# Large-Scale Optical Switch Architectures for Intra-Datacentre Networks

Yojiro Mori[(1)] and Ken-ichi Sato[(2)]

[(1)] Nagoya University, mori@nuee.nagoya-u.ac.jp
[(2)] National Institute of Advanced Industrial Science and Technology (AIST)

**Abstract** The deployment of optical switches inside datacentres will enable substantial reductions in power consumption and the numbers of transceivers and optical link fibres. This paper overviews recent advances in optical-switching architectures for intra-datacentre networks. Several switching technologies and their characteristics are summarized.

## Introduction

Datacentre-related traffic will increase all over the world by more than 20% a year and most of the datacentre-related traffic will terminate inside datacentres. Typical datacentre networks adopt multi-tier electrical switches to interconnect top-of-rack switches mounted on racks containing the computer systems. The multi-tier electrical-switching system offers high fault tolerance; however, such systems are facing critical problems of large power consumption due to broad-bandwidth electrical switches and massive numbers of transponders and optical link fibres[1].

To process the large volumes of traffic in a cost-effective manner, optical switching technologies need to be introduced to future intra-datacentre networks. Among various switching networks, optical and electrical hybrid switching networks will provide the most cost-effective solution[2-7]. In such systems, small or short-live flows are processed by electrical packet switches while large or long-live flows are processed by optical circuit switches. With this scheme, we can offload a part of the traffic volume from power-consuming electrical switches while retaining the desired network functionalities. Eventually, optical switches are expected to replace most electrical switches so as to minimize the power consumption[1]. Realizing such optics-based switching systems demands a high-port-count and fast-changeable optical switch. The use of high-port-count optical switches realizes a flat network in which top-of-rack switches are interconnected using single-hop optical switching, which eliminates the need for multiple-tier electrical switching networks. Furthermore, the fast switching operation allows us to extend kinds of datacentre applications that can be offloaded from the electrical domain. Consequently, the switch power consumption and the numbers of transceivers and optical link fibres can be reduced substantially.

This paper introduces optics-based switching technologies for intra-datacentre networks and discusses their fundamental characteristics. Due to the rapid expansion of this research field, many types of optical switch architectures are being intensively studied[8]–[17]. This paper cannot cover the entire field and hence we will focus on several selected technologies and their characteristics with respect to port count and changing time.

## Optical switch technologies

### *MEMS-based space switch*

Fig. 1(a) illustrates optical switch architectures based on a two-dimensional (2-D) microelectromechanical system (MEMS)[18]. An M×M optical switch based on 2-D MEMS comprises $M^2$ 2×2-switch elements aligned in a crossbar topology with M rows and M columns. Each 2×2-switch element is a bi-stable micro-mirror that can take either bar state or cross state. The attainable maximum port count of 2-D MEMS switches is limited by the fabrication precision of numerous micro-mirrors; as a result, this switch architecture cannot attain high port counts necessary for datacentre applications. Moreover, the switching time is relatively long owing to the mechanical actuation; switching times are typically in the order of milliseconds.

Fig. 1(b) is a schematic of an MxM optical switch using the three-dimensional (3-D) MEMS technology[9,19], where 2M micro-mirrors are typically utilized for beam steering. The input ports and output ports are arranged in two-dimensional arrays. This scheme can realize high-port-count switching; however, the port count and switching time has a trade-off because the number of beam-steering states increases with the square of the switch port count. The largest port count reported so far is 1100[20]. The switching time is around 5 ms when the port count is 256[21]. The switching time of 150 μs was also demonstrated using a 61-port prototype switch [9].

These MEMS-based switches offer broad

bandwidth and low insertion loss; hence, the MEMS-based switch is compatible with most transponder configurations in terms of signal baud, modulation formats, and detection schemes.

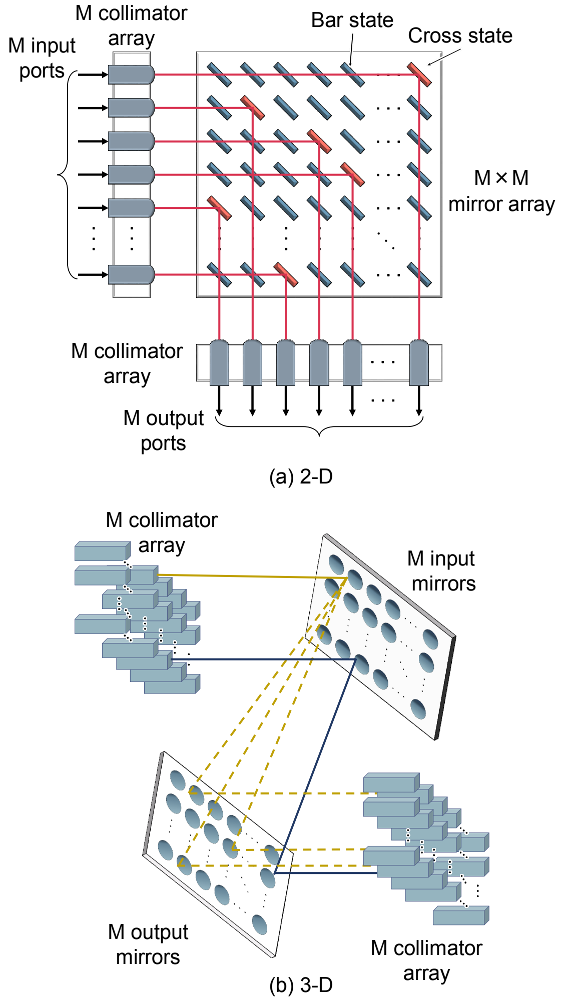


Fig. 1 MEMS-based space switches

*MZI-based space switch*

Fig. 2 depicts an M×M optical switch architecture based on Mach-Zehnder interferometers (MZIs). Here, $M^2$ 2×2 MZIs are arranged to form the path-independent-insertion-loss (PILOSS) topology[8,15]. Each 2×2 MZI can take either bar or cross state. Such MZI-based space switches can be compactly fabricated using silicon-photonics technologies. The thermo-optic phase control attains switching times of around 5 μs[22,23]. With this architecture, a port count of 32 was successfully demonstrated[8,15,23]. The port count can be expanded by bridging multiple switches, e.g. clos networks; however, the insertion loss due to multi-stage fibre-to-chip and chip-to-fibre connections bounds the total switch port count. The MZI-based switch supports broad bandwidth and so can treat most signal forms.

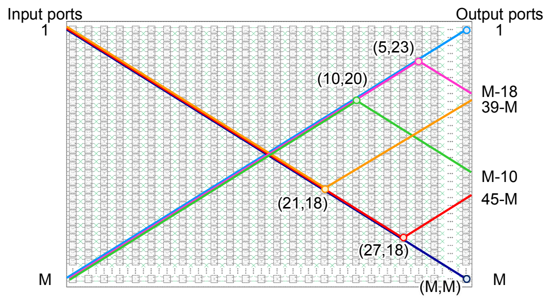


Fig. 2 MZI-based space switch

*Wavelength-routing switch*

Fig. 3(a) illustrates an example of N×N wavelength-routing switch based on tuneable lasers (TLs)[24]. An N×N switch consists of N wavelength-tuneable transponders supporting N wavelengths and an N×N fixed-wavelength router. The fixed-wavelength router can be realized by a single cyclic arrayed-waveguide grating (AWG), multiple cyclic AWGs, or a combination of coupler and non-cyclic AWG[24-26]. The wavelength router directs signals according to the signal wavelengths. By changing transmitter-laser wavelength, an arbitrary output port can be selected. The port count is limited by the available wavelength range of lasers and signal bandwidth. The switching time is determined by the wavelength-tuning time of lasers. For example, a tuning range of 12 nm and a tuning time of less than 130 ns or a tuning range of 55 nm and a tuning time of around 500 μs have been reported[27,28].

Fig. 3(b) shows an example of N×N wavelength-routing switch based on tuneable filters (TFs). The switch comprises N fixed-wavelength transponders and an adaptive wavelength router. The adaptive wavelength router can be constructed by a wavelength multiplexer such as AWG and coupler, a splitter, and tuneable filters. By tuning the passband of the filter, any output port can connect with any input port. The port count of this architecture is limited by the available wavelength ranges of lasers and filters. The switching time is determined by the tuning time of the filter. Silicon-photonic tuneable filters have demonstrated a tuning range of 35 nm and a tuning time of 10 μs[29,30]. If coherent detection is adopted, the local oscillator in the receiver must be a tuneable laser; consequently, the switching time is determined by the wavelength-tuning time of the local oscillator.

In both configurations, optical amplifiers can be utilized as necessary so as to recover the signal power[7]. In such a case, the port count of the wavelength-routing switches can also be limited by the effective bandwidth of the optical amplifier.

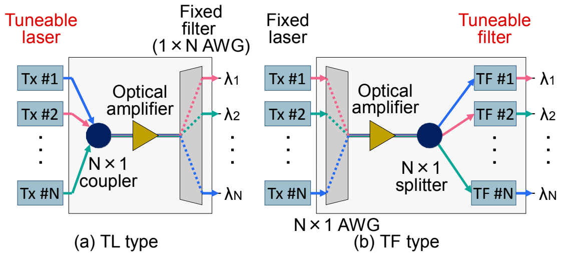


Fig. 3 Examples of wavelength-routing switches

*Combination of space switches and wavelength-routing switches*

Fig. 4 is a schematic of the MN×MN optical switch, where N M×M space switches and M N×N wavelength-routing switches are used as sub-switches. The overall switch port count is given by the product of the two sub-switches, i.e. MN. The switching time is determined by the tuning time of the slowest sub-switch.

Fig. 5(a) depicts an MN×MN optical switch architecture using space switches and TL-based wavelength-routing switches[26]; it includes MN wavelength-tuneable transponders that can treat any one of N wavelengths, N M×M delivery-and-coupling space switches, and M N×N wavelength-routing part; each space switch comprises M 1×M selectors and M M×1 couplers and each aggregation-and-distribution part comprises an N×1 coupler, an optical amplifier (e.g. EDFA), and a 1×N AWG. Here, arbitrary output and input ports can be connected by changing the state of space switch and the wavelength of tuneable laser inside the transponder. With this architecture, the port count of over 1536 was demonstrated[31].

Fig. 5(b) is a schematic of the MN×MN optical switch based on tuneable filters, which is almost mirror symmetrical to the TL-based architecture shown in Fig. 5(a). The switch comprises MN fixed-wavelength transponders, M N×N aggregation-and-distribution parts, N M×M multicast switch, and MN tuneable filters. Here, arbitrary output and input ports can be connected by changing the state of the space switch and the passband of the tuneable filter. If coherent detection is adopted, the local oscillator needs to adopt tuneable lasers. This architecture has attained port counts of 768[29].

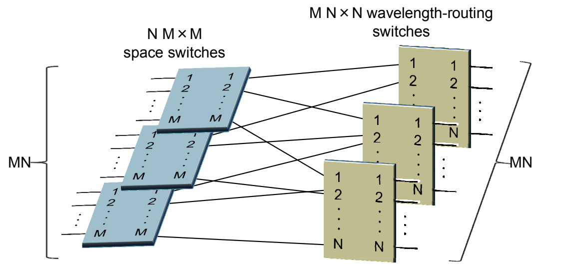


Fig. 4 MN×MN switch using M×M space switches and N×N wavelength-routing switches

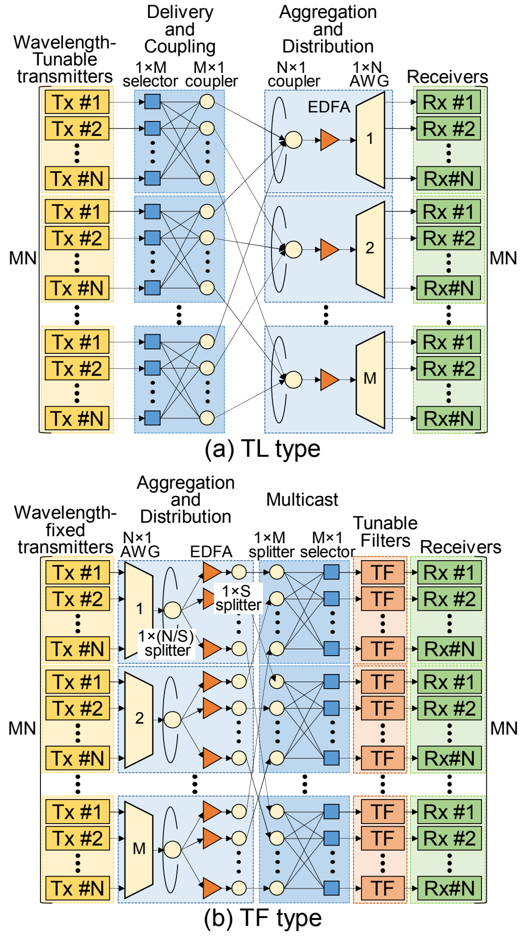


Fig. 5 Optical switch based on space switches and wavelength-routing switches

## Conclusions

The introduction of optics-based switching technologies into datacentres will enable substantial reductions in power consumption, the number of transceivers, and required link fibre resources. This paper overviewed optical-switch candidates for future intra-datacentre applications. The independent use of space switch or wavelength-routing switch may not suitable for inter-rack connections within future hyper-scale datacentres because of their limited switch port counts. The combination of space switches and wavelength-routing switches can easily satisfy the requirement for the port count; however, developing a fast and wide-range tuneable laser is still a challenge, particularly when we apply coherent systems since a wavelength-routing switch relying on coherent technologies will demand wavelength-tuneable lasers. The presentation will discuss architectures and their characteristics in more detail; port count, switching time, hardware cost, and transmission characteristics will be comprehensively evaluated considering appropriate modulation/demodulation schemes.

## References


[1] K. Sato, "Realization and application of large-scale fast optical circuit switch for data center networking", IEEE/OSA J. Lightw. Technol., vol. 36, pp. 1411–1419, Feb. 2018.

[2] N. Farrington et al., "Helios: a hybrid electrical/optical switch architecture for modular data centers", in Proc. ACM SIGCOMM 2010 Conference (SIGCOMM '10), New Delhi, India, Aug. 2010, pp. 339–350.

[3] C. Kachris and I. Tomkos, "A survey on optical interconnects for data centers", IEEE Commun. Surv. Tutorials, vol. 14, pp. 1021–1036, Jan. 2012.

[4] W. M. Mellette et al., "RotorNet: A Scalable, Low-complexity, Optical Datacenter Network", in Proc. ACM SIGCOMM 2017 Conference (SIGCOMM '17), Los Angeles, USA, Aug. 2017, pp. 267–280.

[5] 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", IEEE/OSA J. Opt. Commun. Netw., vol. 10, pp. 1–14, July 2018.

[6] X. Xue et al., "Experimental assessments of a flexible optical data center network based on integrated wavelength selective switch", in Proc. Optical Fiber Communication Conference (OFC), San Diego, USA, Mar. 2020, W1F.5.

[7] K. Sato, "Design and performance of large port count optical switches for intra data centre application", in Proc. 2020 22nd International Conference on Transparent Optical Networks (ICTON), Bari, Italy, Sept. 2020, Tu.C3.2.

[8] K. Tanizawa et al., "Ultra-compact 32 × 32 strictly-non-blocking Si-wire optical switch with fan-out LGA interposer", OSA Opt. Express, vol. 23, pp. 17599–17606, Jun. 2015.

[9] W. M. Mellette et al., "A scalable, partially configurable optical switch for data center networks", IEEE/OSA J. Lightw. Technol., vol. 35, pp. 136–144, Jan. 2017.

[10] Q. Cheng et al., "Demonstration of the feasibility of large-port-count optical switching using a hybrid Mach–Zehnder interferometer–semiconductor optical amplifier switch module in a recirculating loop", OSA Opt. Lett., vol. 39, pp. 5244–5247, Dec. 2014.

[11] A. C. Funnell et al., "Hybrid wavelength switched-TDMA high port count all-optical data centre switch", IEEE/OSA J. Lightw. Technol., vol. 35, pp. 4438–4444, Aug. 2017.

[12] M. Ding et al., "Hybrid MZI-SOA InGaAs/InP photonic integrated switches", IEEE J. Sel. Top. Quantum Electron., vol. 24, pp. 1–8, Oct. 2017.

[13] N. Terzenidis et al., "High-port low-latency optical switch architecture with optical feed-forward buffering for 256-node disaggregated data centers", OSA Opt. Express, vol. 26, pp. 8756–8766, Apr. 2018.

[14] R. Proietti et al., "Experimental demonstration of a 64-port wavelength routing thin-CLOS system for data center switching architectures", IEEE/OSA J. Opt. Commun. Netw., vol. 10, pp. B49–B57, July 2018.

[15] K. Suzuki et al., "Low-insertion-loss and power-efficient 32×32 silicon photonics switch with extremely high-Δ silica PLC connector", IEEE/OSA J. Lightw. Technol., vol. 37, pp. 116–122, Aug. 2018.

[16] Q. Cheng et al., "Ultralow-crosstalk, strictly non-blocking microring-based optical switch", OSA Photonics Res., vol. 7, pp. 155–161, Jan. 2019.

[17] T. J. Seok et al., "Wafer-scale silicon photonic switches beyond die size limit", OSA Optica, vol. 6, pp. 490–494, Apr. 2019.

[18] P. De Dobbelaere et al., "Digital MEMS for optical switching", IEEE Commun. Mag. vol. 40, pp. 88–95, Aug. 2002.

[19] W. M. Mellette and J. E. Ford, "Scaling limits of MEMS beam-steering switches for data center networks", IEEE/OSA J. Lightw. Technol., vol. 33, pp. 3308–3318, Aug. 2015.

[20] J. Kim et al., "1100×1100 port MEMS-based optical crossconnect with 4-dB maximum loss", IEEE Photon. Technol. Lett. vol. 15, pp. 1537-1539, Nov. 2003.

[21] D. T. Neilson et al., "256×256 port optical cross-connect subsystem," IEEE/OSA J. Lightw. Technol., vol. 22, pp. 1499–1509, June 2004.

[22] H. Matsuura et al., "Accelerating switching speed of thermo-optic MZI silicon-photonic switches with "Turbo Pulse" in PWM control", in Proc. Optical Fiber Communication Conference (OFC), Los Angeles, USA, Mar. 2017, W4E.3.

[23] K. Tanizawa et al., "Optical time-slot switching with 5-μs guard time using 32×32 Si TO PILOSS switch", in Proc. European Conference on Optical Communication (ECOC), Dublin, Ireland, Sept. 2019, Th.1.A.

[24] K. Sato et al., "A large-scale wavelength routing optical switch for data center networks", IEEE Commun. Mag. vol. 51, pp. 46-52, Sept. 2013.

[25] K. Ueda et al., "Large-scale optical switch utilizing multistage cyclic arrayed-waveguide gratings for intra-datacenter Interconnection", IEEE Photon. J. vol. 9, #7800412, Feb. 2017.

[26] H. Nagai et al., "Design and verification of large-scale optical circuit switch using ULCF AWGs for datacenter application", IEEE/OSA J. Opt. Commun. Netw., vol. 10, pp. 82–89, July 2018.

[27] R. Maher et al., "Widely tunable burst mode digital coherent receiver with fast reconfiguration time for 112 Gb/s DP-QPSK WDM networks", IEEE/OSA J. Lightw. Technol., vol. 30, pp. 3924–3930, Dec. 2012.

[28] K. Ueda et al., "Demonstration of 1,440x1,440 fast optical circuit switch for datacenter networking," in Proc. Optoelectronics and Communications Conference/International Conference on Photonics in Switching (OECC/PS), Niigata, Japan, July 2016, WF1-3.

[29] Y. Mori, M. E. Ganbold, and K. Sato, "Design and evaluation of optical circuit switches for intra-datacenter networking", IEEE/OSA J. Lightw. Technol., vol. 37, pp. 330–337, Jan. 2019.

[30] H. Matsuura et al., "Fast Frequency Tuning of Silicon-Photonic Thermo-optic MZI Filters using "Turbo Pulse" Method", in Proc. Optical Fiber Communication Conference (OFC), San Diego, USA, Mar. 2018, M4H.2.

[31] E. Honda et al., "High-throughput optical circuit switch for intra-datacenter networks based on spatial super-channels", in Proc. Optical Fiber Communication Conference (OFC), San Diego, USA, Mar. 2020, W1F.2.