

An Ultra-Low Latency Routing Node for Optical Packet Interconnection Networks

Assaf Shacham, Benjamin A. Small, Odile Liboiron-Ladouceur, John P. Mack and Keren Bergman

Department of Electrical Engineering, Columbia University, New York, NY 10027

assaf@ee.columbia.edu; 212-854-2768

Introduction

Photonic interconnection networks are a promising means of routing high bandwidth optical packets in applications ranging from data communications and storage to high-performance computing. Maintaining routed packet traffic in the optical domain alleviates the electronic bottleneck and facilitates low latency and high throughput interconnects which can take full advantage of the bandwidth provided by dense wavelength division multiplexing (DWDM). In this paper, we demonstrate the functionality of a fabricated 2×2 self-routing optical switching node to be used in such a network, and we measure its latency to be 16.4 ns. The packet payload, which can contain multiple 10 Gbps WDM data channels, is transparently routed through the node.

In order to maximize the usefulness of optical communications for interconnection networks, the switching topology should not require buffering [1,2] nor complex address processing [2]. Instead, the routing nodes should require as little decoding of the packet as possible. Nodes should also have free paths available for deflection to realize contention resolution, which eliminates the need for optical buffering [3]. Additionally, the system design must maintain scalability and robustness. While arbiters have often been used in telecommunications networks, this approach often requires buffering and packet queuing, and can be very difficult to scale to large (e.g., 10k-port) interconnection networks [1,2,4].

Self-routing deflection network topologies are indeed very conveniently implemented with fiber optic and lightwave technology [3]. In such networks, the switching is done individually within simple nodes in which the address, or a part of the address, is processed and the routing decision is then made. Deflection paths are guaranteed at every node, and electronic signaling facilitates the deflection decision. These networks are especially appealing for optical packet switching because of the difficulty of implementing optical buffers and complex optical data processing [4].

The simplest possible node that can be used in a self-routing deflection network has two input ports, two output ports, an input for deflection signals, and minimal routing logic, as described in [5,6]. The routing logic must implement some form of address matching and deflection mechanism. Distributed routing eliminates the need for central arbitration, which is often time consuming and computationally complex. For supercomputing systems, one of the key applications of optical interconnection networks, the required data transmission latencies are on the order of nanoseconds [2]. Since deflection networks inherently have larger network diameters [2,4], latency is of paramount importance in their design.

Node Functionality and Design

Upon entering a node, packets with addresses that match the user-programmed value are directed to the primary output port, unless a deflection signal indicates that the next node in that direction is already occupied. When the address does not match or when a deflection signal is received, the packet is directed to the secondary output port, which is always available as a deflection path for properly configured topologies. Thus, the nodes themselves are quite modular, and an entire topology can be constructed by pre-programming the nodes correctly. For example, a butterfly or a Beneš arrangement would reduce the number of hops to approximately the logarithm of the port count [2]. Such an arrangement requires only a binary addressing system. A packet qualifier, or *frame*, is encoded on the packet to prevent noise from being mistakenly propagated through the network. The entire routing header is encoded on multiple designated DWDM channels in parallel with one or more payloads channels, which are transparent to the routing mechanism and thus can be modulated at extremely high data rates [4,5].

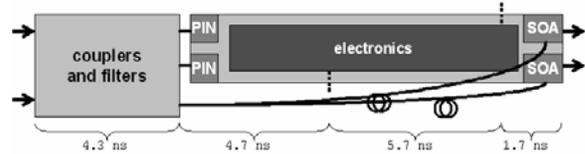


Figure 1. Schematic of the routing node subsystems (dotted lines represent deflection signal paths).

The fabricated switching node contains five major subsystems: the passive optics required to isolate the wavelength routing header channels and to provide the 2×2 pathway configuration; a delay line for the packet while the address is being processed; two photodetectors for the header channels; high-speed electronics for processing the routing logic; and two SOAs to execute the routing decision and to compensate for coupler losses. These subsystems are diagrammed in Figure 1, annotated with the individual latencies. The optoelectronic and electronic devices are integrated onto a single printed circuit board (PCB) in order to reduce latency, size and possible noise. Optical data enters the node at either of the input ports; it is then split at the optical coupler module, and the wavelength-based routing header is filtered off and directed to the photodetectors. The address information is processed, based upon the aforementioned routing logic, and either one or the other of the output SOAs is turned on, directing the packet to either the primary output port or the secondary output port; an appropriate deflection signal is also generated for use in an adjacent node. The packet itself is not affected by the routing procedure (except for a small amount of noise introduced by the SOAs [6]), and

all header information is preserved; the address information carried with a packet remains unchanged throughout the routing path, but the subset of the bits compared can change from node to node, depending on the optical filters used.

The two photodetectors in each node are commercially available *p-i-n* diodes (PINs) with integrated transimpedance amplifiers. The electrical signals are then rectified with a limiting amplifier so that they can be used to drive differential low-voltage positive-referenced emitter coupled logic (LV-PECL) circuitry. The external electronic input and output control lines are connected with microwave cables, and the routing decision is processed electronically with high-speed LV-PECL logic gates. The final stage of electronics switches a current driver which powers only one of the SOAs, setting it to transmit the packet from one of the two output ports. Meanwhile, the optical packet is delayed, having already been split and filtered for the correct header wavelengths. Finally, after the processing has been completed, the entire packet exits from the appropriate port, having been routed transparently.

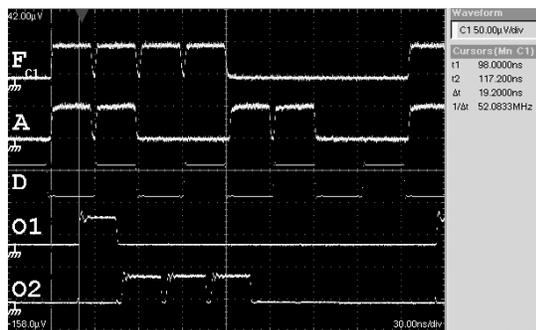


Figure 2. Optical input and successfully routed output of switching node (F: frame, A: address, D: deflection signal [electronic], O1: primary output, O2: secondary output); the 19.2 ns delay includes 70 cm of superfluous fiber pigtail.

The subsystems have each been designed with the goal of minimizing overall switching node latency while maintaining flexibility. Moreover, in order for the node to function properly, the optical delay time for the packet should exactly match the processing time for the headers. Providing all the timing and functional constraints have been satisfied, the node will indeed operate properly as a switching element within a carefully designed photonic distributed deflection-routing interconnection network.

Experimental Results

As illustrated in Figure 2, the delay of the fabricated node is 16.4 ns; the switching transition time is 1.1 ns, and the required deadtime between packets can be as low as 2.0 ns. Figure 2 also demonstrates the correct routing function of the node: when the frame is present and the address bit matches the preset value (logical ‘1’ here), the packet is routed to the primary output port unless a deflection signal blocks it into the secondary output port; when the address bit does not match, the packet is routed to the secondary

output port. When no frame is present, both output ports are disabled. It is further shown that a 10 Gbps NRZ payload is routed transparently by the node, maintaining a bit error rate of better than 10^{-12} (see Fig. 3). The payload may also include more than one wavelength channel, each modulated at 10 Gbps [5].



Figure 3. Eye diagrams of the payload before and after the switching node.

Conclusions

The design of a low-latency optical interconnect switching node has been confirmed. Six of such nodes have been successfully fabricated on a single 6”×19” PCB (see Fig. 4), and have been arranged into a switching fabric subsystem. The latency of each node is 16.4 ns and the required deadtime is 2.0 ns, and because the node itself is payload-transparent, there is almost no limit to the bandwidth of the packets which it can route. Further design improvements will reduce the latency and improve address processing functionality.

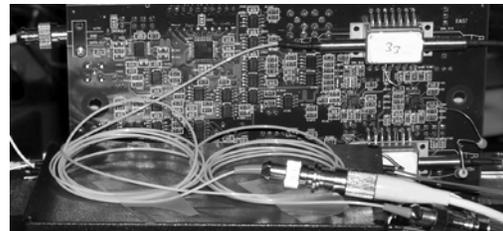


Figure 4. Photograph of an individual switching node (3”×6”) within the six-node PCB setup.

Acknowledgment: This work was supported by the Department of Defense (subcontract B-12-644).

References

- [1] G.I. Papadimitriou, C. Papazoglou, A.S. Pomportsis, “Optical Switching: Switch Fabrics, Techniques, and Architectures,” *J. Lightwave Technol.*, vol. 21, pp. 384-405, February 2003.
- [2] W.J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*, San Francisco: Morgan Kaufmann, 2004.
- [3] J.P. Jue, “An Algorithm for Loopless Deflection in Photonic Packet-Switched Networks,” *Proc. IEEE ICC 2002*, vol. 5, pp. 2776-80, April 2002.
- [4] R. Ramaswami and K.N. Sivarajan, *Optical Networks: A Practical Perspective*, 2nd ed., San Francisco: Morgan Kaufmann, 2002.
- [5] Q. Yang and K. Bergman, “Traffic Control and WDM Routing in the Data Vortex Packet Switch,” *IEEE Photonics Technol. Lett.*, vol. 14, pp. 236-38, February 2002.
- [6] W. Lu, O. Liboiron-Ladouceur, “Cascading Switching Nodes in the Data Vortex Optical Packet Interconnection Network,” to appear in *Electronics Lett.*, submitted 9 April 2004.