Architectural and Technological Issues for Future Optical Internet Networks

Marco Listanti, Vincenzo Eramo, University of Rome "La Sapienza"
Roberto Sabella, Ericsson Lab Italy

ABSTRACT

This article reports a review of the most significant issues related to network architectures and technologies which will enable the realization of future optical Internet networks. The design of such networks has to take into consideration the peculiar characteristics of Internet traffic. Several architectures have been proposed to provide optical networking solutions, based on wavelength-division multiplexing and compatible with the IP world. These architectures are presented briefly, and the main advantages and drawbacks are discussed. Furthermore, advanced network architectures are reported. In particular, two network paradigms are illustrated and discussed: the optical transparent packet network and optical burst switching. Finally, the key technologies are illustrated.

INTRODUCTION

The telecommunications world is evolving dramatically toward challenging scenarios. The convergence of the *telecom* and *datacom* worlds into the *infocom* era is becoming a reality.

New competitive companies are offering lower prices and driving the introduction of new services. New technologies are emerging, government regulations are being relaxed, and the industry is rapidly globalizing. The efficient transport of information is becoming a key element in today's society.

By some estimates, bandwidth usage of the Internet is doubling every six to 12 months. For the first time, data network capacities are surpassing voice network capacities, and the growing demand for network bandwidth is expected to continue in the coming years. Current networks use only a small fraction of the available bandwidth of fiber optic transmission links. The emergence of wavelength-division multiplexing (WDM) technology is now unlocking more of the available bandwidth, leading to lower costs, which can be expected to further fuel the demand for bandwidth.

We now face the near-term prospect of single fibers capable of carrying hundreds of gigabits per second of data. Single optical fibers have the potential for carrying as much as 10 Tb/s. This leads to a serious mismatch with current switching technologies which are not yet capable of switching these high aggregate rates. Emerging asynchronous transfer mode (ATM) switches and IP routers are able to switch data using the individual channels within a WDM link (the channels typically operate at 2.4 or 10 Gb/s), and this implies that tens or hundreds of switch interfaces must be used to terminate a single link with a large number of channels.

Moreover, there can be a significant loss of statistical multiplexing efficiency when parallel channels are used simply as a collection of independent links, rather than as a shared resource. Proponents of optical switching have long advocated new approaches using optical technology in place of electronics in switching systems [1]. Unfortunately, the limitations of optical component technology [1] have largely limited optical switching to facility management applications. While there have been attempts to demonstrate the use of optical switching to directly handle end-to-end user data channels, these experiments have been disappointing. Indeed, they have primarily served to show how crude optical components remain and have done little to stimulate any serious move toward optical switching.

The aim of this article is to provide an overview of the main architectural and technological issues related to optical networking solutions which support the future infocom scenario.

We will review the main characteristics of Internet traffic, which demand a different approach to network design. We will provide a survey of the architectures and technologies for realizing "wavelength-routing-based" optical networks, and introduce some advanced solutions proposed in the literature to better match the requirements of Internet traffic with the facilities offered by optical technology. In particular, two paradigms are discussed: the optical transparent packet network and optical burst switching. The key WDM optical devices' state of maturity is discussed, while the last section derives some concluding remarks and the perspectives envisaged by the authors.

0163-6804/00/$10.00 © 2000 IEEE

Internet Traffic Characteristics

Since Internet traffic will more and more dominate traditional telecom traffic, the understanding of its characteristics is crucial for a reasonable design of the future optical infocom network. Specifically, three main issues are worth highlighting:

- The self-similar nature of Internet traffic
- Routing and data flow asymmetry
- Server-bound congestion

The Self-Similar Nature of Internet Traffic

The self-similar nature of Internet traffic has been demonstrated by several measurements and statistical studies [2]. In particular, it means that traffic on Internet networks exhibits the same characteristics regardless of the number of simultaneous sessions on a given physical link. As an example, Fig. 1 shows self-similar traffic vs. Poisson voice traffic for different numbers of aggregated users. It can easily be seen that as the number of voice flows increases, the traffic becomes more and more smoothed. In other words, the variance of voice traffic (which can be modeled as a Poisson process) rapidly decreases with the increase of flow aggregation. On the other hand, this does not happen with Internet traffic. In fact, the variance of this process decreases with much lower speed. This property is usually known as the *long-range dependence* (LRD) of Internet traffic.

The self-similar nature of Internet traffic has a direct impact on network dimensioning. In particular, buffer sizing is crucial. On one hand, the buffer size should be great enough to absorb very long traffic bursts induced by self-similar characteristics; on the other hand, the size should not be so large as to introduce unacceptable delays. A possible solution for a network designer is to increase the buffer size at admission points into the network in order to smooth out the peaks and valleys; and to dimension the IP link capacities so that the IP network can operate at a higher peak-to-average load than a traditional telecom network.

Routing and Data Flows Asymmetry

The phenomenon of IP data flow asymmetry has been much observed on both national and international links [3]. Such asymmetry is attributed to larger server farms sending out large amounts of data in response to small requests and to the preponderance of users who download Web pages. Web server farms tend to be clustered near large Internet service points, while users are randomly distributed around the edges of the network. Consequently, near large interconnection points where Web servers are located, there is a large asymmetry in transmit/receive (Tx/Rx) data flows in favor of the Tx path exiting the servers. A sketch of the traffic flows in an Internet network is shown in Fig. 2.

The main consequence of asymmetry is that a considerable amount of Internet bandwidth — sometimes close to 50 percent — is idle, and at the same time the bandwidth on the other side of a Tx/Rx is totally congested. It is necessary to highlight that such a condition exists because contemporary telecom systems are still designed to primarily support voice traffic.

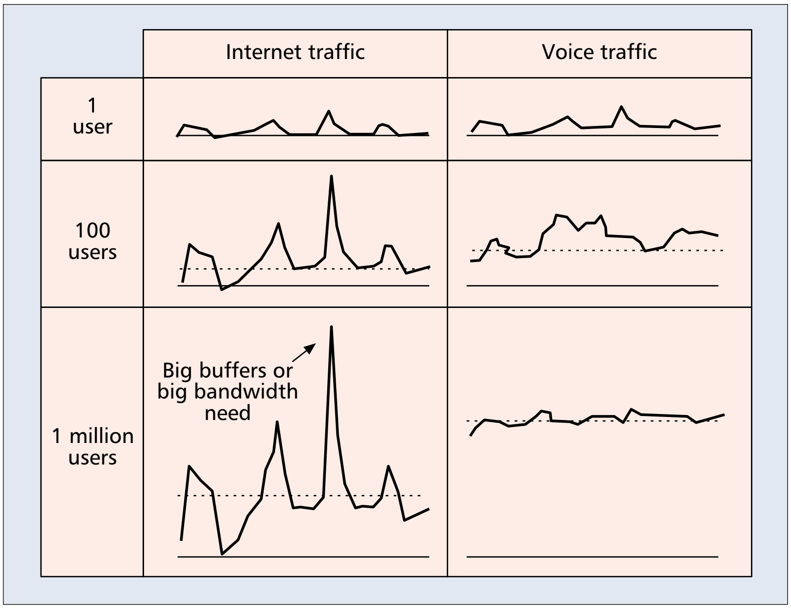

Figure 1. *The self-similar nature of Internet network traffic.*

Server-Bound Congestion

It is widely experienced that traffic flows on the Internet are limited by the servers providing data to requests from users, rather than by the network itself. In the presence of large bandwidth it is increasingly likely that the server flow control window will be the dominant control element in traffic throughput rather than today's congestion window. As a result, with larger pipes Internet throughput will be increasingly server-bound, even beyond the 52 percent server-bound congestion experienced today [3]. Certainly, servers will

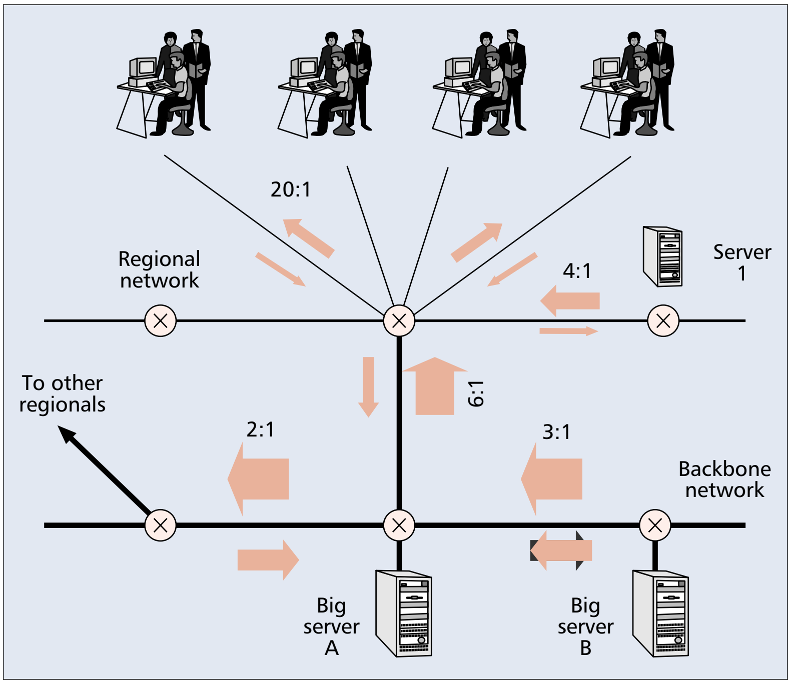

Figure 2. *Asymmetry of the data flows in an Internet network.*

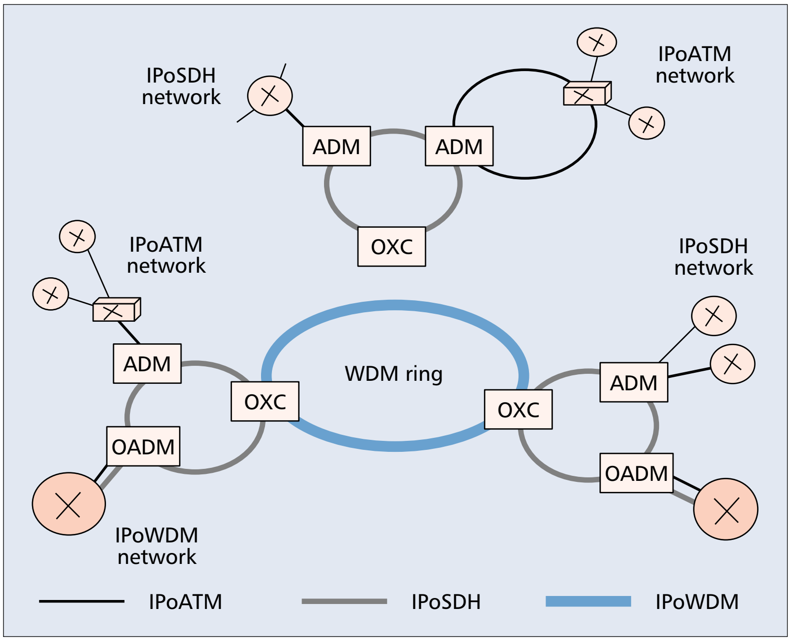

Figure 3. *A relevant example of an optical Internet network. In the picture, different protocol stacks are integrated to provide different-size bandwidth pipes and classes of services.*

increase their flow control window size as they increase in CPU power. However, the overall network capacity increases faster than the average CPU performance of most servers. Ultimately, the question of server-bound vs. network-bound will depend on the relative growth of bandwidth vs. CPU power. If bandwidth growth, principally due to the deployment of WDM, is faster than Moore's law for CPU power and capacity, then ultimately server congestion will be the controlling element in future networks [3].

Wavelength Routing Architectures

WDM networking has been launched by the concept of *wavelength routing*. The principle is that high-speed data flows, which consist of many time-division multiplexed channels, are associated with specific optical wavelengths. Thus, they are routed through the optical network by means of their wavelengths, without necessarily being opto-electronically converted, demultiplexed, and electronically routed. This concept allows the realization of all-optical routers, which can handle many WDM channels simultaneously, without the need for very high-speed electronics. The lower hierarchies are naturally processed by an electronic cross-connect that possibly interoperates with the optical cross-connect (OXC). Thus, wavelength routing consents the realization of optical add-drop multiplexers (OADMs) and OXCs working in a semi-permanent way. The capability to aggregately handle optical bandwidth provides the means to cope with the Internet traffic characteristics previously mentioned.

The main feature of this kind of optical network lies in the possibility of performing these operations directly in the optical domain without requiring costly high-speed electronic equipment, and in its transparency, that is, the possibility of making those functions independent, to some extent, of the signal format. Actually, it is possible to define several degrees of transparency. In fact, absolute transparency is the property of a network in which any signal travels along the network independent of its transmission format, speed (bit rate in case of pulse code modulated, PCM, signals), and so on; that is, only terminal equipment would determine the limitation on such a signal format. However, due to physical limitations of fiber propagation and the physical nature of optical devices traversed by the signal, absolute transparency can never be reached. Thus, it is more useful to specify a certain level of transparency. The simplest degree of transparency is in digital signals (independence of bit rate, format, and protocol). Furthermore, it is possible to define transparency to intensity-modulated signals (both analog and digital). Full transparency would require that a network be transparent to any optical signal, regardless of its amplitude, phase, or frequency modulation.

In practical networks, transparency will allow the handling of different types of data flows simultaneously. In fact, wavelengths can carry either synchronous digital hierarchy/optical network (SDH/SONET) streams, ATM streams, or other possible transport formats. If the degree of transparency is high enough, as in a municipal area network, even analog signals such as analog video signals can be carried without any conversion.

The main issue when designing optical networks for Internet application is the right mode of transport for IP packets. Actually, several transport options have been proposed in the literature, such as IP over ATM over WDM and IP over SDH/SONET over WDM; and recently, a lot of literature has proposed IP over WDM.

The rate of change of technology also impacts the selection of core network technology. For instance, today's implementation of IP networks makes use of different transport techniques, embodying IP, frame relay, ATM, SDH/SONET, and WDM. If one minimizes network elements (e.g., IPoWDM), it may be more cost effective, but there is more risk of obsolescence of investments. However, such a risk may be small if IP still remains the predominant traffic type.

A conceivable infocom network is sketched in Fig. 3. It supports different services and a variety of transport protocols. For instance, IPoWDM could be used for high-volume best-effort computer-to-computer traffic, while IP over ATM could be used to support virtual private networks and mission-critical IP networks. IP over SDH/SONET could instead be used to aggregate and deliver traditional IP network services. Each data flow is carried by a dedicated wavelength.

In any case, OADMs and OXCs represent the key elements of optical networking.

In the following sections, the optical node architectures and the different solutions for transporting IP packets on optical links are briefly described.

Optical Node Architectures

A general scheme of an OADM is depicted in Fig. 4. It can selectively drop or add a specific wavelength on a WDM comb carried by a fiber. The other wavelengths are passed through the node optically. This optical node is characterized by several functionalities. For example, it could be rigid, in the sense of adding/dropping one or more fixed wavelengths; or flexible, adding/dropping any one or more of the wavelengths at its disposal. Technological details on OADM architecture and technologies can be found in [4].

An OXC provides the possibility of routing individual channels coming from any of its input ports to any output port. There are several architectures, depending on whether the OXC is rigid, rearrangeable, or strictly nonblocking. The basic schemes are shown in Fig. 5a–d [4].

The simplest configuration (Fig. 5a) does not give any possibility of rearrangement. A rearrangeable OXC is depicted in Fig. 5b, where a space-division switching function has been introduced by using space switching matrices. Here each wavelength of any input fiber can be routed to any output fiber not already using that wavelength. This OXC presents a bandwidth proportional to N x M x B, where N is the number of input/output fiber ports, M the number of wavelengths carried by each fiber, and B the bit rate per wavelength.

The constraint that two channels carried on two different fibers at the same wavelength cannot be routed simultaneously onto a single outgoing link can be accepted or not, depending on network topology, dimensions, traffic, operations, administration, and maintenance (OAM) functions, and so forth. However, such a constraint can be eliminated by using wavelength translators in conjunction with a large switch inside the optical node, as shown in Fig. 5c. This configuration adds significant complexity to the routing node structure, but permits better wavelength reuse. To avoid large space-division switches, which are impractical, especially for

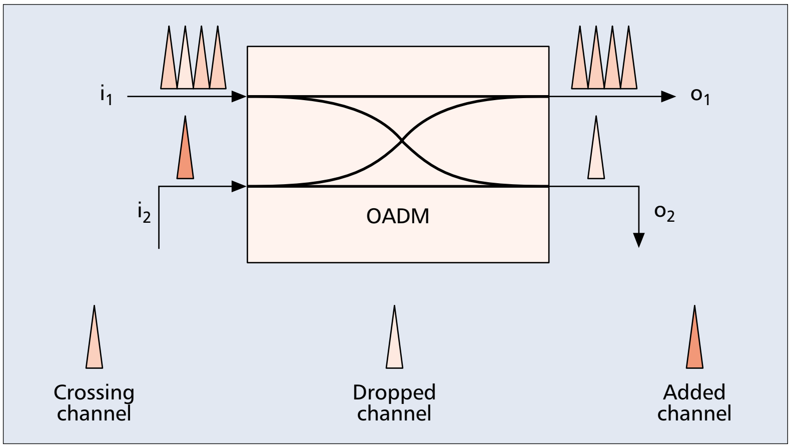

Figure 4. *A generic scheme of an optical add-drop multiplexer.*

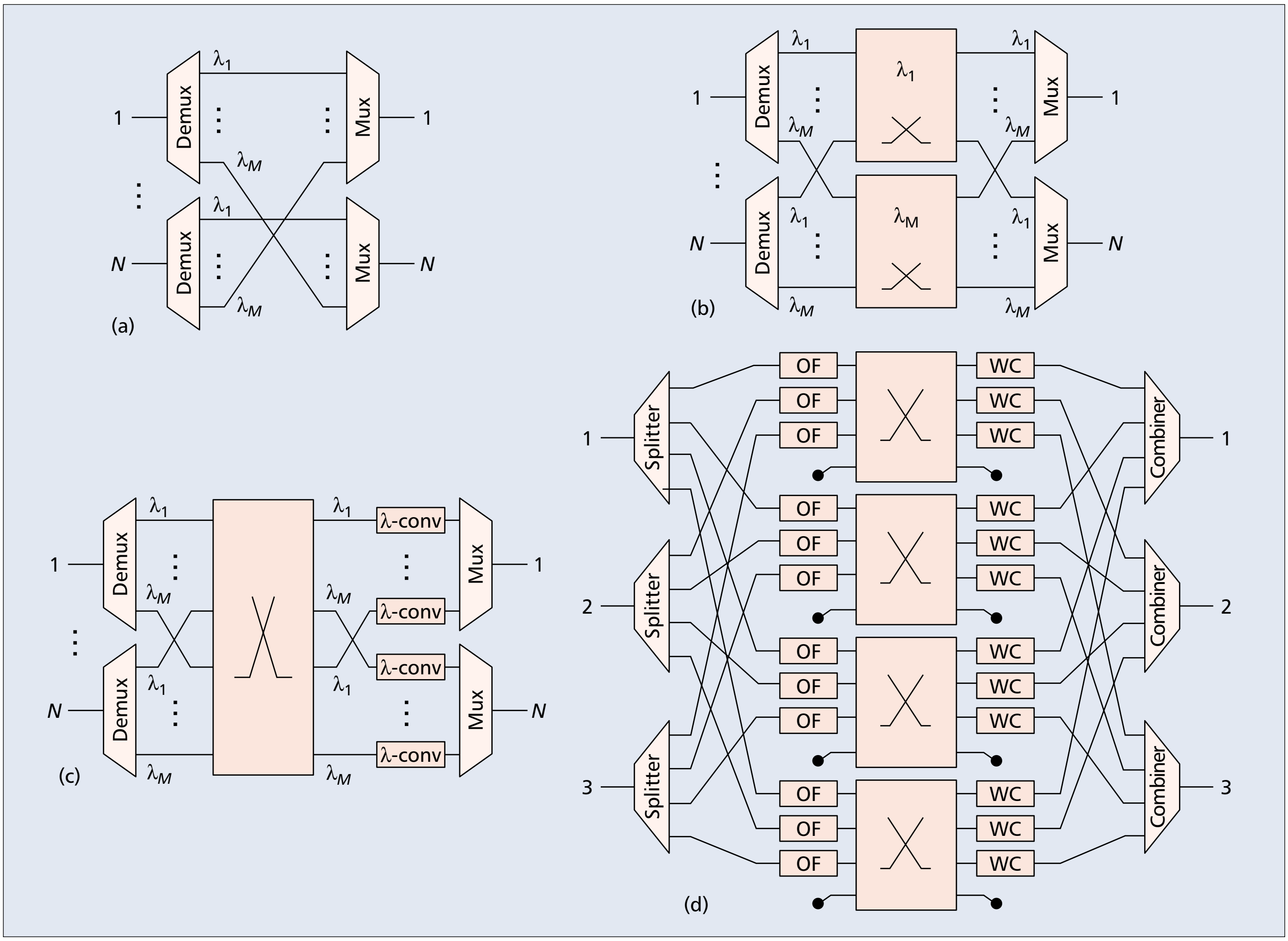

Figure 5. *Basic schemes of optical cross-connect architecture: (a) rigid; (b) rearrangeable; (c) strictly nonblocking; (d) strictly nonblocking with lower-dimension space switches.*

large dimensions (N x M) of the OXC, the architecture shown in Fig. 5d can be adopted, allowing the highest flexibility to be obtained. In this architecture channel selection is accomplished by a combination of passive power splitters and tunable filters. Several low-dimension switch matrices thus substitute for the large switch. Technological issues related to OXCs and several examples of OXC schemes are reported in [4].

The IPoATMoWDM Solution

The ATM networking solution is attractive because it makes it possible to aggregate different traffic types onto the same pipe; thereby it produces significant savings in overall bandwidth over managing services on different networks. In fact, ATM networks are optimized to carry a mix of different service types rather than one specific service type. The other advantage of ATM in this context is that it should be relatively easy to support virtual private networks (VPNs) and classes of service for data. However, if the current Internet trends continue, Internet data will be the predominant service type. Thus, it makes sense to build up a network optimized for delivery of Internet data. The remaining services can then be delivered on top of an IP network or, alternatively, still be delivered over a parallel ATM network.

Actually, ATM networks provide an incredible degree of flexibility in terms of network engineering and design, but this flexibility comes at a cost in terms of complexity. Running IP-over-ATM networks is generally much more complex to manage than traditional IP leased line networks. While ATM provides a powerful set of capabilities in terms of traffic engineering, existing IP routing protocols have limited traffic engineering capabilities in terms of directing traffic across specific links, mainly because the routing metrics are based on the number of routing hops. A promising technique is multiprotocol label switching (MPLS). In fact, MPLS can provide this same traffic engineering capabilities at the IP layer, but MPLS may end up introducing the same level of complexity as currently exists with ATM networks. Another interesting advantage of ATM lies in the possibility to realize fast ATM switches. However, several companies are proposing IP routers with very high throughput (terabit routers).

As a result, since ATM does not offer inherent improvements in throughput over the new class of terabit routers but brings about more complexity, the solution of IP over ATM over WDM makes little sense, particularly in large backbone networks. If the predominant traffic is IP, the ATM network is an added level of complexity that is costly to network providers in terms of management.

The IPoSDH/SONEToWDM Solution

One of the main advantages of SDH/SONET is its restoration capability in the event of a fiber cut or failure in an SDH/SONET node. An SDH/SONET ring network can switch to an alternate fiber or path on the other side of a fiber ring in the event of a fiber cut in less than 50 ms. This restoration property is transparent to the IP network layer.

In a backbone for Internet data, this sophisticated link management in the SDH/SONET layer may not be necessary. Protection and restoration capabilities are part of the Internet's intrinsic distributed survivability characteristic.

One of the big advantages of having routers directly connected to WDM wavelengths is that the router can use wavelengths on both sides of a fiber ring and load share IP traffic, and possibly double the bandwidth utilization of any Internet link at very low incremental cost to the carrier. In the event of a fiber cut it might be possible to throttle back best-effort IP traffic to be routed over the single surviving fiber, or alternatively rerouted over a completely different path to the destination. Since the nature of Internet data is self-similar, the consequences of a fiber cut are less severe in the data networking environment than in the traditional telecom environment. The loss of a fiber may also be compensated for by well-known techniques for flow control, buffering, and rerouting. In addition, in an optical Internet network the router can establish asymmetric transmit/receive wavelengths to balance ingress and egress traffic across the network. SDH/SONET networks are always realized under the main assumption of transmitting and receiving traffic always being in balance, and as such cannot be optimized for asymmetric transmit and receive traffic flows.

Framing: IPoWDM Solutions

There are several approaches proposed to date for framing IP packets. The most important ones are SDH/SONET and Gigabit Ethernet framing [3].

Optical networks can need signal regeneration if the covered distances are long enough (e.g., several hundred kilometers). Most of today's regeneration systems are designed to work with SDH/SONET. In that case it is necessary to packet the IP datagrams into the SDH/SONET frames. Nevertheless, SDH/SONET framing has several limitations related to segmentation and reassembly (SAR) processing, which can be very time-consuming on a router interface card, resulting in a degradation in throughput and performance. Several companies are working on a new framing standard called Fast-IP or Slim SDH/SONET, which will provide for much of the functionality of SDH/SONET framing but use more modern techniques for header placement and matching frame size to packet size. The main advantage of SDH/SONET framing is that it carries signaling and network management information in its header bits. However, SDH/SONET has a large amount of overhead reserved for fault monitoring and operation support systems. This overhead could be minimized if these functions were incorporated into the IP routing devices. On the other hand, a drawback of SDH/SONET framing is the current high cost of SDH/SONET transponders and regeneration equipment.

Several companies are working on a new framing standard called Fast-IP or Slim SDH/SONET, which will provide much of the functionality of SDH/SONET framing but use more modern techniques in terms of header placement and matching frame size to packet size.

The other approach lies in the use of typical LAN equipment for regeneration, such as Gigabit Ethernet (GE). This approach is more suitable for municipal networks, where bandwidth is more available and access systems can have proprietary protocols. GE is not as efficient as SDH/SONET since it uses a simple block coding scheme where every 8 data bits are encapsulated in a 10-bit transmission block. This overhead results in network inefficiency of over 25 percent. However, a number of vendors are working

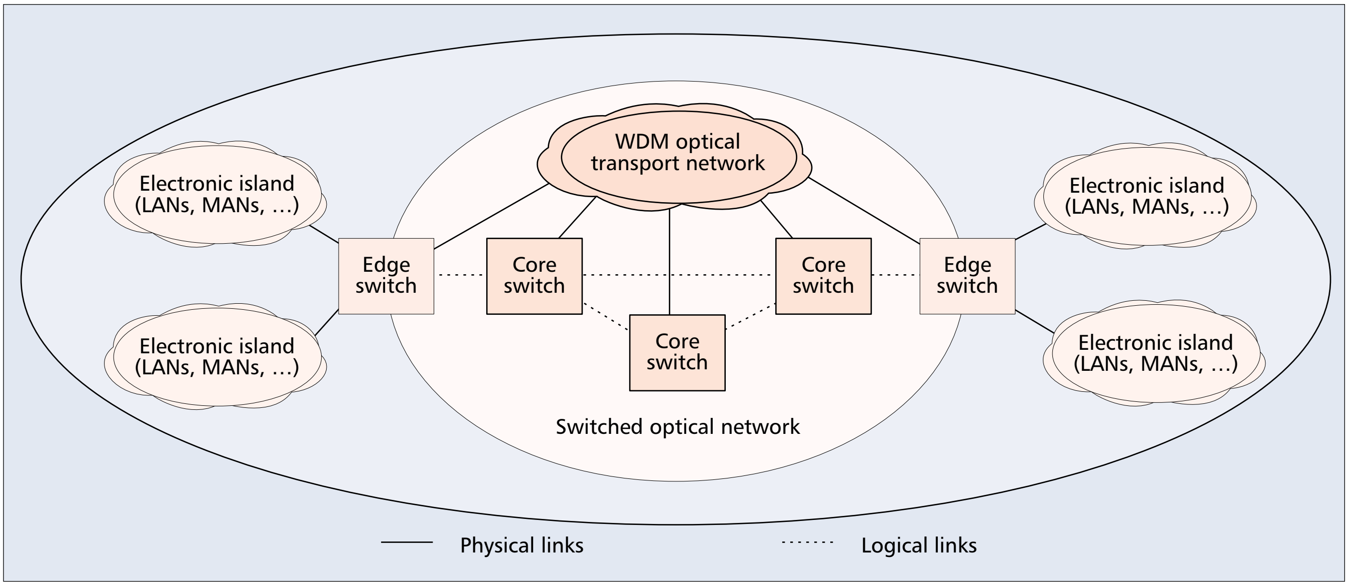
WDM optical transport network
Electronic island (LANs, MANs, ...)
Electronic island (LANs, MANs, ...)
Electronic island (LANs, MANs, ...)
Electronic island (LANs, MANs, ...)
Edge switch
Core switch
Core switch
Core switch
Edge switch
Switched optical network
Physical links
Logical links

Figure 6. *The switched optical network structure.*

on a new 10xGE standard, specifically designed for dense WDM (DWDM) systems. It is expected that the new 10xGE standard will use a much more efficient block coding scheme, perhaps even synchronous coding like SDH/SONET.

GE presents several advantages:

- It has low cost and optimized design to carry the same frames used by most networked computers.
- It uses the same frames originally generated by the hosts on either end of a connection, so there is no remapping to other transport protocols like SDH, and ATM, and as such SAR and bit stuffing operations are not required in the router interface to align the data frame with the transport frame.
- It provides lower cost per tributary delivery.

Advanced Network Architectures

The research carried out so far is driven by the assumption that the future optical telecom network will remain circuit-based. Therefore, applications of WDM as a networking technique [5] focus on the static utilization of single channels resulting in very inefficient optical bandwidth usage. A technological breakthrough in this direction is represented by optical packet switching [6], enabling fast allocation of the WDM channels and their utilization as shared resources in an on-demand fashion with very fine granularities. This leads to a significant increase in statistical multiplexing gain with respect to the case of electronic IP routers which, treating a WDM link as a collection of independent transmission channels, need a high number of physical interfaces to be used to terminate a single fiber, thereby increasing system complexity and hence cost [7].

Together with the efficiency aspects, a further factor fueling the interest in optical packet switching is bit rate transparency [6], entailing intrinsic flexibility to cheaply support incremental increases of the bit rate of transmission links; that is, successive upgrades of the transmission layer capacity can be planned with minor impact on the switching nodes [6].

Currently, the challenge is how to combine the advantages of the relatively coarse-grained WDM techniques with optical switching capabilities to yield a high-throughput optical platform able to efficiently support IP traffic.

A promising direction for network evolution lies in the migration of most of the switching burden into the optical domain in order to exploit the scalability property provided by optical technology to support progressive increments of WDM link transmission capacity. This approach could lead to an effective decoupling between transmission/switching and routing/forwarding functionality. The former should be handled in the optical domain so as to access the huge fiber bandwidth; the latter should be carried out in the electronic domain, where the routing/forwarding functions based on packet header processing should be performed.

The previously delineated network structure is sketched in Fig. 6. Two functional layers are envisaged. The external one is the electronic layer, performing traffic aggregation and main packet routing functions; the internal layer, here called the *switched optical network* (SON), is based on optical technology, and performs transmission and low-layer switching functions.

The *edge switches* (ESs) are located at the boundary between the two layers. IP traffic is injected into ESs by standard electronics networks (i.e., LANs, MANs, etc.). The ESs perform traffic aggregation and basic routing functions; that is, they determine to which ESs the incoming IP packets have to be forwarded. An ES assembles incoming IP packets directed to a given destination ES in an optical packet. A critical issue is the waiting time of the IP packets needed in this assembly process.

Once the optical packet is assembled, it is delivered to the SON. The SON transports optical packets from source to destination ESs. At the destination ES, the traffic is disaggregated and deliv-

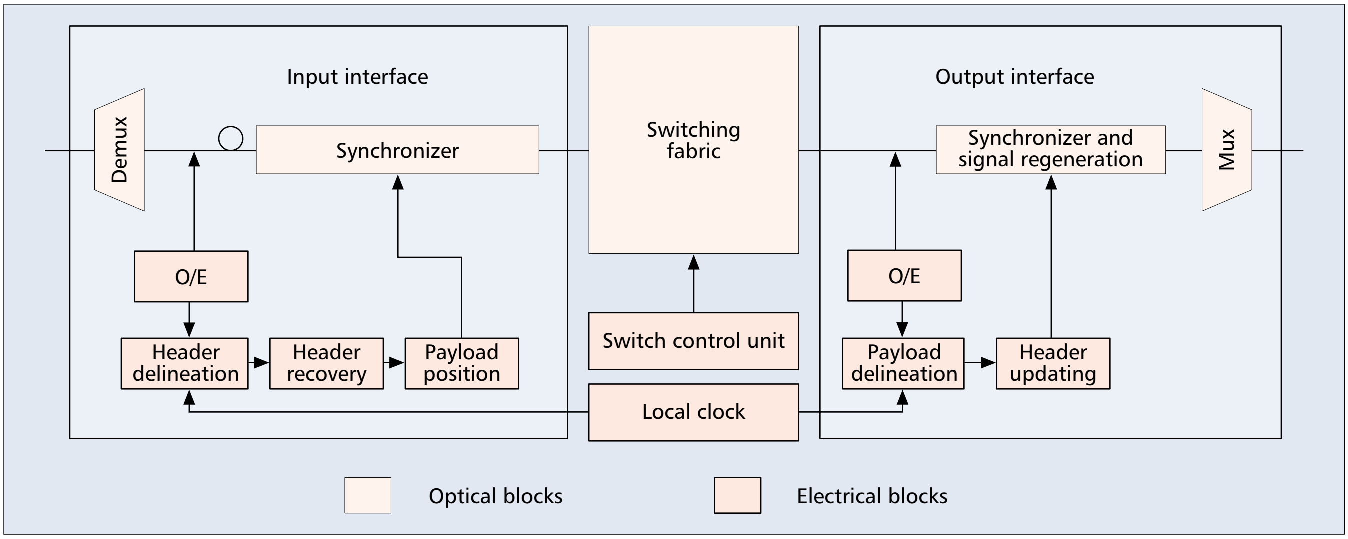

Figure 7. *Optical transparent packet network (OTPN) packet format.*

ered to the destination network. The SON switches, here called *core switches* (CSs), are interconnected via a WDM optical transport network. The CSs perform the forwarding of the optical packets in the optical domain and have the goal of handling the statistical multiplexing over the WDM links.

In order to simplify the packet forwarding process within the optical nodes, MPLS [8, 9] could be used in CSs and ESs. MPLS is a link-level forwarding technique able to provide simpler and faster packet forwarding capability than in traditional schemes like those used in IP networks. As an example, in the IP layer forwarding requires the analysis of a relatively large header and the execution of a longest match algorithm in order to determine the output to which packets have to be forwarded. In MPLS, label-swapping packet forwarding is based on a simple short-label exact match; this results in a simpler forwarding paradigm. According to MPLS, the entire forwarding space is partitioned into *forwarding equivalency classes* (FECs) and the packets belonging to the forwarding subspace relative to a given FEC are forwarded in a similar manner; this happens as follows. A short, fixed-length, locally significant identifier known as a label is assigned to each FEC. A packet is labeled by either encoding a label in an available location in the data link layer or network layer header, or encapsulating the packet with a header specifically for this purpose. In the context of the SON, IP packets arriving at a source ES and directed to the same destination ES are assembled into an optical packet to which is assigned a given label. The next-hop CS uses the label as an index into a table which specifies the next outgoing label and the next hop. The old label is replaced with the new one, and the packet is forwarded to the next hop. This eliminates the need for network-layer lookups from all but the first node in the path from source ES to destination ES.

The crucial aspect to be discussed in this scenario is the choice of the more appropriate packet transfer mode within the SON. Such a transfer mode has to arise from a compromise between the efficiency requirement that has to be guaranteed and the capabilities of current optical technology.

Two possible transfer modes have been proposed in the literature to be implemented within the SON. They basically differ in the structure of the optical packet and the network node operation. The first, the optical transparent packet network (OTPN), studied in ACTS Project KEOPS [6], is based on fixed-length packets with synchronous node operation. The second, optical burst switching (OBS) [7, 10], is based on variable-length packets, indicated as bursts, with asynchronous node operation.

In the next section the principles of OTPN and OBS will be illustrated and their implementation issues discussed.

The Optical Transparent Packet Network Paradigm

The optical packets used in the OTPN are placed in a fixed-duration time slot, allowing for synchronous operation of the switching nodes. The general format of an OTPN packet is shown in Fig. 7 [6]. The packet contains:

- A header at a fixed bit rate that is electronically processed in a node
- A payload with fixed duration and variable bit rate
- Guard times inserted to account for the device switching times, the jitter experienced by the payload within the node, and the nonideality of the synchronization units in the input/output node interfaces

The OTPN switch structure shown in Fig. 8 is composed of three main blocks:

- The input interfaces, performing packet delineation and synchronization functions in order to align the optical packets coming from the various input ports; this is realized by means of an optical synchronizer in order to keep the payload transparency
- The optical switching matrix, with the tasks of routing the optical packets toward the output ports and solving output packet contention

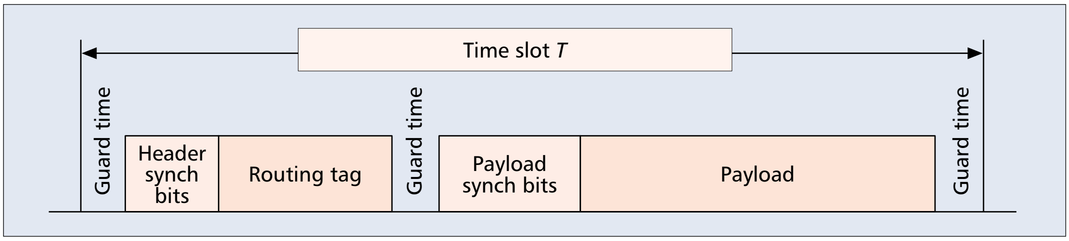

Figure 8. *OTPN switch architecture.*

One of the key problems in the application of packet switching in the optical domain is the handling of packet contention when two or more incoming packets are directed to the same output line.

- The output interfaces, performing the output synchronization operation in order to reduce the jitter due to the different paths followed by the packets inside the switch; in addition, such a block realizes the header rewriting operation and optical regeneration of the signal required to compensate the degradation of both the extinction ratio and optical signal-to-noise ratio introduced by the switching matrix

A critical issue in OTPN is the need to implement packet synchronization in the optical domain in order to guarantee bit-rate-transparent operation of the switch. As aforementioned, two synchronizers are needed:

- An input synchronizer is used in order to compensate for the slow jitter of packets arriving at the same input; these delay variations are due to both temperature variations and chromatic dispersion resulting in different propagation speeds over different wavelengths.
- An output synchronizer, used to compensate for the delay variation of the packets inside the switching matrix; in fact, packets can follow different paths with unequal lengths within the switching fabric, resulting in fast jitter of the packets directed to the same output port.

Since both the synchronizers are realized by means of fiber delay lines and switches of various technologies [11], the hardware complexity of the switch increases, as well as crosstalk and attenuation due to insertion loss, which introduce a degradation of signal quality increasing with the number of cascade nodes.

To take into account the fact that packet jitter can only be partially compensated for by synchronization units, it is necessary to introduce guard times in the optical packet before and after the payload in order to prevent payload damage during header erasure or insertion. The guard times are also introduced in order to take into account the switching time of the opto-electronic devices and the finite precision of fiber delay lines. Notice that when the header is transmitted serially with the payload, the guard times are introduced for both reasons mentioned earlier. On the contrary, they are introduced only for the latter reason if packet optical switching with subcarrier multiplexing (SCM) is adopted. Accordingly, header and payload are multiplexed on the same wavelength, and the current modulating the laser is constituted by a baseband signal that is the payload, in addition to a pass-band signal which is the header. In particular, in [12] header updating implemented by means of SOA is proposed, allowing realization of bit-rate-transparent switching.

Static inefficiency due to the insertion of guard times can be reduced by increasing optical packet duration. However, this can cause dynamic inefficiency if low traffic intensity periods between ESs occur. In these cases the number of IP packets could be insufficient to fill the payload of the optical packet, whose length is constant.

Contention Resolution in OTPN Switches

One of the key problems in the application of packet switching in the optical domain is the handling of packet contention when two or more incoming packets are directed to the same output line. Various techniques have been examined in literature:

- Buffering
- Wavelength translation
- Deflection routing
- Wavelength dimension

The application of the classical buffering technique makes the structure of an optical packet switch strictly close to that of a traditional electronic packet switch. For this reason it has been primarily and extensively studied (e.g., [13]). Unfortunately, at least with current technology, optical buffering can only be implemented via a bundle of fiber delay lines (FDLs) with lengths equal to a multiple of a packet duration. Hence, the buffer capacity of an optical packet switch cannot exceed a few units. Moreover, the number of FDLs is a critical system design parameter because it has an impact on optical hardware volume, switch size, and noise level due to the transit of the optical signal in the FDLs.

As in the case of the electronic switches, different packet buffering techniques have been investigated:

- Shared buffer
- Output queuing
- Partially shared buffer

In an optical packet switch adopting the shared buffer technique [13], if more than one packet is directed to the same output, all but one recirculate from the output to the input line through recirculation loops. Optical amplifiers are used to overcome the signal attenuation introduced by the space switch on the recirculating packets. When the traffic is bursty and the load high, many recirculations are required, causing accumulation of amplifier spontaneous emission (ASE) noise in the loops. To overcome these problems, packets that have to suffer different delays recirculate in fiber delay lines of different lengths.

An optical packet switch adopting the output queuing technique [6], consists of a space switch with a buffer on each output. A buffer consists of fiber delay lines of different lengths. This structure is not limited by ASE

As the number of delay-lines dramatically increases when the load is high and the traffic profile is "critical," a new approach has been proposed that uses the wavelength dimension for contention resolution by converting packets, addressed towards the same output line, to different wavelengths.

noise, but requires more hardware (total length of fiber delay lines) than needed by architectures employing recirculation buffering techniques.

Finally, the partially shared buffer technique [14] dedicates an optical buffer to each output and incorporates an additional common buffer shared among all of the output lines. In the switches adopting this technique, the packets, finding the output buffer full, are routed to the shared buffer for temporary storage. Subsequently, the packets stored in the shared buffer will recirculate back to an input port for further attempts. The partially shared buffer technique combines the advantages of both previously mentioned approaches. In fact, an optical architecture provided with correctly dimensioned output and recirculating buffers uses limited hardware and partly overcomes the problem of signal attenuation.

Since the number of delay lines dramatically increases when the load is high and the traffic profile is critical (burst traffic), a new approach has been proposed [15]. It uses the wavelength dimension for contention resolution by converting packets addressed tow the same output line to different wavelengths. This is accomplished by means of tunable optical wavelength converters (TOWCs). In [15] the improvement in packet loss probability is analyzed when packet wavelength conversion is used. Also investigated is how TOWCs allow reduction of the number of fiber delay lines by storing multiple packets on different wavelengths in the same fiber. The basic result is that an optical packet switch architecture uses a number of TOWCs proportional to the total number of input channels (i.e., the product of the number of input lines and the number of wavelengths).

Deflection routing [16] is simply a multipath routing technique that allows the contention problems to be solved, and the buffer depth and number of optical gates to be reduced with reasonable savings in hardware volume and cost. The effectiveness of this technique critically depends on network topology; as a matter of example, meshed topologies with a high number of interconnections experience the largest gain from deflection routing, whereas minor advantages arise from simpler topologies.

The wavelength dimension technique uses the wavelength dimension as a logical buffer in the WDM optical network layer. In [17] a network solution is proposed that eliminates the need for optical buffers by splitting the traffic load on the wavelength channels by using TOWCs. The proposed scheme solves the problem regarding optical buffering; however, it implies a high number of TOWCs. In fact, one TOWC is needed for each input wavelength channel. For example, in [17] it is shown that, if the intensity traffic per line is 0.8 and the required packet loss probability equals 10^{-10}, a 16 x 16 switch with 11 wavelengths/input requires only 2816 gates, compared to a "traditional" switch with optical buffers employing only one wavelength, which would require 12,288 gates; however, a high number of converters are necessary: 176 TOWCs.

The Optical Burst Switching Paradigm

As previously described, the two major problems of the OTPN are, on one hand, the transmission inefficiency due to the choice of a constant packet size, and on the other the difficulty of realizing optical packet synchronizers. This has led to the definition of a new switching paradigm, optical burst switching (OBS) [7, 8]. OBS is based on:

- Variable-length packets, named bursts
- Asynchronous node operation
- Decoupling of the burst payload from its header, which is transferred on a wavelength different from that of the payload

According to OBS, when an ES has a burst to send, it emits a burst control packet (BCP) on a prefixed control wavelength channel, followed by the data burst on an unused data wavelength channel. Along the path from the source ES to the destination ES, the BCP is processed electronically, and resources are reserved on the data path for the transmission of the burst. The architecture of an OBS CS is shown in Fig. 9.

The figure depicts a node with *N* input and output fibers; each fiber has *W* wavelengths for data channels and one for the control channel. The demux is the first component of the OBS switch; its role is to split the input control channel (ICC) used by the BCPs, and input data channels (IDCs) used by the data bursts. When a BCP reaches a CS it is immediately converted in the electronic domain by the input module (IM) and directed to the BCP router that determines on which output fiber to send the BCP and the related data burst. A fiber delay line is used to delay the data burst in order to process the BCP. Once the BCP has been processed, it is transferred to the output module/transmission (OM/TX), which updates the control fields contained in the BCP and transmits the data burst on the wavelength channel determined by the scheduler. Finally, the mux inserts the control channel in the output fiber.

To efficiently use the switch resources [7], the scheduling of the BCP is not performed when the BCP arrives but an instant before the related data burst arrival; in order to delay the scheduling of the BCP, a reordering buffer is used whose queue is ordered according to the arrival times of the data bursts and whose logic delivers the BCP to the scheduler s before the burst arrival,where is the sum of scheduling and optical switch setting times. The scheduler processes the BCP and reserves resources needed to forward the data burst; furthermore, it reveals any burst contention phenomena on the output wavelength channels.

Note that although OBS represents a simpler solution, it does not solve all the problems of the OTPN: the resolution of burst contention is an open problem, and it is to be expected that performance, in terms of burst loss probability, could degrade on the OTPN. This is due to two main reasons:

- There are more output packet contentions due to the more unpredictable and less regulated burst statistics.
- The FDLs are not efficiently employed, and voids between bursts can take place, as in [18].

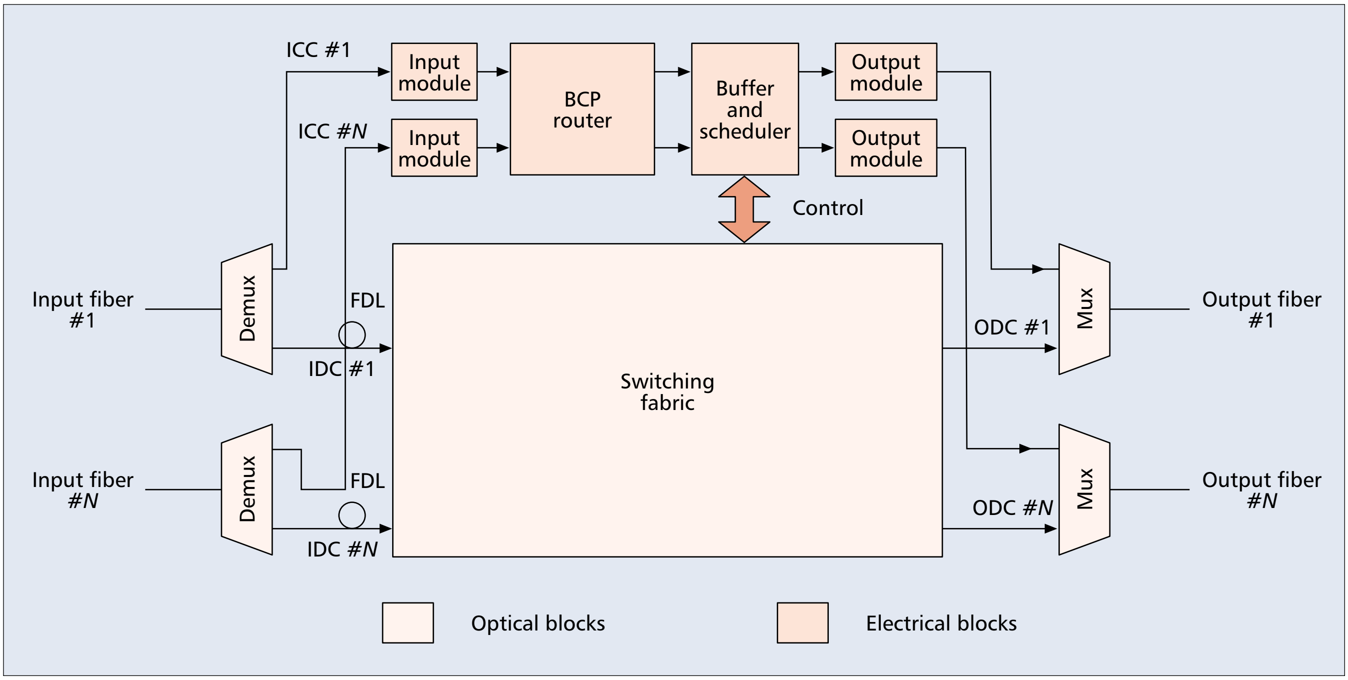

Figure 9. *Optical burst switching (OBS) node architecture.*

As far as the last point is concerned, we can affirm that the main problem is the dimensioning of the basic timescale unit *D* to be used for the FDLs. If the FDL buffer is designed using fibers introducing delays that are consecutive multiples of *D*, small values of *D* lead to high time resolution and poor buffering capacity, while large values of *D* lead to large buffering capacity with poor time resolution. Neither case is optimal, and there is a trade-off between them to provide an acceptable optical burst loss probability. In order to fill the void, in [18] a new scheduling algorithm is proposed; nevertheless, it is only able to solve the problem partially, at the price of an increase in the control complexity of the optical switch. In [19] a switch architecture is proposed, equipped with multistage FDLs allowing a buffer with fine granularity and long delay to be obtained; however, this increases the hardware complexity of the switch in terms of the number of used optical gates.

WDM Technologies

In a recent article published in this magazine [20] the main technologies for DWDM networks were reviewed. In particular, the authors reported several device technologies for the realization of tunable and switched sources, tunable filters, wavelength converters, wavelength routers, and switching elements. The maturity of the technology needed to manufacture these devices is a key issue. For instance, some of these have a high maturity and can easily be inserted in real systems (e.g., tunable filters); some others are still not mature enough for employment in practical systems (e.g., wavelength converters). However, we can expect that in the following two or three years most of them will gain a technological maturity which will allow the implementation of more complex DWDM systems.

Most DWDM network architectures presented so far are based on static or semi-permanent wavelength routing. This means that the status of the network and its devices changes very slowly with time, on the order of hours or days. Nowadays, advanced network architectures are hypothesized, which will allow more flexible and dynamic use of wavelength resources, depending on the variation of traffic dynamics. This is particularly true in the case of optical networks for data traffic (e.g. optical Internet, or optical networks which interconnect several IP routers). A relevant example is represented by multi-protocol lambda-switched optical networks, which are optical networks compatible with the multi-protocol label switching[1] scheme proposed for Internet routing. In this case, dynamic routing in a timeframe of seconds or even less is required. WDM networks which employ such dynamic and flexible routing schemes need *wavelength agility*, that is, the property of optical devices to rapidly change their working conditions.

Wavelength agile devices have already been demonstrated. In particular, burst mode operating receivers and agile wavelength converters have been realized.

The wavelength agility characteristic requires technological efforts to render such devices reliable and well performing. This means that much effort should be made to push the maturity of the technology at a reasonable level. Wavelength agility is the key function for realizing wavelength/time-division multiple access (WDMA/TDMA) access systems. A relevant example is represented by the switchless network concept reported in [21]. It basically consists of a single-hop shared-access network employing time and wavelength agility (a WDMA/TDMA scheme), using fast tunable transmitters and receivers to set up individual customer connections through a single wavelength router (suitably replicated

[1] *The same acronym, unfortunately, is used to identify either the electronic forwarding scheme or the optical network compatible with such an approach.*

for resilience). In such a network connection among users is realized by a double dimension resource: wavelength and time slots.

The role of WDM technology is crucial for providing networking solutions for the future backbone networks, due to the huge bandwidth of the optical fiber, and to the high throughput characteristics provided by the optical routing nodes.

A further step beyond is represented by the realization of devices suitable for optical packet switching (OPS). In this case it is much more difficult to foresee when and if the maturity of these devices will be such that OPS networks can practically be realized.

Conclusions and Perspectives

The convergence between the telecom and datacom worlds into the infocom scenario will dramatically change network design. One of the most significant aspects is related to the dominance of Internet data traffic over traditional voice traffic. The role of WDM technology is crucial in providing networking solutions for future backbone networks, due to the huge bandwidth of optical fiber and the high throughput characteristics provided by optical routing nodes. For the time being several architectural approaches are emerging to provide effective transport of IP packets. Much effort is being made to find solutions that harmonize the statistical multiplexing advantages provided by packet-switching approaches and the intrinsic circuit-switched nature of optical networking. Optical networking solutions that utilize the same control plane as MPLS networks could be a valid solution in the short term. Optical burst switching may represent an interesting method in the mid/long term. Optical packet switching can be seen as the last step of such an evolutionary path. Anyway, some breakthroughs are needed in order to realize mature optical devices which allow dynamic buffering of optical packets.

References

[1] G. Prati, Ed., *Photonic Networks*, Springer Verlag, 1997.

[2] W. E. Leland *et al.*, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Trans. Net*, vol. 2, no.1, 1994, pp. 1–15.

[3] T. W. Chung *et al.*, "Architectural and Engineering Issues for Building an Optical Internet," draft, 1998; http://www.canet2.net

[4] R. Sabella and P. Lugli, *High Speed Optical Communications*, Kluwer, 1999.

[5] R. Ramaswami and K. N. Sivarjan, *Optical Networks*, Morgan Kaufmann, 1998.

[6] C. Guillemot *et al.*, "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach," *IEEE J. Lightwave Tech.*, vol.16, no. 12, Dec. 1998.

[7] J. Turner, "Terabit Burst Switching," *J. High Speed Networks*, vol. 8, no. 1, 1999, pp. 3–16.

[8] A. Viswanathan *et al.*, "Evolution of Multiprotocol Label Switching," *IEEE Commun. Mag.*, May 1998, pp. 165–73.

[9] R. Callon *et al.*, "A Framework for Multiprotocol Label Switching," Internet draft, daft-ietf-mpls-framework-05.txt, Sept. 1999.

[10] C. Qiao and M. Yoo, "Choices, Features and Issues in Optical Burst Switching (OBS)," *Opt. Networking Mag.*, vol. 2, Apr. 1999.

[11] M. Burzio, P. Gambini and L. Zucchelli, "New Solution for Optical Packet Delineation and Syncronization in Optical Packet Switch Networks," *ECOC '96*, Oslo, Norway, Sept. 1996, pp. 3.301–04.

[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 Amplifier," *IEEE Phot. Tech. Lett.*, vol. 9, 1997, pp. 827–29.

[13] D. K. Hunter, M. C. Chia and I. Andonovic, "Buffering in Optical Packet Switches," *IEEE J. Lightwave Tech.*, vol. 16, no. 12, Dec. 1998.

[14] J. Diao and P. L. Chu, "Analysis of Partially Shared Buffering for WDM Optical Packet Switching," *IEEE J. Lightwave Tech.*, vol. 17, no. 12, Dec. 1999, pp. 2461–69.

[15] S. L. Danielsen, P. B. Hansen and K. E. Stubbkyaer, "Wavelength Conversion in Optical Packet Switch," *IEEE J. Lightwave Tech.*, vol.16, no. 12, Dec. 1998.

[16] G. Castanon, L. Tancevski and L. Tamil "Routing in All-Optical Packet Switched Irregular Mesh Networks," *GLOBECOM '99*, vol. b, Rio de Janeiro, Brazil, Dec. 5–9, 1999, pp. 1017–22.

[17] S. L. Danielsen *et al.*, "Optical Packet Switched Network Layer Without Optical Buffers," *IEEE Phot. Tech. Lett.*, vol. 10, no. 6, June 1998.

[18] L. Tancevski *et al.*,"A New Algorithm for Asynchronous Variable Length IP Traffic Incorporating Void Filling," *Proc. OFC '99*, San Francisco, CA, Feb. 1999, pp. 180–82.

[19] F. Callegati, G. Corazza, and C. Raffaelli, "An Optical Packet Switch with a Multi-stage Buffer for IP Traffic," *Optical Networking*, A. Bononi, Ed., Springer Verlag, 1999.

[20] J. M. H. Elmirghani and H. T. Mouftah, "Technologies and Architectures for Scalable Dynamic Dense WDM Networks," *IEEE Commun. Mag.*, Feb. 2000, pp. 58–66.

[21] N. P. Caponio *et al.*, "Single-Layer Optical Platform Based on WDM/TDM Multiple Access for Large-Scale 'Switchless' Networks," *ETT*, vol. 11, no. 1, Jan./Feb. 2000, pp. 73–83.

NOTE: A detailed list of references is contained in the electronic version of the magazine (*IEEE Communications Interactive*).

Biographies

MARCO LISTANTI [M] (marco@infocom.ing.uniroma1.it) received his Dr. Eng. degree in electronics engineering from the University of Rome "La Sapienza" in 1980. He joined the Fondazione Ugo Bordoni in 1981, where he was leader of the TLC Network Architecture group until 1991. In November 1991he joined the Infocom Department of the University of Roma "La Sapienza," where he is associate professor in switching systems. Since 1994, he also collaborates with the Electronic Department of the University of Rome "Tor Vergata," where he holds courses in telecommunication networks. He participated at several international research projects sponsored by EEC and ESA and is the author of several papers published in the most important technical journals and conferences in the area of telecommunications networks. His current research interests focus on traffic control in IP networks and the evolution of techniques for optical networking. He has been representative of Italian PTT administration in international standardization organizations (ITU, ETSI).

VINCENZO ERAMO received the degree in Electronic Engineering from the University of Rome "La Sapienza" in 1996. From June to December 1996 he was a researcher at Scuola Superiore Reiss Romoli. In 1997 he joined Fondazione Ugo Bordoni as a researcher in the Telecommunication Network Planning group. He is pursuing a Ph.D. degree at the University of Rome "La Sapienza." HIs current research interests include network performance and traffic characterization; introduction of new services (video on demand) in the access network; QoS support in the Internet; and optical packet switching.

ROBERTO SABELLA (roberto.sabella@tei.ericsson.se) received his Dr. Eng. degree in electronics engineering from the University of Rome "La Sapienza" in 1987. He then joined Ericsson Telecomunicazioni, Rome, Italy, where he was involved first in hardware design and subsequently in research activities on advanced fiber optic communication systems. His research interests have covered the fields of optical device technology, high-speed optical communication systems, and WDM optical networks. In May 1997 he joined CoRiTel consortium as research technical coordinator. Since 1999 he has been the manager of a research department in Ericsson Lab Italy, Rome. He holds two patents on optical cross-connects, is co-author of a book on high-speed optical communications, and is the author of about 80 papers in scientific/technical journals and international conferences. He has been a lecturer (professore a contratto) at the University of Rome "Tor Vergata" and the Polytechnic of Bari, Italy. He is a member of the IEEE/LEOS Technical Committee on Optical Networks and Systems, and a member of the editorial board of the new international journal *Photonic Network Communications* (New York: Kluwer Academic). For the same journal he has operated as guest editor for special issues on WDM transport networks. He was one of the guest editors of a special issue on optical networks for *Computer Networks* (Elsevier).