Experimental WDM Packet Networks for Metro Applications: the RingO and Wonder Projects

A. Bianciotto, A. Carena, V. De Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione and P. Poggiolini
PhotonLab and LIPAR Advanced Laboratories
Dipartimento di Elettronica, Politecnico di Torino
C.so Duca degli Abruzzi 24, 10129, Torino, Italy
E-mail: [*lastname*]@mail.tlc.polito.it Tel. +39.011.5644172 Fax +39.011.5644099

Abstract

This paper presents the activities carried out at Politecnico di Torino, Italy, on an advanced ring-based WDM optical packet network suitable for a high-capacity metro environment. We present two alternative architectural designs, and elaborate on the effectiveness of optical with respect to electronic technologies, trying to identify an optimal mix of the two technologies [1]. The first presented architecture refers to a research project called RingO, which ended in 2002, while the second one refers to the current ongoing project Wonder. Both projects are funded by the Italian Ministry for Education and Research (MIUR). 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. The main contribution of the 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.

I. Introduction

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 optical and electronic technologies, and new networking paradigms that better suit the constraints dictated by available photonic components and subsystems.

Our research groups at Politecnico di Torino, Italy, have designed and prototyped network architectures for metro applications, taking an approach based 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 and design of two research projects focused on optical network for metro application. The first presented architecture refers to a research project called RingO (Ring Optical) [2], which ended in 2002, while the second one refers to

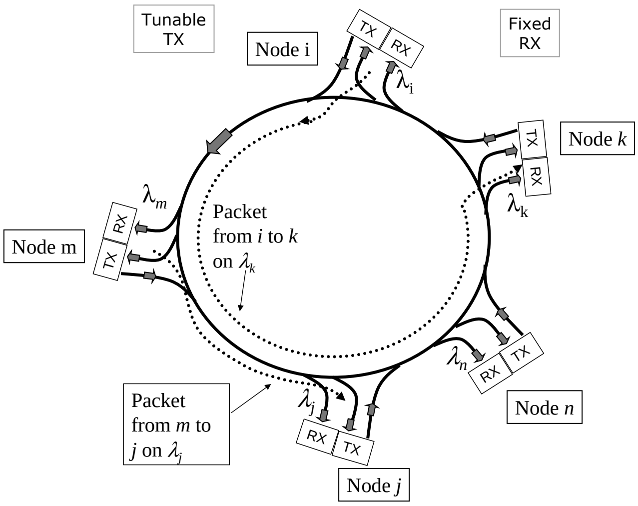

Fig. 1. Architecture of the RingO network.

the current ongoing project Wonder (WDM Optical Network DEmontrator over Rings). Both projects are funded by the Italian Ministry for Education and Research (MIUR) and involve a consortium of Italian Universities coordinated by the Optical Communication and Network groups 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. Related network architectures are described in [3],[4], [5] and [6].

II. RingO Architecture

The general architecture of the RingO network is illustrated in Fig. 1. 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 wavelenghts;
- packets have fixed length corresponding to one time slot: the packet format adaptation, including segmentation and reassembly, is left to higher (electronic) layers of the node protocol and 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;
- each node is able to check the busy/free state of all wavelengths (a feature called λ-monitoring) on a slot-by-slot basis, and avoids collisions and contentions by electronically queuing input packets, and by accessing channels using a suitable access protocol, as discussed in Subsection 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 their 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.

The proposed architecture combines in an efficient manner optical and electronic technologies: the aggregate bandwidth is handled in the photonic domain by working on a wavelength granularity, while packet queueing, Medium Access Control (MAC) protocol, and statistical time multiplexing are handled in the electronic domain at the speed of a single data channel.

Our architecture does not require any advanced optical component, such as fast optical switches or wavelength converters (see [7]). 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.

A. Node Structure

The structure of the node interfaces for the first RingO design is shown in Fig. 2. 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) Tapping 10% of the optical power and demultiplexing the WDM comb. Devices which have been used for this purpose are Arrayed Waveguide Grating (AWG) filters.
3) Monitoring the state of channels on each slot. This is done by sending the power on each fiber at the output of the AWG demultiplexer to a DC-coupled photodiode array.
4) Dropping wavelength λ_{drop} that must be received locally. This is done by an add-drop filter.
5) Burst-mode detection of the incoming data-stream on 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.
6) 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. 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 [8]; this choice was due to the difficulty in finding commercial fast-tunable lasers;
    - 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 [9]. In our situation, multicasting means to replicate the same packet on different wavelengths, possibly in the same time slot. With our structure, multicasting in a single time slot can be implemented without increasing electrical bandwidth requirements, since the “replication” of packets is obtained in the optical domain.

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 one of the advantages of our architectures with respect to current SONET/SDH circuit-switched solutions.

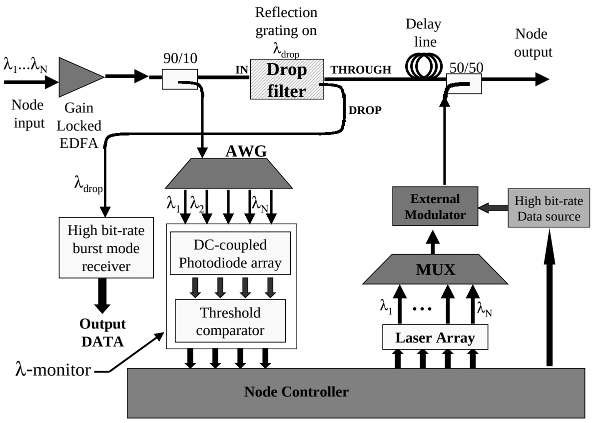

Fig. 2. Structure of RingO nodes, based on fiber-grating add-drops filters.

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 multi-channel 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 a-posteriori [10] packet selection strategy exploiting a Virtual Output Queuing (VOQ) structure. While standard single-channel protocols use a single FIFO (First In First Out) electrical queue, in multi-channel scenarios, where channels are associated with destination nodes, FIFO queuing performs poorly due to the Head-Of-Line (HOL) problem [11]. The HOL problem has been carefully studied, and can be solved using one of the VOQ [11] structures. The basic VOQ idea, applicable to the RingO architecture, consists in storing packets waiting for ring access in separated 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 not difficult to observe that our multi-channel 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. 1 and 2), 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 [11]. 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 Subsection III.

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

III. Experimental Testbed

The RingO network experimental testbed, shown in Fig. 3, was implemented in the PhotonLab [12] at Politecnico di Torino, and was based upon nodes having the structure

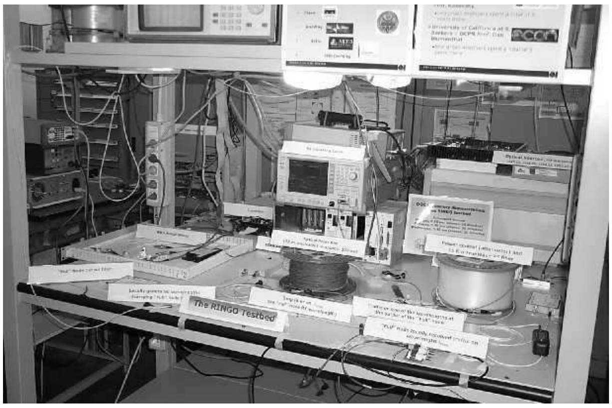

Fig. 3. The RingO testbed implemented in the PhotonLab at Politecnico di Torino.

shown in Fig. 2. The RingO testbed goals are:

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

The testbed developed in the RingO project is based upon two nodes, exchanging information on four different wavelengths, spaced at 200 GHz. The first node is used to generate random packet data traffic, while the second one implements all RingO protocol functions. We were 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. The node controller is based on a high-performance FPGA mounted on a custom-designed electronic board. The FPGA is an Altera APEX20KE600-3, with 600000 gates, 24320 flip-flops, 4 internal PLL's, 588 I/O and works at 133MHz.

The control logic takes about 370 ns to perform all the logic operations required by the MAC protocol. 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. 2). Details of the demonstrator were already shown in [2].

IV. RingO evolution: the Wonder Project

An important limitation of the previously presented RingO architectures is the fact that the number of nodes cannot 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 a second design, which is the basis of the currently ongoing project Wonder. This new architecture 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 discussed below.

A possible physical node architecture, with N nodes and W wavelengths, when $W < N$ is shown in Fig. 5. 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 a reception path, which used disjoint resources with respect to the transmission path, and then received during a second ring traversal. In fact, the ring is transformed in two busses, or into a folded bus by using two counter-rotating fibers, as shown in Fig. 4.

The negative effect of this transmission/reception separation is the loss of the space reuse capability typical of ring topologies. This can causes a significant throughput loss, which depends on the traffic distribution: it is around 50% in uniform traffic with a large number of nodes, less for hot-spot client/server traffic, but more for highly localized traffic.

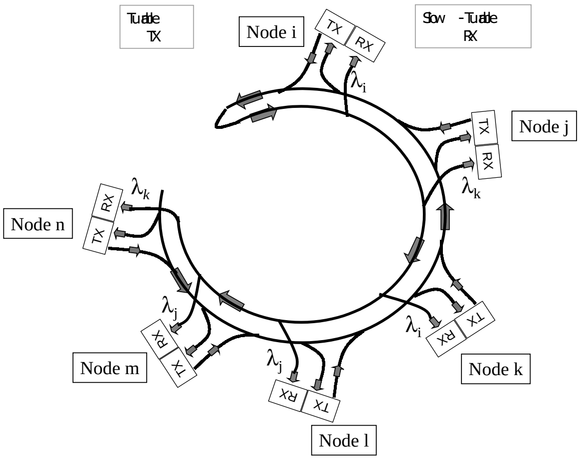

Fig. 4. Scalable architecture of the Wonder network: two fiber rings topology.

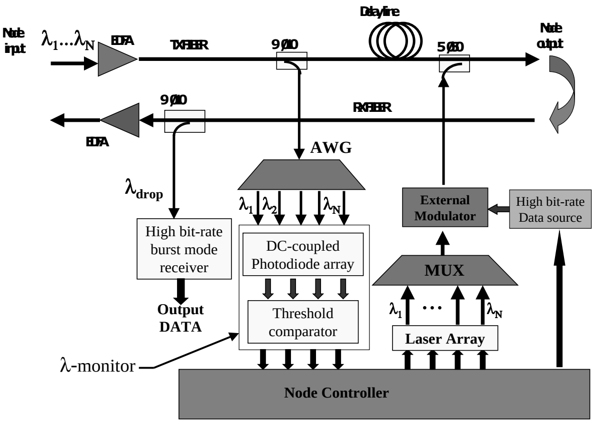

Fig. 5. Third structure of Wonder nodes, based on two fiber rings.

This is the price that has to be paid with multiple receivers per channel if no optical switching in the data path is introduced.

However, this new architecture can provide fault recovery at no extra cost. Single-fault protection can be achieved by logically moving the folding point between the transmission bus and reception bus to just before the fault (which can be a fiber cut or a node failure) on the transmission bus.

Moreover, from a pure physical point of view, the new node architecture shown in Fig. 5 has significant advantages to the previous architecture shown in Fig. 2. In fact, the presence of optical components (filters, splitters, etc) on the node “through” optical path is now really kept to a minimum. We now have only optical amplifiers and 1-2 splitters on the through path, a solution that let such impairments as self-filtering, Polarization Dependent Loss and Dispersion (PDL and PMD) to be minimized.

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 ILP *(Integer Linear Programming)*, and it can be shown to be equivalent to the well-known problem of scheduling jobs on identical parallel

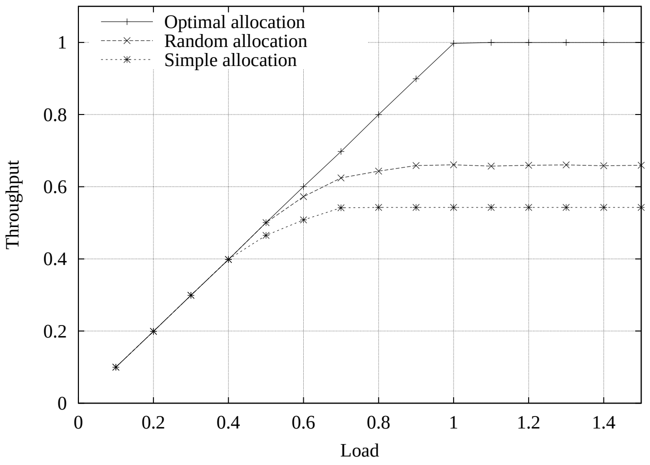

Fig. 6. Normalized network throughput versus input load for three different allocations of node receivers to WDM channels.

machines, which falls in the class of NP-hard problems [13]. The problem states that given W wavelengths and N nodes, the receiver bandwidth load can be expressed as:

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

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:

Minimize $L_{\max}$

subject to the following constraints:

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

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

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

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

Performance results are plotted in Fig. 6, where a scenario with 16 nodes and 4 wavelengths (4 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 $\frac{1}{14}$ of the channel capacity. Hence the input and output load for all servers and for all clients are the same.

In Fig. 6 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 allocation obtained with the algorithm described above with two other allocations. In the first one, called *random allocation*, each node is randomly allocated to one of the W wavelengths, this allocation is fixed ”a priori” at system startup. The second one, called *simple allocation*, is similar to the random allocation, but we also force that the number of allocated nodes on each wavelength is the same. We can observe that a non optimal solution to the allocation problem may lead to significant reductions of the total network throughput.

V. Conclusion

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 and Wonder projects presented in this paper.

References

[1] R. Ramaswami, and K.N. Sivarajan, *Optical Networks - A Practical Perspective*, Morgan Kaufman, 1988.

[2] 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”, *IFIP Optical Network Design and Modeling Conference ONDM 2002*, Turin, Italy, Feb. 2002.

[3] S. Yao, B. Mukherjee, and S. Dixit, “Advances in Photonic Packet Switching: an Overview”, *IEEE Communication Magazine*, Vol. 38, No. 2, 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 Journal of Lightwave Technology*, Vol. 16, No. 12, pp. 2135-2145, Dic. 1998.

[5] 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 Technology Letters*, Vol. 7, No. 5, pp. 555-557, May 1995.

[6] 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 Journal on Selected Areas in Communications*, Vol. 18, No. 10, pp. 2004-2016, Oct. 2000.

[7] 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 Journal of Lightwave Technology*, Vol. 16, No. 12, pp. 2117-2134, Dec. 1998.

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

[9] W. Parkhurst, *Cisco Multicasting Routing & Switching*, McGraw-Hill, 1999.

[10] A. Bianco, E. Di Stefano, A. Fumagalli, E. Leonardi, F. Neri, “A Posteriori Access Strategies in All-Optical Slotted Rings”, *IEEE GLOBECOM’98*, Sydney, Australia, November 1998.

[11] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, “Achieving 100% throughput in an input-queued switch”, *IEEE Transactions on Communications*, Vol. 47, No. 8, pp. 1260-1267, Aug. 1999.

[12] `http:\\www.photonlab.org`

[13] M. Pinedo, “Scheduling : theory, algorithms, and systems”, *Prentice Hall, 2002*