# RSON: an Inter/Intra-Chip Silicon Photonic Network for Rack-Scale Computing Systems

Peng Yang[1], Zhengbin Pang[2], Zhifei Wang[1],
Zhehui Wang[1], Min Xie[2], Xuanqi Chen[1], Luan H. K. Duong[1], Jiang Xu[1]
[1]The Hong Kong University of Science and Technology, Hong Kong
[2]National University of Defense Technology, China

***Abstract*—The increasing demand for more computational power from scientific computing, big data processing, and machine learning is pushing the development of HPC (high-performance computing) systems. As the basic HPC building blocks, modularized server racks with a large number of multicore nodes are facing performance and energy efficiency challenges. This paper proposes RSON, an optical network for rack-scale computing systems. RSON connects processor cores, caches, local memories, and remote memories through a novel inter/intra-chip silicon photonic network architecture. We develop a low-latency scalable channel partition and low-power dynamic path priority control scheme for RSON. Experimental results show that RSON can help rack-scale computing systems achieve up to 6.8X higher performance under the same energy consumption than state-of-the-art systems under the latest APEX (application performance at extreme scale) benchmarks.**

## I. Introduction

Scientific computation, big data, and deep learning applications demand higher and higher performance and energy-efficient exascale computing systems. However, it is widely acknowledged that the way towards exascale systems will be significantly more challenging. Tens of thousands of multicore processors are deployed to undertake large scale applications. To carry out distributed workloads, these chips are then organized in racks, forming cloud computing and HPC infrastructures. The power consumed by this interconnection network is expected to grow correspondingly and can contribute almost one-quarter of overall system power consumption [1]. This energy consumption requirement for exascale system is difficult to be met with the traditional electrical links because of its high power consumption and limited data rate for long distance transmission. Thus, optical interconnect becomes a promising candidate to address the performance and energy challenges for exascale systems because of its ultra-high bandwidth and low energy consumption by CMOS-III-V process in [2].

Rack-scale optical interconnect architecture, therefore, plays an important role in approaching an efficient exascale system by deploying many multicore chips/nodes. [3], [4] targeted on hybrid optical-electrical on-chip interconnects. All optical intra-chip networks, which were studied in [5], [6], [7], removing the power-hungry and latency overhead of O/E/O conversion on the intermediate paths. Optical channel efficiency was also investigated in [8]. The concept of off-chip optical interconnects were investigated by [9], [10]. Although a few works put forward inter/intra-chip networks [11], they did not present details on how to connect these two parts and how they interact with each other. In this paper, we propose RSON, a rack-scale computing system, to consider both inter-chip and intra-chip (exchangeably as ONoC, Optical Network-on-Chip) networks and to provide balanced interconnection for on-chip and off-chip traffic. The integrated inter-node interface is proposed to show how to connect the inter-chip and intra-chip network efficiently with low loss and latency overhead.

RSON targets high performance and efficient rack-scale computing systems equipped with silicon photonic network. The inter-chip network is built by a high-radix optical switch [12] and optical fibers connecting the server nodes. An optical Network-on-Chip (ONoC) is proposed to connect the memory controllers and the inter-node interface. Optical interconnected systems have been extensively studied. For example, in [13], [14], the non-local data access and processor-to-DRAM network were devised for efficient off-chip data access. A scalable photonic interconnection network architecture for data center applications was proposed in [15]. In [16], the authors presented an ASIC switch that is able to sustain 8.2TB/s for rack-scale communication. In [17], the authors experimentally demonstrated a 75 ns end-to-end latency in an optical top-of-rack switch. Wavelength routing characteristics of a switch fabric was adopted by [18] to resolve the switch control based on Arrayed Waveguide Grating Router. Most of these works paid less attention to the optical data channel assignment and control.

RSON adopts circuit switching for the inter-chip and ONoC because of the relatively high overhead on optical path setup/teardown and the difficulty on buffering optical signals. Although centralized optical networks controller was designed [19], it is not suitable for large scale systems, which involves both intra-chip and inter-chip optical networks. Circuit switching requires path reservation in multiple domains, such as server nodes' ONoC and the optical switch. The reservations among different domains could incur significant performance overhead due to the idle cycles on reserved resources. We utilize the inter-node interface as the medium to coordinate the request from both local ONoC and optical switch. This greatly decreases the negotiation procedure overhead. Moreover, we propose channel partition and dynamic path priority control scheme to reduce the control complexity and arbitration overhead. These techniques contribute much to efficient RSON data channel control and improved performance.

In this paper, we make the following major contributions. Firstly, we propose RSON, an optical network for rack-scale computing systems. RSON connects local and remote processor cores, caches, memories through a novel inter/intra-chip silicon photonic network architecture. Secondly, we clearly present how the ONoC and inter-chip optical networks

This work is partially supported by PCF.004.16/17 and FSGRF15EG04 of Hong Kong SAR.

978-3-9819263-0-9/DATE18/©2018 EDAA

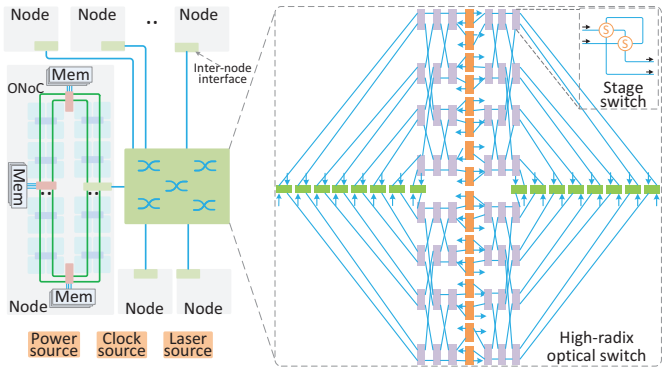


Fig. 1: RSON architecture overview. The inter-chip network is deployed to interconnect tens of server nodes. ONoC (or intra-chip optical network) is designed to connect the on-chip memory controllers and inter-node interface.

are efficiently connected by inter-node interface and optical transceiver, eliminating the unbalanced on-chip and off-chip interconnection performance. Thirdly, we propose the scalable channel partition and dynamic path priority scheme to efficiently conduct channel assignment and path reservation with slight power overhead. This reduces path reservation and control overhead dramatically. Finally, we thoroughly evaluate and compare RSON with existing designs by application performance at extreme scale (APEX) benchmarks [20]. RSON achieves up to 6.8X higher system performance compared to that of InfiniBand-based rack architecture under the same energy consumption.

The rest of this paper is organized as follows. Section II illustrates the RSON's structure and architecture design details. The communication flow and arbitration are presented in Section III. Section IV presents the evaluations of our proposed RSON. We conclude this paper in Section V.

## II. RSON ARCHITECTURE

RSON is an optical intra-rack communication architecture. It aims for high performance and energy efficiency for rack-scale computing systems. RSON covers the design of inter-chip optical network, intra-chip optical network (or ONoC), optical inter-node interface and optical transceiver. In addition to these, RSON also implements a scalable and efficient control scheme and communication flow optimization to improve the functionality of hardware components.

### A. Architecture Overview

The overview RSON computing system structure is shown in Fig. 1. RSON consists of tens of server nodes, and each node is composed of multiple core clusters, several memory controllers and the integrated inter-node interface. These nodes are fully connected by the inter-chip interconnect via integrated silicon photonic switch fabric and optical fibers. This high-radix switch is a staged switch, as shown on the right-hand side of Fig. 1. It is composed of the input stage, middle stage, and output stage with similar stage structure built by waveguides and basic optical switching elements. The switch is able to provide high-throughput, low-latency and low-loss switching. On each server node, a dedicated ONoC is designed to provide high-bandwidth, low-latency communication service for local memory controllers and the inter-node interface, benefiting both local memory copy and remote memory access. The inter-node interface is used to "merge" the inter-chip and ONoC networks, enabling streamed optical data path from local memory to remote memory via the high-radix optical switch without relaying along the path. In our design, the E/O/E conversion only happens in the inter-node interface and the optical switch ports, off-chip laser sources are deployed as the optical power source for sharing among nodes and optical switch and easy implementation.

The detailed server node organization of core clusters, private/share caches, memories, electrical Network-on-Chip (ENoC) and ONoC is shown in Fig. 2. These core clusters, including caches, memory controllers, inter-node interface and other on-chip resources are interconnected by traditional electrical interconnects via corresponding ENoC interfaces. Except for the ENoC, we also propose ONoC to provide high-bandwidth and low-latency interconnection between local memory controllers and inter-node interface for local memory copy and remote memory access. These together builds up the powerful and efficient rack-scale computing systems.

### B. ONoC Design

The on-chip network could be a bottleneck as the inter-chip network evolves to high-bandwidth and low-latency optical interconnect. We propose an Optical Network-on-Chip (ONoC) interconnecting memory controllers and inter-node interface via corresponding ONoC interfaces to eliminates the inter/intra-chip performance bottleneck, as shown in Fig. 2. Each ONoC interface contains the optical transceivers to bridge the on-chip and off-chip optical interconnect, as shown on the right side of Fig. 2. The ONoC data channel is composed of multiple parallel closed-loop bidirectional waveguides. Wavelength-division multiplexing (WDM) technique is used to modulate multiple wavelengths into a single waveguide/fiber. 25 Gbps optical modulator is demonstrated in [21]. We can obtain 400 Gbps using 16 separate wavelengths by extending from [22]. $N$ transceivers work together to conduct communication. This will be illustrated further by channel partition in section III.B. For each communication transaction, the N transceiver blocks within an ONoC interface should work together to provide parallel and high data rate transmission. This design provides two dimensions of flexibility, adjusting bandwidth by changing the number of transceiver blocks and reducing contention by utilizing different data channels for different ONoC interfaces. The number of waveguides $N$ would adjust to the designed overall bandwidth accordingly. This configuration also applies to inter-chip optical links. It avoids as many pins as electrical links needed to provide equivalent chip I/O bandwidth. This can mitigate chip pin struggle and package cost greatly.

### C. Inter-node Interface and Optical Transceiver

The optical inter-node interface is integrated into the chip, working as an interface to connect the inter-chip optical network and the ONoC. Optical switch and traditional server node based rack architecture still needs an extra hop to connect the traditional on-chip network and inter-chip optical

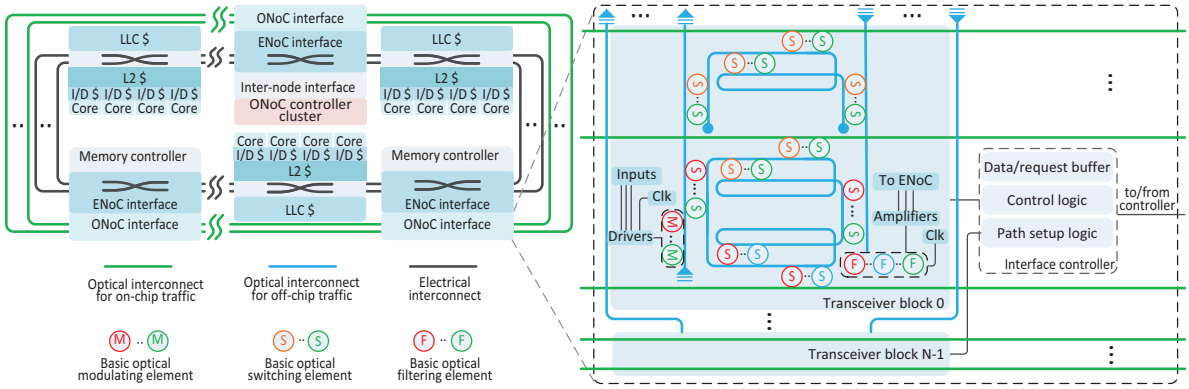


Fig. 2: Server node architecture. The organization of core clusters, private/shared caches, ENoC, ONoC, inter-node interface, ONoC controller cluster is shown on the left. The design of ONoC interface and optical transceivers is shown on the right.

network via network interface controller. This incurs extra latency and power overhead and diminishes the benefits of high-bandwidth ONoC and inter-chip optical interconnect. Our results show that if the performance bottleneck between ONoC and inter-chip networks is not eliminated, only less than 12% performance gain is achieved under APEX [20] benchmarks.

The inter-node interface contains ENoC and ONoC interface to interact with the ENoC and ONoC, as shown on the right-hand side of Fig. 2. It bridges the on-chip and off-chip optical interconnects via the optical transceiver blocks. The interface controller is designed to coordinate signals from ENoC and set up the optical devices for data transmission. Two types of inter-chip traffic are handled by the inter-node interface. The first type is channel semantic, for instance, send/receive operation, which exchanges data from buffers and caches via ENoC first and E/O conversion in the inter-node interface. The data received from the ENoC interface are then modulated into optical signals and sent out under the grant of the inter-chip network. ONoC is not used in this scenario. The second type is memory semantic, remote direct memory access (RDMA) as a common case, which needs the cooperation of local/remote ONoC and inter-chip optical network and realizes direct data movement between local and remote memories. Circuit switching is used for optical interconnect communication and the path in local/remote ONoC and optical switch should be ready before real data transmission. It requires efficient coordination of ONoC and the inter-chip optical switch. That is one of the reasons that we place the ONoC controller cluster in the inter-node interface.

The transceiver block is essential for functions of ONoC and inter-node interface. Its functionality is illustrated in Fig. 3. Channel 0 to $M-1$ are for data transmission. Different communication transactions can employ different channels from each transceiver block. The data originated by local mem can travel from right to left towards remote nodes channel 0. Concurrent transaction from other node can happen from left to right heading to local mem. Bidirectional transmission is also realized by changing relevant optical element status. The data from ENoC can be sent out to local memory or other nodes directly without switching. Each data channel is virtually sectioned and these sections are independent of each other so that they can conduct communication simultaneously. Specially, channel $M$ is implemented for the purpose of control

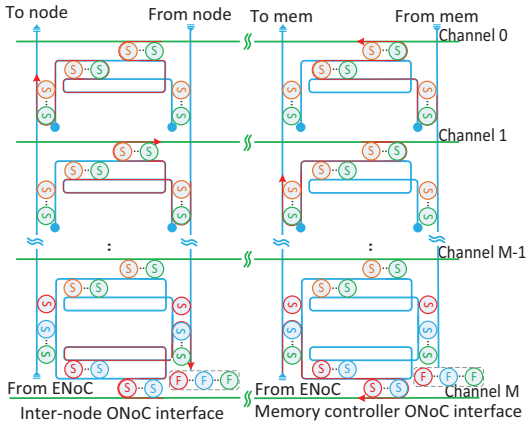


Fig. 3: Illustration of optical transceiver block's functionality.

signals. Each interface has unique wavelength for control, so the control signals from/to different interfaces can be efficiently transmitted without blocking and interfering.

## III. Communication Control Scheme

RSON provides full support for channel and memory semantic, RDMA as a common case, inter-chip/node communications traffic. Circuit switching is used for these two kinds of traffic and path setup is needed but time-consuming. As much as 80% of the traffic generated in cloud computing stays within racks [23], the traffic is handled efficiently by the underlying control schemes to reduce setup overhead and queue delay.

### A. Inter-chip Optical Switch Control

The optical switch handles requests from all inter-node interfaces regardless of channel and memory semantic traffic. As in the ONoC, dedicated wavelengths are used to transmit control signals for efficient signaling. In other words, the inter-node interface can use the same transceivers for inter-chip and intra-chip control purpose, without adding hardware overhead. The switch adopts a tree-based structure to handle all ports' requests. It utilizes a hierarchical round-robin arbitration scheme to provide efficient and fast path control. The synthesis results show that it can do arbitration in 8 cycles for 64-port switch fabric under 15 nm technology node. The power is around 0.26 $W$, which corresponds to a small fraction of the total interconnection energy. Switch setup process for the requested path is necessary and will incur extra latency on data transmission. It is notable that it can be hidden by the

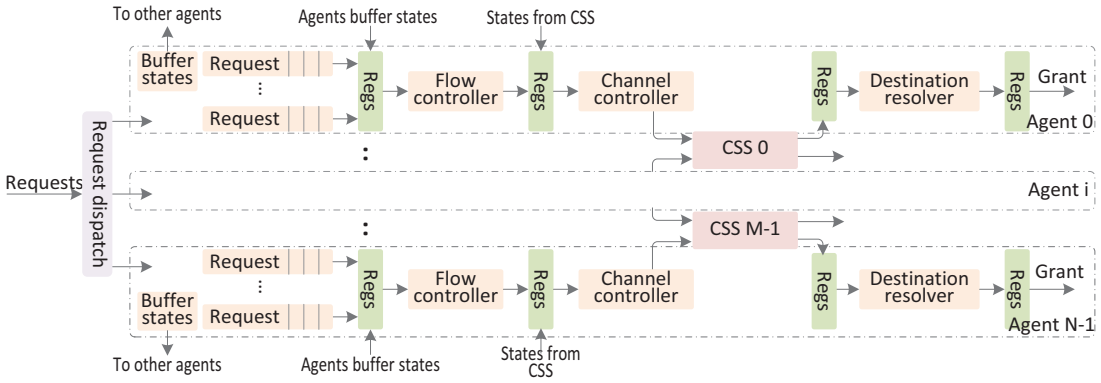


Fig. 4: ONoC controller cluster implementation diagram. There are N agents and M channel section solvers (CSS), where N and M vary according to the number of memory controllers and the ONoC's configuration.

grant to requesters. In most cases, optical switch needs operate collaboratively with the inter-node interface and ONoC to conduct communication. The request and acknowledge packets can be forwarded via internal buffers and links to the targeted output ports and server node's inter-node interface by the dedicated control links. This is critical for efficient optical path reservation and coordination over source/destination node's ONoC and the switch itself.

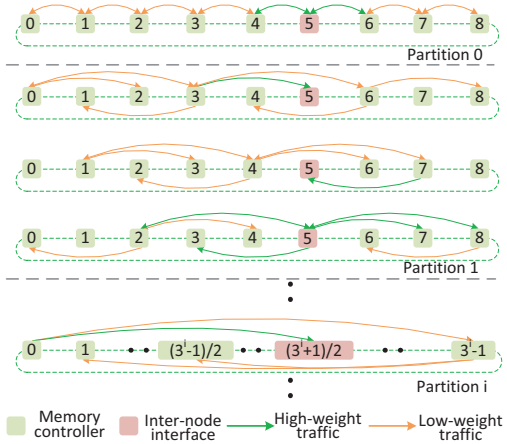


Fig. 5: ONoC channel partition. The square is the ONoC interface for memory controller or optical inter-node interface.

### B. Channel Partition

The aforementioned ONoC design flexibility on channel section and number of data channels could lead to a performance and scalability issues for the controller. To lower the control overhead, we present channel partition to divide data channels into different partitions with specific traffic patterns defined as number of intermediate interfaces, in each partition. In Fig. 5, each arrowed line marks one possible traffic pattern. The allowed traffic in partition $i$ is defined as $((3^i-1)/2, 3^i]$, The starting node of each waveguide is different from each other. Therefore, the maximum traffic distance decides the minimum number of waveguides in partition $i$ to cover all intermediate nodes. For example, in group 1, the longest traffic distance is 3. There should be 3 waveguides to cover all intermediate nodes. In each channel section, only the interfaces at two ends are allowed to issue data transmission request. It means that contention only happens in "neighboring" nodes in each channel section. In this way, there are only limited requesters valid for one channel section, therefore the channel section arbitration and reservation can be as low as $O(1)$ for different channel sections. This alleviates ONoC arbitration.

### C. Dynamic Path Priority ONoC Control Subsystem

The inter-node interface is a hot-spot on the ONoC since all traffic from/to inter-chip network needs its support. It could be a performance bottleneck if not handled. We propose dynamic path priority scheme to initialize inter-node interface with higher path priority, as shown in Fig. 5. The path priority is dynamically updated based on grant results. High priority can decrease upon grant until the same value as low-priority path. This avoids starvation on memory to memory traffic.

We implement ONoC control subsystem as in Fig. 4 based on channel partition and the dynamic path priority. The controller mainly contains interface agents and channel section solvers (CSS). Interface agents are corresponding to ONoC interfaces and deal with requests from that ONoC interface. CSS resolves the contention for the specific channel section. When a request arrives, the interface agent should first assign channel partition and section based on source and destination. The agent will iteratively check the request buffer to issue new requests to flow controller, in a round-robin fashion. Under successful flow control, requests are forwarded to corresponding CSS $i$. The agents for different ONoC interfaces can work relatively independently, only sharing buffer status, CSS status. Each CSS arbitrates among requests with their corresponding priorities. The priority will be updated accordingly. As stated in channel partition, the CSS only needs to resolve contention among limited requests and keep the overhead low. The complete process can finish in 6 clock cycles. The synthesis result shows that the controller power is around 0.06 $W$ when using 16 ONoC interfaces.

## IV. Evaluations

The two most commonly used interconnection solutions in industry and academia for high-performance computing are the InfiniBand-based Top-of-Rack switch and Network Interface Controller connected rack architecture (Trad. IB) and the traditional server nodes with optical switch and fibers connected rack (Trad. OS). In this section, we will conduct evaluations on RSON and compare it with these two designs on a simulation platform [24]. By using this simulation platform, we consider the processor microarchitecture, cache subsystem and memory, other than the interconnect itself. The

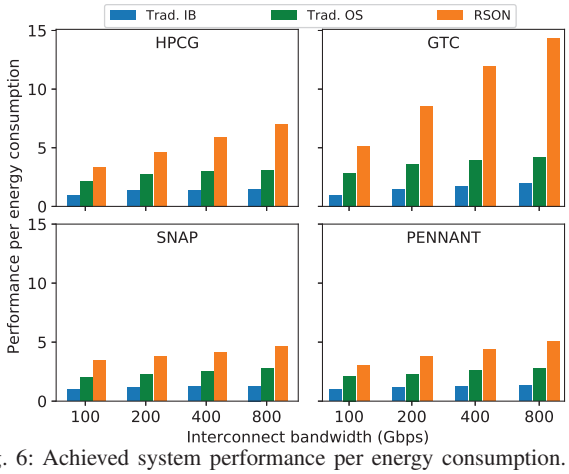


Fig. 6: Achieved system performance per energy consumption. Performance is measured as the execution time for each application.


Fig. 7: Performance comparison under different interconnect bandwidths for different applications.

basic configurations are shown in Table I. Each chip node contains 32 cores with 4 forming a cluster. Eight memory controllers are evenly distributed on the chip and each holds 16 GB DRAM. According to InfiniBand roadmap [25], the link bandwidth is evolving to 200 Gbps when adopting 4x port widths. We scale InfiniBand's interconnect bandwidth beyond 200 Gbps, up to 400 and 800 Gbps for comparison with Trad. OS and RSON. We set the processor frequency to 6 GHz, which will be much more common in the future generation CPUs. The following evaluations are carried out for 64-node/chip system unless otherwise stated. APEX benchmarks, HPCG, GTC, SNAP and PENNANT are used for evaluation.

TABLE I: Basic configurations

| Item | Configuration |
|---|---|
| Core cluster | 8 clusters, 4 cores/cluster |
| L1 I/D $ | 64KB/core, private |
| L2 $ | 512KB/cluster, shared |
| LLC | 2.5MB/slice |
| Coherence protocol | Directory based MOSI |
| ENoC bandwidth | 384Gbps |
| Memory bandwidth | 480GBps |

### A. Performance Evaluation

Firstly, we present the overall system performance per energy consumption (PPE). Higher PPE signifies more performance under the same power consumption. Fig. 6 shows that RSON obtains higher PPE for different applications and interconnect bandwidths. More specially, RSON achieves 6.8X higher PPE than Trad. IB when running GTC under 800Gbps. This is because GTC is communication intensive and is more sensitive to the interconnect technology. It is also interesting to observe that RSON's PPE scales much better than that of Trad. IB and Trad. OS regarding the interconnect bandwidth. A similar trend is observed in performance results, as shown in Fig. 7. These results demonstrate that RSON brings joint benefits on performance and the PPE, which is crucial for the emerging exascale HPC systems to obtain higher performance under strict power constraint.

### B. Energy Efficiency

The interconnect energy efficiency deserves a thorough evaluation as it contributes much to overall system energy consumption. Optical power is measured as minimum laser power generated to compensate the loss induced by optical devices, as shown in Table II with variations from various process technology. It should be noticed that the thermal tuning power for optical devices, SERDES energy consumption and the request/data buffer are also fairly considered in the interconnection. As shown in Fig. 8, both RSON and Trad. OS gain much better energy efficiency, consuming almost 85% less energy for transmitting one bit compared to that of Trad. IB. Moreover, RSON maintains the energy efficiency under 18 pJ/bit even considering optical devices loss variation. It's noteworthy that RSON achieves comparable energy efficiency as Trad. OS. It is because that optical inter-chip interconnect contributes major part of the overall interconnect energy consumption. However, RSON gains more performance benefit by introducing optical interconnect to on-chip communication.

TABLE II: Parameters for optical interconnect

| Item | Value |
|---|---|
| MR insertion loss | 1 $\sim$ 3 dB |
| MR passing loss | 0.06 $\sim$ 0.3 dB |
| Edge coupling loss | 1 $\sim$ 3 dB |
| Waveguide crossing loss | 0.3 $\sim$ 0.8 dB |
| Waveguide Bending loss | 0.05 $\sim$ 0.1 dB/90° |
| Waveguide propagation loss | 0.8 $\sim$ 1.3 dB/cm |
| Photodetector sensitivity | -15 $\sim$ -20 dBm |

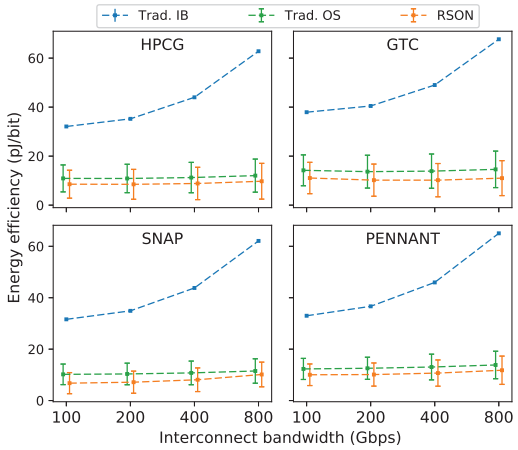


Fig. 8: The interconnect energy efficiency when considering optical parameters' range of variations.

Technology's impact on interconnect energy is shown in Fig. 9. Trad. IB energy efficiency decreases dramatically as

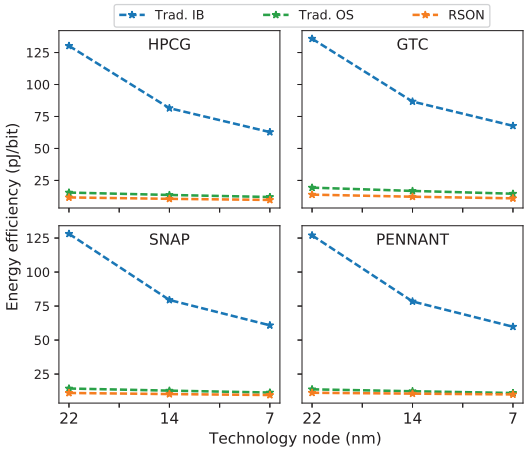


Fig. 9: Energy efficiency vs. Technology nodes.

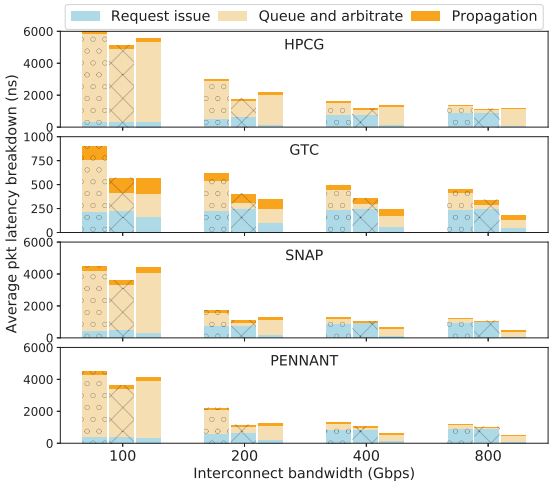


Fig. 10: Latency breakdown. The bars in each group stand for Trad. IB, Trad. OS and RSON from left to right.

process node evolves from 22 nm to 7 nm, while RSON and Trad. OS's trend is slower. The reason is that the advance of technology node brings corresponding power reduction to electrical circuits related interconnect technology, but optical power lowers little since they are relatively independent on technology node. It is harder to realize high-bandwidth and energy-efficient interconnection using the traditional electrical methods as Moore's law stops taking effect. Thus, RSON is a promising design choice to fulfill the performance and energy requirements towards exascale systems.

### C. Latency Analysis

Latency is another important factor for the interconnection architecture and it affects a lot on the task's turnaround and response time. We break the average latency into request issue, queue/arbitrate and propagation latency, which represents the time between initialization and issuance of the switch, time on waiting for the grant from the switch and the packet transmission time, respectively. The results are shown in Fig. 10. The main trend indicates that higher bandwidth can lessen the average latency. However, a closer look reveals that further bandwidth increase will not significantly lower Trad. IB and Trad. OS latency. Request issue dominates latency in higher bandwidth for Trad. IB and Trad. OS. These two designs' limited chip I/O bandwidth and protocol restrict the benefit of high bandwidth inter-chip interconnect. It is reasonable that the queue and arbitration delay takes up a major part of RSON's packet latency since circuit switching requires path setup and release along requested path and the straightforward multi-domain reservation scheme is used. This could be improved with some cooperative and proactive channel reservation protocols.

## V. Conclusions

In this paper, we have introduced RSON, an inter/intra-chip rack-scale optical network for rack-scale computing systems. RSON considers both the inter-chip optical network and ONoC and uses the optical inter-node interface to eliminate performance gap. Scalable and fast control schemes is achieved by the proposed channel partition and dynamic path priority, dealing with channel semantic and memory semantic traffic efficiently. The evaluation demonstrates that RSON is advantageous on both performance and energy consumption compared to Trad. IB and Trad. OS, providing a promising solution for energy-efficient exascale systems.

## References


[1] C. Minkenberg, "Interconnection Network Architectures for High-Performance Computing," tech. rep., IBM Research, Zurich, 2013.

[2] T. Krishna *et al.*, "Automatic place-and-route of emerging LED-driven wires within a monolithically-integrated CMOS- III-V process," in *DATE*, IEEE, 2017.

[3] M. J. Cianchetti *et al.*, "Phastlane: a Rapid Transit Optical Routing Network," in *the 36th ISCA*, 2009.

[4] Y. Pan *et al.*, "FlexiShare: Channel Sharing for an Energy-efficient Nanophotonic Crossbar," in *16th HPCA*, 2010.

[5] S. Koohi *et al.*, "All-Optical Wavelength-Routed Architecture for a Power-Efficient Network on Chip," *Transactions on Computers*, 2012.

[6] S. Bartolini *et al.*, "Olympic: A Hierarchical All-Optical Photonic Network for Low-Power Chip Multiprocessors," in *DSD*, 2013.

[7] N. Kirman *et al.*, "A Power-efficient All-optical On-chip Interconnect Using Wavelength-based Oblivious Routing," *SIGPLAN Not.*, 2010.

[8] S. Le Beux *et al.*, "Chameleon: Channel Efficient Optical Network-on-Chip," in *DATE*, 2014.

[9] D. Vantrease *et al.*, "Corona: System Implications of Emerging Nanophotonic Technology," in *35th ISCA*, 2008.

[10] J. Kim *et al.*, "Cost-efficient Dragonfly Topology for Large-scale Systems," in *OFC*, 2009.

[11] X. Wu *et al.*, "Union: A unified inter/intrachip optical network for chip multiprocessors," *TVLSI*, 2013.

[12] Z. Wang *et al.*, "Low-Loss High-Radix Integrated Optical Switch Networks for Software-Defined Servers," *JLT*, 2016.

[13] D. Whelihan *et al.*, "P-sync: A Photonically Enabled Architecture for Efficient Non-local Data Access," in *IPDPS*, 2013.

[14] C. Batten *et al.*, "Designing Chip-level Nanophotonic Interconnection Networks," pp. 81–135, 2013.

[15] S. Rumley *et al.*, "Low Latency, Rack Scale Optical Interconnection Network for Data Center Applications," in *ECOC*, 2013.

[16] A. Krishnamoorthy *et al.*, "From Chip to Cloud: Optical Interconnects in Engineered Systems for the Enterprise," in *IEEE OI*, 2016.

[17] P. Andreades *et al.*, "Experimental Demonstration of 75 ns End-to-end Latency in an Optical Top-of-rack Switch," in *OFC*, 2015.

[18] X. Ye *et al.*, "DOS: A Scalable Optical Switch for Datacenters," in *ANCS*, 2010.

[19] F. G. de Magalhães *et al.*, "Design and modelling of a low-latency centralized controller for optical integrated networks," *CL*, 2016.

[20] "Benchmark Distribution & Run Rules." http://www.nersc.gov/research-and-development/apex/apex-benchmarks/.

[21] K. Yashiki *et al.*, "25-gbps error-free operation of chip-scale si-photonics optical transmitter over 70 $^{o}$c with integrated quantum dot laser," in *OFC*, 2016.

[22] P. Yang *et al.*, "Inter/intra-chip optical interconnection network: opportunities, challenges, and implementations," in *NOCS*, 2016.

[23] T. Benson *et al.*, "Network Traffic Characteristics of Data Centers in the Wild," in *the 10th SIGCOMM conference on IM*, 2010.

[24] R. K. Maeda *et al.*, "JADE: a Heterogeneous Multiprocessor System Simulation Platform Using Recorded and Statistical Application Models," in *AISTECS*, 2016.

[25] InfiniBand Trade Assocation, "InfiniBand Roadmap."