

WONDER: OVERVIEW OF A PACKET-SWITCHED MAN ARCHITECTURE

A. Bianciotto and R. Gaudino

OptCom Group, PhotonLab

Dipartimento di Elettronica, Politecnico di Torino

C.so Duca degli Abruzzi 24, 10129, Torino, Italy

E-mail: alessandro.bianciotto@polito.it - Tel. +39.011.2276301 - Fax +39.011.2276299

E-mail: roberto.gaudino@polito.it - Tel. +39.011.5644172 - Fax +39.011.5644099

Abstract: This paper presents the architecture of WONDER, an advanced ring-based WDM optical packet network designed for high capacity metro environments. The network prototype is currently being developed at PhotonLab in Torino, Italy, by several Italian research groups, thanks to a financing by the Italian Ministry of University and Research (MIUR). It represents an evolution over a similar ring based prototype, named "RingO" which was previously realized in 2001-2002. The WONDER network architecture elaborates on the effectiveness of optics with respect to electronics, trying to identify an optimal mix of the two technologies. We present network architecture, physical topology and node structure of the WONDER prototype, as well as its MAC protocol. The main contribution of this article is the identification 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.

Key words: Optical Network Testbeds; Packet Switched Networks; Advanced Optical Network Architectures.

1. 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 extensions lowers the impact of fiber transmission impairments such as fiber dispersion and nonlinearities. 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.

The OptCom and Network 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. This approach was already at the base of our previous “RingO” [2] network architecture, also developed in Turin at the end of 2002. The WONDER architecture was designed in order to find the best compromise between optical and electrical technologies. Similar solutions are currently under investigation in several other laboratories, see for instance [3]. To this end, the bulk of raw data transmission is kept in the optical domain, while network control functions are mostly implemented in the electronic domain.

In the following sections we introduce the rationale and design of the WONDER network prototype describing the physical topology, the node structure, the MAC protocol and the fault recovery mechanism, and the plans for the testbed currently under implementation in PhotonLab [1].

2. WONDER ARCHITECTURE

The WONDER network topology, see Fig. 1, is based on two WDM fiber rings connecting N nodes. One of the rings, called the “transmission” ring, is devoted to the transmission of data, while the other, called the “reception” ring, is devoted to the reception of data. The two rings are physically interconnected by a proper fiber shortcut that can be implemented by any node in the network. The shortcut is realized by closing a loop-back fiber between the transmission and reception rings at the output of a node, which is conventionally called the “folding” node (node i in Fig. 1). The node located at the other side of the shortcut turns out to be the first on the transmission ring and the last on the reception ring. This node is conventionally called the “master” node since, by preceding all the other nodes, it is devoted to the transmission of a suitable synchronization signal to the whole network. At any time, the network must have one and only one pair of active master and “folding” nodes in order to work properly. The resulting topology can also be viewed as a folded bus on which each node has two connections in two distinct points.

The transmission of packets is time-slotted and synchronized on all wavelengths, and each packet has a fixed duration corresponding to one time slot. In our current implementation, each slot is $1 \mu s$ long, and carries a Gigabit Ethernet bit stream at 1.25 Gbit/s. Future evolutions at higher bit rates are already planned. The WONDER protocol works as follows: each node i is

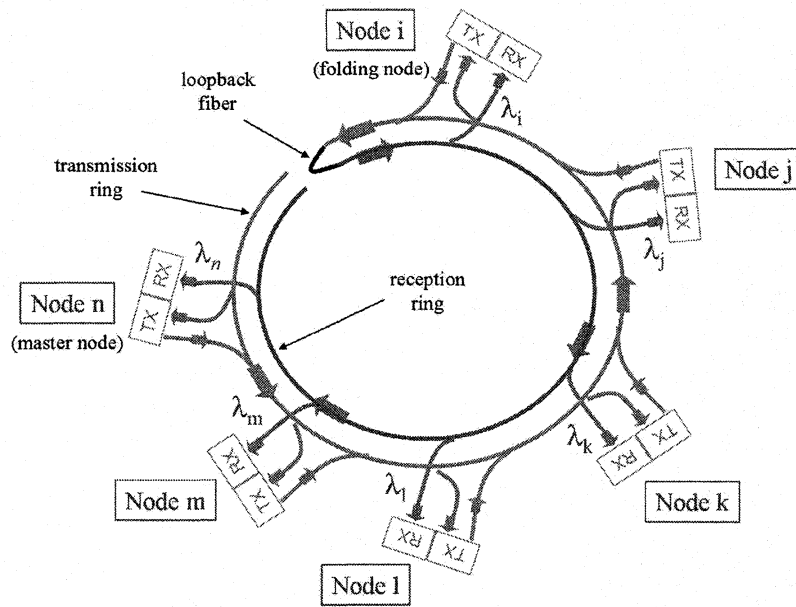


Figure 1. The two fiber rings topology at the basis of the WONDER architecture

equipped with an optical receiver tuned on a certain wavelength λ_i on the ITU-T grid. When a node has a packet to send to another node tuned on λ_i , it must set its transmitter on λ_i and wait for the first available “empty” slot on that wavelength to transmit the packet (see following Sect. 2.2). The WONDER architecture eases some problems of the previous Ringo architecture [2], and in particular:

- Two or more nodes can share the same wavelength in reception, thus allowing a fine granularity in the allocation of the available bandwidth. Thus, a label must be associated to the optical packet to indicate its destination among the various nodes sharing the same wavelength. There are many ways to implement this kind of labeling, both in the electrical and in the optical domain. The pros and cons of each method have been evaluated, but in the end the simpler option has been chosen, consisting in the addition of an apposite field in the packet’s header and a totally electronic recognition.
- Global synchronization is eased by the presence of the head-of-line master node, which sends to all other nodes a “timing” signal containing both bit and slot timing on a dedicated wavelength λ_t .
- The network is able to recover from any single fiber failure between two nodes by simply rearranging them to be “master” and “folding”, i.e., by

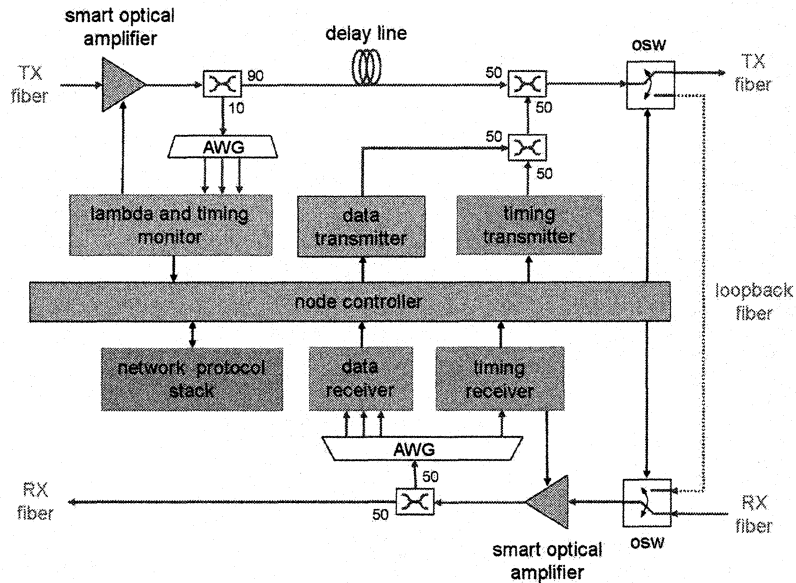


Figure 2. Structure of WONDER nodes.

dynamically re-arranging the fiber shortcut shown in Fig.1. Each node can sense a failure by monitoring the timing signal. If this signal is not sensed on the preceding fiber on the transmission ring, an “up-hill” failure has occurred, and the node must negotiate to become the network master. If the timing signal is not sensed on the preceding fiber on the reception ring, a “down-hill” failure occurred on this fiber, and the node must become the “folding” node. In general, a node can be either in a “master”, “folding” or generic “through” state: the node controller’s logic can use the information about failures to switch among these three states.

2.1 Node structure

The general structure of a WONDER node is shown in Fig. 2. It is logically subdivided into a number of subsystems, each one implementing a subset of the node’s functionalities. In the next subsections we’ll describe the functions, interfaces and the proposed experimental setup for each subsystem.

2.1.1 Smart optical amplifiers. In a metropolitan environment, the separation between two nodes can vary from a few kilometers to some tens kilometers, so the amount of amplification required is unknown, and has to be adapted case by case in order to have fixed optical power per channel at

the output of the amplifiers, thus requiring an automatic “static” gain setting technique. Another kind of problem arises because of the bursty nature of the traffic in the network [5]. In general, when a packet enters an EDFA it sees a dynamic gain that depends on the busy/free state in the preceding slots [4]. The gain transients dynamics can seriously impact the BER performances of the optical receivers, so adequate gain transients suppression techniques are mandatory to guarantee the proper operation of the network. The optical amplifiers in the WONDER network are thus termed “smart” because they must provide both “static” nominal gain setting and fast gain transients suppression features. We envision the use of the timing signal, which is continuous and always on, as a monitoring signal for gain settings. From an implementation point of view, two alternatives are available:

1. Use of EDFAs with a suitable control electronics which varies the pump power injected in the doped fiber as a function of the power sensed on the timing signal [6].
2. Use of Linear Optical Amplifiers (LOAs) [7], which are gain-clamped semiconductor optical amplifiers whose nominal gain can be set by varying the current injected into the active region.

The EDFAs have a very low noise figure and high output power, while the LOAs are noisier and less powerful, but more compact and potentially less expensive. In the WONDER scenario, LOAs are the preferred option for economic reasons, and their performance is currently under investigation.

2.1.2 Lambda and timing monitor. In order to avoid collisions in the access to the transmission medium, each node must monitor the busy/free state on all wavelengths in each time slot. The monitoring extends also to the timing channel to extract slot timing, information about failures on the rings, and to provide optical power monitoring to the “smart” optical amplifiers. In our prototype, the “lambda and timing monitor” optically demultiplexes the WDM channels by an AWG, and then extract digital information on the busy/free state of channels and on eventual failures by means of an array of DC coupled photodiodes followed by threshold comparators.

2.1.3 Node controller. The node controller is at the hearth of the node’s operation and implements all the intelligence of the network. It is realized on an FPGA board exchanging data with all the other subsystems, taking care of the implementation of the MAC protocol, packets queueing, fault recovery mechanism, burst mode reception etc. We decided to base all our (electronic) architecture on an FPGA board due to the great increase in capabilities and reduction in costs that FPGA devices have shown in the last 2/3 years. The FPGA we will use for the testbed is able to handle several input/output data-stream running at up to 3.2 Gbit/s, with advanced Clock and Data Recovery

(CDR) capabilities, very high processing power, and ease of integration with other external devices. The FPGA interfaces with all the upper layers of the WONDER network protocol stack to enable data exchange between nodes by means of standard protocols like FTP, HTTP and SMTP and suitable middle-ware software residing on a workstation.

As an important feature, our architecture requires an electrical data path bandwidth, on the transmitter, receiver and node controller that is equal to a single channel data rate only. In fact, 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.

2.1.4 Data transmitter. A WONDER transmitter must be able to tune on any of the available wavelengths on a slot by slot basis. Conventional transmitters are very different because they work with a continuous flow of data and on a fixed wavelength, so a radically different design is required. We envisioned two options:

1. Use of an array of fixed wavelength DFB lasers on the ITU-T grid, on/off switched by a digital signal coming from the node controller on a slot by slot basis. The outputs of the lasers are coupled together, then externally modulated by an external modulator, which is driven by the data bits.
2. Use of a fast tunable [8] laser which can be tuned on any desired C-band wavelength in a few tenths of nanosecond by injecting a suitable current, followed by an external modulator.

In our prototype, the first option was chosen because of its component simplicity. The second option is today less feasible, but should be preferred as soon as fast tunable lasers become reliable and commercially available [8].

2.1.5 Timing transmitter. When a node becomes the network "master", it has to transmit a timing signal containing both bit and slot time references. These signals must be combined on a single wavelength to minimize costs, so we decided to try an approach where a low extinction ratio square wave with period equal to one slot ($1 \mu s$ in our implementation) is generated by directly modulating the laser with a small-signal current. The bit timing, set to 1.25 Gb/s, is then superimposed by means of an external modulator.

2.1.6 Timing receiver. The timing receiver must extract the bit clock frequency from the timing signal, as well as providing a digital signal indicating failures on the reception fiber. The first function is obtained by passing the electrical timing signal (recovered by a wide-band photodiode) into an high-pass filter with a cutoff frequency of approx. 1 GHz. The resulting signal is

then directly fed to the node controller FPGA CDR circuitry, where a clean bit clock is made available. A failure signal can be issued whenever the power on the received timing signal falls below a predefined threshold.

2.1.7 Data receiver. A WONDER data receiver must work in burst mode on a packet by packet basis, so it is quite different from conventional continuous mode receivers. The main issues in the realization of an high speed burst-mode receiver are the following:

1. The bursty nature of the received signal complicates the recovery of the phase of the clock signal needed to sample data bits at the optimal instant. The clock phase recovery must take at most some tenths of nanoseconds otherwise a significant part of the packet's duration would be wasted for synchronization. The CDRs that are present on-board on the selected FPGA can automatically perform very fast phase recovery whenever the clock nominal frequency is known. This is the main reason why we send a "master" bit clock signal through the network on the timing signal.
2. Conventional high speed optical receivers are usually AC coupled. Their cutoff frequency is typically set quite low (about 30 kHz) to avoid baseline wander. Anyway, in a burst mode receiver, the cutoff frequency must be set to higher values, to have a very fast AC transient at the beginning of a packet. This in turns requires a suitable line coding (we use Ethernet 8B/10B coding).
3. A suitable "training" sequence must precede the packet useful bits (header and payload) to make sure the CDR can perform phase recovery. This is achieved by using a proper sequence, like a Barker sequence, at the beginning of the bit stream. The selected FPGA has built-in functions to implement these synchronization features.

2.2 MAC protocol

Our architecture requires a suitable MAC protocol to allocate time slots to transmitters. From the MAC protocol design perspective, WONDER 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 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 FIFO queuing performs poorly due to the Head-Of-Line (HOL) problem, which has been

carefully studied in the literature, and can be solved using one of the VOQ [9] structures. The basic VOQ idea, applicable also to the WONDER architecture, consists in storing packets waiting for ring access in separated queues, each corresponding to a different destination, and to appropriately select the queue that gains access to the channels for each time slot.

3. 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 (at most) some degree of 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 WONDER project presented in this paper.

REFERENCES

- [1] <http://www.photonlab.org>
- [2] A. Carena, V. Ferrero, 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] 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.
- [4] A. Bononi, L. A. Rusch, "Doped-fiber amplifier dynamics: a system perspective", *IEEE Journal of Lightwave Technology*, Vol. 16, no. 5, pp. 945-956, May 1998.
- [5] L. Tancevski, A. Bononi and L. A. Rusch "Output Power and SNR Swings in Cascades of EDFA's for Circuit- and Packet-Switched Optical Networks", *IEEE Journal of Lightwave Technology*, Vol. 17, No. 5, pp. 733-742, May. 1999.
- [6] L. Tancevski, L. A. Rusch and A. Bononi "Gain Control in EDFA's by Pump Compensation", *IEEE Photonics Technology Letters*, Vol. 10, No. 9, pp. 1313-1315, Sep. 1998.
- [7] D. A. Francis, S. P. DiJaili and J. D. Walker "A single chip Linear Optical Amplifier", *OFC 2001, Anaheim, CA*, paper PD13, 2001.
- [8] Y. Fukashiro, K. Shrikhande et al. "Fast and fine wavelength tuning of a GCSR laser using a digitally controlled driver", *Optical Fiber Communication Technical Digest*, paper WM43, pp. 338-340, Baltimore, MD, March 2000.
- [9] 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.