nature reviews electrical engineering

https://doi.org/10.1038/s44287-025-00223-0

Review article

Check for updates

Integrating silicon photonics with complementary metal–oxide–semiconductor technologies

Yating Wan[1] ✉, William He[1], James Jaussi[2], Ling Liao[3], David Z. Pan[4], John E. Bowers[5] & Haisheng Rong[2] ✉

Abstract

Artificial intelligence, machine learning and high-performance computing workloads are pushing electrical input/output to its limits in signal reach, energy efficiency and bandwidth density, turning optics from option to necessity. Complementary metal–oxide–semiconductor-integrated silicon photonics offers a practical path forward by combining high-volume manufacturing with mature photonic building blocks. This Review presents progress across devices (on-chip lasers and semiconductor optical amplifiers, compact modulators, high-speed photodetectors, low-loss routing and efficient chip–fibre couplers), multimaterial integration (hybrid assembly, heterogeneous wafer bonding, microtransfer printing and monolithic epitaxy) and electronics co-design (digital signal processing, serializer/deserializer, stacked-driver topologies, bias control and thermal tuning) to show how total link energy is being driven towards the sub-picojoule per bit regime. We connect these advances to system architectures that are evolving from pluggables to linear-drive pluggables and co-packaged optics, and we discuss the trade-offs among bandwidth density, thermal design, yield and cost. We identify near-term bottlenecks, notably thermal pathways and manufacturing yield, and highlight technologies most likely to unlock the next jump in performance, including on-chip comb sources for dense wavelength-division multiplexing and wafer-scale 3D electronic and photonic stacks. The same platform is poised to impact optical compute input/output, sensing and quantum photonics, linking device-level innovation to system-level gains across computing and communications.

Sections

[1]Integrated Photonics Lab, King Abdullah University of Science and Technology, Thuwal, Makkah Province, Saudi Arabia. [2]Intel Corporation, Santa Clara, CA, USA. [3]Nvidia Corporation, Santa Clara, CA, USA. [4]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA. [5]Institute for Energy Efficiency, University of California, Santa Barbara, CA, USA. ✉e-mail: yating.wan@kaust.edu.sa; haisheng.rong@intel.com

Key points

- Optical interconnects offer higher bandwidth density and lower energy per bit than copper, and complementary metal–oxide–semiconductor-compatible silicon photonics provides a scalable, cost-effective manufacturing path.
- Core building blocks are now integration-ready for data-centre links: on-chip lasers and semiconductor optical amplifiers, compact modulators, high-speed photodetectors, low-loss waveguide routing and efficient chip-to-fibre couplers.
- Multimaterial and 3D integration using hybrid, heterogeneous wafer bonding, microtransfer printing and monolithic epitaxy expands functionality, density and wavelength coverage, while enabling wafer-scale, multilayer electronic–photonic stacks.
- Tight co-design with advanced electronics, including digital signal processing, serializer/deserializer, stacked-driver topologies, bias control and thermal tuning, aligns voltage and noise budgets, pushing total link energy towards sub-picojoule per bit.
- System architectures are evolving from pluggables to linear-drive pluggables and co-packaged optics, trading electrical complexity, thermal design against bandwidth density, yield and cost.
- Next steps include on-chip comb-enabled dense wavelength-division multiplexing, wafer-level burn-in/testing and wafer-scale 3D electronic–photonic integrated circuit stacks to reach petabit-scale interconnects, while complementary metal–oxide–semiconductor-integrated photonics expands into artificial intelligence/computing, sensing and quantum technologies.

Introduction

The computing industry has achieved sustained performance gains for decades, mainly through continued progress in device integration and scaling, as captured by the Moore's law. Modern data extend this trend by packing more silicon compute capacities onto each package and by deploying ever richer system networks. A prevailing strategy in central processing unit (CPU) and graphical processing unit (GPU) design is therefore to increase the number of compute cores with advances in complementary metal–oxide–semiconductor (CMOS) technology[1]. This scaling of compute cores directly amplifies memory bandwidth requirements, which currently doubles approximately every 2 years[2]. Designers address the issue by boosting the data rate per memory-interface lane and by adding more memory channels. However, both measures increase package size and pin count, which complicates signal routing and power delivery, adds thermal and mechanical stress and raises manufacturing cost. These challenges underscore a fundamental limitation: the shoreline bandwidth density, defined as the data bandwidth that can be delivered along the limited input/output (I/O) perimeter ('shoreline') of a silicon die. This parameter ultimately sets a hard ceiling on achievable bandwidth.

Bandwidth pressure has escalated even further with the rapid growth of artificial intelligence (AI), machine learning (ML) and other high-performance computing (HPC) workloads[3]. Model sizes alone have expanded by five orders of magnitude in the past 5 years[4], sharply increasing the volume of data that must shuffle among compute tiles. Transporting those bits consumes energy in the physical layer, typically expressed in picojoules per bit. For electrical I/O, energy scales with both channel length and data rate. Recent demonstrations achieve 200 Gb s^{-1} over a 1-m-long high-performance cable[5,6], but beyond that distance, the required power rises steeply. To meet bandwidth needs at multimetre to rack scale, the industry is therefore moving towards optical interconnects.

Historically, cost, physical size and power constraints have confined optical modules to the upper tiers of data-centre networks, starting at top-of-rack switches. Advances in photonic devices, packaging and manufacturing are breaking those barriers, enabling short-reach optical links and motivating the integration of photonics directly inside compute packages. Such co-packaged optics (CPO) promises much higher bandwidth density, lower energy per bit and flatter network topologies suited to both scale-up and scale-out applications[7–9].

A major factor of this evolution has been the progress in silicon photonics, which offers seamless integration with electronics and exceptional manufacturing scalability[10,11]. Notable recent progress includes on-chip lasers with precise wavelength control, microring modulators (MRMs) for compact and efficient dense wavelength-division multiplexing (DWDM) and high-speed photodetectors (PDs) co-designed with CMOS electronics. Together, these devices are pushing bandwidth density and energy efficiency towards the requirements of next-generation computing architectures[12–14]. The pace of progress in the past few years, coupled with growing commercial deployments, demands a fresh assessment of the field.

In this Review, we survey the core photonic and electronic building blocks that support silicon photonics-based interconnects, compare state-of-the-art integration strategies and their remaining challenges, highlight emerging applications from on-package optical memory interfaces to rack-scale disaggregated systems and outline future directions that could deliver petabit-per-second bandwidths at sub-femtojoule energy efficiencies.

Key components and evolutionary trends

Silicon photonics has progressed from isolated laboratory demonstrations to high-volume platforms that integrate multiple optical functions on a single wafer. In this section, we examine the essential devices and the breakthroughs in materials and design that support the performance targets of modern data centres.

Fundamental building blocks

Integration-ready light sources and optical amplifiers, compact electro-optic modulators, high-speed PDs and low-loss waveguides with chip-to-fibre couplers form the foundation of silicon-photonics links (Fig. 1). Each of these devices is designed for CMOS-compatible fabrication. In the following sections, we outline their operating principles, highlight state-of-the-art performance and discuss the challenges that remain for scaling, energy efficiency and system-level innovation.

Lasers. Lasers are the engines of any photonic platform; however, the indirect bandgap of silicon has long prevented the realization of fully integrated sources. Early work therefore relied on external lasers, a choice that introduced optical interfaces, alignment steps and thermal overhead. Steady progress now makes on-chip solutions both feasible and preferable. In the 1990s, the development of optically pumped erbium (Er)-doped silica waveguide lasers[15], followed by the first continuous-wave Raman silicon laser in 2005 (ref. 16),

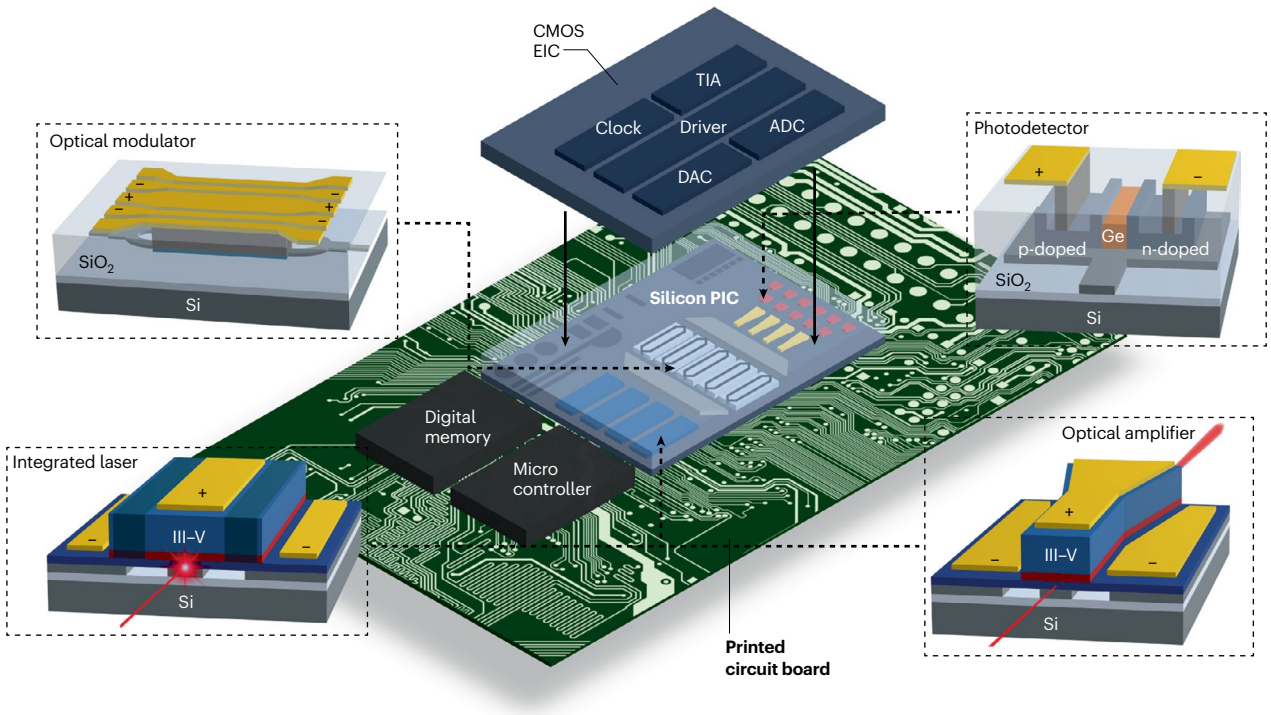

Fig. 1 | Silicon photonics integrated with compatible metal–oxide–semiconductor electronics. The silicon photonic integrated circuit (PIC), mounted on a printed circuit board, incorporates optical components, such as integrated lasers, modulators, optical amplifiers and photodetectors. Vertically stacked above the PIC is a compatible metal–oxide–semiconductor (CMOS) electronic integrated circuit (EIC) comprising drivers for the PIC, supported by transimpedance amplifiers (TIAs), digital-to-analog converters (DACs), analog-to-digital converters (ADCs) and clocking modules.

demonstrated the feasibility of on-chip silicon light sources. However, true commercial impact only arrived with the advent of heterogeneous bonding of III–V gain chips to silicon[17]. These heterogeneous lasers, used in millions of optical transceiver products worldwide[18], routinely beat native-substrate devices in linewidth[19], wavelength tunability[20], isolation-free stability[21,22] and volume uniformity[12].

The next leap came from quantum dot (QD) gain bonded to silicon[22]. QD lasers operate at higher temperatures, show longer lifetimes[23] and achieve lower threshold currents, while also offering ultrafast gain recovery. These attributes make them ideal for on-chip mode-locked combs that drive DWDM links and clock recovery[24]. Their low linewidth enhancement factor further narrows emission spectra and tolerates optical feedback, eliminating bulky isolators and trimming both cost and power[25,26]. Collectively, these advances close the performance gap with off-chip sources while unlocking fully integrated, energy-efficient photonic systems.

Optical amplifiers. On-chip semiconductor optical amplifiers (SOAs) are gaining importance as photonic integrated circuits (PICs) grow larger and introduce higher optical losses. Heterogeneously integrated SOAs on silicon, first demonstrated soon after heterogeneous lasers[27], can now be incorporated into the same process flow. They can be strategically positioned where loss is substantial and therefore reduce laser power budgets and improve system energy efficiency. Recent work targets on lower noise and wider bandwidth. Multistage cascaded SOAs built from low-dimensional quantum structures routinely achieve noise figures around 5 dB (ref. 28) and designs that trade modest gain for noise suppression have reduced this further to 3.7 dB (ref. 29), comparable to the performance of erbium-doped fibre amplifiers. Although these amplifiers still offer higher overall gain, SOAs with gain bandwidths exceeding 120 nm (ref. 30) are getting growing attention in DWDM links.

An alternative gain medium is Er-doped materials, which are fully compatible with CMOS deposition. Erbium-doped aluminium oxide ($Er:Al_2O_3$) offers an ultraviolet–mid-infrared window, low loss and high rare-earth solubility. Its wafer-scale films have produced ultra-narrow linewidth lasers with a linewidth of 1.7 kHz (ref. 31). Erbium-implanted silicon nitride (Si_3N_4) waveguide amplifiers already yield >30 dB small-signal gain and 145 mW output[32]. Additionally, atomic-layer deposition or vapour deposition promises seamless integration without lattice-mismatch concerns. Together, these advances position on-chip amplifiers as essential building blocks for next-generation high-bandwidth PICs.

Optical modulators. The first silicon modulator, reported in 2004, included a metal-oxide–semiconductor (MOS)-capacitor (MOSCAP) Mach–Zehnder interferometer (MZI) and reached 1 Gb s^{-1} (ref. 33). Most commercial devices still rely on MZI phase shifters driven by MOSCAP or carrier-depletion-effect junctions[34]. More recent designs, such as racetrack modulators and MRMs, now deliver the same functionality in footprints of only tens of micrometres, while also improving speed and energy efficiency[35]. MRMs trade size for a narrow optical bandwidth and strong thermo-optic drift. Conventional thermal tuning is one-way and power-hungry, meaning that a mismatch of a few kelvins

between an off-chip laser and an on-chip ring can force a full retune of free spectral range.

Integrating the laser directly with the modulator eliminates this penalty, as both sources share the same temperature and their III–V/silicon (Si) thermo-optic coefficients track closely. This keeps the ring on resonance with minimal bias power. Beyond devices made purely from silicon, researchers have integrated materials such as lithium niobate ($LiNbO_3$)[36], barium titanate ($BaTiO_3$)[37], transparent conductive oxides[38] and III–V compound semiconductors[39] onto silicon, enabling electro-optic bandwidths exceeding 100 GHz. However, realizing wafer-scale, cost-effective flows that integrate these materials with existing silicon lasers and germanium (Ge) detectors remains a key challenge[40].

Photodetectors. Most silicon photonic transceivers use Ge-on-silicon (GeSi) waveguide PDs[41], which provide broad bandwidth and are well established in commercial platforms. More recently, all-silicon waveguide-based PDs[42] have emerged as a low-cost alternative. By exploiting mechanisms such as two-photon absorption, defect-mediated absorption, photon-assisted tunnelling and the avalanche effect, these devices overcome the telecom transparency of silicon and achieve responsivity sufficient for high-speed optical signal detection. This enables high-speed signal detection[43] and in-line power monitoring[44] at lower costs than Ge or III–V-based PDs.

Meanwhile, QD PDs integrated into silicon photonics have achieved dark currents below 0.01 nA (ref. 45) while utilizing the same epitaxial layers and processing steps as QD lasers. This compatibility enables the realization of fully integrated silicon transceivers that combine both lasers and detectors on a single chip without the need for additional lithography masks[46].

Passive waveguides. Silicon waveguides now attain losses as low as 0.2 dB cm^{-1} (ref. 47), owing to the high silicon/silicon oxide index contrast ($\eta_n \approx 3.48/1.44$) and the side-wall-smoothing achieved through CMOS thermal oxidation and oxide stripping. They represent a clear advantage over the ≥1 dB cm^{-1} loss typical of III–V waveguides[48] and foundries routinely integrate silicon waveguides into crossings, bends, tapers, gratings, arrayed waveguide gratings and multimode interference (MMI) devices[49].

For longer paths and high-power routing, Si_3N_4 offers even lower propagation losses, reaching 0.034 dB m^{-1} (ref. 50). In current Si_3N_4 processes, high-temperature hydrogen-out-diffusion anneals (≥1,100 °C) still dominate the thermal budget, but low-temperature deuterated-oxide claddings are emerging as a drop-in alternative[51]. Si_3N_4 further offers a wide transparency window, a small thermo-optic coefficient and negligible nonlinear absorption losses, enabling high-Q filters and low phase-noise cavities[52]. It also withstands >300 mW input while maintaining <0.01 dB cm^{-1} propagation loss, <0.001 dB 90°-bend loss and 0.015 dB interwaveguide transition loss[53].

Chip-to-fibre couplers. Fibre-to-chip coupling largely fixes the loss budget, power draw and packaging cost of a photonic link. Two mature approaches dominate today: grating couplers (GCs) and edge couplers. GCs give surface access, enabling wafer-scale testing, space-division multiplexing and dense multiple-row fibre arrays. Conventional uniform GCs are limited to ≈−2.2 dB efficiencies and 30–40 nm operational bandwidths[54]. Through design innovations such as apodization structures[55], polysilicon overlays[56], sub-wavelength[57] and metal-reflected gratings[58] and dual-level stacks[59], efficiencies <1 dB loss, 1-dB bandwidths >100 nm (ref. 60) and dual-polarization operation have been demonstrated. However, the efficiency–bandwidth product still falls below 60 nm, and many record devices require process steps outside standard foundry flows[61].

Edge couplers achieve <1 dB loss through advanced mode converters[62], index-matching claddings[63], sub-wavelength gratings[64] and trident tapers[65]. They are broadband (>100 nm) and polarization-insensitive, but scale only along one 'beachfront' that limits total fibre rows. To address these trade-offs, hybrid solutions are emerging that combine the advantages of both approaches. On an Si_3N_4 platform, vertically integrated silicon microlenses and 90° beam deflectors have enabled multirow, broadband coupling with 0.08 dB loss over 100 nm bandwidth. This architecture supports four rows of 80 fibres each, and embedding polarization beam splitter-rotators sustain >23 dB extinction ratio from 1,260 nm to 1,360 nm (ref. 53).

Multimaterial integration

Silicon alone cannot supply the gain, electro-optic response or ultra-low-loss routing demanded by next-generation photonic systems. Adding complementary materials, such as III–V components, Si_3N_4, $LiNbO_3$, Ge, 2D layers and rare-earth-doped films, extends functionality while retaining CMOS volume manufacturing. Three integration families have emerged, namely, hybrid (chip-to-chip assembly), heterogeneous (die-to-wafer bonding) and monolithic (single-wafer co-fabrication). Each tackles mismatched lattices, thermal budgets and alignment tolerances in a different way. This section compares their trade-offs in efficiency, scalability and cost.

Hybrid integration. Hybrid integration is currently the most widely used approach in industry, combining independently fabricated chips on separate substrates into a shared platform that delivers the required optical, mechanical and electrical functions. The conventional approach uses fibre-coupled off-chip lasers linked to the PIC through fibre alignment units and polarization-maintaining fibres with optical isolators. Although straightforward, this process faces challenges of precise alignment requirements, coupling losses owing to mode mismatch and bulky packaging that limits scalability and complicates thermal management.

To address these issues, direct chip-to-chip coupling approaches (Fig. 2a) have been developed, including angled facets combined with GCs for vertical optical coupling[66], vertical alignment pillars or intermediate coupling elements for edge coupling[67] and photonic wire bonds to address spatial displacements[68]. Among these methods, flip–chip bonding has gained particular attraction, as it enables device pre-screening, precise placement with micrometre-scale accuracy and dense interconnection through solder or micro-bump arrays, delivering high optical gain and output power[69]. Despite these advances, most hybrid laser attachments still rely on butt-coupling, in which large divergence angle of typical III–V lasers imposes stringent requirements on chip alignment in all dimensions. As a result, intermediate elements such as spot-size converters are typically required to match the laser mode to the silicon waveguide[70].

Heterogeneous integration. Heterogeneous integration, in contrast to hybrid integration, bonds dissimilar materials directly onto a common substrate, enabling wafer-scale lithography and tight optical alignment. The technique widens the wavelength coverage of silicon photonics from gallium arsenide (GaAs)-based gain media with Si_3N_4 waveguides at ~900 nm (ref. 20), extending to quantum cascade lasers at 4.8 µm, while keeping CMOS-level cost and density[71].

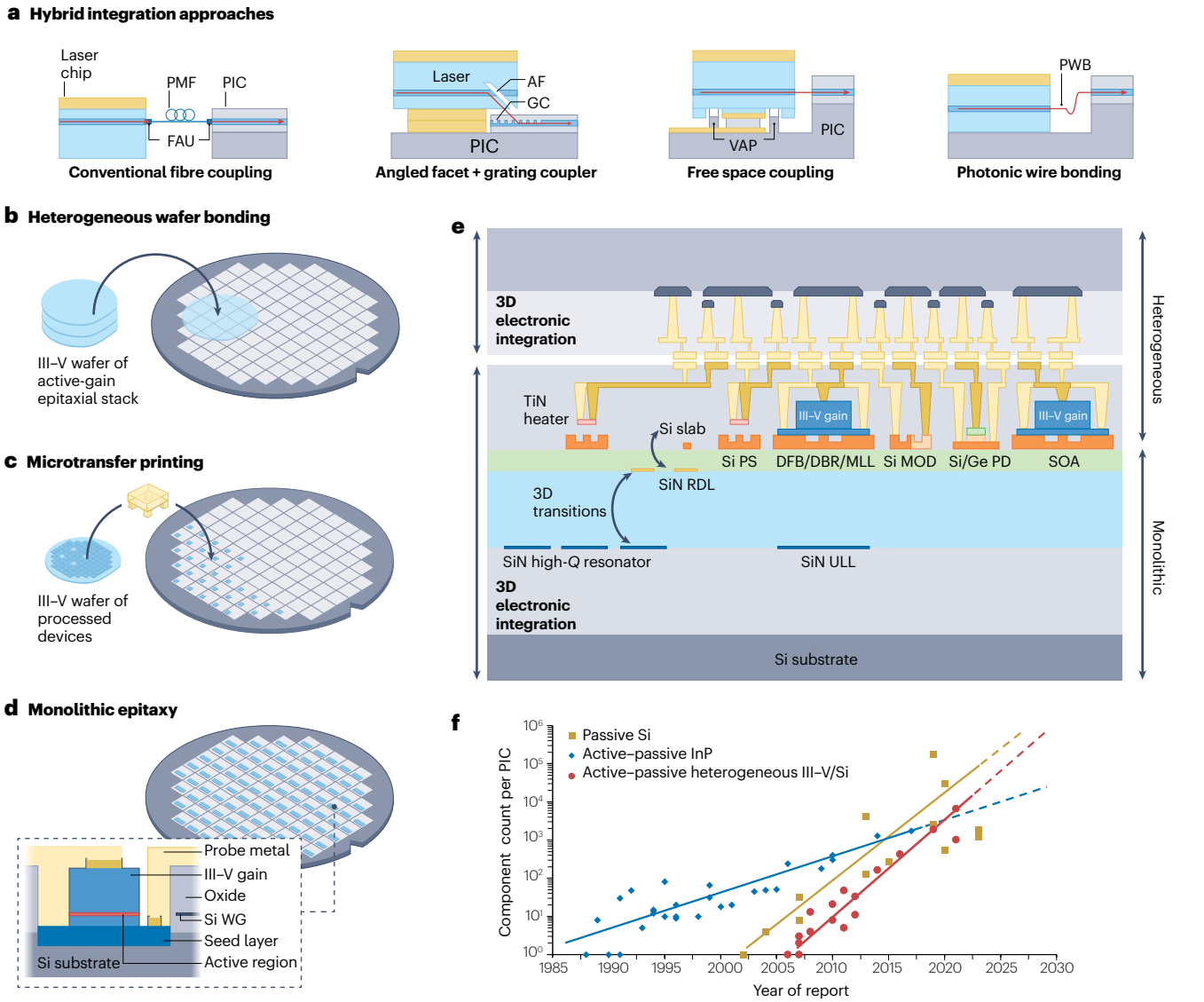

Fig. 2 | Comparative analysis of multimaterial integration techniques in silicon photonics. a, Hybrid integration approaches. **b**, Heterogeneous wafer bonding. **c**, Microtransfer printing. **d**, Monolithic epitaxy. **e**, Cross-sectional depiction of a III–V/silicon (Si)/silicon nitride (Si_3N_4) 3D photonic integration platform. **f**, The increasing number of devices on a single photonic integrated circuit. AF, angled facet; DBR, distributed Bragg reflector; DFB, distributed feedback; FAU, fibre alignment unit; GC, grating coupler; MLL, mode-locked laser; MOD, modulator; PD, photodetector; PIC, photonic integrated circuit; PMF, polarization-maintaining fibre; PS, phase shifter; PWB, photonic wire bond; RDL, redistribution layer; Si, silicon; SOA, semiconductor optical amplifier; ULL, ultra-low loss; VAP, vertical alignment pillar; WG, waveguide. Part **e** is adapted from ref. 21, CC BY 4.0. Part **f** is adapted from ref. 3, CC BY 4.0.

Wafer bonding. One of the most popular heterogeneous integration approaches involves bonding unpatterned thin films of functional materials to silicon or Si_3N_4 wafers early in the process, after which device features are defined through post-bond lithography (Fig. 2b). Compared with hybrid integration approaches, this method shifts device manufacturing to the wafer level, enhancing integration density and cost-effectiveness by eliminating the need for active alignment[72]. It also enables large laser arrays with broad wavelength coverage, precisely aligned to the silicon photonic waveguides with the accuracy of CMOS lithography. Because the laser facets are formed in silicon rather than III–V materials and are protected by deposited dielectric films, they are not exposed to potential damage or contamination. This structural robustness translates into almost lossless optical coupling to silicon photonic circuits and best-in-class reliability[72].

Heterogeneous integration uses various wafer-bonding methods, including polymer, metal and direct bonding. Polymer bonding is noted for exceptional tolerance to surface roughness and good thermal stability. Metal bonding, known for its low resistance and high thermal conductivity, is preferable for thermal management but may introduce optical loss and metal contamination. Direct bonding, forming strong bonds through intermolecular forces without adhesives, has become the preferred technique. It enables the first electrically

pumped, silicon-based heterogeneous integrated laser in 2006 (ref. 17) and has since been adopted in the commercialization of 100 Gb s^{-1} optical transceivers[73]. Building on this foundation, sequential wafer bonding extends the approach to multilayer structures. By mitigating index mismatches across different materials, it further broadens the operational wavelength range and improves device performance. An example is the integration of III–V/Si/Si_3N_4 configurations[19], which combines ultra-low-loss passive routing with broadband, thermally stable gain.

Microtransfer printing. Microtransfer printing (Fig. 2c) represents an alternative heterogeneous integration approach, which adapts the die-level flip–chip integration technique to the device level. In this process, densely packed III–V devices are picked from a source wafer with a poly-dimethylsiloxane stamp and then printed onto a pre-processed silicon photonics wafer through rate-dependent adhesion[74], with entire arrays placed in only 45 s per cycle[75]. Alignment of <50 nm has been achieved[76], but production tools typically guarantee $\lesssim$1 µm (ref. 74). Compared with wafer bonding, transfer printing sacrifices a degree of alignment accuracy but offers exceptional material flexibility and pre-testability, enabling the integration of a wide range of functionalities not inherent to the base material.

Monolithic integration. Monolithic integration embeds III–V gains directly on silicon via epitaxial growth, providing excellent thermal paths at wafer-scale cost. However, differences in lattice constants, thermal expansion coefficients and material polarities generate various defects, including threading dislocations, stacking faults, misfit dislocations and antiphase domains (APDs), all of which supress optical gain and shorten device lifetime. Threading dislocations and stacking faults can be mitigated by thermal cycle annealing, dislocation filter layers and compositionally graded buffers. Asymmetric step-graded filters, for instance, have reduced threading dislocation density to 1×10^6 cm^{-2} in ref. 77. Misfit dislocations can be addressed using strained quantum well (QW) trapping layers, which achieve up to 90% removal[78]. APDs, in turn, can be eliminated through nano-patterned V-grooved surfaces[79] or by using commercially available gallium phosphide on silicon wafers, both of which provide CMOS-compatible substrates[80].

Building on these advances, monolithic III–V/Si growth has progressed steadily. Early achievements included the first continuous-wave lasers on miscut silicon, the development of APD-free buffers and CMOS-compatible substrates and later the introduction of advanced doping and new active layers, each extending device lifetime and output power[26]. A particularly important breakthrough has been the replacement of QW gains with QDs, which improves defect tolerance and has enabled silicon-grown QD lasers that now rival QW devices[81]. Demonstrations span sub-milliamp threshold microrings[82], mode-locked lasers with wide mode-locking regimes[83], as well as distributed feedback (DFB) lasers[84] and tunable lasers[85] with high side-mode suppression ratios. More recent progress includes active–passive co-integration on 300 mm patterned silicon wafers[86,87] and efficient light coupling into silicon waveguides[88] (Fig. 2d). Although monolithic integration remains less mature than hybrid and heterogeneous approaches, the gap is closing quickly.

Looking beyond III–V materials, other materials are now being explored to expand device functionality. Phase-change materials (PCMs) and rare-earth-doped films can be sputtered or thermal evaporated directly onto silicon, adding non-volatile tuning or on-chip gain without bonding or epitaxial growth. The phase-change compound $Ge_2Sb_2Te_5$ and related PCMs combine large refractive index contrast with tunable optical absorption, enabling low-power optical logic, microwave photonics links and neuromorphic or in-memory photonic processors[89–92]. Key challenges, including cycling endurance, switching energy and uniform wafer-scale fabrication, remain active areas of research in material engineering and process optimization.

Trends and prospects

Multimaterial silicon photonics now defines every roadmap for bandwidth scaling. Ultra-low-loss Si_3N_4 resonators enable hertz-linewidth hybrid integrated lasers with CMOS-ready process[51]. Active components such as high-bandwidth modulators[93] and high-speed PDs[94] further increase link density. The shipment of tens of millions of bonded III–V/Si lasers demonstrates manufacturability, yet challenges in yield, thermal design and automated packaging remain.

On-chip laser integration. Silicon photonics has progressed from the first 4×12.5 Gb s^{-1} coarse wavelength-division multiplexing chips in 2011 (ref. 95) to >8 million PICs carrying 32 million wafer-bonded on-chip lasers in 100–400 Gb s^{-1} pluggable optical transceivers in 2022 (ref. 96). This progression shows that integrated light sources can scale with Moore-like economics. Wafer-scale bonding improves cost, reliability, channel density, bandwidth and compactness. It also enables efficient wafer-level burn-in and testing, in which devices are stressed under controlled conditions to screen for early failures and stabilize performance. The same process can co-bond SOAs, supporting $1 \times N$ lasers sharing or boosting power margins. These benefits come with a more complex fabrication flow. High-yield die bonding, III–V substrate removal and robust thermal pathway engineering are essential to keep lasers and SOAs within their lower operating temperature window relative to nearby electronics. Integrating light sources before packaging reduces the dominant packaging cost component but requires that reliability and calibration be secured earlier in the manufacturing line. Wafer-scale burn-in and testing therefore become essential, not only to verify device lifetime and stability but also to improve laser performance by annealing active-region defects during the process.

Towards 3D multilayer silicon photonics. The move from planar PICs to vertically stacked, 3D photonic chips[97] overcomes reticle-size limits. This approach accommodates far more devices per package, a capability that is crucial for data centres and AI workloads. Stacked tiers also unlock new architectures. For example, integrated comb sources increase total throughput while reducing laser power. Wavelength-division multiplexing places modulators, detectors and waveguides onto one vertically unified platform. A flagship demonstration, shown in Fig. 2e, combines monolithic and heterogeneous processes to integrate III–V gain materials with ultra-low-loss Si_3N_4 waveguides (<0.5 dB m^{-1}). The incorporation of high-Q cavities reduces laser noise to levels comparable with fibre lasers, eliminating isolators and supporting low-noise heterodyne microwave generation on a single silicon substrate[21].

Scaling towards optical system-on-chip architectures, from initial intercore links to fully integrated on-chip networks, demands photonic stacks that can operate reliably in hot processor environments while remaining economical at wafer-scale volumes. To meet this demand, both industry and academia are exploiting CMOS manufacturing to integrate more functions per die, with heterogeneous III–V/Si platforms that include on-chip lasers advancing most rapidly (Fig. 2f). Scalability efforts are driven by system-level metrics such as bandwidth

density, energy efficiency and latency, with full-system energy targets of roughly 100 fJ bit^{-1}, of which only 10–20 fJ bit^{-1} can be allocated to the optical source[98]. Meeting these targets is critical for future HPC and data fabrics interconnects to clearly outperform advanced electrical links. Looking ahead, silicon photonics must resolve trade-offs in multimaterial integration: heterogeneous approaches offer flexibility and proven manufacturability, whereas monolithic systems promise higher performance but greater fabrication complexity. Balancing these approaches is necessary to optimize cost, yield and system efficiency for large-scale deployments.

Building blocks of electronics

The speed, energy efficiency and reliability of a photonic link are ultimately determined by the electronics that bias, drive and recover the optical signals. Four functions are paramount: digital signal processing (DSP), serializer/deserializer (SerDes) macros, high-speed drivers and bias control.

Digital signal processing and SerDes

DSP and SerDes technologies manage data flow and signal integrity across multiwavelength links. Long-haul and metro communication systems with ≤40 km rely on coherent transceivers. These transceivers combine advanced DSP, featuring chromatic dispersion, polarization and nonlinear compensation, with SerDes lanes operating at >200 Gb s^{-1}. Together, they provide clock and data recovery, equalization and forward-error correction[99]. Short-range links with <100 m instead use intensity-modulated direct-detection (IMDD) to reduce power by trimming DSP and forward-error correction and still employing adaptive equalization and lightweight coding. Across all regimes, DSP extends reach and capacity, whereas SerDes bridges electrical and optical domains.

Voltage requirements for CMOS compatibility

Realizing the full potential of silicon photonic components requires co-engineering with driving electronic circuits to achieve high-precision modulation, detection and amplification of optical signals. Voltage compatibility is central, with MRMs, PDs, lasers and SOAs operating from CMOS-friendly rails provided by either dedicated power delivery ICs or on-die thick-gate oxide CMOS transistors. High-speed drivers adopt a two-stage topology: a linear pre-driver lifts sub-volt CMOS logic to an intermediate level, and the main driver produces the modulation voltages required for photonic devices. In MRMs, for instance, driver circuits must deliver modulation voltages ranging from 1.8 V to 3.6 V peak-to-peak (Vpp) while minimizing signal distortion and power dissipation. This precision is particularly important in DWDM systems, in which accurate modulation ensures optimal bandwidth utilization and reliable system performance. MRMs are widely used in IMDD transceivers for DWDM systems owing to their ability to precisely align with individual laser wavelengths[100,101]. Frequency tuning of an MRM is typically achieved through thermal adjustment, in which a thick-gate transistor drives a resistive heating element to match the desired laser wavelength[102]. Beyond tuning, these transistors also provide bias voltages for other photonic devices, including PDs, lasers and SOAs, ensuring synchronized operation of the entire photonic system.

High-speed circuit design for modulation

The precise voltage requirements of photonic devices heavily influence the design of electronic circuits. Thick-gate oxide CMOS transistors remain well suited for direct-current and low-frequency biasing where their robustness and tolerance to higher voltages are advantageous. However, high-speed modulation requires advanced CMOS process nodes with thinner gate oxides, shorter channels and faster transistors, which can handle rapid voltage swings and maintain signal integrity. Drivers for multi-tens-of-gigahertz operation must handle transitions across different voltage domains while maintaining signal integrity. A common solution is the stacked transistor topology, which enables output driver to deliver modulation voltages up to twice the nominal CMOS supply[102,103] (Fig. 3). This design introduces dual supply-voltage paths and additional level-shifting circuitry to bridge the normal supply and doubled-voltage domains. Even with this added complexity, it remains CMOS-compatible and meets the rigorous performance demands of high-speed photonic systems.

Link architecture for bandwidth and efficiency

Beyond DSP, SerDes and driver circuits, the overall efficiency and scalability of a silicon photonic system also depend on the architecture of the optical links that connect photonic and electronic components. To meet growing bandwidth demands and reduce power consumption, two primary link architecture strategies are being pursued. One approach drives each channel to higher line rates, thereby reducing the total wavelength count[104]. The other maintains modest per-channel rate but spreads traffic across a larger number of wavelengths[105]. These choices reshape the power budget: at higher data rate, CMOS circuit power dominates, whereas at lower data rates, CMOS circuit power contracts and the fixed laser power becomes more dominant.

Figure 3 illustrates a general *N*-channel DWDM optical link architecture, comprising an integrated *N*-wavelength on-chip integrated DFB laser array, SOAs, MRMs, ring demultiplexers and high-speed PDs. A closer look at the transmitter[35] (Fig. 3a) shows a flip-chip assembly of electronic integrated circuit (EIC) and PIC. In the PIC, an eight-wavelength DFB laser array feeds an optical multiplexer, with each lane modulated by its own MRM that integrates a driver, photocurrent sensor, resonance tuning heater and thermal control unit (TCU).

Transmitter. In Fig. 3b, the optical transmitter (Tx) receives an external half-rate clock. The signal is first converted to CMOS levels and distributed to each Tx channel via on-die transmission lines. It then undergoes duty-cycle correction and division to produce four-phase quarter-rate clocks, which drive the 4:1 serializer. The full-rate pre-driver stage generates the inputs for the doubled-voltage driver, delivering 1.8 Vpp to modulate the MRM.

Thermal control unit. Conventional sensing schemes for MRM thermal control rely on external monitor photodiodes, which add multiple electro-optical interfaces and limit integration. By contrast, the reverse-biased p–n junction within the MRM generates a photocurrent that directly reflects the average absorbed optical power, enabling intrinsic monitoring without additional components. The TCU, in Fig. 3c, uses transimpedance amplifiers (TIAs) to amplify both the ring photocurrent and the input monitor PD photocurrent, and then adjusts their ratio to maintain precise spectral alignment between the ring resonators and the laser array. In addition to the TIA, the TCU incorporates a 32-bit counter, a sigma-delta modulator, a current digital-to-analog converter and a square-root compression circuit. Together, these components drive the power digital-to-analog converter, enabling sub-picometre resonance tuning and ensuring consistent performance across all wavelength channels.

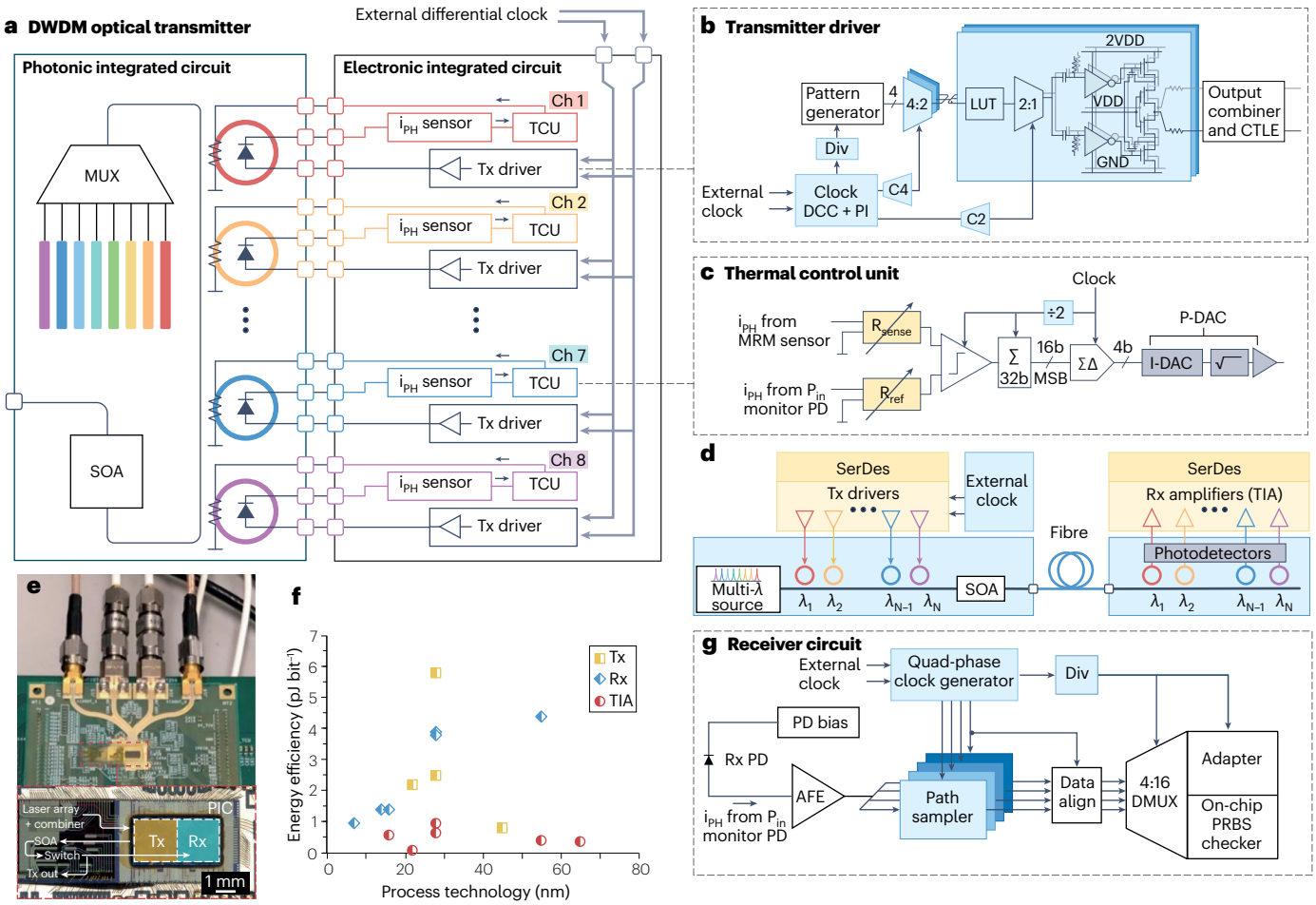

Fig. 3 | Representative architecture of microring modulator-based optical transmitters leveraging dense wavelength-division multiplexing. a, A dense wavelength-division multiplexing (DWDM) optical transmitter (Tx). **b**, Circuit diagram of a per-channel modulator driver. **c**, Circuit schematic of ring thermal control unit (TCU). **d**, Schematic of a complete DWDM optical link. **e**, Photograph of an electronic integrated circuit (EIC) mounted atop a photonic integrated circuit (PIC) on a printed circuit board. **f**, Comparative analysis of energy efficiency achieved in different optical transceiver sub-modules using different process technologies. **g**, Schematic of receiver circuit. AFE, analog front-end; C2, half-rate clock; C4, quarter-rate clock; CTLE, continuous-time linear equalization; DCC, duty-cycle correction; Div, clock divider; DMUX, demultiplexer; GND, ground; i_{PH}, photocurrent; I-DAC, current-output digital-to-analog converter; LUT, lookup table; MRM, microring modulator; MSB, most significant bit; MUX, multiplexer; P-DAC, power digital-to-analog converter; PD, photodetector; PI, phase interpolator; P_{in}, input power; PRBS, pseudorandom binary sequence; R_{sense}, sensor transimpedance amplifier feedback resistance; R_{ref}, reference transimpedance amplifier feedback resistance; Rx, receiver; SERDES, serializer/deserializer; SOA, semiconductor optical amplifier; TIA, transimpedance amplifier; VDD, supply voltage. Parts **a**–**c** are adapted with permission from ref. 102, IEEE. Part **e** is adapted with permission from ref. 104, IEEE. Part **g** is adapted with permission from ref. 14, JSAP.

Receiver. The optical receiver (Rx)[14] in Fig. 3 starts with a Ge-based PD. Its current signal is amplified by a regulated shunt feedback TIA, which determines the sensitivity, noise and high bandwidth of the receiver. Equalizers then correct bandwidth-limited impairments. The signal is translated to CMOS levels by an integrating sample-and-hold circuit, followed by an amplifying stage and regenerative latches. A quadrature phase rotator in the clock path fine-tunes the sampling phase, enabling reliable high-speed data recovery.

Energy efficiency. The energy efficiency of electronic circuits is a key factor in the overall performance of photonic systems. Driver circuits, TIAs, DSP and SerDes all contribute to the power budget, and their optimization is necessary to meet the stringent energy targets of modern applications. Figure 3f compares the energy per bit of TIAs, Rxs and Txs implemented in different processing technologies. At the front end, the TIA, as the first active stage, sets the link noise floor and the minimum optical sensitivity, with recent CMOS advances reducing its consumption to below 0.1 pJ bit^{-1} (ref. 106). Building on this, complete Rxs now consume 0.96–4.4 pJ bit^{-1} at 50–100 Gb s^{-1}, with the best energy efficiency achieved in 7 nm process technology[107]. On the Txs side, DWDM channels operate at 25–112 Gb s^{-1} per wavelength, with state-of-the-art designs at 25 Gb s^{-1} requiring just 0.8 pJ bit^{-1} (ref. 108). Together, these results underscore how aggressive process scaling and co-optimized circuit-photonic design can deliver substantial energy savings across the optical I/O chain.

Integration and development pathway

The integration of photonic and electronic dies, whether monolithically or through hybrid 2D, 2.5D and 3D assembly, has become a pivotal system-level decision. This section compares the leading integration frameworks, highlights their trade-offs in cost, bandwidth density, thermal management and yield and outlines a roadmap towards full-3D electronic–photonic integration.

Monolithic integration of CMOS and silicon PICs

Although conceptually similar to the monolithic integration of additional materials within silicon photonics, discussed earlier in the section 'Multimaterial integration', the efforts to combine EICs and PICs into monolithic electronic–photonic integrated circuits (EPICs) is a separate topic and presents distinct challenges. The mainstream strategy is to design photonic components that can be fabricated directly within existing electronic process nodes, with only minimal modifications to the established fabrication flows (Fig. 4a). At first glance, this seems to be the most straightforward solution. It eliminates the need for interconnection pads and bumps between separate chips, reducing impedance mismatch and notably simplifying packaging. Co-locating driving electronics near active photonic components also enables nearly parasitic-free operation, which greatly improves overall system performance.

In practice, however, fabrication remains complex. Photonic and electronic devices differ in size, geometry, material stacks and thermal budgets, leading to substantial barriers for co-integration. The most

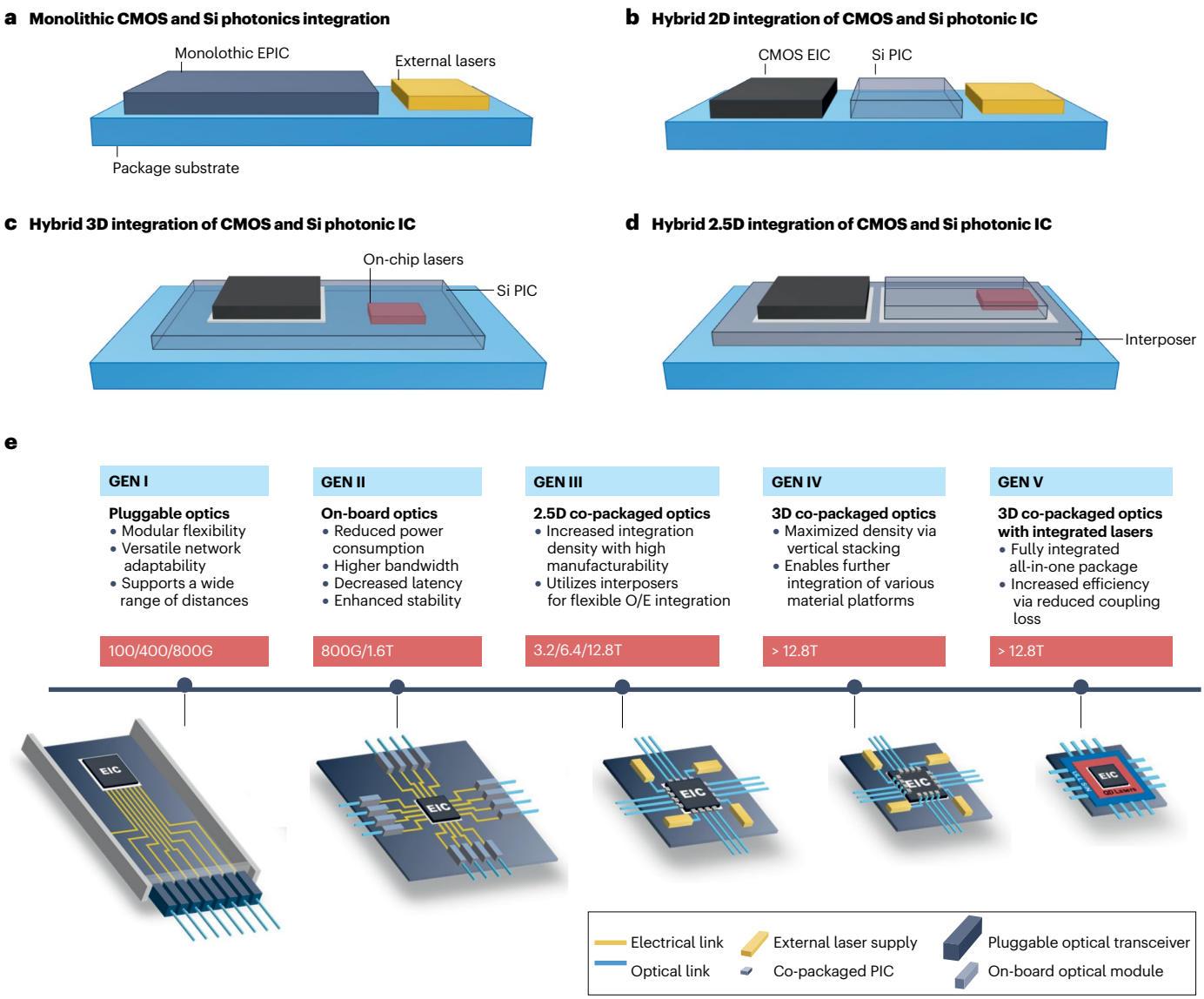

Fig. 4 | Silicon photonic integration approaches and evolutionary pathway. A schematic illustration of the monolithic complementary metal–oxide–semiconductor (CMOS) and silicon electronic–photonic integrated circuit (EPIC) integration approach with external lasers (part **a**) and hybrid 2D integration of CMOS electronic integrated circuit (EIC) with silicon photonic integrated circuit (PIC) with external lasers (part **b**). **c**, 3D assembled CMOS and silicon photonics integration with integrated on-chip laser in the silicon PIC. **d**, 2.5D integration of CMOS EIC and silicon PIC with integrated lasers using an interposer. **e**, Evolution of photonic integration and co-packaging technologies in data centre applications. O/E, optoelectronic. Part **e** is adapted from ref. 3, CC BY 4.0.

notable challenge is scale: photonic components are much larger than electronic devices in advanced nodes. Consequently, monolithic platforms have so far been demonstrated only in older process nodes, such as 45 nm and 32 nm (ref. 109), well behind state-of-the-art fin field-effect transistor technology. This mismatch compromises transistor performance, increasing energy consumption and fanout delays, while offering only modest photonic improvements, such as higher voltage tolerances from thick-gate devices. Reported platforms also lag behind their counterparts in waveguide losses, PD responsivity and PD bandwidth[110]. These issues reflect conflicting trade-offs in process node requirements. Older nodes accommodate the larger footprint of photonic structures but limit transistor performance, whereas advanced nodes improve transistor speed but constrain photonic design. Integrating both domains on a single die further introduces competition for silicon area, increasing the overall chip size. Costs add another constraint: embedding photonics within advanced CMOS nodes, which evolve every 18–24 months, adds notable expense and raises questions about long-term sustainability.

Despite these challenges, progress continues. An early example was a 4×10 Gb s^{-1} DWDM transceiver chip demonstrated in 2007, integrating photonic and electronic functions on a single silicon-on-insulator substrate[111]. More recently, a 45 nm EPIC platform[112] has delivered a dual-channel receiver at 0.9 pJ bit^{-1} energy efficiency up to 40 Gb s^{-1} (ref. 113), and on-chip feedback has stabilized quantum-correlated photon-pair sources in microrings[114]. Meanwhile, industry adoption has leaned towards fabricating discrete PICs and EICs to decouple their process flows, allowing each to use the most suitable node. This approach has already proven successful, with more than two million pluggable optical transceiver modules shipped by 2018 (ref. 115).

Looking ahead, research is targeting tighter co-design of electronics and photonics, smarter power distribution and selective use of hybrid add-ons to bridge remaining optical gaps. Although monolithic EPICs still face strong competition from hybrid assembly, their potential to streamline packaging and reduce costs keeps them under active investigation. Whether this approach can ultimately deliver superior cost-effectiveness across diverse applications remains an open question.

Hybrid assembly of CMOS and silicon PICs

Currently, the hybrid assembly of CMOS EICs and silicon PICs remains the dominant strategy to integrate photonic and electronic functionalities. Unlike monolithic integration, hybrid assembly maintains separate fabrication processes for each technology and then combines them at the packaging level. This separation allows each domain to be independently optimized: electronics can use the most advanced CMOS process nodes, whereas photonics can build on established silicon photonics platforms tailored for optical performance[116]. Over the years, the hybrid method has evolved through multiple integration schemes, including 2D, 3D and 2.5D architectures, each presenting distinct trade-offs in terms of interconnect density, signal integrity, thermal management and cost.

Two-dimensional integration. Early hybrid modules followed a 2D integration approach, in which separate CMOS and PIC chips were mounted side-by-side on the same package substrate, interconnected via wire bonds (Fig. 4b). Examples range from a five-channel microring transmitter at 25 Gb s^{-1} per channel[117] to a 112 Gb s^{-1} pulse-amplitude modulation with four levels MRM transmitter with on-chip lasers and a co-packaged CMOS driver[118]. This method is relatively simple, which eases the packaging process, but suffers from several drawbacks. Wire bonds introduce latency, impedance mismatches and parasitic of 0.5–1.0 nH mm^{-1} (ref. 119), which restricts bandwidth to <30 GHz in high-speed applications. Compared with more advanced 3D integrations, 2D approaches typically exhibit lower energy efficiencies (>5 pJ bit^{-1}) owing to larger interconnect losses, higher driver power requirements and thermal constraints. Additionally, the lateral placement of components limits bandwidth density, as EIC–PIC interconnects can only run along a single chip edge.

Three-dimensional integration. To overcome the limitations of 2D integration, 3D stacking techniques were developed, in which EICs and PICs are vertically integrated using advanced interconnects such as through-silicon vias, copper pillars and microsolder bumps[106] (Fig. 4c). This architecture shortens interconnect lengths, which decreases signal losses and power consumption, enhances bandwidth and reduces latency. The tighter stacking shrinks the overall module footprint while also enabling finer interconnect scaling. Progress in direct bonding, photolithography and wafer-level alignment has pushed bump pitch from 40–50 µm in early multichip modules[120] to 10–40 µm in advanced process lines[121], with prototypes showing 2 µm die-to-wafer bonding[122] and even 400 nm wafer-to-wafer bonding[123].

Despite these advantages, 3D integration presents challenges related to thermal management. Vertical stacking increases power density, which can impair the performance of thermally sensitive devices such as resonators and lasers. Effective thermal dissipation, using advanced thermal interface materials and efficient heat spreaders, is therefore essential for stable operation. Manufacturing also faces sub-micrometre alignment tolerances that reduce yield, along with the need to ensure compatibility across diverse material systems and fabrication processes[97].

Two-and-a-half-dimensional integration. As a compromise between 2D and 3D integration, 2.5D architectures leverage silicon interposers or redistribution layers to bridge CMOS and PIC chips with high-density interconnects (Fig. 4d). In this scheme, EICs and PICs are separately fabricated and then mounted onto a common silicon or glass interposer, which acts as a dense, high-speed interconnection fabric[124,125]. 2.5D assembly offers improved bandwidth density and reduced latency compared with 2D integration. At the same time, it avoids many of the thermal challenges associated with 3D stacking[110]. This has made 2.5D integration attractive for high-performance applications such as optical transceivers and AI accelerators, in which both power efficiency and interconnect density are paramount. The trade-off, however, is that the additional interposer layer increases manufacturing complexity and cost, making it less suitable for cost-sensitive applications.

Market trends and innovations

Advances in integration have turned silicon photonics into a market exceeding $3 billion, projected to grow at a compound annual growth rate of >20% (ref. 126). As bandwidth requirement increases, reducing power losses and signal degradation from copper interconnects has become critical. To address this issue, photonic modules are moving closer to the host application-specific integrated circuit (ASIC), evolving from pluggable modules to on-board optics (OBO) and now to CPO positioned beside the switch or accelerator die (Fig. 4e).

This shift towards tighter integration fundamentally reshapes the role of electronics in silicon photonics. Architectures such as pluggable optics, OBO, CPO and linear drive pluggable optics (LPOs) each impose

unique requirements on integration density, signal integrity and power efficiency. The following section traces how these market drivers and technical advances are steering the roadmap towards ever-deeper photonic–electronic integration.

Pluggable optics. Pluggable transceivers have been the workhorse of high-speed optical communications in data centres and telecommunication networks for more than a decade. State-of-the-art modules, operating at 50–100 Gb s^{-1} per lane under multisource agreements, achieve energy efficiencies of around 15–30 pJ bit^{-1}. However, as data rates increase to 400 Gb s^{-1}, 800 Gb s^{-1} and beyond, pluggable optics face growing challenges. Long copper traces between the ASIC and the pluggable module degrade signal integrity, which requires higher driver power and limits the achievable bandwidth. Thermal management also becomes increasingly difficult at these speeds, with heat dissipation emerging as a critical limit to reliable operation. These challenges are driving the industry to explore alternative packaging solutions that shorten electrical paths and to minimize insertion losses. However, pluggable modules remain widely used, valued for their modular flexibility, high interoperability and mature manufacturing ecosystem. They also continue to evolve: coherent optics and other innovations are extending their capabilities, with 1.6 Tb s^{-1} modules already been introduced to support the scaling demands of AI infrastructure[127].

On-board optics. OBO represents a middle ground between pluggable transceivers and CPO. By placing optical engines directly on the host board, OBO shortens the electrical paths between the transceivers and switch ASICs, reducing signal loss and lowering power consumption. Initial demonstrations are found in HPC systems[128], and a multisource agreement has since defined modules operating at 400 Gb s^{-1} and 800 Gb s^{-1} rates. Despite these advances, OBO provides only limited improvement in switch-to-module channel length and occupies substantial footprint on the main printed circuit board. As a result, industry adoption has been modest, with focus quickly shifting towards CPO solutions that offer tighter integration.

Co-packaged optics. Although pluggable transceivers have long been the mainstay of Ethernet connectivity in data centres, the mid-2010s marked the start of CPO development. CPO improves power efficiency, density and cost-effectiveness by placing the optics close to Ethernet switches, which shortens electrical channels and reduces their associated power. A major milestone came in 2020, when a 12.8 Tb s^{-1} Ethernet switch with CPO transceiver engines was demonstrated. The CPO engines, built from PICs with integrated lasers, delivered 1.6 Tb s^{-1} of optical capacity, achieved a 40-fold improvement in bandwidth density and more than 30% reduction in power consumption[18,129]. By 2023, multiple companies have introduced their own CPO-based Ethernet switches[130–132]. Unlike the 2020 prototype, these implementations relied on discrete EICs and PICs powered by fibre-coupled remote laser modules. Separating the lasers from the engines allows higher bandwidth density but adds complexity, as laser packaging increases assembly cost and introduces optical loss penalty.

Linear drive pluggable optics. LPOs gained momentum in 2023 as an energy-efficient alternative solution for network connectivity. Unlike traditional pluggable modules, LPOs eliminate the bulky and costly retimer DSP inside the optics and instead rely on the switch SerDes to drive signals directly. This reduces latency and lowers power consumption by ~40% (ref. 133). Compared with CPO, LPOs represent a more immediate upgrade path for existing pluggable optics, although they demand more challenging signal integrity design and higher-quality printed circuit board package and connectors. Current implementations mainly use SiGe bipolar CMOS technology[134]. This platform supports the initial demonstrations, but its scalability to higher data rates remains uncertain. The linear drive technique is also being explored in CPO frameworks, with initial demonstrations using embedded continuous-time linear equalizers (ref. 135). Meanwhile, research is advancing to scale optical connectivity for AI/ML fabrics, where DWDM wavelength count, lane rate, fibre count and polarizations are treated as design vectors, each contributing to sustained bandwidth growth[136].

The diversity of applications

As components become more mature, silicon photonics is expanding beyond traditional datacom into computing, sensing and emerging quantum systems. These applications show how tailored photonic–electronic co-integration is turning device-level advances into system-level benefits, ranging from sub-picojoule per bit optical I/O for AI accelerators to neuromorphic processors.

Communications and interconnects

The potential of silicon photonics in communication and interconnect applications is demonstrated through various technologies, including IMDD transceivers based on Mach–Zehnder modulators[73,137] and MRMs[101,138], coherent transceivers[83] and photonic switches[139]. A key advancement in DWDM optical communication links, particularly for emerging optical compute I/O, is the integration of flip–chip assembled CMOS EIC on a PIC. The demonstrated link operates at 32 Gbps per channel with a bit error rate of $<1 \times 10^{-12}$ across eight wavelengths, achieving a total fibre bandwidth of 256 Gbps (ref. 104). The reported Tx, Rx and optical energy efficiencies are 1.35, 3.8 and 1.94 pJ bit^{-1}, respectively, for a total link energy efficiency of 7.08 pJ bit^{-1}. As CMOS process nodes continue to advance, power consumption is expected to decrease further, facilitating broader adoption of photonics in HPC.

Computation expansion with AI/ML integration

Silicon photonics is moving beyond traditional networking into computation, driven by the bandwidth growth of AI/ML workloads[136]. Training large language models already links more than 50,000 tensor processing unit chips[140], demanding robust, high-speed channels that can seamlessly link chip-level, cluster-level and data-centre-level components. Traditional digital electronics, such as GPUs and ASICs[141,142], are facing mounting challenges, such as millisecond-level latency, high energy consumption, excessive heat production and elevated interconnect costs[143,144].

Integrated photonic devices provide a path forward. MMIs and related components offer low-latency, high-bandwidth tensor processing and neuromorphic computing. When co-packaged or chiplet-mounted with CMOS-driven electronic control circuits, they scale to meet AI fabric demands. A recent 4 Tb s^{-1} bidirectional optical-compute-interconnect chiplet combines a laser-integrated PIC with a CMOS EIC via radiofrequency through-silicon vias[96]. Driven directly from fifth-generation peripheral component interconnect express (PCIe Gen 5) ports, it enables CPU-to-CPU optical links without retiming, targeting 5 pJ bit^{-1} energy efficiency and <10 ns latency — well beyond the performance achievable with state-of-the-art pluggable optics[96].

Glossary

Antiphase domains
(APDs). A crystalline defect region in which two adjacent domains of a compound semiconductor are shifted by a lattice translation.

Arrayed waveguide gratings
A photonic device that spatially separates or combines optical signals of different wavelengths using an array of waveguides with incrementally varied lengths.

Asymmetric step-graded filters
An optical filter composed of multiple layers with stepwise and non-uniform refractive index changes, designed to achieve asymmetric transmission characteristics.

Atomic-layer deposition
A vapour-phase thin-film synthesis technique based on sequential, self-limiting surface reactions that enable atomic-scale control of thickness in material growth on substrates.

Avalanche effect
A nonlinear carrier multiplication process in which energetic charge carriers gain sufficient energy to ionize atoms through collisions, generating additional carriers.

Bit error rate
A performance metric in communication systems defined as the ratio of incorrectly received bits to the total transmitted bits.

Burn-in
A reliability testing process in which a device is operated under elevated stress conditions for a defined period to accelerate early failures and ensure long-term stability.

Coarse wavelength-division multiplexing
A technology that multiplexes several widely spaced wavelengths of optical carrier signals onto a single optical fibre or waveguide.

Complementary metal–oxide–semiconductor
(CMOS). A semiconductor technology integrating complementary pairs of p-type and n-type metal–oxide–semiconductor field-effect transistors to implement logic functions.

Continuous-time linear equalizers
An analog filter that compensates for frequency-dependent losses in high-speed electrical interconnects.

Co-packaged optics
(CPO). An integration approach that places optical engines in close proximity to switching application-specific integrated circuits within a single package.

Dark currents
A current that flows through a photosensitive device even in the absence of incident light.

Defect-mediated absorption
A sub-bandgap optical absorption process in which photons are absorbed via electronic states introduced by defects or impurities in a material.

Dense wavelength-division multiplexing
(DWDM). A technology that multiplexes several closely spaced wavelengths of optical carrier signals onto a single optical fibre or waveguide.

Distributed feedback (DFB) lasers
A laser that incorporates a periodic grating within the gain medium to provide wavelength-selective feedback, offering single-frequency emission.

Edge couplers
A photonic device that couples light between an optical fibre and an integrated waveguide through an engineered waveguide facet at the chip edge.

Fin field-effect transistor
A metal–oxide–semiconductor transistor that uses a thin, fin-shaped channel rising above the substrate to improve electrostatic control.

Forward-error correction
A coding technique in which redundant information is added to a transmitted signal, allowing the receiver to detect and correct errors without retransmission.

Free spectral range
A parameter of a resonant optical cavity defined as the frequency or wavelength spacing between consecutive resonant modes.

Grating couplers
(GCs). A photonic device that uses a periodic grating structure on a waveguide surface to diffract and couple light to an optical fibre.

Heterodyne microwave generation
A technique that produces microwave signals by mixing two optical frequencies, with the difference frequency corresponding to the generated microwave output.

Intensity-modulated direct-detection
(IMDD). An optical communication scheme in which data are encoded onto the intensity of a light source and recovered through direct photodetection.

Large language models
A neural network trained on vast text corpora to learn statistical patterns of language, enabling capabilities such as text generation, reasoning and question answering across diverse domains.

Linewidth enhancement factor
A laser parameter describing the coupling between changes in carrier density and refractive index, which influences spectral linewidth, frequency chirping and overall modulation dynamics.

Lithography masks
A patterned template used in photolithography to selectively transmit or block light onto a photoresist-coated substrate, enabling the transfer of microscale and nanoscale features during semiconductor fabrication.

Mach–Zehnder interferometer
(MZI). An optical device that splits light into two paths, introduces a relative phase shift and then recombines them to enable precise control of interference.

Misfit dislocations
A crystalline defect that forms at the interface between two lattice-mismatched materials to relieve strain.

Mode-locked lasers
A laser in which longitudinal modes are phase-locked to generate a stable train of ultrashort pulses or an optical frequency comb.

Multimode interference (MMI) devices
A photonic device that uses the interference of multiple propagating modes in a multimode waveguide region to split, combine or route optical signals.

On-board optics
(OBO). An optical interconnect approach in which optical engines are integrated directly onto line cards or printed circuit board.

Optical isolators
An optical device that allows light transmission in only one direction to prevent unwanted feedback.

Phase-change materials
(PCMs). A class of materials that can reversibly switch between amorphous and crystalline states under external stimuli.

Photon-assisted tunnelling
A quantum transport process in which electrons tunnel through a potential barrier with the simultaneous absorption or emission of photons.

Glossary (continued)

Pluggable optical transceivers
A standardized hot-swappable module for bidirectional conversion between optical and electrical signals in communication systems.

Polarization beam splitter-rotators
A photonic device that separates orthogonal polarization states of light and simultaneously rotates one polarization to match the other.

Pulse-amplitude modulation with four levels
A multilevel modulation format that encodes data into four discrete signal amplitudes.

Quadrature phase rotator
A circuit that shifts the phase of in-phase and quadrature signal components by precise increments.

Quantum cascade lasers
A unipolar semiconductor laser that generates coherent light through intersubband electronic transitions in a repeated quantum well structure for emissions across the mid-infrared to terahertz spectral ranges.

Quantum dot
(QD). A nanoscale semiconductor structure that confines electrons in all three spatial dimensions, producing discrete energy levels and size-tunable optical and electronic properties.

Quantum well
(QW). A nanostructure formed by confining charge carriers in a thin semiconductor layer sandwiched between materials with larger bandgaps.

Serializer/deserializer
(SerDes). A circuit that converts parallel data into serial form for transmission and reconverts it to parallel form at the receiver.

Side-mode suppression ratios
The ratio of power in the dominant lasing mode to that in the strongest side mode, quantifying the spectral purity of a laser.

Sigma-delta modulator
An oversampling analog-to-digital conversion architecture that shapes quantization noise by feeding back the error between the input and quantized output.

Spot-size converters
A photonic device that gradually transforms the mode field diameter of light between components.

Stacking faults
A planar crystalline defect caused by a disruption in the regular stacking sequence of atomic planes.

Tensor processing unit
An application-specific integrated circuit optimized for accelerating machine-learning workloads, particularly matrix and tensor operations in neural network training and inference.

Thermal cycle annealing
A post-growth process in which a material is repeatedly heated and cooled through controlled temperature cycles to reduce defects, relieve strain and improve crystalline quality.

Threading dislocations
A crystalline defect in an epitaxial semiconductor layer where a dislocation line threads through the film thickness.

Through-silicon vias
A vertical electrical interconnect that passes through a silicon wafer or die.

Transimpedance amplifiers
(TIAs). An electronic circuit that converts an input current into a proportional voltage signal with amplification.

Transparent conductive oxides
A class of wide-bandgap oxide materials that combine optical transparency with electrical conductivity.

Two-photon absorption
A nonlinear optical process in which two photons are absorbed simultaneously to excite an electron from a lower to a higher energy state.

Vapour deposition
A thin-film synthesis technique in which material is deposited onto a substrate from the vapour phase.

Neuromorphic photonics offers further speed-and-energy gains by leveraging the inherent advantages of photonics, such as low latency, wide bandwidth and high parallelism[145]. Starting from early explorations, such as free-space optical computing, optical reservoir computing and optical digital computing[146], integrated photonics-based neuromorphic computing has attracted extensive attention owing to its higher compactness, energy efficiency and electronics compatibility. Academic efforts have focused on developing coherent photonic tensor cores based on broadband devices, such as MZIs[147,148], phase shifters[149], MMIs[150], star-couplers[151] and metalenses[152], as well as incoherent designs based on microring resonators[153,154], frequency microcombs[155] and PCMs[91]. Commercial systems already utilize MZI networks for AI acceleration and solving non-deterministic polynomial-time complete problems[156,157]. These advancements underscore the potential of silicon photonics to deliver high-performance, energy-efficient computing that overcomes electronic limits and supports the next-generation AI infrastructure.

Emerging applications and opportunities

Beyond communications and AI, silicon photonics is positioned to transform precision sensing[158,159], quantum computing[160] and biomedical imaging[161]. Chip-scale photonic sensors combine CMOS-compatible fabrication with high sensitivity and compact footprints, enabling applications such as trace-gas monitors, biological marker detection and mechanical vibration analysis with sub-nanometre precision[162]. Emerging materials, including thin-film lithium niobate (TFLN), rare-earth-doped materials and PCMs, provide enhanced modulation, amplification and nonlinear optical functions. Commercially available high-quality TFLN[163] now supports integration-ready building blocks, including polarization management devices[164], electro-optic isolators[165] and femtosecond pulse generators[166], which can be directly integrated with silicon platforms. Scaling these technologies will require innovations in packaging and assembly. Techniques such as 3D integration and wafer-level packaging are being actively developed to enhance performance while controlling costs and ensuring thermal efficiency.

Conclusions

The growth of AI, ML and HPC has exposed fundamental limits of electrical I/O in energy per bit, reach and bandwidth density. CMOS-integrated silicon photonics now provides a practical and scalable path forward, with on-chip lasers and SOAs, compact modulators, high-speed detectors and low-loss routing and efficient chip-to-fibre couplers reaching production-ready maturity for data-centre-class links. Co-designed with CMOS drivers/SerDes/DSP, biasing and thermal control, these building blocks are reducing total link energy towards the sub-picojoule per bit range.

Integration strategies strongly influence performance, cost and manufacturability. Hybrid assembly, heterogeneous bonding, micro-transfer printing and monolithic epitaxy each extend functionality and wavelength coverage, whereas early multilayer 3D stacks show how photonics and electronics can be brought together at wafer scale. In parallel, packaging is shifting from pluggables to LPO and CPO, shortening electrical paths and increasing bandwidth density while introducing new thermal and yield constraints.

Future progress depends on several near-term enablers. On-chip comb sources can scale DWDM capacity. Wafer-level burn-in and testing will secure stability and yield earlier in the flow. Finer-pitch 3D integration can surpass reticle limits while shortening electronic and photonic interconnects. Equally important are manufacturing realities: photonics is analog, so small fabrication inaccuracies accumulate as phase/coupling errors, losses and stray-light scattering across large circuits. Design for tolerance, post-fabrication trimming or tuning, systematic yield improvement and attention to reticle-size limits are essential, which in turn point towards miniaturization and vertical stacking as key directions for future development.

New materials, such as TFLN, Ge-on-Si and selected 2D or magneto-optic films, expand wavelength range, power handling and add functions such as isolators and non-volatile memory. This diversity enables new performance, but it also dilutes volume and therefore yield. To manage this tension, strategies such as mix-and-match high-yield technology modules, chiplet-based back-end assembly, decentralized open-access foundries, data-driven module process design kits, well-specified electrical and optical interfaces and process-control monitors will be needed.

The targets for the next phase are higher symbol rates beyond 100 gigabaud, lower latency, denser interconnects, reduced loss and seamless co-integration of photonics and electronics near the compute die. Standardized reliability screening and practical metrics for energy and bandwidth density will help move optical I/O from a promising add-on to a core infrastructure for computing and communications, with parallel impact in compute I/O, neuromorphic acceleration, precision sensing and quantum technologies.

Published online: 07 November 2025

References

1. Mahajan, R. et al. Co-packaged photonics for high performance computing: status, challenges and opportunities. *J. Light Technol.* **40**, 379–392 (2022).
   Reviews the motivations and technical issues of co-packaged optics for data centres and HPC.
2. Krishnamoorthy, A. V. et al. From chip to cloud: optical interconnects in engineered systems. *J. Light Technol.* **35**, 3103–3115 (2017).
   A systems-level map of where optics wins by explaining link budgets, reach/regime choices (IMDD versus coherent) and implications from packages to data centres.
3. Margalit, N. et al. Perspective on the future of silicon photonics and electronics. *Appl. Phys. Lett.* **118**, 220501 (2021).
   Broad perspective on Si-photonics and electronics co-evolution.
4. Wu, Z. et al. Peta-Scale embedded photonics architecture for distributed deep learning applications. *J. Light Technol.* **41**, 3737–3749 (2023).
5. Kim, J. et al. 8.1 A 224 Gb/s DAC-based PAM-4 transmitter with 8-Tap FFE in 10 nm CMOS. In 2021 *IEEE International Solid-State Circuits Conference* (*ISSCC*), 126–128 (IEEE, 2021).
6. Segal, Y. et al. A 1.41 pJ/b 224 Gb/s PAM-4 SerDes receiver with 31 dB loss compensation. In 2022 *IEEE International Solid-State Circuits Conference* (*ISSCC*), 114–116 (IEEE, 2022).
7. Lim, K. et al. Disaggregated memory for expansion and sharing in blade servers. In *Proceedings of the 36th Annual International Symposium on Computer Architecture* 267–278 (ACM, 2009).
8. Michelogiannakis, G. et al. A case for intra-rack resource disaggregation in HPC. *ACM Trans. Archit. Code Optim.* **19**, 1–26 (2022).
9. Gonzalez, J. et al. Optically connected memory for disaggregated data centers. *J. Parallel Distrib. Comput.* **163**, 300–312 (2022).
10. Sun, J. et al. A 128 Gb/s PAM4 silicon microring modulator with integrated thermo-optic resonance tuning. *J. Light Technol.* **37**, 110–115 (2019).
11. Sun, C. et al. A 45 nm CMOS-SOI monolithic photonics platform with bit-statistics-based resonant microring thermal tuning. *IEEE J. Solid-State Circuits* **51**, 893–907 (2016).
12. Huang, D. et al. 8-Channel hybrid III–V/silicon DFB laser array with highly uniform 200 GHz spacing and power. In 2021 27*th International Semiconductor Laser Conference* (*ISLC*) 1–2 (IEEE, 2021).
13. Kumar, R., Doylend, J., Sakib, M., Sun, J. & Rong, H. Demonstration of an on-chip III–V/Si hybrid semiconductor optical amplifier for photonics integration. In 2018 *IEEE* 15*th International Conference on Group IV Photonics* (*GFP*) 1–2 (IEEE, 2018).
14. Xuan, Z. et al. A 256 Gbps heterogeneously integrated silicon photonic microring-based DWDM receiver suitable for in-package optical I/O. In 2023 *IEEE Symposium on VLSI Technology and Circuits* (*VLSI Technology and Circuits*) 1–2 (IEEE, 2023).
15. Guldberg-Kjær, S. et al. Planar waveguide laser in Er/Al-doped germanosilicate. *Electron. Lett.* **35**, 302–303 (1999).
16. Rong, H. et al. A continuous-wave Raman silicon laser. *Nature* **433**, 725–728 (2005).
17. Fang, A. W. et al. Electrically pumped hybrid AlGaInAs-silicon evanescent laser. *Opt. Expr.* **14**, 9203–9210 (2006).
18. Fathololoumi, S. et al. 1.6 Tbps silicon photonics integrated circuit and 800 Gbps photonic engine for switch co-packaging demonstration. *J. Light Technol.* **39**, 1155–1161 (2021).
    Provides one of the first complete demonstrations of a silicon-photonic engine for co-packaged optics in real systems.
19. Xiang, C. et al. Narrow-linewidth III–V/Si/Si_3N_4 laser using multilayer heterogeneous integration. *Optica* **7**, 20–21 (2020).
20. Tran, M. A. et al. Extending the spectrum of fully integrated photonics to submicrometre wavelengths. *Nature* **610**, 54–60 (2022).
21. Xiang, C. et al. 3D integration enables ultralow-noise isolator-free lasers in silicon photonics. *Nature* **620**, 78–85 (2023).
    Shows system-grade, isolator-free on-chip lasers via multilayer heterogeneous integration, proving that integrated sources can meet coherence and stability demands.
22. Wan, Y. et al. High speed evanescent quantum-dot lasers on Si. *Laser Photon. Rev.* **15**, 2100057 (2021).
    Demonstrates fast, low-threshold isolator-free quantum dot lasers heterogeneously integrated on silicon, showing a realistic path to efficient on-chip light sources.
23. Shang, C. et al. High-temperature reliable quantum-dot lasers on Si with misfit and threading dislocation filters. *Optica* **8**, 749–754 (2021).
24. Rafailov, E. U., Cataluna, M. A. & Sibbett, W. Mode-locked quantum-dot lasers. *Nat. Photon.* **1**, 395–401 (2007).
25. Wan, Y., Norman, J., Liu, S., Liu, A. & Bowers, J. E. Quantum dot lasers and amplifiers on silicon: recent advances and future developments. *IEEE Nanotechnol. Mag.* **15**, 8–22 (2021).
26. Shang, C. et al. Perspectives on advances in quantum dot lasers and integration with Si photonic integrated circuits. *ACS Photon.* **8**, 2555–2566 (2021).
    Summarizes the progress and challenges of quantum dot lasers and their integration with silicon photonics.
27. Park, H. et al. A hybrid AlGaInAs–silicon evanescent amplifier. *IEEE Photon. Technol. Lett.* **19**, 230–232 (2007).
28. Juodawlkis, P. W. et al. High-power, low-noise 1.5-μm slab-coupled optical waveguide (SCOW) emitters: physics, devices, and applications. *IEEE J. Sel. Top. Quant. Electron.* **17**, 1698–1714 (2011).
29. Hasegawa, H., Funabashi, M., Yokouchi, N., Kiyota, K. & Maruyama, K. Design and fabrication of semiconductor optical amplifier with low noise figure. In *OECC* 2010 *Technical Digest* 204–205 (IEEE, 2010).
30. Morito, K., Tanaka, S., Tomabechi, S. & Kuramata, A. A broad-band MQW semiconductor optical amplifier with high saturation output power and low noise figure. *IEEE Photon. Technol. Lett.* **17**, 974–976 (2005).
31. Bernhardi, E. H. et al. Ultra-narrow-linewidth, single-frequency distributed feedback waveguide laser in Al_2O_3:Er^{3+} on silicon. *Opt. Lett.* **35**, 2394–2396 (2010).
32. Liu, Y. et al. A photonic integrated circuit-based erbium-doped amplifier. *Science* **376**, 1309–1313 (2022).
33. Liu, A. et al. A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor. *Nature* **427**, 615–618 (2004).
    Landmark demonstration that established practical, CMOS-compatible high-speed modulation in silicon.
34. Reed, G. T., Mashanovich, G., Gardes, F. Y. & Thomson, D. J. Silicon optical modulators. *Nat. Photon.* **4**, 518–526 (2010).
    Comprehensive early review that organizes modulation mechanisms and trade-offs in silicon, giving newcomers a guide for device selection, drive requirements and integration.
35. Levy, C. et al. A 3D-integrated 8λ×32 Gbps λ silicon photonic microring-based DWDM transmitter. In 2023 *IEEE Custom Integrated Circuits Conference* (*CICC*) 1–2 (IEEE, 2023).
36. He, M. et al. High-performance hybrid silicon and lithium niobate Mach–Zehnder modulators for 100 Gbit s^{-1} and beyond. *Nat. Photon.* **13**, 359–364 (2019).
37. Eltes, F. et al. A $BaTiO_3$-based electro-optic Pockels modulator monolithically integrated on an advanced silicon photonics platform. *J. Light Technol.* **37**, 1456–1462 (2019).
38. Hsu, W.-C. et al. Sub-volt high-speed silicon MOSCAP microring modulator driven by high-mobility conductive oxide. *Nat. Commun.* **15**, 826 (2024).
39. Boeuf, F., Han, J.-H., Takagi, S. & Takenaka, M. Benchmarking Si, SiGe, and III–V/Si Hybrid SIS optical modulators for datacenter applications. *J. Light Technol.* **35**, 4047–4055 (2017).
40. Rahim, A. et al. Taking silicon photonics modulators to a higher performance level: state-of-the-art and a review of new technologies. *Adv. Photon.* **3**, 024003 (2021).

41. Michel, J., Liu, J. & Kimerling, L. C. High-performance Ge-on-Si photodetectors. *Nat. Photon.* **4**, 527–534 (2010).
Showed that germanium detectors grown directly on silicon can work at high speed and low cost, making them the standard choice for receiving light signals in silicon photonic chips.
42. Sakib, M. et al. Demonstration of a 50 Gb/s all-silicon waveguide photodetector for photonic integration. In 2018 *Conference on Lasers and Electro-Optics* (*CLEO*) 1–2 (IEEE, 2018).
43. Peng, Y. et al. High-speed all-silicon double microring avalanche photodetectors. In 2023 *Opto-Electronics and Communications Conference* (*OECC*) 1–4 (IEEE, 2023).
44. Ma, C., Kumar, R., Sakib, M. N., Huang, D. & Rong, H. A precision in-situ waveguide loss measurement technique using in-line silicon photodetectors. In 2021 *IEEE 17th International Conference on Group IV Photonics* (*GFP*) 1–2 (IEEE, 2021).
45. Tossoun, B. et al. Indium arsenide quantum dot waveguide photodiodes heterogeneously integrated on silicon. *Optica* **6**, 1277–1281 (2019).
46. Wan, Y. et al. Monolithically integrated InAs/InGaAs quantum dot photodetectors on silicon substrates. *Opt. Expr.* **25**, 27715–27723 (2017).
47. Wei, J., Zhu, C., Yu, Y., Wang, R. & Yu, S. Low-loss silicon waveguide and an ultrahigh-Q silicon microring resonator in the 2 μm wave band. *Opt. Lett.* **49**, 3271–3274 (2024).
48. Bitincka, E., Gilardi, G. & Smit, M. K. On-wafer optical loss measurements using ring resonators with integrated sources and detectors. *IEEE Photon. J.* **6**, 1–12 (2014).
49. Yeh, S. K. et al. Silicon photonics platform for next generation data communication technologies. In 2024 *International Electron Devices Meeting* (*IEDM*) 1–4 (IEEE, 2024).
50. Liu, K. et al. 720 million quality factor integrated all-waveguide photonic resonator. In 2021 *Device Research Conference* (*DRC*) 1–2 (IEEE, 2021).
51. Jin, W. et al. Hertz-linewidth semiconductor lasers using CMOS-ready ultra-high-Q microresonators. *Nat. Photon.* **15**, 346–353 (2021).
Demonstrates CMOS-compatible ultra-high-*Q* resonators that enable semiconductor lasers with hertz-level linewidths, paving the way for ultra-low-noise sources in advanced communications and precision metrology.
52. Xiang, C., Jin, W. & Bowers, J. E. Silicon nitride passive and active photonic integrated circuits: trends and prospects. *Photon. Res.* **10**, A82–A96 (2022).
53. Hsia, H. et al. EPIC-BOE: an electronic–photonic chiplet integration technology with IC processes for broadband optical engine applications. In 2024 *International Electron Devices Meeting* (*IEDM*) 1–4 (IEEE, 2024).
54. Marchetti, R., Lacava, C., Carroll, L., Gradkowski, K. & Minzioni, P. Coupling strategies for silicon photonics integrated chips [invited]. *Photon. Res.* **7**, 201–239 (2019).
55. Marchetti, R. et al. High-efficiency grating-couplers: demonstration of a new design strategy. *Sci. Rep.* **7**, 16670 (2017).
56. Zhou, X. & Tsang, H. K. Photolithography fabricated sub-decibel high-efficiency silicon waveguide grating coupler. *IEEE Photon. Technol. Lett.* **35**, 43–46 (2023).
57. Halir, R. et al. Waveguide grating coupler with subwavelength microstructures. *Opt. Lett.* **34**, 1408–1410 (2009).
58. Benedikovic, D. et al. Subwavelength index engineered surface grating coupler with sub-decibel efficiency for 220-nm silicon-on-insulator waveguides. *Opt. Expr.* **23**, 22628–22635 (2015).
59. Sacher, W. D. et al. Wide bandwidth and high coupling efficiency Si_3N_4-on-SOI dual-level grating coupler. *Opt. Expr.* **22**, 10938–10947 (2014).
60. Zhou, X. & Tsang, H. K. Photolithography fabricated broadband waveguide grating couplers with 1 dB bandwidth over 100 nm. *IEEE Photon. J.* **16**, 1–6 (2024).
61. Sánchez-Postigo, A. et al. Breaking the coupling efficiency–bandwidth trade-off in surface grating couplers using zero-order radiation. *Laser Photon. Rev.* **15**, 2000542 (2021).
62. Yi, X. et al. Asymmetric bi-level dual-core mode converter for high-efficiency and polarization-insensitive O-band fiber-chip edge coupling: breaking the critical size limitation. *Nanophotonics* **13**, 4149–4157 (2024).
63. Pu, M., Liu, L., Ou, H., Yvind, K. & Hvam, J. M. Ultra-low-loss inverted taper coupler for silicon-on-insulator ridge waveguide. *Opt. Commun.* **283**, 3678–3682 (2010).
64. Papes, M. et al. Fiber-chip edge coupler with large mode size for silicon photonic wire waveguides. *Opt. Expr.* **24**, 5026–5038 (2016).
65. Hatori, N. et al. A hybrid integrated light source on a silicon platform using a trident spot-size converter. *J. Light Technol.* **32**, 1329–1336 (2014).
66. Song, B., Stagarescu, C., Ristic, S., Behfar, A. & Klamkin, J. 3D integrated hybrid silicon laser. *Opt. Expr.* **24**, 10435–10444 (2016).
67. Martir, A. et al. Hybrid silicon photonics flip–chip laser integration with vertical self-alignment. In 2017 *Conference on Lasers and Electro-Optics Pacific Rim* (*CLEO-PR*) 1–4 (IEEE, 2017).
68. Billah, M. R. et al. Hybrid integration of silicon photonics circuits and InP lasers by photonic wire bonding. *Optica* **5**, 876–883 (2018).
69. Lau, J. H. Recent advances and trends in advanced packaging. *Manuf. Technol.* **12**, 228–252 (2022).
70. Bian, Y. et al. 3D integrated laser attach technology on a 300-mm monolithic CMOS silicon photonics platform. *IEEE J. Sel. Top. Quantum Electron.* **29**, 1–19 (2023).
71. Spott, A. et al. Quantum cascade laser on silicon. *Optica* **3**, 545–551 (2016).
72. Xiang, C. et al. High-performance silicon photonics using heterogeneous integration. *IEEE J. Sel. Top. Quant. Electron.* **28**, 1–15 (2022).
Explains how wafer bonding and heterogeneous integration deliver state-of-the-art device and system performance.
73. Jones, R. et al. Heterogeneously integrated InP\/silicon photonics: fabricating fully functional transceivers. *IEEE Nanotechnol. Mag.* **13**, 17–26 (2019).
74. Roelkens, G. et al. Micro-transfer printing for heterogeneous Si photonic integrated circuits. *IEEE J. Sel. Top. Quant. Electron.* **29**, 1–14 (2023).
75. Gomez, D. et al. Micro transfer printing for micro assembly of heterogeneous integrated compound semiconductor components. In *Compound Semiconductor Manufacturing Technology* (*CS MANTECH*) *Conference* 201–205 (CS MANTECH, 2022).
76. McPhillimy, J. et al. Automated nanoscale absolute accuracy alignment system for transfer printing. *ACS Appl. Nano Mater.* **3**, 10326–10332 (2020).
77. Shang, C. et al. A pathway to thin GaAs virtual substrate on on-axis Si (001) with ultralow threading dislocation density. *Phys. Stat. Sol. A* **218**, 2000402 (2021).
78. Selvidge, J. et al. Defect filtering for thermal expansion induced dislocations in III–V lasers on silicon. *Appl. Phys. Lett.* **117**, 122101 (2020).
79. Wan, Y. et al. Optically pumped 13 μm room-temperature InAs quantum-dot micro-disk lasers directly grown on (001) silicon. *Opt. Lett.* **41**, 1664–1667 (2016).
80. Jung, D. et al. High efficiency low threshold current 1.3 μm InAs quantum dot lasers on on-axis (001) GaP/Si. *Appl. Phys. Lett.* **111**, 122107 (2017).
81. Norman, J. C. et al. A review of high-performance quantum dot lasers on silicon. *IEEE J. Quant. Electron.* **55**, 1–11 (2019).
82. Wan, Y. et al. 1.3 μm submilliamp threshold quantum dot micro-lasers on Si. *Optica* **4**, 940–944 (2017).
83. Liu, S. et al. High-channel-count 20 GHz passively mode-locked quantum dot laser directly grown on Si with 41 Tbit/s transmission capacity. *Optica* **6**, 128–134 (2019).
84. Wan, Y. et al. 1.3 μm quantum dot-distributed feedback lasers directly grown on (001) Si. *Laser Photon. Rev.* **14**, 2000037 (2020).
85. Wan, Y. et al. Tunable quantum dot lasers grown directly on silicon. *Optica* **6**, 1394–1400 (2019).
86. Wei, W.-Q. et al. Monolithic integration of embedded III–V lasers on SOI. *Light Sci. Appl.* **12**, 84 (2023).
87. Shang, C. et al. Electrically pumped quantum-dot lasers grown on 300 mm patterned Si photonic wafers. *Light Sci. Appl.* **11**, 299 (2022).
88. Liang, D. et al. High-performance quantum-dot distributed feedback laser on silicon for high-speed modulations. *Optica* **8**, 591–593 (2021).
89. Wei, M. et al. Monolithic back-end-of-line integration of phase change materials into foundry-manufactured silicon photonics. *Nat. Commun.* **15**, 2786 (2024).
90. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
91. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
92. Fang, Z. et al. Ultra-low-energy programmable non-volatile silicon photonics based on phase-change materials with graphene heaters. *Nat. Nanotechnol.* **17**, 842–848 (2022).
93. Han, C. et al. Slow-light silicon modulator with 110-GHz bandwidth. *Sci. Adv.* **9**, eadi5339 (2023).
94. Lischke, S. et al. Ultra-fast germanium photodiode with 3-dB bandwidth of 265 GHz. *Nat. Photon.* **15**, 925–931 (2021).
95. Koch, B. et al. A 4×12.5 Gb/s CWDM Si photonics link using integrated hybrid silicon lasers. In 2011 *Conference on Lasers and Electro-Optics* (*CLEO*) 1–2 (IEEE, 2011).
96. Urricariet, C. Intel® shows OCI optical I/O chiplet co-packaged with CPU at OFC2024, enabling explosive AI scaling. Intel Community. https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/Intel-Shows-OCI-Optical-I-O-Chiplet-Co-packaged-with-CPU-at/post/1582541 (2024).
97. Xiang, C. & Bowers, J. E. Building 3D integrated circuits with electronics and photonics. *Nat. Electron.* **7**, 422–424 (2024).
Outlines the future vision of multilayer 3D electronic–photonic integration as a scalable platform.
98. Miller, D. Device requirements for optical interconnects to silicon chips. *Proc. IEEE* **97**, 1166–1185 (2009).
Defines the performance targets and limits for optical interconnects, shaping how researchers think about energy, loss, coupling and system design.
99. Guansheng, L. et al. 18.1 A 600Gb/s DP-QAM64 coherent optical transceiver frontend with 4x105GS/s 8b ADC/DAC in 16nm CMOS. In 2024 *IEEE International Solid-State Circuits Conference* (*ISSCC*) 338–340 (IEEE, 2024).
100. Sakib, M. et al. A high-speed micro-ring modulator for next generation energy-efficient optical networks beyond 100 Gbaud. In 2021 *Conference on Lasers and Electro-Optics* (*CLEO*) 1–2 (IEEE, 2021).
101. Wade, M. et al. An error-free 1 Tbps WDM optical I/O chiplet and multi-wavelength multi-port laser. In 2021 *Optical Fiber Communications Conference and Exhibition* (*OFC*) 1–3 (IEEE, 2021).
102. Sharma, J. et al. Silicon photonic microring-based 4 × 112 Gb/s WDM transmitter with photocurrent-based thermal control in 28-nm CMOS. *IEEE J. Solid-State Circuits* **57**, 1187–1198 (2022).
103. Li, H. et al. A 3-D-integrated silicon photonic microring-based 112-Gb/s PAM-4 transmitter with nonlinear equalization and thermal control. *IEEE J. Solid-State Circuits* **56**, 19–29 (2021).
104. Huang, D. et al. A silicon photonic 8λ, x 32Gbps/λ, WDM transceiver with integrated laser array and SOA for optical I/O. In 2023 *International Electron Devices Meeting* (*IEDM*) 1–4 (IEEE, 2023).
105. Daudlin, S. et al. Ultra-dense 3D integrated 5.3 Tb/s/mm^2 80 micro-disk modulator transmitter. In 2023 *Optical Fiber Communications Conference and Exhibition* (*OFC*) 1–3 (IEEE, 2023).
106. Daneshgar, S., Li, H., Kim, T. & Balamurugan, G. A 128 Gb/s, 11.2 mW single-ended PAM4 linear TIA with 2.7 μa_{rms} input noise in 22 nm FinFET CMOS. *IEEE J. Solid-State Circuits* **57**, 1397–1408 (2022).

107. Raj, M. et al. A O.96pj/b 7×50Gb/s-per-fiber WDM receiver with stacked 7nm CMOS and 45nm silicon photonic dies. In 2023 *IEEE International Solid-State Circuits Conference* (*ISSCC*) 11–13 (IEEE, 2023).
108. Sun, C. et al. TeraPHY: an O-band WDM electro-optic platform for low power, terabit/s optical I/O. In 2020 *IEEE Symposium on VLSI Technology* 1–2 (IEEE, 2020).
109. Stojanović, V. et al. Monolithic silicon-photonic platforms in state-of-the-art CMOS SOI processes [Invited]. *Opt. Expr.* **26**, 13106–13121 (2018).
Clarifies the 'zero-change CMOS' approach — what co-fabrication buys (and what it costs) in speed, voltage swing, area and manufacturing flow.
110. Abrams, N. C. et al. Silicon photonic 2.5D multi-chip module transceiver for high-performance data centers. *J. Light Technol.* **38**, 3346–3357 (2020).
111. Narasimha, A. et al. A fully integrated 4 × 10-Gb/s DWDM optoelectronic transceiver implemented in a standard 0.13 µm CMOS SOI technology. *IEEE J. Solid-State Circuits* **42**, 2736–2744 (2007).
112. Rakowski, M. et al. 45nm CMOS — Silicon Photonics Monolithic Technology (45CLO) for next-generation, low power and high speed optical interconnects. In 2020 *Optical Fiber Communications Conference and Exhibition* (*OFC*) 1–3 (IEEE, 2020).
113. Movaghar, G. et al. A 40-Gb/s, 900-fJ/bit dual-channel receiver in a 45-nm monolithic RF/photonic integrated circuit. *Process. IEEE Solid-State Circuits Lett.* **5**, 313–316 (2022).
114. Kramnik, D. et al. Scalable feedback stabilization of quantum light sources on a CMOS chip. *Nat. Electron.* **8**, 620–630 (2025).
115. Freeman, M. Cisco Systems is buying Carlsbad chip firm Luxtera for $660 million. The San Diego Union-Tribune. https://www.sandiegouniontribune.com/2018/12/18/cisco-systems-is-buying-carlsbad-chip-firm-luxtera-for-660-million (2018).
116. Moazeni, S. et al. Monolithic integration of O-band photonic transceivers in a 'zero-change' 32nm SOI CMOS. In 2017 *IEEE International Electron Devices Meeting* (*IEDM*) 24.3.1–24.3.4 (IEEE, 2017).
117. Li, H. et al. A 25 Gb/s, 4.4 V-swing, AC-coupled ring modulator-based WDM transmitter with wavelength stabilization in 65 nm CMOS. *IEEE J. Solid-State Circuits* **50**, 3145–3159 (2015).
118. Li, H. et al. A 112 Gb/s PAM4 silicon photonics transmitter with microring modulator and CMOS driver. *J. Light Technol.* **38**, 131–138 (2020).
119. Ndip, I. et al. Modeling and minimizing the inductance of bond wire interconnects. In 2013 *17th IEEE Workshop on Signal and Power Integrity* 1–4 (IEEE, 2013).
120. Rakowski, M. et al. Hybrid 14nm FinFET — silicon photonics technology for low-power Tb/s/mm^2 optical I/O. In 2018 *IEEE Symposium on VLSI Technology* 221–222 (IEEE, 2018).
121. Chen, M.-F., Chen, F.-C., Chiou, W.-C. & Yu, D. C. System on integrated chips (SoIC(TM)) for 3D heterogeneous integration. In 2019 *IEEE 69th Electronic Components and Technology Conference* (*ECTC*) 594–599 (IEEE, 2019).
122. Lin, Y. et al. Direct die-to-wafer hybrid bonding using plasma diced dies and bond pad pitch scaling down to 2 µm. In 2024 *IEEE 74th Electronic Components and Technology Conference* (*ECTC*) 40–44 (IEEE, 2024).
123. Chew, S.-A. et al. The challenges and solutions of Cu/SiCN wafer-to-wafer hybrid bonding scaling down to 400 nm pitch. In 2023 *International Electron Devices Meeting* (*IEDM*) 1–4 (IEEE, 2023).
124. Mahajan, R. et al. Embedded multi-die interconnect bridge (EMIB) — a high density, high bandwidth packaging interconnect. In 2016 *IEEE 66th Electronic Components and Technology Conference* (*ECTC*) 557–565 (IEEE, 2016).
125. Huang, P. K. et al. Wafer level system integration of the fifth generation CoWoS®-S with high performance Si interposer at 2500 mm^2. In 2021 *IEEE 71st Electronic Components and Technology Conference* (*ECTC*) 101–104 (IEEE, 2021).
126. Kozlov, V. May 2023 silicon photonics, linear drive pluggable and co-packaged optics. LightCounting Market Report. https://www.lightcounting.com/report/may-2023-silicon-photonics-linear-drive-pluggable-and-co-packaged-optics-203 (2023).
127. Xenos, H. Coherent comes to the data center: meet Ciena's 1.6 Tb/s Coherent-Lite pluggable. Ciena Blog. https://www.ciena.com/insights/blog/2024/coherent-comes-to-the-data-center-meet-cienas-1.6-tbs-coherent-lite-pluggable (2024).
128. Derradji, S., Palfer-Sollier, T., Panziera, J.-P., Poudes, A. & Atos, F. W. The BXI interconnect architecture. In 2015 *IEEE 23rd Annual Symposium on High-Performance Interconnects* 18–25 (IEEE, 2015).
129. Fathololoumi, S. et al. 1.6 Tbps silicon photonics integrated circuit for co-packaged optical-IO switch applications. In 2020 *Optical Fiber Communications Conference and Exhibition* (*OFC*) 1–3 (IEEE, 2020).
130. Margalit, N. Broadcom's persistent cadence of co-packaged optics innovation. Broadcom B-Connected Blog. https://www.broadcom.com/blog/broadcoms-persistent-cadence-copackaged-optics-innovation (2023).
131. Torza, A. Cisco demonstrates Co-Packaged Optics (CPO) system at OFC 2023. Cisco Blogs. https://blogs.cisco.com/sp/cisco-demonstrates-co-packaged-optics-cpo-system-at-ofc-2023 (2023).
132. Weissberger, A. Ranovus Monolithic 100G Optical I/O Cores for Next-Generation Data Centers. IEEE Communications Society Technology Blog. https://techblog.comsoc.org/2023/03/10/ranovus-monolithic-100g-optical-i-o-cores-for-next-generation-data-centers/ (2023).
133. Chou, E. S. et al. 100 G and 200 G per lane linear drive optics for data center applications. In 2024 *Optical Fiber Communications Conference and Exhibition* (*OFC*) 1–3 (IEEE, 2024).
134. Mahmud, M. H., Rubaye, H. A & Rebeiz, G. M. Broadband linear drivers for 800G/1.6 T energy efficient optical links. In 2024 *IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium* (*BCICTS*) 103–106 (IEEE, 2024).
135. Liu, H. et al. A 4×112 Gb/s PAM-4 silicon-photonic transmitter and receiver chipsets for linear-drive co-packaged optics. *IEEE J. Solid-State Circuits* **59**, 3263–3276 (2024).
136. Liao, L. et al. Silicon photonics for next-generation optical connectivity. In *Optical Fiber Communication Conference* (*OFC*) 2023, 1–3 (IEEE, 2023).
137. Alduino, A. Demonstration of a high speed 4-channel integrated silicon photonics WDM link with hybrid silicon lasers. In 2010 *IEEE Hot Chips22 Symposium*(*HCS*) 1–29 (IEEE, 2010).
138. Fathololoumi, S. et al. Highly integrated 4 Tbps silicon photonic IC for compute fabric connectivity. In 2022 *IEEE Symposium on High-Performance Interconnects*(*HOTI*) 1–4 (IEEE, 2022).
139. Seok, T. J., Kwon, K., Henriksson, J., Luo, J. & Wu, M. C. Wafer-scale silicon photonic switches beyond die size limit. *Optica* **6**, 490–494 (2019).
140. Anantharaman, R. Google Cloud demonstrates the world's largest distributed training job for large language models across 50000+ TPU v5e chips. Google Cloud Blog. https://cloud.google.com/blog/products/compute/the-worlds-largest-distributed-llm-training-job-on-tpu-v5e (2023).
141. Jouppi, N. et al. TPU v4: an optically reconfigurable supercomputer for machine learning with hardware support for embeddings. In *Proceedings of the 50th Annual International Symposium on Computer Architecture* 1–14 (ACM, 2023).
142. Chen, Y.-H., Krishna, T., Emer, J. S. & Sze, V. Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE J. Solid-State Circuits* **52**, 127–138 (2017).
143. Dennard, R. et al. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid-State Circuits* **9**, 256–268 (1974).
144. Waldrop, M. M. The chips are down for Moore's law. *Nature* **530**, 144–147 (2016).
145. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).
Key reference on how photonics can accelerate AI and enable neuromorphic architectures.
146. Ying, Z. et al. Electronic–photonic arithmetic logic unit for high-speed computing. *Nat. Commun.* **11**, 2154 (2020).
147. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
148. Pai, S. et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* **380**, 398–404 (2023).
149. Feng, C. et al. A compact butterfly-style silicon photonic–electronic neural chip for hardware-efficient deep learning. *ACS Photon.* **9**, 3906–3916 (2022).
150. Gu, J. et al. M^{3}ICRO: machine learning-enabled compact photonic tensor core based on programmable multi-operand multimode interference. *APL Mach. Learn.* **2**, 016106 (2024).
151. Zhu, H. H. et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044 (2022).
152. Wang, Z., Chang, L., Wang, F., Li, T. & Gu, T. Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nat. Commun.* **13**, 2131 (2022).
153. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
154. Gu, J. et al. SqueezeLight: a multi-operand ring-based optical neural network with cross-layer scalability. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **42**, 807–819 (2023).
155. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
156. Ward-Foxton, S. Optical chip solves hardest math problems faster than GPUs. EE Times. https://www.eetimes.com/optical-computing-chip-runs-hardest-math-problems-100x-faster-than-gpus/ (2021).
157. Ramey, C. Silicon photonics for artificial intelligence acceleration: Hotchips 32. In 2020 *IEEE Hot Chips32 Symposium* (*HCS*) 1–26 (IEEE, 2020).
158. Rogers, C. et al. A universal 3D imaging sensor on a silicon photonics platform. *Nature* **590**, 256–261 (2021).
159. Westerveld, W. J. et al. Sensitive, small, broadband and scalable optomechanical ultrasound sensor in silicon photonics. *Nat. Photon.* **15**, 341–345 (2021).
160. Vigliar, C. et al. Error-protected qubits in a silicon photonic chip. *Nat. Phys.* **17**, 1137–1143 (2021).
161. Hazan, Y., Levi, A., Nagli, M. & Rosenthal, A. Silicon-photonics acoustic detector for optoacoustic micro-tomography. *Nat. Commun.* **13**, 1488 (2022).
162. Pelucchi, E. et al. The potential and global outlook of integrated photonics for quantum technologies. *Nat. Rev. Phys.* **4**, 194–208 (2021).
163. Loizeau, F. Lightium secures $7 million seed funding to meet explosion in data center growth resulting from AI impact. Lightium Press Release. https://lightium.com/lightium-secures-7-million-seed-funding-to-meet-explosion-in-data-center-growth-resulting-from-ai-impact/ (2024).
164. Lin, Z. et al. High-performance polarization management devices based on thin-film lithium niobate. *Light Sci. Appl.* **11**, 93 (2022).
165. Yu, M. et al. Integrated electro-optic isolator on thin-film lithium niobate. *Nat. Photon.* **17**, 666–671 (2023).
166. Yu, M. et al. Integrated femtosecond pulse generator on thin-film lithium niobate. *Nature* **612**, 252–258 (2022).

Acknowledgements

The research of Y.W. and W.H. is supported by the King Abdullah University of Science and Technology (KAUST) under Award Nos RFS-TRG2024-6196, ORFS-CRG12-2024-6487, RFS-OFP2023-5558 and FCC/1/5939.

Author contributions

Y.W., W.H., J.J., L.L., D.Z.P. and H.R. conducted the literature search and prepared the first draft, with J.E.B. providing feedback and revision support.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Electrical Engineering* thanks Patrick Guo-Qiang Lo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.