Dual-Layer Locality-Aware Optical Interconnection Architecture for Latency-Critical Resource Disaggregation Environments

Nikos Terzenidis, Miltiadis Moralis-Pegios, Theoni Alexoudi, Stelios Pitris, Konstantinos Vyrsokinos, Nikos Pleros

Dual-layer locality-aware optical interconnection architecture for latency-critical resource disaggregation environments

Nikos Terzenidis[1[0000-0001-8180-3953]], Miltiadis Moralis-Pegios[1[0000-0002-9401-730X]], Theonitsa Alexoudi[1[0000-0001-6722-201X]], Stelios Pitris[1[0000-0001-5010-8843]], Konstantinos Vyrsokinos[2] and Nikos Pleros[1[0000-0003-2931-4540]]

[1] Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
[2] Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
nterzeni@csd.auth.gr

Abstract. Significant research efforts, both industrial and academic, have been committed in the direction of Rack-scale computing through resource disaggregation, that aims to increase resource utilization at a reduced energy and cost envelope. However, the realization of resource disaggregation necessitates an underlying network infrastructure that can compete with a challenging set of requirements including low-latency performance and high-port count connectivity, as well as high data-rate operation. At the same time, it is crucial for the interconnection architecture to be able to accommodate efficient delivery of traffic with different locality characteristics. We propose a dual-layer locality-aware optical interconnection architecture for disaggregated Data Centers by combining the STREAMS silicon-based on-board communication paradigm with the disaggregation-oriented Hipoλaos high-port count switch. Simulation evaluation of a 256-node disaggregated system, comprising 32 optically-interconnected 8-socket boards, revealed up to 100% throughput and mean, p99 latencies not higher than 335nsec and 610nsec, respectively, when a 50:50 ratio between on- and off-board traffic is employed. Evaluation of the same layout with 75:25 on-/off-board traffic yields even lower mean and p99 latency at 210ns and 553ns, respectively.

Keywords: Silicon-photonics, Optical switch, Interconnection architecture, Disaggregated data center, Traffic locality.

1 Introduction

The ever-increasing energy consumption of Data Centers (DCs), projected to rise to 3% of the global electricity demand by 2020 [1], along with the huge waste of resources, that is observed in traditional DCs and may reach up to 50% [2]-[3], have forced DC operators to invest in solutions that will considerably improve energy efficiency. In this context, resource disaggregation in computing and network architectures is under heavy research [4]-[6], as a groundbreaking innovation that could amortize the energy and cost impact caused by the vast diversity in resource demand of emerging DC workloads

[7]. This architectural shift breaks the tight co-integration of CPU, memory and storage resources from a single server board, towards disaggregated systems with multiple resources organized in trays and synergized via the DC network. The first promising results of a full-fledged CPU-memory disaggregated prototype, from the EU-funded project dRedBox, suggest an important decrease in Total Cost of Ownership (TCO) [6], while the deployment of partially-disaggregated servers in Intel's DCs contributed to an impressive power usage effectiveness (PuE) rating of 1.06 [8].

However, resource disaggregation, by breaking apart the critical CPU-to-memory path, introduces a challenging set of requirements in the underlying network infrastructure [9], that has to support low latency and high throughput communication, while providing connectivity to an increased number of nodes. At the same time, it is crucial for the interconnection architecture to be able to accommodate efficient unicast and multicast traffic delivery with different locality characteristics. To this end, recent studies [10]-[12] have indicated heavy traffic exchange within the boundaries of a Rack, that can be observed through a variety of emerging DC workloads, while a number of applications span their communication capacity through the entire network hierarchy [10], requiring all-to-all connectivity. The transition to a disaggregation paradigm endorses this type of mixed communication profile and effectively narrows down the locality pattern to board-level, where for example computing resources are synergized in homogeneous pools exhibiting highly localized traffic, while requiring also connectivity with the memory or storage pools located in other trays of the system.

During the first demonstrations of disaggregated systems [6] optical circuit switches (OCS) have been employed to provide rack-level connectivity between resource bricks residing in different trays, due to their high-radix, scaling to hundreds of ports, along with their datarate-transparent operation. However, currently available OCS solutions based either on 3D Micro Electromechanical Systems (MEMS) [13]-[14] or piezoelectric beam steering technology [15] come at the cost of ms switching time values, that effectively limit their employment as slow reconfigurable backplanes [16]. To this end, we have recently demonstrated the Hipoλaos optical packet switch (OPS) architecture [17]-[21], that provides low-latency connectivity between a high number of ports, while at the same time enabling dynamic switching operation with packet granularity. Moreover, Hipoλaos OPS offers up-to 95% throughput for unicast traffic when interconnecting 1024 nodes located in 32 different trays [19], enabling this way efficient communication through the whole Rack hierarchy.

Moving on to the next DC hierarchy layer, the tray, electrical switching solutions have been employed in the first disaggregated prototypes to provide on-board connectivity, following the established paradigm of interconnection architectures for multi-socket boards (MSB), that are currently dominated by electrical point-to-point (P2P) links. While P2P interconnects, like Intel QPI [22] and AMD Infinity fabric, can definitely offer low latency and high bandwidth communication, their scalability is typically limited to 4 endpoints, before requiring extra latency-inducing hops or active switch-based solutions like Oracle Bixby [23] or PCIe switches. Recent advances on board-compatible optical interfaces highlight the significant throughput and delay improvements that can be released through optically-interconnected blade technologies [24], paving the way for their employment in high-density MSB scenarios. To this end,

we have demonstrated, in the context of STREAMS project [25], a Silicon-photonics-based interconnection scheme for up to 40 Gb/s chip-to-chip communication [26] using an integrated Ring-Modulator (RM) transmitter (Tx), an 8×8 Arrayed Waveguide Grating Router (AWGR)-based routing circuitry and a co-packaged InP-based photodiode (PD) and Transimpedance Amplifier (TIA) receiver (Rx). The STREAMS architecture is able to provide high-bandwidth communication to 8 on-board nodes, while the AWGR-based routing ensures ultra-low latency and efficient multicast/broadcast traffic delivery encompassing the requirements of coherency-induced multi- and broadcasting traffic patterns.

In this paper we combine the STREAMS silicon-based on-board interconnection architecture with the disaggregation-oriented Hipoλaos high-port count switch, in a dual-layer locality-aware Rack interconnection scheme that efficiently accommodates the need for transparent on-board local traffic forwarding in a disaggregated environment with hundreds of nodes. We evaluate via an Omnet++ simulation analysis, this novel rack configuration with 256 disaggregated nodes using a number of 32 optically-interconnected 8-socket MSBs. This 256-socket setup can take advantage of traffic localization techniques towards low-latency workload execution, forming a powerful disaggregated rack-scale computing scheme with mean and p99 latencies not higher than 335nsec and 610nsec, respectively, when a 50:50 ratio between on- and off-board traffic is employed. Finally, evaluation of the same system with 75:25 on-:off-board traffic yields even lower mean and p99 latency at 210ns and 553ns, respectively.

2 Hipoλaos and STREAMS Optical Interconnects

2.1 Hipoλaos Rack-level optical switch architecture

Hipoλaos (High port λ routed all optical switch) manages to combine low latency and high-throughput performance into high-port count configurations, through a hybrid scheme exploiting Spanke switching and wavelength routing. The Hipoλaos switch differentiates from alternative OPS layouts by incorporating the multi-lambda routing capabilities of AWGRs, along with careful optimizations in the different stages of the Spanke architecture, that can be classified into four main categories/features:

(i) Distributed control: The Hipoλaos architecture overcomes the scalability limitations of Spanke designs, towards high-port count layouts, by distributing the control and switching functions in small clusters, named as Planes. Moreover, this distributed nature of the architecture minimizes the computational time associated with forwarding and scheduling, contributing this way to lower total latency of the switch fabric.

(ii) Feed-forward buffering: The adoption of small-scale optical buffering [27] enables the realization of high-throughput performance while at the same time reducing latency by avoiding optoelectronic buffering and the associated Optical/Electrical/Optical conversions.

(iii) Advanced Wavelength Conversion schemes: Wavelength converters (WC) on the different stages of the Spanke layout utilize the differentially biasing scheme [28],

that has been shown to operate successfully with up to 40Gb/s NRZ signals, fulfilling the requirement for high datarates.

(iv) Multi-wavelength routing: The architecture utilizes the cyclic routing properties of AWGRs in order to extend the switch radix through a collision-less WDM routing mechanism, that ensures non-blocking forwarding of the different packets to the desired destination node. Moreover, it enables the realization of multicast functionality building upon the proven efficiency of AWGR devices in multicast operations [29].

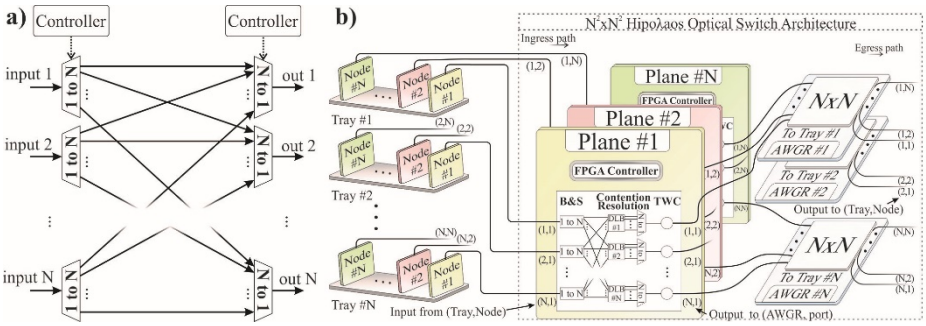

Fig. 1. (a) N×N Spanke switch architecture, (b) Schematic illustration of a N^2-node DC, organized in #N Rack-Trays with #N nodes per tray and interconnected via a $N^2 \times N^2$ ToR Hipoλaos switch

Fig. 1(b) depicts a schematic illustration of the Hipoλaos OPS architecture, configured in a N2×N2 layout, interconnecting #N Rack trays with #N computing nodes in every tray. The switch is organized in #N Planes that are interconnected to #N AWGR devices. On the ingress path, every node communicates with a specific Plane over a different fiber link, while the Plane's role is to aggregate traffic from #N input ports and forward it via a BS scheme, with every 1:N splitting node corresponding to the 1:N switches employed in the Spanke architecture, as depicted in Fig. 1(a). The N:1 switches at the outputs of the Spanke design are modified to incorporate an Optical Delay-Line Bank (DLB), forming this way an optical contention resolution stage comprising #N DLBs. Packets leaving the contention resolution stage are finally delivered to the destination node after passing through a wavelength routing stage comprising tunable WCs and cyclic routing AWGRs.

The feasibility of the Hipoλaos architecture has been validated through a prototype experimental evaluation following the design principles of the 256-port layout [17], while its network performance characteristics have been evaluated through a variety of unicast traffic scenarios [18] revealing a high throughput value of >85%, for up to 100% loads, along with 605nsec latency, even for 2 packet-size optical buffers. An actual photo of the experimental prototype is depicted in the right inset of Fig. 3. A scalability analysis of the Hipoλaos architecture, considering also the main limiting factors, has been presented in [18], concluding to an attainable port-count of at least 1024-ports, with the experimental evaluation in unicast and multicast operational mode presented in [19].

2.2 STREAMS silicon-photonics board-level interconnect

The European H2020 project ICT-STREAMS is currently attempting to deploy the necessary silicon photonic and electro-optical PCB technology toolkit for realizing the AWGR-based MSB interconnect benefits in the O-band and at data rates up to 50Gb/s [25],[30]. It aims to exploit wavelength division multiplexing (WDM) Silicon photonics transceiver technology at the chip edge as the socket interface and a board-pluggable O-band silicon-based AWGR as the passive routing element, as shown in a generic N-socket architecture depicted in Fig. 2. Each socket is electrically connected to a WDM-enabled Tx optical engine equipped with N-1 laser diodes (LD), each one operating at a different wavelength. Every LD feeds a different Ring Modulator (RM) to imprint the electrical data sent from the socket to each one of the N-1 wavelengths, so that the Tx engine comprises finally N-1 RMs along with their respective RM drivers (DR). All RMs are implemented on the same optical bus to produce the WDM-encoded data stream of each socket. The data stream generated by each socket enters the input port of the AWGR and is forwarded to the respective destination output that is dictated by the carrier wavelength and the cyclic-frequency routing properties of the AWGR [31]. In this way, every socket can forward data to any of the remaining 7 sockets by simply modulating its electrical data onto a different wavelength via the respective RM, allowing direct single-hop communication between all sockets through passive wavelength-routing. At every Rx engine, the incoming WDM-encoded data stream gets demultiplexed with a 1:(N-1) optical demultiplexer (DEMUX), so that every wavelength is received by a distinct PD. Each PD is connected to a transimpedance amplifier (TIA) that provides the socket with the respective electrical signaling.

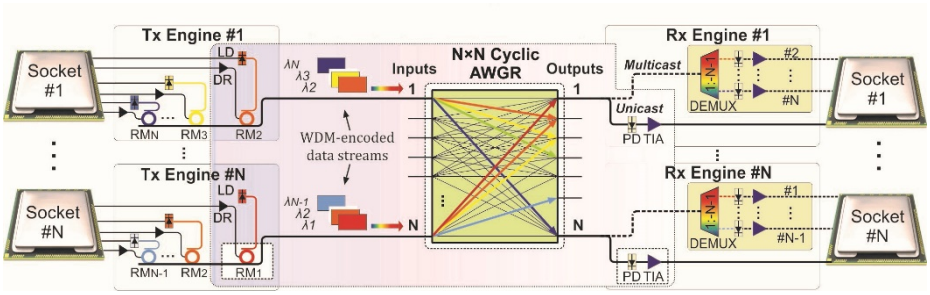

Fig. 2. STREAMS optical N×N AWGR-based interconnection concept for MSB connectivity exploiting WDM-enabled Tx/Rx engines.

This flat-topology AWGR-based interconnect scheme requires a higher number of transceivers compared to any intermediate switch solution, but this is exactly the feature that allows to combine WDM with AWGR's cyclic frequency characteristics towards enabling single-hop communication and retaining the lowest possible latency. Link capacity can be increased in this case by residing on channel bonding through bit-parallel schemes, as already reported in [32], by using AWGR designs for waveband instead of single wavelength routing. Utilizing an 8×8 AWGR, the optically-enabled MSB can allow single-hop all-to-all interconnection between 8 sockets, while scaling the AWGR

to 16×16 layouts can yield single-hop communication even between 16 sockets, effectively turning current "glued" into "glueless" designs. The ICT-STREAMS on-board MSB aims to incorporate 50GHz single-mode O-band electro-optical PCBs [33], relying on the adiabatic coupling approach between silicon and polymer waveguides [34] for low-loss interfacing of the Silicon-Photonics transceiver and AWGR chips with the EO-PCB.

The first 40Gb/s experimental demonstration of the fiber-interconnected integrated photonic building blocks when performing in the AWGR-based 8-socket MSB architecture, presenting the 40Gb/s experimental results have been reported in [26]. The energy efficiency of the proposed 40 Gb/s C2C photonic link is estimated at 24 pJ/bit, but can dramatically go down to 5.95 pJ/bit when transferring the demonstrated fiber-pigtailed layout into an on-board assembled configuration and assuming a 10% wall-plug efficiency for the external laser. This indicates that the on-board version has the credentials to lead to 63.3% reduction in energy compared to the 16.2 pJ/bit link energy efficiency of Intel QPI [35].

3 Dual-layer Locality-aware optical interconnection architecture

Fig. 3 presents a schematic illustration of a 256-node DC system comprising 32 optically-enabled STREAMS boards, with every board incorporating 8 network nodes. Every node in the proposed dual-layer network hierarchy is connected via different optical links to an on-board 8x8 AWGR, serving as the intra-board routing infrastructure, as well as to a Hipoλaos-based 256 port switch, providing inter-board all-to-all connectivity. The internal node architecture is depicted in the left inset of Fig. 3, where a CPU, for example, can communicate with any of the remaining 7 nodes of the board by utilizing links #1 to #7 that effectively forward data via the underlying silicon photonics Tx engine. The WDM-encoded data stream, comprising seven lambdas, is forwarded to the on-board AWGR device where every wavelength channel is finally delivered to a different end node. This first layer of switching can ensure maximum throughput of on-board traffic, being in agreement with the requirement for transparent localized traffic forwarding, while the latency associated with header processing and scheduling is eliminated, since the routing decision is performed in the source node by simply selecting the appropriate link(s). The second layer in the switching topology can be accessed through link #8, that forwards inter-board traffic, via a fixed-wavelength optical data stream, to the Hipoλaos switch. The internal architecture of the 256-port Hipoλaos layout, that has been described in detail in [17], comprises 16 switch Planes with every Plane aggregating traffic from 16 nodes. Due to the mismatch between the number of nodes per board and the number of ports per Plane, the input port allocation per switch plane is performed so that Node#i, $1 \leq i \leq 8$, from the odd-numbered boards#j, j=1,3,..,31, connects to the input#k, k=(j+1)/2, of Plane#l, l=i, denoted as input (l,k) of the switch. Moreover, Nodes#i, $1 \leq i \leq 8$, from the even-numbered boards#j, j=2,4,..,32, connect to the input#k, k=j/2, of Plane#l, l=8+i. The proposed port-allocation scheme groups into the respective contention resolution stages of the switch packets from 2

adjacent boards, that in conjunction with the on-board switching layer ensures minimum contention between the different packets. It should be noted that every node can accommodate either computing resources, as illustrated on the example of Fig. 3, as well as memory, storage and accelerator resources.

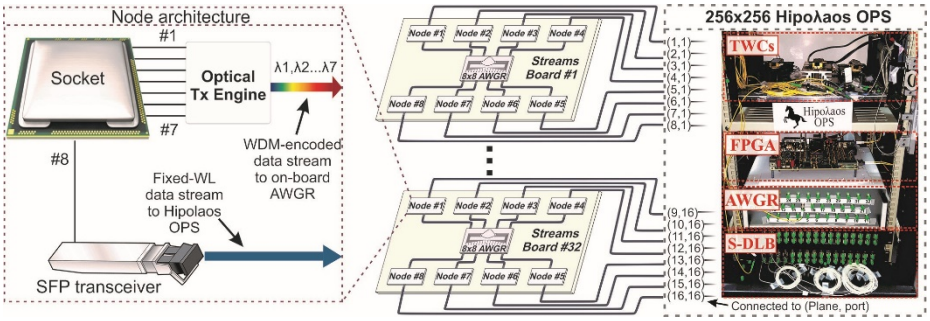

Fig. 3. Illustration of a locality-aware Rack interconnection scheme employing 32 Streams boards, with 8 nodes each, interconnected to a 256x256 Hipoλaos switch. On the left inset, the internal node architecture is presented, while on the right an actual photo of the Hipoλaos experimental prototype is presented.

Having already demonstrated the performance credentials of each constituent interconnection scheme in [17] and [18], we have proceeded with the evaluation of the network throughput and latency performance in a DC system incorporating the proposed dual-layer interconnection architecture. A simulation analysis has been performed for the 256-node system illustrated in Fig. 3, using the Omnet++ discrete event platform. A synchronous slotted network operation has been modelled where the packets are generated at predefined packet-slots, each one lasting for 57.6ns. The traffic profile was customized to distribute a certain percentage of the total traffic generated by every node, uniformly to nodes of the same board (intra-board traffic), while the rest of the traffic was uniformly distributed to nodes from the other 31 boards (inter-board traffic). In order to offer a thorough evaluation of the architecture's performance, in terms of latency, both mean packet delay, as well as p99 delay metrics were collected by the simulation. The switch port allocation to the network nodes was performed according to the procedure described in the previous paragraph and illustrated in the layout of Fig. 3.

The modelled DC system featured node-to-switch and node-to-AWGR channel datarate of 10 Gb/s, along with fixed size packet-length of 72 bytes, comprising 8-bytes for header, synchronization and guard-band requirements and 64 bytes data payload, matching the size of a single cache-line transfer. Regarding the Hipoλaos processing latency, it was modelled to 456ns in accordance with the experimental results [17], while the propagation latency for the various optical components of the switch (fibers, amplifiers, AWGRs), excluding the optical delay lines, was modelled to be 35ns.

In order to perform a versatile evaluation of the proposed architecture under different traffic locality patterns, we have considered in our analysis two different cases for the percentage of the intra-/inter-board traffic; 50/50 and 75/25. Performance has been

evaluated as a function of the available packet-buffers in the Hipoλaos switch, with the number of buffers ranging for 0 to 4 and corresponding to the maximum number of buffers experimentally demonstrated in [18].

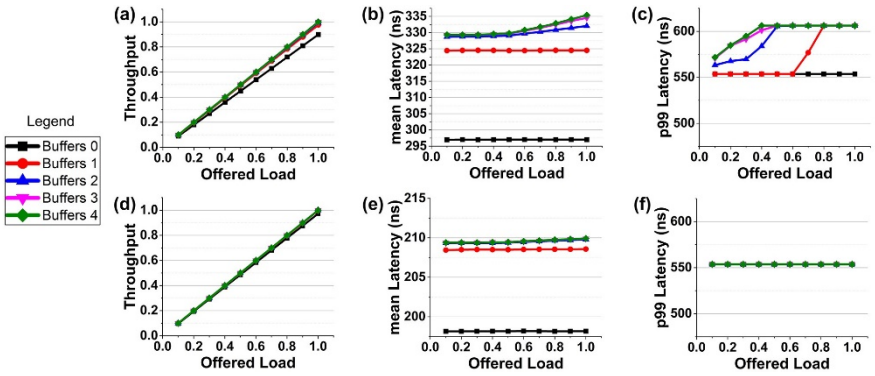

Fig. 4. Simulation results for different number of buffers per DLB (a) Throughput – 50/50, (b) Mean latency – 50/50, (c) P99 latency – 50/50, (d) Throughput – 75/25, (e) Mean latency – 75/25, (f) P99 latency – 75/25

Fig. 4(a) to (c) present the simulation results for the case of 50/50 intra-/inter-board traffic distribution. Figure 4(a) presents the respective throughput versus the offered load results, concerning the total network traffic (both intra- and inter-board) for different numbers of buffers per Hipoλaos DLB. As expected, throughput increases almost linearly with increasing buffer size, approaching 100% for 100% offered load as long as the buffer size equals to more than 2 packet slots. Figure 4(b) presents the mean packet delay versus the offered load results, showing that latency ranges between 295ns and 330ns for a buffer size between 0 and 2 packet slots and for loads until 100%. Mean latency increases slightly as the buffer size increases reaching a maximum value of 335ns for a buffer size of 4 packet slots where throughput reaches also its maximum value. Figure 4(c) presents the p99 delay results vs. the offered load, revealing a p99 value of 610ns for maximum load and 4 buffers per DLB. As can be observed, the p99 delay metrics perform a step-wise "jump" as contention occurs due to the fact that packets are forwarded to longer DLB buffers that introduce delays in packet duration granularity. It is important to note that the only point of congestion in the architecture was identified at the Hipoλaos switch, since the intra-board AWGR switching scheme is able to offer 100% throughput and minimum latency values originating just from the signal's propagation delay.

Fig. 4(d) to (f) present the simulation results for the case of 75/25 intra-/inter-board traffic distribution. As expected, throughput is slightly higher, reaching 100% in all cases, due to the fact that a lower percentage of traffic is headed towards the Hipoλaos switch, where congestion occurs. At the same time mean packet latency is decreased to 210ns, while p99 latency now presents a maximum value of 550ns, as less delay lines have to be utilized in the Hipoλaos DLB blocks.

The vital information that can be easily extracted from Fig. 4(b)-(c) and Fig. 4(e)-(f), is that the proposed dual-layer switching topology does not induce any additional latency compared to the conventional Hipoλaos architecture [17], but rather decreases the mean and p99 packet latency, due to the fact that the 2 switching layers operate in a parallel and functionally isolated manner. With sub-μsec latency considered as the main performance target for current memory disaggregated systems [9], the mean and p99 latency values of this novel Hipoλaos-STREAMS-based architecture with clustered optically-enabled 8-Socket MSBs reveals an excellent potential for a practical interconnect solution that can bring latency down to just a few 100's of nanosecond. Allowing on-board nodes to cluster in single-hop configurations over AWGR-based interconnects can yield minimized latency when combined with proper workload allocation for strengthening board-level traffic localization, while off-board traffic benefits from the latency-optimized dynamic switch characteristics of the Hipoλaos design.

4 Conclusion

The emergence of resource disaggregation in DC architectures is imposing stringent requirements on the interconnection architecture that has to support low-latency, high-throughput and high-radix connectivity, while at the same time accommodating efficient delivery of traffic with different locality characteristics. To this end we have demonstrated a dual-layer locality-aware optical interconnection architecture by combining the ICT-STREAMS silicon-photonics on-board communication paradigm with the intra-DC Hipoλaos high-port count switch. Simulation analysis of a 256-node disaggregated system, comprising 32 optically-interconnected 8-socket boards, revealed up to 100% throughput and mean, p99 latencies not higher than 335nsec and 610nsec, respectively, when a 50:50 ratio between on- and off-board traffic is employed. Finally, this layout could in principle form the basis for replacing the massive QPI "island" interconnection supported by a number of switch technologies like Bixby's [23] and PCI express, yielding a powerful network of cache-coherent islands at a maximum p99 latency value just above 600nsec even when a balanced 50/50 traffic locality pattern is followed.

Acknowledgments. This work has been partially supported by the European H2020 projects ICT-STREAMS (Contract No. 688172) and L3MATRIX (Contract No. 688544).

References

1. Jones, N.: How to stop data centres from gobbling up the world's electricity. Nature. 561, 163-166 (2018).
2. Di, S., Kondo, D., Cappello, F.: Characterizing Cloud Applications on a Google Data Center. 2013 42nd International Conference on Parallel Processing. (2013).

3. Reiss, C., Tumanov, A., Ganger, G., Katz, R., Kozuch, M.: Heterogeneity and dynamicity of clouds at scale. Proceedings of the Third ACM Symposium on Cloud Computing - SoCC '12. (2012).
4. Intel® Rack Scale Design, http://www.intel.com/content/www/us/en/architecture-and-technology/rack-scale-design-overview.html.
5. Open Compute Project. The Open Compute server architecture specications, http://www.opencompute.org.
6. Bielski, M., Syrigos, I., Katrinis, K., Syrivelis, D., Reale, A., Theodoropoulos, D., Alachiotis, N., Pnevmatikatos, D., Pap, E., Zervas, G., Mishra, V., Saljoghei, A., Rigo, A., Zazo, J., Lopez-Buedo, S., Torrents, M., Zyulkyarov, F., Enrico, M., de Dios, O.: dReDBox: Materializing a full-stack rack-scale system prototype of a next-generation disaggregated datacenter. 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE). (2018).
7. Tencent Explores Datacenter Resource Pooling Using Intel® Rack Scale Architecture (Intel® RSA), https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/rsa-tencent-paper.pdf.
8. Disaggregated Servers Drive Data Center Efficiency and Innovation, https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/disaggregated-server-architecture-drives-data-center-efficiency-paper.html.
9. Gao, P. X., Narayan, A., Karandikar, S., Carreira, J., Han, S., Agarwal, R., Ratnasamy, S., Shenker, S.: Network requirements for resource disaggregation. 12th {USENIX} Symposium on Operating Systems Design and Implementation (OSDI). (2016).
10. Roy, A., Zeng, H., Bagga, J., Porter, G., Snoeren, A.: Inside the Social Network's (Datacenter) Network. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication - SIGCOMM '15. (2015).
11. Delimitrou, C., Sankar, S., Kansal, A., Kozyrakis, C.: ECHO: Recreating network traffic maps for datacenters with tens of thousands of servers. 2012 IEEE International Symposium on Workload Characterization (IISWC). (2012).
12. Kandula, S., Sengupta, S., Greenberg, A., Patel, P., Chaiken, R.: The nature of data center traffic. Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference - IMC '09. (2009).
13. S Series Optical Circuit Switch | CALIENT Technologies, https://www.calient.net/products/s-series-photonic-switch/.
14. Glimmerglass Intelligent Optical Systems | Glimmerglass, http://www.glimmerglass.com/index.php/products/intelligent-optical-systems/.
15. Polatis SERIES 7000 - 384x384 port Software-Defined Optical Circuit Switch, https://www.polatis.com/series-7000-384x384-port-software-controlled-optical-circuit-switch-sdn-enabled.asp?.
16. Chen, Q., Mishra, V., Parsons, N., Zervas, G.: Hardware Programmable Network Function Service Chain on Optical Rack-Scale Data Centers. Optical Fiber Communication Conference. (2017).
17. Terzenidis, N., Moralis-Pegios, M., Mourgias-Alexandris, G., Vyrsokinos, K., Pleros, N.: High-port low-latency optical switch architecture with optical feed-forward buffering for 256-node disaggregated data centers. Optics Express. 26, 8756 (2018).
18. Terzenidis, N., Moralis-Pegios, M., Mourgias-Alexandris, G., Alexoudi, T., Vyrsokinos, K., Pleros, N.: High-Port and Low-Latency Optical Switches for Disaggregated Data Centers: The Hipoλaos Switch Architecture [Invited]. Journal of Optical Communications and Networking. 10, B102 (2018).

19. Moralis-Pegios, M., Terzenidis, N., Mourgias-Alexandris, G., Vyrsokinos, K., Pleros, N.: A 1024-Port Optical Uni- and Multicast Packet Switch Fabric. Journal of Lightwave Technology. 37, 1415-1423 (2019).
20. Moralis-Pegios, M., Terzenidis, N., Mourgias-Alexandris, G., Cherchi, M., Harjanne, M., Aalto, T., Miliou, A., Vyrsokinos, K., Pleros, N.: Multicast-Enabling Optical Switch Design Employing Si Buffering and Routing Elements. IEEE Photonics Technology Letters. 30, 712-715 (2018).
21. Terzenidis, N., Moralis-Pegios, M., Mourgias-Alexandris, G., Vyrsokinos, K., Pleros, N.: Multicasting in a High-Port Sub-μsec Latency Hipoλaos Optical Packet Switch. IEEE Photonics Technology Letters. 30, 1535-1538 (2018).
22. An Introduction to the Intel® QuickPath Interconnect, https://www.intel.com/content/www/us/en/io/quickpath-technology/quick-path-interconnect-introduction-paper.html.
23. Wicki, T., Schulz, J.: Bixby: The scalability and coherence directory ASIC in Oracle's highly scalable enterprise systems. 2013 IEEE Hot Chips 25 Symposium (HCS). (2013).
24. Maniotis, P., Terzenidis, N., Siokis, A., Christodoulopoulos, K., Varvarigos, E., Immo-nen, M., Yan, H., Zhu, L., Hasharoni, K., Pitwon, R., Wang, K., Pleros, N.: Applica-tion-Oriented On-Board Optical Technologies for HPCs. Journal of Lightwave Tech-nology. 35, 3197-3213 (2017).
25. Kanellos, G., Pleros, N.: WDM mid-board optics for chip-to-chip wavelength routing interconnects in the H2020 ICT-STREAMS. Optical Interconnects XVII. (2017).
26. Pitris, S., Moralis-Pegios, M., Alexoudi, T., Lambrecht, J., Yin, X., Bauwelinck, J., Ban, Y., De Heyn, P., Pantouvaki, M., Van Campenhout, J., Broeke, R., Pleros, N.: A 40 Gb/s Chip-to-Chip Interconnect for 8-Socket Direct Connectivity Using Integrated Photonics. IEEE Photonics Journal. 10, 1-8 (2018).
27. Moralis-Pegios, M., Mourgias-Alexandris, G., Terzenidis, N., Cherchi, M., Harjanne, M., Aalto, T., Miliou, A., Pleros, N., Vyrsokinos, K.: On-Chip SOI Delay Line Bank for Optical Buffers and Time Slot Interchangers. IEEE Photonics Technology Letters. 30, 31-34 (2018).
28. Spyropoulou, M., Pleros, N., Vyrsokinos, K., Apostolopoulos, D., Bougioukos, M., Petrantonakis, D., Miliou, A., Avramopoulos, H.: 40 Gb/s NRZ Wavelength Conversion Using a Differentially-Biased SOA-MZI: Theory and Experiment. Journal of Lightwave Technology. 29, 1489-1499 (2011).
29. Pitris, S., Mitsolidou, C., Alexoudi, T., Pérez-Galacho, D., Vivien, L., Baudot, C., De Heyn, P., Van Campenhout, J., Marris-Morini, D., Pleros, N.: O-band Energy-efficient Broadcast-friendly Interconnection Scheme with SiPho Mach-Zehnder Modulator (MZM) & Arrayed Waveguide Grating Router (AWGR). Optical Fiber Communication Conference. (2018).
30. Alexoudi, T., Terzenidis, N., Pitris, S., Moralis-Pegios, M., Maniotis, P., Vagionas, C., Mitsolidou, C., Mourgias-Alexandris, G., Kanellos, G., Miliou, A., Vyrsokinos, K., Ple-ros, N.: Optics in Computing: From Photonic Network-on-Chip to Chip-to-Chip Inter-connects and Disintegrated Architectures. Journal of Lightwave Technology. 37, 363-379 (2019).
31. Leijtens, X., Kuhlow, B., Smit, M.: Arrayed Waveguide Gratings. Springer Series in Optical Sciences. 125-187.
32. Grani, P., Liu, G., Proietti, R., Ben Yoo, S.: Bit-Parallel All-to-All and Flexible AWGR-based Optical Interconnects. Optical Fiber Communication Conference. (2017).
33. Lamprecht, T., Betschon, F., Lambrecht, J., Ramon, H., Yin, X., Bruderer, A., Premerlani, R.: EOCB-Platform for Integrated Photonic Chips Direct-on-Board Assembly within Tb/s Applications. 2018 IEEE 68th Electronic Components and Technology Conference (ECTC). (2018).

34. Dangel, R., La Porta, A., Jubin, D., Horst, F., Meier, N., Seifried, M., Offrein, B.: Polymer Waveguides Enabling Scalable Low-Loss Adiabatic Optical Coupling for Silicon Photonics. IEEE Journal of Selected Topics in Quantum Electronics. 24, 1-11 (2018).
35. Gough, C., Steiner, I., Saunders, W.: Energy efficient servers: blueprints for data center optimization. Apress (2015).