DTU Library

Experimental Demonstration of Multidimensional Switching Nodes for All-Optical Data Centre Networks

Kamchevska, Valerija; Medhin, Ashenafi Kiros; Da Ros, Francesco; Ye, Feihong; Asif, Rameez; Fagertun, Anna Manolova; Ruepp, Sarah Renée; Berger, Michael Stübert; Dittmann, Lars; Morioka, Toshio

Total number of authors:
12

Publication date:
2015

Document Version
Peer reviewed version

Experimental Demonstration of Multidimensional Switching Nodes for All-Optical Data Centre Networks

Valerija Kamchevska, Ashenafi K. Medhin, Francesco Da Ros, Feihong Ye, Rameez Asif, Anna M. Fagertun, Sarah Ruepp, Michael Berger, Lars Dittmann, Toshio Morioka, Leif K. Oxenløwe and Michael Galili

DTU Fotonik, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, (vaka@fotonik.dtu.dk)

Abstract *We experimentally demonstrate network nodes that enable SDM/WDM/TDM switching. 1 Tbit/s/core error-free performance is achieved for connections with different granularities being switched between three network nodes interconnected with 7-core multicore fibres.*

Introduction

To enable future-proof data centre network (DCN) architectures that support scaling to thousands of Pbit/s capacity while allowing for low power, low cost and low latency, fundamentally new data plane technologies need to be developed, coupled with a new framework for control and service orchestration. The key challenge of traditional fat-tree architectures[1] is that scaling results in a many-tiered network due to the limited availability of very high radix switches. Several proposed schemes[2-4] have considered adopting optical switching in data centres for this reason, as high-port-count optical switches enable flat architecture with reduced end-to-end latency and additionally favour energy-efficiency[5]. Alternate approaches[6,7] have also revolved around the combination of different multiplexing technologies, such as time division multiplexing (TDM), wavelength division multiplexing (WDM) and space division multiplexing (SDM) allowing for better bandwidth utilization and enhanced scalability.

In order to address the aforementioned challenges of current DCNs, we propose hierarchically built all-optical network nodes (NNs) with multidimensional (SDM/WDM/TDM) switching capabilities, allowing for seamless scaling to enormous switching capacity, while at the same time leveraging the advantages of optical switching technologies such as reduced latency and power consumption. The switching nodes demonstrated here are composed of low loss free-space 3D beam-steering switches, wavelength selective switches and fast electro-optic switches, and interconnected with multi-core fibres (MCFs) to facilitate the link capacity scaling. We experimentally demonstrate communication between three network nodes connected with two 7-core MCFs carrying 1 Tbit/s/core and achieve error-free performance for connections with different granularities being switched in three different dimensions (space, wavelength and time). The demonstration relies on 40 Gbit/s On-Off Keying (OOK) modulated channels enabling simple transmitter design and low-latency direct detection. All connections may be orchestrated and reconfigured in a dynamic fashion by a centralized software defined networking (SDN) controller.

Data centre network architecture

As shown in Fig. 1, NNs are interconnected with MCFs and a single NN serves a number of racks, each composed of servers and Top-of-Rack (TOR) switches. Four different levels exist in the node structure. At the lowest level, switching of whole MCFs is performed; enabling traffic from full MCFs to be dropped at the NNs. On the next level, individual core switching is performed, which allows for switching between cores in the MCFs, as well as add/drop of fibre cores. The top two levels perform wavelength and sub-wavelength TDM-based switching, thus allowing for efficient bandwidth utilization.

The multi-level node structure allows establishing connections with capacity ranging from sub-wavelength to full MCF capacity, enabling any node to drop or add connections at different levels of granularity. Combined with SDN, this adds flexibility in the resource allocation, ultimately leading to reduced cost as well as adaptability to real-time demands, a feature highly desired for server-server and server-storage communication.

The fundamental principle of the proposed node structure is to pass connections to the higher level only if they require switching at a finer granularity, allowing for bypass traffic to be largely unaffected by the presence of the node. This property of the NN translates directly to an inherently large amount of bypass traffic at any node therefore we have initially considered a ring topology as a suitable solution for small DCNs, where a single ring of all-optical NNs will be sufficient for connecting the TOR switches. Shifting towards larger DCNs, different topologies for interconnecting the all-optical NNs may be preferred and remain to be investigated.

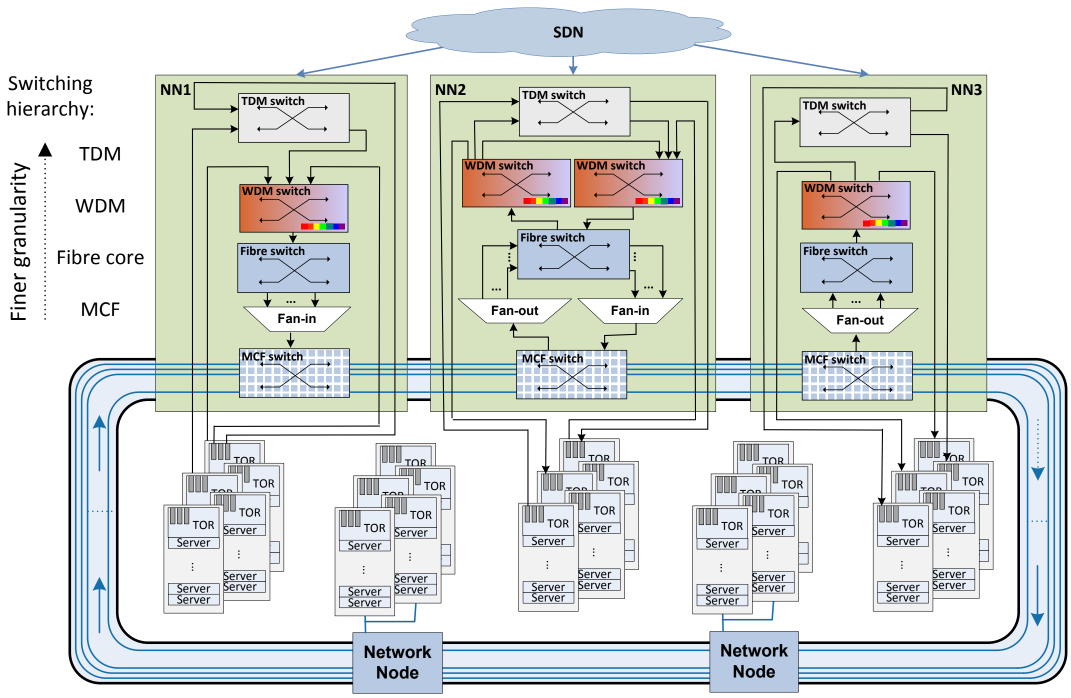

Fig. 1: Proposed DCN architecture composed of network nodes performing all-optical multidimensional switching

Experimental setup and results

For the experimental demonstration as shown in Fig. 2, three nodes are connected with two spans of 2-km 7-core MCFs. NN1 only adds traffic, while NN3 only drops incoming traffic. NN2 is dynamically reconfigured to demonstrate the full functionality of the proposed network node, namely add, drop and bypass traffic at all the different levels. WDM channels on a 100 GHz grid in the C-band carrying 40 Gbit/s OOK modulated data are used as WDM connections. Two alternating bursts of 190 ns with 10 ns gap carried on a single channel and generated using FPGA-based control signals are used as TDM connections. A 5-MHz clock is distributed for TDM synchronization on a separate wavelength. All the data channels and the clock are loaded in all the cores of the first MCF using a free-space fan-in device. At the MCF output, the cores are spatially de-multiplexed using a free space fan-out device and then connected to a 48x48 beam-steering switch used as a combined MCF and fibre switch. The switch is segmented into two 24x24 sections serving both NN2 and NN3. A 1x4 wavelength selective switch (WSS) is used as a WDM switch and a $LiNbO_3$ based 2x2 electro-optic switch is used as a TDM switch.

Two types of scenarios are investigated. In the case of 'full bypass', all connections from NN1 bypass NN2 at the lowest level and are fully received at NN3. In the case of 'add/drop/bypass', connections from NN1 target both NN2 and NN3. NN2 drops the relevant traffic and reuses the vacated resources to establish WDM and TDM connections to NN3. The burst added at NN2 is generated by an FPGA control signal based on the propagated clock, thus providing TDM synchronization. All the cores in the second MCF are then loaded with data bypassed and added by NN2 using a spliced fan-in device. At NN3, the cores are connected to the second segment of the fibre switch, while a WDM and TDM switch are used to drop and receive the individual channels. Three data channels (one carrying TDM slotted data) are initially investigated for characterizing the system performance in the different scenarios, while 25 data channels are used to demonstrate seamless scaling to 1 Tbit/s/core and 7 Tbit/s total throughput per MCF.

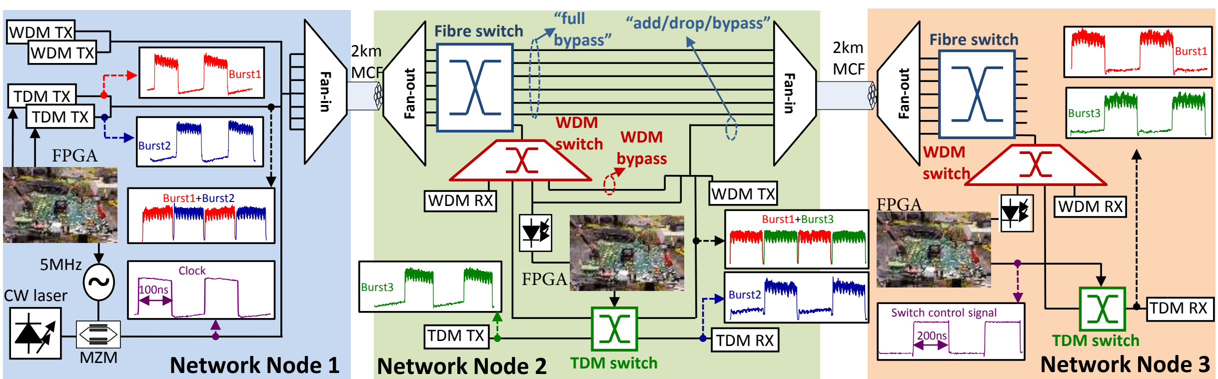

Fig. 2: Experimental setup for communication between three network nodes implementing multidimensional switching

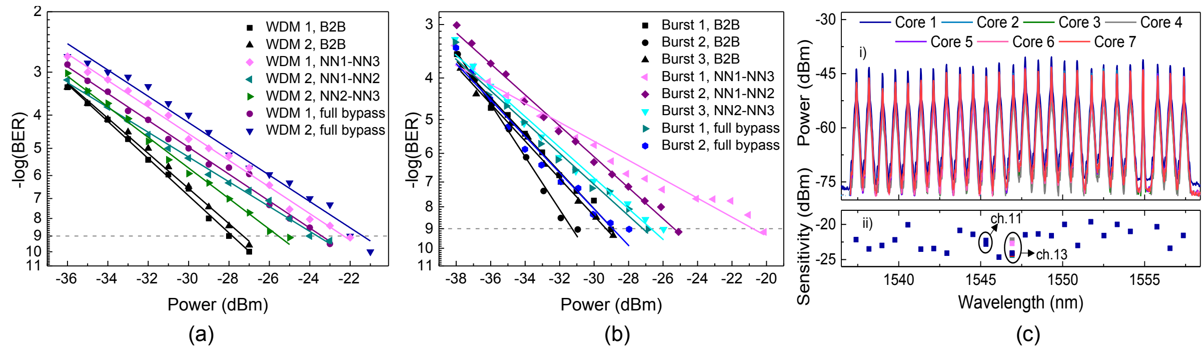

Fig. 3: (a) BER performance of the WDM connections (b) BER performance of the TDM connections (c) 1Tb/s/core system performance: i) Spectra of all cores at the output of the second MCF and ii) Receiver sensitivities of all channels measured on a single core (core 1) and receiver sensitivities in all the cores measured on a single channel (ch.13).

Fig. 3(a) shows the BER results of the WDM connections in the 'full bypass' and 'add/drop/bypass' case. Error-free performance is achieved for all connections. For connections between adjacent nodes, there is a 2 to 3-dB penalty, making it reasonable to consider this type of nodes for small data centres. In addition, the experimental results confirm the preference of bypassing at the lowest level possible for connections between nodes that are not adjacent, although only 1-dB extra penalty is observed when bypassing at WDM compared to SDM level. This could be found useful for time critical connections that in the lack of resources, instead of being delayed may be assigned a path with slightly higher penalty.

Fig. 3(b) shows the BER performance of the TDM connections in the 'full bypass' and 'add/drop/bypass' scenario. All connections have error-free performance. Due to the limited suppression ratio of the TDM switches and the generation of the bursts, there is a 2 to 3-dB penalty as a result of crosstalk whenever bursts are multiplexed. However, as the aim of the node is to bypass at the lowest level possible, the overall penalty experienced for a TDM connection can easily be reduced with a proper path allocation mechanism in which paths with minimum number of TDM switches through which the bursts pass, will be favoured.

Loading the MCFs with 25 WDM channels, yielding 1 Tbit/s/core and a total of 7 Tbit/s is also error-free. Fig. 3(c) shows the spectra and the receiver sensitivities (at BER=1E-9) of all 25 channels in the 'full bypass' scenario. The B2B receiver sensitivity measured on two channels (ch.13 and ch.19) is -27 dBm, while the B2B receiver sensitivity of burst 1 and burst 2 on the TDM wavelength (ch.11) is -31 dBm and -33 dBm, respectively. After transmission, the difference in the sensitivity among the different channels received is within 4.9 dB. Moreover, to confirm that similar performance is expected in all the remaining cores, the receiver sensitivity is measured on a single channel across all cores. Ch.13 has receiver sensitivity of -24 dBm in five cores, while in two cores the sensitivity is -22.2 dBm and -22.7 dBm, as a result of higher insertion loss and higher crosstalk from the coupling devices, respectively.

Conclusions

We have demonstrated highly scalable optical network nodes with multidimensional switching capabilities suitable for data centre networks. By combining optical switching with advanced multiplexing technologies and SDN, connections with different granularity can easily be obtained. We have experimentally demonstrated 1 Tbit/s/core transmission between three NNs interconnected with multicore fibres and achieved error-free performance for all the connections with different granularities being switched at SDM, WDM and TDM level.

Acknowledgements

We would like to thank OFS for fabricating and providing the spliced fan-in/fan-out MCF and Polatis for providing the 48x48 fibre switch. This work was supported by the ECFP7 grant no.619572, COSIGN and InnovationsFonden 0603-00514B, E-SPACE.

References

[1] M. Al-Fares et al., "A scalable, commodity data centre network architecture," ACM SIGCOMM, (2008).

[2] N. Farrington et al., "Helios: A hybrid electrical/optical switch architecture for modular data centers," ACM SIGCOMM, (2010).

[3] G. Wang et al., "c-Through: part-time optics in data centers," ACM SIGCOMM, (2010).

[4] Z. Cao et al., "Experimental demonstration of dynamic flexible bandwidth optical data center network with all-to-all interconnectivity," ECOC, PD.1.1, (2014)

[5] R. S. Tucker, "Scalability and energy consumption of optical and electronic packet switching," J. Lightw. Technol., Vol. **29**, no.16, p.2410, (2011).

[6] S. Yan et al., "First demonstration of all-optical programmable SDM/TDM intra data centre and WDM inter-DCN communication," ECOC, P.D.1.2, (2014).

[7] N. Fontaine et al., "Heterogeneous Space-Division Multiplexing and Joint Wavelength Switching Demonstration," OFC, Th5C.5, (2015).