Ultra-Low Latency Optical Packet Switching Node

Benjamin A. Small, *Student Member, IEEE*, Assaf Shacham, *Student Member, IEEE*, and Keren Bergman, *Member, IEEE*

***Abstract*—A photonic packet switching node is introduced, and its routing latency is shown to be 15.3 ns. The power penalty of the node at a bit-error rate (BER) of 10^{-9} is confirmed to be approximately 0.2 dB across 33 nm of the C-band for 10-Gb/s payload wavelengths. Moreover, multiple-wavelength packets containing 16 payload wavelengths can be switched while maintaining BERs of 10^{-12} or better.**

***Index Terms*—Interconnection networks, packet switching, photonic switching systems, wavelength-division multiplexing (WDM).**

I. Introduction

MODERN high-performance computing systems require interconnection networks with extremely high throughput and low latency in order to pass messages between thousands of processor and memory elements [1]. Optical packet switching (OPS) fabrics offer a potentially viable solution to this requirement as fiber-optic components are capable of carrying many terabits per second of encoded optical data while maintaining near speed-of-light limited transit latencies [2], [3]. Much research has been accomplished in designing and implementing OPS nodes [4]–[6]. The recent advent of semiconductor optical amplifiers (SOAs) as commercially available wide bandwidth amplifying switching elements has enabled new opportunities for implementation of large-scale photonic switching systems [2], [3].

We present an elemental wavelength-division-multiplexing (WDM) wavelength-parallel (bit-parallel) OPS node for use in distributed routing (self-routing) deflection networks. This node has two input ports and two output ports (i.e., 2×2 structure) and is designed for *single-packet* routing. It also accepts a deflection signal input and can generate an output deflection signal, consistent with incorporation in a fully implemented deflection routing network [1]. Throughout the node design and routing structure, absolute minimalism and simplicity are favored in order to reduce routing latency, so that optical packets can be routed on-the-fly and can approach physical time-of-flight limitations.

The resulting SOA-based node implementation completes all necessary routing and packet ejection in less than 15.3 ns for packets with payload bandwidths of 160 Gb/s. Also, the switching speed from one packet to the next can be as low as 1.6 ns. The receiver power penalty induced on the 16-wavelength WDM payload is near 0.2 dB at a bit-error rate (BER) of 10^{-9}.

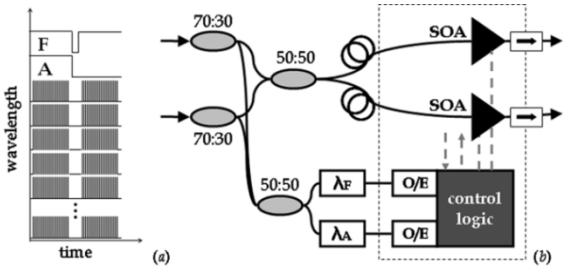

Fig. 1. (a) Illustration of the WDM wavelength-parallel payload and control signal packet format. (b) Schematic of the implemented node (ovals for optical couplers, λ filters, O/E photodetectors, and boxed arrows for isolators, dashed arrows for deflection signal input and output).

II. Functionality

The switching node is designed for use with a particular packet format, which utilizes wavelength-parallel control signals on at least two wavelengths, and contains packet payload information on other wavelengths [7]–[9] (Fig. 1). With the *frame* and *address* control signals encoded with only one bit per wavelength over the duration of the packet, fast routing latencies are easily obtained since routing information is available simultaneously with the packet payload, and because no complex time-division-multiplexing operations are required. In fact, only a simple optical bandpass filter and an optoelectronic (O/E) detector are required to decode each optical control signal [7]–[9]. Cascades of these nodes can be arranged such that each node evaluates on a different address wavelength, resulting in a routing tree similar to a conventional binary banyan network (e.g., shuffle, butterfly).

Based upon the frame (F) and address (A) bits which accompany the packet payload, and from the input deflection signal, the packets are routed from their input port to one of two output ports. In order to be routed to the primary output port, the address bit must match a preset value ("0" or "1") and the deflection signal must be inactive. When the address does not match, or when the deflection signal is active, the whole multiple-wavelength packet is routed to the secondary output port; at the same time, the output deflection signal is activated. When no packet is present, both output ports are deactivated, mitigating the propagation of spurious noise within the system. It is important to reemphasize this node's functionality as a single-packet routing structure. Only one multiple-wavelength packet may be incident on the node in a particular time slot; deflection signaling is utilized to ensure that packet collisions are avoided.

This node's deflection routing structure is fundamentally different from conventional deflection routing implementations. Whereas conventional architectures allow for on-the-fly

Manuscript received November 23, 2004; revised February 22, 2005. This work was supported in part by the National Science Foundation under Grant ECS-0322813 and by the U.S. Department of Defense under Subcontract B-12-644.

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

Digital Object Identifier 10.1109/LPT.2005.848276

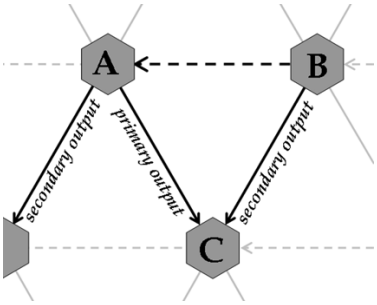

Fig. 2. Diagram of the control signaling relationship (hexagons for nodes, solid lines for fiber-optic pathways, dashed lines for deflection signals). Nodes **A** and **B** compete for injection to node **C**. In order to prevent a collision, **B** sends a deflection signal to **A**, resulting in **A**'s packet being diverted.

Fig. 3. Plot of frame (F) and address (A) optical waveforms, with electrical deflection signal (D), and the resulting packet payload waveforms at the output ports (1 and 2); cursors indicate the 18.9-ns node delay, which includes 73 cm of superfluous fiber pigtail.

contention resolution to occur within the individual switching nodes, the structure presented provides a distributed control mechanism for ensuring that two packets are never simultaneously incident on the same switching node. Nodes must be arranged so that, when two nodes share a common destination node, one must be able to provide the other with a deflection signal in order to prevent collisions (Fig. 2). If a node's output is blocked, the packet is diverted to an alternate destination node (via the secondary output). The deflection signaling structure further guarantees that packets have at least one free output path at every node. A hierarchal routing structure allows packets to be deflected while still maintaining a path to their network destination [7].

A triangular node arrangement (Fig. 2) is necessary to fully utilize the deflection signaling implemented in the node design. This triangular building block can be arranged to form any variety of topological configurations. However, the triangular structure of the deflection signaling implies that a strict but physically reasonable timing relationship between primary output latencies and secondary output latencies must be met, in order for the deflected packet to be given enough time to be routed appropriately: The path following a secondary output port must be longer than the path following a primary output port by an amount equal to the deflection signal transit and processing latencies. This relationship is discussed in more detail in [7].

III. Implementation

The node structure (Fig. 1) is implemented with electronic and O/E devices mounted on a printed circuit board (PCB), and utilizes conventional fiber-optic components; standard p-i-n photodetectors and SOAs bridge the electronic and photonic subsystems [9]. When packets enter the node, the frame and header wavelengths are first filtered off for processing; meanwhile, the packets are delayed for a short time within a length of fiber, until the routing decision is complete and the appropriate SOA is enabled.

The node implementation was designed to be straight-forwardly reproduced to form a complete network. Recently, 36 such nodes were successfully arranged to form a complete 12×12 OPS network [10]. The components are almost entirely interchangeable, with only a few settings based upon the node's location within the switching network topology. For example, the filters for address bit recognition differ from node to node based upon the hierarchal level at which the node is placed, permitting a straightforward implementation of a banyan addressing structure; the frame wavelength, on the other hand, is universal. The PCB electronics can be set with jumpers, configuring the node to match an address bit of either "1" or "0" (or the frame and address pairs "01," "10," or "11") also based upon the node's position within the topological hierarchy. Additionally, the values of the SOA drive current are variable so that empirical propagation losses can be compensated with increased gain. Some degree of variability is also possible in the voltage settings of the two photodetectors, and in the voltage threshold of the limiting amplifiers, allowing for a range of control wavelength signal powers to be used. Also, to ensure that the timing between the node's subsystems is well matched, a programmable delay element is incorporated within the electronics.

In order for this node to be as modular and as scalable as possible, an input packet should be nearly indistinguishable from an output packet: The node should route packets transparently. To this end, it is crucial that the power and wavelength characteristics be unaffected by the node. All insertion losses due to branching and control signal decoding are restored to all wavelengths in the packet by the wide bandwidth SOA amplifiers before packets are ejected. The conventional adding and dropping of control or payload wavelengths is inappropriate for this design. Here, all optical control signals remain with the multiple-wavelength packet in space and in time as it propagates throughout the switching fabric.

IV. Results

A. Routing Confirmation

Confirmation of the node's routing behavior is shown in Fig. 3. With no frame present, no packets exit the node. When the address value does not match, or when the externally generated deflection signal is active, the complete multiple-wavelength packet is routed to the secondary output port (third, fourth, and first packets, respectively). A packet emerges from the primary output port only when the deflection signal is inactive, and when the address value is correct (second packet).

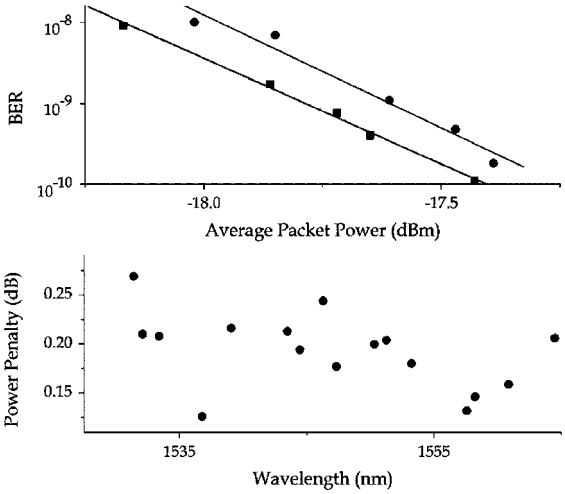

Fig. 4. Plot of signal power required for BER values near 10^{-9} with (•) and without (■) the switching node at the 1552.5-nm (C31) payload wavelength, indicating a receiver power penalty of 0.18 dB at a BER of 10^{-9} (top pane). Plot of payload wavelengths' power penalty against wavelength value, with mean 0.193 dB and sample standard deviation 0.038 dB (bottom pane).

B. Routing Latency

The routing latency can be determined by analyzing the waveforms of Fig. 3. The SOAs are followed by 70 cm of superfluous fiber pigtail, yielding a total packet routing and ejection latency of 15.3 ns, which is equivalent to just 320 cm of conventional single-mode fiber. Attributed to the fiber-optic couplers and filters is 4.3 ns of this latency, with the remaining 11.0 ns due to the PCB electronics and O/E [9].

The node can route 24.1-ns-long packets, which contain 3.2 ns of guardtime and 20.9 ns of 16-wavelength payload. Consecutive packets are spaced by 1.6 ns, which is limited by the switching speed of the SOAs. These figures imply an 80% payload throughput duty efficiency.

C. Routing Transparency

Moreover, the node is almost perfectly transparent to the routed packets. Packets contain 16 WDM payload wavelengths nonreturn-to-zero modulated at 10 Gb/s each with a $2^9 - 1$ pseudorandom binary sequence which are decorrelated by approximately 450 ps/nm, in addition to the two control wavelengths. The wavelengths used conform to the ITU WDM (100 GHz) grid specifications, and some adjacent wavelengths are spaced by just 0.8 nm. All 16 payload wavelengths attain a 10^{-12} BER for the intended optical power levels. BERs are measured only within the packet payloads by precisely utilizing the gating functionality of a conventional BER tester. Because the node contains less than 320 cm of optical fiber, no significant wavelength dispersion was observed.

The receiver power penalty δ induced by the node on each of the 16 payload wavelengths is approximately 0.2 dB at a BER of 10^{-9} across a 33-nm range at the middle of the C-band (Fig. 4). The average power penalty figure is consistent with measurements reported in [11] and [12]. It also agrees with the analytically calculated result for an ideal amplifier with the SOAs' measured noise figure of approximately 7 dB

$$\delta \approx \frac{(r-1)(r-n+1)}{(r+1)(r-n-1)} \tag{1}$$

where r represents the signal extinction ratio and n the noise figure, assuming the signal not to be bandwidth-limited [13].

This low noise figure was obtained by driving Kamelian OPB-10 commercial SOA devices with a current of approximately 35 mA, which provides sufficient gain to compensate for the ∼5 dB of coupler losses while maintaining fairly low amplified spontaneous emission (ASE) noise and minimal nonlinear effects. These devices have a transparency current of about 18 mA, and their ASE spectrum is concentrated near 1465 nm. When in the low-gain operating mode, many SOAs can be cascaded without substantial signal degradation [12].

Additionally, the net power difference induced by the node is measured to be less than 0.9 dB for each of the 16 payload wavelengths, consistent with [12]. The switching contrast ratio of the SOAs is found to exceed 50 dB.

V. Conclusion

We have designed and tested an ultra-low latency OPS node which achieves routing latencies of just 15.3 ns. We demonstrate 160-Gb/s (16×10 Gb/s) WDM throughput and measure 0.2-dB receiver power penalty at a BER of 10^{-9}. This photonic node design paradigm shows promise in becoming a fundamental building block for future large-scale OPS networks.

References

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

[2] 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.

[3] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer, "Photonic packet switches: Architectures and experimental implementations," *Proc. IEEE*, vol. 82, pp. 1650–1667, Nov. 1994.

[4] D. J. Blumenthal, K. Y. Chen, J. Ma, R. F. Feuerstein, and J. R. Sauer, "Demonstration of a deflection routing 2×2 photonic switch for computer interconnects," *IEEE Photon. Technol. Lett.*, vol. 4, no. 2, pp. 169–173, Feb. 1992.

[5] R. M. Fortenberry, Y. Cai, and R. S. Tucker, "Optically transparent nodes for photonic packet-switched ring networks," *Electron. Lett.*, vol. 29, pp. 417–418, Feb. 1993.

[6] *Feature Issue on Optical Interconnection Networks (OIN), J. Opt. Netw.*, vol. 3, pp. 760–836, Nov. 2004.

[7] Q. Yang and K. Bergman, "Traffic control and WDM routing in the data vortex packet switch," *IEEE Photon. Technol. Lett.*, vol. 14, no. 2, pp. 236–238, Feb. 2002.

[8] Q. Yang, K. Bergman, G. D. Hughes, and F. G. Johnson, "WDM packet routing for high-capacity data networks," *J. Lightw. Technol.*, vol. 19, no. 10, pp. 1420–1426, Oct. 2001.

[9] A. Shacham, B. A. Small, O. Liboiron-Ladouceur, J. P. Mack, and K. Bergman, "An ultra-low latency routing node for optical packet interconnection networks," in *Proc. 17th Annu. LEOS Meeting*, Nov. 2004, Paper WM2, pp. 565–566.

[10] 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 *Proc. Optical Fiber Commun. Conf. (OFC)*, Anaheim, CA, Mar. 2005, Paper OWK1.

[11] W. Lu, B. A. Small, J. P. Mack, L. Leng, and K. Bergman, "Optical packet routing and virtual buffering in an eight-node data vortex switching fabric," *IEEE Photon. Technol. Lett.*, vol. 16, no. 8, pp. 1981–1983, Aug. 2004.

[12] W. Lu, O. Liboiron-Ladouceur, B. A. Small, and K. Bergman, "Cascading switching nodes in data vortex optical packet interconnection network," *Electron. Lett.*, vol. 40, pp. 895–96, Jul. 2004.

[13] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd ed. New York: Wiley, 2002.