Layout Guidelines for 3D Architectures including Optical Ring Network-on-Chip (ORNoC)

Sébastien Le Beux, Jelena Trajkovic, Ian O'Connor, Gabriela Nicolescu

Layout Guidelines for 3D Architectures including Optical Ring Network-on-Chip (ORNoC)

Sébastien Le Beux*, Jelena Trajkovic†, Ian O'Connor* and Gabriela Nicolescu†

*Lyon Institute of Nanotechnology,
École Centrale de Lyon, France
Email: {Sebastien.Le-Beux, Ian.OConnor}@ec-lyon.fr

†Software and Computer Engineering Department
École Polytechnique de Montréal, Canada
Email: {Jelena.Trajkovic, Gabriela.Nicolescu}@polymtl.ca

***Abstract*—Trends in design of the next generation of Multi-Processors System on Chip (MPSoC) point to 3D integration of thousand of processing elements, requiring high performance interconnect for high throughput and low latency communications. Optical on-chip interconnects enable significantly increased bandwidth and decreased latency. They are thus considered as one of the most promising paradigms for the design of such system. However, existence of interfaces between electronic and photonic signals implies strong constraints on the layout of the 3D architecture and may impact the architecture scalability. In this paper, we propose and evaluate a possible layout for an optical Network-on-Chip used to interconnect processing elements located on different electrical layers.**

I. Introduction

The latest edition of ITRS (International Technology Roadmap for Semiconductors) [7] emphasizes "More Than Moore's Law" trend. This trend focuses on system integration rather than transistor density, allowing for both functional and technological diversification in integrated systems. The functional diversification allows for non-digital functionalities to migrate from the board level into chip-level. This allows for integration of new technologies that enable high performance, low power, high reliability, low cost and high design productivity.

Moreover, 3D integration paradigm and technology scaling down to ultra deep submicron domain provides for billions of transistors, which enable the integration of hundreds of cores on a single chip. These cores, running at a high clock frequency, create a need for high data bandwidth and increased parallelism. Therefore, the role of interconnect becomes a dominant factor in performance. Designing such systems using traditional electrical interconnect poses a significant challenge: due to capacitive and inductive coupling [6] interconnect noise and propagation delay of global interconnect increase, which puts limits to achievable bandwidth and overall system performance.

Use of Optical Network-on-Chip (ONoC) promises to deliver significantly increased bandwidth, increased immunity to electromagnetic noise, decreased latency, and decreased power. Aside from physical properties, use of wavelength routing and Wavelength Division Multiplexing (WDM) [20] contributes to the valuable properties of optical interconnect by enabling low contention or even contention-free routing. WDM allows for multiple signals to be transmitted simultaneously, facilitating higher throughput. The current technology is mature enough to allow this integration, thanks to CMOS-compatible optical components, such as light sources [10], waveguides [11], modulators [16], [17], and detectors [18], [21].

Defining new architectures while taking advantage of optical interconnects represents today a key issue for 3D MPSoC designers. This paradigm shift requires new methodologies for the efficient design. The design methodologies have to take into account the new constraints specific to optical interconnect. For instance, the number of waveguides and wavelengths used for a design is limited for feasibility, variability, power consumption, area and cost considerations.

In our prior work, a 3D architecture including ONoC (named ORNoC for Optical Ring Network on Chip) was proposed [14]. This architecture is composed of identical electrical layers dedicated to computations and an optical layer dedicated to the communications between the layers (i.e. *inter-layer* communications). ORNoC is contention-free (no need for arbitration) network with high throughput and low latency. In ORNoC, a single wavelength can be reused for multiple communications on a single waveguide while taking into account the design constraints. Consequently, fewer waveguides are required and the scalability is facilitated.

The optical layer is of key importance in 3D architecture including ORNoC, as illustrated in Figure 1. Each electrical layer interfaces with ORNoC through a set of Optical Network Interface (ONI) (represented by arrows on the figure) that basically convert electrical signals into optical signals (and vice versa). ONI parts are located on both electrical and optical layers: dedicated vertical connections between these layers are thus necessary. Managing such vertical connection through TSV (Through Silicon Vias [15]) is obvious solution in case both layers are located adjacent to each other. However, in case electrical layer is not directly closed to the optical layer, vertical connections will necessary have to cross intermediate electrical layers, which require different layout for each electrical layer in order to propagate the vertical connections. This drastically increases the design cost of such an architecture, which is nevertheless intrinsically regular. Proposed architectures allow stacking identical electrical layers [13], but the regular vertical connections are shared by all the crossed layers. Since dedicated vertical connections are required, a new approach is necessary.

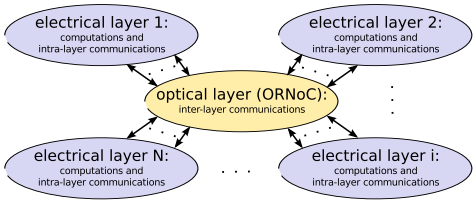


Fig. 1. 3D architecture including ORNoC.

In this paper we propose and evaluate layout for 3D architecture including ORNoC. The main idea is to provide dedicated vertical connections between each electrical layer and the optical layer. By allocating additional area to propagate vertical connections across the electrical layers, the architecture remains regular and the same layout can be used to design all the electrical layer of the architecture, which does not require redesign and therefore does increase the design cost.

The paper is organized as follows. Section II discusses related work. Section III describes the 3D architecture including ORNoC and Section IV presents the design methodology. Section V presents the experimental results and Section VI concludes the paper.

II. Related Work

Several contributions address ONoC design exploiting both electrical and optical NoC technologies. An approach using electrical interconnects for control flow and optical interconnects for data flow was proposed in [27]. The electrical signal precedes the optical one in order to reserve the optical path. Therefore, optical communications may be delayed until an optical path becomes free, resulting in contention delay. Hence, this type of network is not contention-free. The same observation is made for the fat tree and the mesh ONoC proposed in [5], [4]. In [24], [9], electrical interconnects manage local communication while an optical interconnect is responsible for global communications. However, such Single-Write-Multiple-Read (SWMR) implementation implies that each wavelength flowing through the ONoC must be assigned to a given ONI, avoiding parallel communications through the same wavelength. Such technique drastically affects the ONoC scalability. Because we use WDM and reuse wavelengths to realize more than one communication on the same waveguide at the same time, ORNoC does not suffer from this problem.

The Corona architecture [29], [1] follows a Multi-Write-Single-Read implementation that requires arbitration to manage write conflicts. Arbitration is not required in ORNoC. The Firefly architecture [23] extends the prior work by proposing the implementation of reservation-assisted SWMR buses. The main objective is to reduce the power consumption of optical communications by using an initialization packet in charge of turning-on data receiver resources. As a drawback, extra-latency is required compared to the SWMR technique and the network throughput rapidly decreases with the token round-trip latency [22]. FlexiShare architecture [22] reduces this drawback by allowing the injection of a new token each cycle (token stream arbitration). In contrast, ORNoC does not require arbitration.

Only [8] and [2] consider contention-free ONoC, but they do not consider any method to reduce the implementation complexity when total connectivity is not required. The cores of an ATAC processor [12], [25] are connected via an electrical and an optical network. The optical network is used for global broadcasting. Its topology is most similar to the proposed ORNoC, however, the contention-free property is based only on WDM, while in ORNoC it is based on both WDM and wavelengths reuse. Moreover, contrary to our approach, ATAC does not support simultaneous communications between one source and multiple destinations, unless it is broadcast of the same message. Our approach has the potential for fewer waveguides/wavelengths, which eases scaling to large 3D architectures.

3D integration paradigm allows the integration of heterogeneous technologies (e.g. electronic and photonic) in a same system. Basically, 3D architectures consists of stacked 2D layers that are interconnected through TSV. Main advantages of TSVs are their low latency and low power, and main drawbacks are area size and design cost. TSV thus need to be used carefully in order to find best efficiency/area cost trade-off solutions. Several methodologies allow minimizing the number of TSVs [26], [31] and optimize their location on a die in order to maximize their benefits for a given application. While such methodology results in efficient architectures, the resulting layers are application-specific and may be difficult to reuse in other context (e.g. to execute other applications) and to scale. We believe that architecture genericity and scalability comes with the regularity: the more an identical pattern is regularly repeated on a die (e.g. such as in Mesh and Torus networks), the more an architecture is generic and scalable. The same principle can also be applied to 3D architectures: the more an identical layer is regularly repeated, the more the architecture is generic and scalable. The architecture proposed in this paper follows this trend: it is composed of identical electrical layers dedicated to computation and an optical layer dedicated to the communications between the electrical layers. Our approach thus has the potential for scaling to complex systems.

III. 3D Architecture

This section introduces the overall 3D architecture and ORNoC.

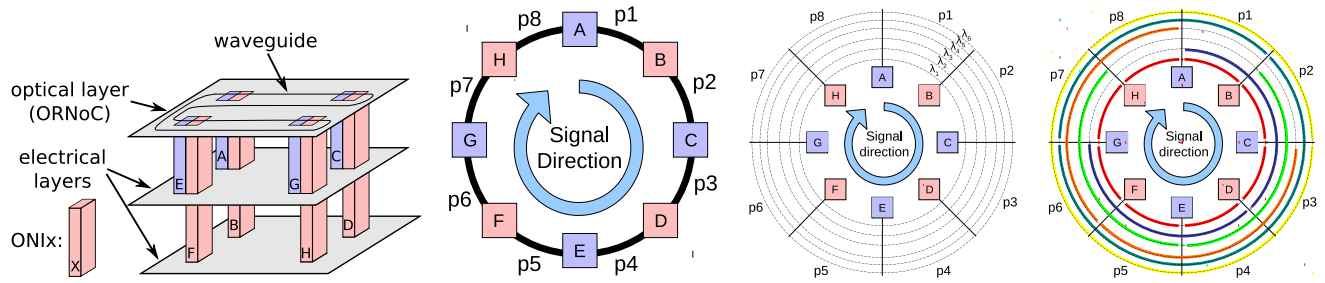
A. Architecture Overview

The architecture is composed of a set of stacked electrical layers and one optical layer. The electrical layers are composed of a set of computing nodes interconnected through a NoC while the optical layer integrates the ORNoC network (i.e. including on-chip lasers, waveguides, etc.). All electrical layers are connected to ORNoC using electrical vertical TSVs (Through Silicon Vias [15]) that upload and download the data between electrical and optical layers. ORNoC is dedicated to *inter-layer* communications (i.e. communications between nodes of different electrical layers) while *intra-layer* communications are locally managed by each layer. Access to ORNoC are provided through Optical Network Interface (ONI) that realize electrical-to-optical and optical-to-electrical conversions. Figure 2(a) illustrates a possible architecture integrating 2 layers and 4 (i.e. 2×2) ONIs per layer: ONIs A, C, E and F provide ORNoC access to one layer and ONIs B, D, F and H provide ORNoC access the other. ORNoC is located on the optical layer, i.e. on top of the 3D architecture, it manages all the inter-layers communications (communication from A to B, from A to D, etc.).

B. ORNoC Architecture

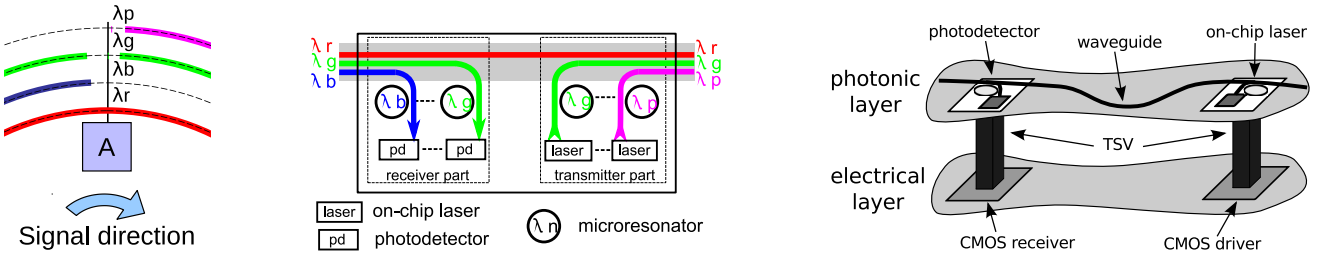
Figure 2(b) represents how the 8 ONIs on the optical layer are connected by the optical ring network, assuming a single waveguide is necessary. The ring is a waveguide crossing all the ONI; each ONI has the potential for injecting and ejecting signals based on predefined wavelengths. We define as *partition* a part of the waveguide located between two ONIs. In this example, the waveguide is composed by 8 partitions (i.e. $p1 \dots p8$). Figure 2(c) illustrates the same network, but from a wavelength point of view. In this representation, the single waveguide is represented as multiple virtual rings, each one being associated to a given wavelength. In this example, we consider that 6 different wavelengths can be used: 6 rings are thus represented ($\lambda_1 \dots \lambda_6$).

The unmatched feature of ORNoC is that a communication is realized by using a given wavelength only on a set of consecutive partitions. As a consequence, the same wavelength may be used on a single waveguide to concurrently realize multiple communications at a same time, without any contention [14]. The wavelength reuse translates into smaller overall number of wavelengths used in the system and, therefore in better scalability. This feature is jointly used with WDM in order to increase the number of possible parallel communications. Figure 2(d) illustrates how the network can be virtually partitioned according to the possible communications schemes (i.e. including communications from ONI_A to ONI_B, from ONI_B to ONI_C, etc). In this representation, each colored wavelength represents such a communication. Figure 3(a) details this behavior of ONI_A.



(a) Example of 3D architecture including ORNoC (b) focus on ORNoC (c) wavelengths in waveguide (d) communications in ORNoC

Fig. 2. Communications in 3D architecture including ORNoC.



(a) ONI behavior (b) Optical part of the ONI (c) A vertical connection in ONI

Fig. 3. Optical Network Interface (ONI) in ORNoC

In order to design an efficient and feasible ORNoC we proposed a design methodology that takes into account design constraints and design/connectivity requirements [14]. The objective of the methodology is to customize a generic ORNoC architecture according to a given scenario by providing a sufficient number of waveguides for realizing all the communications in a contention-free manner. The virtual waveguide partitioning is used in order to minimize the number of required waveguides. The algorithm takes into account the maximum number of wavelengths that is tolerated to satisfy reliability. Finally, in case multiple waveguides are necessarily, we alternate the use of clockwise and counter-clockwise rotation ring in order to minimize the communication length, and thus the power consumption.

C. Optical Network Interface

The waveguide virtual partitioning key feature is made possible by ONI. ONI are composed of a transmitter part and a receiver part. Each part includes electrical and optical components (respectively located on electrical and optical layer). The optical part of ONI is shown in Figure 3(b). ONI's receiver and transmitter parts consist of sets of n_r and n_t microresonators, respectively, each set characterized by its resonance wavelengths: λ_{ri}, for receiver and λ_{tj}, for transmitter part. Trajectory of the signal depends on the value of the signal's wavelength λ_s. Therefore, we define three modes of operation:

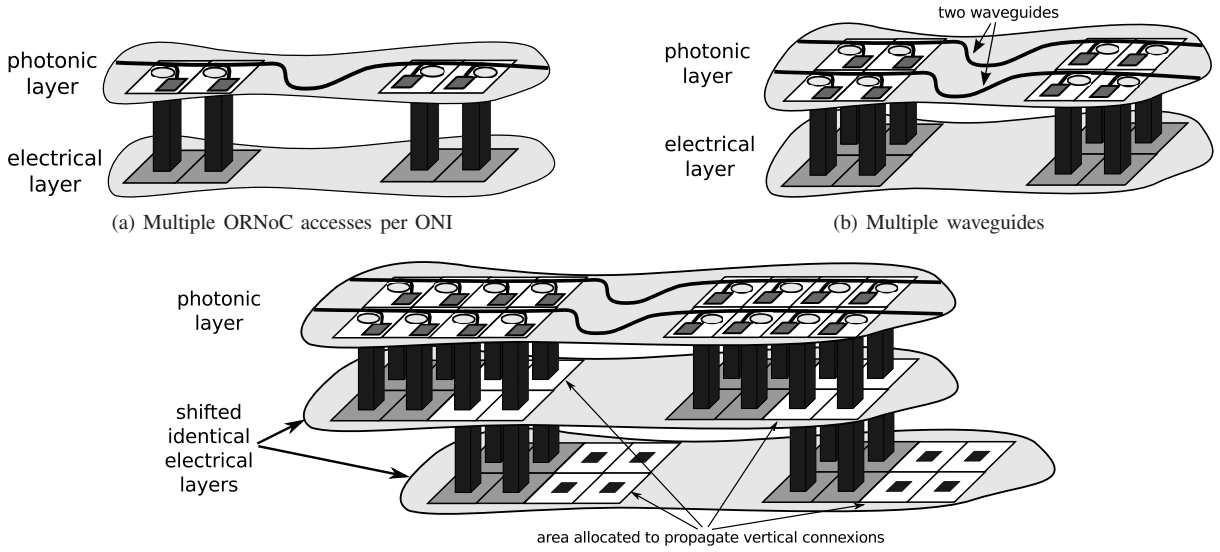
- *inject*: on-chip laser emits a signal which couples in the waveguide.
- *eject*: the received signal couples into a microresonator and then couples out into the perpendicular waveguide towards photodetector.
- *pass through*: the signal propagates along the waveguide.

Figure 3(c) represents both electrical and optical parts of vertical connections used in ONI. In this example, a single electrical layer is considered and the optical layer is located on top of the 3D architecture. From the electrical layer side, CMOS drivers modulate laser signals (*inject* mode) and CMOS receivers convert current transmitted by photodetectors (*eject* mode). TSV are used to realize dedicated vertical connections between electrical and optical layers. CMOS drivers and receivers thus, respectively, initiate and terminate communication through ORNoC. From this figure and the components design characteristics used in the conversion process, the area size required to implement such a vertical connection can be evaluated for both electrical and optical layers. The largest obtained area then gives VC_{area}, the area to allocate in both layers to manage a given vertical connection.

On electrical layer side, CMOS drivers and receivers area size was estimated at $14\mu m \times 12\mu m$ in $0.13\mu m$ CMOS technology by considering $800\mu A$ laser current threshold [19]. By considering $350\mu A$ current threshold observed in [28], with the same technology, the driver area size may drop to approximately $8.5\mu m \times 9.5\mu m$.

On the optical layer side, a microdisk laser with $7.5\mu m$ radius was demonstrated [28]. In ORNoC, this microdisk is coupled to passive microdisk resonator to inject signal in the waveguide. Since the microdisk resonator radius depends on its resonant wavelength, the largest radius is considered, i.e. $10\mu m$. By considering $1\mu m$ diameter for the waveguide [20], the total area size required for injecting signal into the waveguide can be approximated to $10\mu m \times 18.5\mu m$. Since this area is larger than the area allocated for i) electrical components, ii) TSV ($pitch \approx 5\mu m \times 5\mu m$ [3]) and iii) receiver part of the optical layer (area footprint of less than $20\mu m^2$ was reported for the ATAC architecture [12]), it is considered as the basis for evaluating area overhead introduced by each vertical dedicated connection, i.e. $VC_{area} = 10\mu m \times 18.5\mu m$.

The layout guidelines to facilitate such connection are obvious in case the optical layer is directly located on top of the electrical layer. However, in case 3D architecture includes multiple electrical layers, vertical connections will necessary have to cross intermediate layers. Since these connections are dedicated, traditional architectures relying on shared vertical communications [13] cannot be used. Thus, new approach to provide such direct connections is mandatory.



(a) Multiple ORNoC accesses per ONI

(b) Multiple waveguides

(c) Multiple layers

Fig. 4. Possible layouts for 3D architecture including ORNoC.

IV. Layout Guidelines for Layers Crossing Vertical Connections

3D architecture including ONoC that manages inter-layer communications requires dedicated TSV that potentially needs to cross multiple intermediate electrical layers. This constrains the layout of each crossed electrical layer: in addition to allocating area to initiate or terminate vertical communications (by using CMOS drivers and receivers), area needs to be allocated to propagate layers crossing vertical connections (i.e. coming from other electrical layers). Another key characteristic of 3D architecture is the design cost: more regular architectures allow stacking of identical layers, which drives down its cost. However, the electrical layers used in the proposed architecture are intrinsically different since, based on their distance from the optical layer, they are crossed by different number of vertical connections, as illustrated in Figure 2(a). Key issue is, thus, to keep layers identical while they are crossed by different number of vertical connections.

In order to solve this issue, we propose a layout guidelines for designing the electrical layers. In addition to allocating area for initiating and terminating vertical connections, extra area is allocated to propagate crossing layer connections: at the cost of extra area, the electrical layer remains regular and direct access to ONoC is provided to all the layers. This additional area can be aligned with the initiating CMOS driver or the terminating CMOS receiver by shifting electrical layers during the stacking process. By stacking identical electrical layers, a low cost scalable 3D architecture with high performance inter-layer communications can be obtained.

A. Layout Guidelines

The following presents how the approach is extended to support additional access per ONI, additional electrical layers and additional waveguides. The general case and the corresponding area cost model are then presented.

1) Multiple Access per ONI: In case multiple connections between an electrical layer and the photonic layers are required in ONI, the area dedicated to initiate and terminate vertical connections are regularly replicated following a same pattern, as illustrated in Figure 4(a). Having laser sources and photodetectors allocated alongside the waveguide facilitates better layout (fewer waveguide crossings and bends). The limiting factor of the approach is the maximum number of wavelengths to be used in the network; if additional connections are required to satisfy communication scenarios constraints, additional waveguides are necessary in the ONoC [14].

2) Multiple Waveguides: The need for additional waveguides occurs when WDM technique does not provide enough wavelengths for realizing all possible communications or when data are not serialized (e.g. to be compliant with a bus interface [24]). Indeed, one has to take care of the number of wavelengths used per waveguide: for 5nm lithography and 90nm technology, for wavelengths between 1557-1583nm and microresonator radii from 1.0-2.5μm, waveguide with 16 multiplexed wavelengths have been fabricated [30]. In order to scale, ORNoC offers the possibility of multiple waveguides. Since each waveguide is independent from each other, the same set of wavelengths can be reused to realize additional communications. When additional waveguides are used in ORNoC, additional laser sources and photodetectors are required to interface with the new waveguides. Basically, the structure used in the single electrical layer scenario is replicated, as illustrated in Figure 4(b). The pattern is replicated following the waveguide perpendicular direction in order to avoid any waveguide crossing, which contributes to reducing power losses and, thus, to the architecture scalability.

3) Multiple Electrical Layers: The need for ONoC in 3D architecture comes with the growing number of stacked electrical layers. In order to provide direct access to the photonic layer from each electrical layer, additional area dedicated to the propagation of vertical connections is necessary. The allocation of additional area to propagate vertical connections depends on the number of electrical layers. However, assuming regular architecture, the number of vertical interconnects initiated and terminated by each layer is the same from a layer to another. This regularity can be exploited at the layout level. In Figure 4(c), we consider two electrical layers requiring direct connections to a same waveguide. Compared to the scenario illustrated in Figure 4(a), two ONI are required, which doubles the number of laser sources and photodetectors on the photonic layer. By considering the electrical layer adjacent to the optical layer, the only difference from the first scenario is the additional area allocated

to propagate the vertical connections linking the other, non-adjacent, electrical layer to the photonic layer. In order to correctly propagate the vertical connections, the CMOS driver and receiver need to be correctly aligned. For this purpose, the electrical layers can be shifted from each other during the stacking process. The same principle can be applied for any number of electrical layers. The main benefit of this approach is that electrical layers are the same, which contributes to reducing the architecture design cost. This layer reusability comes at the price of area overhead that needs to evaluated.

B. General Case and Area Overhead Estimation

The area overhead requires considering the general case architecture, taking into account the ONoC architecture configuration (i.e. number of wavelengths and number of waveguides), the number of electrical layer and the number of ONI per electrical layer.

The atomic element to consider for evaluating the area overhead is VC_{area}, i.e. the area size of the most area consuming component to realize a single vertical connection (from CMOS driver to laser source, or from photodetector to CMOS receiver). The area size for each ONI, ONI_{area}, is obtained by considering the number of wavelengths WL and the number of waveguides WG in ORNoC:

$$ONI_{area} = 2 \times VC_{area} \times WL \times WG$$

The area overhead $ONI_{area\_overhead}$ used to propagate vertical connections is obtained by considering the number of electrical layers L. Since each ONI needs to be replicated L times:

$$ONI_{area\_overhead} = ONI_{area} \times L$$

Finally, in case multiple ONIs are required on each electrical layer, the total area size T_{area} for allocating ONI is:

$$T_{area} = ONI_{area} \times (nb\_ONI\_per\_layer - 1)$$

and the total area dedicated to propagate vertical connections $T_{area\_overhead}$ is:

$$T_{area\_overhead} = ONI_{area\_overhead} \times (nb\_ONI\_per\_layer - 1)$$

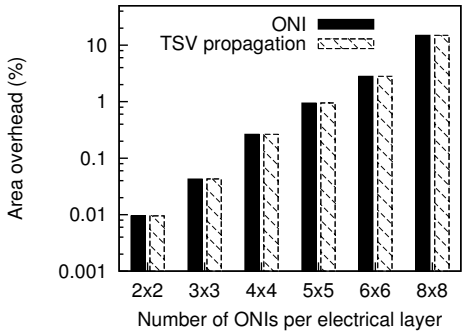
C. Discussion

In the proposed formula, we consider that each wavelength of each waveguide is accessible from any ONI. However, the ONI *pass-through* mode implies that only some wavelengths can be used from a given ONI. The area size overhead estimation thus relies on the worst case scenario. By considering such worst case scenario, we can guaranty that the resulting layout is generic enough to design each electrical layer of the 3D architecture. Only the optical layer needs to be customized based on the considered communication scenario.

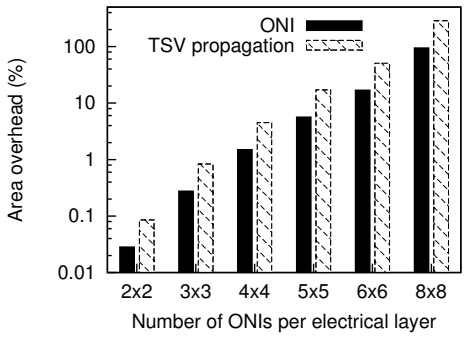
The resulting electrical layers are designed to interface with ORNoC, which is a contention free network. The high throughput provided by ORNoC may come at the price of a large number of laser sources and photodetectors that will impact the size of electrical layer. By sharing network accesses, this area overhead could be reduced. We plan to investigate such performances/area design trade-off in our future work. For this purpose, other components used to interface with optical interconnect (e.g. serializer and deserializer) will be considered. They are not considered in this work since they do not impact the area dedicated to propagate vertical connections.

V. Case Study

In order to evaluate the area size dedicated to vertical connections, we consider the architecture proposed in [29] as the elementary electrical layer. In this architecture, 64 clusters are considered, each cluster being composed of 4 cores. The total die area for this architecture was reported to be $491mm^2$. In the following experiments, we evaluate the additional area cost required for initiating, terminating or propagating vertical connections for interconnecting each electrical layer to the optical layer. For this purpose, we successively consider 2 and 4 identical electrical layers. Obviously, the complexity of the optical interconnect depends on the possible communications schemes. Assuming that ORNoC manages only inter-layers communications, this complexity basically depends on the number of layers and the number of ONIs per layer [14]. In the experiments, from 2×2 to 8×8 ONIs per layer are considered: in the former case, each ONI is shared amongst 16 clusters while, in the latter case, each ONI is dedicated to a single cluster. As per optical technological constraints, we consider a maximum of 16 wavelengths per waveguide.



(a) 2 electrical layers



(b) 4 electrical layers

Fig. 5. Area size dedicated to ONI and area overhead for propagating vertical connections (in percentage of the initial electrical layer die size).

Figure 5(a) illustrates results for 3D architecture including 2 electrical layers. For 2×2 ONIs per layer, 2 waveguides allow interconnecting the 8 ONIs (considering only inter-layers communications). Since each ONI requires 16 CMOS drivers and 16 CMOS receivers per waveguide, we obtain $ONI_{area} = 11840\mu m^2$ which approximately represents 0.0025% of the initial electrical layer die size. Considering the 4 ONIs per layer, 0.01% extra area need to be allocated (as displayed on the figure). Since 2 electrical layers are considered, the same area is allocated to propagate vertical connections. For these scenarios, the area overhead introduced to provide dedicated vertical connections is thus reasonable compared to the initial die size (15.8% for the 8×8 ONIs scenario).

Figure 5(b) illustrates results obtained for 4 electrical layers scenario. Compared to the previous scenario, main differences are i) the larger area size overhead (from 0.08% for the 2×2 ONIs

scenario to 286% for the 8×8 ONIs scenario) and ii) the increased ratio between the area dedicated to ONI and the area dedicated to propagate TSV. The former difference is due to the increased complexity of ORNoC (which needs to manage more communication scenarios) and the latter difference comes from the increased number of electrical layers to be crossed by vertical connections.

VI. Conclusion

We propose a 3D architecture including ORNoC, a contention-free Optical NoC. The unmatched feature of ORNoC is that the same wavelength can be used to realize multiple communications on the same waveguide, at the same time, with no arbitration required. This contention free property requires dedicated vertical connections between each electrical layer and the optical layer. This strongly constrains the electrical layer layout and may results in a different layout for each electrical layer. We proposed layout guideline for keeping all the electrical layers identical, resulting in a cost efficient scalable 3D architecture. The area overhead required by the approach is evaluated through analytical formulation.

References

[1] R. G. Beausoleil, J. Ahn, N. Binkert, A. Davis, D. Fattal, M. Fiorentino, N. P. Jouppi, M. McLaren, C. M. Santori, R. S. Schreiber, S. M. Spillane, D. Vantrease, and Q. Xu. A Nanophotonic Interconnect for High-Performance Many-Core Computation. In *Proceedings of the 16^{th} IEEE Symposium on High Performance Interconnects*, pages 182–189, 2008.

[2] Mark J. Cianchetti, Joseph C. Kerekes, and David H. Albonesi. Phastlane: a Rapid Transit Optical Routing Network. In *Proceedings of the 36^{th} International Symposium on Computer Architecture*, ISCA, pages 441–450, 2009.

[3] Jason Cong and Yan Zhang. Thermal-driven multilevel routing for 3-D ICs. In *Proceedings of the Asia and South Pacific Design Automation Conference*, ASP-DAC, pages 121–126, 2005.

[4] Huaxi Gu, Jiang Xu, and Zheng Wang. A Novel Optical Mesh Network-on-Chip for Gigascale Systems-on-Chip. In *Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems*, pages 1728–1731, December 2008.

[5] Huaxi Gu, Jiang Xu, and Wei Zhang. A Low-Power fat Tree-Based Optical Network-on-Chip for Multiprocessor System-on-Chip. In *Proceedings of the conference on Design, Automation and Test in Europe*, DATE, pages 3–8, 2009.

[6] R. Ho, K.W. Mai, and M.A. Horowitz. The Future of Wires. *Proceedings of the IEEE*, 89(4):490–504, April 2001.

[7] International technology roadmap for semiconductors. http://www.itrs.net/.

[8] Ajay Joshi, Christopher Batten, Yong-Jin Kwon, Scott Beamer, Imran Shamim, Krste Asanovic, and Vladimir Stojanovic. Silicon-Photonic Clos Networks for Global on-Chip Communication. In *Proceedings of the 3^{rd} ACM/IEEE International Symposium on Networks-on-Chip*, NOCS, pages 124–133, 2009.

[9] Nevin Kirman, Meyrem Kirman, Rajeev K. Dokania, Jose F. Martinez, Alyssa B. Apsel, Matthew A. Watkins, and David H. Albonesi. Leveraging Optical Technology in Future Bus-based Chip Multiprocessors. In *Proceedings of the 39^{th} Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO, pages 492–503, 2006.

[10] M. J. Kobrinsky. On-Chip Optical Interconnects. *Intel Technology Journal*, 08(02):129–141, October 2004.

[11] S.J. Koester, G. Dehlinger, J.D. Schaub, J.O. Chu, Q.C. Ouyang, and A. Grill. Germanium-on-Insulator Photodetectors. In *2^{nd} IEEE International Conference on Group IV Photonics*, pages 171–173, 2005.

[12] George Kurian, Jason E. Miller, James Psota, Jonathan Eastep, Jifeng Liu, Jurgen Michel, Lionel C. Kimerling, and Anant Agarwal. ATAC: a 1000-Core Cache-Coherent Processor with on-Chip Optical Network. In *Proceedings of the 19^{th} international conference on Parallel Architectures and Compilation Techniques*, PACT, pages 477–488, 2010.

[13] Walid Lafi, Didier Lattard, and Ahmed Jerraya. An Efficient Hierarchical Router for Large 3D NoCs. In *21st IEEE International Symposium on Rapid System Prototyping*, RSP, Fairfax, VA, 2010.

[14] Sébastien Le Beux, Jelena Trajkovic, Ian O´ Connor, Gabriela Nicolescu, Guy Bois, and Pierre Paulin. Optical Ring Network-on-Chip (ORNoC): Architecture and Design Methodology. In *Proceedings of the conference on Design, Automation and Test in Europe*, DATE, 2011.

[15] Igor Loi, Federico Angiolini, and Luca Benini. Supporting Vertical Links for 3D Networks-on-Chip: Toward an Automated Design and Analysis Flow. In *Proceedings of the 2^{nd} international conference on Nano-Networks*, Nano-Net, pages 1–5, 2007.

[16] Yehia Massoud, Naomi Halas, and Peter Nordlander. Subwavelength Nanophotonics for Future Interconnects and Architectures. Invited talk, NRI SWAN Center, Rice University, 2008.

[17] D. Miller. Device Requirements for Optical Interconnects to Silicon Chips. *Proceedings of the IEEE*, 97(7):1166–1185, 2009.

[18] Jacob R. Minz, Somaskanda Thyagaraja, and Sung Kyu Lim. Optical Routing for 3D System-on-Package. In *Proceedings of the conference on Design, Automation and Test in Europe*, DATE, pages 337–338, 2006.

[19] David Navarro, Matthieu Briere, Ian O'Connor, Fabien Mieyeville, Frédéric Gaffiot, and Laurent Carrel. Quantitative Study of Area and Power Consumption Costs for 3 Gbits/s Optical Communications in a $0.13\mu m$ CMOS Circuit. In 20^{th} *Conference on Design of Circuits and Integrated Systems*, DCIS, Lisbon, Portugal, 2005.

[20] I. O'Connor, F. Mieyeville, F. Gaffiot, A. Scandurra, and G. Nicolescu. Reduction Methods for Adapting Optical Network on Chip Topologies to Specific Routing Applications. In *Proceedings of the Design of Circuits and Integrated Systems*, DCIS, November 2008.

[21] Ian O'Connor and Frédéric Gaffiot. On-Chip Optical Interconnect for Low-Power. In Enrico Macii, editor, *Ultra Low-Power Electronics and Design*, pages 21–39. Springer US, 2004.

[22] Yan Pan, J. Kim, and G. Memik. FlexiShare: Channel Sharing for an Energy-Efficient Nanophotonic Crossbar. In *Proceedings of the 16^{th} IEEE International Symposium on High Performance Computer Architecture*, HPCA, pages 1–12, January 2010.

[23] Yan Pan, Prabhat Kumar, John Kim, Gokhan Memik, Yu Zhang, and Alok Choudhary. Firefly: Illuminating Future Network-on-Chip with Nanophotonics. In *Proceedings of the 36^{th} annual International Symposium on Computer Architecture*, ISCA, pages 429–440, 2009.

[24] Sudeep Pasricha and Nikil Dutt. ORB: an on-Chip Optical Ring Bus Communication Architecture for Multi-Processor Systems-on-Chip. In *Proceedings of Asia and South Pacific Design Automation Conference*, ASP-DAC, pages 789–794, 2008.

[25] J. Psota, J. Miller, G. Kurian, H. Hoffman, N. Beckmann, J. Eastep, and A. Agarwal. ATAC: Improving Performance and Programmability With on-Chip Optical Networks. In *Proceedings of IEEE International Symposium on Circuits and Systems*, ISCAS, pages 3325–3328, 2010.

[26] Seiculescu, Ciprian and Murali, Srinivasan and Benini, Luca and De Micheli, Giovanni. Sunfloor 3D: a Tool for Networks on Chip Topology Synthesis for 3-D Systems on Chips. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29:1987–2000, December 2010.

[27] Assaf Shacham, Keren Bergman, and Luca P. Carloni. Photonic Networks-on-Chip for Future Generations of Chip Multiprocessors. *IEEE Transactions on Computers*, 57:1246–1260, September 2008.

[28] T. Spuesens, Liu Liu, T. de Vries, P.R. Romeo, P. Regreny, and D. Van Thourhout. Improved Design of an InP-Based Microdisk Laser Heterogeneously Integrated with SOI. In *Proceedings of the 6^{th} IEEE International Conference on Group IV Photonics*, GFP, pages 202–204, September 2009.

[29] Dana Vantrease, Robert Schreiber, Matteo Monchiero, Moray McLaren, Norman P. Jouppi, Marco Fiorentino, Al Davis, Nathan Binkert, Raymond G. Beausoleil, and Jung Ho Ahn. Corona: System Implications of Emerging Nanophotonic Technology. In *Proceedings of the 35^{th} Annual International Symposium on Computer Architecture*, ISCA, pages 153–164, 2008.

[30] L. Zhang, M. Yang, Y. Jiang, E. Regentova, and E. Lu. Generalized wavelength routed optical micronetwork in network-on-chip. In *Proceedings of the 18^{th} international conference on Parallel and Distributed Computing and Systems*, PDCS, pages 698–703, November 2006.

[31] Pingqiang Zhou, Ping-Hung Yuh, and Sachin S. Sapatnekar. Application-specific 3D Network-on-Chip design using simulated allocation. In *Proceedings of the Asia and South Pacific Design Automation Conference*, ASPDAC, pages 517–522, 2010.