Advances in Photonic Packet Switching: An Overview

Shun Yao and Biswanath Mukherjee, University of California
Sudhir Dixit, Nokia Research Center

ABSTRACT

The current fast-growing Internet traffic is demanding more and more network capacity every day. The concept of wavelength-division multiplexing has provided us an opportunity to multiply network capacity. Current optical switching technologies allow us to rapidly deliver the enormous bandwidth of WDM networks. Photonic packet switching offers high-speed, data rate/format transparency, and configurability, which are some of the important characteristics needed in future networks supporting different forms of data. In this article we present some of the critical issues involved in designing and implementing all-optical packet-switched networks.

As telecommunications and computer communications continue to converge, data traffic is gradually exceeding telephony traffic. This means that many of the existing connection-oriented or circuit-switched networks will need to be upgraded to support packet-switched data traffic. The concept of wavelength-division multiplexing (WDM) has provided us an opportunity to multiply network capacity. Current optical switching technologies allow us to rapidly deliver the enormous bandwidth of WDM networks. Of all the switching schemes, photonic packet switching appears to be a strong candidate because of the high speed, data rate/format transparency, and configurability it offers. The goal of this article is to discuss some of the critical issues involved in designing and implementing photonic packet-switched networks. We will first discuss the synchronization issues, then the contention resolution and switching strategies, followed by the header and packet format. Finally, we conclude by describing some of the emerging technologies with the potential to revolutionize optical packet switching.

In general, optical packet-switched networks can be divided into two categories: slotted (synchronous) and unslotted (asynchronous). When individual photonic switches are combined to form a network, at the input ports of each node packets can arrive at different times. Since the state of the switch fabric can only be reconfigured at discrete times, it is crucial for the network designer to decide whether to have all the packets aligned before entering the switch fabric. In both these cases, bit-level synchronization and fast clock recovery are required for packet header recognition and packet delineation.

In a slotted network all the packets have the same size. They are placed together with the header inside a fixed time slot, which has a longer duration than the packet and header to provide guard time. Slotted networks have been extensively studied, while optical fiber was being proposed as the buffer in store-and-forward contention resolution. In most cases optical buffering is implemented by using fiber loops or delay lines with a fixed propagation delay equal to a multiple of the time slot duration. This leads to the requirement that all input packets arriving at the input ports have the same size and be aligned in phase with a local clock reference (Fig. 1).

In an unslotted network the packets may or may not have the same size. Packets arrive and enter the switch without being aligned. Therefore, the packet-by-packet switch action can take place at any point in time. Obviously, in unslotted networks the chance of contention is larger because the behavior of the packets is more unpredictable and less regulated. On the other hand, unslotted networks are easier and cheaper to build, more robust, and more flexible than slotted networks. As shown later in the article, with careful design of node architecture and protocols according to the network specifications, satisfactory performance can be achieved.

Before we delve into the details of synchronization schemes and architectures, it would be insightful to first look at the source for delay variation of packets within the network.

Delay Variation Between Nodes — The time for a packet to travel through a certain distance of the fiber depends on fiber length, chromatic dispersion, and temperature variation. The proposal to use managed internode link delays (to make them equal to an integer number of time slots) is not yet reasonably applicable with current technology. When WDM is used the effect of chromatic dispersion has to be taken into consideration. Chromatic dispersion results in different propagation speeds for packets transmitted on different wavelengths; therefore, different propagation delays occur. For example, with a typical fiber dispersion of 20 ps/nm/km (where ps is the time

 0163-6804/00/$10.00 © 2000 IEEE

unit for delay variation, nm the unit for wavelength difference, and km the unit for propagation distance), a wavelength variation of 30 nm (consistent with the typical erbium doped fiber amplifier, EDFA, 1530–1560 nm window), and a propagation distance of 100 km, the propagation delay variation would be about 60 ns. If dispersion compensation fibers are used, the above delay variation can be reduced by one order of magnitude.

The packet propagation speed also varies with temperature, with a typical figure of 40 ps/°C/km. 100 km of fiber under a temperature variation range of 0–25°C means a delay variation of 100 ns.

The delay variations mentioned above are relatively slow in respect to time; they can be compensated for statically instead of dynamically (on a packet-by-packet basis).

Delay Variations Inside the Nodes — What happens to each packet within the node depends on the switch fabric and contention resolution scheme. In a slotted network that uses fiber delay lines as optical buffers, a packet can take different paths with unequal lengths within the switch fabric. All the considerations given in delay variations in the internode links apply here. It is worth mentioning that the fast time jitter (as compared with the slow delay variation above) induced by dispersion between different wavelengths and unequal optical paths varies from packet to packet at the output of the switch; therefore, a fast output synchronization interface might be required. Thermal effects are smaller here because they vary more slowly and can easily be controlled within the node.

In an optical packet-switched network each switching node is operating in reference to its own internal clock. As is common practice in synchronous digital hierarchy/synchronous optical network (SDH/SONET), this clock is derived from a network synchronization signal distributed throughout the network. Phase noise of the oscillators accumulated along the clock distribution and thermal effects on the optical carriers can all contribute to the impairment of the synchronization signal. According to the SDH network synchronization standard, 1 μs is the maximum wander of the local node clock for a time duration larger than 1000 s. Such slow phase variation has to be taken into consideration.

Synchronization of Slotted Networks

The Impact of Packet Format

Figure 2 shows a function block diagram to show the synchronization stages in a node. An optical splitter splits a small amount of power from the incoming packets. The header reading circuits will recognize a stream with a special bit pattern at the beginning of the packet and then prepare to read the header information. It also passes the timing information of the incoming packet to the control unit to configure the synchronization stages and switch fabric. The input synchronization stage aligns packets before they enter the switch fabric. The output synchronization stage, which is not shown in Fig. 1, is to further compensate for the fast time jitter that occurs inside the node. It may or may not be needed depending on the actual packet format and node architecture.

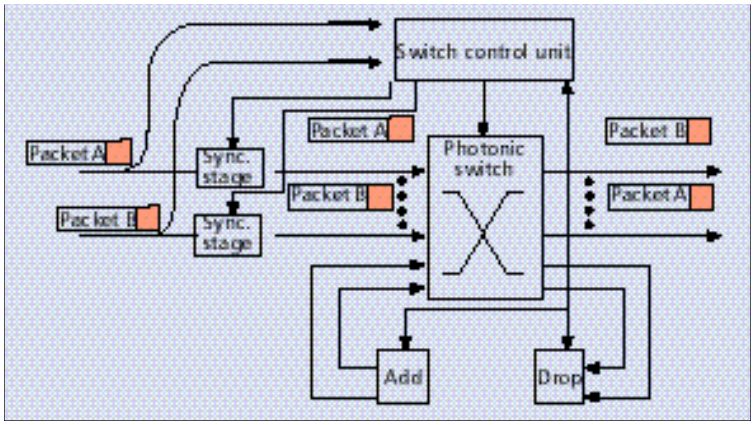

Figure 1. A generic node architecture of the slotted network (contention resolution is not shown here).

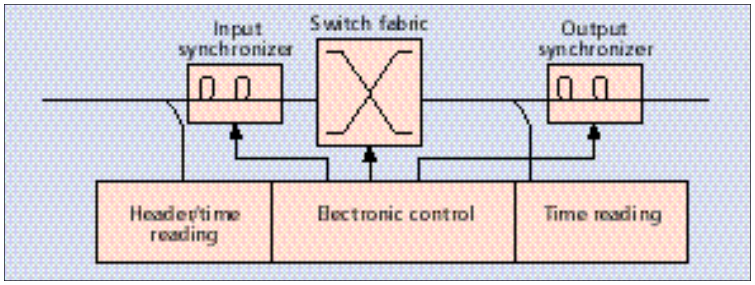

Figure 2. A function block diagram of synchronization of packets.

In general, the required resolution of synchronization (how fine we want to tune the position of each incoming packet) depends on the actual packet format (i.e., the size and position of the header, payload, and guard time). The longer the packet, the more guard times we could put in there without sacrificing link utilization; and more guard time means a less strict requirement for alignment.

With regard to the position of payload, header, and guard times, there are two cases to be considered here, as shown in Fig. 3 [1]:

- Headers define the beginning of time slots. In this case, since we only need to read the information contained in the header, the position of the whole slot will vary slowly with different propagation delay and local clock frequency drift. We do not need to worry about the time jitter before and after the payload.
- A guard time is placed between the header and the beginning of the slot, as well as between the header and the payload. In this case the consecutive packets coming in from the same link can have different misalignments, and fast clock recovery for header reading must be carried out on a packet-by-packet basis.

In the first case the header is aligned precisely at the beginning of the time slot. Therefore, the consecutive packet headers arrive at a node

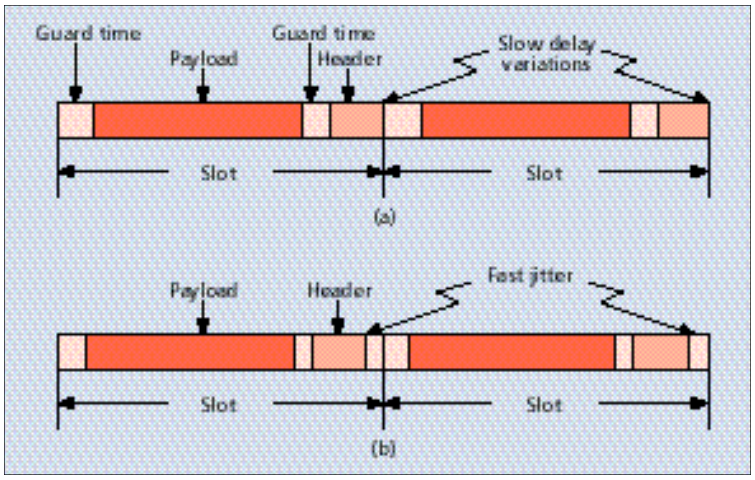

Figure 3. Two possible packet formats that determine the synchronization schemes.

with a constant time interval with respect to the preceding node. Since the header appears exactly at the beginning of each time slot, and we only need to look at the header to switch the packets correctly, the packet delineation and electronic control of the input synchronizer are relatively simple to implement. The input synchronizer stage only has to deal with slow delay variations. However, every effort should be made to keep the header well aligned with the slot boundary. In the diagram in Fig. 2, the output synchronizer stage should be a fast and high-resolution solution to compensate for the jitter of the header occurring in the different optical paths inside the switch fabric on a packet-by-packet basis. In the second case, since a guard time is given between the header and time slot boundary and header jitter is allowed, the header reading electronics has to deal with fast clock recovery of jittered header on a packet-by-packet basis. In other words, the switching node cannot precisely predict the exact arrival time of the header, since it only has knowledge of the time when the slot begins, which is in the middle of a guard time. A fast high-resolution output synchronization stage becomes optional because the header jitter is taken care of by the header reading electronics at the following node's input synchronization stage.

Packet delineation is essential for both slotted and unslotted networks. During packet delineation the incoming bits are locked in phase with the local clock in order for the node to read the header information. As described above, certain packet formats require this bit-level synchronization to be carried out on a packet-by-packet basis. In other words the node should be able to synchronize the header with its clock within several bits. The traditional phase locked loop approach is not applicable here because it requires too many bits to work. Bit-level synchronization is beyond the scope of this article.

Synchronization Schemes

Since packets enter a node from different links, for all the previously stated reasons they can arrive totally out of phase with each other. Figure 4 shows a typical synchronization stage consisting of a series of switches and delay lines, as appears at the input synchronization stage of a node.

Once the bit pattern in the packet header has been recognized and packet delineation has been carried out, the packet start time will be identified, and the control unit will calculate the necessary delay and configure the correct path through these delay lines. The lengths of the delay lines are arranged in an exponential sequence between the 2 x 2 switches; that is, the first delay line is equal to 1/2 time slot duration, the second delay line is equal to 1/4 time slot duration, and so on. The resolution of this scheme is $1/2^n$ time slot duration where n is the number of delay lines. This type of synchronization scheme can be used for both static (slow) and dynamic (fast) synchronization. At system initialization the synchronization is set up to compensate for delay variations between different inputs and to keep this configuration throughout system operation (static). For packet-based (dynamic) synchronization, much faster switches have to be used to operate during the guard time.

From a physical point of view, this scheme introduces insertion loss and crosstalk due to the switches used. Cascading the switches inevitably requires optical amplification, which results in degraded signal-to-noise ratio. Meanwhile, the crosstalk accumulated through the switches also increases the bit error rate. In a multinode network the power penalty caused by all the synchronization stages may significantly impair system performance.

Another approach to synchronization uses a tunable wavelength converter and a piece of highly dispersive fiber (Fig. 5). Since the light propagation speed in the highly dispersive fiber depends on the wavelength of the packet, by properly converting the wavelength of the incoming packet we can achieve a desired delay. It should be noted that the tuning characteristic of the tunable wavelength converter is not continuous, but consists of small steps; therefore, there is a finite resolution for the synchronization.

Unslotted Networks

Figure 6 shows the general node architecture and packet behavior of unslotted networks. (Note the absence of synchronization stages and packet alignment.) The fixed-length fiber delay lines are used only to hold the packet when the header is being processed and the switch fabric reconfigured. There is no packet alignment stage, and all the packets go through the same amount of delay in the same relative position in which they arrived, provided there is no contention. With contention, some kind of contention resolution, such as buffering, space deflection, or wavelength conversion, must be used. We will discuss contention resolution in more detail in the following section. Obviously, unslotted networks are easier to build because there are no complex synchronization stages. On the other hand, given the same traffic load, the link throughput is lower than in slotted networks because contention is more likely to occur.

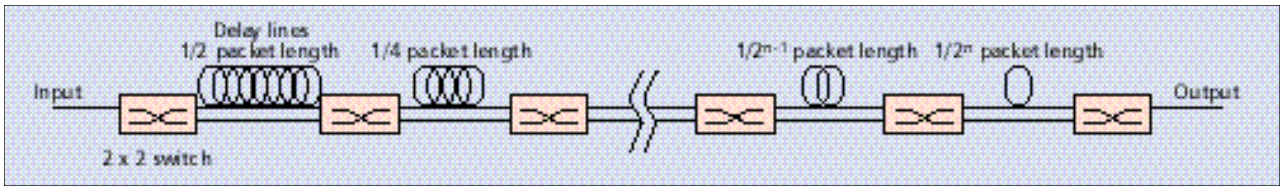

Figure 4. A scheme for the input synchronization stage in a node.

CONTENTION RESOLUTION

In a packet-switched network each packet has to go through a number of switches to reach its destination. When the packets are being switched, contention occurs whenever two or more packets are trying to leave the switch from the same output port. How contention is resolved has a great effect on network performance. Here we will look at three types of contention resolution: optical buffering, deflection routing, and wavelength conversion.

OPTICAL BUFFERING

In electronic routers contention is usually resolved by a store-and-forward technique, which means that the packets in contention are stored in a queue and sent out one by one. This is possible because of the available random access memory (RAM). In an optical switch we have to take a different approach because there is no ready-to-use optical RAM. The main difference between electronic RAM and an optical buffer is that optical buffers must be implemented with delay lines, which are fixed-length fibers. Once a packet has entered the fiber, it must emerge from the other end after a fixed amount of time; there is no way to retrieve the packet anytime earlier (except for recirculation fiber loops, which will be discussed later.)

There are various designs of node architecture applying optical buffers, and there are different ways to categorize them. One way is to compare them to the buffering in electronic switches (input, output, shared, and recirculating buffering). There is a simpler, more direct way to categorize them. In general, the optical buffer can be categorized into single- or multistage, forward or feedback. (A stage is a single continuous piece of delay line.) We will not look at multistage feedback buffering since it is seldom proposed.

The first example, proposed in the European Advanced Communications Technology and Services (ACTS) Keys to Optical Packet Switching (KEOPS), is a broadcast-and-select space switch using single-stage forward buffering for contention resolution (Fig. 7). The wavelength converters encode the packet streams entering each input; therefore, the packets on each input are distinguished by a separate wavelength. The streams are then combined by a multiplexer and distributed to k groups of delay lines of different lengths, which give the packets the necessary delays to resolve contention. By means of semiconductor optical amplifier (SOA) gates and passive couplers, each output port is able to select the packets with proper delays. At the final stage, the demultiplexer, SOA gates, and multiplexer can select one packet from a specific input port. In this architecture there is only

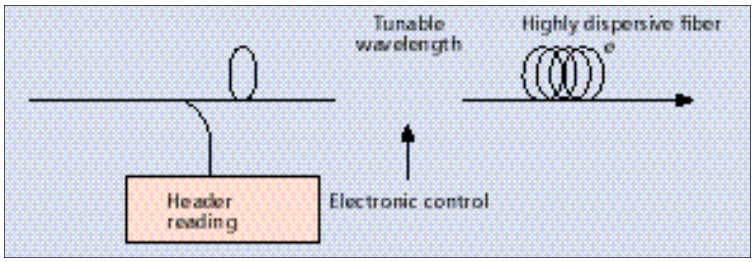

Figure 5. Synchronization using a tunable wavelength converter and high dispersion fiber.

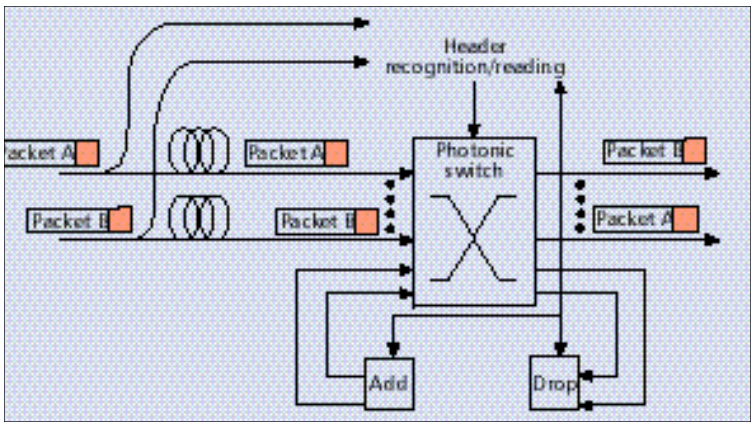

Figure 6. A generic node architecture of the unslotted network (contention resolution is not shown here).

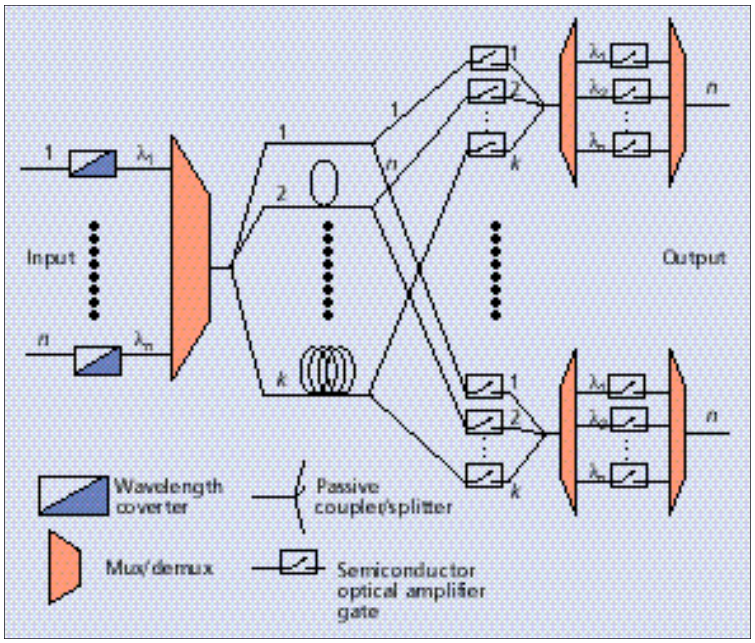

Figure 7. The broadcast-and-select switch proposed in the KEOPS project.

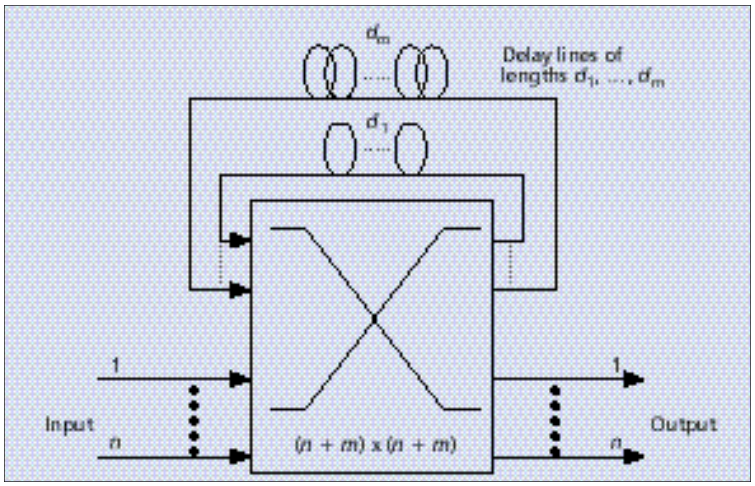

Figure 8. The shared memory optical packet switch.

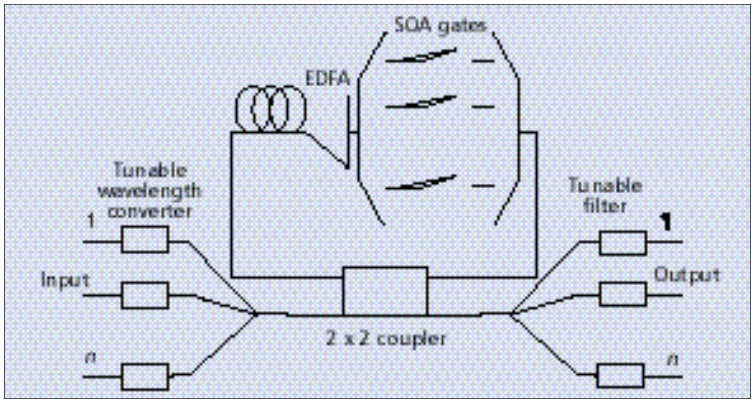

Figure 9. The fiber loop memory switch from ATMOS.

one buffer stage, and each delay line feeds forward to the next part of the switch. Since each packet is broadcast to all delay lines and every output port, it is possible to offer broadcast operation and packet priorities. The drawback is the use of a great number of components and controls, which considerably increases the cost. For example, it needs n wavelength converters, $n \times k + n^2$ SOA gates, and 2n + 1 multiplexers/demultiplexers.

The second example is the shared-memory optical packet (SMOP) switch [2], which belongs to the single-stage feedback category. It is very straightforward to see how the switch works from Fig. 8. The lengths of the delay lines could be 1, 2, 3, …, m packet duration. The (n + m) x (n + m) space switch can switch a packet either directly to an output port or to one of the delay lines, according to how much delay the packet needs. Delay lines of length greater than one packet duration greatly reduce the number of recirculation loops needed, resulting in a reduced need for amplifiers and therefore less noise. This scheme also allows packet priorities since a lower-priority packet may be preempted by being sent to another circulation. Since the number of recirculations a packet is to take is unpredictable, some packets could suffer more power loss than others, making optical amplification necessary. This will inevitably introduce additional signal-to-noise ratio degradation into the recirculating packets.

Another case of single-stage feedback optical buffering is the fiber loop memory switch concept introduced in the Research and Development in Advanced Communications in Europe (RACE) Asynchronous Transfer Mode Optical Switching (ATMOS) project (Fig. 9). The buffer is based on a fiber loop delay line containing multiple wavelength channels. When contention occurs, the input packet is converted to one of the available wavelengths in the loop and kept circulating by activating the corresponding passive fixed filter (i.e., by turning on the related SOA gate). At the input of the loop, half the power enters the loop, and the other half goes toward the outputs through the passive coupler. When the contention is resolved, the packet is switched to the destination link by the proper tuning of the corresponding output tunable filter. At the same time, the passive filter in the loop is turned off to erase the packet in the buffer. It is possible for incoming packets to preempt those that are already waiting; hence, this type of switch can implement packet priorities.

For multistage feed-forward buffering examples, several node architectures applying cascaded 2 x 2 switching elements containing optical buffers [3] have been proposed (Fig. 10). Each of these switching elements provides buffering of one or more packet duration delays in case of contention. A larger switch fabric can be constructed by cascading a number of these 2 x 2 elements in, for example, a Banyan configuration.

There are various designs for optical buffering, for example, the staggering switch [4]; switched fiber delay lines (SDL) such as contention resolution by delay lines (CORD) [5]; and switch with large optical buffers (SLOB) [6]. Packet loss rate, network latency, hardware cost, control circuit complexity, and packet reordering are among the many important issues to be considered in the design, which depends on the network specification — network dimension, topology, traffic load and pattern, and so forth.

Deflection Routing

Optical buffering was, to a great extent, inspired by its conventional electronic network counterparts. In electronic networks, the link bandwidth is much less than today's optical fiber's capacity, and much effort was put into increasing link utilization. In a network deploying optical buffers, each packet is guaranteed to arrive at its destination along the shortest possible path, and for a given connectivity the expected number of hops is minimized. Implementing optical buffers involves a great amount of hardware and complex electronic controls. Another issue that arises with optical buffers is that the optical signal suffers from power loss in the delay lines, and optical amplifiers are often used. The accumulated noise from the cascaded amplifiers can severely limit the network size at very high bit rates, unless expensive signal regeneration is applied. In deflection routing, as the name implies, contention is resolved as follows: if two or more packets need to use the same output

link to achieve minimum distance routing, only one will be routed along the desired link, while others are forwarded on paths which may lead to greater than minimum distance routing. Hence, for each source–destination pair the number of hops taken by a packet is no longer fixed. Deflection routing does not necessarily exclude the use of optical buffers. The most simplification can be obtained with hot-potato routing [7], which is a special case of deflection routing where buffers are not provided at all.

Extensive studies on deflection routing have been carried out in regular network topologies with uniform traffic load. These logical topologies can be built on different physical topologies (e.g., ring, star, or mesh). Figure 11 shows the two most typical logical topologies used for network performance simulation: the Manhattan Street Network (MSN) and ShuffleNet. Each node in these two topologies has two input ports and two output ports. A node has to handle both bypassing and locally generated/terminated packets. Figure 12 shows an example of node architecture for MSN or ShuffleNet.

Studies have been conducted to determine the impact of different routing strategies on network performance, such as delays, average number of hops (i.e., the number of switches a packet has to traverse between the source and destination node) for each packet, and network aggregate capacity (the number of packets a network can process within a certain period of time). A comparison done on the ShuffleNet topology between store-and-forward and hot-potato routing shows that the average number of hops for each packet is larger for hot-potato routing, because not all the packets take the shortest route toward their destinations [7]. As the number of users (or number of nodes) increases, both the average number of hops and aggregate capacities increase for both routing strategies. In multihop networks, where information from a source node to a destination node may be routed through intermediate nodes [8], only a portion of the network capacity is used for newly generated traffic. A certain amount of network capacity is taken up by "bypassing traffic" as packets hop from one node to another to reach the destinations. The overall capacity of the network is inversely proportional to the average number of hops, and proportional to the number of nodes and the capacity of each link between two nodes. Store-and-forward routing can maximize the network capacity as the number of nodes increases. It has also been shown that even for networks containing several thousand nodes, the aggregate capacity of hot-potato routing is not worse than 25 percent of that for store-and-forward routing [9]. Another more intuitive explanation is that in hot-potato routing, the nodes use the whole network as a big buffer and route the packet in contention to the rest of the network. This type of routing trades off network throughput for simpler hardware implementation.

In optical packet-switched networks an all-optical path is provided between the source and destination without complete regeneration; therefore, at very high bit rates the propagation distance, proportional to the number of hops taken by a packet, becomes limited. The distance bit rate product is fixed if we want to keep the packet error rate (PER: the probability of a packet received in error) below a certain threshold. Given the network size and average number of hops, the PER depends on the link bit rate. The maximum bit rate therefore determines the maximum network throughput. There are three characteristics that determine the performance of the network with deflection routing:

- Diameter: The maximum distance in number of hops between any node pair in the network. The diameter is a good indicator of how compact a network is.
- Deflection Cost: ThE maximum increase in path length in number of hops due to a single deflection.
- Don't care nodes: For a given destination, any node that has both its output links as part of the shortest path is a don't care node.

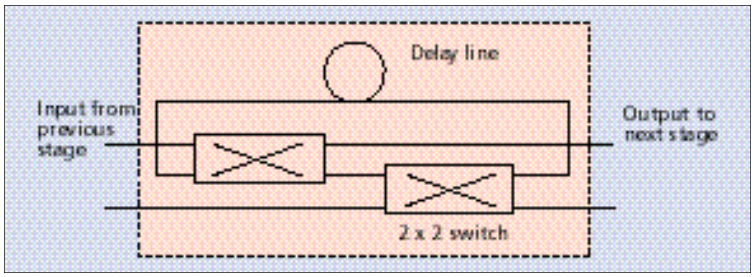

Figure 10. A 2 *x* 2 switching element containing an optical buffer.

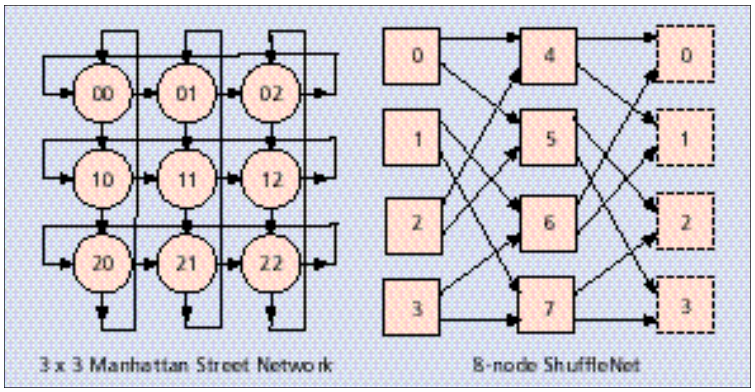

Figure 11. The Manhattan Street Network and ShuffleNet.

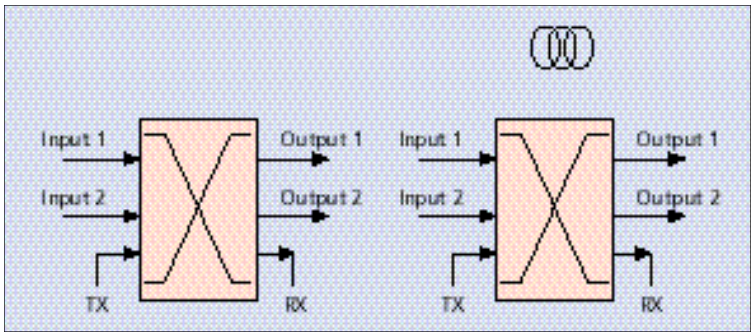

Figure 12. An example of node architecture for hot potato routing or deflection with limited buffer routing in MSN or ShuffleNet.

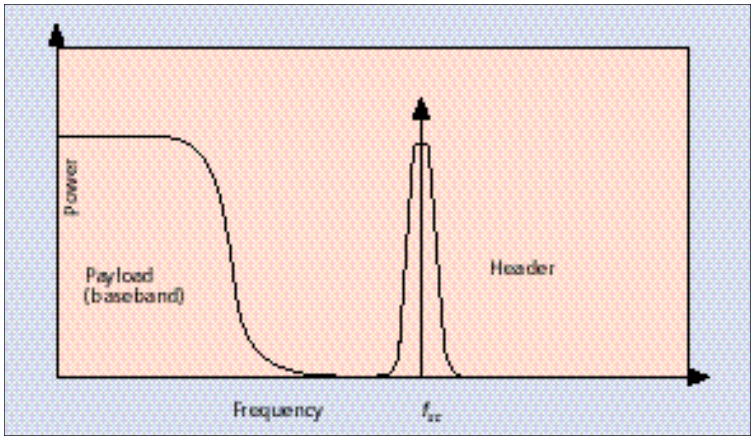

Figure 13. Qualitative power spectrum of the laser modulation current.

A high percentage of don't care nodes helps keep the number of deflections to a low level even at high loads. The performance of ShuffleNet is better because the initial advantage in diameter of ShuffleNet over MSN is preserved under heavy traffic by the high percentage of don't care nodes. The expected number of hops noticeably depends on the routing algorithm. For store-and-forward with infinite buffers the average number of hops is minimum, since the packets always take the shortest path to the destination. However, the queuing delay could diverge to infinity when the network approaches saturation (i.e., when the probability of packet generation in each time slot approaches 1). For deflection routing the average number of hops becomes an increasing function of link load, and the throughput is therefore lower than with store-and-forward.

So far, the routing strategies described have assumed slotted (synchronous) network operation, which, as shown in the previous section, involves complex and expensive packet alignment schemes. Since we are examining deflection routing here, which means no or little optical buffer is used, what will happen if we have an asynchronous network operation? What is the network performance if we take away the packet alignment stages and use deflection routing at the same time? An asynchronous network suffers from severe congestion as the offered load increases, and its throughput collapses completely when the load exceeds a certain threshold. The reason is that with increasing congestion there are more and more packets starting to "wander" around in the network (due to deflection routing), and they further lower the network capacity to process newly generated packets; meanwhile, more packets are being generated. The whole scenario forms a vicious cycle, and as a result the network throughput collapses completely. To avoid such total collapse the number of hops a packet traverses has to be monitored and kept under a maximum value.

One way to improve network throughput and eliminate congestion is to provide limited optical buffering to such asynchronous deflection networks. The corresponding performance improvement is very encouraging in the high-load region. Also, congestion is greatly reduced with more than one recirculating loop. One practical concern of asynchronous deflection routing with limited optical buffering is the number of times a packet is allowed to recirculate in the loop. Optical amplification imposes noise on the signal. Network latency also increases with the number of circulations. Therefore, it is necessary to establish an optimal maximum number of recirculation for packets.

Monitoring the number of hops a packet has taken is essential to avoid congestion, which can be caused by too many packets wandering in the network. Having a time-to-live field in the packet header (as in IP packets) is hard to implement because it requires the header to be rewritten at every node. One possible solution is to have the source node put a time stamp on each packet and the other nodes compare it with the local time when the packet is in transit, provided all the nodes have a globally synchronized clock.

Deflection routing plays a prominent role in many optical network architectures, since it can be implemented with no or modest optical buffering. Asynchronous (unslotted) deflection routing combined with limited buffering can help avoid complex synchronization schemes and provide decent performance with careful design. In general, deflection routing presents more choices to the network designer, while many problems, such as packet reordering and the impact of deflection degree, remain to be more thoroughly studied.

Wavelength Conversion

Optical buffering and deflection routing could be regarded as deflection in general, one in the time domain and the other in the space domain. With today's enabling technology in WDM, the wavelength domain presents one more dimension of solution. Both buffering and deflection have their advantages and disadvantages: buffering offers better network throughput but involves more hardware and controls; deflection is easier to implement, but cannot offer ideal network performance. When combined with wavelength conversion, their disadvantages could be overcome or minimized, therefore giving the network designer more choice and flexibility. In this section we will examine some interesting combinations.

In a switch node applying wavelength conversion and buffering, the input stage demultiplexes wavelength channels and the wavelength converters locate available wavelengths for certain output ports. The nonblocking space switch selects the output port or appropriate delay line. Here the buffer may consist of a series of delay lines with different lengths.

Wavelength conversion combined with optical buffering can also be incorporated in an asynchronous network. An example of an optical packet switch block without packet alignment is described in [10]. It is very similar to the broadcast-and-select architecture proposed in the KEOPS project, except that it is expanded for WDM operation with wavelength conversion. Since the switch is made of optical gates, it is fully nonblocking and can be configured incrementally, making the architecture ready for asynchronous operation.

Wavelength conversion has been shown to reduce the number of optical buffers or reduce packet loss probability. When the nodes are provided with a number of optical receivers/transmitters equal to the number of wavelengths, hot-potato routing in conjunction with wavelength conversion becomes an interesting option for mesh topologies such as MSN and ShuffleNet [11].

It has been shown that delay lines are more effective in solving contention than wavelength conversion. However, wavelength conversion provides noise suppression and signal reshaping. Therefore, whether to use wavelength conversion or not depends on the specific network. In a network with only a small number of wavelengths, buffering might be more desirable. In a network with a large number of wavelengths and full wavelength conversion, buffers may not be necessary.

There are several possible combinations of optical buffering and wavelength conversion with store-and-forward or deflection routing. It presents an open research problem to decide which scheme offers low implementation cost, low packet delay, low packet loss ratio, high network throughput, and so on, depending on the network specifications.

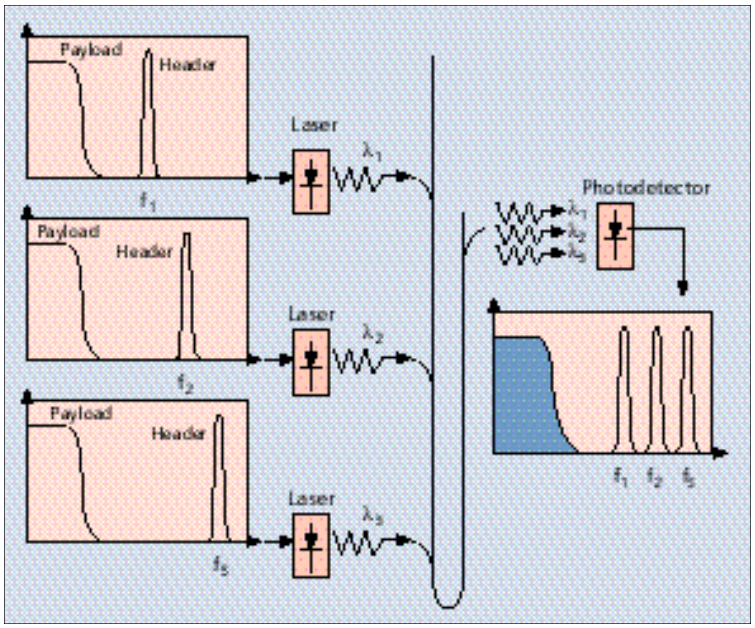

Figure 14. Header retrieval in SCM.

HEADER AND PACKET FORMAT

In electronic networks, the packet header is transmitted serially with the payload data at the same data rate (e.g., IP packets and ATM cells). Electronic routers or switches will process the header information at the same data rate as the payload. In an optical network, the bandwidth is much larger than their electronic counterparts. A typical wavelength channel has a line speed of 2.5 Gb/s (OC 48). Although there are various techniques to detect and recognize packet headers at gigabit-per-second speed, either electronically or optically, it is still difficult to implement electronic header processors operating at such high speed to switch packets on the fly at every node.

Among several different proposed solutions, packet switching with subcarrier multiplexed (SCM) headers is attracting increasing interest. In this approach the header and payload data are multiplexed on the same wavelength (optical carrier). In the current that modulates the laser transmitter, payload data is encoded at the baseband, while header bits are encoded on a properly chosen subcarrier frequency at a lower bit rate, as shown in Fig. 13. The header information on different wavelengths can be retrieved by detecting a small fraction of the light in the fiber with just a conventional photodetector, without any type of optical filtering. In the output current of the photodetector various data streams from different wavelengths jam at baseband, but the subcarrier remains distinct, and the header can be retrieved by electrically filtering out the desired subcarrier (Fig. 14).

Since the laser and photodetector electrical frequency response must extend as far as the highest subcarrier frequency, it is important to keep the subcarrier frequencies as few, low, and closely spaced as possible. Since the minimum subcarrier spacing cannot be less than twice the header bit rate, it is also important to keep the header bit rate low. On the other hand, if the header bit rate is too low, it takes longer to transmit and receive the header information, thus causing longer delays for the payload data.

A nice feature of the SCM header is that it can be transmitted on top of the payload data and can take up the whole payload transmission time, since it does not interfere with the payload. Of course, the header can also be transmitted serially with the payload if so desired. One potential pitfall of the SCM header is its possible limit on the payload data rate. If the payload data rate is increased, the baseband will expand and might eventually overlap with the subcarrier frequency, which is limited by the microwave electronics.

In many of the routing and switching protocols, packet headers have to be updated at each node. There have been several approaches proposed on all-optical same-wavelength header replacement for headers transmitted serially with the payload data stream. All-optical header replacement could be done by blocking the old header with a fast optical switch and inserting the new header, generated locally by another laser, at the proper time. One important issue here is that in WDM networks the new header should be precisely at the same wavelength as the payload data; otherwise, serious problems could arise because of dispersion, nonlinearity, or wavelength-sensitive devices in the network.

It has been proposed that the header updating be done by transmitting the payload and header on separate wavelengths, and demultiplexing the header for optoelectronic conversion, electronic processing, and retransmission on the header wavelength. This approach suffers from fiber dispersion, which separates the header and payload as the packet propagates through the network. SCM headers have far fewer dispersion problems

In the future, optical tag switching, micro-electro-mechanical systems, photonic slot routing, and optical burst switching, among others, will likely play an important part in the architecture and system of photonic packet switched networks.

since they are very close to the baseband frequency. An SCM header could be removed by narrowband optical filters, but would be extremely sensitive to wavelength drift. Previous practical SCM header replacement schemes were limited to full optoelectronic conversion of the entire packet followed by electronic filtering, remodulation, and retransmission on a new laser. Reference [12] proposed a technique to update the SCM header with simultaneous wavelength conversion of baseband payload using SOAs. It involves a two-stage process. First, simultaneous SCM header suppression and wavelength conversion of the baseband payload are achieved due to the low-pass frequency response of cross-gain modulation in the SOAs; then header replacement is achieved by optically remodulating the wavelength converted signal with a new header at the original subcarrier frequency.

Packet length is another issue of concern for network designers. A short packet might not give good throughput because a greater percentage of the bandwidth is given to the header or guard time between time slots. On the other hand, a long packet would need longer optical buffers, and not provide a granularity that is fine enough. From a physical point of view, balancing the PER between payload and header is very important. PER is different from bit error rate (BER); it is the probability of the entire packet being received in error. PER increases with BER and the number of bits contained in the packet. For efficient network operation, the PER for payload and header should be about the same, in order to deliver the packets as successfully as possible [13]. Payload usually contains many more bits than the header. If the header is updated at every traversed node, the bits in the payload will have to suffer more physical impairment than the bits in the header. Another fact is that if SCM is used, the header is usually transmitted at a lower bit rate than the payload data. All these facts lead to a big advantage of lower BER for header bits over the payload bits. Therefore, it is imperative to optimize the amount of power to be tapped from the packet at each node and the packet length in order to achieve a balanced PER for payload and header at the destination node.

Looking into the Future

During the development of optical packet-switched networks, the most prominent and early project was the ACTS KEOPS Project, which addressed the analysis and demonstration of optical transparent packet switching within all-optical network architectures by means of network and system studies, and laboratory demonstrations based on components developed in the project. Since the KEOPS node architecture uses wavelength conversion to achieve switching, it can apply optical buffering or optical buffering plus wavelength conversion for contention resolution. WASPNET [14] is another research collaboration between three British universities. The project involves determining the management, system, and device ramifications of an optical packet network applying wavelength conversion plus buffering for contention resolution. In addition to the above projects, there are several other projects ongoing across the globe. In the future, optical tag switching, micro-electro-mechanical systems (MEMS), photonic slot routing, and optical burst switching, among others, will likely play an important part in the architecture and system of photonic packet switched networks.

Optical Tag Switching

Fast-growing Internet traffic is playing a major role in today's telecommunications infrastructure. The current Internet Protocol requires a complicated IP header to be processed on a hop-by-hop basis. This involves hundreds of lines of software processing, which could impose a bottleneck in the future as fiber link speeds approach terabits per second. Tag switching, as an alternative approach, has been proposed to simplify the packet forwarding process. It assigns a short fixed-length label containing routing information, a so-called tag, to multiprotocol (i.e., IP, ATM, frame relay, etc.) packets for transport across interconnected subnetworks.

A tag switched network consists of:

- Tag edge routers, which are located at the boundaries of the Internet and apply tags to packets
- Tag switches, which switch tagged packets based on the tags
- Tag distribution protocol, which is used to distribute tag information between nodes

The tag switches use the routing table generated by routing protocols to assign and distribute tag information via the tag distribution protocol, while they also receive tag information and build a forwarding table for local switching. When a tag edge router receives a packet for forwarding across the network, it analyzes the network layer header, performs applicable network layer services, selects a route for the packet, and applies a tag to the packet. Then it forwards the packet to the next-hop tag switch. The tag switch receives the tagged packet and switches the packet based on the tag, without reanalyzing the network layer header. The packet reaches the tag edge router at the exit point of the network, where the tag is removed and the packet delivered (Fig. 15).

MEMS Optical Switches

In this article we have not discussed much about the core of the packet switch, the switch fabric. There have been numerous schemes proposed to construct an N x M optical switch in the past few years. The fabrication of even a small switch unit, such as a 1 x 2 or 2 x 2 switch block, involves many physical issues. It is not the goal of this article to present a detailed discussion on switch fabrics; however, we would like to mention an emerging technology which may potentially revolutionize the switch fabrication industry.

Conventional mechanical switches suffer from large size, large element mass, and slow switching time. Guided-wave solid-state switches impose limited cascadability, high crosstalk, and large size. Meanwhile, micro-electro-mechanical-systems (MEMS) technology is beginning to impact many areas of science and industry. It has shown a bright future of achieving high-quality and

high-port-count optical switching. MEMS devices are built in a similar manner to silicon integrated circuits. Various layers of different materials are deposited and patterned to produce complicated, multilayer, three-dimensional structures. At the end of the fabrication process, selective etching removes some of the deposited materials and creates movable parts for the device. Most MEMS switches make use of movable torsion mirrors to redirect the propagation direction of light and achieve the switching functionality. They can provide low loss and low crosstalk while remaining compact in size and providing good economy due to monolithic batch production.

Photonic Slot Routing

A WDM network offers wavelength conversion as one more dimension of switching. On the other hand, taking advantage of this feature requires fast control and wavelength-selective devices, which can dramatically increase the network cost. Photonic slot routing (PSR) was proposed as an alternative to using WDM only as a way to multiply network capacity, thus reducing node complexity and cost, and facilitating network scalability [15]. According to this concept, packets transmitted in the same time slot (photonic slot) on all wavelengths are switched jointly. The switching node is only required to handle each slot as a whole, without having to access and switch packets on different wavelengths individually. At each node, packets destined for a specific node are transmitted on the available wavelength in the slots assigned to that particular node. If a slot is not assigned, it can be assigned by the first packet transmitted in that slot under a certain fairness control protocol. Contention can be resolved using switched delay lines. The PSR approach shifts the burden of wavelength-selective switching to a problem of finding effective access protocols at the source nodes.

Optical Burst Switching

Optical burst switching (OBS) was proposed as another way of implementing packet switching optically to avoid potential electronic bottlenecks. The basic unit of data to be transmitted is a burst, which consists of multiple packets. The data burst is sent after a control packet reserves necessary resources on the intermediate nodes without waiting for acknowledgment from the destination node (as in the virtual circuit setup process in ATM). OBS could achieve high bandwidth utilization with lower average processing and synchronization overhead than pure packet switching since it does not require packet-by-packet operation. It is also possible to implement quality of service (QoS) by manipulating the offset time between the control packet and the data burst [16, 17].

Conclusions

It is impossible to cover every aspect of optical packet switching in one article. This topic involves routing, synchronization, contention resolution, header format/updating, switch fabrics, physical impairment of the devices, network control, protocol, and so on. Only a handful of the important aspects were covered in this article.

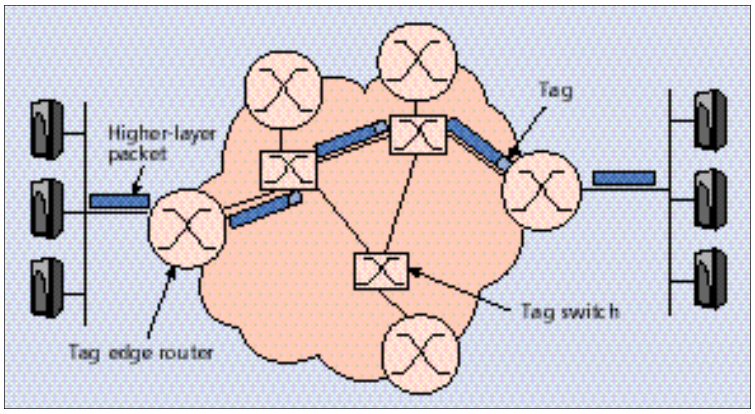

Figure 15. A tag switched network.

Optical packet switching is promising because it offers much higher capacity and data transparency. Progress has been made in several areas, but not all. Meanwhile, there is a tremendous increase in the processing speeds and capacity of electronic switches and routers. It is important for network designers to reduce the number of protocol layers being used in today's networks, while preserving the functionality and making use of the current optical technology.

References

[1] B. Bostica *et al.*, "Synchronization Issues in Opitcal Packet Switched Networks," *Photonic Networks*, G. Prati, Ed., London: Springer Verlag, 1997, pp. 362–76.
[2] M. J. Karol, "A Shared-Memory Optical Packet (ATM) Switch," *Proc. 6th IEEE Wksp. Local and Metro. Area Networks*, 1993, pp. 205–11.
[3] R. L. Cruz and J. T. Tsai, "COD: Alternative Architectures for High-Speed Packet Switching," *IEEE/ACM Trans. Net.*, vol. 4, Feb. 1996, pp. 11–21.
[4] Z. Haas, "The 'Staggering Switch': An Electronically Controlled Optical Packet Switch," *J. Lightwave Tech.*, vol. 11, no. 5/6, May/June 1993, pp. 925–36.
[5] I. Chlamtac *et al.*, "CORD: Contention Resolution by Delay Lines," *IEEE JSAC*, vol. 14, June 1996, pp. 1014–29.
[6] D. K. Hunter *et al.*, "SLOB: A Switch with Large Optical Buffers for Packet Switching," *J. Lightwave Tech.*, vol. 16, Oct. 1998, pp. 1725–36.
[7] A. S. Acampora and I. A. Shah, "Multihop Lightwave Networks: A Comparison of Store-and-Forward and Hot-Potato Routing," *IEEE Trans. Commun.*, vol. 40, June 1992, pp. 1082–90.
[8] B. Mukherjee, *Optical Communication Networks*, McGraw-Hill, 1997, pp. 87.
[9] 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, Jan. 1995, pp. 88-98.
[10] P. B. Hansen *et al.*, "Optical Packet Switching without Packet Alignment," *Proc. ECOC '98*, Madrid, Spain, Sept. 1998, paper WdD13.
[11] A. Bononi, G. A. Castañón, and O. K. Tonguz, "Analysis of Hot-Potato Optical Networks with Wavelength Conversion," *J. Lightwave Tech.*, vol. 17, Apr. 1999, pp. 525–34.
[12] M. D. Vaughn and D. J. Blumenthal, "All-Optical Updating of Subcarrier Encoded Packet Headers with Simultaneous Wavelength Conversion of Baseband Payload in Semiconductor Optical Amplifiers," *IEEE Photon. Tech. Lett.*, vol. 9, 1997, pp. 827–29.
[13] D. Datta *et al.*, "BER-Based Call Admission in Wavelength-Routed Optical Networks," *OFC '98 Tech. Dig.*, 1998, pp. 92–93.
[14] D. K. Hunter *et al.*, "WASPNET: A Wavelength Switched Packet Network," *IEEE Commun. Mag.*, Mar. 1999, pp. 120–29.
[15] I. Chlamtac *et al.*, "Scalable WDM Network Architecture Based on Photonic Slot Routing and Switched Delay Lines," *Proc. IEEE INFOCOM '97*, Kobe, Japan, Apr. 1997, pp. 7–11.

It is important for network designers to reduce the number of protocol layers being used in today's networks, while preserving the functionality and making use of the current optical technology.

[16] M. Yoo and C. Qiao, "Just-Enough-Time(JET): A High Speed Protocol for Bursty Traffic in Optical Networks," *Proc. IEEE/LEOS Tech. for a Global Info. Infrastructure*, Aug. 1997, pp. 26–27.
[17] Y. Xiong, M. Vandenhoute, and H. Cankaya, "Design and Analysis of Optical Burst-Switched Networks," *Proc. SPIE '99 Conf. All Opt. Networking: Architecture, Cont. and Mgmt. Issues*, Boston, MA, vol. 3843, Sept. 1999, pp. 112–19.

BIOGRAPHIES

SHUN YAO (shyao@ece.ucdavis.edu) received his B.E. in electronic engineering in 1997 from Tshinhua University, China. Before joining the Ph.D. program at the University of California-Davis, he studied in the optoelectronics program at the University of Wisconsin-Madison from 1997 to 1999. He is currently working closely with Nokia Research Center Boston on optical packet-switched networks. His research interests include all-optical networks, and network control and management issues.

SUDHIR DIXIT [SM] (sudhir.dixit@nokia.com) received a B.E. degree from Maulana Azad College of Technology (MACT), Bhopal, India, an M.E. degree from Birla Institute of Technology and Science (BITS), Pilani, India, and a Ph.D. degree from the University of Strathclyde, Glasgow, Scotland, all in electrical engineering. He also received an M.B.A. degree from Florida Institute of Technology, Melbourne. He currently heads research in broadband networks at Nokia Research Center, Boston, Massachusetts, specializing in ATM, Internet, all-optical networks, and third-generation mobile networks. From 1991 to 1996 he was a broadband network architect at NYNEX Science and Technology (now Bell Atlantic). Prior to that he held various engineering and management positions at other major companies, such as GTE, Motorola, Wang, Harris, and STL (now Nortel Europe). He has published extensively, and has 13 patents either granted or pending. He has been either conference chair, session chair, and/or on the program committees of several conferences. As an ATM Forum Ambassador, he has presented tutorials on ATM internationally. He was a guest editor for a special issue of *IEEE Network* on digital video dial-tone networks, published in October/November 1995, and a guest editor for a feature topic on service and network interworking in a WAN environment published in *IEEE Communications Magazine* in June 1996. He is a Guest Editor of a special issue of *IEEE Communications Magazine*, "WDM Networking: A Reality Check," to be published in March 2000. He is listed in several national and international Who's Who publications. He is an editor and lightwave series editor of *IEEE Communications Magazine*.

BISWANATH MUKHERJEE (mukherjee@ece.ucdavis.edu) received a B.Tech. (Hons) degree from Indian Institute of Technology, Kharagpur, in 1980, and a Ph.D. degree from the University of Washington, Seattle, in June 1987. At Washington he held a GTE Teaching Fellowship and a General Electric Foundation Fellowship. In July 1987 he joined the University of California-Davis, where he has been professor of computer science since July 1995, and chairman of computer science since September 1997. He is co-winner of paper awards presented at the 1991 and 1994 National Computer Security Conferences. He serves on the editorial boards of *IEEE/ACM Transactions on Networking*, *IEEE Network*, *ACM/Baltzer Wireless Information Networks* (WINET), *Journal of High-Speed Networks*, *Photonic Network Communications*, and *Optical Network Magazine*. He also serves as Editor-at-Large for optical networking and communications for the IEEE Communications Society. He served as Technical Program Chair of IEEE INFOCOM '96. He is author of the textbook *Optical Communication Networks* (McGraw-Hill, 1997), a book which received the Association of American Publishers, Inc.'s 1997 Honorable Mention in Computer Science. His research interests include lightwave networks, network security, and wireless networks.