Optical Burst Transport: A Technology for the WDM Metro Ring Networks

Jaedon Kim, Jinwoo Cho, Saurav Das, David Gutierrez, Mayank Jain, Ching-Fong Su, *Member, IEEE*, Richard Rabbat, *Member, IEEE*, Takeo Hamada, *Member, IEEE*, and Leonid G. Kazovsky, *Fellow, IEEE*

***Abstract*—We propose a sublambda traffic-grooming scheme on wavelength-division-multiplexing ring networks, named optical burst transport. The network protocol and architecture are designed to support dynamic bandwidth allocation, which is more reasonable for bursty data traffic. To verify our network protocol and architecture, we build a testbed which supports burst-mode transmission. Also, we transmit streaming video over Ethernet as an application.**

***Index Terms*—Metropolitan area network (MAN), optical burst transport (OBT), traffic grooming.**

I. Introduction

METROPOLITAN area networks (MANs) interconnecting high-speed backbone and low-speed access networks mostly rely on synchronous optical network (SONET)/synchronous digital hierarchy (SDH). Although SONET/SDH combined with wavelength-division-multiplexing (WDM) technology has increased transmission capacity, the efficiency issue regarding data transmission over a circuit-switched network is still present. In addition, recently, access networks have started to provide a large amount of bandwidth by employing advanced LAN technologies, such as digital subscriber line (xDSL), cable modems, passive optical network, and wireless access. Consequently, metro networks are expected to suffer a lack of bandwidth to deal with access network traffic in the near future. Although optical network solutions would solve such a metro gap problem, it is not simple for service providers to deploy a new optical MAN solution over a legacy MAN. While high bandwidth and reliable service are in users' interests, the reduction of OPEX and CAPEX is an important consideration to the service provider. Moreover, noting the large bandwidth change over short periods of the current data traffic, the effective bandwidth provisioning has become a more and more important issue.

Over the past few years, there have been tremendous efforts to improve legacy networks regarding the above issues. The generic framing procedure [1] and link capacity adjustment scheme (LCAS) [2] have been developed to enhance SONET with data-friendly features. Also, virtual concatenation (VCAT) [3] can increase the network utilization by setting the connection bandwidth at fine granularities. Next-generation SONET will allow dynamic allocation of link capacity for optimization of the overall throughput.

Another approach has been developed to adapt Ethernet to MAN, which is supported by the IEEE Resilient Packet Ring (RPR) Working Group (IEEE 802.17). Compared to the legacy SONET, RPR increases bandwidth efficiency introducing a spatial reuse and a differentiated bandwidth provisioning based on class of frame.

However, current point to point connected networks require every node to process large amount of transit traffic which increases complexity, cost, size, and power requirements. In addition, to support the ever-growing traffic volume in the metro area, networks inherently supporting WDM will become necessary in the near future. In such case, network equipment upgrade costs will increase stepwise in current point to point connected network, because all nodes should have the same number of transceivers. Therefore, a protocol supporting more flexible resource management is also necessary.

Optical packet switching (OPS) [4]–[6] was developed to address such problems. With no optical-electronic conversion and high-speed electronics, these switches could theoretically accommodate a very large amount of traffic and switch data at a very small granularity. The motivation for optical burst switching (OBS) [7], [8] is to alleviate some of the optical problems of OPS as well as to do less optical processing. The objective of OBS is to assemble large bursts of data and switch them optically by looking at the burst header/label. That label can be sent ahead of the burst in order to allow enough time for the switching fabric to reconfigure. It can also be sent out of band on a different control channel and processed electronically. Other solutions have been proposed where IP routers or layer-2 switches are used to groom traffic in optical networks while optical links (e.g., point-to-point WDM transmission systems) interconnect high-speed routers.

However, replacing transport network equipment (e.g., SONET add drop multiplexer) with IP routers is not a very scalable solution because each individual IP packet needs to be processed. On the other hand, simply collapsing the layer between IP routers and WDM networks, as advocated by OPS and OBS technologies, is not necessarily practical today because of immature optical components such as optical buffers and large port-count nanoswitches. Nevertheless, we do need a new optical transport architecture that is more adaptive and flexible

Manuscript received June 30, 2006; revised October 19, 2006.

J. Kim, J. Cho, S. Das, D. Gutierrez, M. Jain, and L. G. Kazovsky are with Stanford University, Stanford, CA 94305 USA (e-mail: jdon@stanford.edu; ledcho@stanford.edu; sd2@stanford.edu; degm@stanford.edu; mayjain@stanford.edu; kazovsky@stanford.edu).

C.-F. Su and T. Hamada are with Fujitsu Laboratories of America, Sunnyvale, CA 94085 USA (e-mail: ChingFong.Su@us.fujitsu.com; richard@us.fujitsu.com; takeo.hamada@us.fujitsu.com).

R. Rabbat is with Google, Inc., Mountain View, CA 94043 USA

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.2006.888483

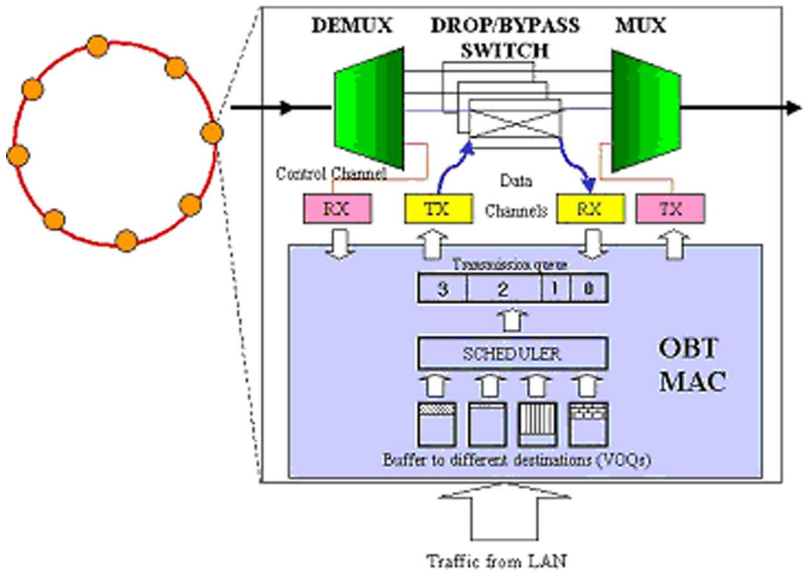


Fig. 1. OBT node architecture.

in order to meet the traffic requirement of current and future Internet applications. Hence, we propose optical burst transport (OBT) [9] to bridge the architectural mismatch between a circuit-based physical transport and the carried bursty packet streams. OBT is based on burst-mode transmission between senders and receivers on a WDM ring topology, and does not require complex electronic processing.

This paper is organized as follows. Section II describes the new OBT architecture and compares OBT with other network architectures. Section III discusses the testbed implementation to verify OBT network. The extension of OBT combined with Ethernet is investigated in Section IV. Section V concludes this paper.

II. OBT Network Architecture and Protocol

A. Network Architecture and Basic Operation

In order to accommodate bursty data traffic at the optical transport layer, we propose a new network architecture that supports burst-mode transmission via fast optical switches on WDM networks. As shown in Fig. 1, the burst-mode transmission is made possible by swift reconfiguration of short-term light paths through the optical switch.

The network topology is based on WDM ring, where one control channel and multiple data channels occupy dedicated wavelengths. Each network node processes incoming control signals on the control channel and reacts accordingly. Control signals include 1) token, 2) control header (CH), and 3) network-management messages. When a token arrives at a node, the node holds it and utilizes the corresponding data channel to transmit data packets only if it has enough data to send. If the amount of data is not enough, the node passes the token to next node. In each node, packets waiting for transmission are stored in several virtual output queues (VOQs) associated with their destination nodes on the ring. When the node has enough data to transmit, the scheduler within the node allocates affordable transmission session for each VOQ based on the length of VOQs. After that, chunks of packets

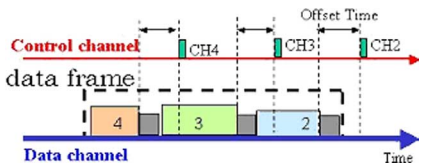


Fig. 2. Traffic on the control channel and data channel at the source node.

are shifted from each VOQ to a transmission queue to form a burst. Therefore, a single burst would have several subbursts for different destinations and all packets in a subburst have the same destination. After building a burst, the node sends a CH to a destination followed by the subburst. Between CH and subburst, it is required to put a constant offset time corresponding to the time to configure the light path and process the control packet. Since control and data channels are separate, we can send a CH to another destination in the middle of data transmission for one destination; this saves bandwidth for data transmission. A timing diagram of CHs and subbursts is illustrated in Fig. 2. Since the burst size is a parameter that can be properly tuned, the OBT protocol allows flexibility between the two extremes of data-oriented or circuit-oriented services for optimal use of network resources.

The traffic load in the network is defined as [10]

$$\rho = \frac{NR}{C\lambda} \tag{1}$$

where N corresponds to the number of nodes on the ring, R the arrival data rate of the traffic to be added onto the OBT ring, C the capacity on each WDM channel, and λ the number of wavelengths. In the case of an ideal burst transport system, the network throughput is proportional to the incoming traffic rate, that is

$$\text{Throughput} = R = \frac{C\lambda}{N}\rho \tag{2}$$

which shows that the network throughput is independent of the burst sizes. Burst size, however, provides an upper bound on

utilization $U_{\max}$. From (2), ρ can be interpreted as the ratio of total incoming traffic to the network to the maximum affordable transmission capacity of the network. When $U \leq U_{\max}$, $\rho \approx U$ in the stabilized network because the blocking probability is far less than 1. Let the token round trip time be D, and the burst transmission time be b. To achieve the maximum utilization, every node always has to be in the ready-to-transmit state. Then, on a single data channel, the token arrives at the specific node every $\mathbf{R} = D + bN$ second, and each node transmits data for b seconds at every $\mathbf{R}$ second. Since the number of node is N, the maximum utilization of the network is

$$U_{\max} = \frac{\text{size of transmitted burst}}{\text{maximum round trip time}} \times N = \frac{bN}{D + bN}. \quad (3)$$

Equation (3) indicates that the maximum utilization becomes 100 percentile as b goes to infinite.

In order to manage multiple WDM data channel, multiple tokens can propagate on the same control channel to carry access grants for respective data channels (wavelengths). Combined with tunable transceivers, multiple tokens can manage the WDM network without stepwise increasing the number of transceivers.

B. Spatial Reuse

Since only one OBT node can access a data channel at a time through a dedicated token, the network does not have collisions on a single data channel. However, as in the other scheduled media-access control (MAC) protocols, the network resource cannot be fully utilized and efficiency is low. For more resource utilization, OBT offers a spatial reuse mechanism. As we discussed in Section II-A, in the OBT network, destination stripping enables a data burst to be terminated at each destination. Scrutinizing the signal channel unidirectional path on the OBT network, when the token is held by a source node for transmission to the destination node, the data path from the source to the destination is occupied, while the data path from the destination to the source is empty and could be used for another transmission. During the time when a destination node is receiving the data, it is also granted an opportunity to transmit its own data. Upon receiving a CH destined to it, the node knows the time duration for transmitting as well as receiving its data at the same time. Therefore, destination nodes are allowed to initiate a secondary transmission without a token to make full use of the available capacity. The size of the secondary transmission is less than the reception time of the subburst because the CH process time would be included in the secondary transmission time slot. Since the secondary transmission uses the timeslot informed by the CH, the size of secondary transmission could have different value every time. Therefore, burst aggregation for the secondary traffic does not follow the burst length-based algorithm but has a maximum allowed time slot. Note that during the secondary transmission, the destination of the secondary transmission may also create another transmission for better transmission performance.

Fig. 3 shows the benefit of spatial reuse. We use two simulation models with or without spatial reuse scheme. In each case, a WDM ring network consisting of five nodes is analyzed, with a circumference of 200 km. Each node has the same fixed burst size of 200 kB. The data rate is 1.25 Gb/s per wavelength on data channel and 625 Mb/s on the control channel. For simplicity, we use a single data channel, but, from (2), it is easily expected that network throughput will linearly increase with the number of WDM data channel. For each traffic load, we run the simulation ten times and take the average value.

As shown in Fig. 3(a), the throughput for the OBT network without spatial reuse is close to the theoretical maximum value, because a deviation for all simulation results are not significant. From the simulation condition

$$N = 5$$
$$b = 200 \times 8 \times 10^3/(1.25 \times 10^9) = 1.28 \times 10^{-3}$$
$$D = 200 \times 10^3/(2 \times 10^8) = 1 \times 10^{-3}.$$

Therefore, from (3)

$$U_{\max} = 0.864$$

and

$$\text{Maximum throughput} = R(\rho = U_{\max}) = 216.2 \text{ Mb/s}.$$

There is a small deviation due to the finite token propagation time between adjacent nodes. On the other hand, the OBT with spatial reuse shows considerable improvement by increasing channel capacity from C to $2C$. In the analysis, the delay measurement consists of the transmission delay, propagation delay and queuing delay of every packet. In Fig. 3(b), the delay remains low when the traffic load is lower than $U_{\max}$ and exhibits a sharp increase when the traffic load close to $U_{\max}$. Another observation about latency is that latency has become large at very low traffic load, and decreases as the traffic load increases. This is because we set the maximum length to assemble a burst. The choice of burst assembly algorithm for OBT, however, is flexible. To compensate the large latency at low traffic load, we can also apply hybrid time and length-based algorithm [12] or dynamic burst assembly algorithm [13]. Due to the effectively enlarged channel capacity, the average delay performance at the node significantly improves. From (1), we can also expect this performance improvement. Since the spatial reuse makes channel capacity from C to $2C$, effectively, ρ axis of Fig. 3 expands twofolds. Therefore, the latency with the spatial reuse can keep a low value even at $\rho = 1$ because, ρ is equivalent to 0.5 without the spatial reuse.

C. Network Resilience

The network topology of OBT is a unidirectional ring. To ensure that the time required enabling protection in the ring is minimized, while keeping the unidirectional approach to make possible a low cost and simple implementation at the node, we chose 1 + 1 optical layer protection. Since the protection scheme can be categorized as optical unidirectional line switched ring (OULSR), it shares the characteristics of OULSR [14]. On top of the optical layer protection, we can add more procedures to increase network resilience. In typical

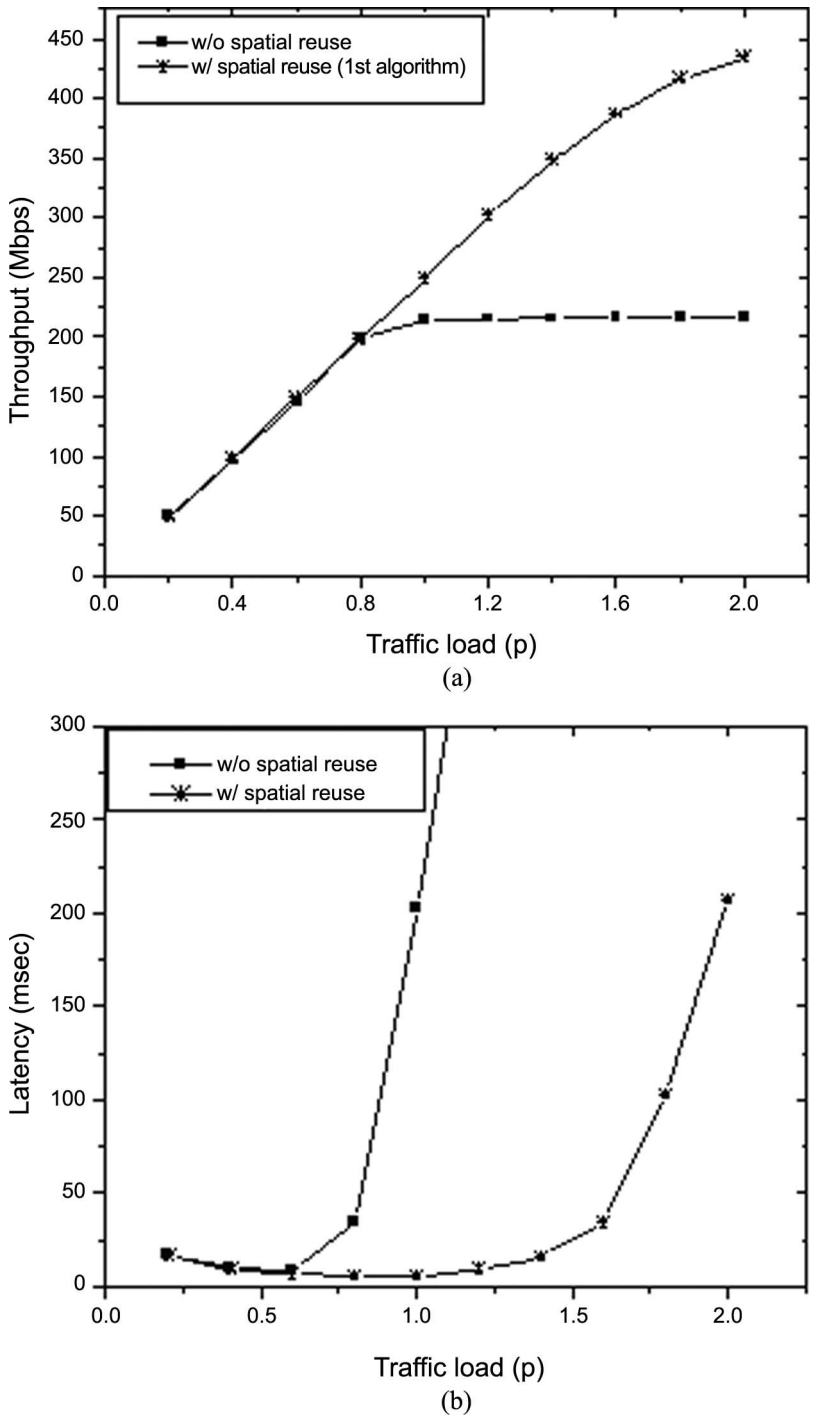


Fig. 3. Comparison of network performance with or without spatial reuse. (a) Throughput. (b) Delay.

OULSR protection, we could not prevent data loss during link failure. In OBT, however, we can reduce the loss by adding additional procedure to the protection.

In OBT with optical link protection, burst will be lost until the token passes through the failed link because not all nodes know the link failure. After the token loss, no node can transmit its burst until the link is recovered and the token is regenerated. The token is managed by a designated master node, which take a charge of the following.

1) Generate token upon network start.
2) Regenerate token when the token is lost.

Therefore, token will be regenerated by a master node every maximum token round trip time, which can cause data loss during the link failure. To reduce data loss as well as increase bandwidth efficiency, we have developed the following procedure.

1) If the node detects the loss of control channel signal, it becomes the node of detection (NOD).
2) The NOD assumes the role of the master node and starts token timeout timers for all data channel tokens.
3) The NOD sends out the fault message on the control channel and initiates the switch of its incoming path to the protection fiber.

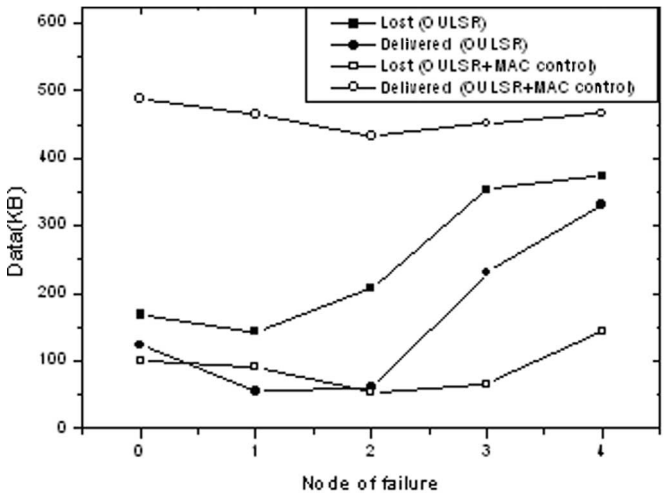


Fig. 4. Delivered and lost data frame with and without MAC control.

4) The fault message informs every node of the existence of the fault as the control channel is processed hop by hop down to the node of failure (NOF).
5) As nodes become "fault aware," they avoid sending any secondary transmission until they receive FaultOK message.
6) When the old master node becomes "fault aware" it immediately stops all its token timeout timers and ceases to be the master node.
7) When the old master node is the NOD, in which case it continues to be the master and does not stop its token timeout timers.
8) Upon timeout of the tokens at the NOD, the NOD regenerates intelligent tokens that require nodes to exclude all nodes beyond the NOF from their destination set.
9) Ultimately when the NOF sends out the FaultOK message, again every node becomes aware that the fault has been bypassed and they resume sending primary and secondary transmissions to all destinations.
10) Finally, the NOD remains as the master node.

The same simulation model as in Section II-B with two different protection mechanisms has been used to observe the effect of link failure. For each simulation, we set master node as node 0, traffic load as 0.8, and run the simulation with different node of failure. With respect to protection time, we consider two main factors—protection switch time and ring latency for the NOF to be informed of the fault. The protection switching time is conservatively assumed to be 10 ms and the ring latency for the above defined ring circumference is 1 ms. Thus, the total protection time is a little over 12 m. For a single node failure, we run ten times and observe lost and delivered bursts for the entire network and take the average value. As can be seen in Fig. 4, with only optical layer protection, the amount of delivered data and lost data becomes larger, as a broken link is further from the master node. This is because token can exist longer in the network, which gives nodes a higher chance to transmit its own burst.

The result becomes different when we add our protection algorithm on the optical layer protection. First, data loss is suppressed for every different location of failure. At the same time, data delivery is also improved for all the location of failure. After adding additional protection algorithm, we can remove fault location dependence because master node will shift to NOD whenever the link is broken.

TABLE I
COMPARISON OF NEXT-GENERATION SONET, RPR, AND OBT

	Type	*Average hop distance	Bandwidth Provisioning Time	Bandwidth granularity
Next-Gen SONET	LMPR	1	A few seconds to a few minutes,	SONET hierarchy
RPR	LRPR	$\frac{N}{2}$	Sub-milliseconds,	SONET, Ethernet frame
OBT	LMPR	1	Sub-milliseconds,	Flexible

*Average physical hop distance over logical topology.

D. Comparison With Other Network Architecture

It is instructive to compare OBT with the current technologies, next-generation SONET with VCAT and LCAS support, and RPR, as listed in Table I. All these designs are based on the ring topology, and are suitable for transmission of data-dominant traffic. However, the logical connection at the electrical layer shows a different connection topology. Logical mesh over physical ring (LMPR) is a logical mesh topology realized over a physical ring.

We consider one-way average hop distance in order to compare both topologies. In physical-ring network, when average hop distance increase, every transmitted data will get through more intermediate nodes. That is, each node will receive more transit traffic, which is not destined to it. As a result, a node will waste more processing time and power on load destined to other nodes, when average hop distance increases. Assuming that there are N nodes on the ring, any one node on the ring has $N-1$ adjacent nodes, and any node on the ring can reach any other node in a single hop. Logical ring over physical ring is another extreme, whose adjacencies are only limited to immediate neighbors. The average hop distance between nodes on the ring is $N/2$. Next-generation SONET and OBT are both LMPR. As a result, next-generation SONET and OBT do not need to backlog transmit traffic because of dealing with transit traffic. Next-generation SONET, however, still relies on a conventional management system and signaling protocols such as generalized multiprotocol label switching to realize dynamic bandwidth provisioning, which typically takes a few seconds up to a few minutes. OBT, on the other hand, uses tokens for controlling optical-burst, which enables submillisecond provisioning time. All these schemes support traffic grooming. Next-generation SONET uses SONET hierarchy so that the granularity of bandwidth is a multiple of virtual tributary groups from VT1.5 to STS-3c. RPR is designed as to handle Ethernet frame as well as SONET frames. OBT is also able to aggregate Ethernet packets, which will be shown in Section IV. Moreover, the OBT protocol is compatible with tunable transceivers for WDM systems, while next-generation SONET and RPR need additional transceivers for underlying WDM channels.

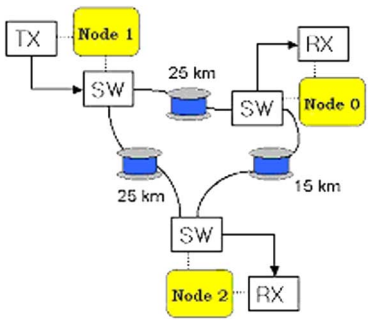


Fig. 5. Testbed configuration.

III. Testbed Implementation

A prototype experimental testbed is constructed for investigation as illustrated in Fig. 5. The testbed consists of three nodes and uses two International Telecommunication Union Telecommunication Standardization Sector (ITU-T) dense WDM wavelengths occupied by the control and data channel. Line rates are 2.5 and 1.25 Gb/s for data and control channel each. An Field Programmable Gate Array (FPGA) (Vertex II-pro, Xilinx) and its evaluation boards serve as processors in each node to implement the MAC protocol and necessary data processing. Each node needs the transceivers to support burst-mode transmission. For the transmitters, we have some alternative choice. Besides the optical switch and laser diode, which are applied to our testbed, we can also think of other combinations such as tunable laser and arrayed waveguide gratings or burst-mode laser driver and laser diode.

On the other hand, the burst-mode receiver is the more challenging of the two aspects. As described in Fig. 6, it is common for high-passing devices to have a low-frequency cancellation part which consists of a differential amplifier and low pass filter feed back loop. Such a high-passing device can suffer from the baseline wandering problem due to consecutive identical digits (CIDs) [11]. In burst-mode transmission, there exists a lot of CIDs because no signal can be transmitted during idle time. Consequently, it is very likely for the conventional receiver to suffer from baseline wandering problem. Therefore, the time constant of the feedback loop, which can be defined by R, C1, and C2 in Fig. 6, should be lowered to mitigate the problem. Care should be taken to lower the time constant, because, if we overreduce the time constant, the signal power would be reduced by increasing the cutoff frequency of the low pass filter, which seriously degrades the signal to noise ratio.

Another challenge in the design of the burst-mode receiver is burst-mode clock recovery. Since burst-mode clock recovery technology is not so mature to support 2.5-Gb/s system, we use conventional phase-locked loops integrated in the FPGA, which guarantees to recover clock signal within 1 μs. Contrary to packet switching, burst switching has coarse granularity for its time frame due to packet aggregation. In our experiments, we use 100 μs (31.25 kB at 2.5 Gb/s) or 48 μs (15 kB at 2.5 Gb/s) as sizes of burst so that the overhead resulting from the clock recovery time is less than 2%. Considering Ethernet packet sizes distributed between 64 and 1500 B, our burst size is moderate in terms of packet aggregation. On the flip side, noticing the maximum frame size of generic frame protocol is 65 kB, our burst size is not big enough as to incur a significant transmission delay.

Protocol operation is monitored using logic analyzer as in Fig. 7(a). In Fig. 7(a), the source transmits 48-μs data burst to first destination and second destination. While first destination receives data for 32 μs, second destination receives data for 16 μs after a corresponding transmission and propagation delay. Propagation delay from source to first destination is 125 and 200 μs from source to second destination. One of the nodes in the ring network, designated as the master node, is responsible for initializing tokens upon the system power-up. In addition, a timer in this node estimates the token round-trip time plus holding time by other nodes and triggers a regeneration process if it does not receive the token back within that specific period of time. The data add and drop is monitored as in Fig. 7(b). For this experiment, we use a smaller frame size because the guard time is too small to be shown in the scale of 50 μs/div. As we discussed in Section II-A, the offset time consists of the switching time of the optical switch and the CH processing time. It takes five clocks for each node to process a CH, which corresponds to 80 ns. As can be seen in Fig. 7(b), switching time of optical switch is 260 ns. Hence, the overall offset time is 340 ns.

The bit-error-rate (BER) test was done to verify the data link performance. As described in Fig. 8(a), the transmitter is directly modulated by a pattern generator which generates a continuous pseudorandom bit sequence $2^{23}-1$ at 2.5 Gb/s. Whenever a token arrives at node 1, it sends the CH and changes the switch to connect between transmitter and ring. When node 2 receives the CH, it changes the switch to receive data for a specific duration, and then switches it back. During the time node 2 receives data, node 2 also enables a gating signal for the bit error detector so that the error detector counts bit errors. In the Fig. 8(b), the circle indicates transmission without switching, which becomes continuous mode transmission. For the burst-mode transmission, we send data for 100 μs to the first destination and 50 μs to the second destination. BER is measured at the first destination. Compared to continuous mode transmission, burst-mode transmission shows about 4.5-dB power gain at the same BER. However, noting the power meter takes the time average to calculate the received power, despite the fact that burst and continuous modes use the same power at the transmitter, the power meter shows a smaller value for the burst-mode transmission, because the power meter receives nothing during the idle time. In our experiment, the round trip time for a token is 475 μs. The time, during which data is transmitted to the first destination, is 100 μs. Therefore, the power meter would receive the real signal only for 100 μs every 475 μs and during the other 375 μs receive nothing. As a result, the power in the burst-mode transmission is measured to be almost $0.315(=100/475)$ times less than that of the continuous mode transmission, which corresponds to 5.016 dB. After the power compensation, which is indicated by the dotted line, we can see that continuous mode transmission and burst-mode transmission use almost the same peak power. The discrepancy may come from the timing mismatch between the gating signal and real received data. However, the result indicates that we can save transmission energy using the burst-mode transmission.

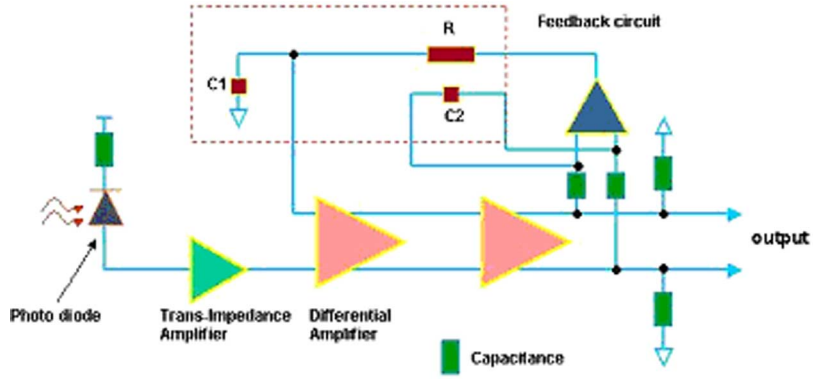


Fig. 6. Optical receiver.

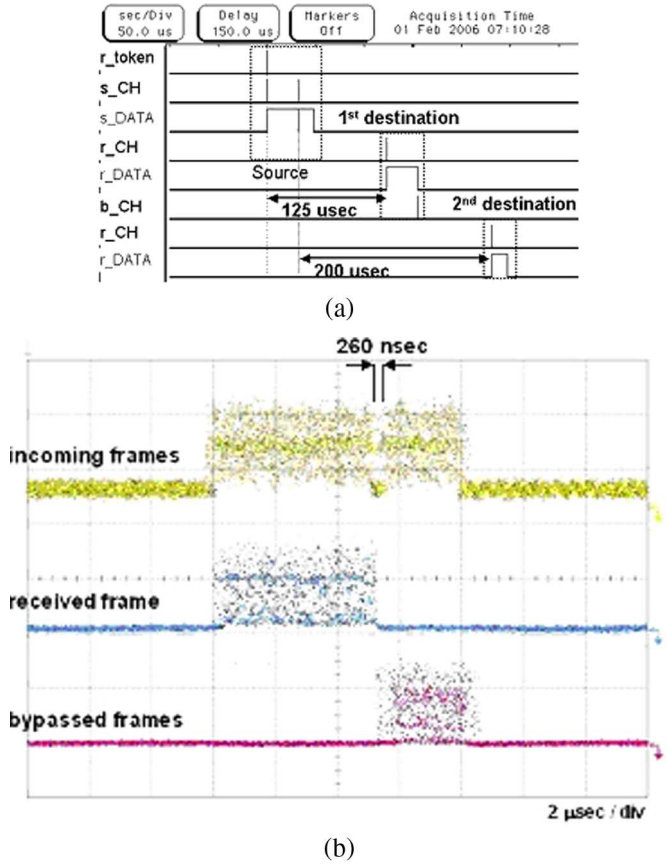


Fig. 7. Testbed configuration. (a) Control signals shown on the logic analyzer. Signal receiving, sending, and bypassing are represented by "r," "s," and "b," respectively. (b) Data add/drop monitored at the first destination.

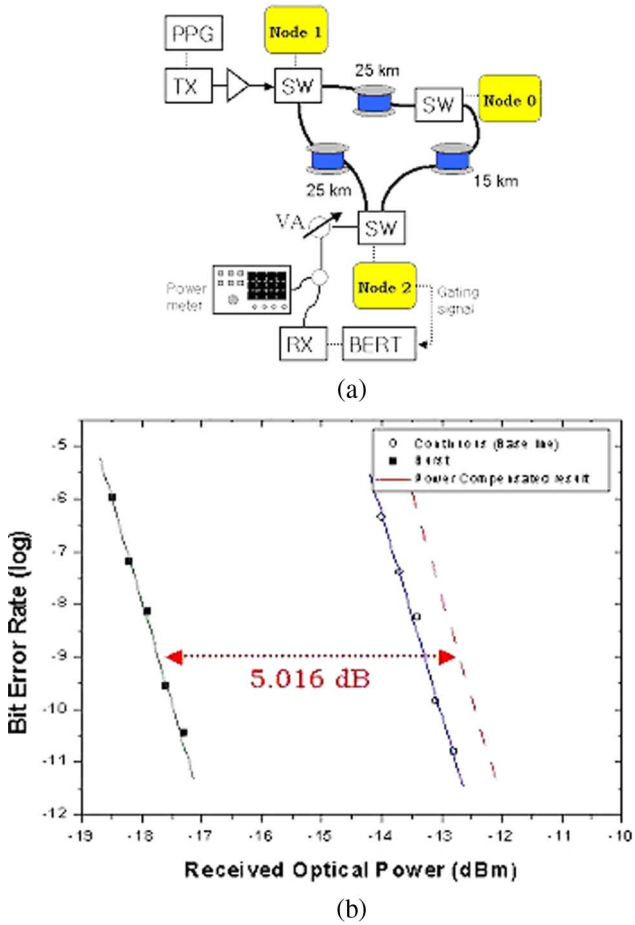


Fig. 8. BER test (a) experiment setup (b) result.

IV. Ethernet With OBT

So far, we assumed incoming traffic from the access network is already in the OBT node. However, it is more reliable to show the capability to deal with incoming packets from the local and access network. Ethernet packets are a proper encapsulation for incoming data traffic since it has become the most popular MAC protocol in local and access area networks. Therefore, it is necessary to consider how to handle Ethernet packets in the OBT network. To investigate the compatibility with Ethernet MAC (EMAC), we attempt to combine Ethernet and OBT MAC protocol. In the system clock point of view, 100-Mb/s EMAC uses 12.5 MHz, while OBT uses 62.5 and 125 MHz to run the control channel process and data channel process. As described in Fig. 9, the OBT control channel process takes a role of combining EMAC and OBT by communicating with the OBT data channel process and interface of EMAC. The communication between control and data process is just to indicate a packet arrival or departure. However, it requires a more sophisticated process for the communication between OBT control channel and interface of EMAC because it needs not only to indicate packet arrival and departure but also to store, process, and encapsulate incoming Ethernet frames.

Ethernet frames can be aggregated at the OBT node as shown in Fig. 10. When an Ethernet frame arrives at the OBT

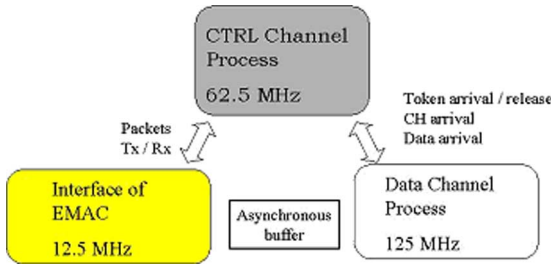


Fig. 9. Simplified very-high-speed-integrated-circuit hardware-description-language structure.

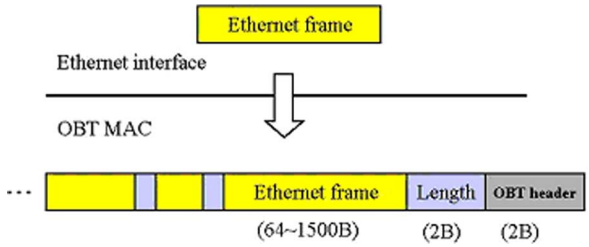


Fig. 10. Frame of EMAC over OBT.

node from the local network, the OBT MAC finds the length information of the incoming packet and puts length bits in front of the incoming frame before the frame is stored to a VOQ. The next new incoming frame should be stored in the same way. When the token reaches the node and that node has enough frames to transmit, it holds the token and generates an OBT frame. The OBT frame has a header which indicates the OBT frame followed by concatenated length bits and Ethernet frame. No extra address section is required in OBT frame because the address is already decided by a separate CH.

After combining EMAC with OBT, we test our network's ability to transmit streaming video files, a test that we successfully demonstrate. For the demonstration setup, we connect two PCs to two dedicated nodes of the same testbed as in Fig. 5 using category five unshielded twisted pair (UTP) cable. We use a Helix DNA Server 9.0 for a streaming server, and Realplayer is used for streaming video player on the client computer. Conceptual packet transport through OBT network is described in Fig. 11. RTSP stands for real-time streaming protocol supported by Helix server at the application layer. When Ethernet packets come from the local area network, the OBT node encapsulates Ethernet packets to build a burst and transmits to a destination node. At the destination node, the OBT header is removed and a pure Ethernet frame would be transmitted to local area network referring to dedicated length bits. Since we do not unpack the Ethernet frames, any other higher layer protocol, i.e., transport layer and application layer can be substituted by transmission-control protocol (TCP) and hypertext transfer protocol or file transfer protocol.

V. Conclusion

We design and implement a new traffic-grooming mechanism for the metro ring network, which supports bandwidth provisioning in short time scales. The OBT MAC also shows

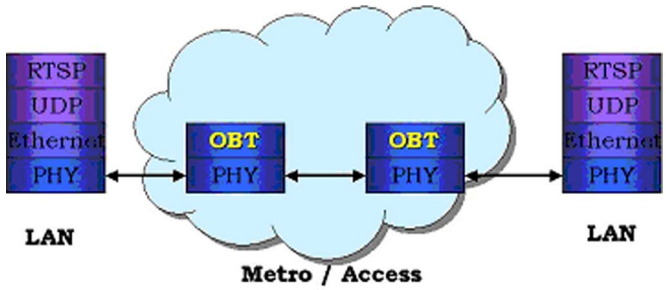


Fig. 11. Ethernet frame transmission through network.

that it can increase bandwidth efficiency with the spatial reuse property and keep resilience using a novel protection algorithm. Moreover, a burst-mode transport physical layer combined with token controlled channel access provides a better way to manage WDM network resource. To verify our reasoning, we build a network testbed providing 2.5-Gb/s data transmission capacity and demonstrate the network's ability to support bursty data traffic through burst-mode BER test. Finally, we test the compatibility of our network architecture with a ubiquitous MAC protocol (Ethernet) by showing real-time streaming video delivery over OBT network. Protocol compatibility with higher layer protocol such as TCP remains for future work. The designed protocol and accompanying physical architecture allow much flexibility to accommodate circuit-oriented and packet-switched data traffic scenarios, making the OBT architecture a promising candidate for future WDM MANs.

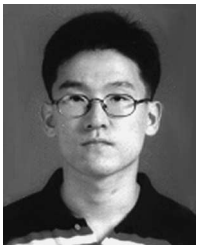
Acknowledgment

The authors would like to thank A. Chugo and Dr. H. Kuwahara of Fujitsu Laboratories, Japan, and Dr. A. Glebov, Dr. M. Lee, Dr. C. Tian, and Dr. T. Naio of Fujitsu Laboratories of America, Inc., for their invaluable advice and discussion. The authors would also like to thank Xilinx University Program and RealNetworks for providing their resources regarding testbed implementation.

References

[1] *Generic Framing Procedure (GFP)*, ITU-T Rec. G. 7041, 2005.
[2] *Link Capacity Adjustment Scheme (LCAS)*, ITU-T Rec. G. 7042, 2001.
[3] *Network Node Interface for the Synchronous Digital Hierarchy.* ITU-T Rec. G. 707, Oct. 2000.
[4] S. Yao, B. Mukherjee, and S. Dixit, "Advances in photonic packet switching: An overview," *IEEE Commun. Mag.*, vol. 38, no. 2, pp. 84–94, Feb. 2000.
[5] C. Guillemot *et al.*, "Transparent optical packet switching: The European ACTS KEOPS project approach," *J. Lightw. Technol.*, vol. 16, no. 12, pp. 2117–2134, Dec. 1998.
[6] D. K. Hunter, D. Cornwell *et al.,* "SLOB: A switch with large optical buffers for packet switching," *J. Lightw. Technol.,* vol. 16, no.10, pp. 1725–1736, Oct. 1998.
[7] D. K. Hunter *et al.*, "WASPNET: A wavelength switched packet network," *IEEE Commun. Mag.*, vol. 37, no. 3, pp. 120–129, Mar. 1999.
[8] C. Qiao and M. Yoo, "Optical burst switching—A new paradigm for an optical internet," *J. High Speed Net.—Special Issue Optical Networks*, vol. 8, no. 1, pp. 69–84, Mar. 1999.
[9] J. Kim, Y.-L. Hsueh, L. G. Kazovsky, C.-F. Su, R. Rabbat, and T. Hamada, "Traffic grooming on WDM rings using optical burst transport," presented at the Optical Fiber Commun. Conf. (OFC), Anaheim, CA, Mar. 2005. Post deadline paper.

[10] Y.-L. Hsueh, J. Kim, C.-F. Su, R. Rabbat, T. Hamada, C. Tian, and L. G. Kazovsky, "Traffic grooming on WDM rings using optical burst transport," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 44–53, Jan. 2006.
[11] *HFAN-09.0.4: NRZ Bandwidth—LF Cutoff and Baseline Wander*. [Online]. Available: http://pdfserv.maxim-ic.com/en/an/4hfan904.pdf
[12] X. Yu, J. Li, Y. Chen, X. Cao, and C. Qiao, "Traffic statistics and performance evaluation in optical burst switched networks," *J. Lightw. Technol.*, vol. 22, no. 12, pp. 2722–2738, Dec. 2004.
[13] X. Cao, J. Li, Y. Chen, and C. Qiao, "Assembling TCP/IP packets in optical burst switched networks," in *Proc. IEEE GLOBECOM*, Nov. 2002, vol. 3, pp. 2808–2812.
[14] O. Gerstel and R. Ramaswami, "Optical layer survivability—A service perspective," *IEEE Commun. Mag.*, vol. 38, no. 3, pp. 104–113, Mar. 2000.



Jaedon Kim received the B.S. degree from Seoul National University, Seoul, Korea, in 2001 and the M.S. degree from Stanford University, Stanford, CA, in 2004, both in electrical engineering.

Since 2003, he has been a Research Assistant at the Photonics and Networking Research Laboratory (PNRL), Stanford University. His current research interest includes fiber optical amplifier, tunable optical transceivers, and optical burst-mode transceivers and MAC protocol for optical metro/access networks.

Jinwoo Cho received the B.S. degree in computer engineering and the M.S. degree in electronic engineering from Kwangwoon University, Seoul, Korea, in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree in electrical engineering at Stanford University, Stanford, CA.

He was with Samsung Electronics from 2000 to 2002 and with Optsys Technology from 2002 to 2005, where he was engaged in research on high-speed optical transmission systems. He is currently a member of the Photonics and Networking Research Laboratory (PNRL), Stanford University. His research focuses on optical burst-mode transceivers, optical burst-mode switching, and high-speed transmission systems.

Saurav Das received the B.E. degree in electronics and communication from the Birla Institute of Technology, Ranchi, India, in 1998 and the M.S. degree in optical sciences from the Optical Sciences Center, University of Arizona, Tucson, in 2000. He is currently working toward the Ph.D. degree in electrical engineering at Stanford University, Stanford, CA.

He was with ANDevices Inc., Fremont, CA, until 2005, where he designed planar waveguide integrated optical components for the optical communications industry. His main research interests are in the areas of architecture and protocols for backbone networks and their implementation in both optical and electronic hardware.

David Gutierrez received the B.Sc. degree in electrical engineering from the Universidad de los Andes, Bogota, Colombia, in 1998 and the M.Sc. degree in electrical engineering from Stanford University, Stanford, CA, in 2002. He is currently working toward the Ph.D. degree at Stanford University in Professor Kazovsky's Photonics and Networking Research Laboratory (PNRL).

His current research focuses on access and metro network architectures, protocols, and algorithms.

Mayank Jain received the B.E. degree in electronics and communications from Delhi University, Delhi, India, in 2002. He is currently working toward the M.S. and Ph.D. degrees in electrical engineering at Stanford University, Stanford, CA.

He has worked as a Design Engineer with Texas Instruments, India, for three years, working on the design of wireless LAN system on chips. His current research focuses on network security for optical access networks.

Ching-Fong Su (S'02–A'02–M'04) received the B.S. degree from the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1991 and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Texas at Austin in 1995 and 1998, respectively.

He is currently a Senior Researcher with Fujitsu Laboratories of America, Sunnyvale, CA, where he has engaged in the research of traffic management, IP/asynchronous-transfer-mode (ATM) network design, and algorithms for high-speed packet processing. Since 2000, he has been working on architecture design and management issues of optical networks. His current research interest includes protection and restoration issues, control planes, and network management of next-generation SONET and wavelength-division-multiplexing (WDM) networks.

Richard Rabbat (S'00–M'01) received the B.S. and M.S. degrees in computer and communications engineering from American University of Beirut, Beirut, Lebanon, in 1994 and 1996, respectively, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1998 and 2001, respectively.

He was previously a Senior Project Manager with Fujitsu Laboratories of America. He is currently an Engineering Project Manager with Google, Inc., Mountain View, CA. He has several patents pending in the area of optical burst switching and multilayer networks. He has authored many papers in a variety of fields and is a frequent contributor to the Internet Engineering Task Force (IETF). His main research interests include generalized-multiprotocol-label-switching (GMPLS) control planes for optical transport networks and advanced optical network architecture.

Takeo Hamada (S'86–M'86) received the B.E. and M.E. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1984 and 1986, respectively, and the Ph.D. degree in computer science, focusing on physical very large scale integration (VLSI) design, from the University of California (UC) at San Diego, La Jolla, in 1992.

He joined Fujitsu Laboratories in 1986. In 1992, he joined the Telecommunication Research Group, Fujitsu Laboratories, Kawasaki, Japan. He spent 1995–1997 at Bellcore, RedBank, NJ, as a part of the Telecommunications Information Networking Architecture (TINA) core team, working on various management architecture of TINA, such as accounting, security, and service management in general. Since 1998, he has been with Fujitsu Laboratories of America, Sunnyvale, CA, where he has been engaged in the study of Internet traffic and its management, and optical networking and its control plane architecture for the multilayer integrated data optical networking.

Leonid G. Kazovsky (M'80–SM'83–F'91) received the M.S. and Ph.D. degrees in electrical engineering from the Electrotechnical Institute of Communications, St. Petersburg, Russia, in 1969 and 1972, respectively.

He was previously with Bellcore (now Telcordia), researching on WDM and high speed and coherent optical-fiber communication systems. While on Bellcore assignments or Stanford sabbaticals, he worked with Heinrich Hertz Institute, Berlin, Germany; Hewlett-Packard Research Laboratories, Bristol, U.K.; Scuola Superiore St. Anna, Pisa, Italy; and Technical University of Eindhoven, Eindhoven, the Netherlands. Through research contracts, consulting engagements, and other arrangements, he worked with many industrial companies and US government agencies, including Sprint, Digital Equipment Corporation (DEC), General Telephone and Electronics (GTE), AT&T, Institutional Venture Partners (IVP), Lucent, Hitachi, Kokusai Denshin Denwa (KDD) Corporation, Furukawa, Fujitsu, Optivision, and Perimeter on the industrial side; and NSF, DARPA, Air Force, Navy, Army, and the Ballistic Missile Defense Organization (BMDO) on the government side. Recent spinoffs of his research idea include such companies as Luminous, Alidian, and Matisse. He joined Stanford University, Stanford, CA, in 1990. He founded the Photonics and Networking Research Laboratory (PNRL), Stanford University, and has led PNRL since then. He has authored or coauthored two books, some 250 journal technical papers, and a similar amount of conference papers.

Prof. Kazovsky is a Fellow of the Optical Society of America. He serves or served on the editorial boards of leading journals (IEEE TRANSACTIONS ON COMMUNICATIONS, *IEEE Photonics Technology Letters*, and *Wireless Networks*) and on program committees of leading conferences (Optical Fiber Communication conference (OFC), Conference on Lasers and Electrooptics (CLEO), IEEE Lasers and Electrooptics Society (LEOS), The international Society of Optical Engineering (SPIE), and GLOBECOM). He also serves or served as a Reviewer for various IEEE and IEE Transactions, Proceedings, and Journals; funding agencies (NSF, OFC, European Research Council (ERC), National Research Council (NRC), etc.), and publishers (Wiley, MacMillan, etc.).