SEVENTH FRAMEWORK PROGRAMME

Project Number: FP7-ICT-2007-1 216863

Project Title: Building the Future Optical Network in Europe (BONE)

CEC Deliverable Number: FP7-ICT-216863/RACTI/R/PU/D14.3

Contractual Date of Deliverable: 15/01/11

Actual Date of Delivery: 15/01/11

Title of Deliverable: D14.3: Final report on integration activities

Workpackage contributing to the Deliverable: WP14 : Virtual Center of Excellence on Optical Switching Systems (VCE S)

Nature of the Deliverable : R (Report)

Dissemination level of Deliverable : PU (Public)

Editors: RACTI / Kyriakos Vlachos
COM-DTU / Martin Nordal Petersen

Abstract:
This document is the final report converting the third year of activities in the VCE-WP14 Virtual Center of Excellence on Optical Switching Systems (VCE S). Being the final report on integration activities it includes the achievements from the last year (Y3).

Keyword list:
Wavelength Converter Usage Reduction, Switching for Network Recovery, Quality of Service in switches, New switching paradigms, Optical Clock Recovery, Wavelength Conversion by nonlinear effects, Optical Multicast Architecture, Switching and network reliability, benchmarking and cost analysis, OCDM encoders/decoders, 2R Regeneration, Optical flip-flops, Optical packet switching, Hybrid Switch Architectures, GMPLS optical switch nodes, Contention Resolution Schemes, Optical Buffering, OTDM time-slot switching, Multi-wavelength regeneration, Optical Cross Connects, Power Issues in Switching Systems.

Disclaimer

*The information, documentation and figures available in this deliverable, is written by the BONE (“*Building the Future Optical Network in Europe*) – project consortium under EC co-financing contract FP7-ICT-216863 and does not necessarily reflect the views of the European Commission*

Table of Contents

1. Executive Summary :

This document is the third and final deliverable from the work package “Virtual Center of Excellence on Optical Switching Systems (VCE-S)”.

This deliverable on “Final report on integration activities” collects the partner expertise and describe the final results and achievements of their planned joint activities in the framework of BONE project. Since the deliverable 2 (Y2), two of the JAs (JA9 and JA11) finished while the rest have run throughout year 3 (2010).

Catching up since year two, we brought attention to the low number of mobility actions that only amounted to 6 in 2009. This seems to have somewhat improved in the final year of the project.

It is the perception of the WP leaders that WP14 has been running very satisfactory with a high level of participation and activity, this accounts for year 3 as well. WP14 has integrated 39 partners and 119 researchers. The overall outcome of the three year BONE project is a much heightened level of collaboration within European Institutions as well as a general awareness and knowledge about research activities taking place throughout EU.

2. Introduction

Recent technology development has unlocked most of the fiber capacity making capacity an abundant resource. However, the majority of WDM (Wavelength Division Multiplexing) deployment has occurred in the form of point-to-point links with amplifiers in between if needed.
Optical WDM light paths are static and are seen as a scarce resource. Once set up, they remain in place, essentially forever. Therefore, switching is only to transforms the raw bit rates into useful bandwidth.

The prime objective of this deliverable is to provide the final results on collaboration within the BONE project WP14 on the topic of photonics in switching - a “Virtual Center of Excellence on Optical Switching Systems (VCE-S)”. Within these directions, VCE-S has crafted R&D directions and defined the position of photonic switching in future optical networks.
The questions that have been researched are how and where photonic switching is positioned in future Internet and the problems that must be solved to reach this goal. The incentive has been to use optical switching to increase spectrum efficiency in terms of switching and not in terms of only capacity. Towards this goal VCE-S has charted a list of key issues and prime research objectives. The actual implementation has been carried out during the project via a number of collaborative projects (JAs).

2.1 VCE-S Participants

The following table displays what partners have participated in WP14.

Participant	Participant
1- IBBT	29 - PoLIMI
2- TUW	31 - SSSUP
3- FPMs	33 - UNIMORE
4 - Fraunhofer	34 - UniRoma1
5 - TUB	35 - TELENOR
6 - UDE	36 - TUE
7 - UST-IKR	37 - IT
8 - COM	38 - AGH
9 - CTTC	39 - PUT
10 - TID	40 - HWDE
11 - UAM	41 - KTH
12 - UC3M	42 - BILKENT
13 - UPC	43 - UniRoma3
14 - UPCT	44 - ORC
15 - UPVLC	45 - UCAM
16 - UVIGO	46 - UCL
17 - FT	47 - UEssex
18 - GET	48 - USWAN
19 - AIT	49 - Ericsson
20 - ICCS-NTUA	50 - UNIBO

3. List of Joint Activities in VCE-S, WP14

JA#	JA title	JA-Leader	Participants	Status
JA1	Power-Cost-Effective Node Architecture for Multicast Light-Tree Routing in WDM Networks.	Gonzalo F.D. Carpio gmfernand@hotmail.com	**UC3M,** UPVLC	Ended
JA2	Feasible parallel schedulers for OBS/OPS nodes.	Pablo Pavón Mariño Pablo.Pavon@upct.es	**UPCT**, UVIGO, DEIS-UNIBO	Ended
JA3	Performance and complexity analysis of optical switching fabrics	Fabio Neri neri@polito.it.	**PoliTo**, UniBO, PoliMI, TUW	Moved to WP25
JA4	Performance of optical switching system architectures with shared wavelength converters.	Nail Akar akar@ee.bilkent.edu.tr	**Bilkent**, UNIBO, UniRoma1, UNIMORE, FT, KTH	Ended
JA5	Code-based optical nodes.	Gabriella Cincotti, cincotti@uniroma3.it	**UNIROMA3**, UNIMORE, UNIBO, UNIROMA1, GET, NICT	Ended
JA6	Repacking and rearranging algorithms for multi-plane banyan type switching fabrics.	Wojciech Kabacinski, wojciech.kabacinski@et.put.poznan.pl	**PUT**, POLIMI	Ended
JA7	Power Consumption and Supply of Individual Network Elements	Slavisa Aleksic slavisa.aleksic@tuwien.ac.at	**TUW**,PoliTo, UNIMORE, UoP, UPCT, BME, UPC, USWAN, ISCOM, FUB	Moved to the new WP21
JA8	All-optical switches utilizing microring resonators.	Adonis Bogris abogris@di.uoa.gr .	**UoA**, UC3M	Moved to WP25
JA9	All-optical label processing techniques for ultra-fast optical packet switches.	Piero Castoldi castoldi@sssup.it	**SSSUP**, UNIROMA3, NICT	Ended in Y2
JA10	Currently inactive / merged into JA12			
JA11	The Optical Switch Architecture with Recirculation Buffer and	Wojciech Kabaciński	**PUT**, RACTI	Ended in Y2

	Wavelength Conversion.	wojciech.kabacinski@et.put.poznan.pl		
JA12	Encompassing switch node impairments and capabilities in dynamic optical networks	Nicola Andriolli nick@sssup.it	**SSSUP**, DTU	Ended
JA13	Reliability of various optical switching technology	Rebecca Chandy rebeccapchandy@gmail.com	**Ericsson**, KTH	Moved to the new WP25
JA14	Photonic code label processors for ultrafast routing	Ian White ihw3@cam.ac.uk	**UCAM**, TUe, Uniroma3	Ended
JA15	Hardware efficient optoelectronic switch fabrics	Ian White ihw3@cam.ac.uk	**UCAM**, TUe, PUT	Moved to the new WP25
JA16	Low-crosstalk optical packet-switching architectures based on wavelength-switching and wavelength-sensitive devices	Achille Pattavina pattavina@elet.polimi.it	**POLIMI**, UPCT, POLITO, OVIGO	Ended
JA17	Novel Multi-granularity Optical Switching Node with Wavelength Management Pool Resources	Yabin Ye yabin.ye@huawei.com	**HUAWEI,** RACTI	Ended
JA18	Comparison of the synchronous/asynchronous operation paradigm in optical switches	Fabio Neri neri@polito.it.	**PoliTO,** RACTI	Ended

4. Technical Report on VCE-S Joint Activities

4.1 JA1 – Power-Cost-Effective Node Architecture for Multicast Light-Tree Routing in WDM Networks

Responsible partner: UC3M

Participants: UPVLC, UC3M

Description of the work carried out in the 3rd year:

During the third year the work has focused on improvements to the multicast architecture proposed in the first two years and on the possibilities of extension of the concept to MPLS networks. Intensive effort was put on modelling a realistic GMPLS/MPLS backbone and on the proposal and analysis of heuristics for multicast tree construction and aggregation.

Collaborative actions carried out:

- mobility actions:

- Jose A. Hernández visit to UPVLC for a joint seminar (one week)
- David Larrabeiti visit to UPVLC for a joint seminar (one week)
- Juan Carlos Guerri visit to UC3M for 2 joint seminars (two weeks)

- number and details of papers:

1 joint paper to be submitted to High Performance Switching and Routing conference in January 2011. “Tree aggregation for scalable multicast service delivery in G/MPLS networks”

Other 2 non-joint papers listed below.

Overall assessment of work carried out within the project duration:

The work carried out has focused on the design of a novel cost-effective multicast-capable optical cross connect (MC-OXC) node architecture that features both tap-and-continue and tap-and-binary-split functionality. This architecture (Fig.1 and Fig.2) intends to provide an interesting balance between simplicity, power efficiency and overall wavelength consumption with respect to models based on TaC (Tap and Continue) or SaD (Split-and-Delivery).

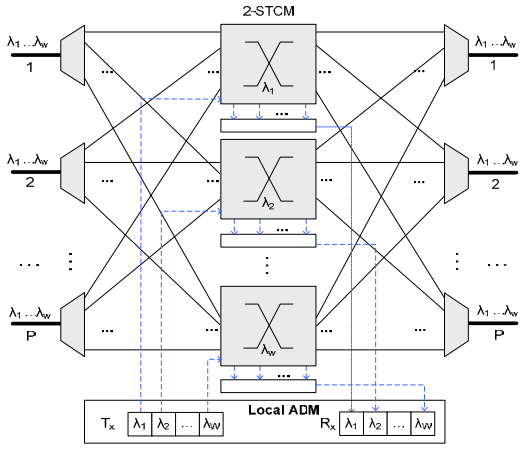
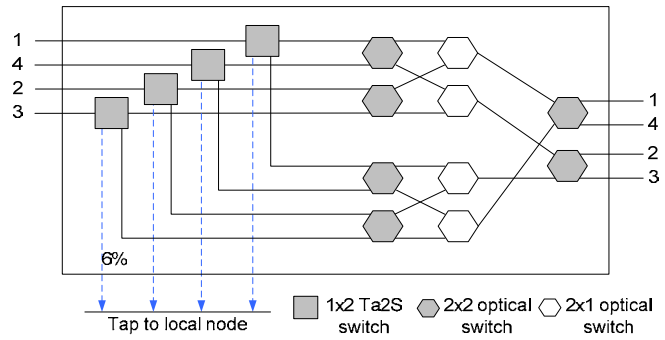
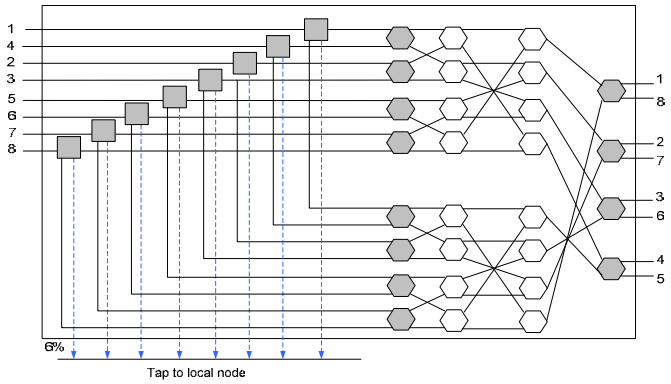


Fig. 1. General architecture of a P x P 2-STC MC-OXC node.



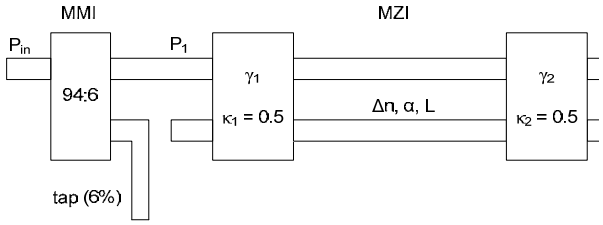
(a) 4 x 4 2-STCM



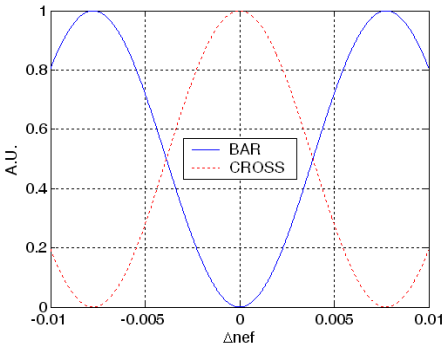
(b) 8 x 8 2-STCM

Fig. 2. Block diagram of the 2-STCM module for 4x4 and 8x8 switches

The main component of this node is a novel Tap-and-2-Split Switch (Ta2S). We proposed and analysed an implementation of this switch based on integrated optics (namely, MMI taps and MZI switches) (Fig.3) , and we characterized and compared it with other alternatives implemented with the same technology.



(a) Schematic of a Ta2S switch composed of a fixed splitter and an MZI.



(b) Bar and Cross Outputs of MZI vs Δn.

Fig. 3. Schematic of a Ta2S switch and output power response when Δ*n* modified.

Accounting the number of components used in the integrated implementation we analysed the scalability as compared to SaD and TaC modules. The study shows that, thanks to the presented Ta2S design, the 2STC node scales better in terms of number of components than the other alternatives.

The number of components (optical switches, Ta2S switches, tunable 2x2 MMI splitters and tap devices) in 2-STCM is much lower than the other proposals (Fig. 4) , which contributes to improve the power efficiency.

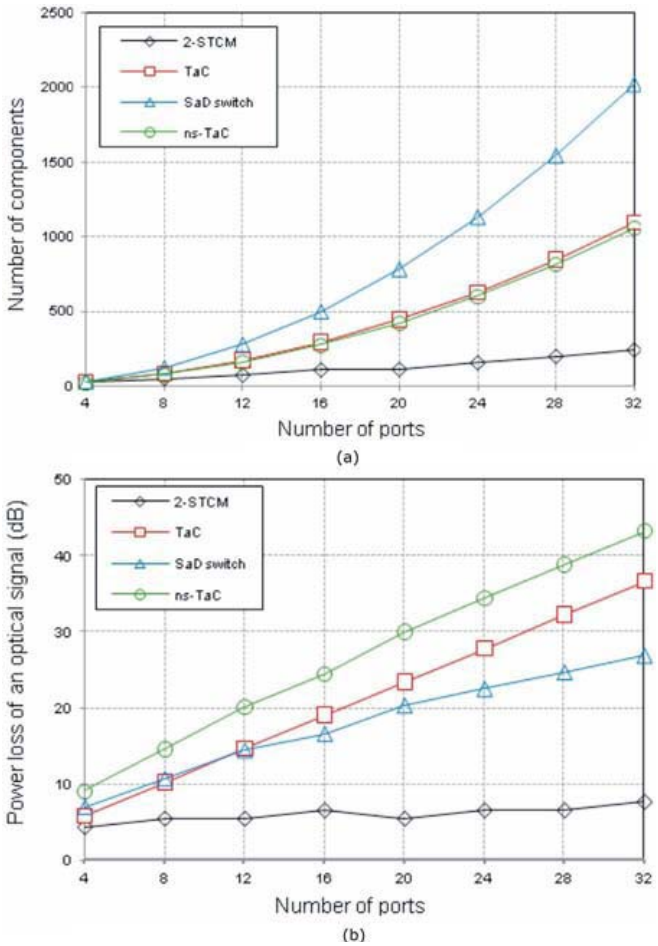


Fig. 4. Comparison of architectures. (a) Total number of components in 2-STCM, TaC and SaD modules. (b) Power loss (in dB) of a single optical signal for 2-STCM, TaC and SaD switch modules when a tap/drop-and-continue action is performed (average case).

Furthermore, it is more power efficient than the SaD design and requires less wavelengths than TaC thanks to the binary split capability (Fig. 5). On the other hand, the simulation results reveal that the 2-split condition does not add a significant additional wavelength consumption in usual network topologies with respect to SaD.

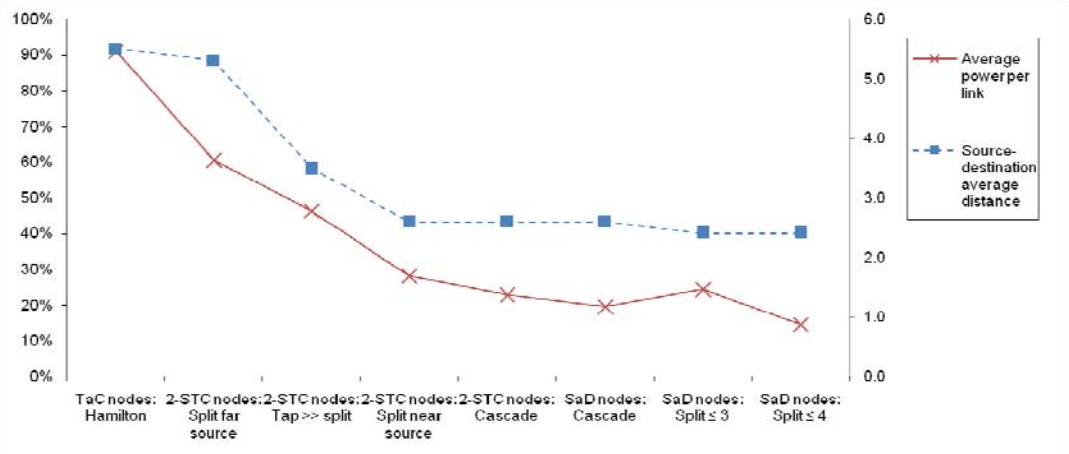


Fig. 5. Link Power and average distance (hops) of different architectures and network settings

The second part of the research of this JA corresponds to an analysis of scalability of the multicast solution by means of the aggregation of multicast *groups*, where the concept of group intends to cover the set of end-systems that make up the leaves of the multicast tree. In reality, the most likely higher layer setting for a "group" is the set of sites in a multicast/broadcast-enabled VPN, be it an Optical VPN or an MPLS VPLS. Through the aggregation procedure, multiple multicast groups are configured to share a single *multicast distribution tree* (MDT) a.k.a. *aggregation tree*. In this way, the number of trees configured in the network is significantly reduced, and consequently the multipoint resources in the node (or the amount of forwarding state information in the case of MPLS) gets also reduced. This improvement is achieved at the expense of bandwidth waste, because aggregation trees may deliver packets to end-nodes with no associated members. Despite this, we prove that the aggregation tree strategy is more efficient than both unicast-VPN and one-tree-per-VPN strategies if intelligent algorithms are in place.

Outcome of the joint research activity:

[P1] Gonzalo M. Fernández, David Larrabeiti, Carmen Vázquez, Pedro Contreras Lallana "Power-Cost-Effective Node Architecture for Light-Tree Routing in WDM Networks". New Orleans, IEEE Globecom 2008

[P2] Gonzalo M. Fernández, Carmen Vázquez, Pedro Contreras Lallana, and David Larrabeiti, "Tap-and-2-Split Switch Design Based on Integrated Optics for Light-Tree Routing in WDM Networks" JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 27, NO. 13, JULY 1, 2009

[P3] G. M. Fernández, J.A. de la Fuente, D. Larrabeiti, J.C. Guerri, T. Cinkler. "Tree aggregation for scalable multicast service delivery in G/MPLS networks" (to be submitted to HPSR2011 in January 2011)

4.2 JA2 – Feasible parallel schedulers for OBS/OPS nodes

Responsible partner: UPCT

Participants: UPCT, UVIGO, DEIS-UNIBO

Description of the work carried out in the 3rd year:

In the third year of the project, UPCT and DEIS/UNIBO collaborated in the emulation of an FPGA implementation of a large-scale version of the PI-OPS scheduler. Results support the feasibility of its implementation, suggesting that it is possible to build large-scale PI-OPS schedulers with response times below 10 microseconds. The results have been published in [Pav10]. During 2010 the UVIGO group completed his work on the design and performance evaluation of different scheduling algorithms for asynchronous (not-aligned) Optical Cell Switching Networks. The main results have been published in [Rod10].

Collaborative actions carried out:

- mobility actions: No mobility actions were completed. However, frequent interactions occurred in multi-conference calls.

- number and details of papers:

Joint publications (1 conference paper, 2 journal papers):

- [Pav09] (UPCT, UNIBO) Pablo Pavon-Mariño, Juan Veiga-Gontan, Alejandro Ortuño-Manzanera, Walter Cerroni, Joan Garcia-Haro, " PI-OBS: a Parallel Iterative Optical Burst Scheduler for OBS networks", IEEE International Workshop on High Performance Switching and Routing 2009 (HPSR 2009), Paris, June 2009.
- [Rod09] (UVIGO, UPCT, UNIBO) M. Rodelgo-Lacruz, P. Pavón-Mariño, F. J. González-Castaño, J. García-Haro, C. López-Bravo, J. Veiga-Gontán, and C. Raffaelli “Guaranteeing packet order in IBWR Optical Packet Switches with Parallel Iterative Schedulers”, European Transactions on Telecommunications, DOI: 10-1002/ett1365.
- [Pav10] (UPCT, UNIBO) P. Pavon-Marino, M.V. Bueno-Delgado, W. Cerroni, A. Campi, F. Callegati, "A parallel iterative scheduler for asynchronous Optical Packet Switching networks", to be published in *Optical Switching and Networking*.

Other:

- [Rod09a] (UVIGO) M. Rodelgo Lacruz, C. López Bravo , F.J. González Castaño, F. J. Gil Castiñeira, H. J. Chao, ”Min-Cost Max-Flow Characterization of Shared-FDL Optical Switches,” IEEE Communications Letters, vol. 13, no. 7, pp. 540-542, August 2009.
- [Rod09b] (UVIGO) M. Rodelgo Lacruz, C. López Bravo, F. J. González Castaño, F.J. Gil Castiñeira, H.J. Chao, “Not-Aligned Optical Cell Switching Paradigm”, IEEE/OSA Journal of Optical Communications and Networking, vol. 1, no. 3, pp. B70-B80, 2009.
- [Rod10] (UVIGO) M. Rodelgo-Lacruz, C. López-Bravo, F. J. González-Castaño, H. J. Chao, Felipe J. Gil-Castiñeira, “Distributed resource scheduling in not-aligned optical cell switching,” IEEE Transactions on Communications 58(4), pp. 1201-1212, April, 2010.

- number of joint experiments: 0

Overall assessment of work carried out within the project duration:

The work in this JA was based on a fruitful collaboration of the three involved partners. This is witnessed by the joint publications, and the significant amount of interactions through conference calls, and direct discussions in the BONE meetings.

One of the objectives of this JA was to design and evaluate scheduling algorithms capable of guarantee wire-speed processing in the worst case. The scheduling algorithms proposed achieved this goal for different switch fabric architectures.

Outcome of the joint research activity:

The activities carried out in this JA can be classified into three main lines:

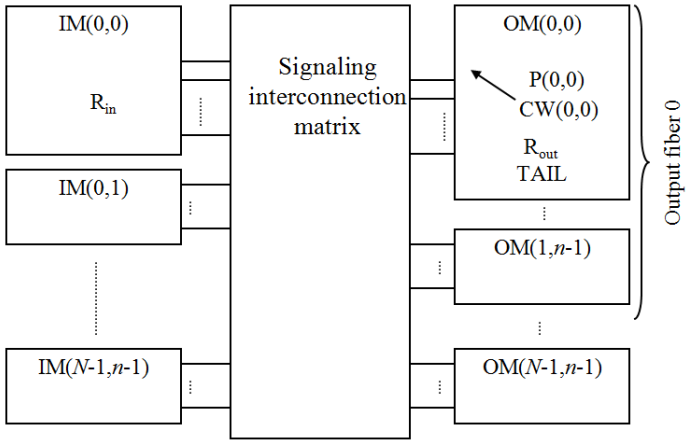
1. (UPCT, UNIBO): Design and performance evaluation of parallel scheduler for asynchronous arrivals, variable size packets in OPS/OBS switches. The PI-OPS scheduler.
2. (UVIGO, UPCT, UNIBO) Design and performance evaluation of a parallel scheduler for the IBWR switch, preserving packet sequence (O-IPDBM scheduler).
3. (UVIGO): Design and performance evaluation of different scheduling algorithms for asynchronous (not-aligned) Optical Cell Switching Networks.

1. PI-OPS algorithm

UPCT and DEIS-UNIBO have investigated a proposal for parallel schedulers suitable for asynchronous arrival of packets of variable size, e.g. for non-slotted OPS networks, or for OBS networks using the ODD (Only Destination Delay) reservation protocol. As far as the partners know, this is the first attempt in this line. We have proposed the scheduler called PI-OPS (Parallel-Iterative Optical Packet Scheduler), which has been presented in [Pav10], as an extension of the work in [Pav09].

The PI-OPS algorithm is designed as a parallel iterative algorithm, which is able to guarantee an upper bound to the response time. Let us denote this response time upper bound as T_A μs. The algorithm is executed periodically, every T_I μs. The constraint $T_I \geq T_A$ ensures that an algorithm execution starts strictly after the previous execution is finished. The algorithm execution starting at time $t=t_0$, is responsible for jointly processing the burst headers asynchronously received during the time interval $[t_0-T_I, t_0]$. We call this interval the *header arrival time window* of the algorithm execution.

Next figure depicts the main building blocks of the proposed scheduler architecture, for a switch fabric with N input and output fibers, and n wavelengths per fiber. It is based on the electronic interconnection of nN input modules (left hand side), and nN output modules (right hand side), connected by means of a crossbar interconnection for inter-module signaling. A sequence of iSLIP-like request-grant iterations are conducted in PI-OPS operation. Bit-mask registers are used in the modules to be able to allocate the appropriate fiber delay line and output port to the incoming optical packets. The PI-OPS has been specifically designed to permit a fast implementation of these operations (see [Pav10] for details)



PI-OPS scheduler architecture.

In [Pav10] it has been shown that:

1) The buffering performance for PI-OPS scheduler is very similar to that of the LAUC-VF scheduler, often considered an upper bound performance limit in OBS/OPS networks.

2) We conduct an emulation of an FPGA implementation of a large-scale version of the scheduler. Results support the feasibility of its implementation, suggesting that it is possible to build large-scale schedulers with response times below 10 microseconds.

2. Packet order in the IBWR switch. The OI-PDBM scheduler

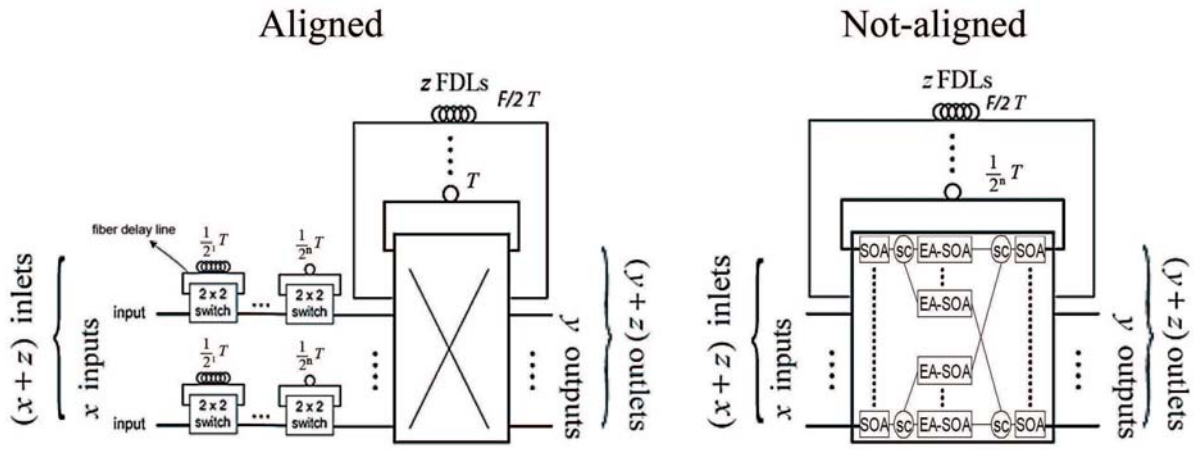
The IBWR switching architecture has been shown to have better hardware scalability properties than other optical switching fabrics for OPS networks. In the past, several algorithms have been proposed for scheduling this switch fabric under slotted traffic: fixed size packets, optically aligned at the switch inputs. PDBM [Pav07] and I-PDBM [Rod07] are two approaches in this line.

In this JA, UVIGO, UPCT and DEIS-UNIBO have investigated a variation of the PDBM-like algorithms to solve the traffic out-of-sequence problem that may arise in these architectures. The scheduler proposed is called OI-PDBM. It follows a modified request-grant scheduling, with respect to I-PDBM algorithm, as shown in [Rod09]. the advantages of OI-PDBM are (i) packet sequence is preserved, (ii) it maintains a simple and scalable implementation, (iii) it has shown good performance in terms of average packet delay, number of FDLs and algorithm convergence.

3. Design and performance evaluation of different scheduling algorithms for asynchronous (not-aligned) Optical Cell Switching Networks.

Most all-optical switching paradigms assume that different wavelengths are switched independently, which limits scalability. In optical cell switching (OCS), time is divided into time slots of fixed size by time-division multiplexing, and the wavelengths in a time slot are all bundled. Thus, each OCS switch has a single switching plane and performs mere time-space switching. Nevertheless, in an OCS network, each switching node (OCX) requires optical slot synchronizers (OSYNs) at all inputs for the arrival slots to be aligned, so that cells can be simultaneously forwarded.

To avoid this harmful process, as it was shown un previous works, UVIGO proposed a new switching paradigm, called Not-Aligned OCS [Rod09b], were the OSYNs and the alignment process are no longer required. Cell shifting still takes place inside the OCXs for minimizing the gaps between cells, but it is not necessary to align them to a reference time. Not-aligned OCS has clear advantages over aligned OCS: the total number of fiber delay loops (FDLs) and the hardware cost are reduced, and the number of switching operations is also lower. Moreover, cell arrival time to the switch is not critical, and the network becomes simpler and more flexible. Next figure shows OCS and Not-Aligned OCS OCXs architecture.



As it can be in the previous figure, in Not-Aligned OCS FDLs are shared among the different input ports. Thus, to avoid possible contention problems, a scheduling algorithm is needed. First, we presented a centralized solution (see [Rod09], for details), now we present a distributed solution (OCX-oriented algorithm), which takes connection blocking probability to reasonable values for practical loads.

Each FDL in Figure 1 delays (buffers) cells for a fixed number of time slots to avoid contention. The scheduling algorithm finds non-blocking FDL routes for the incoming OCX-OCX connections. Once established, the connection is cyclically repeated every frame, until it is released. The OCX-oriented algorithm is executed from one OCX to the next on a sequential basis. If several connections arrive at the same OCX at the same time, they are served one after another. However, it is distributed in the sense that each OCX only maintains its own state information and there is no need for centralized updated information about the whole network. Thus, the execution and state information are distributed among different OCXs. Centralized algorithms that require updated state information about the whole network are unfeasible in practice.

The complexity of the proposed algorithm grows linearly with the number of OCXs in the longest network path, so it is scalable. On the other hand simulations suggests that despite the OCX-oriented algorithm is simple, it is close to the optimum in relatively simple, realistic networks, so more complex strategies are not necessary. In the region of interest (blocking probabilities under 5%), its performance is close to that of the ideal bound given by aligned OCS with unlimited resources, outperforming other strategies as OWS [Zag00].

[Rod09b] M. Rodelgo Lacruz, C. López Bravo, F. J. González Castaño, F.J. Gil Castiñeira, H.J. Chao, “Not-Aligned Optical Cell Switching Paradigm”, IEEE/OSA Journal of Optical Communications and Networking, vol. 1, no. 3, pp. B70-B80, 2009.

[Rod10] M. Rodelgo-Lacruz, C. López-Bravo, F. J. González-Castaño, H. J. Chao, Felipe J. Gil-Castiñeira, “Distributed resource scheduling in not-aligned optical cell switching,” IEEE Transactions on Communications 58(4), pp. 1201-1212, April, 2010.

[Zag00] H. Zang, J. P. Jue, and B Mukherjee, “A review of routing and wavelength assignment approaches for wavelength routed optical WDM networks,” Optical Networks Mag., vol. 1, no. 1, pp. 111-124, Jan. 2000.

4.3 JA4 - Performance of Optical Switching System Architectures with Shared Wavelength Converters

Responsible partner: Bilkent, UNIBO

Participants:Bilkent, UNIBO, UniRoma, FT, FUB, KTH, UNIMORE

Description of the work carried out in the 3rd year:

Shared-Per-Wavelength Asynchronous Optical Packet Switching

The joint work between Bilkent and UNIBO comparatively studies four different schemes to share wavelength converters in **asynchronous** optical packet switching systems with variable length packets. The first two architectures are the well-known Shared-Per-Node (SPN) and Shared-Per-Link (SPL) architectures, while the others are the Shared-Per-Input-Wavelength (SPIW) architecture (recently proposed as an optical switch architecture in synchronous context only, and a new scheme called Shared-Per-Output-Wavelength (SPOW) architecture. In this joint work, analytical models to evaluate packet loss in asynchronous switching systems are proposed for SPIW and SPOW architectures based on Markov chains and fixed-point iterations. The proposed models also account for unbalanced traffic whose impact is thoroughly studied. These models are validated by comparison with simulations which demonstrate that they are remarkably accurate. Most of the work was carried out in the 2nd year of the project and a manuscript entitled "Shared-Per-Wavelength Asynchronous Optical Packet Switching: A Comparative Analysis" was submitted to Computer Networks (Elsevier) and revisions were made in the 3rd year of the project leading to the acceptance of this manuscript.

Shared-Per-Wavelength Multi-fiber Asynchronous Optical Packet Switching

This joint work between Bilkent and UNIBO studies the Shared-Per-Input-Wavelength (SPIW) architecture for an optical packet switch with N input and output interfaces with each interface comprising F fibers each with M wavelengths. We assume that each interface is associated with a specific destination and comprises MF wavelengths. When F = 1, the system reduces to the ordinary N by N switch with M separate wavelength channels per interface/destination. This joint work attempts to analytically model the impact of using multiple fibers, i.e., F > 1, on system performance. The SPIW architecture is illustrated in Figure 1 for the case F = 1. The wavelength conversion ratio is defined as the ratio of all wavelength converters used in the system divided by the overall number of wavelength channels on the switch. For a given wavelenght conversion ratio, we conjecture that increasing F reduces the packet loss probability since for a given packet on a given wavelength channel, its wavelength conversion requirement would be reduced due to the multiplicity of that wavelength on the destined interface. We obtain a new analytical model for this system using Markov chains and fixed-point iterations. Simulation results and analytical results obtained from the proposed analytical model are given in Figure 2 for a switch with 16 interfaces, each interface comprising MF=16 channels configured as F fibers and M=16/F separate wavelength channels per fiber. We assume that traffic is Poisson and

each output interface receives the same amount of traffic. For a given wavelength conversion ratio, the use of multiple fibers significantly reduces the loss probability and our model very accurately captures the system behavior.

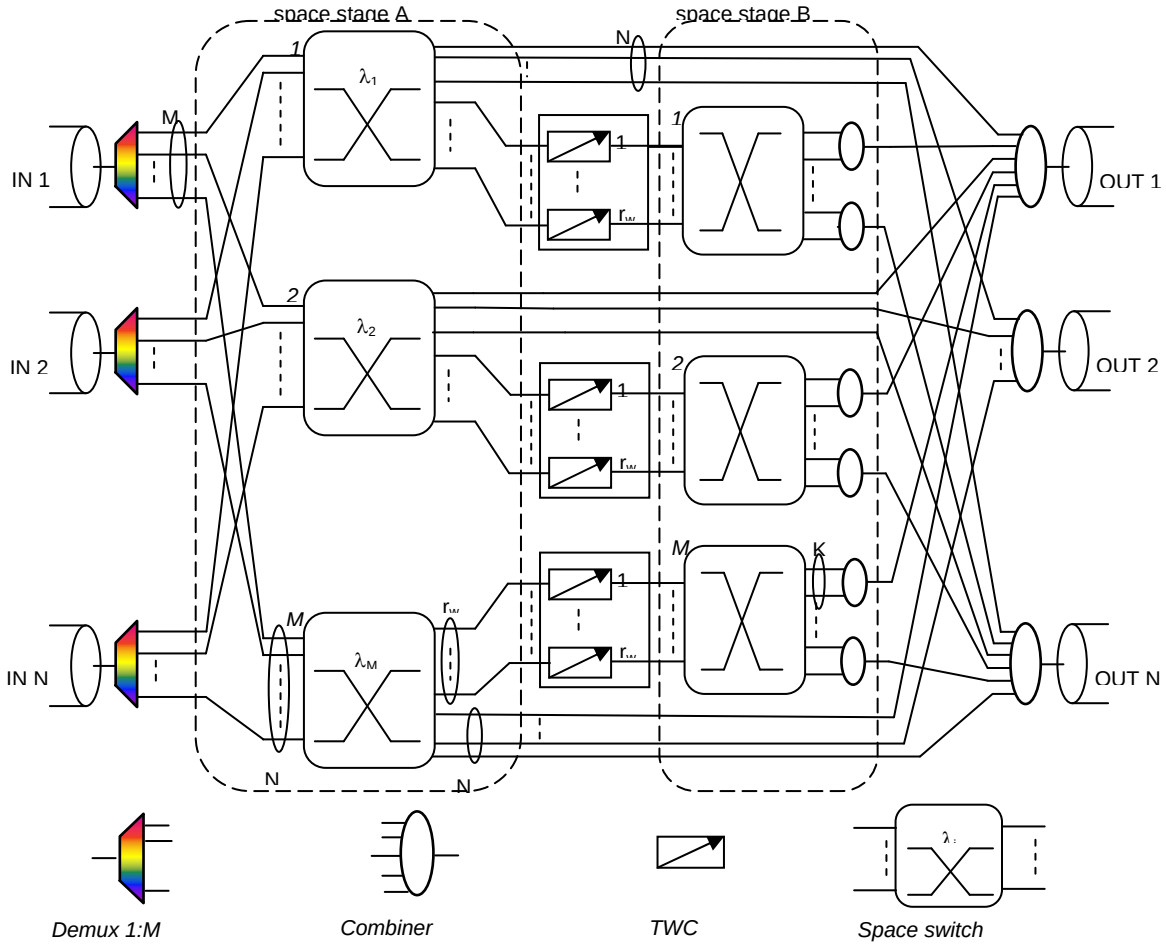


Figure 1 SPIW architecture

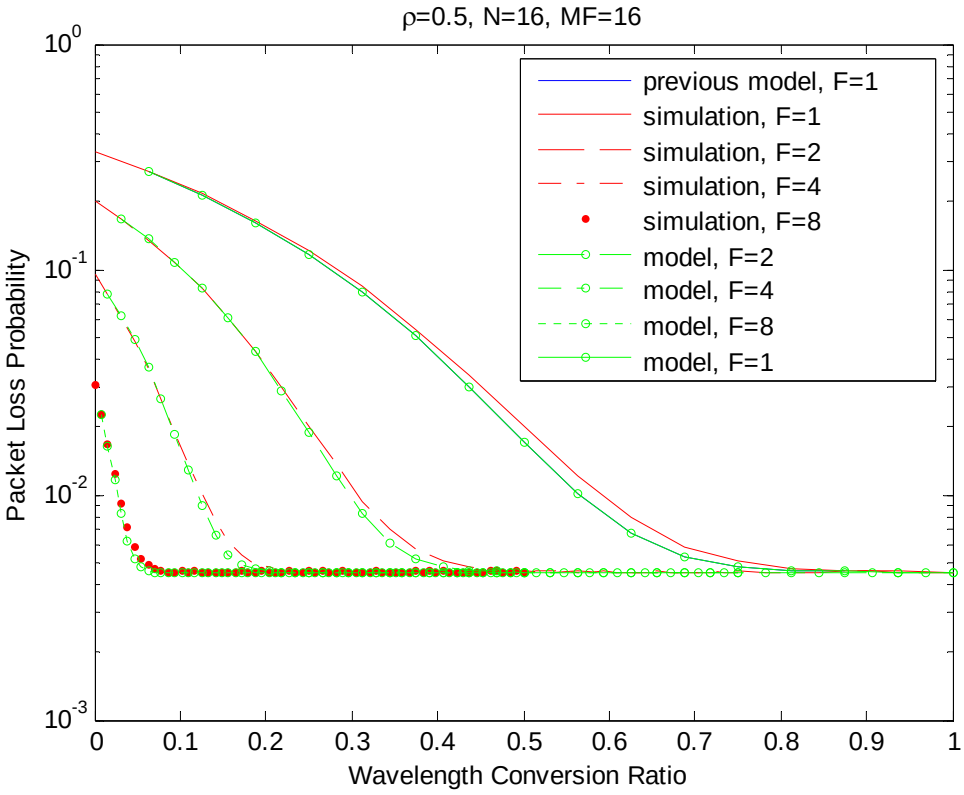


Figure 2 Packet loss probability as a function of the wavelength conversion ratio for N=16, MF=16, ρ=0.5, for various values of F.

Power Consumption in Asynchronous Wavelength-Modular Optical Switching Fabrics

In this joint work among UniRoma1, UNIBO, and Bilkent, we study power consumption in optical switching systems using share-per-wavelength converters in asynchronous and synchronous contexts. The asynchronous share-per-wavelength converter sharing solution is named ASPW, whereas the asynchronous share-per-node solution is named ASPN. The synchronous counterparts of these two schemes are called SSPW, and SSPN, respectively. We propose analytical models to evaluate the power consumption in these four solutions. Our results are depicted in Figure 3.The results we obtain show that the combination of the asynchronous operation with the wavelength based system partitioning in ASPW leads to significant power saving with respect to the other solutions in the range of interest for switch fabric dimensioning. A sample switch design with target packet loss probability performance has confirmed the expectation for different characteristics of commercial components. This work is anyway based on component theoretical modeling only and can further be validated by application of measurement-based evaluation of component behavior in practical test-beds.

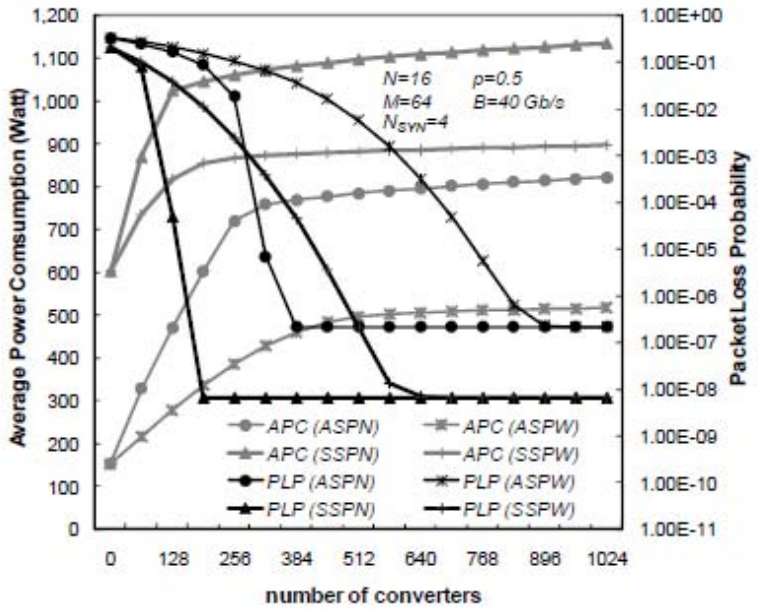


Figure 3 Comparison of the Average Power Consumption (APC) and Packet Loss Probability (PLP) in ASPW, ASPN, SSPW, and SSPN switches as a function of the number of wavelength converters. Switch and traffic parameters are N = 16, M = 64 and ρ = 0.5.

Collaborative actions carried out:

- mobility actions

1 day visit Carla Raffaelli to La Sapienza –Rome to finalize the joint activity on Power consumption evaluation. Meeting with Vincenzo Eramo.

numerous e-mail exchanges Nail Akar – Carla Raffaelli – Vincenzo Eramo

- number and details of papers

5 published joint journal papers, 2 published joint conference papers, one submitted joint paper, one joint paper which is in the process of preparation.

- number of joint experiments

Numerous joint experiments for which analytical and simulation-based results are jointly obtained.

Overall assessment of work carried out within the project duration:

In this joint activity, we studied optical packet switching system architectures with shared wavelength converters. There are different converter sharing schemes ranging from shared-per-wavelength to shared-per-node and shared-per-link architectures. The switching systems can also operate synchronously or asynchronously. Optical links can also comprise multiple fibers or a single fiber. The main goal of this joint activitiy is to study not only by simulations but also analytical modeling these architectures with the variations mentioned above. Besides performance, power consumption implications are also critical in choosing the right architecture for next-generation switching systems. Promising results in terms of accurate analytical models as well as tools for power consumption analysis are provided as an outcome of this joint activity. Moreover, most of the joint results are published in prestigious journals and conference proceedings.

Outcome of the joint research activity:

Joint publications:

- V. Eramo, A. Germoni, C. Raffaelli, M. Savi, "Packet loss analysis of shared-per-wavelength multi-fiber all-optical switch with parallel scheduling", Elsevier Computer Networks, vol. 53, no. 2, pp. 202-216, February 2009.

- C. Raffaelli, M. Savi, A. Stavdas, "Multistage Shared-per-Wavelength Optical Packet Switch: Heuristic Scheduling Algorithm and Performance", IEEE Journal of Lightwave Technology, Vol. 27, No. 5, March 2009.

- N. Akar, E. Karasan, C. Raffaelli, "Fixed point analysis of limited range share per node wavelength conversion in asynchronous optical packet switching systems", Photonic Network Communications, , Volume 18, Number 2, pp. 255-263, October 2009.

- A. Cianfrani, V. Eramo, A. Germoni, C. Raffaelli, M. Savi, "Loss Analysis of Multiple Service Classes in Shared-Per-Wavelength Optical Packet Switches", IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 1, no. 2, pp. 69-80, July 2009.

- V. Eramo, A. Germoni, A. Cianfrani, C. Raffaelli, M. Savi, "Evaluation of QoS Differentiation Mechanism in Shared-Per-Wavelength Optical Packet Switches", ONDM 2009, Braunschweig, Germany.

- C. Raffaelli, M.Savi, 'Hybrid Contention Resolution in Optical Switching Fabric with QoS traffic', WOBS 2009, Madrid, Spain.

- N. Akar, C. Raffaelli, M. Savi, and E. Karasan, "Shared-Per-Wavelength Asynchronous Optical Packet Switching: A Comparative Analysis", *Computer Networks*, vol. 54, no. 13, pp. 2166-2181, 2010.

- V. Eramo, Germoni A., Cianfrani A., Listanti M., Raffaelli C.: Evaluation of Power Consumption in Low Spatial Complexity Optical Switching Fabrics, IEEE Journal of Selected Topics in Quantum Electronics, vol. 17, no. 2, March/April 2011, DOI: 10.1109/JSTQE.2010.2053350.

Other:

– N. Akar, V. Eramo, and C. Raffaelli, "Comparative Analysis of Power Consumption in Asynchronous Wavelength-Modular Optical Switching Fabrics", submitted to Optical Switching and Networking, special issue on Green Communications and Networking.

– N. Akar, C. Raffaelli, M. Savi, "The Shared-per-Wavelength Multi-fiber Asynchronous Optical Packet Switch", in preparation.

4.4 JA5 – Code-based optical nodes

Responsible partner: UNIROMA3
*Participants:*UNIMORE, UNIBO, UNIROMA1, GET, NICT

Description of the work carried out in the 3rd year:
In this JA, UNIROMA1 has investigated the performance of Optical Code Division Multiplexing (OCDM)/Wavelength Division Multiplexing (WDM) Optical Packet Switches (OPS), describing an optimization procedure to set the optimal receiver threshold. UNIMORE has investigated high performance switching systems suitable for core nodes in optical burst switched (OBS) networks, considering, in particular, a novel buffer-less switch architecture based on optical codes. UNIROMA3 and UNIBO have studied an alternative detection mechanism for distinguishing two different types of information flow in the OpMiGua switching architecture.

Collaborative actions carried out:
- mobility actions
Prof. Carla Raffaelli e dr. Michele Savi visited UNIROMA3 to discuss the paper presented in PHOTONICS IN SWITCHING 2010.
- ***joint papers***
[1] V. Eramo, L. Piazzo, M. Listanti, A. Germoni, A. Cianfrani: Performance *Evaluation of an Optical Packet Switch using Wavelength and Code domain to solve Output Contentions*, IEEE ICTON 2010, Munich, (Germany), June 2010.
[2] N. Stol, C. Raffaelli, M. Savi, G. Cincotti, "*Optical codes for packet detection in the OpMiGua switch architecture,"* Photonics in Switching (PS), Monterey, California 2010.
- ***other papers***
[3] V. Eramo: *Impact of the MAI and Beat Noise on the Performance of OCDM/WDM Optical Packet Switches using Gold Codes*, Optics Express, vol. 18, no. 17, August 2010, pp. 17897-17912, ISSN 1094-4087.
[4] L. Piazzo, V. Eramo, A. Aceto: *Performance Analysis of an OCDMA System by Means of the Quasi Analytical and the Gaussian Approaches*, IEEE ICTON 2010, Munich, (Germany), June 2010.
[5] V. Eramo, L. Piazzo: *Optimizing Receiver Threshold in OCDM/WDM Optical Packet Switch*, submitted for publication.
[6] M. Casoni, *"A Novel Photonic Switch Architecture based on Optical Codes for Optical Burst Switched Networks"*, IEEE Workshop on Optical/Burst Switching 2008, London, September 2008.
[7] M. Casoni, A. Sacchi, *"System Design and Evaluation of a Large Photonic Switch based on Optical Codes for Optical Burst Switched Networks"*, 3rd IEEE International Symposium on Advanced Networks and Telecommunication Systems (ANTS), New Delhi, December 2009.
[8] G. Cincotti, *"Source, symbol, channel and signal coding in OCDMA systems,"* invited paper 14th Conference on Optical Network Design and Modelling (ONDM), Kyoto, Japan 2010.

Overall assessment of work carried out within the project duration:
In this JA, UNIROMA1 has investigated the performance of Optical Code Division Multiplexing (OCDM)/Wavelength Division Multiplexing (WDM) Optical Packet Switches

(OPS) [1], [3], [4] when impairment due to both Multiple Access Interference and Beat Noise are taken into account. An optimization procedure evaluating the optimum receiver threshold in an OCDM system has been proposed in the case when Gold codes are used. When the procedure is performed, better performance of the OCDM/WDM switch in terms of Packet Loss Probability due to the output packet contention, can be obtained.

In Fig. 1, the Packet Loss Probability $P_{l,noise}$ due to MAI, Primary Beat Noise (PBN) and Secondary Beat Noise (SBN) versus the number F of Optical Codes (OC) supported at each wavelength are shown. Notice that $P_{l,noise}$ is the probability that at least a bit of a packet is affected by an error due to MAI and beat noise. We have evaluated $P_{l,noise}$ for offered traffic p varying from 0.3 to 0.9 and the packet length H equal to 500 bytes. The code length L is chosen to be 1023.

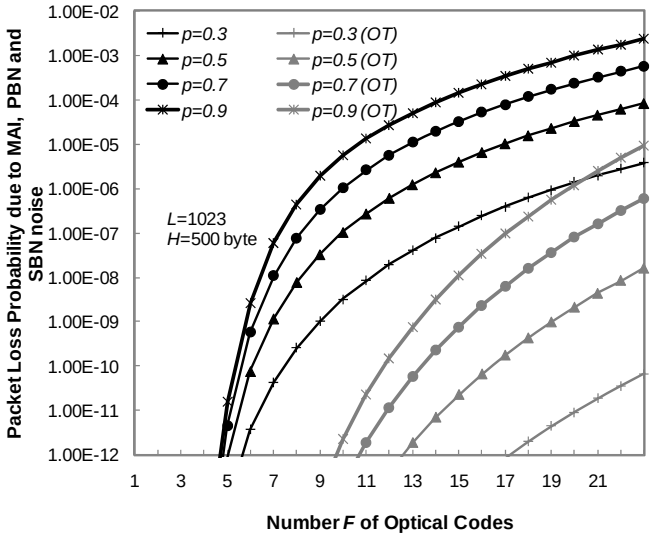


Fig. 1. Packet Loss Probability due to MAI and beat noise as a function of the number F of Optical Codes. The curves labelled with OT report the results when the Optimum Threshold is used.

For each value of traffic, we report two curves: the first is the one obtained when the not optimum receiver threshold evaluated as illustrated in [3] is used, the second one report the results when the optimum threshold is chosen according to the procedure illustrated in [5]. From Fig. 1, you can notice how, when a given $P_{l,noise}$ is fixed, a higher number of codes can be carried out on each wavelength if the optimum receiver threshold is used. For instance, when p equals 0.7 and $P_{l,noise}$ is around 10^{-9}, we have that 6 and 15 codes can be carried out on each wavelength in the cases in which not optimum and optimum threshold are used respectively.

The choice of an optimum receiver threshold impacts on the OCDM/WDM switch performance allowing for a reduction of the Packet Loss Probability $P_{l,opc}$ due to output packet contentions [3]. $P_{l,opc}$ is reported in Fig. 2 as a function of the number r of used WCs for H=500 bytes, number of Input/output Fiber N=8, number of wavelengths M=8, L=1023 and p=0.6,0.7,0.8. The number F of used OCs for each wavelength is chosen so that the threshold Packet Loss Probability $P_{l,noise}$ due to MAI and beat noise is fixed equal to 10^{-9}. For each value of traffic, we report $P_{l,opc}$ when both the not optimum and optimum thresholds are used. We can notice that better performance is obtained when the threshold is chosen according to the procedure illustrated in [5]. For instance when p=0.7, the $P_{l,opc}$ in saturation conditions, equals $3.91*10^{-4}$ and $1.28*10^{-6}$ in the cases in which the not optimum and optimum receiver thresholds are used respectively. Obviously the better performance is due to the higher number of codes that is possible to carry out on each wavelength and consequently a higher probability in solving output packet contentions.

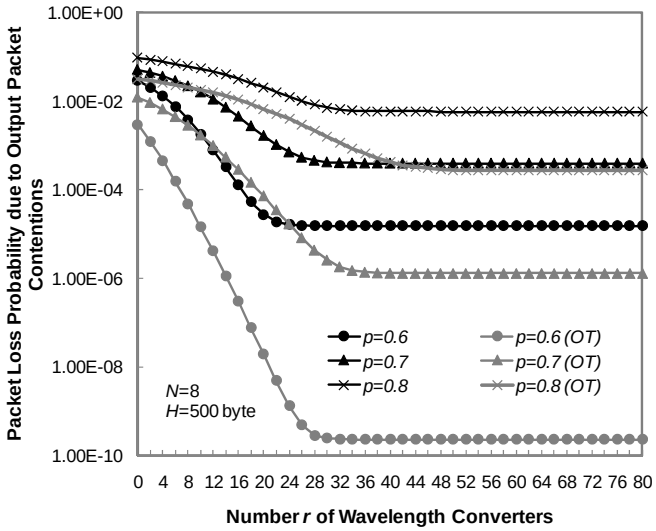


Fig. 2. Packet Loss Probability due to output packet contentions as a function of the number *r* of used WCs. The curves labelled with OT report the results when the Optimum Threshold is used.

In this JA; the research group of University of Modena and Reggio Emilia (UNIMORE) has investigated high performance switching systems suitable for core nodes in optical burst switched (OBS) networks. In particular, a novel buffer-less switch architecture based on optical codes has been proposed and evaluated.

Since in OBS networks resource reservation is typically one way by means of control packets, contention may happen among bursts in core nodes which can lead to losses. Generally speaking, contention resolution can be in different domains, such as time, wavelength and space. Here, buffer-less switches are assumed which means that no fibre delay lines are employed, so that the time domain is not used. Also, space domain is not considered. Thus, contention resolution in core nodes is here supposed to be done through wavelength conversion only.

A possible way to perform wavelength conversion is by means of optoelectronic devices. An incoming optical signal is converted into a electronic signal then it is regenerated and retransmitted employing a laser onto a different wavelength. A fundamental issue is whether the output wavelength is a-priori fixed or variable. In the former case the laser is a fixed-wavelength laser whereas in the latter a tuneable laser is required. Input wavelength is here not a issue to be concerned about. Wavelength converters with variable output are also called Tuneable Optical Wavelength Converters (TOWC) and they represent a critical device for the switch design, in particular considering their complexity and cost. On the other hand, Fixed Output Wavelength Converters (FOWC) employ fixed, not tuneable, lasers so that their complexity and cost is much lower even if, of course, they are not as flexible as TOWC. It is here also assumed that converters are full range. Therefore, important issue is the design of switching architectures which reduce the number of TOWCs as much as possible to make them feasible in shorter times. One of the goals in the design of a novel switching architecture is to avoid to use TOWCs.

Depending on the placement of converters, different switching architectures are possible [9]. In the Share-Per-Node (SPN) architecture there is a shared bank of TOWCs which can be used by any incoming burst on whatever wavelength. In the Share-Per-Input-Link a bank of TOWCs is shared among incoming bursts on the same fibre, while in the Share-Per-Output the bank of TOWCs is shared among bursts addressed to the same output fibre.

An important issue investigated in recent years is the employment of optical codes not just for transmission and access control but also for fast switching in optical core nodes. As a matter

of fact, optical coding can also be used to improve the performance of optical switching when, for instance, MPLS-GMPLS is employed [10]. This can allow to reduce the switching time increasing, correspondingly, the amount of data processed per time unit.
One of the hardest challenges in the field of optical networking is given by the study and the design of fully optical switching systems, which can support very high transmission rates and which employ optical devices of easy use and integration. During the last years, some optical devices such as multiple plane encoder-decoders [11] have been studied and implemented, which are very promising for the design and implementation of fully optical switches and routers [12].
Let us assume now a core optical switch equipped with *MxM* optical interfaces capable of supporting *N* wavelengths each. Considering the availability of optical codes and the related multiple plane encoder-decoders, it is possible to think to switch incoming packets or bursts by encoding them as a function of the requested output or, in other words, to associate a code to each output wavelength. To this end, each incoming wavelength has to be connected to a tuneable encoder driven by the control unit which decides the code to use, depending on the assigned output wavelength. Of course, the control unit, in order to take this decision, has also to perform the output scheduling. The incoming burst is then fully encoded with the assigned code and forwarded through an optical switching matrix to the desired output where a decoder, configured on that code, recovers the burst. However, it may happen that two or more incoming bursts on the same wavelength are addressed to the same output fibre so that somewhere the wavelength conversion function has to be implemented, since no fibre delay lines are used. Each outgoing wavelength is then equipped with a FOWC, which can convert a burst over any wavelength to a fixed wavelength.
The switch is then equipped with FOWCs only, from any of the *N* wavelengths to one. No TOWC is employed in this switch. The overall architecture of the switch is depicted in Figure 1.
The main result of our investigations is that the novel optical code based switch architecture has shown the same performance in terms of burst blocking probability as a SPN architecture based on TOWCs, but at lower costs, since it does not make use of this kind of expensive converters [6].
Then UNIMORE has investigated the performance of a multi-stage Clos architecture having a bufferless switch based on optical codes as basic switching element.
A core node with this multi-stage architecture has been compared with a monolithic single-stage switch, with the final goal to evaluate the performance feasibility of a large Terabit optical switch.
In order to study the performance of this switch architecture, we have used
a modular and flexible simulation tool, that we have developed and we call M_OBS_SIM (Modular OBS Simulator).
The main figure of merit investigated has been the burst blocking probability, as a function of different load conditions and switch size (Figure 2).
The main result obtained is that a 3-stage Clos architecture can be used for the realization of a terabit OBS core switch based on optical codes [7].
In this JA, UNIROMA3 and UNIBO have investigated an alternative detection approach to distinguish two different types of information flow in the OpMiGua switching architecture. Next-generation core networks are required to be able to handle all types of quality of service demands, including the ability to transport packets with guaranteed service quality. To avoid extreme over-provisioning of resources, a hybrid switch architecture, the OpMiGua network has been considered, shown in Fig. 3.

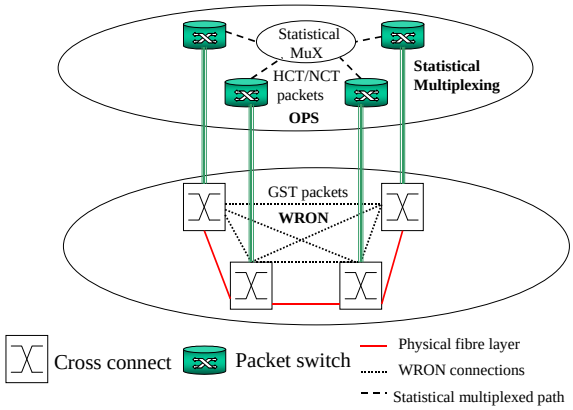


Fig. 3. The OpMiGua hybrid network structure

The information flow denoted as Guaranteed Service Type (GST) packets is transported through the optical network on pre-established lightpaths with absolute priority. Packets to be handled by the Optical Packet Switches (OPS), i.e. being statistically multiplexed are denoted as SM packets and they can subdivided into different priorities, e.g. High or Normal Class Transport (HCT or NCT). The operation is asynchronous, i.e. both GST and SM packets may have variable length. The basis for establishing lightpaths (with or without wavelength conversion capabilities) is a Wavelength Routed Optical Network (WRON) based on Optical Cross Connect switches (OXCs). The wavelengths constituting these lightpaths are, in addition, used to transport SM packets between the OPSs (link by link) when not used for transport of GST packets, thus exploiting available capacity more instantaneously and more efficiently than in older hybrid switch architectures.

The use of polarization to distinguish GST and SM packets suggested in the original OpMiGua network is not straightforward; UNIBO and UNIROMA3 have proposed an alternative mechanism, based on the use of optical codes. In this case, GST packets are not marked by an OC, and the default operation of the node is to route incoming packets to the OXC. On the other hand, all SM packets are marked at the ingress node by an OC, both in the header and tail. At the switching node, the incoming signal is split and a part of it is sent to the decoder. When a header OC is detected, the incoming signal is routed to the OPS by changing the state of two gates, and when a tail OC is detected, the state of the gates are changed back to their default positions. Header OCs can potentially have a range of values which may then also additionally be used for switching and QoS differentiation inside the OPS.

The presented OC-based solution can be more easily implemented with current technology, rather than polarization-based solutions. Also the polarization dimension is not “used up” for detection purposes, so that it would be possible to upgrade the architecture using polarization multiplexing.

REFERENCES

9. V. Eramo, M. Listanti, P. Pacifici, "A Comparison Study on the Number of Wavelength Converters Needed in Synchronous and Asynchronous All-Optical Switching Architectures", Journal of Lightwave Technology, Vol.21, No.2, February 2003, pp.340-355.

10. K-I Kitayama, M. Murata, "Versatile optical code-based MPLS for Circuit, Burst and Packet Switchings", IEEE Journal of Lightwave Technology, vol.21, no.11, November 2003, pp.2753-2764.

11. G. Cincotti, "Full Optical Encoders/Decoders for Photonic IP Routers", IEEE Journal of Lightwave Technology, vol.22, no.2, February 2004, pp.337-342.
12. G. Cincotti, N. Wada, K. Kitayama, "Characterization of a Full Encoded/Decoder in the AWG Configuration for Code-Based Photonic Routers - Part I: Modeling and Design", IEEE Journal of Lightwave Technology, vol.24, no.1, January 2006, pp.103-112.

4.5 JA6 – Repacking and rearranging algorithms for multi-plane banyan type switching fabrics

Responsible partner: PUT

Participants: PUT, POLIMI

Description of the work carried out in the 3rd year:

As a continuation of earlier works on the control algorithms for multi-$\log_2 N$ switching networks several actions were taken. One of them was the hardware implementation of the control algorithm for setting up and disconnecting connections in the switching fabric. The detailed description of this implementation can be found in [1]. The model of working controller has been realized in the FPGA (Field Programmable Gate Array) structure. We have created VHDL (VHSIC (Very High Speed Integrated Circuit) Hardware Description Language) code to describe all internal structures to implement this algorithm. To improve the efficiency and to reduce the complexity of the coding process, the dedicated tool was prepared [2]. The whole design was compiled and implemented in FPGA chip - Virtex 5 from Xilinx Company. The artificial source of requests (connections to be set up) was also prepared and used in the complete test bed. The whole system works with the basic frequency f=100 MHz. In each cycle one request is generated. Thanks to the parallelization all calculations takes less than 10 ns, what is a very good result. Moreover, the sequence of steps in the algorithm was modified so that the decision can be obtained immediately at the beginning of each algorithm's iteration. The controller was realized in ML505 - the demo board prepared for FPGA chip from Xilinx.

Another action realized in this activity concerned the extension of the rearrangement algorithm for switching fabrics with even number of stages to switching fabrics with odd number of stages. These results are described in [3]. The algorithm prepared earlier is dedicated to switching fabrics with even number of stages, it works in that way, that no more than one rearrangement was necessary. The algorithm for rearrangements in switching fabrics with odd number of stages is more complicated, because the structure of such fabrics force more complicated analysis. In this case, the maximal number of rearrangements is equal to three. Both algorithms were modified and prepared to work without any interruption of existing connection. The assumption for hardware realization of this algorithms have been prepared.

Collaborative actions carried out:

- mobility actions

5-9.05.2009 - Guido Maier from Politecnico di Milano was hosted by Poznan University of Technology.

During his stay at PUT Guido Maier has interacted with the PUT research team involved in BONE WP14 and WP25. Maier presented an overview on optical interconnects for high-performance network nodes. In the discussion that followed ideas emerged for new architectures that can be studied as applications of optical interconnects inside high-end switching systems at various levels of interconnections. The visit was also useful to discuss possibilities for joint publications with algorithms for rearrangements.

- number and details of papers

see below

- number of joint experiments

We realized a lot of common face to face discussion during different meetings (conferences, summer/master schools) and also many teleconferences. They were a part of our common work, they have produced a step forward for our works on algorithms. We have prepared several models of developed algorithms. We have also prepared several software tools, which were, still are, and probably will be in the future useful in our works on algorithms and control mechanisms.

Overall assessment of work carried out within the project duration:

The work taken during this project was very interesting and productive. Taking part in common works on the subject of control algorithms for switching fabrics allowed us to realize our goals, improve our knowledge and experience. It has also opened our minds to new ideas and research technologies. As pure results, we have developed and tested algorithms for setting up and disconnecting connections in banyan-type switching networks. This algorithm was implemented in the hardware structures, the programmed FPGA chip realizes controller for real switching network. We have prepared also algorithm for rearrangements with constant number of rearrangements. We have prepared also the core of algorithm for repacking. All prepared mechanism were optimized for hardware realization in very fast FPGA chips.

This works were useful and important, specially for young researchers from PUT, which successfully started their cooperation with foreign scientist and industry.

Outcome of the joint research activity:

1. W. Kabaciński, M. Michalski: *The FPGA Implementation of the $Log_2(N, 0, p)$ Switching Fabric Control Algorithm.* IEEE International Conference on High Performance Switching and Routing, HPSR 2010, Dallas, TX, USA 13-16 June 2010 pp. 114 - 119

2. M. Michalski: *The Graphical Analyzer for Switching Fabrics and VHDL Code Generator for FPGA Controller*. Budapest, Hungary, Bone Summer School 6-7.09.2010

3. W. Kabaciński, M. Michalski: *The algorithm for rearrangements in the $Log_2(N, 0, p)$ fabrics with odd number of stages,* submitted for IEEE International Conference on Communications, ICC 2011, Kyoto, Japan

4. W. Kabaciński, J. Kleban, M. Michalski, M. Żal, A. Pattavina, G. Maier: *Rearranging Algorithms for Log2(N, 0, p) Switching Networks with Even Number of Stages.* International Workshop on High Performance Switching and Routing, Paris, France June 22 - 24, 2009

5. W. Kabaciński, G. Danilewicz, J. Kleban, M. Żal, M. Michalski: *Sieć doskonałości BONE*. Przegląd Telekomunikacyjny LXXXI, nr 8/9, 2008 1437-1441

6. W. Kabaciński, M. Michalski, M. Żal, A. Pattavina, G. Maier: *Algorithms for Banyan-type switching fabrics.* submitted for Future Network & Mobile Summit 2011, Warsaw, Poland

4.6 JA7 – Power Consumption and Supply of Individual Network Elements

Responsible partner: TUW (Slavisa Aleksic)

Participants: TUW (Slavisa Aleksic), PoliTo (Fabio Neri, Guido Gavilanes) UNIMORE (Maurizio Casoni), UoP (Tanya Politi), UPCT (Pablo Pavon), BME (Szilárd Zsigmond), USWAN (Karin Enser), UPC (Josep Prat, Jose A. Lazaro), ISCOM (Giorgio M. Tosi Beleffi) FUB (Francesco Matera).

This joint activity (JA) is under the common umbrella of WP21 and WP14. Therefore, the results of collaborative actions within JA2 are reported in both work packages. There is a clear statement for each particular contribution about whether it is rather related to WP14 or to WP21.

Description of the work carried out in the 3rd year:

Contribution from UPC, ISCON and TUW (mainly within WP21)

Energy efficiency of long-reach optical access networks

Within this JA, TUW, UPC and ISCON carried out a study on energy efficiency of long-reach optical access networks. Power efficiency of various optical access networks was modelled and investigated in three scenarios that refer to different uplink capacities of the central office (CO), namely for a theoretically unlimited uplink and considering two limitations of 320 Gbit/s and 100 Gbit/s. The considered access networks are reach-extended versions of conventional passive optical networks (PONs) and optical P-t-P Ethernet networks providing 1 and 10 Gbit/s data rates. A special focus was dedicated to the SARDANA network, whose power efficiency is analyzed through two cases. The first one includes the measured power consumption of the existing SARDANA test bed, while the second one is an estimation including more market-ready devices with optimized power consumption values. The second case (SARDANA 2) behaves similarly to the 10G-EPON offering very good efficiency. According to the results, it can be observed that the power consumption of 1G PONs, i.e., GPON and EPON, is not being affected by considered uplink limitation for up to 1,000 users. However they provide the lowest data rate per user among all considered access options.

Description of the model

All considered networks are long-reach access options. Figure 4 depicts basic topology differences of considered networks. All networks are FTTH variants including various realizations of passive optical networks, optical point-to-point (P-t-P) Ethernet connections and the hybrid ring-tree TDM/WDM PON as proposed in the SARDANA project.

We consider reach extension of access networks by means of optical amplifiers and optoelectronic repeaters. All types of considered PON options use doped fibre amplifiers, which can extend the transmission span from 20 km that are specified in PON standards to an extended distance of 100 km. For 1G and 10G P-t-P Ethernet networks we take into account power efficient optical-electrical-optical (OEO) repeaters and semiconductor optical

amplifiers (SOA), respectively [1]. Thus, the reach of each separate P-t-P Ethernet line is extended with one bidirectional doped fibre amplifier (DFA), while in PON, reach extender is connected to only one shared feeder fibre, as shown in Figure 4.

The power efficiency of the whole network is estimated by taking into account all contributing elements and peculiarities of each technology such as topology, typical configuration and realization, and maximum data rates in downstream (DS) and upstream (US) directions. Additionally, we model the bandwidth per user as a function of several parameters. These parameters include the number of users connected to the access network, the maximum available DS and US data rates and the aggregated uplink throughput in the central office. These factors influence the performance of the system and set some limitations for considered scenarios. In this paper, we especially investigate how the uplink capacity in the central office affects the power efficiency, which we define as power consumption per Gbit/s of user's bandwidth. We consider three different scenarios:

1. Uplink bandwidth in the CO is theoretically unlimited, so that every user can achieve the maximum possible upstream and downstream data rate.
2. Uplink capacity (CU) is limited to 320 Gbit/s.
3. Uplink capacity (CU) is limited to 100 Gbit/s.

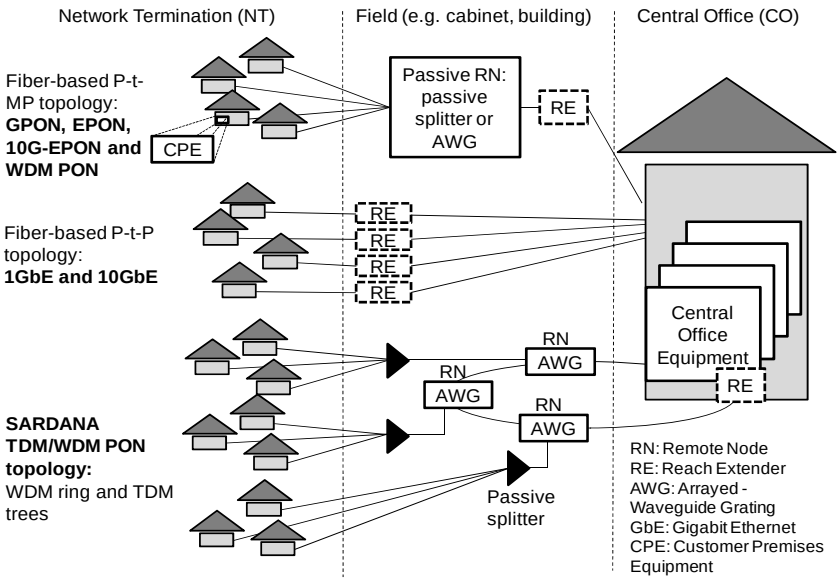


Figure 4 Considered access network topologies

The network elements are modelled at the chip level. That means first generic structures for both network-side and user-side elements are specified and then their total power consumption is calculated by summing up values for consumption of individual functional blocks. The CO equipment is modelled as an aggregation switch with a number of downlink ports, Ethernet or PON ports, and a number of 10G Ethernet uplink ports, as shown in Figure 4. Detailed structures of considered OLT and Ethernet line cards are depicted on the right side of Figure 4. It shows main functional blocks of optical line terminals firstly in the case of TDM PON, i.e., GPON, EPON and 10G-EPON, secondly in the case of WDM PON, thirdly for the SARDANA network, and finally for P-t-P Ethernet. Additionally, the figure depicts the considered structure of customer premises equipment (CPE) with a user interface consisting of a number of Fast Ethernet ports and/or a number of 1 GE ports, which are

modelled on the chip level analogue to the CO equipment. The consumption values of network elements, number of users per CO and uplink capacity (CU), are the input to the model used for calculation of power efficiency. The output of the model is power efficiency presented by Watt per user per bit/s of user's data rate. Due to the fact that users of passive optical networks connected to the same OLT have to share both US and DS data rates, they can not reach the maximum data rates unless there is only one user per OLT.

In all three scenarios we assume that the number of users connected to a CO can rise up to 1,000 and that all users are active and consume the maximum data rate they can get. The considered data rates for different technologies are net line data rates exclusive coding overhead, for instance in the case of P-t-P 1GE we assume 1 Gbit/s (w/o the overhead associated with the 8B/10B coding) in both directions, unless it is limited by the uplink of the aggregation switch in the CO. In other words, both protocol inefficiency and user behaviour are not taken into account because it would make the comparison much more complex. In the case of PONs, we assume that every user can receive the maximum downstream data rate, while upstream data rate is shared among all users connected to the particular OLT, i.e., to a TDM tree. The average data rate per user is then calculated according to:

$$R_{user} = R_{DS,\max} + R_{US,\max} \frac{N_{OLT}}{N_{user}} \quad (1)$$

where $R_{DS,max}$ and $R_{US,max}$ are the maximum downstream and upstream data rates, respectively. N_{user} is the total number of active users connected to a CO and N_{OLT} denotes the number of OLTs needed for the given number of users, while according to the standardized splitting ratio, 32 users can be connected to one OLT. For P-t-P networks, we assume that all users can achieve the maximum data rate unless there is another limiting factor. The data rate per user in P-t-P options can be only limited by the maximum capacity of the aggregation switch to which they are connected and by the maximum uplink capacity of the corresponding CO.

For the SARDANA network, we assume 32 wavelength channels, each with a TDM tree comprising 32 ONUs, which could provide access to 1024 users. However, we assume the same maximum number of users as for other access network options, i.e., up to 1,000 users per CO. SARDANA provides 10 Gbit/s of DS and 2.3 Gbit/s of US net data rate per TDM tree. In our power efficiency analysis, we introduce two SARDANA cases. The first one refers to the existing test bed in which values of power consumption are real measured values of the devices that are deployed in the test bed. The power consumption of the test bed equipment is not optimized in terms of power savings and improved efficiency as it would be in some degree done in commercial equipment. In the second case, we assume that the network is equipped with market-ready components, i.e., with components which can provide required performance and are optimized for deployment in the field. Such components would provide remarkably lower power consumption. However, electronic devices for 10G PONs are still hardly commercially available, therefore we collected available data about power consumption and formed a database that refers to many scientific papers and some product data sheets. According to those values, we were able to estimate the power consumption of future 10G PON devices that could be used for SARDANA. For more detailed information about the model and the values used as the input to the model please refer to [1].

The following diagrams show the results obtained for the three previously mentioned scenarios. Figure 5a shows the power consumption per Gbit/s of the considered technologies in the case of an unlimited uplink in the CO. Generally, the highest power consumption is observed for the 1 GE P-t-P network. For a large number of users and unlimited uplink, the power efficiency of 1 GE P-t-P, WDM PON and EPON becomes similar. This is evident from

Figure 5a as the three highest curves are closely placed around 10 W/Gbit/s. Opposite to these curves, there is the lowest band of curves referring to 10 GE P-t-P network, SARDANA 2, and 10G-EPON. They are the most power efficient networks in the given scenario. SARDANA 1 and GPON are in the middle range.

Figure 5b shows the results for the case where the limitation in the CO is set to 320 Gbit/s. In this scenario, a remarkable increase in power per user per Gbit/s can be observed for higher numbers of users. This effect can be explained as follows. A growing number of users imply an increase in required number of network terminals, i.e., higher total power consumption. At the same time, the data rate per user decreases due to the bandwidth limitation in the CO, i.e., users can not exploit their maximum data rates anymore. Consequently, an increase of power per Gbit/s, or equivalently, a decrease of power efficiency is caused.

Although 10 GE P-t-P network performs very well for up to several tens of users, it becomes the most inefficient one for more than 60 users due to the bandwidth limitation as shown in Figure 5b. Also 1 GE P-t-P network and WDM PON, which are relatively power-inefficient for small number of users, become even less efficient when the number of users increases. The power efficiency of EPON and GPON remains the same as in the previous scenario because they are not affected by the limitation, i.e., the CO uplink with 320 Gbit/s is sufficiently large to provide the maximum possible data rate to 1,000 users. SARDANA is generally not affected by this limitation until 800 users per CO. For 1,000 users, SARDANA 2 is the most power efficient network providing 3 W/Gbit/s. Apart from that, it bears great resemblance to 10G-EPON regarding power efficiency over a large interval.

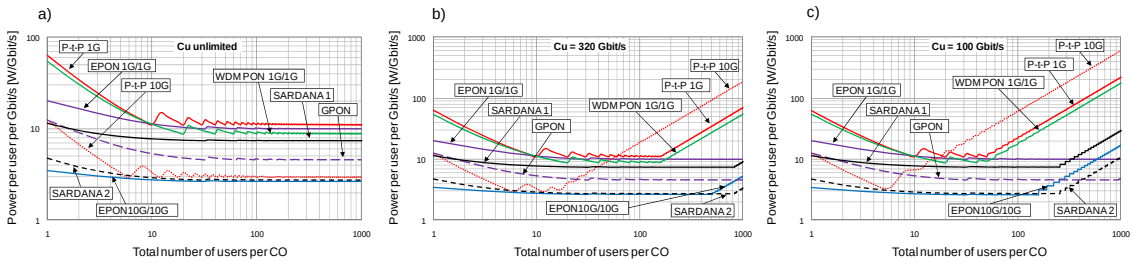


Figure 5: Results for a) unlimited uplink, b) uplink limitation of 320 Gbit/s and c) uplink limitation of 100 Gbit/s

Finally, Figure 5c represents the results obtained for the case where we set an even stronger uplink limitation, i.e., 100 Gbit/s. The general trend of increasing power consumption per bandwidth consumed can be observed for all technologies except for EPON and GPON, which still remain unaffected by the limitation even for 1,000 users. 10 GE P-t-P network becomes strongly limited in terms of available bandwidth, and accordingly least power efficient, already for more than 5 users per CO. For a very high number of users it can be seen that GPON becomes the most power efficient, while SARDANA 2 shows the highest efficiency up to 400 users.

Contribution from UNIMORE and TUW (mainly within WP14)

Optical hybrid switching node

The collaboration between UNIMORE and TUW has concentrated on developing and application of a simulator for an optical hybrid core node that can be used to study both node performance (achievable throughput and loss probabilities) and power consumption. Thus, the simulator is able to provide an estimation of power consumption depending on different

factors such as node architecture, realisation, aggregate capacity, achievable throughput, different load levels and traffic patterns. An integrated control plane for the considered optical hybrid switching node has been defined and implemented [2]. In the following, we briefly describe the integrated control plane, considered node architectures, implementation of the simulator and some preliminary results. This collaboration resulted in a mobility action between UNIMORE and TUW, in which Matteo Fiorani form UNIMORE spent more than four months at TUW in order to complete the above mentioned tasks described.

Definition of a novel integrated control plane for HOS

HOS is a switching paradigm that aims to combine optical circuit switching (OCS), optical burst switching (OBS) and optical packet switching (OPS) on the same network. In the first part of the project we defined: the format of the data (circuits, bursts and packets) handled by the HOS node, the format of a unified control packet able to carry the control information for all data types, and the structure of the scheduler. Also, we defined three possible architectures for the node to be managed by the proposed control plane.

Format of data units: The node is supposed to handle time division multiplex (TDM) circuits. In a TDM-circuit, time is divided in frames, each of which is divided in a fixed number of time-slots. Different traffic flows, sharing the same circuit, use different time-slots in a time-domain multiple access (TDMA) manner. The core node can fill unused slots in a frame with optical packets of suitable length that have the same destination as the circuit.

Aiming to increase the network efficiency, an offset time was defined for circuits. This offset time informs the nodes along the path about the exact time in which data are going to arrive at the input ports of the switch, so that the nodes reserve the resources only for the actual duration of the circuit. The circuit offset time is defined to be greater than the sum of the maximum burst-offset and the maximum burst-length, thus circuits results to have higher priority over bursts and packets.

Regarding switching of optical bursts, the JET reservation mechanism was chosen for its high efficiency. The offset time gives bursts a prioritized handling in comparison to packets.

Format of the control packet: The format of a control packet able to carry the control information for all the considered data types was introduced. The two main fields of the control packet are the routing information and the bandwidth reservation. The routing information field is used to route the control packet from source to destination node. The bandwidth reservation field carries the information needed by the core nodes to deploy the data scheduling, and it is divided into five subfields: type, length, offset, free-slots, and slot-length.

The type subfields inform the node about the type (circuit, burst or packet) of the incoming data. The length and the offset subfields contain the length (expressed in kByte) and the offset time (expressed in μs) of the incoming data, respectively. The free-slots and the slot-length subfields are used to schedule optical packets into unused slots of circuits. The free-slots subfield contains number and position of the unused slots in the frame. The slot-length contains the length of a slot.

Scheduler: The proposed control plane employs a suitable scheduling algorithm for each incoming data type.

Circuit scheduling: because circuits have the highest priority, a new circuit is scheduled on the first output channel in which no other circuit has been previously scheduled.

Burst scheduling: two different algorithms for burst scheduling have been implemented and compared, the first fit unscheduled channel with void filling (FFUC-VF) and the best fit with void filling (BF-VF). In FFUC-VF, the scheduler keeps track of all the void intervals and assigns a new burst to the first suitable void, which is found by checking all the available wavelength channels in a sequential manner. In BF-VF, when a suitable void is found, the scheduler computes the difference between the arriving time of the burst and the starting time of the void, and the difference between the ending time of the void and the ending time of the burst. The BF-VF algorithm selects the void in which the sum of these values is lower. BF-FV is more efficient than FFUC-VF but introduces higher complexity and processing time.

Packet scheduling: when a new packet reaches the node, firstly the scheduler tries to insert it on a free TDM-slot of an already established circuit. If there is no circuit able to host it, the packet is then scheduled on the first output channel in which no reservation has been made for the time required for transmission. Firstly, the performance has been evaluated without providing any buffering space for packets contention resolution. Afterwards, aiming to reduce the packet loss probability, optical buffers have been implemented.

Considered node architectures

We proposed three node architectures: all-optical hybrid, optical/electronic hybrid and all-electronic.

All the architectures are composed of an electronic control unit and a switching fabric. The control information associated to each wavelength channel is extracted and sent to the control unit, while the corresponding data are sent towards the switching fabric. The control unit converts the control signal into the electronic domain and processes it in order to select the proper output port for the corresponding data. It is divided in two functional blocks: the routing unit and the scheduler.

In the hybrid architectures, the switching fabric is divided in two parts: a slow and a fast switch. The slow switch is used for switching “slow” traffic (i.e. circuits and long bursts), the fast switch is used for switching “fast” traffic (i.e. packets and short bursts). In the all-optical hybrid architecture both the slow and fast switch are realized using optical switching elements. The fast switch is realized using semiconductor optical amplifiers (SOA), while the slow switch is realized using micro electro-mechanical systems (MEMS). Also in the optical/electronic hybrid architecture the slow switch is realized using MEMS, whereas the fast switch is realized using fast electronic switching elements. Finally, the all-electronic architecture is given by a unique fast electronic switch.

Development of the simulation model

In the second part of the project, we developed a simulation model for the analysis of the proposed node. The simulation model is a time-discrete C++ simulator. Time is divided in simulator clock-times, each of which has been set to be equal to the time needed for the transmission of 1 kByte of data. Considering a data-rate of 40 Gbps, each clock-time corresponds to 0.2 μs. One traffic source is connected to each input port of the node.

In each clock-time sources generate a new control packet with a probability belonging to a Poisson distribution with mean defined by the user. After generating a control packet, the source goes in an idle state for a fixed period of time, in which it does not generate any new control packet. The scheduler receives the new generated control packets and schedules the corresponding data on the output channels using the scheduling algorithms previously

described. The simulator produces outputs for both performance and power consumption analysis.

Performance and power consumption evaluation

The evaluation of performance and power consumption of the node was carried out by changing one parameter per time. The parameters that have been varied are: input load, traffic pattern, aggregate node capacity and wavelength converters capacity.

The figures utilized to measure the node performance are: normalized throughput, packet loss probability, burs loss probability and circuit establishment failure probability.

To compare the power consumption of the different proposed architectures the concept of "increase in power efficiency" was introduced. Consequently, the figures utilized in the power consumption analysis are the increase in power efficiency between all-optical hybrid and optical/electronic hybrid architectures ($IE_{O,O/E}$) and the increase of efficiency between optical/electronic hybrid and all-electronic architectures ($IE_{O/E,E}$).

The results regarding packet and burst loss ratios as well as node's power consumption for different realization possibilities and varying traffic pattern were obtained. An example of such a result for different loads and percentages of resources reserved for the circuit-oriented traffic is shown in Figure 6.

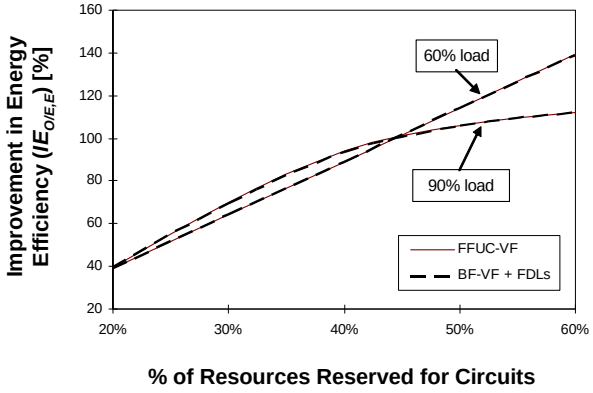


Figure 6: Improvement of energy efficiency when a dynamically adaptive hybrid optical/electronic switch is used instead of a pure electronic packet switch (a conventional IP router).

From the figure it can be observed that the increase in energy efficiency between a hybrid optical/electronic and a pure electronic implementation, $I_{O/E,E}$, exceeds 100% if in average more than 45% of the active switching ports are slow ones, i.e., if 45% of the available resources is reserved for circuits. At moderate loads of 60% and higher percentages of circuit-oriented traffic, the $IE_{O/E,E}$ increases up to about 140%. For very high loads of 90% and above, the achievable improvement of energy efficiency is limited to approximately 110% because both burst and packet losses become high at very high loads and when a large portion of the resources is reserved for circuits. Although an implementation of FDLs and efficient scheduling algorithms may have a strong impact on switch performance, i.e. on the burst and packet loss probabilities, it has, however, a very low, almost negligible influence on the achievable improvement of energy efficiency as it is evident from Figure 6. This is because an addition of optical buffers and more effective scheduling algorithms usually leads to a higher achievable throughput, but at the same time it causes an increase in the total power consumption.

Contribution from USWAN and TUW (mainly within WP21)

Techno-economic feasibility studies for greener Access Network

This contribution presents a techno-economic feasibility study to use solar energy as a renewable energy to power the passive optical network. The focus is on the requirements and sources of power for the active components of the Passive Optical Network (PON). To ensure continuous reduction of the carbon footprint while advancing towards purely passive optical networks, we have emphasized the need for alternative, renewable and greener sources to supply the required power. Solar power would be used as the default power source while mains are used as backup or a standby source. This is because solar power can be harvested and stored with less carbon footprint.

Broadband access network growth and demand

As the demand for speed and bandwidth rise, so are the requirements on network elements. Jointly with this growth and demand come the requirements to support and maintain the operations and running of access networks. It becomes crucial to consider the power needed to run and meet the growing demands of the infrastructure. The global growth for demand of Internet, Broadband and associated services continues to grow exponentially and the number of Internet users is estimated to reach 120 Millions per annum [3].

Methods to reduce the power consumption in broadband access networks

There are a number of methods proposed for reduction of power consumption in access networks. Especially for mobile devices, a very low consumption is of great importance because the batteries have limited capacity, size and weight. Additionally, because wireless technologies use shared transmission medium, there is a strong interference at the receivers that needs to be mitigated by using advanced modulation formats and coding schemes as well as intelligent control of the transmitter power by combining efficient channel selection algorithms together with power-aware transmission and routing protocol [4][5]. Similar to wireless sensor networks, also radio access networks are designed a priori for low consumption. Although there has been a lot of progress towards energy efficient mobile devices and base stations [6][7][8] there is still some room for additional improvement; especially when looking into radio base stations, a huge reduction of energy losses of about 75% may be possible [6].

The requirements on low power consumption in wired access networks are not as high as in wireless networks. Here, network terminals are usually supplied by the power grid and capacity, size and weight of batteries does not play a significant role. The most common options for a wired access are different versions of digital subscriber line (xDSL) systems, hybrid fiber/coax (HFC) network or an optical fiber-based solution (FTTx). Although radio and wireless access solutions have made a lot of progress towards energy efficiency and there are many new highly efficient components and systems for power supply and transmission in copper-based technologies such as xDSL, the most promising technology concerning high performance and low power consumption for the access area seems to be a solution based on optical fibers [9][10]. Since the contribution of access networks to the total power consumption of global networks is large, the deployment of energy-efficient components and systems for access area will have a large impact on the overall power consumption. Therefore, one should concentrate on methods for improving the energy efficiency of

components and systems within the access area.

Optical transmission and processing technologies are generally able to provide both low power consumption and high data rates. For a large number of users, high access rates, and delivering of broadcast services such as standard-definition (SD) and high-definition (HD) television, which is a realistic situation in the access area today and in the near future, the passive optical networks is one of the most efficient solution. Indeed, very large number of subscribers in urban and suburban areas can be connected to a single central office of a network provider when using a reach-extended PON. Additional to the very large number of users connected to a single OLT port (> 1000) and long reaches (~ 100 km), reach-extended PONs also provide the highest energy efficiency [11]. Different methods such as remote amplification and remote powered extender box [12] can be used to further improve the already high energy efficiency. Protocols for access networks can be adapted to allow dynamic resource allocation and to make possible to adjust performance and power consumption to actual needs by taking into account the actual network utilization. The energy savings can be then achieved by either using low-power modes or switching off inactive devices. It has been shown recently [13] that if 60% of end users are not active then up to 27% (or up to 58%) of power can be saved when inactive equipment is operating in the low-power state (or CPE is switched off).

Solar power energy and system selection

The equipment at the subscriber premises represents a small value of power consumption in the overall domestic energy consumption. Therefore the study focuses on supplying renewable solar energy for the Optical Line Terminal (OLT) which belongs to the operators. The OLTs are usually placed in the metropolitan area and connected to the core network by the Point-of-Presence (POP). This equipment belongs to the operator and when summed all the OLT's power consumption of the entire network the total value is significant large.

To power the 200W OLT to supply 256 subscribers with broadband, a minimum of a 200W solar power module/panel would be required. A possible module to use is a module supplied by Sanyo, which supplies 215 W at 42 V. This module is approximately 16.1 kg in weight and has dimensions of 1580 mm tall by 798 mm wide and a depth of 46 mm [14]. The cost of this module is $900 (approximately £605). This module would require a regulator that is capable of controlling up to 48 V. The batteries required for this task would need to have a current of at least 4.93A. A Surrette battery has a 100 hours capacity rate; at 5.32A (amps); with an overall capacity of 532 AH (ampere-hours). This would mean that the battery would be capable of powering the system for 100 hours when fully charged [15]. The cost of this battery is $343 (approximately £230), which has a weight of 58 kg and an expected lifetime of ten years.

Discussions and conclusions

The number of solar panel modules required per OLT depends on the solar energy harvest. As an example, we consider the city Swansea in United Kingdom. Let's consider the energy harvest by the 215W solar panel over the year in Swansea. The minimum number of modules required per OLT is six; this would power the OLT continuously for the months May, June and July. However in December there would be required thirty-three modules [16]. This indeed will vary with the geographic location and period of the year. One could install an average number of modules to cover partially the yearly needs and obtain a reduction of carbon emissions. Assuming that each OLT has 250 subscribers in operation to serve 1

Million users, it will be necessary 4000 OLTs. If an average number of panels per OLT is used, e.g., 13 panels, the total panels required will be a total of 52000 solar panels.

The cost of implementing solar power includes the fittings, the solar system and associated components including a battery is roughly £6200 per OLT [19]. It is important to recognize that this is a one-off cost and the price will potentially be reduced when large quantities are bought.

There are various methods of generating electricity, each producing a different amount of CO_2 per kWh. Therefore, by calculating a weighted average, the amount of CO_2 released in to the atmosphere is 0.54kg for every kWh of electricity generated [17]. The carbon emission of Photovoltaic cells is approximately 58g /kWh. This value is exclusively from the manufacture of the cells, as there are no carbon emissions when the system is in use [18]. This leads to a significant reduction of carbon emission by using solar panels to feed the telecom equipments.

Solar panels are designed to be installed outdoors and have weather resistant covers which help to retain their reliability. Another relevant issue is to maximize the system efficiency. There are many ways however the best is to check the cost per kWh. To increase the efficiency of the whole system the modules should be closed to the OLT to avoid lost of energy during travelling.

Vandalism is always a concern when installing anything in metropolitan areas where there are higher population densities. A possible solution is to install special designed modules with an impact resistant acrylic. It is recommended that the OLT is connected to the photovoltaic system and the national grid. This will allow the system to continue operating even when the solar system stops or the mains has a fault.

In conclusion, the use of solar power within broadband network is theoretically feasible. At present there is a large initial cost associated with the implementation of the solar system. This could be reduced by increasing the efficiency of the solar technology and its demand. The benefits for the environment are significant due to very large carbon emission reduction.

Contribution from PoliTo (mainly within WP 14)

Frame Scheduling for Input-Queued Switches with Energy Reconfiguration Cost

The growth of energy consumption in core router/switches is one of the most critical design problems, mainly due the thermal issues that require complex cooling systems. As a consequence the next generation of high-end core routers/switches will have to deliver maximum performance at minimum power requirements. In this research work, PoliTo focused on minimizing the energy consumption in a slotted input-queued switch with a crossbar-like switching fabric while preserving high throughput.

Typically a switching fabric is placed on a single chip and its energy consumption can be strictly related on the technology used. On the one hand, the energy consumption can be *i)* bit rate dependent (as it typically happens for electronic switching fabric families) or *ii)* it can be independent of the number of transported bits (as it can be assumed for optical switching fabric families). The family of optical switching architectures is very promising to face energy scalability issues because the energy consumption depends on the number of

rearrangements in the switching configuration between consecutive time-slots, and not on the volume of switched traffic.

The energy aware scheduling problem can be addressed through the fulfilment of the following targets:

- Change the switching fabric configuration to serve all queues in order to maximize the throughput.
- Minimize the variations between two consecutive switch configurations to minimize the energy consumption.

PoliTo [20] proposed and compared a set of scheduling algorithms to solve this problem through a frame-based approach. For each frame the problem can be solved combining two independent classes of algorithms: i) a configuration selection algorithm to maximize the throughput and ii) a frame sorting algorithm to reduce the variations between two configurations so as to minimize the energy consumption.

PoliTo considered four different algorithms to define the set of matchings composing a frame. The first three are iterative algorithms (BvN, GMax, GMin), exploiting the proposed generic decomposition algorithm Gen-DEC; they are based on the Birkhoff-Von Neumann decomposition, and on maximum and minimum weight matching respectively; the last one (Diag) is a simple decomposition algorithm based on a precomputed set of diagonals.

A sorting algorithm (aware of the energy/reconfiguration trade-off) is used to order the previously selected matchings with the aim of minimizing the energy consumption due to reconfigurations in consecutive time-slots. This problem reduces to a Travelling Salesman Problem (TSP) [21], which is NP-complete, but can be well approximated with heuristics. The effect of the different sortings was examined using 3 cases: no sorting at all (NS), Best sort (BS) and Worst sort (WS), corresponding to lowest and highest energy cost.

The impact of frame sorting algorithms is shown in Table 1 each row represents the achievable (normalized) energy consumption per packet obtained from a configuration selection algorithm with a particular frame sorting algorithm for a 64×64 input-queued switch. The impact of frame sorting allows reducing the achievable energy per packet consumption.

Configuration Select Alg	*Worst Sort*	*No Sort*	*Best Sort*
BvN	0.87	0.81	0.74
GMax	0.45	0.41	0.37
GMin	0.72	0.045	0.05
GExa	0.87	0.02	0.02
Diag	0.02	0.02	0.02

Table 1. Energy consumption comparison per packet for different algorithms

The energy performance was assessed with an ad-hoc event driven simulator. Figure 7 shows the energy-throughput trade-off obtained by GMin-NS by varying the number of switch ports *N*. It is interesting to observe that, regardless of the switch size, the maximum sustainable load is always significant, and the increase in the energy per packet as a function of *N* is marginal. Figure 8 shows the trade-off between the maximum sustainable load and the achievable energy consumption per packet. Each point represents the average value

considering the best combination between matching selection and a frame sorting selection algorithms; the horizontal and vertical bars show the maximum and the minimum values obtained.

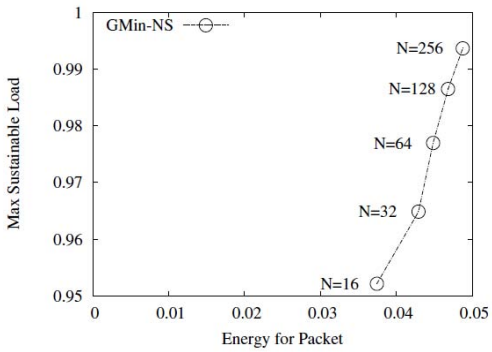


Figure 7: Throughput and energy trade-off for the GMin algorithm (with no sorting) for Uniform traffic.

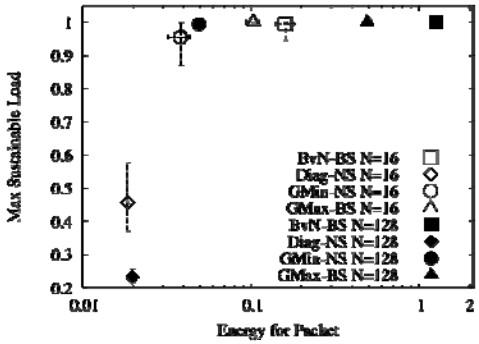


Figure 8: Throughput and energy trade-off for a number of ports N=16 and 128

It can be observed that the energy consumption per packet may vary by almost two orders of magnitude depending on the chosen algorithm. Even if BvN with best sort is a well-known optimal algorithm for energy-oblivious frame decomposition, it is also the least efficient among the studied algorithms in terms of energy consumption, further requiring a high computational complexity. GMin with no sorting appears to provide the best compromise between sustainable load (always very close to the maximum) and energy consumption (very close to the observed minimum). Diag is very good in energy consumption, but provides too low throughputs. In general this work assesses the reduction of power consumption of switching fabrics, by properly controlling the scheduling applied to control the switch, through a proper separation in two individual simpler problems.

The power efficiency of AWG-based optical switching architectures

PoliTO has studied several Optical Switching Fabrics (OSFs) for high performance packet switches in [22]. Three OSF architectures were introduced: Multiplane-Couple-Amplify-Demultiplex (MCAD), Wavelength-SelectiVe (WSV) and Wavelength-Routing-Space (WRS) architecture, which were reported in [23] and [24]. In the research work reported in [22], PoliTO focused on the WRS architecture, which is based on Arrayed Waveguide Gratings (AWGs); AWGs are passive optical devices and can have large aggregated bandwidth and lower power requirements with respect to other optical interconnection architectures based on active components.

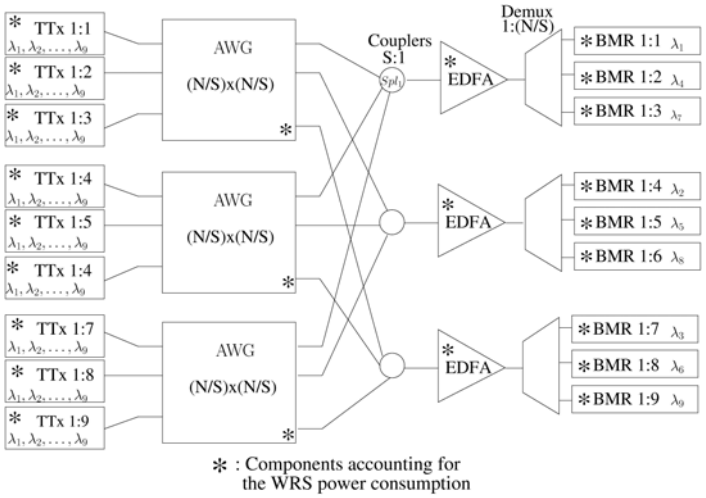


Figure 9: RS switching fabric architecture.

The WRS architecture is depicted in Figure 9; it is based on the AWG cyclic property, by which wavelengths in homologous positions of different Free Spectral Ranges (FSRs) lead to the same input/output signal routing. This property is used to perform both the plane and the destination selection. Being S the number of planes and *N* the number of linecards, each Tunable Transmitter (TTx) needs to tune *N*/*S* wavelengths on *S* different FSRs; hence requiring a total tunability range equal to *N*. 0shows a 9x9 WRS OSF with *S* = 3.

In [23] it is showed that the scalability of AWG-based OSFs can be limited by coherent crosstalk when too many AWG inputs use the same wavelength at the same time. These crosstalk limitations can however be solved either at *i)* an architectural level, or *ii)* by using a proper scheduling algorithm. In [24] a version of the WRS architecture named WRS-zc (zero crosstalk) was analyzed, that takes the former approach. In WRS-zc, by equipping each linecard with a TTx spanning over a different set of *N* wavelengths, the reuse of the same wavelength in different parts of the fabric can be avoided. On the other hand, in [25], the authors proved that it is possible to design scheduling algorithms able to select input-output matchings that minimize wavelength reuse.

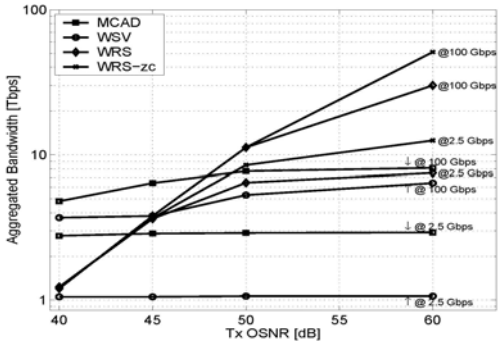


Figure 10: Aggregated bandwidth as a function of the transmitter $OSNR_{TX}$

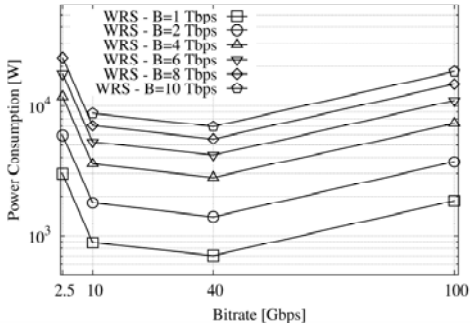


Figure 11: Power consumption trends of WRS OSF.

PoliTO has assessed the scalability and the power consumption for each one of the WRS, MCAD and WSV architectures. Physically the WRS is highly penalized by the lasers' optical noise ($OSNR_{TX}$) as shown in Figure 10, since AWGs accumulate *N* source contributions

instead of *N*/*S* for the other architectures; however, typical values of $OSNR_{TX}$ near 50 dB are possible and this allows the WRS architecture to scale in aggregate capacity comparably to MCAD and WSV. By imposing realistic sensitivity and signal-to-noise-ratio constraints at receivers, it was observed that OSFs based on commercially available components of today achieve aggregate bandwidths of several Tb/s.

Regarding the power consumption of WRS, 0reports the total power consumption as a function of the bitrate for the WRS architecture when $OSNR_{TX}$ = 50 dB for several aggregate bandwidths. The behaviour of the power consumption strongly depends on the bitrate, hence on the technology used. Indeed, technologies for low bitrates (e.g. 2.5 Gb/s) can be now considered consolidated and almost obsolete, whereas for highest bitrates (as 100Gb/s) technologies are not mature yet. The minimum power consumption occurs when using available and mature technologies.

The MCAD and WSV OSFs exhibit power requirements similar to the WRS OSF for the same aggregate bandwidth, but both the MCAD and the WSV architectures support lower switching capacity than the WRS solution.

The AWG is therefore a promising device to build OSFs able to scale to ultra-high capacities while ensuring significant power savings. Comparing the power consumption of the WRS architecture and the power consumption of the CRS-1 Cisco router (which is one of the largest packet switching devices offered on the market today), the power consumed by WRS OSF is lower than the power needed by the CRS-1 switching fabric, which is declared to be the 15% of total router power consumption. It is important to note that this gain in power efficiency is expected to be much higher moving from architectures based on discrete components (as PoliTo considered in this work) to integrated designs and properly engineered implementations.

References

[1] A. Lovrić, S. Aleksić, J. A. Lazaro, G. M. Tosi Beleffi, J. Prat, and V. Polo “Power Efficiency of SARDANA and Other Long-Reach Optical Access Networks”, accepted for publication in 15th International Conference on Optical Network Design and Modeling (ONDM 2011), Bologna, Italy, February 2011.

[2] M. Fiorani, “Performance and Power Consumption Analysis of a Hybrid Optical Core Node”, MSc Thesis, University of Modena e Reggio Emilia, Italy, pp. 1-111, 2010,

[3] www.allaboutmarketresearch.com/internet.htm.

[4] J. P. Monks, V. Bharghavan and W.-M.W. Hwu, “A power control multiple access protocol for wireless packet networks”, IEEE INFOCOMM, Feb 2001, pp. 219- 228.

[5] N.-C. Wang, J.-S. Chen, Y.-F. Huang and Y.-L. Su, “A Power- Aware Multicast Routing Protocol for Mobile Ad Hoc Networks With Mobility Prediction”, Springer, Journal Wireless Personal Communications, Vol. 43, No. 4, December 2007, pp. 1479 - 1497.

[6] T. Edler, S. Lundberg, “Energy Efficiency Enhancements in Radio Access Networks”, Ericsson Review No. 1, 2004, pp. 42 - 51.

[7] R. Min and A. Chandrakasan, “Top Five Myths about the Energy Consumption of Wireless Communication”, ACM SIGMOBILE Mobile Communications Review, Vol. 6, No.4, 2002, pp. 128 - 133.

[8] E. Shih, P. Bahl and M. Sinclair, “Wake on Wireless: An Event Driven Energy Saving Strategy for Battery Operated Devices”, Proceedings of the 8th Annual ACM Conference on Mobile Computing and Networking, September 2002, pp. 160 - 171.

[9] S. Aleksić, A. Lovrić, "Power Consumption of Wired Access Network Technologies", In Proceedings of CSNDSP 2010, Newcastle, U.K., July 2010, pp. 154-158.

[10] J. Baliga, R. Ayre, W. V. Sorin, K. Hinton and R. S. Tucker, "Energy Consumption in Access Networks", OFC/NOFC 2008, San Diego, CA, USA, February 2008, paper OThT6.

[11] S. Aleksić, A. Lovrić, "Power Efficiency in Wired Access Networks", in Elektrotechnik und Informationstechnik (e&i), Springer, Vol. 127, November 2010, Issue 11.

[12] F. Saliou, P. Chanclou, F. Laurent, N. Genay, J. A. Lazaro, F. Bonada, J. Prat, "Reach Extension Strategies for Passive Optical Networks [Invited]", IEEE/OSA Journal of Optical Communications and Networking,, Vol. 1, No. 4, Sepr. 2009, pp. C51 - C60.

[13] S. Aleksić, "Technologies and Approaches for Improving Energy Efficiency of Network Elements", (invited), Photonics in Switching (PS 2010), Monterey, California, USA, July 2010, paper PTuB3.

[14] ASANYO Energy (U.S.A.) Corp. HIT-Power-215N. *wholesalesolar*. [Online] 1 September 2009. http://www.wholesalesolar.com/pdf.folder/module%20pdf%20folder/HIT-Power-215N.pdf.

[15] Surrette. S-530__6volt. *wholesalesolar*. [Online] 2 January 2003. http://www.wholesalesolar.com/pdf.folder/battery-folder/Surrette%20S-530_6volt.pdf. DWG 039, BD-S30.

[16] K. Ennser, B. Devlin and S. Mangeni, "Towards greener optical access networks", Proc. International Conference on Transparent Optical Networks, Munich, Germany, 2010, invited paper Tu.B1.1.

[17] W.A. Steer, Energy Efficiency Calculations. *techmind.org*. [Online] 23 June 2008. http://www.techmind.org/energy/calcs.html.

[18] S. Baldwin, Carbon Footprint of Electricity Generation. London : Parliamentary Office of Science and Technology, 2006, Number 268.

[19] B Devlin, "Solar-powered optical nodes in the metropolitan area", Thesis, Swansea university, May 2010.

[20] A. Bianco, P. Giaccone, M. Ricca, "Frame-scheduling for Input-queued Switches with Energy Reconfiguration Costs", *IEEE GLOBECOM 2009*, pp. 1-6, Honolulu, December 2009.

[21] V. V. Vazirani, *Approximation Algorithms*, Springer, March 2004.

[22] BONE Deliverable D21.2b: *Intermediate TP report on first research results and planned activities*, Section 5.2, JA2 Power Consumption and Supply of Individual Network Elements.

[23] J.M. Finochietto, R. Gaudino, G. A. Gavilanes Castillo, F. Neri, "Simple Optical Fabrics for Scalable Terabit Packet Switches", *IEEE ICC*, Beijing, May 2008.

[24] D.Cuda, R.Gaudino, G.A.Gavilanes, G.Maier, F.Neri , C.Raffaelli, M.Savi , "Capacity/Cost Tradeoffs in Optical Switching Fabrics for Terabit Packet Switches", *ONDM 2009*, Braunschweig, Germany, Feb. 2009.

[25] A.Bianco, D.Hay, F.Neri, "Crosstalk-Preventing Scheduling in AWG-Based Cell Switches", *GLOBECOM'09*, Honolulu (Hi), USA, 2009.

Collaborative actions carried out:

- mobility actions: one mobility action between UNIMORE and TUW

- number and details of papers: three joint publications and ten single-partner papers

Overall assessment of work carried out within the project duration:

There have been a number of collaborative actions carried out in JA2. The most of the collaborating partners took part in informal meetings and discussions that have been held at the conferences ICTON 2010 in Munich, Germany, OFC 2010 in San Diego, CA, USA, EU Future Networks & Mobile Summit, Florence, Italy as well as at the BONE Summer and

Master School in Budapest, Hungary. Some results of JA2 were also presented at the EU Future Networks & Mobile Summit. Since JA2 is under common umbrella of WP21 and WP14, some of the activities are more related to WP21 and other to WP14. The joint work resulted in three joint publications and ten single-paper papers of which two papers are published in refereed journals (see Section 0). Additionally, three other joint journal papers are under preparation for submission. The collaborative work concentrated mainly on three particular areas, namely on energy efficiency in optical access networks (TUW, UPC, and ISCON), power consumption and performance evaluation of optical hybrid switching nodes (UNIMORE and TUW) and solar-powered optical passive networks (USWAN and TUW). The first and the third collaborative works are mainly within the scope of WP21 while the second work is mainly related to WP14. Also contribution of PoliTo is mainly related to WP14, in which a set of scheduling algorithms were proposed and compared in order to solve the energy aware scheduling problem through a frame-based approach. Additionally, PoliTo studied the power efficiency of AWG-based optical switching architectures. A mobility action was performed between UNIMORE and TUW, in which MSc student Matteo Fiorani spent more than four months at TUW working on implementation of a model for evaluation of energy efficiency in optical hybrid switching networks.

Outcome of the joint research activity:

Publications

1. (joint paper) A. Lovrić, S. Aleksić, J. A. Lazaro, G. M. Tosi Beleffi, F. Bonada, and J. Prat, Influence of Broadcast Traffic on Energy Efficiency of Long-Reach SARDANA Access Network", accepted for publication in OFC/NFOEC 2011, Los Angeles, CA, USA, March 2011.
2. (joint paper) A. Lovrić, S. Aleksić, J. A. Lazaro, G. M. Tosi Beleffi, J. Prat, and V. Polo "Power Efficiency of SARDANA and Other Long-Reach Optical Access Networks", accepted for publication in 15th In-ternational Conference on Optical Network Design and Modeling (ONDM 2011), Bologna, Italy, February 2011.
3. K. Ennser, B. Devlin and S. Mangeni, "Towards greener optical access networks", in Proc. International Conference on Transparent Optical Networks – ICTON'10, Munich, Germany, 2010, invited paper Tu.B1.1.
4. (joint paper) K. Ennser, S. Mangeni, S. Taccheo, S. Aleksic, "Techno-economic feasibility studies for solar powered passive optical network", in Proc Conf. Broadband Access Communication Technologies V as part of SPIE Photonic West Symposium on OPTO: Optical Communications: Devices and Systems, USA, 2011, paper 7958-21.
5. S. Aleksić, A. Lovrić, "Power Efficiency in Wired Access Networks", in Elektrotechnik und Informationstechnik (e&i), Springer, Vol. 127 (2010), Issue 11.
6. S. Aleksić, „Energy-efficiency of Electronic and Optical Network Elements", in Journal of Selected Topics in Quantum Electronics, IEEE, (invited), scheduled for publication in March 2011.
7. S. Aleksić, "Design Considerations toward Low-Power-Consuming Optical Network Elements", (invited), to be published in Asia Communications & Photonics Conference & Exhibition (ACP 2010), Shanghai, China, paper SI2, December 2010.

8. Lovrić and S. Aleksić, “Influence of Uplink Limitation and Broadcast Traffic on Power Efficiency in Long-Reach Optical Networks”, to be published in Asia Communications & Photonics Conference & Exhibition (ACP 2010), Shanghai, China, paper SuP4, December 2010.
9. S. Aleksić, “Technologies and Approaches for Improving Energy Efficiency of Network Elements”, (invited), in Proceedings of Photonics in Switching (PS 2010), Monterey, California, USA, July 2010, paper PTuB3.
10. S. Aleksić, A. Lovrić, “Power Consumption of Wired Access Network Technologies”, (best paper award), in Proceedings of 7th IEEE/IET International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP 2010), Newcastle, June 2010.
11. Bianco, P. Giaccone, M. Ricca, “Frame-scheduling for Input-queued Switches with Energy Reconfiguration Costs”, IEEE GLOBECOM 2009, pp. 1-6, Honolulu, December 2009.
12. E. Bonetto, D. Cuda, G. Gavilanes, F. Neri, “The Role of Arrayed Waveguide Gratings in Energy-Efficient Optical Switching Architectures”, 34th Optical Fiber Communication (OFC/NFOEC 2010) Conference, invited paper, San Diego, CA, USA, March 2010.
13. N. Fehratovic, S. Aleksic, “Power Consumption and Scalability of Optically Switched Interconnects for High-Capacity Network Elements”, accepted for publication in OFC/NFOEC 2011, Los Angeles, CA, USA, March 2011.

4.7 JA12 - Encompassing switch node impairments and capabilities in dynamic networks

Responsible partner: SSSUP-COM

Participants: SSSUP: Nicola Andriolli, Nicola Sambo, Mirco Scaffardi, Piero Castoldi;

COM: Martin Nordal Petersen, Sarah Ruep

Description of work carried out.

In the third year we have been working on a novel light path provisioning scheme based on a PMD prediction model which accounts for PMD temporal correlation properties. The proposed PMDtemporal-correlation (PTC) based light path provisioning scheme is compared with a scheme based on a classical PMD model. Simulation results show that PTC scheme significantly reduces the light path blocking probability with respect to the classical scheme. Moreover, PTC demonstrates that, by considering PMD temporal correlation, the transparency domain size can be increased, since paths that would be rejected by a classical model can be actually accepted within specific time ranges.

Introduction

Transparent optical networks are affected by optical layer physical impairments that cumulate along the paths. Therefore, some all-optical paths in the network may be unfeasible in terms of quality of transmission (QoT). In a dynamic transparent optical network, the QoT must be on-line evaluated upon lightpath request by means of estimation [1-6] or measurements [7, 8]. To cope with this issue, physical impairments must be modeled, monitored, and considered during routing and wavelength assignment, and lightpath set up [2, 4]. Moreover, in case of distributed GMPLS-controlled networks, GMPLS routing protocol [2] or signaling protocol [4] must be extended to distribute physical impairment information needed to perform QoT estimation at each node. The solutions exploiting routing protocol (e.g., OSPF-TE) extensions require that QoT parameters related to links and node ports are flooded in the whole network and stored at each node. This represents a heavy burden to the control plane, although in a network only few paths (typically the longest ones) have to be avoided because of their unacceptable QoT. For this reason, the Internet Engineering Task Force (IETF) is discussing more scalable solutions based on the signaling protocol [9] (e.g., RSVP-TE). In this case each node needs to store QoT parameters related only to local ports and attached links, and QoT parameters, gathered along the computed path by the extended signaling protocol, allow to estimate the QoT during lightpath set up. Then, if the estimated QoT is acceptable, the lightpath is established, otherwise another set up attempt over a different path is performed. A further approach exploits the path computation element (PCE) architecture to perform the QoT estimation [10], with QoT parameters obtained from other sources (e.g., Network Management System) than the control plane. However, this scenario is outside the scope of the paper, which focuses on distributed schemes.

QoT in dynamic GMPLS-controlled transparent networks can be also evaluated through measurements on probe traffic [8]. In this way, the lightpath QoT is assessed during the lightpath establishment by sending probe traffic along the candidate path. Then, if the

measured QoT on probe traffic is acceptable, the path is validated and client data traffic can be finally transmitted, otherwise another set up over a different path is attempted.

Considering QoT estimation, the authors in [3] identify amplified spontaneous emission (ASE), polarization mode dispersion (PMD), chromatic dispersion (CD), and self phase modulation (SPM) as the most relevant physical impairments to be modeled. Among them, PMD is a time-variant stochastic effect that depends on environmental (e.g., temperature, stress) and aging effects. Therefore, PMD may cause a channel outage along a path in specific time intervals, while it could be not so detrimental in other time intervals. In [3], a path is feasible in terms of PMD if the cumulated average differential group delay (DGD) $\bar{\tau}$ is below a maximum value, which ensures that the outage probability is acceptable. However, this condition can be pessimistic since, given a high $\bar{\tau}$, the instantaneous DGD may be not excessively detrimental in specific time intervals. Moreover, in [11] measurements on a 150 km span with $\bar{\tau} = 0.64$ps show that during quiescent periods, the DGD does not change more than 0.05 ps. Temporal dynamics of the PMD vector are also investigated in [12-14].

In this paper a QoT-aware lightpath provisioning scheme, based on both QoT estimation through RSVP-TE extensions and probe measurements, is proposed for GMPLS-controlled transparent optical networks. The novelty of the proposed scheme consists of improving QoT estimation by means of a PMD temporal correlation model. The adopted model, valid for any bit rate and WDM or single channel scenarios, aims at predicting PMD first-order dynamics and thus the lightpath outage probability in PMD-uncompensated systems. In particular, given the average path DGD and the measured DGD at time t_0, the model provides an estimation of the outage probability due to PMD within a time range $[t_0; t_0 + \Delta t]$. The proposed PMD-temporal-correlation (PTC) based lightpath provisioning scheme for WDM networks works as follows. First, the optical signal to noise ratio (OSNR) for the lightpath is estimated through RSVP-TE extensions. Then, the lightpath is set up, and probe traffic is sent along the lightpath for measuring the instantaneous DGD [15-17]. The measured DGD value is used by the proposed model, which accounts for PMD temporal evolutions, to assess the effects of PMD on QoT. The PTC scheme is compared with a classical RSVP-TE extensions based scheme [4] that utilizes an uncorrelated PMD model as in [3]. Simulation results show that PTC allows to significantly reduce the blocking probability. Moreover, PTC permits to establish lightpaths along paths which are rejected by the uncorrelated model, thus increasing the transparency domain size.

PMD Model accounting for PMD temporal correlation

We assume that a path is feasible in terms of PMD if, during the lightpath holding time, the outage probability---that we define as the probability that the instantaneous DGD $\tau(t)$, i.e., the modulus of the PMD vector $\vec{\tau}$, exceeds a maximum value $\tau_{\max}$---is below a given threshold P_{th}. In a classical analysis, the instantaneous DGD is assumed to be a stationary process with Maxwellian distribution and mean value $\bar{\tau}$ [18]. The corresponding outage probability is

$$P_{\text{out}} @ \Pr\{\tau > \tau_{\max}\}$$

$$= \mathrm{erfc}\left(\frac{2}{\sqrt{\pi}}\frac{\tau_{\max}}{\bar{\tau}}\right) + \frac{4}{\pi}\frac{\tau_{\max}}{\bar{\tau}} e^{-4\tau_{\max}^2/(\pi\bar{\tau}^2)} \quad (1)$$

In our approach, the idea is that, given a measure of the DGD at time t_0, we can derive a more accurate prediction of the statistical impact of PMD at time $t > t_0$ by exploiting some knowledge about the temporal dynamics of PMD. To this aim, we use the simple stochastic model in [12], in which the PMD vector $\vec{\tau}$ is modeled as an Ornstein-Uhlenbeck process in time, with drift coefficient $1/T_c$ and diffusion coefficient $(\bar{\tau}/2)\sqrt{\pi/T_c}$, where the PMD correlation time T_c and the mean DGD $\bar{\tau}$ are two given parameters. The resulting conditional distribution of the PMD vector $\vec{\tau}(t)$ at time t given the PMD vector $\vec{\tau}(t_0)$ at time t_0 is Gaussian. In particular, its components $\tau_i(t)$, $i = 1,2,3$, are three independent, jointly normal random variables, with mean values and variances

$$\begin{aligned} m_i &= \tau_i(t_0) e^{-(t-t_0)/T_c} \\ \sigma_i^2 &= \sigma^2 = (\pi\bar{\tau}^2/8)\left[1 - e^{-2(t-t_0)/T_c}\right] \end{aligned} \quad (2)$$

Therefore, the conditional distribution of the instantaneous DGD

$$\tau(t) = \sqrt{\tau_1^2(t) + \tau_2^2(t) + \tau_3^2(t)} \quad (3)$$

given $\tau(t_0)$ is a non-central chi distribution (also known as generalized Rice distribution) with three degrees of freedom [19]. As $t \to \infty$, the mean values in (2) vanish, the PMD vector approaches a stationary zero-mean Gaussian distribution with variance (per each component) $\pi\bar{\tau}^2/8$, and the DGD approaches its well-known stationary Maxwellian distribution (the unconditional distribution) with mean DGD $\bar{\tau}$. We define the *instantaneous* conditional outage probability at time t given the instantaneous DGD at time t_0

$$P_{\mathrm{out}}(t \mid \tau(t_0)) @ \Pr\{\tau(t) > \tau_{\max} \mid \tau(t_0)\} \quad (4)$$

Accounting for the non-central chi distribution of $\tau(t)$ given $\tau(t_0)$, the outage probability in (4) can be expressed as

$$P_{\mathrm{out}}(t \mid \tau(t_0)) = Q_{1.5}\left(\frac{\eta}{\sigma}, \frac{\tau_{\max}}{\sigma}\right) \quad (5)$$

where $Q_M(x, y)$ is the generalized Marcum Q-function of order M, and

$$\eta = \sqrt{\sum_{i=1}^{3} m_i^2} = \tau(t_0) e^{-(t-t_0)/T_c} \quad (6)$$

is the non-centrality parameter of the chi distribution. By using the following result on the generalized Marcum Q-function of fractional order [20]

$$Q_{1.5}(x, y) = \frac{1}{2}\mathrm{erfc}\left(\frac{y-x}{\sqrt{2}}\right) + \frac{1}{2}\mathrm{erfc}\left(\frac{y+x}{\sqrt{2}}\right) +$$

$$+\frac{1}{x\sqrt{2\pi}}\left[e^{-(y-x)^2/2}-e^{-(y+x)^2/2}\right] \tag{7}$$

and substituting (2) and (6) in (5), the conditional outage probability can be finally expressed as

$$P_{\text{out}}(t\mid\tau(t_0))=\frac{1}{2}\operatorname{erfc}(\alpha)+\frac{1}{2}\operatorname{erfc}(\beta)+$$
$$+\frac{1}{(\beta-\alpha)\sqrt{\pi}}\left(e^{-\alpha^2}-e^{-\beta^2}\right) \tag{8}$$

where

$$\beta,\alpha=\frac{2}{\sqrt{\pi}}\frac{\tau_{\max}}{\bar{\tau}}\left[\frac{1\pm\tau(t_0)/\tau_{\max}e^{-t'}}{\sqrt{1-e^{-2t'}}}\right]$$

and $t'=(t-t_0)/T_c$ is the normalized time.

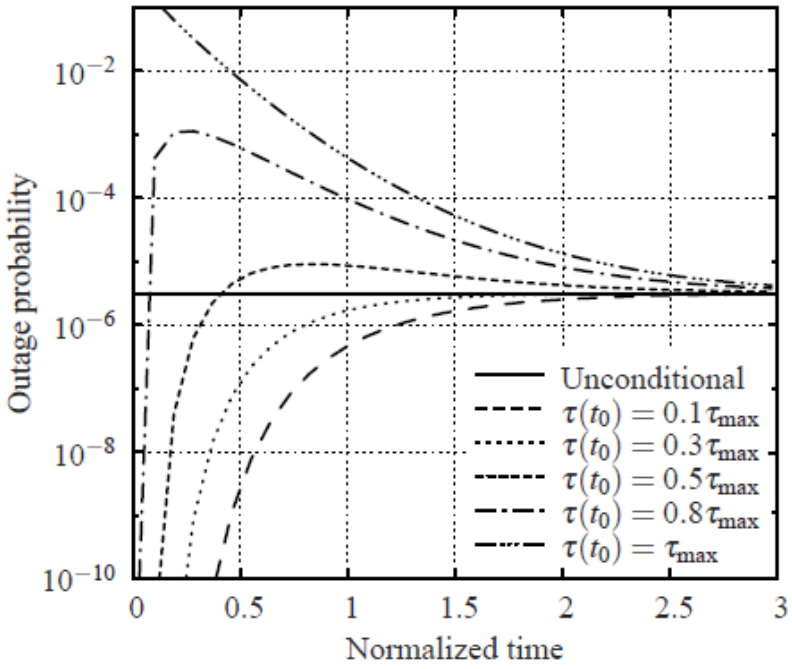


Figure 1: Instantaneous conditional outage probability vs. normalized time for a mean DGD $\bar{\tau}=0.3\tau_{\max}$ and different values of the initial DGD $\tau(t_0)$. The unconditional outage probability is also shown as a reference.

Fig. 1 shows the unconditional and instantaneous conditional outage probability as a function of the normalized time t', for different values of the initial DGD $\tau(t_0)$. The mean DGD is set to $\bar{\tau}=0.3\tau_{\max}$. The initial DGD determines the transient values of the conditional outage probability. For $t'=0$ the outage probability is either 0 or 1, depending on whether $\tau(t_0)$ is

lower or higher than $\tau_{\max}$. On the other hand, for $t' \to \infty$, the non-centrality parameter η in (6) vanishes and, as expected, the conditional outage probability (8) converges to the unconditional one given by (1), regardless of the initial DGD. In particular, for small values of the initial DGD (e.g., $\tau(t_0)/\tau_{\max} = 0.1,0.3$), (8) is always lower than (1) and converges to it from below, while for $\tau(t_0) \geq \tau_{\max}$, it is always higher and converges from above. Finally, for intermediate values of the initial DGD (e.g., $\tau(t_0)/\tau_{\max} = 0.5,0.8$), the conditional outage probability given by (8) is lower than the unconditional one given by (1) during an initial period of time; then, it becomes higher than (1) and exhibits a maximum before converging to it. It is therefore clear that, for specific holding times, depending on the initial DGD, the conditional outage probability can be significantly lower than the unconditional one.

Since we are interested in setting up a lightpath during a finite time interval $[t_1, t_2]$, we define the *average* conditional outage probability in such an interval as

$$\overline{P}_{\text{out}}(t_1, t_2 \mid \tau(t_0)) @ \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} P_{\text{out}}(t \mid \tau(t_0)) dt \tag{9}$$

Since the variation time scale of α and β in (8) is of the order of T_c, when $t_2 - t_1 = T_c$ the average outage probability (9) approaches the instantaneous one (8); on the other hand, when $t_1 - t_0 ? T_c$ or $t_2 - t_0 ? T_c$, the average outage probability (9) approaches the unconditional one (1). More in general, (9) is numerically evaluated by using (8) and resorting to some quadrature rule.

We conclude this Section with a few notes about the definition of outage probability and the dynamic PMD model adopted in this work. As regards the outage definitions in (1) and (4), a more accurate analysis should include other PMD parameters (e.g., the power-splitting ratio and second-order PMD) [21, 22]. However, as pointed out in [21], a tight upper bound for uncompensated systems can be obtained by integrating the joint distribution of the DGD τ and power splitting ratio ρ in the rectangular region $\{\rho_0 \leq \rho \leq 1-\rho_0, \tau \geq \tau_{\max}\}$, where $\tau_{\max}$ is the maximum instantaneous DGD value tolerated for $\rho = 0.5$ (worst-case power splitting ratio), and ρ_0 $(1-\rho_0)$ is the minimum (maximum) power-splitting ratio value for which outages may occur. Since no additional information about ρ is gained by measuring the instantaneous DGD, the conditional distribution of ρ is equal to the unconditional one, i.e., uniform in $\{0,1\}$ and independent of τ. Therefore, the unconditional/conditional outage probability definition in (1)/(4), based on the first-order PMD approximation and worst-case power-splitting condition, coincide with the upper bound given in [21] for uncompensated systems, except for a multiplicative factor $1-2\rho_0$, which can be practically neglected (slightly loosening the bound) or absorbed into the threshold outage probability value P_{th}.

As regards the adopted dynamic PMD model, underlying the results in (5)-(9), its check against experimental data has shown a reasonable accuracy [12]. However, more complex and accurate models may be required to better represent PMD dynamics in installed networks [11, 23, 24]. They could be adopted in the provisioning scheme proposed in the next Section without significant changes, provided that computationally treatable formulas for the conditional outage probability can be derived and replaced in (8),(9).

Lightpath Provisioning based on PMD Correlation

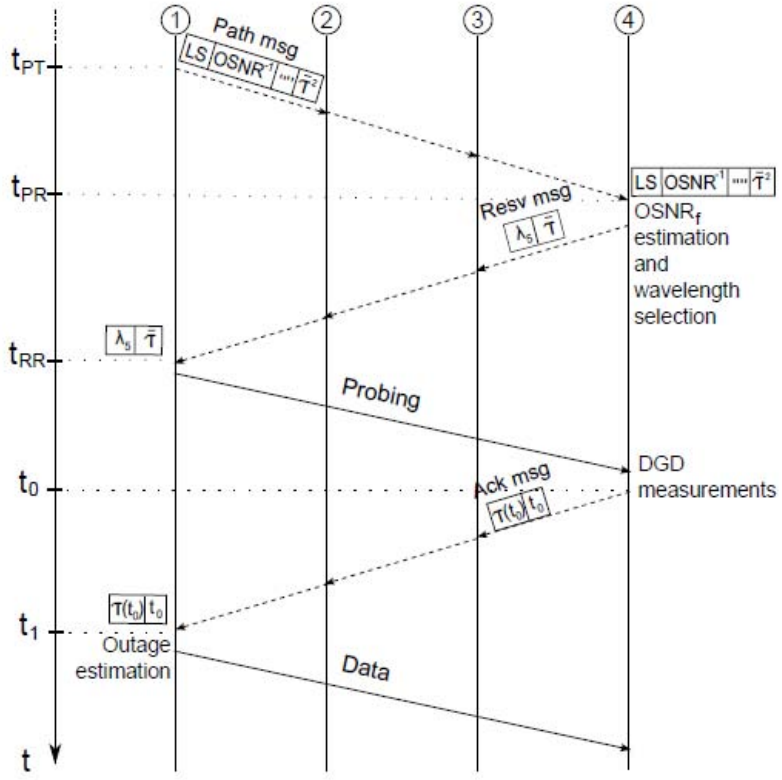


Figure 2: Lightpath provisioning time chart.

A GMPLS-controlled transparent optical network is considered. RSVP-TE signaling protocol is extended to collect additive QoT parameters along a path, as in [4]: the inverse of OSNR to account for ASE, the square of the average DGD, the residual CD, and the non linear phase shift ϕ_{NL} to account for SPM. Each node has the knowledge of QoT parameters related to local (i.e., attached) links and ports. Each node can measure the DGD.

The proposed *PMD-temporal-correlation* (PTC) based lightpath provisioning scheme is described with a time chart in Fig. 2. Upon a lightpath request from source s to destination d with a holding time Δt, s computes a path p to d (p is 1-2-3-4 in Fig. 2). Then, at time t_{PT}, s transmits an RSVP-TE Path message to d to gather wavelength availability (in the standard Label Set object --- LS) and QoT parameters (OSNR^{-1}, CD, ϕ_{NL} and $\bar{\tau}^2$). At time t_{PR}, d receives the Path message containing the list of wavelengths satisfying the wavelength continuity constraint, and the QoT parameters related to p. Then, the OSNR of the path is estimated [4]. For this purpose, the CD and ϕ_{NL} parameters contained in the Path message are computed as penalty to the OSNR. Moreover, the non-modeled effects (e.g., cross-talk, polarization dependent loss) are accounted for with worst-case margins. Penalty and margins are subtracted from the received OSNR parameter. The computed OSNR$_f$ (i.e., OSNR$_f$ =OSNR--Penalty--Margins) is compared with a threshold OSNR$_{\min}$.

If $OSNR_f < OSNR_{min}$ (i.e., QoT blocking), or if no wavelength satisfies the wavelength continuity constraint (i.e., wavelength blocking), an RSVP-TE PathErr message is sent to s. Otherwise, besides the OSNR, the outage probability due to PMD has to be estimated. First, an available wavelength is selected (e.g., λ_5). Then, an RSVP-TE Resv message, carrying also $\bar{\tau}$ of p, is sent from d to s to reserve the selected wavelength and configure the optical cross-connects for probing. When, at time t_{RR}, the Resv message reaches s, the lightpath is established for probing, and s has the knowledge of $\bar{\tau}$ of p.

Then, the instantaneous DGD of p is measured by sending probe traffic from s along p (in the data plane). At time t_0, d measures the instantaneous DGD $\tau(t_0)$ and, along the control channels, it informs s about: i) the measured DGD value $\tau(t_0)$ along p; ii) the time of measurement t_0. At time t_1, s receives the DGD measurement information. If $\tau(t_0) > \tau_{max}$ (i.e., QoT blocking) s sends an RSVP-TE ResvErr message to d to free resources along p. Otherwise, given $t_2 = t_1 + \Delta t$ (i.e., the lightpath tear down time), by exploiting $\bar{\tau}$ and $\tau(t_0)$, s computes $\bar{P}_{out}(t_1, t_2 \mid \tau(t_0))$ as in (9). If $\bar{P}_{out}(t_1, t_2 \mid \tau(t_0)) > P_{th}$ (i.e., QoT blocking), a ResvErr is sent from s to d to free resources along p. Otherwise, the lightpath is finally activated and data traffic can be sent along p. If the lightpath set up fails due to QoT or wavelength blocking, s performs further set up attempts along alternate paths.

Performance analysis

In this section, the proposed PTC lightpath provisioning scheme is compared with the *signaling approach with path feedback information (SAP)* lightpath provisioning scheme [4]. SAP uses the model in (1), which does not consider PMD temporal correlation. With SAP, QoT blocking occurs if: i) $OSNR_f < OSNR_{min}$ as with PTC; ii) P_{out}, obtained from (1), is higher than P_{th}. P_{out} is estimated at d at time t_{PR}, i.e., when the Path message is received. With SAP, if no wavelength or QoT blocking is experienced, data traffic starts to be transmitted at time t_{RR} without probing.

PMD model analysis

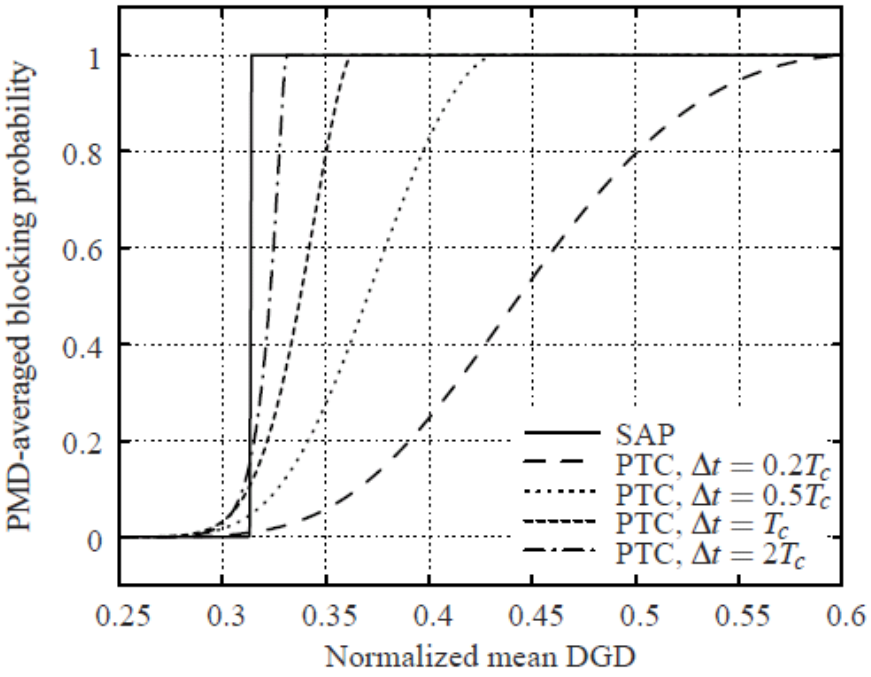


Figure 3: QoT blocking probability due to PMD vs. normalized mean DGD $\bar{\tau}/\tau_{max}$ of the path with SAP and with PTC for different holding times Δt, averaged on $\tau(t_0)$.

To highlight the different behavior of the two considered lightpath provisioning schemes, Fig. 2 shows the *PMD-averaged blocking probability* (i.e., the QoT blocking probability due to PMD averaged on $\tau(t_0)$) for SAP and PTC as a function of the normalized mean DGD $\bar{\tau}/\tau_{\max}$ of the path. The considered threshold outage probability is $P_{\text{th}} = 10^{-5}$. For PTC, t_1 is set equal to t_0 and different holding times $\Delta t = t_2 - t_1$ are considered.

With SAP, QoT depends only on $\bar{\tau}$. In this case, QoT blocking is a deterministic event that occurs when $P_{\text{out}} > P_{\text{th}}$, that is when $\bar{\tau}$ is higher than some threshold value, and the blocking probability is either 0 or 1 depending on whether $\bar{\tau}$ is lower or higher than that threshold value.

With PTC, QoT depends also on the initial DGD $\tau(t_0)$ and holding time Δt; in this case, QoT blocking is a stochastic event that occurs when $\overline{P}_{\text{out}}(t_0, t_0 + \Delta t \mid \tau(t_0)) > P_{\text{th}}$) and the PMD-averaged blocking probability---i.e., the probability that $\overline{P}_{\text{out}}(t_0, t_0 + \Delta t) > P_{\text{th}}$, computed by averaging out the Maxwellian distribution of the initial DGD $\tau(t_0)$---can assume intermediate values between 0 and 1. Therefore, there is a finite probability that a path, that would be blocked by SAP, could be accepted by PTC, and a finite probability that a path, that would be accepted by SAP, could be blocked by PTC. However, Fig. 3 clearly shows that on average, for specific holding times, PTC permits to establish lightpaths along paths with a

significantly larger mean DGD compared to SAP. As expected, for long holding times, PTC converges to SAP.

Network analysis

The performance of the proposed PTC scheme is compared to the performance of SAP in a network scenario by means of a custom C++ event-driven simulator. A Pan-European topology with 33 links and 17 nodes is considered [4]. Each link supports 40 wavelengths with 10-Gb/s NRZ modulation format. OSNR penalty due to CD and ϕ_{NL} is modeled as in [4]. For each set up attempt, the measured DGD $\tau(t_0)$ is drawn from a Maxwellian distribution with mean value $\bar{\tau} = (\sum_{i \in p} D_i^2 L_i)^{1/2}$, where $D_i = 0.27\ \mathrm{ps}/\sqrt{\mathrm{km}}$ is the PMD coefficient and L_i the length of the i-th link in p. The correlation time is $T_c = 1000\,\mathrm{s}$, the maximum DGD $\tau_{\max} = 40\,\mathrm{ps}$, and the threshold outage probability is P_{th}. Lightpath requests follow a Poisson process and are uniformly distributed among all node pairs. Both lightpath inter-arrival and holding times are exponentially distributed with an average of $1/\lambda$ and $1/\mu$, respectively. Traffic load is expressed in Erlang as the ratio λ/μ. p is randomly selected among a set $P_{s,d}$, which contains all the paths within one hop from the shortest path.

PTC and SAP are compared in terms of *lightpath blocking probability* after n set up attempts, defined as the ratio between the number of blocked lightpaths after up to n set up attempts, and the number of lightpath requests. Lightpath set up attempts can fail due to QoT blocking or wavelength blocking. Up to $n = 3$ set up attempts are performed along maximally link-disjoint paths.

Table I
DGD PARAMETERS AND ESTIMATED OUTAGE PROBABILITY FOR TWO SAMPLE PATHS

p	$\bar{\tau}$ [ps]	Δt [s]	$\tau(t_0)$ [ps]	P_{out}	$\bar{P}_{\mathrm{out}}(t_1, t_2 \vert \tau(t_0))$	QoT with SAP	QoT with PTC
p_1	13.1	654	22.6	$2.8 \cdot 10^{-5}$	$4.3 \cdot 10^{-5}$	unacceptable	unacceptable
p_1	13.1	577	3.2	$2.8 \cdot 10^{-5}$	$2.0 \cdot 10^{-8}$	unacceptable	acceptable
p_2	11.8	1200	24.8	$1.8 \cdot 10^{-6}$	$1.6 \cdot 10^{-5}$	acceptable	unacceptable
p_2	11.8	3376	17.2	$1.8 \cdot 10^{-6}$	$1.9 \cdot 10^{-6}$	acceptable	acceptable

Tab. I shows a set of cases in which two sample paths p_1 and p_2 (listed together with their DGD parameters) are either accepted or rejected due to PMD with SAP and PTC. Four lightpath holding time $\Delta t = t_2 - t_1$ are considered and P_{th} is set to 10^{-5}. Both paths do not experience QoT blocking due to low OSNR. The path p_1 has $\bar{\tau} = [13.1]ps$, which gives $P_{\mathrm{out}} = 2.8 \cdot 10^{-5}$ higher than P_{th}. Therefore, p_1 is unacceptable with SAP. Indeed, as shown in Fig. 3, paths with $\bar{\tau} > [12.5]ps$ (i.e., $\bar{\tau}/\tau_{max} > 0.31$), such as p_1, are always unacceptable with SAP. On the other hand, with PTC, p_1 can be either unacceptable (first line, $\Delta t = [654]s$ and $\tau(t_0) = [22.6]ps$) or acceptable (second line, $\Delta t = [577]s$ and $\tau(t_0) = [3.2]ps$).

Regarding the path p_2, it has $\bar{\tau} < [12.5]ps$, thus it is always acceptable with SAP. Again, with PTC, p_2 can be either unacceptable (third line, $\Delta t = [1200]s$ and $\tau(t_0) = [24.8]ps$) or acceptable (fourth line, $\Delta t = [3376]s$ and $\tau(t_0) = [17.2]ps$). In particular,

the third line of the table shows a case where PTC, exploiting $\tau(t_0)$ as additional information with respect to SAP, refuses the lightpath along p_2 because of the excessive $\tau(t_0)$.

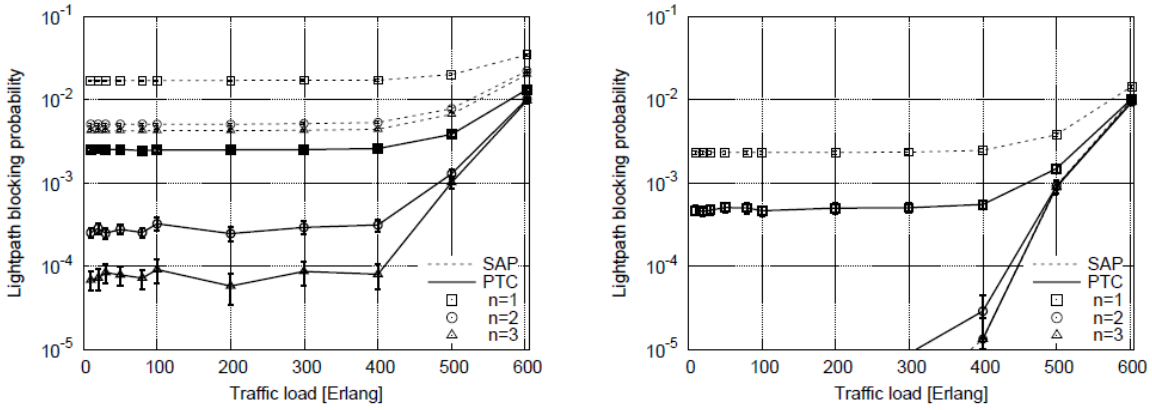


Figure 4. Lightpath blocking probability vs. traffic load [Erlang] at varying $1/\lambda$ with $1/\mu = 500\,\mathrm{s}$

Fig. 4 shows the lightpath blocking probability versus the network load, obtained by varying $1/\lambda$ when $1/\mu = [500]s$, i.e., with a fixed average holding time. P_{th} is set to 10^{-6} in Fig. 4a, and to 10^{-5} in Fig. 4b.

Lightpath blocking probability decreases while n increases, as more feasible paths are tested. Fig. 0 shows that PTC significantly reduces the blocking probability with respect to SAP at each n-th set up attempt. At loads lower than 400 Erlang, SAP achieves a constant and high blocking probability. For these loads QoT blocking is the dominant contribution, which is not influenced by lightpath mean inter-arrival and holding times with the uncorrelated PMD model. In particular, the uncorrelated PMD model causes the rejection of all paths longer than 1586 km (as reported by Tab. II).

Table II
MAXIMUM LENGTH OF AN ACCEPTED LIGHTPATH $[Km]$

P_{th}	SAP	PTC
10^{-6}	1586	2104
10^{-5}	1927	2267

On the contrary, PTC allows to accept lightpaths with a length up to 2104 km (as reported by Tab. II). PTC experiences an almost constant blocking probability at loads lower than 400 Erlang, since with the PMD correlation based model QoT blocking is influenced only by the holding time and not by the inter-arrival time, as shown in (9). For loads higher than 400 Erlang, the blocking probability of both schemes further increases since the wavelength blocking becomes dominant.

By comparing Figs. 4a and 4b, it can be noticed that the blocking probability with both SAP and PTC when $P_{\mathrm{th}} = 10^{-5}$ is lower than the one experienced when $P_{\mathrm{th}} = 10^{-6}$. Fig. 4b shows that PTC is able to significantly reduce the blocking probability at the first set up attempt. Differently from Fig. 4a, for medium loads SAP experiences a null blocking when $n = 2$, since paths with acceptable QoT are discovered in successive set up attempts. This is due to the less stringent limit on the outage probability, i.e., $P_{\mathrm{th}} = 10^{-5}$, compared to

$P_{\text{th}} = 10^{-6}$ of Fig. 4a. Fig. 4b confirms that both models are not influenced by the average inter-arrival time. For medium loads, PTC experiences a non-null blocking probability for $n = 2$. This is explained with Fig. 2, which shows a region in which the PMD-averaged blocking probability predicted with PTC (by exploiting the knowledge of $\tau(t_0)$) is actually higher than that predicted with SAP (which does not exploit $\tau(t_0)$). For $n = 3$ both SAP and PTC experience a null QoT blocking. Finally, with $P_{th} = 10^{-5}$, paths with length up to 1927 and 2267 Km are acceptable with SAP and PTC, respectively (as reported by Tab. II).

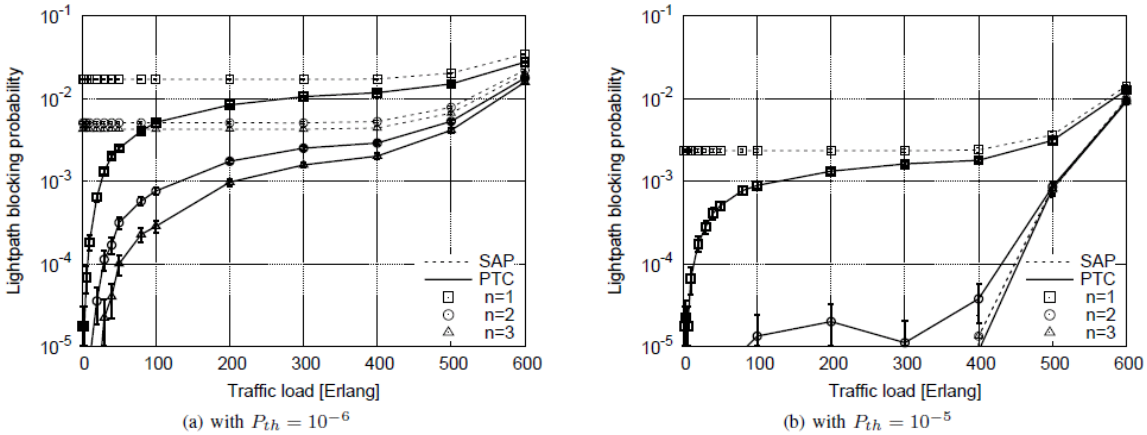


Figure 5. Lightpath blocking probability vs. traffic load [Erlang] at varying $1/\mu$ with $1/\lambda = 10\,\text{s}$

Fig. 5 shows the blocking probability versus the network load, obtained by varying $1/\mu$ when $1/\lambda = [10]s$, i.e., with a fixed average inter-arrival time. P_{th} is set to 10^{-6} in Fig. 5a, and to 10^{-5} in Fig. 5b.

Fig. 5a confirms that when the limit on the outage probability is stringent, PTC significantly decreases the blocking probability with respect to SAP at each n-th set up attempt. In particular, for low loads (corresponding to short holding times) QoT blocking approaches zero even with stringent $P_{th} = 10^{-6}$. Then, for increasing load, PTC blocking probability increases and approaches SAP blocking, since the lightpath holding time is more likely to be higher than T_c, where both models give similar outage probability estimates. Again, for loads higher than 400 Erlang, the blocking probability of both schemes increases since the wavelength blocking becomes dominant. Similarly to Fig. 4b obtained with the same P_{th}, in Fig. 5b PTC overcomes SAP at the first attempt, while SAP achieves a null QoT blocking at the second attempt.

Tab. II refers to Figs. 4 and 5, and summarizes the maximum length values of the accepted lightpaths (here named as network transparency size), for both SAP and PTC. The table shows that PTC permits to increase the network transparency size, for both P_{th} values. In particular, PTC permits to increase the transparency domain size for the considered scenario with respect to SAP by 30% and 18% when P_{th} is 10^{-6} and 10^{-5}, respectively.

Conclusion

In this paper, a lightpath provisioning scheme, called PMD-temporal-correlation based (PTC) lightpath provisioning scheme, has been proposed for GMPLS-controlled dynamic transparent optical networks. PTC exploits a PMD prediction model, based on the fact that DGD evolution is correlated in time. Therefore, if a lightpath does not experience a channel outage at a specific time, it is supposed that the outage is not experienced for a subsequent time range.

PTC exploits RSVP-TE extensions to estimate physical impairment parameters during lightpath set up. Additionally, PTC enforces a monitoring phase in which DGD is measured in order to estimate the outage probability within the required lightpath holding time.

Simulation results show the benefits of the proposed PTC lightpath provisioning scheme with respect to a classical QoT-aware lightpath provisioning scheme. In particular, PTC strongly reduces the lightpath blocking probability with respect to the other scheme. Moreover, PTC allows an increase in the transparency domain size, since the PMD temporal correlation model permits establish of lightpaths along paths that would be rejected by the classical model.

References

[1] B. Ramamurthy, D. Datta, H. Feng, J. P. Heritage, and B. Mukherjee, ``Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks," J. Lightw. Technol., vol. 17, no. 10, pp. 1713--1723, Oct. 1999.

[2] J. Strand, A. L. Chiu, and R. Tkach, ``Issues for routing in the optical layer," IEEE Commun. Mag., vol. 39, no. 2, pp. 81--87, Feb. 2001.

[3] D. Penninckx and C. Perret, ``New physical analysis of 10-Gb/s transparent optical networks," IEEE Photon. Technol. Lett., May 2003.

[4] F. Cugini, N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, E. Le Rouzic, and J. Poirrier, ``Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks," J. Lightw. Technol., vol. 26, no. 19, Oct. 2008.

[5] A. Morea, N. Brogard, F. Leplingard, J. Antona, T. Zami, B. Lavigne, and D. Bayart, ``QoT function and A* routing: an optimized combination for connection search in translucent networks," OSA Journal of Optical Networking, vol. 7, no. 1, pp. 42--61, 2008.

[6] Y. Pointurier, M. Brandt-Pearce, and S. Subramaniam, ``Analysis of blocking probability in noise- and cross-talk-impaired all-optical networks," IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 1, no. 6, Nov. 2009.

[7] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, and A. E. Willner, ``Optical performance monitoring," J. Lightw. Technol., vol. 22, no. 1, Jan. 2004.

[8] N. Sambo, C. Pinart, E. Le Rouzic, F. Cugini, L. Valcarenghi, and P. Castoldi, ``Signaling and multi-layer probe-based schemes for guaranteeing QoT in GMPLS transparent networks," in Proc. of OFC 2009, San Diego, CA, Feb. 2009.

[9] G. Martinelli and A. Zanardi, ``GMPLS signaling extensions for optical impairment aware lightpath setup," IETF draft-martinelli-ccamp-optical-imp-signaling-02.txt, July 2009.

[10] G. Bernstein, Y. Lee, and W. Imajuku, ``Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks," draft-bernstein-ccamp-wavelength-switched-02.txt, Oct. 2007, Internet Draft.

[11] M. Brodsky, M. Boroditsky, P. Magill, N. Frigo, and M. Tur, ``A hinge model for the temporal dynamics of polarization mode dispersion," in Proc. of Lasers and Electro-Optics Society, 2004. LEOS 2004. The 17th Annual Meeting of the IEEE, Nov. 2004.

[12] C. Antonelli, A. Mecozzi, M. Brodsky, and M. Boroditsky, ``A simple analytical model for PMD temporal evolution," in Proc. OFC 2006, March 2006, paper OWJ4.

[13] O. Karlsson, J. Brentel, and P. Andrekson, ``Long-term measurement of PMD and polarization drift in installed fibers," Lightwave Technology, Journal of, vol. 18, no. 7, pp. 941 --951, jul. 2000.

[14] C. Allen, P. Kondamuri, D. Richards, and D. Hague, ``Measured temporal and spectral PMD characteristics and their implications for network-level mitigation approaches," Lightwave Technology, Journal of, vol. 21, no. 1, pp. 79 -- 86, jan. 2003.

[15] M. Brodsky, N. Frigo, M. Boroditsky, and M. Tur, ``Polarization mode dispersion of installed fibers," J. Lightw. Technol., Dec. 2006.

[16] J. Jiang, S. Sundhararajan, D. Richards, S. Oliva, M. Sullivan, and R. Hui, ``PMD monitoring in traffic-carrying optical systems," in Proc. of European Conference on Optical Communication, ECOC 2008, Sep. 2008.

[17] M. N. Petersen, N. Sambo, N. Andriolli, and M. Scaffardi, ``PMD monitoring using optical sideband filtering," in Proc. of Lasers and Electro-Optics Society, 2009. LEOS 2009., Oct. 2009.

[18] G. J. Foschini and C. D. Poole, ``Statistical theory of polarization dispersion in single mode fibers," J. Lightwave Technol., vol. 9, no. 11, 1991.

[19] J. G. Proakis, Digital Communications, 4th ed. 1em plus 0.5em minus 0.4em New York: McGraw-Hill, 2001.

[20] A. Annamalai, C. Tellambura, and J. Matyjas, ``A new twist on the generalized marcum Q-function QM(a, b) with fractional-order M and its applications," in Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE, Jan. 2009, pp. 1--5.

[21] E. Forestieri and G. Prati, ``Exact analytical evaluation of second-order PMD impact on the outage probability for a compensated system," Lightwave Technology, Journal of, vol. 22, no. 4, pp. 988--996, April 2004.

[22] P. Winzer, H. Kogelnik, and K. Ramanan, ``Precise outage specifications for first-order PMD," Photonics Technology Letters, IEEE, vol. 16, no. 2, pp. 449 --451, feb. 2004.

[23] C. Antonelli, C. Colamarino, A. Mecozzi, and M. Brodsky, ``A model for temporal evolution of PMD," Photonics Technology Letters, IEEE, vol. 20, no. 12, pp. 1012 --1014, jun. 2008.

[24] =2 plus 4 3 minus 4 D. Yevick, M. Reimer, and M. O'Sullivan, ``Simplified transition matrix analysis of the hinge model," J. Opt. Soc. Am. A, vol. 26, no. 3, pp. 710--714, 2009. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-26-3-710

Outcome of Joint Activity

- Publication in Journal of Lightwave Technology, Vol.28 Issue.24-24, page 3307-, 2010:

“Accounting for PMD Temporal Correlation during Lightpath Set Up in Transparent Optical Networks”

- Conference contribution to Optical Fiber Communication Conference and Exposition,OFC 2010, OMU2 :

“Accounting for PMD Temporal Correlation during Lightpath Set up in GMPLS-Controlled Optical Networks”

4.8 JA 14 – Photonic code label processors for ultrafast routing

Responsible partner: University of Cambridge

Participants: UNIROME and TU/E

Description of the work carried out in the 3rd year:

We have been developed a technique for DPSK high speed signals which are encoded by specially design SAC code labels for self routing using external commercial devices. HNLF and nonlinear SOA components are used to perform all-optical self-switching and label swapping of spectrally label encoded 40 Gb/s signals. A Q-factor of > 6 is obtained after self-switching and label swapping for 21 channels.

Collaborative actions carried out:

- mobility actions

We have a visit from Dr Kevin Williams in April 2010 where we discussed the presented paper in ECIO conference.

- number and details of papers

We published the paper "All Optical Switching and Label Swapping Using HNLF and SOAs for 40 Gb/s Spectrally Encoded Systems", in ECIO 2010. BONE project was acknowledged in the paper.

Overall assessment of work carried out within the project duration:

For the this work developed in the 3rd year of BONE, it was developed this novel ultrafast self routed CDMA labelled system. Recently, label swapping has become an important function for future networks based on optical label switching. Here, one or more labels are renewed or added in an intermediate node of a multi-hop switched network. The labels are used as headers to determine the forwarding path or for the management of other networking functions such as quality of service (QoS). Previously, we developed spectral amplitude code (SAC) labels for label switching since they have benefits such as fast label recognition and the potential for a large label set of codes. Most coded label systems needs a bank of optical decoders to recognize the SAC labels and multi-wavelength conversion for label swapping since the SAC labels have weight > 2.To alleviate the use of a bank of correlators, multiple SAC labels can be recognized by a single correlator using four wave mixing (FWM) sideband (SB) allocation and selective optical filtering. Label switching of signals can be separate payload. Futhermore, recently, we developed a control-less fast self switching system that self-switches the signal to a FWM sideband,

Importantly, this approach does not need switch fabrics or other complex items, such as complex de/encoders, short pulses, envelope detection, etc. So the below experiment shows a novel all-optical self switching system with label swapping for the two-weight code self switching system using FWM and cross gain modulation (XGM) respectively. Since the codes are built with only two wavelengths, this alleviates the cost, power consumption and complexity of the label swapping section since multi-wavelength conversions for swapping one code to another are not necessary. Here, new bins are formed by converting the wavelength of the switched signal to a new bin via XGM whilst the second bin is a local cw

signal to form the new label. A schematic of the self switched network with label swapping is shown in fig. 1.

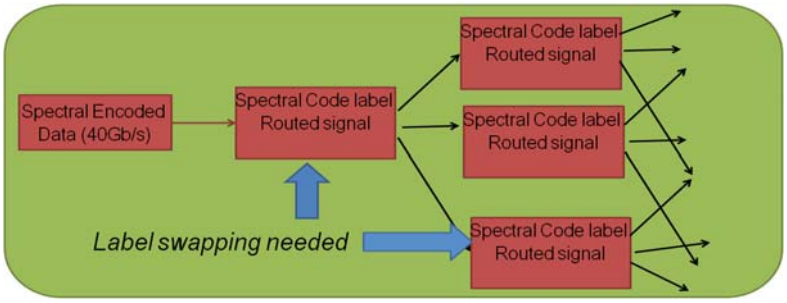


Fig.1. Multi-hop spectral coded label self switched network

Operation and experimental demonstration

Fig. 2 shows the optical packet format. Each SAC label is modulated by the data payload. Thus, the label is implicit in the payload bits (each bit has two wavelengths) allowing the payload to be self-switched by FWM according to table 1 (which is developed using an algorithm similar to the one used for FWM long distance transmission). The code family used has a length of 7, a weight of 2 and a total set of 21 codes. The bins are separated by a multiple of a frequency slot (f), in this case 100 GHz, which is also the minimum bin separation. The frequencies that overlap with others cannot be used for self-switching and these are highlighted in grey in table 1. It is important to notice that in, a nonlinear SOA was used and a shorter table was demonstrated since its FWM efficiency response less than the HLNF. The HNLF allows having enough FWM reponse for about 15nm bandwidth.

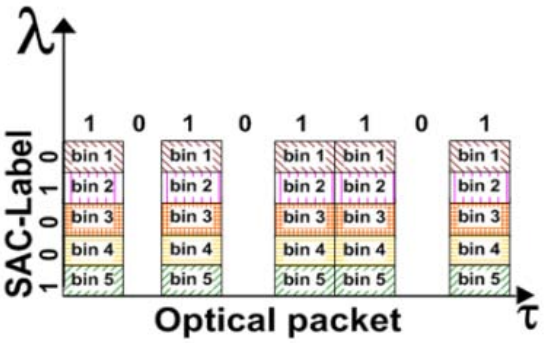


Fig.2. Optical packet format

TABLE 1. WAVELENGTH BIN ALLOCATION FOR SAC-LABEL FAMILIES

in nm	0	2	6	9	13	17	20	Δf=100GHz	
Code	1544.1	1545.7	1548.9	1551.3	1554.5	1557.8	1560.2	FWM 1	FWM 2
1	1	1	0	0	0	0	0	1547.32	1542.5
2	1	0	1	0	0	0	0	1539.37	1553.7
3	1	0	0	1	0	0	0	1537	1558.9
4	1	0	0	0	1	0	0	1533.86	1565.1
5	1	0	0	0	0	1	0	1530.72	1571.7
6	1	0	0	0	0	0	1	1528.38	1576.6
7	0	1	1	0	0	0	0	1542.54	1552.1
8	0	1	0	1	0	0	0	1540.16	1557
9	0	1	0	0	1	0	0	1537	1563.5
10	0	1	0	0	0	1	0	1533.86	1570
11	0	1	0	0	0	0	1	1531.51	1575
12	0	0	1	1	0	0	0	1546.52	1553.7
13	0	0	1	0	1	0	0	1543.33	1560.2
14	0	0	1	0	0	1	0	1540.16	1566.7
15	0	0	1	0	0	0	1	1537.79	1571.7
16	0	0	0	1	1	0	0	1548.11	1557.8
17	0	0	0	1	0	1	0	1544.92	1564.3
18	0	0	0	1	0	0	1	1542.54	1569.2
19	0	0	0	0	1	1	0	1551.32	1561
20	0	0	0	0	1	0	1	1548.91	1565.9
21	0	0	0	0	0	1	1	1555.34	1562.6

Figure 3 shows the experimental setup. Two tunable lasers, TL1 and TL2, generate the spectral code labels and they are modulated by the data, thus producing the spectrally encoded bits. As an example we use code 20 in table 1. After 25 km of SMF transmission and the corresponding DCF, at the intermediate node we implement all optical self switching and label swapping. Firstly, the 40Gb/s spectrally encoded signal is propagated through a highly nonlinear fibre (HNLF) where it is wavelength switched to only one FWM SB (output port) following table 1 - in this case to a wavelength of 1548.91 nm (port 20). The HNLF has a zero dispersion wavelength at 1552nm, a length of 135 m and a nonlinear coefficient of 20 (W.km)-1. Suppose we desire to transmit to the next node identified by Code 19, we need to implement a label swapping section. This is achieved using two local lasers in the intermediate nodes (TL3 and TL4), and a SOA which has 15 dB gain and saturation power of 12 dBm. XGM in the SOA is used to convert the signal from 1548.91nm (incoming signal from Port 20) to 1554.54nm using TL3 (this signal becomes one bin of new Label 19). The cw signal from TL3 is transmitted in the counter propagation direction in the SOA so filtering of this signal is avoided after the XGM process.

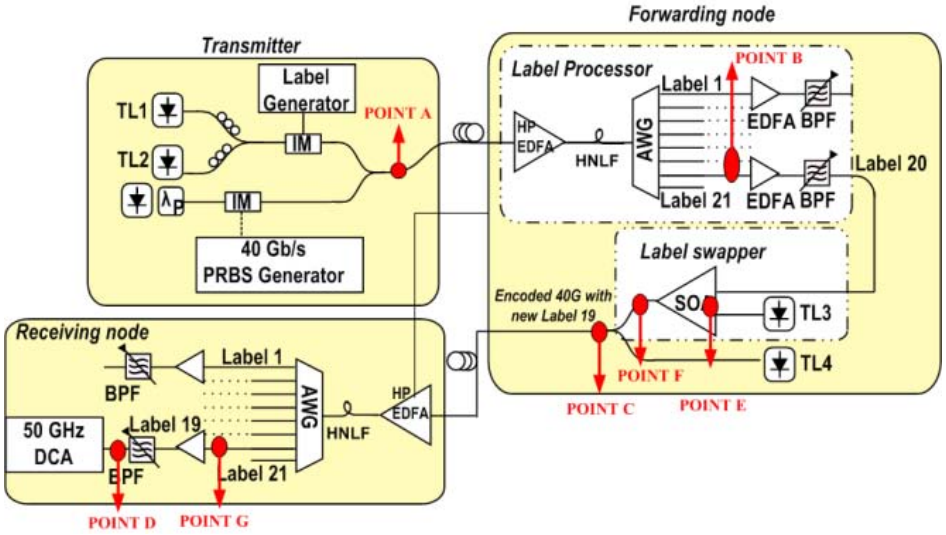


Fig.3. Experimental setup

After the self-switched data signal has been converted, it is combined with TL4, which provides the second bin of the new SAC label. To build label 19, it is tuned to 1557.8 nm. At

the receiver node, the Code label 19 is self routed to Port 19 and its Q factor is measured using a 50 GHz digital communications analyser (DCA).

Results

Figures 4(a) to 4(d) show the eye diagrams of the 40 Gb/s signal at different stages in the system (the stages are labelled in Fig. 3). Fig 4(a) shows the transmitted eye of the original label code 20 which has with Q-factor of 10.1 (point A in the setup). Fig. 4(b) shows the eye diagram of the Code 20 after FWM self switching, with Q=7.33 (point B). Fig. 4(c) and 4(d) show the wavelength routed signals converted from 1548.91 nm to 1554.54 nm, with Q factors of 6.5 and 6.02, respectively (points C and D). Fig. 4(e) to Fig 4(h) show the spectrum of the signal at the corresponding stages. Fig. 4(e) shows the Code 20 spectrum after FWM at the intermediate node (point B), where we can see the input bins and the two SBs. The SB at 1548.91nm is filtered and input to the swapping processor. Fig. 4(f) shows the incoming Code 20 data together with the CW signal used in the XGM wavelength conversion process (point E). It should be noted that the data is inverted from the original. The signal after wavelength conversion and filtering is shown in Fig 4(g) (point F). Finally, the spectrum of the 40 Gb/s signal encoded with the new Code 19 is shown in Fig 4(h) (point G).

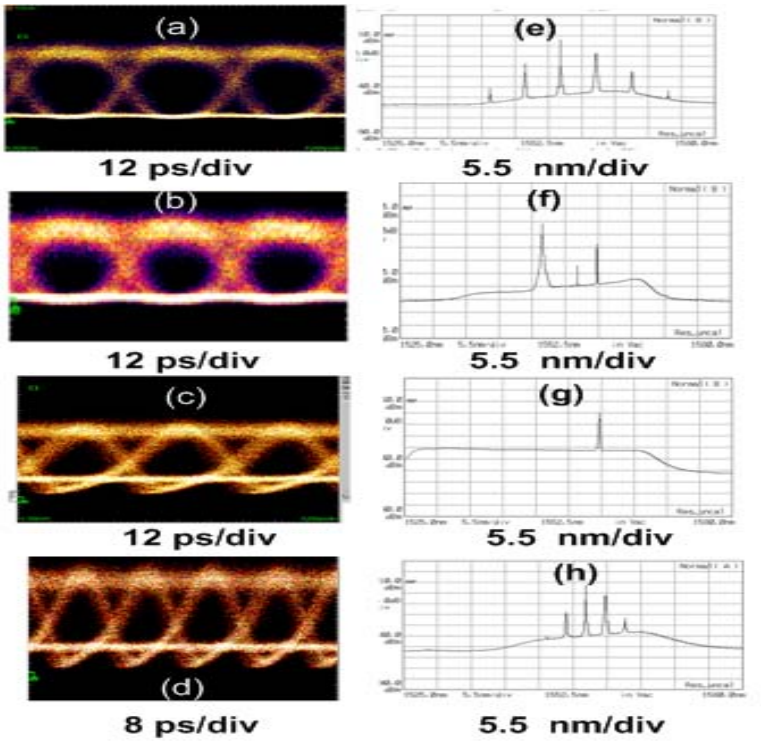


Fig. 4. (a) Original eye diagram of the 40 Gb/s Code 20 spectrally encoded data; (b) 40 Gb/s encoded signal after self routing by FWM; (c) 40 Gb/s encoded signal after wavelength conversion by XGM; (d) 40 Gb/s encoded signal with Code 19 after self routed by FWM; (e) Spectrum of Code 20 after FWM in the forwarding node; (f) spectrum of filtered 40 Gb/s signal and the CW used in the XGM process; (g) converted 40Gb/s data (h) 40Gb/s coded signals after FWM at the receiver.

Outcome of the joint research activity:

It has been developed a novel self switching system based on two weight codes and FWM in a HNLF with long bandwidth response. Furthermore, it is demonstrated all optical label swapping of spectrally label encoded 40 Gb/s signals is optimized using XGM in a nonlinear SOA for a 21-port self routing system. In contrast to previous SAC label swapping demonstrations, multi-wavelength conversion is not needed. A Q-factor of >6 is obtained

after self-switching by FWM and label swapping. It is possible to route 21 different 40 Gb/s signals in a multi-hop self switched label system with a total throughput of 0.84 Tb/s.

4.9 JA16 - Low-crosstalk optical packet-switching architectures based on wavelength-switching and wavelength-sensitive devices.

Responsible partner: PoliMi

Participants: PoliMi, PoliTO, UPCT, UVIGO

Description of the work carried out in the 3rd year:

This Joint Activity is focused on the design and evaluation of architectures based on wavelength-sensitive devices. On this account the main objective is to study the crosstalk impairment on these architectures and to limit its impact on them. The effort applied from the participants to the JA in the third year comprises two different aspects of the topic. In particular, PoliMi investigated the characteristics of the Arrayed Waveguide Grating (AWG) based structure that represents the wavelength-switching core of its architecture, while PoliTO and UVIGO strengthened their work on the design and evaluation of crosstalk limiting schedulers for AWG-based optical switches. Even if the two contributions covers two aspects of the same topic, they will be presented in the following in two different sections.

Scalability of Optical Interconnections based on the Arrayed Waveguide Grating

The contribution from PoliMi to this Joint activity is based on a feasibility and scalability analysis of switching architectures with overall capacity in the order of Tb/s [1]. The work has been focused on the issue of interconnecting line cards of the routers exploiting wavelength agility at transmitters to control signals switching by means of a Arrayed Waveguide Grating (AWG) based structure.

In Y3, starting from the architecture described in the previous deliverable (Fig. 2), PoliMi deeply investigated the characteristics of the aforementioned structure that is adopted as the central stage of the switching backplane. Moreover, a comparison with a single plane architecture has been conducted.

The simplest conceivable AWG-based backplane architecture is shown in Fig. 1, which has been already presented in Ref. [1]. In this architecture all the *N* transmitters (TX) of the line cards are directly connected to a single AWG and thereby to the *N* receivers (RX).

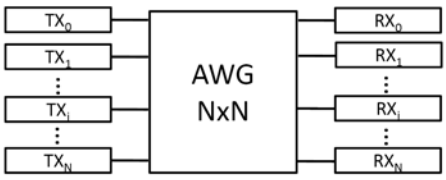


Fig. 1 *AWG based single-plane architecture.*

Here the routing property of the AWG is exploited to connect TX and RX. The architecture proposed by PoliMi is depicted in Fig. 2. The fabric can be divided in switching planes to which different subsets of transmitters and receivers are connected. Be *S* the number of switching planes and *N/S* the number of line-cards for each plane; in Fig. 2 an example with S=3 and N/S=4 is provided. The architecture is so composed:

- *N* line-cards (disposed over *S* planes) consisting of a tunable transmitter operating at the bit rate of a single WDM channel and a receiver.

- *S* AWGs (one per each plane). Routing properties of AWG to connect every ingress input ports to the desired output port.
- A stage of *S* couplers that collect channels coming from different switching planes.
- *S* EDFA amplifiers.
- *S* demultiplexers (*1:N/S*) that address each channel to a corresponding output port within each plane. These demultiplexers can be realized with a broadcast and select structure made of a coupler and a filter.

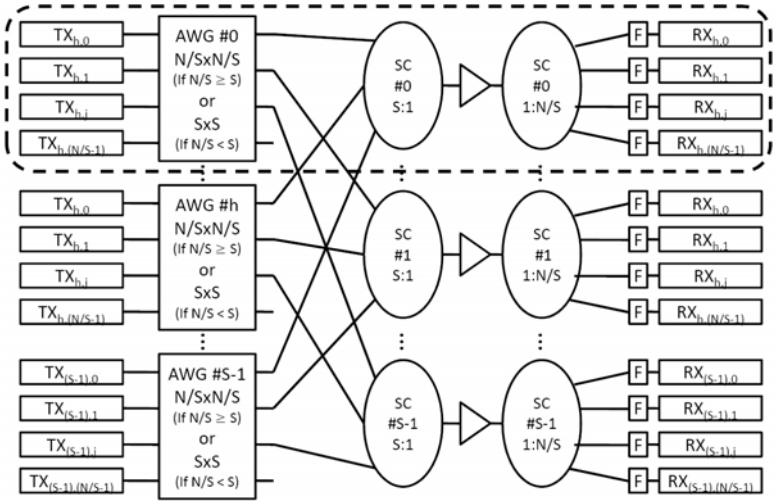


Fig. 2. *AWG based multi-plane architecture.*

In a generic optical system several degradation effects are to be taken into account to analyze the physical layer. In our board-to-board Optical Interconnection (OI), however, given the typical interconnection distances involved and the presence of the AWGs, the most relevant of such effects is the optical crosstalk. In our AWG-based architecture crosstalk is present at each AWG output port and its origin is due to the interfering channels contributions caused by the non-ideal routing properties of the device. Crosstalk can be classified as in-band (coherent) crosstalk and out-of-band (incoherent) crosstalk. In the first case crosstalk is due to interfering channels at the same wavelength of the useful signal. It is particularly harmful as it can not be eliminated by optical filtering and may cause intensity noise beating with the signal. Conversely, out-of-band (non-coherent) crosstalk is originated by interfering channels on wavelengths different from that of the signal. Thus, during Y3 PoliMi investigated how these crosstalk impairments can be mapped into the architecture performance evaluation.

Let us better define the crosstalk impairment. Given a pair of input/ouput ports (*k*,*j*) of an AWG and an input wavelength λ_p on the port *k*, be λ_t the primary lobe wavelength of the AWG transfer function evaluated between port *k* and port *j*. If λ_p is equal to λ_t, the signal will pass through the AWG with no attenuation (we are not considering AWG insertion loss at this step). Otherwise it has to be considered as crosstalk for all the other signals passing at the port *j*. Moreover, as it can be seen in Fig. 3(a), if λ_p is adjacent to λ_t the crosstalk contribution is adjacent (X_A), otherwise it is non-adjacent (X_N).

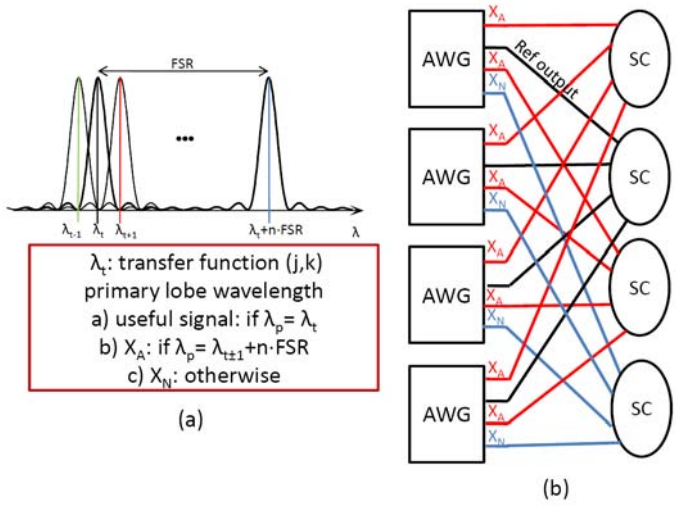


Fig. 3. *Crosstalk definition in a single AWG (a) and between different AWGs on different planes (b).*

Given the aforementioned definition of adjacent crosstalk, taking into account the reference signal sent to a plane *h*, we showed that the adjacent crosstalk contributions are generated from signals sent to the neighboring planes to *h* (Fig. 3(b)). We finally showed that the number of contributions of adjacent crosstalk (at the receivers) is always equal to min{*S*-1; 2}, while all the other contributions are non-adjacent. Then we calculated, in the worst case, how many crosstalk contributions are coherent and non-coherent.

In order to perform the analytical worst case scalability analysis PoliMi exploited the model already described in the previous deliverable. The design of the single AWG for each configuration bit/rate-channel spacing has been improved (resulting coefficients are shown in Tab. 1).

	25 GHz	50 GHz	100 GHz	200 GHz	ref.	
2.5 Gb/s	−29 dB	−36 dB	−42 dB	−48 dB	**-25 dB**	X_A
	−36 dB	−42 dB	−48 dB	−54 dB	**-30 dB**	X_N
10 Gb/s	−18 dB	−23 dB	−29 dB	−36 dB	**-25 dB**	X_A
	−23 dB	−30 dB	−36 dB	−42 dB	**-30 dB**	X_N
40 Gb/s	–	−11 dB	−18 dB	−24 dB	**-25 dB**	X_A
	–	−18 dB	−23 dB	−30 dB	**-30 dB**	X_N

Tab. 1. *Adjacent and non-adjacent crosstalk coefficients for different values of channel spacing and bit rate.*

Moreover, the same crosstalk reduction techniques described in the previous deliverable have been adopted to reduce the crosstalk impairment impact on the architecture. To recall them, strategy *S1* provides the reduction of the adjacent crosstalk, while strategy *S2* suppresses the coherent crosstalk. By combining the crosstalk analysis conducted at the Y3 with the previous years scalability models, PoliMi computed total aggregate bandwidth by exploring five different solutions:

A) Commonly-used commercial reference values of crosstalk coefficient (X_A =-25 dB and X_N =-30 dB);

B) Values of crosstalk coefficient resulting from the ad-hoc optimization of the AWG, according to actual channel spacing and channel bit-rate (X_A and X_N from Tab. 1);

C) Reduced adjacent crosstalk exploiting strategy *S1*;

D) Suppressed in-band (coherent) crosstalk exploiting strategy *S2*;

E) C and D strategies combined together.

Total aggregate bandwidth evaluation has been carried out considering $OSNR_{TX}$ values at the transmitter ranging from 40 dB to 70 dB.

We first choose a channel spacing value equal to 100 Ghz which corresponds to the standard DWDM channel spacing (0.8 nm) and we show scalability results regarding the various AWG designs for a bit rate equal to 40 Gb/s (Fig. 4). We evaluate the results in terms of architecture aggregated bandwidth. We make a comparison between the multi plane architecture (MP) and a single plane architecture (SP) in which a single AWG with *N* inputs is deployed. AWG dimension is assumed to be bounded at 1080 inlets/outlets, according to the record results presented in Ref. [3]. This bound has obviously a direct impact on the scalability of the single plane architecture.

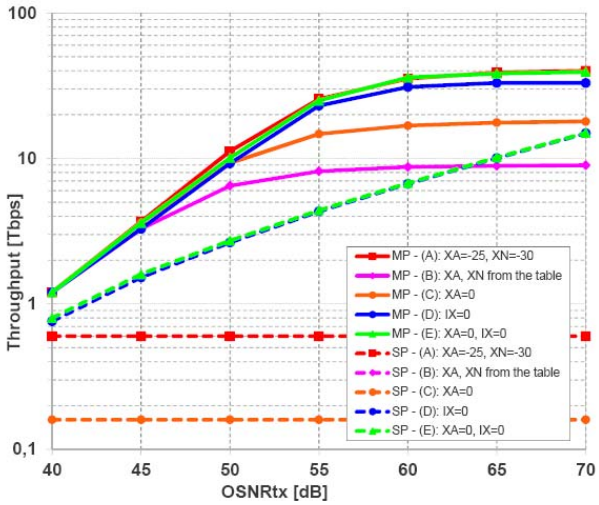


Fig. 4. *Maximum throughput VS $OSNR_{TX}$ for bit rate equal to 40 Gb/s, channel spacing of 100 GHz and different AWG designs.*

Let us now point out some observations:

- the multi plane architecture is more scalable than the single plane architecture and can reach a maximum scalability around tens of Tb/s;
- the scalability of the AWG based architecture is strongly limited, as well as for every completely passive structure, by the value of transmission OSNR because of the high noise accumulated throughout the architecture. Thus, a lot of solutions show similar aggregate bandwidths for low values of $OSNR_{TX}$.
- On the opposite, for high values of $OSNR_{TX}$ the limiting factor to the overall scalability becomes the power level at receivers. Best results are obtained when the coherent crosstalk elimination strategy is exploited (solutions D and E).

Crosstalk values in Tab. 1 are very high. In this scenario, the crosstalk given by the reference values underestimates the real crosstalk value which is provided by the design. As it should be noted, the solution *A* which employs the reference crosstalk values (X_A =-25 dB and X_N =-30 dB), shows better results compared to the solution *B*, which refers to the real crosstalk values. Hence, the adoption of these reference values should be avoided. Moreover, when the

real crosstalk values coming from the AWG design are considered, the application of the crosstalk reduction techniques is fruitful.

Finally, we realized a modular simulator working with the transfer functions of the various devices of the architecture. This simulator can be used to validate results coming from the analytical worst case estimation. Results of the comparison between the Analytical Model (AM) and the simulator (SIM) are shown in Fig. 5. The adopted design solution is *B*. As it can be appreciated the simulator validates our analytical model and the worst case estimation, for the printed outputs, is always conservative.

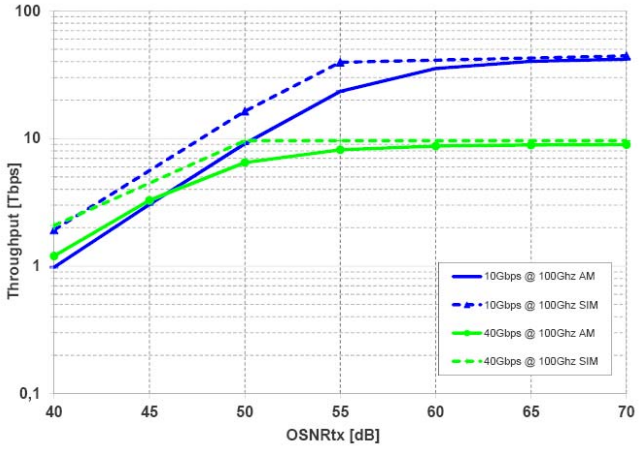


Fig. 5. *Comparison between the aggregated bandwidth obtained through the Analytical Model (AM) and the simulator (SIM); channel spacing of 100 GHz and AWG design B.*

References:

[1] R. Gaudino, G. Gavilanes Castillo, F. Neri, and J. Finochietto, "*Simple optical fabrics for scalable terabit packet switches*", in IEEE International Conference on Communications, 2008. ICC '08, May 2008, pp. 5331-5337.

[2] D. Cuda, R. Gaudino, G. Gavilanes Castillo, F. Neri, G. Maier, C. Raffaelli and M. Savi, "*Capacity/cost tradeoffs in optical switching fabrics for terabit packet switches*", in International Conference on Optical Network Design and Modeling, 2009. ONDM 2009, Feb. 2009, pp. 1-6.

[3] Y. Hibino, "*Recent advances in high-density and large-scale AWG multi/demultiplexers with higher index-contrast silica-based PLCs*", IEEE Journal of Selected Topics in Quantum Electronics, vol. 8, no. 6, pp. 1090-1101, nov. 2002.

Crosstalk-limiting and crosstalk-preventing scheduling in AWG-based switches

During this third year the partnership between PoliTO and UVIGO has strengthened, through the work on two joint papers on the design and evaluation of crosstalk limiting schedulers for AWG-based optical switches. The first paper has already been submitted and accepted at the IEEE Global Communication Conference (GlobeCom 2010) and the second one is under development.

Optical switching fabrics are gaining interest as forwarding backplanes in multi Terabit switching devices; the advantages over their electronic equivalents are mainly their information density and power consumption. AWGs are promising optical devices proposed by the academic and industrial community to build optical switching fabrics. Because of their

wavelength routing property, AWGs allow wavelength reuse over different ports introducing in-band crosstalk, which strongly limits scalability of AWG-based backplanes. However, this effect can be mitigated or even completely avoided by means of proper scheduling algorithms. Continuing with the work initiated on 2009, we modify several well-known maximal size matching scheduling algorithms [1][2][3] which limit, or even eliminate, the effect of coherent crosstalk in AWG-based switching fabrics while achieving good performance in terms of throughput and delay.

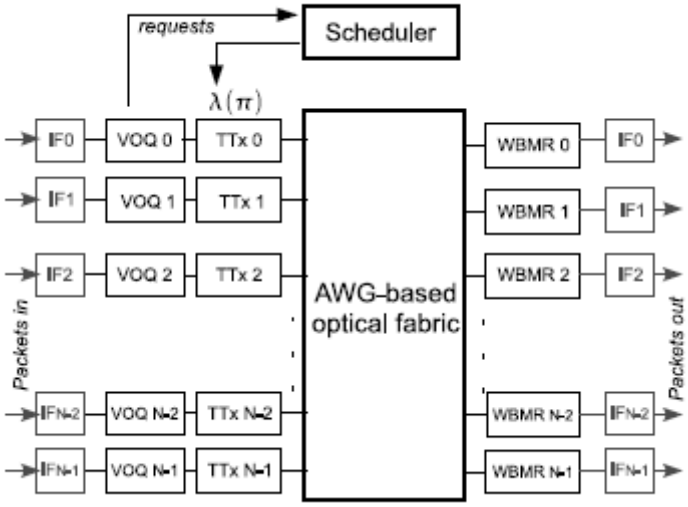


Figure 1 *Reference architecture.*

Figure 1 shows the optical switching fabric we are considering. It consists of an AWG connecting *N* tunable transmitters (TTx) and *N* fixed receivers. This architecture uses a single wavelength per port, derived from the single transceiver architecture, but given that cells can be received on any wavelength, since they can come from any port, receivers must be wideband and operate in burst mode (Wideband Burst Mode Receiver – WBMR). The optical fabric does not include any active switching element: packet switching is actually controlled by the tunable transmitters (implemented with fast tunable lasers) and exploits the wavelength routing property of the AWG-based optical fabric.

The switch is assumed to be synchronous and line cards process packets electronically by implementing input queues and other packet operations at the line card bitrate; this keeps electronic complexity within the single port-speed limit. The scheduler uses Virtual Output Queuing (VOQ) at inputs to avoid the Head of the Line blocking problem. At each time slot, a scheduler controls the switch queues transferring fixed size cells from at the most *N* inputs to *N* outputs taking up to one scheduling decision per time slot, since no speedup is available. At each scheduling decision a matching is defined such that at most one cell is transferred from each input port and to each output port. Thus, each scheduling decision is an input-output permutation. If an optical switching fabric based on AWG is used to forward packets, the scheduler must provide both the input-output permutation and its wavelength assignment.

We say that a certain input-output permutation is *k*-legal if no wavelength is used more that *k* times according to the given wavelength assignment. Hence, *k* represents the maximum number of times a single wavelength can be used at different input ports during the same time slot. The problem of constraining wavelength reuse to control crosstalk in AWG-based switching fabric has been addressed and solved in [4].

Crosstalk limiting schedulers in AWG-based optical switches (PoliTO, UVIGO)

The focus of this work is on modifications of iterative maximal size matching algorithms 2DRR[1], iSLIP4.9[2], RDSRR [3], which usually ensure good performance and are simple enough to be implemented in hardware. In order to cope with the wavelength constraint, the proposed schedulers introduce a λ-phase that ensures that the selected permutations are *k*-legal, i.e., that each wavelength can be used to transmit packets from inputs to outputs at the most k times during the same time slot.

In the algorithms that UVIGO and PoliTO proposed, each input (output) is identified by a pointer, or arbiter, p^I_i (p^O_j), with $0 \le i \le N-1$ ($0 \le j \le N-1$). In addition, wavelengths are identified by λ-pointers (or λ-arbiter) p^W_w and by λ-counters c^W_w (first initialized to zero), with w being the wavelength identifier ($0 \le w \le N-1$). The detail of the algorithms is explained in the following.

The λ-2DRR scheduler follows the same logic of the 2DRR algorithm, but it exploits the N generalized anti-diagonals to scan over the request matrix. The selection of the anti-diagonals as a covering set of the request matrix implies that all the input-output pairs belonging to the same anti-diagonal can be transmitted using a different wavelength. Thus, the λ-2DRR minimizes the probability of contentions when choosing a wavelength. To ensure fairness, at each time slot, the initial anti-diagonal changes according to a fixed round-robin scheme. At the beginning of each time slot all the λ-counters are initialized to $c^W_w = 0 \ \forall$ w and each time wavelength w is used to match an input to an output c^W_w is incremented by one. The λ-2DRR selects input-output permutations (i, j) which belong to the chosen generalized anti-diagonal, provided that, the element (i, j) of the request matrix is set to 1, input i and output j are both free and their selection does not violate the *k*-legal constraint, i.e., the usage of wavelength $w = j - i \bmod N$ can be granted if $c^W_w < k$.

The Centralized-iSLIP and Distributed-iSLIP scheduler are modified versions of iSLIP. In both cases, the Request and Accept phases of iSLIP remain unchanged, whereas the Grant phase is modify to cope with wavelength constraints.

The main gear of the *centralized-iSLIP* (C-iSLIP) is the λ-vector, an array containing the N λ-counters c^W_w. Thus, each of its elements is associated with a specific wavelength and records the number of times that a wavelength has been used during the current time slot. All the elements of the λ-vector are initialized to $c^W_w = 0$ at the beginning of a time slot. The Request and Accept phases do not change. However, the Grant phase changes as follows: sequentially, each unmatched output j selects, among all the requests, the one coming from the highest priority input i (indicated by p^O_j) and it sends a request to the λ-vector for λ_w, with $w = j - i \bmod N$. The scheduler checks c^W_w and it grants λ_w to the requiring output if $c^W_w < k$, with k being the maximum number of times a wavelength can be used at each time slot; if λ_w is granted, c^W_w is incremented by one. If λ_w can not be used because its granting violates the *k*-legal constraint the λ-vector does not grant the requiring output indicating that the wavelength request is Not Acknowledged (NACK).

Once an output receives the ACK/NACK for the requested wavelength, the next output of the round-robin scheme is served. To be fair, at each time slot, the first output sending a request to the λ-vector is selected according to a round-robin scheme. After all the outputs have sent a request to the λ-vector, outputs which are still not matched to any wavelength (received a NACK) send another request for the wavelength corresponding to the input with the next highest priority. The Grant phase terminates either when all the outputs are matched to one wavelength, or when there are no requests from inputs left. The algorithm can perform several

iterations, repeating the same procedure considering only unmatched inputs and outputs during previous iterations. The main issue of C-iSLIP is that it cannot be executed in parallel (all the outputs need to access sequentially the λ-vector). The Distribute-iSLIP and the λ-RDSRR schedulers solve this problem.

In the *Distributed-iSLIP* (D-iSLIP) each of the *N* wavelengths is associated with a λ-pointer (p^W_w) and a λ-counter (c^W_w). Each output *j* is equipped with an additional pointer q^O_w that manages priorities among the different wavelengths. In this case, the Grant phase is divided in two steps. During the first step (from outputs to wavelengths), the outputs forward the requests received (during the Request phase) to the corresponding λ-pointer according to this expression: $w=(j-i) \bmod N$. In the second step (from wavelengths to outputs and from outputs to inputs), each output collects the grants or NACKs coming from the λ-pointers. If an output receives a grant from more than one wavelength, it selects the grant corresponding to the first wavelength with the highest priority in the round-robin scheme (indicated by q^O_w). Then, each output sends a grant to the input associated with this wavelength. Pointers p^W_w and q^W_w are incremented (modulo *N*) one location above the granted wavelength if the grant is accepted during the Accept phase (Accept messages are forwarded by output pointers to wavelength arbiters). The two steps of the Grant phase are interleaved with the λ-phase. During the λ-phase, each pointer p^W_w receives requests forwarded by outputs during the first step of the grant phase. To respect the *k*-legal constraint, each pointer $p^W w$ can send up to *k* grants back to the outputs starting from the one with the highest priority according to the round-robin scheme. Each counter c^W_w tracks the number of times each wavelength is used. For all the other requests, the λ-pointers do not grant the requesting outputs, indicating a NACK.

Finally, the *λ-RDSRR* algorithm is an adaptation of the Rotating Double Static Round-Robin (RDSRR) scheduler [3] which differs from iSLIP in the pointer updating rule and because unlike the D-iSLIP, λ-RDSRR selects the wavelength at the inputs. Again, each wavelength is associated with a λ-pointer p^W_w and with λ-counter c^W_w, but the wavelength selection is performed at the inputs. Initially, the RDSRR initializes the input and the output pointers to $p^I_i = (-i) \bmod N$ and $p^O_j = (-j) \bmod N$, respectively. λ-pointers are set to $p^W_w = w$. Note that each input points to an output that reciprocally points to that same input, so that wavelength assignment is selected to give priority to those pairs. At each time slot the λ-RDSRR performs five phases. The Request and Grant phases are identical to iSLIP (except that pointer updates are independent of assignations) and the Accept phase is postponed to the λ-phase, which is composed by the following two steps. The first step is the Wavelength-Request step: if an input receives one or more grant, it selects the one which appears next in a fixed round-robin schedule, starting from the input pointer p^I_i, and requests the associated transmission wavelength λ_w to the corresponding wavelength pointer according to expression $w=(j-i) \bmod N$. The second step of the λ-phase is the Wavelength-Grant step; according to the *k* value, each wavelength pointer p^W_w grants the first $k–c^W_w$ requests following the fixed round-robin schedule starting from the first highest priority input (as indicated by p^W_w). The counter c^W_w is updated to reflect packet assignations, i.e., it is updated only if the input-output permutation has been selected by the scheduler during the Accept phase. The scheduler can run more iterations, at each iteration considering only inputs and outputs which are still free and wavelengths which are still compliant with the *k*-legal constraint ($c^W_w < k$). To maintain the initial pointers' scheme which gives priority to input-output pairs on the anti-diagonal, at each time slot p^I_i and p^O_j pointers are incremented (modulo *N*) while p^W_w pointers are decremented (modulo *N*) regardless of the packet assignments. Furthermore, the search direction is reversed each time slot to improve fairness in case of non-uniform traffic.

The performance of these four algorithms was evaluated by simulation. Figure 2 shows the results, under the following assumptions: Each input is equipped with N queues, one for each output, with a capacity of 10.000 cells at each queue. C-iSLIP, D-iSLIP and λ-RDSRR iterates $\log_2(N)$, which usually is enough to ensure that the found matching is maximal. By definition, λ-2DRR iterates *N* times. In the figure, *N* indicates the number of switch ports, *k* indicates the *k*-legal constraint and *I* denotes the number of iterations each algorithm runs at each time slot.

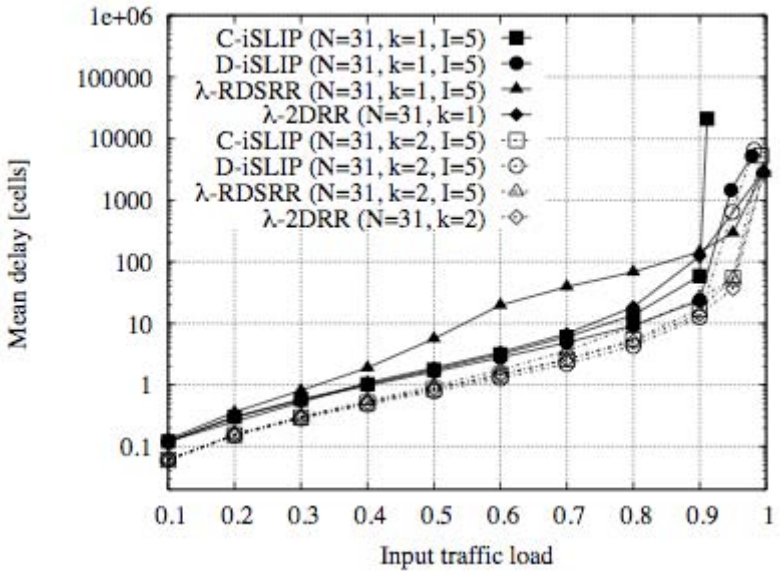


Figure 2 *Delay performance of the different algorithms under uniform traffic when either 1-legal and 2-legal permutations are allowed.*

The relaxation of the *k*-legal constraint reduces delays since it becomes easier for the scheduler to find a feasible matching. λ-RDSRR is the algorithm whose performance improves the most passing from $k = 1$ to $k = 2$, since it becomes easier to find input-output permutations which do not belong strictly to the anti-diagonals. Indeed, the λ-RDSRR pointer updating rule is fixed and does not depend on the packets forwarded by the switch. C-iSLIP performance improvement is remarkable too, since it is not able to achieve 100% when $k = 1$, while it ensures 100% of throughput when $k = 2$. The D-iSLIP and the λ-2DRR schedulers proved themselves less sensitive to the *k*-legal constraint, indeed, only a marginal improvement in the delay can be observed.

As a conclusion, the proposed heuristics allow to drastically overcome the in-band crosstalk limitation to scalability of AWG-based switching fabrics, without worsening performance neither increasing significantly the scheduler complexity; thus making AWG a viable solution to build future all-optical switching fabrics.

Once we have established these research bases, in the near future we plan to extend the evaluation to compare the different algorithms not only in terms of delay but also in terms of "fairness" (which algorithm is more equitable with all input queues); study the physical scalability, and, finally, evaluate performances when iterations are also included in the λ-phase. We expect to publish new research results along these lines in a joint paper.

In the same framework of the AWG-based switch in Figure 1, PoliTO studied [2] the properties and the existence conditions of switch configurations able to control coherent crosstalk. The presented results show that, by running a properly constrained scheduling algorithm to avoid or minimize crosstalk, it is possible to operate an AWG-based switch with large port counts without significant performance degradation.

Firstly the crosstalk-constrained scheduling problem was formally formulated. By choosing a proper value for k, crosstalk figures can be controlled by a proper switch control and large port counts become feasible; so PoliTO explored the basic properties of the permitted switch configurations.

For 1-legal permutations, PoliTO showed that a difference exists between odd and even values of N and studied the performance obtainable with single-stage AWG switching architectures. PoliTO showed that uniform traffic patterns can be scheduled using 1-legal permutations with no speedup for odd N and with a small speedup with even N.

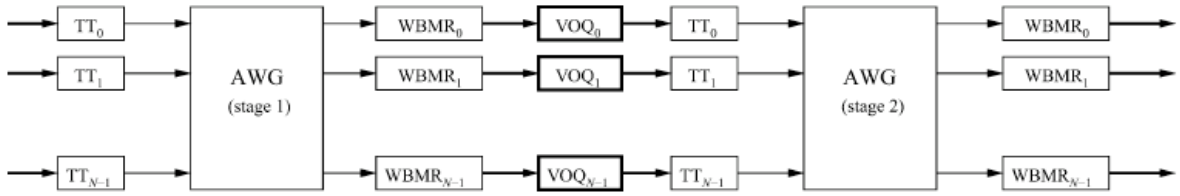


Figure 3 *Two-stage AWG-based switch. Electronic paths and components are highlighted with thick lines*

Considering general traffic patterns, PoliTO showed that they can be scheduled instead with 1-legal permutations using two-stage switches (see Figure 3) using the same small speedup; the use of VOQs between the two switching stages, and cell resequencing at outputs is required. A decomposition algorithm proposed by PoliTO would allow any scheduler output to be decomposed into two 1-legal permutations corresponding to each stage. This study also showed that 2-legal permutations permit to avoid intermediate VOQs and resequencing problems for small values of N. Finally, PoliTO formally proved that a 2-stage AWG-based switch could be configured with pairs of 4-legal permutations (or 3-legal permutations in case that N is a prime number) with no buffering between the two stages; the switch control scheme proposed is based on a quadratic decomposition algorithm. This is a remarkable result, since it means that the physical limitations caused by coherent crosstalk can be limited to a very low value using in a two-stage AWG-based switches with any number of ports. This contribution was published in IEEE/ACM Transactions on Networking [3].

Nanosecond Optical Switch built on Commercial Devices (UVIGO)

Once we know that with the proper scheduling algorithms AWG-based fabrics were able to provide good performances in terms of throughput, delay and crosstalk, we worked on a proposal to build all-optical AWG-based switches fully made from commercial available devices.

Commercial devices introduce power losses on signals that degrade the maximum rate at which we can transmit over fibers. Furthermore, using AWGs to multiplex signals on different wavelengths in one fiber in a DWDM scenario is limited to 16 or 40 wavelengths due to commercial specifications. However, new commercial devices in optical technologies, such as Lithium Niobate Interferometric switches, are able to switch two optical signals below nanoseconds. This feature let us build large switches that fit requirements of new

feasible optical switching paradigms, as Optical Cell Switching [6]. Figure 4 shows the architecture of the proposed switch.

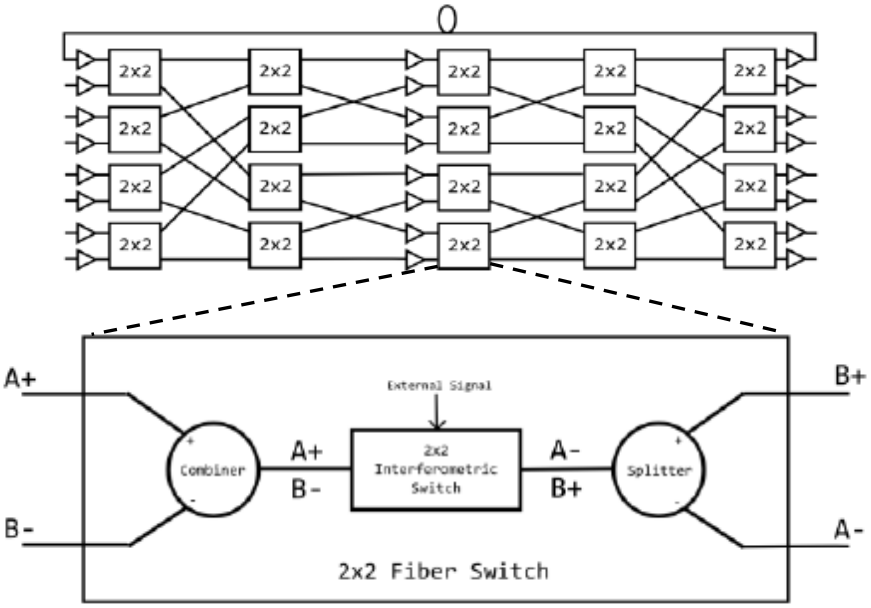


Figure 4 *Architecture of a 8x8 switch built from 2x2 interferometric switches clos network.*

We characterized the architectural physical devices from commercial datasheet values, calculated the maximum baud rates achievable for different switch sizes and showed the maximum performance achievable by the proposed all-optical network. As conclusion, we found that in well-known network topologies like NFS-Net and Geant2, is possible to reach nearly 1 Tbps per fiber using WDM technology and the proposed commercially available switches. That is, it is possible to have 16- port core switches that can support up to 16 Tbps of aggregated bandwidth.

This work has been submitted to IEEE International Communications Conference (ICC' 2011) [7].

References

[1] R.0. LaMaire and D.N. Serpanos. Two-dimensional round-robin schedulers for packet switches with multiple input queues. *IEEE/ACM Transactions on Networking*, 2(5):471–482, Oct 1994.

[2] N. McKeown. islip: A scheduling algorithm for input-queued switches. *IEEE/ACM Transactions on Networking*, 7(2):188–201, Apr 1999.

[3] Y. Jiang and M. Hamdi. A fully desynchronized round-robin matching scheduler for a voq packet switch architecture. In *IEEE workshop on high performance switching and routing (HPSR)*, pages 407–412, Paris, France, Jun 2001.

[4] J.M. Finochietto, R. Gaudino, G.A. Gavilanes, and F. Neri. Simple optical fabrics for scalable terabit packet switches. In *International Conference on Communications (ICC)*, Beijing, China, 2008.

[5] M. Rodelgo-Lacruz, C. Lopez-Bravo, F.J. Gonzalez-Castan, and H.J. Chao, Practical scalability of wavelength routing switches. In *International Conference on Communications (ICC)*, Dresden, Germany, 2009.

[6] H.J. Chao, S.Y. Liew, "A New Optical Cell Switching Paradigm", *International Workshop on Optical Burst Switching*, Dallas, TX, Oct. 2003.

[7] D. Fernández Hermida, M. Rodelgo Lacruz, C. López Bravo, F.J. González Castaño, "Nanosecond Optical Switch built on Commercial Devices" submitted to *IEEE International Conference on Communications ICC 2011*.

Collaborative actions carried out:

Meetings (including tele-conferences)

- 8 March teleconference. Participants: Miguel Rodelgo (UVIGO), Guido Gavilanes (PoliTo) and Davide Cuda (PoliTo). Preparing the first joint paper, accepted at GlobeCom.
- 26 March teleconference. Participants: Miguel Rodelgo (UVIGO), Guido Gavilanes (PoliTo) and Davide Cuda (PoliTo). Preparing the first joint paper, accepted at GlobeCom.
- 29 March teleconference. Participants: Miguel Rodelgo (UVIGO), Guido Gavilanes (PoliTo) and Davide Cuda (PoliTo). Preparing the first joint paper, accepted at GlobeCom.
- 5 October teleconference. Participants: Miguel Rodelgo (UVIGO), David Fernández Hermianda (UVIGO), Guido Gavilanes (PoliTo) and Davide Cuda (PoliTo). Preparing the extention of the paper, for a new joint publication.
- 20 September, Technical Meeting. Partecipants: Achille Pattavina (PoliMi), Guido Maier (PoliMi), Domenico Siracusa (PoliMi), Guido Gavilanes (PoliTo), Davide Cuda (PoliTo) and Fabio Neri (PoliTo).

Mobility actions

1. David Fernández Hermida, PhD Student (UVIGO) visited PoliTo from 22/11/2010 to 30/11/2010. The purpose of this mobility action was to continue the collaboration initiated with PoliTO two years ago concerning the problem of preventing coherent crosstalk impairments in AWG-based optical switches. In this occasion, new schedulers, new traffic scenarios and new convergence analysis, were considered.

Overall assessment of work carried out within the project duration:

It can be stated that the objective of the Joint Activity has been reached. All the partners involved have investigated the topic of low-crosstalk optical packet-switching architectures based on wavelength-switching and wavelength-sensitive devices. PoliMi focused on card-to-card low-crosstalk optical interconnections based on Arrayed Waveguide Grating; PoliTO and UVIGO focused on crosstalk-limiting and crosstalk-preventing scheduling in AWG based switches; UPCT evaluated the effects of coherent crosstalk in the input-buffered wavelength-routed switch (first two years).

Outcome of the joint research activity:

- Joint Publications

- Bianco, D. Cuda, G. Gavilanes Castillo, F. Neri, M. Rodelgo Lacruz, F.J. González Castaño, C. López Bravo, M. Salvat, "Crosstalk limiting schedulers in AWG-based optical switches", *IEEE Global Communications Conference*, Miami, FL, Dec. 2010.

- Partner Publications

- D. Siracusa, G. Maier, V. Linzalata, A. Pattavina, "*Scalability of Optical Interconnections based on the Arrayed Waveguide Grating in High Capacity Routers*", 15th conference on Optical Network Design and Modeling (ONDM 2010), Bologna, *In printing*.
- D. Hay, A. Bianco, F. Neri, "Crosstalk-Preventing Scheduling in AWG-Based Cell Switches", *IEEE Global Communications Conference (GLOBECOM'09)*, Optical Networks and Systems Symposium,} Best Paper Award, Honolulu, Hawaii, Dec. 2009
- A. Bianco, D. Hay, F. Neri, "*Crosstalk-Preventing Scheduling in Single- and Two-Stage AWG-Based Cell Switches*," IEEE/ACM Transactions on Networking, to be published, doi: 10.1109/TNET.2010.2054105
- D. Fernández Hermida, M. Rodelgo Lacruz, C. López Bravo, F. J. González Castaño, "Nanosecond Optical Switch built on Commercial Devices" submitted to IEEE International Conference on Communications ICC'2011.

- Planned Joint Publications

A joint paper on crosstalk-preventing heuristics is planned between PoliTo and UVigo.

4.10 JA17 Experimental investigation of multi-format regeneration using a large-scale hybridly integrated photonic device

Responsible partner: ICCS/NTUA

Participants: AIT, ICCS/NTUA

Description of the work carried out in the 3rd year:

The joint activity focuses on the demonstration of a large-scale silica-on-silicon photonic integrated circuit provided by CIP Technologies for multi-format regeneration and wavelength conversion. Figure 1 depicts the layout of the photonic chip comprising a nested structure of two Mach-Zehnder Interferometers (MZIs) with four hybridly integrated Semiconductor optical amplifiers (SOAs) (numbers 1-4) at every interferometer branch, 14 on-chip phase-shifters (letters A-O), four 1-bit delay line interferometers (DLIs) with the outlined free spectral ranges (22 and 44 GHz) and a complex network of waveguides. The novel photonic platform is capable to operate with OOK and PSK signals at 22 and/or 44 Gb/s and additionally with DQPSK modulated data at 44 GBaud by simply selecting the appropriate ports of the chip for introducing the input signals and by properly setting the phase shifters of the DLIs (in the case of PSK and DQPSK) and the nested MZIs.

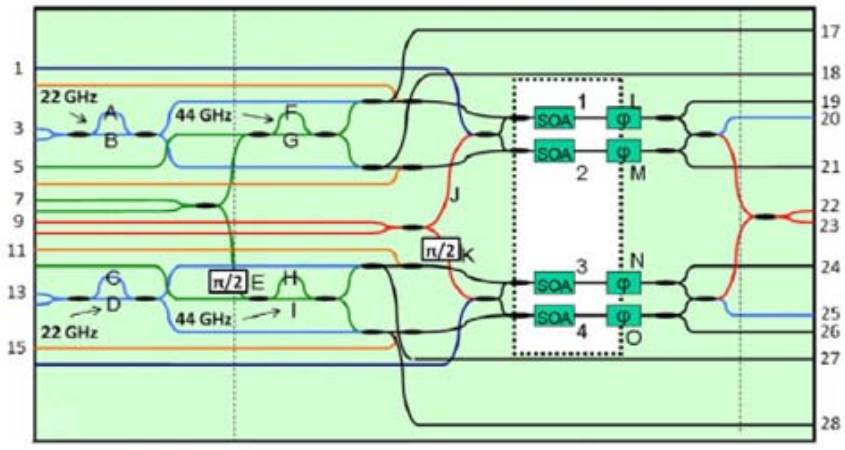


Figure 1. Layout of the processing chip

Here, we briefly report on the wavelength conversion performance of the photonic integrated chip for OOK and PSK at 22 Gb/s illustrating regenerative characteristics with respect to signal distortion by means of, (i) ASE noise and (ii) phase-noise. **Figure 2** illustrates the experimental setup. The outputs of two DFB lasers at 1551 nm (λ1) and 1558 nm (λ2) pass through an electro-absorption modulator (EAM) for pulse carving at 22 GHz generating two clock signals with 6 ps pulse width. The λ1 clock is forwarded to an InP integrated IQ modulator (provided by HHI) driven by a 2^{31}-1 long pseudo-random bit sequence (PRBS) for OOK or PSK modulation depending on the bias conditions, and the λ2 clock passes through a pulse compression stage to reduce its pulse width down to 4.6 ps. The data signal (OOK or PSK) at the output of the InP integrated IQ modulator is degraded before entering the multi-functional regenerator: (i) an attenuator followed by an EDFA are used to degrade the optical signal-to-noise ratio (OSNR) by adding ASE noise to the data signal and (ii) a distorting phase modulator driven by a 5.2 GHz sinusoidal signal with its peak-to-peak value defining

the level of distortion is used to add phase-noise to the PSK signal. In the case of OOK signals, the signal is driven to the port 2 of the device with 0.73 mW and the clock to port 9 with 0.38 mW. The output is taken from port 20, filtered, amplified and driven to a photodiode for inspection and BER evaluation after 4:1 electrical demultiplexing. In the case of PSK signals, the signal is driven to port 3 with 8.6 mW and the clock to port 9 again with 0.29 mW average power. The PSK output from port 20 is launched back to the processor through port 27 for decoding by the C-D DI (see Figure 1). The complementary decoded streams are taken from ports 13 and 14 to the evaluation unit. In all cases, the on-chip SOAs operate with ~300 mA.

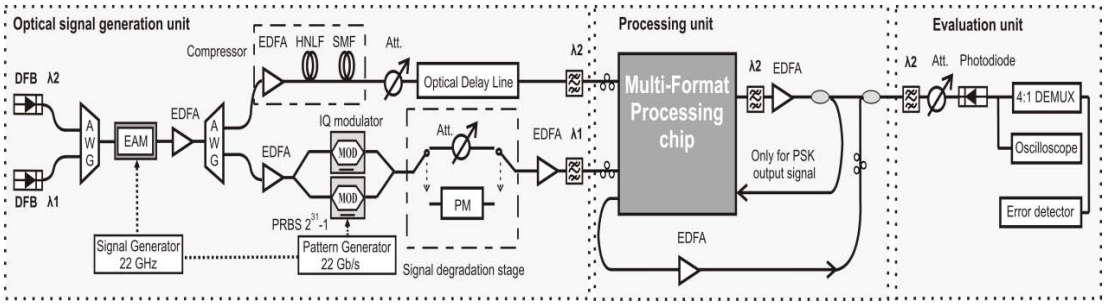


Figure 2. Experimental setup

The performance of the processor was evaluated through BER measurements, presented in Figures 3a-c. For OOK signals, we investigated the operation with ASE noise-loaded inputs, whereas for PSK signals, we studied the performance of the device with ASE and phase noise-loaded inputs. Each curve in Figure 3a corresponds to the worst performing between the four 5.5 Gb/s tributaries, whereas each curve in Figures 3b and 3c corresponds to the worst performing between the two complementary decoded streams and their four tributaries that relate to a 22 Gb/s PSK signal. The variation both between the time tributaries and the decoded streams in the case of a PSK input was negligible though due to the absence of patterning effects and the symmetry of each SOA-MZI gate, respectively. Power penalty improvement of ~1.5 dB at the level of 10^{-9} BER was obtained for an OOK input of reduced (18.3 dB) OSNR. For an input PSK signal of 18.4 dB OSNR, the power penalty improvement was ~1.1 dB, whereas it was higher (~1.4 dB) for phase noise-loaded signals. Figures 3d-p present the corresponding eye-diagrams for these measurements referring to the undegraded inputs (B-2-B), the degraded inputs, the regenerated outputs and the respective decoded streams of PSK signals. Regeneration in the case of OOK signals originates mainly from the suppression of the peak power fluctuations, but also from noise suppression at the space level. For PSK signals with ASE noise, the regeneration is associated again with the suppression of the peak power fluctuations, which is also observed in the decoded streams. On the other hand, for the PSK signal with phase noise only, the regenerator suppresses sufficiently this type of noise and allows for a decoded signal with higher extinction ratio but slightly enhanced peak power variation due to the partial conversion of the phase- into amplitude-noise.

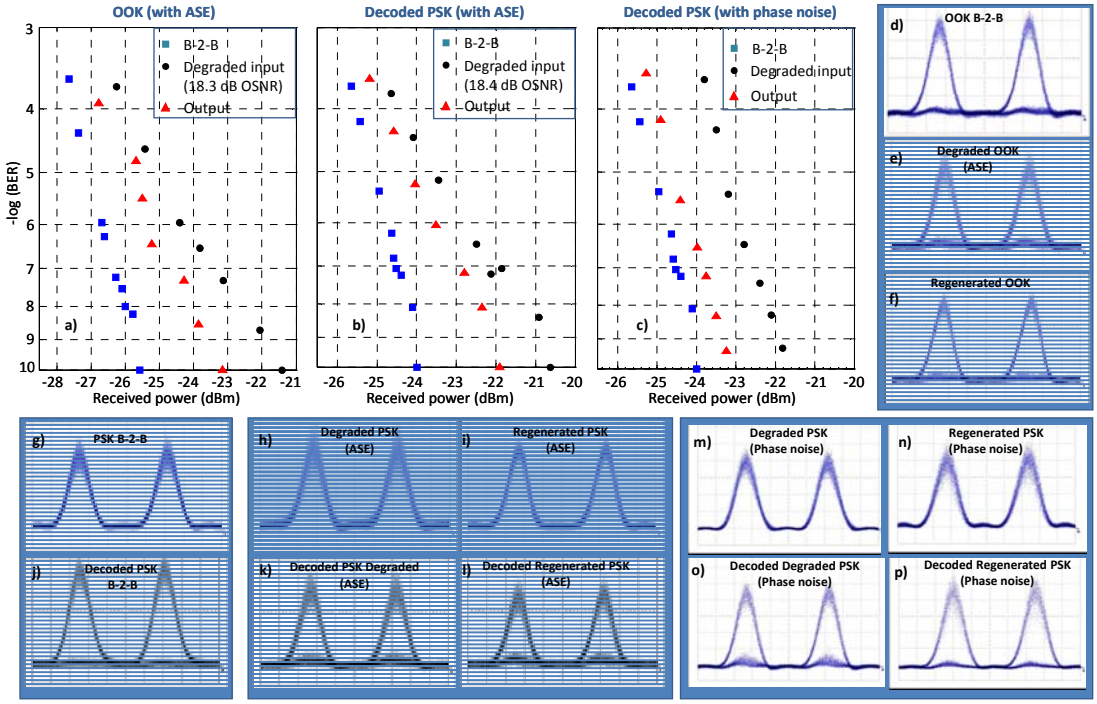


Figure 3. BER curves for: a) OOK and b) PSK with ASE noise, and c) PSK signals with phase noise. (d)-(p): Eye diagrams of OOK, PSK and the respective decoded signals for the B-2-B, the inputs and the regenerated outputs. Time scale: 10 ps/div.

Collaborative actions carried out:

- mobility actions

One researcher from AIT, Dr. Dimitrios Klonidis, visited ICCS/NTUA to experimentally validate the performance of the regenerator in collaboration with the researchers of ICCS/NTUA.

- number and details of papers 0

- number of joint experiments 1

Overall assessment of work carried out within the project duration:

During the project duration ICCS/NTUA has had active participation in total 5 mobility actions. A significant number of technical papers have been submitted as a result of joint work that contributed to the dissemination of the project outcome.

Outcome of the joint research activity:

List of papers where at least an author is from ICCS-NTUA:

1 D. Apostolopoulos (ICCS-NTUA), H. Simos (UoA), D. Petrantonakis (ICCS-NTUA), A. Bogris (UoA), M. Spyropoulou (AUTh), M. Bougioukos (ICCS-NTUA), K. Vyrsokinos (ICCS-NTUA), N. Pleros (AUTh), D. Syvridis (UoA), H. Avramopoulos (ICCS-NTUA), *A New Scheme for Regenerative 40 Gb/s NRZ Wavelength Conversion using a Hybrid Integrated SOA-MZI*, OFC 2010 Conference proceedings, Vol. OThS6, San Diego, CA, USA, March 2010.

2 D. Apostolopoulos (ICCS-NTUA), D. Klonidis (AIT), P. Zakynthinos (ICCS-NTUA), K. Vyrsokinos (ICCS-NTUA), N. Pleros (AUTh), I. Tomkos (AIT), H. Avramopoulos (ICCS-NTUA), *Cascadability Performance Evaluation of a New NRZ SOA-MZI Wavelength Converter*, IEEE PHOTONICS TECHNOLOGY LETTERS, Vol. 21, No. 18, pp. 1341-1343, September 2009.

3 K. Vlachos (RACTI), C. Raffaelli (UNIBO), S. Aleksic (TUW), N. Andriolli (SSSUP), D. Apostolopoulos (ICCS-NTUA), H. Avramopoulos (ICCS-NTUA), D. Erasme (GET), D. Klonidis (AIT), M. N. Petersen (COM), M. Scaffardi (SSSUP), K. Schulze (UPVLC), M. Spiropoulou (AIT), S. Sygletos (AIT), I. Tomkos (AIT), C. Vazquez (UC3M), O. Zouraraki (ICCS-NTUA), F. Neri (PoliTO), *Photonic in Switching: Enabling Technologies and Subsystem Design*, OSA Journal of Optical Networks, Vol. 8, No. 5, pp. 404-428 , May 2009.

4 P. Bakopoulos (ICCS-NTUA), E. Kehayas (ICCS-NTUA), A. Oehler (), T. Sudmeyer (), K. Weingarten (), K. Hansen (), U. Keller (), H. Avramopoulos (ICCS-NTUA), *An Agile Multi-Wavelength Optical Source with configurable channel spacing and CW or pulsed operation for High-Capacity WDM Optical Networks*, OFC09 Conference procedings, Vol. OWB6, USA, March 2009.

5 D. Apostolopoulos (ICCS-NTUA), D. Klonidis (AIT), P. Zakynthinos (ICCS-NTUA), K. Vyrsokinos (ICCS-NTUA), N. Pleros (ICCS-NTUA), I. Tomkos (AIT), H. Avramopoulos (ICCS-NTUA), *Demonstration of 8 Error-free Cascades of 2R NRZ SOA-MZI Wavelength Converter*, OFC09 Conference procedings, Vol. OThS2, San Diego, CA, U.S.A, March 2009.

6 D. Apostolopoulos (ICCS-NTUA), P. Zakynthinos (ICCS-NTUA), L. Stampoulidis (ICCS-NTUA), E. Kehayas (ICCS-NTUA), R. McDougal (CIP), R. Harmon (CIP), A. Poustie (CIP), G. Maxwell (CIP), R. Van Caenegem (IBBT), D. Colle (IBBT), M. Pickavet (IBBT), E. Tangdiongga (TUE), H. Dorren (TUE), H. Avramopoulos (ICCS-NTUA), *Contention Resolution for Burst-Mode Traffic Using Integrated SOA-MZI Gate Arrays and Self-Resetting Optical Flip-Flops*, Photonic Technology Letters, Vol. 20, No. 24, pp. 2024-2026, December 2008.

7 P. Zakynthinos (ICCS-NTUA), D. Apostolopoulos (ICCS-NTUA), L. Stampoulidis (ICCS-NTUA), E. Kehayas (ICCS-NTUA), A. P. Poustie (CIP), G. Maxwell (CIP), R. Van Caenegem (IBBT), D. Colle (IBBT), M. Pickavet (IBBT), E. Tangdiongga (TUE), H. Dorren (TUE), H. Avramopoulos (ICCS-NTUA), *Successful Interconnection of SOA-MZI Arrays and Flip-Flops to Realize Intelligent, All-optical Routing*, ECOC 08, Brussels, September 2008.

8 H. Teimoori (GET), D. Apostolopoulos (ICCS-NTUA), K. Vlachos (RACTI), C. Ware (GET), D. Petrantonakis (ICCS-NTUA), L. Stampoulidis (ICCS-NTUA), H. Avramopoulos (ICCS-NTUA), D. Erasme (GET), *Optical logic gate aided packet switching in transparent optical networks*, IEEE/OSA Journal of Lightwave technology, Vol. 26, No. 16, pp. 2848-2856 , August 2008.

9 C. Raffaelli (UNIBO), K. Vlakos (RACTI), P. Pavon Marino (UPCT), N. Andriolli (SSSUP), D. Apostolopoulos (ICCS-NTUA), J. Buron (COM), R. van Caenegem (COM), G. Danilewicz (PUT), J. G. Garcia-Haro (UPCT), D. Klonidis (AIT), M. O'Mahony (UEssex), G. Maier (PoliMI), A. Pattavina (PoliMI), S. Ruepp (COM), M. Savi (UNIBO), M. Scaffardi (SSSUP), I. Tomkos (AIT), A. Tzanakaki (AIT), L. Wosinska (KTH), O. Zouraraki (AIT), F. Neri (PoliTO), *Photonics in Switching: Architectures, Systems and enabling Technologies*, Elsevier Computer Networks, Vol. 52, No. 10, pp. 1873-1890, The Netherland, July 2008.

10 H. Teimoori (GET), D. Apostolopoulos (ICCS-NTUA), K. Vlachos (RACTI), C. Ware (GET), D. Petrantonakis (ICCS-NTUA), L. Stampoulidis (ICCS-NTUA), H. Avramopoulos (ICCS-NTUA), D. Erasme (GET), *Physical architectures for packet-switching network nodes based on non-linear logic gates*, 6th International Symposium on Communication Systems, Networks and Digital Signal Processing, CNSDSP 2008, Vol. 1, No. 1, pp. 676 – 679, June 2008.

4.11 JA 18 - Comparison of asynchronous vs synchronous paradigms on optical switches (single partner activity).

Description of the work carried out in the 3rd year:

Local research related to the topics of WP14 was performed at PoliTO with BONE support. During the last reporting period, PoliTO mainly worked on the comparison of asynchronous vs synchronous paradigms on optical switches. Different directions were explored, such as the performance evaluation by means of simulation, and the dimensioning of electronic and optical resources at an architectural level. In this final deliverable, three research activities performed at PoliTO within BONE WP14 are briefly outlined.

Asynchronous vs Synchronous Input-Queued Switches

In [1] PoliTO compared the performance of asynchronous (ASY) switches with respect to synchronous (SYN) switches under realistic traffic scenarios. The results show that throughput gains/losses depend on the considered traffic scenario; but ASY can also outperform SYN depending on the traffic conditions. These results were submitted for publication to Globecom 2010.

On an asynchronous switch, at any time slot, a centralized scheduler chooses the cells to transfer from the available input queues, satisfying the crossbar constraints: (i) no more than one cell per input and (ii) no more than one cell per output. The scheduler can operate either in cell mode (CM) or in packet mode (PM). In CM, scheduling decisions are taken on each individual cell, so that packets can be interleaved when traversing the crossbar. In PM, scheduling decisions are taken on groups of cells belonging to the same packet, and packets cannot be interleaved at the outputs.

PoliTO compared the ASY and SYN performance by means of queuing models and simulation using realistic traffic variances of the packet size, that is, considering the variation coefficient of the packet length α as a parameter. It can be shown that for real packet size distributions, as those obtained from traffic traces, $\alpha < 2.3$.

This study was performed comparing different schedulers to select the input packet to be transferred to a free output port: random (RND) scheduler and round robin (RR). When comparing the type of switch, PoliTO tested bufferless switches, single FIFO and VOQs. Figure 12 shows the maximum throughput reached as a function of α for a single FIFO switch. In the case α is very low, all the packet have practically the same size, and the maximum throughput for an ASY switch is near 58% as in the SYN architecture. This is not surprising, since even if the arrivals in an ASY switch are time-continuous, the queuing effect tends to synchronize the services among all the outputs and, after a transient period, the system behaves like a SYN switch in saturation. As α becomes larger the maximum throughput goes to zero. This result highlights that throughput degradations due to ASY mode can be very high, but this happens for a large α: only for $\alpha > 2$ the throughput remains smaller than 30%.

For any realistic value of α, the estimated throughput is larger than 0.4. So in practical, these results show that, depending on the traffic conditions, an ASY switch can be better or worse than a SYN switch and, in the worst case, the throughput degradation due to the ASY behavior is limited.

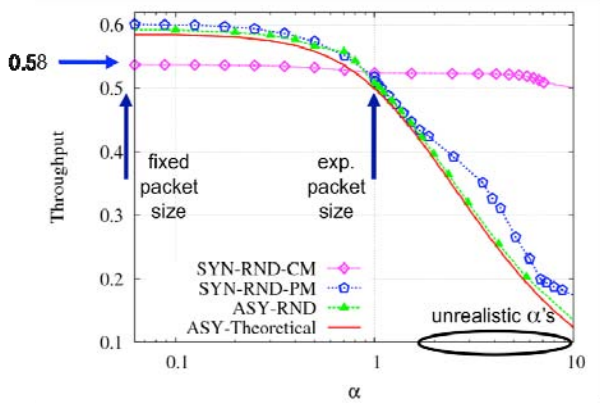


Figure 12. Maximum throughput for single-FIFO ASY and SYN switches under uniform traffic for a 100 x 100 switch.

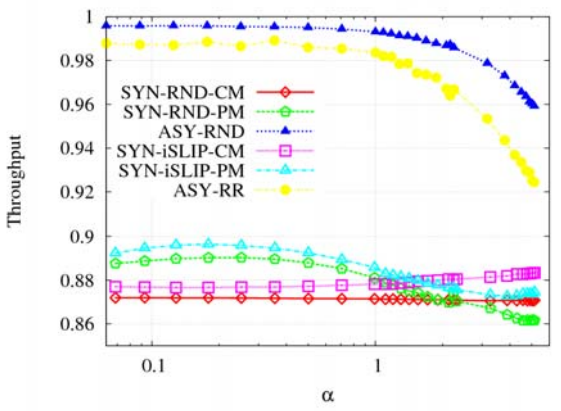


Figure 13. Maximum throughput under bidiagonal traffic vs. packet length variation coefficient, with

Figure 13 shows the throughput performance for non-uniform bi-diagonal traffic, that is, from input *i* only to outputs *i* and *i*+1. PoliTO considered the RND and RR schedulers in ASY switches and the corresponding schedulers in SYN switches, in both CM and PM. The main result is that ASY switches are always outperforming SYN switches, for any $\alpha \leq 5$. It is also worth to note that ASY, for fixed sized packets ($\alpha = 0$), achieves almost the maximum throughput. Asynchronous schemes prove to be intrinsically more capable of adapting to unbalanced traffic conditions. This is an interesting result, also considering that bidiagonal traffic is known for being a worst-case test for scheduling algorithms. This result holds even when using random schedulers, which are much simpler than the iSLIP or maximum weight matching MWM schedulers, normally considered for SYN switches.

More details on this activity can be found in publication [1] .

Optics vs. Electronics in High Capacity Switches

In the work [2] PoliTO analyzed a hybrid (electro/optic) switching architecture, relying on an all-optical switching fabric, with the aim of understanding the trade-offs between current photonic and electronic technologies. PoliTO considered both the fixed-size synchronous paradigm typical of high-end electronic switches and the asynchronous operation mode, which may be easier to implement, especially in a full photonic switch. Buffers organized according to either an electronic FIFO paradigm or a photonic FDL mechanisms are analyzed and compared.

The considered architecture is depicted in the Figure 14. The architecture comprises *N* input and *N* output ports, which, for simplicity, are assumed to operate at the same speed. *M* re-circulating delay lines, including buffers, are available. Thus, the switching fabric has *N*+*M* input/output lines.

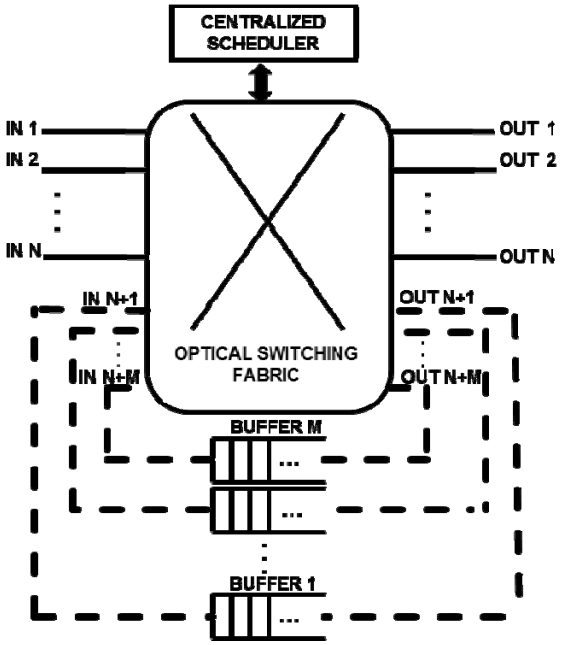


Figure 14. Considered switching architecture

The packet switch is built around an all-optical non-blocking switching fabric, which may be implemented using SOA gates or other fast photonic technologies. This helps in reducing the required energy and the power dissipation; furthermore, the switching complexity does not depend on the transmission bit rate. Assuming that the optical fabric scales well in the number of I/O lines, *M* extra lines can be used to interconnect buffers, thereby providing a form of spatial speed-up with respect to the *NxN* fabric. The overall switch capacity is increased from *N* to *N+M* packets per packet time; hence, a *(N+M)/N* speed-up us offered.

Although the basic switch architecture was defined, several design choices are still available, leading to a set of possible switch configurations. The two paradigms SYN and ASY were considered and two different re-circulation buffers: photonic Fiber Delay Lines (FDLs) vs. electronic FIFO. Also two buffer management policies (dedicated vs. shared) were considered.

However, in all architectures, the existence of a centralized scheduler was assumed to select packets to be transferred to output ports. When contentions arise, priority is always given to packets available in the re-circulating buffers.

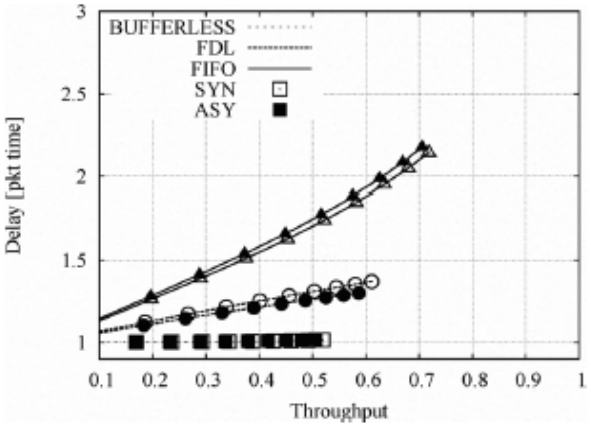


Figure 15. Delay vs. throughput when buffers are managed using a dedicated buffer strategy

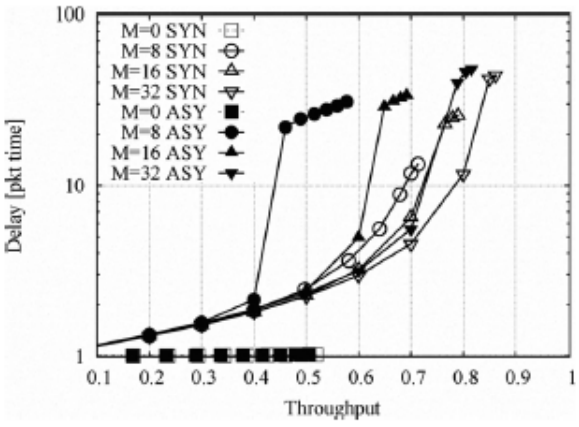


Figure 16. Delay vs. throughput when FIFO buffers exploits a shared buffer strategy

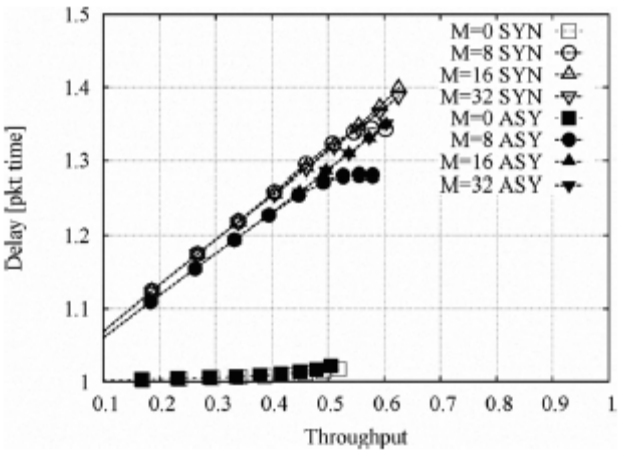


Figure 17. Delay vs. throughput when FDL lines exploits a shared buffer strategy

Some performance results were obtained by simulation considering the architecture described in Figure 14. The number of inputs/outputs ports *N* is set to 16 and each port runs at 10 Gpbs. The number of re-circulating lines *M* is set equal to 16, in the case of dedicated buffers, or 8, 16 and 32 in the case of shared buffers.

FIFOs were assumed have a storing capacity equal to 20 packets of average size, and FDLs delay a packet for a time equal to the time needed to transmit an average packet size at the port line rate. Figure 15 shows the average and maximum FIFO occupancy (values are normalized to the buffer size) when the dedicated buffer policy is adopted. On average, queues are relatively empty, even if the maximum occupancy is high. Indeed, performance is mainly limited by contentions and not by buffer overflow.

Figure 16 shows delay vs. throughput performance when the shared buffer strategy is adopted and FIFOs are employed. Compared to the dedicated strategy, the shared policy ensures higher throughput for both SYN and ASY switching modes. Note that, when the shared buffer policy is employed, packets are discarded only when buffers are completely full. Indeed, if more packets aim at the same output, one packets is transmitted, while the other packets select the buffer with minimum occupancy among available buffers (buffers to which no packet is already being transferred). Delay increases because the number of contentions increases; thus, packets can spend more time in memory.

Figure 17 shows the results when FDLs are employed. The shared buffer strategy shows no significant performance gain. The persistence of contention and the void filling problem still limit the switch throughput. Delays are kept low because packets, in the worst case, are either transmitted after a delay equal to the transmission time of an average size packet. Performance difference between SYN and ASY switching modes are less evident.

In summary, PoliTO studied in [2] a switch architecture exploiting electronic FIFO re-circulation buffers and compared its performance with the classical “all-optical” architecture, where buffers are implemented by means of FDLs. The considered switch architectures, built around an optical switching fabric, exhibits interesting performance while keeping low control complexity. Electronic FIFOs always exhibit better performance (especially if the shared strategy is adopted). The ASY switching mode does not penalize significantly switch performance with respect to the SYN case. Thus, it may be considered as a more promising architecture, given its simplicity. Furthermore, the minor throughput loss should not be

considered as a major issue in the context of optical technology given the huge amount of available bandwidth in the optical domain.

Multi-hop Scheduling in Switches with Reconfiguration Latency

In [3] PoliTO considered optical switching fabrics (OSFs) and a multihop approach to schedule packet transfers, i.e., packets are sent to their final destination port exploiting transmission through intermediate ports. The multi-hop approach is a promising technique to control the trade-off between delay and throughput, in particular in the case of slow switching technologies, as it permits to partly decouple the switch reconfiguration rate from the packet duration.

PoliTO also proposed a general framework to solve the issue of multi-hop transmission in IQ packet switches. Furthermore, the multi-hop approach was examined when the direct exchange of packets among ports is based on multi-dimensional regular topologies.

The basic idea of this study is that switch ports can be considered as input/output nodes. The multi-hop switch control scheme identifies a sequence of switching configurations that are cyclically repeated to best serve a given traffic matrix. The multi-hop switch control problem can be described as a sequence of steps.

In the first step, a virtual topology is defined as an interconnection pattern among switch ports. Given the virtual topology, a routing scheme defines the paths followed in the virtual topology by each possible flow. In the second step, a set of "covering" matchings is chosen to permit packet transmission among ports on all the logical links of the virtual topology. Finally, a sequence among the identified matchings is selected to transfer the packets across the switching fabric through the multi-hop approach. A specific example is shown in

Figure 18; a 6-port switch is assigned an arbitrary topology $\hat{G}_\sigma$; note that a direct (i.e., single-hop) packet transfer is not possible for all node pairs, e.g. between node 0 and node 2 in

Figure 18.

Given a virtual topology (a directional graph), a deterministic routing algorithm computes the shortest paths (the path length is measured in terms of number of hops to connect any pair of nodes). When more than one path may connect two nodes, the algorithm would be forced to choose always one path, to prevent mis-sequenced delivery of packets belonging to the same flow. The choice of the topology is somehow arbitrary, although other criteria to optimize the virtual topology according to traffic needs can be defined; these considerations were also discussed in this research.

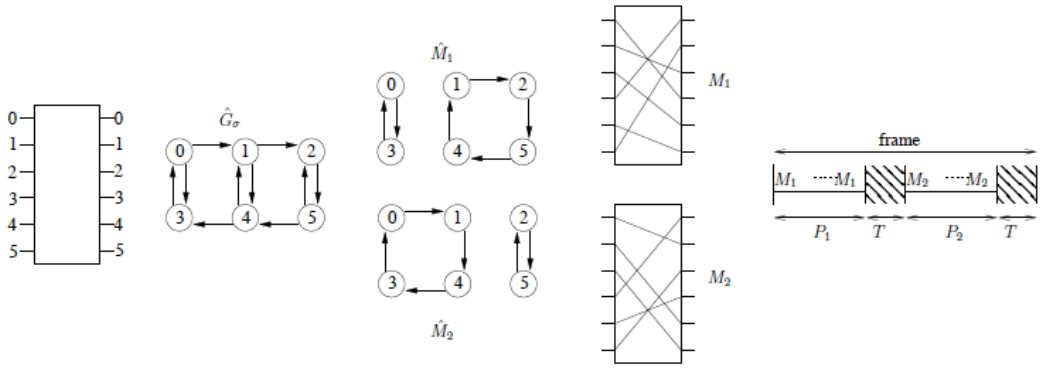


Figure 18. Example of multi-hop scheduling for a 6 × 6 switch. $\hat{G}_\sigma$ is covered by two di-matchings $\hat{M}_1$ and $\hat{M}_2$

The algorithm computes a minimal set C of directed di-matchings on $\hat{G}_\sigma$, such that the union of all the di-edges in C covers all di-edges in $\hat{G}_\sigma$. This condition guarantees that each di-edge in the virtual topology is covered by at least one di-matching. Hence, any routing path can be mapped into a sequence of di-edges belonging to di-matchings. Equivalently, throughout any sequence of all covering di-matchings, it is possible to ensure full connectivity among the nodes in $\hat{G}_\sigma$. The decomposition in di-matchings can be computed, for example by using the classical Birkhoff-von Newmann decomposition algorithm. The final scheduling (in the case depicted in

Figure 18) would be given by a frame, which will be composed by $\hat{M}_1$ for P1 time-slots and by $\hat{M}_2$ for P2 time-slots, with the corresponding reconfiguration time T between di-matchings.

Among all the topologies, in this work PoliTO analyzed in some detail multidimensional bidirectional Manhattan topologies (MH), since they permit to easily define good routing strategies, and they provide a good delay/throughput trade-off, and other highly connected single-hop topologies (SH). The performance estimation was done analytically and by means of simulation.

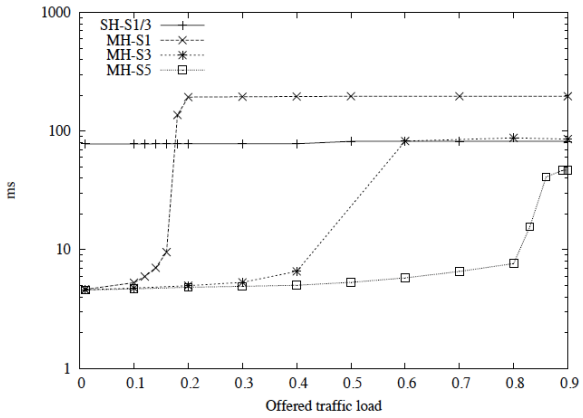


Figure 19. Average delay under uniform

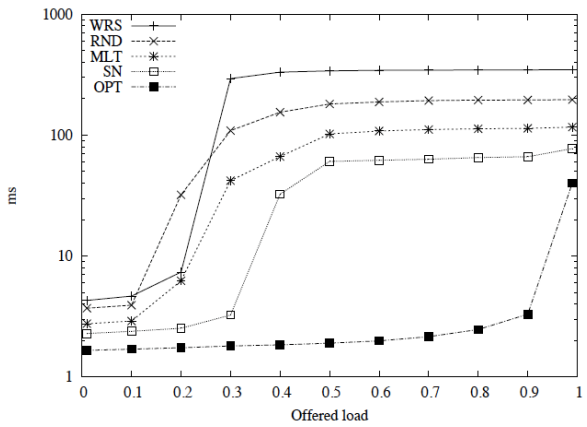


Figure 20. Average delay for node placement

traffic, for switch size N = 121, 2D multi-hop topologies, and variable speedup S.

algorithms

The throughput reduction due to multi-hop schemes can be compensated by some speedup. PoliTO considered explicitly the effect of temporal speedup. Figure 19 shows the average delay for a large switch (N = 121) for single-hop (SH) and 2D multihop (MH) schemes, in the case of variable speedup S = 1, 3, 5. PoliTO proposed approximated models to bound the average delay for low load and the maximum throughput. It is important to observe that the average delay for single-hop is not affected by the speedup, because of the scheduling approach based on frames. Throughput for multi-hop is affected by speedup: a speedup S is able to increase the maximum achievable throughput by a factor S. The delay for low load is independent from the speedup.

Three heuristics were proposed to compute the node placements in a Manhattan topology: Minimum Link Traffic (MLT), Swap Neighbors, (SN) and Random (RND) placement. Figure 20 shows the throughput and the average delay obtained by the different node placement algorithms under localized traffic for a switch size N = 25. OPT and WRS correspond to the optimum and worst-case theoretical bounds respectively. The best heuristic is SN, guaranteeing slightly more than 50% of the bandwidth available when using the optimal placement. RND provides performance very close to WRS.

Figure 20 shows the delay/throughput trade-off. This adversarial traffic scenario shows that a clever node placement algorithm may provide non-marginal performance benefits. However, from a practical point of view, it is important to understand the behaviour of the placement algorithms under more general traffic scenarios. The tests were repeated over 1000 doubly stochastic traffic matrices and running the placement algorithm on them. Although the SN heuristic still provides the best performance, the difference between the algorithms was not so significant. Thus from a practical point of view, simple node placement algorithms may only slightly reduce switch performance and are a viable solution to exploit multi-hop schemes.

The multi-hop approach is the only viable switch control technique to overcome the large reconfiguration penalty of optical devices in large-size IQ switches. Although many degrees of freedom are available in the definition of the switch control scheme, which would ask to optimize performance for particular traffic matrices, a simple approach defined for uniform traffic was shown to be robust enough to provide good performance for general traffic scenarios.

More details can be found in [3] .

Collaborative actions carried out:

- number and details of papers

[1] Bianco, D. Cuda, P. Giaccone, F. Neri, “Asynchronous vs. Synchronous Input-Queued Switches”, in *Proc. IEEE Global Communications Conference (GLOBECOM 2010)*, Next Generation Networking Symposium, Miami, FL, USA, December 2010.

[2] Bianco, D. Camerino, D. Cuda, F. Neri, “Optics vs. Electronics in Future High-Capacity Switches/Routers”, in *Proc. IEEE Workshop on High Performance Switching and Routing (HPSR 2009)*, Paris, June 2009.

[3] V. Alaria, A. Bianco, P. Giaccone, E. Leonardi, F. Neri, “Multi-hop Scheduling Algorithms in Switches with Reconfiguration Latency”, *IEEE/OSA Journal on Optical Comunications and Networking*, Vol. 1, No. 3, August 2009.

Overall assessment of work carried out within the project duration:

PoliTO has contributed actively to BONE on switching systems, since the very early stages of the project. In addition to the contributions to WP14 summarized above, PoliTO contributed to WP25, analysing the complexity and scalability of optical switching fabrics in collaboration with other partners, such as UNIBO and PoliMI, and developing new scheduling schemes, such as the crosstalk-avoiding optimal and heuristic schedulers and switch paradigms based upon AWGs, in collaboration with UVIGO.