Design of Transparent Optical Multihop Shuffle Networks

Seung-Woo Seo, Andrew Myers*, Chiangling Ng, Hisashi Kobayashi, Paul R. Prucnal
Department of Electrical Engineering
Department of Computer Science*
Princeton University
Princeton, NJ 08544

Abstract

Recent advances in device technologies have opened new opportunities for implementing high-performance optical networks. The potentially multi-terahertz bandwidth made available with the advent of optical fibers can be exploited by eliminating the bottleneck caused by electro-optic conversion. In this paper, we present some of the considerations in the design of all-optical multihop networks. Each node of the network is composed of several components including low-loss photonic switches, a switch controller, an optical data processing unit, and a user interface. Although an overall performance of multihop networks is determined by a number of factors, we focus on two of them: network topology and node architecture. As for network topology, we provide comprehensive simulation results on the effects of topology variations to the performance of networks. As for node architecture, we propose a new switch architecture called Butterfly switch, and analyze its connection and permutation capability.

1. Introduction

High-capacity, bandwidth-critical networks are crucial to many emerging applications in broadband communication and multicomputer interconnects. Although the optical fiber medium provide several times larger capacity than traditional transmission media, the bandwidth bottleneck lies in the opto-electronic (O/E) or electro-optic (E/O) conversion required at each of the geographically distributed nodes in the network. In *transparent optical networks* (TONs)[1,2] the electronic bottleneck may be overcome by allowing signals to remain in an optical format until they arrive at their destination. Routing control is simplified by processing only a small fraction of the information flowing through at each node (e.g., packet header). In packet-switching multihop TONs, several basic functions are required including routing of packets from source to destination and synchronization of packets at multiple switch input ports[3].

In this paper, we consider some key design issues of multihop TONs. Although an overall performance of multihop networks is determined by various factors, we focus on two of them: network topology and node architecture. These two factors are particularly important when the deflection routing algorithm is used as packet routing control. Deflection routing (which is called "hot-potato" routing in case of no buffer[4]) is a technique that maintains a limited buffer size, while providing a network performance comparable to the store-and-forward scheme which requires an ample buffer.

The paper is organized as follows: In Section 2, we discuss the characteristics of multihop networks. We present a performance analysis of a new class of shuffle networks in Section 3. In Section 4, we introduce a node structure of the network for deflection routing as well as a new switch architecture called "Butterfly switch." In Section 5, conclusions are offered with some remarks.

2. Space-switched Multihop Networks

So far, multihop TONs with space-division switches have been studied by many researchers. The number of inputs and outputs at each node is limited to a fixed number, thus packets arrive at their destinations after multiple hops through intermediate nodes. Optical channels are provided on dedicated fibers between nodes, whereby each fiber carries time-division multiplexed (TDM) or wavelength-division multiplexed (WDM) data, allowing simultaneous transfer of high-speed data and video signals.

Space-switched multihop networks provide an inherent architecture for statistical multiplexing. Each node can generate and inject packets whenever channels are available. Space-division switches allow a complete set of concurrency among multiple users. Throughput-delay performance is improved by eliminating conventional bandwidth-distance limited multiple access protocols. Photonic space-division switches can provide full transparency and scalability in multihop optical networks[3]. Space-switched multihop networks have several fundamental advantages including modularity, robustness, and survivability in the presence of link failure.

For multihop network topologies, regular two-connected networks (i.e., each node has two inputs and two outputs), such as the Manhattan Street Network (MSN) and the shuffle network, have been considered [9]. In multihop networks, sophisticated routing and flow control requires dynamic space-division switching. In TONs, routing and flow control at each node should be as simple as possible since data rate is extremely high. This

implies that each node needs to process only a small fraction of the information (header of a packet) flowing through it in real time, which in turn simplify node structure and speed up its operation.

Deflection routing is a technique that maintains a bounded buffer size, while providing a compatible network performance to the store-and-forward scheme. By utilizing only the local information at each node, this technique can adapt to load or topology variations of the network. This may be compared with the store-and-forward algorithm which cannot adapt quickly to network status. Comprehensive simulation and analysis of the effect of buffers in deflection routing have been reported (for example, [9]).

If deflection routing is used as a routing principle, network throughput may be contingent upon network topology. Multihop network performance is optimized by minimizing the probability of deflection, and the number of extra hops caused by a deflection. Hence, when designing a multihop network, it is important to optimize its topology to achieve minimum diameter and deflection cost, and a maximum percentage of "don't care" nodes.

Many research results have been reported on the effect of network topology to the performance of deflection routing[5]. In [6], we examined a way to improve network performance through topology variations by using a new definition of a shuffle network (called generalized shuffle networks). In our preliminary study of the generalized shuffle networks, an average number of hops in various topologies was derived[6].

3. Optimal Multihop Network Topology

In the generalized shuffle networks, the tight relationship (i.e., $N=kn$, and $n=p^k$ with some prime number p for a given N) between the number of stages (k) and the number of nodes per stage (n) is removed. The network topology becomes more flexible by allowing n to be independent of k, i.e., $N=kn$. This relaxed constraint enables us to select more distinct values of N. In the following, we assume that each node has two input/outputs, one local memory, and one TX/RX.

At each network cycle, the packets incoming and stored in memory are serviced for the desired outputs. The service can be done based on two priority schemes: standard and equal priority. In a standard priority scheme, all packets arriving at each node are scheduled by the following priority rules: (1) The packet in the memory is serviced first; (2) The packet in inputs for RX is serviced next; (3) If input packets are contending for the same output, one of them will be selected randomly; (4) The packet in TX has the lowest priority. On the contrary, in an equal priority scheme all packets are scheduled based on equal priority rules.

We conducted an intensive simulation as well as an analysis of throughput-delay performance of the generalized shuffle networks. The results are shown in Figs. 1 and 2 for N=256. When N=256, the generalized shuffle networks can be realized in eight different ways, and the best performance is obtained if the number of stages (k) is equal to 4. Although it is intuitively clear that the equal priority scheme results in a worse performance than the standard priority scheme due to possible packet losses, the simulation results quantitatively verify that the standard priority scheme provides up to a 30% improvement in throughput over the equal priority scheme under heavy traffic conditions. It is also shown that the packet loss significantly (up to 70%) relieves a network congestion, and reduces packet delays in the network, especially when the number of stages is large.

Other performance measures such as normalized packet delay, TX delay and packet loss probability are also compared for various shuffle network topologies. (The previous studies[5,8,9] did not consider the effect of TX delays, but assumed only network delays in calculating the overall delays. Instead, we note the fact that the delay in TX buffer before a packet is injected into the network is not negligible if the network is heavily congested.) In calculating normalized packet delays, a theoretical bound on the minimum number of hops in two-connected networks that holds independently of network topology was used to normalize the packet delays in each case. In the case of N=256, this value was obtained as 6.055 [9].

The results show how many packets should be dropped to make the network delay satisfactory in heavily congested networks. The results also imply that a multihop network can be designed in different manners depending on the required application, e.g., in voice transmission where a small amounts of packet loss may be tolerated, the equal priority scheme can be adopted for reducing the delay, while in data transmission where real time delivery is not of primary importance, the standard priority scheme may be used for high throughput. For a proper network management, these two priority schemes can be used together in a flexible manner.

4. All-Optical Node Structure for a Multihop TON

Building a multihop network for deflection routing requires an optimal low-loss photonic space-switched node architecture as well as an efficient network topology. The node modules we investigate have two input ports, two output ports, one transmitter (TX), one receiver (RX) and two additional ports connected to a fiber-loop buffer. Fig. 3 shows a general two-connected node structure when one shared memory is introduced. Each node is comprised of all-optical switches, switch controller, optical data processing unit (optical packet generation and compression unit, optical routing controller, and all-optical demultiplexer), and user interfaces.

To accommodate ultra-fast bit rates and scalability, the node structure should be simple and low-loss, contain a minimum number of photonic crossbar switches and use non-priority deflection routing with one or zero fiber-loop buffers. The switches connected to the TX and RX ports in Fig. 4 are called add and drop switches, respectively. The remaining switches are for routing[8]. If a buffer is not used, simple hot-potato routing can be implemented. With hot-potato routing, the main routing switch in a node is simply implemented by

a 2×2 crossbar. It is noted that for the single-buffer case, a rearrangeable non-blocking 4×4 switch requires at least five 2×2 switches. Simple node designs with fewer than five 2×2 switches are not non-blocking, and provide suboptimal performance.

Node control must successfully handle destination addresses of input packets, locally generated packets for transmission, as well as the packets in buffers. The packets arriving in each time slot can be empty, destined for that node, destined to exit at output 1 or 2, or can be "don't care" packets (when both outputs provide equivalent shortest paths to the destination). Deflection occurs when two or more packets contend for the same output and there is not enough memory to store the losers. The algorithm should minimize the number of missed packets as well as the average number of hops to a destination. The routing algorithm is executed by the node controller, which may be implemented with all-optical processing or with slower hybrid optical/electronic processing.

By simply adding a single-packet optical delay-line buffer, the number of deflections can be significantly reduced. It is noted that although the single-buffer routing scheme yields a significant performance improvement over hot potato, adding more memory does not appreciably improve the performance, but introduces complexity to a node structure and a routing algorithm[8]. This result in [8] shows a trade-off between the structural complexity of a deflection routing node and network performance.

4.1. Butterfly Switch Architecture

For deflection routing, various all-optical switch architectures for multihop networks have been proposed. Chlamtac and Fumagalli[7] discuss packet deflection switches for two-connected multihop networks using optical delay lines for storing and switching packets. Bononi and Prucnal[8] propose various node structures with single transmitter/receiver and single buffer (Fig. 4 (a) and (b)). In both schemes of [7] and [8], electronically controlled 2×2 $LiNbO_3$ crossbar switches are used as basic building blocks. In these schemes, arriving packets and packets at a node are sorted according to priority rules to set up the 2×2 switches. However, the proposed architectures suffer from a large signal loss (about 12 dB per stage) due to multiple stages of 2×2 switches (more than three stages). Furthermore, the routing algorithm becomes very complex due to irregularity of the switch.

Unlike the switch configurations proposed so far, a simple switch architecture called a Butterfly switch provides acceptable performance at a reasonable cost for two-connected multihop networks with single-buffer deflection routing. The Butterfly switch has a regular structure with four 2×2 switches in two stages, and two adjacent stages are connected in a shuffle manner as shown in Fig. 5. The switch can be configured by a very simple control algorithm requiring only bit-level manipulations. In addition, the packets arriving at two input ports experience an equal power loss, and this loss is less than other schemes due to the regular structure of the switch.

If multiple packets want to exit through the same output port at the same time, some kinds of contention control must be used. Otherwise, the simultaneous connection capability of a switch is determined by the permutation capability which allows multiple distinct connections to be realized at the same time within the switch. Since the Butterfly switch has only four switches, it is internally blocking. However, the permutation realizability of the Butterfly switch is maximized by considering that some of the input/output connections in the switch may not always be used without causing much performance degradation:

1. The packet which is currently in the buffer is not sent to the buffer again.
2. The node inserting a packet does not send the packet to itself.
3. The node inserting a packet does not send the packet to its own local buffer.
4. The packets arriving at both inputs are not originally destined to memory.

Regulating the first connection is related to a priority control, which implies that the optically stored packet has to be sent out first without being recirculated in the delay line, since this may cause the optical power to be reduced to undetectable levels, or reduce the signal-to-noise ratio to an unacceptable level due to accumulated noises from optical amplifiers. With these considerations, it is shown that the Butterfly switch can realize about 80% of the permutations compared to the ideal five switch case.

The realizability of all possible permutations can easily be checked. Assuming the sequence (0 1 2 3) represents the output sequence (NM O_1 O_2 RX) in a permutation, i.e.,

$$P = \begin{pmatrix} I_1 & I_2 & M & TX \\ NM & O_1 & O_2 & RX \end{pmatrix},$$

most of the sequences except for (2 3 0 1), (3 2 0 1), (2 3 1 0), (3 2 1 0), and (1 0 3 2) can be realized in the Butterfly switch. Note that input ports of the Butterfly switch are arranged in the order of (I_1 I_2 M TX) so that two input ports I_1 and I_2 have an equal priority (i.e., fair) in contention. Similarly, M and TX are connected to the same input switch so that the packet in memory can always have a higher priority than the packet in TX. For example, if the permutation is (3 2 1 0), I_1 is contending with I_2 for the same output port of the first stage switch. Since I_1 wants to be connected to RX and has a higher priority than I_2, I_2 is stored in the buffer for the next cycle. In this permutation, the packet from TX is not sent to a local memory buffer, which enables the connection between M and O_1 to be realized without any contention.

5. Conclusions

In this paper, we presented two design considerations in all-optical multihop networks: the network topology and node architecture. It was shown that the generalized shuffle networks can offer a wide range of selections of multihop network topology. The Butterfly switch we propose can achieve a satisfactory

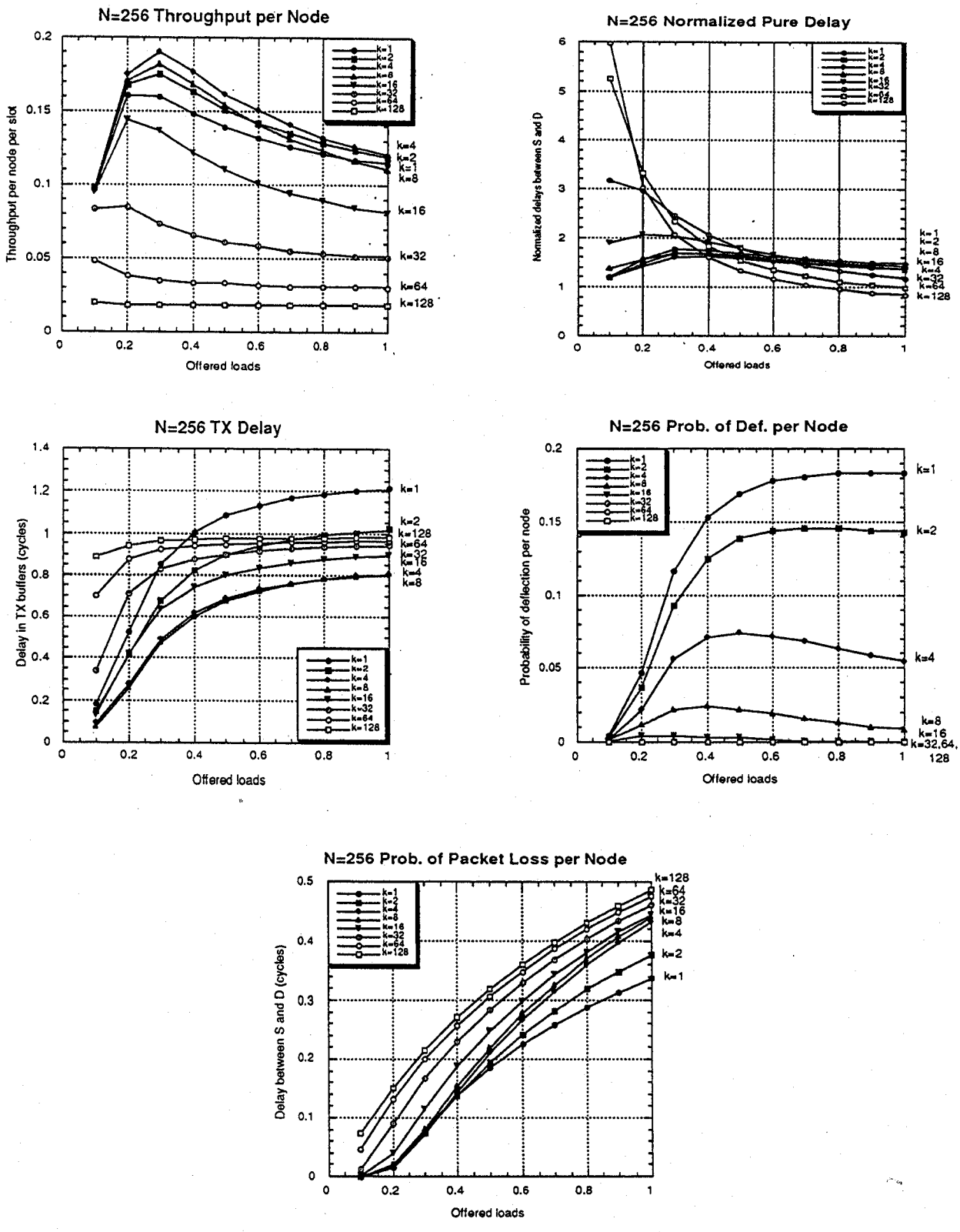


Fig. 2: Simulation results with equal priority for N=256

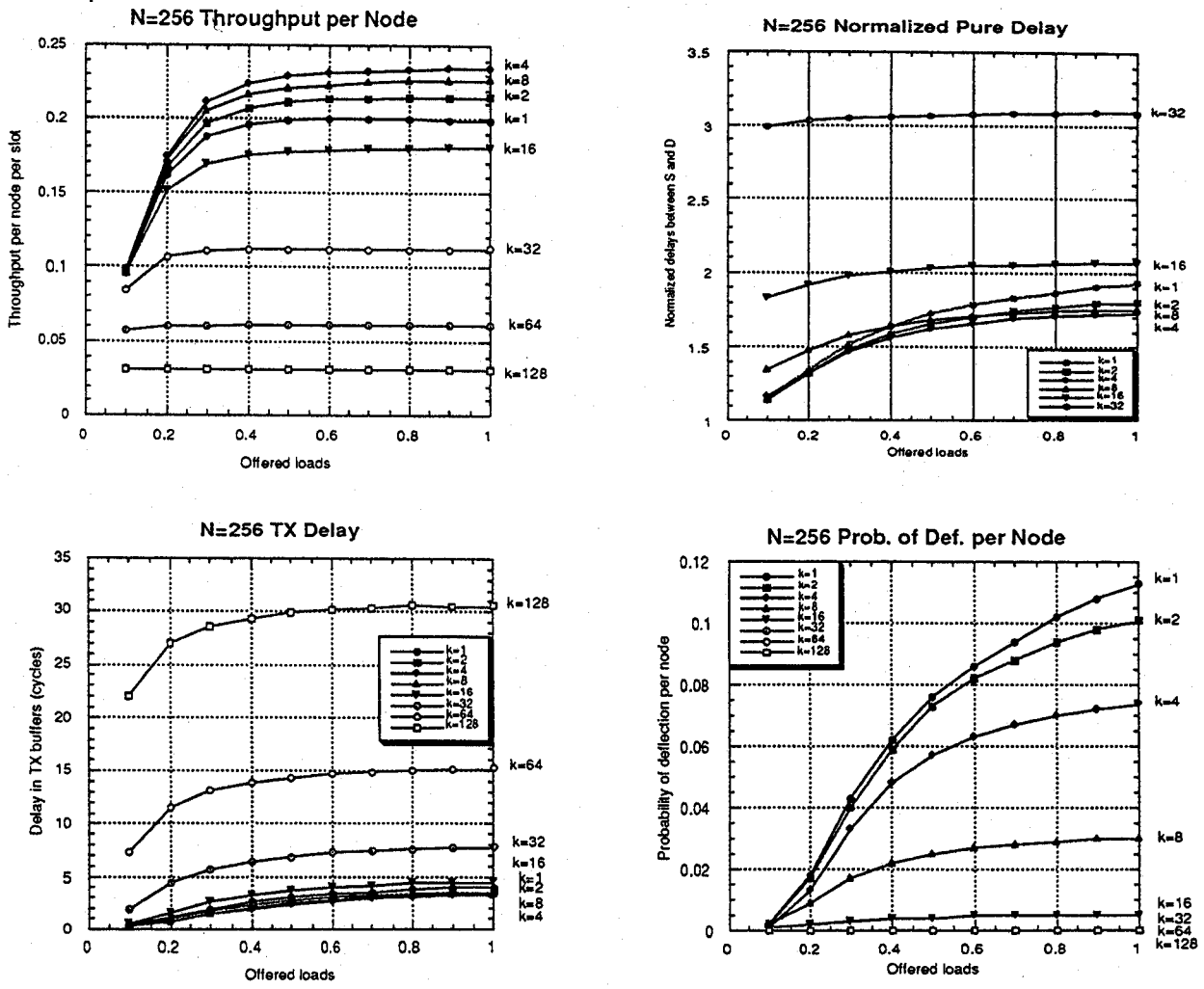


Fig. 1: Simulation results with standard priority for N=256

throughput at the reasonable expense of additional hardware, cost, and control complexity. We have successfully designed a switch controller and its peripheral board which are run by a Pentium-based PC. This controller uses the standard priority scheme in arbitrating possible contentions among arriving packets.

References

1. A.S. Acampora, M.J. Karol and M.G. Hluchyj, "Terabit lightwave networks: The multihop approach," *AT&T Tech. Journal*, Vol. 66, pp. 21-34, Nov./Dec., 1985.
2. M.G. Hluchyj and M.J. Karol, "ShuffleNet: An application of generalized perfect shuffles to multihop lightwave networks," *Proc. of IEEE INFOCOM*, pp. 4B.4.1-4B.4.5, 1988.
3. P.R. Prucnal, "Optically-process self-routing, synchronization and contention resolution for 1D and 2D photonic switching architectures," *IEEE J. Quantum Electronics*, 29, Vol. 2, pp. 600-612, 1993.
4. P. Baran, "On distributed communication networks," *IEEE Trans. Commun. Syst.*, Vol. 12, pp. 1-9, Mar. 1964.
5. A. Krishna and B. Hajek, "Performance of shuffle-like networks with deflection," *Proc. of IEEE INFOCOM*, Vol. 2, pp. 473-480, 1990.
6. S.-W. Seo, P.R. Prucnal, H. Kobayashi, and J.B. Lim, "On the Performance of a Class of Multihop Shuffle Networks," *Proc. of 95' International Conference on Communications*, Vol. 2, pp. 1211-1215, Seattle, June 1995.
7. I. Chlamtac and A. Fumagalli, "An optical switch architecture for Manhattan networks," *IEEE Journal of Selected Areas on Communications*, Vol. 11, pp. 550-559, May 1993.
8. A. Bononi and P.R. Prucnal, "New structures of the optical node in transparent optical multihop networks using deflection routing," *Proc. of IEEE INFOCOM*, pp. 415-422, 1994.
9. N.F. Maxemchuk, "Comparison of deflection and store-and-forward techniques in the Manhattan street and shuffle-exchange networks," *Proc. of IEEE INFOCOM*, pp. 800-809, 1989.

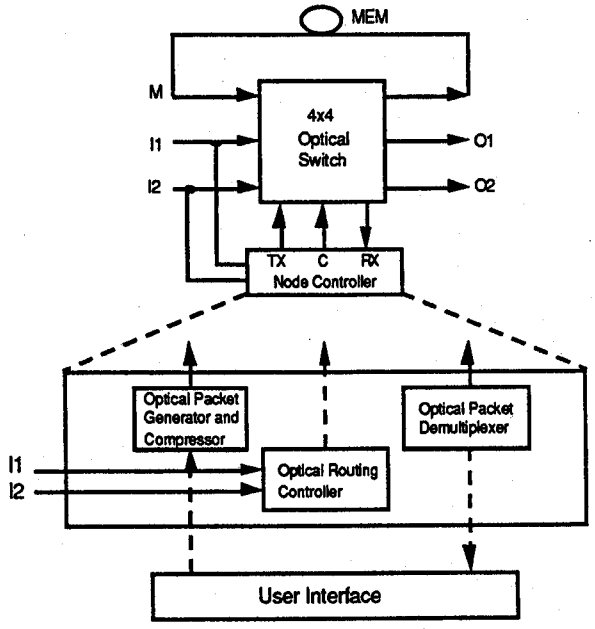
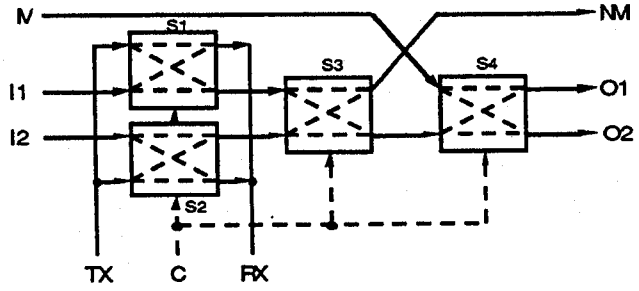
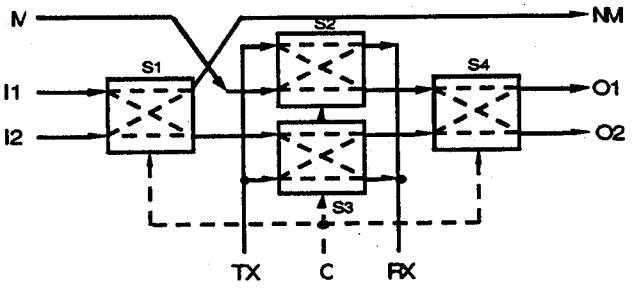


Fig. 3: Complete node structure



(a)



(b)

Fig. 4: Four-switch architectures [8]: (a) Add/drop switch block precedes buffer; (b) the other case

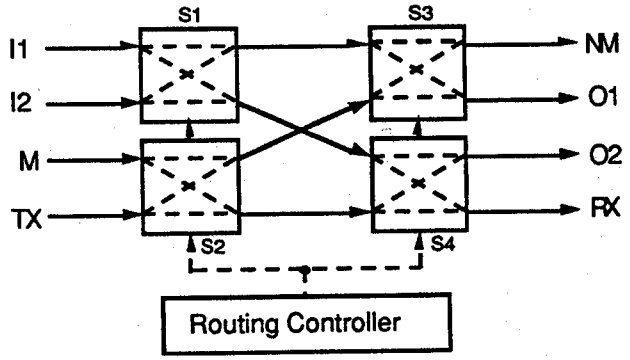


Fig. 5: Butterfly switch with routing controller