Chapter 4

Optical Burst Switching and Optical Packet Switching

Steinar Bjørnstad, Christoph M. Gauger, Martin Nord (editors), Elise Baert,
Franco Callegati, Davide Careglio, Walter Cerroni, Jan Cheyns, Didier Colle,
Piet Demeester, Chris Develder, Dag R. Hjelme, Gabriel Junyent,
Mirosław Klinkowski, Marian Kowalewski, Marko Lacković,
Marian Marciniak, Mario Pickavet, Carla Raffaelli, Josep Solé-Pareta,
Norvald Stol, Erik Van Breusegem, and Paolo Zaffoni

4.1 Optical Burst Switching and Optical Packet Switching Concepts

Most existing wide area telecommunication networks (WAN) have an SDH based, electronically circuit switched transport core. Connection set-up or tear down may require days or weeks and multiplexing/demultiplexing always require complex optical/electro/optical (O/E/O) conversions. Nowadays, the operators and vendors are working on an optical control plane, which should control set-up and tear down of connections. Work on automatically switched optical network (ASON) and generalised multi-protocol label switching (GMPLS) takes place within ITU and IETF, respectively. Resulting optically circuit switched (OCS) networks can offer explicit transfer guarantees, since circuit establishments are confirmed. However, this generates a delay equal at least to the round-trip time, typically several ms. Even though OCS networks will offer more flexibility than today's solution, the access to the optical bandwidth will still be provided with fibre/wavelength granularity.

Future networks should be able to serve a client layer that includes packet-based networks, such as the Internet, which may have a highly dynamic connection pattern with a significant portion of bursty traffic between the communicating pairs. In this case, OCS transport may not be flexible enough. It would require over-dimensioning of the number of connections and of the bandwidth reservation of each connection, to avoid excessive delay and extensive buffering at the ingress router. Here is when Optical Packet Switching (OPS) and Optical Burst Switching (OBS) come into play, with the goal of reducing delays and improving the utilisation of the network's resources through statistical multiplexing. This comes at the expense of not being able to offer explicit transfer guarantees. However, suitable node design and proper dimensioning of network resources may enable support of most services over the same network. OPS and OBS logical performance in relation to contention resolution and quality of service (QoS) differentiation will be discussed in 4.3 and 4.4. Moreover, hybrid schemes are possible where OPS and OBS share the WDM layer with an OCS scheme, serving applications with the need for explicit transfer guarantees.

Europe has been very active in OPS research, especially through projects like RACE ATMOS [1], ACTS KEOPS [2], IST DAVID [3], IST STOLAS [4] and COST 266. OBS was proposed very recently [5], but different research groups in Europe have already made significant contributions to this field, as later sections in this chapter show.

Inevitably, there are some differences in terminology within the research community; we here explicitly describe some concepts and terms used in this chapter. Both *optical packet switching* and *optical burst switching* are based on the idea of separating forwarding from switching in the network nodes. Forwarding decisions are taken by means of a burst control packet (BCP) or packet header that undergoes O/E conversion and electronic processing at the nodes. On the other hand, the burst/packet payload is optically switched, thus avoiding the costly O/E/O conversion, and in principle simplifying the interface cards to the optical fibres. OBS and OPS, as intended here, are hence different from *all-optical* packet- and burst switching approaches, where the headers are processed in the optical domain, thereby controlling a switch. Due to their increased optical complexity, we consider that such concepts are further away from implementation.

Commercial transponders performing O/E/O conversions are optimised for a specific signal bitrate and transmission format. Networks based on optical switching may avoid transponders, which opens up for network *transparency*, here meaning design of a network that readily handles any signal format and

bitrate. This is often cited as an attractive property of optical switching, together with the potential for high bit-rate systems (40 Gbit/s and above). Nevertheless, realisation of such *fully transparent* networks requires a high number of adaptive components, as discussed in [6]. There is hence a trade-off between the flexibility and the complexity/cost in network design. S*emi-transparent* networks, here meaning optical networks optimised for a certain signal format and bitrate may therefore be attractive. Furthermore, such a network, due to the fixed signal format, may allow O/E/O conversion, e.g. to perform 3R regeneration, wavelength conversion and buffering; whenever a function is less costly to perform in the electronic domain. A comparison of optical and electronic buffering is made in section 4.3.

4.1.1. Packet/Burst Handling Schemes

Both in OBS and in OPS the switching matrix is required to be reconfigurable on the burst/packet time scale to allow the packets/burst to efficiently share node and fibre resources. As will be discussed in 4.2, the switching operation is hence more demanding in OPS than in OBS. The basic principles for switching architectures and functionalities are independent of the packet/burst handling scheme [6]. In Figure 4.1 we report the four potential packet/burst handling schemes, classified according to synchronisation and size of the data units.
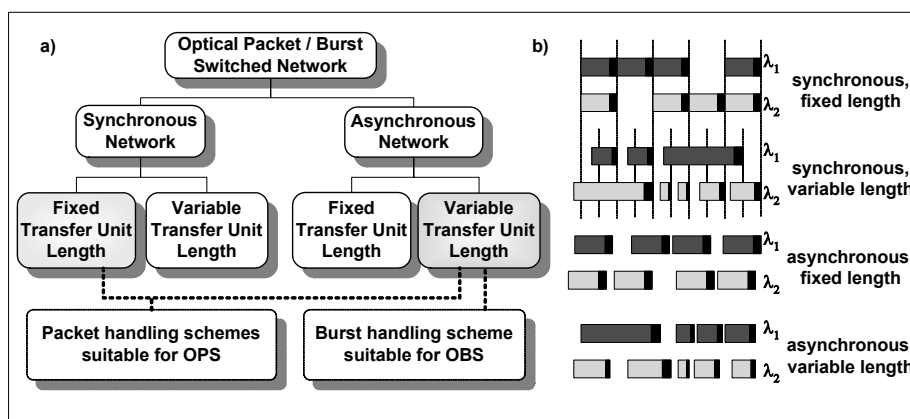

Figure 4.1. - Potential packet/burst handling schemes in OPS and OBS

Figure 4.1 shows that asynchronous and variable length data units are considered more suitable for OBS. This is motivated by decreased complexity at the optical layer, and OBS will also benefit from a large degree of freedom in the burst assembly mechanism. This is the case most widely studied in the literature and to which we refer in the following.

On the other hand, the best scheme for OPS is a matter of debate. The most studied case is that of synchronous, fixed length packet (FLP) [1][7][8], but more recently, work on variable length packets (VLP) have also been studied, with both asynchronous [9][10][11][12][13] and synchronous operation (typically *trains* of packets) [14]. We here discuss aspects relevant for the choice of packet handling scheme in OPS.

Synchronous operation requires optical synchronisers at the switch interfaces and a global network clock for practical realisation. Using FLP requires fragmentation of client packets and padding to fill the optical packet. Both factors increase the overhead, i.e. bandwidth consumption for successful transmission through the OPS network (c.f. 4.2.2), and thus the blocking ratio for the same client load. In addition, fragmentation calls for reassembly of client packets, which increases egress node complexity. Fibre delay line (FDL) based buffer design and management is simpler for FLP than for VLP, and it is simpler to maintain the packet order. The complexity of other contention resolution methods may be independent of the packet length, as discussed in 4.3. In general, a switch matrix operating in synchronous, FLP mode will have less contention than when operating in asynchronous mode [15]. Furthermore, in this mode, a re-arrangeably non-blocking may have equal performance to a strictly non-blocking switch matrix [16].

In most cases, the synchronous, FLP mode or the asynchronous, VLP mode is considered to be the better choice for OPS. This is since the former minimises blocking, enables simpler FDL based buffers and only requires re-arrangably non-blocking switch matrices. The latter has less overhead and avoids the need for a network global clock, optical synchronisation and reduces edge node assembly time. However, at a given technology status, and in a given network context (client layer characteristics and service requirements), a performance and cost evaluation is needed to answer which packet handling scheme is more suitable.

4.1.2. OPS/OBS Properties and Comparison

We consider OPS and OBS for application in a mesh-based WAN or "core network", context. Mesh networks minimise switch matrix sizes and propagation distance, at the same time enabling flexible load balancing and link-protection, whilst avoiding single-points of failure. Ring based OPS for metro area networks (MAN) are considered in 4.5. Considering transmission systems, high-capacity WDM systems are considered, and the interplay between channel count and network performance is discussed in section 4.3.

Node functionalities

The network nodes combine edge router and core router functionalities (Figure 4.2). Edge (ingress/egress) routers modules are capable of communicating with the electrical layer, and may also form the interfaces to adjacent OPS/OBS layers. The main functionalities of an ingress router are to aggregate electrical packets into optical packets/bursts as defined by the assembly algorithm. Note that the aggregation process at an ingress node will shape the traffic flowing into the core, which may be beneficial to network performance. To some degree, aggregation may reduce the degree of self-similarity [17][18], although not at larger time-scales [19]. Its routing table and scheduling algorithm determine what control information to encode and when to send the optical packets/bursts. The egress router reads the optically encoded control information, defragments arriving packets/bursts, and forwards the client packets, after reassembly if required, to the client layer.
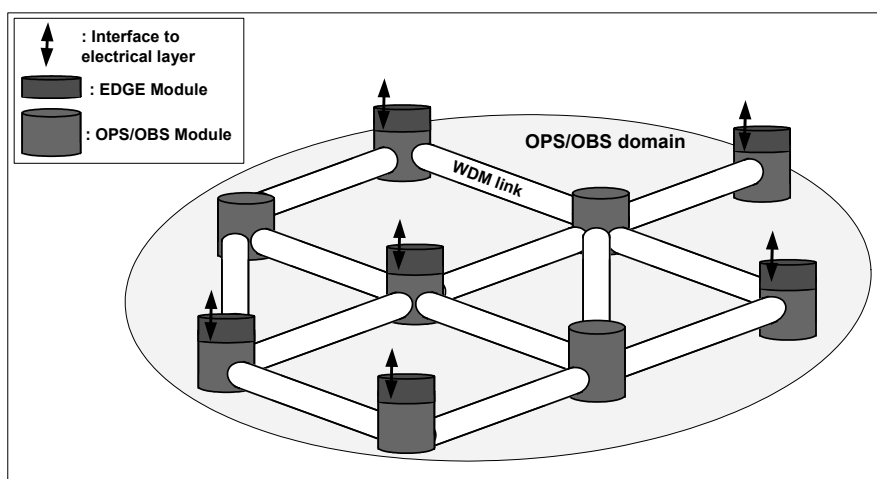

Figure 4.2. - Network consisting of nodes with edge and core routers functionalities

A core router module reads the control information after O/E conversion. It then processes this information electronically to configure the switch matrix and contention resolution resources. When required, control information should be re-written, otherwise the old header continues with the packet.

Comparison of concepts

OPS and OBS concepts were compared in [6][13][20]; these studies identified some fundamental differences, briefly discussed here. Differences in node realisations, contention resolution and QoS differentiation are discussed in the next chapters.

Most OPS proposals assume packet durations in the 0.1-1 μs range. We consider client packets representative of today's Internet traffic, which is in the 40 to 1500 bytes length [21]. In that case, OPS typically requires a very low degree of client layer packet assembly, at a bitrate of 10 Gbit/s. However, the level of aggregation needed for the same packet duration, increases with the bitrate. The payload length and relation to packet format has an impact on the node design, as discussed in section 4.2. Furthermore, control information is typically encoded "in-band". Here, in-band means that the header is transmitted on the same wavelength and either overlaps the payload in time (as for sub-carrier modulation), or it is transmitted just ahead of the payload (serial header).

OBS assumes more extensive burst aggregation, to realise bursts with payloads typically carrying tens of kbytes. OBS uses out-of-band encoding of control information, which is realised by sending BCPs on a common BCP wavelength on that link. Each BCP is transmitted ahead of its corresponding burst, and contains the timing information on the burst. The BCP information and the time offset, in combination with delayed reservation (DR) principle, enable advanced burst scheduling. This can optimise bandwidth usage and enable QoS differentiation methods as discussed in 4.3.

4.1.3. Summary

OPS and OBS are optical transport networks architectures that have a finer granularity than OCS and are expected to better support dynamic traffic patterns. One main advantage is the improved utilisation of network resources that can be expected. OBS assumes asynchronous operation with variable burst length, whilst OPS typically operate in either synchronous FLP mode or in asynchronous VLP mode. Different packet and burst granularities, control encoding and appropriate scheduling mechanisms are hence the main differentiators of OPS and OBS concepts. We develop these aspects in later sections, and discuss their impact when it comes to node design (section 4.2), contention resolution (section 4.3) and QoS differentiation mechanisms (section 4.4). Application of the OPS concepts to a metro environment is discussed in section 4.5. In the last section (4.6), we draw a conclusion based on this study.

4.2 Design and Analysis of OPS Nodes

This section describes the main functionalities of OPS and OBS nodes and makes some design considerations. A brief overview of technology status is

given before presenting OPS node architectures proposed within COST266. In general, the level of details in work on node design varies greatly. Some work is closer to implementation and considers e.g. realisation of clock recovery and control unit, whilst other focus on packet/burst switch architectures to ensure correct forwarding, without details of realisation.

4.2.1. OPS and OBS Nodes

The OPS/OBS node design space is broad. However, in the general case, an OPS or OBS node consists of four main building blocks. These are illustrated in Figure 4.3, through the example of an OPS node in slotted operation.
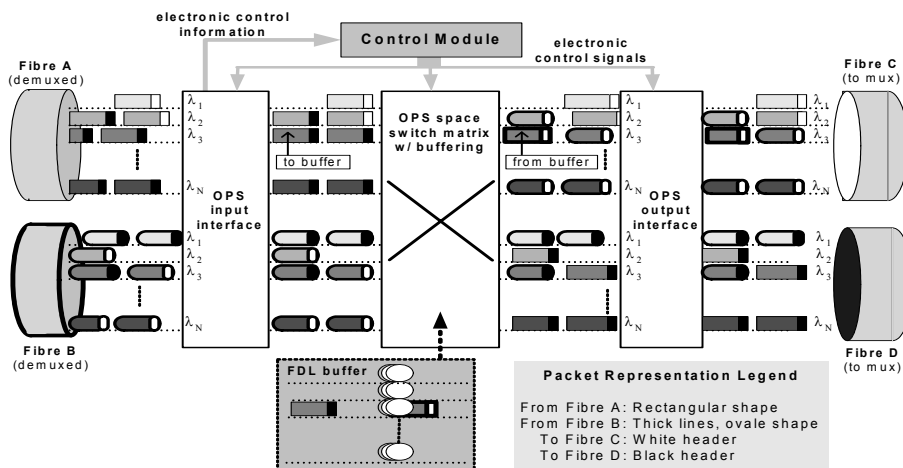

Figure 4.3. - Generic OPS node with FDL buffers in slotted operation. Packets may be followed as they transit through the switch

The input interface detects a preamble/synchronisation pattern, marking the packet arrival and enabling clock recovery. It monitors incoming signals and conditions the incoming packets as required, e.g. through power equalisation and alignment of packets in slotted operation. Furthermore, it retrieves control information, encoded in the packet headers or in the BCPs, and transmits it to the control unit in electronic form. To accommodate processing delay, packets may be delayed using fixed length FDLs. The electronic control unit makes look-ups in the routing table, and is responsible for implementing the scheduling policy. It considers the control information and current resource configurations to identify a suitable output port/wavelength, or buffer allocation. If needed, the control unit identifies new control information and updates timing information (in OBS). The switch matrix, buffers and interfaces respond to the electronic reconfiguration signals from the control unit. The switch matrix

influences the node performance by its switching time, maximum throughput, internal blocking properties and signal degradation. The output interface updates the headers or BCPs and conditions the signal by e.g. 3R regeneration, if required.

4.2.2. General Design Considerations

The technology required to realise OPS/OBS nodes is currently not mature. Therefore cost considerations are hard to make at present; developments within technology are described in 4.2.3. We will in this paragraph focus on relevant performance parameters and bandwidth efficiency.

The OPS/OBS network should be designed to offer its client layers throughput in the Tbit/s range to meet expected increases in traffic. The throughput of an internally strictly non-blocking switch matrix is the capacity switched per time unit in the ideal case, when input and output channels are fully loaded. In realistic network scenarios, however, external blocking or *contention* occurs, leading to loss of packets/bursts, which depends on the packet/burst switch architecture/designs, including contention resolution resources. Typically, simulations are used to study different designs' logical performance as a function of node degree, load and traffic characteristics. In addition to acceptable loss rates, the packet/burst switch should support QoS differentiation, as discussed in section 4.4. The delay of an OPS/OBS node is in most cases negligible compared to the transmission delay in the network. However, if jitter caused e.g. by buffering is significant, it may lead to reordering of packets/bursts, which increases egress node complexity.

In a packet/burst switched network paradigm, each network layer encapsulates higher layer packets, thereby adding an overhead. In the OPS/OBS layer, successful processing and switching typically dictates packet fields for control information, synchronisation pattern(s) and optical guard bands (OGBs). OPS/OBS networks should be designed to handle a certain load offered by the client layer, but the overhead creates a need for the OPS/OBS network to actually be designed for a higher *optical* load.
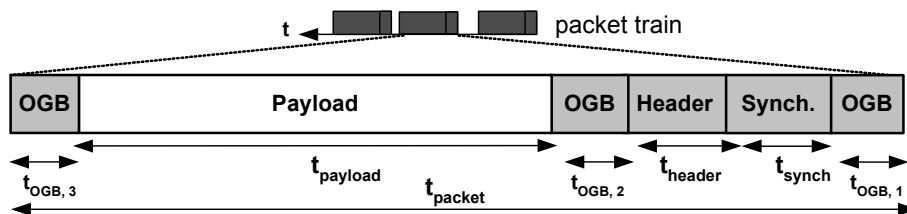

Figure 4.4. - Example of optical packet format

Considering a single node, the overhead, defined in equation 4.1, describes what proportion of time the switch matrix spends settling the switch and transmitting non-payload fields relative to the payload duration. The durations of these fields are technology dependent. E.g. the header duration depends on the header encoding method, and the packet synchronisation/preamble field must contain a pattern long enough to allow a stable clock-recovery with unambiguous start-of-packet detection. OGBs are required to accommodate jitter in e.g. the header insertion process and between packets in a packet train. In addition comes the switching time, during which the considered switch matrix path cannot be exploited.

$$Overhead_{OPS} = \frac{t_{packet} + t_{switch} - t_{payload}}{t_{payload}} = \frac{\left(t_{header} + t_{OGB} + t_{synch} + t_{switch}\right)}{t_{payload}} \qquad (4.1)$$

Even with 1 ns switching time, it is very hard to limit overhead to e.g. 10 % at 10 Gbit/s payload bitrate, when payload length is in the lower region of typical distribution of IP packet length. The interplay between packet format and switching time on overhead is further discussed in [22]. To reach this threshold typically requires some aggregation at the OPS ingress nodes, to minimise the fraction of optical packets with payloads in the 40-200 bytes range. For payloads around 1500 bytes, switching times around 100 ns can in general be accepted.

In OBS, each burst on a data wavelength needs a fraction of the control wavelength bandwidth to transmit its BCPs, similar to the header field in OPS. OGBs are still needed to accommodate finite switching times, but the impact decreases due to increased payload durations.

4.2.3. Optical Technology Status

Most optical packet/burst switch designs require advanced technology for realisation. Among the essential challenges are control unit processing and optical switching technologies, which are briefly discussed below. Detailed studies of all enabling technologies have not been the focus of the COST 266 action, and are beyond the scope of this document, overview of key technology realisation can e.g. be found in [2].

The control unit should maintain an accurate routing table. It can typically be updated on a relatively slow time-scale, based on control plane information on topology changes and load balancing factors. On the other hand, electronic processing of packet headers must be performed for a high number of packets simultaneously. The operation typically includes table look-up to identify

output port, scheduling of output wavelength or buffer input in agreement with packets' service class. As a maximum bound, the sum of processing time and switch fabric reconfiguration time should not exceed the packet duration. Otherwise a data flow bottleneck at the switch input occurs. Hence, electronic processing may become a bottleneck in optical networks. Increasing packet duration gives more time for processing, so that OBS eases the requirement on total processing time. However, since the scheduling is expected to be more advanced, e.g. to implement a just enough time (JET) based QoS differentiation scheme; there will still be stringent requirements on the speed [23] and/or processing parallelisation degree of the control unit.

Switch fabrics for OPS and OBS can be categorised as *space switches*, *broadcast-and-select* and *wavelength routers*. This review considers maximum switching times for OPS of 100 ns, and is based on [22]. OBS may accept larger switching times, and may have more alternatives for switching fabrics. However, considering burst durations up to 100 μs, the switching time should not exceed 10 μs, excluding e.g. the use of the relatively mature micro electro-machine systems (MEMS) switches. All architectures considered are strictly non-blocking, unless otherwise stated.

N×N *space switches* are configured by setting basic solid-state optical switches according to the switch architecture. *Crossbar* switches are based on 2×2 switches, and suffer from poor cross-talk properties. They are wide-sense non-blocking, thus an intelligent switch path selection can completely avoid internal blocking. *Router-Selector* switches use 1×2 switches and exploit this architecture's excellent crosstalk properties. Crossovers and bends in the interconnection shuffle may however give differential loss and cause crosstalk. In these architectures the number of basic switches scales as N^2 and $2N^2$, respectively. Reported matrices with less than 100 ns switching time are rare, and these matrices are limited to port counts of 16×16.

The standard *Broadcast-and-Select* switch is based on passive $1/(N \cdot W)$ splitting of the N WDM input signals, each with W wavelengths, followed by active selection at each of the $N \cdot W$ outputs. A high splitting ratio gives SNR reduction, and the loss is compensated for by EDFAs which further decrease signal quality by adding the ASE noise. The total number of SOA gates needed is $W^2 \cdot N + W \cdot N^2$. For a given size ($N \cdot W$), SOA count is minimised when $N = W$. An integrated board compatible with a 16×16 space switch with $W = 16$ (throughput of 2.56 Tbit/s) has been reported [24]. 512 of the total 8192 SOAs would be in the on-state at any moment, and power consumption becomes an issue. Using sub-equipped versions of this board, an OPS experiment has demonstrated a throughput of 640 Gbit/s in asynchronous operation

[25]. An adaptation proposed in the context of OBS is called Tune-and-Select and has been shown to scale to an effective throughput in the range of Tbit/s [26].

Wavelength-routers are based on a passive fabric, usually array waveguide gratings (AWGs), with preconfigured input-output paths depending on the input wavelength. Reported in [27] is e.g. a demonstration of use of such a 42×42 AWG for packet switching. In general, spectral filtering and channel cross-talk may cause signal degradation, but AWGs have a very good potential for use in OPS switches.

4.2.4. Packet/Burst Switch Architectures

In general, except for the control information encoding and control unit scheduling, OPS node architectures for asynchronous operation will also be suitable for OBS.

Example of a blocking OPS architecture

A very basic node design consists of having the demultiplexed WDM signals connected to an AWG via tuneable wavelength converters (TWCs). This is very simple switch architecture, however, it has the drawback that it is blocking, since no wavelength conversion is performed at the output of the switch, the converters at the input can only use the same wavelength set as in the input fibre. This has as a consequence that this switch architecture is internally blocking.

Important in such a switch design is how the output ports of the AWG are combined into the output fibres [28]. If this is done properly and intelligent choices on the wavelength conversion are made at the input, the performance of this blocking node can approximate the performance of a non-blocking node. In asynchronous mode, however, it is a lot harder to emulate a non-blocking node using this architecture. The problem is that once a decision is taken for a TWC this cannot be reverted, but when a future packet arrives it might become clear that another choice would have been better. Thus a possible way of improving performance is using a windowed scheduling mechanism [29]. This scheduling in the switch increases the time separation of header and payload by an extra FDL at the input. In this way there is a form of prediction of which packets will block each other, so that the converter decision for these overlapping packets can be coupled, which will result in a lower blocking probability. We show simulation results for this switch design in Figure 4.5.
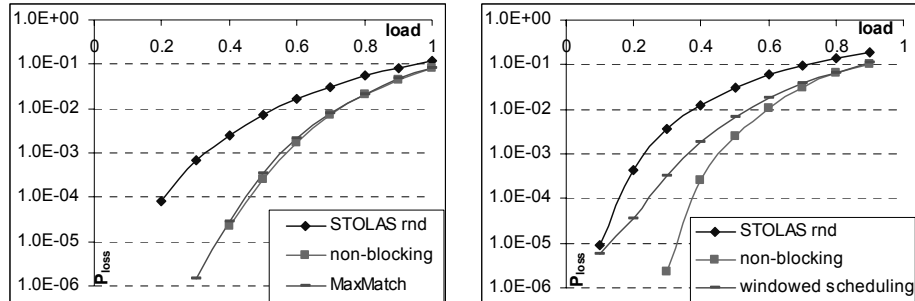
Figure 4.5. - Packet loss simulations for a blocking optical packet switch (3 fibres in and out, 15 wavelengths per fibre) in slotted (left) and asynchronous mode (right). A non-blocking node is shown as reference, STOLAS rnd is performance using no specific TWC assignment algorithm. In slotted operation, MaxMatch TWC assignments allow to emulate a non-blocking node. In asynchronous mode windowed scheduling improves the performance, but there still exists a serious gap with the ideal non-blocking node.

Example of non-blocking OPS architecture

We here describe a switch design, suitable for asynchronous packet switching, shown in Figure 4.6, proposed in [11]. The input WDM signal of each fibre is demultiplexed to its corresponding wavelengths and fed to the input of the TWCs. The outputs of each TWC are then fed to the AWG inputs. By tuning the TWCs output wavelength, packets can be sent to either of the AWG outputs. The packet will be sent directly to the scheduled output, if vacant. If no output with correct destination is available, the packet will be sent to one of the buffer inputs, if a vacant buffer input can be found. If not, the packet will be dropped.

Buffered packets are clocked out of the buffer and sent back to an AWG input as soon as a wavelength output to the destination becomes available. At the buffer output, the wavelength, and thus AWG output, is set by tuning a tunable laser. This type of architecture is called a feedback design, and has the benefit of supporting packet priority, also when FDLs are used for buffering [30], [31]. When a packet is leaving the AWG for the output, the signal is converted to the desired wavelength before it is multiplexed onto the correct output fibre.

The design is suitable for the given COST scenario described in Chapter 1 of this report, where node degree is set to typically a maximum of 5. However, a drawback with this design is that it does not scale well to a high node degree. The total number of switch inputs n, is given as $n = N \cdot w$, where N is the number of input/output fibres and w is the number of wavelengths in each fibre. The total number of channels needed in the AWG is given as $n+k$, where k is the number of buffer inputs. An AWG with size $(n+k) \times (n+k)$ is therefore

required. Since *n* increases both with the number of fibres and the number of WDM link wavelengths, the maximum switch size is limited by the size of the AWG, which is currently reported to be a maximum of 400 channels [32]. However, a scalable design based on the same principles, scaling to a very high number of wavelengths, and a high node degree, can be found in [11].
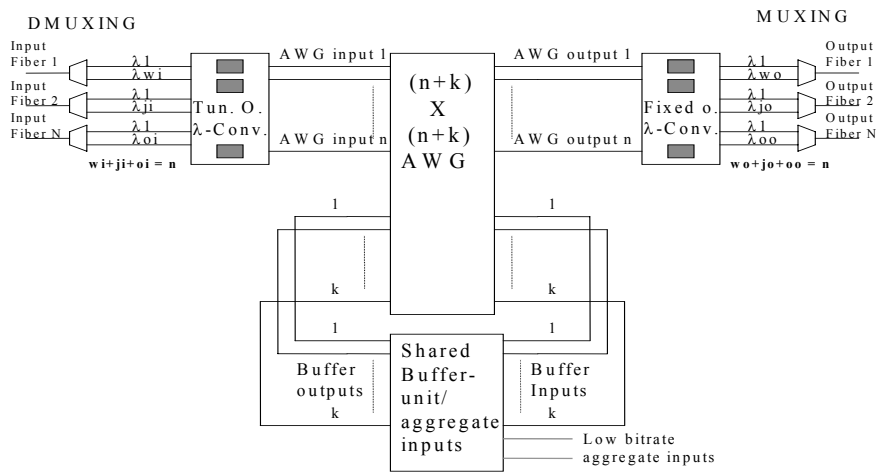

Figure 4.6. - An optical packet switch with shared buffers and low bit rate aggregate inputs

4.2.5. Summary

This section introduced the topic of node design in OPS/OBS networks, and pointed out important design factors. Currently, node design for OPS/OBS is focused on design principles and logical performance. For implementation in networks, the goal is to optimise the network's performance/cost ratio. Hence, one should also consider impact from client layer requirements, technology status and the existing WDM layer's topology and transmission system. OPS/OBS node design requires several performance/complexity trade-offs that will be further studied in relation to contention resolution and QoS differentiation, in sections 4.3 and 4.4, respectively.

4.3 Contention Resolution

4.3.1. Motivation and Overview

Optical burst and packet switching inherently rely on statistical multiplexing in order to achieve good utilisation in presence of bursty traffic. As a conse-

quence, temporary overload situations called contention situations occur and have to be resolved. A reservation or transmission conflict, which leads to burst or packet loss, exists if the wavelength on the designated output fibre is blocked by a different burst or packet. Such a contention situation can be resolved in one or several of the following three domains:

- **Wavelength domain:** By means of wavelength conversion, a burst or packet can be transmitted on a different wavelength channel of the designated output fibre. Thus, all wavelength channels of an output fibre can be considered a single shared bundle of channels.

- **Time domain:** A burst or packet can be delayed until the contention situation is resolved by applying a buffer. In OPS or OBS, either simple FDL buffers or electronic buffers can be used. While FDL buffers only provide a set of fixed delays an electronic buffer can provide virtually unlimited delay and random access. Complexity issues of FDL and electronic buffers will be discussed in the next subsection.

- **Space domain**: In deflection routing, a burst or packet is sent to a different output fibre of the node and consequently on a different route towards its destination node. As contention is not resolved locally in a single node but by rerouting over-load traffic to neighbouring nodes, this scheme depends heavily on network topology and routing strategy. Deflection routing results in only limited improvement for variable length bursts or packets [33] and has not been investigated within this COST Action. Furthermore, as a consequence, packets can arrive out of order at the egress node.

   Space domain can be exploited differently in case of multi-fibre networks, i.e. several fibres are attached to an output interface. In this case, a burst can also be transmitted on a different fibre of the designated output interface without wavelength conversion.

In the following sections, contention resolution in wavelength and time domain is discussed and results of a joint comparative performance evaluation are presented.

4.3.2. Wavelength Conversion

WDM not only provides increased transmission capacity but also allows for highly effective contention resolution. If wavelength converters are employed, all wavelengths on a fibre (or within a certain waveband) can be considered a bundle of channels shared by all bursts or packets to be transmitted over this fibre (waveband). In teletraffic theory, it is well known that a bundle of n parallel servers each with capacity c has a smaller blocking probability and thus a

higher utilization than a single server of capacity $n \cdot c$. This is called economies of scale.

Therefore, most node designs for optical burst and packet switching assume full wavelength conversion, i.e. every incoming or outgoing wavelength is equipped with a wavelength converter. However, as wavelength converters are technologically complex and rather expensive, shared converter pool concepts have been proposed and investigated [34][35]. In this case, the ratio of the number of wavelength converters in the pool and the number of wavelength converters in case of full conversion is referred to as the conversion ratio r_c.

4.3.3. FDL Buffer Architectures

Since traditional queuing is not feasible in all-optical burst or packet switches, contention resolution in the time domain may be provided by using Fibre Delay Lines (FDLs), which imitate conventional queuing by delaying packets that are forced to go through an optical fibre of a given length. In literature different kinds of FDL buffer architectures (either in a single or multistage configuration) have been proposed [36], which may be basically classified into

- *feed-forward* buffer, where a packet coming out of the buffer goes directly to one of the output ports of the switch;

- *feedback* buffer, where a packet coming out of the buffer either goes to the output or re-enters the delay lines.

Since it is here assumed that the output wavelength is reserved at packet arrival, both the above buffer configurations are equivalent to an output queue.

On the other hand, in a DWDM network contention resolution may also exploit the wavelength domain, by sharing the wavelength pool of a fibre and then by transmitting contending packets on different wavelengths.

As a consequence, when a packet needs to be forwarded to an output fibre specified in the routing table, the Wavelength and Delay Selection problem (WDS) arises. In fact, these two actions are somewhat correlated, because the need to delay a packet is related to the availability of the wavelength selected. The WDS problem becomes also more complex in case of asynchronous, variable-length optical packets, since some gaps may appear between queued packets inside the FDL buffer due to the discrete number of available delays [9]. In order to solve the WDS problem under these traffic assumptions, a few resource allocation policies have been proposed [37][10]. Here we consider the following ones:

- *Random Non-Full queue* (RNF): the wavelength is chosen randomly excluding those that will be busy beyond the maximum available delay (full queues).

- *MINimum Length queue* (MINL): the wavelength that will be free as soon as possible (the shortest queue) is chosen.

- *MINimum Gap queue* (MING): the choice this time falls on the wavelength that introduces the smallest gap between the current packet and the last queued one.

In case all queues are full, no choice is made and the packet is lost.

In optical burst switching, JET reservation scheme offers the flexibility to reserve newly arriving bursts in gaps left by already reserved bursts. Thus, JET provides another solution for the problem of gaps induced by FDL buffers [38]. Also, regarding the WDS, the sequence in which wavelength conversion and buffering are applied can be exploited to trade off wavelength converter and FDL buffer usage [39].

4.3.4. Electronic Buffering

Optical memory with random access in the time-domain is known to still be immature, and FDLs are therefore used as optical buffers [30]. Since FDLs give fixed delays, random access is not possible, i.e., in contrast to electronic memory, FDLs cannot provide access to a specific data packet at an arbitrary access time.

First in first out (FIFO) operation of the buffer is desirable in order to avoid reordering of packets on their way to the destination. Although this can be achieved using FDLs, storing variable length packets brings up the need for buffering all, or most of the, packets using many different delays and thereby a high number of FDLs. As an alternative to FDLs, the use of simple electronic FIFO memory with few opto-electronic interfaces is suggested in [11]. When using electronic memory, fast random access with respect to time in a FIFO buffer can be obtained. A random access in space to a random storage unit is more complicated since addressing the storage unit before readout of the data is then needed. However, since FIFO buffering is used, access to a random storage unit in the buffer is not required.

Like when using FDLs, data-format transparency is obtainable in electronic memory, however bit rate transparency is more complicated since clock recovery circuits recognizing the bit rates is then needed.

4.3.5. Comparative Performance Evaluation

In a common evaluation scenario, the impact of different WDS algorithms and individual FDL delays in an FDL buffer is compared for OBS and OPS both assuming asynchronous operation and variable length bursts or packets. Then, both approaches are compared to OPS with electronic buffers based on the number of buffer interfaces.

Bursts and packets arrive at a node with 4 input and output fibres according to a Poisson process with an offered load of 0.8. Burst and packet length is negative exponentially distributed with mean 100 kbit (bursts) and 4 kbit (packets) which translates into an average transmission time $h = 10\,\mu s$ (bursts) or $h = 0.4\,\mu s$ (packets) for a 10 Gbit/s line-rate. Unless stated differently, 16 wavelengths are assumed on each fibre and FDL. For the FDL buffer, the length of FDL i can be calculated as $i{\cdot}b$ with respect to a basic delay b. For OBS, JET is applied for fibre and FDL reservation. FDL reservation is performed at time of burst arrival (*PreRes* in [38]).

Figure 4.7 compares the three WDS policies for OPS with 8 FDLs in the buffer (circles) and OBS with 4, 6 and 8 FDLs in the buffer (triangles). Burst/packet loss probability is plotted as a function of the basic buffer delay b normalized to the average packet length.
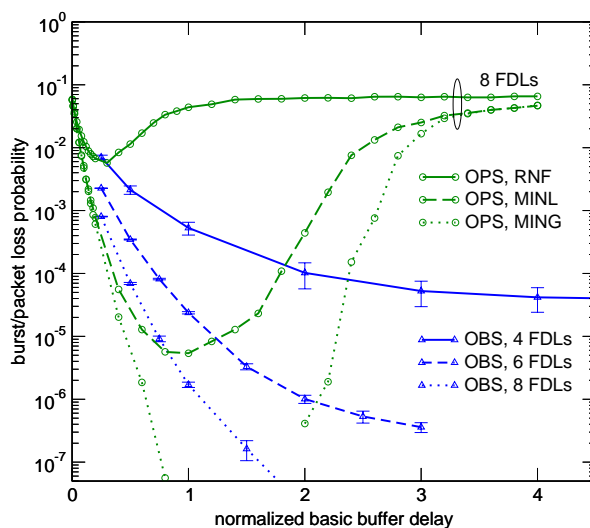

Figure 4.7. - Burst/Packet loss probability versus normalized buffer delay

The RNF choice gives the worst performance since such policy does not detect the wavelengths immediately available, which do not insert gaps in the buffer. More intelligent mechanisms, such as MINL and MING, provide a strong

improvement. In particular MING outperforms MINL because it aims, first of all, at reducing the gaps, leading to a more efficient buffer utilization and therefore to shorter queues overall and better performance. All three curves clearly show a minimum of the packet loss rate as a function of the delay unit, with the position of such minimum depending on the wavelength selection policy adopted. These results demonstrate that a smart WDS policy together with an accurate dimensioning of the buffer parameters allow to achieve very good performance with only a limited number of FDLs.

For OBS, increasing the delay significantly reduces loss probability for different number of FDLs in the buffer. In contrast to OPS and due to the more flexible but also more complex JET reservation mechanism, no minimum appears but a saturation effect can be observed from a basic delay of 2-3 times the mean burst length on.
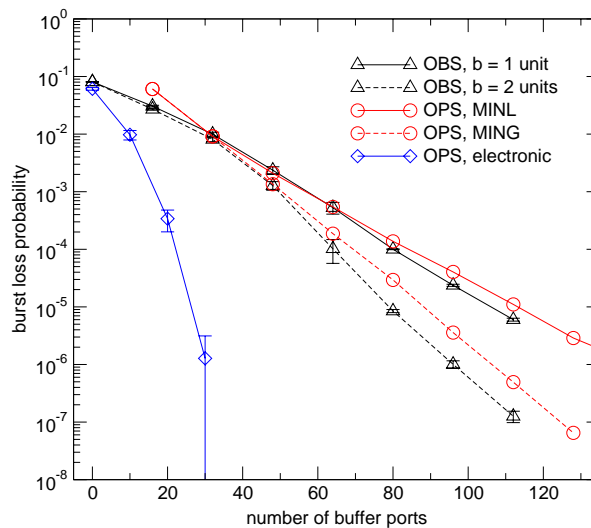

Figure 4.8. - Burst/Packet loss probability versus number of buffer ports

Figure 4.8 depicts the impact of the number of buffer ports for FDL and electronic buffers. For FDL buffers, the number of buffer ports is the product of number of wavelengths and number of FDLs in the buffer. Basic delay of FDL buffers is chosen to be the optimal one in case of OPS (Figure 4.7), and one and two times the mean burst transmission time in case of OBS, respectively. It can be seen that increasing the number of buffer ports greatly reduces blocking for all scenarios. While all results for FDL buffers show comparable trends, electronic buffers need a significantly smaller number of buffer inter-

faces. However, as electronic buffers need O/E/O interfaces the number of buffer ports has to be minimized. For a node with 8 input and output fibres and different numbers of wavelengths, Figure 4.9 shows that a decreasing number of buffer ports are needed for an increasing number of link-wavelengths.
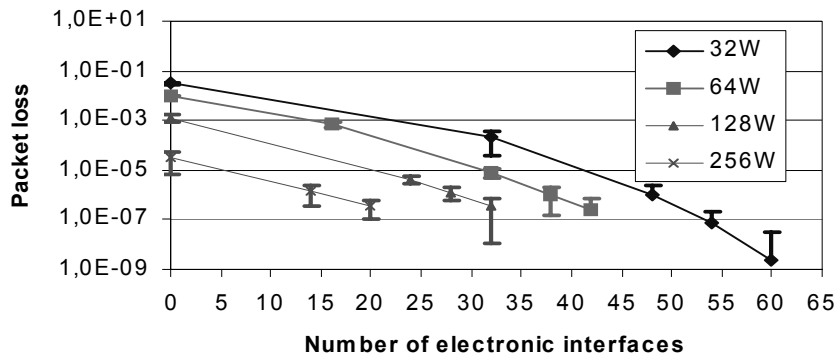

Figure 4.9. - Packet loss when using electronic buffering, 32, 64, 128 or 256 wavelengths (W)

4.3.6. Summary

In this section a performance evaluation based on simulations from several partners in COST 266 has been presented. Different approaches for OPS/OBS contention resolution in time, assuming contention resolution also in the wavelength domain is shown. We conclude that both FDL and electronic buffers can significantly reduce loss probability when used together with wavelength conversion. The comparison shows that electronic buffers need fewer but potentially more expensive O/E/O interfaces to reach the same loss rates. For all the compared buffering schemes delay is negligible compared to typical end-to-end delays.

4.4 Quality of Service in OBS/OPS

4.4.1. Introduction

Introduction of multimedia applications in the Internet, which may have strict real time and information loss demands like a high quality video component, have increased the need for service quality differentiation. We expect IP to be the converging protocol layer, but the IP protocol itself does not support QoS differentiation. When implementing an OBS or an OPS layer, the quality of the service offered will be influenced by the amount of resources, like buffering and wavelength converters, spent in the network nodes. An OPS layer

should therefore be able to support QoS differentiation, preventing over-dimensioning of the network nodes and delivering QoS differentiation to the IP-layer.

ITU-T Recommendation Y.1541 [40] defines some provisional IP network QoS class definitions and end-to-end network performance objectives. The highest demand of any class with respect to IP packet transfer delay is 100 ms, with an allowed delay variation of 50 ms. As shown in the analysis in this report's "contention resolution" chapter, delay in OPS and OBS networks is negligible. The lowest specified IP packet loss ratio for any class is 10^{-3}. One should however notice that this value is for an end-to-end IP relation. When specified for real-time services, it is based solely on observation of high quality voice applications and voice codecs (see e.g. [41] for a laboratory study of four VoIP gateways demonstrating that even an IP packet loss ratio of 0.1 may be acceptable). It is however made very clear that some of the values in Y.1541 are too relaxed. Appendix IX (Informative) states: "The Classes in Table1/Y.1541 are intended to cover a broad range of applications for which the transport requirements are known. Examples of applications not covered by these classes are broadcast TV distribution, program audio, Digital Cinema, and compressed HDTV transport, where very low loss may be needed and possibly low network delay". It is also clear from other sources that some applications based on MPEG2 video coding can not tolerate IP packet loss ratios above 10^{-5}, e.g. in [42] a packet loss ratio of 10^{-6} is used for the highest priority QoS class. We therefore expect that a packet loss through an OBS/OPS node better than 10^{-6} should be sufficient to service even the most demanding video-services.

4.4.2. Quality of Service in Optical Burst Switching

In order to provide service differentiation directly in the optical layer several approaches have been proposed and investigated for optical burst switching. They take advantage of burst reservation, burst assembly or a combination of both and can be classified based on their key mechanism as follows [6].

- Differentiating offset values,

- Preemption (composite burst switching),

- Intentional dropping of (low priority) bursts,

- (Re-)scheduling in core nodes, and

- Access control and bandwidth reservation.

Offset-based schemes rely on the concept of delayed reservation, i.e. the burst control packet is separated from the data burst by an offset time. In the JET reservation scheme the exact start and end times of burst transmission are considered for reservation. Due to this detailed reservation information, bursts can be reserved in between two already reserved bursts.
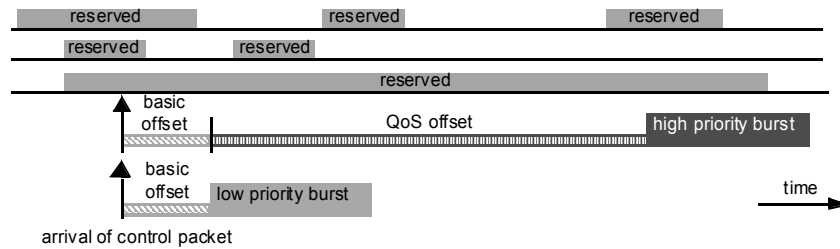

Figure 4.10. - Reservation scenario for bursts of different classes

Service differentiation is achieved by allowing early reservation of high priority bursts by assigning an extra offset time [43] - called *QoS offset*. Therefore, high priority bursts make their reservation in a rather lightly loaded system and have a smaller loss probability while low priority bursts experience the total system load and have a higher loss probability. Figure 4.10 illustrates this effect for three wavelength channels on which some bursts are already reserved, and high priority and low priority bursts with different QoS offsets arrive at the same time.
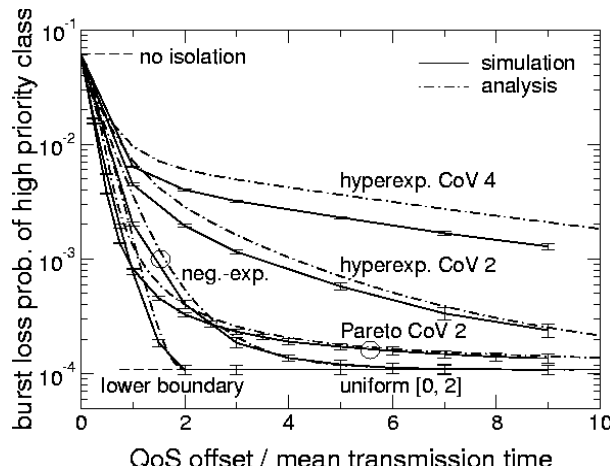

Figure 4.11. - Impact of QoS offset on burst loss probability of high priority bursts

The impact of QoS offsets on differentiation in loss probability has been analysed in [43] and [44]. Figure 4.11 depicts the impact of the QoS offset on the

burst loss probability of the high priority class. As the mean and the distribution of low priority bursts have significant impact, the QoS offset is normalized by its QoS mean burst length, and different burst length distributions of the low priority class are included.

Offset-based QoS has the same total blocking probability with or without service differentiation, i.e. the overall performance is not reduced significantly ([43] reported a slight increase in overall loss rate for low loads). However, this scheme has two severe drawbacks [44]: First, the loss probabilities of high and low priority classes are highly dependent on burst characteristics. Second, basic offset adaptation in core nodes can change the differentiation in offset, which leads to undesirable subclasses as they introduce unfairness [45].

4.4.3. Quality of Service in Optical Packet Switching

Fibre Delay Line buffering

It is important to underline that the QoS management techniques in optical packet switches must be kept very simple due to the delay-oriented nature of FDL buffers. In particular, it is not possible to change the order of packets coming out of the delay lines, thus making pre-emption based techniques not applicable. Therefore, mechanisms based on a-priori access control of packets to the WDM buffers are necessary [46]. The intent here is to improve the WDS policies mentioned in 4.3.3 in order to differentiate the QoS by allowing different degrees of choice to different policies. The objective is to apply some form of reservation of the resources managed by the WDS policies, i.e. the available wavelength and delay, in order to privilege one traffic class over the other. The following alternatives have been investigated:

- *Time-threshold-based technique*: the resource reservation is applied to the time domain, and a delay threshold T_{low} lower than the maximum available delay is defined. The WDS policy for low-priority packets cannot choose delays that are above threshold; therefore, a low-priority packet cannot be accepted if the current buffer occupancy is greater than T_{low}, leaving the remaining buffer space to high-priority packets which see the whole buffer capacity. This causes packets belonging to different classes to suffer different loss rates.

- *Wavelength-based technique*: the resource reservation is applied to the wavelength domain. The WDS algorithm for high-priority packets can send packets to any wavelengths of a fibre, while low-priority packets are allowed to access only a subset (w_{low}) of the wavelength resources, which in any case is shared with high-priority packets. Low-priority packets are

expected to suffer higher congestion and typically to experience higher delays than high-priority ones.

These two concepts have been applied to the MING WDS policy discussed in section 4.3.3, leading to two new QoS-oriented policies named MING-D and MING-LIM which use the time-threshold-based and the wavelength-based technique respectively. As an example, in Figure 4.12 the performance of the MING-D policy is shown for the same node configuration as in section 4.3.3, providing a good separation between the high-priority class (grey curves) and the low-priority class (black curves).
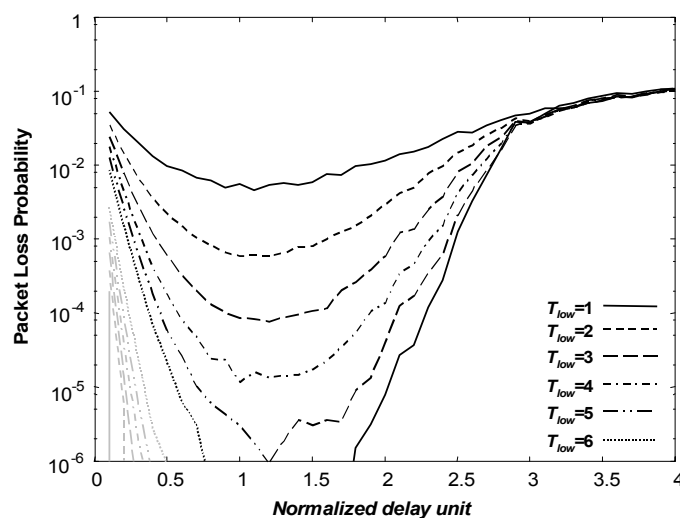

Figure 4.12. - Service differentiation with the MING-D policy

Electronic buffering

Using the OPS switching architecture described in 4.2.4, packet loss ratio differentiation dependent on the service classes is achieved by reserving parts of the buffer inputs for specified service classes. As argued in [47], we expect two service classes to be sufficient for service differentiation in an optical packet switched network. Therefore we have chosen to evaluate the packet loss when the buffer resources are divided into two different blocks of inputs, allowing two service classes. If the packet belongs to the High Class Transport bearer service (HCT), any available buffer input can be used. If the packet belongs to the Normal Class Transport bearer service (NCT), only a limited number of buffer inputs can be used, if one of them is available. If no buffer input is available, the packet will be dropped.

To evaluate the described principle, we have done simulations quantifying to which extent buffer inputs should be reserved when two traffic classes are assumed. The share of traffic belonging to the HCT class will be set to 10 and 50 % respectively, and the number of wavelengths in the links to 32.
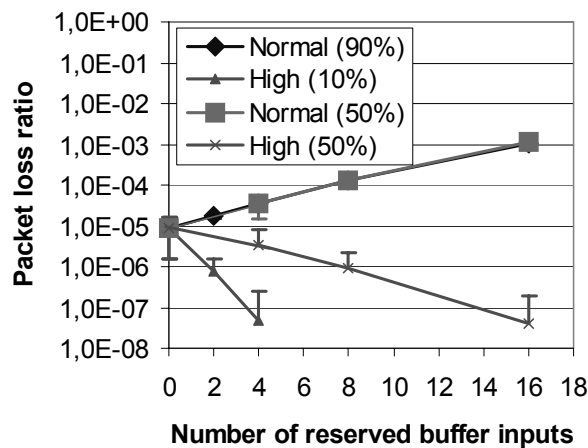

Figure 4.13. - Packet loss as a function of number of buffer inputs reserved for HCT packets, 32 available wavelengths, load 0.8, 8 input/output fibres, variable length Poisson packet arrival. A total of 42 buffer inputs are available. Curves are given for 10 % HCT (High) and 90 % NCT (Low) traffic and the two traffic classes having equal share of traffic (50 %). The two curves for the NCT traffic are coinciding and are therefore seen as only one curve. The error bars marks the limits within a 95 % confidence interval. Where only the upper limit is given, the lower limit is lacking. Higher precision can be achieved, making simulation time excessively long.

In Figure 4.13, the total number of buffer inputs is set to 42. Number of reserved buffer inputs is varied from 0 to 16. The share of HCT traffic is set to 10 % and 50 % of the total traffic load, while the rest of the traffic consists of NCT traffic. The figure shows clearly that reserving buffer inputs gives a decrease in packet loss ratio for the HCT packets, while the NCT packets pay the price with a higher packet loss ratio. It is also confirmed that when the fraction of HCT traffic is increased to 50 %, the number of buffer inputs reserved for the HCT traffic has to be increased significantly in order to obtain the same packet loss ratio as for the case when HCT traffic load is 10 %.

When the number of reserved buffer inputs is set to 4 and the fraction of HCT traffic is set to 10 %, packet loss ratio is approximately three orders of magnitude higher for the NCT traffic than for the HCT traffic. The obtained PLR of $< 10^{-7}$ satisfies the demands for the HCT class, while PLR of $< 10^{-4}$ satisfies the demands for the NCT class. Assuming a 50/50 split of the traffic load

between the two QoS classes, PLR's that satisfies the QoS demands is obtained by reserving 8 buffer inputs.

4.4.4. QoS in MPLS Optical Networks

In the first phase of building optical packet networks, QoS mechanisms can be based on MPLS, which offers resource reservation with proper control algorithms to support Class of Services and Traffic Engineering enhancing network efficiency.

MPLS aspects of optical packet networks

MPLS is a suitable technique for packet routing in optical packet networks. Optical packet nodes use information in packet labels to decide how to forward a packet. Within the optical node, the optical packet label is read and compared with a look up table. The payload is then transparently routed in the optical domain to the appropriate output port with a new label attached. Core MPLS optical nodes can process labels more rapidly than traditional address headers (i.e. IP headers), therefore network performance may be improved. This is especially desirable in optical packet networks, due to the weak buffering capabilities.

MPLS control unit

An MPLS controller should perform the following functions: (*i*) Building and maintaining the Label Information Base (LIB), (*ii*) MPLS signalling with support for CR-LDP and RSVP Tunnels for Label Distribution, (*iii*) forwarding which includes processing of incoming packets, making of forwarding decisions, packet shaping and delivering the packets to the outputs. Edge nodes should also be equipped with adaptation functions for incoming/outgoing traffic.

QoS tasks in an MPLS controller consist of packets classification in edge nodes, differentiated packets servicing in core nodes and Traffic Engineering that performs operations on traffic vectors such as merging, comparing, summarising and subtracting flows as well as LSP admission control and traffic load management.

Support for QoS

In order to address the QoS issue, the ability to introduce connection-oriented forwarding techniques to connectionless optical packet network is necessary. MPLS offers such functionality by establishing the label switched paths (LSPs). In effect, this allows optical network to reserve resources, such as buffers or wavelengths over predetermined paths for service differentiation,

providing QoS guarantees. Currently two main MPLS features for supporting QoS are used, differentiated servicing [48, 49] and Traffic Engineering.

Differentiated Servicing – QoS in nodes

The MPLS Class of Service (CoS) feature enables support for differentiated types of services across an MPLS optical network. The differentiated services model defines a variety of mechanisms for classifying traffic into a small number of service classes. Once packets are classified at the edge of the network, specific forwarding rules are applied in each node. This combination of packet marking and specific servicing procedure results in a scalable QoS solution for any traffic flow.

Traffic Engineering – QoS in networks

MPLS Traffic Engineering is a process of routing data traffic in order to balance the traffic load on the various links and nodes in the network. It assures e.g. better utilization of available bandwidth (e.g. automatically increasing or decreasing the bandwidth reserved for an MPLS TE tunnel based on measured traffic load), accommodation of high class traffic load to buffering capabilities in nodes, routing around failed links/nodes (reliability) and capacity planning. All these functions improve QoS of the networks.

Summary

In burst switching, service differentiation is achieved by e.g. allowing early reservation for high priority bursts through assigning them an extra offset time - called *QoS offset*. Therefore, high priority bursts make their reservation in a rather lightly loaded system and have a smaller loss probability while low priority bursts experience the total system load and have a higher loss probability. However, drawbacks of this principle are that the loss probabilities of high and low priority classes are highly dependent on burst characteristics and basic offset adaptation in the core nodes can change the differentiation in offset and lead to undesirable subclasses.

In OPS, when using FDL's, the QoS management techniques must be kept very simple due to the delay-oriented nature of the FDL buffers. In particular, if a feed-forward FDL buffer is used like in [8], it is not possible to change the order of packets coming out of the delay lines, thus making the pre-emption based techniques not applicable. Therefore, mechanisms based on a-priori access control of packets to the WDM buffers are necessary. Two techniques have been studied: 1) The *time-threshold-based technique*, where the resource reservation is applied to the time domain, and a delay threshold T_{low} is defined. In contrast with the high priority packets, the low-priority packets cannot

choose delays that are above the threshold, thereby allowing these packets to access only parts of the buffer. 2) The *wavelength-based technique* where the resource reservation is applied to the wavelength domain. Low-priority packets are allowed to access only a subset of the wavelength resources, while high-priority packets can access all the wavelength resources.

As opposed to using FDLs, if electronic buffers are used in combination with OPS, advanced techniques for QoS differentiation are feasible. Still, if efficient QoS differentiation can be achieved using simple principles, the number of components and thereby the costs can be reduced. In this report, a principle based on QoS differentiation in an optical packet switch with a limited number of electronic buffer interfaces has been studied. Differentiation on packet loss is achieved by giving high priority packets access to all buffer inputs, while low priority packets only can access parts of the buffer inputs. Obviously, this approach can be easily adapted to optical buffering with FDLs.

OPS can use MPLS for support of control and QoS differentiation. MPLS core nodes can process labels more rapidly than the traditional address headers (i.e. IP headers); therefore, network performance may be improved. This is desirable especially in optical packet networks, due to their weak buffering capabilities. By establishing label switched paths (LSPs), resources can be reserved in order to provide QoS guarantees, such as buffers, or wavelengths over predetermined paths.

It is shown that all the described reservation techniques for QoS differentiation have the capability of supporting the PLR performance and the QoS differentiation suggested in the introduction. However, the burstiness of the traffic pattern will influence the performance of the suggested techniques. The design of the OPS/OBS nodes and of the network, as well as of the reservation parameters, will have to be decided according to the traffic pattern.

4.5 Optical Packet Switching in Metro Networks

Nowadays, the optical equipment vendors show increasing interest in enhancing existing and developing new packet-based technologies. The trend in networking is migration of the packet-based technologies from access networks to Metropolitan Area Networks (MANs). The special attention paid to MANs is a result of the rapidly increasing data traffic volume in the metro networks, which is challenging the capacity limits of the existing transport infrastructures such as SONET/SDH and ATM. In such a situation, packet-based transport technology, a natural fit with the now ubiquitous IP protocol,

appears to be one of the best choices for meeting the cost/efficiency trade-off in the metro networks. In this context, researchers world-wide are making great efforts to investigate new MAN architecture solutions (e.g. Resilient Packet Ring [50]).

OPS based on packet switching at the optical level that can also provide high throughput appears to be a good candidate for future MAN architectures (Optical MANs, O-MANs).

In this subsection we identify the main requirements for O-MANs and use them to compare five different O-MANs architectures: three currently under development in different research/university centres ([51], [52] and [53]), and two solutions investigated within the COST 266 Action [54] and IST project DAVID [3].

4.5.1. O-MAN Properties

In the current networking context, a MAN (and consequently an O-MAN) has to meet the following requirements:

- *Flexibility*. It must be able to handle different granularities of bandwidth and to support a wide range of protocols.

- *Cost-effectiveness*. It must beat decisively the current technologies both in CAPEX and OPEX costs.

- *Upgradability*. It has to be able to incorporate new technologies in an easy and non-disruptive manner.

- *Scalability*. It must be possible to remove and add network devices in an easy and non-disruptive way

- *Efficiency*. It must provide high throughputs and short delays.

- *Fairness*. Starvation of nodes must be avoided through a regulation of bandwidth usage.

- *Multicasting*. It must allow multicasting in order to efficiently support applications such as videoconferences or distributed games.

- *Quality of Service*. It must have rapid provisioning capabilities and provide service guarantees to mission-critical data and delay-sensitive applications.

- *Reliability*. The network elements must offer a high degree of reliability. This mandates that critical sub-systems be fully protected and capable of in-service upgrade.

Clearly, these requirements are better met by packet-based technologies than by circuit-based technologies such as SONET/SDH rings.

HORNET (Figure 4.14a), which stands for Hybrid Opto-Electronic Ring NETwork, is a WDM time-slotted ring developed at the Stanford University [51]. HORNET has a tunable transmitter, fixed receiver design (TTFR), and the nodes use a CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) MAC protocol to govern access to the wavelengths (see [51] for more details on the access protocol). HORNET can use either slotted or variable-length packets- a characteristic which can provide flexibility. It can be based on two counter-rotating rings to offer cut-fibre protection; nevertheless, a failure will result in halving the available bandwidth. Multicast can be provided via node-by-node re-transmission, but no protocol is included and evaluated to handle multicast traffic and incoming traffic at the nodes. Neither QoS strategies nor fairness mechanisms are implemented. No performance evaluations to assess the merits of such architecture are available.
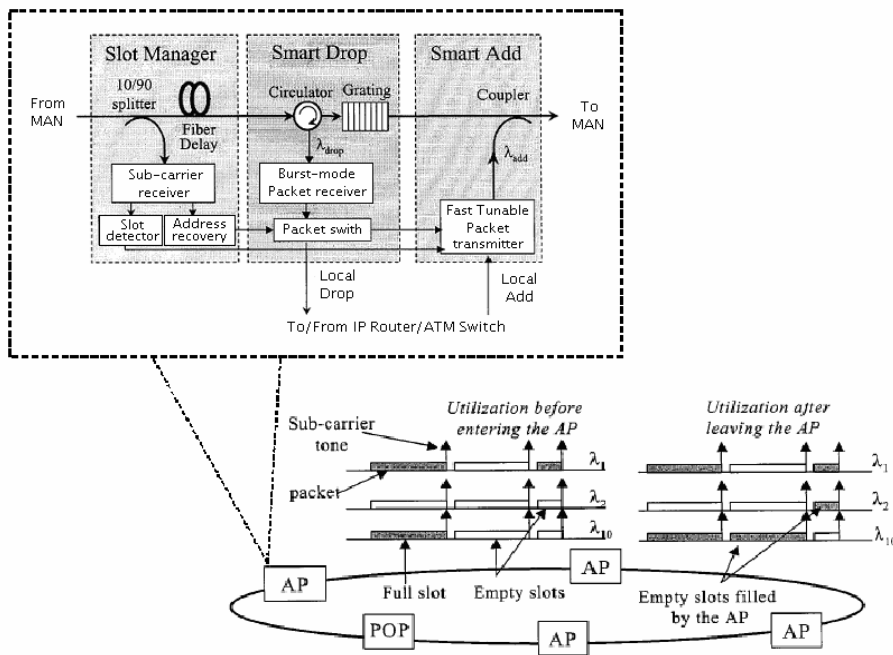

Figure 4.14.a. - O-MAN architectures: HORNET

RingO (Figure 4.14b) is a WDM time-slotted ring developed at the Politecnico of Torino [52]. The peculiarity of the network is that the number of nodes is

equal to the number of wavelengths. In this way, each node is equipped with a laser array (for supporting multicast traffic), and a fixed receiver operating on a given wavelength that identifies the particular node. This wavelength is extracted from the ring by the same node: in order to communicate with node k, a node uses λ_k. The MAC protocol is based on a generalisation of the empty-slot approach, where each node is able to check the state of the wavelength occupancy on a slot-by-slot basis. No other features are currently implemented in the test-bed, nevertheless a fairness mechanism based on a generalization of the SAT token, a QoS strategy based on differentiated service, and several methods for supporting variable-length packets have been suggested and evaluated by simulations. No protection mechanisms have been implemented and, currently, the scalability is largely affected by the imposition of the number of wavelengths equal to the number of nodes.
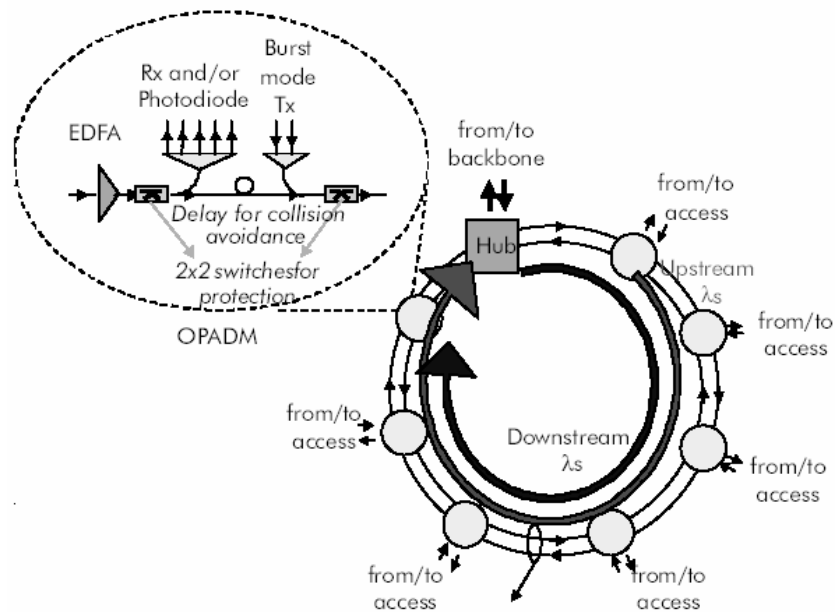

Figure 4.14.b. - O-MAN architectures: RingO

DBORN (Figure 4.14c), which stands for Dual Bus Optical Ring Network, is based on a unidirectional fibre ring organized around a Hub developed at Alcatel Research & Innovation [53]. This architecture uses a spectral separation of upstream and downstream flows from/toward the Hub, forming a dual logical bus structure. Nodes dynamically read data on the downstream bus and write on the upstream bus, while the Hub interconnects the buses through a

wavelength conversion. The spectral separation avoids the use of erasing functionality at the nodes increasing the nodal cascadability. A simple collision avoidance MAC is implemented through power detection utilizing a photodiode and a fixed-length delay line. The network can support any client packets and makes it easy to add/remove nodes to/from the network. Some cost studies have shown the benefits of this architecture [53]. A performance evaluation is not available; neither QoS strategies nor fairness mechanisms have been implemented. The current protection mechanism is based on duplicating the network components.
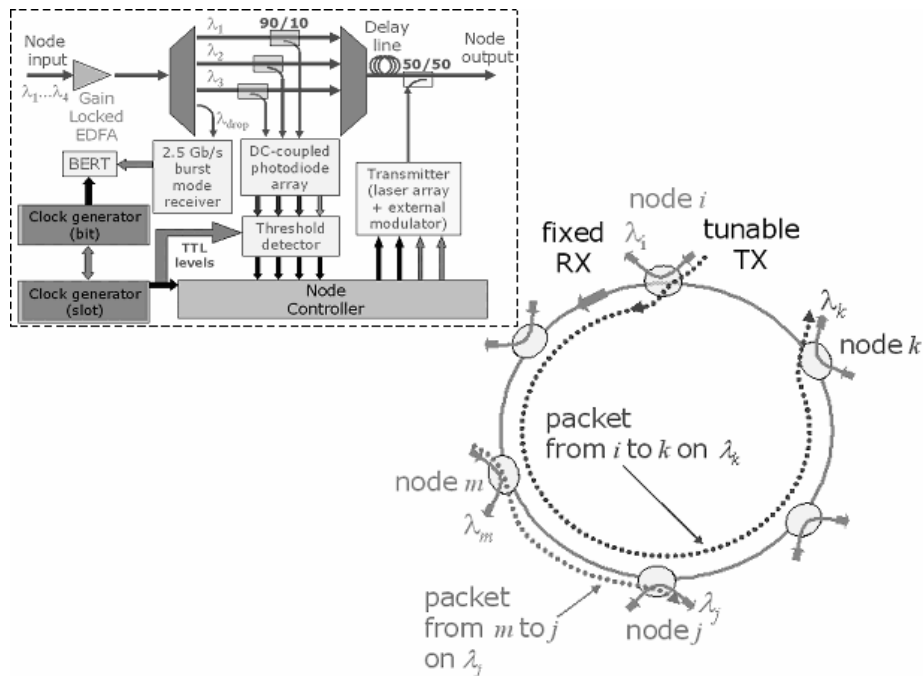

Figure 4.14.c. - O-MAN architectures: DBORN

All the above proposals are based on simple ring architectures where the overall throughput is limited to hundreds of Gbit/s and the access networks attached to the metro nodes are based on copper technologies. A Fibre-To-The-Home scenario would significantly increase the traffic demand up to a few Tbit/s throughputs. Therefore, our research has been concentrated on studying and evaluating novel advanced optical architectures based on multiple rings or multiple trees able to achieve more than 1 Tbit/s throughput.

We have theoretically studied the architecture shown in Figure 4.14d developed within the IST project DAVID. It consists of several optical slotted

packet unidirectional rings interconnected by an optical packet switch called Hub [3]. The performance of such architecture shows good performance in efficiency (both throughput and end-to-end delay) and effective fairness [55]. Moreover, several QoS mechanisms have been proposed and evaluated based both on differentiated service scheme and on guaranteed service provisioning. Some protection schemes have been designed and compared [3].
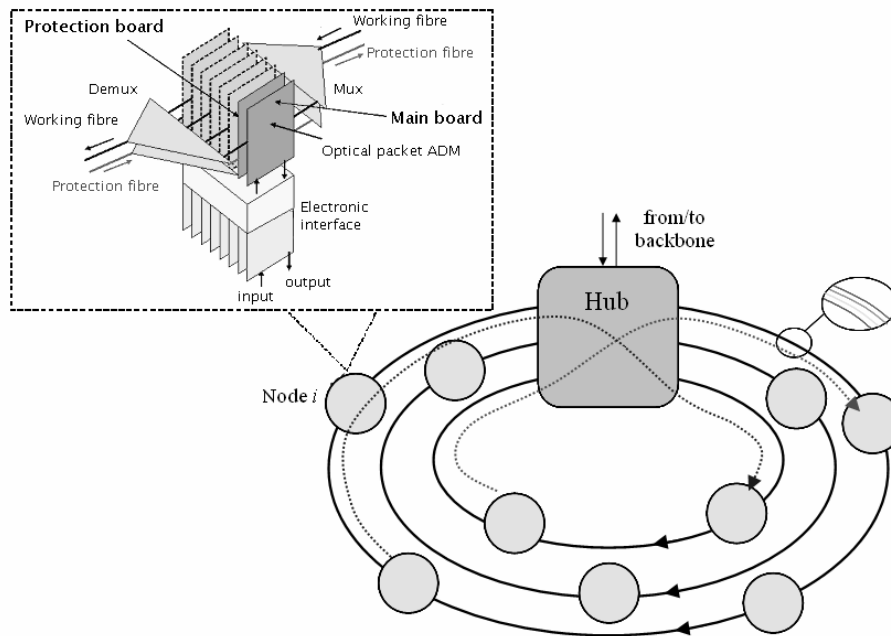

Figure 4.14.d. - O-MAN architectures: DAVID

The O-MAN architecture shown in Figure 4.14 e) consists of several slotted optical packet unidirectional trees interconnected by an array wavelength grating (AWG) which provides a static wavelength routing [54]. A Network Controller (NC) manages the network resources through a proper scheduling algorithm. This architecture recalls the well-known SONATA network. Two different scheduling solutions have been evaluated by simulation [56] [57], and a QoS provisioning scheme providing differentiated service has been proposed in [57]. A protection mechanism based on coupling an AWG with a Passive Star Coupler and a multicast support using optical splitters at the output of AWG have been suggested and evaluated in [58]. Cost study including scalability, upgradeability and reliability issues are in progress. Preliminary results show that this solution seems less costly than the DAVID solution [59].
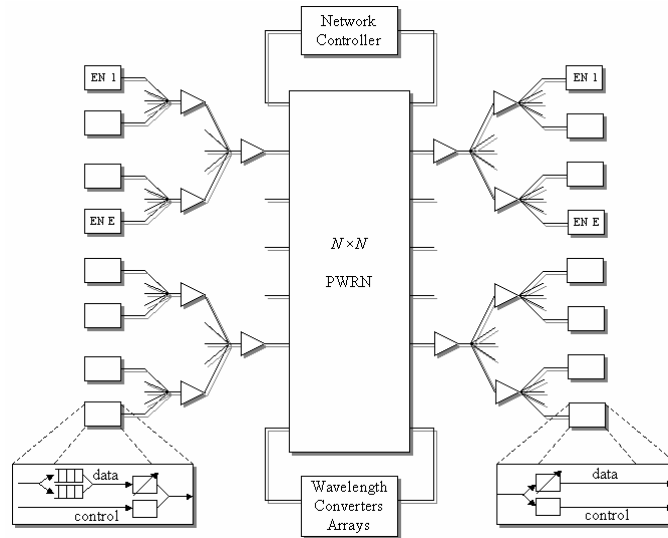
Figure 4.14.e. - O-MAN architectures: DAVID

Table 4.1 shows comparison of the above-described O-MAN architectures. The symbol – indicates that the O-MAN does not implement and/or consider that feature.

Table 4.1. - Comparison of O-MAN architectures

Feature	HORNET	RingO	DBORN	SONATA	DAVID
Flexibility	medium	medium	high	medium	medium
Cost-effective	high	medium	high	high	medium
Upgradability	high	high	medium	–	–
Scalability	medium	low	high	medium	medium
Efficiency	medium	medium	low	high	high
Fairness	–	high	–	high	high
Multicasting	medium	high	–	high	–
QoS	–	medium	–	medium	high
Reliability	medium	–	–	medium	medium

4.6 Conclusion and Outlook

In COST 266, optical burst and packet switching have been studied with respect to performance and complexity. Node designs, contention resolution strategies and QoS architectures in OBS have received special attention.

The work done in COST 266 and recently published papers on OPS/OBS show an emerging trend towards asynchronous packet and burst switching. Comparative concept studies show that packet and burst switching share several properties, but also have some important differences like in the control signalling schemes, the granularity and the overhead.

Key design parameters, as well as a detailed node design of OPS/OBS networks, are described in this chapter. In order to achieve a sufficiently low burst or packet loss rate in the described node designs, contention resolution schemes must be employed, both in OBS and OPS. In combination with the wavelength dimension, both FDL and electronic buffering are compared for OBS and OPS in a common scenario. It is shown that both buffering technologies are able to significantly reduce loss probability. While electronic memory has the potential of using fewer buffer interfaces, O/E/O interfaces are expected to be a major contributor to the cost of the buffer.

For OBS, a classification of several QoS schemes is given, and the offset-based QoS is studied in more detail. For support of control and QoS differentiation, OPS can use MPLS reserve resources and provide QoS guarantees, such as buffers, or wavelengths over predetermined paths.

Finally, metropolitan area networks (MAN) have been investigated. In the MAN environment, the advantages of OPS solutions are highlighted, and the main requirements have been identified and described. These requirements are used to compare optical metro network architectures that are currently under development in research/university centres with two new advanced architectures that have been studied within the COST 266 action and the IST DAVID project.

The performance of a node depends on the available resources, which are the results of several trade-offs between performance and complexity. In addition, client layer requirements, technology status, as well as the topology and transmission systems of existing WDM networks influence the network design.

Further work should therefore include network and end-to-end client layer performance studies, comparing the OCS and OBS/OPS techniques, with respect to performance and CAPEX/OPEX. Scalability is an issue gaining increased importance as the need for large networks increases. Further work on node and network design should take into account the need for network scalability, both with respect to throughput, signal quality and cost efficiency. Topics suitable for further studies are e.g. end-to-end QoS schemes, contention resolution schemes and their influence on the general performance of

OPS/OBS networks, both in the backbone and metropolitan environments. How these schemes can be combined with a GMPLS control scheme should also be investigated.

4.7 References

[1]   F. Masetti *et al*., "High speed, high capacity ATM optical switches for future telecommunication transport networks," *IEEE Journal of Selected Areas in Communications (JSAC)*, 14(5), June 1996.

[2]   C. Guillemot *et al*., "Transparent Optical Packet Switching: the European ACTS KEOPS project approach," *IEEE Journal of Lightwave Technology (JLT)*, 16(12), pp. 2117-2134, Dec. 1998.

[3]   L. Dittman *et al*., "The European IST Project DAVID: a Viable Approach towards Optical Packet Switching," accepted for *IEEE JSAC,* 2003.

[4]   K. Vlachos *et al*., "An Optical IM/FSK Coding Technique for the Implementation of a Label-Controlled, Arrayed Waveguide Packet Router", to appear in *IEEE JLT*, Nov. 2003.

[5]   M. Yoo, M. Jeong, and C. Qiao. "A High Speed Protocol for Bursty Traffic in Optical Networks," *SPIE's All-Optical Communications Systems*, vol. 3230, pp. 79-90, Nov. 1997.

[6]   M. Nord, S. Bjørnstad, and C. Gauger. "OPS or OBS in the Core Network?" *Proc. of Optical Network Dimensioning and Modelling (ONDM) Conference*, Feb. 2003, Budapest.

[7]   I. Chlamtac *et al*., "CORD: contention resolution by delay lines," *IEEE JSAC*, 14(5), pp. 1014-1029, June 1996.

[8]   F. Callegati *et al*., "Architecture and performance of a broadcast and select photonic switch," *Optical Fiber Technology*, 4(3), pp. 266-284, July 1998.

[9]   F. Callegati, "Optical buffers for variable length packets," *IEEE Comm. Letters*, 4(9), pp. 292-294, Sept. 2000.

[10]   L. Tančevski *et al*.,"Optical routing of asynchronous, variable length packets", *IEEE JSAC*, 18(10), pp. 2084-2093, Oct. 2000.

[11]   S. Bjornstad, D. R. Hjelme, and N. Stol, "An optical packet switch design with shared electronic buffering and low bit rate

add/drop inputs," *Proc. of Int. Conference on Transparent Optical Networks (ICTON) 2002*, vol.1, pp. 69-72, Warzaw, Poland, 2002.

[12]   S. Bjornstad, D. R. Hjelme, and N. Stol, "A scalable optical packet switch for variable length packets employing shared electronic buffering", *Proc. European Conference on Optical Communication, ECOC 2002*, Copenhagen, Sept. 2002.

[13]   C. Develder *et al*., "Node architectures for optical packet and burst switching," *Proc. Int. Topical Meeting on Photonics in Switching (PS2002)*, (invited) paper PS.WeA1, pp. 104–106, Cheju Island, Korea, July 2002.

[14]   F. Callegati and W. Cerroni, "Time-wavelength exploitation in optical feedback buffer with trains of packets," *Proc. SPIE OptiComm 2002*, vol. 4874, pp. 274-285, Boston, USA, July 2002.

[15]   P. B. Hansen, S. L. Danielsen, and K. E. Stubkjaer, "Optical packet switching without packet alignment," *Proc. ECOC 1998*, pp. 591-592, Madrid, Spain, 1998.

[16]   Z. Haas, "The 'staggering switch': An electronically controlled optical packet switch," *JLT*, May 1993, pp. 925-936.

[17]   A. Ge, F. Callegati, and L. S. Tamil, "On optical burst switching and self-similar traffic", *IEEE Comm. Letters*, 4(3), Mar 2000, pp. 98 -100.

[18]   F. Xue and S.J. Ben Yoo, "Self-similar traffic shaping at the edge router in optical packet-switched networks", *Proc. International Conference on Communications, ICC 2002*, vol. 4, pp. 2449-53.

[19]   G. Hu, K. Dolzer, and C.M. Gauger. "Does burst assembly really reduce the self-similarity?" *Proc. of Optical Fiber Communication (OFC) conference 2003,* Atlanta, USA, 2003.

[20]   Technical Committee Telecommunications COST 266, Workgroup 2, Advanced Infrastructure for Photonic Networks, Optical Packet and Burst switching Report. (Ed:S. Bjørnstad) Ch. 2. (S. Vandweerd *et al*.) "Optical burst and packet switching research", www.ure.cas.cz/dpt240/cost266/, 2002.

[21]   K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: Recent traffic measurements from an Internet backbone," *Proc. the annual meeting of the Internet Society 1998.* (*INET'98).*

http://www.caida.org/outreach/papers/1998/Inet98/index.xml
Caida 1998.

[22]   Technical Committee Telecommunications COST 266, work-group 2, *Advanced Infrastructure for Photonic Networks, Optical Packet and Burst switching Report. (ed. S. Bjørnstad) Ch 4. (M. Nord) "Analysis of required switching times and identification of suitable switching technologies in optical packet switching,"* www.ure.cas.cz/dpt240/cost266/, 2002.

[23]   S. Junghans and C. Gauger "Architectures for Resource Reservation Modules for Optical Burst Switching Core Nodes," *Proceedings of the 4th ITG Workshop on Photonic Networks*, Leipzig, May 2003.

[24]   D. Sahri *et al*. "A Highly integrated 32-SOA gates optoelectronic module suitable for IP multi-terabit optical packet routers," *Proc. OFC'01* PD32, Anaheim, USA, 2001.

[25]   F. Masetti *et al*. "Design and Implementation of a Multi-Terabit Optical Burst/Packet Router prototype," *Proc. OFC'02*. vol.1, FD11 - FD13. Anaheim, USA, 2002.

[26]   H. Buchta, C.M. Gauger, E. Patzak, and J. Saniter, "Limits of effective throughput of optical burst switches based on seminconductor optical amplifiers," *Technical Digest of the Optical Fiber Communication Conference* (OFC 2003), Atlanta, March 2003, pp. 215-217.

[27]   J.Gripp *et al*. "Demonstration of a 1.2 Tb/s Optical Packet Switch Fabric (32 * 40 Gb/s) based on 40 Gb/s Burst-Mode Clock-Data-Recovery, Fast Tuneable Lasers, and a high-performance NxN AWG," *ECOC 2001*, PD A.1.7, Amsterdam, the Netherlands, 2001.

[28]   J. Cheyns, C. Develder, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, and P. Demeester, "Routing in an AWG based optical packet switch," *Photonic Network Commun. (PNET)*, Kluwer Academic Publishers, 5(1), pp. 69 - 80., January 2003.

[29]   J. Cheyns, E. Van Breusegem, C. Develder, A. Ackaert, M. Pickavet, and P. Demeester, "Performance improvement of an internally-blocking optical packet/burst switch," accepted for ICC 2003, , Anchorage, Alaska, USA, 11-15 May 2003.

[30]   M. C. Chia *et al*. "Packet Loss and Delay Performance of Feedback and Feed-Forward Arrayed-Waveguide Gratings-Based Optical Packet Switches With WDM Inputs-Outputs," *IEEE JLT* 19(9), September 2001.

[31]   C. Develder, M. Pickavet, and P. Demeester, "Strategies for an FDL based feed-back buffer for an optical packet switch with QoS differentiation", *Proc. Int. Conf. on Optical Internet (COIN2002),* paper COIN.TuD1, pp. 114–116, Cheju Island, Korea, 21–25 Jul. 2002.

[32]   Y. Hida *et al*. "400-channel 25-GHz spacing arrayed waveguide grating covering a full range of C and L-bands," *OFC'01*, vol. 3, WB2, 1-3, Anaheim, USA, 2001.

[33]   Yao, B. Muhkerjee, S. J. B. Yoo, and S. Dixit, "All-optical packet-switched networks: a study of contention-resolution schemes in an irregular mesh network with variable-sized packets," *Proc. OptiComm 2000*, (Dallas), October 2000.

[34]   S. L. Danielsen, B. Mikkelsen, C. Joergensen, and T. Durhuus, "WDM packet switch architectures and analysis of the influence of tuneable wavelength converters on the performance," *IEEE JLT,* vol. 15, pp. 219–227, Feb. 1997.

[35]   V. Eramo and M. Listanti, "Packet loss in a bufferless optical WDM switch employing shared tunable wavelength converters," *IEEE JLT*, vol. 18, pp. 1818–1833, Dec. 2000.

[36]   D. K. Hunter, M. C. Chia, and I. Andonovic, "Buffering in Optical Packet Switches," *IEEE JLT*, 16(12), pp. 2081-2094, Dec. 1998.

[37]   F. Callegati, W. Cerroni, and G. Corazza "Optimization of Wavelength Allocation in WDM Optical Buffers," *Optical Networks Magazine*, 2(6), pp. 66-72, Nov. 2001.

[38]   C. M. Gauger, "Dimensioning of FDL buffers for optical burst switching nodes, *Proc. ONDM 2002*, Torino, 2002.

[39]   C. M. Gauger, "Performance of converter pools for contention resolution in optical burst switching," *Proc. SPIE OptiComm 2002*, Boston, July 2002.

[40]   ITU-T Recommendation Y.1541: Series Y: " *Global information infrastructure and internet protocol aspects. Internet protocol aspects – Quality of service and network performance: Network*

performance objectives for IP-based services," Prepublished Recommendation (05/2002).

[41]   Black, Uyless, "*Advanced Internet Technologies*," Prentice Hall, pp. 123-128, 1999.

[42]   R.J. Gibbens *et al.* "An Approach to Service Level Agreements for IP networks with Differentiated Services," Statistical Laboratory, University of Cambridge and Internet and Data Networks, British Telecommunications plc. (Article submitted to Royal Society).

[43]   M. Yoo, C. Qiao, and S. Dixit, "QoS Performance in IP over WDM Networks", *IEEE JSAC- Special Issue on the Protocols for Next Generation Optical Internet*," 18(10), pp. 2062-2071, Oct. 2000.

[44]   K. Dolzer, C. Gauger, "On Burst Assembly in Optical Burst Switching Networks – A Performance Evaluation of Just-Enough-Time," *Proc. 17th International Teletraffic Congress (ITC 17)*, Salvador/Brazil, 2001.

[45]   C. Gauger, "Determining offset times in optical burst switching networks", Presentation at COST 266 MCM, Zagreb, June 2001.

[46]   F. Callegati, G. Corazza, and C. Raffaelli, "Exploitation of DWDM for Optical Packet Switching with Quality of Service Guarantees," *IEEE JSAC*, 20(1), pp. 190-201, Jan. 2002.

[47]   S. Bjørnstad, N. Stol, and D. R. Hjelme. "Quality of service in optical packet switched DWDM transport networks," *Proc. Asia Pacific Optical and Wireless communications conference and exibition 2002* (APOC 2002, SPIE conference), Shanghai, China, 2002.

[48]   E. Horlait and N. Rouhana, "Differentiated Services and Integrated Services Use of MPLS," *Proc. 5th IEEE Symposium on Computers and Communications (ISCC 2000)*, pp 1-6, 2000.

[49]   M. Klinkowski and M. Marciniak, "Services differentiation in MPLS photonic packet networks," *Proc. ONDM 2003*, Budapest, Hungary, Feb. 2003.

[50]   IEEE 802.17, Resilient Packet Ring Working Group, http://grouper.ieee.org/groups/802/17/

[51]   K.V. Shrikhande *et al*., "HORNET: a packet-over-WDM multiple access metropolitan are ring network," *IEEE JSAC*, 18(10), Oct. 2000.

[52]   R. Gaudino, *et al*., "RINGO: a WDM ring optical packet network demonstrator," *Proc. ECOC 2001* Th.L.2.6, Amsterdam, Netherlands, Oct. 2001.

[53]   N. Le Sauze, *et al*., "A novel, low cost optical packet metropolitan ring architecture," *Proc. ECOC 2001*, PD, Amsterdam, Netherlands, October 2001.

[54]   D. Careglio, J. Solé-Pareta, S. Spadaro, and G. Junyent, "Performance Evaluation of Interconnected WDM PONs Metro Networks with QoS Provisioning," *Proc. 2nd Workshop on All-Optical Network (WAON 2001)*, Zagreb, Croatia, June 2001.

[55]   D. Careglio, J. Solé-Pareta, S. Spadaro, and G. Junyent, "Evaluación de prestaciones del la red metropolitana del proyecto DAVID", *12th Telecom I+D*, Madrid, Spain, Nov. 2002 (in Spanish).

[56]   D. Careglio, J. Solé-Pareta, S. Spadaro, and G. Junyent, "Performance Evaluation of Interconnected WDM PONs Metro Networks with QoS Provisioning," *Proc. ONDM 2003*, Budapest, Hungary, Feb. 2003.

[57]   D. Careglio, A. Rafel, J. Solé-Pareta, S. Spadaro, and G. Junyent, "Quality of service strategy in an optical packet network with a multi-class frame-based scheduling," Proc. HPSR 2003, Turin, Italy, June 2003.

[58]   C. Fan, M. Maier, and M. Reisslein, "The AWG‖PSC network: a performance enhanced single-hop WDM network with heterogeneous protection," *Proc. of IEEE Infocom 2003*, San Francisco, CA, April 2003.

[59]   A. M. Hill, D. Careglio, J. Solé-Pareta, and A. Rafel, "Relative costs of WDM rings and PONs for metro optical packet networks," *Proc. PS 2003*, Paris, France, Sept. 2003.