RingO: An Experimental WDM Optical Packet Network for Metro Applications

Andrea Carena, *Member, IEEE*, Vito De Feo, *Student Member, IEEE*, Jorge M. Finochietto, *Student Member, IEEE*, Roberto Gaudino, *Member, IEEE*, Fabio Neri, *Member, IEEE*, Chiara Piglione, *Student Member, IEEE*, and Pierluigi Poggiolini, *Member, IEEE*

***Abstract*—This paper presents Ring Optical Network (RingO), a wavelength-division-multiplexing (WDM), ring-based, optical packet network suitable for a high-capacity metro environment. We present three alternative architectural designs and elaborate on the effectiveness of optic with respect to electronic technologies, trying to identify an optimal mix. We present the design and prototyping of a simple but efficient access control protocol, based upon the equivalence of the proposed network architecture with input-buffering packet switches. We discuss the problem of node allocation to WDM channels, which can be viewed as a particular optical network design problem. We, finally, briefly illustrate the fault protection properties of the RingO architecture.**

The main contribution of this paper is the identification and experimental validation of an innovative optical network architecture, which is feasible and cost effective with technologies available today, and can be a valid alternative to more consolidated solutions in metro applications.

***Index Terms*—Metropolitan area networks, optical packet networks, optical testbeds, wavelength-division-multiplexing (WDM) rings.**

I. Introduction

THE MARKET segment of metropolitan high-speed networks is alive despite the current telecom crisis. According to several studies, the provision of low-cost broadband access in metropolitan areas has the potential for fast returns on investments, and can foster the development of new bandwidth-hungry applications, which in turns should lead to the long-sought return to the fast increase of user demands that can revitalize the telecom market.

Metro networks are characterized by high dynamism of traffic patterns, relatively high aggregate bandwidths, and relatively short covered distances. Technical solutions for metro architectures are far from being consolidated, and range from classical circuit-switched synchronous optical network/synchronous digital hierarchy (SONET/SDH) rings, to extensions of traditional high-speed local area networks (LANs), such as the resilient packet ring IEEE 802.17, to switched (multi) Gigabit Ethernets, to Broadband Passive Optical Networks ITU-T G.983, to more innovative optical packet switching proposals. The latter are considered by many researchers the only approach capable of withstanding in the long term the continuous growth of aggregate capacities.

Wavelength-division multiplexing (WDM) is today a well-established technique to exploit the fiber bandwidth in both core and metro networks, and all major vendors in this field offer a wide range of products and commercial solutions. The development of optical technologies for applications beyond point-to-point transmission has instead suddenly slowed down due to the telecom market downfall. Nevertheless, at research and standardization levels, a large effort is being devoted to exploit optical technologies also for the implementation of network functions such as switching, protection, and restoration [1].

Nowadays, the most advanced products essentially provide optical *circuit* switching at the wavelength level (see, for example, [2]), in the sense that end-to-end optical lightpaths are dynamically set up and torn down upon network, or even user requests. On the contrary, the implementation of optical *packet* switching functions [3] (i.e., of an optical layer that can handle and switch data traffic on time scales in the order of microseconds or less) is still at an earlier development stage, although several prototypes and testbeds have already been demonstrated [4]–[6]. This is certainly due to the high technological challenges inherent in dealing with packets directly at the optical level. Indeed, although optical devices allow huge potential in terms of available bandwidth, they do not easily offer substantial features in terms of very fast switching, processing speed, and storage of digital signals, which are instead necessary for packet switching and are very natural and easy to implement in the electronic domain.

Metropolitan area networks are one of the best arenas for an early penetration of advanced optical technologies. Indeed, their large traffic dynamism requires packet switching to efficiently use the available resources; their high-capacity requirements justifies WDM use; and their limited geographical distances lowers the impact of fiber transmission impairments. From a research view point, designing innovative architectures for metro networks often means finding cost-effective combinations of optic and electronic technologies and new networking paradigms that better suit the constraints dictated by available photonic components and subsystems.

Our research group has designed and prototyped network architectures for metro applications, taking an approach based

Manuscript received July 30, 2003; revised March 10, 2004. This work was supported in part by the Italian Ministry for Education, University and Research (MIUR) under the PRIN Projects "RingO" and "Wonder," in part by the FIRB Project "Adonis," and in part by the European Commission under the FP5 IST Project "David."

A. Carena, R. Gaudino, and P. Poggiolini are with the PhotonLab, Dipartimento di Elettronica, Politecnico di Torino, Torino 10129, Italy (e-mail: andrea.carena@polito.it; roberto.gaudiono@polito.it; pierluigi.poggiolini@polito.it).

V. De Feo, J. M. Finochietto, F. Neri, and C. Piglione are with the Dipartimento di Elettronica, Politecnico di Torino, Torino 10129, Italy (e-mail: vito.defeo@polito.it; jorge.finochietto@polito.it; fabio.neri@polito.it; chiara.piglione@polito.it).

Digital Object Identifier 10.1109/JSAC.2004.830479

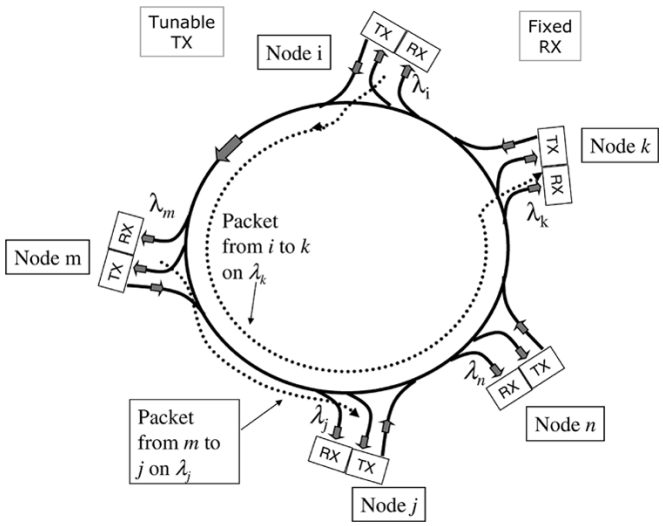


Fig. 1. Architecture of the RingO network.

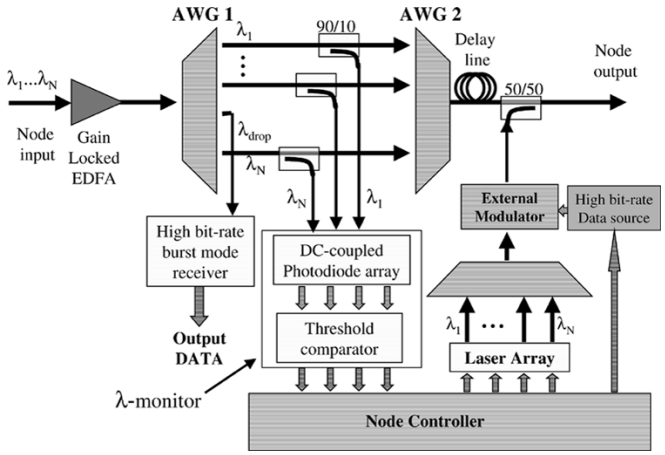


Fig. 2. First structure of RingO nodes.

upon optical packets, but limiting optical complexity to a minimum and trying to use only commercially available components. To best exploit the advantages of available technologies, the bulk of raw data is kept in the optical domain, while more complex network control functions are mostly implemented in the electronic domain. Likewise, neither distributed resource allocation nor contention resolution is performed in the optical domain, thereby taking a radically different perspective with respect to traditional electronic packet-switched architectures.

In this paper, we introduce the rationale, the network architecture and design of the ring optical network (RingO) project, carried out by a consortium of Italian Universities coordinated by the Optical Communications Group (OPTCOM) and the Telecomunication Network Group (TNG) of Politecnico di Torino. The RingO project is focused on experimentally studying the feasibility of a WDM optical packet network based on a ring topology. The presentation will evolve through three different network designs, both to follow the project history, and to ease the description for the reader.

The paper is organized as follows. The RingO general architecture, medium access control (MAC) protocol and node structure are explained in Section II. Section III briefly overviews physical-layer issues related to transmission impairments and network scalability. Then, in Section IV, we describe the current RingO experimental setup, presenting the demonstrator and some details of the node controller hardware implementation. In Section V, we present an interesting evolution of the node design, and discuss problems related to allocating nodes to the available WDM channels. Finally, in Section VI, we briefly discuss fault recovery mechanisms.

II. RingO Architecture

The general architecture of the RingO network is illustrated in Fig. 1, while the structure of a node is depicted in Fig. 2 (and described in more detail in Section II-A). As mentioned above, we will step through three different versions of the network architecture in this paper; they all preserve the same rationale and basic subsystems design.

The first version of the RingO network is based on a unidirectional WDM fiber ring with N network nodes equipped with an interface between the electronic domain and the optical domain. The main features of this first RingO architecture are the following:

- packets transmission is time-slotted and synchronized on all wavelengths; as reference values in RingO, the slot duration is 1 μs and the transmission bit rate is 2.5 Gb/s;
- packets have fixed length corresponding to one time slot: the packet format adaptation, possibly including segmentation/reassembly, or concatenation, is left to higher (electronic) layers of the node protocol, it is outside the scope of this paper;
- the number N of nodes in the network in this first design is equal to the number W of wavelengths (which will be often indicated in the following as "channels"): a given node i is, thus, identified by a wavelength λ_i, it is the only node able to receive this wavelength, and it is also responsible for physically removing it from the ring, using a fixed-wavelength optical drop filter;
- each node is equipped with a tunable transmitter since, in order to communicate to node k, a node must tune its transmitter to send a packet on λ_k, as shown in Fig. 1; tuning times are assumed to be short with respect to the slot duration;
- each node is able to check the state (busy/free) of all wavelengths (a feature called λ-monitoring) on a slot-by-slot basis, and avoids collisions and contentions by electronically queueing input packets, and by accessing channels using a suitable access protocol, as discussed in Section II-B.

In the architecture described above, the fixed relation between a destination node and a wavelength allows a significant simplification on the optical hardware with respect to most of other packet network proposals. First, packet headers are not required, at least for addressing functions, since the destination address is "coded" into the used wavelength. Second, packets do not need to be actively routed along the network, but are simply passively dropped at the destination by the node optical drop filter. As a result, our proposal is able to take advantage of packet statistical multiplexing without requiring optical switches. Third,

λ-monitoring can be obtained by simply measuring the power level in each slot and wavelength, without again requiring the presence of an optical header.

For what regards wavelength tunability requirements, it is easy to understand that full tunability either at the transmitter, or at the receiver, is required to provide full node-to-node connectivity avoiding a multihop operation, that would increase the amount of electronic processing in the network. Tunability is a characteristic feature of optical networks, leading to interesting and well-understood logical topology design and fault recovery approaches, but it is still very costly, specially if very high switching rates are necessary. All RingO designs chose to have fast-tunable transmitters, which are considered to be easier to implement than tunable receivers. The tunable transmitters are made by an array of ON–OFF switchable fixed lasers in the lab testbed, as discussed later.

The proposed architecture combines, in an efficient way, optic and electronic technologies: the aggregate bandwidth is handled in the photonic domain by working on a wavelength granularity, while packet queueing, MAC protocol, and statistical time multiplexing are handled in the electronic domain at the speed of a single data channel. Due to the *optical* simplicity of our solution, the resulting architecture does not offer all the networking features of other more complex optical network proposals [7], like a large Internet protocol (IP)-like addressing base, label swapping, arbitrary mesh topology, etc. However, we carefully selected a solution that is only based on commercially available optical components, and that at the same time offers a set of interesting features for metropolitan area networks connecting a limited number of very high-capacity nodes over a ring.

Our architecture does not require any advanced optical component, such as fast optical switches or wavelength converters. Moreover, it does not require at all optical buffering. In fact, packet buffering is implemented in the electronic domain at the boundary of the optical cloud. In our opinion, this is an important aspect, since it allows to both reduce optical complexity *and* to implement electronically efficient access algorithms.

The RingO structure is an evolution of certain WDM ring packet network proposals presented in the mid 1990s. In 1993, the first such proposal appeared in [8]. It featured a WDM ring architecture with time slotting and as many wavelengths as the number of nodes. It relied on fixed transmitters and tunable receivers, as opposed to later proposals that did the opposite. Shortly afterwards, in [9], a new network structure, using instead tunable transmitters and fixed receivers, was proposed. It already featured a λ-monitor-based protocol to avoid collisions, based on subcarriers: each wavelength carried a different and unique subcarrier frequency, which could be probed in a given time slot to assess the presence or absence of a packet on that wavelength. Later, the same basic ideas of [9] were independently brought to actual implementations at Stanford University, in the hybrid opto-electronic ring network (HORNET) project [5], and at Politecnico di Torino, in the RingO project. The two groups comprise some of the original authors of [9].

A. Node Structure

The structure of the first RingO node design is shown in Fig. 2. Some basic subsystems are common to all node architectures presented in this paper. Scanning the node structure from input to output, the main functions supported by the node are the following.

1) Amplification of the optical signals in order to compensate for the losses of the node passive elements and of the downstream fiber link.
2) Demultiplexing of the WDM comb after the amplifier. Devices which have been used for this purpose for the first RingO testbed are arrayed waveguide grating (AWG) filters.
3) Monitoring the state of channels on each slot. This is done by tapping a fraction of the power on each fiber at the output of the demultiplexer and by sending it to a DC-coupled photodiode array. This λ-monitoring electronics requires a much smaller bandwidth than the data bit-rate, since it should only detect the received average power on a slot-by-slot basis. Our solution for packet detection is easier to implement than other approaches, such as the often proposed subcarrier-tone detection [10].
4) Burst-mode detection of the incoming data-stream on the wavelength λ_i associated with node i. Note that the shift from continuous-wave operations of traditional optical network to our burst-mode operation is a major increase in complexity, but it is a price that we chose to pay to allow high efficiencies in resource utilization via statistical multiplexing.
5) Local packet traffic generation. We used a laser array driven by the node controller. The lasers are turned on for each time slot by direct current injection when a packet has to be generated. Data bits are then "written" inside the packet by an external modulator. This transmitter architecture has several motivations:
   - the use of an array of lasers, rather than a single fast tunable laser, allows using commercial and reliable devices on the ITU wavelength grid [2]; this choice was due to the difficulty in finding commercial fast-tunable lasers, and to the interesting opportunity to implement multicasting (see next item);
   - to allow efficient multicast, i.e., to send the same packets to multiple destinations. Multicasting is currently seen as an important requirement, since it is crucial to video conferencing and groupware, and indeed it is implemented in most of today commercial top-level routers [11]. In our situation, multicasting means to replicate the same packet on different wavelengths, possibly in the same time slot. With our structure, bits can be written *simultaneously* by the external modulator on all wavelengths that are generated by the laser array in a given time slot. In this way, multicasting in a single time slot can be implemented without increasing electrical bandwidth requirements at the transmitter, since the "replication" of packets is obtained in the optical domain (in which bandwidth efficiency is less critical). For what regards the electronic part of the transmitter, the cost of sending a packet to multiple destinations is the same as for sending a packet to a single destination.

As it can be seen from the description above, our architecture requires an electrical data path bandwidth, on both the transmitter and receiver side, that is equal to a single channel data rate. In fact, even when multicasting is implemented, the high-speed electrical interface of the transmitter and receiver need only to handle data traffic carried by a single wavelength, and not the aggregate bit rate of all wavelengths passing through the node. This is true for all RingO designs, and part of the RingO rationale: a metro architecture capable of scaling at large aggregate capacities must avoid to process at each node the whole network bandwidth. This was one of the problems that prevented a straightforward extension of the original LAN protocols and architectures (which assume to process the entire network bandwidth at each node interface) to metros. This is also one of the advantages of our architectures with respect to current SONET/SDH circuit-switched solutions.

B. MAC Protocol

Our architecture requires a suitable MAC protocol to allocate time slots to transmitters. From the MAC protocol design perspective, RingO is a multichannel network, in which packet collisions must be avoided and some level of fairness in resource sharing must be guaranteed together with acceptable levels of network throughput.

A collision may arise when a node inserts a packet on a time slot and wavelength which have already been used. This is avoided by giving priority to upstream nodes, i.e., to in-transit traffic, via the λ-monitoring capability.

Fairness is obtained by implementing an efficient *a posteriori* [12] packet selection strategy exploiting a virtual output queueing (VOQ) structure. While standard single-channel protocols use a single first-in–first-out (FIFO) electrical queue, in multichannel scenarios, where channels are associated with destination nodes, FIFO queueing performs poorly due to the head-of-line (HOL) problem [13]: a packet at the head of the queue may block following packets which could be transmitted on other channels. The HOL problem has been carefully studied and can be solved using one of the VOQ [13] structures. The basic VOQ idea, applicable to the RingO architecture, consists in storing packets waiting for ring access in separated FIFO queues, each corresponding to a different destination (or to a different set of destinations), and to appropriately select the queue that gains access to the channels for each time slot. It is worth noting that VOQ was demonstrated to be able to achieve 100% throughput for uniform and unicast traffic when suitable packet selection algorithms are implemented [13].

Another problem common to ring and bus topologies is the fact that an upstream node can "flood" a given wavelength, as shown in [14], reducing (or even blocking) the transmission opportunities of downstream nodes, thus generating a significant fairness problem. The fairness problem has also been investigated in detail in several previous papers, where it was shown that, again, it can be solved by using separate input queues, by selecting them with some form "round-robin" strategy (called antiresonant ring (ARR) or split ring resonator (SRR) in [14]), and by using a fairness control algorithm suited for this multichannel setup (called multimetaring).

It is not difficult to observe that our multichannel ring is equivalent to a distributed input-queued packet switch, in which node interfaces correspond to input/output line cards, and the fiber ring behaves as a distributed switching fabric. When one wavelength channel is associated with each receiver (as in Figs. 2 and 3), this switching fabric is functionally equivalent to a crossbar, capable in each time slot of delivering at most one packet to each destination, and of allowing at most the transmission of one packet from each source. In other words, in each time slot at most an input/output permutation can be served. Building upon this equivalence, the optimal packet selection criteria would be the outcome of a centralized maximal weight matching (MWM) algorithm, with weights equal to queue sizes [13]. Since this would have led to excessive complexities, our packet selection criteria is a distributed heuristic maximal approximation of MWM: each node transmits in a given slot the packet at the head of the longest of its several queues, neglecting queues whose HOL packets could not be transmitted because of the λ-monitor information. The implementation of the MAC protocol in RingO is further described in Section IV.

The complexity of the proposed MAC algorithm is mainly confined to the electronic domain, without stringent requirements on optical devices.

III. Transmission Issues in RingO

We studied RingO physical-layer design and performance in a previous paper [15], by detailed simulative analysis. Although the full set of results cannot be shown in this paper due to space limitations, the major results are briefly outlined in the following. We assume that the system is limited by the accumulation of ASE noise along the ring. We require a 2-dB margin over a reference bit-error rate equal to 10^{-9} for any receiver. Under reasonable assumptions, the cascadability of the node structure (shown in Fig. 2) was verified to reach 16 nodes, using 16 wavelength-division-multiplexing (WDM) channels each working at 10 Gbit/s, with a distance of 25 km between nodes. This transmission limit comes from ASE noise accumulation, and it is determined by the combination of the high insertion loss at each optical node, mainly due to the presence of two AWGs, and the limited signal power level at the output of the Erbium-doped fiber amplifiers (EDFAs). This cascadability is not large, but should be sufficient in a metro environment, which is the target of the RingO project.

Anyway, our paper showed that any optical impairment on the node input–output path is critical, since signals propagate all-optically along the ring without 3R regeneration. For example, in order to reach 16 nodes, polarization dependent loss (PDL) must be less than 0.6 dB per node. In addition, self-filtering effects can be critical: with the commercial AWGs used in our experiment, the required wavelength alignment accuracy for a 16 node ring should be better than 30 GHz for a 200-GHz WDM spacing.

The node structure based on AWGs, see Fig. 2, was first proposed because of the major flexibility given by fully demultiplexing all channels on separate fibers. Although some more advanced network functionalities could be envisioned with

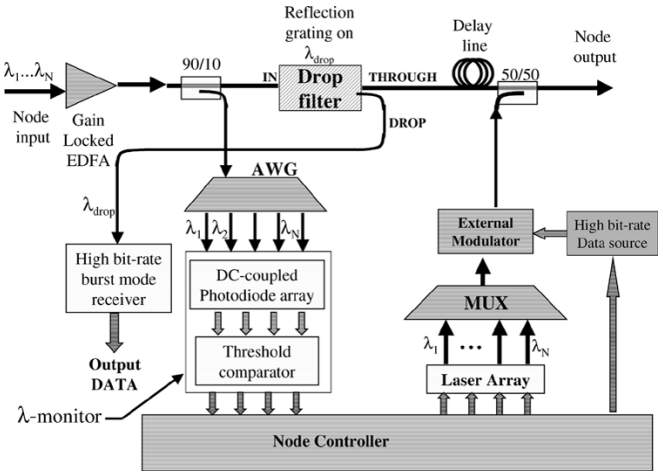


Fig. 3. Second structure of RingO nodes based on fiber-grating add-drops filters.

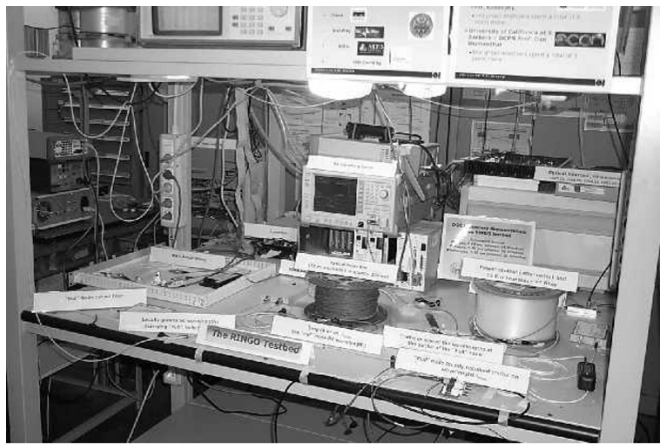


Fig. 4. RingO testbed in the PhotonLab at Politecnico di Torino.

such a structure, our simulative analysis and experimental measurements showed significant physical-layer performance limitations. In order to increase the scalability of the proposed network in terms of maximum number of nodes, we need to reduce the node insertion loss, PDL, and self-filtering effects.

These results motivate the introduction of our second node design, which is based on an add–drop filter, see Fig. 3, allowing for better cascadability and less stringent physical constraints.

While this structure is similar to the previous one for network functionalities, and most subsystems are directly derived from Fig. 2, it is significantly different from the physical layer point of view. The input–output optical path is greatly simplified, and now consists only of a passive optical splitter and a fixed add–drop filter tuned on the wavelength λ_{drop} that must be received locally. This setup greatly reduces node attenuation, self-filtering, and PDL effects, and allows a higher node cascadability.

IV. Experimental Testbed

The RingO network experimental testbed, shown in Fig. 4, was implemented in the PhotonLab at Politecnico di Torino, and was based upon nodes having the structure shown in Fig. 3. The RingO testbed goals are:

- the demonstration of the proposed architecture and MAC protocol;
- the availability of an experimental setup, where RingO physical transmission properties can be easily studied.

The testbed is currently based on two nodes, as depicted in Fig. 5, exchanging information on four different wavelengths, spaced at 200 GHz. The first one is used to generate random packet data traffic, while the second one implements all RingO protocol functions. We are, thus, able to generate an arbitrary stream of packets on any wavelength using the first node, and to demonstrate the MAC protocol operations in the second one. Since the *optical* details of the demonstrator were already shown in [15], we focus in this paper on the implementation of the node controller, which is based on a high-performance FPGA mounted on a custom-designed electronic board. The FPGA is an Altera APEX20KE600–3, with 600 000 gates, 24 320 flip-flops, four internal PLLs, 588 I/O. The working frequency can be set in the range of 30–133 MHz, thanks to the aid of the internal broadband PLL.

The node controller logical structure is shown in Fig. 6. In the following, we present the node functionality, focusing on multicast transmission. Unicast transmission can be seen as a particular case of multicast transmission in our architecture, requiring only a subset of the described logic functions.

When a packet arrives from the PCI bus (we assume that segmentation/reassembly, or packet concatenation, if necessary, occurs in higher layers, typically in the operating system of the attached PC), it is stored into an input FIFO buffer. This buffer is needed to separate the activity of the PCI bus from the activity of the on-board logic, which are not synchronous. Every packet is formatted in a fixed size protocol data unit (PDU), which contains the payload bits [service data unit (SDU)], and a fan-out set, which contains packet destination information. Since four wavelengths are used in this first prototype, the fan-out set is composed of 4 bits; a bit set to 1 means that the packet must be transmitted on the corresponding wavelength. Eight FIFO queues store packets waiting for ring access, four unicast queues, and four multicast queues. The chosen number of queues stems from our previous studies in [16]. A reference fan-out set is associated with each queue, and a simple criterion based on the minimum Hamming distance is used to build a lookup table which associates all possible multicast fan-out sets with one of the eight destination queues. The reference fan-out sets for the eight queues are shown at queue-heads in Fig. 6. For example, queue Q6 stores packets whose fan-out sets are at minimum distance from the fan-out set comprising destinations 2 and 4. For each packet entries in the queues comprise a pointer to the SDU and the corresponding fan-out set. The fan-out set of the HOL packet can be a residual, since a fan-out-splitting service policy [17] is implemented according to which destinations in the fan-out set may be reached with more than one transmission.

The length of each queue is stored in a special register file $(L0, \ldots, L7)$. On the rising edge of the slot synchronization signal, the channel state is acquired by the λ-monitor. In Fig. 6, an available wavelength is coded by a logic "1", in this example,

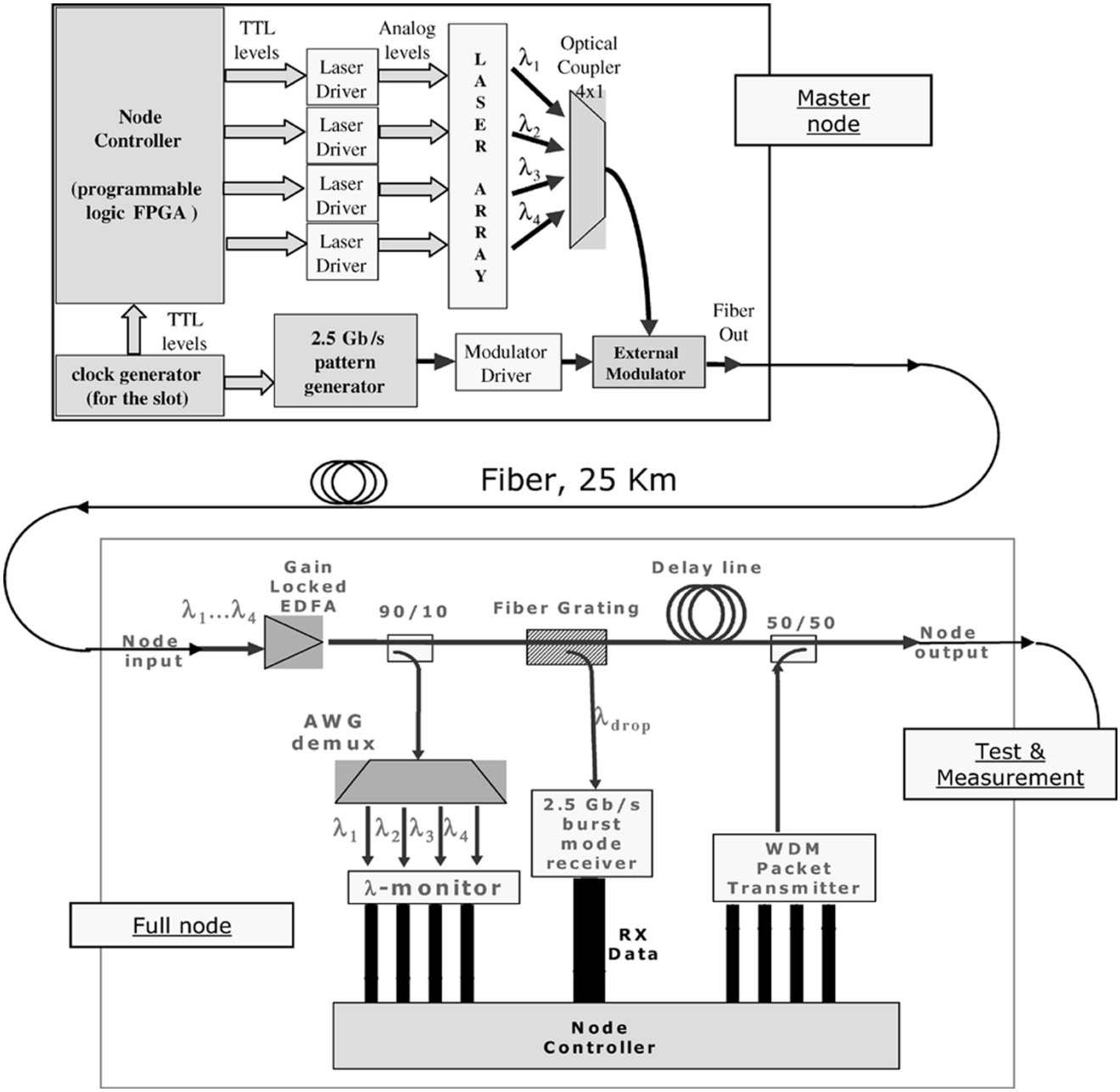


Fig. 5. Topology of the RingO testbed.

λ_2 and λ_4 are not available. A bitwise AND operation is computed between the channel state vector and the residual fan-out set vector of the HOL packet of each queue. The result is loaded into a support vector. Some queues cannot be chosen for transmission (in the example Q1, Q3, Q6) because all the wavelengths of their fan-out set are not available. The next step is to find the queue with maximum length among queues having transmission possibility. The maximum length is found by a tournament algorithm. Transmission requests of the "winner" queue, in the example Q4, have to be served in accordance with the channel wavelength availability. Our *a posteriori* packet selection policy, thus selects, among HOL packets that can be transmitted to at least one destinations in the fan-out set, the packet belonging to the longest queue. It is demonstrated that this choice guarantees the maximum throughput when the inputs and outputs are not overloaded [13], i.e., when no node is transmitting nor receiving more than the capacity of one channel. In the example, the node controller enables laser 1 and laser 3 to transmit, and sends the PDU of the first packet in queue 4 to the external laser modulator. The last step is to refresh the queue content. In the example, the fan-out set of the first packet in queue 4 has to be changed from 1110 to 0100 because λ_2 is the only transmission request not served. When all transmission requests are served, the packet is removed from the queue-head.

The control logic takes about 370 ns to do all these operations. Hence, an optical delay line of about 75 m was placed in the node demonstrator between the point where λ-monitoring is performed and the point where the locally generated packets are inserted (see Fig. 3).

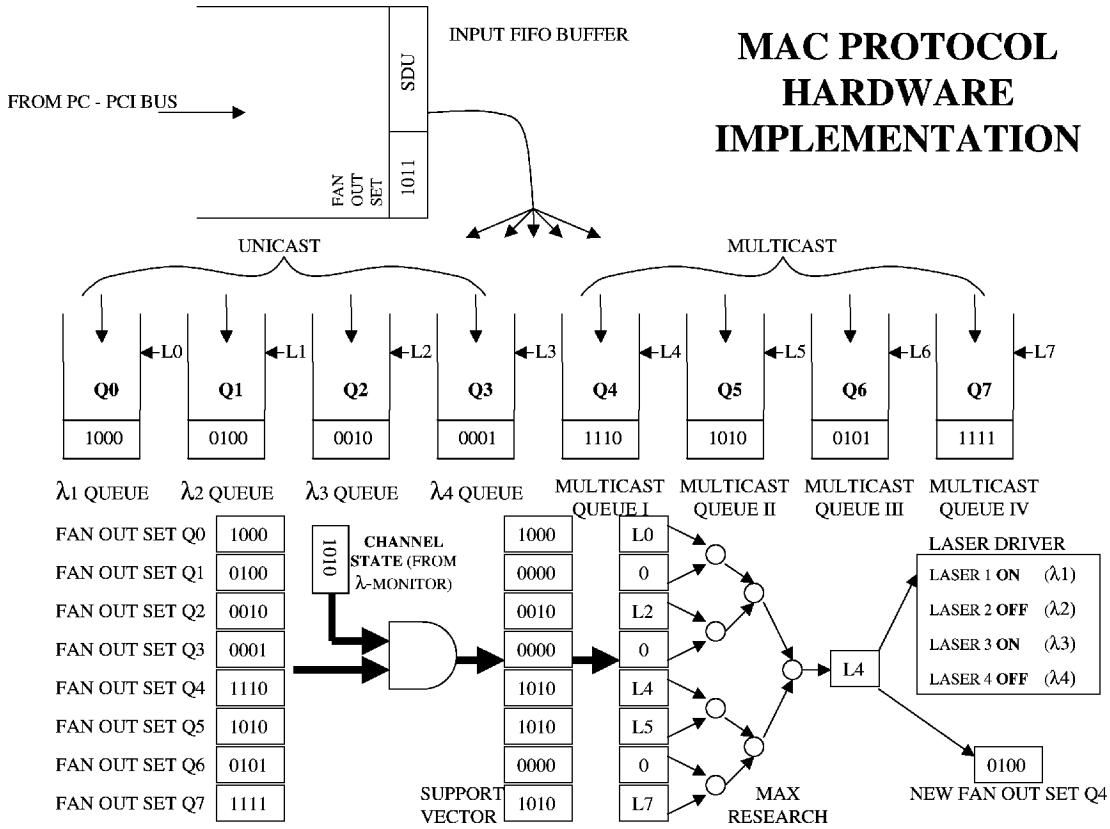


Fig. 6. Node controller logical structure.

V. RingO Scalable Architecture

An important limitation of the two previously presented RingO architectures is the fact that the number of nodes must not be greater than the number of wavelengths available on the ring, i.e., $N = W$. This largely impairs the scalability and the flexibility of our proposal. This observation leads us to the introduction of the third design for RingO nodes, which overcomes the above limitation by means of statistically time multiplexing packets to several destinations on the same wavelength channel (that is, the same wavelength can be used to transmit to different nodes). This can be achieved without changing the node's hardware in a significant way: the same basic node architecture, with fast tunable transmitter and a fixed receiver can be used, as discuss below.

A possible physical node architecture, with N nodes and W wavelengths, when $W < N$ are shown in Fig. 8. The major difference of this new design is the separation between resources devoted to transmission and resources devoted to reception. Transmitted packets traverse the ring a first time, are switched to the reception path, and then received during a second ring traversal. The transmission/reception separation can be obtained in wavelength (using different wavelength bands), in time (using different time frames), or in space (using two different fibers). We pick here the third option because it is easier to implement. This means that two physical fiber rings are used (see Fig. 7): packet transmissions occur on the first ring and receptions occur on the second ring. At some point, the two rings are interrupted and a connection between the transmission ring and the reception ring is done. This means that the ring is indeed transformed into two busses or into a folded bus, with

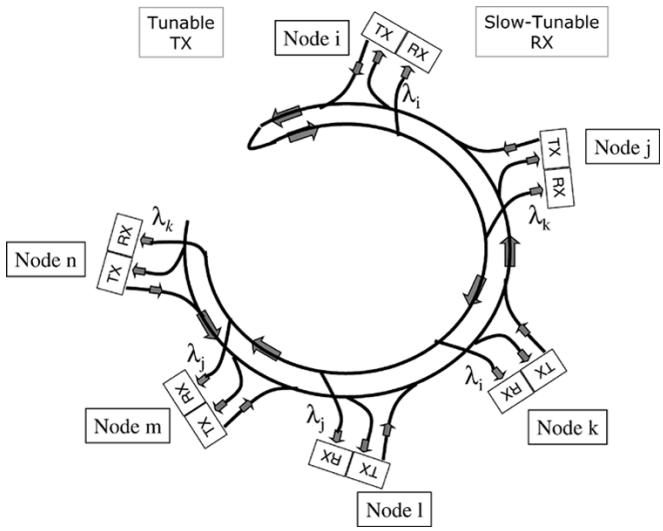


Fig. 7. Scalable architecture of the RingO network: two fiber rings topology.

significant advantages from the optical transmission viewpoint. Note that also in previous node designs the ring was broken into a set of staggered busses, one per wavelength, terminating at different receivers (each receiver terminates one wavelength). A given node must not drop from the ring the packets carried on its own receiver wavelength, and should select them (possibly in the electronic domain) according to a destination address.

The architecture of Fig. 8 requires extra optical capacity in the network, but no increase in the node complexity, nor in the capacity of transmitter and receivers, and of the data path toward applications. A negative effect of this transmission/reception separation is the loss of the space reuse capability typical of

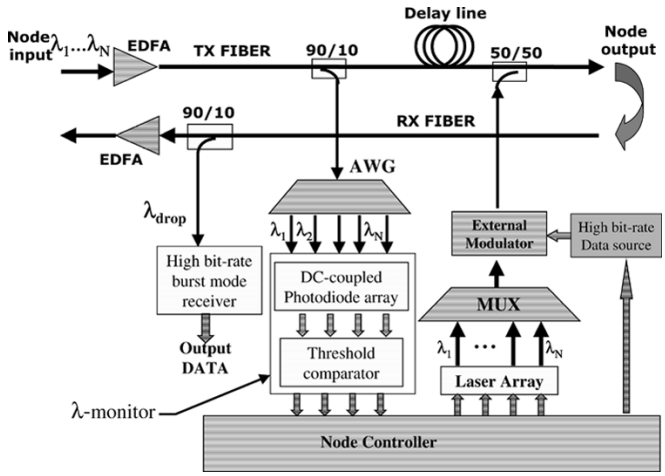


Fig. 8. Third structure of RingO nodes based on two fiber rings.

ring topologies. The two previous RingO architectures did not exploit space reuse on WDM channels due to the presence of a single receiver per channel. Space reuse becomes instead possible with multiple receivers per WDM channel. Space reuse can bring a significant throughput gain, which depends on the traffic distribution, it is around 100% in uniform traffic with a large number of nodes, less for hot-spot client/server traffic, but more for highly localized traffic.

The loss of the space reuse opportunity is the price that has to be paid with multiple receivers per channel if no optical switching in the data path is introduced. Indeed, another possible RingO architecture is currently under investigation, which keeps the single-ring topology of Fig. 1, and selectively drops packets at a receiver depending on destination addresses, allowing space reuse. To this end, at least two significant extra features should be added to the architecture of Fig. 8: an optical packet header, to carry the information on the packet destination, and a fast optical switching functionality to select packets to be dropped. Such an approach needs much more careful design of the physical layer. Moreover, fast optical switches are far from being mature components, so that this solution has been not further analyzed in this paper.

Another important feature of the architecture considered in this section is the fact that single-fault recovery comes at no extra cost, as discussed later in Section VI.

From a physical layer perspective, the architecture shown in Fig. 8 simplifies even more the node input–output optical path, which now consists only in (possibly sparse) EDFAs and optical splitter/combiners, while optical filters or add-drops are avoided. As discussed in Section III, the reduction of the complexity of optical components on the data path greatly reduces physical impairments such as PDL and self-filtering. Moreover, the two-fiber topology shown in Fig. 7 does not generate optical loops, thus avoiding potentially detrimental effects such as ASE noise recirculation, ring lasing, etc.

A. Allocation of Receivers to WDM Channels

From a network dimensioning perspective, since more than one node can receive on the same wavelength, a decision problem arises concerning the allocation of the different receivers to WDM channels. Good solutions to this problem should aim at equalizing the load on the different channels, that is the maximum load among all channels must be minimized.

It is straightforward to notice that the solution of the node allocation problem depends on the traffic on the network. Although this traffic matrix could be dynamically estimated, we suppose for simplicity that the traffic matrix is known.

The problem can be formalized in terms of integer linear programming (ILP), and it can be shown to be equivalent to the well-known problem of scheduling jobs on identical parallel machines, which falls in the class of NP-hard problems [18]. The problem states that given W wavelengths and N nodes, the receiver bandwidth load can be expressed as

$$l_i = \sum_{j=1}^{N} p_{ji} r_j \qquad \forall i,\ 1 \le i \le N$$

where r_j represents the transmission rate of node j and, p_{ji} its transmission probability to node i. A set of control variables x_{ik} can be defined, where

$$x_{ik} = \begin{cases} 1, & \text{iff node } i \text{ receives on wavelength } k \\ 0, & \text{otherwise} \end{cases}.$$

Receivers allocation is to be done trying to minimize $L_{\max}$, i.e., the load on the most loaded wavelength $L_{\max} = \max_k \sum_{i=1}^{N} l_i x_{ik}$. Thus, our problem formulation becomes

$$\text{Minimize } L_{\max}$$

subject to the following constraints:

$$L_{\max} \ge \sum_{i=1}^{N} l_i x_{ik} \qquad \forall k,\ 1 \le k \le W \tag{1}$$

$$\sum_{k=1}^{W} x_{ik} = 1 \qquad \forall i,\ 1 \le i \le N \tag{2}$$

$$x_{ik} \in \{0,1\} \qquad \forall i,\ 1 \le i \le N \quad \forall k,\ 1 \le k \le W. \tag{3}$$

Equation (1) ensures that no wavelength has a load larger than $L_{\max}$. Equation (2) ensures that each receiver must be allocated to only one wavelength.

Performance results are plotted in Fig. 9, where a scenario with 16 nodes and 4 wavelengths (four for each fiber ring, since we obtain transmission/reception separation using separated fiber rings) was simulated. In this simple scenario, two nodes named servers transmit at high load, equal to the capacity of one wavelength per server, with equal probability to the remaining 14 nodes, called clients. Client nodes transmit only to servers at a lower rate, equal to 1/14 of the channel capacity. Hence, the input and output load for all servers and for all clients are the same.

In Fig. 9, we show the throughput versus input load (both normalized to the available network capacity) for three different modes of allocating nodes to wavelengths. In particular, we compare the optimal receiver allocation obtained with the ILP

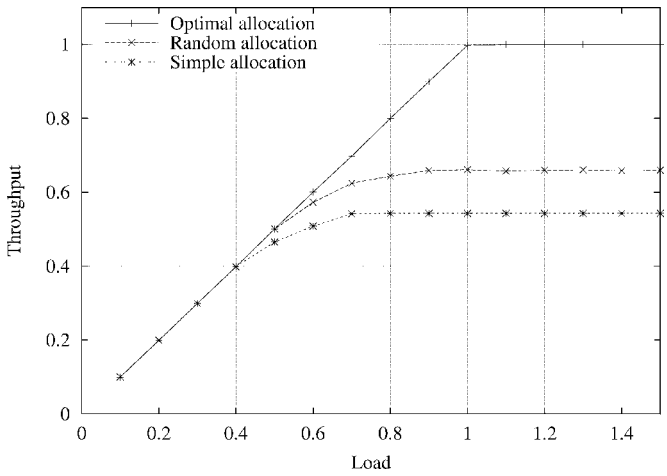


Fig. 9. Normalized network throughput versus input load for three different allocations of node receivers to WDM channels.

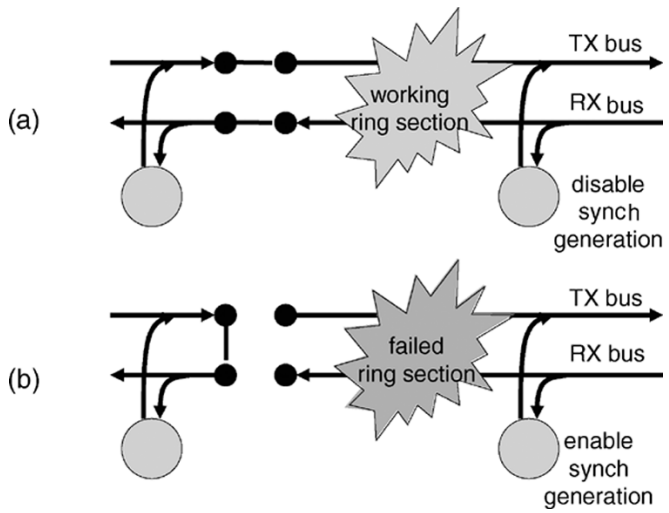


Fig. 10. Logical functionalities (a) for normal nodes and (b) for the nodes around the bus folding point (or the faulty section).

model described above with two other allocations. In the first one, called *random allocation*, each node is randomly allocated to one of the available wavelengths. In the second one, called *simple allocation*, we force that the number of allocated nodes on each wavelength is the same. We can observe that a nonoptimal solution to the allocation problem may lead to significant reductions of the total network throughput.

The complexity of the optimal solution may be too large. We sketch a simple but effective heuristic (for the scenario of Fig. 9 it provides the same solution of the ILP model) that solves the problem of receiver allocation with a low complexity [it can be shown to be $O(NW\log(W))$]. The algorithm follows the next three steps.

Step 1) Order all receiver loads as a nonincreasing sequence.
Step 2) Allocate the first receiver of the sequence on the least loaded wavelength, and delete it from the sequence.
Step 3) If the sequence is empty, then EXIT; else GOTO Step 2).

This algorithm is known, in operational research, as longest processing time (LPT) [18].

Despite the fact that the traffic matrix upon which the receiver allocation is chosen must be known *a priori*, it can show variations over time, i.e., it can behave as a dynamic matrix. In this case, it may be worthwhile to reallocate receivers dynamically in order to keep the network in an optimal operation point. One elegant way of achieving this result is to introduce (slow) tunability in node receivers. This tunability does not need to be fast, and does not need to track packet-by-packet variations. Low-cost devices available today (e.g., mechanical or thermo-optic filters) can be suitable to implement this slow receiver tunability feature. It is out of the scope of this paper to deal with problems concerning reconfiguration issues, but it should be clear that a tradeoff arises between keeping the receiver allocation well matched to the dynamic traffic matrix to optimize performance, and throughput losses due to blackouts when receivers are tuning.

VI. Fault Recovery

The node architecture described in the previous section has interesting fault recovery properties. The capability of recovering faults is considered an essential feature in all high-speed networks. Recovering from single faults requires biconnected topologies and bidirectional rings are the simplest such topologies. Ring-based networks typically require two counterrotating fiber rings to be able to protect single faults.

In the design of Fig. 7, we already have two counterrotating fiber rings, and single-fault protection can be provided by logically moving the folding point between the transmission bus and the reception bus to just before the fault (which can be either a fiber cut or a node failure) on the transmission bus. Although we did not discuss synchronization issues in this paper, around the folding point between two busses, and more precisely at the beginning of the transmission bus, the slot synchronization information must be injected in all wavelength channels. Fault protection implies equipping all nodes with this capability, and enabling it only at the first node of the transmission bus. Fig. 10 logically depicts the switching and synch signal generation capabilities that should be available at all nodes to recover from single faults. Note that switching must not necessarily be very fast: a reasonable target for error recovery is the SONET/SDH 50 ms figure. We do not further discuss on this issue due to space limitations, but proper fault detection procedures, fault signaling protocols, and fault recovery algorithms must be identified and implemented. Although this architecture does not allow space reuse, we observe that, in presence of a fault, all current traffic can be rerouted in the restored topology (i.e., overloading is avoided); the same is in general not true when rings exploiting space reuse are to be protected.

In absence of faults, the position of the folding point can be selected according to straightforward algorithms, and the configuration of the nodes around the folding point are exactly the same as for the nodes around a faulty network section.

As previously noted, however, the nice property of the architecture depicted in Fig. 8 is the sharing of network resources between the multireceiver per wavelength feature, and the fault

recovery mechanisms. We also remark that no additional transceivers are required for fault protection, and that the amount of fiber for protection is minimal.

VII. Conclusion

Metropolitan area networks are an arena where researchers and network architects have the opportunity to speculate on the best utilization of optical technologies in the implementation of switching and control functions.

Our work was motivated by the trust that optical packet transmission, though not yet standardized and commercially available, may become in the medium term a promising alternative to the current approach of building WDM networks with a high degree of fast circuit-switching reconfigurability, but where packet switching is still completely handled at the electronic level. At the same time, we do not believe that all packet switching functions can be *completely* moved from the electrical to the photonic domain in a reliable way without fundamental improvements in optical components technology. A good compromise between the two domains (optical and electrical) is the major goal of the RingO project presented in this paper.

We have presented three alternative node designs, which exhibit several common features, but trace an evolutionary path toward a final design that can be engineered in a successful and cost-effective manner. Several important issues (e.g., signaling for fault recovery, synchronization) were not discussed in this paper due to space limitations.

An interesting contribution of the RingO architecture was the definition of the access protocol, which builds upon previous experiences in scheduling packets in input-queued switching architectures, and offers good performance at complexities that are compatible with available technologies, as proved in our lab experiments.

References

[1] R. Ramaswami and K. N. Sivarajan, *Optical Networks–A Practical Perspective*. San Mateo, CA: Morgan Kaufman, 1988.

[2] *ITU-T Recommendation G.872*.

[3] S. Yao, B. Mukherjee, and S. Dixit, "Advances in photonic packet switching: An overview," *IEEE Commun. Mag.*, vol. 38, pp. 84–94, Feb. 2000.

[4] A. Carena, M. Vaughn, R. Gaudino, M. Shell, and D. J. Blumenthal, "OPERA: An optical packet experimental routing architecture with label swapping capabilities," *IEEE/OSA J. Lightwave Technol.*, vol. 16, pp. 2135–2145, 1998.

[5] K. V. Shrikhande, I. M. White, D. Wonglumsom, S. M. Gemelos, M. S. Rogge, Y. Fukashiro, M. Avenarius, and L. G. Kazovsky, "HORNET: A packet-over-WDM multiple access metropolitan area ring network," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2004–2016, Oct. 2000.

[6] C. Guillemot, M. Renaud, P. Gambini, C. Janz, I. Andonovic, R. Bauknecht, B. Bostica, M. Burzio, F. Callegati, M. Casoni, D. Chiaroni, F. Clerot, S. L. Danielsen, F. Dorgeuille, A. Dupas, A. Franzen, P. B. Hansen, D. K. Hunter, and A. K. Kloch, "Transparent optical packet switching: The European ACTS KEOPS project approach," *IEEE/OSA J. Lightwave Technol.*, vol. 16, pp. 2117–2134, Dec. 1998.

[7] D. J. Blumenthal, B. E. Olsson, G. Rossi, T. E. Dimmick, L. Rau, M. Masanovic, O. Lavrova, R. Doshi, O. Jerphagnon, J. E. Bowers, V. Kaman, L. A. Coldren, and J. Barton, "All-optical label swapping networks and technologies," *IEEE/OSA J. Lightwave Technol.*, vol. 18, pp. 2058–2075, Dec. 2000.

[8] I. Chlamtac, A. Fumagalli, L. K. Kazovsky, and P. Poggiolini, "A multi-Gbit/s WDM optical packet network with physical ring topology and multi-subcarrier header encoding," in *Proc. Eur. Conf. Optical Communications*, Montreux, Switzerland, Sept. 1993, pp. 121–124.

[9] ——, "A contention/collision free WDM ring network for multi-Gbit/s packet switched communication," *J. High Speed Networks*, vol. 1, no. 4, pp. 1–19, Apr. 1995.

[10] M. Cerisola, T. K. Fong, R. T. Hofmeister, L. G. Kazovsky, C. L. Lu, P. Poggiolini, and D. J. M. Sabido IX, "CORD-A WDMA optical network: Subcarrier-based signaling and control scheme," *IEEE Photonics Technol. Lett.*, vol. 7, pp. 555–557, May 1995.

[11] W. Parkhurst, *Cisco Multicasting Routing & Switching*. New York: McGraw-Hill, 1999.

[12] A. Bianco, E. Di Stefano, A. Fumagalli, E. Leonardi, and F. Neri, "*A posteriori* access strategies in all-optical slotted rings," in *Proc. IEEE GLOBECOM*, Sydney, Australia, Nov. 1998, pp. 300–306.

[13] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," *IEEE Trans. Commun.*, vol. 47, pp. 1260–1267, Aug. 1999.

[14] M. A. Marsan, A. Bianco, E. Leonardi, M. Meo, and F. Neri, "MAC protocols and fairness control in WDM multi-rings with tunable transmitters and fixed receivers," *J. Lightwave Technol.*, vol. 14, pp. 1230–1244, June 1996.

[15] A. Carena, V. Ferrrero, R. Gaudino, V. De Feo, F. Neri, and P. Poggiolini, "RingO: A demonstrator of WDM optical packet network on a ring topology," in *Proc. IFIP Optical Network Design and Modeling Conference ONDM 2002*, Turin, Italy, Feb. 2002, pp. 183–197.

[16] A. Bianco, P. Giaccone, E. Leonardi, F. Neri, and C. Piglione, "On the number of input queues to efficiently support multicast traffic in input queued switches," in *Proc. IEEE Workshop High Performance Switching Routing*, Turin, Italy, June 2003, pp. 111–116.

[17] R. Ahuja, B. Prabhakar, and N. McKeown, "Multicast scheduling for input-queued switches," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 855–866, June 1997.

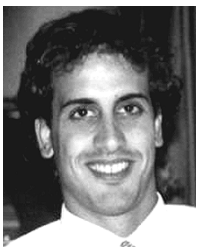
[18] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

Andrea Carena (M'98) was born in Carmagnola, Torino, Italy, on October 7, 1970. He received the Laurea degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in electrical engineering (optical communications) from Politecnico di Torino, Torino, Italy, in 1995 and 1999, respectively.

In 1998, he spent a year as a Visiting Researcher, first, in the Optical Communications and Photonic Network (OCPN) Group, Georgia Institute of Technology, Atlanta, and then at the University of California at Santa Barbara working in the realization of OPERA, an optical label swapping network testbed. He has collaborated in the development and implementation of OptSim, an optical transmission system simulator. He is currently an Assistant Professor in the Optical Communication Group, Politecnico di Torino. He has coauthored more than 40 papers in the field of optical fiber transmission. His research interests are in the field of new modulation formats, long-haul DWDM systems, fiber nonlinearity, performance analysis, and computer simulation of lightwave transmission systems.

Vito De Feo (S'03) was born in Salerno, Italy, in 1972. He received the Laurea degree in electronic engineering (*summa cum laude*) from Politecnico di Torino, Torino, Italy, in 2001 (thesis on the electronic control logic of an all-optical network). In 2001, he began working toward the Ph.D. degree in the Optical Communications Group and in the Telecommunication Network Group, Department of Electronics, Politecnico di Torino. He is currently a Visiting Ph.D. Student working toward the Ph.D. degree in the Optical Communication and Optical Network Group, Department of Electrical Engineering, Stanford University, Stanford, CA.

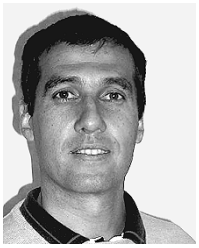
In 1997, he was a Visiting Student at Trinity College, Dublin, Ireland. His interests involve experimental demonstration of optical networks, subsystems for next-generation all-optical networks, and scheduling in optical network.



Jorge M. Finochietto (S'99) was born in Buenos Aires, Argentina, in 1978. He received the degree in electronics engineering from the Universidad Nacional de Mar del Plata, Mar del Plata, Argentina, in 2000. Since 2002, he has been working toward the Ph.D. degree in the Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy.

From 2000 to 2001, he was with the Engineering Group of Techtel, Buenos Aries, Argentina, as a South American Network Operator, in the areas of routing performance, quality-of-service (QoS), and ATM. His research interests include the design of all-optical networks and switch architectures.

Roberto Gaudino (M'98) is currently an Assistant Professor in the Optical Communications Group, Politecnico di Torino, Torino, Italy, where he works on several research topics related to optical communications. His main research interest is in the metro and long-haul DWDM systems, fiber nonlinearity, modeling of optical communication systems, and on the experimental implementation of optical networks. Currently, he is investigating new optical modulation formats, such as polarization or phase modulation, and packet switched optical networks. In 1997, he spent one year as a Visiting Researcher in the Optical Communication and Photonic Network (OCPN) Group, Georgia Institute of Technology, Atlanta, where he worked in the realization of the MOSAIC optical network testbed. From 1998, for two years, he has been with the team that coordinates the development of the commercial optical system simulation software OptSim. He is author or coauthor of more than 60 papers in the field of optical fiber transmission and optical networks. He is a consultant for several companies of the optical sector, and he is also involved in professional continuing education.



Fabio Neri (M'98) received the Dr.Ing. and Ph.D. degrees in electrical engineering from Politecnico di Torino, Torino, Italy, in 1981 and 1987, respectively.

He is a Full Professor in the Electronics Department, Politecnico di Torino. His teaching includes graduate-level courses on computer communication networks and on the performance evaluation of telecommunication systems. He leads a research group on optical networks at Politecnico di Torino. He has coauthored over 100 papers published in international journals and presented in leading international conferences. His research interests are in the fields of performance evaluation of communication networks, high-speed and all-optical networks, packet switching architectures, discrete event simulation, and queueing theory.

Dr. Neri was General Co-Chair of the 2001 IEEE Local and Metropolitan Area Networks (IEEE LANMAN) Workshop and General Chair of the 2002 IFIP Working Conference on Optical Network Design and Modeling (ONDM).

Chiara Piglione (S'03) was born in Racconigi, Italy, in 1977. She received the degree in telecommunication engineering (*summa cum laude*) from Politecnico di Torino, Torino, Italy, in July 2002. In 2003, she began working toward the Ph.D. degree in the Electronics Department, Politecnico di Torino. She is currently a Visiting Ph.D. Student working toward the Ph.D. degree in the Department of Electrical Engineering, Arizona State University, Tempe, AZ.

From October 2002 to December 2002, she had a research contract with the Italian National Inter-University Consortium for Telecommunications (CNIT). Her research interests include the study of multicast traffic, input queued switches, and all-optical networks.

Pierluigi Poggiolini (S'90–M'93) was born in Torino, Italy, in 1963. He received the M.S. (*cum laude*) and the Ph.D. degrees from Politecnico di Torino, Torino, Italy, in 1988 and 1993, respectively.

From 1990 to 1992, he was a Visiting Scholar with the Optical Communications Research Laboratory, Stanford University, Stanford, CA, where he mainly worked on the STARNET optical network research project. In 1994 and 1995, he was again at Stanford University, as a Postdoctoral Fellow, working on the ARPA-funded CORD all-optical packet network project. Since 1998, he has been an Associate Professor at Politecnico di Torino. He is the Coordinator of the Optical Communications Group, Politecnico di Torino. He has coauthored over one hundred papers in leading journals and conferences. His research interests include new modulation formats for optical transmission, optical packet networks, nonlinear fiber effects and modeling, and simulation of optical communications systems.

Dr. Poggiolini was awarded the International Italgas Prize for Scientific Research and Technological Innovation in 1998.