OPTICAL PACKET SWITCHING OF IP TRAFFIC

Stefano Bregni, Giacomo Guerra, Achille Pattavina
Politecnico di Milano, Milano, Italy
{bregni,guerra,pattavina}@elet.polimi.it

Abstract To efficiently support the fast-growing demand for transmission capacity, optical packet-switched systems seem to be strong candidates as they allow fast dynamic allocation of WDM channels combined with a high degree of statistical resource sharing. In this work, we propose the architecture of an optical packet switching node, equipped with a shared fiber delay line buffer, for all-optical switching of IP traffic flows. Some traffic simulations results of node operation are also presented.

1. Introduction

In the latest years, telecommunication networks have been demanding an unprecedented, dramatic increase of capacity, fostered mostly by the exponential growth of Internet users and by the introduction of new broadband services. The IP architecture is being seen as the unifying paradigm for a variety of services and for making real the Broadband Integrated Services Digital Network (B-ISDN).

To face this challenge, considerable research is currently devoted to design IP fully-optical backbone networks, in order to relieve the capacity bottleneck of classical electronic-switched networks.

A single optical fibre offers a potentially huge transmission capacity: just in the III wavelength window, something like tens of Terahertz are there to be mined, if only we could be able to exploit such tremendous bandwidth with adequate technology. In the last ten years, optical Dense Wavelength Division Multiplexing (DWDM) has been developed, which made available commercial systems providing impressive transmission capacities: one Terabit per second and per fibre, over distances on the order of 100 km, are feasible nowadays.

Moreover, recently DWDM has evolved to support some network functions as circuit routing and wavelength conversion and assignment. In WDM-routed networks, a wavelength is assigned to each connection in such a way that all traffic is handled in the optical domain, without any electrical processing on transmission. The established connections are called lightpaths: each of them

A. Bianco et al. (eds.), *Next Generation Optical Network Design and Modelling*

occupies only one wavelength per link. The established lightpaths forms the virtual topology, or logical topology, opposed to the network physical topology made of nodes and fibres. Different lightpaths on the same fibre must use different wavelengths.

Unfortunately, today optical devices used in market equipment allow slow switching times, suitable for the circuit routing mentioned above. Therefore, packet switching is still performed by electronics. The extension of optics from transmission and circuit switching to packet switching is thus the second step needed to realize the high-capacity backbone transport infrastructure. In this context, all-optical packet switches play a central role and will be a significant breakthrough on this way. This equipment should allow switching of datagram, variable-length IP-like packets directly in the optical domain, avoiding the need of several optical-electrical-optical conversions. In this work we present the performance evaluation of an input-buffered optical packet switching node, equipped with packet recirculation ports, used as a shared buffer for optical packets. The study of the same node architecture, without packet recirculation ports, was carried out in [1].

The paper is organized as follows. Section 2 briefly reviews the technical literature available in the field of optical packet switching systems. Sections 3 and 4 describe the optical network architecture we envision and the proposed architecture of an optical packet switching node. Traffic performance results attainable with this node are described in section 5. Some conclusions are finally given.

2. Background Literature in Optical Packet Switching

Optical packet switching allows to exploit single wavelength channels as shared resources, with the use of statistical multiplexing of traffic flows, helping to efficiently manage the huge bandwidth of WDM systems. Although many solutions have been proposed to this aim ([2]), basically two approaches can be distinguished: with and without recirculation lines internal to the node.

An architecture for an optical packet switching node without recirculation lines has been developed as a part of the KEOPS project [3]. It is an input-buffered architecture composed of two stages, an optical buffering stage and a switching stage. In the first one contending packets are delayed of a suitable amount of time in order to avoid collisions at the output ports, while, in the second one, they are routed to the correct output fiber.

Since this solution does not allow packet recirculation, it can't efficiently support different packet priorities, because, once a packet has been sent to a delay line, it can't be stored longer than the fiber delay to eventually transmit a new packet with higher priority. This is a crucial shortcoming of this solution, since the need for some methods of providing differentiated classes of service

for Internet traffic is growing, with the explosion of new possible applications. Actually the IPv4 TOS field or the IPv6 Traffic Class field are already used to give packets a particular forwarding treatment at each network node, and the possibility of handle this matter is a fundamental requirement.

An example of an optical node with recirculation lines is the WASPNET switch ([4]). It is composed of two stages, an input and a switching stage. The input stage is used to route packets to the delay lines buffer, for contention resolution, or to the switching stage. The switching stage, then, is used to properly route packets to the desired output port. In both stages an AWG (Arrayed Waveguide Grating) is used to switch packets depending only on their transmission wavelength. This architecture allows packet recirculation, but the need for a second AWG to route packets to their output link yields a considerable hardware overhead. Moreover both of the architectures presented above are used to transmit fixed length optical packets.

The systems presented carry out header processing and routing functions electronically, while the switching of optical packet payloads takes place directly in the optical domain. This eliminates the need for many optical-electrical-optical conversions, which call for the deployment of expensive opto-electronic components.

The transmission of fixed-length optical packets yet implies the development of complex segmentation and reassembly protocols for their generation. The solution we propose in this work is to use variable-length optical packets, in order to directly interface the IP layer with the optical layer, to avoid a heavy packet processing overhead at the optical transport network edges.

3. Optical Transport Network Architecure

The architecture of the optical transport network we propose consists of $M = 2^m$ *optical packet-switching nodes*, each denoted by an optical address made of $m = \log_2 M$ bits, which are linked together in a mesh-like topology. A number of *edge systems* (ES) interfaces the optical transport network with IP legacy (electronic) networks (see Fig. 1).

An ES receives packets from different electronic networks and performs *optical packets* generation. The optical packet is composed of a simple optical header, which comprises the m-bits long 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 flows it receives to their destination electronic networks.

At each intermediate node in the transport network packet headers are received and electronically processed, in order to provide routing information

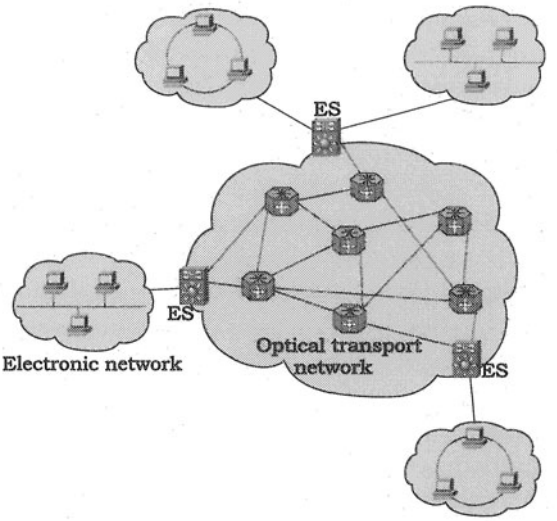

Figure 1. The optical transport network architecture

to the control electronics, which will properly configure the node resources to switch packet payloads directly in the optical domain.

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 *synchronous* (slotted). In the model we propose, the time slot duration, T, is equal to the amount of time needed to transmit an optical packet, with a 40-bytes long payload, from an input WDM channel to an output WDM channel. Supposing a bit rate of 10 Gbps 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.

The operation of the optical nodes is slotted since the behavior of packets, in an unslotted node, is less regulated and more unpredictable, resulting in a larger contention probability.

4. Node Architecture

The general architecture of a network node is shown in Fig. 2. It consists of N incoming fibers with W wavelengths per fiber. The incoming fiber signals are demultiplexed and G wavelengths, from each input fiber, are then fed into one of the W/G switching planes, which constitute the switching fabric core. Once signals have been switched in one of the second-stage parallel planes, packets can reach every output port on one of the G wavelengths that are directed to

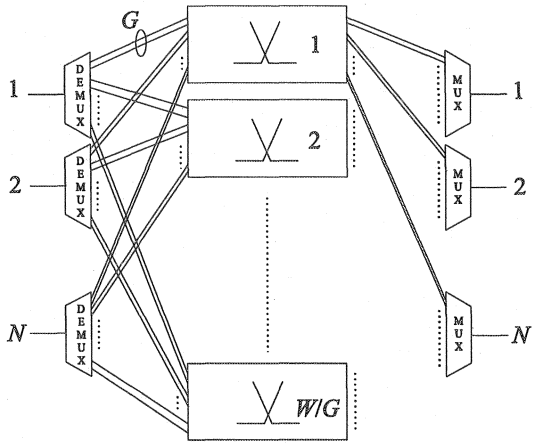

Figure 2. Optical packet-switching node architecture

each output fiber. This allows the use of wavelength conversion for contention resolution, since 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 Fig. 3. It consists of three main blocks: an input *synchronization*

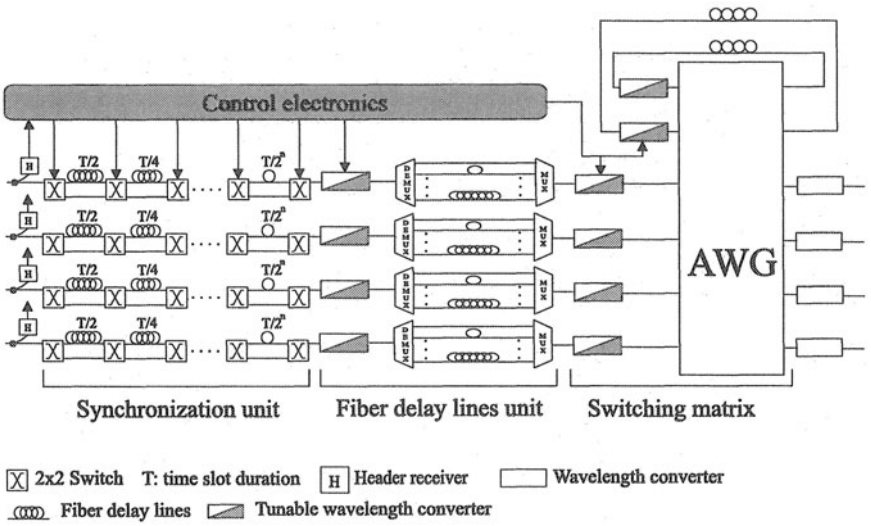

Figure 3. Detailed structure of one of the W/G parallel switching planes

unit, as the node is slotted and incoming packets need to be aligned, a *fiber*

delay lines unit, used to store packets for contention resolution, and a *switching matrix unit*, 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 in order to properly delay and route incoming packets.

The switching matrix is used to switch packets to the output ports or, if necessary, to the recirculation ports, in order to store them for an additional amount of time, to avoid collisions. Moreover recirculation ports allow the switch to support different priority classes, with service preemption. In fact, an optical packet, travelling through a recirculation port delay line, can always be preempted by a higher priority packet and be redirected to a recirculation port, instead of being transmitted.

We will now describe the second-stage switching units mentioned above, detailing their implementation.

4.1. Synchronization Unit

This unit consists of a series of 2×2 optical switches interconnected by fiber delay lines of different lengths. These are arranged in a way that, depending on the particular path set through the switches, the packet can be delayed of a variable amount of time, ranging between $\triangle t_{min} = 0$ and $\triangle t_{max} = (1 - (1/2)^n) \cdot T$, with a resolution of $T/2^n$, where T is the time slot duration and n the number of delay lines.

The synchronization is achieved as follows: once the packet header has been recognized and packet delineation has been carried out, the packet start time is identified and the control electronics can calculate the necessary delay and configure the correct path of the packet through the synchronizer.

Due to the fast reconfiguration speed needed, fast 2×2 switching devices, such as 2×2 semiconductor optical amplifier (SOA) switches [5], which have a switching time in the nanosecond range, must be used.

4.2. Fiber Delay Lines Unit

After packet alignment has been carried out, the routing information carried by the packet header allows the control electronics to properly configure a set of tunable wavelength converters, in order to deliver each packet to the correct delay line to resolve contentions. An optical packet can be stored for a time

slot, with a 40 ns duration, in about 8 meters of fiber at 10 Gbps. To achieve wavelength conversion several devices are available [6], [7], [8].

The delay lines unit was used as an *optical scheduler*, by proper operation of the control electronics. This means that the delay lines are used in order to schedule the transmission of the maximum number of packets onto the correct output link. This implies that an optical packet P_1, entering the node at time αT from the i-th WDM input channel, can be transmitted after an optical packet P_2, entering the node on the same input channel at time βT, being $\beta > \alpha$. For example, suppose that packet P_1, of duration $l_1 T$, must be delayed of d_1 time slots, in order to be transmitted onto the correct output port. This packet will then leave the optical scheduler at time $(\alpha + d_1)T$. So, if packet P_2, of duration $l_2 T$, has to be delayed for d_2 slots, it can be transmitted before P_1 if $\beta + d_2 + l_2 < \alpha + d_1$ since no collision will occur at the scheduler output.

4.3. Switching Matrix Unit

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 [9].

The AWG is used 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. Moreover AWG routers are strictly non-blocking and offer high wavelength selectivity. Commercially available 40 channel devices have a channel spacing of 100 Ghz and show a typical insertion loss of 7.5 dB.

As we said before, to improve the system performance and to eventually support different priority classes with service preemption, some of the AWG ports are reserved to allow packet recirculation (see Fig. 3). To this purpose R AWG output ports are connected, via fiber delay lines, to R input ports. Packet recirculation is then managed using tunable wavelength converters.

After crossing the three stages previously described, packets undergo a final wavelength conversion, to avoid collisions at the output multiplexers, where W WDM channels are multiplexed on each output link.

5. Simulation results

In this section, we present some simulation results of the operation of an optical transport network node with a single switching plane, with an 8×8 and a 16×16 AWG (see Figs. 4 and 5)

These results have been obtained assuming that the node receives its input traffic directly from N edge systems. The edge systems buffers capacity is

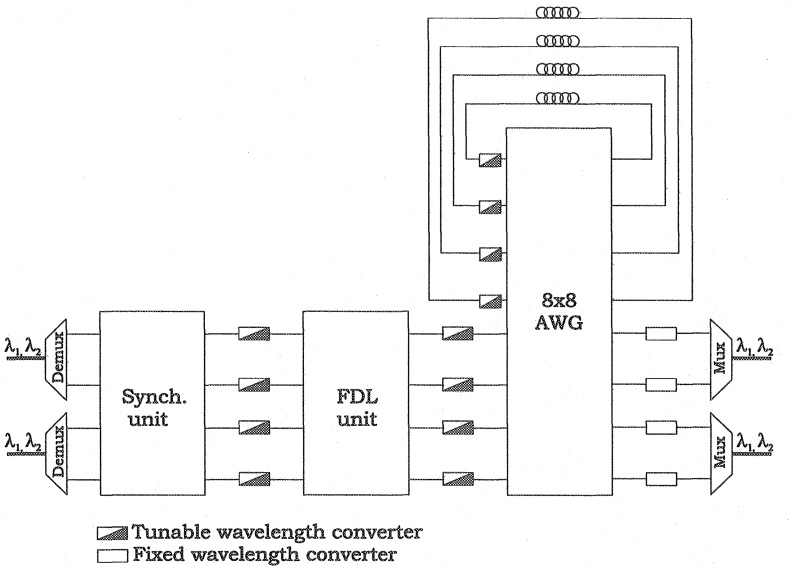

Figure 4. Optical node with an 8 × 8 AWG, R=4.

supposed to be large enough to make packet loss negligible. Each WDM channel is supposed to have a dedicated buffer in the edge system. Moreover, since the node has a single switching plane, W always equals; the value of G.

The packet arrival process has been modelled as a Poisson process, with packet interarrival times having a negative exponential distribution. As the node operation is slotted, the optical packets duration was always assumed to be multiple of the time slot duration T, which is equal to the amount of time needed to transmit an optical packet, with a 40-bytes long payload, from an input WDM channel to an output WDM channel.

As far as packet length is concerned, the following probability distributions was considered:

1 *Empirical distribution.* Based on real measurements on IP traffic [10], [11], we assumed the following probability distribution for the packet length L:

$$\begin{cases} p_0 = P(L = 40 \text{ bytes}) = 0.6 \\ p_1 = P(L = 576 \text{ bytes}) = 0.25 \\ p_2 = P(L = 1500 \text{ bytes}) = 0.15 \end{cases} \quad (1)$$

In this model, packets have average length equal to 393 bytes. Since a 40-bytes long packet is transmitted in one time slot of duration T, the average duration of an optical packet is approximatively $10T$. Moreover,

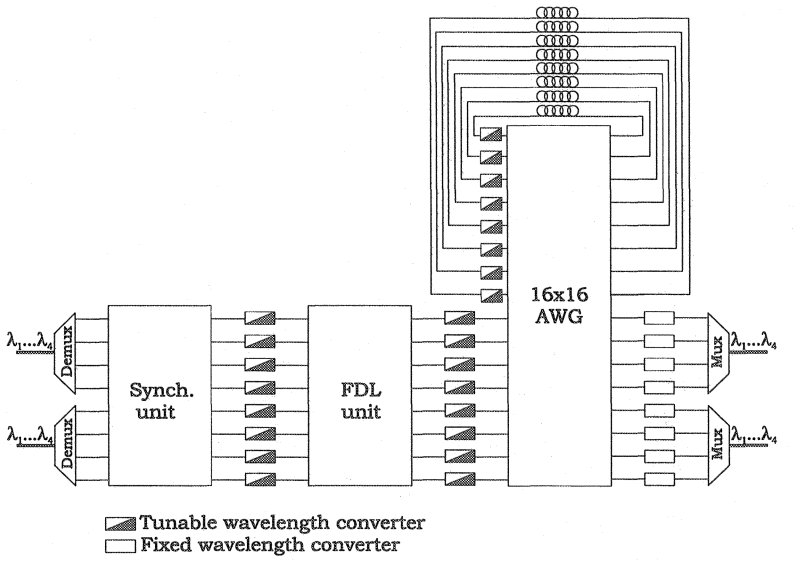

Figure 5. Optical node with a 16 × 16 AWG, R=8.

p_0, p_1 and p_2 represent the probability that the packet duration is T, $15T$ and $38T$ respectively.

In this system a packet is supposed to be lost if it can't be delayed of a suitable amount of time, in order to be transmitted onto the correct output port, on any of the G available wavelengths.

Two different structures, for the recirculation delay lines, were tested: the constant delay recirculation (CDR) and the variable delay recirculation (VDR). Multiple recirculation of a packet are allowed only if the packet duration LT is lower than the recirculation delay, to prevent long packets to occupy more than one recirculation port at a time.

In the CDR structure all the recirculation ports delay each packet of the same amount of time, $D_{rec} = kT$, while in the VDR structure D_{rec} doubles every two ports. The first couple of ports will then have a recirculation delay of T, the second couple of $2T$, and so on.

Figure 6 shows the packet loss probability for the 16 × 16 AWG, R=8, by comparing the CDR with $D_{rec} = T$ and VDR structures, for different values of the maximum delay achivable in the fiber delay lines stage, D_{max}. The VDR structure always gives a better performance than the CDR structure since, as the maximum delay of its recirculation ports is $8T$, packets can be buffered longer to avoid collisions.

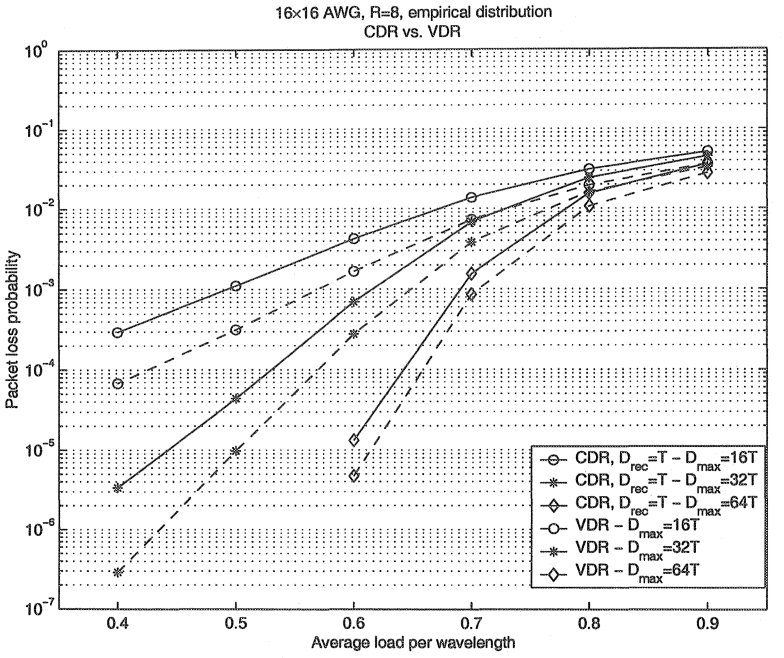

Figure 6. Packet loss probability of 16 × 16 AWG, R=8: CDR, $D_{rec} = T$ vs. VDR.

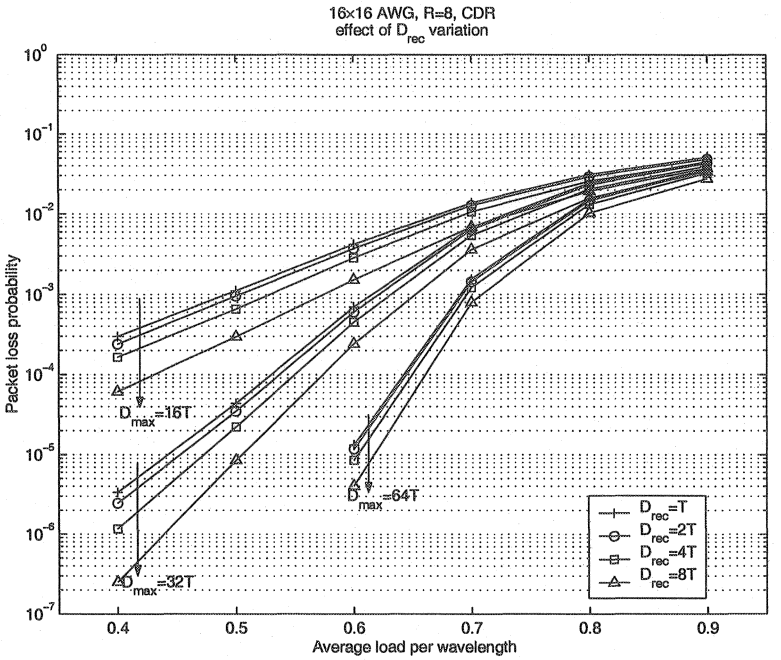

Figure 7. Packet loss probability of 16 × 16 AWG, R=8: effect of D_{rec} variation.

The effect of the maximum recirculation port delay can also be seen in Fig. 7, where the packet loss probability for the 16 × 16 AWG, R=8 CDR is shown at different traffic loads per wavelength, for different values of D_{rec}. It can be seen that the switch performance is improved as D_{rec} grows, since, as we said before, packets can be stored longer. However, the lower loss probability yields an increase of the average packet delay in the node, as shown in Fig. 8, which is noticeable only for high levels of the offered load. In general we observe that, due to the recirculation process, the average packet delay grows almost linearly with the offered load.

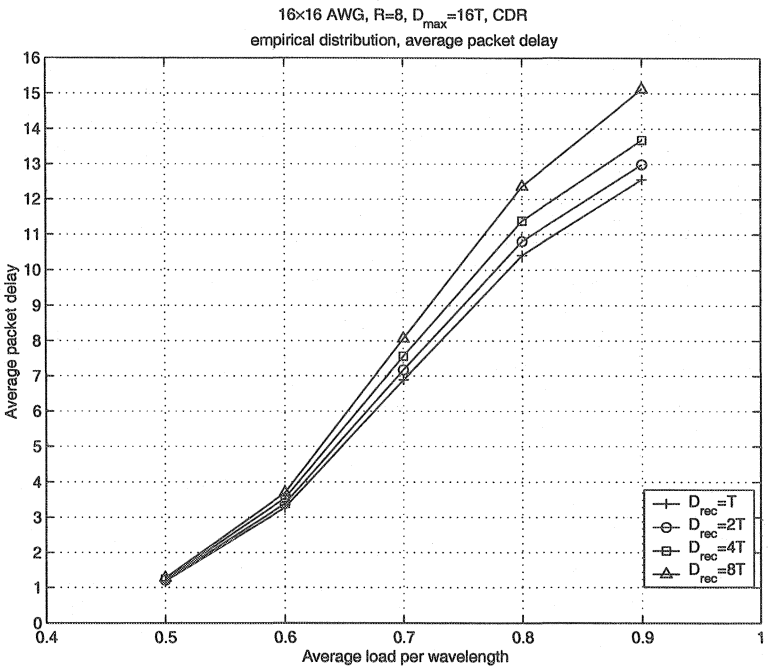

Figure 8. Average packet delay for a 16 × 16 AWG, R=8, CDR.

Figure 9 plots the packet loss probability, at different traffic loads per wavelength, comparing an 8 × 8 AWG, R=4 with a 16 × 16 AWG, R=8. Thus the n/R ratio is constant, being n the AWG dimension. The first solution reduces the hardware complexity but gives a higher packet loss probability. This is due to the reduction of the grouping factor G, which leads to a reduction of the number of packets that can be transmitted, at the same time, on the same output port.

Finally, to show a comparison with the empirical distribution described above, we have modelled the optical packet length as a stochastic variable,

uniformly distributed between 40 bytes (duration T) and 760 bytes (duration $19T$). Also in this model, packets have average duration of $10T$. Figure 10 plots a comparison between the empirical and uniform distribution packet loss.

It can be seen that the traffic performance of this node is influenced by the maximum optical packet length L_{max}, as in the case of the node without recirculation ports (see [1]), regardless of packet length distribution. In fact, for the same value of D_{max}, uniformly distributed packets show a lower loss probability, as they have a maximum length of $19T$, being $38T$ the maximum packet length of the empirical distribution.

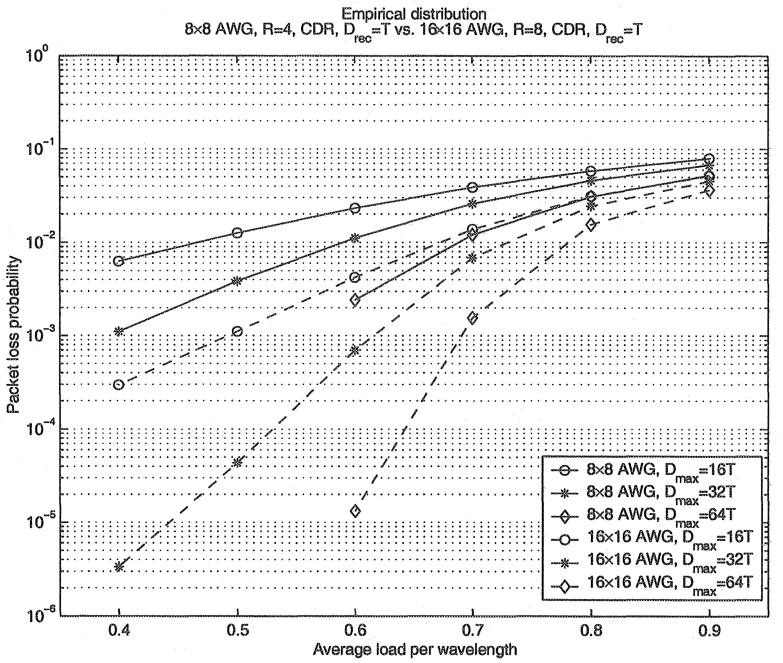

Figure 9. Effect of the grouping factor reduction.

6. Conclusions

In this work, we have proposed the architecture of an input-buffered optical node with packet recirculation ports. The simulation results show that the system performance (packet loss probability) is improved as the delay enabled by recirculation ports increases.

Yet, the average packet delay in the node also grows proportionally with the offered load. A trade-off between low packet losses and high end-to-end delays is then required, especially for real-time applications.

It has been also shown that a reduction of hardware complexity yields a higher loss probability if the grouping factor G is reduced.

Many issues will have to be addressed in the future, such as the detailed study of the improvement attainable with the use of recirculation ports upon the introduction of different priority classes, with service preemption. Moreover, the behavior of an optical transport network, as a whole, will have to be investigated since a single node operation has been simulated for this work.

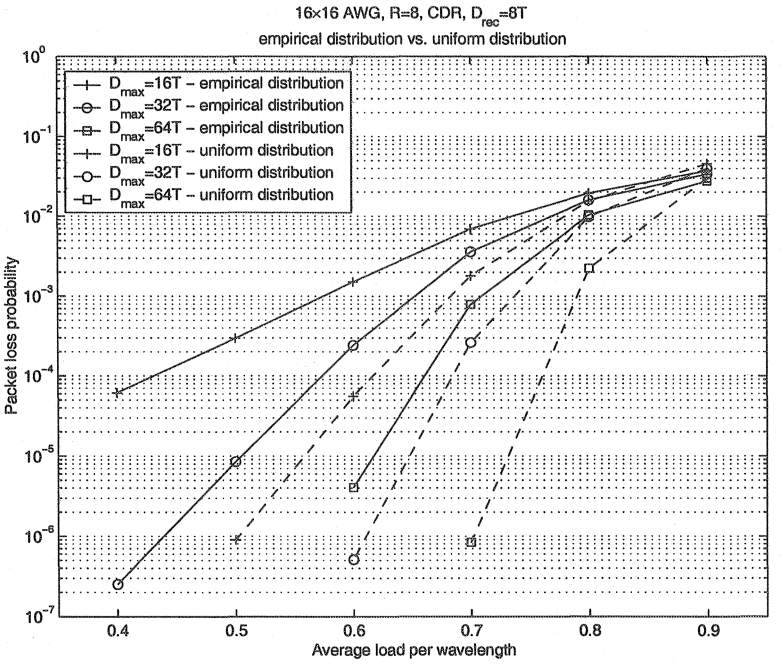

Figure 10. 16×16 AWG, R=8, $D_{rec} = 8T$, CDR: empirical distribution vs. uniform distribution

References

[1] Bregni, S., Guerra, G., Pattavina, A.:Optical Packet Switching for IP-over-WDM Transport Networks, Proc. of 2001 Int. Workshop on Digital Commun., Taormina, Sept. 2001, pp. 10-25.

[2] Xu, L., Perros, H.G., Rouskas, G.:Techniques for Optical Packet Switching and Optical Burst Switching. IEEE Commun. Mag. (Jan 2001) 136-142

[3] Renaud, M., Masetti, F., Guillemot, C., Bostica, B.: Network and System Concepts for Optical Packet Switching. IEEE Commun. Mag. (Apr. 1997) 96-102

[4] Hunter, D.K. et al.: WASPNET: A Wavelength Switched Packet Network. IEEE Commun. Mag. (Mar. 1999) 120-129

[5] Dorgeuille, F., Mersali, B., Feuillade, M., Sainson, S., Slempkès, S., Foucher, M.: Novel Approach for Simple Fabrication of High-Performance InP-Switch Matrix Based on Laser-Amplifier Gates. IEEE Photon. Technol. Lett., Vol. 8. (1996) 1178-1180

[6] Tzanakaki, A., O'Mahony, M.J.: Analysis of Tunable Wavelength Converters Based on Cross-Gain Modulation in Semiconductor Optical Amplifiers Operating in the Counter Propagating Mode. IEE Proc.-Optoelectron., Vol. 147., No.1, (Feb. 2000) 49-55

[7] Mak, M.W.K., Tsang, H.K.: Polarisation-insensitive Widely Tunable Wavelength Converter Using a Single Semiconductor Optical Amplifier. IEEE Electron. Lett., Vol. 36. (2000) 152-153

[8] Tzanakaki, A. et al.: Penalty-Free Wavelength Conversion Using Cross-Gain Modulation in Semiconductor Laser Amplifiers with no Output Filter. Elec. Lett., Vol. 33. (1997) 1554-1556

[9] Parker, C., Walker, S.D.: Design of Arrayed-Waveguide Gratings Using Hybrid Fourier-Fresnel Transform Techniques. IEEE J. Selec. Topics Quant. Electron., Vol. 5. (1999) 1379-1384

[10] Thompson, K., Miller, G.J., Wilder, R.: Wide-area Internet Traffic Patterns and Characteristics. IEEE Network, Vol 11. (1997) 10-23

[11] Generating the Internet Traffic Mix Using a Multi-Modal Length Generator. Spirent Communications white paper. http://www.netcomsystems.com