Highly Reliable and Large-Scale Optical Circuit Switch for Intra-Datacentre Networks

Takumi Mitsuya(1), Takuro Ochiai(1), Takuma Kuno(1), Yojiro Mori(1), Hiroshi Hasegawa(1), Ken-ichi Sato(2)

(1) Nagoya University, mitsuya.takumi.a9@s.mail.nagoya-u.ac.jp
(2) The National Institute of Advanced Industrial Science and Technology (AIST)

Abstract *We propose a novel optical circuit switch architecture offering high reliability and high capacity. The proposed scheme substantially reduces the annual downtime of the switch with little additional hardware cost. Its transmission performance is experimentally confirmed by constructing part of a 1,536×1,536 optical switch. ©2022 The Authors*

Introduction

With the proliferation of cloud-based services, intra-datacentre traffic is rapidly increasing. In typical datacentres, electrical switches and optical fibres are adopted to connect top-of-rack switches [1]. The use of electrical switching demands a large number of power-consuming optical-to-electrical and electrical-to-optical convertors. To solve this problem, the hybrid switching network architecture employing both optical and electrical switches is being extensively studied [2-13]. In such switching networks, large-scale optical circuit switches are essential for offloading signals from power-consuming electrical switches. Many types of optical circuit switches have already been reported [14]. The combination of delivery-and-coupling (DC) switches and wavelength-routing (WR) switches is a promising candidate because of its scalability and cost-effectiveness [5]. However, erbium-doped fibre amplifiers (EDFAs) used in the WR-switch part are likely to fail; failure of a single EDFA makes multiple switch ports unavailable. Therefore, designs that address failure are necessary to construct reliable intra-datacentre networks.

In this paper, we propose a highly reliable optical circuit switch architecture, where redundancy is introduced with no additional system loss. Numerical simulations confirm that the annual downtime caused by EDFA failure is reduced to 0.0014% while the number of EDFAs per port is increased by only 0.01. To prove the feasibility of the proposed architecture, we experimentally construct part of a 1,536×1,536 optical switch.

Proposed Optical Circuit Switch Architecture

Basic Switch Structure

Figure 1 illustrates the basic switch structure that combines multiple DC switches and WR switches [15]. The overall switch system comprises MN transmitters supporting N wavelengths, N M×M DC switches, M N×N WR switches, and MN receivers, where M denotes the port count of the DC switch and N represents the port count of the WR switch. The M×M DC switch comprises M 1×M selectors and M M×1 couplers, whereas the N×N WR switch comprises n (N/n)×1 couplers, n EDFAs, n 1×2 splitters, and a pair of n×(N/2) uniform-loss and cyclic-frequency arrayed waveguide gratings (ULCF AWGs) whose passbands are interleaved [15,16]. Here, n is a design parameter that determines the system loss. The degree of the aggregation coupler is

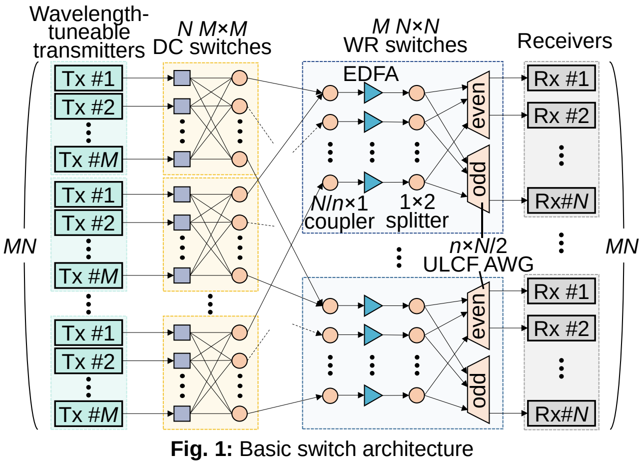

Fig. 1: Basic switch architecture

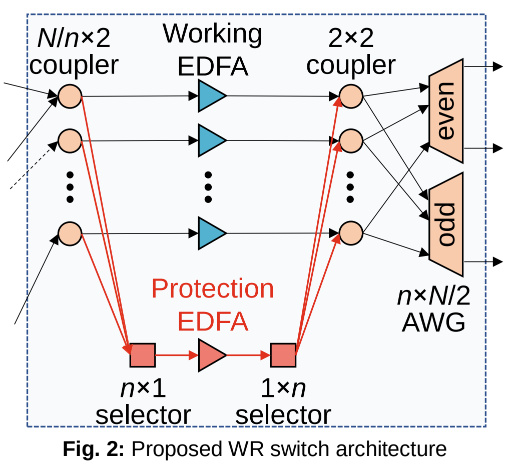

Fig. 2: Proposed WR switch architecture

reduced by using AWGs with *n* inputs, i.e., the loss of the WR-switch part is reduced by a factor of 1/*n*.

The switching process is as follows: a target signal is created by a transmitter; the wavelength is determined to suit the target output port. The signal is then delivered by the *M*×*M* DC switch and directed to the WR-switch part. Multiple signals having different wavelengths are aggregated by an (*N*/*n*)×1 coupler. After amplification with an EDFA, the signals are routed by the combination of 1×2 splitter and paired *n*×(*N*/2) ULCF AWGs. The target signal is finally recovered by the target receiver.

Note that the combination of *n* 1×2 splitters and a pair of interleaved *n*×(*N*/2) ULCF AWGs can be replaced by a single *n*×*N* ULCF AWG. However, the use of interleaved ULCF AWGs avoids the spectrum narrowing caused by AWG traversal [17].

Reliable Architecture

EDFAs used in the WR-switch part are likely to fail, and the failure of an EDFA makes *N*/*n* ports unavailable. To solve this problem, we modify the WR switch structure as shown in Fig. 2. A protection EDFA can substitute for any one of the *n* working EDFAs via couplers and selectors.

The protection process is as follows: a signal input to the WR switch is broadcast to the working port and protection port by an (*N*/*n*)×2 coupler. In the protection part, an *n*×1 selector selects the coupler connected to the failed EDFA. After amplification by the protection EDFA, the signal is delivered to the 2×2 coupler connected with the failed EDFA by a 1×*n* selector. Note that replacing (*N*/*n*)×1 couplers and 1×2 splitters with (*N*/*n*)×2 couplers and 2×2 couplers can completely prevent an increase in system loss. As a result, redundancy is realized with no additional system loss.

Simulations

To confirm the effectiveness of the proposed switch architecture, we evaluate the annual downtime caused by EDFA failure and hardware requirements. Let MTBF be mean time between failures and MTTR be mean time to repair. EDFA availability is defined as MTBF/(MTBF+MTTR). Referring to [18,19], we set MTBF=1,000,000 hours and MTTR=4 hours.

Figure 3 shows the annual downtime of the overall switch system, where *N*=96, *n*=6, and *M* is a parameter. The downtime is reduced to 0.0014% compared to the basic switch structure. Figure 4 shows the per-port number of necessary protection EDFAs; only 0.01 additional EDFAs are necessary for each port. Figure 5 shows the per-port number of 1×2 selectors needed for introducing protection, where we assume a 1×*n* selector is composed of (*n*-1) 1×2 selectors. The proposed switch architecture needs 0.1 additional 1×2 selectors for each port. It is noteworthy that the additional per-port hardware costs do not increase with switch port count as seen in Fig. 4 and Fig. 5. Therefore, our architecture offers scalability against the switch port count.

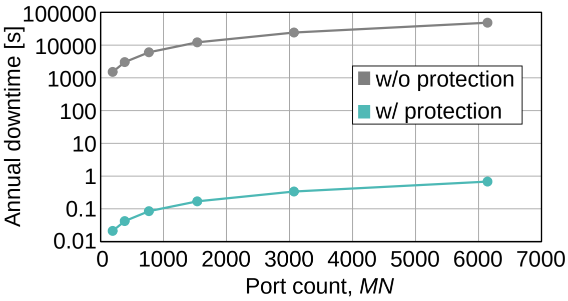

Fig. 3: The annual downtime vs. switch port count (*N*=96, *n*=6, and *M* is a parameter)

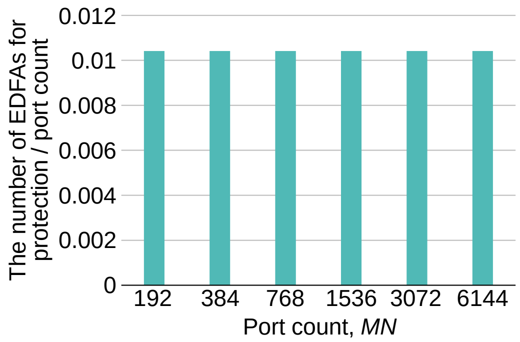

Fig. 4: The per-port number of necessary protection EDFAs vs. switch port count (*N*=96, *n*=6, and *M* is a parameter)

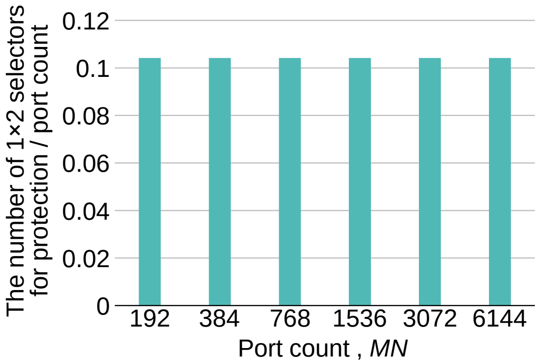

Fig. 5: The per-port number of 1×2 selectors needed for introducing protection vs. switch port count (*N*=96, *n*=6, and *M* is a parameter)

Experiments

We performed experiments to verify the feasibility of our proposed switch architecture. Figure 6 depicts the setup for the experiments. A continuous wave (CW) was output from a tuneable laser and then modulated by a lithium-niobate IQ modulator. The IQ modulator was driven by a waveform generator to form a 32 Gbaud 16QAM signal. Polarization division multiplexing (PDM) was then executed by a PDM emulator. After that, the signal power was adjusted by an EDFA followed by a variable optical attenuator (VOA). The signal thus obtained was split by a 1×2 splitter; one was used as the target signal while the other was used as the intra-band crosstalk. Power of each tributary was set to 2 dBm. The signal was then delivered to our switch for testing. We constructed part of a 1,536×1,536 optical circuit switch, where the DC-switch scale *M* was 16 and the WR-switch scale *N* was 96; the design parameter *n* was set to 6 considering hardware cost and transmission performance. In the switch system, the signal first traversed the DC-switch part consisting of a 1×16 selector and a 16×1 coupler, where fibre delay lines were inserted to emulate intra-band crosstalk. The following WR-switch part consisted of a 16×2 coupler, a working EDFA, a 2×2 couplers, and a pair of 6×48 ULCF AWGs; a 6×1 selector, a protection EDFA, and a 1×6 selector were additionally introduced to realize EDFA protection. In the WR-switch part, one of the output signals from the 16×2 coupler was delivered to the working port and the other was delivered to the protection port. To emulate inter-band crosstalk, 95 non-target signals on different wavelengths were added using the 16×2 aggregation coupler; consequently, 96 wavelengths were multiplexed. Figure 7 shows the signal spectra input to the EDFA. Additionally, 5 signals were incident on the residual input ports of the ULCF AWG to emulate intra-band crosstalk with the same wavelength as the target signal. Finally, the target signal was recovered by a typical digital coherent receiver [20].

Figure 8 illustrates constellation maps of the recovered 16QAM signal and measured BERs for both working port and protection port. We can observe that symbol states of the recovered signals are distinctly separated; the working port and protection port show almost the same performance. We verified that all BERs attained the target value of 10^{-2} for both the working and protection ports.

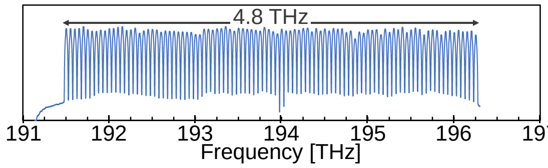

Fig. 7: The signal spectra input to the EDFA

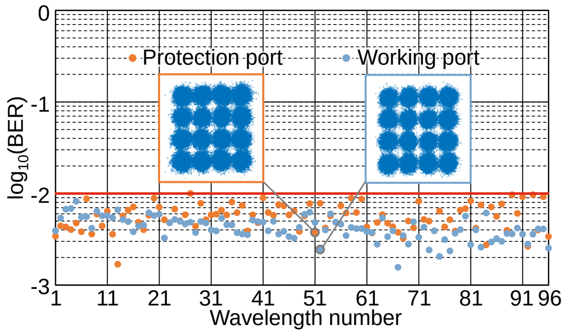

Fig. 8: The constellation maps and BERs

Conclusion

We proposed a highly reliable and large-scale optical switch architecture for intra-datacentre networks. Numerical simulations clarified that the proposed switch architecture significantly improves reliability at little additional cost. To confirm the technical feasibility, we successfully demonstrated part of a 1,536×1,536 optical switch and evaluated BERs of 32-Gbaud DP-16QAM signals, i.e., the net capacity of each port is 200 Gbps. The switch throughput was 393.2 Tbps.

Acknowledgements

This paper is based on results obtained from a project JPNP16007 commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

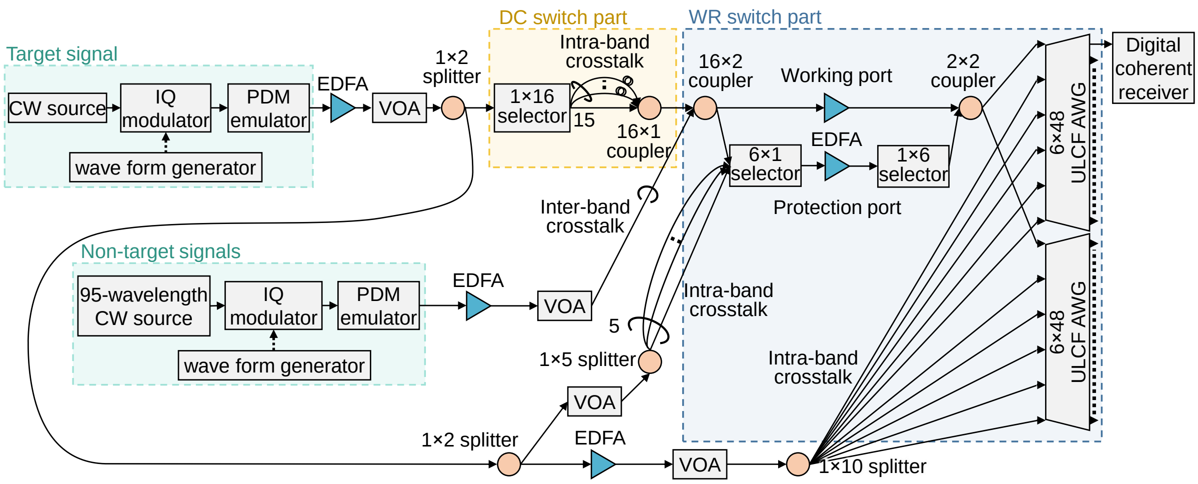

Fig. 6: Experimental setup

References

[1] C. Kachris, K. Bergman, and I. Tomkos, *Optical Interconnects for Future Data Center Networks*, New York, NY: Springer, 2012. DOI: 10.1007/978-1-4614-4630-9

[2] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: a hybrid electrical/optical switch architecture for modular data centers," in *Association for Computing Machinery*, New York, USA, 2010, pp. 339-350. DOI: 10.1145/1851275.1851223

[3] C. Kachris and I. Tomkos, "A survey on optical interconnects for data centers," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 1021-1036, January 2012. DOI: 10.1109/SURV.2011.122111.00069

[4] K. Christodoulopoulos, D. Lugones, K. Katrinis, M. Ruffini, and D. O'Mahony, "Performance evaluation of a hybrid optical/electrical interconnect," *Journal of Optical Communication Networking*, vol. 7, no. 3, pp. 193–204, 2015. DOI: 10.1364/JOCN.7.000193

[5] K. Sato, "Realization and application of large-scale fast optical circuit switch for data center networking," *Journal of Lightwave Technology*, vol. 36, no. 7, pp.1411-1419, 2018. DOI: 10.1109/JLT.2018.2801308

[6] F. Yan, X. Xue, and N. Calabretta, "HiFOST: a scalable and low-latency hybrid data center network architecture based on flow-controlled fast optical switches", *Journal of Optical Communication Networking*, vol. 10, no. 7, pp. 1–14, 2018. DOI: 10.1364/JOCN.10.0000B1

[7] M. Balanici and S. Pachnicke, "Hybrid electro-optical intra-data center networks tailored for different traffic classes," *Journal of Optical Communications and Networking*, vol. 10, no. 11, pp. 889-901, 2018. DOI: 10.1364/JOCN.10.000889

[8] A. Minakhmetov, C. Ware, and L. Iannone, "Data Center's Energy Savings for Data Transport via TCP on Hybrid Optoelectronic Switches," *IEEE Photonics Technology Letters*, vol. 31, no. 8, pp. 631-634, 2019. DOI: 10.1109/LPT.2019.2902980

[9] H. Fang, W. Lu , Q. Li, J. Kong, L. Liang, B. Kong, and Zuqing Zhu, "Predictive analytics based knowledge-defined orchestration in a hybrid optical/electrical datacenter network testbed," *Journal of Lightwave Technology*, vol. 37, no. 19, pp.4921-4934, 2019. DOI: 10.1109/JLT.2019.2925827

[10] X. Xue, F. Nakamura, K. Prifti, B. Pan, F. Yan, F. Wang, X. Guo, H. Tuda, and N. Calabretta, "Experimental assessments of a flexible optical data center network based on integrated wavelength selective switch", presented at *Optical Fiber Communication Conference (OFC)*, San Diego, USA, 2020. DOI: 10.1364/OFC.2020.W1F.5

[11] R. R. Reyes, S. Sultana, V. V. Pai, and T. Bauschert, "Techno-economic assessment of hybrid electrical-optical intra-data centre networks," presented at *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, Bari, Italy, 2020. DOI: 10.1109/ICTON51198.2020.9203166

[12] M. Murakami, H. Kubokawa, K. Sugiura, E. Oki, S. Okamoto, and N. Yamanaka, "Online parameter tuning of the flow classification method in the energy-efficient data center network "HOLST"," *Journal of Optical Communications and Networking*, vol. 12, no. 11, pp. 344-354, 2020. DOI: 10.1364/JOCN.395798

[13] M. Balanici and S. Pachnicke, "Classification and forecasting of real-time server traffic flows employing long short-term memory for hybrid E/O data center networks," *Journal of Optical Communications and Networking*, vol. 13, no. 5, pp. 85-93, 2021, DOI: 10.1364/JOCN.411017

[14] Y. Mori and K. Sato, "High-port-count optical circuit switches for intra-datacenter networks[Invited Tutorial]," *Journal of Optical Communications and Networking*, vol. 13, no. 8, pp. D43-D52, 2021. DOI: 10.1364/JOCN.425929

[15] H. Nagai, Y. Mori, H. Hasegawa, and K. Sato, "Design and verification of large-scale optical circuit switch using ULCF AWGs for datacenter application," *Journal of Optical Communications and Networking*, vol. 10, no. 7, pp. 82-89, 2018. DOI: 10.1364/JOCN.10.000B82

[16] K. Okamoto, T. Hasegawa, O. Ishida, A. Himeno, and Y. Ohmori, "32 × 32 arrayed-waveguide grating multiplexer with uniform loss and cyclic frequency characteristics," *Electronics Letters*, vol. 33, pp. 1865-1866, 1997. DOI: 10.1049/el:19971246

[17] T. Kuno, T. Mitsuya, Y. Mori, H. Hasegawa, and K. Sato, "Demonstration of high-throughput intra-datacenter switches using interleaved AWGs for nyquist WDM," presented at *Optical Fiber Communication Conference (OFC)*, San Diego, USA, 2022. DOI: 10.1364/OFC.2022.Th3A.2

[18] A. Oosenbrug, C. S. Harder, A. Jakubowicz, and P. Roentgen, "Power integrity of 980 nm pump lasers at 200 mW and above," in *Conference Proceedings LEOS'96 9th Annual Meeting IEEE Lasers and Electro-Optics Society*, Boston, USA, 1996, pp. 348-349. DOI: 10.1109/LEOS.1996.565276

[19] OASE, "Technical Assessment and Comparison of Next-Generation Optical Access System Concepts," https://cordis.europa.eu/docs/projects/cnect/5/249025/080/deliverables/001-OASED421WP4UEssex31Oct2011V10.pdf, accessed on 10 May 2022.

[20] M. S. Faruk, and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1125-1141, 2017. DOI: 10.1109/JLT.2017.2662319