An Optimization Framework for Routing on Optical Network-on-Chips (ONoCs) from a Networking Perspective

Yawen Chen, Haibo Zhang and Feiyang Liu
Department of Computer Science
The University of Otago, New Zealand
{yawen, haibo, feiyang}@cs.otago.ac.nz

Huaxi Gu
Key Laboratory of ISN
Xidian University, China
hxgu@xidian.edu.cn

Abstract—As increasing numbers of CPU cores are placed on single-processor chips, conventional electronic interconnects for on-chip communications are becoming a bottleneck, due to difficulties in meeting ever-increasing demands on throughput and increasingly undesirable energy consumption. A potential solution is offered by Optical Networks-on-Chips (ONoCs): an emerging communication architecture for new generation multicore systems. Optical interconnection among cores at the chip level can offer ultra-high communication bandwidth, low latency, and high energy efficiency. This paper provides a brief review and sheds some insights on the unique characteristics and constraints of ONoCs for designing efficient routing schemes, especially from the networking perspective. Based on the properties of ONoCs, we propose an optimization framework for routing and wavelength assignment on ONoCs, which can be used to guide the upgrade from un-optimized ONoCs to optimized ONoCs. This paper provides the insight and guidance for high-level routing design in future research.

Keywords—Optical Network-on-Chips; routing and wavelength assignment; interconnection networks; multicore systems; energy efficient computer architecture

I. INTRODUCTION

A. Advancement of Optical Network-on-Chips (ONoCs)

Today’s processor development has moved to multiple cores on a single chip (e.g., 80 cores in Intel Teraflops Chip [1] and 192 cores in CSX700 Processor [2]). As hundreds of interconnected cores will fit within one chip, the ITRS Roadmap [3] has predicted that conventional electrical interconnect on the chips, Network-on-Chips (NoCs), will not be able to meet the high communications demands, thus becoming a primary bottleneck with performance limits and increasingly unacceptable power consumption. Recent advances [4-10] in nanoscale silicon photonics and optical devices have led to the development of Optical Network-on-Chips (ONoCs) [11-14], a silicon-based optical interconnection among cores at the chip level, an attractive candidate to overcome the limitations of conventional electronic interconnects. Recent advances in optical devices substantially improve the feasibility of ONoCs, and leading chip manufacturers (IBM and Intel) are investing heavily in this new technology. In Dec. 2012, IBM announced a disruptive technology, silicon nanophotonics [4], which uses light to transfer large amount of data at very high speed with extremely low power over a thin optical link (waveguide). In Jan. 2013, Intel announced the use of a silicon photonic architecture to define the next generation of servers [5], and demonstrated its first inexpensive, functional 100Gb/s optical chip in Apr. 2013 [6]. While industry is making efforts to push ONoCs into reality, designing efficient ONoCs demands extensive research - “research focusing on on-chip optical interconnect as the most probable commercial intercept” will be an “emerging change” for interconnects (ITRS Roadmap 2012 December Meeting [15]).

B. Topology and routing on ONoCs

Topology and routing algorithms play a primary role in achieving high performance in ONoCs. Topology defines how routers are placed and connected, and routing algorithm determines the path taken by a communication. Recently, many studies have been conducted for designing ONoC architectures based on various topologies, such as Mesh [16-20], Fat tree [21-23] and Ring [24-25]. However, most existing designs for ONoCs adopt the routing policies initially developed for electronic NoCs, which cannot fully take advantage of the unique benefits and address the special requirements of ONoCs. These special characteristics of communications on ONoCs can be summarized as follows.

• Low energy consumption by end-to-end bufferless communication: Unlike electronic Network-on-Chips, ONoCs afford high energy-efficiency by transmitting large volumes of data end-to-end without the need for buffering and switching at intermediate nodes [26]. Some of the conventional criteria for routing design, such as shortest path or minimum hop-count, are not suitable to ONoCs.

• Ultra-high bandwidth by concurrent communications: ONoCs offer orders-of-magnitude bandwidth improvement by leveraging on Wavelength-Division Multiplexing (WDM) technology, which allows...
multiple signals to be transmitted concurrently using different wavelengths (“light colors”) through a single waveguide (optical link) [27-30].

- **New on-chip communication requirements:** ONocS also differ from traditional large-scale optical networks and even from multi-chip interconnects, since design methodologies have to take into account new constraints specific to ONoCs, such as chip area/space, heat distribution, implementation complexity, thermal effects and hardware cost [31].

C. Constraints for routing design on ONoCs

ONoCs’ unique characteristics and constraints pose new challenges, with important ramifications on solutions to traditional networking problems in a new context.

- **Wavelength constraint:** Though WDM can significantly reduce communication contention, the number of wavelengths in ONocS is very limited in realistic scenarios (only 4 wavelengths by Intel’s recent report [29]). Thus the wavelength resources used for a design are limited for feasibility, power consumption, chip area and hardware cost considerations [24].

- **Reliability constraint:** Heat distribution is one of the key factors affecting ONoC reliability, as uneven heat causes chip instability [32]. In contrast to electronic NoCs, the heat dissipated on an optical link is independent of the transmission distance, but mainly generated by the processing elements [32]. Another important reliability issue is the signal loss caused by the insertion of optical routers (insertion loss [33]).

- **Complexity constraint:** As complexity is a critical design factor for ONoCs, routing mechanisms on ONoCs have stringent constraints on complexity. For simplicity, XY routing (first in x-direction and then in y-direction) is commonly adopted for Mesh topologies [19,20]. However, XY routing can cause non-uniform contention and non-uniform heat distribution, reducing performance and causing overheating at the central nodes.

These constraints are tightly coupled and can even conflict with each other (e.g. increasing the number of wavelengths can mitigate congestion but may increase router complexity; balancing heat can enhance reliability but may increase routing complexity). Existing ONoC research [16-25, 42, 47-48] has rarely taken into account all the above challenges from a networking perspective, and no comprehensive theoretical models have yet been used to analyse the tradeoffs among the performance, requirements and constraints of ONoCs. Desirable routing solutions, which can be used to guide ONoC design, have not been well studied. In this paper, we propose an optimization framework for routing problems on ONoCs from a networking perspective, which can be used to guide the establishment of the analytical models for routing problems on ONoCs, contributing to the high-level designing of the emerging ONoCs architectures.

II. Optimization Framework for Routing and Wavelength Assignment on NoCs

A. Preliminaries

An ONoC topology can be represented by a graph $G = (V, E)$, where nodes in $V$ are optical routers, and edges in $E$ are the optical links (silicon waveguides) connecting the optical routers. The most commonly used topology for ONoCs is Mesh (Fig.1(a)), as its grid-type shape and regular structure are most appropriate for the two dimensional layout on a chip. Optical routers are used to route optical signals between a set of input and output ports. Much research [34-37] has been conducted on designing the routers (Fig. 1(b)) based on Micro-ring Resonator (MR: filtering a specific wavelength and changing the direction of light). The MR is powered into on-state (Fig.1 (c)) when the signal is turning around, causing more signal loss than the off-state (Fig.1 (d)) when the signal is passing by.

B. Optimization Framework for Routing and Wavelength Assignment

The optimization for routing and wavelength assignment problem on ONoCs is: given an ONoC, how can we design low-complexity routing algorithms and wavelength assignment schemes that use a limited number of wavelengths to maximise performance while satisfying the constraints (e.g. heat balancing and signal loss constraint). This problem has more constraints than the routing and wavelength assignment problem in conventional large-scale optical networks, which is known to be NP-hard [38]. To achieve different objectives and satisfy different constraints on ONoCs, the setup of the optimization models can be different. Based on the properties and constraints of ONoCs, we propose the optimization framework as follows:

- **Achieving high throughput and low communication delay** by minimizing the maximum contention probability (the probability that communications conflict) among the links. As the wavelength resources are limited, the optimization challenge is how to optimise the use of limited wavelength resources to maximize performance (e.g. minimize contention).
• **Enhancing reliability** by balancing heat distribution and limiting the insertion loss (optical power loss resulting from the insertion of the on-chip optical router). The optimization challenge is how to design routing schemes that can optimally balance heat distribution.

• **Reducing complexity** by using regular topology, low-complexity routing algorithms, limited numbers of wavelengths, and low-complexity wavelength assignment approaches. The optimization challenge is how to design low-complexity routing algorithms to achieve the tradeoffs between the performance and complexity.

In this paper, we use the following optimization model to illustrate the optimization framework by using contention probability as the optimization objective, number of wavelengths as the constraint, and Mesh topology as the target architecture.

C. Optimization Modelling Example: Routing-aware wavelength-constrained assignment on ONoCs

The optimization problem is: for a given routing scheme (e.g. XY, YX or its variations), how can a limited number of wavelengths be assigned in such a way to minimise the maximum contention probability on Mesh ONoCs?

![Fig. 3. Illustration of different number of wavelengths (λ1, λ2,...)](image)

(a) single wavelength (high contention)  (b) 16 wavelengths (contention-free)  (c) 3 wavelengths (contention probability?)

Figure 3. Illustration of different number of wavelengths (λ1, λ2,...)

The details of the setup for the optimization modeling are illustrated as follows:

• **Constraint: Wavelength constraint on ONoC**

As illustrated in Fig. 3, most existing schemes regarding to the usage of wavelengths fall into two categories: (1) single wavelength design [41], which suffers from low link utilization and high contention probability; (2) contention-free design [24,43,46], but the scalability is limited by the available wavelengths. In practical implementations, only a limited number of wavelengths are available due to concerns on complexity, reliability and cost.

In the optimization model, the number of wavelengths is used as the constraint in this optimization model.

• **Performance Modeling: ONoC Performance modelled by contention probability**

Unlike electronic interconnects where communication delay is dominated by the transmission distance and the queuing delay at routers, the optical end-to-end throughput is most affected by the amount of time that a shared link in the routing path is occupied by other conflicting communications. As shown in Fig. 2, two concurrent communications c1 and c2 will

(a) **conflict** by XY routing if they are assigned the same wavelength due to the shared link along their routing paths,

(b) **not conflict** by YX routing in the same wavelength as they do not share any links, and

(c) **not conflict** by XY routing if they are assigned different wavelengths even if they share a common link.

![Fig. 2. Illustration of communications conflict (c1, c2), routing (XY,YX) and wavelength assignment](image)

As the conflict situation depends not only on the routing algorithm but also on the wavelength assignment approach, the analysis of contention probability needs to consider both factors. In this optimization model, we propose to set the objective as achieving high throughput and low communication delay by minimizing the maximum contention probability among the links. Given a routing scheme and a traffic pattern (e.g. uniform traffic pattern: each node has the same probability to communicate with all other nodes), the contention probability on a directional link can be computed based on all possible conflicting communications that request to pass through the given directional link at the same time. As illustrated in Fig. 4, the conflicting concurrent communications on the given directional link are communications that use the same wavelength and happen between nodes in the shaded box and nodes on the dashed line by XY or YX routing.

• **Optimization Objective: minimizing the maximum contention probability**

Since different links may experience different contention probability for a given routing and wavelength assignment approach, the links with high contention probabilities will become bottlenecks, resulting in long and uneven end-to-end
communication delay. Hence a good wavelength assignment scheme should not only minimise the average contention but also balance the contention among all links. Therefore, we choose to minimize the maximum contention probability subject to the constraint on the number of wavelengths.

\[
\begin{align*}
(a) & \quad \text{XY/vertical link} \\
(b) & \quad \text{XY/horizontal link} \\
(c) & \quad \text{YX/vertical link} \\
(d) & \quad \text{YX/horizontal link}
\end{align*}
\]

Figure 4. Illustration of analyzing contention probability on directional links of Mesh

**Optimal Solutions**

This problem can be formulated as a Min-Max Integer Programming (IP) problem and use IP solvers such as CPLEX[49] to compute the optimal solution. This approach may not be feasible for large problem instances (e.g. 512×512 with 8 wavelengths) even on a supercomputer due to the demanding requirement on memory. However, analysing the optimal solutions for Meshes with smaller sizes (e.g. 16×16) will enable us to identify the properties of the optimal solution for large problem instances, which can be used to guide the designing of low-complexity near-optimal solutions. Such optimal solutions can also be used as benchmarks to evaluate the performance of low-complexity heuristic schemes. For the future work, we will design a near-optimal wavelength assignment scheme with low complexity, and investigate the relationships between contention probability, low-complexity routing, and wavelength assignment scheme.

D. Simulator for ONoCs

While current commercial optical chips are still not widely used, conducting experiments on real optical chips is very expensive. The performance of the optimization modeling on ONoCs can be evaluated through ONoC simulators, and analyzed under different traffic patterns (e.g. uniform traffic pattern and realistic synthetic traffic patterns) and different network sizes. To evaluate the performance of Optical Network on Chip, some new methodologies and tools should be designed since optical devices are fundamentally different from the electrical ones. PhoenixSim simulator is mainly focusing on physical-layer properties of the photonic network, such as optical insertion loss, crosstalk noise, and energy dissipation [50]. The foundation of this simulator is a Photonic Device Library, which records the physical characteristics of some basic photonic devices. Similarly, another simulator, LioeSim, is proposed for optical-electrical hybrid architecture with extended device library and system-level evaluation for mesh-based optical-electrical hybrid NoC [51]. Zhang et. al. proposed a full-system simulator by integrating PhoenixSim into a system-level simulator with self-correction trace model [52]. To some extent, these simulators have ignored the unique properties of optical communications, which are only suitable for some specific topologies. Many optical NoC architectures, with wavelength routing and all optical topologies (crossbar/ring, multistage optical network), still urgently call for a highly modularized, highly configurable, and highly scalable full-system simulator.

### III Conclusion

With the increasing number of cores on a chip, conventional interconnects for on-chip communications are becoming a bottleneck. As an emerging architecture, Optical Networks-on-Chips (ONoCs) can offer ultra-high bandwidth, low latency, and high energy-efficiency. This paper provided a special view of the unique characteristics and special constraints of routing problems on ONoCs from the networking perspective, and proposed an optimization framework for designing efficient routing and wavelength assignment schemes on ONoCs, by taking advantage of their unique characteristics, and accommodating their special constraints. In our future work, we will fine tune the optimization modeling to investigate schemes that optimally/near-optimally maximise performance by modelling, analysing and evaluating tradeoffs among performance, reliability and complexity. The optimization on ONoCs can guide high-level ONoC design, and help manage the transition from electronic NoCs to ONoCs and from un-optimized ONoCs to optimized ONoCs, thus advancing the development of high performance computing and environment-friendly computer architectures.

### REFERENCES


[34] H. Gu, K. Mo, J. Xu, W. Zhang, A Low-power Low-cost Optical Router for Optical Networks-on-Chip in Multiprocessor Systems-on-Chip, in Proceedings of IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2009. (Best Paper Honorable Mention Award)


