

**UC Davis**

**UC Davis Electronic Theses and Dissertations**

**Title**

Silicon Photonic Switching Fabrics and Synaptic Interconnections for High Performance Computing and Neuromorphic Computing Systems

**Permalink**

<https://escholarship.org/uc/item/4730p864>

**Author**

Xiao, Xian

**Publication Date**

2021

Peer reviewed|Thesis/dissertation

Silicon Photonic Switching Fabrics and Synaptic Interconnections for  
High Performance Computing and Neuromorphic Computing Systems

By

XIAN XIAO  
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical and Computer Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

---

S. J. Ben Yoo, Chair

---

Erkin Şeker

---

Josh Hihath

Committee in Charge

2021

Copyright © 2021 by

Xian Xiao

*All rights reserved.*

*To my parents, Zhiting Xiao and Lingying Wu,  
my grandparents, Huanru Xiao, Shuiyu Wu, Shengquan Wu, and Hezhi Zhu,  
and my love, Weier Guo*

# CONTENTS

List of Figures . . . . .	vi
List of Tables . . . . .	xvi
Abstract . . . . .	xvii
Acknowledgments . . . . .	xix
<b>1 Background and Motivation</b>	<b>1</b>
1.1 End of the Moore’s Law and Optical Interconnects . . . . .	1
1.2 Applications and Challenges of Optical Interconnects in High Performance Computing and Neuromorphic Computing Systems . . . . .	3
1.3 Silicon Photonic Technology and Heterogeneous Integration . . . . .	6
1.4 Scope of This Dissertation . . . . .	7
<b>2 Si-LIONS For Scalable All-to-All Optical Interconnects</b>	<b>8</b>
2.1 Principle and Architecture . . . . .	8
2.1.1 Arrayed Waveguide Grating Router (AWGR) . . . . .	8
2.1.2 Silicon Photonic Low-Latency Interconnect Optical Network Switch (Si-LIONS) . . . . .	10
2.2 Impact of Intra-Band Crosstalk on Scalability of All-to-All Optical Interconnects . . . . .	11
2.2.1 Crosstalk penalty for OOK modulation format . . . . .	11
2.2.2 Worst-Case Crosstalk Penalty for AWGR . . . . .	14
2.2.3 Worst-Case Crosstalk Penalty for MRR Crossbar . . . . .	15
2.2.4 Comparison Based on the State-of-the-Art Parameters . . . . .	20
2.2.5 Experimental measurements of the crosstalk penalty of AWGR . . . . .	21
2.3 Foundry-Enabled Si-LIONS Using SiN AWGR and SiPh Transceivers . . . . .	24
2.3.1 $8 \times 8$ Si-LIONS Chip Design, Layout and Fabrication . . . . .	24
2.3.2 SiN AWGR Characterization . . . . .	26
2.3.3 Microdisk Modulator Characterization . . . . .	26

2.3.4	Wavelength Routing Experiments . . . . .	28
2.3.5	Initial Designs and Results on $16 \times 16$ and $32 \times 32$ SiN AWGR . . . . .	31
2.4	Experimental Demonstration of a 64-Port Thin-CLOS Wavelength-Routing System . . . . .	32
2.4.1	Thin-CLOS AWGR Architecture for Datacenter Switching . . . . .	32
2.4.2	Design, Fabrication, and Experimental Demonstration of a 64-port Thin-CLOS . . . . .	35
2.4.3	Crosstalk and Power Budget Analysis . . . . .	42
<b>3</b>	<b>Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable All-to-All Optical Interconnects</b>	<b>45</b>
3.1	System Demonstration of Flex-LIONS With Off-The-Shelf Components . . . . .	46
3.1.1	Architecture and experimental setup . . . . .	46
3.1.2	Measurement results: 16-Port Flex-LIONS . . . . .	48
3.1.3	Scalability study: $32 \times 32$ all-to-all interconnect experiment . . . . .	50
3.2	Silicon Photonic Flex-LIONS for Bandwidth Reconfigurable Optical Interconnects . . . . .	52
3.2.1	SiPh Flex-LIONS Architecture and Principle . . . . .	52
3.2.2	Comparison With the State-of-the-Art Reconfigurable Switching Fabrics . . . . .	54
3.2.3	Design, Fabrication, and Single Component Characterization of $8 \times 8$ Silicon Photonic Flex-LIONS . . . . .	57
3.2.4	Experimental Demonstration of Optical Reconfiguration . . . . .	60
3.2.5	Scalability of Flex-LIONS . . . . .	63
3.3	Multi-FSR Silicon Photonic Flex-LIONS Module for Bandwidth-Reconfigurable All-to-All Optical Interconnects . . . . .	67
3.3.1	Principle of Multi-FSR Flex-LIONS . . . . .	69
3.3.2	Design, Fabrication, and Packaging of $8 \times 8$ Silicon Photonic Flex-LIONS Module . . . . .	69

3.3.3	Experimental Demonstration of Bandwidth-Reconfigurable All-to-All Optical Interconnects . . . . .	73
3.4	O-Band Silicon Photonic Flex-LIONS With ns Switching Speed . . . . .	83
3.4.1	Single Elements Design . . . . .	83
3.4.2	O-band 16×16 SiPh Flex-LIONS Design and Layout . . . . .	87
3.4.3	O-band 16×16 SiPh Flex-LIONS Fabrication Process . . . . .	90
<b>4</b>	<b>Scalable and Compact Tensorized Photonic Neural Networks</b>	<b>92</b>
4.1	Tensor-Train Decomposed Synaptic Interconnects . . . . .	92
4.1.1	Conventional PNN Architecture . . . . .	94
4.1.2	Tensorized PNN Architecture . . . . .	95
4.1.3	Comparison Between Tensorized and Conventional PNNs . . . . .	97
4.1.4	Example: 1024×1024 Tensorized PNN in a Single Plane . . . . .	99
4.2	Silicon Photonic Tensor Core Design and Layout . . . . .	100
<b>5</b>	<b>Light Sources for Computing Systems</b>	<b>102</b>
5.1	Transfer-Printed III-V-on-Si Quantum Dot Lasers . . . . .	102
5.1.1	III-V-on-Si Hybrid Optical Amplifier Design . . . . .	102
5.1.2	Transfer Printing Fabrication Process . . . . .	105
5.2	Optical in-Phase-Quadrature Modulator Based on Injection-Locked VCSEL Phase Array . . . . .	110
5.2.1	Principle and Experimental Setup . . . . .	112
5.2.2	OIL VCSEL Properties: Bandwidth Enhancement, AM and PM . . . . .	113
5.2.3	Demonstration of IQ Modulator . . . . .	116

## LIST OF FIGURES

1.1	(a) All classical technological drivers to further performance improvements end in approximately 2025. (b) Three potential paths forward to realize continued performance improvements for Si microelectronics. (Courtesy of John Shalf, Lawrence Berkeley National Laboratory) . . . . .	2
1.2	(a) A fat tree topology using electronic switches at the core and at the aggregation edges of the network. (b) A flattened optically interconnected network example utilizing a passive optical fabric or a reconfigurable optical switching fabric with electronic switches at the edges. . . . .	3
1.3	(a) PNN architecture consisting of an input neuron layer, many hidden neuron layers, an output neuron layer, and synaptic interconnections. (b) Mars chip from Lightmatter with a photonic core can implement $64 \times 64$ photonic synaptic interconnections (Courtesy of Lightmatter). . . . .	5
2.1	All-to-all optical interconnects based on: (a) waveguides (b) wavelength-routing device. . . . .	8
2.2	(a) Schematic of AWGR. (b) Wavelength routing function of AWGR. (c) Wavelength routing table of AWGR. . . . .	9
2.3	Si-LIONS architecture. . . . .	11
2.4	(a) Power penalty versus varied crosstalk for AWGR. (b) Number of scalable nodes versus varied crosstalk with power penalty constraint of 1, 3, 6 dB. . . . .	14
2.5	Schematic of: (a) MRR (b) Conventional MRR-based crossbar architecture.	15
2.6	(a) Schematic of uniform-loss MRR-based crossbar architecture. (b) Comparison of the number of MRRs between conventional crossbar (dashed line) and uniform-loss crossbar (solid line). . . . .	17



2.7	(a) Power penalty versus varied crosstalk for conventional crossbar (solid lines) and uniform-loss crossbar (dashed lines). (b) Number of scalable nodes versus varied crosstalk with power penalty constraint of 1 dB. (c) Number of scalable nodes versus varied crosstalk with power penalty constraint of 3 dB. . . . .	19
2.8	Comparison of BER floor among the state-of-the-art silicon conventional crossbar, silicon uniform-loss crossbar, silicon AWGR, and silicon nitride AWGR. . . . .	21
2.9	Experiment testbed using 32-port AWGR with 100 GHz spacing AWGR. EDFAs: erbium doped fiber amplifier; VOA: variable optical attenuator. . . . .	21
2.10	Picture of the experiment testbed. . . . .	22
2.11	(a) Measured crosstalk from input port 2-32 to output port 32 at $\lambda_{1-32}$ . (b) Measured BER with fixed decision-threshold setting for random polarization (green) and aligned polarization. . . . .	23
2.12	(a) Optical microscope picture of the $8 \times 8$ Si-LIONS system with SiN AWGR and SiPh transmitters and receivers. Zoom-in pictures of (b) a silicon microdisk modulator, (c) the $8 \times 8$ SiN AWGR and (d) a silicon microdisk filter and Ge photodetector pair. . . . .	24
2.13	Measured transmission spectra of the $8 \times 8$ SiN from (a) central waveguide input (input waveguide 5) and (b) side waveguide input (input waveguide 1). . . . .	25
2.14	Measured transmission spectra of (a) a 4-channel modulator array with different resonance wavelength and (b) a 9-channel modulator array with same resonance wavelength. . . . .	27
2.15	(a) Measured transmission spectra of a microdisk modulator upon different heating power and (b) Measured and fitted heater efficiency. . . . .	28
2.16	Measured transmission spectra of a microdisk modulator upon different bias voltage. Inset: 10 Gb/s eye diagram of a modulator upon 1Vpp swing. . . . .	29
2.17	Experimental setup for the routing demonstration on the fabricated chip. . . . .	29

2.18	Simulated eye diagrams for (a) electrical input and (b) Ge PD output. Measured eye diagrams for (c) input 5 to output 5, (d) input 5 to output 1, (e) input 1 to output 1 and (f) input 1 to output 5. (g) Measured BER curves as a function of received power. . . . .	30
2.19	Optical microscope picture of: (a) a $16 \times 16$ SiN AWGR (b) a $32 \times 32$ SiN AWGR. Measured transmission spectra from the central input of: (c) a $16 \times 16$ SiN AWGR (d) a $32 \times 32$ SiN AWGR. . . . .	31
2.20	(a) $N$ -port Thin-CLOS AWGR architecture. (b) Table shows a comparison in terms of number of wires, number of AWGRs, and number of wavelengths for all-to-all implementation with optical fibers only, one AWGR and with Thin-CLOS. (c) Use of Thin-CLOS AWGR with WDM TRXs (passive LIONS). (d) Use of Thin-CLOS AWGR with tunable TRXs (active LIONS). . . . .	33
2.21	Experiment setup using Enablence 32-port OSD based on single 100 GHz spacing AWGR. PM, power monitor; PBS, polarization beam splitter; EDFA, erbium-doped fiber amplifier; VOA, variable optical attenuator; PC, polarization controller. . . . .	36
2.22	Crosstalk measurements at output 32 from input 1 at $\lambda_{1-32}$ (worst-case loss from input 1). . . . .	38
2.23	Analytical results of power penalty as a function of AWGR crosstalk for the different AWGR port count with random polarizations (polarization state is unknown) and worst-case aligned polarizations (polarizations aligned in parallel). . . . .	38
2.24	End-to-end link experiment with 32 signals at the same wavelength and worst-case polarization alignment for: (a) Input port 1 to output port 32; (b) Input port 17 to output port 31; Input port 1 to output port 31. . . .	39
2.25	Connectivity between MTP1 and nodes 1 to 4. The MTP cable carries two eight-fiber ribbons, which break out into single LC fiber cables for connection with WDM muxes and demuxes at each node. . . . .	40

2.26	(a) Experimental testbed for verifying Thin-CLOS functionality. (b) Picture of the fabricated Thin-CLOS system in 1U rack size. (c) Measured BER with different input and output node combinations. . . . .	43
3.1	Modern HPC systems with heterogeneous processor and memory nodes. .	46
3.2	(a) State 1: Fully all-to-all topology based on AWGR. (b) State 2: Partially all-to-all topology in which some pairs of nodes can be reconfigured to directly connect using all the wavelengths. . . . .	47
3.3	$N$ -node Flex-LIONS using two $(N+m)$ -node spatial switches and one $N \times N$ AWGR . . . . .	48
3.4	Experimental setup of 16-node Flex-LIONS. EDFA: erbium doped fiber amplifier; MUX: AWG multiplexer; DeMUX: AWG demultiplexer; Inset is the optical spectrum after the multiplexer. . . . .	49
3.5	Picture of the testbed of 16-node Flex-LIONS. . . . .	49
3.6	Topology diagram and BER curves of Flex-LIONS: (a) State 1, fully all-to-all interconnects ( $1 \times$ bandwidth for all interconnected nodes); (b) State 2, partially all-to-all interconnects ( $16 \times$ bandwidth for traffic between selected nodes). . . . .	50
3.7	(a) Power penalty of 32-node all-to-all interconnects under worst-case crosstalk condition. (b) Power penalty at $10^{-12}$ under different OSNR value. (c) BER curves with different OSNR value for output port 1 at wavelength channel 22 (1552.52 nm). . . . .	51
3.8	(a) $N \times N$ Flex-LIONS architecture with $N \times N$ AWGR, $b$ MRR add-drop filters at each input and output ports, and $N \times N$ multi-wavelength MRR crossbar switch. (b) Schematic of multi-wavelength MRR crossbar switch. (c) Schematic of the wavelength relation between the WDM channels and the resonances of MRR add-drop filters and multi-wavelength MRR crossbar switch. . . . .	53
3.9	(a) Cross section of the multi-layer platform. (b) Schematic of the $8 \times 8$ SiPh Flex-LIONS layout. . . . .	56

3.10	(a) Schematic of edge coupler design. (b) Schematic of MMI waveguide crossing design. (c) Simulated insertion loss of MMI waveguide crossing with various taper and multimode region lengths. (d) Schematic of Si-to-SiN evanescent coupler design. (e) Simulated insertion loss of evanescent coupler with various gap values. . . . .	57
3.11	(a) Fabrication flowcharts for the $8 \times 8$ SiPh Flex-LIONS. (b) Microscope image of the fabricated $8 \times 8$ SiPh Flex-LIONS ( $N = 8$ , $b = 3$ ) chip. (c) Microscope image of MRR add-drop filter. (d) Microscope image of multi-wavelength MRR switch. . . . .	59
3.12	Transmission spectra of (a) $8 \times 8$ AWGR from input port 4 and (b) through and drop ports of multi-wavelength MRR switch with different TO tuning power. Thermal tuning efficiency of (c) multi-wavelength MRR switch and (d) MRR add-drop filter. . . . .	61
3.13	Experimental setup. TLD: tunable laser diode; EDFA: erbium-doped fiber amplifier; MZ: Mach Zehnder; DAC: digital to analog converter; DUT: device under test; VOA: variable optical attenuator; PD: photodetector; EA: error analyzer. . . . .	61
3.14	(a) Transmission spectrum from input port 4 to output port 5 before re-configuration. (b) BER curves of input port 4 to different output ports. (c) Eye diagrams of input port 4 to output port 8, 1, 4, and 5 before reconfiguration. . . . .	62
3.15	(a) Transmission spectrum from input port 4 to output port 5 after re-configuration. (b) BER curves of input port 4 to output port 5. (c) Eye diagrams of input port 4 to output port 5 using $\lambda_3$ , $\lambda_5$ , $\lambda_6$ , and $\lambda_7$ after reconfiguration. . . . .	64

3.16	(a) Simulated power penalty versus varied crosstalk for AWGR. (b) Calculated number of scalable nodes versus varied crosstalk with power penalty constraint of 1, 3, 6 dB. (c) End-to-end link experiment with eight input signals at the same wavelength and aligned polarization for the worst-case crosstalk scenario. . . . .	65
3.17	(a) Schematic of $N \times N$ Thin-CLOS Flex-LIONS architecture ( $N = M \times W$ ). (b) Layout of $16 \times 16$ Thin-CLOS Flex-LIONS with four $8 \times 8$ Flex-LIONS ( $N = 16, M = 2, W = 8$ ). . . . .	66
3.18	Heterogeneous processor and memory nodes with: (a) LIONS (all-to-all interconnects); (b) Single-FSR Flex-LIONS (bandwidth reconfigurable interconnects); (c) Multi-FSR Flex-LIONS (bandwidth-reconfigurable all-to-all interconnects). (d) $N \times N$ multi-FSR Flex-LIONS architecture with $N \times N$ AWGR, $b$ MRR add-drop filters at each input and output ports, and $N \times N$ Beneš MZS network. FSR <sub>0</sub> is used for maintaining all-to-all interconnectivity and FSR <sub>1</sub> is used for bandwidth reconfiguration. . . . .	68
3.19	(a) Cross section of the multi-layer platform. (b) $8 \times 8$ SiPh Flex-LIONS layout. . . . .	70
3.20	(a) Design of MMI based waveguide crossing. (b) Layout of MRR add-drop filter. (c) Layout of $2 \times 2$ MZ switching element (arm length not to scale). . . . .	71
3.21	(a) Fabrication flow charts for the $8 \times 8$ SiPh Flex-LIONS. (b) Microscope image of the fabricated $8 \times 8$ SiPh Flex-LIONS ( $N = 8, b = 3$ ) chip. (c) Microscope image of MRR add-drop filter. (d) Microscope image of part of $2 \times 2$ MZS. . . . .	72
3.22	(a) Layout of the co-designed PCB. (b) Photograph of the integrated Flex-LIONS module with lid-less PM fiber arrays on a co-designed PCB. (Courtesy of Optelligent, LLC). . . . .	74

3.23	(a) Transmission spectra of $8 \times 8$ SiN AWGR from input port 4. (b) Linear fitting of the normalized transmission of Si MMI waveguide crossing for insertion loss calculation. . . . .	75
3.24	(a) Transmission spectra of through and drop ports of MRR add-drop filter with different TO tuning power. (b) TO tuning efficiency of MRR add-drop filter. (c) Transmission spectra of $2 \times 2$ MZS at different TO tuning power for the cross port and the bar port. . . . .	77
3.25	Experimental setup. SFP: small form pluggable; MUX: multiplexer; MZ: MachZehnder; EDFA: erbium-doped fiber amplifier; DAC: digital to analog converter; VOA: variable optical attenuator; DeMUX: demultiplexer; PD: photodetector; EA: error analyzer. . . . .	78
3.26	(a) Transmission spectrum from input port 4 to output port 8 before reconfiguration. (b) BER curves of all-to-all interconnects through FSR <sub>0</sub> before reconfiguration. (c) BER curves of all-to-all interconnects through FSR <sub>1</sub> before reconfiguration. (d) 25 Gb/s eye diagrams for back-to-back and selected input and output ports. . . . .	79
3.27	(a) Transmission spectrum from input port 4 to output port 8 after reconfiguration. (b) BER curves of all-to-all interconnects through FSR <sub>0</sub> after reconfiguration. (c) BER curves of input port 4 to output port 8 after reconfiguration ( $\lambda_8$ in FSR <sub>0</sub> , $\lambda_{10}$ , $\lambda_{12}$ , $\lambda_{14}$ , $\lambda_{16}$ in FSR <sub>1</sub> ). (d) Eye diagrams of input port 4 to output port 8 using $\lambda_8$ , $\lambda_{10}$ , $\lambda_{12}$ , $\lambda_{14}$ , and $\lambda_{16}$ after reconfiguration. . . . .	80
3.28	Time-domain optical response of the switch element. (a) Applied square-wave electrical drive signal. (b) Measured optical waveform for MRR add-drop filter. (c) Measured optical waveform for $2 \times 2$ MZS. . . . .	82
3.29	(a) Layout of the $16 \times 16$ 80-GHz-spacing SiN AWGR. (b) Measured transmission spectrum from the center input (input port 8) of the AWGR. . .	84
3.30	(a) Cross section of the designed O-band Si PIN junction. (b) Simulated excess loss with varied offset values. . . . .	84

3.31	Simulated ion range for (a) P <sup>++</sup> (b) N <sup>++</sup> Si using TRIM. . . . .	85
3.32	Carrier concentration profile under different forward bias voltages simulated by Silvaco Athena and Atlas. . . . .	85
3.33	(a) Simulated phase shift with different forward bias voltages and different phase shifter length. (b) Simulated insertion loss of the phase shifter with different forward bias voltages and different phase shifter length. . . . .	86
3.34	(a) O-band MRR add-drop filter structure. (b) Summary of the simulated insertion loss and 3-dB bandwidth of the drop port at 0V bias. (c-f) Simulated transmission spectra of the through and drop port at different bias voltages and varied self-coupling coefficient $r$ . . . . .	88
3.35	(a) Layout of O-band 16×16 SiPh Flex-LIONS (7.6 mm × 21.8 mm). (b) Layout of O-band MRR add-drop filter with EO and TO tuning. (c) Layout of O-band 2×2 MZS with EO and TO tuning. (d) Design of O-band MMI based waveguide crossing. (e) Design of O-band 2×2 MMI coupler. . . . .	89
3.36	Fabrication flow charts for the O-band 16×16 SiPh Flex-LIONS. . . . .	91
4.1	Schematic of the ANN architecture with input layer, hidden layers, output layers, and synaptic interconnections. Each synaptic interconnection is a linear operation represented by an arbitrary weight matrix $W$ . . . . .	93
4.2	(a) $N \times N$ unitary matrix represented by a ‘rectangular’ MZI mesh. (b) $N \times N$ unitary matrix represented by a ‘triangular’ MZI mesh. (c) 2×2 MZIs as the building blocks of the MZI mesh. ( $N$ is assumed to be an odd number here) . . . . .	94
4.3	Conventional PNN architecture with $N$ neurons at layer $t$ , $M$ neurons at layer $t+1$ and an $M \times N$ synaptic interconnections. . . . .	95
4.4	Weight matrix TT-decomposition for parameter compression. . . . .	96
4.5	Tensor-train layer. . . . .	97

4.6	Tensorized PNN architecture with cascaded TT cores implemented by 3D $R_{k-1}M_k \times N_kR_k$ MZI meshes and cross-connects. Alternatively, the 3D photonics can be replaced by putting MZI meshes side by side in a single plane. . . . .	98
4.7	Comparison of number of MZIs as a function of the radix $N$ with $N_1 = \dots = N_d = 2$ (a) and 4 (b). Comparison of insertion loss as a function of the radix $N$ with with $N_1 = \dots = N_d = 2$ (a) and 4 (b). . . . .	99
4.8	Schematic of the device architecture of the $1024 \times 1024$ tensorized PNN. .	100
4.9	(a) Layout of an $9 \times 9$ ‘rectangular’ MZI mesh. (b) Layout of a $8 \times 8$ ‘triangular’ MZI mesh. . . . .	100
5.1	(a) Cross-section of the designed III-V-on-Si hybrid SOA. (b) Simulated mode profile and $\Gamma_{\text{core}}$ with the Si waveguide width of $0.6 \mu\text{m}$ . (c) Simulated mode profile and $\Gamma_{\text{core}}$ with the Si waveguide width of $3 \mu\text{m}$ . . . . .	103
5.2	III-V-on-Si taper design for optical transitions. . . . .	104
5.3	(a) Simulated 3D mode profile of the $L_1$ taper with the length of $16 \mu\text{m}$ . (b) Simulated 3D mode profile of the $L_2 + L_3 + L_4$ tapers with the length of $5 \mu\text{m}$ . (c) Simulated 3D mode profile of the $L_5$ taper with the length of $20 \mu\text{m}$ . (d) Simulated transmission and reflection of the $L_2 + L_3 + L_4$ tapers with different lengths. (d) Simulated transmission and reflection of the $L_5$ taper with different lengths. . . . .	104
5.4	Schematic of the transfer printing process. . . . .	105
5.5	Fabrication flow charts for the transfer printing target wafer. . . . .	106
5.6	Layer stacks of the QD epitaxial wafer from Innolume GmbH. . . . .	107
5.7	Fabrication flow charts for the transfer printing source wafer. . . . .	108
5.8	(a) Coupon layer photomask showing the AA (tether) and BB areas. (b) Microscope image of the source wafer after coupon layer lithography. (c) SEM image of the tether. (d) FIB-SEM image of the AA area after sacrificial layer wet etching. (e) FIB-SEM image of the BB area after sacrificial layer wet etching. . . . .	109



5.9	Principle of our proposed IQ modulator shown by constellation map: (a) push-pull operation of two phase modulated OIL VCSELs as one quadrature. (b) Two quadratures are perpendicularly superposed to build IQ modulator. . . . .	112
5.10	(a) Microscope image of packaged VCSEL array. (b) Schematic of experimental setup. (c) Microscope image of experimental setup. . . . .	113
5.11	(a) Voltage (b) optical power (c) wavelength of free-running VCSEL with varied bias current. . . . .	114
5.12	(a) Detuning range of injection-locked VCSEL with varied injection ratio and indicated bias current. (b) Small-signal electrical to optical frequency response ( $S_{21}$ ) under 6-mA bias and the indicated injection ratio. . . . .	114
5.13	(a) Variation of AM and PM with RF frequency between 2 GHz and 20 GHz. Bias current 6 mA, injection ratio -8 dB, wavelength detuning 0.21 nm. (b) Variation of AM and PM with bias current from 4 mA to 18 mA. Injection ratio -2 dB, detuning wavelength 0.16 nm, RF frequency 10 GHz.	115
5.14	Variation of (a) AM and (b) PM with detuning wavelength at 6-mA bias current. The RF frequency is 10 GHz. . . . .	116
5.15	Large-signal modulation response of a single VCSEL for $2^{10}-1$ PRBS with 400 mV $V_{pp}$ drive: (a) AM and PM time plots (b) IQ histogram of the optical field. Bias current, injection ratio, wavelength detuning are 6 mA, -2 dB, and 0.15 nm, respectively. . . . .	117
5.16	Single-quadrature modulator output power versus phase shift between the two VCSELs. . . . .	117
5.17	Constellation of 20-GBd BPSK after decision directed equalizer and sampling with OSNR of: (a) 15.4 dB (b) 8.8 dB. . . . .	118
5.18	OSNR sensitivity of our synthesized 20-GBd BPSK. . . . .	119
5.19	Chirp-free modulation driven by Nyquist pulse shaped waveforms at 10, 30, and 40 GBd: (a) constellation after decision directed equalizer and sampling. (b) optical spectrum. . . . .	119

## LIST OF TABLES

2.1	Useful signal power of uniform-loss MRR-based crossbar architecture for different input and output ports. . . . .	18
2.2	Crosstalk power of uniform-loss MRR-based crossbar architecture for the worst-case path. . . . .	18
2.3	Worst-case insertion loss of the 32-port AWGR from center input (17) and side input (1). . . . .	37
2.4	Required connections between MTPs in group 1 (MTP 1 to 8) and group 2 (MTP 9 to 16) and the AWGR input and output fibers. . . . .	41
2.5	Wavelength values used in the Thin-CLOS routing experiment. . . . .	42
2.6	Typical power budget related values measured in the experiment. . . . .	44
3.1	Comparison among Flex-LIONS and the state-of-the-art bandwidth-reconfigurable switching fabrics. (* For $N \times N$ scale. ** By using dual MRR switches.) .	55

## ABSTRACT

### **Silicon Photonic Switching Fabrics and Synaptic Interconnections for High Performance Computing and Neuromorphic Computing Systems**

The explosive growth of the data traffic in today's high performance computing (HPC) and datacenter systems demands massive-scale and energy-efficient data interconnections. On the manufacturing side, as the lithography reaches atomic scale and fabrication costs continue to rise, Moore's law is barely maintaining its trend of continuing increases in transistor densities so that electrical integrated circuits (ICs) alone are reaching its bottleneck. Optical interconnects on silicon photonic (SiPh) platform is an attractive solution due to the capabilities of transmitting data at the speed of light without having a length-dependent impedance. On the architecture side, today's datacenters heavily rely on cascaded stages of many power-hungry electronic packet switches interconnected with a fixed networking topology. The fixed topology cannot dynamically adapt the bandwidth to the varied traffic patterns, which prevents better utilization of computing resources. Optical wavelength-and-space selective switching fabrics is a particularly suitable solution due to the ability of dynamic configuration and reconfiguration in both spectral and spatial domains. On the other hand, neuromorphic computing have gained growing popularity compared with traditional von Neumann computing due to its superior performance in various tasks such as computer vision, speech recognition, machine translations, medical diagnoses, and the game of Go. Photonic neural networks (PNNs) consisting of synaptic interconnections and neurons with nonlinear activation functions can improve both energy efficiency and throughput significantly compared with electronic artificial neural networks (ANNs). Recently, one crucial piece of research is focused on implementing high radix (e.g.,  $1024 \times 1024$ ) photonic synaptic interconnections.

This dissertation firstly presents the demonstration of arrayed waveguide grating router (AWGR)-based all-to-all optical interconnects using silicon photonic low-latency interconnect optical network switch (Si-LIONS). The impact of intra-band crosstalk on scalability is investigated and experimentally verified. An  $8 \times 8$  chip-scale Si-LIONS sys-

tem with integrated silicon nitride (SiN) AWGR and SiPh transceivers are taped out and fabricated by foundry multi-project-wafer (MPW) run. Wavelength routing functionality is demonstrated with error-free data transmission at 10 Gb/s using the on-chip modulators and SiN AWGRs. A  $64 \times 64$  wavelength routing Thin-CLOS system with significantly improved scalability is also experimentally demonstrated in a 1U rack enclosure. Second, the dissertation proposes a bandwidth-reconfigurable SiPh switching fabric called Flex-LIONS (Flexible Low-Latency Interconnect Optical Network Switch). Flex-LIONS architecture is enabled by combining an AWGR-based all-to-all interconnection, microring resonator (MRR) add-drop filters, and multi-wavelength spatial switches. Flex-LIONS exhibits  $21 \times$  fewer number of switching elements and  $2.9 \times$  lower on-chip optical losses for 64 ports than the state-of-the-art architectures. A multi-free-spectral-range (FSR) integrated  $8 \times 8$  SiPh Flex-LIONS module has been designed, fabricated, and packaged. Successful system testing demonstrates error-free all-to-all interconnects for both  $\text{FSR}_0$  and  $\text{FSR}_1$  with a 5.3-dB power penalty induced by AWGR intra-band crosstalk under the worst-case polarization scenario. After reconfiguration in  $\text{FSR}_1$ , the bandwidth between the selected pair of nodes is increased from 50 to 125 Gb/s while maintaining a 25 Gb/s/ $\lambda$  all-to-all interconnectivity in  $\text{FSR}_0$ . The design, layout, and fabrication of an O-band  $16 \times 16$  SiPh Flex-LIONS chip with ns switching speed are also presented. Third, the dissertation proposes a PNN architecture based on tensor-train decomposed synaptic interconnections. The device implementation design shows that high-radix ( $1024 \times 1024$ ) synaptic interconnections can be enabled by cascaded small-radix ( $16 \times 16$ ) photonic tensor-train cores. At  $1024 \times 1024$ , the proposed tensorized PNNs reduce the insertion loss by 171.8 dB and the number of Mach-Zehnder interferometers (MZIs) by  $582 \times$  compared with conventional PNNs. In the end, the design, layout, and fabrication process development of transfer-printed III-V-on-Si quantum dot (QD) lasers are presented. Also, chirp-free optical in-phase-quadrature (IQ) modulators based on injection-locked VCSEL phase array are experimentally demonstrated. 20-GBd BPSK signal is synthesized with a peak-to-peak drive voltage of only 400 mV. Nyquist pulse shaped drive signals at 10, 30, and 40 GBd indicates the modulator's chirp-free operation by showing the flat top of the optical spectrum.

## ACKNOWLEDGMENTS

Pursuing my PhD is a fruitful and unforgettable experience in my life. I want to express my gratitude to everyone who walked through this long journey with me.

First of all, I'd like to thank my advisor, Distinguished Professor S. J. Ben Yoo, for his helpful guidance and continuous support. But, more importantly, his rigorous and hardworking attitude towards research will keep inspiring me throughout my academic career.

Secondly, it's pleasant to work with all the fantastic colleagues in the Next Generation Networking and Computing Systems (NGNCS) Group. I want to thank Roberto Proietti for his guidance and discussion on the system testbed for Flex-LIONS and Thin-CLOS LIONS, thank Yu Zhang for his advice on the SiPh device design and foundry PDK layout, thank Gengchen Liu and Hongbo Lu for their help on the BER measurements, thank Pouya Fotouhi and Sebastian Werner for their brainstorming on the Flex-LIONS architecture, thank Yi-Chun Ling for the device packaging of SiPh Flex-LIONS, thank Mehmet Berkay On and Yun-Jhu Lee for the collaboration on the neuromorphic project, thank Rijuta Ravichandran for the cooperation on transfer printing process development. I'd also like to thank former NGNCS group members including Binbin Guan, Shaoqi Feng, Weicheng Lai, Mathias Prost, Tiehui Su, Chuan Qin, Kuanping Shang, Guangyao Liu, Yumeng Zhang, Zhao Dong, Siwei Li, and Burak Tekcan, for their encouragements.

Third, I am so grateful for all the professional development opportunities of being a research intern with Nokia Bell Labs, Lawrence Berkeley National Laboratory, and Hewlett-Packard Labs during my PhD. I want to thank Nicolas K. Fontaine, Haoshuo Chen, Roland R. Ryf, Georgios Michelogiannakis, Dilip Vasudevan, Di Liang, and Geza Kurczveil for their guidance and help during my internships.

Lastly, I'd like to thank my parents for supporting me in pursuing my academic goals abroad. Without your care, love, education, and company since childhood, I won't be what I am now. I'd also like to thank my fiancée, Weier Guo, for always being there, understanding, and encouraging me unconditionally. Your love takes me through the most challenging times during my PhD.

# Chapter 1

## Background and Motivation

### 1.1 End of the Moore's Law and Optical Interconnects

For the past 50 years, the Moore's Law has successfully enabled the information technology industry to double the performance and functionality of silicon microelectronics roughly every two years within a fixed cost, power, and area [1]. However, as the 2D lithography capability approaches the atomic realm and the fabrication cost continues to rise, the classical technological drivers that have supported Moore's Law for decades are failing by the mid-2020s [2, 3], as shown in Figure 1.1(a). On the other hand, the explosive growth of the data traffic in today's high performance computing (HPC) and datacenter systems demands the continuing realizations of faster, more energy-efficient, and cheaper computing performance. Then it is necessary to develop successor technologies to the current complementary metal oxide semiconductor (CMOS) logic. Three primary paths are shown in Figure 1.1(b), including new materials and devices, more efficient architectures and packaging, and new models of computation.

Optical interconnects are one of the potential successor technologies to mitigate the cost of data movement using electrical interconnects. Unlike electrical wires, which have a distance-dependent impedance (i.e., energy cost), optical waveguides can transmit data at the speed of light with the costs nearly independent of the distance, allowing them to overcome the wire-resistance limitation. As the rate of increase in the integration

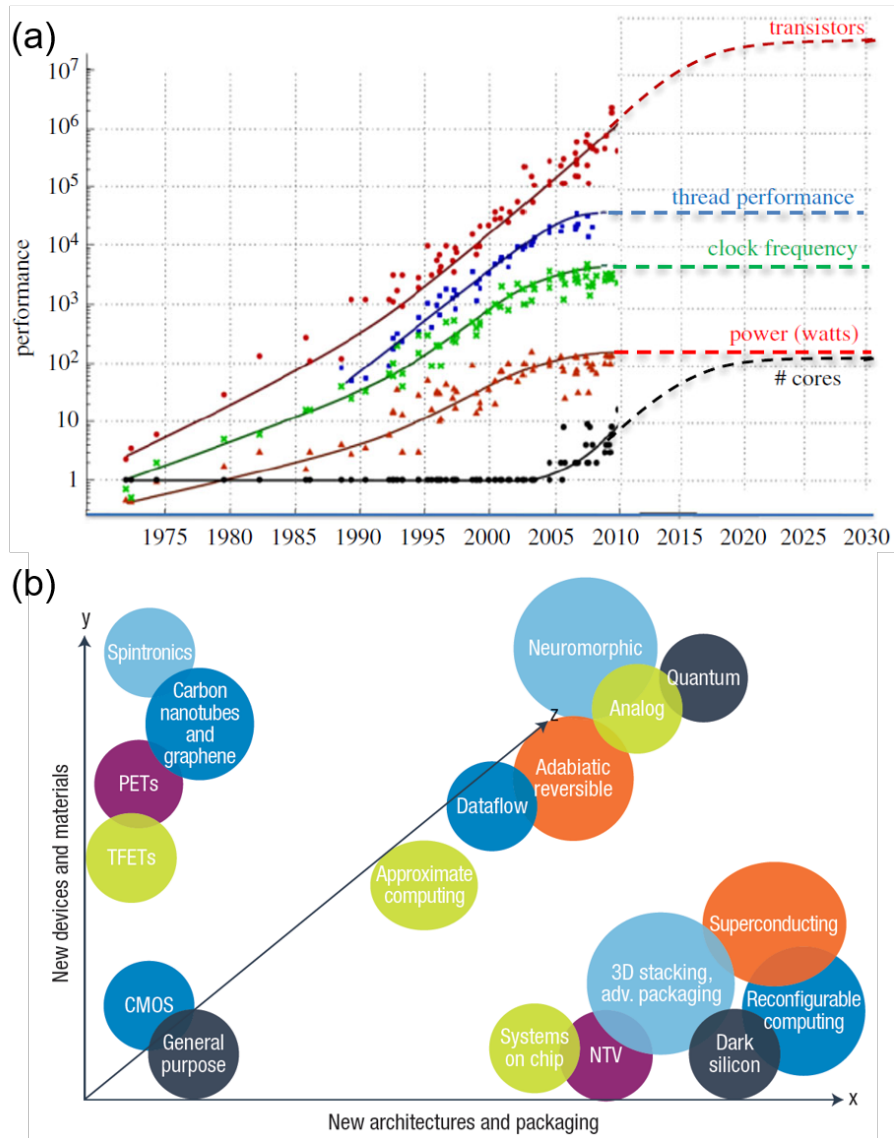


Figure 1.1. (a) All classical technological drivers to further performance improvements end in approximately 2025. (b) Three potential paths forward to realize continued performance improvements for Si microelectronics. (Courtesy of John Shalf, Lawrence Berkeley National Laboratory)

density of photonic integrated circuits (PICs) ('photonic Moore's Law' [4]) is now twice as fast as that of the electronic integrated circuits (EICs), optical interconnects on various PIC platforms are more and more applicable to replacing the electrical interconnects. For instance, silicon CMOS photonic electronic-photonic-integrated-circuits (EPICs) (e.g., GF9WG and GF45CLO from GlobalFoundries and SG25 EPIC from iHP) are now commercially offered, and co-integration of tens of thousands of PICs with EICs on a single

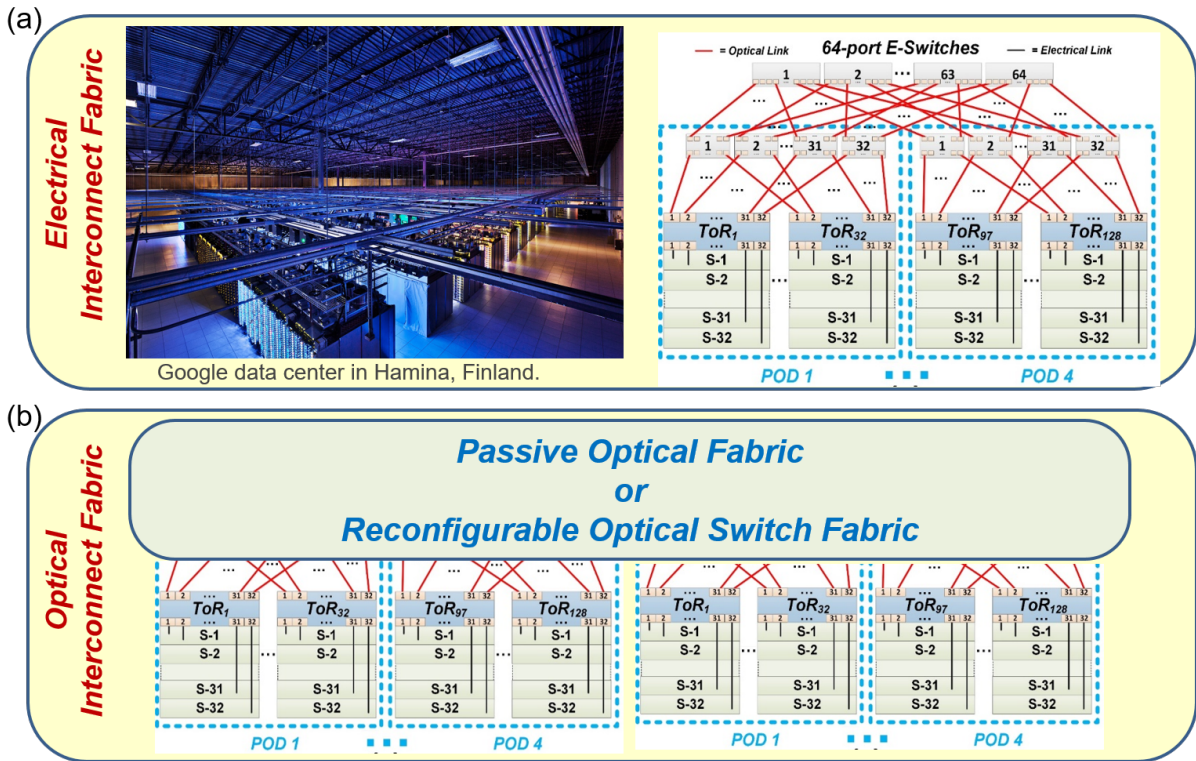


Figure 1.2. (a) A fat tree topology using electronic switches at the core and at the aggregation edges of the network. (b) A flattened optically interconnected network example utilizing a passive optical fabric or a reconfigurable optical switching fabric with electronic switches at the edges.

die manufactured on 300-mm wafers is possible.

## 1.2 Applications and Challenges of Optical Interconnects in High Performance Computing and Neuromorphic Computing Systems

Due to the rise of machine learning and artificial intelligence, today's computing systems are already emerging von Neumann and non-von Neumann architectures. Future computing systems are expected to work like human brains which means the left brain mainly deals with digital computations or the logic tasks while the right brain mainly deal with analog computations or the recognition tasks.

For the von Neumann computing, today's HPC and datacenter architectures heavily rely on cascaded stages of many power-hungry electronic packet switches interconnected



across the datacenter network in fixed hierarchical communication topologies such as fat-tree, as shown in Figure 1.2(a). Due to the limited radix and bandwidth of the electronic switches, warehouse-scale datacenter networks suffer from high energy consumption and latency due to repeated ‘store-and-forward’ electronic processes. Another characteristic of today’s HPC systems is that there are hot-spot and cold-spot links simultaneously created in the different locations of the network. This will make the communication patterns in such systems spatially and temporally non-uniform. However, the electrical interconnection fabrics have fixed topologies that cannot be dynamically reconfigured to match bursty communication patterns. Thus, employing a passive optical fabric or a reconfigurable optical switching fabric with distributed electronic switches (e.g., top-of-rack (ToR) switches) is desirable.

There has been a significant amount of architectural and experimental works on optical switching fabrics for HPC and datacenter systems, including semiconductor optical amplifier (SOA)-gate based optical switches [5], silicon photonic (SiPh) Mach-Zehnder interferometers (MZI)-based optical switches [6, 7], SiPh microring resonator (MRR)-based optical switches [8, 9], SiPh microelectromechanical systems (MEMS) switches [10–12], and SiPh arrayed waveguide grating router (AWGR)-based switches [13]. However, most of these switching fabrics only have single wavelength connectivity between I/O ports, and the bandwidth cannot be steered.

In the past few years, several integrated bandwidth-reconfigurable switching fabrics have been demonstrated by using wavelength-and-space selective optical switching [14–16]. However, all these reported switching fabrics exhibit poor scalability and low energy efficiency due to either high insertion losses induced by power splitters [14],  $O(N^2)$  waveguide crossings in the worst-case path [15], or large number ( $O(N^3)$ ) of required switching elements [15, 16]. Thus the challenge is to design a more scalable bandwidth-reconfigurable optical switching fabric.

Neuromorphic computing processors such as IBM TrueNorth [17] and Intel Loihi [18] have shown significantly superior performance compared with traditional central processing units (CPUs) for specific neural network tasks. A majority of the power consumption

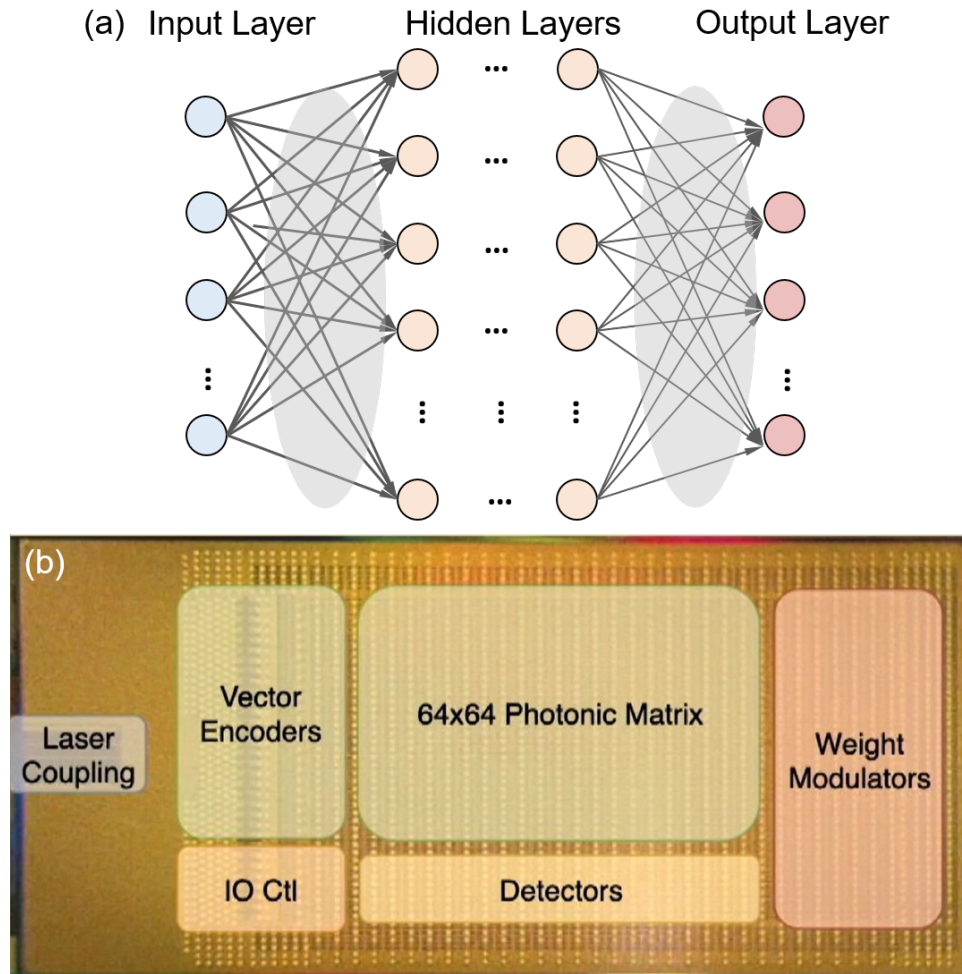


Figure 1.3. (a) PNN architecture consisting of an input neuron layer, many hidden neuron layers, an output neuron layer, and synaptic interconnections. (b) Mars chip from Lightmatter with a photonic core can implement  $64 \times 64$  photonic synaptic interconnections (Courtesy of Lightmatter).

of the electrical neuromorphic computing processors comes from data movement in the synaptic interconnections. Photonic neural networks (PNN), as shown in Figure 1.3(a), consisting of optical neurons and photonic synaptic interconnections, can significantly improve both energy efficiency and throughput compared with electrical neuromorphic computing processors. The typical architecture of a PNN contains one input layer of neurons, many hidden layers of neurons, and one output layer of neurons. There is a weight matrix between each adjacent layer of neurons that can be abstracted as an arbitrary unitary matrix. MZI meshes typically implement this matrix.

One major challenge of PNNs is that the MZI mesh-based photonic synaptic inter-

connections are challenging to scale up to high radix (e.g.,  $1024 \times 1024$ ) like the electronic neuromorphic computing hardware. The main reason is that the number of MZIs (as well as the chip size and number of electrical I/Os) in an MZI mesh scale as  $O(N^2)$ . For instance, Lightmatter company presented their Mars chip at the 2020 Hot Chips conference as shown in Figure 1.3(b) [19]. There is a photonic core in their chip that can implement  $64 \times 64$  weight matrix. However, the chip size is already  $150 \text{ mm}^2$ . Thus it is desirable to propose novel PNN architectures that can mitigate the scalability issues of the MZI mesh-based photonic synaptic interconnections.

### 1.3 Silicon Photonic Technology and Heterogeneous Integration

Silicon photonics (SiPh) exploiting silicon as the waveguide core and thermally-grown oxide as the cladding realized extremely low-loss and low-defect optical interfaces while offering strong confinement of optical modes enabling relatively compact low-loss bending of waveguides. Furthermore, the SiPh platform shares the same fabrication process steps in most parts as the silicon CMOS platform, leading to the commercial large-volume and cheap silicon CMOS EPIC Manufacturing. Compared with other material platforms, the silicon platform offers advantages such as the commercial availability of 300 mm wafers, superior thermo-mechanical characteristics, abundant material availability, and the well-established end-to-end ecosystem from design, multi-project-wafer (MPW) run, to testing and packaging. When considering large-scale integration, SiPh offer superior integration density and yield thanks again to the high-quality, high-density, and low-defect passivation and the low-loss (0.03 dB/cm) high-contrast silicon waveguides.

However, silicon is an indirect bandgap material, so that the silicon light emitter is inefficient. Secondly, silicon is centrosymmetric, so that the SiPh modulator is inefficient. Lastly, silicon is reciprocal, so SiPh isolator and circulator are impossible. Thus it's necessary to integrate other materials on silicon heterogeneously, including III-V, Ce:YIG, and periodically poled lithium niobate (PPLN), and so on, on silicon. The main methods for heterogeneous integration include wafer-bonding, epitaxially growth, and transfer

printing.

## 1.4 Scope of This Dissertation

This dissertation addresses the challenges mentioned in Section 1.2 as follows:

Chapter 2 presents the demonstration of AWGR-based all-to-all optical interconnects using Si-LIONS. An  $8\times 8$  chip-scale Si-LIONS system with integrated SiN AWGR and SiPh transceivers are taped out and fabricated by foundry MPW runs. Wavelength routing functionalities are demonstrated with error-free data transmission at 10 Gb/s using the on-chip modulators and SiN AWGRs. A  $64\times 64$  wavelength routing Thin-CLOS system with significantly improved scalability is also experimentally demonstrated in a 1U rack enclosure.

Chapter 3 proposes the Flex-LIONS architecture, which is enabled by combining an AWGR-based all-to-all interconnection, MRR add-drop filters, and multi-wavelength spatial switches. A multi-FSR integrated  $8\times 8$  SiPh Flex-LIONS module has been designed, fabricated, and packaged. Successful system testing demonstrates error-free all-to-all interconnects for both  $\text{FSR}_0$  and  $\text{FSR}_1$  with a 5.3-dB power penalty induced by AWGR intra-band crosstalk under the worst-case polarization scenario. After reconfiguration in  $\text{FSR}_1$ , the bandwidth between the selected pair of nodes is increased from 50 to 125 Gb/s while maintaining a 25 Gb/s/ $\lambda$  all-to-all interconnectivity in  $\text{FSR}_0$ . The design, layout, and fabrication of an O-band  $16\times 16$  SiPh Flex-LIONS chip with ns switching speed are also presented.

Chapter 4 proposes a PNN architecture based on tensor-train decomposed synaptic interconnections. The device implementation design shows that high-radix ( $1024\times 1024$ ) synaptic interconnections can be enabled by cascaded small-radix ( $16\times 16$ ) photonic tensor-train cores.

Chapter 5 presents the design, layout, and fabrication process development of transfer-printed III-V-on-Si quantum dot (QD) lasers. Also, chirp-free optical in-phase-quadrature (IQ) modulators based on injection-locked VCSEL phase array are experimentally demonstrated.

# Chapter 2

## Si-LIONS For Scalable All-to-All Optical Interconnects

### 2.1 Principle and Architecture

#### 2.1.1 Arrayed Waveguide Grating Router (AWGR)

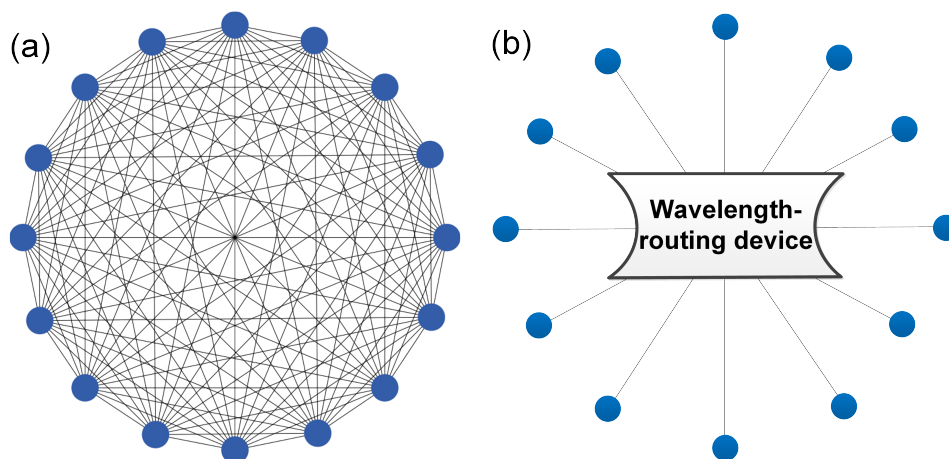


Figure 2.1. All-to-all optical interconnects based on: (a) waveguides (b) wavelength-routing device.

Optical network on chip (NOC) and chip-to-chip interconnects are emerging in modern computing systems with chip-multicore-processors (CMP) and a system-in-package (SIP) due to the high demand in latency, throughput, and power consumption [20–22]. All-to-all optical interconnects is defined as each node in an interconnection network can communicate with all the other nodes simultaneously. All-to-all interconnections are

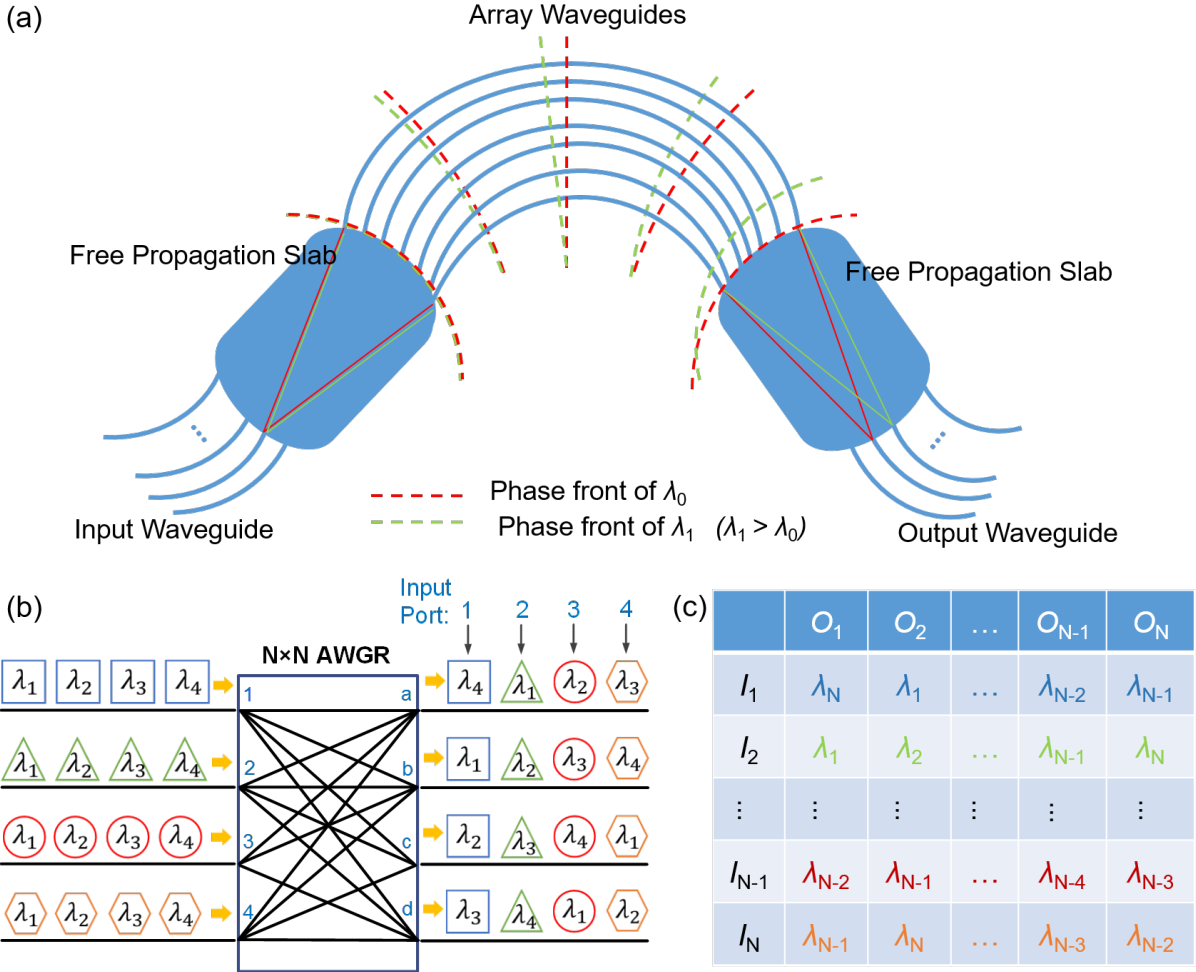


Figure 2.2. (a) Schematic of AWGR. (b) Wavelength routing function of AWGR. (c) Wavelength routing table of AWGR.

essential for many applications, including map-reduce based applications [23], parallel sorting applications [24], and deep neural network (DNN) applications [25, 26]. Such interconnection topology is contentionless so that a simpler control plane can be employed. However, traditional waveguide-based all-to-all optical interconnects are becoming more difficult to deploy in high-radix topology since an  $N$ -node network requires  $N(N-1)$  optical wires and countless crossings (Figure 2.1 (a)).

One solution is to utilize wavelength-routing devices (Figure 2.1 (b)) such as AWGR [13, 27–29] and MRR based crossbar [16, 30]. The advantage of using AWGR is firstly reflected by the fact that it does not require tuning, while an  $N \times N$  MRR-based crossbar requires  $N^2$  tuning elements. Another important factor needs to be considered is the

scalability of the two architectures which is mainly affected by the intra-band signal-crosstalk beat noise [27]. Based on the state-of-the-art crosstalk values, the AWGR is proven to be more scalable than the MRR-based crossbar architecture [31]. The detailed calculation and discussion will be included in Section 2.2.

The AWGR consists of input and output waveguides, two free-propagation slab regions, and arrayed waveguides where each neighboring waveguide has a constant path length difference as shown in Figure 2.2(a). The input light beam diffracts at the left slab region, travels through the arrayed waveguides, and interferes at a particular focal point in the second slab region. The locations of the focal points depend on the signal wavelength and the corresponding relative phase delay in each arrayed waveguide. Therefore, AWGR can provide all-to-all interconnects in a flat topology without contention, connectivity between input and output nodes can be achieved by injecting different wavelength (Figure 2.2(b) and (c)).

### **2.1.2 Silicon Photonic Low-Latency Interconnect Optical Network Switch (Si-LIONS)**

Integrated photonics, particularly silicon photonics, offers an attractive platform for such AWGR-based all-to-all interconnect systems with advantages of: (1) significant reductions in size, weight, and power (SWaP) compared to the standalone devices with fiber connections; (2) and facilitating high-radix all-to-all interconnections. The Si-LIONS is an on-chip interconnect architecture with the wavelength routing functionality of AWGR (Figure 2.3) [13]. In order to support all-to-all operation, each node has a multi-wavelength transmitter and a multi-wavelength receiver. An off-chip comb laser source provides the  $N$  wavelength division multiplexing (WDM) wavelengths which match with the AWGR channels. The cyclic frequency feature of AWGR guarantees the same set of wavelengths can be used at each node. A centralized control plane is assumed here for simplicity. Alternatively, a distributed control plane with the all-optical TOKEN technique can be used. Si-LIONS also allows close integration with silicon photonic transceivers, CMOS ICs, and nanoelectronics.

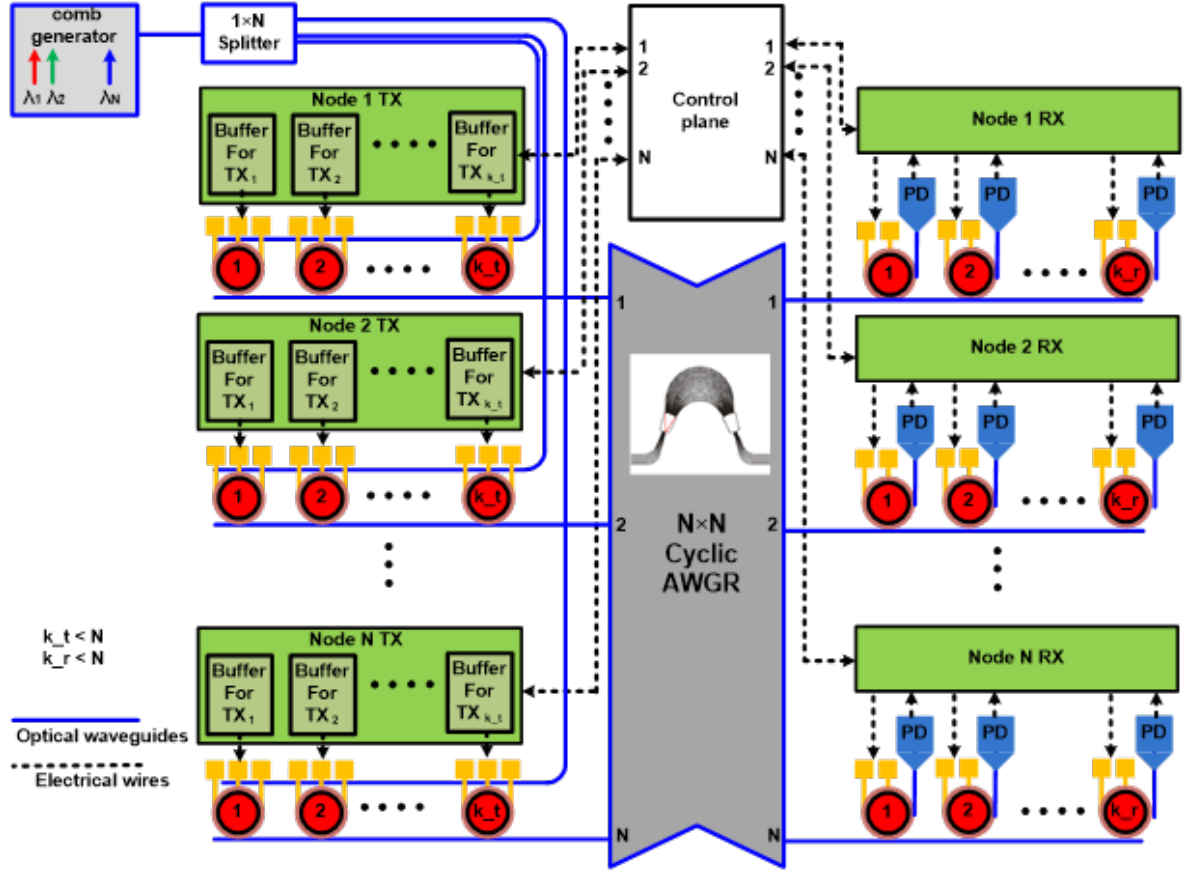


Figure 2.3. Si-LIONS architecture.

## 2.2 Impact of Intra-Band Crosstalk on Scalability of All-to-All Optical Interconnects

### 2.2.1 Crosstalk penalty for OOK modulation format

The scalability of crossbar is mainly affected by intra-band signal-crosstalk beat noise since it cannot be removed by demultiplexer at the receiver end. The bit error rate (BER) of the on-off keying (OOK) modulation format can be expressed by:

$$BER = P(1)P(1|0) + P(0)P(0|1) \quad (2.1)$$

where  $P(1|0)$  is the probability of misinterpretation of '1' as '0', and  $P(0|1)$  is the probability of misinterpretation of '0' as '1'. For OOK format, '1' and '0' are equally probable



so that:  $P(1)=P(0)=1/2$ . Assuming additive white Gaussian noise (AWGN) model:

$$P(1|0) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{I_{th}} e^{-\frac{(y-I_1)^2}{2\sigma_1^2}} dy = \frac{1}{2} \operatorname{erfc}\left(\frac{I_1 - I_{th}}{\sqrt{2}\sigma_1}\right) \quad (2.2)$$

$$P(0|1) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{I_{th}}^{\infty} e^{-\frac{(y-I_0)^2}{2\sigma_0^2}} dy = \frac{1}{2} \operatorname{erfc}\left(\frac{I_{th} - I_0}{\sqrt{2}\sigma_0}\right) \quad (2.3)$$

where  $I_{th}$  is the decision threshold which can be manually or automatically set by the receiver,  $\sigma_1^2$  and  $\sigma_0^2$  are the variance (noise) of the distribution at '1' and '0'. By substituting Equation 2.2 and 2.3 into 2.1:

$$BER = \frac{1}{4} \operatorname{erfc}\left(\frac{I_1 - I_{th}}{\sqrt{2}\sigma_1}\right) + \frac{1}{4} \operatorname{erfc}\left(\frac{I_{th} - I_0}{\sqrt{2}\sigma_0}\right) \quad (2.4)$$

For the case of fixed decision threshold, assuming the receiver noise is mainly thermal noise for simplicity ( $\sigma_0 \approx \sigma_1$ ), then:

$$I_{th} = \frac{\sigma_0 I_1 + \sigma_1 I_0}{\sigma_0 + \sigma_1} = \frac{I_0 + I_1}{2} \quad (2.5)$$

By substituting Equation 2.5 into 2.4:

$$BER = \frac{1}{2} \operatorname{erfc}\left(\frac{I_1 - I_0}{2\sqrt{2}\sigma_0}\right) \quad (2.6)$$

Assuming high extinction ratio:  $I_0 = 0, I_{th} = I_1/2$ . Then:

$$BER = \frac{1}{2} \operatorname{erfc}\left(\frac{I_1}{2\sqrt{2}\sigma_0}\right) \quad (2.7)$$

$$Q = \frac{I_1}{2\sigma_0} \quad (2.8)$$

where Q is the Q-factor without crosstalk being considered.

The signal-crosstalk beat noise only occurs when the signal is at '1', so that:

$$\sigma'_0 = \sqrt{\sigma_0^2 + \sigma_{RIN}^2 I_1'^2} \quad (2.9)$$

where  $\sigma_{RIN}$  is the relative intensity noise (RIN). By substituting Equation 2.8 into 2.4, we have the updated BER as:

$$BER' = \frac{1}{4} \operatorname{erfc}\left(\frac{I_1'}{2\sqrt{2}\sqrt{\sigma_0^2 + \sigma_{RIN}^2 I_1'^2}}\right) + \frac{1}{4} \operatorname{erfc}\left(\frac{I_1'}{2\sqrt{2}\sigma_0}\right) \quad (2.10)$$

$$Q' = \frac{I'_1}{2\sqrt{\sigma_0^2 + \sigma_{\text{RIN}}^2 I_1'^2}} \quad (2.11)$$

Since  $\frac{1}{4}\text{erfc}(\frac{Q'}{\sqrt{2}}) < \text{BER}' < \frac{1}{2}\text{erfc}(\frac{Q'}{\sqrt{2}})$ , it can be easily proved that  $Q' \approx Q$  especially when BER value is low. Based on Equation 2.8 and 2.11, the crosstalk penalty for fixed decision threshold can be calculated by:

$$\text{Penalty} = 10 \log \frac{I'_1}{I_1} = 10 \log \frac{Q'}{Q} - 5 \log(1 - 4\sigma_{\text{RIN}}^2 Q'^2) \approx -5 \log(1 - 4\sigma_{\text{RIN}}^2 Q'^2) \quad (2.12)$$

For optimized decision threshold which means the  $I_{\text{th}}$  can be automatically set:

$$I'_{\text{th}} = \frac{\sigma_0 I'_1}{\sigma_0 + \sqrt{\sigma_0^2 + \sigma_{\text{RIN}}^2 I_1'^2}} \quad (2.13)$$

By substituting Equation 2.13 into 2.4:

$$\text{BER}' = \frac{1}{2} \text{erfc}\left(\frac{I'_1}{\sqrt{2}(\sigma_0 + \sqrt{\sigma_0^2 + \sigma_{\text{RIN}}^2 I_1'^2})}\right) \quad (2.14)$$

$$Q' = \frac{I'_1}{\sigma_0 + \sqrt{\sigma_0^2 + \sigma_{\text{RIN}}^2 I_1'^2}} \quad (2.15)$$

Comparing Equation 2.7 and 2.14,  $Q' = Q$ . Based on Equation 2.8 and 2.15, the crosstalk penalty for optimized decision threshold can be calculated by:

$$\text{Penalty} = 10 \log \frac{I'_1}{I_1} = 10 \log \frac{Q'}{Q} - 10 \log(1 - \sigma_{\text{RIN}}^2 Q'^2) = -10 \log(1 - \sigma_{\text{RIN}}^2 Q'^2) \quad (2.16)$$

In  $N$ -node all-to-all wavelength-routing optical interconnects, for each pair of input and output nodes, there is an assigned wavelength for the signal and  $N-1$  crosstalk components at the same wavelength contributed by other  $N-1$  nodes. To take the worst case into account, we assume that the polarization states of all the crosstalk sources are the same as that of the signal, and the whole noise power is inside the receiver bandwidth. Then the RIN can be expressed by  $\sigma_{\text{RIN}}^2 = \sum_i R_{xi}$ , where  $R_{xi}$  is the optical power ratio of the  $i$ -th crosstalk component to the signal. Assuming the input power of each port (node) are equalized as  $P_{\text{in}}$ , then  $R_{xi} = P_{xi}/P_{\text{out}}$  where  $P_{xi}$  is the crosstalk power from input port  $i$  and  $P_{\text{out}}$  is the useful signal power.

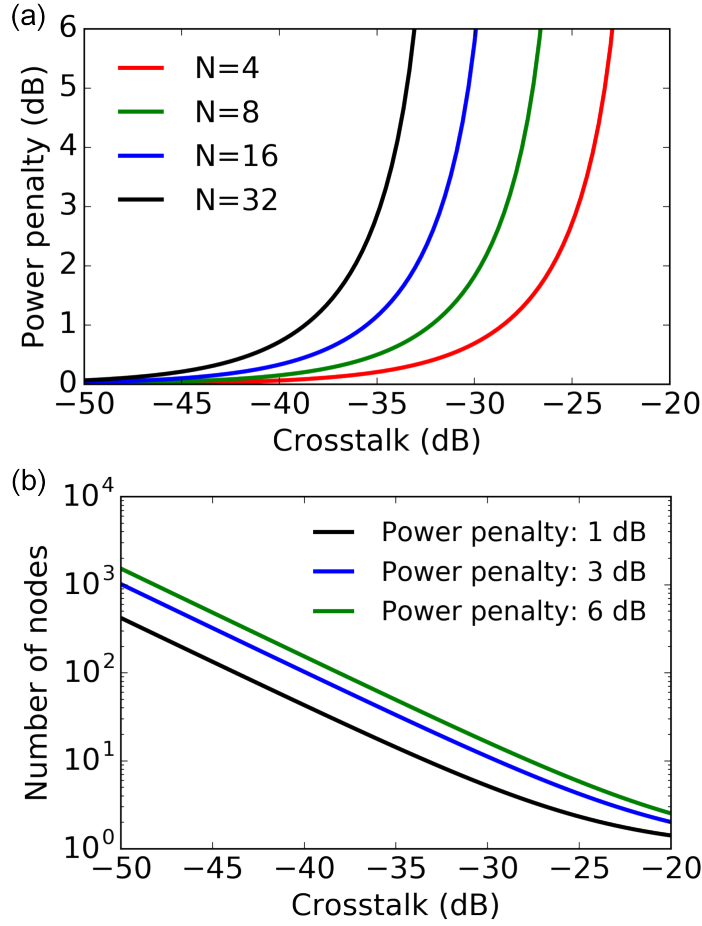


Figure 2.4. (a) Power penalty versus varied crosstalk for AWGR. (b) Number of scalable nodes versus varied crosstalk with power penalty constraint of 1, 3, 6 dB.

## 2.2.2 Worst-Case Crosstalk Penalty for AWGR

For AWGR, the variations in the waveguide geometry and the thickness of the material stack make the average effective indices for each waveguide arms different. This results in phase errors on the waveguide arms. Since AWGR is an interferometric device, the phase errors cause a degraded crosstalk. Normally, the crosstalk value varies among adjacent channels and non-adjacent channels. For convenience, we use  $X_{\text{AWGR}}$  to notate the average per-port crosstalk contribution so that the RIN of AWGR is:

$$(\sigma_{\text{RIN}}^2)_{\text{AWGR}} = (N - 1) \cdot X_{\text{AWGR}} \quad (2.17)$$

By substituting Equation 2.17 into 2.16, the power penalty due to the intra-band crosstalk of AWGR can be calculated with optimized decision-threshold setting. For all

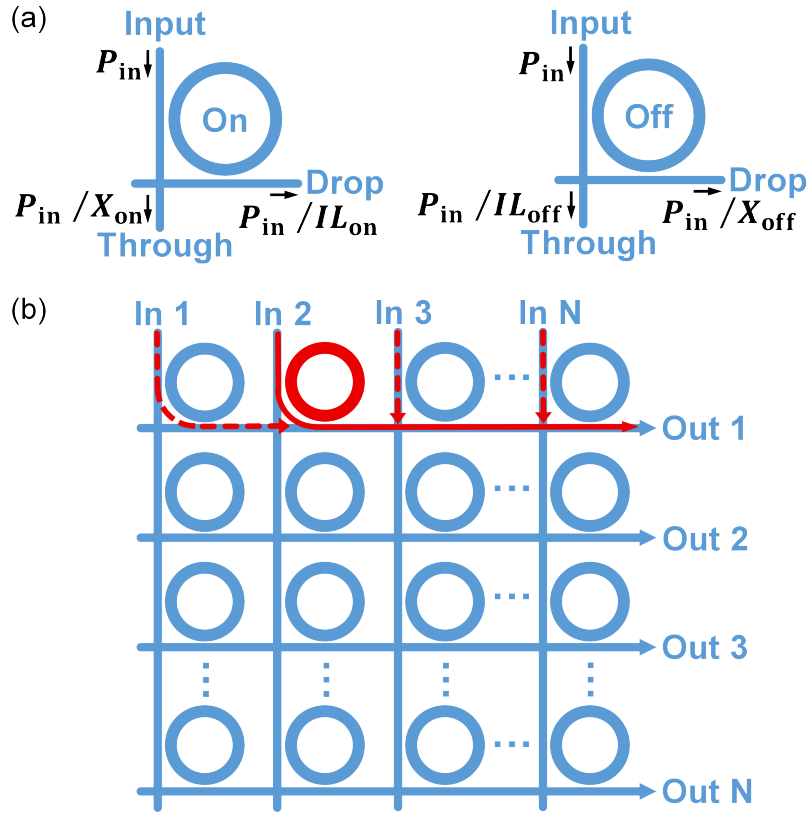


Figure 2.5. Schematic of: (a) MRR (b) Conventional MRR-based crossbar architecture.

the calculations in this section,  $Q$  is set as 7 to obtain a BER of  $10^{-12}$ .

Figure 2.4(a) shows the power penalty of AWGR with varied crosstalk for  $N = 4, 8, 16, 32$ . The results show that, with -35 dB crosstalk, the power penalty for 4-, 8-, 16-, and 32-node interconnects are 0.21, 0.50, 1.15, and 2.84 dB, respectively. For more clarity, the number of scalable nodes versus varied crosstalk for the power penalty constraints of 1, 3, and 6 dB are plotted in Figure 2.4(b). With 1-, 3-, and 6-dB power penalty constraint, the AWGR can scale to 32 nodes when the crosstalk is less than -38.7, -34.9, -33.1 dB, respectively.

### 2.2.3 Worst-Case Crosstalk Penalty for MRR Crossbar

Figure 2.5(a) shows the schematic of an add-drop MRR at the designed resonance wavelength. The input signal can resonantly drop (shown as ‘On’) or to go through (shown as ‘Off’). For on-resonance (off-resonance), the incident light outputs from the drop port

(the through port) with an insertion loss of  $IL_{\text{on}}$  ( $IL_{\text{off}}$ ) and leaks to the through port (the drop port) with a crosstalk of  $X_{\text{on}}$  ( $X_{\text{off}}$ ). In practice,  $X_{\text{on}}$  is typically smaller than  $X_{\text{off}}$  due to the asymmetric structure.

Conventional MRR-based crossbar architecture can be formed by a matrix of  $N^2$  MRRs as shown in Figure 2.5(b). The routing mechanism is similar to AWGR where the WDM signals from each input node can be routed to different output nodes by wavelength-selective dropping at assigned on-state MRRs. In order to provide arbitration-free full connectivity, each column and each row of MRR matrix should all have different resonance wavelengths which correspond to the WDM channels.

Since the conventional crossbar architecture is asymmetric, the path with maximum RIN should be used for power penalty calculation. Ref. [30] states that the worst-case path is the one connecting *In 1* to *Out N* where the transmission loss for the signal is minimum. However, the output signal power is in the range of  $P_{\text{in}}/(IL_{\text{on}} \cdot IL_{\text{off}}^{2(N-1)})$  to  $P_{\text{in}}/IL_{\text{on}}$  and  $IL_{\text{off}}$  is normally less than 0.1 dB, thus the total crosstalk power is more dominant in RIN. However, the actual worst-case path should be the one connecting *In 2* to *Out 1* (red solid line in Figure 2.5(b)), which yields the useful signal power  $P_{\text{out}} = P_{\text{in}}/(IL_{\text{on}} \cdot IL_{\text{off}}^{N-2})$ . For *In 3* to *In N*, the crosstalk signal goes through only one crosstalk of MRR before reaching the *Out 1* (red dashed lines in Figure 2.4(b)), which gives a crosstalk power  $P_{xi} = P_{\text{in}} \cdot X_{\text{off}}/IL_{\text{off}}^{N-i}$ , where  $i = 3, 4, \dots, N$ . Then  $R_{xi} = P_{xi}/P_{\text{out}} = X_{\text{off}} \cdot IL_{\text{on}} \cdot IL_{\text{off}}^{i-2}$ . Higher-order crosstalk components with two or more crosstalk of MRR are ignored, since the crosstalk is normally smaller than -20 dB and the number of first-order and higher-order crosstalk components are in the same order of magnitude. The path connecting *In 1* to *Out 1* is not considered since each node does not need to connect to itself. Therefore, the RIN for the worst-case path is:

$$(\sigma_{\text{RIN}}^2)_{\text{C-Crossbar}} = \sum_{i=1}^N R_{xi} = \frac{X_{\text{off}} \cdot IL_{\text{on}} \cdot IL_{\text{off}} \cdot (IL_{\text{off}}^{N-2} - 1)}{IL_{\text{off}} - 1} \quad (2.18)$$

Figure 2.6(a) shows the schematic of uniform-loss MRR-based crossbar architecture [32] where columns with  $N/2$  and  $N/2-1$  MRRs are located alternately. Comparing to conventional crossbar, the required number of MRRs is reduced from  $N^2$  to  $N(N-1)/2$  as shown in Figure 2.6(b). In order to enable all-to-all interconnects, each column should

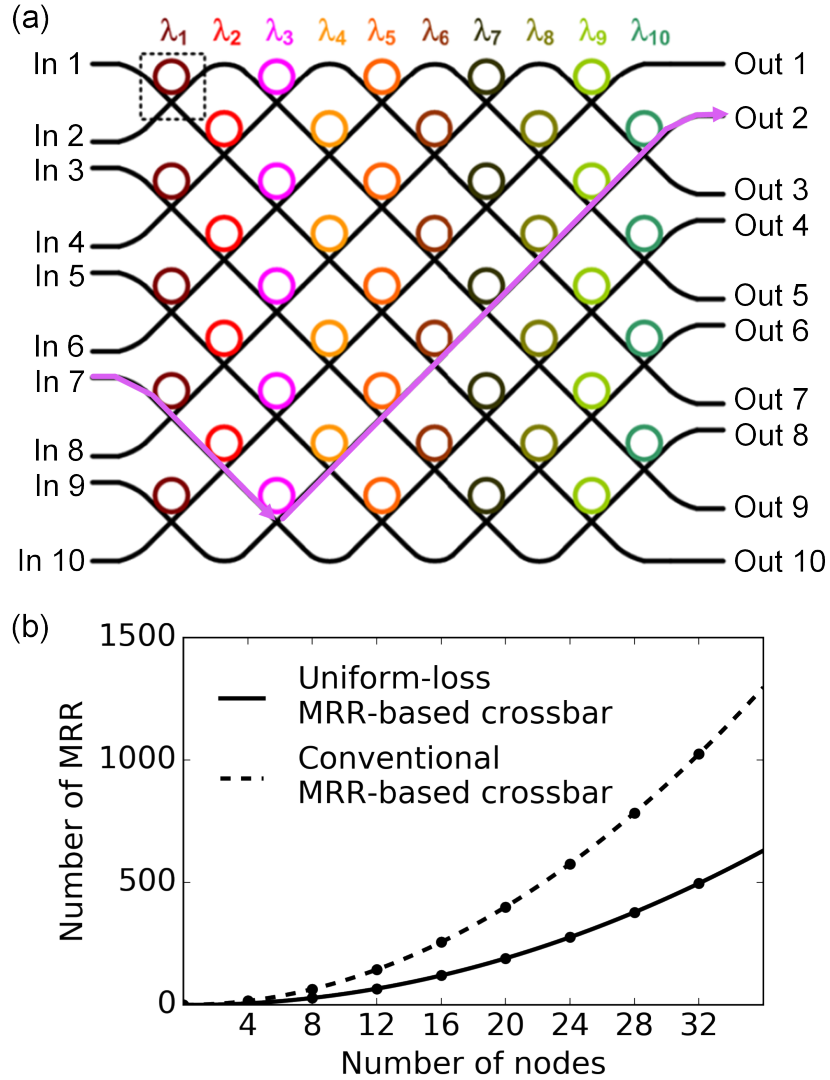


Figure 2.6. (a) Schematic of uniform-loss MRR-based crossbar architecture. (b) Comparison of the number of MRRs between conventional crossbar (dashed line) and uniform-loss crossbar (solid line).

have different resonance wavelengths.

Although the uniform-loss MRR-based crossbar architecture is also not symmetric, the insertion loss is more uniform since the signals go through nearly same number of MRRs before output. For instance, the signal power from  $In\ i$  to  $Out\ N-i+1$  is  $P_{in}/IL_{off}^{N-1}$ . Other than this, the signal power from different even and odd input and output port combinations are listed in Table 2.1 with a maximum variation of only  $IL_{off}^2$ .

The worst-case path is the one connecting  $In\ N-3$  to  $Out\ 2$  using  $\lambda_3$  since it has the

Input port	Output port	Signal power $P_{\text{out}}$
Odd	Odd	$P_{\text{in}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-2})$
Odd	Even	$P_{\text{in}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-1})$
Even	Odd	$P_{\text{in}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-3})$
Even	Even	$P_{\text{in}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-2})$

Table 2.1. Useful signal power of uniform-loss MRR-based crossbar architecture for different input and output ports.

Input port $i$	Crosstalk power $P_{xi}$	$R_{xi} = P_{xi}/P_{\text{out}}$
1, 3, 5, ..., $N-7$	$P_{\text{in}} \cdot X_{\text{off}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-3})$	$X_{\text{off}} \cdot (IL_{\text{off}})^2$
2	$P_{\text{in}} \cdot X_{\text{off}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-4})$	$X_{\text{off}} \cdot (IL_{\text{off}})^3$
4, 6, 8, ..., $N-4$	$P_{\text{in}} \cdot X_{\text{off}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-2})$	$X_{\text{off}} \cdot IL_{\text{off}}$
$N-5$	0	0
$N-2, N$	$2P_{\text{in}} \cdot X_{\text{off}}/(IL_{\text{on}} \cdot (IL_{\text{off}})^{N-2})$	$2 \cdot X_{\text{off}} \cdot IL_{\text{off}}$
$N-1$	$P_{\text{in}} \cdot X_{\text{on}}/(IL_{\text{off}})^{N-2}$	$X_{\text{on}} \cdot IL_{\text{on}} \cdot IL_{\text{off}}$

Table 2.2. Crosstalk power of uniform-loss MRR-based crossbar architecture for the worst-case path.

maximum number of first-order crosstalk components (pink solid line in Figure 2.6(a)). The corresponding signal power is  $P_{\text{out}} = P_{\text{in}}/(IL_{\text{on}} \cdot IL_{\text{off}}^{N-1})$ . Ignoring all the higher-order crosstalk components, the crosstalk power of different input ports and the corresponding ratio to signal power are listed in Table 2.2. Thus the RIN for the worst-case path is:

$$(\sigma_{\text{RIN}}^2)_{\text{U-Crossbar}} = \sum R_{xi} = \frac{N-6}{2} \cdot X_{\text{off}} \cdot IL_{\text{off}}^2 + \frac{N+2}{2} \cdot X_{\text{off}} \cdot IL_{\text{off}} + X_{\text{off}} \cdot IL_{\text{off}}^3 + X_{\text{on}} \cdot IL_{\text{on}} \cdot IL_{\text{off}} \quad (2.19)$$

Compared with conventional MRR-based crossbar, uniform-loss MRR-based crossbar exhibits a scalability improvement especially for high-radix interconnects. Figure 2.7(a) plots the power penalty versus varied crosstalk for conventional (solid lines) and uniform-loss crossbar (dashed lines). Here, the crosstalk refers to  $X_{\text{off}}$ ,  $X_{\text{on}}$  is assumed to be 5 dB smaller than  $X_{\text{off}}$ , as measured in [33],  $IL_{\text{off}}$  is assumed to be 0.1 dB. It can be seen that uniform-loss crossbar has lower power penalty than conventional crossbar for 8-, 16-, and

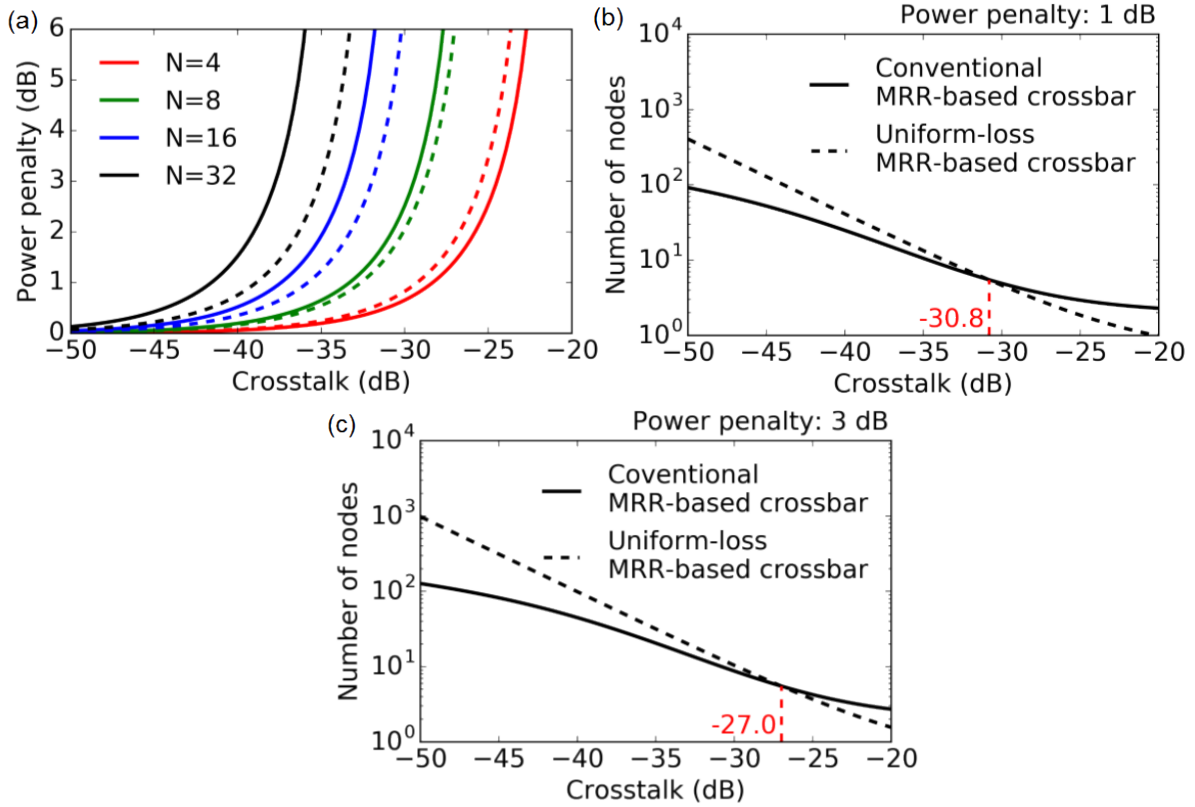


Figure 2.7. (a) Power penalty versus varied crosstalk for conventional crossbar (solid lines) and uniform-loss crossbar (dashed lines). (b) Number of scalable nodes versus varied crosstalk with power penalty constraint of 1 dB. (c) Number of scalable nodes versus varied crosstalk with power penalty constraint of 3 dB.

32-node interconnects. This can be explained by Equation 2.18 and 2.19 in which the RIN of uniform-loss crossbar is linearly associated with  $N$  while the RIN of conventional crossbar is exponentially associated with  $N$ . That is why the RIN, i.e., the power penalty of uniform-loss crossbar is smaller than conventional crossbar and such improvement is even more significant with higher  $N$ .

For more clarity, the number of scalable nodes versus varied crosstalk are plotted with power penalty constraint of 1 dB (Figure 2.7(b)) and 3 dB (Figure 2.7(c)). With crosstalk of -35 dB and power penalty lower than 1 dB, conventional crossbar and uniform-loss crossbar can support 10- and 13-node interconnects. If the power penalty constraint is loosened to 3 dB, the corresponding number of supportable nodes can be expanded to 20 and 31. The intersections of the two curves in Figure 2.7(b) and (c) are also marked



to indicate that with crosstalk lower than -30.8 dB and -27.0 dB, uniform-loss crossbar can support more nodes than conventional crossbar for power penalty constraint of 1 dB and 3 dB.

## 2.2.4 Comparison Based on the State-of-the-Art Parameters

For comparison, the BER floor of all-to-all optical interconnects with different number of nodes are calculated and plotted in Figure 2.8 based on the state-of-the-art technologies.

Thanks to the silicon CMOS ecosystem, silicon photonics is a relatively mature photonics platform for implementing wavelength-routing devices since it allows high-density and low-cost PICs fabrication leveraging the CMOS process flow. For silicon MRR,  $X_{\text{on}} = -23.1$  dB,  $X_{\text{off}} = -18.1$  dB, as measured in [33]. As shown by the black line in Figure 2.8, the BER floor of the state-of-the-art silicon conventional MRR-based crossbar is unacceptable ( $10^{-6}$ ) even for  $4 \times 4$  interconnects. Moreover, the BER floor of silicon uniform-loss MRR-based crossbar (red curve) is slightly lower but still far from acceptable values. To make MRR-based crossbar a competitive architecture, one approach is to decrease the crosstalk by cascading more MRRs in a dilated switching element [34]. However, such an approach will enlarge the footprint and narrow the passband. For silicon AWGR, the crosstalk can be as low as -25 dB with a comprehensive optimal design of arrayed waveguide width and bi-level tapers in the free propagation region [35].

Silicon nitride (SiN) AWGR is superior to silicon AWGR in mitigating the degraded crosstalk since the variation of the average effective index of the arrayed waveguides is lower [36]. Consequently, the induced phase error is easier to control. Additionally, as a widely applied material in CMOS fabrication, SiN can be equally valuable for PIC fabrication. With ultra-thin (50 nm) core layer design, the transmission loss of the arrayed waveguides can be extremely low (0.4-0.8 dB/cm) due to reduction of scattering caused by side-wall roughness and the crosstalk from adjacent and non-adjacent channels at 1550 nm can be as low as -25 dB and -30 dB, respectively [37]. Alternatively, another paper reports adjacent and non-adjacent crosstalk values as -39 dB and -33.5 dB by optimizing the deposition condition of cladding oxide and cross-section structure [38]. The corresponding BER floor curves as a function of the number of nodes for the three

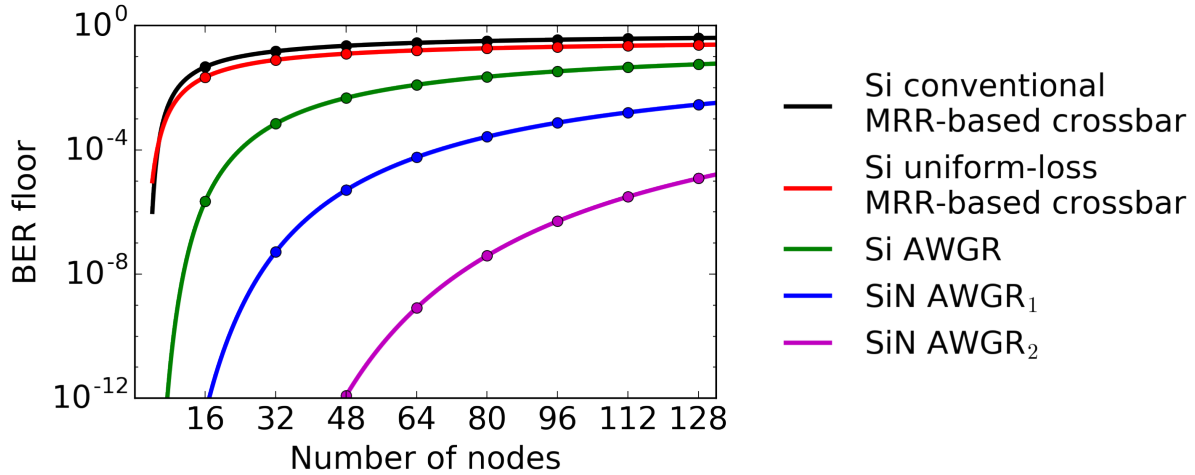


Figure 2.8. Comparison of BER floor among the state-of-the-art silicon conventional crossbar, silicon uniform-loss crossbar, silicon AWGR, and silicon nitride AWGR.

state-of-the-art AWGRs are also plotted in Figure 2.8. It can be seen that the BER floor of SiN AWGR can be  $10^{-12}$  with 48 nodes.

### 2.2.5 Experimental measurements of the crosstalk penalty of AWGR

Figure 2.9 shows the experimental testbed to verify the analytical power penalty induced by crosstalk for all-to-all interconnects using a 32-port silica AWGR from Enablence Technologies Inc. with 100-GHz channel spacing.

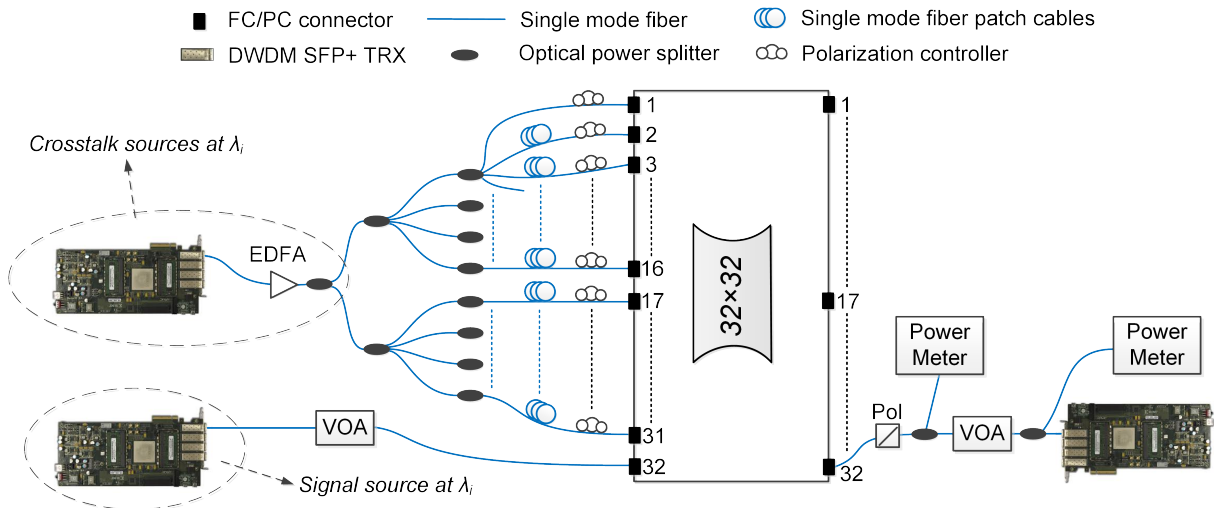


Figure 2.9. Experiment testbed using 32-port AWGR with 100 GHz spacing AWGR. EDFA: erbium doped fiber amplifier; VOA: variable optical attenuator.

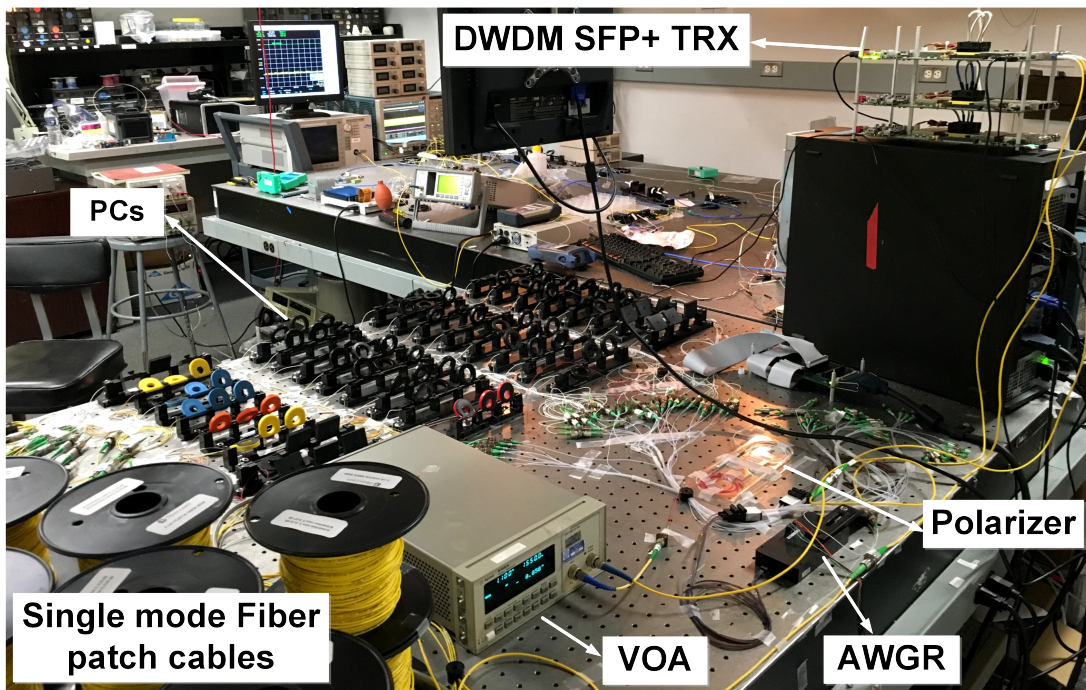


Figure 2.10. Picture of the experiment testbed.

One field-programmable gate array (FPGA) evaluation board with wavelength-specific small form pluggable (SFP+) transceivers at 10 Gb/s acted as the transmitter and receiver to measure the BER at specific wavelength values. A second FPGA board with another SFP+ module at the same wavelength was used to emulate the 31 in-band crosstalk components in an actual all-to-all scenario. The output of this second SFP+ was split into 31 copies and single mode fiber (SMF) patch cables of 50 to 80 meters were used to assure that different copies were decorrelated. It needed to be mentioned that the length of these fiber cables were significantly higher than the coherent length of the SFP+ (5 m for 20 MHz linewidth) guaranteeing that the different copies acted as they were generated from different lasers. Polarization controllers (PCs) placed at each AWGR input were used to align all the polarizations at the output of the AWGR and study the worst-case for the crosstalk scenario (Figure 2.10). A polarization beam splitter at the output port of AWGR assured that the polarizations of the signal and crosstalk components were aligned. The first power meter at the polarization beam splitter (PBS) output was used to maximize the crosstalk power by tuning the PCs at each AWGR input. The crosstalk-to-signal-

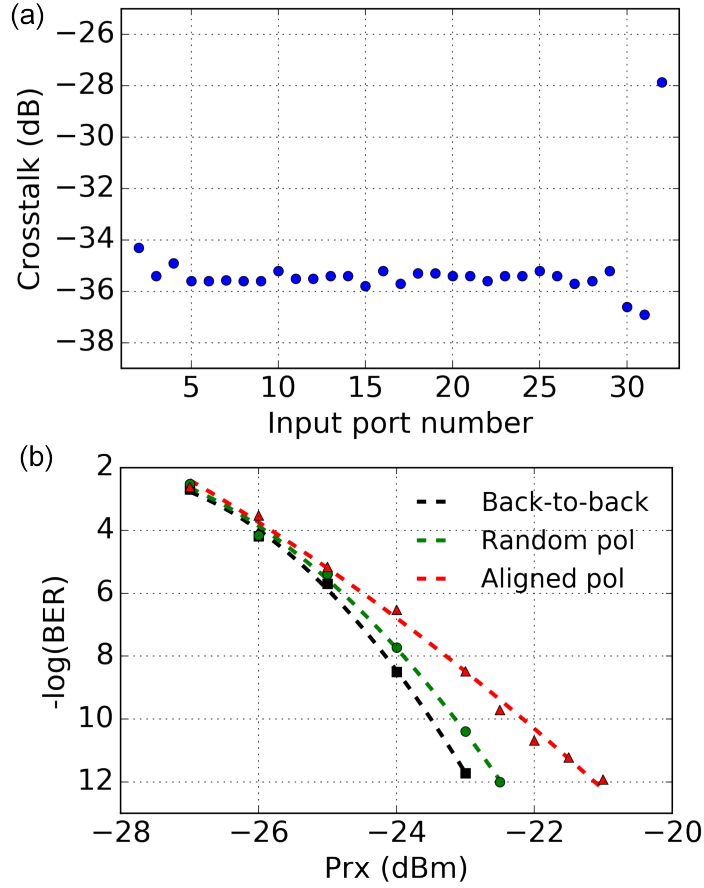


Figure 2.11. (a) Measured crosstalk from input port 2-32 to output port 32 at  $\lambda_{1-32}$ . (b) Measured BER with fixed decision-threshold setting for random polarization (green) and aligned polarization.

power-ratio was measured before and after the PBS to avoid polarization misalignment. The power meter after the variable optical attenuator (VOA) monitored the power at the receiver input of the FPGA measuring the BER.

The insertion loss of the AWGR is not uniform and the worst-case loss is measured to be from input port 1 to output port 32 with a maximum value of 5.6 dB. The corresponding signal wavelength is 1559.8 nm and notated as  $\lambda_{1-32}$ . Additionally, the crosstalk from input port 2-32 to output port 32 at  $\lambda_{1-32}$  is measured and normalized to the insertion loss, as shown in Figure 2.11(a). The average per-port crosstalk contribution is calculated as -35 dB. Figure 2.11(b) plots the measured BER curves for back-to-back, random polarization and aligned polarization settings. The power penalty for BER of  $10^{-12}$  is 2 dB with aligned polarization setting which well matches with the analytical results showing in Figure 2.4.

## 2.3 Foundry-Enabled Si-LIONS Using SiN AWGR and SiPh Transceivers

### 2.3.1 $8 \times 8$ Si-LIONS Chip Design, Layout and Fabrication

SiN/SiO<sub>2</sub> waveguides, compared to silicon/SiO<sub>2</sub> waveguides, offer lower index contrast and lower thermo-optical coefficient [39]. Therefore, they are less sensitive to fabrication imperfections and environmental temperature variations, thus they are more desirable for low-loss [40, 41] and high port count AWGRs.

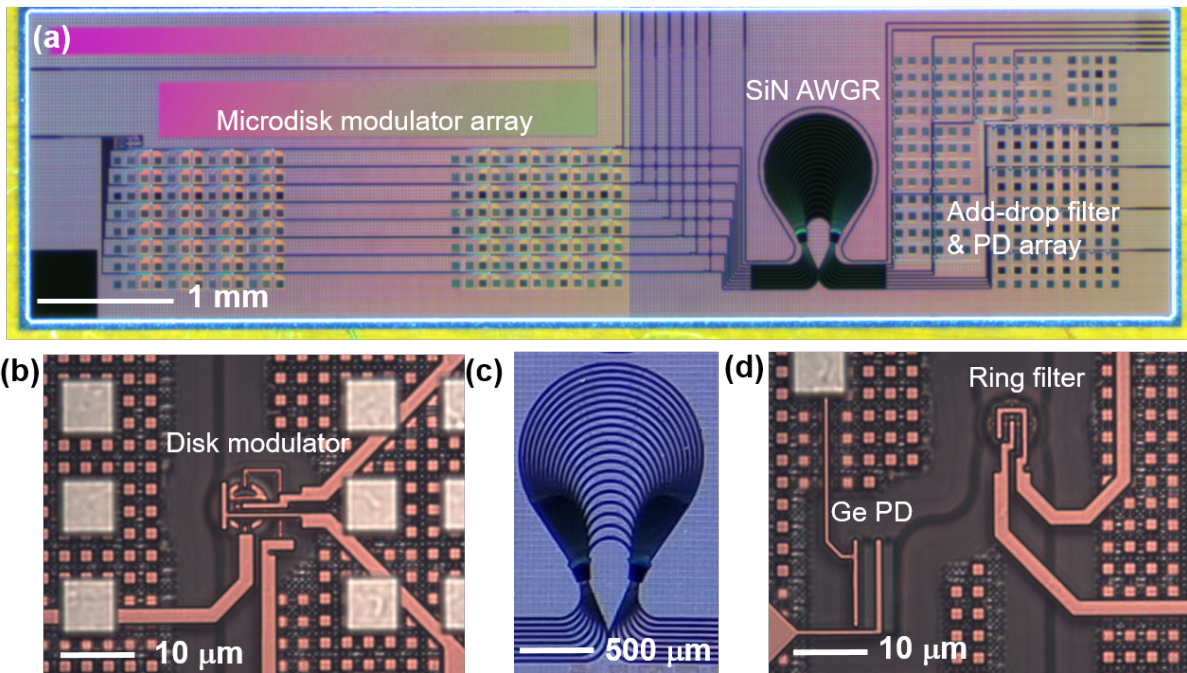


Figure 2.12. (a) Optical microscope picture of the  $8 \times 8$  Si-LIONS system with SiN AWGR and SiPh transmitters and receivers. Zoom-in pictures of (b) a silicon microdisk modulator, (c) the  $8 \times 8$  SiN AWGR and (d) a silicon microdisk filter and Ge photodetector pair.

Figure 2.12(a) shows the microscope image of the  $8 \times 8$  Si-LIONS chip which was laid out using AIM Photonic process design kit (PDK) v2.0 and fabricated through an AIM Photonic MPW run. An eight-wavelength laser emission was coupled into the chip through an edge coupler from the left side of the chip and then split equally into the 8 input waveguides. For each input ports of the  $8 \times 8$  SiN AWGR, there is an array of 8 microdisk modulators as SiPh transmitters and at each output ports, there is an array

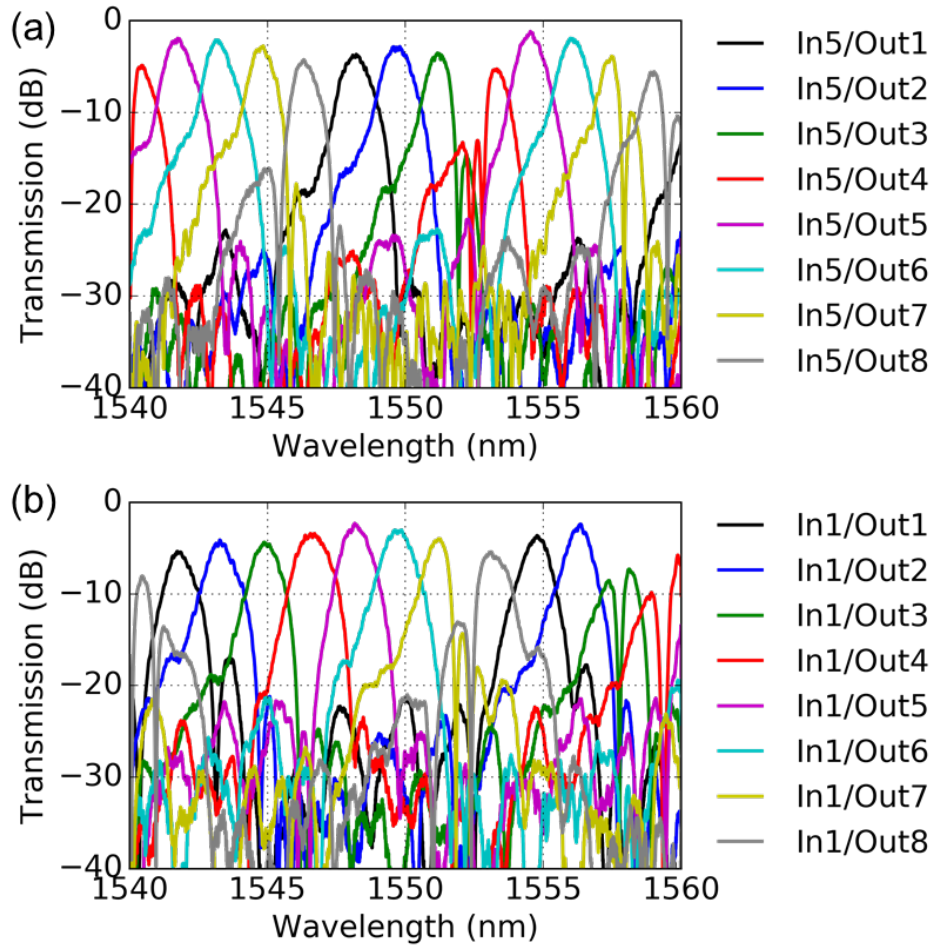


Figure 2.13. Measured transmission spectra of the  $8 \times 8$  SiN from (a) central waveguide input (input waveguide 5) and (b) side waveguide input (input waveguide 1).

of 8 microdisk add-drop filters and photodetectors as the SiPh receiver. The microdisk modulator array and add-drop filters array are designed to have the high-speed metal pads arrangement and spacing to be compatible with a 65 nm technology node electronic driver ICs [42, 43]. A 90/10 coupler is used to monitor the resonance alignment of the modulators at each of the AWGR's input ports.

Figure 2.12 (b)(d) show the zoom-in photographs of a silicon microdisk modulator, the  $8 \times 8$  SiN AWGR and a silicon microdisk filter and Ge photodetector (PD) pair. The microdisk modulator has a diameter less than  $10 \mu\text{m}$ . Therefore, the capacitance is only a few fF with  $\sim 1\text{V}$  drive voltage, which helps to reduce the power consumption compared to popularly used MZ modulators. The designed  $8 \times 8$  SiN AWGR has a footprint of

1.3mm  $\times$  0.9 mm. The channel spacing is designed to be 200 GHz and FSR is designed to 1.6 THz for cyclic wavelength routing performance. The add-drop filter and Ge PD pair works as a wavelength selective receiver.

### 2.3.2 SiN AWGR Characterization

We measured the transmission spectra of the SiN AWGR after diced off 1-to-8 splitters and modulator arrays. The input port end facets are polished after the dicing. Figure 2.13 (a) and (b) show the measured transmission spectra of the  $8 \times 8$  SiN AWGR from central input and side input. Measured spectra are normalized to the wrapped around waveguide with similar length. We extract a 1.8 dB insertion loss and a 13 dB crosstalk for the central input. There is an additional  $\sim 1$  dB loss for the side input. We attribute the relatively high crosstalk to the unoptimized SiN AWGR design. In our previous design with narrower input/output waveguide spacing, we were able to achieve 18 dB crosstalk [44]. The sharp dips in the measured spectra originate from the add-drop filters at the output ports.

### 2.3.3 Microdisk Modulator Characterization

Figure 2.14(a) shows the measured transmission spectra of four modulators at different resonance wavelengths from the test structure. The resonance wavelengths are designed to have 800 GHz spacing. We studied the resonance wavelength uniformity on a single die using a cascaded nine microdisk modulator structure. All nine modulators are designed to have the same resonance wavelength. Figure 2.13(b) shows the measured transmission spectra from the nine-modulator array. We observed a maximum wavelength deviation of  $\pm 1$  nm (125 GHz). Therefore, resonance wavelength tuning is required to align the modulator operating wavelength with the SiN AWGR channel passband.

The modulator element from the AIM photonic foundry’s PDK has a built-in thermal tuner. Figure 2.15(a) shows the resonance red-shift upon different heating power. We extract a tuning efficiency of 0.38 nm/mW (Figure 2.15(b)), corresponds to 65 mW for tuning across a full free-spectral-range (FSR). Further reductions in this thermo-optical tuning power consumption can rise from selectively etching the oxide layer underneath

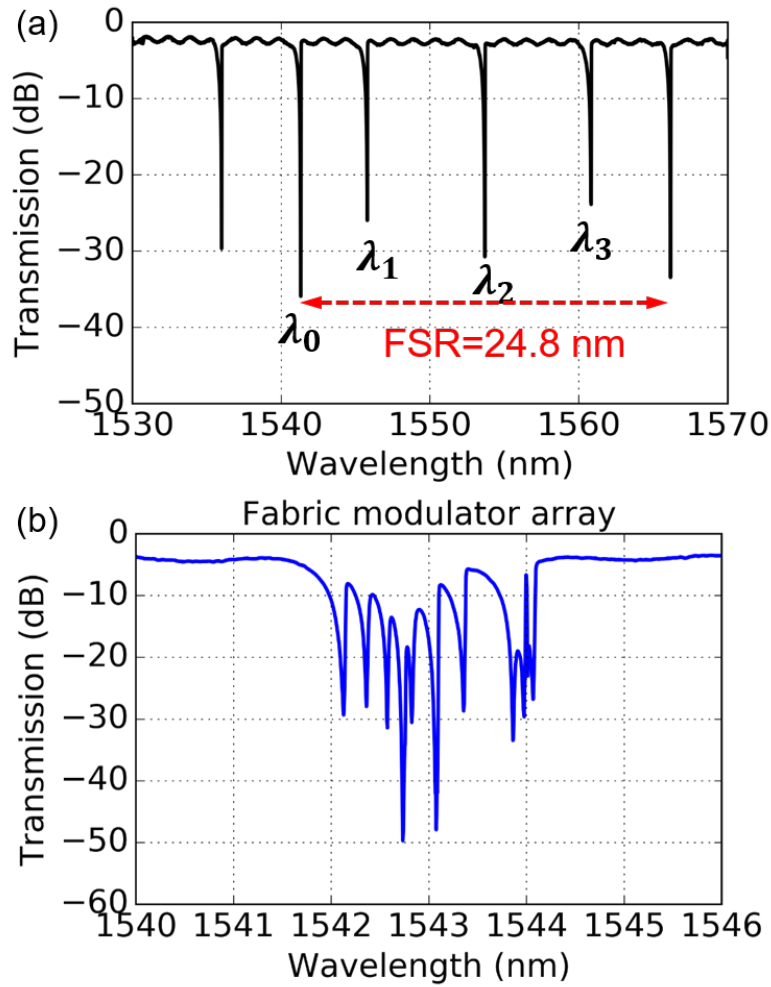


Figure 2.14. Measured transmission spectra of (a) a 4-channel modulator array with different resonance wavelength and (b) a 9-channel modulator array with same resonance wavelength.

[45].

Figure 2.16 shows the measured electro-optical response of the modulator upon different bias voltage on the p-n diode. With 1V swing (0.6V to 0.4V), the modulator reveals an extinction ratio (ER)  $>20$  dB with an insertion loss  $<3$  dB. The inset shows the 10 Gb/s eye diagram from the modulator with 1Vpp. Current modulation speed is limited by our pattern generator and according to the PDK performance, it can operate at a data rate up to 40 Gb/s.



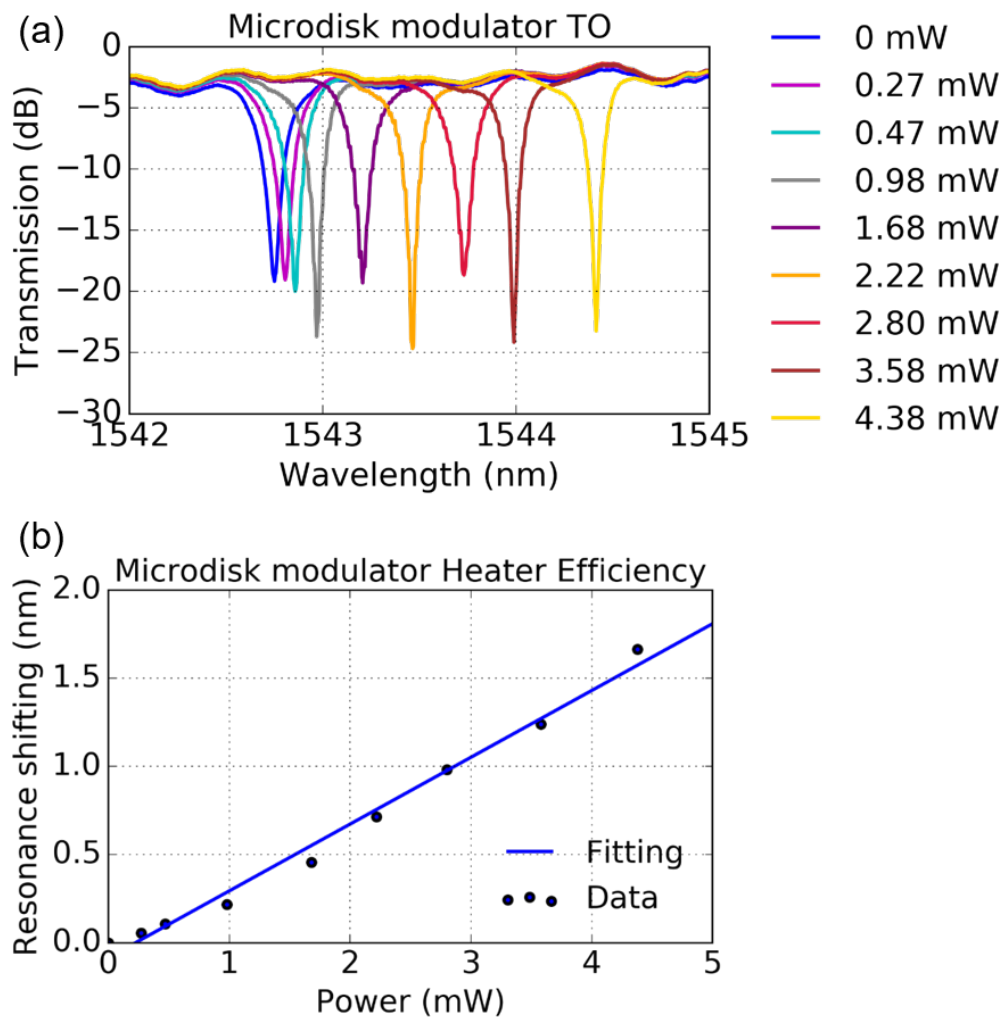


Figure 2.15. (a) Measured transmission spectra of a microdisk modulator upon different heating power and (b) Measured and fitted heater efficiency.

### 2.3.4 Wavelength Routing Experiments

We demonstrated proof-of-concept wavelength routing interconnects using a setup shown in Figure 2.17. A PC is used to ensure TE polarization input and a pair of lensed fibers are used for coupling light into and out of the Si-LIONS chip. The input light of the chip is modulated by one of the on-chip microdisk modulator and routed by the AWGR. The optical signals at the AWGR output is amplified by another EDFA and detected by an external PD. In this experiment we did not use the on-chip Ge PD due to the lack of external transimpedance amplifier (TIA). An RF probe array is used to provide DC signals to the heaters of the microdisk for resonance alignment and inject RF signals to

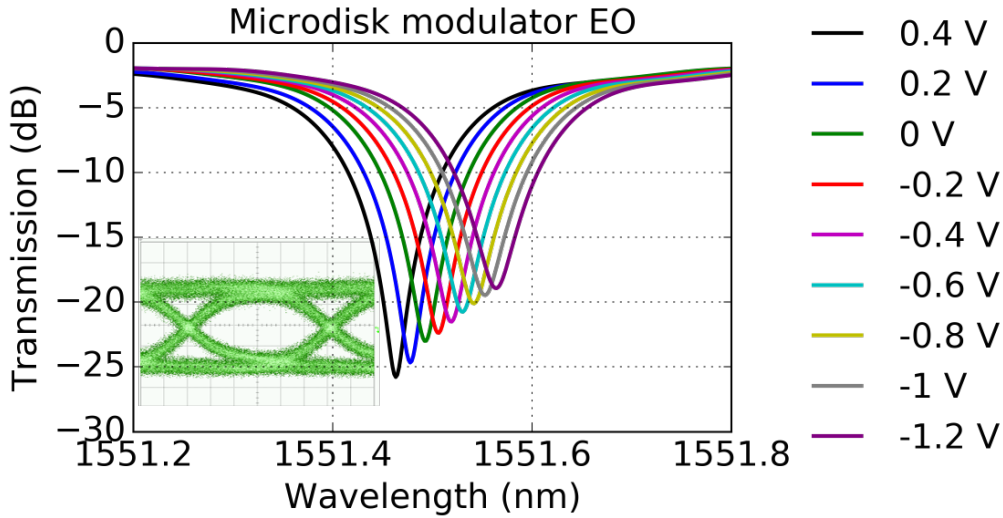


Figure 2.16. Measured transmission spectra of a microdisk modulator upon different bias voltage. Inset: 10 Gb/s eye diagram of a modulator upon 1Vpp swing.

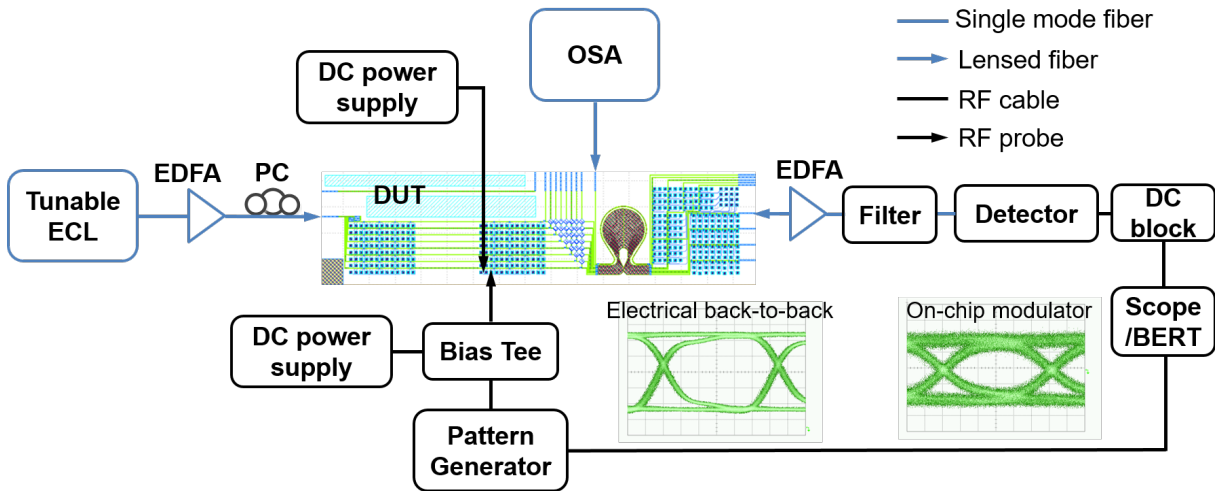


Figure 2.17. Experimental setup for the routing demonstration on the fabricated chip.

the diodes of the microdisk for modulation. Resonance alignment is monitored through the top 90/10 coupler coupled to an optical spectrum analyzer (OSA). The eye diagram is measured by using an oscilloscope.

Figure 2.18 (a) and (b) shows the simulated eye-diagram from the electrical input and from on-chip Ge PD using our Verilog-A based models. The output eye diagram is mainly deteriorated by the coherent crosstalk from the AWGR passband. Figure 2.18 (c)(f) show the eye diagrams of 10 Gb/s OOK transmission from input port 5 to output port 5, input

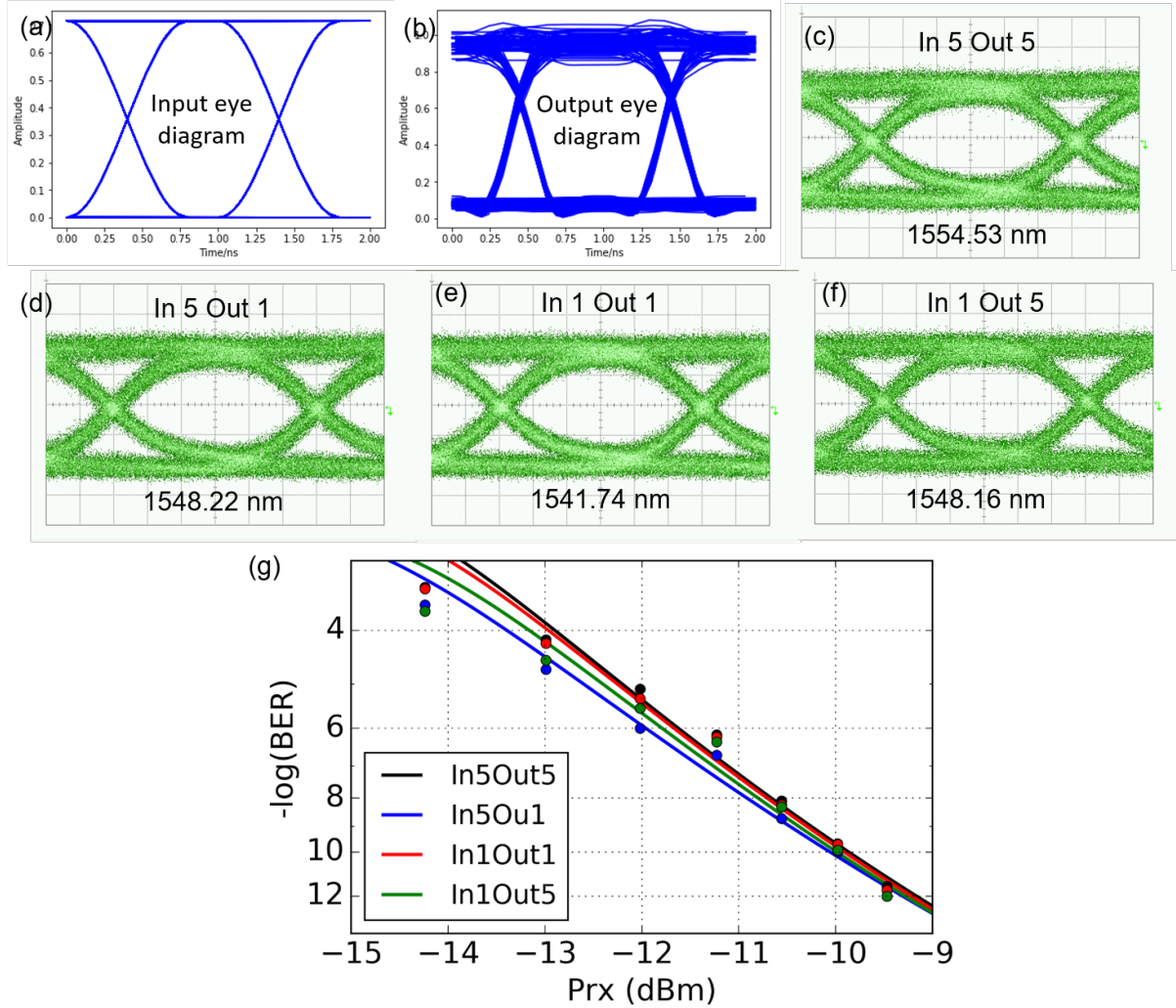


Figure 2.18. Simulated eye diagrams for (a) electrical input and (b) Ge PD output. Measured eye diagrams for (c) input 5 to output 5, (d) input 5 to output 1, (e) input 1 to output 1 and (f) input 1 to output 5. (g) Measured BER curves as a function of received power.

port 5 to output port 1, input port 1 to output port 1 and input port 1 to output port 5. The modulation signal is a  $2^{31} - 1$  pseudorandom binary sequence (PRBS) produced by a pattern generator. The bias voltage is optimized to 0.9 V and the peak-to-peak voltage is set as 1 V. Fig. 8(e) shows the measured BER curves of all four paths. Error-free operations are achieved with received power larger than 10 dBm. BER curves from the on-chip Ge PD will be measured after driver ICs integration.

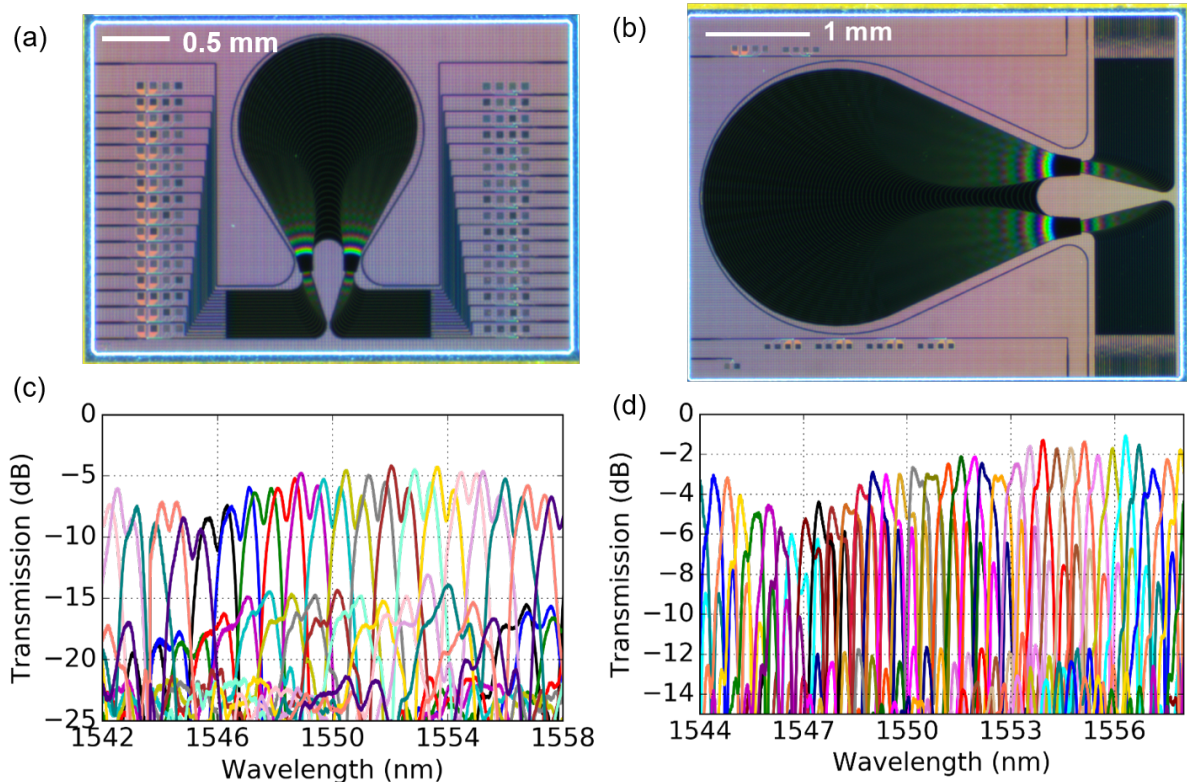


Figure 2.19. Optical microscope picture of: (a) a  $16 \times 16$  SiN AWGR (b) a  $32 \times 32$  SiN AWGR. Measured transmission spectra from the central input of: (c) a  $16 \times 16$  SiN AWGR (d) a  $32 \times 32$  SiN AWGR.

### 2.3.5 Initial Designs and Results on $16 \times 16$ and $32 \times 32$ SiN AWGR

For HPC applications, massive parallelism is preferred to connect thousands of GPUs, CPUs, FPGAs and etc. in flexible and scalable architectures. This requires a large port count possibly exceeding  $1000 \times 1000$ . In the same MPW run, we have taped-out  $16 \times 16$  and  $32 \times 32$  SiN AWGR to explore the scalability of our AWGR devices. Figure 2.19 (a) and (b) shows the optical microscope images of the  $16 \times 16$  and  $32 \times 32$  SiN AWGR with a footprint of  $2.4 \text{ mm} \times 1.6 \text{ mm}$  and  $4.6 \text{ mm} \times 2.9 \text{ mm}$ . Figure 2.19 (c) and (d) shows the measured transmission spectra of the  $16 \times 16$  and  $32 \times 32$  SiN AWGR. We extracted an insertion loss of 5 dB and a crosstalk of 10 dB for the  $16 \times 16$  SiN AWGR and an insertion loss of 2 dB and a crosstalk of 10 dB for the  $32 \times 32$  SiN AWGR. We attribute the relatively high crosstalk partly due to the phase error induced from the fabrication imperfections (sidewall roughness, film thickness variation) in the relatively

large device region. We believe this can be reduced using improved lithography method from a commercial foundry.

## 2.4 Experimental Demonstration of a 64-Port Thin-CLOS Wavelength-Routing System

### 2.4.1 Thin-CLOS AWGR Architecture for Datacenter Switching

Three limiting factors prevent AWGR-based all-to-all interconnects from being practically deployed in a large scale ( $\geq 32$ ) network. First, intra-band (coherent) crosstalk increases significantly as the number of wavelength channels grows [46]. This can significantly impact the BER of the optical links. Second, device size and fabrication constraints can also limit the port count of a single AWGR. Limitations are mainly due to the highly precise control needed for the channel spacing during fabrication as well as accurate wavelength registration required for all channels after fabrication [47]. Third, increasing the port count linearly increases the number of wavelengths, but the limited spectral range results in narrow channel spacing. Therefore, to achieve scalability, it would be desirable to use many smaller AWGRs using a smaller number of wavelengths that can be combined to provide the same interconnectivity offered by a single larger AWGR. This can be achieved by using the proposed Thin-CLOS AWGR.

Figure 2.20(a) shows a generic  $N$ -port Thin-CLOS architecture achieving the same functionality of a single  $N$ -port AWGR with  $N$  wavelengths by using smaller  $W$ -port AWGRs. The architecture is strictly nonblocking and consists of a single layer of  $M$  groups of  $M$  AWGRs with  $W$  ports, subject to  $N = M \times W$ . In summary, as shown in Figure 2.20(b), there are  $M^2$  AWGRs and  $2 \times M^2 \times W$  fiber connections in a Thin-CLOS architecture ( $M^2 \times W$  input ports and  $M^2 \times W$  output ports). Although this translates to a larger number of fibers, connectors, and more complex fiber management inside the enclosure, there are several significant advantages compared with the single AWGR solution:

1. it greatly reduces the in-band crosstalk due to the lower port count of the AWGRs

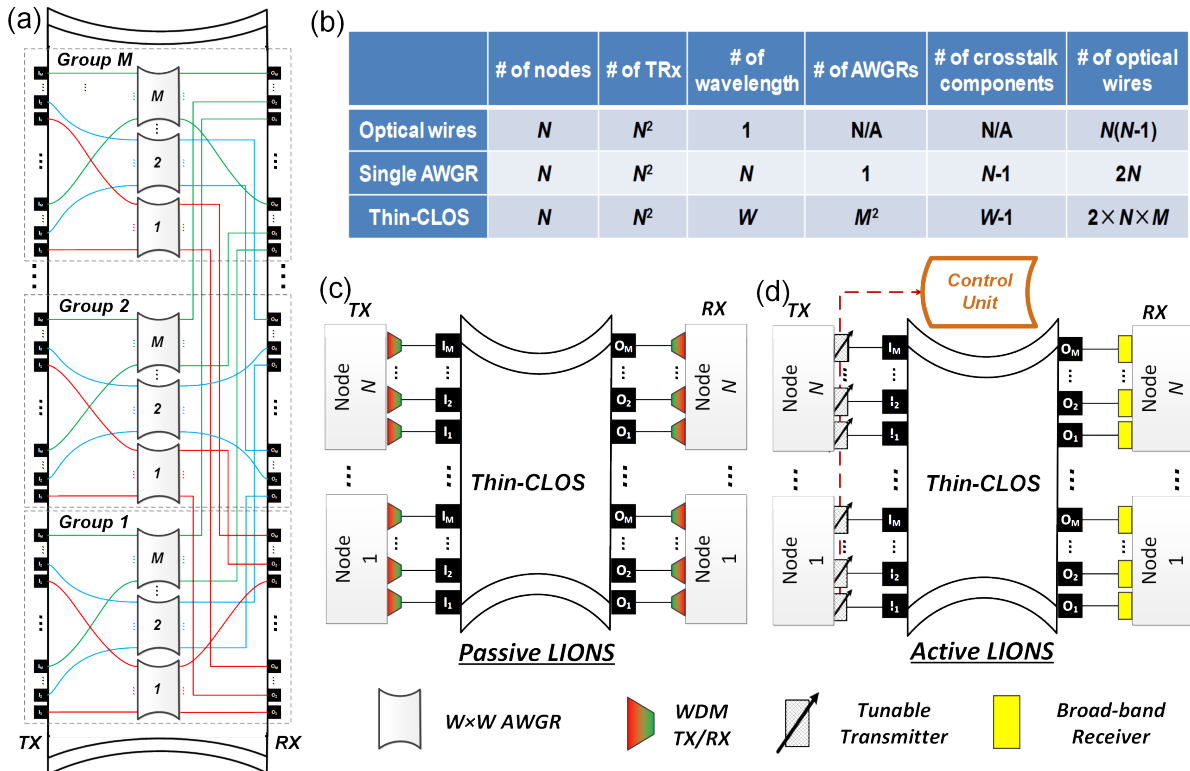


Figure 2.20. (a)  $N$ -port Thin-CLOS AWGR architecture. (b) Table shows a comparison in terms of number of wires, number of AWGRs, and number of wavelengths for all-to-all implementation with optical fibers only, one AWGR and with Thin-CLOS. (c) Use of Thin-CLOS AWGR with WDM TRXs (passive LIONS). (d) Use of Thin-CLOS AWGR with tunable TRXs (active LIONS).

- ( $W-1$  number of crosstalk sources instead of  $N-1$ );
- 2. it offers lower optical losses and reduced optical loss nonuniformity and frequency deviation issues;
- 3. it allows larger channel spacing, relaxes manufacturing tolerance, and relaxes the temperature control requirement (thus reduces power consumption); and
- 4. it improves the yield and reduces the manufacturing cost of the many, identical, and smaller AWGRs.

As explained above, Thin-CLOS architecture mimics the functionality of an  $N$ -port AWGR and can, therefore, can be used as enabling technology to implement different AWGR-based interconnection and switching solutions already proposed in the literature [13, 48–53], which could benefit from a large port-count wavelength routing system. The

Thin-CLOS LIONS can be used as a passive all-to-all wavelength interconnection with WDM lasers or an optical switch with tunable lasers.

The first approach requires each node to use  $M$  banks of  $W$ -wavelength WDM TRXs (Figure 2.20(c)). This architecture is all-to-all; therefore, there is no contention in the optical domain. Switching operation is done at the edges, in the electronic switches (these switches could be either ToR switches or embedded switches in a computing node). This realization has been named a passive AWGR switch or passive low-latency interconnect optical network switch (LIONS) because no optical reconfiguration is necessary. One limitation of this passive approach is the fact that the number of TRXs required to interconnect  $N$  nodes is  $N^2$ . This can be expensive in terms of energy and manufacturing cost, especially when using off-the-shelf WDM TRXs. A more suitable approach for this solution would be to use emerging SiPh WDM TRXs with integrated optical frequency combs [54]. In terms of performance, in the case of uniform random traffic, the throughput is 100%, and the latency is constant with the load and is simply equal to the transmission time and propagation delay because the interconnection is flat and single-hop without contention among the  $N$  nodes. To scale the number of nodes beyond what is possible with a single Thin-CLOS AWGR, a hierarchical approach could be used.

In the second approach with tunable lasers (Figure 2.20(d)), each node requires using only  $M$  tunable TRXs. Therefore, the solution is more affordable because the total number of TRXs is now  $M \times N$ . However, this solution requires careful centralized scheduling or contention resolution schemes in the optical domain (indicated as a generic block called control unit in Figure 2.20(d)). Overall, the complexity of this approach is higher. The switch performance in terms of latency and throughput depends on the specific contention resolution scheme, laser switching time, and packet sizes. Assuming fast tunable lasers with switching time in the order of a few nanoseconds and burst-mode RXs, the throughput can be still above 70% for packet sizes as small as 256 bytes, even if the number of RXs per node ( $M$  in this case) is  $\ll N$ . This is because the AWGR wavelength routing principle naturally implements output queuing architecture with a speedup of  $M$ . Therefore, under a uniform random traffic profile, the contention probability reduces significantly.

## 2.4.2 Design, Fabrication, and Experimental Demonstration of a 64-port Thin-CLOS

As mentioned above, Thin-CLOS can reduce the number of wavelengths and crosstalk components at the expense of increasing the number of fibers and AWGRs. As shown in Figure 2.20(b), the number of fibers scales as  $2 \times M^2 \times W$ . Therefore, to limit the number of fibers and AWGRs, we chose to investigate a design with  $M^2 = \text{four } 32 \times 32$  AWGRs, so that all the components could fit into a 1U rack unit. The 64 ports are divided into two groups: 1 to 32 and 33 to 64. Each port has two pairs of fibers that would connect to a node.

Each node has two sets of transmitters (Tx) and receivers (Rx) with 32 wavelengths. Intra-group interconnects are implemented by the first and fourth AWGRs while cross-group interconnects are supported by the second and third AWGRs.

While  $M = 2$  represents the best solution regarding the number of AWGRs and fibers required, it is essential to verify that the 32-port AWGR from Enablence Technologies, Inc. could guarantee error-free conditions under the worst-case in-band crosstalk scenario in all-to-all configuration. Towards this aim, we performed a detailed experimental analysis of a crosstalk-induced power penalty. The section below reports the experimental setup and results showing that Enablence 32-port AWGR can deliver error-free performance for all-to-all communication. Note that the crosstalk penalty in a 64-port Thin-CLOS system solely depends on the 32-port AWGRs inside the Thin-CLOS enclosure. Our experiment used 32 signals at the same wavelength with aligned polarization and for worst-case loss paths. Therefore, the power penalty measurements can be considered representative of a 64-port system at full-scale.

Figure 2.21 shows the experimental setup used to assess the crosstalk-induced power penalty when using Enablence 32-port AWGR. One FPGA evaluation board with 10 Gbs DWDM SFP+ TRX acted as a transmitter and receiver to measure the BER at a given wavelength (we can choose the wavelength by using different TRX modules). To emulate the 31 crosstalk components at the same wavelength of the signal under test, we split the output of the SFP+ TX. In this way, we assured that the crosstalk wavelength



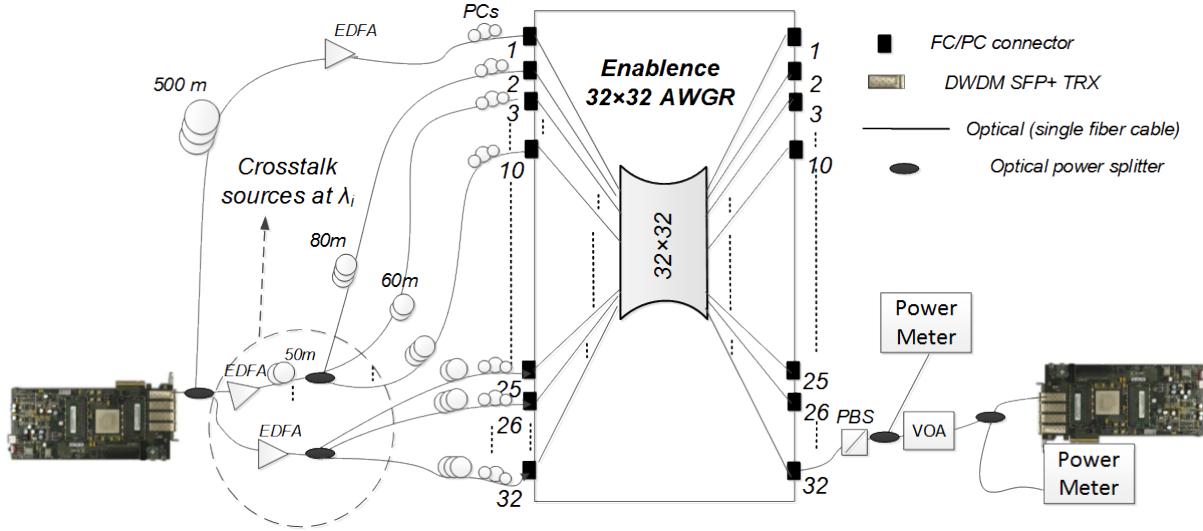


Figure 2.21. Experiment setup using Enablence 32-port OSD based on single 100 GHz spacing AWGR. PM, power monitor; PBS, polarization beam splitter; EDFA, erbium-doped fiber amplifier; VOA, variable optical attenuator; PC, polarization controller.

signals were exactly matching with the one of the test signals for maximum interference condition. Fiber delay lines of several tens of meters (much longer than the coherence length of the SFP+ laser -5 m for 20 MHz linewidth) were used to guarantee that the crosstalk signals acted as independent laser sources. These delay lines also assure that the 31 different copies were decorrelated. Polarization controllers were used together with a polarizer placed at the AWGR output to align all the polarizations for the worst-case crosstalk scenario. The crosstalk-to-signal ratio was measured before and after the PBS to assure the one after the PBS was not improved due to polarization filtering caused by polarization misalignment. When choosing the wavelength of the signal under test, we considered the fact that Enablence AWGR loss is not uniform. Therefore, as the first step, we determined the worst-case loss when using side and center inputs of the AWGR (inputs 1 and 17). When using input 17, the worst-case loss is when the signal must reach output 31. We measured a loss of 4.7 dB, which is in agreement with the data sheet provided by Enablence.

However, when using input 1, we measured a worst-case loss of 6 dB when using output port 32 (Table 2.3). Thus, we selected the wavelengths associated with these input and output combinations to test the BER performance under these worst-case loss

and crosstalk scenarios. We also measured the in-band crosstalk contribution given by each of the other 31 ports when using the wavelengths selected. Figure 2.22 shows the results of these measurements. The average per-port crosstalk contribution (normalized to the signal output power or AWGR loss) was -34.85 dB and -37.39 dB at outputs 31 and 32 for  $\lambda_{1-32}$  and  $\lambda_{17-31}$ , respectively.

Input	Output	Wavelength (nm)	Insertion loss (dB)
1	1	1558.983	5.35
1	2	1558.173	3.95
1	16	1546.917	3.17
1	17	1546.119	3.23
1	31	1560.614	5.39
1	32	1559.800	6.00
17	30	1548.515	3.16
17	31	1547.715	4.68
17	32	1546.917	3.76
17	-	Data sheet	1.52-4.53

Table 2.3. Worst-case insertion loss of the 32-port AWGR from center input (17) and side input (1).

According to the analytical results in Figure 2.23, the crosstalk values above should guarantee error-free operation with limited power penalty. We confirmed this by using the selected wavelengths for the experiment shown in Figure 2.21. Figure 2.24 report the BER measurements for the worst cases discussed above. We considered the worst-case path loss from the AWGR datasheet provided for center input 17 as well as the worst-case loss path measured in the lab for side input 1. The power penalty difference is 0.5 dB. We also plotted the BER curve in the case of nonaligned polarizations to emphasize the importance of designing the system for the worst-case scenario.

Note that each measurement has been carried out using a different SFP+ DWDM TRXs for the specific wavelength needed to communicate between two specific ports. We noticed that different TRXs have slightly different RX sensitivities. This explains why

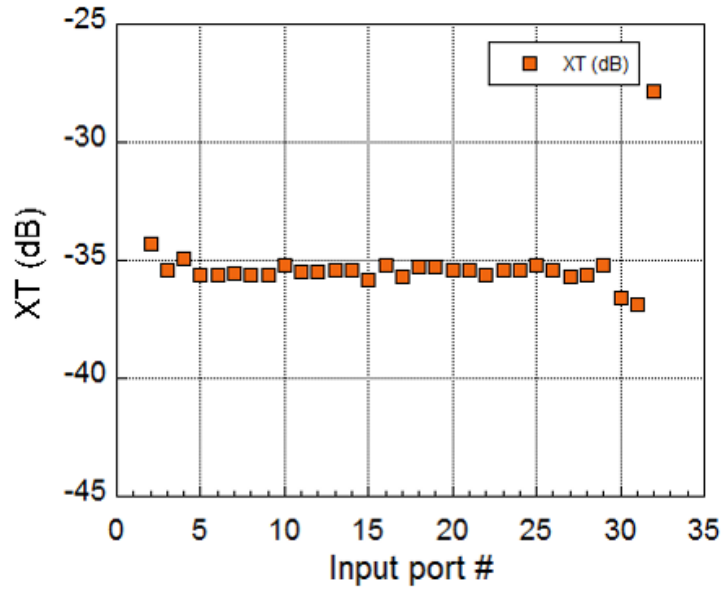


Figure 2.22. Crosstalk measurements at output 32 from input 1 at  $\lambda_{1-32}$  (worst-case loss from input 1).

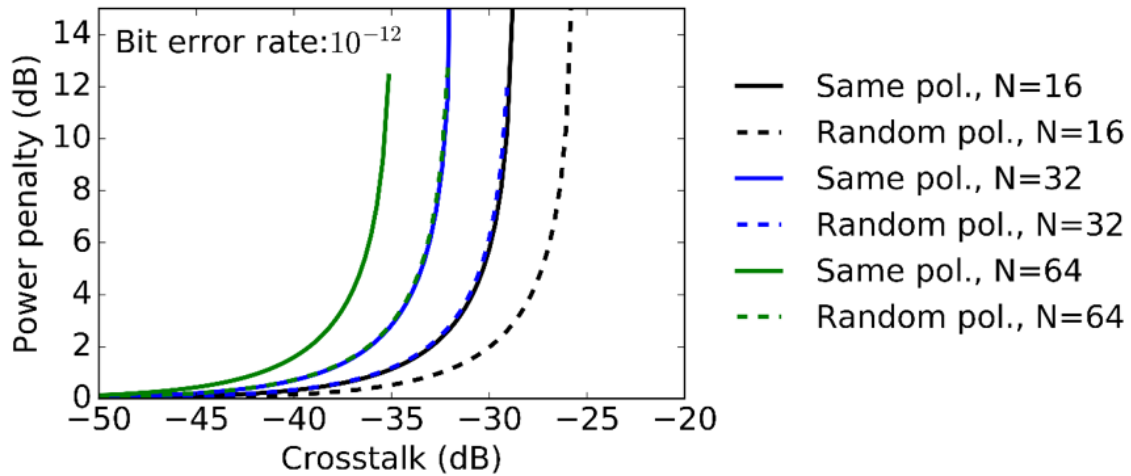


Figure 2.23. Analytical results of power penalty as a function of AWGR crosstalk for the different AWGR port count with random polarizations (polarization state is unknown) and worst-case aligned polarizations (polarizations aligned in parallel).

the back-to-back curves are slightly different. In any case, the power penalty is always  $< 3$  dB for  $\text{BER} = 10^{-12}$  with PRBS testing sequence of  $2^{31} - 1$ .

The experiment results reported above demonstrated that it is feasible to use 32-port AWGRs to build a 64-port Thin-CLOS architecture. Here, we discuss details and solutions adopted regarding connectors and fiber management necessary to guarantee the correct

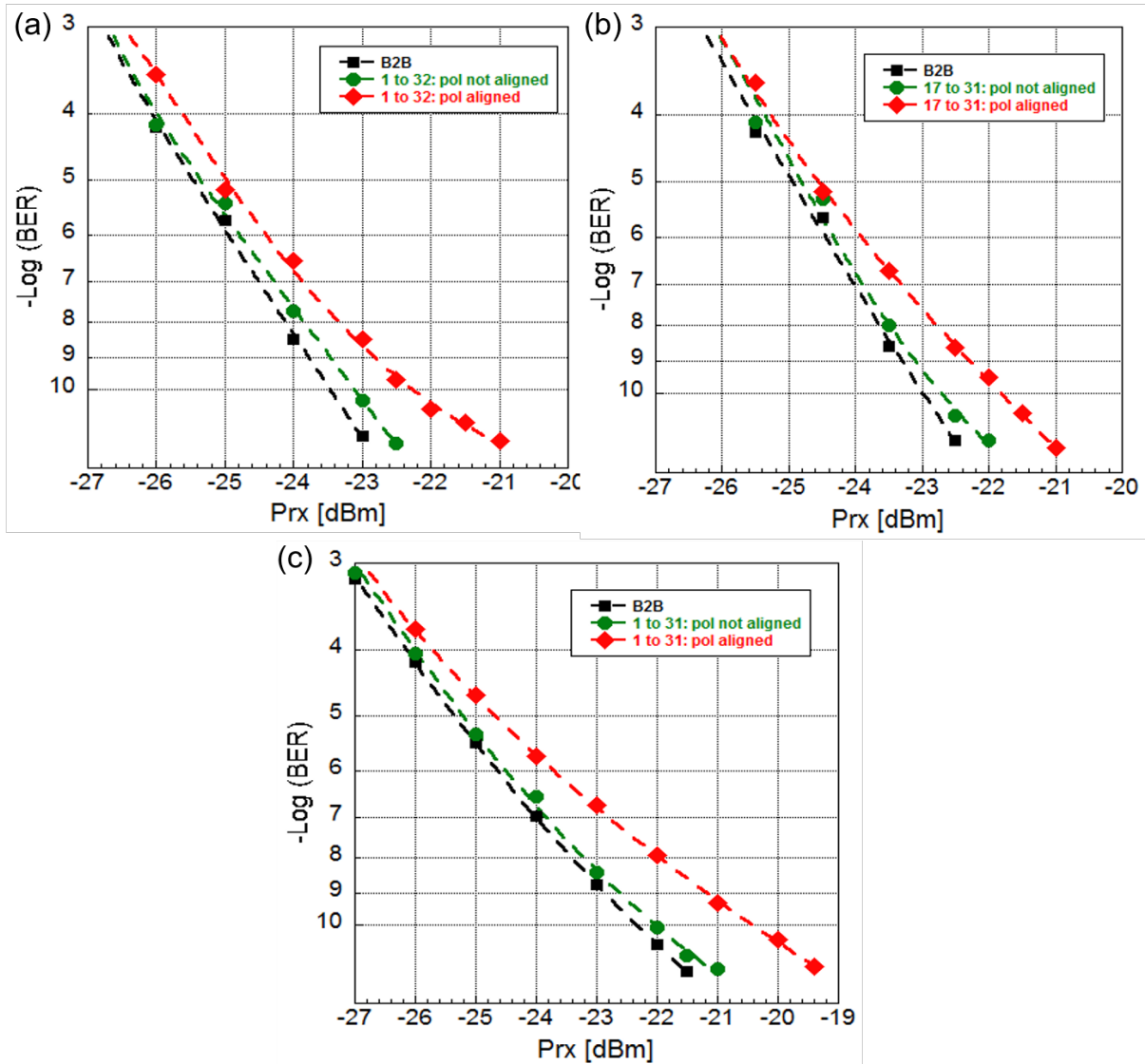


Figure 2.24. End-to-end link experiment with 32 signals at the same wavelength and worst-case polarization alignment for: (a) Input port 1 to output port 32; (b) Input port 17 to output port 31; Input port 1 to output port 31.

functionality and connectivity required for a Thin-CLOS with  $M = 2$  that fits in a 1U rack enclosure.

First, a 64-port all-to-all Thin-CLOS architecture has 128 input fibers and 128 output fibers. To accommodate these 256 connections on the front panel of a 1U rack enclosure, we determined that it was necessary to make use of high-density connectors, i.e., MTP connectors. We used 16 MTP connectors and cables, each one carrying a bundle of 16 fibers (these are custom-made MTP cables because the legacy commercial solutions carry

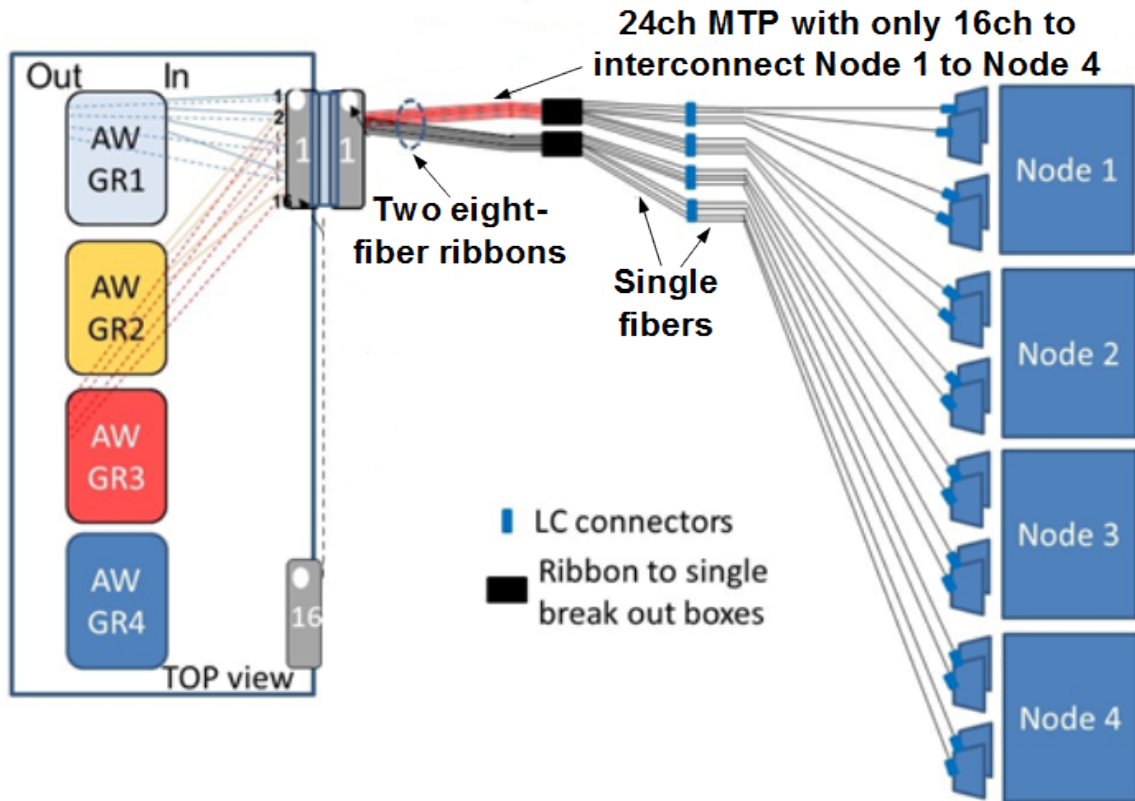


Figure 2.25. Connectivity between MTP1 and nodes 1 to 4. The MTP cable carries two eight-fiber ribbons, which break out into single LC fiber cables for connection with WDM muxes and demuxes at each node.

24 fibers). Each MTP connector carries input and output fibers for four nodes (four fibers per nodes), as shown in Figure 2.25. In fact, each node connected to the designed 64-port Thin-CLOS architecture would have  $M = 2$  WDM transmitters and receivers, requiring then  $M = 2$  input and  $M = 2$  output fibers. Thus, in each MTP connector, the first group of four pins are for the first node, the second group of four pins are for the second node, and so on. Also, pins with odd index numbers (i.e., 1, 3, 5, etc.) connect to the nodes' TXs, while pins with even index numbers (i.e., 2, 4, 6, etc.) connect to the nodes' RXs. Figure 2.25 shows an example of connectivity between MTP1 and nodes 1 to 4.

Once the MTPs' pins assignment for the connectors facing the end-nodes was determined, it was necessary to carefully determine the connections between each MTP pin inside the enclosure and the AWGRs input and output fibers. Note that, as explained above, in a Thin-CLOS with  $M = 2$ , the ports (nodes) are organized into two groups.

The first group of 32 makes use of AWGR 1, 2, and 3. The second group of 32 makes use of AWGR 4, 3, and 2. Therefore, because each MTP serves four nodes, MTP 1 to 8 will belong to the first group, while MTP 1 to 9 will belong to the second group. Table IV explicates the rule that needs to be followed to connect MTPs in the first and second groups to the AWGR fibers inside the enclosure. The example in Table 2.4 is given for MTP 1 (node 1 and node 2, group 1) and MTP 9 (node 33 and 34, group 2).

Node	MTP 1 Pin#	AWGR I/O type	AWGR I/O#	AWG#
1	1	I	1	1
1	2	O	1	1
1	3	I	1	2
1	4	O	1	3
2	5	I	2	1
2	6	O	2	1
2	7	I	2	2
2	8	O	2	3
-	MTP 9 Pin#	AWGR I/O type	AWGR I/O#	AWG#
33	1	I	1	4
33	2	O	1	4
33	3	I	1	3
33	4	O	1	2
34	5	I	2	4
34	6	O	2	4
34	7	I	2	3
34	8	O	2	2

Table 2.4. Required connections between MTPs in group 1 (MTP 1 to 8) and group 2 (MTP 9 to 16) and the AWGR input and output fibers.

The intra-band worst-case crosstalk for the fabricated 64-port Thin-CLOS is the same as the worst-case crosstalk scenario for one of the  $32 \times 32$  AWGRs inside the enclosure. The measurements above already show that the 32-port silica AWGR crosstalk rejection

value guarantees error-free operation under a worst-case scenario. The goal of the experiment setup of Figure 2.26(a) is to verify and demonstrate the correct routing operation in the 64-port Thin-CLOS system. One FPGA evaluation board with wavelength-specific SFP+ transceivers at 10 Gbs acts as the TX and RX to measure the BER curves for different input and output nodes combinations. The FPGA connects to specific ports of the Thin-CLOS enclosure, as shown in Figure 2.26(a). A power meter is used to monitor the signal power at the receiver. Figure 2.26(c) plots the BER curves for intra-group (Node 1 to Node 32, Node 64 to Node 63) and cross-group (Node 1 to Node 64, Node 64 to Node 32) interconnects using different AWGRs inside the 1U enclosure. The difference between the four BER curves is simply related to the different sensitivity values of the commercial SFP+ DWDM TRXs used in this experiment. The wavelength values used in the experiment are summarized in Table 2.5 and are determined by the wavelength routing table of the four AWGRs (all the AWGRs have the same specifications and wavelengths).

Outputs Inputs	Node 32	Node 63	Node 64
Node 1	1557.41 nm	1555.79 nm	1556.59 nm
Node 64	1535.98 nm	1560.76 nm	1561.56 nm

Table 2.5. Wavelength values used in the Thin-CLOS routing experiment.

### 2.4.3 Crosstalk and Power Budget Analysis

The system crosstalk power penalty discussed above is of fundamental importance for performing a power budget analysis and determining whether the average optical power at the RX will be enough to meet the requirements in terms of BER (i.e.,  $\text{BER} \leq 10^{-12}$ ). The equation below represents the lower bound for the average optical power of each optical transmitter connecting to the all-to-all system in dB scale:

$$P_{\text{TX}} \geq P_{\text{RXSens}} + AWGR_{\text{IL}} + AWGR_{\text{muxIL}} + AWGR_{\text{demuxIL}} + \text{Connector}_{\text{IL}} + \text{Penalty}_{\text{XT}} \quad (2.20)$$

where  $P_{\text{RXSens}}$  is the RX sensitivity, defined as the minimum required optical power to guarantee error-free performance in the absence of any source of impairments other than

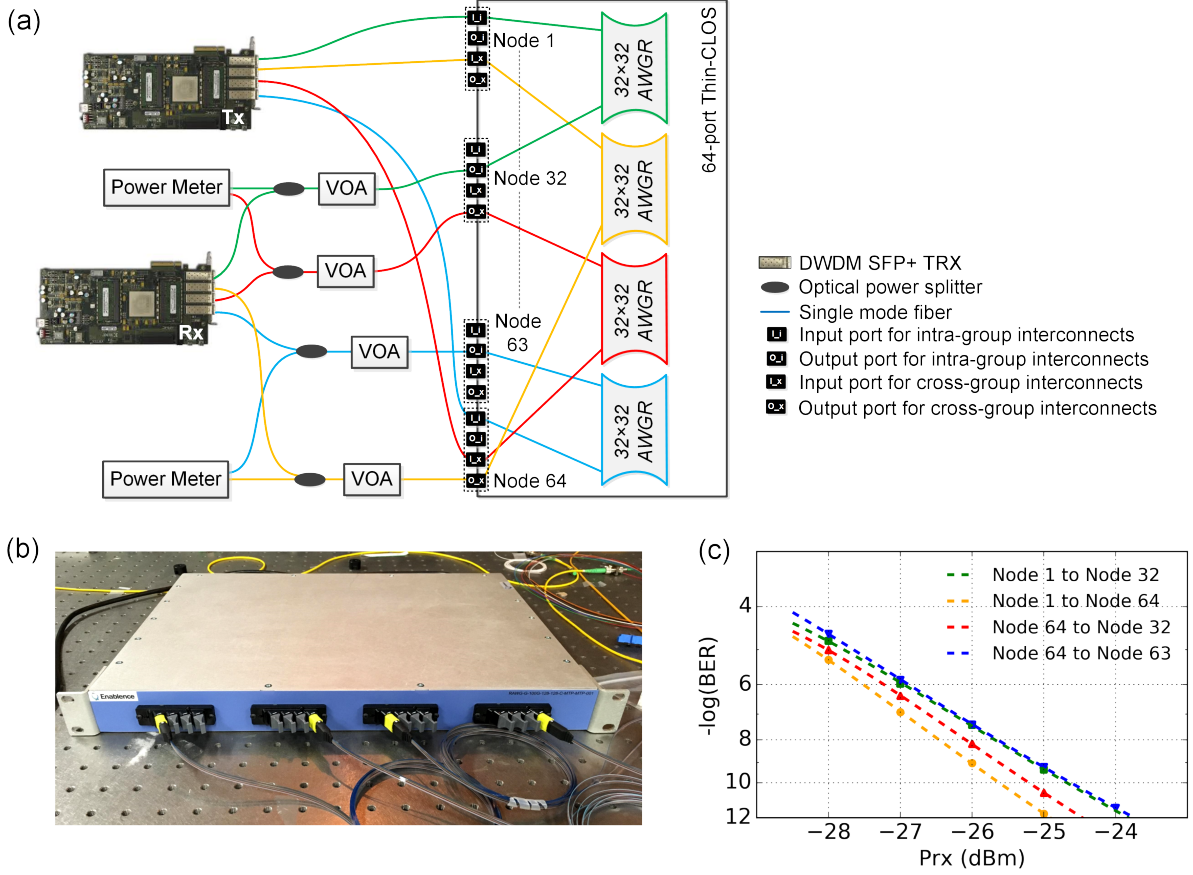


Figure 2.26. (a) Experimental testbed for verifying Thin-CLOS functionality. (b) Picture of the fabricated Thin-CLOS system in 1U rack size. (c) Measured BER with different input and output node combinations.

shot noise and thermal noise at the RX;  $AWGR_{IL}$  is the AWGR insertion loss between each input and output port;  $AWG_{mux_{IL}}$  is the insertion loss of the optical wavelength multiplexer located at each end-point;  $AWG_{demux_{IL}}$  is the insertion loss of the optical wavelength demultiplexer located at each end-point;  $Connector_{IL}$  is the insertion loss of all the connectors located between any pair of TX and RX;  $Penalty_{XT}$  is the power penalty as defined in the above sections.

Based on the values reported in Table 2.6, it is easy to calculate that the required  $P_{TX}$  is only -9 dBm for the fabricated  $K_{64,64}$  OSD with  $W = 32$ ,  $M = 2$ . This requirement is about 10 dB lower than the typical output power of the commercial SFP+ DWDM TRXs used in the experiment.



-	Definition	Typical Values
$P_{TX}$	SFP+ transmitter optical power	0 dBm
$P_{RXSens}$	Receiver sensitivity	-25 dBm
$P_{RXMax}$	Maximum receiver optical input power	-10 dBm
$AWGR_{IL}$	AWGR insertion loss	4 dB
$AWGmux_{IL}$	Multiplexer insertion loss	2 dB
$AWGdemux_{IL}$	Demultiplexer insertion loss	2 dB
$Penalty_{XT}$	Power penalty induced by in-band crosstalk	<3 dB
$Connector_{IL}$	Loss due to connectors	$\leq 1$ dB

Table 2.6. Typical power budget related values measured in the experiment.

## Chapter 3

# Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable All-to-All Optical Interconnects

Today's HPC and datacenter systems are growingly adopting heterogeneous memory and processor nodes (Figure 3.1) to better utilize resources for various tasks. The communication patterns in such systems driven by modern workloads tend to be temporally bursty and spatially nonuniform. The hotspots and coldspots simultaneously created in different locations in the network can lead to heavy congestions in some links, while others are poorly utilized, negatively affecting the overall throughput and energy efficiency performance. However, today's interconnection networks based on electronic switches and optical fibers are inherently rigid, incapable of changing the network topology or link bandwidth, while adaptive routing techniques cannot adequately cope with the significant variations of traffic patterns. On the other hand, all-to-all interconnections are essential for many applications, including map-reduce based applications, parallel sorting applications, and DNN applications. It would then be desirable to design a bandwidth-reconfigurable interconnection network that can support all-to-all connectivity and adapt its connectivity to the traffic demand of hotspots when necessary.

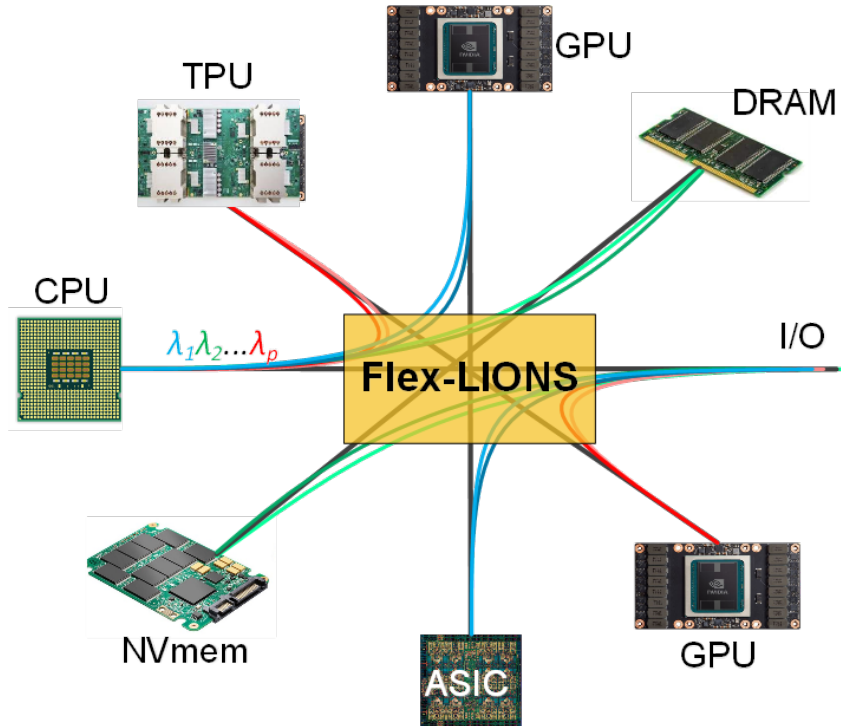


Figure 3.1. Modern HPC systems with heterogeneous processor and memory nodes.

### 3.1 System Demonstration of Flex-LIONS With Off-The-Shelf Components

In this section, we propose and experimentally demonstrate a 2.56Tb/s 16-port Flex-LIONS whose optical interconnection topology can be mutually reconfigured between fully all-to-all and partially all-to-all using off-the-shelf components.

#### 3.1.1 Architecture and experimental setup

The wavelength routing property of AWGR can significantly reduce the complexity of all-to-all topologies by taking advantage of WDM technology as shown in Figure 3.2(a). Low-Latency Interconnect Optical Network Switch (LIONS) is the implementation of such all-to-all optical interconnected networks, in which each node uses different wavelengths to communicate with all the other node. As for Flex-LIONS (Flexible Low-Latency Interconnect Optical Network Switch), we are aiming to switch the topology to partially all-to-all in which some pairs of nodes can be reconfigured to directly connect using all the wavelengths as showing in Figure 3.2(b).

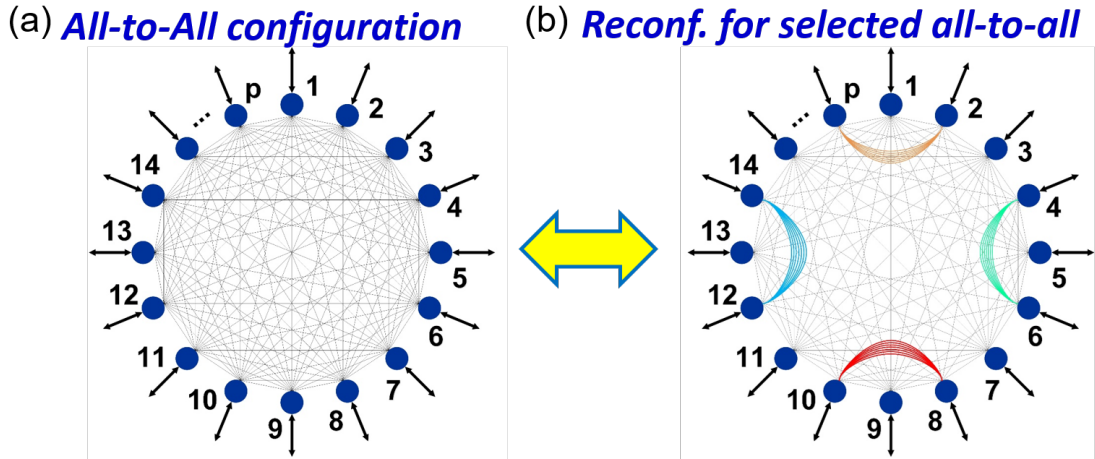


Figure 3.2. (a) State 1: Fully all-to-all topology based on AWGR. (b) State 2: Partially all-to-all topology in which some pairs of nodes can be reconfigured to directly connect using all the wavelengths.

The reconfigurability of an  $N$ -node Flex-LIONS is enabled by using two optical switches with  $N + m$  ports and one  $N \times N$  AWGR, as shown in Figure 3.3. For State 1, all the connections between each pair of nodes are through the AWGR and the inter-node communication bandwidth,  $B$ , is limited by the bandwidth of a single WDM transceiver. For State 2, the wavelengths from some nodes ( $m$  nodes at maximum) can be switched to bypass the AWGR and directly connect to other nodes using the two optical switches. The total bandwidth can then be increased by a factor of  $N$ .

Figure 3.4 shows the experimental setup of the 16-port Flex LIONS ( $N=16$ ,  $m=16$ ). Four FPGA evaluation boards with 10 Gb/s DWDM SFP+ transceivers were used to generate 16 signals with 100-GHz channel spacing (from 1548.51 nm to 1560.61 nm). Then, the generated signals were combined by a  $16 \times 1$  AWG multiplexer, amplified by an EDFA, and split into 16 copies to represent 16-wavelength signals from 16 different nodes. SMF patch cables of 40 to 80 meters were used to assure that different copies were decorrelated. The lengths of these fiber cables are significantly higher than the coherent length of the SFP+ (5 m for 20 MHz linewidth) guaranteeing that the different copies acted as if they were generated from different lasers. Then, the 16 copies entered the left 32-port optical spatial switch and were switched to either go through the  $16 \times 16$  AWGR or to directly connect to the second optical spatial switch, whose outputs were connected

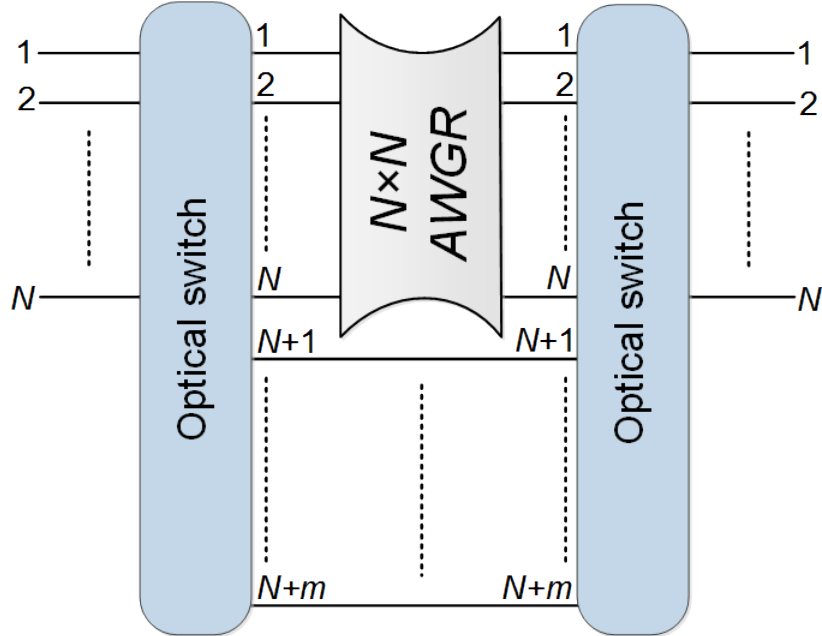


Figure 3.3.  $N$ -node Flex-LIONS using two  $(N + m)$ -node spatial switches and one  $N \times N$  AWGR

to the receiver. The receiver included an EDFA, a  $1 \times 16$  AWG demultiplexer, and one SFP+ receiver. We used the same SFP+ module to guarantee the same BER receiver sensitivity for all the measurements. Figure 3.5 shows the picture of the testbed.

The inset of Figure 3.4 shows the optical spectrum after the  $16 \times 1$  multiplexer. The power per wavelength is nearly 0 dBm. Note that the two EDFAs compensated the splitting loss introduced by the  $1 \times 16$  splitter and ensured that the power per wavelength at the receiver was higher than the receiver sensitivity. In an actual deployment, each of the 16 nodes would have its own set of 16 transceivers and no power splitters nor EDFAs would be needed.

### 3.1.2 Measurement results: 16-Port Flex-LIONS

First, we demonstrated the error-free operation and reconfiguration capability of the 16-node Flex-LIONS. Figure 3.6(a) (left) shows the fully all-to-all topology (State 1) and the red, green, blue, and yellow lines indicate that the interconnections between node 2 to 4, 6 to 8, 10 to 12, and 14 to 16 are through the AWGR using a single wavelength. Figure 3.6(a) (right) shows the BER curves of these interconnections and the power penalty is less than 1 dB at BER of  $10^{-12}$  compared to back-to-back (measured by directly connecting the

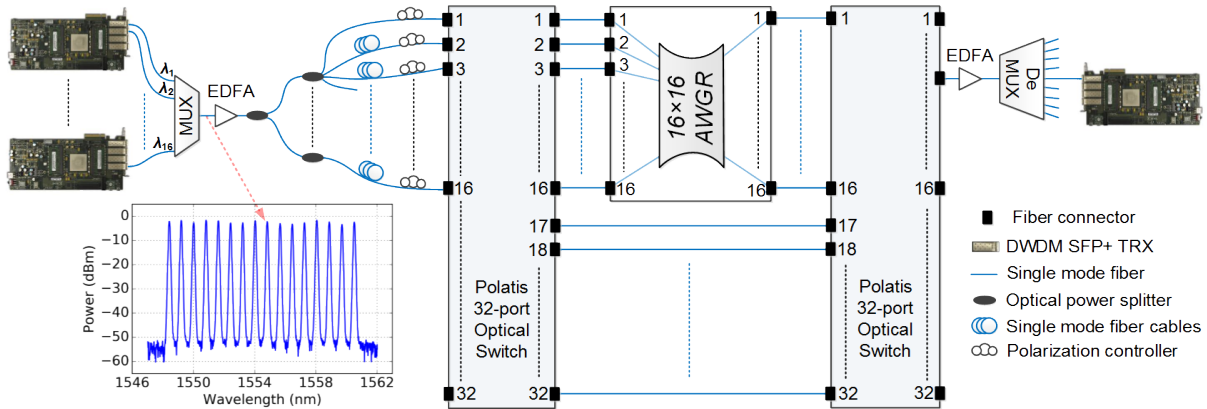


Figure 3.4. Experimental setup of 16-node Flex-LIONS. EDFA: erbium doped fiber amplifier; MUX: AWG multiplexer; DeMUX: AWG demultiplexer; Inset is the optical spectrum after the multiplexer.

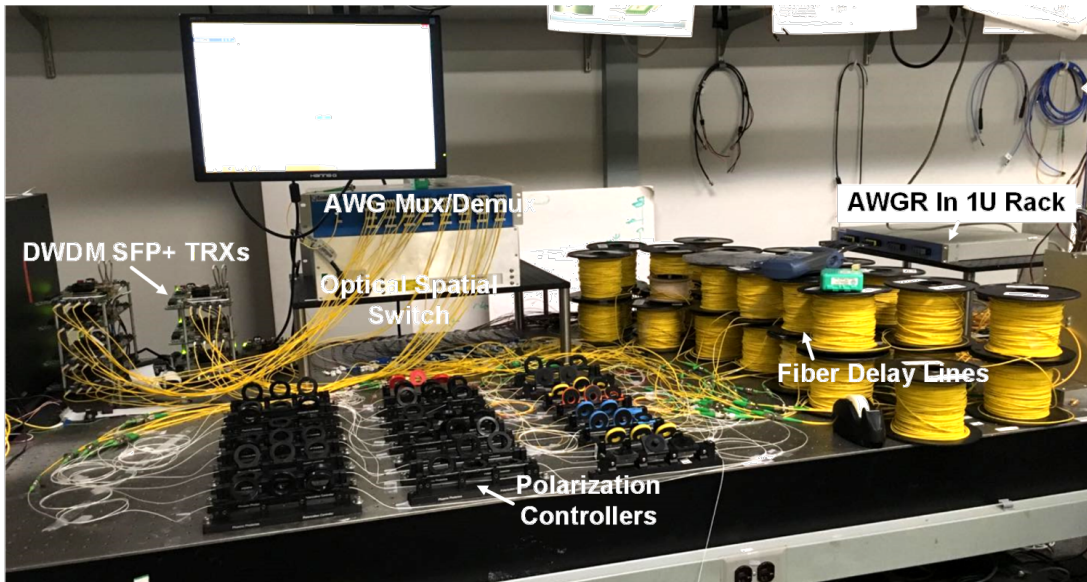
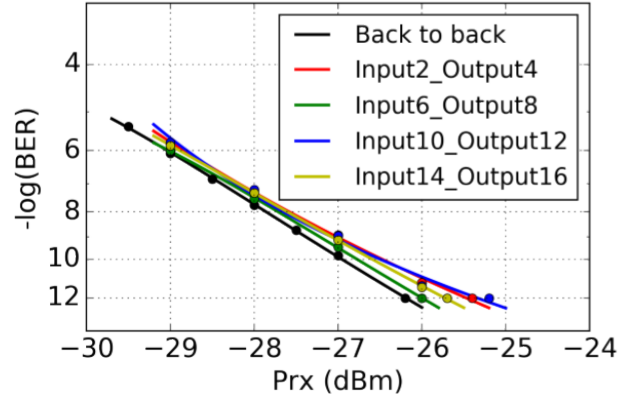
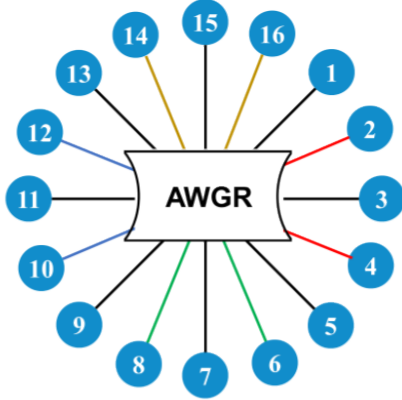


Figure 3.5. Picture of the testbed of 16-node Flex-LIONS.

transmitter and receiver).

By switching the left  $32 \times 32$  switch, all the 16 wavelengths from nodes 2, 6, 10, and 14 can be rerouted to bypass the AWGR and directly connect to nodes 4, 8, 12, and 16, respectively, as shown in Figure 3.6(b). The BER curves indicates error-free transmission for all the 16 wavelengths so that the total transmission bandwidth  $B$  is increased by a factor of 16.

(a) State 1: fully all-to-all



(b) State 2: partially all-to-all

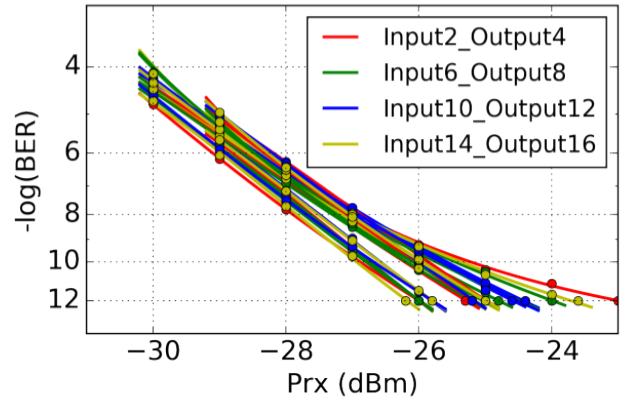
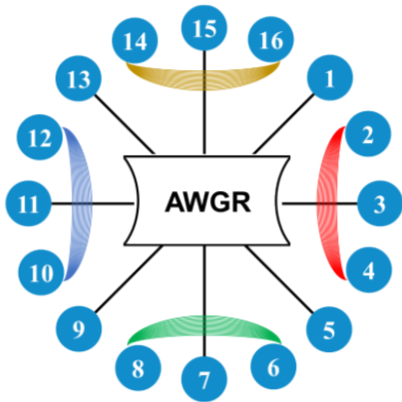


Figure 3.6. Topology diagram and BER curves of Flex-LIONS: (a) State 1, fully all-to-all interconnects ( $1\times$  bandwidth for all interconnected nodes); (b) State 2, partially all-to-all interconnects ( $16\times$  bandwidth for traffic between selected nodes).

### 3.1.3 Scalability study: $32\times 32$ all-to-all interconnect experiment

The intra-band crosstalk is the main impairment affecting the scalability of our proposed Flex-LIONS since it cannot be removed by the demultiplexer at the receiver side. For the worst case, the Flex-LIONS is in fully all-to-all scenario (State 1) and each output signals will have  $N-1$  intra-band crosstalk components. To verify the scalability of Flex-LIONS to 32 ports, we demonstrated 32-node fully all-to-all interconnects using 32 SFP+ modules and  $32\times 32$  100-GHz AWGR with average per-port crosstalk contribution as -40 dB.

Since 32 wavelengths from 32 nodes are fully loaded to the AWGR inputs, there are 1024 simultaneous links for a total aggregated bandwidth of 10.24 Tb/s. Figure 3.7(a)

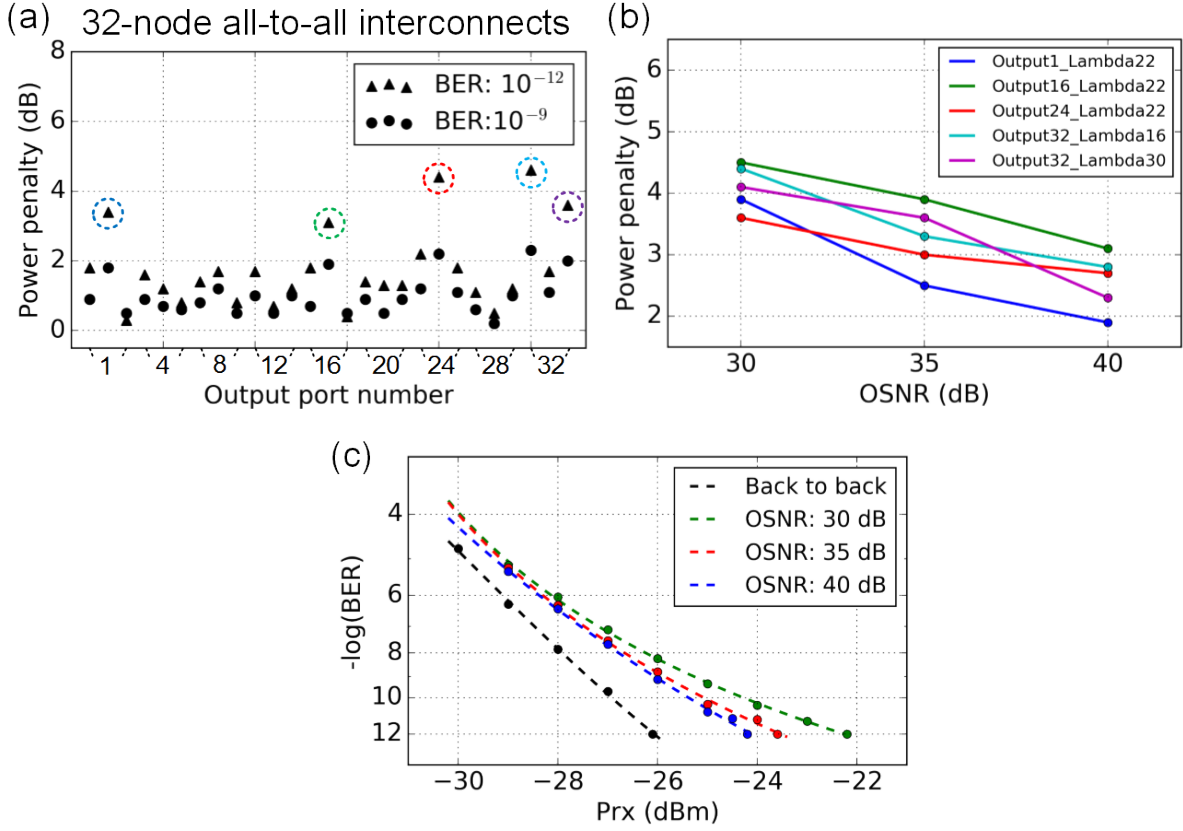


Figure 3.7. (a) Power penalty of 32-node all-to-all interconnects under worst-case crosstalk condition. (b) Power penalty at  $10^{-12}$  under different OSNR value. (c) BER curves with different OSNR value for output port 1 at wavelength channel 22 (1552.52 nm).

shows the power penalty at BER of  $10^{-9}$  and  $10^{-12}$  (error-free) for output ports 1, 4, 8, 12, 16, 20, 24, 28, and 32 at wavelength channels 16, 22, and 30 (1547.72 nm, 1552.52 nm, and 1558.98 nm) for the worst-case crosstalk condition. Most of the power penalty values are lower than 2 dB, as also expected from the theoretical results [27, 31]. The circled points have higher power penalty since the optical signal-to-noise ratio (OSNR) is degraded to 30 dB by the EDFAs [55], while other points are with OSNR of  $> 35$  dB. Figure 3.7(b) shows that the power penalty for these five points increases by 1-2 dB at  $10^{-12}$  BER when the OSNR rises from 30 dB to 40 dB. Figure 3.7(c) shows the BER curves with different OSNR values for output port 1 at wavelength channel 22 (1552.52 nm). As mentioned in previous section, EDFAs are not necessary in an actual deployment so the power penalty degradation caused by the EDFAs will not be present.



## 3.2 Silicon Photonic Flex-LIONS for Bandwidth Reconfigurable Optical Interconnects

In the past few years, there has been significant attention to the application and development of optical switching fabric for bandwidth reconfiguration between computing nodes or Top-of-Rack switches [52, 56]. At the physical layer, SiPh offers a variety of integrated devices with the capability of wavelength routing and space switching, thereby support dynamic configuration and reconfiguration in both spectral and spatial domains. Indeed, wavelength-and-space selective switching fabrics that can reconfigure the bandwidth between selected pair of input and output ports have been demonstrated with InGaAsP/InP AWGR + SOAs [14], SiPh echelle gratings + MEMS arrays [15], and SiPh multi-wavelength selective crossbar [16]. However, all these reported switching fabrics exhibit poor scalability and low energy efficiency due to either high insertion losses induced by power splitters [14],  $O(N^2)$  waveguide crossings in the worst-case path [15], or large number ( $O(N^3)$ ) of required switching elements [15, 16].

Here, we proposed a bandwidth-reconfigurable all-to-all interconnection switch, ‘Silicon Photonic Flexible Low-Latency Interconnect Optical Network Switch (SiPh Flex-LIONS),’ enabled by combining an AWGR-based all-to-all interconnection, MRR add-drop filters, and multi-wavelength spatial switches [57]. The multi-wavelength spatial switches which can be wide-band MEMS switches [10–12], wide-band Beneš Mach-Zehnder switch (MZS) networks [6, 7], or multi-wavelength MRR crossbar switches [33], are required to switch all the WDM signals simultaneously. The Flex-LIONS architecture has the lowest number of switching elements and insertion loss, enabling better scalability and energy efficiency when compared with other wavelength-and-space selective switching fabrics.

### 3.2.1 SiPh Flex-LIONS Architecture and Principle

Figure 3.8(a) illustrates the architecture of SiPh Flex-LIONS with multi-wavelength MRR crossbar. It has an  $N$ -port cyclic AWGR at the core and includes  $b$  MRR add-drop filters at each AWGR input/output port. For uniform-random traffic, all MRR add-drop filters

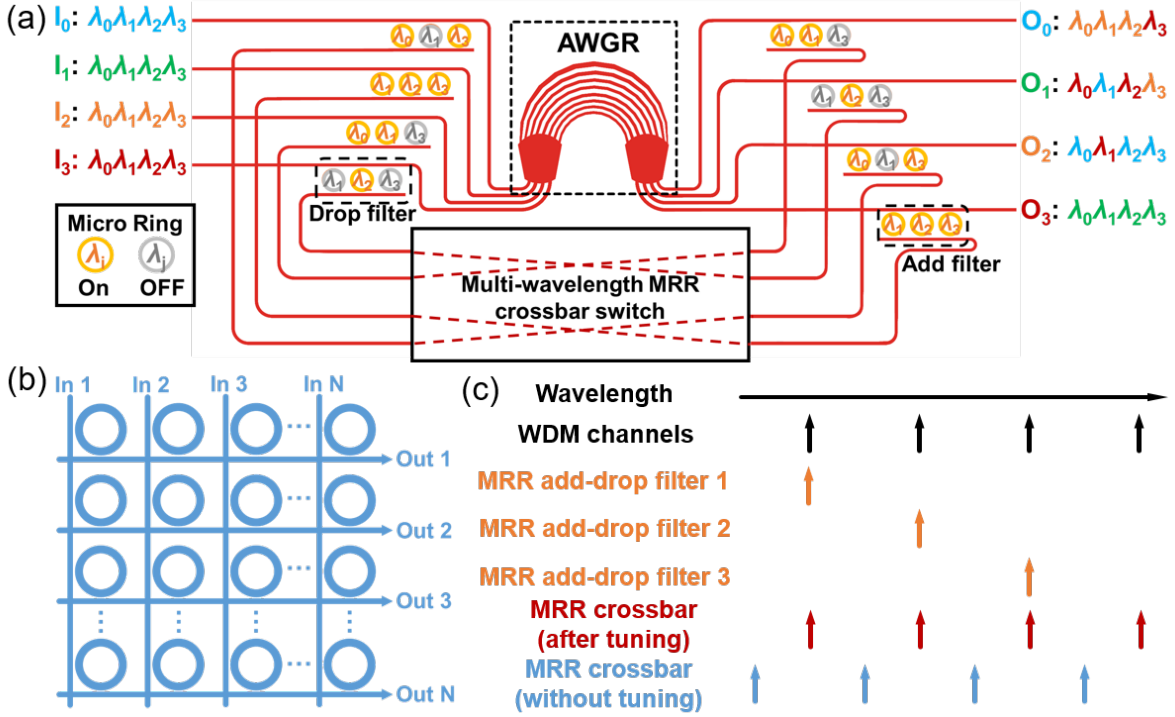


Figure 3.8. (a)  $N \times N$  Flex-LIONS architecture with  $N \times N$  AWGR,  $b$  MRR add-drop filters at each input and output ports, and  $N \times N$  multi-wavelength MRR crossbar switch. (b) Schematic of multi-wavelength MRR crossbar switch. (c) Schematic of the wavelength relation between the WDM channels and the resonances of MRR add-drop filters and multi-wavelength MRR crossbar switch.

can be set off-resonance so that each input port provides  $N$  WDM signals to interconnect with all the  $N$  output ports according to the all-to-all wavelength routing property of the AWGR [13, 28, 58, 59]. For non-uniform traffic or for resolving hot-spots, the MRR add-drop filters can be tuned in resonance to selected wavelengths channels so that those channels can be spatially switched by the multi-wavelength MRR crossbar switch shown in Figure 3.8(b). For instance,  $b$  of the  $N$  wavelengths from input port  $i$  can be dropped and then routed to the desired output port  $j$  by the  $N \times N$  multi-wavelength MRR crossbar switch. Here, the FSR of the multi-wavelength MRR is designed to match with the AWGR channel spacing (i.e., WDM channel spacing) so that all the  $b$  wavelength can be simultaneously routed by tuning the desired multi-wavelength MRR in the crossbar (Figure 3.8(c)). In this way, the bandwidth between input port  $i$  and output port  $j$  is effectively increased by a factor of  $b$ .

### 3.2.2 Comparison With the State-of-the-Art Reconfigurable Switching Fabrics

There has been a significant amount of architectural and experimental works on optical switching fabrics for HPC and datacenter systems, including semiconductor optical amplifier (SOA)-gate based optical switches [5], silicon photonic (SiPh) Mach-Zehnder interferometers (MZI)-based optical switches [6, 7], SiPh microring resonator (MRR)-based optical switches [8, 9], SiPh microelectromechanical systems (MEMS) switches [10–12], and SiPh arrayed waveguide grating router (AWGR)-based switches [13]. However, most of these switching fabrics only have single wavelength connectivity between I/O ports and the bandwidth cannot be steered to match the interconnections with specific application and traffic patterns.

On the other hand, multi-wavelength wavelength-and-space selective bandwidth reconfigurable switching fabrics exhibit better flexibility in interconnection patterns thanks to the ability to reconfigure the connectivity between I/O ports using any combination of the input wavelengths. Table 3.1 compares FlexLIONS with various state-of-the-art approaches including InP AWGRs + SOA gates [14], SiPh echelle gratings + MEMS arrays [15], and SiPh multi-wavelength selective crossbar [16]. In particular, the comparison study takes into account worst-case on-chip loss, crosstalk, footprint, and the number of switching elements (SOA gates, MEMS switches, MRRs). Here we assume all the wavelength channels can be reconfigured ( $b = N$  for Flex-LIONS).

To evaluate the scalability to high radix, we use the number of switching elements and the on-chip loss as a function of the number of ports ( $N$ ) as primary metrics. Typical switch architectures [14] have the problem of high on-chip insertion losses when the port number increases. Although the SOA gates can be used to compensate for such high losses, the low energy efficiency prevents [14] SOA-based switch architectures from scaling up to high radix. Ref. [15] architecture suffers not only from the high number of switching elements (scales as  $N^3$ ) but also from high on-chip insertion loss since the number of waveguide crossings increases by  $\sim N^2$ , while the number of waveguide crossings in FlexLIONS increases only by  $\sim N$ . Ref. [16] architecture also has the issue of a high number of

Architecture	Port Count	On-Chip Loss (dB)	Crosstalk (dB)	Footprint ( $mm^2$ )	Number of Switching Elements*	On-Chip Loss* (dB)
InP AWGRs+ SOA Gates [14]	4×4	23.7	-12	4.2×3.6	$2N^2$	$(N-1) \times 0.5 + \log_2 N \times 7 + 8.5$
Si Echelle Gratings+ MEMS Arrays [15]	8×8	16	Adj. ch.: -17 Non-adj. ch.: -30	9.7×6.7	$N^3$	$N \times 0.18 + N(N-1) \times 0.034 + 12.6$
Multi-Wavelength Selective MRR Crossbar [16]	8×4	14	-32**	1.92×4.15	$N^3$	$(N-1) \times 1.2 + 4.7$
Flex-LIONS with Multi-Wavelength MRR Crossbar [57]	8×8	6	Adj. ch.: -15 Non-adj. ch.: -28	10×5.6	$3N^2$	$(2N+5) \times 0.1 + (2N-2) \times 0.09 + 3.5$
Flex-LIONS with Beneš MZS Network [60]	8×8	8.4	Adj. ch.: -18 Non-adj. ch.: -28	10×4	$2N^2 + N \log_2 N - N/2$	$(N-1) \times 0.16 + N \times 0.5 + 4$

Table 3.1. Comparison among Flex-LIONS and the state-of-the-art bandwidth-reconfigurable switching fabrics. (\* For  $N \times N$  scale. \*\* By using dual MRR switches.)

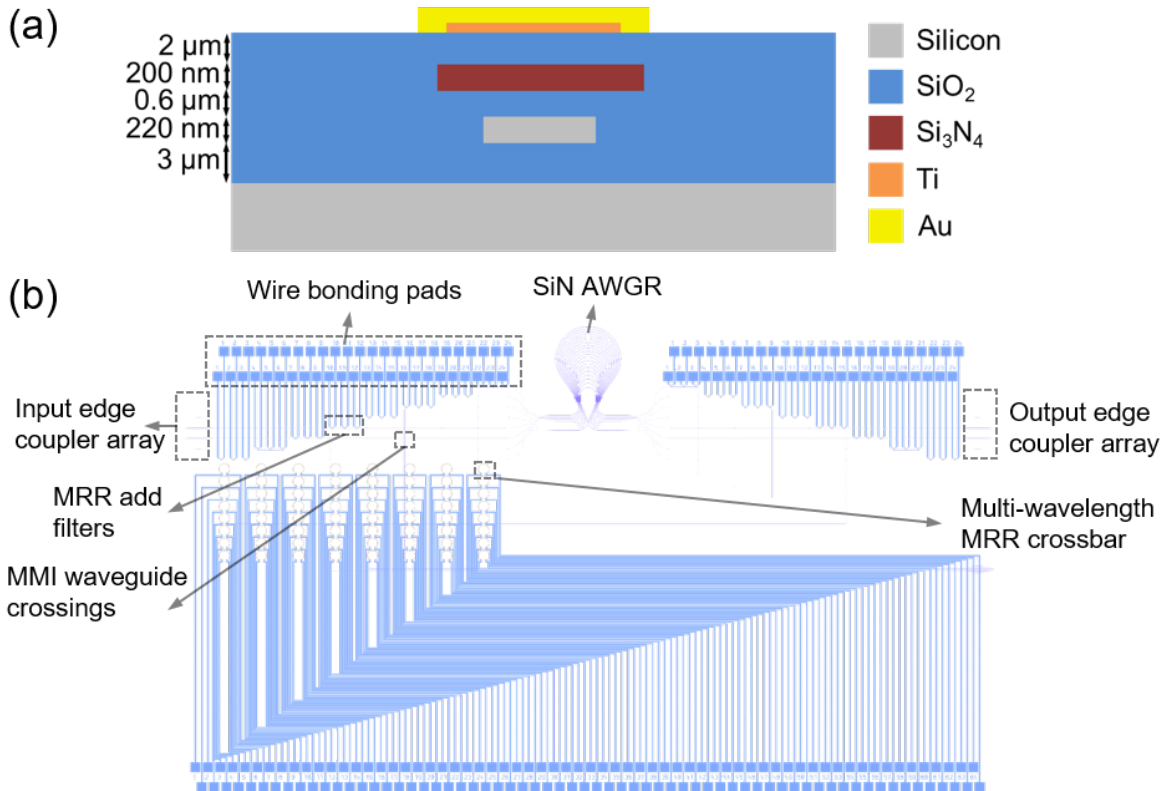


Figure 3.9. (a) Cross section of the multi-layer platform. (b) Schematic of the  $8 \times 8$  SiPh Flex-LIONS layout.

switching elements which limits the scalability. Taking  $N = 64$  for example, our proposed Flex-LIONS can save  $21\times$  in the number of switching elements compared with Ref. [15] and Ref. [16] architectures, with  $2.9\times$ ,  $5.7\times$ , and  $2.8\times$  lower on-chip losses compared to the switch architectures in Ref. [14], [15], and [16], respectively.

Compared to Flex-LIONS with multi-wavelength MRR crossbar [57], the number of cascaded MRRs on the path of the reconfigured channels in Flex-LIONS with Beneš MZS network [60] is reduced from three to two so that the bandwidth-narrowing effect is reduced. Besides, Flex-LIONS with Beneš MZS network has lower architectural complexity since the number of switching elements of the Beneš MZS network is  $N\log_2 N - N/2$  while that of the multi-wavelength MRR crossbar is  $N^2$ .

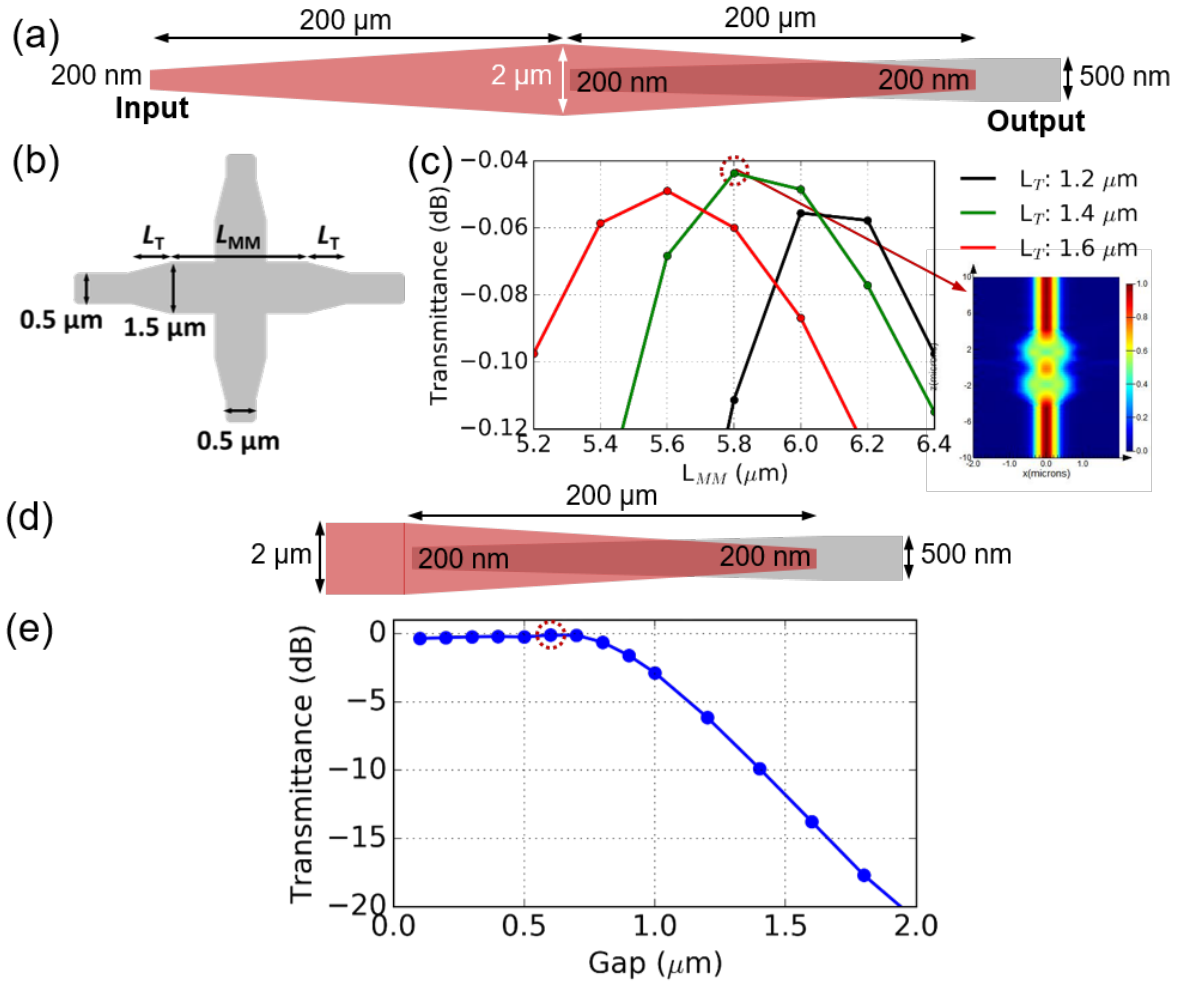


Figure 3.10. (a) Schematic of edge coupler design. (b) Schematic of MMI waveguide crossing design. (c) Simulated insertion loss of MMI waveguide crossing with various taper and multimode region lengths. (d) Schematic of Si-to-SiN evanescent coupler design. (e) Simulated insertion loss of evanescent coupler with various gap values.

### 3.2.3 Design, Fabrication, and Single Component Characterization of $8 \times 8$ Silicon Photonic Flex-LIONS

This section presents a detailed description of the design, fabrication, and single component characterization of the SiPh  $8 \times 8$  Flex-LIONS ( $N = 8$ ,  $b = 3$ ).

#### 3.2.3.1 Design

We designed our SiPh Flex-LIONS device on a multi-layer platform with silicon-dioxide cladding on silicon-on-insulator (SOI) wafers, as shown in Figure 3.9(a) [61]. The bottom layer is the silicon (Si) waveguide layer, which contains the MRR add-drop filters and

the multi-wavelength MRR crossbar switch. Ridge Si waveguides with 220-nm height and 500-nm width are used for low propagation loss. 600 nm above the Si waveguide layer is the SiN waveguide layer which contains the 200-GHz-spacing  $8 \times 8$  cyclic AWGR (the detailed design procedures can be found in [62]). Ridge SiN waveguides with 200-nm height and  $2\text{-}\mu\text{m}$  width are used for low propagation loss and relatively large bending radius. On top of the  $2\text{-}\mu\text{m}$ -thick silicon dioxide cladding are the 100-nm-thick titanium (Ti) heater layer and 800-nm-thick contact metal layer for thermo-optical (TO) tuning of the MRRs.

Figure 3.9(b) shows the schematic of the  $8 \times 8$  SiPh Flex-LIONS layout. Edge coupler arrays with  $127\text{-}\mu\text{m}$ -pitch are used to reduce the coupling loss from the fiber to the chip, as shown in Figure 3.10(a). The designed radii of the two MRRs are  $4.75\ \mu\text{m}$  and  $63\ \mu\text{m}$  corresponding to FSRs of 19 nm and 1.6 nm, respectively. The gap between the bus waveguides and the MRRs are fabrication-calibrated to be 300 nm and 450 nm to minimize the insertion loss for dropping.

SiPh low-loss and low-crosstalk multimode interference (MMI) crossings are essential components to keep the overall insertion loss low [63]. Figure 3.10(b) presents the physical dimensions of our crossing design. Figure 3.10(c) shows the FDTD simulations of insertion loss with various taper length ( $L_T$ ) and multimode region lengths ( $L_{MM}$ ). With the optimal design ( $L_T = 1.4\ \mu\text{m}$ ,  $L_{MM} = 5.8\ \mu\text{m}$ ), the simulated insertion loss is 0.04 dB.

The SiN AWGR vertically interfaces with the Si layer through inverse-tapered evanescent couplers [61, 64, 65]. As shown in Figure 3.10(d), the Si waveguide is tapered from 500 nm to 200 nm over a length of  $200\ \mu\text{m}$ , while the SiN waveguide is tapered from 200 nm to  $2\ \mu\text{m}$ . Figure 3.10(e) shows the FDTD simulation of inverse-tapered evanescent coupler transmission with a varied interlayer gap. The optimal gap value is 600 nm, with an insertion loss of 0.1 dB.

### 3.2.3.2 Fabrication

Utilizing micro and nanoscale fabrication facilities at the University of California at Davis and Berkeley, we fabricated the device on a 220-nm SOI wafer with  $3\text{-}\mu\text{m}$ -thick buried oxide, as shown in Figure 3.11(a). The silicon layer was defined by deep-UV projec-

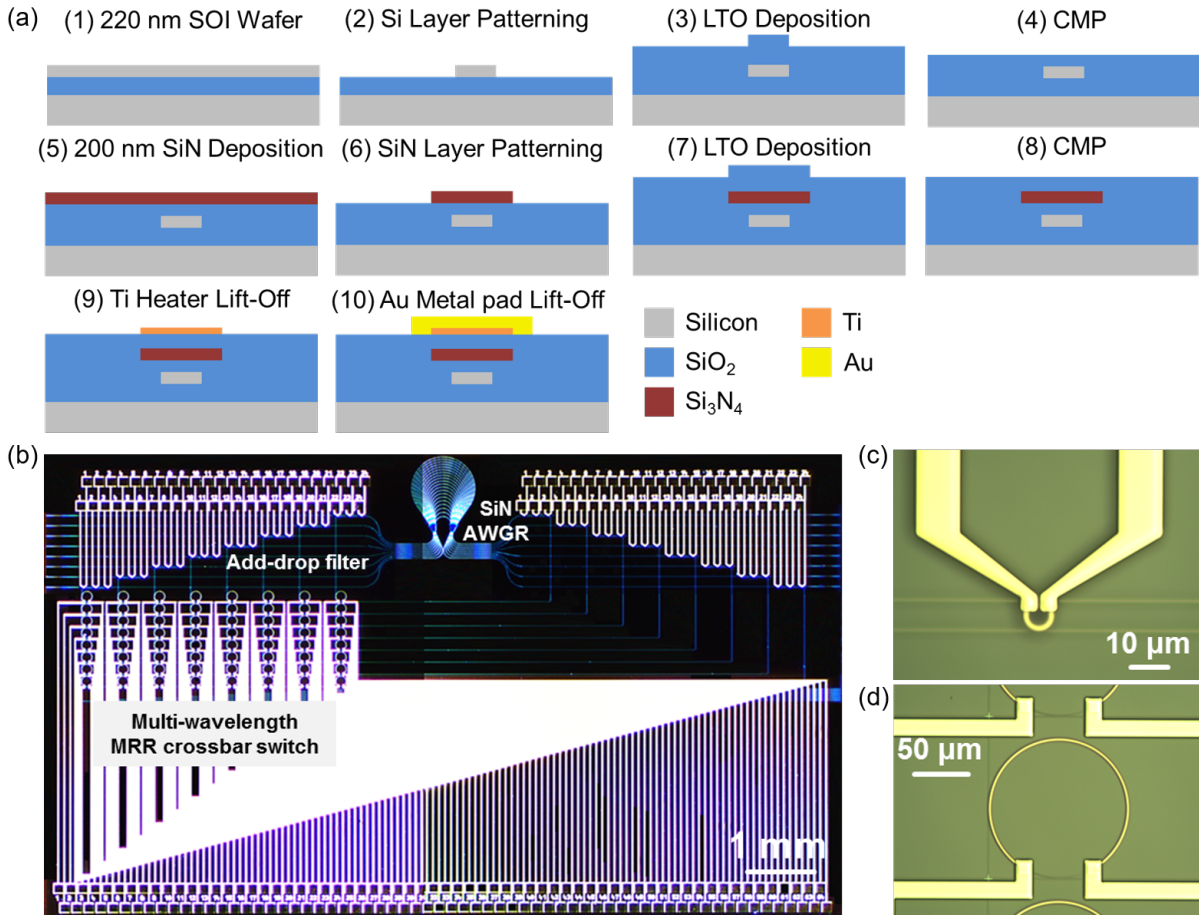


Figure 3.11. (a) Fabrication flowcharts for the  $8 \times 8$  SiPh Flex-LIONS. (b) Microscope image of the fabricated  $8 \times 8$  SiPh Flex-LIONS ( $N = 8$ ,  $b = 3$ ) chip. (c) Microscope image of MRR add-drop filter. (d) Microscope image of multi-wavelength MRR switch.

tion lithography and inductively coupled plasma (ICP) etching. Then a 1000-nm-thick low-temperature oxide (LTO) was deposited by low-pressure chemical vapor deposition (LPCVD) and then planarized to 800 nm by chemical mechanical planarization (CMP). Following the deposition of a 200-nm-thick SiN layer, the AWGR was patterned by deep-UV lithography and ICP etching, followed by a 2- $\mu\text{m}$ -thick LTO deposition and planarization. A 100-nm-thick Ti was then deposited on top of the cladding and along the MRR to act as a heater for TO tuning. Finally, a 20-nm-Ti and 800-nm-Au were deposited to form the contact metal layer. Figure 3.11(b-d) show the microscope images of the fabricated chip, MRR add-drop filter, and the multi-wavelength MRR switch.



### 3.2.3.3 Single Component Characterizations

Figure 3.12(a) shows the transmission spectra of the  $8 \times 8$  SiN AWGR from input port 4 measured by an optical vector network analyzer (OVNA) system. The AWGR is cyclic with an FSR of 12.8 nm, channel spacing of 1.6 nm (200 GHz) and full-width-at-half-maximum (FWHM) of 1.2 nm. The adjacent channel crosstalk is  $<15$  dB, the non-adjacent channel crosstalk is  $<28$  dB, and the insertion loss is  $<5.1$  dB. Figure 3.12(b) shows the transmission spectra of the through and drop ports of multi-wavelength MRR switch with different TO tuning power values. The insertion loss for the drop port at the resonance is 1 dB, and the corresponding FWHM is 0.24 nm. All the spectra are normalized to the reference waveguide. The TO tuning efficiencies of multi-wavelength MRR switch and MRR add-drop filter are 0.03 nm/mW and 0.15 nm/mW, respectively, as shown in Figure 3.12(c) and (d). Higher TO tuning efficiency can be achieved by using waveguide microheaters, and faster reconfiguration can be obtained by electro-optical (EO) tuning.

### 3.2.4 Experimental Demonstration of Optical Reconfiguration

Figure 3.13 shows the experimental setup we used to demonstrate the optical reconfiguration capabilities of the SiPh Flex-LIONS. The light sources are eight tunable laser diodes (TLDs) which provide the 200-GHz-spacing WDM grid of the Flex-LIONS. All the WDM wavelengths are multiplexed, amplified by a booster EDFA, and modulated by a Mach-Zehnder (MZ) modulator at 25 Gb/s. The driven signals are  $2^{11} - 1$  PRBS signals generated by a high-speed DAC. The modulated WDM signals are coupled in/out the Flex-LIONS chip using lensed fibers. The output signal from the Flex-LIONS chip is then received by an optically pre-amplified receiver (RX). A real-time error analyzer (EA) performs BER measurements as a function of the RX input power, which is measured by the built-in optical power monitor of the VOA. The Flex-LIONS chip was wire-bonded on a printed circuit board (PCB) and driven by a multi-channel DAC controller. The driving signals were used to tune the MRR add-drop filters as well as the multi-wavelength MRR crossbar switch, responsible for switching all the wavelength dropped to the desired output port.

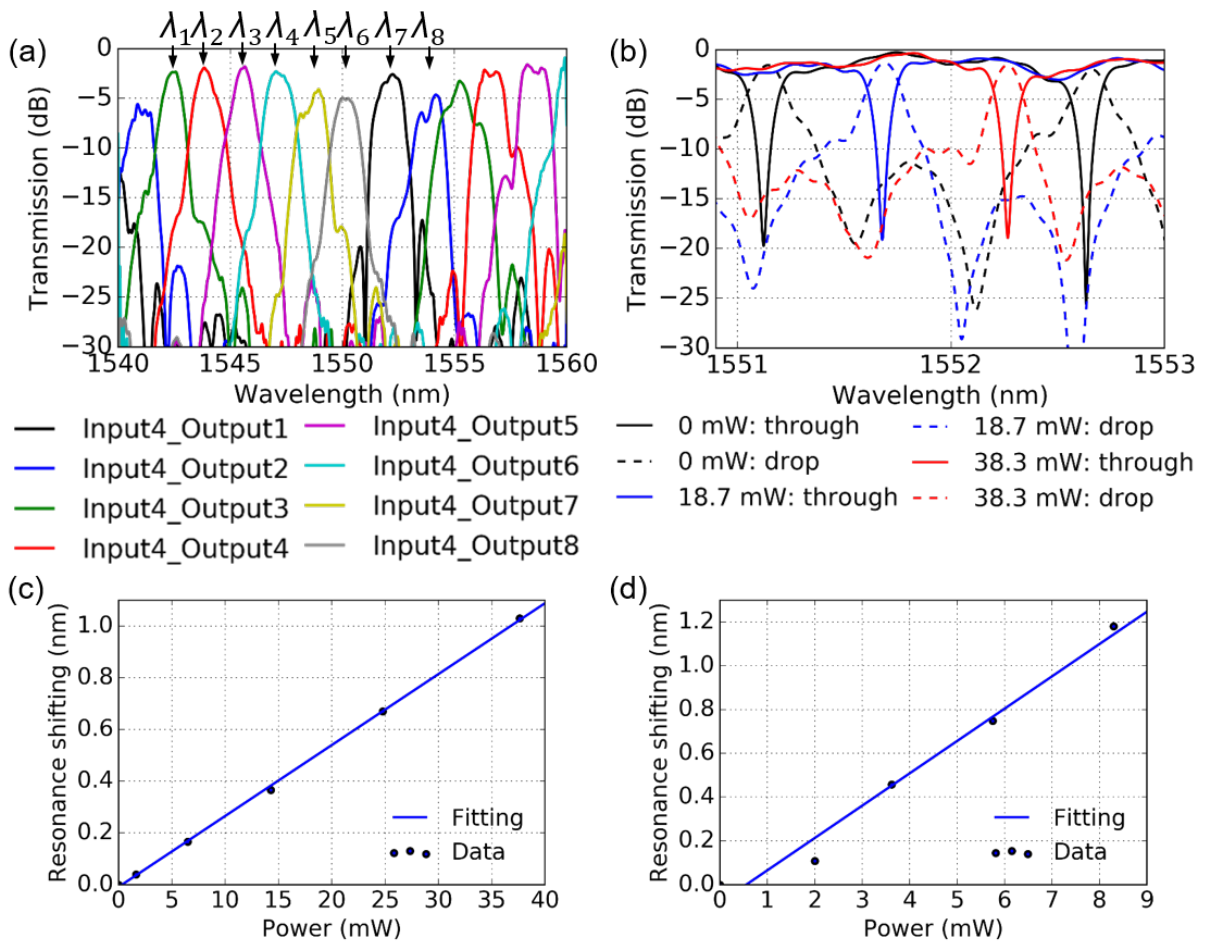


Figure 3.12. Transmission spectra of (a)  $8 \times 8$  AWGR from input port 4 and (b) through and drop ports of multi-wavelength MRR switch with different TO tuning power. Thermal tuning efficiency of (c) multi-wavelength MRR switch and (d) MRR add-drop filter.

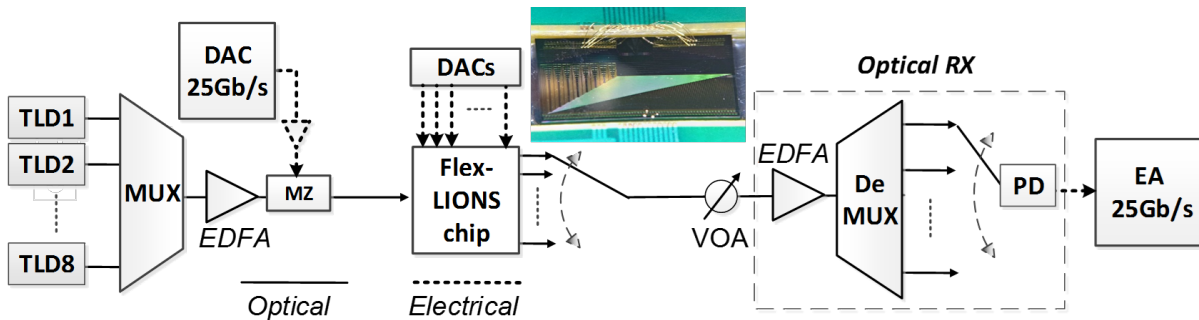


Figure 3.13. Experimental setup. TLD: tunable laser diode; EDFA: erbium-doped fiber amplifier; MZ: Mach Zehnder; DAC: digital to analog converter; DUT: device under test; VOA: variable optical attenuator; PD: photodetector; EA: error analyzer.

**Before reconfiguration (Input 4 to Output 5)**

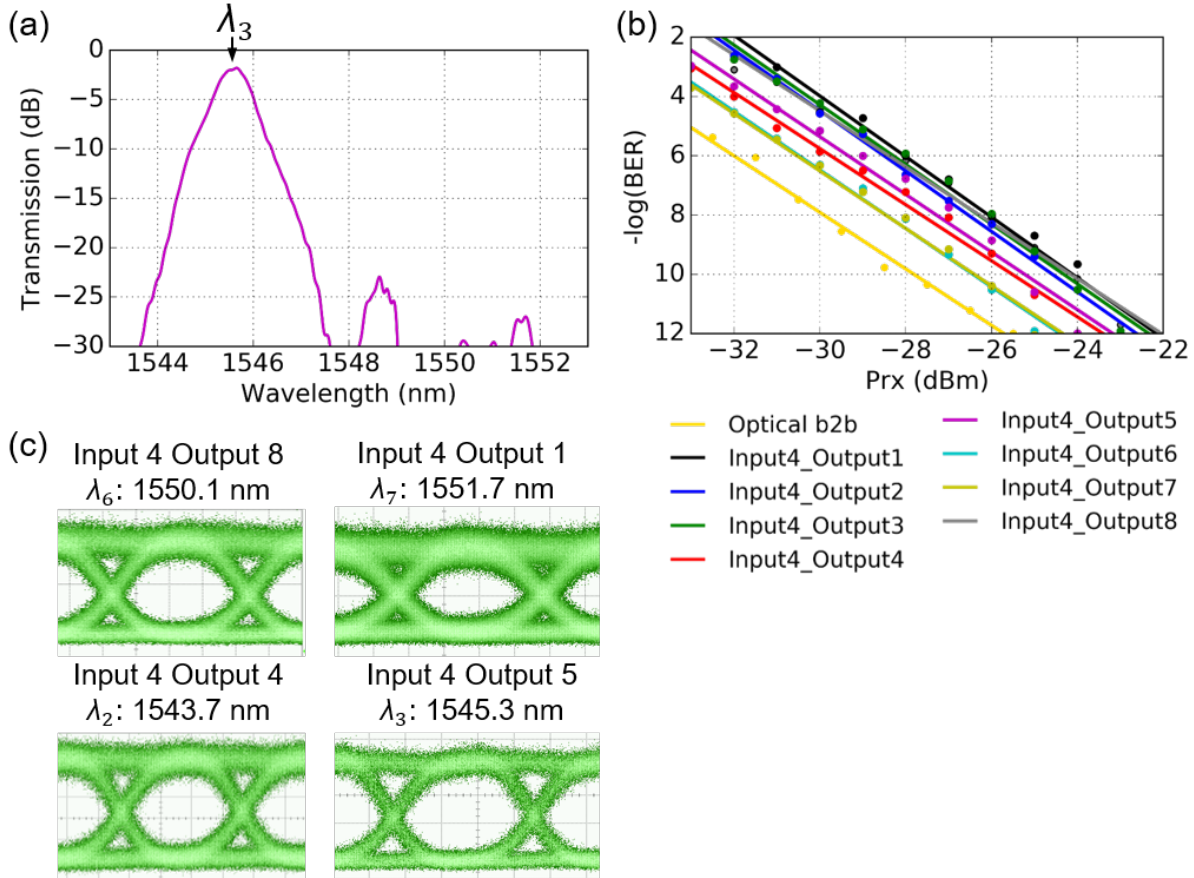


Figure 3.14. (a) Transmission spectrum from input port 4 to output port 5 before reconfiguration. (b) BER curves of input port 4 to different output ports. (c) Eye diagrams of input port 4 to output port 8, 1, 4, and 5 before reconfiguration.

Before reconfiguration, the device implements all-to-all connectivity, and the 8-channel WDM signal applied at the Flex-LIONS input 4 is demultiplexed to the eight Flex-LIONS output ports (one wavelength per port) according to the wavelength routing table of the AWGR. Figure 3.14(a) shows the bandwidth available between input 4 and outputs 5, which is single-channel ( $\lambda_3$ ) bandwidth of the AWGR. Figure 3.14(b) and (c) shows the measured BER curves and eye diagrams of the signals at the eight different output ports, demonstrating 25 Gb/s error-free operation with limited power penalty compared with the back-to-back BER curve. The total system capacity is  $25 \text{ Gb/s} \times 8 \times 8 = 1.6 \text{ Tb/s}$ .

After reconfiguration of the Flex-LIONS, four wavelengths from input port 4 are routed to output port 5. One wavelength is going through the AWGR while the other three

wavelengths are dropped, switched by the spatial switch and added to output port 5, effectively increasing the bandwidth between port 4 and port 5 by  $4\times$  (from 25 Gb/s to 100 Gb/s). Figure 3.15(a) shows that there are now four frequency slots available between input 4 and outputs 5. One of the four frequency slots is the passband of AWGR ( $\lambda_3$ ) while the other three ( $\lambda_5, \lambda_6, \lambda_7$ ) are from the cascaded MRR add-drop filters and multi-wavelength MRR crossbar switch. The maximum baud rate per wavelength channel is mainly limited by the compound cavity effect of cascaded MRRs for the signals dropped at AWGRs' inputs and going through the multi-wavelength MRR crossbar. To reduce the cascaded MRR filtering effect it could be possible to use wide-band Benes MZS networks as the multi-wavelength spatial switch to reduce the number of cascaded MRRs on the path of the reconfigured channels from three to two. Another method would be to employ flat-passband coupled MRRs [66] at the expense of incorporating more complicated tuning methods. While all the signals reach error-free condition (Figure 3.15(b) and (c)), the power penalty for one of the signal is significantly higher. We attribute this larger penalty to the frequency deviation between the WDM wavelengths and compound MRR cavity resonance.

### 3.2.5 Scalability of Flex-LIONS

Although our proposed Flex-LIONS architecture exhibits the least number of switching elements comparing with the state-of-the-art architectures, the scalability to larger port count (up to  $1024 \times 1024$  for example) is still limited by the power penalty induced by AWGR crosstalk. In this section, we calculate and experimentally measured the impact of intra-band crosstalk of AWGR on the scalability of Flex-LIONS, and discuss the use of Thin-CLOS Flex-LIONS architecture for port-count scaling.

#### 3.2.5.1 AWGR Crosstalk Power Penalty

Several factors limit the scalability of AWGRs, including insertion loss, loss non-uniformity, and channel spacing. Besides, the intra-band crosstalk is primary main impairment mechanism that affects the scalability of AWGRs since the signal-crosstalk beat noise cannot be removed by filters or de-multiplexers after the output ports [27].

Figure 3.16(a) shows the crosstalk power penalty of AWGR versus different crosstalk

### After reconfiguration (Input 4 to Output 5)

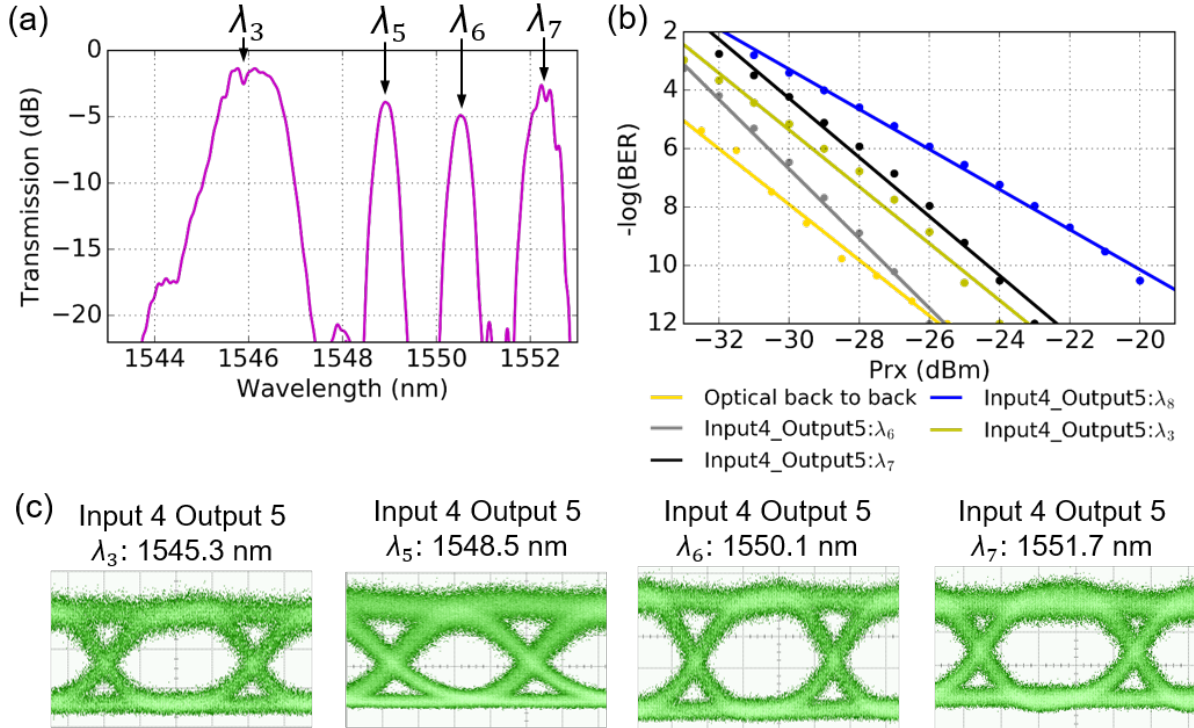


Figure 3.15. (a) Transmission spectrum from input port 4 to output port 5 after reconfiguration. (b) BER curves of input port 4 to output port 5. (c) Eye diagrams of input port 4 to output port 5 using  $\lambda_3$ ,  $\lambda_5$ ,  $\lambda_6$ , and  $\lambda_7$  after reconfiguration.

values with BER of  $10^{-12}$ . Here we assume optimized decision-threshold setting and aligned polarization for the worst case [31]. The results show that with 35 dB crosstalk the power penalty for 4-, 8-, 16-, and 32-node interconnects are 0.21, 0.50, 1.15, and 2.84 dB, respectively. For more clarity, the numbers of scalable nodes versus varied crosstalk values for different power penalty constraints are plotted in Figure 3.16(b). With 1-, 3-, and 6-dB power penalty constraint, the AWGR can scale to 32 nodes when the crosstalk is less than 38.7, 34.9, 33.1 dB, respectively.

The power penalty induced by AWGR intra-band crosstalk is experimentally measured using the  $8 \times 8$  SiPh Flex-LIONS chip. The Flex-LIONS chip is aligned and packaged with two 16-channel  $127\text{-}\mu\text{m}$ -pitch polarization-maintaining (PM) fiber arrays on both the input and output sides. The modulated 25 Gb/s/ $\lambda$  WDM signal is firstly split by a  $1 \times 8$  splitter. Then the eight signals are decorrelated by single-mode fiber catch

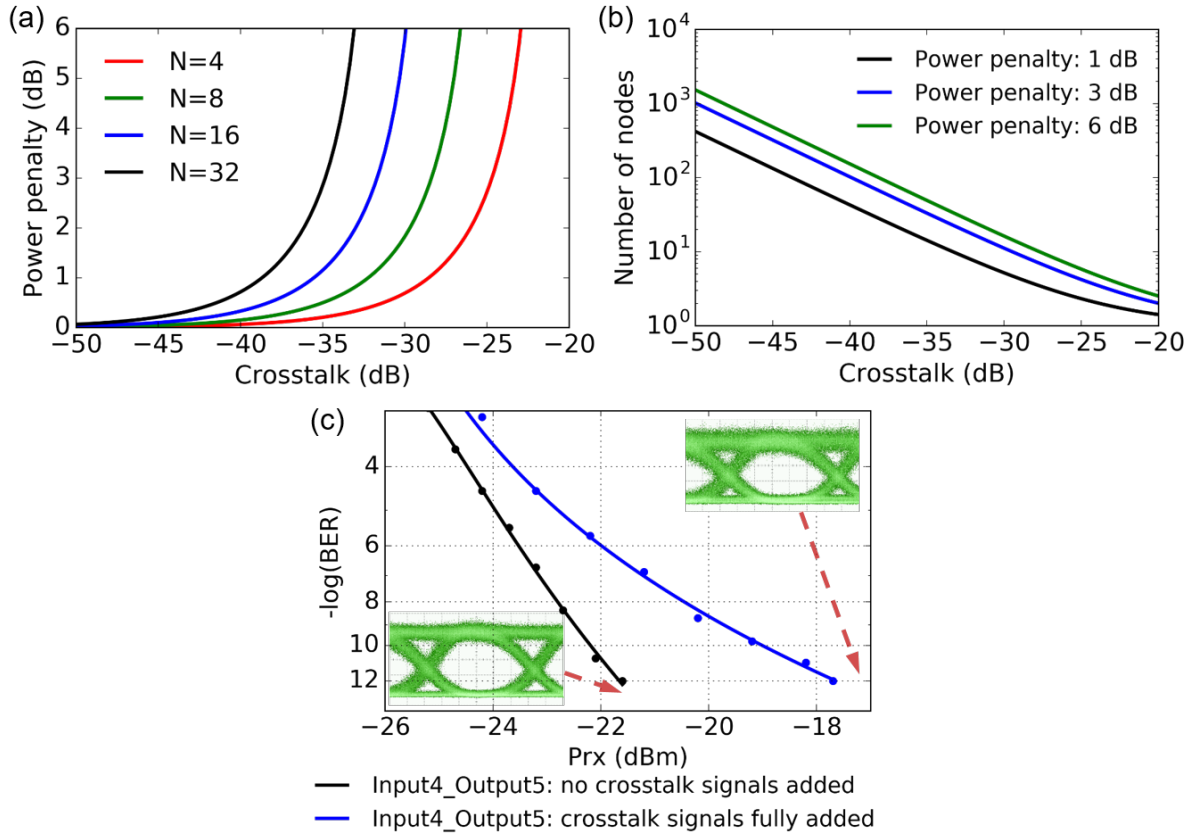


Figure 3.16. (a) Simulated power penalty versus varied crosstalk for AWGR. (b) Calculated number of scalable nodes versus varied crosstalk with power penalty constraint of 1, 3, 6 dB. (c) End-to-end link experiment with eight input signals at the same wavelength and aligned polarization for the worst-case crosstalk scenario.

cables with different lengths. Before input into all the eight input waveguides of the chip, the polarization of each signal is aligned by a polarization controller. Note that, in our future work, SiPh polarization splitters and rotators could be included to transform the polarization of the input signal from the single-mode fiber into fundamental TE mode [67] (this is important as datacom systems do not make use of PM fibers). The blue curve in Figure 3.16(c) shows the BER measurements of the signal going from input 4 to output 5 at  $\lambda_3$  for the worst-case crosstalk scenario (with all the input signals at  $\lambda_3$  are aligned in polarization). Comparing with no crosstalk signal added (black curve), the measured power penalty is 3.9 dB at  $\text{BER} = 10^{-12}$  which is slightly lower than the theoretically calculated value most likely due to the polarizations of the input signals not being perfectly aligned. The insets of Figure 3.16(c) show the eye diagrams of the

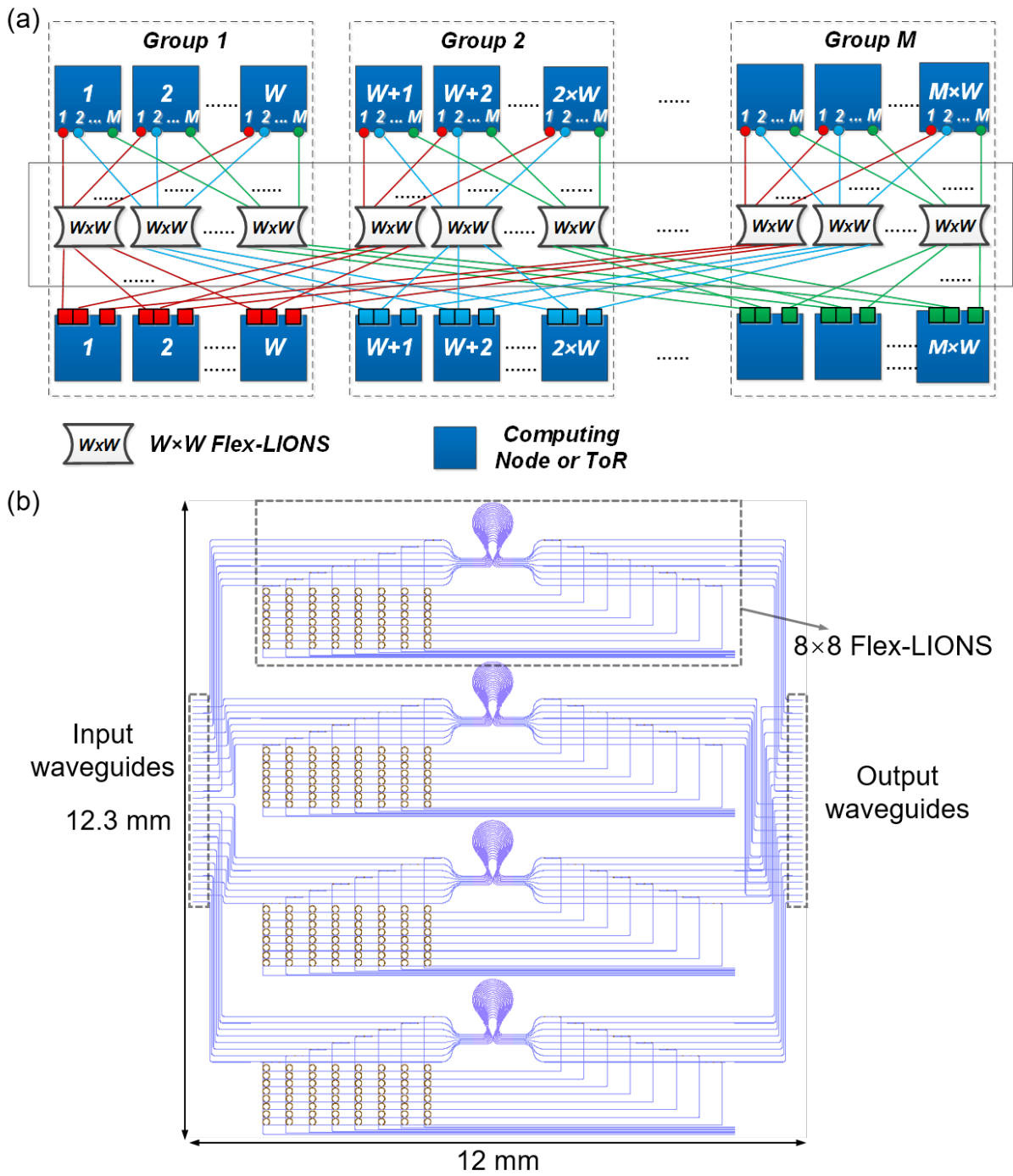


Figure 3.17. (a) Schematic of  $N \times N$  Thin-CLOS Flex-LIONS architecture ( $N = M \times W$ ). (b) Layout of  $16 \times 16$  Thin-CLOS Flex-LIONS with four  $8 \times 8$  Flex-LIONS ( $N = 16, M = 2, W = 8$ ).

transmitted signal with and without crosstalk signals added.

SiN AWGRs are superior to Si AWGR in mitigating crosstalk since the lower index

contrast makes SiN/SiO<sub>2</sub> waveguides less sensitive to fabrication imperfections. As a result, SiN AWGRs have smaller phase errors and consequently lower crosstalk [36]. In Ref. [29], 16 × 16 and 32 × 32 SiN AWGRs have been fabricated and characterized with crosstalk value of -10 dB. Alternatively, Ref. [38] reports SiN AWG with adjacent and non-adjacent crosstalk value as -39 dB and -33.5 dB by optimizing the deposition condition of cladding oxide and cross-section structure [38], which indicates the possibility to scale AWGRs beyond 32 × 32.

### 3.2.5.2 Thin-CLOS Flex-LIONS

As discussed in Section 2.4, Thin-CLOS Flex-LIONS is a promising architecture that can enable  $N \times N$  bandwidth-reconfigurable switching fabric using  $M^2$  number of  $W \times W$  Flex-LIONS instead of a single  $N \times N$  Flex-LIONS as shown in Figure 3.17(a) [68–70]. In this case, the number of intra-band crosstalk components is decreased from  $N - 1$  to  $W - 1$  so that the crosstalk power penalty can be significantly reduced. Other than that, smaller AWGRs also means lower insertion losses, loss non-uniformity, and larger channel spacing in a fixed spectral range. Figure 3.17(b) shows the schematic of the layout of a 16 × 16 SiPh Thin-CLOS Flex-LIONS with four 8 × 8 Flex-LIONS ( $N = 16$ ,  $M = 2$ ,  $W = 8$ ). The overall size is 12.0 mm × 12.3 mm. The waveguide crossings can be addressed by an additional SiN layer, and the chip can be flip-chip bonded on an optical interposer for the electrical fan-out [29, 44]. We believe such an approach paves the way to realizing large-scale Flex-LIONS with a limited number of wavelengths (e.g.,  $W = 64$ ).

## 3.3 Multi-FSR Silicon Photonic Flex-LIONS Module for Bandwidth-Reconfigurable All-to-All Optical Interconnects

One limitation of all the state-of-the-art bandwidth-reconfigurable switching fabrics, including the Flex-LIONS works in Section 3.2, is that the reconfigured bandwidth is ‘borrowed’ from the other optical links, negatively affecting the connectivity between the other nodes in the network [57, 71, 72]. This could lead to higher latency for the traffic between node pairs that are not part of the hotspot due to the additional number of hops required



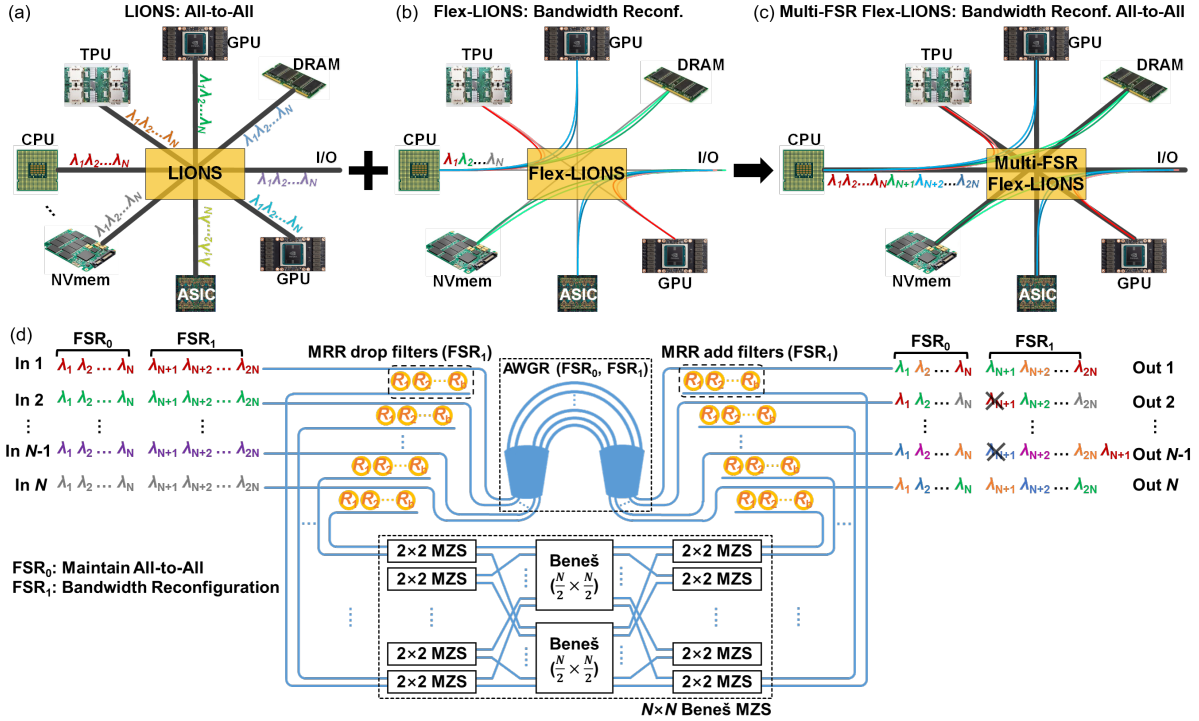


Figure 3.18. Heterogeneous processor and memory nodes with: (a) LIONS (all-to-all interconnects); (b) Single-FSR Flex-LIONS (bandwidth reconfigurable interconnects); (c) Multi-FSR Flex-LIONS (bandwidth-reconfigurable all-to-all interconnects). (d)  $N \times N$  multi-FSR Flex-LIONS architecture with  $N \times N$  AWGR,  $b$  MRR add-drop filters at each input and output ports, and  $N \times N$  Beneš MZS network. FSR<sub>0</sub> is used for maintaining all-to-all interconnectivity and FSR<sub>1</sub> is used for bandwidth reconfiguration.

to reach the destination nodes.

Here, we propose to leverage multiple free spectral ranges (FSRs) in a Flex-LIONS architecture to address the above-mentioned issues [60, 73]. The multi-FSR operation of AWGR is firstly proposed and demonstrated in [59]. Due to the cyclic nature of AWGRs, with the same device, the connectivity between each pair of nodes can be easily increased by exploiting multiple FSRs. Some FSRs of the core AWGR (e.g., FSR<sub>0</sub>) guarantees a minimum diameter all-to-all topology among the  $N$  interconnected nodes before and after reconfiguration as shown in Figure 3.18(a), while the other FSRs (e.g., FSR<sub>1</sub>) can be freely used to boost the bandwidth between specific node pairs as shown in Figure 3.18(b). In this case, bandwidth-reconfigurable all-to-all interconnects can be achieved as shown in Figure 3.18(c).

### 3.3.1 Principle of Multi-FSR Flex-LIONS

Figure 3.18(d) shows the architecture of the  $N \times N$  multi-FSR Flex-LIONS which contains an  $N \times N$  cyclic AWGR at the core,  $b$  MRR add/drop filters at the input/output ports of the AWGR ( $b < N$ ), and a broad-band  $N \times N$  Beneš MZS network [6, 7, 74] (rearrangeably non-blocking) at the bottom.  $2N$  WDM signals within two adjacent FSRs of the core AWGR (FSR<sub>0</sub> and FSR<sub>1</sub>) are loaded into each input port. For uniform-random traffic, both the FSR<sub>0</sub> ( $\lambda_1, \lambda_2, \dots, \lambda_N$ ) and FSR<sub>1</sub> ( $\lambda_{N+1}, \lambda_{N+2}, \dots, \lambda_{2N}$ ) are used for all-to-all interconnects based on the wavelength routing function of the AWGR so that the bandwidth between each pair of nodes is  $2B$  ( $B$  is the bandwidth carried by single wavelength) for Flex-LIONS using two FSRs. For resolving hotspots, up to  $b$  of the  $N$  wavelengths in FSR<sub>1</sub> from each input port can be dropped by the MRR drop filters and spatially switched to a selected output port by the Beneš MZS network so that the bandwidth between a specific node pair can be increased to  $(b + 2) \cdot B$ . Since FSR<sub>0</sub> is untouched, even if some pairs of nodes lose connectivity due to the reconfiguration in FSR<sub>1</sub>, they can still maintain a minimum bandwidth of single- $\lambda$  interconnections through FSR<sub>0</sub>. In other words, all-to-all connectivity based on one set of wavelengths will always be guaranteed. For example, assuming  $\lambda_{N+1}$  from input port 1 (which is initially used for interconnecting with output port 2) is reconfigured to output port  $N-1$ , both input port 1 to output port 2 and input port 4 to output port  $N-1$  will lose one wavelength ( $\lambda_{N+1}$ ) as shown in Figure 3.18(d). However, the connectivity between these two pairs of nodes is maintained by using  $\lambda_1$  in FSR<sub>0</sub>.

### 3.3.2 Design, Fabrication, and Packaging of $8 \times 8$ Silicon Photonic Flex-LIONS Module

#### 3.3.2.1 Design

The SiPh Flex-LIONS device is designed on a multi-layer platform on an SOI wafer as shown in Figure 3.19(a) [61]. The buried oxide of the SOI wafers is  $3\text{-}\mu\text{m}$  thick. The bottom 220-nm Si layer contains MRR add-drop filters and Beneš MZS network. Ridge Si waveguides with 500-nm width are used for single transverse electric (TE) mode transmission and low propagation loss. Above the Si layer is the 200-nm SiN waveguide

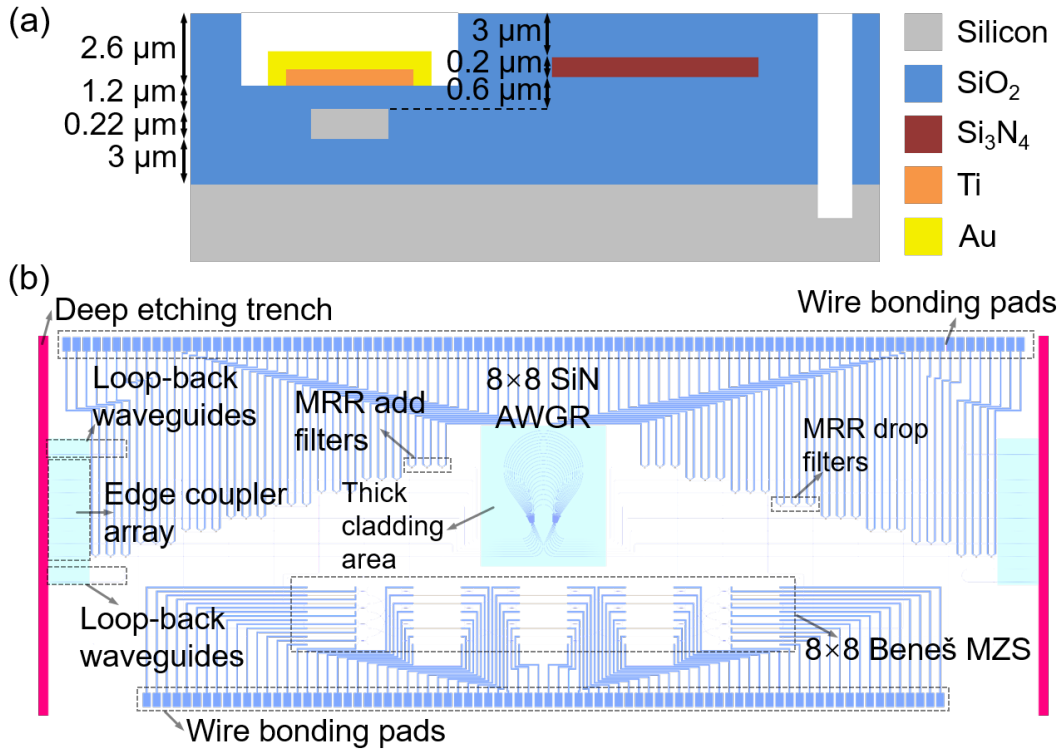


Figure 3.19. (a) Cross section of the multi-layer platform. (b)  $8 \times 8$  SiPh Flex-LIONS layout.

layer which contains the 200-GHz-spacing  $8 \times 8$  low-crosstalk SiN AWGR. The SiN layer vertically interfaces with the Si layer through inverse-tapered evanescent couplers with a 600-nm gap [61, 64]. Ridge SiN waveguides with  $2\text{-}\mu\text{m}$  width are used for low propagation loss and a relatively large bending radius. The silicon oxide cladding of the SiN layer is  $3\text{-}\mu\text{m}$  thick. An oxide cladding window is etched to  $1.2\ \mu\text{m}$  above the Si layer for higher thermo-optical (TO) tuning efficiency of the switching elements. On top of the oxide cladding are the 400-nm-thick Ti-heater layer and 800-nm-thick Au contact metal layer.

Figure 3.19(b) shows the  $8 \times 8$  SiPh Flex-LIONS chip layout. Edge coupler arrays with  $127\text{-}\mu\text{m}$ -pitch are used for low coupling loss from the fiber array to the chip. The edge coupler contains a SiN inverse taper from  $2\ \mu\text{m}$  to 200 nm and an evanescent coupler from the SiN layer to the Si layer. Two loop-back waveguides are placed on both sides of the edge coupler array for fiber array alignment. SiPh MMI waveguide crossings are designed to lower the overall insertion loss as shown in Figure 3.20(a). The detailed design

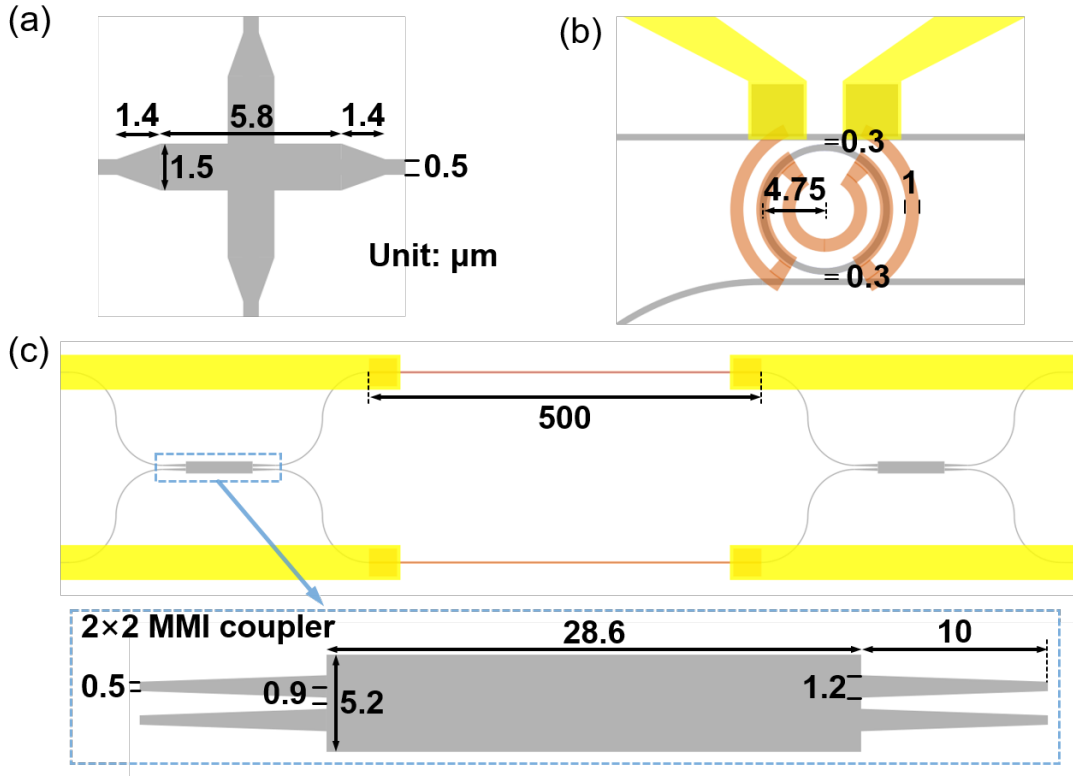


Figure 3.20. (a) Design of MMI based waveguide crossing. (b) Layout of MRR add-drop filter. (c) Layout of  $2 \times 2$  MZ switching element (arm length not to scale).

and simulation results of the MMI waveguide crossing can be found in Section 3.2.3.

The radius and the gap of the MRR add-drop filters are fabrication-calibrated to be 4.75  $\mu\text{m}$  and 0.3  $\mu\text{m}$ , respectively. Spiral resistive heaters along the MRR waveguide are designed to increase the TO tuning efficiency as shown in Figure 3.20(b). The width of the heaters is 1  $\mu\text{m}$ . Figure 3.20(c) shows the layout of  $2 \times 2$  MZS as the building block of the Beneš MZS network. The  $2 \times 2$  MZS contains two  $2 \times 2$  MMI couplers and two 500- $\mu\text{m}$ -long arms. The  $2 \times 2$  MMI couplers are designed and fabrication-calibrated for low insertion loss and high power balance [63]. The width and length of the  $2 \times 2$  MMI couplers are 5.2  $\mu\text{m}$  and 28.6  $\mu\text{m}$ , respectively. The center-to-center distance between the two access waveguides is optimized to be 1.8  $\mu\text{m}$ . The input and output waveguides are linearly tapered to 1.2  $\mu\text{m}$  in a length of 10  $\mu\text{m}$ . In order to achieve minimum TO tuning power, heaters are placed on both arms of the MZS. The width and length of the heaters are 1  $\mu\text{m}$  and 500  $\mu\text{m}$ , respectively.

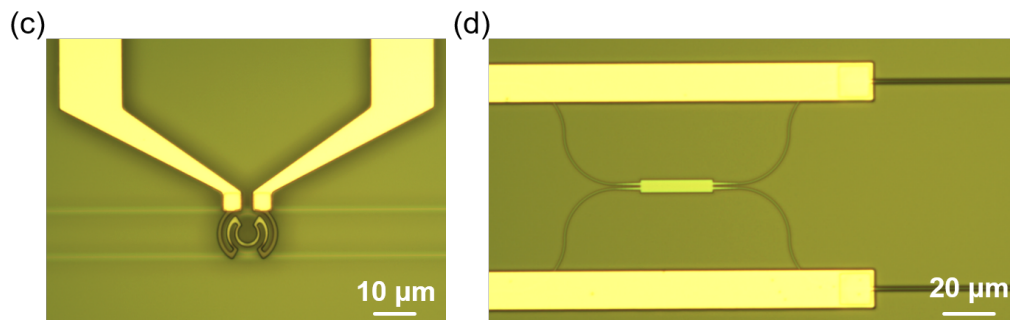
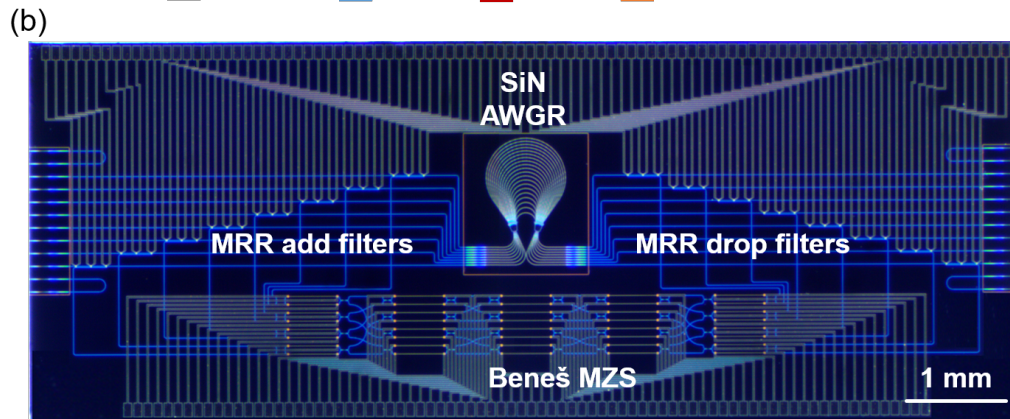
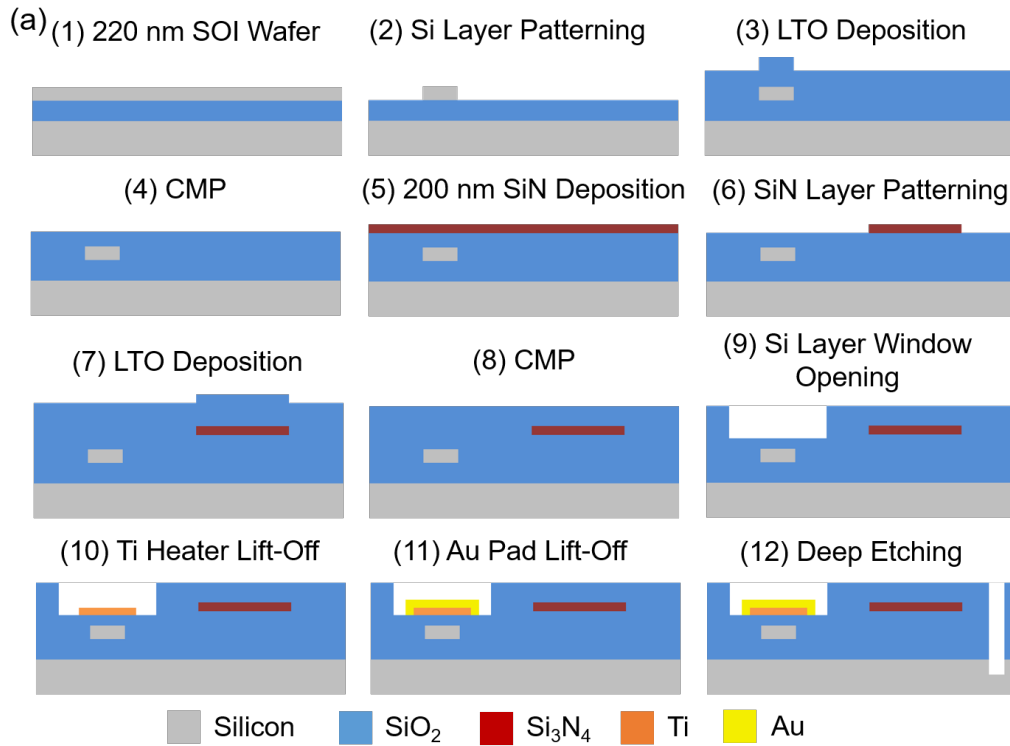


Figure 3.21. (a) Fabrication flow charts for the  $8 \times 8$  SiPh Flex-LIONS. (b) Microscope image of the fabricated  $8 \times 8$  SiPh Flex-LIONS ( $N = 8$ ,  $b = 3$ ) chip. (c) Microscope image of MRR add-drop filter. (d) Microscope image of part of  $2 \times 2$  MZS.

### 3.3.2.2 Fabrication

The Flex-LIONS chip was fabricated on a 220-nm SOI wafer with 3- $\mu\text{m}$ -thick buried oxide using the micro and nanoscale fabrication facilities at the University of California at Davis and Berkeley. Figure 3.21(a) shows the fabrication flow charts. Firstly, the Si layer is defined by deep-UV projection lithography and ICP etching. Then a 1000-nm LTO was deposited by LPCVD and then planarized to 800 nm by CMP. Following the deposition of a 200-nm SiN layer by LPCVD, the AWGR was patterned by deep-UV lithography and ICP etching. Then a 3- $\mu\text{m}$  LTO cladding was deposited and planarized. Subsequently, the oxide cladding window is opened by ICP etching. The 400-nm-thick Ti-heater layer and 800-nm-Au contact metal layer were then fabricated by E-beam evaporation and lift-off. Finally, a 140- $\mu\text{m}$  deep etching trench is fabricated using ICP etching. Figure 3.21(b-d) show the microscope images of the fabricated chip, MRR add-drop filter, and the  $2 \times 2$  MZS. The total chip size is 10 mm  $\times$  4 mm.

### 3.3.2.3 Packaging

The fabricated chip with 176 100- $\mu\text{m}$ -pitch electrical pads on the edge was wire-bonded to a co-designed PCB for electrical fan-out. Figure 3.22(a) shows the layout of the co-designed PCB with a total size of 120 mm  $\times$  50 mm. Two lid-less 16-channel 127- $\mu\text{m}$ -pitch polarization-maintaining (PM) fiber arrays were attached to the input and output of the chip using index-matching UV epoxy. Flexible flat cable (FFC) connectors are surface-mounted on the PCB for a compact footprint. The coupling loss from the PM fiber array to the chip after packaging is 4.75.7 dB/facet. Figure 3.22(b) shows the photograph of the integrated Flex-LIONS module.

## 3.3.3 Experimental Demonstration of Bandwidth-Reconfigurable All-to-All Optical Interconnects

This section presents the detailed characterization of the single switching elements and an experimental demonstration of bandwidth-reconfigurable all-to-all optical interconnects using the fabricated Flex-LIONS module and two FSRs.

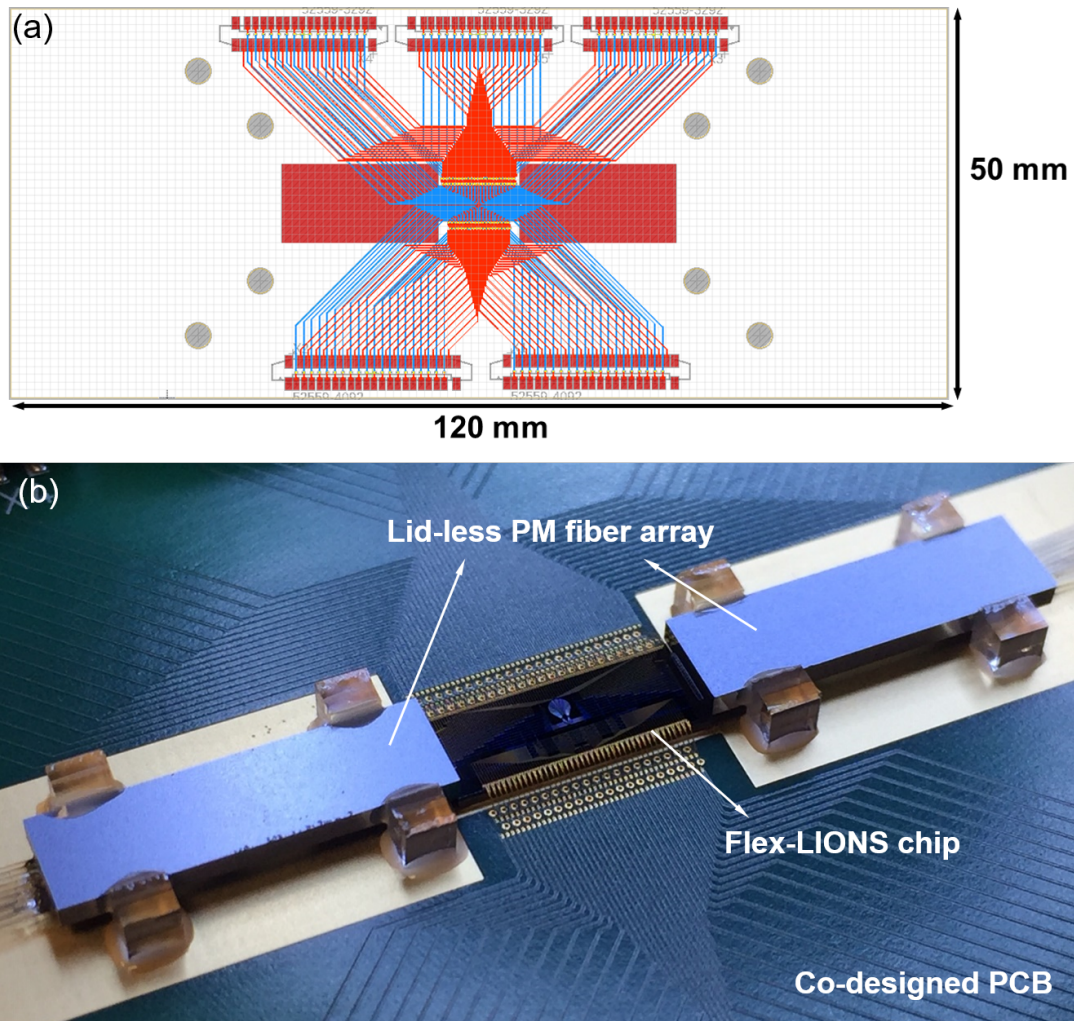


Figure 3.22. (a) Layout of the co-designed PCB. (b) Photograph of the integrated Flex-LIONS module with lid-less PM fiber arrays on a co-designed PCB. (Courtesy of Optelligent, LLC).

### 3.3.3.1 Single Elements Characterization

The transmission spectra of the  $8 \times 8$  SiN AWGR within two FSRs are measured by an optical vector network analyzer (OVNA) system as shown in Figure 3.23(a). The FSR, channel spacing, and full-width-at-half-maximum (FWHM) of the AWGR is 12.8 nm, 1.6 nm (200 GHz), and 1.07 nm respectively. The adjacent channel crosstalk is  $< -18$  dB, the non-adjacent channel crosstalk is  $< -28$  dB, and the insertion loss is  $< 3.5$  dB. The eight wavelength channels in  $\text{FSR}_1$  ( $\lambda_9, \lambda_{10} \dots, \lambda_{16}$ ) are for bandwidth reconfiguration while the eight wavelength channels in  $\text{FSR}_0$  ( $\lambda_1 = \lambda_9 - \text{FSR}, \lambda_2 = \lambda_{10} - \text{FSR} \dots, \lambda_8 = \lambda_{16} -$

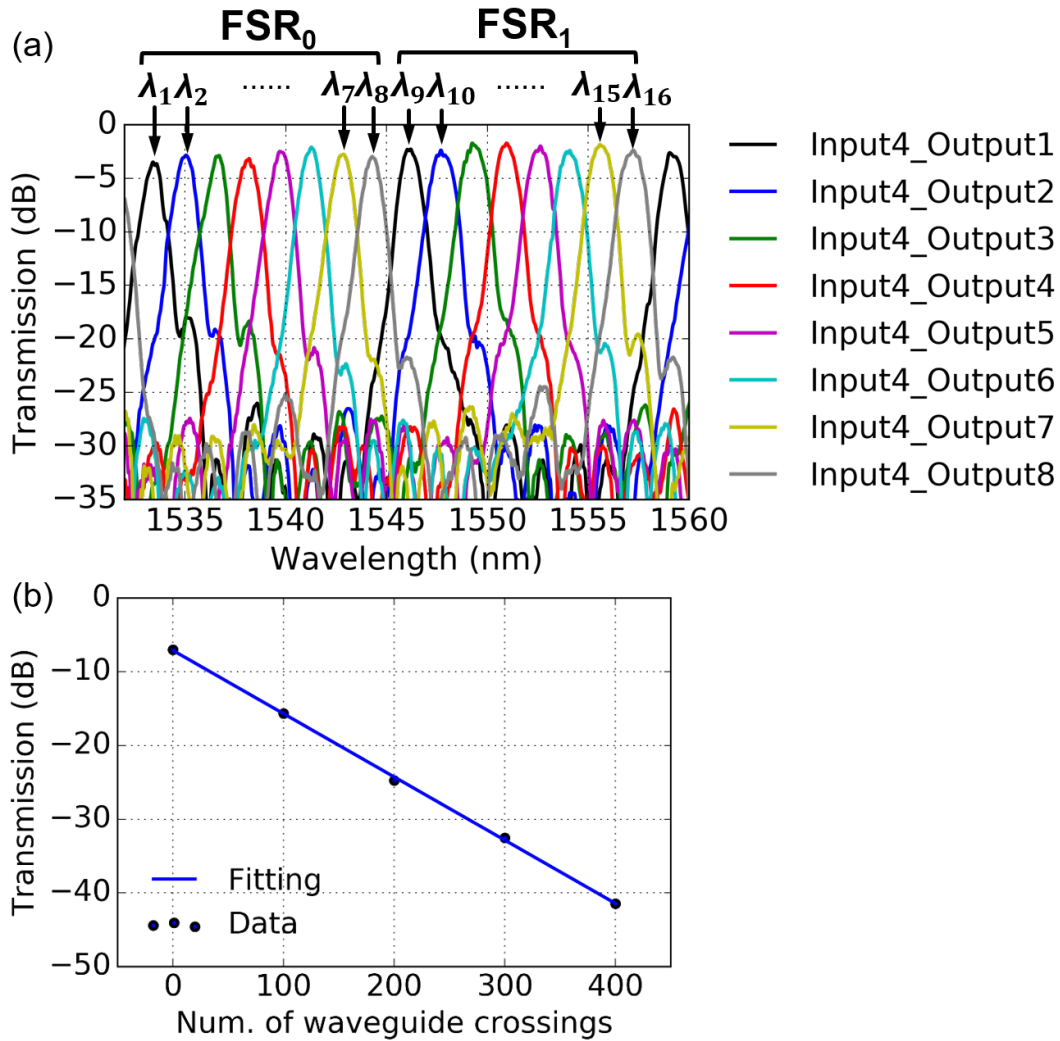


Figure 3.23. (a) Transmission spectra of  $8 \times 8$  SiN AWGR from input port 4. (b) Linear fitting of the normalized transmission of Si MMI waveguide crossing for insertion loss calculation.

FSR) are for maintaining basic all-to-all connectivity. All the wavelength channels match with the dense wavelength division multiplexing (DWDM) ITU grid. The insertion loss of the Si MMI waveguide crossing is measured as 0.08 dB through the linear fitting of the normalized transmission of four cascaded waveguide crossing structures as shown in Figure 3.23(b).

Figure 3.24(a) shows the transmission spectra of the through and drop ports of MRR add-drop filters with different TO tuning power. All the spectra are normalized to the reference waveguide. The insertion loss for the drop port, FWHM, and FSR are 1.4 dB,



0.71 nm, and 20.2 nm, respectively. Figure 3.24(b) shows the linear fitting of the resonance wavelength shifting with TO tuning power. The measured TO tuning efficiency of the MRR add-drop filter is 0.3 nm/mW (67 mW/FSR). Figure 3.24(c) shows the transmission spectra of MZS at the bar and cross port with different TO tuning power applied on the upper heater. An initial bias of 0.87 mW is required to achieve the cross state due to phase errors induced by fabrication imperfection. The insertion loss is 0.3 dB and the TO power to switch between cross and bar state is 16.5 mW. The crosstalk in the wavelength range of 20 nm is lower than -20 dB while the minimum crosstalk is lower than -40 dB.

### 3.3.3.2 Experimental Demonstration of Bandwidth Reconfiguration Using Two-FSR Flex-LIONS

Figure 3.25 shows the experimental setup for demonstrating bandwidth-reconfigurable all-to-all optical interconnects using the integrated SiPh Flex-LIONS module. Here, two-FSR Flex-LIONS is demonstrated so that  $\text{FSR}_1$  can be used for bandwidth steering while  $\text{FSR}_0$  maintains basic all-to-all connectivity after reconfiguration.

Sixteen DWDM SFP lasers provide the sixteen 200-GHz-spacing WDM signals ( $\lambda_1 = 1533.47$  nm,  $\lambda_2 = 1535.04$  nm ...,  $\lambda_{16} = 1557.36$  nm). All the WDM signals are multiplexed and modulated by a MZ modulator at 25 Gb/s. The electrical driving signals are  $2^{11} - 1$  PRBS signals generated by a high-speed digital to analog converter (DAC). Sixteen PCs before the multiplexer (MUX) and a polarizer before the MZ modulator are used for polarization alignment. The modulated signal is boosted by an erbium-doped fiber amplifier (EDFA) and then split by a  $1 \times 8$  splitter. The eight split signals are decorrelated by single-mode fiber catch cables with different lengths and aligned to the polarization of the PM fiber array by a PC before entering the Flex-LIONS module. The output signals from the chip are received by an optically pre-amplified receiver (RX). A real-time error analyzer (EA) performs BER measurements as a function of the RX input power, which is measured by the optical power monitor of the VOA.

Before bandwidth reconfiguration, both FSRs implement all-to-all optical interconnects based on AWGR's wavelength routing property so that the bandwidth between each pair of input and output ports is  $2\lambda \times 25 \text{ Gb/s}/\lambda = 50 \text{ Gb/s}$ . The total system

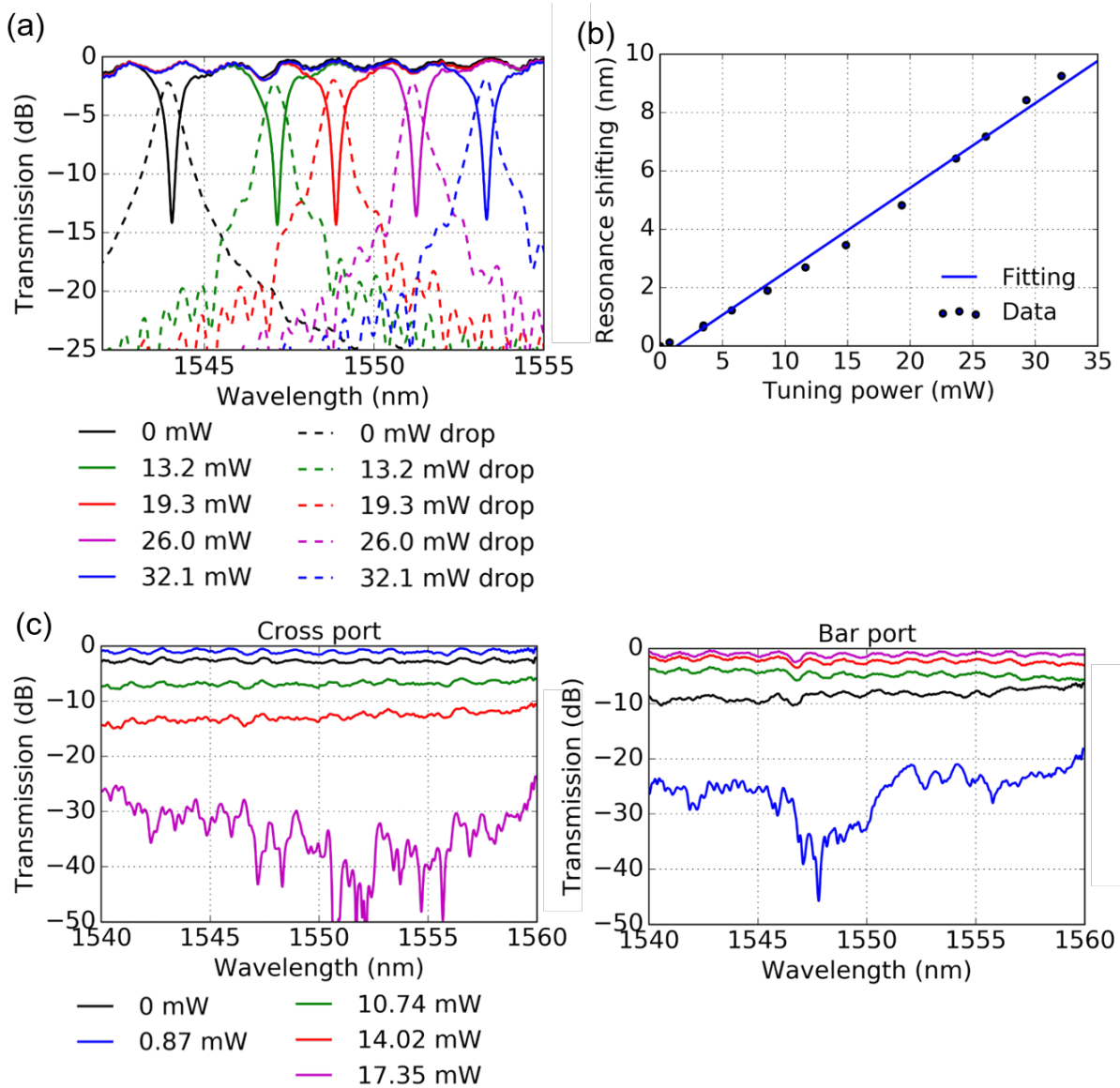


Figure 3.24. (a) Transmission spectra of through and drop ports of MRR add-drop filter with different TO tuning power. (b) TO tuning efficiency of MRR add-drop filter. (c) Transmission spectra of  $2 \times 2$  MZS at different TO tuning power for the cross port and the bar port.

capacity is  $25 \text{ Gb/s}/\lambda \times 16 \lambda \times 8 = 3.2 \text{ Tb/s}$ . Figure 3.26(a) shows the transmission spectrum from input port 4 to output port 8 with AWGR channel  $\lambda_8$  and  $\lambda_{16}$ . Figure 3.26(b) and (c) show the BER curves from center and side input ports through  $\text{FSR}_0$  and  $\text{FSR}_1$  which both demonstrates error-free all-to-all optical interconnects. Comparing with the back-to-back curve (no crosstalk signals added), the measured power penalty

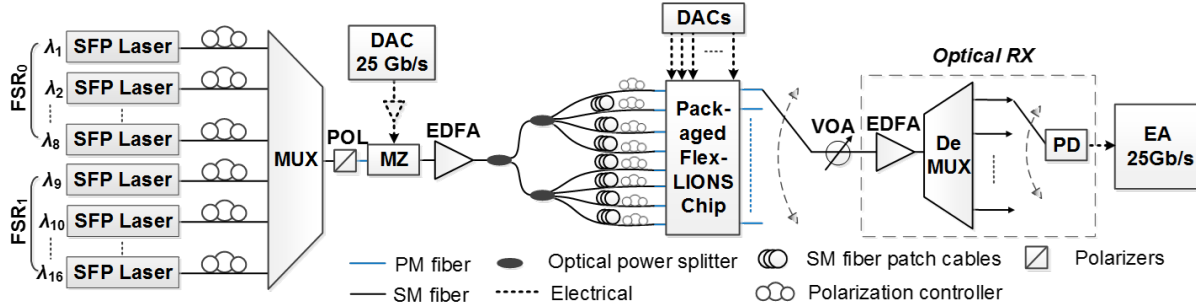


Figure 3.25. Experimental setup. SFP: small form pluggable; MUX: multiplexer; MZ: MachZehnder; EDFA: erbium-doped fiber amplifier; DAC: digital to analog converter; VOA: variable optical attenuator; DeMUX: demultiplexer; PD: photodetector; EA: error analyzer.

under the worst-case crosstalk scenario (aligned polarization for all the input signals) is in the range of 3.9 to 5.3 dB at  $\text{BER} = 10^{-12}$ . Such power penalty is mainly induced by the intra-band crosstalk of the AWGR since the crosstalk from cascaded MRR add-drop filters is a second-order crosstalk. The measured power penalty is slightly lower than the theoretically calculated value [27] due to the polarization of the input signals not being perfectly aligned. Lower crosstalk penalty can be achieved by optimized AWGR design and fabrication [38]. Figure 3.26(d) shows the eye diagrams for the back-to-back and selected input and output ports.

After bandwidth reconfiguration, three wavelengths in  $\text{FSR}_1$  from input port 4 ( $\lambda_{10}$ ,  $\lambda_{12}$ , and  $\lambda_{14}$ ) are dropped by the MRR add-drop filter and then routed to output port 8 by the Beneš MZS network. Together with two wavelength channels from the AWGR ( $\lambda_8$  and  $\lambda_{16}$ ), the total number of wavelengths channels from input port 4 to output port 8 is increased to 5 as shown in Figure 3.27(a). Note that, the dropping of any wavelength in  $\text{FSR}_1$  will not cause any unwanted wavelength drop in  $\text{FSR}_0$  since the FSR of the MRR add-drop filter is 12.6 times the channel spacing of the AWGR. The FWHM of AWGR channels ( $\lambda_8$  and  $\lambda_{16}$ ) are 1.05 nm and the FWHM of reconfigured channels ( $\lambda_{10}$ ,  $\lambda_{12}$ , and  $\lambda_{14}$ ) are narrower (0.42 nm) due to the filtering effect of two cascaded MRR add-drop filters. The insertion loss of the reconfigured channels is  $< 8.4$  dB which consists of: 2.8 dB ( $2 \times 1.4$  dB) from the drop loss of the MRR add-drop filters, 0.32 dB ( $4 \times 0.08$  dB) from the insertion loss of the MMI waveguide crossings, 4.1 dB from the Beneš

## Before reconfiguration

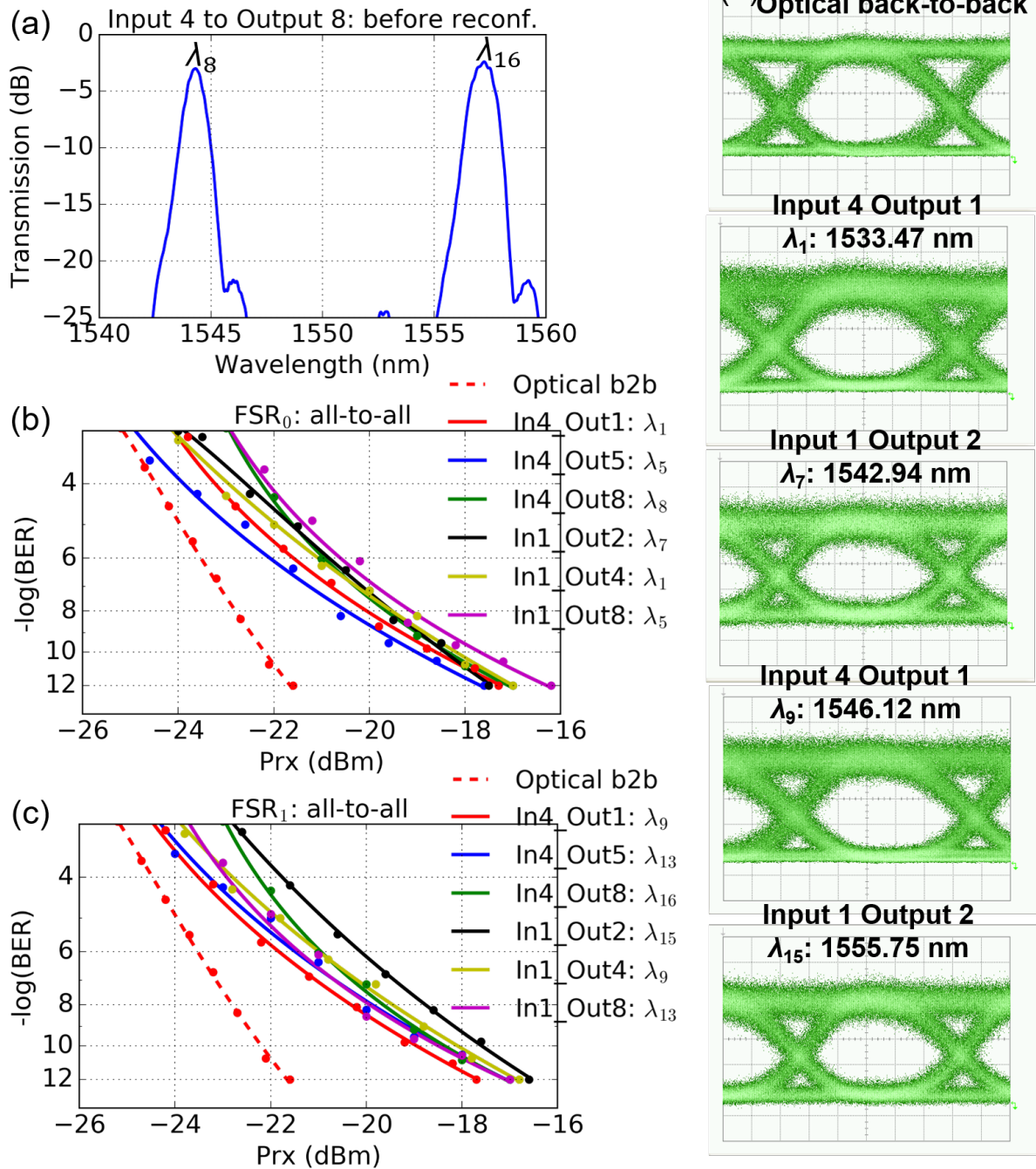


Figure 3.26. (a) Transmission spectrum from input port 4 to output port 8 before reconfiguration. (b) BER curves of all-to-all interconnects through  $FSR_0$  before reconfiguration. (c) BER curves of all-to-all interconnects through  $FSR_1$  before reconfiguration. (d) 25 Gb/s eye diagrams for back-to-back and selected input and output ports.

## After reconfiguration

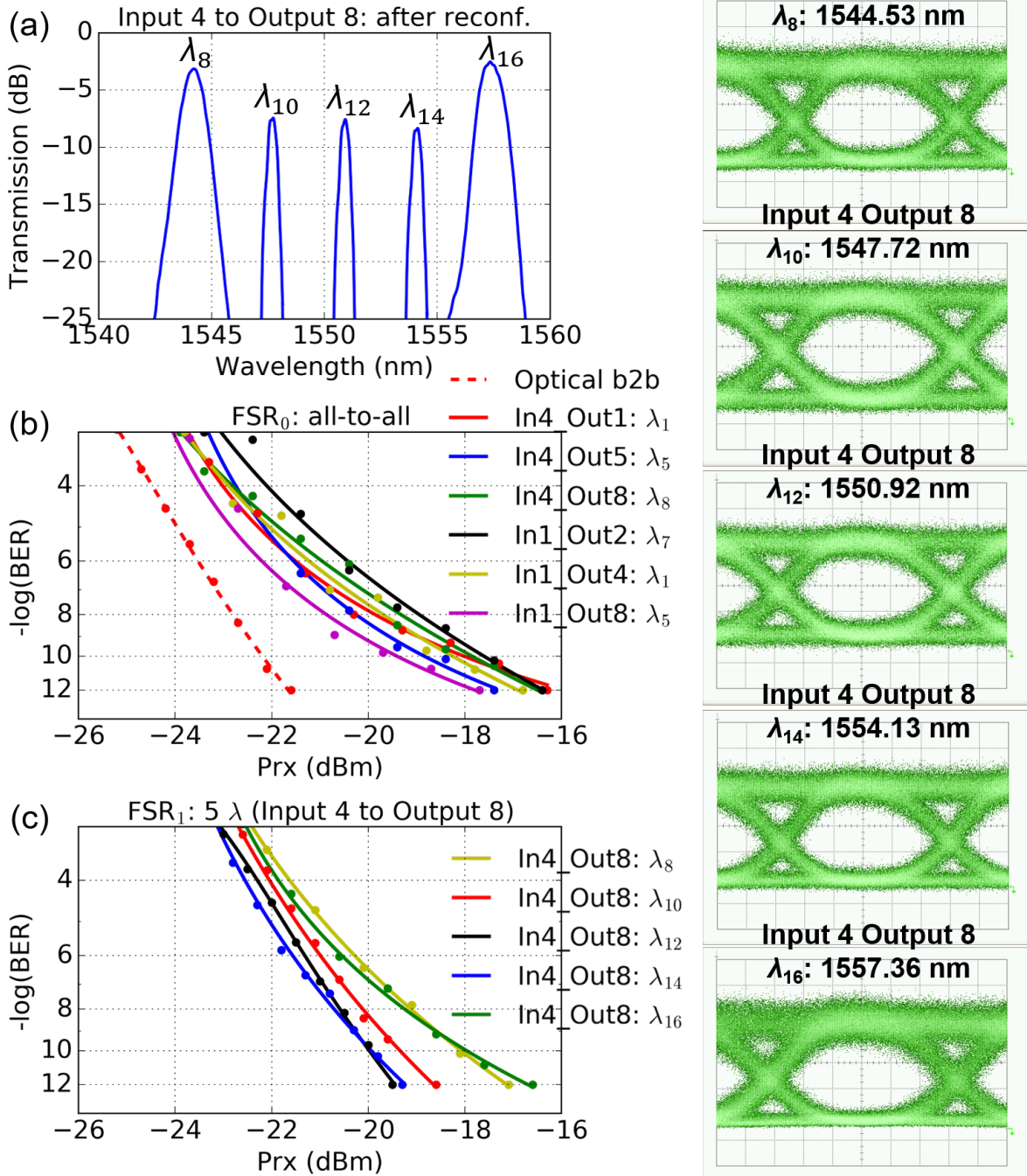


Figure 3.27. (a) Transmission spectrum from input port 4 to output port 8 after reconfiguration. (b) BER curves of all-to-all interconnects through  $FSR_0$  after reconfiguration. (c) BER curves of input port 4 to output port 8 after reconfiguration ( $\lambda_8$  in  $FSR_0$ ,  $\lambda_{10}$ ,  $\lambda_{12}$ ,  $\lambda_{14}$ ,  $\lambda_{16}$  in  $FSR_1$ ). (d) Eye diagrams of input port 4 to output port 8 using  $\lambda_8$ ,  $\lambda_{10}$ ,  $\lambda_{12}$ ,  $\lambda_{14}$ , and  $\lambda_{16}$  after reconfiguration.

MZS network, and 1.2 dB from the propagation loss of routing waveguides. Error-free operations of all the five wavelength channels show that the bandwidth between input port 4 and output port 8 is increased by  $2.5\times$  (50 Gb/s to 125 Gb/s) as shown in Figure 3.27(b). Figure 3.27(d) shows the eye diagram of these five channels. Note that  $\lambda_{10}$ ,  $\lambda_{12}$ , and  $\lambda_{14}$  from input port 4 are initially used for interconnecting with output port 2, 4, and 6 before reconfiguration, respectively. Although these three wavelengths are routed to output port 8 after reconfiguration, all-to-all interconnects through  $\text{FSR}_0$  are maintained (as shown in Figure 3.27(c)) so that input port 4 can still interconnect with output port 2, 4, and 6 at 25 Gb/s through  $\lambda_2$ ,  $\lambda_4$ , and  $\lambda_6$ , respectively.

### 3.3.3.3 Switching Speed Characterization

The switching speed of the Flex-LIONS chip is characterized by measuring the temporal response of the switching elements. Figure 3.28(a) shows the 5-kHz square-wave electrical driving signals that are applied to the MRR add-drop filters and the upper heater of the  $2 \times 2$  MZS. The peak-to-peak drive voltage is 2 V. Figure 3.28(b) and (c) show the measured optical waveform for the MRR add-drop filters and the  $2 \times 2$  MZS, respectively. The dashed lines mark the 10% and 90% power levels. The measured rise/fall time of the MRR add-drop filters and  $2 \times 2$  MZS are 7.6/13.6  $\mu\text{s}$  and 13.2/11.2  $\mu\text{s}$ , respectively. Faster switching speed can be obtained by using electro-optical (EO) tuning in the future [6, 75].

### 3.3.3.4 Power Consumption

Without tuning, the resonance of the MRR add-drop filters is designed to be located between  $\lambda_8$  in  $\text{FSR}_0$  and  $\lambda_9$  in  $\text{FSR}_1$  so that the required TO tuning power for reconfiguration is minimum. The average power consumption to correct the fabrication variation for each MRR add-drop filter is 4.23 mW. For the case shown in Section 3.3.3.2, the total power consumption is 141.81 mW, which includes 137.46 mW for tuning six MRR add-drop filters and 4.35 mW for switching five MZSs to the cross state. In the worst case, the total power consumption to reconfigure three wavelength channels between a pair of input and output ports is 320.81 mW, assuming the six MRR add-drop filters are tuned to drop the longest wavelength channels in  $\text{FSR}_1$  and the five MZSs on the path

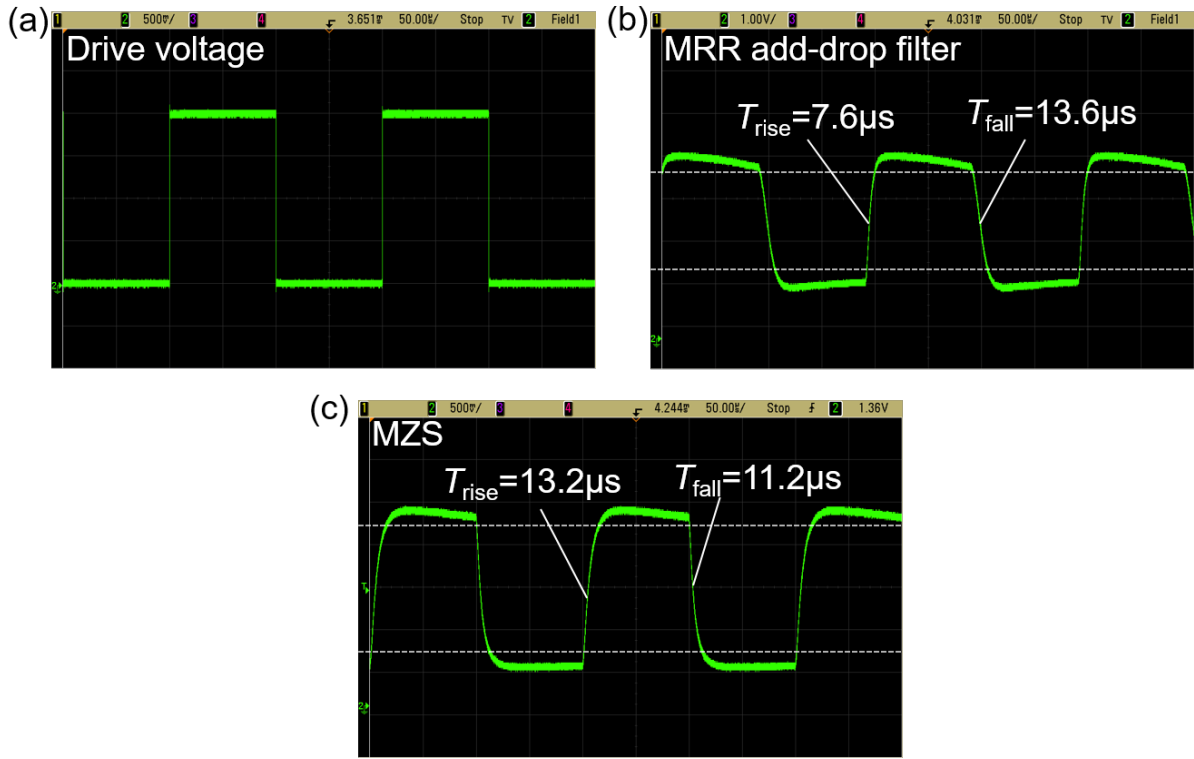


Figure 3.28. Time-domain optical response of the switch element. (a) Applied square-wave electrical drive signal. (b) Measured optical waveform for MRR add-drop filter. (c) Measured optical waveform for  $2 \times 2$  MZS.

are switched to bar state.

The TO tuning efficiency of the MRR add-drop filters and MZSs can be further improved by reducing the heater-waveguide distance [76], using silicon doped heater [77], or removing the waveguide substrate and adding air trenches [45]. In addition to the power required for tuning the resonant wavelength, MRR add-drop filters also consume power for wavelength stabilization. A recent work in [42] reported a 65 nm CMOS circuit for MRR resonance auto-alignment and tracking that consumed 5.17 mW. Further reduction in power consumption can be achieved by replacing thermo-optical tuning elements with electro-optical tuning elements [6].

## 3.4 O-Band Silicon Photonic Flex-LIONS With ns Switching Speed

This section presents the design, layout, and fabrication process of the O-band  $16 \times 16$  two-FSR silicon photonic Flex-LIONS ( $N = 16$ ,  $b = 4$ ) with EO-tuned switching elements.

### 3.4.1 Single Elements Design

#### 3.4.1.1 O-band $16 \times 16$ 80-GHz-Spacing SiN Cyclic AWGR Design

The  $16 \times 16$  80-GHz-spacing AWGR is designed on a 200 nm thickness low-loss SiN platform with silicon dioxide ( $\text{SiO}_2$ ) cladding. The input and output waveguides of the AWGR are ridge waveguides with the dimension of  $0.2 \mu\text{m} \times 1.2 \mu\text{m}$ . The AWGR is designed to be box-shaped with a center wavelength of 1280 nm. The width of the input/output port of the star coupler is optimized to be  $4.0/3.0 \mu\text{m}$ . The width of the 56 arrayed arms is tapered to  $1.9 \mu\text{m}$ . Figure 3.29(a) shows the layout of the  $16 \times 16$  80-GHz-spacing AWGR with a footprint of  $1.0 \text{ mm} \times 4.0 \text{ mm}$  (not including the input and output waveguide routings). Figure 3.29(b) shows the measured transmission spectrum from the center input (input port 8) of the AWGR. The measured center wavelength is 1280.55 (0.04% offset). The AWGR is cyclic with a channel spacing of 80.97 GHz (1.21% offset). The measured adjacent channel crosstalk is  $< -12 \text{ dB}$ , and the non-adjacent channel crosstalk is  $< -15 \text{ dB}$ . The measured 3-dB-bandwidth of the AWGR channel is 0.34 nm with a passband/channel-spacing ratio of 77.8%.

#### 3.4.1.2 O-band Si PIN Junction Design

Figure 3.30(a) shows the cross-section of the designed O-band (1280 nm) Si PIN junction. The Si rib waveguide is  $220 \text{ nm} \times 400 \text{ nm}$  with a slab thickness of 70 nm. The effective index and group index of the Si waveguide are simulated by Lumerical MODE as:  $n_{\text{eff}} = 2.69$  and  $n_g = 4.11$ . The Si wafer has an original p-type doping concentration of  $1 \times 10^{15} \text{ cm}^{-3}$ . The doping concentrations of the  $\text{P}^{++}$  and  $\text{N}^{++}$  contact areas are  $1 \times 10^{20} \text{ cm}^{-3}$ . The excess loss of the Si waveguide with varied offset values (distance from the  $\text{P}^{++}$  and  $\text{N}^{++}$  areas to the waveguide) is simulated using Lumerical Mode Solutions. The absorption of the  $\text{P}^{++}$  and  $\text{N}^{++}$  areas is calculated by the plasma dispersion effect of Si.



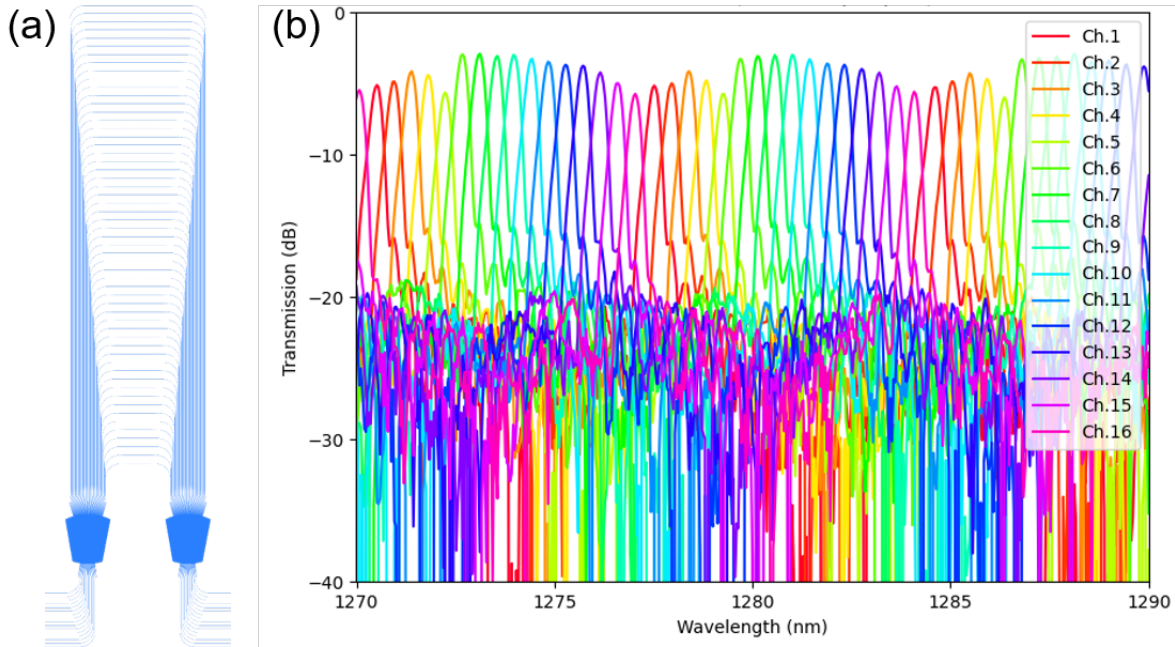


Figure 3.29. (a) Layout of the  $16 \times 16$  80-GHz-spacing SiN AWGR. (b) Measured transmission spectrum from the center input (input port 8) of the AWGR.

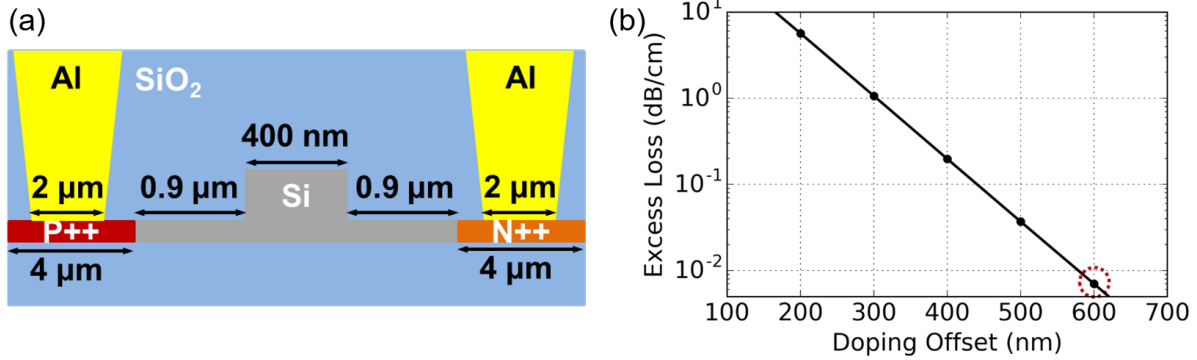


Figure 3.30. (a) Cross section of the designed O-band Si PIN junction. (b) Simulated excess loss with varied offset values.

As shown in Figure 3.30(b), the excess loss is lower than 0.01 dB/cm when the offset is higher than 600 nm. Considering the lateral diffusion of the dopants during the dopant activation, the offset value is set to be 900 nm.

The ion implantation energy and dose of the P<sup>++</sup> and N<sup>++</sup> Si are simulated using TRIM software. The ions used for P<sup>++</sup> and N<sup>++</sup> are boron and phosphorus, respectively. The simulated ion energies for P<sup>++</sup> and N<sup>++</sup> are 9 keV and 22.5 keV for a projected range of 35 nm. Figure 3.31 shows the simulated ion ranges. For a doping concentration of  $1 \times$

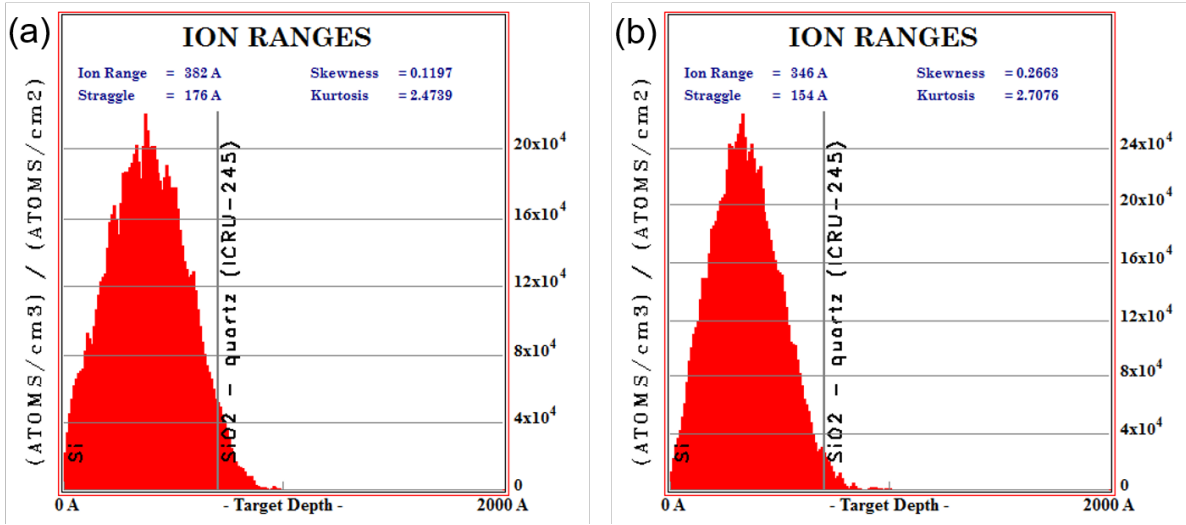


Figure 3.31. Simulated ion range for (a) P<sup>++</sup> (b) N<sup>++</sup> Si using TRIM.

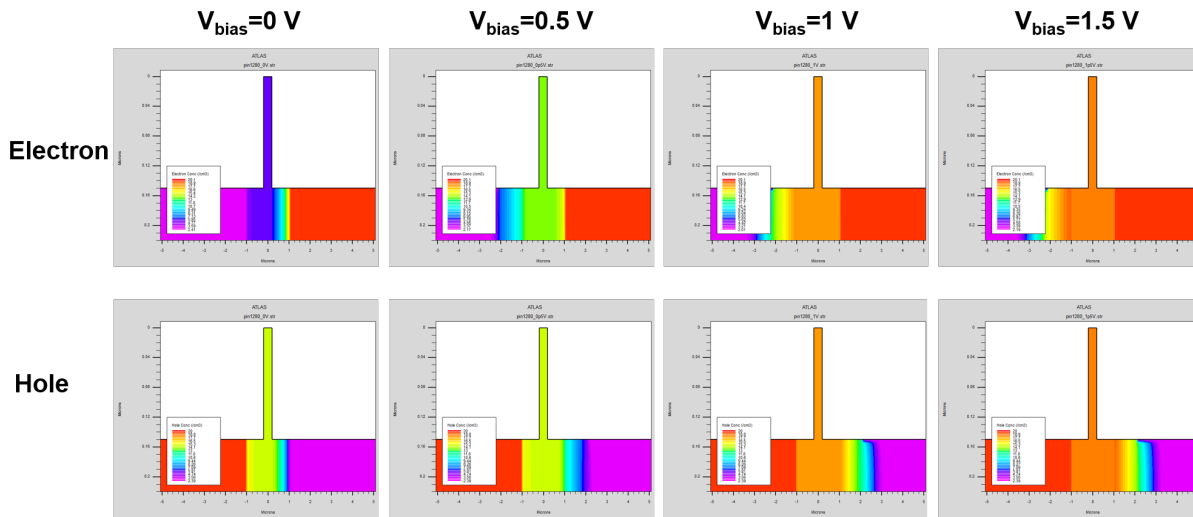


Figure 3.32. Carrier concentration profile under different forward bias voltages simulated by Silvaco Athena and Atlas.

$10^{20}\text{ cm}^{-3}$  at the surface (0-nm depth), the doses for P<sup>++</sup> and N<sup>++</sup> are calculated as  $4.5 \times 10^{15}\text{ cm}^{-2}$  and  $7 \times 10^{15}\text{ cm}^{-2}$ , respectively.

Figure 3.32 shows the carrier concentration (electrons and holes) profile under different forward bias voltages simulated by Silvaco Athena and Atlas. The fabrication process is included in the simulation with the annealing condition of 1050 degrees for 30 s. The source code is included in the Appendix.

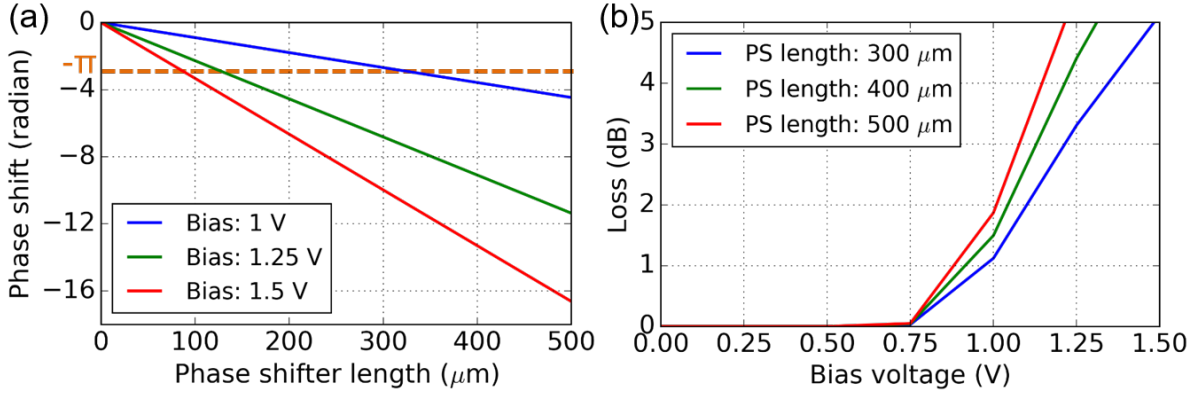


Figure 3.33. (a) Simulated phase shift with different forward bias voltages and different phase shifter length. (b) Simulated insertion loss of the phase shifter with different forward bias voltages and different phase shifter length.

### 3.4.1.3 O-band EO-Tuned Si MRR Add-Drop Filter and MZ Switch Design

Based on the simulated carrier concentrations in Section 3.4.1.2, the refractive index and material absorption change of Si can be calculated based on the plasma dispersion effect (at 1.3 μm):

$$\Delta n = -6.2 \times 10^{-22} \Delta n_e - 6.0 \times 10^{-18} \Delta n_h^{0.8} \quad (3.1)$$

$$\Delta \alpha = 6.0 \times 10^{-18} \Delta n_e + 4.0 \times 10^{-18} \Delta n_h \quad \text{cm}^{-1} \quad (3.2)$$

where  $\Delta n_e$  and  $\Delta n_h$  are the free-electron and free-hole carrier concentrations, respectively.

The effective index and waveguide propagation loss change of the Si waveguide can be calculated using the following equations:

$$\Delta n_{\text{eff}} = \frac{\int \int \Delta n(x, y) n(x, y) |E(x, y)|^2 dx dy}{\int \int n(x, y) |E(x, y)|^2 dx dy} \quad (3.3)$$

$$\alpha = \frac{\int \int \alpha(x, y) n(x, y) |E(x, y)|^2 dx dy}{\int \int n(x, y) |E(x, y)|^2 dx dy} \quad (3.4)$$

For the O-band MZS,  $\pi$ -phase shift needs to be provided if phase shifters are placed on both arms of the MZS. The phase shift of the phase shifter can be calculated by:  $\Delta \phi = \frac{2\pi \Delta n_{\text{eff}} L}{\lambda}$  where  $L$  is the length of the phase shifter. The insertion loss of the phase shifter can be calculated by:  $IL[\text{dB}] = -10 \log_{10}(\exp(-\alpha L))$ . Figure 3.33(a) shows the simulated phase shift with varied forward bias voltages and varied phase shifter length.

It can be seen that with a length of 500  $\mu\text{m}$ , the required bias voltage is lower than 1 V. Figure 3.33(b) shows the simulated insertion loss of the phase shifter with varied forward bias voltages and varied phase shifter length. With the phase shifter length of 500  $\mu\text{m}$ , the insertion loss is lower than 1.87 dB.

Figure 3.34(a) shows the O-band MRR add-drop filter architecture, which contains one MRR coupled to two bus waveguides. The FSR of the MRR is designed as 1320 GHz (7.2 nm), which is  $16.5\times$  the channel spacing of the AWGR (i.e., 80 GHz). In this way, the filter only drops one wavelength channel in the two FSRs of the AWGR at a time. Then the radius of the MRR can be calculated by:  $R_{\text{MRR}} = \frac{\lambda^2}{2\pi n_g \text{FSR}_{\text{MRR}}} = 8.80\mu\text{m}$ . The transmission spectrum of the through and drop port of the MRR add-drop filter can be calculated by:

$$T_t = \frac{r_2^2 a^2 - 2r_1 r_2 a \cos\phi + r_1^2}{1 - 2r_1 r_2 a \cos\phi + (r_1 r_2 a)^2} \quad (3.5)$$

$$T_d = \frac{(1 - r_1^2)(1 - r_2^2)a}{1 - 2r_1 r_2 a \cos\phi + (r_1 r_2 a)^2} \quad (3.6)$$

where  $r_1$  and  $r_2$  are the self-coupling coefficient at the upper and lower bus waveguides,  $\phi = \frac{2\pi(n_{\text{eff}} + \Delta n_{\text{eff}})L}{\lambda}$ ,  $L$  is the round trip length,  $a$  is the single-pass amplitude transmission which can be calculated by  $a^2 = \exp(-\alpha L)$ .

The critical coupling of the MRR happens at  $r_2 a = r_1$ . Here, we assume  $r_2 = r_1$  since  $a \approx 1$ . Figure 3.34(c-f) shows the Simulated transmission spectra of the through and drop port at different bias voltages and varied self-coupling coefficient  $r$ . Figure 3.34 (b) summarizes the simulated insertion loss and 3-dB bandwidth of the drop port at 0V bias.

### 3.4.2 O-band $16\times 16$ SiPh Flex-LIONS Design and Layout

Figure 3.35(a) shows the O-band  $16\times 16$  SiPh Flex-LIONS chip Layout ( $N = 16$ ,  $b = 4$ ). The bottom 220-nm Si layer contains MRR add-drop filters and a Beneš MZS network. Above the Si layer is the 200-nm SiN waveguide layer, which contains the 80-GHz-spacing  $16\times 16$  low-crosstalk SiN AWGR. The SiN layer vertically interfaces with the Si layer through inverse-tapered evanescent couplers. The MRR add-drop filters and MZSs are EO tuned by PIN junctions in order to provide ns switching speed. Thermal waveguide

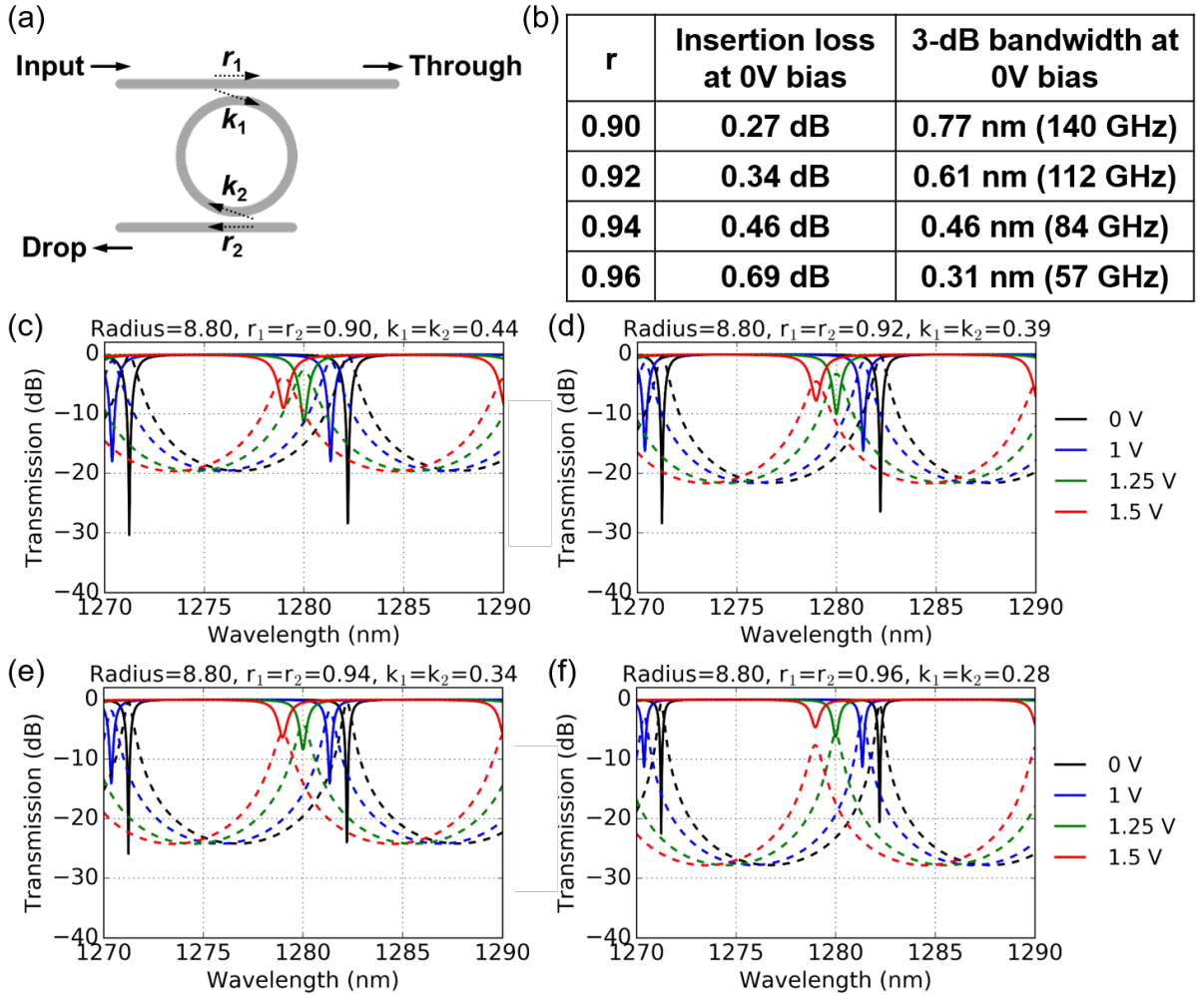


Figure 3.34. (a) O-band MRR add-drop filter structure. (b) Summary of the simulated insertion loss and 3-dB bandwidth of the drop port at 0V bias. (c-f) Simulated transmission spectra of the through and drop port at different bias voltages and varied self-coupling coefficient  $r$ .

heaters are used for resonance alignment in MRRs and fabrication error compensation in MZSs. Edge coupler arrays with 127- $\mu\text{m}$ -pitch are used for low coupling loss from the fiber array to the chip. The edge coupler contains a SiN inverse taper from 1.2  $\mu\text{m}$  to 200 nm and an evanescent coupler from the SiN layer to the Si layer. Two loop-back waveguides are placed on both sides of the edge coupler array for fiber array alignment. O-band MMI waveguide crossings are designed to lower the overall insertion loss as shown in Figure 3.35(d).

The radii of four MRR add-drop filters are 8.80  $\mu\text{m}$ , 8.82  $\mu\text{m}$ , 8.84  $\mu\text{m}$ , and 8.86  $\mu\text{m}$ ,

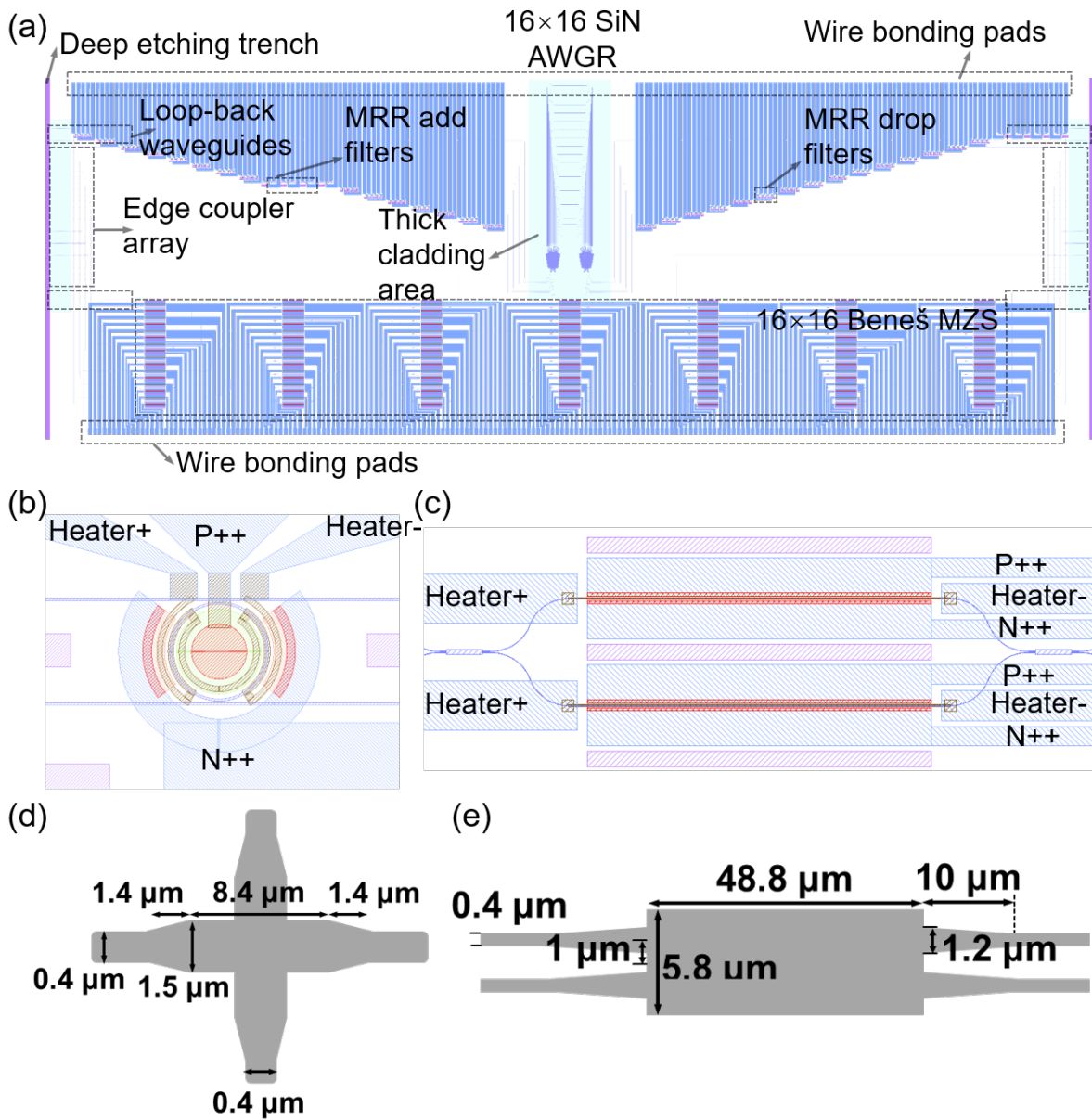


Figure 3.35. (a) Layout of O-band  $16 \times 16$  SiPh Flex-LIONS ( $7.6 \text{ mm} \times 21.8 \text{ mm}$ ). (b) Layout of O-band MRR add-drop filter with EO and TO tuning. (c) Layout of O-band  $2 \times 2$  MZS with EO and TO tuning. (d) Design of O-band MMI based waveguide crossing. (e) Design of O-band  $2 \times 2$  MMI coupler.

respectively. The gap of the MRR add-drop filters is  $300 \text{ nm}$ . PIN diodes and spiral resistive heaters are designed to provide both EO and TO tuning for the MRR add-drop filters as shown in Figure 3.35(b). The width of the heater is  $1 \mu\text{m}$ . Figure 3.35(c) shows the layout of O-band  $2 \times 2$  MZS with EO and TO tuning as the building block of the Beneš

MZS network. In order to achieve minimum TO tuning power, heaters are placed on both arms of the MZS. The width and length of the heaters are  $2\ \mu\text{m}$  and  $500\ \mu\text{m}$ , respectively. Figure 3.35(e) shows the layout of O-band  $2\times 2$  MMI coupler. Optical power taps are placed along the reconfiguration path of the chip for power monitoring. The  $\text{P}^{++}$ ,  $\text{N}^{++}$ , heater+, and heater- traces of both the MRR add-drop filters and  $2\times 2$  MZSs are routed to the top and bottom edge of the chip for wire-bonding. The wire-bonding pads are  $80\ \mu\text{m} \times 140\ \mu\text{m}$  with  $100\text{-}\mu\text{m}$  pitch.

### 3.4.3 O-band $16\times 16$ SiPh Flex-LIONS Fabrication Process

Figure 3.36 shows the fabrication flow charts for the O-band  $16\times 16$  SiPh Flex-LIONS chip. The fabrication process starts with a 220-nm SOI wafer with  $3\text{-}\mu\text{m}$ -thick buried oxide. Firstly, the Si layer is defined by deep-UV projection lithography and ICP etching. Then the  $\text{N}^{++}$  and  $\text{P}^{++}$  areas are implanted with a photoresist mask based on the ion energy and dose values simulated in Section 3.4.1.2. Then a 1000-nm  $\text{SiO}_2$  is deposited by plasma-enhanced chemical vapor deposition (PECVD) and then planarized to 650 nm by CMP. Following the deposition of a 200-nm SiN layer by high density plasma chemical vapor deposition (HDPCVD), the AWGR is patterned by deep-UV lithography and ICP etching. Then a  $2.5\text{-}\mu\text{m}$   $\text{SiO}_2$  cladding is deposited and planarized. Subsequently, the oxide cladding window and contact via opening are opened by ICP etching. The aluminum (Al) via for  $\text{P}^{++}$  and  $\text{N}^{++}$  contacts is fabricated by E-beam evaporation and lift-off. Then the 400-nm-thick platinum (Pt) heater layer and 800-nm-thick gold (Au) metal pad layer are fabricated by E-beam evaporation and lift-off. Finally, a  $140\text{-}\mu\text{m}$  deep etching trench is fabricated using ICP etching.

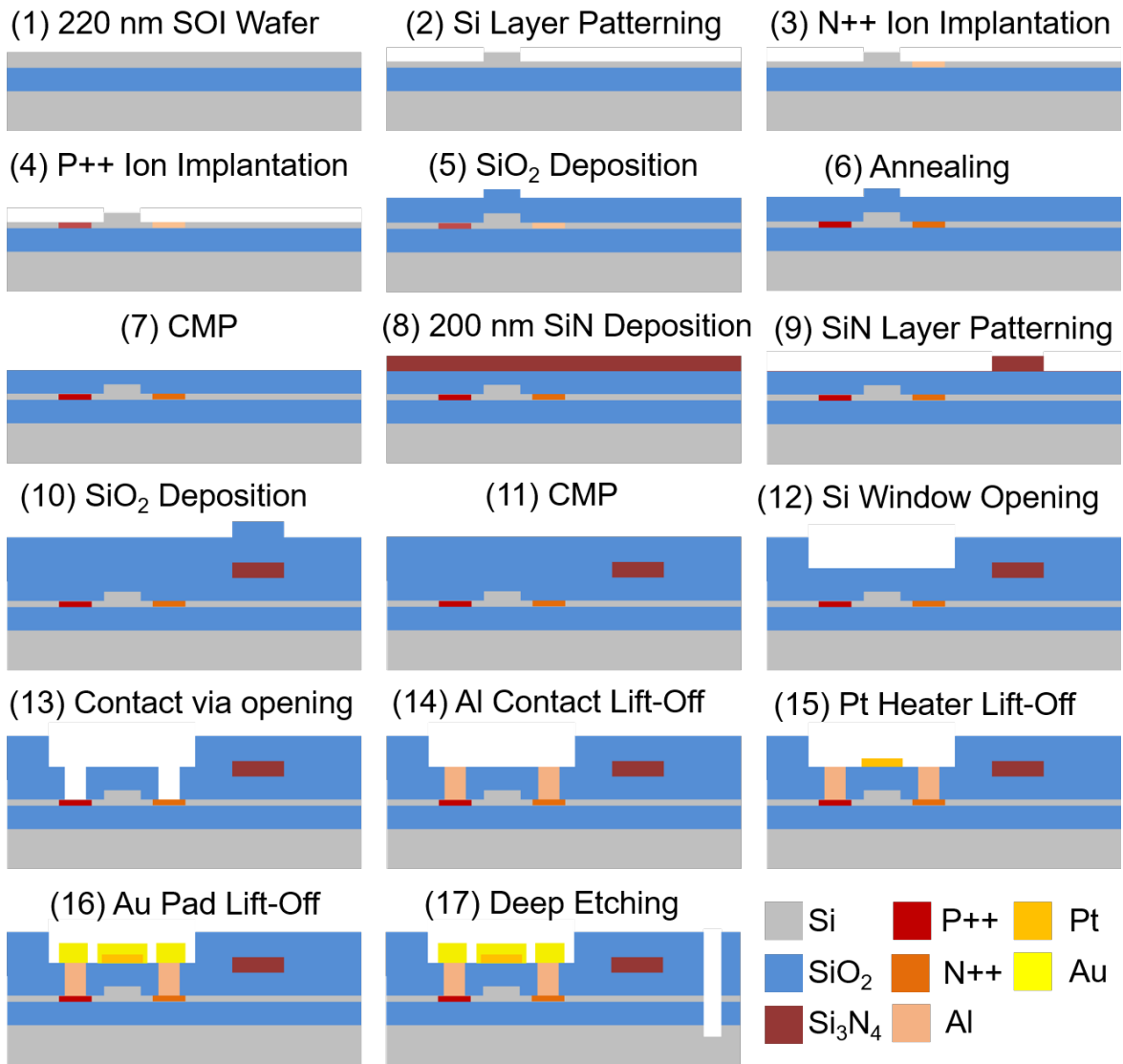


Figure 3.36. Fabrication flow charts for the O-band 16×16 SiPh Flex-LIONS.



# Chapter 4

## Scalable and Compact Tensorized Photonic Neural Networks

Artificial neural networks (ANN) have proven their remarkable capabilities in various tasks including computer vision, speech recognition, machine translations, medical diagnoses, and the game of Go [78]. Neuromorphic computing processors such as IBM TrueNorth [17] and Intel Loihi [18] have shown significantly superior performance compared with traditional central processing units (CPUs) for specific neural network tasks. A majority of the electrical ANN hardware’s energy consumption comes from data movement in the synaptic interconnections. Photonic neural networks (PNNs) are expected to improve the energy efficiency and throughput significantly compared with electrical ANNs due to the capabilities of transmitting data at the speed of light without having a length-dependent impedance [79]. On the other hand, a shallow network with large fully-connected layers is proved to achieve almost the same accuracy as an ensemble of deep convolutional neural networks (CNNs) [80]. Thus it is desirable to implement high-radix (e.g.,  $1024 \times 1024$ ) photonic synaptic interconnections.

### 4.1 Tensor-Train Decomposed Synaptic Interconnects

PNNs typically consist of an input neuron layer, many hidden neuron layers, an output neuron layer, and synaptic interconnections which can be abstracted by arbitrary weight matrices  $W$  (Figure 4.1). As shown in Figure 4.2(a) and (b), the synaptic interconnections

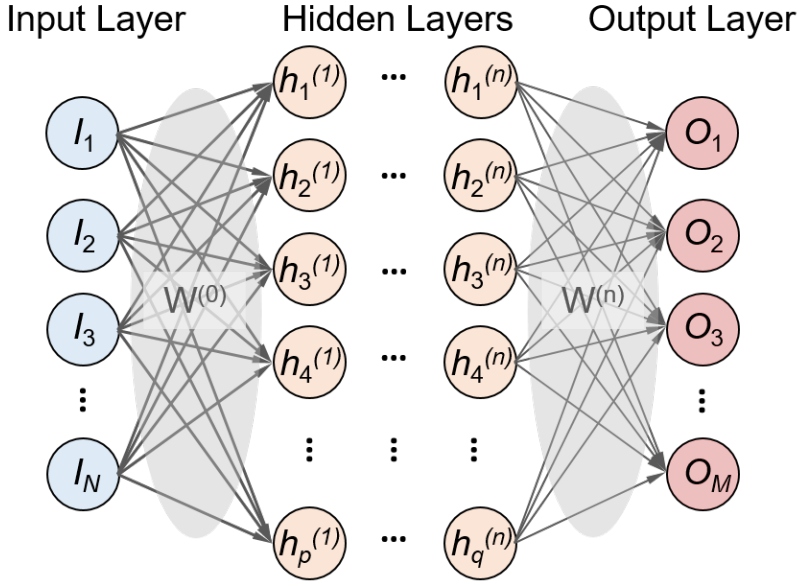


Figure 4.1. Schematic of the ANN architecture with input layer, hidden layers, output layers, and synaptic interconnections. Each synaptic interconnection is a linear operation represented by an arbitrary weight matrix  $W$ .

in PNN are typically enabled by a ‘rectangular’ [81] or ‘triangular’ MZI mesh [82] which can be abstracted by a unitary arbitrary weight matrix. As the building block, each  $2 \times 2$  MZI contains two phase shifters as shown in Figure 4.2(c). The ‘rectangular’ MZI mesh is more uniform in the insertion loss while the ‘triangular’ MZI mesh is more friendly for self-configuring [83].

However, an  $N \times N$  MZI mesh requires  $O(N^2)$  MZIs and  $O(N)$  cascaded stages [81]. Although MEMS [84] or non-volatile [85] technologies can be used to reduce the length of the phase shifters, the MZI meshes are still difficult to scale to high radix. For example, Lightmatter’s Mars device integrates  $64 \times 64$  MZI meshes with Nano Optical Electro Mechanical System (NOEMS) as tuning elements on a  $150 \text{ mm}^2$  chip [19]. The predicted chip size for  $1024 \times 1024$  will be  $384 \text{ cm}^2$ , which is extremely difficult to implement. To address the scalability issue of the PNN, we propose to use tensor-train (TT) decomposed synaptic interconnections, which can realize large-scale PNNs with reduced hardware resources [86, 87].

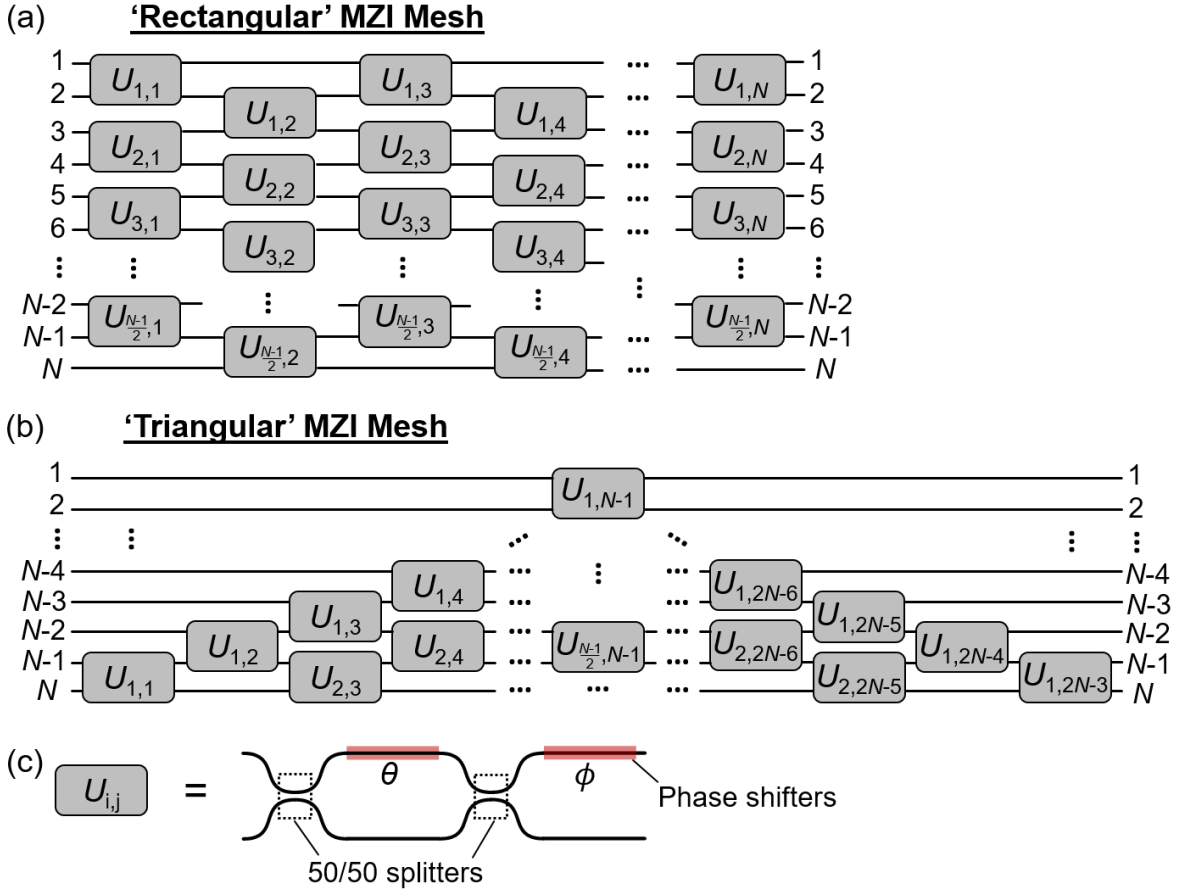


Figure 4.2. (a)  $N \times N$  unitary matrix represented by a ‘rectangular’ MZI mesh. (b)  $N \times N$  unitary matrix represented by a ‘triangular’ MZI mesh. (c)  $2 \times 2$  MZIs as the building blocks of the MZI mesh. ( $N$  is assumed to be an odd number here)

#### 4.1.1 Conventional PNN Architecture

For conventional PNNs, the  $N$  neuron signals at the neuron layer  $t$  are converted to optical signals by  $N$  optical modulators at the wavelength of  $\lambda_0$ , fully interconnected by an  $M \times N$  MZI mesh, and detected by  $M$  photodetectors at the neuron layer  $t+1$  as shown in Figure 4.3. The MZI mesh can be abstracted by a unitary weight matrix  $W^{(t)}$  that operate linear transformation of an input vector  $x^{(t)}$  by  $x^{(t+1)} = W^{(t)}x^{(t)}$  where  $W^{(t)} \in R^{M \times N}$ ,  $x^{(t)} \in R^N$ , and  $x^{(t+1)} \in R^M$ . Letting  $M = N$ , the  $N \times N$  MZI mesh can be implemented by a ‘rectangular’ fabric with  $N(N-1)/2$  MZIs as building blocks in  $N$  cascaded stages. Letting  $N=1024$ , the MZI mesh requires 523,776 MZIs and 1024 cascaded stages. Assuming the insertion loss of each MZI is 0.2 dB as in [88], the total

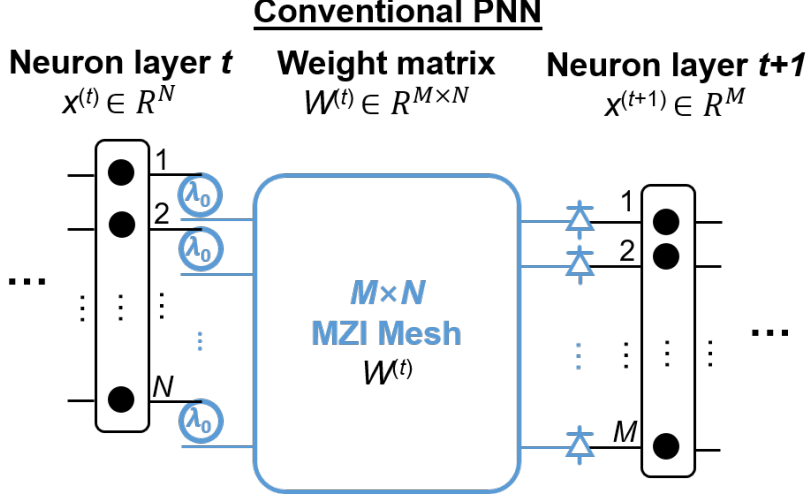


Figure 4.3. Conventional PNN architecture with  $N$  neurons at layer  $t$ ,  $M$  neurons at layer  $t+1$  and an  $M \times N$  synaptic interconnections.

insertion loss is 204.8 dB which is difficult to be compensated by optical amplifiers.

#### 4.1.2 Tensorized PNN Architecture

To represent weight matrix  $W^{(t)} \in R^{M \times N}$  in the tensor-train format, we firstly assume  $M$  and  $N$  can be factored as  $M = \prod_{k=1}^d M_k$ ,  $N = \prod_{k=1}^d N_k$ . Letting  $\mu$  and  $\nu$  be the natural bijections from indices  $(i, j)$  of  $W^{(t)}$  to indices  $(\mu_1(i), \nu_1(j), \dots, \mu_d(i), \nu_d(j))$  of an order- $2d$  weight tensor  $\mathbf{W}^{(t)}$ . Then  $W^{(t)}(i, j) = \mathbf{W}^{(t)}(\mu_1(i), \nu_1(j), \dots, \mu_d(i), \nu_d(j))$ . As shown in Figure 4.4, the TT-decomposition expresses the tensor  $\mathbf{W}^{(t)}$  as a series of tensor products [89–91]:

$$\mathbf{W}^{(t)}(\mu_1(i), \nu_1(j), \dots, \mu_d(i), \nu_d(j)) = \prod_{k=1}^d \mathbf{G}_k(\cdot, \mu_k(i), \nu_k(j), \cdot) \quad (4.1)$$

where the 4-way tensor  $\mathbf{G}_k \in R^{R_{k-1} \times M_k \times N_k \times R_k}$  is the TT core, the vector  $R_{TT} = (R_0, R_1, \dots, R_d)$  is the TT-rank, and  $R_0 = R_d = 1$ . In this way, the total number of parameters can be reduced from  $M \times N$  into the summation of the parameters of each small TT cores, i.e.,  $\sum_{k=1}^d R_{k-1} M_k N_k R_k$ .

To achieve the tensor-train layer as shown in Figure 4.5, the vector  $x^{(t)}$  and  $x^{(t+1)}$  should also be reshaped into the tensor format by:  $x^{(t)}(j) = \mathbf{x}^{(t)}(\mu_1(j), \dots, \mu_d(j))$ ,  $x^{(t+1)}(i) = \mathbf{x}^{(t+1)}(\nu_1(i), \dots, \nu_d(i))$  where  $x^{(t)} \in R^N$ ,  $x^{(t+1)} \in R^M$ ,  $\mathbf{x}^{(t)} \in R^{N_d \times \dots \times N_1}$ ,  $\mathbf{x}^{(t+1)} \in R^{M_d \times \dots \times M_1}$ . Then in TT-format, the linear transformation by the weight matrix tensor  $\mathbf{W}^{(t)}$  can be

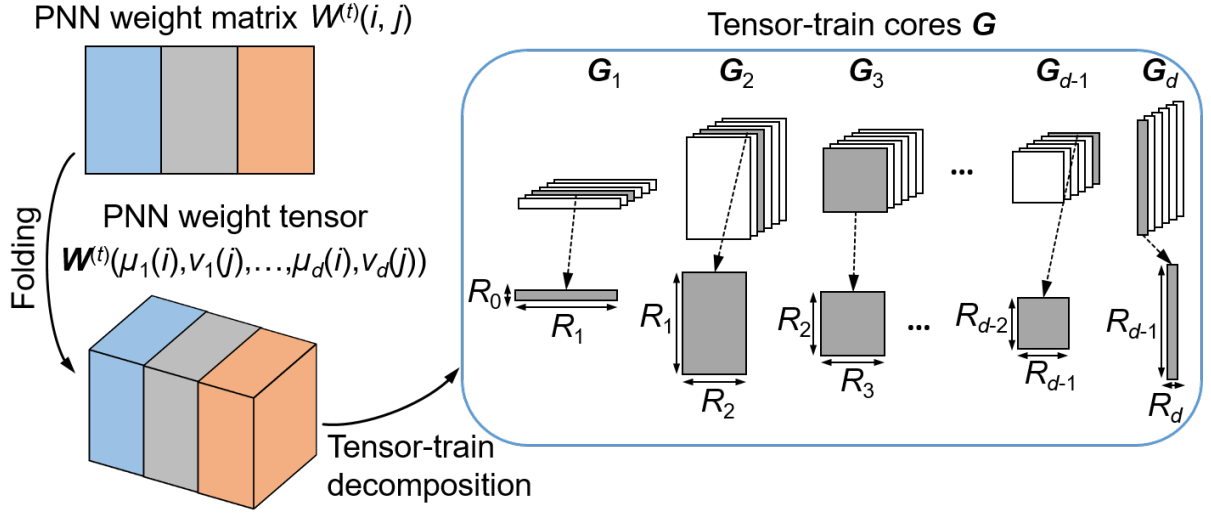


Figure 4.4. Weight matrix TT-decomposition for parameter compression.

represented by [90]:

$$\mathbf{x}^{(t+1)}(\nu_1(i), \dots, \nu_d(i)) = \prod_{k=1}^d \mathbf{G}_k(:, \mu_k(i), \nu_k(j), :) \mathbf{x}^{(t)}(\mu_1(j), \dots, \mu_d(j)) \quad (4.2)$$

Figure 4.6 shows the detailed device implementation of the tensorized PNN architecture [87]. For convenience, we let  $M = N$ ,  $M_1 = \dots = M_d = N_1 = \dots = N_d$ ,  $R_1 = \dots = R_{d-1} = R$ . To emulate the tensor products, the input signal  $\mathbf{x}^{(t)}$  at the neuron layer  $t$  is firstly encoded by  $N_{d-1}$  layers of  $N_d R_d$  groups of WDM microring modulator arrays with the wavelengths of  $\lambda_1, \dots, \lambda_g$ , where  $g = N/(N_{d-1} N_d)$ . Since there is no time-domain multiplexing involved, the total number of modulators of the tensorized PNNs is the same as the conventional PNNs. Then the  $\mathbf{x}^{(t)}$  is multiplied by each TT cores in the sequence of  $\mathbf{G}_d, \mathbf{G}_{d-1}, \dots, \mathbf{G}_1$ . At the TT core  $\mathbf{G}_k \in R^{R_{k-1} \times M_k \times N_k \times R_k}$  where  $k = d, \dots, 1$ , the input optical signals are  $\mathbf{G}_{k+1} \dots \mathbf{G}_d \mathbf{x}^{(t)} \in R^{R_k \times N_k \times N_{k-1} \times \dots \times N_1 \times M_{k+1} \times \dots \times M_d}$ . The  $\mathbf{G}_{k+1} \dots \mathbf{G}_d \mathbf{x}^{(t)}$  is represented by  $N_{k-1}$  layers of  $N_k$  groups of  $R_k$  inputs, and the TT core  $\mathbf{G}_k$  is represented by  $N_{k-1}$  layers of identical  $R_{k-1} M_k \times N_k R_k$  MZI meshes. In this way the product between  $\mathbf{G}_{k+1} \dots \mathbf{G}_d \mathbf{x}^{(t)}$  and  $\mathbf{G}_k$  is implemented and the output signals are  $\in R^{R_{k-1} \times M_k \times N_{k-2} \times \dots \times N_1 \times N_{k-1} \times M_{k+1} \times \dots \times M_d}$ . Then 3D photonic cross-connects are used to switch the index of  $M_k$  and  $N_{k-1}$  so that the output signals are  $\mathbf{G}_k \dots \mathbf{G}_d \mathbf{x}^{(t)} \in R^{R_{k-1} \times N_{k-1} \times N_{k-2} \times \dots \times N_1 \times M_k \times M_{k+1} \times \dots \times M_d}$ , which is ready to serve as the input signals of the

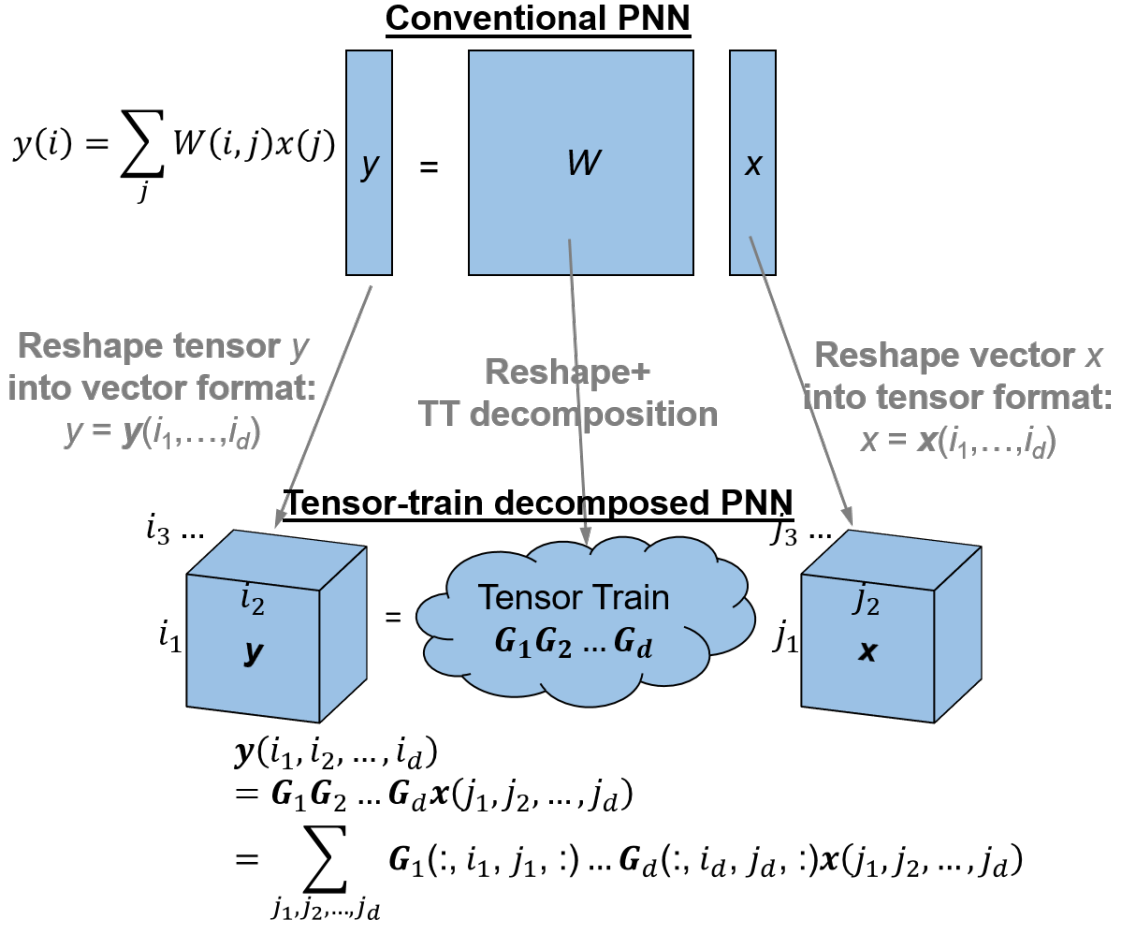


Figure 4.5. Tensor-train layer.

next TT core  $\mathbf{G}_{k-1}$ . After products with all the TT cores, the  $\mathbf{G}_1 \dots \mathbf{G}_d \mathbf{x}^{(t)} = \mathbf{x}^{(t+1)}$  is demultiplexed and received by WDM microring add-drop filter and photodetector arrays.

By adding parallelism in both wavelength domain using WDM technology and in space domain using 3D photonics, our proposed tensorized PNN architecture can achieve all-optical tensor core products and index manipulations without optical-to-electrical-to-optical (O/E/O) conversions. Thus all the benefits of the conventional PNNs, including low latency and high throughput, can be maintained. Alternatively, 3D photonics can be replaced by 2D photonics with MZI meshes placed side by side in a single plane.

### 4.1.3 Comparison Between Tensorized and Conventional PNNs

To evaluate the scalability of the tensorized PNNs, the number of MZIs and the on-chip insertion loss as a function of the radix  $N$  are the two primary metrics. Letting  $M = N$

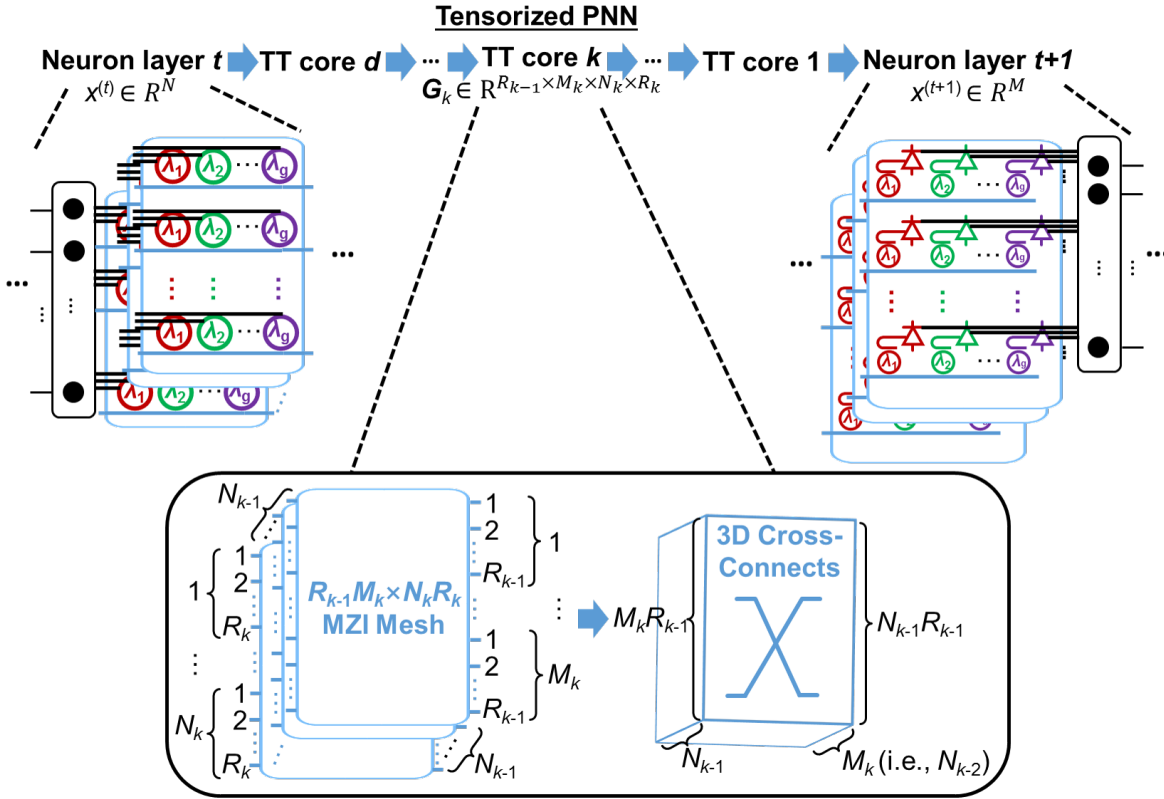


Figure 4.6. Tensorized PNN architecture with cascaded TT cores implemented by 3D  $R_{k-1}M_k \times N_kR_k$  MZI meshes and cross-connects. Alternatively, the 3D photonics can be replaced by putting MZI meshes side by side in a single plane.

and  $R_0 = R_1 = \dots = R_d = R$ , the tensorized PNNs reduce the number of MZIs from  $N(N-1)/2$  to  $\sum_{k=1}^d N_{k-1}(RN_k - 1)RN_k/2$  compared with conventional PNNs. The insertion loss of the tensorized PNNs is contributed by  $R \sum_{k=1}^d N_k$  cascaded stages of MZIs and  $d$  cross-connects:  $IL_{\text{tensorized\_PNN}} = R \sum_{k=1}^d N_k \cdot IL_{\text{MZI}} + d \cdot IL_{\text{cross-connect}}$ . The insertion loss of conventional PNNs can be calculated by  $IL_{\text{conventional\_PNN}} = N \cdot IL_{\text{MZI}}$ . Here we assume the insertion loss of each MZI is  $IL_{\text{MZI}} = 0.2$  dB as in [88], and the insertion loss of each vertical-optical-via based cross-connect is  $IL_{\text{cross-connect}} = 1.3$  dB as in [92].

Letting  $N = 2^n$ ,  $M_1 = \dots = M_d = N_1 = \dots = N_d = 2^l$ , and  $d = n/l$ . Figure 4.7(a) and (b) plot the number of MZIs as a function of the radix  $N$  with  $l = 1$  and  $l = 2$ , respectively. At the radix of  $N = 1024$ , for TT-rank  $R = 5$  and  $l = 1$ , the tensorized PNNs require  $582\times$  fewer MZIs, which shows a significant improvement in compactness

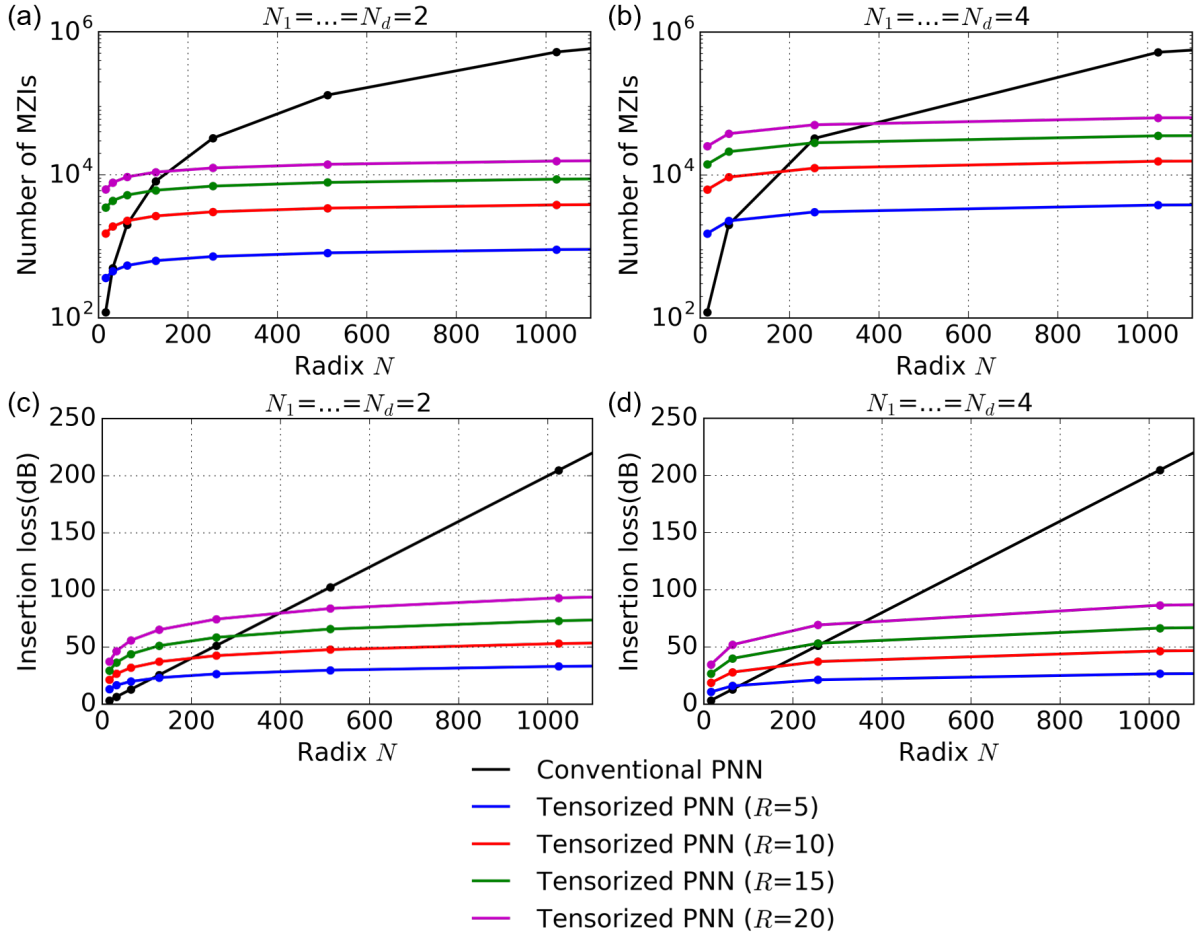


Figure 4.7. Comparison of number of MZIs as a function of the radix  $N$  with  $N_1 = \dots = N_d = 2$  (a) and 4 (b). Comparison of insertion loss as a function of the radix  $N$  with  $N_1 = \dots = N_d = 2$  (a) and 4 (b).

and energy efficiency since the power consumption and device footprint of the synaptic interconnections are proportional to the number of MZIs. Figure 4.7(c) and (d) plot the insertion loss as a function of the radix  $N$  with  $l = 1$  and  $l = 2$ , respectively. At the radix of  $N = 1024$ , conventional PNNs have an insertion loss of 204.8 dB, which requires a huge amount of optical gain to compensate. With  $R = 5$  and  $l = 1$ , our approach has an insertion loss of 33 dB, which is 171.8 dB lower than the conventional PNNs.

#### 4.1.4 Example: $1024 \times 1024$ Tensorized PNN in a Single Plane

Figure 4.8 shows the device architecture example of the  $1024 \times 1024$  tensorized PNN. Here  $1024 \times 1024$  is factorized as  $4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4$ , which has the potential to achieve  $> 95\%$  training accuracy [93]. As a result, the weight matrix is decomposed



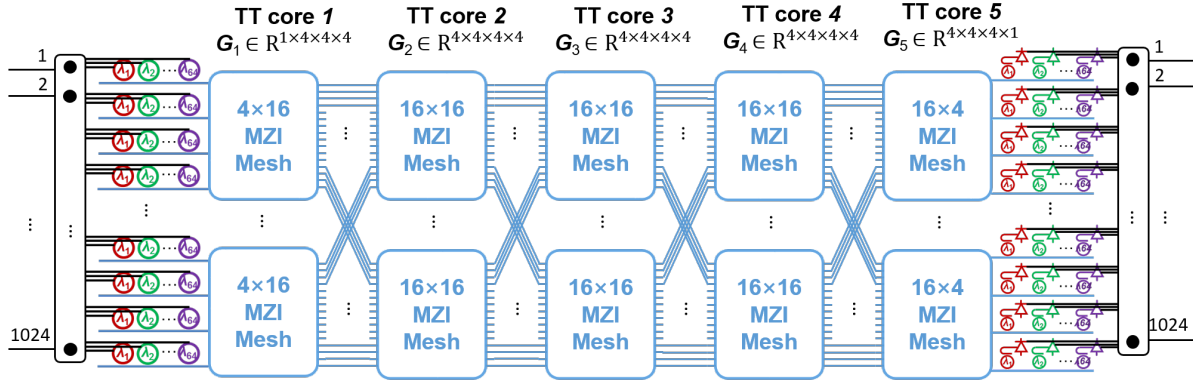


Figure 4.8. Schematic of the device architecture of the  $1024 \times 1024$  tensorized PNN.

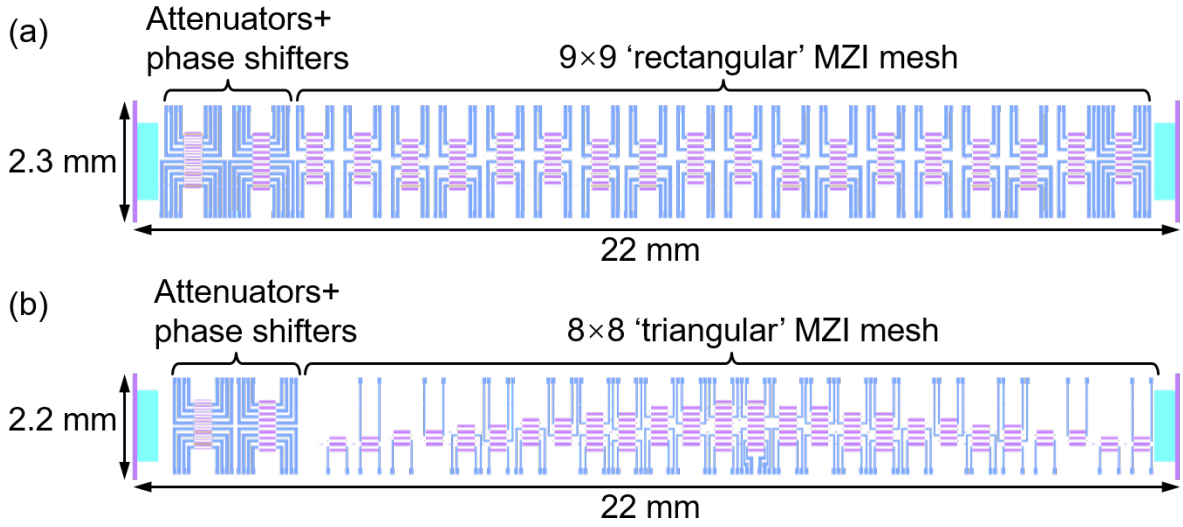


Figure 4.9. (a) Layout of an  $9 \times 9$  ‘rectangular’ MZI mesh. (b) Layout of a  $8 \times 8$  ‘triangular’ MZI mesh.

into five TT-cores. Each TT core contains four  $4 \times 16$  or  $16 \times 16$  MZI meshes side-by-side and cross-connections. The input neuron signals are firstly modulated by sixteen 64-wavelength WDM microring modulators arrays, then multiplied by each TT-cores, and finally detected by sixteen 64-wavelength WDM microring add-drop filter and detector arrays.

## 4.2 Silicon Photonic Tensor Core Design and Layout

Figure 4.9 (a) and (b) show the layout of a silicon photonic  $9 \times 9$  ‘rectangular’ MZI mesh and a silicon photonic  $8 \times 8$  ‘triangular’ MZI mesh, which can serve as the TT cores of the

tensorized PNN. The silicon waveguide is designed as a  $400 \text{ nm} \times 220 \text{ nm}$  rib waveguide with a slab thickness of  $70 \text{ nm}$ . Each MZI contains one phase shifter on the upper arm and one phase shifter at the upper output waveguide. The phase shifters are TO-tuned with micro-heaters on top of the waveguides. Edge coupler arrays with  $127\text{-}\mu\text{m}$ -pitch are used for low coupling loss from the fiber array to the chip. The edge coupler contains a SiN inverse taper from  $1.2 \mu\text{m}$  to  $200 \text{ nm}$  and an evanescent coupler from the SiN layer to the Si layer. Two loop-back waveguides are placed on both sides of the edge coupler array for fiber array alignment. Each MZI mesh has an MZI-based attenuator array and a phase shifter array at the input for experimental measurements. The chip sizes of the  $9 \times 9$  ‘rectangular’ MZI mesh and the  $8 \times 8$  ‘triangular’ MZI mesh are  $2.3 \text{ mm} \times 22 \text{ mm}$  and  $2.2 \text{ mm} \times 22 \text{ mm}$ , respectively. The fabrication of the SiPh tensor cores is in line with the process in Section 3.4.3.

# Chapter 5

## Light Sources for Computing Systems

### 5.1 Transfer-Printed III-V-on-Si Quantum Dot Lasers

Heterogeneously integrated III-V-on-Si lasers have attracted a lot of interest in the past 15 years [94]. The wafer bonding technique has successfully been used to transfer III-V materials onto the SOI substrates. However, such a technique suffers from the wafer size mismatch and inefficient use of the III-V layers. Transfer printing is a promising technique since it can be done at a wafer-scale and make the best use of the epitaxial materials [95]. Then the laser fabrication can be performed after the printing process using standard lithography and etching technologies.

#### 5.1.1 III-V-on-Si Hybrid Optical Amplifier Design

Figure 5.1(a) shows the cross-section of the designed III-V-on-Si hybrid SOA. The III-V stack is transfer printed on top of the Si waveguide wafer. The Si rib waveguides are 500-nm-thick with a slab thickness of 220 nm. The quantum dot (QD) core region is defined by hydrogen implantation with a width of 3  $\mu\text{m}$ . The optical mode in the hybrid SOA is shared by both the Si and QD core regions. Thus two parameters named core confinement factor  $\Gamma_{\text{core}}$  and Si confinement factor  $\Gamma_{\text{Si}}$  are used to represent the amount of optical mode in the core and Si regions, respectively. Here, we fix the III-V wafer stack design, QD core region width, silicon height, and etch depth. Figure 5.1(a) and (b)

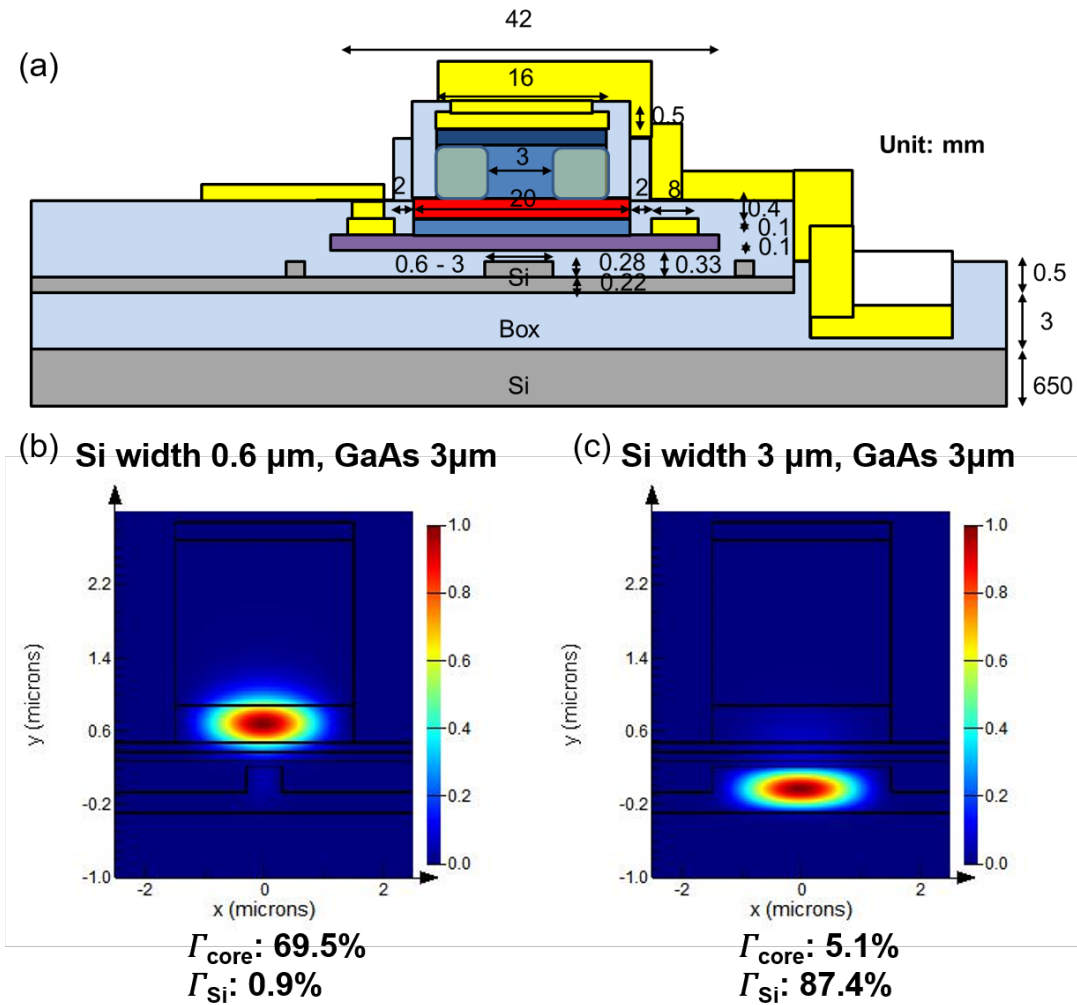


Figure 5.1. (a) Cross-section of the designed III-V-on-Si hybrid SOA. (b) Simulated mode profile and  $\Gamma_{\text{core}}$  with the Si waveguide width of 0.6 μm. (c) Simulated mode profile and  $\Gamma_{\text{core}}$  with the Si waveguide width of 3 μm.

show the simulated optical mode profiles with the Si waveguide width of 0.6 μm and 3 μm, respectively. The corresponding  $\Gamma_{\text{core}}$  values are 69.5% and 5.1%, respectively. This allows the transition from the amplifier region to the Si waveguide region by tapering the Si waveguide width from 0.6 μm to 3 μm.

The III-V-on-Si taper for optical transitions is divided into five regions as shown in Figure 5.2. The functionalities of different taper regions are:

- $L_1$ : Pre-transition taper to push mode down.
- $L_2$ : P-cladding taper.
- $L_3$ : Core taper (un-pumped QD).

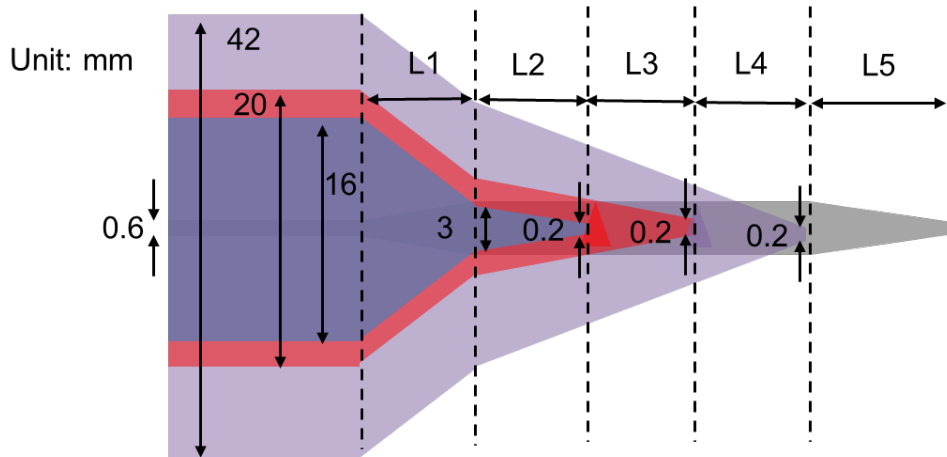


Figure 5.2. III-V-on-Si taper design for optical transitions.

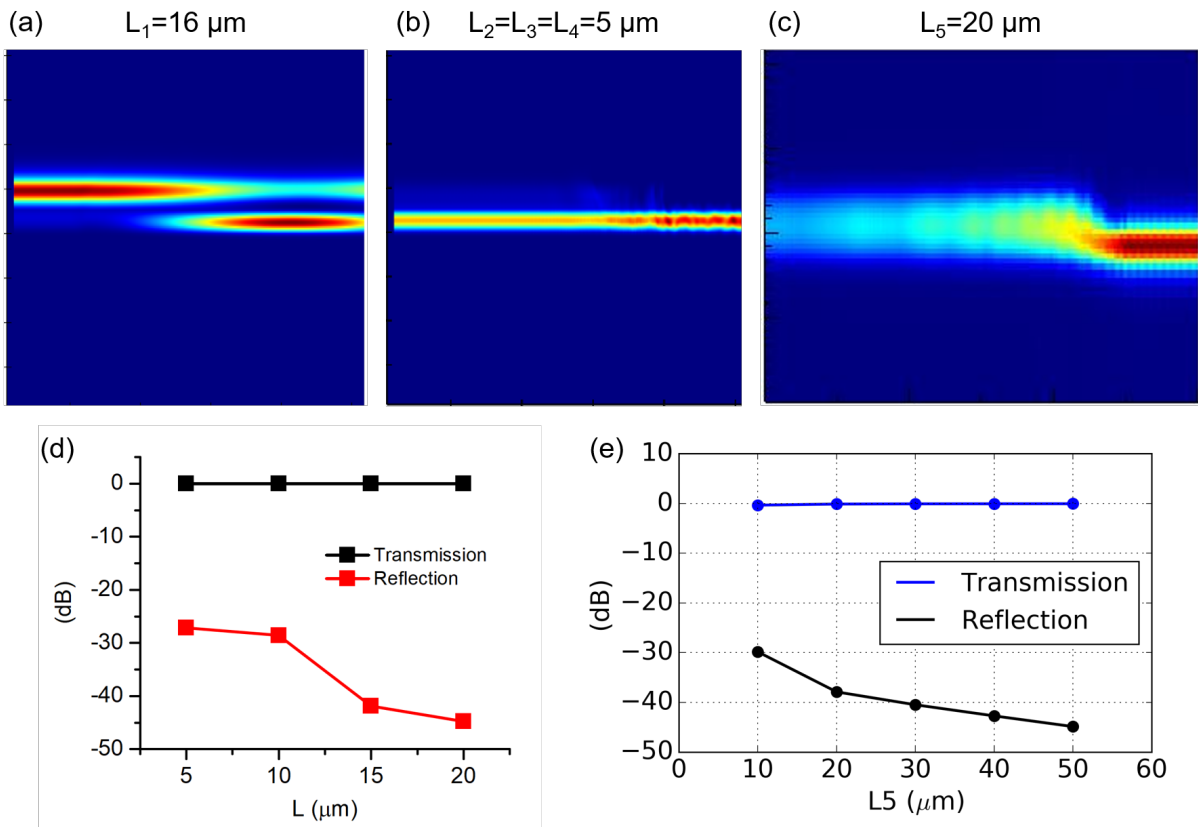


Figure 5.3. (a) Simulated 3D mode profile of the  $L_1$  taper with the length of  $16 \mu\text{m}$ . (b) Simulated 3D mode profile of the  $L_2 + L_3 + L_4$  tapers with the length of  $5 \mu\text{m}$ . (c) Simulated 3D mode profile of the  $L_5$  taper with the length of  $20 \mu\text{m}$ . (d) Simulated transmission and reflection of the  $L_2 + L_3 + L_4$  tapers with different lengths. (e) Simulated transmission and reflection of the  $L_5$  taper with different lengths.

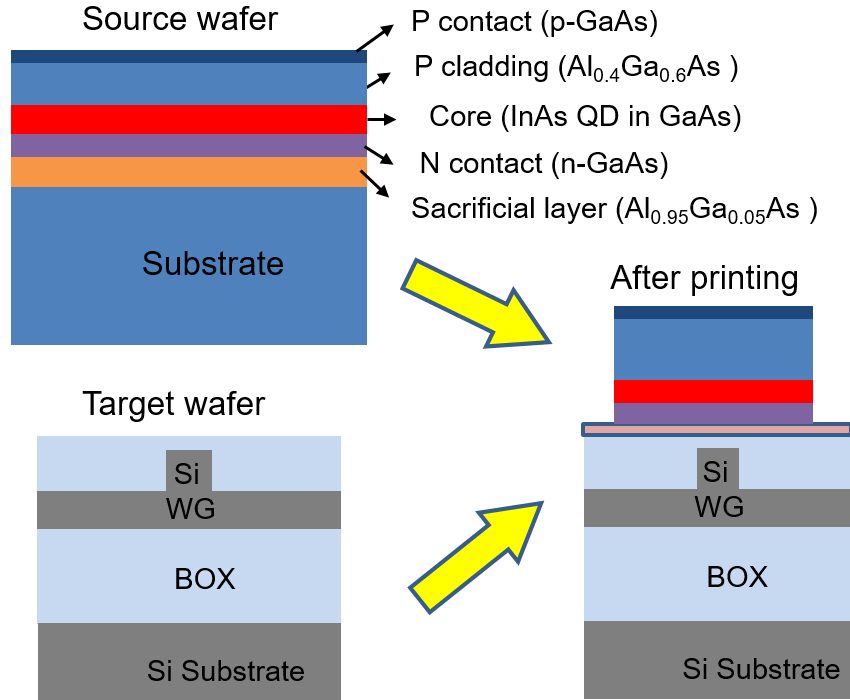


Figure 5.4. Schematic of the transfer printing process.

$L_4$ : N-contact taper (highly doped n-GaAs).

$L_5$ : 500-nm-thick 220-nm-slab Si rib waveguide to 220-nm-thick Si ridge waveguide taper.

Figure 5.3(a), (b), and (c) shows the simulated 3D mode profile of the  $L_1$  taper,  $L_2 + L_3 + L_4$  tapers, and  $L_5$  taper, respectively. For  $L_1$  taper, the optical mode can be pushed down with -52 dB reflection and 0.0002 dB insertion loss in  $L_1 = 16 \mu\text{m}$ . Figure 5.3(d) shows the simulated transmission and reflection of the  $L_2 + L_3 + L_4$  tapers with different lengths. It can be seen that the reflection can be  $< -40$  dB and the insertion loss can be  $< 0.005$ , with the length of  $L_2=L_3=L_4=15 \mu\text{m}$ . Figure 5.3(e) shows the simulated transmission and reflection of the  $L_5$  taper. With  $L_2 > 20 \mu\text{m}$ , the reflection can be  $< -38$  dB and the insertion loss can be  $< 0.116$  dB.

### 5.1.2 Transfer Printing Fabrication Process

The transfer printing process started with the preparation of target and source wafers. Secondly, the sacrificial layer on the source wafer was wet-etched so that the photoresist tethers hold the coupons. Then the coupons were picked up and printed on the source

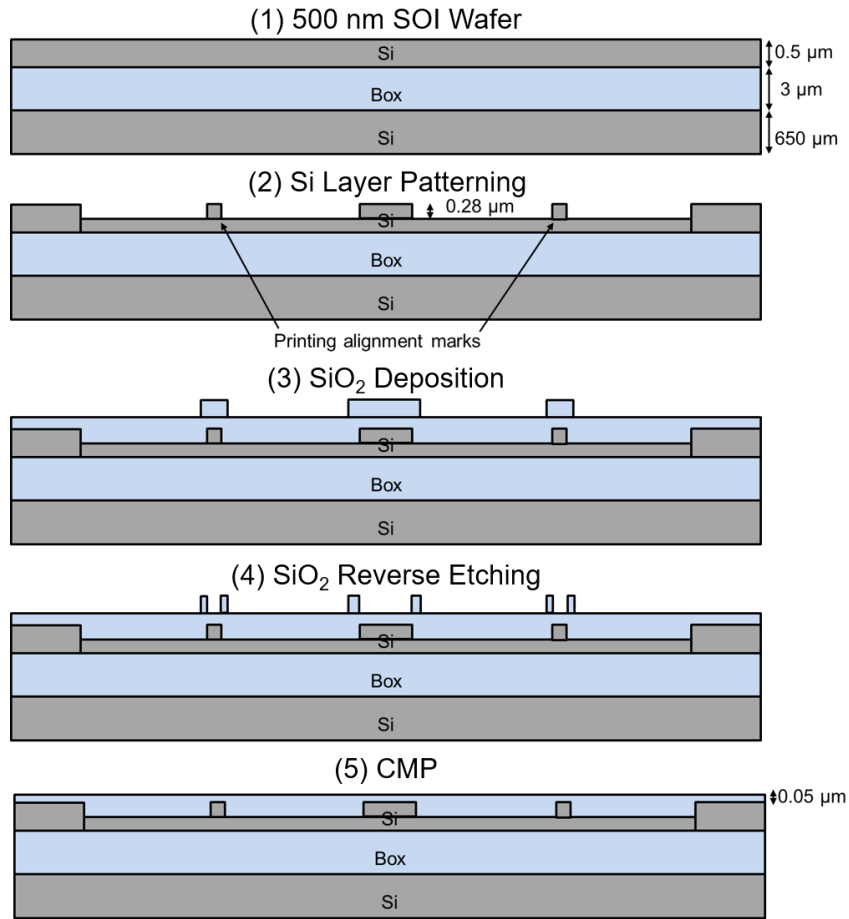


Figure 5.5. Fabrication flow charts for the transfer printing target wafer.

wafer with resin, as shown in Figure 5.4.

### 5.1.2.1 Target Wafer Fabrication

The target wafer fabrication was started with a 500-nm SOI wafer with 3- $\mu\text{m}$ -thick buried oxide as shown in Figure 5.5. Firstly, the Si layer was defined by deep-UV projection lithography and partially etched by ICP etching. Then a 385-nm-thick LTO was deposited by LPCVD. Following the  $\text{SiO}_2$  reverse etching by ICP etching, the  $\text{SiO}_2$  cladding on top of the Si waveguide was planarized to 50 nm by CMP.

### 5.1.2.2 Source Wafer Fabrication

Figure 5.6 shows the layer stacks of 3-inch 1310-nm GaAs QD epitaxial wafers from Innolume GmbH. The epitaxial wafer is grown on a GaAs substrate. An etch stop layer with high Al content ( $\text{Al}_{0.95}\text{Ga}_{0.05}\text{As}$ ) is grown for the sacrificial wet etching. Then n-

Layer	Material	Thickness (nm)	Doping (/cm <sup>3</sup> )	Comments
Substrate	GaAs	---	uid	
Etch stop	Al <sub>0.95</sub> Ga <sub>0.05</sub> As	500	uid	High Al content for etch stop
n-contact	GaAs	100	Si: 5.0E+18	Heavily n-doped for n contact
n-clad	Al <sub>0.4</sub> Ga <sub>0.6</sub> As	100	Si: 1.0E+18	Lightly n-doped
Core (QD > 9x)	InAs QD in GaAs	~400	C: 0.1E+18 (? Innolume will decide)	9 stacks to have sufficient optical gain (dope in GaAs, not in InAs)
p-clad	Al <sub>0.4</sub> Ga <sub>0.6</sub> As	250	C: 0.25E+18	Graded doping to balance loss and serial resistance
p-clad	Al <sub>0.4</sub> Ga <sub>0.6</sub> As	250	C: 0.5E+18	Graded doping to balance loss and serial resistance
p-clad	Al <sub>0.4</sub> Ga <sub>0.6</sub> As	1300	C: 1.0E+18	Graded doping to balance loss and serial resistance
P-digitized gradient	GaAs/Al <sub>0.4</sub> Ga <sub>0.6</sub> As	10	C: 1.0E+18	Digitally graded composition of GaAs and Al <sub>0.4</sub> Ga <sub>0.6</sub> As
p-contact	GaAs	200	C: 5.0E+19	Heavily p-doped for p contact

Growth Direction ↓

Figure 5.6. Layer stacks of the QD epitaxial wafer from Innolume GmbH.

contact, n-cladding, core, p-cladding, and p-contact layers are grown sequentially. The core layer (active region) is with 10-nm InAs QDs suspended in 400-nm-thick GaAs.

Figure 5.7 shows the fabrication flow charts for the transfer printing source wafer. The fabrication started with the E-beam evaporation and lift-off of the 100-nm-thick Au alignment mark. Then a 50-nm-thick SiO<sub>2</sub> hardmask is deposition by HDPCVD and patterned by contact lithography and ICP etching using CF<sub>4</sub> and CHF<sub>3</sub> gases. With SiO<sub>2</sub> hardmask, the mesa area was etched by ICP with Cl<sub>2</sub>, BCl<sub>3</sub>, and Ar gases. The mesa etching stopped above the *n*-GaAs layer to isolate the *n*- and *p*-contact layers to the wet etchant in the sacrificial wet etching step. After removing the SiO<sub>2</sub> hardmask using ICP, a 180-nm-thick SiO<sub>2</sub> encapsulation layer was deposited by HDPCVD. Then the pedestal layer was patterned by contact lithography and two steps (SiO<sub>2</sub> and III-V) of ICP etching. The pedestal was over-etched into the GaAs substrate so that the tethers were adhesive to the substrate after sacrificial layer wet etching. Following the coating of a 3.5- $\mu$ m-thick photoresist, the 2- $\mu$ m-wide tethers were patterned by contact lithography. Lastly, the sacrificial layer was wet etched with the following steps [96]:



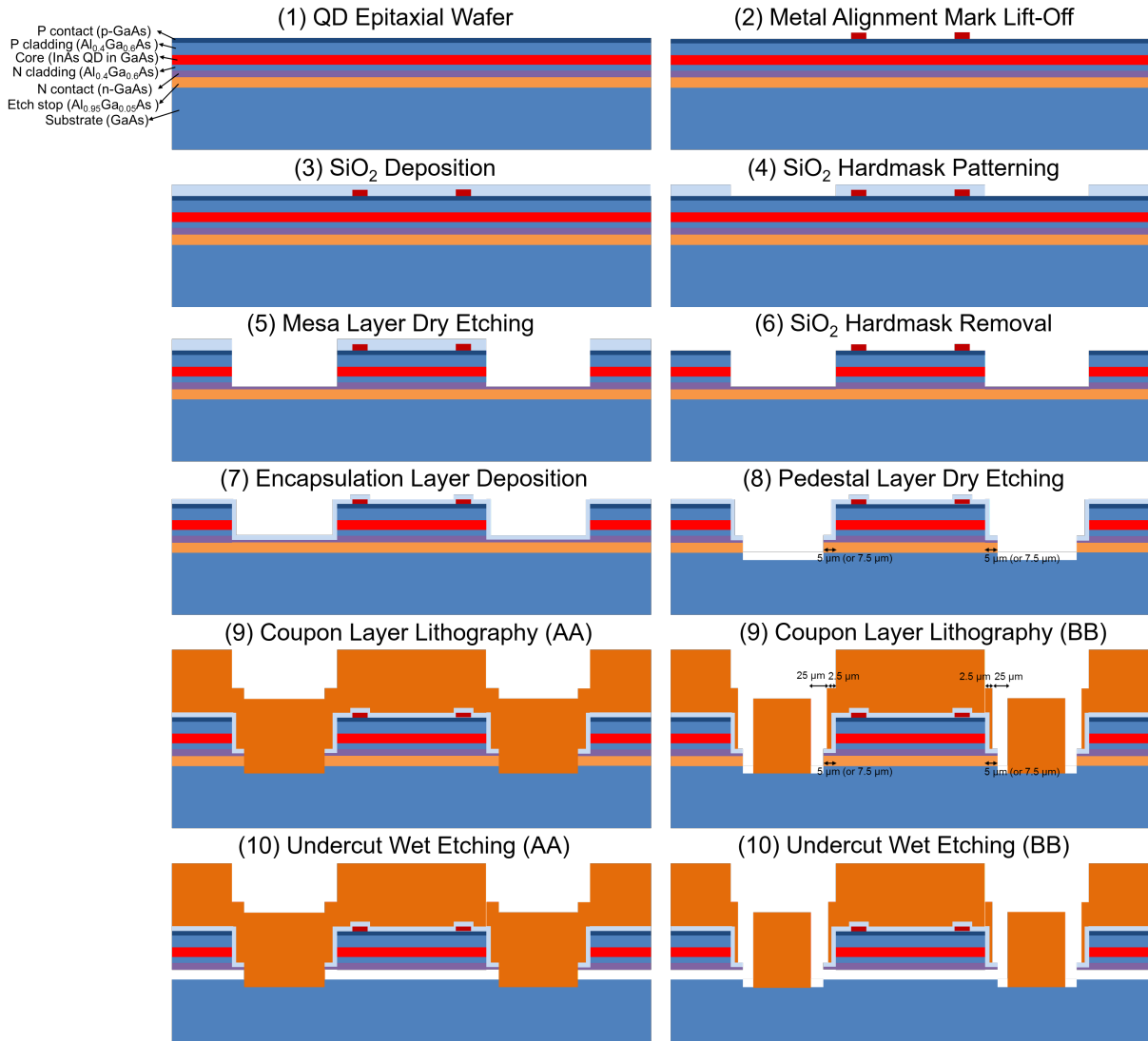


Figure 5.7. Fabrication flow charts for the transfer printing source wafer.

1.  $\text{NH}_4\text{OH}:\text{DI}$  (1:9) for 30s to prevent delaminating
2. DI water dunking
3.  $\text{HCl}:\text{DI}$  (1:1) for 1 hour, heated to  $25^\circ$  and with mild agitation
4. DI water bath to remove acid residue
5. Air dry

The fabrication quality of the source wafer highly affects the yield of the transfer printing. Here are some tips for the source wafer preparation:

1. The p-contact surface needs to be cleaned using acetone, methanol, isopropanol,

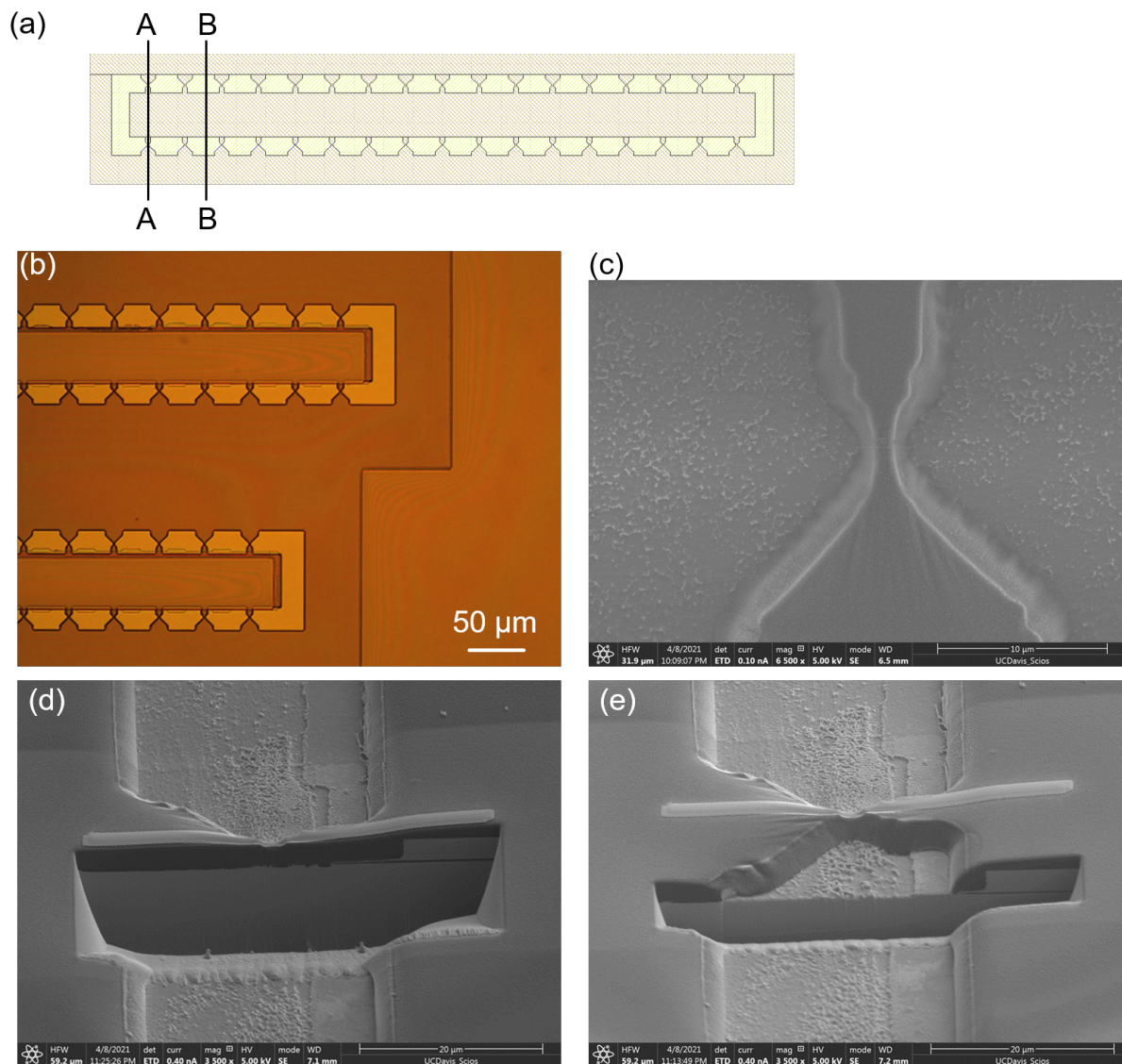


Figure 5.8. (a) Coupon layer photomask showing the AA (tether) and BB areas. (b) Microscope image of the source wafer after coupon layer lithography. (c) SEM image of the tether. (d) FIB-SEM image of the AA area after sacrificial layer wet etching. (e) FIB-SEM image of the BB area after sacrificial layer wet etching.

concentrated sulfuric acid (30s), and DI water. The surface cleaning can increase the adhesiveness of the tether photoresist so that the coupon will be easier to pick up.

2. The sacrificial layer needs to be completely etched. Otherwise, the coupon would not be able to be picked up, or the coupon can break during the picking up or bonding. Additionally, suppose the sacrificial layer is partially etched. In that case, the coupon may not stick to the target wafer since the smoothness of the bottom of the source wafer

directly impacts the bonding quality. Thus the controlling of the wet etching time need to be carefully calibrated.

3. The photoresist tether lithography needs to be optimized. On the one hand, the tether must be strong enough to hold the coupon during the sacrificial layer etching. On the other hand, the tether also needs to be sufficiently brittle for the stamp to pick the coupon from the source wafer.

Figure 5.8(a) shows the Coupon layer photomask indicating the AA (tether) and BB areas. Figure 5.8(b) shows the microscope image of the source wafer after coupon layer lithography before the sacrificial layer wet etching. Figure 5.8(c) shows the scanning electron microscopy (SEM) image of the photoresist tether. Figure 5.8(d) and (e) shows the focused ion beam scanning electron microscopy (FIB-SEM) images of the AA and BB areas after sacrificial layer wet etching, respectively.

## 5.2 Optical in-Phase-Quadrature Modulator Based on Injection-Locked VCSEL Phase Array

Directly modulated vertical cavity surface emitting lasers (VCSELs) are promising light sources for short-reach communications in supercomputers and datacenters because of their low cost, low power consumption, compact size, capability of large-scale fabrication, and simple fiber coupling [97]. Moreover, such approach do not require additional modulators and the low modulation voltage swings can be directly provided by CMOS. However, to apply VCSELs in long-haul high-capacity transmissions, two obstacles need to be overcome which are frequency chirping and limited modulation bandwidth. Furthermore, lack of independent amplitude modulation (AM) and phase modulation (PM) makes VCSELs infeasible to produce high-spectral-efficiency modulation formats such as quadrature amplitude modulation (QAM).

For long-haul high-capacity transmission, QAM signals are normally generated by free-running laser sources combined with external in-phase-quadrature (IQ) modulators to control the full field of optical waveform and shape the spectra [98–100]. One of the most important features of IQ modulators is the ability to map electrical signals onto

the optical carriers without creating additional frequencies which broaden the spectra (i.e. chirp). This is achieved by putting a pair of phase modulators in a Mach-Zehnder interferometer with push-pull operation so that the chirp from the two phase modulators can be cancelled. However, the state-of-the-art phase modulators based on lithium niobate [101], silicon [102, 103], indium phosphide [104, 105], plasmonics [106, 107], polymer [108, 109], organic [110] are mostly long and require large drive voltages ( $> 1V$ ).

Fortunately, optical injection locking (OIL) has been proved to be an effective method to suppress the frequency chirping and enhance the modulation bandwidth of directly modulated VCSELs by transferring not only wavelength but also intrinsic linewidth, frequency and phase stability of the ‘master’ lasers to the ‘slave’ lasers [111–116]. More importantly, by proper tuning the OIL parameters, nearly pure AM as well as pure PM can be attained which pave the way for building IQ modulators using VCSELs [117, 118]. The AM and PM behavior of OIL VCSELs have been characterized with the RF frequency of kilohertz to gigahertz [117]. However, higher RF frequency range still needs to be investigated in order to meet the capacity demand of modern data transmission systems.

In this section, we have demonstrated an IQ modulator for QAM signal synthesis based on OIL VCSEL phase array. Firstly, the AM and PM of OIL VCSELs up to tens of gigahertz are experimentally measured using high-speed digital-to-analog converter (DAC) and coherent receiver. Nearly pure PM with depth of 0.55 rad is achieved by properly controlling the bias current, injection ratio, and wavelength detuning. By placing integrated OIL VCSEL phase array in a reflective multi-arm interferometer which is formed by liquid crystal on silicon (LCoS), 20-GBd BPSK signals are generated with a peak-to-peak modulation voltage of 400 mV and the measured OSNR sensitivity exhibits good performance. Additionally, Nyquist pulse shaped waveforms are used to demonstrate the spectrum shaping capability of our modulator. Optical spectra with flat top are observed at 10, 30, and 40 GBd indicating chirp-free operation. Our approach maintains all the benefits of directly driven VCSELs and inspires a novel method for dense array integration of cost-effective, high-performance flat IQ modulators.

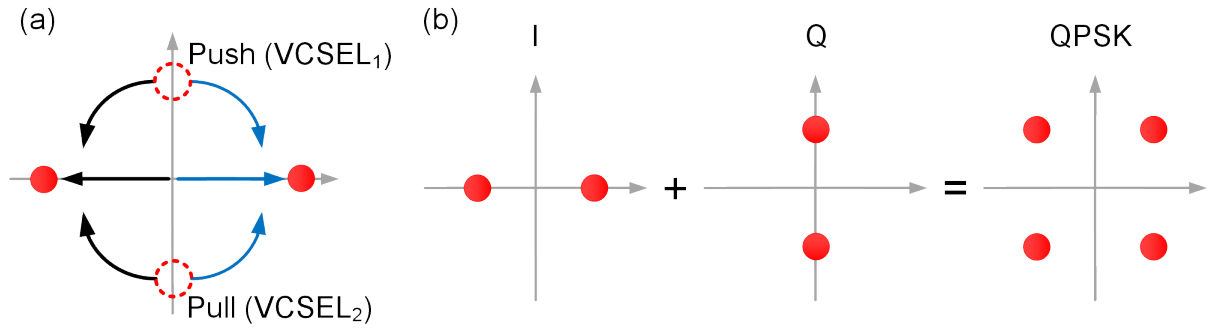


Figure 5.9. Principle of our proposed IQ modulator shown by constellation map: (a) push-pull operation of two phase modulated OIL VCSELs as one quadrature. (b) Two quadratures are perpendicularly superposed to build IQ modulator.

### 5.2.1 Principle and Experimental Setup

The principle of our proposed IQ modulator is sketched in Figure 5.9 through constellation map [119]. The PM of OIL lasers is theoretically limited to less than  $\pi$  rad, which is not enough for binary ( $0, \pi$  rad) optical PSK modulation. Thus two phase modulated OIL VCSELs are push-pull operated and their outputs are coherently superposed with 180-degree phase difference to form one quadrature of the IQ (Figure 5.9(a)). The mutual coherence between the VCSELs is guaranteed by using the same master laser. Then, another two VCSELs can similarly form the other quadrature and the two quadratures are perpendicularly superposed to obtain QPSK (Figure 5.9(b)). Such method can be easily adapted to arbitrary QAM signals, for example using 4-level RF signal to drive each OIL VCSELs can generate 16-QAM. Alternatively, two amplitude modulated OIL VCSELs with 90-degree shift and carrier suppression can be another approach to build IQ modulator [120, 121].

A monolithically integrated long-wavelength VCSEL array from Vertilas is wire-bonded to do RF fan-out (Figure 5.10). The VCSELs are spaced with a pitch of  $250 \mu\text{m}$  and the technical details can be found in [122]. Each of the VCSEL are biased by two DC sources and directly modulated by the outputs of a 60 GS/s DAC. The DC and RF signals are combined by a 50-GHz bias tee. The schematic and microscope image of the experimental setup are shown in Figure 5.10(b) and (c). As a proof-of-principle experiment, two slave VCSELs are injection-locked by an external cavity laser (ECL) which has a linewidth of

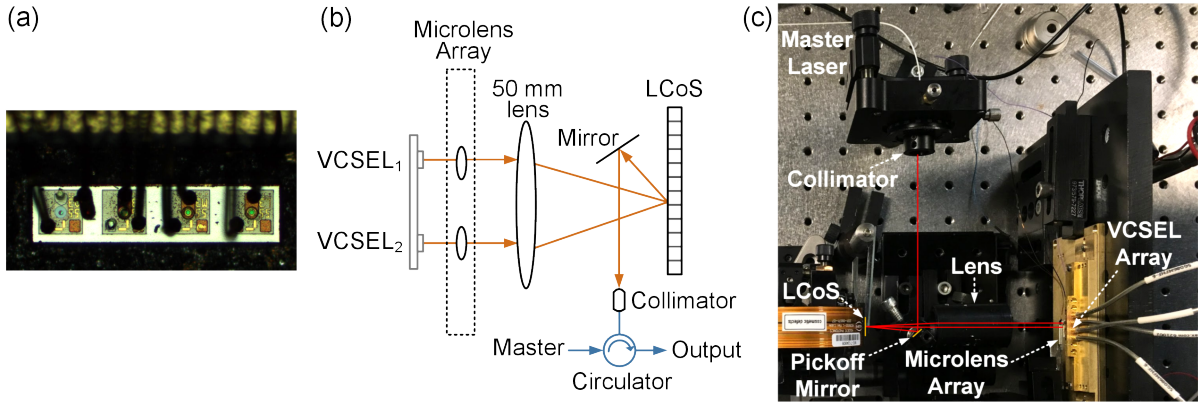


Figure 5.10. (a) Microscope image of packaged VCSEL array. (b) Schematic of experimental setup. (c) Microscope image of experimental setup.

30 kHz and inputs through a circulator. The optical outputs of the VCSELs are firstly collimated by a microlens array, then focused on the LCoS by a 50-mm lens, and lastly reflected into a fiber collimator. The LCoS acts as a beam splitting hologram which can evenly split the incident master beam and combine the outgoing slave beams with full control of their relative phase. Alternatively, a planar lightwave circuit (PLC) can be directly coupled to the VCSELs to provide the beam splitting functionality. The coupling loss from the VCSELs to the fiber collimator is 3 dB which includes 1.5-dB loss from the LCoS. The complex output optical field is captured by a 160-GS/s single-polarization digital coherent receiver with a different ECL as the local oscillator (LO).

## 5.2.2 OIL VCSEL Properties: Bandwidth Enhancement, AM and PM

The voltage, emitting power and wavelength of free-running VCSEL are firstly measured with different bias current, as shown in Figure 5.11. The static resistance is  $50 \Omega$  and the wavelength difference of the VCSELs under identical bias current is within 0.3 nm which indicates the uniformity across the array. It can be seen from Figure 5.11(b) that there is a linear region for the P-I curve and the rollover occurs at 16 mA.

There are mainly three tuning parameters for OIL VCSELs which are bias current, injection ratio, and detuning wavelength. The injection ratio denotes the ratio of the incident power from the master laser to the emitting power of free-running slave VCSEL.

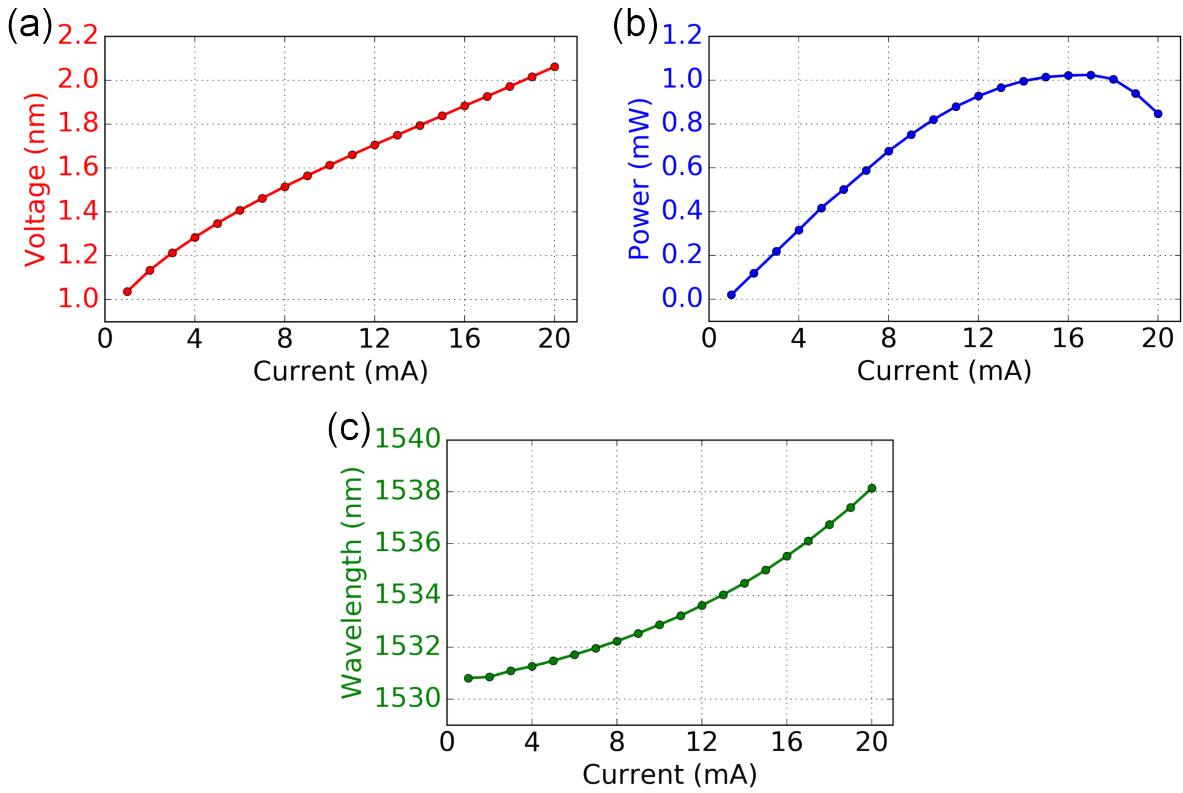


Figure 5.11. (a) Voltage (b) optical power (c) wavelength of free-running VCSEL with varied bias current.

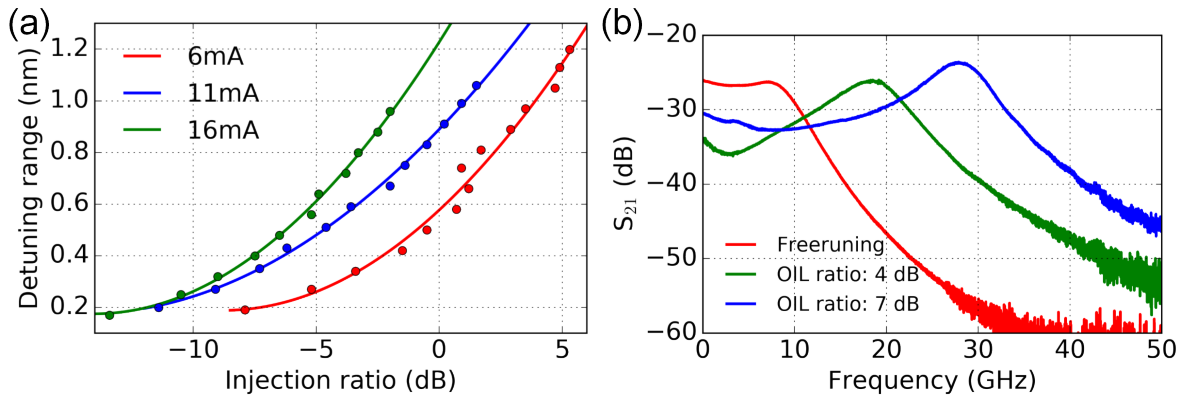


Figure 5.12. (a) Detuning range of injection-locked VCSEL with varied injection ratio and indicated bias current. (b) Small-signal electrical to optical frequency response ( $S_{21}$ ) under 6-mA bias and the indicated injection ratio.

The detuning wavelength represents the wavelength difference between master laser and slave VCSEL. As the master laser red shifts, there is an edge that the locking status of the slave VCSEL is switched to unlocking status and the corresponding detuning wavelength

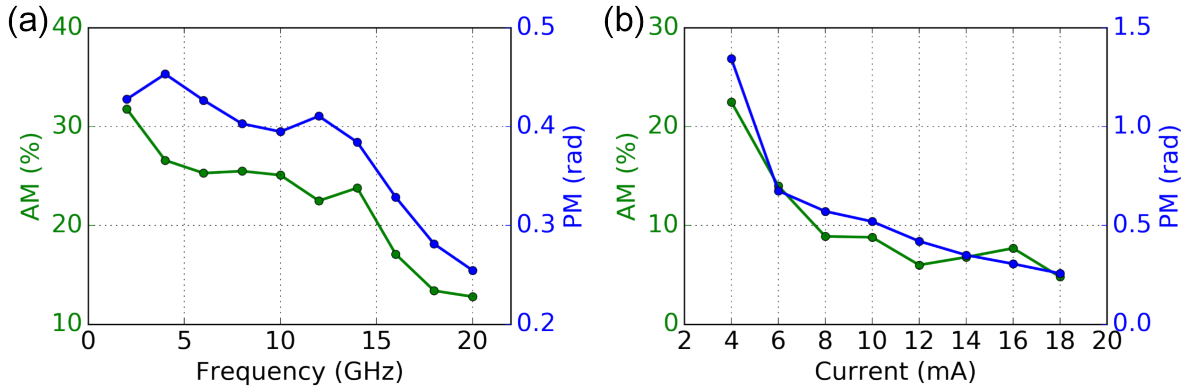


Figure 5.13. (a) Variation of AM and PM with RF frequency between 2 GHz and 20 GHz. Bias current 6 mA, injection ratio -8 dB, wavelength detuning 0.21 nm. (b) Variation of AM and PM with bias current from 4 mA to 18 mA. Injection ratio -2 dB, detuning wavelength 0.16 nm, RF frequency 10 GHz.

is notated as detuning range. Figure 5.12(a) shows the detuning range of OIL VCSELs under different bias current and injection ratio. It is observed that higher injection ratio as well as lower bias current result in larger detuning range.

The small-signal frequency response of VCSEL with and without injection locking are measured using a 50-GHz vector network analyzer (VNA). The VCSEL is probed using a 40-GHz RF probe and the 6-mA bias current is delivered by the bias tee inside the VNA. The wavelength of master laser is fixed at the same as free-running VCSEL so that no wavelength detuning is added. The modulation bandwidth of OIL VCSEL can increase approximately as square root of the injection ratio due to enhancement of the relaxation resonance [123]. For instance, the modulation bandwidth increase from 10 GHz to over 30 GHz at 7-dB injection ratio as shown in Figure 5.12(b). Additionally, the frequency roll-up caused by relaxation oscillation can act as built-in pre-emphasis to help compensate for the strong frequency roll-off observed in electronic components such as DAC. It needs to be mentioned that, there is a bandwidth to modulation depth trade-off when practically choosing the injection ratio value since higher injection ratio not only increase the modulation bandwidth but also reduces the modulation depth.

The large-signal amplitude and phase response of OIL VCSELs are described by AM index and PM index which equals the peak variation of AM divided by the mean and the peak excursion of the PM in radians, respectively. Figure 5.13(a) shows the measured



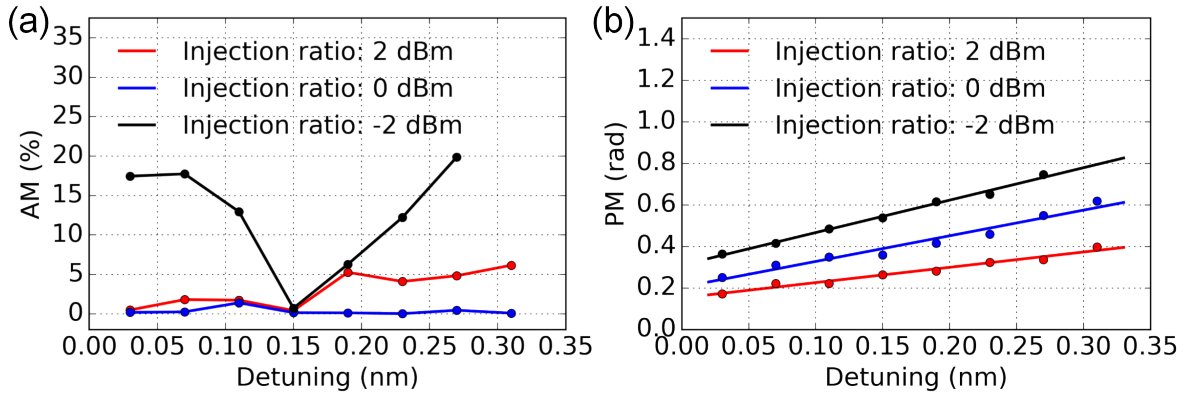


Figure 5.14. Variation of (a) AM and (b) PM with detuning wavelength at 6-mA bias current. The RF frequency is 10 GHz.

AM and PM by scanning the RF frequency from 2 GHz to 20 GHz. The drive signal is a PRBS with length of  $2^{10}-1$  and peak-to-peak voltage of 400 mV. The bias current, injection ratio, and wavelength detuning are fixed at 6 mA, -8 dB, and 0.21 nm, respectively. In such frequency region, temperature modulation is suppressed and electronic modulation dominates so that both AM and PM drop with increasing frequency due to bandwidth limitation. Moreover, the bias current is also scanned from the beginning of the linear region (4 mA) to post rollover point (18 mA) on the P-I curve, as shown in Figure 5.13(b). The injection ratio, detuning wavelength, RF frequency are fixed at -2 dB, 0.16 nm and 10 GHz, respectively. It can be seen that higher bias current results in lower AM and PM so that we choose 6 mA to obtain higher modulation depth.

Setting bias current as 6 mA, the other two tuning parameters (injection ratio and detuning wavelength) are scanned to attain pure PM. It is shown that lower injection ratio and higher detuning wavelength result in higher PM (Figure 5.14(a)). On the other hand, with injection ratio of -2 dB and wavelength detuning of 0.15 nm, the AM reaches the minimum value and the modulation is closest to pure PM (Figure 5.14(b)). Such set of tuning parameters is used to build our proposed IQ modulator.

### 5.2.3 Demonstration of IQ Modulator

Figure 5.15 shows the large-signal modulation response of a single VCSEL which exhibits nearly pure PM modulation with proper tuning. The RF signal is  $2^{10}-1$  PRBS with 400

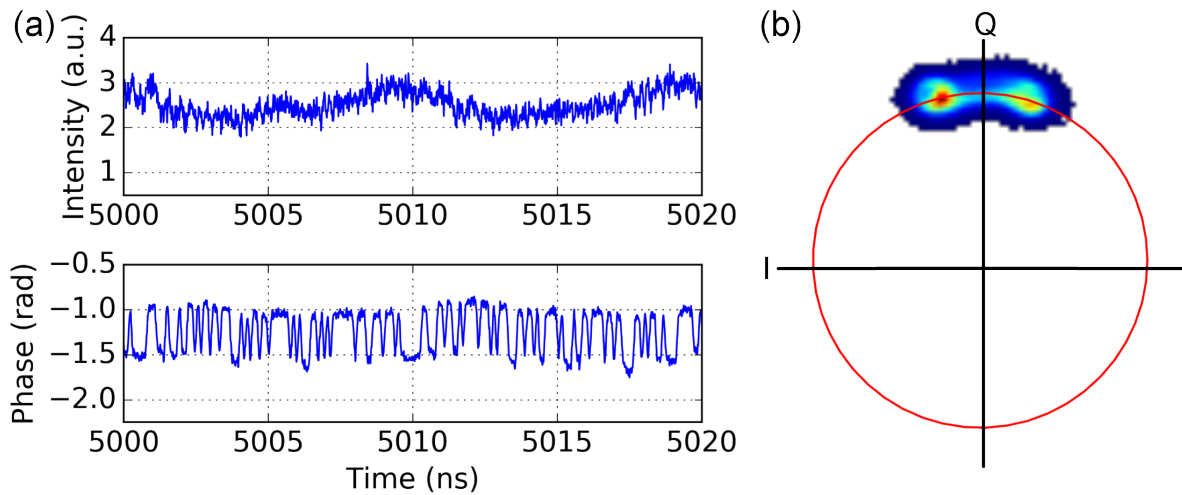


Figure 5.15. Large-signal modulation response of a single VCSEL for  $2^{10}-1$  PRBS with 400 mV  $V_{pp}$  drive: (a) AM and PM time plots (b) IQ histogram of the optical field. Bias current, injection ratio, wavelength detuning are 6 mA, -2 dB, and 0.15 nm, respectively.

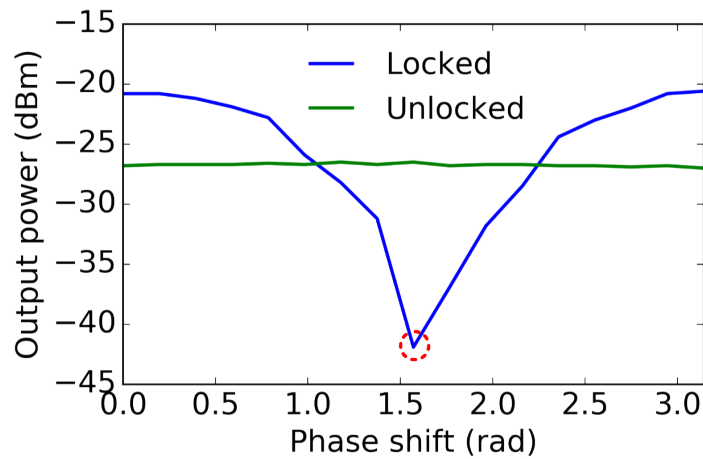


Figure 5.16. Single-quadrature modulator output power versus phase shift between the two VCSELs.

mV peak-to-peak voltage. The bias current, injection ratio, and wavelength detuning are 6 mA, -2 dB, and 0.15 nm, respectively. For the amplitude and phase time plot as shown in Figure 5.15(a), 500 waveforms are averaged together to improve the signal to noise ratio (SNR). At the same time, the IQ histogram of the optical field is plotted on a logarithmic scale without averaging (Figure 5.15(b)). The AM index and PM index are 0.007 and 0.55 rad, respectively, which can act as the building blocks of the IQ modulator.

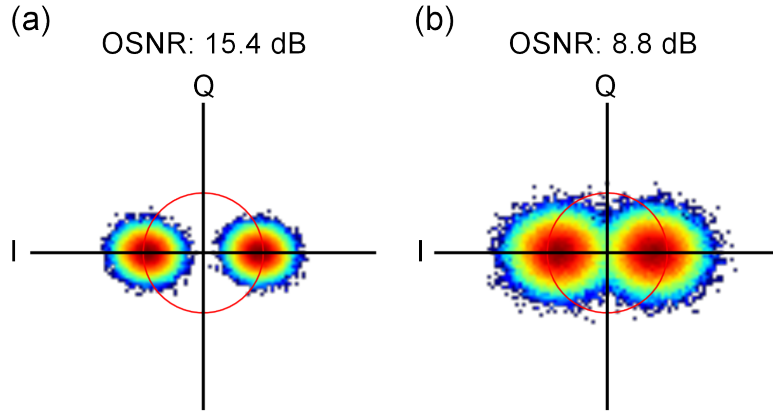


Figure 5.17. Constellation of 20-GBd BPSK after decision directed equalizer and sampling with OSNR of: (a) 15.4 dB (b) 8.8 dB.

Figure 5.16 shows the output power of the two VCSELs combined by the LCoS with varied phase shift between them. Without OIL, the free-running VCSELs are incoherently added so that the output power is constant. With OIL, the output power changes with phase shift in a sinusoidal shape indicating that the two output beams of VCSELs are coherently superposed. The optimal phase-shifting condition is the null point (red dashed circle in Figure 5.16) providing 180-degree phase difference which is necessary for push-pull operation.

After fine tuning the parameters, 20-GBd BPSK signal is generated by push-pull operation of the two VCSELs. The RF signals are two complimentary  $2^{10}-1$  PRBS produced by DAC with peak-to-peak voltage of 400 mV which is compatible with CMOS circuitry. The bias currents are 5.8 mA and 6 mA, injection ratios are 4.6 dB and 2.5 dB, and detuning wavelengths are 0.26 nm and 0.37 nm, respectively. The difference in the tuning parameters are due to fabrication variation of the VCSELs. Figure 5.17 plots the BPSK constellations of the optical field after a decision directed adaptive equalizer to sample the waveform at the baud rate. The optical field locates in one quadrature indicating chirp-free operation.

Figure 5.18 shows the BER as a function of OSNR for the 20-GBd BPSK signal. The BER at OSNR of 15.4 dB and 8.8 dB are  $2.3 \times 10^{-5}$  and  $1.7 \times 10^{-2}$ , respectively. The theoretical BER curve (dashed red curve) is also plotted for comparison. It can be seen that there is about 6.7 dB implementation penalty at BER of  $1 \times 10^{-3}$ . The relatively

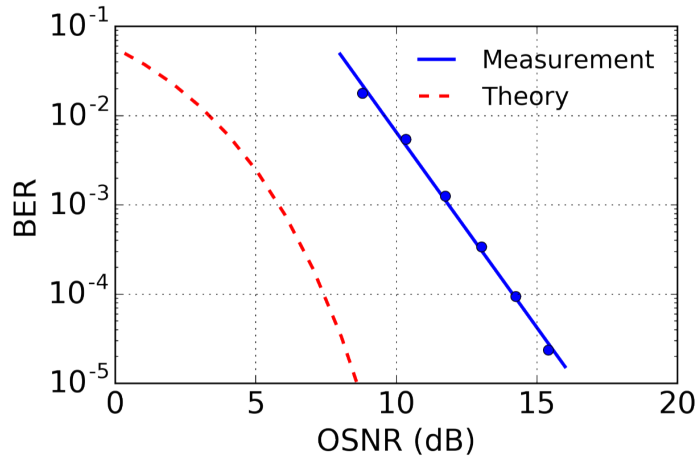


Figure 5.18. OSNR sensitivity of our synthesized 20-GBd BPSK.

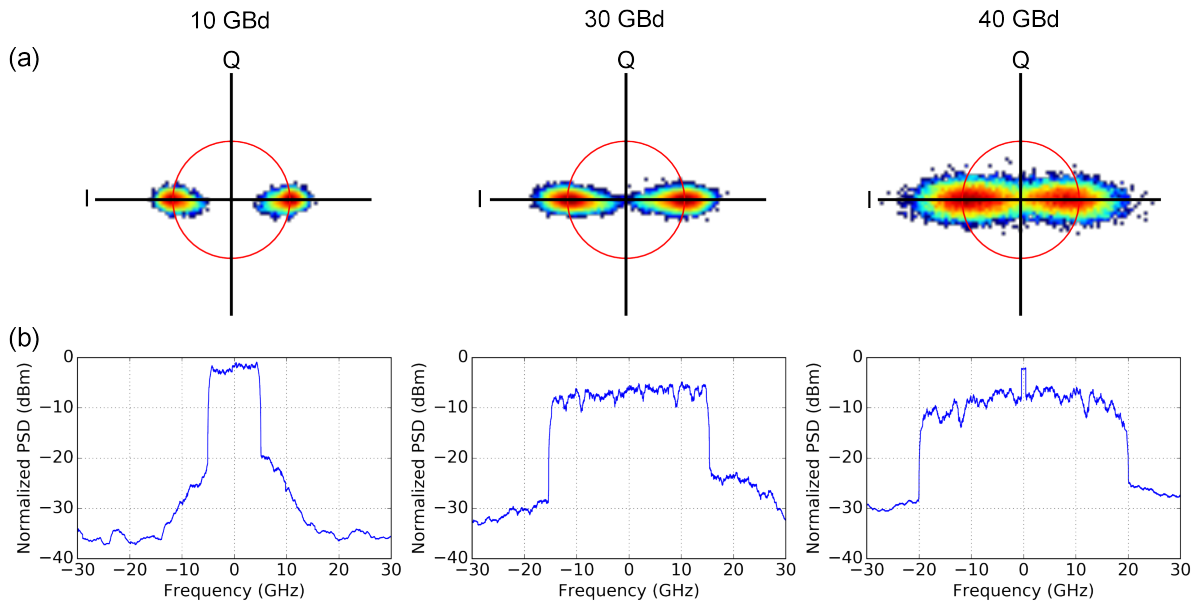


Figure 5.19. Chirp-free modulation driven by Nyquist pulse shaped waveforms at 10, 30, and 40 GBd: (a) constellation after decision directed equalizer and sampling. (b) optical spectrum.

large penalty is due to the fact that no application-specific optimization is done for the free-space elements and transmitter components in this proof-of-principle experiment.

To evaluate the capability of spectra shaping, two complimentary Nyquist pulse shaped waveforms at 10, 30, and 40 GBd with pattern length of  $2^{10}-1$  are programmed by DAC and used to drive the two VCSELs in our push-pull modulator. The Nyquist shaping uses root raised cosine (RRC) filter with roll-off factor of nearly 0 to obtain an almost ideal

rectangular spectrum. The peak-to-peak voltages of the shaped waveforms are 400 mV. Figure 5.19(a) shows the recovered BPSK constellation after a decision directed adaptive equalizer. The modulation bandwidth enhancement of OIL enables the generation of 30 GBd and 40 GBd signals even though of the intrinsic bandwidth of free-running VCSEL are 15 GHz and 20 GHz, respectively. Figure 5.19(b) shows the optical spectrum with flat top and suppressed carriers indicating chirp-free operation of the modulator.

## Appendix: The Silvaco Source Code for the PIN Junction Simulations

```
go athena
# Define the mesh
line x loc=-5.10 spac=0.2
line x loc=-1.10 spac=0.02
line x loc=-0.20 spac=0.01
line x loc=0.20 spac=0.01
line x loc=1.10 spac=0.02
line x loc=5.10 spac=0.2
line y loc=0.00 spac=0.005
line y loc=0.15 spac=0.005
line y loc=0.22 spac=0.005
# Define initial substrate
init silicon c.boron=1.0e15 orientation=100 two.d
save outf=PIN1280_initial.str
# Si WG etching
etch silicon start x=-5.10 y=0.00
etch cont x=-5.10 y=0.15
etch cont x=-0.20 y=0.15
etch done x=-0.20 y=0.00
etch silicon start x=5.10 y=0.00
etch cont x=5.10 y=0.15
etch cont x=0.20 y=0.15
etch done x=0.20 y=0.00
save outf=PIN1280_after_Si_etching.str
# Barrier deposition (P++)
deposit barrier thick=0.2
# Barrier etching (P++)
```

```

etch barrier start x=-5.10 y=0.15
etch cont x=-1.10 y=0.15
etch cont x=-1.10 y=-0.05
etch done x=-5.10 y=-0.05
save outf=PIN1280_before_P++_implantation.str
# P++ implantation
implant boron dose=9e14 energy=6 tilt=0 rotation=0 crystal
# Remove barrier (P+)
etch barrier all
save outf=PIN1280_after_P++_implantation.str
# Barrier deposition (N++)
deposit barrier thick=0.2
# Barrier etching (N++)
etch barrier start x=5.10 y=0.15
etch cont x=1.10 y=0.15
etch cont x=1.10 y=-0.05
etch done x=5.10 y=-0.05
save outf=PIN1280_before_N++_implantation.str
# N++ implantation
implant phosphor dose=1e15 energy=23 tilt=0 rotation=0 crystal
# Remove barrier (N+)
etch barrier all
save outf=PIN1280_after_N++_implantation.str
# RTA annealing
diffus time=30 seconds temp=1050 nitro
save outf=PIN1280_after_RTA.str
# Metal deposition and etching
#deposit aluminum thick=0.1
#etch aluminum start x=-1.20 y=0.15

```

```

#etch cont x=1.20 y=0.15
#etch cont x=1.20 y=-0.10
#etch done x=-1.20 y=-0.10
#save outf=PIN1280_after_metal_etching.str
# Electrode
#electrode name=anode x=-3.70
#electrode name=cathode x=3.70
#save outf=PIN1280_after_electrode.str
go atlas
# Define electrode
electrode name=anode x.min=-4.1 x.max=-2.1 y.min=0.15 y.max=0.15
electrode name=cathode x.min=2.1 x.max=4.1 y.min=0.15 y.max=0.15
# Define contact
contact name=anode
contact name=cathode
# Define Interface
interface qf=3e10
# Specify physical models
models bipolar print
impact selb
# Define method
method gummel newton
# Define output
#output j.electron j.hole
# 0V bias
solve init
save outf=pin1280_0V.str
# 0.25V bias
solve prev

```



```

solve vanode=0.25 name = anode
save outf=pin1280_0p25V.str
# 0.5V bias
solve prev
solve vanode=0.5 name = anode
save outf=pin1280_0p5V.str
# 0.75V bias
solve prev
solve vanode=0.75 name = anode
save outf=pin1280_0p75V.str
# 1V bias
solve prev
solve vanode=1 name = anode
save outf=pin1280_1V.str
# 1.25V bias
solve prev
solve vanode=1.25 name = anode
save outf=pin1280_1p25V.str
# 1.5V bias
solve prev
solve vanode=1.5 name = anode
save outf=pin1280_1p5V.str
extract init inf=pin1280_1p5V.str
extract name="iv" curve(v."anode", i."anode") outfile="pin1280_IV_1p5V.dat"
# 2V bias
solve prev
solve vanode=2 name = anode
save outf=pin1280_2V.str
extract init inf=pin1280_2V.str

```

```
extract name="iv" curve(v."anode", i."anode") outfile="pin1280_IV_2V.dat"  
quit
```

## REFERENCES

- [1] G. E. Moore, “Cramming more components onto integrated circuits,” 1965.
- [2] J. M. Shalf and R. Leland, “Computing beyond Moore’s Law,” *Computer*, vol. 48, no. 12, pp. 14–23, 2015.
- [3] J. Shalf, “The future of computing beyond Moore’s law,” *Philosophical Transactions of the Royal Society A*, vol. 378, no. 2166, p. 20190061, 2020.
- [4] M. Smit, J. Van der Tol, and M. Hill, “Moore’s law in photonics,” *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 1–13, 2012.
- [5] A. Wonfor, H. Wang, R. Penty, and I. White, “Large Port Count High-Speed Optical Switch Fabric for Use Within Datacenters [Invited],” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, no. 8, pp. A32–A39, 2011.
- [6] L. Lu, S. Zhao, L. Zhou, D. Li, Z. Li, M. Wang, X. Li, and J. Chen, “16 × 16 non-blocking silicon optical switch based on electro-optic Mach-Zehnder interferometers,” *Optics Express*, vol. 24, no. 9, p. 9295, 2016. [Online]. Available: <https://www.osapublishing.org/abstract.cfm?URI=oe-24-9-9295>
- [7] S. Zhao, L. Lu, L. Zhou, D. Li, Z. Guo, and J. Chen, “16×16 silicon Mach-Zehnder interferometer switch actuated with waveguide microheaters,” *Photon. Res.*, vol. 4, no. 5, pp. 202–207, oct 2016. [Online]. Available: <http://www.osapublishing.org/prj/abstract.cfm?URI=prj-4-5-202>
- [8] Q. Cheng, L. Y. Dai, N. C. Abrams, Y.-H. Hung, P. E. Morrissey, M. Glick, P. O’Brien, and K. Bergman, “Ultralow-crosstalk, strictly non-blocking microring-based optical switch,” *Photon. Res.*, vol. 7, no. 2, pp. 155–161, feb 2019.
- [9] Q. Cheng, M. Bahadori, Y. Hung, Y. Huang, N. Abrams, and K. Bergman, “Scalable Microring-Based Silicon Clos Switch Fabric With Switch-and-Select Stages,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–11, 2019.
- [10] S. Han, T. J. Seok, N. Quack, B.-W. Yoo, and M. C. Wu, “Large-scale silicon photonic switches with movable directional couplers,” *Optica*, vol. 2, no. 4, pp. 370–375, apr 2015.

- [11] T. J. Seok, N. Quack, S. Han, R. S. Muller, and M. C. Wu, “Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers,” *Optica*, vol. 3, no. 1, pp. 64–70, 2016. [Online]. Available: <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-3-1-64>
- [12] S. Han, T. J. Seok, K. Yu, N. Quack, R. S. Muller, and M. C. Wu, “Large-Scale Polarization-Insensitive Silicon Photonic MEMS Switches,” *Journal of Lightwave Technology*, vol. 36, no. 10, pp. 1824–1830, 2018.
- [13] R. Yu, S. Cheung, Y. Li, K. Okamoto, R. Proietti, Y. Yin, and S. J. B. Yoo, “A scalable silicon photonic chip-scale optical switch for high performance computing systems,” *Opt. Express*, vol. 21, no. 26, pp. 32 655–32 667, dec 2013.
- [14] R. Stabile, A. Rohit, and K. A. Williams, “Monolithically Integrated 8 × 8 Space and Wavelength Selective Cross-Connect,” *Journal of Lightwave Technology*, vol. 32, no. 2, pp. 201–207, 2014.
- [15] T. J. Seok, J. Luo, Z. Huang, K. Kwon, J. Henriksson, J. Jacobs, L. Ochikubo, R. S. Muller, and M. C. Wu, “Silicon photonic wavelength cross-connect with integrated MEMS switching,” *APL Photonics*, vol. 4, no. 10, p. 100803, 2019.
- [16] A. S. P. Khope, M. Saeidi, R. Yu, X. Wu, A. M. Netherton, Y. Liu, Z. Zhang, Y. Xia, G. Fleeman, A. Spott, S. Pinna, C. Schow, R. Helkey, L. Theogarajan, R. C. Alferness, A. A. M. Saleh, and J. E. Bowers, “Multi-wavelength selective crossbar switch,” *Opt. Express*, vol. 27, no. 4, pp. 5203–5216, feb 2019.
- [17] M. V. DeBole, B. Taba, A. Amir, F. Akopyan, A. Andreopoulos, W. P. Risk, J. Kunitz, C. O. Otero, T. K. Nayak, R. Appuswamy, P. J. Carlson, A. S. Cassidy, P. Datta, S. K. Esser, G. J. Garreau, K. L. Holland, S. Lekuch, M. Mastro, J. McKinstry, C. di Nolfo, B. Paulovicks, J. Sawada, K. Schleupen, B. G. Shaw, J. L. Klamo, M. D. Flickner, J. V. Arthur, and D. S. Modha, “TrueNorth: Accelerating From Zero to 64 Million Neurons in 10 Years,” *Computer*, vol. 52, no. 5, pp. 20–29, 2019.
- [18] M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. K. Lin, A. Lines, R. Liu, D. Mathaikutty,

- S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.
- [19] C. Ramey, "Silicon Photonics for Artificial Intelligence Acceleration : HotChips 32," in *2020 IEEE Hot Chips 32 Symposium (HCS)*, 2020, pp. 1–26.
- [20] M. A. Taubenblatt, "Optical Interconnects for High-Performance Computing," *Journal of Lightwave Technology*, vol. 30, no. 4, pp. 448–457, 2012.
- [21] A. Benner, "Optical interconnect opportunities in supercomputers and high end computing," in *OFC/NFOEC*, 2012, pp. 1–60.
- [22] M. A. Taubenblatt, "Optical Interconnects for Large Scale Computing Systems: Trends and Challenges," in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPCom, SOF)*, ser. OSA Technical Digest (online). Zurich: Optical Society of America, 2018, p. NeTh3F.2. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=Networks-2018-NeTh3F.2>
- [23] Y. H. Ezzeldin, M. Karmoose, and C. Fragouli, "Communication vs distributed computation: An alternative trade-off curve," in *2017 IEEE Information Theory Workshop (ITW)*, 2017, pp. 279–283.
- [24] M. Ghaffari, "Distributed MIS via All-to-All Communication," in *Proceedings of the ACM Symposium on Principles of Distributed Computing*, ser. PODC '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 141–149.
- [25] H. Kwon, A. Samajdar, and T. Krishna, "A Communication-Centric Approach for Designing Flexible DNN Accelerators," *IEEE Micro*, vol. 38, no. 6, pp. 25–35, 2018.
- [26] K.-C. J. Chen, M. Ebrahimi, T.-Y. Wang, and Y.-C. Yang, "NoC-Based DNN Accelerator: A Future Design Paradigm," in *Proceedings of the 13th IEEE/ACM International Symposium on Networks-on-Chip*, ser. NOCS '19. New York, NY, USA: Association for Computing Machinery, 2019.
- [27] H. Takahashi, K. Oda, and H. Toba, "Impact of crosstalk in an arrayed-waveguide multiplexer on  $N \times N$  optical interconnection," *Journal of Lightwave Technology*, vol. 14, no. 6, pp. 1097–1105, 1996.

- [28] P. Grani, R. Proietti, S. Cheung, and S. J. B. Yoo, “Flat-Topology High-Throughput Compute Node With AWGR-Based Optical-Interconnects,” *Journal of Lightwave Technology*, vol. 34, no. 12, pp. 2959–2968, 2016.
- [29] Y. Zhang, X. Xiao, K. Zhang, S. Li, A. Samanta, Y. Zhang, K. Shang, R. Proietti, K. Okamoto, and S. J. B. Yoo, “Foundry-Enabled Scalable All-to-All Optical Interconnects Using Silicon Nitride Arrayed Waveguide Router Interposers and Silicon Photonic Transceivers,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–9, 2019.
- [30] A. Bianco, D. Cuda, R. Gaudino, G. Gavilanes, F. Neri, and M. Petracca, “Scalability of Optical Interconnects Based on Microring Resonators,” *IEEE Photonics Technology Letters*, vol. 22, no. 15, pp. 1081–1083, 2010.
- [31] X. Xiao, R. Proietti, and S. J. B. Yoo, “Scalability of microring-based crossbar for all-to-all optical interconnects,” in *2017 IEEE Optical Interconnects Conference (OI)*, 2017, pp. 39–40.
- [32] L. Zhou, S. S. Djordjevic, R. Proietti, D. Ding, S. J. B. Yoo, R. Amirtharajah, and V. Akella, “Design and evaluation of an arbitration-free passive optical crossbar for on-chip interconnection networks,” *Applied Physics A*, vol. 95, no. 4, pp. 1111–1118, 2009.
- [33] B. G. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, “All-Optical Comb Switch for Multiwavelength Message Routing in Silicon Photonic Networks,” *IEEE Photonics Technology Letters*, vol. 20, no. 10, pp. 767–769, 2008.
- [34] J. R. Ong, R. Kumar, and S. Mookherjea, “Ultra-High-Contrast and Tunable-Bandwidth Filter Using Cascaded High-Order Silicon Microring Filters,” *IEEE Photonics Technology Letters*, vol. 25, no. 16, pp. 1543–1546, 2013.
- [35] J. Wang, Z. Sheng, L. Li, A. Pang, A. Wu, W. Li, X. Wang, S. Zou, M. Qi, and F. Gan, “Low-loss and low-crosstalk  $8 \times 8$  silicon nanowire AWG routers fabricated with CMOS technology,” *Optics Express*, vol. 22, no. 8, pp. 9395–9403, 2014. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-22-8-9395>
- [36] A. Rahim, E. Ryckeboer, A. Z. Subramanian, S. Clemmen, B. Kuyken, A. Dhakal,

- A. Raza, A. Hermans, M. Muneeb, S. Dhoore, Y. Li, U. Dave, P. Bienstman, N. Le Thomas, G. Roelkens, D. Van Thourhout, P. Helin, S. Severi, X. Rottenberg, and R. Baets, “Expanding the Silicon Photonics Portfolio With Silicon Nitride Photonic Integrated Circuits,” *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 639–649, 2017. [Online]. Available: <http://jlt.osa.org/abstract.cfm?URI=jlt-35-4-639>
- [37] D. Dai, Z. Wang, J. F. Bauters, M.-C. Tien, M. J. R. Heck, D. J. Blumenthal, and J. E. Bowers, “Low-loss  $\text{Si}_3\text{N}_4$  arrayed-waveguide grating (de)multiplexer using nano-core optical waveguides,” *Optics Express*, vol. 19, no. 15, pp. 14 130–14 136, 2011. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-19-15-14130>
- [38] M. Piels, J. F. Bauters, M. L. Davenport, M. J. R. Heck, and J. E. Bowers, “Low-Loss Silicon Nitride AWG Demultiplexer Heterogeneously Integrated With Hybrid III-V/Silicon Photodetectors,” *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 817–823, 2014.
- [39] A. Arbabi and L. L. Goddard, “Measurements of the refractive indices and thermo-optic coefficients of  $\text{Si}_3\text{N}_4$  and  $\text{SiO}_x$  using microring resonances,” *Optics Letters*, vol. 38, no. 19, pp. 3878–3881, 2013. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-38-19-3878>
- [40] D. Dai, J. Bauters, and J. E. Bowers, “Passive technologies for future large-scale photonic integrated circuits on silicon: polarization handling, light non-reciprocity and loss reduction,” *Light: Science & Applications*, vol. 1, p. e1, mar 2012. [Online]. Available: <https://doi.org/10.1038/lisa.2012.1> <http://10.0.4.14/lisa.2012.1>
- [41] K. Shang, S. Pathak, C. Qin, and S. J. B. Yoo, “Low-Loss Compact Silicon Nitride Arrayed Waveguide Gratings for Photonic Integrated Circuits,” *IEEE Photonics Journal*, vol. 9, no. 5, pp. 1–5, 2017.
- [42] H. Li, Z. Xuan, A. Titriku, C. Li, K. Yu, B. Wang, A. Shafik, N. Qi, Y. Liu, R. Ding, T. Baehr-Jones, M. Fiorentino, M. Hochberg, S. Palermo, and P. Y. Chiang, “A 25 Gb/s, 4.4 V-Swing, AC-Coupled Ring Modulator-Based WDM Transmitter with Wavelength Stabilization in 65 nm CMOS,” *IEEE Journal of Solid-State Circuits*,

- vol. 50, no. 12, pp. 3145–3159, dec 2015.
- [43] K. Yu, C. Li, H. Li, A. Titriku, A. Shafik, B. Wang, Z. Wang, R. Bai, C. Chen, M. Fiorentino, P. Y. Chiang, and S. Palermo, “A 25 Gb/s Hybrid-Integrated Silicon Photonic Source-Synchronous Receiver With Microring Wavelength Stabilization,” *IEEE Journal of Solid-State Circuits*, vol. 51, no. 9, pp. 2129–2141, 2016.
- [44] X. Xiao, Y. Zhang, R. Proietti, and S. J. B. Yoo, “Scalable AWGR-based All-to-All Optical Interconnects with 2.5D/3D Integrated Optical Interposers,” in *2018 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, 2018, pp. 161–162.
- [45] P. Dong, W. Qian, H. Liang, R. Shafiq, D. Feng, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and M. Asghari, “Thermally tunable silicon racetrack resonators with ultralow tuning power,” *Opt. Express*, vol. 18, no. 19, pp. 20 298–20 304, sep 2010. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-18-19-20298>
- [46] E. L. Goldstein, L. Eskildsen, and A. F. Elrefaie, “Performance implications of component crosstalk in transparent lightwave networks,” *IEEE Photonics Technology Letters*, vol. 6, no. 5, pp. 657–660, 1994.
- [47] I.-F. Jang and S.-L. Lee, “Simple approaches of wavelength registration for monolithically integrated DWDM laser arrays,” *IEEE Photonics Technology Letters*, vol. 14, no. 12, pp. 1659–1661, 2002.
- [48] R. Proietti, Y. Yin, R. Yu, X. Ye, C. Nitta, V. Akella, and S. J. B. Yoo, “All-Optical Physical Layer NACK in AWGR-Based Optical Interconnects,” *IEEE Photonics Technology Letters*, vol. 24, no. 5, pp. 410–412, 2012.
- [49] Y. Yin, R. Proietti, X. Ye, C. J. Nitta, V. Akella, and S. J. B. Yoo, “LIONS: An AWGR-Based Low-Latency Optical Switch for High-Performance Computing and Data Centers,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 2, p. 3600409, 2013.
- [50] Z. Cao, R. Proietti, and S. J. B. Yoo, “HALL: A hierarchical all-to-all optical interconnect architecture,” in *2014 Optical Interconnects Conference*, 2014, pp. 73–74.



- [51] J. Chen, Y. Gong, M. Fiorani, and S. Aleksic, “Optical interconnects at the top of the rack for energy-efficient data centers,” *IEEE Communications Magazine*, vol. 53, no. 8, pp. 140–148, 2015.
- [52] Z. Cao, R. Proietti, M. Clements, and S. J. B. Yoo, “Experimental Demonstration of Flexible Bandwidth Optical Data Center Core Network With All-to-All Interconnectivity,” *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1578–1585, 2015.
- [53] H. Tsuda, “Large-scale arrayed-waveguide grating based photonic router using T- and O-band,” in *2016 IEEE 6th International Conference on Photonics (ICP)*, 2016, pp. 1–3.
- [54] J. Müller, J. Hauck, A. Moscoso-Mártir, N. Chimot, S. Romero-García, B. Shen, F. Merget, F. Lelarge, and J. Witzens, “High speed WDM interconnect using silicon photonics ring modulators and mode-locked laser,” in *2015 European Conference on Optical Communication (ECOC)*, 2015, pp. 1–3.
- [55] H. K. Kim and S. Chandrasekhar, “Dependence of coherent crosstalk penalty on the OSNR of the signal,” in *Optical Fiber Communication Conference. Technical Digest Postconference Edition. Trends in Optics and Photonics Vol.37 (IEEE Cat. No. 00CH37079)*, vol. 2, 2000, pp. 359–361 vol.2.
- [56] K. Wen, P. Samadi, S. Rumley, C. P. Chen, Y. Shen, M. Bahadori, K. Bergman, and J. Wilke, “Flexfly: Enabling a Reconfigurable Dragonfly through Silicon Photonics,” in *SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2016, pp. 166–177.
- [57] X. Xiao, R. Proietti, G. Liu, H. Lu, P. Fotouhi, S. Werner, Y. Zhang, and S. J. B. Yoo, “Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 2, pp. 1–10, 2020.
- [58] R. Proietti, Z. Cao, C. J. Nitta, Y. Li, and S. J. B. Yoo, “A Scalable, Low-Latency, High-Throughput, Optical Interconnect Architecture Based on Arrayed Waveguide Grating Routers,” *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 911–920, 2015. [Online]. Available: <http://jlt.osa.org/abstract.cfm?URI=jlt-33-4-911>

- [59] P. Grani, G. Liu, R. Proietti, and S. J. B. Yoo, “Bit-parallel all-to-all and flexible AWGR-based optical interconnects,” in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.
- [60] X. Xiao, R. Proietti, G. Liu, H. Lu, Y. Zhang, and S. J. B. Yoo, “Multi-FSR Silicon Photonic Flex-LIONS Module for Bandwidth-Reconfigurable All-To-All Optical Interconnects,” *Journal of Lightwave Technology*, vol. 38, no. 12, pp. 3200–3208, 2020.
- [61] K. Shang, S. Pathak, B. Guan, G. Liu, and S. J. B. Yoo, “Low-loss compact multilayer silicon nitride platform for 3D photonic integrated circuits,” *Opt. Express*, vol. 23, no. 16, pp. 21 334–21 342, aug 2015.
- [62] S. Cheung, T. Su, K. Okamoto, and S. J. B. Yoo, “Ultra-Compact Silicon Photonic 512 512 25 GHz Arrayed Waveguide Grating Router,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 4, pp. 310–316, 2014.
- [63] L. B. Soldano and E. C. M. Pennings, “Optical multi-mode interference devices based on self-imaging: principles and applications,” *Journal of Lightwave Technology*, vol. 13, no. 4, pp. 615–627, 1995.
- [64] W. D. Sacher, Y. Huang, G. Lo, and J. K. S. Poon, “Multilayer Silicon Nitride-on-Silicon Integrated Photonic Platforms and Devices,” *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 901–910, 2015.
- [65] K. Shang, S. Pathak, G. Liu, S. Feng, S. Li, W. Lai, and S. J. B. Yoo, “Silicon nitride tri-layer vertical Y-junction and 3D couplers with arbitrary splitting ratio for photonic integrated circuits,” *Opt. Express*, vol. 25, no. 9, pp. 10 474–10 483, may 2017.
- [66] T. Barwicz, M. A. Popovic, P. T. Rakich, M. R. Watts, H. A. Haus, E. P. Ippen, and H. I. Smith, “Microring-resonator-based add-drop filters in SiN: fabrication and analysis,” *Optics Express*, vol. 12, no. 7, pp. 1437–1442, 2004. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-12-7-1437>
- [67] D. Dai and H. Wu, “Realization of a compact polarization splitter-rotator on silicon,” *Opt. Lett.*, vol. 41, no. 10, pp. 2346–2349, may 2016. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-41-10-2346>

- [68] Y. Yin, R. Proietti, C. J. Nitta, V. Akella, C. Mineo, S. J. B. Yoo, and K. Wen, “AWGR-based all-to-all optical interconnects using limited number of wavelengths,” in *2013 Optical Interconnects Conference*, 2013, pp. 47–48.
- [69] X. Xiao, R. Proietti, K. Zhang, G. Liu, H. Lu, J. Messig, and S. J. B. Yoo, “Experimental Demonstration of 64-Port Thin-CLOS Architecture for All-to-All Optical Interconnects,” in *2018 Conference on Lasers and Electro-Optics (CLEO)*, 2018, pp. 1–2.
- [70] R. Proietti, X. Xiao, K. Zhang, G. Liu, H. Lu, P. Fotouhi, J. Messig, and S. J. B. Yoo, “Experimental demonstration of a 64-port wavelength routing thin-CLOS system for data center switching architectures,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 7, pp. 49–57, 2018.
- [71] X. Xiao, R. Proietti, S. Werner, P. Fotouhi, and S. J. B. Yoo, “Flex-LIONS: A Scalable Silicon Photonic Bandwidth-Reconfigurable Optical Switch Fabric,” in *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, 2019, pp. 1–3.
- [72] X. Xiao, R. Proietti, G. Liu, H. Lu, Y. Zhang, and S. J. Ben Yoo, “Experimental Demonstration of SiPh Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects,” in *2019 European Conference on Optical Communication (ECOC)*, 2019, pp. 1–4.
- [73] R. Proietti, X. Xiao, M. Fariborz, P. Fotouhi, Y. Zhang, and B. SJ, “Flex-LIONS: A Silicon Photonic Bandwidth-Reconfigurable Optical Switch Fabric,” *IEICE Transactions on Communications*, vol. 103, no. 11, pp. 1190–1198, 2020.
- [74] N. Dupuis and B. G. Lee, “Impact of Topology on the Scalability of MachZehnder-Based Multistage Silicon Photonic Switch Networks,” *Journal of Lightwave Technology*, vol. 36, no. 3, pp. 763–772, 2018.
- [75] C. Wang, M. Zhang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer, and M. Lončar, “Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages,” *Nature*, vol. 562, no. 7725, pp. 101–104, 2018. [Online]. Available: <https://doi.org/10.1038/s41586-018-0551-y>

- [76] P. Pintus, M. Hofbauer, C. L. Manganelli, M. Fournier, S. Gundavarapu, O. Lemonnier, F. Gambini, L. Adelmini, C. Meinhart, C. Kopp, F. Testa, H. Zimmermann, and C. J. Oton, “PWM-Driven Thermally Tunable Silicon Microring Resonators: Design, Fabrication, and Characterization,” *Laser & Photonics Reviews*, vol. 13, no. 9, p. 1800275, sep 2019. [Online]. Available: <https://doi.org/10.1002/lpor.201800275>
- [77] M. R. Watts, W. A. Zortman, D. C. Trotter, G. N. Nielson, D. L. Luck, and R. W. Young, “Adiabatic Resonant Microrings (ARMs) with directly integrated thermal microphotronics,” in *2009 Conference on Lasers and Electro-Optics and 2009 Conference on Quantum electronics and Laser Science Conference*, 2009, pp. 1–2.
- [78] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, 2016. [Online]. Available: <https://doi.org/10.1038/nature16961>
- [79] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, “Deep learning with coherent nanophotonic circuits,” *Nature Photonics*, vol. 11, no. 7, pp. 441–446, 2017. [Online]. Available: <https://doi.org/10.1038/nphoton.2017.93>
- [80] J. Ba and R. Caruana, “Do deep nets really need to be deep?” in *Advances in neural information processing systems*, 2014, pp. 2654–2662.
- [81] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, “Optimal design for universal multiport interferometers,” *Optica*, vol. 3, no. 12, pp. 1460–1465, 2016. [Online]. Available: <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-3-12-1460>
- [82] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, “Experimental realization of any discrete unitary operator,” *Physical Review Letters*, vol. 73, no. 1, pp. 58–61, jul 1994. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.73.58>

- [83] D. A. B. Miller, “Self-configuring universal linear optical component [Invited],” *Photonics Research*, vol. 1, no. 1, pp. 1–15, 2013. [Online]. Available: <http://www.osapublishing.org/prj/abstract.cfm?URI=prj-1-1-1>
- [84] S. Hamann, A. Ceballos, J. Landry, and O. Solgaard, “High-speed random access optical scanning using a linear MEMS phased array,” *Optics Letters*, vol. 43, no. 21, pp. 5455–5458, 2018. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-43-21-5455>
- [85] M. Miscuglio and V. J. Sorger, “Photonic tensor cores for machine learning,” *Applied Physics Reviews*, vol. 7, no. 3, p. 031404, jul 2020. [Online]. Available: <https://doi.org/10.1063/5.0001942>
- [86] X. Xiao and S. J. B. Yoo, “Tensor-Train Decomposed Synaptic Interconnections for Compact and Scalable Photonic Neural Networks,” in *2020 IEEE Photonics Conference (IPC)*, 2020, pp. 1–2.
- [87] —, “Scalable and Compact 3D Tensorized Photonic Neural Networks,” in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.
- [88] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O’Brien, and A. Laing, “Universal linear optics,” *Science*, vol. 349, no. 6249, pp. 711 LP – 716, aug 2015. [Online]. Available: <http://science.sciencemag.org/content/349/6249/711.abstract>
- [89] I. V. Oseledets, “Tensor-train decomposition,” *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [90] A. Novikov, D. Podoprikin, A. Osokin, and D. P. Vetrov, “Tensorizing neural networks,” in *Advances in neural information processing systems*, 2015, pp. 442–450.
- [91] C. Hawkins and Z. Zhang, “Bayesian tensorized neural networks with automatic rank selection,” *arXiv preprint arXiv:1905.10478*, 2019.
- [92] Y. Zhang, Y. Ling, Y. Zhang, K. Shang, and S. J. B. Yoo, “High-Density Wafer-Scale 3-D Silicon-Photonic Integrated Circuits,” *IEEE Journal of Selected Topics*

- in *Quantum Electronics*, vol. 24, no. 6, pp. 1–10, 2018.
- [93] M. B. On, Y.-J. Lee, X. Xiao, R. Proietti, and S. J. Ben Yoo, “Analysis of the Hardware Imprecisions for Scalable and Compact Photonic Tensorized Neural Networks,” in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.
- [94] D. Liang and J. E. Bowers, “Recent Progress in Heterogeneous III-V-on-Silicon Photonic Integration,” *Light: Advanced Manufacturing*, vol. 2, no. 1, pp. 1–25, 2021. [Online]. Available: <http://www.light-am.com//article/id/67462e78-bcbe-4bdc-89a6-26beb867fe7a>
- [95] J. Justice, C. Bower, M. Meitl, M. B. Mooney, M. A. Gubbins, and B. Corbett, “Wafer-scale integration of group IIIIV lasers on silicon using transfer printing of epitaxial layers,” *Nature Photonics*, vol. 6, no. 9, pp. 610–614, 2012. [Online]. Available: <https://doi.org/10.1038/nphoton.2012.204>
- [96] A. R. Clawson, “Guide to references on IIIIV semiconductor chemical etching,” *Materials Science and Engineering: R: Reports*, vol. 31, no. 1, pp. 1–438, 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0927796X00000279>
- [97] J. A. Tatum, D. Gazula, L. A. Graham, J. K. Guenter, R. H. Johnson, J. King, C. Kocot, G. D. Landry, I. Lyubomirsky, A. N. MacInnes, E. M. Shaw, K. Balemarthy, R. Shubochkin, D. Vaidya, M. Yan, and F. Tang, “VCSEL-Based Interconnects for Current and Future Data Centers,” *Journal of Lightwave Technology*, vol. 33, no. 4, pp. 727–732, 2015.
- [98] A. Chiba, T. Sakamoto, T. Kawanishi, K. Higuma, M. Sudo, and J. Ichikawa, “16-level quadrature amplitude modulation by monolithic quad-parallel Mach-Zehnder optical modulator,” *Electronics letters*, vol. 46, no. 3, pp. 227–228, 2010.
- [99] P. Dong, X. Liu, S. Chandrasekhar, L. L. Buhl, R. Aroca, and Y. Chen, “Monolithic Silicon Photonic Integrated Circuits for Compact 100+ Gb/s Coherent Optical Receivers and Transmitters,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 4, pp. 150–157, 2014.
- [100] P. Dong, C. Xie, L. L. Buhl, Y. Chen, J. H. Sinsky, and G. Raybon, “Silicon In-

- Phase/Quadrature Modulator With On-Chip Optical Equalizer,” *Journal of Lightwave Technology*, vol. 33, no. 6, pp. 1191–1196, 2015.
- [101] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, “A review of lithium niobate modulators for fiber-optic communications systems,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 1, pp. 69–82, 2000.
- [102] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, “A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor,” *Nature*, vol. 427, no. 6975, pp. 615–618, 2004.
- [103] A. Liu, L. Liao, D. Rubin, H. Nguyen, B. Ciftcioglu, Y. Chetrit, N. Izhaky, and M. Paniccia, “High-speed optical modulation based on carrier depletion in a silicon waveguide,” *Optics Express*, vol. 15, no. 2, pp. 660–668, 2007.
- [104] C. R. Doerr, L. Zhang, P. J. Winzer, J. H. Sinsky, A. L. Adamiecki, N. J. Sauer, and G. Raybon, “Compact High-Speed InP DQPSK Modulator,” *IEEE Photonics Technology Letters*, vol. 19, no. 15, pp. 1184–1186, 2007.
- [105] L. Zhang, J. Sinsky, D. V. Thourhout, N. Sauer, L. Stulz, A. Adamiecki, and S. Chandrasekhar, “Low-voltage high-speed travelling wave InGaAsP-InP phase modulator,” *IEEE Photonics Technology Letters*, vol. 16, no. 8, pp. 1831–1833, 2004.
- [106] A. Melikyan, L. Alloatti, A. Muslija, D. Hillerkuss, P. C. Schindler, J. Li, R. Palmer, D. Korn, S. Muehlbrandt, D. Van Thourhout, B. Chen, R. Dinu, M. Sommer, C. Koos, M. Kohl, W. Freude, and J. Leuthold, “High-speed plasmonic phase modulators,” *Nature Photonics*, vol. 8, no. 3, pp. 229–233, 2014.
- [107] C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, and J. Leuthold, “All-plasmonic MachZehnder modulator enabling optical high-speed communication at the microscale,” *Nature Photonics*, vol. 9, no. 8, pp. 525–528, 2015.
- [108] D. Chen, H. R. Fetterman, A. Chen, W. H. Steier, L. R. Dalton, W. Wang, and

- Y. Shi, “Demonstration of 110 GHz electro-optic polymer modulators,” *Applied Physics Letters*, vol. 70, no. 25, pp. 3335–3337, 1997.
- [109] Y. Enami, C. T. Derose, D. Mathine, C. Loychik, C. Greenlee, R. A. Norwood, T. D. Kim, J. Luo, Y. Tian, A. K.-Y. Jen, and N. Peyghambarian, “Hybrid polymer/solgel waveguide modulators with exceptionally large electrooptic coefficients,” *Nature Photonics*, vol. 1, no. 3, pp. 180–185, 2007.
- [110] L. Alloatti, R. Palmer, S. Diebold, K. P. Pahl, B. Chen, R. Dinu, M. Fournier, J.-M. Fedeli, T. Zwick, W. Freude, C. Koos, and J. Leuthold, “100 GHz siliconorganic hybrid modulator,” *Light: Science & Applications*, vol. 3, no. 5, pp. e173–e173, 2014.
- [111] T. B. Simpson, J. M. Liu, and A. Gavrielides, “Bandwidth enhancement and broadband noise reduction in injection-locked semiconductor lasers,” *IEEE Photonics Technology Letters*, vol. 7, no. 7, pp. 709–711, 1995.
- [112] L. Chrostowski, X. Zhao, and C. J. Chang-Hasnain, “Microwave performance of optically injection-locked VCSELs,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 2, pp. 788–796, 2006.
- [113] E. K. Lau, X. Zhao, H.-K. Sung, D. Parekh, C. Chang-Hasnain, and M. C. Wu, “Strong optical injection-locked semiconductor lasers demonstrating  $\geq$  100-GHz resonance frequencies and 80-GHz intrinsic bandwidths,” *Optics Express*, vol. 16, no. 9, pp. 6609–6618, 2008.
- [114] D. Parekh, X. Zhao, W. Hofmann, M. C. Amann, L. A. Zenteno, and C. J. Chang-Hasnain, “Greatly enhanced modulation response of injection-locked multimode VCSELs,” *Optics Express*, vol. 16, no. 26, pp. 21 582–21 586, 2008.
- [115] E. K. Lau, L. J. Wong, and M. C. Wu, “Enhanced Modulation Characteristics of Optical Injection-Locked Lasers: A Tutorial,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 15, no. 3, pp. 618–633, 2009.
- [116] W. D. Sacher, E. J. Zhang, B. A. Kruger, and J. K. S. Poon, “High-speed laser modulation beyond the relaxation resonance frequency limit,” *Optics express*, vol. 18, no. 7, pp. 7047–7054, 2010.



- [117] S. P. Bhooplapur and P. J. Delfyett, “Characterization of the Phase and Amplitude Modulation of Injection-Locked VCSELs at 1550 nm Using Coherent Optical Demodulation,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 6, pp. 6–13, 2013.
- [118] P. Guo, T. Sun, W. Yang, D. Parekh, C. Zhang, X. Xie, C. J. Chang-Hasnain, A. Xu, and Z. Chen, “Optical phase modulation based on directly modulated reflection-mode OIL-VCSEL,” *Optics Express*, vol. 21, no. 19, pp. 22 114–22 123, 2013.
- [119] N. K. Fontaine, X. Xiao, H. Chen, B. Huang, D. T. Neilson, K. W. Kim, J. H. Sinsky, R. R. Ryf, G. Raybon, P. Winzer, A. Daly, C. Neumeyr, and M. Ortsiefer, “Chirp-Free Modulator using Injection Locked VCSEL Phase Array,” in *ECOC 2016 - Post Deadline Paper; 42nd European Conference on Optical Communication*, 2016, pp. 1–3.
- [120] Z. Liu, J. Kakande, B. Kelly, J. O’Carroll, R. Phelan, D. J. Richardson, and R. Slavík, “Modulator-free quadrature amplitude modulation signal synthesis,” *Nature Communications*, vol. 5, no. 1, p. 5911, 2014.
- [121] X. Xiao, N. K. Fontaine, H. Chen, B. Huang, D. T. Neilson, K. W. Kim, J. H. Sinsky, R. R. Ryf, G. Raybon, P. Winzer, A. Daly, C. Neumeyr, M. Ortsiefer, and S. J. B. Yoo, “High-Speed IQ Modulator Based on Injection-Locked VCSEL Array,” in *Conference on Lasers and Electro-Optics*, ser. OSA Technical Digest (online). San Jose, California: Optical Society of America, 2017, p. STu1M.6.
- [122] R. Shau, M. Ortsiefer, J. Rosskopf, G. Boehm, C. Lauer, M. Maute, and M.-C. Amann, “Long-wavelength InP-based VCSELs with buried tunnel junction: properties and applications,” in *Proc.SPIE*, vol. 5364, jun 2004.
- [123] A. Murakami, K. Kawashima, and K. Atsuki, “Cavity resonance shift and bandwidth enhancement in semiconductor lasers with strong light injection,” *IEEE Journal of Quantum Electronics*, vol. 39, no. 10, pp. 1196–1204, 2003.