Photonic Switching Technologies, Architectures, and Integrated-Systems for Future Disaggregated and Optically Reconfigurable Data Centers

S. J. Ben Yoo
Department of Electrical and Computer Engineering
University of California
Davis, California 95616, USA
Email: sbyoo@ucdavis.edu

***Abstract*— This Tutorial covers technologies, architectures, and system-integration for future data centers with optical reconfigurability. Optical interconnects allow disaggregation of computing resources in the data centers thanks to distance-independent energy-efficient and high-throughput communications of photonics. Photonic switching can provide additional benefits of reconfigurability of the interconnection topologies without requiring electronic-switches that accompany store-and-forward mechanisms. Hence, the primary motivation for considering photonic switching in data centers rises from the need for energy-efficient and scalable intra-data center networks to meet rapid increases in data traffic driven by emerging applications, including machine learning. To accommodate such traffic, today's large-scale data centers employ cascaded stages of many power-hungry electronic packet switches interconnected across the data center network in fixed hierarchical communication topologies. Numerous research papers have predicted significant benefits in scalability, throughput, and power efficiency from deploying photonic switches in data centers. However, photonic switching is not yet widely deployed in commercial warehouse-scale data centers at the time of writing this Tutorial due to significant challenges. They are related to (1) cross-layer issues involving control and management planes together with data integrity during switching, (2) scalability to > 5000 racks (> a quarter-million servers), (3) performance monitoring required for reliable operation, (4) currently existing standards allowing limited power margin (3 dB), and (5) other practical (technology-dependent) issues relating to polarization sensitivity, temperature sensitivity, cost, etc. We will discuss possible solutions for future data centers involving cross-layer methods, new topologies, and innovative photonic switching technologies. Furthermore, the Tutorial broadly surveys state-of-the-art photonic switching technologies, architectures, and experimental results, and further covers the details of arrayed-waveguide-grating-router-based switch fabrics offering hybrid switching methods with distributed control planes towards scalable data center networking.**

Keywords*— *data center networking, switching, optical switching, photonics, optical packet switching, optical burst switching, silicon photonics, electronic switching

I. Introduction

Our daily lives critically depend on data communications. Global data center IP traffic grew 11-fold over the past eight years [6] at a Compound Annual Growth Rate (CAGR) of 25%, exceeding 20 Zettabytes per year by 2021 [1]. More recently, driven by the rapid increases in AI and machine learning related traffic, some estimates indicate that the annual data traffic will increase by over 400× over the next 10 years corresponding to CAGR of 82%. At the same time, the global energy consumption in data centers reached 200 TWh in 2020 [2] with a CAGR of 4.4% [3].

Today's data center network architectures heavily rely on cascaded stages of many power-hungry electronic packet switches interconnected across the data center network in fixed hierarchical communication topologies such as Fat-Tree within the data center (see Fig. 1(a))[4]. Due to the limited radix and bandwidth of the electronic switches, warehouse-scale data centers involve a large number of cascaded electronic switches where high energy consumption and latency compound due to repeated '*store-and-forward'* electronic processes. These architectures are also designed with a fixed topology at fixed data rates. On the other hand, as Fig. 1(b) illustrates, employing a passive optical fabric or a reconfigurable optical switch fabric with distributed electronic switches (e.g. ToR) could greatly improve (a) scalability of the capacity and the number of compute nodes (or racks with ToRs), (b) energy-efficiency of the network, (c) modular upgradeability, and (d) cost savings by eliminating many large and power-hungry core electronic switches at the core while keeping the smaller and disaggregated electronic switches (e.g. ToR) at the edge nodes. This transformation not only flatten the interconnect topology of the data center networks with a reduced number of hierarchies, but it also brings the possibility of optical re-configurability enhanced by wavelength division multiplexing (WDM) and silicon photonics. Fundamentally, an all-to-all interconnect topology (shown in Fig. 1 (c)[Left]) can offer uniform and contentionless interconnections between the compute nodes. As actual data centers must handle data movements of nonunifrom and dynamically changing traffic patterns, threfore, their interconnection toplogies and bandwidth assignments should closely reflect those driven by the workflow. 'Application-aware' networking [5] of data centers, would then benefit from a reconfigurable interconnection plaform which can, for example, represent a low-latency all-to-all topology (e.g. Fig. 1 (c)[Left]) at certain

978-3-903176-44-7

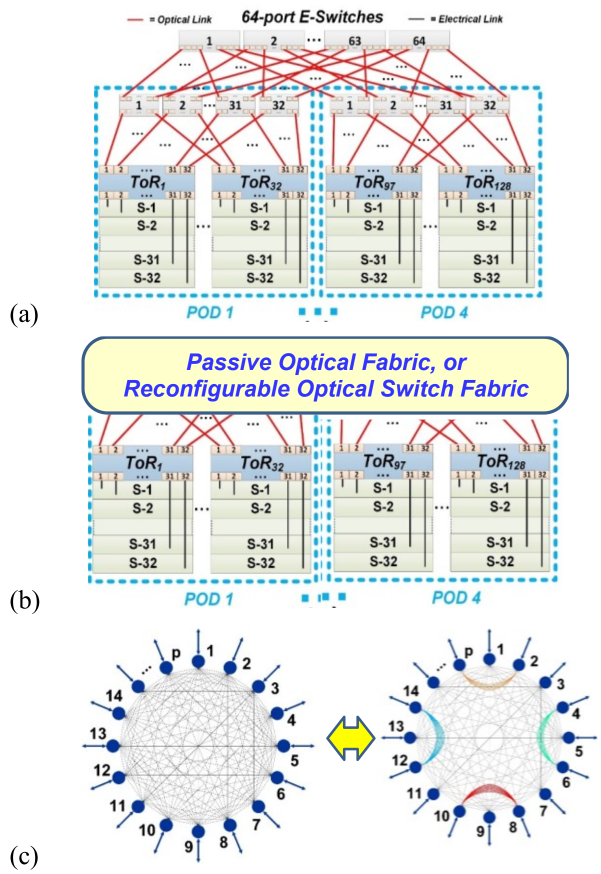

Fig. 1. (a) A fat tree topology using electronic switches at the core and at the aggregation edges of the network, (b) a flattened optically interconnected network example utilizing a passive optical fabric (such as arrayed waveguide routers) or a reconfigurable optical switch with electronic switches at the edges (e.g. ToR), and (c) [Left] all-to-all interconnection and [Right] bandwidth-steered interconnected topology after reconfiguration [4].

times and high-bandwidth neighbor communication at other times (e.g. Fig. 1 (c) [Right]) [4]. For example, as Fig. 2 illustrates, workloads involving deep neural networks, map-reduce, sorting, and many others from time to time require all-to-all communications, which typical data center interconnect topologies cannot support easily unless they involve numerous sequential and hierarchical communications between the nodes.

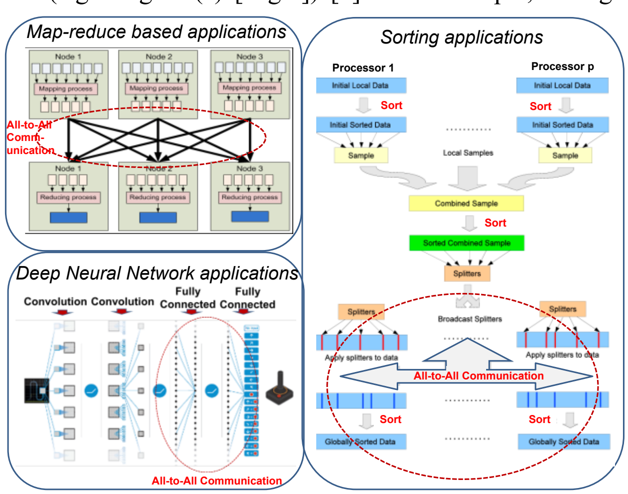

Fig. 2. Examples of applications showing very different data movement patterns. Map-reduce, deep neural network, and sorting applications all show all-to-all communication pattern at some point of the workflow but not throughout the entire workflow.

Hot spots formed during such dynamically changing bursty communication patterns further aggravates the throughput and energy-efficiency of the data centers. Reducing congestion and latency to speed up the execution time of different application threads is a key aspect for energy saving, since a significant amount of power is consumed in the servers and not only in the communication network [4].

II. Electrical vs. Optical, and Hybrid Switches

Typical electronic switch ASICs used in today's datacenters are built with shared-buffer architectures[6]–[8]. The key features are as follows. First, contention resolution and arbitration exploit the time-domain store-and-forward process. Second, even if the switching speed of the electronic switching fabric is much slower than the bit rate, there is no need to include a '*guard time*' and there will be no concerns for losing bits of information during the switching. This contrasts with optical switches where the optical guard time must be included in the data stream to avoid losing meaningful bits during the transition of the switching states. Third, the electronic switches can conduct sophisticated electronic processing, queuing, and DPI on the packets while the packets are in the buffer memory.

On the other hand, electronic switches relying on the store-and-forward methods have limitations in data rates, energy-efficiency, and switch radix. Beyond the electronic switch capacity limit, the electronic switches need to be clustered to achieve higher capacity at the expense of consuming additional I/O ports just for interfacing between the switches in the cluster. Considering that more than half of the switch fabric energy consumption comes from the linecards, increasing the linecards just for clustering is detrimental to realizing energy-efficient data centers at large scale.

All-optical switches consisting of passive photonic components such as optical MEMS, Mach-Zehnder switches, and even semiconductor optical amplifier (SOA) arrays need not have components that must respond to every bit of the datagram at the bit rate, and the energy consumption in such devices are relatively modest and independent of data rates. In many cases, high-radix switches are more easily realizable with optics compared to electronics as optical switches do not have the same limitations imposed in electronic switch fabrics.

A class of hybrid switches utilizing electronic switches and optical switches (or passive optical interconnects) can exploit high-efficiency of small electronic switches and high-scalability of optical interconnects at the same time. It is possible to create an arbitration and contention-free all-to-all $N\times N$ switch fabric by combining simple 1:N selection electronic switches and a passive $N\times N$ wavelength-routing optical device such as an arrayed-waveguide-grating-router (AWGR) simultaneously supporting N^2 circuits or data flows without contention [9]–[12]. Additional advantages of such a hybrid switch are that it can scale easily while retaining distributed control plane in the electronic switches and that it can benefit from a rich set of processing capability by the electronic switch without having to include the guard time for switching.

Lack of viable optical buffers add challenges to both the control plane and the data plane of optically reconfigurable data center networks. The electrically reconfigurable data center networks can mitigate such challenges thanks to electronic buffers, despite inferior scalability and energy-efficiency. The hybrid switching networks with optically interconnected distributed electronic switches can possibly achieve the benefit of scalability, energy-efficiency, and agile reconfigurability. The question then is whether to add reconfigurability to this optical interconnection of distributed electronic switches.

III. Data Plane, Control Plane, and Management Plane

The introduction of TCP/IP and the availability of Layer 2 and Layer 3 protocols such as Ethernet, ATM, SONET, and OTN meant that hardware switches with proper protocols embedded in the linecards can readily achieve network switching since the control plane and management plane protocols can run on those protocol-specific linecards (and the switch fabric). Such switches employed distributed control and management planes, using the protocol-specific information embedded in the datagram.

Recent electronic switches have evolved towards better programmability, reconfigurability, and protocol independency. Adoptions of an open source programming language such as P4 [13], which allows fast reconfiguration and software-level programmability, greatly facilitates deployment of large-scale data centers and computing clusters, making it easy for operators to control and manage these complex systems. This also meant that optical switches developed for 2nd generation and 3rd generation optical networking could play an active role in data center networks with the SDN paradigm. Optical MEMS switches already developed for telecom more than a decade ago could be readily deployed in data centers. However, scalability of the centralized control and management planes becomes challenging if rapid reconfigurations are required at high load in a network with a large number of nodes. For this reason, optical switches face challenges if optical reconfigurations are required rapidly and frequently in a large data center network, while electronic switches can more readily support such reconfigurations due to the integrated electronic memory and switch fabric despite their high-power consumption and capacity limitations. Thus, for dynamic optical circuit switching data center networks with reconfigurable optical switches, cross-layer issues involving the application, transport, network, link, and physical layers inevitably become extremely important. However, this cross-layer issue remains as unsolved and too challenging for data center networks at scale.

IV. Scaling and Disaggregation of Data Switching

Today's data centers often employ many thousands of racks, and the scalability of the data centers is a compelling requirement while networking such a large-scale data center becomes an immense challenge seen both from the data plane and the control plane perspectives. Further, the multi-tenant data centers are becoming more popular running heterogeneous applications simultaneously. Hence, a scalable and disaggregated switching network is desired in the data plane, while controllability, manageability, and virtualization [14] of the data center network are necessary.

In an electronic data center network shown in Fig. 1 (a), scaling-up becomes challenging due to the limitations in the bandwidth, radix, and switching capacity of the electronic switches if electronics-only solutions are sought. In scale-up data center networks, each individual network devices must increase its capacity and bandwidth, which is difficult to achieve with electronics-only solutions. Scaling-out data center networks utilizing commodity electronics is far more attractive from both flexibility and energy-efficiency perspectives [15], as demonstrated by Facebook's F16 networks [16].

On the other hand, as Fig. 1(b) illustrates, employing a passive optical fabric or a reconfigurable optical switch fabric with distributed electrical switches (e.g. ToR) at the edges could greatly facilitate scalability and disaggregation of the data center networks while offering significant energy, modular upgradeability, and cost savings by eliminating large and power-hungry electronic switches at the core. Fig. 3shows one such example employing a $N \times N$ cyclic arrayed waveguide grating router (AWGR) with all-to-all interconnection capability by optical wavelength routing. As we will discuss later, since such an AWGR supports N^2 simultaneous optical circuits without contention, the switch capacity can scale to, for example 26.2 Pb/s interconnection capacity using 100 Gb/s transceiver per port for N=512 [17] (512^2*100 Gb/s=26.2 Pb/s).

Scalability of all-optical switch fabrics are typically limited by the number of required switching elements (that may scale as $O(N^2)$, $O(N\log_2 N)$, or $O(2N)$) or by the number of cascaded stages of optical switches. In some cases, such as c-Through [18] or Helios [19] networks, the authors proposed to use optical switches to supplement or partially replace electronic switches to improve communications in data centers.

Alternatively, hybrid switching consisting of wavelength routing and electronic switches achieves this arbitration-free all-to-all interconnection. Then, each node is interconnected in an all-to-all topology as shown in Fig. 3(a) where P nodes are directly optically interconnected to each other. Physically, such a network would require $\frac{P(P-1)}{2}$ pairs of optical fibers. As Fig. 3 (b) illustrates, this interconnection can be greatly simplified by introduction of WDM and a wavelength routing device such as an AWGR with the well-known cyclic frequency routing characteristic, where an $N\times N$ AWGR interconnects [10], [11],[22] for p number of nodes emitting N wavelengths, where $N = p + \mu$. Hierarchical switchingcan be achieved as illustrated in Fig. 3 (c) [22]. Fig. 3 (d) and (e) illustrate wavelength routing properties of cyclic frequency routing AWGRs (shown is an N=5 example) [23][24], and Fig. 3 illustrates how a data center network with a passive optical wavelength routing device, such as a cyclic frequency $N\times N$ AWGR [23][24], to interconnect N racks with a Top of the Rack (ToR) switch with N wavelength WDM ports.

The scalability of supporting many compute nodes can be achieved in three ways. The first method is to introduce one very large $N\times N$ AWGR. Although silicon photonic 512×512

AWGRs [17] and other large-scale cyclic-frequency AWGRs on PLCs have been demonstrated, this method is considered impractical because it would require a large number of wavelengths (N) and TRXs and it would induce a substantial amount of crosstalk. To address the wavelength and crosstalk issues, a Thin-CLOS architecture [25], [26] has been designed to achieve the same all-to-all interconnection by using many small $W \times W$ AWGRs so that the number of wavelengths and the amount of crosstalk would reduce significantly.

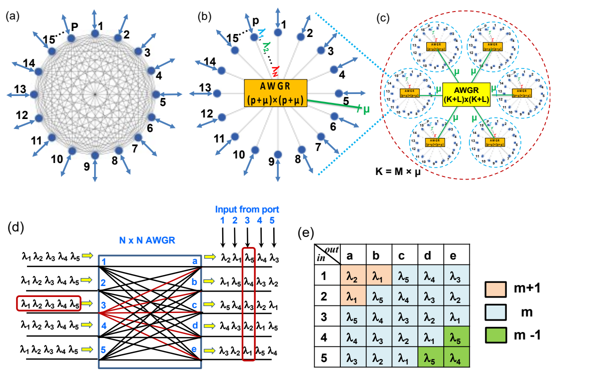

Fig. 3. a) Fully connected all-to-all interconnection network, (b) fully-connected all-to-all interconnection network utilizing wavelength routing by an Arrayed Waveguide Grating Router (AWGR), (c) Hi-LIONS with fully connected subnetworks that are interconnected with a reconfigurable optical switch), (d) all-to-all wavelength routing interconnection pattern of a $N \times N$ cyclic AWGR using N wavelengths ($N = 5$ example), (e) wavelength routing property of the $N \times N$ cyclic AWGR ($N = 5$ example) offering all-to-all interconnects using N wavelengths.

V. Interconnection Network Topologies

Ideally, the interconnection topology of the data center network should closely match the data flow pattern according to the workload of the data center at any given time. In practice, as Fig. 2 illustrates, the data flow pattern changes from workload to workload, and from one phase of the workload to another even within the same workload. Fig. 2 also shows that all-to-all communication is necessary at some point in time in all three application examples but not necessary all the time. On the other hand, supporting all-to-all communications in a network topology realized by interconnection of low-radix switches often cause elevated congestion and latency. For these reasons, a reconfigurable optical switch capable of supporting any arbitrary connectivity including all-to-all interconnection is desirable. Various interconnection topologies: Flattened Butterfly, FatTree, Dragonfly, 3D Torus, 3-stage CLOS, Hypercube, and SlimFly are considered typically for electronic switches, and hybrid data center interconnection topology involving both electronic and reconfigurable optical switches including are possible for c-Through [18], Helios [19], and Optical Switching Architecture (OSA) [27].

VI. Time Scales for Reconfiguration and Limitations Imposed by the Control Plane

The previous section compels us to consider optical switches capable of configuring the data center interconnection topology that would be optimally matched to the data flow pattern for the given workload, or even reconfiguring during the run time of the application as the data flow pattern changes within the run time. As we will see later in this section, it is extremely challenging to realize scalable and low-latency control planes for such a reconfigurable optical circuit switching driven by the dynamicity of the changing traffic patterns.

In standard or dynamic circuit networks' physical layers, 100 μs or longer timescale reconfiguration may be sufficient, but flow-switching or burst-switching should achieve reconfigurations at below 100 μs, while packet-switched networks must achieve switching at much faster time-scale than the length of the packets (< 1 ns). Seen from the applications or workload, job-level reconfigurations can be at timescales longer than 1 ms, while flow-level reconfigurations and packet-level reconfigurations should achieve < 100 μs and < 10 ns respectively. In terms of the control plane, depending on the scale of the data center network and the scheduling algorithm, the centralized SDN control plane may be able to achieve and complete reconfiguration of the data center networks at 1 ms or longer, while faster reconfiguration should resort to distributed hardware control using FPGAs or ASICs.

The benefit of reconfigurations of communication networks in data and computing systems have been discussed from the perspective of efficiently and effectively utilizing the available resources (processing, memory, and communications) [27]–[30]. In particular, mitigating hot-spot creations in data centers [28], [31], [32] can be where optical reconfigurations can prove to be very useful.

These findings indicate that a static optical circuit network may not be effective for future data centers and a dynamically reconfigurable optical network should be considered. But the time scale of these bursts at < 25 μs casts serious challenges for scheduling and for control plane designs across the data center network. Even if optical switches can reconfigure in less than 1 ns, the control plane cannot achieve coordination between all the compute nodes in the data center in such a short amount of time. In a high-performance computing system running a single threaded application with predictable changes in traffic patterns (e.g. map-reduce application transition from map-phase to reduce-phase), such a reconfiguration is conceivable if a guard time is incorporated, but it is difficult to predict such traffic patterns in a data center running many heterogeneous workloads simultaneously. Some studies are underway to apply machine-learning methods to statistically predict data flow patterns within the data centers [33][34].

The challenges in implementing centralized control planes for optical reconfiguration also raise an interesting question regarding the viability of optical-packet switching (OPS) and optical-burst switching (OBS), in addition to dynamic optical circuit switching (OCS) in future data centers. For OCS based intra-datacenter networks, we assumed the central control plane triggering reconfiguration of optical switches for dynamical reconfiguration of optical circuits. The scalability of this method depends greatly on scheduling algorithms and on the dynamicity and the load of the traffic pattern. The NEPHELE project [35] introduced WDM ring network with optical reconfiguration based on TDMA time-slot allocated by a SDN (OpenFlow) control plane to schedule these applications, and showed that the makespan can reach 48% when short-term load dynamicity is high [36]. Such centralized schedulers inevitably add scheduling delays which can become unacceptably high in large data center networks. Just to poll the traffic demands, the delay T_D scales as $O(N^2)$ [37], which corresponds to 4 ms for N=5000 racks at C_{BW}=100 Gb/s. Ref [38] suggested, with some optimism, an *observe-analyze-act* framework including a number of intelligent algorithms while recognizing unsolved problems, while Ref. [37] declared that the central scheduling a dead-end unless (a) fixed scheduling without considering application awareness and with additional latency, or (b) distributed scheduling with less accurate or no coordination is adopted.

Table 1. Summary of various optical switching technologies and their comparisons. (SNB: Strictly Nonblocking, WNB: Wide-sense Nonblocking, RNB: Rearrangeable Nonblocking).

	Switching Time	Scalability	Crosstalk for each stage [dB]	PDL [dB] for the Fabric	Optical Losses per stage [dB]	3dB Optical Pass-band Bandwidth [GHz]	Switch Fabric Topology	Blocking
Free-space opto-mechanical	~ 4 ms	384x384	< -55	< 0.1	< 2 [fiber-to-fiber]	> 10,000	Point-to-Point Mesh	SNB
Free-Space Optical MEMS	3D: ~10 ms; 2D: ~5 ms	3D: 1296x1296; 2D: 32x32	3D: < -60; 2D: < -50	3D: < 0.1; 2D: < 0.3	3D: < 2; 2D < 3.5 [fiber- fiber]	> 10,000	Point-to-Point Mesh	SNB
Waveguide Optical MEMS	~ 0.9 µs	240x240	-70	> 3	Thru: 0.026; Drop: 0.47 [on-chip]	> 10,000	Crossbar	SNB
PLZT Switches	350 ns	4x4	-25	> 3	> 1 (est) [on-chip]	> 10,000	Point-to-Point Mesh	SNB
PILOSS Switches	~10 µs	32x32	Si: -20; SiO2: -56	< 2	Si: 19.7; SiO_2: 6.6 [fiber- fiber]	> 10,000	Cylindrical Spanke-Benes	SNB
Mach-Zehnder Switches (excluding PILOSS)	TO: ~10 µs; EO: ~10 ns	16x16	Single: -23; Nested: -35	Single Pol.	~ 1 [on-chip]	> 10,000	Benes, Dilated-Benes, Dilated-Banyan, Crossbar	SNB; RNB
Liquid Crystal Switches	~ 5 ms	2x2 WOXC; 1x9 WSS	< -35	0.2	2 [fiber-to-fiber]	> 10,000	2x2 WOXC; 1x9 WSS	For 2x2: SNB;
Micro Resonator Ring Switches	TO: ~10 µs; EO: ~10 ns; MO: ~100 ns	8x8 (crosstalk limited)	-28	Single Pol.	TO Thru/Drop: =0.2/ 0.6; EO Thru/Drop: = 0.33/1.64 MO Thru/Drop: = 0.33/2 [on-chip]	1st order: ~30; 2nd order: ~60; 8thorder: ~100	Crossbar, Benes, Dilated-Benes, etc.	Crossbar: SNB; Others: RNB
Wavelength Routing Switches	~ 1 ns	512x512	-25	< 0.1 (Pol. Ind. detector)	~3 [on-chip]	~200 GHz or ~70% of channel spacing	Wavelength Routing Star	SNB

VII. Switching Technologies

In considering optical switching technologies for data centers, there are countless attributes that must be considered. As discussed in [8], these attributes can be summarized in three categories: signal quality, configuration, and performance. Table 1 summarizes various photonic switching technologies. It is also possible to integrate the switching functions to all-to-all interconnects discussed as LIONS so that reconfiguration from all-to-all to arbitrary interconnection is possible. This new reconfigurable wavelength routing switch are called Flex-LIONS [39]–[40].

VIII. Summary and Future Prospects

The accelerating trend of exponential growth in data traffic and the fact that the large portion of the data traffic reside in the data centers imply that photonic switching could play increasingly important roles in future scalable data and computing systems. However, there are significant challenges relating to scheduling of many concurrent applications with dynamically changing traffic patterns when attempting to introduce photonic switching technologies in large-scale data centers. Cross-layer design of scheduling and control will be important. Centralized control plane is effective only if it can handle dynamic high-capacity applications in a scalable manner. A combination of distributed and centralized control planes is expected to be necessary. Further, recently developed silicon photonic switches and the availability of foundry-based manufacturing and packaging exploiting CMOS electronic industry ecosystem can accelerate electronic-photonic integration and development of photonic switching embedded in compute nodes, backplanes, and racks.

ACKNOWLEDGMENT

This work was supported in part by DoD #H98230-16-C-0820 and NSF grant # 1611560. The author would like to thank the many researchers around the world who contributed to this paper, especially R. Proietti, X. Xiao, G. Liu, and Yu Zhang.

IX. REFERENCES

[1] Cisco, "Global data center IP traffic from 2012 to 2021, by data center type (in exabytes per year)," 2019. [Online]. Available: https://www.statista.com/statistics/227268/global-data-center-ip-traffic-growth-by-data-center-type/.

[2] N. Jones, "The Information Factories," *Nat. Mag.*, no. 561, 2018.

[3] A. Shehabi *et al.*, "United States Data Center Energy Usage Report," United States, 2016.

[4] S. J. B. Yoo, "Prospects and Challenges of Photonic Switching in Data Centers and Computing Systems," *J. Light. Technol.*, pp. 1–1, 2021.

[5] 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.

[6] M. Kalkunte, "Design of a Switch Chip," 2019. [Online]. Available: http://web.stanford.edu/class/cs349f/slides/StanfordGuestLecture-April-08-2019.pdf.

[7] "Cisco Nexus 3000 Series Switches." [Online]. Available: https://www.cisco.com/c/en/us/products/collateral/switches/nexus-3548-switch/white_paper_c11-715262.html.

[8] S. J. B. Yoo, "Optical packet and burst switching technologies for the future photonic Internet," *J. Light. Technol.*, vol. 24, no. 12, pp. 4468–4492, 2006.

[9] 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, 2013.

[10] Y. W. Yin, R. Proietti, X. H. Ye, C. J. Nitta, V. Akella, and S. J. B. B. Yoo, "LIONS: An AWGR-Based Low-Latency Optical Switch for High-Performance Computing and Data Centers," *IEEE J. Sel. Top. Quantum Electron.*, vol. 19, no. 2, 2013.

[11] S. J. J. Ben Yoo, R. Proietti, and P. Grani, "Photonics in Data Centers," in *Optical Switching in Next Generation Data Centers*, F. Testa and L. Pavesi, Eds. Cham: Springer International Publishing, 2018, pp. 3–21.

[12] X. H. Ye, S. J. B. Yoo, and V. Akella, "AWGR-Based Optical Topologies for Scalable and Efficient Global Communications in Large-Scale Multi-Processor Systems," *J. Opt. Commun. Netw.*, vol. 4, no. 9, pp. 651–662, 2012.

[13] P. Bosshart *et al.*, "P4: programming protocol-independent packet processors," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 87–95, 2014.

[14] M. F. Bari *et al.*, "Data Center Network Virtualization: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 2, pp. 909–928, 2013.

[15] A. Greenberg *et al.*, "VL2: A scalable and flexible data center network," *Computer Communication Review*, vol. 39, no. 4. pp. 51–62, 2009.

[16] Facebook, "Reinventing Facebook's data center network," 2019. [Online]. Available: https://engineering.fb.com/data-center-engineering/f16-minipack/.

[17] S. Cheung, T. Su, K. Okamoto, and S. J. B. Yoo, "Ultra-compact Silicon Photonic 512x512 25-GHz Arrayed Waveguide Grating Router," *Sel. Top. Quantum Electron. IEEE J.*, vol. PP, no. 99, p. 1, 2013.

[18] G. Wang *et al.*, "c-Through: part-time optics in data centers," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 327–338, 2010.

[19] N. Farrington *et al.*, "Helios: a hybrid electrical/optical switch architecture for modular data centers," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 339–350, 2010.

[20] R. Proietti, C. J. Nitta, Y. W. Yin, R. X. Yu, S. J. B. Yoo, and V. Akella, "Scalable and Distributed Contention Resolution in AWGR-Based Data Center Switches Using RSOA-Based Optical Mutual Exclusion," *IEEE J. Sel. Top. Quantum Electron.*, vol. 19, no. 2, 2013.

[21] X. Ye *et al.*, "DOS: a scalable optical switch for datacenters," *Proc. 6th ACM/IEEE Symp. Archit. Netw. Commun. Syst. (ANCS 2010)*, pp. 1–12, 2010.

[22] Z. Cao, R. Proietti, and S. J. B. Yoo, "Hi-LION: Hierarchical Large-Scale Interconnection Optical Network With AWGRs [Invited]," *J. Opt. Commun. Netw.*, vol. 7, no. 1, pp. A97–A105, 2015.

[23] K. Okamoto, T. Hasegawa, O. Ishida, A. Himeno, and Y. Ohmori, "32×32 arrayed-waveguide grating multiplexer with uniform loss and cyclic frequency characteristics," *Electron. Lett.*, 1997.

[24] P. Bernasconi, C. Doerr, C. Dragone, M. Cappuzzo, E. Laskowski, and A. Paunescu, "Large N×N waveguide grating routers," *J. Light. Technol.*, 2000.

[25] R. Proietti *et al.*, "Experimental Demonstration of a 64-Port Wavelength Routing Thin-CLOS System for Data Center Switching Architectures," *J. Opt. Commun. Netw.*, vol. 10, no. 7, pp. B49–B57, 2018.

[26] R. Proietti, Z. Cao, C. J. Nitta, Y. Li, and S. J. Ben 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.

[27] K. Chen *et al.*, "OSA: An Optical Switching Architecture for Data Center Networks With Unprecedented Flexibility," *IEEE/ACM Trans. Netw.*, vol. 22, no. 2, pp. 498–511, 2014.

[28] G. Porter *et al.*, "Integrating microsecond circuit switching into the data center," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 447–458, 2013.

[29] G. Wang, T. S. E. Ng, and A. Shaikh, "Programming your network at run-time for big data applications," *Proceedings of the first workshop on Hot topics in software defined networks*. ACM, Helsinki, Finland, pp. 103–108, 2012.

[30] R. Tessier, K. Pocek, and A. DeHon, "Reconfigurable Computing Architectures," *Proc. IEEE*, vol. 103, no. 3, pp. 332–354, 2015.

[31] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, Melbourne, Australia, pp. 267–280, 2010.

[32] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 92–99, 2010.

[33] M. Balanici and S. Pachnicke, "Machine Learning-Based Traffic Prediction for Optical Switching Resource Allocation in Hybrid Intra-Data Center Networks," in *2019 Optical Fiber Communications Conference and Exhibition, OFC 2019 - Proceedings*, 2019.

[34] X. Chen *et al.*, "Machine-Learning-Aided Cognitive Reconfiguration for Flexible-Bandwidth HPC and Data Center Networks [Invited]," *J. Opt. Commun. Netw.*, vol. 13, no. 6, pp. C10–C20, 2020.

[35] P. Bakopoulos *et al.*, "NEPHELE: An End-to-End Scalable and Dynamically Reconfigurable Optical Architecture for Application-Aware SDN Cloud Data Centers," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 178–188, 2018.

[36] K. Kontodimas, K. Christodoulopoulos, E. Zahavi, and E. Varvarigos, "Resource allocation in slotted optical data center networks," in *2018 International Conference on Optical Network Design and Modeling (ONDM)*, 2018, pp. 248–253.

[37] E. Zahavi, "ODCNs Architectures Fundamental Limits," *Optical Fiber Communication Conference*. Optical Society of America, San Diego, California, 2019.

[38] H. H. Bazzaz *et al.*, "Switching the optical divide: fundamental challenges for hybrid electrical/optical datacenter networks," *Proceedings of the 2nd ACM Symposium on Cloud Computing*. ACM, Cascais, Portugal, pp. 1–8, 2011.

[39] 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.

[40] X. Xiao *et al.*, "Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 2, pp. 1–10, Mar. 2020.