A Table-Free Approximate *Q*-Learning-Based Thermal-Aware Adaptive Routing for Optical NoCs

Wenfei Zhang and Yaoyao Ye, *Member, IEEE*

***Abstract*—Optical networks-on-chips (NoCs) based on silicon photonics have been proposed as an emerging communication architecture for many-core chip multiprocessors. However, the thermal sensitivity of silicon photonics is one of the major challenges. *Q*-learning-based adaptive routing has been proposed in related work to mitigate the thermal issue. However, table overhead of the traditional table-based *Q*-routing would scale up quickly with the increase of network size. In this article, we propose a table-free approximate *Q*-learning-based thermal-aware adaptive routing to find optimal low-loss paths in the presence of on-chip temperature variations. The simulation results show that the proposed table-free approximate *Q*-learning-based adaptive routing can converge faster and it can achieve similar optimization effect as compared to the best optimization effect of the traditional table-based *Q*-routing. The performance gap between the proposed approximation method and the traditional table-based *Q*-routing expands when the network size increases.**

***Index Terms*—Function approximation, optical network-on-chip, *Q*-routing, temperature sensitivity.**

I. Introduction

With the development of nanoscale silicon photonic technologies, optical networks-on-chip (NoC) have been widely proposed as a new generation of on-chip communication architecture for future chip multiprocessors [1], [2]. In the last decade, many optical NoC architectures have been proposed in literature, where silicon microresonators (MRs) are widely used as a wavelength-selective optical switching element to perform the switching function. However, because of thermo-optic effects, silicon MRs suffer from temperature-dependent wavelength shifts which would result in thermal-induced optical power loss in switching [3].

With the shrinkage of the chip area, the thermal issue will become more serious for optical NoCs. Several device-level thermal compensation techniques have been proposed to reduce temperature-dependent wavelength shift of silicon MRs, including thermal tuning by local microheaters, and low-temperature-dependence MRs with polymer materials as upper cladding. However, thermal tuning is relatively slow and power inefficient, while there are still compatibility issues when fabricating a thermal MRs with CMOS technology. Some efforts have also been made to overcome the thermal challenges in optical NoCs from network-level perspectives [3]–[6].

Furthermore, adaptive routing algorithms have been proposed in optical NoC designs to dynamically make routing decisions according to run-time on-chip temperature variations. A thermal-sensitive source routing was proposed in [7], which uses global temperature information to find optimal paths with the minimum optical power loss. Li *et al.* [5] proposed to reroute packets through cooler regions to destinations. In [8], an adaptive fault-tolerant routing algorithm was proposed to dynamically route packets in the presence of on-chip temperature variations. Considering an optical NoC in the presence of on-chip temperature variations, how to select a best path with the objective to optimize the optical power loss can be formulated as an optimization problem. Ye *et al.* [9] proposed a *Q*-learning-based thermal-aware adaptive routing which is able to converge quickly to optimal selections. However, such a table-based *Q*-routing suffers from bad scalability, in which the *Q*-table required in each node would scale up quickly with the increase of network size.

In this article, we propose a table-free approximate *Q*-learning-based thermal-aware adaptive routing algorithm to optimize thermal-induced optical power loss in optical NoCs. By taking on-chip temperature gradient into account, the proposed table-free approximate *Q*-learning-based routing can predict the best alternative paths with the optimal thermal-induced optical power loss. With a linear function approximation method, the proposed approximate *Q*-learning-based routing can achieve very close optimization effect as compared to the traditional table-based *Q*-routing. To our knowledge, this is the first work to apply approximate *Q*-learning in the NoC domain. This article is extended from our previous conference paper [10]. Additional contributions include comparisons for two types of action-value approximation methods in Section IV-B; explorations for effect of different learning rates in Section IV-C; experiments for the convergence under dynamically changing temperature distributions in Section IV-D; discussions for the scalability in Section IV-G.

The remainder of this article is organized as follows. In Section II, we give a brief introduction about the basics of *Q*-learning. In Section III, we present the proposed approximate thermal-aware *Q*-routing in details. In Section IV, we present simulation results and comparisons. Finally, Section V draws the conclusion of this article.

II. Preliminary

Q-routing is an adaptive distributed routing originating from *Q*-learning model, which is a reinforcement learning model. *Q*-learning requires a table (named *Q*-table) to maintain *Q*-values which indicate the quality of an action the agent selects in certain state. Suppose that the agent is in state S and it chooses action A using policy derived from the *Q*-table. Then the agent takes action A, gets a reward, and reaches a new state S'. After that, the agent chooses the action of maximum *Q*-value in state S' but not to take this action. Instead, the agent uses the *Q*-value and the reward as a new estimate of state S to update the *Q*-table. In this way, the agent will get a policy according to which the agent can do the best in any sequence and maximize the total reward. *Q*-routing makes every node a *Q*-learning model and passes *Q*-values to each other.

Q-routing has been utilized in the traditional electronic NoC domain for latency optimization under congestion, as well as for QoS enhancement [11]. In general, each node in the network learns the network conditions by receiving learning packets from

Manuscript received December 2, 2019; revised February 23, 2020; accepted April 7, 2020. Date of publication April 14, 2020; date of current version January 11, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61602298. This article was recommended by Associate Editor L. P. Carloni. *(Corresponding author: Yaoyao Ye.)*

The authors are with the Department of Micro/Nano Electronics, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yeyaoyao@sjtu.edu.cn).

Digital Object Identifier 10.1109/TCAD.2020.2987775

0278-0070 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: National Institute of Technology Karnataka Surathkal. Downloaded on September 21,2024 at 11:21:18 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Approximate Q-Learning-Based Thermal-Aware Adaptive Routing Algorithm

```
Q_{i,out}(d, i, h) is abbreviated as Q_i.
Q_{i-1,out}(d, i, h) is abbreviated as Q_{i-1}.
initialize all coefficients of all nodes
send packet from source node s to destination node d
for node i in all nodes do
    receive setup packet from last node i-1
    if i != s then
        calculate l
        if i == d then
            receive optical loss to the destination l_processor
            newQ_{i-1} = l + l_processor
        else
            calculate min_out Q_{i,out}(d, i, h)
            newQ_{i-1} = l + min_out Q_{i,out}(d, i, h)
        end if
        send newQ_{i-1} in learning packet back to node i-1
    end if
    if i != d then
        select output port arg min_out Q_{i,out}(d, i, h)
        or randomly select output port with ε probability
        send setup packet to next node i+1
        receive learning packet from node i+1
        for θ_j and scaled feature f_j in node i do
            θ_j = θ_j + α(newQ_i − Q_i)f_j
        end for
    end if
end for
```

neighboring nodes. In order to reduce the table overhead in Q-routing, in [12], clustering-based regional Q-routings were proposed which only requires to maintain cluster-level tables. However, Q-tables are still required and there are still scalability issues when core counts get larger.

III. Approximate Q-Learning-Based Thermal-Aware Adaptive Routing With Linear Function Approximation

We assume a mesh-based optical NoC works in a circuit switching mechanism. Before each payload transmission in optical domain, a setup packet is first routed in the electrical control network to reserve an optical path for the payload transmission. In order to take advantage of the optimization effect of Q-routing but without the table overhead, in this section, we propose a table-free approximate Q-learning-based adaptive routing by using the linear function approximation. Algorithm 1 presents the proposed routing algorithm from the perspective of any single node.

In the following, we use an example to explain the procedure of path selection. We summarize notations used in this section in Table I. For a packet to be sent from the source node s to the destination node d, assume that its setup packet is currently at node x. Suppose that $A_x(d)$ represents the set of feasible output ports of node x found along the shortest paths, where A represents the action space. We use the resending mechanism to avoid deadlock in the network [13]. The proposed approximate Q-learning-based routing can also be implemented based on deadlock-free routings (e.g., odd-even turn model). We define $Q_{x,\text{out}}$ as the Q-value calculated by the linear approximation model of Q-function [14] representing the estimated thermal-induced optical power loss from input port of node x through any feasible output port out $\in A_x(d)$ to the destination d. Actually, many models can approximate the Q-value but the linear method has convergence guarantees [15]. We choose the linear function model also for reasons of efficiency and simplicity of implementation.

TABLE I
NOTATIONS

Notation	Explanation
d	destination id (feature)
i	input port id of the current node (feature)
h	hop number of the path (feature)
f_j	the j_{th} scaled feature
$A_x(d)$	the set of feasible output ports in node x
A	action space
$Q_{x,out}(d,i,h)$	Q-value calculated by Q-function in node x
$newQ_{x,out}(d,i,h)$	new estimate of Q-value
θ_j	the j_{th} coefficient of Q-function
out_i	output port to the node i
l	optical power loss from input port of current node to input port of the next node
ϵ	probability
γ	discount factor

We select four features to represent states, including constant term 1, destination node d, input port of local node i, and the hop number h from local node to destination node. The output port *out* represents the action. In Q-function, coefficients can get too large as updating. We use the feature scaling method to avoid this problem. In (1), we define f_j as the jth scaled feature, and θ_j as the jth coefficient of the Q-function. Here, all features are standardized by dividing by their range

$$Q_{x,\text{out}}(d,i,h) = \theta_0 \cdot f_0 + \theta_1 \cdot f_1 + \theta_2 \cdot f_2 + \theta_3 \cdot f_3. \tag{1}$$

Node x selects an output port $\text{out}_y \in A_x(d)$ which is with the minimum estimated thermal-induced optical power loss, or randomly selects an output port from set $A_x(d)$ with a very small probability ϵ (even 0). Node x then forwards the setup packet to the next node y, and the setup packet carries l (2) representing the real loss from input port of node x to input port of node y. l is calculated according to the models proposed in [3], as the addition of loss inserted by every component. The first term of (2) is the temperature-dependent insertion loss of active MRs in the switching, where N_1 is the number of active MRs. N_1 is zero for switchings along the same dimension in the mesh network, and it is one for switchings in a turn to another dimension. ΔT is temperature variation, ρ_{MR} is the MR temperature-dependent wavelength shift coefficient, 2δ is the 3-dB bandwidth, κ^2 is the fraction of power coupling between the ring and the waveguide, and κ_p^2 is the power loss per round-trip of the ring. The second term is the insertion loss of passive MRs, where N_2 is the number of passive MRs, and we assume the insertion loss of each passive MR is L_{passive}. The third term is optical waveguide crossing loss, where N_c is the number of waveguide crossing, and L_c is the insertion loss of each waveguide crossing. The fourth term is the waveguide propagation loss, where L_w is the waveguide length (in *mm*) from the input port of node x to input port of node y, and L_p is the waveguide propagation loss per *mm*

$$l = N_1 \cdot 10\log\left(\left(\frac{2\kappa^2+\kappa_p^2}{2\kappa^2}\right)^2 \cdot \frac{(\rho_{\text{MR}}\cdot\Delta T)^2+\delta^2}{\delta^2}\right) + N_2 \cdot L_{\text{passive}} + N_c \cdot L_c + L_w \cdot L_p. \tag{2}$$

Assume that output port out_z meets the requirement of the smallest $Q_{y,\text{out}}(d,i,h)$, out $\in A_y(d)$. Node y sends a learning packet with a new Q-value $\text{new}Q_{x,\text{out}}(d,i,h)$ back to its previous node x, which indicates the new minimum estimated thermal-induced optical power loss from input port of node x to the node d. As shown in (3),

Authorized licensed use limited to: National Institute of Technology Karnataka Surathkal. Downloaded on September 21,2024 at 11:21:18 UTC from IEEE Xplore. Restrictions apply.

new$Q_{x,\text{out}}(d, i, h)$ is equal to the sum of $\gamma Q_{y,\text{out}_z}(d, i, h)$ and l. Since routing is an episodic task [15], the discount factor γ will be one

$$\text{new}Q_{x,\text{out}}(d, i, h) = l + \gamma \cdot Q_{y,\text{out}_z}(d, i, h). \quad (3)$$

When the node x receives the new estimated Q-value new$Q_{x,\text{out}}(d, i, h)$ from the node y, it will update every local coefficient θ_j of feature f_j according to (4), where α is the learning rate. After sending the learning packet, node y would act like the previous node x

$$\theta_j = \theta_j + \alpha \cdot \left(\text{new}Q_{x,\text{out}}(d, i, h) - Q_{x,\text{out}}(d, i, h)\right) \cdot f_j. \quad (4)$$

IV. Simulation Results and Comparisons

A. Simulation Setup

We develop a SystemC-based cycle-accurate optical NoC simulator for a mesh-based optical NoC implementing the proposed approximate Q-learning-based thermal-aware routing. We use Hotspot [16] as the on-chip temperature simulator and McPAT [17] as the power simulator. We assume the processor core model is ARM Cortex-A9, and each core works at 1.25 GHz clock frequency at 45 nm technology. We assume the 8×8 mesh-based chip multiprocessor occupies 10×10 mm^2. In the following evaluation part, we use an 8×8 mesh-based optical NoC as a case study and conduct simulations under five typical temperature distributions, including center block, corner block, narrow strait, winding path, and side block [5]. A set of synthetic traffic patterns and real applications are used in the simulation as traffic loads. For synthetic traffic patterns, we assume the injection rate is 0.1, and the packet size is 512 B. For an optical signal transmitting in the mesh network, the thermal-induced optical power loss per hop is evaluated according to the model in (2). We assume the initial resonant wavelength of MRs at room temperature is 1550 nm, the 3-dB bandwidth is 0.62 nm, and the temperature-dependent wavelength shift is 0.06 nm/$^{\text{o}}$C. We assume the insertion loss of an active MR is 0.5 dB at peak resonance. We assume the waveguide crossing loss is 0.12 dB per crossing, and the waveguide propagation loss is 0.17 dB/mm. The average optical power loss is defined as the average loss of all transmitted packets over the entire network. The proposed routing approach can be utilized together with device-level compensation techniques such as thermal tuning, especially for scenarios with large temperature variations. To focus on the optimization effect of the proposed routing approach, we assume no thermal tuning is used in the following evaluation part.

B. Two Types of Action-Value Function Approximation

In the previous section, the output port *out* (the action feature) is treated, especially to store four groups of four coefficients of features. Another possible solution is to consider the output port *out* as the fifth feature, with one group of five coefficients in the approximation. Fig. 1 shows that the approximation with four groups of four coefficients performs better than another type under the corner block temperature distribution. More simulation results under other temperature distributions also demonstrate this.

C. Effect of Different Learning Rates

To determine the learning rate α, we have conducted comprehensive simulations for a wide range of learning rates under the five typical temperature distributions. Fig. 2 shows one case study for the average optical power loss over the entire network under uniform traffic with the corner block temperature distribution. More case studies under other temperature distributions confirm that 0.01 is a universal value which can result in a good optimization effect for all the case studies we conducted.

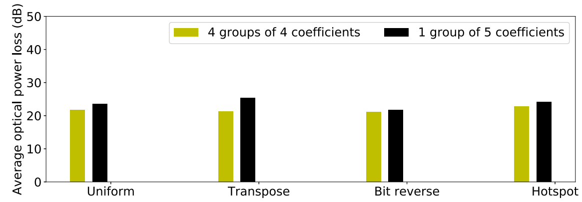

Fig. 1. Comparisons for effect of two approximation.

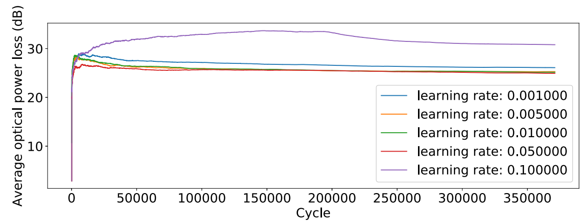

Fig. 2. Comparison for effect of different learning rates.

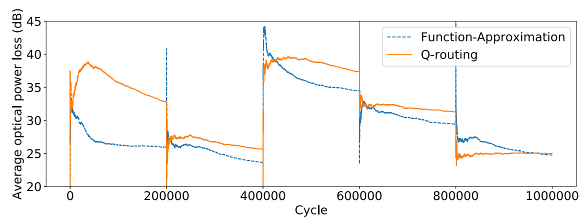

Fig. 3. Convergence under dynamic temperature distributions.

D. Comparisons for Convergence Speed

Fig. 3 shows the convergence of the proposed table-free approximate Q-routing is faster than the traditional table-based method under dynamically changing temperature distributions. In Fig. 3, we assume the initial temperature distribution is set as center block, then it updates to corner block, narrow strait, winding path, and side block in a fixed time interval.

E. Comparisons for Optical Power Loss

Fig. 4(a) shows the comparison of the average thermal-induced optical power loss under the center block temperature distribution. On average of the four synthetic traffic patterns, the proposed routing reduces the average optical power loss by 28.94%, 36.19%, and 30.81%, respectively, as compared to the traditional negative-first routing, the odd-even routing, and the west-first routing. As compared to the traditional table-based Q-routing, the proposed routing mechanism sacrifices 2.03%–7.01% of optimization effect. Fig. 4(b)–(e) shows the same trend under other temperature distributions. In some case, e.g., for the bit reverse traffic under the corner block temperature distribution, the proposed approximate Q-learning-based thermal-aware adaptive routing could outperform the traditional table-based Q-routing because of its faster convergence and less exploration of bad situations.

To further verify the proposed routing, we also conducted comprehensive simulations under a set of real applications. Under the center block temperature distribution, the proposed routing sacrifices 2.47%–6.97% of optimization effect as compared to the traditional table-based Q-routing. Meanwhile, it reduces the average optical power loss by 13.16%, 18.99%, and 13.13%, respectively, as compared to the negative-first, the odd-even, and the west-first routing. The simulations under other temperature distributions show similar results.

Authorized licensed use limited to: National Institute of Technology Karnataka Surathkal. Downloaded on September 21,2024 at 11:21:18 UTC from IEEE Xplore. Restrictions apply.

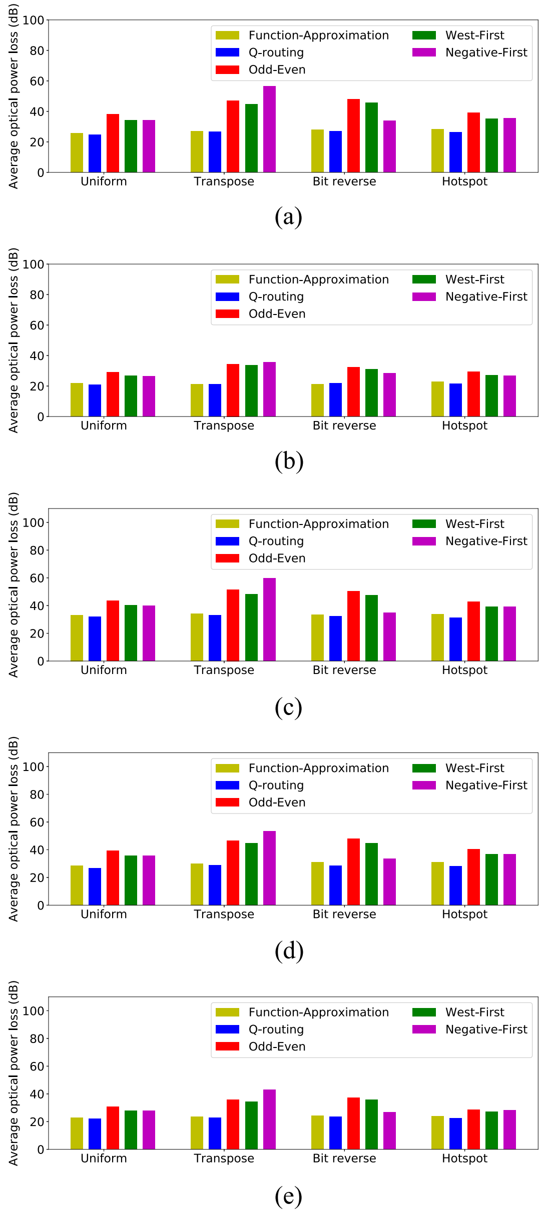

Fig. 4. Average optical power loss under synthetic traffic. (a) With center block temperature distribution. (b) With corner block temperature distribution. (c) With narrow strait temperature distribution. (d) With winding path temperature distribution. (e) With side block temperature distribution.

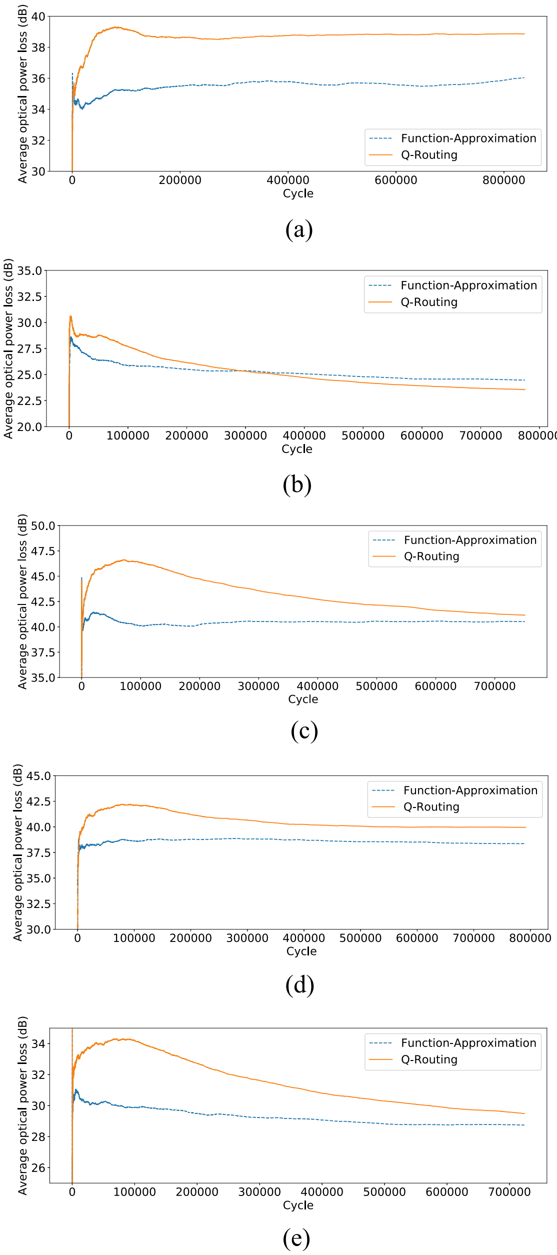

Fig. 5. Comparisons for optimization effect in a 16×16 mesh. (a) With center block temperature distribution. (b) With corner block temperature distribution. (c) With narrow strait temperature distribution. (d) With winding path temperature distribution. (e) With side block temperature distribution.

F. Comparisons for Network Performance

In this section, we study the overhead of the proposed routing on network performance. For uniform traffic, the average network throughput of the proposed network varies from 204.11 Gb/s to 205.47 Gb/s with different temperature distributions, which is 0.1%–0.15% lower than the other four baseline networks. For bit reverse traffic, the network throughput of the proposed network varies from 165.83 Gb/s to 179.56 Gb/s. It is about $2.06\%-2.43\%$ lower than the four baseline networks. For hotspot traffic, the network throughput of the proposed network varies from 195.12 Gb/s to 201.85 Gb/s with the five temperature distributions. It is $0.67\%-1.01\%$ lower than the negative-first routing, west-first routing, and the traditional table-based Q-routing, while it is 2.79% higher than the baseline network with Odd-Even routing. For transpose traffic, the network throughput of the proposed network varies from 154.94 Gb/s to 179.42 Gb/s. It is $1.83\%-3.45\%$ lower than the baseline networks. On average of the three real applications, the proposed design achieves almost the same performance with the baseline networks.

G. Scalability

The traditional table-based Q-routing suffers from a bad scalability due to the large table cost. For an $N \times N$ mesh-based optical NoC under uniform traffic, each node in the network needs to keep a table of $6(N-1)^2$ entries on average. When the network size increases from 8×8 to 16×16, the total number of entries required in the network increases from 18 816 to 345 600. In comparison, the proposed routing method only needs to maintain 16 coefficients in each node. To further demonstrate the scalability, Fig. 5 shows the optimization effect of the proposed routing in a 16×16 mesh-based

Authorized licensed use limited to: National Institute of Technology Karnataka Surathkal. Downloaded on September 21,2024 at 11:21:18 UTC from IEEE Xplore. Restrictions apply.

optical NoC. The traditional Q-routing becomes quite difficult to achieve the optimal state, while the proposed method can quickly reach a suboptimal state.

V. Conclusion

To eliminate the thermal issue of optical NoCs in the presence of on-chip temperature variations, we proposed a table-free approximate Q-learning-based thermal-aware adaptive routing for reducing thermal-induced optical power loss in optical NoCs. As compared to the traditional table-based Q-routing, the main advantage of the proposed routing is that it can achieve very close optimization effect with faster convergence speed but does not need to keep a big table in each node of the network.

References

[1] A. Shacham, K. Bergman, and L. Carloni, "On the design of a photonic network-on-chip," in *Proc. 1st Int. Symp. Netw. Chip*, Princeton, NJ, USA, 2007, pp. 53–64.

[2] T. Alexoudi *et al.*, "Optics in computing: From photonic network-on-chip to chip-to-chip interconnects and disintegrated architectures," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 363–379, Jan. 15, 2019.

[3] Y. Ye *et al.*, "System-level modeling and analysis of thermal effects in optical networks-on-chip," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 21, no. 2, pp. 292–305, Feb. 2013.

[4] S. V. R. Chittamuru, I. G. Thakkar, and S. Pasricha, "LIBRA: Thermal and process variation aware reliability management in photonic networks-on-chip," *IEEE Trans. Multi Scale Comput. Syst.*, vol. 4, no. 4, pp. 758–772, Oct.–Dec. 2018.

[5] Z. Li, A. Qouneh, M. Joshi, W. Zhang, X. Fu, and T. Li, "Aurora: A cross-layer solution for thermally resilient photonic network-on-chip," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 1, pp. 170–183, Jan. 2015.

[6] J. L. Abellán *et al.*, "Adaptive tuning of photonic devices in a photonic NoC through dynamic workload allocation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 5, pp. 801–814, May 2017.

[7] K. Yao, Y. Ye, S. Pasricha, and J. Xu, "Thermal-sensitive design and power optimization for a 3D torus-based optical NoC," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Irvine, CA, USA, Nov. 2017, pp. 827–834.

[8] M. Yang and P. Ampadu, "Thermal-aware adaptive fault-tolerant routing for hybrid photonic-electronic NoC," in *Proc. 9th Int. Workshop Netw. Chip Architect.*, New York, NY, USA, 2016, pp. 33–38.

[9] Y. Ye, W. Zhang, and W. Liu, "Thermal-aware design and simulation approach for optical NoCs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, early access, Aug. 14, 2019, doi: 10.1109/TCAD.2019.2935407.

[10] W. Zhang and Y. Ye, "An approximate thermal-aware Q-routing for optical NoCs," in *Proc. IEEE/ACM Workshop Photon. Opt. Technol. Orient. Netw. Inf. Comput. Syst. (PHOTONICS)*, Denver, CO, USA, Nov. 2019, pp. 22–27.

[11] M. Ebrahimi *et al.*, "HARAQ: Congestion-aware learning model for highly adaptive routing algorithm in on-chip networks," in *Proc. IEEE/ACM 6th Int. Symp. Netw. Chip*, Copenhagen, Denmark, May 2012, pp. 19–26.

[12] F. Farahnakian, M. Ebrahimi, M. Daneshtalab, P. Liljeberg, and J. Plosila, "Bi-LCQ: A low-weight clustering-based Q-learning approach for NoCs," *Microprocess. Microsyst.*, vol. 38, no. 1, pp. 64–75, Feb. 2014.

[13] Y. Ye, J. Xu, X. Wu, W. Zhang, and L. Weichen, "A torus-based hierarchical optical-electronic network-on-chip for multiprocessor system-on-chip," *J. Emerg. Technol. Comput. Syst.*, vol. 8, no. 1, pp. 1–26, Feb. 2012.

[14] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," in *Proc. 6th Int. Conf. Adv. Neural Inf. Process. Syst.*, 1994, pp. 671–678.

[15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (Adaptive Computation and Machine Learning Series), 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[16] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan, "HotSpot: A compact thermal modeling methodology for early-stage VLSI design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 5, pp. 501–513, May 2006.

[17] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proc. 42nd Annu. IEEE/ACM Int. Symp. Microarchitect. (MICRO)*, New York, NY, USA, Dec. 2009, pp. 469–480.

Authorized licensed use limited to: National Institute of Technology Karnataka Surathkal. Downloaded on September 21,2024 at 11:21:18 UTC from IEEE Xplore. Restrictions apply.