Performance of Optical Packet Switching Nodes in IP Transport Networks

Maurizio Aste, Achille Pattavina
Dept. of Electronics and Information,
Politecnico di Milano
e-mail:pattavina@elet.polimi.it

Abstract

This paper[1] deals with optical packet switching in a full-IP transport network scenario. For the switching of IP packet flows different node architectures are considered that are based on current optical routing devices. The traffic performance of a mesh network is evaluated with the various node structures, assuming that nodes employ either shortest path routing or deflection routing to forward packets to the addressed destinations. The paper shows how the different node structures behave in terms of packet loss probability with different network configurations when the node parameters are varied.

1 Introduction

In the latest years, the outbreak of Internet and broadband services for the development of electronic commerce, entertainment and education has involved the increase of demand for transmission bandwidth. Nowadays, routing of traffic flows in transport networks occurs by processing electronically data and transmitting them in optical fibers; optics is therefore used exclusively at the physical level. Advent of the wavelength multiplexing technologies (WDM) has concurred an increase of the transmission capacity per fiber but at the same time has shown the limits of actual network infrastructures based on electronic devices. Today networks use only a small fraction of the large-capacity made available by each fiber (in the order of terabits per second), since electronic switching, processing and storage technologies do not allow to manage fully the huge size of data transported by fibers. At the same time the current need for transporting very large amounts of traffic based on the IP protocol has shown all the limits of SONET/SDH technologies, born as solutions for circuit

switching and thus unable to guarantee efficient management of the IP traffic flows.

The development of new switching systems is therefore important in order to face such new networking scenario. The advent of optical switching devices is going to define a new generation of network elements in which routing is operated without optical/electronic conversions. In this context optical circuit and packet switching technologies play a different role. With *optical circuit switching* every connection needs the reservation of an entire WDM channel in order to realize *end-to-end* circuits, with the inefficient bandwidth usage typical of circuit switching. *Optical packet switching* on the other hand enables a high and efficient exploitation of the available capacity thanks to the bandwidth sharing typical of statistical multiplexing. This latter technology would moreover be consistent with the new *paradigma* of an IP protocol that supports any kinds of telecommunication service.

Unfortunately, today optical devices used in market equipment are still too crude to allow packet-by-packet operation. An interesting solution which tries to represent a balance between circuit switching low hardware complexity and packet switching efficient bandwidth utilization is *optical burst switching* [10], [8]. In an optical burst switching system, the basic units of data are bursts, made up of multiple packets, which are sent after control packets, carrying routing information, whose task is to reserve electronically the necessary resources on the intermediate nodes of the transport network.

This paper addresses the future-looking scenario of optical packet switching by exploiting the optical switching technologies available today. In particular it considers the architecture of optical packet switching nodes already proposed in [1][2][3], which exploit arrayed waveguide grating devices for packet routing and are equipped with fiber delay lines used either for input buffering or for shared buffering of optical packets. The paper is organized as follows. Sections 2 describes the envisioned optical network architecture, while section 3 details the proposed structures of the

---

optical packet switching nodes. Finally section 4 provides a traffic performance comparison of the different node architectures in a mesh network with different packet routing strategies.

2 Network Architecture

The architecture of the optical transport network we propose consists of *optical packet-switching nodes*, which are mutually connected in a mesh-like topology. A number of *edge systems* (ES) interfaces the optical transport network with IP legacy (electronic) networks (see figure 1). An ES receives packets from different electronic networks (or local hosts) and performs *optical packet* generation. The optical packet is composed of a simple optical header, which comprises the destination address, and of an optical payload made of a single IP packet, or, alternatively, of an aggregate of IP packets. The optical packets are then buffered and routed through the optical transport network to reach their destination ES, which delivers the traffic it receives to its intended destination in the electronic domain. At each intermediate node in the transport network, packet headers are received and electronically processed, in order to provide routing information to the control electronics, which will properly configure the node resources to switch packet payloads directly in the optical domain.
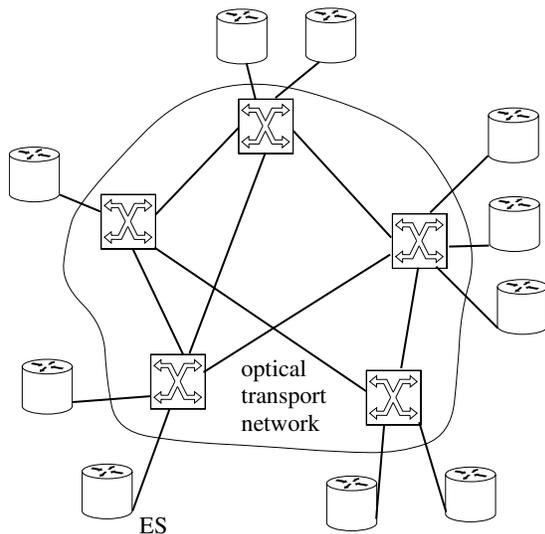

Figure 1. The optical transport network architecture

The transport network operation is *asynchronous*; that is, packets can be received by nodes at any instant, with no time alignment. The internal operation of the optical nodes, on the other hand, is assumed to be *synchronous*,

or *slotted*, since the switching of packets in an unslotted node is less regulated and more unpredictable, resulting in a larger contention probability. In our model we propose, the time slot duration, T, to be equal to the amount of time needed to transmit from an input WDM channel to an output WDM channel an optical packet with a 40-bytes payload which corresponds to the smallest payload enabled by an IP packet. Supposing a bit rate of 10 Gbit/s per wavelength channel, a 40 ns slot duration seems appropriate, since the 40-bytes payload is transmitted in 32 ns, and the additional time can be used for the optical packet header transmission and to provide guard times. For a deeper discussion about this issue the reader is referred to [3].

In order to relieve the complexity of the problem related to the arbitrary mesh topology of the transport network, we assume here that all nodes of the transport network have the same *nodal degree* N_h, that is the same number of adjacent nodes. Since we have selected eight optical nodes in the transport network, we are going to examine here the four network topologies represented in figure 2, where the number N_h of adjacent nodes is N_h = 2, 3, 5, 7. Hence these regular topologies span from the ring network (N_h = 2) up to the full-mesh network (N_h = 7). The other two networks identify topologies with intermediate meshing degree. If we define the connectivity factor α as the ratio between the number of links in the network and that of a full-mesh network, it follows that the eight-node networks in figure 2 are characterized by a connectivity factor α_8 = 0.29, 0.43, 0.71, 1.

3 Node Architecture

The general architecture of a network node is shown in figure 3. It consists of three stages: a first stage of channel demultiplexing, a second stage of switching and a third stage of channel multiplexing. The node is fed by N incoming fibers each having W wavelengths. In the first stage the incoming fiber signals are demultiplexed and G wavelengths from each input fiber are fed into each one of the $N_p = W/G$ second-stage switching planes, which constitute the switching fabric core. Once signals have been switched in one of the parallel planes, packets can reach every output port through multiplexing carried out in the third stage using any of the G wavelengths that are directed to each output fiber. We note that the number of inlets of each third-stage multiplexer varies, depending on the specific structure of the switching planes. Wavelength conversion must be used for contention resolution, since at most G packets can be concurrently transmitted by each second-stage plane on the same output link.

The detailed structure of one of the W/G parallel switching planes is presented in figure 4. It consists of three main blocks: an input *synchronization unit*, as the packet switch-
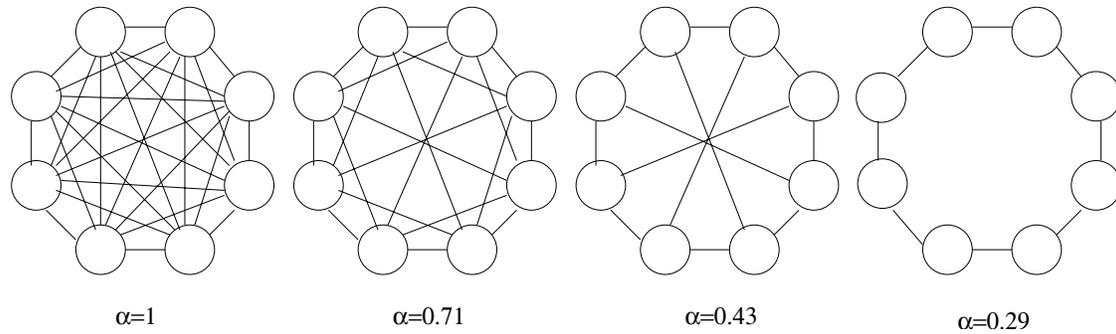
α=1 α=0.71 α=0.43 α=0.29

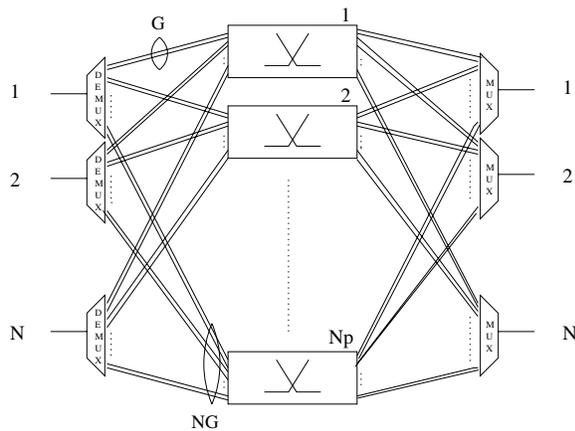**Figure 2. Network topologies**

Figure 3. Optical packet-switching node general architecture

ing is slotted and incoming packets need to be slot-aligned, a *fiber delay lines unit*, used to store packets for contention resolution, and a *switching matrix unit*, adopted to achieve the switching of signals.

These three blocks are all managed by an *electronic control unit* which carries out the following tasks:

- optical packet header recovery and processing;

- managing the synchronization unit in order to properly set the correct path through the synchronizer for each incoming packet;

- managing the tunable wavelength converters inside the switching matrix, in order to properly delay and route incoming packets.

The details of the synchronization and of the fiber delay unit can be found in [3]. We simply recall here that the delay lines are used as an *optical scheduler* that, thanks to the

electronic control, maximizes the number of packets transmitted on the requested output links. Given the maximum achievable delay D_{max} slot, for each switch input $D_{max}+1$ delay lines are needed in each plane, with delays growing from 0 to D_{max}. Once packets have crossed the fiber delay lines unit, they enter the switching matrix stage in order to be routed to the desired output port. This is achieved using a set of tunable wavelength converters combined with an arrayed waveguide grating (AWG) wavelength router [7].

The AWG is used here as it gives better performance than a normal space switch interconnection network, as far as insertion losses are concerned. This is due to the high insertion losses of all the high-speed all-optical switching fabrics available at the moment, that could be used to build a space switch interconnection network. Commercially available 40 channel devices have a channel spacing of 100 GHz and show an insertion loss of less than 7.5 dB [5]. Other proposals of switching nodes based on AWGs can be found in [4][6].

Three different structures are proposed here for the implementation of the switching matrix unit, referred to as structures Basic (B), Enhanced (E) and Optimized (O). In all these structures a *shared buffer* is implemented through recirculation lines in order to enable a much more effective contention resolution. R denotes the number of AWG ports destined to recirculation lines, each one delaying the packets by a fixed amount D_{ric} slot and R_{max} denotes that maximum number of recirculations allowed to a packet in the node.

3.1 Basic structure (B)

The simplest switching matrix (Basic structure), first proposed in [1], is shown in figure 5, referred to a single plane. It consists of $2NG + R$ tunable wavelength converters and an AWG with size $(NG + R) \times (NG + R)$. Only one packet is routed to each AWG outlet and this packet must finally be converted to one of the wavelengths used in the WDM channel, paying attention to avoid contention
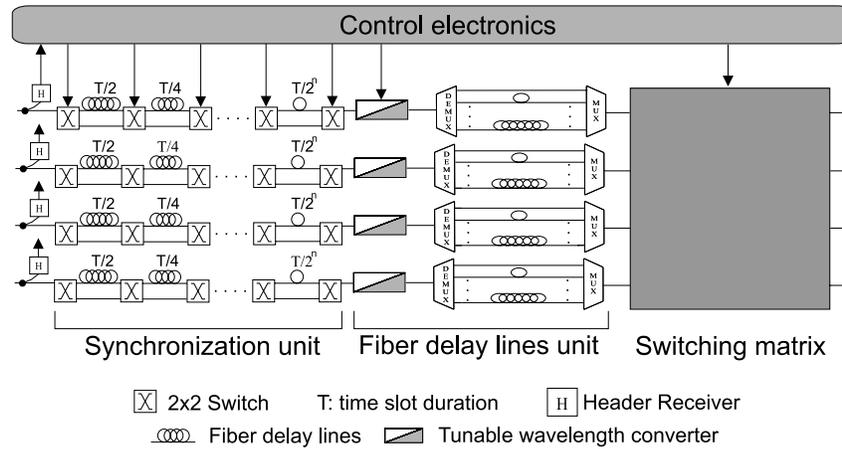
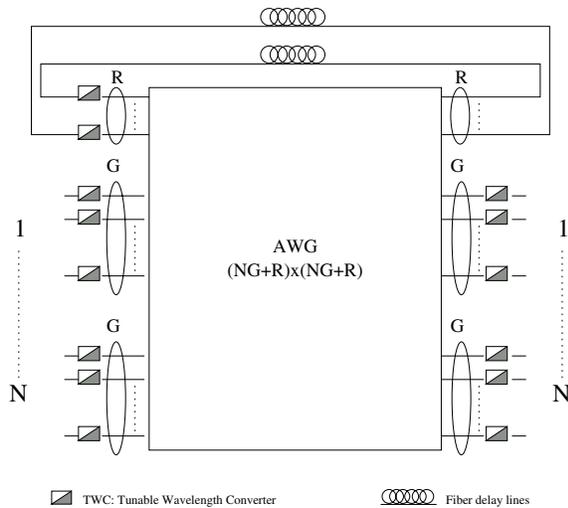**Figure 4. Detailed structure of one of the W/G parallel switching planes**

Figure 5. Basic switching matrix

with other packets of the same channel.

3.2 Enhanced structure (E)

In order to reduce the number of planes of the node and thus to better exploit the "channel grouping" effect (i.e. the sharing of different channels for transmitting a large number of packets, the load per channel being constant), more than one packet can be routed through each AWG inlet; apparently the packets sharing the same input must be transmitted on different wavelengths. The structure of the AWG is such that different wavelengths entering the same input port will emerge on different output ports.

In the Enhanced switching matrix structure illustrated in

figure 6, up to k different packets are sent to the same AWG inlet using different wavelengths. A simple node design requires k to be an integer that divides G. From AWG input port i, the output channel j can be reached by G/k different packets, since there are exactly G/k AWG outlets connected to that channel. During each time slot, up to G packets can be routed to the same AWG outlet using different wavelengths. Hence, demultiplexers are needed to split the different signals and to route them to the last stage of wavelength converters. If $k \leq G/k$, no contention can happen in the multiplexing stage, so this structure behaves exactly as a structure Basic with size $NG \times NG$. On the other hand, when $k > G/k$, events of packet blocking occur, considering the fact that G/k paths are available to reach a tagged output for up to k packets per inlet. So, when more than G/k packets in the same AWG inlet are destined to the same output channel, a contention happens, even if the total number of packets addressed to that output is smaller than G.

3.3 Optimized structure (O)

The Enhanced node structure be simplified by selecting $k = N$, so that each AWG input can receive up to N packets using different wavelengths, thus implying that the node structure includes just one plane ($N_p = 1$). Therefore, the number of AWG inlets is now exactly W. In this structure, shown in figure 7 and referred to as Optimized structure, the last TWC stage isn't needed anymore, provided the employed AWG works on the same wavelengths used in the outgoing fibers. In fact, if the electronic controller takes care of avoiding wavelength contention between AWG outlets connected to the same output channel, packets are ready to be transmitted as soon as they exit the AWG. Therefore, a packet entering the AWG inlet i and destined to the output
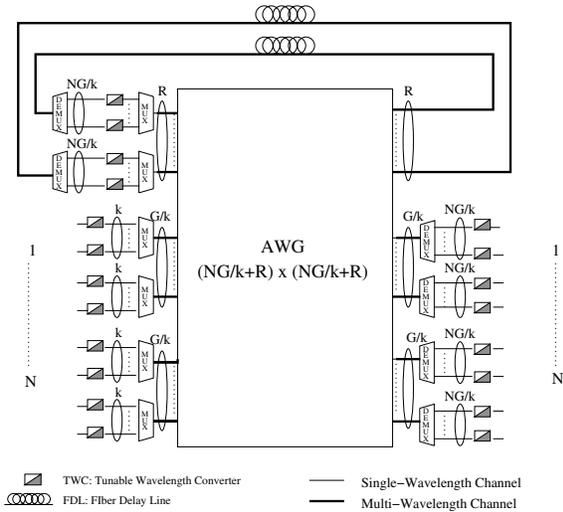
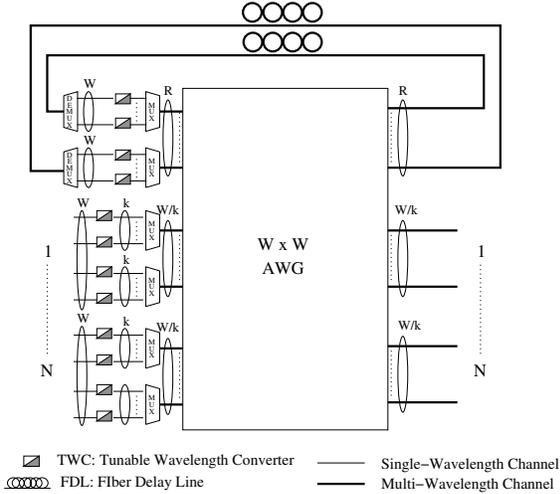**Figure 6. Enhanced switching matrix**

Figure 7. Optimized switching matrix

WDM channel j can not be transmitted using every color in the WDM channel, but only using a subset which consists of the W/N wavelengths through which the packet can reach the desired output channel, thus reducing the benefits of *channel grouping* (when N and W are kept constant).

4 Performance Evaluation

We show now some traffic performance results given by the different node architecture configurations obtained through computer simulation. In order to evaluate the performance provided by the three different switching struc-

tures, we will first examine the packet loss probability in 8-node networks with different connectivity factors. In particular we first consider the case of *shortest path routing*, in which each optical packet must exit the node on a specific output link (either to an adjacent node or to a local edge system). Then we will examine the effect of adopting *deflection routing*, in which an optical packet is forwarded to a link different from that identifying the shortest path if the intended link is already busy in spite of the input buffering possibilities enabled by the node. Such deflection routing is operated choosing random the output link. If the node is equipped also with shared buffering, this last storage possibility is exploited if all the outgoing links are busy. Finally we will examine the effect of varying one single network parameter, that is the number of nodes, the network connectivity factor α and the AWG size.

In order to obtain performance results mutually comparable in spite of the different connectivity factors, we assume that each network node interfaces a number N_h of edge systems (ES) generating the same amount of traffic ρ. In such a way the total number of ESs interfacing a node is equal to the number of network links outgoing from the node and the total (normalized) load offered to the network is ρ.

Traffic levels and shared buffer capacities are assumed equal for all nodes, so that Basic and Enhanced nodes require more planes ($N_p > 1$). The AWG size is set equal in all nodes and this value depends on the network connectivity. The number of wavelengths per plane has been set to $G = 2$ for the Basic node with shared buffer and the number of recirculation lines R equals half of the AWG size (if $R = 0$ only input buffering is exploited by the node). Given the previous assumption about the ESs (the internode links equal the number of local ESs), it follows that we are assuming an AWG size $8N_h \times 8N_h$. Then it follows that the AWG sizes are 16×16, 24×24, 40×40 and 56×56 for the nodal degrees $N_h = 2, 3, 5, 7$, respectively. Furthermore the number of wavelengths W equals the AWG size, since the Optimized node includes just one plane.

The buffering capacity of each node has been set according to the following parameters: $D_{max} = 8$ slot in input buffers, while $D_{ric} = 2$ slot and $R_{max} = 4$ when shared buffering is equipped ($R > 0$).

As far as the offered traffic distribution is concerned, packet interarrivals for each ES wavelength has been modelled as a Poisson process with negative exponential distribution. Based on measurement of real IP traffic [9], the following distribution of packet length L has been assumed

$$\begin{cases} p_0 = Pr(L = 40 \text{ bytes}) = 0.6 \\ p_1 = Pr(L = 576 \text{ bytes}) = 0.25 \\ p_2 = Pr(L = 1500 \text{ bytes}) = 0.15 \end{cases}$$

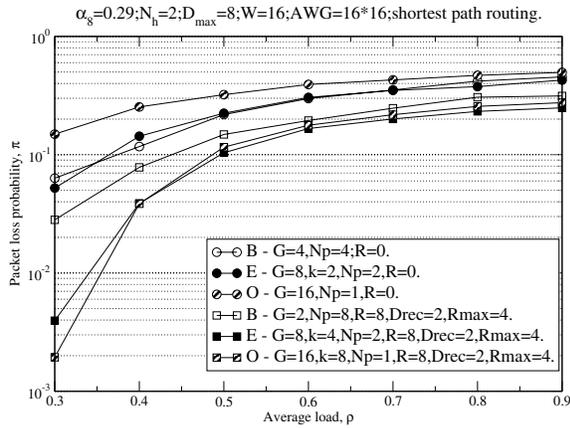so that the resulting average packet length is 393 bytes.

$\alpha_8=0.29; N_h=2; D_{max}=8; W=16; AWG=16*16;$ shortest path routing.

Figure 8. Packet loss performance with shortest path routing and $\alpha = 0.29$

$\alpha_8=0.71; N_h=5; D_{max}=8; W=40; AWG=40*40;$ shortest path routing.
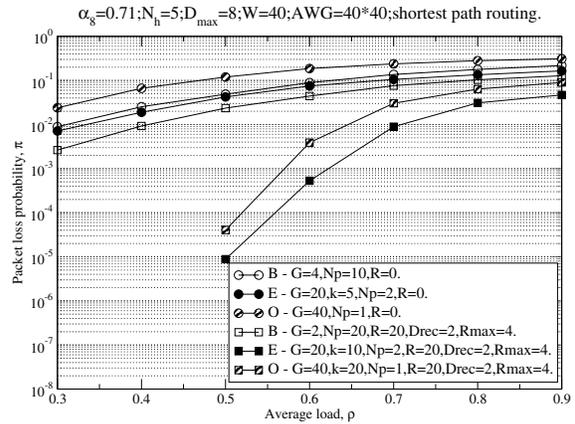
Figure 10. Packet loss performance with shortest path routing and $\alpha = 0.71$

$\alpha_8=0.43; N_h=3; D_{max}=8; W=24; AWG=24*24;$ shortest path routing.
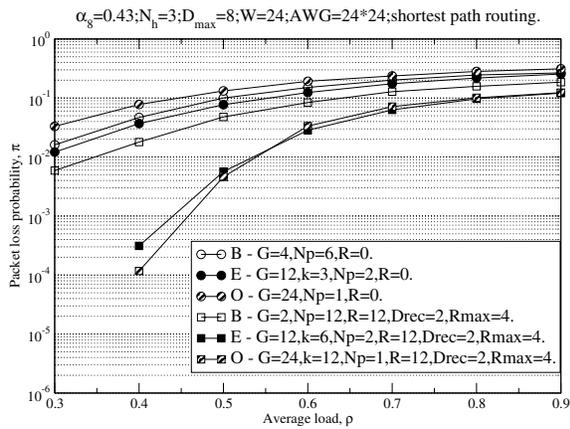
Figure 9. Packet loss performance with shortest path routing and $\alpha = 0.43$

$\alpha_8=1; N_h=7; D_{max}=8; W=56; AWG=56*56;$ shortest path routing.
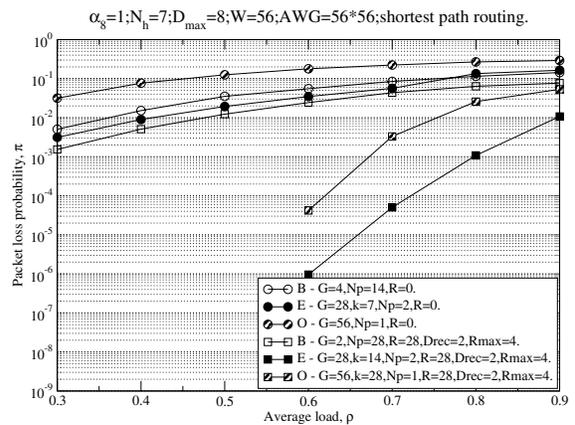
Figure 11. Packet loss performance with shortest path routing and $\alpha = 1$

Packets are assumed to be equally likely to be addressed to any destination ES.

4.1 Shortest path routing

The packet loss probability with shortest path routing is shown in figures 8, 9, 10, 11, for $\alpha_8 = 0.29, 0.43, 0.71, 1$, respectively. We notice that, as we might expect, solutions with shared buffering give better performance than without it. The Optimized node gives the worst performance without shared buffers, the other two solutions providing similar behaviour. An explanation of this is that packets coming from a local ES or a node cannot be routed to the same node outlet (ES or node). So this implies that with the Enhanced and Optimized solutions more contentions arise for accessing each single node outlet compared to a case in which all node outlets are equally addressable, as assumed in [3].

On the other hand when shared buffering is employed, the Enhanced and Optimized structures outperform the Basic one especially for large network connectivity factors. This is due to the fact that the shared buffer is distributed among a much larger number of planes in the former node structures and then becomes less efficient than a larger buffer in a single plane (Optimized solution).
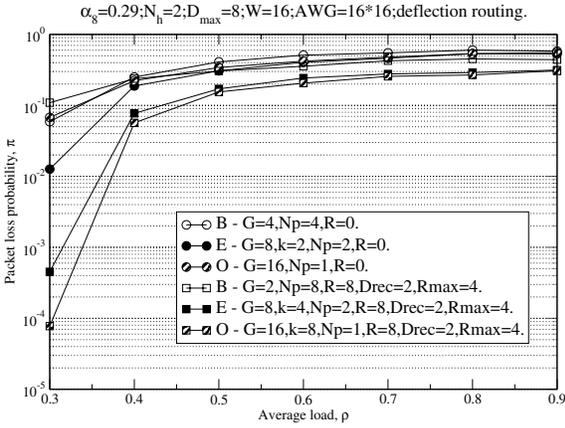
$\alpha_8=0.29;N_h=2;D_{max}=8;W=16;AWG=16*16;$deflection routing.

Figure 12. Packet loss performance with deflection routing and $\alpha = 0.29$

$\alpha_8=0.71;N_h=5;D_{max}=8;W=40;AWG=40*40;$deflection routing.
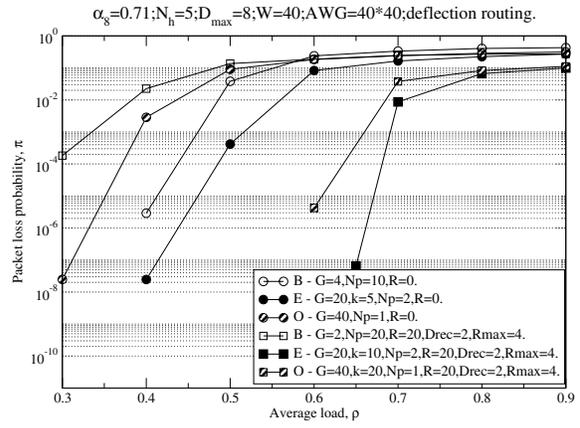
Figure 14. Packet loss performance with deflection routing and $\alpha = 0.71$

$\alpha_8=0.43;N_h=3;D_{max}=8;W=24;AWG=24*24;$deflection routing.
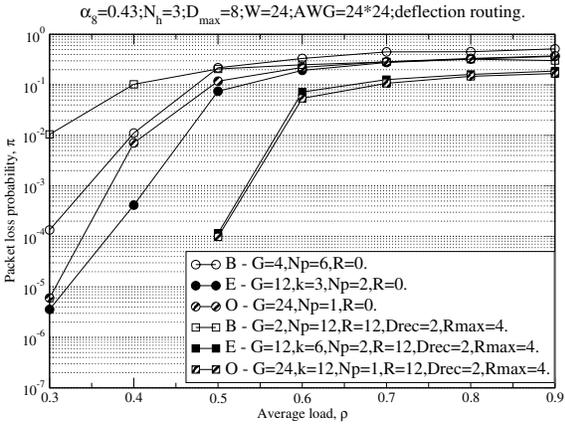
Figure 13. Packet loss performance with deflection routing and $\alpha = 0.43$

$\alpha_8=1;N_h=7;D_{max}=8;W=56;AWG=56*56;$deflection routing.

Figure 15. Packet loss performance with deflection routing and $\alpha = 1$

4.2 Deflection routing

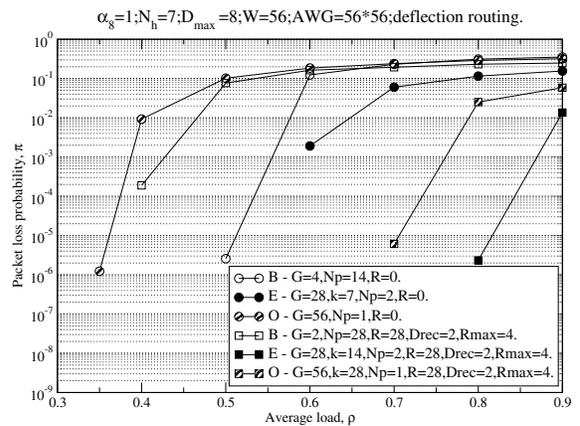The three switching node structures are now compared with deflection routing for the same values of the network connectivity and the corresponding results are shown in figures 12, 13, 14, 15.

It is quite interesting to note that now the network performance improves significantly as the network connectivity grows due to the fact that deflection routing can be exploited better with more output links from the node. When shared buffering is employed Enhanced and Optimized node structures give again the best performance due to the larger buffers they can exploit, as with shortest path routing. Without shared buffering, the Enhanced node gives the best performance, whereas the Optimized node behaves the worst. The reason is analogous to that given for shortest path routing. In fact the load offered to the node outputs is not evenly distributed, since now packets entering the node from local outlets cannot exit the node on the same outlet. Note that now, due to deflection routing, packets entering the node from an upstream node can be routed on any output. Hence it follows that the load offered to inter-node outlets is larger than that received from local outlets of the node. Notice that such behaviour applies to all node structures.

Unlike networks based on Enhanced and Optimized nodes, adding shared buffering with the Basic solution provides worse loss performance. As already observed in [3], shared buffering with Basic node is beneficial only if the (fixed) delay of recirculation lines exceeds a given threshold, for example 16 slots (recall that we have assumed here a delay $D_{ric} = 2$ slots).

5 Conclusions

In this work, we have analyzed and compared the packet loss performance of three types of optical nodes that differ substantially for the switching matrix. We have found that the Enhanced node performs always better than the others without or with shared buffering. In the former case the Basic node provides better results, whereas in the latter case the situation is reversed. Adding shared buffering to a node by keeping the same AWG size (and hence increasing the number of planes with Basic solution or increasing the multiplexing factor in Enhanced and Optimized solutions) provides large benefits with all nodes except in the case of deflection routing and Basic structure. Adopting deflection routing improves significantly the loss performance especially for low traffic loads and high connectivity factors.

References

[1] S. Bregni, G. Guerra, and A. Pattavina. Optical Packet Switching of IP Traffic. In *Proceedings of* 6th *Working Conference on Optical Network Design and Modeling (ONDM)*, 2002.

[2] S. Bregni, G. Guerra, and A. Pattavina. Optical Switching of IP Traffic Using Input Buffered Architectures. *Optical Network Magazine*, 3(6):20–29, 2002.

[3] S. Bregni, A. Pattavina, and G. Vegetti. Architectures and Performances of AWG-based Optical Switching Nodes for IP Networks. *IEEE Journal on Selected Areas in Communications*, 21(7):1113–1121, 2003.

[4] J. Cheyens, J. Jennen, E. V. Breusegem, M. Pickavet, and P. Demeester. Optical Packet Switches Based on a Single Arrayed Waveguide Grating. *Conference on High Performance Switching and Routing (HPSR)*, pages 5–9, Jun. 2003.

[5] S. Electronics. Fiberoptics products: Arrayed Waveguide Grating.
http://www.samsungelectronics.com/fiberoptics/.

[6] L. Li, S. Scott, and J. Deogun. Performances Analysis of WDM Optical Packet Switches with a Hybrid Buffering Architecture. *Conference on Optical Networking and Communications (OptiComm)*, 5285:346–356, Oct. 2003.

[7] C. Parker and S. Walker. Design of Arrayed-Waveguide Gratings Using Hybrid Fourier-Fresnel Transform Techniques. *IEEE Journal on Selected Topics in Quantum Electronics*, 5:1379–1384, 1999.

[8] C. Qiao. Labeled Optical Burst Switching for IP-over-WDM Integration. *IEEE Communications Mag.*, pages 104–114, Sep. 2000.

[9] K. Thompson, G. J. Miller, and R. Wilder. Wide-Area Internet Traffic Patterns and Characteristics. *IEEE Network Magazine*, pages 10–23, Nov. 1997.

[10] M. Yoo and C. Qiao. Just-Enough-Time(JET): A High Speed Protocol for Bursty Traffic in Optical Networks. In *Proc. IEEE/LEOS Tech. for a Global Info Infrastructure*, pages 26–27, Aug. 1997.