Buffering in Optical Packet Switches

David K. Hunter, *Member, IEEE,* Meow C. Chia, and Ivan Andonovic, *Senior Member, IEEE, Member, OSA*

(Invited Paper)

***Abstract*— This paper consists of a categorization of optical buffering strategies for optical packet switches, and a comparison of the performance of these strategies both with respect to packet loss/delay and bit error rate (BER) performance. Issues surrounding optical buffer implementation are discussed, and representative architectures are introduced under different categories. Conclusions are drawn about packet loss and BER performance, and about the characteristics an architecture should have to be practical. It is shown that there is a strong case for the use of optical regeneration for successful cascading of these architectures.**

***Index Terms*— Buffer memories, communication switching, modeling, optical fiber communication, optical fiber delay lines, packet switching, photonic switching systems.**

I. Introduction

OPTICAL buffering is fundamental to many optical packet switch implementations which have been proposed worldwide to overcome anticipated future problems with large electronic packet switches (including ATM switches). Although significant advances have been made in integration and packaging, electromagnetic interference (EMI) is nevertheless perceived to be a major problem in electronic systems, manifesting itself as crosstalk when conductors are placed close together. Furthermore, interconnecting many chips, each having a very large number of pins, may be a major difficulty with future large electronic switch cores; optics can overcome this problem by a combination of using wavelength division multiplexing (WDM) and high speeds. For these practical reasons, optical packet switching is a strong contender for future telecommunications systems [1].

This paper centers on a discussion of buffering in optical packet switches, in which some representative optical buffering mechanisms are discussed and compared. Their fundamental limitations and difficulties will be identified, as will the performance and relative merits of each scheme. Packet loss and delay performance will be considered, together with their optical performance with respect to crosstalk and noise. The discussion will consider some simple electronic packet switches and show how these have influenced optical packet switch design.

Manuscript received June 9, 1998. This work was supported in part by the Fujitsu Europe Telecom R&D Center under Project EPSRC WASPNET. The work of D. Hunter was supported by EPSRC in the form of an Advanced Fellowship.

The authors are with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW U.K. (e-mail: d.hunter@eee.strath.ac.uk).

Publisher Item Identifier S 0733-8724(98)09301-3.

Throughout this paper, the discussion focuses on the implementation of packet buffering by optical fiber delay-line memories [2] (although silica-on-silicon technology [3] is also attractive for very high bitrates of 50–100 Gb/s and over, where the short delay-line lengths implied are compatible with this technology). The fundamental difficulty facing the designer of an optical packet switch is that variable-length buffers must be implemented with delay-lines; by their nature, these optical delay-lines are of fixed length. The delay-line is simply a length of fiber, and once a packet has entered it, it must emerge a fixed length of time later; it is impossible to remove the packet from the delay-line before that time.

II. Introduction to Packet Switches

Throughout the paper it is assumed that time is divided into equal timeslots, each containing one packet. The development of optical packet switching has involved fixed-length packets, simplifying the design and operation of the network and its switching nodes, and allowing transport of larger entities (e.g., IP datagrams) by splitting them up into smaller, equal packets. It is assumed that packets entering the switch are aligned with respect to their boundaries, so that each packet is aligned with its timeslot. Such synchronization is generally a requirement for correct switch operation; approaches to achieving packet synchronism at the inputs to the switches constitute a topic in its own right [4] which will not be discussed further. Each packet consists of a header and a payload. The header contains (among other things) information pertaining to the packet destination while the payload carries the information itself. In an optical implementation, the header may be at a much lower bitrate than the payload to facilitate the electronic decoding and interpretation of header information.

Such a packet switch has three principal functions: *switching*, *buffering* and, optionally, *header translation. Switching* ensures that each packet emerges at the correct output, depending on the information contained in the packet header. Although packets arriving on the inputs are synchronized, there is no coordination between packet streams arriving on different inputs. Hence one or more packets may arrive during the same timeslot on different inputs wishing to go to the same output. For this reason, *buffering* is required—all but one packet is held up, and subsequently transmitted to the output.

The last function, *header translation*, is part of the ATM (asynchronous transfer mode) standard [5]. ATM switching is a particular type of packet switching; in a conventional TDM system, the channel that each timeslot belongs to is dictated by its position within a frame. In ATM, each packet (or "cell" in ATM parlance) has its channel identified by

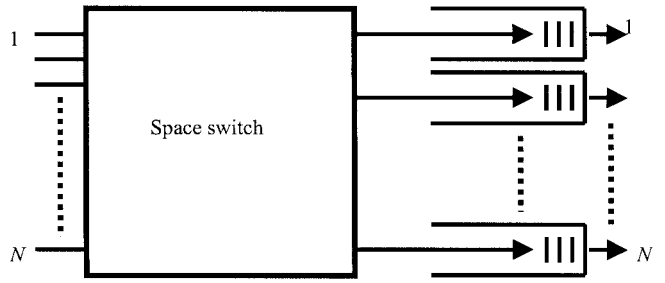


Fig. 1. An output buffered packet switch. Packets enter the inputs of the space switch on the left, aligned on timeslot boundaries, but with no coordination between inputs. They are often directed by the space switch to the appropriate output buffer, depending on their desired output.

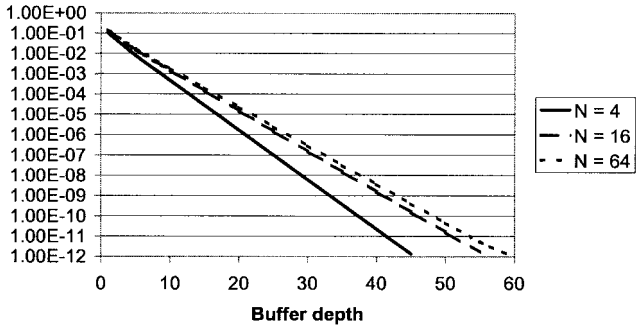


Fig. 2. Packet loss in an output buffered switch for different switch sizes with a load of 0.8. For most practical applications, a packet loss of 10^{-10} or 10^{-11} is required. As expected, the buffer depth requirement increases as the packet loss probability is decreased. There is a slight dependence of packet loss probability on the number of inputs and outputs, with the former converging as N becomes very large.

the virtual channel indicator (VCI) in the header. At every switch in a packet's route, the VCI value is looked up in a table, and from the table, the new value of the VCI and the output port that the packet must go to is determined. The new VCI is substituted for the old one in the packet header. While this scheme offers great functionality and flexibility, it is not used in every optical packet switching system due to its complexity; less complex but less efficient schemes with a longer header can be used instead.

Electronic control is assumed throughout. The devices that make up the architectures are controlled electronically, as the technology for implementing optical control is as yet immature. Importantly, the electronic control circuitry need only operate at the packet rate. For example, if each packet has 1000 bits, and the channel bit rate is 100 Gb/s, the packet rate will be 100 Mpacket/s, i.e., a control signal bitrate of 100 Mb/s.

III. Approaches to Optical Packet Switching

Before detailing the range of optical buffering strategies, it is instructive to examine techniques originally proposed for electronic packet buffering, and see how they are either successful or unsuccessful in mapping onto a photonic implementation. Buffered packet switches are essentially classified according to the position of the buffers. Here, the four most common configurations are examined: output buffering, *shared buffering*, *recirculation buffering*, and *input buffering*. Each has particular advantages and disadvantages. Throughout this paper, a packet switch is assumed to have N inputs and the same number of outputs.

A. Output Buffering

An output buffered switch consists of a space switch with a buffer on each output (Fig. 1). On each timeslot, zero or more packets arrive destined for a particular output (and hence a particular output buffer), all being placed in the appropriate buffer simultaneously. If a packet arrives at a buffer, and the buffer is full, the packet is discarded and packet loss is said to have occurred. Typical acceptable probabilities for a specified packet being lost are usually quoted as being 10^{-10} to 10^{-11}, depending on application, although the figure is correspondingly higher for a circuit or path consisting of many switch traversals [5]. If the buffer size is arbitrarily large (i.e., infinite), then packet loss is zero and the switch performance regarding delay and throughput is optimal [6], [7];[1] packets are delayed only by the unavoidable contention caused by more than one packet being destined for the same output at once. The delay is simply the average delay experienced in a buffer, whereas throughput, ρ, is the average number of packets per timeslot, per input or output (a number between 0–1) which can be sustained before the delay becomes arbitrarily large.

When first analyzing a packet switch, it is usual to assume the simplest possible traffic scenario. In this case, there is a probability ρ (equal to the throughput) that on any timeslot, any input will be receiving a packet. This probability is constant and independent of whether there have been packets on previous timeslots and/or other inputs. With $\rho = 0.5$, this can be likened to repeatedly tossing a coin; the probability of "heads" appearing is always 0.5, independent of whether "heads" came up on any previous tosses of the coin. This type of traffic is known as Bernoulli traffic.

Another simplifying assumption, which is often made simultaneously, is that the traffic is uniformly distributed over the outputs (uniform traffic). In this case, if there is a packet entering any input, there is a probability of $1/N$ that it will be destined for any given output.

Such a simplistic model clearly has limitations, since real traffic (particularly data traffic) is often bursty (i.e., the packets tend to group together in time), and traffic may concentrate on a certain output or group of outputs. Nevertheless, it is amenable to analysis and allows the essential character of various buffer configurations to be determined relatively easily. This model will be assumed throughout.

Fig. 2 shows the packet loss in an output buffer switch under a load of 0.8 with uniform Bernoulli traffic [6], [7]. This packet loss originates when packets are discarded having arrived at a full buffer, and it varies according to the number of inputs and outputs, N. A packet loss of 10^{-11} and a load of 0.8 implies a buffer depth requirement of 55 packets per output, whereas a load of 0.9 with the same packet loss requires a buffer depth of 110 packets. For bursty traffic, buffer depths of thousands have been quoted [8]. The delay/throughput performance is shown

[1] One could, in a sense, reduce delay by throwing away packets but it is best to regard the delay of these discarded packets as being infinite since they never reach their destination.

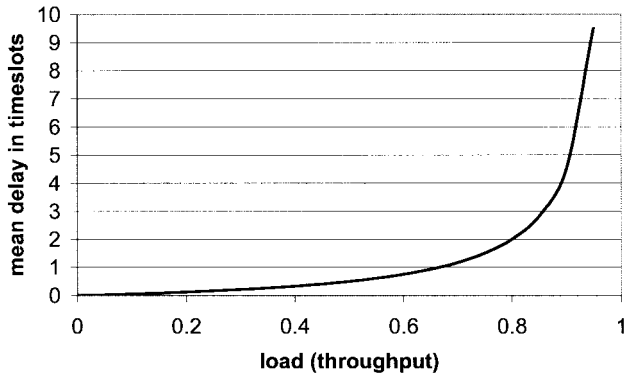


Fig. 3. Delay versus throughput performance of an arbitrarily large output buffered switch. The average packet delay increases as the load increases, rising indefinitely as the load increases towards one.

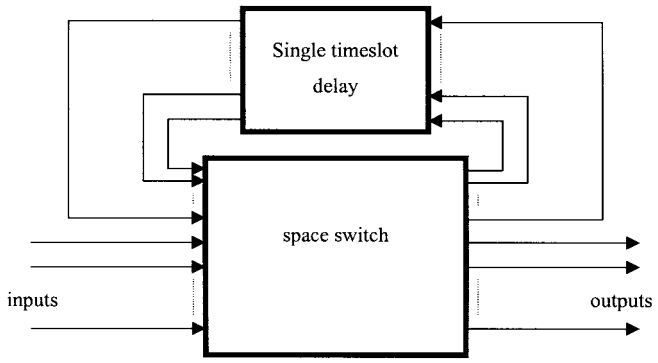


Fig. 4. The STARLITE switch. If multiple packets arrive at the inputs of the space switch destined for the same output, all but one is sent around to the single timeslot delay.

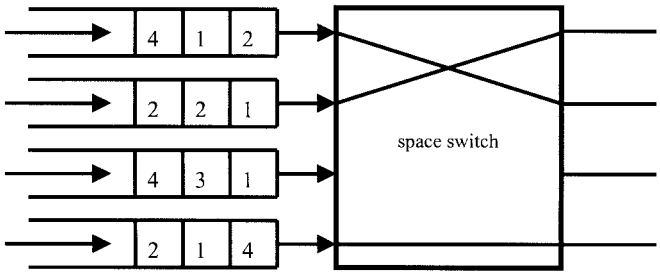


Fig. 5. An input buffered switch and head-of-line (HoL) blocking. This illustrative example has four inputs and outputs. The second packet in the third queue is prevented from reaching output three, even though it is free, since the packet ahead of it is experiencing contention for output 1.

in Fig. 3 [6], [7]; the delay tends to infinity as the throughput reaches unity.

Output buffering is the basis of many optical packet switches, although it is usual for an optical packet switch to emulate an output-buffered switch rather than have separate identifiable output buffers. Once a packet is in a delay-line, it is not possible to change its delay by removing it before it reaches the end; this suits output buffering where the delay in each output buffer can be determined before the packet enters it.

B. Shared Buffering

Electronic shared buffering may be regarded as a form of output buffering, where all the output buffers share the same RAM memory area [9]. Hence, the capacity restriction is not on the number of packets in each individual buffer but on the total number of packets in all buffers. Shared buffering is one of the most common methods of implementing electronic ATM switches, usually implemented as electronic random access memory (RAM). It is presently not possible to implement it optically in this form, because an equivalent optical RAM does not yet exist, and due to the complexity of the switch. However, many optical packet switches may be said to use shared buffering when emulating output buffering, since the delay-lines are shared among multiple output buffers. This will be discussed in detail later.

C. Recirculation Buffering

In recirculation buffering (for example, the STARLITE switch [10]—Fig. 4), a number of recirculation loops from the output of a space switch feed back into the input. Each loop has a delay of one packet. If more than one packet arrives at the space switch input for a particular packet switch output, all but one are placed into the recirculation loops.

For example, in a 64×64 switch for a loss of 10^{-10}, 237 recirculation loops are required if the load is 0.8 [7, Section III–D]. When implemented optically with unity delays, many recirculations are required, implying high loss and accumulation of amplifier noise in the loops (see the multiwavelength fiber loop in Section VI–C). To overcome these problems, one can have delays of different lengths, which will be discussed later.

D. Input Buffering

Input buffering is the final scheme that will be examined (Fig. 5) [11]; this consists of a space switch with buffers on the inputs. The fundamental difficulty with this is head-of-line (HOL) blocking, which limits the throughput to a maximum of 58% for uniform traffic [6]. For example, in Fig. 5, the second packet in the third queue (destined for output 3) cannot reach that output because the packet ahead of it in the queue, going to output 1, is blocked. This is because the second buffer, feeding into output 1, is already transmitting to that output.

Input buffering is never proposed for purely optical implementation, primarily because of its poor performance. It is possible to improve performance by selecting packets other than those at the head of a buffer to go to the output, but this is too complex for optical implementation. It is also not easy to realize optically since a packet's delay in an input buffer is not determined before entering it.

E. Approaches to Optical Packet Buffering

Various approaches are prevalent for optical buffering.

- *Imitate Electronics:* Here the objective is to implement large optical buffers, perhaps with header translation. This approach has been followed in the RACE ATMOS [12], [13] and ACTS KEOPS [1] programs, and also in an approach to obtain very deep buffers (SLOB [14]).
- *Deflection Routing:* Here there is no buffering [15]. If more than one packet arrives for a given output, all but one are *deflected*, i.e., they are sent to the wrong output

and must try to reach the correct destination node via an alternative route. This will not be discussed further, since, because it does not use optical buffering, it is outside the scope of this paper.

- *A Small Amount of Buffering with Deflection Routing:* This is a compromise between the above two approaches [16], [17]. If a buffer overflows, the packet is sent to the wrong output, and must find its way to the correct destination node (as with deflection routing).

In Section V, specific discussions on optical buffering begin. Various schemes will be compared, with regard to cell loss/delay but also taking into account optical performance in terms of crosstalk and noise. This latter topic is introduced in Section IV.

IV. Optical Performance Evaluation Methodology

The bit error rate (BER) performance of each optical buffering scheme is determined using analytical models, evaluated by computer. The analytical models for each device exist in the literature. In each case, a path through the packet switch is traced, and the noise and crosstalk components introduced in each device are modeled. Multiple interfering terms of low intensity (60 dB less than the main signal [18]) may be safely ignored. The gain of each switch architecture is adjusted to 0 dB; one or more extra amplifier stages may be added to facilitate this.

It will be assumed that fiber delay-line loss may be neglected. As an example to justify this assumption, assume that a relatively large buffer of 250 packets is to be implemented at a bit rate of 10 Gb/s. If ATM cells are buffered, then the buffer must hold $53 \times 8 \times 250 = 106\,000$ bits. Since each bit lasts 0.1 ns, this corresponds to a delay of 10.6 μs, i.e., a delay-line length of approximately 2.12 km. Assuming typical fiber loss, such a delay-line has insignificant attenuation which does not affect the models, so delay-line loss may be neglected. Variation in optical path length due to temperature variation may be overcome by thermal stabilization of the delay-lines.

Power penalties are obtained at a BER of 10^{-14}; this supports virtually error-free transmission and facilitates the support of a reasonable end-to-end BER. It represents the BER that would be required of a commercial product, and while higher BER's (e.g., 10^{-9}) might be acceptable for merely comparing performance, 10^{-14} will certainly be required in practice if these systems are to be commercially deployed. Electronic switches fundamentally have a zero BER and hence any new technology must approach this performance.

Semiconductor optical amplifiers (SOA's)—used for switching and amplification—introduce additional spontaneous noise [19], [20], that is detrimental and is represented in the model, resulting in eye closure when compared to the original signal. Noise power accumulates due to amplifier cascading, limiting the possible size of each switch and the number that may be cascaded. Devices such as combiners and AWG's (arrayed waveguide gratings—used to route packets according to their wavelength) introduce crosstalk, which degrades the signal and, for the worst case, is all assumed to be at the same wavelength and same linear polarization as the signal [18]. Several simplifying assumptions are made in the calculations:

- the SOA's have uniform gain over the optical bandwidth [19];
- the input signal to an SOA is centered in frequency over the passband of the SOA [19];
- second harmonic terms in the SOA's are neglected in the derivation [19];
- there is no SOA saturation due to spontaneous noise from previous SOA's, but there is gain saturation due to the signal—this derives from the observation that the signal power is much greater than the spontaneous noise;
- tunable wavelength converters have zero loss—this is achievable in reality [21];
- tunable wavelength converters do not add noise—this is justified since they have zero power penalty; although they do add noise, they improve the signal shape due to their nonlinear transfer characteristic [22];
- the AWG filter profile is not important; it is modeled by the loss when a signal goes to the correct output (2 dB) and when it goes to a wrong output ($35+2=37$ dB); the device has crosstalk of 35 dB less than the signal [23];
- the lasers are assumed to be stable with respect to frequency;
- dispersion shifted fiber is used for the delay-lines, avoiding unwanted dispersion effects;
- the input extinction ratio is 20 dB;
- each $1 \times n$ passive splitter or $n \times 1$ passive combiner has a loss of $10 \log_{10} n + 2$ dB (where 2 dB is a pessimistic value for the excess loss) thus there is a direct correlation between the number of output or inputs n and the loss;
- each SOA has a noise figure of 6 dB and a saturation power of 10 dBm;
- the combined coupling loss in and out of a SOA is 2 dB;
- the extinction ratio of the SOA's is 60 dB [24];
- each EDFA has a noise figure of 5 dB.

TABLE I
Categorization of Optical Delay-Line Buffering Strategies

	Single-Stage	Multi Stage
Feed Forward Delays	• OASIS • Broadcast and Select Switch	• Wave-Mux: scheduled center stage • Cascaded 2×2 switches • SLOB: cascaded OASIS switches
Feed Back Delays	• Multiwavelength Loop: unity delays • SMOP: non-unity delays	

V. Classification of Optical Buffering

Optical packet switches may be categorized in two fundamental ways (Table I).

- First, there can be a single stage of delay-lines or multiple stages. A single stage is generally easier to control, but as will be shown later, with multiple stages it is possible to economize on the amount of hardware required for large buffer depths.
- Second, the delays may be connected in either a feed-forward or feedback configuration.

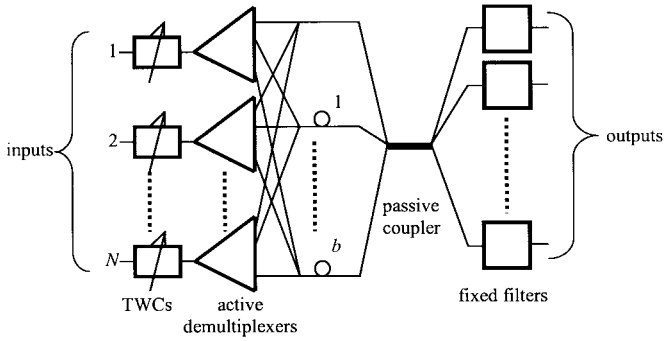


Fig. 6. An OASIS switch based upon a passive coupler. The tunable wavelength converters (TWC's) and the active demultiplexers ($1 \times (b+1)$ switches) select the appropriate output and delay line, respectively, for each incoming packet.

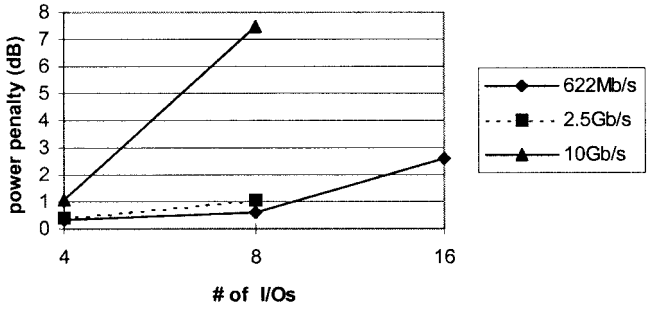


Fig. 7. Power penalty of single OASIS switches based on passive couplers. Due to greater splitting and combining losses, the power penalty rises with the number of inputs and outputs.

— In the former, each delay-line feeds forward to the next stage of the switch—traveling from one end of the switch structure to the other (by whatever route) involves a constant number of delay-line traversals.
— In a feedback configuration, a delay-line sends each packet back to the input of the same stage, implying that the number of delay-line traversals generally differs between packets.

Two types of single stage feed-forward switches are described here, namely, OASIS (from the RACE ATMOS program [12]—Fig. 6) and the broadcast-and-select switch (from ACTS KEOPS [25]—Fig. 9). The latter is a development of the former, which, as its name implies, allows broadcasting to take place. There are two single-stage feedback architectures—one (the multiwavelength loop [13]—Fig. 11) uses a unity delay whereas the other (SMOP-shared memory optical packet switch [26]) consists of a space switch with feedback delay-lines of differing lengths.

Multistage feedback architectures have rarely been proposed and none are considered here. Of the multistage feed-forward architectures, Wave-Mux [27] uses a central space switch with surrounding buffers, which schedule packets to be transmitted at the correct preassigned time over the space switch. A number of proposals have been published for linear arrays of cascaded 2×2 switches and delay-lines. These are discussed here, as is SLOB (switch with large optical buffers [14]—Fig. 22) which cascades OASIS switches to obtain very deep optical buffers.

VI. Single-Stage Switches

A. OASIS (Single Stage, Feed-Forward Delays)

The performance, scalability and cascadability of two types of OASIS switch [12] are studied here—one has a passive coupler, in conjunction with filters, to provide the routing function while the other reduces the amount of loss inherent in this configuration by using an arrayed waveguide (AWG) device. In common with several of the other switches that follow, OASIS emulates an output-buffered switch, by delaying each packet by the same number of timeslots as an output-buffered switch, and directing it to the correct output. This will be referred to as *output buffer simulation*, and is described below.

1) Output Buffer Simulation: In output buffer simulation, the switch is configured so that each packet entering it can suffer any delay between 0 and b timeslots, where b is the depth of each output buffer. Each packet may also be directed to any output, under the condition that no more that one packet may leave an output at once. To emulate an output-buffered switch, N counters, one associated with each switch output, calculate the packet delays. Each counter holds the number of packets in an imaginary FIFO (first-in–first-out) output buffer, decremented by 1 each time a packet leaves its output, then incremented by 1 for each arriving packet destined for its output. A packet leaves an output on each new timeslot, unless the corresponding counter is already zero, when the imaginary buffer is empty. If multiple packets arrive for an output on the same timeslot, they are assigned successive delays, corresponding to them being put into the output buffer one after another.

The total delay (in timeslots) that a packet experiences is the value of the designated output counter upon its arrival. When a counter is already set to b, arriving packets are discarded before entering the switch since buffer overflow has occurred. By delaying packets in this way, the switch effectively buffers them, since the same delay would be experienced in a real output buffered switch. Thus the OASIS switch has exactly the same performance as an output buffered switch with buffer depth b (Section III–A).

2) OASIS with a Passive Coupler: In this version of OASIS (Fig. 6), the TWC's (tunable wavelength converters) encode incoming packets with the wavelength corresponding to the filter at the desired output. Each packet then enters an optical switch (active demultiplexer) with one input and $b+1$ outputs which directs it to one of the delay-lines having length $1, 2, \cdots, b$ timeslots or to the null delay with an effective length of zero. The output buffer simulation algorithm discussed above calculates the packet delays; it is then relatively straightforward to determine the device control signals since the delay to be experienced by each packet is known.

Using the assumptions of Section IV, the optical power penalty of this architecture was determined (Fig. 7). Throughout this paper, it is assumed that the buffer depth is equal to the number of inputs and outputs, giving a good appreciation of optical performance without making the discussion overly complex. Hence, as the switch size (number of inputs and outputs) increases, the buffer depth will also increase and hence the packet loss rate will decrease. Also, this represents a likely

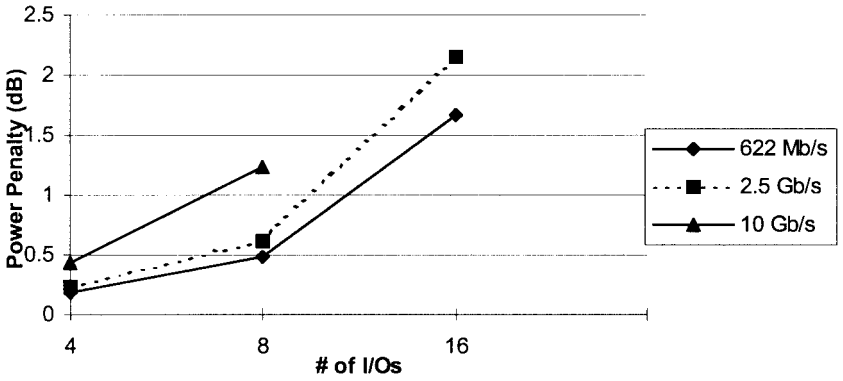


Fig. 8. Power penalty of OASIS switches based on AWG's. Again, the power penalty rises withthe number of inputs and bit rate, but is smaller than for the OASIS without an AWG.

scenario for demonstrator construction. Bit-rate influences the performance because SOA noise increases rapidly with bit rate. The performance was evaluated at 622 Mb/s, 2.5 Gb/s, and 10 Gb/s (this is done throughout the paper). Even for a 16×16 switch, a BER of 10^{-14} is not reached at 2.5 or 10 Gb/s, and the power penalty is unacceptably high (well over 2 dB) at 10 Gb/s for 8×8. Including the SOA already in the signal path as part of the active demultiplexer, two SOA's are in the signal path to overcome splitting and combining losses; these SOA's introduce noise, which is detrimental to the performance of the switch. SOA's are used throughout rather than EDFA's wherever possible, as they permit the possibility of integration, by means of silicon motherboards, for example. Under the assumptions of Section IV, they yield a worst-case power penalty due to the slight difference in noise figure between EDFA's and SOA's. As with all the buffering strategies in this paper, larger switches yield larger power penalties because of the greater need for amplification implied by the higher splitting and combining losses.

3) OASIS with an AWG: This is a modification of the previous architecture (Fig. 6), the passive coupler and filters being replaced by an AWG to reduce loss. It operates much as before, except that the tunable wavelength converters encode each packet with the wavelength for the desired output, corresponding to the delay-line that will be traversed (i.e., the AWG input that will be used). Fig. 8 shows the performance of this modified version of OASIS, a BER of 10^{-14} now being reached at 2.5 Gb/s with 16 inputs and outputs, and the power penalty at 10 Gb/s for an 8×8 switch being less than 1.5 dB. This is because lower ASE results at the lower SOA gain, required in order to overcome the loss in the signal path, which is lower due to the use of an AWG device.

With both versions of OASIS, it is not possible to implement packet priorities since a packet already in a queue (i.e., waiting in a delay-line) cannot be preempted by one with a higher priority. The OASIS architecture has the same packet loss and delay as an output-buffered switch. Other architectures considered in this paper could also be realized utilizing AWG's, although it is not always beneficial to do so. For example, with the multiwavelength loop switch of Section VI-C, the use of an AWG actually increases the power penalty, except for 622 Mb/s, where electronic switching would doubtless be a better option.

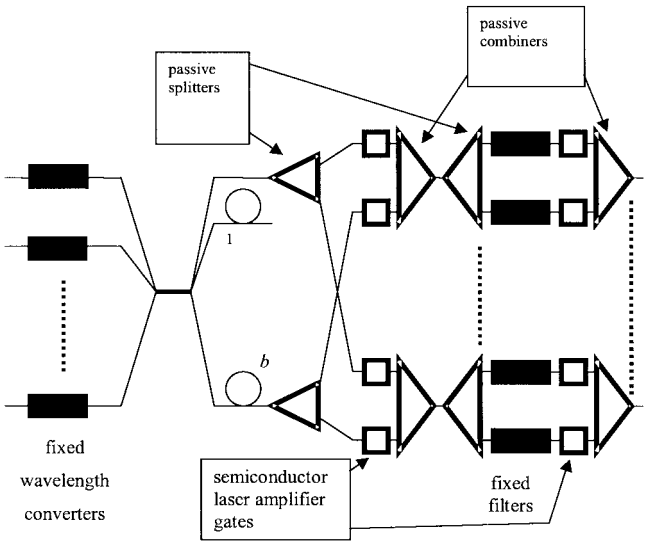


Fig. 9. The broadcast-and-select switch. The wavelength of each incoming packet determines which input it arrived on, so each output can select a packet on each timeslot corresponding to the correct delay-line (i.e., originating from the correct timeslot) and the correct input.

B. Broadcast-and-Select Switch (Single Stage, Feed-Forward Delays)

In the broadcast-and-select switch [25] (Fig. 9) fixed wavelength converters encode the packet streams entering each input; so packets on each input are distinguished by a separate wavelength. The streams are then combined and distributed to the delay-lines. By means of SOA switches and a passive combiner, each output can select the signal from one delay-line. Then the components at each output select, using a bank of filters, the packet from the correct input. In conjunction with output buffer simulation, this switch has the same performance as an output-buffered switch. Moreover, since all packets are broadcast to all outputs, with all possible time delays, it offers the possibility of broadcast operation, and implementation of packet priorities. The latter feature is possible because each packet is broadcast to all delay lines.

Fig. 10 shows the power penalties for various sizes of this switch, at bit rates of 622 Mb/s, 2.5 Gb/s, and 10 Gb/s. There are two SOA's in each path through the switch, and an EDFA is placed on each switch output to overcome losses.

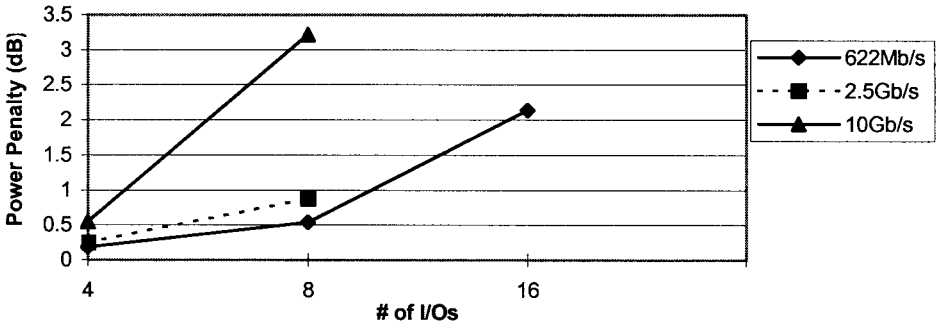


Fig. 10. Power penalty of single broadcast-and-select switches. The switch is only usable with four, eight, or 16 inputs and outputs at and below 10 Gb/s and 622 Mb/s, respectively.

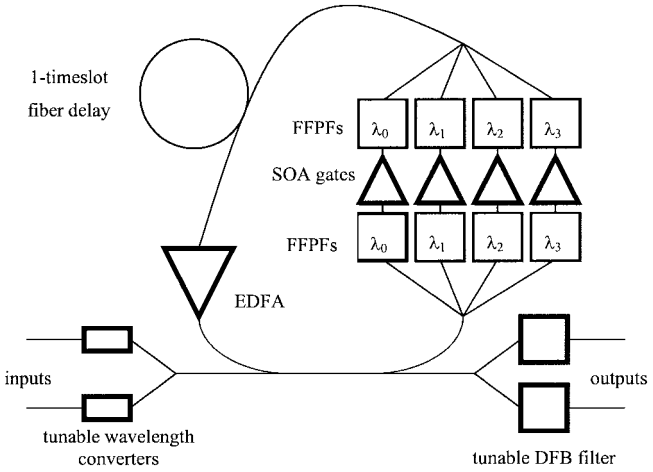


Fig. 11. The multiwavelength loop switch—this particular example has just two inputs and outputs with four wavelengths. FFPF = fiber Fabry–Perot filter. Each packet is assigned a wavelength and circulates in the loop until the required output is free. They are gated by the FFPF and the SOA gates.

For 622 Mb/s and 2.5 Gb/s, the performance is similar to the splitter/combiner based OASIS. At 10 Gb/s, the performance is appreciably worse than OASIS with an AWG but better than the splitter/combiner version of OASIS.

C. Multiwavelength Loop (Single Stage, Feedback Delays)

In this switch, multiple packets are stored on the same fiber feedback loop at different wavelengths (Fig. 11, [13], [28], [29]). Each TWC is tuned every timeslot so that the incoming packet does not contend with other packets when placed on the loop. Once a packet is to be removed from the loop, the tunable DFB filter at the relevant output is tuned to its wavelength, and one of the SOA's in the loop gate out the packets. Experiments have demonstrated the validity of a packet performing over ten recirculations at 622 Mb/s in a two-input, two-output configuration [28].

In addition to the assumptions of Section IV, it was assumed that the tunable filters have noise figures of 6 dB with gain from 21 to 25 dB. Fig. 12 shows the power penalty at both 2.5 and 10 Gb/s, which not surprisingly increases with the number of recirculations. Even for as few as seven recirculations, the power penalty is as large as 8 dB at 2.5 Gb/s.

This switch emulates an output-buffered switch, and yields the same cell loss and delay performance, with a buffer depth equal to the maximum permissible number of recirculations. It is possible for incoming packets to preempt those that are already waiting hence this type of switch can implement packet priorities.

D. SMOP (Single Stage, Feedback Delays)

SMOP (shared memory optical packet switch) [26] is a recirculating loop switch with delay-lines of length $1, 2, 3, \cdots, m$, and a central $(N + m) \times (N + m)$ space switch (compare with Fig. 4 which has unity delays throughout). Having delay-lines of length greater than one reduces the number of feedback loops and amplifiers, and reduces the size of the space switch. When multiple cells arrive at the space switch destined for a particular output, all but one are sent round the recirculating delay-lines. The control algorithm is based upon output buffer simulation, and proceeds on each timeslot as follows.

- Packets arriving from the delay-lines ready to go to an output are routed.
- Any packets arriving at the SMOP inputs that can go directly to the SMOP outputs are routed, providing this does not, in each case, violate the FIFO queuing discipline.
- If any packets can reach an output after just one recirculation, they are scheduled to go round just one recirculation and then exit the switch. Priority is given to packets with the shortest remaining delay.
- The remaining packets are allocated to delay-lines, trying to avoid having more than one packet destined for a particular output reaching the space switch at once after the recirculation.

For full details, refer to [26]. Throughout, packets are kept in a FIFO sequence, and, based upon simulation, the maximum number of recirculations required is ten. This architecture permits packet priorities, since a lower priority packet may be preempted by sending on another recirculation. The packet loss probability is close to that possible for shared memory switches. For $N = 8$ and $m = 13$, the packet loss is 10^{-6} at a uniform Bernoulli load of 0.9. At the same load, the same packet loss is obtained.[2] For $N = 16$ and $m = 18$. As with

[2] Figures for packet loss of less than 10^{-6} are not available [26].

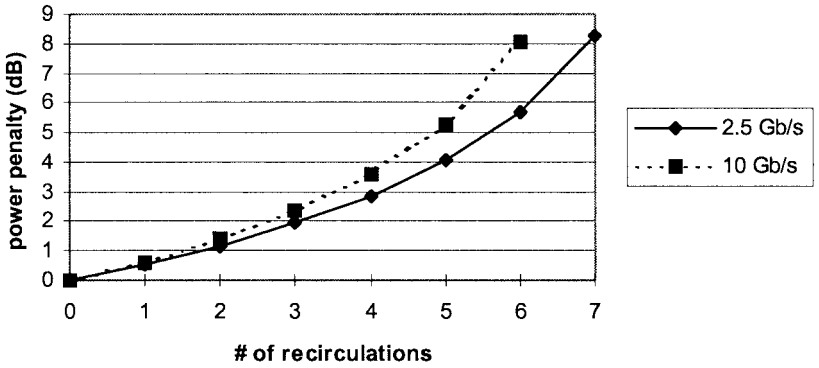


Fig. 12. Power penalties for an 8 × 8 multiwavelength loop switch with differing numbers of recirculations. As expected, the power penalty rises as the number of recirculations increases.

the multiwavelength recirculating loop switch, it is possible for higher priority packets to preempt those that are already waiting in the switch, thus permitting a priority scheme to be implemented.

In the simulations, one SOA in each loop and one on each output are required to compensate the losses within a stage, and $N = m$ throughout. Fig. 13 shows the power penalty at both 2.5 and 10 Gb/s, increasing with the number of recirculations. To maximize its practicality, this architecture must be implemented with a space switch design that has low noise and crosstalk. The space switch is implemented as a tree architecture [30], exhibiting excellent crosstalk performance. Each 1 × 2 or 2 × 1 switch has an extinction ratio of −35 dB and a loss of 2 dB. It is clear from these results that carrying out ten recirculations is entirely practical.

E. Conclusions on Single-Stage Switches

SMOP yields excellent power penalty performance, since its use of a space switch allows a clever architectural technique—the tree architecture—to be employed, reducing crosstalk. With OASIS, the use of AWG's also yields an improvement in performance, due to the reduction in loss. All the switches in this section have the same packet loss and delay performance as an output buffered switch (Figs. 2 and 3), except for SMOP which allows the delay-line memory to be shared and performs almost as well as a shared-buffer switch. Preemption of packets in the switch by higher priority packets is possible in all of these switches except for OASIS. This is because OASIS does not use recirculating loops and does not broadcast all packets over all its feed-forward delay-lines.

VII. Multistage Switches

This section is concerned with buffering schemes having multiple stages of delay-lines. Wave-Mux [27] uses electronics to perform the bulk of the buffering, with optical buffering performing the scheduling necessary to facilitate transport within the switch. The 2 × 2 packet switches are introduced [16], [17], [31]–[33] which may have certain specialized applications, the SLOB [14] being a generalization and extension of this concept.

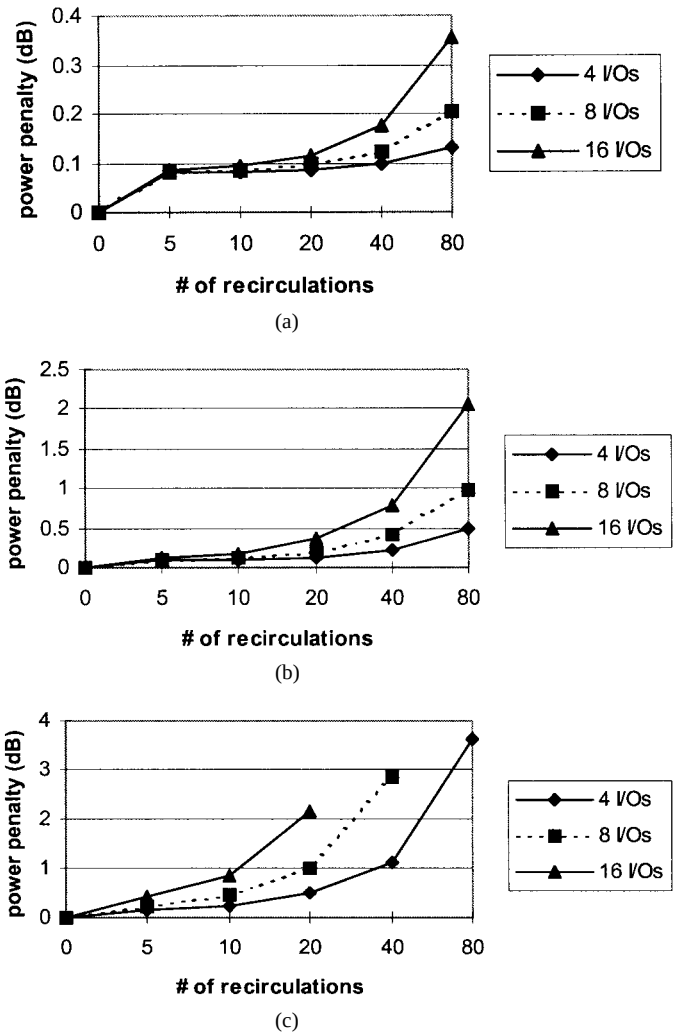


Fig. 13. Power penalties for the SMOP switch with differing numbers of recirculations and with bit rates of (a) 622 Mb/s, (b) 2.5 Gb/s, and (c) 10 Gb/s. Although only ten recirculations are required with the algorithm described in the text, the graphs indicate that a much larger number of recirculations is necessary to obtain a power penalty of over 2 dB.

A. Wave-Mux (Multistage, Feed-Forward Delays)

In Wave-Mux [27], a center-stage switch that changes state every timeslot, operates in a repeating, scheduled manner [Fig. 14(a)]. The inputs are divided up into groups, each of

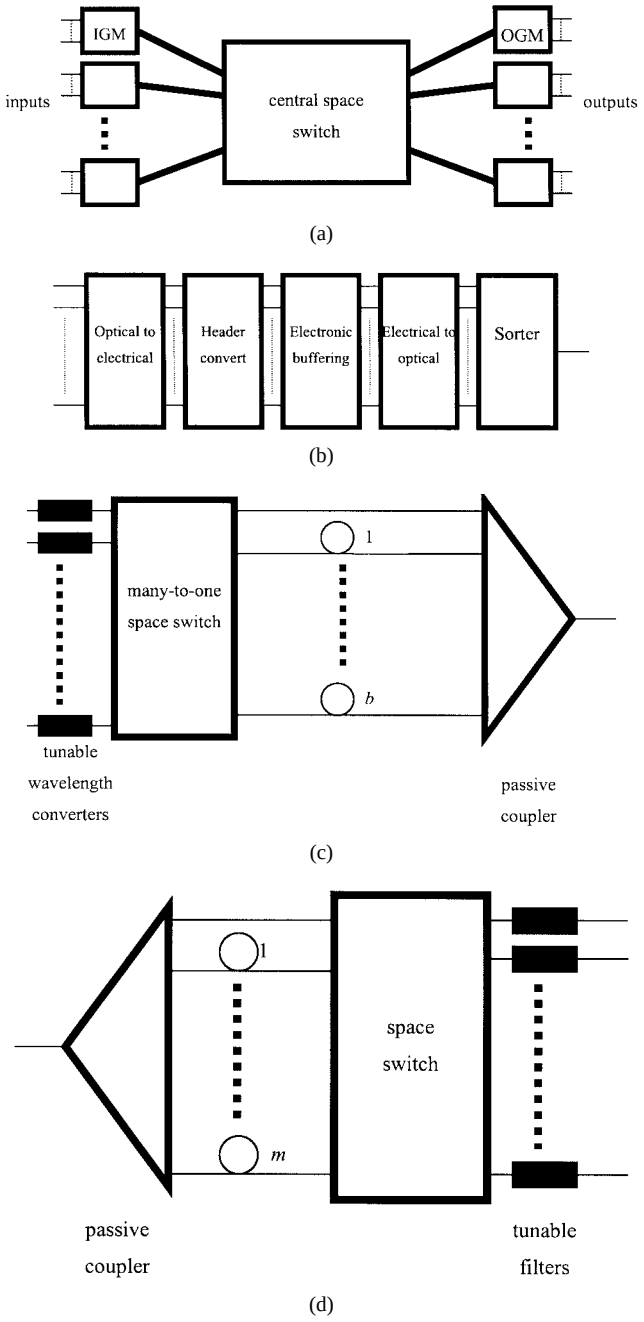


Fig. 14. (a) The Wave-Mux architecture block diagram. The central space between groups of inputs (each represented by an IMG or input group module) and groups of outputs (each represented by and OGM or output group module). Because there is no buffering in the central space switch, the IGM's must schedule the transmission packets. (b) A complete IGM. Packets are converted to electronic form prior to header translation and buffering. Sorting is necessary prior to sending each packet over the central space switch to avoid contention, since there is no buffering there. (c) A sorter in an IGM. Each packet is sent by the many-to-one space switch to the correct delay line so that it arrives at the central space switch at the right time. Wavelength conversion is necessary so that multiple packets can travel over the same fiber on the same timeslot. (d) An OGM operates much like an output of a broadcast-and-select switch. By using the space switch, each output is fed all the packets from the correct timeslot, and each tunable filter selects the appropriate packet at the correct wavelength.

which has its own input group module (IGM); likewise, each group of outputs has its own output group module (OGM). Each IGM/OGM pair is assigned a particular timeslot and the IGM's buffer incoming packets until they are transmitted over the center stage switch on the correct timeslot. Each cell on a given timeslot going over the central space switch on a specific path is given a different wavelength.

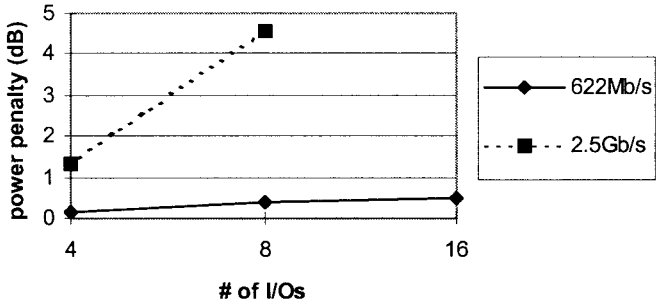


Fig. 15. Power penalty of single Wave-Mux switches. Each IGM has four inputs and each OGM has four outputs. The architecture does not operate at 10 Gb/s due to an error in flooring.

Packets entering an IGM are first converted to electronic form to undergo header conversion and buffering before being converted back to optical form [Fig. 14(b)]. Electronic buffering is used since a large amount of buffering is required to yield an acceptable packet loss performance. The buffer memory is organized so that multiple packets from one input can enter the sorter simultaneously [Fig. 14(c)], overcoming HOL blocking. The sorter is used to ensure that packets enter the center stage switch on the correct timeslot to go to the correct output. Wavelength converters allow multiple packets to traverse the central space switch over the same route simultaneously. The OGM [Fig. 14(d)] sends each packet to its correct output in sequence, by means of delay-lines, a space switch and tunable filters.

One advantage of Wave-Mux is that the size of space switches required is fairly small, even for a large packet switch. With 128 inputs and output, eight inputs on each IGM and 16 outputs on each OGM, 128 packets per line in the electronic buffers and a load of 0.8, a cell loss ratio of 10^{-10} is achieved. With these parameters, the cell loss ratio increases very rapidly for higher loads. Preemption by higher priority packets to implement packet priorities could be implemented in the electronic input buffers, however this has not been studied.

When modeling this architecture, one SOA gate and two EDFA's (one for 4×4 architectures) are required (per optical path) to compensate all the losses, and, as with SMOP, all space switches were assumed to be tree architectures [30]. In each IGM sorter, passive couplers are used at the output of the space switch to permit many-to-one operation. Fig. 15 shows how the power penalty of this architecture scales with respect to size and bitrate. The architecture does not yield a power penalty at 10 Gb/s, due to error flooring, amplified spontaneous emission (ASE) being the principal contributor to performance degradation. As expected, the power penalty increases with bitrate and size of switch.

The Wave-Mux switch uses optics and optical switching to provide interconnection and switching, rather than buffering. Due to the amount of optical manipulation (delaying and switching) required to implement these functions, the power penalty exhibited is in fact worse than for the other architectures considered here. However, this does not take account of

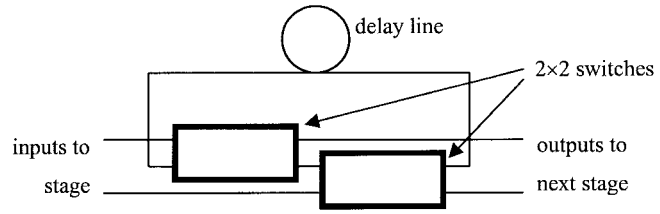


Fig. 16. A track changer (TC) module for COD. As many stages as are required of this module are cascaded, to obtain the desired packet loss. To preserve packet order, all delay lines must be of unity length; however, longer delay lines may be used whcih upset packet order, but with fewer stages.

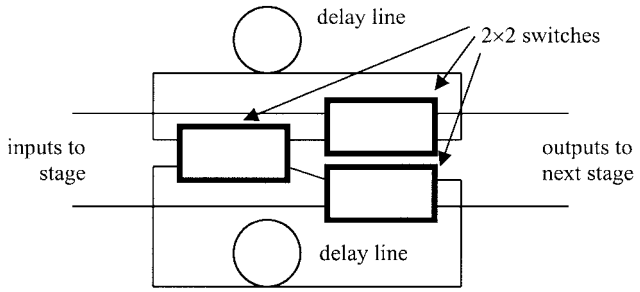


Fig. 17. A twin track changer (TTC) module for COD. The same comments apply as for Fig. 16, although the TTC yields lower packet loss than the TC for the same number of stages.

the size of the electronic buffer, which may be as large as required.

B. Cascaded 2 × 2 Switches (Multistage, Feed-Forward Delays)

A number of 2 × 2 optical buffered switching node architectures, consisting of cascaded 2 × 2 switch devices and delay-lines have been proposed [16], [17], [31]–[34]. Each 2 × 2 switch may change state between timeslots, under electronic control. While being too small to be used in a conventional telecommunications network, these may find application in rings and local area networks. The implementation of a packet priority mechanism in these switches has not been studied and would probably be difficult to implement due to the way packets are scheduled to pass over the switch at preordained times. Optical simulation results for these architectures are discussed in Section VII-B5.

1) COD (Cascaded Optical Delay-Lines): In COD [31], each 2 × 2 switching device sorts the packets entering it and is called a "smart crossbar," implementing distributed control (Figs. 16 and 17). A smart crossbar sorts packets destined for the upper output of the architecture to its upper output whenever possible; likewise for packets destined for the lower output of the architecture. Each stage is either a track changer (TC—Fig. 16) or a twin track changer (TTC—Fig. 17). Either type is connected in a chain, with a smart crossbar on the input. If the delays in each stage are unity length, then packet ordering is preserved. However, the buffer depth only grows linearly with the number of stages.

With nonunity delays, a more economical hardware implementation is possible, but the packet order is upset. A TTC chain has a lower cell loss than one using TC's, even for the same number of smart crossbars. For example, with 5 TC stages and delays of 1, 7, 49, 686, and 14 406, the packet loss is 10^{-11} for a load of 0.8, using a total of 11 smart crossbars. Four TTC modules, with delay-lines of length 1, 7, 98, and 4116 yield a packet loss of 10^{-24} for a load of 0.8, using 13 smart crossbars. Thus very few extra crossbars yield much better performance with TTC. The delay-line lengths were chosen to yield low values of packet loss and also to facilitate analysis.

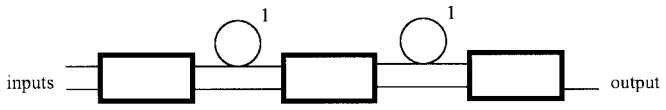


Fig. 18. The switched delay line (SDL) switch. If two contending packets arrive at the inputs, one is placed in the first delay, and if contention persists, one new packet is delayed in the second delay line.

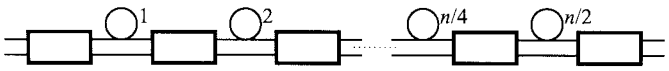


Fig. 19. The logarithmetic delay-line switch. It emulates an output buffered switch and the number of the packets in both emulated output buffers must always total $n - 1$.

2) SDL (Switched Fiber Delay-Lines): This 2 × 1 architecture [32] consists of two stages of delay-lines (Fig. 18), originally proposed for alleviating receiver contentions in WDM packet systems. It has been implemented with InP modal evolution switches and fiber delay-lines. If two contending packets arrive at the inputs, one is placed in the first delay, and, if contention persists, one new packet is delayed in the second delay-line. Numerous control strategies exist and it gives slightly better performance when the FIFO discipline is disrupted.

It has been proposed for use in the CORD (contention resolution by delay lines) project, where equal delay-lines of length 1 are used [32]. However, unequal delays of 1 and 10 have been shown to implement, for example, a packet loss of 0.0469 at a load of 0.5 on both inputs [33]. All packets entering both inputs are directed to one output since it is a 2 × 1 architecture. The packet switch is controlled so as to avoid more than one packet leaving both delay-lines at once. This is done by controlling what incoming packets are sent into the shorter delay-line, in view of what is already in the longer delay.

3) Logarithmic Delay-Line Switch: This structure [16] offers logarithmic growth of hardware requirements with buffer depth (Fig. 19). It employs output buffer simulation, and functions much as a 2 × 2 output-buffered switch but with nonoptimal delay; output buffer simulation is used to control the architecture. If the traffic load is 100% then it can be proven that it exhibits no internal blocking when used as an output-buffered switch [16]. However, 100% loading implies instability (i.e., "infinite" delay under statistical equilibrium—see Fig. 3). To circumvent this problem, each empty timeslot entering the switch is placed in the emptiest queue, forcing the system toward equilibrium and avoiding instability. The empty timeslots in the queues are responsible for the nonoptimal delay. Fig. 20 shows the packet loss (deflection) performance of this architecture.

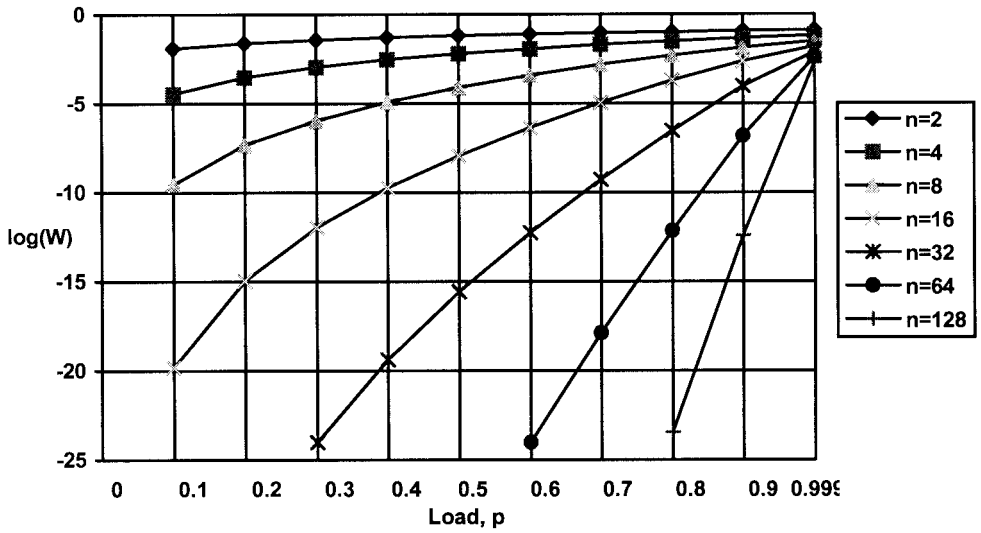


Fig. 20. Packet loss (deflection) performance of the logarithmetic delay line switch. Log(W) is the logarithm of deflection probability, and p is the load.

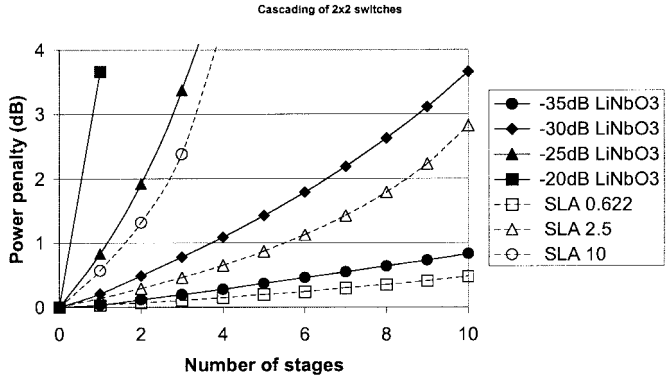


Fig. 21. Performance of a chain of 2 × 2 switches at622 Mb/s, 2.5 Gb/s, and 10 Gb/s for both SOA and lithium niobate technologies. The performance of the lithium niobate cascade does not alter with bit rate. The three traces for semiconductor laser amplifiers (SLA 0.622, SLA 2.5, and SLA 10) also show that high-performance switches are required to cascade many stages.

4) Single-Buffer Deflection Routing Switch: This switch [17], [34] is essentially a special case of the logarithmic delay-line switch, with $n = 2$, consisting of two 2 × 2 switches and one delay-line. The switch may be controlled slightly differently[3] so that the deflection probability is reduced slightly, although this effect is only noticeable for $n = 2$. Deployment of this single-buffer switch in regular networks has been extensively investigated [17]. Due to its very small buffer depth, this architecture is only suitable for applications where a very high deflection probability can be tolerated, and is hence still more specialized in application than the logarithmic delay-line switch itself, upon which it is based. Its performance is shown in Fig. 20, where $n = 2$.

5) Optical Modeling of Cascaded 2 × 2 Switches: All the switch architectures proposed in this section involve cascades of 2 × 2 switches so the graphs presented here (which depict the power penalty of such cascades) are helpful in determining performance. The results for cascading 2 × 2 switches are obtained at 622 Mb/s, 2.5 Gb/s, and 10 Gb/s (Fig. 21), covering both SOA and $LiNbO_3$ technologies. For SOA's, each SOA compensates all the insertion loss of 10 dB within each stage (predominantly due to coupling, but also due to bends, etc.). The input power is −20 dBm. The SOA's are assumed to exhibit negligible crosstalk, and perform in keeping with the assumption of Section IV. The lithium niobate switches have crosstalk varying from −20 to −35 dB, but loss is neglected since it may be overcome by amplification, and crosstalk is being focused upon here. The graphs show that high-performance switches are required to cascade many stages.

6) Conclusions on Cascaded 2 × 2 Switches: CORD and the single-buffer deflection routing switch are small switches, which can only be used for specific applications. COD offers less efficient hardware usage than the logarithmic delay-line switch. For example, a packet loss of 10^{-24} requires 13 smart crossbars in the TTC configuration, whereas the logarithmic delay-line switch requires only eight 2 × 2 switches for similar packet loss (Fig. 20). In addition, CORD introduces a large packet delay and does not preserve packet order.

C. Switch with Large Optical Buffers (SLOB) (Multistage, Feed-Forward Delays)

The switch with large optical buffers (SLOB) [14] (Fig. 22) cascades many small switches, forming a larger switch with a greater buffer depth. The OASIS switch is chosen as the basic element, although other switches could be used. SLOB is electronically controlled but has an optical packet data path, with N inputs and outputs. The delay-line lengths increase exponentially from left to right along the structure, and each buffer has a depth of $N^k - 1$ packets, where k is the number of stages in the architecture; hence the number of stages increases with the logarithm of buffer depth. It emulates an output-buffered switch, thus exhibiting optimal delay/throughput performance. Studies have shown that buffer depths of thousands are feasible [14], thus making it feasible to handle bursty traffic. Implementing a packet switch with a comparable buffer depth would not be feasible with the other designs discussed in this paper (apart from Wave-Mux, where the buffering is elec-

[3] For a discussion, see [16, the Appendix].

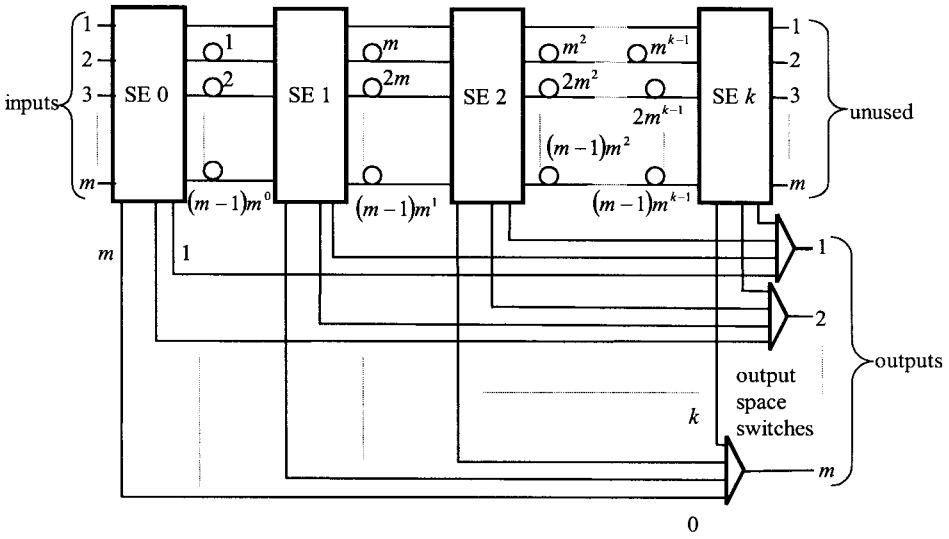


Fig. 22. The switch with large optical buffers (SLOB's). Each switching element (SE) is an optical packet switch, such as the OASIS switch. Once packets have experienced the correct delay, they are sent to the relevant output via the output space switches. The switch is controlled by output buffer simulation.

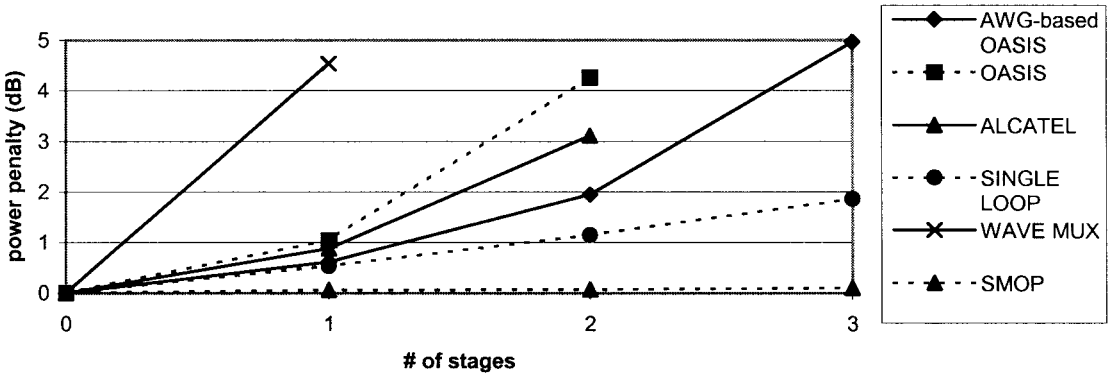


Fig. 23. A cascadability comparison at 2.5 Gb/s. The Wave-Mux switch gives the worst power penalty while the SMOP switch gives the best performance.

tronic). This is due to economic reasons, since cost does not scale well with buffer depth in these architectures, and also due to considerations of splitter power loss. SLOB offers logarithmic (and hence economical) growth of hardware with buffer depth, while still being feasible from the viewpoint of optical performance [14]. It has been shown that the architecture can be controlled using current technology [14]. Packet preemption and priorities are not possible with SLOB as it stands.

VIII. Conclusions

All the packet switches considered here, except for COD, CORD, SMOP, and Wave-Mux, simulate output-buffered switches and as such, have the same delay/throughput performance. Wave-Mux is essentially an input-buffered switch. Figs. 23 and 24 show how well these architectures cascade, when taking crosstalk and noise performance into account. The number of packet switches that are cascaded in a path or circuit is highly dependent on the network architecture and routing algorithm; discussion of this topic in any detail is outside the scope of this paper. Bearing in mind that a BER of 10^{-14} represents a very stringent operating condition, the results show in many cases that cascading the architectures is not practical. This provides a strong case for optical regeneration; indeed, researchers at Alcatel have shown the feasibility of cascading 40 broadcast and select switches by experiment [35], in conjunction with optical regeneration. Operation at 622 Mb/s is not considered, as this is well within the capabilities of electronics, indeed monolithic electronic ATM switches operating at 622 Mb/s have been reported [36]. The SMOP and Wave-Mux switches make use of internal space switches, which allow techniques to reduce crosstalk within the space switches to be employed, such as the tree architecture. This implies that SMOP in particular yields a very low power penalty. The SLOB switch is a technique for cascading existing optical packet switches, permitting very deep buffers (thousands of packets) to be produced with favorable optical performance and hardware utilization.

Table II summarizes the principal characteristics of the buffering architectures considered in this paper. Priorities and preemption of packets are not permitted in OASIS or any of the multistage switches. In many of the multistage architectures it is not clear how preemption would be implemented, and this is an area that remains to be studied. Many of the switches make internal use of wavelength, to facilitate internal switching and buffering. In this paper, the possibility of using multiple wavelengths for transport between nodes has

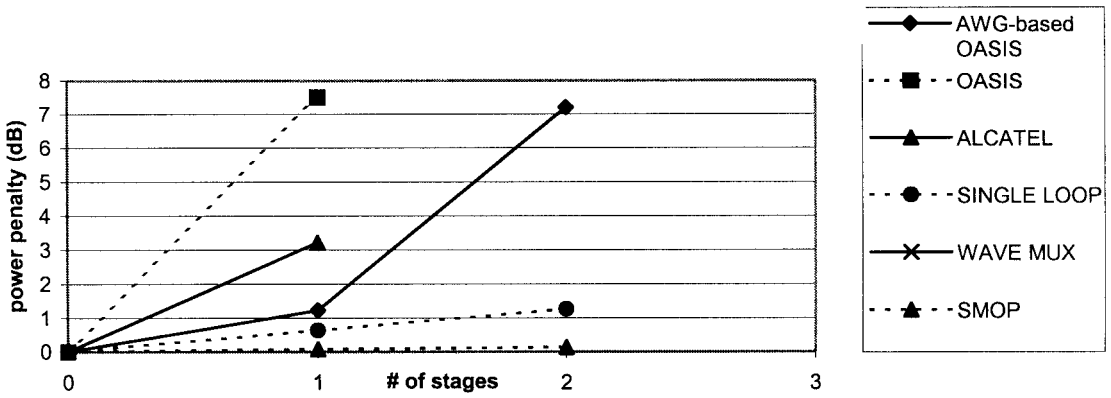


Fig. 24. A cascadability comparison at 10 Gb/s.

TABLE II
COMPARISON OF OPTICAL PACKET SWITCH ARCHITECTURES. FOR BUFFER SIZE, "SMALL" IS LESS THAN 10, "MEDIUM" IS 10–100, AND "LARGE" IS GREATER THAN 100

Switch	Buffer depth	Packet priorities	Internal use of wavelength	Control
OASIS (AWG)	Medium	No	Yes	Output buffer simulation
Broadcast and select	Medium	Yes	Yes	Output buffer simulation
Multiw'length loop	Small	Yes	Yes	Shared buffering
SMOP	Medium or large	Yes	No	Output buffer simulation plus scheduling
Wave-Mux	Large	No	Yes	Scheduled packet transport over central space switch
CORD	Small	No	No	Various
COD	Medium or large	No	No	Self-routing
Logarithmic Delay-Line	Medium or large	No	No	Output buffer simulation
SLOB	Large	No	Yes	Output buffer simulation

not be considered; this would entail a modification of the architectures considered here.

For general small to medium buffer depth applications, the OASIS or broadcast and select switches perform well. (Note that the AWG version of OASIS was used in the table due to its superior performance.) Due to its use of space switching, and the availability of techniques to minimize crosstalk in space switches, SMOP may be useful when medium to large buffer depths are required. Wave-Mux is useful for constructing large switches, with a large number of inputs and outputs, although the use of electronic buffering defeats some of the objectives discussed in the introduction. For specialized applications requiring two inputs and outputs, the logarithmic delay-line switch (and the single-buffer deflection routing switch) offer the ability to construct buffers while using hardware efficiently. SLOB may be useful if very large buffer depths of thousands are required.

In conclusion, it is clear that many differing proposals have been presented for undertaking optical packet switching. Many of the details concerning how these switches might be used, and what their impact will be on the network topology, routing, control and services, remains to be studied in depth. Finally, it is worth remembering that despite the progress described in this paper, the advances in high-speed electronic packet switching—both current and projected—are impressive with single-chip 16×16 ATM switches with 622 Mb/s ports [36], and three-chip 8×8 switches at 2.5 Gb/s [37], having been reported.

ACKNOWLEDGMENT

The authors wish to thank P. Ball and I. Wright of the Fujitsu Europe Telecom R&D Center and R. Gardner of the University of Strathclyde, U.K., for their comments on an earlier versions of the manuscript.

REFERENCES

[1] C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clérot, S. L. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P. B. Hansen, D. K. Hunter, A. Kloch, R. Krähenbühl, B. Lavigne, A. Le Corre, C. Raffaelli, M. Schilling, J.-C. Simon, and L. Zucchelli, "Optical packet switching: The European ACTS KEOPS project approach," *J. Lightwave Technol.*, this issue, pp. 2117-2134.

[2] D. B. Sarrazin, H. F. Jordan, and V. E. Heuring, "Fiber optic delay line memory," *Appl. Opt.*, vol. 29, no. 5, pp. 627–637, Feb 10, 1990.

[3] J. R. Bonar and J. S. Aitchison, "Co-doping effects in rare-earth-doped planar waveguides," in *Proc. IEEE Optoelectron.*, Oct. 1996, vol. 143, no. 5, pp. 293–297.

[4] A. Franzen, D. K. Hunter, and I. Andonovic, "Synchronization in optical packet switched networks," in *Proc. WAON'98, Workshop All-Optical Networks*, Zagreb, Croatia, May 6–8, 1998.

[5] R. O. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues.* Norwood, MA: Artech House, 1994.

[6] M. J. Karol and M. G. Hluchyj, "Input versus output queuing on a space-division packet switch," *IEEE Trans. Commun.*, vol. 35, pp. 1347–1356, Dec. 1987.

[7] M. G. Hluchyj and M. J. Karol, "Queuing in high-performance packet switching," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 1587–1597, Dec. 1988.

[8] J. Dupraz, "ATM: Current status and perspectives of evolution," in *Proc. ECOC'94*, Firenze, Italy, Sept. 26–29, 1994, pp. 555–562.

[9] R. H. Hofmann and R. Muller, "A multifunctional high-speed switch element for ATM applications," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1036–1040, July 1992.

[10] A. Huang, "STARLITE: A wideband digital switch," in *Proc. GLOBECOM'84*, Atlanta, GA, paper 5.3, pp. 121–125.

[11] J. Hui and E. Arthurs, "A broadband packet switch for integrated transport," *IEEE J. Select. Areas Commun.*, vol. 5, pp. 1264–1273, Oct. 1987.

[12] J. M. Gabriagues and J. B. Jacob, "OASIS: A high-speed photonic ATM switch—Results and perspectives," in *Proc. 15th Int. Switching Symp. (ISS'95)*, Berlin, Germany, Apr. 1995, paper C8.4, pp. 457–461.

[13] G. Bendelli, M. Burzio, P. Gambini, and M. Puleo, "Performance assessment of a photonic ATM switch based on a wavelength-controlled fiber loop buffer," in *Proc. OFC'96*, pp. 106–107.

[14] D. K. Hunter, W. D. Cornwell, T. H. Gilfedder A. Franzen, and I. Andonovic, "SLOB: A switch with large optical buffers for packet switching," *J. Lightwave Technol.*, vol. 16, pp. 1725–1736, Oct. 1998.
[15] D. J. Blumenthal, R. J. Feuerstein, and J. R. Sauer, "First demonstration of multihop all-optical packet switching," *IEEE Photon. Technol. Lett.*, vol. 6, pp. 457–460, Mar. 1993.
[16] D. K. Hunter, D. Cotter, R. B. Ahmad, W. D. Cornwell, T. H. Gilfedder, P. J. Legg, and I. Andonovic, "Buffered switch fabrics for traffic routing, merging, and shaping in photonic cell networks," *J. Lightwave Technol.*, vol. 15, pp. 86–101, Jan. 1997.
[17] F. Forghieri, A. Bononi, and P. R. Prucnal, "Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks," *IEEE Trans. Commun.*, vol. 43, pp. 88–98, Jan. 1995.
[18] P. J. Legg, M. Tur, and I. Andonovic, "Solution paths to limit interferometric noise-induced performance degradation in ASK/direct detection lightwave networks," *J. Lightwave Technol.*, vol. 14, pp. 1943–1954, Sept. 1996.
[19] A. Olsson, "Lightwave systems with optical amplifiers," *J. Lightwave Technol.*, vol. 7, pp. 1071–1082, July 1989.
[20] J. Yao, "Impact of input signals on BER performance of semiconductor laser amplifier based switch systems," in *Proc. IEEE Optoelectron.*, Aug. 1994, vol. 141, no. 4, pp. 265–270.
[21] S. L. Danielsen, C. Joergensen, M. Vaa, B. Mikkelsen, K. E. Stubkjaer, P. Doussiere, F. Pommerau, L. Goldstein, R. Ngo, and M. Goix, *Electron. Lett.*, vol. 32, no. 18, pp. 1688–1690, Aug. 29, 1994.
[22] B. Mikkelsen, S. L. Danielsen, C. Joergensen, R. J. S. Pedersen, H. N. Poulsen, and K. E. Stubkjaer, "All-optical noise reduction capability of interferometric wavelength converters," *Electron. Lett.*, vol. 32, no. 6, pp. 566–567, Mar. 14, 1996.
[23] H. Yamada, K. Takada, Y. Inoue, K. Okamoto, and S. Mitachi, "Low-crosstalk arrayed-waveguide grating multi/demultiplexer with phase compensating plate," *Electron. Lett.*, vol. 33, no. 20, pp. 1698–1699, 1997.
[24] A. Kelly, BT Labs, personal communication, Feb. 1997.
[25] F. Masetti, M. Sotom, D. De Bouard, D. Chiaroni, P. Parmentier, F. Callegati, G. Corazza, C. Rafaelli, S. L. Danielsen, and K. E. Stubkjaer, "Design and performance of a broadcast and select photonic packet switching architecture," in *Proc. European Conf. Optic. Commun. (ECOC'96)*, Oslo, Norway, Sept. 15–19, 1996.
[26] M. J. Karol, "Shared-memory optical packet (ATM) switch," in *Proc. SPIE: Multigigabit Fiber Commun. Syst.*, 1993, vol. 2024, pp. 212–222.
[27] Y. Nakahira, H. Inoue, and Y. Shiraishi, "Evaluation of photonic ATM switch architecture—Proposal of a new switch architecture," in *Proc. 15th Int. Switching Symp. (ISS'95)*, Apr. 1995, vol. 2, pp. 128–132.
[28] M. Calzavara, P. Gambini, M. Puleo, B. Bostica, P. Cinato, and E. Vezzoni, "Simultaneous buffering of ATM packets in a multiwavelength optical-fiber-loop memory," in *OFC'94*, San Jose, CA, Feb. 20–25, 1994, paper ThD2, pp. 195–196.
[29] G. Bendelli, M. Burzio, M. Calzavara, P. Cinato, P. Gambini, M. Puleo, E. Vezzoni, F. Delorme, and H. Nakajima, "Photonic ATM switch based on a multiwavelength fiber loop buffer," in *Proc. OFC'95*, San Diego, CA, Feb.–Mar. 1995.
[30] R. A. Spanke, "Architectures for large nonblocking optical space switches," *IEEE J. Quantum Electron.*, vol. 22, pp. 964–967, June 1986.
[31] R. L. Cruz and J.-T. Tsai, "COD: Alternative architectures for high-speed packet switching," *IEEE/ACM Trans. Networking*, vol. 4, pp. 11–21, Feb. 1996.
[32] I. Chlamtac, A. Fumagalli, L. G. Kazovsky, P. Melman, W. H. Nelson, P. Poggiolini, M. Cerisola, A. N. M. M. Choudhury, T. K. Fong, R. T. Hofmeister, C.-L. Lu, A. Mekkittikul, D. J. M. Sabido, IX, C.-J. Suh, and E. W. M. Wong, "CORD: Contention resolution by delay lines," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 1014–1029, June 1996.
[33] I. Chlamtac, A. Fumagalli, and C.-J. Suh, "A delay line receiver architecture for all-optical networks," in *Proc. INFOCOM'96*, San Francisco, CA, paper 4a.3
[34] A. Bononi, F. Forghieri, and P. R. Prucnal, "Throughput limitations in ultrafast all-optical soliton mesh networks using deflection routing," in *Proc. OFC/IOOC'93*, San Jose, CA, Feb. 21–26, 1993, paper ThI2, pp. 197–198.
[35] D. Chiaroni, B. Lavigne, L. Hamon, A. Jourdan, F. Dorgeuille, C. Janz, E. Grard, M. Renaud, R. Bauknecht, C. Graf, H. P. Schneibel, and H. Melchior, "Experimental validation of an all-optical network based on 160 Gbit/s throughput packet switching nodes," in *Proc. 24th European Conf. Optic. Commun. (ECOC'98)*, Madrid, Spain, Sept. 20–24, 1998.
[36] D. Weil, A. Botta, A. Chemarin, P. Gallay, J. Majos, and M. Servel, "A 16 × 622 Mb/s ATM switch: PRELUDE switch architecture integrated into a 6-million transistor monochip," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1108–1114, July 1997.î
[37] H. Yamada, M. Tsunotani, F. Kaneyama, and S. Seki, "20.8 Gb/s GaAs LSI's self-routing switch for ATM switching systems," *IEEE J. Solid-State Circuits*, vol. 32, pp. 31–37, Jan. 1997.

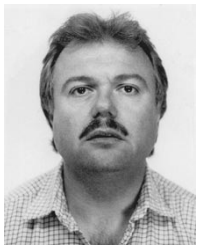
David K. Hunter (S'88–M'90) was born in Glasgow, Scotland, U.K. He received the B.Sc. degree in electronics and microprocessor engineering with first class honors from the University of Strathclyde, Scotland, in 1987 and the Ph.D degree from the same university in 1991 for research into switch architectures for optical TDM switching.

Since 1991, he has been a Research Fellow, then Senior Research Fellow, at the University of Strathclyde. From 1992 to 1993, he was a Visiting Researcher at BT Laboratories, Ipswich, U.K., and in 1994, he was a Visiting Scholar at the University of Pittsburgh, Pittsburgh, PA. Since 1995, he has been holding a five-year EPSRC Advanced Fellowship. His research interests include optical switch architectures, telecommunications switch modeling, and network routing and survivability.

Dr. Hunter is an Associate Member of the Institution of Electrical Engineers (IEE).

Meow C. Chia was born in Malaysia in 1970. She received the Bachelor's degree in electronic and electrical engineering with first class honors from the University of Strathclyde, Scotland, U.K., in 1995. She is currently working towards the Ph.D. degree in electronic and electrical engineering at the University of Strathclyde.

Her interests include optical packet switching node design, packet loss performance, and traffic analysis.



Ivan Andonovic (S'79–SM'97) has been with the Department of Electronic and Electrical Engineerng, University of Strathclyde, Scotland, U.K., since 1985, and is currently a Professor of Broadband Networks. Previously, he was a Research Scientist for three years at Barr and Stroud, where his responsibilities included design, manufacturer, and test of guided wave devices for a variety of applications. His main interests center on the development of guided wave architectures for implementing optical signal processing, optical switching and routing as applied in next-generation optical networks. He held a two-year Royal Society Industrial Fellowship in collaboration with BT Laboratories, Ipswich, U.K., during which time he was tasked with investigating novel approaches to optical networking. He has edited two books and authored/coauthored five chapters in books and over 140 journal and conference papers.

Dr. Andonovic is a fellow of the Institution of Electrical Engineers (IEE) and a member of the Optical Society of America (OSA). He has been Chairman of the Institution of Electrical Engineers (IEE) Professional Group E13, has held a BT Laboratories Short-Term Fellowship, and is Editor of the *International Journal of Optoelectronics*.