is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

Wavelength-Reused Hierarchical Optical Network on Chip Architecture for Manycore Processors

Feiyang Liu, Haibo Zhang, Yawen Chen, Zhiyi Huang, and Huaxi Gu

Abstract—Manycore processor becomes the mainstream platform for cloud computing applications. However, the design of high-performance and sustainable inter-core communication network is still a challenging problem. Optical Network on Chip (ONoC) is a promising chip-scale optical communication technology with high bandwidth capacity and energy efficiency. In this paper, we present a Wavelength Reused Hierarchical ONoC architecture, WRH-ONoC. It leverages the nonblocking wavelength-routed λ-router and the hierarchical networking to reuse the limited number of available wavelengths. In WRH-ONoC, all the cores are grouped to multiple subsystems, and the cores in the same subsystem are directly interconnected using a single λ-router for nonblocking communication. For inter-subsystem communication, all subsystems are further connected through multiple λ-routers and gateways in a hierarchical manner. Thus, the available wavelengths can be reused in different λ-routers. Furthermore, WRHm-ONoC, an efficient extension with multicast ability is also proposed. Given the numbers of cores and available wavelengths, we derive the minimum hardware requirement, the expected end-to-end delay, and the maximum data rate. Theoretical analysis and simulation results indicate WRH-ONoC achieves prominent improvement on communication performance and sustainability, *e.g.,* 46.0% of reduction on zero-load delay and 72.7% of improvement on throughput for 400 cores with modest hardware cost and energy overhead.

Index Terms—Manycore processor, Optical Network on Chip (ONoC), sustainability, routing protocol, wavelength reuse, hierarchical network.

1 Introduction

The development of today's processor has moved to an era with many micro-cores in a single chip, *e.g.*, 72 cores in Tilera Tile-Gx72 [2], 80 cores in Intel Teraflops [3], and 256 cores in Kalray MPPA [4]. Manycore processor is becoming the mainstream platform for streaming applications, cloud computing, data center, and supercomputing systems [5]. It was predicted that thousands of or even more cores will be integrated in a manycore processor in the near future [6]. With the rapid development of high-speed cores and the large requirement of data communication between cores [7] [8], an inter-core communication network should achieve high sustainability, *i.e.*, obtaining low delay and high bandwidth capacity with low hardware cost and energy overhead. Even though the conventional electrical interconnect is efficient for small-scale processors with tens of cores, in the future it is hard to meet the communication requirement and energy efficiency of large-scale manycore processors with more than hundreds of cores, due to deep submicron effects of metallic interconnects, *e.g.*, increased link delay and leakage power [9]. The design of sustainable inter-core network architecture and communication scheme is a challenging problem for manycore processors [10].

Recent advances in nanophotonics have led to the development of Optical Network on Chip (ONoC), a silicon-compatible optical interconnection network for the cores at chip level, as an attractive solution to overcome the limitations of conventional electrical interconnects [11][12]. In an ONoC architecture, data packets are transmitted between cores by using modulated optical signals. With the benefits of Wavelength Division Multiplexing (WDM), multiple optical signals can be transmitted simultaneously through the same waveguide using different wavelengths, thereby enabling extremely high throughput and low communication delay [13] [14]. In industry, Intel announced the use of silicon photonic architecture to define the next-generation multicore processors and revealed its first inexpensive 100 *Gbps* optical chip in 2013 [15]. IBM advanced a significant step by integrating photonic chip on the same package as CPUs in 2015 [16]. However, from the manufacture engineering point of view, at present optical interconnects also suffer from many difficulties, such as the lack of optical buffer and processing logics, and the limited number of wavelength channels (*e.g.*, 62 wavelengths in maximum with 19 *Gbps* bandwidth and -20 *dB* noise tolerance [17]). Existing electrical-based schemes cannot be simply exploited in ONoC due to different physical properties [6] [9]. Efficient wavelength reused ONoC architecture is a promising but challenging research problem for the high-performance and sustainable inter-core communication of manycore processors [1] [18].

In this paper, we propose a Wavelength-Reused Hierarchical ONoC architecture, namely WRH-ONoC, for manycore processors. The design principle of WRH-ONoC is to achieve high-performance and sustainability communication for large-scale manycore processors by exploring the benefits of hierarchical networking and wavelength reused routing. In our scheme, all the cores in a manycore processor are grouped into subsystems. Then, wavelength-routed λ-router is employed to provide non-blocking optical communication for each subsystem [19], while multiple λ-

F. Liu, H. Zhang, Y. Chen, and Z. Huang are with the Department of Computer Science, University of Otago, Dunedin 9016, New Zealand, e-mail: (feiyang, haibo, yawen, hzy@cs.otago.ac.nz). H. Gu is with the State Key Laboratory of ISN, Xidian University, Xi'an 710071, China, e-mail: (hxgu@xidian.edu.cn).
This paper is an extension of previous work published in INFOCOM 2015 [1].

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

routers and gateways are further connected in a hierarchical manner to provide high-bandwidth inter-subsystem communication. The available wavelengths are reused in different λ-routers through wavelength reassignment in gateways. Hence, WRH-ONoC is capable of interconnecting even thousands of cores by using a limited number of wavelengths. The key contributions are listed as follows:

- We propose a wavelength-reused hierarchical architecture that sustains the strength of λ-router in high-throughput communication, but offsets its weakness in scalability. By dividing the cores into subsystems and connecting them using λ-routers in a hierarchical network, available wavelengths can be reused in all λ-routers by wavelength reassignment in gateways.
- We design unicast and multicast routing schemes. Intra-subsystem communication is fully nonblocking via one hop of λ-router, while high-throughput inter-subsystem communication is achieved through paralleled wavelength reassignment in each gateway and dynamic load balance among sibling gateways. WRHm-ONoC is a system-level multicast extension.
- We derive the expected end-to-end communication delay, the maximum data rate, and the average energy consumption for the unicast communication, assuming that the traffic follows Uniform distribution in space and Poisson distribution in time.
- We analyse the hardware requirements of WRH-ONoC with given number of cores. The results indicate that optical devices can be reduced by $\sim$90% in comparison with the λ-router scheme. Meanwhile, compared with existing ONoC schemes, the overall hardware cost measured in chip area can also be reduced.
- We carry out extensive simulations to evaluate WRH-ONoC using both real data traces and synthetic traffic patterns. Simulation results indicate that it is efficient for both unicast and multicast communications. Compared with existing schemes, it can achieve significant improvement on the performance and scalability, *e.g.*, with a decrease of 46.0% on zero-load delay and an increase of 72.7% on throughput for 400 cores.

The rest of this paper is organized as follows. Section 2 introduces the background and related work. Section 3 presents the network architecture and its communication schemes. Section 4 theoretically analyses the performance, and Section 5 presents the simulation results. Finally, Section 6 concludes the paper and discusses future work.

2 Background

2.1 Optical Network on Chip

The fundamental components of an ONoC include light sources, silicon waveguides, optical routers, modulators, and photodetectors. Light source provides optical signals on which information is carried, and it can be implemented by using an off-chip laser coupled with power waveguides [14] or on-chip lasers for each core [20]. Waveguide is the optical transmission medium whose propagation loss can be less than 1 *dB/cm*. Optical router conducts high-speed switching for optical signals when its connection is configured in advance. Modulator converts electrical signals into optical signals (on-off keying) and injects them to waveguides, and photodetector receives optical signals and converts them back to electrical signals. Microring resonator (MR) is the basic element of optical routers, modulators, and photodetectors. It is a compact and energy-efficient optical filter which is designed to pass optical signals with a specific wavelength [21]. As shown in Fig. 1(a), when the wavelength of an input signal λ_i equals to the resonant wavelength λ_r of MR, the optical signal couples into the MR and changes its direction; otherwise the optical signal keeps in its original direction. Thus, in the optical switch, optical signals with different wavelengths can be filtered and routed to different outputs in parallel based on the specific resonant wavelength of MRs, as shown in Fig. 1(b).

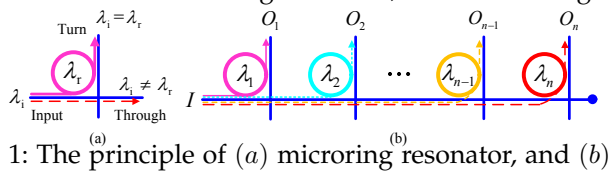


Fig. 1: The principle of (a) microring resonator, and (b) optical switch by filtering signals with different wavelengths.

2.2 λ-Router

λ-router is a wavelength-routed optical crossbar which can provide nonblocking communication among all the connected cores by using totally different wavelengths [19]. Fig. 2(a) illustrates the connection scheme of an 8-inputs×8-outputs λ-router. The key component in a λ-router is the 2-inputs×2-outputs optical switching element (OSE) designed by using two MRs which have the same resonant wavelength λ_r, as shown in Fig. 2(b). When the wavelength of input signal λ_i equals to λ_r of the OSE, the optical signal will be coupled into an MR of OSE and output from one port; otherwise the optical signal will pass the OSE and output from the other port. OSE is an area-compact optical device, and its footprint is less than $10\times10\mu m^2$ when the radius of each MR is smaller than $3\mu m$ in general [22].

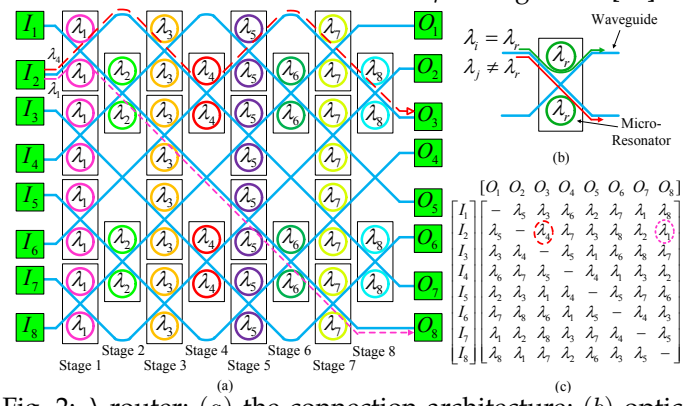


Fig. 2: λ-router: (a) the connection architecture; (b) optical switching element (OSE); (c) wavelength routing matrix.

For the nonblocking optical communication among the connected cores, an $N\times N$ λ-router needs to employ wavelength routing by using N waveguides and N wavelengths. Since there is no communication between input I_i and output O_i as they connect to the same core $\{i\}$, the minimum number of MRs and OSEs required are $N(N-2)$

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

and $\lceil\frac{N(N-2)}{2}\rceil$, respectively [1]. As illustrated in Fig. 2(a), the layout of OSEs in a λ-router is in N stages where the number of OSEs in stage i is $\lfloor\frac{N}{2}\rfloor$ if i is odd and $\lfloor\frac{N-3}{2}\rfloor$ if i is even [22]. In each stage, all the OSEs share the same resonant wavelength λ_r. The wavelength used for the communication between I_i and O_j is determined by $M_{i,j} \in \boldsymbol{M}$, where $\boldsymbol{M}$ is the wavelength routing matrix. For example, according to the matrix given in Fig. 2(c), $M_{2,3}$=λ_4 and $M_{2,8}$=λ_1. The communication paths from I_2 to O_3 and I_2 to O_8 are determined only by the wavelengths of λ_4 and λ_1 respectively, as highlighted by dashed lines in Fig. 2(a).

λ-router has several distinctive *advantages*: fully non-blocking communication, multicast capability through WDM, and low latency and high bandwidth. However, the only *drawback* of λ-router is its *poor scalability*. The number of waveguides and wavelengths are linearly proportional to the size of λ-router (the number of ports), but the number of OSEs and MRs increase *quadratically*. That makes it not suitable for large-scale ONoCs, *e.g.*, to connect 100 cores, a λ-router requires as many as 100 wavelengths, 100 waveguides, and 4900 OSEs (9800 MRs) cascaded in 100 stages.

2.3 Related Work

Existing ONoC architectures can be generally classified to two categories: *all-optical* [19] [23] [24] and *electrical-optical hybrid* [12] [25] [26] [27]. All-optical architectures connect all the cores with solely optical components, *e.g.*, λ-router [19], ORNoC [23], and QuT [24]. Most of all-optical architectures can provide the nonblocking communication through fixed wavelength allocation, namely assigning different optical routing paths or wavelengths fixedly between any two cores, similar to λ-router. However, their main drawback is *poor scalability*, due to the limited available wavelengths and quadratic increase of required optical devices.

This paper focuses on the research of electrical-optical hybrid ONoC to achieve high scalability. Electrical-optical hybrid architectures combine the properties of electrical and optical interconnects. One type of hybrid architecture is to employ the optical circuit-switching scheme, namely using an extra electrical control network to conduct buffering and processing for the optical data network, *e.g.*, PNoC [12], HOME [28], and 3D-DMONoC [29]. The end-to-end optical routing path is dedicatedly reserved by the electrical network before a specific communication in a hop-by-hop way. These schemes can provide guaranteed communication for data-intensive applications after the optical path reservation. However, the time-consuming electrical path-setup process leads to long preparation delay and low link utilization for short packets due to severe contentions, especially for connecting up to hundreds of cores. Moreover, for N cores, PNoC and 3D-DMONoC [29] need N electrical and optical routers in a 2D/3D mesh/torus topology, while HOME shares one optical router among four cores so that it is a 2D mesh with $\frac{N}{4}$ nodes. These schemes can lead to longer communication paths in average for large-scale manycore systems. Partially considering the potential hardware costs, WDM is not employed in these schemes. In our scheme, the nonblocking λ-routers with wavelength routing is utilized, thus it needs no optical path reservation in advance. Moreover, the communication path is significantly shortened due to the hierarchical networking of λ-routers and the fast transmission in each λ-router.

Another type is to construct a hierarchical network with multiple electrical local interconnects and one or more optical global interconnects. For instance, Corona [25], ATAC [26], Firefly [27], OCMP [9], and LumiNOC [13] divide all the cores in a large system into several small clusters, then use an electrical network in each cluster and connect all the clusters through ring-like optical crossbars. In these schemes, the size of optical global network is constant according to the number of available wavelengths, namely 64. For intra-cluster traffic, they use different electrical interconnects, *e.g.*, crossbar for Corona, mesh for ATAC, Firefly, OCMP, and LumiNOC. For inter-cluster traffic, the optical crossbar utilizes WDM to provide separate wavelength channels among clusters, *e.g.*, multi-write-single-read for Corona, OCMP, and LumiNOC, while single-write-multi-read for ATAC and Firefly. However, an arbitration scheme is required to solve the access contentions to the global optical network, since the wavelength need to be dedicatedly reserved and cannot be reused. In terms of scalability, the electrical intra-cluster interconnects are not efficient, since in most cases the cores in the same cluster are communication-intensive [30]; while the contentions in the global optical network can also become their bottleneck, especially when multiple cores share the same access point in the optical network. Firefly and LumiNOC can only reduce the contentions by increasing same redundant optical interconnects. In our scheme, the optical interconnects are used for both intra- and inter-subsystem traffics with nonblocking λ-routers, in which each input use different wavelengths to different outputs, so it needs no extra arbitration for the optical routing path. Moreover, the available wavelengths can be reused in our scheme by wavelength reassignment in gateways to achieve high scalability.

3 Architecture and Communication

The design principle of our scheme is to sustain the strength of λ-router in nonblocking routing but offset its weakness in poor scalability. In our scheme, the optical communication is provided for both local traffic (in one hop) and global traffic (in multiple hops), and no extra arbitration is needed for optical interconnects due to the wavelength routing.

3.1 Hierarchical Networking

We aim to design an ONoC architecture to interconnect N cores by using only $W_{\max}$ wavelengths where $N \gg W_{\max}$. As illustrated in Fig. 3, all the cores are grouped into multiple subsystems. For generality, in this paper we suppose each subsystem interconnects the same number of cores. Within each subsystem, the cores are connected using a small λ-router with sufficient wavelengths, thereby providing nonblocking optical communication in each subsystem. Communication between cores in different subsystems is done through the network hierarchy which is constructed by using multiple λ-routers and gateways

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

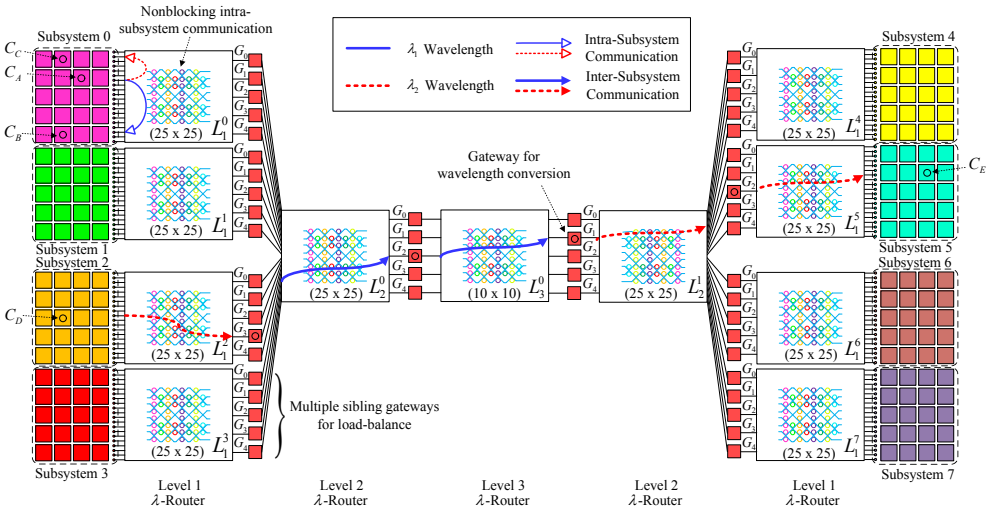


Fig. 3: An example of WRH-ONoC for interconnecting 160 cores using 25 wavelengths with 3 levels of λ-routers.

organised in a hierarchical manner. The gateways serve as the bridges between λ-routers to achieve wavelength reuse, where optical signals can change their wavelengths. Multiple *sibling gateways*, which connect with the same two λ-routers, are used to obtain high throughput by providing paralleled wavelength reassignment between λ-routers in adjacent levels in the hierarchical network. Fig. 3 gives an example of 160 cores using only 25 wavelengths via a three-level hierarchical λ-router network. 160 cores are divided to 8 subsystems each with 20 cores, and 5 sibling gateways are used to connect a λ-router to the next-level λ-router.

Through the wavelength reassignment in gateways, all $W_{\max}$ available wavelengths can be reused by each λ-router, which only connects with a group of cores instead of the whole system, thereby overcoming the drawback of scalability for using a large λ-router directly. An off-chip laser with multiple wavelengths is coupled to several power waveguides which provide light source for each core and gateway [25]. Currently, due to technology limitations of chip-scale optical wavelength converter, the optical signal needs to be converted to electrical signal, temporarily buffered, and then retransmitted using another wavelength.

From the perspective of networking, WRH-ONoC is *similar to a tree network*, *i.e.*, the top-level λ-router is the root and cores are leaves. However, the *most important difference* is that WRH-ONoC has multiple redundant channels between each parent and child λ-routers by using sibling gateways, thus it can provide higher capacity between λ-routers in different levels. Load balance can be achieved by uniformly choosing a sibling gateway for wavelength reassignment between λ-routers in a dynamic way. For instance, in Fig. 3, there are 625 paths between C_D and C_E, depending on which gateways are used in each hop.

3.2 Gateway Architecture

As shown in Fig. 4, each gateway consists of input/output ports, optical-to-electrical (O/E) and electrical-to-optical (E/O) converters, buffer queues, packet dispatchers, and wavelength matrices. Each gateway has two separate pairs of input/output ports: one pair for upward traffic from a lower-level λ-router to a higher-level λ-router (upward direction), and the other for downward traffic from a higher-level λ-router to a lower-level λ-router (downward direction). Each pair of input/output ports has independent buffer queues. Let w_l and w_h be the number of wavelengths used in the lower-level and higher-level λ-routers which are connected by a gateway, respectively.

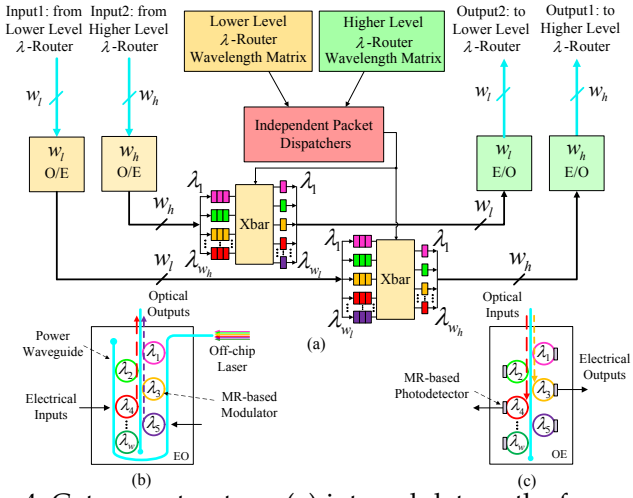


Fig. 4: Gateway structure: (a) internal data paths for wavelength assignment; (b) E/O converter with MR-based modulators; (c) O/E converter with MR-based photodetectors.

There are w_l MR-based photodetectors in O/E converter for *Input1* with each receiving for a specific wavelength. The optical signal filtered by the MR with a resonant wavelength λ_i is converted to electrical signal by the associated photodetector and buffered in the input queue for λ_i. Similarly, there are w_h MR-based modulators in E/O converter for *Output1* with each operating for a specific wavelength. The packet which is allocated for λ_j in the output buffer is modulated by the MR with wavelength λ_j to optical signal and then transmitted. The input queues and output buffers are fully connected using an internal crossbar for wavelength switching. For paralleled wavelength reassignment in gateways, each input queue is associated with a packet dispatcher which is responsible for dispatching packets from input queues to corresponding outputs based on the wavelength to be used in the next hop, according

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

to the wavelength matrices. The design for the downward direction is similar.

The gateway generally operates as follows: when a set of WDM-based optical signals enter an input, each single-wavelength signal is filtered by a specific MR, converted to electrical signal and written to the corresponding input queue based on the receiving wavelength, as shown in Fig. 4(c). Each packet dispatcher works for one wavelength and it continually dispatches packets from input queue to output port. According to the destination of a packet, the dispatcher determines the next-hop λ-router for each packet (details on choosing the next hop will be given in Section 3.3), and looks up either the lower-level or higher-level wavelength matrix to determine the carrier wavelength. As shown in Fig. 4(b), the gateway continuously forwards the packets stored in output buffers to MR-based modulators, and multiple optical signals are multiplexed and transmitted in a waveguide with different wavelengths.

The main advantages of gateway design include that multiple optical signals with different wavelengths can be processed concurrently, and there is no blocking in each data channel thereby maximizing the wavelength utilization. Although the signal conversion and packet buffering introduce some delay, they have negligible impact on the performance of manycore systems, because: (*i*) E/O and O/E conversions can be done at very high speed (less than 100 *ps*) and with low energy cost (100 *fJ/bit*) [9]; (*ii*) in general a large portion of traffic in an ONoC occurs locally due to application properties or task mapping algorithm [30].

3.3 Communication Scheme

The foundation of packet routing is the positional prefix address. Each core has a unique address in the form of $\{networkID; coreID\}$. The *coreID* represents the unique identification of each core in the subsystem it belongs to, and it has $\lceil \log_2 n \rceil$ bits where n is the maximum number of cores in a subsystem. The *networkID* is composed of several fields for subnetworks $\{s_l, ..., s_2, s_1\}$ where l is the number of levels of λ-routers. Fig. 5 illustrates the address assignment for the network given in Fig. 3. The number of bits for *coreID* field is 5 as there are 20 cores in each subsystem. In our design, the routing structure is a tree which rooted at the top-level λ-router, and all cores and gateways in a subtree that rooted at one λ-router forms a subnetwork. Let $|s_i|$ be the number of bits in the s_i field where $0 < i \leq l$. $|s_l|$ is 1 bit as there must be one λ-router in the top level. $|s_i|$ is $\lceil \log_2 r_i \rceil$ for $i < l$ where r_i is the maximum number of level i λ-routers which connect to a λ-router at level $i+1$. Thus, different level i λ-routers can be uniquely denoted by different addresses, as the example shown in Fig. 5.

3.3.1 Unicast Communication

Unicast communication happens between a source core and a destination core, such as reading a remote cache line. In our design, the unicast communication can be classified into *intra-subsystem unicast* and *inter-subsystem unicast*. If the *networkID* of source and destination addresses are equal, it is an intra-subsystem unicast packet; otherwise it is an inter-subsystem unicast packet.

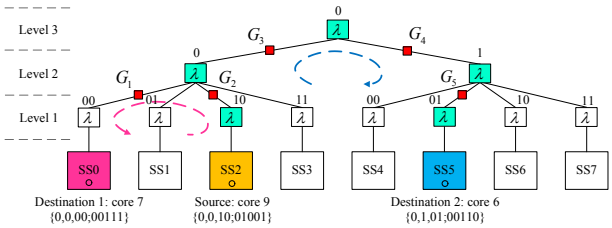


Fig. 5: Positional prefix address for the network hierarchy.

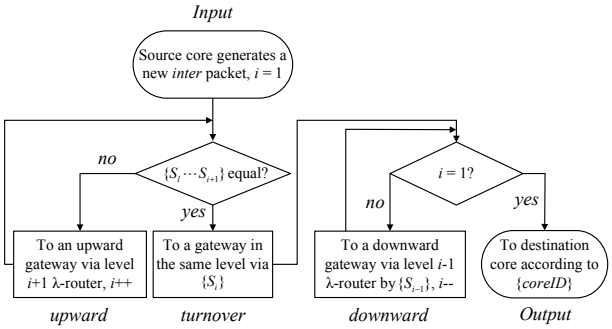


Fig. 6: Routing process of an inter-subsystem packet.

For an intra-subsystem unicast packet, the source core directly looks up its local wavelength matrix to determine the right wavelength to be used, and it sends the packet with this wavelength via the connected λ-router.

For an inter-subsystem unicast packet, the source core uniformly chooses a gateway from the sibling gateways that connect to the same subsystem to achieve load balance, and it sends the packet to the chosen gateway through the bottom level λ-router. Then the packet is routed in gateways and delivered to the destination in a multi-hop manner. The gateway determines the next-hop λ-router and the wavelength used for next-hop transmission. Fig. 6 shows the general process of routing an inter-subsystem packet.

(I) Upward Transmission: When an inter-subsystem unicast packet is generated, it needs to be transmitted upward in the hierarchy according to the location of its destination. Suppose a gateway receives the packet from a level i λ-router. If the $\{s_l, ..., s_{i+1}\}$ fields of source and destination addresses are not equal, the destination must be located outside of the subnetwork which is rooted at the level $i+1$ λ-router. Then, the gateway needs to transmit the packet further upward. For example, in Fig. 5, a packet generated by core 9 in $SS2$ with *networkID* $\{0, 0, 10\}$ is to be routed in the gateway G_2. If the destination is core 6 in SS_5 with *networkID* $\{0, 1, 01\}$, the packet is routed upward to gateway G_3 since the $\{s_3, s_2\}$ fields are unequal.

(II) Turnover: If the $\{s_l, ..., s_{i+1}\}$ fields of source and destination addresses are equal while the $\{s_i, ..., s_1\}$ fields are different, the destination must be located in the same subnetwork rooted at the level $i+1$ λ-router. The packet is stopped from forwarding upward, and it is routed to a sibling gateway which connects to another level i λ-router encoded with $\{s_i\}$. For example in Fig. 5, if the destination of the packet generated by core 9 is core 7 in $SS0$ with *networkID* $\{0, 0, 00\}$. Since the $\{s_3, s_2\}$ fields of the source and destination addresses are the same, G_2 routes this packet to a sibling gateway G_1 which connects to another λ-router encoded as $\{s_1\}$=$\{00\}$ at level 1.

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

(III) Downward Transmission: After the turnover λ-router, the packet is routed in the downward direction. The rules for transmitting a downward packet are: (*i*) if the gateway connects to a level i+1 λ-router and a level i λ-router where $i > 1$, the packet is routed to a gateway that connects the level i λ-router with the λ-router encoded with $\{s_{i-1}\}$ at level i–1; (*ii*) if the gateway connects to the destination subsystem, it transmits the packet directly to the destination core via the bottom level λ-router according to *coreID*. For example, in Fig. 5, if G_4 needs to route the packet to core 6 in $SS5$, it sends the packet to gateway G_5 which connects to the λ-router at level 1 with $\{s_1\}$=$\{01\}$. Then gateway G_5 routes it to the destination core 6 by *coreID*=$\{00110\}$.

3.3.2 Multicast Communication

Multicast communication occurs when transmitting the same data from a source core to multiple destination cores simultaneously. It is quite common in manycore systems for cooperative computing and cache coherence [18] [31]. In this paper, we also present an extension with subsystem-level multicast ability, namely WRHm-ONoC. For intra-subsystem multicast, the packet is modulated using multiple wavelengths according to the destinations' *coreID*s, and separate copies are sent to each destination through the nonblocking λ-router. For example, in Fig. 3, core C_A can multicast packets to C_B and C_C at the same time via two distinct wavelengths. Inter-subsystem multicast is conducted through the combination of inter-subsystem unicasting and intra-subsystem multicasting. Specifically, for each inter-subsystem multicast packet, the source core unicasts a packet copy to every subsystem which contains at least one destination. Once the packet copy arrives at the gateway which connects to the destination subsystem, the gateway uses intra-subsystem multicasting to send the packet copy to all the destination cores in the subsystem through different wavelengths at the same time.

To distinguish between multicast packets and unicast packets, the packet header contains a *multi_flag* field (1 *bit*), *i.e.*, '1' for multicast and '0' for unicast. The multicast address is expressed in the form of {*networkID; bit-string*}, where {*networkID*} is used for routing the packet copy to the destination subsystem, and {*bit-string*} labels all the destination cores in the subsystem. {*bit-string*} contains n bits where n is the number of cores in each subsystem, and the i^{th} bit is marked as '1' only if the i^{th} core is a destination. Compared with the traditional tree-based multicast routing [32], our WRHm-ONoC scheme is simple but very efficient. In tree-based schemes, packet copies are dynamically generated in routing nodes based on the distribution of destination cores. However, this requires each multicast packet to carry the address information of all the destinations, which increases not only the complexity of routing algorithm but also the communication overhead. In WRHm-ONoC, each packet copy only needs to carry the addresses of destinations in one subsystem where it is destined to, and no special routing policy is required for multicast packets in the intermediate gateways.

Note that WRHm-ONoC employs a subsystem-level multicast scheme to converge the bursty multicast traffics, and in the worst case (*e.g.*, broadcast), the number of copies need to be sent via the inter-subsystem unicasting for each multicast packet is no larger than the total number of subsystems. Moreover, since the λ-router can provide non-blocking routing for any connected core, intra-subsystem multicast will not lead to heavy congestion even using multiple wavelengths at the same; while since the inter-subsystem multicast traffic is converged and each copy is transmitted independently similar to an unicast packet, the load balance can also be achieved through uniformly selecting a sibling gateway in the routing path.

3.3.3 Wavelength-Level Flow Control

Since the buffer size of each input is strictly limited [33], a wavelength-level credit-based flow control scheme is used in gateways to prevent the buffer overflow. When the buffer queue for a specific wavelength is full, the gateway postpones any request for using this wavelength until there is a new vacancy for the incoming packet. Moreover, since there is no cyclic dependent path in WRH-ONoC and every packet can be delivered to its destination in a limited number of hops, it is deadlock- and livelock-free.

4 Modelling and Analysis

This section theoretically analyses the performance with a typical traffic pattern which follows Uniform distribution in space and Poisson distribution in time. Because of the inter-subsystem unicasting property of multicast scheme, we only analyse the case of unicast communication.

4.1 Network Hierarchy

To interconnect N cores using $W_{\max}$ wavelengths, we assume the hierarchy of λ-routers has L levels, and each λ-router at level i connects to a level i+1 λ-router via g sibling gateways, where $1 \le i < L$. The following theorem gives the minimum number of λ-routers required at each level to interconnect all the cores and gateways.

Theorem 1. *The minimum number of λ-routers required at level i, denoted by R_i, can be calculated as*

$$R_i = \begin{cases} \lceil \frac{N}{W_{\max}-g} \rceil, & i = 1; \\ \lceil \frac{gR_{i-1}}{W_{\max}-g} \rceil, & i \in (1, L); \\ 1, & i = L. \end{cases} \quad (1)$$

Proof: Since each λ-router can directly connect to $W_{\max}-g$ cores and g gateways using $W_{\max}$ wavelengths, the minimum number of λ-routers required at bottom level is $\lceil \frac{N}{W_{\max}-g} \rceil$. For λ-routers at level i, they need to connect to R_{i-1} λ-routers at level $i-1$ via gR_{i-1} gateways. Then if $gR_{i-1} \le W_{\max}$, level i is the top level L and only one λ-router is required, $R_L = 1$; otherwise more than one λ-routers are required, and each λ-router at level i also needs to connect to one λ-router at the higher level via g sibling gateways. Thus, gR_{i-1}+$gR_i \le R_i W_{\max}$, and we have $R_i \ge \frac{gR_{i-1}}{W_{\max}-g}$. Hence, the minimum number of λ-routers required at level i where $1 < i < L$, is $R_i = \lceil \frac{gR_{i-1}}{W_{\max}-g} \rceil$. □

From Theorem 1, we can achieve the minimum number of λ-routers and gateways required for N cores with $W_{\max}$ wavelengths, details in our conference paper [1].

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

4.2 Communication Delay

In our analysis, the communication delay is defined as the amount of time taken for a packet traversing through the hierarchical network from the source core to its destination. Depending on the location of destination cores, different packets go through different number of hops. Let $\overline{D}$ denote the average communication delay, we have:

$$\overline{D} = \alpha \overline{D}_{intra} + (1-\alpha)\overline{D}_{inter}, \tag{2}$$

where $\overline{D}_{intra}$ and $\overline{D}_{inter}$ represent the average delivery delays for intra-subsystem traffic and inter-subsystem traffic, respectively. α is the proportion of intra-subsystem traffic.

For intra-subsystem traffic, each packet only traverses a specific λ-router at bottom level. Due to the high-speed optical transmission and nonblocking routing, any packet from different inputs to different outputs has the same delay to pass through the λ-router. Thus, the delay for an intra-subsystem packet is constant and can be computed as:

$$\overline{D}_{intra} = D_{E/O} + D_{\lambda R} + D_{O/E}, \tag{3}$$

where $D_{E/O}$ and $D_{O/E}$ are signal conversion delays, and $D_{\lambda R}$ is the propagation delay over one hop of λ-router.

For inter-subsystem traffic, the average communication delay $\overline{D}_{inter}$ can be computed as:

$$\overline{D}_{inter} = D_{E/O} + \underbrace{\overline{N}_{hop} \times D_{\lambda R}}_{I} + \underbrace{(\overline{N}_{hop} - 1) \times D_{GW}}_{II} + \overline{D}_Q + D_{O/E}, \tag{4}$$

where $\overline{N}_{hop}$ is the expected number of hops that a packet needs to traverse. Part I of Eq. (4) is the expected accumulated delay for the packet to traverse all λ-routers on the routing path. Part II of Eq. (4) is the accumulated delay incurred by packet processing in gateways excluding the queuing delay, where D_{GW} is the packet processing delay in a gateway and it can be modelled as:

$$D_{GW} = D_{E/O} + 2D_{buf} + D_{xbar} + D_{wl} + D_{O/E}, \tag{5}$$

where D_{buf} is the delay for reading and writing the sliced buffer, D_{xbar} is the delay for copying a packet from input to output through an internal crossbar, and D_{wl} is the delay incurred by looking up the wavelength routing matrix to get the right wavelength for the next hop. $\overline{D}_Q$ in Eq. (4) is the expected accumulated queuing delay in all the gateways which are on the routing path. Actually, $\overline{D}_Q$ is the only uncertain part of packet delay incurred by gateways. We model it separately so that Part II only depends on the average number of hops $\overline{N}_{hop}$.

From Eq. (4), $\overline{D}_{inter}$ is a function of $\overline{N}_{hop}$ and $\overline{D}_Q$, which depend on network structure and traffic patterns. In the following, we model $\overline{D}_{inter}$ for WRH-ONoC with L levels of λ-routers and g sibling gateways under Uniform-Poisson traffic. Assume each core generates packets following Poisson distribution with the same traffic rate θ, and each core sends packets to other cores with the same probability. Hence, the traffic rate from any core i to other core j is $\frac{\theta}{N-1}$, and the proportion of intra-subsystem traffic α is $\frac{W_{\max}-g-1}{N-1}$.

Lemma 1. *The probability that an inter-subsystem packet traverses $2i-1$ hops of λ-routers is $P(2i-1) = (\frac{1}{R_i} - \frac{1}{R_{i-1}}) \times \frac{N}{N-1}$, where $i = 2, ..., L$. The expected number of hops that an inter-subsystem packet traverses is $\overline{N}_{hop} = \sum_{i=2}^{L}(2i-1)P(2i-1)$.*

The detailed proof of Lemma 1 is given in our conference paper [1]. In our design, each input buffer slice in the gateway is a separate FIFO queue. To derive the maximum injection rate and the relationship between the gateway buffer size and traffic rate, we first assume the infinite input buffer in following analysis. Queuing delay with limited buffer size is analysed in Appendix and evaluated in our simulations. Hence, each data path in the gateway is an independent M/M/1 queuing system. We also assume the routing structure as illustrated in Fig. 5 is a balanced and complete tree, so that in each level all λ-routers connect to the same number of cores or gateways and every input/output of λ-routers are used. The following gives the expected end-to-end queuing delay for an inter-subsystem packet. If the routing tree is imbalanced or incomplete, *i.e.*, the network is not fully utilized, Theorem 2 gives an upper bound of the queuing delay.

Theorem 2. *The average accumulated packet queuing delay is*

$$\overline{D}_Q = \sum_{i=2}^{L} \left(P(2i-1) \times \sum_{j=1}^{i-1} \frac{2\theta_j}{\mu_j(\mu_j - \theta_j)} \right), \tag{6}$$

where $\theta_j = \frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2} \times \frac{\theta}{g(W_{\max}-g)}$ and $\mu_j = \frac{s_p}{t_d}$. s_p is the packet size and $t_d = 2D_{buf} + D_{xbar} + D_{wl} + D_{E/O}$.

Proof: For each inter-subsystem packet that traverses $2i-1$ hops of λ-routers, the packet is routed by $i-1$ gateways in both upward and downward directions. Let T_i be the queuing delay in the i^{th} hop gateway. According to Lemma 1, the probability that a packet passes $2i-1$ hops of λ-routers is $P(2i-1)$. Then, the expected accumulated queuing delay $\overline{D}_Q$ can be computed by

$$\overline{D}_Q = \sum_{i=2}^{L} \left(P(2i-1) \times \sum_{j=1}^{2i-2} T_j \right). \tag{7}$$

Since the upward and downward traffics are processed separately in two independent data paths, we analyse the queuing delay T_i of upward and downward respectively.

(I) T_i in Upward Direction: Since each core generates traffic following Poisson distribution in time and Uniform distribution in space with the same traffic rate θ, the traffic injection rate in a gateway that bridges a level $j-1$ λ-router and a level j λ-router can be computed as follows: the traffic rate from a level $j-1$ λ-router to a level j λ-router is $\theta \times \frac{N}{R_{j-1}} \times \frac{N - \frac{N}{R_{j-1}}}{N-1} = \theta \times \frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2}$. Since the packets from a level $j-1$ λ-router to a level j λ-router are evenly routed via g sibling gateways that connect to these two λ-routers, the traffic injection rate in each gateway that connects a level $j-1$ λ-router to a level j λ-router is $\frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2} \times \frac{\theta}{g}$. According to our gateway design, each gateway needs to have $W_{\max}-g$ input queues to buffer the packets received from the level $j-1$ λ-router using $W_{\max}-g$ wavelengths, since there is no communication among g sibling gateways. Thus, the upward traffic injected to each gateway is dispatched to $W_{\max}-g$ input queues following Uniform distribution. Let θ_j denote the traffic injection rate at one input queue

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

in a gateway that connects a level $j-1$ λ-router to a level j λ-router. We have $\theta_j = \frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2} \times \frac{\theta}{g(W_{\max}-g)}$.

In the gateway, input queues are fully connected with output buffers, and each input queue has an independent packet dispatcher. Let t_d represent the packet dispatch delay which is defined as the average time interval between two adjacent packets that are sent out from the same output buffer. Then $t_d = 2D_{buf} + D_{xbar} + D_{wl} + D_{E/O}$. Each input queue can be modelled as a FIFO queuing system with the injection rate of θ_j and service rate of $\mu_j = s_p/t_d$ where s_p is the average packet size. Each input queue is subjected to the Birth-Death process according to the queuing theory. Assume the probability that q packets stay in the queuing system is $P_j(q)$ with an initial state of $P_j(0)$, and we have $P_j(q) = P_j(0) \times (\frac{\theta_j}{\mu_j})^q$ for $q \geq 0$. To achieve a stable queuing system, we should guarantee $\frac{\theta_j}{\mu_j} < 1$, and thus the stable average queue length, denoted by Q_j, is $Q_j = \sum_2^\infty [(q-1)P_j(q)] = \sum_2^\infty [(q-1)(1-\frac{\theta_j}{\mu_j})(\frac{\theta_j}{\mu_j})^q] = \frac{\theta_j^2}{\mu_j(\mu_j-\theta_j)}$. According to the Little's Law theorem, the stable queue delay of each input queue is $T_j = \frac{Q_j}{\theta_j} = \frac{\theta_j}{\mu_j(\mu_j-\theta_j)}$.

(II) T_i in Downward Direction: Similar to upward direction, the traffic rate from a level j λ-router to a level $j-1$ λ-router is $\theta \times (N - \frac{N}{R_{j-1}}) \times \frac{\frac{N}{R_{j-1}}}{N-1} = \theta \times \frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2}$, which is equal to the traffic rate from a level $j-1$ λ-router to a level j λ-router because the downward process is symmetrical to upward process. Since the routing of downward packets is similar to upward packets through another independent path, we have $T_k = T_{2i-k-1}$, $k \in [1, i-1]$. Hence, $\overline{D}_Q = \sum_{i=2}^{L} \left(P(2i-1) \times \sum_{j=1}^{i-1} 2T_j \right) = \sum_{i=2}^{L} \left(P(2i-1) \times \sum_{j=1}^{i-1} \frac{2\theta_j}{\mu_j(\mu_j-\theta_j)} \right)$, where $\theta_j = \frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2} \times \frac{\theta}{g(W_{\max}-g)}$, and $\mu_j = s_p/t_d$. □

Corollary 1. *If WRH-ONoC uses the minimal number of λ-routers and in each λ-router all available wavelengths are fully utilized, the maximal data rate can be achieved is $\theta_{\max} = \frac{s_p(N-1){W_{\max}}^2}{t_d N^2}$, which only relates to the number of cores N, number of available wavelengths $W_{\max}$, and delay in gateway $\frac{s_p}{t_d}$.*

Proof: In order to ensure the network stability, each input queue in a gateway between level $j-1$ and level j should satisfy $\frac{\theta_j}{\mu_j} < 1$, *i.e.*, $\frac{N^2}{N-1} \times \frac{R_{j-1}-1}{R_{j-1}^2} \times \frac{\theta}{g(W_{\max}-g)} < \frac{s_p}{t_d}$. Then, we can have $\theta < \min_{j \in [2,L]} \frac{s_p g(W_{\max}-g)}{t_d} \times \frac{(N-1)R_{j-1}^2}{N^2(R_{j-1}-1)}$. When $1 \leq R_j < R_{j-1}$, we always have $\frac{R_j^2}{R_j-1} < \frac{R_{j-1}^2}{R_{j-1}-1}$ for $\forall j \in [1, L]$. Thus, $\theta_{\max} = \frac{s_p g(W_{\max}-g)}{t_d} \times \frac{(N-1)R_{L-1}^2}{N^2(R_{L-1}-1)}$, when $j = L$. If WRH-ONoC is constructed by using the minimal number of λ-routers and in each λ-router all available wavelengths are fully utilized, the level L λ-router interconnects with $R_{L-1} = \frac{W_{\max}}{g}$ level $(L-1)$ λ-routers. Hence, $\theta_{\max} = \frac{s_p g(W_{\max}-g)}{t_d} \times \frac{(N-1)}{N^2} \times (\frac{W_{\max}}{g})^2 \times \frac{g}{W_{\max}-g} = \frac{s_p(N-1){W_{\max}}^2}{t_d N^2}$, so it only relates to the number of cores N, number of available wavelengths $W_{\max}$, and gateway processing delay $\frac{s_p}{t_d}$. □

Corollary 2. *If data rate is $\theta = t\theta_{\max}$, $0 < t < 1$, the maximum buffer size required in a gateway is $\lceil \frac{t^2}{1-t} \rceil$ in packets/wavelength.*

Proof: According to Corollary 1, if the traffic rate $\theta \leq \theta_{\max}$, we always have $\theta_j \leq \mu_j$ in any gateway at level j, $\mu_j = \frac{s_p}{t_d}$. Since θ_j is linearly related to the data rate θ, then if $\theta = t\theta_{\max}$, $0 < t < 1$, we have $\theta_j \leq t\mu_j$ in gateways at level j. According to the proof for Theorem 2, we have the average queue length for each buffer slice in a level j gateway to be $Q_j = \frac{\theta_j^2}{\mu_j(\mu_j-\theta_j)}$. Thus, if the actual data rate is $\theta = t\theta_{\max}$, we have the maximum buffer size required in the gateway to be $Q_{\max} = \lceil \frac{(t\mu_j)^2}{\mu_j(\mu_j - t\mu_j)} \rceil = \lceil \frac{t^2}{1-t} \rceil$ in *packets/wavelength*. □

4.3 Energy Consumption

Let $\overline{E}$ represent the average energy consumption for transmitting a packet. It can also be computed from two aspects:

$$\overline{E} = \alpha \overline{E}_{intra} + (1-\alpha)\overline{E}_{inter}, \tag{8}$$

where $\overline{E}_{intra}$ and $\overline{E}_{inter}$ stand for the average energy consumption for intra- and inter-subsystem traffics, and α is the proportion of intra-subsystem traffic.

The average energy consumption for intra-subsystem traffic $\overline{E}_{intra}$ is composed of the energy consumption for data modulation $E_{E/O}$ and demodulation $E_{O/E}$, and the optical power provided by the light source E_{ls}. We have:

$$\overline{E}_{intra} = E_{E/O} + E_{O/E} + E_{ls}. \tag{9}$$

With a given data rate, $E_{E/O}$ and $E_{O/E}$ are constant for specific O/E and E/O converters. Meanwhile, the light source should provide sufficient optical power to ensure the photodetector can correctly decode the optical signals. E_{ls} is determined by the light source efficiency η, insertion loss of waveguides and λ-routers IL, and photodetector sensitivity P_d [9]. Since a separate light source is provide for each core and gateway, with given clock frequency of f_{clk}, E_{ls} can be:

$$E_{ls} = \frac{1}{\eta f_{clk}} \times P_d \times 10^{\frac{IL}{10}}, \tag{10}$$

where the insertion loss IL (in dB) is the power attenuation of optical signal when it propagates through all the optical devices. In our design, insertion loss occurs in waveguides (transmit through and cross) and λ-routers (drop in or pass an MR). Note that WRH-ONoC only uses passive MRs in λ-routers and the resonant wavelengths are determined by geometric parameters, thus it has no static energy cost for tuning MRs like in ONoC using active MRs [9].

The average energy consumption for inter-subsystem communication includes an electrical part (*i.e.*, gateway) and an optical part (*i.e.*, light source, O/E and E/O converters) as shown in Eq.(11). For the electrical part, each packet transmits through $\overline{N}_{hop} - 1$ hops of gateways. For the optical part, each packet should pass $\overline{N}_{hop}$ λ-routers, and experience $\overline{N}_{hop}$ times of O/E and E/O conversions in the interface of source/destination core and the gateways.

$$\overline{E}_{inter} = \overline{E}_{GW}(\overline{N}_{hop} - 1) + (E_{E/O} + E_{O/E} + E_{ls})\overline{N}_{hop}. \tag{11}$$

The energy consumption of gateway $\overline{E}_{GW}$ is caused by the sliced input/output buffers, wavelength dispatchers, crossbars, and static energy cost (*e.g.*, leakage).

5 Performance Evaluation

This section evaluates the performance extensively with both real data traces and synthetic traffic patterns. Four typical hierarchical ONoC schemes are also compared, including PNoC [12], HOME [28], ATAC [26], and Firefly [27].

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

TABLE 1. Simulation Settings

Clock frequency	1 *GHz*	Packet size	64 *bits*
Channel bandwidth	10 *Gbps/wl*	Gateway delay	5 *cycles*
λ-router delay	8 *stages/cycle*	Router delay	2 *cycles*
Router/crossbar delay	8 *hops/cycle*	Overall buffer size	*Equal*

5.1 Simulation Setup

To make fair comparison, we implement WRH-ONoC and all other schemes based on Noxim [34]. Table 1 summarizes the parameter settings. All optical devices, including E/O&O/E converters and routers, work with a bandwidth of 10 *Gbps* for each wavelength channel [21]. All electrical devices use a system clock of 1 *GHz*, thus each clock cycle is 1 *ns*. Each packet has a constant size of 64 *bits*, the same size of a control packet for cache coherence [35]. Thus, multiple successive data packets need to be transmitted for a single cache line, *e.g.*, 64 bytes. According to [36], the optical signal can go through 8 routers in an optical mesh, thus in our simulation we also configure the delay for optical signal passing through a λ-router to 8 *stages/cycle* and 8 *hops/cycle* in ring-like crossbars for fair comparison. The processing delay of gateway incurred by buffering, wavelength look-up, and packet dispatching takes 5 *cycles* for every packet. While in other schemes, the processing delay of electrical router is set to their minimum, 2 *cycles*. Since Noxim is a simulator for mesh-based electrical NoCs, we use it to simulate the path setup process for PNoC and HOME, and intra-cluster packet delivery for ATAC and Firefly. Thus, the end-to-end packet delay is the electrical network delay plus the optical delivery delay. Moreover, since the buffer size in gateways and electrical routers have significant influence on energy consumption and area cost, we configure the overall buffer size in each gateway and electrical router to be exactly equal. To ensure the accuracy of simulation results, each simulation run lasts for 500,000 *cycles* with 10,000 *cycles* for warmup. The λ-router scheme was not simulated since it is nonblocking with a constant transmission delay between any pair of connected cores.

5.2 Comparison with Theoretical Results

Fig. 7 compares the average end-to-end delay obtained from the theoretical analysis and simulation results with the variation of average data rate when the buffer size of gateway is set to infinite. This simulation indicates the upper-bound performance for a certain network configuration. We conduct two sets of simulations in which $\{N, W_{\max}, g\}$ are configured to $\{400, 25, 5\}$ and $\{480, 30, 6\}$, respectively. In the simulations, we set the buffer size to be 1000 packets

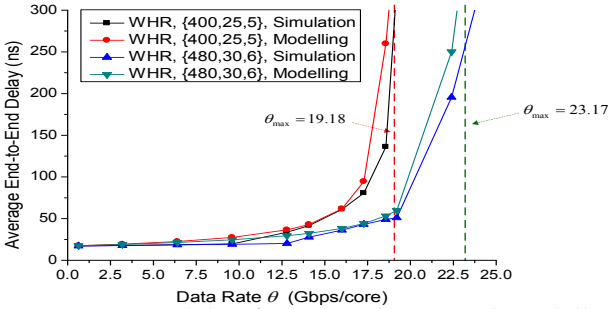


Fig. 7: Average delay from simulation and modelling.

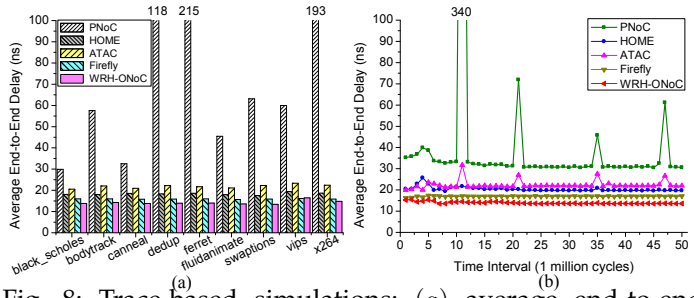


Fig. 8: Trace-based simulations: (a) average end-to-end delay; (b) delay variations over time (*black_scholes* trace).

to approximate the infinite buffer. It can be seen that the average delay measured in simulations keeps close to the theoretical results. When the average data rate θ is small (*e.g.*, $\theta \leq 10$ *Gbps/core*), the average end-to-end delay is also small ($\sim$25 *ns*) and remains stable because most of the packets do not experience much queuing delay in the gateways. When the average data rate θ approaches the maximum data rate $\theta_{\max}$, the average end-to-end delay increases dramatically because the network becomes saturated and most of the packets need to be queued. This also validates the correctness of our theoretical analysis.

5.3 Simulation with Data Traces

We firstly evaluate the performance using real data traces. The data traces are obtained from Netrace [37] which constructs a 64-core system running PARSEC benchmarks with 2-level cache and MESI coherence protocol. Even though the average data rates in these traces are low (less than 0.24 *packets/cycle*), they contain bursty (massive data in a short time) and multicast traffics. For 64 cores, WRH-ONoC is configured to a 2-level hierarchical network in the simulations, with $\{N, W_{\max}, g\}$=$\{64, 20, 4\}$, *i.e.*, 16 cores in each subsystem, 20 wavelengths used in λ-routers, and 4 sibling gateways between two λ-routers. We also compare WRH-ONoC with other schemes. For the regularity, PNoC is configured to an 8×8 mesh network, and HOME uses a 4×4 mesh with four cores share one optical router. These two schemes do not employ WDM and need to reserve the optical path by the electrical network in advance. ATAC and Firefly are configured with 16 clusters, 4 cores for each cluster in a 2×2 electrical mesh. In ATAC each cluster has a fixed point to access the global optical ring, while in Firefly the cores at the same relative position of different clusters are connected by an optical ring thus with higher path diversity compared with the ATAC scheme.

Fig. 8(a) shows the average end-to-end delay by running different data traces. We can see WRH-ONoC achieves the lowest packet delay ($\sim$15 *ns*) compared with the other schemes. It is worth noting the average end-to-end delay achieved by WRH-ONoC does not vary much for different traces. That indicates the communication capacity of WRH-ONoC can accommodate the traffic variations in all these traces, since it can process multiple communications in parallel through different channels via wavelength multiplexing and has better load-balance capability to deal with the bursty traffic and multicast traffic. Since PNoC requires the end-to-end optical path reservation and has longer

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

communication distance using 8×8 mesh, its packet delay is much higher and varies a lot even with a low data rate. Since HOME has lower communication distance within a 4×4 mesh, the path reservation delay is not significant. In ATAC, the packets need to be transmitted in an electrical network at both the source and destination sides, and all of the cores need to compete for the global optical ring, hence its average delay is a little higher. Firefly can also achieve better performance compared to others by using more optical rings, in which the cores at the same position of each cluster are connected by a separate optical ring.

To further demonstrate the detailed delay variations, Fig. 8(b) shows the short-term average delay that during a fixed time interval of 1 million *cycles* for these schemes using *black_scholes* data trace. It can be seen the average delay for WRH-ONoC remains low and stable, while PNoC has lots of fluctuation in average delay over time. Even though ATAC and Firefly schemes also have a low fluctuation, their average packet delays are a little higher than WRH-ONoC.

In trace based simulations, since there are only 64 cores and the average data rate is low, the performance improvement of WRH-ONoC over other schemes is not significant. Thus, we will evaluate WRH-ONoC in larger network sizes and with variable data rates in synthetic traffic patterns.

5.4 Simulation with Synthetic Traffic

We use synthetic traffic to further evaluate the performance and scalability of large-scale WRH-ONoCs, which interconnect more than hundreds of cores. With the synthetic traffic, we can evaluate the saturation data rate and the maximum throughput by gradually increasing the average data rate of each core. Also the performance for unicast and multicast communication can be simulated separately.

5.4.1 Comparison with Existing Schemes

We use the unicast traffic which follows Uniform-Poisson pattern to evaluate the performance of different ONoC schemes. The total number of cores N is 400. In WRH-ONoC, the cores are divided into 20 subsystems and organized as $\{N, W_{\max}, g\}$=$\{400, 21, 1\}$. In this configuration, WRH-ONoC has 2 levels of λ-routers with g=1 sibling gateway. Meanwhile, for regularity, in PNoC 400 cores are connected by a 20×20 electrical-optical mesh network, while HOME interconnects all the four-core clusters using a 10×10 electrical-optical mesh. PNoC and HOME use only one wavelength without WDM. In ATAC and Firefly, 400 cores are grouped into 25 clusters connected by WDM-based optical crossbar with a little higher maximum number of wavelengths than WRH-ONoC (25 to 21), and 4×4 electrical mesh for the intra-cluster communication.

Fig. 9(a) shows the average packet delay with the variation of average data rate. It can be seen that our scheme can achieve the lowest end-to-end delay and highest saturation data rate. When the data rate θ is small, most of packets do not suffer from queuing delay in the routers and gateways. We use *zero-load delay*, which is the delay without contention and is a lower bound of the average delay, to compare these schemes. Since WRH-ONoC has only 2

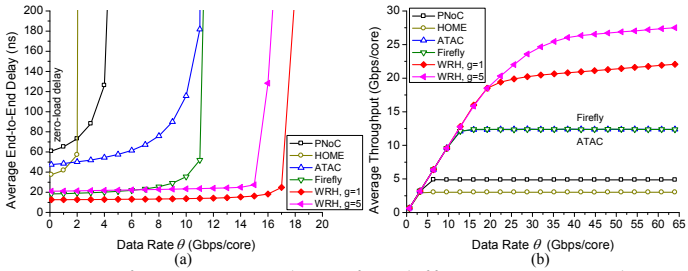


Fig. 9: Performance analysis for different ONoC schemes: (a) average end-to-end delay; (b) throughput per core.

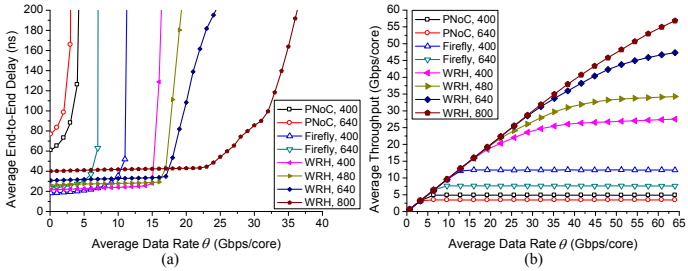


Fig. 10: Performance analysis with different network sizes: (a) average end-to-end delay; (b) throughput per core.

levels of λ-routers when g=1 and most of packets do not experience many wavelength reassignments in gateways, it can obtain the lowest zero-load delay, around 12.6 *ns*. The zero-load delay of Firefly is also low, about 18.4 *ns*, owing to its low average distance and path diversity. But when compared with other schemes, WRH-ONoC has significant advantages. PNoC has the highest zero-load delay, about 60.8 *ns*, due to the large average length of routing path and the time-consuming path reservation. The zero-load delays of HOME and ATAC are 37.4 *ns* and 47.5 *ns*. *Hence, WRH-ONoC has at least* 46.0% *of reduction on the zero-load delay even in comparison with Firefly, the best of other schemes.*

As the data rate increases, the average end-to-end delay increases as well because of the increasing contentions occurred in the electrical routers or gateways. Whereas, our scheme can achieve very low end-to-end delay (<20 *ns*) even with the data rate of 16 *Gbps/core*. HOME gets saturated quickly at 2 *Gbps/core* due to heavy traffic loads on the global mesh network with 4 cores in a cluster sharing one access point. PNoC saturates at around 4 *Gbps/core* due to the path reservation and poor scalability of mesh. ATAC and Firefly achieve higher saturation data rate of ∼11 *Gbps/core* since they both utilize the optical ring for all 25 clusters. The saturation data rate of WRH-ONoC (∼17 *Gbps/core*) is about 8.5, 4.3, 1.55 and 1.55 times of that for HOME, PNoC, ATAC and Firefly owing to the nonblocking property of λ-routers and the fast wavelength assignment in gateways. Fig. 9(b) shows the average throughput and our scheme can achieve much higher throughput than others, ∼22.1 *Gbps/core*. *Generally, WRH-ONoC has at least* 72.7% *of improvement on throughput, compared to the best of other schemes (12.8 Gbps/core in Firefly).*

5.4.2 Impact of Network Size

We evaluate the scalability of WRH-ONoC using different network sizes. Four groups of simulations are carried out with 400, 480, 640, and 800 cores. PNoC and Firefly are

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

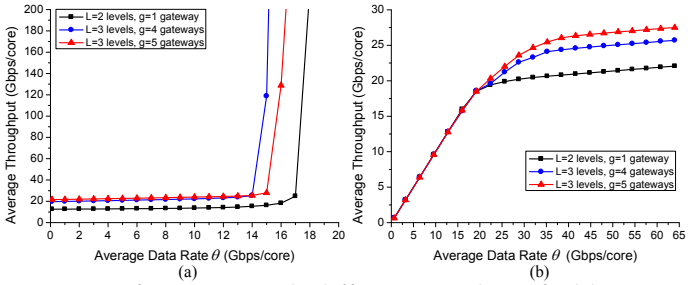


Fig. 11: Performance with different number of sibling gateways: (a) average packet delay; (b) throughput per core.

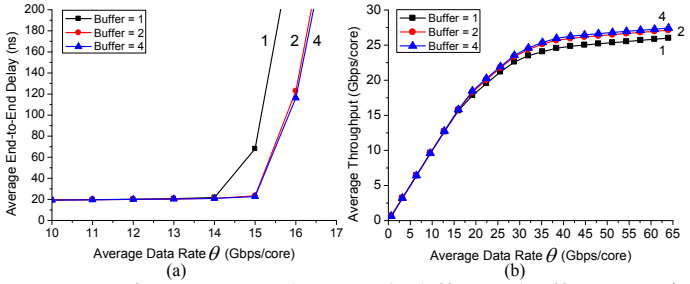


Fig. 12: Performance analysis with different buffer sizes: (a) average end-to-end packet delay; (b) throughput per core.

also compared with 400 and 640 cores. PNoC is configured as 20×20 and 20×32 mesh networks. Firefly uses the same number of wavelengths as WRH-ONoC in the global optical ring, and 16 cores in each cluster are connected by 4×4 mesh. As shown in Fig. 10, our scheme can achieve much better performance than PNoC and Firefly when connecting the same number of cores. It is worth noting that (i) the performance of PNoC and Firefly deteriorate as the network size expands. That is because the linearly increased average hops and severe contentions in the path reservation for PNoC, and the inefficient electrical local network and contentions on the optical global network for Firefly. Similar results can be achieved in HOME and ATAC; (ii) benefiting from nonblocking λ-routers and load balance among the sibling gateways, our scheme can achieve higher saturation data rate and average throughput for larger scale systems by increasing the number of wavelengths and gateways. For instance, when $W_{\max}$ is increased from 25 to 50 and g is increased from 5 to 10, the saturation data rate and average throughput are more than doubled, even though the number of cores is also doubled (from 400 to 800). This is because the proportion of intra-subsystem traffic increases with the increase of $W_{\max}$, and the paralleled data paths between λ-routers increases with the increase of g. For Firefly, even the optical network can provide higher bandwidth with more wavelengths, the electrical local network will become its bottleneck.

5.4.3 Impact of Gateway and Buffer Size

Fig. 11 shows the performance with different number of sibling gateways. Simulations are carried out for a system with 400 cores. WRH-ONoC has 2 levels of λ-routers when g=1, and 3 levels when $g=4$ and 5. It can be seen that when $g=1$, WRH-ONoC has less levels of λ-routers, thus it can achieve lower packet delay due to less wavelength reassignments, as shown in Fig. 11(a); when $g=4$ and 5, WRH-ONoC has

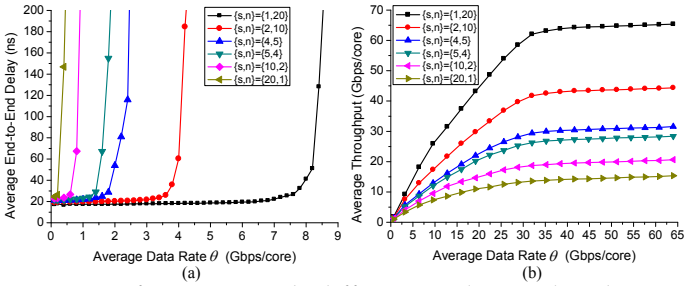


Fig. 13: Performance with different multicast distributions: (a) average end-to-end delay; (b) average throughput.

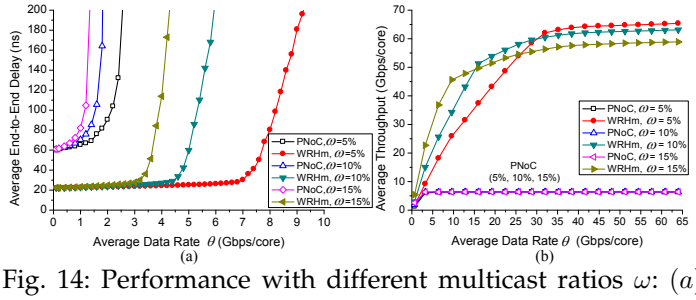


Fig. 14: Performance with different multicast ratios ω: (a) average end-to-end delay; (b) average throughput.

more alternative channels and better load balance ability, thus it can achieve higher saturation throughput, as shown in Fig. 11(b). Hence,*WRH-ONoC can be configured, according to the performance requirement of specific applications, on the number of sibling gateways to achieve a better tradeoff between the average packet delay and throughput*.

Fig. 12 shows the influence of gateway's buffer size. Simulations are carried out for a system with $\{N, W_{\max}, g\}$= $\{400, 25, 5\}$. The buffer size in each gateway increases from 1 to 4 *packets/wavelength*. It can be seen that the performance is improved by increasing the buffer size from 1 to 2, because larger packet buffer in the gateways can better deal with the bursty traffic. However, further increasing the buffer size to 4 does not greatly enhance the performance. This is because the average queuing delay in gateways is very small when the network is not saturated due to the high-speed transmission in λ-routers and the paralleled packet dispatching in gateways. Even with a large buffer, in most cases only a small portion of the buffer is used.

5.4.4 Impact of Multicast Traffic

We evaluate the performance of multicast routing with the parametrized destination's distribution and multicast traffic ratio. In our simulation, the destinations of each multicast packet can be distributed in different subsystems. The multicast traffic ratio, denoted by ω, is the proportion of multicast packets over the total number of packets.

The simulation results in Fig. 13 reveal the influence of different distributions of destinations for a 400-core system with $\{N, W_{\max}, g\}$=$\{400, 25, 5\}$. The traffic follows Uniform-Poisson distribution in which $\omega = 5\%$ of traffic is multicast. The parameters $\{s, n\}$ indicate the destinations for each multicast are randomly distributed in s different subsystems with n randomly chosen cores per subsystem. In the simulation, we set the total number of destinations for a multicast packet to 20, which is the number of cores in a subsystem, so that 6 combinations of s and n can be

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

evaluated. WRHm-ONoC can achieve much lower delay and much higher throughput by placing the destinations in as less subsystems as possible, *e.g.*, the same subsystem or neighbouring ones, because less packet copies are required to transmit in the hierarchical network. Very low packet delay (∼20 *ns*) is achieved when all 20 destination cores are in the same subsystem with up to 8 *Gbps/core* data rate. Since for a multicast packet, as many as 20 packet copies are generated at last-hop gateways, the saturated data rate is relatively lower compared to unicast simulations, while saturated throughput is higher due to multiple copies.

Fig. 14 shows the performance with different multicast traffic ratios. We also compare WRHm-ONoC with PNoC which uses a tree-based multicast routing [32]. WRHm-ONoC is configured with $\{N, W_{\max}, g\} = \{400, 25, 5\}$, while PNoC is in a 20×20 mesh. For each multicast packet, there are 20 randomly chosen destinations. The traffic follows Uniform-Poisson distribution with different proportions of multicast traffic. It can be seen that, as ω increases, the maximum data rate and throughput decrease significantly owing to much heavier bursting multicast traffic especially with as many as 20 destinations. However, WRHm-ONoC can always achieve better performance than PNoC, *over 3 times*. These results can also be achieved from HOME, ATAC, and Firefly, especially since multicast traffic can lead to much heavier congestion on their global optical network.

5.5 Hardware Cost Analysis

In Table 2, we further analyse the requirement of optical devices (in MRs) of WRH-ONoC and make a comparison with λ-router, since WRH-ONoC is proposed on the basis of λ-router and to solve its scalability problem for large-scale manycore systems. From this table, it can be seen that the number of wavelengths, MRs required in E-O converters (O/E&E/O) and λ-routers in WRH-ONoC are much smaller than those in a global λ-router, more than 90% of reduction, (details of the calculation are in our conference paper [1]). This is because each core in the λ-router needs to directly connect to all the other cores, while in WRH-ONoC each core only directly connects with the cores or gateways in the same subsystem. Even though the use of gateways will increase some hardware costs on the electrical devices, the number of gateways is much fewer than the connected cores, *e.g.*, 20 gateways for a 400-core system when $W_{\max}$=25, g=1, and the buffer space is much smaller than in the traditional electrical routers due to the high-speed optical transmission. Without mature 3D integration technology, all the electrical gateways can be integrated in a separate layer. In a manycore processor chip with a size of 20×20 mm^2, the overall footprint of each gateway can be less than 0.16 mm^2 with 45 *nm* process according to Orion 3.0 [38].

Table 3 shows the comparison on hardware cost between WRH-ONoC and the other schemes by using Orion 3.0 [38], which is often used to estimate the chip area of electrical routers for Network on Chips with the commercial-popular 45 *nm* process. According to the table, WRH-ONoC consumes only the second smallest area of chip among existing schemes even when g=5. For the electrical part, PNoC, ATAC, and Firefly need as many as 400 routers, while WRH-ONoC needs only 125 gateways when g=5 with larger crossbars (25×25). Due to the fast wavelength reassignment and abundant optical channels, WRH-ONoC has significantly reduced the size of buffers which could potentially consume a lot of chip area. However, in our simulation, to make fair comparison, we set the overall buffer size for the gateways and routers to be exactly the same among all compared schemes, a decision that has somewhat disadvantaged WRH-ONoC. It can be seen that even though WRH-ONoC uses larger crossbar, the total area cost is smaller than most other schemes due to fewer gateways used. The overall chip area of our scheme using 5 sibling gateways (the second last row in the table) is only higher than HOME which has the lowest throughput capacity. Moreover, if we decrease the number of sibling gateways from 5 to 1, the number of gateways will be significantly reduced and the overall area cost is only about 2.41 mm^2, which is the smallest among all the schemes (see the last row in the table). For the optical part, since without available synthesis tool at present, we estimate the chip area by assuming each MR takes 3×3 μm^2 and each OSE takes 10×10 μm^2 according to [22]. PNoC and HOME use a large number of small optical routers in mesh topology, while ATAC and Firefly use a large optical ring for global interconnection. WRH-ONoC uses multiple λ-routers and thus more optical devices. However, since the overall chip area of optical devices only occupies a small proportion on an ONoC, WRH-ONoC still consumes much less chip area.

From the comparison in Table 3 and simulation results in previous subsections, it can be seen WRH-ONoC can achieve better tradeoff between the hardware cost and communication performance than the other existing schemes.

TABLE 2. Comparison the Requirement on MRs to λ-router

Architecture	Configuration			Requirements of MRs			
	N	$W_{\max}$	g	in E-Os	Reduction	in Routers	Reduction
λ-router	400	400	-	159600	-	159200	-
WRH-ONoC	400	25	5	14600	90.85%	13950	91.23%
λ-router	480	480	-	229920	-	229440	-
WRH-ONoC	480	30	6	21120	90.81%	20340	91.13%
λ-router	640	640	-	408960	-	408320	-
WRH-ONoC	640	40	8	37760	90.77%	36720	91.01%

TABLE 3. Hardware Costs Comparison (area in mm^2)

Scheme	$W_{\max}$	Electrical router/gateway				Optical router/crossbar			
		count	size	per area	sum area	count	size	per area	sum area
PNoC	1	400	5×5	0.0606	24.23	400	5×5	0.0015	0.6
HOME	1	100	8×8	0.0782	7.82	100	5×5	0.0015	0.15
ATAC	25	400	5×5	0.0606	24.23	1	25×25	0.0325	0.0325
Firefly	25	400	6×6	0.0665	26.59	8	25×25	0.0325	0.26
WRH (g=5)	25	125	25×25	0.1614	20.17	26	25×25	0.0325	0.845
WRH (g=1)	21	20	21×21	0.1204	2.41	21	21×21	0.0221	0.4641

5.6 Energy Efficiency

We evaluate the energy efficiency of WRH-ONoC with data traces and synthetic traffics, and compare it with the other schemes. The energy efficiency is defined as the amount of energy consumed for transmitting one packet in average, in *pJ/packet*. In the simulation, we consider both the dynamic energy consumption (*e.g.*, routing, forwarding) and the static energy consumption (*e.g.*, leakage). Table 4 lists the

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEI Transactions on Sustainable Computing

TABLE 4. Optical Energy Parameters

Parameter	Value	Unit	Parameter	Value	Unit
Laser efficiency	30 [14]	%	Receiver	100 [9]	*fJ/bit*
Modulator	100 [9]	*fJ/bit*	Photodetector sensitivity	-26 [9]	*dBm*
MR drop	0.5 [35]	*dB*	Waveguide transmitting	1 [9]	*dB/cm*
MR pass	0.05 [35]	*dB*	Waveguide crossing	0.5 [9]	*dB*

TABLE 5. Electrical Energy Parameters (45*nm* Process)

Wavelengths (*max*)	20 (*for 64 cores*)				25 (*for 400 cores*)				
Router/gateway size	5	6	8	20	5	6	8	21	25
Routing (*pJ/packet*)	8.69	9.83	11.99	20.86	12.12	13.51	16.40	33.51	45.74
Leakage (*pJ/packet*)	0.36	0.39	0.45	0.62	0.51	0.56	0.65	0.96	1.28

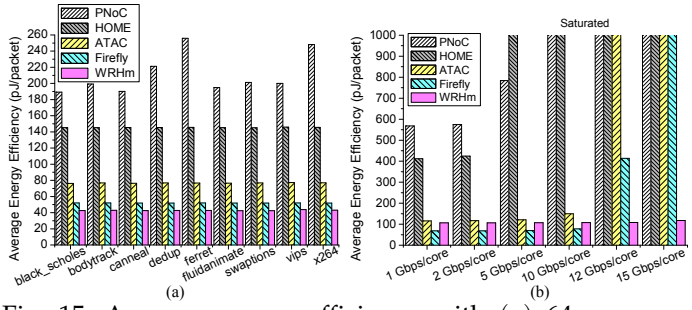


Fig. 15: Average energy efficiency with (a) 64 cores and data traces; (b) 400 cores and synthetic traffics.

optical energy parameters used in the simulation, which are cited from existing works. The electrical energy cost of gateways in WRH-ONoC and routers in the other schemes are obtained from Orion 3.0 [38] and given in Table 5.

Since the buffer size in electrical routers and gateways has huge influence on the energy cost, we set the overall buffer size to be equal in each router and gateway. With different number of wavelengths in trace based and synthetic simulations, *i.e.*, 20 wavelengths are used for connecting 64 cores in trace-based simulation and 25 wavelengths are used for 400 cores in synthetic-based simulation, even though the overall buffer size is the same, their energy costs are different and they are listed separately in Table 5.

Fig. 15(a) compares the average energy efficiency with different data traces. The network configurations are the same as in Section 5.3. We can see that WRH-ONoC has the lowest energy cost than other schemes, due to the optical transmission for both intra-subsystem and inter-subsystem traffics, and no path reservation through an electrical control network. PNoC and HOME consume much higher energy for transmitting a packet, because they need the hop-by-hop path reservation and release processes in the electrical network. ATAC and Firefly have relatively lower energy cost because the optical transmission is used for global communication, while the electrical routing is still used for local communication. It can be seen that *WRH-ONoC only consumes about 44 pJ for transmitting a packet, the overall power consumption is only 1.41 W for 64 cores even with a data rate of 0.5 packets/cycle.*

Fig. 15(b) evaluates the energy efficiency for a 400-core system with synthetic traffics. It can be seen WRH-ONoC consumes the second lowest energy when the data rate is low compared with other schemes. Since WRH-ONoC requires to go through about 2.8 hops of λ-routers and larger crossbar in gateways, the energy cost is a little bit higher than Firefly which uses only one hop of optical network and small electrical routers in the local network. ATAC needs to transmit packets in an electrical local network at both the source and destination side, so the energy cost is a little higher. PNoC and HOME consume much higher energy due to the electrical path reservation process in the large-scale mesh network. However, as the increases of data rate, the energy cost for transmitting a packet also increase. That is because much more static energy is consumed when the end-to-end delay is increased and the packets keep being buffered due to contentions in gateways or routers, even if the static energy cost is low in device-level. Since WRH-ONoC can achieve much lower communication delay even at the data rate of 15 *Gbps/core* as shown in Fig. 9, it consumes less static energy with slight increase.

From the comparison on energy overhead, it can be seen WRH-ONoC is more energy efficient compared with the other schemes, especially in manycore systems with high bandwidth and low delay requirements, such as *on-chip cloud computing* [10] and *memory system* [14].

6 Conclusions

This paper presents an optical inter-core communication network, WRH-ONoC. It combines the hierarchical networking with wavelength reuse to achieve both high-performance and sustainable communication for manycore processors with more than hundreds of cores. WRH-ONoC has several advantages: (*i*) reusing a limited number of wavelength channels for large-scale manycore systems; (*ii*) nonblocking optical local communication and high-throughput global communication; (*iii*) efficient unicast and multicast routing schemes. Moreover, WRH-ONoC is a highly configurable architecture, *e.g.*, the number of sibling gateways and the size of λ-routers, through which it can achieve better tradeoff between performance and hardware cost according to the requirements of specific applications. Simulation results indicate WRH-ONoC can obtain much lower end-to-end delay and higher throughput for both unicast and multicast communications with modest hardware cost and energy overhead. In our future work, we will further analyse the optimal configuration for WRH-ONoC and investigate the optimal gateway architecture.

References

[1] F. Liu, H. Zhang, Y. Chen, Z. Huang, and H. Gu, "WRH-ONoC: a wavelength-reused hierarchical architecture for Optical Network on Chip," in *IEEE Proc. INFOCOM*, 2015, pp. 1912–1920.
[2] Tilera, "72-core processor for intelligent networking and video processing," in *TILE-Gx72 Products*, Feb. 2013.
[3] Intel, "Intel research advances 'era of tera'," in *News Release*, 2007.
[4] Kalray, "MPPA: the supercomputing on a chip solution," 2013.
[5] H. Wei, J. Yu, H. Yu, M. Qin, and G. Gao, "Software pipelining for stream programs on resource constrained multicore architectures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 12, pp. 2338–2350, 2012.
[6] G. Nychis, C. Fallin, T. Moscibroda, O. Mutlu, and S. Seshan, "On-chip networks from a networking perspective: congestion and scalability in many-core interconnects," in *ACM Proc. SIGCOMM*, 2012, pp. 407–418.
[7] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "Optimally removing intercore communication overhead for streaming applications on MPSoCs," *IEEE Trans. Comput.*, vol. 62, no. 2, pp. 336–350, 2013.
[8] Y. Wang, Z. Shao, H. Chan, D. Liu, and Y. Guan, "Memory-aware task scheduling with communication overhead minimization for streaming applications on bus-based multiprocessor System-on-Chips," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, pp. 1797–1807, 2014.

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2733551, IEEE Transactions on Sustainable Computing

[9] R. Morris, A. Kodi, A. Louri, and R. Whaley, "Three-dimensional stacked nanophotonic Network-on-Chip architecture with minimal reconfiguration," *IEEE Trans. Comput.*, vol. 63, pp. 243–255, 2014.

[10] H. Lu, B. Fu, Y. Wang, Y. Han, G. Yan, and X. Li, "RISO: enforce noninterfered performance with relaxed Network-on-Chip isolation in many-core cloud processors," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 12, pp. 3053–3064, 2015.

[11] N. Kirman, M. Kirman, R. Dokania, J. Martinez, A. Apsel, M. Watkins, and D. Albonesi, "Leveraging optical technology in future bus-based chip multiprocessors," in *IEEE/ACM Proc. MICRO*, 2006, pp. 492–503.

[12] M. Petracca, B. Lee, K. Bergman, and L. Carloni, "Photonic NoCs: system-level design exploration," *IEEE Micro Magazine*, vol. 29, no. 4, pp. 74–85, 2009.

[13] C. Li, M. Browning, P. Gratz, and S. Palermo, "LumiNOC: A Power-Efficient, high-performance, Photonic Network-on-Chip," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, pp. 826–838, 2014.

[14] R. Morris, E. Jolley, and A. Kodi, "Extending the performance and energy-efficiency of shared memory multicores with nanophotonic technology," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, 2014.

[15] Intel, "Intel silicon photonics demonstrated at 100 Gbps," in *Intel News Release*, Apr. 2013.

[16] J. Hruska, "IBM to demonstrate first on-package silicon photonics," in *ExtremeTech*, Mar. 2015.

[17] K. Preston, N. Droz, J. Levy, and M. Lipson, "Performance guidelines for WDM interconnects based on silicon microring resonators," in *IEEE Proc. CLEO*, 2011, pp. 1–2.

[18] F. Liu, H. Zhang, Y. Chen, Z. Huang, and H. Gu, "Dynamic ring-based multicast with wavelength reuse for optical network on chips," in *IEEE Proc. MCSOC*, 2016, pp. 153–160.

[19] I. O. Connor, "Optical solutions for system-level interconnect," in *ACM Proc. SLIP*, 2004, pp. 79–88.

[20] C. Chen, T. Zhang, P. Contu, J. Klamkin, A. Coskun, and A. Joshi, "Sharing and placement of on-chip laser sources in silicon-photonic NoCs," in *IEEE/ACM Proc. NOCS*, 2014, pp. 88–95.

[21] A. Liu, L. Liao, Y. Chetrit, J. Basak, N. Hat, D. Rubin, and M. Paniccia, "Wavelength division multiplexing based photonic integrated circuits on silicon-on-insulator platform," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 23–32, 2010.

[22] A. Kazmierczak, M. Briere, E. Drouard, P. Bontoux, P. Rojo-Romeo, I. Connor, X. Letartre, F. Gaffiot, R. Orobtchouk, and T. Benyattou, "Design, simulation, and characterization of a passive optical add-drop filter in silicon-on-insulator technology," *IEEE Photon. Technol. Lett.*, vol. 17, no. 7, pp. 1447–1449, 2005.

[23] S. Beux, J. Trajkovic, I. Connor, G. Nicolescu, G. Bois, and P. Paulin, "Optical ring network-on-chip (ORNoC): architecture and design methodology," in *IEEE Proc. DATE*, 2011, pp. 1–6.

[24] P. Hamedani, N. Jerger, and S. Hessabi, "QuT: a low-power optical network-on-chip," in *IEEE/ACM Proc. NoCS*, 2014, pp. 80–87.

[25] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. Beausoleil, and J. Ahn, "Corona: system implications of emerging nanophotonic technology," in *IEEE Proc. ISCA*, 2008, pp. 153–164.

[26] G. Kurian, J. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, L. Kimerling, and A. Agarwal, "ATAC: a 1000-core cache-coherent processor with on-chip optical network," in *ACM Proc. PACT*, 2010, pp. 477–488.

[27] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary, "Firefly: illuminating future Network-on-Chip with nanophotonics," in *IEEE Proc. ISCA*, 2009, pp. 29–40.

[28] K. Mo, Y. Ye, X. Wu, W. Zhang, W. Liu, and J. Xu, "A hierarchical hybrid optical-electronic Network-on-Chip," in *Proc. ISVLSI*, 2010.

[29] J. Zhao, Y. Gong, W. Tan, and H. Gu, "3D-DMONoC: a new topology for optical network on chip," in *IEEE Proc. ICOCN*, 2016, pp. 1–3.

[30] A. Guerre, N. Ventroux, R. David, and A. Merigot, "Hierarchical Network-on-Chip for embedded many-core architectures," in *IEEE/ACM Proc. NOCS*, 2010, pp. 189–196.

[31] A. Karkar, T. Mak, N. Dahir, R. Al-Dujaily, K. Tong, and A. Yakovlev, "Network-on-Chip multicast architectures using hybrid wire and surface-wave interconnects," *IEEE Trans. Emerg. Topics Comput.*, 2016.

[32] F. Samman, T. Hollstein, and M. Glesner, "Adaptive and deadlock-free tree-based multicast routing for Networks-on-Chip," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst*, vol. 18, no. 7, pp. 1067–1080, 2010.

[33] A. Ruadulescu, K. Goossens, G. D. Micheli, S. Murali, and M. Coenen, "A buffer-sizing algorithm for networks on chip using TDMA and credit-based end-to-end flow control," in *IEEE Proc. CODES+ISSS*, 2006, pp. 130–135.

[34] "Noxim: the Network on Chip simulator." [Online]. Available: https://github.com/davidepatti/noxim.

[35] Y. Kao and H. Chao, "BLOCON: a bufferless photonic Clos Network-on-Chip architecture," in *IEEE/ACM Proc. NOCS*, 2011, pp. 81–88.

[36] J. Liu, J. Yang, and R. Melhem, "GASOLIN: global arbitration for streams of data in optical links," in *IEEE IPDPS*, 2015, pp. 93–102.

[37] J. Hestness, B. Grot, and S. Keckler, "Netrace: dependency-driven trace-based Network-on-Chip simulation," in *ACM Proc. NoCArc*, 2010, pp. 31–36.

[38] A. Kahng, B. Lin, and S. Nath, "ORION3.0: a comprehensive NoC router estimation tool," *IEEE Embedded Syst. Lett.*, vol. 7, no. 2, pp. 41–45, Jun. 2015.



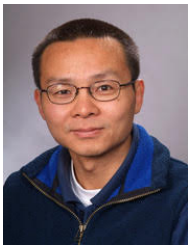
Feiyang Liu is a PhD candidate from Department of Computer Science, University of Otago, New Zealand. He obtained Master degree of communication and information systems in 2012 and Bachelor degree of telecommunication engineering in 2009 from Xidian University, China. His research interests include Network on Chip, Optical Network on Chip (ONoC), multicore systems, communication network, and routing algorithm, etc.



Haibo Zhang received the M.Sc. degree in computer science from Shandong Normal University, China, in 2005 and the Ph.D. degree in computer science from the University of Adelaide, Australia, in 2009. From 2009 to 2010, he was a Postdoctoral Research Associate with the Automatic Control Lab at Royal Institute of Technology in Sweden. He is currently a senior lecturer with the Department of Computer Science, University of Otago, New Zealand. He has served on a number of technical program committee boards and co-chaired several international conferences. His current research interests include real-time industrial wireless communication, optical network-on-chips, wireless body sensor networks, delay-tolerant networks, green computing, distributed algorithms, and protocol design.



Yawen Chen obtained her PhD degree in Computer Science from The University of Adelaide in Australia in 2008, and Bachelor degree in Computer Science in 2002 and Master degree in 2004 from Shandong Normal (Teachers) University in China. She worked as a researcher in Japan Advanced Institute of Science and Technology (JAIST) in 2005. She has been a Lecturer in the University of Otago in New Zealand since 2011. Her research interests include resource optimization and performance evaluation in computer networking and computer architecture (optical network, interconnection network, green computing).



Zhiyi Huang is an Associate Professor at the Department of Computer Science, University of Otago. He received the BSc degree in 1986 and the PhD degree in 1992 in computer science from the National University of Defense Technology (NUDT) in China. He was a visiting professor at EPFL and Tsinghua University in 2005, and a visiting scientist at MIT CSAIL in 2009. His research fields include parallel/distributed computing, multicore architectures, operating systems, signal processing, green computing, cluster/grid/cloud computing, high-performance computing, bio-engineering, and computer networks. He has more than 100 publications.



Huaxi Gu received his PhD degree in telecommunication and information system from Xidian University in 2005. He is currently a professor in the state key lab of ISN, Xidian University, Xian, Shaanxi, China. His current research interests include interconnection networks, networks-on-chip and optical interconnect, and data center networks. He has more than 100 publications in many international journals and conferences.

2377-3782 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.