Practical solutions for optical packet switches with shared wavelength converters: performance and complexity evaluations

Carla Raffaelli, Michele Savi

D.E.I.S.- University of Bologna - Bologna, Italy.
email:{craffaelli,msavi}@deis.unibo.it

***Abstract*— Multi-stage optical packet switching fabric based on the broadcast-and-select principle is considered to implement the shared-per-node wavelength conversion scheme. This approach fairly allows to combine contention resolution in the wavelength and space domains. A scheduling algorithm is proposed to control optical packet forwarding in synchronous context. The results aim at showing how the sharing of wavelength converters impacts on node performance, the degree of sharing achieved with respect to the reference shared-per-node architecture, and at providing a meaningful support for cost-performance benchmarking studies. An important aspect dealt with in the paper is the complexity evaluation in terms of expensive optical components, tunable wavelength converters (TWCs) and semiconductor optical amplifiers (SOAs) used as optical gates. The multi-stage switching fabric is compared with the reference shared-per-node scheme in terms of complexity, thus showing the better scalability property of the proposed architecture, given the reduced number of optical gates employed.**

Index terms: Optical packet switching, Contention resolution, Tunable wavelength converters, Optical gates, Switch complexity, Cost evaluation

I. INTRODUCTION

Photonic packet switches [1], [2], [3] are key devices in the design of dynamically reconfigurable optical networks to support statistical multiplexing in the optical layer with finer switching granularity with respect to optical lighpaths for the proliferation of future Internet based services [4], [5], [6]. One of the main and still open issues in optical packet switches is contention resolution, which arises when packets compete for the same resource at the same time. Space, time, wavelength and (now) code [7] domains are the contexts where contention can be solved. While in electronic packet switches contention is managed in the time domain through random access memories used as queues, in the optical domain this approach has limited applicability since it relies on fiber delay lines (FDLs) that offer only discrete and limited delay values while they are contributing to further signal quality degradation. On the other hand, contention resolution can be performed in the wavelength domain taking advantage of wavelength conversion [8].

Since tunable wavelength converters (TWCs) are expensive components [9], switch architectures where TWCs are shared among optical channels through a node have been proposed in literature [10], [11], [12]. Here the shared-per-node (SPN) scheme, also shown in figure 1, is considered as a reference where a pool of TWCs serves all input fibers to properly forward incoming packets to the output wavelengths [12].

It has been demonstrated that such architectures with a limited number of full-range tunable-input tunable-output TWCs can provide the same performance as the fully equipped case, thus assuring relevant cost saving [12]. The sharing of TWCs typically requires additional space switching capability to allow packets that need conversion to access the wavelength converter pool and then to reach the proper output fiber. In any case most works refer to general switch architectures as in figure 1, while only recent investigations refer to practical switch implementations and related control algorithms [12], [13]. In this paper, the activity developed in the framework of the e-photon/ONe and OSATE projects is presented to discuss practical issues to implement the shared-per-node concept by combining wavelength and space domains exploitation. A buffer-less multi-stage switch architecture is first presented.

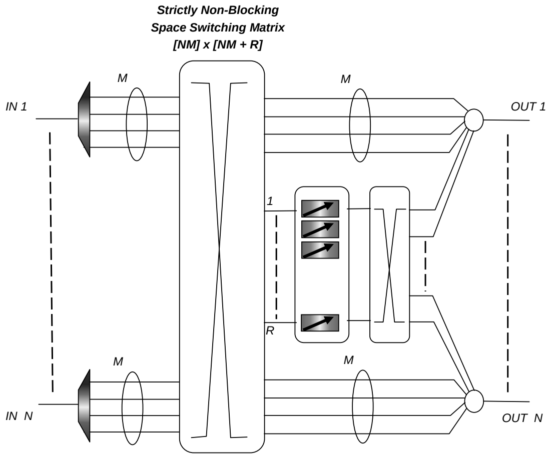

Fig. 1. SPN switch architecture with N input and output fibers, M wavelengths per fiber and limited number R of TWCs shared-per-node.

Similarly to their electronic counterparts, single stage optical cross-connects require either a large number of space switching elements, like optical gates (OGs), or tunable wavelength converters that are tuned over a large number of wavelength channels like in [14]. Both these approaches do not scale easily in terms of capacity. A multi-stage architecture facilitates in overcoming these limitations. In addition, it is taken as axiomatic that future nodes should be cross-connecting an unspecified mixture of both optical circuit switched (OCS) and optical packet switched (OPS) traffic where the OCS can be constructed by means of sequential concatenation of slots. The proposed switch could perform this task if it is operated in a synchronous packet/slot context, meaning that optical packets/slots should have fixed length and are packet level synchronized at switch input like in [15]. Moreover, since the switch is operated in a slot synchronized mode, a Rearrangeable Non-Blocking (RNB) fabric could ensure high throughput, providing that a suitable scheduling algorithm is adopted, thus leading to further hardware saving (compared to strictly non-blocking architectures).

The multi-stage switch considered here is organized as space-lambda-space (S-λ-S) architecture to keep complexity low. It employs fast optical technology such as tunable transmitters and optical gates that are expected to mature in the near future, as shown in [15]. Particularly, semiconductor optical amplifiers (SOAs) used as optical gates (OGs) are very attractive due to the tuning time in the range of the nanosecond and sufficient maturity. A scheduling algorithm is presented to assign switch resources on a time slot basis to incoming optical packets and avoid contention. Switch performance has been evaluated by means of an analytical model in [16], that takes into account the particular organization of shared wavelength converters related to the multi-stage architecture.

Here extended results, obtained by simulation, are presented to evaluate the complexity in terms of the optical components (TWCs and OGs) of the multi-stage architecture in comparison with the reference SPN architecture (figure 1). The obtained results can be used for any meaningful cost-performance benchmarking study.

The paper is organized as follows. Section II presents background and related works. Section III describes the multi-stage architecture. Section IV presents the traffic assumptions and an example of heuristic scheduling algorithm to control optical packet forwarding. Section V illustrates a comparison between the reference shared-per-node and multi-stage architecture in terms of both performance and complexity (i. e. the number of optical components needed). Finally, section VI illustrates the conclusions of this work.

II. BACKGROUND AND RELATED WORKS

The sharing of TWCs in general all-optical switch architectures based on strictly non-blocking space switching matrix has been extensively studied in the past [10], [11], [12], with the aim to demonstrate that architectures with limited number of TWCs can provide the same performance as fully equipped architectures. In particular two types of TWCs sharing have been investigated, the shared-per-node (SPN), where a pool of TWCs is shared between all input fibers and channels, and the shared-per-link (SPL), where each output fiber has a dedicated pool of TWCs. Here the focus is pointed in SPN architectures. In this kind of switches, a scheduling algorithm is needed to manage the packet forwarding. In particular in synchronous context, algorithms with the aim to maximize the number of packets forwarded in a time slot are required. In [12], a scheduling algorithm to manage packet forwarding in synchronous environment in general shared-per-node architecture is presented. This scheduling algorithm aims at minimizing the number of wavelength conversion, maximizing the number of packets forwarded in a time slot. The scheduling algorithm is composed by three steps. In the first step, the input channels (wavelengths) are scanned, and the packets carried by the same wavelength and directed to the same output fiber are grouped. Packets in the same group contend for the same output channels, while packets on different groups are output contention free. For this reason, in the second step one packet from each set, randomly chosen, is sent without wavelength conversion and removed from the set. In this way the maximum number of packets are sent without conversion. In the third step, the output fibers are randomly selected and the remaining packets in the groups are sent exploiting wavelength conversion, until there are both free output channels and free TWCs, other packets are lost.

In any case, most works in literature don’t consider the practical implementation of the switch and the related number of optical components, as well as needed scheduling algorithms. In this paper, a practical implementation of the shared-per-node concept is illustrated. The proposed architecture, starting from the all-optical switches presented in [17], [18], implement the sharing of TWCs in a multi-stage broadcast-and-select S-λ-S architecture. Multi-stage architectures are of interest because they allow to save a large number of optical components with respect to general architectures based on strictly non-blocking space switching matrices.

When S-λ-S architecture is taken in account, a proper scheduling algorithm is needed given the multi-stage organization and the related particular organization of TWCs sharing.

Here, a scheduling algorithm suitable to manage packet forwarding in the multi-stage architecture is presented. This algorithm is similar to the scheduling algorithm presented in [12] and previously recalled, in the sense that it aims at maximizing the number of packets forwarded without conversion, but it is modified to taking in account the particular organization of the TWCs in the S-λ-S architecture. This algorithm has been presented for the first time in [16]. The scheduling algorithm is presented in section IV.

III. MULTI-STAGE BROADCAST-AND-SELECT ARCHITECTURE WITH SPARSE WAVELENGTH CONVERTERS

The practical implementation of the shared per node concept here proposed, called MS-SPN, is illustrated in figure 2. It is composed by three stages and organized as S-λ-S architecture. The first and third stages (S) are identical. These space switches exploit the broadcast-and-select principle and were reported for the first time in [17], [18]; since then they have been extensively considered in literature. The principle of operation is the following: at each node input, after optical amplification by means of an EDFA (Erbium-Doped Fiber Amplifier), a power coupler is used to generate multiple copies of the multi-wavelength bundle of channels entering the node from this input. The power coupler should have $N+1$ outlets where one outlet per incoming fiber is reserved for a local drop, while N copies for each input fiber are directed to a group of N wavelength selectors (only the N outlets for the transit traffic are shown in the figure). Each wavelength selector (WS) consists of two grating Mux/Demuxes (or any device with equivalent functionality) in tandem separated by an array of M optical devices (each one dedicated to one wavelength) that are able to operate as a 'shutter' (optical ON/OFF gate). At the output side, the WSs are interconnected by means of a $N:1$ power combiner in such a way that only one WS from the same input fiber might be coupled to the same output power combiner. Again, there might be $N+1$ branches in the coupler to serve for local add. Switching is achieved when the optical device is turned to the on-state, letting the wavelength to pass through, whilst when a particular wavelength should not appear in a particular output fiber, the 'gate' is switched to the off-state, blocking further propagation of the wavelength. The on-off ratio of the gate determines the level of in-band/out-of-band crosstalk. In [18] detailed information about this architecture is given. Since the architecture belongs to the broadcast-and-select family of switching fabrics it easily allows broadcasting and multicasting.

The intermediate stage (λ) represents the conversion stage and it is composed by B TWC blocks and $N-B$ optical fibers. Each TWC block is equipped with M wavelength converters. In the same block, each TWC is dedicated to a single wavelength, so that in this architecture tunable-input tunable-output or fixed-input tunable-output wavelength converters can be used indifferently. Being the TWCs in the same block dedicated to different wavelengths, packets carried by the same wavelength can exploit B TWCs placed in B different blocks. For this reason, incoming packets carried by the same wavelength (up to N) can exploit $N-B$ optical channels without TWC and B TWCs. The connectivity of the switch is maintained, so the packets that don't need conversion are sent on channels without conversion capability (in the optical fibers) and other packets (those that need conversion) are sent on blocks with TWCs. When TWC blocks are removed, the cost of the switch drops but blocking increases. The optimum number of TWCs is obtained by evaluating this cost-performance trade-off.

In each TWC block, the outputs of the wavelength converters are interfaced to the λ-module. This module can be physically constructed from a variety of devices/sub-systems e.g. a) an $M:1$ power coupler b) an $M \times K$ passive AWG router and $K:1$ coupler c) an $M \times M$ passive router and an array of M wavelength converters with fixed output wavelength followed by a grating multiplexer. The role of the λ-module is to group all the M wavelengths to a single fiber. Although the options (a-c) are identical in terms of logical performance, their physical layer performance is radically different. Further, the cost difference between these three options is enormous.

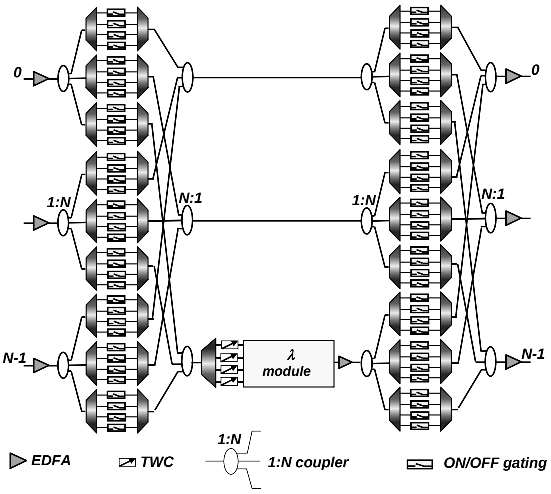

Fig. 2. Multi-stage switching node architecture with $N=3$ input/output fibers, $M=4$ wavelengths per fiber with sparse TWCs shared-per-node (MS-SPN).

By observing the multi-stage (figure 2) and the shared-per-node (figure 1) architectures, it is possible to observe that the differences between them are:

- the organization of the TWCs. In the SPN the TWCs are grouped in a single pool, if tunable-input TWCs are taken in account, each incoming packets can exploit each TWCs. In MS-SPN the TWCs are partitioned in B blocks of M TWCs, in a block each TWC is dedicated to one specific wavelength, so that even if tunable-input TWCs are considered, an incoming packet can exploit only B TWCs placed in different blocks (those TWC dedicated to the wavelength the packet is carried)
- packet grooming. In the SPN the packets at the output of the TWC pool are sent directly to the proper output fiber using a strictly non-blocking space switching matrix. In MS-SPN the wavelengths at the output of each TWC block are grouped in a single fiber by the λ-module. This can limit the possibility to find a match on the wavelength to forward a packet (i. e. two packets in the same TWC block cannot be converted in the same wavelength even if they are directed to different output fibers)

The first situation is the most evident and leads to the fact that the employment of tunable-input TWCs in MS-SPN is useless from logical point of view. The second effect is negligible especially when the load per wavelength is not very high.

A space equivalent of the proposed architecture helps to understand contention occurrences and to propose a scheduling algorithm that optimize resource utilization. It is presented in figure 3, for $N=3$ input and output fibers, $M=4$

wavelengths per fiber and $B = 1$ TWC block. It is composed by 3 stages, being the first and the third stages identical and consisting of M $N \times N$ cross-bars, each one representing contention on the same wavelength. The middle stage (B TWC blocks equipped with M TWCs dedicated each one to one specific wavelength and $N - B$ groups of M optical channels (each dedicated to one specific wavelength) is equivalent to B $M \times M$ cross-bars and $N - B$ groups of M simple channels. By observing in figure 3 the first cross-bar in the first stage, which concerns λ_0, it is possible to understand that a packet carried by λ_0 can exploit $N - B = 2$ optical channels and $B = 1$ TWC. In this situation, if three packets carried by λ_0 are directed to the same output fiber k, one packet is sent on an optical channel, one packet is sent on TWC and one packet is lost because no further TWC block is available for wavelength conversion. When the number of TWC blocks B increases, the number of optical fibers $N - B$ decreases, so that if B is high, it is possible that some packets that do not need conversion cannot exploit optical channels and must use TWCs even if they don't need conversion. The proposed switch can be used in virtual wavelength path (VWP) networks offering both shared wavelength path (SHWP) and scattered wavelength path (SCWP) functionality which according to [14] means wavelength-to-wavelength and fiber-to-fiber switching respectively. Therefore, with OCS traffic and the associated network-wide decisions for connection set-up, the proposed switch is non-blocking. This is not the case with OPS traffic in a multi-hop scheme and the blocking probability will be assessed here.

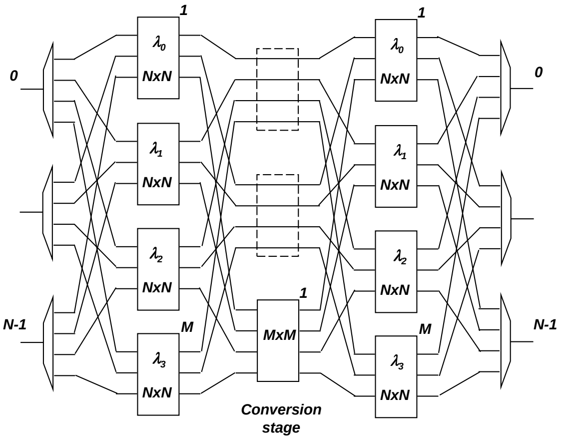

Fig. 3. Space equivalent of multi-stage architecture with N=3 input/output fibers, M=4 wavelengths per fiber and B=1 block of TWCs.

To resolve contentions, a scheduling algorithm is presented and applied to the architecture with a limited number of TWCs shared-per-node. The fully equipped switch can be dealt with as a special limiting case.

IV. TRAFFIC ASSUMPTIONS AND SCHEDULING ALGORITHM FOR FIBER-TO-FIBER SWITCHING

Two main different traffic assumptions are considered regarding the arrivals on switch input channels:

- Bernoulli arrivals, meaning that arrivals in different slots are independent and characterized by the average arrival rate p packets/slot
- Admissible traffic, meaning that arrivals are still characterized by mean p but no more than M packets arrive in a slot for the same output fiber

Bernoulli traffic can be considered as representative of the traffic in connection-less optical packet-switched networks as the result of statistical multiplexing of an high number of optical packets generated by the edge assembly units [19]. Admissible traffic could, on the other hand, be considered as the result of the admission operation performed on optical bursts that makes the traffic at each node to avoid switch output overbooking in each time slot, that is no more than M bursts are admitted on the same output fiber. In any case also admissible traffic needs wavelength conversion to resolve contention in the wavelength domain and could run into switch internal blocking due to switch resource unavailability. For both kinds of traffic fiber-to-fiber switching is considered meaning that a packet arriving on an input fiber k and wavelength j could in principle be forwarded to any output l and wavelength m. The proposed scheduling algorithm is able to manage both kinds of traffic.

The proposed switching architecture can be used in different network contexts ranging from wavelength switching to optical packet switching. Here the attention is focused on synchronous optical packet-switched networks with fixed-sized optical packets transferred through the network using a slotted statistical multiplexing scheme.

A heuristic scheduling algorithm called SB is here briefly recalled (a detailed description of the SB scheduling algorithm may be found in [16]). The scheduling algorithm operates on the particular organization of TWCs, that are grouped in B blocks, and on the multi-stage organization of the proposed architecture. The algorithm aims at maximizing the number of packets forwarded without conversion. It is an extension to the multi-stage architecture of the idea presented in [12]. Let us introduce the following definitions:

- L_k^j the sets that contain the packets carried by wavelength j and directed to output fiber k in a time slot ($k = 0, \cdots, N-1; j = 0, \cdots, M-1$).
- O_k the sets that contain the packets directed to output fiber k in a time slot ($k = 0, \cdots, N-1$).

The SB scheduling algorithm consists of two steps executed at each time slot. The first step is the creation of sets L_k^j and O_k. To this end, the input fibers are scanned and, if a packet carried by wavelength j and directed to output fiber k is found, it is added to L_k^j and O_k. In the second step the output fibers are scanned sequentially starting to the one indicated by a round robin counter RRO (modulo N). The wavelength on each fiber are considered starting from that one indicated by RRW (modulo M) as many times until all packets directed to that output fiber are served. At the first iteration related to an output fiber one packet, if present, from each set L_k^j is sent without conversion for each output fiber, given that packets on different groups are output contention free. When wavelength j on output fiber k is taken in account, the algorithm randomly chooses a packet from L_k^j (while $L_k^j \neq \oslash$) and calls the procedure sel_lambda_SB that returns a wavelength to encode the packet. The pseudo code of the proposed scheduling algorithm is here presented:

```
step 1:
for ($i = 0$; $i < N$; $i++$)
  for ($j = 0$; $j < M$; $j++$)
    if (there is a packet on fiber 'i' carried by $\lambda_j$ directed to output 'k')
      add the packet to $L_k^j$ and $O_k$;
step 2:
for ($k = 0$ ; $k < N$ ; $k++$)    {
    $out = (k + RRO)$ mod $N$;
    $j = 0$;
    while ($O_{out} \neq \oslash$)    {
      $\lambda = (j + RRW)$ mod $M$;
      if ($L_{out}^{\lambda} \neq \oslash$) {
        $\lambda_o$=$sel\_lambda\_SB(\lambda, out)$;
        select randomly a packet from $L_{out}^{\lambda}$;
        if ($\lambda_o \geq 0$)    the packet is sent on output fiber 'out' on $\lambda_o$;
        else                the packet is lost;
        remove the packet from $L_{out}^{\lambda}$ and $O_{out}$;
      }
      $j = (j + 1)$ mod $M$;
    }
}
$RRO = (RRO + 1)$ mod $N$;
if ($RRO == 0$)    $RRW = (RRW + 1)$ mod $M$;
```

The procedure sel_lambda_SB tries to forward the packet without conversion. If the wavelength the packet is carried is free on the destination output fiber (the packet does not need conversion), the packet is sent exploiting one of the fibers in the second stage, if there is at least one fiber where the that wavelength is free. In this case sel_lambda_SB returns the same wavelength the packet is carried. If no fiber is available, the packet is sent to the first TWC block where the corresponding TWC is free, and forwarded without conversion if the corresponding wavelength is free at the output of the TWC block (sel_lambda_SB returns the same wavelength the packet is carried). Otherwise sel_lambda_SB returns the first available wavelength on the TWC block. If the wavelength the packet is carried is not available on the destination output fiber, sel_lambda_SB looks for a TWC block where the TWC dedicated to that wavelength is free, and returns the first available wavelength to forward the packet. After having considered all available TWC blocks without finding any wavelength to match the preceding conditions, sel_lambda_SB returns a null value (-1) and the packet is lost. The pseudo-code and a detailed description of sel_lambda_SB may be found in [16].

By observing the architecture implementation (figure 2), a path between input and output channel (wavelength) must be established to forward a packet. Suppose that the scheduling algorithm decides to forward a packet from wavelength $j = 4$ of the input fiber $k = 2$ to the wavelength $m = 3$ on the output fiber $l = 1$. In addition suppose that the scheduling algorithm decides to forward the packet exploiting the TWC block $i = 1$ in the conversion stage. In the first stage the ON/OFF gate $j = 4$, in the WS connecting input fiber $k = 2$ and TWC block $i = 1$ (third WS starting from the top), has to be turned on. In the second stage, the TWC $j = 4$ on TWC block $i = 1$ is used to convert a packet on the proper output wavelength $m = 3$. Finally, in the WS connecting the TWC block $i = 1$ and the output fiber $l = 1$ (seventh WS), the ON/OFF gate $m = 3$ must be turned on.

V. PERFORMANCE AND COMPLEXITY COMPARISON

The aim of this section is to compare the MS-SPN implementation presented in this paper with the reference SPN scheme as regards performance and complexity in terms of expensive optical components.

A. Performance comparison between multi-stage and shared-per-node architectures

In this subsection, MS-SPN and SPN architectures are compared by simulation using SB algorithm for MS-SPN and a similar scheduling algorithm, already presented in [12] and briefly illustrated in II, for SPN. This algorithm allows to forward the maximum number of packets in a time slot, so that results evaluated with this scheduling algorithm provide a lower bound of packet loss probability obtainable with the SPN architecture. Simulation results are obtained with confidence interval at 95% less than or equal to the 5% of the mean.

First, the two architectures are compared when tunable-input tunable-output TWCs are employed. The shared-per-node architecture with this kind of TWCs is here called ID-SPN. Figure 4 shows a comparison between MS-SPN and ID-SPN architectures, equipped with the same number of TWCs ($R = BM$), in terms of packet loss probability under Bernoulli traffic. The packet loss is plotted as a function of the number of TWCs varying load in case ($N = 64, M = 8$). Both architectures lead to the same asymptotic value of packet loss, due to output blocking, while ID-SPN architecture provides better performance than the MS-SPN when the number of TWCs is limited. This is due to the intrinsic characteristics of the multi-stage organization that does not allow the perfect sharing of available wavelength converters, that are partitioned among the input wavelengths. Instead, in ID-SPN architecture equipped with tunable-input tunable-output TWCs, each incoming packet can exploit whatever TWC.

The same comparison with admissible traffic is shown in figure 5 in case $N = 64, M = 8$. Packet loss probability is plotted as e function of the number of TWCs varying load. When the load is very high $p = 1$, the packet loss of MS-SPN presents an asymptote. This is due to the packet grooming after each TWC block, that reduces the possibility to find a path between input and output channels. In fact, in a TWC block, two packets cannot be converted in the same wavelength even if they are directed to different output fibers, given that they are put in the same fiber at the output of the TWC block. This effect can be reduced by adding some additional fibers in the second stage [16]. In this manner, as it can be seen in figure

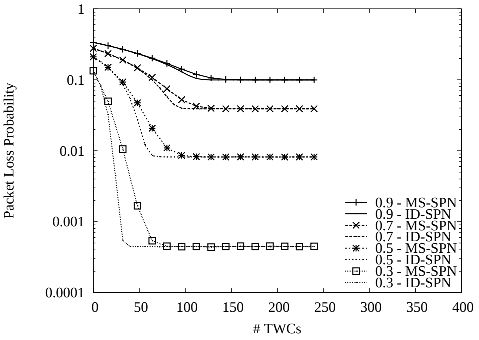

Fig. 4. Comparison between shared-per-node architecture with tunable-input TWCs (ID-SPN) and multi-stage architecture (MS-SPN) under Bernoulli traffic. Packet loss probability as a function of the number of TWCs varying load in case $N = 64, M = 8$.

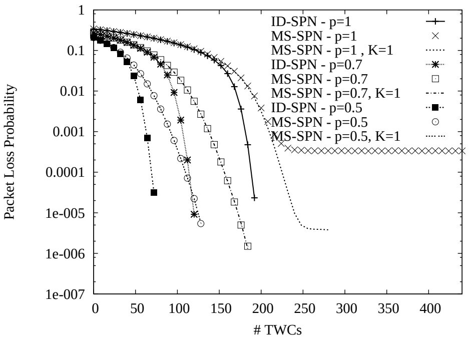

Fig. 5. Comparison between shared-per-node architecture with tunable-input TWCs (ID-SPN) and multi-stage architecture (MS-SPN) with admissible traffic. Packet loss probability as a function of the number of TWCs varying load in case $N = 64, M = 8$ and $K = 0, 1$ additional links in the MS-SPN.

5, where the case with $K = 1$ additional fiber is also plotted, the asymptotic value of packet loss can be lowered. Anyway, also with admissible traffic SPN architecture provides better performance than the MS-SPN, due to the complete sharing of TWCs.

Now, the two architectures are compared when fixed-input tunable-output TWCs are considered. The SPN with this kind of TWCs is called FI-SPN (fixed-input shared-per-node). FI-SPN is equipped with B TWCs dedicated to each wavelength (like in the MS-SPN) for a total amount of $R = BM$ TWCs. Figure 6 presents a comparison between MS-SPN and FI-SPN in terms of packet loss as a function of the total number of TWCs in case $N = 64, M = 8$ varying load per wavelength under Bernoulli traffic. MS-SPN and FI-SPN architectures provide the same performance. Only when load is high, there is a very little difference between the two architectures, due to the packet grooming that limit the capability of the scheduling algorithm to find a path to forward a packet. This constraint is due to the particular implementation, and it is not present in FI-SPN architecture. Anyway, the differences due to this constraint are negligible as shown in figure 6. In conclusion, under Bernoulli traffic, performance of MS-SPN architecture is the same with fixed-input or tunable-input TWCs, due to the particular organization, while the SPN performs better with tunable-input TWCs, because they allows to achieve more TWCs sharing.

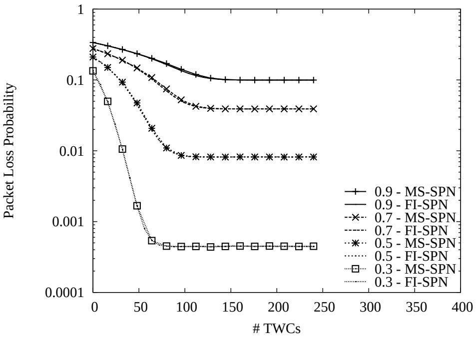

Fig. 6. Comparison between shared-per-node architecture with fixed-input TWCs (FI-SPN) and multi-stage architecture (MS-SPN) under Bernoulli traffic. Packet loss probability as a function of the number of TWCs varying load in case $N = 64, M = 8$.

In figure 7 the same comparison is illustrated when admissible traffic is taken in account. Packet loss probability is illustrated as a function of the total number of TWCs varying load, in case $N = 64, M = 8$. Further, packet loss probability when $K = 1$ additional optical fiber is added in the middle stage is plotted. In this situation, with FI-SPN architecture the packet loss is only due to the lack of TWCs, and this packet loss tends to zero when the number of TWCs is enough high. Instead in the MS-SPN, when the load is high, the packet loss tends to an asymptotic value when the number of TWCs increases. This is due to the constraint introduced by the packet grooming after each TWC block, as illustrated in IV. Also in this case the asymptote can be lowered by adding a number of optical fibers in the middle stage, as shown in figure 7. By adding some simple optical fibers, the packet loss of MS-SPN tends to the packet loss of FI-SPN, also when $p = 1$. The differences between MS-SPN and FI-SPN are not negligible only when the load is high ($p = 0.9, 1$).

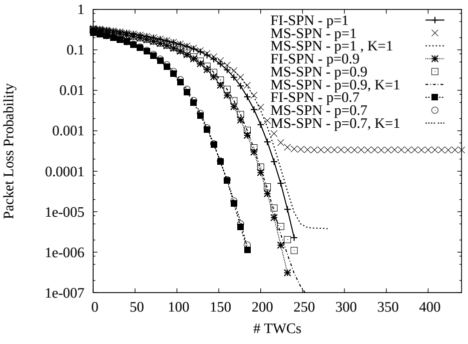

Fig. 7. Comparison between shared-per-node architecture with fixed-input TWCs (FI-SPN) and multi-stage architecture (MS-SPN) with admissible traffic. Packet loss probability as a function of the number of TWCs varying load in case $N = 64, M = 8$ and $K = 0, 1$ additional links in the MS-SPN.

B. Complexity comparison between multi-stage and shared-per-node architectures

In this section the two architectures are compared by evaluating the number of expensive optical components needed, i.e. optical gates (SOA) and TWCs. To this end the shared-per-node architecture is assumed to rely on the same SOA technology as the multi-stage architecture to implement the switching fabrics. The number of TWCs and optical gates needed for SPN and MS-SPN is evaluated.

The total number of TWCs in the SPN is:

$$\#TWC_{\text{SPN}} = R \tag{1}$$

The SPN architecture consists of two space switching matrices, the first with size $NM \times (NM + R)$, the second with size $R \times N(M-1)$, being $M-1$ the maximum number of packets per output interface which may require wavelength conversion. The total number of optical gates in the SPN architecture results in:

$$\begin{aligned} \#OG_{\text{SPN}} &= NM(NM+R) + RN(M-1) = \\ &= N^2M^2 + RN(2M-1) \end{aligned} \tag{2}$$

The number of optical gates needed in the SPN depends on the number of TWCs employed, R.

In the MS-SPN architecture each of the B TWC blocks is equipped with M TWCs, so the total number of TWCs is:

$$\#TWC_{\text{MS-SPN}} = MB \tag{3}$$

MS-SPN is equipped with M optical gates in each wavelength selector (WS). In the first stage, there are N WSs for each input interface, so the total number of WSs in the first stage is N^2, and the number of optical gates in the first stage is N^2M. The third stage is identical to the first one, so the total number of optical gates in the MS-SPN architecture is:

$$\#OG_{\text{MS-SPN}} = 2N^2M \tag{4}$$

The number of optical gates does not depend on the number of TWCs. In comparison with the number of OGs in the SPN, it is possible to see that in this case the dependence from M is linear instead of squared.

Table I shows a comparison between SPN and MS-SPN in terms of main optical components (optical gates, TWCs). Now, some complexity evaluations as a function of the main switch parameters (N, M, number of TWCs employed) are presented. First of all, in figure 8 a comparison of the number of optical gates needed for SPN and MS-SPN architecture as a function of the total number of TWCs employed is presented. The number of optical gates needed in the SPN is bigger than in the MS-SPN, independently on the number of TWCs. It means that, even if MS-SPN would require a larger number of TWCs, it allows to obtain a saving in terms of optical gates employed. In figure 9 the number of optical gates needed in SPN and MS-SPN architectures is plotted as a function of the number of wavelengths per fiber varying the total number of TWCs employed, in the case $N = 64$. For MS-SPN, only one curve is plotted, given that the number of optical gates employed in this architecture is independent on the number of TWCs, in the SPN instead the number of optical gates needed increases with the increase in the number of TWCs. In the SPN the dependence from M is squared while in the MS-SPN is linear. This means that the difference in terms of optical components enormously increases as the number of wavelengths per fiber increases (that is expecting for the future). In figure 10 the number of optical components in the SPN and MS-SPN architectures is presented as a function of the number of input/output fibers N, varying the percentage of TWCs with respect to the fully equipped architecture (NM

	MS-SPN	**SPN**
# TWC	BM	R
# ON/OFF gates	$2N^2M$	$N^2M^2 + 2NMR$

TABLE I

NUMBER OF OPTICAL COMPONENTS NEEDED IN MS-SPN AND SPN ARCHITECTURES

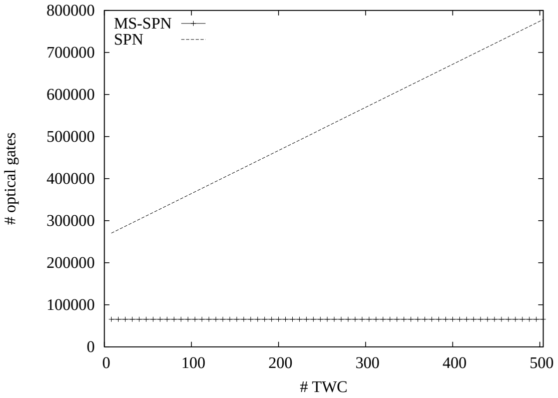

Fig. 8. Number of optical gates needed in the SPN and MS-SPN architectures as a function of the number of TWCs employed in case $N = 64$, $M = 8$.

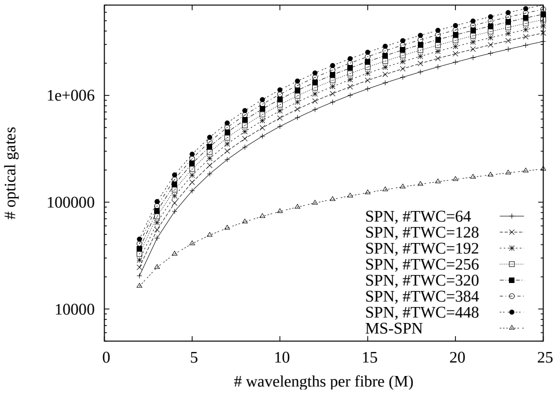

Fig. 9. Number of optical gates needed in the SPN and MS-SPN architectures as a function of the number of wavelengths M varying the total number of TWCs employed in case $N = 64$.

TWCs), in case $M = 8$. The number of optical gates needed for both SPN and MS-SPN is a function of N^2. Anyway, the SPN would require a bigger number of optical gates than the MS-SPN, with all value of N.

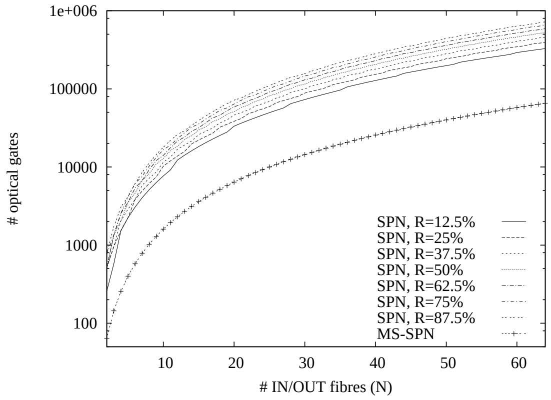

Fig. 10. Number of optical components needed in SPN and MS-SPN architectures as a function of the number of input/output fibers N varying the percentage of TWCs employed with respect to the maximum number NM, in case $M = 8$.

Finally, when the MS-SPN architecture with K additional fibers in the middle stage is taken in account (in order to reduce the packet loss with Admissible traffic and high load), the number of WSs connected to each input fiber becomes $(N + K)$, and the number of WSs connected to each output fiber also becomes $N + K$. The total number of WSs in the MS-SPN with K additional fibers is $2N(N+K)$, and the total number of optical gates is:

$$\#OG'_{MS-SPN} = 2NM(N + K) = 2N^2M + 2NMK \quad (5)$$

The K additional fibers require $2NK$ additional WSs, so the complexity in terms of optical gates is increased by $2NK$ with respect to the MS-SPN.

In figure 11 is presented the number of optical gates needed for the MS-SPN as a function of the switching node size (N) varying the number of additional fibers K, in case $M = 8$ and the number of TWC blocks equal to 50% of the maximum value (NM). The influence of K is relevant only when the switching size is low. In addition, the number of additional optical fibers needed is very narrow, so the complexity of the MS-SPN with some additional fibers is in the same magnitude of MS-SPN, except when the switching size is very limited.

In conclusion, the MS-SPN allows to save many optical gates with respect the SPN, especially when the switching size increases in both number of links (N) and wavelengths per fiber (M).

C. Switch complexity evaluation: numerical example

To have an idea of the switching cost, the proposed architectures have to be compared in both number of optical gates and TWCs in a common scenario. The number of TWCs needed depends on the input traffic characteristics. In fact the packet loss depends on the input traffic.

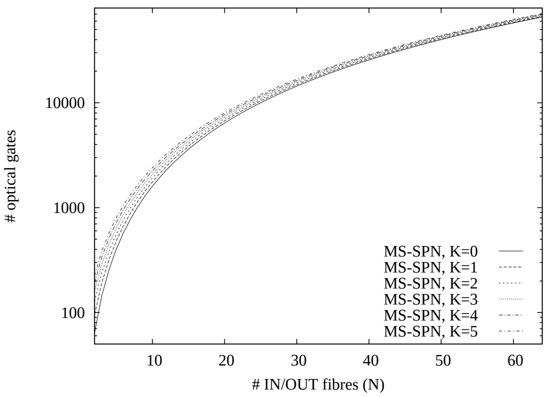

Fig. 11. Number of optical gates needed in the MS-SPN architecture as a function of the number of input/output fibers N varying the number of additional fibers on the second stage, K, in case $M = 8$, $B = 50\%$ of the maximum value, NM

When Bernoulli traffic is considered, the target is to obtain the asymptotic value of the packet loss (see figure 4, 6) with the minimum number of TWCs.

In figure 4 MS-SPN and SPN architectures are compared when tunable-input tunable-output TWCs are employed under Bernoulli traffic. By observing this figure it is possible to evaluate the number of TWCs needed to obtain the asymptotic value of the packet loss for both architectures. They result in $\#TWC_{\text{SPN}} = 72$ for SPN and $\#TWC_{\text{MS-SPN}} = 112$ for MS-SPN when the load per wavelength is $p = 0.5$. The number of the optical gates needed for SPN and MS-SPN can be obtained by applying (2), (4). They result in $\#OG_{\text{SPN}} = N^2M^2 + RN(2M - 1) = 331264$, $\#OG_{\text{MS-SPN}} = 2N^2M = 65536$ respectively. The number of optical components (optical gates and TWCs) for both architectures with Bernoulli traffic and $p = 0.5$ is presented in the table II

p=0.5	**MS-SPN**	**SPN**
# TWC	112	72
# ON/OFF gates	65536	331264

TABLE II

NUMBER OF OPTICAL GATES AND TWCS NEEDED IN MS-SPN AND SPN ARCHITECTURES WITH BERNOULLI TRAFFIC AND LOAD $p = 0.5$

When admissible traffic is considered, the packet loss tends to zero when the number of TWCs increases. In this case a suitable value of the packet loss that can not be overcome is chosen, then the minimum number of TWCs that allows to satisfy this condition can be found. For example, if the maximum tolerated value of packet loss is $1e^{-5}$, for $p = 0.7$, the number of TWCs needed for SPN and MS-SPN is $\#TWC_{\text{SPN}} = 120$ and $\#TWC_{\text{MS-SPN}} = 176$ (22 TWC blocks with $M = 8$ TWCs each) respectively (deduced from the figure 5). Then, the number of optical gates is evaluated by applying (2), (4), as in Bernoulli case. The total number of optical components needed with admissible traffic, load $p = 0.7$, maximum packet loss $1e^{-5}$ can be found in table

III

p=0.5	**MS-SPN**	**SPN**
# TWC	176	120
# ON/OFF gates	65536	377344

TABLE III

NUMBER OF OPTICAL COMPONENTS NEEDED IN MS-SPN AND SPN ARCHITECTURES WITH ADMISSIBLE TRAFFIC, LOAD PER WAVELENGTH $p = 0.5$ AND PACKET LOSS LOWER THAN $1e^{-5}$.

By observing the tables, it is possible to see that the MS-SPN requires a bigger number of tunable-input tunable-output TWCs than the SPN, but it allows to save a very relevant number of optical gates (SOA). So, the proposed multi-stage architecture demonstrates the feasibility of SPN concept with limited complexity.

VI. CONCLUSIONS

In this paper a multi-stage broadcast-and-select architecture that implements the shared-per-node wavelength sharing has been described. It has been thought to be implemented with available photonic devices and to ensure future proof concepts by taking advantage of new components with enhanced functionalities, with minimum disruption. A scheduling algorithm has been proposed to exploit switch resources in a synchronous packet context. Comparisons with the ideal shared-per-node scheme in terms of performance and complexity have been also presented showing the main differences related to the multi-stage switch organization. It has been demonstrated that the multi-stage architecture allows to obtain a relevant saving in terms of optical gates, which are expensive components. In addition, the multi-stage architecture provides the same performance as the SPN equipped with fixed-input tunable-output TWCs with Bernoulli traffic. With admissible traffic, when the load is very high the multi-stage architecture has to be modified by adding some additional fibers in the second stage to obtain the same performance as the SPN equipped with fixed-input tunable-output TWCs. Further work concerns the study of the proposed architecture as an hybrid opto-electronic switch equipped with electronic buffers for contention resolution also in the time domain.

ACKNOWLEDGMENT

The authors would like to thank the Commission of European Union and the Italian Ministry of Research and Education (MIUR) for funding this work through the IST-FP6 Network of Excellence *e-Photon/ONe* and the PRIN Project OSATE. They also wish to thank Prof. Alexandros Stavdas, University of Peloponnese, Greece, for the useful discussions about switch architectures and implementation.

REFERENCES

[1] M. J. O'Mahony, D. Simeonidou, D. K. Hunter, A. Tzanakaki, ”The application of Optical Packet Switching in Future Communications Networks”, *IEEE Communication Magazine*, vol. 39, no. 3, pp.128-135, Mar. 2001.

[2] M. Listanti, V. Eramo, R. Sabella, ”Architectural and Technological Issues for Future Optical Internet Networks”, *IEEE Communication Magazine*, Vol. 38, No. 9, pp. 82-92, Sept. 2000.

[3] B. Li, Y. Qin, X. Cao, K. Sivalingam, ”Photonic Packet Switching: Architectures and Performance”, *Opt. Network Magazine*, pp. 27-39, Jan/Feb. 2001.

[4] S Yao, B. Mukherjee, S. Dixit, ”Advances in Photonic Packet Switching: An Overview”, *IEEE Comm. Magazine*, Vol. 38, pp. 84-94, Feb. 2000.

[5] A. Jourdan, D. Chiaroni, E. Dotaro, G. Eilenberger, F. Masetti, M. Renaud, ”The Perspective of Optical Packet Switching in IP-Dominant Backbone and Metropolitan Networks”, *IEEE Comm. Magazine*, Vol.39 pp. 136-141, Mar. 2001.

[6] Xueli Hou and H.T. Mouftah, ”Towards optimal design of wavelength-convertible optical switches for the all-optical next-generation internet”, *IEEE CCECE 2001*, Page 0441

[7] K. Kitayama, M. Murata ”Versatile Optical Code-Based MPLS for Circuit, Burst and Packet Switchings”, Journal of Lightwave Technoloogy, Vol. 21, No 11, November 2003.

[8] S. L. Danielsen, P. B. Hansen, K.E. Stubkjaer, ”Wavelength Conversion in Optical Packet Switching”, *Journal Lightwave Technology*, vol. 16, no. 9, pp. 2095-2108, September 1998.

[9] G. Maier, A. Pattavina, ”Generalized Space-Equivalent Analysis of Optical Cross-Connect Architectures”, *IEEE INFOCOM 2001*, Vol. 1, April 2001, pp.159-168.

[10] F. A. Al-Zaharani, A.A. Habiballa, A. G. Fayoumi, A. P. Jayasumana ”Performance Tradeoffs of Shared Limited Range Wavelength Conversion Schemes in Optical WDM Networks”, *Second IFIP International Conference on Wireless and Optical Communications Networks*, 2005. WOCN 2005, March 2005, pp. 18 - 22.

[11] V. Eramo, M. Listanti, M. Spaziani, ”Resource Sharing in Optical Packet Switches With Limited-Range Wavelength Converters”, *Journal of Lightwave Technology*, Vol.23, No. 2, Feb. 2005.

[12] 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, Feb. 2003.

[13] C. Raffaelli, M. Savi, A. Stavdas, ”Sharing Wavelength Converters in Multistage Optical Packet Switches”, *proceeding of hpsr 2006*, Poznan, Poland, June 2006, page(s): 203-208.

[14] D. Hunter et al., ”WASPNET: A Wavelength Switched Packet Network”, *IEEE Communication Magazine*, March 1999, pp. 120-129.

[15] L. Dittmann et al. ”The European IST project DAVID: a viable approach toward optical packet switching”, *IEEE Journal on Selected Areas in Communications*, Volume 21, Issue 7, Sept. 2003,pp. 1026 - 1040.

[16] C. Raffaelli, M. Savi, A. Stavdas ”Performance of Scheduling Algorithms in Multi-Stage Optical Packet Switches with Sparse Wavelength Converters”, *Proceedings of Globecom 2006*, San Francisco, U.S.A., November 2006.

[17] A. Stavdas: Design of multiplexer/demultiplexer for dense WDM wavelength routed optical networks, *Ph.D. Thesis, University of London*, 1995.

[18] A. Stavdas, H. Avramopoulos, E. Protonotarios, J. E. Midwinter ”A novel architecture for a wavelength crossconnect node”, *Photonic Network Communications*, Vol. 1 (1), pp.77-88, 1999.

[19] M. Izal, J. Aracil, ” On the Influence of Self Similarity on Optical Burst Switching Traffic”, Proc. IEEE Globecom 2002.