Alexoudi, T., Mitsolidou, C., Pitris, S., Han, J. H., Farhadi Beldachi, A., Manihatty Bojan, N., ... Kanellos, G. (2018). WDM routing for edge data centers and disaggregated computing. In *Photonic Networks and Devices 2018: Zurich Switzerland 2–5 July 2018* [NeM3F. 2] Optical Society of America (OSA). https://doi.org/10.1364/NETWORKS.2018.NeM3F.2

WDM routing for edge data centers and disaggregated computing

Theoni Alexoudi[2,3], Charoula Mitsolidou[2,3], Stelios Pitris[2,3], Jong Hun Han[1], Arash Farhadi Beldachi[1], N Manihatty-Bojan[1], Y. Ou[1], E. Hugues-Salas[1], R. Broeke[4], R. Nejabati[1] and D. Simeonidou[1], and George T. Kanellos[1]

(1) High Performance Networks group, School of & Computer Science, Electrical and Electronic Engineering, and Engineering Maths, University of Bristol, BS8 1UB, UK,

(2) Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece,

(3) Information Technologies Institute, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece

(4) Bright Photonics B.V., 5612 AX Eindhoven, The Netherlands

Author e-mail address: gt.kanellos@bristol.ac.uk

Abstract: In the present communication we review recent advances in o-band silicon photonics transceivers and wavelength routers and demonstrate their potential application in board-level and rack level interconnection for edge and disaggregated data centers.

OCIS codes: (060.4265) Networks, wavelength routing; (250.5300) Photonic integrated circuits

1. Introduction

Cisco predicts that Annual global cloud IP traffic will reach 8.6 ZB while the data created globally by Internet of Everything (IoE) devices will reach 507.5 ZB per year by 2019. The two predictions alone are enough to underline two transformative trends in data centers: 1) the rise of hyperscale data centers, expected to account for almost half of the total installed data center (DC) servers worldwide by 2020 [1], and 2) the emergence of edge data centers as distributed compute nodes placed near the network access in the next generation 5G network deployments, to serve applications with ultra-low latency requirements. Optical interconnects have been a major drive in the data center networks (DCN) evolution, have now reached the form of 100Gb PSM4 and CWDM4 transceiver to penetrate all levels of hierarchical DCN interconnection [2], while focusing already on scaling to 400Gbps networks [3].

However, scaling up the DCN and designing it to perform with minimal latency will take more than simply scale the transceivers and switches capacity, already a major challenge, and will require radically new architectural approaches that effectively take advantage of the progress in PIC technologies. Google recently indicated that near future hyperscale data centers will soon require a 5Pb/s network capacity, which is roughly an order of magnitude greater than the total bisection of the internet [4]. To address such network scaling, Facebook has already turned its interest towards the highest-performing bandwidth and reach technologies deployed for telecom — particularly switching, routing, and transport DWDM technologies [5]. Facebook and Acacia are not alone, with other key industrial representatives from DC operators such as Google [7], to equipment vendors such as Rockley [6] and other major academic institutions [8][9], to propose their versions of DWDM technologies for scaling data centers. Their common ground is on full scale DWDM networks, as this inherent ability of light for channel parallelism is the only practical approach for efficiently scaling the future DC networks.

Following this rationale, in the present communication we present recent advancements in key-enabling WDM routing components, such as a Silicon Photonic integrated Arrayed Waveguide Grating Routers – AWGR operating in the O-band and discuss new WDM concepts for future data centers networks, such as 1) the deployment of transceivers with broadcasting functionality for processor-to-processor or board-to-board interconnection and 2) the use of broadcast-and-select rack-level topologies towards flatter, faster and more energy efficient interconnections.

2. 8x8 Silicon Photonic AWGR

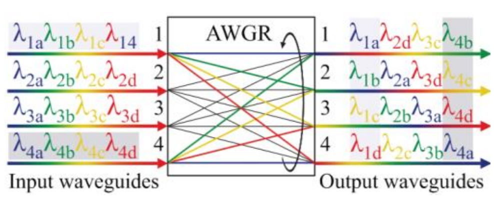

Fig. 1. Cyclic frequency operation in a N×N AWGR (N=4)

AWGR are essential components in wavelength routing architectures offering all-passive non-blocking, all-to-all interconnection schemes that support N^2 interconnections using N wavelengths when employed as N×N routers.

However, their deployment in data centers suggests AWGRs are designed for operation in the O-band in order to be compatible with the QSFP transceivers. To this end, we review on a novel silicon integrated AWGR device [10], where the 8 channels are aligned to match the CLR4/CWDM optical interface specifications but with a denser channel spacing. To meet these requirements, the AWGR was designed with 10nm-channel spacing [1261 nm, 1271 nm…1331 nm], 400 GHz 3-dB bandwidth and an Free Spectral Range (FSR) of 80 nm.

3. Broadcast-friendly optical transceivers for AWGR-based routing

We propose a new WDM routing scheme that proves advantageous in cases where broadcasting is evident, such as processor to processor communication in multi-socket server boards. The scheme employs WDM parallel transceivers with multiple Ring Modulators (RMs) for the unicast and multicast communication between the multiple nodes and a broadband Mach-Zehnder Modulator (MZM) in the common WDM multiplexed transmitter nodes path to enable broadcast operation in an energy-efficient manner. In this architecture, direct wavelength cycling routing is achieved with an AWGR platform, allowing for simultaneous direct all-to-all interconnection between the multiple nodes through a collision- and buffer-less scheme. Experimental validation of single optical links and 1-to-2 broadcasting operation has been demonstrated with 10Gb/s On/Off Keyed (OOK) optical data signals, exploiting silicon photonics building blocks. The results reveal that optical links have the potential to outperform their electronics counterparts for p2p interconnection in terms of bandwidth and energy efficiency.

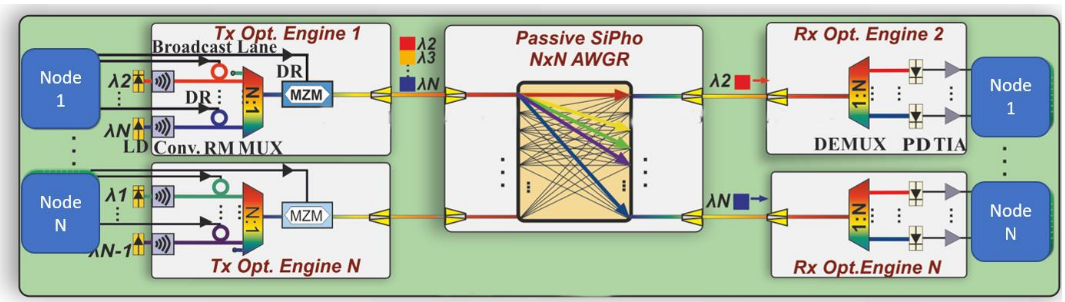

Fig. 2. Optical Interconnection architecture for unicast and broadcast communication for up to N nodes, comprising WDM transceivers (Txs-Rxs) and the AWGR routing platform

4. Rack-level WDM broadcast-and-filtering topologies

For the broadcast mode, we propose an architecture that can be implemented in a hybrid mode of optical and Ethernet that employs a broadcast-and-filter topology. In this case, the receiver module on each socket has a filter for look-up whether the broadcasting packets need to be forwarded into the operation system. The filtering process can be implemented with a parser and a CAM (Content Addressable Memory) or TCAM (Ternary Content Addressable memory) for look-up. The filtering processor consists of a parser, a match table incorporating the TCAM or CAM and a Finite State Machine (FSM) to control whether it forwards or drops the packets. The prototype hardware can be implemented on a FPGA platform feeding four SFP+ ports.

The broadcast-and-filtering topology can be combined with the WDM broadcast friendly concepts to provide a flatter board- or rack- level connectivity with improved latency characteristics and enhanced energy efficiency.

5. References

[1]."Cisco Global Cloud Index Projects Cloud Traffic to Quadruple by 2019", Cicso News, Oct. 2015 link
[2].“100 Gigabit Ethernet optical transceiver sales topped $1B in 2016”, LightCounting Lightw. News 2017,link
[3].http://www.ethernetalliance.org/subcommittees/roadmap-subcommittee/
[4].A. Vahdat,"A look inside Google's data center network",Open Networking Summit,Santa Clara, June 2015
[5].I. Lyubomirsky, B. Taylor, H.-J. W. Schmidtke, "An open approach for switching, routing, and transport," Telecom Infra Project, Facebook, Nov. 2016
[6]."Optoelectronic switch", WO 2016050868 A2 Patent, Rockley Photonics, Sep. 2015 link
[7]."Optoelectronic packet switch/routers", US 8792787 B1 Patent, Google Inc, June 2014 link
[8].A. Saleh, "Scaling-out Data Centers Using Photonics Technologies," Advanced Photonics for Communications, JM4B.5, San Diego, USA, July 2014
[9].Proietti, R.; Zheng Cao; Nitta, C.J.; Yuliang Li; Yoo, S.J.B., "A Scalable, Low-Latency, High-Throughput, Optical Interconnect Architecture Based on Arrayed Waveguide Grating Routers," Lightwave Technology, Journal of , vol.33, no.4, pp.911,920, Feb.15, 15 2015
[10] Stelios Pitris, George Dabos, Charoula Mitsolidou, Theoni Alexoudi, Peter De Heyn, Joris Van Campenhout, Ronald Broeke, George T. Kanellos, and Nikos Pleros, "Silicon photonic 8 × 8 cyclic Arrayed Waveguide Grating Router for O-band on-chip communication," Opt. Express 26, 6276-6284 (2018)