Agelis, S. and M. Jonsson, "Optoelectronic router with a reconfigurable shuffle network based on micro-optoelectromechanical systems," *Journal of Optical Networking*, Optical Society of America (OSA), vol. 4, no. 1, pp. 935-944, Jan. 2005.

Optoelectronic Router with MOEMS–Based Reconfigurable Shuffle Network

Sacki Agelis and Magnus Jonsson

CERES – CENTRE FOR RESEARCH ON EMBEDDED SYSTEMS

School of Information Science, Computer and Electrical Engineering, Halmstad University,

Box 823, S-301 18, Halmstad, Sweden.

Sacki.Agelis@ide.hh.se

An optoelectronic router with a shuffle exchange network is presented and enhanced by adding micro-optical-electrical mechanical systems (MOEMS) in the network to add the ability to reconfigure the shuffle network. The MOEMS described here are fully connected any-to-any crossbar switches. The added reconfigurability gives the opportunity to adapt for different common application characteristics. Two representative application models are described, where the first has symmetric properties and the second has asymmetric properties. The router system is simulated with the specified applications and an analysis of the results is carried out. By using MOEMS in the optical network, and thus reconfigurability, over 50% increased throughput performance and decreased average packet delay is obtained for the given application. Network congestions are shown to be avoided throughout the system if reconfigurability is used.

KEYWORDS: MOEMS, VCSEL, switch, router, application, reconfigurability, simulation.

1. Introduction

In the community of reconfigurable systems, optics are still relatively unproven, particularly in the domain of telecommunications. Optics are in the verge of achieving the requirements that are needed in order to be considered in reconfigurable computer and communication architectures, especially when it comes to micro-optical-electrical mechanical systems (MOEMS). For quite some time MOEMS devices has been regarded too unstable, too expensive etc. Reports and articles [1-5] now claim that the former flaws of MOEMS are changed to strengths instead. Even the avionics industry is now contemplating to use MOEMS devices in next generation reconfigurable computer architectures [6] as reliability reports prove less than 37 failures over 1 million operating hours [7].

This article describes how MOEMS crossbar devices can be used in optoelectronic routers to avoid traffic

congestions for certain application domains, i.e. increase the overall throughput. The application domain targeted spans from homogenously distributed traffic patterns to unevenly distributed traffic patterns, e.g. video-on-demand, embedded parallel processing (radar signal processing and radio base stations signal processing), etc. These applications have a strong tendency of being heavily congested at some specific areas in the system as the traffic loads increase. In order to prove the gain of adding reconfigurability by using MOEMS in the backbone network a simulation and an analysis of the router system has been carried out. The optical MOEMS crossbar switch is used to reconfigure the shuffle exchange network, as seen in Figure 1, to fit the application characteristics. Hence, through reconfigurability, the router system has the ability to adapt for the current traffic conditions to increase the overall throughput. As an example, assume that the traffic from an electronic crossbar on the input side changes to be mostly destined to a specific electronic crossbar on the output side. The shuffle network can then be adapted to have more lines than normally between this pair of input and output crossbars. By simulating both the network architecture with reconfigurability and without, a comparison of the two is obtained.

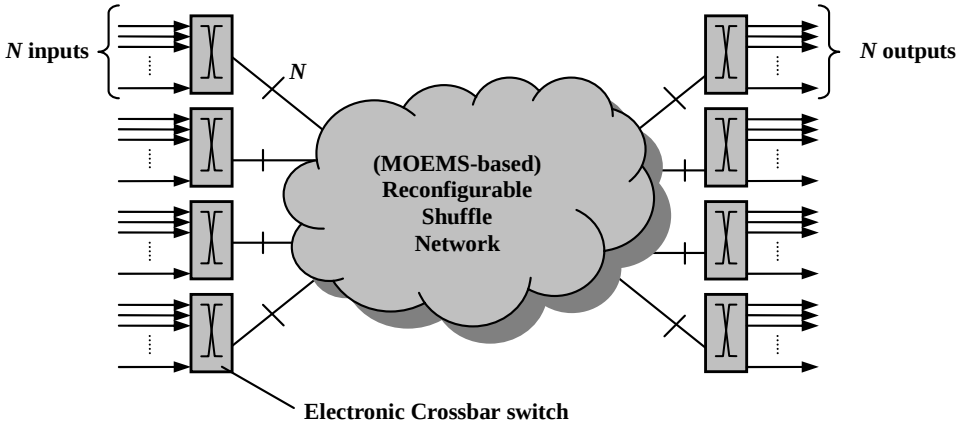

Figure 1: The general overview of the router.

Often, network traffic with asymmetric traffic patterns are relayed over routers and switches. These traffic patterns, or applications, do not change dramatically with each packet that is sent. *High Definition-TV* is one representative example of this kind of application, where an established connection lasts over a relatively long time, and the traffic might be concentrated to specific living areas or only from a few video sources. The overall communication pattern can change depending on, e.g., what the time of the day it is or depending on who is broadcasting the most popular TV program for the moment. The applications considered here follow a dataflow pattern that typically congests a local area in the system and a dataflow that does not change very often over the duration of a larger time span. Many types of these traffic patterns exist but for clarity we have chosen regular, but still asymmetric traffic patterns for the analysis in this paper, as follows.

In order to model the asymmetry in the application behavior we have a model for the dataflow. First, consider an application that has a homogenous distributed traffic pattern, a so called *symmetric* application. By symmetric we imply that all the traffic that enters an input switch board, in overall, is equally spread to all of the output switch boards. Hence, at full traffic load the ideal backplane should have all network resources evenly distributed because the application follows the appropriate behavior for that particular network type, or vice versa. However, as the application becomes more asymmetric, i.e. the board-to-board (or input to output crossbar) load becomes more imbalanced, the backplane ceases to suite the application and the throughput performance will drop as the traffic intensity increases. The bandwidth required may be extremely high or it may be very moderate. This property of the application is said to be *asymmetric*. In turn, this asymmetry will lead to problems such as; buffer overloads, backplane links working at maximum while other backplane links are used very modestly, etc. To avoid these problems the backplane needs to be adjusted or *reconfigured*, to fit the current traffic flow. This reconfigurability is achieved, as described in Section 3, by use of MOEMS devices which are described in Section 2.

All these different demands present a delicate problem since it is very hard to satisfy all of the varying demands with one static network type. This makes the use of MOEMS components very appealing, since they can provide robustness and compact solutions to a low cost and adds support for the varying asymmetric traffic patterns that can emerge in the application. In our previous work [8, 9] we have investigated how optical network architectures can be incorporated into parallel computing and communication systems. We have showed that the power of reconfigurability in communication equipment is best applied in the backplane where the highest gain is achieved to the lowest cost [9].

2. Optical Components

The switch boards are interconnected by using standard multimode fiber ribbons [10] or laminated fibers [11-14] for easy mounting and coupling. The fibers would then not have to wear their thick surrounding jackets and all the fibers would be placed in one physical plane, i.e. a foil. The advantage of using this technology is to save space and make coupling easier. On the other hand, standard fiber ribbon solutions can instead be used where up to 72 fibers can be integrated in one physical cable [15].

VCSELs [16] are used for transmission of light and equal amount of photo-detectors is used at the receiver end, i.e. N unidirectional optical communication lines from the input switch boards to the optical MOEMS switch, and then from the MOEMS crossbar to the receiving output switch boards at the output stage.

Optical transmissions are made by using standard low-cost VCSELs with a typical wavelength of 850 nm and with a maximum bit rate of 10 Gbit/s [11, 17].

The optical components are pop-up MOEMS with 8×8, 16×16 or 32×32 mirrors, as illustrated in Figure 2. The actuated flip-flop mirrors are monolithically integrated on a silicon chip. By pushing a selected

mirror, the collimated light is redirected to the desired output port as shown in Figure 2. The micro-actuated flip-flop switch mirror consists of a mirror (usually gold or aluminum) that provides a reflective surface to the incoming light beam. This mirror is in turn anchored on the silicon substrate. In the MOEMS discussed here, the optical beams move in a 2-dimensional space where each mirror is controlled to be in either down-position (no beam bounce) or up-position (beam bounce). This technology is seen to achieve low light-loss, low crosstalk, and low polarization dependence. One disadvantage with this technology is the relatively long reconfiguration time, which is in the order of milliseconds. The MOEMS should, however, in our case only be used to reconfigure the network, they should not be regarded as packet switches.

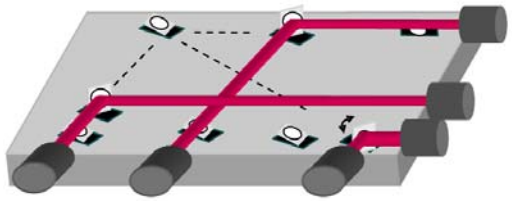

Figure 2: An $N \times N$ optical switch with pop-up mirrors, i.e. N^2 MOEMS mirrors.

Because these devices steer light beams they are bit rate independent, wavelength (i.e. channel) independent as well as protocol independent. This means that 80 or more wavelengths are possible at 10 Gbit/s per fiber wire, i.e. 800 Gbit/s in total per fiber wire [18]. Significantly important is that each device in a MOEMS system is relatively large so this technology does not push the limits of lithography and wafer fabrication processes. In general, last generation semiconductor processing equipment can be used.

The 2-dimensional MOEMS described here brings compact solutions for small or medium-sized systems. The disadvantage with the 2-dimensional MOEMS is that N^2 mirrors are needed in an $N \times N$ switch. A more scalable solution is to let the beams travel in 3-dimensional space and to use mirrors for which the angle can be controlled in two axes. In an $N \times N$ switch, two arrays of size $\sqrt{N} \times \sqrt{N}$ are used. Each beam will first bounce on an input specific mirror in the first array and then on an output specific mirror in the second array. Reports show that 112×112 switches with the capacity of 35.8 Tbit/s are possible and reliable [19]. Agilent announced a fully non-blocking 32×32 switch in March 2000 [20] and Lucent Technologies announced a 238×238 switches based on the 3-dimensional MOEMS technology with 56 644 possible connections [21], while switch sizes of 1024×1024 are prototyped and 4096×4096 port MEMS are considered to be reality in the next few following years. Due to the relatively low signal attenuation of 2 dB in today's MOEMS devices, G. Shen et al. [20] have showed that multistage MOEMS switches also are possible. This low attenuation is significantly important since it reduces the power requirements of the optical transmitters and receivers; hence the cost is also reduced [21].

3. Architectural Description

In this section we present the basic router architecture to then extent it with reconfigurability next. The router mainly consists of switch boards mounted and cooled in a large rack. The switch boards are interconnected to each other through a backplane interconnection network with a kind of all-to-all network topology. Generally, a switch board consists of several smaller network switches which in turn are interconnected internally on one board. As the number of I/Os increases several crossbars might be used to implement a single switch board. Here, an entire switch board is approximated to one single electrical crossbar switch for the following reason: in comparison to the backplane the communication internally on one switch board is faster. Hence, for the ease of reasoning a switch board will be defined as a single crossbar switch throughout this paper, as depicted in Figure 3. A switch board is modeled as an ideal crossbar with zero-delay internal communication transfers and infinite output port buffer sizes within the scope of this paper. All packets from the inputs are transferred to the output buffer (output port queue on the board) with the correct routing to an output switch board and with the least number of packets in the queue, i.e. least occupied buffer slots.

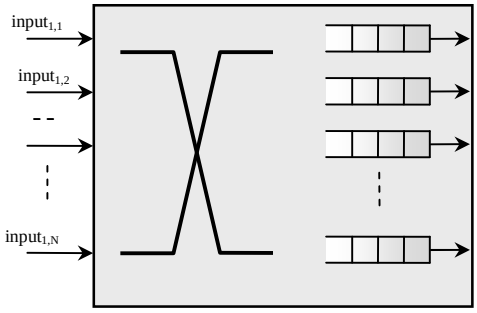

Figure 3: The switching board with buffers (queues).

As illustrated in Figure 4, the input switch boards are interconnected through a backplane network to the output switch boards. Hence, a shuffle exchange network is created with M input switch boards and M output switch boards, where each switch board holds N physical input wires and N physical output wires. Here, we will assume a router size of $M = 4$, i.e. two columns of switch boards where each column hold M switch boards, and $N = 16$, i.e. 64×64 inputs/outputs for the complete system. The interconnection network between the two stages is hereafter denoted as the *backplane* or *network*.

For a switch node with $N = 16$ outputs, four physical wires are connected to each of the output switches. This intuitively implicates that the router is optimized for evenly distributed applications, so called *symmetric* applications. If the network traffic has *asymmetric* properties as described in Section 4, the performance will

decrease drastically as the traffic load increases. Reconfigurability in the backbone network can avoid these drawbacks. The symmetric network will be used for comparisons reasons, in order to see what gains can be achieved.

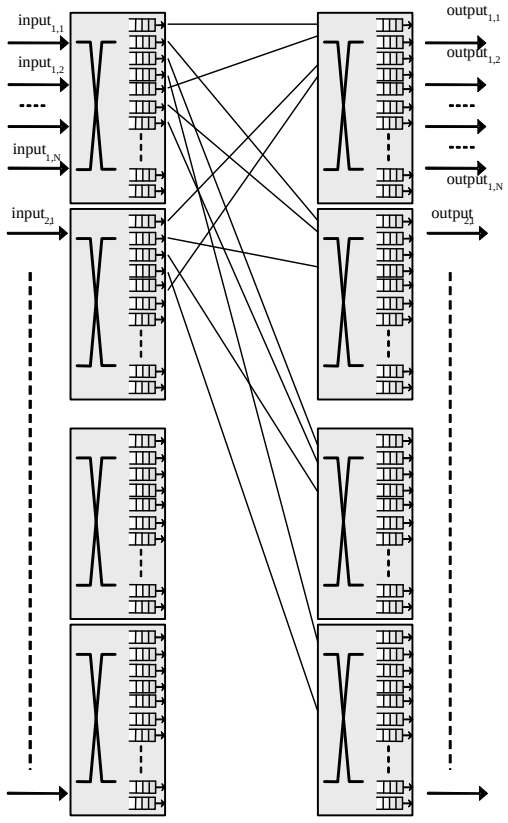

Figure 4: The symmetric topology for the backbone network.

As described earlier the interconnection network is of all-to-all type. What distinguishes our network is that reconfigurability is attained by using MOEMS devices. This gives the shuffle exchange network the ability to adapt to the traffic conditions and application characteristics. Reconfigurability by use of MOEMS devices can significantly improve performance as the network is given the possibility to adapt for the asymmetry in the application. By applying a MOEMS switch to the network backbone as showed in Figure 5, increased throughput capacity can be obtained as further explained in Section 4. If $N = 16$, the router size is 64 input ports times 64 output ports. Four 32×32 two dimensional (planar) MOEMS switches can be used to form a 64×64 non-blocking optical switch and for larger port counts (outside the scope of this paper) three dimensional MOEMS switches can be used.

Two different traffic patterns (applications) are modeled and applied to the different system solutions which allow us to extract some conclusions about the designs. If the overall traffic tendency is evenly

distributed, i.e. strictly symmetric, 25% of the *overall* traffic on one single input switch board will go straight forward to its neighboring (with the same index number) output switch board. This application is denoted as the symmetric application. Similarly, if the overall traffic from one input switch board where $N - M + 1$ network links are routed to the corresponding output switch board with the same index number *M*, and the remaining traffic is equally distributed to the rest of the outgoing switch boards. This is an extremely asymmetric application, hereafter denoted as the asymmetric application. In this case where $M = 4$ and $N = 16$ we will in overall have 13 network links going to the output switch board with the same index number *M*, as clarified in Figure 5, when the system has maximum traffic load.

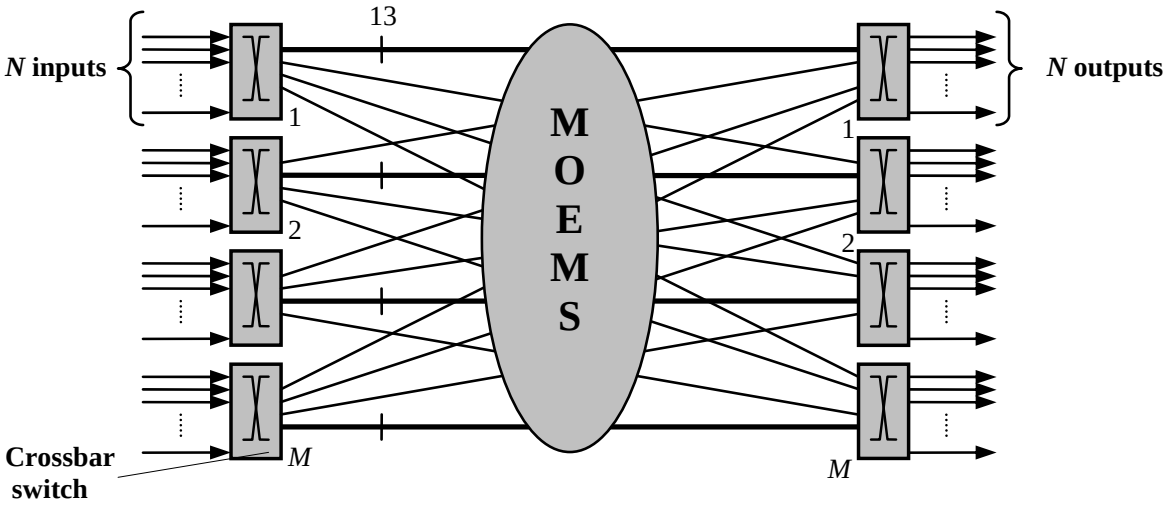

Figure 5: The reconfigured shuffle network, now with asymmetric properties.

In reality, the asymmetric cases can represent, e.g., networks in which the proposed system is a network router where a video on demand server sends mostly to some specific parts of the network for the moment. The reconfigurable network can adapt to the applications described above by reconfiguring the optical crossbar switch to fit the current traffic conditions. This inherently implies that for the symmetric application the MOEMS is configured according to the basic all-to-all setup since the symmetric application represents a traditional network type that works sufficiently well for most applications. If the traffic characteristics are according to the asymmetric application the MOEMS will be configured so that 13 wires from an input switch board are routed to its neighboring output switchboard (with the same index number) and one network link to each of the remaining three output switch boards. Similarly, the MOEMS is configured accordingly throughout the backplane.

4. A Case Study Analysis

The application models chosen for the analysis consist of periodic traffic with a period T_{period}, and where the traffic intensity is increased by increasing the number of logical channels traversing the system. All physical input wires are subjected with the same number of logical channels and as the number of channels per wire increases the traffic intensity also increases proportionally. Already established logical channels are preserved unchanged as more channels are added, i.e. when simulating higher traffic intensity.

The simulator works on the packet transmission level, where buffer sizes are regarded infinite. The packet size is constant and the transmission time of one packet is defined as one time unit. At the start of the simulation each logical channel is assigned an offset delay before the first period start. Each channel's offset value is randomly generated, with an even distribution, within the range between zero and T_{period}.

The two applications described previously are targeted to the system. *Throughput*, *average packet delay* and *average delay per input buffer* are measured and analyzed (see definitions below). These parameters are not measured from the simulation start since the traffic congestion has not settled. Instead they are measured from simulation time unit 400 up to 1000. The throughput is defined as total number of packets on average that has exited the router per time unit and physical wire. The period $T_{period} = 20$, while the amount of generated packets per period is set to $C = 1$. Inherently, a maximum of 20 logical channels are possible on one single physical input wire. The average packet delay is defined as the average delay of all packets delivered to any output. Finally, the average delay per input buffer is defined as the average delay of all packets from *one* specific input buffer that has exited on *any* output port.

In Figure 6 it can be clearly noted that when the application is asymmetric the backplane network is congested at around 6 logical channels per physical link where each physical wire can hold a maximum of 20 channels. I.e. the backplane is congested at around 30% of the theoretical maximum link capacity. At this point the throughput clearly increases with a slower rate. In fact, it never reaches above 45% throughput.

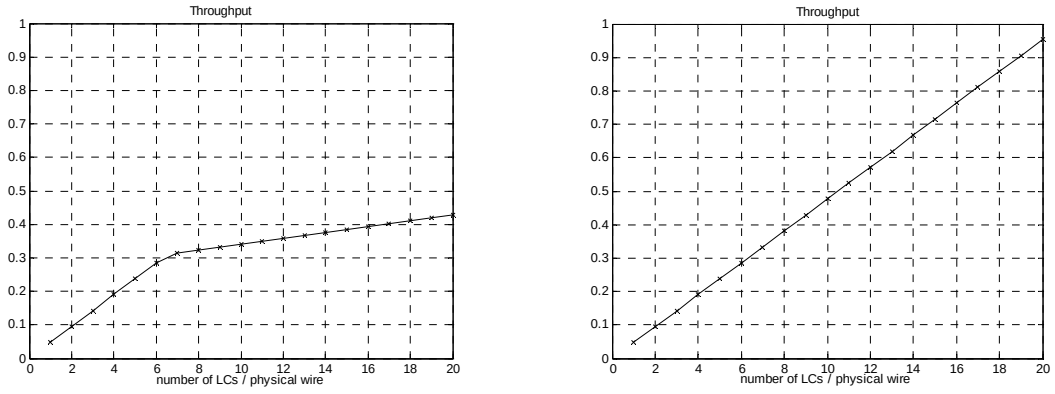

Figure 6: Throughput for the symmetric network (left) and for asymmetric network (right).

The throughput is normalized where 1 (or 100%) represents maximum load. Further, when the network is reconfigured to fit the asymmetry of the application the throughput increases linearly all the way to nearly 100% as shown to the right in Figure 6. Hence, a dramatic performance gain is achieved after 30% network load. At maximum load more than 50% performance gain can be attained when our reconfigurable shuffle network is used.

If we look at the average packet delay when the symmetric network is used it can be noted that the delay escalates heavily at 30% traffic intensity (6 logical channels per physical wire), see Figure 7. That indicates that some buffers are overloaded and the system is saturated. When the network is of asymmetric type according to the current application model the average packet delay increases fairly linear, and never grows as rapidly as in the symmetric case.

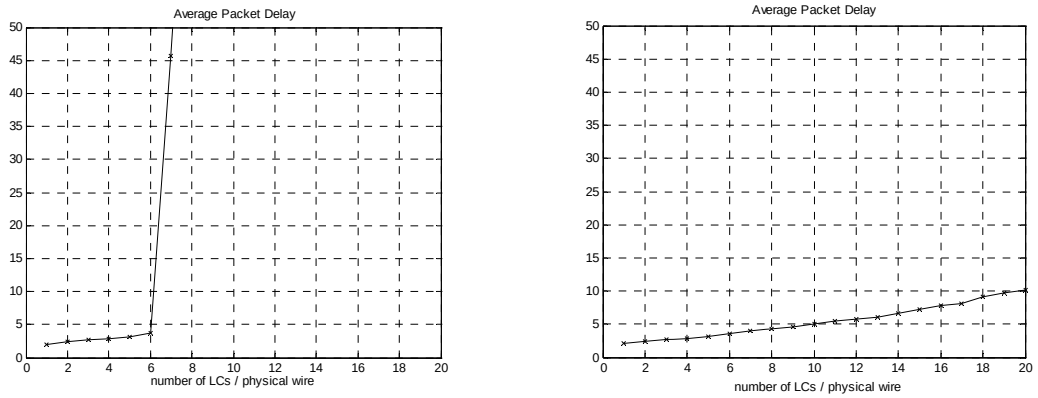

Figure 7: Packet delay for the symmetric network (left) and the asymmetric network (right).

It is also interesting to see what input buffers stalls (i.e. heavily congests) the system first, since a drastic increase in occupied slots in an input buffer is an indicator of what bandwidth bottlenecks the network has. In other words, if a particular buffer increases radically in number of packets held, the corresponding network link is congested itself and the board-to-board bandwidth is insufficient. In Figure 8 it can be seen that some of the buffers are congested earlier then others and these buffers are the ones that increases the average packet delay as well as decreases the throughput. For the symmetric network, again, it can be seen that at around 30% traffic load some of the buffers are saturated. Actually, there are 16 links (4 × 4) that are saturated. Those are the ones that are not sufficient to meet the overall board-to-board bandwidth that the pipelined traffic requires. In the case of using the asymmetric network the buffers are more evenly filled and no buffer becomes so populated that it can not recover from the situation. The network is more evenly congested throughout its resources. Therefore, it can be concluded that the resources for the asymmetric network are more properly used, for the given application, than in the symmetric case.

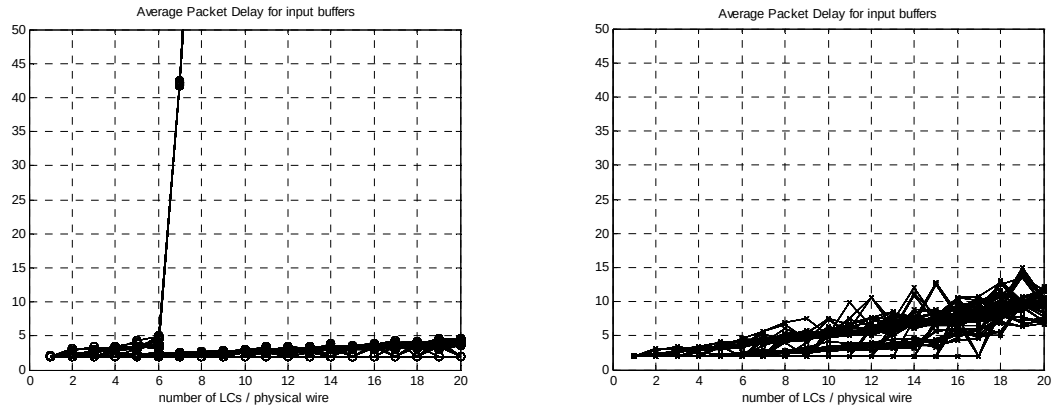

Figure 8: Average buffer delay for the symmetric (left) and the asymmetric network (right).

5. Conclusions

An optoelectronic switch and router system has been described. The system uses MOEMS devices to achieve reconfigurability in the backplane network which gives the opportunity to adapt the shuffle exchange network to asymmetric application characteristics. Models of the applications have been created and simulations have been carried out. From the analysis of these results it is showed that significant performance increase can be obtained if the backbone network is reconfigurable. This conclusion is made by measuring throughput, average packet delay and investigating what network links are subjected to severe congestion. This shows the potential of MOEMS devices in reconfigurable networking for the next generation high capacity routers.

6. References and Links

1. "*In-Stat MDR*", http://www.instat.com/press.asp?ID=949&sku=IN0401416EA, (current May 2004).
2. "*The Chalmers Centre for High Speed Technology*", http://www.nt.chalmers.se/chach/ (current February 2004).
3. *"Reconfigurable System Architecture"*, http://www.hh.se/ide/forskning/compsyseng/rsa.htm (current February 2004).
4. A. Pattavina, M. Martinelli, G. Maier, and P. Boffi, "Techniques and technologies towards all-optical switching," *SPIE Optical Networks Magazine*, vol. 1, no. 2, 2000, pp. 75-93.
5. Magnus Jonsson, "Optical interconnection technology in switches, routers and optical cross connects," *SPIE Optical Networks Magazine*, vol. 4, no. 4, 2003, pp. 20-34.
6. *"SafSec - Safety and Security"*, http://www.safsec.com/safsec_files/resources/asaac_~1.doc, (current May 2004).
7. X. Ma and G. Kuo, "Optical switching technology comparison: optical MEMS vs. other technologies," *IEEE Communications Magazine*, vol. 41, no. 11, 2003, pp. 16-23.

8. S. Agelis, S. Jacobsson, M. Jonsson, A. Alping, and P. Ligander, “Modular interconnection system for optical PCB and backplane communication,” *Proc. Workshop on Massively Parallel Processing (WMPP'2002) in conjunction with International Parallel and Distributed Processing Symposium (IPDPS'02)*, Fort Lauderdale, FL, USA, 2002, pp. 245-250.
9. S. Agelis and M. Jonsson, “System-level runtime reconfigurability – optical interconnection networks for switching applications,” *Proc. Engineering of Reconfigurable Systems and Algorithms*, Las Vegas, NV, USA, 2004.
10. L. Y. Lin, E. L. Goldstein, and R. W. Tkach, “Free-space micromachined optical switches for optical networking,” *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 5, no. 1, 1999, pp. 4-9.
11. “*Lucent Technologies*”, http://www.lucent.com/, (current February 2004).
12. P. Eriksen, K. Gustafsson, M. Niburg, G. Palmskog, M. Robertsson, and K. Åkermark, “The Apollo demonstrator – new low-cost technologies for optical interconnects,” *Ericsson Review*, vol. 72, no. 2, 1995.
13. M. Robertsson, K. Engberg, P. Eriksen, H. Hesselbom, M. Niburg, and G. Palmskog, “Optical interconnects in packaging for telecom applications,” *Proc. of the 10th European Microelectronics Conference*, 1995, pp. 580-591.
14. M. A. Shahid and W. R. Holland, ”Flexible optical backplane interconnections,” *Proc. 3rd International Conference on Massively Parallel Processing using Optical Interconnections (MPPOI’96)*, Maui, HI, USA, 1996, pp. 178-185.
15. *“Xanoptics Inc.”*, http://www.xanoptics.com/ (current February 2004).
16. F. Mederer, I. Ecker, R. Michalzik, G. Steinle, H. Riechert, B. Lunitz, J. Moisel, and D. Wiedenmann, “VCSEL transmitters for 10-Gigabit Ethernet: 1.3 µm wavelength VCSELs for metropolitan area networks and TO-packaged 850 nm wavelength VCSELs for data transmission over multimode fibers and optical backplane waveguides,” *Proc. 52nd Electronic Components and Technology Conference*, 2002, pp. 5 -11.
17. “*Zarlink Semiconductor Inc.*”, http://www.zarlink.com/, (current February 2004).
18. P. B. Chu, S. Lee and, S. Park, “MEMS: the path to large optical crossconnects,” *IEEE Communications Magazine, IEEE*, vol. 40, no. 3, 2002, pp. 80 – 87.
19. D. T. Nielson et al., “Fully provisioned 112 × 112 micro-mechanical optical crossconnect with 35.8 Tb/s demonstrated capacity,” *Optical Fiber Communication Conference (OFC’2000)*, vol. 4, 2000, pp. 202 – 204.
20. G. Shen, T. H. Cheng, S. K. Bose, C. Lu and T. Y. Chai, “Architectural design for multistage 2-D MEMS optical switches,” *Journal of Lightwave Technology*, vol. 20, no. 2, 2002, pp. 178 – 187.
21. V. A. Aksyuk et al., “238 × 238 micromachined optical crossconnect with 2dB maximum loss,” *Optical Fiber Communication Conference* (*OFC’ 2002)*, 2002, pp. FB9-1 - FB9-3.