APL Photonics

PERSPECTIVE | JUNE 09 2025

The perspective of all-silicon photonics and systems

Y. Yuan ; Y. Peng ; S. Cheung ; X. Xiao ; W. V. Sorin ; Z. Huang; D. Liang ; A. Kumar ; R. Liu; Y. Hu; S. Hooten ; S. Palermo ; M. Fiorentino ; R. G. Beausoleil

APL Photonics 10, 060901 (2025)
https://doi.org/10.1063/5.0255608

AIP Publishing

The perspective of all-silicon photonics and systems

Cite as: APL Photon. 10, 060901 (2025); doi: 10.1063/5.0255608
Submitted: 30 December 2024 • Accepted: 14 May 2025 •
Published Online: 9 June 2025

Y. Yuan,[1,2,a)] Y. Peng,[1] S. Cheung,[1,3] X. Xiao,[1] W. V. Sorin,[1] Z. Huang,[1] D. Liang,[4] A. Kumar,[5] R. Liu,[5] Y. Hu,[1] S. Hooten,[1] S. Palermo,[5] M. Fiorentino,[1] and R. G. Beausoleil[1]

AFFILIATIONS

[1] Hewlett Packard Labs, Hewlett Packard Enterprise, Milpitas, California 95035, USA
[2] Department of Electrical and Computer Engineering, Northeastern University, Oakland, California 94613, USA
[3] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695, USA
[4] Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, Michigan 48109, USA
[5] Analog and Mixed Signal Center, Texas A&M University, College Station, Texas 77843, USA

[a)] **Author to whom correspondence should be addressed:** y.yuan@northeastern.edu

ABSTRACT

Silicon photonics has emerged as a transformative solution to address the energy and bandwidth challenges of modern computing and communication systems. While integrating diverse materials with silicon has enhanced the functionality of photonic integrated circuits, these hybrid approaches often face challenges related to scalability, cost, and compatibility with CMOS processes. We provide a comprehensive perspective on advancing high-performance all-silicon photonic devices and systems. By fully leveraging the inherent potential of silicon, diverse functionalities have been demonstrated, including Raman lasers, modulators, photodiodes, and optical memories. This approach outlines a pathway toward fully integrated electronic–photonic circuits on the silicon platform, seamlessly aligned with the existing CMOS infrastructure.

I. INTRODUCTION

Silicon (Si) has been the foundation of some of the most groundbreaking technological achievements in history. By the end of the last millennium, Si had ignited the digital revolution, powering transistors, microchips, and integrated circuits that transformed the way we compute, communicate, and innovate. These technologies gave rise to small, affordable, and energy-efficient devices capable of astonishing computational feats, revolutionizing industries and enabling an explosion of data generation and exchange. Yet, this rapid growth has brought with it a critical challenge: heat. The relentless flow of electrons through microchip interconnects generates pervasive Joule heating, an issue that looms large, especially in data centers where 35%–65% of operating costs are devoted to cooling systems.[1] Moreover, modern computer systems demand an explosive increase in bandwidth to keep pace with growing data processing needs. The compound annual growth rate is around 16% for the PCIe interface and 32% for the NVLink interface.[2] Meeting this demand is increasingly challenging due to the inherent limitations of electronic clock signal frequencies. In the last two decades, Si photonics has emerged as a transformative field, leveraging Si as an optical medium to address these challenges. Advances in high-precision etching have enabled the fabrication of Si waveguides with dimensions comparable to the wavelength of light, facilitating high-speed data transmission with significantly reduced heat and energy losses. Furthermore, Si photonics has enabled the development of optical modulators,[3,4] resonators,[5,6] and on-chip spectrometers,[7,8] allowing for precise light manipulation and analysis within compact device architectures.

While the vision of integrated circuits operating entirely with photons instead of electrons has long captivated researchers, the hybrid integration of photonic and electronic integrated circuits (EICs) represents a more practical and effective approach.[9] Photonic integrated circuits (PICs) excel in high-speed, low-power data

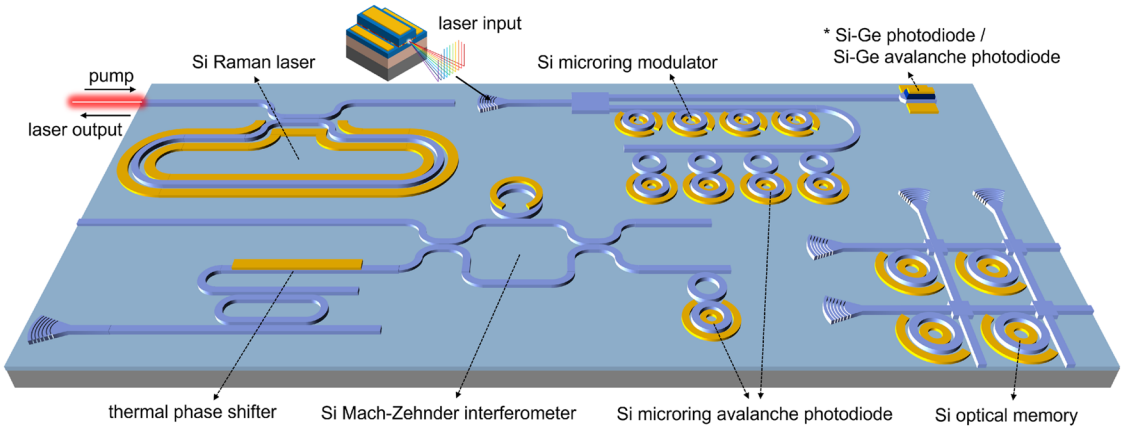

FIG. 1. Schematic diagram of an all-silicon (Si) photonic integrated circuit, illustrating key components, including a Si Raman laser,[10,11] Si microring modulator,[12,13] Si Mach–Zehnder interferometer,[14] Si microring avalanche photodiode,[15,16] and Si optical memory.[17,18]

transmission, while EICs are unparalleled in computational logic and signal processing. The combination of these technologies capitalizes on their respective strengths, creating a versatile platform capable of addressing the diverse requirements of modern computing and communication systems. The ultimate goal of Si photonics is to achieve co-packaging with networking, storage, and compute application-specific integrated circuits (ASICs), enabling optical input/output (I/O) directly from the Si chip itself. In this context, Si stands out as the most promising material platform for PICs over III–V semiconductors, owing to its compatibility with existing electronic infrastructure and CMOS fabrication processes. This compatibility ensures high production volumes, cost efficiency, and seamless integration of photonic and electronic components. However, Si is not without limitations. Its indirect bandgap prevents lasing, and its inability to detect light in the guided wavelength range of its waveguides necessitates the integration of various materials to achieve certain functionalities. Unfortunately, integrating multiple materials compromises the core advantages of Si photonics. Mismatches in lattice constants and thermal expansion coefficients result in defects such as dislocations or cracks during material growth or bonding, which can severely degrade performance and yield. Moreover, the added fabrication complexity and increased costs create significant barriers to scalability and seamless CMOS compatibility, limiting the widespread adoption of these hybrid solutions.

From this perspective, we appreciate the opportunity to review our recent progress toward realizing all-Si photonic integrated devices and systems, as illustrated in Fig. 1, which seek to address these challenges without sacrificing the benefits of the Si platform. To overcome the intrinsic limitations of Si, we introduce novel designs and mechanisms that fully exploit Si's potential. With the exception of the laser light source, which is implemented using other materials via single heterogeneous integration or external lasers, all remaining functionalities are implemented on an all-Si platform, striking an optimal balance between integration complexity and scalability. First, we present all-Si devices with a detailed analysis of their key performance metrics. Then, we explore all-Si systems built upon these devices, demonstrating their viability and potential to advance the next generation of computing and communication technologies. This approach aims to chart a pathway toward high-performance, scalable, and cost-effective photonic solutions grounded entirely in the Si platform.

II. ALL-SILICON PHOTONIC INTEGRATED DEVICES

A. Si lasers

The ideal scenario for Si photonics is to achieve a seamlessly integrated Si-based light source on the same chip. However, Si's indirect bandgap poses a fundamental challenge by inhibiting stimulated emission. Free electrons primarily occupy the X valley of the conduction band, where non-radiative Auger recombination dominates, leading to extremely poor light emission efficiency. The light emission efficiency can be described as $\eta_l = \tau_{non}/(\tau_{non} + \tau_{rad})$, where τ_{non} and τ_{rad} are the lifetimes for non-radiative and radiative processes, respectively.[19] Given that non-radiative processes occur far more readily in Si, η_l is exceptionally low, on the order of 10^{-6}. In 1991, Cullis and Canham showed that highly porous Si can emit efficient, multicolor visible light.[20] This light emission is facilitated by the quantum size effect within nanostructured crystalline formations under photoexcitation. Later, in 2000, Pavesi *et al.* demonstrated that light amplification is possible using Si itself by introducing nanocrystals, achieving gain comparable to that of direct-bandgap quantum dots.[21] At the nanoscale, holes and electrons are confined within the structure, significantly reducing the recombination time from milliseconds to microseconds. It enables population inversion between the fundamental state and a radiative state associated with the Si/SiO_2 interface. The first true Si laser was demonstrated in 2004 by Boyraz and Jalali using a different mechanism: stimulated Raman scattering.[22] Due to Si's strong Raman gain coefficient, achieving gain over a waveguide length is feasible. The pulsed operation was employed to avoid losses from the free carrier accumulation induced by two-photon absorption (TPA). One year later, Rong *et al.* from Intel developed the first continuous-wave Si Raman laser.[10] They

showed that TPA-induced free carrier absorption in the Si waveguide can be significantly reduced by introducing a reverse-biased P–I–N junction to sweep free carriers away. The threshold pump power of the Si Raman laser can be further reduced using a resonant cavity, achieving a low lasing threshold of 20 mW, nearly a tenfold improvement.[11]

These Si nanostructures and Raman lasers conclusively proved that Si can amplify light and lase efficiently despite having an indirect bandgap. However, they are not fully compatible with Si PICs primarily due to the requirement of an optical pump. It is essential to have electrically pumped lasers for PICs to avoid carrying an external light source. Due to this limitation, III–V compound lasers remain the preferred choice for superior performance. Significant progress has been made in integrating III–V lasers onto Si PICs using various techniques, including heteroepitaxy, wafer bonding, transfer printing, and photonic wire bonding, among others.[19,23–27] Either of the mentioned heterogeneous integration techniques will involve III–V or III-N materials, which can be patterned or unpatterned. Typical O- and C-band materials involve InP and GaAs with their ternary or quaternary counterparts. These III–V/Si lasers debuted commercially in Intel's Parallel Single Mode 4 (PSM4) and Coarse Wavelength Division Multiplexing 4 (CWDM4) 100 Gb/s transceivers in 2016.[28] The four distributed feedback (DFB) lasers were spaced 20 nm apart from 1290 to 1350 nm with a standard deviation of 0.3 nm, a relative intensity noise (RIN) < −150 dB/Hz, a side-mode suppression ratio of 50 dB, and a wall-plug efficiency >15% at 80 °C.[28] Hybrid III–V/Si lasers based on quantum dots are an active area of research due to frequency noise suppression and near-zero α_H factor, which is essential for minimizing coherence collapse due to external optical feedback.[29] In turn, this allows the removal of bulky and expensive optical isolators. Furthermore, these lasers have shown low transparency current density,[30] low optical gain thermal stability, and low RIN.[31] For dense wavelength division multiplexing (DWDM) communication, heterogeneous quantum dot comb lasers have been demonstrated with 15.5 GHz channel spacing over a 25 nm width.[32] The Fabry–Pérot version of these lasers has a low continuous wave (CW) threshold current density of 670 A/cm^2 at 100 °C.[32] Hybrid III–V/Si optical amplifiers have also been demonstrated with a wall-plug-efficiency of 12.1% using flared geometries.[33,34] A comparison of various types of lasers integrated on Si PICs is presented in Table I.

B. Si microring modulators

Optical modulation is a critical function in optical systems, allowing electrical signals to be encoded into the optical domain. This is achieved by altering the refractive index of a material through an applied electric field, thereby modulating the optical beam passing through it. Changes in the real part of the refractive index (Δn), known as electro-refraction, occur via mechanisms such as the Pockels effect and the Kerr effect. Meanwhile, changes in the imaginary part ($\Delta\alpha$) enable electro-absorption, exemplified by the Franz–Keldysh effect and quantum-confined Stark effect. Unfortunately, it has been shown that both changes are weak in pure Si at the telecommunications wavelengths of 1.31 and 1.55 μm. In 1987, Soref and Bennett evaluated the change in Si refractive index due to carriers,[42] a phenomenon termed the plasma dispersion effect. By varying carrier concentrations through applied voltage or current, the Si refractive index can be modified indirectly. The plasma dispersion effect has been commonly used in Si modulators and phase shifters as the most prominent, high-speed effect. The refractive index change at 1.31 μm is described by[3]

$$\begin{aligned}\Delta n &= -6.2\times10^{-22}\Delta N - 6.0\times10^{-18}\Delta P^{0.8},\\ \Delta\alpha &= 6.0\times10^{-18}\Delta N + 4.0\times10^{-18}\Delta P,\end{aligned} \tag{1}$$

where ΔN and ΔP are the carrier concentration changes of electrons and holes, respectively, in units of cm^{-3}. This effect is relatively weak; for example, the carrier concentration changes of 5×10^{17} cm^{-3} lead to a Δn of ~ −0.001. To amplify this effect, the microring resonator (MRR) structure has been widely adopted in Si modulators, enabling more pronounced optical changes with Δn. Compared to the millimeter-scale lengths of silicon Mach–Zehnder modulators (MZMs), silicon microring modulators (MRMs) offer a dramatically smaller footprint, with radii ranging from tens to a few micrometers. Resonance enhances the effective interaction length between propagating light and free carriers while also introducing intrinsic wavelength selectivity for wavelength division multiplexing (WDM). Their compact size enables higher integration density, reduced capacitance for greater bandwidth, and lower power consumption, making them an efficient choice for modern photonic systems.

One of the most significant applications of Si photonics is optical interconnects, a market witnessing unprecedented growth

TABLE I. Comparison of integrated lasers on Si PICs.

Type	Threshold	Optical power (mW)	Linewidth (kHz)	Footprint (mm^2)	Integration
Si Raman[11]	20 mW (optical)	50	<100	~300	Monolithic
InAs/GaAs QD FP[35]	65 mA (electrical)	6.8	···	~9×10^{-3}	Monolithic
InAs/GaAs QD DFB[36]	20 mA (electrical)	4.4	480	~5×10^{-3}	Monolithic
InGaAsP QW DFB[37]	10.5 mA (electrical)	>12	···	~4×10^{-4}	Heterogeneous
InAs/GaAs QD comb[32,38]	20 mA (electrical)	~1	···	~3×10^{-2}	Heterogeneous
InGaAsP QW comb[39]	60 mA (electrical)	~0.6	<400	~0.4	Heterogeneous
InAs/GaAs QD ring[40]	2.4 mA (electrical)	~0.6	···	~2.5×10^{-3}	Heterogeneous
InP DBR[41]	23 mA (electrical)	2	···	~0.375	Hybrid

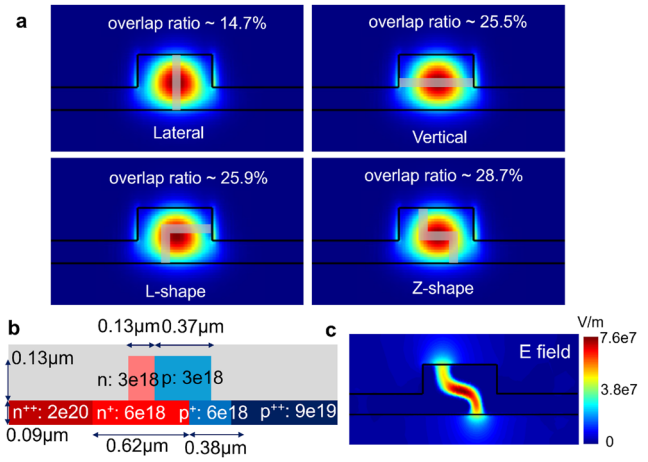

FIG. 2. (a) Overlap between the optical TE mode in a Si waveguide and the depletion regions of various junction configurations. (b) Doping profile of the Z-shaped junction achieved through six standard implantation steps. (c) Electric field distribution at −3 V corresponding to the Z-shaped junction.[13]

fueled by the need for ultra-fast, energy-efficient data transmission in response to the insatiable demands of artificial intelligence (AI). To satisfy the speed requirement, P–N junction-based Si MRMs, known as depletion mode MRMs, are a suitable choice. They mitigate the minority carrier lifetime limitation of the P–I–N junction-based MRMs (injection mode)[43,44] while avoiding high capacitance associated with the oxide capacitor-based MRMs (accumulation mode).[45,46] The price of the high bandwidth is the relatively weaker modulation efficiency compared to other types of Si MRMs, which remains a key bottleneck of the depletion mode MRMs. To address this, various junction configurations have been explored to improve the overlap between optical mode and carrier density variation, including the lateral junction,[12,47] vertical junction,[48] L-shaped junction,[49] and Z-shaped junction.[13] As shown in Fig. 2(a), advanced junction configurations can substantially enhance the overlap between the optical mode and the depletion regions. For example, the Z-shaped junction achieves nearly twice the overlap integral of the conventional lateral junction, enabling a larger Δn for the same carrier density variation. The Z-shaped MRM in Ref. 13 achieves a modulation efficiency $V_\pi \cdot L$ of around 0.6 V cm, a ~ 67% improvement over the lateral junction MRM's 1.0 V cm in Ref. 12. On the other hand, extended junction configurations result in increased capacitance, preserving the trade-off between modulation efficiency and RC time-limited bandwidth. Fortunately, the flexible doping concentrations enabled by the Z-shaped junction offer a solution to alleviate the trade-off: the slab region can be doped more heavily to reduce series resistance, while the top core region can be doped slightly less to strike a balance between modulation efficiency and free carrier absorption loss, as illustrated in Fig. 2(b). Its corresponding electric field distribution at −3 V is depicted in Fig. 2(c), where the depletion region is well-aligned with the region of the highest optical mode density. By reducing series resistance, the Z-shaped junction enhances its RC bandwidth despite having a larger capacitance. Compared to the lateral junction MRM in Ref. 12, the Z-shaped MRM in Ref. 13 achieves an ~30% increase in RC bandwidth. In essence, the Z-shaped junction design redistributes the doping profile to maximize the overlap between carrier modulation and the optical mode while preserving a comparable RC bandwidth. This enables a theoretical enhancement in modulation efficiency of up to twofold without compromising bandwidth, thereby further mitigating the fundamental limitations imposed by the inherently weak plasma dispersion effect in Si.

In addition, a two-segment design is incorporated into the Z-shaped MRM to further enhance its performance, as shown in Fig. 3(a). The MRM consists of two independent junctions with a length ratio of ~1:2, corresponding to the least significant bit (LSB) and the most significant bit (MSB). This configuration enables the MRM to function as an optical digital-to-analog converter (DAC), capable of generating four-level pulse amplitude modulation (PAM4) directly within the modulator. By eliminating the need for a power-intensive electronic DAC, the Z-shaped MRM simplifies the generation of high-speed, equally spaced PAM4 signals. As a result, PAM4 modulation can be achieved using two

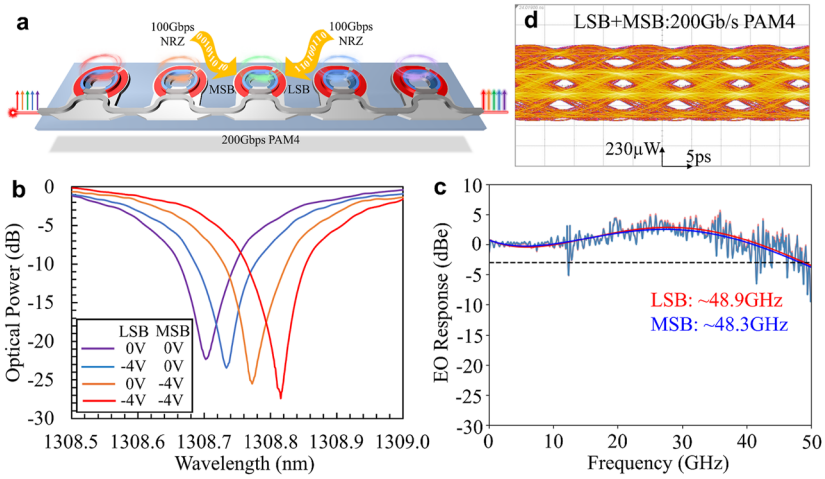

FIG. 3. (a) Schematic diagram of the five-channel Z-shaped MRMs featuring a two-segment design. (b) Measured transmission spectrum of the MRM under four different bias conditions. (c) Measured frequency response of the MRM's LSB and MSB segments. (d) Measured 200 Gb/s eye diagram combing LSB and MSB NRZ eyes for PAM4 modulation.[13]

basic non-return-to-zero (NRZ) signals without requiring complex equalization. Furthermore, the two-segment design offers an additional advantage by further enhancing the RC bandwidth. While the product of junction capacitance and series resistance remains constant with respect to junction length, shorter junctions enhance the RC bandwidth of the overall equivalent circuit when accounting for parasitic parameters. Compared to a single-segment design, the RC bandwidth increases by >20%. The two-segment, Z-shaped MRM can thereby achieve state-of-the-art performance. As shown in Fig. 3(a), a five-channel DWDM MRM array has been successfully demonstrated. By independently applying voltages to the LSB and MSB segments, four equally spaced wavelength shifts are achieved, as depicted in Fig. 3(b). The enhanced RC bandwidth enables both LSB and MSB junctions to achieve an overall bandwidth >48 GHz, as shown in Fig. 3(c). This allows each MRM to deliver a clear 200 Gb/s PAM4 eye diagram with a transmitter dispersion eye closure quaternary (TDECQ) of ~ 1 dB at a symbol error rate of 10^{-2}, resulting in a total modulation data rate of 1 Tb/s for the array. This DWDM Si MRM array exhibits minimal channel crosstalk, measuring < −33 dB between the two closest channels without thermal tuning. With thermal tuning, the crosstalk can be further reduced to approximately −57 dB. Leveraging the efficient Z-shaped junction and two-segment design, the MRM achieves exceptional energy efficiency with an energy consumption of 6.3 fJ/bit.

C. Si Mach–Zehnder interferometers

Si MZMs, particularly traveling-wave variants, have been widely adopted in commercial Si photonics transceivers developed by companies such as Cisco, Marvell, and Intel.[50] Their ability to support high-speed data rates of up to 200 Gb/s, combined with mature fabrication processes and robust performance, has made them a reliable solution for high-performance optical communication systems. However, Si MRMs offer several compelling advantages over MZMs, including a substantially smaller footprint, lower power consumption, and inherent compatibility with DWDM. These characteristics make MRMs particularly well-suited for scalable, high-throughput photonic systems, such as photonic chip I/O. Nonetheless, Mach–Zehnder interferometers (MZIs) continue to offer distinct advantages in specific application domains. Their superior linearity and broader dynamic range make them ideal for scenarios requiring high signal fidelity, such as advanced modulation formats and precision analog photonic signal processing. This advantage arises from their elegant transfer function,

$$U_{MZI} = \begin{bmatrix} e^{i\theta}(e^{i\phi} - 1) & ie^{i\theta}(1 + e^{i\phi}) \\ i(e^{i\phi} + 1) & 1 - e^{i\phi} \end{bmatrix}, \tag{2}$$

where ϕ and θ represent the phase shifts within and after the MZI arms, respectively. It allows 2×2 unitary transformation on the input state, serving as a fundamental building block for constructing any $m \times m$ unitary matrix. Through singular value decomposition (SVD), MZI meshes integrated with optical attenuation or amplification devices can execute matrix multiplication, making MZI a core component of coherent optical neural networks (ONNs).[51,52] ONNs are another potential application that has been revitalized in the last decade with the development of PICs. Leveraging the high bandwidth of photonic devices (tens of GHz), ONNs can, in principle, operate up to two orders of magnitude faster than electronic neural networks with significantly reduced latency.

Despite the sinusoidal transfer function of the MZI being more linear than the Lorentzian-shaped transfer function of the MRM, it is constrained by limited bit resolution. Consequently, power-hungry DACs remain necessary, undermining the potential energy efficiency and latency advantages of ONNs. To address this, a passive nonlinear device, an extremely overcoupled microring, has been integrated into the MZI to reshape the transfer function, as shown

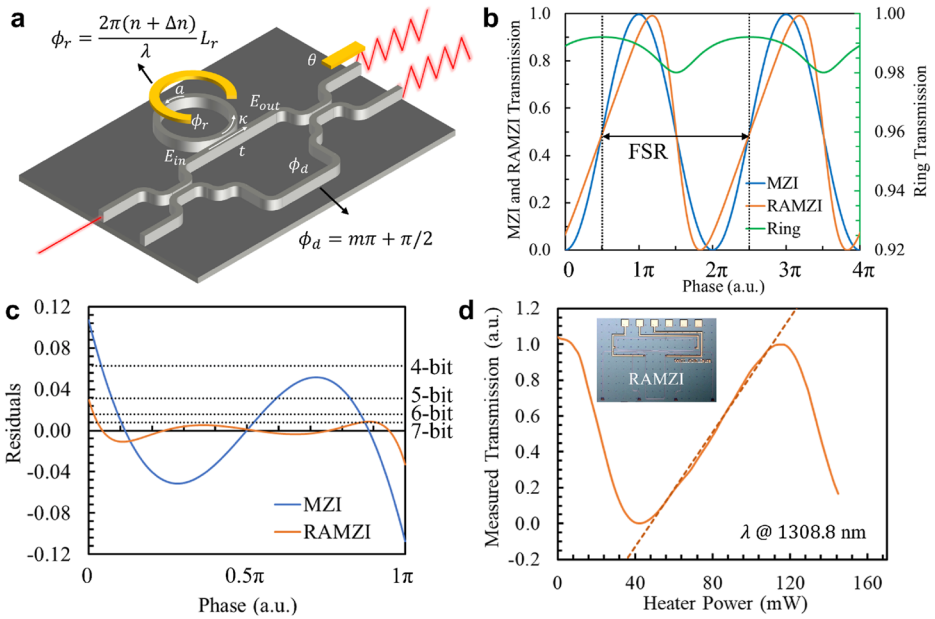

FIG. 4. (a) Schematic diagram of the RAMZI. (b) Transmission spectrum of the MZI (blue), RAMZI (orange), and its overcoupled microring (green). (c) Residuals from linear regressions of the MZI and RAMZI transitions from levels "0" to "1." (d) Measured transmission as a function of phase (heater power) for the RAMZI.[14,54]

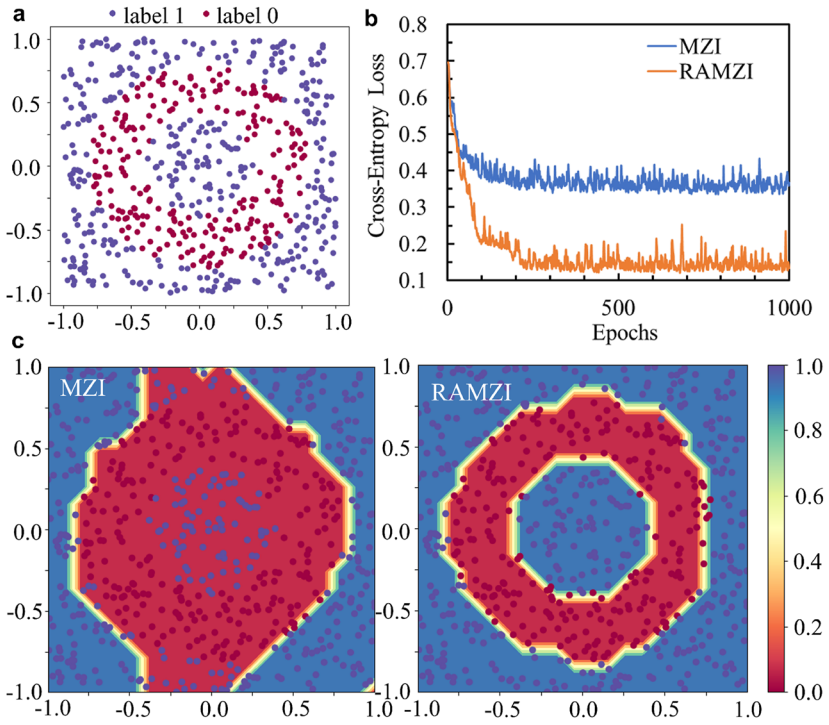

FIG. 5. (a) Randomly generated ring-shape planar dataset. (b) Cross-entropy loss $\mathcal{L}$ vs training epochs. (c) Predicted decision boundaries for the MZI and RAMZI meshes on the planar dataset.[14,54]

in Fig. 4(a). The overcoupled microring is designed to produce a substantial phase change with a minimal impact on amplitude. The transfer function of the all-pass microring is

$$U_R = e^{i(\pi+\phi_r)}(a - te^{-i\phi_r})/(1 - ate^{i\phi_r}), \tag{3}$$

where ϕ_r is the round trip phase change of the microring, a represents the round trip field transmission of the microring, and t signifies the field transmission of the bus-microring coupler. Consequently, the transfer function of the ring-assisted MZI (RAMZI) is

$$U_{RAMZI} = \begin{bmatrix} e^{i\theta}(U_R - e^{i\phi_d}) & ie^{i\theta}(e^{i\phi_d} + U_R) \\ i(U_R + e^{i\phi_d}) & e^{i\phi_d} - U_R \end{bmatrix}, \tag{4}$$

where ϕ_d denotes the phase difference between the two arms of the MZI excluding the microring and θ represents the phase shift at output, as shown in Fig. 4(a). By introducing a quarter-period delay in the MZI arms, $\theta_d = m\pi + \pi/2$, where m is an integer, the gradual phase change slope of the microring (in the off-resonant region) aligns with the rising edge of the MZI's sinusoidal function. Here, the overcoupled microring is configured with a round trip field transmission a of 0.99 and an optimized coupler field transmission t of ~0.391. This optimization ensures that the phase variation of the microring reduces the phase change at the center of the rising edge compared to the sides, compensating for the nonlinearity of the sinusoidal function. The transmission spectra of the conventional MZI, the overcoupled microring, and the proposed RAMZI are shown in Fig. 4(b), where the RAMZI demonstrates a triangular-like shape. The linearity of the device is evaluated using linear regression between levels "0" and "1." The standard deviation of the transfer function for the RAMZI is significantly reduced, decreasing from ~0.043 in the conventional MZI to about 0.007, a reduction of over 6 × . The residuals from the linear regressions are plotted in Fig. 4(c). The RAMZI (orange curve) exhibits much smaller residuals, with most confined within 7-bit intervals, whereas the MZI (blue curve) achieves only about 4-bit resolution. A resolution of 6 or 7 bits is sufficient for deep learning models to achieve good accuracy without degradation.[53] The measured transmission curve of the experimentally demonstrated RAMZI is presented in Fig. 4(d), showing good linearity that aligns well with the theoretical predictions.[14]

The impact of the enhanced linearity was evaluated using a chip-level ONN simulator, Neuroptica.[55] A ring-shaped planar classification task, depicted in Fig. 5(a), was employed for testing the ONN. Two ONNs, each with five layers comprising 5 × 5 MZI Clements meshes, were compared. The only difference between the two was the fundamental building block: one used the conventional MZI, while the other employed the linearized RAMZI. The cross-entropy loss $\mathcal{L}$ as a function of training epochs is shown in Fig. 5(b). The RAMZI reduced $\mathcal{L}$ from ~0.36 to 0.14, demonstrating an improvement of >61%. The predicted decision boundaries of the MZI and RAMZI meshes are illustrated in Fig. 5(c), where the RAMZI demonstrates more accurate predictions. By incorporating the designed overcoupled microring, the Si MZI achieves significantly improved linearity, enhancing the bit resolution from 4 to 7 bits with the same amount of phase shifters. This improvement facilitates faster convergence and higher accuracy for ONNs and expands its applicability to a wide range of high-precision tasks.

D. Si O-band photodiodes

High-speed and high-responsivity photodetectors (PDs) are other critical components in optical interconnects.[56,57] Unfortunately, the intrinsic cutoff wavelength of the bulk Si is ~1.1 μm, making photodetection at telecommunication wavelengths dependent on the heteroepitaxial growth of germanium (Ge). However, this growth often suffers from high dark currents and reliability challenges owing to defects at the Si/Ge interface.[58,59] Furthermore, the high-temperature epitaxial growth of Ge is complex, costly, and prone to non-uniformity across the wafer. This process can contribute ~40% of the total chip cost and is not widely supported by many CMOS foundries. In light of these limitations, there has been growing interest in all-Si PDs for telecommunication wavelengths.[60–64] By reverse biasing the Si junction, the all-Si MRR PDs can absorb sub-bandgap wavelengths >1.1 μm to generate considerable responsivity. Experimental analysis reveals four contributing mechanisms: (1) photon-assisted tunneling (PAT), (2) two-photon absorption (TPA), (3) resonance enhancement, and (4) avalanche gain. The band diagrams of the Si junction are shown in Fig. 6(a). Increasing the reverse bias voltage steepens the depletion region, effectively reducing the triangular barrier width w_b. The PAT probability is exponentially increased with w_b,

$$T \approx exp\left(-\frac{4\sqrt{2m_e^*}}{3}\sqrt{E_b}w_b\right), \tag{5}$$

where m_e^* is the effective mass of electrons and E_b represents the energy barrier height, determined by the energy difference between the Si bandgap and the input photon energy. At high reverse bias, w_b is reduced, significantly increasing PAT and enabling carrier generation from sub-bandgap wavelengths.[65] Meanwhile, the high bias leads to a strong electric field. The electric field distribution of a Si lateral junction is shown in Fig. 6(b). At −6.4 V, the electric field exceeds 5×10^7 V/m, which is sufficient to trigger impact ionization. Therefore, avalanche gain amplifies the photocurrent. In addition, resonance enhancement in the MRR boosts the light intensity within the Si absorption waveguide, further improving the responsivity. Finally, the TPA mechanism, as discussed in Sec. II A, induces free carrier absorption in Si, which is typically detrimental to Si Raman lasers. However, on the detection side, TPA proves beneficial as it contributes to carrier generation, thereby improving the responsivity. Both TPA and avalanche gain have been experimentally verified. At −4 V, the responsivity vs input optical power is shown in Fig. 6(c), where the responsivity increases linearly, indicating that TPA contributes to the responsivity. In contrast, at −6.4 V, the responsivity shows a saturation trend with increasing optical power, as illustrated in Fig. 6(d). This is attributed to the saturation of avalanche gain, which dominates the observed trend. Quantitatively, the combined contributions of PAT and TPA result in ~1% absorption per round trip at the O-band. This absorption is further enhanced by multiple round trips due to resonance and amplified through avalanche gain. With a high-Q cavity and optimized avalanche gain, the overall responsivity can be >65 A/W.[63] The responsivity of the Si MRR APD can be expressed as[15]

$$\begin{aligned} R &= RE \cdot M \cdot \eta \frac{q}{h\upsilon}, \\ \eta &\approx \frac{\alpha_p}{\alpha_{tot}}\left[1 - \exp\left(-\delta_r\right)\right] \approx \frac{\alpha_p}{\alpha_{tot}}\delta_r, \\ \alpha_p &= \Gamma \cdot \alpha_t + \alpha_2 = \Gamma \cdot \alpha_t + \beta_2 \frac{RE \cdot P_i}{A}, \end{aligned} \tag{6}$$

where RE is the resonance enhancement, M is the avalanche gain, η denotes the internal quantum efficiency for a single round trip, $h\upsilon$ is

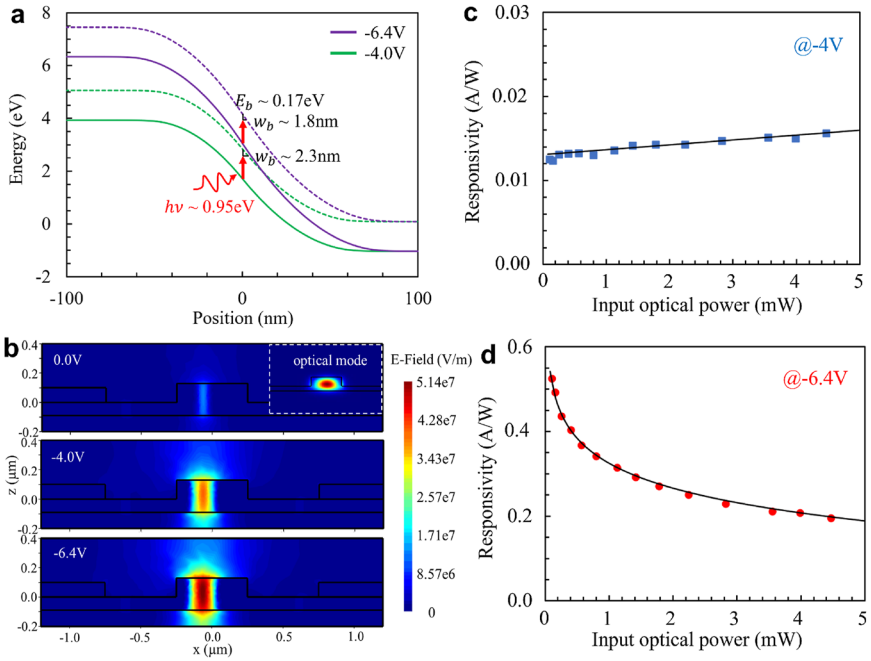

FIG. 6. (a) Energy band diagrams, (b) electric field distribution, and (c) and (d) measured responsivity as a function of input optical power at bias voltages of −4 V and −6.4 V, respectively.[15]

photon energy, α_p is the absorption coefficient contributing to photocurrent, α_{tot} is the total absorption coefficient, and δ_r is the MRR propagation loss. The α_p comprises two parts: $\Gamma \cdot \alpha_t$ for PAT and α_2 for TPA, where Γ is the optical mode-depletion mode confinement factor, β_2 is the TPA constant, P_i is the input optical power, and A is the effective waveguide cross-sectional area.

Nevertheless, these MRR APDs face two inherent trade-offs: one between bandwidth and responsivity and the other between channel capacity and crosstalk. To address these challenges, we demonstrated a novel all-Si receiver (Rx) comprising eight double-MRR APDs.[16] As illustrated in the schematic of a double-MRR APD in Fig. 7(a), the device comprises two microring resonators of identical radii. The first resonator is undoped, ensuring low propagation loss, while the second resonator incorporates a P–N junction that functions as a PD. It should be noted that, apart from the lateral junction, a Z-shaped P–N junction can also be implemented in the second resonator. This alternative design has been characterized to provide enhanced modal overlap with the optical field. To analyze the structure, the small-signal response of the double-MRR PD can be derived from rate equations, as shown in Ref. 66. The system exhibits two complex poles, which contribute to frequency response peaking and enhance the bandwidth of the double-MRR

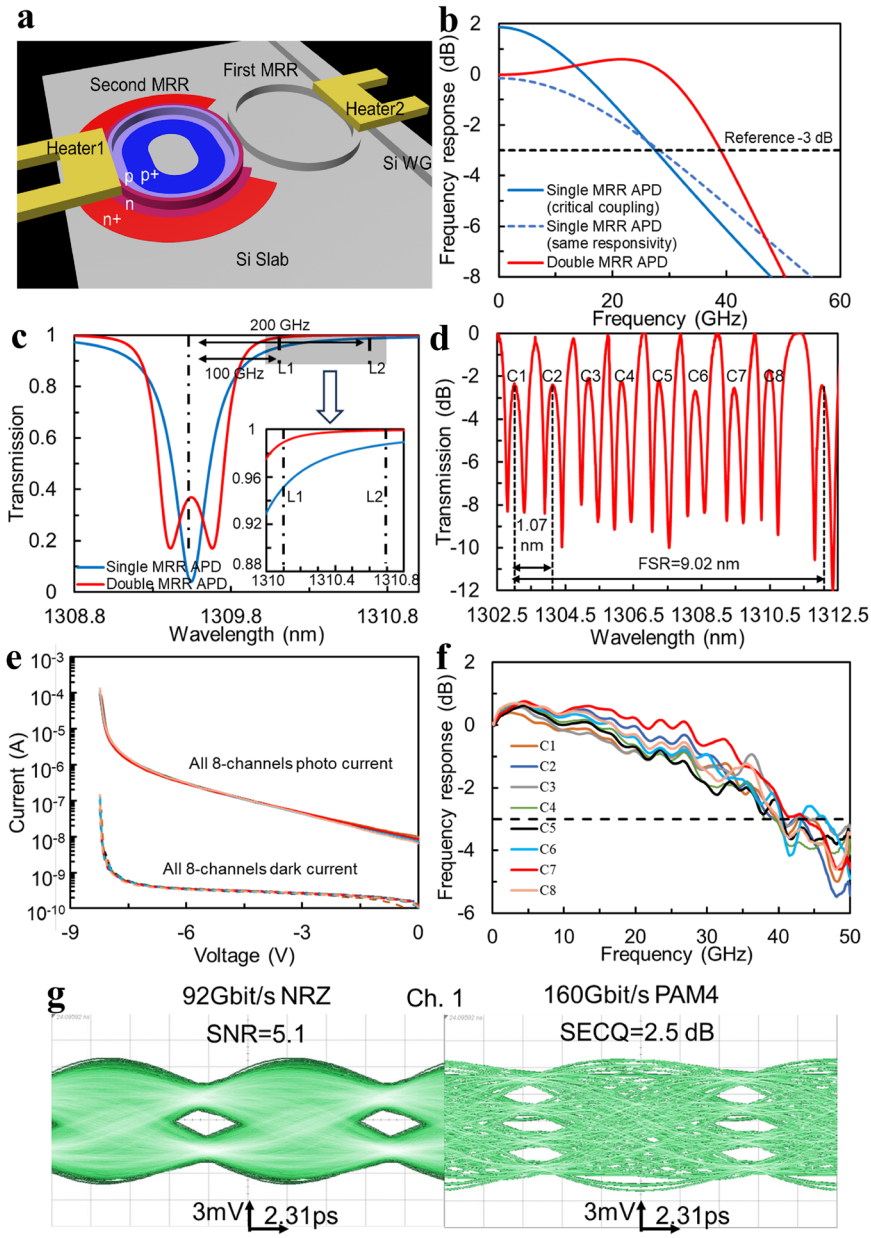

FIG. 7. (a) 3D schematic of the double-MRR APD. (b) Simulated frequency response. (c) Simulated spectrum and crosstalk of double- and single-MRR APD; inset: enlarged spectrum. (d) Measured spectrum. (e) Measured photocurrent and dark current for eight channels. (f) Measured frequency response for eight channels. (g) Measured 92 Gb/s NRZ and 160 Gb/s PAM-4 eye diagrams.[16]

PD. Under high bias conditions, the avalanche buildup time is included to account for the effects of impact ionization. At −8 V and a gain of 6, As depicted in Fig. 7(b), the double-MRR APD achieves a bandwidth of 40 GHz, representing a 37% improvement over the 29 GHz bandwidth of the single-MRR APD at the same responsivity. In addition to the speed benefits, the double-MRR structure significantly improves the sharpness of the response roll-off and neighboring channel rejection, as illustrated in Fig. 7(c). The enhanced transmission spectrum at the through port demonstrates that the double-MRR structure achieves electrical channel crosstalk levels of −41 to −66 dB, compared to −26 to −37 dB for the single-MRR structure, for channel spacings ranging from 100 to 200 GHz. For example, to enable −40 dB electrical crosstalk, channel spacing can be reduced from 245 to 96 GHz, enabling nearly three times more channels to fit within a given free spectral range. With its broadband spectrum, high spectral efficiency, and excellent neighboring channel rejection, the double-MRR structure is well-suited for DWDM Rx applications.

The all-Si APDs were fabricated using a standard Si photonics process at the Advanced Micro Foundry, ensuring full compatibility with standard MPW runs without requiring any process modifications. Figure 7(d) shows the measured transmission spectrum of the eight-channel Rx with heater adjustment to compensate for the fabrication error and match the resonance. The MRR Rx features a free spectral range of 9.02 nm and a channel spacing of 1.07 nm. Neighboring channel crosstalk was measured to be less than −50 dB across all frequencies, which is effectively negligible. The I–V characteristics of the APDs were measured at the center resonant wavelength, as illustrated in Fig. 7(e). Leveraging the mature silicon process, all eight channels exhibit excellent uniformity, with dark currents as low as ∼1 nA at a responsivity of 0.4 A/W at −8 V. This dark current is two orders of magnitude lower than that of conventional Si–Ge APDs with similar responsivity. This improvement arises from the use of an all-Si structure, which inherently contains fewer defects. In contrast, Si and Ge suffer from a lattice mismatch of ∼4.2%, leading to increased defect densities and thereby higher dark current, particularly under high electric fields in the APD gain region. In addition to suppressing channel crosstalk, the double-MRR structure also delivers high bandwidth performance. Figure 7(f) shows the measured frequency response for all eight channels, which also exhibit excellent uniformity. At a gain of 5.9 (−8 V), all devices achieve a 3 dB bandwidth of ∼40 GHz, nearly double the bandwidth of single-MRR APDs.[67] The experimental results align closely with theoretical predictions, and the gain-bandwidth product (GBP) is measured at 236 GHz at −8 V. To further investigate the dynamic properties of APD, we measure NRZ and PAM4 eye diagrams with the eight-tap feed-forward equalization (FFE) without a *trans*-impedance amplifier. Figure 7(g) shows the performance of one channel as a reference. For the NRZ signal, the SNR of all channels is >4.9, which is affected by the circuit noise from the oscilloscope and the limited sample rate of the arbitrary waveform generator. At a forward error correction (FEC) threshold of 3.8×10^{-3}, the sensitivity of this all-Si APD is approximately −2 dBm for 92 Gbit/s NRZ signals. To further enhance high-speed signal reception capabilities, we explored the PAM4 format. The stress eye closure quaternary (SECQ) metric was used to calculate the penalties for receivers. In this study, the SECQ was measured at the soft decision FEC threshold of symbol error rate of 4.8×10^{-4}. The SECQ penalties for the 160 Gb/s PAM4 eye diagrams range between 2.5 and 5.2 dB. Overall, this eight-channel receiver supports a total data rate of 1.28 Tb/s per fiber, demonstrating the potential of the all-Si double-MRR APD for next-generation 1.6 Tb/s Ethernet applications.

E. Si optical memories

On-chip non-volatile optical memories have long been a coveted breakthrough for PICs, holding the promise of revolutionizing data storage and processing with unparalleled speed and efficiency. While previous efforts have focused on integrating advanced materials such as phase-change materials (PCMs),[68,69] ferroelectric materials,[70,71] and memristors[72,73] into the Si photonics platform, these approaches come with significant trade-offs. They demand additional fabrication steps, escalating complexity, cost, and yield challenges while posing risks to reliability and process compatibility. In a bold departure from these constraints, we introduce, for the first time to our knowledge, an all-Si non-volatile optical memory.[17,18] This innovation requires zero changes to existing foundry processes, offering a game-changing solution that combines scalability, simplicity, and unprecedented integration potential, heralding new potential for Si photonics.

The Si optical memory is implemented using a MRR structure, as shown in Fig. 8(a), where resonance amplifies the non-volatile changes in its optical spectrum. Similar refractive index non-volatile changes can also be achieved in other device structures. Unlike existing optical memory mechanisms, this device leverages a brand new mechanism named optical avalanche-induced trapping memory (ATM). The device is based on a standard Si P–N junction; its cross section is illustrated in Fig. 8(c). The design employs a Z-shaped doping configuration, akin to that used in modulators, but operates under distinctly different conditions. To set the memory, the junction is reverse-biased near its avalanche breakdown region, with a laser injected to generate a substantial photocurrent of ∼100 μA. Under this condition, traps at the Si–SiO_2 interface gradually capture excess holes produced by the unbalanced carrier dynamics intrinsic to Si. As depicted in Fig. 8(b), these interface traps result from the imperfect transition between crystalline Si and amorphous SiO_2. To reset the memory, the junction is forward-biased without laser illumination, generating a forward current of several hundred μA. This condition equalizes the concentrations of p-type and n-type carriers, thereby emptying the filled traps at the interface. The set and reset behaviors of the memory can be mathematically described using the following rate equation:[74]

$$\frac{dF_{tD}}{dt} = v_P\sigma_P\left[P(1-F_{tD}) - F_{deg}F_{tD}n_i\,\exp\left(\frac{E_{tD}-E_i}{k_BT}\right)\right] - v_N\sigma_N\left[NF_{tD} - \frac{(1-F_{tD})n_i}{F_{deg}}\exp\left(\frac{E_i-E_{tD}}{k_BT}\right)\right], \quad (7)$$

where F_{tD} is the trap occupation probability, v_P and v_N are the thermal velocities of holes and electrons, while σ_P and σ_N represent their respective trap capture cross sections. P and N are the concentrations of free holes and electrons, and F_{deg} is the degeneracy factor, n_i indicates the intrinsic carrier concentration, corresponding to the energy levels of donor-like traps and intrinsic carriers, respectively.

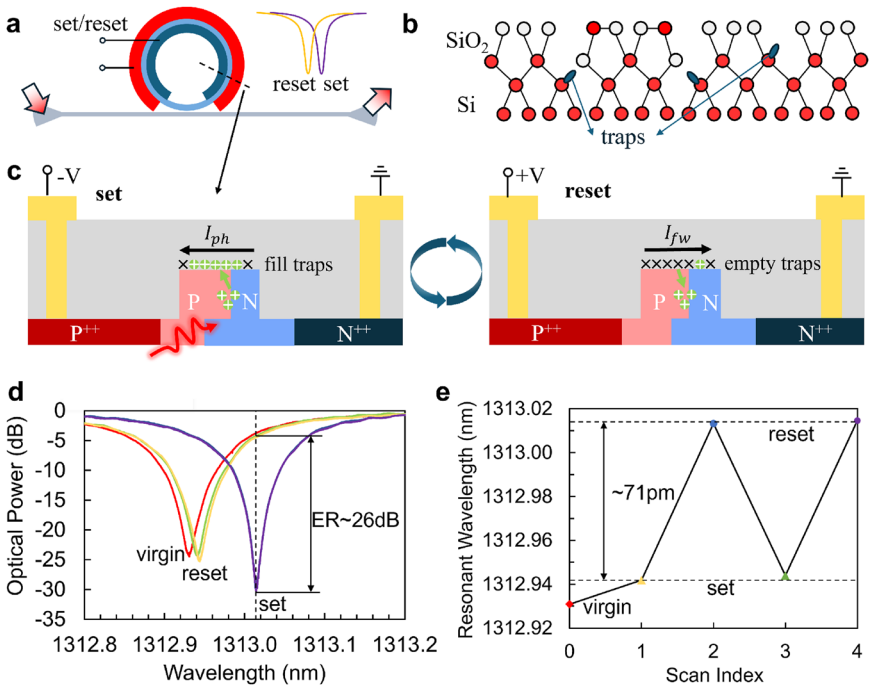

FIG. 8. (a) Schematic diagram of the all-Si non-volatile optical memory. (b) Illustration of traps at the Si–SiO_2 interface. (c) Cross section of the device. (d) Measured optical spectrum and (e) resonant wavelength of the Si optical memory under virgin, erase, and program states.[17,18]

The detailed analysis is provided in Ref. 17. The charged and discharged traps cause a redistribution of doping within the Si P–N junction, which alters the refractive index via the plasma dispersion effect. The resulting non-volatile shifts in the optical spectrum and resonant wavelength are illustrated in Figs. 8(d) and 8(e), showing a resonant wavelength shift of ~71 pm, corresponding to an extinction ratio (ER) of about 26 dB.

Further characteristics of the Si ATM are illustrated in Fig. 9. Since the traps gradually capture excess holes, different setting times result in varying resonant wavelength shifts. As shown in Fig. 9(a), the resonant wavelength of the ATM redshifts progressively with increased setting time, enabling the optical memory to exhibit multiple non-volatile states. Specifically, it supports at least 26 distinct states with a standard deviation of ~0.004, as depicted by the optical

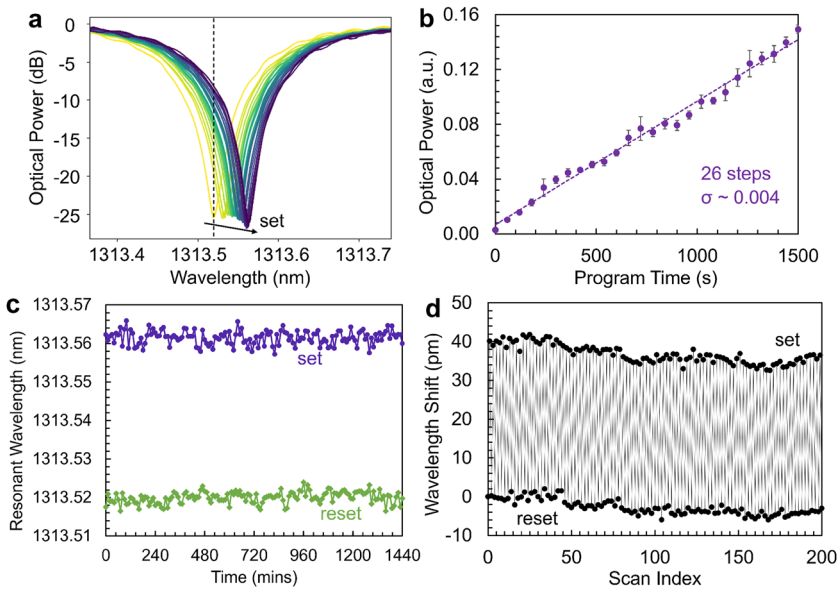

FIG. 9. (a) Measured spectrum of the Si optical memory vs setting time. (b) Demonstration of multiple memory states, (c) retention characteristics, and (d) cycling performance of the Si optical memory.[17,18]

power changes in Fig. 9(b). In addition to multi-state functionality, the Si ATM demonstrates excellent retention and cycling performance. Both set and reset states remain stable without any decay over 24 h, and up to 100 cycles of set–reset switching have been successfully demonstrated. The endurance measurement is primarily constrained by the relatively slow speed, which is currently the main limitation of this Si optical memory. Enhancing the oxide conditions atop the Si waveguide and introducing a vertical electric field could improve the rate dF_{tD}/dt in Eq. (7), thereby increasing the operational speed. This first demonstration of the Si ATM presents a most convenient method for integrating non-volatile memory into Si photonics with competitive performance. While the speed remains a bottleneck, the memory can be directly applied in existing Si PICs for applications that do not require frequent switching, such as system trimming and inference in ONNs. This innovative optical memory also holds great potential for further optimization in the future.

III. OPTICAL INTERCONNECTS

Driven by the ever-growing computational demands of the AI era, the need for high-speed and seamless communication across multi-node, multi-chip systems has become increasingly critical. Traditional electrical interconnects are no longer sufficient to address the escalating requirements for bandwidth, density, and energy efficiency. Si photonics have been recognized as a natural solution for chip interconnects, offering higher speed and compatibility with CMOS technology. The adoption of the co-packaged optics (CPO) approach further elevates this technology by merging electronics and photonics onto a shared substrate, significantly reducing link lengths, enhancing data density, and minimizing latency and power consumption.[75] Yet, Si alone is not a universal solution for all photonic device functionalities. Considerable efforts have been made in this direction through hybrid, monolithic, and heterogeneous integration methods. However, the current integration techniques are not without challenges; they are constrained in the variety of materials they can accommodate. Complex processes such as multiple bonding and selective regrowth add layers of fabrication difficulty, driving up costs and diminishing reliability. These methods also face challenges such as limited compatibility with high-temperature processes, increased device footprints, and reduced yield. Consequently, while Si photonics presents immense promise, achieving its full potential requires innovation to simplify and refine the integration of material systems.

In response to this challenge, we have developed novel designs and mechanisms that enable Si alone to efficiently perform various key functionalities such as modulation, detection, and non-volatile shifts. Consequently, only a single bonding process, monolithic growth, or selective regrowth is necessary for the light source, with all remaining components implemented using Si. For example, the high-temperature growth of germanium for photodiodes (600–750 °C) often poses thermal budget challenges for bonding layers, an issue that can be mitigated by our all-Si approach. This solution significantly enhances the flexibility and adaptability of Si PICs while also improving yield and cost-efficiency. The process flow of Si PICs can be streamlined, as illustrated in Fig. 10. The optical interconnect can thus be revolutionized, as depicted in Fig. 11. A single quantum dot comb laser serves as the laser source,[76] with Si MRMs on the transmitter (Tx) side and Si double-ring photodiodes on the Rx side. Both devices inherently support DWDM functionality, eliminating the need for additional optical filters and further enhancing circuit density. In addition, thanks to the universal Si device design,[77] both modulators and photodiodes have the potential for non-volatile wavelength shifts. This capability is particularly beneficial for device trimming in DWDM systems, where the resonant wavelengths of microrings vary due to unavoidable fabrication variances. It will reduce the constant

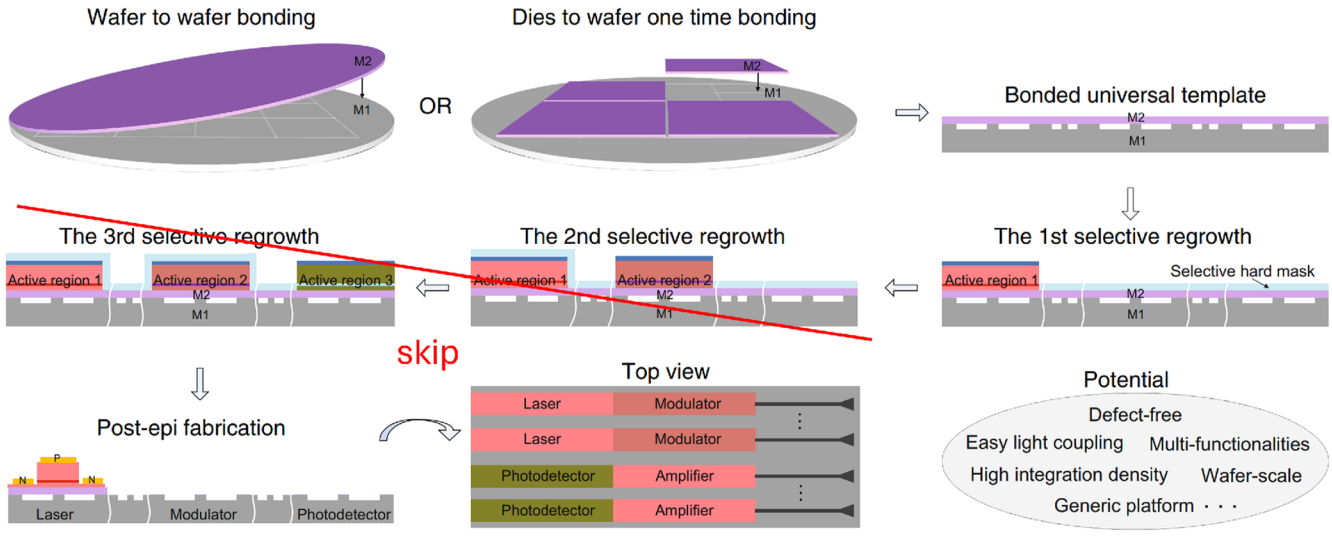

FIG. 10. Schematic of the heterogeneous integration process for lasers with all-Si technologies, minimizing the need for multiple bonding steps or epitaxial growth.[26]

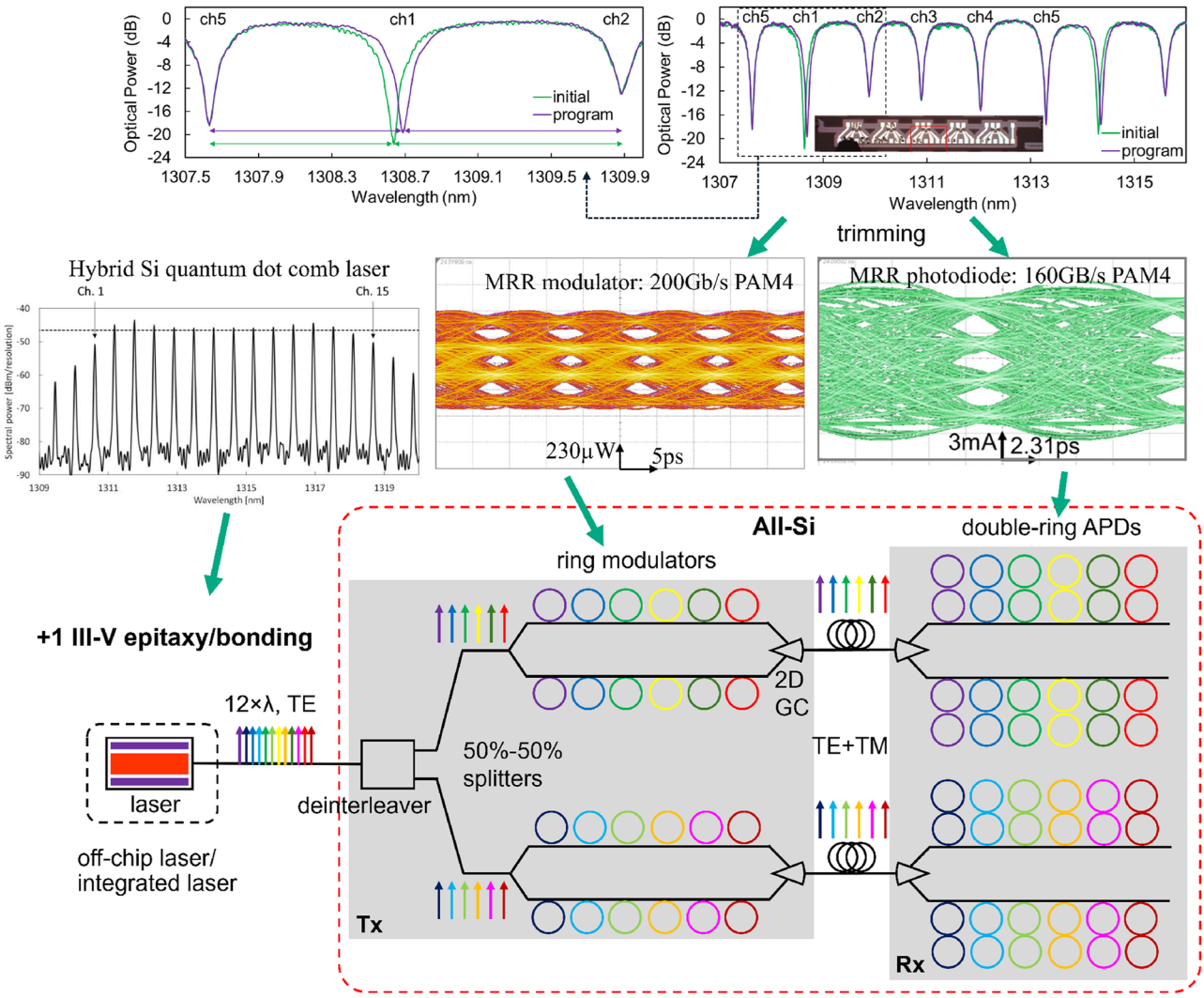

FIG. 11. Optical interconnect architecture based on a hybrid Si quantum dot comb laser source and all-Si TRX components: Si microring modulators, Si double-ring avalanche photodiodes, and Si optical memory for DWDM trimming.[13,16,17,76,78]

thermal energy consumption for the wavelength offsets. As demonstrated at the top of Fig. 11, a five-channel DWDM MRM array was used to showcase this trimming capability. By setting the channel 1 MRM, the channel spacing difference between ch5–ch1 and ch1–ch2 is reduced by about 0.1 nm, making a more uniform channel spacing, and the channel crosstalk improves by about 2.5 dB. The eye diagrams of the Si MRM and double-ring photodiode are shown in Fig. 11, supporting PAM4 data rates of at least 160 Gb/s. Utilizing the power of polarization division multiplexing (PDM) and DWDM, the optical link will support an overall data rate exceeding 3.6 Tb/s.

To verify practical implementation, we have demonstrated Tx and Rx with Si PICs using wire bonding, as shown on the left of Fig. 12. The Tx features a 28 nm quarter-rate CMOS chip integrated with two-segment Si MRMs. An AC-coupled pulse-based driver supplies a high swing voltage of 3.42 V_{ppd} to both segments, enabling independent level tuning, edge-rate control, and asymmetric Feed-Forward Equalization (FFE). This configuration achieves a clear eye diagram at 80 Gb/s, with a TDECQ of 1.75 dB, an outer ER of 5.4 dB, and an energy efficiency of ~3.66 pJ/bit.[79] The Rx similarly integrates a 28 nm CMOS chip with the Si PIC through wire bonding. The output photocurrent from the photodiode flows into an analog front end (AFE), comprising a transimpedance amplifier (TIA), continuous-time linear equalizer (CTLE), and variable gain amplifier (VGA). Following this, clock-driven Schinkel comparators digitize the PAM4 signals using three comparators for analog-to-digital conversion. At a data rate of 56 Gb/s for PAM4 signals, the Rx achieves an optical modulation amplitude (OMA) of −7 dBm at

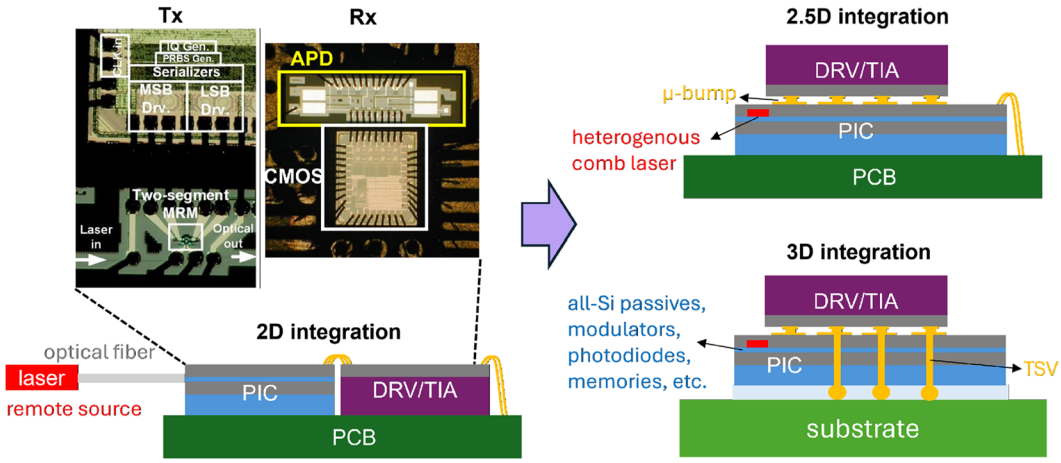

FIG. 12. Schematic and demonstration of 2D integration of Si photonics for optical interconnects (left)[79,80] and 2.5D and 3D heterogeneous integration of Si photonics for optical interconnects (right).

a bit error rate (BER) of 10^{-4}, with an energy efficiency of about 1.61 pJ/bit.[80] While effective, this 2D integration is limited by parasitics introduced through wire bonding, restricting data rates. Advanced integration techniques, such as 2.5D integration with flip-chipped CMOS or interposers[50,81] and 3D integration using through-silicon vias (TSVs),[82] as shown on the right of Fig. 12, can overcome these limitations. These approaches bring the CMOS chip closer to the PIC, minimizing parasitics and enabling higher speeds for datacom and telecom applications.

In such electronic–photonic integrated circuits (EPICs), heat generated by the laser sources, drivers, and TIA circuits induces resonant wavelength shifts in the MRRs, one of the primary challenges for the MRR-based DWDM system. The thermo-optic coefficient of Si is $\eta_{Si} \approx 1.95 \times 10^{-4}$/K for the O band and $\eta_{Si} \approx 1.85 \times 10^{-4}$/K for the C band. In comparison, the coefficient of silicon oxide is $\eta_{SiO\_2} \approx 1 \times 10^{-5}$/K for oxide cladding layers.[83] Due to the strong optical confinement within the Si core and the fact that η_{SiO_2} is more than an order of magnitude smaller than η_{Si}, the contribution of the silicon dioxide cladding to the overall thermo-optic response is negligible. Based on these values, the effective index shift of the Si waveguide with temperature, $\partial n_{eff}/\partial T$, can be calculated. For MRR-based devices, the corresponding shift in resonant wavelength with respect to temperature is given by[84]

$$\frac{d\lambda}{dT} = \left(n_{eff}\alpha_{sub} + \frac{\partial n_{eff}}{\partial T}\right)\frac{\lambda_0}{n_g}, \tag{8}$$

where α_{sub} is the substrate expansion coefficient and n_g is the group index. For an MRR using a 220 nm-thick, 500 nm-wide Si waveguide, the temperature-dependent resonance shift in the O-band is approximately $d\lambda/dT \approx 60$ pm/K, making MRR devices incompatible with the typical temperature range of microelectronic environments. To address this issue, feedback control systems incorporating optical power monitors and integrated heaters are typically employed to dynamically track and stabilize the resonant wavelength. Various approaches have been proposed such as proportional-integral-derivative (PID) controllers, OMA-based controllers, and balanced homodyne CMOS control circuits.[85–87] As a result, MRR-based Si interconnects have been successfully implemented in commercial products.[88] From a system-level perspective, this thermal shift issue is manageable, albeit with certain overheads, such as additional footprint for optical monitors, extra pads for heaters, increased circuit complexity, and higher power consumption. The proposed all-Si devices solution offers key advantages that help alleviate these overheads in the feedback control loop. For example, the integration of all-Si photodiodes enables *in situ* optical power monitoring without the need for heterogeneous materials. In addition, the incorporation of non-volatile Si memory can compensate for static resonance drift, thereby reducing the required heater power and improving overall system efficiency.

Consequently, the all-Si PIC platform provides further advantages by minimizing material heterogeneity, streamlining integration, and enabling versatile device functionalities. By employing a quantum dot comb laser as the sole heterogeneously integrated component, this approach eliminates the need for external laser sources, enabling compact, low-loss integration. As depicted in Fig. 11, this streamlined interconnect offers superior scalability, cost efficiency, and miniaturization. As optical interconnects are adopted over shorter distances, these advances position all-Si PIC as a critical enabler for chip-to-chip connectivity, addressing the bottlenecks of modern computing hardware.

IV. OPTICAL NEURAL NETWORKS

Optical neural networks (ONNs) are expected to outperform their electronic counterparts due to the low computational latency, ultra-high throughput, high energy efficiency, and high parallelism.[89] Many integrated ONNs have been reported, such as the MZI meshes,[51,52] MRR weight banks,[90] MRR crossbar,[91] directional coupler crossbar,[92] balanced homodyne detection,[93] and integrated chip diffractive neural networks.[94] However, the limited scalability prevents inference and training on real-size ONNs. This is due to

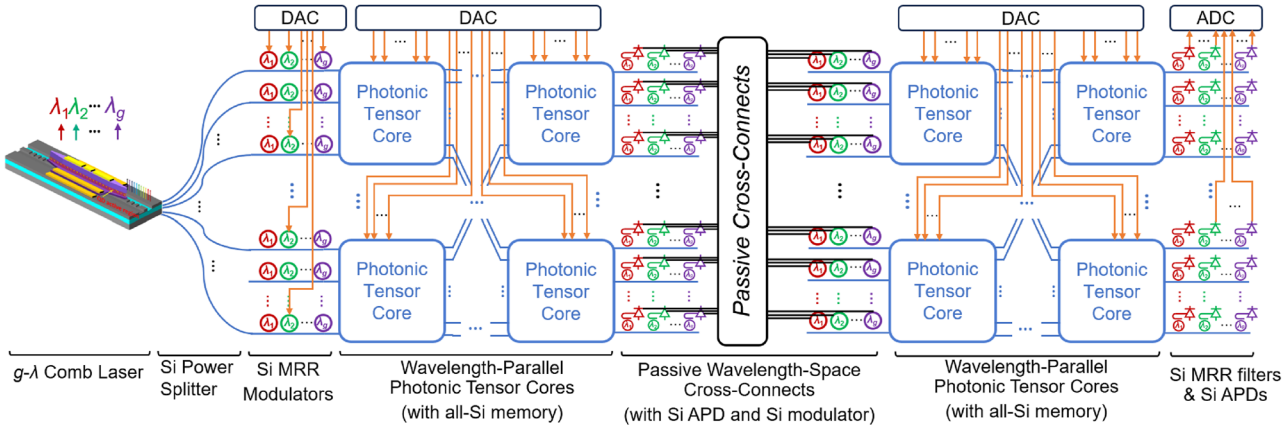

FIG. 13. TONN on all-Si photonic integrated circuit platform.[100]

photonic integrated circuits' large device footprints and low integration density. For instance, MRRs and MZIs are much larger (~10–100 s of micrometers) than CMOS transistors. A real-size ONN with $>10^5$ parameters can easily exceed the available chip size with the quadratic scaling rule, where an $N \times N$ weight matrix requires $O(N^2)$ MRRs or MZIs. As a result, the state-of-the-art photonic AI accelerator is 64×64.[95] As a result, large-scale optical matrices are computed by tiles or blocks with time multiplexing, demanding intensive memory access to store the intermediate data. That means E/O and O/E conversions and high-speed high-bit-resolution DACs/ADCs are involved during memory access, limiting the overall power efficiency. The study[96] shows that in an ONN, only ~10% of the overall power is consumed in optical devices, while the major part of the power is consumed in memory access and data movement.

At Hewlett Packard Labs, we address the scalability and partially address the associated memory issue by deploying hardware-algorithm co-design for in-memory computation. In detail, we have developed a scalable and energy-efficient tensorized optical neural network (TONN) architecture[97] by leveraging the tensor-train (TT) decomposition algorithm to compress the less important parameters in ONNs. In TONN, the large-scale matrix multiplications are emulated by representing the tensor indices in the wavelength and space domains and multiplying them with an array of small-scale (e.g., 8×8) wavelength-parallel photonic tensor cores based on either MZI meshes or MRR crossbar arrays,[98] as shown in Fig. 13. The TT decomposition enables scaling to 1024×1024 and beyond, which is extremely difficult for conventional ONNs. On the other hand, in-memory computation brings the data memory closer to where the computation happens, mitigating the bottlenecks of memory access in ONNs. For example, during the inference where high-speed input data multiplies with static optical weight matrices, photonic memory enables a true "set-and-forget" operation for the phase shifters in photonic tensor cores, eliminating the requirement of external memory to store the weight values. For the training of ONNs, assuming a weight stationary (WS) setting, the weight matrices are programmed into the phase shifters in the photonic tensor cores, and the input vectors are encoded in the high-speed (~10 GHz) optical signals. In each training iteration, the same weights (phases) are multiplied with batched (e.g., 1000[99]) input data. As a result, the update rate of the phase shifters is three orders of magnitude lower than that of the data encoding, so the power consumption and latency for each training iteration can be reduced by applying photonic memory to the trainable phase shifters.

The all-Si PIC platform is particularly suitable for implementing TONN with in-memory computation since it can integrate all the required devices in TONN with all-Si optical memory, as in Sec. II E. As shown in Fig. 13, a multiple-wavelength light source (either external or on-chip III–V-on-Si comb laser[32,101]) is split using a Si power splitter. The input data for the TONN are then encoded in the split light by Si DWDM MRM arrays.[12,13] After that, the encoded light multiplies with layers of wavelength-parallel photonic tensor cores enabled by MZI meshes[14] or MRR crossbar arrays. The phase tuners are implemented by on-chip non-volatile all-Si optical memories.[17,18] At the end of TONN, Si MRR filters acting as DWDM de-multiplexers and Si APDs[16] are used for data receiving. Such an all-Si platform exhibits superior power efficiency compared with other conventional photonic device platforms for the following reasons: (1) heterogeneous integration of all optical components eliminates the coupling loss between discrete chips; (2) all-Si phase tuners have negligible static phase tuning energy consumption, while other photonic technologies require power or voltage to maintain the phase tuning; and (3) the all-Si PIC enables the implementation of larger ONNs with improved uniformity and yield, easing scalability challenges in handling larger matrix multiplications. These advantages make the all-Si PIC platform a highly efficient and practical solution for large-scale TONNs and other advanced photonic computing applications.

V. CONCLUSION

This Perspective redefines the potential of all-Si photonics, showcasing its ability to transcend the limitations traditionally associated with the Si photonics platform. By innovating within the boundaries of Si, we demonstrate the cohesive integration of

optoelectronic devices, such as Raman lasers, microring modulators, Mach–Zehnder interferometers, avalanche photodiodes, and optical memories, all operating harmoniously within a monolithic framework. Through precise engineering of Si junctions and operating voltage ranges, Si is shown to perform a wide range of functions, from light amplification and modulation to detection and memory. These functionalities are further enhanced through optimized optical waveguide designs tailored to specific applications. These advancements chart a visionary course for the future of photonic systems, providing unparalleled scalability, energy efficiency, cost-effectiveness, and manufacturability without the intricacies of multi-material integration. Leveraging Si platforms wherever feasible also facilitates easier integration of EPICs, whether through monolithic or hybrid approaches. It reduces optoelectronic connection distances, enhancing package density and taking full advantage of light's inherent properties, such as low latency and high-speed transmission. The applications investigated, from optical interconnects to neural networks, reveal the untapped capabilities of an all-Si paradigm to address the pressing demands of the AI and high-performance computing eras. As this approach continues to evolve, it promises not only to elevate photonic integration to unprecedented levels but also to inspire a paradigm shift in how we envision the role of Si in shaping the next generation of computational and communication infrastructures.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Y. Yuan: Conceptualization (lead); Data curation (lead); Writing – original draft (lead). **Y. Peng**: Data curation (equal); Writing – original draft (equal). **S. Cheung**: Writing – original draft (equal). **X. Xiao**: Writing – original draft (equal). **W. V. Sorin**: Writing – review & editing (equal). **Z. Huang**: Writing – review & editing (equal). **D. Liang**: Writing – review & editing (equal). **A. Kumar**: Writing – review & editing (equal). **R. Liu**: Writing – review & editing (equal). **Y. Hu**: Writing – review & editing (equal). **S. Hooten**: Writing – review & editing (equal). **S. Palermo**: Writing – review & editing (equal). **M. Fiorentino**: Funding acquisition (equal); Project administration (equal); Writing – review & editing (equal). **R. G. Beausoleil**: Funding acquisition (equal); Project administration (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]H. Rong, H. Zhang, S. Xiao, C. Li, and C. Hu, "Optimizing energy consumption for data centers," Renewable Sustainable Energy Rev. **58**, 674–691 (2016).
[2]M. Minkov, P. Sun, B. Lee, Z. Yu, and S. Fan, GPU-accelerated photonic simulations (2024).
[3]G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, "Silicon optical modulators," Nat. Photonics **4**, 518–526 (2010).
[4]G. Li, A. V. Krishnamoorthy, I. Shubin, J. Yao, Y. Luo, H. Thacker, X. Zheng, K. Raj, and J. E. Cunningham, "Ring resonator modulators in silicon for interchip photonic links," IEEE J. Sel. Top. Quantum Electron. **19**, 95–113 (2013).
[5]W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," Laser Photonics Rev. **6**, 47–73 (2012).
[6]S. F. Preble, Q. Xu, and M. Lipson, "Changing the colour of light in a silicon resonator," Nat. Photonics **1**, 293–296 (2007).
[7]L. Zhang, M. Zhang, T. Chen, D. Liu, S. Hong, and D. Dai, "Ultrahigh-resolution on-chip spectrometer with silicon photonic resonators," Opto-Electron. Adv. **5**, 210100 (2022).
[8]Q. Qiao, X. Liu, Z. Ren, B. Dong, J. Xia, H. Sun, C. Lee, and G. Zhou, "MEMS-enabled on-chip computational mid-infrared spectrometer using silicon photonics," ACS Photonics **9**, 2367–2377 (2022).
[9]N. Margalit, C. Xiang, S. M. Bowers, A. Bjorlin, R. Blum, and J. E. Bowers, "Perspective on the future of silicon photonics and electronics," Appl. Phys. Lett. **118**, 220501 (2021).
[10]H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave Raman silicon laser," Nature **433**, 725–728 (2005).
[11]H. Rong, S. Xu, Y.-H. Kuo, V. Sih, O. Cohen, O. Raday, and M. Paniccia, "Low-threshold continuous-wave Raman silicon laser," Nat. Photonics **1**, 232–237 (2007).
[12]Y. Yuan, W. V. Sorin, Z. Huang, X. Zeng, D. Liang, A. Kumar, S. Palermo, M. Fiorentino, and R. G. Beausoleil, "A 100 Gb/s PAM4 two-segment silicon microring resonator modulator using a standard foundry process," ACS Photonics **9**, 1165–1171 (2022).
[13]Y. Yuan, Y. Peng, W. V. Sorin, S. Cheung, Z. Huang, D. Liang, M. Fiorentino, and R. G. Beausoleil, "A 5 × 200 Gbps microring modulator silicon chip empowered by two-segment Z-shape junctions," Nat. Commun. **15**, 918 (2024).
[14]Y. Yuan, S. Cheung, T. Van Vaerenbergh, Y. Peng, Y. Hu, G. Kurczveil, Z. Huang, D. Liang, W. V. Sorin, X. Xiao *et al.*, "Low-phase quantization error Mach–Zehnder interferometers for high-precision optical neural network training," APL Photonics **8**, 040801 (2023).
[15]Y. Yuan, W. V. Sorin, D. Liang, S. Cheung, Y. Peng, M. Jain, Z. Huang, M. Fiorentino, and R. G. Beausoleil, "Mechanisms of enhanced sub-bandgap absorption in high-speed all-silicon avalanche photodiodes," Photonics Res. **11**, 337–346 (2023).
[16]Y. Peng, Y. Yuan, W. V. Sorin, S. Cheung, Z. Huang, C. Hong, D. Liang, M. Fiorentino, and R. G. Beausoleil, "An 8 × 160 Gb s^{-1} all-silicon avalanche photodiode chip," Nat. Photonics **18**, 928–934 (2024).
[17]Y. Yuan, Y. Peng, S. Cheung, W. V. Sorin, S. Hooten, Z. Huang, D. Liang, J. Zhang, M. Fiorentino, and R. G. Beausoleil, "All-silicon non-volatile optical memory based on photon avalanche-induced trapping," Commun. Phys. **8**, 39 (2025).
[18]Y. Yuan, Y. Peng, S. Cheung, W. V. Sorin, Z. Huang, D. Liang, M. Fiorentino, and R. G. Beausoleil, "Silicon non-volatile optical memory and all-silicon photonics," in *2024 IEEE Photonics Society Summer Topicals Meeting Series (SUM)* (IEEE, 2024), pp. 1–3.
[19]D. Liang and J. E. Bowers, "Recent progress in lasers on silicon," Nat. Photonics **4**, 511–517 (2010).
[20]A. G. Cullis and L. T. Canham, "Visible light emission due to quantum size effects in highly porous crystalline silicon," Nature **353**, 335–338 (1991).
[21]L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzò, and F. Priolo, "Optical gain in silicon nanocrystals," Nature **408**, 440–444 (2000).
[22]O. Boyraz and B. Jalali, "Demonstration of a silicon Raman laser," Opt. Express **12**, 5269–5273 (2004).
[23]Y. Wan, J. Norman, Q. Li, M. J. Kennedy, D. Liang, C. Zhang, D. Huang, Z. Zhang, A. Y. Liu, A. Torres *et al.*, "1.3 μm submilliamp threshold quantum dot micro-lasers on Si," Optica **4**, 940–944 (2017).
[24]H. Yang, D. Zhao, S. Chuwongin, J.-H. Seo, W. Yang, Y. Shuai, J. Berggren, M. Hammar, Z. Ma, and W. Zhou, "Transfer-printed stacked nanomembrane lasers on silicon," Nat. Photonics **6**, 615–620 (2012).
[25]J. Justice, C. Bower, M. Meitl, M. B. Mooney, M. A. Gubbins, and B. Corbett, "Wafer-scale integration of group III–V lasers on silicon using transfer printing of epitaxial layers," Nat. Photonics **6**, 610–614 (2012).

[26]Y. Hu, D. Liang, K. Mukherjee, Y. Li, C. Zhang, G. Kurczveil, X. Huang, and R. G. Beausoleil, "III/V-on-Si MQW lasers by using a novel photonic integration method of regrowth on a bonding template," Light: Sci. Appl. **8**, 93 (2019).
[27]N. Lindenmann, G. Balthasar, D. Hillerkuss, R. Schmogrow, M. Jordan, J. Leuthold, W. Freude, and C. Koos, "Photonic wire bonding: A novel concept for chip-scale interconnects," Opt. Express **20**, 17667–17677 (2012).
[28]R. Jones, P. Doussiere, J. B. Driscoll, W. Lin, H. Yu, Y. Akulova, T. Komljenovic, and J. E. Bowers, "Heterogeneously integrated InP/silicon photonics: Fabricating fully functional transceivers," IEEE Nanotechnol. Mag. **13**, 17–26 (2019).
[29]B. Dong, J. Duan, H. Huang, J. C. Norman, K. Nishi, K. Takemasa, M. Sugawara, J. E. Bowers, and F. Grillot, "Dynamic performance and reflection sensitivity of quantum dot distributed feedback lasers with large optical mismatch," Photonics Res. **9**, 1550–1558 (2021).
[30]G. Park, O. B. Shchekin, D. L. Huffaker, and D. G. Deppe, "Low-threshold oxide-confined 1.3-μm quantum-dot laser," IEEE Photonics Technol. Lett. **12**, 230–232 (2000).
[31]A. Capua, L. Rozenfeld, V. Mikhelashvili, G. Eisenstein, M. Kuntz, M. Laemmlin, and D. Bimberg, "Direct correlation between a highly damped modulation response and ultra low relative intensity noise in an InAs/GaAs quantum dot laser," Opt. Express **15**, 5388–5393 (2007).
[32]G. Kurczveil, A. Descos, D. Liang, M. Fiorentino, and R. Beausoleil, "Hybrid silicon quantum dot comb laser with record wide comb width," in *Frontiers in Optics/Laser Science, OSA Technical Digest*, edited by B. Lee, C. Mazzali, K. Corwin, and R. Jason Jones (Optical Society of America, Washington, DC, 2020), p. FTu6E.6.
[33]S. Cheung, Y. Kawakita, K. Shang, and S. J. Ben Yoo, "Highly efficient chip-scale III-V/silicon hybrid optical amplifiers," Opt. Express **23**, 22431–22443 (2015).
[34]S. Cheung, Y. Kawakita, K. Shang, and S. J. B. Yoo, "Theory and design optimization of energy-efficient hydrophobic wafer-bonded III–V/Si hybrid semiconductor optical amplifiers," J. Lightwave Technol. **31**, 4057–4066 (2013).
[35]W.-Q. Wei, A. He, B. Yang, Z.-H. Wang, J.-Z. Huang, D. Han, M. Ming, X. Guo, Y. Su, J.-J. Zhang, and T. Wang, "Monolithic integration of embedded III-V lasers on SOI," Light: Sci. Appl. **12**, 84 (2023).
[36]Y. Wan, J. C. Norman, Y. Tong, M. J. Kennedy, W. He, J. Selvidge, C. Shang, M. Dumont, A. Malik, H. K. Tsang *et al.*, "1.3 μm quantum dot-distributed feedback lasers directly grown on (001) Si," Laser Photonics Rev. **14**, 2000037 (2020).
[37]J. Rahimi, J. Van Kerrebrouck, B. Haq, J. Bauwelinck, G. Roelkens, and G. Morthier, "Demonstration of a high-efficiency short-cavity III-V-on-Si C-band DFB laser diode," IEEE J. Sel. Top. Quantum Electron. **28**, 1–6 (2022).
[38]G. Kurczveil, X. Xiao, A. Descos, S. Srinivasan, D. Liang, and R. G. Beausoleil, "High-temperature error-free operation in a heterogeneous silicon quantum dot comb laser," in *Optical Fiber Communication Conference* (Optica Publishing Group, 2022), p. Tu2E-2.
[39]Z. Wang, K. Van Gasse, V. Moskalenko, S. Latkowski, E. Bente, B. Kuyken, and G. Roelkens, "A III-V-on-Si ultra-dense comb laser," Light: Sci. Appl. **6**, e16260 (2016).
[40]C. Zhang, D. Liang, G. Kurczveil, A. Descos, and R. G. Beausoleil, "Hybrid quantum-dot microring laser on silicon," Optica **6**, 1145–1151 (2019).
[41]B. Song, C. Stagarescu, S. Ristic, A. Behfar, and J. Klamkin, "3d integrated hybrid silicon laser," Opt. Express **24**, 10435–10444 (2016).
[42]R. Soref and B. Bennett, "Electrooptical effects in silicon," IEEE J. Quantum Electron. **23**, 123–129 (1987).
[43]Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "125 Gbit/s carrier-injection-based silicon micro-ring silicon modulators," Opt. Express **15**, 430–436 (2007).
[44]R. Wu, C.-H. Chen, J.-M. Fedeli, M. Fournier, K.-T. Cheng, and R. G. Beausoleil, "Compact models for carrier-injection silicon microring modulators," Opt. Express **23**, 15545–15554 (2015).
[45]W. Zhang, M. Ebert, K. Li, B. Chen, X. Yan, H. Du, M. Banakar, D. T. Tran, C. G. Littlejohns, A. Scofield *et al.*, "Harnessing plasma absorption in silicon MOS ring modulators," Nat. Photonics **17**, 273–279 (2023).
[46]W.-C. Hsu, N. Nujhat, B. Kupp, J. F. Conley, Jr, H. Rong, R. Kumar, and A. X. Wang, "Sub-volt high-speed silicon MOSCAP microring modulator driven by high-mobility conductive oxide," Nat. Commun. **15**, 826 (2024).
[47]Y. Zhang, H. Zhang, J. Zhang, J. Liu, L. Wang, D. Chen, N. Chi, X. Xiao, and S. Yu, "240 Gb/s optical transmission based on an ultrafast silicon microring modulator," Photonics Res. **10**, 1127–1133 (2022).
[48]E. Timurdogan, C. M. Sorace-Agaskar, J. Sun, E. Shah Hosseini, A. Biberman, and M. R. Watts, "An ultralow power athermal silicon modulator," Nat. Commun. **5**, 4008 (2014).
[49]J. Sun, R. Kumar, M. Sakib, J. B. Driscoll, H. Jayatilleka, and H. Rong, "A 128 Gb/s PAM4 silicon microring modulator with integrated thermo-optic resonance tuning," J. Lightwave Technol. **37**, 110–115 (2019).
[50]R. Nagarajan, L. Ding, R. Coccioli, M. Kato, R. Tan, P. Tumne, M. Patterson, and L. Liu, "2.5D heterogeneous integration for silicon photonics engines in optical transceivers," IEEE J. Sel. Top. Quantum Electron. **29**, 8200209 (2022).
[51]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," Nat. Photonics **11**, 441–446 (2017).
[52]W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walsmley, "Optimal design for universal multiport interferometers," Optica **3**, 1460–1465 (2016).
[53]B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 2704–2713.
[54]Y. Yuan, S. Cheung, T. Van Vaerenbergh, Y. Peng, Y. Hu, G. Kurczveil, Z. Huang, D. Liang, W. V. Sorin, M. Fiorentino *et al.*, "A 7-bit precision linearized Mach-Zehnder interferometer for high accuracy optical neural networks," in *2023 Opto-Electronics and Communications Conference (OECC)* (IEEE, 2023), pp. 1–3.
[55]B. Bartlett, M. Minkov, T. Hughes, and I. A. Williamson, "Neuroptica: Flexible simulation package for optical neural networks," GitHub, 2019, https://github.com/fancompute/neuroptica.
[56]Z. Huang, C. Li, D. Liang, K. Yu, C. Santori, M. Fiorentino, W. Sorin, S. Palermo, and R. G. Beausoleil, "25 Gbps low-voltage waveguide Si–Ge avalanche photodiode," Optica **3**, 793–798 (2016).
[57]Y. Yuan, S. Srinivasan, Y. Peng, D. Liang, Z. Huang, W. V. Sorin, S. Cheung, M. Fiorentino, and R. G. Beausoleil, "OSNR sensitivity analysis for Si-Ge avalanche photodiodes," IEEE Photonics Technol. Lett. **34**, 321–324 (2022).
[58]L. Yi, D. Liu, D. Li, P. Zhang, B. Tang, B. Li, W. Wang, Y. Yang, and Z. Li, "Waveguide-integrated Ge/Si avalanche photodiode with vertical multiplication region for 1310 nm detection," Photonics **10**, 750 (2023).
[59]S. Y. Siew, B. Li, F. Gao, H. Y. Zheng, W. Zhang, P. Guo, S. W. Xie, A. Song, B. Dong, L. W. Luo *et al.*, "Review of silicon photonics technology and platform development," J. Lightwave Technol. **39**, 4374–4389 (2021).
[60]M. Sakib, P. Liao, R. Kumar, D. Huang, G.-l. Su, C. Ma, and H. Rong, "A 112 Gb/s all-silicon micro-ring photodetector for datacom applications," in *Optical Fiber Communication Conference* (Optica Publishing Group, 2020), p. Th4A-2.
[61]Y. Peng, S. Hooten, Y. Yuan, Z. Huang, S. Cheung, W. V. Sorin, D. Liang, M. Fiorentino, and R. G. Beausoleil, "Polarization-insensitive high-speed all-silicon double-microring avalanche photodiodes," in *2024 IEEE Silicon Photonics Conference (SiPhotonics)* (IEEE, 2024), pp. 1–2.
[62]Y. Yuan, Y. Peng, Z. Huang, J. Hulme, S. Cheung, W. V. Sorin, D. Liang, M. Fiorentino, and R. G. Beausoleil, "An O-band all-silicon microring avalanche photodiode with> 38 GHz rf bandwidth," in *2023 IEEE Silicon Photonics Conference (SiPhotonics)* (IEEE, 2023), pp. 1–2.
[63]Y. Peng, Y. Yuan, W. V. Sorin, S. Cheung, Z. Huang, M. Fiorentino, and R. G. Beausoleil, "All-silicon microring avalanche photodiodes with a >65 A/W response," Opt. Lett. **48**, 1315–1318 (2023).
[64]Y. Peng, W. V. Sorin, S. Cheung, Y. Yuan, Z. Huang, M. Fiorentino, and R. G. Beausoleil, "Small-signal analysis of all-Si microring resonator photodiode," Electronics **11**, 183 (2022).
[65]J.-B. You, H. Kwon, J. Kim, H.-H. Park, and K. Yu, "Photon-assisted tunneling for sub-bandgap light detection in silicon pn-doped waveguides," Opt. Express **25**, 4284–4297 (2017).
[66]V. Van, *Optical Microring Resonators: Theory, Techniques, and Applications* (CRC Press, 2016).
[67]Y. Yuan, Y. Peng, Z. Huang, J. Hulme, S. Cheung, W. V. Sorin, D. Liang, M. Fiorentino, and R. G. Beausoleil, "A 4 × 100 Gbps DWDM receiver using all-Si

microring avalanche photodiodes," in *Optical Fiber Communication Conference* (Optica Publishing Group, 2023), p. W1A–5.
[68]C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "Integrated all-photonic non-volatile multi-level memory," Nat. Photonics **9**, 725–732 (2015).
[69]J. Zheng, A. Khanolkar, P. Xu, S. Colburn, S. Deshmukh, J. Myers, J. Frantz, E. Pop, J. Hendrickson, J. Doylend *et al.*, "GST-on-silicon hybrid nanophotonic integrated circuits: A non-volatile quasi-continuously reprogrammable platform," Opt. Mater. Express **8**, 1551–1561 (2018).
[70]J. F. Scott, "Applications of modern ferroelectrics," Science **315**, 954–959 (2007).
[71]J. Geler-Kremer, F. Eltes, P. Stark, A. Sharma, D. Caimi, B. J. Offrein, J. Fompeyrine, and S. Abel, "A non-volatile optical memory in silicon photonics," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)* (IEEE, 2021), pp. 1–3.
[72]S. Cheung, B. Tossoun, Y. Yuan, Y. Hu, G. Kurczveil, Y. Peng, D. Liang, and R. G. Beausoleil, "Non-volatile memristive III-V/Si photonics," in *2023 IEEE Silicon Photonics Conference (SiPhotonics)* (IEEE, 2023), pp. 1–2.
[73]B. Tossoun, D. Liang, S. Cheung, Z. Fang, X. Sheng, J. P. Strachan, and R. G. Beausoleil, "High-speed and energy-efficient non-volatile silicon photonic memory based on heterogeneously integrated memresonator," Nat. Commun. **15**, 551 (2024).
[74]Simulating the hysteresis effects of Si/SiO_2 interface traps, Silvaco Inc, 2010.
[75]C. Minkenberg, R. Krishnaswamy, A. Zilkie, and D. Nelson, "Co-packaged datacenter optics: Opportunities and challenges," IET Optoelectron. **15**, 77–91 (2021).
[76]G. Kurczveil, C. Zhang, A. Descos, D. Liang, M. Fiorentino, and R. Beausoleil, "On-chip hybrid silicon quantum dot comb laser with 14 error-free channels," in *2018 IEEE International Semiconductor Laser Conference (ISLC)* (IEEE, 2018), pp. 1–2.
[77]J. Zhu, W. Zhang, K. Li, B. Pant, M. Ebert, X. Yan, M. Banakar, D. T. Tran, C. G. Littlejohns, F. Gan *et al.*, "Universal silicon ring resonator for error-free transmission links," Photonics Res. **12**, 701–711 (2024).
[78]Y. Yuan, Y. Peng, Z. Huang, S. Cheung, W. V. Sorin, D. Liang, M. Fiorentino, and R. G. Beausoleil, "All-silicon microring transceivers enabling single-lane throughput exceeding 128 Gb/s," in *Frontiers in Optics* (Optica Publishing Group, 2023), p. FM5D-4.
[79]A. Kumar, Y. Yuan, R. Liu, Z. Huang, I. Kim, M. Fiorentino, R. Beausoleil, and S. Palermo, "A 80 Gb/s PAM4 CMOS transmitter integrated with two-segment silicon microring modulator," in *2024 IEEE European Solid-State Electronics Research Conference (ESSERC)* (IEEE, 2024), pp. 189–192.
[80]R. Liu, Y. Peng, A. Kumar, Y. Yuan, Y.-H. Fan, Y. Zhu, T. Liu, H. Kang, I. Kim, P. Yan *et al.*, "A 56Gb/s PAM4 optical receiver integrated with Si-Ge APD," in *2023 IEEE Photonics Conference (IPC)* (IEEE, 2023), pp. 1–2.
[81]N. C. Abrams, Q. Cheng, M. Glick, M. Jezzini, P. Morrissey, P. O'Brien, and K. Bergman, "Silicon photonic 2.5D multi-chip module transceiver for high-performance data centers," J. Lightwave Technol. **38**, 3346–3357 (2020).
[82]S. Shekhar, W. Bogaerts, L. Chrostowski, J. E. Bowers, M. Hochberg, R. Soref, and B. J. Shastri, "Roadmapping the next generation of silicon photonics," Nat. Commun. **15**, 751 (2024).
[83]D.-X. Xu, A. Delâge, P. Verly, S. Janz, S. Wang, M. Vachon, P. Ma, J. Lapointe, D. Melati, P. Cheben, and J. H. Schmid, "Empirical model for the temperature dependence of silicon refractive index from O to C band based on waveguide measurements," Opt. Express **27**, 27229–27241 (2019).
[84]K. Padmaraju and K. Bergman, "Resolving the thermal challenges for silicon microring resonator devices," Nanophotonics **3**, 269–281 (2014).
[85]K. Padmaraju, J. Chan, L. Chen, M. Lipson, and K. Bergman, "Thermal stabilization of a microring modulator using feedback control," Opt. Express **20**, 27999–28008 (2012).
[86]S. Agarwal, M. Ingels, M. Pantouvaki, M. Steyaert, P. Absil, and J. Van Campenhout, "Wavelength locking of a Si ring modulator using an integrated drop-port OMA monitoring circuit," IEEE J. Solid-State Circuits **51**, 2328–2344 (2016).
[87]S. Lin, X. Zheng, P. Amberg, S. S. Djordjevic, J.-H. Lee, I. Shubin, J. Yao, Y. Luo, J. Bovington, D. Y. Lee *et al.*, "Wavelength locked high-speed microring modulator using an integrated balanced homodyne CMOS control circuit," in *Optical Fiber Communication Conference* (Optica Publishing Group, 2016), p. Th3J-4.
[88]NVIDIA Announces Spectrum-X Photonics, Co-packaged Optics Networking Switches to Scale AI Factories to Millions of GPUs.
[89]P. L. McMahon, "The physics of optical computing," Nat. Rev. Phys. **5**, 717–734 (2023).
[90]A. N. Tait, A. X. Wu, T. F. De Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Microring weight banks," IEEE J. Sel. Top. Quantum Electron. **22**, 312–325 (2016).
[91]S. Ohno, R. Tang, K. Toprasertpong, S. Takagi, and M. Takenaka, "Si microring resonator crossbar array for on-chip inference and training of the optical neural network," ACS Photonics **9**, 2614–2622 (2022).
[92]J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," Nature **589**, 52–58 (2021).
[93]R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).
[94]H. H. Zhu, J. Zou, H. Zhang, Y. Z. Shi, S. B. Luo, N. Wang, H. Cai, L. X. Wan, B. Wang, X. D. Jiang *et al.*, "Space-efficient optical computing with an integrated chip diffractive neural network," Nat. Commun. **13**, 1044 (2022).
[95]C. Ramey, "Silicon photonics for artificial intelligence acceleration: HotChips 32," in *2020 IEEE Hot Chips 32 Symposium (HCS)* (IEEE, 2020), pp. 1–26.
[96]C. Demirkiran, F. Eris, G. Wang, J. Elmhurst, N. Moore, N. C. Harris, A. Basumallik, V. J. Reddi, A. Joshi, and D. Bunandar, "An electro-photonic system for accelerating deep neural networks," ACM J. Emerg. Technol. Comput. Syst. **19**, 1–31 (2023).
[97]X. Xiao, M. B. On, T. Van Vaerenbergh, D. Liang, R. G. Beausoleil, and S. J. B. Yoo, "Large-scale and energy-efficient tensorized optical neural networks on III–V-on-silicon MOSCAP platform," APL Photonics **6**, 126107 (2021).
[98]X. Xiao, S. Cheung, S. Hooten, Y. Peng, B. Tossoun, T. Van Vaerenbergh, G. Kurczveil, and R. G. Beausoleil, "Wavelength-parallel photonic tensor core based on multi-FSR microring resonator crossbar array," in *Optical Fiber Communication Conference* (Optica Publishing Group, San Diego, CA, 2023), p. W3G.4.
[99]Y. Zhao, X. Xiao, X. Yu, Z. Liu, Z. Chen, G. Kurczveil, R. G. Beausoleil, and Z. Zhang, "Real-time FJ/MAC PDE solvers via tensorized, back-propagation-free optical PINN training," in NeuRIPS Workshop on Machine Learning with Non-Conventional Computing Paradigm, 2023.
[100]X. Xiao, Y. Zhao, Y. Yuan, G. Kurczveil, M. Fiorentino, R. Beausoleil, and Z. Zhang, "TOMFuN: A tensorized optical multimodal fusion network," APL Mach. Learn. **3**, 016121 (2025).
[101]G. Kurczveil, M. A. Seyedi, D. Liang, M. Fiorentino, and R. G. Beausoleil, "Error-free operation in a hybrid-silicon quantum dot comb laser," IEEE Photonics Technol. Lett. **30**, 71–74 (2018).