Optical Packet Routing and Virtual Buffering in an Eight-Node Data Vortex Switching Fabric

W. Lu, *Member, IEEE*, B. A. Small, *Student Member, IEEE*, J. P. Mack, *Student Member, IEEE*,
L. Leng, *Member, IEEE*, and K. Bergman, *Member, IEEE*

***Abstract*—We demonstrate the routing of optical packets containing an eight-wavelength-division-multiplexing (WDM) channel 10-Gb/s payload through an eight-node subnetwork of the data vortex packet switching fabric. The single-pulse WDM encoding employed for the header and frame bits greatly simplifies the routing control functions and reduces the switching latency. The unique synchronous and distributed control mechanism eliminates the requirement of optical buffering and avoids packet contention. We demonstrate the virtual buffering capability of the switching fabric through a cascade of seven node hops, achieving a power penalty of less than 1.5 dB for each of the eight payload channels.**

***Index Terms*—Optical buffering, photonics packet switching, semiconductor optical amplifiers (SOAs).**

I. Introduction

In CURRENT large-scale performance computing systems, an emerging need for ultrahigh capacity low latency communications between processors and shared memory has led investigators to consider insertion of photonic interconnection networks. It is well recognized that although processor speeds and memory densities continue to grow at immense rates, processor-memory access latency is forming an increasing communications bottleneck [1]. In particular, for important computing applications such as cryptography and data mining, which exhibit low data reference locality, standard methods of repositioning data are ineffective. To efficiently execute such applications, a high-capacity low-latency processor-memory interconnection network is required. The use of optical technology for the physical switching fabric within a high-performance computing system is clearly advantageous in providing maximum data bandwidth per cable and in lacking the need for signal regeneration for interconnects exceeding tens of meters. Furthermore, the transparency offered in the optical domain allows wide flexibility in the data encoding and protocols [2]–[5]. Exploiting wavelength-division-multiplexing (WDM) data encoding can enable efficient realization of ultrahigh throughput capacities. However, the key challenges to implementing an optical packet switching fabric include contention resolution and packet buffering, which often necessitate conversion to the electronic domain [6]. The use of optical buffering typically eliminates the advantage of data format transparency and presents a severe practical challenge if the data is encoded in several wavelength channels [7].

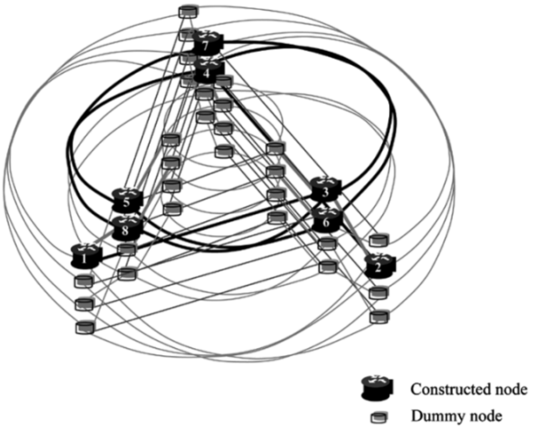

Fig. 1. Schematic of the data vortex topology with $A = 3$, $H = 4$, and $C = 3$. Numbered cylindrical nodes are experimentally implemented.

Many switch fabric architectures that exhibit excellent performance in electronics are not implemented well in the optical domain. Since such switch fabrics were designed for electronic implementation, complex routing control and abundant buffering are generally necessary to achieve high performance and provide the necessary services and protection. The architecture of the data vortex switching fabric has been specifically proposed to address the unique issues associated with inserting an optical packet interconnection network in high-performance computing systems [8], [9]. To avoid packet contention, the data vortex employs a synchronous and distributed control mechanism for the packet flow. As a result, each switch node encounters at most one packet in a given packet time window, and no optical buffering is required to mitigate contention. This traffic control mechanism leads to packet deflection. However, the probability of multiple deflections for any one packet is minimized since packets are provided multiple paths to the destination [10]. An open path provided by the "angle" dimension of the architecture is always available to a deflected packet and, thus, provides a virtual buffering mechanism internal to the switching fabric.

Fig. 1 shows the schematic of the data vortex switch architecture. The size of the data vortex can be described by three parameters: C, A, and H, corresponding to the cylinder, angle, and height parameters, respectively. Shown in Fig. 1 is a complete switching fabric of size $C = 3$, $A = 3$, and $H = 4$ which would support 12 input and 12 output ports. The number of cylinder levels (C) scales as $C = \log_2 H + 1$. Data packets ingress from input nodes at the outermost cylinder and exit from output nodes at the innermost cylinder. The innermost cylinder ($c = \log_2 H$) also provides virtual buffering, allowing packets

Manuscript received November 26, 2003; revised April 9, 2004. This work was supported in part by the National Science Foundation (ECS-0 322 813) and by the Department of Defense (Subcontract B-12-644).

W. Lu, B. A. Small, J. P. Mack, and K. Bergman are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: bergman@ee.columbia.edu).

L. Leng is with the New York City College of Technology, CUNY, Brooklyn, NY 11201 USA.

Digital Object Identifier 10.1109/LPT.2004.829762

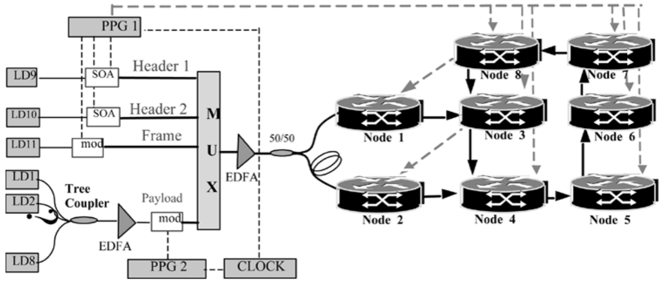

Fig. 2. Experimental setup of eight-node data vortex subnetwork. LD: laser diode. EDFA: Erbium-doped fiber amplifier. Optical path (solid line). Electrical path (dashed line). Control line (bold dash). MOD: modulator.

to circulate when the output electronic buffers are busy. The connection links within the cylinders cross in a binary tree fashion to fix the next most significant bit of the header address before the packet is forwarded to an inner cylinder level. The connections between cylinders simply forward packets maintaining the packet height.

The switching fabric operates in a synchronized manner. Input packets are generated synchronously in a fashion similar to messages sent by processor or memory terminals controlled by a global clock. Each packet moves one angle forward within each packet time cycle. As the packets propagate from the outer cylinder toward the inner cylinder, only a single bit of their target height address is decoded within each switching node. Thus, switching nodes in the data vortex are extremely simple: They consist of two input ports, one from an outer cylinder (north) and one from the same cylinder (west); and two output ports, one to the same cylinder (east) and one to an inner cylinder (south), as shown in Fig. 2(a). The routing logic truth table is shown in Fig. 2(b). In each switching node, when a packet is deflected to the east, a control bit ZERO is simultaneously sent to the corresponding node at the outer cylinder to prevent contention. Therefore, only a single packet enters a node in any given packet time cycle, eliminating the need to perform contention resolution.

We have recently demonstrated routing and traffic control between two data vortex switching nodes with WDM encoded optical packets [11], [12]. In this letter, we report on the implementation of an eight-node data vortex subnetwork within a 12×12 switching fabric that demonstrates the critical switch fabric functionalities: packet generation and insertion through two input nodes, which are interconnected to six switch nodes; packet flow control among two cylinder levels; and packet deflection and virtual buffering at the inner cylinder. The routed packets contain eight WDM payload channels modulated at 10 Gb/s per channel. The packets' header and framing bits are encoded in a bit-parallel WDM fashion to simplify the routing and to reduce latency [13]. Each switch node includes two semiconductor optical amplifiers (SOAs) that switch the packets and compensate for the coupling losses. The SOAs in each node consume only tens of milliamperes of current, eliminating the need for continuous temperature control. The virtual optical buffering is realized by properly programming the control signal of each switching node. The power penalty incurred after the WDM packets traverse seven hops is shown to be approximately 1.5 dB for each of the eight payload channels.

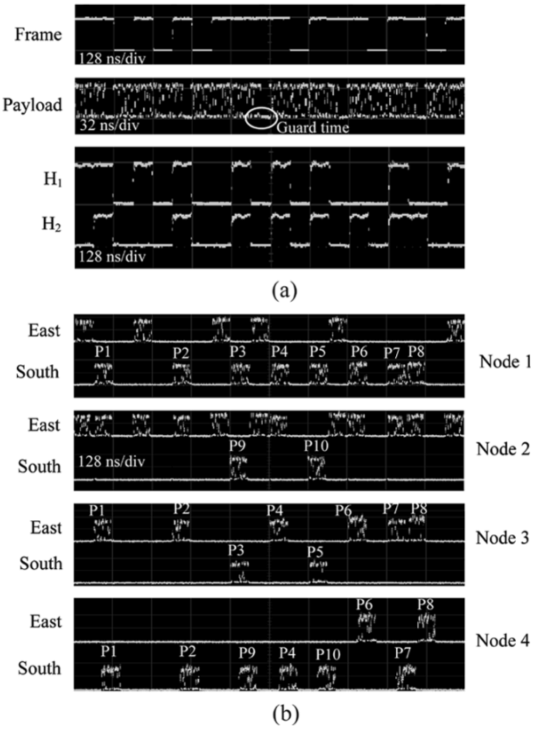

Fig. 3. (a) Data format of frame, headers, and payload. (b) Routing results at four routing nodes. The top trace in each window shows the packets out from east port of the node and the bottom trace shows the south port of the node.

II. Experiment

The arrangement of the experimentally implemented eight-node subnetwork within the data vortex architecture is shown by the highlighted numbered nodes in Fig. 1. Fig. 3(a) details the experimental setup. Each packet consists of two header channels, one frame channel, and eight payload channels. We use SOAs to modulate the header wavelength channels so that extinction ratios of larger than 40 dB can be obtained. This prevents the rise of accumulated noise in the ZERO bit of the header address and avoids critical logic errors during the routing process. Nodes 1 and 2 are input nodes at the outer cylinder. Nodes 3 through 8 are switching and output nodes at the inner cylinder. Nodes 8 and 3 provide control signals to Nodes 1 and 2, respectively, so that the contention at Nodes 3 and 4 is avoided, as illustrated in Fig. 3(b). The control signals for Nodes 3 through 8 are generated by a multichannel pulse pattern generator (PPG 1). In each node, 10% of the optical power is tapped off to extract the header and frame information. The control logic board processes the header, frame, and control bits to generate the gating signals for switching the SOAs. The control logic board also produces a gating signal directed toward the appropriate outer cylinder node to prevent packet contention. Careful timing design is required, since this control signal must reach the outer node before the packet is processed. The total latency in each of the implemented nodes is 30 ns. It is worth noting that the latency is dominated by optical delay due to long fiber pigtails of individual components. If all the optical components were integrated into one chip, the optical delay can be substantially decreased. Taking into account the electrical delay of the control logic board, the overall latency of each switch node could be as low as 2 ns.

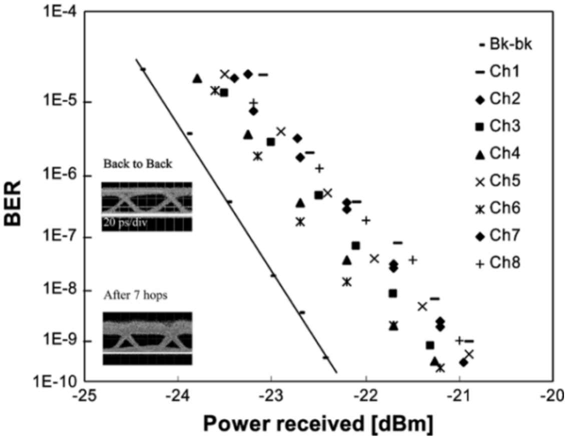

Fig. 4. Measured BER performance for back-to-back and after seven hops.

Fig. 4 shows the packet structure. The header and frame bits are encoded at separate wavelengths as single bits along the entire length of the 64-ns-long packet. To achieve low latency, these bits are extracted by simple wavelength filters and processed in parallel as the packet traverses through the node. Eight WDM payload channels, located between 1530 and 1565 nm, are encoded with 10-Gb/s nonreturn-to-zero data streams. The encoded data was a $2^9 - 1$ pseudorandom bit sequence. A guard time of 6.4 ns is inserted between adjacent packets to allow for routing transients. We programmed a data sequence of 20 packets, each carrying address destinations (H_2H_1), as shown in Fig. 5(a). Although the packets' payloads are continuously generated, packets with a ZERO frame bit are treated as empty and blocked at the input nodes, as indicated by the logic truth table. Therefore, there are 14 valid packets (with frame bits ONE) initially injected into Nodes 1 and 2. The packets injected through input Node 2 are delayed relative to the packets at Node 1 by 30 ns in order to achieve more independent packet sequencing. Fig. 5(b) shows the routing results for one example routing sequence. Initially, the control signal sent to Node 1 by Node 8 is ONE, since no packet is passing through Node 8. Thus, in accordance with the generated routing destinations and frame bits, eight packets (P1 to P8) correctly propagate from Node 1 to inner cylinder Node 3, while the remaining six exit Node 1 through its east port. At Node 3, the control signal in each sequence is 1011001110, 0 110 010 111. As packets traverse Node 3, only two packets (P3 and P5) exit from its south port; the remaining six packets emerge from its east port toward the west port of Node 4. Meanwhile, Node 3 sends a control signal to Node 2 in order to prevent contention at Node 2. As a result, only two packets (P9 and P10) of the 14 from Node 2 propagate to Node 4. The control sequence at Node 4 is programmed as 0100100010, 1 010 000 100. Six packets then exit through the south port of Node 4. The remaining two packets propagate to Nodes 5 and eventually exit from the south port of Node 5.

In the above sequencing example, the maximum number of hops any packet travels is three. In order to test the virtual buffering capability of the data vortex, we specifically programmed a sequence that directs packets to circulate through the testbed inner cylinder. We then measure the bit-error rate (BER) for each of the eight payload channels under two cases: back-to-back and after seven hops. Because the total fiber length traversed by a packet is less than 100 m, dispersion is insignificant and the clock recovery at the receiver is not necessary for the measurement of BER. Additionally, because the effects of cross modulation are cumulative, the nonlinearity is strongest when channels are modulated identically, as we have done [14]. As shown in Fig. 6, the worst channel power penalty compared to the back-to-back case is approximately 1.5 dB at a BER of 10^{-9}. The eye diagrams corresponding to error-free operation are shown as the insets of Fig. 6. For a large scale data vortex switching fabric (e.g., 2 × 2 k), the average number of packet node hops is about 20 [9], [10]. Thus, the node cascading results demonstrate the potential for successfully scaling the switch fabric.

III. Summary

Successful routing of high-capacity WDM optical packets through an eight-node data vortex subnetwork demonstrates the key functionalities of the switching fabric. Single-pulse WDM encoding for the header and frame bits enables low latency and simplifies the packet traffic control. Virtual buffering of the WDM packets is shown through seven node hops incurring a power penalty of 1.5 dB.

References

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

[2] D. J. Blumenthal, B.-E. Olsson, G. Rossi, T. E. Dimmick, L. Rau, M. Mašanović, O. Lavrova, R. Doshi, O. Jerphagnon, J. E. Bowers, V. Kaman, L. A. Coldren, and J. Barton, "All-optical label swapping networks and technologies," *J. Lightwave Technol.*, vol. 18, pp. 2058–2075, Dec. 2000.

[3] B. Meagher, G. K. Chang, G. Ellinas, Y. M. Lin, W. Xin, T. F. Chen, X. Yang, A. Chowdhury, J. Young, S. J. Yoo, C. Lee, M. Z. Iqbal, T. Robe, H. Dai, Y. J. Chen, and W. I. Way, "Design and implementation of ultra-low latency optical label switching for packet-switched WDM networks," *J. Lightwave Technol.*, vol. 18, pp. 1978–1987, Dec. 2000.

[4] C. Guillemot, M. Henry, F. Clerot, A. Le Corre, J. Kervaree, A. Dupas, and P. Gravey, "KEOPS optical packet switch demonstrator: Architecture and performance," in *Optical Fiber Communications (OFC 2000)*, Paper ThO1-1, pp. 204–206.

[5] D. Wonglumsom, I. M. White, S. M. Gemelos, K. Shrikhande, and L. G. Kazovsky, "HORNET—A packet-switched WDM network: Optical packet transmission and recovery," *IEEE Photon. Technol. Lett.*, vol. 11, pp. 1692–1694, Dec. 1999.

[6] *Special Issue on Optical Networks, J. Lightwave Technol.*, vol. 18, pp. 1603–2223, Dec. 2000.

[7] D. K. Hunter, M. C. Chia, and I. Andonovic, "Buffering in optical packet switches," *J. Lightwave Technol.*, vol. 16, pp. 2081—2094, Dec. 1998.

[8] C. Reed, "Multiple Level Minimum Logic Network," U.S. Patent 5 996 020, Nov. 30, 1999.

[9] Q. Yang, K. Bergman, G. D. Hughes, and F. G. Johnson, "WDM packet routing for high-capacity data networks," *J. Lightwave Technol.*, vol. 19, pp. 1420–1426, Oct. 2001.

[10] Q. Yang and K. Bergman, "Performance of the data vortex switch architecture under nonuniform and bursty traffic," *J. Lightwave Technol.*, vol. 20, pp. 1242–1247, Aug. 2002.

[11] ——, "Traffic control and WDM routing in the data vortex packet switch," *IEEE Photon. Technol. Lett.*, vol. 14, pp. 236–238, Feb. 2002.

[12] W. Lu, K. Bergman, and Q. Yang, "WDM routing with low cross-talk in the data vortex packet switching fabric," in *Optical Fiber Communications (OFC 2003)*, 2003, Paper FS4, pp. 795–797.

[13] O. H. Adamczyk, M. C. Cardakli, J.-X. Cai, M. I. Hayee, C. Kim, and A. E. Willner, "Coarse and fine bit synchronization for WDM interconnections using two subcarrier-multiplexed control pilot tones," *IEEE Photon. Technol. Lett.*, vol. 11, pp. 1057–1059, Aug. 1999.

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