Characterization of A Shared Buffer Optoelectronic Packet Router

†Shunyuan Ye, ‡Marina Thottan, ‡Jesse E. Simsarian, †Shivendra Panwar
†Department of ECE, Polytechnic Institute of NYU
‡Bell Laboratories, Alcatel-Lucent
e-mail: sye02@students.poly.edu, {marina.thottan, jesse.simsarian}@alcatel-lucent.com, panwar@catt.poly.edu

***Abstract*—The rapid increase in Internet traffic is forcing packet routers to grow in capacity to meet the demand. Optical packet routers with less buffering and a greater degree of optical transparency are actively being researched as a way to improve energy efficiency and capacity scaling over traditional electronic routers. Since it is difficult to buffer packets in the optical domain, in this paper we analyze the performance of a hybrid optoelectronic packet router. The router architecture has multiple optical switch planes and a shared electronic buffer to resolve output-port contention. By using multiple ports on the switch planes for each input and output fiber, and by using some switch-plane ports to inter-connect the planes, we can achieve a relatively low packet loss ratio in a router with no buffer. In this case, most traffic can be switched using only the through optical paths of the router without entering the shared buffer. The shared electronic buffer is primarily used to reduce the packet drop ratio under periods of heavy loads and occasionally for optical regeneration of a packet. We run extensive simulations to evaluate the performance of the router with varying number of switch plane ports, number of connections to the electronic buffer, and number of interconnections between the switch planes. We show that the router can provide good throughput, with realistic on-off bursty traffic and asynchronous packet arrivals.**

I. Introduction

Internet traffic has been rapidly increasing thereby driving research in energy-efficient routing technologies to meet bandwidth demands without large increases in the power consumption of networking equipment. While there has been substantial research in all-optical routers, they remain impractical due to the difficulty of buffering packets optically. Devices like *fiber delay lines (FDLs)* have been used to buffer optical packets by delaying optical signals for a specific amount of time that is proportional to the length of the FDL. However, FDLs are bulky and inflexible in their buffering capabilities. Moreover, it is inefficient to use FDLs with asynchronous packet arivals. Therefore, researchers have pursued hybrid optoelectronic architectures [1]–[7] that exploit advantages of both electronic and photonic technologies: packets are switched over an optical fabric, and electronic buffers are used to resolve output port contentions when needed.

The OSMOSIS (Optical Shared Memory Supercomputer Interconnect System) project [2], [3] at IBM adopted an input-queued architecture, in which every packet arriving at the switch is buffered by *virtual output queues (VOQs)*. To eliminate the head-of-line blocking problem, each input has to maintain N VOQs, where N is the number of ports. Packets are then switched over an optical fabric. Optical-to-electrical and electrical-to-optical conversions are required at the inputs and outputs of the switch, respectively, and known scheduling algorithms for input-queued switches, i.e. *maximum weight matching (MWM)* [8], [9] and iSLIP [10], can be directly applied. This architecture requires every packet to be buffered, even when there is no packet contention.

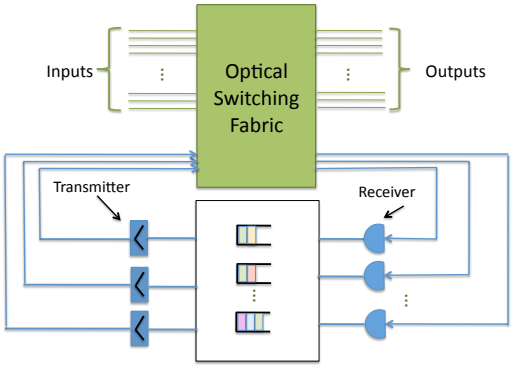

Fig. 1: Model of an optoelectronic shared-buffer router architecture

To reduce electronic buffering, a shared-buffer architecture, as shown in Fig. 1, has been proposed by many researchers [1], [4]–[7]. Different from an *input-queued (IQ)* or an *output-queued (OQ)* switch, an electronic buffer is placed in the loopback path to resolve output-port contentions. This architecture has been used in all-optical switch designs [11], [12], where the buffer is implemented by FDLs. By introducing the loopback path, a packet can either go to the output directly, or first be sent to the buffer and then back to the fabric to be switched to the output. For hybrid shared-buffer architectures, where the buffer is electronic, packets are only sent to the buffer either when there is output-port contention or for packet regeneration, thus reducing power consumption and packet queuing delay as compared to switches that electronically buffer every packet.

This work uses the hybrid shared-buffer architecture of Ref. [1] that has some of the features shown in Fig. 1 but with some modifications: As with typical optical transmission systems, each incoming fiber to the router uses wavelength-division multiplexing (WDM) for increased capacity. Also, instead of using one single large switch fabric to route the packets, we first de-multiplex the incoming wavelengths and

then use multiple switch planes with one wavelength at each switch-plane port to route the packets. A multiplexer is used at each output fiber to multiplex the signals from different planes. The main advantage of this architecture is that we can build a router that scales to a high capacity using smaller-size switching fabrics. We use a single shared buffer for all the planes, allowing packets that arrive at the buffer from one plane to be switched to other planes, enabling load-balancing over multiple planes. While an 8 x 8 prototype of the hybrid optoelectronic router has been demonstrated [1], in this paper, we characterize the performance of the router to find the best configurations as the router is scaled to higher capacity. The main contributions of this paper are as follows:

- Methods for scaling a hybrid optoelectronic shared-buffer router are proposed.
- A strategy for managing the shared buffer is developed.
- A simulated model of the router with different size switch planes, number of connections between the switch planes, and number of connections to the shared buffer along with an evaluation of its performance under different traffic patterns and loading conditions.

The remainder of this paper is organized as follows. We will introduce the router architecture and analyze its performance in Section II. Simulation results will be presented in Section III, with different traffic patterns and router configurations. Section IV concludes the paper.

II. Router Architecture

A. Architecture Overview

The hybrid router is shown in Fig. 2. There are X incoming fibers and X output fibers. Each fiber uses WDM with up to L multiplexed channels. The router has P switch planes, where each plane is an $N \times N$ arrayed waveguide grating (AWG) that routes optical packets from any input port to any output port depending on the wavelength [13]. There is a $1:L$ demultiplexer at each input fiber. After the demultiplexing, $M = L/P$ wavelengths are sent to each plane. For example, wavelengths λ_1 to λ_M are sent to plane one, λ_{M+1} to λ_{2M} are sent to plane two, and so on. There is an $L:1$ multiplexer at each output to combine all the received wavelengths into the output fibers.

Each input and output port of a switch plane are equipped with a wavelength converter except the *to next plane* ports, *to buffer*, and *from buffer* ports that we will describe in following paragraphs. The purpose of the wavelength converters is to translate the optical packets from one wavelength to another and they may be implemented with various optical or opto-electronic technologies [1], [14]. At the input of the switch planes are tunable wavelength converters (TWCs) that determine the output port of the packet from the switch plane depending on the wavelength transmitted by the converter. The TWCs require nanosecond-scale wavelength-tuning times to route on a packet-by-packet basis. At the output of the switch plane are fixed wavelength converters that always transmit at the same wavelength so that the packets will be routed to the output fibers by the optical multiplexers.

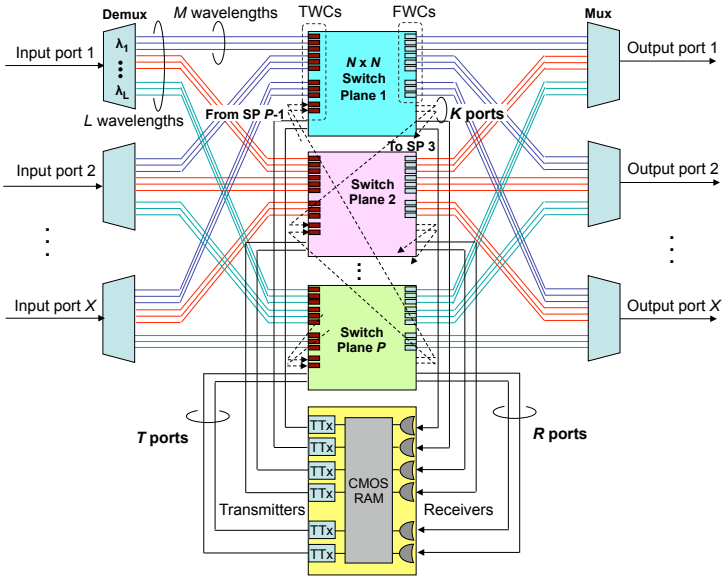

Fig. 2: The router architecture used for the simulations. SP is a switch plane, FWCs are fixed wavelength converters and TTx are tunable transmitters.

A more detailed model of the switch planes in the router is illustrated in Fig. 3. Besides the $X \times M$ ports that connect from the input fibers and to the output fibers, there are K input and output ports per plane to interconnect the switch planes. For example, at plane one, the $next_1$ output port connects to plane two, the $next_2$ port connects to plane three, and so on. Similarly, at the input, there is a $previous_1$ port connecting the plane P to plane one, and $previous_2$ connecting plane $(P-1)$ to plane one, and so on. The *next/previous* ports are introduced to resolve contention and improve the switch throughput. For example, it is possible that some packets destined for output fiber i are blocked at plane j, but at the same time a connection to output fiber i is free at plane $j+1$. By using the *next/previous* ports, some of the blocked packets at plane j can be delivered to the correct output fiber over plane $j+1$.

There are also multiple ports to connect each switch plane to and from the shared electronic buffer. Let R represent the number of *to buffer* ports per plane, and T represent the number of *from buffer* ports per plane. Each *to buffer* port requires a receiver to convert the optical packet to the electronic domain. Similarly, a tunable-wavelength transmitter is needed to send data to the *from buffer* port.

By adding these ports to resolve contentions, either to other planes or to the shared buffer, the number of ports on each switch plane is increased. The number of input and output ports on each switch plane can be calculated using following the equation:

$$N = M \times X + K + R, \tag{1}$$

For simplification, we assume here that $T = R$.

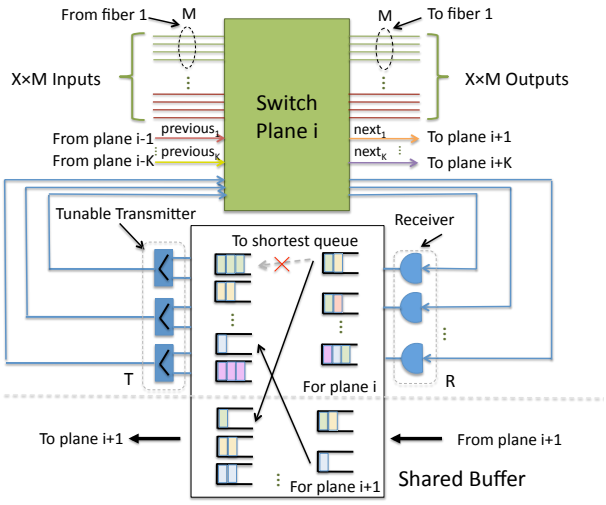

Fig. 3: Diagram of a switch plane and shared buffer.

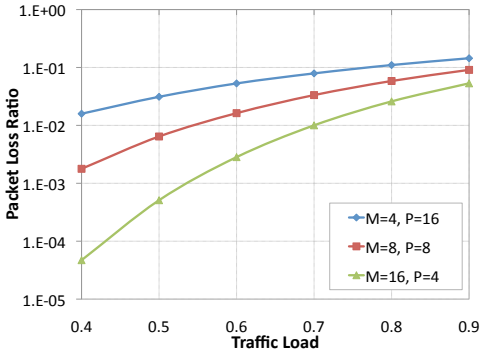

Fig. 4: Impact of the switch size on the packet loss ratio with no *next/previous* ports, $K = 0$, M is the number of wavelengths from each fiber per plane, and P is the number of switch planes and no electronic buffering, $R = 0$

.

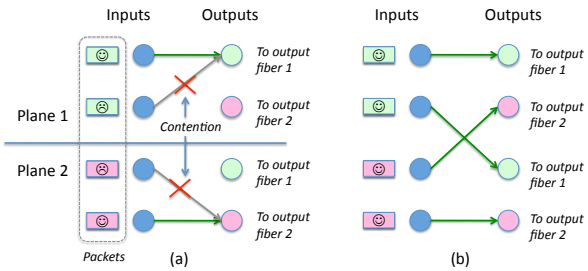

Fig. 5: Multiplexing gain for large switches

B. Scheduling Algorithm

We assume that the system is time-slotted, and each arriving packet has a fixed size. To begin with a simplified analysis, we initially assume that the system is synchronous, but later we will evaluate the performance of the router with asynchronous packet arrivals.

The packets are switched with the following algorithm: after a packet arrives at an input, if there is no contention for its destination output, it will be delivered directly. Otherwise, if the packet cannot be delivered due to contention, it first tries to use an available output to the same fiber at other planes. If the output is free at some plane that the current switch plane is connecting to, the packet can be sent to that plane through the "to next plane" port and be switched to the correct fiber by that plane. Implicit in this algorithm is the fact that we are only concerned with routing packets to the correct fiber, regardless of the wavelength on which it will leave the router. If the packet cannot be delivered to the output fiber over other planes, it may then be sent to the shared buffer. When the number of buffer receivers is less than the number of packets to be simultaneously buffered, some packets are dropped. To empty the shared buffer, stored packets can only be sent back to a switch plane to get switched over the fabric if (1) the destination outputs are free (after new arriving packets have been scheduled), and (2) there is a transmitter available.

This packet scheduling algorithm is simple to implement and since incoming packets always have a higher priority than packets in the shared buffer, unnecessary buffering is avoided. However, packets may not be delivered strictly in order, since a packet arriving late to the switch that does not experience contention may be delivered before an earlier packet that experiences buffering.

C. Number of Switch Planes

Each incoming fiber has L wavelengths, and we are interested in finding an optimal number of switch planes, P, and an optimal number of connections from each fiber to a plane, $M = L/P$.

We developed a software model of the router and ran simulations with randomly generated packet input traffic. The router parameters for all of the simulations are 8 input and output fibers, $X = 8$, with 64 wavelengths each, $L = 64$. We first investigate the performance of the router with varying M and the results are presented in Fig. 4. As M, the number of wavelengths from each fiber that connect to a single plane increases, we reduce the blocking probability and thus the packet loss ratio. The result can be explained with the example illustrated in Fig. 5. Assume there are two different routers: one router has two planes, each with one port to an output fiber, and the other router only has one plane with two ports to each output fiber. Also assume there are two simultaneous input packets that have the destination of output fiber 1 and two simultaneous input packets with the destination of output fiber 2. For the two-plane router of Fig. 5a, only one packet can be delivered in each plane due to the contention. But for the larger single-plane router of Fig. 5b, all four packets can be delivered, and thus has a better throughput performance due to the *multiplexing* gain. As shown in the example, doubling the size of the switching fabric, can double the throughput in the best case.

We have shown that a larger switch fabric can lead to reduced packet loss, and a reduction in the number of planes

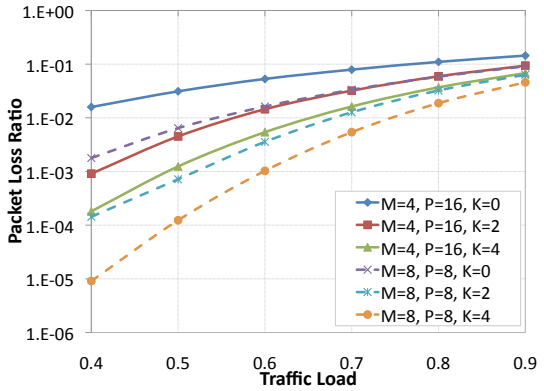

Fig. 6: Impact of adding *next/previous* ports with no electronic buffering, $R = 0$.

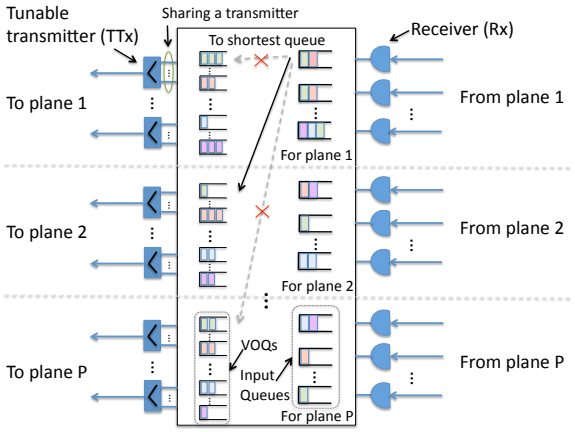

Fig. 7: Detail of the shared electronic buffer

needed. However, increasing the size of a switch plane means that the AWG switch needs more ports and the tunable wavelength converter needs to convert the incoming packets to a wider range of wavelengths. Due to the physical limitations of AWGs and tunable lasers we cannot arbitrarily increase the number of ports on the switch plane to get an acceptable packet loss ratio. Also, as we increase the number of ports on a plane, the implementation of the scheduler becomes more computationally intensive. To solve this problem, we use multiple switch planes and assume a practical number of ports on each plane, which may be 40 and up to 80 ports [13].

D. Number of Next/Previous Ports

Since there is a limit on the switch plane size, it is difficult to achieve an acceptable packet loss ratio by only increasing the AWG size. Reduced packet loss can only be obtained when the following conditions are both satisfied: 1) some packets are blocked on a switch plane; 2) the destined output ports of these blocked packets are idle on another plane. When the traffic is light, the probability that condition 1) holds is low, and when the traffic is heavy, the probability that condition 2) holds is low. Therefore, most of the time, there are only a few packets that can be sent between planes to utilize the idle ports. Therefore, instead of increasing the switch plane sizes, we connect the switch planes using some additional ports. As shown in Fig. 3, we introduce *next/previous* ports, through which packets that arrive at a particular switch plane can be sent to other planes directly.

Since these ports also increase the size of the fabric, there is a trade off between the number of *next/previous* ports and the packet loss ratio. To understand the impact of adding *next/previous* ports on the router performance, we run simulations also with $X = 8$ fibers and $L = 64$ wavelengths. Fig. 6 shows the results for the cases when $M = 4, P = 16$ and $M = 8, P = 8$. We see that by adding *next/previous* ports the packet loss ratio can be substantially reduced. When $M = 4$ and $K = 2$ ($N = 8 \times 4 + 2 = 34$), the performance is almost the same as the case that $M = 8$ and $K = 0$ ($N = 8 \times 8 + 0 = 64$). So by adding just a few *next/previous* ports, we can use more of the smaller switch fabrics to provide nearly the same throughput as fewer large switch fabrics.

E. Shared Electronic Buffer

As we see from Fig. 6, when $M = 4, K = 4$ the loss probability is around 6% when the load is 0.9, which is still high. To further resolve the contentions, a shared electronic buffer can be used to buffer the packets that cannot be delivered. We need R *to buffer* output ports and T *from buffer* input ports on the switch fabric, and we assume that $T = R$. Note that these ports are expensive in power consumption and electronic component requirements and each has to be equipped with an optical receiver and a tunable transmitter. There is a trade off to be made between the number of ports, R, and the packet loss ratio. From previous sections, we know that by having multiple ports for each fiber on a switch plane and adding a few *next/previous* ports, the fraction of packets that cannot be delivered is relatively low. Therefore, it is possible that only a small number of *to buffer* ports is needed to achieve a high throughput. We now analyze the impact R by simulations.

Fig. 7 shows the architecture of the shared electronic buffer. There is an input FIFO queue for each receiver and each plane has multiple *virtual output queues (VOQs)*. Let VOQ_{pj} represent the VOQ for output j at plane p and Q_{pj} represent the occupancy of VOQ_{pj}. Incoming packets for an output j can be switched to any VOQ_{pj}, where $p = 1, 2 \cdots P$. So packets coming from a switch plane can be switched to other planes in the buffer.

To better utilize the lasers at each plane and have load balancing over the planes, we use a simple algorithm called *shortest queue first (SQF)* to switch new incoming packets to the buffer. For a packet destined to output fiber j, the algorithm picks plane p, which is the solution to $\min_p\{Q_{pj}\}$. The shared buffer sends packets back to the switching fabrics when there are free output ports. Since there are only a few transmitters that empty the buffers, contentions may occur when there are multiple outputs available on a plane. The *longest queue first (LQF)* algorithm is used to resolve the contentions. When there

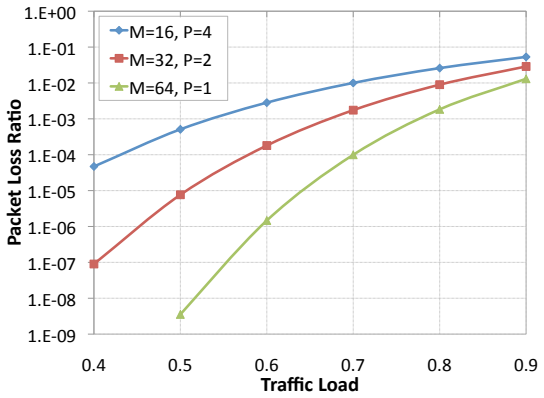

Fig. 8: Packet loss ratio for an all-optical router with large switch fabrics and no *next/previous* ports, $K = 0$, or electronic buffering, $R = 0$. M is the number of wavelengths from each fiber connected to a plane and P is the number of planes.

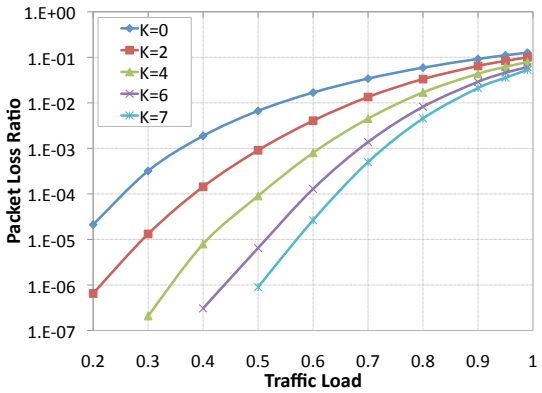

Fig. 9: Packet loss ratio for an all-optical router with $M = 8$ wavelengths from each fiber per plane, $P = 8$ planes, K *next/previous* ports per plane, and no electronic buffering, $R = 0$.

are multiple packets to be sent to a switch plane, the VOQ with a longer queue size is served first.

We assume that the shared buffer has a size of S, including all the input queues and VOQs. When the incoming traffic is heavy, traffic sent to the buffer is also heavy, and new arriving packets are dropped if the buffer is full.

III. Simulations

To evaluate the performance of the router, we ran extensive simulations, each for millions of time slots. We study the router's performance with different switch configurations and under different traffic patterns, including Bernoulli i.i.d. (*identically and independently distributed*), and more realistic On-Off bursty traffic. For each traffic setting, we also subjected the switch to varying loading conditions.

A. Optical Router

As we previously showed, the packet blocking probability can be reduced by increasing the switch size and reducing the number of planes. Even though it is difficult to build optical switching fabrics with a large size, it is still interesting to see in simulation whether it is possible to design an all-optical router by only increasing the switch size. We again assume that there are $X = 8$ fibers, each with 64 incoming wavelengths. Fig. 8 shows the results. Note that the switch size is $N = X \times M$.

As we can see, when the switch size is large, for example $N = 64 \times 8 = 512$, the packet loss ratio can be reduced to less than 1% even when the traffic load is 0.9. So if it is possible to build large switch fabrics in the future, an all-optical router can be built to achieve a low loss ratio without any buffers.

Since an optical fabric size is physically limited, it is impractical to build an all-optical router by only increasing the switch size. Another approach is to connect the switch planes by *next/previous* ports. We limit the switch size to less than 80 input ports, which has been demonstrated previously [13], and increase the number of *next/prevous* ports. Fig. 9 shows the results. When the traffic load is light, the packet loss ratio can be reduced significantly by adding *next/previous* ports. But when the traffic is heavy, the packet loss ratio is still high. This is because when traffic is heavy, it is more difficult to find an idle port on other planes, and increasing the number of *next/previous* ports does not help. Fig. 10 shows the router's performance under uniform and non-uniform traffic:

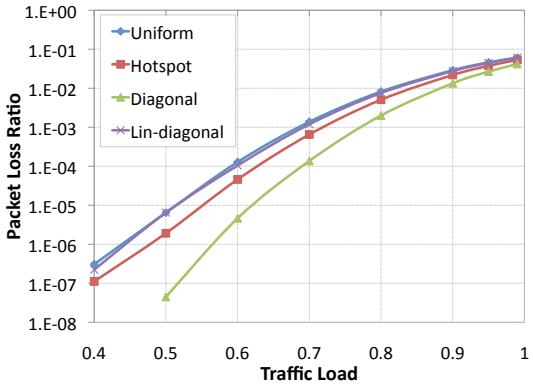

Fig. 10: Packet loss ratio for an all-optical router with non-uniform Bernoulli Traffic with $M = 8$ wavelengths from each fiber per plane, $P = 8$ planes, $K = 6$ *next/previous* ports per plane, and no electronic buffering, $R = 0$.

- Hotspot: a packet at input fiber i is destined to output fiber i with probability 0.5, and to other outputs with equal probabilities, which is $\frac{1}{2(X-1)}$.
- Diagonal: a packet at input fiber i is destined to output fiber i and $i+1$ only, with equal probabilities.
- Lin-diagonal: a packet at input i is destined to outputs with probabilities differing linearly: $r_{i(i+j \pmod X)} - r_{i(i+j+1 \pmod X)} = 2r/X(X+1)$.

As we can see, non-uniform traffic does not increase the packet loss ratio. For diagonal traffic, there is even a significant drop on the packet loss.

B. Hybrid Router with an Electronic Buffer

As shown in the previous section, an electronic buffer is needed in order to achieve an acceptable loss ratio when the

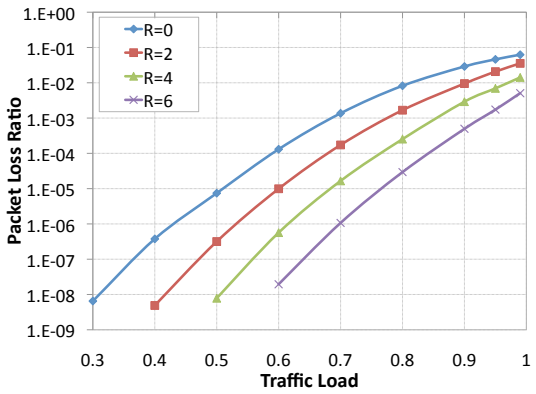

Fig. 11: Impact on the packet loss ratio of adding buffer transceivers to the hybrid router when $M = 8$, $P = 8$, and $K = 6$.

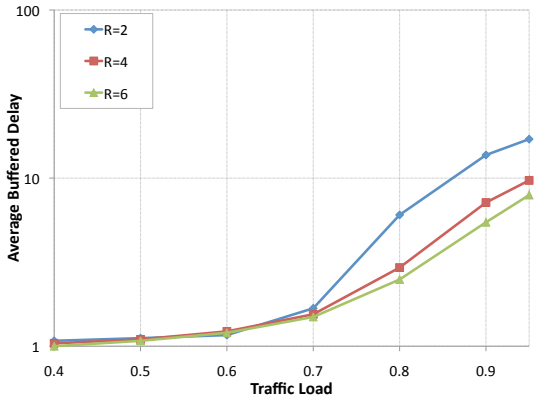

Fig. 12: Average buffered delay as buffer transceivers are added to the hybrid router planes with $M = 8$, $P = 8$, and $K = 6$.

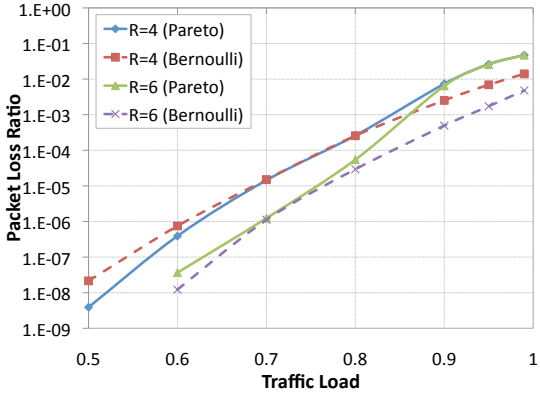

Fig. 13: Packet loss ratio comparison for the hybrid router with bursty and non-bursty traffic and different number of buffer transceivers per plane, R, with $M = 8$, $P = 8$, and $K = 6$.

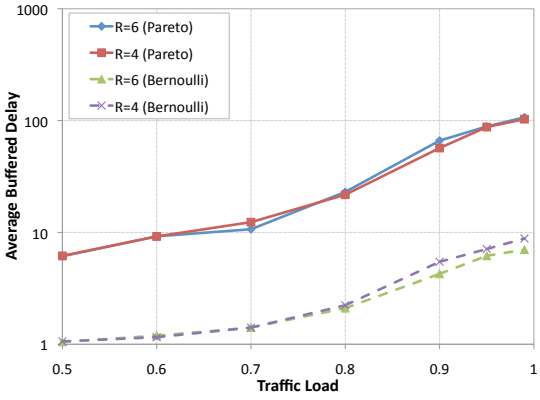

Fig. 14: Average buffered delay comparison for the hybrid router with bursty and non-bursty traffic and different number of buffer transceivers per plane, R, with $M = 8$, $P = 8$, and $K = 6$.

traffic is heavy. Therefore, each switch plane should have some ports connecting to/from the electronic buffers. We assume that there are R receivers and transmitters at the shared buffer for each plane. Since each optical buffer port requires electronic hardware to electronically multiplex/de-multiplex, process, and buffer the packets, we want to keep the number of buffer ports small. We ran the simulations with varying R to find the best configuration assuming that the buffer size is $S = 1000$.

From the results of Fig. 11, we see that, when $R = 6$ the packet loss ratio can be less than 0.1% even when the traffic is heavy. The switch plane size in this case is only $N = X \times M + K + R = 64 + 6 + 6 = 76$. Therefore, a hybrid optoelectronic router can be built using practically feasible switching fabrics with a low packet loss ratio.

We also show the queuing delay that a packet experiences after being sent to the electronic buffer, Fig. 12. We find that a packet waits for just a few time slots in the buffer. With such a small buffer delay, it is easy for the destination node to do packet reordering.

C. Bursty Traffic

In reality, traffic is not as ideal as Bernoulli i.i.d. traffic. On-off bursty traffic patterns are more typical for the Internet data or for traffic within a data center [15]. In the bursty-traffic simulations we assume that the packet arrival process at an input is characterized by a two-state on-off model. When it is in the ON state, traffic arrives in a bursty mode. The number of time slots spent in the ON state is distributed over $[1, 1000]$, following a truncated Pareto distribution:

$$P(l) = \frac{c}{l^{\alpha}}, \; l = 1, 2, ..., 1000, \qquad (2)$$

where l is the burst length, α is the Pareto parameter and c is the normalization constant. No packets arrive during the OFF state. The number of slots spent in the OFF state is geometrically distributed. In the simulation, we set $\alpha = 1.7$, and the expectation of burst length is $E[l] = 10.6$

We compare the packet loss ratio, buffered delay, and maximum buffer occupancy of the router under Bernoulli i.i.d. and bursty traffic in Fig. 13, Fig. 14 and Fig. 15, respectively. We assume that $M = 8$, $P = 8$, $X = 8$, $K = 6$, and the

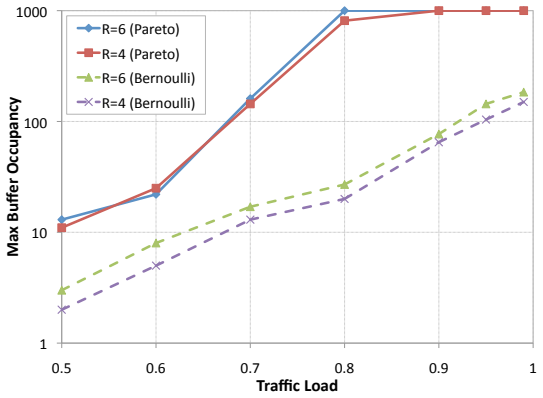

Fig. 15: Maximum buffer occupancy comparison for the hybrid router with bursty and non-bursty traffic and different number of buffer transceivers per plane, R, with $M = 8$, $P = 8$, and $K = 6$.

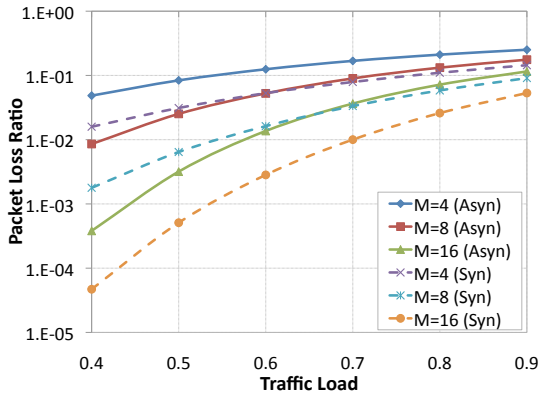

Fig. 16: Packet loss ratio for an all-optical router with no *next/previous* ports, $K = 0$, or electronic buffering, $R = 0$, with synchronous as well as asynchronous packet arrivals.

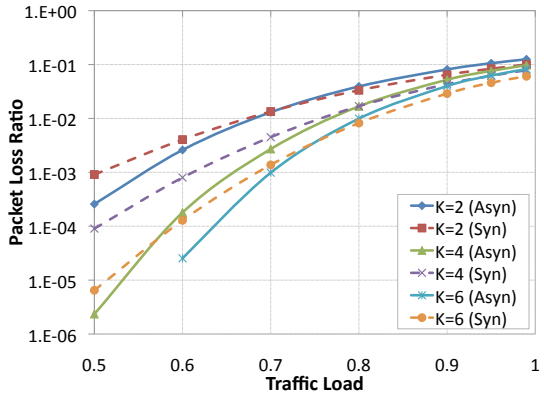

Fig. 17: Packet loss ratio for an all-optical router with varying number of *next/previous* ports, K, no electronic buffering, $R = 0$, $M = 8$ wavelengths from each fiber per plane, and synchronous as well as asynchronous packet arrivals.

buffer size is $S = 1000$, including all VOQs. As we can see, the bursty traffic does increase the packet loss ratio though the increase is not significant except when the traffic is heavy. Note that when traffic is bursty, arrivals at the buffer are also bursty. If the traffic is heavy, the buffer overflows and more packets are dropped at the buffer, leading to a higher packet loss ratio. Increasing the number of receivers R can reduce the blocking probability, but will increase the packet drops due to buffer overflow. When the load is $r = 0.95$, the ratio of packets dropped at the buffer is 1.94% when $R = 4$, and 2.59% when $R = 6$. When $r = 0.99$, the ration of packets dropped at the buffer for $R = 4$ is 3.35%, and 4.31% for $R = 6$. So, if we want to reduce the packet loss ratio for heavy bursty traffic, a larger buffer is needed.

The bursty traffic has a large impact on the buffer occupancy, as shown in Fig. 14 and Fig. 15. The results can be explained by a simple example. Assume that there are two bursts destined for the same output. Since only one of them can be delivered at every time slot, one packet has to be sent to the buffer. Therefore, packet arrivals at the buffer are also bursty. After the packets are sent to the buffer, they have to wait until one of these two bursts stops and no other new bursts destined for the same output arrive. Therefore, on average, packets have to wait in the buffer for a longer time when the traffic arrivals are bursty.

From the results, we can also see that when the traffic is bursty and heavy, packets in the buffer have to wait for longer periods, 10 to 100 slots. This delay may make it more difficult for the end nodes to do packet reordering.

D. Asynchronous Transmission

So far, we have assumed that the system is time-slotted and synchronous, but optical networks are difficult to synchronize. Therefore, we simulate the router performance with asynchronous packet arrivals. The results are shown in Fig. 16, Fig. 17 and Fig. 18.

As we can see from Fig. 16, the packet drop ratio of the router is higher when the traffic is asynchronous. To explain this, we consider a simple example. Assume that there is only one $N \times N$ plane with $K = 0$, $R = 0$ and $M = 1$. We assume that the traffic is uniform with $r = 1$. If the traffic is synchronous, an output port is idle only when all of the arriving packets are destined for other outputs. So we have:

$$B_{syn} = 1 - (1 - \frac{1}{N})^N, \tag{3}$$

where B_{syn} is the probability that an output port is busy. If the traffic is asynchronous, a packet gets dropped when its destination output is busy after it arrives. Let B_{asyn} represent the probability that an output is busy at any time. The probability that its destination output is busy when a packet arrives is then $\frac{N-1}{N} B_{asyn}$, as the output is definitely not busy transmitting from the input on which the new packet arrives. So we have $r(1 - \frac{N-1}{N} B_{asyn}) = B_{asyn}$, from which we can derive:

$$B_{asyn} = \frac{r}{1 + r\frac{N-1}{N}} = \frac{1}{2 - \frac{1}{N}}. \tag{4}$$

Compare Eq. (3) and (4) for any value of N, we always have $B_{syn} > B_{asyn}$. Note that the throughput of an output port

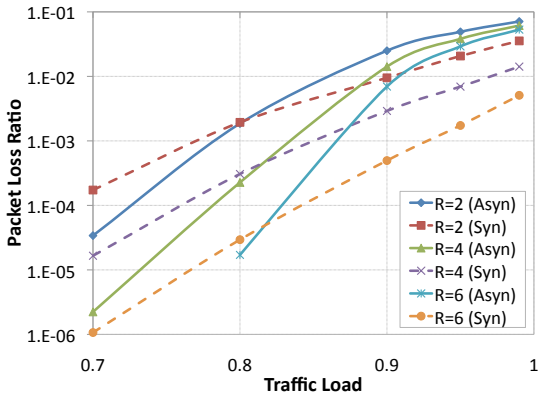

Fig. 18: Packet loss ratio for the hybrid optical router with varying number of electronic buffer transceivers, R, and asynchronous packet arrivals with $M = 8$ wavelengths from each fiber per plane, and $K = 6$ *next/previous* ports per plane. For $R = 6(Asyn)$ with $r = 0.7$, the loss ratio is too small and thus too difficult to get good simulation results.

is equal to: $B_{syn} \times b$ or $B_{asyn} \times b$, where b is the number of packets arrive per second. Therefore, the throughput under synchronous traffic is always higher when the load is heavy.

From Fig. 17, we see that, when *next/previous* ports are used without buffering, the packet loss ratios for asynchronous traffic is less than for synchronous arrivals when the load is not heavy ($r < 0.8$). This means that it is easier to find an idle output port on other planes with light, asynchronous traffic. The same holds true when buffering is included as in Fig. 18. But when the traffic becomes heavy ($r > 0.9$), the asynchronous system will have larger packet drop ratios both with and without buffering. Therefore, to provide an acceptable packet loss ratio, the traffic load should not be too heavy, which may be enforced by techniques such as admission control and traffic shaping at the edge of the network.

IV. Conclusion

In this paper, we analyzed the capacity scaling of a new hybrid shared buffer optoelectronic router. We evaluated the performance of the shared-buffer router and show that a well-configured multiple-plane architecture can achieve good performance characteristics under different traffic patterns and traffic loading conditions.

We show that smaller sized switching fabrics can provide good throughput by using multiple ports per switching plane per fiber and multiple ports to connect the switching planes. The shared buffer configurations are also reasonable since the maximum buffer size is of the order of 1000 packets and the number of shared buffer ports that have to be equipped with optical receivers and tunable lasers are 6 per plane or fewer. As shown in Fig. 11, when the number of ports per plane for each fiber, $M = 8$, the number of planes, $P = 8$, with $K = 6$ ports connecting to other planes, and $R = 6$ buffer receivers per plane, the packet loss ratio with synchronous Bernoulli traffic is less than $1E - 6$ when the traffic load is $r = 0.7$, and even the traffic load is as heavy as $r = 0.9$, the packet loss ratio is still less than $1E - 3$.

Based on this characterization study we conclude that the ideal operating environment for this switch is in scenarios where the arriving traffic is relatively smooth, like the aggregated traffic in the Internet core network or the aggregated traffic from multiple racks of servers in a data center. If the traffic is bursty or asynchronous, a larger buffer is needed to achieve a good packet loss ratio. In addition, when the traffic is bursty, the end systems should support packet reordering. This is not unlike the case of conventional high speed routers that use hardware parallelism for capacity scaling.

V. Acknowledgement

This work was performed under the joint NTT Corporation and Alcatel-Lucent Optical Packet Switching Collaboration.

References

[1] H. Takenouchi, R. Urata, T. Nakahara, T. Segawa, H. Ishikawa, and R. Takahashi, "First Demonstration of a Prototype Hybrid Optoelectronic Router," in *35th European Conference and Exhibition on Optical Communication*, September 2009.

[2] R. Hemenway, R. Grzybowski, C. Minkenberg, and R. Luijten, "Optical-packet-switched Interconnect for Supercomputer Applications," *Journal of Optical Networks*, vol. 3, no. 12, pp. 900–913, 2004.

[3] C. Minkenberg, F. Abel, R. Krishnamurthy, M. Gusat, P. Dill, I. Iliadis, R. Luijten, B. R. Hemenway, R. Grzybowski, and E. Schiattarella, "Designing a Crossbar Scheduler for HPC Applications," *IEEE Micro*, vol. 26, pp. 58–71, May-June 2006.

[4] Z. Zhang and Y. Yang, "Performance Analysis of Optical Packet Switches Enhanced with Electronic Buffering," in *23th IEEE International Parallel and Distributed Processing Symposium*, May 2009.

[5] L. Liu, Z. Zhang, and Y. Yang, "Packet Scheduling in a Low Latency Optical Packet Switch," in *16th IEEE International Conference on High Performance Switching and Routing*, June 2010.

[6] A. Bianco, D. Camerino, D. Cuda, and F. Neri, "Optics vs. Electronics in Future High-capacity Switches/Routers," in *15th IEEE International Conference on High Performance Switching and Routing*, June 2009.

[7] X. Ye, P. Mejia, Y. Yin, R. Proietti, S. J. B. Yoo, and V. Akella, "DOS - A Scalable Optical Switch for Datacenters," in *ACM/IEEE Symposium on Architectures for Networking and Communications Systems*, October 2010.

[8] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queuing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks," *IEEE Transactions on Automatic Control*, vol. 37, pp. 1936–1949, December 1992.

[9] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% Throughput in an Input-Queued Switch," *IEEE Transactions on Communications*, vol. 47, pp. 1260–1267, August 1999.

[10] N. Mckeown, "The iSLIP Scheduling Algorithm for Input-Queued Switches," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 188–201, 1999.

[11] S. Y. Liew, G. Hu, and H. J. Chao, "Scheduling Algorithms for shared Fiber-Delay-Line Optical Packet Switches - Part I: The Single-Stage Case," *Journal of Lightwave Technology*, vol. 23, no. 4, pp. 1586–, 2005.

[12] T. Zhang, K. Lu, and J. R. Jue, "Shared Fiber Delay Line Buffers in Asynchronous Optical Packet Switches," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 118–127, April 2006.

[13] P. Bernasconi, C. Doerr, C. Dragone, M. Cappuzzo, E. Laskowski, and A. Paunescu, "Large NxN Waveguide Grating Routers," *Journal of Lightwave Technology*, vol. 18, pp. 985–991, 2000.

[14] A. Bhardwaj, J. Simsarian, J. LeGrange, L. Zhang, P. Bernasconi, N. Sauer, L. Buhl, and D. Neilson, "Wavelength Conversion using Semiconductor Optical Amplifiers in a Differential Mach-Zehnder Interferometer with a Tuneable Input Coupler," *Electronics Letters*, vol. 45, no. 4, pp. 225–227, 2009.

[15] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding Data Center Traffic Characteristics," in *Proceedings of ACM Sigcomm*, August 2009.