Loopback Interconnect Datacenter Optical Network for Intra- and Inter-Pod Switching in Datacenters

Yifeng Gao, Gordon Ning Liu* and Gangxiang Shen
School of Electronic and Information Engineering
Soochow University
Suzhou, China
*gordonnliu@suda.edu.cn.

Xiangyong Hao, Fuhan Wang
Hengtong Optic-Electronic Co., Ltd
Suzhou, China
haoxy@htgd.com.cn, wangph@htgd.com.cn

***Abstract*—A loopback interconnect datacenter optical network is proposed for Intra- and Inter-Pod switching with commercial devices. Its feasibility was demonstrated by experiments since the filtering performance is similar at loopback and common ports.**

Keywords—Datacenter network, intra- and inter-Pod switching, wavelength selective switch

I. INTRODUCTION

The emerging services such as 5G and Internet of Things require low latency interconnects between terminals and datacenters [1], which drives the traffic growth in datacenters (DCs). Existing electrically switched datacenter networks, such as the spine leaf architecture created by commodity Ethernet switches and routers, have introduced exorbitant rise in cost and power consumption for these massively increasing bandwidth demands [2]. The electronics bottleneck limits the further developments of the datacenters. Due to the features of low latency, high bandwidth, and low power consumption, employing optical switching in DCs has attracted wide attention recently. The optical datacenter network for nanosecond switching is proposed, while being able to solve the contention problems and implement the nanosecond recovery of the optical data packets [3]. A flat datacenter network, Sirius, uses a combination of tunable lasers and passive gratings that route light based on its wavelength to achieve a scalable network that can offer high bandwidth and very low end-to-end latency [4]. Also, wavelength-division-multiplexing (WDM) and high-radix switching can scale to provide ~Tb/s/link transmission capacity with potential interconnectivity for multiple nodes [5-7]. While there have been considerate efforts in designing innovative optical switching architecture, key challenges remain on how to deploy kinds of optical switches within the datacenter network cost-efficiently. An improved optical switching architecture, optical tunnel network system (OPTUNS), has been proposed with more reasonable cost [8]. It consists of arrayed waveguide grating (AWG) and wavelength selective switch (WSS) based optical add/drop subsystem (OADS) modules, and N×M WSS based optical switch interconnect subsystem (OSIS) modules. The architecture enables optical intra- and inter-Pod switching in DCs with a low power consumption and low latency. Combining with the software defined network (SDN), it achieves high wavelength utilization and good tolerance for errors.

In this paper, we proposed an innovative optical intra- and inter-Pod switching architecture in DCs: Loopback Interconnect Datacenter Optical Network (LIDON). Due to its modularized, flexibly scalable and path diversified architecture, LIDON fits demands of DCs very well, sharing same advantages of low latency, low power consumption, good scalability, high wavelength utilization, and robust fault tolerance with OPTUNS. Moreover, compared with OPTUNS, LIDON is realized with just commercial devices utilized in reconfigurable optical add-drop multiplexer (ROADM) networks, instead of any specialized N×M WSS, making LIDON cost-efficient and easy to achieve. For some architectures, signals of intra- and inter-Pod switching need to be operated through different devices and topologies, introducing additional device requirements and increased power consumption. While in this architecture, we exploit the additional port of a new type of WSS [9] to conduct the intra- and inter-Pod switching simultaneously with the wavelength selection restriction. Furthermore, LIDON supports flexible grid since the fixed AWGs are not required. Unlike applying dedicated expensive devices in the previous architectures, LIDON is able to act against device failures with commercial devices. The features will be described in details in Section 2. Then, in Section 3, we will show the feasibility of LIDON by testing specifications of the new type of commercial WSS employed in LIDON.

II. ARCHITECTURE OF LIDON

The proposed LIDON architecture at one Pod (Pod0) is showed in Fig. 1. Each Pod in LIDON consists of two types of optical switching modules: the Rack Expansion Module (REM) and the Loopback Interconnect Module (LIM). Switching among servers within one rack is conducted electrically via a Top of Rack (ToR) switch [10]. Then the wavelength specific or tunable transceivers connect the ToR switch to a REM. Due to the modularity of the REM, the rack numbers within the Pod can be expanded smoothly according to the traffic demand as the amount of REMs increases. REM is based on a conventional multicast switch (MCS). Therefore, a REM can connect a ToR switch to several LIMs contentionlessly, i.e., different transceivers on the ToR switch can use a same wavelength to connect REM without the restriction of wavelength collision in a fiber. The other module, LIM, is based on one commercial twin 2×N WSS. Compared to the conventional 1×N WSS, 2×N WSS is a new type of WSS and its 2-port side has a loopback port except a traditional common port. The functionality of loopback port is almost same as the one of common port. Any wavelengths within the loopback port can be selectively switched to any ports at N-port side of 2×N WSS, and vice versa. The difference is that its filtering performance is not optimized since it is originally designed for monitoring purpose only. Additionally, the loopback port and common port cannot transmit or receive signals with the same wavelength simultaneously.

This work is partially supported by the National Natural Science Foundation of China (62271340).

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

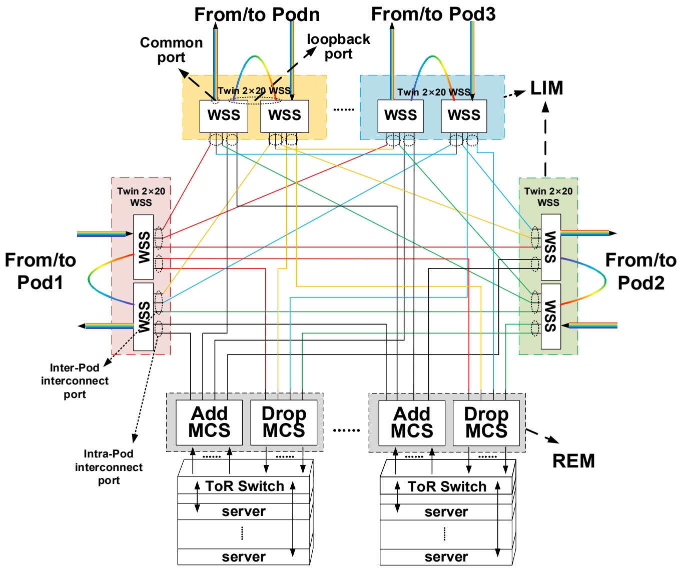

Fig. 1. The architecture of LIDON at Pod0.

In LIDON, we explored a new application of loopback ports. As shown in Fig. 1, for the 2×N WSS of each LIM, ports at N-port side are divided into intra-Pod interconnect ports and inter-Pod interconnect ports, and individually connected to different REMs and other LIMs. Note that the port number N of WSS within one LIM limits the total number of REMs and other LIMs which can be connected simultaneously. The capacity of a Pod and total Pod quantity within one datacenter are related to REM and LIM numbers, respectively. With the development of the optical switch port count scale, the volume of LIDON will be further improved.

For the intra-Pod interconnection, the uplink side of REM routes WDM signals emitted from a corresponding ToR switch to different LIMs. By connecting two loopback ports within a LIM, this LIM forwards the intra-Pod switching signals from one REM to another REM within the same Pod. The receiving REM routes WDM signals to different transceivers of the receiving ToR switch. When there are multiple Pods within the datacenter, a LIM is also responsible for receiving or sending inter-Pod switching signals from one Pod to another at the same time through the common ports of 2×N WSS. Consequently, such an architecture supports interconnections between any two ToR switches within one Pod or among different Pods, and achieves intra- and inter-Pod switching optically. Like OPTUNS, LIDON has wider bandwidth and lower power consumption than electrical switching schemes. Also, multiple generations of optical switches can be deployed in LIDON for the forward compatibility of optical switching, enhancing its cost and performance advantages. Moreover, LIDON is based on the commercial 2×N WSS and MCS, not the AWG and specialized N×M WSS. That makes LIDON a cost-effective architecture. Besides, an EDFA array used in conjunction with the MCS ensures that no other optical amplifier setup is required inside the node. As all the switching devices of LIDON support the flexible grid, this architecture is compatible with its contained transponders to transport with different bandwidths. Higher channel data rates and greater spectral efficiency requirements can be adapted through a higher modulation format and variable bandwidth.

The entire datacenter consists of several Pods connected by LIMs with different topologies. As one of the topologies, the full-mesh connected topology provides a dedicated link between any Pod pairs, creating wide bandwidth and low latency for inter-Pod switching. The parallel channels of full-mesh interconnect allow the switching between different sources and destinations to use the same wavelength without wavelength competition, improving wavelength utilization. The WDM signals from different Pods are transmitted on different links reliably, reducing wavelength contention during multipoint communication. Furthermore, the reliability is also identified by the fact that every LIM provides the loopback functionality independently. Since LIM is a key component of intra- and inter-Pod switching, we need to ensure that the failure of any one device has limited impact on the whole network. When a LIM is out of work, the intra-Pod switching can be performed through other LIMs. And the related inter-Pod switching can be carried out through a relay node. Meanwhile, the hop number is limited to one, which mitigates the passband narrowing due to the multiple cascaded WSS, and then reduces the filtering penalty [11]. On the other hand, different Pods can be connected by partial mesh or ring topology for cost-effectiveness. For example, in a datacenter with ring topology, a Pod only needs two LIMs connected to Pod1 and Pod2, as depicted in the red and green parts of Fig. 1. The required LIM quantities and port counts of REM/LIM used in each node are decreased, thus significantly reducing the overall cost. But it also increases the impact of wavelength contention and device failures as a result.

III. Experimental Investigations

For verifying the performance of WSS loopback ports meet the intra-Pod switching requirement, we conducted an experiment to compare the filtering performance between loopback ports and common ports of a commercial 2×20 WSS [9]. By launching a wideband Amplified Spontaneous Emissions (ASE) source into the common/loopback port, we measured the output spectrum at a port of 20-port side with an optical spectrum analyzer (OSA). Comparing this spectrum with the measured ASE source spectrum, we can calculate the transfer functions of WSS. In this way, physical properties such as bandwidth and insertion loss can be obtained.

In Fig. 2, the transfer functions have been shown while the passing channel is centered at 192.15THz with 50 GHz bandwidth. The transfer functions measured at the loopback and common ports are plotted by blue and red solid curves, respectively. For comparing the intra- and inter-Pod interconnect performance and simulating the transmission of optical signals from a relay node, the transfer functions of two cascaded WSS are also plotted by the dash curves. We observed that the bandwidth through the loopback port is almost the same as that through the common port. The 3dB bandwidth of single-stage loopback port filtering is 47.09GHz, while the one of two-stage filtering is compressed to 44.04GHz. Test results at other channels do not show any significant difference between the loopback and common ports. In addition, the insertion loss (IL) of loopback port is slightly larger by 0.37 dB compared to the one of common port. But the insertion loss ripple is not significantly different between the two ports. The port performance differs depending on its port location. While ports physically located in middle of WSS have better performance, those on two ends show worse performance. But all of them meet the expected specifications.

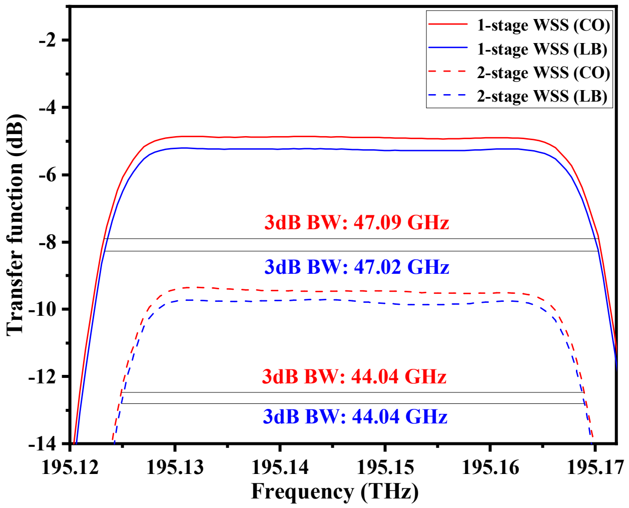

Fig. 2. Transfer functions of the loopback (LB) port and common (CO) port.

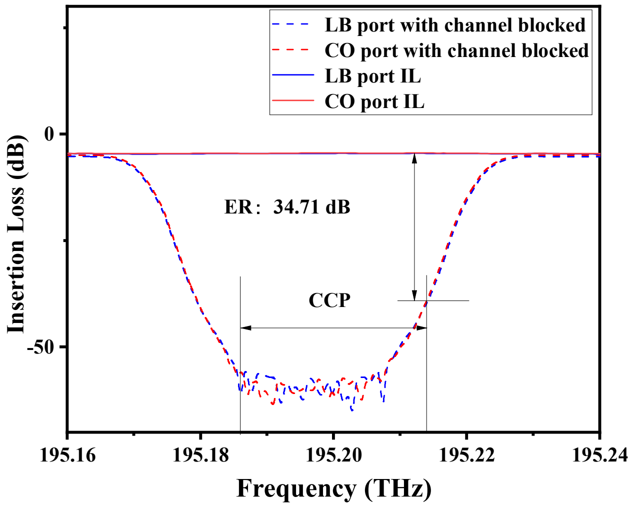

Fig. 3. The channel extinction ratio of the loopback (LB) port and common (CO) port.

Moreover, the channel extinction ratio (ER) is measured at the loopback and common ports and plotted in Fig. 3. Here, ER is defined as the ratio between the channel insertion loss and the maximum leakage power on the evaluated port within the clear channel passband (CCP) when the channel turns on and off. CCP is ±14GHz according to the WSS specifications. According to the test results, the ER of the loopback port calculated through the blue solid and dash curves is almost same as that of the common port calculated through the red solid and dash curves. Both ERs are about 34.71dB. Finally, we measured the wavelength dependent loss (WDL) using tunable laser source and optical power meter. Fig. 4 has shown the insertion loss of the loopback and common ports by blue and red curves within the entire C-band. We observed a WDL difference of about 0.15 dB between the loopback port and common port. Overall, it seems that the loopback port is not optimized, especially in terms of insertion loss, since it is originally designed for monitoring purpose. But the bandwidth of the loopback port is almost same as the one of common port. Such a small performance difference in bandwidth, insertion loss and ER, between loopback ports and common ports of 2×20 WSS will not affect the optical intra- and inter-Pod switching and the 2×20 WSS meets the demand of our proposed LIDON architecture.

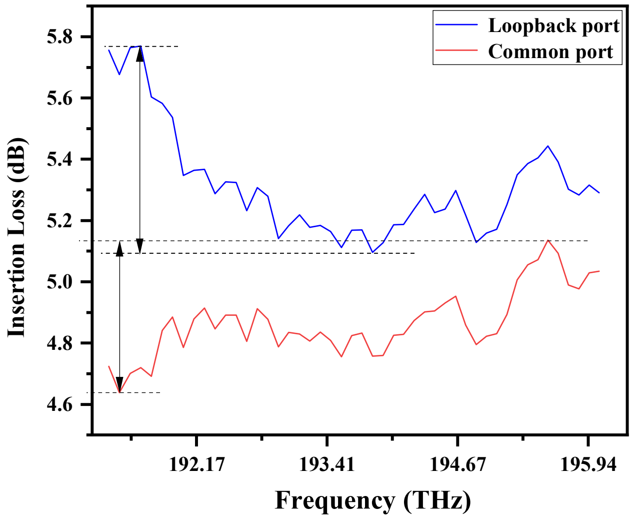

Fig. 4. The wavelength dependent loss of the loopback (LB) port and common (CO) port within the entire C-band.

IV. Conclusions

We proposed a new datacenter architecture, LIDON, with the intra- and inter-Pod optical switching. Based on commercial devices, LIDON has features of low cost, good scalability, high wavelength utilization, and flexible grid. The feasibility was demonstrated by a WSS experiment that the performance of loopback ports has little difference from common ones.

References

[1] W. Shi et al., “Edge computing: vision and challenges,” IEEE Internet Things J., vol. 3, no. 5, pp. 637-646, Oct. 2016.

[2] K. -i. Sato, "Design and Performance of Large Port Count Optical Switches for Intra Data Centre Application," in ICTON, Bari, Italy, 2020, pp. 1-4.

[3] X. Xue et al. "Nanosecond optical switching and control system for data center networks," Nat Commun 13, 2257 (2022).

[4] B. Hitesh et al., "Sirius: A Flat Datacenter Network with Nanosecond Optical Switching." SIGCOMM '20, pp. 782-797.

[5] Y. Mori et al., "High-port-count optical circuit switches for intra-datacenter networks [Invited Tutorial]," J. Opt. Commun. Netw., vol. 13, no. 8, pp. D43-D52, August 2021.

[6] M. Xu et al., "A Hierarchical WDM-Based Scalable Data Center Network Architecture," in ICC, Shanghai, China, 2019, pp. 1-7.

[7] L. Poutievski et al. "Jupiter evolving: transforming google's datacenter network via optical circuit switches and software-defined networking. " SIGCOMM '22, New York, NY, USA,2022, pp.66–85.

[8] M. Yuang *et al.*, “OPTUNS: optical edge datacenter network architecture and prototype testbed for supporting 5G,” J. Opt. Commun. Netw., vol. 12, no. 1, pp. A28-A37, January 2020.

[9] https://www.molex.com/molex/products/family/wavelength_selective_switches

[10] *Distributed Cloud Computing and its Impact on the Cabling Infrastructure within a Data Center*, Cisco Systems, Inc., San Jose, CA, USA, 2021.

[11] M. Filer et al., “N-degree ROADM architecture comparison: broadcast-and-select versus route-and-select in 120 Gb/s DP-QPSK transmission systems,” in OFC, San Francisco, CA, USA, 2014, pp. 1-3.