Wavelength-Routed Optical NoCs: Design and EDA — State of the Art and Future Directions

(Invited Paper)

Tsun-Ming Tseng†, Alexandre Truppel†, Mengchu Li†, Mahdi Nikdast‡, and Ulf Schlichtmann†

†Chair of Electronic Design Automation, Technical University of Munich, Arcisstraße 21, 80333 München, Germany
‡Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA

Abstract—Wavelength-routed optical network-on-chip (WRONoC) design consists of topological and physical synthesis. It covers many interacting design aspects such as wavelength assignment, message routing, network construction, component placement, and waveguide routing. Due to the high complexity of the design problem, current manual design usually trades optimality for scalability and feasibility, which results in performance degradation and waste of resources. In this paper, we will present an overview of the existing design automation approaches that have demonstrated their effectiveness in customizing and optimizing application-specific WRONoC designs, and of the potential design automation directions to address a wider range of design challenges. We will also discuss the advantages of comprehensive optimization considering multiple design aspects simultaneously, and the possible barriers that need to be removed to achieve this goal.

I. INTRODUCTION

Silicon photonics has rapidly evolved over the past decade, and aroused great interest in both academia and industry [1]–[5]. With the growing maturity of the technology, optical network-on-chip (ONoC) emerges as a next-generation infrastructure for the data transmission in multiprocessor systems-on-chip (MPSoC) [6]. As the name suggests, ONoCs convert electrical signals into optical signals on different wavelengths and transmit them through optical waveguides, which are the dual of electrical wires, from initiators (senders) to targets (receivers) in the network. With the use of wavelength-division multiplexing (WDM) that allows a single waveguide to accommodate multiple wavelengths, ONoCs are able to deliver energy-efficient on-chip communication with high bandwidth and low latency [7], [8].

In particular, wavelength-routed optical networks-on-chips (WRONoCs) further reduce the latency overhead and deliver on-chip communication in a highly predictable fashion. In WRONoCs, a dedicated data transmission path is statically reserved for each initiator/target pair at design time so that all initiators can communicate with all targets concurrently without data-collision. This arbitration-free communication comes however at the expense of a relatively large amount of optical resource usage, including more microring resonators (MRRs) and wavelengths, which then lead to higher MRR tuning power [9] and laser power [10], respectively. The excessive resource usage is considered to be the major challenge for WRONoCs to be applied to large-scale networks.

To reduce resource usage, researchers have proposed several WRONoC topologies [11]–[13] where the MRRs and wavelengths can be efficiently shared among different signal paths. Nevertheless, two challenges remain before the scalability concerns can be alleviated.

First, the topology needs to be tailored to suit various communication scenarios. Standard WRONoC topologies are designed to support full connectivity, i.e. each initiator is connected to each target. This over-conservative reservation leads to significant resource waste in application-specific systems where all-to-all communication is not required [14].

Second, while topologies only focus on the logic schemes, network components such as MRRs and waveguides need to be placed and routed in a physical plane considering layout constraints. In particular, shorter waveguides with fewer bendings and crossings are preferable for minimizing the laser power. Due to the large design space, handcrafting the layout is highly inefficient and the quality of the results is unpredictable.

Design automation for WRONoCs emerges in this context to reduce design difficulty and to enhance design quality. Over the past years, remarkable progress has been made in both topological and physical synthesis. By exploring the design space in a systematic manner, networks can be automatically customized, and a significant reduction in resource usage and laser power can be achieved.

Despite the inspiring advances, design automation for WRONoCs is still in its infancy. There are essential performance factors to be investigated, and gaps in the design flow to be bridged. In this paper, we will briefly characterize the interacting design aspects of WRONoCs, and present an overview of the state-of-the-art design automation approaches as well as the open challenges. With this discussion, we want to address the great potential of design automation in promoting the development of WRONoCs, and identify the possible barriers that need to be removed to reach this goal.

II. BACKGROUND

A. WRONoC Working Mechanism

A typical setting for WRONoC application is a 3D-integrated chip consisting of vertically stacked photonic layer and electronic layer [13], [14], such as the processor-memory network shown in Figure 1. On the electronic layer there can be multiple clusters of processors, each of which is connected to a hub in the photonic layer by through-silicon vias (TSVs). With laser sources providing wavelengths, electrical signals are converted to optical signals and vice versa at electrical/optical (E/O) and optical/electrical (O/E) interfaces. Data transmission on the photonic layer among hubs and between hubs and memory controllers (MCs) of off-chip memories can be managed by a WRONoC router.
signals passively. For each initiator-target pair, the dedicated routing path is determined at design time.

The message routing mechanism in WRONoCs relies on microring resonators (MRRs). An MRR consists of a microring and a coupling mechanism to access the microring. The radius of the microring defines the resonant wavelengths of the MRR [15]. Figure 2 shows two MRRs that have the same resonant wavelengths but support different switching mechanisms: the parallel switching element (PSE) shown in Figure 2(a) consists of a microring between two parallel waveguides, and the crossing switching element (CSE) shown in Figure 2(b) consists of a microring nearby a waveguide crossing. When signals modulated on a resonant wavelength of the MRR approach the PSE/CSE through a nearby waveguide, they will be coupled to the microring and then switched to an opposite/orthogonal direction on the other waveguide. In these cases, we refer to the signals as being “on-resonance” with the MRRs. On the other hand, when signals modulated on some wavelengths other than the resonant wavelengths approach the MRRs, they will just ignore the microrings and keep their original directions. In this case, we refer to the signals as being “off-resonance” with the MRRs.

In WRONoCs, multiple MRRs of different resonant wavelengths are applied to construct different paths for signals on different wavelengths.

B. Topological and Physical Design Aspects

Compared to other ONoC architectures, WRONoCs require relatively more resource usage, because every communicating initiator/target pair requires a dedicated signal path. The available optical resources are however constrained by technology and power. For example, the maximum number of wavelengths that can be achieved with a 50-nm WDM band is limited to 62 or fewer if lower crosstalk is required [16]; and the usage of wavelengths and MRRs positively correlates with the static power consumption [17]. Thus, it is essential to share the wavelengths and the MRRs among different signal paths in WRONoCs.

Figure 3 shows two WRONoC topologies that both support 16 signal paths but with different resources [13]. In the folded crossbar shown in Figure 3(a), each signal path consists of a dedicated MRR for resonance, and thus 16 MRRs are used in total. But in the Snake topology shown in Figure 3(b), some signal paths do not rely on MRR for data transmission, and only 12 MRRs are used in total. This comparison demonstrates the influence of the topology design on the network performance.

WRONoC topology design consists of three interacting aspects:

- Network construction. The switching mechanisms of MRRs depend on the nearby waveguides. Current WRONoC topologies are mostly based on two waveguide structures: crossbar [18], which naturally supports both PSE and CSE, and ring [19], which avoids waveguide crossings and does not rely on MRR for signal routing.
- Wavelength assignment. At least one wavelength needs to be assigned to each signal path. A wavelength can be shared among multiple signal paths, but the following constraints must be satisfied to avoid data-collision: 1) wavelengths assigned to the signal paths between the same initiator and different targets must be different; and 2) wavelengths assigned to the signal paths between different initiators and the same target must be different.
- Message routing. The MRR usage and the logic connections among the MRRs, initiators, and targets need to be determined, so that signals can be routed from their initiators to their designated targets based on their wavelengths. In particular, the paths of signals on the same wavelengths but between different initiator/target pairs must not overlap.

To design a WRONoC topology with as few wavelengths and MRRs as possible, the three aspects should be considered comprehensively with respect to the network communication requirements. The design space is however huge and increases drastically as the network size increases.

While topology design focuses on the logic schemes, the network
components such as MRRs and waveguides need to be placed and routed in a physical plane considering layout constraints such as the locations of the initiators/targets, the routability of the waveguides, the size of the components, etc. For example, Figure 4 shows a \( \lambda \)-router topology and a physical layout that maps this topology to the optical layer of a 3D-stacked chip, where the locations of the hubs and memories are fixed and the size of the chip is limited. We can clearly see the increased complexity of the final layout and envision the huge design space.

WRONoC physical design consists of placement of the network components and routing of the waveguides. In particular, the insertion loss (signal degradation) in a signal path positively correlates with the lengths of waveguides (propagation loss), the number of waveguide bends (bending loss) and the number of waveguide crossings (crossing loss). Since the signal path that introduces the highest insertion loss determines the laser output power per wavelength [9], the physical design has a great impact on the network performance.

III. DESIGN AUTOMATION: THE STATE-OF-THE-ART

Design automation for WRONoCs has emerged in the last decade to ease design difficulty and to enhance design quality. Remarkable progress has been made in both topological and physical design. A recent approach has also demonstrated the possibility to combine the topological and physical synthesis and optimize all the interacting design aspects comprehensively. A review of the state-of-the-art design automation approaches will be given in this section.

A. Topology Customization

Standard WRONoC topologies focus on a specific subset of communication graphs that require full connectivity among all initiators and targets [11]–[13]. However, fully connected WRONoC topologies suffer from quadratic scaling of the number of MRRs with respect to initiator/target nodes [9]. Since the number of MRRs positively correlates with the manufacturing cost and the static power consumption, fully connected topologies are limited only to small networks [14], [21].

[22] lays the ground work for automatic topology synthesis based on crossing switching elements (CSE). It identifies that for an \( n \times n \) topology with \( n \) initiators and \( n \) targets, \( n(n-1) \) CSEs are required to route the signal paths. It also identifies that by combining two CSEs of the same resonant wavelengths into a 2-input \( \times \) 2-output add-drop filter (ADF), \( [(n-1)*(n-2)*\cdots*(n-n+2)]^n \) different topologies can be found in the design space. It proposes a four-step methodology to exhaustively explore the design space. However, the proposed method is limited to \( n \times n \) topologies with full connectivity, and for \( n > 4 \), the enumeration of all possible design points becomes computationally unaffordable.

Topology customization for WRONoCs is first mentioned in [14]. It proposes a method to remove the redundancy in the \( \lambda \)-router topology for application-specific designs. Though the performance of the proposed method is limited by the \( \lambda \)-router topology which is designed to provide full-connectivity, it demonstrates the potential benefit that one can expect from topology customization.

[23] proposes CustomTopo, a design automation approach to synthesize WRONoC topologies with customized connectivity. Figure 5 shows the synthesis flow of CustomTopo. The input of CustomTopo is a communication graph, as shown in Figure 5(a). Based on the communication graph, it synthesizes a communication matrix where the entry in the \( i \)-th column and in the \( j \)-th row represents the wavelength assigned to the signal path from initiator \( i \) to target \( j \), as shown in Figure 5(b). The communication matrix is optimized to form as many as “ADF-sharing structures” as possible, so that the MRR usage can be minimized. Based on the communication matrix, it then synthesizes a WRONoC topology with the objective of minimizing the insertion loss, as shown in Figure 5(c). CustomTopo proposes the first mathematical formulation of the topology synthesis problem, and demonstrates that the resource usage in customized topologies does not necessarily increase with the network size, i.e. the number of nodes in the network, but depends on the communication density, i.e. the number of required signal paths.

Design automation for topology customization has shown significant advantages in resource usage. Take a “Screen Savor” multimedia network [24] that consists of 12 nodes and 26 signal paths for example: compared with the fully-connected \( \lambda \)-router [12] that uses 132 MRRs, the method proposed in [14] reduces the MRR usage to 48 (64% reduction), and the topology synthesized by CustomTopo only consists of 26 MRRs (80% reduction).

B. Physical Design

The first automatic physical synthesis tool for WRONoCs is called PROTON and proposed in [25]. PROTON and its follow-up works
PROTON+ [26] and PLATON [27] can automatically determine a geometrical description of all optical devices (ADFs) and waveguides of a WRONoC topology on a physical plane, where the locations of initiator/target nodes are predetermined. Placement and routing are performed as two sequential steps considering the minimization of the insertion loss in the worst-case signal path. Compared to handcrafted physical layouts for λ-router [12] and GWOR [11] topologies, PROTON reduces the worst-case insertion loss by nearly a half, which allows the laser power to be reduced by up to 99%.

[28] proposes a thermal-aware physical design approach based on PLATON. Silicon photonic devices are thermal-sensitive. Temperature fluctuation can cause the resonant wavelengths of MRRs to drift [29] and degrade the efficiency of on-chip lasers [30]. [28] calculates the thermal distribution based on the distribution of the cores on the electrical layer and places the optical devices away from the hot areas.

[31] figures out that most standard WRONoC topologies are planar, i.e. there is no extra waveguide crossings outside the ADFs. However, when designing the physical layout, the PROTON family of tools introduces a large number of additional waveguide crossings. This leads to high insertion loss. [31] thus proposes a three-step physical synthesis approach named PlanarONoC, which is able to maintain the planarity of a planar logic topology. PlanarONoC first constructs a connection graph considering the orientations of ADFs for a given topology. If the connection graph is not planar, it extracts a maximally planar sub-graph. Based on the connection graph, it then applies a sophisticated algorithm to determine the routing strategy, and finally performs concurrent placement and routing to synthesize the layout. Compared to the PROTON tools, PlanarONoC trades an increase in the waveguide lengths off for a minimized number of waveguide crossings. Since the crossing loss plays a more dominant role than the propagation loss with current WRONoC technology parameters, PlanarONoC enables further reduction of the worst-case insertion loss.

C. Comprehensive Optimization

In the state-of-the-art flow, to confine the huge design space of WRONoCs, logic synthesis and physical design are separated into two sequentially-performed steps. This separation however implies an optimization gap, since physical design greatly depends on the logic topology, but in the logic synthesis step, the physical features of the optical layer are mostly ignored and thus the network performance cannot be accurately predicted and optimized.

[32] (PSION) makes the first attempt to bridge this optimization gap by combining logic synthesis and physical design as a single step. To deal with the resulting increase of the synthesis complexity, it proposes the use of a physical layout template, which can be considered as a collection of predefined waveguide- and MRR-placeholders. Instead of exhaustively exploring the whole design space, PSION only explores the placeholders provided by the template, and models the logic synthesis and physical design problems together using mixed-integer linear programming considering insertion loss. Figure 6 shows a WRONoC design synthesized by PSION, where MRRs are selectively placed in the placeholders and configured to various resonant wavelengths, so that a collision-free data transmission path is reserved for each communicating initiator-target pair. Compared to PROTON+ and PlanarONoC which perform physical synthesis based on a given topology, PSION is able to reduce the worst-case insertion loss by up to 40%, and the number of MRRs by up to 28%.

Though the benefits of comprehensive optimization have been demonstrated by PSION, there are still some barriers to be removed. First, compared to the design automation approaches that perform logic synthesis and physical design separately, PSION is more computationally expensive. For example, PSION needs a few minutes to synthesize a design with 8 nodes, while PlanarONoC can synthesize the physical layout within a second. To deal with large-scale networks, the computational efficiency must be improved. Second, the performance of PSION strongly depends on the physical layout templates. Compared to a general-purpose template, a customized template allows the MRR usage to be further reduced by 20%. However, automatic template customization calls for more research efforts.

IV. DESIGN AUTOMATION: THE CHALLENGES

Though significant progress has been made, design automation for WRONoCs is still in its early stage. There are important performance factors that have been overlooked, and a large portion of design space that remains unexplored. In this section, we will introduce some design automation challenges that need to be tackled in the future.

A. Crosstalk

Crosstalk is an intrinsic characteristic of MRRs and waveguide crossings [33]. When two signals reach an MRR or a waveguide crossing simultaneously, a small portion of the power of one signal is directed to the other signal and becomes noise. Crosstalk noise degrades the signal-to-noise ratio (SNR) and in turn requires higher laser power [16]. In ONoCs that consist of multiple MRRs and waveguide crossings, crosstalk noise will accumulate and become a severe issue that constrains the network scalability.

Current design automation approaches mostly focus on insertion loss and overlook crosstalk noise. However, [33] demonstrates that the path with the lowest SNR may not be the path that suffers the most insertion loss. Thus, crosstalk noise needs to be considered as a separate performance factor and carefully addressed in both topological and physical design.

Crosstalk noise has been analyzed for mesh-based [33], folded-torus-based [34], and fat-tree-based [35] ONoC architectures, and a formal study of crosstalk noise in an arbitrary WDM-based ONoCs...
is proposed in [36], which provides the necessary theoretical foundation for developing crosstalk-aware design automation methods.

B. Bit-Level Parallelism

So far most WRONoC topology designs assume that each initiator/target pair communicates with one bit at a time using one wavelength. However, MRRs have a periodic transmission characteristic which allows them to support multiple resonant wavelengths [15]. Thus, the communication parallelism of a signal path can be increased by assigning multiple wavelengths to this path. [10] proposes the first approach to increase the bit-level parallelism by optimizing the selection of MRR radii in a given topology. However, [37] points out that the achievable bit-level parallelism does not only rely on the device parameters but also depends on the topology.

As many NoC systems have multiple different bandwidth requirements [24], [38], topology customization considering bit-level parallelism will substantially enlarge the application scope of WRONoCs. Besides, since increasing the bit-level parallelism indicates using more wavelengths, which further results in more laser power and crosstalk noise, a comprehensive study of the trade-off needs to be performed to ensure the practicability of the designs.

C. Alternative MRR Switching Elements

Current design automation research focuses on the crossing switching element (CSE), as shown in Figure 2(b), and overlooks the parallel switching element (PSE), as shown in Figure 2(a).

Compared to CSE, PSE avoids waveguide crossing and allows signals to be coupled to an MRR between parallel waveguides. As waveguide crossings are an important source of insertion loss and crosstalk noise, PSE shows potential advantage in lower laser power and higher SNR. Some manually designed ONOC routers [33], [39] demonstrate the combined usage of both PSE and CSE, which achieves promising results in insertion loss and crosstalk noise reduction.

Taking PSE into consideration implies more topological variants and thus significantly extended design space, which calls for more research efforts.

D. Alternative Topological Structures

Besides optical crossbar structures relying on MRR coupling, a ring-based topology named ORNoC has been proposed in [19] where no waveguide crossing is introduced and no MRR is used for signal routing. Compared with the crossbar-based topologies such as λ-router, GWOR, and Snake, ORNoC significantly reduces the drop loss and completely avoids the crossing loss, which leads to better energy-efficiency in small networks [18]. However, since ORNoC uses longer waveguides which result in more propagation loss, it does not scale well and is outperformed by the crossbar-based routers in large networks [13].

ORNoC demonstrates a structural alternative worthy of investigation beyond crossbar, which has not yet been systematically explored with design automation methods.

E. Power Distribution Network

Current WRONoC lasers can be classified into two categories: on-chip and off-chip. On-chip lasers are under rapid development and show potential advantages in energy efficiency [40], but they are more temperature sensitive [30] and their technology is not yet mature. Therefore, off-chip lasers are still considered as the more practical solution by many researchers [41]–[43]. Compared to on-chip lasers, off-chip lasers require an additional power distribution network (PDN) to split the laser power from the off-chip lasers to the corresponding waveguides connected to the initiators.

So far the PDN has been designed manually [44] and not yet been considered in the existing automatic physical design approaches. However, the splitting and routing of waveguides in the PDN can also have a substantial influence on the WRONoC performances. In particular, additional waveguide crossings may be introduced and degrade the signals. Thus, it is necessary to add PDN to the automatic synthesis flow considering the routing of PDN waveguides, the placement of waveguide splitters, and the strategy for waveguide splitting.

V. Conclusion

WRONoC design covers many interacting aspects which imply a huge problem space. However, advances in the design automation field demonstrate that this space can be systematically explored with the aid of algorithms and computational power. In this paper, we gave a brief overview of the WRONoC design problems and introduced the state-of-the-art design automation approaches that address these problems. As design automation for WRONoCs is still in its infancy, we also discussed some performance factors and design aspects that have been so far overlooked and need to be tackled in the future.

VI. Acknowledgments

The authors would like to give their sincere thanks to their long-term collaborator Prof. Davide Bertozzi from the University of Ferrara. As an insightful WRONoC designer, he always appreciates the capability of design automation and promotes the development of this field.

References
