Physically Aware Wavelength-Routed Optical NoC Design for Customized Topologies with Parallel Switching Elements and Sequence-Based Models

Wei-Yao Kao, Tai-Jung Lin and Yao-Wen Chang

ASPDAC'25, January 20–23, 2025, Tokyo, Japan

National Taiwan University

臺灣大學

Physically Aware Wavelength-Routed Optical NoC Design for Customized Topologies with Parallel Switching Elements and Sequence-Based Models

Wei-Yao Kao, Tai-Jung Lin and Yao-Wen Chang

ASPDAC'25, January 20–23, 2025, Tokyo, Japan

National Taiwan University

臺灣大學

Outline

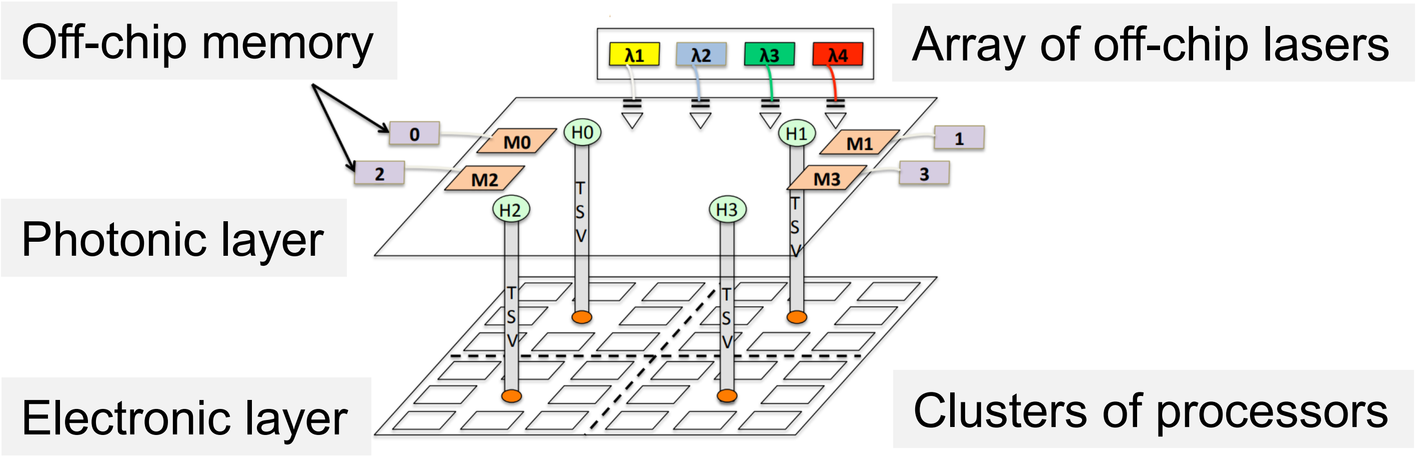
- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

Outline

- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

Wavelength-Routed Optical Networks-on-Chips (WRONoCs)

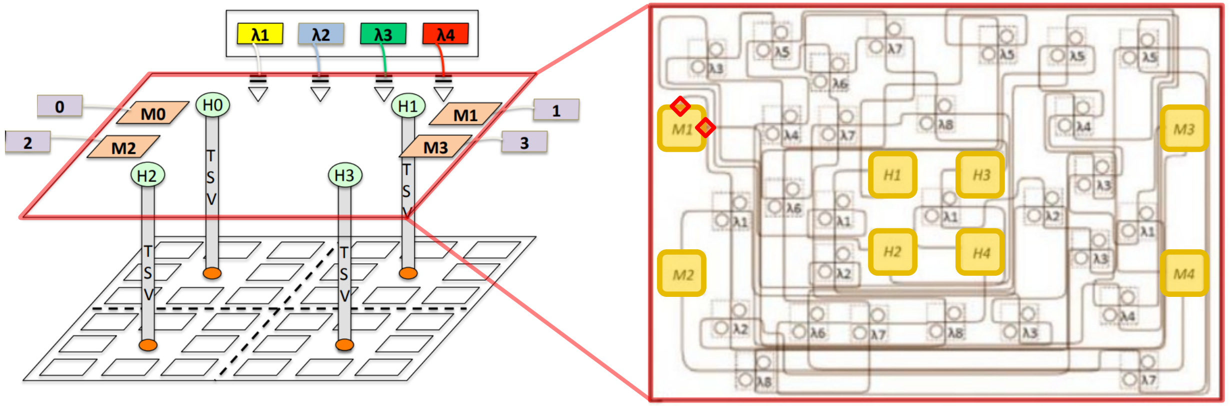
- Architecture
  - The photonic layer is stacked vertically above the electronic layer
  - Communicate by through-silicon vias (TSVs)
  - Integrate within 3D-stacked multiprocessor systems
- Mechanism
  - Convert electrical signals into optical signals (lasers)
  - Transmit signals through optical waveguides with different wavelengths
- Advantages
  - Power efficiency
  - Low latency
  - High bandwidth



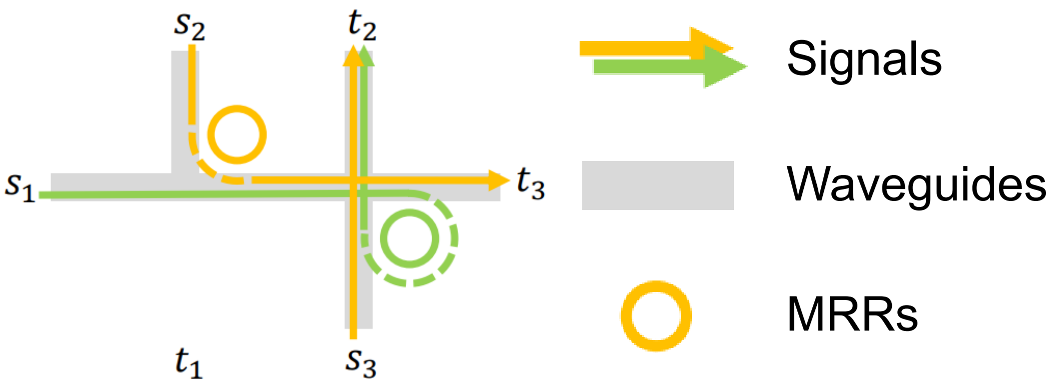
3D-stacked multiprocessor systems [Beuningen, ISPD'16]

WRONoC Design

- Objective
  - Construct **node** communications: transmitting signals from sources (s) to targets (t)
  - **Customized topologies**: for applications that do not need all-to-all communications
- Core Components
  - **Optical waveguides**: guide different wavelengths of signals, allow crossings
  - **Microring resonators (MRRs):** switch signals from one waveguide to another
- Concerns
  - Signals suffer from power loss (**insertion loss**)



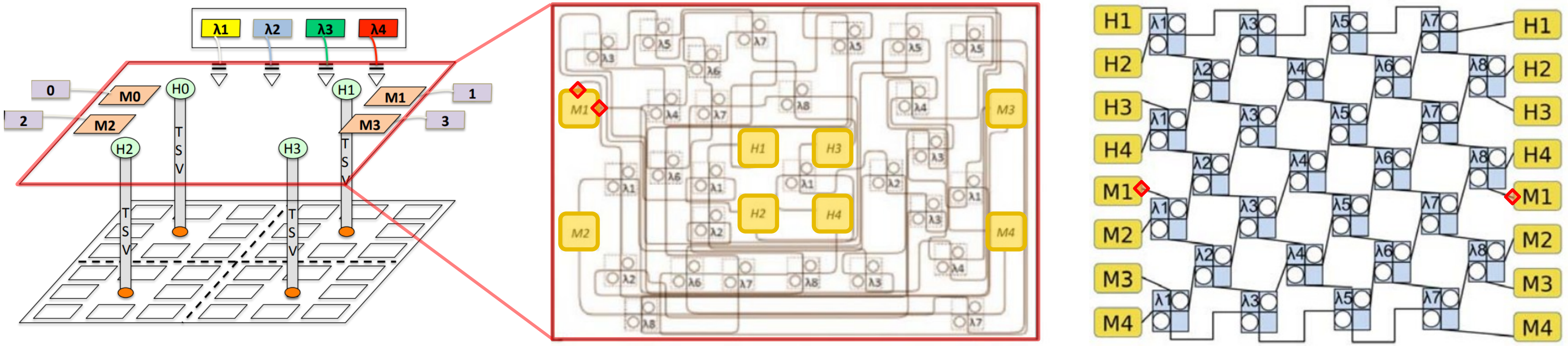
WRONoC layout result [Tseng et al., ICCAD'19]



WRONoC core components

WRONoC Design Flow

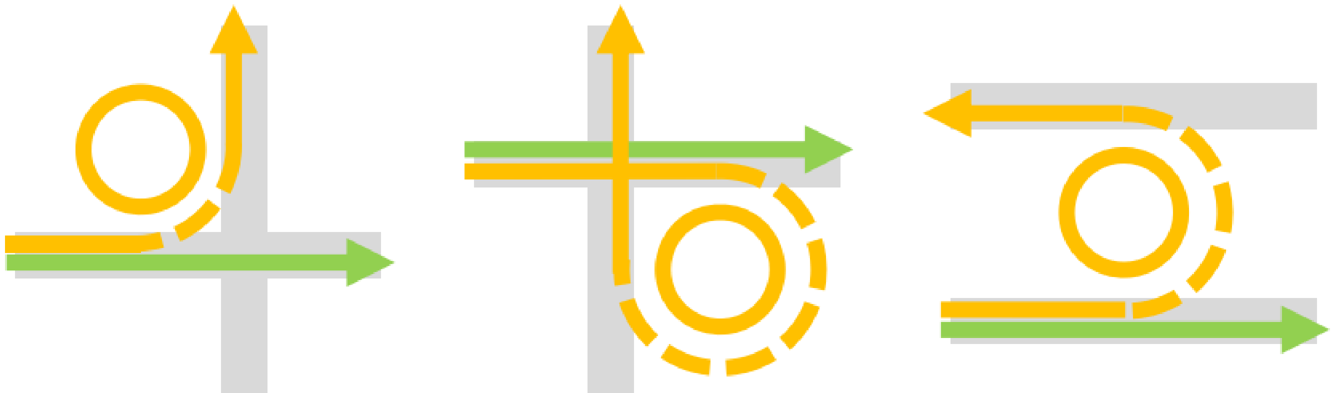
- Topology design stage
  - Manage the logical communication transmission among nodes
  - Involve **waveguides and MRRs utilization** and **wavelengths assignment**
- Physical design stage
  - Translate the information from topologies
  - Realize the optical components placement and the actual waveguide routing



WRONoC **layout result** and **topology** from separate stages [Tseng et al., ICCAD’19]

Topology Design – Utilizing MRRs

- Mechanism
  - Route each signal to its intended destination with some specific wavelengths
  - Signal **activate** the MRR and then **drop** to an adjacent waveguide
- MRR switching structures
  - Crossing switching element (CSE)
  - Parallel switching element (PSE)



90° CSE　　270° CSE　　180° PSE

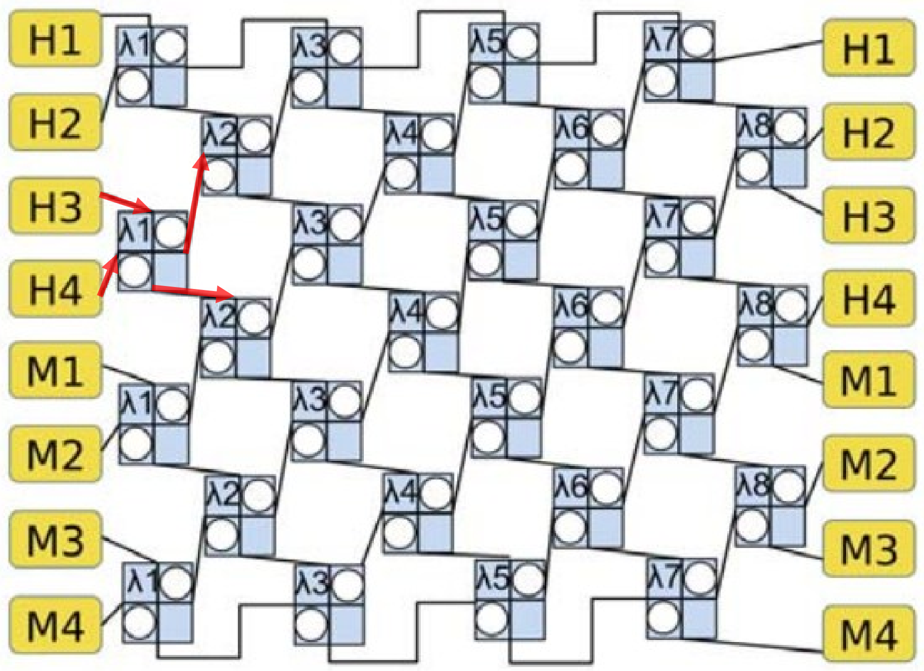
Waveguides　　Signals　　MRRs

Topology Design – Communication Constructions

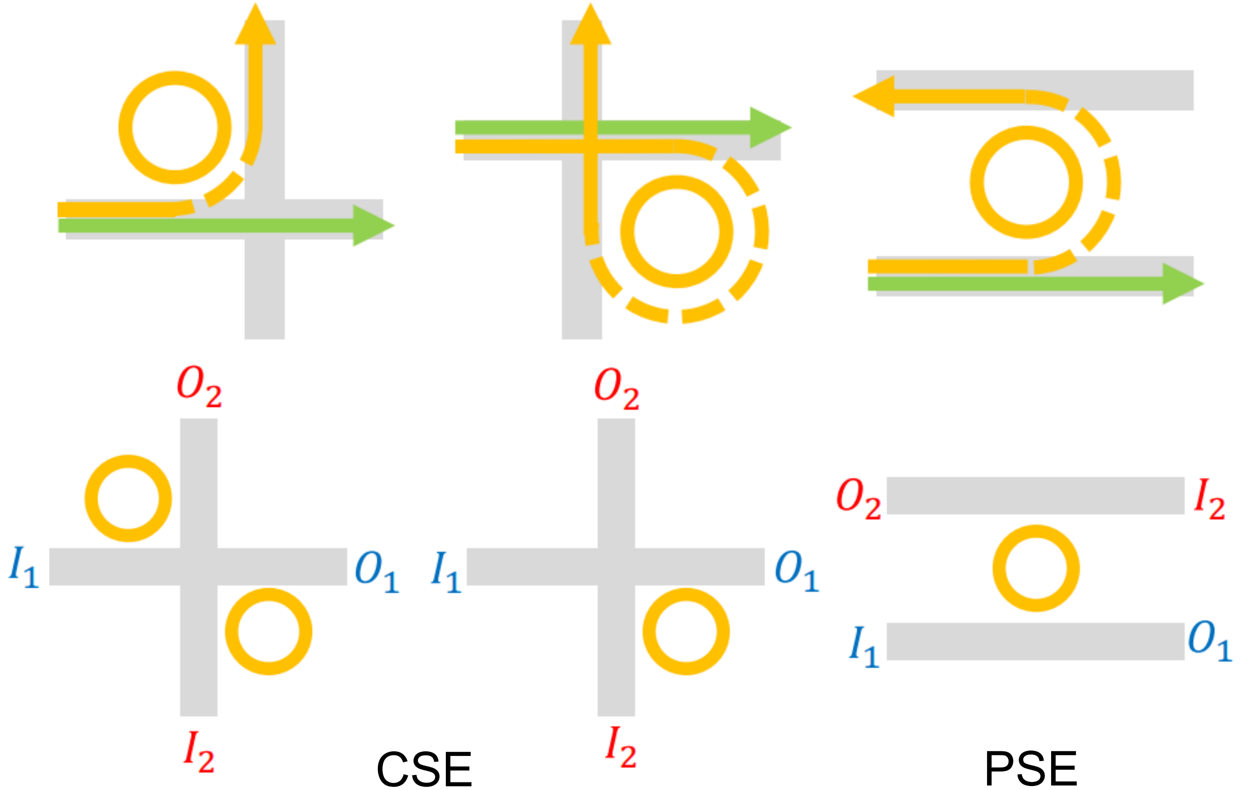
- Construct communications using waveguides $(s_1 \rightarrow t_3)$
  - **Default paths**: directly connect the signal sources to the signal targets
  - **Default communications**: communications supported by the default paths
- Construct communications by MRRs $(s_1 \rightarrow t_2)$
  - Viewpoint 1: insert MRR to the default path
  - Viewpoint 2: interact the two default paths

Topology Design – Building Block

- Building block: **add-drop-filter (ADF)**
  - The structure of the interaction of two default paths
  - Support communications of up to **2-input by 2-output** $(I_1 I_2 \times O_1 O_2)$



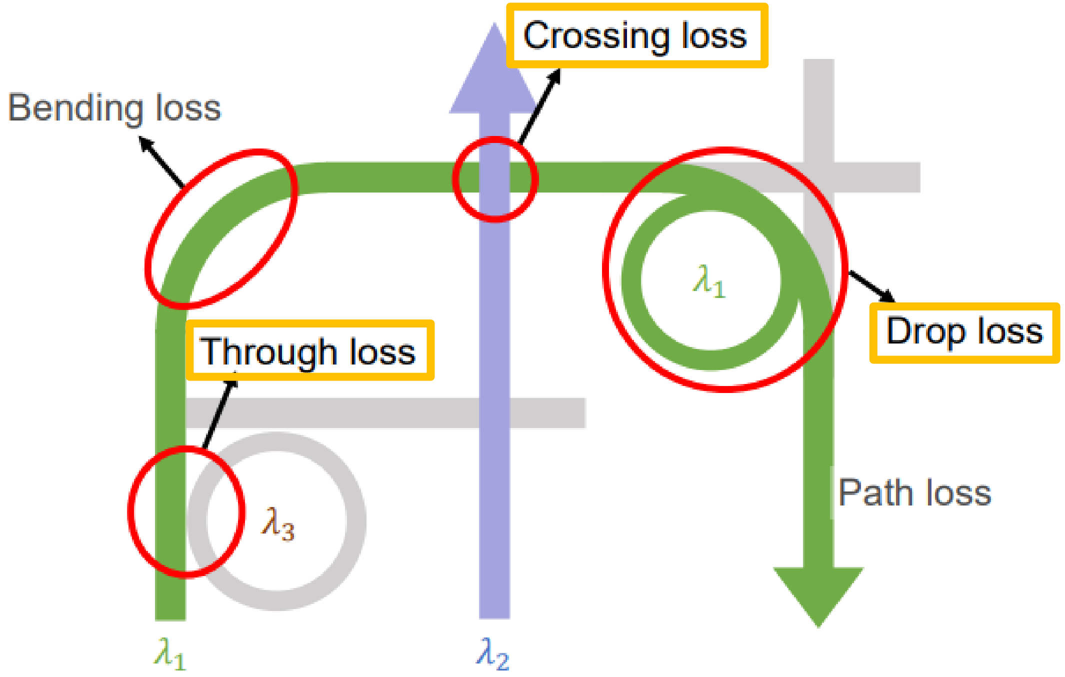
Sources Targets



CSE PSE

Topology Design – Insertion Loss

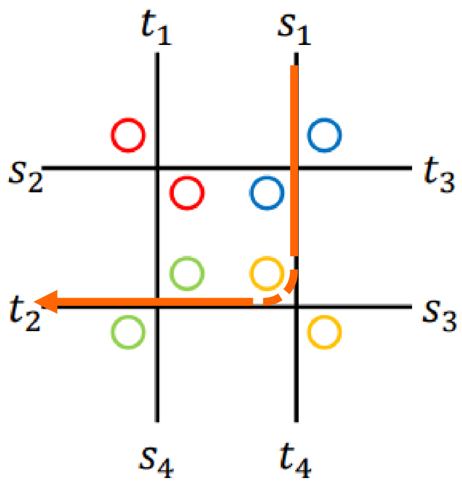
- In the topology design, we shall consider
  - Drop loss, through loss, and crossing loss
- Objective: minimize the **maximum insertion loss**
  - Maximum insertion loss among all signals determines the **minimum required power**



Insertion loss includes five types of transmission losses [Lu et al., TCAD'22]

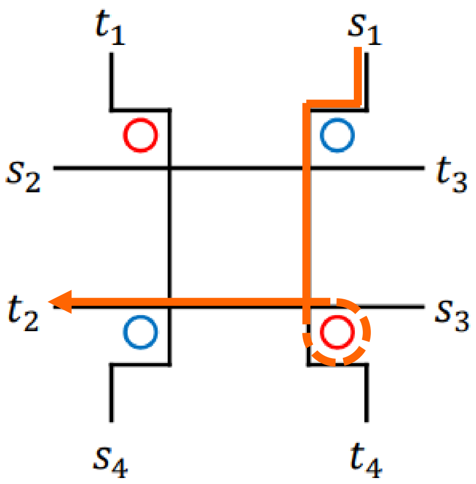
Motivation (1/2) – Utilizing CSEs and PSEs

- Previous works adopt **predefined grid-like template** and **CSE structures**
  - Draw a planar grid-like template
  - Determine whether to put MRRs by the waveguide intersections
- Cons
  - Constrain the solution space of the topology
  - PSE is expected to result in lower MRR usage and fewer waveguide crossings



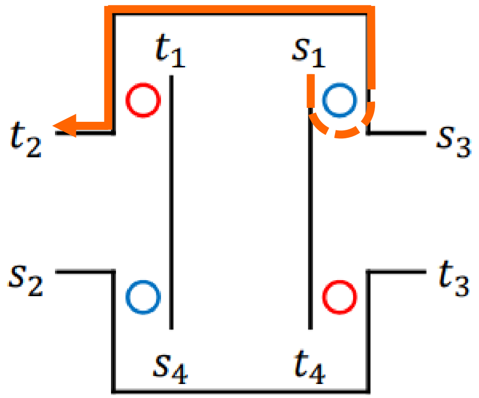
GWOR: 0.600 dB

[Tseng et al., ICCAD'19]



HASH: 0.670 dB

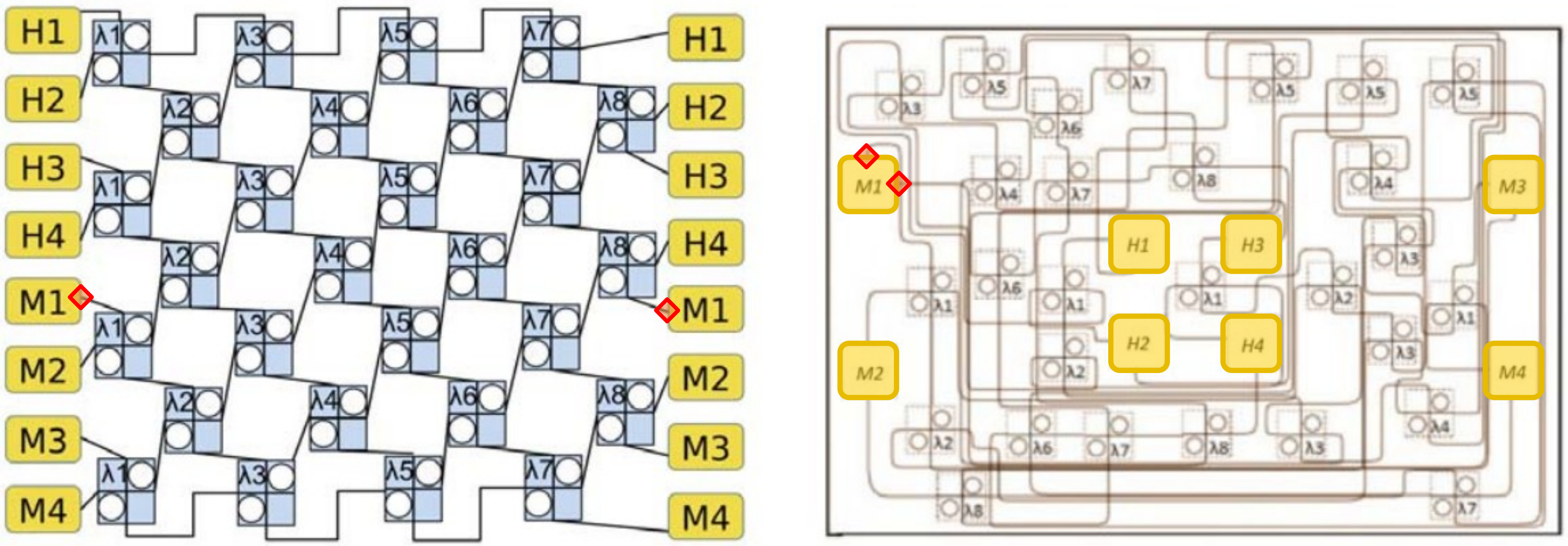
[Zheng et al., ASPDAC'21]



Our 4x3 one: 0.505 dB

Motivation (2/2) – Considering Physical Specifications

- Node positions are not considered sufficiently
  - Lead to a translating mismatch between separate design stages
  - Incur inaccurate loss calculation when translating topology design to physical design

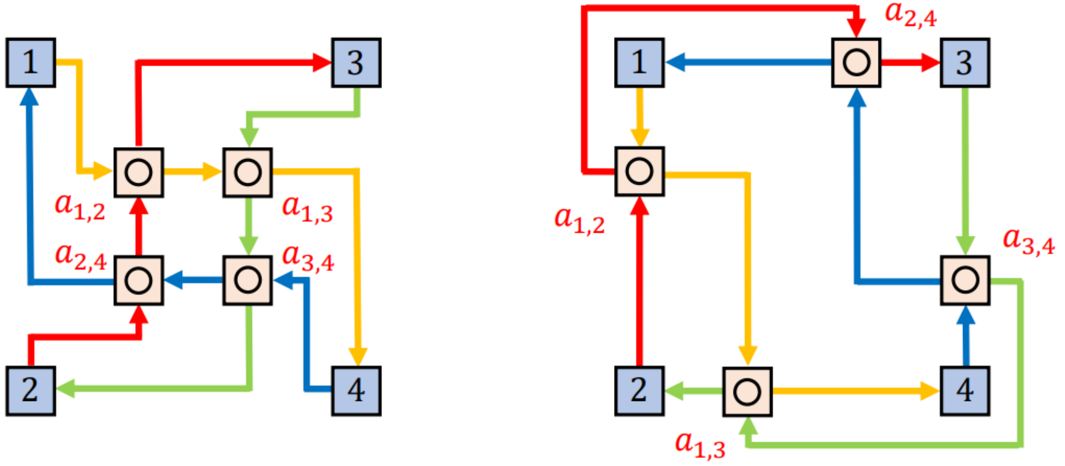


WRONoC topology and layout result from separate stages [Tseng et al., ICCAD'19]

Our Solution

- Construct logical communication with **sequence-based models**
  - Recall: signals can switch to the other default paths by ADFs
  - Each sequence $\langle Q_i \rangle$ represents a default path (d_i)
  - Determine the ADFs $(a_{i,j})$ each sequence should pass through
- Draw the sequences on the plane
  - Planar graph drawing and consider fix-nodes locations

$$\langle Q_1 \rangle = \langle a_{1,3}, a_{1,2} \rangle$$
$$\langle Q_2 \rangle = \langle a_{2,4}, a_{1,2} \rangle$$
$$\langle Q_3 \rangle = \langle a_{3,4}, a_{1,3} \rangle$$
$$\langle Q_4 \rangle = \langle a_{3,4}, a_{2,4} \rangle$$



ADFs sequences | Initial placement | Optimized solution

Our Contributions

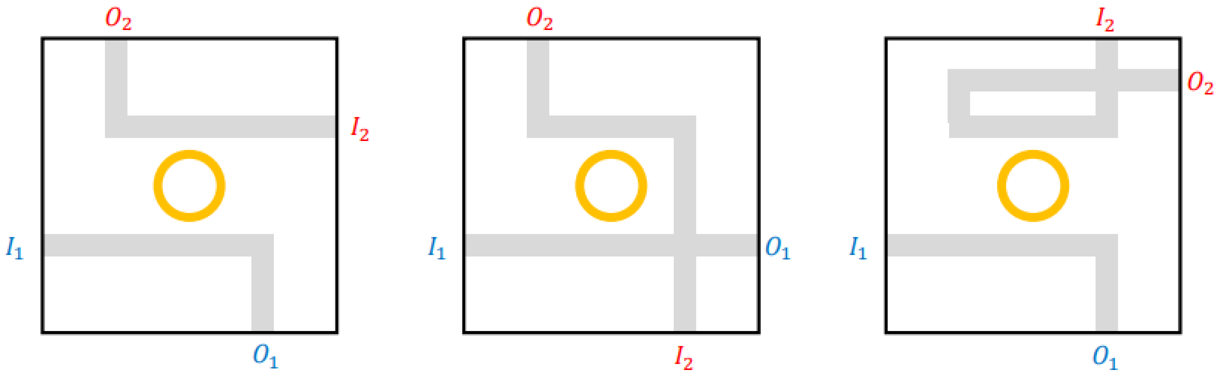
- We propose a fully automated topology design flow
  - Utilize **PSE structures** for customized WRONoC topology designs
- Our algorithm flow incorporates an **ADF sequence model**
  - Expand the solution space beyond the constraints of a grid-like template
- Our **fixed-node crossing-aware edge routing**
  - Effectively minimizes the waveguide crossings
  - Our A*-search preserves the admissibility property and guarantees an optimal routing solution
- Our design flow directly **considers physical specifications**
  - Minimize potential mismatches caused by separate design stages
- Experimental results show that our method substantially outperforms the existing design flows for customized topology designs

Outline

- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

Our ADF with PSE Structures

- In this work, we apply PSE structures as the ADF building block
- ADFs can be classified into 3 types (permutations of 2-input and 2-output)
  - Type 1 $(IOIO)$: utilizes the PSE structure optimally (no waveguide crossing)
  - Type 2 $(I_1 I_2 O_1 O_2)$
  - Type 3 $(I_1 I_2 O_2 O_1)$



	O_1	O_2
I_1	0	0
I_2	0	0

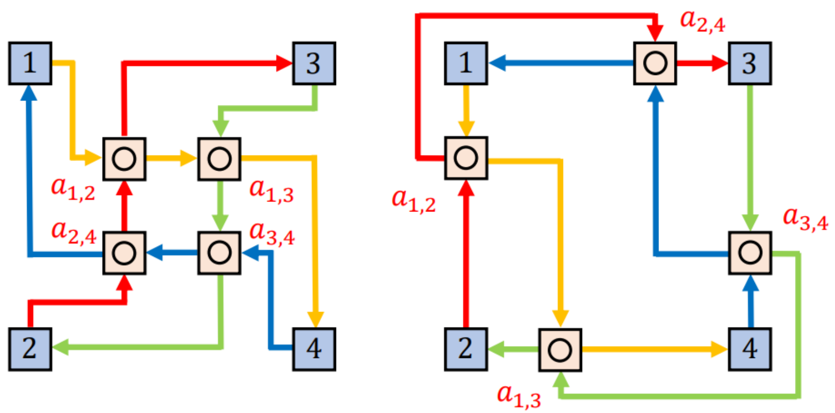
Type 1

	O_1	O_2
I_1	1	0 or 2
I_2	2 or 0	1

Type 2

	O_1	O_2
I_1	0 or 2	1
I_2	1	2 or 0

Type 3



Type 2 Type 1

Three types of ADFs and the number of crossings for communications

Problem Formulation

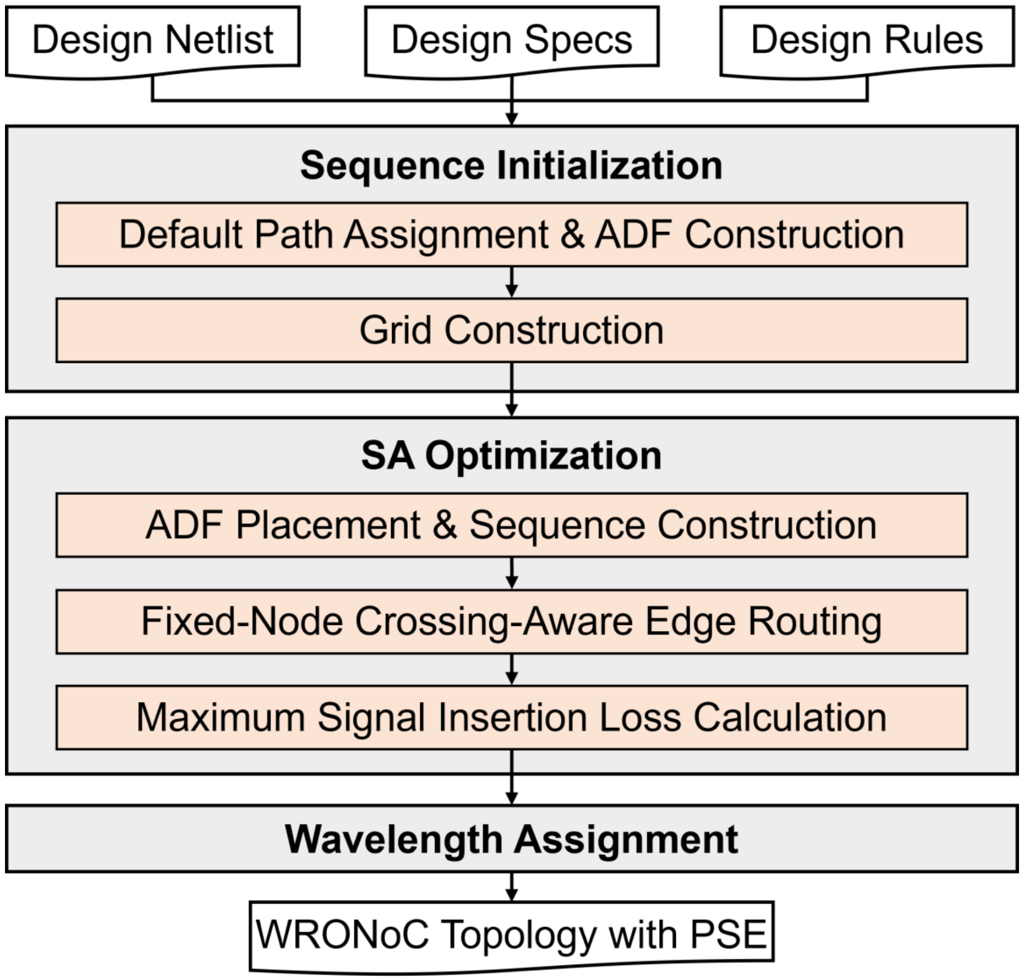
- Problem
  - Physically Aware WRONoC Topology Design Problem
- Given
  - A WRONoC netlist
  - Node locations
- Output
  - A WRONoC topology
- Objective
  - Minimize the MRR usage
  - Minimize the maximum insertion loss
  - Minimize the wavelength usage

Outline

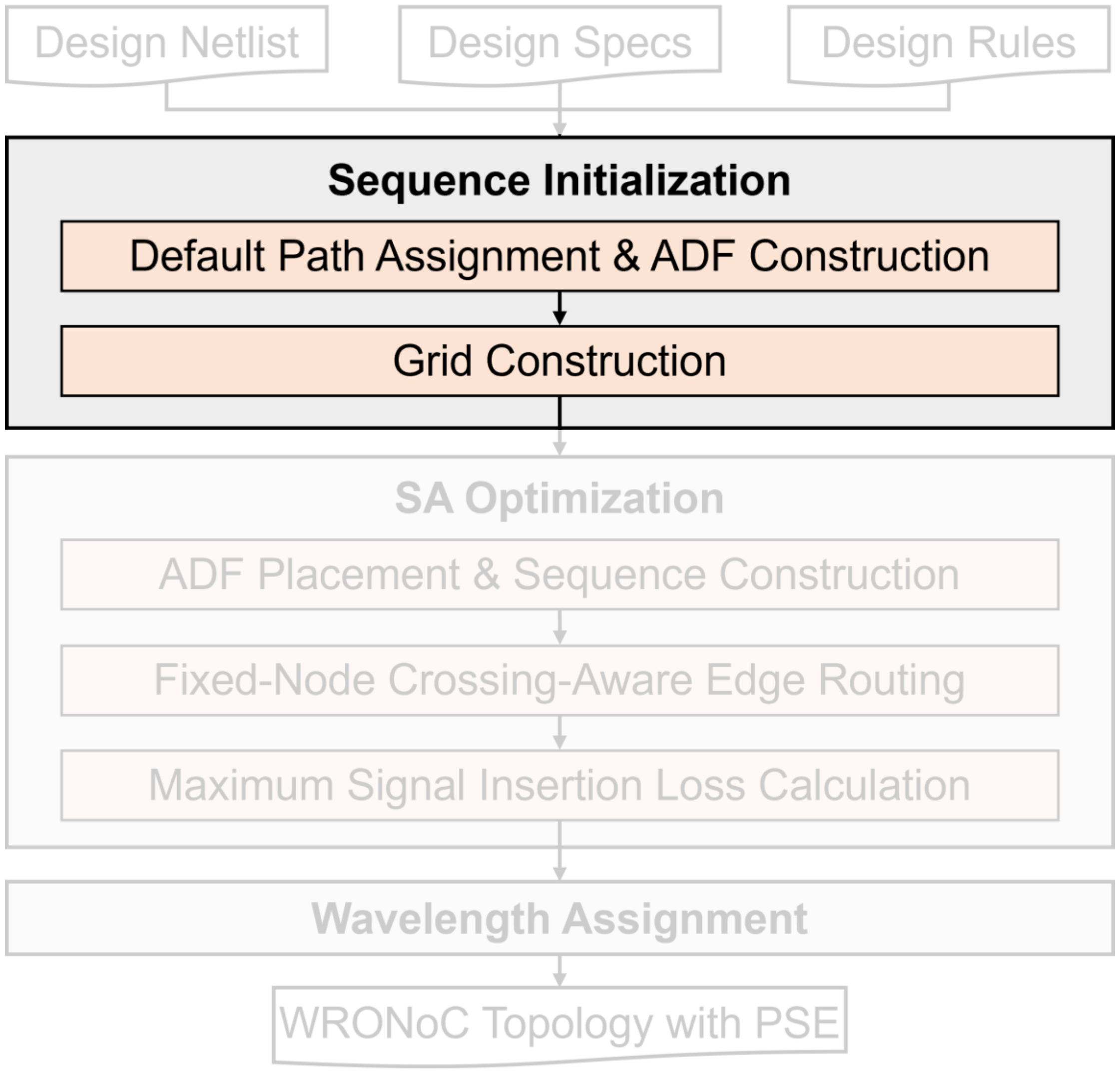
- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

An Overview of the Algorithm Flow

- Sequence initialization
  - **Minimize the MRR usage**
  - Construct ADF sequences for the netlist
- Simulated Annealing (SA) optimization
  - Planarize the topology from ADF sequences
  - Apply fixed-node crossing-aware edge router on a specially designed grid structure
  - Search for topologies with **lower maximum insertion loss** using the SA algorithm
- Wavelength assignment
  - **Optimize the wavelength usage** using ILP
  - Assign wavelength for all signals and MRRs

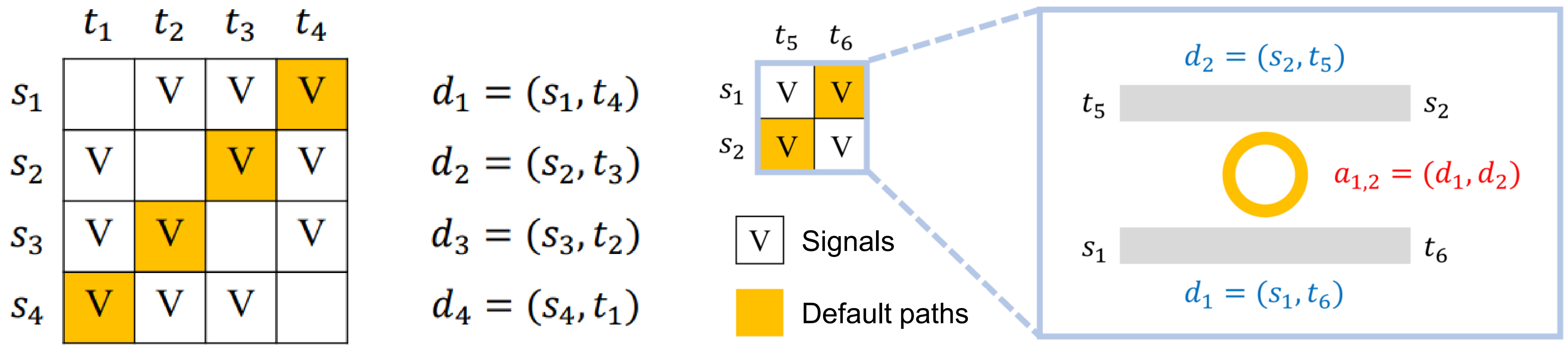


Algorithm Flow



Default Path Assignment

- Assign default paths to minimize MRR usage
  - Assign n default paths for n nodes
  - Maximize the default communication usage
  - Enable ADF communication sharing
- Once the default paths were assigned, the ADFs can be constructed



Default path assignment for the communication matrix and the according ADF structure

Default Path Assignment for Minimizing the MRR Usage

- Select the one with the lowest MRR usage
  - Anti-diagonal assignment: assign default path $d_i = (s_i, t_{n-i})$ for $i = 1\ to\ n$
  - Diagonal assignment: assign default path $d_i = (s_i, t_i)$ for $i = 1\ to\ n$
  - Maximum flow assignment: adapted from [Chen et al., ICCAD'23]

	t_1	t_2	t_3	t_4	t_5	t_6
s_1		V	V	V	V	V
s_2	V		V	V	V	V
s_3	V	V		V	V	V
s_4	V	V	V		V	V
s_5	V	V	V	V		V
s_6	V	V	V	V	V	

Anti-diagonal: 12 MRRs

	t_1	t_2	t_3	t_4	t_5	t_6
s_1					V	
s_2					V	
s_3					V	V
s_4					V	V
s_5	V	V	V	V		V
s_6			V	V	V	

Diagonal: 7 MRRs

	t_1	t_2	t_3	t_4	t_5	t_6
s_1					V	
s_2					V	
s_3					V	V
s_4					V	V
s_5	V	V	V	V		V
s_6			V	V	V	

Maximum flow: 10 MRRs

ADF Construction

- In this step, we establish communications for the netlist
  - Assume each communication requires at most one ADF drop to reach the target
- Procedure
  - Construct ADFs for non-default communications
  - Obtain the MRR usage and a set of ADF sequences

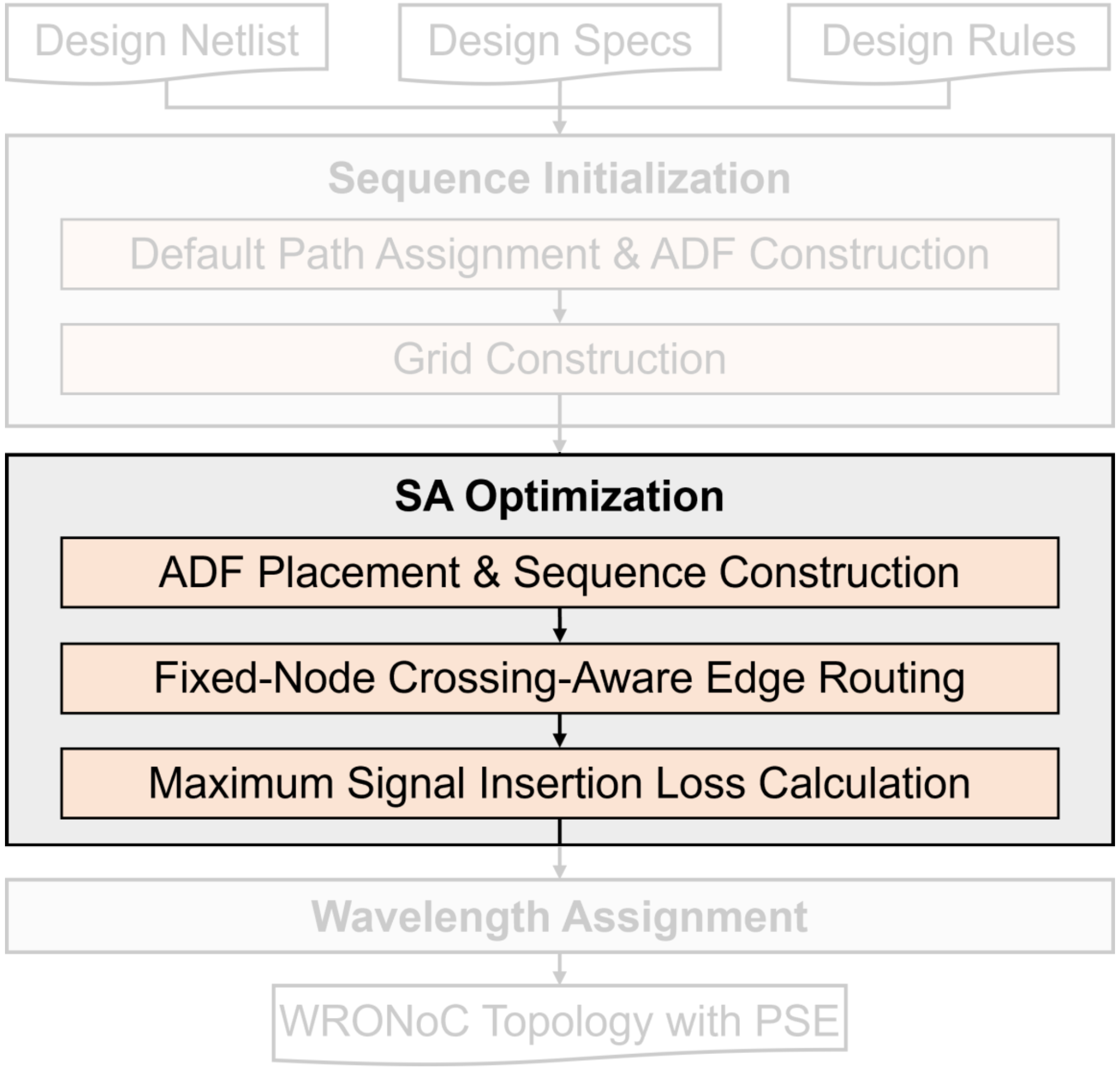
	t_1	t_2	t_3	t_4	
s_1		V	V	V	$d_1 = (s_1, t_4)$
s_2	V		V	V	$d_2 = (s_2, t_3)$
s_3	V	V		V	$d_3 = (s_3, t_2)$
s_4	V	V	V		$d_4 = (s_4, t_1)$

Default path assignment

	t_1	t_2	t_3	t_4	
		$a_{1,3}$	$a_{1,2}$	d_1	$\langle Q_1 \rangle = \langle a_{1,3}, a_{1,2} \rangle$
	$a_{2,4}$		d_2	$a_{1,2}$	$\langle Q_2 \rangle = \langle a_{2,4}, a_{1,2} \rangle$
	$a_{3,4}$	d_3		$a_{1,3}$	$\langle Q_3 \rangle = \langle a_{3,4}, a_{1,3} \rangle$
	d_4	$a_{3,4}$	$a_{2,4}$		$\langle Q_4 \rangle = \langle a_{3,4}, a_{2,4} \rangle$

ADFs construction: 4 MRRs

Algorithm Flow



SA Optimization – Overview

- Topology planarization (planar graph drawing)
  - Planarize the topology from ADF sequences to get **accurate** maximum insertion loss
  - Include the ADF placement and the waveguide routing
- SA optimization objective
  - Find a suitable ADF placement (good **“relative position”** between nodes and ADFs)
  - Route waveguides with minimal crossings and thus results in a lower SA cost

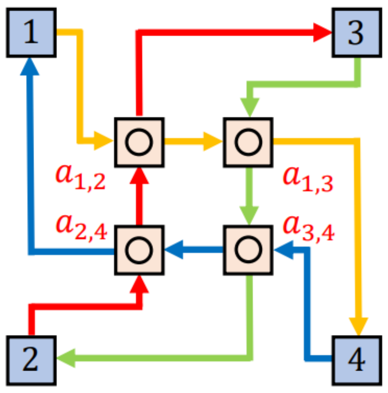
$\langle Q_1 \rangle = \langle a_{1,3}, a_{1,2} \rangle$

$\langle Q_2 \rangle = \langle a_{2,4}, a_{1,2} \rangle$

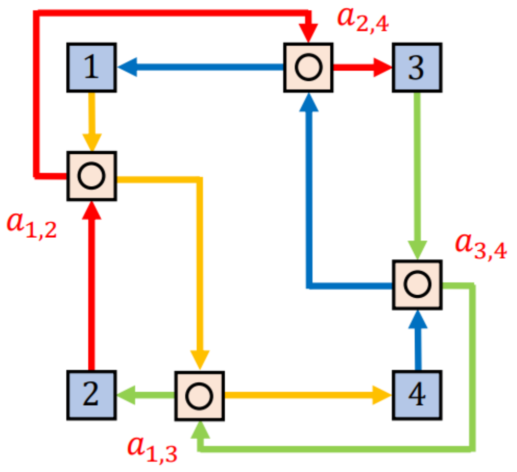
$\langle Q_3 \rangle = \langle a_{3,4}, a_{1,3} \rangle$

$\langle Q_4 \rangle = \langle a_{3,4}, a_{2,4} \rangle$

ADFs sequences



Initial placement

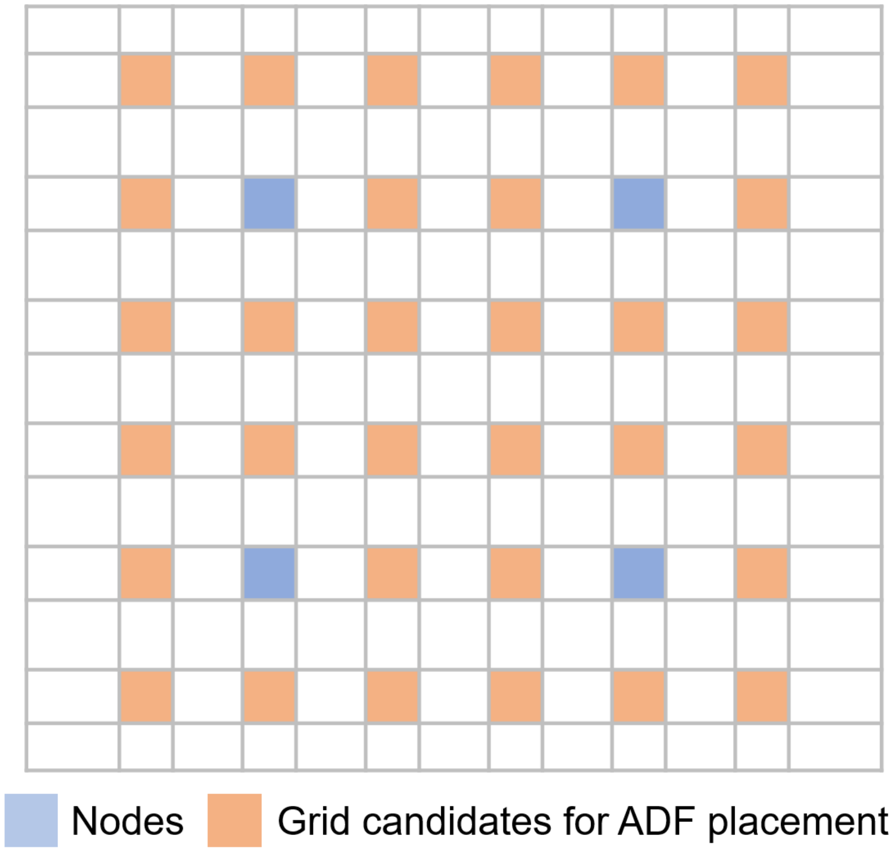


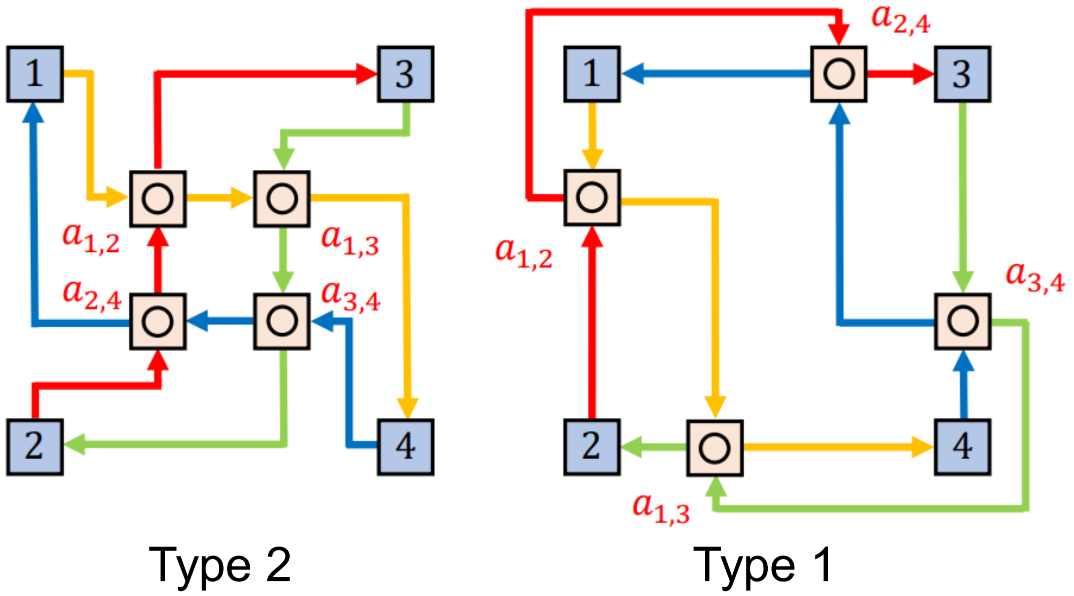
Optimized solution

Nodes ADFs Waveguides

Grid Construction

- Explore possible positions for ADF to utilize the advantages of crossing-free PSE structures (Type 1 ADF)
- Consider the efficiency of finding a solution (placement and routing) from the SA neighboring structure





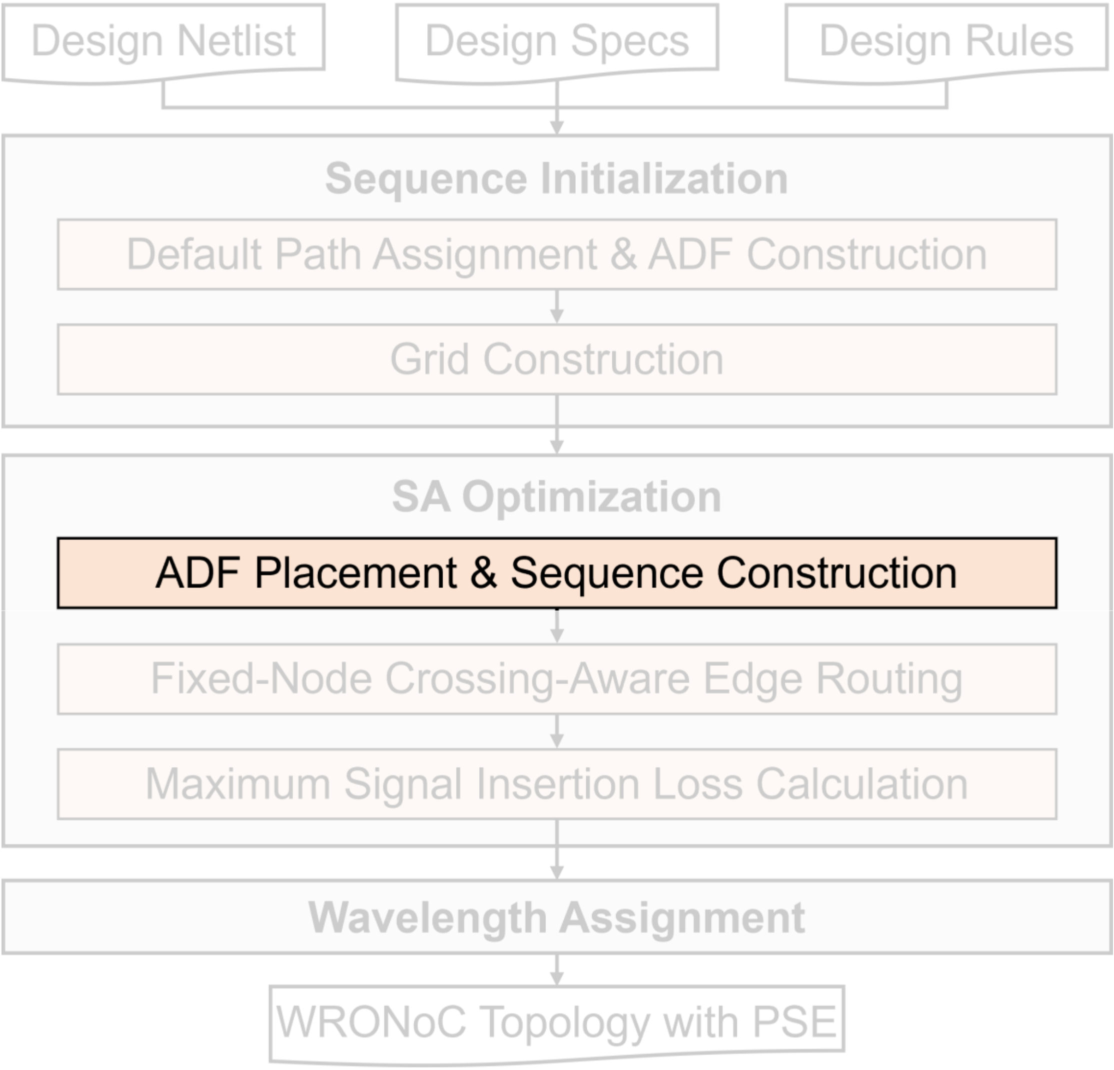
SA Optimization

- Two operations
  - Op1: Move an ADF to a new empty grid
  - Op2: Swap two ADF grids
- In the SA cost function, consider
  - Accurate maximum insertion loss cost (I)
  - Estimated wavelength usage cost (W)
- SA cost of an ADF placement p can be calculated as:

$$\phi(p) = \alpha I^* + \beta W^*$$

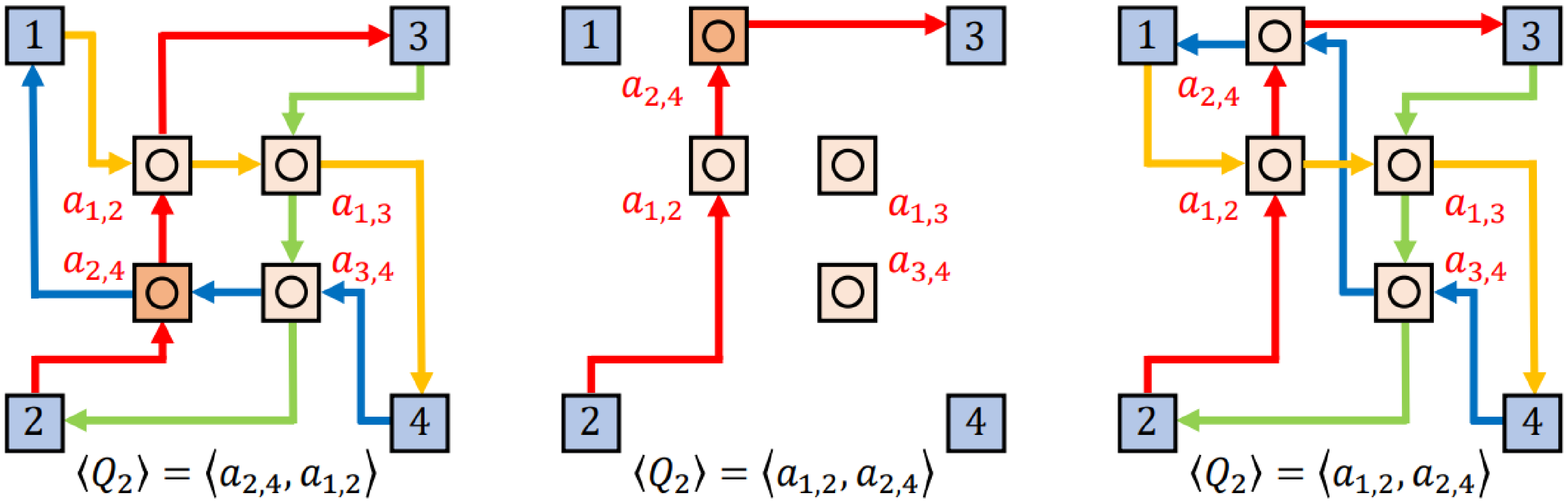
  - I^* and W^* are normalized individual costs
  - α and β are weights of the costs

Algorithm Flow



ADF Placement and Sequence Construction

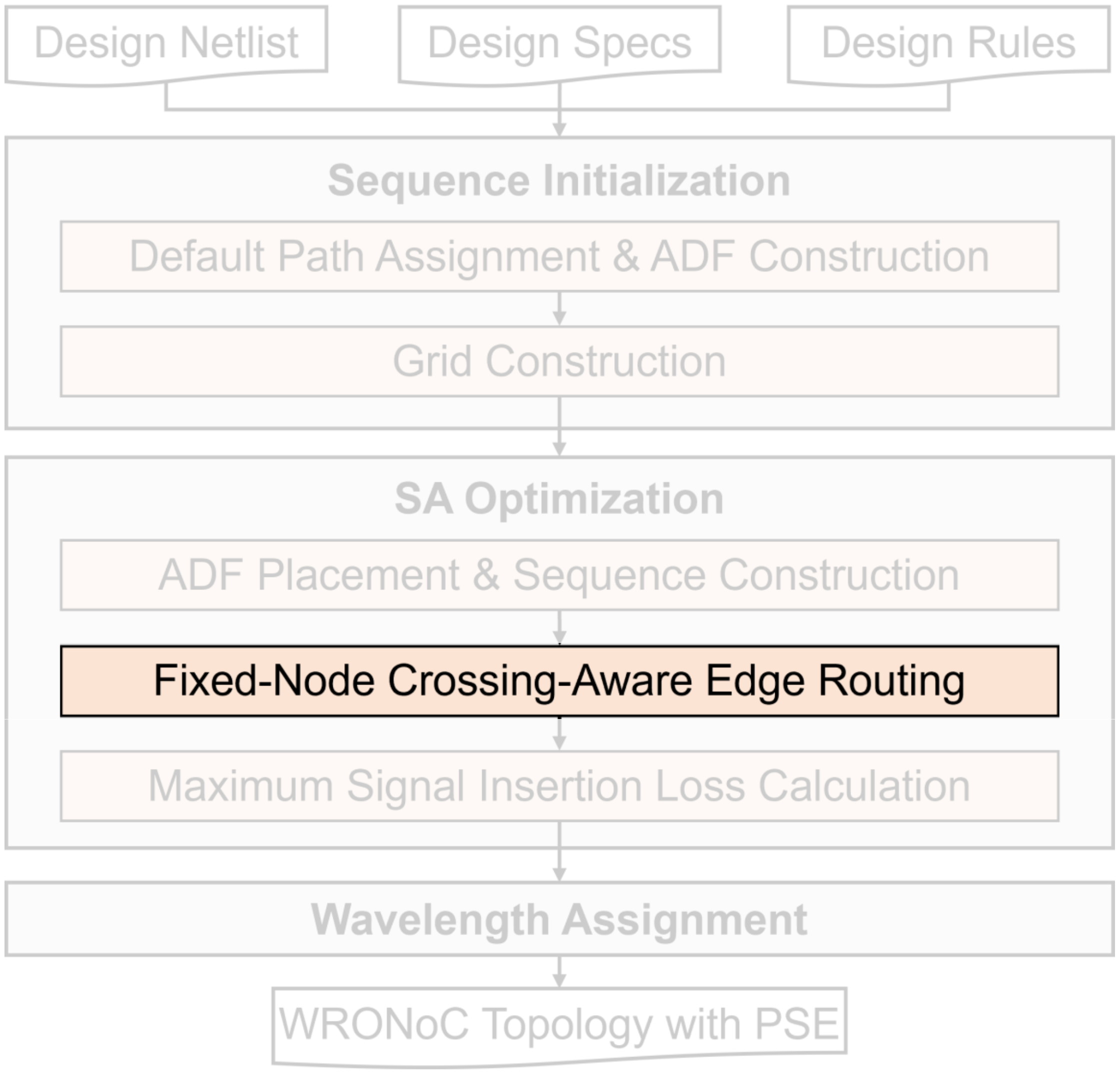
- Initial placement
  - Place ADF to the candidate grid according to its default paths' sources and targets
- Sequence construction
  - Whenever a new placement is obtained, we reorder the sequences to facilitate a smooth connection from the source to the target



Initial placement

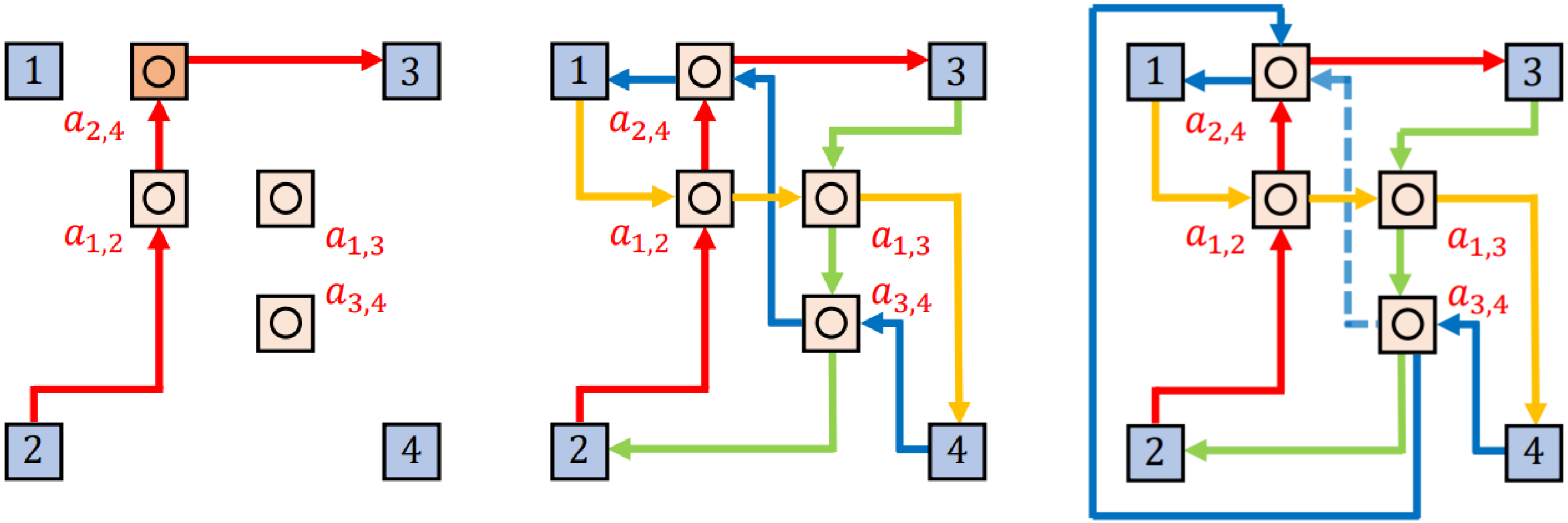
Move the ADF $a_{2,4}$ and reorder the sequences

Algorithm Flow



Fixed-Node Crossing-Aware Edge Routing

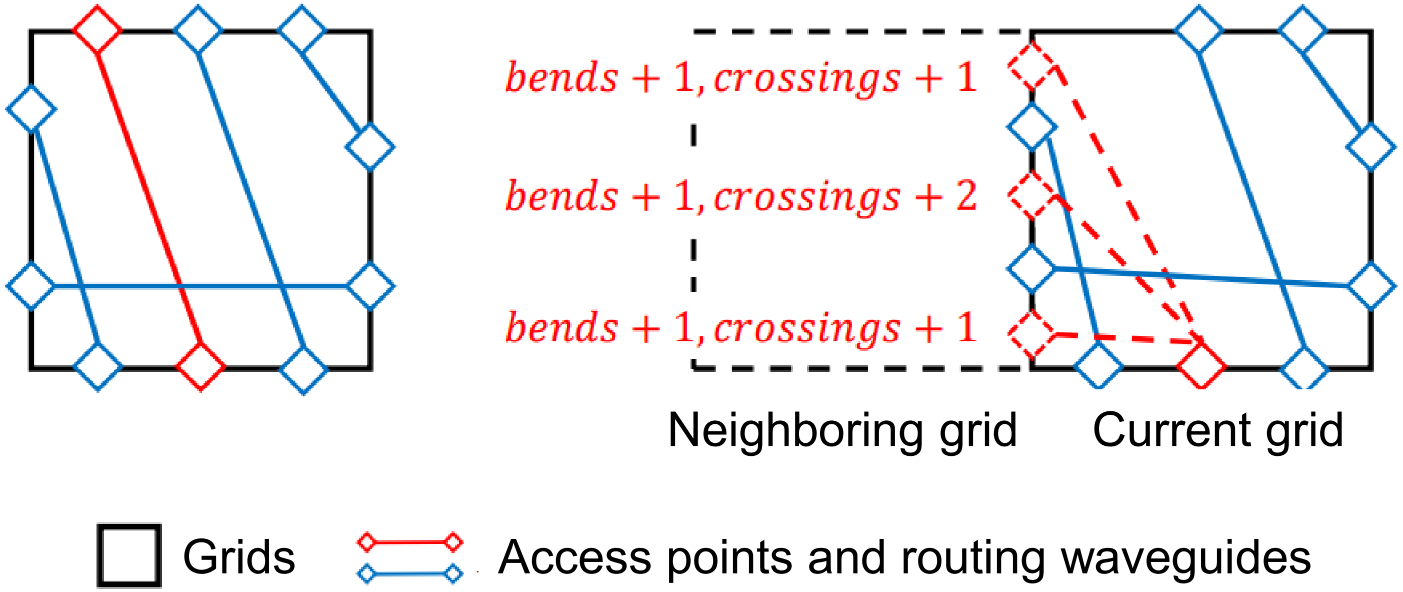
- Objective
  - Build waveguides according to the placement and the ADF sequence order
  - Route the waveguides with minimal crossings
- Fixed-node crossing-aware routing
  - Routing order: non-decreasing order of the grid distance
  - Apply crossing-aware A*-routing to route waveguides



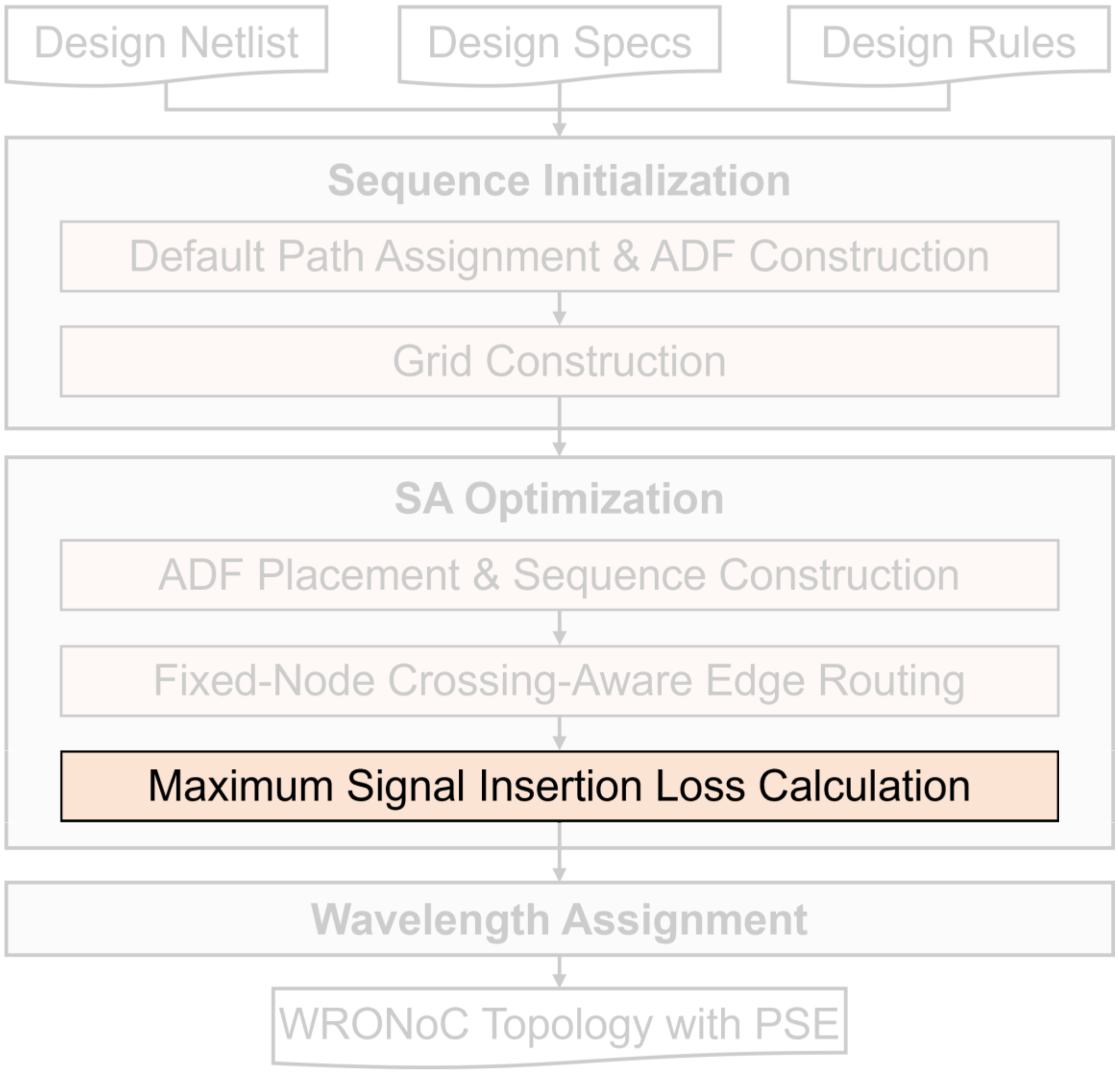
Finding a path that goes around the periphery of the topology

Crossing-Aware A*-Routing

- Cost function $f(n) = g(n) + h(n)$
  - $g(n)$: path cost, bending cost, and crossing cost
  - $h(n)$: the Manhattan distance to the target grid
- Features
  - Generate access points at the edges of the grids to form the waveguides
  - Calculate the crossing numbers with the access points

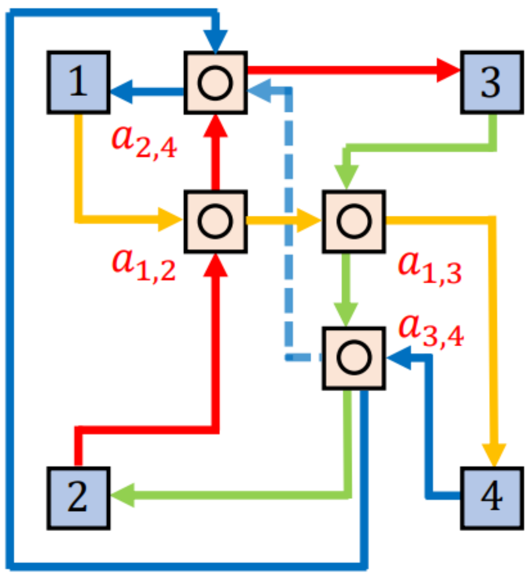


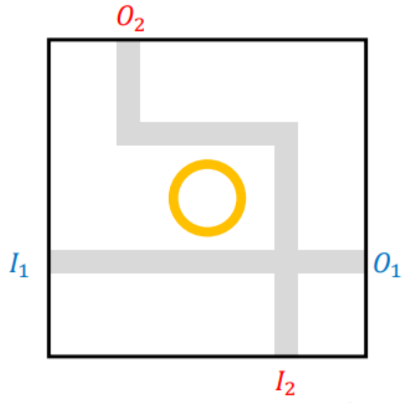
Algorithm Flow



Maximum Signal Insertion Loss Calculation

- Calculate accurate insertion loss for each signal
  - **Drop loss** and **through loss** can be calculated by tracing the **ADF sequences**
  - **Crossing loss** includes the crossing in the waveguides and within the ADFs
- Crossing assignments for type 2 and type 3 ADFs
  - Calculate **fixed number** of crossings for signals
  - Assign the **two-crossing** to the path with a smaller insertion loss

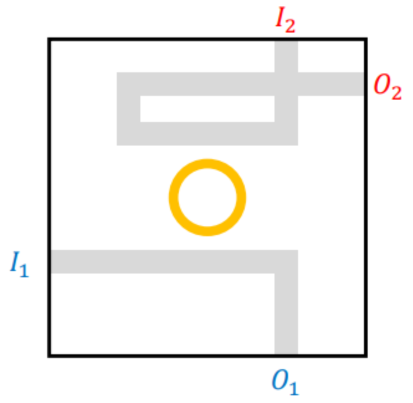




$I_1 \rightarrow O_2: 0.7dB$

$I_2 \rightarrow O_1: 0.6dB$

	O_1	O_2
I_1	1	0
I_2	2	1

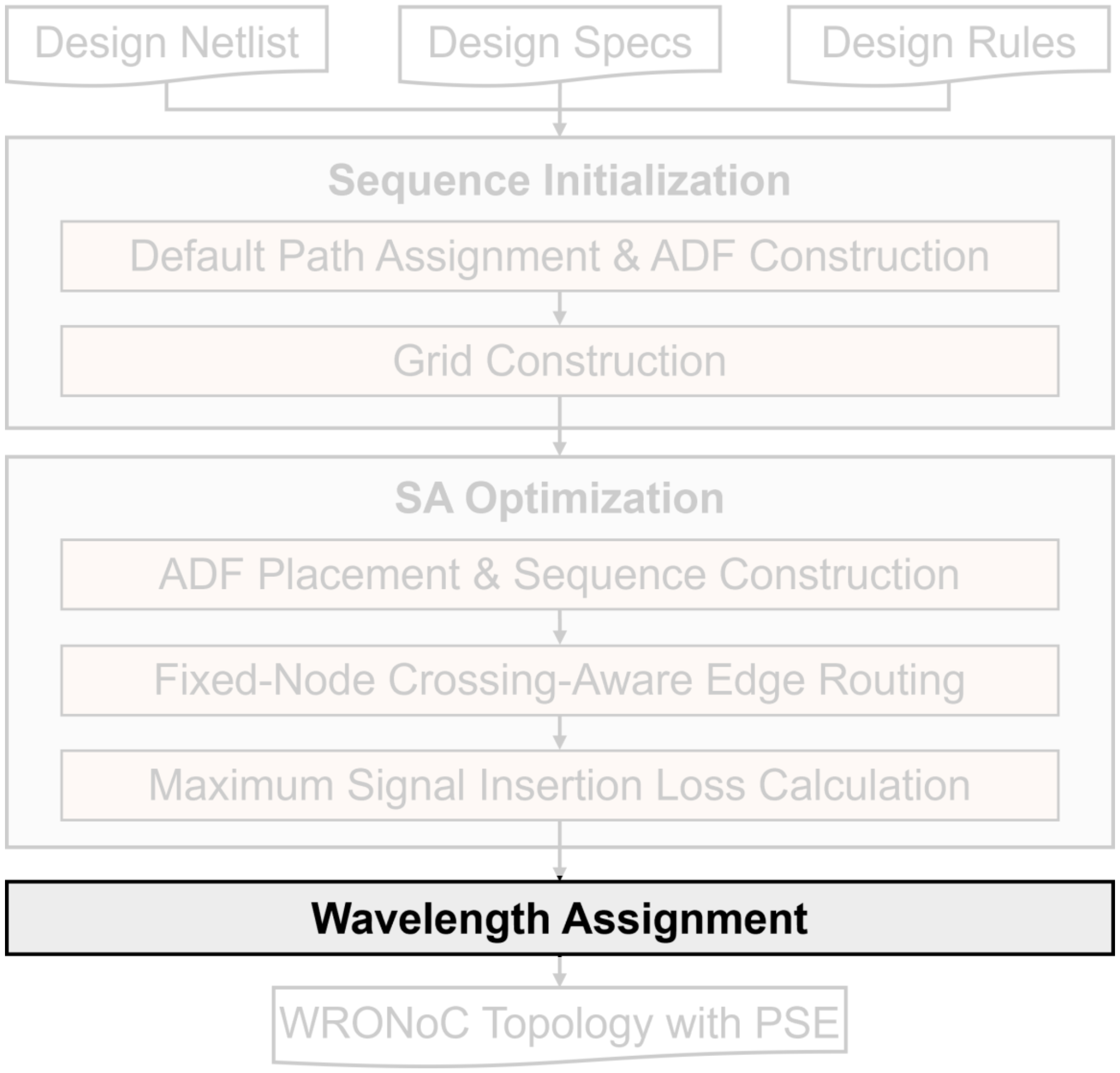


$\max(I_1 \rightarrow O_1): 0.7dB$

$\max(I_2 \rightarrow O_2): 0.6dB$

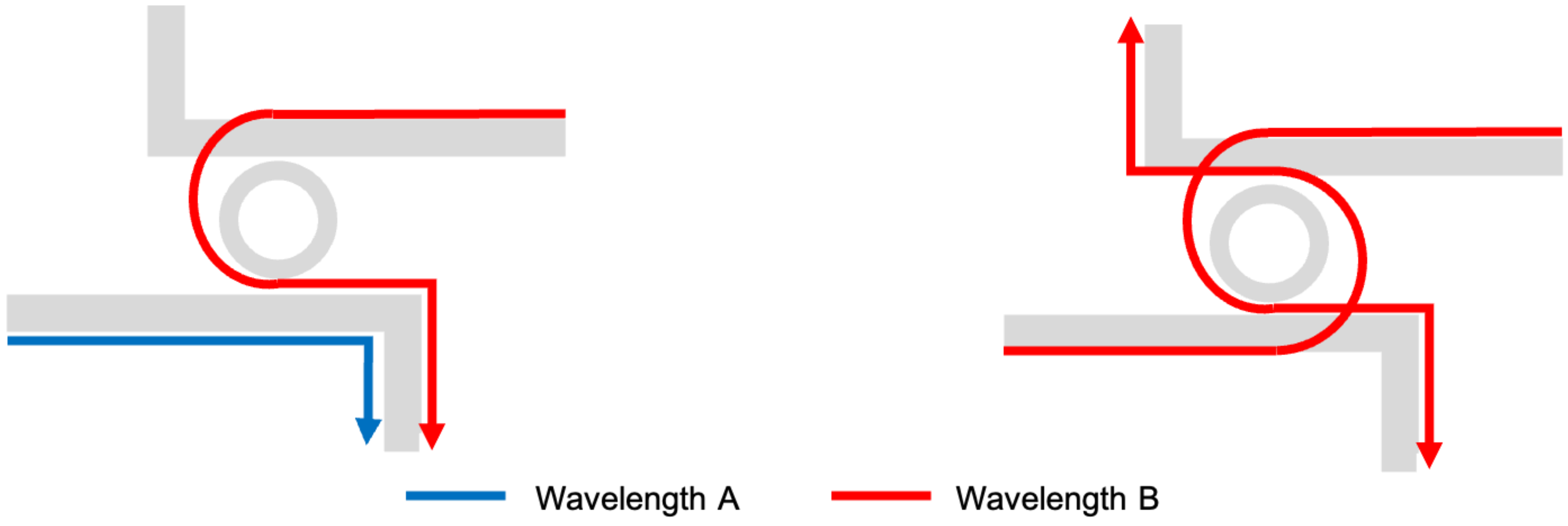
	O_1	O_2
I_1	0	1
I_2	1	2

Algorithm Flow



Wavelength Assignment

- Assign a wavelength to each signal to avoid conflict
  - Two signals using the same waveguide must use different wavelengths
  - Two signals using the same ADF to drop must use identical wavelength
- Modify the previous ILP approach [Lu et al., TCAD'22]
  - Conflict graph creation
  - ILP graph coloring



Outline

- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

Experimental Setup

- Environmental settings
  - C++ programming language
  - AMD Ryzen 2.9GHz workstation with 128GB memory
  - Gurobi ILP solver

- Comparison targets for customized topology design
  - FAST [Xiao et al., DATE'21], SA-FTONoC [Chen et al., ICCAD'23]

- Insertion loss value
  - $0.5dB$ per drop, $0.04dB$ per cross, and $0.005dB$ per through

Experimental Results – Customized (1/2)

- Compared to FAST and SA-FTONoC
  - 44.5% and 36.6% reduction in MRR usages

Benckmark	#Node	#Signal	FAST		SA-FTONoC		Ours	
			#MRR	#WL	#MRR	#WL	#MRR	#WL
Case1	8	44	36	7	36	7	20	7
Case2	12	26	24	7	21	7	13	7
Case3	12	20	14	5	12	5	10	5
Case4	16	22	19	7	11	7	10	7
Case5	8	48	40	6	40	6	20	6
Case6	8	24	20	6	20	6	12	6
Case7	8	24	24	6	20	6	12	6
Comp.	**-**	**-**	**1.80**	**1.00**	**1.58**	**1.00**	**1.00**	**1.00**

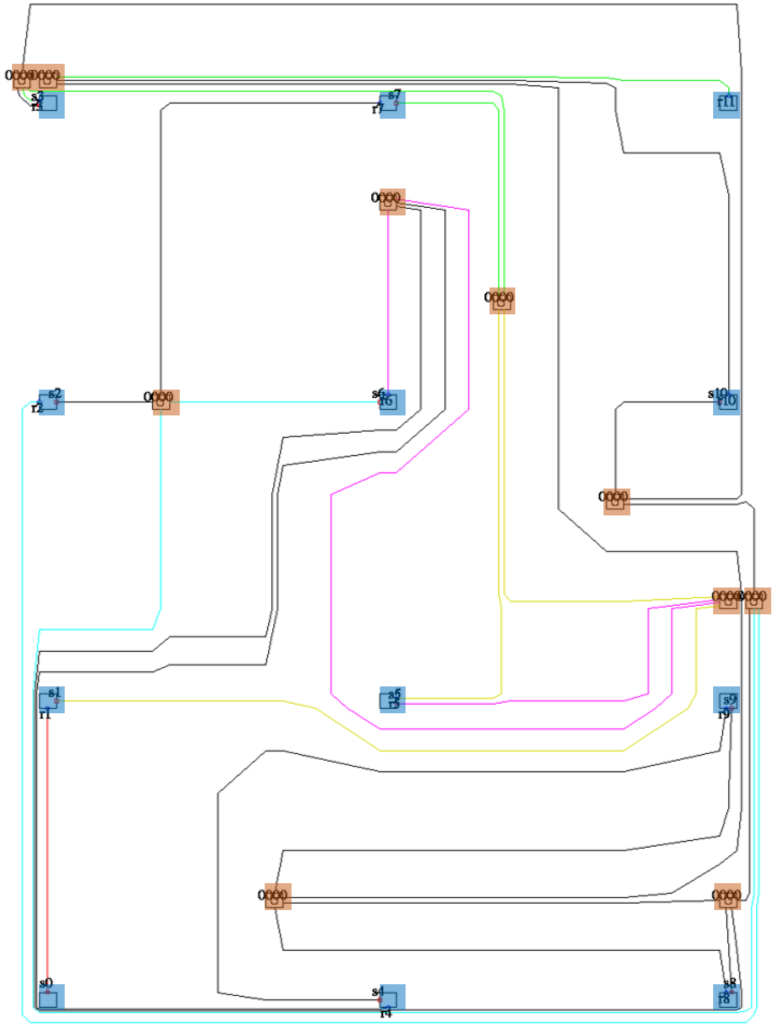
Experimental Results – Customized (2/2)

- Compared to FAST and SA-FTONoC
  - 24.7% and 15.7% reduction in maximum insertion loss
  - No waveguide crossings in 4 out of 7 cases
  - No crossings within ADF in 3 out of 7 cases

Benckmark	FAST		SA-FTONoC		Ours	
	Max.IL (dB)	Time (s)	Max.IL (dB)	Time (s)	Max.IL (dB)	Time (s)
Case1	1.055	0.04	0.945	11.60	0.810	26.30
Case2	0.725	1.25	0.725	12.90	0.530	16.60
Case3	0.755	1.65	0.635	3.70	0.525	14.40
Case4	0.760	1.50	0.725	0.70	0.530	44.30
Case5	0.980	0.04	0.800	11.60	0.850	48.10
Case6	0.730	0.31	0.680	3.90	0.605	16.80
Case7	0.840	0.03	0.680	2.80	0.605	27.60
Comp.	**1.33**	**0.03**	**1.18**	**0.28**	**1.00**	**1.00**

Topology Layout for Case3

- The topology is totally planarized (12 nodes, 20 signals, 10 ADFs)



	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}
s_0		λ_0										
s_1				λ_1		λ_0	λ_2					
s_2								λ_0				
s_3					λ_0				λ_1			
s_4										λ_0		
s_5							λ_0					
s_6			λ_0					λ_2	λ_3			
s_7				λ_0					λ_2			
s_8					λ_3		λ_1					
s_9									λ_0		λ_2	λ_1
s_{10}									λ_4			
s_{11}												

Nodes ADFs Default path assignment

Outline

- Introduction
- Preliminaries
- Proposed Approach
- Experimental Results
- Conclusion

Conclusion

- We have proposed a fully automated topology design flow
  - Utilize PSE structures for customized WRONoC topologies
- Our ADF sequence model
  - Expand the solution space and leverages the advantage of crossing-free PSE structure
- Our design flow
  - Thoroughly considers the physical layout information
  - Minimizes the waveguide crossings by the fixed-node crossing-aware edge routing
- Experimental results have shown that our method substantially outperforms the existing design flows

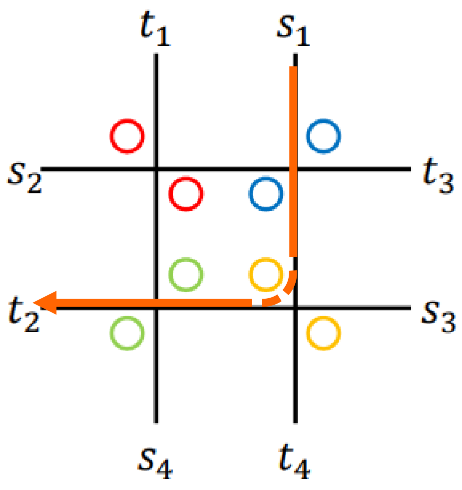
Thank You!

National Taiwan University

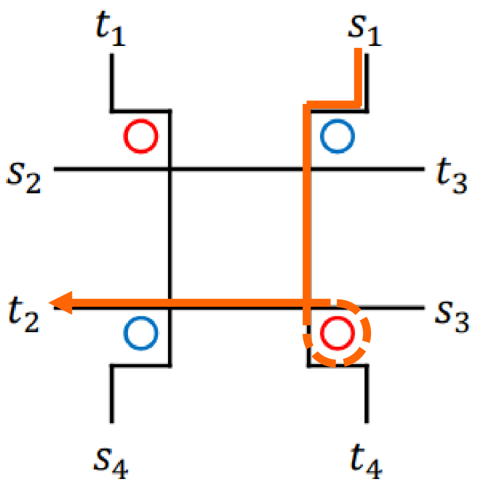
Backup Sildes

Challenges of Adopting the PSE Structure

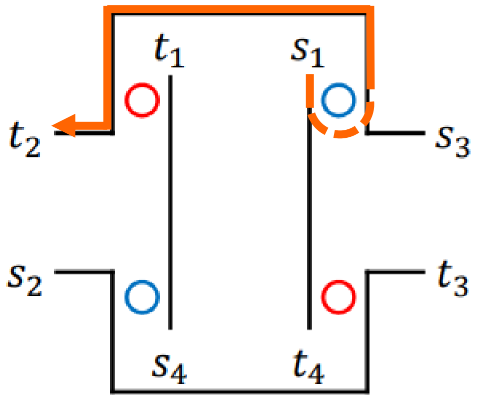
- Change of transmission direction
  - Orientation of the two outputs is different to the CSE structure
  - Grid-like topology inevitably introduces extra waveguide bends and crossings
- Trade-off between MRR usage and maximum insertion loss
  - When waveguide crossing occurs, one of the drop signals must experience two extra waveguide crossings during the MRR drop ($s_1 \rightarrow t_2$ in HASH)
  - May lead to a higher maximum insertion loss



GWOR: 0.600 dB



HASH: 0.670 dB



Our 4x3 one: 0.505 dB