Contention-aware routing for thermal-reliable optical networks-on-chip

Li, Mengquan; Liu, Weichen; Duong, Luan H. K.; Chen, Peng; Yang, Lei; Xiao, Chunhua

2020

Li, M., Liu, W., Duong, L. H. K., Chen, P., Yang, L., & Xiao, C. (2021). Contention-aware routing for thermal-reliable optical networks-on-chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(2), 260-273. doi:10.1109/TCAD.2020.2994261

<https://hdl.handle.net/10356/146079>

<https://doi.org/10.1109/TCAD.2020.2994261>

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TCAD.2020.2994261>

Downloaded on 26 Apr 2023 21:08:00 SGT

Contention-aware Routing for Thermal-Reliable Optical Networks-on-Chip

Mengquan Li, Weichen Liu, Luan H.K. Duong, Peng Chen, Lei Yang, Chunhua Xiao

Abstract—As an emerging communication infrastructure, optical network-on-chip (ONoC) offers ultrahigh bandwidth, low latency, and low power dissipation for new-generation manycore systems. However, the benefits in communication performance and energy efficiency will be diminished by communication contention. The intrinsic thermal susceptibility is another challenge for ONoC designs. Under on-chip temperature variations, core functional devices suffer from significant thermal-induced optical power loss, which seriously threatens ONoCs' reliability. In this paper, we develop novel routing techniques to resolve both issues for ONoCs. By analyzing the thermal effect in ONoCs, we first present a routing criterion at the network level. Combined with device-level thermal tuning, it can implement thermal-reliable ONoCs. Two routing approaches, including a mixed-integer linear programming (MILP) model and a heuristic algorithm (called CAR), are further proposed to minimize communication conflicts based on guaranteed thermal reliability, and meanwhile, maximize the communication energy efficiency in the presence of on-chip thermal variations. By applying the criterion, our approaches achieve excellent performance with largely reduced complexity of design space exploration. Evaluation results based on both synthetic traffic patterns and realistic benchmarks validate the effectiveness of our approaches with an average of 126.95% improvement in communication performance and 16.12% reduction in energy overhead compared to state-of-the-art techniques. CAR only introduces 7.20% performance difference compared to the MILP model and is more scalable to large-size ONoCs.

Index Terms—Optical network-on-chip, Thermal reliability, Contention-aware routing, Design space exploration.

I. INTRODUCTION

OPTICAL network-on-chip (ONoC) architecture [1] provides an innovative solution for new-generation manycore systems to satisfy the communication bandwidth and latency requirements with low power dissipation. It typically comprises a photonic network for inter-processor data transmission and an electronic network to perform logical control. An approach of *optical circuit switching* is employed on ONoCs for routing. Data transmission is preceded by a path reservation process, in which a 'path-setup' control packet is routed in the electronic network to establish an optical path; once the path is acquired, optical messages are

transmitted end-to-end in the photonic network without in-flight buffering and processing. This approach provides excellent communication performance for unconflicted messages, in which dedicated optical circuits guarantee the full bandwidth of the channels and remain connected for the duration of the communication sessions. However, conflicted messages have to be blocked with extra delays. It significantly degrades the communication performance and energy efficiency of ONoCs, especially for communication-intensive applications.

To address this problem, contention-free ONoC architectures have been widely employed [2], [3]. They possess high efficiency but introduce additional hardware overheads. Contention-aware routing is an attractive solution for ONoCs, which mitigates communication conflicts by utilizing the inherent flexibility of adaptive routing. It requires no extra hardware but performs lightweight computation over the existing generic network architecture for routing control. Existing adaptive routing techniques [4], [5] make routing decisions for every communication pair individually, which does not facilitate contention avoidance. For large-scale manycore systems where thousands and millions of processor cores are executed in parallel, multiple inter-processor communications are often conducted in NoCs at any given time. Therefore, we focus on contention-aware adaptive routing by considering multiple communications in this work.

Thermal reliability is another concern for ONoC designs [7], [8]. Optical switches, as the versatile components in ONoCs, are highly susceptible to ambient temperature fluctuations. On every routing path in ONoCs, multiple switches switch signals in stages to implement data transmission. However, under chip thermal variations, the resonant wavelength of a switch red-shifts with increasing temperatures and blue-shifts with decreasing ones. The undesired mismatch between the wavelength of optical signals and the central wavelength of the switch will result in additional optical power loss, resulting in network performance degradation and even functional failures. To guarantee the thermal reliability of ONoCs, device-level thermal tuning is widely employed for optical switches to compensate their temperature-dependent wavelength shift [9]. However, this technique is energy-hungry, such that the energy consumed for thermal regulation may null the benefits of ONoCs in energy efficiency [10]. The problem becomes more acute in the ONoCs who have large on-chip temperature variations and a large number of switching stages.

Separate performance improvement or thermal management is insufficient for deep optimization on ONoCs. However, there are apparent contradictions between minimizing contention and guaranteeing thermal reliability. Device-level thermal tun-

M. Li and P. Chen are with the College of Computer Science, Chongqing University, China and also with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. E-mail: mengquan@cqu.edu.cn.

W. Liu and L. H.K. Duong are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.

L. Yang is with the Department of Computer Science and Engineering, University of Notre Dame, USA.

C. Xiao is with the College of Computer Science, Chongqing University, China.

Fig. 1: The schematic of a mesh-based ONoC architecture and a tile in logical view.

Fig. 2: Router *Cygnus* [6].

ing techniques guarantee the thermal reliability of generic ONoCs but introduce heavy energy overhead. In order to minimize the energy consumed by thermal tuning, messages are required to be routed along the path that has the lowest thermal gradient, whereas communication pairs prefer to select different routing paths that do not conflict with one another for better communication performance. Therefore, strategies to collaboratively solve these contradictions are urgent.

In this paper, we develop novel routing techniques for ONoCs to minimize communication contention and maximize energy efficiency with guarantee thermal reliability. Analyzing the thermal effects in ONoC, we model the thermal sensitivity of optical switches, and theoretically formulate a boundary condition for the number of switching stages in routing paths. Both are principal factors influencing the thermal reliability of ONoCs. A routing criterion at the network level is further presented, combined with device-level thermal tuning, it can implement thermal-reliable ONoC. We further propose two contention-aware adaptive routing approaches, including a mixed-integer linear programming model (MILP) model and a heuristic algorithm (called *CAR*), to minimize communication conflict based on the guaranteed thermal reliability, and meanwhile, mitigate the communication energy overheads in the presence of chip thermal variations. By applying the criterion, our approaches can achieve excellent performance with largely reduced complexity of design space exploration. Based on synthetic communication traces and realistic benchmarks, evaluation results verify the effectiveness of our approaches. They averagely improve communication performance by 126.95% and reduce energy overhead by 16.12% in reliability-guaranteed ONoCs, compared with state-of-the-art techniques. *CAR* only introduces 7.20% performance difference compared to the MILP model and is more scalable to large-size ONoCs.

II. PRELIMINARY AND RELATED WORK

ONoC architecture. We take a 2D-mesh based ONoC as an example in Fig. 1. Vertically on top of processor cores, a photonic network provides optical interconnects between optical routers for inter-processor bulk data transmission. An electronic network comprised of control routers and metallic interconnects is designed to perform logical control. The photonic and electronic networks together constitute an ONoC. In this paper, we consider ONoCs with 2D-mesh and 2D-torus topologies as the target platforms. Mesh topology has been widely used by ONoC designers thanks to its simplicity and good scalability. By adding long loop links between boundary nodes, torus topology keeps the simplicity of mesh while offers

lower network diameter and lower communication contention probability with better path diversity [11].

Optical router. As the core functional component on ONoCs, an optical router is comprised of three parts: a *transmitter*, a *receiver*, and a *switching network*, shown in Fig. 1. The transmitter (i.e., E-O module) converts electrical signals into optical signals. Built-in microlaser sources, such as VCSELs [12], can be implemented in it. The receiver (i.e., O-E module) uses high-resolution photodetectors (PDs) [13] to convert optical signals into electrical signals. We provide the specific switching network design of a nonblocking optical router *Cygnus* in Fig. 2. There are five ports. The injection and ejection ports connect the transmitter and the receiver, respectively. The other four bidirectional ports, including the west, south, east and north ports, connect to its neighbor routers. Optical switches and waveguides compose the switching network. For every specific transmission, either no or only one switch is active in a router. Optical router designs follow this principle to minimize optical power loss for optimized communication performance and guaranteed system reliability.

Contention-aware communication in ONoCs. Contention-free ONoC architectures using wavelength division multiplex (WDM) routing have been proposed in literature [3]. Bypass links are also exploited by some researches for conflict circumvention [14], [2]. In an advanced ONoC architecture, *Amon* [2], besides adopting WDM routing, bypass links are added for the nodes sending data with the same wavelength to alleviate conflicts. Using dedicated network architectures and wavelength routing, these works efficiently achieve contention-free but introduce heavy extra hardware costs. In contrast, contention-aware routing [4], [5] is an attractive solution, which mitigates contention by utilizing the inherent flexibility of adaptive routing. It requires no extra hardware but conducts lightweight software routing computation based on existing generic network architectures. However, the routing techniques proposed for traditional electronic NoCs are usually unsuitable to be transplanted directly into ONoCs. On the one hand, the recently-proposed techniques [5] make routing decision for NoCs based on the buffer utilization information, whereas there is no photonic equivalent of a buffer in ONoCs. On the other hand, typical adaptive routing approaches, such as DyXY [4], are simple enough to be used in ONoCs but the unique characteristics of optical interconnects have not been taken into account. In this paper, we propose two contention-circumvented adaptive routing approaches for ONoCs. Our approaches are orthogonal to wavelength routing and are considered

a good complement to be integrated into it, because they are to avoid conflicts between the data transmissions using the same wavelength, and the data modulated in different wavelengths are naturally contention-free. Instead of using extra bypass links as [14], [2] do, we employ adaptive routing for contention avoidance.

Thermal Reliability in ONoCs. Chip temperature fluctuates temporally and spatially as a result of uneven chip power density and limited heat dissipation techniques. In general, it can vary by up to 30 K across the chip under typical operating conditions [15]. Under these circumstances, the thermal susceptibility of ONoCs has aroused great concern recently [7]. Control-based solutions to compensating for the thermal effects on nanophotonic devices have been widely explored at the device level [9]. These techniques enhance the ONoCs' thermal reliability at the cost of heavy thermal regulation energy consumption [16]. The problem will be more acute in the ONoCs who have large on-chip temperature variations and a large number of switching stages. To improve the communication energy efficiency, thermal-aware routing mechanisms have been presented [17], [18]. In the presence of on-chip temperature variations, the latest thermal-sensitive routing algorithm [18] selects the route with the minimum thermal-induced optical loss out of all the alternative shortest paths for every communication, in order to minimize the total routing energy overhead.

III. SYSTEMATIC OVERVIEW

The prior works focus on either alleviating communication contention or guaranteeing thermal reliability for ONoCs. However, separate communication performance improvement or thermal reliability management is insufficient for deep optimization. In this work, we develop novel routing techniques to make systematic optimization on communication performance, energy efficiency, and system reliability for ONoCs. There are threefold critical challenges. First, how to alleviate the thermal susceptibility of optical devices and route bulk messages in a reliable way, such that the thermal reliability of ONoCs can be guaranteed. Second, based on the thermal-reliable ONoCs, how to explore contention-aware routing to minimize communication conflicts. The last but not the least, how to further maximize the communication energy efficiency with considering both contention and thermal regulation overheads in the presence of on-chip thermal variations.

Figure 3 shows an overview of our work to address the above challenges. Based on the analysis of the intrinsic thermal sensitivity of switching elements, we theoretically formulate a boundary condition for the number of switching stages in routing paths. A network-level routing criterion is first presented (Section IV-C). The criterion, combined with device-level thermal tuning, is able to guarantee the thermal reliability of ONoCs. On top of the thermal-reliable ONoC platforms, we then model the problem by MILP formulations to find the optimal routing solution for the given communication demand (Section V-B), which minimizes communication conflicts and maximizes communication energy overheads in the presence of on-chip thermal variations. A polynomial-time heuristic

algorithm for contention-aware routing, called *CAR*, is further proposed (Section V-C). By applying the criterion, our approaches can achieve excellent performance with largely reduced complexity of design space exploration. Based on synthetic traffic patterns and realistic benchmarks, comprehensive evaluations have been conducted to verify the effectiveness of our approaches in network performance, communication energy consumption, thermal reliability, and technique efficiency, compared with state-of-the-art techniques.

Fig. 3: Contention-circumvented routing for thermal-reliable ONoCs: a systematic overview.

IV. THERMAL RELIABLE ONO C

A. The thermal sensitivity of switching elements

Optical switches play a key role in implementing high-level routing policies in ONoCs. There are two different switch designs shown in Fig. 4: parallel switching element (PSE) and crossing switching element (CSE). Both consist of one ring and two straight waveguides. When a switch is configured to be switched *off*, the optical signal from the *Input* port will be delivered to the *Through* port. Otherwise, the signal is resonated into the ring and delivered to the *Drop* port.

Fig. 4: Optical switching elements.

A switch resonates with light whose single-pass phase shift is a multiple of 2π , formulated as $a \cdot \lambda_{SE} = 2\pi R \cdot n_{eff}$, where λ_{SE} is the vacuum resonant wavelength of the switching element; R is the average bending radius of the ring; n_{eff} is the effective index of the resonator and a is a positive integer.

The resonant wavelength of switching elements is sensitive to ambient temperature changes. A linear relationship between resonant wavelength shifts and temperature variations is widely recognized. Given the initial resonant wavelength, λ_0 , at the nominal operating temperature, T_0 , the relation between the resonant wavelength of a switching element, λ_{SE} , and its ambient temperature, T_{SE} , can be expressed as follows:

$$\lambda_{SE} = \lambda_0 + \rho \cdot (T_{SE} - T_0) \quad (1)$$

where ρ is the temperature-dependent resonant wavelength shift coefficient of the switching element, and is about 0.06 nm/K at the 1,550 nm wavelength range [19].

The thermal-induced wavelength shift causes the resonances of switches to deviate from the expected peaks. The undesired mismatch between the signal wavelength and the switches resonant wavelength would result in additional optical power loss. According to the traveling wave theory [20], the optical power loss of an optical switching element at the drop-port, OL_{SE} , can be formulated as follow:

$$OL_{SE} = 10 \log \left(\frac{2\kappa^2 + \kappa_p^2}{2\kappa^2} \right)^2 \left(1 + \frac{4(\lambda_{in} - \lambda_0 - \rho(T_{SE} - T_0))^2}{\theta^2} \right) \quad (2)$$

where κ^2 is the fraction of optical power coupling between the waveguides and the ring. κ_p^2 is the fraction of intrinsic power losses (such as bending, absorption and surface scattering due to roughness) per round-trip in the ring. λ_{in} is the wavelength of the input signal. θ is the -3dB bandwidth of the drop-port power transfer spectrum. These models provide a theoretical foundation for communication reliability analysis in ONoCs.

B. Two methods for reliable optical communication

Despite architecture diversity, in ONoCs, an optical connection for data transmission is generally composed of an optical transmitter (in the source router), an optical receiver (in the destination router), and a routing path connecting them. On every routing path, optical switches, combined with optical waveguides, transmit signals from the transmitter to the receiver. To ensure that ONoC functions properly, the signal power reaches the receiver on each routing path should be guaranteed not lower than the receiver's sensitivity. Otherwise, the fluctuations in received power may lead to dramatic reliability degradation and even communication failure [21]. We formulate the necessary condition as (3), where OP_{TX} is the optical power output from the transmitter on a path; $\sum_{i=1}^m OL_{SE_i}$ is the total optical power loss of all the active switches in the path under temperature variations; m is the number of switching stages in the path (i.e., the number of active switches in this route); OL_{WG} includes the loss due to associated passive switches and waveguides, and S_{RX} is the receiver sensitivity.

$$OP_{TX} - \sum_{i=1}^m OL_{SE_i} - OL_{WG} \geq S_{RX} \quad (3)$$

According to the formula (3), there are two methods to achieve reliable data transmission on ONoCs under temperature fluctuations. One is to increase the output power of the transmitter to compensate for the power loss on the path such that enough optical power reaches the receiver. The other is to employ device-level thermal tuning control for active switches whose power loss is the main contributor to the total loss of the routing path. Both methods enhance ONoCs' thermal reliability at the expense of extra energy consumption. To compare their energy costs, we conduct quantitative analysis with consideration of on-chip thermal variations.

Ideally, in every optical interconnection on ONoCs, the output wavelength of the transmitter equals the initial resonant wavelengths of the switches, (i.e., $\lambda_{in} = \lambda_0$); consequently, sufficient signal power reaches the receiver after suffering a

small amount of loss. Nevertheless, under thermal variations, more optical power is lost due to the resonance shift of active switches. Assumed the ambient temperature of the i -th switching element changes from the initial value of T_0 to T_{SE_i} , to compensate for the thermal-induced power loss for guaranteeing reliable communication, more input signal power is needed, which increases the transmitter energy consumption. Based on (2) and (3), we model the additional energy consumption required by a single-wavelength transmitter as (4). The impacts of thermal variations on waveguide propagation loss and receiver sensitivity are neglected because they are trivial compared to that on high-Q active switches [7].

$$\Delta E_{TX} \geq \frac{1}{\eta} \prod_{i=1}^m \left(1 + \frac{4\rho^2 \cdot (T_{SE_i} - T_0)^2}{\theta^2} \right) \cdot t_{TX} \quad (4)$$

where ΔE_{TX} is the difference between the electrical energy consumed by the transmitter with and without temperature variations; η is the wall-plug efficiency of the transmitter, which is the conversion efficiency of electrical power to optical power and is defined as the ratio of the generated optical power to the input electrical power; and t_{TX} is the duration time for which the transmitter remains turned on.

Formula (4) indicates that, it is very energy-inefficient to increase the transmitters' output power to compensate for the thermal-induced power losses in routing paths. There are three reasons. First, optical losses in routing paths are typically multiplicative in nature [16]. Second, the wall-plug efficiency of transmitters is also an obstacle. It is reported that VCSEL lasers can only achieve around 30% wall-plug efficiencies, whereas they are already much better than other competing off-chip lasers [22]. Third, the transmitter need be active most of the time because it requires a long time (hundreds of ns) to turn a laser on again if we turn it off for energy-saving [16]. To maintain the benefits of low latency for ONoCs, this delay should not be incurred. In addition to the heavy energy overhead, hardware cost also increases potentially. When employing VCSELs as the laser sources in transmitters, more VCSEL lasers are required to be integrated into a chip to generate strong optical signals, due to the limited produced optical power of a single VCSEL.

By contrast, control-based thermal tuning is widely used for active switches to compensate for their power losses [9]. Using integrated microheaters and PDs that are readily available in typical silicon photonic platforms, the tuning techniques dynamically maintain the local temperature of the switches and compensate for their thermal-induced power losses throughout the operation duration, via feedback control logic. Because of lacking effective cooling solutions, optical switches are designated to operate at a higher temperature during initialization [7], [9]. By leveraging the repeated nature of the switches optical resonances, the decreases or increases in ambient temperature can be corrected by increasing or decreasing respectively the power delivered to the integrated heater. We formulate the tuning energy consumption in a routing path as (5), where ε is the tuning efficiency in mW/nm; t_{dur} denotes the duration time of the switches' operation, and T_{tgt} is their target operating temperature. Note that the target operating temperature of a switch is not necessarily equal to

its initial temperature T_0 because it is theoretically possible to tune the switch by one or multiple free-spectral-range (FSR) at the cost of distance-related tuning energy consumption.

$$E_{\text{thermal-tuning}} = \sum_{i=1}^m \varepsilon \cdot \rho \cdot (T_{SE_i} - T_{tgt}) \cdot t_{dur} \quad (5)$$

According to (4) and (5), unlike the former method whose energy consumption grows exponentially with the increasing number of active switching elements (m) and their temperatures (T_{SE_i}) in a routing path, the energy consumed by thermal tuning only increases linearly. Besides, numerous techniques to improve the tuning efficiency and tuning speed have been explored [7], [9], which further enhance the energy efficiency of the second method. Therefore, we employ the device-level thermal tuning technique to implement reliable optical communication in this paper.

C. Routing criterion: I/L/Z-shaped routing path

Nevertheless, besides thermal variations, the communication reliability of ONoCs is also threatened by the large number of switching stages in routing paths. Even in the cases where every switch works stably and properly via thermal tuning, a large number of switching elements in a routing path would increase the total power loss on the path, resulting in a risk of unreliability. To ensure that the signal power received by the receiver on an optical path is not lower than its sensitivity, according to (3), we derive a boundary condition to limit the number of switching stages in routing paths as follow:

$$m \leq (OP_{TX} - OL_{WG} - S_{RX}) / OL_{SE_i}^{ideal} \quad (6)$$

where $OL_{SE_i}^{ideal}$ is the optical loss of the switches at their desired operating temperature. The power loss due to the switches under temperature fluctuations is always greater than the loss of ideal switches (i.e., $\sum_{i=1}^m OL_{SE_i} \geq m \cdot OL_{SE_i}^{ideal}$).

To show the impact of the number of switching stages on the power losses of routing paths more intuitively, we conduct a group of experiments to simulate the transmission processes of optical signals along routing paths that have different numbers of switching stages, based on current silicon photonics manufacturing level. To simulate the associated photonic components and integrated circuits, we used Lumerical FDTD and INTERCONNECT simulation infrastructure [23].

TABLE I: Average loss of an optical router *Cygnus*.

Incoming Port(s)	Outgoing Port(s)	Average power loss (dB)
Inj	N, S, E, W	3.3172
N	E, W	3.7248
S	E, W	3.5828
E	N, S	3.5623
W	N, S	3.3386
N, S, E, W	Ej	3.5196
Usage of the router in a routing path		Average power loss (dB)
Sender		3.3172
Receiver		3.5196
Intermediate turning node		3.5521

A recognized optical router *Cygnus* (see Fig. 2) is employed in the simulations. Assumed that all the switching elements in routers work at their desired operating temperature by applying thermal tuning, Table I summarizes the average optical

power loss of different input-output port pairs in a *Cygnus*. With consideration of the insertion losses of optical devices including active and passive switches, crossing, straight and bending waveguides, we calculate the average loss of each input-output port pairs for *Cygnus*. Furthermore, we give the average loss of a *Cygnus* according to its role played in a communication path. Slight loss differences exist when a *Cygnus* is employed as a sender, a receiver and an intermediate turning node in a routing path. On the path between the sender and the receiver, multiple intermediate optical routers implement signal transmission. We call a router as a turning node at which the signal direction changes (i.e., with one active switch). Otherwise, it is a passthrough node (i.e., without signal directional change). Note that compared with the loss of a *Cygnus* as a turning node, the power loss of a *Cygnus* used as a passthrough node is negligible. It is considered to approximate zero for simplicity. These results are consistent with those reported in [24].

Fig. 5: The impact of the # of switching stages on the total loss of a routing path.

Fig. 6: Three types of routes: I/L/Z-shaped routing paths.

According to the simulation results shown in Fig. 5, the total power loss of a routing path is averagely proportional to the number of switching stages in it. State-of-the-art PDs at the receiver side achieve a sensitivity of -14.2 dBm for a BER of 10^{-12} [25]. Assumed the input power is 1 mW, we observe that if the number of switching stages in a route is more than 4, the power reached to the PD is lower than its sensitivity. This problem will be more serious under large temperature variations. Therefore, in this paper, we restrict the number of switching stages to 4 for guaranteed system reliability¹ and, as a side benefit, energy-saving. The less number of switching stages a route has, the less input energy the transmitter requires for reliable transmission. Besides, the energy consumed by thermal tuning decreases with the reduced number of switching elements.

For the ONoCs with 2D-mesh and 2D-torus topologies, there are three types of shortest routing paths² for every communication pair according to their shape: I-shaped, L-shaped, and Z-shaped routes, as illustrated in Fig. 6. To take advantage of the loop links in torus topology, for every communication pair in a 2D-torus network, we consider all the routing paths that are not longer than the shortest routes

¹With the development of silicon photonic technologies, this restriction can be relaxed by virtue of the low-loss optical device fabric.

²Minimal routing considers the route length and path utilization, which will facilitate to mitigate link delay and the risk of contention.

in a mesh counterpart as its candidate paths. More alternative paths can be obtained. This is an important observation that contributes to contention avoidance for ONoCs.

V. CONTENTION-AWARE ROUTING

Through device-level thermal tuning and the proposed network-level routing criterion, we achieve thermal-reliable ONoCs. However, it brings other challenges. To minimize the regulation energy consumed by thermal tuning, messages are required to be routed along the path that has the lowest thermal gradient, while communication pairs prefer to select different routing paths without contention for better communication performance. In addition, we focus on mitigating communication contention by utilizing the inherent flexibility of adaptive routing in this paper. While the flexibility would decrease when the number of switching stages is restricted, which is adverse to contention reduction.

To collaboratively solve the above contradictions, in this section, we formulate the problem and further propose two contention-aware routing approach to achieve communication optimization based on guaranteed thermal reliability.

A. Problem formulation

1) *Inter-processor communication in ONoC*: In ONoCs, the latency for inter-processor communication includes the path-setup latency in the electronic network, $L_{path-setup}^e$, which is the time required by the control packet 'path-setup' to reserve an optical path, and the data-transmission latency in the photonic network, $L_{payload}^o$. It can be formulated as follow:

$$L_{comm} = L_{path-setup}^e + L_{payload}^o \quad (7)$$

The control packet is routed in the electronic network in a packet-switching manner. It travels a number of electronic routers and undergoes several processes in each hop. After activating m associated switching elements, as well as their tuning control, in corresponding optical routers, an optical communication path is reserved/established. Local thermal tuning is adopted for optical switches for reliable data transmission. It works for the duration of the switches' operation with a tuning speed as fast as $\sim 1 \mu s$ [9]. Negligible delay but extra tuning energy is incurred. We formulate the path-setup latency as follow:

$$L_{path-setup}^e = L_R^e \cdot (h+1) + (W_{ctrl}^e - 1)/b^e + m \cdot L_{SE}^e + L_{contention}^e \quad (8)$$

where L_R^e is defined as the number of stages in each control router (typically 2 or 3 stages) under the clock frequency of f^e ; h denotes the number of hops from the source to the destination; W_{ctrl}^e is the size of the control packet in the electrical control network; b^e is the bandwidth of electronic channels; L_{SE}^e is the latency of activating a switch; $L_{contention}^e$ is the additional delay caused by communication contention during the path-reservation phase, which is non-neglectable especially for communication-intensive applications.

Once the optical path is established, the photonic network can be perceived as a fast circuit-switched network. The data transmission latency in the photonic network is formulated as

$$L_{payload}^o = W_{payload}^o / R_{oeo} + d \cdot n / c \quad (9)$$

where $W_{payload}^o$ refers to the volume of data transmitted in the photonic network; R_{oeo} is the O-E and E-O interface data rate; d is the travel distance of optical data; n is the refractive index of silicon waveguide and c is the light speed in vacuum.

Likewise, the total communication energy consumption has two portions shown in (10), where $E_{payload}^o$ is the energy consumed for data transmission and E_{ctrl}^e is the control overhead.

$$E_{comm} = E_{ctrl}^e + E_{payload}^o \quad (10)$$

E_{ctrl}^e can be calculated by Eq. (11), where E_{int}^e is the average energy required to transfer a single bit through an electrical interconnection; E_{cu}^e denotes the average energy required by the control unit to make a decision for a single packet; $E_{contention}^e$ is the extra energy incurred by communication contention during path setup.

$$E_{ctrl}^e = E_{int}^e \cdot W_{ctrl}^e \cdot h + E_{cu}^e \cdot (h+1) + E_{contention}^e \quad (11)$$

The energy overhead of optical data transmission includes the energy consumed by E-O and O-E conversion, switch activation, and local thermal tuning, formulated as:

$$E_{payload}^o = E_{oeo}^o \cdot W_{payload}^o + m \cdot P_{SE}^o \cdot t_{dur} + E_{thermal-tuning} \quad (12)$$

where E_{oeo}^o is the energy required by a single bit for O-E and E-O conversions; P_{SE}^o is the average power consumption for activating an optical switch; $t_{dur} = W_{payload}^o / R_{oeo} + d \cdot n / c$, which calculates the duration time of the switches' operation.

2) *Motivational example*: These enlightening models reveal that, in addition to the number of switching stages in routing paths, communication contention and on-chip thermal variations are another two key factors influencing the communication performance and energy efficiency of ONoCs. To show their impacts more intuitively, we continue to give two examples of communication routing on a 3×3 2D-mesh ONoC. The values of the parameters used in Eq. (7)–(12) are listed in Table II.

TABLE II: Parameters used in this paper (ref. [26], [27]).

Parameter	Value	Parameter	Value
f^e	1 GHz	b^e	32-b bidirectional
L_{SE}^e	30 ps	R_{oeo}	12.5 Gbit/s
n	3.48	c	3×10^8 m/s
W_{ctrl}^e	8+1 bits	$W_{payload}^o$	512 bits
E_{int}^e	0.52 pJ/bit	E_{cu}^e	1 pJ/pkt
P_{SE}^o	20 μ W	E_{oeo}^o	1 pJ/bit
ϵ	1.10 mW/nm	ρ	0.06 nm/K

Fig. 7: Contention and on-chip thermal variations have significant impacts on ONoCs' communication efficiency.

Case 1: As shown in Fig. 7(a), $\langle S_1, D_1 \rangle$ and $\langle S_2, D_2 \rangle$ are two pairs that are requiring data transmission. Conflict emerges when employing XY routing for them. The pair $\langle S_2, D_2 \rangle$ is blocked. According to Eq. (7)–(9) and Table II, the communication latencies of the two pairs are 48.10 ns

and 93.17 ns. The latency of the pair $\langle S_2, D_2 \rangle$ contains the contention-induced waiting time. By contrast, if employing contention-aware adaptive routing, we select a routing path ($S_2 \rightarrow R_5 \rightarrow D_2$) for the pair $\langle S_2, D_2 \rangle$. No conflict exists. The latencies of these pairs are reduced to 48.10 ns and 45.07 ns, achieving an average of 34.05% reduction in latency and 93.70% improvement in network throughput. The network throughput is defined as the rate of optical data packets transmitted through an ONoC over a period of time. It is measured in packet-per-second in this work.

Case 2: As shown in Fig. 7(b), for a pair $\langle S, D \rangle$, assumed there are two alternative routes that have the same manhattan distance and number of switching stages. The chip thermal distribution is shown on the right side. As the temperature around node R_9 is about 10 K higher than that around node R_1 , the total optical loss on the red path (i.e., $S \rightarrow R_8 \rightarrow R_9 \rightarrow R_6 \rightarrow D$) is distinctly larger than that on the blue path (i.e., $S \rightarrow R_4 \rightarrow R_1 \rightarrow R_2 \rightarrow D$). According to Eq. (10)–(12) and Table II, when employing local thermal tuning for every optical switch, the total communication energy consumption of the red and blue routing paths are 770.41 pJ/pkt and 667.65 pJ/pkt, respectively. Compared to the red one, the energy overhead of the blue path is reduced by 13.34%. It indicates that routing approaches that take on-chip thermal variations into account can improve the energy efficiency of ONoCs.

3) *Problem definition:* Based on the above analysis, we define the problem addressed in this section as follow: *Given the floorplan of a thermal-reliable ONoC that is implemented using the solutions in Section IV, and a communication demand represented by a set of pairs that require inter-processor data transmission, determine a routing path for every pair to minimize conflicts and maximize communication energy efficiency in the presence of on-chip thermal variations.*

B. MILP-based routing approach

TABLE III: Variables used in the MILP model.

Notation	Definition
D	The communication demand.
rP_i	The set of candidate routing paths for the comm. pair i .
L	The set of optical links in ONoC.
$reql_{i,k}$	The set of required links to construct the routing path k for pair i .
T_r	The ambient temperature of the router r .
$SE_{i,k}$	The set of switching elements in the path k of pair i .
$X_{i,k}$	Binary variable, equals to 1 if communication pair i routes along the path k .
$N_{c \times}$	The number of communication conflicts.
E_i	The total communication energy consumption of pair i .

We first model the problem by MILP formulations. Our MILP-based routing approach includes two phases: the main objective is to minimize communication conflicts; based on it, we minimize the total communication energy consumption with consideration of on-chip thermal variations. The notations used in MILP formulations are summarized with detailed definition in Table III. In the MILP model, by applying the routing criterion proposed in Section IV-C for guaranteed thermal reliability, the set of alternative routing paths of every

communication pair is a set of I/L/Z-shaped routing paths. It exponentially reduces the design search space and thus enhance the computation efficiency of our approaches.

In the first phase, we minimize communication contention. The main objective is shown as follow:

$$\text{Objective 1: minimize}(N_{c \times}) \quad (13)$$

The key constraints are described as follows:

- (1) *Constraint for communication demand:* Equation (14) guarantees that at most one routing path is selected for every communication pair.

$$\sum_{k \in rP_i} X_{i,k} \leq 1 \quad \forall i \in D \quad (14)$$

- (2) *Constraint for route overlap:* For each optical link $l \in L$, it can be occupied by zero or one routing path. If link l is occupied by a route, $C_l = 1$. Then $\sum_{k \in rP_i} X_{i,k} * reql_{i,k} \leq 1$, which guarantees no route overlap.

$$\sum_{k \in rP_i} X_{i,k} * reql_{i,k} \leq C_l \quad \forall i \in D \quad (15)$$

- (3) *Constraint for communication conflict:* Equation (16) calculate the number of conflicts for the given communication demand. $|D|$ is the size of the communication demand.

$$N_{c \times} = |D| - \sum_{i \in D} \sum_{k \in rP_i} X_{i,k} \quad (16)$$

In the second phase, on the basis of the minimal communication conflicts obtained from the first phase, we minimize the total communication energy consumption with consideration of on-chip thermal variations, shown as follow:

$$\text{Objective 2: minimize}(\sum(E_i)) \quad i \in D \quad (17)$$

In addition to the constraints (14)–(15), we add two key constraints in this phase as follows:

- (4) *Constraint for minimum conflict:* Equation (18) guarantees that we optimize the communication energy efficiency based on the minimized communication conflicts.

$$\sum_{i \in D} \sum_{k \in rP_i} X_{i,k} = |D| - N_{c \times} \quad (18)$$

- (5) *Constraints for thermal-induced energy consumption:* In this paper, we consider a set of tasks running in a relatively stable power profile. The processor cores and thus the ONoC are all in a steady temperature state during the time period we concern. Equation (19) estimates the total communication energy consumption for each pair by taking on-chip thermal variations into account.

$$E_i = X_{i,k} * \left(E_{int}^e \cdot W_{ctrl}^e \cdot h + E_{cu}^e \cdot (h + 1) + E_{oeo}^o \cdot W_{payload}^o + m \cdot P_{SE}^o \cdot t_{dur} + \sum_{r \in SE_{i,k}} \varepsilon \rho (T_r - T_{tgt}) \cdot t_{dur} \right) \quad (19)$$

According to the MILP formulations, we optimize communication performance by minimizing contention in the first phase, and further explore the design space to find the optimal routing solution with minimized energy consumption based on the minimal communication conflicts. The theoretically guaranteed optimal result obtained by our MILP model can be converted into an equivalent communication routing solution, which serves as the baseline solution for the heuristic algorithm proposed in the next section.

C. Heuristic algorithm for contention-aware routing

As the problem addressed in this paper is NP-hard, the running time of the proposed MILP model grows exponentially with the increasing sizes of network and communication demand. To efficiently solve this complex issue, we further propose a polynomial-time algorithm, *CAR*.

Algorithm 1: Contention-aware routing (*CAR*) algorithm.

Data: The *ONoC*, the communication demand *D*
Output: Routes for every communication pair *rPs*

```

1 idleLs = GetIdleLinks(ONoC);
2 for cPair ∈ D do
3   cRegion = GetCommPathCandidates(cPair, idleLs);
4   # A set of I-shaped, L-shaped, and Z-shaped paths;
5 IndePairs = FindIndependPairs(D.cRegion);
6 for cPair ∈ IndePairs do
7   cPair.rP = GetRouteWithMaxEnergyEfficiency(cRegion);
8 Update(D);
9 D = sort(D, D.cRegion, Ascend);
10 for cPair ∈ D do
11   idleLs = GetIdleLinks(ONoC);
12   if src.x = des.x or src.y = des.y then
13     cPair.rP = (src → des); #I-shaped path;
14     Continue;
15   cPair.rP = GetRouteWithMaxEnergyEfficiency(cRegion ∩ idleLs);
16 return (rPs);

```

As shown in Algorithm 1, we can firstly obtain the set of all the available candidate routing paths for every communication pair (i.e., a set of I/L/Z-shaped routes) by applying the routing criterion proposed in Section IV-C, which constitutes its communication region (denoted as *cRegion*) (Line 2–3). For example, in a 2D-mesh based *ONoC*, the communication region of a pair $\langle src, des \rangle$ contains at most $|des.x - src.x| + |des.y - src.y|$ paths in total. The routing path for each pair is selected from its communication region. For the pairs whose communication regions are independent of other regions, we are free to select a routing path based on the routes' temperature distribution, without the risk of contention (Line 4–6). The function *GetRouteWithMaxEnergyEfficiency()* is designed to find the route that maximizes energy efficiency among the communication region. For the pairs that there are overlaps between their regions, we should consider the risk of link conflict, potentially resulting in communication contention. To handle the potential conflict regions, we sort pairs according to the sizes of their communication regions, from small to large (Line 8). The pairs with small communication regions are prior to the pairs with large regions because it has less alternatives. It is easier for the pairs with large communication regions to find feasible routes even if some of links are occupied by other communications. In the most extreme case, no feasible route can be obtained for the pairs that there is only one available routing path and it has been occupied by others. We make routing decisions for the pairs in the sorted order. For every communication pair, we select the route that achieves maximal energy efficiency on the basis of no communication contention (*cRegion* ∩ *idleLs*), which can be realized by labeling the occupied/idle links (Line 9–14).

Figure 8 gives an example on a 4×4 mesh-based *ONoC* to illustrate our *CAR* algorithm. There are three communication pairs (*Pair*₁: $\langle R_{13}, R_{10} \rangle$, *Pair*₂: $\langle R_7, R_1 \rangle$ and *Pair*₃: $\langle R_{10}, R_4 \rangle$) and we can obtain their communication regions (i.e., *cRegion*₁, *cRegion*₂ and *cRegion*₃ respectively). As *cRegion*₁ is independent of the other two regions, we are free to select a route for *Pair*₁ according to the temperature distributions of the routes *rP*_{1,1} and *rP*_{1,2}. However, the communication regions of *Pair*₂ and *Pair*₃ are partly overlapped. There may be contention on the links (*R*₆ → *R*₂) and (*R*₇ → *R*₃) when deciding routes for them. In this case, we give priority to the pair that has smaller region and select a routing path for it first (*Pair*₂ in this example). Assume the routing path *rP*_{2,2} achieves maximum energy efficiency for *Pair*₂, the link (*R*₇ → *R*₃) is occupied; consequently, the available communication region of *Pair*₃ is reduced to {*rP*_{3,1}, *rP*_{3,2}, *rP*_{3,3}} and we can select one of routing paths based on their temperature distributions without the risk of communication contention.

Fig. 8: An example of the *CAR* algorithm.

Complexity analysis: The complexity of Algorithm 1 is $O(M \times N)$, where *M* is the number of communication pairs and *N* is the square root of the *ONoC* size (which is N^2). The complexity of looping through the alternative paths to find a route with the maximal energy efficiency for a pair is $O(N)$ in the worst case, and there are *M* communication pairs in total.

Deadlock is avoided by our approaches. The routing paths for a set of communication pairs are constructed in batch without deadlock by using the centralized approaches. Besides, considering the hybrid network architecture of *ONoCs*, the path establishment process is performed via the electronic network. A deadlock-free electronic control network can be achieved using virtual channel flow control [19], and the photonic network is inherently deadlock-free due to circuit switching and predetermined routing path.

D. The applicability of our techniques on different topologies

Our approaches are designed for mesh and torus by considering their grid-type shapes. Nevertheless, there are two differences when applying the approaches in these two topologies. One lies in the process of obtaining candidate routing paths for communication pairs, which is formulated as the function *GetCommPathCandidates()* in Algorithm 1.

As shown in Function 1, we first obtain the source and the destination of the given communication pair (Line 1). If they have the same x or y coordinates, we construct a I-shaped alternative routing path (Line 2–3). The function *XTraverse()* and *YTraverse()* are respectively designed to traverse the

X- and the Y-axis directions and find a path that connects the source and the destination nodes based on available optical links. To construct the L- or Z-shaped candidate routing paths, we first obtain the intermediate tuning nodes for this pair (Line 5). Considering the grid-type shapes of mesh and torus topologies, each tuning node has the same x or y coordinates with the source or the destination of the pair. Therefore, the method of constructing I-shaped paths can be employed segment by segment to obtain the L- and Z-shape routes (Line 7–14). Due to the extra loop links between boundary routers in tori, there could be two directions when conducting $X\text{Traverse}()$ and $Y\text{Traverse}()$, while there is only one direction in meshes. For 2D-torus ONoCs, a lightweight preprocessing is involved (Line 1*–2*). By flattening the loop links of a torus, an equivalent interconnection with 2D-mesh logical topology can be obtained; consequently, this function is applicable to torus-based ONoCs as well.

Function 1: GetCommPathCandidates($c\text{Pair}$, $idleLs$)

```

# For 2D-mesh ONoCs;
1  $cRegion = \emptyset$ ;  $\langle src, des \rangle = \text{GetSrcAndDes}(c\text{Pair})$ ;
2 if  $src.x = des.x$  or  $src.y = des.y$  then
3    $cRegion += Y\text{Traverse}(src, des, idleLs)$ 
   or  $X\text{Traverse}(src, des, idleLs)$ ; # I-shaped paths;
4 else
   # L/Z-shaped candidate paths;
5    $turnNs = \text{GetIntrmTuningNodes}(c\text{Pair})$ ;
6   for  $tN \in turnNs$  do
7      $aP = \emptyset$ ;  $curN = src$ ;  $nextN = tN$ ;
8     while  $nextN \neq des$  do
9       if  $curN.x = nextN.x$  then
10         $aP += Y\text{Traverse}(curN, nextN, idleLs)$ ;
11         $curN = nextN$ ;  $nextN = [des.x, nextN.y]$ ;
12       else if  $curN.y = nextN.y$  then
13         $aP += X\text{Traverse}(curN, nextN, idleLs)$ ;
14         $curN = nextN$ ;  $nextN = [nextN.x, des.y]$ ;
15      $cRegion += aP$ ;
16 return ( $cRegion$ );

# For 2D-torus ONoCs;
1*  $ONoC_e = \text{GetEquivalentMeshNet}(ONoC)$ ;
2*  $idleLs = \text{GetIdleLinks}(ONoC_e)$ ;
3* go to Line 1; # invoke the function above for the logical
   mesh-based ONoCs;

```

The other difference is, there are more waveguide crossings in torus topology than in mesh as a result of the loop links. In torus networks, when transmitting optical data along the same routing path as that in meshes of the same network sizes, it experiences more insertion losses as well as crosstalk noise from the link crossings. To guarantee enough optical power reaching the receiver in torus-based ONoCs, a common solution is to increase the input optical power, at the expense of more energy consumption for the transmitter. Therefore, when analyzing and evaluating the performance of our approaches in 2D-torus ONoCs, a larger value for E_{ooo}^o is needed.

In this paper, we neglect this part of the extra energy overhead because it can be largely compensated. Our approaches are based on the shortest routing paths. In a $M \times N$ torus-based ONoC with our routing approaches, assume both M and N are odd, it takes $(M + N)/2 - 1$ hops in the

longest routing paths [28] (when both M and N are even, the number of hops in the longest paths is $(M + N)/2$). We set the input optical power for every communication as 1 mW. As there are up to three link crossings for each node in torus topology, for a communication path, the additional input optical power required by torus topology over mesh is lower than $L_c^{3(M+N)/2} - \sum_{i=0}^{3(M+N)/2-1} K_c \cdot L_c^i$ where L_c and K_c are the power loss and crosstalk coefficient per crossing, respectively [29]. This is quite a loose upper bound based on pessimistic assumptions. Quantitative analysis results show that, small additional input power is required in tori compared to meshes even for large-size ONoCs. In addition, optimization techniques for topology layout can largely reduce this overhead. Optimized floorplans for torus topology have been proposed with the minimum number of waveguide crossings [28]. The total number of waveguide crossings in a $M \times N$ torus topology can be reduced from $3MN - 4M - 4N + 8$ down to $MN - 2\max(M, N)$ after optimization (e.g., reduced from 563 down to 195 in a 15×15 torus, with a reduction of over 65%).

The discussion on the proposed approaches being generally applied to other ONoC platforms with different topologies (such as ring-based and high-radix indirect networks) is temporarily beyond the scope of this paper. It needs further extensions. Our routing criterion is closely related to network topology and router structure because both of them would limit the allowable number of switching stages in routing paths. Based on different routing criteria, Function 1 (and thus our routing approaches) need corresponding modifications for these platforms. So does the crosstalk issue. The resultant crosstalk noise in different ONoC topologies are different and extra platform-specific investigations are required.

E. Technique implementation and overhead analysis

Our approaches require centralized control based on the global information of chip thermal distribution, resource occupation, and inter-processor communication requests. It helps to achieve global optimization in both contention avoidance and communication energy efficiency.

Fig. 9: The logical design for the centralized routing controller.

We provide the logical design of our routing control unit in Fig. 9. The communication request(s) are stored in a request buffer³. After fetching the requests from the buffer, the control unit will select an optimized routing path for each request according to our routing approaches. The path information is then stored to a path buffer and wait to be sent to the corresponding communication source nodes. Based on the routing information, inter-processor communications are conducted in source routing manners. Like in [18], the global

³Both of our approaches are designed to process multiple communication requests concurrently, but are also applicable to the cases with a single communication request.

information of chip temperature distribution is stored in a temperature table that will be dynamically updated with chip thermal fluctuations. Similarly, the link occupation information can also be stored in a table (or in special registers as in [30]). To perform routing calculation, three key components are included in the controller. The candidate path unit is responsible for obtaining the thermal-reliable candidate routes for each request by employing the routing criterion proposed in Section IV-C. From these alternatives, the contention detection unit selects the routing paths that do not conflict with the existing paths and do not collide with each other, based on the link occupation information. Among these unconflicted paths, we calculate their communication energy consumption in the energy estimation unit based on the global temperature information, and select the most efficient routing path for every request. Finally, we update the link information based on the selected routing paths. If the required resources of a request are being occupied by other communications, the controller would delay the request and store it back to the request buffer until the resources are available.

The cost of global information collection and broadcast is a general issue in centralized routing management. Similar overhead also exists when employing other centralized routing techniques, where the information of chip thermal distribution, link occupation, and inter-processor communication requests is collected by a control unit and the unit sends the routing paths to the routers. The main overhead incurred by our approaches is their execution time, which is acceptable due to three reasons. First, as communication contention introduces heavy overhead in ONOCs, it is urgent to employ effective routing strategies to minimize conflicts though it may cause more control overhead for making routing decisions. Unlike electronic channels where electronic packets can be buffered in situ, if an optical path is interrupted due to contention, the source node has to send a control packet to teardown the path, and a small acknowledgment packet may be sent from the destination node to the source for guaranteed delivery. After resolving the contention, the optical path is re-established and the large volumes of data will be resent then. Significant delays and energy costs are incurred by these operations and increase exponentially under complex communication patterns and heavy network traffic. To effectively address this problem, our MILP model, as a formal method, can perform systematic exploration in the design space and provide a theoretically guaranteed optimal solution. And the heuristic CAR can obtain a near-optimal solution within polynomial time. Second, our approaches are able to process multiple communication requests concurrently, thus the control overhead involved in per communication would be small due to cost-sharing. Last, additional techniques such as pipelining can further compensate for the overheads incurred by our approaches.

VI. PERFORMANCE EVALUATION

A. Experimental Setup

Our approaches are evaluated by comparing with two recognized techniques, including an efficient contention-aware routing approach (DyXY) [4] and a thermal-sensitive routing

algorithm proposed in [18]. These two techniques achieve excellent performance in contention avoidance or thermal-induced energy overhead reduction.

We consider the 2D-mesh and 2D-torus based ONOCs with sizes ranging from 8×8 to 15×15 as the target platforms. A high-level simulator in MATLAB is built to evaluate our approaches. In the simulator, we use the out-of-order Alpha 21346 core array in 22 nm technology for computation. Vertically on top of the processing layer where the cores are located, ONOCs with 2D-mesh and 2D-torus topologies are employed for inter-processor communication. Considering the DVFS capability of modern processor cores, we assume that each core can operate at four different voltage and frequency levels: 1.250V at 2.4GHz, 1.210V at 2.1GHz, 1.187V at 1.6GHz, and 1.06V at 1.0GHz [15]. Under different operating voltage/frequency levels, we can model the power consumption of processor cores by McPAT v1.0 [31]. Most of power consumed by the processors for task execution would be converted into heat, resulting in uneven chip thermal distribution. We use HotSpot v5.02 [32] to obtain the steady-state chip thermal profiles, based on the processor cores' power traces generated from McPAT. Inter-processor communications are conducted under this chip temperature profile. Based on extensive synthetic communication traces and a set of real-world applications, we simulate the inter-processor communications by employing different routing techniques (including our approaches, the algorithm in [18], and DyXY) for effectiveness comparison. The routing criterion proposed in Section IV-C is applied in the simulations to perform reliable on-chip communications. Gurobi optimization [33] with CVX v2.1 is employed as the MILP solver in the simulations, which is designed to be the fastest, most powerful solver available for MILP problems. Experiments are run on a workstation with Intel(R) Xeon(R) CPU E5-1650 at 3.60 GHz and 32 GB memory.

B. Comparison of communication performance and energy consumption

We first evaluate the routing techniques based on extensive *synthetic traffic patterns* (including Uniform, Bit Complement, Bit Reverse, and Hotspot). All of them are common traffic patterns used to evaluate interconnection networks. Bit Complement and Bit Reverse are two permutation traffic models. Hotspot is an extension of the Uniform pattern but is closer to reality. In this model, one or some of the nodes are chosen as 'hot' nodes and receive messages with a greater probability of h than the uniform nodes that are with an equal probability. We use Hotspot 15% (i.e., set h as 15%) in the experiments. Under each pattern, we conduct 100 groups of experiments for every size of ONOC (including both mesh and torus networks). In every group of experiment, the source, destination and the volume of communications among processor cores are randomly generated and follows the traffic model. To focus on the effectiveness of these techniques in contention reduction, we set a high load rate for the networks in the experiments. Therefore, the size of the generated communication demand increases with the growing size of ONOCs. The voltage and

Fig. 10: Comparison of evaluation results obtained by different approaches under Uniform traffic pattern (The solid curves represent the results on 2D-mesh ONoCs while the dashed lines represent those based on 2D-torus ONoCs).

Fig. 11: Comparison of evaluation results obtained by different approaches under Bit Complement traffic pattern.

Fig. 12: Comparison of evaluation results obtained by different approaches under Bit Reverse traffic pattern.

Fig. 13: Comparison of evaluation results obtained by different approaches under Hotspot traffic pattern.

frequency levels of processor cores are randomly assigned, consequently resulting in a randomly-generated chip thermal distribution. Based on the generated chip thermal variation and communication demand, our approaches, work [18], and DyXY algorithm are separately employed to perform routing control. Evaluation results on average communication latency, network throughput, link utilization, and communication energy consumption are compared. The link utilization is defined as the ratio of busy links to the total links in an ONoC during communication sessions.

Figure 10-13 illustrate the experimental results obtained by the techniques under the synthetic traffic patterns on mesh- and torus-based ONoCs. The solid curves represent the results on meshes while the dashed lines represent those based on tori. We set the results obtained by the MILP in 8×8 mesh-

based ONoC as the baseline and normalize the other results accordingly. The results slightly vary in different traffic models but show a similar trend in optimization gains. Owing to the minimum communication conflicts, the MILP-based approach achieves an average of 24.78% and 17.64% reduction in communication latency and an average of 126.95% and 93.18% improvement in network throughput, respectively, compared to the work [18] and DyXY. The link utilization is increased by 51.79% and 50.99% as well. Results obtained by CAR are close to the optimal results obtained by the MILP approach, with a 7.32% difference on average. A stable performance gain is achieved and is scalable to large-scale ONoCs with heavy traffic. Based on the preserved communication performance in terms of communication latency and network throughput, the link utilization decreases with the increasing network size;

Fig. 14: Comparison of evaluation results obtained by different approaches based on the realistic applications.

consequently, more communication requests are potentially affordable. In addition, compared to work [18] that obtains minimal communication energy consumption, the MILP and the CAR only consume an average of 0.01 pJ/bit and 0.07 pJ/bit more energy, respectively, and they reduce 16.12% energy overhead than the DyXY. Tori achieve better tolerance to communication contention with better path diversity than meshes. It further improves the communication efficiency of ONoCs by about 1% on average while occupies much fewer optical channels. The performance gains obtained from torus topology would be further increased if more boundary routers are involved in on-chip communications.

Work [18] achieves excellent energy efficiency by selecting the routing paths with the minimum communication energy consumption for communication pairs. Significant delays are introduced to wait for the occupied optical channels to become available, thus increasing the communication latency and reducing the network throughput and link utilization. The adaptive DyXY is a contention-aware routing algorithm, which makes decisions by monitoring the optical channel occupancy status in the proximity. By selecting the next hop with more available optical outgoing links, this technique effectively reduces the possibility of communication contention for ONoCs. Therefore, compared to the algorithm in [18], DyXY achieves better communication performance with low latency, high throughput and link utilization, but is more energy-consuming because it does not take the thermal susceptibility of ONoCs into account. However, restricted by the routing criterion applied in ONoCs and using only local information to make decisions for communication pairs individually, the benefits of DyXY in communication performance optimization is diminished. As shown in Fig. 13, under the Hotspot traffic where communication contention is highly acute, it only achieves slightly better performance than work [18] does, in terms of latency, throughput, and link utilization. Compared with them, our approaches implement centralized control based on global information, which facilitates both contention avoidance and energy efficiency optimization.

To evaluate the feasibility of our approaches to real-world applications, we further conduct experiments based on a set of *realistic benchmarks*, including *autocor*, *audiobeam*, *tde_pp*, *fmradio*, *filterbank* and *beamformer* in StreamIt [34] and *IIR filter* and *4_stagelattice* in DSP-stone benchmark [35]. The data communication traces are generated by the task mapping algorithm in [35] and the chip thermal distributions are generated by McPAT and HotSpot simulations, similar to the experiments that use synthetic communication traces.

Figure 14 illustrates the results obtained by these approaches based on the realistic applications in mesh- and torus-based ONoCs. On mesh-based ONoCs, compared to the work [18] and DyXY, our MILP model achieves an average of 22.86% and 13.84% reduction in communication latency, an average of 123.47% and 85.75% improvement in network throughput, and an average of 50.93% and 31.63% improvement in link utilization, respectively. The heuristic CAR only has an average of 7.20% performance difference compared to the MILP. Regarding the communication energy, our approaches consume an average of < 0.05 pJ/bit more energy than work [18] but improve the energy efficiency by 8.10% compared with DyXY. The torus topology further improve the communication efficiency than meshes. The experimental results based on the realistic benchmarks are consistent with those based on the synthetic communication traces.

TABLE IV: The performance results of the MILP model.

Performance metrics	ONoC topology	Synthetic traffic patterns				Realistic benchmarks
		Uniform	Bit Comp.	Bit Rev.	Hotspot	
Latency (s)	Mesh	4.53E-8	4.61E-8	4.82E-8	4.56E-8	4.60E-8
	Torus	4.52E-8	4.57E-8	4.80E-8	4.55E-8	4.58E-8
Throughput (pkt/s)	Mesh	4.57E8	3.96E8	2.62E8	3.52E8	4.16E8
	Torus	4.58E8	3.98E8	2.66E8	3.54E8	4.18E8
Utilization	Mesh	48.30%	22.06%	23.54%	19.80%	26.25%
	Torus	43.44%	19.70%	21.14%	17.64%	22.64%
Energy (pJ/bit)	Mesh	2.65	2.68	2.81	2.51	2.74
	Torus	2.62	2.64	2.66	2.50	2.73

Finally, we record the experimental results of the MILP in Table IV. These optimal results are served as the baseline solutions in the simulations. Using the MILP for effective contention circumvention, the average communication latency is only about 46.24 ns with a high throughput of about 3.8×10^8 packet-per-second. The time spent for E-O and O-E conversion is a major contributor to the total communication latency. As shown in Table II, we conservatively set the O-E and E-O interface data rate as 12.5 Gbit/s in the simulations, and thus it costs 40.96 ns for optical data modulation when transferring a 512-bit payload packet. By employing advanced E-O/O-E interfaces with a higher data rate of 25 and even 50 Gbit/s, the communication latency could be further reduced exponentially. Compared to the non-uniform patterns including Bit Complement, Bit Reverse and Hotspot, our approaches achieve better throughput and link utilization under uniform traffic. In addition, with fine-grained consideration of the thermal effects in ONoCs, it costs only < 3 pJ/bit for data transmission in the presence of on-chip thermal variations.

Fig. 15: The proportion of I/L/Z-shaped routing paths obtained by our approaches under the synthetic traffic patterns and the realistic applications (The cross-hatched bars represent the results based on 2D-torus ONoCs).

Compared to the meshes, an improvement in communication efficiency is obtained on torus-based ONoCs.

C. Comparison of thermal reliability

It has been theoretically verified that all of the routing paths determined by our approaches are reliable, by applying the routing criterion proposed in Section IV-C. To further experimentally show the effectiveness of the criterion, we calculate the proportion of different types of routing paths (i.e., the I/L/Z-shaped routes) obtained by our approaches on various sizes of ONoCs under the synthetic traffic patterns and the realistic applications. As shown in Fig. 15, in most cases, I- and L-shaped routing paths are already able to avoid communication contention efficiently, not to mention Z-shaped routes. By virtue of the loop links in torus topology, an average of 1.6% Z-shaped routing paths is avoided, which have been highlighted using red circles in Fig. 15. It helps to improve the ONoC communication efficiency on the basis of guaranteed reliability. The practicality of our routing criterion is validated. It is true that we can relax the restriction on the switching stage numbers when making routing decisions, but it would increase communication latency and energy consumption correspondingly.

D. Comparison of technique running time

(a) Simulation time for an entire communication demand. (b) Avg. simulation time for each communication request.

Fig. 16: The running time of the proposed approaches.

We experimentally compare the running efficiencies of our approaches based on high-level simulations in MATLAB. The simulation time of the MILP and the CAR are illustrated in Fig. 16. Although the running time of our approaches in the high-level simulators is not the actual execution time in real systems, it reflects their technique efficiencies to some extent and can be used for comparison.

Our approaches are designed to process multiple concurrent communication requests at a time. We provide the average

running time required for the entire communication demand as well as that required for a single request in Fig. 16(a) and (b), respectively. There are three observations. First, as a formal method, the time complexity of the MILP model is exponentially increased with the growth of the search space. Nevertheless, it is worthwhile taking efforts to perform systematic exploration in the search space for an optimal solution, not only due to the huge performance gain that could be expected, but also because it is running at design-time only once and then will probably be applied to millions of platforms for run-time implementation. Compared with the MILP, the polynomial-time algorithm CAR is able to obtain a near-optimal solution with negligible running time, especially in large-scale ONoCs. Second, as our approaches are able to process multiple requests concurrently, the execution time cost involved in a single communication is small through cost-sharing, shown in Fig. 16(b). Third, compared with meshes, the loop links in tori provide better path diversity for contention circumvention but also increase the design space. It takes more time to find optimized solutions in 2D-torus ONoCs, especially when employing the MILP model.

VII. CONCLUSION

To co-optimize the thermal reliability and communication efficiency, in this paper, we novelly proposed a routing criterion and two routing approaches (including a MILP model and a heuristic CAR algorithm) for ONoCs. The network-level routing criterion, combined with device-level thermal tuning, achieves thermal-reliable ONoCs. Based on it, the communication performance and energy efficiency are optimized through contention-circumvented and thermal-aware routing. Extensive evaluation results verified the applicability and effectiveness of our approaches, compared with state-of-the-art techniques.

ACKNOWLEDGMENT

This work is partially supported by the the National Natural Science Foundation of China (NSFC No. 61772094), and partially supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE2019-T2-1-071) and Tier 1 (MOE2019-T1-001-072), and Nanyang Technological University, Singapore, under its NAP (M4082282) and SUG (M4082087).

REFERENCES

- [1] S. Werner *et al.*, "A survey on optical network-on-chip architectures," *ACM Computing Surveys*, vol. 50, no. 6, p. 89, 2017.

- [2] S. Werner, J. Navaridas, and M. Lujan, "Amon: An advanced mesh-like optical noc," in *HOTI*, 2015, pp. 52–59.
- [3] X. Wang, H. Gu, Y. Yang, K. Wang, and Q. Hao, "RPNoc: A ring-based packet-switched optical network-on-chip," *IEEE Photonics Technology Letters*, vol. 27, no. 4, pp. 423–426, 2015.
- [4] M. Li, Q. A. Zeng, and W. B. Jone, "DyXY - a proximity congestion-aware deadlock-free dynamic routing method for network on chip," in *DAC*, 2006, pp. 849–852.
- [5] C. Li *et al.*, "CCAS: Contention and congestion aware switch allocation for network-on-chips," in *ICCD*, 2016, pp. 444–447.
- [6] H. Gu *et al.*, "A low-power low-cost optical router for optical networks-on-chip in multiprocessor systems-on-chip," in *ISVLSI*, 2009, pp. 19–24.
- [7] Y. Ye, J. Xu, X. Wu, W. Zhang, X. Wang, M. Nikdast, Z. Wang, and W. Liu, "System-level modeling and analysis of thermal effects in optical networks-on-chip," *IEEE TVLSI*, vol. 21, no. 2, pp. 292–305, 2013.
- [8] M. Li *et al.*, "Lightweight thermal monitoring in optical networks-on-chip via router reuse," in *DATE*, 2020.
- [9] K. Padmaraju and K. Bergman, "Resolving the thermal challenges for silicon microring resonator devices," *Nanophotonics*, vol. 3, no. 4–5, pp. 269–281, 2014.
- [10] W. Liu *et al.*, "Thermal sensing using micro-ring resonators in optical network-on-chip," in *DATE*, 2019, pp. 1611–1614.
- [11] W. J. Dally and B. P. Towles, *Principles and practices of interconnection networks*. Elsevier, 2004.
- [12] S. Spiga *et al.*, "Single-mode high-speed 1.5- μm vcsels," *IEEE JLT*, vol. 35, no. 4, pp. 727–733, 2016.
- [13] S. Saeedi, S. Menezes, G. Pares, and A. Emami, "A 25 gb/s 3d-integrated cmos/silicon-photonics receiver for low-power high-sensitivity optical communication," *IEEE JLT*, vol. 34, no. 12, pp. 2924–2933, 2015.
- [14] S. J. Hollis *et al.*, "Exploiting emergence in on-chip interconnects," *IEEE TC*, vol. 63, no. 3, pp. 570–582, 2014.
- [15] M. Li, W. Liu, L. Yang, P. Chen, and C. Chen, "Chip temperature optimization for dark silicon many-core systems," *IEEE TCAD*, 2017.
- [16] J. Bashir, E. Peter, and S. R. Sarangi, "A survey of on-chip optical interconnects," *ACM Computing Surveys*, vol. 51, no. 6, p. 115, 2019.
- [17] Z. Li *et al.*, "Aurora: A cross-layer solution for thermally resilient photonic network-on-chip," *IEEE TVLSI*, vol. 23, no. 1, pp. 170–183, 2015.
- [18] K. Yao *et al.*, "Thermal-sensitive design and power optimization for a 3d torus-based optical noc," in *ICCAD*, 2017, pp. 827–834.
- [19] M. Li, W. Liu, N. Guan, Y. Xie, and Y. Ye, "Hardware-software collaborative thermal sensing in optical network-on-chip-based manycore systems," *ACM TECS*, vol. 18, no. 6, p. 118, 2019.
- [20] W. Liu *et al.*, "Quantitative modeling of thermo-optic effects in optical networks-on-chip," in *GLSVLSI*, 2017, pp. 263–268.
- [21] M. Mohamed, Z. Li, X. Chen, L. Shang, and A. R. Mickelson, "Reliability-aware design flow for silicon photonics on-chip interconnect," *IEEE TVLSI*, vol. 22, no. 8, pp. 1763–1776, 2014.
- [22] J.-F. Seurin *et al.*, "Progress in high-power high-efficiency vcsel arrays," in *Vertical-Cavity Surface-Emitting Lasers XIII*, vol. 7229. SPIE, 2009, p. 722903.
- [23] "Lumerical inc." <https://www.lumerical.com/>, [Online].
- [24] L. H. Duong *et al.*, "Crosstalk noise reduction through adaptive power control in inter/intra-chip optical networks," *IEEE TCAD*, vol. 38, no. 1, pp. 43–56, 2018.
- [25] G. Masini, G. Capellini, J. Witzens, and C. Gunn, "A 1550nm, 10gbps monolithic optical receiver in 130nm cmos with integrated ge waveguide photodetector," in *IEEE Inter. Conf. on Group IV Photonics*, 2007.
- [26] H. Gu *et al.*, "A low-power fat tree-based optical network-on-chip for multiprocessor system-on-chip," in *DATE*, 2009, pp. 3–8.
- [27] Y. Ye, J. Xu *et al.*, "3-d mesh-based optical network-on-chip for multiprocessor system-on-chip," *IEEE TCAD*, vol. 32, no. 4, pp. 584–596, 2013.
- [28] K. Feng, Y. Ye, and J. Xu, "A formal study on topology and floorplan characteristics of mesh and torus-based optical networks-on-chip," *Microprocessors and Microsystems*, vol. 37, no. 8, pp. 941–952, 2013.
- [29] M. Nikdast, J. Xu *et al.*, "Crosstalk noise in wdm-based optical networks-on-chip: A formal study and comparison," *IEEE TVLSI*, vol. 23, no. 11, pp. 2552–2565, 2014.
- [30] X. Wu *et al.*, "Union: A unified inter/intrachip optical network for chip multiprocessors," *IEEE TVLSI*, vol. 22, no. 5, pp. 1082–1095, 2013.
- [31] S. Li *et al.*, "Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures," in *MICRO*, 2009, pp. 469–480.
- [32] W. Huang *et al.*, "Hotspot: A compact thermal modeling methodology for early-stage vlsi design," *IEEE TVLSI*, vol. 14, no. 5, pp. 501–513, 2006.

- [33] "Gurobi optimization," <http://www.gurobi.com/>, [Online].
- [34] W. Liu *et al.*, "Energy-efficient application mapping and scheduling for lifetime guaranteed mpsoes," *IEEE TCAD*, vol. 38, no. 1, pp. 1–14, 2019.
- [35] L. Yang, W. Liu, P. Chen, N. Guan, and M. Li, "Task mapping on smart noc: Contention matters, not the distance," in *DAC*, 2017, p. 88.

Mengquan Li received the B.E. degree from Chongqing University in 2015 and is currently pursuing the Ph.D. degree from the College of Computer Science at Chongqing University, China. She is now a visiting scholar in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include optical network-on-chip, on-chip temperature modeling and optimization, and nanophotonic accelerator design for deep learning.

Dr. Weichen Liu (S07-M11) is a Nanyang Assistant Professor at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He received the Ph.D. degree from the Hong Kong University of Science and Technology, Hong Kong and the BEng and MEng degrees from Harbin Institute of Technology, China. His research interests include embedded and real-time systems, multiprocessor systems, network-on-chip and machine learning acceleration.

Dr. Luan H.K. Duong (S14-M17) is now a postdoctoral research fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He received the Ph.D. degree from the Hong Kong University of Science and Technology, Hong Kong. His research interests include optical interconnect networks, optical computing, embedded systems, and many-core systems.

Peng Chen is currently pursuing the Ph.D. degree from the College of Computer Science at Chongqing University, China. He is now a visiting scholar in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His current research interests include embedded and real-time systems, multicore scheduling.

Dr. Lei Yang is now a post-doctoral research associate in the Department of Computer Science and Engineering, University of Notre Dame, USA. She received the Ph.D. degree from Chongqing University, China. Her current research interests include Network-on-chip optimization, multi-core architecture design and high-performance parallel computing in Nonvolatile Memory (NVM) based NoCs.

Dr. Chunhua Xiao is currently an associate professor at the College of Computer Science, Chongqing University, China. She received the Ph.D. degree from the Beijing University of Technology, China. Her research interests include MPSoCs, hardware and software co-design, and embedded systems.