2×2 Buffered Switch Fabrics for Traffic Routing, Merging, and Shaping in Photonic Cell Networks

David K. Hunter, *Member, IEEE,* David Cotter, R. Badlishah Ahmad, W. David Cornwell, Tim H. Gilfedder, Peter J. Legg, and Ivan Andonovic, *Member, IEEE*

***Abstract*—An approach to optical packet switching is discussed, which uses small, simplified optical elements for traffic routing, merging, and shaping. The elements are constructed from 2×2 switches and optical delay lines, and may be implemented in a variety of technologies. They are designed for use with deflection routing, and even when using only six switches in a module, a deflection probability of 2.8×10^{-7} is possible with a load of 0.8. The modules may also be used as 2×1 mergers, where a deflection probability of 10^{-12} is possible with six switches and a total load of 0.8. The BER performance of the modules is simulated with respect to crosstalk, with even relatively poor switch devices of -18.5 dB isolation yielding a power penalty of less than 1 dB. A networking strategy radically different from today's is discussed, driven by the need to reduce hardware, software and operating costs.**

I. Introduction

OPTICAL PACKET switching fabrics have been proposed to overcome the anticipated problems of scaling electronic telecommunications switches to large sizes and capacities; EMI becomes a problem at very high speeds, the size of switches being restricted by chip pinout. One solution is to replicate the functionality of an electronic switch in optics [1]; here, an alternative approach is discussed using small simple optical nodes [2] in a network radically different from the present telecommunications network.

Such a network would [3]:

- be composed of elements with simple functionality;
- use simple control systems that are minimal, localized and autonomous;
- exhibit a degree of self-organization;
- be based on a pragmatic combination of photonics and high-speed electronics;
- have little or no software base;
- have higher intelligence distributed to the periphery.

This paper concentrates on the design of small simplified switching nodes that would be used in such a photonic network, which is assumed to carry fixed-length packets or cells. The switching nodes perform the dual functions of contention resolution and routing, and may be used either as 2×2 buffered switches or to merge two streams of traffic. Because the amount of buffering that can be used inside the switching nodes is limited in practice, strategies are needed to cope with buffer overflow; one strategy to prevent cell loss directs the overflow cells to the wrong switch output. The packet then finds its way to the correct destination by another route. This strategy, known as deflection routing [4], [5], is well suited to data traffic, particularly in a local-area network. However, for networks that extend over a wider geographical area, deflection routing suffers the penalty of increased delay which may be highly significant for delay-sensitive traffic such as that serving real-time broadband applications. For such traffic, it may be acceptable and preferable to tolerate occasional cell loss in order to minimize delay. The geographical diameter of the network beyond which delay becomes significant depends not only on the traffic type and application, but also on the bandwidth employed. As the bandwidth used by an application is increased, the effect of delay can become more pronounced. Conventional data traffic is less sensitive to delay, but the loss of a single cell may be sufficient to corrupt an entire message or file. Therefore, depending on the geographical size of the network, the bandwidth and traffic type, a choice must be made between cell loss or delay as the penalty for buffer congestion.

While buffered optical switch modules composed of 2×2 optical switches and delay lines have been proposed before [6]–[9], the importance of our design is that it is readily scaleable to large buffer depths with economical (logarithmic) growth in the amount of hardware needed. The new method of analysis presented here, based on recursive decomposition of the buffered switch structure, applies to switch modules of any size. We also present, for the first time, the results of simulations giving the power penalty of such switch modules, taking into account the effects of homodyne interferometric interference.

Section II describes the elementary 2×2 buffered switch which uses deflection routing to resolve contention. The deflection probability is analyzed, and also the power penalty in various fabrics under differing conditions are determined. In Section III, the modification of these modules for merging two traffic streams is discussed; again, the traffic performance and power penalties are modeled. It is also shown how merging modules can be combined to form an alternative 2×2 buffered switch architecture which guarantees no cell deflections, at the expense of finite cell loss. Section IV shows how the merging

Manuscript received February 5, 1996; revised September 3, 1996. The work of D. K. Hunter, P. J. Legg, and T. H. Gilfedder was supported by BT Labs.

D. K. Hunter, R. B. Ahmad, W. D. Cornwell, T. H. Gilfedder, and I. Andonovic are with the Department of Electrical and Electronic Engineering, University of Strathclyde, Glasgow G1 1XW, Scotland.

D. Cotter is with the Networks Technology Centre, BT Laboratories, Martlesham Heath, Ipswich IP5 7RE, UK.

P. J. Legg is with Motorola GPD, Swindon SN5 8YQ, UK.

Publisher Item Identifier S 0733-8724(97)00695-6.

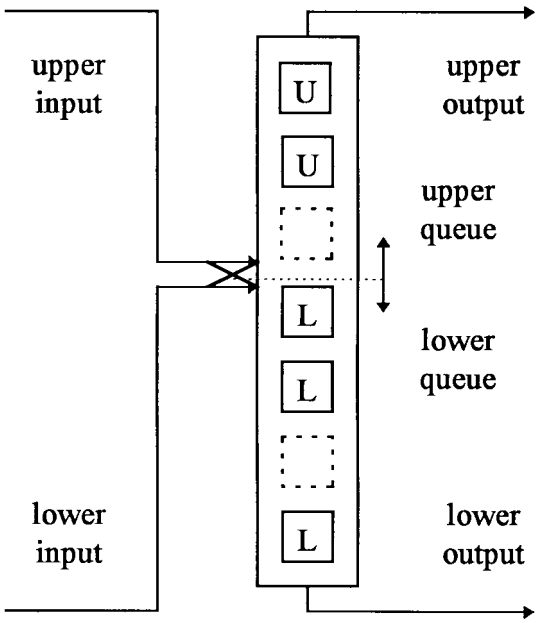

Fig. 1. Representation of the function of a 2×2 optical contention resolution and buffering module. There are two FIFO queues; one queue is associated with the upper (U) output and the other is associated with the lower (L) output. The two FIFO queues are of variable length, but they share a common fixed length buffer and the sum of their lengths is equal to the size of the buffer. The whole module thus functions as a 2×2 routing switch.

module may be used for shaping, to avoid congestion in critical parts of the network and to perform policing functions. Finally, Section V contains the conclusions. Proofs and analysis are confined to the Appendixes, although they are essential to this paper.

II. Binary Routing Switches

This Section describes the optical contention resolution and buffering module used throughout the network. The architecture consists of a chain of 2×2 optical switches and delay lines, and is quite similar in function to a 2×2 routing switch with two "first-in first-out" (FIFO) output queues sharing a single buffer (Fig. 1). The two input traffic streams are assumed to be synchronized in time so that, at each input, either one or no cell may appear in a given time slot. The time slot consists of the duration of one cell plus a time guard band during which the optoelectronic switches are reconfigured as necessary. The modules can be made from lithium niobate directional coupler switches [10] or indium phosphide laser amplifier switches [11], together with optical fiber delay lines [12]. The architecture has a low component count which is logarithmically related to the buffer depth; to double the depth of the buffers, only one extra switch is required. Furthermore, it may be controlled electronically, using a simple control architecture consisting of a counter and associated random logic (Appendix A). Feedforward delays are used throughout, implying uniform attenuation and superior crosstalk performance [13].

The switches change state only at the cell rate. Thus, with the inclusion of a suitable guard band, the electronic control rate is effectively much lower than the optical data rate. Owing to the transparent nature of optical switches, the data rate is not limited in principle, and data rates of 100Gb/s should be possible if attention is paid to synchronization issues [14], [15]. The data rate can be increased while keeping the delay lines (and hence also the delays) the same size simply by increasing the number of bits per timeslot. At such high data rates, optical header recognition [16] must be used.

A. Architecture

The architecture is referred to as $B(n)$ where $n-1$ is the depth of the shared output buffer. For example, in Fig. 1 the shared buffer depth is 7. Cells arriving at the inputs of the switching module are designated by their preference to emerge on the upper (U) or lower (L) output (in Fig. 1 cells are labeled "U" or "L" according to their routing preference). The shared buffer provides two FIFO queues, consisting of one queue for each of the two outputs. The queues are so depicted to emphasize that the sum of their lengths is always equal to the buffer size. A "payload" is defined as the contents of a timeslot i.e. either a blank space, a "U" cell or an "L" cell. On each timeslot, either:

- one payload enters the upper queue and the other payload enters the lower queue,
- both payloads enter the upper queue, or
- both payloads enter the lower queue.

This is determined by the routing preferences of the cells entering the architecture. Because blank spaces cannot be discarded when placing payloads in each queue, the module does not behave exactly as a classical output-buffered switch, and cells may experience additional delay due to the blank positions ahead of them in the queue. Also, in each time slot the upper output emits the cell at the head of the upper queue and the lower output emits the cell at the head of the lower queue. Each queue can contain up to $n-1$ payloads, but the combined number of payloads in the two queues is always equal to the shared buffer size $n-1$; if one queue becomes full the other must become empty. However, if one of the output queues becomes filled (for example, if all of the $n-1$ slots in the shared buffer contains a cell destined for the upper output), then an overflow could occur in the next time slot if two input cells are both forwarded to the full queue. This contention is resolved by deflecting one of the input cells to the wrong output queue (in the example described, both of the inputs are "U" cells that would prefer to emerge on the upper output, but one of them is sent to the lower queue). This strategy avoids cell loss (the shared buffer never overflows), but this is achieved at the expense of misrouting a cell.

Each module consists of a chain of $\log_2\ n+1$ switches and delay lines, and may be defined recursively (Fig. 2). Some examples of these fabrics, up to size $B(8)$, are shown in Fig. 3; the structure of $B(8)$ may be deduced by first starting with $B(1)$, then applying the recursive construction of Fig. 2 three times to produce $B(2)$, $B(4)$, and finally $B(8)$. To appreciate how these modules perform the buffered switching function, refer to the Appendixes.

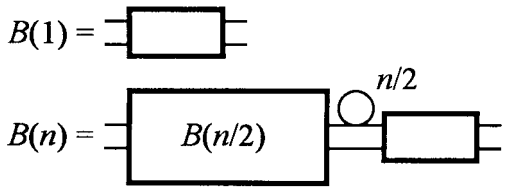

Fig. 2. Definition of $B(n)$.

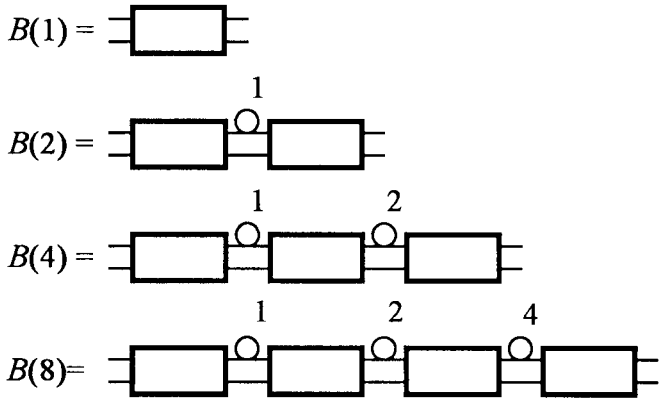

Fig. 3. Examples of buffered switching modules.

B. Deflection Performance

The control logic system for the buffered switch module reads the header address on each cell as it enters the switch, determines the preferred routing direction, and directs the cell to the appropriate output port. A given cell may be deflected to the wrong output with a certain probability W corresponding to a queue overflow. Markov chain analysis is used to evaluate the following deflection (routing error) probability (Appendix B):

$$W = \frac{p}{8}\frac{a-1}{a^{n/2}-1}a^{n/2-1}$$

where

$$a = \frac{p^2}{p^2+4p+4}$$

and p is the load (i.e. the probability that a timeslot is full); the distribution of traffic is assumed to be uniform. The assumption of uniform traffic is at worst, a useful first approximation which gives an over-optimistic estimate of deflection probability for asymmetric or bursty loads. However, it is an adequate model for traffic composed of messages emanating from a large number of users, provided that the instantaneous activities of those end users are mainly uncorrelated, that the traffic is well mixed, and that the traffic load associated with any message is small compared to the total capacity. The traffic load is assumed to fluctuate slowly on the time scale of many tens or hundreds of timeslots, a good approximation to the traffic statistics in a large telecommunications network.

Ideally, a network design should take into consideration the overall performance and interactions between all the components and subsystems. Packet deflections cause an increase in the link loadings, increased end-to-end delay and a reduction in the overall network throughput [5]. These parameters are difficult to evaluate for real and evolving networks—particularly irregular networks incorporating splitters and combiners. This is a topic for further research which is outside the scope of this paper. Nevertheless, the results here for the packet loss of single switching fabrics provide a useful indication and comparison of performance.

Fig. 4 depicts deflection probability W against load p for various sizes of fabric. As expected, the deflection probability becomes lower as the switch fabric is made larger; for a large $B(128)$ fabric, the loss is as low as 3.9×10^{-13} for a high channel load of 0.9. Even for $B(16)$, the deflection probability is 1.2×10^{-8} for a load of 0.5.

C. Loss and Crosstalk Performance

The attenuation of a $B(n)$ fabric is $A = L(\log_2 n+1)$ where L is the insertion loss of a single switch, including coupling loss. In a lithium niobate implementation, $L = 5$ dB is typical, implying that $B(16)$ has 25 dB attenuation; clearly optical amplification would be required to overcome the loss of each module, regardless of the link loss between nodes. Lithium niobate switches also have the disadvantage of being sensitive to polarization state; polarization insensitive devices are available but they require a higher drive voltage, implying lower speed operation. InP semiconductor laser amplifier switches, on the other hand, exhibit a much higher isolation,[1] and due to their implicit amplification, may exhibit gain [11]. Here, the key performance issue is buildup of amplified spontaneous emission (ASE). This is outside the scope of this paper and the reader is referred to [17] for an analytical treatment of similar systems. It is clear that crosstalk performance depends on the switch technology, and that crosstalk issues will be much less important if InP semiconductor laser amplifier devices are used.

To estimate the effect of crosstalk on the performance of the buffering module, the power penalties for various sizes of fabric and values of switch device isolation were calculated using a customized software package [18],[2] assuming the use of lithium niobate technology. Just as for the deflection probabilities, ideally the network designer should take into consideration the power penalty for an entire network, rather than a single switching fabric as is done here. Such a calculation would necessarily be specific to a given network layout of splitters and combiners, and so is outside the scope of this paper. Nevertheless, the present simulations give a useful comparison of node performance and an approximate indication of overall network performance.

Pipelined architectures of this sort suffer from interferometric noise (homodyne crosstalk) effects [19]–[21], due to

[1] Here "isolation" is defined as the level of the crosstalk signal relative to the correctly routed signal. For example, if a 2×2 has a 0 dBm signal on one input, with a signal at the correct output of -3.5 dBm and a crosstalk signal at the wrong output of -35 dBm, then the isolation is $3.5 - 35 = -31.5$ dB.

[2] xHatch is a powerful interactive computer modeling tool for determining the effect of noise and crosstalk on the performance of optical systems. It was originally developed at the University of Colorado at Boulder for studying timing in optical systems before being developed into its present form at the University of Strathclyde.

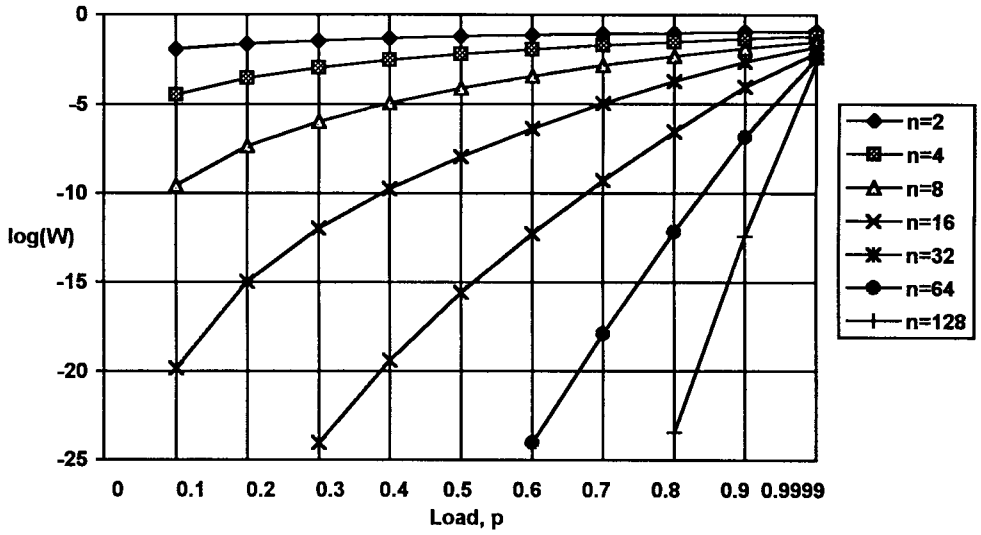

Fig. 4. Log of cell deflection probability $\log W$ against load p for various sizes of buffering modules.

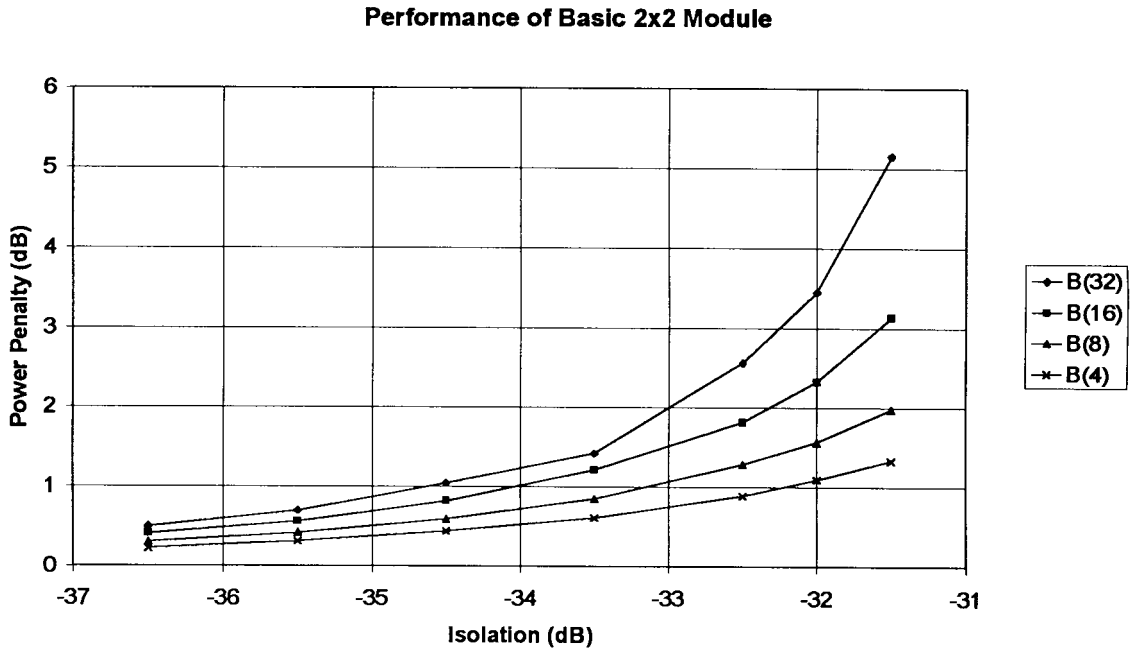

Fig. 5. Power penalty performance of the basic 2×2 buffered switch module for BER $= 10^{-9}$ and full loading plotted versus isolation performance of the switch devices.

crosstalk optical field addition at each switch device, which ultimately limits the size of the buffer that can be implemented. Fig. 5 shows the optical power penalty for different sizes of fabric, as a function of switch device isolations from −36.5 dB to −31.5 dB. Full loading (i.e., a traffic level of $p = 1$) is assumed; while this would be an unrealistic assumption for traffic modeling, it provides a worst-case assumption for modeling the optical performance. From the graph, $B(32)$ becomes impractical for −31.5 dB isolation; indeed for −31 dB it was found that the error floor was above 10^{-9}. It is also clear, that even for modest fabrics, switch devices of extremely high performance would be required; one way of solving this problem is to use dilation [13], [22].

Before discussing the issue of dilation, consider Fig. 6. This shows the performance of the basic switching module for traffic levels from 0.0 to 1.0, assuming the use of switch devices with a −31.5 dB isolation. As expected, the power penalty increases with load; this is because as the load increases, the probability of any given cell meeting another cell in a switch also increases, increasing the probability of crosstalk from adjacent cells. It also illustrates that a load of $p = 1.0$ represents the worst case.

The results of Figs. 5 and 6 may also be overpessimistic because they assume that all the cells are derived from the same optical source and are therefore at the same wavelength. Since in practice, cells come from different laser sources and tend to

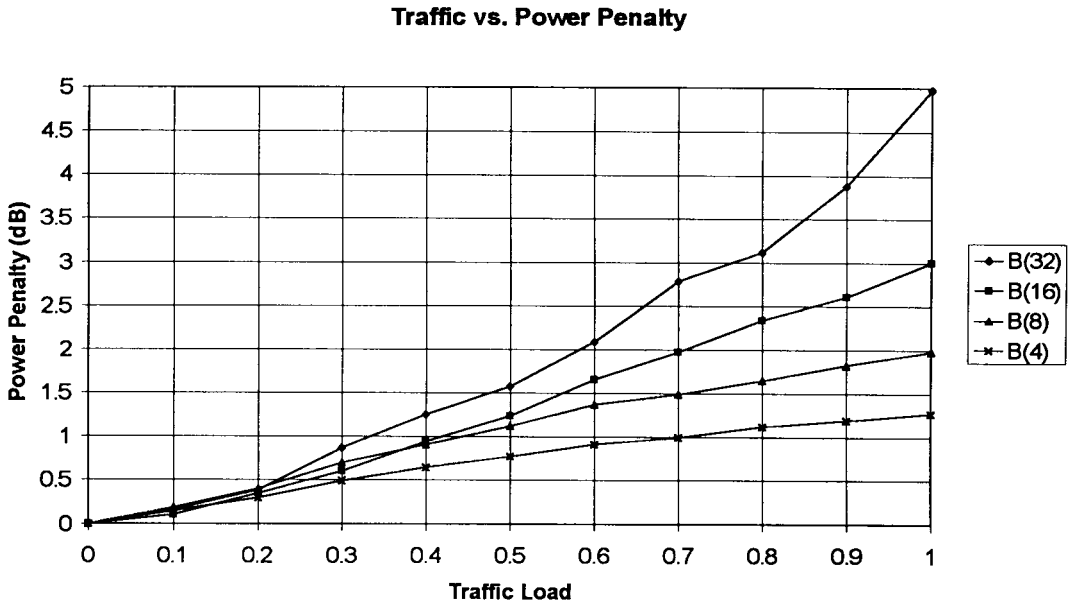

Fig. 6. Power penalties for the basic switching module for BER $= 10^{-9}$ versus traffic load, assuming switch components with -31.5 dB isolation.

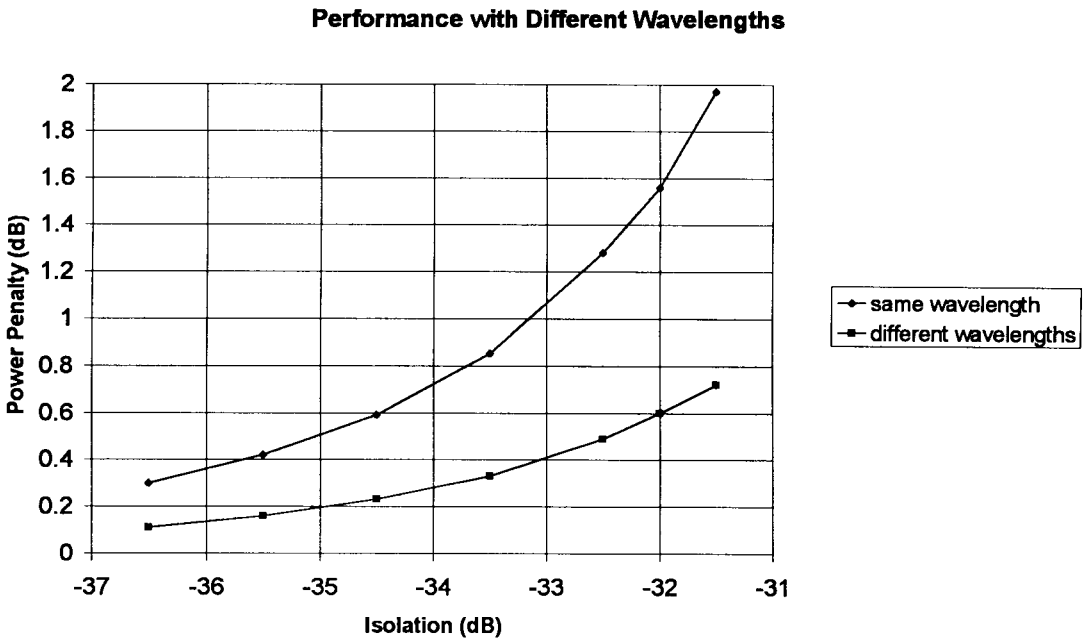

Fig. 7. Power penalties for a fully loaded $B(8)$ plotted versus device isolation assuming BER $= 10^{-9}$. One curve assumes the two inputs are at different wavelengths (1550 and 1551 nm) and the other curve assumes the input wavelengths are the same.

be on slightly different wavelengths, a simulation was carried out where all cells on the upper input were at 1550 nm and all those on the lower input were at 1551 nm. While this scenario is not exactly that which would be encountered in practice, it allows an insight into the effect of different wavelengths. For a $B(8)$ fabric (Fig. 7), the power penalty for the two wavelength case is less than half (in dB) than for the same wavelength case. This is to be expected, since when two signals of different wavelengths beat together at the detector, the beat noise terms are outside the bandwidth of the detector. Again, this confirms that the results in Fig. 5 are a worst case calculation.

To relax the demands on device crosstalk performance, it is possible to dilate the switch fabric. This involves increasing the number of devices employed to ensure that at most one cell goes through a switch device at once, thus reducing crosstalk [13], [22]. Fig. 8 compares an undilated fabric and the equivalent fully dilated fabric, showing how the dilated fabric could be controlled using the same control signals as the original fabric but with a small amount of additional electronic logic.

The crosstalk performance for dilated switch fabrics is shown in Fig. 9. Again, the fabric is fully loaded and compar-

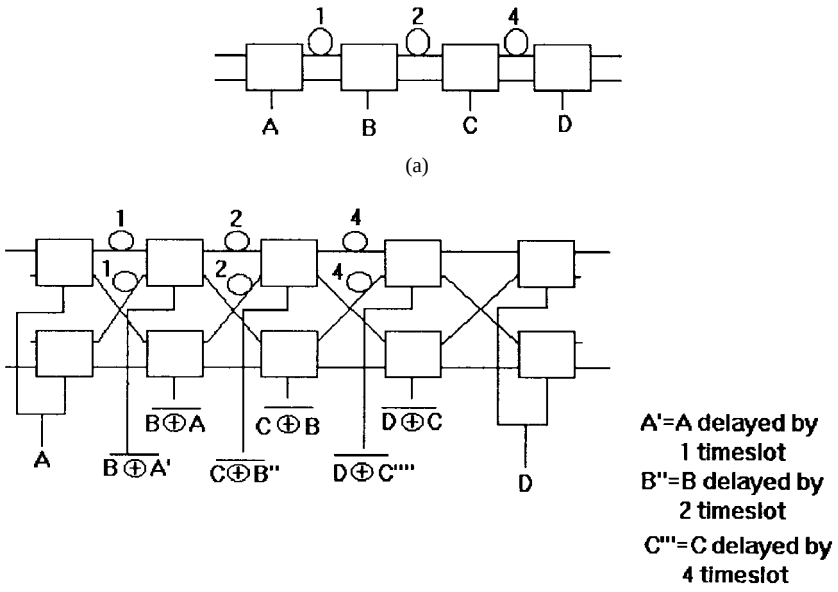

Fig. 8. The $B(8)$ fabric with control scheme (a) undilated (b) fully dilated. The symbol "$\oplus$" represents the exclusive-OR logic function.

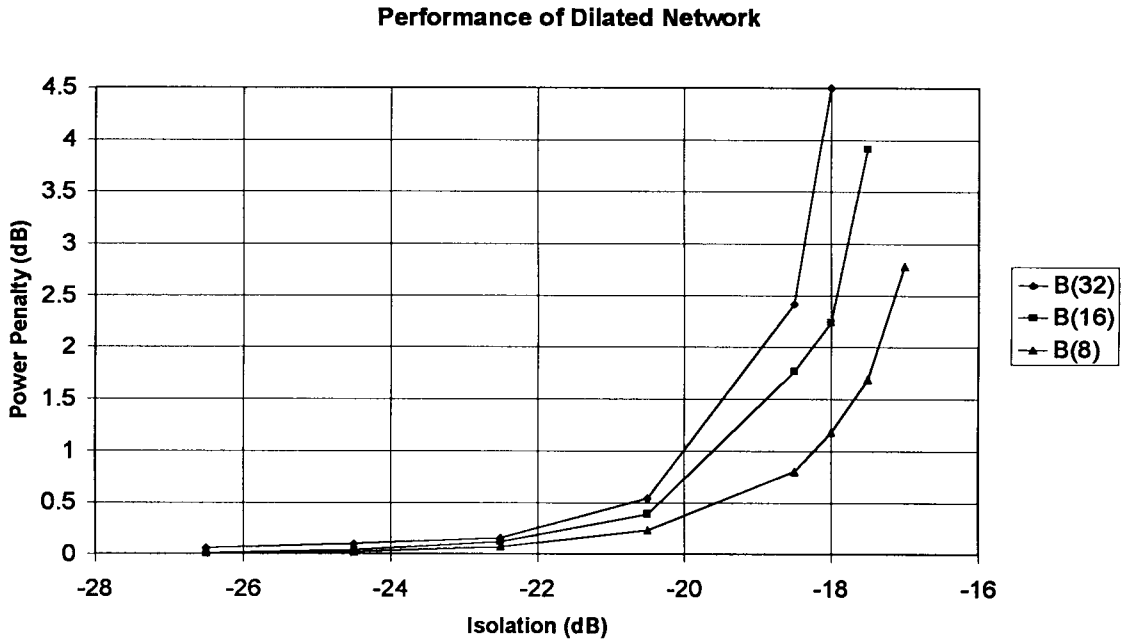

Fig. 9. Power penalties for dilated fabrics under full load assuming BER $= 10^{-9}$ versus isolation of the switch devices.

ing with the original geometry (Fig. 5), it is clear that dilation offers a dramatic improvement in performance. Indeed, for $B(8)$ and $B(16)$, the power penalty is negligible for device isolations in the range -36.5 to -26.5 dB.

III. Traffic Merging Junctions

A. Architecture

Binary routing switches of the type described above can be interconnected by joining each input of every switch fabric to the output of another switch fabric, so as to form a mesh network with degree of connectivity equal to two (sometimes referred to as a "two-connected network"). Well-known examples of regular, two-connected mesh networks are the Manhattan Street Network [23] and the Highway-Transfer Network [24], which both have a toroidal topology. However, these highly regular networks may be realistic only for specialized local-area applications, such as interconnection backplanes in distributed computing environments. More typically, networks evolve and grow in an irregular fashion, and this is especially true as the geographical size increases. Therefore there is a need for switch fabrics that perform the functions of splitting and merging traffic at irregular nodes (ones with an unequal number of inputs and outputs).

Traffic splitting can be performed very simply using a binary routing switch device in which there is one input and two outputs. Buffering and contention resolution are not required in such a switch, and no cell loss is incurred. On the other hand, a "merging junction" is a switch that combines the traffic from two inputs (a "main" input and a "tributary" input) onto a single main output. If the traffic capacity of the output link is less than the sum of the capacities of the two input links then buffering is required, and to avoid cell loss due to buffer overflow the input traffic must be held within limits. In this section we consider how the 2×2 buffered switch module can be used as a merging junction, and in Section IV this is extended to perform traffic shaping.

The main difference in operation between the merging junction and the 2×2 buffered switch (Fig. 3) is that in the junction, *all* the incoming traffic is designated to exit from one port (the main output). A merging junction is thus simply a 2×2 buffered switch fabric with one output unused, or rather, used as an overflow; the other output is the main output. Header detection and recognition are unnecessary since the intended exit route of every cell is known. The junction acts as a single output buffer of depth $n-1$ containing a single queue (c.f. a single output buffer of depth $n-1$ shared by two output queues in the 2×2 switch module of Fig. 1), and if the buffer overflows, surplus cells are deflected to the overflow output. As well as merging traffic at irregular network nodes, the merging junction can form part of a larger 2×2 switch fabric (as we shall see shortly).

The order in which the two inputs to the junction are designated as being the "main input" and "tributary" is arbitrary. Neither input need have higher priority than the other, and all cells may be processed equally and indistinguishably within the junction, regardless of origin, in the manner described. However, if required, the control logic may be wired or programmed in a simple way so as to give higher priority to the designated "main" traffic; thus, should the switch buffer overflow (by the arrival of two cells simultaneously whilst the fabric is in state $n-1$—see Appendix C), the cell incident at the "main" input would be selected preferentially to exit by the main output port, whilst the cell incident simultaneously at the "tributary" input would be directed to the overflow exit.

B. Deflection Performance

Again, random uniform traffic is assumed; with this assumption, Markov chain analysis can be used to evaluate the probability W of a given cell being directed to the overflow exit (Appendix C)

$$W = \frac{p_T p_M}{p_T + p_M}\left(\frac{1-\beta}{1-\beta^n}\right)\beta^{n-1}$$

where

$$\beta = \frac{p_T p_M}{(1-p_T)(1-p_M)}$$

and p_M and p_T are, respectively, the traffic load (i.e., the probability that a timeslot is full) for the main and tributary input. Note that these expressions are symmetrical in p_M and p_T. If one input is turned off (p_M or $p_T = 0$), the cell-overflow probability is zero, thus maintaining the transparency of the switch fabric for the other input.

Fig. 10 shows the calculated cell-overflow probability W against the tributary traffic load p_T for various sizes of switch fabric. In the example it is assumed that the main input traffic load p_M is 0.4. As expected, the cell-overflow probability becomes smaller as the size of the switch fabric is increased. However, when the total input traffic load $p_M + p_T$ exceeds 1, the cell-overflow probability increases to a large value, regardless of how deep the buffer capacity may be. As either p_M or p_T (or both) approaches 1, $W \rightarrow p_M p_T/(p_M + p_T)$. In the region below over-capacity ($p_M + p_T < 1$), useful low cell-overflow probabilities (e.g., $W < 10^{-12}$) can be obtained.

Fig. 11 shows the calculated maximum tributary traffic load $(p_T)_{\max}$ that can be merged with a main traffic stream of input load p_M, resulting in a cell-overflow probability $W \leq 10^{-12}$, for various sizes of the switch fabric. These results show that substantial tributary traffic loads can be accommodated using a switch fabric of modest size. For example, with a main input traffic load $p_M = 0.5$, a tributary loading as high as $(p_T)_{\max} = 0.16$ can be accommodated with low cell overflow probability ($W \leq 10^{-12}$), using a switch fabric of size $B(16)$ which comprises four delay lines and five switches. With the same main input load $p_M = 0.5$, a switch fabric having just one more switch device and delay line, $B(32)$, can accommodate a tributary load $(p_T)_{\max}$, resulting in an output-load capacity $((p_T)_{\max} + p_M)$ greater than 80%. As the switch fabric is increased further in size, the output-load capacity approaches 100%, regardless of the main input load p_M (Fig. 11).

C. Zero-Deflection Switch

The basic 2×2 buffered routing switch described in Section II (and shown in Fig. 2) has the property that it *guarantees* no loss of cells, although this guarantee is obtained at the expense of a finite probability of cell deflection. Fig. 12 shows another design of 2×2 routing switch which differs in that the output buffers are entirely separated. This architecture consists of a fabric of two traffic splitters and two merging junctions. A traffic splitter is a 1×2 switch device which directs each incoming cell to the output dictated by the controller; these are the two leftmost switch devices in Fig. 12. The merging junctions (to the right of the traffic splitters in Fig. 12) act as independent output buffers. This design of 2×2 buffered routing switch fabric contains more optical components than the basic module described in Section II. Nevertheless it has an interesting and useful complementary property; the switch architecture *guarantees* there will be no cell deflections. This zero-deflection guarantee is obtained at the expense of a finite probability of cell loss; cell loss can occur if either of the output buffers overflow. As discussed in the Introduction, this type of buffered switch with zero deflections could be valuable for networks carrying traffic that would be disrupted by delays cause by deflection routing, if that traffic were already tolerant to cell loss. The probability of cell loss is easily derived from

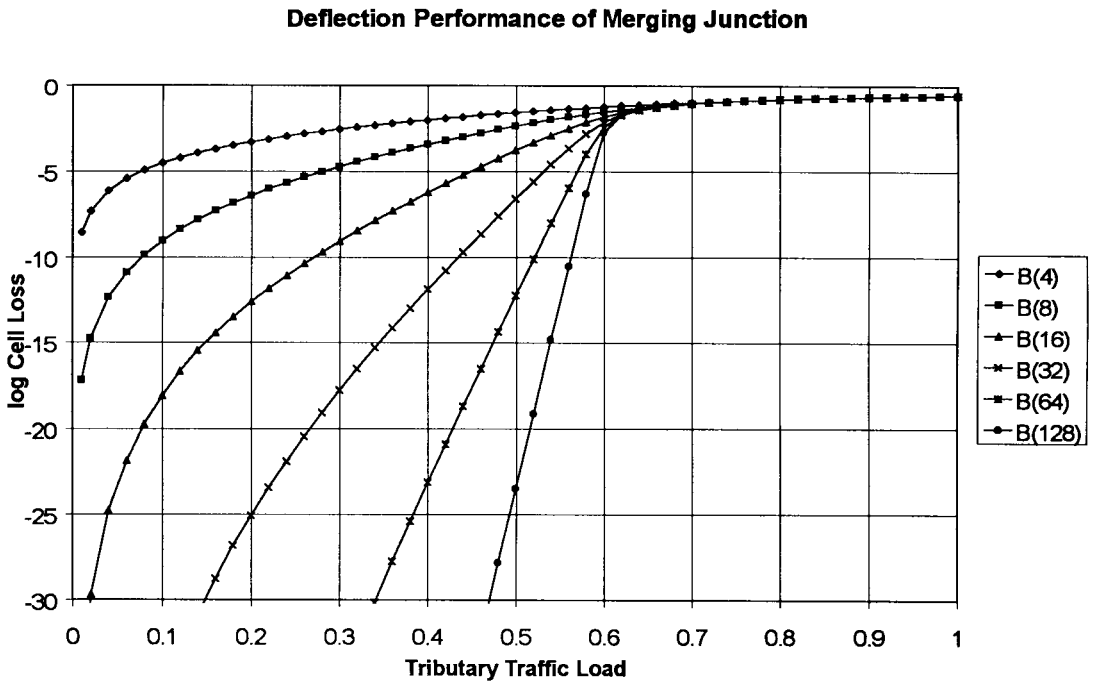

Fig. 10. Calculated deflection probability against tributary traffic load for the merging junction.

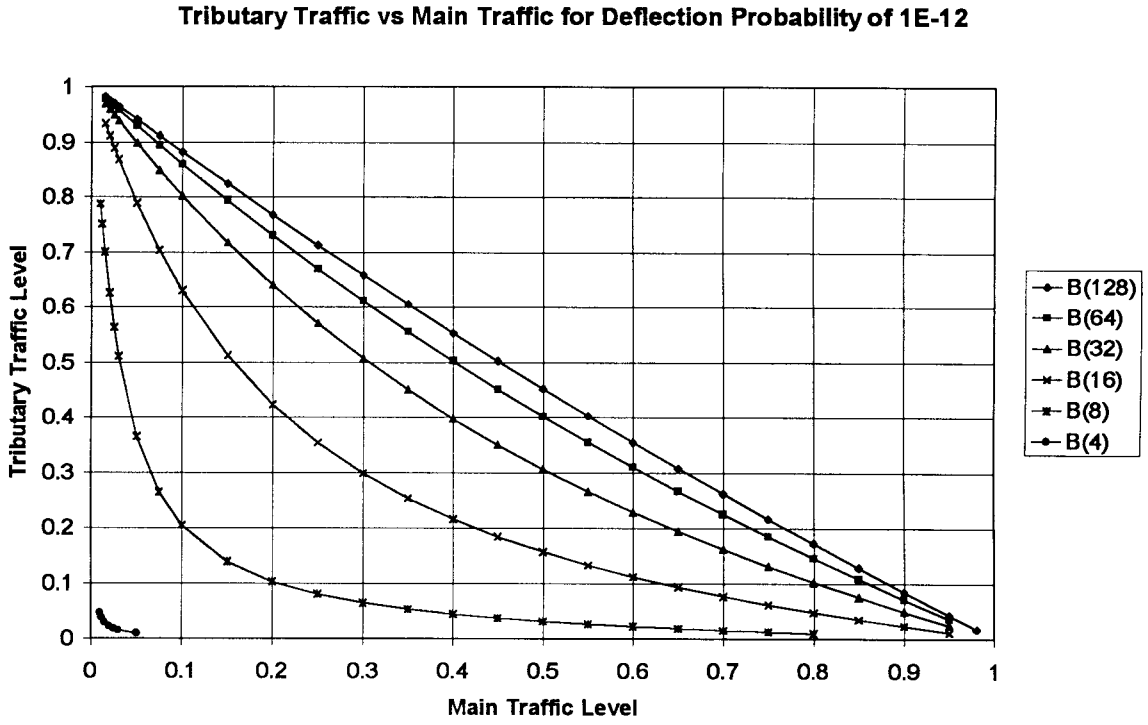

Fig. 11. Calculated maximum tributary traffic load against main traffic load for deflection probability $<10^{-12}$.

the deflection probability shown in Fig. 10 and derived in Appendix C.

D. Crosstalk Performance

The crosstalk performance of the traffic-merging module was simulated. Fig. 13 shows the power penalty for the 2×2 buffered routing switch architecture of Fig. 12. Again, lithium niobate technology has been assumed, with the parameters given in Section II-C. This simulation provides a measure of performance for the merging junctions. Since traffic of $p = 1.0$ is entering each input to the 2×2 routing switch, and the traffic is uniform, the traffic entering each input to the merging junctions will be $p = 0.5$. Note also that the traffic splitters (i.e., the 1×2 switches at the left of Fig. 12) do not contribute to the optical power penalty since

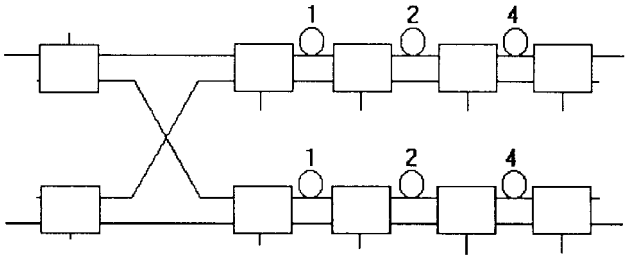

Fig. 12. A buffered switch fabric of depth eight, made up of two splitters and two combiner modules. This architecture guarantees that there are no cell deflections, but incurs a finite cell loss probability.

they only cause loss and no crosstalk. Hence these results may be viewed as the performance of an individual merging junction with $p_T = p_M = 0.5$. Again, since $p_T + p_M = 1$, this is not a realistic scenario for traffic modeling, but nonetheless provides us with a useful worst case for power penalty comparisons. In comparison with Fig. 9, which uses a similar number of devices, the crosstalk performance for switch architecture with separate output buffers (Fig. 12) is much poorer, indicating that lithium niobate technology (with its relatively poor crosstalk performance) would not be best suited to this switch implementation.

IV. Traffic Shaping

Traffic shaping is an important part of the overall scheme to prevent network congestion [25]. In high-speed photonic cell networks it would be attractive to implement traffic shaping in hardware, using modest amounts of buffering in the optical domain. In this section we consider how the traffic-merging junction described in Section III can be adapted in a simple way to perform a type of traffic shaping.

In the normal operation of the merging junction, tributary traffic is accepted and merged with a main traffic flow, although the overflow will rise to a large value if the sum of the input main and tributary traffic flows exceeds the capacity of the main traffic output. As already described in Section III, the switch control logic can be arranged so that the overflow traffic is taken only from the tributary input, and safe passage of the main input traffic is guaranteed. We now describe how the traffic-merging junction can be configured in such a way that if the tributary traffic exceeds a certain critical level, the excess tributary traffic will be forced to "overflow" (in effect, the switch module forces the excess traffic to be deflected). This could be used in an elementary "policing" operation in high-speed photonic cell networks.

The special feature of the traffic-merging junction that will be exploited here is the phase transition that occurs when the sum of the input traffic flows equals the capacity of the main output. This phase transition can be seen in Fig. 10 when $p_M + p_T = 1$, and it is clear that the transition becomes increasingly abrupt as the switch size is increased. To adapt the merging junction to perform traffic shaping, this phase transition must be forced to a different position on the input-output curve. A way of doing this is to trick the switch control logic into believing that the main input traffic level is greater than it actually is. Suppose we wish to ensure that the contribution to the main output traffic from the tributary will not exceed a certain level $p_{T\max}$ (and if the input tributary traffic load p_T should exceed $p_{T\max}$ then we shall impose the penalty that the excess $p_T - p_{T\max}$ will be deflected). For normal operation of the buffered switch module, the inputs to the switch control logic are two continuous binary strings (one representing the main input traffic and the other representing the tributary) in which a value 1 denotes a time slot is filled and 0 denotes the time slot is empty. To perform traffic shaping, the binary string representing the main input traffic is replaced by a "synthesized" binary string with mean value $p_S = 1 - p_{T\max}$. This synthesized string is created by taking the binary string representing the main input traffic flow and converting certain of the 0 values to 1. Fig. 14 shows a simple algorithm for generating this synthesized string. The binary random number generator produces one number per time slot, with mean value p_S. We have verified that the values in the synthesized string generated using this algorithm have the correct Bernoulli statistics, under the assumption that the input traffic is uniform. In this way some of the empty time slots in the main input traffic are "reserved", so that they cannot be filled by tributary cells. This has the effect of limiting the number of tributary cells that can be merged with the main traffic. In concept, this has some similarity with the well-known "leaky bucket" [25].

Adapting the final result of Appendix C, the probability that a time slot in the shaped output traffic is filled is given by

$$p_0 = p_M + p_T\left(1 - (1 - p_{T\max})\frac{1-\gamma}{1-\gamma^n}\gamma^{n-1}\right)$$

where

$$\gamma = \frac{p_T(1 - p_{T\max})}{p_{T\max}(1 - p_T)}.$$

Using these expressions, Fig. 15 shows the flow-controlled output and overflow traffic against tributary load p_T assuming a given main input load p_M. It can be seen that with a switch fabric as small as $B(8)$ (consisting of just three delay lines and four switches), there is well-defined limiting action when the input tributary load p_T exceeds the set maximum value $p_{T\max}$. Fig. 16 depicts the shaped output traffic against tributary traffic load for various levels of main input traffic load p_M. These results show that the operation of the device is independent of the main input load p_M, provided $p_M + p_{T\max} < 1$. The traffic-shaping device can therefore provide a well-defined limiting action for the tributary traffic whilst allowing the main input traffic load to fluctuate freely. Because the control mechanism is localized, it is inherently fast-acting and responsive. The maximum permitted tributary load $p_{T\max}$ is defined by setting the value of $p_S = 1 - p_{T\max}$ in the algorithm used to generate the synthesized number string, and if required this could be set and controlled by the overall network management system.

The analysis above has assumed that the traffic is uniform, at least on the time scale of the buffer depth. It is important also to consider the performance with bursty and nonuniform traffic, which will have the effect of increasing the overflow probability. One difficulty in analyzing nonuniform traffic is

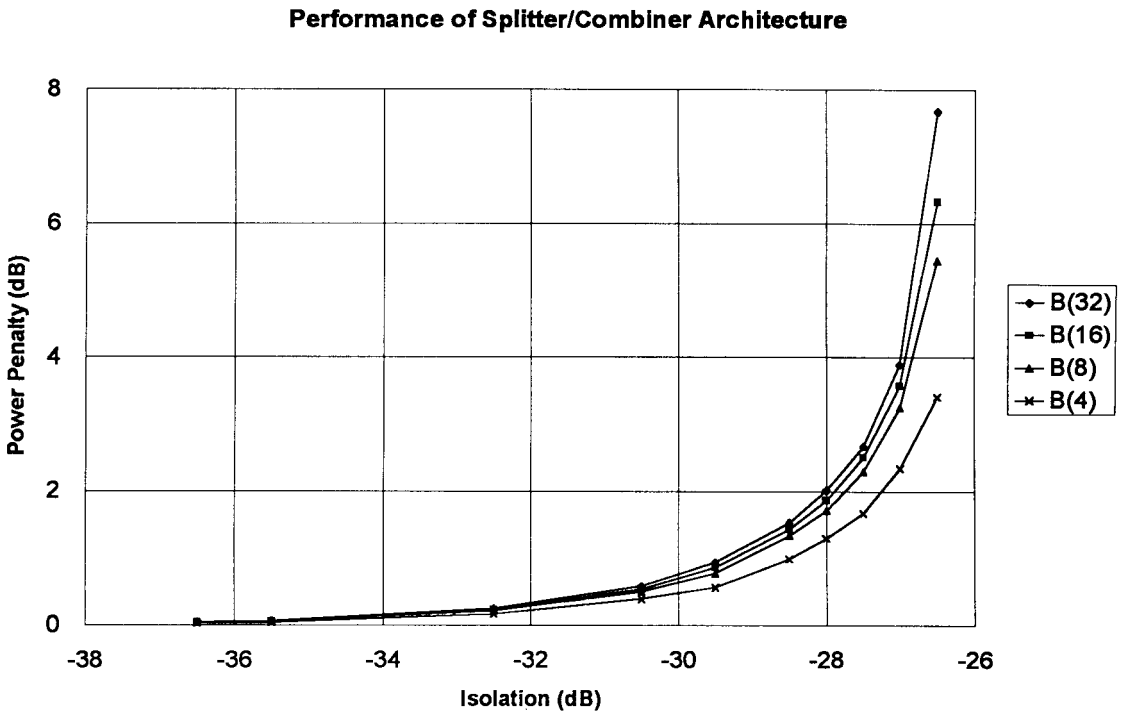

Fig. 13. Power penalties for the 2×2 switch architecture of Fig. 12 versus isolation of the switch devices, assuming input traffic loads $p_M = p_T = 0.5$ and BER $= 10^{-9}$.

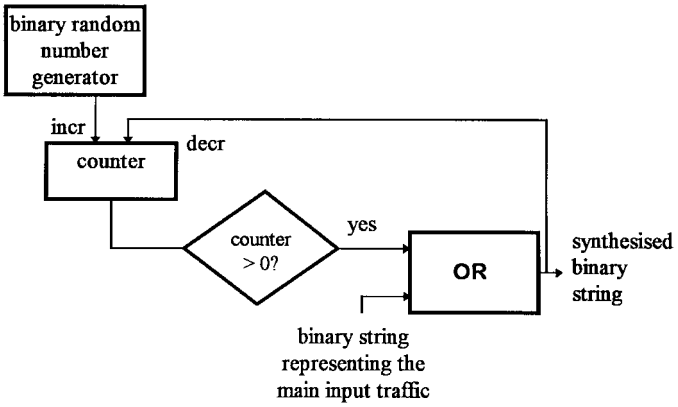

Fig. 14. Logic diagram of the algorithm to create the synthesised binary string used in the operation of the traffic shaping device. In the input and output strings, a value 1 denotes that a timeslot is filled by a cell and 0 denotes that the timeslot is empty. The binary random number generator produces one number per time slot with mean value $p_S = 1 - p_{T\max}$. The integer counter can cover over positive and negative values.

that it is not easy to provide a useful specification of the traffic using a compact notation. However useful insight can be gained by considering the following simple example.

Suppose that the tributary traffic is composed of bursts of consecutive cells, interspersed by empty gaps. During such a gap, the state diagram for the traffic shaper is as shown in Fig. 19, except that the transitions labeled "2L" are absent (because no cells are arriving on the tributary input), and the transitions "N" and "L" represent the absence or presence of cells in the synthesized stream with mean value p_S. It can be seen, therefore, that provided the gaps in the tributary traffic are much longer than the depth of buffer in the traffic-shaping module and provided also p_S is not close to 1, the system can reach stable equilibrium in state 0 before the arrival of the next burst. Then, during a burst, the state diagram for the traffic shaper is again as shown in Fig. 19, except that the transitions labeled "N" are absent (because cells are continually arriving on the tributary input). Starting in state 0, the system will tend to move toward higher states; if the burst is long enough, the system will eventually reach state $n-1$, and overflow occurs. The probability that a burst of b cells on the tributary will be accepted successfully by the traffic shaper without overflow is given by

$$P(b) = 1 - \sum_{i=n}^{b} \binom{b}{i} p_S^i (1 - p_S)^{b-i}.$$

The summation term is the probability of overflow—the probability that in b timeslots there are n or more (i.e. between n and b) cells in the synthesized string on the main input to the shaper, given by a sum of Binomial distributions. It is assumed that during these timeslots there is always a cell on the tributary input. This expression shows that bursty traffic on the tributary, even if well dispersed, can be guaranteed acceptance only if the burst length b does not exceed the buffer depth $n-1$.

This simple example highlights an important aspect of photonic cell switches and networks. The limitations of current technology are such that photonic cell buffers are much smaller than their electronic counterparts. This hampers the ability of photonic networks to handle bursty traffic without cell loss or deflection when the peak burst rate is comparable to the bandwidth of the network links. However, with current technology, high-capacity photonic cell networks should be

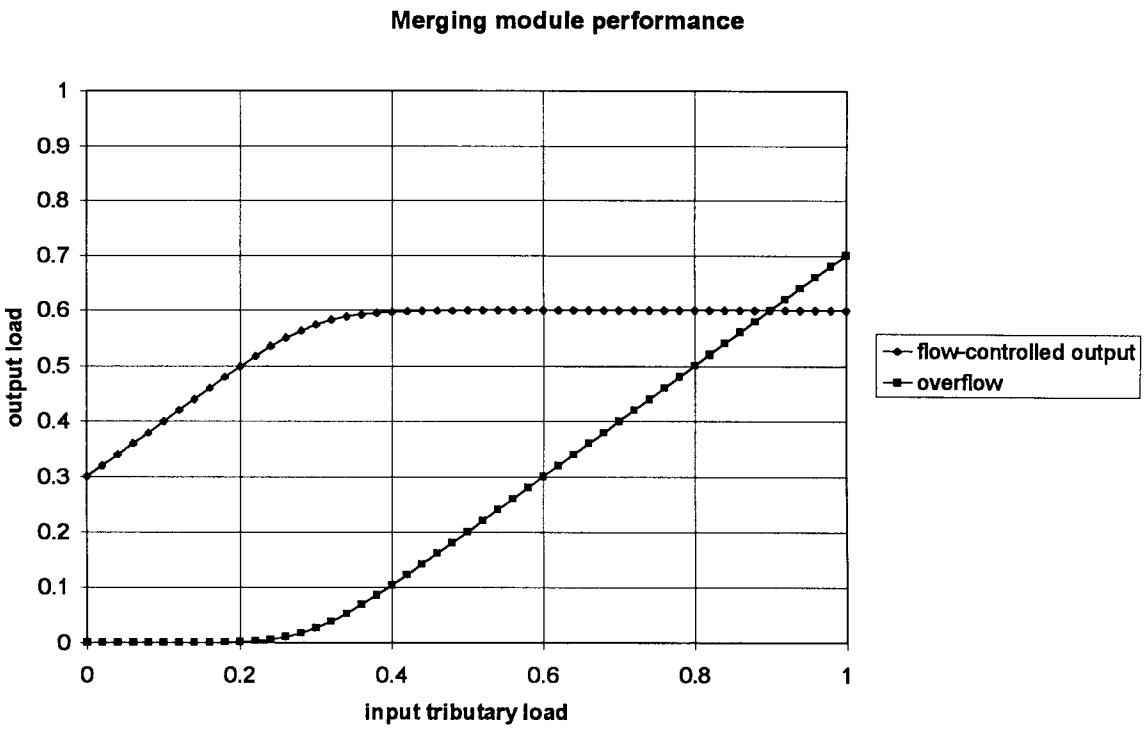

Fig. 15. Performance of a traffic-merging module ($n = 8$) configured as a traffic shaping device, showing the shaped output and overflow versus input tributary traffic load. Excess tributary traffic (exceeding a preset maximum $p_T = 0.3$) is diverted to the output overflow port. An arbitrary value of 0.3 is assumed for the main input traffic p_M.

well able to transport composite traffic consisting of a large number of well-distributed, concurrent flows (a "flow" being the end-to-end traffic associated with a particular process or application), when the peak rate in any flow is a small fraction of the link bandwidth. As mentioned earlier, the uniform traffic model is acceptable in this case.

V. Conclusions

In this paper a new type of optical 2×2 buffered switch module for photonic cell networks has been described and analyzed. It uses a feed-forward architecture and consists of a linear chain of 2×2 optical switches and delay lines. A single module can be used as an output-buffered routing switch, in which contention due to full buffers is resolved by deflecting cells to the wrong output. Even for relatively small fabrics, low deflection probabilities can be achieved; a $B(32)$ module (consisting of six switch devices and five delay lines) with a loading of $p = 0.8$ achieves a deflection probability of 2.8×10^{-7}. In practice, higher deflection probabilities (and hence smaller switch modules) could be used in conjunction with deflection routing. We have described how the buffered switch module can be modified to allow two traffic streams to merge where a deflection probability of 10^{-12} is possible with $B(32)$ and a total load of 0.8. It was shown that two such traffic-merging devices can be combined to form a composite 2×2 buffered switch architecture. This design guarantees there will be no cell deflections, but at the expense of finite cell loss. The optimum choice of switch architecture depends on factors such as the geographical size of the network, the bandwidth and dominant traffic type. Future work will concentrate on analyzing the effects of bursty and nonuniform traffic, which will increase the deflection probability for any given size of module.

The performance of these modules was simulated with respect to crosstalk, and it was concluded that if lithium niobate switches are used, the modules must be dilated to yield an acceptable power penalty. For example, in a dilated module, switches with an isolation of -18.5 dB yield a power penalty of less than 1 dB. Future work involves looking at networks of these nodes and determining their performance.

A networking concept radically different from that presently in use was discussed for use with these switching modules, driven by the need to reduce hardware, software and operating costs. Apart from using these modules in their obvious application (i.e., routing), their use for traffic "policing" in photonic cell networks was proposed, which is necessary to avoid links becoming overloaded, preventing congestion. Although the concepts introduced here are obviously quite different from today's approaches, it is clear that they will be possible in the early part of next century.

Appendix A
Switch Module Control

Each switch module has two inputs and two outputs; each cell entering the fabric is designated "U" or "L" depending on whether it is destined for the upper or lower output, respectively. It will be assumed that on every timeslot, two cells enter the fabric. In practice, there may be timeslots on which one or both of the inputs do not carry a cell; this possibility will be dealt with later.

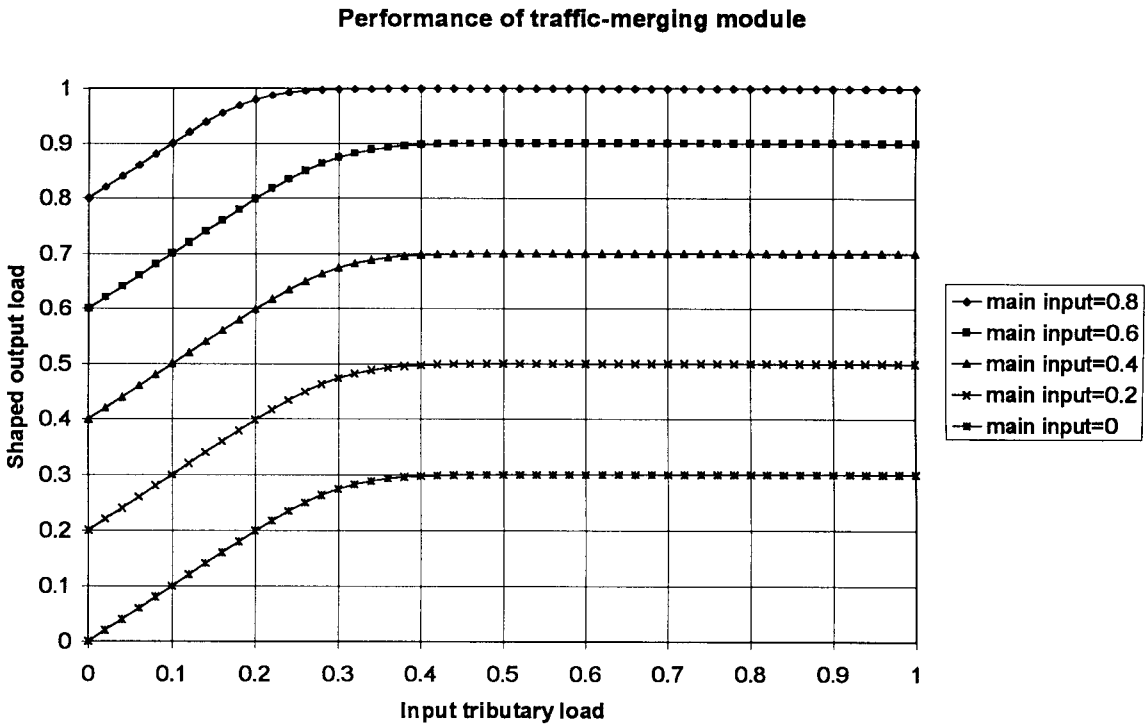

Fig. 16. Performance of a traffic-merging module ($n = 8$) configured as a traffic shaping device with $p_T = 0.3$ showing the shaped output versus input tributary traffic load, for various values of main input traffic p_M.

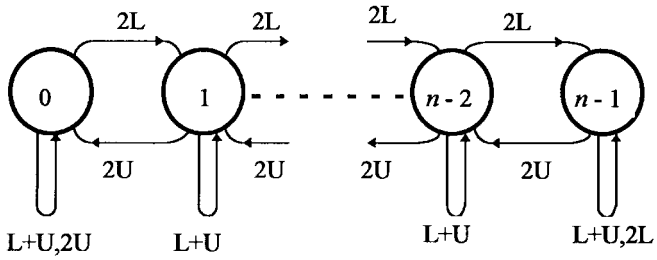

Fig. 17. State diagram of a buffered 2×2 switch module assuming the arrival of two cells per timeslot.

The objective is to ensure that on each timeslot, it is impossible for two "U" cells or two "L" cells to reach the outputs simultaneously. If this objective is met, the module is effectively buffering the cells, since each cell is stored until a free output is available to transmit it. There are three possibilities for the cells entering the input:

1) two "L" cells—this will be called "2L;"
2) two "U" cells—this will be called "2U;"
3) one "L" and one "U" cell—this will be called "L+U."

As discussed in Section II-A, the modules are referred to as $B(n)$ where n is an integral power of two and refers to the number of states in the Markov chain described below.

In addition to the switches and delay lines, two header detectors are required on the inputs to each module [16]. These extract the routing information from the header, and pass their values on to the control logic. These values determine whether each cell is a "U" or a "L." The control logic consists of a $\log_2 n$-bit up/down counter to keep track of the switch fabric state plus associated random logic.

The state of the fabric is an integer from 0 to $n-1$ which represents how many cells are in the buffer feeding to the lower output. Its value (which is held in the up/down counter) varies as cells enter and leave the module. Suppose that the fabric state is represented by an integer s (taking on values from 0 to $n-1$). If the fabric is in state s and "L+U" cells arrive then the "L" cell will be delayed by s timeslots before emerging on the lower output, and the "U" cell will be delayed by $n-1-s$ timeslots. The fabric remains in state s.

Again, suppose that the fabric is in state s, but two "L" cells arrive at the input. One will be delayed by s timeslots and the other will be delayed by $s+1$ timeslots. The fabric goes into state $s+1$. Conversely, if the fabric is in state s and two "U" cells arrive, one cell will be delayed by $n-1-s$ timeslots and the other will be delayed by $n-s$. The fabric goes into state $s-1$.

It can be seen that this ensures that one "U" cell and one "L" cell arrive at the output on every timeslot, preserving the ordering of the cells. This is equivalent to performing buffering of the cells. If two "L" cells arrive in state $n-1$, or two "U" cells arrive in state 0, one cell will be deflected to the wrong output since the buffer has overflowed. In the case of two "L" cells, one will be delayed by $n-1$ timeslots and sent to the lower output, while the other will be delayed by 0 timeslots and sent to the upper output. Thus in state $n-1$, "2L" is treated as "L+U," i.e., one of the "L" cells is treated as a "U" cell. Similarly, in state 0, "2U" is treated as "L+U" also. The cell which is treated as "L" or "U," respectively, must be labeled as such by the controller. It is helpful to represent this by a state diagram (Fig. 17), where deflections are shown. Since it has been possible to reduce the description to a single

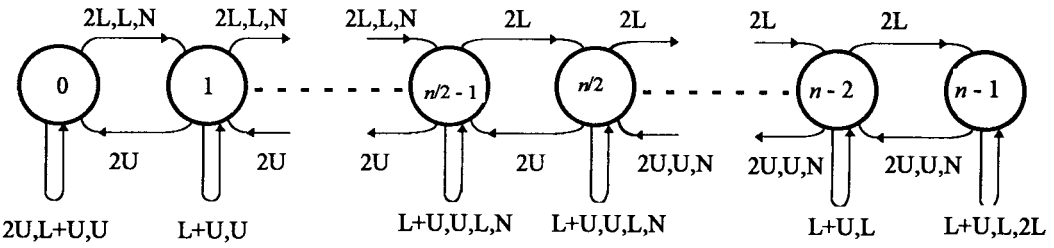

Fig. 18. State diagram for the switching module amended to allow for the possibility of fewer than two cell arrivals in a timeslot.

state variable, it is not necessary to have a two-dimensional state diagram with separate dimensions for both "L" and "U" packets.

It has been assumed until now that on every timeslot, two cells enter the module. If we allow one or more inputs to be free on some timeslots, this introduces three new possibilities which are listed as follows:

1) "L"—cell for lower output only,
2) "U"—cell for upper output only, and
3) "N"—no cells.

Clearly, "L" can be substituted for "L+U" or "2L" in the state diagram, and "U" can be substituted for "L+U" or "2U." "N" can be substituted for anything. The state diagram must be manipulated so that the module goes toward one of the central two states whenever possible. This minimizes the probability of state 0 or $n-1$ being reached, and possible deflection occurring. The amended state diagram is shown in Fig. 18.

There is a minor optimization[3] which may be carried out for small fabrics where the cell ordering is not important. Although there are two cases where deflection may take place (state 0 and state $n-1$), state 0 will be considered here, the extension to the other state being obvious. Suppose two "U" cells arrive in state 0. Then one of them must exit immediately and the other must withstand a delay of in $n-1$ order to stay in state 0. If the cell currently leaving from the "U" queue was originally a "U" cell, then everything proceeds as before. However, if it was originally a dummy cell, the new cell which has a delay of zero can exit on the "U" output (overtaking other "U" cells in the buffer and upsetting packet ordering) with the dummy cell being sent to the "L" output instead. (Without the optimization the dummy cell would go to the "U" output and the zero-delay "U" cell would go to the "L" output.) Thus by using this technique, the deflection probability is reduced, at the expense of upsetting packet ordering. The only case where packet ordering is not changed occurs in the trivial case of $n=2$. Although this effect is significant for $B(2)$, it is negligible for larger fabrics, particularly when $p>0.4$, as in practice; hence it will not be considered further here.

Appendix B
Defleciton Performance of the Switching Module

As mentioned in the text, it is assumed that the traffic is uniform. To evaluate the cell loss and mean delay of the fabric, the labels in the state diagram of Fig. 18 may be replaced by their probabilities

$$\begin{aligned} P(N) &= (1-p)^2 \\ P(L) &= p(1-p) \\ P(U) &= p(1-p) \\ P(U+L) &= \frac{p^2}{2} \\ P(2L) &= \frac{p^2}{4} \\ P(2U) &= \frac{p^2}{4} \end{aligned}$$

where p is the probability of a cell occupying any given timeslot. The resulting Markov chain is symmetrical and may be replaced by a new chain half the size where each state $i(i=0,\cdots,n/2-1)$ represents both states $n/2+i$ and $n/2-1-i$ in the original Markov chain.

Let π_i be the probability of being in state i. Since the system is in statistical equilibrium

$$\tfrac{1}{4}p^2\pi_{i-1} = (\tfrac{1}{4}p^2 - p + 1)\pi_i$$

where $i=1,\cdots,n/2-1$. Also,

$$\sum_{i=0}^{n/2-1} \pi_i = 1.$$

Solving these two equations in the usual way yields

$$\pi_i = \frac{a-1}{a^{n/2}-1}a^i$$

where

$$a = \frac{p^2}{p^2-4p+4}.$$

Let ρ be equal to the probability of being in state 0 in Fig. 18 which is also equal to the probability of being in state $n-1$. By symmetry, $\rho = \pi_{n/2-1}/2$. In a given switching module, the probability of deflection is $2p^2\rho/4$ since, in Fig. 18, if "2L" arrive in state $n-1$ or "2U" arrive in state 0, one cell must be deflected. $p^2/4$ is the probability of "2L"—also equal to the probability of "2U"—and the factor of 2 accounts for there being two states (0 and $n-1$) in which deflection may occur. For a given cell, the probability of deflection is half that because one out of two cells is being misdirected, assuming it is chosen randomly. The deflection probability is thus

$$W = \frac{p\pi_{n/2-1}}{8}.$$

[3]This was pointed out by Alberto Bononi of SUNY at Buffalo.

A factor of p is used above rather than p^2, since the probability of any particular cell being deflected is desired rather than the probability of a cell being deflected on a given link in a given timeslot.

Since Fig. 18 is symmetrical, the average delay through a switch module is $(n-1)/2$.

APPENDIX C
ANALYSIS OF THE MERGING MODULE

Here, we consider the problem of unequal traffic on the inputs i.e. the probability of a cell being present on the top (tributary) input is p_T and on the bottom (main) input it is p_M. All the cells are directed to the lower output port; the upper output port is reserved for those that are deflected. In these circumstances, it can be shown that

$$\begin{aligned} P(N) &= (1-p_T)(1-p_M) \\ P(L) &= p_T + p_M - 2p_Tp_M \\ P(2L) &= p_Tp_M \\ P(U) &= 0 \\ P(2U) &= 0 \\ P(U+L) &= 0. \end{aligned}$$

The state diagram of Fig. 17 may then be modified by making substitutions as before (Fig. 19). Observe that the diagram is contrived so that the state will tend, if possible, to decrease toward zero. If the state is zero, and no packets arrive, it will remain in that state. Conversely, if two packets arrive (2L) in state $n-1$, the merger stays in that state, and one packet is deflected to the overflow output immediately, while the other is delayed by $n-1$ timeslots and sent to the main output. In statistical equilibrium, $P(N)P_i = P(2L)P_{i+1}$. Substituting and solving in the normal way yields

$$P_{n-1} = \beta^{n-1}\left(\frac{1-\beta}{1-\beta^n}\right)$$

where

$$\beta = \frac{p_Tp_M}{(1-p_T)(1-p_M)}.$$

The probability of packet loss in a given time interval is $P_I = P_{n-1}p_Tp_M$, so the probability of a particular cell in the main input being lost to the overflow is $P_I/2p_M$; likewise, for the tributary input, $P_I/2p_T$. The average probability of any input cell being lost (i.e. deflected) is thus

$$W = \frac{p_Tp_M}{p_T+p_M}\left(\frac{1-\beta}{1-\beta^n}\right)\beta^{n-1}.$$

If the main traffic is given priority such that none of its cells overflow, the overall cell-overflow probability (experienced only by the tributary) is increased. The new expression is

$$W = p_M\left(\frac{1-\beta}{1-\beta^n}\right)\beta^{n-1}.$$

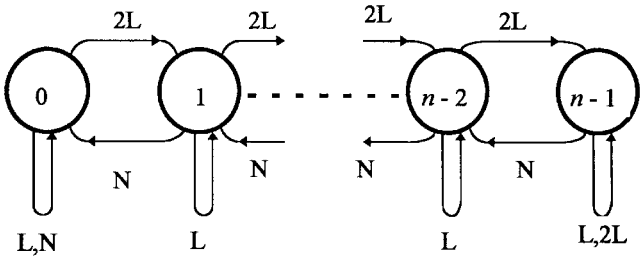

Fig. 19. State diagram of a traffic-merging module.

APPENDIX D
PROOF OF MODULE OPERATION

It is sufficient to assume that two cells enter the switching module on each timeslot, as discussed in Appendix A. To show that the switching modules operate as stated, consider their definition in Fig. 2. Define a **connection** as either a delay line or a link between switch devices in the finished module. Here, a **virtual connection** is an ordered pair of a connection and a timeslot. As discussed earlier, the delay to be experienced by a cell is decided before it enters the fabric. Also, for any input/output pair and any given delay, there is only one possible path through the module. A new cell can be thought on as reserving virtual connections for its use, where each virtual connection represents a connection and the timeslot on which the cell will enter that connection. The reserved set in a particular state is the set of virtual connections reserved for all the cells currently in the module. When two new cells enter the module, the appropriate virtual connections are added to the reserved set. To demonstrate the validity of the modules, it will be necessary to show that no conflict exists between the reserved set and the virtual connections for each two new cells entering the module.

The proof is by induction. Consider Fig. 2, and suppose that the fabric $B(n/2)$ operates correctly and that there is a particular reserved set which is always associated with each state. This is obviously true for the trivial case of $B(1)$ and to complete the induction, it will be necessary to show that if it is true for $B(n/2)$, it is true for $B(n)$ as well. Suppose that the right-hand switch device in the lower half of Fig. 2 is in the bar-state. Then it can be seen that if the $B(n/2)$ fabric is in a state s where $s = 0, \cdots, n/2-1$, the whole $B(n)$ fabric will also be in state s. Also, if $B(n/2)$ moves between these states, so will $B(n)$ and the reserved set associated with any state of $B(n)$ will always be dependent only on s. Likewise, if the switch device is in the cross-state, it can be shown that if $B(n/2)$ is in state s then $B(n)$ will be in state $s+n/2$. The "U" and "L" cells emerging from $B(n/2)$ must be exchanged, and this can be done by inverting the state of the rightmost switch in $B(n/2)$.

It now remains to show that transition from state $n/2-1$ to $n/2$ is possible; the proof for the converse is similar. First, observe that the reserved sets for state s and state $n-1-s$ are the same because virtual connections reserved by "U" cells become reserved by "L" cells and vice-versa. So if the module is in state $n/2-1$, it may be put into state $n/2$ without fear

of any conflict arising. Suppose it is in state $n/2-1$, but two "L" cells arrive. Normally, if an "L" and "U" cell arrived, "L" would be delayed by $n/2-1$ timeslots and "U" would be delayed by $n/2$. The two "L" cells may be delayed by $n/2$ and $n/2-1$ without altering the reserved set. The reserved set is the same as for state $n/2$, so the fabric may be put into that state.

This completes the proof since it has been shown that $B(n)$ functions correctly with a particular reserved set for each state.

ACKNOWLEDGMENT

The authors would like to thank M. Tatham of BT Labs for valuable comments on the manuscript, and also wish to acknowledge the original developers of xHatch, H. Jordan, M. Salerno, V. Heuring, and J. Pratt of the University of Colorado at Boulder.

REFERENCES

[1] J. M. Gabriagues and J. B. Jacob, "OASIS: A high-speed photonic ATM switch—Results and perspectives," in *15th Int. Switching Symp.*, Berlin, Germany, Apr. 1995, vol. 2, paper C8.4.
[2] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer, "Photonic packet switches—Architectures and experimental implementations," *Proc. IEEE*, vol. 82, pp. 1650–1667, Nov. 1994.
[3] D. Cotter, I. Hawker, A. Hill, and V. Tandon, "New network infrastructures for the 21st century," *Brit. Telecommun. Eng. J.*, vol. 13, no. 3, Oct. 1994, pp206–212.
[4] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.*, vol. 12, Mar. 1964, pp. 1–9.
[5] F. Borgonovo, "Deflection routing," in *Routing in Communications Networks*, M. Steenstrup, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1995, ch. 9, pp. 263–305.
[6] I. Chlamtac and A. Fumagalli, "An optical switch architecture for Manhattan networks," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 550–559, May 1993.
[7] A. Bononi, F. Forghieri, and P. R. Prucnal, "Soliton ultrafast all-optical mesh networks," *IEE Proc.-J.*, vol. 140, no. 5, pp. 285–290, Oct. 1993.
[8] 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.
[9] 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.
[10] H. Okayama and M. Kawahara, "Low-crosstalk 2×2 digital optical switch," *Electron. Lett.*, vol. 30, no. 5, pp. 403–405, Mar. 1994.
[11] G. Sherlock, J. D. Burton, P. J. Fiddyment, P. C. Sully, A. E. Kelly, and M. J. Robertson, "Integrated 2×2 optical switch with gain," *Electron. Lett.*, vol. 30, no. 2, pp. 137–138, 1994.
[12] D. B. Sarrazin, H. F. Jordan, and V. E. Heuring, "Fiber optic delay line memory," *Appl. Opt.*, vol. 29, no. 5, 10 Feb. 1990, pp627–637.
[13] D. K. Hunter and D. G. Smith, "New architectures for optical TDM switching," *J. Lightwave Technol.*, vol. 11, pp. 495–511, Mar. 1993.
[14] C. E. Love and H. F. Jordan, "Format insensitive packet synchronizer," in *Proc. 33rd Annu. Allerton Conf. Commun., Contr. and Comput.*, Oct. 4–6, 1995, pp. 568–577.
[15] M. Shabeer, J. K. Lucek, K. Smith, D. Cotter, and D. C. Rogers, "Self-Synchronization scheme for high-speed photonic networks," *Electron. Lett.*, vol. 31, no., 17, pp. 1476–1477, Aug. 1995.
[16] D. Cotter, J. K. Lucek, M. Shabeer, K. Smith, D. C. Rogers, P. Gunning, and D. Nesset, "100 Gbit/s packet self-routing using all-optical 6-Bit 'keyword' address recognition," *Electron. Lett.*, vol. 31, no. 25, pp. 2201–2202, Dec. 1995.
[17] D. K. Hunter and I. Andonovic, "Architectures for synchronous optical TDM switching employing semiconductor laser amplifiers," *IEE Proc. Part J*, vol. 142, no. 3, pp. 132–142, June 1995.
[18] T. H. Gilfedder, P. J. Legg, D. K. Hunter, I. Andonovic, and M. Shabeer, "Modeling of the performance of optical TDM switching networks in the presence of interferometric noise due to interchannel crosstalk," in *European Symp. Advanced Networks and Services*, EOS/SPIE/EC, Amsterdam, The Netherlands, paper 2450–10, Broadband Strategies & Technologies for Local, Metropolitan and Optical Access Networks, Mar. 20–23, 1995, SPIE, vol. 2450, pp. 484–494.
[19] P. J. Legg, D. K. Hunter, I. Andonovic, and P. E. Barnsley, "Inter-channel crosstalk phenomena in optical time division multiplexed switching networks," *IEEE Photon. Technol. Lett.*, vol. 6, pp. 661–663, May 1994.
[20] P. J. Legg, "The impact of interferometric noise on the performance of optical communication networks," Ph.D. dissertation, Univ. of Strathclyde, Glasgow, Scotland, 1995.
[21] E. L. Goldstein, L. Elskildsen, and A. F. Elrefaie, "Performance implications of component crosstalk in transparent lightwave networks," *IEEE Photon. Technol. Lett.*, vol. 6, pp. 657–660, May 1994.
[22] K. Padmanabhan and A. N. Netravali, "Dilated networks for photonic switching," *IEEE Trans. Commun.*, vol. COM-35, pp. 1357–1365, Dec. 1987.
[23] N. F. Maxemchuk, "Regular mesh topologies in local and metropolitan area networks," *AT&T Tech. J.*, vol. 64, no. 7, pp. 1659–1685, 1985.
[24] T. Kubo and K. Yoguchi, "Highway transfer: A new forwarding technique for real-time applications," in *Proc. INFOCOM'90*, 1990, pp. 403–408.
[25] C. Partridge, *Gigabit Networking.* Reading, PA: Addison-Wesley, 1994, ch. 11, pp. 253–263.

David K. Hunter (S'88–M'90) was born in Glasgow, Scotland. 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 Labs, Ipswich, England, and in 1994, he was a Visiting Scholar at the University of Pittsburgh, Pittsburgh, PA. Since 1995, he has held a five-year EPSRC Advanced Fellowship. His research interests include optical switch architectures, telecommunications switch modeling, network routing, and network survivability.

Dr. Hunter is an Associate Member of the IEE.

David Cotter received the B.Sc. degree in electronic science and the Ph.D. degree in nonlinear optics from the University of Southampton, England, in 1973 and 1976, respectively.

In 1976–1977, he was the 1851 Royal Exhibition Fellow and during 1977–1980, he was with the Max–Planck Institute for Biophysical Chemistry, Goettingen, Germany. Since 1980, he has been with BT Labs, where he is currently the Advisor on Photonic Networks and has research interest in developing radically new networks using ultrafast photonic technologies. He has published numerous journal and conference papers, holds several patents, and is a coauthor of *Nonlinear Optics of Free Atoms and Molecules* (New York: Springer-Verlag, 1979) and *The Elements of Nonlinear Optics* (New York: Cambridge University Press, 1990).

Dr. Cotter is a Fellow of the Institute of Physics.

R. Badlishah Ahmad was born in Selangor, Malaysia. He received the B.Eng. degree in electronic engineering with honors from the University of Glasgow, Scotland, in 1994, and the M.Sc. degree in optical electronics in 1995 from the University of Strathclyde, Scotland. He is currently pursuing the Ph.D. degree at the University of Strathclyde in ultrahigh-speed optical packet switching.

W. David Cornwell was born in Glasgow, Scotland. He received the B.Eng. degree in electrical and electronic engineering with first class honors from the University of Strathclyde, Scotland, 1994. He is currently pursuing the Ph.D. degree at the University of Strathclyde in optical interferometric noise models and their experimental verification.

Mr. Cornwell is a Student Member of the IEE.

Tim H. Gilfedder was born in 1968 and received the Bachelor's degree in theoretical physics from Exeter University, England, in 1989. He received the M.Sc. degree in optical electronics from the University of Strathclyde, Scotland, in 1992 for which the modeling of optical networks was investigated. This work was subsequently extended to form the basis of his Ph.D.

In 1994, he spent nine months at British Telecom Laboratories, Ipswich, England, to investigate interferometric noise effects in a variety of optical switching applications. He is a graduate member of the British Institute of Physics.

Peter J. Legg received the M.A. degree in physics from Cambridge University, Cambridge, England, in 1988 and the Ph.D. degree in electronic engineering at the University of Strathclyde, Scotland, in 1995.

For seven years, he has undertaken research into fiber optic communication at the device, system and network level—his doctoral dissertation addressed interferometric noise in optical networks. He now works within the GSM Research Department of Motorola (UK), considering high bandwidth wireless ATM networks.

Ivan Andonovic (M'79) received the B.Sc. degree with honors in electrical and electronic engineering from the University of Strathclyde in 1978. He also received the Ph.D. degree in optical waveguide modulator devices on $LiNbO_3$ in collaboration with nearby Glasgow University, Scotland.

Following a three-year period as a Research Scientist responsible for the design, manufacture, and test of guided-wave devices for a variety of applications, he joined the electronic and Electrical Engineering Department at Strathclyde University in 1985. His main interests center on the development of guided-wave architectures for implementing optical signal processing and switching functions for optical networks. He has recently returned from a two-year Royal Society Industrial Fellowship in collaboration with BT Laboratories during which time he was investigating novel approaches to networking. He has edited two books and authored/coauthored four chapters in books and over 90 journal and conference papers.

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