

First Demonstration of a Cross-Layer Enabled Network Node

Caroline P. Lai, *Member, IEEE*, Daniel Brunina, *Member, IEEE*, Brandon W. Buckley, Cedric Ware, Wenjia Zhang, Ajay S. Garg, *Student Member, IEEE*, Bahram Jalali, *Fellow, IEEE*, and Keren Bergman, *Fellow, IEEE*

Abstract—Exploding traffic demands and increasing energy consumptions facing today’s networks are driving the designs of next-generation networking technologies. Cross-layer enabled approaches will allow for the packet-level control of the optical layer, to enable dynamic resource allocation and traffic engineering at the physical layer. We demonstrate an intelligent cross-layer enabled network node that can support high-bandwidth, all-optically routed packets, using emerging photonic technologies including optical packet switched fabrics and packet-scale performance monitoring. Using a cross-layer control and management plane, the node can dynamically optimize optical switching based on higher-layer constraints such as quality-of-service and energy consumption, as well as quality-of-transmission metrics such as link integrity and bit-error rates.

We demonstrate a first-generation prototype of the cross-layer node, outlining its architecture and major implemented subsystems. The packet-rate physical-layer reconfiguration of the node’s fabric is shown using an implemented performance monitor and control plane. The realized node supports 8×40 -Gb/s wavelength-striped optical packets with pseudorandom data with error-free transmission (bit-error rates less than 10^{-12}), in conjunction with the heterogeneous transmission of video traffic using 10-Gigabit Ethernet optical network interface cards based on field-programmable gate arrays.

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

Manuscript received August 14, 2012; revised October 31, 2012 and March 04, 2013; accepted March 05, 2013. Date of publication March 13, 2013; date of current version March 29, 2013. This work was supported in part by the NSF Engineering Research Center on Integrated Access Networks (CIAN) under Grant EEC-0812072, and in part by the NSF Future Internet Design (FIND) 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, Lee Maltings, University College Cork, Cork, Ireland (e-mail: caroline.lai@tyndall.ie).

D. Brunina was with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA. He is now with the Photonic Systems Group, Tyndall National Institute, Lee Maltings, University College Cork, Cork, Ireland (e-mail: daniel.brunina@tyndall.ie).

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

B. W. Buckley and B. Jalali are with the Department of Electrical Engineering, University of California Los Angeles, Los Angeles, CA 90095 USA (e-mail: bbuckley@ucla.edu; jalali@ucla.edu).

C. Ware is with the Institut Mines-Télécom, Télécom ParisTech, CNRS LTCl, 75634 Paris CEDEX 13, France (e-mail: cedric.ware@telecom-paristech.fr).

W. Zhang was with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA. He is now with SMART, Singapore 138602 (e-mail: wenjia@smart.mit.edu).

A. S. Garg was with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA. He is now with MIT Lincoln Laboratory, Lexington, MA 02421 USA (e-mail: ajay.sinclair.garg@ieee.org).

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.2013.2252421

I. INTRODUCTION

REPUTABLE traffic forecasts show that networks will be required to support exponentially-increasing bandwidths [1], with growth rates of $\approx 60\%$ per year [2]. This trend is driven primarily by high-bandwidth applications, such as video and media services [3]. Challenges facing today’s infrastructure emphasize the importance of designing a future Internet that can seamlessly support the necessary data rates. Additionally, the network’s power consumption is predicted to grow rapidly in the next decade and become a key bottleneck [4], [5]. The current network cannot viably sustain this massive traffic growth [6], nor can following incremental technology advances continue to meet demands at sustainable energy consumptions [7], [8]. These factors emphasize the need for new energy-efficient, high-bandwidth optical technologies and devices.

The ultimate goal of this work is to enable high-speed, agile network capabilities while simultaneously maintaining low power consumptions [9]. To this end, we propose the design of a cross-layer communications framework (Fig. 1). Our cross-layer approach will enable greater intelligence and network functionality on the optical layer through the development of a *cross-layer optical switching node* [10]. The cross-layer node, hereafter also referred to as the *cross-layer box (CLB)*, will use advanced optical switching technologies to enable a more intelligent optical layer with greater flexibility in switching. In addition, the node will use packet-scale measurement and performance monitoring (PM) subsystems to analyze the health of the optical channel, and feedback this information on the optical quality-of-transmission (QoT) to higher routing layers. This will create an ‘introspective’ access to the physical layer, allowing the routers and switches to be aware of physical-layer impairments (PLIs). Thus, our cross-layer signaling approach will offer real-time knowledge of the physical layer.

In addition to allowing the optical signals’ QoT to be sent upward to the routing layers, the bidirectional cross-layer signaling design also enables higher-layer attributes to flow down to the optical layer. These higher-layer parameters can consist of quality-of-service (QoS) metrics and energy constraints that may be set by higher network layers, e.g. the IP layer (layer 3). The cross-layer design allows these parameters to flow down to the physical layer, thus supporting a flexible optical layer where optical switching can be optimized while accounting for these higher-layer parameters.

This approach ultimately yields a multi-layer optimized networking framework that can incorporate knowledge of the optical signals, with the resulting optical layer also having

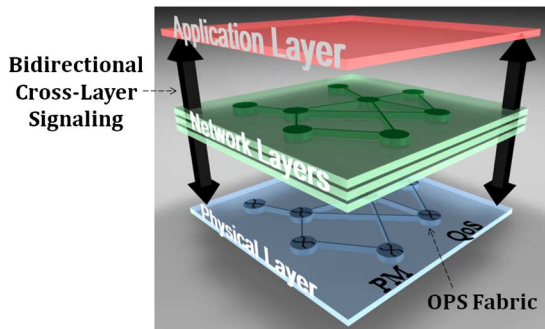


Fig. 1. Block schematic of the envisioned cross-layer-enabled network stack, which can support a bidirectional exchange of control signals between the network layers. Optical-layer switching algorithms can then account for higher-layer network parameters such as quality-of-service (QoS). The QoS-aware physical layer uses optical packet switched (OPS) fabrics and integrated performance monitoring (PM) devices.

advanced higher-layer switching functionalities. The CLB will be able to dynamically optimize optical switching based on both real-time PM measurements as well as the higher-layer attributes. Optical switching and multi-layer traffic engineering can then be executed with minimal energy consumption, while maximizing delivered bandwidth and maintaining quality guarantees. In summary, the cross-layer framework will allow for dynamic physical-layer introspection to detect optical signal degradations on a packet-by-packet basis, while providing the optical layer with awareness of higher-layer constraints. The CLB achieves high-bandwidth bit rates (using optical switching) with high optical signal quality (using advanced PM techniques).

Though the CLBs may be deployed in the core, we plan to leverage this technology in access/aggregation networks [11]. These are traditionally implemented with layer-3 (IP) routers; however, electronic switching is reaching fundamental limits with respect to achievable bandwidths and energy [12]. Though passive optical networks (PONs) are feasible in access networks, active opto/electronic switches with advanced cross-layer capabilities may open new avenues to achieve high energy efficiencies.

One of the major subsystems of the CLB is a high-bandwidth optical packet switch. Optical packet switching (OPS) [13] can facilitate the broadband transmission of wavelength-parallel optical packets via wavelength-division multiplexing (WDM), with fast switching speeds and data-rate transparency. Here, we aim to demonstrate OPS with a high level of network capability using an optical switching fabric with advanced photonic switching functionalities, such as packet multicasting and the support for optical QoS constraints. Broadband applications can be supported with lower cost by migrating these higher-layer capabilities lower in the network protocol stack to the optical layer [14].

Although OPS may allow future networks to reach ultrahigh capacities, by reducing the number of optical/electrical/optical (O/E/O) conversions and using fewer electronic components, the system loses access to advantages provided by electronic regeneration and grooming. Since these are important to preserve adequate signal integrity for end-to-end links, this evolution results in the overall network becoming more sensitive to

PLIs. For the CLB's cross-layer signaling, fast PM techniques are essential to quickly detect PLIs. These subsystems monitor the optical-layer performance to capture the optical signal quality, e.g. by evaluating the bit-error rate (BER) and/or other optical properties (e.g. loss, optical power, optical-signal-to-noise ratio (OSNR), etc.). Based on these measurements, which can feedback to the higher-layer routing layers, as well as on the higher-layer (IP) constraints, the dynamic management of optical switching at the scale of both packets and flows can be performed, and complete optical switching affected. A distributed control plane architecture and routing protocols can then use these inputs to enable cross-layer functionality.

In this paper, we present for the first time the complete architecture of the CLB, and demonstrate a first-generation prototype that can support heterogeneous traffic. To the authors' knowledge, this comprises the first realization of a cross-layer optimized node. Not all aspects of the node have been realized at this time; however, several fundamental subsystems have been implemented in this demonstrator. The design, development, and implementation of the optical switching fabric is outlined. The experimental demonstration of this initial prototype shows a dynamic network element with distributed control plane management, fast packet-rate optical switching capabilities, and embedded physical-layer PM modules.

The CLB supports heterogeneous traffic, comprised of multiwavelength optical packets with 8×40 -Gb/s wavelength-striped pseudorandom payloads and 4×2.5 -Gb/s circuit-switched high-definition (HD) video traffic streaming from a 10-Gigabit Ethernet (10GE) optical network interface card (O-NIC) that uses field-programmable gate arrays (FPGAs). The fabric's packet-rate reconfiguration is shown using an exemplary PM that monitors the optical signal quality (via the BER). Error-free transmission (confirmed with BERs less than 10^{-12} on all payload channels) is obtained for the wavelength-striped messages. Finally, depending on the optical QoT, the HD video's bit rate can be varied. Typical error-free BERs in the 10^{-12} range are envisioned for the CLB system, since no forward-error correction (FEC) is assumed in the network interfaces, the optical devices can seamlessly operate with this high-performance requirement, and the implemented PM can measure these low BERs. The requirement for these low BERs can be easily relaxed by setting less stringent BER or OSNR constraints and assuming the use of FEC.

The remainder of this paper is organized as follows. The CLB's architecture and prototype is outlined in Section II. Section III details the complete experimental demonstration, including the OPS fabric and test-bed. The PM system and accompanying demonstration is described in Section IV. The CLB's ability to support an application-driven demonstration (i.e. video streaming) is discussed in Section V. Section VI concludes the paper.

II. CLB ARCHITECTURE AND SUBSYSTEMS

A. Overview

The proposed CLB is a novel, intelligent optical aggregation network node which can enable OPS while simultaneously delivering high optical QoT and maintaining application-specific

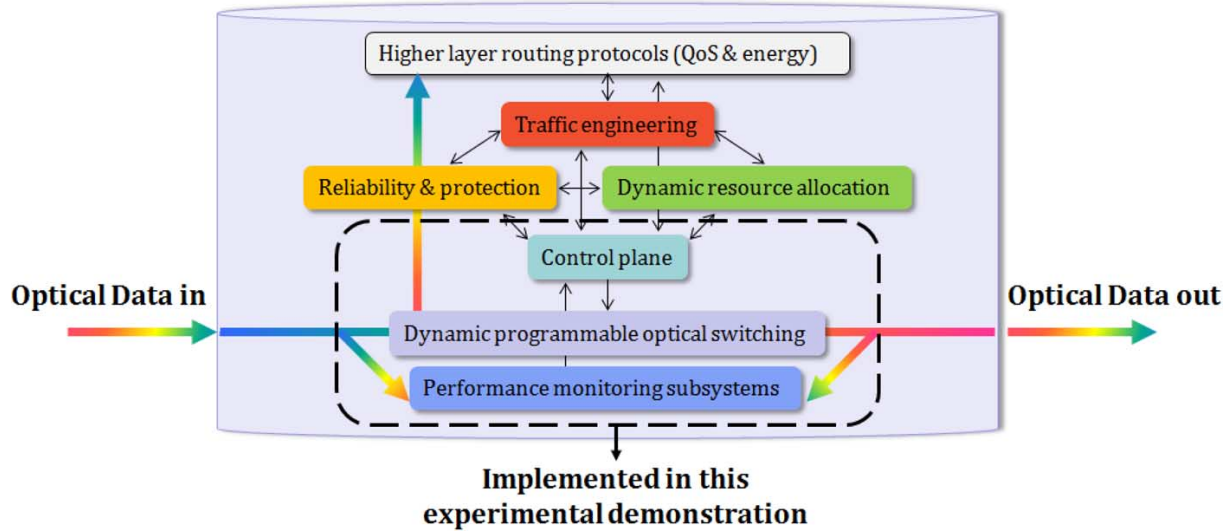


Fig. 2. Block schematic of the CLB, detailing the various subsystems that compose the box and showing the components that have been realized in this experimental demonstration.

QoS constraints. The complete envisioned cross-layer platform will realize various dynamic routing applications and support various multi-layer optimization and traffic engineering protocols, to allow for the co-optimization of QoS and QoT with energy awareness [9], [15].

The CLB prototype demonstrated here supports heterogeneous aggregation traffic and high-bandwidth applications, while optimizing the performance of the switched optical data. The node allows optical packet switching to be triggered by optical signal degradation measurements. The option to react to the awareness of the optical channel properties and performance at a packet-rate timescale can also be dependent on energy metrics via cross-layer control plane algorithms.

Fig. 2 depicts a high-level block schematic of the CLB. The node is composed of the following key subsystems:

- an opto-electronic switching fabric with packet-rate reconfiguration, all-optical data switching, and advanced physical-layer functionalities;
- dynamic performance measurement subsystems;
- a distributed cross-layer control plane; and
- cross-layer network routing protocols enabled by higher-layer interactions.

Compared to conventional network nodes in which control signals and rerouting optimization algorithms are primarily unidirectional, the proposed CLB supports more advanced routing techniques that can actuate packet-level or flow-based rerouting based on PM measurements, as well as requirements set by high-bandwidth applications (e.g. HD video transmission shown here).

B. Current Prototype

A first-generation benchtop prototype of the CLB is described here for the first time. The current prototype is constructed using commercially-available, off-the-shelf components; future work involves designing a node with integrated functionalities and a smaller footprint. As shown in Fig. 2 and further discussed

in this section, the focus of the experiment is on the following components that have been realized in the test-bed:

- an all-optical switching fabric;
- the TiSER monitoring subsystem; and
- a cross-layer control and management plane.

C. Optical Switching Fabric

A crucial subsystem is a fabric that supports dynamic all-optical data switching (shown in Fig. 2). The basic requirement for the fabric is to perform bit-rate-transparent all-optical data switching, specifically supporting both wavelength-striped optical packets and longer video-based flows.

This current implementation uses a multi-terabit-capacity optical switching fabric composed of 2×2 broadband non-blocking photonic switching elements (PSEs), which are organized as a transparent multi-stage 4×4 interconnect and controlled in a distributed manner using complex programmable logic devices (CPLDs) [16], [17]. The fabric provides a means of interconnecting multiple high-bandwidth input/output ports in next-generation Internet routers and switches [18]. The complete switching fabric is opto-electronic in the sense that the PSEs leverage the optical domain for high-bit-rate data transmission and the electronic domain for packet-level control processing. The PSEs are composed of an all-optical switch with a dedicated electronic control structure, where the data remains entirely in the optical domain end-to-end.

As depicted in Fig. 3, each 2×2 PSE is constructed using macro-scale components, including four semiconductor optical amplifier (SOA) gates, which are organized in a broadcast-and-select topology. The SOAs provide a wide wavelength band (approximately the International Telecommunication Union (ITU) C-band), in addition to transparency to the optical packets' data format and bit rate, fast nanosecond-scale switching speeds, and built-in optical gain. In this demonstrator, the optical messages have lengths that are longer than the PSEs; thus, no optical storage or buffering is available. Packets are dropped in the case of message contention within the fabric by the PSE's CPLD

TABLE I
LIST OF ACRONYMS

Acronym	Definition
10GE	10-Gigabit Ethernet
A/D	Analog-to-digital
ASE	Amplified spontaneous emission
BER	Bit-error rate
BERT	Bit-error-rate tester
CLB	Cross-layer box
CPLD	Complex programmable logic device
CW	Continuous wave
DCA	Digital communications analyzer
DFB	Distributed feedback
DTG	Data timing generator
EDFA	Erbium-doped fiber amplifier
FEC	Forward-error correction
FPGA	Field-programmable gate array
GPIO	General purpose input/output
HD	High definition
IP	Internet Protocol
ITU	International Telecommunication Union
LA	Limiting amplifier
MLL	Mode-locked laser
MZ	Mach-Zehnder
NIC	Network interface card
NRZ-OOK	Nonreturn-to-zero on-off-keying
O/E/O	Optical/electrical/optical
O-NIC	Optical network interface card
OPS	Optical packet switching
OSA	Optical spectrum analyzer
OSNR	Optical-signal-to-noise ratio
PC	Polarization controller
PD	Photodetector
PLI	Physical-layer impairment
PM	Performance monitoring
PON	Passive optical network
PPG	Pulse pattern generator
PRBS	Pseudorandom bit sequence
PSE	Photonic switching element
QoS	Quality-of-service
QoT	Quality-of-transmission
QSFP	Quad Small Form-factor Pluggable
RBS	Real-time burst sampling
ROADM	Reconfigurable optical add-drop multiplexer
RZ-DQPSK	Return-to-zero differential quaternary phase-shift keying
SOA	Semiconductor optical amplifier
TIA	Transimpedance amplifier
TiSER	Time-stretch enhanced recording
UDP	User Datagram Protocol
VBR	Variable bit rate
VOA	Variable optical attenuator
WDM	Wavelength-division multiplexing
XAUI	10-Gigabit Attachment Unit Interface

control logic. Although contention is resolved here by simple packet dropping, future work with advanced control plane protocols can allow for the fabric to deflect one of the contending packets to an available output port of the PSE.

Several PSEs are connected to create a multistage fabric topology. As shown in Fig. 4, four PSE building block structures are arranged to realize a two-stage, 4×4 switching fabric for this experiment. The switching control logic is synthesized within a CPLD located within each PSE, providing a high

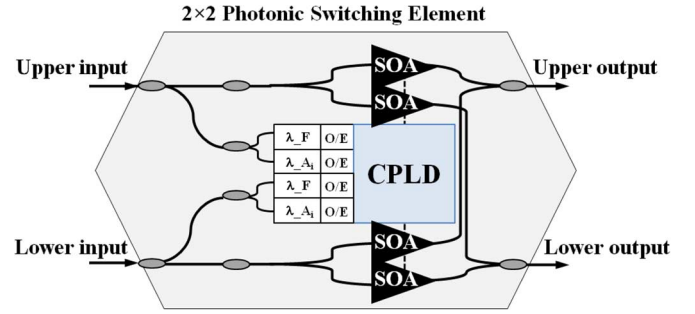


Fig. 3. Schematic of the 2×2 photonic switching element building block. The element supports two independent inputs and outputs, and uses one CPLD to realize the electronic logic.

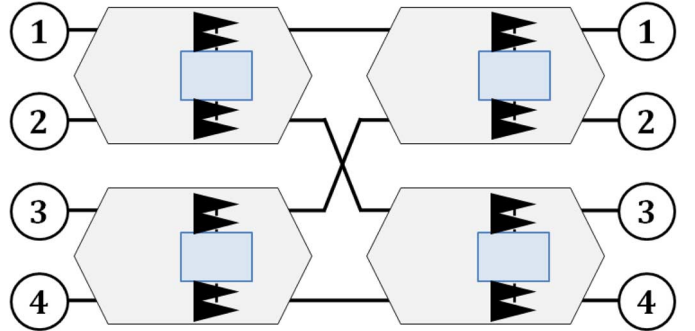


Fig. 4. Two-stage, 4×4 switching fabric implementation.

level of programmability to reconfigure the physical connections between PSEs. This basic topology uses a multistage binary banyan design, that requires $\log_2(N)$ of identical stages to create a $N \times N$ interconnect to map a large number of ports [19]. Each stage consists of $N/2$ photonic switching elements, connected in a perfect-shuffle arrangement. In the simple topology in Fig. 4, the 4×4 switching fabric requires $\log_2(N) = 2$ stages of $N/2 = 2$ PSEs (i.e. $N = 4$). The use of banyan network topologies allows for a high level of scalability, since these designs can achieve large port counts with many fabric stages. Messages are injected using the input terminals of the fabric, ingressing via the independent input ports, and are transparently and all-optically routed at each PSE.

In this experiment, the implemented hybrid opto-electronic switching fabric supports synchronous, time-slotted operation, with packets of equal lengths arriving at the PSEs in fixed timeslots. The fabric thus enables the fast, synchronous all-optical data switching of wavelength-striped messages; the optical packet structure and time-slotted approach are shown in Fig. 5. This wavelength-striping approach allows messages to achieve high aggregate transmission bandwidths by leveraging WDM's large bandwidth potential and allocating the message data to parallel wavelengths that simultaneously contain payload data. These multiwavelength messages consist of control header information (i.e. the frame, address, and QoS bits), which are encoded on a subset of dedicated frequencies, modulated at a single bit per wavelength per timeslot. The control includes: a frame signal F , denoting the presence of a packet and spanning the entire length of the packet; address signals (represented by A_i, A_j in Fig. 5), denoting the packet's destination port within

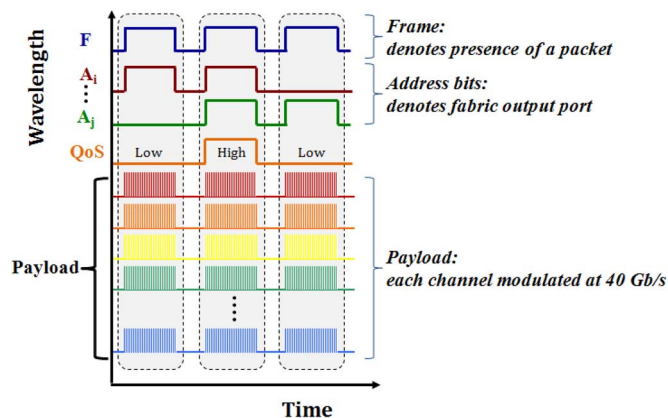


Fig. 5. Supported wavelength-striped optical packet format. The control signals include the frame, address bits, and QoS; the payload channels are each modulated at 40 Gb/s in this experiment. The control and payload signals have equal lengths and are aligned to the start of fixed timeslots (here, packets are 32 μ s long).

the switching fabric; and a QoS information bit, denoting the packet's priority class (as indicated by a higher-layer protocol). By allowing the control wavelengths to remain high for the duration of the optical message, the PSE's switching state remains constant as messages propagate through the fabric. Simultaneously, the packet's payload data is fragmented and modulated a high data rate (e.g. at 40 Gb/s per data payload channel) on the rest of the supported frequency band (here, the ITU C-band). In a single packet, the control and payload signals on all the wavelength channels are of equal length and are synchronized in time using fixed timeslots ('cells'), i.e. the start of the control bits is aligned with the first data bit of the packet. Thus, the hybrid opto-electronic switching fabric allows the payload data to remain completely in the optical domain as it traverses the PSE, whilst the switching decisions are performed electronically. Any skew between wavelengths in a packet that arises from chromatic dispersion may be mitigated in future implementations by including a longer guardtime between packets; this ensures that the control headers are maintained over the payload duration even after some dispersion.

The OPS design enables packet-rate control header processing, wherein the message header is instantaneously decoded at each PSE and the routing control decision can be made upon reception of the packet's leading edge [16]. The elements' electronic control logic is distributed among the individual PSEs using high-speed programmable logic (i.e. the CPLDs), yielding a high level of routing flexibility. The message payload data and routing control headers are transmitted concurrently to the PSEs and propagate together end-to-end in the fabric. At each of the 2×2 PSEs, the actual routing decision is based on the control header extracted from the packet. The leading edge of the optical packet is detected and received at one of the input ports. The framing and address bit signals are extracted immediately using fixed wavelength filters and low-speed *p-i-n* optical receivers. The PSE's switching state is based solely on the information encoded in the optical header, which is recovered from the incoming packet and processed by high-speed electronic circuitry. The CPLD electronically

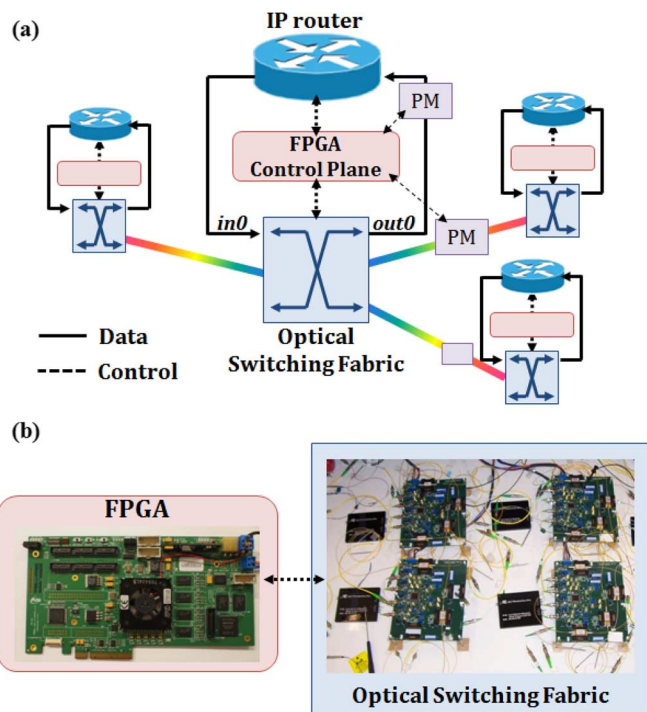


Fig. 6. (a) Systems architecture depicting four network nodes, each composed of an IP router, an optical switching fabric, a distributed FPGA-based control and management plane, and PM systems. In order to dynamically manage optical switching, a cross-layer interface is offered by the FPGA, which supports the exchange of control signals between the router, the optical network layer, and PM modules; (b) Photographs of the FPGA circuit board and OPS fabric that were realized as one network node in this experiment.

drives the appropriate SOA gates, and the optical messages are then routed to their encoded destination (or dropped upon contention). The switching control is distributed among the PSEs using simple, combinational logic, and no additional signals are exchanged between them. The elements also do not add (or subtract) information to/from the optical messages. The payload information is not decoded by the PSE logic and is simply routed transparently using one of the four SOAs. Successfully switched messages set up end-to-end transparent lightpaths between fabric terminals. The use of reprogrammable CPLDs results in straightforward reconfigurability and the potential for supporting different routing protocols and logic. The fabric has also been shown to operate in an asynchronous fashion with a simple modification to the CPLD control logic [20].

The fast speeds associated with the SOAs' switching and the electronic logic allows for the optical fabric to exhibit nanosecond-scale reconfiguration response times. This is key for future switching fabrics to perform fast switching and path provisioning in the case of router failure or link degradation, in order to recover and potentially route around PLIs. The network architecture for the switching fabric reconfiguration is shown in Fig. 6(a).

D. Performance Monitoring Subsystem

The second key component in the CLB is a packet-level performance monitor to allow for real-time evaluation of the optical data on a packet-by-packet basis (shown in Fig. 2 as 'Performance monitoring subsystems'). In this experiment, the realized

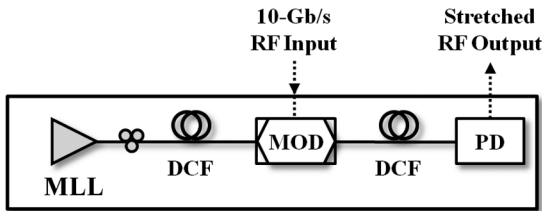


Fig. 7. Block diagram illustrating the physics of the time-stretch pre-processor used by TiSER.

PM is the photonic time-stretch enhanced recording (TiSER) oscilloscope [21], [22], which can characterize PLIs and realize a diagnostic, PM tool for the lightpaths in the CLB. TiSER provides the real-time digitization of high-speed signals, and can extrapolate the optical packets' BER on a message timescale to allow for dynamic cross-layer interactions. The CLB can also use other previously-demonstrated PM systems that can monitor the packet-rate OSNR [23].

Here, TiSER is inserted in the CLB and is used to generate real-time eye diagrams and monitor the BER [24]. Cross-layer algorithms can then use the BER measurements to reconfigure optical switching with rapid capacity provisioning. The system uses real-time burst sampling (RBS) [21] to effectively slow down electronic signals for high-speed digitizers, mitigating potential bandwidth limitations of analog-to-digital (A/D) converters in future receivers and allowing the capture of the optical eye diagrams of the packets' 40-Gb/s payload channels. In order to realize performance monitoring, the signals' quality (Q) factor are extracted on a packet timescale from the eye diagrams. TiSER allows each data channel in the multiwavelength packet to scale to higher data rates with low BERs.

The TiSER oscilloscope uses photonic time-stretch pre-processing (Fig. 7) to perform RBS of high-speed data signals. TiSER captures a burst of samples in real-time and reconstructs the corresponding eye diagrams in equivalent-time mode. It enables the capture of fast non-repetitive dynamics at the modulation rate, comprising a rapid monitoring solution for high-data-rate optical links and a means of providing cross-layer signaling. TiSER has been shown to capture data signals up to 45 Gb/s [21], as well as high-speed signals with advanced modulation formats (e.g. 100-Gb/s return-to-zero differential quaternary phase-shift keying (RZ-DQPSK)) [25]. RBS captures bursts of measurement samples in real-time in each sampling period [21], yielding high bandwidth performance and real-time sampling within the captured bursts. By capturing high-speed signals using slower commercially-available digitizers, TiSER bridges the gap in measurement functionality and performance between sampling oscilloscopes and real-time digitizers.

In its current implementation, TiSER (Fig. 7) uses a mode-locked laser (MLL) that generates 36-MHz ultra-short optical pulses. A -20 -ps/nm dispersion-compensating fiber (DCF) then creates chirped pulses with a sufficient time aperture to support 40-Gb/s RF data rates. A 40-Gb/s Mach-Zehnder (MZ) intensity modulator encodes the 40-Gb/s data signal over the chirped pulses. Propagation through a span of -1310 -ps/nm DCF stretches the modulated optical pulses in time, realizing a stretch factor of approximately 70. A 40-Gb/s photodetector

(PD) receives the pulses and creates an RF signal that is a stretched version of the original with reduced bandwidth. A commercial A/D digitizer with 2-GHz bandwidth is used and the eye diagram is constructed using the recorded data by removing an integral number of data periods from the stretched timescale.

The first TiSER prototype [26] is implemented in a 19-inch rackmount chassis, which can accommodate the electronic A/D converter. All of the pre-processor components are integrated in the TiSER chassis. The inputs consist of a RF signal, a RF trigger, and a MZ modulator voltage (approximately 4 Vdc). The output ports include the stretched RF signal, the digitized data, and clock.

The use of TiSER can allow for the real-time estimation of the packets' BER from the eye diagram. This may not only be faster than a conventional BER tester (BERT) (since the number of bit errors does not need to be counted), but also allows for the BER to be extrapolated on a packet-by-packet basis. The rapid QoT computation on OPS packet timescales can be achieved using fast FPGAs and/or custom-designed electronic circuitry; TiSER is key to enabling these real-time estimation algorithms by effectively 'slowing down' the high-speed optical signals to accommodate the data rate of the FPGA's transceivers. In this specific work, offline signal processing (in Matlab) is first leveraged to generate the eye diagram using the acquired data from the A/D digitizer, then implements a BER estimation algorithm to ascertain a relatively accurate value for the BER from the opening of the eye using an eye-mask method.

E. Control Plane

The third component of the prototype that we demonstrate is a control plane that would interface the CLBs and physical-layer PM devices with higher-layer router nodes ('Control Plane' in Fig. 2). The control plane enables packet-rate reconfiguration and feedback from the optical layer.

In this implementation, an external FPGA controller acts as the fabric's control and management plane. Control signals from the higher layers and/or embedded physical-layer PM devices can trigger reconfiguration and reroute messages on the optical layer. This allows the fabric to be reconfigured based on communication between the optical and network layers. The use of the FPGA controller, together with SOA-based nanosecond switching, enables fast cross-layer fabric recovery [27].

Fast optical-layer reconfiguration is necessary for the underlying optical network to account for higher-layer parameters (e.g. QoS) in our cross-layer approach. In the case of IP router failures (or if a router is placed into sleep mode to reduce energy consumption [28]), lightpaths between end nodes in an all-optical network can be maintained by reprovisioning the optical connections around the failures (or sleeping routers). The packet-rate reconfiguration of the switching fabric also allows more seamless optical lightpath bypass [29].

In this first prototype realization, one FPGA device implements the control plane for one CLB node. In a future multi-CLB network, we envision a distributed control plane composed of multiple FPGAs to manage the control signals for multiple CLBs. The FPGA controller in this demonstrator accepts inputs consisting of QoT feedback from the optical

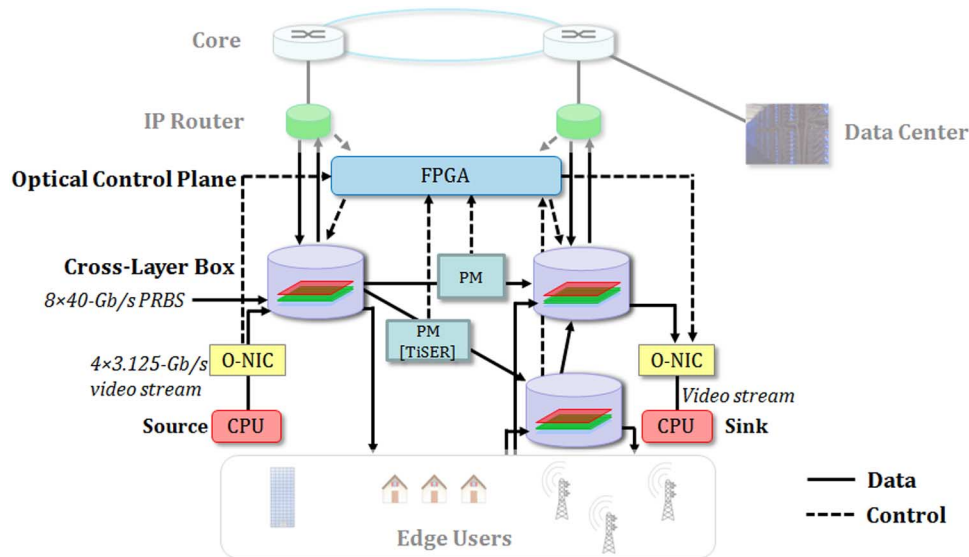


Fig. 8. Overview of the CLB experimental demonstration, depicting the cross-layer bidirectional signaling infrastructure. The FPGA represents the control plane and PM blocks denote the performance monitoring subsystems. The node supports $8 \times 40\text{-Gb/s}$ optical packets with PRBS, and 10GE-based video streams using an optical network interface card (O-NIC) connected to two computers (CPUs). The faded blocks correspond to future work and serve to show the context of this experiment.

channel to implement optical path computation and light-path/packet rerouting (if necessary). When a degraded link is detected, the FPGA-based control plane is notified, which can compute the back-up path(s) to route around the failure. The controller then signals the optical fabric (i.e. gates on/off the appropriate SOAs) to ensure the packets and/or circuit streams can be successfully routed.

For a multi-CLB network (as depicted in Fig. 8), the distributed control plane would also realize an advanced failure detection and localization mechanism. The control plane allows for the crucial advance of being able to reconfigure the CLB's fabric at the packets' nanosecond timescale upon the detection of either a failed higher-layer router and/or degraded optical signals. In this way, the optical-layer data can be rerouted within the fabric to maintain a high QoT as determined by the embedded PM, minimizing traffic loss and packet dropping. In comparison, typical current aggregation systems exhibit millisecond-scale operation.

III. EXPERIMENTAL IMPLEMENTATION

A. Overview of Experiment

The currently CLB demonstrator is composed of a reconfigurable multi-terabit optical switching fabric, packet-level TiSER performance monitor, and FPGA control plane. With these subsystems, we experimentally demonstrate the CLB's optical switching capabilities with both packet- and circuit-switched data. The current prototype allows the optical packets' BER to be measured to enable packet protection switching and message rerouting.

In this experiment, the first-generation CLB is demonstrated to showcase a select number of example capabilities. Specifically, the switching fabric can support the aggregation of multiple data rates via the simultaneous transmission of both pseudorandom and real video data traffic. The prototype supports:

- $8 \times 40\text{-Gb/s}$ wavelength-striped optical packets, with each payload wavelength carrying a 40-Gb/s nonreturn-to-zero on-off-keyed (NRZ-OOK) signal with pseudorandom bit sequence (PRBS) data; and
- $4 \times 3.125\text{-Gb/s}$ 10GE-based HD video data streams.

In this way, the joint transmission of packet data from a high-bandwidth source and circuit video data from the O-NIC provides a viable demonstration of realistic network operation [30]. This particular dimension also shows the support for concurrent packet- and circuit-switched lightpaths within the switching fabric at a given time (as in [31]).

Fig. 8 depicts the complete network architecture, with several CLBs in a mesh, and shows the underlying control and data links connecting the CLBs and control plane. The implemented experiment is also shown, indicating the two different data streams (the blocks that were not realized are faded). The fabric is confirmed to transmit both data streams successfully, with error-free operation.

We experimentally demonstrate the fast per-packet reconfiguration of the switching fabric using the FPGA controller using the two distinct traffic loads, at a nanosecond packet rate. A two-part experiment is performed, with both parts occurring simultaneously. A detailed experimental setup for the complete demonstration is shown in Fig. 9.

The first part of the demonstration leverages the large multi-terabit capacity of the switching fabric. The QoT of the $8 \times 40\text{-Gb/s}$ optical packets is assessed using TiSER, monitoring one of the 40-Gb/s payload channels at the output of the fabric. This is indicated by the blue/upper region of Fig. 9. Upon the detection of a failure or a degraded link (as indicated by TiSER), the control plane can then signal the switching fabric to modify its switching state to reroute the optical packets and dynamically avoid the PLI.

The second part of the demonstration utilizes a custom-designed 10GE O-NIC interface [32] to support the transmission

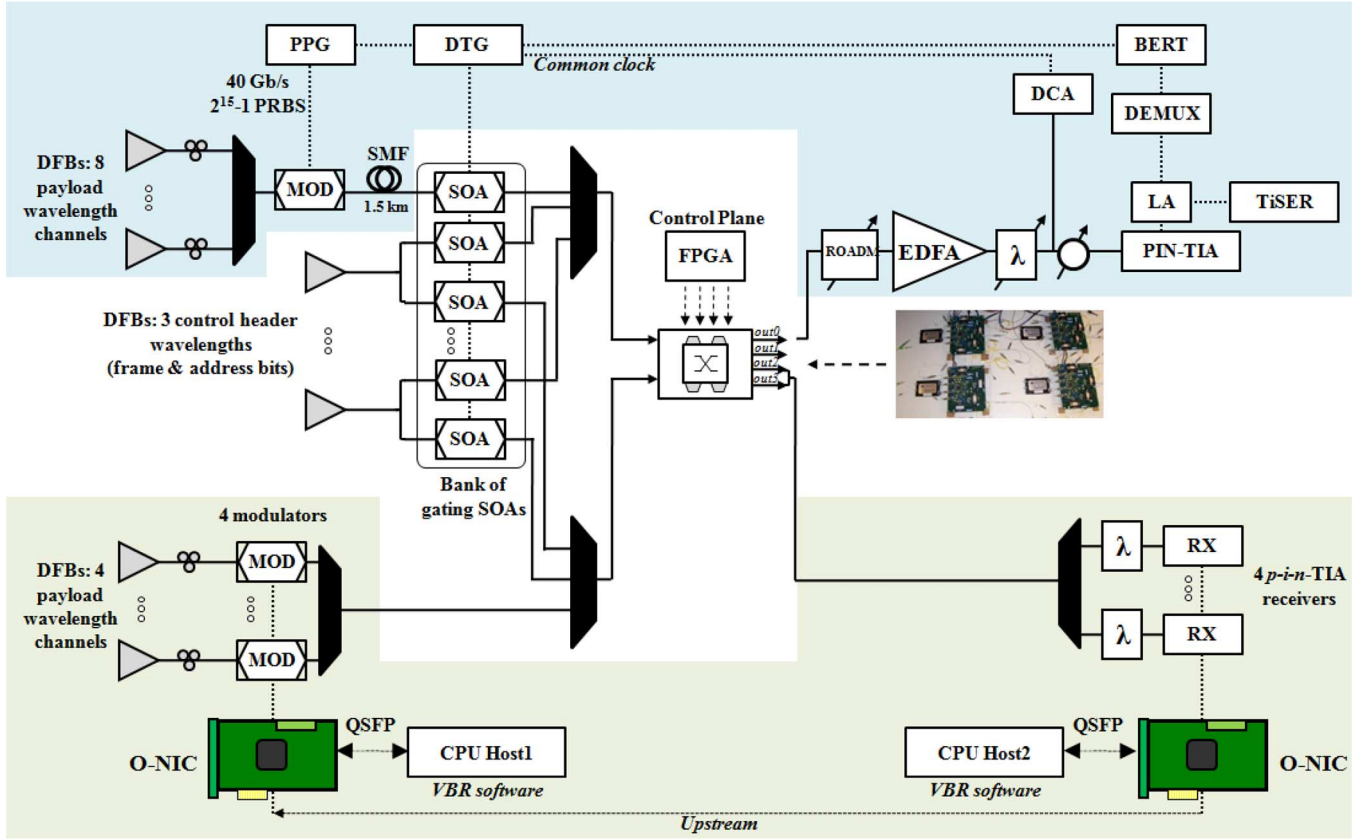


Fig. 9. Detailed experimental setup diagram of the complete demonstration. The blue/upper region denotes the setup associated with the 8×40 -Gb/s packet generation; the green/lower region denotes the setup associated with the 4×3.125 -Gb/s video data.

of circuit-switched HD video data through the fabric without frame loss or observable distortion. This is shown by the green/lower region of Fig. 9. Again, in the face of a higher-layer router failure and/or the detection of optical signal degradation, the FPGA control plane signals the fabric to perform a nanosecond-scale reconfiguration and allows the video data to be transmitted seamlessly upon restoration of the optical link. Further, the cross-layer adaptability of the application layer to the physical layer is demonstrated using variable-bit-rate (VBR) video transmission over the fabric.

The details of these two parts are discussed in Sections IV and V, respectively.

B. Optical Fabric Setup

The experiment shows that the optical fabric can switch optical packets based on the higher-layer failure state denoted by the control plane. The photograph in Fig. 10(a) shows our implemented test-bed environment with the 4×4 optical switching fabric and the FPGA-based control plane.

We use a two-stage, 4×4 fabric design built using four PSEs. Each element uses commercially-available, off-the-shelf components, including four individually-packaged SOAs, passive optical devices and couplers, fixed wavelength filters, low-speed 155-Mb/s *p-i-n* photodetectors, and the required electronic circuitry. The electronic routing decision logic is synthesized in high-speed Xilinx CPLDs. The PSEs are able to decode optical control bits and maintain their routing state

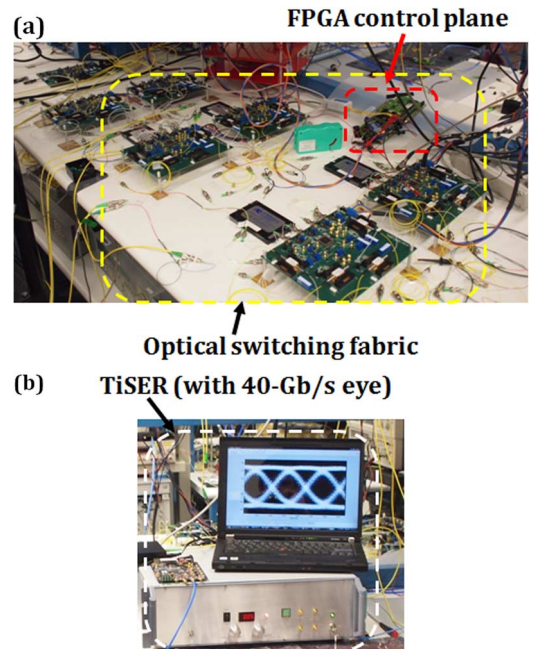


Fig. 10. (a) Photograph showing the implemented 4×4 optical switching fabric and FPGA-based control plane; (b) Photograph showing the implemented TiSER scope chassis that is realized within the CLB.

based on the extracted headers while simultaneously handling wavelength-striped data transparently in the optical domain.

At each switching stage, the wavelength-based routing signals are extracted, with each PSE decoding four control header bits (two per input port) for routing: one frame and one address bit. The CPLD uses the header bits as inputs in a programmed routing truth table, then gates on the appropriate SOAs. At each 2×2 PSE, the extracted frame bit denotes the presence of a wavelength-stripped packet; then, according to the detected address signal, the CPLD gates the suitable SOA for the packet to be routed to the upper (or lower) output port of the PSE (Fig. 3). The combinational logic synthesized in the CPLD uses the two-bit control header as follows: upon the presence of the frame bit (F), the logic then examines the address bit. If the address bit is low, the message is directed to the upper output port; if the address is high, the message is transmitted to the lower output port. The PSE supports the detection of messages ingressing on two input ports.

The SOAs are operated in the linear regime, and their inherent optical gain compensates for the insertion losses of the passive optical components. The SOAs are mounted on a custom-designed electronic circuit board (Fig. 10(a)) with the required electronic components, current driver chips, and low-speed optical receivers.

Lastly, the optical fabric has been previously reported to have advanced photonic switching capabilities, such as packet multicasting (i.e. simultaneous packet transmission from one input to multiple output ports) and optical QoS/priority encoding (i.e. to allow packet switching to be based on QoS constraints) [33]. Though these specific functionalities were not implemented in this experiment, they can be easily realized in future demonstrators. Multicasting can be achieved by using additional PSEs within the fabric to increase path diversity [34]. Also, the QoS wavelength header that can be encoded in the packets' wavelength-stripped format (as described above and in Fig. 5) can be seamlessly extracted by the FPGA control plane using the same low-speed photodetectors as above. The switching algorithms and decisions can then account for the required priority class. For example, the fabric can support a packet-protection switching mechanism that can be triggered both on PM measurements and the QoS class encoded in the optical packet [23].

C. Reconfiguration Scheme

The experiment uses a recovery scheme that allows the optical fabric's 2×2 PSEs to account for higher-layer failures. If a failure or degraded link is detected, the control plane signals the fabric to modify its switching state to route around the failure and avoid additional degraded packets. The fabric can be rapidly reconfigured and its resources reprovisioned while operating under the two different traffic loads.

Two explicit cases are shown: (1) an online router (i.e. when packets are correctly switched to their desired output ports), and (2) an offline router (i.e. the router or following optical link is down and/or impaired, thus packets are rerouted according to predetermined recovery switching logic). As shown in Fig. 6, for an offline router, packets that would have been transmitted to the router are instead rerouted to another output port if there is no contention; otherwise, they are dropped. The recovery scheme deflects packets to an alternate port of the same PSE to mitigate failure on a given link. This deflection routing can

be dependent on QoS requirements, so that the control plane ensures high-QoS packets are given priority over low-QoS messages in the case of contention within a PSE; in this way, the deflection scheme can match packets' encoded QoS.

The FPGA control plane is implemented using an Altera Stratix II (Fig. 10(a)). The FPGA can accept external inputs (e.g. electronic signals from a router and/or PM modules), and then generates failure signals for the PSEs. The routing logic synthesized in the CPLDs is adapted to accept these electronic failure signals to either route normally (for an online router) or route around the failure (for an offline/failed router, to ensure no messages are transmitted on a degraded link). The Altera FPGA circuit board contains eight flip switches and 28 general purpose input/output (GPIO) pins. The FPGA is programmed to receive input from the flip switches, indicating the presence of a router failure, and to signal the appropriate PSEs using the GPIO pins. In this way, the FPGA performs a packet-scale lightpath computation in order to ensure successful optical message transmission.

IV. FABRIC RECONFIGURATION DEMONSTRATION WITH TiSER

A. Setup

Packet-rate monitoring and fast reconfiguration is shown with the fabric operating under a multi-terabit PRBS load. As mentioned above, the OPS fabric supports 8×40 -Gb/s wavelength-stripped optical packets, which switched by the fabric depending on the router failure state as signaled by the control plane. TiSER is used as a PM module to monitor the link and indicate whether the fabric has successfully reconfigured its switching state.

The multiwavelength packets' payload is composed of data encoded on eight separate payload wavelength channels, each modulated at 40 Gb/s. The 8×40 -Gb/s optical packets thus have a total aggregate bandwidth of 320 Gb/s (per fabric input port), indicating the fabric's multi-terabit capacity.

The blue/upper region in Fig. 9 depicts the setup for the 8×40 -Gb/s packet generation and signal integrity analysis. The payload channels are generated using eight separate continuous-wave (CW) distributed feedback (DFB) lasers each connected to a polarization controller (PC). The payload wavelengths range from 1533.12 nm (ITU C58) to 1560.61 nm (ITU C21), with a minimum frequency spacing between two adjacent payload channels of 100 GHz. The outputs of all eight lasers are passively combined onto a fiber using an optical coupler and then modulated simultaneously with a high-speed radio frequency (RF) signal, namely a 40-Gb/s NRZ-OOK signal that carries a $2^{15} - 1$ PRBS. A single commercial 40-Gb/s LiNbO₃ amplitude modulator is utilized, which is driven by the 40-Gb/s RF signal that is generated using a high-speed pulse pattern generator (PPG). The multiwavelength channels are then passed through a 1.5-km span of SMF-28 fiber to decorrelate the data and subsequently to an external SOA for packet gating. The packets are 32- μ s in duration, allowing for TiSER to acquire a sufficient number of samples (1500 sample points) to capture the eye diagram of a single packet. Although the packet lengths used here are relatively long, future integrated TiSER implementations may require fewer data samples, thus allowing shorter packets to be supported.

The control header signals are created independently using three separate CW-DFB laser sources at the suitable wavelengths for the frame (1555.75 nm (C27)), and two switching fabric address bits for the two-stage topology (1531.12 nm (C58), and 1543.73 nm (C42)). Each of the control DFB lasers are connected to a separate packet gating SOA. The control and multiwavelength payload channels are then gated into the 32- μ s long packets using a data timing generator (DTG) and the bank of gating SOAs. The DTG acts as a programmable electronic pattern generator and is synchronized with the 40-Gb/s clock. The address bits are encoded appropriately high or low for each packet to ensure correct switching through the fabric. The channels are then multiplexed together using a passive combiner, yielding wavelength-striped optical packets including three control bits and eight 40-Gb/s data streams. A similar packet-generation setup may be used concurrently for each set of control and payload signals to form a distinct packet pattern for the each of the fabric's input ports.

The wavelength-striped optical messages are switched within the fabric and correct path routing is verified. At the output of the realized switching fabric, the multiwavelength packet is monitored and examined using an optical spectrum analyzer (OSA) and high-speed sampling oscilloscope (i.e. a digital communications analyzer (DCA)). The packet analysis system also allows the wavelength-striped packet to propagate to a tunable grating filter (here, a narrow-band reconfigurable optical add-drop multiplexer (ROADM in Fig. 9), selecting one 40-Gb/s payload stream for signal integrity analysis and rejecting the SOAs' accumulated amplified spontaneous emission (ASE). The payload channel is then sent to an erbium-doped fiber amplifier (EDFA), another tunable filter, and a variable optical attenuator (VOA). The signal is then received by a 40-Gb/s *p-i-n* photodiode followed by a transimpedance amplifier (TIA). A limiting amplifier (LA) is also used with a differential output.

The first port of the LA is connected to an electrical demultiplexer, which time-demultiplexes the signal such that the BER can be evaluated using a commercial 10-Gb/s BERT. The DTG is used to gate the BERT to measure the errors over the duration of the packet. No clock recovery is performed in this system, and a common clock synchronizes the DTG, pattern generator, BERT, and electrical demultiplexer.

The LA's second differential output is connected to TiSER, which allows the capture of 40-Gb/s eye diagrams. In order to avoid a dispersion penalty, this implementation uses less dispersive fiber than previously [24]. Chromatic dispersion causes interference between the sidebands of the 40-Gb/s signal, giving rise to a slight low-pass filtering effect and consequently a dispersion-related penalty.

TiSER monitors one 40-Gb/s channel at the fabric's output at a time; by adjusting the tunable filters at the output, TiSER can then evaluate the eight 40-Gb/s channels. Fig. 10 shows a photograph of the TiSER system. The data is sampled using a commercial A/D digitizer with 2-GHz bandwidth, capturing up to 20 GSamples/s using a real-time scope.

B. Results

The experimental demonstration shows correct functionality of the switching fabric, with correct addressing and switching; wavelength-striped optical packets with 8×10 -Gb/s payloads

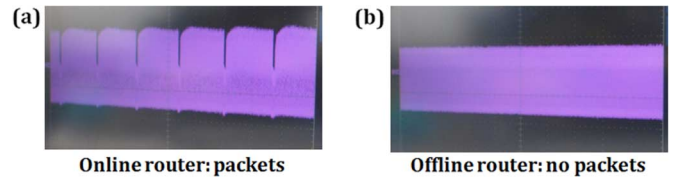


Fig. 11. Photographs showing two cases: (a) an online router: the optical packet flow egresses at its desired port (out0); (b) an offline router: the lack of packet flow at the original output port of the fabric indicates that the packets have been rerouted to an alternate port (out1) to avoid the point of failure.

are correctly routed through the fabric. Further, TiSER allows the QoT of an egressing optical packet to be evaluated offline using advanced signal processing techniques. At the output of the CLB's fabric, the QoT of a high-bandwidth optical packet is determined by assessing one of the 40-Gb/s optical payload channels. TiSER obtains a sufficient number of samples to generate a 40-Gb/s eye diagram from a single optical packet. Using the sampled eye, the BER is then estimated by a calibrated signal processing algorithm that rapidly determines the resulting Q factor.

Here, the TiSER scope is used to monitor the egression of optical packets from the CLB's switching fabric and allows the observation of the fabric's fast reconfiguration. As outlined above, the control plane informs the fabric of a failure so the optical packet stream can be rerouted accordingly to preserve the end-to-end lightpath. The monitoring and fabric recovery capability uses the 40-Gb/s channels, and the signal from the higher-layer router to the control plane is by means of a manual adjustment of a flip switch on the FPGA circuit board. Thus, here, we use offline signal processing to estimate the BER. In the future, a circuit board with on-board FPGA and low-speed A/D can also be used to enable the real-time, online BER extrapolation. This real-time estimation of the packets' QoT has potential to be more rapid than a traditional BERT. This packet-scale BER estimation can then be leveraged in the cross-layer infrastructure to denote the optical signal quality on a packet rate.

TiSER is connected one of the output ports of the switching fabric (specifically out0). Fig. 11(a) depicts the reconfiguration experiment state of an online router, as shown by the A/D. Using the low-speed digitizer realized with TiSER, the optical packet stream is seen to be transmitted to the desired router link (out0). Correspondingly, Fig. 11(b) depicts the reconfiguration experiment state of an offline router. In this case, the TiSER digitizer's output thus displays no packets, since they are rerouted to alternate port (out1) within the switching fabric to avoid the packet loss of transmitting to a failed/degraded link.

The eight payload wavelength channels were evaluated consecutively. In Fig. 12, we show representative 40-Gb/s eye diagrams of a single optical packet (at $\lambda = 1538.98$ nm) as captured by TiSER during the fabric reconfiguration experiment. Fig. 12(a) depicts the 40-Gb/s TiSER-measured eye diagram at the fabric port corresponding to the router (out0) in the case that the router is online, while Fig. 12(b) depicts the 40-Gb/s TiSER-captured eye diagram at the rerouted fabric port (here, out1). When the router is offline or the following link is shown to be degraded, the cross-layer platform signals the optical packets to be redirected to an available output in the switching fabric (e.g. out1). As expected, we observe minimal

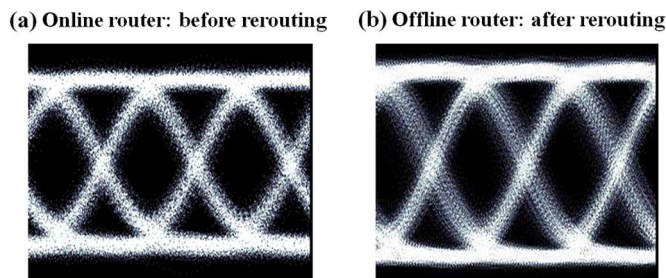


Fig. 12. The TiSER-captured 40-Gb/s optical eye diagrams for the cases of an (a) online router and (b) offline router ($\lambda = 1538.98$ nm).

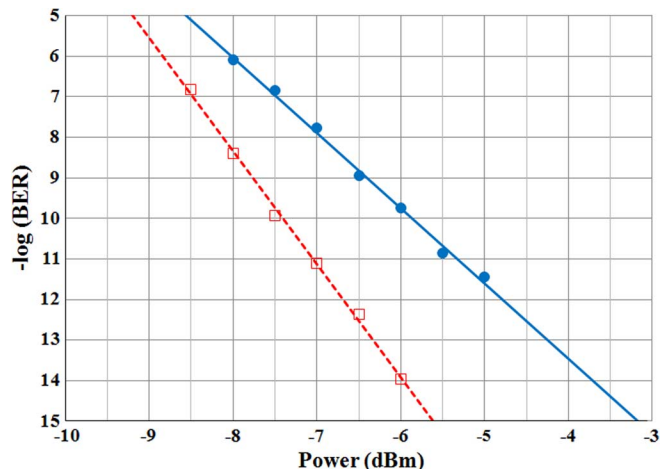


Fig. 13. TiSER-captured 40-Gb/s sensitivity curves for one representative payload channel in the online router scenario (the red/dashed line refers to the back-to-back measurements at the fabric input; the blue/solid line refers to measurements at the router output port, out0 ($\lambda = 1538.98$ nm)).

performance degradation due to switching and rerouting, as indicated by the eye diagrams in Fig. 12. The BER estimation algorithms also show that the rerouted packets exhibit better BER performance in the offline router case, compared to the online router scenario.

Using the packet analysis system outlined above, our BER measurements with a commercial BERT show that all packets are switched through the fabric with error-free performance, attaining BERs less than 10^{-12} on all eight payload wavelength channels.

To demonstrate TiSER's packet-level BER extrapolation algorithms, we then estimate the signal's BER using TiSER alone in lieu of using the conventional BERT system. TiSER samples the data at varying optical power levels, then employs offline signal processing techniques to estimate the Q factor; a sufficient sample size is used to account for any statistical variations in the noise distribution on a packet timescale. As indicated by TiSER, we confirm the error-free transmission and plot the resulting TiSER-generated BER data with respect to the received power. As shown in Fig. 13, we obtain 40-Gb/s sensitivity curves resulting from the TiSER-estimated data. A 1.3-dB power penalty is obtained for the complete system.

This first part of the experiment highlights the fast, multi-terabit, hybrid opto/electronic switches that will be required to be reconfigured in the face of failures, and that can be seamlessly and transparently integrated with real-time PM. TiSER acts as

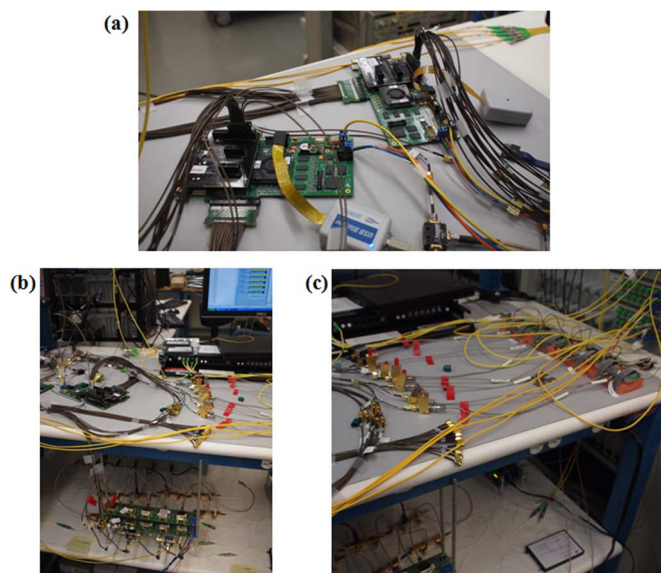


Fig. 14. Photographs of the experimental O-NIC setup, showing the (a) two Altera FPGA development boards, (b) four LiNbO₃ MZ modulators, and (c) four *p-i-n* receivers with TIA and LA pairs.

the embedded PM; its rapid BER extrapolation capabilities allows the monitoring of 40-Gb/s channels and thus the fast measurement of the optical QoT at the packet rate.

V. VIDEO STREAMING DEMONSTRATION

A. Setup

We then demonstrate the CLB's ability to support multimedia/video applications via the transmission of 10GE-based HD video traffic using 4×3.125 -Gb/s streams, which occurs simultaneously to the wavelength-striped PRBS data operation. We use a custom-designed 10GE-based O-NIC to enable Ethernet-based video traffic through the CLB's switching fabric without frame loss or noticeable distortion. In response to router failure and/or optical link impairments, the cross-layer FPGA control plane allows for the switching fabric to reconfigure with a nanosecond timescale. This allows the video data to be recovered and to be transmitted seamlessly upon restoration of the optical network link.

We also show cross-layer interactions between the application and physical layers using a VBR operation of the data switched by the fabric. The VBR demonstration allows the network to exploit PLI information to adapt the application's transmission rate. The time constant associated with this decision process is largely related to the electronic logic/circuitry; however, the performance gain will be quantitatively obtained through the improvement of the overall network's resiliency to impairments. This would also allow the application layer to dynamically allocate the appropriate resources depending on the performance of the optical layer (e.g. the source node can potentially transmit more lower-bit-rate video streams as required).

The green/lower region in Fig. 9 depicts the setup for generating the 4×3.125 -Gb/s wavelength-striped video streams as required by this part of the experiment. Our custom O-NIC uses commercial 10GE network interface cards (NICs) in the



Fig. 15. Screenshots of the video streaming demonstration, showing the two hosts connected over the 10GE optical network link. The source host1 (right) transmits the real-time images from an HD camera of several authors of this work, to the destination host2 (left).

two computer end nodes (host1 and host2), connected by Quad Small Form-factor Pluggable (QSFP) cables. The NICs are extended by high-speed FPGA devices, which are connected using the 10-Gigabit Attachment Unit Interface (XAUI) interface. The XAUI interface allows the complete system to support four separate lanes of 8b/10b-encoded 3.125-Gbaud signals with an effective data rate of 10 Gb/s. Ethernet packets are transmitted via the end hosts to the O-NIC. Then, the logic within the FPGA deserializes and aligns the data, and adds the 8b/10b encoding at the transceiver. The information is then passed to several self-defined modules in the FPGA in order to parse the Ethernet header information and buffer the effective data packets. The XAUI-based Ethernet payload is then converted to the optical domain, exploiting the wavelength parallelism provided by WDM. The O-NIC produces 4×3.125 -Gb/s Ethernet-based video streams end-to-end.

Four CW-DFB lasers at the following optical payload wavelength channels: 1548.51 nm (C36), 1547.72 nm (C37), 1546.92 nm (C38), and 1546.12 nm (C39), are used to create the optical link. As described above and shown in Fig. 9, the Ethernet data is generated by the source host (host1) and corresponding FPGA, which drive four separate LiNbO₃ MZ modulators. The setup uses two 10GE NICs connected to 64-bit computers (CPUs), and the two O-NICs are implemented using development boards with embedded Altera Stratix II GX FPGAs and transceivers configured with the XAUI protocol above.

The multiwavelength data is then combined with the appropriate control headers and injected in the CLB's switching fabric; circuit-switched paths are established for the video streams, connecting one input port (in3) with one output port (out2). At the output of the fabric, each of the four data streams is appropriately filtered and received using four p - i - n receivers with TIA and LA pairs, and transmitted to the transceivers on the destination host's FPGA board. The upstream traffic is looped back electronically. The photographs in Fig. 14 show some of the hardware used to implement the O-NIC.

B. Results

Simultaneous to the PRBS traffic, the O-NIC generates HD video streams and transmits them over the two-stage switching fabric. The video is set to play on the source host CPU, passes through the O-NIC, all-optically transmitted through the fabric, then plays on the monitor connected to the destination host CPU. We observe that the video is transmitted to the destination without distortion or the loss of frames.

Subsequently, the CLB's switching fabric's reconfiguration is again shown for the video data streams: the control plane signals the switching fabric to reroute video data upon the detection of an optical link degradation. During the lightpath rerouting, the video pauses for a short time while the Ethernet link is restored, then is shown to continue playing with minimal disruption.

Finally, in order to demonstrate the cross-layer adaptability of the application layer with the optical physical layer, a VBR transmission is set up over the CLB's switching fabric. The two host computers are connected through the optical fabric using the 10GE O-NIC interface described above, effectively creating a two-host private IP network using the 10GE-based optical network link. The source host (host1) is physically connected to a HD web camera, as well as to the input of the fabric. The images originating from the camera are transmitted on the optical network and are displayed on the monitor physically connected to the destination host (host2). The transmitted video is encoded using software based on modified FFmpeg [35] and streamed over the fabric in the form of User Datagram Protocol (UDP) packets. Fig. 15 shows the real-time streaming-over-optics of the camera images from host1 to host2, depicting several of this work's authors. Video images without observable distortion can be seen at the destination.

Additionally, the video encoding is customized such that the codec parameters can be modified experimentally on-the-fly. The source in the system switches between high bit rates (supporting high-quality video) and degraded bit rates (supporting low-quality video) upon receiving signaling commands embedded in specific UDP packets. The signals are sent from host2 (destination) to host1 (source), allowing the source node to be informed of the link's QoT. In future implementations, this information could be carried using out-of-band signaling to another network interface on the source host.

Fig. 16 shows the screenshots of the high-quality and low-quality video images that result from the VBR demonstration. We note that an application may wish to transmit a high-quality video as a result of the measured link quality; if a more degraded link is measured, the cross-layer interaction allows for the application to dynamically adjust the bit rate of its transmitted video to cater to the link quality.

This second part of the experiment highlights the CLB's ability to support heterogeneous traffic through the transmission of HD video streams. The cross-layer signaling was manual, whereby the control UDP packets are sent upon user command. In a practical networking scenario, PM subsystems can detect

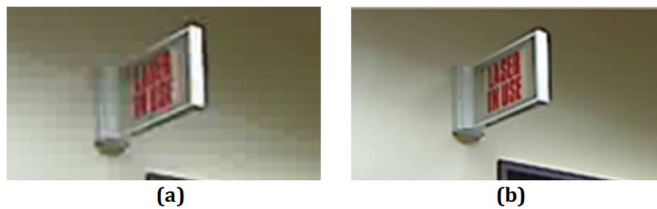


Fig. 16. Screenshots resulting from the variable-bit-rate transmission experiment, showing the support for low-quality video streams (left) and high-quality video data (right).

the QoT degradations and/or increases in BER on a given link, and subsequently signal the QoT to the control plane. The control plane can then instruct the transponders at the sending and/or receiving terminals to reduce the link's bit rate for improved impairment resiliency, and inform the application layer of these changes to allow for the network to cope with reduced resources.

VI. CONCLUSIONS

We have designed an intelligent cross-layer network node that can enable packet-scale reactive switching by exploiting both physical-layer awareness and the knowledge of higher-layer network parameters such as QoS. The CLB uses packet-level monitoring and a distributed control plane to realize cross-layer functionalities, multi-layer traffic engineering, and fast optical switching. Here, we report on the architecture and initial demonstration of a first-generation benchtop prototype of the CLB. The node's implemented subsystems include a high-capacity optical switching fabric, a TiSER performance monitor, and an FPGA control plane. In our test-bed, we demonstrate fast packet-scale reconfiguration of the switching fabric, with error-free transmission of multiwavelength optical packets and of 10GE-based video traffic without noticeable distortion. Cross-layer interactions between the application and physical layers are exemplified by varying the effective bit rate of the video stream depending on link quality.

This work highlights the importance of designing a network node that can optimize optical switching based on parameters from multiple network layers. Some of the key challenges related to timing and packet-scale performance monitoring may be addressed in future practical implementations by enabling a clear path towards integration; this will result in a successful, commercially-viable cross-layer system. An integrated CLB will reduce the lengths of the electronic traces connecting discrete components and replace fibers between optical devices with waveguides, which may affect packet lengths and the timescales upon which the performance monitoring subsystems must operate. As a result, fewer issues with timing and skew may arise with integrated CLB versions. Thus, future work within the CIAN project [11] aims to realize an integrated node with embedded PM modules, achieving small footprints and low energy consumptions, to be deployed in mesh access/aggregation networks. In the current prototype, the switching fabric uses discrete SOAs as switching gates. With the goals of commercialization and meeting stringent cost and energy requirements, future prototypes may feature similar optical elements that will be packaged with other photonic devices and

electronic control plane circuitry. Thus, we will explore the cost-effective packaging of various CLB components using a hybrid integration platform with the required silicon and III-V elements. This will allow the next-generation CLB systems to potentially be more tolerant of thermal and fabrication variations, while reducing power consumption, footprint, and assembly cost. Hybrid integration may also facilitate higher port count fabrics. The technologies currently under development, as well as possible forms of integration, will determine the size and performance specifications of future systems with respect to port count, number of channels, bit rates, etc. These metrics will largely depend on the application space, which is still open for exploration.

Other future studies include: implementing different flavors of cross-layer enabled routing algorithms, exploring different means of contention resolution, investigating advanced photonic switching schemes, and realizing various PM techniques. Here, the first TiSER prototype showcases the CLB's potential to measure the QoT on a message timescale; in future generations of the CLB, TiSER may be replaced by either an integrated device or another packet-scale PM currently being developed by the project. These many avenues of future work will ultimately contribute to a commercial cross-layer system.

The cross-layer node can dynamically co-optimize switching with the ingressing data's QoS and the physical-layer QoT, whilst realizing high bandwidths, low cost, and reduced energy consumptions. The novel cross-layer framework will provide new ways to incorporate packet-level measurements, techniques for monitoring the health of optical channels, and performance prediction in next-generation multi-terabit networks.

ACKNOWLEDGMENT

The authors acknowledge valuable discussions with B. G. Bathula, D. C. Kilper, M. S. Wang, J. W. Wissinger, and G. Zussman.

REFERENCES

- [1] A. Odlyzko, Minnesota Internet Traffic Studies [Online]. Available: <http://www.dtc.umn.edu/mints/>
- [2] R. W. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Labs Tech. J.*, vol. 14, pp. 3–9, Winter, 2010.
- [3] B. Swanson and G. Gilder, "Estimating the Exaflood: The Impact of Video and Rich Media on the Internet a Zettabyte by 2015," Discovery Institute, Jan. 2008 [Online]. Available: <http://www.discovery.org/a/4428>
- [4] 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.
- [5] GreenTouch. [Online]. Available: <http://www.greentouch.org/>
- [6] P. J. Winzer, "Challenges and evolution of optical transport networks," in *Proc. ECOC*, Sep. 2010, Paper We.8.D.1.
- [7] D. C. Kilper, G. Atkinson, S. K. Korotky, S. Goyal, P. Vetter, D. Suvakov, and O. Blume, "Power trends in communication networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 275–284, Mar.–Apr. 2011.
- [8] O. Tamm, C. Hermsmeyer, and A. M. Rush, "Eco-sustainable system and network architectures for future transport networks," *Bell Labs Tech. J.*, vol. 14, no. 4, pp. 311–327, Winter 2010.
- [9] C. P. Lai, C. Ware, B. G. Bathula, D. Brunina, and K. Bergman, "Intelligent highly-functional cross-layer optimized interfaces for future access/aggregation networks," presented at the 13th Int. Conf. Transparent Optical Networks (ICTON), Jun. 2011, Paper We.B3.2.

- [10] C. P. Lai, F. Fidler, and K. Bergman, "Experimental demonstration of QoS-aware cross-layer packet protection switching," in *Proc. ECOC*, Sep. 2009, Paper 2.5.3.
- [11] NSF Engineering Research Center for Integrated Access Networks (CIAN). [Online]. Available: <http://cian-erc.org/>
- [12] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proc. IEEE*, vol. 88, no. 6, pp. 728–749, Jun. 2000.
- [13] R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2002.
- [14] S. Perrin, "The Need for Next-Generation ROADM Networks [White Paper]," Heavy Reading, Sep. 2010.
- [15] C. Ware, C. P. Lai, D. Brunina, W. Zhang, A. S. Garg, B. G. Bathula, and K. Bergman, "Cross-layer reconfigurable optical network: Fast failure recovery in testbed for routing algorithms," *Proc. 13th Int. Conf. Transparent Optical Networks (ICTON)*, Jun. 2011, Paper We.B3.3.
- [16] A. Shacham, B. G. Lee, and K. Bergman, "A wide-band nonblocking 2×2 switching node for a SPINet network," *IEEE Photon. Technol. Lett.*, vol. 17, no. 12, pp. 2742–2744, Dec. 2005.
- [17] 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.
- [18] 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 Annual Meeting IEEE Photonics Soc.*, Nov. 2010, Paper ThQ2.
- [19] W. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, CA, USA: Morgan Kaufmann, 2003.
- [20] 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.
- [21] S. Gupta and B. Jalali, "Time stretch enhanced recording oscilloscope," *Appl. Phys. Lett.*, vol. 94, no. 4, pp. 041 105–041 105-3, Jan. 2009.
- [22] B. W. Buckley, A. Fard, and B. Jalali, "Time-stretch analog-to-digital conversion using phase modulation and broadband balanced coherent detection for improving resolution," in *Proc. OFC/NFOEC*, Mar. 2011, Paper OThW4.
- [23] C. P. Lai, J.-Y. Yang, A. S. Garg, M. S. Wang, M. R. Chitgarha, A. E. Willner, and K. Bergman, "Experimental demonstration of packet-rate 10-Gb/s OOK OSNR monitoring for QoS-aware cross-layer packet protection," *Opt. Exp.*, vol. 19, no. 16, pp. 14 871–14 882, Aug. 2011.
- [24] C. P. Lai, A. Fard, B. Buckley, B. Jalali, and K. Bergman, "Cross-layer signal monitoring in an optical packet-switching test-bed via real-time burst sampling," in *Proc. 23rd Annual Meeting IEEE Photon. Soc.*, Nov. 2010, Paper ME3.
- [25] A. Fard, J.-Y. Yang, B. Buckley, J. Wang, M. Chitgarha, L. Zhang, A. E. Willner, and B. Jalali, "100-Gb/s RZ-DQPSK signal monitoring using time-stretch enhanced recording oscilloscope," in *Proc. QELS*, May 2011, Paper CFP1.
- [26] S. Gupta, B. Jalali, and A. Motafakker-Fard, "Time Stretch Enhanced Recording Scope," U.S. Patent 0 201 345, Aug. 12, 2010.
- [27] C. P. Lai, D. Brunina, C. Ware, B. G. Bathula, and K. Bergman, "Demonstration of failure reconfiguration via cross-layer enabled optical switching fabrics," *IEEE Photon. Technol. Lett.*, vol. 23, no. 22, pp. 1679–1681, Nov. 2011.
- [28] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy efficiency in telecom optical networks," *IEEE Commun. Surveys Tutorials*, vol. 12, no. 4, pp. 441–458, Fourth Quarter, 2010.
- [29] G. Shen and R. S. Tucker, "Energy-minimized design for IP over WDM networks," *J. Opt. Commun. Netw.*, vol. 1, no. 1, pp. 176–186, Jun. 2009.
- [30] P. Winzer, G. Raybon, H. Song, A. Adamiecki, S. Corteselli, A. Gnauck, D. Fishman, C. Doerr, S. Chandrasekhar, L. Buhl, T. Xia, G. Wellbrock, W. Lee, B. Basch, T. Kawanishi, K. Higuma, and Y. Painchaud, "100-Gb/s DQPSK transmission: From laboratory experiments to field trials," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3388–3402, Oct. 2008.
- [31] H. Wang, A. S. Garg, K. Bergman, and M. Glick, "Design and demonstration of an all-optical hybrid packet and circuit switched network platform for next generation data centers," in *Proc. OFC/NFOEC*, Mar. 2010, Paper OTuP3.
- [32] W. Zhang, A. S. Garg, H. Wang, C. P. Lai, J. Wu, J. Lin, and K. Bergman, "Experimental demonstration of 10 Gigabit Ethernet-based optical interconnection network interface for large-scale computing systems," in *Proc. IPC*, Oct. 2011, pp. 443–444, Paper WG3.
- [33] C. P. Lai and K. Bergman, "Implementing an optical QoS encoding scheme in an optical packet switching fabric test-bed," *IEEE Photon. Technol. Lett.*, vol. 22, no. 20, pp. 1518–1520, Oct. 2010.
- [34] C. P. Lai and K. Bergman, "Broadband multicasting for wavelength-striped optical packets," *J. Lightw. Technol.*, vol. 30, no. 11, pp. 1706–1718, Jun. 2012.
- [35] FFmpeg [Online]. Available: <http://www.ffmpeg.org/>

Author biographies not included by author request due to space constraints.