Photonic networks with nanoseconds switching and control for distributed machine learning systems

Nicola Calabretta
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
n.calabretta@tue.nl

Xuwei Xue
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
x.xue1@tue.nl

Henrique Freire Santana
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
h.freire.santana@tue.nl

Aref Rasoulzadeh Zali
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
a.rasoulzadehzali@tue.nl

Xiaotao Guo
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
x.guo@tue.nl

Marijn Rombouts
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
m.p.g.rombouts@tue.nl

Bitao Pan
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
b.pan@tue.nl

Fulong Yan
ECO Group, Eindhoven Hendrik Casimir Institute
Eindhoven University of Technology
Eindhoven, The Netherlands
f.yan@tue.nl

***Abstract*—To flexibly accommodate diverse cloud workloads and AI applications, a novel disaggregated DCN architecture based on photonic integrated switches prototype with low latency and high bandwidth to interconnect computing and memory nodes is presented.**

Keywords— Photonic integrated switch, data center networks, disaggregated networks, distributed machine learning systems

I. INTRODUCTION

Big data applications (like cloud computing, media streaming, and cloud storage) and AI applications, different resource requirements of different applications demand a more flexible, low latency and high capacity next generation data centers (DCs). Current issues in server-centric architectures is that the shared hardware resources cannot meet the application requirements and a large amount of hardware resources (processor, memory, and storage) are frequently underutilized [1]. Moreover, the different progress and rate in upgrading different hardware (processor or memory) in a server causes a significantly impact to the whole server. Therefore, a novel DC architecture consisting of disaggregated hardware nodes (hence naming disaggregated DCN) is developed to provide flexible resource provision and boosting performance for various processor (memory) intensive applications [2-6]. In order to implement the disaggregated DCN, there are several issues to be addressed, in which the disaggregation of the processor and memory is a key challenge. Due to the resource disaggregation, the on-board high-speed point-to-point bus between the processor and memory is replaced by the network interconnection. This difference requires a network interconnection of high bandwidth and low latency to sustain the communication among disaggregated processor and memory nodes. The network switching may also lead to signal impairments (like channel cross-talk, OSNR degradation, signal distortion, and power budget); hence degrading the performance of disaggregated DCNs. Therefore, a stable, error-free, and low packet loss network is also significantly important for disaggregated DC architectures.

Several experimental studies have been conducted to investigate the feasibility of disaggregated DCNs. According to the analysis in [7], the extra network latency performs the most impact to the performance of disaggregated DCNs, while the low packet loss and long-term network stability are also critical. Some disaggregated architectures were proposed based on current servers and hierarchical electrical network to implement decoupled memory blades [8, 9]. However, the low speed peripheral bus in servers and intrinsically large latency of the multi-tier network interconnection result in the application performance deterioration, while frequent O/E/O conversions increase the power consumption of disaggregated DCNs. Disaggregated DCNs based on optical switches were proposed and experimental investigated in [10, 11] exploiting high aggregation bandwidths and transparent switching, but the high switching delay (milliseconds) of the optical switch technology cannot provide nanoseconds scale communication between processor and memory nodes. A disaggregated DCN architecture was proposed in our previous work based on nanoseconds optical switches (NOS) [12]. The NOS is based on a broadcast and select switch architecture employing semiconductor optical amplifiers (SOA) based optical gates Exploiting the nanoseconds switching of SOA-based gates and parallel optical flow control, the proposed architecture can potentially provide an optical network interconnection of high bandwidth and low latency for disaggregated DCNs. However, only application performance of the proposed disaggregated architecture was numerically assessed, and the impact of optical network switching based on NOS as well as network performance must be experimentally investigated to validate the feasibility and scalability of the disaggregated DCN architecture.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

In this work, we present an experimental demonstration of a distributed NOS based disaggregated architecture prototype. The hardware resource nodes (processor and memory) are implemented by field-programmable gate array (FPGA) platforms. The network traffic is processed based on two parallel data-path and control-path channels. The optical flow control of the control-path channel is exploiting to solve contention due to the lack of optical buffer. The disaggregated architecture prototype is experimentally assessed with a 4×4 NOS and 4 processor/ memory nodes. The prototype provides an error-free operation with a power penalty of 0.5dB at BER of 1E-9. Meanwhile, it is shown in the network performance assessment that the prototype achieves a minimal node-to-node latency of 122.3ns and no packet loss during the stable operation. Scalability of the NOS based disaggregated architecture is also evaluated in this work. Scaling the NOS port count to 64, an error-free operation with power penalty of 1.5dB are achieved.

II. Disaggregated Data Center Network

The NOS based high-bandwidth, low-latency, and scalable disaggregated DCN architecture is shown in Fig. 1. The disaggregated DCN architecture consists of N racks, and each rack groups p processor nodes (PN) and q memory nodes (MN) (p+q=N). Hardware nodes per rack are connected by an intra-rack NOS (RNOS). The i-th (i= 1, 2, ···, p) processor nodes (PN_i) across N racks are interconnected by i-th NOS for PN ($PNOS_i$), whereas j-th (j= 1, 2, ···, q) MNs (MN_j) per rack are also connected to j-th NOS for MN ($MNOS_j$). The hardware nodes are interconnected with the NOS based on two types of channels: data-path (solid line) and control-path (dashed line) channels. The data-path channel connects to input/ output ports of the NOS to transmit optical packet payloads, while the control-path channel interconnects to the switch controller for sending optical packet labels. The packet processing delay in the NOS is minimized exploiting two parallel data-path and control-path channels. It is shown in Fig. 1 that only a single hop is required for the intra-rack hardware node communication, while at most two hops for the inter-rack hardware node communication. In addition, it is also a single hop for the traffic among processor nodes (memory nodes) in different racks via PNOS (MNOS), which can be applied in specific scenarios like direct memory access and hot virtual machine migration. Therefore, it is important for the scalability of the network to implement a NOS with large divided into M groups, while each group consists of F hardware nodes. More details about the hardware nodes grouping rule are reported in [13]. Based on 2N NOSs with a port count of N, the disaggregated DCN architecture can interconnect up to N^2 hardware nodes, which guarantees the scalability of disaggregated DCNs based on the NOS of a moderate port count. Note that the amount of processor and memory nodes per rack can be different, thus the amount of hardware nodes can be flexibly configured according to the requirement. The schematic of the NOS architecture is depicted in Fig. 2. Leveraging distributed processing modules, the NOS can process multiple optical packets from different hardware nodes in parallel. When the FPGA-implemented switch controller analyzes the packet label from the control-path channel, the packet payload from the data-path channel is broadcasted to SOA-based gates of the 1×F switch using a splitter. The hardware node grouping can scale the port count of the NOS utilizing multiple 1×F switch with smaller port counts. SOA gates can provide nanoseconds switching time and compensate the splitting power loss. Based on the packet destination information in the packet label, the switch controller sets the on/off state of the SOA gates, so that the packet payload is forwarded to the target hardware node. The switch controller is also responsible for solving potential contentions before forwarding the optical packets. If contention occurs, the optical packet with the highest priority is forwarded, while the other packets are blocked by setting the target SOA gate off. The switch controller then sends the acknowledgment signals to corresponding hardware nodes for successfully forwarding the packet (ACK) or retransmitting the packet (NACK) [14]. Benefiting from the structure of distributed modules and nanoseconds switching time of SOA gates, the NOS can provide a scalable network with the minimal network latency and high aggregation bandwidth for the disaggregated hardware nodes interconnection. More details about the NOS can be found in [15]. The functional diagram of processor node is illustrated in Fig. 3(a). There is still a minimal memory (local memory) integrated in the processor node for the operation system processing and necessary data caching. Same as current server-centric architectures, processors first access the static random-access memory (SRAM) based processor cache when processing application data [16]. Once data missing in the processor cache, the instruction and logical data address are sent to memory management unit (MMU). The MMU performs the translation of virtual data addresses to physical addresses. If the physical address of target data locates in the dynamic random-access memory

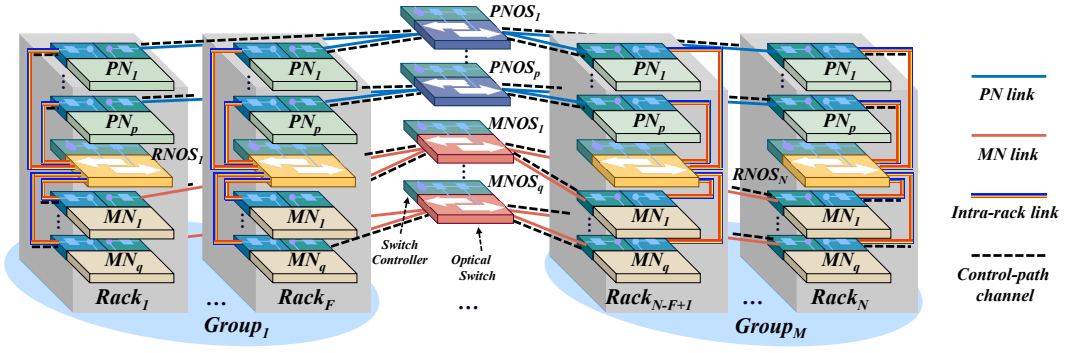

Fig. 1. A disaggregated DCN architecture based on nanoseconds optical switches and optical flow control

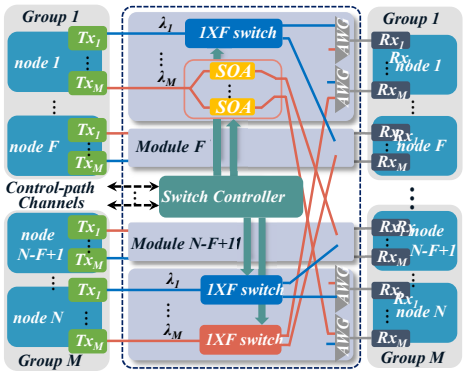

Fig. 2. Schematic of the NOS architecture.

forwarded to the flow controller. Based on the hardware node address containing the target data address, the instruction is packaged in the network interface, and then sent to the corresponding transmitter (TX in Fig. 3). The functional diagram of memory node is shown in Fig. 3(b). The on-board memory resource is based on Double Data Rate 4 (DDR4) DRAM or Hybrid Memory Cube (HMC), managed by the memory controller. Similarly, to the processor node, there are also the flow controller and network interface in the memory node to process packets from other processor/memory nodes. The memory node not only processes packets of reading/writing data from processor nodes, but packets from other memory nodes for the direct memory access.

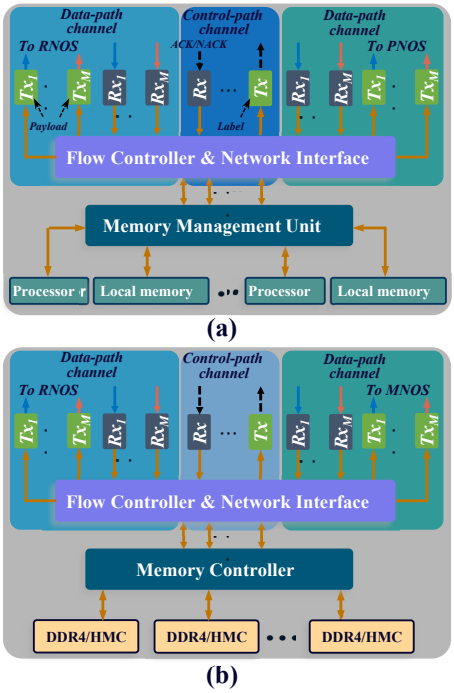

Fig. 3. Functional diagram of (a) processor node and (b) memory node.

III. Assessment of the Disaggregated Prototype

The experimental setup of the NOS based disaggregated DCN architecture prototype is shown in Fig. 4. It consists of 4 FPGA-implemented hardware nodes (2 processor nodes and 2 memory nodes) and 1 4×4 NOS. All the hardware nodes are based on Xilinx Vertex UltraScale VU095 platform [17], while the switch controller is based on Xilinx Vertex-7 VC709 platform [18]. Each hardware node is connected to the NOS by two commercial 10Gb/s SFP+ transceivers operating at 1550nm: one to the SOA based optical switch for the data-

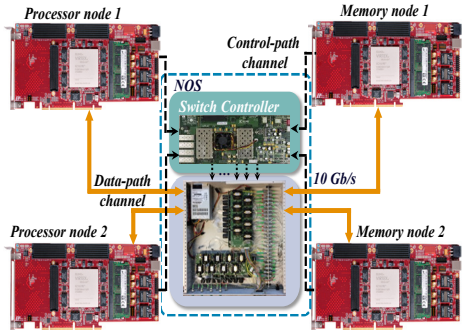

Fig. 4. Experimental setup consisting of 4 hardware nodes and 1 NOS.

path channel, and the other one to the FPGA based switch controller for the control-path channel. The MMU, flow controller, and network interface are implemented in the FPGA programmable chip with the processing clock of 322.3 MHz. The processor node processes the data with a 32-bit width, while the length of optical packet is 64 bytes because this is the typical value length of cache line in current computer architecture. The output power of the SFP transceivers is 1.2 dBm, and the interconnection link between FPGA based nodes and the NOS is 2m. In the NOS, the SOA gates also have 1m fiber pigtails. A pre-ordered logical address-to-physical address table is loaded to processor nodes in the experiment to decide the destination memory node of the target data. The 4x4 NOS consists of 4 1×4 splitter, sixteen SOA gates, and four 4×1 coupler, and it is controlled by the FPGA based switch controller. The SOAs in the NOS not only operate as optical gates but also compensate the switching losses and therefore no extra amplifiers are used between the CPU and memory nodes.

As the NOS is based on broadcast and select architecture, optical power of the signal is split by N, where N is the amount of output port, which results in power losses. The SOA-based optical gates can compensate the splitting losses, but also introduce noise that degrades the quality of output signals. Therefore, we investigate the signal impairments introduced by the SOA gates of NOS. First, we measured the gain spectra of the SOA gate with different input signal wavelength ranging from 1520nm to 1630nm. The employed SOA has a polarization dependent gain less than 1dB. The input signal power is 1.2dBm which is the output power of the commercial SFP transceiver of 10Gb/s. The SOA employed has a broadband gain of up to 9.72dB with a driving current of 60mA, while a gain of up to 14.4dB at 150mA driving current.

Due to the NOS structure of broadcast and select, only one of N SOA-based optical gates is ON state when forwarding optical packets to the specific output port. However, the rest of SOA-based gates at OFF state may not completely block the optical signal, which results in the channel cross-talk when coupling output signals of SOA-based gates. Thus, to quantify the signal impairment introduced by the cross-talk, the ON/OFF ratio of SOA gates is measured under different driving currents as shown in Fig. 5 (a). It is shown that SOA gates achieve a better ON/OFF ratio under a higher driving current, and a ON/OFF ratio higher than 60dB is measured under driving current of larger than 30mA. The output spectra of SOA gate

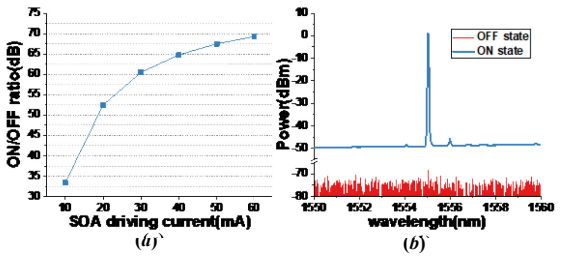

Fig. 5. (a) ON/OFF ratio under different driving currents (b) Output spectra of the SOA gate under ON and OFF states.

are illustrated in Fig. 5(b) under ON and OFF states, in which the driving current is set to 60mA at the ON state. This demonstrates that the SOA gate can compensate the splitting loss with a gain of up to 14.4dB and low cross-talk, which guarantees the physical performance of the NOS.

To quantify the physical performance and feasibility of the NOS based interconnection network, the Bit Error Rate (BER) curve for the disaggregated architecture prototype is measured under a SOA driving current of 60mA, as shown in Fig. 6. The Xilinx IBERT IP Core [19] is utilized in each hardware node to measure the BER. The BER is measured node-to-node and the FPGA based hardware node is equipped with commercial plug-in SFP transceivers. Note that only the SOA is employed in the experiment to amplify the optical signals and compensate splitting losses of broadcast & select architecture, and no EDFA is applied for the BER measurement of disaggregated prototype. The back-to-back BER curve (B-to-B in Fig. 6) is also recorded as the benchmark. Results indicate that an error-free operation is achieved with a power penalty of 0.5dB at BER of 1E-9 under the SOA driving current of 60mA.

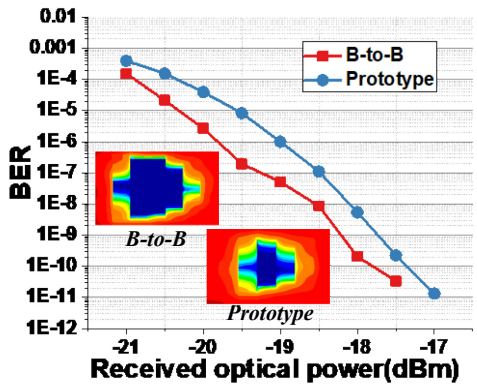

Fig. 6. BER curve and eye diagram of the disaggregated prototype.

This section reports the network performance of the disaggregated architecture prototype. To guarantee the communication among processor and memory nodes, the network latency and packet loss should be minimized. The network latency of the disaggregated prototype includes four components: the processing delay in flow controller and network interface of source and destiantion hardware nodes, packet switching in the NOS, and fiber transmission delay. The network latency of the disaggregated architecture prototye and its components are shown in Fig. 7 when the processor node accesses data of the memory node. The processing delay inside the processor node is reduced to 20.63ns by applying the custimized protocol for the disaggregated architecture insead of classical Ethernet protocol. The 34.3ns switching delay in the NOS consists of 12.4 ns label processing time, 3 ns switch driver delay, 6 ns switch rising time, 6 ns switch falling time and 16 ns margin time of the switch control signal with respect to the optical packet. Thus, a minimal network latency of 122.3ns is achieved accessing the data in memory nodes when no retransmission in the optical network switching. This latency can be further reduced by exploiting the Application Specific Integrated Circuit (ASIC) for the processing at the transmitter and receivers.

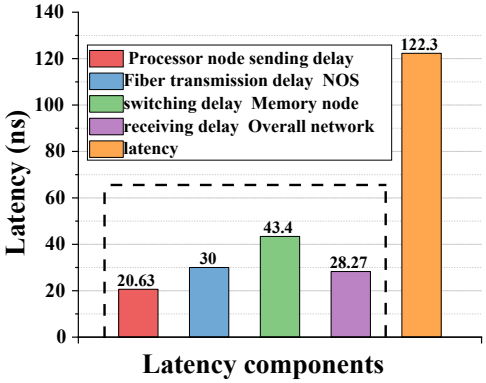

Fig. 7. Network latency components when accessing memory node.

To investigate the stability and packet loss of the disaggregated architecture prototype, the cumulative packet loss is continuously recorded over a runtime of 80 hours. The processor node sends out 1.16E12 packets per hour during the measurement. It is illustrated in Fig. 8 that the disaggregated architecture prototype achieves a cumulative packet loss rate of less than 9.5E-12 after an operation time of 80 hours. There are only 11 lost packets occurring in the initialization procedure. The reason is that the transmission receiver side (memory nodes) needs to perform the clock and data recovery (CDR). After finishing the CDR procedure, the NOS based optical network can provide a stable and no packet loss interconnection for the disaggregated hardware nodes.

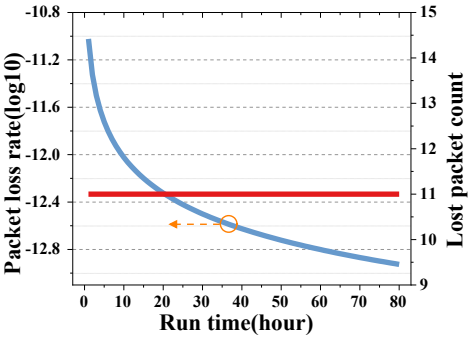

Fig. 8. Cumulative network packet loss over 80 hours.

Besides the network latency and packet loss of the data-path channel packets, signals of the control-path channel are also investigated in this section. Figure 9 depicts the signals of control-path channel among four hardware nodes, in which the packet destination of two processor nodes (P-node) is one of two memory nodes (M-node) and vice versa. Each hardware node sends request signal (Rq) consisting of the optical packet destination and priority to the switch controller

via the control-path channel. When the packet contention occurs, the packet with higher priority is forwarded, and the response signal (Rs) of ACK (request=response) is sent back to the source hardware node via the control-path channel. Meanwhile, the packet with lower priority is blocked, and the response signal of NACK (request ≠ response) is sent back. For example, in the time slot *N*+1, both the processor node 1 and 2 send the request signal with the destination memory node of 2, and the packet of processor node 1 has a higher priority. Consequently, the packet from process node 1 is forwarded, and the processor node 1 receives the response signal of ACK. The packet of processor node 2 is blocked, while the processor node 2 receives the response signal of NACK.

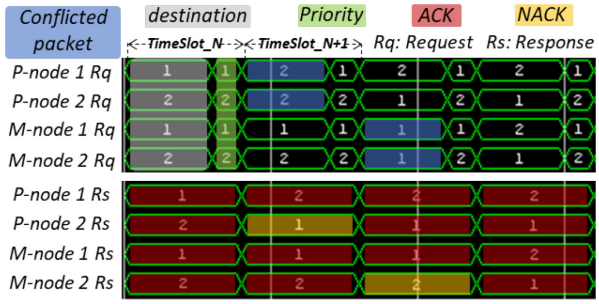

Fig. 9. Cumulative network packet loss over 80 hours.

The disaggregated architecture prototype has been experimentally validated with four hardware nodes. To better investigate the scalability of NOS based disaggregated architecture, the physical performance of NOS based disaggregated architecture are evaluated under different network scales. For the physical performance evaluation, the output OSNR and power penalty for error-free operation are experimentally assessed as a function of the NOS port count. As the NOS port count increases from 4 to 64, the amount of splitting losses in the broadcast and select architecture increases, and thus the input optical power fed into the SOA optical gates decreases (down to 18 dB extra losses for a 64x64 NOS). Thus, we investigate the NOS output OSNR under different NOS port counts (from 4 to 64 ports).

The input OSNR of NOS based interconnection network (output OSNR of hardware nodes) is measured as 62.3dB. Figure 10 reports the NOS output signal OSNR as functions of the NOS port count and SOA driving current ranging from 60mA to 150mA. The NOS output signal achieves an OSNR of 43.3dB with a port count of 4 and SOA driving current of

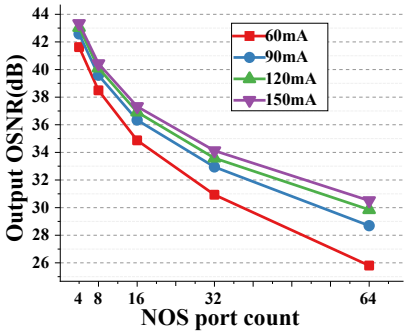

Fig. 10. NOS output OSNR with different port counts.

150mA, while an OSNR of 30.5dB at a port count of 64. Moreover, it can be observed that the SOA driving current, for the range employed in the experiment (60 mA - 150 mA), has limited impact on the output signal OSNR with a NOS port count of less than 16. Under a NOS port count of more than 16, the output signal OSNR only has a slow increase when the driving current is larger than 90mA.

Besides the output OSNR of NOS based interconnection network, the required power penalty for an error free operation at BER of 1E-9 is also experimentally assessed under different NOS port counts. We consider NOS port counts increasing from 4 to 64, and the required power penalty is illustrated in Fig. 11. It is shown that the NOS based disaggregated architecture achieves an error-free operation with power penalty of 0.9dB at BER of 1E-9 under a NOS port count of 8. When increasing the NOS port count to 64, the result indicates a slight performance degradation with power penalty of 1.5dB. This is because that a lower input power at SOA gates may degrade the performance of SOA gates and introduce more noise. Those results validate that the NOS can provide a scalable and reliable network interconnection for disaggregated DCN architectures.

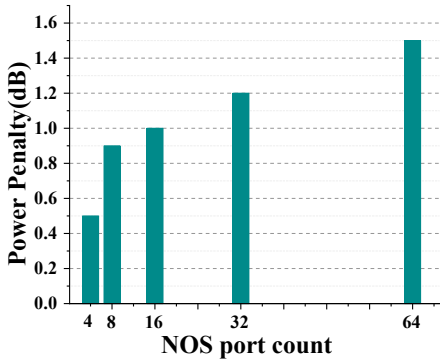

Fig. 11. Power penalty for an error-free operation under different NOS port counts.

IV. Conclusion

We have demonstrated and experimentally investigated a disaggregated DCN architecture prototype based on the NOS. Leveraging two parallel data-path and control-path channels as well as the distributed structure of the NOS, an optical network of low network latency and high bandwidth is provided for the communication among processor and memory nodes. Experimental results show that the NOS based 4x4 disaggregated architecture prototype performs an error-free operation with a power penalty of 0.5dB at BER of 1E-9. An ON/OFF switch ratio higher than 60dB is measured in the prototype, and thus the low channel cross-talk is guaranteed. In the network performance assessment, the NOS based disaggregated architecture prototype achieves a node-to-node network latency of 122.3 ns and no packet loss during the stable operation over 80 hours. The scalability of the NOS based disaggregated architecture is also assessed. Error-free operation with a power penalty of 1.5 dB at BER of 1E-9 is measured for an emulated 64-port NOS, and an output signal OSNR of 30.5dB. Numerical investigation using the experimentally measured parameters show that the NOS based disaggregated architecture requires a link bandwidth of

40Gb/s to outperform current server-centric architectures, which performs an intra-rack node-to-node latency of 148.5ns and an inter-rack latency of 265.7ns under the 4096-node network and MNAR of 0.9. For the prefetched cache line, with a link bandwidth of 40Gb/s, an intra-rack latency of 755.8ns and an inter-rack latency of 1244.9ns are achieved under the 4096-node network and MNAR of 0.9. These results validate the feasibility and scalability of the proposed NOS based disaggregated DCN architecture.

ACKNOWLEDGMENT

This work has been partially supported by the EU Horizon 2020 research and innovation programme under the ADOPTION project with grant agreement 780326 and the EU project ALLEGRO with grant agreement 101092766.

REFERENCES

[1] J. Guo et al., “Who limits the resource efficiency of my datacenter: An analysis of alibaba datacenter traces,” in Proc. 27th Int. Symp. Qual. Service, 2019, pp. 1–10.

[2] R. Lin, Y. Cheng, M. D. Andrade, L. Wosinska, and J. Chen, ‘‘Disaggregated data centers: Challenges and trade-offs,’’ IEEE Commun. Mag., vol. 58, no. 2, pp. 20–26, Feb. 2020.

[3] S. Han, N. Egi, A. Panda, S. Ratnasamy, G. Shi, and S. Shenker, “Network support for resource disaggregation in next generation datacenters,” in 12th ACM Workshop on Hot Topics in Netwworks (HotNets-XII), College Park, Maryland, 2013, article 10.

[4] C.-S. Li, H. Franke, C. Parris, B. Abali, M. Kesavan, and V. Chang, “Composable architecture for rack scale big data computing,” Futur. Gener. Comp. Syst., vol. 67, pp. 180–193, Feb. 2017.

[5] M. Bielski et al., “dReDBox: Materializing a full-stack rack-scale system prototype of a next-generation disaggregated datacenter,” in Proc. Design, Automat. Test Eur. Conf. Exhib., 2018, pp. 1093–1098.

[6] T. Alexoudi et al., “Optics in Computing: From Photonic Network-on-Chip to Chip-to-Chip Interconnects and Disintegrated Architectures,” J. Lightwave Technol. 37(2), 363–379 (2019).

[7] P. X. Gao, “Network requirements for resource disaggregation,” in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Savannah, Georgia, 2016, pp. 249–264.

[8] K. Lim, Y. Turner, J. R. Santos, A. AuYoung, J. Chang, P. Ranganathan, and T. F. Wenisch, ‘‘System-level implications of disaggregated memory,’’ in Proc. IEEE Int. Symp. High-Performance Comp Archit., LA, USA, Feb. 2012, pp. 1–12.

[9] Yizhou Shan, Yutong Huang, Yilun Chen, and Yiying Zhang, “LegoOS: A Disseminated, Distributed OS for Hardware Resource Disaggregation.” OSDI 18, USENIX Association, CA, 2018, 69–87.

[10] H. M. Mohammad Ali, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, “Future energy efficient data centers with disaggregated servers,” J. Lightw. Technol., vol. 35, no. 24, pp. 5361–5380, Dec. 2017.

[11] A. Saljoghei et al., "MCF-SMF Hybrid Low-Latency Circuit-Switched Optical Network for Disaggregated Data Centers," in J. Lightwave Technol, vol. 37, no. 16, pp. 4017-4029, Aug. 2019.

[12] X. Guo, F. Yan, J. Wang, G. Exarchakos, Y. Peng, X. Xue, B. Pan, and N. Calabretta, “RDON: a rack-scale disaggregated data center network based on a distributed fast optical switch,” J. Opt. Commun. Netw. 12(8), 251–263, 2020.

[13] F. Yan, W. Miao, O. Raz, and N. Calabretta, “OPSquare: a flat DCN architecture based on flow-controlled optical packet switches,” J. Opt. Commun. Netw. 9, 291–303, 2017.

[14] W. Miao, J. Luo, S. D. Lucente, H. Dorren, and N. Calabretta, “Novel flat datacenter network architecture based on scalable and flow controlled optical switch system,” Opt. Express 22, 2465–2472, 2014.

[15] X. Xue, F. Yan, K. Prifti, F. Wang, B. Pan, X. Guo, S. Zhang, and N. Calabretta, "ROTOS: A Reconfigurable and Cost-Effective Architecture for High-Performance Optical Data Center Networks," in J. Lightwave Technol, vol. 38, no. 13, pp. 3485-3494, 2020.

[16] U. Drepper, “What every programmer should know about memory,” Technical report, Red Hat, Inc., 2008.

[17] [Online.] Available: https://www.xilinx.com/products/silicon-devices/fpga/virtex-ultrascale.html

[18] [Online.] Available: https://www.xilinx.com/products/boards-and-kits/dk-v7-vc709-g.html

[19] [Online.] Available: https://www.xilinx.com/products/intellectual-property/ibert_ultrascale_gth.html

[20] Hennessy, J.L. and Patterson, D.A., “Computer architecture: a quantitative approach”, 6th edition, Elsevier, 2018.

[21] A. K. Shah and M. D. Matani, “An O(1) algorithm for implementing the lfu cache eviction scheme,” in Technical report, 2010. [Online]. Available: http: //dhruvbird.com/lfu.pdf