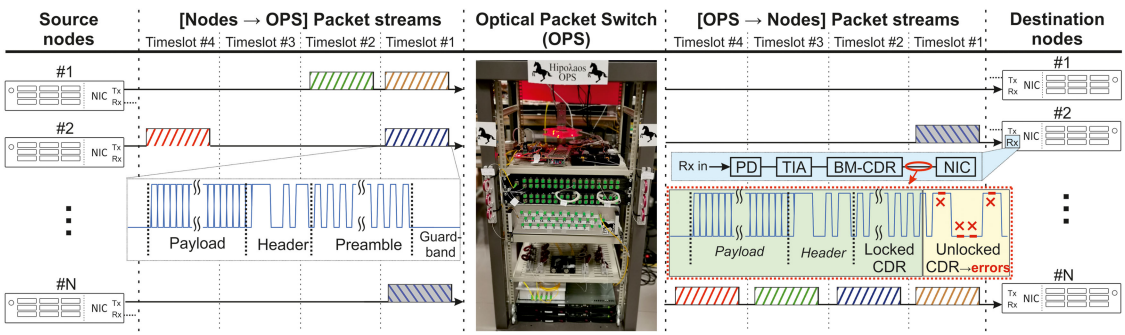
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

End-to-End 1024-Port Optical Packet Switching With 25 Gb/s Burst-Mode Reception for Data Centers

Volume 13, Number 3, June 2021

Nikolaos Terzenidis
Apostolos Tsakyridis
George Giamougiannis
Joris Van Kerrebrouck
Marijn Verbeke
Guy Torfs
Miltiadis Moralis-Pegios
Nikos Pleros



DOI: 10.1109/JPHOT.2021.3082642

End-to-End 1024-Port Optical Packet Switching With 25 Gb/s Burst-Mode Reception for Data Centers

Nikolaos Terzenidis,[1] **Apostolos Tsakyridis**,[1]
George Giamougiannis,[1] **Joris Van Kerrebrouck**,[2]
Marijn Verbeke,[2] **Guy Torfs**,[2]
Miltiadis Moralis-Pegios,[1] **and Nikos Pleros**[1]

[1]Department of Informatics, Center for Interdisciplinary Research and Innovation, Aristotle, University of Thessaloniki, Thessaloniki 54124, Greece
[2]IDLab, INTEC, Ghent University imec, Ghent B-9052, Belgium

DOI:10.1109/JPHOT.2021.3082642

Manuscript received April 20, 2021; revised May 15, 2021; accepted May 18, 2021. Date of publication May 21, 2021; date of current version June 4, 2021. The work was supported by the H2020 NEBULA (Contract No. 871658) and the 5GPPP Phase II Project 5G-PHOS (Contract No. 761989). Corresponding author: Nikolaos Terzenidis (e-mail: nterzeni@csd.auth.gr).

Abstract: Despite the fact that Optical Packet Switching (OPS) emerges as a promising solution for future Data Center (DC) networks, towards increasing capacity and radix, while retaining sub-μs latency performance, the requirement for ultra-fast burst-mode reception has been a serious restraining factor. We attempt to overcome this limitation and demonstrate, for the first time to our knowledge, an end-to-end optical packet switch link through the 1024-port 25.6 Tb/s Hipoλaos OPS, featuring burst-mode reception with $<$50 ns locking time. The switch performance for unicast traffic is evaluated via Bit-Error-Rate measurements and error-free performance at 10^{-9} is reported for all validated port-combinations, with a mean power penalty of 2.88 dB. Moreover, multicast flows from two different ports of the switch were successfully received validating the architecture's credentials for efficient multicast packet delivery. Taking one step further towards a realistic evaluation of an OPS-enabled DC, a simulation analysis was conducted, proving that low-latency performance, including the burst-mode reception time-overhead, can be successfully realized in a Hipoλaos-switched DC with up-to 100% throughput for a variety of traffic profiles.

Index Terms: All-digital clock and data recovery (AD-CDR), burst mode reception, Data Centers, optical multicast, Optical Buffering, Optical Packet Switching, optical switches.

1. Introduction

At the dawning of the exaflop era, the growing demand for ubiquitous high-bandwidth processing and cloud computing applications, along with the boost of IoT and big-data analytics, has stimulated an unprecedented increase on the data traffic residing within Data Centers (DC) [1]. At the same time, an analysis on the implications of the covid-19 pandemic on Internet, revealed a further traffic upsurge by 20% [2], [3], within just weeks, as a result of the confinement measures and the huge uptake in remote-working, -education, e-commerce and e-entertainment. This relentless growth of data traffic is currently pushing DC interconnects to their capacity limits, while novel DC architectures pursued by cloud operators, are simultaneously pushing the envelope on other network metrics as well, such as latency and high-radix connectivity. In this context rack-scale

disaggregation [4], [5], that is endorsed as a promising paradigm shift for tackling the resource underutilization and energy problems through the highly granular synergy of standalone CPU, GPU and memory components, urges for Tb/s capacity, sub-μs latency and hundreds of port connectivity [6].

As a result of this challenging set of performance requirements, the most crucial part of the DC network, i.e., its switching fabric, has been brought at the epicenter of the attention. The DC switching landscape, currently dominated by electronic packet switches, has been largely defined by the switching ASIC and SerDes capabilities, that are expected to hit up-to 25.6 Tb/s capacity [7], [8] and 112 Gb/s line-rates [9] by 2021. The development of 25.6 Tb/s fabrics has mainly built upon the advances on 7nm CMOS processing technologies, along with a novel technology trend that emerged as an alternative to the bulky and energy-consuming frontpanel pluggable optics, namely the co-packaging of high-speed optical engines with switching ASICs [10]. However, scaling beyond this capacity peak, is expected to face severe challenges, mainly due to the high-power Long-Reach (LR) SerDes and DSP interfaces required to route 224G signals from the ASIC package to the optics site. To this end, scale-out designs where multiple ASICs are interconnected in a Folded-Clos layout to form higher capacity non-blocking switches, have already been demonstrated [11], but this capacity boost comes at the cost of increased latency, energy and cost figures.

All this indicates that conventional electronic switches are approaching the end of the roadmap that was followed for many years and are most probably on the verge towards a new paradigm for future capacity scalings. At the same time, this may designate an opportunity for optical switch technologies to take advantage of the significant developments pursued during the last years and transform into a tangible DC solution. High-radix Optical Circuit Switches (OCS), for example, proved successful in scaling port-number and capacity while retaining line-rate transparency, but suffer from msec reconfiguration times [12], [13]. On the other hand, a number of novel Optical Packet Switch (OPS) architectures has been proposed, leveraging space- and wavelength-routing [14]–[20], or even coherent technologies [21], [22], in order to combine high-radix and high-capacity connectivity with low latency operation. Within this frame, we have also recently introduced the Hipoλaos OPS architecture that supports up-to 1024-port and 25.6 Tb/s capacity configurations in conjunction with sub-μsec latency values [23]–[28]. Still, all these demonstrations have been limited to the realization of the optical forwarding plane assuming synchronized source and destination nodes and ignoring the requirement for asynchronous packet traffic between the OPS network nodes. This would in turn necessitate the employment of a Burst-Mode Clock & Data Recovery (BM-CDR) circuitry with ns-scale locking time in order to handle the phase-mismatch of the packets emerging from different nodes. While it has long become apparent that the realization of BM-CDR circuits, supporting both high-datarate operation and fast locking times, would face several challenges, during the past years a number of promising demonstrations has been reported [29]–[32], achieving down-to 6.8 ns locking times, for up-to 56 Gb/s rates [30]. However, the advances in BM-CDR circuits have not yet been reflected in respective OPS architectures, with demonstrations showcasing end-to-end asynchronous packet operation just recently reported [33], [34], but with limited port count and capacities well below the 25.6 Tb/s target for next-generation switches.

In this paper, we extend our previous work [27], [35] and demonstrate for the first time, to the best of our knowledge, end-to-end optical packet switching supported by BM-CDR functionality over the recently reported 1024-port Hipoλaos OPS layout [27]. We present, in this way, a 25.6 Tb/s capacity OPS setup where an all-digital BM-CDR circuit at the receiver end allows the establishment of packet-level communication between non-synchronized source and destination nodes. We demonstrate successful reception of 25 Gb/s optical burst-data packets that were routed through a fully functional Hipoλaos Plane, followed by a 32 $\times$ 32 AWGR and were finally received by a BM-CDR [29] with just $<$50 ns locking time. The unicast performance of the OPS architecture was assessed via Bit Error Rate (BER) measurements, revealing error-free operation for the whole range of switch output ports with a mean power penalty of 2.88 dB. True end-to-end OPS was demonstrated also in multicasting packet operation by exploiting Hipoλaos' AWGR-enabled multicast credentials [36], successfully transmitting 2 optical multicast flows to 2 different switch

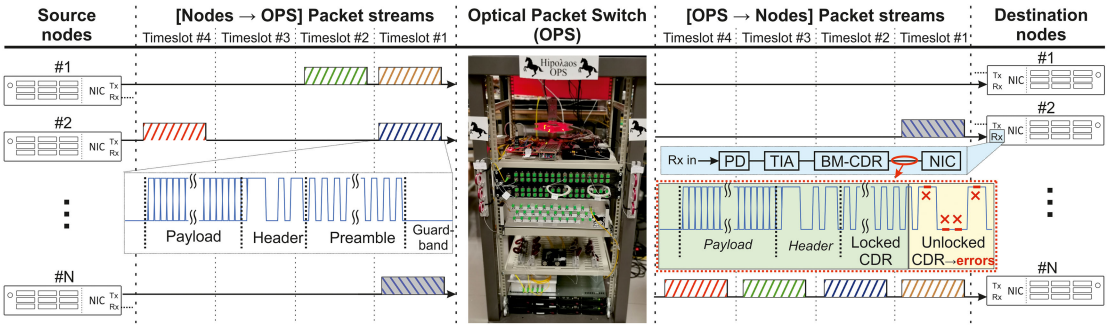


Fig. 1. Forwarding operation example and optical frame structure for an Optical Packet Switched network with #N nodes. In the middle, photo of the Hipoλaos prototype developed for our evaluation.

outputs. The throughput and latency metrics of the BM-supported Hipoλaos OPS have been evaluated through in-depth network simulations that take into account respective experimental results, revealing that low-latency performance, as required by disaggregated DCs, can be successfully realized in a Hipoλaos-switched DC with up-to 100% throughput for a variety of traffic profiles.

2. Optical Packet Switching (OPS) Schemes and the Hipoλaos Approach

The path towards the deployment of Optical Packet Switches in DC networks has been covered, until now, with thorns and spikes, due to the technological challenges encountered in critical technologies such as BM-CDR circuits, along with the immaturity of respective protocols that will support this deployment. These challenges in the protocol stack mainly stem from the working principle of OPS fabrics.

2.1 Principle of Operation and Deployment of OPS in DC Networks

The main difference of OPS schemes compared to the electronic packet switching networks lies in the absence of continuous data transmission between any two pair of nodes. It should be noted that in conventional electronic networks every server/node is statically connected either with a switch, or directly with another node and a continuous data transmission is established between the respective Tx/Rx circuits, according to the protocol used. This is not the case, however, in OPS networks where the connection between 2 nodes can be disrupted after a packet transmission, as the optical switch is reconfigured and new optical paths are established. In this context, time-slotted operation has been favored as a solution that tackles some of the complications arising from the discontinuous data transmission, by determining the exact time that each node begins the packet transmission in order to avoid collisions and eventually increase the network's throughput. At the same time, the lack of continuous data results in packets arriving from different source nodes and having inevitably clock phase mismatches, forcing the use of a receiver CDR circuit that needs to resettle on a per packet basis. It is, of course, vital to keep this settling time overhead as short as possible in order to ensure the minimum hit in the network's throughput performance. Considering the current line-rates of DC interconnects and the distribution of packet sizes [37], the requirement for a fast settling time tightens to just a few ns. The employment however of burst-mode reception enables at the same time more efficient network links, in terms of energy, since parts of the network can operate in low-power mode when no actual traffic is present, avoiding the need for dummy packet transmission in order to keep the link synced.

A schematic representation of the principle of operation of OPS networks and the need for BM reception is illustrated in Fig. 1, along with an example of the forwarding operation for 5 packets from different nodes. Taking into consideration 4 distinct timeslots, the packet streams at the output of #N servers are depicted on the left side of Fig. 1. The timeslots are common for the entire network

and each server is allowed to start the packet transmission at the beginning of the timeslot, with server #1 in this example transmitting packets at slots #1 and #2, server #2 transmitting at slots #1 and #4 respectively and server #N at slot #1. In addition to the packet streams, the generic frame structure employed in OPS networks is depicted as an inset below the second stream. The frame starts and ends with an idle guardband that mainly accounts for the rise/fall times of the optical switching elements incorporated in an OPS layout. Subsequently, in order to accommodate the settling process of the BM-CDR, a long sequence of alternating "1" and "0", the so-called "preamble", is added before the packet header ensuring that enough fast transitions are present for the BM-CDR to extract the correct clock frequency. Finally, the header and payload parts of the frame follow the preamble, with a sufficiently short, compared to the preamble, CDR settling time being required so as to ensure the header and payload data are correctly recovered.

Moving to the right side of Fig. 1, the resulting packet streams at the outputs of the respective OPS fabric are illustrated, when considering that the packet from servers #1 and #2 are destined towards server #N, and the packet from server #N is destined towards server #2. For this specific example, the packet scheduling is based on the Hipoλaos contention resolution scheme that manages to forward packets received on the same timeslot and destined to different nodes by exploiting the delay-line-based buffering capabilities of the switch, as described in detail in [26]. As a result, the first packet from server #1 is forwarded on timeslot #1, while the first packet from server #2 is buffered and forwarded on the next timeslot (#2) considering for example a higher priority for server #1. The second packet from server #1 is also buffered for one timeslot duration in order to avoid collision, while the second packet from server #2 is directly forwarded towards its destination. At this point, it should be noted that alternative contention resolution schemes have been also employed in OPS demonstrations, such as drop and retransmit [18], inducing however increased latency and decreased throughput performance.

The inset referring to the Rx part of Destination node #2 illustrates in more detail the constituent Rx functional blocks explaining also the working principle of the BM-CDR. As long as the CDR is unlocked (illustrated with yellow background color), errors occur at received bits of the preamble. From the timepoint that the CDR is locked to the data, the rest of the packet stream is successfully recovered, showcasing the necessity for a fast settling time that has to be shorter from the preamble duration. As can be easily perceived, the shorter the BM-CDR locking time, the less throughput hit that occurs due to the frame preamble. Moreover, as current and future network protocols feature short data-bursts or radio on/off functionality, BM-CDR devices with ns-scale settling are being actively developed supporting rates up-to or above 25 Gb/s in order to keep pace with the respective protocols' lanerates [30], [38], [36], [39].To achieve this datarate scaling researchers have relied on different techniques and technologies, such as advanced FinFet CMOS platforms, additional equalizers that compensate the bandwidth limitations introduced by optical (de)modulators, or even clean clock sources originating from an external source or PLL. On the other hand, in order to decrease the CDR locking time, additional circuits are added, such as two types of phase detectors towards satisfying the contradicting requirements of fast-locking and stable, low jitter operation when phase lock is obtained [30], [38]. By splitting these two functionalities in two dedicated blocks, i.e., a burst-mode phase detector and a traditional bang-bang phase detector, both requirements can be achieved in one design, while a finite state machine controls which phase detector is active. Finally, by introducing metastability detection [39], the meta-stable conditions, that arise due to the uncorrelation of the recovered clock and the incoming burst, are avoided and the maximum lock time is reduced.

Taking into consideration that the maturity of BM-CDR solutions with ns-locking time has advanced substantially over the years, the actual deployment of an OPS-enabled system could make a leap forward towards innovative DC architectures. The incorporation of OPS in a DC environment could potentially bring significant performance and energy/cost benefits, as indicated in several studies [19], [40], provided that some additional, yet limited, modifications are enforced in the protocol stack, as has been the case with the recent admission of OCS solutions, where SDN orchestration ensured interoperability with already deployed equipment.

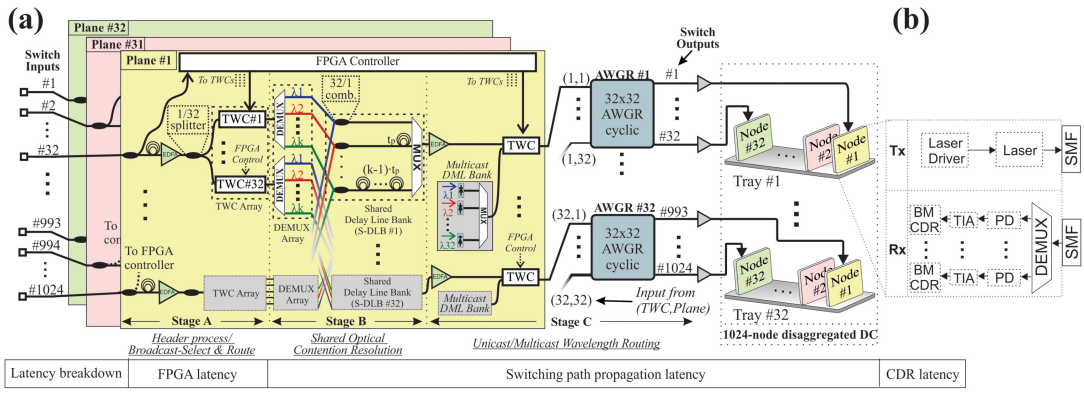


Fig. 2. (a) Generic layout of the 1024x1024 Hipoλaos switch architecture, (b) Node transceiver interface, SMF: Single-Mode Fiber, TIA: Transimpedance Amplifier.

2.2 The Hipoλaos Data- and Control-Plane Architecture for End-to-End OPS Links

The Hipoλaos OPS architecture has been developed as a disaggregation-oriented switch fabric, able to offer scalability in all three crucial performance factors: capacity, latency and port-count. It comprises an optimized Spanke layout overcoming the radix scalability limitations of conventional OPS layouts, by distributing the control and switching functions in separate clusters, named as Planes. The Planes forward traffic via a Broadcast and Select (B&S) scheme realized through wavelength converters (WC), while the incorporation of optical feed forward buffering enables the realization of high-throughput and low-latency performance. Despite the fact that the utilization of WC renders as quite challenging the employment of PAM-4 signaling, the Hipoλaos switch manages to achieve high-data rates, through a differentially-biased scheme [41] at every WC stage that allows for NRZ operation up-to 40 Gb/s [42]. Finally, the architecture exploits at its last stage, WCs along with cyclic AWGRs, in order to extend the switch port-count through a collision-less WDM routing mechanism.

The 1024-port Hipoλaos OPS is illustrated in Fig. 2(a), interconnecting a DC rack system, that is composed of #32 trays with each tray hosting #32 nodes. The switch is organized in 32 Planes with 32 ports/Plane, followed by 32 32 $\times$ 32 AWGR devices, for traffic aggregation and wavelength routing, respectively. The main latency contributing factors on the Hipoλaos architecture are depicted at the bottom of Fig. 2 in accordance with the Stage of the switch that each latency figure occurs. The switch architecture follows the design principles described in detail in [24], where header processing was performed by an FPGA, contention resolution by Shared Delay Line Banks (S-DLB) and wavelength routing to the destination port was achieved by tunable wavelength converters in conjunction with AWGRs. On every Plane, the FPGA controller undertakes the crucial role of orchestrating the forwarding operation throughout the switching Stages, making sure that packet collisions at Stage B of the switch are avoided by allowing on each timeslot only one of the available TWCs to convert signals on a specific wavelength [24]. More specifically, a routing table lookup is carried out for every incoming packet in order to determine the S-DLB to be forwarded. Subsequently, the controller examines the current state of the switch so as to check for available buffer delay lines on that specific S-DLB. In the case of at least 1 free buffer, the first available buffer is marked as occupied, the switch state is updated accordingly and an activation signal is generated towards the appropriate WC, triggering the conversion to the wavelength that corresponds to this buffer. In the case of no available buffers, the packet is dropped at the switch, by means of not activating any WC. Finally, the controller implements a Quality of Service (QoS) scheduler that grants buffer access to the packet with the highest priority, when considering the case of multiple packets requesting simultaneously buffers from the same S-DLB.

Regarding the supported packet forwarding modes, the Hipoλaos architecture has been developed in order to facilitate both unicast and multicast traffic, enabling this way a constantly

increasing number of DC workloads that inherently require multicast traffic delivery as part of their communication patterns. Taking advantage of the Hipoλaos' layout, efficient optical multicasting is realized through multi-λ-routing at its final AWGR-based switching stage, with the addition of a multicast laser bank at every Plane's Output, as described in detail in [36]. A photo of the actual developed Hipoλaos switch prototype can be seen in the middle of Fig. 1, with the first rack level including the FPGA control-plane along with the Tunable WCs that are realized by SOA-MZI devices. The second rack level hosts the S-DLB comprising three fiber delay lines that offer up to three packet buffering capacity. The third rack level comprises a 32x32 AWGR, undertaking the wavelength routing, while the fourth rack level includes the EDFAs and the other optical components required (polarization controllers, optical delay lines etc.).

However, constructing a 1024-port optical switch using discrete fiber-interconnected components raises a number of challenges associated with size and complexity. As a result, the adoption of integrated photonic circuitry, at least for the deployment of footprint-critical individual subsystems, can provide a viable path towards the realization of practical Hipoλaos switch implementations. In this context, the mature Indium Phosphide (InP) platform appears to have an advantage, since all the components required by the Hipoλaos architecture have been already demonstrated. The cyclic wavelength-routed functionality can be provided by AWGR configurations [43], [42], the (de)multiplexing functionality can be provided by AWGs [44], buffering can be provided by integrated delay lines [45], while the amplifiers and WC arrays required both in Stages A and C of the Hipoλaos switch can rely on SOA and SOA-MZI configurations [46]. At the same time, the InP platform has already proven its credentials in the integration of a high number of components on the same PIC [46]. On a different aspect, the um-SOI platform could also form a promising candidate for integrating delay line banks, cyclic wavelength-routing blocks and WC arrays on the same chip, but significant progress is required in each individual technology block until reaching the maturity level of the InP platform. In this context, a small-scale layout of the Hipoλaos architecture have been already validated with the delay-line components implemented in the um-SOI platform, along with an Echelle grating router to provide the wavelength routing functionality [23]. In order to provide even greater port-scalability the Hipoλaos architecture can be deployed in cascaded configurations [24], in line with the electrical switch configurations, concluding to scaled-up layouts that provide significantly higher connectivity.

Finally, the respective interface of a node in a Hipoλaos-switched network is depicted in Fig. 2(b). Each node is connected to the switch by a pair of fibers, utilizing a single optical link at a fixed wavelength on the transmitter side. At the receiver side every node is considered to have the capability of receiving multiple packets at different wavelengths by employing a demultiplexer along with a photodiode (PD) and a transimpedance amplifier (TIA). The integration of a BM-CDR at the TIA output is the last step that expands the capabilities of the architecture from a lab-scale prototype towards a system credible for evaluation in realistic DC conditions. It is important to note that the Hipoλaos architecture can be in principle extended towards supporting a simpler single-channel and colorless receiver architecture at each node, by incorporating an additional distributed contention resolution mechanism at its outgoing stage.

3. Experimental Demonstration of End-to-End Optical Switching With Burst-Mode Reception At 25 Gb/s

In order to demonstrate an end-to-end true packet switched link through the 1024-port Hipoλaos architecture operating with 25 Gb/s burst-mode optical packets, a fully functional Plane has been experimentally validated, comprising an S-DLB with three delay lines (direct, tp, 2tp), interconnected to a 32 $\times$ 32 AWGR along with a BM-CDR circuit. Our evaluation followed a stepwise approach starting with the validation of successful switching, utilizing unicast burst packets, followed by a multicast packet delivery experiment and finally BER performance assessment for the entire range of switch output ports.

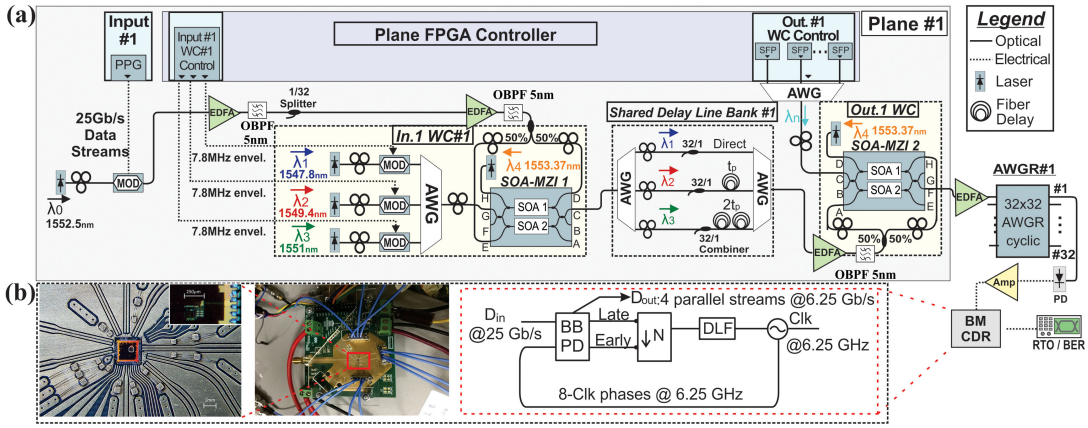


Fig. 3. (a) Experimental setup for the evaluation of Hipoλaos architecture with a BM-CDR, (b) Layout and photo of the BM-CDR.

3.1 Evaluation With Unicast Burst-Mode Packets

The experimental setup used for the evaluation of the proposed switch is illustrated in Fig. 3(a). A Pulse Pattern Generator (PPG) was used for data generation at 25 Gb/s while a Xilinx FPGA was employed for controlling the switching functionalities with a processing latency of 97.28 ns [27]. Due to the fact that a second BM-CDR device was not available for packet reception at the FPGA, one FPGA transceiver was employed to generate the exact same packets, transmitting them through an internal loopback at the Rx side of the transceiver for header processing and control signal generation. Moreover, in order to match the packet slots generated by the FPGA and the PPG, a common reference clock was distributed to both devices in order to synchronize them. A CW beam at $\lambda 0 = 1552.5$ nm was launched into a LiNbO3 modulator driven by the PPG to produce 26.112 μs NRZ data packets at 25 Gb/s, comprising a 153.6 ns preamble, a 25.9584 μs payload and an inter-packet guardband of 153.6 ns. The guardband was chosen in accordance with the BM-CDR operating conditions, making sure that the gap between two consecutive packets is long enough so that at the beginning of every packet the CDR is no longer in lock with the generator [29]. However, the Hipoλaos architecture has been already validated with inter-packet guardband of just 3.5 ns [24]. The payload size was defined as the maximum supported by the employed PPG device, in order to stress the capability of the employed BM-CDR to retain its frequency-lock even for extremely long packet sizes.

The modulated signal was injected to an EDFA for amplification, then filtered by a 5 nm Optical Bandpass Filter (OBPF) and fed to the Hipoλaos switch where a 1/32 splitter was used to emulate the actual losses of a 32 $\times$ 32 Plane within a 1024 $\times$ 1024 switch. The resulting signal after a second stage of amplification and filtering, was launched to a 50/50 splitter and split into two identical signals feeding the ports D and E of the SOA-MZI #1, realizing the differentially-biased scheme, whereas a CW beam at $\lambda 4 = 1553.37$ nm was injected into port H as an auxiliary holding beam, forcing the SOAs to operate in their deeply saturated regime [47], [48]. At this point it should be noted that the FPGA in every timeslot, according to the state of the S-DLB and the free buffering lines, determines the desirable TWC output wavelength and subsequently "activates" only one of the wavelength channels in the laser bank by providing an envelope signal in the modulator of this specific channel. The envelope signal is practically represented by a sequence of consecutive "1" bits with duration of 26.115 μs, that after being transcripted on the respective optical channel is injected on port G, defining this way the conversion wavelength. The output signal from port C of SOA-MZI #1 was then injected to the S-DLB #1, where 1/32 combiners were used to emulate the 1024 $\times$ 1024 combining ratio. The signal exiting the S-DLB, after being amplified by an EDFA and filtered by a 5 nm OBPF, was injected into the Out. 1 WC and at ports A and H of the SOA-MZI #2. A CW holding beam at $\lambda 4 = 1553.37$nm was launched into port D, while SFP modules emitting at different wavelengths were used to generate the envelope signal (at λn) injected at port C. The

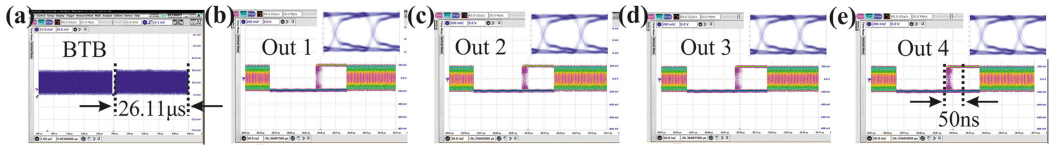


Fig. 4. Unicast Experimental results. (a) Modulated output signal (BTB) @25 Gb/s with 153.6 ns guardband, (b)–(e) the 4 outputs of the BM-CDR with ∼50 ns worst-case locking time and respective eye diagrams, X-axis scale for (a): 1 μs/div, (b)–(e): 50 ns/div, Y-axis scale for (a)–(e): 200 mV/div.

optical signal, converted at λn, that emerged at output G of SOA-MZI #2, after amplification, was injected to input port #1 of the AWGR and then it was sequentially switched to each AWGR output by enabling a different SFP+ at the FPGA. It should be noted that only one SFP+ module was activated at any given time by the FPGA controller, since unicast operation was targeting for this evaluation scenario. The signal at the AWGR outputs was transferred to the electrical domain by a photodiode and after amplification, was launched into the BM-CDR to handle the variation in optical power and phase-mismatch on a packet-by-packet basis.

Fig. 3(b) depicts the BM-CDR that was employed in our demonstration. It comprises a 25 Gb/s all-digital clock and data recovery (AD-CDR) circuit implemented in a 40 nm CMOS process, incorporating a bang-bang phase detector (BB-PD), a subsampling block, a digital loop filter (DLF) and a digitally controlled oscillator (DCO). Thanks to the all-digital implementation, the circuit occupies a compact active area of 0.050mm2 and consumes only 46 mW, resulting in an energy-per-bit of just 1.8 pJ/bit at 25Gb/s rates. To this end, considering also the energy efficiency of the Hipoλaos switch [27], the total efficiency of a Hipolaos end-to-end link with BM-CDR reception would result in ∼222 pJ/bit at 25 Gb/s rates. Its wide-band lock loop is capable of phase-locking an incoming burst in 37.5 ns and requires neither a system clock nor a start-of-burst signal. In order to relax the circuit requirements, the output of the BM-CDR was demultiplexed in 4 data signals at a clock rate of 6.25 GHz that were evaluated separately in this scenario. The AD BM-CDR has a minimum input driving voltage requirement of ∼300 mVpp for error-free operation at 10^{-12}. Taking also into account the PD and amplifier that were employed in our evaluation, the minimum optical power at the BM-CDR front-end was ∼2 dBm. A more detailed description of the BM-CDR can be found in [29].

The experimental results for 25 Gb/s unicast burst-mode packets routed to output port #1 of the switch, are presented in Fig. 4(a)–(e). More specifically, Fig. 4(a) depicts the output trace of the modulator (BTB) at 25 Gb/s, showing two consecutive 26.112 μs long data packets that are subsequently routed through the Hipoλaos switch and received at the BM-CDR circuit. Fig. 4(b)–(e) illustrate the obtained traces from a real-time scope and eye-diagrams from the 4 $\times$ 6.25 Gb/s demultiplexed signals at the BM-CDR output. At this point it should be mentioned that when the "1010 …" preamble is demultiplexed by four, the output should stay either low or high [29]. In case a transition occurs during this preamble, an error has occurred. As a consequence, the settling time of the AD-CDR can be measured by recording when a transition occurs in the subsampled preamble at the output of the AD-CDR. In our evaluation, after capturing 500.000 packets the worst locking time was measured at ∼50 ns. That corresponds to an increase of only 12.5 ns compared to the lowest lock time achieved by the same device with packets received directly from a generator [29]. In this realm, the minimum preamble size required for successful CDR locking is 50 ns, while there is no constraint on the payload size, as long as it is up-to 26.112 μs. Considering an end-to-end latency value of 182.28 ns for our prototype, the BM-CDR contribution to the total latency is 27.5%. Finally, the captured traces from all CDR outputs were subsequently evaluated off-line via the Matlab software, revealing error free operation for 106 bits.

3.2 Evaluation of Multicast Packet Delivery

During the next step of our evaluation a multicast laser bank was emulated in the experimental testbed of Fig 3(a) in order to perform a multicast performance analysis. While the experimental

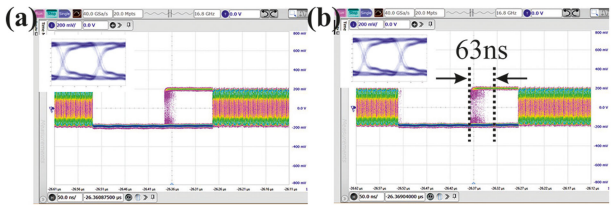


Fig. 5. Multicast Experimental results. (a) Captured traces from output#1 of AWGR, (b) Captured traces from output#2 of AWGR X-axis scale: 50 ns/div, Y-axis scale: 200 mV/div.

setup remained the same, the only difference occurred in the input signal of the SOA-MZI #2. In order to create identical multicast copies of the same packet in two different wavelength channels, two SFP+ modules were simultaneously activated at the output #1 WC control block, producing 25.9664 μs envelopes at λn1 = 1540.56 nm and λn2 = 1541.44 nm. After combining the modulated envelopes in an AWG mux, they were launched as input signal on the SOA-MZI #2 through port C. The optical signal originating from stage B, comprised the same 26.112 μs long 25 Gb/s NRZ packets, described in Section 3(a) as the output of Stage B. This optical packet stream is copied on the multiwavelength signal of port C, generating two multicast copies of the same packet at λn1 and λn2 that emerge at output F of SOA-MZI #2. The WDM stream, after being amplified on an EDFA, is injected at port #1 of the AWGR. The cyclic AWGR routes each wavelength channel to output ports #1 and #2, achieving this way multicast delivery of the same optical packet towards two different destination nodes.

Fig. 5(a), (b) illustrate the respective captured traces in the real-time scope, along with eye-diagrams from output #1 of the BM-CDR circuit. As can be observed, the CDR for both channels successfully locks before the end of the preamble. The worst-case locking time was observed for AWGR output #2 with a value slightly higher than the unicast scenario at ~63ns. It is worth noting that this is, to the best of our knowledge, the first successful reception of burst-mode optical multicast packets at 25Gb/s.

3.3 Scalability to 1024 Switching Ports

In order to validate the scalability of the Hipoλaos architecture towards a 1024-node OPS-switched system a more in-depth performance evaluation has been performed for all possible routing paths through the Hipoλaos switch, i.e., for all possible wavelength combinations that correspond to the 32 AWGR ports. However, due to the fact that the received 25 Gb/s signal is demultiplexed at four CDR outputs, there was an objective difficulty of performing BER measurements with the employed pattern on commercial BER testing equipment. As a consequence, the BER evaluation was performed with a continuous 25 Gb/s PRBS7 pattern, since PRBS signals still demux back down to PRBS signals.

Fig. 6(a) shows at the top the optical eye diagram at the modulator output, along with its electrical spectrum (middle) at the BM-CDR output revealing successful locking at 6.25 GHz, with the respective electrical eye diagram at 6.25 Gb/s at one of the 4 outputs of the BM-CDR being depicted at the bottom. Fig. 6(b)–(i) depict the eye diagrams (top) and spectrums of the optical signal (middle) after being routed to outputs #1, #5, #9, #11, #16, #25, #28 and #32 of the AWGR, along with the corresponding electrical eye diagrams obtained at output #1 by the BM-CDR (bottom). In this case, different SFP+ were utilized to evaluate the broadband operation of the switch, by selecting equally distributed wavelengths from 1540.56 nm to 1565.4 nm to cover the wavelength range supported by the AWGR channels. The same performance was observed also for the rest of the AWGR ports.

Finally, the BER performance was measured for 8 AWGR output ports distributed along the complete spectral range supported by the AWGR device, so as to validate its capability to perform successfully along all possible Hipoλaos i/o port combinations. BER measurements at every output

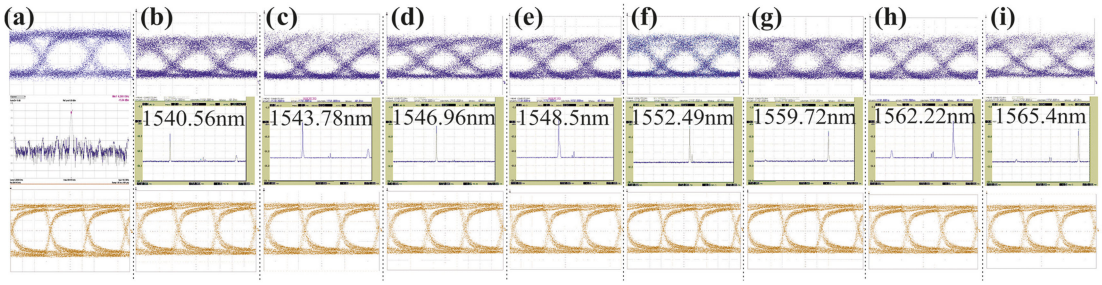


Fig. 6. Scaling to 1024 ports Experimental results. (a) BTB optical eye diagram (top) - electrical spectrum of the BM-CDR locked @6.25 GHz (mid) -electrical eye diagram of a BM-CDR output (bottom), (b)–(i) Optical eye diagrams (top) and optical spectrums (mid) of AWGR outputs #1, #5, #9, #11, #16, #25, #28 and #32 - electrical eye diagrams of a BM-CDR output (bottom). X-axis scale for (a)–(i): optical eye diagrams: 10ps/div, electrical spectrum 10 MHz/div, optical spectrums: 4 nm/div, electrical eye diagrams: 50 ps/div, Y-axis scale for (a)–(i): optical eye diagrams: 5 mV/div, electrical and optical spectrums 10 dB/div.

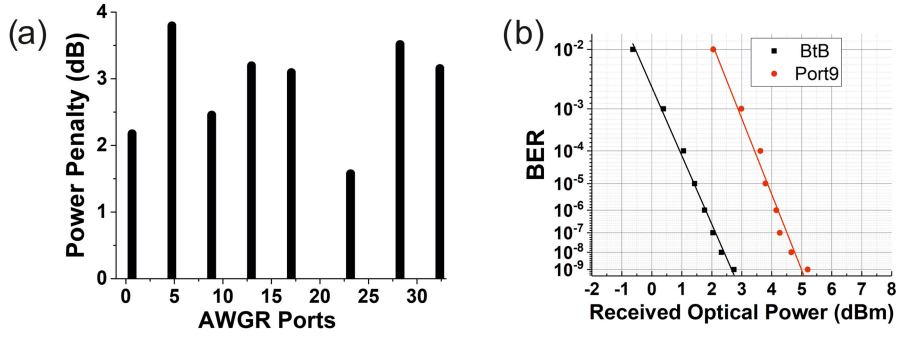


Fig. 7. (a) Power Penalty at error-free operation for 8-channels of the AWGR, (b) BER measurements for packets routed to AWGR output port 9.

port were performed by evaluating the received signal from output #1 of the BM-CDR. Error-free operation (10^{-9}) was obtained for all channels, with the power penalty (PP) values for these 8-indicative output-channels being depicted in Fig. 7(a). The power penalty here is defined as the extra power required in the TWCs, while preserving the same operational conditions at the residual components, to obtain error-free operation in the receiver site, with the mean PP compared to the BtB scenario being 2.88 dB while the max value was $<$4 dB. Fig. 7(b) illustrates the illustrates the BER curve obtained for AWGR channel #9, revealing an average power penalty of 2.46 dB against the BtB measurement.

4. Performance Analysis of 25.6 Tb/s Optically-Connected Data Center

Taking one step further towards the evaluation of an OPS-interconnected Data Center in terms of throughput and packet end-to-end latency, we have proceeded with the development of an OMNeT++ [49] simulation model. For our evaluation the network nodes/servers were modelled taking into account the BM-CDR requirements on the receive side of their transceivers, while the switch model followed the principle of operation of the Hipoλaos OPS, as described in Section II (a), and in more detail in [24].

The model was further refined to the requirements of OPS networks by defining a slotted network operation, with the traffic generating nodes transmitting their packets at the beginning of the respective slots. The traffic profiles that were implemented in the simulation model were

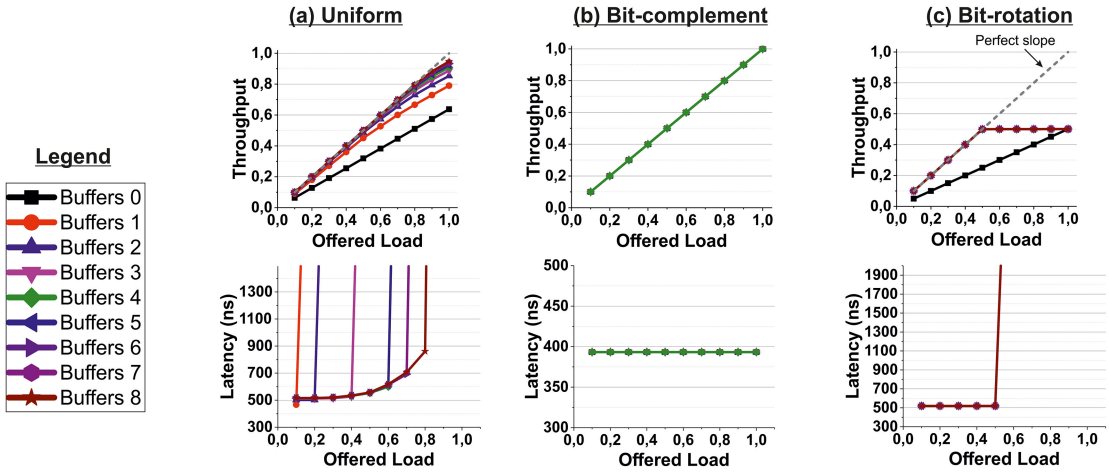


Fig. 8. Simulation performance analysis for a 1024-port Hipoλaos-switched network at 25 Gb/s linerate. Throughput and Mean latency results vs offered load for: (a) Random Uniform, (b) Bit-Complement, (c) Bit-Rotation.

based on 3 well-known synthetic patterns [50] for unicast traffic evaluation: 1) Random-Uniform, 2) Bit-Complement, 3) Bit-Rotation. It should be noted that the Bit-Complement and Bit-Rotation patterns generate permutation traffic in order to stress the topology with each source node sending all of its traffic to a single destination node, according to different permutation functions. More specifically, the destination node address is computed by permuting and selectively complementing the bits of the source node address.

For our analysis, a DC network with 1024 computing nodes-servers was evaluated, with the linerate between the server transceivers defined at 25 Gbps, in accordance with our experimental evaluation. The DC system was modelled according to the schematic representation of Fig. 2. The nodes were distributed among 32 trays, with 32 nodes incorporated in each tray, generating this way a 2-dimensional topology. At the same time, each one of the 1024 nodes was assigned a sequential address in the range from 0-1023, with the address assignment based on the nodes position in the topology. As an example, the first node on the first tray (position 11) was assigned address “0”, the second node on the first tray (position 12) address “1” and so on. A re-transmission mechanism was considered in the simulation model, where the dropped packets at the switch were re-added on the source node queue and re-transmitted with the latency statistic calculated using the original packet creation time.

The Hipoλaos switch featured 32 Planes, along with 32 $\times$ 32 AWGRs, resembling the architecture depicted in Fig. 2. As far as the frame size is concerned, we have considered the median of the Ethernet frame (800 byte) for the payload, taking into account the studies in [37], [51] where the majority of flows within DCs have been found to utilize frame sizes of 576, 800 and 1500 bytes. In order to account for the preamble, header and guardband parts in each frame, the frame size was expanded to 950 bytes. Considering these values, the overhead of the preamble and guardband would be $\sim$17% and the network traffic efficiency for this specific case would be $\sim$83%. An FPGA processing latency of 97.28 ns was considered in the simulations, complying with the respective experimentally measured performance of the FPGA control plane, while the propagation latency for the various optical components of the switch, excluding the buffer delay lines, was modelled to 35 ns.

The performance of the switch architecture was evaluated for all 3 traffic patterns, as a function of the available buffer delay lines per S-DLB block, with the number of buffers ranging from 0 up-to 8. Fig. 8(a) presents the respective throughput and mean latency results versus the offered load for the uniform pattern. The throughput increases almost linearly until $\sim$80% load, with the maximum value reaching $\sim$96% when utilizing 8 buffers. The latency in all buffer-cases ranges between 460

and 860 ns before the network saturation point that depending on the number of buffers arises at a different offered-load setting. As an example, in the case of 8 buffers the saturation point arises at ∼80% and from that point on the latency increases significantly and becomes practically unbounded. It should be noted that in the case of 0 buffers the network is saturated even at 10% load. Fig. 8(b) presents the respective results for the Bit-complement pattern, where it is observed that 100% throughput is obtained for all different buffering options, while the mean packet delay is maintained at 390 ns. Considering the described network topology and address assignment policy, this behavior can be explained by the principle of operation of the Hipoλaos switch, since for this pattern the packets in every timeslot are evenly distributed at the outputs of the respective Planes. Moving to the Bit-Rotation pattern in Fig. 8(c), a linear throughput increase is observed until 50% load, with the maximum throughput reaching ∼50% in all cases, while the mean latency remains at 518 ns until 50% load is reached and then becomes unbounded due to the network saturation. This performance owes to the inherent properties of this traffic profile where packets in every timeslot are destined in pairs towards the same Plane output, yielding the 50% load as the ultimate limit for the switch capability to resolve contention.

5. Conclusion

Despite the fact that hyperscale DC switches have been able to push the capacity envelope up-to 25.6 Tb/s, scaling beyond that point is expected to build upon novel concepts with Optical packet switching promising a viable solution towards increasing capacity and radix, while retaining sub-μs latency performance. In this realm, we have presented the principle of operation of OPS networks along with the requirement of a BM-CDR at the receiver input with ns-scale locking times that will account for the phase variation between different generating nodes. We evaluated, for the first time, a 25.6 Tb/s OPS layout, based on the Hipoλaos architecture in conjunction with a BM-CDR and demonstrated end-to-end 25 Gb/s true optical packet switching featuring burst-mode reception with $<$50 ns locking time. Error-free performance at 10-9 was obtained for all validated port-combinations. The multicast functionality of the Hipoλaos switch was also evaluated through the BM-CDR, concluding to successful multicasting to 2 output ports of the switch. Finally, a simulation performance analysis highlighted that low-latency performance, as required by disaggregated DCs, can be successfully realized in a Hipoλaos-switched DC with up-to 100% throughput for a variety of traffic profiles. Our evaluation showed that OPS solutions, like Hipoλaos can be a cornerstone towards future disaggregated DC systems, offering the required bandwidth, latency and radix credentials.

Acknowledgment

The authors would like to acknowledge Keysight for supporting the experiments with measurement equipment.

References

[1] "Global cloud index projects cloud traffic to represent 95 percent of total data center traffic by 2021," Newsroom.cisco.com, 2020. [Online]. Available: https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1908858
[2] A. Feldmann *et al.*, "The lockdown effect," in *Proc. Assoc. Comput. Machinery Internet Meas. Conf.*, 2020, pp. 1–12.
[3] T. Böttger, G. Ibrahim, and B. Vallis, "How the internet reacted to Covid-19," in *Proc. Assoc. Comput. Machinery Internet Meas. Conf.*, 2020, pp. 34–41.
[4] K. Katrinis *et al.*, "Rack-scale disaggregated cloud data centers: The dReDBox project vision," in *Proc. Des., Automat. Test Europe Conf. Exhib.*, 2016, pp. 690–695.
[5] G. Zervas, H. Yuan, A. Saljoghei, Q. Chen, and V. Mishra, "Optically disaggregated data centers with minimal remote memory latency: Technologies, architectures, and resource allocation," *J. Opt. Commun. Netw.*, vol. 10, pp. A270–A285, 2018.
[6] P. X. Gao *et al.*, "Network requirements for resource disaggregation," in *Proc. 12th USENIX Conf. Operating Syst. Des. Implementation*, Berkeley, CA, 2016, pp. 249–264.

[7] Broadcom Ships Tomahawk 4, Industry's Highest Bandwidth Ethernet Switch Chip at 25.6 Terabits per Second. [Online]. Available: https://www.broadcom.com/company/news/product-releases/52756

[8] The world's highest performance programmable switch –8 Tbps through 25.6 Tbps featuring industry-leading analytics, largest on-chip buffer and lowest-latency. [Online]. Available: https://www.innovium.com/wp- content/uploads/Innovium_TL8_Product_Brief_v1.0.pdf

[9] "112G XSR serdes PHY - Rambus," Rambus, 2020. [Online]. Available: https://www.rambus.com/interface-ip/serdes/112g-xsr-phy/

[10] "Rockley photonics demos co-packaged optics OptoASIC switch system with multiple partners," Lightwave, 2020. [Online]. Available: https://www.lightwaveonline.com/optical-tech/components/article/14170143/rockley-photonics-demos-copackaged-optics-optoasic-switch-system-with-multiple-partners

[11] "Flattening networks – and budgets – with 400G ethernet," *The Next Platform*, 2020. [Online]. Available: https://www.nextplatform.com/2018/01/20/flattening-networks-budgets-400g-ethernet/

[12] Calient.net, "S series optical circuit switch/CALIENT technologies," 2018 [Online]. Available: https://www.calient.net/products/s-series-photonic-switch/

[13] Polatis, "SERIES 7000 - 384 $\times$ 384 port software-defined optical circuit switch," 2016 [Online]. Available: https://www.polatis.com/series-7000-384x384-port-software-controlledoptical-circuit-switch-sdn-enabled.asp

[14] R. Hemenway, R. Grzybowski, C. Minkenberg, and R. Luijten, "Optical-packet-switched interconnect for supercomputer applications," *J. Opt. Netw.*, vol. 3, pp. 900–913, 2004.

[15] K. I. Sato, H. Hasegawa, T. Niwa, and T. Watanabe, "A largescale wavelength routing optical switch for data center networks," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 46–52, Sep. 2013.

[16] X. Kang, Y.-H. Kao, and H. J. Chao, "A petabit bufferless optical switch for data center networks," in *Optical Interconnects for Future Data Center Networks*. New York, NY, USA: Springer, 2013, pp. 135–154.

[17] R. Proietti *et al.*, "Scalable optical interconnect architecture using AWGR based TONAK LION switch with limited number of wavelengths," *J. Lightw. Technol.*, vol. 31, no. 24, pp. 4087–4097, Dec. 2013.

[18] W. Miao, S. Di Lucente, J. Luo, H. Dorren, and N. Calabretta, "Low latency and efficient optical flow control for intra data center networks," *Opt. Exp.*, vol. 22, pp. 427–434, 2014.

[19] J. L. Benjamin, T. Gerard, D. Lavery, P. Bayvel, and G. Zervas, "PULSE: Optical circuit switched data center architecture operating at nanosecond timescales," *J. Lightw. Technol.*, vol. 38, no. 18, pp. 4906–4921, Sep. 2020.

[20] 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, no. 7, pp. 1–14, Jul. 2018.

[21] Y. Mori, M. Ganbold, and K. Sato, "Design and evaluation of optical circuit switches for intra-datacenter networking," in *J. Lightw. Technol.*, vol. 37, no. 2, pp. 330–337, Jan. 2019.

[22] E. Honda, Y. Mori, H. Hasegawa, and K. Sato, "Feasibility test of large-scale (1,424$\times$1,424) optical circuit switches utilizing commercially available tunable lasers," in *Proc. 24th Optoelectron. Commun. Conf. 2019 Int. Conf. Photon. Switching Comput.*, Fukuoka, Japan, 2019, pp. 1–3.

[23] M. Moralis-Pegios *et al.*, "Multicast-enabling optical switch design employing Si buffering and routing elements," in *IEEE Photon. Technol. Lett.*, vol. 30, no. 8, pp. 712–715, Apr.2018.

[24] N. Terzenidis, M. Moralis-Pegios, G. Mourgias-Alexandris, K. Vyrsokinos, and N. Pleros, "High-port low-latency optical switch architecture with optical feed-forward buffering for 256-node disaggregated data centers," *Opt. Exp.*, vol. 26, pp. 8756–8766, 2018.

[25] M. Moralis-Pegios, N. Terzenidis, G. Mourgias-Alexandris, K. Vyrsokinos, and N. Pleros, "A 1024-port optical uni- and multicast packet switch fabric," *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1415–1423, Feb. 2019.

[26] N. Terzenidis, M. Moralis-Pegios, G. Mourgias-Alexandris, T. Alexoudi, K. Vyrsokinos, and N. Pleros, "High-port and low-latency optical switches for disaggregated data centers: The Hipoλaos switch architecture," *J. Opt. Commun. Netw.*, vol. 10, no. 7, pp. B102–B116, 2018.

[27] A. Tsakyridis, N. Terzenidis, G. Giamougiannis, M. Moralis-Pegios, K. Vyrsokinos, and N. Pleros, "25.6 Tbps capacity and SUB-μSEC latency switching for datacenters using $>$1000-port optical packet switch architectures," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 2, pp. 1–11, Mar./Apr. 2021.

[28] G. Giamougiannis *et al.*, "Demonstration of low-latency ETH-switched datacenter and 5G fronthaul networks using the 1024-port hipoλaos optical packet switch," in *Proc. Eur. Conf. Opt. Commun.*, Brussel, Belgium, 2020, pp. 1–4.

[29] M. Verbeke *et al.*, "A 25 Gb/s all-digital clock and data recovery circuit for burst-mode applications in PONs," *J. Lightw. Technol.*, vol. 36, no. 8, pp. 1503–1509, Apr. 2018.

[30] I. Ozkaya *et al.*, "A 56 Gb/s burst-mode NRZ optical receiver with 6.8ns power-on and CDR-Lock time for adaptive optical links in 14 nm FinFET CMOS," in *Proc. IEEE Int. Solid - State Circuits Conf.*, San Francisco, CA, USA, 2018, pp. 266–268.

[31] K. Shi *et al.*, "System demonstration of nanosecond wavelength switching with burst-mode pam4 transceiver," *Proc. 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.

[32] K. Clark *et al.*, "Sub-nanosecond clock and data recovery in an optically-switched data centre network," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.

[33] A. Forencich *et al.*, "A dynamically-reconfigurable burst-mode link using a nanosecond photonic switch," *J. Lightw. Technol.*, vol. 38, no. 6, pp. 1330–1340, Mar. 2020.

[34] P. Bakopoulos *et al.*, "NEPHELE: An end-to-end scalable and dynamically reconfigurable optical architecture for application-aware SDN cloud data centers," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 178–188, Feb. 2018.

[35] A. Tsakyridis *et al.*, "End-to-end optical packet switching with burst-mode reception at 25 Gb/s through a 1024-port 25.6 Tb/s capacity Hipoλaos optical packet switch," in *Proc. Eur. Conf. Opt. Commun.*, Brussel, Belgium, 2020, pp. 1–4.

[36] N. Terzenidis, M. Moralis-Pegios, G. Mourgias-Alexandris, K. Vyrsokinos, and N. Pleros, "Multicasting in a high-port Sub-μsec latency Hipoλaos optical packet switch," *IEEE Photon. Technol. Lett.*, vol. 30, no. 17, pp. 1535–1538, Sep. 2018.

[37] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," in *Proc. Sigcomm Workshop: Res. Enterprise Netw.*, 2009, pp. 92–99.

[38] A. Cevrero *et al.*, "4×40 Gb/s 2 pJ/bit optical RX with 8ns Power-on and CDR-Lock time in 14nm CMOS," in *Proc. Opt. Fiber Commun. Conf. Expo.*, San Diego, CA, USA, 2018, pp. 1–3.
[39] J. Im *et al.*, "A 0.5-28GB/S wireline tranceiver with 15-Tap DFE and fast-locking digital CDR in 7 nm FinFET," in *Proc. IEEE Symp. VLSI Circuits*, Honolulu, HI, USA, 2018, pp. 145–146.
[40] W. Mellette *et al.*, "RotorNet," in *Proc. Conf. ACM Special Int. Group Data Commun.*, 2017, pp. 267–280.
[41] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, "An all-optical neuron with sigmoid activation function," *Opt. Exp.*, vol. 27, pp. 9620–9630, 2019.
[42] M. Spyropoulou *et al.*, "40 Gb/s NRZ wavelength conversion using a differentially-biased SOA-MZI: Theory and experiment," *J. Lightw. Technol.*, vol. 29, no. 10, pp. 1489–1499, May 2011.
[43] M. Zirngibl, C. Dragone, and C. Joyner, "Demonstration of a 15 $*$ 15 arrayed waveguide multiplexer on InP," *IEEE Photon. Technol. Lett.*, vol. 4, no. 11, pp. 1250–1253, Nov. 1992.
[44] X. Zheng, N. Calabretta, O. Raz, D. Zhao, R. Lu, and Y. Liu, "40 Gb/s all-optical unicast and multicast wavelength converter array on an InP monolithically integrated chip fabricated by MPW technology," *Opt. Exp.*, vol. 25, no. 7, 2017, Art. no. 7616.
[45] S. Stopinski, M. Malinowski, R. Piramidowicz, E. Kleijn, M. K. Smit, and X. J. M. Leijtens, "Integrated optical delay lines for time-division multiplexers," *IEEE Photon. J.*, vol. 5, no. 5, Oct. 2013, Art. no. 7902109.
[46] N. Calabretta, W. Miao, K. Mekonnen, and K. Prifti, "SOA based photonic integrated WDM cross-connects for optical metro-access networks," *Appl. Sci.*, vol. 7, no. 9, pp. 865, 2017.
[47] A. Tsakyridis *et al.*, "Theoretical and experimental analysis of Burst-mode wavelength conversion via a differentially -biased SOA-MZI," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4607–4617, Sep. 2020.
[48] A. Tsakyridis, T. Alexoudi, A. Miliou, N. Pleros, and C. Vagionas, "10 Gb/s optical random access memory (RAM) cell," *Opt. Lett.*, vol. 44, pp. 1821–1824, 2019.
[49] A. Varga, "The OMNeT++ discrete event simulation system," in *Proc. Eur. Simul. Multiconference*, Prague, Czech Republic, 2001, pp. 1–7.
[50] W. J. Dally, and B. Towles, *Principles and Practices of Interconnection Networks*. San Fransisco, CA, USA: Morgan Kaufmann, 2004.
[51] K. Clark *et al.*, "Sub-nanosecond clock and data recovery in an optically-switched data centre network," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.