

PSION 2: Optimizing Physical Layout of Wavelength-Routed ONoCs for Laser Power Reduction

Alexandre Truppel
Technical University of Munich
Munich, Germany
alex.truppel@tum.de

Tsun-Ming Tseng
Technical University of Munich
Munich, Germany
tsun-ming.tseng@tum.de

Ulf Schlichtmann
Technical University of Munich
Munich, Germany
ulf.schlichtmann@tum.de

ABSTRACT

Optical Networks-on-Chip (ONoCs) are becoming increasingly attractive for intra-chip communications due to their low power-per-bit requirements and high bandwidth. Wavelength-Routed ONoCs (WRONoCs), a subtype of ONoCs, further reduce network latency. Recently, tools to design WRONoCs have been developed, but these tools are still incomplete as they do not yet consider key design aspects such as the type of laser source used and the impact of the laser Power Distribution Network (PDN) on the laser power consumption. In this work we propose the first design automation tool to combine awareness of both on-chip and off-chip lasers with optimization of both the logical topology and the physical layout of WRONoCs for application-specific designs. Compared to previous works, the incorporation of the type of laser and the PDN into the optimization process combined with a new Generic Routing Unit (GRU) placement method leads to a laser power reduction of up to 20%.

ACM Reference Format:

Alexandre Truppel, Tsun-Ming Tseng, and Ulf Schlichtmann. 2020. PSION 2: Optimizing Physical Layout of Wavelength-Routed ONoCs for Laser Power Reduction. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD '20)*, November 2–5, 2020, Virtual Event, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3400302.3415655>

1 INTRODUCTION

With the development of ever more complex System-on-Chip (SoC) designs, Optical Networks-on-Chip (ONoCs) have become increasingly attractive for intra-chip communications. ONoCs are expected to consume less power per bit while providing lower latency and higher bandwidth than conventional Electrical Networks-on-Chip [8]. Passive ONoCs, also called Wavelength-Routed ONoCs (WRONoCs), further decrease signal delay by passively routing optical signals based on their wavelength and eliminating contention between signals [3, 10].

Conceptually, a WRONoC is composed of the following top-level elements, in order: *laser source(s)*, a *Power Distribution Network* (when necessary), *modulator arrays*, a WRONoC *router*, and *demodulator arrays*. A WRONoC node exists for each SoC element connected by the WRONoC, and each WRONoC node consists of

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

ICCAD 2020, DOI: [10.1145/3400302.3415655](https://doi.org/10.1145/3400302.3415655)

<https://ieeexplore.ieee.org/document/9256654>

<https://dl.acm.org/doi/10.1145/3400302.3415655>

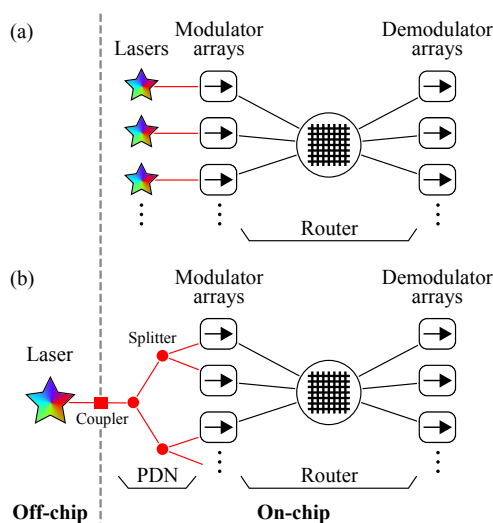


Figure 1: Conceptual diagram of the complete WRONoC optical architecture for (a) on-chip laser sources and (b) off-chip laser sources. All on-chip elements are located on the optical layer of the die.

zero or more modulator and demodulator arrays¹. The WRONoC router itself contains *waveguides* and *optical routing elements* to perform wavelength-based signal routing [12]. Each modulator array must be connected to a source of laser power, which can be provided either by *on-chip* or *off-chip* laser sources. In either case, the connection between the source and the modulator array is also done through waveguides. For on-chip laser sources this connection is trivial since multiple sources are used and each source is placed close to the modulator array it powers. However, for off-chip sources the power is first transferred to the die by couplers on the edge of the die. It must then be routed through the optical layer to each modulator array, which requires a *Power Distribution Network* (PDN). This network, composed of waveguides and *laser power splitters*, is placed on the same layer as the WRONoC router and is structured as a binary tree. Figure 1 shows a conceptual diagram of the complete WRONoC optical architecture for (a) on-chip sources and (b) off-chip sources.

The Electronic Design Automation field has seen a growing number of tools to design WRONoCs. The current state-of-the-art toolset includes methods to optimize WRONoC logical topologies

¹For example, if an SoC element receives information from the WRONoC but does not send information through it, the corresponding WRONoC node must have at least one demodulator array but no modulator arrays for it are required.

for incomplete communication matrices [1, 7] and to perform physical design (placing and routing) of those topologies [2, 4–6, 14, 15]. More recently, a new design methodology has been proposed that seeks to combine logical topology generation and physical design optimization into one optimization step for any logical topology and communication matrix [11, 12]. This methodology has been proven to outperform other tools created to achieve the same goals, and so it is considered the state-of-the-art in this work.

However, so far these general tools seek to minimize optical insertion loss values at the modulators, that is, the insertion loss created by the router only. Thus, they ignore the details of how the connection between the laser sources and the router influences total laser power (for example, through a PDN when off-chip laser sources are used) [13]. This is a sub-optimal procedure since it has been shown that the PDN can cause a sharp increase in the amount of laser power required once it is added to the WRONoC design [8]. Some other works have tackled the laser power supply aspect of the WRONoC, but have done it manually or with specific algorithms for select designs only [8, 9].

In this work we propose a new design tool containing the following major contributions: (1) we fully model the effects of the different types of on-chip and off-chip laser sources on the total laser power, including using accurate optimization functions and modeling the PDN, when off-chip laser sources are used; (2) we develop a new placement and routing method which takes advantage of (1) to further decrease total laser power compared to the state-of-the-art. In doing so, we create the first design automation tool that considers the complete WRONoC design from laser source to demodulator for any logical topology and communication matrix.

In Section 2 we highlight key WRONoC design aspects and explain the motivation behind this work. The work itself is detailed further in Section 3 and Section 4. In Section 5 we compare our work against the state-of-the-art design tool PSION+. Finally, conclusions are drawn in Section 6.

2 WRONOC DESIGN

2.1 Generic Routing Unit

Micro Ring Resonators (MRRs) are the basic optical routing elements of WRONoCs. MRRs route optical signals between waveguides based on wavelength, as shown in Figure 2(a). A Generic Routing Unit (GRU) is a structured collection of MRRs and short waveguide fragments that connects to up to four external waveguide sections and performs routing of optical signals between those sections. A WRONoC router has multiple interconnected GRUs, with each one helping to perform part of the total routing requirements of the router.

Each GRU in a WRONoC router can be configured at design time with its own internal structure from a set of possible structures [12]. Figure 2(b) shows examples of possible GRU internal structures. The selection of the best internal structure for each GRU in the router is an important goal of the WRONoC design and optimization process.

2.2 Insertion loss

The insertion loss through a path in the WRONoC describes the amount of power attenuation (in dB) that an optical signal suffers

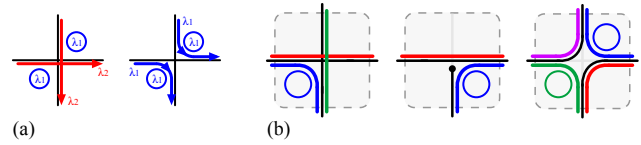


Figure 2: (a) Example influence of an MRR next to a waveguide crossing on optical signal path depending on MRR wavelength and optical signal wavelength. (b) Examples of possible GRU internal structures. Color indicates wavelength.

when going through that path. All WRONoC elements cause insertion loss: *MRRs* and optical power *splitters* (used in PDNs, see Section 2.4) cause various types of loss, *waveguides* cause propagation loss proportional to their length and bending loss linked to their bends, crossings between waveguides (inside GRUs or otherwise) cause crossing loss, and both *modulators* and *demodulators* are also sources of loss. The sum of all the losses on a path is the (total) insertion loss of that path.

Commonly, insertion loss is calculated on a path from some location in the WRONoC (for example, a modulator array or a laser source) to a demodulator array. In this case it can be used to calculate the minimum amount of optical power required at that location for optical signals going through that path to still be detected correctly at the optical-electrical interface of the demodulator array (see Section 2.3). Thus, the insertion loss value becomes analogous to an optical power amount and can be interpreted as such.

2.3 Laser sources

A laser source is an element that outputs a certain amount of optical power for a certain set of predefined wavelengths for which it is configured. Here we categorize laser sources into two fundamental types based on their location:

- **On-chip:** These sources are placed on the die next to the modulator arrays they power. The connection between the source and the modulators powered by it is trivial to design and places no burden on the design of the WRONoC itself.
- **Off-chip:** The source is placed outside of the chip it powers. The laser power is transmitted into the chip through optical couplers. Once inside the chip, the power must be routed to each modulator array of each node of the WRONoC on the same optical layer as the WRONoC router itself. This is done by a PDN (explained in Section 2.4).

We also categorize laser sources into two other conceptual types, orthogonal to *On-chip* and *Off-chip*:

- **Type X:** The optical power output can be controlled (during design/fabrication) *per* wavelength emitted by the source.
- **Type Y:** The source emits the same amount of optical power on *all* wavelengths, and only this one value can be controlled.

If a laser source must supply wl wavelengths and each wavelength λ suffers a total insertion loss (in dB) of il_{λ}^L , then the total

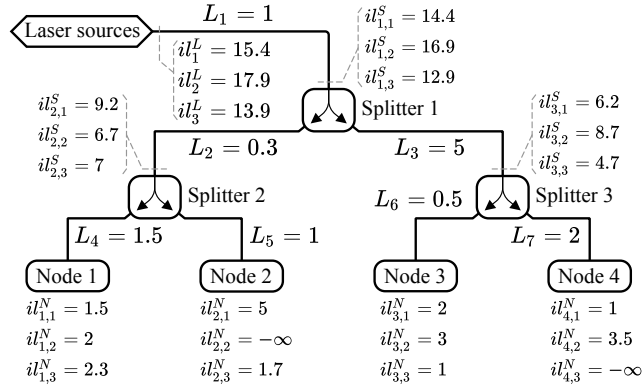


Figure 3: Example of a PDN distributing three wavelengths to the four nodes of a WRONoC. Each modulator array must receive enough power to overcome the insertion loss values caused by the WRONoC router, $il_{n,\lambda}^N$. Splitters have a 50% splitting ratio.

optical power (in Watt) required for each type of laser source is:

$$P_{optical}^X = \sum_{\lambda=1}^{wl} \frac{1}{K} 10^{\frac{il_{\lambda}^L + S}{10}} \quad (1)$$

$$P_{optical}^Y = wl * \frac{1}{K} 10^{\frac{\max_{\lambda=1}^{wl} \{il_{\lambda}^L\} + S}{10}} \quad (2)$$

where K is a combination of technology and fabrication parameters and S is the sensitivity of the demodulator combined with the desired Bit Error Rate (BER)² [14]. Note that for *type Y* the max function is used since the source must output enough power on each one of the wavelengths. This leads to power loss as inevitably some wavelengths will receive more power than the bare minimum dictated by their insertion loss values.

2.4 Power Distribution Network

For off-chip laser sources, a PDN is required. Its goal is to receive the optical power emitted by the sources and distribute enough power to each modulator in the modulator's wavelength. Any excess power is considered wasted. PDNs are implemented as a binary tree [8, 9] consisting of optical power splitters at the branching nodes and of WRONoC nodes at the leaf nodes. A conceptual example of a PDN powering four WRONoC nodes is shown in Figure 3.

To better explain the PDN, we introduce the following notation used in Figure 3 from the bottom to the top:

- $il_{n,\lambda}^N$ – Insertion loss on node n for wavelength λ . This includes the insertion loss caused by the modulator itself, plus the insertion loss caused by the router (from the modulator to the demodulator, excluding both), plus the insertion loss caused by the demodulator. Router insertion loss depends on the exact design of the router whereas (de)modulator losses are only technology dependent constants.

- L_e – Insertion loss caused by PDN waveguide (edge of the PDN tree) e of the PDN. This loss is created by a combination of propagation, bending and crossing losses in the waveguide and depends on the exact routing of the waveguide through the optical layer.
- $il_{s,\lambda}^S$ – Insertion loss up to the input of splitter s for wavelength λ .
- il_{λ}^L – Insertion loss up to the laser source for wavelength λ .

To show how il_{λ}^L is calculated, we provide an example explained from the bottom-up. At the *left output* of splitter 2, power must be available to overcome $il_{1,\lambda}^N + L_4$ of insertion loss on each wavelength λ . Similarly, at the *right output* of splitter 2, power must be available to overcome $il_{2,\lambda}^N + L_5 \forall \lambda$. Ideally, splitter 2 would receive the exact amount of total power required for both branches and then be constructed with the exact splitting ratio between outputs to satisfy both power requirements. This would minimize power losses in the PDN. However, unbalanced power ratios for splitters come with fabrication problems due to unavoidable process variations [8]. Thus, we consider that all splitters have 50% splitting ratios. We reach the following value for $il_{2,\lambda}^S \forall \lambda$:

$$il_{2,\lambda}^S = \max\{il_{1,\lambda}^N + L_4, il_{2,\lambda}^N + L_5\} + 10 \log_{10}(2) + L^S \quad \forall \lambda \quad (3)$$

L^S is the extra insertion loss caused by the splitter itself ($L^S = 0.2$ dB in this example) and the max operation guarantees each output of the splitter will have enough power for one of the child branches. Naturally, the splitter must receive enough power for both child branches together, which is double the power given by the max operation. We double the input power requirement from one child branch to two by adding the constant term $10 \log_{10}(2) \approx 3$ dB [8].

To calculate $il_{1,\lambda}^S \forall \lambda$ (one level above in the tree), we proceed identically:

$$il_{1,\lambda}^S = \max\{il_{2,\lambda}^S + L_2, il_{3,\lambda}^S + L_3\} + 10 \log_{10}(2) + L^S \quad \forall \lambda \quad (4)$$

Finally, calculating $il_{\lambda}^L \forall \lambda$ is straightforward:

$$il_{\lambda}^L = il_{1,\lambda}^S + L_1 \quad \forall \lambda \quad (5)$$

Note that if node n does not send an optical signal on wavelength λ then it does not require power on that wavelength and thus $il_{n,\lambda}^N = -\infty$ dB. Equations (1)–(3) are still valid in this case. For example, in Figure 3, the required power at the input of splitter 3 for wavelength 3 is:

$$\begin{aligned} il_{3,3}^S &= \max\{il_{3,3}^N + L_6, il_{4,3}^N + L_7\} + 10 \log_{10}(2) + L^S \\ &= \max\{1 + 0.5, -\infty + 2\} + 10 \log_{10}(2) + 0.2 \\ &= 1.5 + 10 \log_{10}(2) + 0.2 \approx 4.7 \text{ dB} \end{aligned} \quad (6)$$

Likewise, if $il_{s,\lambda}^S = -\infty$ dB for some λ then the laser source does not need to output power on wavelength λ .

When on-chip laser sources are used there is no need for a PDN because a laser source exists for each node, each source is next to the node it powers and is connected directly to the node's modulators. Thus, the power requirements of the laser sources depend not only on the wavelength λ , but also on the node n they belong to, which

²Details on how K and S are calculated are left out of this work because these constant parameters are irrelevant for the purpose of minimizing $P_{optical}$ and can thus be ignored.

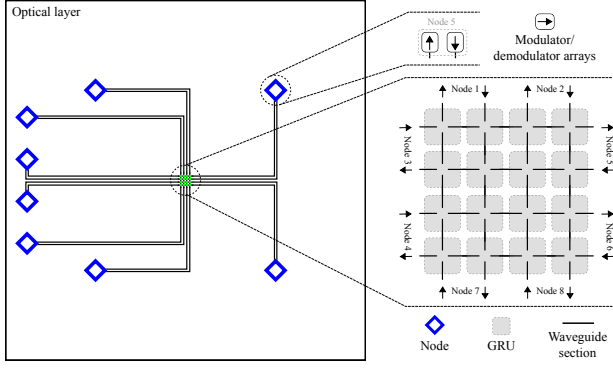


Figure 4: Example of a physical layout template, a grid template with 16 GRUs, connecting 8 nodes on a 9 mm die.

means il_{λ}^L must be replaced by $il_{n,\lambda}^L$. Finally, because there is no insertion loss between the source and the modulators, we have simply $il_{n,\lambda}^L = il_{n,\lambda}^N$.

2.5 State-of-the-art

PSION+ [12] is a state-of-the-art tool for WRONoC design based on the idea of a physical layout template. A physical layout template is a collection of WRONoC design elements placed in specific locations on the optical layer. These elements are modulator and demodulator arrays, waveguide sections and GRUs. Figure 4 is an example of a physical layout template as it contains the aforementioned elements in their physical locations. The existence and locations of the modulator and demodulator arrays in the template are defined by the SoC nodes in the electronic layer. The amount of GRUs, their locations and interconnections, and the routing of the waveguide sections are decided by the WRONoC designer.

The designer must specify this template along with a communication matrix between the nodes of the WRONoC. The communication matrix is a binary matrix $C_{i,j}$ with size $N \times N$ where N is the number of nodes in the WRONoC and $C_{i,j} = 1$ if and only if node i sends an optical signal to node j . Those two inputs are used to build a Mixed Integer Programming (MIP) model and the model is solved to output the optimal design considering the specified optimization function. The output design is described by the internal configuration of all GRUs and the wavelengths of optical signals sent between each pair of nodes in the communication matrix.

2.6 Motivation

The optimization framework of PSION+ has two major shortcomings. Firstly, it optimizes insertion loss by minimizing the maximum insertion loss ($\max\{il_{n,\lambda}^N\}$) among all optical signals in the router. This optimization function is inadequate and leads to sub-optimal solutions since it does not take into account the different types of lasers nor the PDN when one exists. Here we present the first work to optimize laser power considering all types of lasers with and without a PDN.

Secondly, it considers that templates are static during optimization, i.e. GRU locations and waveguide paths are fixed. However, this need not be the case. In this work we propose to optimally

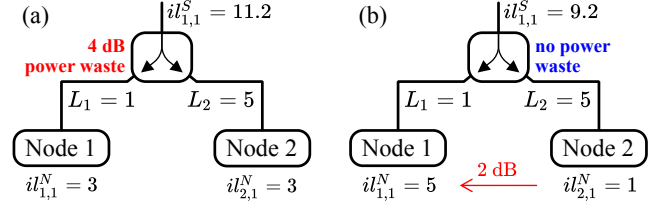


Figure 5: Example of how changing the design of the router may lower power requirements. (a) 50% splitting ratio in the splitter causes Node 1 to receive 4dB extra due to asymmetries in the PDN. (b) Changes in the design of the router can shift insertion loss amounts between nodes to compensate for the PDN and thus result in lower total insertion loss. This compensation is only possible if the PDN is considered in the optimization.

place GRUs (defined in the template given as input) for power optimization using a GRU placement and waveguide routing strategy for two reasons. Firstly, this reduces total propagation loss in the router, which leads to lower power consumption. Secondly, this helps rebalance the router when a PDN is required. To understand why this also reduces power consumption, we use the example in Figure 5. All max operations, both in the PDN and in $P_{optical}^Y$ (from Equation 2), have the potential to lead to power losses if the operands of max have very disparate values – in Figure 5(a), the max operation in the splitter causes 4dB of extra power to be sent to Node 1. By allowing GRUs to be moved, the solver can relocate insertion loss loads between nodes – in Figure 5(b), both branches of the splitter now require the same amount of power, which leads to no power waste and lower power requirements. This can lead to substantial reductions in laser power.

3 LASER POWER OPTIMIZATION

In this section we will explain the proposed new optimization functions required to accurately model on-chip and off-chip lasers of both types in our MIP model. Here we assume for simplicity that each node contains only one modulator array and, if on-chip sources are used, that only one source (of multiple wavelengths) exists per node. A more general case is straightforward to develop based on this work.

3.1 Calculating $il_{n,\lambda}^L$ (off-chip) and $il_{n,\lambda}^L$ (on-chip)

We adopt the MIP model from the state of the art [12] as a foundation. Variables defined in it – which detail the insertion loss through the router of each optical signal sent by each node – are used to set the value of the $il_{n,\lambda}^N$ variables. Specifically, we use constraints to set $il_{n,\lambda}^N$ equal to the insertion loss of the optical signal sent by node n on wavelength λ (if no optical signal is sent by node n on wavelength λ then $il_{n,\lambda}^N = -\infty$ dB).

We then create variables $il_{n,\lambda}^L$ for on-chip designs or variables $il_{s,\lambda}^L$ and $il_{s,\lambda}^S$ for off-chip designs, and set their values with constraints that mirror the PDN calculations explained in Section 2.4.

3.2 Minimizing $P_{optical}$

With the values of il_{λ}^L and $il_{n,\lambda}^L$ established, we now define optimization functions to minimize $P_{optical}$ according to equations (1) and (2) for on-chip and off-chip lasers which are more accurate than the state-of-the-art (minimization of $\max\{il_{n,\lambda}^N\}$).

The challenge with minimizing $P_{optical}$ is to synchronize the measures of power consumption (in Watt) and insertion loss (in dB). The relationship between both is non-linear and thus strategies to convert $P_{optical}$ into a linear function of variables il_{λ}^L and $il_{n,\lambda}^L$ must be adopted. The essence of the employed strategy is to do the following approximation:

$$\begin{aligned} \min \sum x_i &\approx \min \prod x_i \\ \Leftrightarrow \min \log_{10}(\prod x_i) &= \min \sum y_i \end{aligned} \quad (7)$$

where x_i are measures of power consumption in Watt which cannot be represented precisely with the available MIP model variables and $y_i = \log_{10}(x_i)$ are measures of insertion loss in dB which can be represented by a linear combination of the available MIP model variables. In equations (7), (8), (11)–(13), $=$ indicates that the two surrounding expressions are mathematically equivalent, \Leftrightarrow indicates that the minimization of the two surrounding expressions is exactly equivalent, i.e. leads to the same optimal assignment of values to model variables even though the expressions themselves are not equivalent, and \approx indicates that the minimization of the two surrounding expressions is not exactly equivalent but is still likely to lead to the same or a similar optimal assignment of values to model variables.

3.2.1 Off-chip lasers. For type Y off-chip lasers, the following simplification can be done:

$$\begin{aligned} \min P_{optical}^Y &= \min wl * \frac{1}{K} 10^{\frac{\max_{\lambda=1}^{wl} \{il_{\lambda}^L\} + S}{10}} \\ \Leftrightarrow \min wl * 10^{\frac{\max_{\lambda=1}^{wl} \{il_{\lambda}^L\}}{10}} \\ \Leftrightarrow \min 10 \log_{10}(wl * 10^{\frac{\max_{\lambda=1}^{wl} \{il_{\lambda}^L\}}{10}}) \\ \Leftrightarrow \min 10 \log_{10}(wl) + \max_{\lambda=1}^{wl} \{il_{\lambda}^L\} \end{aligned} \quad (8)$$

Since wl is an integer variable, the expression $10 \log_{10}(wl)$ can be precisely linearized into variable $\log wl$ with the use of M auxiliary binary variables α_i , where M is the maximum number of wavelengths available (that is, M is a constant and $wl \leq M$ is always true), and the following $M + 1$ constraints:

$$\log wl = \sum_{i=1}^M \alpha_i * 10 \log_{10}(i) \quad (9)$$

$$(wl = i) \Rightarrow (\alpha_i = 1) \quad \forall i = 1 \dots M \quad (10)$$

Thus Equation 8 is precisely equivalent to minimizing $P_{optical}^Y$. In this case, no approximation is made.

For type X off-chip lasers, the same simplification will now be only an approximation since the summation must be changed to a

product before applying the log function:

$$\begin{aligned} \min P_{optical}^X &= \min \sum_{\lambda=1}^{wl} \frac{1}{K} 10^{\frac{il_{\lambda}^L + S}{10}} \\ \Leftrightarrow \min \sum_{\lambda=1}^{wl} 10^{\frac{il_{\lambda}^L}{10}} &\approx \min \prod_{\lambda=1}^{wl} 10^{\frac{il_{\lambda}^L}{10}} \\ \Leftrightarrow \min 10 \log_{10} \left(\prod_{\lambda=1}^{wl} 10^{\frac{il_{\lambda}^L}{10}} \right) &\Leftrightarrow \min \sum_{\lambda=1}^{wl} il_{\lambda}^L \end{aligned} \quad (11)$$

3.2.2 On-chip lasers. For on-chip lasers the simplifications for both types are only approximations since both contain a summation (of the optical power over all laser sources) which must be converted to a product before applying the log function.

For type Y on-chip lasers we have:

$$\begin{aligned} \min \sum_{n=1}^N P_{optical_n}^Y &= \min \sum_{n=1}^N wl_n * \frac{1}{K} 10^{\frac{\max_{\lambda=1}^{wl_n} \{il_{n,\lambda}^L\} + S}{10}} \\ \approx \min \prod_{n=1}^N wl_n * 10^{\frac{\max_{\lambda=1}^{wl_n} \{il_{n,\lambda}^L\}}{10}} \\ \Leftrightarrow \min \sum_{n=1}^N \left(10 \log_{10}(wl_n) + \max_{\lambda=1}^{wl_n} \{il_{n,\lambda}^L\} \right) \end{aligned} \quad (12)$$

where $P_{optical_n}^Y$ and wl_n are respectively the power consumption and the number of wavelengths of the laser on the node n and N is the total number of nodes.

For type X on-chip lasers we have:

$$\begin{aligned} \min \sum_{n=1}^N P_{optical_n}^X &= \min \sum_{n=1}^N \sum_{\lambda=1}^{wl_n} \frac{1}{K} 10^{\frac{il_{n,\lambda}^L + S}{10}} \\ \approx \min \prod_{n=1}^N \prod_{\lambda=1}^{wl_n} 10^{\frac{il_{n,\lambda}^L}{10}} &\Leftrightarrow \min \sum_{n=1}^N \sum_{\lambda=1}^{wl_n} il_{n,\lambda}^L \end{aligned} \quad (13)$$

4 GRU POSITIONING OPTIMIZATION

As explained in Section 2.6, good router design is essential for balancing power usage in the PDN and reducing propagation loss in the router. We propose GRU positioning as the key to provide sufficient flexibility during router optimization. In this section the modeling of GRU positioning is described. First we define the feasible area in which GRUs and waveguides can exist. Then we explain the constraints for GRU positioning. Finally, we clarify how waveguide section routing is modeled, including how the resulting changes in the propagation and bending loss of the router waveguides are taken into consideration during optimization.

4.1 Feasible router area

Both the router and the PDN, when one exists, are placed on the optical layer. This can lead to crossings between PDN waveguides and router waveguides which substantially increase total laser power [8, 9]. PSION+ avoids this problem by dividing the optical layer into two areas, one for the PDN and one for the router. As long as the two areas are mutually exclusive, no crossings between router and PDN are possible. An example of these two areas for a

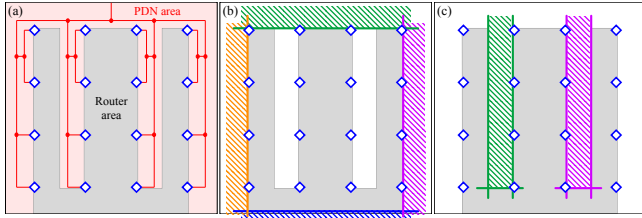


Figure 6: (a) Router area and PDN area for a 16 node WRONoC. Router area must be concave in this case. (b) The four convex edges for this router area. (c) The two concave edge sets for this router area. Striped areas are infeasible areas for the router.

16 node WRONoC with a PDN is shown in Figure 6(a). Both areas are continuous and we consider both to be polygons with only horizontal and vertical edges as this reduces modeling complexity significantly and we don't foresee more complex cases to be necessary in practice.

For node placements such as in Figure 4, a PDN can be designed such that the router area is convex. In other cases, for instance the one in Figure 6(a), the router area is always concave. For modeling purposes, we decompose any router area into convex edges and concave edge sets. A convex edge is an edge where one of the sides is an infeasible area for the router. The four convex edges for the example in Figure 6(a) are shown in Figure 6(b). A concave edge set is a set of three edges (two parallel and one perpendicular) where the area within the edges is an infeasible area for the router, as shown in Figure 6(c). The union of the infeasible areas defined by each convex edge and each concave edge set results in a complete definition of the total infeasible area of the router, whose complement is the feasible router area. If no PDN exists, the feasible router area is the entire optical layer, which can be modeled with just four convex edges on the boundary of the optical layer.

To avoid the insertion loss penalty caused by crossings between router and PDN, our GRU positioning and waveguide routing method must comply with the defined router area, that is, GRUs must be placed and waveguides must be routed within the feasible area defined by the convex edges and the concave edge sets.

4.2 GRU positioning

Our GRU placement method is designed to be simple to model in order to keep solver run-time overhead as low as possible. We also strive to keep the topology of the physical layout template intact, i.e. when optimizing GRU positions we do not create crossings between waveguides that didn't exist in the original physical layout template already. We achieve these goals with a straightforward set of constraints that keeps relative positions between interconnected pairs of GRUs and between interconnected GRUs and (de)modulator arrays.

For GRU–GRU connections, we consider all possible ways GRUs can be connected together (left port of GRU A with left port of GRU B, left port of GRU A with top port of GRU B, etc). We categorize those 16 ways into three types, as shown in Figure 7: *type I* (4 ways), *type L* (8 ways) and *type D* (4 ways). We add variables gpx_g, gpy_g

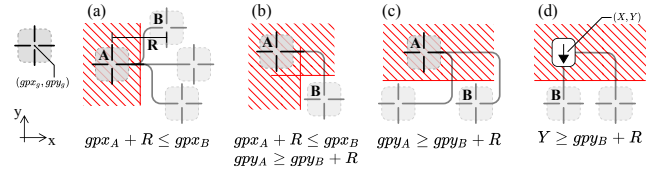


Figure 7: Positioning constraints for GRU–GRU and GRU–array connections. Striped areas are infeasible locations for GRU B. (a) Type I: ports on opposite sides. (b) Type L: ports on orthogonal sides. (c) Type D: ports on the same side. (d) Type A: ports on all four sides.

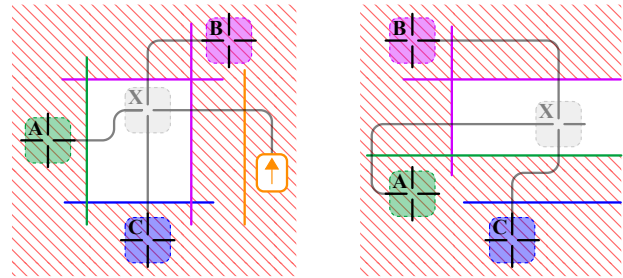


Figure 8: Two examples of how the location of GRU X is constrained by the locations of the elements connected to it, in this case GRUs A, B, C and a (de)modulator array. Striped areas are infeasible areas for GRU X.

for each GRU g to the MIP model, where (gpx_g, gpy_g) is the (x, y) coordinate of the center of GRU g , and constrain them appropriately for each pair of interconnected GRUs based on each connection type present in the template while always keeping a minimum distance R between each GRU pair.

For GRU–(de)modulator array connections, there are only four possible ways (left port of GRU connected to array, top port of GRU connected to array, etc) and all four ways are categorized into the same type, *type A*, as shown in Figure 7. We add constraints for gpx_g and gpy_g based on the port of the GRU that connects to the array while keeping a minimum distance R between the GRU and the array.

The aforementioned set of constraints results in an infeasible area for each GRU which is defined by its connected elements. Two examples are shown in Figure 8. Note that all constraints in an MIP model are considered together during optimization. Thus, if a location of a GRU changes during optimization, so will the infeasible area (and consequently possibly also the location) of its connected GRUs.

Finally, to ensure that all GRUs are placed within the router area, a set of constraints for each GRU g is created for each convex edge and each concave edge set such that GRU g cannot be in any of their infeasible areas. Thus, the complete infeasible area for a GRU is the union of the infeasible areas defined by the convex and concave edges with the infeasible areas created by the GRU's connected elements.

4.3 Waveguide section routing

Since we keep the topology of the physical layout template intact when moving GRUs, waveguide routing is straightforward: waveguides are routed vertically and horizontally while minimizing their length and number of bends.

If no concave edge sets exist (only convex edges are present), the length of the waveguides can be approximated by the Manhattan distance between their starting and ending locations, which are given by the gpx_g and gpy_g variables and the fixed positions of the (de)modulator arrays. For example, the length of a waveguide section connecting GRUs g_1 and g_2 is $|gpx_{g_1} - gpx_{g_2}| + |gpy_{g_1} - gpy_{g_2}|$ for any type of GRU-GRU connection³. This is because no GRUs or (de)modulator arrays are ever in the infeasible area of convex edges. Linearization techniques are applied here to model the absolute value expressions. The number of bends is also easily modeled based on the type (I, L, D or A) and the relative positions of the starting and ending locations.

If concave edge sets exist, more complex modeling is required. If the starting and ending locations of a waveguide are such that the Manhattan routing of the waveguide goes through the infeasible area of a concave edge set, then the waveguide must make a detour around the edges of the concave edge set. The total length of the waveguide is now the Manhattan distance as written above plus an extra length caused by the detour. This detour also adds extra bends. We model these cases accordingly such that the correct length and number of bends is always considered by the MIP solver during optimization.

The changes in length and number of bends for each waveguide section induce changes in the amount of propagation and bending loss for optical signals going through that waveguide section. The constraints that calculate the insertion loss for each optical signal take into consideration the effects of GRU positioning and waveguide routing. These alterations to the amount of propagation and bending loss of optical signals not only reduce total insertion loss in the router but also help avoid power losses due to the max operations by rebalancing the router design.

5 RESULTS

We tested our new methodology, which includes PDN modeling, GRU positioning and accurate optimization functions, against the same WRONoC test cases from multiple other works [2, 4, 8, 11, 12, 14, 15] to ensure a fair comparison. These are an 8 node WRONoC and a 16 node WRONoC. PSION+ [12] was shown to outperform the other optimization tools and manual designs in both test cases so it will be used as the baseline for this comparison.

5.1 8 node WRONoC

This test case contains 8 nodes on a 9 mm die: four computing hubs and four memory controllers for off-chip memory. The hubs are equally spaced on a 2×2 grid and the memory controllers are next to the edges of the die. PSION+ tested five physical layout templates, shown in Figure 9(a)–(e): a non-expanded Centralized Grid Template (CGT-e0), an expanded Centralized Grid Template

³This is only an approximation as some waveguide sections may be slightly longer. For example, with Type D GRU-GRU connections, or to add sufficient spacing between parallel waveguides. But the error is small enough to be ignored for modeling purposes.

Table 1: Laser power reduction using GRU positioning and accurate optimization functions for 8 node WRONoC.

Laser & PDN		Laser power reduction				
Location	Type	CGT-e0	CGT-e6	DGT	Ring	Custom
Off-chip &	X	13.6%	17.6%	22.5%	11.8%	3.5%
Sym. PDN	Y	7.9%	4.5%	11.5%	10.0%	1.8%
Off-chip &	X	19.6%	20.4%	20.5%	13.1%	4.9%
Asym. PDN	Y	19.3%	15.1%	0.0%	15.0%	11.5%
On-chip &	X	10.7%	11.5%	3.6%	2.1%	6.7%
no PDN	Y	9.1%	6.6%	4.5%	1.8%	7.1%

with 6 extra waveguide sets (CGT-e6), a Distributed Grid Template (DGT), a ring template with 3 rings and a custom template. To test off-chip lasers we created two PDN designs that connect to those 8 nodes, shown in Figure 9(f)(g): one PDN is symmetric in its construction while the other is asymmetric.

This test case yielded 30 design configurations where each is a combination of template design, laser location, laser type and PDN design (when off-chip lasers are used). To compare our methodology against the state-of-the-art PSION+ we optimized each configuration twice: as a baseline we optimized each configuration using PSION+ (i.e. no GRU positioning and $\max\{il_{n,\lambda}^N\}$ as the optimization function) and for our methodology we used GRU positioning and the accurate optimization function for each laser type and location as explained in Section 3.2.

The results are shown in Table 1. On average we reduce laser power by **10.2%**. The following conclusions can be drawn:

- For 4 out of 5 templates, off-chip lasers (i.e. with a PDN) produce a higher average improvement compared to on-chip lasers (no PDN). This is because considering the PDN in the optimization allows the solver to compensate for its effects, which is something that PSION+ cannot do.
- The asymmetric PDN yields higher reductions compared to the symmetric PDN on 80% of the cases. This is expected since the asymmetric PDN has a higher imbalance in the distribution of insertion loss between its branches, which leads to more potential for rebalancing like in Figure 5.
- Type X lasers yield higher reductions compared to type Y lasers on 73% of the cases. This is because both the optimization function used by PSION+ ($\max\{il_{n,\lambda}^N\}$) and the optimization function for type Y lasers contain a max operation, whereas the optimization function for type X lasers does not. This makes the optimization function used by PSION+ a worse approximation to the function for type X lasers than for type Y. Hence, using the accurate optimization function for type X lasers results in a higher improvement.
- Even when on-chip laser sources are used and no PDN is required, we still reduce significantly the amount of laser power required. Here the reduction is achieved not only because of the accurate optimization functions, but also by optimally positioning the GRUs. Running the test case for on-chip lasers with the highest improvement (CGT-e6 type X, improvement of **11.5%**) without GRU positioning reduces the improvement to only **7.6%**.

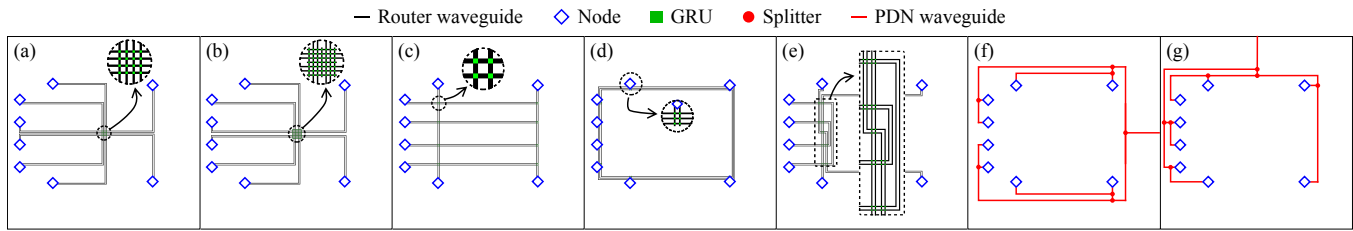


Figure 9: Physical layout templates and PDN designs for an 8 node WRONoC. (a) Non-expanded Centralized Grid Template (CGT-e0). (b) Expanded CGT with 6 extra waveguide sets (CGT-e6). (c) Distributed Grid Template (DGT). (d) Ring template with 3 rings. (e) Custom template. (f) Symmetric PDN. (g) Asymmetric PDN.

We also found that different templates achieve their power reduction results in different ways. For example:

- The CGT templates are originally concentrated on the center of the die, so GRU positioning is paramount to achieving high amounts of power reduction by scattering the GRUs to their optimal locations. Running the test case for the CGT template with the highest improvement (CGT-e6 off-chip type X asymmetric PDN, improvement of **20.4%**) without GRU positioning reduces the improvement to only **12.5%**.
- The DGT template already has the GRUs distributed throughout the die, so the gains due to GRU positioning are much smaller. For these cases, the power reduction is mostly due to the inclusion of the PDN in the optimization process and the use of better optimization functions. In fact, these two aspects are so crucial that the DGT off-chip type X symmetric PDN case is able to achieve **22%** reduction based solely on them, i.e. without moving any GRUs.

Here we used the same technology parameters as [14]. We tested other sets of technology parameters and obtained similar results. More specifically, we found that technology parameter sets with high propagation loss coefficients lead to higher improvements. From this we deduce that our GRU positioning method is effective at reducing propagation loss in the router. On the whole, we can conclude that all aspects of this work (accurate optimization functions, consideration of the PDN and GRU positioning) play a decisive role in reducing laser power for WRONoCs.

These test cases took on average about $5\times$ longer to run than PSION+. We consider this a worthwhile quality/run-time trade-off as all of PSION+'s cases, except for the Ring template, complete in under two minutes.

Finally, to understand how the reduction in laser power depends on the density of the communication matrix, we chose, as an example, the CGT-e0 asymmetric PDN test case. With this case we performed random testing for communication matrices from one to 56 optical signals to obtain average laser power reduction values for all four combinations of laser locations and types⁴. The results are shown in Figure 10. Even with very sparse communication matrices our methodology achieves substantial reductions in optical power. Over the full range of number of optical signals we obtain an improvement between **7%** and **20%**. We can observe that the improvement for off-chip lasers is in general higher than for on-chip

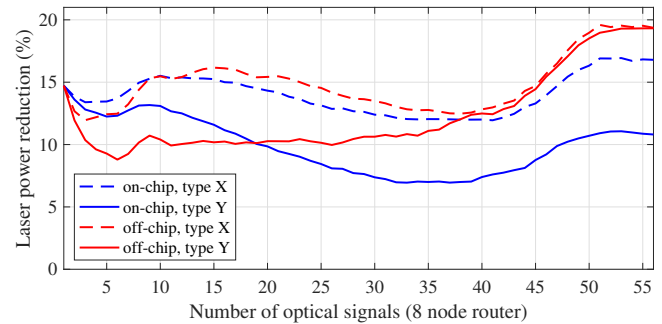


Figure 10: Laser power reduction for different laser configurations and communication matrix densities for an 8 node CGT-e0 with Asymmetric PDN.

lasers. This is again due to the presence of the PDN, which provides more opportunities for insertion loss rebalancing.

5.2 16 node WRONoC

This test case contains 16 equally spaced nodes in a 4×4 grid configuration on a 16 mm die [8, 12]. Here the laser source is off-chip, so a PDN is required. Since the nodes are equally spaced on a grid pattern, the PDN is symmetric. The node placement along with the PDN design for this test case is equal to those of Figure 6(a).

PSION+ uses an expanded CGT-e10 for this test case. It outperforms the manual designs from [8] and so it is the baseline for our comparison. By using our methodology with GRU positioning, PDN aware optimization and accurate optimization functions to optimize the exact same template we are able to reduce optical power consumption by **20.2%** and **8.0%** for type X and type Y laser sources respectively.

6 CONCLUSION

This paper proposes new design automation methods for WRONoC design which feature accurate, laser type aware, optimization functions, and GRU positioning capabilities. We have shown that these improvements contribute significantly to reduce laser power consumption.

ACKNOWLEDGMENTS

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 146371743 – TRR 89 Invasive Computing.

⁴For on-chip laser tests the PDN was removed.

REFERENCES

- [1] S. Le Beux et al. 2013. Reduction methods for adapting optical network on chip topologies to 3D architectures. *Microprocessors and Microsystems* (Feb. 2013), 87–98. <https://doi.org/10.1016/j.micpro.2012.11.001>
- [2] A. Boos et al. 2013. PROTON: An automatic place-and-route tool for optical Networks-on-Chip. In *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. 138–145.
- [3] M. Briere et al. 2007. System Level Assessment of an Optical NoC in an MPSoC Platform. In *2007 Design, Automation Test in Europe Conference Exhibition*.
- [4] Y. Chuang et al. 2018. PlanarONoC: Concurrent Placement and Routing Considering Crossing Minimization for Optical Networks-on-chip. In *Proceedings of the 55th Annual Design Automation Conference*. 151:1–151:6.
- [5] D. Ding et al. 2009. O-Router: An optical routing framework for low power on-chip silicon nano-photonic integration. *Proceedings of Design Automation Conference*, 264–269. <https://doi.org/10.1145/1629911.1629983>
- [6] F. Jiao et al. 2018. Thermal-Aware Placement and Routing for 3D Optical Networks-on-Chips. In *2018 IEEE International Symposium on Circuits and Systems*. 1–4. <https://doi.org/10.1109/ISCAS.2018.8351101>
- [7] Mengchu Li et al. 2018. CustomTopo: A Topology Generation Method for Application-Specific Wavelength-Routed Optical NoCs. In *Proceedings of the 37th International Conference on Computer-Aided Design*. 1–8. <https://doi.org/10.1145/3240765.3240789>
- [8] M. Ortín-Obón et al. 2017. Contrasting Laser Power Requirements of Wavelength-Routed Optical NoC Topologies Subject to the Floorplanning, Placement, and Routing Constraints of a 3-D-Stacked System. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (July 2017), 2081–2094. <https://doi.org/10.1109/TVLSI.2017.2677779>
- [9] M. Ortín-Obón et al. 2017. A tool for synthesizing power-efficient and custom-tailored wavelength-routed optical rings. In *Asia and South Pacific Design Automation Conference*. 300–305. <https://doi.org/10.1109/ASPDAC.2017.7858339>
- [10] X. Tan et al. 2011. On a Scalable, Non-Blocking Optical Router for Photonic Networks-on-Chip Designs. In *2011 Symposium on Photonics and Optoelectronics*. 1–4. <https://doi.org/10.1109/SOPO.2011.5780550>
- [11] A. Truppel et al. 2019. PSION: Combining Logical Topology and Physical Layout Optimization for Wavelength-Routed ONoCs. In *Proceedings of the 2019 International Symposium on Physical Design*. 49–56. <https://doi.org/10.1145/3299902.3309747>
- [12] A. Truppel et al. 2020. PSION+: Combining logical topology and physical layout optimization for Wavelength-Routed ONoCs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* (2020).
- [13] T. Tseng et al. 2019. Wavelength-Routed Optical NoCs: Design and EDA – State of the Art and Future Directions: Invited Paper. In *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. 1–6.
- [14] A. von Beuningen et al. 2015. PROTON+: A Placement and Routing Tool for 3D Optical Networks-on-Chip with a Single Optical Layer. *J. Emerg. Technol. Comput. Syst.* (Dec. 2015), 44:1–44:28. <https://doi.org/10.1145/2830716>
- [15] A. von Beuningen et al. 2016. PLATON: A Force-Directed Placement Algorithm for 3D Optical Networks-on-Chip. In *Proceedings of the 2016 International Symposium on Physical Design*. 27–34. <https://doi.org/10.1145/2872334.2872356>