

Chapter 3

Optical Switching in Data Centers: Architectures Based on Optical Packet/Burst Switching

Nicola Calabretta and Wang Miao

3.1 Introduction

Data centers (DC) are phenomenally growing in size and complexity, to satisfy the demands for more powerful computational performance driven by the data-intensive applications as well as high-density virtualizations [1]. High-performance and energy-efficient multi-core processors are developed aggressively to provide higher processing capability [2]. Foreseeing the preservation of Moore's law through chip-level parallelism, the multi-core product is expected to scale unabated in computing systems [3], which exposes more pressure on the interconnects and switching elements of the intra-DC network to guarantee a balanced I/O bandwidth performance [4].

Current data center networks (DCNs) are organized in a hierarchical tree-like topology based on the bandwidth-limited electronic switches. A certain degree of oversubscription is commonly enforced [5], resulting in the bandwidth bottleneck and large latency especially for inter-rack/cluster communications. High power consumption related to the O/E/O conversion and format-dependent front end is another issue, limiting the power-efficient and cost-efficient scaling to higher capacity. Therefore, optical switching technology has been considered as a promising candidate for intra-DC networking solutions.

Compared with the optical circuit switching (OCS), the OPS and OBS based on fast optical switches could provide on-demand resource utilization, highly flexible connectivity to effectively cope with bursty traffic features, as well as high fan-in/fan-out hotspot patterns in DCNs. Many techniques have been actively developed, each exhibiting advantages and disadvantages when considered for DCN scenarios. In this chapter, various classes of optical switching technologies in implementing

N. Calabretta (✉) • W. Miao

Electro-Optical Communication Group, Institute for Photonic Integration (IPI), Department of Electrical Engineering, Technical University of Eindhoven, Groene Loper 5, 5612 AE, Flux 9.087, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: n.calabretta@tue.nl

OPS and OBS nodes have been briefly introduced. Then several typical optical DCN architectures based on OPS and OBS are presented, followed by the discussion on the performance focusing on different attributes. In the last section, a novel DCN architecture employing fast optical switches is reported which shows potential settlement to the scalability challenges faced by traditional solutions.

3.2 Data Center Networks: Requirements and Challenges

Data centers consist of a multitude of servers as computing nodes and storage subsystems interconnected with the appropriate networking hardware and accompanied by highly engineered power and cooling subsystems [6]. The DCN is responsible to support the large amounts of workload exchanged among the parallel server machines. The traditional DCN uses a multitier architecture, with tens of servers housed in individual racks, and racks are grouped into clusters. The top-of-the-rack (ToR) switches interconnect the servers via copper or optical links, and the inter-rack communication is handled by layers of electronic switches. Ideally, the DCN should provide a full bisection bandwidth, and thus the oversubscription ratio is 1:1 indicating high server utilization and computation efficiency. However, due to the super linear costs associated with scaling the bandwidth and port density of electronic switches, such a design would be prohibitively expensive for a large-scale DCN. In practice, DCs tend to enforce an oversubscription 1:4 to 1:10 [7]. There is more bandwidth available for intra-rack communication than inter-rack communication, and similar trend would be found at higher switching layers.

A set of stringent requirements are imposed on the DCNs, a few key points of which are listed as follows.

- *Capacity*: An increasing fraction of data centers is migrating to warehouse scales. Although substantial traffic will continue to cross between users and data centers, the vast majority of the data communication is taking place within the data center [8]. Recent studies have shown the continuous increase of the inter-rack traffic with a clear majority of traffic being intra-cluster (> 50%) [9]. Higher bandwidth interconnects in combination with high-capacity switching elements are required especially for inter-rack and inter-cluster communications, to avoid the congestion drops caused by the inherent burstiness of flows and intentionally oversubscribed network [10].
- *Latency*: Packet latency is defined as the time it takes for a packet to traverse the network from the sender to the receiver node (end-to-end latency) which includes both the propagation and switch latency. When it comes to the closed environment like DCs, the latency is dominated by the switch latency mainly contributed by the buffering, routing algorithm, and arbitration. Low latency is a crucial performance requirement especially for mission-critical and latency-sensitive applications where microseconds matter (e.g., financial networking).

- *Interconnectivity*: The servers in data centers have 10s–100s concurrent flows on average [9]. Considering the small fraction of intra-rack traffic, almost all flows will traverse an uplink at the ToR switch as inter-rack communication. Therefore, the degree of the interconnectivity supported by the switching network should be large enough to accommodate the number of concurrent flows. Moreover, considering most flow is short and tends to be internally bursty, fast and dynamic reconfiguration of such interconnectivity (e.g., statistical multiplexing) is also needed to guarantee the efficient bandwidth utilization and timely service delivery.
- *Scalability*: The network architecture should enable scaling to large number of nodes to address future capacity needs in a cost-efficient manner. Extension of an existing network in terms of both node count and bandwidth in an incremental fashion is preferable, i.e., without having to replace a disproportionate amount of the installed hardware.
- *Flexibility*: Data centers are expected to adopt technologies that allow them to flexibly manage the service delivery and adapt to changing needs. To this end, the resources (such as computing, storage and network) are pooled and dynamically optimized by the control plane through software configuration. In addition, open standards, open protocols, and open-source development are more and more involved to facilitate and speed up the deployment operation as well as management in the application- and service-based environment.
- *Power/cost efficiency*: A data center represents a significant investment in which the DCN occupies a significant portion [11]. Besides the costs for hardware and software installation, running a large-scale data center is mainly a power consumption matter. Power efficiency is a key target for reducing the energy-related costs and scaling the bandwidth by improving the power density performance. In this sense, significant efforts have been made toward employment of optical technology and virtualization leading to enhancements in power and cost efficiency [12].

As can be seen from these requirements, high-capacity switching networks with low switching latency and fine switching granularity (e.g., deploying statistical multiplexing) are necessary to effectively improve the bandwidth efficiency and handle the burstiness of the DC traffic flows. The large number of concurrent flows makes large interconnectivity as well as fast reconfiguration a necessity for the switches, in which case the circuit-based approaches may be challenging to employ. The pairwise interconnection and tens of milliseconds reconfiguration time have strictly confined the applications to well-scheduled and long-lived tasks.

With the increasing number of server nodes and rapid upgrade in I/O bandwidth, the abovementioned requirements would be quite challenging for current DCN, in terms of both switching node and network architecture.

First, it is difficult for the electronic switch to satisfy the future bandwidth need. The increasing parallelism in microprocessors has enabled continued advancements in computational density. Despite the continuous efforts from merchant silicon providers toward the development of application-specific integrated circuits (ASICs),

the implementation of high-bandwidth electronic switch node is limited by the switch ASIC I/O bandwidth (to multi-Tb/s) due to the scaling issues of the ball grid array (BGA) package [13]. Higher bandwidth is achievable by stacking several ASICs in a multitier structure but at the expense of larger latency and higher cost. Another limiting factor is the power consumption. As electronic switch has to store and transmit each bit of information, it dissipates energy with each bit transition, resulting in power consumption at least proportional to the bit-rate of the information it carries. In addition, the O/E/O conversions and format-dependent interfaces need to be further included as front end, greatly deteriorating the power-efficiency and cost-efficiency performance.

Interconnecting thousands of ToRs, each with large amount of aggregated traffic, would put an enormous pressure on the multitier treelike topology employed by the current DCNs. Due to the limiting performance in terms of bandwidth and port density of conventional electronic switches, network is commonly arranged with oversubscription. Consequently, data-intensive computations become bottlenecked especially for the communication between servers residing in different racks/clusters. The multiple layers of switches also bring large latency when a packet traverses the all DCN to reach its destination, mainly caused by the queueing delay of buffer-related processing. Therefore, to effectively address the bandwidth, latency, scalability, and power requirements imposed by the next-generation DCNs, innovations in the switching technology and network architecture are of paramount significance.

3.3 Optical Data Center Networks

With the prevalence of the high-capacity optical interconnects, optically switched DCNs have been proposed as a solution to overcome the potential scaling issues of the electronic switch and traditional tree-like topology [14, 15]. The switching network handles the data traffic in optical domain thus avoiding the power-consuming O/E/O translations. It also eliminates the dedicated interface for modulation-dependent process, achieving better efficiency and less complexity. Moreover, benefiting from the optical transparency, the switching operation (including the power consumption) is independent of the bit-rate of the information. Scalability to higher bandwidth and employment of WDM technology can be seamlessly supported, enabling superior power-per-unit bandwidth performance.

Various optical switching techniques have been investigated for DC applications, among which the OPS, OBS, and OCS are the most prominent ones. With respect to the requirements for the DCNs, optical switching technologies and the potential of photonic integration can support high-capacity and power-/cost-efficient scaling. Software-defined networking (SDN) is also seeing penetration into the newly proposed optical DCNs to facilitate the flexible provisioning and performance enhancement. However, concerns regarding the limited interconnectivity and handling of applications with fast-changing traffic demands still exist. OCS networks

employing slow switches (tens of milliseconds reconfiguration time) have strictly confined the applications to well-scheduled and long-lived tasks. The static-like and pairwise interconnections would only be beneficial as supplementary elements. The OPS and OBS with fast optical switches allowing for on-demand resource utilization and highly flexible connectivity enabled by the statistical multiplexing are becoming the appealing switching schemes for DCNs.

3.4 Optical Packet and Burst Switching Technologies

OPS and OBS technology makes it possible to achieve sub-wavelength bandwidth granularity exploiting statistical multiplexing of bursty flows. The OPS/OBS network consists of a set of electronic edge nodes interconnected by optical switches. At the edge nodes, electrical data packets from the client network with similar attributes are aggregated in an optical packet/burst. It goes through the optical switches transparently without O/E/O conversion. After arriving at the destined edge node, it is disassembled and forwarded to the client network. The switching operation of the packet/burst (usually referred as payload) is determined by a packet header/burst control header (BCH), which is optically encoded but undergoes O/E conversion and electronic processing at the optical switch node. The main differences between OPS and OBS are:

- In OPS networks, the packet durations are in the hundreds of nanoseconds to microseconds range. The packet header is transmitted in the same channel with respect to the payload and either overlaps the payload in time or sits ahead of it. Advance reservation for the connection is not needed, and the bandwidth can be utilized in the most flexible way. These features make OPS a suitable candidate for data center applications which requires transmission of small data sets in an on-demand manner.
- OBS uses more extensive burst aggregation in the order of tens to thousands of microseconds. The BCH is created and sent toward the destination in a separate channel prior to payload transmission. The BCH informs each node of the arrival of the data burst and drives the allocation of an optical end-to-end connection path. OBS enables sub-wavelength granularity by reserving the bandwidth only for the duration of the actual data transfer.

Note that reconfiguration time of the optical switch including the control operation should be much smaller than the duration of the packet/burst, to ensure a low-latency delivery at a fast arrival rate as well as optimized bandwidth utilization. Practical realization of OPS/OBS relies heavily on the implementation of controlling technique and scheme adopted for contention resolution [16]. Table 3.1 summarizes some examples of different classes of optical switching technologies for OPS/OBS node where comparisons of different attributes in terms of switching performance are also reported. A broad range of technologies has been developed for OPS and OBS systems. The space optical switches based on piezoelectric beam

Table 3.1 Optical switching technologies for implementing OPS and OBS node

	Switching time	Transparency	Scale	Loss	Application	Ref.
Piezoelectric beam steering	~10 ms	Good	384 × 384	Low	OBS	[17]
3D optical MEMS	~10 ms	Good	320 × 320	Low	OBS	[18]
2D optical MEMS	~50 μs	Good	50 × 50	Fair	OBS	[19]
LCoS WSS	~10's ms	Good	1 × 40	Fair	OBS	[21]
Mach-Zehnder (thermo-optic)	~10' μs	Good	32 × 32	High	OBS	[20]
Micro-ring resonator (thermo-optic)	~10' μs	Fair	8 × 8	Fair	OBS	[22]
PLZT MZI	~10 ns	Good	8 × 8	High	OPS, OBS	[23]
InP MZI	2.5 ns	Good	8 × 8	High	OPS, OBS	[24]
LiNbO ₃ MZI	~1 ns	Fair	32 × 32	High	OPS, OBS	[25]
MZI + EAM	< 10 ns	Good	8 × 8	Fair	OPS, OBS	[26]
TWC + AWGR	~10's ns	Poor	10's × 10's	Fair	OPS, OBS	[27]
TL + AWGR	~10's ns	Good	100' × 100's	Low	OPS, OBS	[28]
EAM B&S	~1 ns	Good	8 × 8	High	OPS, OBS	[29]
SOA B&S	~1 ns	Good	64 × 64	Fair	OPS, OBS	[30]
SOA multistage	< 10 ns	Good	16 × 16	Fair	OPS, OBS	[31]
Semiconductor optical phase array	~20 ns	Good	64 × 64	Fair	OPS, OBS	[32]

steering and 3D MEMS as well as wavelength selective switch (WSS) based on Liquid Crystal on Silicon (LCoS) have tens of milliseconds of switching time, which are more suitable for long burst operations. The rest of the techniques are mainly based on interferometric and gating switch elements, holding the potential of photonic integration to further scale the capacity. Large interconnectivity can be enabled by cascading 1×2 or 2×2 switching elements such as 2×2 Mach-Zehnder interferometer (MZI) and micro-ring resonator (MRR). Mach-Zehnder switches with electrooptic switching offer faster reconfiguration than the thermo-optic tuning, and extra optical amplifier is normally needed due to the relatively high insertion loss, and therefore scalability can be compromised by the OSNR degradation. Another category of the fast (nanoseconds) optical switches is implemented by arrayed waveguide grating router (AWGR) along with tunable lasers (TLs) or tunable wavelength converters (TWCs). The interconnection scale and performance is largely dependent on the capability of the TL and TWC. Note that WSS, MRR, and AWGR are all wavelength dependent. For the broadcast-and-select (B&S) architecture, the semiconductor optical amplifier (SOA) and electro-absorption modulator (EAM) are commonly used as the gating elements. The broadcast stage introduces high splitting loss, in which case the SOA can provide loss compensation which is essential in realizing a large connectivity. In the practical implementation of the OPS/OBS network, the techniques (or in combination) listed here can be further included in a network as a basic switching unit [33, 34].

3.4.1 Technical Challenges in OPS/OBS Data Centers

Despite the advantages of increased capacity and power efficiency brought by the optical transparency, employing OPS/OBS is actually facing several challenges which need to be carefully considered for DC networking applications.

- *Lack of optical memory*: As no effective optical memories exist, contention resolution is one of the most critical functionalities that need to be addressed for OPS/OBS node. Several approaches have been proposed to resolve the contention in one or several of the following domains:
 - Time domain: the contending packet/burst can be stored in fixed fiber delay line (FDL) or electronic buffer.
 - Wavelength domain: by means of wavelength conversion, the packet/burst can be forwarded in alternative wavelength channels.
 - Space domain: contending burst is forwarded to another output port (deflection routing).

The techniques based on FDL, wavelength conversion, and deflection routing increase significantly the system complexity in terms of routing control and packet synchronization. Moreover, the power and quality of the signal are degraded which results in limited and fixed buffering time. A promising solution is to exploit the electronic buffer at the edge nodes [35]. To minimize the latency, the optical switch should be as close as possible to the edge nodes and fast decision-making is required. This is feasible in a DC environment with interconnects ranging from few to hundreds of meters.

- *Fast reconfiguration and control mechanism*: To fully benefit from the flexibility enabled by the statistical multiplexing, fast reconfiguration of the optical switch is a key feature. Although OBS is less time demanding, slower OBS can cause inefficiency and unpredictability especially under high network load. Therefore, the optical fabrics with fast-switching time together with fast-controlling mechanism are desired. Regarding the DCN applications, the implementation of the controlling technique should follow the increase of the network scale and optical switch port count and, more importantly, occupy as least resources as possible.
- *Scalability*: Depending on the design and technology employed in optical switches, signal impairment and distortion are observed due to noise and optical nonlinearities. Consequently, the optical switches are realized with limited port count. Scaling the network interconnectivity and maintaining the performance would require the switches to have port count as large as possible and to be intelligently connected to avoid the hierarchical structure. The flat topology also brings the benefits of simplified controlling and buffering which may be problematic for fast optical switches. On the other hand, optical transparency and WDM technology would benefit the DCN in the context of scaling up the bandwidth density. Further improvements could be made by means of photonic integration, which greatly reduces the optical switch footprint and power consumption.

3.5 Optical DCN Architecture Based on OPS/OBS

OPS and OBS technologies providing high bandwidth and power efficiency have been adopted in the recent researches for optical DCNs [36, 37]. This section gives an overview and a general insight on the optical DCN architectures based on OPS/OBS that have been recently proposed, which can be classified into different categories according to the switching technologies used.

3.5.1 Based on OBS

3.5.1.1 OBS with Fast Optical Switches

In [44], a DCN consisting of ToR switches at the edge and an array of fast optical switches at the core to perform optical burst forwarding on the pre-configured light paths has been proposed. It has separate data and control planes as shown in Fig. 3.1. Two-way reservation OBS is implemented, facilitated by the single-hop topology with configuration of only one switch per request. It achieves zero burst loss with slight degradation of the latency owing to the limited round-trip time in DC environment. The centralized control plane is responsible for the routing, scheduling, and switch configuration. It processes the control packets from the ToRs sent through a dedicated optical transceiver, finds appropriate path to the destination ToR, and configures optical switches as allocated in the control packets. Since the fast optical switch connects to every ToRs, scalability is challenging in terms of achievable port count for large number of ToRs. The resulted complexity in the control plane may be another bottleneck in scaling up the network.

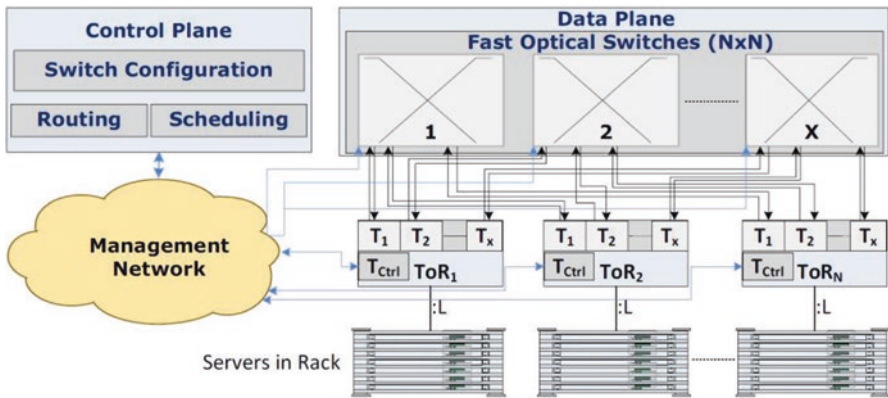


Fig. 3.1 DCN based on OBS with fast optical switches

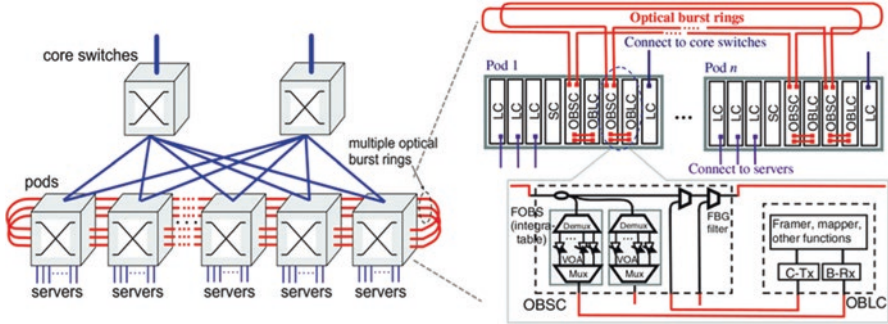


Fig. 3.2 Multiple optical burst rings and internal architecture of the pod

3.5.1.2 Optical Burst Rings

The OBS is utilized in [45] to improve the inter-pod communications in DCNs. The network architecture and the pod are depicted in Fig. 3.2. The pods are connected through the multiple optical burst rings. Bursty and fast-changing inter-pod traffic is handled by the core switches, while the relatively stationary traffic is handled via the optical burst rings. Some line cards (LCs) are configured for connecting the servers, and others are used to access the core switches. The switch cards (SCs) aggregate the traffic and together with the control unit make decision to forward the traffic to the LCs connecting to the core switches or to an optical burst line card (OBLC) which sends the traffic in form of burst to the optical rings. The optical burst switch cards (OBSCs) perform optical burst add/drop to/from the optical burst rings, as shown in Fig. 3.2. The advantages of this architecture are the high inter-pod transmission bandwidth and large number of interconnectivity (>1000 pods). Much shorter connection reconfiguration time is offered compared with OCS-based solutions, achieving better bandwidth utilization.

3.5.1.3 HOS Architecture

An optical switched interconnect based on hybrid optical switching (HOS) has been proposed and investigated in [46]. HOS integrates optical circuit, burst, and packet switching within the same network, so that different DC applications are mapped to the most suitable optical switching mechanisms. As shown in Fig. 3.3, the HOS is arranged in a traditional fat-tree three-tier topology, where the aggregation switches and the core switches are replaced by the HOS edge and core node, respectively. The HOS edge nodes are electronic switches which perform the traffic classification and aggregation. The core node has parallel optical switches composed of switching elements (SEs). A slow optical switch based on 3D MEMS handles circuits and long bursts, and a fast SOA-based optical switch with a three-stage Clos network deals with packets and short bursts. The HOS control plane manages the scheduling

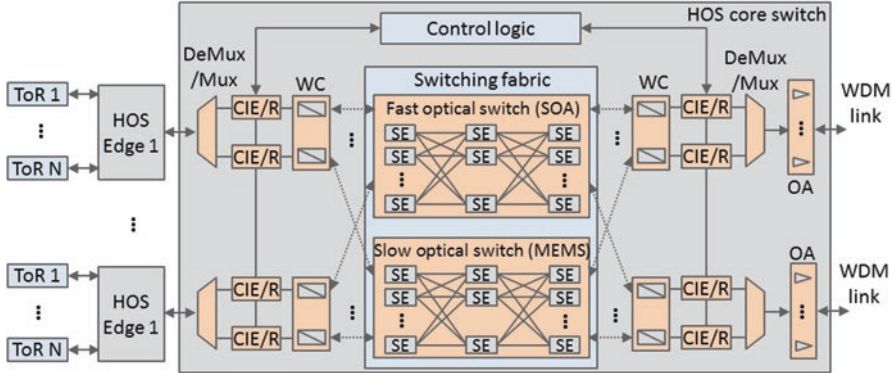


Fig. 3.3 HOS interconnection network

and transmission of the optical circuits, bursts, and packets. Wavelength converters (WCs) are used to solve the possible contentions. Numerical studies show low loss rates and low delays although the practical implementation of a large-scale network remains challenging.

3.5.1.4 HOSA Architecture

HOSA, shown in Fig. 3.4, is another DCN architecture that employs both fast and slow optical switches [47]. Different with the previous work that uses only fast optical switches [44], slow MEMS optical switches are added to exploit the benefits of both types of fabrics. The traffic assembling/disassembling and classification is implemented at the newly designed ToR switch. The control plane still uses a centralized controller which receives connection requests and configures the data plane through a management network. The array of fast optical switches operates in an OBS manner, forwarding the data burst on the predefined connection path. The evaluation results show low-latency and high-throughput performance with low power consumption, assuming large port counts of slow/fast optical switches are deployed in the single-stage network.

3.5.1.5 Torus-Topology DCN

Figure 3.5 shows the Torus DCN [48] based on co-deployment of OPS and OCS. The architecture features with a flat topology where each hybrid optoelectronic router (HOPR), that interconnects a group of ToR switches, is connected to the neighboring HOPRs. The traffic from server is converted into the optical packet and fed into the corresponding HOPR attached with a fixed-length optical label. HOPR uses fast

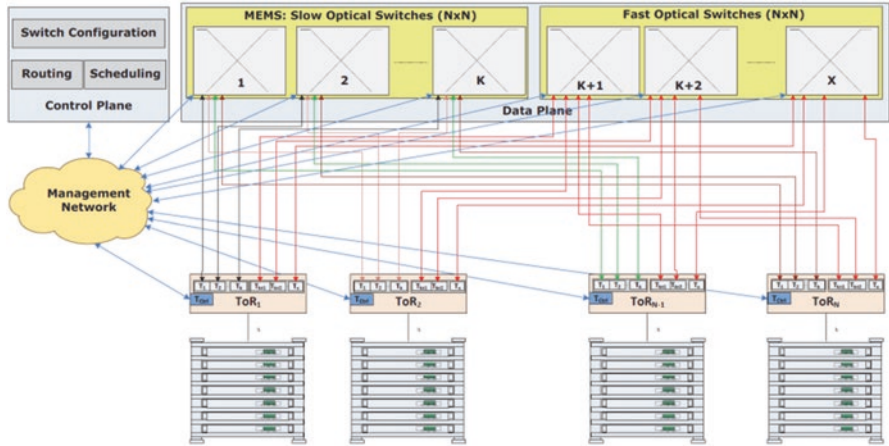


Fig. 3.4 HOSA DCN architecture

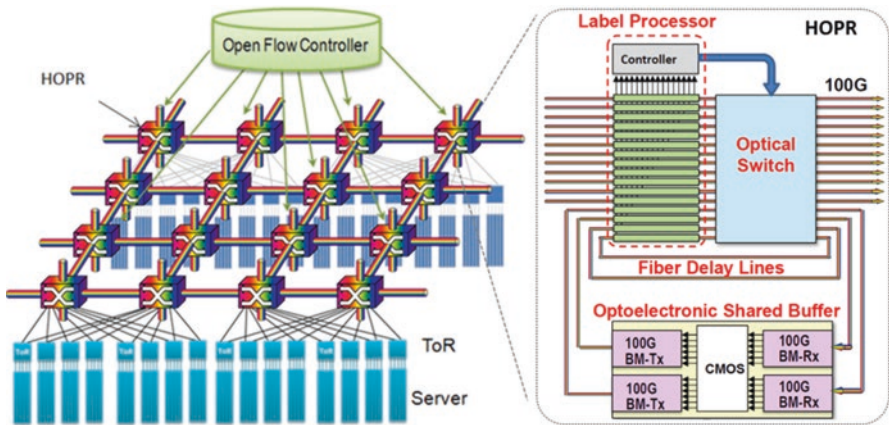


Fig. 3.5 Torus DCN employing hybrid optoelectronic routers (HOPRs)

optical fabric (EAM-based broadcast-and-select structure) which supports both packet operation and circuit operation (express path). The packet contention, which happens when a link is desired by more than one packets or it is reserved by an express path, is solved by different schemes (i.e., deflection routing, FDLs, and optoelectronic shared buffer). The enabling technologies for implementing an HOPR have been detailed, aiming at a high energy efficiency and low latency of 100 ns regime. For an efficient transfer of high-volume traffic, a flow management has been implemented with OpenFlow-controlled express path. Despite the multi-hop transmission may be needed for interconnecting the ToRs, Torus provides the advantages of superior scalability and robust connectivity.

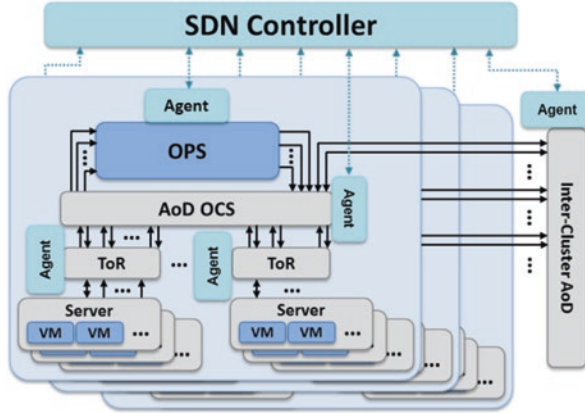


Fig. 3.6 The LIGHTNESS DCN architecture

3.5.1.6 LIGHTNESS DCN Architecture

A flat DCN architecture integrating both OPS and OCS switching technologies to deal with the inconsistent application requirements has been investigated in LIGHTNESS project [49]. The hybrid network interface card (NIC) located in each server supports the switching of the traffic to either OPS or OCS resulting in an efficient utilization of the network bandwidth. As illustrated in Fig. 3.6, the SOA-based OPS which employs broadcast-and-select architecture is plugged into the Architecture on Demand (AoD) backplane as a switching module to handle short-lived data packets. The AoD itself is a large port count fiber switch which can be configured to support OCS function for long-lived data flows. The network can be scaled by interconnecting multiple intra-cluster AoDs with an inter-cluster AoD. Another innovation made by LIGHTNESS is the fully programmable data plane enabled by the unified SDN control plane. It is worth noting that the switching operation of the OPS is controlled by the local switch controller based on the in-band optical labels, which is decoupled from the SDN-based control (e.g., look-up table update and statistic monitoring). Similar schemes have been found in Archon [50] and burst-over-circuit architecture [51] where the OPS is replaced by PLZT-based optical switch and AWGR incorporating with TWC, respectively.

3.5.2 Based on OPS

3.5.2.1 IRIS Project: Photonic Terabit Routers

The IRIS project has developed a photonic packet router that scales to hundreds of terabit/s capacity [38]. As shown in Fig. 3.7, the router employs a load-balanced multistage architecture. Each node (e.g., ToR switch) is connected to an input port

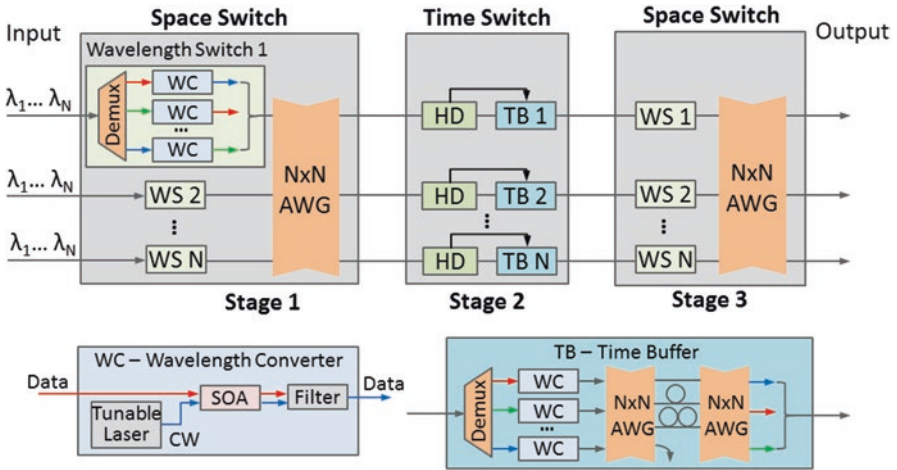


Fig. 3.7 IRIS photonic terabit router

of the first stage using N WDM wavelength channels each carrying synchronous fixed-length data packets. The wavelength switch is based on an array of all-optical SOA-based wavelength converters to set the wavelength routing. The second stage is time switch which contains N time buffers consisting of shared ODLs.

The wavelength is configured in the way that the packet entering on port of the time buffer always exits on the corresponding output port. The third stage then forwards the packet to the desired destination. Due to the periodic operation of the third space switch, the scheduling is local and deterministic to each time buffer which greatly reduces control-plane complexity. The IRIS project has demonstrated the operation of a partially populated router with integrated photonic circuits and developed interoperability card that can connect electronic routers with 10 Gb Ethernet interfaces to the IRIS router. Using 40 Gb/s data packets and 80×80 AWGs allows this architecture to scale to 256 Tb/s capacity.

3.5.2.2 Petabit Optical Switch

The petabit optical switch is based on tunable lasers (TLs), tunable wavelength converters (TWCs), and AWGRs, as schematically shown in Fig. 3.8 [39]. The ToR switches are connected to the optical switch, which is three-stage Clos network including input modules (IMs), central modules (CMs), and output modules (OMs). Each module uses an AWGR as the core. The SOA-based TWCs as well as the TLs in the line cards are controlled by the scheduler. A prominent feature of the switch is that packets are buffered only at the line cards, while the IMs, CMs, and OMs do not require buffers. This helps to reduce implementation complexity and to achieve low latency. Performance of the petabit optical switch is evaluated with simulation which shows high throughput benefited from the efficient scheduling algorithm.

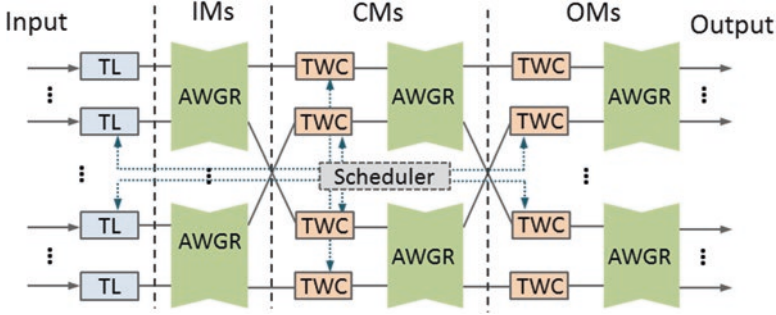


Fig. 3.8 The petabit optical switch architecture

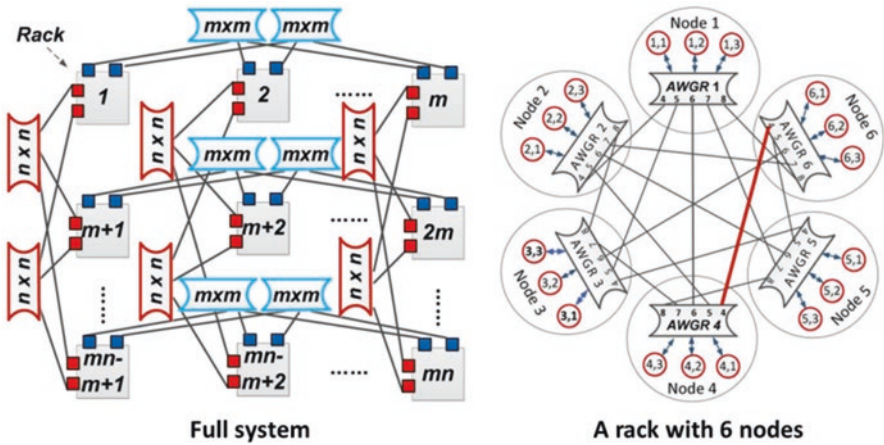


Fig. 3.9 Hi-LION full system with inter-/intra-rack AWGRs communication

3.5.2.3 Hi-LION

A large-scale interconnect optical network Hi-LION has been proposed in [28]. It exploits fast-tunable lasers and high-radix AWGRs in hierarchy to achieve very large-scale and low-latency interconnection of computing nodes. The architecture of the full system and an example of 6-node rack is depicted in Fig. 3.9. The essence is to rely on the unique wavelength routing property assisted by electrical switching embedded in the node to provide all-to-all flat interconnectivity at every level of hierarchy (node-to-node and rack-to-rack). As shown in the Fig. 3.9, the local AWGRs and global AWGRs are used to handle the intra-rack and inter-rack communications, respectively. Single-hop routing in the optical domain also avoids the utilization of optical buffers. However, the maximum hop count for inter-rack

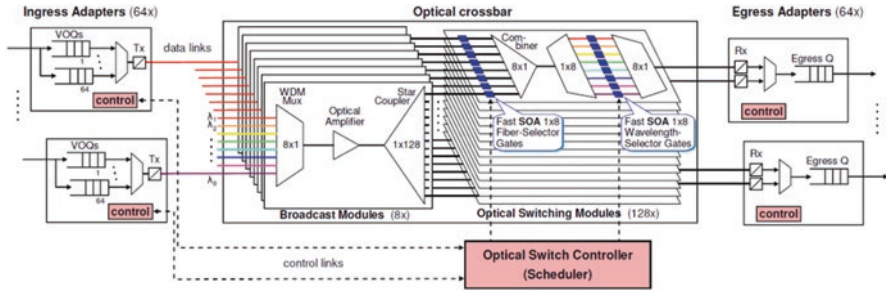


Fig. 3.10 Single-stage OSMOSIS switching system

communication can be seven including the intra-rack forwarding. Compared with previous AWGR-based solutions like DOS (LIONS) [27] and TONAK LION [40], interconnectivity of more than 120,000 nodes can be potentially connected.

3.5.2.4 OSMOSIS Optical Packet Switch

OSMOSIS project targets for accelerating the state of optical switching technology for use in supercomputers [41]. The architecture of the implemented single-stage 64-port optical packet switch is illustrated in Fig. 3.10. It is based on a broadcast-and-select architecture and the switching modules consist of a fiber and a wavelength selection stage, both built with SOAs as the gating elements. The switching of the synchronous fixed-length optical packets is controlled via a separate central scheduler. The performance studies of the OSMOSIS demonstrator confirm high-capacity and low-latency switching capabilities. Two-level fat-tree topology can be potentially built, further scaling the network to 2048 nodes.

3.5.2.5 Data Vortex

The Data Vortex is a distributed interconnection network which is entirely composed of 2×2 switching nodes arranged as concentric cylinders [42]. As illustrated in Fig. 3.11, the Data Vortex topology integrates internalized virtual buffering with banyan-style bitwise routing specifically designed for implementation with fiber-optic components. The 2×2 node uses SOA as the switching element. The broadband operation of SOA allows for successful routing of multichannel WDM packets. Contentions of packet are resolved through deflection routing. The hierarchical multistage structure is easily scalable to larger network size. However, the practical scale is limited by the increased and nondeterministic latency, as well as deteriorated signal quality.

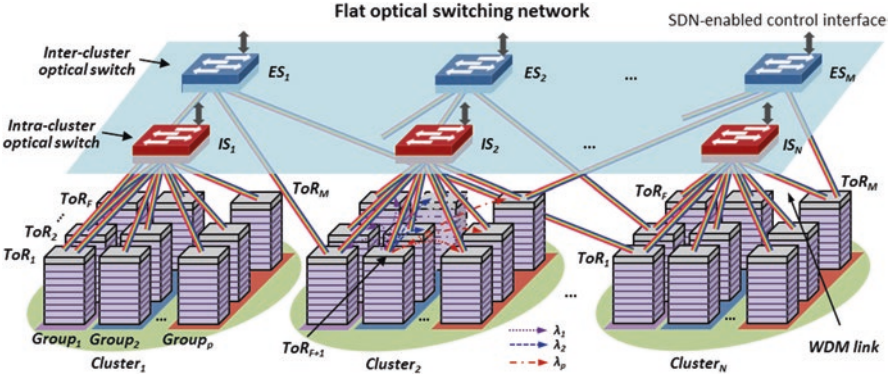


Fig. 3.12 OPSquare DCN architecture built on fast optical switches

is sent to the destination ToR via the fast optical switch. An optical in-band RF tone label is attached to the packet which will be extracted and processed at the fast optical switch node. Multiple (p and q) WDM transceivers allow for scaling the communication bandwidth between the ToRs and the optical network. Moreover, each of the WDM transceivers is dedicated for the communication with a different group of ToRs. For intra-cluster network, the M ToRs are thus divided into p groups, and each group contains $F = M/p$ ToRs. One of the p WDM TXs addresses F (instead of M) possible destination ToRs, in combination with the $1 \times F$ switch at IS. The structure and operation of the inter-cluster interface are similar to the intra-cluster ones.

The schematic of the fast optical switch node acting as IS/ES is shown in Fig. 3.13. The optical switching is realized by SOA-based broadcast-and-select architecture. The fast optical switch node has a modular structure, and each module consists of F units each of which handling the WDM traffic from one of the F ToRs in a single group. The WDM inputs are processed in parallel, and the label extractor (LE) separates the optical label from the payload. The extracted label is processed by the switch controller. The SOA has nanoseconds switching speed and can provide optical amplification to compensate the losses caused by the broadcasting stage. The contention is solved through the optical flow control, according to which ToR releases the packets stored in the buffers (for ACK) or triggers the retransmission (for NACK). Different class of priority can be applied to guarantee traffic with more stringent QoS requirement. The priority is defined by provisioning the look-up table in the switch controller through the SDN control interface [53]. In addition, the SDN control plane can create and manage multiple virtual networks over the same infrastructure by configuring the look-up table and make further optimization through the developed monitoring functions. Note also that benefiting from the WDM TX wavelength assignment for the grouped ToRs and the wavelength switching capability enabled by the optical switch node, the splitting loss of the broadcast-and-select switch is $3 \times \log_2 F$ dB ($F = M/p$ for the IS switch and $F = N/q$ for the ES switch), which is much less than $3 \times \log_2 M$ dB and $3 \times \log_2 N$ dB required in the original configuration. Lower splitting losses lead to less OSNR degradation and

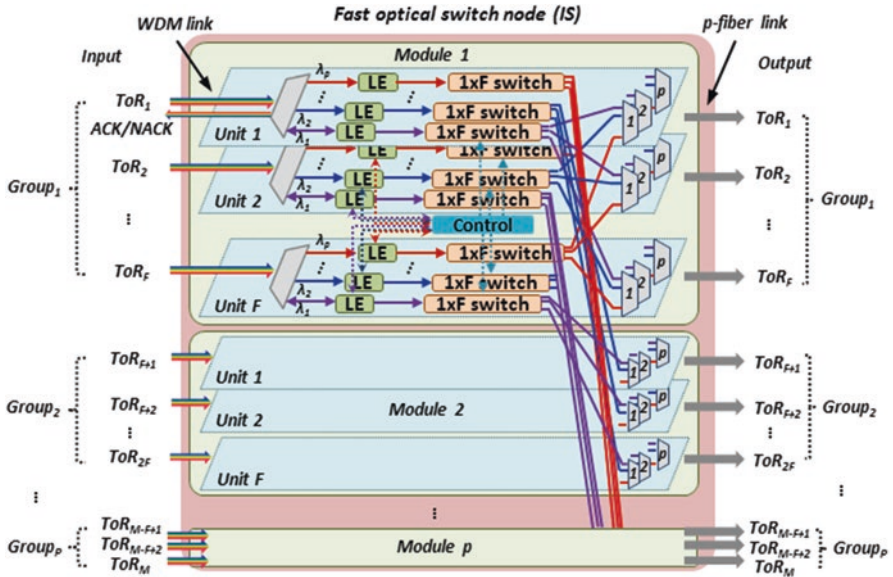


Fig. 3.13 Flow-controlled fast optical switch node

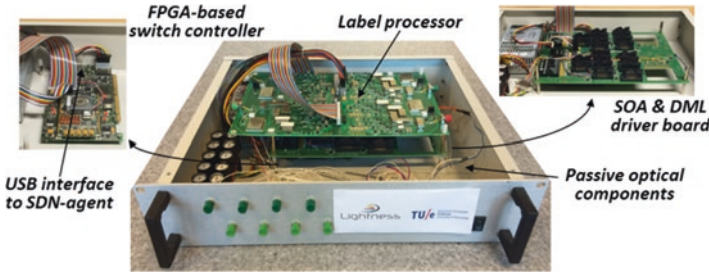


Fig. 3.14 Flow-controlled fast optical switch node prototype

significant improvement of the scalability and feasibility of the network. The developed 4×4 optical switch prototype integrating the FPGA-based switch controller (with interface to SDN-agent), label processor, SOA drivers, and passive optical components (circulators, couplers, etc.) is shown in Fig. 3.14.

3.6.1 Performance Investigation

The performance studies of the OPSquare architecture are reported in Fig. 3.15. The DCN includes 40 servers per rack each with 10 Gb/s uplink programmed to create Ethernet packets with 40–1500 bytes length at a certain load [43]. The round-trip

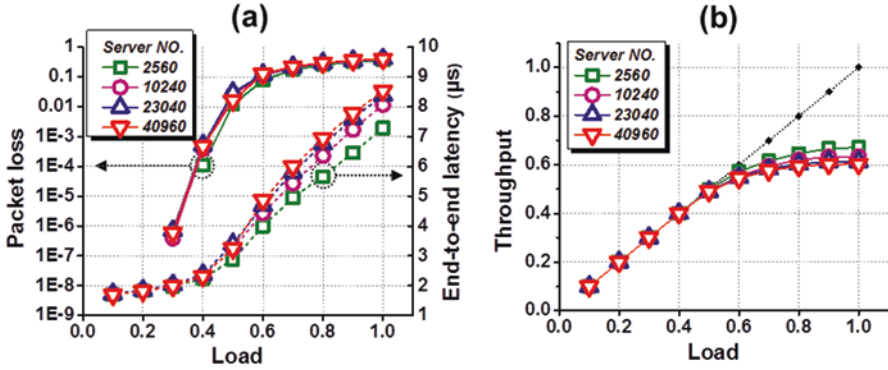


Fig. 3.15 Performance of (a) packet loss ratio and server end-to-end latency and (b) throughput

time between the ToR and optical switch node is 560 ns (2×50 m distance and 60 ns delay caused by the label processor as well as flow control operation). Considering that most of the traffic resides inside the cluster, four transceivers have been assigned to the IS and one for the ES ($p = 4$ and $q = 1$). For the inter-rack communication, the data packets will be forwarded to the ports associated with the intra-/inter-cluster network interface and aggregated to compose a 320 byte optical packet to be transmitted in the fixed 51.2 ns time slot. The delay caused by the head processing and buffering at the ToR input is taken as 80 ns and 51.2 ns, respectively. DCNs with variable amount of servers (from 2560 to 40,960) and racks (from 64 to 1024) with 3:1 intra-/inter-cluster traffic have been investigated. Thus, ES and IS optical switches with port count of 8×8 to 32×32 are needed to build up the desired DCN size. The buffer is set as 20 KB for each TX. The packet loss ratio and server end-to-end latency reported in Fig. 3.15a as a function of the load indicate almost no performance degradation as the number of servers increases. The packet loss ratio is smaller than 1×10^{-6} , and the server end-to-end latency is lower than $2 \mu\text{s}$ at load of 0.3 for all scales, which indicates the potential scalability of the OPSquare architecture. Similar results have been achieved for the throughput performance, as clearly shown in Fig. 3.15b.

The performance of the OPSquare architecture in terms of scalability and capacity is also investigated by using different modulation formats. The transparency to the data rate/format enabled by the fast optical switches allows for the immediate capacity upgrade maintaining the same switching infrastructure, without dedicated parallel optics and format-dependent interfaces to be further included as front end. In this respect, three types of directly modulated traffic, namely, 28 Gb/s PAM4, 40 Gb/s DMT, and 4×25 Gb/s NRZ OOK featuring with IM/DD, have been investigated. Exploiting the modular structure of the optical switch, the switching performance in OPSquare would mainly depend on the $1 \times F$ broadcast-and-select switch

and will be limited by the splitting loss experienced by the payload. Using the prototyped optical switch shown in Fig. 3.14, the switching performance and the port count scalability for realizing a large-scale OPSquare DCN have been assessed. Details on the experimental setup are reported in [43].

The power penalty at $\text{BER} = 10^{-3}$ measured at different input optical powers within scale of 32×32 and 64×64 optical switch for 28 Gb/s PAM4 and 40 Gb/s DMT is depicted in Figs. 3.16a and b, respectively. An example of the optimal bit allocation after bit loading is included in Fig. 3.16b. In 32×32 scale, 10 dB dynamic range has been measured with <3 dB penalty, while for 64×64 , 8 dB dynamic range has been obtained for both traffic with 3 dB penalty. With 4 WDM transceivers ($p = q = 4$) per ToR each operating at 40 Gb/s and 64×64 optical switches (1×16 broadcasting ratio) used, an OPSquare DCN comprising 4096 ToRs each with 320 Gb/s aggregation bandwidth would have a capacity >1.3 Pb/s. Larger interconnectivity can be achieved either by increasing the broadcasting ratio of the $1 \times F$ switch with limited performance degradation or increasing the number of transceivers per ToR which could also improve the bandwidth performance.

Then the waveband switching of 4×25 Gb/s data payload enabled by the broadband operation of the SOA-based switch is analyzed. The power penalty at $\text{BER} = 10^{-9}$ with different input optical powers at scales of 32×32 , 64×64 , and 128×128 optical switch when employing 4 wavebands ($p = q = 4$) is reported in Fig. 3.17c, respectively. The 16 dB input dynamic range is achieved with less than 2 dB power penalty. Here each waveband has a 100 Gb/s capacity which can be increased by inserting more wavelength channels. With four 100 Gb/s wavebands per ToR and 64×64 optical switches (1×16 broadcasting ratio), an interconnectivity of $64^2 = 4096$ ToRs and a capacity >3.2 Pb/s can be achieved benefiting from the transparency and TX wavelength assignment for the grouped ToRs featured by the OPSquare DCN.

The experimental assessments of the OPSquare DCN reported so far are based on discrete components which would result in power-inefficient bulky systems in practical implementations. Photonic integrated circuits (PICs) can reduce the footprint and the power consumption. In view of this, a 4×4 WDM fast optical switch PIC has been designed and fabricated exploiting the modular architecture as shown in Fig. 3.17. The modular photonic chip shown in Fig. 3.17 is a 4×16 port (the combiners shown in the schematic on the left side of Fig. 3.17 were not integrated in this photonic circuit for lack of space) and integrates four optical modules, in which each module includes four WSSs. More than 100 components including the SOAs, AWGs, and couplers are integrated in the same chip. As reported in [54, 55], the compensation of the losses offered by the SOAs allowing for large dynamic range, the low cross talk, and the wavelength, time, and switch nanosecond switch operation indicate the potential scalability to higher data rate and larger number of port counts of the optical switch PIC and potential enhancement of the OPSquare DCN performance.

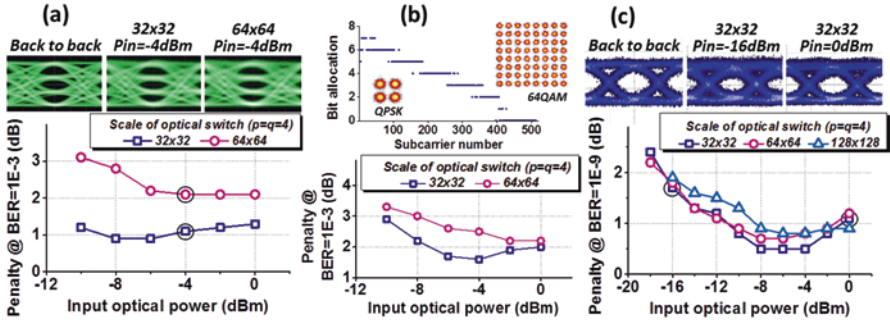


Fig. 3.16 Power penalty vs. input optical power for (a) 28 Gb/s PAM4, (b) 40 Gb/s DMT, and (c) waveband 4 × 25 Gb/s traffic

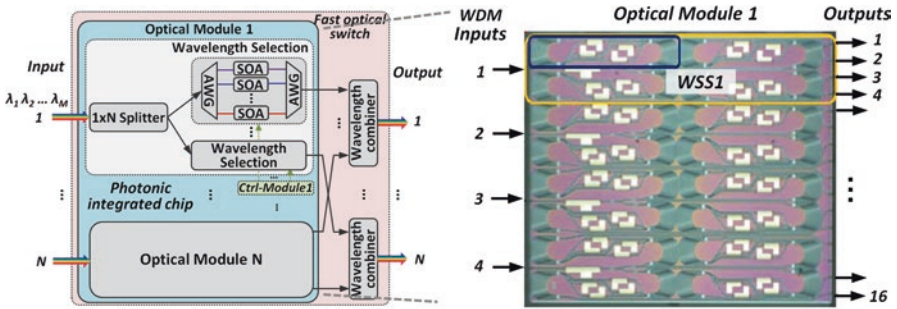


Fig. 3.17 Schematic of the fabricated 4 × 4 fast optical switch PIC

3.7 Conclusions and Discussions

The never-ending growth in the demand for high bandwidth in data centers is accelerating the deployment of more powerful servers and more advanced optical interconnects. To accommodate the increasing volume of traffic with low communication latency and high power efficiency, technological and architectural innovations of the DCNs are required. OPS/OBS based on fast optical switches is an attractive solution, by providing efficient statistical multiplexing and transparent high-capacity operation and eliminating the O/E/O conversions as well as opaque front ends. However, the lack of optical memory, the limited scalability due to the relative low port count of fast (nanoseconds) optical switches, the inefficient and no scalable centralized scheduler/control system capable to fast (tens of nanoseconds) control and configure the overall optical data plane based on fast optical switch, and the compatibility of the OPS technology with commercial Ethernet switches and protocol are some of the practical hurdles to exploit OPS and OBS in DCN. Solving those problems require complete solutions from the DCN architecture down to the devices. Promising results have been shown in the recent investigations to solve

those issues. Optical DCN architectures based on OPS and OBS have been presented in this chapter, and the different characteristics in terms of scalability, flexibility, and power/cost efficiency are summarized in Table II. As can be seen in the table, for the contention resolution, most of the schemes use practical electronic buffer (EB) placed at the edge side, either waiting for the command of the scheduler or retransmitting the packet/burst in case of contention. The efficiency of the scheduling, the configuration time of the switch, and the round-trip time would play an important role in reducing the processing latency and the size of the costly buffer. It is difficult for the architectures with a single switching element to scale to large number of interconnections. In this respect, multistage and parallel topologies have been adopted by many solutions. The fast reconfiguration of the optical switches used for OPS and OBS has allowed for flexible interconnectivity which is a desired feature for DC applications. Relatively lower power/cost efficiency is the price need to pay compared with OCS technology, mainly due to the active components and the loss experienced in the switch fabrics. Performance improvement is expected, with the maturing of the fast optical switching technologies in combination with photonic integration (Table 3.2).

Table 3.2 Optical DCN architectures based on OPS and OBS technologies

	Technology	Contention resolution	Scalability	Flexibility	Power/cost efficiency
IRIS	OPS	ODL	Fair	Good	Fair
Petabit	OPS	EB at edge	Fair	Good	Fair
DOS (LIONS)	OPS	EB	Poor	Good	Fair
TONAK-LION	OPS	EB at edge	Poor	Good	Fair
Hi-LION	OPS	EB at edge	Good	Fair	Fair
OSMOSIS	OPS	EB at edge	Fair	Good	Fair
Data Vortex	OPS	Deflection	Fair	Good	Fair
OPSquare	OPS	EB at edge	Good	Good	Fair
OBS with fast optical switch	OBS	EB at edge	Poor	Fair	Good
Optical burst ring	OBS + EPS	EB at edge	Fair	Fair	Good
HOS	OPS + OBS+ OCS	EB at edge	Fair	Good	Fair
HOSA	OBS + OCS	EB at edge	Fair	Fair	Good
Torus	OPS + OCS	Deflection +ODL+ EB	Good	Good	Fair
LIGHTNESS	OPS + OCS	EB at edge	Good	Fair	Fair
Archon	OPS + OCS	EB at edge	Good	Fair	Fair
Burst over circuit	OBS + OCS	EB at edge	Poor	Fair	Good

References

1. International Data Corporation, The new need for speed in the datacenter. in *2015. International Data Corporation, "The New Need for Speed in the Datacenter*, 2015
2. S. Shah, N. Guenov, Multicore Processing: Virtualization and Data Center, Freescale, 2012
3. B. Bland, Titan – early experience with the Titan system at Oak Ridge National Laboratory. in *2012 SC Companion: High Performance Computing, Networking, Storage and Analysis*, 2012
4. G. Bell, J. Gray, A. Szalay, Petascale computational systems. *IEEE Comput.* **39**(1), 110–112 (2006)
5. T. Benson, A. Akella, D.A. Maltz, Network traffic characteristics of data centers in the wild. in *10th ACM SIGCOMM Conference on Internet Measurement*, 2010
6. L.A. Barroso, J. Clidaras, U. Hölzle, The datacenter as a computer: an introduction to the design of warehouse-scale machines. *Synth. Lect. Comput. Architect.* **8**(3), 1–154 (2013)
7. A. Greenberg, J.R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D.A. Maltz, P. Patel, S. Sengupta, VL2: a scalable and flexible data center network. *ACM SIGCOMM Comp. Commun. Rev.* **39**(4), 51–62 (2009)
8. Cisco, *Cisco Global Cloud Index: Forecast and Methodology*, 2014–2019, 2015
9. A. Roy, H. Zeng, J. Bagga, G. Porter, A.C. Snoeren, Inside the social network's (datacenter) network. *ACM SIGCOMM Comput. Commun. Rev. SIGCOMM'15* **45**(4), 123–137 (2015)
10. A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead, R. Bannon, S. Boving, G. Desai, B. Felderman, P. Germano, A. Kanagala, J. Provost, J. Simmons, E. Tanda, J. Wanderer, U. Hölzle, S. Stuart, A. Vahdat, Jupiter rising: a decade of clos topologies and centralized control in Google's datacenter Network. *ACM SIGCOMM Comput. Commun. Rev. SIGCOMM'15* **45**(4), 183–197 (2015)
11. A. Greenburg, J. Hamilton, D.A. Maltz, P. Patel, The cost of a cloud: research problems in data center networks. *ACM SIGCOMM Comput. Commun. Rev.* **39**1, 68–73 (2009)
12. A. Hammadi, L. Mhamdi, A survey on architectures and energy efficiency in Data Center Networks. *Comput. Commun.* **40**, 1–21 (2014)
13. A. Ghiasi, Large data centers interconnect bottlenecks. *Opt. Express* **23**(3), 2085–2090 (2015)
14. C. Kachris, I. Tomkos, A roadmap on optical interconnects in Data Centre Networks. in *International Conference on Transparent Optical Networks*, 2015.
15. K. Aziz, M. Fayyaz, Optical interconnects for Data Center Networks, in *Handbook on Data Centers*, (Springer, New York, 2015), pp. 449–483
16. S.J. Ben Yoo, Optical packet and burst switching technologies for the future photonic Internet. *J. Lightw. Technol.* **24**(12), 4468–4492 (2006)
17. Polatis Series 7000, [Online]. Available: <http://www.polatis.com/series-7000-384x384-port-software-controlled-optical-circuit-switch-sdn-enabled.asp>
18. CALIENT [Online]. Available: <http://www.calient.net/products/s-series-photonic-switch/>
19. S. Han, T.J. Seok, K. Yu, N. Quack, R.S. Muller, M.C. Wu, 50x50 polarization-insensitive silicon photonic MEMS switches: design and experiment. in *42nd European Conference on Optical Communication*, Paper Th.3.A. 5, Dusseldorf, Germany, 2016
20. K. Tanizawa, K. Suzuki, M. Toyama, M. Ohtsuka, N. Yokoyama, K. Matsumaro, M. Seki, K. Koshino, T. Sugaya, S. Suda, G. Cong, T. Kimura, K. Ikeda, S. Namiki, H. Kawashima, Ultra-compact 32 × 32 strictly-non-blocking Si-wire optical switch with fan-out LGA interposer. *Opt. Express* **23**, 17599–17606 (2015)
21. M. Iwama, M. Takahashi, M. Kimura, Y. Uchida, J. Hasesawa, R. Kawahara, N. Kagi, LCOS-based flexible grid 1×40 wavelength selective switch using planar lightwave circuit as spot size converter. in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, 2015
22. F. Testa et al, Design and implementation of an integrated reconfigurable silicon photonics switch matrix in IRIS project, *J. Select. Top. Quant. Electron.*, vol. 22, n. 6, pp. 155-168, 2016.
23. EpiPhotonics, [Online]. Available: <http://epiphotonics.com/products1.htm>

24. H. Kouketsu, S. Kawasaki, N. Koyama, A. Takei, T. Taniguchi, Y. Matsushima, K. Utaka, High-speed and Compact Non-blocking 8×8 InAlGaAs/InAlAs Mach-Zehnder-Type Optical Switch Fabric. *Opt. Fiber Commun. Conf. M2K*(3) (2014)
25. H. Okayama, M. Kawahara, Prototype 32×32 optical switch matrix, *Electron. Lett.*, **30**(14), 1128–1129, 1 1994
26. Y. Muranaka, T. Segawa, R. Takahashi, Integrated fat-tree optical switch with Cascaded MZIs and EAM-gate array. in *21st OptoElectronics and Communications Conference/International Conference on Photonics in Switching 2016 (OECC/PS 2016)*, paper WF3–2, Niigata, Japan, July 2016
27. Y. Yin, R. Proietti, X. Ye, C.J. Nitta, V. Akella, S.J.B. Yoo, LIONS: an AWGR-based low-latency optical switch for high-performance computing and data centers. *IEEE J. Select. Top. Quant. Electron.* **19**(2), 3600409 (2012)
28. Z. Cao, R. Proietti, S.J.B. Yoo, Hi-LION: hierarchical large-scale interconnection optical network with AWGRs. *J. Opt. Commun. Netw.* **7**(1), A97–A105 (2015)
29. T. Segawa, M. Nada, M. Nakamura, Y. Suzuki, R. Takahashi, An 8×8 broadcast-and-select optical switch based on monolithically integrated EAM-gate array. in *European Conf. Exhibition Optical Communication*, Paper TuT4.2, London, UK, 2013
30. R. Luitjen, R. Grzybowski, The OSMOSIS optical packet switch for supercomputers. in *Optical Fiber Communication Conference*, 2009
31. H. Wang, A. Wonfor, K.A. Williams, R.V. Penty, I.H. White, Demonstration of a lossless monolithic 16x16 QW SOA switch. in *35th European Conference on Optical Communication*, Vienna, 2009
32. T. Tanemura, I. Soganci, T. Oyama, T. Ohyama, S. Mino, K. Williams, N. Calabretta, H.J.S. Dorren, Y. Nakano, Large-capacity compact optical buffer based on InP integrated phased-array switch and coiled fiber delay lines, *IEEE/OSA. J. Lightwave Technol.* **29**(4), 396–402 (2011)
33. C. Raffaelli, K. Vlachos, N. Andriolli, D. Apostolopoulos, J. Buron, R. van Caenegem, G. Danilewicz, J.M. Finochietto, J. Garcia-Haro, D. Klonidis, M. O’Mahony, G. Maier, A. Pattavina, P. Pavon-Marino, S. Ruepp, M. Savi, M. Scaffardi, I. Tomkos, A. Tzanakaki, L. Wosinska, O. Zouraraki, F. Neri, Photonics in switching: architectures, systems and enabling technologies. *Comput. Netw.* **52**(10), 1873–1890 (2008)
34. R. Stabile, A. Albores-Mejia, A. Rohit, K.A. Williams, Integrated optical switch matrices for packet data networks. *Microsyst. Nanoeng.* **2**, 15042 (2016)
35. M. Glick, M. Dales, D. McAuley, T. Lin, K. Williams, R. Penty, I. White, SWIFT: a testbed with optically switched data paths for computing applications. In *Proceedings of 2005 7th International Conference Transparent Optical Networks*, 2005
36. C. Kachris, I. Tomkos, A Survey on Optical Interconnects for Data Centers. *IEEE Commun. Surv. Tut.* **14**(4), 1021–1036 (2012)
37. C. Kachris, K. Bergman, I. Tomkos, *Optical Interconnects for Future Data Center Networks* (Springer, New York, 2013)
38. J. Gripp, J. E. Simsarian, J.D. LeGrange, P. Bernasconi, D.T. Neilson, Photonic terabit routers: the IRIS project. in *Optical Fiber Communication Conference*, 2012.
39. K. Xi, Y.-H. Kao, H.J. Chao, A Petabit bufferless optical switch for data center networks, *Optical Interconnects for Future Data Center Networks*, Springer New York, 2013, pp. 135–154.
40. R. Proietti, Y. Yawei, Y. Runxiang, C.J. Nitta, V. Akella, C. Mineo, S.J.B. Yoo, Scalable optical interconnect architecture using AWGR-based TONAK LION switch with limited number of wavelengths. *J. Lightw. Technol.* **31**, 4087–4097 (2013)
41. R. Luitjen, C. Minkenberg, R. Hemenway, M. Sauer, R. Grzybowski, Viable opto-electronic HPC interconnect fabrics. in *Proceedings of Supercomputing 2005*, Seattle, 2005.
42. O. Liboiron-Ladouceur, A. Shacham, B.A. Small, B.G. Lee, H. Wang, C.P. Lai, A. Biberman, K. Bergman, The Data Vortex optical packet switched interconnection network. *J. Lightw. Technol.* **26**(13), 1777–1789 (2008)

43. W. Miao, F. Yan, N. Calabretta, Towards petabit/s all-optical flat data center networks based on WDM optical cross-connect switches with flow control. *J. Lightw. Technol.* **34**(17), 4066–4075 (2016)
44. M. Imran, M. Collier, P. Landais, K. Katrinis, Software-defined optical burst switching for HPC and cloud computing data centers. *J. Opt. Commun. Netw.* **8**(8), 610–620 (2016)
45. C.Y. Li, N. Deng, M. Li, Q. Xue, P.K.A. Wai, Performance analysis and experimental demonstration of a novel network architecture using optical burst rings for interpod communications in data centers. *IEEE J. Select. Top. Quant. Electron.* **19**(2), 3700508–3700508 (2013)
46. M. Fiorani, S. Aleksic, M. Casoni, Hybrid optical switching for data center networks. *J. Electric. Comput. Eng.* **2014**, 1–13 (2014)
47. M. Imran, M. Collier, P. Landais, K. Katrinis, HOSA: hybrid optical switch architecture for data center networks. in *12th ACM International Conference on Computing Frontiers*, 2015
48. K.-I. Kitayama, Y.-C. Huang, Y. Yoshida, R. Takahashi, T. Segawa, S. Ibrahim, T. Nakahara, Y. Suzuki, M. Hayashitani, Y. Hasegawa, Y. Mizukoshi, A. Hiramatsu, Torus-Topology Data Center Network Based on Optical Packet/Agile Circuit Switching with Intelligent Flow Management. *J. Lightw. Technol.* **33**(5), 1063–1071 (2015)
49. LIGHTNESS Project, [Online]. Available: <http://www.ict-lightness.eu/>
50. S. Yan, E. Hugues-Salas, V.J.F. Ranaño, Y. Shu, G.M. Saridis, B.R. Rofoee, Y. Yan, A. Peters, S. Jain, T. May-Smith, P. Petropoulos, D.J. Richardson, G. Zervas, D. Simeonidou, Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking. *J. Lightw. Technol.* **33**(8), 1586–1595 (2015)
51. Q. Huang, Y. Yeo, and L. Zhou, Optical burst-over-circuit switching for multi-granularity traffic in data centers. in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013*, paper OW3H.5, 2013
52. W. Miao, S. Di Lucente, J. Luo, H. Dorren, N. Calabretta, Low latency and efficient optical flow control for intra data center networks. *Opt. Express* **22**(1), 427–434 (2014)
53. W. Miao, F. Agraz, S. Peng, S. Spadaro, G. Bernini, J. Perello, G. Zervas, R. Nejabati, N. Ciulli, D. Simeonidou, H. Dorren, N. Calabretta, SDN-enabled OPS with QoS guarantee for reconfigurable virtual data center networks, *IEEE/OSA. J. Opt. Commun. Netw.* **7**(7), 634–643 (2015)
54. N. Calabretta, K. Williams, H. Dorren, Monolithically integrated WDM cross-connect switch for nanoseconds wavelength, space, and time switching. in *2015 European Conference on Optical Communication (ECOC)*, 2015
55. N. Calabretta, W. Miao, K. Mekonnen, K. Prifti, K. Williams, Monolithically integrated WDM cross-connect switch for high-performance optical data center networks. in *Optical Fiber Communication Conference 2017 (OFC 2017)*, paper Tu3F.1, 2017