Scalable Silicon Photonic Receivers for Next-Generation AI Computing Fabrics

Yang, Seung-Jae

Department of Electrical and Electronic Engineering

Graduate School

Yonsei University

Scalable Silicon Photonic Receivers for Next-Generation AI Computing Fabrics

Advisor: Prof. Choi, Woo-Young

A Dissertation Submitted to the Department of Electrical and Electronic Engineering and the Committee on Graduate School of Yonsei University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Yang, Seung-Jae

August 2026

Scalable Silicon Photonic Receivers

for Next-Generation AI Computing Fabrics

This certifies that the dissertation
of Yang, Seung-Jae is approved

Committee Chair	**Woo-Young Choi**
Committee Member	**Sang-Kook Han**
Committee Member	**Myung-Jae Lee**
Committee Member	**Kwanseo Park**
Committee Member	**Younghyun Kim**

Department of Electrical and Electronic Engineering

Graduate School

Yonsei University

August 2026

TABLE OF CONTENTS

List of Figures ..... **iii**

List of Tables ..... **vii**

Abstract ..... **viii**

CHAPTER 1 Introduction **1**

1.1 Motivation 1

1.2 Optical Interconnects and the Wide-and-Slow Paradigm 3

1.3 Industrial Landscape 5

1.4 Emerging Technical Bottlenecks and Thesis Contribution 7

CHAPTER 2 DWDM Crosstalk Cancellation **11**

2.1 Previous Limitation 11

2.2 Crosstalk Cancellation Technique 13

2.2.1 Principle 13

2.2.2 Architecture 16

2.2.3 Measurement Results 20

2.3 Temperature Control Method 23

2.3.1 Principle 23

2.3.2 Architecture 27

2.3.3 Measurement Results 28

2.4 Conclusion 30

CHAPTER 3 Monolithic Silicon Optical Receiver **32**

3.1 Previous Limitation 32

3.2 Monolithic Si APD Optical Receiver in Standard CMOS Technology 34

3.3 Si APD Design 36
3.3.1 Si APD Fabrication 36
3.3.2 Si APD Modeling 40
3.4 Receiver Architecture 49
3.5 Experimental Results 54
3.6 Conclusion 59
CHAPTER 4 Roadmap for Scalable Optical Interconnect 61
4.1 Dissertation Contributions 61
4.2 Remaining Challenges 62
4.3 Industry Roadmap and Future Directions 64
References 66
Abstract in Korean 70
List of Publications 72
International Journal Papers 72
International Conference Presentations 73

List of Figures

Fig. 1-1. LLM parameter outpacing memory bandwidth [1] 1

Fig. 1-2. Interconnect Wall 2

Fig. 1-3. TeraPHY solution of AyarLabs [7] 5

Fig. 1-4. MicroLED solution of Avicena [8] 6

Fig. 2-1. MRR filtering characteristic with FWHM of 369 pm 11

Fig. 2-2. (a) Diagram and (b) chip photograph of WDM receiver with MRR crosstalk cancellation. 17

Fig. 2-3. Schematics of the EIC data lane. 18

Fig. 2-4. Drop-port transmission spectrum as a function of wavelength offset from the resonance. 20

Fig. 2-5. Through-port transmission spectra and 28 Gbps eye diagrams with and without crosstalk cancellation for different channel spacing. 21

Fig. 2-6. 4-ch 28 Gbps eye diagrams with 250pm channel spacing. 22

Fig. 2-7. (a) Filtering characteristics of cascaded MRR array, (b-c) measured drop-port optical power of CH_1 MRR and the derivative for 400 pm channel spacing with varying CH_1 on-chip heater power and (d-e) for 250 pm channel spacing. 24

Fig. 2-8. (a) Block diagram and (b) chip photograph of demonstrated WDM receiver with temperature control 27

Fig. 2-9. (a) Monitored optical powers of channel 1 & 2 as heater 1 sweep, derivatives of monitored heater power of (b) channel 1 & 2 as heater 1 sweep, (c) channel 2 & 3 as heater 2 sweep, and (d) channel 3 & 4 as heater 3 sweep. 28

Fig. 2-10. (a) Heater voltage showing thermal stress compensation, and (b) measured 28Gb/s 4-channel eye diagrams without and with the heater control 29

Fig. 3-1. Cross-section of fabricated Si APDs. 38

Fig. 3-2. (a) I-V characteristics and (b) optical responsivity of the 28-nm Si APD under varying incident power levels as a function of applied reverse bias. 40

Fig. 3-3. Opto-electric 3-dB bandwidth measurements and equivalent-circuit simulation

results of the Si APD at three reverse bias conditions. 42

Fig. 3-4. Measured and simulated electrical reflection coefficients of the Si APD at three reverse bias conditions. 42

Fig. 3-5. (a) Measured signal photocurrent and noise current and (b) resulting SNR of the Si APD as a function of reverse bias voltage 44

Fig. 3-6. Equivalent circuit model of the Si APD. 46

Fig. 3-7. Top-level architecture of the monolithic 850-nm optical receiver 52

Fig. 3-8. Schematic diagram of TIA. 53

Fig. 3-9. Simulated frequency response of the monolithic Si optical receiver showing gain peaking at 9.8 GHz..... 53

Fig. 3-10. 20 Gb/s PRBS31 eye diagrams of the monolithic Si optical receiver at V_R = 9.4 V for (a) simulated and (b) measured results. 55

Fig. 3-11. Measured BER versus received optical power of the monolithic Si optical receiver for 20 Gb/s PRBS31 data 56

List of Tables

Table 2-1 IM/DD DWDM Comparison Table 30

Table 3-1 Extracted parameter values of the equivalent circuit model of the Si APD. . 48

Table 3-2 Comparison of monolithic optical receivers. 57

Abstract

Scalable Silicon Photonic Receivers for Next-Generation AI Computing Fabrics

Explosive growth in AI model complexity has exposed a critical mismatch between computational throughput and the bandwidth that conventional copper-based interconnects can sustain. This dissertation confronts that mismatch by developing scalable silicon photonic receiver circuits organized around a “Wide-and-Slow” philosophy, wherein aggregate capacity is built from a large number of low-rate optical lanes rather than from a small number of high-speed serial links.

Channel-count scaling can be pursued along two orthogonal axes. Spectral parallelism multiplies the number of WDM wavelengths carried within a fixed optical passband, as exploited by co-packaged optics solutions currently entering production. Spatial parallelism tiles 2D arrays of independent optical lanes across the full chip face rather than confining I/O to the peripheral shoreline. Either direction demands compact, high-density receiver circuits that integrate seamlessly with advanced CMOS processes.

The dissertation targets one critical bottleneck in each direction. On the spectral side, a DWDM receiver employing electronic inter-channel interference cancellation is

presented. Packing more wavelengths into a fixed free spectral range (FSR) mandates narrower channel spacing, yet narrower spacing increases spectral leakage among adjacent MRR filters. An analog cancellation circuit operating in the electrical domain neutralizes this deterministic interference, enabling error-free 28 Gb/s transmission at 250 pm (31.2 GHz) channel spacing across all four channels.

On the spatial side, a fully monolithic 850 nm optical receiver fabricated within an unmodified standard 28 nm CMOS process is reported. Realizing the silicon avalanche photodetector (Si APD) and the transimpedance amplifier (TIA) on the same die eliminates the interfacial parasitics that hybrid wire-bond or flip-chip assembly inevitably introduces, yielding a compact and area-efficient receiver topology. The fabricated device sustains 20 Gb/s operation at an optical sensitivity of -4 dBm sensitivity with an energy efficiency of 0.567 pJ/bit in 0.034 mm^2 area.

Taken together, the two contributions confirm that electronic interference cancellation and monolithic integration of photonic and electronic components are complementary and practically realizable routes to the scalable, sub-pJ/bit optical I/O demanded by next-generation AI computing fabrics.

Keywords: silicon (Si) photonics, micro-ring modulator (MRM), micro-ring resonator (MRR), optical interconnect, thermal calibration, wavelength-division multiplexing (WDM), crosstalk cancellation, monolithic integration, high-speed driver, high-performance computing (HPC), artificial intelligence (AI) systems.

CHAPTER 1 Introduction

1.1 MOTIVATION

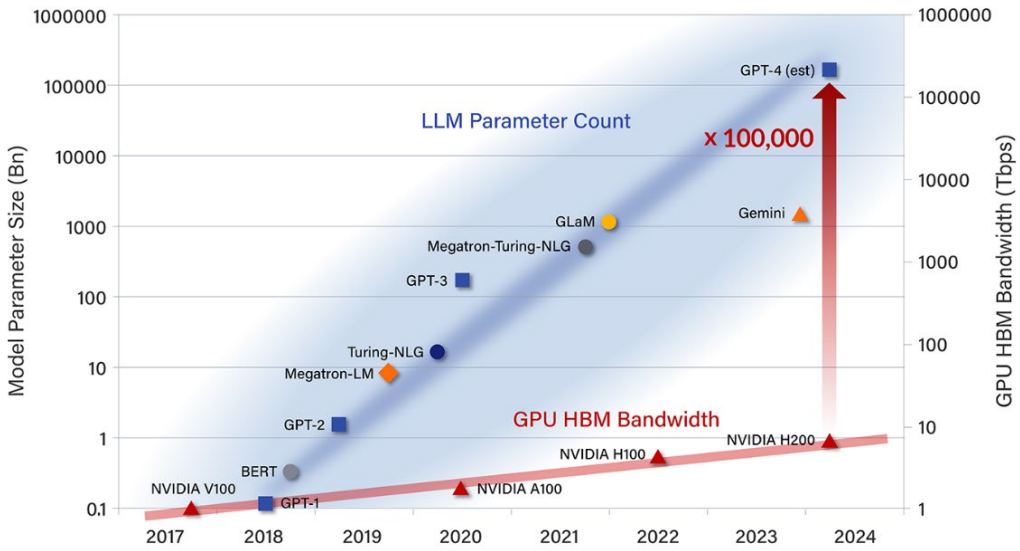


Fig. 1-1. LLM parameter outpacing memory bandwidth [1]

The computing infrastructure underpinning modern artificial intelligence is undergoing a fundamental architectural transformation. Rather than collections of loosely coupled servers, next-generation systems are converging toward tightly integrated "Computing Fabrics" in which accelerators and memory pools are interconnected so densely that they behave as a single logical resource. The primary catalyst for this shift is

the exponential growth of Large Language Models (LLMs). As illustrated in Fig. 1-1, parameter counts have expanded by more than five orders of magnitude since GPT-1 was introduced in 2017. While model complexity has grown without bound, the bandwidth available between accelerators (GPUs/NPUs) and High-Bandwidth Memory (HBM), the essential infrastructure underpinning these fabrics, has scaled far more slowly. This growing disparity has produced a severe bandwidth bottleneck that now constitutes the primary performance limiter for large-scale AI training and inference.

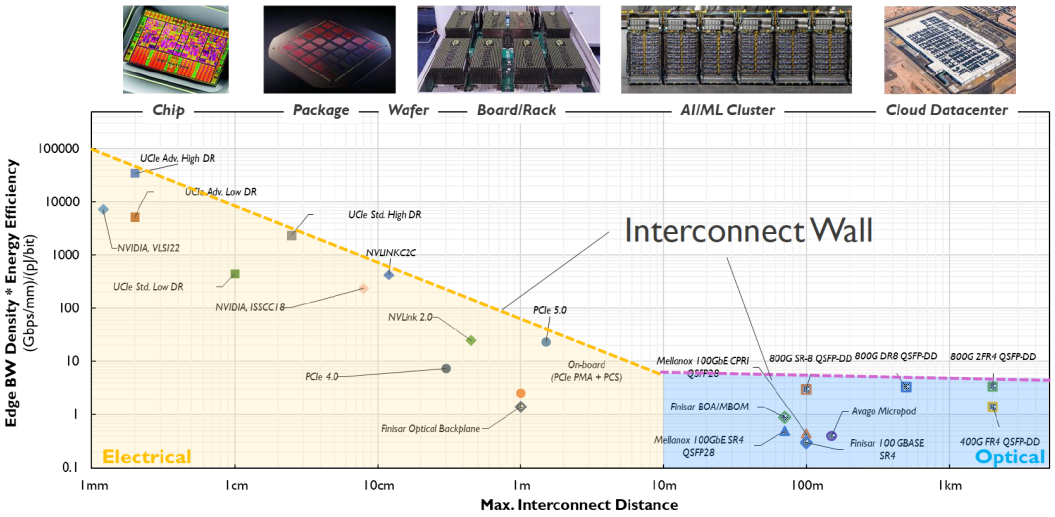


Fig. 1-2. Interconnect Wall

Copper-based PCB traces face hard physical ceilings rooted in frequency-dependent

attenuation mechanisms, including the skin effect and dielectric absorption [2]. As per-lane signaling rates advance from 112 Gbps toward 224 Gbps and beyond, PCB trace reach collapses sharply. The OIF CEI-224G specification illustrates this trend clearly. A 112 Gbps lane can traverse roughly 11 inches of board trace, whereas a 224 Gbps (56 GHz Nyquist) lane is limited to approximately 5 inches. This channel-shortening trend, captured in the "Interconnect Wall" of Fig. 1-2, renders copper wiring impractical for rack-scale AI clusters where intra-rack link distances routinely exceed what electrical SerDes can support without prohibitive power overhead.

1.2 Optical Interconnects and the Wide-and-Slow Paradigm

Optical interconnects have emerged as the primary solution to overcome the physical limitations of electrical links. Optical links provide low propagation loss, immunity to electromagnetic interference, and superior energy efficiency at medium and long distances. Silicon photonics (SiPh) is particularly attractive because it enables dense integration of optical components through CMOS-compatible fabrication processes, making it a strong candidate for next-generation AI computing fabrics [3], [4].

Scale-up interconnects, which connect accelerators within a single compute node or

rack, must achieve low latency, high energy efficiency, and high edge bandwidth density simultaneously. The industry is pursuing a wide range of architectural approaches to meet these requirements, from conventional electrical NVLink to silicon photonics WDM, MicroLED arrays, and VCSEL-based links.

The first generation of co-packaged optics solutions, exemplified by NVIDIA's Spectrum-X series, pursued what can be termed a "Fast-and-Narrow" (FaN) philosophy, wherein aggregate bandwidth grew by pushing individual serial lanes to ever-higher data rates, targeting configurations such as 8×100 Gbps for 800G links. This approach runs into a hard power ceiling. A state-of-the-art 200 Gbit/s PAM4 pluggable module dissipates roughly 13.5 pJ/bit across the full link, of which SerDes and digital signal processing (DSP) circuitry alone account for approximately 3.5 pJ/bit. Retiming, nonlinearity compensation, and forward error correction overhead compound nonlinearly as lane rates climb, undermining the scalability of future AI clusters [5]. This power pressure is driving the industry toward a "Wide-and-Slow" (WaS) architecture in which hundreds of parallel low-rate optical lanes replace a handful of fast serial channels, simplifying the electronic front-end and eliminating the need for power-hungry DSPs and forward error correction (FEC) engines [5], [6].

1.3 INDUSTRIAL LANDSCAPE

To realize Wide-and-Slow architectures, the industry has pursued various physical-layer approaches. These efforts fall into two broad directions. The first is spectral parallelism, which increases the number of WDM wavelength channels within a fixed optical bandwidth. The second is spatial parallelism, which deploys two-dimensional arrays of parallel optical lanes.

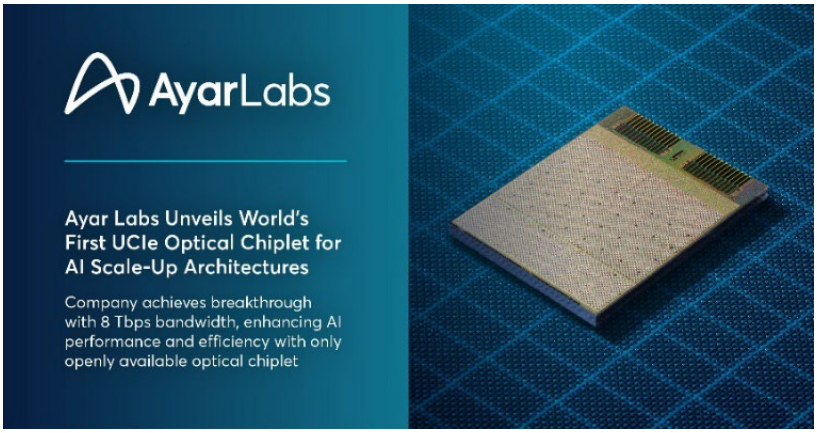


Fig. 1-3. TeraPHY solution of AyarLabs [7]

On the spectral parallelism front, Ayar Labs' TeraPHY chiplet multiplexes 16 wavelengths at 32 Gbps each and is engineered to stay below a 5 pJ/bit link budget [7]. Lightmatter's L200 photonic interposer scales each wavelength to 106 Gbps for a 2.98 Tbps aggregate throughput, and NVIDIA's 2026 silicon photonics roadmap targets 8

wavelengths at 32 Gbps with sub-pJ/bit efficiency. Despite their capacity advantages, all WDM solutions share a common vulnerability. As channel spacing shrinks, inter-channel spectral crosstalk intensifies and wavelength stabilization becomes increasingly demanding.

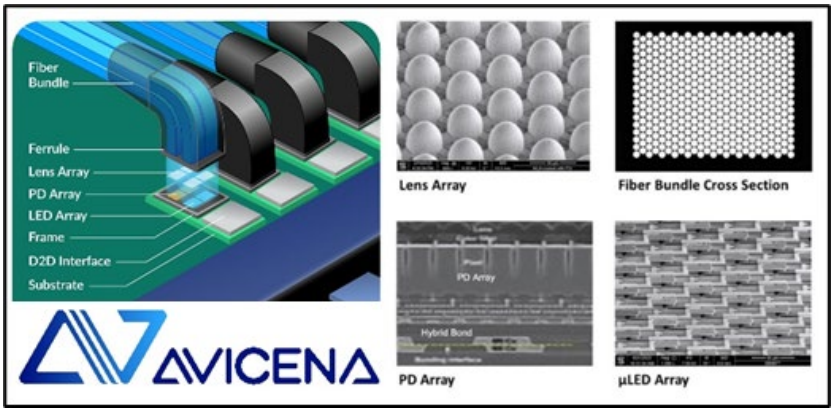


Fig. 1-4. MicroLED solution of Avicena [8]

Representative spatial-parallelism demonstrations include Avicena, which drives an array of roughly 300 GaN MicroLEDs at 3.3 Gbps each to deliver 1 Tbps of aggregate capacity within a 1 mm × 1 mm optical aperture. Microsoft's MOSAIC program takes a different tack, operating 100 low-rate optical lanes at 2 Gbps apiece through an analog-only backend to project 800 Gbps per link [6]. VCSEL-based solutions exploit their inherently high wall-plug efficiency, which exceeds 20 percent, to push link energy well below competing technologies. Intel's 2026 VCSEL CPO engine delivers 108 Gbps PAM4

at 0.9 pJ/bit, while arrays optimized for 32 Gbps NRZ achieve approximately 0.75 pJ/bit at the full link level [5], [9].

1.4 EMERGING TECHNICAL BOTTLENECKS AND THESIS CONTRIBUTION

As WaS architectures push for higher bandwidth density, the spectral spacing between channels must be aggressively reduced to accommodate more wavelengths within the fixed free spectral range of an on-chip MRR array. This spectral congestion affects the transmitter and receiver sides asymmetrically. On the transmitter side, micro-ring modulators (MRMs) benefit from operating at a relatively high Q factor, since a sharper resonance improves modulation efficiency and reduces the drive power required to achieve a given extinction ratio. On the receiver side, however, the MRR filter must simultaneously pass the full data bandwidth of the target channel while rejecting the adjacent channel, which places a lower bound on the filter bandwidth and consequently an upper bound on the usable quality factor. As channel spacing narrows, satisfying this constraint becomes increasingly difficult, and the receiver-side MRR reaches its crosstalk-limited performance ceiling before the transmitter-side MRM does. In dense WDM systems, this spectral leakage from neighboring wavelengths through the overlapping Lorentzian filter skirts of

adjacent MRRs causes deterministic inter-channel interference in the electrical domain, closing the eye diagram and creating a BER floor that cannot be overcome by simply increasing the laser power [10], [11]. The optical receiver therefore emerges as the first bottleneck in the path toward denser spectral scaling, motivating a receiver-centric approach to crosstalk mitigation.

This crosstalk induced by dense channel spacing is further exacerbated by thermal drift and fabrication tolerances, which cause unstable shifts in resonance. Similarly, in high-density spatial arrays like MicroLED or VCSEL clusters, light leakage between adjacent pixels, known as spatial crosstalk, degrades the signal-to-noise ratio (SNR). These interference effects collectively close the eye diagram and severely increase the bit error rate (BER). Consequently, the crosstalk challenge creates a performance ceiling, potentially nullifying the energy-efficiency gains intended by adopting lower-speed parallel transmission.

The scalability of parallel optical interconnects is further hindered by a combination of physical parasitic constraints and prohibitive manufacturing costs associated with traditional packaging. Current hybrid integration methods, which rely on wire bonding or micro-bumping to connect photonic components with electronic ICs, introduce significant

parasitic capacitance and inductance at the interface. These parasitics act as low-pass filters that attenuate high-frequency components and degrade receiver sensitivity, necessitating power-hungry circuitry and preventing the achievement of sub-pJ/bit energy efficiency.

Simultaneously, the economic and structural burden of scaling these interfaces presents a major hurdle. While 3-D hyper integration and advanced packaging are essential for achieving high density and reduced power in micro-nano systems, they face critical challenges regarding yield, cost, and complex thermal-mechanical constraints [12]. These complexities are particularly pronounced in optical interconnects, where packaging and fiber assembly costs now constitute the dominant fraction of the total module. The requirement for high-precision alignment across hundreds of channels, coupled with the inevitable increase in parasitics from complex assembly, makes traditional hybrid techniques both physically and economically unsustainable for mass-scale AI clusters.

To overcome these multifaceted bottlenecks, this thesis presents two primary research thrusts aimed at enabling the next generation of AI scale-up networks. First, we propose a system-level strategy for DWDM crosstalk management, employing active electronic cancellation and resilient wavelength locking to maintain signal integrity in ultra-dense spectral environments. Second, we introduce the design and implementation of a

monolithic silicon optical receiver, which integrates photodetectors and front-end circuitry on a single die to eliminate packaging-induced parasitics and reduce assembly complexity. Together, these approaches provide a scalable pathway to achieving the extreme bandwidth density and energy efficiency required for future computing fabrics.

CHAPTER 2 DWDM Crosstalk Cancellation

2.1 PREVIOUS LIMITATION

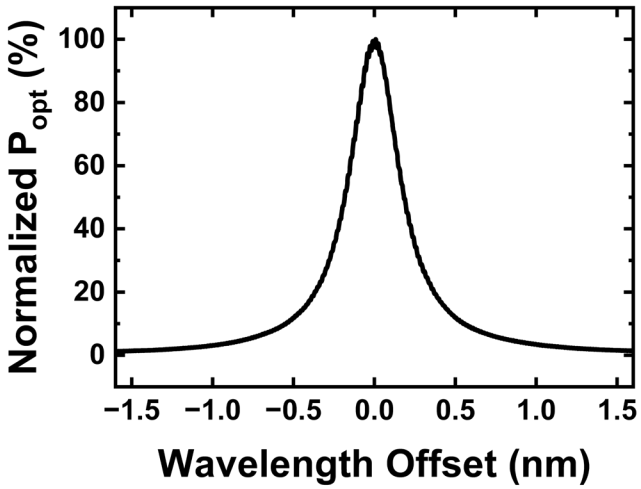


Fig. 2-1. MRR filtering characteristic with FWHM of 369 pm

The shift toward high-density WDM is a prerequisite for satisfying the escalating bandwidth demands of next-generation AI computing fabrics. As discussed in Chapter 1, the "Wide-and-Slow" strategy necessitates packing a greater number of optical channels into a limited spectral window to maximize aggregate throughput while maintaining energy efficiency. However, increasing the spectral density inevitably leads to severe inter-channel crosstalk, particularly when utilizing silicon photonic MRRs as wavelength-selective filters.

MRRs have emerged as the preferred wavelength-selective filters for dense WDM systems due to their ultra-compact footprint ~100 μm^2 and high quality factor. It can be integrated in massive arrays within a single chip, facilitating CPO architectures. Their Lorentzian-shaped transfer function provides spectral selectivity, enabling high spectral efficiency and a scalable I/O density as shown in Fig. 2-1.

Current commercial WDM transceivers defined under IEEE 802.3df, such as 800G FR4 and 1.6T FR8 modules, carry each lane at 200 Gbps PAM4 over a coarse wavelength grid with channel spacings of approximately 20 nm, corresponding to roughly 3.5 THz in the frequency domain. At this spacing, the inter-channel spectral overlap at the MRR filter skirt is entirely negligible, and crosstalk has not been a design concern. However, the industry trajectory points clearly toward much tighter channel spacings. The recently established Optical Compute Interconnect (OCI) Multi-Source Agreement, co-founded by NVIDIA, AMD, Broadcom, Meta, Microsoft, and OpenAI, defines a scale-up optical interconnect standard for AI clusters that already places four O-band wavelengths at approximately 2.3 nm spacing within a single bidirectional fiber, with its roadmap scaling wavelength counts further toward 3.2 Tbps per fiber. As this scaling progresses, channel spacings will be driven into the sub-nanometer regime, where MRR crosstalk transitions from a theoretical consideration into a fundamental performance limiter.

As channel spacing is aggressively reduced to achieve "Ultra-Dense" WDM, the spectral overlap between adjacent channels increases significantly. This overlap leads to optical crosstalk, where a portion of the signal from an adjacent wavelength leaks into the target receiver. This phenomenon causes deterministic inter-channel interference in the electrical domain, severely degrading the SNR and creating a BER floor that cannot be overcome by simply increasing the laser power.

2.2 CROSSTALK CANCELLATION TECHNIQUE

2.2.1 Principle

To address crosstalk, the MRR quality factor can be increased, thereby reducing the MRR filter bandwidth. However, this also limits the data bandwidth that can go through the filter. Multiple rings can be used so that a sharper filter response can be achieved, but this makes precise control of MRR characteristics more challenging [38], [39]. The interleaver can be utilized to isolate the neighboring channel, but it leads to a larger footprint [13], [14].

In this work, an analog crosstalk cancellation technique is proposed which mitigates the impact of spectral overlap. The core principle is to exploit the deterministic nature of the interference. Since the crosstalk in the victim channel is a filtered version of the signal

from the known adjacent aggressor channel, the interference can be neutralized in the electrical domain.

For generalization, let x be the channel spacing over the MRR bandwidth. The aggregate throughput = Data rate×Number of channels = Data rate×FSR/channel spacing = Data rate×FSR×MRR bandwidth/x. The optical power filtered by the neighboring channel MRR can be represented using a function as follows.

$$\mathrm{f(x)} = \frac{1}{1 + \mathrm{x}^2} \tag{2.1}$$

Assuming the MRRs are equally spaced as, the channel after that shows following characteristic.

$$\mathrm{f}(2x) = \frac{1}{1 + 4x^2} \tag{2.2}$$

Inversely, the optical power left after the preceding MRR filtering is

$$1 - f(x) = 1 - \left(\frac{1}{1 + x^2}\right) \tag{2.3}$$

Let the optical signal for channel n be P_n, and the signal filtered by channel n MRR as S_n. Then the Sn is the weighted sum of all P_n.

$$S_n = \sum_{k=0}^{N} C_k P_{n-k} \tag{2.4}$$

Where the Ck is determined by MRR characteristics.

The coefficient Ck represents the contribution to the nth MRR after the optical power has been depleted by all preceding MRRs from the first to the (n−1)th. Specifically, it quantifies the residual power that reaches the nth MRR. Preceding MRRs take all of the optical signal corresponding to them, so the Ck becomes 0 for k < n.

To generalize,

$$C_k = f(kx) \times \prod_{i=k+1}^{n} \big(1 - f(ix)\big), \quad 0 \; for \; k < 0 \tag{2.5}$$

For example, if we consider only 1 and 2 channel spacing crosstalk for ease,

$$C_0 = f(0) \times \big(1 - f(x)\big) \times \big(1 - f(2x)\big) \tag{2.6}$$

$$C_1 = f(x) \times \big(1 - f(2x)\big) \tag{2.7}$$

$$C_2 = f(2x) \tag{2.8}$$

The crosstalk cancellation technique used on the system is subtracting the following channel filtered signal with a weight.

$$S_n^{XTC} = S_n - w \times S_{n+1} \tag{2.9}$$

Then it can be expended only considering 1 and 2 channel spacing crosstalk as,

$$S_n^{XTC} = \mathrm{P_n} \times \mathrm{C_0} + \mathrm{P_{n+1}} \times (\mathrm{C_1} - \mathrm{w} \times \mathrm{C_0}) + \mathrm{P_{n+2}} \times (\mathrm{C_2} - \mathrm{w} \times \mathrm{C_1}) + \mathrm{P_{n+3}} \times (-\mathrm{w} \times \mathrm{C_2}) \tag{2.10}$$

The crosstalk can be calculated from the equation as the ratio of undesired power to desired power

$$XT = \frac{\sqrt{\{\mathrm{P_{n+1}} \times (\mathrm{C_1} - \mathrm{w} \times \mathrm{C_0})\}^2 + \{\mathrm{P_{n+2}} \times (\mathrm{C_2} - \mathrm{w} \times \mathrm{C_1})\}^2 + \{\mathrm{P_{n+3}} \times (-\mathrm{w} \times \mathrm{C_2})\}^2}}{\mathrm{P_n} \times \mathrm{C_0}} \tag{2.11}$$

The weight w should be taken to minimize the crosstalk.

2.2.2 Architecture

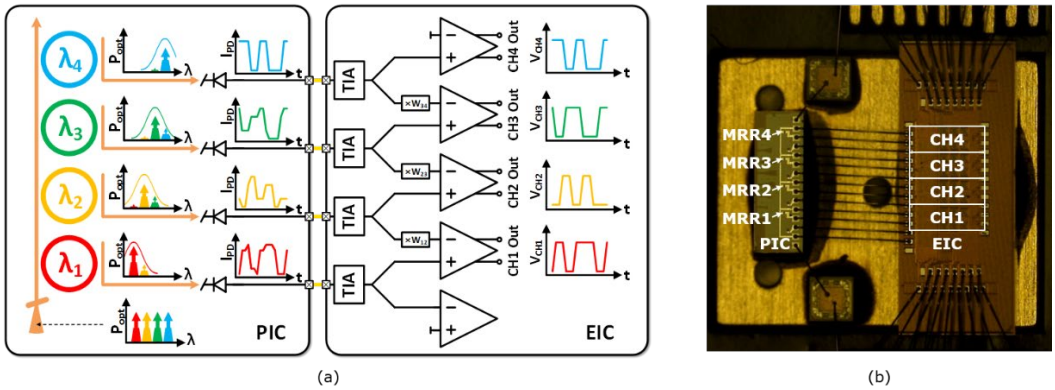


Fig. 2-2. (a) Diagram and (b) chip photograph of WDM receiver with MRR crosstalk cancellation.

Fig.2-2(a) schematically illustrates how crosstalk arises in cascaded MRRs and how it can be cancelled through signal processing in the receiver. As an example, consider two adjacent wavelengths, λ_1 and λ_2, shown in Fig.2-2(a). Ideally, each MRR filters only its designated wavelength. However, if the filter for λ_1 is not sufficiently selective, a portion of the λ_2 signal can leak into the λ_1 path, introducing crosstalk. This crosstalk can be eliminated by subtracting a weighted (W_{12}) electrical signal corresponding to λ_2 from the signal corresponding to λ_1, as shown in the figure. This subtraction function can be performed for each WDM channel, resulting in much improved received data quality. Although three subtractors each for λ_1 to λ_2, λ_2 to λ_3, and λ_3 to λ_4 are required for four WDM channels, two additional subtractors are added at the first and the last WDM channels, so that the symmetry

in the circuit layout can be maintained.

Fig.2-2(b) presents the implementation of the 4-channel WDM optical receiver featuring MRR-based crosstalk cancellation, which integrates a wire-bonded Si photonic integrated circuit (PIC) with a custom-designed electronic integrated circuit (EIC). The PIC incorporates four identical 8 μm radius add-drop MRR filters, each equipped with an integrated metal heater for thermal tuning. At the drop port of every MRR, a germanium (Ge) photodetector (PD) is situated, providing a bandwidth of 28 GHz. The MRRs exhibit a quality factor of approximately 4200, yielding a filter bandwidth of 23 GHz, and demonstrate a measured free spectral range (FSR) of 12.29 nm.

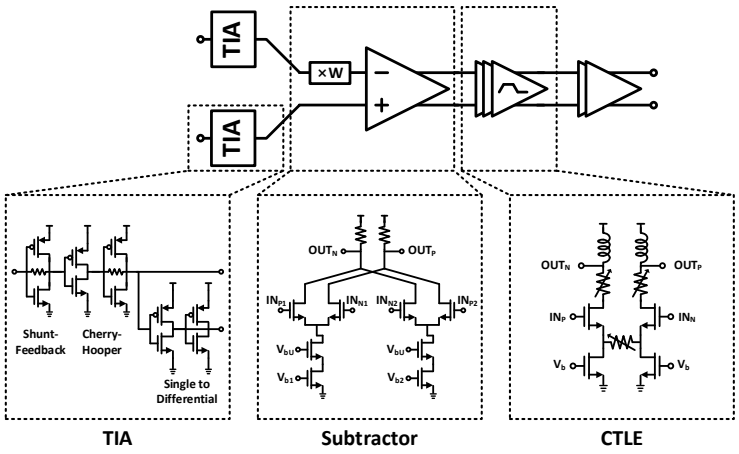


Fig. 2-3. Schematics of the EIC data lane.

The EIC, fabricated using a 28nm CMOS process, incorporates four distinct data lanes, as illustrated in Fig. 2-2(a). Each lane comprises a trans-impedance amplifier (TIA), a weighted subtractor for effective crosstalk cancellation, a three-stage equalizer to optimize receiver bandwidth, and a two-stage output buffer. The underlying circuit schematics are presented in Fig. 2-3. Specifically, the TIA architecture integrates a shunt-feedback inverter and an inverter-based Cherry-Hooper amplifier, concluding with a single-to-differential conversion stage. Crosstalk mitigation is achieved via a current-mode logic (CML)-based differential current summing architecture, where the subtraction weight is regulated by tail current control. Signal conditioning is performed by a source-degenerated continuous-time linear equalizer (CTLE) with peaking inductors, while a CML-based output buffer ensures precise 50 Ω impedance matching. [41]

To minimize distortion, symmetry in the subtraction path is prioritized to prevent timing skews. Because the PIC-induced phase offset is sufficiently small compared to the data rate, symmetry is enforced by utilizing symmetric subtractor topologies, placed strategically at the midpoint between adjacent TIAs in the EIC layout. This physical arrangement effectively mitigates timing offset. Additionally, to deal with signal distortion arising from the beat frequency components generated by closely spaced WDM signals at the PD, the equalizer is utilized to tune the EIC bandwidth. This approach selectively

attenuates the interference components while maintaining the spectral quality of the main data channel.

2.2.3 Measurement Results

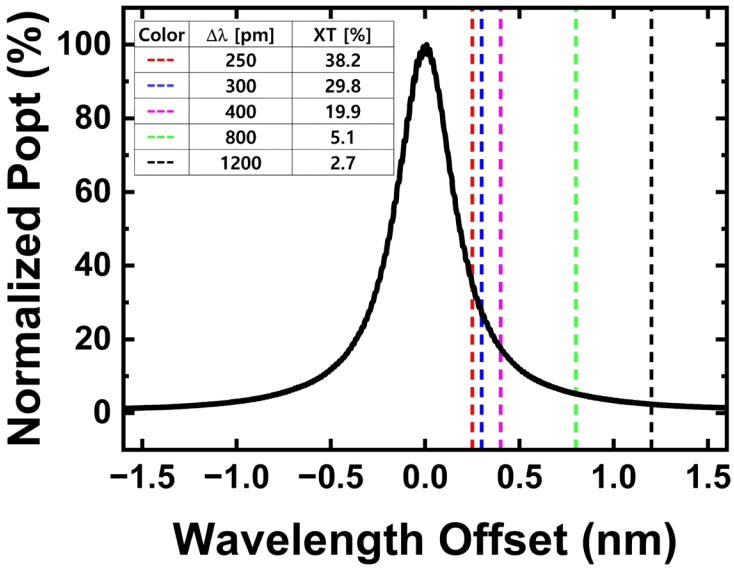


Fig. 2-4. Drop-port transmission spectrum as a function of wavelength offset from the resonance.

Fig. 2-4 illustrates the measured drop-port transmission characteristics of the MRR employed in the WDM receiver. The inset details the crosstalk (XT) levels resulting from

an aggressor channel at varying wavelength offsets (Δλ) of 250, 300, 400, 800, and 1200 pm from the MRR resonance peak. In a cascaded architecture, signals intended for a specific MRR may be partially dropped by preceding MRRs, thereby attenuating the desired signal power. Consequently, the actual interference experienced is more severe than that indicated by isolated drop-port measurements. To address this, we define the effective crosstalk as $XT_{eff} = XT / (1 - XT)$, which normalizes the crosstalk against the diminished target signal power. Based on this definition, the effective crosstalk values are calculated as 61.7%, 42.4%, 24.8%, 5.3%, and 2.8% for the corresponding wavelength offsets

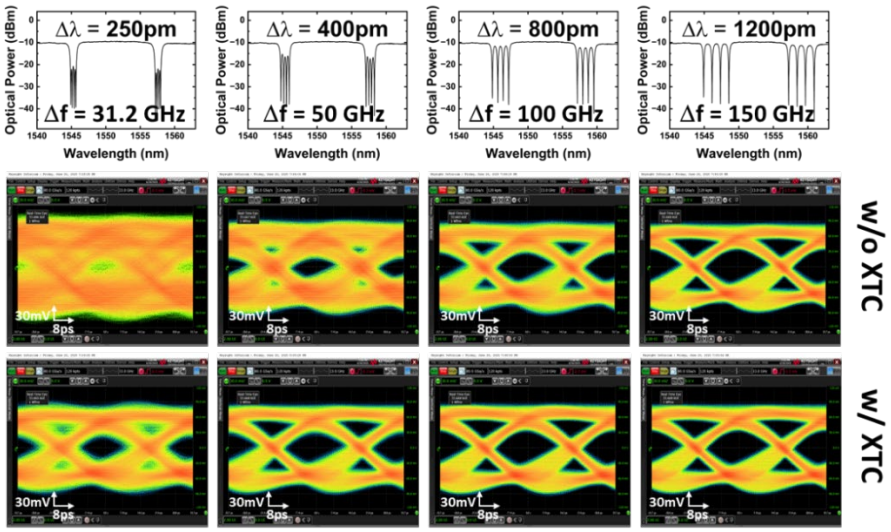


Fig. 2-5. Through-port transmission spectra and 28 Gbps eye diagrams with and without crosstalk cancellation for different channel spacing.

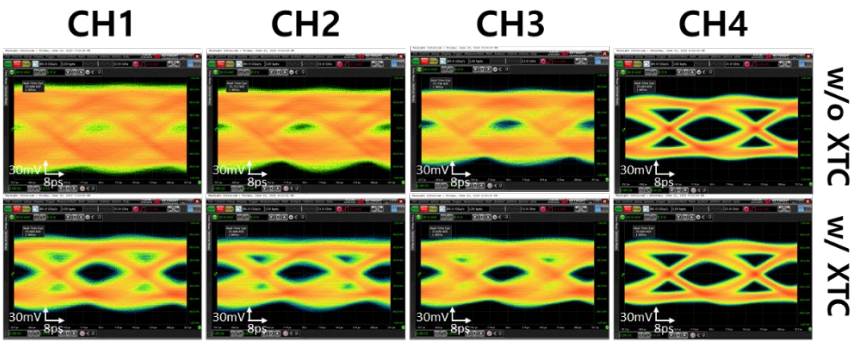


Fig. 2-6. 4-ch 28 Gbps eye diagrams with 250pm channel spacing.

Fig. 2-5 displays the measured through-port transmission spectra across various channel spacings, alongside the eye diagrams for Channel 1 (λ_1), demonstrating the system performance both with and without crosstalk cancellation. The eye diagrams were captured using a 28 Gbps PRBS31 signal generated via an external Mach-Zehnder modulator (MZM), with an input optical power of -3 dBm per WDM channel. Channel spacing was precisely controlled by adjusting the MRR temperatures through integrated on-chip heaters. As shown in Fig. 2-6, eye diagrams for all four channels at a $\Delta\lambda$ of 250 pm (corresponding to 31.2 GHz) exhibit clearly open eyes despite the significant spectral overlap. This successful crosstalk mitigation, combined with an FSR of 12.29 nm, validates the scalability of this architecture to 48 WDM channels at 250 pm spacing, yielding an aggregate throughput of 1.344 Tb/s in a single link.

2.3 TEMPERATURE CONTROL METHOD

2.3.1 Principle

Although the XTC can successfully suppress crosstalk between closely spaced WDM channels, locking the MRR resonance wavelength to the desired value for each WDM channel against any external perturbation presents a significant challenge. In this section, I present a new technique to address this issue.

The resonant wavelength of silicon MRRs is highly sensitive to temperature variations due to the large thermo-optic coefficient of silicon ~$1.8\times10^{-4}/K^{-1}$. In ultra-dense WDM systems, even a sub-degree temperature shift can cause the MRR resonance to drift away from the laser carrier wavelength, leading to signal fading or increased crosstalk. Previous solutions using global temperature control are power-intensive and lack the granularity required for dense arrays. The proposed principle is based on a feedback loop that monitors the receiver's performance, either through the average photocurrent or a dedicated monitor photodiode (MPD).

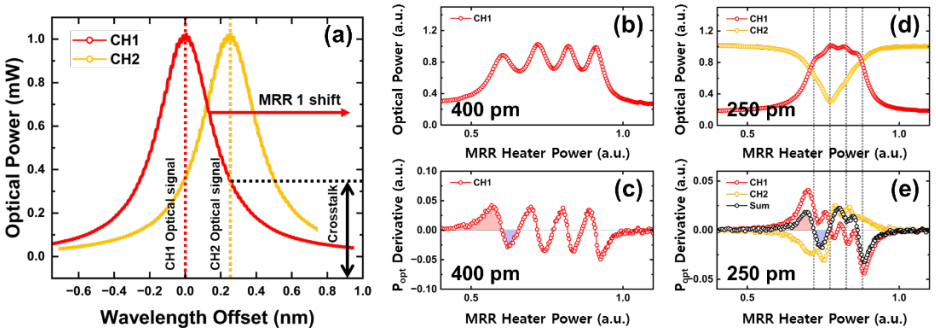


Fig. 2-7. (a) Filtering characteristics of cascaded MRR array, (b-c) measured drop-port optical power of CH_1 MRR and the derivative for 400 pm channel spacing with varying CH_1 on-chip heater power and (d-e) for 250 pm channel spacing.

Fig.2-7(a) shows the filtering characteristics of the cascaded MRR array. Fig.2-7(b) shows the drop-port optical power of CH_1 corresponding to channel 1 when the WDM channel spacing is 400 pm, as the power supplied to the CH_1 on-chip heater is scanned while all four wavelength input signals are introduced to the receiver. As can be seen in the figure, it is still possible to identify four distinctive peaks corresponding to four input wavelengths, but as shown in Fig.2-7(d) it is not possible at 250 pm channel spacing due to the relatively small channel spacing compared to the transmission bandwidth of the MRR. Although initial matching of each MRR with the target wavelength is achieved by sequentially turning one laser on at a time, it is not possible to lock this condition against

any external perturbation.

The core idea of the proposed thermal control algorithm is to consider the filtered optical power of the following MRR. The goal of the temperature control is to make the MRR filter the intended input optical signal maximally. The heater sweep response error occurs because a single MRR filters more than one optical signal at a narrow channel spacing, and this can be modeled as follows.

$$M_n = P_{n,n} + \mathrm{XT} \times P_{n,n+1} \tag{2.12}$$

To detect the resonance wavelength of MRR_n that maximizes the $P_{n,n,}$ the channel n signal filtered by MRR_n, maximizing the M_n, channel n monitored optical power, is not enough because of $P_{n,n+1,}$ the crosstalk from the channel n+1 filtered by MRR_n.

We add channel n and n+1 monitored optical power as follows.

$$M_n + M_{n+1} = P_{n,n} + \mathrm{XT} \times P_{n,n+1} + P_{n+1,n+1} + \mathrm{XT} \times P_{n+1,n+2} \tag{2.13}$$

In here, the channel n+1 optical signal filtered by MRR_n is ‘stolen' light that should have reached MRR_{n+1}. It leads directly to a decrease in the filtered optical power of MRR_{n+1} when λ_n, the resonance of MRR_n, sweeps. The derivative of the equation follows.

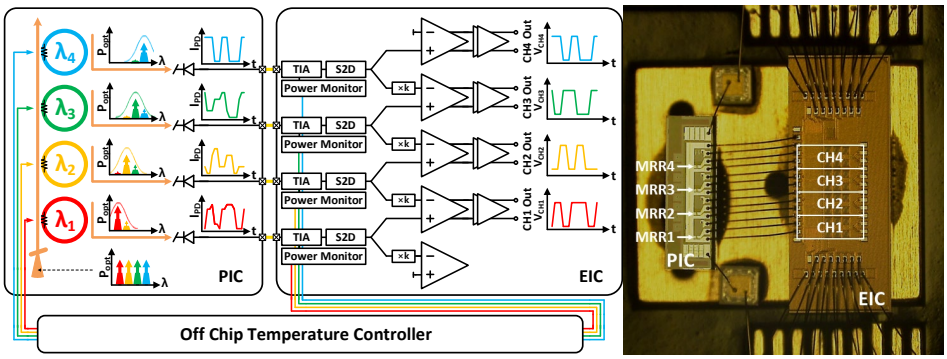
$$\frac{\partial M_n}{\partial \lambda_n} + \frac{\partial M_{n+1}}{\partial \lambda_n} = \frac{\partial P_{n,n}}{\partial \lambda_n} + \frac{\partial (XT \times P_{n,n+1})}{\partial \lambda_n} + \frac{\partial (P_{n+1,n+1})}{\partial \lambda_n} + \frac{\partial (XT \times P_{n+1,n+2})}{\partial \lambda_n} \quad (2.14)$$

The derivative has the same magnitude as the crosstalk with opposite direction, and as following equation it canceled out. Furthermore, $P_{n+1,n+2}$ is independent of λ_n, so it is also canceled out.

$$\frac{\partial M_n}{\partial \lambda_n} + \frac{\partial M_{n+1}}{\partial \lambda_n} = \frac{\partial P_{n,n}}{\partial \lambda_n} \quad (2.15)$$

Unlike simply monitoring M_n, the derivative sum crosses exactly zero at the point where the desired signal $P_{n,n}$ is maximized. Therefore, this property enables precise identification of the optimal heater temperature for each MRR channel. The proposed temperature control method utilizes the derivative sum to find the resonance wavelength that maximizes the intended signal power. [42]

2.3.2 Architecture



(a) (b)

Fig. 2-8. (a) Block diagram and (b) chip photograph of demonstrated WDM receiver with temperature control

Fig. 2-8(a) shows the block diagram of the 4-channel WDM receiver with the crosstalk cancellation and wavelength locking unit. It includes the crosstalk cancellation circuits of the previous section, heater driver, and filtered optical power monitor, which are connected to off chip temperature controller. The controller is implemented with the proposed algorithm, which utilizes the derivative sum to search and maintain the proper resonance wavelength of each MRR. Fig. 2-8(b) shows the chip photograph of the implemented 4-channel WDM optical receiver. The photonic integrated chip (PIC) and electrical integrated chip (EIC) are wire-bonded and mounted on a PCB. The EIC includes an optical power

monitor that converts optical power to voltage, which is fed into an external wavelength locking unit. This unit controls the heater driver on the EIC to tune the resonance wavelengths of MRRs in the PIC based on the derivative sum method.

2.3.3 Measurement Results

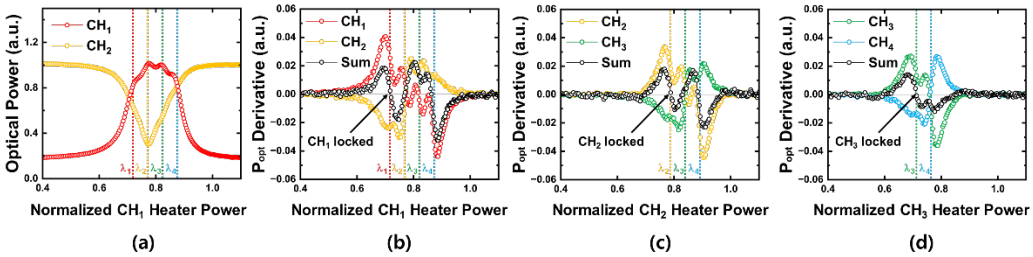


Fig. 2-9. (a) Monitored optical powers of channel 1 & 2 as heater 1 sweep, derivatives of monitored heater power of (b) channel 1 & 2 as heater 1 sweep, (c) channel 2 & 3 as heater 2 sweep, and (d) channel 3 & 4 as heater 3 sweep.

Fig. 2-9(a) shows the drop-port optical power of channel 1 to when the WDM channel spacing is 250 pm, as the power supplied to the channel 1 on-chip heater is scanned while all four wavelength input signals are introduced to the receiver. As can be seen in the figure, it is not possible to clearly identify four distinctive peaks corresponding to four input wavelengths due to the relatively small channel spacing compared to the transmission bandwidth of the MRR. Although initial matching of each MRR with the target wavelength

is achieved by sequentially turning one laser on at a time, it is not possible to lock this condition against any external perturbation. But as shown in Fig. 2-9(b), (c), and (d), the controller gets a proper metric to identify the heater power maximizing desired signal by utilizing the derivative sum.

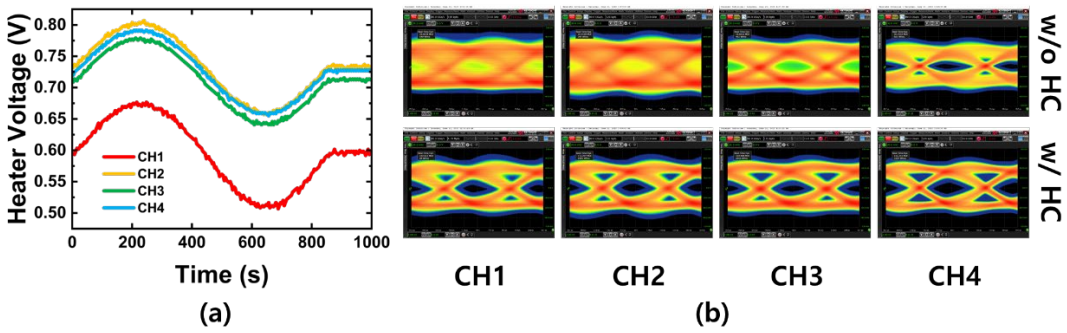


Fig. 2-10. (a) Heater voltage showing thermal stress compensation, and (b) measured 28Gb/s 4-channel eye diagrams without and with the heater control

Measurement results verify that the proposed method can maintain resonance stability over a wide temperature range. Fig 2-10(a) shows the heater voltages compensating the 5°C thermal stress for 1000 seconds, and Fig 2-10(b) shows the measured 28Gb/s eye diagrams during the measurement without and with the proposed control architecture. All of these are measured at 250 pm channel spacing and -3 dBm input optical power with PRBS31 pattern.

These results represent a spectral efficiency of 0.89 b/s/Hz, which is more than 5× higher than the prior state-of-the-art for IM/DD DWDM systems (0.16 b/s/Hz). The four-channel receiver achieves 112 Gb/s aggregate throughput; with a 48-channel configuration the same approach projects to 1.3 Tb/s aggregate capacity [42].

2.4 CONCLUSION

Table 2-1
IM/DD DWDM Comparison Table

	HP OFC 2023	Univ. CAS CICC 2024	Columbia Univ. Nat. Phot. 2023	Columbia Univ. CICC 2024	AMD ISSCC 2024	Yonsei Univ. This work
Carrier band (nm)	1310	1310	1550	1550	1310	**1550**
Channel spacing (GHz)	1600	1600	200	100	262	**31.2**
Number of channels	4	2	32	32	2	**4**
Datarate (Gb/s)	100	100	16	16	45	**28**
Spectral Efficiency (bps/Hz)	0.06	0.06	0.08	0.16	0.17	**0.897**
Crosstalk mitigation	-	Cascading MRR	Interleaving	Interleaving	Cascading MRR & Interleaving	**Electrical Analog Cancelling**
Thermal Control	-	Conditional peak searching	-	-	Conditional peak searching	**Derivative sum**

This table shows papers demonstrating dense WDM with IM/DD system, including the receiver. The spectral efficiency is defined as the data rate over channel spacing. The maximum aggregate throughput is obtained by multiplying the spectral efficiency by FSR. The two techniques presented in this chapter collectively push IM/DD DWDM spectral

efficiency to previously unreported levels. The electronic XTC circuit suppresses inter-channel spectral leakage in the electrical domain, permitting a four-channel receiver to operate at 250 pm (31.2 GHz) spacing with clearly open eyes at 28 Gb/s per lane. The derivative-sum wavelength locking algorithm addresses a complementary challenge. At spacings this tight, conventional peak-search routines cannot resolve individual MRR resonances from the monitored optical power alone. By exploiting the equal-and-opposite influence that a heater sweep on MRR×n exerts on its two neighboring monitor photodiodes, the derivative sum crosses exactly zero at the heater set-point where the intended signal power is maximized, providing an unambiguous locking reference regardless of inter-channel spectral overlap. Together, both mechanisms sustain robust eye quality through a 5°C ambient thermal excursion over 1000 seconds, achieving a spectral efficiency of 0.897 b/s/Hz, which is the highest on record for IM/DD DWDM systems and more than five times the prior state of the art. This superiority is confirmed by comparison with the 4λ × 28 Gb/s/λ WDM receiver of Kim et al. [36], the 4×112 Gb/s photocurrent-controlled WDM transmitter of Sharma et al. [37], and the 8λ × 50 Gbps heterogeneous Si-Ph DWDM transmitter of Levy et al. [40].

CHAPTER 3 Monolithic Silicon Optical Receiver

3.1 PREVIOUS LIMITATION

The conventional architecture of silicon photonic interconnects primarily relies on edge couplers or grating couplers to interface optical signals with the chip. Due to the stringent requirements for high-precision optical alignment and specific waveguide coupling structures, these I/O ports are traditionally confined to the chip's periphery, or "shoreline." This spatial constraint results in the shoreline bottleneck, where the aggregate escape bandwidth is strictly limited by the chip's perimeter. To satisfy the terabit-per-second bandwidth density demanded by next-generation AI clusters and exascale computing, the transition from one-dimensional shoreline scaling to two-dimensional surface-normal (vertical) coupling is becoming an architectural necessity.

Vertical-incidence optical solutions, particularly those based on 850 nm VCSELs and MMF, have been widely adopted for short-reach high-density applications due to their maturity and cost-effectiveness [9], [15], [16]. A quantitative technology comparison reported in [5] underscores why VCSELs are the preferred transmitter for this architecture. GaN microLED arrays are constrained to roughly 3.6 pJ/bit at the optical-engine level, a

consequence of their low wall-plug efficiency (∼1.2%) and the high drive voltage (4.4 V) required by their large capacitive load. In contrast, a Wide-and-Slow VCSEL array operating at 32 Gb/s per lane exploits wall-plug efficiency exceeding 20% to reach a projected total-link energy of approximately 0.75 pJ/bit, making VCSELs the only technology currently demonstrated to deliver a sub-1 pJ/bit co-packaged optics solution viable for AI scale-up links [5].

However, implementing these solutions on a silicon platform presents a fundamental material challenge. Silicon exhibits a very low absorption coefficient at the 850 nm wavelength, making efficient photodetection intrinsically difficult.

3.2 Monolithic Si APD Optical Receiver in Standard CMOS Technology

To address the aforementioned constraints, this research introduces a vertical-incidence silicon avalanche photodetector (Si APD) that is fully realized within a standard 28-nm CMOS process environment. A key advantage of this approach is the absence of any process modifications or design rule violations, ensuring seamless integration with existing foundry ecosystems. The Si APD leverages intrinsic avalanche multiplication to compensate for the characteristically low absorption coefficient of silicon at near-infrared wavelengths, particularly at 850 nm. By bypassing the complexities and signal integrity issues associated with heterogeneous bonding, this monolithic architecture provides a scalable, low-parasitic solution capable of supporting the massive throughput required for next-generation AI computing fabrics.

Prior work on monolithic Si APD receivers spans three technology generations. These include 0.25 μm BiCMOS reaching 12.5 Gb/s [17], 0.13 μm CMOS reaching 10 Gb/s [18], and 65 nm CMOS reaching 18 Gb/s [19]. In these previous implementations, the Si APD was strategically utilized to enhance responsivity while maintaining sufficient photodetection bandwidth. However, scaling these monolithic solutions to more advanced

technology nodes introduces significant performance trade-offs. As CMOS technology scales, doping concentrations are increased to mitigate short-channel effects in MOS transistors. This trend leads to a narrower depletion region width, which reduces the effective volume for photon absorption and avalanche gain, subsequently degrading the overall responsivity of the photodetector [20].

Despite these device-level challenges, advanced nodes offer the significant benefit of higher-speed, lower-power CMOS electronics. Achieving an optimal performance balance between the scaled Si APD and the high-speed CMOS front-end is the fundamental challenge for realizing high-performance monolithic receivers. In this work, the 28-nm Si APD's photodetection characteristics at 850 nm are meticulously investigated and represented through a newly developed equivalent circuit model that faithfully emulates both frequency response and noise characteristics.

Utilizing this precise model, a monolithic optical receiver was designed featuring an underdamped TIA architecture specifically engineered to compensate for the Si APD’s bandwidth limitations. The resulting fabricated receiver achieves a robust 20 Gb/s operation with a bit error rate (BER) of less than 10^{-12} for PRBS31 data, even with an incident optical power as low as -4 dBm.

3.3 SI APD DESIGN

3.3.1 Si APD Fabrication

The implementation of high-performance APDs within a standard CMOS environment is inherently challenging, as these processes are optimized for digital and analog transistors rather than optoelectronic conversion. Consequently, extensive research has focused on identifying the most effective junction configurations to maximize both responsivity and bandwidth without necessitating process modifications.

The most elementary configuration for a CMOS-integrated APD involves an N-well/P-substrate junction. This structure benefits from a relatively wide depletion region, which typically yields higher responsivity compared to other available PN junctions in standard CMOS. However, this approach suffers from two critical drawbacks. First, the lack of a compatible guard ring structure leads to localized electric field enhancement at the junction boundaries, causing premature edge breakdown and severely limiting the achievable avalanche gain. Second, the photodetection bandwidth is significantly constrained by the slow diffusion of carriers generated deep within the charge-neutral P-substrate, where no electric field is present to facilitate drift transport.

In contrast, P+/N-well junction-based APDs offer a more robust alternative. By utilizing the Shallow Trench Isolation (STI) as a natural guard ring, these devices can suppress edge breakdown and achieve much higher stable avalanche gains. Furthermore, the presence of the N-well restricts the charge-neutral region, thereby reducing the impact of slow carrier diffusion and enhancing the overall bandwidth. [33]

To further optimize performance, this work utilizes an N+/P-well junction isolated by a Deep N-well (DNW) [21]. The inclusion of the DNW is crucial. It serves as a physical barrier that isolates the active P-well region from the P-substrate. This isolation effectively blocks the collection of slow-diffusing carriers from the substrate and prevents noise coupling from the underlying silicon [22], [23]. By strategically selecting this multi-junction architecture, the proposed Si APD achieves an optimal balance between high responsivity through avalanche multiplication and the high-speed operation required for 20 Gb/s data links.

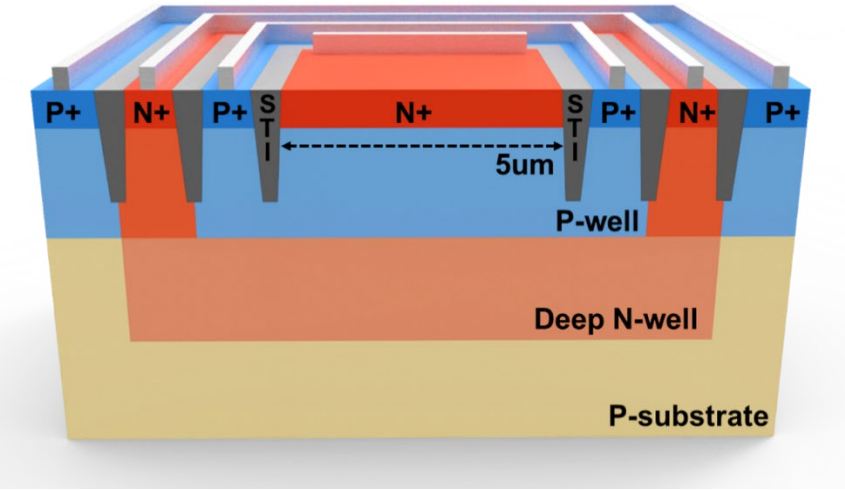


Fig. 3-1. Cross-section of fabricated Si APDs.

Fig. 3.1 illustrates the structural cross-section of the Si APD developed using the 28-nm CMOS process node. The device architecture utilizes a vertical N+/P-well junction, a configuration strategically selected to maximize the photodetection bandwidth. This choice leverages the superior mobility of minority carriers within the N+/P-well structure compared to the P+/N-well alternative, thereby facilitating faster carrier transport.

To ensure high-speed performance and signal purity, a deep N-well (DNW) is incorporated beneath the active P-well region. The DNW serves as an essential isolation barrier, preventing the collection of slow-diffusing photogenerated carriers from the P-

substrate, which would otherwise degrade the temporal response of the receiver. Additionally, this isolation layer effectively blocks noise coupling from the underlying silicon substrate, enhancing the overall signal-to-noise ratio.

The reliability of the N+/P-well junction under high electric fields is maintained by a shallow trench isolation (STI) guard ring. This structure is implemented to suppress localized electric field enhancement at the junction boundaries, thereby preventing premature edge breakdown and enabling stable avalanche operation. To optimize optical coupling while maintaining low electrical parasitics, a salicide-blocking layer is utilized to define a 5 μm × 5 μm optical window [24]. This layout ensures that incident light can penetrate the active region while minimizing parasitic resistance in the surrounding contact areas.

The operation of the Si APD is achieved by applying a reverse bias voltage across the N+ and P-well junction, with the resulting photocurrent extracted directly through the N+ terminal. To reinforce the isolation from the substrate, the junction between the P-well and the DNW is also reverse-biased. This dual-bias configuration ensures that the influence of the P-substrate is effectively neutralized, allowing the receiver to operate with high sensitivity and bandwidth even in the presence of complex substrate-level dynamics.

3.3.2 Si APD Modeling

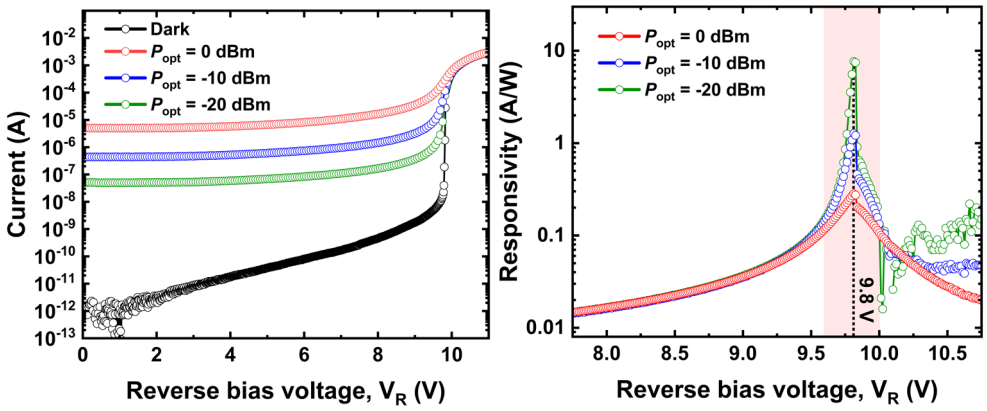


Fig. 3-2. (a) I-V characteristics and (b) optical responsivity of the 28-nm Si APD under varying incident power levels as a function of applied reverse bias.

The static performance of the fabricated 28-nm Si APD was evaluated through current-voltage measurements under varying optical power levels, as illustrated in Figure 3.2(a). For this characterization, an 850-nm laser diode served as the optical source, with light coupled into the device on-wafer via a lensed fiber featuring a 2.5 μm spot diameter. To ensure accurate analysis, an estimated coupling loss of approximately 3 dB was accounted for in the calculations.

As shown in the I-V curves, the Si APD exhibits a distinct avalanche breakdown at a reverse bias voltage of approximately 9.82 V. Fig. 3.2(b) presents the responsivity of the

device, calculated by normalizing the net photocurrent, defined as the difference between the total measured current under illumination and the dark current, to the optical power effectively incident on the active area. The device achieves its peak responsivity at a reverse bias of approximately 9.8 V, just below the breakdown threshold, where the avalanche gain is maximized.

A significant observation in the device's behavior is the dependency of responsivity on the incident optical power. The maximum responsivity exhibits a clear downward trend as the optical power increases, with 7.65 A/W for -20 dBm, 1.26 A/W for -10 dBm, and 0.28 A/W for 0 dBm. This reduction in responsivity at higher intensities is attributed to the saturation of avalanche multiplication within the high-field region of the Si APD. This power-dependent gain characteristic is a critical factor that must be considered when designing the subsequent TIA to ensure a wide dynamic range and stable receiver operation.

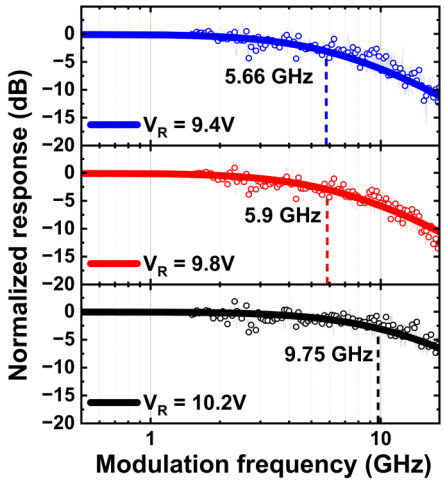


Fig. 3-3. Opto-electric 3-dB bandwidth measurements and equivalent-circuit simulation results of the Si APD at three reverse bias conditions.

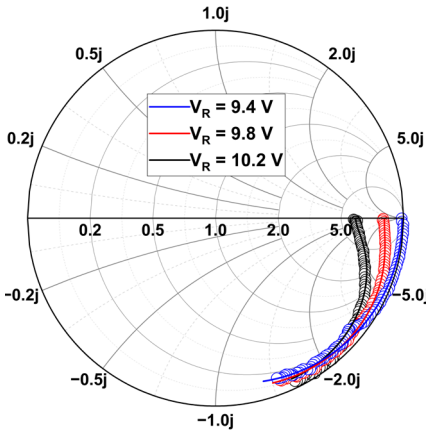
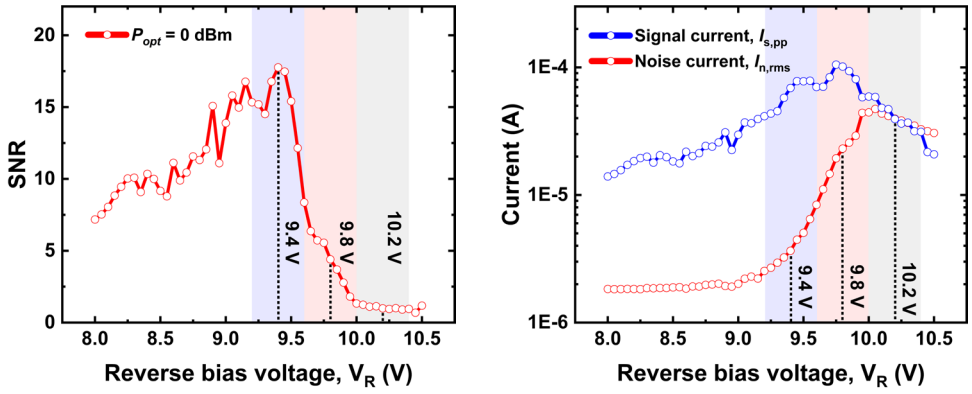


Fig. 3-4. Measured and simulated electrical reflection coefficients of the Si APD at three reverse bias conditions.

The dynamic performance of the monolithic Si APD was characterized by measuring

its small-signal frequency response across various reverse bias conditions. As depicted in Figure 3.3, the photodetection frequency responses were obtained using a characterization setup comprising an electro-optical modulator and a vector network analyzer (VNA). To ensure the accuracy of the high-frequency data, rigorous calibration was performed to de-embed the parasitic effects of RF cables and connectors. All measurements in this section were conducted with a constant average incident optical power of 0 dBm. The experimental results demonstrate that the photodetection bandwidth is highly dependent on the applied reverse bias voltage. A notable observation is that the Si APD achieves a maximum 3-dB bandwidth of 9.75 GHz at a reverse bias of 10.2 V. This bandwidth enhancement is primarily attributed to the enhanced inductive peaking effect that manifests as the device operates near the avalanche breakdown regime [25]. This inductive component effectively compensates for the junction capacitance, extending the frequency response beyond the conventional RC-limited cutoff and enabling the high-speed performance required for 20-Gb/s operation. To further investigate the device's electrical behavior and the mechanisms underlying the bandwidth extension, the electrical reflection coefficients were measured at various reverse bias voltages, as shown in Fig. 3.4. These electrical reflection coefficients provide essential information regarding the complex impedance of the Si APD under high-field conditions. The measured reflection data serve as a critical foundation for developing

an accurate equivalent circuit model, which is necessary for the co-design and optimization of the integrated TIA in the following sections.



(a) (b)

Fig. 3-5. (a) Measured signal photocurrent and noise current and (b) resulting SNR of the Si APD as a function of reverse bias voltage

To determine the optimal operating point for the integrated receiver, the signal and noise characteristics of the Si APD were analyzed across various reverse bias conditions. Fig. 3.5(a) presents the measured signal and noise currents at a frequency of 1 GHz. For the signal current characterization, a 0 dBm optical carrier was externally modulated at 1 GHz, and the resulting electrical power was measured using a spectrum analyzer. For the noise measurement, the same optical power was injected without modulation, and the power spectral density (PSD) at 1 GHz was recorded.

The noise current was derived by multiplying the measured PSD by the effective noise bandwidth, which was determined from the device's photodetection frequency response. To ensure high measurement precision and overcome the resolution floor of the spectrum analyzer, the Si APD output was amplified using commercial low-noise amplifiers (LNAs) with a combined gain of 46 dB and a noise figure of 3 dB. The influence of the LNAs was subsequently de-embedded from the final data to isolate the intrinsic noise performance of the Si APD.

Fig. 3.5(b) illustrates the SNR derived from these measurements. The results reveal a critical trade-off in the avalanche multiplication regime. While the device achieves its maximum responsivity at a reverse bias of 9.8 V, the peak SNR is observed at a lower bias of 9.4 V, where the responsivity is 0.067 A/W. Beyond 9.4 V, the noise current increases at a faster rate than the signal current due to the rapid escalation of the avalanche-induced excess noise. Consequently, the SNR begins to degrade as the bias approaches the breakdown voltage. Identifying this optimal bias point is essential for maximizing the sensitivity of the monolithic receiver, as it defines the threshold where the benefits of avalanche gain are most effectively utilized before being offset by excessive noise [34]. In this analysis, the noise contribution of the subsequent TIA circuit is not included. This simplification is justified because the Si APD noise current is significantly larger than the

input-referred noise of the TIA, making the circuit noise negligible in the overall SNR budget. The dominant noise source is therefore the Si APD itself, and the measured SNR in Fig. 3-5(b) accurately represents the photodetector-limited system noise performance.

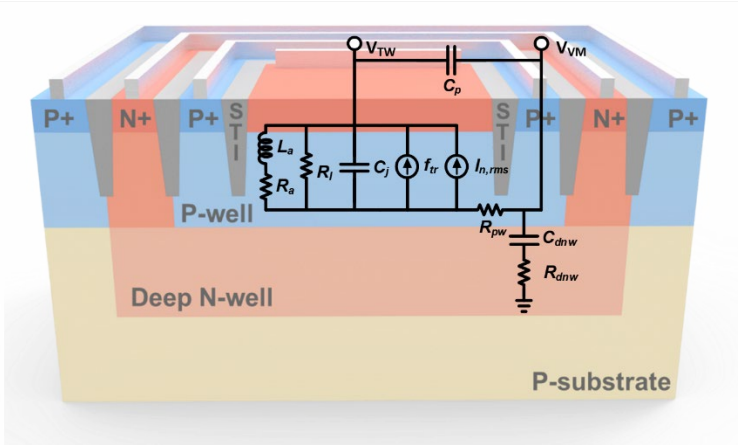


Fig. 3-6. Equivalent circuit model of the Si APD.

To facilitate the co-design and optimization of the TIA, an equivalent circuit model of the Si APD was developed, as illustrated in Fig. 3.6. This model is designed to represent not only the fundamental photodetection process but also the complex high-frequency parasitic network and the unique avalanche-induced dynamics of the Si APD. The core of the APD model consists of the active junction components. C_j represents the junction capacitance between the N-well and P-well, which is a primary factor determining the RC-limited bandwidth. The resistive elements R_a and R_l are incorporated to account for lossy

characteristics arising from the finite reverse saturation current and the field-dependent carrier velocity, respectively. Notably, the inductor L_a is included to model the phase delay in the photocurrent caused by the impact ionization process. This inductance is the physical basis for the inductive peaking effect observed in the frequency response measurements, allowing the model to accurately predict the bandwidth extension near the breakdown. The extrinsic parasitic network is modeled to reflect the specific layout and substrate architecture of the CMOS process. R_{pw} represents the resistance of the inactive P-well region, while C_p captures the parasitic capacitance between the N+ and P+ electrodes. Furthermore, the model accounts for the DNW isolation layer through R_{dnw} and C_{dnw}, which represent the DNW resistance and the junction capacitance between the DNW and the P-well, respectively [25], [26]. The intrinsic signal generation is represented by a frequency-dependent current source. The bandwidth of this source, denoted as f_{tr}, models the transit time limitations of the photogenerated carriers. Additionally, the noise characteristics of the device are integrated into the model through an RMS noise current source, $I_{n,rms}$, which is derived from the PSD measurements discussed in the previous section. By integrating these physical parameters into a unified SPICE-compatible netlist, the model enables precise prediction of the monolithic receiver's sensitivity, bandwidth, and stability during the circuit design phase.

Table 3-1

Extracted parameter values of the equivalent circuit model of the Si APD.

	9.4V	9.8V	10.2V
C_p [fF]	5		
L_a [nH]	20	14	4.3
R_a [Ω]	12800	825	240
R_l [kΩ]	20	12	8
C_j [fF]	19	16	12
R_{pw} [Ω]	100		
R_{dnw} [Ω]	220		
C_{dnw} [fF]	190	165	140
f_{tr} [GHz]	5.8	5.9	9
$I_{n,rms}$ [μA]	3.6	23	39.6

Table 3.1 summarizes the numerical values for all circuit elements across three distinct bias conditions. The parameters for the passive elements were determined by fitting the measured reflection coefficients with the simulated results, as shown in Fig. 3-4. Subsequently, the f_{tr} values were identified by correlating the measured Si APD frequency responses with simulations of the equivalent circuit using the previously extracted passive values. Finally, the $I_{n,rms}$ values were derived from the experimentally measured noise PSDs.

3.3.3 Wavelength Dependence of Si APD Performance

There is growing interest in extending VCSEL-based parallel links to shorter, visible wavelengths. AlGaInP-based red VCSELs emitting at approximately 650 nm have been explored for optical data links, with devices demonstrating 4.7 Gb/s transmission at 2 pJ/bit energy efficiency [27]. However, the AlGaInP material system presents inherent challenges compared to the GaAs-based 850 nm counterpart, most notably a smaller conduction band offset of approximately 0.17 eV at 650 nm versus 0.35 eV at 850 nm, which promotes thermal carrier overflow and limits the maximum operating temperature and output power [27]. In the blue spectral band, GaN-based laser diodes and micro-LEDs operating near 450 nm have demonstrated direct modulation exceeding 4 Gb/s, with the latter already being evaluated as transmitters in chip-scale parallel interconnects where a standard CMOS silicon photodetector serves as the receiver [8]. In the microLED case, the higher silicon absorption coefficient at visible wavelengths makes a simple Si PD without avalanche gain sufficient for detection, but the limited per-lane modulation bandwidth of microLEDs constrains achievable data rates well below those of VCSELs. A visible-wavelength VCSEL-based link retains the high modulation bandwidth of the laser source while also benefiting from the improved silicon absorption at shorter wavelengths, making the Si APD a natural receiver candidate for this configuration.

The structural parameters that are fixed by the device geometry and doping profile of the CMOS process and remain constant across all bias conditions are C_p, R_{pw}, and R_{dnw}, and these are likewise independent of the illumination wavelength. The remaining parameters L_a, R_a, R_l, C_j, C_{dnw}, f_{tr}, and $I_{n,rms}$ all vary with the applied bias voltage, and since the avalanche field distribution is set by the bias rather than the optical wavelength, they too are not directly affected by a change in the illumination wavelength. The transit time parameter f_{tr} is the one parameter that could in principle be affected, since shorter wavelengths confine photogenerated carriers closer to the junction. However, because the DNW isolation already effectively suppresses the slow-diffusing substrate carriers that would otherwise introduce a diffusion-limited bandwidth component, the contribution of carrier generation depth to ftr is already small at 850 nm, and the incremental improvement at visible wavelengths is expected to be modest. The dominant effect of operating at shorter wavelengths is therefore simply an increase in the responsivity, which follows from the higher quantum efficiency of silicon in the visible band.

This observation carries a practically useful implication. Since the equivalent circuit model remains structurally intact across wavelengths, with only responsivity changing as an input parameter, the design methodology established in this work for 850 nm can be applied directly to visible-wavelength VCSEL links without modification. A higher R_0

improves the peak SNR at the same bias and multiplication factor, which would allow the receiver to achieve comparable sensitivity at a reduced optical input power or, alternatively, to relax the avalanche gain requirement and operate at a lower reverse bias voltage than the 9.4 V optimum established at 850 nm. The monolithic Si APD receiver architecture is therefore well positioned to serve as a receiver for emerging visible-wavelength VCSEL-based parallel links with minimal redesign effort.

3.4 RECEIVER ARCHITECTURE

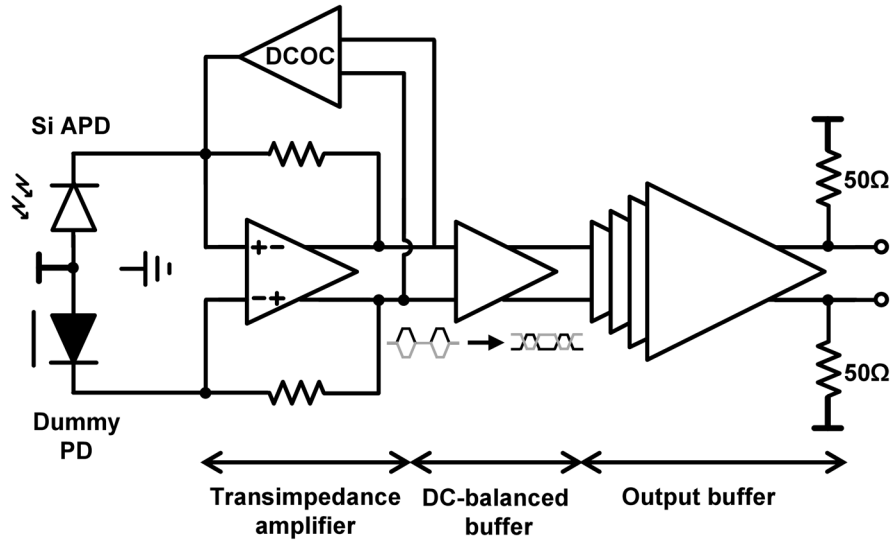


Fig. 3-7. Top-level architecture of the monolithic 850-nm optical receiver

The block diagram of the integrated monolithic optical receiver is illustrated in Fig. 3-7. The architecture comprises the Si APD paired with a dummy photodiode, a TIA, a DC offset cancellation (DCOC) loop, and a DC-balanced buffer, followed by an output stage designed for 50 Ω termination. To establish appropriate operating conditions, the cathode of the Si APD is biased through the common-mode feedback circuit of the TIA, while the anode is connected to an external DC supply.

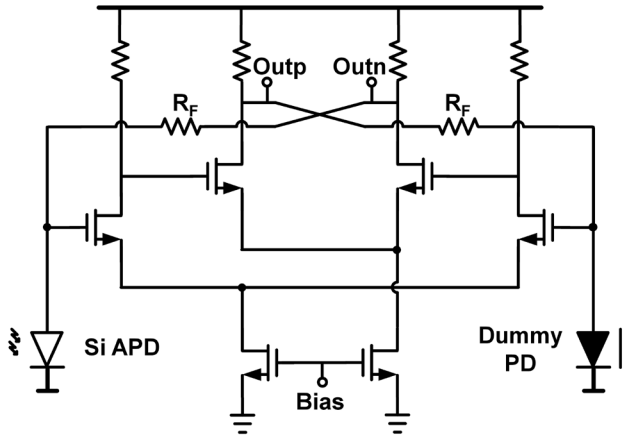


Fig. 3-8. Schematic diagram of TIA.

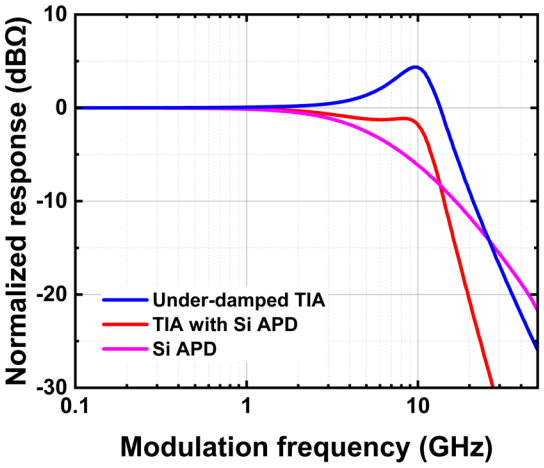
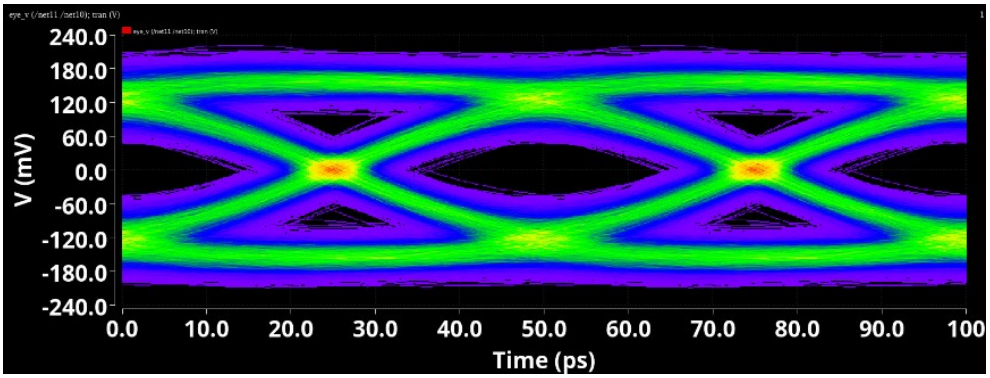


Fig. 3-9. Simulated frequency response of the monolithic Si optical receiver showing gain peaking at 9.8 GHz.

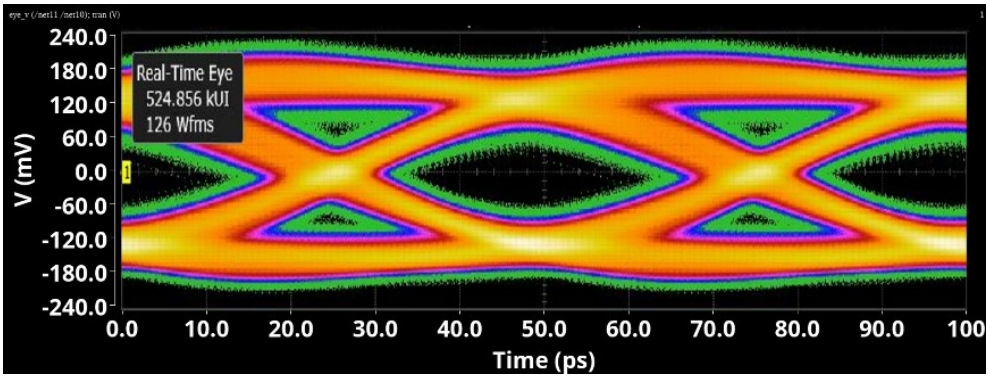
The detailed schematic of the TIA is presented in Fig. 3-8. The circuit is specifically engineered to exhibit an under-damped frequency response, which serves to compensate for the intrinsic bandwidth limitations of the Si APD. For this monolithic design, the equivalent circuit model at a reverse bias of 9.4 V is utilized, as this operating point provides the peak SNR identified in the previous characterization. By integrating the Si APD equivalent model into the design process, the TIA parameters were meticulously tuned to achieve the targeted peaking response for optimized system-level performance. [35]

3.5 EXPERIMENTAL RESULTS

With 1 kΩ shunt-feedback resistors, the TIA is tuned to produce 4.36 dB of intentional gain peaking at 9.8 GHz, visible as the blue curve in Fig. 3-9. Combining the TIA response with the measured Si APD characteristic yields a total transimpedance gain of 58 dBΩ and a system-level 3-dB bandwidth of 10.9 GHz, nearly doubling the standalone photodetector bandwidth of 5.9 GHz. DC wander is suppressed by two on-chip low-pass filters in the DC-balanced buffer stage, with their lower cutoff set at 1 MHz.



(a)



(b)

Fig. 3-10. 20 Gb/s PRBS31 eye diagrams of the monolithic Si optical receiver at V_R = 9.4 V for (a) simulated and (b) measured results.

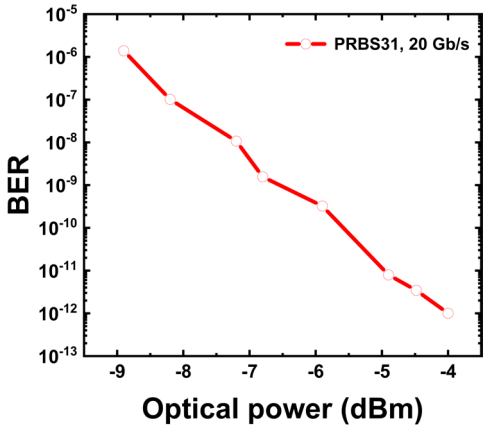
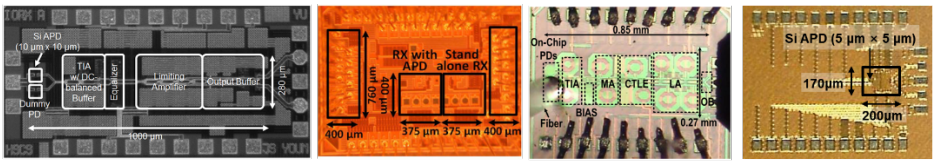


Fig. 3-11. Measured BER versus received optical power of the monolithic Si optical receiver for 20 Gb/s PRBS31 data

Agreement between simulated and measured eye diagrams, presented in Fig. 3-10(a) and (b) for 20 Gb/s PRBS31 data at 0 dBm incident optical power and a 9.4 V APD bias, validates the accuracy of the equivalent circuit model. The simulation jointly embeds the photodetector's frequency-response and noise characteristics together with the TIA transfer function, and closely reproduces the observed eye opening and crossing levels. Fig. 3-11 presents the measured BER as a function of received optical power at 20 Gb/s. The receiver reaches a BER below 10^{-12} at an incident power of −4 dBm, confirming that circuit-level bandwidth compensation fully overcomes the photodetector's intrinsic frequency-response deficit and enables error-free 20 Gb/s operation within an unmodified 28 nm CMOS process.

Table 3-2

Comparison of monolithic optical receivers.



<OE 2012, Yonsei Univ. [17]> <TCAS-I 2019, Semtech [18]> <JSTQE 2016, HKUST [19]> <JLT 2024, Yonsei Univ.>

	[17]	[18]	[19]	This work
Process	0.25 μm BiCMOS	0.13 μm CMOS	65 nm CMOS	28 nm CMOS
Data rate	12.5 Gb/s	10 Gb/s	18 Gb/s	20 Gb/s
PD BW	5 GHz	3.5 GHz	1.1 GHz	5.9 GHz
PD responsivity	0.07 A/W	3.92 A/W	0.272 A/W	0.28 A/W
Sensitivity	-7 dBm	-18.8 dBm	-4.9 dBm	-4 dBm
Area	0.28 mm^2	0.15 mm^2	0.23 mm^2	0.034 mm^2
Power*	59 mW	5.7 mW	48 mW	11.34 mW
Energy efficiency*	4.72 pJ/b	0.57 pJ/b	2.7 pJ/b	0.567 pJ/b

* Excluding output buffer power

Table 3-2 benchmarks the fabricated receiver against three prior monolithic optical receivers. The earliest work [17], built in 0.25 μm BiCMOS, requires a post-TIA limiting amplifier to recover sufficient gain. The compounding power overhead of that extra stage and the BiCMOS technology itself push total consumption to 59 mW, corresponding to 4.72 pJ/bit. The 0.13 μm CMOS implementation [18] is capped at 10 Gb/s because its 3.5 GHz photodetection bandwidth cannot be bridged to higher data rates even with supplementary equalization circuitry. The 65 nm CMOS design [19] illustrates the scaling penalty. As the technology node advances, heavier channel doping shrinks the depletion region, narrows the Si APD bandwidth, and forces multi-stage inductive peaking compensation that pushes power to 48 mW and degrades efficiency to 2.7 pJ/bit. The 28

nm receiver proposed here avoids this penalty by exploiting a more favorable APD-to-circuit bandwidth ratio. The 5.9 GHz photodetector bandwidth is close enough to 20 Gb/s that a single underdamped TIA stage, without any discrete inductors or additional equalizer blocks, is sufficient to extend the system bandwidth to 10.9 GHz. This architectural simplicity, combined with the elimination of bonding parasitics, simultaneously yields the smallest active footprint (0.034 mm^2) and the best energy efficiency (0.567 pJ/bit) in the comparison, while maintaining a −4 dBm sensitivity under the demanding PRBS31 stimulus.

3.6 CONCLUSION

The work reported in this chapter establishes a viable path toward high-density vertical-incidence optical receivers by realizing a fully monolithic 850 nm receiver within a production-standard 28 nm CMOS process, requiring no modifications to the foundry process flow or violations of any design rules. Placing the Si APD and TIA on the same substrate eliminates the parasitic capacitance and inductance that wire-bond and flip-chip assembly inevitably introduce at the interface between the photodetector and the front-end circuit.

A central contribution of this chapter is the measurement-based equivalent circuit model that captures both the frequency response and noise characteristics of the 28 nm Si APD across multiple bias conditions. This model guided the design of an intentionally underdamped TIA whose controlled inductive peaking extends the receiver's 3-dB bandwidth to 10.9 GHz, compensating for the 5.9 GHz bandwidth limit of the standalone photodetector. The fabricated receiver sustains 20 Gb/s PRBS31 transmission at a −4 dBm optical sensitivity, consuming 11.34 mW for an energy efficiency of 0.567 pJ/bit within a 0.034 mm^2 active footprint. These metrics collectively validate that monolithic co-integration of the photodetector and circuit in an unmodified digital CMOS process is a

practical and scalable strategy for high-density optical I/O, and they establish one of the enabling technical foundations for the spatially parallel interconnect roadmap examined in the following chapter.

CHAPTER 4 Roadmap for Scalable Optical Interconnect

4.1 DISSERTATION CONTRIBUTIONS

This dissertation addressed two receiver-side bottlenecks that arise when scaling silicon photonic interconnects under a Wide-and-Slow architecture for AI computing fabrics.

Chapter 2 demonstrated an analog electronic crosstalk cancellation technique for MRR-based DWDM receivers. As WDM channel spacing is reduced below the MRR filter bandwidth, spectral leakage from adjacent channels degrades signal integrity. The proposed XTC circuit estimates this interference at the adjacent TIA output and subtracts it electrically, without modifying the photonic integrated circuit. A complementary crosstalk-resilient wavelength locking method, based on the derivative sum of adjacent monitor photocurrents, maintains MRR resonance alignment under thermal perturbation. Together, these techniques achieved 4 × 28 Gb/s transmission at 250 pm (31.2 GHz) channel spacing with a spectral efficiency of 0.897 b/s/Hz, which is more than five times the prior IM/DD DWDM state of the art. With 48-channel operation over the 12.29 nm MRR FSR, the approach projects to approximately 1.3 Tb/s aggregate throughput per chip.

Chapter 3 demonstrated a monolithic 850 nm optical receiver in standard 28 nm CMOS without process modifications. By fabricating a silicon avalanche photodetector (Si APD) directly within the CMOS substrate, the design eliminates the parasitic capacitance and inductance introduced by conventional hybrid wire-bond or flip-chip photodetector assembly. An underdamped TIA with controlled inductive peaking compensates for the Si APD bandwidth limitation, yielding 20 Gb/s operation at −4 dBm sensitivity, 0.567 pJ/bit energy efficiency, and an active area of 0.034 mm². This result was obtained from a single-channel measurement. The compact per-lane footprint it demonstrates is a meaningful data point for evaluating the scalability potential of the monolithic approach in high-density spatially parallel link architectures, though multi-channel array validation remains future work.

4.2 REMAINING CHALLENGES

For the XTC cancellation technique, three challenges must be resolved before production deployment. The first is signal power recovery. When channel spacing is extremely tight, a fraction of the target channel's optical power is pre-empted by adjacent MRR filters before detection. The XTC circuit removes the electrical interference but cannot recover the attenuated optical signal, setting an SNR floor at very high crosstalk

levels. The second challenge is non-nearest-neighbor interference. The current implementation targets only N ± 1 crosstalk, but at spacings below approximately 200 pm, N ± 2 contributions become non-negligible and require a wider cancellation matrix. The third challenge is autonomous weight calibration. The experimental results of Chapter 2 relied on manually set weights, whereas production systems require a closed-loop calibration engine on the EIC to continuously update weights as temperature, PIC aging, and laser power fluctuate.

For the monolithic Si APD receiver, the primary practical limitation is the high reverse bias voltage required for avalanche gain (approximately 9.4 V), which far exceeds the nominal CMOS supply. The on-chip charge pump that generates this voltage occupies area and introduces noise, and its burden increases as processes scale to lower supply voltages. A secondary constraint concerns wavelength. The Si APD responds at 850 nm and is therefore incompatible with the 1310/1550 nm bands used in WDM silicon photonic systems, which limits direct co-deployment with the Chapter 2 technique. Bridging both contributions would require a Ge-on-Si photodetector compatible with standard silicon photonics foundry processes.

4.3 INDUSTRY ROADMAP AND FUTURE DIRECTIONS

In the WDM domain, the spectral efficiency ceiling of photonic-domain approaches is becoming visible. Lightmatter's L200 achieves approximately 0.30 b/s/Hz using cascaded MRR filters, representing a 3.75× improvement over Ayar Labs TeraPHY (0.08 b/s/Hz), but cascading higher-order filters provides diminishing returns due to increasing thermal control complexity and fabrication tolerances. The XTC technique demonstrated in this work, at 0.897 b/s/Hz, is approximately 3× beyond L200 and operates on the same PIC architecture without changes. This positions electronic XTC cancellation as a natural continuation of the WDM spectral density roadmap once photonic-domain filter improvements are exhausted. The most immediate next step is autonomous weight calibration to enable field-deployed operation, followed by extension to N + 2 matrix coverage for spacings below 200 pm. In the longer term, co-integration with Kerr frequency comb sources [28], which can supply dozens of equidistant carriers from a single micro resonator, would remove the discrete laser array as a footprint and power bottleneck.

In the spatial-parallelism domain, Avicena and Microsoft have validated the WaS philosophy at scale with their respective μLED demonstrations [6], [8], and both groups have explicitly noted that per-lane data rates can be increased with improved receiver

circuit integration. Coherent's 2026 VCSEL 2D-array CPO demonstration points to VCSEL-based parallel links as another emerging platform for near-package AI interconnects [5]. In this landscape, the compact monolithic Si APD receiver of Chapter 3 is relevant as a potential approach to simplifying receiver-side assembly in 850 nm VCSEL-based systems. Eliminating the bonded photodetector assembly removes a source of parasitic overhead that the cited groups identify as limiting. This possibility warrants investigation in a multi-channel co-fabricated array, which is the primary future work direction for Chapter 3. A reduction in APD bias voltage, either through structural Si APD optimization or alternative high-sensitivity detector designs, is a prerequisite for practical deployment at advanced CMOS nodes.

Together, the two contributions of this dissertation demonstrate that circuit-level interference management and monolithic integration of photonic devices and circuits are viable paths toward the bandwidth density and energy efficiency required by next-generation AI computing fabrics, and that the techniques developed here are well-positioned to extend the industry roadmap in both the spectral and spatial parallelism directions.

References

[1] Avicena, "AI and LLMs Limited by Memory Bandwidth," *Avicena Technology*. [Online]. Available: https://avicena.tech/technology_bak/. [Accessed: Apr. 1, 2026].

[2] E. I. Green, F. A. Leibe, and H. E. Curtis, "The Proportioning of Shielded Circuits for Minimum High-Frequency Attenuation," *Bell Syst. Tech. J.*, vol. 15, no. 2, pp. 248–283, Apr. 1936.

[3] B. Jalali and S. Fathpour, "Silicon photonics," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4600–4615, Dec. 2006, doi: 10.1109/JLT.2006.885782.

[4] D. Thomson *et al.*, "Roadmap on silicon photonics," *J. Opt. (United Kingdom)*, vol. 18, no. 7, Jul. 2016, doi: 10.1088/2040-8978/18/7/073003.

[5] M. Kohli and J. Teissier, "VCSEL-based CPO for scale-up in A.I. datacenter—status and perspectives," arXiv:2601.14342 [eess.SP], Jan. 2026. [Online]. Available: https://arxiv.org/abs/2601.14342.

[6] K. Benyahya *et al.*, "MOSAIC: Breaking the optics versus copper trade-off with a wide-and-slow architecture and MicroLEDs," in *Proc. ACM SIGCOMM 2025*, New York, NY, USA: ACM, Aug. 2025, pp. 234–247, doi: 10.1145/3718958.3750510.

[7] M. Wade *et al.*, "TeraPHY: A chiplet technology for low-power, high-bandwidth in-package optical I/O," *IEEE Micro*, vol. 40, no. 2, pp. 63–71, Mar. 2020, doi: 10.1109/MM.2020.2976067.

[8] B. Pezeshki *et al.*, "10-Gb/s NRZ per lane data links with transferred micro-LEDs with record BER and energy efficiency," *IEEE Photon. Technol. Lett.*, vol. 37, no. 8, pp. 453–456, Apr. 2025, doi: 10.1109/LPT.2025.3552708.

[9] S. Mondal *et al.*, "A 108-Gb/s PAM-4 VCSEL-based direct-drive optical engine for co-packaged optics applications," *IEEE J. Solid-State Circuits*, vol. 61, no. 1, pp. 33–46, Jan. 2026, doi: 10.1109/JSSC.2025.3607311.

[10] Y. Wang *et al.*, "Scalable architecture for sub-pJ/b multi-Tbps comb-driven DWDM silicon photonic transceiver," in *Proc. SPIE Photonics West*, vol. 12429, San Francisco, CA, USA: SPIE, Mar. 2023, p. 55, doi: 10.1117/12.2649506.

[11] S. Chen *et al.*, "A 2λ×100-Gb/s optical receiver with Si-photonic micro-ring resonator and photo-detector for DWDM optical-IO," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2024, doi: 10.1109/CICC60959.2024.10529008.

[12] J. Q. Lu, "3-D hyper integration and packaging technologies for micro-nano systems," *Proc. IEEE*, vol. 97, no. 1, pp. 18–30, Jan. 2009, doi: 10.1109/JPROC.2008.2007458.

[13] Y. Wang *et al.*, "Silicon photonics chip I/O for ultra high-bandwidth and energy-efficient die-to-die connectivity," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2024, doi: 10.1109/CICC60959.2024.10529018.

[14] Y. Yuan *et al.*, "A 4×100-Gbps DWDM receiver using all-Si microring avalanche photodiodes," in *Proc. IEEE Opt. Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, 2023.

[15] J. He *et al.*, "Design of a PAM-4 VCSEL-based transceiver front-end for beyond-400G short-reach optical interconnects," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 11, pp. 4345–4357, Nov. 2022, doi: 10.1109/TCSI.2022.3179678.

[16] P. J. Peng *et al.*, "A 56-Gb/s PAM-4 transmitter/receiver chipset with nonlinear FFE for VCSEL-based optical links in 40-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 57, no. 10, pp. 3025–3035, Oct. 2022, doi: 10.1109/JSSC.2022.3192711.

[17] J.-S. Youn *et al.*, "An integrated 12.5-Gb/s optoelectronic receiver with a silicon avalanche photodetector in standard SiGe BiCMOS technology," *Opt. Express*, vol. 20, no. 27, pp. 28153–28162, Dec. 2012, doi: 10.1364/OE.20.028153.

[18] S. Nayak *et al.*, "A 10-Gb/s −18.8-dBm sensitivity 5.7-mW fully-integrated optoelectronic receiver with avalanche photodetector in 0.13-μm CMOS," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 8, pp. 3162–3173, Aug. 2019, doi: 10.1109/TCSI.2019.2909284.

[19] Q. Pan *et al.*, "An 18-Gb/s fully integrated optical receiver with adaptive cascaded equalizer," *IEEE J. Sel. Topics Quantum Electron.*, vol. 22, no. 6, pp. 361–369, Nov. 2016, doi: 10.1109/JSTQE.2016.2574567.

[20] B. Nakhkoob *et al.*, "High speed photodiodes in standard nanometer scale CMOS technology: a comparative study," *Opt. Express*, vol. 20, no. 10, pp. 11256–11270, May 2012, doi: 10.1364/OE.20.011256.

[21] M.-J. Lee *et al.*, "A silicon avalanche photodetector fabricated with standard CMOS technology with over 1-THz gain-bandwidth product," *Opt. Express*, vol. 18, no. 23, pp. 24189–24194, Nov. 2010, doi: 10.1364/OE.18.024189.

[22] M. -J. Lee *et al.*, "Performance optimization and improvement of silicon avalanche photodetectors in standard CMOS technology," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 2, Mar. 2018, doi: 10.1109/JSTQE.2017.2754359.

[23] B. Razavi, "Prospects of CMOS technology for high-speed optical communication circuits," *IEEE J. Solid-State Circuits*, vol. 37, no. 9, pp. 1135–1145, Sep. 2002, doi: 10.1109/JSSC.2002.801195.

[24] M. -J. Lee *et al.*, "Effects of parasitic resistance on the performance of silicon avalanche photodetectors in standard CMOS technology," *IEEE Electron Device Lett.*, vol. 37, no. 1, pp. 60–63, Jan. 2016, doi: 10.1109/LED.2015.2496559.

[25] M. -J. Lee *et al.*, "Equivalent circuit model for Si avalanche photodetectors fabricated in standard CMOS process," *IEEE Electron Device Lett.*, vol. 29, no. 10, pp. 1115–1117, Oct. 2008, doi: 10.1109/LED.2008.2000717.

[26] Y.-C. Wang, "Small-signal characteristics of a Read diode under conditions of field-dependent velocity and finite reverse saturation current," *Bell Syst. Tech. J.*, vol. 44, no. 6, pp. 1149–1163, Jul.–Aug. 1965.

[27] N. Almaymoni *et al*., “650-nm vertical-cavity surface-emitting lasers (VCSELs) for plastic optical fiber communication” *Opt. Express*, vol. 31, no. 25, pp. 41361–41373, Dec. 2023, doi: 10.1364/OE.506616.

[28] A. Rizzo *et al.*, "Massively scalable Kerr comb-driven silicon photonic link," *Nat. Photon.*, vol. 17, no. 9, pp. 781–790, Sep. 2023, doi: 10.1038/s41566-023-01244-7

Abstract in Korean

차세대 AI 컴퓨팅 패브릭을 위한 확장 가능한 실리콘 포토닉 수신기

인공지능 워크로드의 급격한 성장으로 인해 기존 전기적 인터커넥트는 대역폭과 에너지 효율 측면에서 물리적 한계에 직면하고 있다. 본 논문은 이러한 한계를 극복하기 위해 "Wide-and-Slow" 전략에 기반한 확장 가능한 실리콘 포토닉스 수신기 구조를 제안한다. Wide-and-Slow 전략은 저속 광 채널을 대규모로 병렬화함으로써 높은 집계 처리량을 달성하는 방식이다.

채널 수를 확장하는 방식으로는 스펙트럼 병렬화와 공간 병렬화가 있다. 스펙트럼 병렬화는 고정된 광학 대역폭 내에 더 많은 파장분할다중화(WDM) 채널을 집적하는 방식이며, 공간 병렬화는 칩의 가장자리에 국한되지 않고 칩 전체 면적에 걸쳐 2차원 광 레인 배열을 구성하는 방식이다. 두 방식 모두 집적도 높은 수신기 구현을 필요로 한다.

본 논문은 각 방식에서 발생하는 핵심 병목 문제를 해결하는 두 가지

연구를 제시한다.

첫째로 더 높은 수준의 스펙트럼 병렬화를 위해 전자 회로 기반 누화 제거 기법을 적용한 고밀도 파장분할다중화(DWDM) 수신기를 제안한다. 고정된 자유 스펙트럼 영역(FSR) 내에 더 많은 채널을 집적하기 위해서는 채널 간격을 줄여야 하지만, 간격이 좁아질수록 인접 마이크로링 공진기(MRR) 채널로부터 스펙트럼 누화가 발생한다. 본 논문은 인접 채널 신호로부터 간섭을 추정하여 전기 영역에서 차감하는 아날로그 누화 제거(XTC) 기법을 적용하였다. 이를 통해 31.2 GHz 채널 간격에서 채널당 28 Gb/s의 안정적인 전송을 실현하였으며, 0.897 b/s/Hz의 스펙트럼 효율을 달성하였다. 이는 IM/DD DWDM 시스템에서 기존 최고 성능 대비 5배 이상 높은 수치이다. 또한, 초고밀도 채널 간격에서 기존 피크 탐색 방식으로는 정확한 공진 파장 추적이 불가능한 문제를 해결하기 위해, 인접 채널 모니터 광전류의 미분합(derivative sum)을 이용한 파장 잠금 기법을 제안하였다. 이 기법은 열적 스트레스에 대해 MRR 공진 안정성을 유지하며, 31.2 GHz의 좁은 채널 간격에서도

동작한다.

둘째로 공간 병렬화를 위해 표준 28 nm CMOS 공정에서 단일 칩으로 구현된 850 nm 단일칩 광 수신기를 제시한다. 공정 변형 없이 실리콘 애벌란치 포토다이오드(Si APD)와 트랜스임피던스 증폭기(TIA)를 동일 칩에 단일 집적함으로써 하이브리드 패키징에서 발생하는 기생 성분을 제거하고 복잡도를 낮추었다. 해당 수신기는 -4 dBm 감도에서 20 Gb/s 동작을 달성하였으며, 에너지 효율은 0.034 mm^2 면적에서 0.567 pJ/bit이다.

결론적으로 본 논문은 회로 레벨의 간섭 관리와 소자-회로 간 모놀리식 통합이 미래 컴퓨팅 인프라를 위한 고확장성 광학 I/O 구현의 핵심 경로임을 실증하였다. 본 연구에서 제시된 방법론은 향후 Kerr 주파수 콤(Kerr frequency comb) 기반 광원 기술 등과 결합하여, 지속 가능한 AI 컴퓨팅 생태계를 구축하기 위한 기술적 가교 역할을 할 것으로 기대된다.

List of Publications

International Journal Papers

[1] Jae-Ho Lee*, **Seung-Jae Yang***, Myung-Jae Lee, Woo-Young Choi, “25 Gb/s 850 nm Monolithic Optical Receiver With a Si APD Bias Controller Using Histogram Monitoring”, *IEEE Journal of Solid-State Circuits (JSSC, 2026)*

[2] Jae-Koo Park*, Dae-Won Rho*, **Seung-Jae Yang**, Woo-Young Choi, “An 80-Gb/s/pin Single-Ended Voltage-Mode PAM-4 Transmitter With a Pulse width Pre-Emphasis and a 4-Tap FFE in 28-nm CMOS” *IEEE Journal of Solid-State Circuits (JSSC, 2025)*

[3] **Seung-Jae Yang***, Jae-Ho Lee*, Myung-Jae Lee, Woo-Young Choi, “20 Gb/s CMOS Single-Chip 850 nm Optical Receiver” *Journal of Lightwave Technology (JLT, 2024)*

International Conference Presentations

[1] **Seung-Jae Yang**, Yongjin Ji, Daewon Rho, Jae-Ho Lee, Woo-Young Choi, “Crosstalk-Resilient Wavelength Locking for Si Micro-Ring-Resonator-Based Ultra-Dense WDM Receivers” *Optical Fiber Communication Conference (OFC, 2026)*

[2] Yongjin Ji, Daewon Rho, **Seung-Jae Yang**, Woo-Young Choi, “Modulation Crosstalk

Cancellation for Ultra-Dense WDM Silicon Photonic MRM Transmitters" *Optical Fiber Communication Conference (OFC, 2026)*

[3] Daewon Rho, **Seung-Jae Yang**, Yongjin Ji, Jae-Koo Park, Woo-Young Choi, "A 224-Gb/s Si-Photonic WDM Transmitter with Code-Based Calibration for Simultaneous OMA Locking and RLM Optimization" *Optical Fiber Communication Conference (OFC, 2026)*

[4] **Seung-Jae Yang**, YongJin Ji, Dae-Won Rho, Jae-Ho Lee, Woo-Young Choi, "A Si Photonic WDM Receiver with Micro-Ring Resonator Crosstalk Cancellation" *European Conference on Optical Communication (ECOC, 2025)*

[5] Woo-Young Choi, Dae-Won Rho, Jae-Koo Park, **Seung-Jae Yang**, Jae-Ho Lee, and YongJin Ji, "Si Photonic Ring-Resonator-Based WDM Transceivers" *2025 30th Asia and South Pacific Design Automation Conference (ASP-DAC), Tokyo, Japan (2025)*

[6] Dae-Won Rho, Jae-Koo Park, YongJin Ji, **Seung-Jae Yang**, Woo-Young Choi, "A 4λ× 50-Gb/s Si Photonic WDM Transmitter with Code-Based Wavelength Calibration and Locking" *Optical Fiber Communication Conference (OFC, 2025)*

[7] Dae-Won Rho, Jae-Koo Park, **Seung-Jae Yang**, Woo-Young Choi, “A 80Gb/s/pin Single-Ended PAM-4 Transmitter With an Edge Boosting Auxiliary Driver and a 4-Tap FFE in 28-nm CMOS” *IEEE Asian Solid-State Circuits Conference (A-SSCC, 2023)*