Simulation and Optimized Design of High Density Optical Crossconnect Systems for Massively Parallel Computing Architectures

Ulrich Lohmann[1,3], Jürgen Jahns[1], Steffen Limmer[2], and Dietmar Fey[2]

[1] Fernuniversität Hagen
Optical Information Technology
[2] Friedrich Alexander University Erlangen Nürnberg
Department of Computer Science 3
[3] Fernuniversität Hagen
Optische Nachrichtentechnik
Universitätsstrasse 27 /PRG, 58097 Hagen, Germany
Ulrich.lohmann@fernuni-hagen.de

Abstract. We demonstrate the simulation results of a skewless high density approach for a multi-channel optical cross connect using integrated free-space optics. Using the 3D nature of free-space optics, this approach is able to solve geometrical problems with static crossings of the signal paths that occur in waveguide optical and electrical interconnection, especially for large number of connections.

Keywords: High density I/O, high performance planar optical interconnection, Clos networks, modeling simulation and evaluation techniques, optical interconnections for parallel computing, parallelization of simulation, Evolutionary Algorithm, optimization, skewless parallel gigabit/s interconnections.

1 Introduction

Optical chip-level interconnections in the data-rate range of ≥ 10 *Gb/s* are becoming more and more interesting in the "short haul" range both for closely-coupled parallel computing and data center applications. In comparison to conventional electrical interconnections over metallic wires, optical connections have advantages with regards to the signal consistency in data rates up to several *Gb/s* [6,11].

Additionally, the problem of geometrical signal path crossing, which is essential in the case of Clos networks with large numbers of channels, can be circumvented by using 3D optical interconnection. Here, we consider in particular the implementation of the interconnection by using the concept of planar-integrated free-space optics (PIFSO) [9,10] combined with MT fiber-optics technology. The free-space approach offers the possibility of a high interconnection density and a significant reduction of crosstalk problems [2].

The goals that we pursue with this approach are: reduction of the physical dimension of a cross connection segment and elimination of skew. The focus of this

S. Dolev and M. Oltean (Eds.): OSC 2010, LNCS 6748, pp. 42–52, 2011.

work is to present a logical model for the cross interconnect and an optimization approach that allows easy implementation and management.

This holds e.g. for the topology of the links in Clos networks which are considered in this paper. In [1] we investigated the possibilities and benefits for a system architecture using wavelength-division-multiplexing techniques for optical Clos networks. In this paper we now move a step towards the hardware realization and present a solution for designing a concrete geometrical layout for the optical links.

The design of such an optical interconnection system on hardware layer is a complex procedure. Hardware parameters have to be optimized which are inversely correlated, i.e. optimizing one parameter can cause a degradation of another parameter.

The rest of the paper is structured as follows. Chapter 2 shows a brief review of Clos networks. Chapter 3 presents the mathematical modeling of Clos networks realized in PIFSO technology. This model is the base for the optimization of some PIFSO technology parameters for which we used an Evolutionary Algorithm (EA) due to the complexity of that task. This EA approach and the yielded design results are the topic of chapter 4. Finally chapter 5 closes the paper with an outlook.

2 Clos Networks

Clos networks are a form of multistage switching networks, first presented by Clos in 1953 ([3]). Clos networks are very important for parallel computer architectures. They have been used in former parallel architectures, e.g. in the MasPar [12], and they are found as well in the networks of current parallel computer, e.g. in InfiniBand [13] switches. A Clos network operates semantically like the well-known crossbar interconnect in the sense that both networks offer universality, i.e. all connection requirements between input and output ports can be switched conflict-free in principal. The advantage of a Clos network is that the number of required crosspoints can be much fewer compared to a network implemented with just one big crossbar.

An $n \times n$ Clos network has three stages (see Figure 1). The first stage consists of switches of size $r \times k$. Since we have n inputs in total, there must be n/r of those switches. In the second stage are k switches each of size $n/r \times n/r$. Every switch in stage two has a connection to every switch in stage one. The third stage has n/r switches of size $k \times r$. Again every switch in stage two has a connection to every switch in stage three. Between each of the subsequent stages (i.e. between the first and second stage, and the second and the third stage) the topology of the links is as follows. The i^{th} output port of the j^{th} switch in the input stage has a 1-to-1 link to the j^{th} input port of the i^{th} switch in the subsequent stage.

It can be shown that for $k \geq 2n\text{-}1$ the network is nonblocking and for $k \geq n$ it is rearrangeable nonblocking. That means, an unused input can always be connected to an unused output, but a rearrangement of the existing connections may be necessary. The number of crosspoints that can be saved compared to a network of just one crossbar depends on the exact choice of n and r.

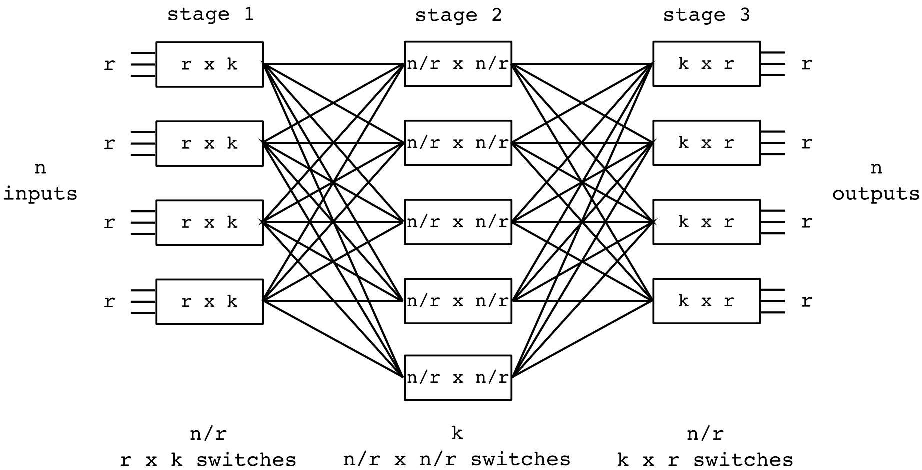

Fig. 1. Topology of an $n \times n$ Clos network

3 Mathematical Modeling of PIFSO *n* × *n* Cross Connects

Like described in Section 1 the realization of the cross interconnections between the stages of a Clos network in PIFSO technology can offer many advantages compared to classical fiber optical or electronic interconnections. The geometrical relations inside the optical plane are defined as shown in Figure 2. With the condition, that the angle of incident is the same as the angle of exit, we can write the total length of the needed interconnection from point A to point B as Equation 1.

$$L_{AB} = \sqrt{(\Delta x_{AB})^2 + (\Delta y_{AB})^2} = n_{FI} \bullet 2 \bullet \frac{H}{\tan \beta} \tag{1}$$

The terms under the square root are the geometric distances between the in- and out-coupling point at the plane $z = H$. The optical path length L_{opt} of a channel can be computed as shown in Equation 2. Equation 3 shows the ratio of L_{opt} to L_{AB}.

$$L_{opt} = n_{FI} \bullet 2 \bullet \frac{H}{\sin \beta} \tag{2}$$

$$\frac{L_{opt}}{L_{AB}} = \frac{1}{\cos \beta} \tag{3}$$

Prior to the realization of a Clos network in PIFSO technology it is necessary to determine concrete values for the deflection angle and the number of folding intervals for every connection. These values should be constituted in such a way, that

1. The light rays hit the output fibers with a minimal position aberration as possible. That means Equation 1 should be satisfied as best as possible.

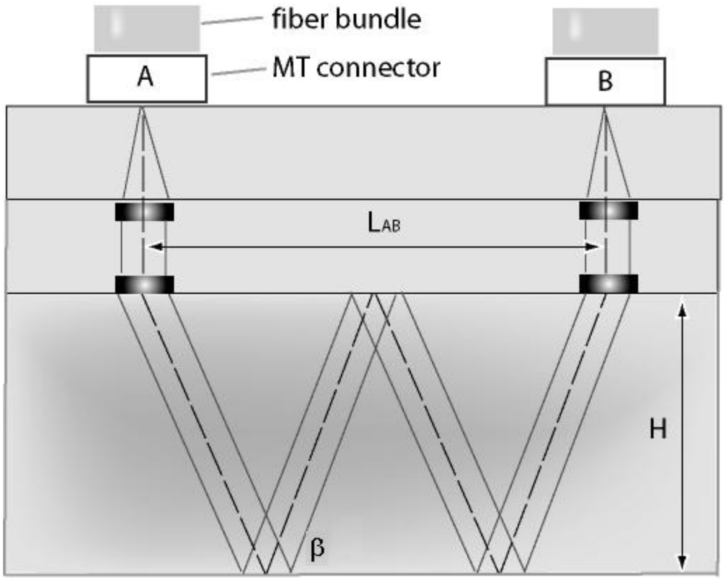

Fig. 2. Geometry of the PIFSO system. The light gray areas indicate optical substrate. The light paths are "folded" into the bottom substrate and propagate under an angle β relative to the substrate surface. n_{FI} denotes the number of double paths (also described here as "folding interval").

2. The aberrations of the optical path lengths L_{opt} of the rays from the average optical path length are as small as possible to minimize the skew effect for parallel lines. For example, a path length difference of 10 *mm* in SiO_2 causes a signal delay of 50 *ps*. This is a quarter of a 10 *Gb/s* digital signal pulse width.

3. The deflection angles lie in the range between 32° and 43° to insure total reflection.

The second condition is the reason for the zig-zag paths of the beams. In this way it is possible to adjust the optical path lengths by an appropriate choice of the deflection angle values and the number of folding intervals. Figure 3 shows a 4×4 Clos network with optimized values as described above.

If one has given the number of folding intervals for one connection, then all the other wanted values can be computed analytically, so that there are no position aberrations and the sum of the aberrations of the optical path lengths are minimal (under the condition that the position aberrations are zero). But with this approach the third condition is not regarded.

Since it is hard or even impossible to find an analytical approach for solving this problem with regard to all three conditions, we use Evolutionary Algorithms (EAs) for this task. The following section gives a short introduction to EAs.

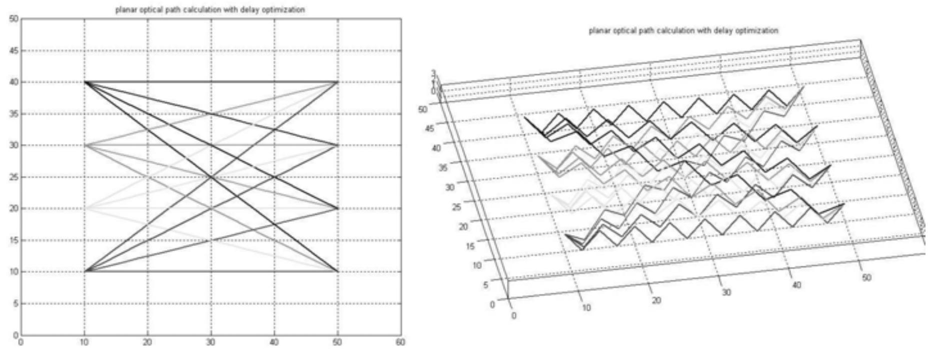

Fig. 3. Top and side view of a skewless optical *4 × 4* Cross interconnect as a part of a Clos network (Example from Figure 1). In the case of integrated free space optic, the path crossing problem is not relevant and allows high density network constellations with more channels.

4 Optimizing PIFSO Clos Networks with Evolutionary Algorithms

4.1 Evolutionary Algorithms

EAs are a popular form of iterative heuristic optimization algorithms ([4]). The process of an EA is oriented in biological evolution. The basis is always a (mostly randomly chosen) population of individuals, where an individual is a possible solution of the problem for which an optimum is searched. The individuals are encoded in an adequate form (in general number values). Such an encoded individual is called a *chromosome*. An individual is in general composed from several parameters. Such a parameter in encoded form is called a *gene*. Every individual will be associated with a real valued *fitness* which expresses how optimal the individual for the examined problem is.

The objective of the algorithm is to create new generations of populations, containing individuals with a higher or respectively smaller fitness and to find in this way an individual with a fitness as maximal or minimal as possible. For the creation of new populations, *genetic operators* are used. These operators are *selection*, *recombination* and *mutation*. Selection is the task of selecting a fixed number of so called parent individuals from the given population. This happens mostly randomized where individuals with a higher fitness have a higher chance to be chosen. Using recombination, offspring are generated from the parent individuals. Mostly one or more offspring are created from two parent individuals. It is desired that parents with a good fitness yield offspring with good fitness values. Subsequent mutation is carried out on the offspring, e.g. random modifications on genes of the offspring will be performed (with a small probability). After that the fitness values of the produced offspring will be computed and they are inserted in the old population according to a certain strategy. By repeating the described procedure continuously, new generations of populations are produced until a certain break condition is met.

4.2 Optimizing $n \times n$ Clos Networks

We implemented an EA that computes the optimal deflection angles and number of folding intervals. An individual consists of a sequence of n folding interval numbers n_{Fi} (one for each of the n rays) that are encoded as integers. The fitness of an individual is computed as the sum of the aberrations of the optical path lengths from the average optical path length and should be minimized in our case. The deflection angles for the different connections are computed directly from the number of folding intervals so that the position aberrations at the outputs are zero. Thus, the position aberrations do not have to be considered for the fitness evaluations. The algorithm turned out to be robust and fast. Robust in that sense, that it yields the same results in almost every run. So the number of runs we performed for obtaining the below mentioned results is not relevant. A mutation rate of $3/n$ was used, so that in average three folding interval numbers of an individual are mutated. We used a population size of 100 and a crossover rate of 0.9. Several thousand generations were computed per run..

It is assumed that there are $\sqrt{n}$ in- as well as output MT connectors with $\sqrt{n}$ fibers each. The connectors are arranged on "top" of each other, as shown in Figure 4. Between two fibers of a connector is a pitch of 0.25 *mm* and between two connectors is a pitch of 4 *mm*. The length of the transmission distance is 100 *mm*. For the height of the PIFSO module we chose 1 *mm*.

For $n = 16$ the algorithm yielded a result where all delays of the optical paths lie in tolerable ranges (under 1 *mm*). However, for higher dimensions like $n = 256$ this is not the case. For $n = 256$ the result has a fitness of 1248.08 (the sum of the delays of all paths in *mm*) and there are several paths with a delay of the optical path larger than 1 *mm*, what is not acceptable.

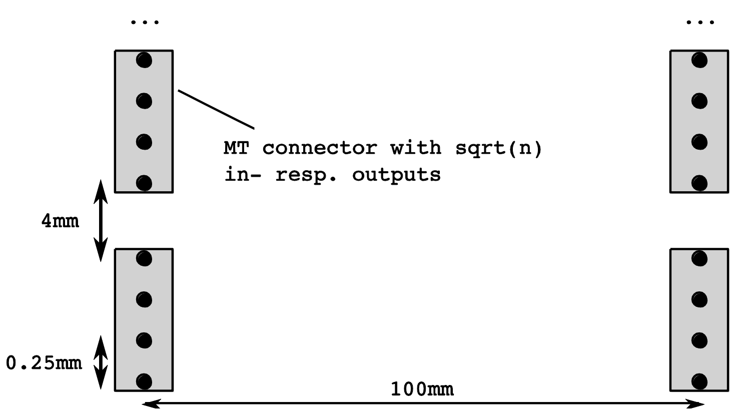

Fig. 4. Schematic illustration of the geometry of a 16 × 16 network

Better results can be found for other arrangements of the connectors. Figure 5 shows different grid-like arrangements of the 16 input connectors (output connectors are arranged in the same way) and the resulting fitness values.

We can see that the best result was achieved with an *8 × 2* arrangement, but here, too, the computed delays are not adequate, yet. From all results, we got so far, it is predictable that the connectors should be arranged as regular and near to each other as possible. Thus, "displacing" individual connectors in the grid-like arrangement is not promising. But by displacing all outputs (resp. analogue the inputs) in x- and/or y-direction (see Figure 6) the result can be further improved.

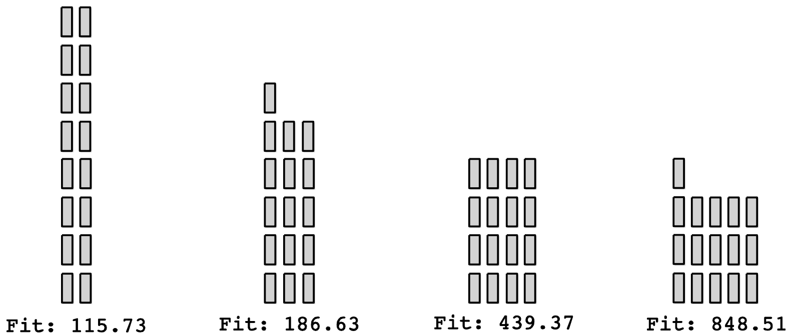

Fig. 5. Four different grid-like arrangements of 16 connectors and the fitness values that were achieved with them

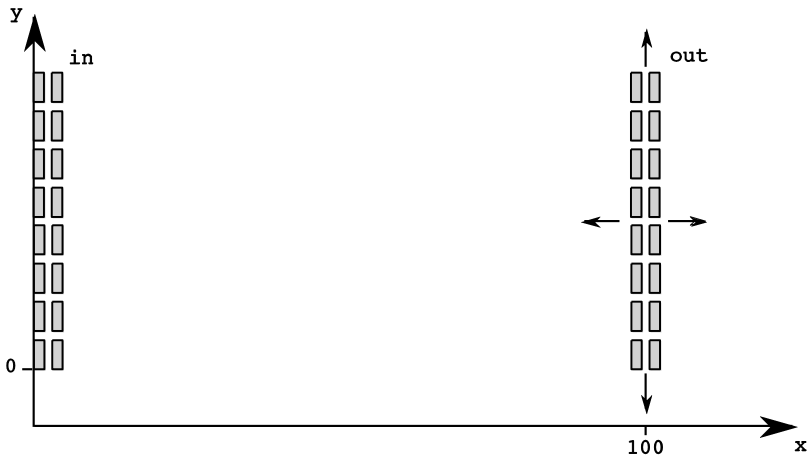

Fig. 6. Displacement of the output connectors from the initial position in x- and y-direction

Figure 7 shows the fitness values dependent on the displacement of the outputs in x-direction. A displacement by +7.1 *mm* yields a fitness of 88.225 and a displacement by +13.45 *mm* a fitness of 74.96. However, the resultant increase of the transmission distance may be problematic for the production. Via a displacement by -1.92 *mm* in x-direction and in addition by +0.2 *mm* in y-direction a fitness of 111.9995 can be achieved, the best value for a transmission distance of max. 100 *mm*, so far. But the results can be further improved, namely by rotating the connectors by 90°. This yields a fitness of 91.284, which can be improved to even 75.88 via displacement of the outputs by -0.7 *mm* in x-direction and +4.1 *mm* in y-direction. Here the optical paths of only 4 out of the 256 connections are not in the desired range. That is the best result, we achieved so far.

Better results are possible for smaller pitches between the in- resp. output fibers, a bigger interval of permitted deflection angles or a longer transmission distance. But all this may cause problems in implementation. A possible alternative is the employment of other connectors, like MPO Connectors ([7]) that allow an arrangement of the fibers of the connector.

4.3 Conclusions for the Design

In Figure 8a-c the results of the performed optimization are shown. The best founded values for deflection angle and number of folding intervals generate the shown values for optical path length variances. In the diagrams all 256 optical channels and their deviation values are shown. This optimization is performed for the rotated MT-connector arrangement.

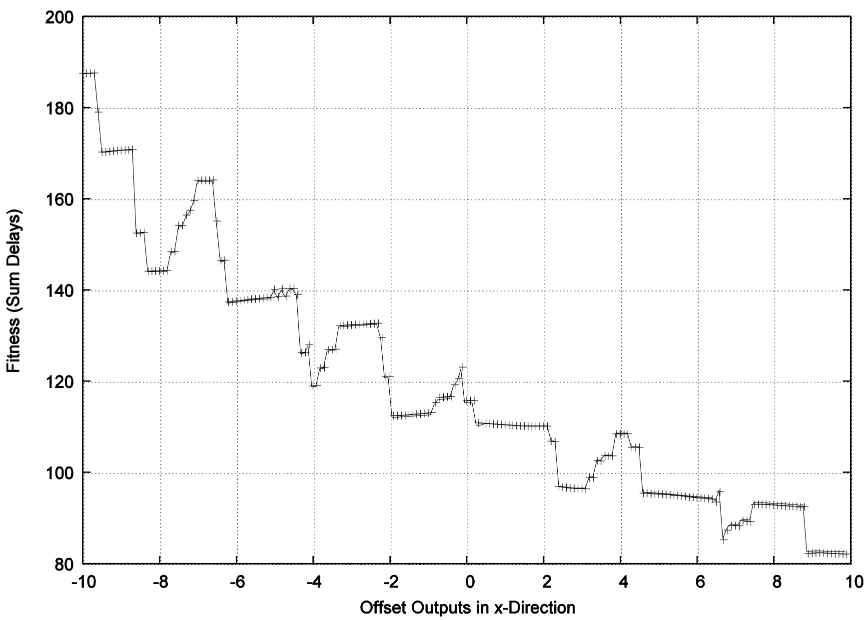

Fig. 7. Optimal fitness value subject to the displacement of the outputs in x-direction

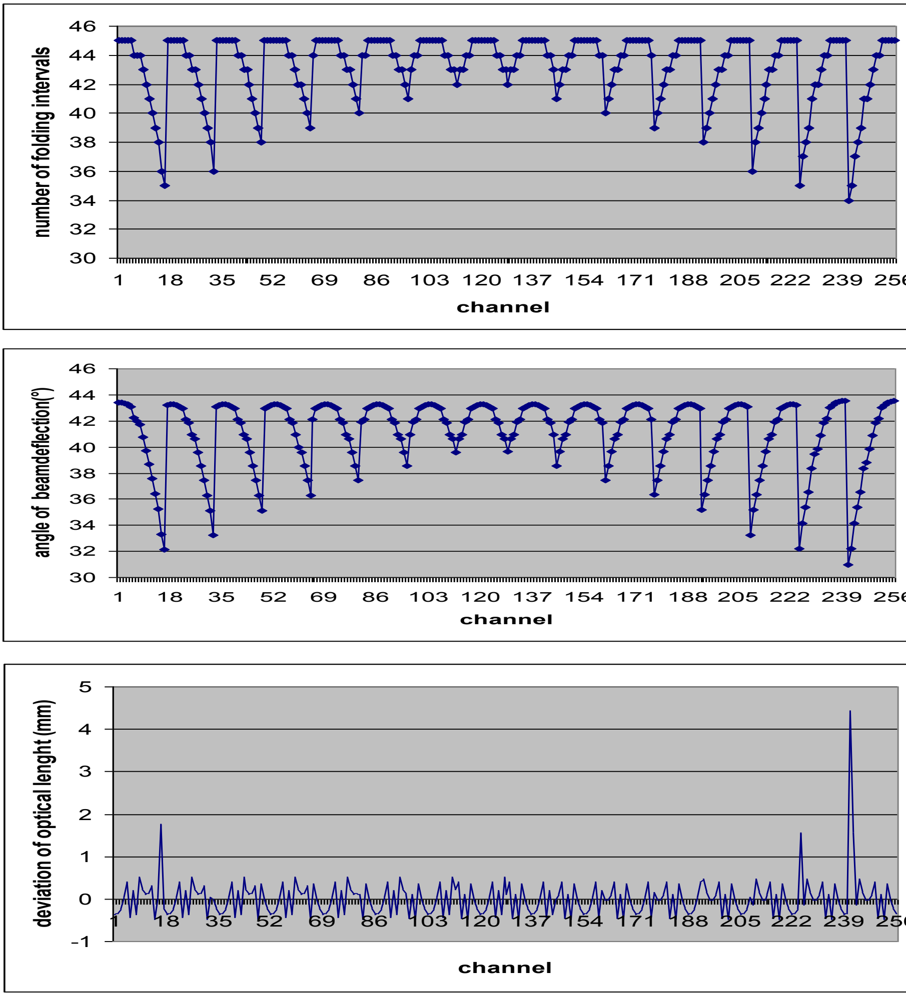

Fig. 8 a-c. The results of a) number of folding intervals, b) angle of beam deflection c) the deviation of the optical path length

5 Outlook

We have presented a novel approach to implement passive free-space optical cross interconnections to realize a high density Clos network. Although the results of the optimizations were not completely as desired, they can be used for an initial guess for design parameters.

In the next step we make use of the large number of individual mirror elements in a DMD array [8] to implement dynamic routing following a photonic network on chip. In Figure 10 the scheme of the experimental setup for the DMD approach is shown. The used optimization algorithm can be modified to calculate this DMD based dynamic routing with defined optical relations inside the PIFSO cuboid.

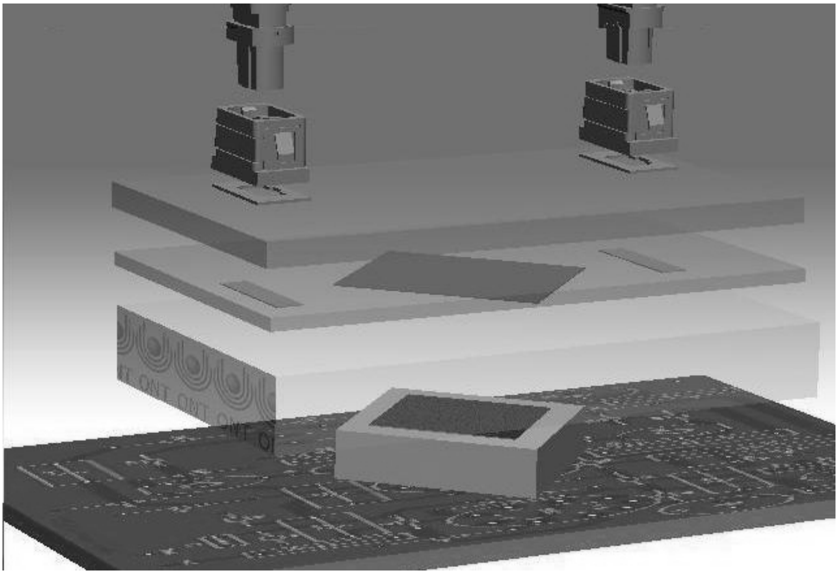

Fig. 9. Exemplary PIFSO & MEMS implementation of high density planar optical switch matrix Clos configurations (Photonic network on chip)

The results presented here, are first results that were obtained by using various simplifying assumptions. Next steps will include adapting the model step-by-step to realistic assumptions. Furthermore, our model and simulation tool will allow us to include modifications of the setup such as a variation of the pitch in the input and output fiber bundles and the geometrical relations for the optical beam deflection.

References

1. Fey, D., Schneider, M., Jahns, J., Knuppertz, H.: Optical Multiplexing Techniques for Photonic Clos Networks in High Performance Computing Architectures. In: Dolev, S., Oltean, M. (eds.) OSC 2009. LNCS, vol. 5882, pp. 110–123. Springer, Heidelberg (2009)
2. Jahns, J.: Digital Optical Computing and Interconnection. In: The Handbook of Photonics, 2nd edn. CRC Press, Boca Raton (2007)
3. Clos, C.: A Study of Non-Blocking Switching Networks. Bell Sys. Tech. J. 32, 406–424 (1953)
4. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
5. Miller, D.A.B.: Rationale and Challenges for Optical Interconnects to Electronic Chips. Proc. IEEE 88, 728–749 (2000)
6. Saleh, B.E.A., Teich, M.C.: Fundamentals of Photonics. Wiley, Chichester (2007)
7. Euromicron Corp. Datasheet: Specification of Standard MPO-Connectors. IEC (1996)
8. Texas Instrument Corp. Datasheet: Specification of Digital Mirror Devices (2009)
9. Jahns, J., Huang, A.: Planar integration of free-space optical components. Appl. Opt. 28, 1602–1605 (1989)

10. Gruber, M., Jahns, J., Sinzinger, S.: Planar-integrated optical vector-matrix multiplier. Appl. Opt. 39, 5367–5373 (2000)
11. Baudet, D., Braux, B., Prieur, O., Hughes, R., Wilkinson, M., Latunde-Dada, K., Jahns, J., Lohmann, U., Fey, D., Karafolas, N.: Innovative On Board Payload Optical Architecture for High Throughput Satellites. In: Proc. ICSO 2010 (2010)
12. El-Ghazawi, T.A.: Characteristics of the MasPar parallel I/O system, frontiers. In: Fifth Symposium on the Frontiers of Massively Parallel Computation (Frontiers 1995), pp. 265–272 (1995)
13. InfiniBand Trade Association, http://www.infinibandta.org (accessed November 2010)