100G Beyond Ethernet Transport for Inter- and Intra-DCN communication with Solutions and Optical Enabling Technologies in the ICT STRAUSS Project

S. Yan, S. Peng, Y. Yan, B. R. Rofoee, Y. Shu, E. Hugues-Salas, G. Zervas, D. Simeonidou
High Performance Networks Group
University of Bristol, Bristol, UK

M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal
Optical Networks and Systems Department
CTTC, Castelldefels (Barcelona), Spain

Y. Yoshida, P.J. Argibay-Losad, K. Kitayama
Dept. of Electrical, Electronic and Information Eng.
Osaka University, Osaka, Japan

M. Nishihara, R. Okabe, T. Tanaka, T. Takahara, J. C. Rasmussen
Network Products Business Unit
FUJITSU LIMITED, Kawasaki, Japan

C. Kottke, M. Schlosser
Photonic Networks and Systems Department
Fraunhofer Heinrich Hertz Institute, Berlin, Germany

F. Jimenez Arribas, V. López
Scalable Multilayer Photonics Networks
Telefónica I+D, Madrid, Spain

***Abstract*—A multi-domain optical infrastructure with end-to-end Ethernet transport capability can deliver Ethernet services over a large scale and provide a promising solution for inter data center networks (DCN) communication. The already existed metro and core networks should be evolved both in data plane and control plane towards to support the heterogeneous and dynamic Ethernet traffic environment.**

In this paper, we report the work carried out in the ICT STRAUSS project to provide Ethernet connections for intra-DCN and inter-DCN over metro and core networks. The key technologies for intra- and inter- DCN communications are reported with experimental validation.

Keywords—Optical communication; data center network;

I. INTRODUCTION

The Strauss project aims to define a highly efficient and global multi-domain optical infrastructure for Ethernet transport. The multi-domain optical infrastructure covers optical packet switching networks (OPS) for DCN, metro and core networks. In such a multi-domain, multi-technology network, software defined networking (SDN) principles could provide network orchestration and enable an end-to-end optical transport service [1]. In terms of the data plane, "Layer 2" Ethernet service will be provided in all the optical domains. As the optical Ethernet technologies provide network versatility in a simple, speed, reliable and cost-efficient manner, it has gained great success in different optical domains. So far, the optical Ethernet has been deployed over fiber for short-range communications, over resilient packet ring to provide a flexible, multi-service implementations span over thousands of kilometers for metro networks [2], and even over flex-grid-based elastic optical core networks. In STRAUSS project, the multi-domain optical infrastructure aims to enable end-to-end Ethernet service delivery on a global scale, which offers some attractive advantages in terms of service provision and network management.

Recently data center networks (DCNs) have experienced a remarkable growth both in scale and in numbers of new deployment. Ethernet services are widely used both for intra- and inter-DCN communications. In future, the large-scale network services will require several data centers work together to provide smooth user experiences. Thus the inter-DCN traffic will soar, not only to provide optical links connected remote located DCNs, but also to enable possible server-to-server low latency communications. The direct server-to-server cross-DCN connections enable the remote distributed DCNs to appear as one big data center, which could enhance scalability and minimize latency and cost.

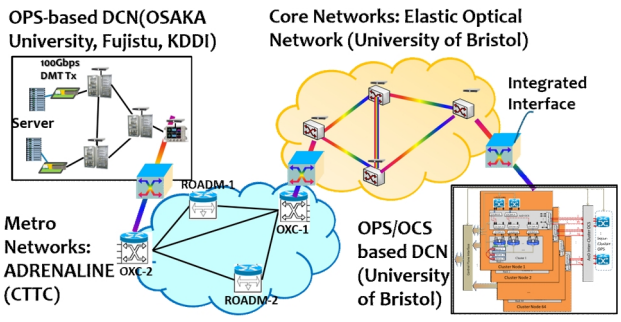

Fig. 1. Use case for end-to-end transparent optical networks: inter-DCN communication in the ICT STRAUSS project

The mainframe of STRAUSS project provides an efficient network infrastructure for Ethernet-based inter- and inter-DCN

communications with capability to deliver Ethernet service globally. Fig. 1 shows the main infrastructure for both intra- and inter-DCN communications provided by STRUASS project. The OPS architecture proposed in STRAUSS aims to provide the granularity and dynamicity required by intra-data center networking. Two distributed DCNs, one of which is an OPS-based DCN with Discrete Multi-Tone (DMT) and OFDM technologies, while another one is adopting both OPS and optical circuit switching (OCS) technologies, are interconnected by means of an optical transport infrastructure with metro and core networks to realize an end-to-end Ethernet transport. The unified Ethernet service is jointly provided over multiple optical domains with simplified network managements. In this paper, the recent progress of the developed multi-domain, multi-technology network infrastructure in the ICT STRAUSS project is reported to deliver Ethernet services over several testbeds for inter-DCN communications.

II. Solutions and Optical Enabling Technologies

A. Optical Packet Switching for DCN

Offloading the traffic from existing electrical packet switching core to optical switching network layer seems the only solution to realize powerful and sustainable warehouse-scale-DCs. So far, several intra-DCN architectures employing optical switching technologies have been reported [3]–[5]. Among the architectures, the optical interconnect networks based on packet-based optical switching technology (or its hybrid) has received continuous attention due to its efficiency in network utilization based on the statistical multiplexing effect. There has been a growing demand for agile reconfigurability in DCNs, which need for supporting multi-tenancy via network virtualization by using virtual machines. OPS-assisted DCNs, which transport optical packets in the self-routing manner, potentially offer the finest reconfigurability by removing the overhead for the optical path provisioning in OCS-based DCNs.

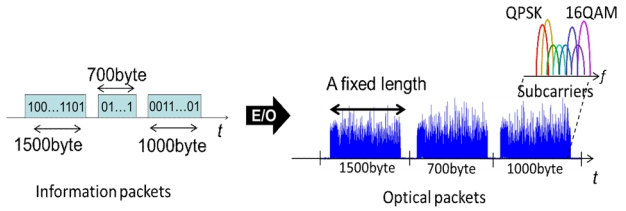

Fig. 2. Fixed-length variable-capacity payload packet based on the multi-carrier adaptive modulation technique

In the STRAUSS project, we develop a novel optical payload format for the OPS network based on DSP-enabled optical multi-carrier modulation technique, such as OFDM and DMT. Fig. 2 shows the conceptual image of the proposed fixed-length variable-capacity (FL-VC) payload packet, wherein incoming variable-length electrical packets are *packed* into fixed-length optical packets by adapting the modulation format packet-by-packet. The employment of the fixed-length optical packets can ease optical buffer implementation and scheduling. Moreover, the concept fits in well with SDN, since the adaptation is based on high-speed DSP circuit and hence programmable. Depending on bandwidth demands and/or network architectures, the SDN-controller can handle the adaptation policy to further improve the network resource usage. In [1], we have demonstrated the DMT-based FL-VC packet switching and its OpenFlow control in our preliminary OPS network testbed, where the 283.5-ns-long optical DMT payloads successfully carried 1500 to 3500 byte depending on their transmission distances (Fig. 3). Although the size of the preliminary testbed was limited, our theoretical analysis shows that the proposed distance-adaptive FL-VC OPS network can achieve 1.5 (w/ uniform traffic distribution, UDP, 28-node ring) to 30 (w/ localized traffic, the modified TPC, the optimized packet length, 28-node ring) times larger application throughput than the conventional OPS networks with a fixed payload bitrate [6].

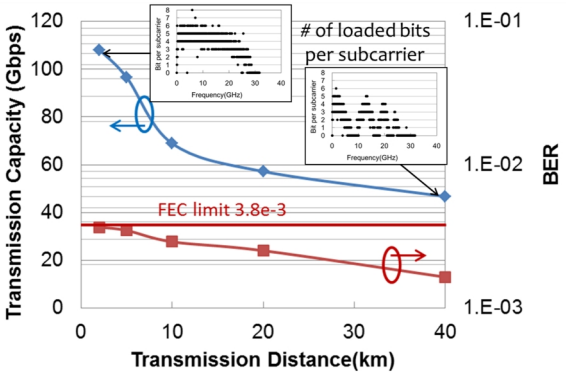

Fig. 3. BER performance and achievable capacity versus transmission distance of DMT-based FL-VC optical packet [1]

B. Multicarrier technologies for intra/inter DCN

a)Discrete Multi-tone technology for high capacity transmission

Discrete Multi-Tone technology is OFDM-based multicarrier modulation format, which transmits data by intensity domain of optical carrier signal and does not use phase domain. The bit allocation of each subcarrier is determined by transmission characteristics obtained by transmission of probe signal. DMT technology realizes high spectral efficiency with simple configuration based on direct detection and it is attractive especially for short reach application, such as intra-data center network. DMT technology is expected as the key technology for the next generation Ethernet transceiver and discussed as one of strong candidates in 400Gb/s Ethernet Task Force of IEEE802.3 [7].

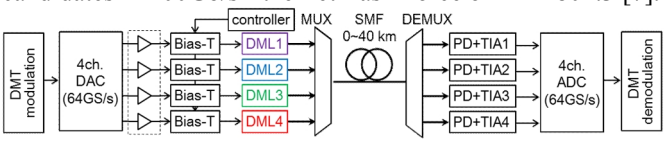

Fig. 4. An Experimental setup of 400G (116 Gbps x 4λ) DMT transmission

Configuration of 100 Gb/s × 4 wavelengths had been proposed for 400 Gb/s DMT transmission [8], [9]. Fig. 4 shows the experimental setup. 4 channels DMT signals were generated by offline processing and converted to analog

signals by a 64 GSA/s, 8bits digital-to-analog converter (DAC). They modulate the directly modulated lasers (DMLs) whose wavelengths were on LAN-WDM grid (1295.0, 1299.5, 1304.1, and 1309.0 nm). The optical DMT signals were wavelength division multiplexed (WDM) by optical multiplexer (MUX) and transmit over transmission fiber. Transmitted signal was demultiplexed to each wavelength by an optical demultiplexer (DEMUX) and received by the optical receiver (PD+TIA). The received signals were converted to digital signal by a 64 GSa/s, 8bits analog-to-digital converter (ADC) and demodulated by offline processing. The bit rate of each wavelength was 116 Gb/s corresponding the payload of 103.125 Gb/s with overhead of 12.5 % forward error correction (FEC). The transmission characteristics of 400G DMT signal are shown in Fig. 5. Horizontal axe is transmission distance and vertical axe is bit error rate (BER) of each wavelength. BER under the FEC limit ($3x10^{-3}$) was achieved for the transmission distance up to 30 km.

DMT technology is an attractive technology for 100G beyond Ethernet transport from its high spectral efficiency with simple configuration realized by multicarrier modulation.

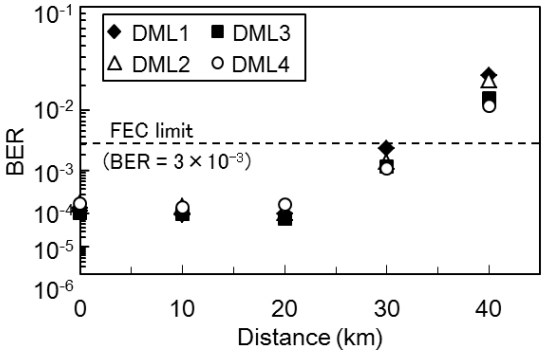

Fig. 5. Transmission distance vs. BER of 400G DMT transmission

b)Bandwidth variable transceiver based on OFDM technology

Multicarrier technologies, and particularly OFDM, are promising candidates for the design of programmable adaptive bandwidth variable transceivers (BVT) targeting inter-DCN connectivity. In fact, they enable implementing rate/distance adaptive transceivers thanks to the unique ability of manipulating the individual subcarriers, which can be independently loaded with the suitable number of bits and gain coefficients. Specifically, using adaptive algorithms at the digital signal processing (DSP) allows mitigating both the chromatic dispersion and the components limitation, according to the channel estimation (SNR profile) performed at the BVT receiver. In our investigation within the STRAUSS project, it has been evidenced that using margin adaptive (MA) algorithms, either optimal, such as Levin Campello, or suboptimal, like Chow Cioffi Bingham, outperform the use of variable bandwidth uniform loading in terms of achievable rate and reach [10]; thus, enabling high data rate transmission over an extended link between data centers.

Combining this powerful DSP with direct detection (DD), we propose to simplify the optoelectronic front-end for a cost-effective transceiver implementation. At the transmitter side an external Mach Zehnder modulator (MZM) driven by a tunable laser is used to optically modulate the OFDM digital signal. Thus, a fine tuning of optical carrier and electrical subcarriers is possible. Additionally, by properly selecting the MZM bias point, the transceiver can be suitably adapted/configured to the targeted transmission rate and distance. Conversely to DMT, which is severely affected by chromatic dispersion, OFDM allows enhancing the reach, even if DD is adopted for cost-effectiveness issue. In fact, it can be combined with single-sideband (SSB) transmission at the expense of slightly more complex transceiver architecture, including an additional optical filter to select the sideband. Particularly, SSB OFDM allows increasing the BVT robustness against transmission impairments for extending the achievable optical reach up to cover the metro segment.

Preliminary assessment within the 4-node mesh network of the ADRENALINE testbed (shown in Fig. 1) has been performed using a DAC with maximum sample rate of 24 GSa/s. We have analyzed a DSP-enabled BVT based on m-QAM adaptive loading (with m up to 256) and complex FFT processing for the OFDM modulation. The obtained net rate/distance adaptive figure has been 20 Gb/s over 185 km (2 hops ADRENALINE path) and 25 Gb/s over 35 km (single hop) path, considering a maximum bandwidth occupancy of 12.5 GHz per channel. The performance of the analyzed BVT in terms of supported data rate and bandwidth can be enhanced using a DAC with increased speed at the programmable transceiver.

Furthermore, the proposed OFDM transceiver architecture can be seen as the building block of a sliceable BVT [11], supporting multiple-format, multiple-rate, multiple-reach, multiple-flow transmission, and characterized by unique scalability and granularity (from the sub- to the super-wavelength level) thanks to the multicarrier technology concept.

c) Real-time OFDM transmitter

In the project also a real-time OFDM transmitter for Ethernet transport is evaluated. This is based on real-time FFT processing. Specifically, the transmitter is realized as an FPGA based realtime implementation. Using a highly parallel architecture, and a very efficient use of hardware resources, the OFDM transmitter fits onto a single Virtex-6 FPGA, including the interface to two digital-to-analog converters yielding the inphase (I) and quadrature (Q) signals in real-time. 1024 OFDM subcarriers are realized and can be clocked at up to 16 GHz. This can achieve a gross data rate up to 64 Gb/s. Real-time implementation enables the exploitation of the inherent flexibility of OFDM depending on the individual channel properties of multiple users and their instantaneous traffic load.

Bit loading (BL) is implemented so that the rate can be dynamically adapted using an individually selectable modulation scheme (None, BPSK, QPSK, 16-QAM) and transmitter power for each sub-carrier.

Within the FEC limit in an electrical back to back scenario 48 Gb/s could be realized. For now in a transmission over 36

km installed fibre 14.7 Gb/s could be achieved. We expect an improved performance with a redesign of the receiver.

C. *Flexible/adaptable optical nodes for flexi-grid DWDM networks*

The Ethernet traffic will be highly variable and complex. The varied requirements of different applications, changes in customer behavior, uneven traffic growth, or network failures will lead to the uncertainty in traffic demands, granularity, and geographic and temporal distributions. To handle the uncertainty nature of Ethernet traffic, an optical switching node in core networks should be evolved to provide the multidimensional switching capability and more flexibility in switching and even in the main architecture.

To handle the heterogeneous and dynamic traffic environment in flexi-grid DWDM networks, architecture-on-demand (AoD) based flexible modular and scalable optical node is proposed using a large-port-count fiber switch (referred as AoD switch). Fig. 6 shows the design of the proposed AoD-based flexible modular and scalable optical node.

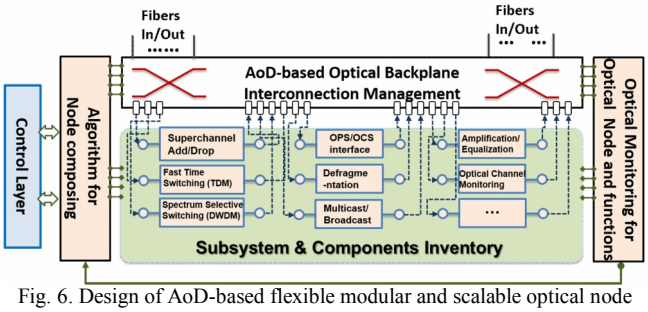

Fig. 6. Design of AoD-based flexible modular and scalable optical node

The proposed optical node comprises of four main parts. A large port count fiber switch (LPFS), used as an optical backplane, provides a large scale switch matrix and enables the programmable feature of the proposed optical node. Several technologies, such as 3D MEMS (micro-electro-mechanical system) and beam steering technology, provide an optical fiber switch matrix up to 320×320. The available LPFSs enable the synthesis of large scale AoD optical node with the subsystem and component inventory. The inventory as other key parts of the node design shown in Fig. 6, is composed of several stand-alone subsystems to provide network functions. These network functions can be deployed in the synthesized AoD node, e.g., superchannel add/drop, OPS/OCS interface, amplification equalization, etc. The inventory also consists of some common optical components, such as optical multiplexers, optical demultiplexers, optical couplers and optical attenuators to compose network function modules when needed. The optical backplane manages the interconnections between the function modules and also handles both the input and output ports of the optical node. This multi-dimensional switching technologies, i.e., frequency, time and space, can be deployed to any optical link when required. The node composing module talks to the control layer and response the network requests, including node functions, wavelength allocation, and bandwidth requirements, then synthesizes optical node architecture by configuring the LPFS based optical backplane. Some node composing algorithms will be deployed here to balance the payload of the optical node to use the network hardware efficiently.

Another important part of the optical node is the optical monitoring blocks, which will monitor the optical node with a set of optical channel monitoring technologies. The monitoring information will be feedback to the node composing module for network performance optimization. The monitor information can also be sent to the control layer for network optimization.

The AoD-based flexible and adaptable optical node enables a high level flexibility with multi-dimensional optical switching capability, and provide a promising solution for Ethernet transport in Core networks.

D. *Integrated OPS/OCS interface*

Ethernet have been widely deployed in different optical domains, however, different Ethernet standards are adopted to provide variable link capacities. In addition, an integrated interface is needed between optical packet switching networks and other optical domains. Thus an FPGA-based OPS/OCS interface is developed to bridge OPS and OCS optical domains and also provide Ethernet aggregation functions for different Ethernet standards.

Fig. 7 shows the internal function blocks in our designed OPS/OCS integrated interface. The FPGA-based OPS/OCS interface receives all the OPS/OCS traffic, process and store the data, and send OCS/OPS out as controlled by the control plane. Incoming traffic is received using multiple SFP+ (small factor pluggable) interfaces. Then the ingress traffic is aggregated and groomed based on its destination MAC address and the virtual network slices they belong to. The designed OPS/OCS interface can handle variable size of Ethernet traffic (with VLAN tag) from 64 Bytes up to 1522 Bytes and flew-in on different bit rates up to 10Gbit/s.

The current platform used is the Hitech Global HTG-V6HXT-100GIG-380 board. It features with Xilinx Virtex 6 HX380T chip and high throughput IOs. Due to the limitation of the chip size, currently the FPGA-based design architecture, supports 2 OPS interfaces, and 1 OCS interface.

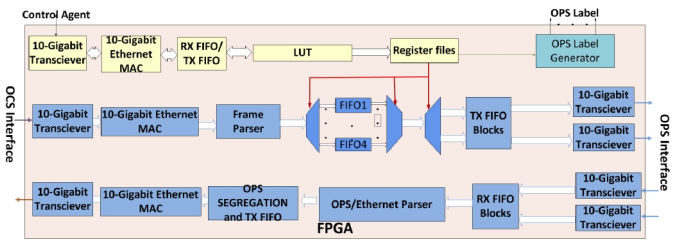

Fig. 7. Internal blocks of the designed FPGA-based OPS/OCS interface

To provide further aggregation, an optical bandwidth variable transmitter is connected to the designed OPS/OCS interface [12]. As shown in Fig. 8, the traffic can be further aggregated to generate either 16QAM or QPSK signal for transmission in core network.

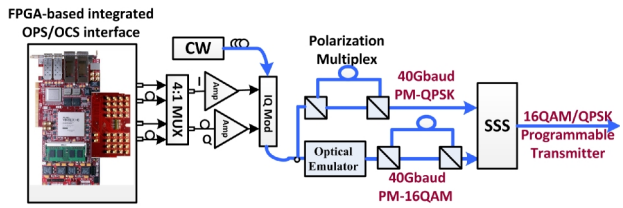

Fig. 8. Experimental setup of QPSK/16QAM multi-format transmitter with OPS/OCS interface

E. *OPS/OCS based DCN*

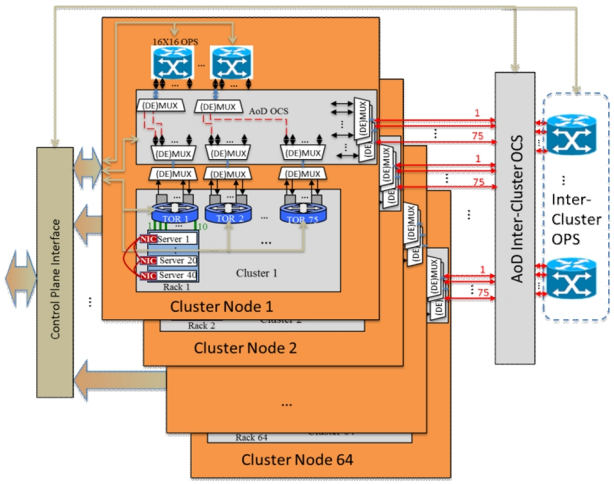

Fig. 9. LIGHTNESS DC Architecture

In order to accommodate the ever-growing workload driven by new and emerging high-performance DC and cloud applications, the EU FP7 project LIGHTNESS designs and implements a new flat DCN architecture, as shown in Fig. 9 [13]. LIGHTNESS DCN comprises all-optical switching technologies (OPS and OCS), hybrid Top-of-the-Rack (ToR) switches and programmable NICs, controlled and operated by a SDN based control plane. The combination of OPS and OCS makes it possible to switch traffic in space, frequency and time, therefore to enhance the capability to handle diverse traffic flows in DCs. Instead of a hardwired interconnection of different DCN devices, an AoD node is adopted to achieve a more flexible DCN architecture. The AoD node consists of an optical backplane (e.g. a 192x192 Polatis switch) with passive and/or active network elements (e.g. Mux/DeMux, OPS, ToR switches) plugged into it. With this AoD-based DCN architecture, different arrangements of inputs/outputs and network elements in between can be dynamically constructed by setting up appropriate cross-connections on demand in the optical backplane to support various applications' requirements. The SDN-based control plane interacts with the optical data plane devices through the southbound interfaces (e.g. extended OpenFlow protocol) to perform capabilities and statistics collection, device configurations as well as network resources monitoring. At the northbound, a set of open application programming interfaces is provided, which enables the integration with external orchestration functions (e.g. OpenStack) and also opens the opportunity to implement enhanced network functions as applications (e.g. DC virtualization [14]) running on top of the SDN controller.

III. CONCLUSIONS

This paper reported the STRAUSS data plane solutions to deliver both intra- and inter-DCN Ethernet services over metro and core networks. A number of key technologies in different domains are developed to handle the dynamic Ethernet traffics. With the developed solutions and optical enabling technologies, the joint testbed of STUAUSS project will demonstrate a global scale Ethernet interconnections for future DCN communication in the next year.

ACKNOWLEDGMENT

This work was partly supported by the FP7 EU STRAUSS project (n°608528) and LIGHTNESS (n°318606). This work is also partly funded by the Japanese Ministry of Internal Affairs and Communications (MIC) and National Institute of Information and Communications Technology (NICT) through the EU-Japan coordinated project STRAUSS.

REFERENCES

[1] Y. Yoshida, et al., "SDN-based Network Orchestration of Variable-capacity Optical Packet Switching Network over Programmable Flexi-grid Elastic Optical Path Network," *J. Light. Technol.*, vol. PP, no. 99, pp. 1–1, 2014.

[2] V. Ramamurti and G. Young, *Ethernet Transport over RPR*. IEEE, 2001.[online]www.ieee802.org/17/documents/presentations/sep2001/vr_ethman_02.pdf

[3] W. Zhang, H. Wang, and K. Bergman, "Next-generation optically-interconnected high-performance data centers," IEEE/OSA J. Lightwave Technol., Vo.30, No.24, pp.3836-3844, 2012..

[4] C. Kachris, K. Kanonakis, and I. Tomkos,, "Optical interconnection networks in data centers: Recent trends and future challenges," IEEE Commun. Magazine, pp.39-45, Sept. 2013

[5] K. Kitayama and et. al., "Torus-topology Data Center Network Based on Optical Packet / Agile Circuit Switching with Intelligent Flow," *IEEEOSA J Light. Technol*.

[6] P. Argibay-Losada, Y. Yoshida, A. Maruta, and K. Kitayama, "Throughput analysis of distant-adaptive, fixed-length, variable-capacity packets in OPS networks," to appear in IEEE ICC 2015.

[7] IEEE P802.3bs 400 Gb/s Ethernet Task Force, http://www.ieee802.org/3/bs/index.html.

[8] T. Tanaka, M. Nishihara, T. Takahara, W. Yan, L. Li, Z. Tao, M. Matsuda, K. Takabayashi, and J.C. Rasmussen, "Experimental Demonstration of 448-Gbps+ DMT Transmission over 30-km SMF," OFC2014, M2I.5, 2014.

[9] R. Okabe, T. Tanaka, M. Nishihara, Y. Kai, T. Takahara, H. Chen, W. Yan, Z. Tao, and J.C. Rasmussen, "Investigation of Fiber Dispersion Impairment in 400GbEb Discrete Multi-tone System for Reach Enhancement up to 40 km," PhotonicsWest2015, 9388-15, 2015.

[10] L. Nadal, et al., "DMT Modulation with Adaptive Loading for High Bit Rate Transmission Over Directly Detected Optical Channels", IEEE/OSA J. Lightwave Technol., vol. 32, no. 21, pp. 3541-3551.

[11] M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal, and F. J. Vílchez, "Optical transceiver technologies for inter-data center connectivity", in Proc. ICTON 2014, paper Mo.D1.4..

[12] S. Yan, et al., "Real-time Programmable Ethernet to Reconfigurable Superchannel Data Conversion," *J. Light. Technol.*, vol. PP, no. 99, pp. 1–1, 2015.

[13] S. Peng, et al., "A novel SDN enabled hybrid optical packet/circuit switched data centre network: The LIGHTNESS approach," IEEE European Conference on Networks and Communications (EuCNC), Bologna, Italy, June 2014.

[14] S. Peng, et al., "Enabling multi-tenancy in hybrid optical packet/circuit switched data center networks," in *2014 European Conference on Optical Communication (ECOC)*, sep. 2014