First Demonstration of All-Optical Programmable SDM/TDM Intra Data Centre and WDM Inter-DCN Communication

S. Yan(1), E. Hugues-Salas(1), V. J. F. Rancaño(2), Y. Shu(1), G. Saridis(1), B. R. Rofoee(1), Y. Yan(1), A. Peters(1) S. Jain(2), T. C. May-Smith(2), P. Petropoulos(2), D. J. Richardson(2), G. Zervas(1), D. Simeonidou(1)

(1) High Performance Networks Group, University of Bristol, UK (Shuangyi.Yan@bristol.ac.uk)
(2) Optoelectronics Research Centre, University of Southampton, UK (vjfr1u10@orc.soton.ac.uk)

Abstract *We successfully demonstrate a flat-structured DCN powered by large-port-count fibre-switch-based OCS, PLZT-switch enabled TDM and MEFs supported SDM. The inter-DCN ToR-to-ToR direct optical connections are setup through metro/core networks using all-optical SDM/WDM converters.*

Introduction

Optical interconnection is the most promising technology to provide power-efficient, high-bandwidth connections in large scale data centre networks (DCN). Large-port-count fibre switches (LPFS) enable flat-structured low-latency DCNs[1,2]. In such DCNs, a large quantity of links need to be setup between the Top-of-Racks (ToRs) and the centralized optical switches to realize an optical circuit switching(OCS)network. One approach is to explore DWDM technologies. However, ubiquitous connections between ToRs require frequent wavelength reconfiguration and switching. Thus, a DWDM-based OCS would first have to demultiplex and switch individual wavelength channels. Thanks to recent advances in fibre technologies, space division multiplexing (SDM) is now possible, employing multicore fibres (MCF) or multi-element fibres (MEF) which provide as many as 19 parallel links in dimensions similar to those of a typical SMF[3,4]. Furthermore, VCSEL/PD arrays can potentially be matched to MCF/MEF, to provide increased capacity with small footprints and low power-consumption[5]. The characteristics of SDM make it suitable for DCN applications.

In this paper, we demonstrate for the first time an all-optical OCS-based DCN, combining the benefits of both SDM and TDM technologies for intra-DCN communications using MEFs. A beam-steering 192×192 LPFS interconnects *a*) ToRs via MEF links, *b*) optical functional elements, i.e., PLZT TDM switch, SDM/WDM converter and *c*) inter-DCN links. The LPFS can realize different connection matrices, e.g., a single hop OCS, to serve the long-lived data flows for intra-DCN communications. The TDM switches, incorporating OCS/TDM reconfigurable transceivers on ToRs, serve low-capacity, highly-connected data flows from 100 Mbit/s to 5 Gbit/s for intra-cluster communications. Bandwidth variable transmitters (BVTs) on ToRs provide inter-cluster links of up to 320 Gbit/s. In addition, by introducing the architecture-on-demand (AoD) concept[6] in DCN, network function programmability (NFP) is enabled to support variable traffic patterns as well as broadcasting and aggregation.

We also demonstrate cross-DCN ToR-to-ToR all-optical connections through the metro/core networks, using reversible SDM/WDM converters. Three SDM signals (up to 320 Gbit/s/channel) on the same-λ, originating from one or multiple ToRs, are converted optically to WDM signals to be transferred through the core networks. The direct optical ToR-to-ToR cross-DCN connections enable the remote distributed DCNs to appear as one big data centre, which could enhance scalability and reduce latency and cost.

Intra- and Inter-DCN Communications

Fig. 1 shows the proposed solutions for both intra- and inter-DCN communications. Each DCN consists of clusters with tens/hundreds of racks networked together and each rack is filled up with tens of servers. Servers are interconnected to ToRs via 10GE optical links. In our design, ToRs play a pivotal role in both inter- and intra-DCN communications. Fig. 2 shows the design of the proposed FPGA-based ToR. The ToR, implemented using FPGA optoelectronics (HTG Xilinx V6 board), parsed the input traffic from servers and sent them out through different transmitters according to their destination. Programmable slotted-TDM/Ethernet over SDM signals are sent out through two SFP+ transceivers for intra-cluster communication. Another four transmitters feed the traffic to a BVT to provide a link of up to 320 Gbit/s for high-capacity inter-cluster and inter-DCN communications. All the transmitters

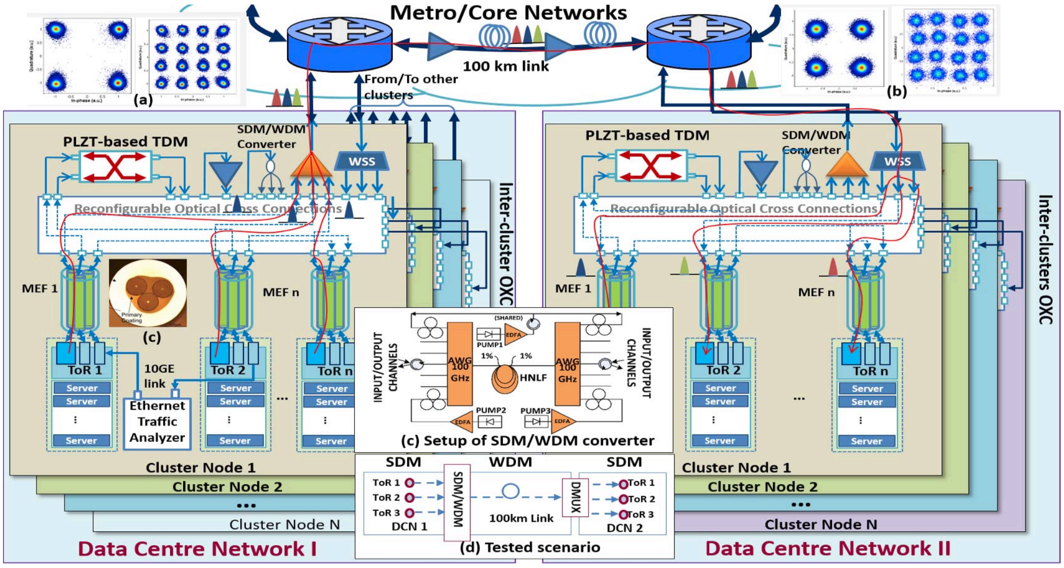

Fig. 1: Proposed solution for intra-DCN communication (SDM+TDM) and inter-DCN communication (SDM+WDM)

use fixed wavelength lasers to avoid expensive temperature control. Then all the transceivers on the ToRs are connected to a LPFS with MEFs. The clusters are networked by interconnecting the centralized LPFS in each cluster using SMFs or MEFs to form a mesh network. The inter-DCN communication is realized by transferring SDM signals through the metro/core networks with SDM/WDM all-optical converters.

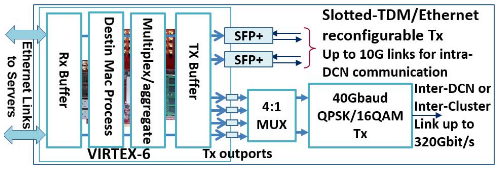
Ethernet Links to Servers
Rx Buffer
Destin Mac Process
Multiplex/aggregate
TX Buffer
VIRTEX-6
SFP+
SFP+
Slotted-TDM/Ethernet reconfigurable Tx
Up to 10G links for intra-DCN communication
4:1 MUX
40Gbaud QPSK/16QAM Tx
Inter-DCN or Inter-Cluster Link up to 320Gbit/s
Tx outports

Fig. 2: FPGA-based ToR provide slotted-TDM/Ethernet signals for intra-DCN and BVT up to 320 Gbit/s for inter-DCN

For the experimental setups, three 3-element MEFs of different lengths connect all the transmitters and receivers in two ToRs to a 192×192 fibre switch. The 4×4 PLZT TDM switch, SDM/WDM converter, WSS, EDFAs, couplers and splitters are all ready connected to the fibre switch for function programmability. A 100 km SMF link is setup for metro/core transmission.

a) DWDM-based Inter-DCN communications

Directed cross-DCN ToR-to-ToR connections are setup through the metro/core network using an all-optical SDM/WDM converter. The experimental setup of the SDM/WDM converter is shown in Fig. 1 (c). The converter is a fibre-based dual-pump four-wave-mixing device[7], and uses a shared-pump bi-directional configuration that allows conversion of two SDM channels at the same time. It provides polarisation-, rate- and modulation format-independent operation. Using this converter, contiguous 3-carrier superchannel signals are obtained from three same-λ SDM signals (either PM-QPSK or PM-16QAM signals were used in our experiments), which are subsequently launched into the metro/core networks. Then the superchannel signal is dropped at the edge node and sent to another DCN. By adopting optical demultiplexing, each carrier is sent to different ToRs. The tested inter-DCN scenario is shown in Fig. 1 (d). The inter-DCN link adopts either PM-16QAM or PM-QPSK signals at 40 Gbaud to trade off capacity against transmission distance. 16QAM/QPSK constellations of the CH1 signals are shown in Fig. 1 (a) for back-to-back and (b) after 100 km transmission.

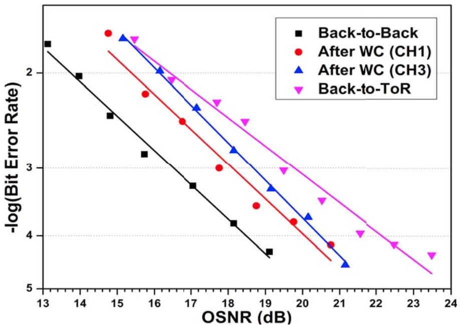

Fig. 3: OSNR vs. BER for 40 Gbaud PM-QPSK signals for inter-DCN communication

For the 40 Gbaud QPSK signal, we analyse the performance of the setup by measuring the BER to OSNR curve. The results are shown in Fig. 3. The SDM/WDM converter introduces an OSNR penalty of about 1.61 dB and 2.53 dB at 1e-3 for the two converted channels: CH1 and CH3. Another SDM signal without wavelength conversion is put in the CH2 wavelength slot. After 100 km transmission, the CH1 signals are transferred back to the ToR in another DCN. The penalty of the ToR-to-ToR connection is about 3.28 dB.

b) TDM/SDM Intra-DCN communications

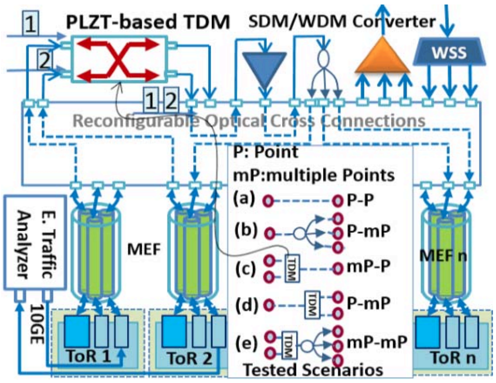

Fig. 4: Experimental Setup of intra-cluster network

For intra-cluster communication, OCS-based SDM (Ethernet) and TDM technologies are used to provide a range of connections of bandwidth and capacity services. Fig. 4 shows the demonstration setup. The single-hop SDM with an element capacity of 10 Gbit/s is realized by reconfiguring the interconnection of the LPFS to supports high throughput Ethernet transport, i.e., the dominant traffic in DCNs. A synchronized (4×4) PLZT TDM switch is connected to the LPFS to realize TDM connections with variable capacity from 100 Mbit/s to 5 Gbit/s. Optical components, such as couplers and splitters are connected to the LPFS, and can be programmed to support aggregation, broadcasting and other network functionalities, as in the examples shown in the inset of Fig. 4.

To test the performance of intra-cluster networks, an Ethernet Traffic Analyser is used to generate Ethernet traffic between servers and ToR. In scenario (a), the latency of the P-P transmission is measured to be about 70 μs, where the FPGA-based ToR contributes about 4.2 μs (5% of the link latency). Fig. 5 presents the results for both SDM-Ethernet network in scenarios (a, b) and TDM network in scenarios (c, d, e). The maximum Ethernet capacity/port is about 9.8 Gbps. In the OCS-based SDM network, the broadcast operation introduces a power penalty of about 1.3dB at 1e-9 due to the noise introduced by the EDFA used to compensate the loss of the splitter. In scenario (e), two time-slot data flows (2.5 Gbit/s) from two ToRs are aggregated to 5 Gbit/s and further broadcast to three other ToRs. The PLZT TDM switch introduces a power penalty of about 2.1 dB at 1e-9 due to its cross talk (20 dB).The received TDM capacity for different received optical power after broadcasting is also shown in Fig. 5.

SDM signals are used for inter-cluster communication. The BVTs on ToRs generate 10 Gbit/s OOK, 40 Gbaud PM-QPSK (160 Gbit/s) or PM-16QAM (320 Gbit/s) signals to setup bandwidth-variable links between clusters.

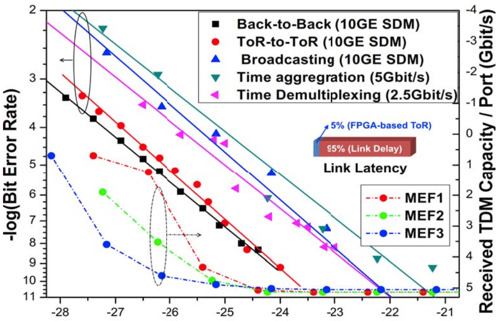

Fig. 5: Experimental results of intra-cluster network

Conclusions

An all-optical multi-dimensional and programmable solution for both intra- and inter-DCN communications is proposed and successfully demonstrated . The MEF-based SDM technology provides low-loss and easy to handle links between ToRs and cluster LPFS. SDM and TDM technologies are used in an OCS manner to realize intra-DCN communications. The inter-DCN communication is realized by converting SDM signals to WDM signals using a multi-λ SDM/WDM converter, and then transferring them to another DCN. The NFP enables the DCN to realize different network functions.

Acknowledgements

This work is supported by the EPSRC grant EP/I01196X The Photonics Hyperhighway, and the ECFP7 grant no. 619572, COSIGN.

References

[1] N. Farrington et al., "Helios: a hybrid electrical optical switch architecture for modular data centers," SIGCOMM'10, New Delhi(2010).

[2] G. Wang et al., "Programming your network at run-time for big data applications," Proc. HotSDN'12,Helsinki, Finland.

[3] S. Matsuo et al., "Recent progress on multi-core fiber and few-mode fiber," Proc. OFC, OM3I.3, Anaheim(2013)

[4] S. Jain et al., "Multi-Element Fiber Technology for Space-Division Multiplexing Applications," Opt. Express, vol. **22**, no. 4, p. 3787(2014)

[5] F. Doany, "Power-Efficient, High-Bandwidth Optical Interconnects for High Performance Computing," IBM, Hot Interconnects, Aug. 2012.

[6] G. Zervas et al.,"Network Function Programmability and Software-Defined Synthetic Optical Networks for Data Centres and Future Internet," Proc. PS, PM4C-3, San Diego(2014).

[7] V. J. F. Rancano et al., "100GHz grid-aligned reconfigurable polarization insensitive black-box wavelength converter," Proc. OFC, JTh2A.19, Anaheim(2013).