Broadband Multicasting for Wavelength-Striped Optical Packets

Caroline P. Lai, *Member, IEEE, Member, OSA*, and Keren Bergman, *Fellow, IEEE, Fellow, OSA*

***Abstract*—Wavelength-striped optical packet multicasting comprises a potentially important functionality for future energy-efficient network applications. We report on two multicast-capable architectures to experimentally demonstrate multiwavelength packet multicasting in an optical switching fabric testbed. The first design uses programmable packet-splitter-and-delivery that simultaneously supports the nonblocking unicast, multicast, and broadcast of high-bandwidth optical packets with parallel switches. This realization achieves the error-free multicasting of optical messages with 8×10 Gb/s payloads, with confirmed bit-error rates less than 10^{-12}, and scalability of per-channel data rates to 40 Gb/s. We then introduce a second multistage multicasting architecture with lower hardware and energy costs, with the design trade-off of more complex routing logic; the experimental demonstration shows the successful switching and error-free multicasting of 8×10 Gb/s optical packets. The energy costs in terms of the capital and operational expenditures are then compared for the two designs, showing the benefits of the second multicast architecture.**

***Index Terms*—Future internet, multicast networks, optical communication, photonic switching systems, routing, wavelength-division multiplexing (WDM).**

I. Introduction

NEXT-GENERATION networks will need to support bandwidth-intensive applications with agile functionalities and low energy consumptions. Traffic trends show annual growths of approximately 60% [1], [2], while the power consumption of telecommunication networks is predicted to grow exponentially [3]–[6]. The current infrastructure cannot viably sustain these traffic and energy trends [7], thus driving the need for novel low-energy optical technologies and high-bandwidth networks [8].

Optical packet switching (OPS) offers a unique, scalable approach for improving the bandwidth, latency, and power consumption performance of next-generation data-centric networks [9]. OPS enables the low-latency transmission of wavelength-striped optical messages in routers' switching fabrics through wavelength-division multiplexing (WDM) [10]. The switching fabric can establish end-to-end transparent lightpaths between network terminals with packet-rate switching speeds and data-rate transparency. We envision a fabric architecture that can leverage optical devices as accessible components, while optimizing performance in a cross-layer way [11], [12]. The cross-layer architectures may provide flexible, quality-of-service-aware and energy-aware capabilities [11], [13], [14], as well as increased control over multiwavelength optical messages with a packet-level granularity.

Demonstrating a high level of network functionality on the optical layer is a challenge for practical optical switching fabrics. As an example application, *optical packet multicasting* can enable greater programmable flexibility for OPS networks [15].

Multicasting is an Internet Protocol operation that allows one source to simultaneously transmit data packets to multiple endpoints. The need for multicasting is driven by high-bandwidth, point-to-multipoint applications such as distributed computing, high-definition video streaming, teleconferencing, gaming, and storage area networking. Traditionally, multicasting is implemented electronically in routers' IP layer by replicating and storing data packets in buffers [16]. However, by migrating the multicast operation lower in the network stack to the optical layer, packet-based functionalities may be supported at lower cost [17].

All-optical multicasting solutions can realize a more intelligent, transparent network [18]. IP multicasting utilizes many optical-electrical-optical (O/E/O) conversions, especially at the intermediate nodes. Data must be converted to the electronic domain, copied, and then converted back to optical signals. In contrast, optical multicasting supports the simultaneous transmission of multiple packets, where the data remain entirely optical from source to destination with no intermediate O/E/Os [16], [19]. The optical approach eliminates redundant O/E/Os, potentially reducing router overload and latency, and thus avoiding expensive IP router retransmissions [20]. Optical multicasting can ultimately minimize the total electrical energy consumption associated with transmission. It does not require buffers, provides network transparency to bit rate and modulation formats, and, hence, may be more powerful than the IP-layer's store-and-forward approach [21], [22]. Fig. 1 shows how we envision optical multicasting to be integrated in a cross-layer stack.

Here, we focus explicitly on broadband *optical packet multicasting* in OPS fabrics. Previous multicast scheduling investigations assume an N-input, N-output port packet switch

Manuscript received May 12, 2011; revised November 20, 2011; accepted February 03, 2012. Date of publication February 16, 2012; date of current version April 06, 2012. This work was supported in part by the National Science Foundation (NSF) Engineering Research Center on Integrated Access Networks under Grant EEC-0812072, and by the NSF Future Internet Design program under Grant CNS-837995.

C. P. Lai was with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA. She is now with the Photonic Systems Group, Tyndall National Institute, University College Cork, Cork, Ireland (e-mail: caroline.lai@tyndall.ie).

K. Bergman is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: bergman@ee.columbia.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2012.2188276

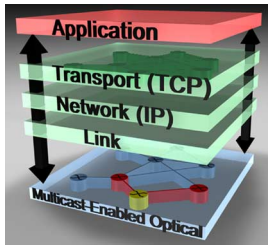

Fig. 1. Illustration of the future cross-layer optimized network stack, with bidirectional signaling between the network layers. The optical physical layer can provide an integrated optical packet multicast operation, where one network node (yellow) can simultaneously transmit to multiple nodes (red).

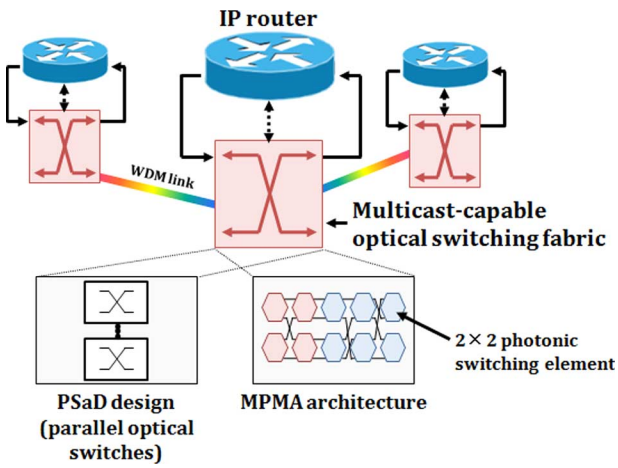

Fig. 2. Diagram of envisioned network nodes, depicting the integration of the multicast fabric designs with IP routers. Two multicasting architectures are discussed: PSaD, comprised of multiple internal optical switches; and MPMA. Both fabric designs use building blocks consisting of 2×2 PSEs.

that is capable of multicast and broadcast [23], [24]; stochastic and deterministic performance analyses show scheduling to be NP-hard. Other OPS designs either focus on wavelength multicasting [25], [26] with wavelength converters, or use impractical optical buffers/fiber delay lines [27], [28]. To the authors' knowledge, there has been little previous work on demonstrating optical packet multicasting in an experimental testbed.

In this paper, we leverage our fabric's distributed electronic routing logic control to seamlessly multicast optical packets. We propose and demonstrate two packet multicast-capable fabric designs on an experimental OPS testbed. Wavelength-striped packets are multicasted without wavelength converters or buffers. Multiwavelength packets can easily support high bandwidths with per-channel rates of 40 Gb/s or higher, as required by future applications [29]. Fig. 2 shows how multicast fabrics can be incorporated in future network structures.

The first design is based on splitter-and-delivery (SaD), proposed in [30] for wavelength multicasting. We modify the original SaD design to enable wavelength-striped packet-splitter-and-delivery (PSaD), providing a higher degree of connectivity and enabling packet multicasting (PaM) [31]. Experimentally, we implement two parallel OPS switches, each supporting a nonblocking unicast of 8×10 Gb/s packets, to show error-free, two-way multicasting with bit-error rates (BERs) less than 10^{-12}.

The second proposed design is the multistage packet multicasting architecture (MPMA), which is an improved topology with lower hardware costs [32] to realize a higher radix multicast. MPMA is a multistage design that uniquely capitalizes on the reprogrammable 2×2 photonic switching elements (PSEs) in the fabric testbed. Error-free multicasting of 8×10 Gb/s optical packets is achieved.

In the two distinct experiments, the switching fabric architectures seamlessly support the unicast, multicast, and broadcast operations. The two approaches showcase the design trade-offs that exist between multicast routing complexity, hardware cost, and possible energy metrics in terms of both capital (CAPEX) and operational expenditures (OPEX).

The remainder of this paper is organized as follows. Section II provides an overview of the basic fabric architecture and testbed realization. Section III discusses the PSaD design, implementation, and experimental results. The MPMA architecture is described in Section IV, outlining the design and experimental results. The two designs are compared in Section V. Section VI presents the conclusion.

II. Optical Switching Fabric Architecture and Implementation

The multiwavelength packet multicast investigations are performed on a switching fabric implementation that is based on a previously demonstrated OPS architecture [33], [34]. The fabric aims to optically interconnect access network edge users or computing network ports. The fundamental architecture is based on a transparent multistage network topology; the building blocks are 2×2 broadband, nonblocking, bufferless PSEs [see Fig. 3(a)] [35]. Each of the PSEs switches optical messages using four semiconductor optical amplifier (SOA) gates. The SOAs support a wide frequency band for transmission, data-format transparency, and packet-rate granularity, as well as offer subnanosecond switching speeds.

Here, the architecture can support both synchronous and asynchronous transmission [36] without any additional hardware. This flexibility alleviates the need for complex synchronization tools or modules; also, since the switching fabric is positioned within a local network node, clock distribution is not an issue for asynchronous operation. Here, the experimental demonstration supports synchronous operation. At the start of the timeslot, each terminal can begin transmission without prior requests from a centralized controller. Messages are injected via the fabric's input terminals and transparently routed by each PSE.

The switching fabric's wavelength-striping approach uses WDM to achieve high aggregate bandwidths by allocating the message data to parallel wavelengths that simultaneously contain payload data. The packet structure encodes control header information (i.e., frame and address bits) on a subset of dedicated frequencies, modulated at a single bit per wavelength per timeslot. Since the control wavelengths remain high for the optical message duration, the PSE's switching state remains constant as the message propagates through the PSE. The payload data are fragmented and modulated a high data rate (e.g., 10 Gb/s, 40 Gb/s, or higher) on the rest of the supported wavelength band. The OPS design enables a fast header processing

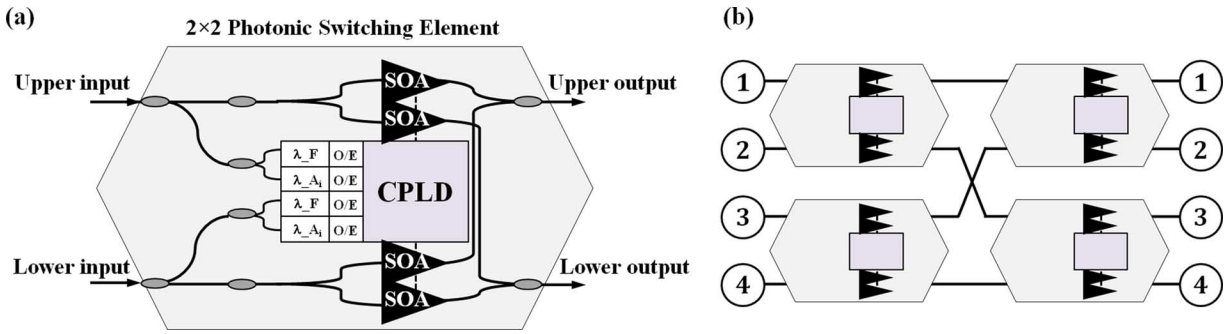

Fig. 3. (a) Schematic of the 2 × 2 PSE. (b) Example of how the PSEs may be arranged in a two-stage, 4 × 4 implementation.

that allows the message to capitalize on the abundant frequency spectrum provided by the wideband SOAs. Each 2×2 PSE uses a single header bit, which is a scalable way to achieve a switching fabric with many ports. Indeed, multiple stages of PSEs can be used, realizing greater than 10 000 number of switching fabric ports, while still maintaining sufficiently low BERs (i.e., BERs less than 10^{-9}) for an 8-channel payload [37]. The number of control wavelengths needed for addressing and switching is dictated by the number of stages; here, the SOAs' wide gain spectrum sufficiently accommodates all the data and control signals.

At each PSE, the leading edge of the optical packet is detected and received at one of the inputs. The electronic control logic decodes the control headers (the frame bit and one address bit) using fixed wavelength filters and low-speed optical receivers, and a routing decision is made immediately using high-speed electronic logic. The PSE's routing decision is based on the information encoded in the optical header. The message payload data are not decoded by the PSE and is transmitted concurrently with the control information using the SOA gates. The routing logic is distributed among the PSEs and experimentally implemented using reprogrammable logic devices, resulting in easy reconfigurability and the potential for supporting different routing protocols and algorithms. Each switching element uses simple, combinational logic, and no centralized fabric control and management plane is necessary. No additional signaling is required between PSEs, nor do the elements add or subtract information to/from the optical messages.

The message is, then, routed accordingly to its encoded destination. No optical buffering is realized in the PSEs; hence, packets are dropped in the case of message contention within the fabric. Though this topology is blocking, it is significantly advantageous since the individual switching elements can be simply realized at low cost, without the added complexity of buffers or wavelength converters. Since each PSE (i.e., routing stage) has identical propagation delays, the leading edges of messages injected in the same timeslot reach the PSEs in a given stage simultaneously. Successfully routed messages set up end-to-end lightpaths.

The SOAs are operated in their linear, small-gain regime, and are electrically driven with low currents ($\approx$50 mA). The SOAs' optical gain compensates for the insertion losses of the passive optical devices, so no net optical power gain/loss is incurred by the optical message. The SOAs provide a low-power switching gate over a wide frequency band such that thermal variations do not significantly affect performance. The SOAs used here are commercially available from Kamelian (Amphotonix) with rated noise figures of $\sim$6.5 dB; each SOA provides $\approx$8.5 dB of gain. The optical powers of the input packets are set such that the SOAs do not add nonlinearities to the packet. Due to the SOAs' linear mode of operation, no pattern-dependent effects were observed; all the BER measurements and experimental results exhibit no dependence on the pseudorandom bit sequence (PRBS) lengths utilized. In future implementations, other low-power switching devices may also be considered.

The architecture is experimentally implemented as a 4×4 fabric, using the 2×2 PSE building blocks. The PSEs use discrete macroscale commercially-available components, such as SOAs, passive optical devices and couplers, fixed wavelength filters, 155-Mb/s low-speed *p-i-n* photodetectors, and electronic circuitry. The high-speed electronic decision logic is synthesized in Xilinx complex programmable logic devices (CPLDs). Each PSE has four SOA gates in a broadcast-and-select topology and organized in a 2×2 matrix. The PSE decodes the optical control bits and maintains a routing state based on the extracted headers while simultaneously handling wavelength-striped data transparently in the optical domain. Each PSE decodes four control header bits (two for each input port); at each switching stage, the wavelength-based routing information is extracted. The CPLD uses the headers as inputs in a programmed routing truth table, and gates the appropriate SOAs. The extracted frame denotes the presence of a wavelength-striped packet; then, according to the low (or high) detection of the address signal, the CPLD gates the suitable SOA for the packet to be routed to the upper (or lower) output port of the 2×2 PSE [see Fig. 3(a)]. Fig. 3(b) provides a typical example of how two stages of PSEs may be connected to realize a 4×4 switching fabric.

The basic switching fabric design leverages a multistage banyan network topology that requires only $\log_2(N)$ stages to create a $N \times N$ interconnect [38]. Each stage of the banyan switch consists of $N/2$ PSEs. Banyan networks require fewer components as compared to a full-interconnected topologies, which may scale as N^2. Thus, a high degree of port scalability can be achieved using a low number of stages. Extending this notion to the following proposed multicasting designs: by leveraging the basic banyan design within either the PSaD or MPMA designs, a relatively low-cost PaM functionality can be achieved (compared to other potential topologies).

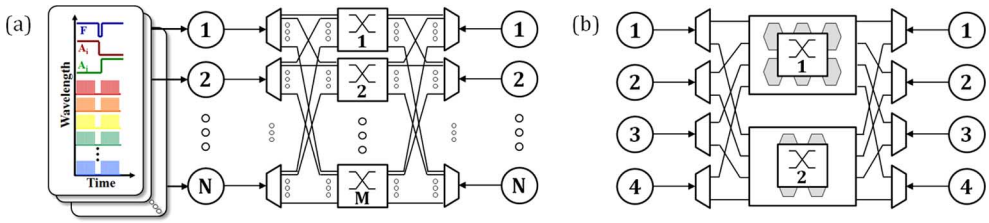

Fig. 4. (a) Proposed PSaD architecture that supports wavelength-striped optical messages ingressing on N input ports, using M optical packet switches. The design uses optical packet switches that operate in parallel. (b) Block diagram of PSaD design demonstrated in the experiment, supporting $N = 4$ input ports with $M = 2$ internal switches. The optical packets pass through optical splitters and combiners at the input and output of the fabric, respectively.

III. PSaD Architecture

A. Overview

The first multicast-capable architecture discussed here uses the PSaD design that can support the simultaneous transmission of multiple broadband, wavelength-striped messages to multiple outputs [31]. The basic initial SaD architecture presented in [30] is nonblocking and demultiplexes the incoming lightpath into M wavelengths to deliver the channels to M separate destination endpoints, realizing wavelength multicasting. Here, we modify the original SaD design to create a PSaD system with a higher level of connectivity, where the input wavelength-striped packet can be spatially split and multicasted to multiple outputs. This is then similar to the splitter and combiner structure discussed in [39]; however, since this system supports multiwavelength PaM, it does not need wavelength converters.

The switching fabric is internally composed of M parallel optical packet switches interconnecting N network terminals [see Fig. 4(a)]. Each source input is connected to each destination output using M separate switch entities that operate in parallel. The PSaD architecture creates M distinct and independent paths between each source and destination, in a nonblocking fashion. Each path (optical switch) supports the multiwavelength optical packet format. One clear advantage is that the optical switching fabric can either handle a unicast using a single switch, or multicast using combinations of several switches.

For the purposes of the experimental fabric demonstration, a two-way multicasting is realized using a complete 4×4 optical switching fabric that is comprised of two parallel OPS switches (i.e., $N = 4$ and $M = 2$) [see Fig. 4(b)]. The two internal switches provide two independent paths to multicast to two distinct destinations. The incoming multiwavelength packet is injected into the switching fabric via propagation on a single fiber. The optical packet is split using a 1:2 passive coupler to create two replicas of the multiwavelength packet, each of which propagates to the first respective routing stages of each switch. At the output, another 2:1 passive optical coupler is used to combine the packets egressing from both switches. In future realizations with $M > 2$ (i.e., more than two parallel optical switches), we can envision using a $1 \times M$ SOA-based switch to provide the gain to compensate for the insertion loss of splitting the signal M ways (in place of the optical couplers). A similar $M \times 1$ switch can be used as a combiner at the output. Here, the upper parallel switch is based on a three-stage banyan architecture, organized in an Omega network topology, with a distribution stage [33], using a total of six PSEs. The lower parallel switch is based on a two-stage banyan topology, using four 2×2 PSEs. Thus, using the three-stage switch in parallel with the two-stage switch, the complete PSaD design is experimentally implemented using ten 2×2 PSEs with five routing stages, with two PSEs in each stage. Due to the differing round-trip times of the two internal switches, packets may have different transit times through the switching fabric. In order to support the cascade of several PSaD fabrics within a real-world, multinode network, the switching fabric will have to be operated in asynchronous mode [36], which does not require any additional hardware. No additional synchronization switching modules would be required.

Each of the distributed switches uniquely supports a unicast; thus, by operating several of these switches in parallel, the complete architecture supports the desired PaM operation. This implementation interconnects $N = 4$ input and output ports using ten PSEs, which is not the absolute minimum; a similar connectivity for a $M = 2$ fanout multicast could be achieved using eight PSEs (using two two-stage switches in parallel). In this realization, we emphasize the fabric's flexibility by demonstrating that different switch topologies can be deployed in the M parallel switches with little effect on the overall multicasting functionality.

Here, as the leading edges of the two optical packets reach the first stages' PSEs, the messages are switched according to the optical headers encoded in the packet. For the five total routing stages, the optical packet leverages five distinct address bits. Within each of the two switches in the complete switching fabric, the PaM operation is realized since each switch supports the high-bandwidth wavelength-striped optical packets, which are switched entirely in the optical domain. The wideband nature of the SOA allows for a straightforward spatial multicasting of packets composed of multiple wavelengths. The multicasted multiwavelength optical packets are delivered to their desired (multiple) destinations at the output of the complete fabric, as per the required multicasting request encoded in the headers. The injected optical packet can be unicasted on a single switch (either upper or lower), or multicasted by traversing both entities simultaneously to reach multiple, distinct output destinations.

The PaM leverages the unique programmability of the individual PSEs. By letting the PSEs act as identical building blocks that can be arranged in different topologies, the multicasting operation can be straightforwardly realized by placing multiple switch entities in parallel. Furthermore, taking into account the distributed nature of the PSEs' routing logic and the fact that there is no signaling required between PSEs, or between the

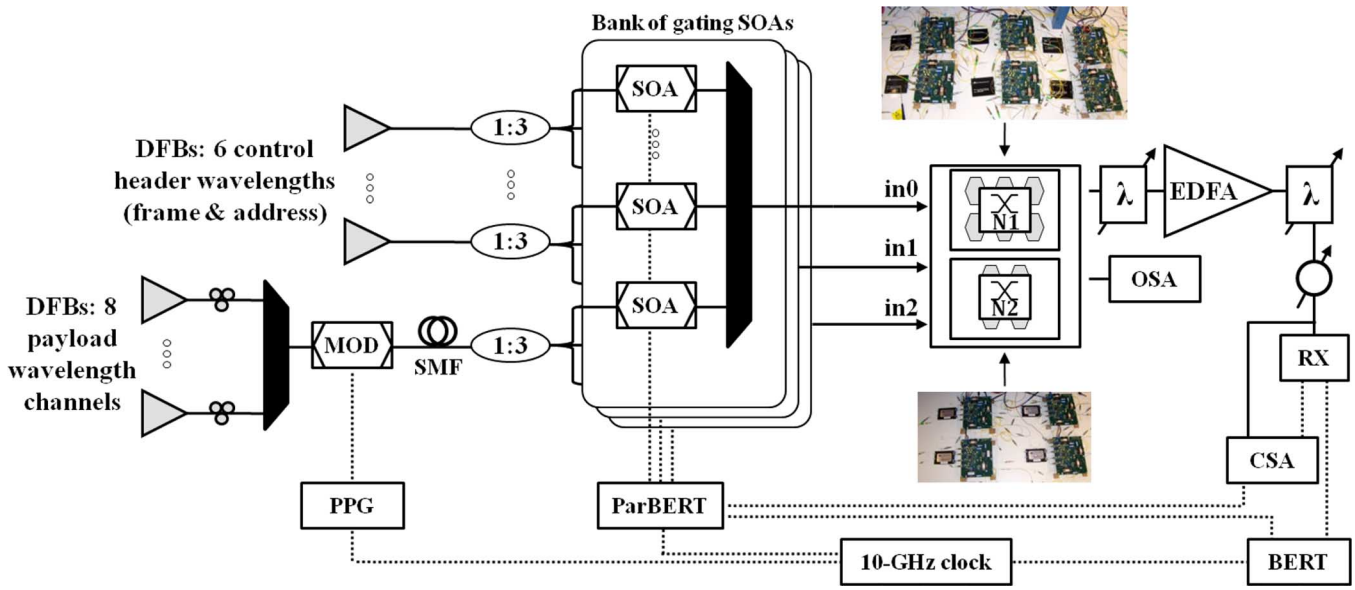

Fig. 5. Experimental setup of PSaD demonstration, with photographs of the two switches internal to the optical switching fabric.

PSEs and a centralized control plane, implementing several of the OPS switches in parallel is a scalable way to enable PaM. The number of realizable switches does not grow with a required controller unit or logic. The simple architecture shows multicasting with limited added routing complexity. Here, buffers and wavelength converters are not required to support the PaM. One may note that for a M-way multicasting, M parallel optical switches are required, which is potentially costly to implement; MPMA presents a possible solution to this issue.

B. Experimental Demonstration and Results

The experimental demonstration of the multicast-capable PSaD design shows the correct routing of multiwavelength optical packets incorporating 8×10 Gb/s wavelength-striped payloads. The packets are multicasted error-free to multiple destination ports using an implemented optical switching fabric with BERs confirmed less than 10^{-12}. A pattern of wavelength-striped optical packets is injected in the fabric, and thus simultaneously into both optical switches (Fig. 5). The testbed connects four independent inputs to four distinct outputs using the two parallel switches, offering a path diversity of two. The payload data for the wavelength-striped packets are generated using eight continuous-wave (CW) distributed feedback (DFB) lasers ranging from 1533.18 to 1564.39 nm, which are combined onto a single fiber using an 8:1 optical combiner. The eight wavelength channels are then simultaneously modulated with a 10 Gb/s nonreturn-to-zero (NRZ) signal that carries a $2^7 - 1$ PRBS using an $LiNbO_3$ modulator. The modulator is driven by a high-speed electrical signal from a pulse pattern generator (PPG). The wavelength channels are decorrelated by 25 km of single-mode optical fiber. The payload wavelengths are then split using a passive 1:3 optical coupler to create three modulated wavelength-striped data flows for injection in three fabric ports. Each set of payload wavelength signals are then transmitted to external gating SOAs.

This system here creates 8×10 Gb/s wavelength-striped packets with a six-wavelength control header and an eight-wavelength payload. The control wavelengths are generated using separate CW-DFB lasers, including one frame at 1555.75 nm and five address bits, ranging from 1531.12 to 1550.92 nm. The DFB lasers are split using 1:3 couplers and sent to a set of gating SOAs. The control header and payload data signals are then gated into packets using an array of gating SOAs, encoding the appropriate addressing information for each packet to be routed through the testbed. The control headers and the payload signals are then passively combined together to create a multiwavelength data stream. A similar packet-generation setup is used concurrently for each set of control and payload to form three distinct packet patterns for the three fabric inputs. An Agilent ParBERT is used as a nanosecond-scale electronic signal generator that controls the gating SOAs for packet gating and fabric addressing. The ParBERT is preprogrammed with test packet patterns that are custom designed for each experimental demonstration. There is no synchronization between the start of the packets and the start of the PRBS patterns. These packets are then injected into the active ports of the switching fabric, simultaneously in both parallel OPS switches. The switch selection (i.e., whether packets are transmitted on the upper or lower optical switch within PSaD) is based on an *a priori* knowledge of the address wavelength space and the destination availability during the custom test pattern. The experiment supports timeslots that are 128 ns long, using optical packets with 115.2 ns durations. The header wavelengths for the switching/multicasting of optical packets are predetermined for each experiment and set by the ParBERT.

At the fabric's output, the multiwavelength packet is monitored using an optical spectrum analyzer and high-speed sampling oscilloscope. A packet analysis system allows the wavelength-striped packet to propagate to a tunable grating filter (λ in Fig. 5). The filter selects one payload channel for signal in-

tegrity analysis and rejects the accumulated amplified spontaneous emission (ASE) from the SOAs. The payload channel is then sent to an erbium-doped fiber amplifier (EDFA), another tunable filter to reduce the ASE from the EDFA, and a variable optical attenuator. The payload wavelength channel is then received by a DC-coupled 10 Gb/s *p-i-n* photodiode with a transimpedance and limiting amplifier pair (RX). The received electrical signals are sent to a BER tester (BERT) that is synchronized with the PPG and gated to analyze the packets with the ParBERT.

In this way, the wavelength-striped optical packets are generated, injected in the fabric, and switched through both parallel OPS switches. The messages are multicasted to two different destinations (if desired) by unicasting on each switch. The waveform traces associated with the optical packet sequence in this experiment are given in Fig. 6. The resulting packets egressing from the switching fabric in the testbed are also given. The waveforms in Fig. 6 provide the frame bit of the packet (set to high for the duration of the packet), as well as one of the payload wavelength channels, for each of the ingressing and egressing fabric ports. The two-bit binary addresses are indicated for each optical message in Fig. 6. Since one address bit is required for each routing stage in the optical switch entity, packets routed through the three-stage N1 switch require three bits, while messages routed using the two-stage N2 switch require two address bits.

All of the 8×10 Gb/s multiwavelength optical messages are correctly routed through the complete switching fabric, and accurately emerge at the destinations that are encoded in the control address headers. The multicasting operation is clearly validated, as the wavelength-striped packets are successfully routed from one fabric input port to multiple output ports. The packet sequence shows that the switching fabric seamlessly supports both the unicast operation using a single switch entity, in addition to the multicasting operation with both switches. In the first active timeslot depicted in Fig. 6 (denoted as A), the sources at two independent input ports transmit only on the upper fabric switch (N1, using in0 and in2). The optical packet from in0 has an encoded address of 001, which represents its output port (out1); the packet clearly emerges from this port in timeslot B. Simultaneously, in timeslot A, a packet appears at the in2 port, with address 010 (addressed for output port out2), and the packet emerges from out2. Similarly, during the second active timeslot, all three sources transmit packets to three distinct output ports, each unicasting on the second/lower optical switch (N2). Packets have a two-bit address header, indicating their desired destinations: the packet from in0 (address 11) wishes to be routed to out3, the packet from in1 (address 00) has out0 as its required output, and the packet from in2 (address 00) has out0 as its desired destination. One of the contending packets from in1 and in2 is dropped and is retransmitted at a later timeslot. Note that due to the fact that the three-stage N1 has an additional routing stage as compared to the two-stage N2 (and thus larger transmission latency), packets that are routed through N2 appear one timeslot earlier than those routed through N1. Correspondingly, in the third timeslot C, a single source at the input of the switching fabric (in0) attempts to perform a packet multicast to two distinct destinations. This

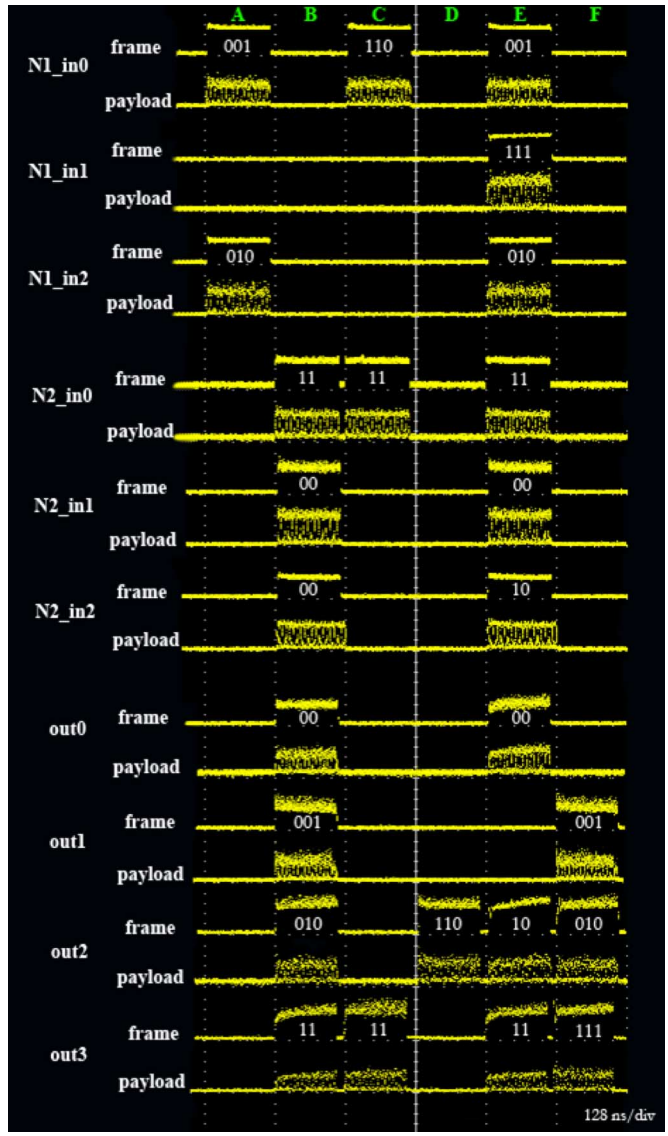

Fig. 6. PSaD experimental waveform traces of the input and output optical packets, with labels referring to the address information encoded in the optical packets.

is done via two simultaneous unicast operations using both of the fabric's switches: packets are routed to out2 using N1 and to out3 using N2. These packets emerge from the output of the fabric in timeslots D and E, respectively. During the fourth active timeslot for packet injection (timeslot E), all available sources attempt to multicast wavelength-striped optical messages to multiple output destination ports by simultaneously transmitting using both switching fabric entities, the upper N1 and the lower N2. These packets egress from the fabric testbed after the appropriate delay. Packets routed through N2 appear at the output ports one timeslot earlier than those routed through N1, and all packets injected in timeslot E are correctly routed, appearing in timeslots E and F.

BER measurements confirm that all packets are received are error-free, achieving BERs less than 10^{-12} on all eight payload wavelengths. Fig. 7 shows the BER sensitivity curves corresponding to the back-to-back operation, as well as transmission through the lower N2 (two-stage switch), taken for the pay-

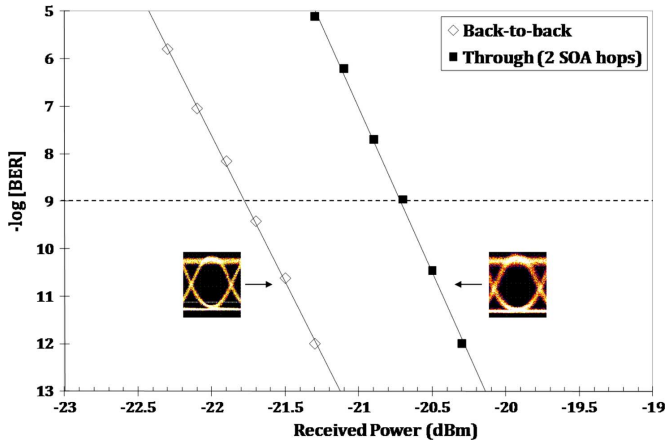

Fig. 7. BER sensitivity curves corresponding to the PSaD demonstration, with insets showing the 10 Gb/s back-to-back and through eye diagrams associated with one payload wavelength channel. The unfilled points correspond to packets before injection in the fabric, while the filled points correspond to packets that emerge from the fabric's output from the two-stage N2 optical switch.

load channel at $\lambda = 1541.05$ nm. The insets show the optical eye diagrams for the same 10 Gb/s channel. We see an approximate 1 dB power penalty for a two-SOA hop system (N2), corresponding to a 0.5 dB penalty performance for each SOA transversal.

Furthermore, we capitalize on the inherent bit-rate transparency of the switching fabric to demonstrate the scalability of the packet payload channels' data rates to higher modulation rates. The bit rates of the modulated data streams are scaled from the initial 10 Gb/s per payload to 40 Gb/s. The switching fabric supports an aggregate packet bandwidth of 250 Gb/s, composed of six 40 Gb/s and one 10 Gb/s multiplexed channels. The 10 Gb/s channel is used to demonstrate error-free performance, using the packet analysis system described previously.

The experimental setup for the combined 10 Gb/s and 40 Gb/s demonstration is similar to the 8×10 Gb/s setup outlined previously. A portion of the packets is generated using six CW-DFB lasers, ranging from 1533.18 to 1564.39 nm, whose outputs are multiplexed onto a single fiber. The six wavelength channels are then modulated using a 40 Gb/s $LiNbO_3$ modulator with a $2^7 - 1$ PRBS signal in an NRZ format. A single 10 Gb/s channel is also simultaneously generated using a separate CW-DFB and 10 Gb/s $LiNbO_3$ modulator. The control and payload signals are gated with external SOAs in a similar fashion as described previously. All modulated payload channels are then multiplexed together with the appropriate control header signals.

Correct routing is achieved with these high-bandwidth multiwavelength packets. Fig. 8 provides the optical eye diagrams of one of the filtered 40 Gb/s data streams (at $\lambda = 1558.24$ nm) for the back-to-back input packet that is injected into the testbed (as seen directly after the gating SOA) and for the output packet (as observed directly after the three-stage upper N1 optical switch). Error-free performance with BERs less than 10^{-12} is obtained on the 10 Gb/s stream. BER packet analysis for the 40 Gb/s payload channels is not feasible due to experimental limitations. However, by leveraging the multiplexed combination of the 10 Gb/s and 40 Gb/s payload channels, we show the successful transmission of 250 Gb/s aggregate bandwidth optical packets.

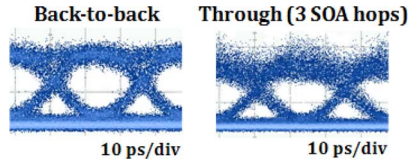

Fig. 8. 40 Gb/s eye diagrams verify the feasibility to scaling the data rates of the individual payload channels to 40 Gb/s. (Left) The 40 Gb/s back-to-back eye is taken directly after the gating SOA (before injection in the fabric). (Right) Th through 40 Gb/s eye is captured at the output of the fabric (after the three-stage N1 optical switch).

This packet multicast architecture uses several parallel optical switches that operate in a distributed fashion. One can then design the complete switching fabric with differing topologies deployed in the internal switches that have various features, including a completely nonblocking design or one with low latency. Then, a traffic classifier can be realized at the source to switch messages according to the requirements of the source applications, e.g., optical packets that must be routed with minimal latency can be routed through the switch with this attribute, or high-priority messages can be sent on a nonblocking switch to ensure transmission without contention.

IV. MPMA

A. Overview

The PSaD design uses PSEs with identical routing logic and multicasts packets by simply adding replicas of the unicast-capable OPS switch; this may be costly to implement for high multicast fanouts. Hence, we now introduce MPMA, an improved switching fabric topology whose basic design is itself optimized for optical packet multicasting, minimizing the added hardware. This architecture enables PaM using a single optimized topology, requiring fewer additional components, and thus could potentially offer a low-cost and energy-efficient solution. MPMA capitalizes on the distributed nature of the PSEs' electronic logic and particularly on their unique reprogrammability. By slightly increasing the complexity of the fabric's routing logic, we can achieve multicasting that may be more hardware-cost efficient than PSaD.

MPMA uses two distinct PSE routing truth tables. Both logic tables are based on simple combinational logic such that no centralized control is required, i.e., no signaling is required between the PSEs and a control unit to execute multicasting. Here, the fabric's PSEs do not contain identical routing control logic. This small increase in routing complexity allows a single switching topology that is designed to use more routing stages to realize PaM. Using the same experimental hardware as PSaD, we can effectively increase the achievable multicasting fanout to four. Similar to PSaD, the design also easily supports unicasting, multicasting, and broadcasting functionalities. However, one trade-off is the decreased path diversity to manage contention. Messages that contend within the switching fabric are dropped and thus must be retransmitted in a subsequent timeslot.

The MPMA design is comprised of a subset of fabric stages that is used for packet routing (PaR), followed by a subset used for PaM. The PSEs in the PaR and PaM stages contain differing

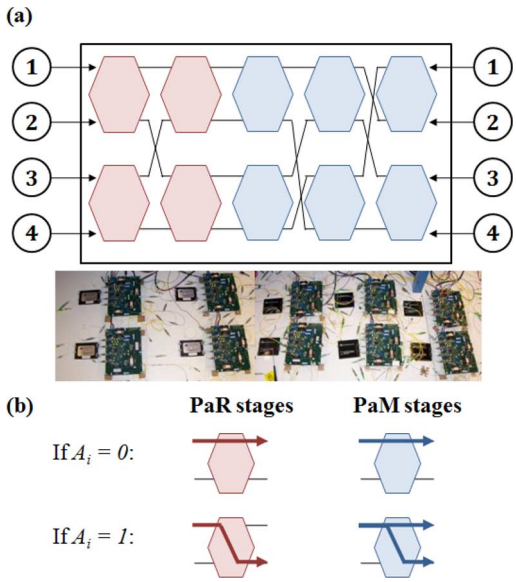

Fig. 9. (a) Block schematic of the proposed multicast-enabled fabric architecture, supporting four inputs/outputs with PSEs realizing PaR (red) stages and PaM (blue) stages. A photograph of the fabric is also provided. (b) Block diagrams showing the routing logic in the PaR and PaM stages; packets are routed depending on the low or high value of the stage's address bit.

control logic via the PSEs' CPLD (see Fig. 9). The logic determines whether the ingressing packet will be sent to one or both of the PSE's output ports. For the PaR stages, a multiwavelength packet that reaches one of the two input ports is routed directly to one of the output ports, i.e., depending on the address bits decoded by the electronic logic, the CPLD will gate one SOA (or none if contention occurs). For the PaM stages, depending on the recovered control bits, an incoming optical packet can be routed to either one or both of the PSE's available outputs. The CPLD will gate either one or two SOAs associated with the message. In both cases, the routing and multicasting operations depend solely on the optical packet headers that are extracted from the message using fixed wavelength filters and low-speed optical receivers. By cascading combinations of PaR and PaM stages, various multicasting topologies can be realized to enable varying multicast fanouts.

Here, we implement one distinct MPMA design to create a 4×4 optical fabric. The goal is to create an architecture that allows any input to transmit to any single port (unicast or one-way multicast), as well as to a subset of output ports (two-way or four-way multicast). The implemented 2×2 PSE hardware allows the PSEs in one routing stage to extract one address bit; also, the PSEs in each stage use the same wavelength addressing. These factors limit the possible MPMA designs that are feasible in this testbed. In order to simultaneously support the unicast and multicast operations, a 4×4 optical fabric using the realized PSEs requires a minimum of five stages, with two PaR stages and three PaM stages [see Fig. 9(a)]. Two different routing decision logic tables are distributed among the ten required PSEs. The optical message's headers indicate whether the packet is required to unicast, multicast, or broadcast using the single optimized fabric. If the PSE hardware supported extracting two distinct address bits at each stage (in addition to the frame), it would then be feasible to realize the 4×4 multicast fabric design using fewer (namely two) stages, in a broadcast-and-select topology, with a integrated version of PaR/PaM routing logic. This alternate topology would result in a more complex fabric addressing scheme, as well as a more complicated PSE hardware design; however, fewer stages would be required (translating to fewer SOA hops), thereby increasing the scalability of the multicast-capable topology.

B. Experimental Results

By simply reprogramming the control logic synthesized in the CPLDs of the ten 2×2 PSEs in the testbed, MPMA can be straightforwardly implemented without additional hardware as compared to PSaD. While a maximum fanout of two is supported in the PSaD implementation, a maximum multicasting fanout of four is achieved with MPMA with the equivalent hardware.

The MPMA experimental setup is generally similar to the first PSaD demonstration. The ten PSEs are arranged in the proposed multistage topology, with two PaR stages followed by three PaM stages. The routing control logic in the 2×2 PaR stages is as described previously, using a two-bit control input [see Fig. 9(b)]: the frame and address bits. When the address bit is low, the packet is routed to the upper port, and when the address bit is high, the packet is routed to the lower port. For the PaM stages [see Fig. 9(b)], the routing control logic was modified by reprogramming the CPLDs to realize the PaM logic. The programmable logic also uses a two-bit control input: when the address bit is low, the packet ingressing on the 2×2 PSE is transmitted across (upper input to upper output, and lower input to lower output), and when the address bit is high, the packet is transmitted to both of the output ports (upper/lower input to both outputs).

To show multicasting using MPMA, an experimental predetermined pattern of 8×10 Gb/s multiwavelength optical messages is generated and injected in the implemented 4×4 fabric. In addition to the eight 10 Gb/s payload channels, each message uses a six-wavelength control header, with one frame bit and five address bits (one address bit per stage). Depending on the high/low levels of the address bits encoded by the transmitter in the packet header and the type of logic encoded within the stage, the CPLD within the 2×2 PSE either routes or multicasts the wavelength-striped packet by gating on one or two SOAs. The payload information for the optical packets is generated similarly to the aforementioned demonstration, using eight CW-DFB lasers. The wavelength channels are concurrently modulated at 10 Gb/s with a single $LiNbO_3$ modulator with a $2^{11} - 1$ PRBS with an NRZ format. The eight-channel modulated payload is then transmitted to a discrete SOA that is gated using the ParBERT. The six control headers are generated separately and the ParBERT provides the corresponding addressing for the PaR and PaM stages. The control signals are then combined with the payload, creating the 179.2 ns long 8×10 Gb/s optical packets using 192 ns timeslots.

The packets are injected in the switching fabric, and are distributedly routed (PaR stages) or multicasted (PaM stages) according to the encoded optical headers. All the unique header combinations are shown to demonstrate a one-way, two-way,

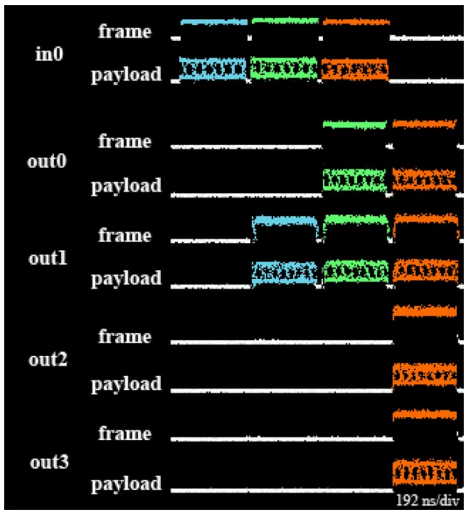

Fig. 10. Waveforms corresponding to the MPMA experimental optical packet sequence, which exemplify the multicasting operation executed by the realized topology.

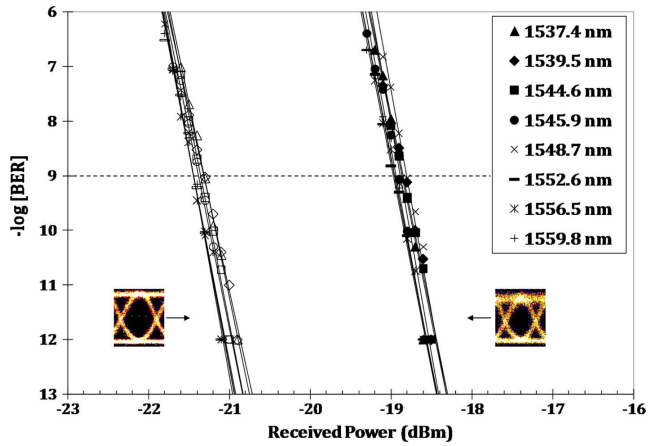

Fig. 11. BER sensitivity curves for MPMA, for all eight payload channels: open data points refer to measurements for packets at the input, and filled points correspond to measurements taken for packets at the output of the fabric. Insets show the input and output optical eye diagrams for the one 10 Gb/s payload channel of an optical packet ($\lambda = 1556.6$ nm).

or four-way multicast. The optical waveforms corresponding to ingressing and egressing optical packets are shown in Fig. 10, providing the traces for the frame bit and one modulated payload channel. In the first active timeslot, a packet (blue in Fig. 10) is injected and is transmitted to one output according to the encoded addressing. The first packet has address information 00000, sending the wavelength-striped message from in0 to out1 according to the distinct PaR and PaM logic tables. In the second timeslot, the packet (green in Fig. 10) has an encoded address of 00001, which routes the message simultaneously from in0, to out0 and out1. The multicasting is initiated by the fifth PaM stage (fifth address bit is high), indicating that the message should be transmitted to both output ports of the fifth stage. The green packet is successfully multicasted to the desired destinations. In the third timeslot, a packet (red in Fig. 10) with address 00011 is injected. The encoded control header shows that the packet wishes to be multicasted to all four available output ports. The fourth and fifth bits are high, such that the PSEs in the fourth and fifth PaM stages multicast the packet to both of their output ports. This allows an optical multicast (broadcast) of the packet from in0 to four outputs; thus, the red packet is transmitted to all four outputs simultaneously.

The waveforms show that all of the 8×10 Gb/s multiwavelength messages are correctly routed, egressing at the destinations designated by the control. We verify that the MPMA design seamlessly enables unicasting to one output, multicasting to two ports, and broadcasting to all four outputs.

The packet analysis system is identical to the prior setup, using a 10 Gb/s RX and BERT synchronized with the ParBERT. The BERT is gated over more than 80% of the packet (over 150 ns). BER measurements confirm the error-free transmission of the wavelength-striped packets at the output. BERs less than 10^{-12} are achieved on all eight payload wavelength channels. Sensitivity curves of the eight payload wavelengths of a packet emerging from the fabric are given in Fig. 11 and show a power penalty of 2.5 dB for the five-stage fabric (taken at a 10^{-9} BER), resulting in a 0.5 dB penalty for each SOA traversal. This result is similar to the PSaD experiment.

Thus, we show that MPMA can successfully realize error-free multicasting, with potentially improved scalability and reduced cost as compared to PSaD. It can be noted that the multistage switching design gives rise to greater packet contention. In the future, a control plane may be required to achieve nonblocking lightpaths for the optical messages. However, the management and scheduling of PaM from the higher layers is considered difficult (NP-hard) [24]. In this paper, we aim to reduce the control complexity on the optical layer and do not address the difficult multicast scheduling problem.

V. Analysis: Comparison of Multicast Designs

A. CAPEX

The PSaD and MPMA designs are first compared with respect to the hardware required to realize each in our testbed. This allows the CAPEX of the two architectures to be evaluated in terms of the number of deployed devices.

In this analysis, we assume the minimum number of components required by each architecture to enable a $M = 4$-way multicast fanout in a 4×4 (i.e., $N = 4$) OPS fabric. This assumes the use of the current 2×2 PSE realization, where each PSE extracts two control header bits.

For the case of PSaD: the aforementioned PSaD realization connects $N = 4$ ports with only a maximum of $M = 2$-way multicast; the first parallel optical switch is a three-stage switch and the second is a two-stage switch. In this case study, we instead assume a PSaD design connecting $N = 4$ ports to enable $M = 4$-way multicast, with each of the four parallel optical switches consisting of two-stage topologies [see Fig. 12(a)]. Hence, PSaD is assumed to use the minimum number of additional components. PSaD also requires several $1 \times M$ SOA-based switches at the fabric input and output, respectively, to compensate for the optical splitting and combining power losses.

For the case of MPMA: the aforementioned architecture is assumed, deploying five stages of 2×2 PSEs [see Fig. 12(b)].

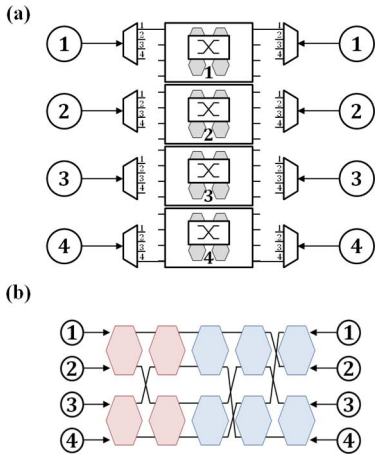

Fig. 12. Diagrams of the compared (a) PSaD and (b) MPMA topologies.

TABLE I
CAPEX ANALYSIS PARAMETERS

Parameter	PSaD	MPMA
Input/output ports (N)	4	4
Maximum multicast fan-out	4	4
Number of parallel switches (M)	4	N.A.
Number of 1:4 switches/couplers	8	N.A.
SOAs required for implementing switches	32	N.A.
Total number of routing stages in fabric	8	5
Total number of 2×2 PSEs	16	10
SOAs required for PSEs	64	40
Total number of SOAs	96	40

This five-stage implementation is needed to provide a $M = 4$-way multicast for a 4×4 fabric.

Table I depicts the results of the CAPEX analysis, depicting the hardware cost associated with this specific case of a 4×4 fabric with a four-way multicast. The main CAPEX cost metric consideration is the number of SOAs required under both PSaD and MPMA approaches. Since the $M = 4$-way PSaD design requires a greater number of additional PSEs (in addition to 1:4 switches/couplers), the number of SOAs needed to implement PSaD far outnumbers that of MPMA. This difference grows with port count and multicast fanout. Thus, we see that MPMA exhibits a reduced hardware cost to support the equivalent multicast fanout.

Future work involves the design of a single-stage PSE using a broadcast-and-select topology, which will require only 16 SOAs to achieve a four-way fanout multicast. This single 4×4 PSE can then comprise the entire 4×4 fabric with much simpler control logic.

B. OPEX

We further consider the issue of energy consumption from the perspective of the energy consumed during operation (i.e., the OPEX) for the two designs. For this analysis, a multilayer, IP-over-WDM perspective is adopted. Each network node is assumed to use a higher layer IP router, which generates the data traffic in the electronic domain. Each node is also composed of several interfacing WDM transponders that effect the

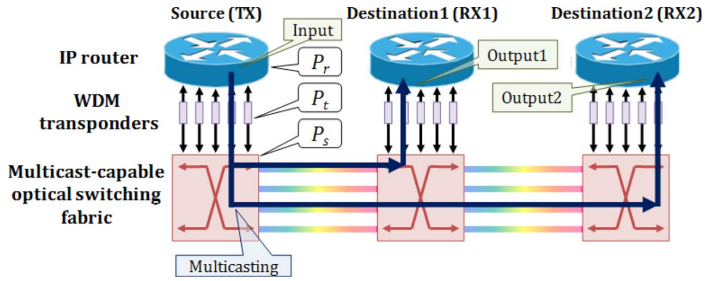

Fig. 13. Diagram depicting the composition of each network node, including an IP router, WDM transponders, and the multicast-capable optical switching fabric. A multicasting request is illustrated. The contributing OPEX power consumption factors are also shown in this multilayer, IP-over-WDM network consideration.

E/O and O/E conversions, transitioning the electrical data signals to/from multiwavelength optical packets. Under the all-optically switched assumption, the transponders are assumed to be used only at the transmitter (TX) and receiver (RX) side and not at intermediate nodes. The physical layer leverages one of the SOA-based multicast-capable optical switching fabrics presented previously, allowing the data to be transparently switched directly in the optical domain. Fig. 13 details the composition of the network node. The analysis takes into account the energy consumption of the router (P_r), the WDM transponders (P_t), and the switching fabric (P_s); Fig. 13 additionally depicts how these contributing energy consumption metrics play a role in this case study.

The unicast fabric is a basic two-stage, 4×4 topology, which uses four PSEs. The assumptions associated with the two multicast designs are identical to the aforementioned CAPEX analysis (i.e., a $M = 4$-way multicast fanout in a 4×4 OPS fabric), with the current 2×2 PSE implementation. During timeslots when the PSEs are not switching optical messages, the electronic control logic does not gate on the SOAs; thus, the PSEs consume negligible power.

For the higher layer energy parameters, we assume a Cisco CRS-1 router [40], which is the premier packet-switched core router available commercially today. The maximum configuration of the router yields a total capacity of 92 Tb/s [40], achieved using 72 linecard chassis, each supporting 1.28 Tb/s. The router is connected to edge nodes via 10 Gb/s packet-over-SONET WDM links. The complete CRS-1 system consumes 1 020 kW, with each 16-slot single shelf linecard consuming 13.6 kW (accounting for both dc and ac power supplies) [41]. The average energy consumption of each router port is then $\approx$850 W (derived indirectly from the Cisco router data sheets [40], similarly as [20]). The router consumes the greatest percentage of energy in communication.

As in [20], the power consumption of the WDM transponder providing the E/O and O/E conversions between the router and the optical layer is assumed to be 34.5 W (obtained also from [42]). We assume that the transponders do not consume any additional energy to support our wavelength-striped format.

One can estimate the power of the physical-layer switching fabric from the energy consumed by the active switching elements (i.e., the SOAs). The power consumption of the SOAs deployed in bufferless packet switches can be evaluated

TABLE II
OPEX ANALYSIS PARAMETERS

Parameter	Unicast	PSaD	MPMA
Switching fabric port count		4	
Maximum supported multicast fan-out	1	4	4
Request: multicast 2 packets from 1 source to 2 destinations	Transmit 2 packets serially (2 timeslots)	Transmit 2 packets once (1 timeslot)	Transmit 2 packets once (1 timeslot)
Power per IP router port (P_r) [W]		850	
Number of IP ports used (1 TX, 2 RX)	4	3	3
Total IP router power ($P_r total$) [W]	3400	2550	2550
Power per WDM transponder (P_t) [W]		73	
Number ofWDM transponders used	4	3	3
Total transponder power ($P_t total$) [W]	292	219	219
Power per PSE (empirical) [W]		6.33	
Power per SOA [W]		1.58	
Total number of PSEs	4	16	10
Total number of SOAs	16	96	40
Total switching fabric power ($P_s total$) [W]	25.32	151.84	63.3
Total power consumption [W]	3717.32	2920.84	2832.3

through analytical modeling [43], or through experimentally verified data, which show SOAs consume power in the range of tens-to-hundreds of milliwatts (in [44], the best-case SOA gate consumes ≈ 750 mW at 160 Gb/s).

Here, the energy consumption performance values for the multicast-capable OPS designs are acquired empirically. Specifically, we experimentally examine the independent power sources of the complete PSE node to determine the energy consumption required by the circuit board and its components (including the four SOA gates, *p-i-n* photodetectors, the supporting electronic circuitry, CPLD logic, etc.) The experimentally obtained power consumption for each PSE is 6.33 W. One can note that the current PSE implementation was not originally designed to optimize power consumption; the circuit board uses several electronic components for debugging and voltage probing purposes that are not integral to packet switching functionality, but still consume excess power. In the PSaD case, we further assume that the 1×4 SOA-based switch realizations will also use additional discrete SOAs and electronic circuitry (such as current drivers, etc.) Thus, the power required to supply each SOA is a quarter of the total PSE power, namely 1.58 W. These values contribute to P_s. The unicast-only and MPMA designs do not require the 1×4 SOA-based switches.

Table II shows the results of the OPEX analysis, comparing the estimated power consumption in the case of: 1) not realizing a multicast-capable fabric, 2) a $M = 4$-way multicast fanout with 4×4 PSaD, and 3) $M = 4$-way multicast fanout with 4×4 MPMA. We assume the transmission request of two packets from one source to two distinct destination ports. In the case of unicast-only fabric, the router must transmit these packets serially using two separate timeslots. In the case of PSaD and MPMA, the router only needs to transmit once in one timeslot, and the difference in power consumptions in the multicast designs comes directly from the physical layer P_s. We assume that the components consume no power when not transmitting. In this specific example, the approximated power consumption values associated with the three cases show that MPMA exhibits the best OPEX energy performance when considering the IP-layer power consumption. As expected, the unicast realization consumes the most energy due to the need for multiple router transmissions.

VI. CONCLUSION

We report on the multicasting of broadband multiwavelength optical packets for future optical networks. The two proposed designs seamlessly support the unicast and multicast of optical messages depending on the encoded packet header. A PSaD architecture is implemented using two switches, realizing a 4×4 OPS fabric. 8×10 Gb/s wavelength-striped optical messages are successfully multicasted error-free with BERs less than 10^{-12} for all payload wavelengths. We then introduce a second optimized fabric architecture for PaM that leverages the unique programmability of the basic fabric elements and is more hardware efficient. The second architecture is experimentally implemented in the 4×4 testbed, and 8×10 Gb/s optical packets are multicasted error-free with a 2.5 dB power penalty for all eight payload wavelength channels.

Using the demonstrated multicast architectures, we see the design trade-offs in enabling optical packet multicasting related to both the added routing complexity, hardware costs (CAPEX), and energy consumption (OPEX). The complexity overhead associated with multicasting can be minimized without regard to cost (i.e., by realizing a design similar to PSaD), or optimized in terms of additional hardware and components at the expense of more complex routing implementations (i.e., the proposed multistage architecture).

Realizing broadband packet multicast-capable optical fabrics may allow for energy-efficient applications, specifically in next-generation core and access/aggregation networks. As [16], we endeavor to allow multiwavelength optical packet multicasting within the core (and possibly extended to the access), while an electronic IP-layer multicast may be more efficient at the edge. This allows the core to achieve the required high-bandwidth connectivity and avoids having to deploy additional multiwavelength transceivers, which could potentially be quite costly to

realize for many edge users. These architectures may be applied to a range of fabric designs using forward-looking, low-energy optical components (even beyond SOAs). In a realistic network environment, a packaged solution with integrated SOAs (or switching devices), optical components, and electronic circuitry may serve to be more robust to thermal and environmental variations. This may also reach a higher port count, with a smaller footprint and lower energy consumption. Ultimately, we envision developing a black-box optical-multicast-capable module that packages one of the aforementioned switching fabrics for easy deployment in next-generation networks.

ACKNOWLEDGMENT

The authors would like to thank D. Brunina and H. Wang for valuable discussions.

REFERENCES

[1] R. W. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Lab. Tech. J.*, vol. 14, pp. 3–9, Feb. 2010.

[2] J. D'Ambrosia, "40 gigabit Ethernet and 100 gigabit Ethernet: The development of a flexible architecture," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. S8–S14, Mar. 2009.

[3] GreenTouch. [Online]. Available: http://www.greentouch.org/

[4] D. Kilper, G. Atkinson, S. Korotky, S. Goyal, P. Vetter, D. Suvakovic, and O. Blume, "Power trends in communication networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 275–284, Mar./Apr. 2011.

[5] W. Vereecken, W. Van Heddeghem, B. Puype, D. Colle, M. Pickavet, and P. Demeester, "Optical networks: How much power do they consume and how can we optimize this?," in *Proc. 36th Eur. Conf. Exhib. Opt. Commun.*, Sep. 2010, pp. 1–4, Paper Mo.1.D.1.

[6] C. Lange, D. Kosiankowski, R. Weidmann, and A. Gladisch, "Energy consumption of telecommunication networks and related improvement options," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 285–295, Mar./Apr. 2011.

[7] G. Raybon and P. J. Winzer, "100 Gb/s challenges and solutions," in *Proc. Opt. Fiber Commun. Conf.*, Feb. 2008, pp. 1–35, Paper OTuG1.

[8] A. Stavdas, C. T. Politi, T. Orphanoudakis, and A. Drakos, "Optical packet routers: How they can efficiently and cost-effectively scale to petabits per second [invited]," *J. Opt. Netw.*, vol. 7, no. 10, pp. 876–894, Oct. 2008.

[9] R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2002.

[10] S. J. B. Yoo, "Optical packet and burst switching technologies for the future photonic internet," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4468–4492, Dec. 2006.

[11] C. P. Lai, F. Fidler, and K. Bergman, "Experimental demonstration of QoS-aware cross-layer packet protection switching," in *Proc. 35th Eur. Conf. Opt. Commun.*, Sep. 2009, pp. 1–2, Paper 2.5.3.

[12] C. P. Lai, M. S. Wang, A. S. Garg, K. Bergman, J.-Y. Yang, M. Chitgarha, and A. Willner, "Demonstration of QoS-aware packet protection via cross-layer OSNR signaling," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2010, pp. 1–3, Paper OTuM2.

[13] NSF Engineering Research Center for Integrated Access Networks (CIAN). [Online]. Available: http://cian-erc.org/

[14] I. Baldine, "Unique optical networking facilities and cross-layer networking," in *Proc. IEEE/LEOS Summer Top. Meet.*, Jul. 2009, pp. 145–146, Paper TuD4.2.

[15] G. N. Rouskas, "Optical layer multicast: Rationale, building blocks, and challenges," *IEEE Netw.*, vol. 17, no. 1, pp. 60–65, Jan./Feb. 2003.

[16] H. S. Chung, S. H. Chang, and K. Kim, "Experimental demonstration of layer-1 multicast for WDM networks using reconfigurable OADM," *Opt. Fiber Technol.*, vol. 15, no. 5–6, pp. 431–437, Jul. 2009.

[17] S. Perrin, "White Paper: The Need for Next-Generation ROADM Networks," San Francisco, CA, Heavy Reading, Tech. Rep., 2010.

[18] Z. Pan, H. Yang, J. Yang, J. Hu, Z. Zhu, J. Cao, K. Okamoto, S. Yamano, V. Akella, and S. Yoo, "Advanced optical-label routing system supporting multicast, optical TTL, and multimedia applications," *J. Lightw. Technol.*, vol. 23, no. 10, pp. 3270–3281, Oct. 2005.

[19] H. Hamza, "Convert-and-deliver: A scalable multicast optical cross-connect with reduced power splitting fan-out," *J. Supercomput.*, pp. 1–24, Feb. 2011.

[20] G. Shen and R. Tucker, "Energy-minimized design for IP over WDM networks," *IEEE/OSA J. Opt. Commun. Network.*, vol. 1, no. 1, pp. 176–186, Jun. 2009.

[21] L. Sahasrabuddhe and B. Mukherjee, "Light trees: Optical multicasting for improved performance in wavelength routed networks," *IEEE Commun. Mag.*, vol. 37, no. 2, pp. 67–73, Feb. 1999.

[22] M. Ali and J. Deogun, "Power-efficient design of multicast wavelength-routed networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 10, pp. 1852–1862, Oct. 2000.

[23] M. A. Marsan, A. Bianco, P. Giaccone, E. Leonardi, and F. Neri, "Multicast traffic in input-queued switches: Optimal scheduling and maximum throughput," *IEEE/ACM Trans. Netw.*, vol. 11, no. 3, pp. 465–477, Jun. 2003.

[24] M. Andrews, S. Khanna, and K. Kumaran, "Integrated scheduling of unicast and multicast traffic in an input-queued switch," in *Proc. 18th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Mar. 1999, vol. 3, pp. 1144–1151.

[25] N. Yan, I. T. Monroy, H.-D. Jung, T. Koonen, A. Teixeira, and T. Silveira, "Optical multicast technologies by multi-wavelength conversion for optical routers," in *Proc. Int. Conf. Commun. Technol.*, Nov. 2006, pp. 1–4.

[26] Q. Huang and W.-D. Zhong, "Wavelength-routed optical multicast packet switch with improved performance," *J. Lightw. Technol.*, vol. 27, no. 24, pp. 5657–5664, Dec. 2009.

[27] D. Tutsch and G. Hommel, "Performance of buffered multistage interconnection networks in case of packet multicasting," in *Proc. Adv. Parallel Distrib. Comput.*, Mar. 1997, pp. 50–57.

[28] X. Liu, H. Wang, and Y. Ji, "Hybrid multicast mode in all-optical networks," *IEEE Photon. Technol. Lett.*, vol. 19, no. 16, pp. 1212–1214, Aug. 2007.

[29] P. J. Winzer, "Beyond 100G Ethernet," *IEEE Commun. Mag.*, vol. 48, no. 7, pp. 26–30, Jul. 2010.

[30] W. S. Hu and Q. J. Zeng, "Multicasting optical cross connects employing splitter-and-delivery switch," *IEEE Photon. Technol. Lett.*, vol. 10, no. 7, pp. 970–972, Jul. 1998.

[31] C. P. Lai and K. Bergman, "Demonstration of programmable broadband packet multicasting in an optical switching fabric test-bed," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2009, pp. 1–3, Paper OTuA5.

[32] C. P. Lai and K. Bergman, "Network architecture and test-bed demonstration of wavelength-striped packet multicasting," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2010, pp. 1–3, Paper OWI4.

[33] A. Shacham and K. Bergman, "An experimental validation of a wavelength-striped, packet switched, optical interconnection network," *J. Lightw. Technol.*, vol. 27, no. 7, pp. 841–850, Apr. 2009.

[34] C. P. Lai, D. Brunina, and K. Bergman, "Demonstration of 8×40-Gb/s wavelength-striped packet switching in a multi-terabit capacity optical network test-bed," in *Proc. 23rd Annu. Meet. IEEE Photon. Soc.*, Nov. 2010, pp. 688–689, Paper ThQ2.

[35] A. Shacham, B. G. Lee, and K. Bergman, "A wideband nonblocking 2×2 switching node for a SPINet network," *IEEE Photon. Technol. Lett.*, vol. 17, no. 12, pp. 2742–2744, Dec. 2005.

[36] C. P. Lai, A. Shacham, and K. Bergman, "Demonstration of asynchronous operation of a multiwavelength optical packet-switched fabric," *IEEE Photon. Technol. Lett.*, vol. 22, no. 16, pp. 1223–1225, Aug. 2010.

[37] O. Liboiron-Ladouceur, B. A. Small, and K. Bergman, "Physical layer scalability of WDM optical packet interconnection networks," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 262–270, Jan. 2006.

[38] W. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, CA: Morgan Kaufmann, 2003.

[39] S. Okamoto, A. Watanabe, and K.-I. Sato, "Optical path cross-connect node architectures for photonic transport network," *J. Lightw. Technol.*, vol. 14, no. 6, pp. 1410–1422, Jun. 1996.

[40] Cisco, Cisco CRS-1 Data Sheets [Online]. Available: http://www.cisco.com

[41] G. Epps, D. Tsiang, and T. Boures, "System Power Challenges," presented at the Cisco Routing Res. Semin., San Jose, CA, Aug. 2006.

[42] Alcatel-Lucent, Alcatel-Lucent WaveStar OLS 1.6T Product Specification [Online]. Available: http://www.alcatel-lucent.com

[43] V. Eramo and M. Listanti, "Power consumption in bufferless optical packet switches in SOA technology," *IEEE/OSA J. Opt. Commun. Network.*, vol. 1, no. 3, pp. B15–B29, Aug. 2009.

[44] J. Sakaguchi, F. Salleras, K. Nishimura, and Y. Ueno, "Frequency-dependent electric DC power consumption model including quantum-conversion efficiencies in ultrafast all-optical semiconductor gates around 160 Gb/s," *Opt. Exp.*, vol. 15, no. 22, pp. 14887–14900, Oct. 2007.

Caroline P. Lai (S'07–M'12) received the B.A.Sc. degree (with Hons.) from the University of Toronto, Toronto, ON, Canada in 2006, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, in 2008 and 2011, respectively, all in electrical engineering.

She is currently a Postdoctoral Researcher in the Photonic Systems Group, Tyndall National Institute, University College Cork, Cork, Ireland. Her current research interests include energy-efficient photonic technologies and cross-layer communications for future optical access and metro transport networks, in addition to optical interconnects, optically-connected memory, and power-efficient optical links for high-performance computing systems.

Dr. Lai is a member of the IEEE Photonics Society and the Optical Society of America.

Keren Bergman (S'87–M'93–SM'07–F'09) received the B.S. degree from Bucknell University, Lewisburg, PA, in 1988, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively, all in electrical engineering.

She is currently a Professor and Chair of electrical engineering at Columbia University, New York, NY, where she also directs the Lightwave Research Laboratory. She leads multiple research programs on optical interconnection networks for advanced computing systems, data centers, optical packet switched routers, and chip multiprocessor nanophotonic networks-on-chip.

Dr. Bergman is a Fellow of the Optical Society of America and currently serves as the Co-Editor-in-Chief of the *IEEE/OSA Journal of Optical Communications and Networking*.