UNIVERSITY OF LEEDS

Proceedings Paper:
Hammadi, AA, El-Gorashi, TEH and Elmirghani, MH (2015) High Performance AWGR PONs in Data Centre Networks. In: International Conference on Transparent Optical Networks. 17th International Conference on Transparent Optical Networks (ICTON), 2015, 05-09 Jul 2015, Budapest, Hungary. IEEE . ISBN 978-1-4673-7880-2

https://doi.org/10.1109/ICTON.2015.7193567

High Performance AWGR PONs in Data Centre Networks

Ali Hammadi, Taisir E.H. El-Gorashi, and Jaafar M.H. Elmirghani
School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT, United Kingdom

ABSTRACT
The unprecedented advances in Passive Optical Networks (PONs) and their proven performance in access networks have encouraged researcher to investigate the use of passive optics to address the limitations of conventional data centre architectures. In this paper, we present a scalable, high capacity and energy efficient arrayed wave guide grating router (AWGR) PON based data centre architecture to facilitate inter and intra racks communication within data centres. The proposed architecture can be scaled up efficiently to hundreds of thousands of servers and has shown energy savings of 45% and 80% compared to the Fat-Tree and BCube architectures, respectively.

Keywords: Passive Optical Network (PON), Data Centre, Energy Efficiency, Arrayed Waveguide Grating Routers (AWGRs).

1. INTRODUCTION

Significant research efforts have been devoted over the last decade to the design of efficient data centre architectures to host the ever growing computing intensive and high bandwidth cloud applications. Major concerns, however, are still raised about the power consumption of data centres and its impact on global warming in the first place and on the electricity bill of data centres in the second place. In addition, conventional data centres networking architectures are not handling the steadily increasing number of servers and the exponentially growing traffic inside data centres gracefully, resulting in links oversubscription and inefficient load balancing [1]. These limitations have stimulated the search for new low cost, scalable, energy-efficient architectures to efficiently serve the increasing demands [2].

The advantages of Passive Optical Network (PON) technologies, proven in residential access networks, such as high capacity links to end users, cost efficiency, and low power consumption [3], have encouraged investigating the use of passive optics in the fabric interconnection for future data centres to address the limitation of conventional data centres. Previous research work has considered partial implementation of PONs in data centres [4]-[7].

In [8], [9] we proposed and compared five novel designs for PON deployment in future data centres to handle intra and inter rack communications. In this paper, we further investigate one of these designs. The investigated design is an Arrayed Waveguide Grating (AWG)-based PON interconnection. We optimize the fabric configuration and topology interconnection within the Arrayed Waveguide Grating Routers (AWGRs) for wavelength routing and assignment to facilitate the inter PON communication among servers within the PON cell. We also compare the power consumption of this to conventional data centre architectures.

The remainder of this paper is organized as follows: In Section 2, we present our proposed PON data centre architecture. In Section 3, we present a benchmark study to compare the power consumption of proposed PON architecture with the conventional Fat-tree and BCube data centre architectures. Finally we conclude the paper in Section 4.

2. THE PROPOSED PON DATA CENTRE ARCHITECTURE

Our objective is to design a high speed energy efficient fabric interconnection mainly relying on PON technologies to replace power hungry devices in conventional data centre architectures such as access and aggregation switches without sacrificing performance [8].

Figure 1 shows the connectivity created by deploying a typical PON architecture to provide interconnections in a data centre. The OLT switch consists, typically of 8 chassis, hosting up to 16 cards each. Each OLT card has the capacity to connect 8 ports, each of which provides a transmission rate up to 10 Gb/s. With a split ratio of 128, a single card port can connect 128 servers in a PON cell, and therefore one card can connect 1024 servers, and one chassis can provision connection to 16,384 servers. Using this architecture to connect Intel core 980x servers of 8 GB RAM memory and 147,600 MIPS processing capability, one OLT chassis can provision a total processing of 2,418,278 GIPS and memory of 131,072 GB. This large connectivity and capacity furnish a substantial infrastructure for cloud applications and services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

In the proposed PON design, shown in Figure 2(a), the PON Cell is equipped with two AWGRs to provision full interconnection between PON racks. All the servers within a group share a single wavelength for inter rack and OLT communication. Therefore the number of wavelengths used for inter rack communications is proportional to the number of racks in the PON cell. This design is similar to the cellular wireless network, as the same wavelengths can be reused in other PON cells connected to different OLT ports. Servers can be either connected to a tuneable ONU or equipped with a network interface card with an array of fixed tuned receivers and a tuneable laser for wavelength detection and selection, respectively. The connections between the racks and the OLT are established via a 1: N AWGR. The architecture can be scaled up to host hundreds of thousands of servers.

Inter-rack communication can be provisioned either via the OLT switch or directly through the AWGR where a wavelength is selected for transmission based on the location of the destination server. A server can reach servers in other racks within the PON Cell by tuning its transceiver to the proper wavelength that matches the AWGR wavelength routing map. Alternative routes facilitate multi-path routing and load balancing at high traffic loads, however; forwarding traffic through the OLT switch should be avoided if possible to reduce delay and power consumption resulting from O/E/O conversions, queuing, and processing.

For intra rack communication, one of three designs depicted in Figure 2(b) based on optical passive devices can be used. The terabit capacity passive polymer optical backplane in [10] for example can provide non-blocking full mesh connectivity with 10 Gb/s rates per waveguide, exhibiting a total capacity of 1 Tb/s. A passive optical star reflector [11] and Fibre Bragg Grating (FBG) [11] can support intra rack communication.

All communications are coordinated via the OLT. Servers with demands send a request control message (Report) containing the destination address and resources requirements to the OLT. If the OLT grants the request, it replies with a gate messages to the source and the destination ONUs informing the servers about the wavelength assignment both servers' ONUs need to tune to and the time slot. Initially, all ONUs transmitters and receivers are tuned to a designated wavelength connecting them with the OLT switch.

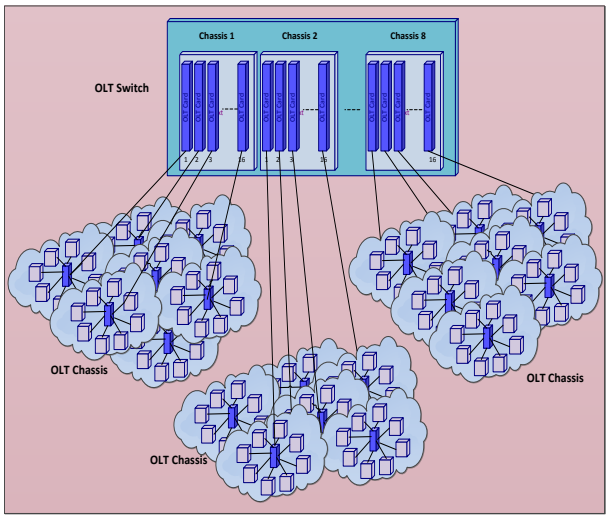

Figure. 1 An OLT switch with 8 chassis for a possible PON deployment in data centers

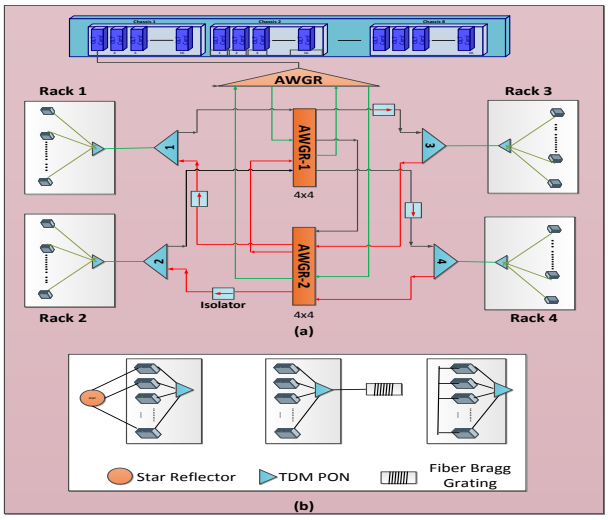

Figure. 2 (a) AWG-based PON cell with servers equipped with tuneable lasers, (b) Alternative technologies for Intra-rack communication

We developed a Mixed Integer Linear Programming (MILP) model to optimize the interconnection of the AWGRs for wavelength routing and assignment to facilitate the inter PON communication among all servers located in different racks within the PON cell.

The model ensures that only a single wavelength is assigned to establish a connection between two groups of servers. A wavelength continuity constraint is used to ensure that the wavelength going into a node (a node can be a server, PON coupler, or AWGR port) is the same wavelength leaving it for all nodes except the source and destination. Also we ensure the AWGR directionality property is satisfied by ensuring flows are only directed from inputs to outputs of the AWGRs devices.

We used the MILP model to optimize the AWGR interconnections of the PON architecture depicted in Figure 3(a). The PON cell consists of 64 servers arranged in 8 PON groups. The model results are shown in Figure 3; where 3(a) presents fabric interconnection design for a PON cell with tuneable lasers for 8 PON racks, 3(b) shows the obtained

MILP configuration for AWGRs interconnection for wavelength routing and assignment, and 3(c) shows the wavelength assignment table for inter rack and racks to OLT switch communications.

According to the wavelength routing table, if a server in PON-1 wants to communicate with a server in PON-4, a control message is sent to the OLT using wavelength 1 routed through AWGR-1 input port 1 to output port 5 that connects with the OLT switch. If the request is granted, the OLT replies with a control message to PON-1 and PON-4 using wavelengths 4 and 5 respectively. The destination server at PON-4 tunes its receiver to wavelength 2 to receive data from the source server in PON-1. Idle servers by default should be tuned to wavelengths connecting them with the OLT.

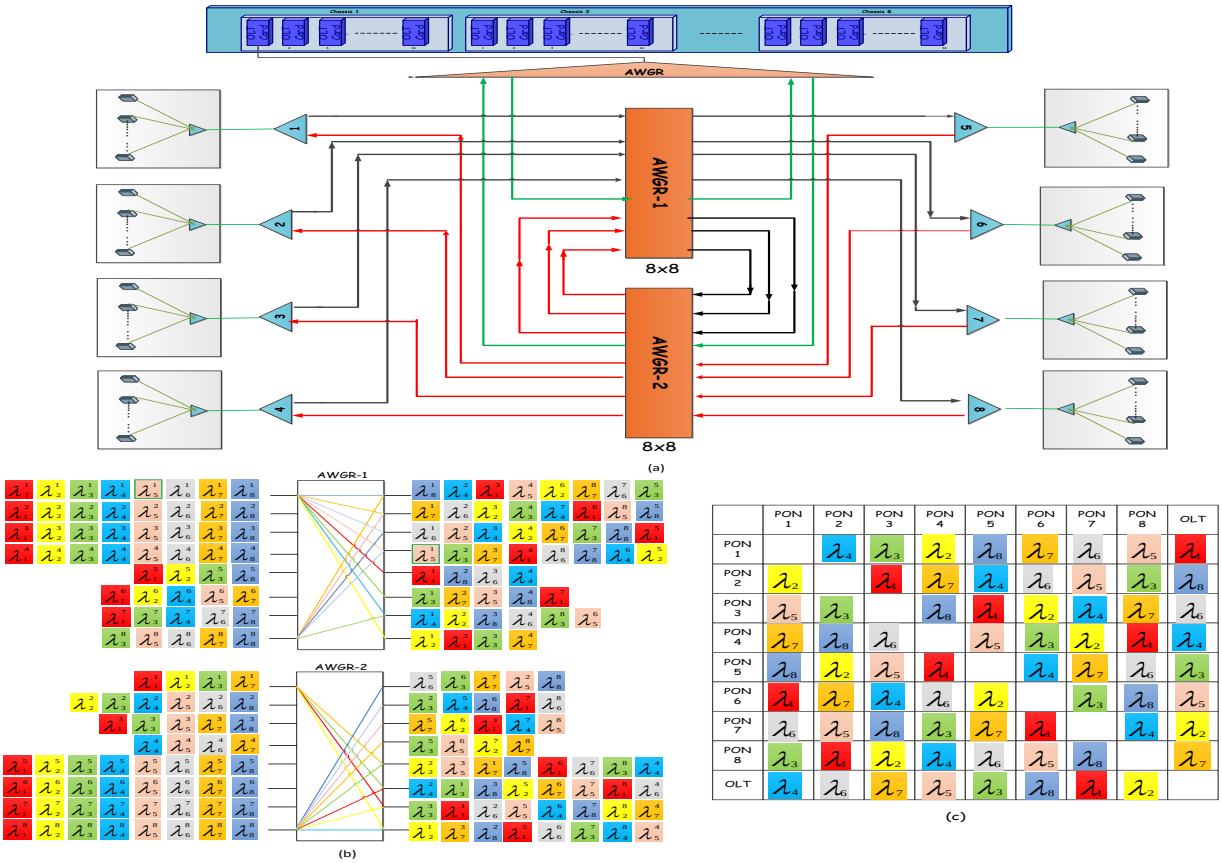

Figure. 3 (a) Fabric interconnection design for a PON cell with tuneable lasers for 8 PON groups, (b) Obtained MILP configuration for AWGRs interconnection for wavelenght routing and assignment, (e) Obtained MILP wavelength routing table for inter rack communication

3. POWER CONSUMPTION BENCHMARKING

In this section, we present a benchmarking study that compares the power consumption of our proposed PON data center architecture to the most common conventional data center architectures; Fat-tree [12] and BCube [13].

The Fat-Tree data center topology, depicted in Figure 4, is built using identical n-port commodity switches in access, aggregation and core layers. The Fat-Tree topology consists of n pods, each of which consists of (n/2) access switches and (n/2) aggregation switches and hosts n servers. Therefore the Fat-Tree topology with n-pods consists of $(n/2)^2$ core switches and can support $(n^3/4)$ servers.

The BCube data center topology, shown in Figure 5, is constructed in a recursive manner starting at $BCube_0$ as its basic building block. $BCube_0$ consists of an n-port commodity switch connecting n-servers. $BCube_1$ is then constructed from n-$BCube_0$ with a total of n n-port commodity switches. The architecture in general is denoted as $BCube_k$ where k+1 defines the number of levels. For $k \geq 1$, the BCube topology is constructed from n $BCube_{(k-1)}$s, n^k n-port switches and servers with k + 1 ports. In $BCube_k$, the total number of servers is n^{k+1} and the total number of levels is k+1with each level consisting of n n-port switches. Figure 5 presents an example of $BCube_k$ structure with k=1 and n=4.

We evaluated power consumption for the Fat-tree and BCube architectures for different fabric configurations. We considered Fat-Tree architectures with 24 and 48 pods to provision connectivity to 3456 and 27648 servers, respectively. The BCube architecture was evaluated with k=2, 3 and 4 for n= 8 to provision connectivity to 512, 4096 and 32768

servers, respectively. We evaluated the power consumption of the PON architecture depicted in Figure 3(a) with 64 servers arranged in 8 PON groups.

Table 1 shows the power consumption of the equipment used for the benchmark evaluations. The Cisco 2960-24TC-L and Cisco 2960-48TC-L switches are used to build the 24 and 48 pods Fat-Tree architectures and the Cisco 2960-8TC-L switch is used to build the BCube architecture. To the best of our knowledge, no studies or vendors have provided power consumption specification for OLT PON transceivers that supports rates of 10Gb/s. The GPON NEC CM7700S OLT [14] supports 1G data rate for typical distances of PON architectures in access network (20km) and consumes 12.5W per port. We estimated the power consumption of the 10Gb/s OLT port based on the GPON NEC CM7700S OLT assuming a linear power profile. A 10 Gb/s OLT port consumes, therefore, 125W. As a conservative estimate we do not consider the reduction in power consumption due to the limited transmission distance in data centers.

Table 1. Equipment power consumption

Equipment	*Power consumption*
Cisco 2960-48TC-L	39W [15]
Cisco 2960-24TC-L	27W [16]
Cisco 2960-8TC-L	12W [17]
10 Gb/s Tuneable ONU	2.5W [18]
10 Gb/s OLT Port	125W
Server's Transceiver	3W [19]

Figure 6 shows the power consumption savings achieved by deploying our proposed PON architecture compared to the Fat-tree and BCube architectures. The high energy consumption of BCube and Fat-Tree architectures is mainly due to the high number of switches used for the interconnections. As discussed above, these switches are eliminated from the PON design and replaced by passive optical devices. Therefore the proposed PON architecture has reduced the power consumption by 45% and 80% compared to the Fat-tree and BCube architectures for 3,456 and 32,768 servers, respectively. The BCube architecture has the highest power consumption as it is a server centric architecture where servers are equipped with multiple transceivers needed to establish connectivity with all the levels. As the levels increase, the architecture can be scaled up to host more servers and the number of transceivers increases as each server needs to have connections with a switch in every level, hence the power consumption increases. The Fat-tree architecture is a switch centric architecture and has lower power consumption as it is designed to have servers with single transceivers to connect to the Top of Rack (ToR) switch. The savings achieved by the proposed PON architecture compared to the Fat-Tree architecture decrease as of the number of server increases. This because the power consumption of the switches used to build the 24 pods and 48 pods Fat-Tree architectures does not increase linearly as the number of pods increases. On the other hand the power consumption savings achieved compared to the BCube architecture increases as the number of servers increases as a result of the increase in the number of transceivers needed as the architecture is scaled up to host a higher number of servers.

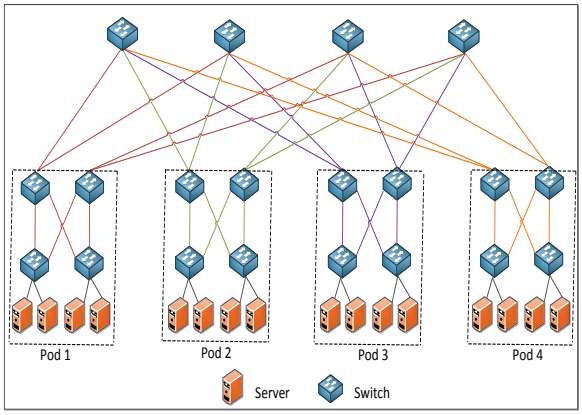

Figure. 4 Fat-Tree data center topology with n=4

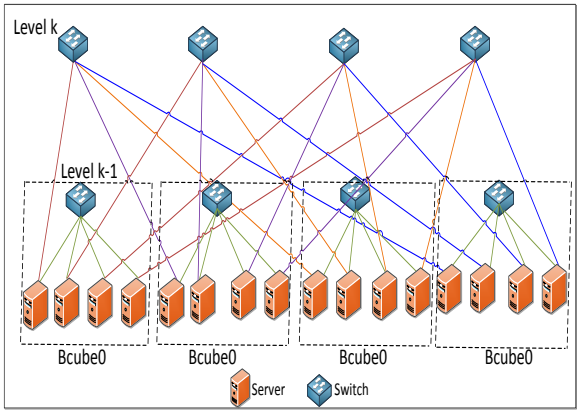

Figure. 5 BCube data center topology ($BCube_1$) with n=4 and k=1

4. CONCLUSIONS

This paper has proposed a PON architecture that eliminates the need for high cost and power consuming access and aggregation switches used in conventional data centre architectures. The proposed architecture is based on passive optical devices such as AWGRs, Couplers, and FBGs and is used to establish inter and intra communications within the PON cell. The interconnection of the AWGR are optimized to facilitate inter rack communication among servers within

the same PON cell. The architecture can be scaled up to host hundreds of thousands of servers. A benchmarking study has shown that our proposed PON architecture energy efficient design can reduce the power consumption of data centres by 45% and 80% compared to Fat-tree and BCube architectures, respectively.

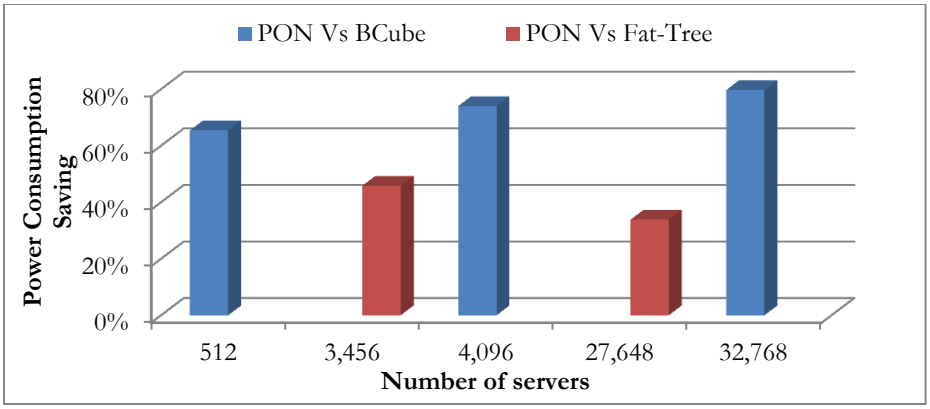

Figure. 6 Power consumption savings for BCube and Fat-tree architectures against PON architecture

ACKNOWLEDGEMENTS

The authors would like to acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC), INTERNET (EP/H040536/1) and STAR (EP/K016873/1).

REFRENCES

[1] A. Hammadi and L. Mhamdi: Review: A survey on architectures and energy efficiency in Data Center Networks, Computer Communication, vol. 40, pp. 1-21, 2014.

[2] Z. Yan, N. Ansari: On Architecture Design, Congestion Notification, TCP Incast and Power Consumption in Data Centers, *Communications Surveys and Tutorials, IEEE,* vol. 15, pp. 39-64, 2013.

[3] M. Nakamura, H. Ueda, S. Makino, T. Yokotani, K. Oshima: Proposal of networking by PON technologies for full and ethernet services in FTTx, IEEE Journal of Lightwave Technology, vol. 22, pp. 2631-2640, 2004.

[4] P. Ji, D. Qian, K. Kanonakis, C. Kachris, I. Tomkos : Design and evaluation of a flexible-bandwidth OFDM-based intra data center interconnect, IEEE Journal of Selected Topics in Quantum Electronics, vol.19, 2013.

[5] L. Yuanqiu, F. Effenberger, S. Meng: Cloud computing provisioning over Passive Optical Networks*, 1st IEEE International Conference in China (ICCC)*, 2012, pp. 255-259.

[6] C. Kachris, I. Tomkos: Power consumption evaluation of hybrid WDM PON networks for data centers: in *Networks and Optical Communications (NOC), 16th European Conference*, 2011, pp. 118-121.

[7] K. Wang, L. Zhao, H. Gu, X. Yu, G. Wu, J. Cai: ADON: a scalable AWG-based topology for datacenter optical network, *Optical and Quantum Electronics,* pp. 1-14, 2015.

[8] J. M. H. Elmirghani, Hammadi, A. and El-Gorashi, T.E. : Data Centre Networks, UK patent, 26 November 2014.

[9] A. Hammadi, T. E. El-Gorashi, and J. M. H. Elmirghani: PONs in Future Cloud Data Centers, *IEEE Communications Magazine.*

[10] J. Beals IV, N. Bamiedakis, A. Wonfor, R. Penty, I. White, J. DeGroot Jr*, et al.*: A terabit capacity passive polymer optical backplane based on a novel meshed waveguide architecture, *Applied Physics A,* vol. 95, pp. 983-988, 2009.

[11] J. C. Palais, *Fiber optic communications*: Prentice Hall, 1988.

[12] M. Al-Fares, A. Loukissas, and A. Vahdat: A scalable, commodity data center network architecture, in Proc. ACM SIGCOMM, Seattle, WA, USA, 2008.

[13] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi*, et al.*: BCube: a high performance, server-centric network architecture for modular data centers, *SIGCOMM Comput. Commun. Rev.,* vol. 39, pp. 63-74, 2009.

[14] J. Baliga, R. W. A. Ayre, W. V. Sorin, K. Hinton, and R. Tucker: Energy Consumption in Access Networks," in *Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, 2008, pp. 1-3.

[15] Cisco: Data sheet of Cisco-2960-48TC-L DataSheet, [Online]. Avaiable : http://www.cisco.com/c/en/us/support/switches/catalyst 2960g-48tc-l-switch/model.html#DataSheets

[16] Cisco: Data sheet of Cisco-2960-24TC-L DataSheet, [Online]. Avaiable : http://www.cisco.com/c/en/us/support/switches/catalyst-2960-24tc-l-switch/model.html

[17] Cisco: Data sheet of Cisco-2960-8TC-L DataSheet, [Online]. Avaiable : http://www.cisco.com/c/en/us/support/switches/catalyst-2960-8tc-l-compact-switch/model.html

[18] K. Grobe, M. Roppelt, A. Autenrieth, J. P. Elbers, M. Eiselt, :Cost and energy consumption analysis of advanced WDM-PONs, *Communications Magazine, IEEE,* vol. 49, pp. s25-s32, 2011.

[19] Gyarmati, T. A. Trinh: How can architecture help to reduce energy consumption in data center networking?, in Proc. 1st International Conference on Energy-Efficient Computing and Networking, Passau, Germany, 2010.