A Wide-Band Nonblocking 2×2 Switching Node for a SPINet Network

Assaf Shacham, *Student Member, IEEE*, Benjamin G. Lee, *Student Member, IEEE*, and Keren Bergman, *Member, IEEE*

***Abstract*—We present SPINet (Scalable Photonic Integrated Network), a novel optical packet switching architecture that is chiefly designed for implementation with photonic integrated circuit technology, where fiber delay lines cannot be employed. A novel physical layer acknowledgment protocol is introduced to mitigate the penalty associated with message dropping. The architecture's concepts are described and the fabrication of a prototype semiconductor optical amplifier-based switching node is reported. Routing of 160-Gb/s (16×10 Gb/s) optically addressed messages with bit-error rates better than 10^{-12} and the successful back-propagating transmission of acknowledgment pulses are experimentally demonstrated.**

***Index Terms*—Multiprocessor interconnection, optical interconnections, optical packet switching (OPS), photonic switching systems.**

I. Introduction

PHOTONIC interconnection technologies provide an attractive solution that can potentially overcome some limitations of electronics in communications systems where intensive data exchange at high bandwidths and low latencies are required [1]–[4]. Dynamic power consumption, wiring density, and signal distortion are fundamental impediments to the scaling of electronic interconnection networks [5], [6]. The advantages offered by optical networks, namely wavelength-division multiplexing (WDM), bit-rate transparency, and low propagation loss can facilitate an interconnect solution that overcomes these problems. Moreover, improvements in the compactness of photonic interconnects and expected advances in large-scale integration show promise of placing an entire interconnection network on a single photonic integrated circuit (PIC) [7]. Such PIC networks will enable compact optical interconnects that deliver immense bandwidths at latencies approaching optical time of flight.

The Scalable Photonic Integrated Network (SPINet) photonic switching architecture presented in this letter is specifically designed for integration as it does not employ fiber delay lines (FDLs) for message buffering or deflection. In SPINet, port-to-port messages are self-routed through an optical multistage interconnection network, constructed from 2×2 wide-band photonic switching nodes, while the payload is maintained in the optical domain across the network. The messages are constructed in a manner that takes advantage of WDM to achieve high bandwidth and simplify the node design. Due to the high level of integration, FDLs cannot be utilized for message buffering and contentions are resolved by dropping contending messages. A novel *physical-layer acknowledgment* protocol takes advantage of the bidirectional transparency and ultrashort propagation delay to facilitate immediate message retransmissions when necessary. SPINet, therefore, offers a comprehensive optical switching solution that is integratable, does not require external management that impedes scalability, and includes an inherent contention resolution mechanism.

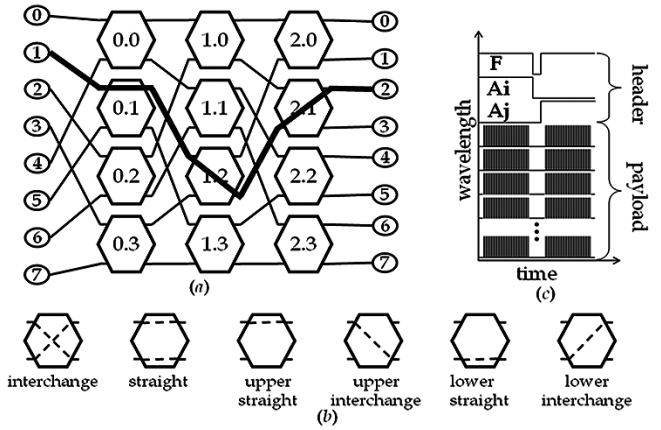

Fig. 1. (a) Omega network with a lightpath across it. (b) Each node has six switching states. (c) The wavelength parallel messages.

Initial research on the performance of SPINet-based networks has been performed and reported in [8]. In this letter, we report the fabrication of a semiconductor optical amplifier (SOA)-based wide-band nonblocking 2×2 switching node using passive optic elements along with electronic and optoelectronic devices integrated on a printed circuit board (PCB). The fabricated switching node, used as a building block for a SPINet network, is dynamically programmable to follow different routing rules and can therefore be used in different locations in the network [8].

II. Architecture Overview

A SPINet network is a binary butterfly-class network, comprised of 2×2 photonic switching nodes, where every user has access to one input and one output terminal. An Omega network [9] [Fig. 1(a)] modified to support six switching states [Fig. 1(b)] is one possible network topology [8].

The network is designed for integration of many switching nodes and interconnecting waveguides on a single PIC. The propagation latency across such an integrated network is expected to be well below 1 ns. In this realm even very short messages (tens of nanoseconds) will appear as continuous lightpaths

Manuscript received June 4, 2005; revised August 3, 2005. This work was supported in part by the National Science Foundation under Grant CCF 0523771.

The authors are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: assaf@ee.columbia.edu; benlee@ee.columbia.edu; bergman@ee.columbia.edu).

Digital Object Identifier 10.1109/LPT.2005.859169

that stretch across the entire network [Fig. 1(a)]. It is impractical in this regime to buffer messages within the network. We, therefore, assume that the duration of the messages exceeds the round-trip propagation delay of light through the network.

Messages in the SPINet architecture are encoded on multiple wavelengths to fully exploit the bandwidth offered by WDM [2]. Each message includes a header, encoded on several dedicated wavelengths, a single bit per wavelength, which may contain framing, address, and priority information. The header field wavelengths remain constant throughout the duration of the message to allow for simple extraction and decoding. The message payload is divided into segments that are modulated, at a much higher rate, on multiple wavelengths that occupy the rest of the transmission band. PIC technology can simplify the encoding and decoding of the multiple-wavelength messages [7]. The header and payload wavelengths propagate together as a unit in the network [Fig. 1(c)]. Group velocity dispersion and resulting wavelength walkoff can be neglected since the propagation distances associated with the networks considered here are less than hundreds of meters for the largest scale systems.

The system is time-slotted and synchronous. At the beginning of each slot, terminals start transmitting messages without a prior request or grant and the leading edges of the messages start propagating across the network (see Fig. 1). In every switching node, a few header bits are extracted to make a local routing decision. The messages are routed accordingly from stage to stage and lightpaths are formed across the network. In the case of contention (i.e., two messages addressed to the same port in a switching node), one of the messages is dropped.

When the leading edges are received at the destination terminals, an acknowledgment (ack) optical pulse is generated by the receiving module to acknowledge the reception of the message. The ack pulses are transmitted in the reverse direction along the transparent lightpaths established by the messages to the appropriate source terminals, signaling the successful reception of the message. This occurs within the single time slot, simultaneously with the transmission of the remainder of the messages. Thus, the ack pulses are received before the end of the slot, to allow for a timely retransmission decision. Dropped messages will obviously not lead to an ack generation and the sources of these dropped messages will not receive an ack and may choose to retransmit at the next slot.

When the slot ends, all the sources cease transmission, the lightpaths are torn down, and the system is ready for a new slot. Since the slot length is constrained by the round-trip propagation delay through the PIC and the fiber connecting it to the source terminal, it can be kept as short as a typical optical packet in an optical packet switching (OPS) system (tens of nanoseconds). The SPINet can, therefore, be viewed as a hybrid approach that uses ideas from circuit switching to implement an OPS solution. The performance parameters of the system (acceptance rate, throughput, etc.) have been studied in [8] and a 40-Gb/s average bandwidth per port has been confirmed for a 64-port network [8]. In this letter, we report on the implementation of a prototype system.

III. 2 × 2 Switching Node

Among several implementation technologies, a structure of four SOA gates (Fig. 2) is an attractive option. SOAs offer the required switching time (<1 ns), bandwidth, and integrability,

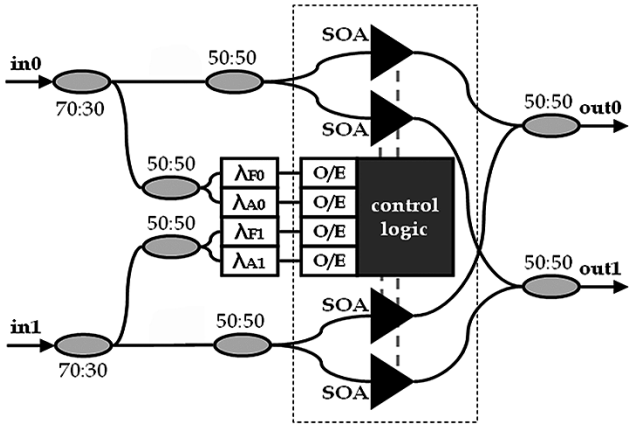

Fig. 2. Experimental node is comprised of SOAs, optical couplers (ellipses, with coupling ratios), wavelength filters (λ), p-i-n receivers (O/E), optical fibers, and an electronic control circuit implemented by a CPLD.

as well as a substantial gain to compensate for coupling losses in the node. Studies of multihop SOA-based networks have shown that multiple-wavelength optical signals can undergo tens of node-hops while maintaining adequate signal quality [10]. Noise and gain compression effects can also be mitigated if the SOAs are operated in the linear regime by controlling the input power [11].

Although designed for integration, most of SPINet's functionality can be modeled and tested using individually packaged components, interconnected to construct a switching node. A prototype switching node has been built using SOAs, p-i-n photodetectors, and electronic circuitry, mounted on a custom-designed PCB, and passive optical elements (couplers and wavelength filters). The electronic control logic is implemented as a Xilinx complex programmable logic device (CPLD) that provides reprogrammable sequential logic with a fast propagation delay (<5 ns).

When the messages enter the node, 30% of their power is tapped off, while the rest of the signal is propagating in a parallel fiber path. The tapped power is then split by a 3-dB coupler and directed into wavelength filters to extract the frame bit ($\lambda_F = 1555.75$ nm, denoting message existence) and the address bit ($\lambda_A = 1532.68$ nm, requested output port). The four bits (two bits from each message) are detected using 115-MHz p-i-n-TIA receivers and are driven into the CPLD. The CPLD controls four SOA drivers, and turns them on according to a routing decision based on a preprogrammed truth-table. When turned on, the Avanex A1901 SOAs are driven with approximately 60 mA and produce the 8.5 dB of gain required to compensate for the coupling and connector losses. At the end of the message, when the optically encoded header bits turn off, the SOAs are switched off accordingly.

The prototype node allows us to verify the feasibility of the address decoding, the correct routing, and the data integrity of the routed messages.

IV. Experimental Results

A testbed (Fig. 3) has been designed to construct the multiple-wavelength messages, to inject them into the switching node, and to analyze them after they are routed. Sixteen distributed feedback lasers ($\lambda = 1531.6$ nm through 1560.2 nm, minimum spacing <0.8 nm) are coupled together on a fiber, modulated

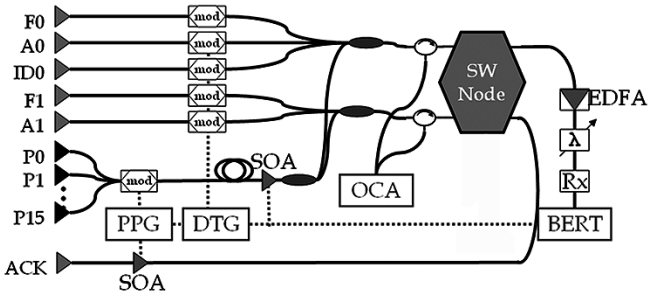

Fig. 3. Optical testbed, designed to construct and analyze the wavelength-parallel messages and *ack* pulses. (PPG: pulse pattern generator. DTG: data timing generator. OCA: optical communications analyzer.)

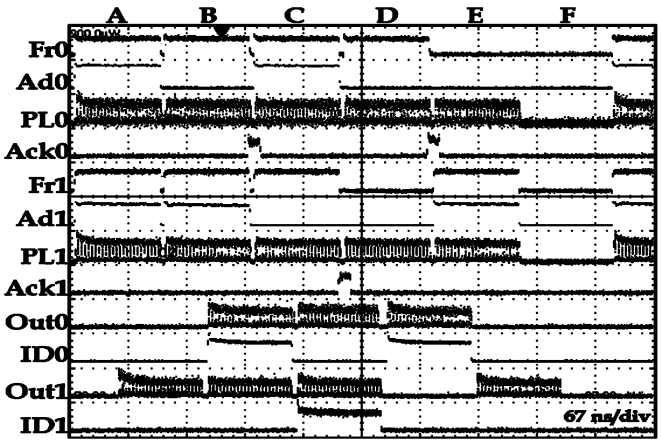

Fig. 4. Optical signals at the inputs (per input: frame, address, payload, and ack) and the outputs (per output: payload and source id) demonstrate the correct routing functionality and reception of acks over six slots (A through F). Acks are generated only in output Port 0. Four messages are sent from each input port (the same payload is driven to both ports, but the frame qualifies only four messages per port).

by a pulse pattern generator (PPG) at 10 Gb/s, and decorrelated by 450 ps/nm in fiber to form a 160-Gb/s wavelength-parallel payload. The payload is then segmented to packets by an SOA, split to two input ports and coupled with the frame and address wavelengths which are modulated separately per port. An ID wavelength ($\lambda = 1533.47$) is coupled to messages injected into Input Port 0, allowing for source identification at the outputs.

The output of the node is directed to an EDFA followed by a filter for wavelength selection, a 10-Gb/s p-i-n receiver, and a bit-error-rate tester (BERT). The BERT is synchronized with the PPG and is gated to measure packet bit-error rate (BER).

Acknowledgment pulses are modulated on a separate wavelength ($\lambda = 1550.92$) using an SOA and transmitted in the back propagating direction via Output Port 0 when messages are received. Optical circulators are located at the node inputs to allow for the extraction of the ack pulses to an optical communication analyzer.

A demonstration of the correct functionality is shown in Fig. 4, where four messages are driven into each port over a six-slot time window. The slot length is 102.4 ns, and the message length is 97.6 ns, leaving a deadtime of 4.8 ns between messages. In Slot A, both inputs attempt transmission to Output Port 1 ($\mathrm{Ad0} = 1$, $\mathrm{Ad1} = 1$), but only the message from Input Port 1 is received ($\mathrm{ID1} = 0$), while the other message is dropped. In Slot C, Input 0 is transmitting to Output 1 while

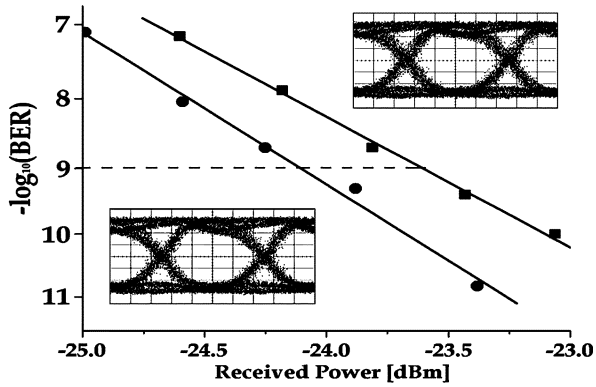

Fig. 5. BER curves and eye diagrams are given for back-to-back (•, left eye) and node output (■, right eye). The power penalty of the node is measured to be less than 0.5 dB at BER $= 10^{-9}$ ($\lambda = 1554.42$ nm).

Input 1 is transmitting to Output 0. Both messages are received and an ack pulse from Output 0 is received in Input 1.

The integrity of the routed data is also verified. BER $< 10^{-12}$ is measured on all 16 payload wavelengths on all four input–output paths. The power penalty incurred by the node is measured to be lower than 0.5 dB (Fig. 5).

V. Conclusion

The SPINet architecture, which makes use of wide-band photonic 2×2 switching nodes and features a unique physical layer acknowledgment protocol, has been presented. The requirements from the switching nodes (optical address decoding, fast switching, uniform response across a wide band, and error-free transmission) have been defined, realized, and verified on a prototype node. Correct routing of back-propagating ack optical pulses has also been shown. Future work will include performance research of the SPINet architecture as well as fabrication of arrays of interconnected photonic nodes according to the discussed principles.

References

[1] "The Future of Supercomputing: An Interim Report," NRC, National Academies Press, 2003.

[2] B. A. Small, O. Liboiron-Ladouceur, A. Shacham, J. P. Mack, and K. Bergman, "Demonstration of a complete 12-Port terabit capacity optical packet switching fabric," in *Optical Fiber Commun. Conf. (OFC)*, Anaheim, CA, 2005, Paper OWK1.

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

[4] G. I. Papadimitriou, C. Papazoglou, and A. S. Pomportsis, "Optical switching: Switch fabrics, techniques, and architectures," *J. Lightw. Technol.*, vol. 21, no. 2, pp. 384–405, Feb. 2003.

[5] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proc. IEEE*, vol. 88, no. 6, pp. 728–749, Jun. 2000.

[6] W. J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, CA: Morgan Kaufmann, 2004.

[7] R. Nagarajan *et al.*, "Large-Scale photonic integrated circuits," *IEEE J. Sel. Topics Qunatum Electron.*, vol. 11, no. 1, pp. 50–65, Jan. 2005.

[8] A. Shacham, B. G. Lee, and K. Bergman, "A scalable, self-routed, terabit capacity, photonic interconnection network," in *13th Annu. IEEE Symp. High Performance Interconnects*, Stanford, CA, 2005.

[9] D. H. Lawrie, "Access and alignment of data in an array processor," *IEEE Trans. Comput.*, vol. 24, no. 12, pp. 1145–1155, Dec. 1975.

[10] O. Liboiron-Ladouceur, W. Lu, B. A. Small, and K. Bergman, "Physical layer scalability demonstration of a WDM packet interconnection network," in *Proc. LEOS 2004*, Rio Grande, PR, pp. 567–568.

[11] M. J. Connelly, *Semiconductor Optical Amplifiers*. Boston, MA: Kluwer Academic, 2002.