Optimal wavelength-space crossbar switches for supercomputer optical interconnects

Ioannis Roudas,[1,2,*] B. Roe Hemenway,[2] Richard R. Grzybowski,[2] and Fotini Karinou[1]

[1]*Department of Electical & Computer Engineering, University of Patras, Rio 26504, Greece*
[2]*Corning Inc., Corning, New York 14831, USA*
[*]*roudas@ece.upatras.gr*

Abstract: We propose a most economical design of the *Optical Shared MemOry Supercomputer Interconnect System (OSMOSIS)* all-optical, wavelength-space crossbar switch fabric. It is shown, by analysis and simulation, that the total number of on-off gates required for the proposed $N \times N$ switch fabric can scale asymptotically as $N \ln N$ if the number of input/output ports N can be factored into a product of small primes. This is of the same order of magnitude as Shannon's lower bound for switch complexity, according to which the minimum number of two-state switches required for the construction of a $N \times N$ permutation switch is $\log_2 (N!)$.

OCIS codes: (200.4650) Optical interconnects; (200.6715) Switching; (250.5980) Semiconductor optical amplifiers.

References and links

1. The Top 500 Supercomputer Sites, http://www.top500.org.
2. R. Hemenway, R. R. Grzybowski, C. Minkenberg, and R. Luijten, "Optical-packet-switched interconnect for supercomputer applications," J. Opt. Netw. **3**(12), 900–913 (2004).
3. R. Luijten, W. E. Denzel, R. R. Grzybowski, and R. Hemenway, "Optical interconnection networks: The OSMOSIS project," in *Proceedings of IEEE Lasers and Electro-Optics Society 17th Annual Meeting*, Rio Grande, Puerto Rico, (2004), 563–564.
4. R. Luijten, C. Minkenberg, R. Hemenway, M. Sauer, and R. Grzybowski, "Viable opto-electronic HPC interconnect fabrics," in *Proceedings of ACM/IEEE SC2005 Conference on High Performance Networking and Computing*, Seattle, WA, USA, (2005).
5. R. Luijten, C. Minkenberg, B. R. Hemenway, and R. R. Grzybowski, "Implementation challenges in the OSMOSIS optical HPC switch," in *Proceedings of IEEE Lasers and Electro-Optics Society 19th Annual Meeting*, Montreal, CA, (2006), 623–624.
6. M. Sauer, R. Hemenway, R. Grzybowski, D. Peters, J. Dickens, and R. Karfelt, "A scaleable optical interconnect for low-latency cell switching in high-performance computing systems," Proc. SPIE **6124**, 61240N, 61240N-12 (2006).
7. I. Roudas, B. R. Hemenway, and R. R. Grzybowski, "Optimization of a supercomputer optical interconnect architecture," in *Proceedings of IEEE Lasers and Electro-Optics Society 20th Annual Meeting*, Orlando, FL, (2007), 741–742.
8. R. P. Luijten and R. Grzybowski, "The OSMOSIS Optical Packet Switch for Supercomputers," in *Proceedings of IEEE/OSA Optical Fiber Communication Conference*, San Diego, CA, (2009), pp. 1–3.
9. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing*, 2nd ed. (Cambridge University Press, 1992).
10. M. Zirngibl, C. H. Joyner, and B. Glance, "Digitally tunable channel dropping filter/equalizer based on waveguide grating router and optical amplifier integration," IEEE Photon. Technol. Lett. **6**(4), 513–515 (1994).
11. L. Kleinrock and F. Kamoun, "Hierarchical routing for large networks–performance evaluation and optimization," Comput. Netw. **1**, 82–92 (1977).
12. O. Ishida, H. Takahashi, and Y. Inoue, "Digitally tunable optical filters using arrayed-waveguide grating (AWG) multiplexers and optical switches," J. Lightwave Technol. **15**(2), 321–327 (1997).
13. A. Misawa, K. Sasayama, and Y. Yamada, "WDM knockout switch with multi-output-port wavelength channel selectors," J. Lightwave Technol. **16**(12), 2212–2219 (1998).
14. N. Kikuchi, Y. Shibata, H. Okamoto, Y. Kawaguchi, S. Oku, H. Ishii, Y. Yoshikuni, and Y. Tohmori, "Monolithically integrated 64-channel WDM channel selector with novel configuration," Electron. Lett. **38**(7), 331–332 (2002).
15. N. Kikuchi, Y. Shibata, H. Okamoto, Y. Kawaguchi, S. Oku, H. Ishii, Y. Yoshikuni, and Y. Tohmori, "Monolithically integrated 64-channel WDM wavelength selective receiver," Electron. Lett. **39**(3), 312–314 (2003).

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

16. N. Kikuchi, Y. Shibata, H. Okamoto, Y. Kawaguchi, S. Oku, Y. Kondo, and Y. Tohmori, "Monolithically integrated 100-channel WDM channel selector employing low-crosstalk AWG," IEEE Photon. Technol. Lett. **16**(11), 2481–2483 (2004).
17. G. Arfken, *Mathematical Methods for Physicists*, 3d ed. (Academic Press, 1985).
18. G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization* (Wiley, 1999).
19. M. A. Saad, *Thermodynamics: Principles and Practice* (Prentice Hall, 1997).
20. P. E. Green, *Fiber-Optic Networks* (Prentice Hall, 1993).
21. F. Karinou, I. Roudas, B. R. Hemenway, and R. R. Grzybowski, "Physical layer performance of HPC optical interconnect architectures," J. Lightwave Technol. **29**(21), 3167–3177 (2011).

1. Introduction

High-performance computing (HPC) systems typically use thousands of microprocessors grouped into clusters, connected among themselves and with several TBytes of distributed memory using an interconnection network. Contemporary HPC system interconnects are mostly based on electronic switch fabrics [1]. Given that HPC systems' performance is anticipated to approach 1 Exaflop by 2020 [1], it is necessary to employ some form of photonic switching, in order to achieve a low-cost, low-latency, high throughput interconnection between the microprocessors and the shared memory.

The Optical Shared MemOry Supercomputer Interconnect System (OSMOSIS) project [2–8] was a joint research effort between Corning, Inc. and IBM. The purpose of the project was to design and demonstrate a scalable, commercially available, hybrid HPC interconnect architecture, combining electronics for buffering and storing with a $N \times N$ wavelength-space crossbar switch fabric for optical cell switching. The latter was implemented using a broadcast-and-select architecture, with fixed-wavelength transmitters and discretely-tunable, direct-detection receivers. Receiver tunability was achieved using optical multiplexer-demultiplexer (MUX/DMUX) pairs and $2N^{3/2}$ semiconductor optical amplifiers (SOAs) as on-off gates.

Since the future viability of HPC systems' optical interconnects critically depends on the drastic reduction of the cost per input/output port, a drastic reduction of SOA-based on-off gates would be clearly beneficial.

In this paper, we propose a most economical design of a fast $N \times N$ all-optical, wavelength-space crossbar switch fabric for optical interconnects. The total number of on-off gates required for the proposed $N \times N$ switch fabric can scale asymptotically as $N \ln N$ if the number of input/output ports N can be factored into a product of small primes. This is of the same order of magnitude as Shannon's lower bound for switch complexity, according to which the minimum number of two-state switches required for the construction of a $N \times N$ permutation switch is $\log_2 (N!)$. This result is reminiscent of the reduction in computational complexity achieved by highly efficient algorithms, e.g., Quicksort [9].

The reduction in the number of on-off gates is achieved by hierarchical multiplexing and by optimizing the design of tunable, direct-detection receivers. Hierarchical multiplexing first groups wavelengths into wavebands, then, groups wavebands into second-order wavebands, and so forth. Hierarchical demultiplexing is performed at the tunable, direct-detection receivers using several successive selection stages, with progressively finer selectivity, in tandem. More specifically, each tunable direct-detection receiver contains a fast, lossless, discretely-tunable optical bandpass filter (also referred to as *wavelength selector*). The latter can be constructed using several stages of optical multiplexer/demultiplexer (MUX/DMUX) pairs and SOAs as on-off gates [10]. We show that an optimal multi-stage wavelength selector configuration exists, which exploits the periodicity of arrayed waveguide grating (AWG) MUX/DMUX transfer functions, in order to reduce the number of SOAs to a minimum.

Hierarchical multiplexing/demultiplexing, in conjunction with AWG MUX/DMUXs, is key to the minimization of SOA on-off gates, in the same way that hierarchical routing is used to minimize the length of routing Tables [11].

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

A drawback of the aforementioned configuration is that the optical signal passes through a longer concatenation of AWG MUX/DMUXs and SOAs than in the original OSMOSIS optical interconnect architecture [2]. Consequently, the proposed interconnect is more vulnerable to transmission effects, esp. in SOAs, i.e., self-gain and cross-gain modulation, and polarization dependent gain. However, by proper selection of the system design parameters, the impact of the aforementioned transmission impairments on the proposed optical interconnect can be limited.

The rest of the paper is organized as follows: In Section 2, the operating principle of the proposed optimal crossbar switch fabric architecture for optimal interconnects is illustrated. In Section 3, we present an approximate analytical model for the calculation of the optimal number of stages of the multiplexing hierarchy and the optimal partition of tributaries per stage, for a given number of channels, in order to minimize the number of SOAs. In the same Section, the validity of the analytical model is checked by exhaustive search of all possible switch realizations and different optical crossbar switch sizes. It is shown that the number of SOAs per receiver card can very closely approach the analytically calculated minimum, when the number of channels can be factored into a product of primes. Section 4 presents a simplified control algorithm for setting the states of the on-off gates of the optimal wavelength selector. Finally, in Section 5, we present an improved analytical model for the calculation of the optimal number of stages of the multiplexing hierarchy and the optimal partition of tributaries per selection stage, by taking into account both the cost of SOAs, as well as of AWG MUX/DMUXs.

2. Operating principle of the optical interconnect

The block diagram of the proposed wavelength-space crossbar switch is shown in Fig. 1(a). Each transmitter uses a different carrier frequency (from a set of N equally spaced frequencies). All channels are broadcasted to all receivers using a $N \times N$ star coupler. Each receiver can select a different channel from the received WDM optical signal.

A most rudimentary implementation of the $N \times N$ star coupler and the tunable receivers in the proposed architecture is shown in detail in Fig. 1(b). Each receiver card contains a discretely tunable channel selector [10], which is used to discriminate one out of the N received carrier frequencies, followed by a direct-detection receiver with no inherent frequency selectivity. The channel selector is composed of a MUX/DMUX pair and N SOAs as on-off gates. By biasing one out of the NSOAs above the transparency current level, it is possible to select one channel, while attenuating all others. Since there are, in total, N SOAs per channel selector and one channel selector per receiver card, the total number of SOAs required in this particular implementation of the optical interconnect is N^2.

As an example, assume $N = 16$. The schematic of the 16-channel selector is shown in Fig. 2(a). Figure 2(b) illustrates the operation of the 16-channel selector in the frequency domain.

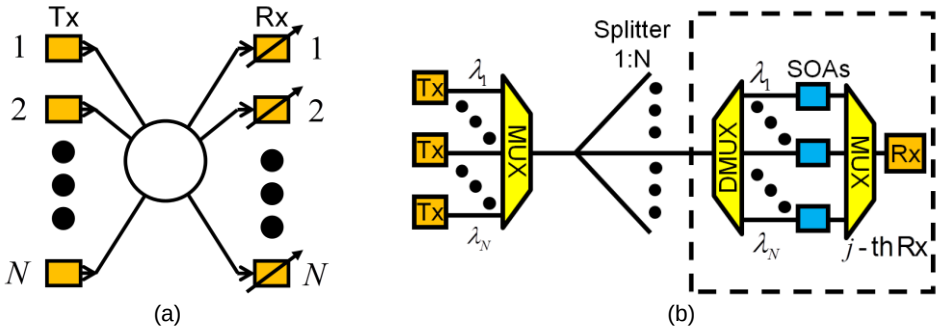

Fig. 1. (a) Broadcast-and-select architecture using a star coupler, N fixed transmitters and N tunable receivers. (b) Actual implementation. (Symbols: Tx = Transmitter, Rx = Receiver, SOA = Semiconductor optical amplifier, MUX = Multiplexer, DMUX = Demultiplexer).

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

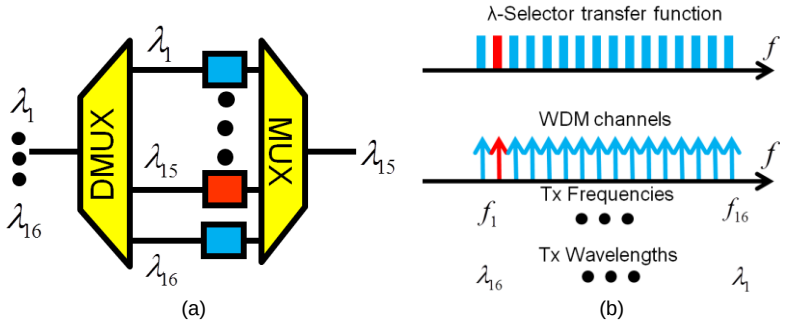

Fig. 2. (a) Schematic diagram illustrating the operation of a single-stage 16-channel selector using an arrayed waveguide grating (AWG) demultiplexer/multiplexer (MUX/DMUX) pair and 16 semiconductor optical amplifiers. (b) Visualization of the functionality in the frequency domain.

The 1 × 16 DMUX spatially separates 16 channels at carrier frequencies f_1 through f_{16} (corresponding to wavelengths λ_{16} through λ_1, respectively), so that only one of the input channels appears at each of its output arms. The idealized power transfer function of all arms is shown as a series of rectangles. The shaded (red) rectangle shows the power transfer function of the selected channel. Broken (blue) rectangles indicate the power transfer functions of all other channels. The channel selected is denoted with a solid (red) arrow, whereas all other channels are denoted with broken (blue) arrows.

Several variants [12–16] of the original single-stage channel selector proposed in [10] appeared in the literature, which take advantage of the periodic nature of the transfer function of arrayed waveguide grating (AWG) MUX/DMUXs, in order to achieve the same functionality as in Fig. 2, while using a smaller number of SOAs.

For instance, a two-stage channel selector [14–16] was proposed, in order to reduce the number of required SOAs. The configuration of the aforementioned selector is illustrated in Fig. 3(a), for the case of 16 channels. The operating principle can be better understood in the frequency domain (Fig. 3(b)): First, notice that the carrier frequencies of the 16 transmitted channels are slightly different, compared to the channel plan of Fig. 2(b), i.e., they are grouped in four wavebands of four channels each. Wavelengths within each waveband and successive wavebands have equal spacing. However, there is a guardband between adjacent wavebands. The MUX/DMUXs of the first selection stage of the channel selector do not have to be AWGs. The bandwidth of their transfer function is $B_W \cong 4\Delta f$, where Δf is the channel spacing within a waveband. The second selection stage uses an AWG MUX/DMUX pair with

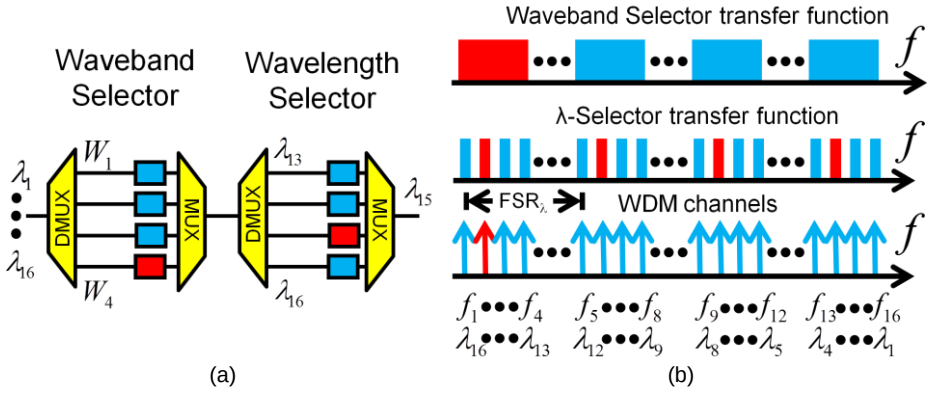

Fig. 3. (a) Schematic diagram illustrating the operation of a two-stage 16-channel selector, which is functionally equivalent to the single-stage 16-channel selector shown in Fig. 2(a). The two-stage selector uses two arrayed waveguide grating (AWG) demultiplexer/multiplexer

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

(MUX/DMUX) pairs and 8 semiconductor optical amplifiers. (b) Visualization of the functionality in the frequency domain.

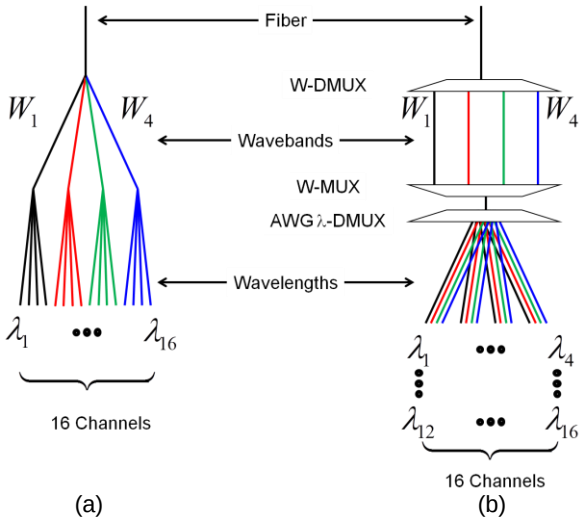

Fig. 4. Multiplexing/demultiplexing functionalities of the 16 × 16 optical interconnect of Fig. 3. (a) At the transmitter, the 16-channel multiplexing hierarchy can be represented by a tree structure. The root tributary is composed of four wavebands of four wavelengths each. (Symbols: The four wavebands are denoted from left to right with black, red, green, and blue lines. Same color code is used for the wavelengths within each waveband). (b) At the receiver, it is possible to reduce the number of SOAs by folding the tree structure of the multiplexing hierarchy, exploiting the periodicity of the AWG transfer function. The folded tree, as it appears after a Waveband MUX/DMUX pair and an AWG Wavelength DMUX, is shown here. By placing SOAs at the arms of the DMUXs at the waveband- and wavelength-level (i.e., the branches and the leaves of the folded tree), it is possible to reduce the number of SOAs by half. In order to achieve this, the wavelength DMUX must be an AWG.

free spectral range FSR_λ equal to the waveband spacing. The bandwidths of the individual arms are wide enough to pass a single wavelength channel without significant distortion. At the first stage of the channel selector, one of the four wavebands is selected by appropriate biasing of one out of the four SOAs (red rectangle). At the second stage of the channel selector, one of the four wavelengths of the selected waveband is chosen (red arrow) by appropriate biasing of one out of the four SOAs (red rectangle).

In conclusion, the two-stage, 16-channel selector is functionally equivalent to the single-stage, 16-channel selector shown in Fig. 2(a). However, the two-stage selector of Fig. 3(a) uses only half of the SOAs required in the single-stage 16-channel selector shown in Fig. 2(a). Figure 4 and Fig. 5 illustrate why clustering/partitioning of wavelengths into wavebands and the aforementioned structure of the 16-channel selector in Fig. 3(a) are beneficial.

In Fig. 4(a), the hierarchical multiplexing shown in Fig. 3(b) is represented by a tree diagram. The highest-ranking tributary is the root of the tree, whereas the smallest-ranking tributaries (i.e., the individual wavelengths) are the leaves of the tree. Then, wavelength selection performed by the discretely-tunable, direct-detection receivers of the optical interconnect is analogous to the traversal of a tree from root to leaves.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

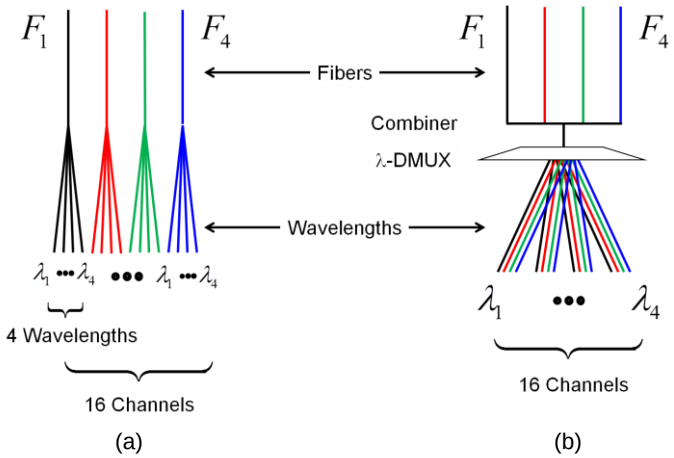

Fig. 5. Alternative implementation of the multiplexing/demultiplexing functionalities of Fig. 4. (a) At the transmitter, hierarchical multiplexing is performed using four fibers carrying the same set of four wavelengths, instead of one fiber carrying four wavebands of four distinct wavelengths each (i.e., a total of 16 wavelengths) as in Fig. 4; (b) Receiver: The Waveband MUX/DMUX pair at the receiver of Fig. 4 is omitted here. The Waveband MUX at the receiver of Fig. 4 is substituted by a combiner. The folded tree, as it appears after a combiner and a Wavelength DMUX, is shown here. By placing SOAs at the arms of the DMUXs at the fiber- and wavelength-level (i.e., the branches and the leaves of the folded tree), it is possible to reduce the number of SOAs by half. The wavelength DMUX does not have to be AWG.

Figure 4(b) shows why optical demultiplexers with periodic transfer functions are key components, in order to exploit the hierarchical organization of wavelengths. Due to the periodicity of the optical demultiplexer transfer functions, the original tree representing the multiplexing hierarchy collapses into an equivalent folded tree, where the number of nodes and edges is significantly reduced. While retrieving a single wavelength, the receiver has to make as many binary decisions as the number of edges emanating from all the nodes of the folded tree. A minimum number of SOAs is required when the original tree (before folding) is maximally symmetric.

Figure 5 shows an alternative implementation of the multiplexing/demultiplexing hierarchy of Fig. 4. At the transmitter, hierarchical multiplexing uses four fibers, carrying the same set of four wavelengths, instead of one fiber carrying four wavebands of four distinct wavelengths each. This reduces the number of used wavelengths and enables wavelength reuse. The waveband MUX/DMUX pair at the receiver of Fig. 4 is omitted here. The waveband MUX at the receiver of Fig. 4 is substituted by a combiner.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

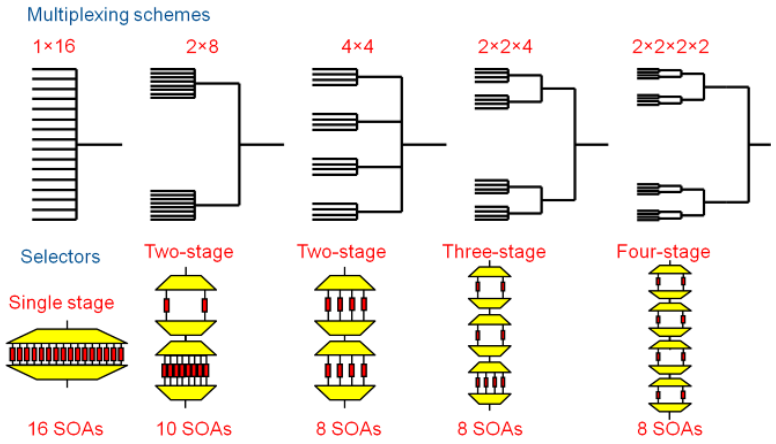

Fig. 6. All possible 16-wavelength multiplexing hierarchies and their corresponding discretely-tunable, multi-stage, channel selector layouts.

For a given number of wavelengths, there are several alternative hierarchical multiplexing schemes with different number of levels and multiplexing granularities (Fig. 6). These alternative hierarchies can be represented by trees with a different number of levels, nodes, and edges, given the same number of leaves. From Fig. 6 we observe that the optimal multiplexing granularity is achieved when an equipartition or a quasi-equipartition of tributaries into the wavebands of different orders exists. Then, wavelength selection (i.e., tree traversal) can be extremely efficient.

It is worth pointing out several important details in the above example:

- MUXs used at all selection stages of Fig. 2(a), Fig. 3(a) could be substituted by combiners at the expense of insertion loss, with a deterioration in performance due to increased insertion loss and adjacent channel crosstalk.
- DMUXs spatially separate the constituents of the WDM signal they receive at their input. This implies that the role of a DMUX is to transform a WDM signal into a spatial division multiplexed (SDM) signal. Alternatively, for the last multiplexing stage at the transmitter, one could use SDM of several fibers. In this case, the DMUX corresponding to the first channel selector must be omitted.
- When SDM of several fibers is used in place of the last level of WDM multiplexing at the transmitter, all fibers can carry the same set of wavelengths (wavelength reuse). This is desirable, since it reduces the number of different transmitters required for the implementation of the optical interconnect architecture. However, it is not necessary for the fibers to carry the same set of wavelengths. They can carry adjacent wavelength sets, as long as the spacing of subsequent wavelength sets is equal to the FSR of the MUX/DMUXs of the subsequent selection stage of the discretely tunable receiver.
- The channel selection stages act as composite filters (if we neglect the nonlinear behavior of the SOAs). Therefore, their order can be interchanged without significant change in the performance of the multi-stage channel selector.
- The DMUX used at the first selection stage of a multi-stage channel selector can always be aperiodic. When SDM of several fibers is used at the last multiplexing level of the transmitter instead of WDM, and if the same set of wavelengths is repeated in all fibers, the DMUX that follows the fiber selection stage can be aperiodic.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

3. Formulation of the problem

A natural extension of the previous example is to further increase the stages of the channel selector until the number of required SOAs is reduced to a minimum. The question arises as to what the optimum number of stages is and the optimum number of tributaries (i.e., subsets of channels) selected by each consecutive stage of the channel selector is, in order to minimize the number of the required SOAs.

The aforementioned question can be formulated as a constrained minimization problem, where the cost function to minimize is the number of SOAs, the constraint is the number of channels, and the (discrete) variables to optimize are the number of selection stages and the number of tributaries per stage. The originality of the current study lies in the mathematical formulation of this optimization problem and its approximate analytical solution, using the method of Lagrange multipliers [17]. The analytical results are verified by exhaustive search of all possible realizations of 64 × 64, 72 × 72, 96 × 96 and 256 × 256 switch fabrics and by direct enumeration of the number of SOAs required in each realization.

3.1 Analytical model: Optimization of the number of tributaries

Consider that the transmitted channels are grouped into a multiplexing hierarchy of K levels (Fig. 7(a)). At each level i of the multiplexing hierarchy, n_i tributaries (i.e., wavelengths, wavebands of different orders, optical fibers) are grouped together (Fig. 7(a)).

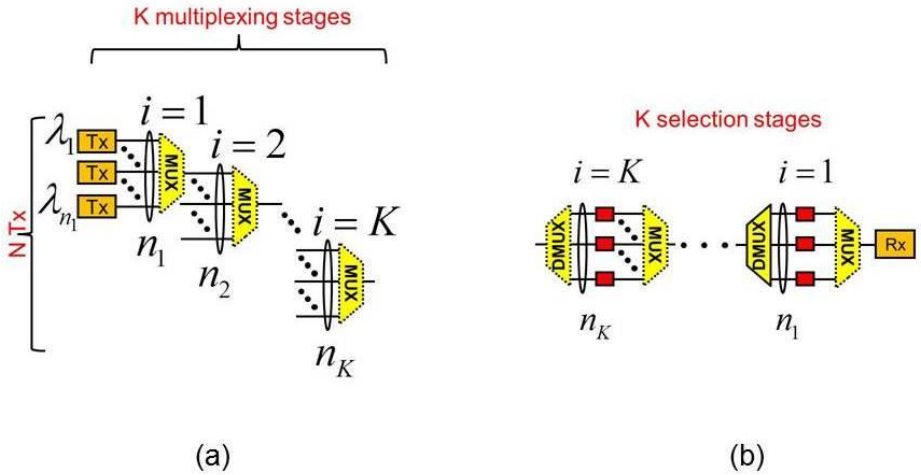

Fig. 7. (a) Example of K multiplexing stages with n_1 channels per waveband of order one (i = 1), n_2 wavebands of order one per waveband of order two (i = 2), and so on. The K-th multiplexing stage can be implemented using spatial division multiplexing (SDM) of n_k optical fibers, in order to allow wavelength reuse and hence, reduce the total number of wavelengths required. In this case, the K-th multiplexer should be omitted. The rest of the multiplexers (denoted by broken trapezoids) are optional, if a N × N star coupler is used instead of a splitter, as in Fig. 1(b). However, in practice, multiplexers might be required, in order to reduce adjacent channel crosstalk. (b) Layout of a receiver card with K channel selection stages. The numbering of the selection stages is done in reverse order from left to right. The MUX/DMUX pair corresponding to the K-th order channel selector (denoted by broken trapezoids) might be omitted if the K-th multiplexing stage at the transmitter is implemented using spatial division multiplexing (SDM) of n_k optical fibers.

Then, the total number of transmitters N is factored as follows:

$$N = \prod_{i=1}^{K} n_i \tag{1}$$

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

The number of tunable receiver cards is equal to the number of transmitters N. At each receiver card, there are K selection stages with n_i, $i = 1,\ldots,K$, SOAs each (Fig. 7(b)).

The number of SOAs per receiver card Ω is

$$\Omega = \sum_{i=1}^{K} n_i \tag{2}$$

The total number of SOAs that must be used in the optical interconnect is then

$$\Omega_{tot} = N\,\Omega \tag{3}$$

We want to minimize Eq. (2) subject to the constraint Eq. (1), given a fixed number of channel selection stages K.

There are several different approaches for solving integer optimization problems [18]. For instance, for multiplexing hierarchies up to two levels, the integer optimization problem under study can be solved graphically (Fig. 8). In this Section, we apply a relaxation technique, following a methodology common to statistical physics [19]. In Section 3, we will apply an enumerative technique.

Relaxation techniques drop the integrality conditions and solve the resultant continuous optimization problem. In our case, this means that although n_i are positive integers, for the minimization, we assume they take values in a continuous range. Despite this approximation,

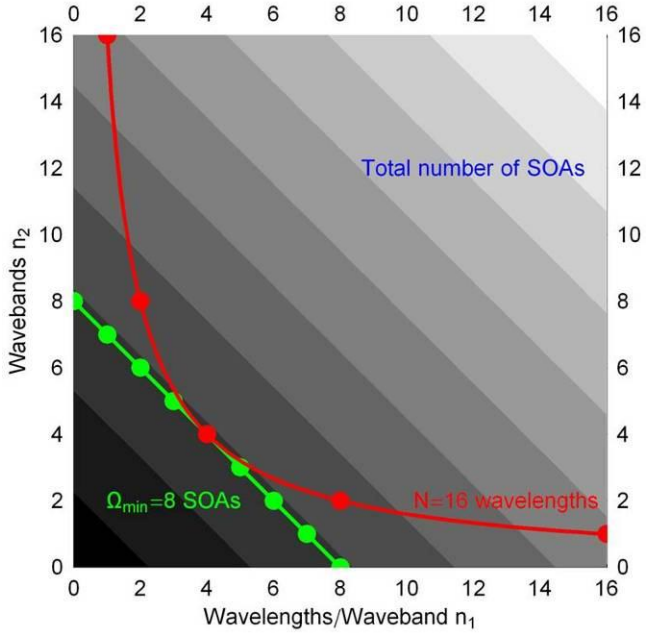

Fig. 8. Graphical solution of the constrained optimization problem regarding the optimal two-stage multiplexing hierarchy for 16 channels. (Symbols: Green line: cost function Eq. (2), Red line: constraint Eq. (1), Points: feasible integer solutions. Shades of gray: contours of the cost function. Shading becomes darker as we move to smaller values).

it will be shown by example that the results hold in the discrete case as well, at least for the order of magnitude of N of practical interest.

Taking the differential of Eq. (2) yields:

$$\delta\Omega = \sum_{i=1}^{K} \delta n_i = 0 \tag{4}$$

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

Taking the differential of Eq. (1) yields:

$$\sum_{i=1}^{K}\left(\delta n_i \prod_{\substack{j=1 \\ j\neq i}}^{K} n_j\right)=0 \tag{5}$$

Interchanging the order of summation and multiplication in Eq. (5) and assuming $n_i \neq 0$ yields

$$\sum_{i=1}^{K}\frac{\delta n_i}{n_i}=0 \tag{6}$$

We combine Eq. (4) and Eq. (6) using the Lagrange multiplier λ [17]

$$\sum_{i=1}^{K}\delta n_i-\lambda\sum_{i=1}^{K}\frac{\delta n_i}{n_i}=0 \tag{7}$$

or, equivalently,

$$\sum_{i=1}^{K}\delta n_i\left(1-\frac{\lambda}{n_i}\right)=0 \tag{8}$$

By equating the terms inside the brackets to zero, we obtain the set of K equations

$$n_i=\lambda \tag{9}$$

where $i = 1,\ldots,K$.

This important result indicates that the number of SOAs is minimized when the number of tributaries in all K channel selection stages is equal (*equipartition of tributaries*).

The value of the Lagrange multiplier λ can be calculated by replacing Eq. (9) in Eq. (1)

$$N \overset{(1)}{=} \prod_{i=1}^{K} n_i \overset{(9)}{=} \lambda^K \Leftrightarrow \lambda = \sqrt[K]{N} \tag{10}$$

The minimum number of SOAs per receiver card for K channel selection stages is calculated by replacing Eq. (10) into Eq. (2)

$$\Omega_{\min|K} \overset{(2),(10)}{=} K\sqrt[K]{N} \tag{11}$$

The minimum total number of SOAs required for K channel selection stages is

$$\Omega_{tot|K}^{\min} \overset{(3),(11)}{=} NK\sqrt[K]{N} \tag{12}$$

3.2 Analytical model: Optimization of the number of channel selection stages K

Equation (11) and Eq. (12) were derived assuming a fixed number of channel selection stages K. It would be desirable to optimize the number of channel selection stages K in order to achieve a global minimum of the number of SOAs.

Although K is a positive integer, in the following, it is assumed to take values in a continuous range. The optimum number of channel selection stages K_{opt} can be calculated by taking the derivative of Eq. (11) or Eq. (12) with respect to K. It is straightforward to show that

$$K_{opt}=\ln N \tag{13}$$

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

The global minimum number of SOAs per receiver card and total number of SOAs required are, respectively,

$$\Omega_{\min} \overset{(11),(13)}{=} \ln Ne \tag{14}$$

$$\Omega_{tot}^{\min} \overset{(12),(13)}{=} N \ln Ne \tag{15}$$

where e is the constant $2.718\ldots$.

There is a geometric interpretation of the results of the analytical model. Assume we have a K-dimensional rectangular hyperbox with edge lengths $n_1,\ldots n_K$. Equation (1) (i.e., the product of the length of all edges) can be interpreted as the "volume" of a hyperbox, whereas Eq. (2) (i.e., the sum of the length of all edges) can be interpreted as the "perimeter" of the hyperbox. Then, the constrained optimization problem under study can be stated as follows: we want to minimize the "perimeter" of a rectangular hyperbox in K-dimensions whose "volume" is given. In two dimensions, the answer to this problem is well known: the rectangle of minimum perimeter, when the area is fixed, is the square. The generalization of the problem in K dimensions shows that the hyperbox of minimum perimeter, when the volume is fixed, is the hypercube.

What is rather surprising is that the formal optimization indicates that among all hypercubes of the same volume N but with different number of dimensions K, the one with the globally minimum perimeter is when there are $K = \ln N$ dimensions and the edge length is equal to e.

The reader is cautioned that Eqs. (13)–(15) are approximate because of the discrete nature of K. In addition, the global minimum might not be achievable, because (i) K_{opt} is not, in general, an integer; and (ii) an equipartition of the tributaries into K_{opt} channel selection stages might not exist, as illustrated in the numerical examples in Section 3.

In spite of their inaccuracy, Eq. (14) and Eq. (15), should be considered to be lower bounds to the number of SOAs required per receiver card and per interconnect, respectively.

3.3 Analytical model: Figures of merit

For a given number of input/output ports (i.e., number channels N), numerous alternative discretely tunable receiver architectures may exist using different numbers of selection stages and tributaries per stage. The number of SOAs Ω required for each receiver architecture can be used as such a figure-of-merit.

To quantify the benefits of receiver architectures with a given number of input/output ports, one should adopt a normalized figure-of-merit. We propose the *optimality factor* defined as

$$F = \frac{\Omega_{\min}}{\Omega} \tag{16}$$

where $\Omega_{\min}$ is the theoretically minimum (albeit non-attainable) number of SOAs per receiver card given by Eq. (14), and Ω is the actual number of SOAs per receiver card given by Eq. (2). Obviously, $0 \le F \le 1$.

The number of SOAs required per receiver card, when using a single-stage channel selector, is equal to the number of channels N. The number of required SOAs per receiver card when using a multi-stage channel selector Ω is given by Eq. (2). An alternative figure-of-merit could be the *gain* in the number of SOAs defined as

$$G = \frac{N}{\Omega} \tag{17}$$

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

3.4 Model validation by direct enumeration

Optimal architecture of a 64 × 64 optical crossbar switch fabric

In the following, without loss of generality, we assume that N, n_i are powers of two, i.e.,

$$N = 2^{N'} \tag{18}$$

$$n_i = 2^{n_i'} \tag{19}$$

Using Eq. (18) and Eq. (19), the constraint Eq. (1) is rewritten in the form

$$N' = \sum_{i=1}^{K} n_i' \tag{20}$$

Equation (20) is the mathematical definition of an additive *partition*: an additive partition of a positive integer N' is a set of K strictly positive integers, whose sum is N'. Obviously, additive partitions which correspond to a permutation of the n_i' yield the same number of SOAs. Therefore, the order of terms in Eq. (20) is disregarded (*unrestricted partitions*). The 11 unrestricted additive partitions of 64 channels and the number of required SOAs given by Eq. (2) are calculated using *Mathematica* and are listed in reverse lexicographic order in Table 1 below. It is observed that, for a given number of channel selection stages K, equipartition of tributaries, when possible, yields the minimum number of SOAs, as predicted.

The number of possible equipartitions of tributaries can be calculated by finding the divisors m of N' so that

$$mK = N' \tag{21}$$

The divisors of 6 are $m = \{1,2,3,6\}$. Neglecting the trivial case $m = 6$, which corresponds to a single selection stage $K = 1$, the other three possible equipartitions of tributaries are

$$n_1 = n_2 = \cdots = n_6 = 2 \qquad (m = 1)$$
$$n_1 = n_2 = n_3 = 4 \qquad (m = 2)$$
$$n_1 = n_2 = 8 \qquad (m = 3)$$

From Table 1, it is observed that, from the aforementioned equipartition cases, the minimum number of SOAs occurs for $m = 1(K = 6)$ and $m = 2(K = 3)$ and is equal to $\Omega = 12$ SOAs per receiver card. The approximate formulae Eq. (13), Eq. (14) predict $K_{opt} = 4.16$, $\Omega_{\min} = 11.3$, respectively, which cannot be achieved in practice. It is worth noting that there are non-equipartition cases for $K = 4,5$ that also yield a minimum number of SOAs per receiver card ($\Omega = 12$).

Table 1. Unrestricted additive partitions of 64 channels in tributaries of K channel selection stages and number of SOAs required per receiver card for each realization.

Additive Partitions Eq. (20)						**Number of required SOAs/Rx Eq. (**2)
n_1'	n_2'	n_3'	n_4'	n_5'	n_6'	
6						64
5	1					34
4	2					20
4	1	1				20
3	3					16
3	2	1				14
3	1	1	1			14
2	2	2				12
2	2	1	1			12
2	1	1	1	1		12
1	1	1	1	1	1	12

In the following graph (Fig. 9), we compare the number of SOAs given by direct enumeration (Table 1) (points) and the theoretical formula Eq. (11) (solid curve). It is

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

observed that Eq. (11) is accurate when the number of selection stages is K = 1,2,3,6 because then, equipartitions of tributaries occur.

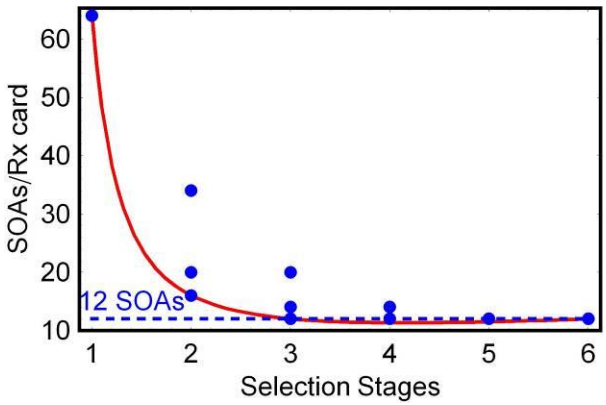

Fig. 9. Number of SOAs required per receiver card for 64 channels partitioned in tributaries of K selection stages. Solid line: expression (11), Points: last column of Table 1.

In the above numerical example, it is observed that there are four (sub-optimal) minima for K = 3 – 6 that correspond to Ω = 12 SOAs per receiver card. Obviously, it is preferable to organize the tributaries into tetrads (K = 3), since this reduces the numbers of required MUX/DMUX pairs, as well. This implementation is shown in Fig. 10(a) and 10(b).

In conclusion, when the number of channels and the number of tributaries per stage are restricted to powers of two, the optimal number of tributaries is:

- Four, when $\log_2 N$ is even.

- Two, for one selection stage and four, for all other selection stages, when $\log_2 N$ is odd.

In both cases, the minimum total number of SOAs required for the optical crossbar switch fabric architecture is $\Omega_{tot}^{\min} = 2N \log_2 N$, which is sub-optimal (i.e., slightly higher than the minimum given by Eq. (15)).

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

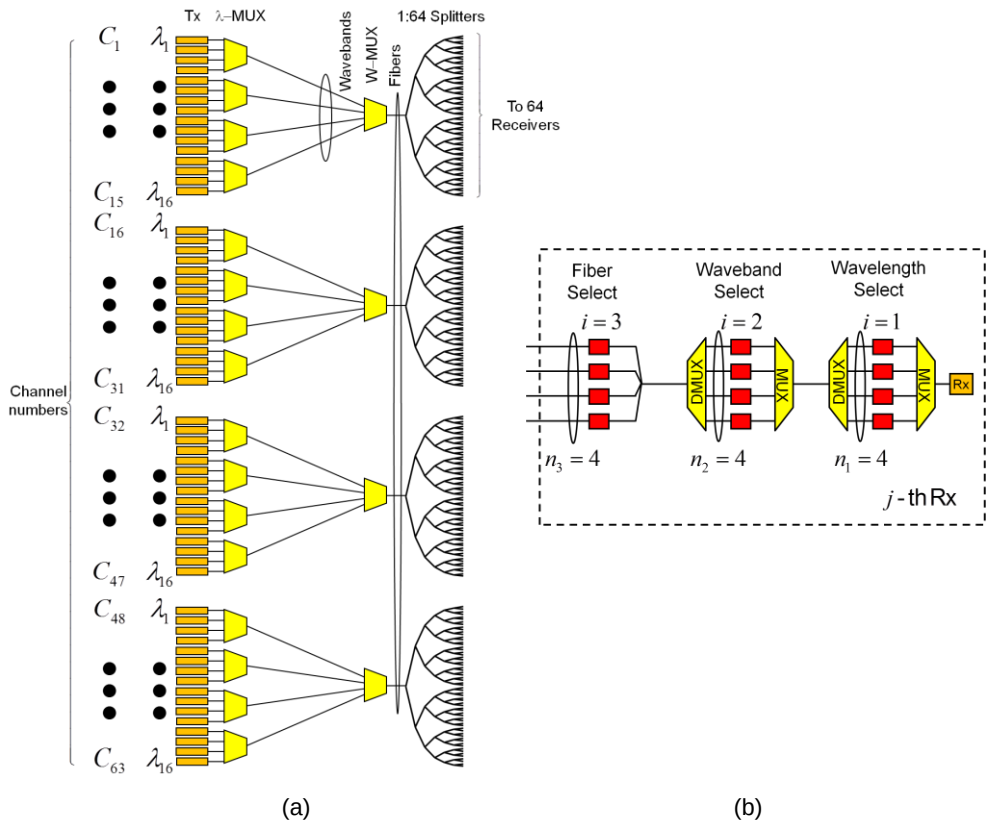

Fig. 10. (a) Implementation of the optimal, 64-channel, 3-level multiplexing hierarchy for the proposed optical crossbar switch fabric, and (b) Implementation of an optimal, 64-channel, discretely-tunable direct-detection receiver, for the multiplexing hierarchy shown in Fig. 10(a).

Further validation of the theoretical model is done for the following cases: (i) when the number of ports is a realistically large power of two (256 × 256); and (ii) when the number of ports is not a power of two and there are no possible equipartitions of tributaries (72 × 72 and 96 × 96 optical crossbar switch fabrics). In Fig. 11(a)-11(c), we compare the number of SOAs given by direct enumeration (points) and Eq. (11) (solid curve). It is observed that Eq. (11) is accurate when the number of selection stages K is such that equipartitions of tributaries occur.

The optimality factor is used to compare the results of direct enumeration for 64 × 64, 72 × 72, 96 × 96 and 256 × 256 optical crossbar switch fabrics (Fig. 11(d)). It is observed that in all aforementioned cases, it is possible to approach within 95% of the optimum, by appropriately choosing the number of selection stages. This is due to the fact that, for all the above cases, it is possible to partition the number of channels N into a product of small primes.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

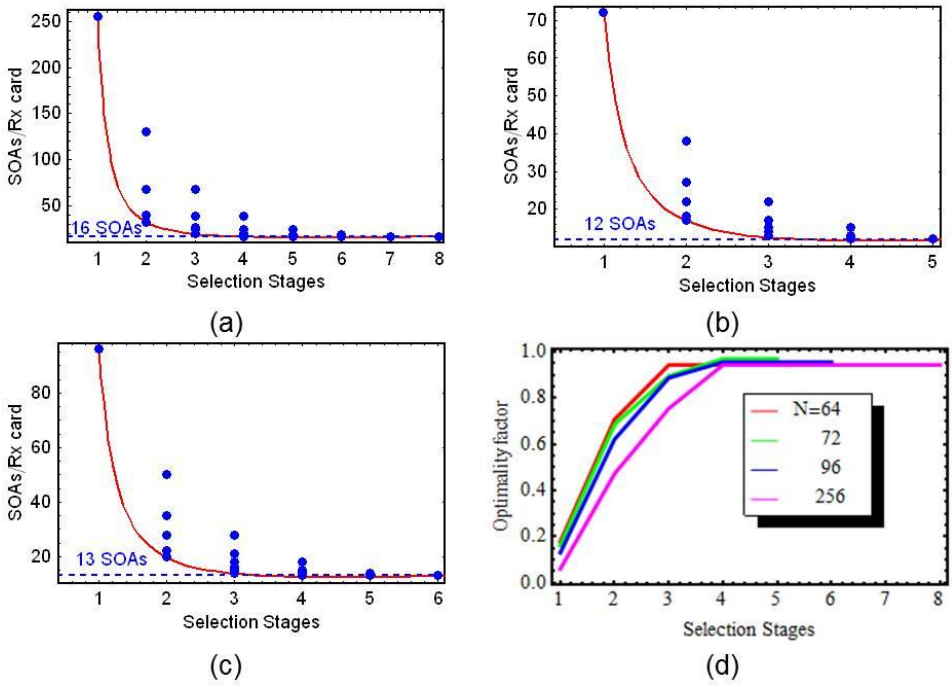

Fig. 11. Number of SOAs required per receiver card for (a) 256, (b) 72, and (c) 96 channels partitioned in tributaries of K selection stages. Solid line: Eq. (11). (d) Comparison of the optimality of the best channel selectors for 64, 72, 96 and 256 channels, as a function of the number of selection stages

Optimality comparison: General case

In the general case, the number of channels N can be factored into a product of M primes $p_1,\ldots,p_M$ as follows

$$N = \prod_{i=1}^{M} p_i^{m_i} \tag{22}$$

where m_i denotes the multiplicity of the prime factor p_i.

The minimum number of SOAs is achieved by using K selection stages, where

$$K = \sum_{i=1}^{M} m_i \tag{23}$$

Then, the minimum number of SOAs is

$$\Omega = \sum_{i=1}^{M} m_i p_i \tag{24}$$

Substituting Eq. (24) into Eq. (16) and Eq. (17), it is possible to calculate the optimality factor and the corresponding gain in the number of SOAs as a function of the number of channels for the optimal architectures, without performing an exhaustive search of all possible realizations of the switch fabric (Fig. 12(a) and Fig. 12(b), respectively). The results vary greatly, depending on if the number of channels N can be factored into a product of small primes or not. As we can see from Fig. 12(a), the worst performance is obtained for large primes. In practice, it is rather unlikely that the number of ports of an optical crossbar switch

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

will ever be a large prime. Nevertheless, one could achieve a reduction of the number of SOAs given by Eq. (24), even in this unlikely case, by using the same multiplexing hierarchy as for the smallest following integer with a high optimality factor and simply leaving some wavelength slots unpopulated. Figure 12(a) indicates that all powers of three approach within 1% of the optimum, whereas powers of two approach within 5% of the optimum. This somewhat surprising result in favor of the powers of three is counteracted by the fact that the powers of two allow one to organize the tributaries into tetrads, and this reduces the numbers of required MUX/DMUX pairs, as well. Therefore, minimization of a more fair cost function, that would factor the relative cost of MUX/DMUX pairs as well, would favor the powers of two, as shown in Section 5. It is worth noting that, with the current cost function, values of N which can be factored into a product of powers of two and three, exhibit a better optimality factor than the powers of two. Finally, Fig. 12(b) indicates that the use of an optimized multichannel selector generally leads to larger SOA savings as the number of channels increases, provided that the latter can be factored into a product of small primes.

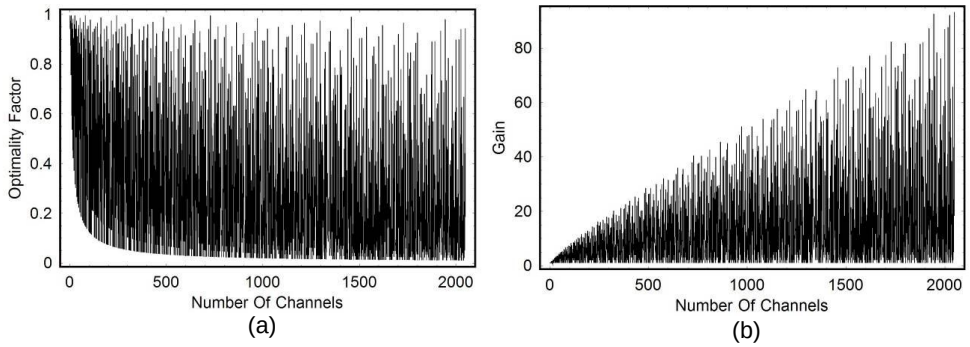

Fig. 12. (a) Optimality factor and (b) Gain in the number of SOAs as a function of the number of channels, using a multi-stage channel selector.

4. Simplified control algorithm

In the case of K channel selection stages and equipartition of tributaries into λ tributaries/selection stage, it is possible to use the following simplified control algorithm for setting the states of the switches:

i. Sequentially number the transmitters and receivers in the decimal numeral system starting from zero, using the tags l, m, respectively (i.e., $l, m \in [0,\ldots,N-1]$.

ii. Sequentially number the SOAs of each selection stage in the decimal numeral system starting from zero, i.e., in the range $[0,\ldots,\lambda - 1]$.

iii. Express the transmitter tags l in base-λ form, i.e., (λ_K … λ_1), where λ_K is the most significant digit and λ_1 is the least significant digit.

iv. Represent the currents of the SOA array of each selection stage by a unit column vector with zeros everywhere except for a ONE in the $(j + 1)$ – th position

$$\upsilon_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \qquad i = K,\ldots,1$$

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

A ONE at the (j + 1) – th row of the i – th unit column vector υ_i means that the j – th SOA on–off gate of the i – th channel selection stage is ON, whereas all other SOA on–off gates are OFF.

v. For an interconnection between the transmitter-receiver pair (l, m), set the on-off SOA gates of the K channel selection stages of the m – th receiver using the following rule:

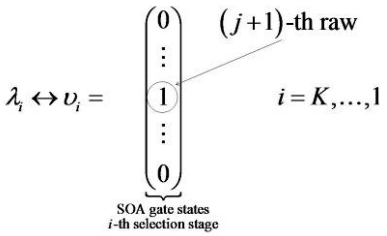

Fig. 13. The rule used for setting the on-off SOA gates of the K channel selection stages of the m-th receiver.

According to this rule, depicted in Fig. 13, there is a ONE at the (j + 1) – th row of the unit column vector υ_i if the number λ_1 corresponds to the (j + 1) – th position of the discrete interval $\lambda_i \in [0,\ldots,\lambda-1]$.

It is worth noting that there is a similarity between this algorithm and one previously proposed for the control of the Mach-Zehnder interferometer chain [20].

The implementation of the aforementioned control algorithm, in the optimal 64 × 64 optical crossbar switch fabric, is shown in Fig. 14(a) and 14(b). In order to establish a connection between the transmitter-receiver pair (37, 16), first the transmitter index is written in base-4 (i.e., 37 = $(211)_4$). Then, each one of the three digits of the base-4 representation is used to set the state of the SOAs of a corresponding selection stage (shaded (red) rectangles in Fig. 14(b) denote SOAs at the ON state). Figure 15 shows that, eventually, the desired channel is selected.

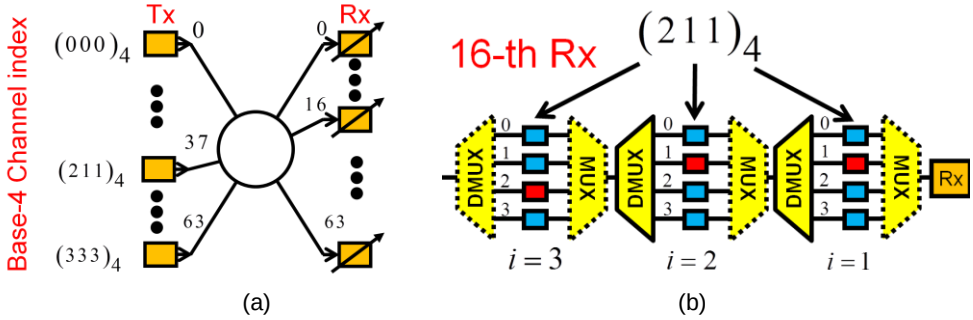

Fig. 14. (a) Example of transmitter/receiver numbering in a 64 × 64 broadcast and select architecture. In parentheses, channel indices are expressed in base-4. In this example, the transmitter with index #37 [$(211)_4$in base-4] is connected to the receiver with index #16. (b) Layout of the 16-th receiver card with 3 channel selection stages. The SOA numbering of each selection stage is shown. The three digits of the base-4 transmitter index are used to set the shaded (red) SOAs ON at the three selection stages.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

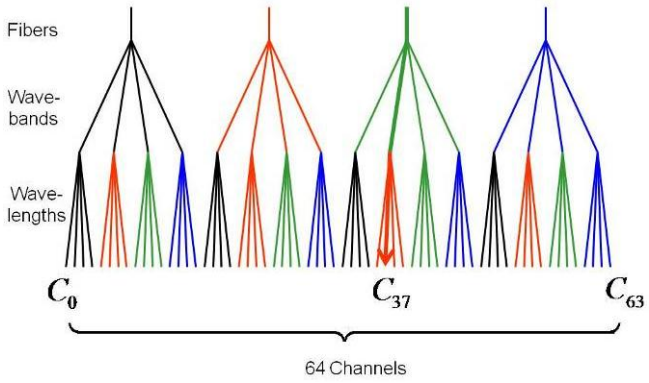

Fig. 15. The SOA setting shown in Fig. 14(b) selects the channel with index #37 out of the 64 wavelength tree, by first selecting the third fiber, then, the second waveband of this fiber, and finally, the second channel in this waveband (thick solid line).

5. Generalization for other cost functions

The theoretical model of Section 3 can be generalized for a cost function that takes into account the price of a MUX or DMUX *A* and an SOA *B*

$$\Omega = \underbrace{2KA}_{\text{MUX/DMUX cost}} + \underbrace{B\sum_{i=1}^{K} n_i}_{\text{SOA cost}} \tag{25}$$

Expression (25) assumes that a MUX/DMUX pair is used at all selection stages of the discretely tunable receiver. However, as pointed out in the introduction, MUXs used at all selection stages could be substituted by combiners, with a slight deterioration in performance, due to adjacent channel crosstalk. In addition, when SDM of several fibers is used in place of the last level of WDM multiplexing at the transmitter, the DMUX of the first selection stage, at the discretely tunable receiver is omitted. Furthermore, the MUX/DMUX pair of the first selection stage does not have to be AWG. These alternative designs have different cost functions than Eq. (25). Finally, Eq. (25) does not take into account the MUX cost at the transmitter side.

Using the method of Section 3, it is straightforward to show that Eq. (9) still holds, i.e., the new cost function is minimized when the number of tributaries in all *K* channel selection stages is equal (*equipartition of tributaries*).

The minimum cost is calculated by replacing Eq. (10) into Eq. (25)

$$\Omega_{\min|K} \overset{(10),(25)}{=} 2KA + BK\sqrt[K]{N} \tag{26}$$

The optimum number of channel selection stages K_{opt} can be calculated by taking the derivative of Eq. (26), with respect to *K* It is straightforward to show that

$$K_{opt} = \frac{1}{x}\ln N \tag{27}$$

where *x* is the root of the transcendental equation

$$x - 1 = re^{-x} \tag{28}$$

and $r = (2A / B)$ is the relative cost between a MUX/DMUX pair and a SOA.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

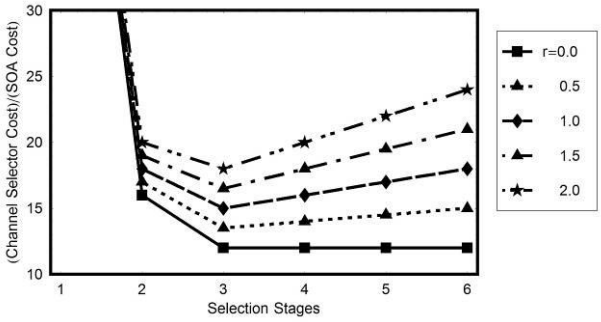

Fig. 16. Normalized cost function (per unit SOA cost) for N = 64 and r = 0 – 2 using direct enumeration. (Symbols: expression (25) and best realizations of Table 1).

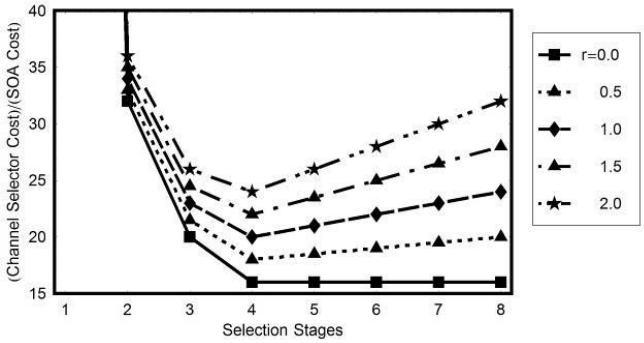

Fig. 17. Normalized cost function (per unit SOA cost) for N = 256 and r = 0 – 2 using direct enumeration.

The global minimum number of SOAs per receiver card and total number of SOAs required per crossbar switch fabric are, respectively:

$$\Omega_{\min} \overset{(26),(27)}{=} y \ln N \tag{29}$$

$$\Omega_{tot}^{\min} \overset{(3),(29)}{=} yN \ln N \tag{30}$$

where y denotes the auxiliary variable

$$y = B \frac{r + e^{x}}{x} \tag{31}$$

Figure 16 shows the normalized channel selector cost (per unit SOA cost) for N = 64 and r = 0 – 2 for the most efficient (in terms of SOAs) realizations listed in Table 1. For r = 0, it is assumed that the MUX/DMUX pairs have negligible cost. Then, channel selectors with K = 3 – 6 stages cost the same (i.e., the equivalent price of 12 SOAs, in agreement with Fig. 9). For r > 0, this degeneracy is lifted. It is observed that for all values of r > 0, the most cost-efficient, discretely tunable receiver should have three selection stages.

Figure 17 shows similar results for N = 256 and r = 0 – 2 for the most efficient (in terms of SOAs) realizations listed in Table 1. It is observed that for all values of r > 0, the best discretely tunable receiver should have four selection stages.

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012

6. Conclusions

The purpose of the present theoretical study was to optimize the architecture of an all-optical, wavelength-space crossbar switch fabric for HPC optical interconnects. The latter was implemented using a broadcast-and-select architecture, with fixed-wavelength transmitters and discretely-tunable, direct-detection receivers. Discretely-tunable, direct-detection receivers were implemented using multiple stages of AWG MUX/DMUXs, with progressively decreasing bandwidths and free spectral ranges, and semiconductor optical amplifiers (SOAs) as on-off gates. A significant reduction of the number of semiconductor optical amplifiers was achieved by exploiting the periodicity of the transfer functions of the AWG MUX/DMUXs. The total number of on-off gates required for the proposed switch fabric can scale asymptotically as $N \ln N$, if the number of input/output ports can be factored into a product of small primes. This is of the same order of magnitude as Shannon's lower bound for switch complexity.

The feasibility of the aforementioned optimal optical interconnects depends on the transmission impairments induced by the optical components. Nonlinear effects, esp. self-gain and cross-gain modulation, and polarization dependent gain in semiconductor optical amplifiers are the major limiting factors of the maximum number of channel selection stages. A separate article addressed these issues [21].

Acknowledgments

The authors are indebted to Prof. S. Daskalakis, University of Patras, Greece, for her remarks on integer optimization, which helped to improve the clarity of the mathematical model, and to Mrs. L. M. Szemiot, for corrections in the syntax of the manuscript. The OSMOSIS project was sponsored by the U.S. Department of Energy as part of the Accelerated Strategic Computing Initiative (ASCI).

#162358 - $15.00 USD Received 1 Feb 2012; revised 9 Apr 2012; accepted 22 Apr 2012; published 21 Aug 2012