CMOS-Compatible Scalable Photonic Switch Architecture Using 3D-Integrated Deposited Silicon Materials for High-Performance Data Center Networks

Aleksandr Biberman, Gilbert Hendry, Johnnie Chan, Howard Wang, and Keren Bergman
Department of Electrical Engineering, Columbia University, 500 West 120th Street, New York, New York 10027
biberman@ee.columbia.edu

Kyle Preston, Nicolás Sherwood-Droz, Jacob S. Levy, and Michal Lipson
School of Electrical and Computer Engineering, Cornell University, 428 Phillips Hall, Ithaca, New York 14853

Abstract: We propose a novel fully-integrated scalable photonic switch architecture for data center networks, sustaining nonblocking 256×256 port size with nanosecond-scale switching times, interconnecting 2,560 server racks with 51.2-Tb/s bisection bandwidth.

OCIS codes: (200.4650) Optical interconnects; (130.4815) Optical switching devices

1. Introduction

As high-performance computing systems and data centers continue to achieve performance gains through ever-increasing processor parallelism, more emphasis is placed on developing novel techniques for adaptable interconnection networks capable of sustaining new classes of bandwidth requirements. These interconnection networks must offer unprecedented capacity, energy efficiency, compactness, and flexibility, while minimizing the overall cost of the system. Recent efforts in addressing these challenges have adopted optical communication technology within the traditionally-electrical interconnection networks in data centers, augmenting circuit-switched optical technology at the core of the network to route aggregated optical data between server racks [1,2]. The inherent bandwidth transparency of optical transmission, combined with the ability to densely aggregate bandwidths using wavelength-division multiplexing (WDM), naturally places optics at the forefront of key enabling technologies in next-generation data center networks. However, these recent efforts have employed commodity optical circuit switches with reconfiguration times on the order of milliseconds, limiting the benefits of optics to only a subset of applications that can take advantage of such massive circuits—the shorter packets are still routed through a bandwidth-limited power-hungry electrical interconnection network [1,2].

We present in this work a highly-scalable and energy-efficient switch architecture based on high-performance silicon photonic technology, capable of routing densely-aggregated wavelength-parallel optical data at the core of the optical data center network. By leveraging ultra-broadband electro-optic switching building blocks with sub-nanosecond switching capabilities [3], we develop the switch architecture with reconfiguration times improved up to six orders of magnitude compared with current commodity optical switches. This approach enables the ideal flexibility of optical communication to support arbitrary-length optical packets and circuits—potentially facilitating an all-optical interconnection network at the core of the data center network. Furthermore, we demonstrate how the scalability of this architecture is enhanced with the adoption of novel multi-layer deposited silicon materials, comparing its performance with the single-layer approach. An example layout of the data center optical interconnection network is depicted in Fig. 1, interconnecting server rack clusters using this scalable photonic switch architecture. Most photonic components in this network, including those critical for network ingression and egression, may be implemented using CMOS-compatible high-performance silicon photonic devices such as waveguides [4], modulators [5], filters [6], switches [3], and photodetectors [6].

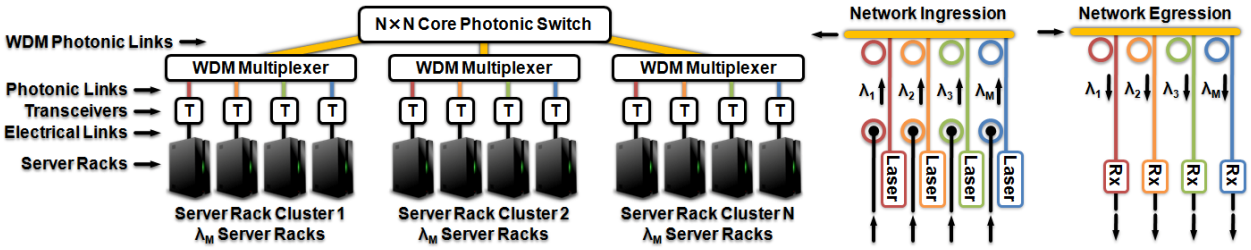
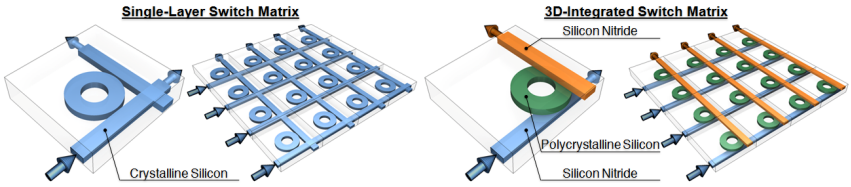


Fig. 1. Layout of the data center optical interconnection network, densely aggregating server rack clusters using WDM, and routing high-bandwidth optical data between server rack clusters using a high-radix photonic switch (left). The switch, as well the ingression and egression components (right), are employed using silicon photonics.



Insertion Loss Parameters

Parameter	Single-Layer	3D-Integrated
Waveguide Propagation	0.5 dB/cm	0.1 dB/cm
Waveguide Crossing	0.05 dB	0 dB
Off-Chip Lateral Coupler	0.5 dB	0.5 dB
Waveguide Pitch (L_{PITCH})	75 μm	75 μm
Waveguide Maximum Injected	20.6 dBm	30 dBm
Switch Maximum Injected	17 dBm	17.8 dBm

Switch Matrix Size	Total Microring Resonators	Any Path Off-Chip Lateral Couplers	Longest Path Propagation Distance	Longest Path Waveguide Crossings	Longest Path Through Ports	Any Path Drop Ports
16×16	256	2	0.24 cm	30	30	1
64×64	4,096	2	0.96 cm	126	126	1
256×256	65,536	2	3.84 cm	510	510	1
N×N	N^2	2	$2N \times L_{PITCH}$	2N–2	2N–2	1

Fig. 2. Silicon photonic 1×2 switch building blocks used for the switch matrix, with corresponding insertion loss parameters, for both types of integration (top), and number of components in each switch matrix size (bottom).

2. Silicon Photonic Materials, Components, and Systems

Silicon photonic devices and systems compatible with the well-refined CMOS platform offer the clearest path toward adaptability and commercialization. Crystalline silicon offers the best set of electrical and optical properties for integrated silicon photonics, but can only be grown from another silicon crystal, making it impossible to deposit multiple layers. Polycrystalline silicon, commonly found in CMOS fabrication as a transistor gate material, is capable of being deposited as an electrically-active light-guiding material [5], but the propagation losses are intolerable over centimeter-scale distances. Silicon nitride, commonly used in CMOS fabrication as a masking layer, is also capable of being deposited, and has been demonstrated with extremely low propagation losses [4], making it an ideal candidate for a passive transport medium. However, silicon nitride does not produce efficient electrically-active devices.

We have recently demonstrated sub-nanosecond electro-optic switching of optical data with up to 40-Gb/s data rates per wavelength channel, observing 12-dB extinction ratios with 1.3-V_{PP} driving voltages [3]. In this work, we build a scalable switch matrix by concatenating 1×2 switches in two dimensions (Fig. 2). The single-layer switch matrix, based on crystalline silicon, comprises unavoidable waveguide crossings, each producing a corresponding insertion loss. Using three distinct optical layers, we can have two independent optical waveguide bus planes with minimal field interaction, coupled together by a middle optical plane for switching (Fig. 2). Here, the 3D-integrated 1×2 switches are located in the electrically-active polycrystalline silicon layer, coupled between two silicon nitride waveguides at opposing layers. The resulting 3D-integrated switch matrix avoids insertion loss due to waveguide crossings, strictly performing all the waveguide crossings out of plane. In both configurations, the electro-optic switches route high-speed optical data between the two perpendicularly-routed low-loss silicon nitride waveguides.

We perform a detailed scalability analysis of the switch matrix for both types of integration by first quantifying all the corresponding insertion loss parameters for each component, as well as developing a model for the number of these components as we vary the size of the switch matrix (Fig. 2).

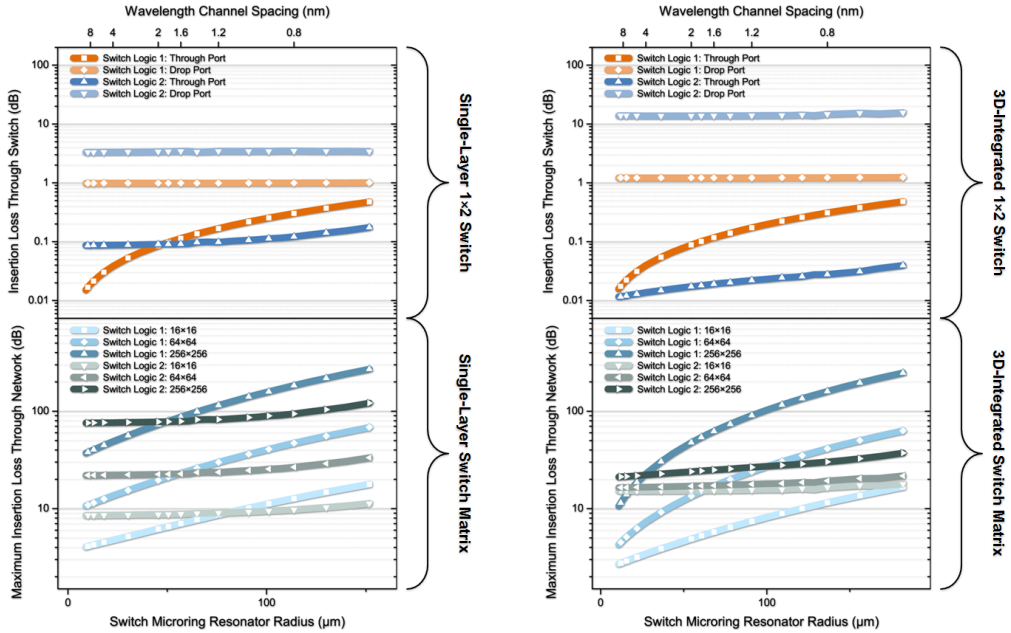


Fig. 3. Insertion loss for both output ports of 1×2 switch, for each type of switch logic and integration, for varying switch microring resonator radius (top), and maximum insertion loss for various sizes of the switch matrix for each type of switch logic and integration, for varying switch microring resonator radius (bottom).

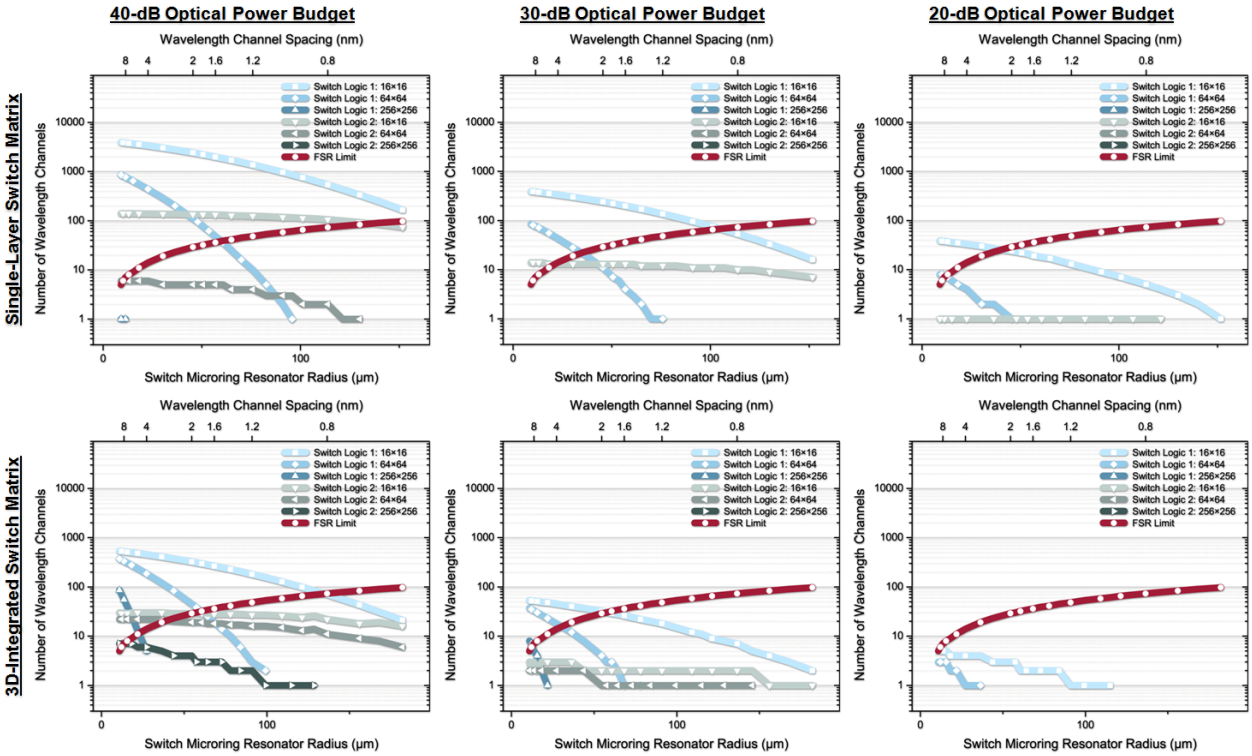


Fig. 4. Number of wavelength channels supported by the optical communication link for various sizes of switch matrix, type of switch logic, integration, and switch microring resonator radius, for three optical power budgets.

3. Scalability Analysis and Conclusions

By finely tuning key physical parameters of the 1×2 switch, including the microring resonator radius and coupling parameters, we are able to optimize each switch for the specific switch matrix architecture. This switch can also be designed to operate in one of two possible modes: *Switch Logic 1* (*Switch Logic 2*), where the diode is *OFF* (*ON*) when the signal is sent to the drop port, and is switched *ON* (*OFF*) to change to the through port. Using these techniques, we optimize insertion loss for the two output ports of the switch to achieve the highest-performing switch matrix.

In Fig. 3, we demonstrate the insertion loss for both output ports of the 1×2 switch for each type of switch logic and integration, as we vary the switch microring resonator radius. Combining these results with those in Fig. 2, we model the optical path with the maximum insertion loss for various sizes of the switch matrix for each type of switch logic and integration, for the same varying switch microring resonator radius (Fig. 3). For each aforementioned configuration, we then quantify the maximum number of wavelength channels supported by the communication link at the core of the network, for three different optical power budgets (Fig. 4). Since the total bandwidth available in our wavelength-parallel approach is limited by the free spectral range (FSR) of the modulator microring resonator (about 59 nm), we plot this limit to determine the number of feasible wavelength channels the system can support (Fig. 4).

Leveraging CMOS-compatible 3D-integrated deposited silicon materials, we demonstrate a fully-integrated scalable photonic core switch architecture feasibly sustaining more than ten (forty) wavelength channels for the switch matrix size of 256×256 (64×64), optically interconnecting at least 2,560 data center server racks and enabling the network bisection bandwidth of 51.2 Tb/s (20 Gb/s per wavelength channel). Furthermore, this architecture represents switching times improvement of up to six orders of magnitude compared with other switching methods, demonstrating the necessary capacity, energy efficiency, compactness, flexibility, and cost reduction, in next-generation data centers.

We acknowledge support from the NSF and Semiconductor Research Corporation under grant ECCS-0903406 SRC Task 2001. This work was part of the National Science Foundation CAREER Program under grant 0446571. This work was performed in part at the Cornell NanoScale Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by the NSF under grant ECS-0335765.

4. References

[1] G. Wang *et al.*, "c-Through: part-time optics in data centers," Proc. Special Interest Group on Data Communication (SIGCOMM), 327–338 (2010).
[2] N. Farrington *et al.*, "Helios: a hybrid electrical/optical switch architecture for modular data centers," Proc. Special Interest Group on Data Communication (SIGCOMM), 339–350 (2010).
[3] A. Biberman *et al.*, "Broadband CMOS-compatible silicon photonic electro-optic switch for photonic networks-on-chip," Proc. Conference on Lasers and Electro-Optics (CLEO), CPDA11 (2010).
[4] M. J. Shaw *et al.*, "Fabrication techniques for low loss silicon nitride waveguides," *Proc. SPIE* **5720,** 109–118 (2005).
[5] K. Preston *et al.*, "Deposited silicon high-speed integrated electro-optic modulator," *Opt. Express* **17,** 5118–5124 (2009).
[6] L. Chen *et al.*, "Ultra-low capacitance and high speed germanium photodetectors on silicon," *Opt. Express* **17,** 7901–7906 (2009).