Reduction methods for adapting optical network on chip topologies to specific routing applications

Ian O'Connor[1], Fabien Mieyeville[1], Frédéric Gaffiot[1], Alberto Scandurra[2], Gabriela Nicolescu[3]
ian.oconnor@ec-lyon.fr

[1]University of Lyon
Lyon Institute of Nanotechnology
Ecole Centrale de Lyon
36 avenue Guy de Collongue
F-69134 Ecully cedex, France

[2]STMicroelectronics
Stradale Primosole, 50
95123 Catania, Italy

[3]École Polytechnique de Montréal
Computer Science Department
Montréal,
QC, H3C 3J7, Canada

***Abstract*— Optical network on chip (ONoC) architectures are emerging as potential contenders to solve congestion and latency issues in future computing architectures. In this work, we examine how a scalable and fully connected ONoC topology can be reduced to fit specific connectivity requirements in MPSoCs and heterogeneous SoCs. Through such techniques, it is possible to reduce the number of required wavelengths and routing elements, thereby relaxing constraints on source wavelength accuracy and passive filter selectivity, and also alleviating power and area issues by reducing the number of active devices. Using this method, we show that it is possible to reduce the number of required wavelengths and routing elements by 38% and 72% respectively, when mapping a full 8×8 ONoC to an 8-node SSTNoC architecture.**

I. INTRODUCTION

The shift to very high performance distributed Multi-Processor Systems-on-Chip (MPSoC) as mainstream computing devices is the recognized route to address, in particular, power issues by reducing individual processor frequency while retaining the same overall computing power. This rationale answers the need for flexible and scalable computing platforms capable (i) of achieving future required application performance in terms of resolution (audio, video and computing) and CPU power / total MIPS (real-time encoding-decoding, data encryption-decryption), and (ii) of working with multiple standards and with constrained power, which are both particularly important for mobile applications.

However, the move to such architectures requires organized high-speed communication between processors and therefore has an impact on the interconnect structure. It clearly relies upon the existence of an extremely fast and flexible interconnect architecture, to such a point that the management of communication between processors will become key to successful development. Aggregated on-chip data transfer rates in MPSoC, such as the IBM Cell processor [1], is critical and is expected to reach over 100Tb/s in the coming decade. As such, interconnects will play a significant role for MPSoC design in order to support these high data rates.

[1] This work was partially funded by FP7-ICT-2007-1-216405 "WADIMOS"

At the architectural level, networks on chip (NoC) overcome the limitations of bus-based platforms by providing each IP block, interfaced towards the network, with one or more reconfigurable channels of high-speed communication. NoC architectures are based on multiple data links interconnected by routers implementing packet switching for resource multiplexing. At the physical communication level, it is increasingly recognized that electrical interconnect will be highly inefficient in NoCs due to increasing power and silicon real estate concerns. One of the main replacement technologies currently under development consists of using integrated optical interconnect. Besides a huge data rate, optical interconnects also allow for additional flexibility through the use of wavelength division multiplexing. Exploring this aspect is necessary since it is not clear that a *direct* (single-wavelength) replacement of electrical links between switchboxes in a NoC topology by optical interconnect will achieve a significant performance gain, since this would require conversion between optical and electrical domains *at each switchbox*. Instead, through a shift in the routing paradigm (where the address of the target is not contained in the data packet but rather in the wavelength of the optical signal), it is possible to exploit this additional flexibility to design more intelligent interconnect systems, such as passive, wavelength-reconfigurable optical networks on chip (ONoC).

Previous work on ONoC architectures has focused on regular structures offering total connectivity between IP blocks communicating through it, or balanced communication between groups of IP blocks of the same number. However, practical applications rarely call for such communication scenarios, and the application of the generic ONoC structure in this context leads to a wasteful use of resources (in terms of silicon area and power consumption), and to stringent constraints on the individual devices within the ONoC. There is thus a need for a means to adapt (or reduce) ONoC structures to real-world communication requirements, building on formal analyses of their specific properties and mathematical models such as that developed in [2].

In this paper, we present a method for reducing the generic ONoC structure according to a given connectivity matrix. Section II details the architecture and principle of operation of

the initial structure, while section III covers its limits. The reduction method itself is presented in section IV, and results are shown in systematic mapping scenarios, as well as in the context of an actual 8-node NoC topology.

II. SCALABLE ONOC ARCHITECTURE

An N×N ONoC, from a functional point of view, has the same behavior as an electrical N-port NoC: each initiator port (among N) can communicate simultaneously with one (or more, and possibly any number up to N) of N target ports. In this work, the quantity N represents the number of IP blocks to be connected through the communication structure; hence each IP block sends data through an initiator port and receives data through a target port. The ONoC is composed of a set of N transmitters and N receivers (one for each initiator port and target port respectively), and a scalable passive integrated photonic routing structure (which we call a λ-router). In this section, we will cover the principle of operation of this architecture and present results of physical and architectural evaluations from previous work.

A. *Principle of operation*

Figure 1 shows an example of an 8×8 ONoC architecture. In this representation, each initiator port I_i ($\forall i \in \{1,N\}$) consists of a network interface (NI) and transmitter; and each target port T_j ($\forall j \in \{1,N\}$) consists of a receiver and NI. Data is sent through the passive λ-router optically from each initiator to one or more targets by selecting a specific wavelength (for each initiator-target pair); in fact, only one physical path associated with a single wavelength exists between I_i and T_j. At any one time, a maximum of N connections can exist in the network if each transmitter is equipped with a single, tunable-wavelength source; and a maximum of N^2 connections can exist in the network if each transmitter is equipped with N single-wavelength sources.

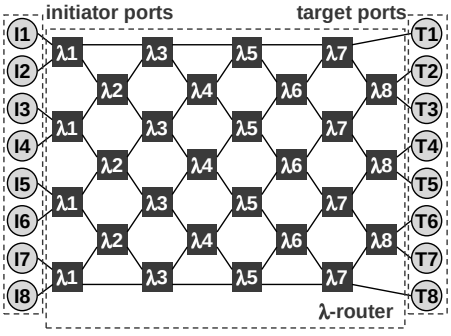


Figure 1 Full 8x8 ONoC topology schematic

In the figure, each box containing λ_x represents a passive photonic component called an "add-drop filter" which can realize the key functionality of selecting and redirecting a signal based on its wavelength. There are many ways of realizing a photonic add-drop filter. In our work, we consider the use of passive microdisk resonators as shown in Figure 2 [3], for which the overall footprint can be considered to be approximately $10\times10\mu m^2$. Resonance in the individual microdisks occurs whenever the wavelengths of the optical signal carried by the neighboring waveguide corresponds to an integer number of lobes around the circumference of the microdisk, i.e. when the energy is distributed in the disk in *whispering gallery* modes. Because of this, the resonant wavelengths of a microdisk depend, for a given technology (and material parameters), on the radius of the microdisk.

As shown in the figure, the switching direction depends on the input wavelength λ and its relation to the resonant wavelength of the add-drop filter:

- when $\lambda=\lambda_n$ (within a given tolerance range depending on the quality factor of the microdisk) the signal will couple into the microdisk and then couple out into the waveguide in the same plane as the input. This is the *straight*, or bar, state.
- when $\lambda\neq\lambda_n$ the signal will propagate along the same waveguide and outputs in a different plane to the input according to the geometry of the waveguide. This is the *diagonal*, or cross, state.

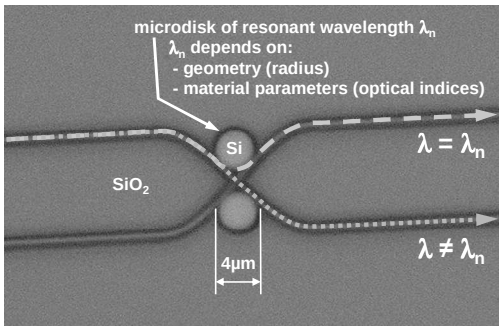


Figure 2 Si/SiO_2 microdisk-resonator based add-drop filter

When the WDM[2] technique is used, i.e. when multiple signals of various wavelengths are injected at the input (which is usually the case to increase the global throughput of the network), a cumulative state occurs, where individual signals simultaneously obey the routing characteristics of the add-drop filter according to their individual wavelengths. Because of this property and the fact that the four add-drop ports can be used simultaneously, a contention-free network can be built.

The overall passive λ-router network consists of N stages of alternately N/2 and (N/2)–1 add-drop filters (or, more generically, routing elements). Using microdisk resonators, the overall area required for the 8×8 passive network is around $3000\mu m^2$. The path followed by the optical signal in the overall network shown in Figure 1 depends only on the wavelength and is described in equation (1). As an example, if the block at initiator port I_3 is to communicate with the block at target port T_5, then I_3 must send data through the λ-router with wavelength λ_1. It is thus clear that each IP block can "reconfigure" its communication paths by using different wavelengths.

The matrix shown in equation (1) displays two interesting properties. Firstly, it is symmetrical around both diagonals. This means that the set of communication properties of the top half of the network is the flipped mirror image of that of the

[2] Wavelength Division Multiplexing

bottom half of the network; and that the return path for communication is exactly the same as the transmission path. The second noteworthy property is the existence of non-resonant wavelengths in certain communication paths (shown in bold in the matrix). While specific wavelengths have been assigned in the matrix to these communication paths, *any* wavelength (other than the wavelengths used by the other communication paths) can be used. This is the case since these communication paths do not actually pass through a routing element corresponding to the assigned wavelength at all – they cross the (N/2)-1 routing stages at the top or at the bottom of the network and thus only pass through a waveguide, rather than a resonant routing element. In the full ONoC, the unused wavelengths are assigned to these communication paths in order to exploit the resources – however during reduction, this can be exploited to reduce the number of wavelengths used.

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ T_7 \\ T_8 \end{bmatrix} = \begin{bmatrix} \lambda_4 & \lambda_5 & \lambda_3 & \lambda_6 & \lambda_2 & \lambda_7 & \lambda_1 & \boldsymbol{\lambda_8} \\ \lambda_5 & \lambda_6 & \lambda_4 & \lambda_7 & \lambda_3 & \lambda_8 & \boldsymbol{\lambda_2} & \lambda_1 \\ \lambda_3 & \lambda_4 & \lambda_2 & \lambda_5 & \lambda_1 & \boldsymbol{\lambda_6} & \lambda_8 & \lambda_7 \\ \lambda_6 & \lambda_7 & \lambda_5 & \lambda_8 & \boldsymbol{\lambda_4} & \lambda_1 & \lambda_3 & \lambda_2 \\ \lambda_2 & \lambda_3 & \lambda_1 & \boldsymbol{\lambda_4} & \lambda_8 & \lambda_5 & \lambda_7 & \lambda_6 \\ \lambda_7 & \lambda_8 & \boldsymbol{\lambda_6} & \lambda_1 & \lambda_5 & \lambda_2 & \lambda_4 & \lambda_3 \\ \lambda_1 & \boldsymbol{\lambda_2} & \lambda_8 & \lambda_3 & \lambda_7 & \lambda_4 & \lambda_6 & \lambda_5 \\ \boldsymbol{\lambda_8} & \lambda_1 & \lambda_7 & \lambda_2 & \lambda_6 & \lambda_3 & \lambda_5 & \lambda_4 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \\ I_8 \end{bmatrix} \quad (1)$$

B. *Evaluated performance metrics*

In prior work [4], a 4×4 passive λ-router was fabricated and measurements show that its operation agrees with the theory. Resonant wavelengths were measured between 1547-1583nm for Si/SiO_2 microdisks of radii from 1.0-2.5μm. The minimum free spectral range (FSR[3]) was measured to be 50nm, and quality factors around 500-800.

In parallel work, the design of a 16×16 ONoC virtual prototype was carried out at various abstraction levels using a top-down approach [5] from architecture to physical design, enabling an accurate estimation of various performance metrics. The source and detector characteristics were extracted from III-V device data, and transistor-level interface circuits sized with a 0.13 μm CMOS technology. In this context, the ONoC can achieve a data rate of up to 3.2Gb/s per port with a latency of 420ps and power consumption of 10mW per unidirectional link. The ONoC data rate is in fact limited by the interface circuits, mainly at the receiver. The SERDES circuits contribute greatly to power consumption at these frequencies.

More recently in [6], the impact of the low latency and absence of contention in the ONoC interconnect architecture was assessed for an 8-processor SoC running an MPEG-4 algorithm. When comparing a 100MHz ONoC against 200MHz STBus [7] and 2- and 5-CCL[4] crossbars, the ONoC demonstrated speedup factors of between 1.5 and 3.2, i.e. better performance, in terms of processing time, than any traditional electrical interconnect, even at half the operational frequency.

[3] FSR is defined as the difference between resonant wavelengths of a passive resonator. In the Si/SiO_2 microdisk resonators, FSR ≈ 50nm.

[4] Clock Cycle Latency

III. LIMITS TO THE FULL ONOC ARCHITECTURE

In this section, we cover the uses of ONoC in actual communication scenarios, and explain the rationale for reducing the size of the ONoC to adapt its functionality to specific connectivity requirements.

A. *Communication scenarios*

The optical waveguides within the ONoC are bidirectional. However, two-way communication between 2N IP blocks over a single ONoC is not feasible since this would require optical detectors and sources with identical wavelength selectivity to lie on the same waveguide with no interaction, which is clearly impossible. Additionally there can be no communication among IP blocks which have been assigned ports situated (physically) on the same side of the passive routing network. In fact there are two scenarios for the use of the N×N ONoC, both using the ONoC for communication in a single direction only:

- in the first scenario, shown in Figure 3(a) for 8 IP blocks, we consider that each IP block is assigned a pair of initiator / target ports. This leads to total connectivity between all N IP blocks, and to the non-use of wavelengths corresponding to communication paths I_i-T_j when i=j.
- in the second scenario, shown in Figure 3(b) for 8 IP blocks, we consider that two identical (N/2)×(N/2) ONoCs are used for request/response type communications between two sets of N/2 IP blocks. In this case, no communication is possible between IP blocks in the same set, but this scenario does lead to reduced requirements on the overall number of wavelengths and routing elements.

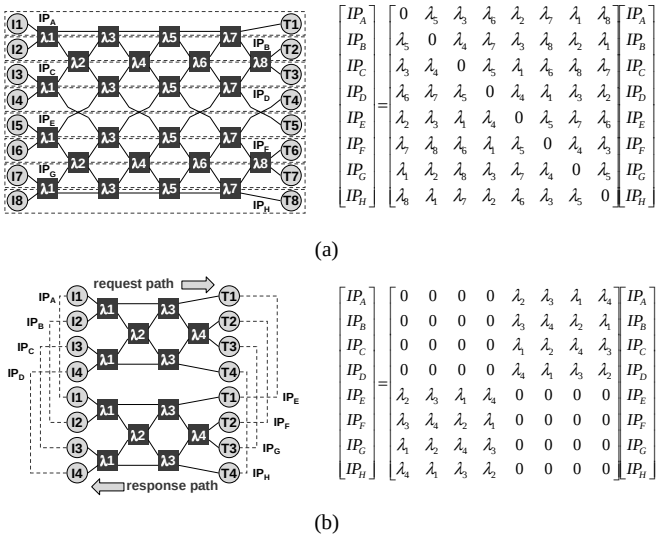


Figure 3 Communication scenarios and corresponding connectivity matrices for ONoC in 8-IP block scenarios (a) single 8×8 ONoC for total connectivity between 8 IP blocks (b) 2 4×4 ONoCs for request/response connectivity between 2 groups of 4 IP blocks

In Table I, a comparison is made between various performance metrics for each scenario. These represent extremes for (a) total connectivity and (b) balanced communication between groups of IP blocks of equal numbers. In practice, it is unlikely that the required system connectivity

will fall into either of these scenarios. However, the total connectivity scenario represents the default or reference scenario, while the grouped connectivity scenario makes clear that if total connectivity is not required in the system, significant reductions in complexity can be achieved.

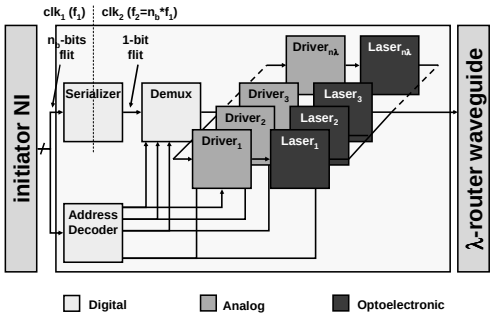
TABLE I. COMPARISON BETWEEN PERFORMANCE METRICS FOR TOTAL CONNECTIVITY AND GROUPED CONNECTIVITY ONOC SCENARIOS

	(a) Total connectivity	(b) Grouped connectivity
IP blocks	N	N
Connections	N(N-1)	$(N/2)^2$
Required wavelengths per IP block n_λ	N-1	N/2
Number of routing elements n_r	N(N-1)/2	N(N-1)/4

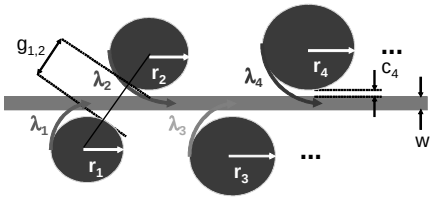
B. Rationale for reduction

The reductions mentioned in Table I are important since firstly, the number of routing elements n_r impacts directly on the overall size and complexity of the passive routing network. The size of the photonic communication layer is limited by the size of the CMOS chip. If several parallel λ-routers can fit into this area, then data rate could be increased (or power consumption reduced by running at a lower clock frequency).

Secondly, the required number of wavelengths n_λ per IP block will impact directly on the number of transmitters (and sources and wavelength multiplexers) and receivers (and wavelength demultiplexers and detectors) per IP block. The schematic of the transmitter structure and corresponding geometrical representation for the set of microdisk laser sources is shown in Figure 4(a) and Figure 4(b) respectively.



(a)



(b)

Figure 4 n_λ-laser source transmitter structure (a) schematic (b) corresponding geometrical representation for the set of microdisk laser sources

Since the laser source drivers are based on current modulation schemes, each source costs, in terms of static and dynamic power consumption, its bias current and modulation current respectively. As a consequence, the overall static and dynamic power consumption increases linearly with n_λ. In terms of the geometry and its impact on the size of the transmitter on the photonic layer, its area A_t can be expressed as

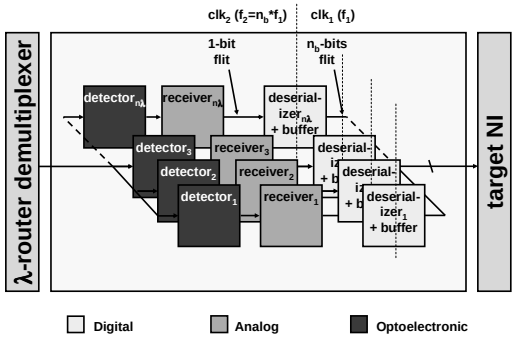
$$A_t = \left((n_\lambda - 1)\sqrt{(2\bar{r}+\bar{g})^2 - (w+2(\bar{c}+\bar{r}))^2} + 2\bar{r}\right)(2(2\bar{r}+\bar{c})+w) \quad (2)$$

where

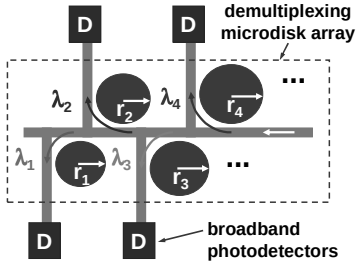
$$\bar{c} = \frac{\sum_{n=1}^{n_\lambda} c_n}{n_\lambda} \; ; \; \bar{r} = \frac{\sum_{n=1}^{n_\lambda} r_n}{n_\lambda} \; ; \; \bar{g} = \frac{\sum_{n=1}^{n_\lambda} g_{n,n+1}}{n_\lambda}$$

and c_n represents the nominal source$_n$-waveguide distance (between 0.4-0.6μm), r_n the nominal microdisk laser radius (between 1-10μm), $g_{n,n+1}$ is the minimum source-source spacing (typically 3μm), and w the waveguide width (under 1μm).

At the target end, each IP block requires a separate receiver path for each wavelength received, in order to identify the origin of each incoming data flit and also in order to be able to buffer flits incoming simultaneously from different initiator ports. The schematic of the receiver structure and corresponding geometrical representation for the set of microdisk demultiplexers and broadband photodetectors is shown in Figure 5(a) and Figure 5(b) respectively.



(a)



(b)

Figure 5 n_λ-demultiplexing receiver structure (a) schematic (b) corresponding geometrical representation for the set of microdisk demultiplexing elements and broadband photodetectors

Finally, as shown in Figure 6, since the maximum WDM window is approximately equivalent to the FSR of the microdisk resonators, a larger number of wavelengths will also lead to more stringent constraints on the selectivity (Q factor) of each resonator, and on the accuracy of the lithography techniques used to define the radius (and resonant wavelength)

of each passive microdisk in the λ-router. With current process technology characteristics, a maximum of around 16 distinguishable and stable wavelengths can be achieved.

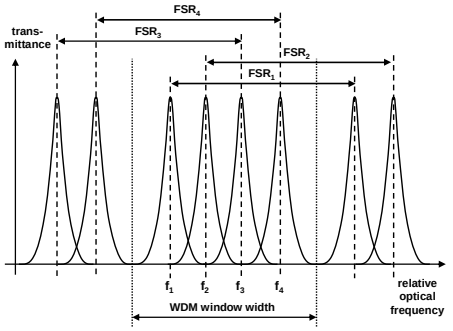


Figure 6 Relationship between microdisk resonator free spectral range and WDM window width

IV. Reduction Method

In this section, we develop a method to adapt from a full ONoC architecture to a specific interconnect requirement which is not covered by the total or grouped connectivity scenarios described in the previous section. This principally occurs when total connectivity is not required in the network and/or when the communicating IP blocks can be grouped into sets of unequal numbers.

A. Method theory

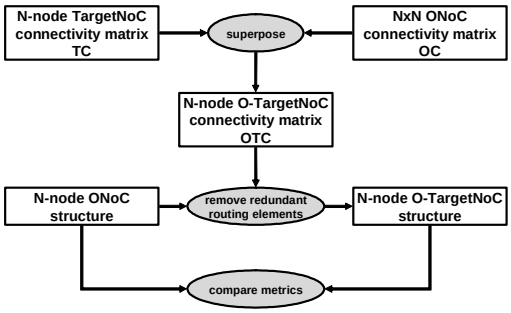


Figure 7 N-node ONoC reduction method overview

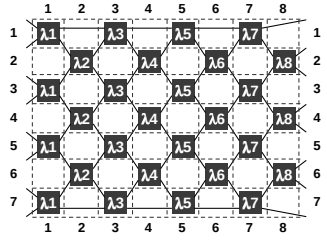
The input to our method, shown in Figure 7, is the N×N connectivity matrix TC of the target N-node interconnect structure. This is superposed with the full N×N (i.e. same matrix dimensions) ONoC connectivity matrix OC (the 8×8 variant of which was shown in equation (1)), such that the elements of the resulting connectivity matrix OTC are:

$$OTC_{ij} = TC_{ij} * OC_{ij} \quad \forall i,j \in \{1,N\} \qquad (3)$$

In order to reduce the full N×N ONoC structure, we now need to define which routing elements are redundant, i.e. which routing elements, required for resonance and "straight" connections through them, correspond to zero connection elements in OTC. To evaluate this, we rely upon the regular structure of the ONoC and establish the geometric position of each routing element corresponding to a connection within the ONoC structure. This can be done through the observation that:

- for the first connections in each row (i.e. column 1), the routing element making the straight connection is at position $x_1y_j=((N+j/2),(N-j/2))$ for even rows j, and at position $((N-(j-1))/2,(N+(j-1))/2)$ for odd rows j
- for the first $(N-j)$ connections in row j, the routing element making the straight connection is at position $x_iy_j=(x_1,y_j)+((i/2),(i/2))$ for even columns i, and at position $x_iy_j=(x_1,y_j)-(((i-1)/2),((i-1)/2))$ for odd columns i
- the matrix is symmetrical about the axis $(N,1)-(1,N)$ and the coordinates of the top left triangular matrix (using the above two rules) can be transposed directly to those of the bottom right triangular matrix

For the full 8×8 ONoC structure, these rules give the matrices shown in Figure 8 (general and explicit forms). It is worth noting that the x-coordinate of each identified routing element corresponds to the wavelength number used.



(a)

$$\begin{bmatrix} 4,4 & 5,5 & 3,3 & 6,6 & 2,2 & 7,7 & 1,1 & - \\ 5,3 & 6,4 & 4,2 & 7,5 & 3,1 & 8,6 & - & 1,1 \\ 3,5 & 4,6 & 2,4 & 5,7 & 1,3 & - & 8,6 & 7,7 \\ 6,2 & 7,3 & 5,1 & 8,4 & - & 1,3 & 3,1 & 2,2 \\ 2,6 & 3,7 & 1,5 & - & 8,4 & 5,7 & 7,5 & 6,6 \\ 7,1 & 8,2 & - & 1,5 & 5,1 & 2,4 & 4,2 & 3,3 \\ 1,7 & - & 8,2 & 3,7 & 7,3 & 4,6 & 6,4 & 5,5 \\ - & 1,7 & 7,1 & 2,6 & 6,2 & 3,5 & 5,3 & 4,4 \end{bmatrix}$$

(b)

$$\begin{bmatrix} \frac{N}{2},\frac{N}{2} & \frac{N+2}{2},\frac{N+2}{2} & \frac{N-2}{2},\frac{N-2}{2} & \frac{N+4}{2},\frac{N+4}{2} & \frac{N-4}{2},\frac{N-4}{2} & \frac{N+6}{2},\frac{N+6}{2} & \frac{N-6}{2},\frac{N-6}{2} & - \\ \frac{N+2}{2},\frac{N-2}{2} & \frac{N+4}{2},\frac{N}{2} & \frac{N}{2},\frac{N-4}{2} & \frac{N+6}{2},\frac{N+2}{2} & \frac{N-2}{2},\frac{N-6}{2} & \frac{N+8}{2},\frac{N+4}{2} & - & \frac{N-6}{2},\frac{N-6}{2} \\ \frac{N-2}{2},\frac{N+2}{2} & \frac{N}{2},\frac{N+4}{2} & \frac{N-4}{2},\frac{N}{2} & \frac{N+2}{2},\frac{N+6}{2} & \frac{N-6}{2},\frac{N-2}{2} & - & \frac{N+8}{2},\frac{N+4}{2} & \frac{N+6}{2},\frac{N+6}{2} \\ \frac{N+4}{2},\frac{N-4}{2} & \frac{N+6}{2},\frac{N-2}{2} & \frac{N+2}{2},\frac{N-6}{2} & \frac{N+8}{2},\frac{N}{2} & - & \frac{N-6}{2},\frac{N-2}{2} & \frac{N-2}{2},\frac{N-6}{2} & \frac{N-4}{2},\frac{N-4}{2} \\ \frac{N-4}{2},\frac{N+4}{2} & \frac{N-2}{2},\frac{N+6}{2} & \frac{N-6}{2},\frac{N+2}{2} & - & \frac{N+8}{2},\frac{N}{2} & \frac{N+2}{2},\frac{N+6}{2} & \frac{N+6}{2},\frac{N+2}{2} & \frac{N+4}{2},\frac{N+4}{2} \\ \frac{N+6}{2},\frac{N-6}{2} & \frac{N+8}{2},\frac{N-4}{2} & - & \frac{N-6}{2},\frac{N+2}{2} & \frac{N+2}{2},\frac{N-6}{2} & \frac{N-4}{2},\frac{N}{2} & \frac{N}{2},\frac{N-4}{2} & \frac{N-2}{2},\frac{N-2}{2} \\ \frac{N-6}{2},\frac{N+6}{2} & - & \frac{N+8}{2},\frac{N-4}{2} & \frac{N-2}{2},\frac{N+6}{2} & \frac{N+6}{2},\frac{N-2}{2} & \frac{N}{2},\frac{N+4}{2} & \frac{N+4}{2},\frac{N}{2} & \frac{N+2}{2},\frac{N+2}{2} \\ - & \frac{N-6}{2},\frac{N+6}{2} & \frac{N+6}{2},\frac{N-6}{2} & \frac{N-4}{2},\frac{N+4}{2} & \frac{N+4}{2},\frac{N-4}{2} & \frac{N-2}{2},\frac{N+2}{2} & \frac{N+2}{2},\frac{N-2}{2} & \frac{N}{2},\frac{N}{2} \end{bmatrix}$$

(c)

Figure 8 8×8 ONoC identification of resonant routing elements for each connection (a) division of ONoC geometry into indexed matrix elements (b) explicit matrix showing position of resonant routing elements (c) general form

The final step in the reduction method is thus to superpose this matrix (Figure 8(b)) onto the matrix OTC, and retain in the final structure only those elements identified in the final matrix. From there it is a simple matter to extract the salient performance metrics (number of wavelengths, number of routing elements).

B. Method test

We tested this method by applying it to the reduction of a totally connected ONoC (shown in Figure 3(a)) when the required connectivity matrix displays group connectivity characteristics between unbalanced sets of initiator IP blocks and target IP blocks (e.g. 6-2, 5-3). All possible combinations of target IP blocks were tested, and the results for the 6-2 case are shown in Figure 9. It is clear from this graphs that (i) the

assignment of ports to IP blocks is critical to achieving significant reduction figures and is thus a vector for optimization, and (ii) significant reduction is only achievable when some form of symmetry (around the diagonal) is present in the required connectivity matrix. The symmetric cases {1,6} {2,5} and {3,4} demonstrate a 25% reduction in the number of required wavelengths, and a 58% reduction in the number of routing elements. For the 5-3 group connectivity case, a maximum reduction of 13% in the number of required wavelengths can be achieved with a 28% reduction in the number of routing elements.

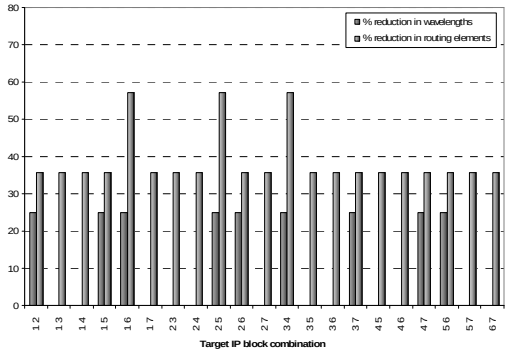
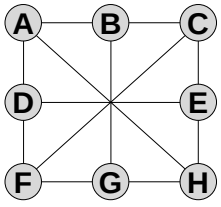


Figure 9 Extracted performance metrics for 8×8 ONoC structures reduced from total connectivity to unbalanced group connectivity (6 initiators, 2 targets)

C. Application to SST-NoC topology

We applied this method to a specific routing topology, called SSTNoC [8]. The 8-node variant is shown in Figure 10(a), with its corresponding extracted connectivity matrix shown in Figure 10(b).



(a)

$$\begin{bmatrix} A_T \\ B_T \\ C_T \\ D_T \\ E_T \\ F_T \\ G_T \\ H_T \end{bmatrix} = \begin{bmatrix} 0&1&0&1&0&0&0&1 \\ 1&0&1&0&0&0&1&0 \\ 0&1&0&0&1&1&0&0 \\ 1&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&1 \\ 0&0&1&1&0&0&1&0 \\ 0&1&0&0&0&1&0&1 \\ 1&0&0&0&1&0&1&0 \end{bmatrix} \begin{bmatrix} A_I \\ B_I \\ C_I \\ D_I \\ E_I \\ F_I \\ G_I \\ H_I \end{bmatrix}$$

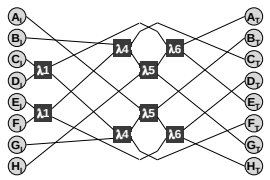
(b)

Figure 10 (a) 8-node SST-NoC topology (b) Corresponding connectivity matrix

As can be seen from Figure 11(a), the matrix is symmetrical and many of the connections lie on the non-resonant diagonal axis. This means that wavelengths only present on this axis are not required and can be eliminated. The final O-SSTNoC structure is shown in Figure 11(b) and has been verified against the required functionality. Table II shows a summary of the performance metrics compared to the reference total connectivity structure. The number of required wavelengths and routing elements are reduced by, respectively, 38% and 72%. By using the equations developed in section III.B, we can estimate the overall reduction in power at 63%, and in silicon surface (combining both transistor and photonic layers) at 66%.

$$\begin{bmatrix} A_T \\ B_T \\ C_T \\ D_T \\ E_T \\ F_T \\ G_T \\ H_T \end{bmatrix} = \begin{bmatrix} 0&\lambda_5&0&\lambda_6&0&0&0&\lambda_8 \\ \lambda_5&0&\lambda_4&0&0&0&\lambda_2&0 \\ 0&\lambda_4&0&0&\lambda_1&\lambda_6&0&0 \\ \lambda_6&0&0&0&\lambda_4&\lambda_1&0&0 \\ 0&0&\lambda_1&\lambda_4&0&0&0&\lambda_6 \\ 0&0&\lambda_6&\lambda_1&0&0&\lambda_4&0 \\ 0&\lambda_2&0&0&0&\lambda_4&0&\lambda_5 \\ \lambda_8&0&0&0&\lambda_6&0&\lambda_5&0 \end{bmatrix} \begin{bmatrix} A_I \\ B_I \\ C_I \\ D_I \\ E_I \\ F_I \\ G_I \\ H_I \end{bmatrix}$$

(a)



(b)

Figure 11 (a) 8-node O-SSTNoC connectivity matrix (b) Corresponding structure

TABLE II. COMPARISON BETWEEN PERFORMANCE METRICS FOR 8×8 TOTAL CONNECTIVITY ONOC AND O-SSTNOC

	(a) Total connectivity	(b) O-SSTNoC
IP blocks	8	8
Connections per IP block	8	3
Total number of required wavelengths	8	5
Wavelengths per IP block	8	3
Total number of routing elements	28	8

V. CONCLUSIONS

In this work, we have described a method to reduce scalable and fully connected N×N ONoC topologies in order to fit specific connectivity requirements. We have covered the main issues with various scenarios for fully connected topologies (constraints on source wavelength accuracy and passive filter selectivity depending on the number of required wavelengths, and power and area issues depending on the number of active and passive devices), and shown how it is possible to exploit the regular properties of ONoC topologies to reduce the number of required wavelengths and routing elements. We have applied this method to map a full 8×8 ONoC to an 8-node SSTNoC architecture, and demonstrated a 38% and 72% reduction in the number of required wavelengths and routing elements, respectively.

REFERENCES

[1] M. Gschwind, "Chip multiprocessing and the Cell broadband engine," *ACM Int. Conf. Computing Frontiers*, Ischia, Italy, May 2006

[2] L. Zhang *et al.*, "Generalized Wavelength Routed Optical Micronetwork in Network-on-chip", Proc. 18th Int. Conf. Parallel and Distributed Computing and Systems (IASTED), pp. 698-703, Dallas, TX, Nov. 2006

[3] A. Kazmierczak *et al.*, "Design, simulation, and characterization of a passive optical add-drop filter in silicon-on-insulator technology," *IEEE Photon. Technol. Lett.*, vol. 17, no. 7, pp. 1447-9, July 2005

[4] A. Kazmierczak *et al.*, "Design and characterisation of optical networks on chip," *Proc. Eur. Conf. Integrated Optics (ECIO)*, Grenoble, France, April 2005

[5] M. Briere *et al.*, "Heterogeneous modelling of an Optical Network-on-Chip with SystemC," *Proc. Rapid System Prototyping Workshop (RSP)*, June 2005

[6] M. Brière *et al.*, "System Level Assessment of an Optical NoC in an MPSoC Platform," *Proc. Design Automation and Test in Europe (DATE)*, Nice, France, April 2007

[7] A. Bona, V. Zaccaria, and R. Zafalon, "System level power modeling and simulation of high-end industrial network-on-chip," *Proc. Design, Automation and Test in Europe (DATE) Conference and Exhibition*, Munich, Germany, February 2004

[8] M. Coppola, "iNoC: An innovative EDA flow for on-chip communication infrastructure," *7th Int. Forum on Application Specific Multi-Processor SoC*, Hyogo, Japan, June 2007