Contents lists available at ScienceDirect

Optical Switching and Networking

journal homepage: www.elsevier.com/locate/osn

A large-scale nesting ring multi-chip architecture for manycore processor systems

Wenzhe Li [a,b], Bingli Guo [a,*], Xin Li [a], Yu Zhou [a], Shanguo Huang [a], George N. Rouskas [b]

[a] *State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, China*
[b] *Department of Computer Science, North Carolina State University, USA*

ARTICLE INFO

Keywords:
ONoC
Large-scale
Routing algorithm
Resource reservation

ABSTRACT

The optical network on chip (ONoC) paradigm has emerged as a promising solution to multi-core/many-core processor systems for offering enormous bandwidth and low power consumption. As chip multiprocessors (CMPs) scale to unprecedented numbers of cores, the performance of next-generation CMPs will be bounded by the process yield and power density of single chip. In earlier work we proposed a multi-chip ONoC architecture that scales to large numbers of CMPs and delivers high performance in terms of delay and throughout. Building on that work, in this paper we propose an optimized architecture for integrating a large number of cores into chips with a novel control strategy, including a contention resolution scheme and a resource reservation scheme. The proposed control strategy is crucial to large scale ONoCs, because the resource reservation scheme ensures efficient wavelength allocation for the traffic while the contention management scheme is effective in reducing the impact of contentions. To sustain good performance and energy efficiency of large-scale ONoC, the topology is optimized to reduce the average transmission distance with minimum increase of power consumption. We evaluate the proposed architecture within a 1000-core processor system and compare it with CMesh and several previously proposed topologies with different control strategies. The simulation results show that, our new large-scale architecture can achieve better performance on throughput and delay.

1. Introduction

As integration capacity of transistors keeps increasing with Moore's Law [1] and more sophisticated tasks pose a continuous computation resources consumption, major manufacturers projected that hundreds or even thousands of cores will be integrated on multi-core processor systems within the next decade. Systems with such a large number of cores present significant challenges in network architecture design as their performance is increasingly limited by communication rather than computation [2]. In particular, the communication demands on many-core processor systems cannot be satisfied with conventional electronic interconnects that suffer from high power dissipation and limited bandwidth. Fortunately, recent advances in the fabrication of nanophotonic devices, especially in silicon photonics, have made photonic interconnects, which can provide significantly higher bandwidth and lower power consumption, a promising solution for many-core processor systems [3–5]. Specially, Microring resonator (MR) becomes one of widely employed components which are suitable for multi-core processor systems with low footprint and power consumption. MR can be used for multiplexing optical signals, and to multiplex N wavelengths into a waveguide, N MRs operating at different wavelengths are cascaded. While MRs also can be used for demultiplexing, then individual channels are detected with CMOS-compatible germanium detector arrays [6,7]. Furthermore, since MRs can change the direction of optical signal, they are also used to implement optical routers (ORs) for optical network on chip (ONoC).

Many ONoC architectures have been proposed in recent years that use optical interconnections between cores in on-chip systems [8]. Passive ONoC architectures are based on passive OR (passive MRs based optical routers) that do not need additional electronic/thermal control and can automatically route optical signals to different destinations according to their wavelength. The Corona architecture [9] uses an optical crossbar to interconnect 64 clusters on a single chip and avoids contentions by token-ring arbitration. ORNoC [10] is a contention-free

[*] Corresponding author.
E-mail addresses: lwz@bupt.edu.cn (W. Li), guobingli@bupt.edu.cn (B. Guo), xinli@bupt.edu.cn (X. Li), zhouyu0909@bupt.edu.cn (Y. Zhou), shghuang@bupt.edu.cn (S. Huang), rouskas@ncsu.edu (G.N. Rouskas).

https://doi.org/10.1016/j.osn.2018.10.004
Received 4 June 2018; Received in revised form 16 October 2018; Accepted 17 October 2018
Available online 20 October 2018

ring-like ONoC that does not need any arbitration and allows a wavelength to be shared among multiple communications. Fully connected ONoC architectures without a control network and based on all-optical wavelength-routing and various types of passive ORs have been proposed in Refs. [11,12]. Passive ONoCs usually have high design complexity, and their scalability is constrained by the number of wavelengths. At the same time, a wide range of hybrid ONoC architectures have been proposed for high flexibility and scalability, including Torus [3], 2D Mesh, 3D Mesh [13,14] and NRO (nesting ring ONoC) [15,16]. Most hybrid ONoCs are based on optical circuit-switching (OCS) mechanism [17,18], and the process of setting up the optical paths is implemented in the electronic network. For example, NRO employs both an optical and an electrical layer which have identical topology. After the optical path is successfully reserved by electronic control unit in the electrical layer, the optical layer begins payload transmission. However, the setup process suffers from severe congestion under heavy loads, leading to high delay and low resource utilization. The work in Ref. [19] uses a ring-based optical network for the path-setup procedure in ONoC, which may reduce the path-setup delay by simultaneous configuration of optical switches. This method is difficult to implement in large-scale ONoCs because of the need for a centralized arbiter.

Other designs are based on multi-chip architectures due to their low power density and high scalability. Firefly [20] employs an optical crossbar as the inter-chip network and an electrical concentrated mesh (CMesh) network for intra-chip communication; however, the latter will suffer from token delay problems when scaled to large ONoC systems. The study in Ref. [11] proposes a WDM-routed all-to-all network, but this architecture may lead to low resource utilization and a large number of waveguide crossing for the optical signals.

With an increasing number of cores to be integrated on the chip, the scalability of a single chip designs is limited by the low process yield and high power density. A multi-chip system which aggregates several individual smaller chips together in a package can overcome the area constraint of single chip [2]. Optical interconnects can provide an enormous bandwidth for both intra-chip and inter-chip large data transfers with low power consumption. In previous work, we presented a multi-chip architecture for ONoC which we refer to as “nesting ring ONoC” (NRO) [16]. The NRO architecture achieves good performance in terms of throughput and delay and has good scalability properties as the number of chips to be used and the number of small rings on each chip are considered jointly to interconnect a given number of cores. However, to integrate a large number of cores (in the order of 1000 cores), the NRO architecture should be further optimized to guarantee good performance on throughput and delay. Specifically, as the number of cores increases to 1000 or more, the resulting increase in network diameter of the NRO architecture causes higher delays overall and increased congestion for traffic along longer paths.

To address this problem, in this paper we present a large-scale NRO (LSNRO) architecture that extends NRO in two dimensions. First, we modify the original NRO topology by strategically introducing new links between certain nodes so as to reduce the network diameter and average cluster-to-cluster distance. Despite these modifications, LSNRO maintains the important feature of NRO that the specific topology may be designed according to the number of cores with the aim of minimizing the average cluster-to-cluster distance. We also notice that the control strategy plays an important part in the communication performance of ONoC especially electronic-controlled ONoC, which has not been investigated thoroughly in previous works. With so many cores to be integrated on chips, the performance of LSNRO will be limited by both the available optical resources and the resource allocation strategy. Therefore, to address these challenges, we also present a comprehensive control strategy that consists of three components: a resource planning scheme, a resource allocation strategy, and a contention management scheme. The schematic of control strategy is shown Fig. 1. The resource planning scheme adds wavelengths to nodes with heavy traffic to ensure that these nodes do not suffer congestion and delays due to lack of transmission resources. The resource allocation strategy employs a backward reservation (backward resources reservation) scheme to select a wavelength by taking into account the state of all links along the path, not just the state of the source link. Backward reservation alleviates one of the major contributors to congestion for most electronically controlled ONoC architectures with backward reservation, including NRO. Finally, to mitigate the impact of the inevitable contentions on the performance on LSNRO, we also propose a contention management scheme that uses *k*-path routing to reduce the time to resolve contentions and the delay of blocked traffic. Overall, LSNRO represents a substantial improvement over NRO in the following dimensions:

Table 1
The parameters of candidate topologies.

Parameter	Topology I	Topology II
A	6	26
B	10	2
C	5	5
D(NRO)	8.75	9.77
D(LSNRO)	5.74	8.25

1. LSNRO is optimized for large networks and the specific topology is designed by minimizing average cluster-to-cluster distance rather than the longest cluster-to-cluster distance.
2. LSNRO employs a novel and power-efficient resource planning scheme.
3. LSNRO makes use of backward resource reservation, a more efficient scheme than NRO’s first-fit resource reservation, which eliminates most contentions.
4. LSNRO employs a comprehensive contention management scheme based on *k*-path that utilizes a range of strategies to handle contentions.

As a conclusion, in this work, we exploit the advantages of multi-chip architectures and hybrid ONoC to integrate a large number of IP cores on chips. In Section 2 we describe the LSNRO architecture and demonstrate it for a design with 1000 cores. We present the comprehensive control strategy for LSNRO in Section 3 and Section 4. In Section 5, we evaluate the performance of the 1000-core LSNRO in terms of throughput, delay, and power consumption relative to two other architectures, CMesh and NRO. Finally, we conclude the paper in Section 6.

2. 1000-core LSNRO topology

NRO [15] is a multi-chip architecture that consists of two components (shown in Fig. 2): a novel nesting ring topology as the intra-chip network, and ring topology as the inter-chip network. The nesting ring topology consists of a series of small rings that make up a larger ring, which is designed to avoid many contentions caused by the bus-like communication channel of the conventional ring topology. In addition, the nesting ring topology has shorter average transmission distance compared with the ring or mesh topologies, leading to better performance in terms of throughput and delay. For higher scalability, every four processor cores in NRO are interconnected by electronic links and are grouped together as a cluster, sharing an OR and an electronic control unit. Furthermore, considering that there is much more intra-chip communication than inter-chip communication, only intersection clusters are interconnected by an inter-chip network to reduce the usage of high-radix optical routers, waveguides and fabrication complexity.

Nevertheless, when the network is expanded to a large size, traffic over longer transmission paths is likely to encounter congestion and consume a significant amount of optical resources. Therefore, our objective with LSNRO is to further optimize the topology so as to decrease the average cluster-to-cluster distance. In NRO, the intersection clusters are important switching nodes in that they interconnect different chips and small rings. Based on this observation, we augment the NRO topology

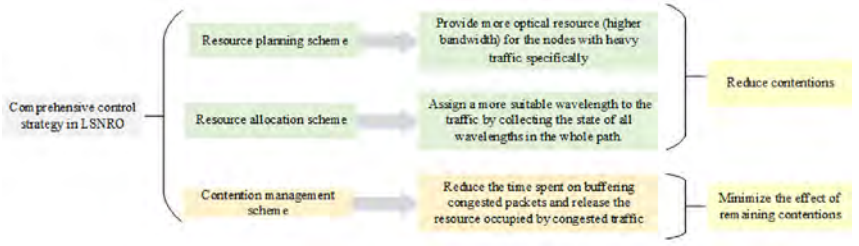

Comprehensive control strategy in LSNRO
Resource planning scheme
Provide more optical resource (higher bandwidth) for the nodes with heavy traffic specifically
Resource allocation scheme
Assign a more suitable wavelength to the traffic by collecting the state of all wavelengths in the whole path
Reduce contentions
Contention management scheme
Reduce the time spent on buffering congested packets and release the resource occupied by congested traffic
Minimize the effect of remaining contentions

Fig. 1. The schematic of the comprehensive control strategy.

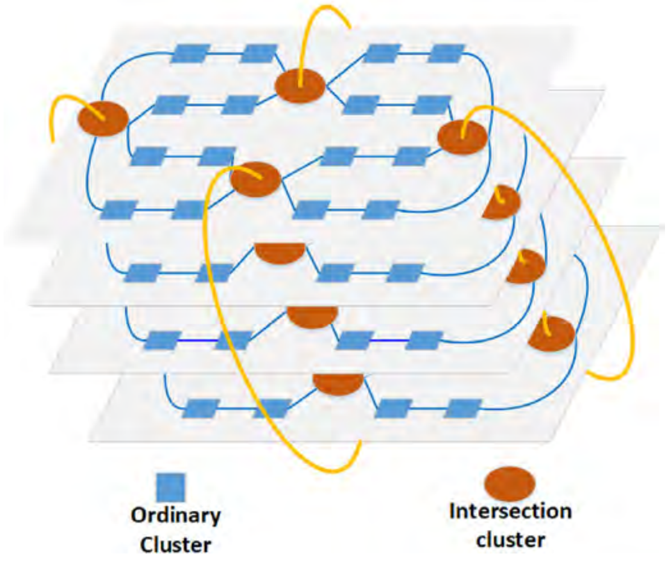


Fig. 2. The original topology of NRO [16].

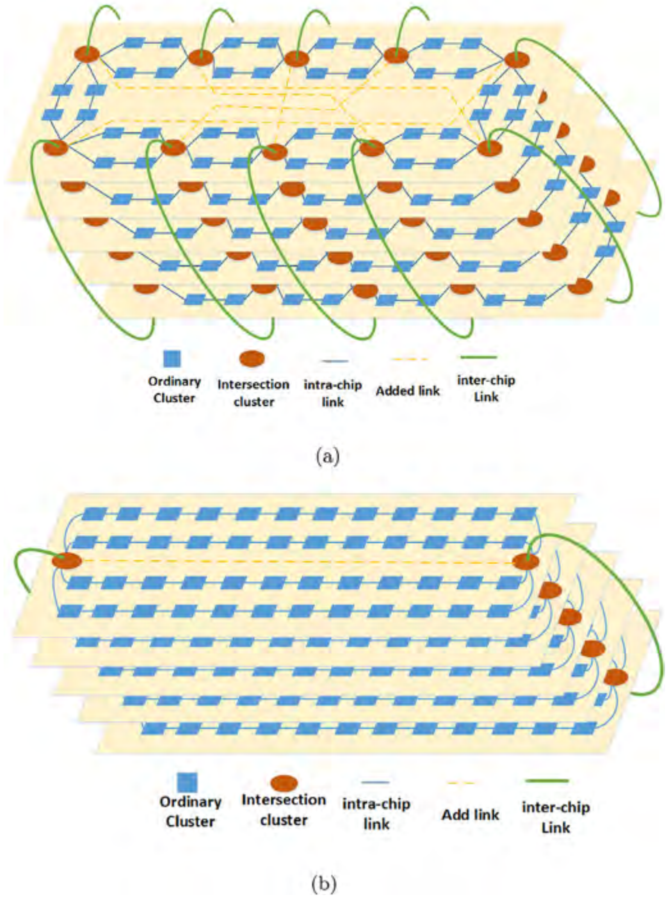


Fig. 3. The architecture of 1000-core LSNRO. (a) Topology I (b) Topology II.

by adding links between each pair of intersection nodes separated by the longest distance (indicated by the dot lines in Fig. 3), while keeping ordinary clusters unchanged. As a consequence, the average cluster-to-cluster distance in LSNRO decreases significantly. For instance, consider Topology I in Fig. 3(a). The average cluster-to-cluster distance in the original NRO topology (i.e., without the dot line links) is 8.75 hops while that of the LSNRO topology (i.e., with the dot line links) is 5.74 hops, a decrease of 34.4%. Since the path-setup delay mainly depends on the propagation delay of control packets and their processing delay on all intermediate routers, we expect that path setup time will be lower for the LSNRO topology. At the same time, shorter paths imply lower resource use, hence we expect an improvement on the throughput of the LSNRO topology.

To implement this enhancement in connectivity, we replace the 7-port ORs at intersection clusters with 8-port ORs. Following the universal method for constructing ORs proposed in Ref. [21], an 8-port OR requires an additional 1/3 of MRs compared to a 7-port OR. Consequently, using higher radix ORs results in higher power consumption, mainly due to the increased thermal power consumed by the additional MRs. On the other hand, switch power is expected to decrease due to the shorter transmission distance. The effect on power consumption and the tradeoff between power efficiency and throughput/delay performance are investigated in Section 5.

The NRO and LSNRO topologies consist of clusters of four cores each, and are parameterized using three parameters that may be selected so as to minimize the average cluster-to-cluster distance, denoted as D:

A denotes the number of clusters on each small ring of the topology;
B denotes the number of small rings on each chip; and
C denotes the number of chips.

Given the number N of cores, the number of clusters $N_{cluster} = N/4$. Furthermore, the relationship between A, B, C and $N_{cluster}$ is given as:

$$N_{cluster} = (A-1) \times B \times C \tag{1}$$

In the above expression, we assume that A and B are even integers with $A >= 4$. We impose the additional constraint on the number of chips $C < 10$, as a large number of chips leads to a loose and redundant network and imposes a burden on the limited inter-chip links. For a 1000-core LSNRO, there are only two sets of values for the three parameters that satisfy the above expression. These sets of values are listed in Table 1, and the corresponding candidate topologies, Topology I and Topology II, are shown in Fig. 3. As we can see in the table, Topology I achieves the lowest average cluster-to-cluster distance, for both the NRO and LSNRO architectures. Therefore, in the remainder of this work we will only consider Topology I.

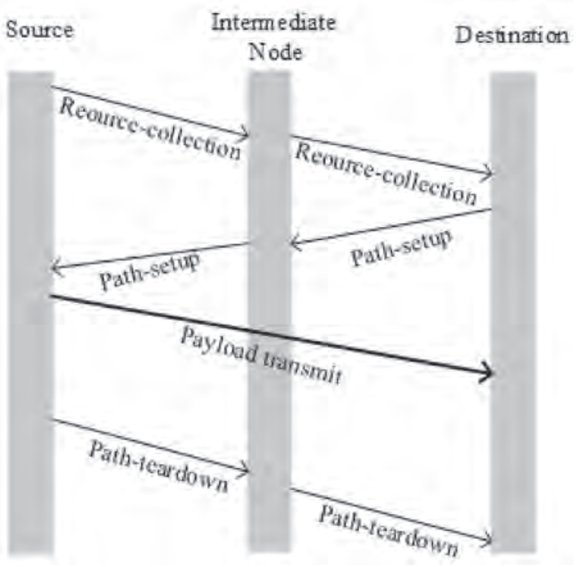


Fig. 4. The process of a successful communication.

3. Resource scheme in LSNRO

3.1. Resource allocation scheme

Optical wavelength allocation in ONoC is crucial in terms of avoiding contentions and guaranteeing high resource utilization, but has not been adequately investigated in the literature. In most previously proposed ONoC architectures, wavelength assignment only consider the states of wavelengths at the source, such that generally the first available wavelength of the source is assigned to the traffic. As a consequence, the first few wavelengths are likely to experience severe contention at subsequent nodes along the path, creating high levels of congestion in hybrid ONoC architectures. Since additional time and energy must be introduced to resolve contentions, the overall performance in terms of delay and energy efficiency will deteriorate. At the same time, unreasonable wavelength assignment leads to low resource utilization, and, in turn, lower network-wide throughput. In LSNRO, we apply a backward resource allocation method as the resource allocation scheme, so as to consider the states of wavelengths along the whole path instead of just the source node. The message passing process of a successful communication is shown in Fig. 4. To set up a path, the source sends a resource-collect packet along the calculated path towards the destination node. The resource-collect packet records the availability of wavelengths at all links of the path. Upon receiving the resource-collect packet, the destination selects a wavelength randomly that is free along all links of the path, and transmits a path setup message towards the source to reserve the selected wavelength hop-to-hop. Upon receipt of the path setup message, the source may start transmitting its data. Considering the cost of path-setup process, We set the minimum supported size of payload as 1000 bits. With this resource allocation scheme, the contentions can be significantly decreased compared with conventional resource allocation. With the increasing workload injected into network, the major contributor to the path-setup delay is the time spent on resolving contentions. Although the process of transmitting the resource-collect packet introduces an extra path setup latency equal to the round-trip time, it may reduce significantly the path-setup delay compared with conventional resource allocation due to a decrease in the number of contentions.

3.2. Resource planning scheme

Since intersection nodes connect different small rings and different chips, they have more traffic passing through them than other nodes. Consequently, the majority of contentions may occur at intersection nodes due to shortage of optical resources. Since the number of optical wavelengths needed by ordinary clusters is generally smaller than that needed by intersection clusters, and since increasing the number of wavelengths for the whole network will increase the total energy consumed due to the additional MRs, we should only add the number of wavelengths of intersection clusters instead of all clusters.

To address this problem and further improve the performance of LSNRO, we propose a resource planning scheme we refer to as “stagger wavelength planning”, which can reduce the blocking rate with minimal increase in power consumption. In the stagger wavelength planning scheme for 1000-core LSNRO, we set the number of wavelengths for intersection nodes at 18 (wavelength 0–17) and that for ordinary nodes at 16. As an example shown in Fig. 5, the ordinary nodes on the two sides of the intersection nodes use two different wavelength groups (wavelengths 0–15 or wavelengths 2–17), but some wavelengths in two groups overlap; the wavelength group in intersection nodes are the union of these two wavelength groups, so there are two extra wavelengths that can be used by the intersection nodes. More optical resource (higher bandwidth) is provided for intersection nodes, fulfilling the traffic requirements while keeping power consumption low. Note that only the wavelengths 2–15 can be used for all paths in LSNRO, because all nodes can receive and switch the wavelengths 2–15; the wavelengths 0–1 and 16–17 can only be used for some specific paths, where all the intermediate nodes can receive and switch wavelengths 0–1 or 16–17. In this paper, depending on the routing and the stagger wavelength planning scheme, the wavelengths 0,1,16, and 17 can be used for the traffic in the following situations.

(1) The source and the destination are located in the same small ring.
(2) The source and the destination are located in two small rings that have the same position in the different layer of chips. For instance, the cluster F in small ring 1 needs to connect the cluster H in small ring 2, and the path can be F-A-E-H, where all clusters support wavelengths 0 and 1.
(3) The source and the destination are located in two small rings that are interconnected by added link (dot line in Fig. 5). For instance, the cluster F in small ring 1 needs to connect the cluster L in small ring 3, and the path is F-G-B-D-K-L; or the cluster F in small ring 1 requests the cluster J in small ring 4, and the path can be F-A-C-I-J.

Hence, the wavelengths 2–15 have higher traffic load than wavelengths 0, 1, 16 and 17. To balance the traffic load on all wavelengths and reduce contentions on wavelengths 2–15, for these specific scenarios, the probability of selecting the wavelengths 0, 1, 16 and 17 should be set higher than for wavelengths 2–15. With this strategy, when wavelengths 0, 1, 16 and 17 are available for the traffic, they will have higher probabilities to be selected than other wavelengths. In this paper, we set the probability at 80%.

4. Contention management scheme in LSNRO

With a conventional OCS communication mechanism, a path-setup packet is sent by the source through the electronic network to the destination, and this path-setup packet reserves the optical path and configures the optical routers. After the path-setup packet arrives at the destination, an ACK packet is sent to the source to inform it that the path has been established and it may begin payload transmission in the optical network. When there is a contention in the path-setup process, the electronic network buffers the blocked message until the requested wavelength is available or the buffering time exceeds a preset limit. The buffering time of the blocked message is one of major contributors to the path-setup delay when the traffic load is relatively heavy; moreover, optical channels that have been reserved by the path-setup packet on upstream links will be unavailable during the buffering time, resulting

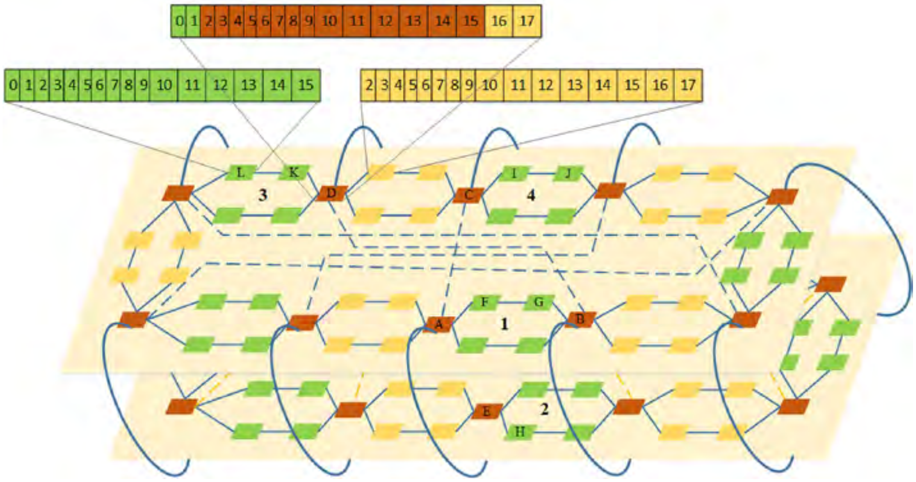


Fig. 5. The Resource Planning in LSNRO (only 2 chips are shown for simplicity).

in lower resource utilization and throughput.

To alleviate the problem, a contention management scheme based on k-path routing algorithm is employed in LSNRO. Specifically, we calculate k shortest paths ($k = 3$ in this paper) as candidate paths between each source-destination pair, and first transmit the setup message along the shortest one. If there contention occurs in the first path, the other two shortest paths are used. The message flow of a congested path-setup process is shown in Fig. 6. Specifically, when contention arise, the wavelength selected by the destination is not available (i.e., it has been reserved by another connection), so the control unit drops the path-setup packet immediately instead of buffering it, and sends a contention packet back to the destination to releases the reserved optical channels. The k-path routing algorithm can exploit the advantage of the nesting ring topology, which can provide multiple paths that have same or similar distance for the traffic. The time overhead of rebuilding the path $T_{rebuild}$ can be calculated with Eq. (2), where $T_{hop}$is the hop latency of control packets, including the processing time of control unit and link traversal time, $H_{contention}$ represents the number of the hops of the contention packet (from the contention node to the source), and $H_{nextpath}$ denotes the number of the hops of the next transmission path. Compared with buffering a blocked message, using the candidate path may reduce the path-setup time at higher traffic loads.

$$T_{bulid} = T_{hop} \times (H_{contention} + 2H_{nextpath}) \tag{2}$$

In addition, as a main constraint of the future large-scale ONoC, the power efficiency should be taken into consideration as well. In LSNRO, the electronic control units of inter-chip network are placed in a central arbiter and are interconnected with short electronic links; on the other hand, the intersection clusters are connected to the central arbiter with optical links, because the longer optical links do not incur additional latency or power dissipation. The setup process of inter-chip communication needs to be implemented with extra EO/OE conversion. The energy of a path reestablishment process, given in Eq. (3), mainly consists of the energy required for control packets traveling through electronic links and routers and the energy consumed for OE/EO conversion of the control packet. In Eq. (3), E_{hop} represents the energy expended to transmit the packet across a link and a router, and it is about 235 pJ according to ORION 2.0 [22]; E_{EOE} represents the energy consumed for EO and OE conversion of the control packet, and it is about 2000; while N denotes the number of control packets through the EO/OE conversion. A cost of the k-path routing algorithm is that the transmission of control packets (including contention packet, path-setup packet and resource packet) will introduce extra energy spent on EO/OE conversion of control packet if the blocking path is across chips. So there is a tradeoff between power consumption and network performance in this situation. Based on the location of a blocked path-setup packet along a path across chips, we distinguish two cases:

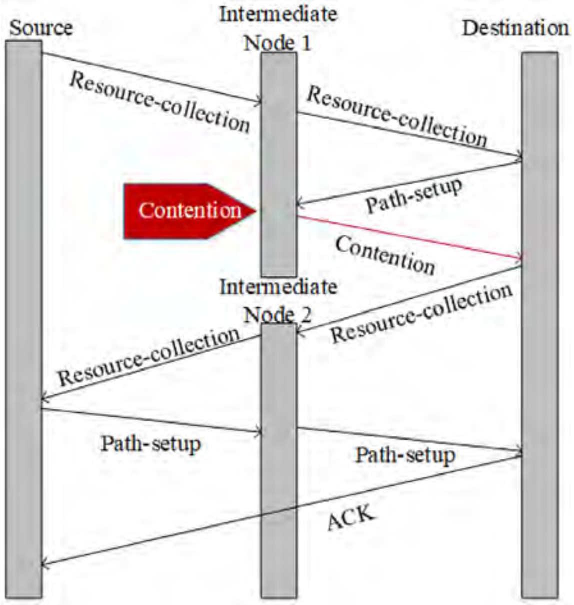


Fig. 6. The process of contention.

a) If the path-setup packet is blocked after it has passed through OE/EO conversion, rebuilding the path will spend power on 5 OE/EO conversions in total, compared with 2 OE/EO conversions without reestablishment; therefore buffering is a better option in this case.
b) If the path-setup packet has not passed through the OE/EO conversions, there is only one extra OE/EO conversion needed for implementing path reestablishment; therefore the k-path algorithm should be applied to achieve better performance in this case.

Table 2
Comparison of optical links and MRs.

Optical Resource	CMesh	NRO1000	LSNRO
The number of optical links	465	310	335
The number of MRs	3688W[a]	1500W[a]	1800W[a]

[a] W represents the number of wavelengths in WDM.

Table 3
The description of the simulated networks.

Network	Routing algorithm	Resource allocation	Resource planning	Topology
LSNRO	K-path based	Backward resource reservation	Add Resources in intersection nodes	Optimized
LSNRO-OnlyBuffer	Dijkstra	First-fit	None	Optimized
LSNRO-BRR	Dijkstra	Backward resource reservation	None	Optimized
LSNRO-NoAddingResource	K-path based	Backward resource reservation	None	Optimized
NRO1000	Dijkstra	First-fit	None	Nesting Ring
CMesh	Dijkstra	First-fit	None	10 × 25 CMesh

$$E_{reestablish} = E_{hop} \times (H_{contention} + 2H_{nextpath}) + E_{EOE} \times N \tag{3}$$

So in our contention management mechanism for LSNRO, we use both multiple candidate paths and buffering of blocked messages, depending on the situation, so as to achieve a good balance between delay and power efficiency. The *k*-path communication mechanism is only triggered when the node where contention occurs is on same chip with the source of the path-setup packet. The communication process of is summarized in following steps.

Step 1: The source calculates k shortest paths, and sends the resource-collect packet through the electronic network to the destination along the first path; if wavelengths are available, then the destination allocates the channel with the information of resource-collect packet.

Step 2: The destination sends a path-setup packet to the source to establish a path.

Step 3: When contention occurs, if the node encountering contention is on a different chip from the source, the control unit buffers the blocked packet; otherwise, the control unit drops the path-setup packet immediately and sends a contention packet back to release the reserved path. The source repeats the process using the next shortest path.

Step 4: The source transmits the payload after the establishment of the path.

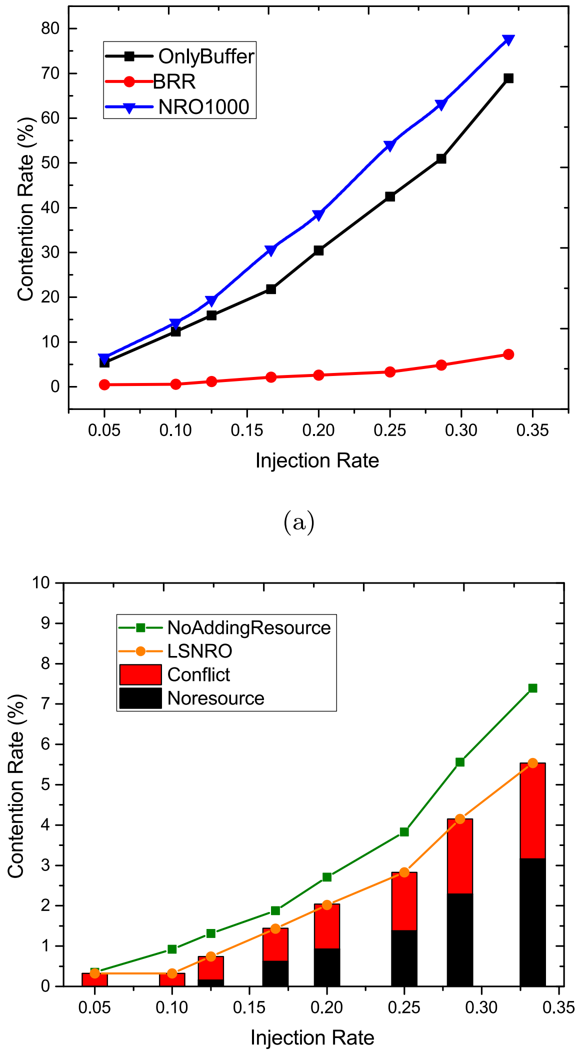
Step 5: After the payload transmission is completed, the source sends a teardown packet to free the resources.

5. Simulation and results

To evaluate the performance of LSNRO and verify the proposed contention management and resource reservation schemes, we simulated the LSNRO and other several ONoC architectures in 1000 cores using on OMNeT++ simulator [23]. Since most existing ONoC architectures were not designed to scale to topologies of 1000 cores or larger, we compare the LSNRO architecture variants (i.e., differing in the control strategies deployed) to the CMesh and NRO architectures that may be extended to large topologies. Table 2 compares the three architectures for the same topology size (1000 cores) in terms of their optical resources (i.e., optical links and MRs). As we can see, LSNRO uses 20% more MRs and 8% more links than NRO, which is expected given our discussion in the previous section of how LSNRO extends NRO. On the other hand, LSNRO requires 51% fewer MRs and 28% fewer links than a straightforward extension of CMesh to 1000 cores, a significant savings in cost and footprint.

The simulation parameters and the description of simulated architectures are shown in Table 3. The clock frequency is set at 5 GHz. The electrical traverse time is modeled as 2 cycles and processing time in router is modeled as 3 cycles. So one hop delay should be 5 cycles.

The networks are simulated with synthetic benchmarks that are a uniform traffic pattern and a 75% localized traffic pattern. In our simulation, 16 wavelengths are multiplexed into a waveguide in ordinary nodes and 18 wavelengths in intersection nodes for the resource planning scheme, while the data rate of each wavelength is set at 10 Gbps, a rate that has been demonstrated in experiments for ONoC [24]. The payload size is assumed to be 125 Bytes, so the transmission delay of payload is a constant value (100 ns) in this simulation. The electrical traverse time is modeled as 2 cycles and processing time in router is modeled as 3 cycles. So one hop delay should be 5 cycles. The contention rates shown in Fig. 7(a) and (b) are the ratio of the number of packets which encounter contentions to the amount of all sent packets within the simulation time. In addition, we evaluated the performance of the architectures in terms of throughput and path-setup delay. The path-setup delay includes the transmission time, processing time, and OE and EO conversion of all control packets in the path-setup process. All metrics are measured as a function of a given injection rate *α*, which



(a)

(b)

Fig. 7. The comparison of contention rates under uniform traffic pattern and the compositions of contention rate of LSNRO. (a) LSNRO-OnlyBuffer and LSNRO-BRR (b) LSNRO-NoAddingResource and LSNRO.

Table 4
The retransmission times of lost packets.

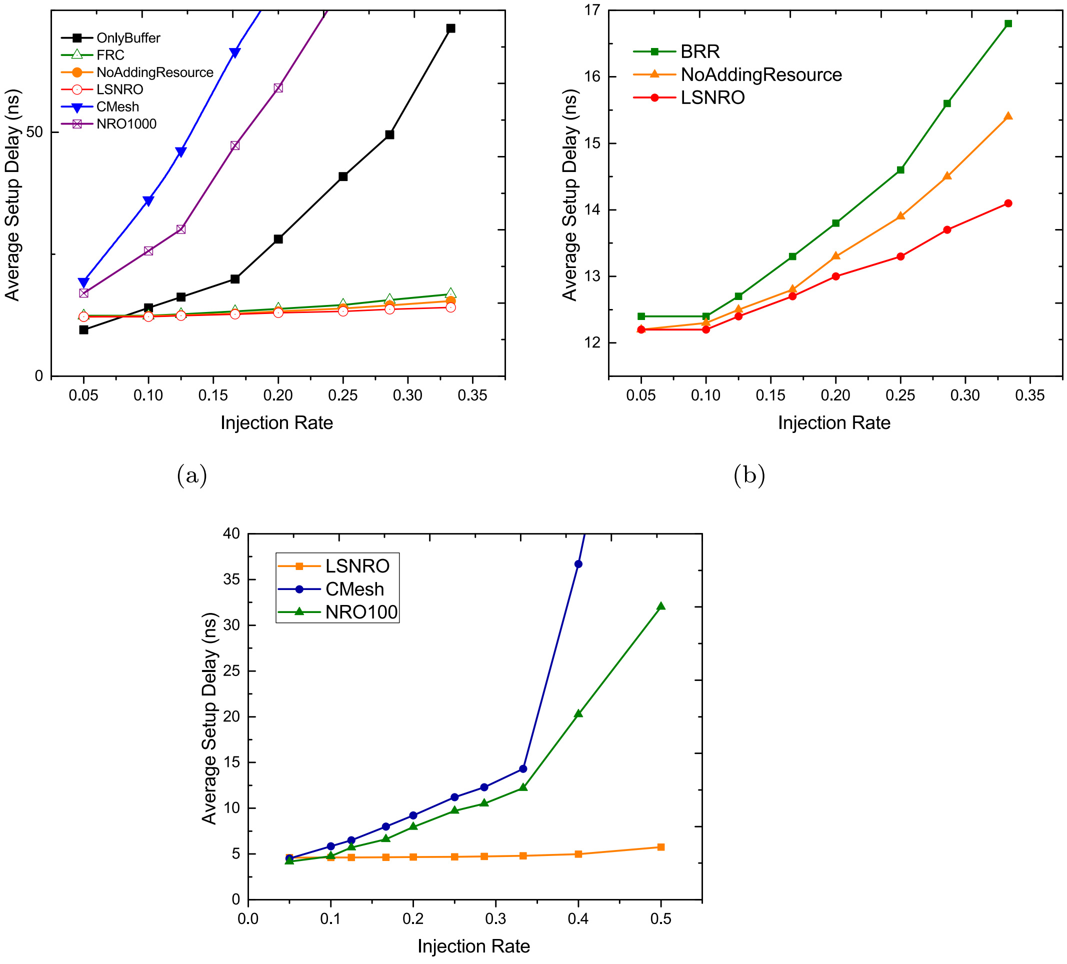
Injection rate	0.2	0.33
Drop rate (%)	0.106	0.187
Success after first retransmission	100%	93%
Success after second retransmission	N/A	7%

is defined by Eq. (4). In Eq. (4), the average interval time between two successive packets is $T_{interval}$, assumed to follow a negative exponential distribution, and $T_{transmit}$ is the duration time of payload transmission.

$$\alpha = \frac{T_{transmission}}{T_{transmission} + T_{interval}} \tag{4}$$

5.1. Contention rate

The comparison of contention rates is shown in Fig. 7. The network performance on delay and throughput is tightly tied to contentions, hence the contention rate may be used to directly evaluate the influence of the proposed method on LSNRO. To this end, we first evaluate several networks with respect to contention rates so as to investigate the impact of the topology, the resource allocation scheme, and the resource planning scheme. As depicted in Table 3, the only difference between LSNRO-OnlyBuffer and LSNRO-BRR is in the resource allocation method. Fig. 7(a) illustrates that s reservation can dramatically decrease the contention rate of the architecture, up to 89% at an injection rate of 0.333 compared with that without it. Since a simplistic resource allocation is a major contributor to congestion in ONoC, most contentions may be avoided by employing the backward resources reservation scheme. Also, from Fig. 7(a), NRO1000 has higher contention rate than LSNRO-OnlyBuffer; this may be explained by the fact, that traffic in NRO1000 takes longer paths on average, and hence it occupies more optical resources leading to more contentions. Furthermore, the topology of NRO1000 has fewer optical interconnections than LSNRO, which causes more contentions. Fig. 7(b) illustrates that the proposed resource planning scheme may decrease the contention rate by up to 25%. This verifies that the proposed resource planning can alleviate the contention problem caused by the shortage of the optical wavelengths. Although the contention rate of LSNRO-NoAddingReousrce is already low, the reduction of contentions is also important to large-scale ONoC, which can reduce the delay and power spent on dealing with contentions. Fig. 7(b) also illustrates the compositions of the contention rate of LSNRO. The Conflict indicates the contention caused by the situation that multiple setup packets try to reserve the same wavelength, while the Noresource indicates the contention caused by the situation that no wavelength is available for the traffic with the information of resource-collect packet. The reason to the conflict is that architecture multiple concurrent resource-collect packets may record same wavelengths due to the distributed network. At the



(a)

(b)

(c)

Fig. 8. The comparison of the path-setup delay. (a) all architectures under uniform pattern (b) LSNRO-BRR, LSNRO-NoAddingResource and LSNRO under uniform pattern (c) LSNRO, NRO, and CMesh under localized traffic pattern.

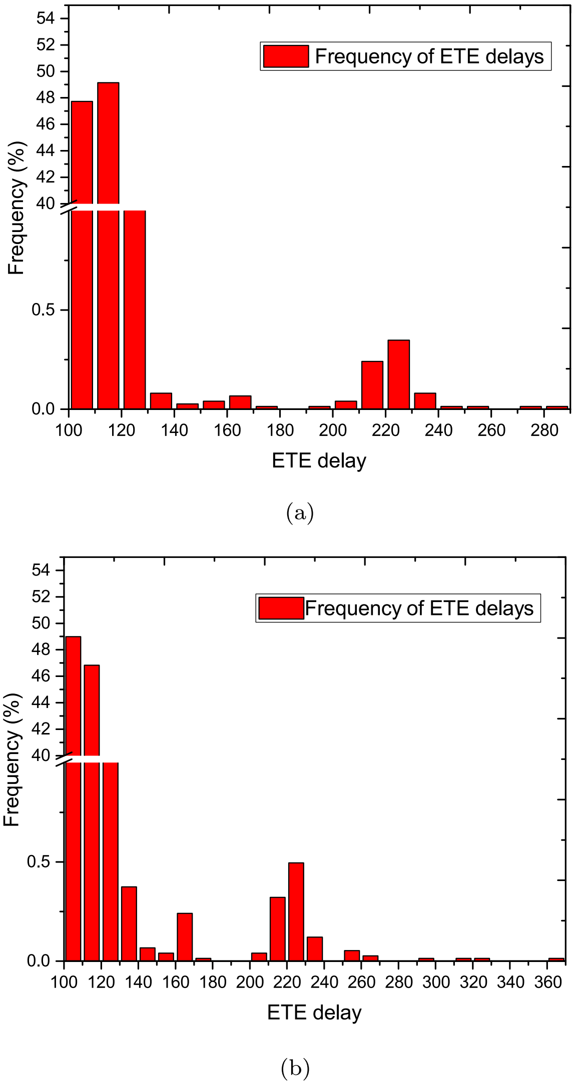


Fig. 9. The frequency distribution of the ETE delays under uniform traffic. (a) With injection rate of 0.2. (b) With injection rate of 0.33.

injection rate of 0.33, less than 50% of contentions caused by conflicts, and the conflict rate is kept under 2.5%, which means the conflicts occurs with a low probability.

5.2. *Lost packets*

To analyzes the starvation phenomenon in proposed architecture, we simulated the drop rate and the retransmission times under injection rate of 0.2 and 0.33, shown in Table 4. In our simulation, we set the time for resending the message after a congested message is dropped at 50 ns. When the network is lightly loaded, the dropped rate is 0.106%, and all messages reach the destination after the first retransmission. When the network is moderately loaded, the drop rate is 0.187% and about 93% of message will reach the destination after the first retransmission and 7% of dropped message will arrive after the second retransmission.

5.3. *Delay*

Fig. 8 shows the path-setup delay of all simulated architectures. As shown in Fig. 8(a), obviously CMesh has the worst performance on delay, while NRO1000 performs the second worst because its topology has longer average transmission distance and fewer interconnections than LSNRO. Compared with NRO1000, LSNRO-OnlyBuffer reduces setup delay by 52.45% at the injection rate of 0.2 due to the optimized topology. The optimized topology in LSNRO can dramatically reduce the average transmission distance, which can directly reduce the setup delay, besides it can reduce much time spent on dealing with contentions. LSNRO, LSNRO-BRR and LSNRO-NoAddingResource (all use backward resource reservation) have much lower setup delay compared with LSNRO-OnlyBuffer after injection rate 0.05, and they are relatively stable under the increasing injection rate. The employment of backward resource reservation needs extra time to transmit resource-collection packet, so setup delay of the networks with backward resource reservation is higher than LSNRO-OnlyBuffer at an extremely low injection rate. However, with increasing injection rate, contentions become the major contributor to the setup delay. As noted above, backward resource reservation can reduce the contention rate dramatically, so the networks with backward resources reservation have much better performance on setup delay that others.

Fig. 8(b) emphasizes the comparison of LSNRO, LSNRO-BRR and LSNRO-NoAddingResource. LSNRO-NoAddingResource reduces setup delay by about 8% compared with LSNRO-BRR at injection rate of 0.33 due to the contention management scheme. For the blocked traffic, it needs about averaged 100 ns to handle contentions with the buffer method, while just needs about averaged 12 ns with *k*-path based method, which proves the effectiveness of the proposed contention management scheme in maintaining a good delay performance of the blocked traffic. Moreover, LSNRO reduces setup delay by about 8.4% compared with LSNRO-NoAddingResource due the proposed resource planning scheme, which can reduce the contentions by providing more wavelengths for intersection clusters. With all proposed strategies, LSNRO obtains a very low setup delay, which is below 14 ns before injection rate 0.33. At the injection rate of 0.33, the average setup delay of LSNRO is 14.3 ns, and the transmission time of payload is 100 ns, so the average end-to-end delay of the LSNRO should be 114.3 ns. Fig. 8(c) shows under localized traffic pattern LSNRO also keep a constant and low delay, while NRO and CMesh increase sharply after saturation points.

We also evaluate the frequency distribution of end-to-end (ETE) delays under injection rates of 0.2 and 0.33, shown in Fig. 9. Under injection rate of 0.2, there are 47.7% of ETE delays between 100ns and 110 ns, while there are 49.1% of ETE delays between 110ns and 120 ns. Longer delays occur in a low probability. When the injection rate increases to 0.33, although the maximum value of the ETE delay increases, the ETE delay of most traffic (95.8%) is between 100 ns and 120 ns. These simulation results demonstrate that LSNRO can fulfill the requirement for low delay in real applications.

In general, we expect that typical ONoC applications will generate localized traffic, and we use the neighborhood pattern to represent these scenarios. We use the uniform pattern to represent applications where there is little information about how traffic is exchanged between nodes. The two patterns account for a variety of traffic patterns that may arise in ONoC applications. Given the consistency of the results regarding the relative performance of the schemes we study, we believe that other traffic patterns will not affect this relative performance and our conclusions.

5.4. *Throughput*

The throughput of the simulated networks is shown in Fig. 10. LSNRO achieves significantly higher throughput than CMesh (up to six times as much, for high loads), while it employs 49% of MRs and 72% of optical links of CMesh (see Table 2). Furthermore, the increase in maximum throughput between LSNRO-OnlyBuffer and NRO1000 is up to 76%, whereas LSNRO uses 20% more MRs and 8% more links compared to NRO1000 (refer to Table 2). This result indicates that the improvement in performance is mainly due to the optimized topology, not just the additional optical resources. LSNRO-OnlyBuffer reaches

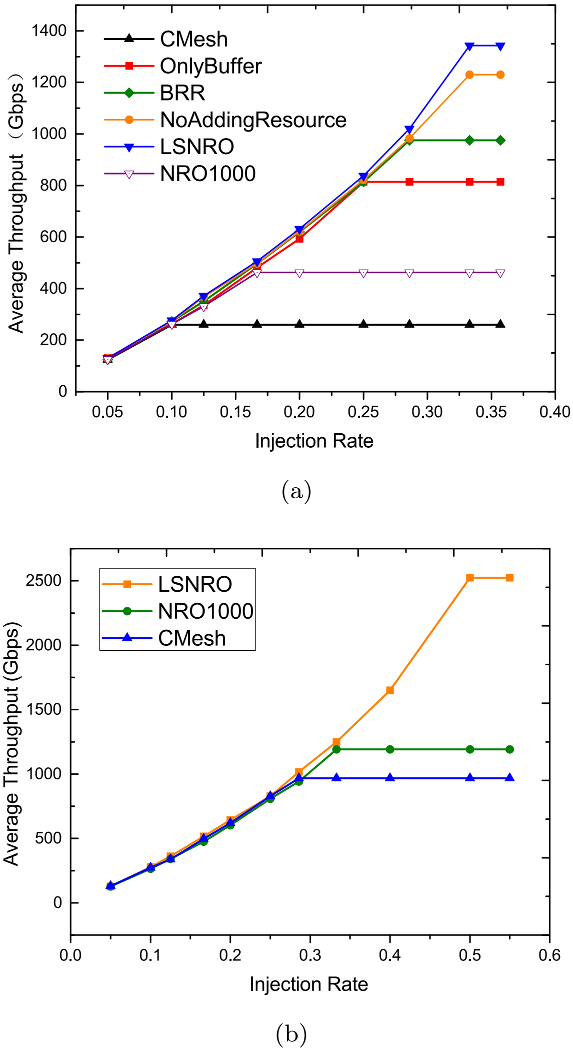


Fig. 10. The comparison of the average throughput. (a) all networks under uniform traffic pattern (b) LSNRO NRO and CMesh under localized traffic pattern.

the maximum throughput at 814 Gbps when the injection rate reaches 0.25, which is referred as saturation point; while, the saturation point of LSNRO-BRR appears around 0.29, therefore LSNRO-BRR can reach a higher maximum throughput than LSNRO-onlyBuffer. As mentioned above, the employment of backward resource reservation in LSNRO-BRR reduces the most contentions when the injection rate increase, which delays the appearance of the saturation point and increases the maximum throughout. In addition, LSNRO-NoAddingResource has higher maximum throughput (1230 Gbps) than LSNRO-BRR due to the contention management scheme, which can help avoid the situation where the blocked traffic will occupy the optical path for a long time and waste a large amount of resources in conventional control strategy. With the proposed resource planning scheme, LSNRO can further improve the maximum throughput, reaching at 1343 Gbps. These simulation results verify that the optimized topology and the proposed control strategy can improve the performance of LSNRO on throughput. To further verify the performance of LSNRO in terms of throughput, we report the LSNRO, NRO, and CMesh under localized traffic pattern shown in Fig. 10(b). The maximum throughput of LSNRO is more than twice as much as that of NRO or CMesh, and it is around 2525 Gbps.

5.5. Power analysis

In this section, we make a simple analysis of optical power consumption of LSNRO with 1000 cores under the uniform traffic pattern.

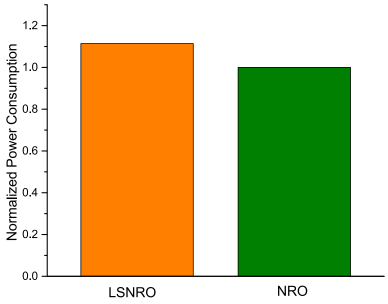


Fig. 11. The comparison of the normalized power consumption between LSNRO and NRO.

Fig. 11 shows the comparison of the power consumption of LSNRO and NRO under the injection rate 0.1. The power dissipated by photodetectors and modulators is the largest contributor to the overall power dissipation. Compared with NRO, the power dissipation of the photodetectors in LSNRO increases due to added wavelengths in intersection clusters. At the same time, the static power consumed on thermal tuning of MRs is the second-largest contributor to the power consumption, and it is related to the number of MRs. As we mention in Section 2, the optimized topology employs higher radix ORs for intersection clusters, which will increase the number MRs by about 23%. So the power consumed on thermal tuning increase by 23% compares with the origin topology of NRO with 1000 cores correspondingly. In addition, a fraction of the power consumption is dissipated by switches and EO/OE conversions of the control packets. Thanks to shorter average transmission distance, LSNRO consumes less switch power than NRO, exactly LSNRO can reduce switch power by about 34%. While LSNRO will spend extra power on EO/OE conversions of the control packet when rebuild the path with the *k*-path based routing algorithm, and the congestion of network is the key factor determining this part of the power consumption. Therefore, when the injection rate is low, there is no extra conversion power consumption needed to deal with congestion. When the injection rate is 0.1, there is about 0.95% growth in the power consumption of conversion.

6. Conclusion

This paper proposes a high-performance large-scale ONoC architecture LSNRO, which is based on our previously proposed architecture NRO. To integrate a large amount of cores, the topology is optimized with minimum increase in power consumption, aiming at reducing the average transmission distance. Furthermore, a series of control strategies are proposed correspondingly to achieve better performance in terms of delay, throughput, and power efficiency. With proposed resource allocation scheme, the majority of contentions in LSNRO can be avoided. Meanwhile, the resource planning scheme can reduce the contentions caused by shortage of wavelengths. The contention management scheme can alleviate the effect of the inevitable contentions and decrease the delay of the blocked traffic. The simulation results show that LSNRO with 1000 cores has better performance on delay and throughput compared with NRO and CMesh, and it has reasonable power consumption as well, which verifies the advantage of LSNRO in future large-scale multi-chip processor systems.

Acknowledgment

This work is supported in part by the NSF of China (No.61601054, 1331008, 61701039, 61571058 and 61771074), the Project of China Scholarship Council (No. 201706470047), the NSF for Outstanding Youth Scholars of China (No.61622102) and Youth Research and Innovation Program of BUPT, State Key Lab. of Adv. Optical Commun. Systems & Networks.

References

[1] G. Kurian, J.E. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, L.C. Kimerling, A. Agarwal, Atac: a 1000-core cache-coherent processor with on-chip optical network, in: Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques, ACM, 2010, pp. 477–488.

[2] I. O'Connor, Optical solutions for system-level interconnect, in: Proceedings of the 2004 International Workshop on System Level Interconnect Prediction, ACM, 2004, pp. 79–88.

[3] A. Shacham, K. Bergman, L.P. Carloni, Photonic networks-on-chip for future generations of chip multiprocessors, IEEE Trans. Comput. 57 (9) (2008) 1246–1260.

[4] H. Wang, M. Petracca, A. Biberman, B.G. Lee, L.P. Carloni, K. Bergman, Nanophotonic Optical Interconnection Network Architecture for On-chip and Off-chip Communications, 2008.

[5] L. Guo, Z. Ning, W. Hou, X. Hu, P. Guo, Quick answer for big data in sharing economy: innovative computer architecture design facilitating optimal service-demand matching, IEEE Trans. Autom. Sci. Eng.

[6] R. Hendry, D. Nikolova, S. Rumley, K. Bergman, Modeling and evaluation of chip-to-chip scale silicon photonic networks, in: High-performance Interconnects (HOTI), 2014 IEEE 22nd Annual Symposium on, IEEE, 2014, pp. 1–8.

[7] W. Hou, Z. Ning, X. Hu, L. Guo, X. Deng, Y. Yang, R. Y. Kwok, On-chip hardware accelerator for automated diagnosis through human-machine interactions in healthcare delivery, IEEE Trans. Autom. Sci. Eng..

[8] S. Hesham, J. Rettkowski, D. Goehringer, M.A.A. El Ghany, Survey on real-time networks-on-chip, IEEE Trans. Parallel Distr. Syst. 28 (5) (2017) 1500–1517.

[9] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N.P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R.G. Beausoleil, J.H. Ahn, Corona: system implications of emerging nanophotonic technology, in: ACM SIGARCH Computer Architecture News, vol. 36, IEEE Computer Society, 2008, pp. 153–164.

[10] S. Le Beux, J. Trajkovic, I. O'Connor, G. Nicolescu, G. Bois, P. Paulin, Optical ring network-on-chip (ornoc): architecture and design methodology, in: Design, Automation & Test in Europe Conference & Exhibition (DATE), vol. 2011, IEEE, 2011, pp. 1–6.

[11] I. O'Connor, F. Mieyeville, F. Gaffiot, A. Scandurra, G. Nicolescu, Reduction methods for adapting optical network on chip topologies to specific routing applications, in: Proceedings of DCIS, 2008.

[12] X. Tan, M. Yang, L. Zhang, Y. Jiang, J. Yang, A generic optical router design for photonic network-on-chips, J. Lightwave Technol. 30 (3) (2012) 368–376.

[13] Y. Ye, J. Xu, B. Huang, X. Wu, W. Zhang, X. Wang, M. Nikdast, Z. Wang, W. Liu, Z. Wang, 3-d mesh-based optical network-on-chip for multiprocessor system-on-chip, IEEE Trans. Comput. Aided Des. Integrated Circ. Syst. 32 (4) (2013) 584–596.

[14] P. Guo, W. Hou, L. Guo, Q. Yang, Y. Ge, H. Liang, Low insertion loss and non-blocking microring-based optical router for 3d optical network-on-chip, IEEE Photon. J. 10 (2) (2018) 1–10.

[15] W. Li, S. Huang, Y. Zhou, S. Yin, J. Zhang, W. Gu, A nesting ring optical network on chip (onoc) architecture for multi-chip systems, in: Asia Communications and Photonics Conference, Optical Society of America, 2015, ASu1H–1.

[16] W. Li, B. Guo, X. Li, S. Yin, Y. Zhou, S. Huang, Nesting ring architecture of multichip optical network on chip for many-core processor systems, Opt. Eng. 56 (3) (2017) 035106.

[17] S. Bartolini, L. Lusnig, E. Martinelli, Olympic: a hierarchical all-optical photonic network for low-power chip multiprocessors, in: Digital System Design (DSD), 2013 Euromicro Conference on, IEEE, 2013, pp. 56–59.

[18] P. Grani, S. Bartolini, Scalable path-setup scheme for all-optical dynamic circuit switched nocs in cache coherent cmps, ACM J. Emerg. Technol. Comput. Syst. 14 (1) (2018) 12.

[19] P. Grani, S. Bartolini, Simultaneous optical path-setup for reconfigurable photonic networks in tiled cmps, in: High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 2014 IEEE Intl Conf on, IEEE, 2014, pp. 482–485.

[20] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, A. Choudhary, Firefly: illuminating future network-on-chip with nanophotonics, in: ACM SIGARCH Computer Architecture News, vol. 37, ACM, 2009, pp. 429–440.

[21] R. Min, R. Ji, Q. Chen, L. Zhang, L. Yang, A universal method for constructing n-port nonblocking optical router for photonic networks-on-chip, J. Lightwave Technol. 30 (23) (2012) 3736–3741.

[22] A.B. Kahng, B. Li, L.-S. Peh, K. Samadi, Orion 2.0: a power-area simulator for interconnection networks, IEEE Trans. Very Large Scale Integr. Syst. 20 (1) (2012) 191–196.

[23] A. Varga, Discrete event simulation system, in: Proc. Of the European Simulation Multiconference (ESM'2001), 2001.

[24] A. Biberman, B.G. Lee, K. Bergman, Demonstration of all-optical multi-wavelength message routing for silicon photonic networks, in: Optical Fiber Communication\National Fiber Optic Engineers Conference, 2008. OFC\NFOEC 2008. Conference on, IEEE, 2008, pp. 1–3.