{MM}). With the optimal design (L_T=1.4 μm, L_{MM}=5.8 μm), the simulated insertion loss is 0.04 dB.

The SiN AWGR vertically interfaces with the Si layer through inverse-tapered evanescent couplers [26,29,30]. As shown in Fig. 2(f), the Si waveguide is tapered from 500 nm to 200 nm over a length of 200 μm, while the SiN waveguide is tapered from 200 nm to 2 μm. Fig. 2(g) shows the FDTD simulation of inverse-tapered evanescent coupler transmission with a varied interlayer gap. The optimal gap value is 600 nm, with an insertion loss of 0.1 dB.

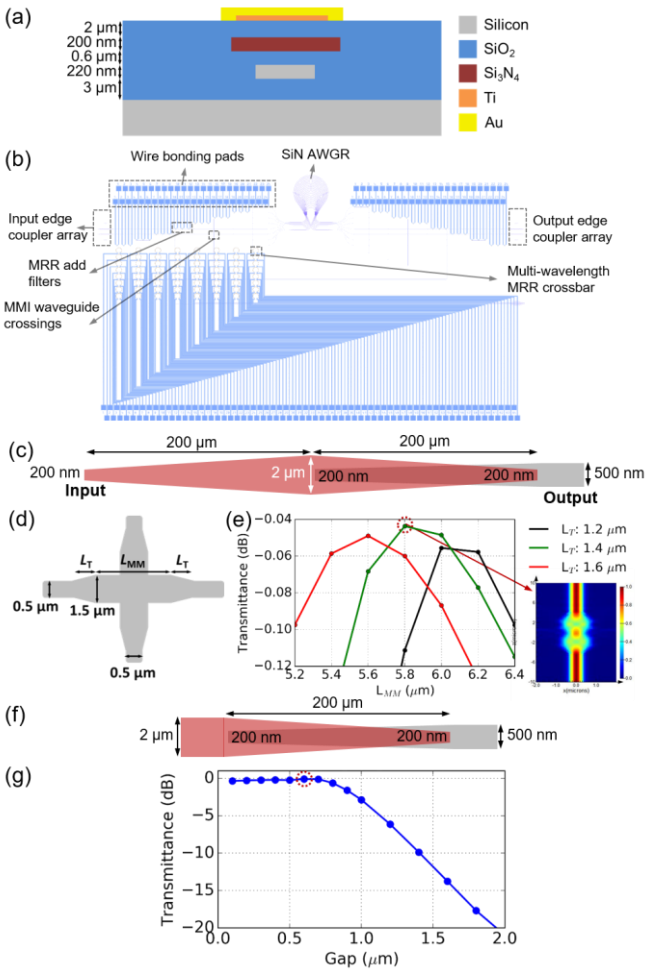

Fig. 2. (a) Cross section of the multi-layer platform. (b) Schematic of the 8×8 SiPh Flex-LIONS layout. (c) Schematic of edge coupler design. (d) Schematic of MMI waveguide crossing design. (e) Simulated insertion loss of MMI waveguide crossing with various taper and multimode region lengths. (f) Schematic of Si-to-SiN evanescent coupler design. (g) Simulated insertion loss of evanescent coupler with various gap values.

B. *Fabrication*

Utilizing micro and nanoscale fabrication facilities at the University of California at Davis and Berkeley, we fabricated the device on a 220-nm SOI wafer with 3-μm-thick buried oxide, as shown in Fig. 3(a). The silicon layer was defined by deep-UV projection lithography and inductively coupled plasma (ICP) etching. Then a 1000-nm-thick low-temperature oxide (LTO) was deposited by low-pressure chemical vapor deposition (LPCVD) and then planarized to 800 nm by chemical mechanical planarization (CMP). Following the deposition of a 200-nm-thick SiN layer, the AWGR was patterned by deep-UV lithography and ICP etching, followed by a 2-μm-thick LTO deposition and planarization. A 100-nm-thick Ti was then deposited on top of the cladding and along the MRR to act as a heater for TO tuning. Finally, a 20-nm-Ti and 800-nm-Au were deposited to form the contact metal layer. Fig. 3(b-d) show the microscope images of the fabricated chip, MRR add-drop filter, and the multi-wavelength MRR switch.

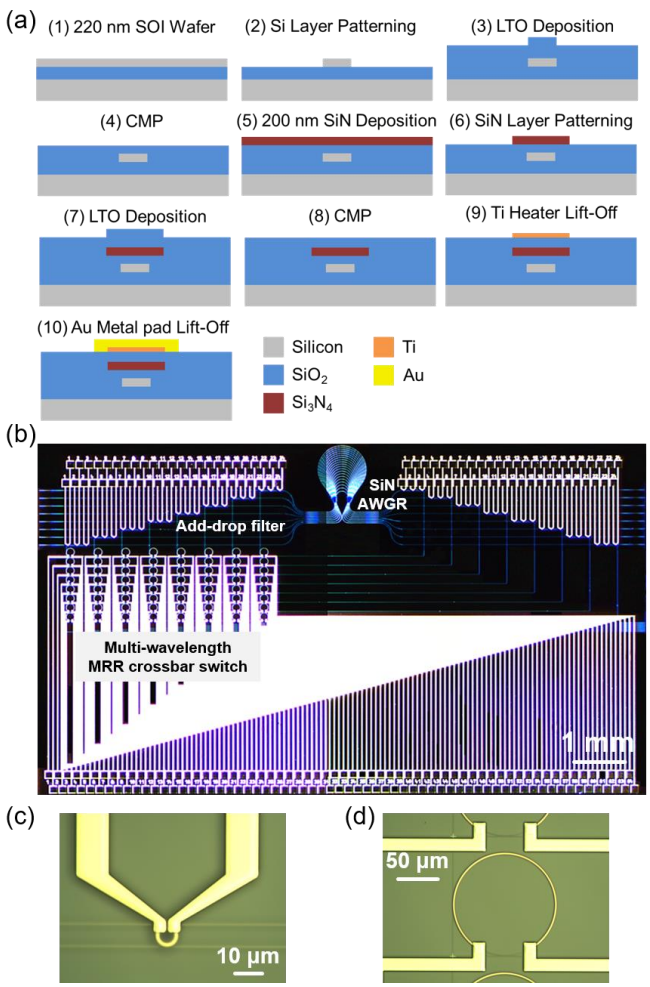

Fig. 3. (a) Fabrication flow charts for the 8×8 SiPh Flex-LIONS. (b) Microscope image of the fabricated 8×8 SiPh Flex-LIONS (N=8, b=3) chip. (c) Microscope image of MRR add-drop filter. (d) Microscope image of multi-wavelength MRR switch.

C. *Single component characterizations*

Figure 4(a) shows the transmission spectra of the 8×8 SiN AWGR from input port 4 measured by an optical vector network analyzer (OVNA) system. The AWGR is cyclic with an FSR of 12.8 nm, channel spacing of 1.6 nm (200 GHz) and full-width-at-half-maximum (FWHM) of 1.2 nm. The adjacent channel crosstalk is < -15 dB, the non-adjacent channel crosstalk is < -28 dB, and the insertion loss is < 5.1 dB. Fig. 4(b) shows the transmission spectra of the through and drop ports of multi-wavelength MRR switch with different TO tuning power values. The insertion loss for the drop port at the resonance is 1 dB, and the corresponding FWHM is 0.24 nm. All the spectra are normalized to the reference waveguide. The TO tuning efficiencies of multi-wavelength MRR switch and MRR add-drop filter are 0.03 nm/mW and 0.15 nm/mW, respectively, as shown in Fig. 4(c) and (d). Higher TO tuning efficiency can be achieved by using waveguide microheaters [17], and faster reconfiguration can be obtained by electro-optical (EO) tuning [16].

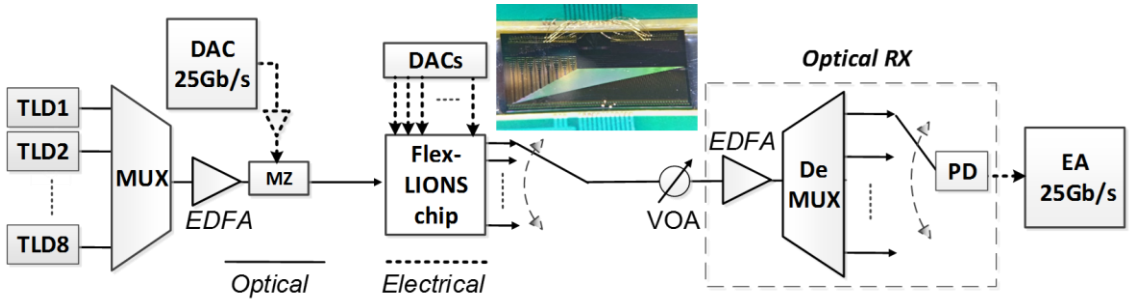

Fig. 5. Experimental setup. TLD: tunable laser diode; EDFA: erbium-doped fiber amplifier; MZ: Mach Zehnder; DAC: digital to analog converter; DUT: device under test; VOA: variable optical attenuator; PD: photodetector; EA: error analyzer.

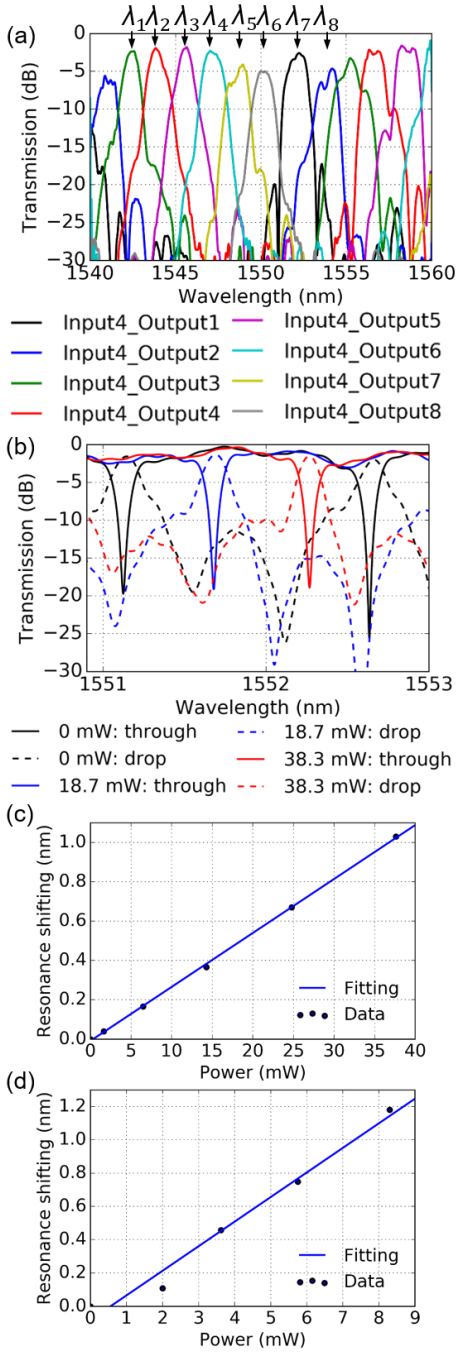

Fig. 4. Transmission spectra of: (a) 8×8 AWGR from input port 4; (b) through and drop ports of multi-wavelength MRR switch with different TO tuning power. Thermal tuning efficiency of: (c) multi-wavelength MRR switch; (d) MRR add-drop filter.

D. Experimental Demonstration of Optical Reconfiguration

Fig. 5 shows the experimental setup we used to demonstrate the optical reconfiguration capabilities of the SiPh Flex-LIONS. The light sources are eight tunable laser diodes (TLDs) which provide the 200-GHz-spacing WDM grid of the Flex-LIONS. All the WDM wavelengths are multiplexed, amplified by a booster EDFA, and modulated by a Mach-Zehnder (MZ) modulator at 25 Gb/s. The driven signals are 2^{11}-1 PRBS signals generated by a high-speed DAC. The modulated WDM signals are coupled in/out the Flex-LIONS chip using lensed fibers. The output signal from the Flex-LIONS chip is then received by an optically pre-amplified receiver (RX). A real-time error analyzer (EA) performs BER measurements as a function of the RX input power, which is measured by the built-in optical power monitor of the VOA. The Flex-LIONS chip was wire-bonded on a PCB and driven by a multi-channel DAC controller. The driving signals were used to tune the MRR add-drop filters as well as the multi-wavelength MRR crossbar switch, responsible for switching all the wavelength dropped to the desired output port.

Before reconfiguration, the device implements all-to-all connectivity, and the 8-channel WDM signal applied at the Flex-LIONS input 4 is demultiplexed to the eight Flex-LIONS output ports (one wavelength per port) according to the AWGR routing table. Fig. 6(a) shows the bandwidth available between input 4 and outputs 5, which is single-channel (λ_3) bandwidth of the AWGR. Fig. 6(b) and (c) shows the measured BER curves and eye diagrams of the signals at the eight different output ports, demonstrating 25 Gb/s error-free operation with limited power penalty compared with the back-to-back BER curve. The total system capacity is 25 Gb/s × 8 × 8 = 1.6 Tb/s.

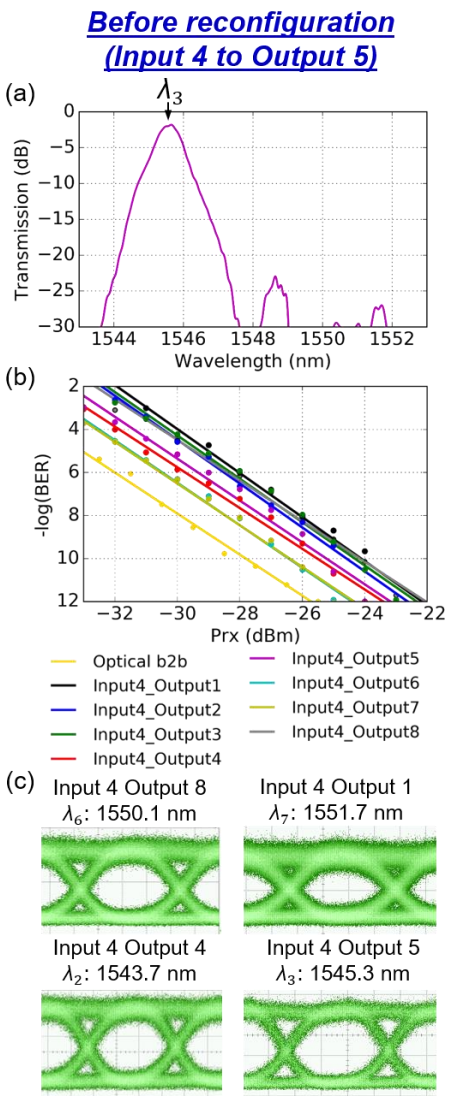

Fig. 6. (a) Transmission spectrum from input port 4 to output port 5 before reconfiguration. (b) BER curves of input port 4 to different output ports. (c) Eye diagrams of input port 4 to output port 8, 1, 4 and 5 before reconfiguration.

After reconfiguration of the Flex-LIONS, four wavelengths from input port 4 are routed to output port 5. One wavelength is going through the AWGR while the other three wavelengths are dropped, switched by the spatial switch and added to output port 5, effectively increasing the bandwidth between port 4 and port 5 by 4× (from 25 Gb/s to 100 Gb/s). Fig. 7(a) shows that there are now four frequency slots available between input 4 and outputs 5. One of the four frequency slots is the passband of AWGR (λ_3) while the other three (λ_5, λ_6, λ_7) are from the cascaded MRR add-drop filters and multi-wavelength MRR crossbar switch. The maximum baud rate per wavelength channel is mainly limited by the compound cavity effect of cascaded MRRs for the signals dropped at AWGRs' inputs and going through the multi-wavelength MRR crossbar. To reduce the cascaded MRR filtering effect it could be possible to use wide-band Benes MZS networks as the multi-wavelength spatial switch to reduce the number of cascaded MRRs on the path of the reconfigured channels from three to two. Another method would be to employ flat-passband coupled MRRs [31] at the expense of incorporating more complicated tuning methods. While all the signals reach error-free condition (Fig. 7(b) and (c)), the power penalty for one of the signal is significantly higher. We attribute this larger penalty to the frequency deviation between the WDM wavelengths and compound MRR cavity resonance.

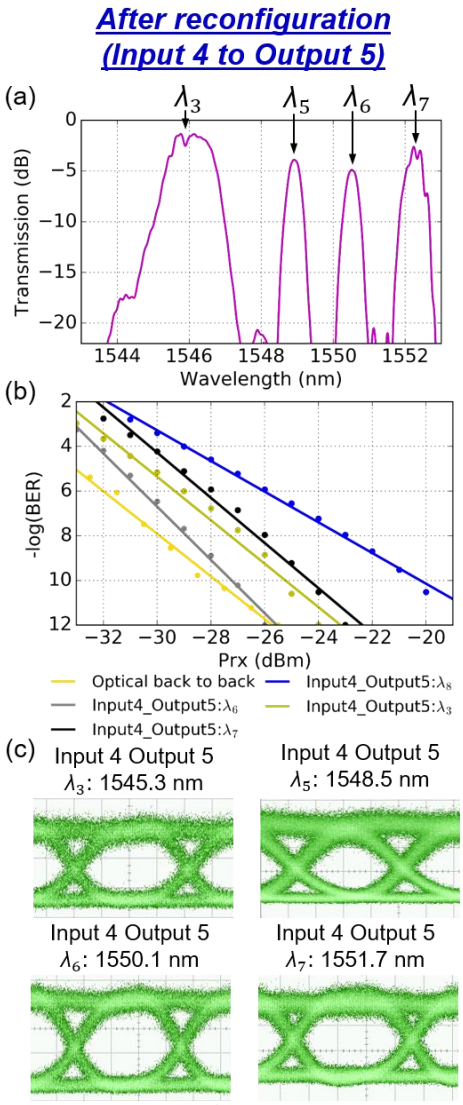

Fig. 7. (a) Transmission spectrum from input port 4 to output port 5 after reconfiguration. (b) BER curves of input port 4 to output port 5. (c) Eye diagrams of input port 4 to output port 5 using λ_3, λ_5, λ_6, and λ_7 after reconfiguration.

V. Performance Study of Flex-LIONS

The potential benefits of reconfiguration are evaluated by using a 16-port Flex-LIONS to replace the core switches of a Fat-Tree architecture with sixteen 32-port Top-of-Rack switches and a total of 256 nodes. To compare Flex-LIONS with the most aggressive baseline, we modelled the power consumption and latency of the switches based on state-of-the-art commercially-available data center switches [32], which consume 95W power and offer a 100ns switch traversal latency. For both the Fat-Tree and Flex-LIONS architecture, we assume 14-nm technology photonic transceiver models which are scaled down from a 65-nm experimental demonstration [33,34] using SPICE models. The optical loss of Flex-LIONS with

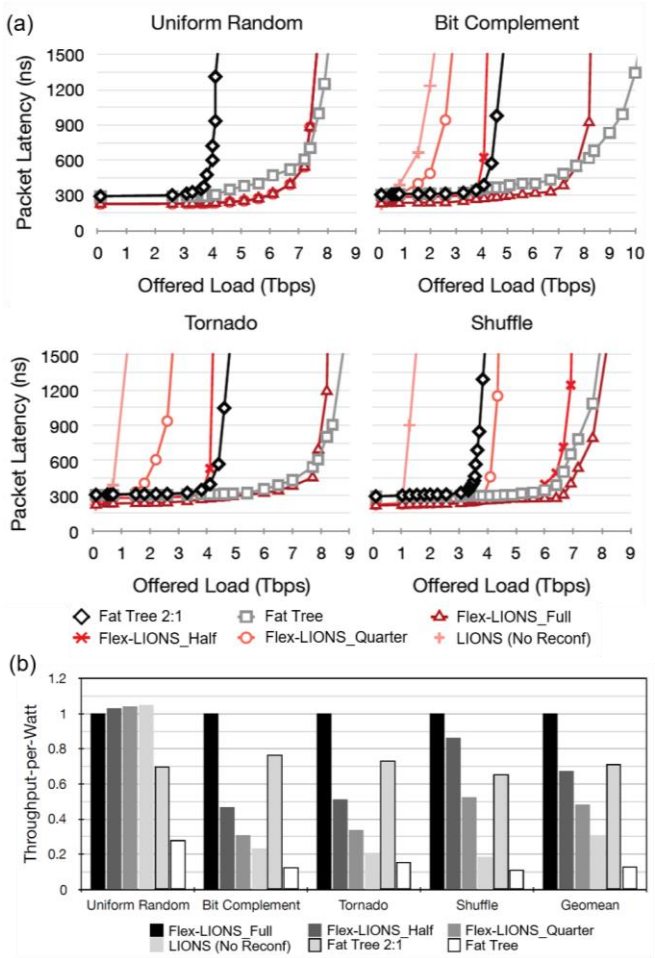

Fig. 8. (a) Average packet latency (ns) versus offered load (Tbps) for different synthetic traffic patterns. (b) Throughput-per-Watt for different network designs normalized to Flex-LIONS_full.

different degrees of reconfigurability is considered so that the power consumption of the photonic transceivers is accordingly calculated. Figure 8 shows the performance results for the different synthetic traffic patterns for different offered network load. LIONS without network reconfiguration performs poorly for each traffic pattern besides uniform random where the traffic is evenly distributed across all links. For the different Flex-LIONS reconfiguration capabilities, we observe that the more flexible the band-width assignment is, the higher the total accepted traffic gets, which indicates that finely adjusting the bandwidth to links improves the performance. On the other hand, compared with Fat-Tree, which offers the same bisection bandwidth as Flex-LIONS, Flex-LIONS reduces the average packet latency before network saturation by 1.33× on average. Figure 8(b) plots throughput-per-Watt (TPW) for different network architectures normalized to Flex-LIONS with full flexibility (Flex-LIONS_full). Though supporting much less maximum throughput, Fat-Tree 2:1 is more power efficient than a full Fat-Tree as it saves a lot of energy through implementing fewer switches and transceivers. As a result, Fat-Tree 2:1 exhibits 0.7× of Flex-LIONS's TPW while full Fat-Tree only shows 0.15× of Flex-LIONS's TPW. It is also found that unless traffic is perfectly uniformly distributed which is very uncommon in both HPC and data center networks, the energy efficiency of Flex-LIONS with full flexibility in reconfiguration is always the best among all architectures.

While the proposed device enables reconfiguration at the physical layer with reconfiguration times of ~10 μs, specific reconfiguration policies and algorithms at the network and application layers are needed. The most straightforward approach would be to perform reconfiguration in between different applications and workloads [35] running sequentially on an HPC node or cluster, while a more complex scenario would include performing reconfiguration on the fly (e.g. in between phases of a specific application or, in a datacenter scenario, based on network traffic congestion). In the latter case, more advanced control strategies, including machine-learning aided traffic prediction [36], and fast and scalable control plane solutions are needed. However, these aspects are beyond the scope of this paper, which focuses on the first experimental demonstration of the Flex-LIONS device.

VI. Scalability of Flex-LIONS

Although our proposed Flex-LIONS architecture exhibits the least number of switching elements comparing with the state-of-the-art architectures, the scalability to larger port count (up to 1024×1024 for example) is still limited by the power penalty induced by AWGR crosstalk. In this section, we calculate and experimentally measured the impact of intra-band crosstalk of AWGR on the scalability of Flex-LIONS, and discuss the use of Thin-CLOS Flex-LIONS architecture for port count scaling.

A. AWGR crosstalk power penalty

Several factors limit the scalability of AWGRs, including insertion loss, loss non-uniformity, and channel spacing. Besides, the intra-band crosstalk is primary main impairment mechanism that affects the scalability of AWGRs since the signal-crosstalk beat noise cannot be removed by filters or de-multiplexers after the output ports [37].

Fig. 9(a) shows the crosstalk power penalty of AWGR versus different crosstalk values with BER of 10^{-12}. Here we assume optimized decision-threshold setting and aligned polarization for the worst case [38]. The results show that with -35 dB crosstalk the power penalty for 4-, 8-, 16-, and 32-node interconnects are 0.21, 0.50, 1.15, and 2.84 dB, respectively. For more clarity, the numbers of scalable nodes versus varied crosstalk values for different power penalty constraints are plotted in Fig. 9(b). With 1-, 3-, and 6-dB power penalty constraint, the AWGR can scale to 32 nodes when the crosstalk is less than -38.7, -34.9, -33.1 dB, respectively.

The power penalty induced by AWGR intra-band crosstalk is experimentally measured using the 8×8 SiPh Flex-LIONS chip. The Flex-LIONS chip is aligned and packaged with two 16-channel 127-μm-pitch polarization-maintaining (PM) fiber arrays on both the input and output sides. The modulated 25 Gb/s/λ WDM signal is firstly split by a 1×8 splitter. Then the eight signals are decorrelated by single-mode fiber catch cables with different lengths. Before input into all the eight input waveguides of the chip, the polarization of each signal is aligned by a polarization controller. Note that, in our future work, SiPh polarization splitters and rotators could be included

to transform the polarization of the input signal from the single-mode fiber into fundamental TE mode [39] (this is important as Datacom systems do not make use of PM fibers). The blue curve in Fig. 9(c) shows the BER measurements of the signal going from input 4 to output 5 at λ_3 for the worst-case crosstalk scenario (with all the input signals at λ_3 are aligned in polarization). Comparing with no crosstalk signal added (black curve), the measured power penalty is 3.9 dB at BER=10^{-12} which is slightly lower than the theoretically calculated value most likely due to the polarizations of the input signals not being perfectly aligned. The insets of Figure 9(c) show the eye diagrams of the transmitted signal with and without crosstalk signals added.

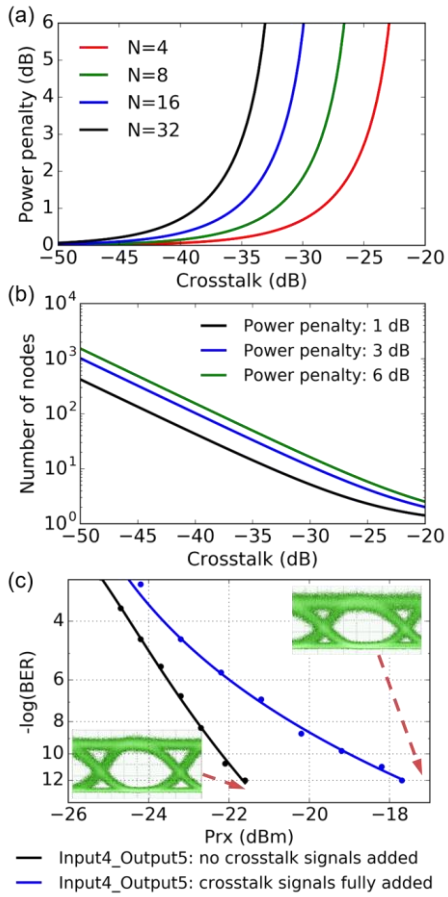

Fig. 9. (a) Power penalty versus varied crosstalk for AWGR. (b) Number of scalable nodes versus varied crosstalk with power penalty constraint of 1, 3, 6 dB. (c) End-to-end link experiment with eight input signals at the same wavelength and aligned polarization for the worst-case crosstalk scenario.

SiN AWGRs are superior to Si AWGR in mitigating crosstalk since the lower index contrast makes SiN/SiO_2 waveguides less sensitive to fabrication imperfections. As a result, SiN AWGRs have smaller phase errors and consequently lower crosstalk. In our previous work [40], 16×16 and 32×32 SiN AWGRs have been fabricated and characterized with crosstalk value of -10 dB. Alternatively, another paper reports SiN AWG with adjacent and non-adjacent crosstalk value as -39 dB and -33.5 dB by optimizing the deposition condition of cladding oxide and cross-section structure [41], which indicates the possibility to scale AWGR beyond 32×32.

B. *Thin-CLOS Flex-LIONS*

To further improve the scalability, Thin-CLOS Flex-LIONS is a promising architecture that can enable N×N bandwidth-reconfigurable switching fabric using M^2 number of W×W Flex-LIONS instead of a single N×N Flex-LIONS as shown in Fig. 10(a) [42-44]. In this case, the number of intra-band crosstalk components is decreased from N-1 to W-1 so that the crosstalk power penalty can be significantly reduced. Other than that, smaller AWGRs also means lower insertion losses, loss non-uniformity, and larger channel spacing in a fixed spectral range. Fig. 10 (b) shows the schematic of the layout of a 16×16 SiPh Thin-CLOS Flex-LIONS with four 8×8 Flex-LIONS (N=16, M=2, W=8). The overall size is 12.0 mm × 12.3 mm. The waveguide crossings can be addressed by an additional SiN layer, and the chip can be flip-chip bonded on an optical interposer for the electrical fan-out [40,45]. We believe such an approach paves the way to realizing large-scale Flex-LIONS with a limited number of wavelengths (e.g., W=64).

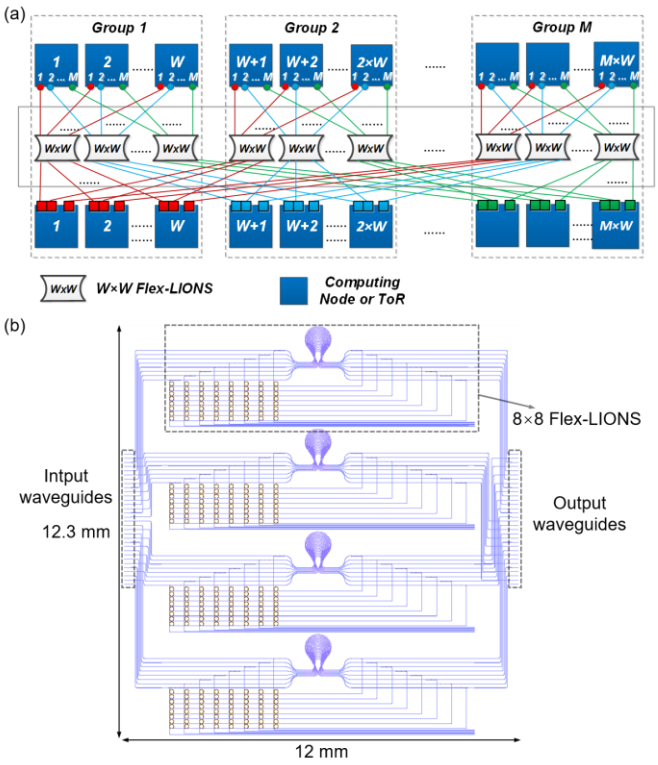

Fig. 10. (a) Schematic of N×N Thin-CLOS Flex-LIONS architecture (N=M×W). (b) Layout of 16×16 Thin-CLOS Flex-LIONS with four 8×8 Flex-LIONS (N=16, M=2, W=8).

VII. Conclusion

We experimentally demonstrated the first SiPh AWGR-based 8-port bandwidth-reconfigurable switching fabric, with 21× less switching elements and 2.9× lower on-chip loss when compared with the state-of-the-art switching fabrics at a scale of 64 ports. Successful Flex-LIONS design, fabrication, and system testing show error-free operation of bandwidth reconfiguration from 25 Gb/s to 100 Gb/s between selected pairs of input and output ports. Benchmarking simulation results show a reduction of 1.33× for average packet latency and

improvements of 1.5× for energy efficiency when compared with Fat-Tree topologies. In addition, Thin-CLOS Flex-LIONS architecture is proposed to scale beyond the port count limitation imposed by coherent crosstalk in AWGRs.

ACKNOWLEDGMENT

Fabrication of the devices utilized the facilities at the Marvell Nanofabrication Laboratory (Berkeley, CA) and the Center for Nano-Micro Manufacturing (Davis, CA). The authors acknowledge technical support and assistance from Yi-Chun Ling, Guangyao Liu, and Rijuta Ravichandran from the Next Generation Network System (NGNS) Laboratory at UC Davis.

REFERENCES

[1] S. Mittal and J. S. Vetter, "A Survey of CPU-GPU Heterogeneous Computing Techniques," *ACM Comput. Surv.*, vol. 47, no. 4, pp. 69:1--69:35, Jul. 2015.

[2] M. J. Schulte *et al.*, "Achieving Exascale Capabilities through Heterogeneous Computing," *IEEE Micro*, vol. 35, no. 4, pp. 26–36, 2015.

[3] T. Benson, A. Akella, and D. A. Maltz, "Network Traffic Characteristics of Data Centers in the Wild," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, 2010, pp. 267–280.

[4] A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren, "Inside the Social Network's (Datacenter) Network," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 123–137, Aug. 2015.

[5] Q. Zhang, V. Liu, H. Zeng, and A. Krishnamurthy, "High-resolution Measurement of Data Center Microbursts," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 78–85.

[6] Z. Cao, R. Proietti, M. Clements, and S. J. B. Yoo, "Experimental Demonstration of Flexible Bandwidth Optical Data Center Core Network With All-to-All Interconnectivity," *J. Light. Technol.*, vol. 33, no. 8, pp. 1578–1585, 2015.

[7] K. Wen *et al.*, "Flexfly: Enabling a Reconfigurable Dragonfly through Silicon Photonics," in *SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2016, pp. 166–177.

[8] R. Stabile, A. Rohit, and K. A. Williams, "Monolithically Integrated 8 × 8 Space and Wavelength Selective Cross-Connect," *J. Light. Technol.*, vol. 32, no. 2, pp. 201–207, 2014.

[9] T. J. Seok *et al.*, "MEMS-Actuated 8×8 Silicon Photonic Wavelength-Selective Switches with 8 Wavelength Channels," in *2018 Conference on Lasers and Electro-Optics (CLEO)*, 2018, pp. 1–2.

[10] A. S. P. Khope *et al.*, "Multi-wavelength selective crossbar switch," *Opt. Express*, vol. 27, no. 4, pp. 5203–5216, Feb. 2019.

[11] X. Xiao, R. Proietti, K. Zhang, and S. J. Ben Yoo, "Experimental Demonstration of Flex-LIONS for Reconfigurable All-to-All Optical Interconnects," in *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.

[12] R. Proietti, G. Liu, X. Xiao, S. Werner, P. Fotouhi, and S. J. B. Yoo, "FlexLION: A Reconfigurable All-to-All Optical Interconnect Fabric with Bandwidth Steering," in *Conference on Lasers and Electro-Optics*, 2019, p. SM3G.2.

[13] S. Han, T. J. Seok, N. Quack, B.-W. Yoo, and M. C. Wu, "Large-scale silicon photonic switches with movable directional couplers," *Optica*, vol. 2, no. 4, pp. 370–375, Apr. 2015.

[14] T. J. Seok, N. Quack, S. Han, R. S. Muller, and M. C. Wu, "Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers," *Optica*, vol. 3, no. 1, pp. 64–70, Jan. 2016.

[15] S. Han, T. J. Seok, K. Yu, N. Quack, R. S. Muller, and M. C. Wu, "Large-Scale Polarization-Insensitive Silicon Photonic MEMS Switches," *J. Light. Technol.*, vol. 36, no. 10, pp. 1824–1830, 2018.

[16] L. Lu *et al.*, "16 × 16 non-blocking silicon optical switch based on electro-optic Mach-Zehnder interferometers," *Opt. Express*, vol. 24, no. 9, pp. 9295–9307, May 2016.

[17] S. Zhao, L. Lu, L. Zhou, D. Li, Z. Guo, and J. Chen, "16 × 16 silicon Mach–Zehnder interferometer switch actuated with waveguide microheaters," *Photon. Res.*, vol. 4, no. 5, pp. 202–207, Oct. 2016.

[18] B. G. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, "All-Optical Comb Switch for Multiwavelength Message Routing in Silicon Photonic Networks," *IEEE Photonics Technol. Lett.*, vol. 20, no. 10, pp. 767–769, 2008.

[19] R. Yu *et al.*, "A scalable silicon photonic chip-scale optical switch for high performance computing systems," *Opt. Express*, vol. 21, no. 26, pp. 32655–32667, Dec. 2013.

[20] R. Proietti, Z. Cao, C. J. Nitta, Y. Li, and S. J. B. Yoo, "A Scalable, Low-Latency, High-Throughput, Optical Interconnect Architecture Based on Arrayed Waveguide Grating Routers," *J. Light. Technol.*, vol. 33, no. 4, pp. 911–920, 2015.

[21] P. Grani, R. Proietti, S. Cheung, and S. J. Ben Yoo, "Flat-Topology High-Throughput Compute Node With AWGR-Based Optical-Interconnects," *J. Light. Technol.*, vol. 34, no. 12, pp. 2959–2968, 2016.

[22] P. Grani, G. Liu, R. Proietti, and S. J. Ben Yoo, "Bit-parallel all-to-all and flexible AWGR-based optical interconnects," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.

[23] A. Wonfor, H. Wang, R. Penty, and I. White, "Large Port Count High-Speed Optical Switch Fabric for Use Within Datacenters [Invited]," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 3, no. 8, pp. A32–A39, 2011.

[24] Q. Cheng *et al.*, "Ultralow-crosstalk, strictly non-blocking microring-based optical switch," *Photon. Res.*, vol. 7, no. 2, pp. 155–161, Feb. 2019.

[25] Q. Cheng, M. Bahadori, Y. Hung, Y. Huang, N. Abrams, and K. Bergman, "Scalable Microring-Based Silicon Clos Switch Fabric With Switch-and-Select Stages," *IEEE J. Sel. Top. Quantum Electron.*, vol. 25, no. 5, pp. 1–11, 2019.

[26] K. Shang, S. Pathak, B. Guan, G. Liu, and S. J. B. Yoo, "Low-loss compact multilayer silicon nitride platform for 3D photonic integrated circuits," *Opt. Express*, vol. 23, no. 16, pp. 21334–21342, Aug. 2015.

[27] S. Cheung, T. Su, K. Okamoto, and S. J. B. Yoo, "Ultra-Compact Silicon Photonic 512 × 512 25 GHz Arrayed Waveguide Grating Router," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 4, pp. 310–316, 2014

[28] L. B. Soldano and E. C. M. Pennings, "Optical multi-mode interference devices based on self-imaging: principles and applications," *J. Light. Technol.*, vol. 13, no. 4, pp. 615–627, 1995.

[29] W. D. Sacher, Y. Huang, G. Lo, and J. K. S. Poon, "Multilayer Silicon Nitride-on-Silicon Integrated Photonic Platforms and Devices," *J. Light. Technol.*, vol. 33, no. 4, pp. 901–910, 2015.

[30] K. Shang *et al.*, "Silicon nitride tri-layer vertical Y-junction and 3D couplers with arbitrary splitting ratio for photonic integrated circuits," *Opt. Express*, vol. 25, no. 9, pp. 10474–10483, May 2017.

[31] T. Barwicz et al., "Microring-resonator-based add-drop filters in SiN: fabrication and analysis," *Opt. Express*, vol. 12, no. 7, pp. 1437–1442, 2004.

[32] Intel, "Intel Omni-Path Edge Switch Products," https://www.intel.com/content/www/us/en/products/network-io/high-performance-fabrics/omni-path-edge-switch-100-series.html, 2019.

[33] H. Li et al., "A 25 Gb/s, 4.4 V-Swing, AC-Coupled Ring Modulator-Based WDM Transmitter with Wavelength Stabilization in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 50, no. 12, pp. 3145–3159, 2015.

[34] K. Yu et al., "A 25 Gb/s Hybrid-Integrated Silicon Photonic Source-Synchronous Receiver With Microring Wavelength Stabilization," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2129–2141, 2016."

[35] G. Yuan *et al.*, "ARON: Application-Driven Reconfigurable Optical Networking for HPC Data Centers," in *ECOC 2016; 42nd European Conference on Optical Communication*, 2016, pp. 1–3.

[36] R. McKenna, S. Herbein, A. Moody, T. Gamblin, and M. Taufer, "Machine Learning Predictions of Runtime and IO Traffic on High-End Clusters," in *2016 IEEE International Conference on Cluster Computing (CLUSTER)*, 2016, pp. 255–258.

[37] H. Takahashi, K. Oda, and H. Toba, "Impact of crosstalk in an arrayed-waveguide multiplexer on N/spl times/N optical interconnection," *J. Light. Technol.*, vol. 14, no. 6, pp. 1097–1105, 1996.

[38] X. Xiao, R. Proietti, and S. J. Ben Yoo, "Scalability of microring-based crossbar for all-to-all optical interconnects," in *2017 IEEE Optical Interconnects Conference (OI)*, 2017, pp. 39–40.

[39] D. Dai and H. Wu, "Realization of a compact polarization splitter-rotator on silicon," *Opt. Lett.*, vol. 41, no. 10, pp. 2346–2349, May 2016.

[40] Y. Zhang *et al.*, "Foundry-Enabled Scalable All-to-All Optical Interconnects Using Silicon Nitride Arrayed Waveguide Router Interposers and Silicon Photonic Transceivers," *IEEE J. Sel. Top. Quantum Electron.*, vol. 25, no. 5, pp. 1–9, 2019.

[41] M. Piels, J. F. Bauters, M. L. Davenport, M. J. R. Heck, and J. E. Bowers, "Low-Loss Silicon Nitride AWG Demultiplexer

Heterogeneously Integrated With Hybrid III–V/Silicon Photodetectors," *J. Light. Technol.*, vol. 32, no. 4, pp. 817–823, 2014.
[42] Y. Yin *et al.*, "AWGR-based all-to-all optical interconnects using limited number of wavelengths," in *2013 Optical Interconnects Conference*, 2013, pp. 47–48.
[43] X. Xiao *et al.*, "Experimental Demonstration of 64-Port Thin-CLOS Architecture for All-to-All Optical Interconnects," in *2018 Conference on Lasers and Electro-Optics (CLEO)*, 2018, pp. 1–2.
[44] R. Proietti *et al.*, "Experimental demonstration of a 64-port wavelength routing thin-CLOS system for data center switching architectures," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 10, no. 7, pp. 49–57, 2018.
[45] X. Xiao, Y. Zhang, R. Proietti, and S. J. B. Yoo, "Scalable AWGR-based All-to-All Optical Interconnects with 2.5D/3D Integrated Optical Interposers," in *2018 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, 2018, pp. 161–162.

Xian Xiao received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2012 and 2015. He has been working toward the Ph.D. degree in electrical and computer engineering at the University of California, Davis, CA, USA, since 2015.

He was a research intern with Nokia Bell Labs in the summer of 2016 and 2017, with Lawrence Berkeley National Laboratory from 2017 to 2018, and with Hewlett-Packard Labs in the summer of 2018.

His current research interest includes silicon photonics, optical interconnects, 2.5D/3D photonic integration, neuromorphic computing.

Roberto Proietti received the M.S. degree in telecommunications engineering from the University of Pisa, Pisa, Italy, in 2004, and the Ph.D. degree in electrical engineering from Scuola Superiore Sant Anna, Pisa, in 2009. He is a Project Scientist with the Next Generation Networking Systems Laboratory, University of California, Davis. His research interests include optical switching technologies and architectures for supercomputing and data center applications, high-spectral-efficiency coherent transmission systems, and elastic optical networking.

Gengchen Liu was born in Beijing, China, in 1993. He received the B.S. degree in Electrical Engineering from Central Michigan University, Mount Pleasant, MI, USA, in 2015. He is currently pursuing the Ph.D. degree in Electrical and Computer Engineering at University of California, Davis, CA, USA.

His research interest focuses on high-speed optical transceiver design, digital signal processing, and machine learning.

Hongbo Lu was born in Shanghai, China, in 1989. He received the B.S. degree in Communication Engineering from Fudan University, Shanghai, China, in 2011, and the M.S. degree in Electrical Engineering from Tokyo University, Tokyo, Japan, in 2014. He is currently pursuing the Ph.D. degree in Electrical and Computer Engineering at University of California, Davis, CA, USA.

His research interest focuses on coherent optical communication and optical-wireless integrations.

Pouya Fotouhi received the B.Sc. in Electrical Engineering from the Isfahan University, and M.Sc. degree in Computer Engineering from University of Delaware in 2017. He is currently pursuing his Ph.D. degree in Computer Engineering from the Department of Electrical and Computer Engineering at University of California, Davis, CA, USA.

His research interest includes optical interconnects, flat memory systems, and heterogenous computing.

Sebastian Werner is a postdoctoral research scientist at the Next Generation Networking Systems Laboratory at the University of California, Davis. He received the B.S. and M.S. degree in Computer Engineering from the Darmstadt University of Technology, Germany, and his Ph.D. in Computer Science from the University of Manchester, UK. His current research interests include computer architecture, memory systems, optical interconnects, and networks-on-chip.

Yu Zhang received his B. S. degree of Optoelectronics from Huazhong University of Science and Technology, Wuhan, China, in 2010 and the Ph.D. degree in Electronic and Computer Engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2016. He is now an assistant project scientist in electrical and computer engineering at the University of California, Davis, CA.

His current research interest focuses on 3D integrated silicon photonic optical phased array, hybrid silicon laser, amplifier and phase modulators and silicon photonic integrated circuit enabled all-to-all interconnection for scalable high-performance computing systems.

S. J. Ben Yoo [S'82, M'84, SM'97, F'07] currently serves as a distinguished professor of electrical engineering at UC Davis. His research at UC Davis includes 2D/3D photonic integration for future computing, communication, imaging, and navigation systems, micro/nano systems integration, and the future Internet. Prior to joining UC Davis in 1999, he was a Senior Research Scientist at Bellcore, leading technical efforts in integrated photonics, optical networking, and systems integration. His research activities at Bellcore included the next-generation Internet, reconfigurable multiwavelength optical networks (MONET), wavelength interchanging cross connects, wavelength converters, vertical-cavity lasers, and high-speed modulators. He led the MONET testbed experimentation efforts, and participated in ATD/MONET systems integration and a number of standardization activities. Prior to joining Bellcore in 1991, he conducted research on nonlinear optical processes in quantum wells, a four-wave-mixing study of relaxation mechanisms in dye molecules, and ultrafast diffusion-driven photodetectors at Stanford University. Prof. Yoo received his B.S. degree in electrical engineering with distinction, his M.S. degree in electrical engineering, and his Ph.D. degree in electrical engineering with a minor in physics, all from Stanford University, California, in 1984, 1986, and 1991, respectively. He is Fellow of IEEE, OSA, NIAC and a recipient of the DARPA Award for Sustained Excellence (1997), the Bellcore CEO Award (1998), the Mid-Career Research Faculty Award (2004 UC Davis), and the Senior Research Faculty Award (2011 UC Davis).