Towards a Centralized Controller for Silicon Photonic MZI-based Interconnects

Fei Lou[1], Monireh Moayedi Pour Fard[1], Peicheng Liao[1], Md Shafiq Hai[1], Rubana Priti[1], Yourui Huangfu[2], Chen Qiu[2], Qinfen Hao[2], Zhulin Wei[2], Odile Liboiron-Ladouceur[1]

[1]Dept. of Electrical and Computer Eng., McGill University, Montreal, Canada; [2]Huawei Technologies Co., Ltd, Hangzhou, China;
Email: fei.lou@mcgill.ca

Abstract: An FPGA-based centralized controller for a low-port 4×4 SiP multi-stage MZI-based switch is experimentally investigated with 10 Gb/s payload data. Packet contention resolutions and scheduling latencies are studied.

OCIS codes: (200.4650) Optical interconnects; (250.5300) Photonic integrated circuits; (040.6040) Silicon.

1. Introduction

With the bandwidth scaling in datacenters and high performance computers, traditional copper-based electrical interconnects are reaching the physical limits in aspects of power consumption and cost per bit. Optical interconnects provide large bandwidth-distance product, relatively more transparent networking, and potentially low power consumption, which may enable a solution to the challenges [1]. Particularly, CMOS-compatible Silicon photonics (SiP) is of interest for its lower cost and dense integration opportunities. Moreover, the capability of hybrid integrations of Silicon photonics with III-V active photonics as well as the CMOS drivers and the controlling ASICs promise versatile functionalities in future high-speed networking systems [2]. In the past few decades, the architecture for the datacenter has been developed from tight-coupled silos to computer virtualization enhancing data processing capability. To further improve the throughput and resource utilization, a high throughput computing (HTC) architecture for the datacenter, which is based on resource disaggregation and interface-unified interconnects, has been proposed **Error! Reference source not found.**. Optical interconnects with higher bandwidth density and longer reach will play an important role in future disaggregated HTC platform. For dynamically reconfiguring the resources in data center, large port count and nano-seconds level latency optical switching system needs to be employed to reconfigure the optical interconnection.

In this work, we demonstrate a prototype of an FPGA-based centralized controller of a 4×4 SiP Mach-Zehnder interferometer (MZI) multi-stage switch. The optical payload has a data speed of 10 Gb/s and the FPGA-based controller operates at a 100 Mb/s clock. Beyond the switch, the work integrates the controller to assess the overall packet switching performances, *i.e.* extinction ratios, contention resolutions and network latencies.

2. The 4×4 MZI-based Switch with Centralized Controller

Figure 1(a) shows the schematic diagram of the experimental demonstrator that includes the FPGA-based centralized switch controller and the transmitter (TX) and receiver (RX) nodes. The SiP chip is a 4-port multi-stage switch configured as Beneš topology with six 2×2 MZIs directly controlled by the centralized switch controller. Specifically geared towards rapid and efficient switching, carrier injection is employed as the tuning mechanism by biasing one of the MZI's arms instead of thermo-optics or carrier depletion approaches [4]. Fabricated by the IME foundry, the MZI features a $V\pi \cdot L$ of ~ 0.26 V·mm and a switching time of ~ 6 ns [5]. The flow chart in Figure 1(b) shows the decision making process of the controller. Namely, TXs firstly send link requests (LinkReq) to the controller, which then checks the availability of the requested optical paths. In case no path is available, the data packets at TXs wait in queue. In case of a granted request, the controller sets the multiple MZIs to the desired states and an acknowledgement signal (Ack) is fed back to the TXs enabling the optical packet generations. The packet of 10 Gb/s payload then transmits through the SiP switch fabric via the grating coupler interfaces, and converted into electrical signals at the RX side. At the end of the packet transmission, the controller releases the previously set optical paths and updates the network status.

As a proof of concept, the demonstrator considers two transmitter nodes (TX1 and TX3) which are injecting packets, and two destination nodes (RX1 and RX2), as shown in Figure 1(a). Experimentally, the optical packets generated by TX1 are injected by gating a semiconductor optical amplifier (SOA) while TX3 is not injected due to component limitations. At the destination node RX, the signal quality of the output packets is monitored to verify the switching performances. Table 1 lists the studied traffic (column) injected by TX1 (blue row) and TX3 (grey row), where TX1 and TX3 are denoted by [0000] and [0010] and RX1 and RX2 by [0100] and [0101], respectively. In certain instance, both TX1 and TX2 request to send to a same RX node, resulting in network contentions. When contention arises, TX3 has higher priority such that TX3's request is granted while TX1 waits in queue.

3. Results and Discussion

The timing diagrams are shown in Figure 2, where the injected traffic is the same as listed in Table 1, and the MZI's states [00], [10] and [11] denote idle, bar and cross state of the MZI, respectively. The packets have 1200 bits with time duration of 120 ns, *i.e.* 12 clock cycles of the FGPA-based controller. The minimal inter-packet gap is 80 ns (8 cycles) due to the setup and hold times of the MZIs, the controller delay in assessing the network status, and feeding back signals. As the main latency contributor is in the decision process, transmitting long packets into the system becomes efficient in terms of latencies and throughputs. From Figure 2 in the case of a contention where the RX destination node of both TX1 and TX3 is the same, the packet from TX1 is delayed until TX3 finishes its concurrent transmission and the network contention is resolved. The eye diagrams of the optical packets at TX1, RX1 and RX2 are shown in Figure 3, with different signal level due to path dependent insertion loss, but clear eye opening. The BER measurement of the SiP switch was reported in [5].

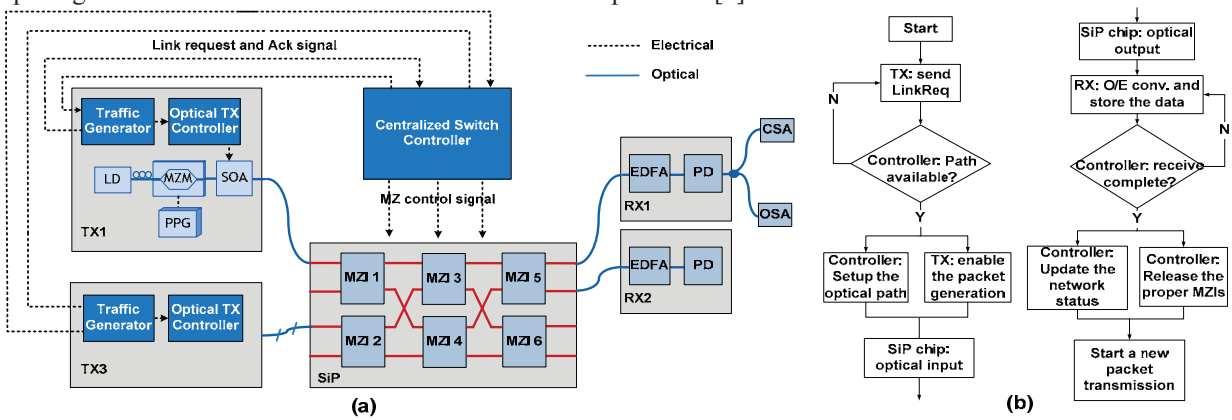

Figure 1. (a) Schematics of the FPGA-based switch controller. (b) Flow chart of the controller.

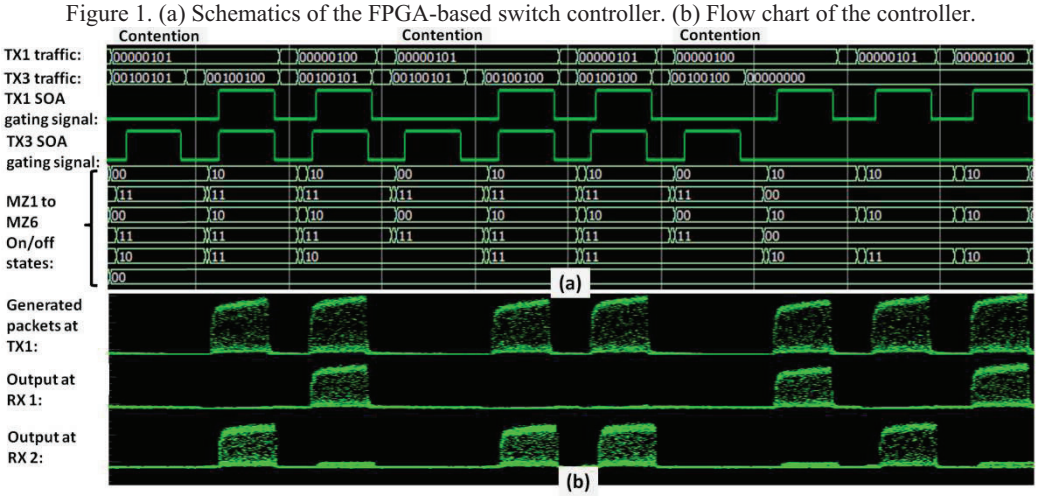

Figure 2. Timing diagrams. (a) Main electrical signals of the switch controller. (b) The generated optical packets at TX1 and the received optical packets at RX1 and RX2. The payload speed is 10 Gb/s.

Table 1. The traffic injected by TX1 and TX3.

TX1	0000 0101	0000 0100	0000 0101	0000 0101	0000 0100	0000 0101	0000 0100
	RX2	RX1	RX2	RX2	RX1	RX2	RX1
TX3	0010 0101	0010 0100	0010 0101	0010 0101	0010 0100	0010 0100	0010 0100
	RX2	RX1	RX2	RX2	RX1	RX1	RX1

Figure 3. The eye diagrams of the optical packets at TX1, RX1 and RX2. The datarate is 10 Gb/s.

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) training program in Silicon Photonics (Si-EPIC) and the Canadian Research Chairs Program. The support from CMC Microsystems for fabrication is also appreciated.

[1] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips", Proc. IEEE, vol. 88, pp.728-749, 2000.
[2] A.V. Krishnamoorthy, "Computer Systems Based on Silicon Photonic Interconnects", Proc IEEE, vol. 97, no. 7, 2009.
[3] "High Throughput Computing Data Center Architecture," www.huawei.com/ilink/en/download/HW_34960, Jun. 2014 [Jan 30, 2015].
[4] R. Soref, "The Past, Present, and Future of Silicon Photonics", IEEE JSTQE, vol. 12, no. 6, pp.1678-1687, 2006.
[5] M. S. Hai, *et al.,* "MZI-based Non-blocking SOI Switches", at ACP, paper ATh3A.147, 2014.