

EXPERIMENTAL DEMONSTRATION OF THE SIMPLEST SINGLE-BUFFER DEFLECTION ROUTING TRANSPARENT OPTICAL NODE

A. Bononi, R. K. Boncek[†], P. R. Prucnal, J. L. Stacy[†], and H. F. Bare[†]

Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA

[†] Rome Laboratories, Photonics Center, Griffiss AFB, NY 13441-4515, USA

Abstract — A new single-receiver/single-transmitter/single-buffer node structure for fast packet-switching two-connected transparent optical networks, using only three 2×2 crossbar switches is theoretically analyzed and experimentally demonstrated at 1.244 Gbit/s.

THEORY

Extremely high transmission rates can be provided to each node in distributed multi-hop packet-switching networks in which the optical nodes are connected by dedicated fibers and can perform space-switching of fixed-length packets between their input and output fibers without electro-optic conversions. However, a copy of each packet's header has to be electronically detected and processed for routing purposes. This electronic processing of the header may limit the bit rate in the packets, since routing computations must be performed within a packet's duration. It is thus desirable to have

node structures and control algorithms that are very simple while still providing good throughput-delay performance.

A minimal node structure with a single optical transmitter (*TX*), a single optical receiver (*RX*), employing single-buffer deflection routing [1] is proposed here for two-connected slotted networks. The structure is shown in Fig. 1. The node has two input fibers *I1*, *I2*, and two output fibers *O1*, *O2*. The optical buffer *M* is a one-packet fiber delay loop. *S1*, *S2*, and *S3* are *LiNbO3* crossbar switches. Part of the energy of the incoming packets is tapped out to provide a copy of the packet's header which is photodetected and electronically processed in the Node Controller.

If the buffer *M* were not present, the node would be a strictly non-blocking memoryless 3×3 optical switch with inputs $\{I1, I2, TX\}$ and outputs $\{O1, O2, RX\}$. Such a switch requires a theoretical minimum of three crossbar switches. Contentions between flow-through packets from *I1* and *I2* that vie for the same output fiber could only be resolved by assigning one packet to the desired fiber and deflecting the other to the second output fiber. This is known as hot-potato routing [2].

The buffer *M* prevents deflections, as long as it does not contain a packet in conflict with the input packets. The choice of the node interconnection pattern is crucial for this single-buffer routing to be effective.

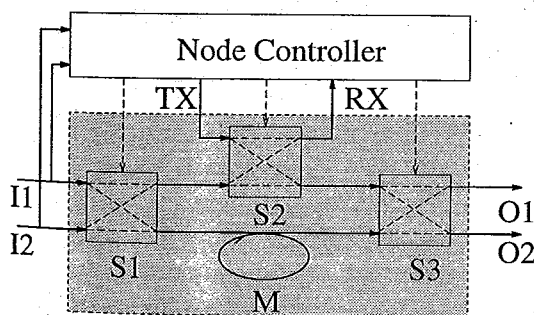


Fig. 1. Node structure.

A topology that provides many alternative paths between each source-destination pair is a good one. In this case it often happens that flow-through packets at $I1$ and $I2$ and packets in the loop M do not care which output fiber they are assigned because both output fibers lead to the packet's destination in the same number of hops. These packets, which we call *don't care packets* (DC), are completely equivalent to empty packets (E) as far as routing is concerned. The strategy of the node controller is to keep the loop M filled with empty or don't care packets as often as possible. This strategy decreases the deflection probability and hence the average number of hops between source-destination pairs, thereby maximizing the network throughput.

Since only one optical receiver RX is available, when $I1$ and $I2$ both contain a packet for the node (FN), one is received and the other is missed. A miss is practically equivalent to a deflection in terms of hops added to the packet's path.

Reception and Access operations are coordinated with buffering operations through switch $S1$. Let $C1$ and $C2$ indicate packets that wish to exit on output $O1$ and $O2$ respectively. At each clock cycle, the two inputs $I1$ and $I2$ and the memory M can take five different values: E, $C1$, $C2$, DC, FN. We detail here the control algorithm of switch $S1$ that provides the highest node throughput:

```

if ( $I1 = I2$ )  $\Rightarrow$  Randomize  $S1$ 
elseif ( $I? = FN$ )  $\Rightarrow$  set  $S1$  to receive it

```

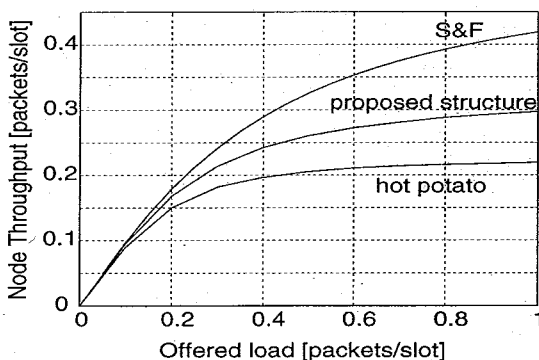


Fig. 2. Node Throughput in a 64-node ShuffleNet in uniform traffic.

```

elseif ( ( $I1, I2$ ) = ( $C1, C2$ ) or ( $C2, C1$ ) ) and
( $M = E$  or DC)  $\Rightarrow$  Randomize  $S1$ 

```

```

elseif ( $I? = E$ ) and ( $TX = full$ )  $\Rightarrow$  set  $S1$  to receive E

```

```

elseif ( $M, I?$ ) = ( $C2, C2$ ) or ( $C1, C1$ )  $\Rightarrow$  store that input

```

```

else store Es or DCs

```

where $I?$ stands for *one of the two inputs*. Randomization in the first line ensures equal treatment of both channels. Absorption of FN packets is done first, as seen in line 2. Line 3 accounts for the fact that two care non-conflicting packets cannot be routed out directly, as the memory cannot be bypassed. In line 4, empty slots are routed to the TX for an injection. In line 5, conflicts with the memory are resolved by storing the conflicting input. Finally, Es and DCs are stored when possible to avoid deflections at the next slot.

Fig. 2 shows node throughput vs offered load (i.e. the probability of having a packet ready at the TX at each clock) for a 64-node ShuffleNet topology [3] in uniform traffic. ShuffleNet is a well-known topology suitable for deflection routing. For comparison, the throughput of nodes with no delay loop (hot-potato routing) and with infinite-buffers (store-and-forward (S&F)) are also shown. The proposed structure yields 71% of the maximum S&F throughput at full load, while the node without the M loop yields only 52%. This is a 37% increase in throughput at the cost of a slight increase in the controller's complexity.

EXPERIMENT

We built an experimental prototype of the node shown in Fig. 1. Three Crystal Technology SW313P $LiNbO_3$ directional couplers with an average fiber-to-fiber loss of 6.4 dB and extinction ratio of -20 dB at $1.3 \mu m$ were used for $S1$, $S2$ and $S3$, for an overall worst-case flow-through loss of 19.2 dB. Such a loss could be more than halved by integrating the three switches. The power levels can be restored by providing an optical amplifier at

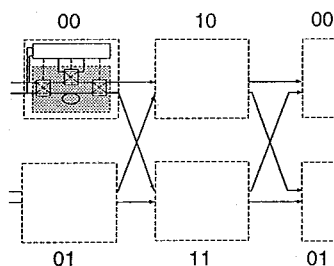


Fig. 3. Experimental node interconnection.

each node output. However, a second important limiting factor on the maximum usable bit rate is the optical noise introduced by these amplifiers [4], which is proportional to the gain. The minimum loss feature of the proposed node well matches the requirements for ultrahigh speed networks.

Packets had an ATM structure of 48-byte payload and 5-byte header. Packets were ASK modulated at $1.3 \mu\text{m}$ at a bit rate of 1.24416 Gbit/s (SONET OC-24 hierarchy). In the header, only five bits were used. These were: a) two consecutive framing bits, always set to 1, used for self-clocking; b) an activity bit (A), set to 0 when the packet contains an empty payload, and c) column and row address bits (C,R).

A span of 78.44 m of conventional single-mode fiber was used for the one-slot fiber loop M . This span accounted for the packet duration, plus a guard-band time of 51.44 ns allowing for the switching time of the LiNbO_3 directional couplers.

Optical header detection was implemented only for input port $I1$. Port $I2$ was either idle or fed with a CW beam for BER crosstalk measurements. Packets from $I2$ were electronically simulated by sending directly the bandbase header bits to the node controller. Header detection and self-clocking from input $I1$ was achieved by tapping off a copy of the incoming signal with a 3 dB splitter. The copy was sent to an optical 1×4 divide-and-delay structure for parallel photodetection of the five useful header bits.

The controller at each node must have information about the global network topology to determine the best output for each packet

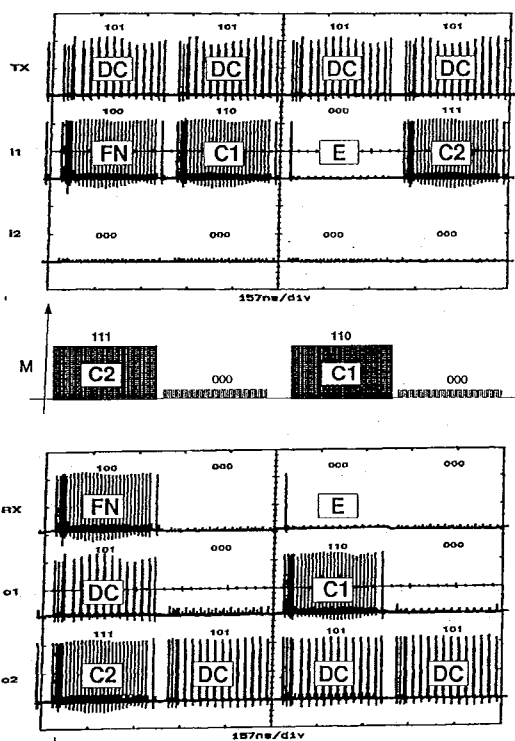


Fig. 4. Measured traces of TX , $I1$, $I2$, RX , $O1$, $O2$ demonstrating the correct implementation of the controller algorithm. The trace of the memory M was not measured, but is sketched here for ease of interpretation.

destination. In the experiment, node 00 of the four-node banyan interconnect shown in Fig. 3 was built. From node 00, destination 10 is reachable preferably from $O1$, i.e. it is $C1$ with respect to 00, destination 11 is $C2$, and destination 01 is DC . The control algorithm was programmed on an off-the-shelf Intel 5C090 CMOS programmable logic array (PLA). The measured header recognition time was 92 ns. Of these, approximately 83 ns were associated with the setup and hold time requirements of the PLA. The PLA inputs were the three header information bits (A,C,R) from each of the two inputs $I1$, $I2$, from the memory M , and from the transmitter TX , as well as two bits (00) denoting the address of the node.

Experimental results of switching at node 00, with input $I2$ idle and empty $I2$ packets simulated at the PLA, are shown in Figure 4.

The three bits over each packet indicate the (A,C,R) contents of its header, which is not visible in the scale shown. These are the three bits feeding the PLA for each packet. For example, (101) indicates a packet to destination 01, i.e. a DC packet. The kind of packet, DC in the example, is also written on the measurements for ease of interpretation. The top three traces represent the incoming packets at TX , $I1$ and $I2$. These traces were taken by biasing switches $S1$, $S2$ and $S3$ in the bar state and measuring the outputs at ports RX , $O1$ and $O2$ respectively. For clarity of presentation, the payload of the TX packets had half the duty-cycle of the payload of the $I1$ packets. In these measurements the bias of the $LiNbO_3$ switches was not optimized to minimize the crosstalk. The bottom three traces are measurements of the received signals at RX , $O1$ and $O2$ with the same input packet streams as before, but when the $LiNbO_3$ switches are driven by the node controller. The trace of the memory M was not measured, but its contents is sketched for clarity in the figure. The memory M starts at time-slot 1 with a C2 packet. The figure shows that switching is performed according to our control algorithm. For instance, at time-slot 2, we have: $TX=DC$, $I1=C1$, $I2=E$, $M=E$, so that C1 is sent to the buffer to make room for the injection of the TX don't care packet. The same happens at time-slot 4.

Finally, Figure 5 shows BER measurements at output port $O2$ relative to a 1.244 Gbit/s pseudorandom $2^{23}-1$ bit stream from input port $I2$, with and without (Baseline) crosstalk interference from the other input ports $I1$ and TX . A Penalty of a fraction of dB is observed.

CONCLUSIONS

In conclusion, we have proposed and experimentally demonstrated at 1.244 Gbit/s the simplest node structure for single-buffer deflection routing in two-connected transparent optical networks. Except for the fiber delay loop, the structure could be integrated to

reduce the overall power loss to below 10 dB. The per-packet processing time was 92 ns by using a commercially available CMOS PLA. Given the simplicity of the routing and access algorithm, much shorter processing times can be achieved by using a more sophisticated electronic controller. Although simple, the routing algorithm yields more than 70% of the maximum achievable throughput in uniform traffic. Less benign traffic patterns, however, may degrade this throughput figure.

REFERENCES

- [1] F. Forghieri, A. Bononi and P. R. Prucnal, "Analysis and comparison of hot-potato and single buffer deflection routing in very high bit rate optical mesh networks," to be published in *IEEE Trans. Commun.*
- [2] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.* vol. 12, pp. 1-9, Mar. 1964.
- [3] A. S. Acampora, M. J. Karol, and M. G. Hluchyj, "Terabit lightwave networks: the multihop approach," *AT&T Tech. J.*, vol. 66, pp. 21-34, Nov./Dec. 87.
- [4] A. Bononi, F. Forghieri, and P. R. Prucnal, "Design and channel constraint analysis of ultra-fast multihop all-optical networks with deflection routing employing solitons," *IEEE J. Lightwave Technol.*, vol. 11, no. 12, pp. 2166-2176, Dec. 1993.

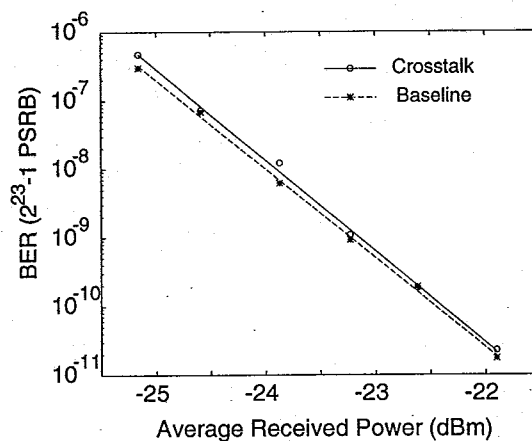


Fig. 5. Measured BER from input port $I2$ to output port $O2$ with and without (baseline) crosstalk from the other input ports TX and $I1$. Input power levels all matched.