

# Autonomous Dynamic Bandwidth Steering with Silicon Photonic-Based Wavelength and Spatial Switching for Datacom Networks

Yiwen Shen, Alexander Gazman, Ziyi Zhu, Min Yee Teh,  
Maarten Hattink, Sebastien Rumley, Payman Samadi, Keren Bergman

*Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA*  
ys2799@columbia.edu

**Abstract:** We present an autonomous SDN network architecture that leverages the spatial and wavelength switching capabilities of silicon photonics microring-based circuits for self-adaptive bandwidth steering. These functionalities are seamlessly integrated and demonstrated in a datacom testbed.

**OCIS codes:** (060.4510) Optical communications, (060.4253) Networks, Circuit Switched

## 1. Introduction

As new data-intensive applications emerge and grow, the current static network architectures of data center (DC) and high performance computing (HPC) systems become increasingly inadequate. DC providers need new ways of meeting bandwidth and latency requirements with minimal cost and energy usage. The solution is the development of flexible, reactive networks that can utilize network resources efficiently and at low cost. This is enabled by the advent of software-defined networking (SDN) which disseminates control plane intelligence to the physical network entities from the higher layers. At the physical layer, novel interconnect technologies such as silicon photonics (SiP) enable the movement of large amounts of traffic while providing low power consumption and low fabrication costs at large scales with CMOS compatibility. While the integration of silicon photonic switching technologies into basic network systems has seen experimental prototypes [1–4] and theoretical works that examine the scalability of SiP switches [5, 6], there has been little study on a control plane to integrate and synergistically utilize SiP switching within a large DC or HPC packet-switched network environment.

In our previous work, we explored the feasibility of integrating silicon photonic switches into a conventional electronic datacom network [7]. Here we extend our network architecture to (1) enable adaptive bandwidth steering using SDN traffic monitoring, and (2) combine it with the versatile spatial and wavelength switching capabilities of a SiP microring resonator (MRR) based device. We fully implement a network testbed and perform simulations to compare its performance to a conventional static network.

## 2. System Architecture and Operation

Fig.1(a) shows our network architecture, consisting of the control and data planes. The control plane consists of the Ryu-based SDN controller, which acts as the top-level management of both the electronic packet switches (EPSs) through the OpenFlow protocol, as well as the physical layer switching enabled by the SiP devices. Our in-house application called the FPGA Controller communicates with FPGAs to provide the necessary bias signals onto the attached digital-to-analog converters (DACs) to cause a state change in the SiP devices. A more in-depth description of the control plane components can be found in our previous work [7].

The data plane shows the connections of the servers and top-of-rack (ToR) EPSs to our SiP MRR device. Each server transmits at a unique wavelength, which are connected to the EPS and subsequently multiplexed together as input to the bus waveguide of the MRR device. The resonances of each MRR are tuned to drop one of input wavelengths where it is then connected to the receiving port of another server. Depending on which wavelengths are being dropped at each MRR, we can set up a bidirectional link between our choice of servers, allowing the SiP device to act as a spatial and wavelength switch. To illustrate this concept more clearly, we show our device here with 4 input wavelengths (1546.92 nm, 1550.12 nm, 1554.94 nm, and 1556.55 nm) and 4 MRRs, allowing the device to operate under three possible  $2 \times 2$  switching configurations, shown on the far right of Fig.1(b). The plots in Fig.1(b) display the output spectrums without amplification on each MRR as a result of tuning the resonances for each switching configuration. In each plot there are four peaks corresponding to the four signals from the servers. The highest peak labeled in each plot is the wavelength that the MRR is tuned to, while other peaks are at least 10 dB lower in power. Fiber-to-fiber loss of our device ranges from 25 to 30 dB, so pre-amplification is required. One key advantage of our system is that all input and output signals enter the same port of the SiP device, allowing for only a single optical pre-amplifier needed to amplify

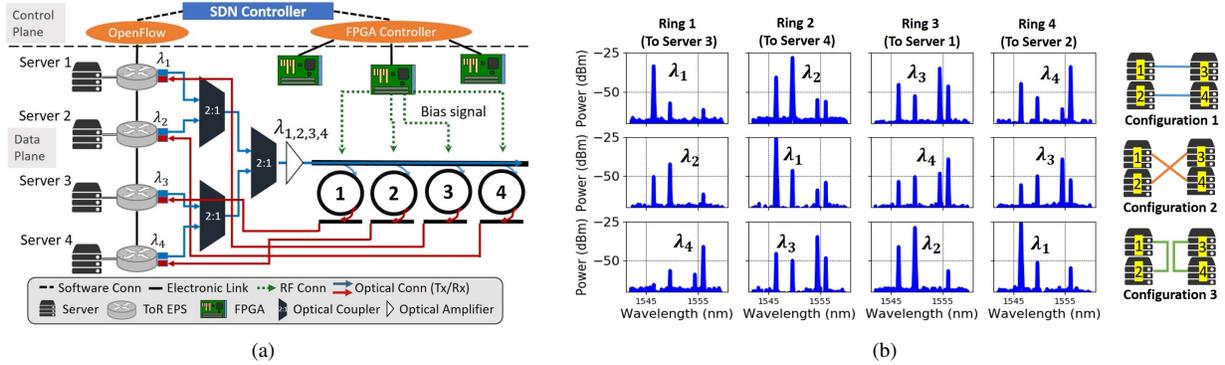


Fig. 1. (a) Network architecture showing the connections of the servers and top-of-rack EPSs to the SiP MRR device, (b) output spectrums of each MRR showing the resonance wavelength (labeled) and crosstalk from other signals for all switching configurations

all signals. This is in contrast to an Mach-Zehnder Interferometer (MZI) based switch, which requires an amplifier for every input and output port. This advantage becomes increasingly prominent as the number of ports increases.

### 3. Control Plane Workflow

We extended our SDN control plane from our previous work to provide autonomous bandwidth steering through two modules: (1) the Traffic Monitor and (2) the Network Optimizer, both used to determine the best switching configuration the SiP device should be set to during each monitoring timeframe. The Traffic Monitor periodically obtains flow statistics from both the static and dynamic flows on each EPS in the network. The Network Optimizer organizes the flows into groups corresponding to Configurations 1, 2, and 3, (shown on the right of Fig.1(b)) and then sums up the amount of bytes transferred by each group of flows, which are then subtracted from the byte count values obtained in the previous monitoring iteration. This difference is divided by the monitoring time to obtain the throughput. The total throughput values for each switching configuration are compared, and the largest value will dictate which configuration the SiP device is set to for this current iteration. The corresponding bias signals to tune each MRR of the SiP device to the proper wavelength, and the flow rule changes to the EPSs are then applied by the control plane. The process then repeats itself for the next monitoring iteration.

### 4. Simulation and Experimental Results

*Testbed:* Fig.2(a) shows the topology of the testbed used to demonstrate the feasibility of our concept on a real network. Our testbed is a Dragonfly topology consisting of 4 groups of 4 EPSs each, with 2 servers connected to each EPS. The intra-group connections are 10G Direct-Attached copper cables and the inter-group connections are optical links with 10G DWDM SFP+ transceivers with 24 dB power budget. The EPSs are bridges created on two Pica8 Ethernet switches. The SiP MRR device used contains 8 MRRs, and we use four of the MRRs to connect the four dynamic inter-group links. The resonance responses of each MRR are separated by 1.27 nm with an FSR of 13 nm and 3 dB bandwidth of 0.7 nm [2]. The switching latency of the MRRs is approximately 60  $\mu$ s (Fig.2(b)).

*Simulation Results:* We performed simulations using SST/macro to measure the throughput for a 6-group and 8-group dynamic network using a simulated SiP device with 6 MRRs and 8 MRRs, respectively, for automated reconfiguration. The topology used is similar to that of the experimental testbed, which consists of static 10G links between each group arranged in a ring, and each group also has a dynamic 10G link to the output port of one of the MRRs of the SiP device. Each group sends a continuous data stream to every other group in a uniformly distributed manner every 1 second. We used 0.2 and 0.5 second end-to-end switching latencies for each configuration change, according to experimental results [7]. The performance of these dynamic networks were compared to that of static networks, which had the same topology and link bandwidths but with static links everywhere. Fig.2(d) shows the percentage improvement in overall network throughput by dynamic networks over static networks. For a 0.5 second switching time, the 6-group network achieves 33% more throughput on average per group, while the 8-group network achieves 12% more throughput on average per group. For a 0.2 second switching time, the 6-group network achieves 53% more throughput on average per group, while the 8-group network achieves 33% more throughput on average per group.

*Experimental Results:* Fig.2(c) shows the latencies associated with the control plane. The traffic monitoring module takes up 64% of the total control plane latency, at 714  $\mu$ s, while the network optimization step requires 177  $\mu$ s and time to set the switching configuration requires 223  $\mu$ s. Fig.2(e) illustrates the self-adaptive bandwidth steering capability of our control plane. The monitoring interval is set to 2 seconds. Initially, Server 5, 6, 7, and 8 are sending background

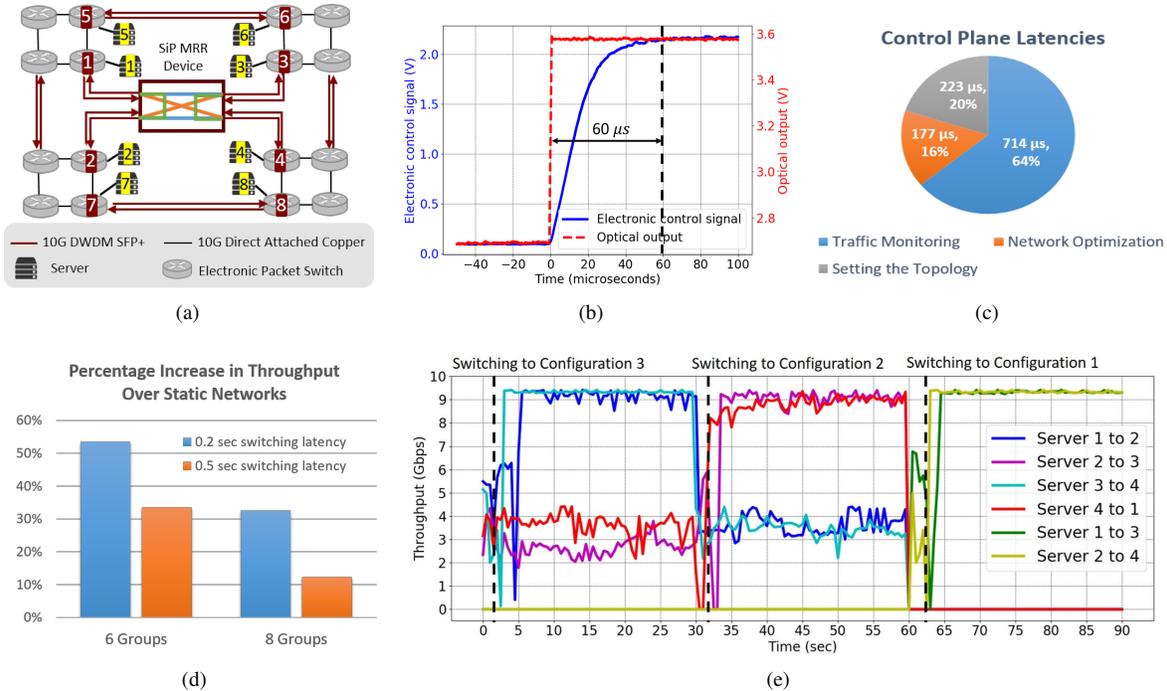


Fig. 2. (a) Testbed consisting of a Dragonfly topology and a SiP MRR switch for inter-group reconfiguration, (b) switching time of an MRR, (c) breakdown of control plane latencies, (d) percentage increase in overall throughput for 6-group and 8-group dynamic networks over static networks obtained through simulations, (e) throughput of various inter-group flows over time demonstrating control plane bandwidth steering capabilities

traffic to each other over the inter-group links that are not connected to the SiP device. Then we allow more traffic to be sent between Server 1 to 2 and Server 3 to 4, compared to traffic from Server 2 to 3 and from Server 4 to 1. The control plane detects this change and initiates the first configuration change at approximately 2 seconds to Configuration 3. This configuration provides a direct connection for traffic between Server 1 to 2 and Server 3 to 4 so that they are able to reach near full link capacity, while the other two flows (2 to 3 and 4 to 1) must compete with the background traffic and are therefore limited in throughput. At 30 seconds, the data transfer from Server 1 to 2 and 3 to 4 decreases, while the traffic between Server 2 to 3 and 4 to 1 increases, causing a change to Configuration 2 to provide a direct connection for these flows, allowing them to transfer at near full link capacity. Finally at 60 seconds, these 4 flows end their data transfer, and we allow a new flow of traffic from Servers 1 to 3 and 2 to 4 to increase, causing a change to Configuration 1. Our system design proves that our control plane algorithm operates as expected. We also demonstrate that our SiP MRR device is able to switch and transfer data reliably at the full 10G capacity.

## 5. Conclusion

We introduce a network architecture that merges the spatial and wavelength switching capabilities of an MRR device with a self-adaptive control plane for bandwidth steering. Results from our simulations and experimental testbed demonstrate that our network architecture is seamlessly integratable into conventional datacom networks, and also show the benefits of its autonomous bandwidth steering capabilities to meet the performance demands of next generation datacom applications.

## References

1. J. M. Estaran *et al.*, "Cloud-BOSS Intra-Data Center Network: on-Demand QoS Guarantees via  $\mu$ s Optical Slot Switching", in ECOC 2017
2. A. Gazman *et al.*, "Software-defined control-plane for wavelength selective unicast and multicast of optical data in a silicon photonic platform," Opt. Exp. Vol. 25, p. 232-242 (2017).
3. K. Wen *et al.*, "Flexfly: Enabling a Reconfigurable Dragonfly through Silicon Photonics," Proc. SC, No. 15 (2016).
4. R. Aguinaldo *et al.*, "Wideband silicon-photonic thermo-optic switch in a wavelength-division multiplexed ring network," Opt. Exp. Vol. 22, p.8205-8218 (2014).
5. K. Ishii *et al.*, "Toward exa-scale optical circuit switch interconnect networks for future datacenter/HPC," Proc. SPIE Vol. 10131 (2017).
6. D.Nikolova *et al.*, "Scaling silicon photonic switch fabrics for data center interconnection networks," Opt. Express 23, 1159-1175 (2015)
7. Y. Shen *et al.*, "Software-Defined Networking Control Plane for Seamless Integration of Silicon Photonics in Datacom Networks", in ECOC 2017