

# Integrated photonic synapses, neurons, memristors, and neural networks for photonic neuromorphic computing

Shufei Han, Weihong Shen\*, Min Gu\* and Qiming Zhang\*

<https://doi.org/10.29026/oet.2025.250011>





## Integrated photonic synapses, neurons, memristors, and neural networks for photonic neuromorphic computing

Shufei Han, Weihong Shen\*, Min Gu\* and Qiming Zhang\*

**Citation:** Han SF, Shen WH, Gu M, Zhang QM. Integrated photonic synapses, neurons, memristors, and neural networks for photonic neuromorphic computing. *Opto-Electron Technol* **1**, 250011 (2025).

<https://doi.org/10.29026/oet.2025.250011>

Received: 19 November 2025; Accepted: 15 December 2025; Published online: 25 December 2025

### Related articles

All-optical digital logic and neuromorphic computing based on multi-wavelength auxiliary and competition in a single microring resonator

Qiang Zhang, Yingjun Fang, Ning Jiang et al

*Opto-Electronic Science* 2025, **4**(11): 250003 doi: [10.29026/oes.2025.250003](https://doi.org/10.29026/oes.2025.250003)

Ultrahigh performance passive radiative cooling by hybrid polar dielectric metasurface thermal emitters

Yinan Zhang, Yinggang Chen, Tong Wang et al

*Opto-Electronic Advances* 2024, **7**(4): 230194 doi: [10.29026/oea.2024.230194](https://doi.org/10.29026/oea.2024.230194)

Multi-dimensional multiplexing optical secret sharing framework with cascaded liquid crystal holograms

Keyao Li, Yiming Wang, Dapu Pi et al

*Opto-Electronic Advances* 2024, **7**(1): 230121 doi: [10.29026/oea.2024.230121](https://doi.org/10.29026/oea.2024.230121)

More related articles in Opto-Electronic Journals Group website 



# Integrated photonic synapses, neurons, memristors, and neural networks for photonic neuromorphic computing

Shufei Han<sup>1,2</sup>, Weihong Shen<sup>1,2\*</sup>, Min Gu<sup>1,2\*</sup> and Qiming Zhang<sup>1,2\*</sup>

**Abstract:** Rising demands for bandwidth, speed, and energy efficiency are reshaping the landscape of computing beyond the limits of von Neumann electronics. Neuromorphic photonics—using light to emulate neural computation—offers ultrafast, massively parallel, and low-energy information processing, positioning integrated photonic neural networks (IPNNs) as promising hardware for next-generation artificial intelligence (AI). By combining the architectural efficiency of neuromorphic models with the physical advantages of integrated photonics, IPNNs enable high-speed and programmable linear operations during the in-plane optical transmission, while leaving room for compact and reconfigurable on-chip optical nonlinearities and memory functions. Firstly, we review the concepts and principles of key building blocks in IPNN, that are photonic synapses, neurons, and photonic memristors which offer optical memory and storage capabilities. And then, we summarize the representative IPNN architectures and their recent advances, including coherent, parallel, diffractive, and reservoir computing, for photonic neuromorphic computing with high throughput and high efficiency. Finally, we outline practical considerations—calibration and stability of large-scale networks, routes toward co-integration with electronics, diffractive–interferometric hybrid architectures, and programmable photonic architectures for general AI purposes. We highlight a forward outlook on enabling IPNN with low energy consumption, robust photonic operations, and efficient training strategies, aiming to guide the maturation of general-purpose, low-power photonic AI.

**Keywords:** neuromorphic photonics; integrated photonic neural networks (IPNNs); photonic integrated circuits (PICs)

## Introduction

Integrated photonic neural networks (IPNNs), as an important direction of neuromorphic computing, are gradually becoming an ideal choice for efficiently processing large-scale artificial intelligence tasks. In recent years, with the rapid development of artificial intelligence (AI), the bottlenecks of traditional computing architectures in speed, latency, bandwidth, and power consumption have become increasingly prominent, especially for processing deep neural networks (DNNs). In the von Neumann architecture, the separation of computation and memory leads to inefficient data movement, which severely constrains the scalability and real-time performance of AI applications<sup>1–4</sup>. At present, global AI compute demand continues to grow rapidly, while traditional hardware architectures are approaching their limits<sup>5–7</sup>, this problem is particularly

evident in edge-computing scenarios, where reliance on power-hungry graphics processing units (GPUs) and bandwidth-heavy workloads is no longer sustainable. Meanwhile, the computational paradigm of biological neural systems—characterized by outstanding energy efficiency, parallelism, and adaptive learning—offers a new direction to address these challenges. Neuromorphic computing, which emulates the structure and function of biological neural systems, exhibits clear advantages in energy efficiency, computational speed, and dynamic learning capability; however, existing electronic hardware still suffers from low integration density, high power consumption, and electrical interference, limiting further development<sup>8–10</sup>.

Photonics, by virtue of its high bandwidth, low latency, and low energy consumption, has become an ideal platform for neuromorphic computing<sup>11–15</sup>. Photonic computing leverages optical signals for parallel processing,

Received: 19 November 2025

Accepted: 15 December 2025

Published online: 25 December 2025

<sup>1</sup>School of Artificial Intelligence Science and Technology, University of Shanghai for Science and Technology, Shanghai 200093, China; <sup>2</sup>Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai 200093, China.

\*Correspondence: WH Shen, E-mail: [shenweihong@usst.edu.cn](mailto:shenweihong@usst.edu.cn); M Gu, E-mail: [gumin@usst.edu.cn](mailto:gumin@usst.edu.cn); QM Zhang, E-mail: [qimingzhang@usst.edu.cn](mailto:qimingzhang@usst.edu.cn)

handling multiple information dimensions and achieving near-lossless data transmission and in-flight accumulation during propagation. Studies show that, under the same power budget, photonic processors deliver far greater computational throughput than traditional electronic processors<sup>16,17</sup>. At the same time, photonic technology supports multi-threading across dimensions such as wavelength and spatial mode<sup>18–21</sup>, enabling significant acceleration of computation without adding proportional overhead.

Despite theoretical and experimental advances, integrated photonic neural networks still face several challenges<sup>11,22,23</sup>. For example, implementing low-power nonlinear activation in hardware and realizing efficient photonic storage media remain focal and difficult research topics<sup>24–27</sup>. In addition, optimizing the architectural design of photonic neural networks to achieve higher computational density and stronger processing capability is an urgent problem to solve<sup>28–31</sup>. As integrated photonic technology matures, future photonic neural networks are expected to realize large-scale, low-power intelligent inference and real-time processing through new architectural designs and more efficient photonic components, thereby promoting the broad application of AI in edge computing, autonomous driving, intelligent manufacturing, and other fields.

In this work, we first introduce photonic synapses and photonic neurons and discuss how they implement weight loading and nonlinear activation functions, which are two critical links in neuromorphic computing. Photonic synapses realize weighted operations by modulating the intensity and phase of light, while photonic neurons process signals through nonlinear activation functions. Optical memristors, as key components in photonic neural networks, provide nonvolatile memory and high-speed processing capabilities, supporting the weight loading and storage functions of photonic synapses, thereby providing strong support for the efficient operation of photonic neural networks. The efficient cooperation of these photonic building blocks forms the foundation of photonic neural networks and propels the rapid development of photonic neuromorphic systems. We then delve into recent progress in integrated photonic neuromorphic systems with different architectures, analyzing their application potential and challenges across various AI tasks. Through this review, we aim to provide theoretical support and practical guidance for the development of integrated photonic neural networks, paving the way for next-generation, efficient, low-power AI computing platforms.

## Integrated photonic neuromorphic devices

### Photonic synapses

Biological neuromorphic systems consist of collections of highly interconnected neurons, computational units and synapses (memory units). Each neuron receives weighted outputs from adjacent neurons via dendrites with specific

synaptic weights, and these signals are summed in the soma, equivalent to a linear combination of all inputs—a process that can be modeled as matrix-vector multiplication (MVM)<sup>32</sup>. In integrated photonic neuromorphic systems, synaptic interconnections are typically implemented by optical waveguides, and the weighting process is usually performed by optical intensity modulators. In other words, the weight is loaded onto the optical carrier's intensity via all-optical or electro-optic modulation<sup>33–35</sup>.

Integrated intensity modulators can be realized in two ways. One directly modulates intensity along the waveguide, and the other converts phase modulation into intensity via resonators or interferometers. Waveguide-based photonic synapses typically use optical attenuators with tunable absorption loss (e.g., devices based on P-type / Intrinsic / N-type (PIN) junctions or phase-change materials<sup>33</sup>) or optical amplifiers with tunable gain (e.g., semiconductor optical amplifiers, SOAs<sup>36</sup>). Such photonic synapses can perform dot products as light propagates and require synaptic arrays to perform MVM operations.

For the other class of photonic synapses, the most common structures are microring resonators (MRRs)<sup>37</sup> and Mach-Zehnder interferometers (MZIs)<sup>28</sup>, as shown in Fig. 1(a) and Fig. 1(b). Photonic weight matrices based on MZI meshes rely on coherent propagation and interference. This architecture typically has a complex network topology<sup>38,39</sup>, and therefore requires precise matrix-decomposition algorithms to configure the phase parameters of each MZI element; backpropagation-based training likewise depends on these configurations. The number of MZIs required by an  $N$ -port neural network grows quadratically with  $N$ , which poses challenges for compactness and power at large scales.

MRRs are significantly smaller than MZIs, with the diameter of MRRs typically in the range of tens of micrometers, whereas the overall dimensions of MZIs are generally hundreds of micrometers scale. The resonant characteristics of MRRs can be combined with wavelength-division multiplexing (WDM) to realize parallel and direct weight loading—often referred to as the "broadcast-and-weight" scheme<sup>40</sup>. Weights are applied to the optical intensity of WDM signals by MRRs operating at different resonant wavelengths. Because optical intensity cannot be negative, balanced photodetection of the through and drop ports of split MRRs is employed to realize real-valued (signed) weights. However, due to inevitable fabrication variations and thermal sensitivity, the calibration and stabilization of large-scale microring arrays remain challenging.

As an alternative, phase-change materials (PCMs) offer broad prospects for nonvolatile photonic synapses with memory effects and can also be used for photonic in-memory computing. Chalcogenide PCMs can be switched nonvolatily and repeatedly between crystalline and amorphous states under optical or electrical pulses, producing significant changes in both the real and imaginary parts of



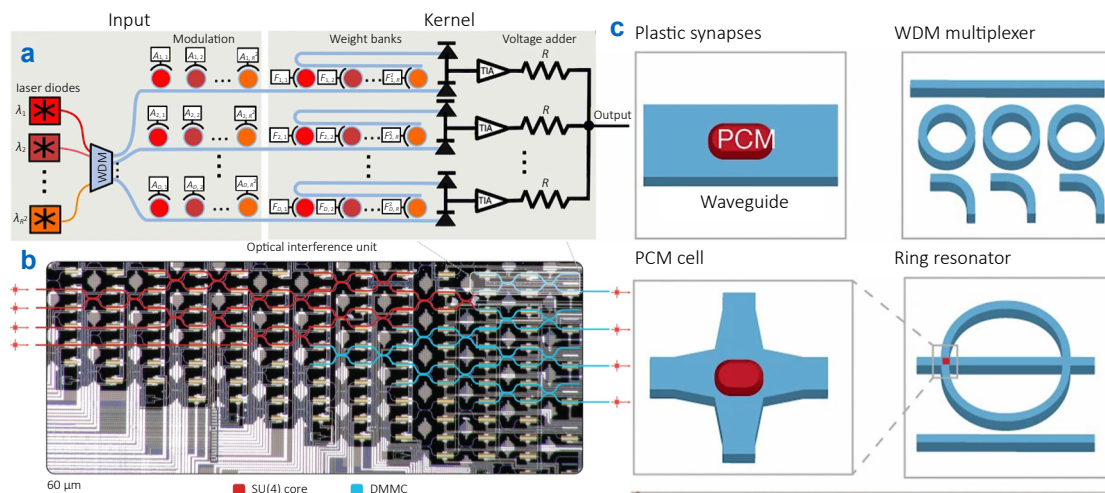


Fig. 1 | Photonic integrated synapses based on (a) MRR, (b) MZI, and (c) PCM. Figure reproduced with permission from: (a) ref.<sup>37</sup>, IEEE; (b) ref.<sup>28</sup>, (c) ref.<sup>41</sup>, Springer Nature.

the refractive index. In PCM-integrated waveguide devices<sup>41</sup>, as shown in Fig. 1(c), the phase-change process induces a change in the refractive index, which modulates both the intensity and phase of optical signals, thereby enabling synaptic weighting. Their nonvolatile nature makes them excellent candidates for optical memory. By finely controlling intermediate states between crystalline and amorphous phases, multi-bit storage can be realized. Since the first fully optical PCM synapse was proposed in 2017<sup>34</sup>, many reports on PCM-based photonic synapses have appeared, enabling in-memory photonic computing, unsupervised spiking neuromorphic networks<sup>41</sup>, and tensor-core-like convolution operations<sup>30</sup>.

### Photonic neurons

In integrated photonic neuromorphic systems, photonic neurons are mainly divided into two types: so-called continuous-time neurons, which communicate using optical continuous-wave analog signals, and spiking neurons, which encode information in the timing of spikes. Most traditional photonic artificial neural networks adopt the continuous-time neuron structure. As long as the nonlinear activation function is differentiable and integrable, gradient descent can be used for effective training. By contrast, spiking neurons face greater difficulty and complexity in signal generation, detection, and network training. Even so, spiking neurons provide a more natural way to represent certain classes of information and are more energy-efficient than non-spiking neurons.

Photonic implementations of artificial neurons include two solutions, namely opto-electro-optic (OEO) conversion and fully optical solutions. For OEO continuous-time neurons, nonlinear responses arise from self-intensity modulation or from the intrinsic electro-optic transfer functions of modulators or photodetectors. In the first case,

as shown in Fig. 2(a) and Fig. 2(b), the optical signal after linear computation is proportionality-converted into an electrical signal by an on-chip photodetector and then fed to an electro-optic modulator (typically an MZI<sup>42</sup> or MRR<sup>43</sup>) to modulate the intensity of a new optical carrier or of the original optical signal. The nonlinear function of an OEO neuron can be programmed via the modulator's voltage bias. Using combined MZI and MRR architectures, as many as 14 nonlinear functions have been reported, showing strong diversity and reconfigurability<sup>43</sup>. However, the time delay introduced by OEO conversions is inevitable and requires corresponding on-chip optical delay lines to synchronize optical and electrical signals.

In addition, the intrinsic electro-optic/optical-electrical transfer function of EO/OE devices can be directly used as the nonlinear activation, as shown in Fig. 2(c). Electro-absorption modulators (EAMs) based on indium tin oxide (ITO) layers monolithically integrated into silicon waveguides have been shown to act as nonlinear photonic neurons<sup>44</sup>. Nonlinear germanium-silicon photodiodes (Ge-Si PDs) exploiting free-carrier absorption have been proposed to realize on-chip optical neuron activation with self-monitoring of optical power<sup>45</sup>, providing a fully integrated all-optical neural network capable of in-situ training. Thanks to the maturity of high-speed EO/OE devices on silicon photonic platforms, OEO neurons offer high feasibility for complementary metal-oxide-semiconductor (CMOS)-compatible, large-scale neuromorphic systems.

OEO schemes for spiking neurons mainly use excitable semiconductor lasers, such as two-section gain/saturable-absorber lasers<sup>46</sup> or electrically injected excitable distributed-feedback lasers<sup>47</sup>. In 2023, Xiang SY et al. proposed and experimentally validated a photonic spiking-neuron chip based on an integrated Fabry-Perot laser with a saturable absorber, exploiting the device's excitable dynamics to realize neuronal functions such as spike generation

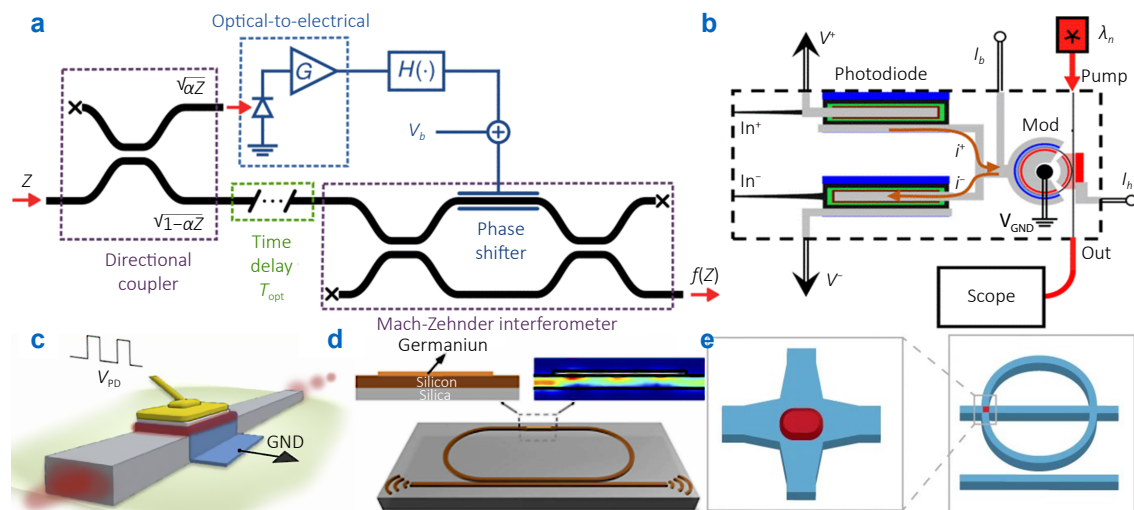


Fig. 2 | Photonic integrated neurons based on OEO conversion based on (a) MZI or (b) MRR, OE neurons based on (c) ITO-EAM, and all-optical neurons based on (d) Ge- and (e) PCM-loaded MRR. Figure reproduced with permission from: (a) ref.<sup>42</sup>, IEEE; (b) ref.<sup>43</sup>, APS; (c) ref.<sup>44</sup>, AIP Publishing; (d) ref.<sup>50</sup>, under a Creative Commons Attribution License; (e) ref.<sup>41</sup>, Springer Nature.

and temporal integration<sup>48</sup>. By further incorporating time-multiplexed spatiotemporal spike encoding and a supervised learning strategy, they demonstrated hardware-algorithm co-inference in cascable photonic spiking neural networks. In the same year, the group extended this approach by developing an indium-phosphide photonic integrated distributed-feedback-laser-saturable-absorber neuro-synaptic chip, monolithically merging synaptic linear weighting and spiking-neuron nonlinear firing within a single device<sup>49</sup>. Using this integrated core, they achieved on-chip hardware inference of convolutional spiking neural networks, thereby substantially advancing the integration and system-level scalability of photonic spiking computing engines. Fully optical schemes rely on nonlinear materials (such as Ge<sup>50</sup> or PCMs<sup>41</sup>) and on the nonlinear dynamics of MRRs or photonic-crystal nanocavities, as shown in Fig. 2(d) and Fig. 2(e). In 2019, Feldmann et al. proposed a fully optical spiking neuron based on a passive MRR structure with PCM-loaded crossings<sup>41</sup>, in which the optical signal after the PCM synapse is applied to the MRR, and the neuron is activated through the phase-change process. In 2022, a passive silicon MRR demonstrated another fully optical integrated spiking neuron via self-pulsation behavior<sup>50</sup>, with observed sub-GHz operating speeds and excitability akin to resonate-and-fire neurons. These two studies reveal new ideas for realizing fully optical integrated spiking neurons and show compatibility with WDM<sup>30</sup>, mode-division multiplexing<sup>51</sup>, and the execution of fully optical plasticity and self-learning.

### Photonic memristors

Photonic memristors, as the optical memory components of IPNN, have attracted considerable attention in recent years. Originating from Chua's proposal of the electrical memris-

tor as the fourth fundamental circuit element in 1971, their defining property is the ability to modulate and retain a resistance state conditioned on the history of the input signal<sup>52</sup>. As the optical counterparts in photonic neuromorphic systems, photonic memristors offer optical memory and processing by tuning the transmission properties of light in non-volatile or volatile forms, thereby providing innovative solutions for photonic in-memory neuromorphic computing.

The photonic non-volatile memristors modulate the optical response in integrated circuits while maintaining the modulated state afterwards achieving high-speed operation together with long retention<sup>53</sup>. The nonvolatile, multilevel storage afforded by optical memristors has been deployed in photonic integrated circuits (PICs) with functional materials such as PCMs and zirconate titanate (PZT)<sup>54</sup>, thereby co-locating memory and computation within the same device and substantially improving the efficiency of IPNN. Nevertheless, widespread adoption remains hindered by challenges including material failure mechanisms, fabrication complexity, and excessive switching energy, all of which warrant further investigation.

Compared to the static and long-term memory of non-volatile memristors, photonic volatile memristors offer short-term memory and rich dynamical characteristics with nonlinearities, leading to adaptive neuromorphic computing with high complexity<sup>55</sup>. Nonlinear optical cavities with nonlinear dynamics like optical bi-stability, self-oscillation and chaos, are the optimal candidates of photonic volatile memristors<sup>32</sup>, and ease to be implemented in integrated photonics, fiber optics or free-space optics. Concurrently, the concept of photonic volatile memristors has been explored in the field of quantum photonics. In 2022, Spagnolo M et al. experimentally demonstrated a photonic quantum memristor capable of processing quantum



information at the single-photon level<sup>56</sup>, underscoring the potential of quantum neuromorphic architectures, as shown in Fig. 3. This innovation not only broadens the application scope of memristive concepts but also suggests a pathway for bridging photonic and quantum computing. Future work will focus on enhancing switching speed, energy efficiency, and scalability to meet the demands of large-scale photonic neuromorphic systems and quantum computing to solve general AI proposes and high-complexity problems like nondeterministic polynomial time (NP)-hard and combinatorial optimization<sup>57</sup>.

Photonic synapses, photonic neurons, and photonic memristors are key components in photonic neuromorphic computing systems, each possessing distinct characteristics, advantages, and limitations. Photonic synapses implement synaptic interconnections via optical waveguides and weight signals using optical intensity modulators, offering high bandwidth, low energy consumption, and parallel computing capabilities. They can efficiently perform matrix-vector multiplication, making them suitable for large-scale neural network applications, and support non-volatile storage. However, challenges remain in terms of system integration, stability, and power consumption. Photonic neurons, including continuous-time and spiking

neurons, exhibit advantages in high-speed computation and low energy consumption. Continuous-time neurons are compatible with CMOS technology and are suitable for large-scale integration, but their implementation is relatively complex. Spiking neurons, on the other hand, encode information through spike signals, offering superior energy efficiency, but their training and implementation are more intricate. Photonic memristors modulate light intensity and phase to enable volatile and non-volatile in-memory computing, allowing in-memory computing tasks to be processed within the same device, thus enhancing efficiency. However, photonic memristors are still in the early stages of development, with significant potential for exploration and discovery. Overall, these components present promising prospects for photonic neuromorphic computing, yet the realization and stability of large-scale systems still face significant technical challenges.

## Integrated photonic neural networks

### Coherent photonic neural networks

Coherent photonic neural networks (PNNs) use a single light source to provide a coherent optical field for the entire network and realize signal distribution, routing, and

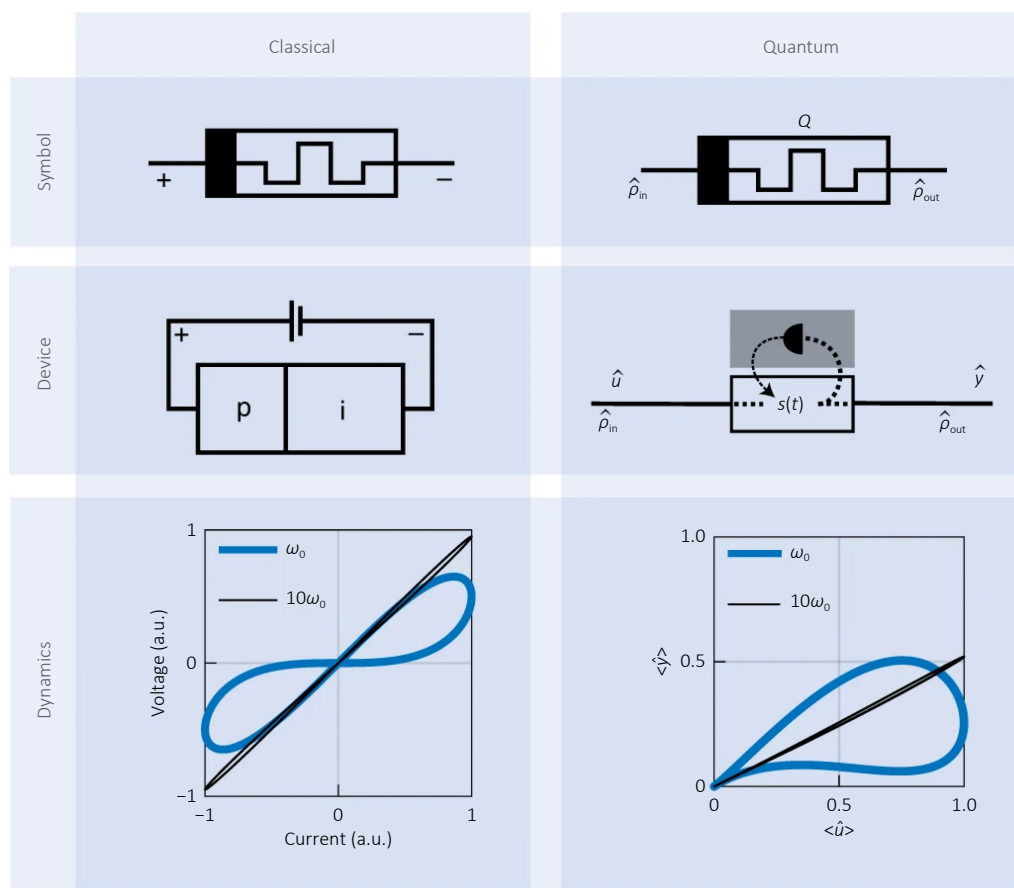


Fig. 3 | Comparison of classical and quantum memristors. Figure reproduced with permission from ref.<sup>56</sup>, under a Creative Commons Attribution License.

weighting through optical interference. Their linear MVM operation is performed entirely in the optical domain at a single carrier frequency, without multi-wavelength signals or O/E conversions, thereby making the system more neuromorphic and enabling planarization and cascade across layers. In 2017, Shen YC et al. proposed a coherent nanophotonic deep-learning processor on a silicon PIC<sup>28</sup>. The processor comprised a cascaded array of 56 programmable MZIs and demonstrated its application to vowel recognition for the first time. Although on-chip optical nonlinear units and fully integrated multiply-accumulate (MAC) operations had not yet been realized at that time, this work marked the starting point of neuromorphic computing on large-scale PIC platforms. Building on this, coherent PNNs have made further breakthroughs in recent years. In 2023, Pai S et al. proposed a new method addressing the training difficulties of coherent neural networks<sup>58</sup>. In 2024, Bandyopadhyay S et al. designed and implemented a fully integrated coherent optical neural network architecture (FICONN)<sup>59</sup>. Coherent PNNs can be applied not only to typical tasks such as image and digit recognition but also to communication channel optimization and reinforcement learning. By 2025, coherent PNNs were capable of realizing truly large-scale integrated photonic devices, validating their feasibility and superiority for practical AI and scientific computing.

In 2023, Pai S et al. used thermo-optic phase shifters as tunable weight elements<sup>58</sup>, as shown in Fig. 4(a). By introducing forward and backward optical-field interference measurements, gradient information can be obtained directly in the optical domain, avoiding frequent OEO conversions and external numerical computation. This enabled the first in-situ, end-to-end training of PNNs and greatly improved the flexibility and usability of optical computing systems. In 2024, Bandyopadhyay S et al. monolithically integrated four key subsystems onto a commercial silicon-photonics chip (6.0 mm × 5.7 mm)<sup>59</sup>. Input vectors were encoded as 6-channel coherent optical fields (amplitude and phase) and passed through an MZI mesh to implement a programmable 6 × 6 matrix multiplication, completing the linear transforms of each layer. The system achieved, for the first time on a single chip, a multilayer coherent all-optical DNN with programmable nonlinearities for inference and in-situ training, demonstrating that efficient on-hardware training is possible without conventional backpropagation and greatly reducing training latency and energy. The power consumption is still limited by thermal phase shifters; lower-power phase-shift technologies are expected to further improve efficiency. In 2025, Liao K et al. integrated near-infrared perovskite LEDs<sup>60</sup>, a Si<sub>3</sub>N<sub>4</sub> photonic processing network (13 layers, locally connected), and perovskite photodetectors on the same chip to construct a heterogeneous "emit-compute-detect" integrated system. Perovskite materials, with low-temperature processing, tunable bandgaps, and tolerance to lattice

mismatch, are suitable for heterogeneous integration with Si<sub>3</sub>N<sub>4</sub> platforms, thereby addressing long-standing cross-material co-integration bottlenecks among sources, processors, and detectors.

Based on the basic MZI structure, in 2023, SeyedMohammad S et al. proposed a programmable MZI mesh integrated with arrays of surface grating couplers<sup>61</sup>. Even when the numbers of apertures at the transmitter and receiver are mismatched or irregularly arranged, or when amplitude occlusion or phase distortions are introduced along the propagation path, the system can still find optimal orthogonal modes. This study finds the optimal communication channel in real time without system modeling or measurement, opening a new frontier for PNN applications. In 2024, Li XK et al. built a hybrid photonic computing platform to directly emulate reinforcement learning (RL) processes<sup>62</sup>, as shown in Fig. 4(b). The system computes the similarity of 3472-dimensional atomic vectors for perovskite synthesis tasks and accelerates Q-learning by using similarity-based rewards, achieving efficiency improvements of 30.6%, 12.2%, and 56.0% in benchmark tasks and real materials design, thereby demonstrating the feasibility and advantages of optoelectronic co-acceleration for large-scale RL.

In 2025, Hua SY et al. proposed a truly large-scale, low-power, high-accuracy, and integrable photonic matrix accelerator<sup>31</sup>, as shown in Fig. 4(c). The system comprises three core modules: an input modulation unit (modulating vector elements into optical signals), a matrix core (a reconfigurable MZI mesh for multiply-accumulate operations), and an optoelectronic detection unit (implementing weighted optical accumulation and electrical readout). The accelerator adopts a non-blocking reconfigurable interferometric structure supporting arrays of 64 × 64 and beyond, improving interconnect flexibility and area efficiency, and innovatively introduces an optical-register mechanism to avoid frequent OEO conversions. With reconfigurability, scalability, and ease of integration, it represents an important advance in optical computing and provides strong hardware support for future on-chip optoelectronic hybrid AI systems.

As application scenarios grow in complexity, hardware architectures featuring large scale, high parallelism, and high integration become increasingly critical for PNNs. In 2024, Dong BW et al. proposed using partially coherent light sources to replace traditional fully coherent sources to reduce crosstalk and boost parallel compute capability<sup>63</sup>. By controlling the coherence length, crosstalk can be reduced by nearly two orders of magnitude. Comparative experiments using fully coherent (laser), partially coherent (amplified spontaneous emission (ASE)/superluminescent diode (SLD)), and incoherent (LED) sources show that partially coherent sources maintain high accuracy even when executing up to nine groups of parallel matrix



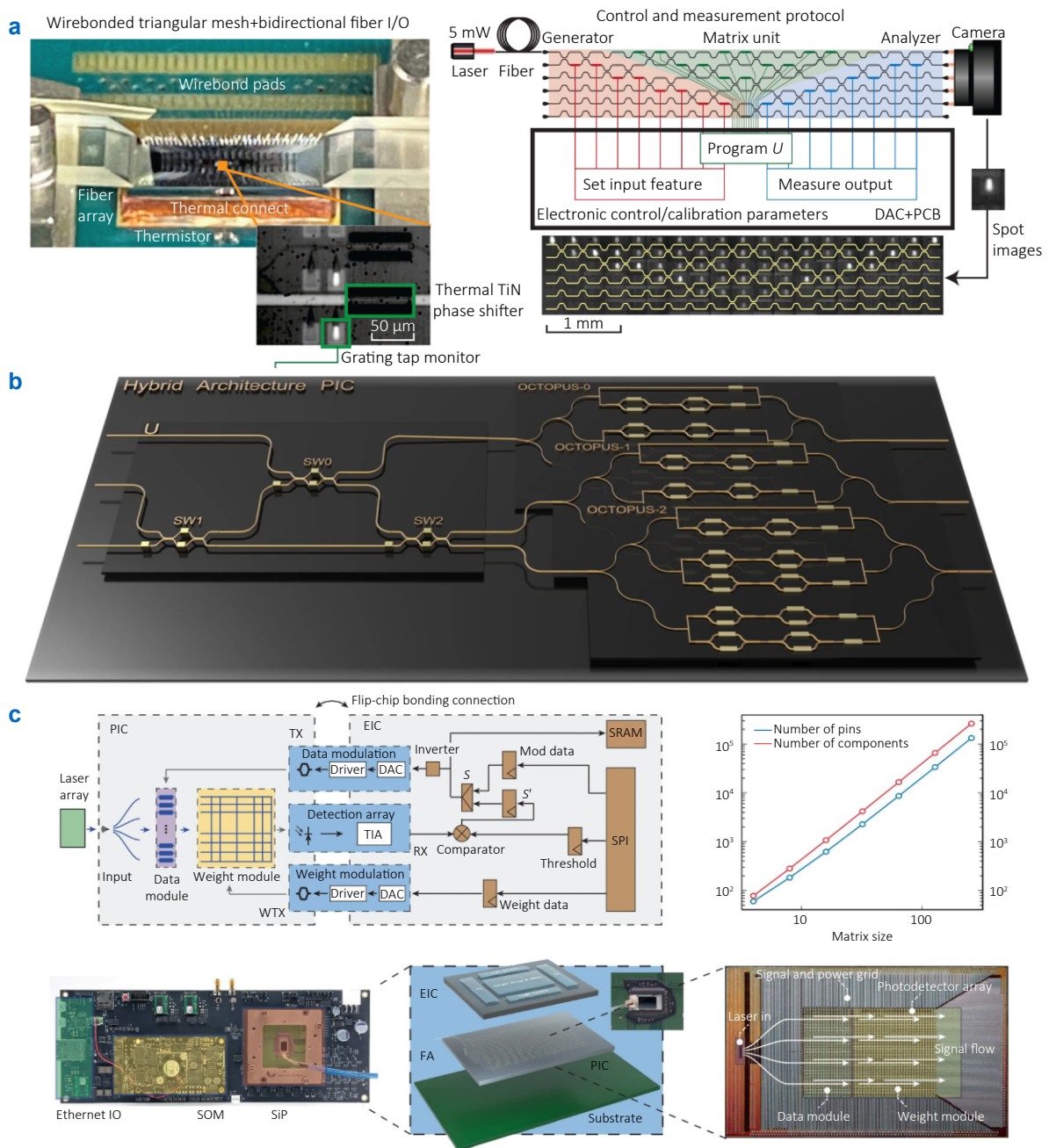


Fig. 4 | Coherent photonic neural network based on (a) thermo-optic phase shifters, (b) emulate reinforcement learning, (c) large-scale. Figure reproduced with permission from: (a) ref.<sup>58</sup>, (b) ref.<sup>62</sup>, (c) ref.<sup>31</sup>, under a Creative Commons Attribution License.

multiplications. This study systematically verifies, for the first time, that "partial coherence" is a key resource for enhancing the parallelism and anti-crosstalk performance of photonic computing, breaking the traditional design paradigm that relied solely on fully coherent sources and offering a new approach for high-precision, high-throughput, and scalable optical computing platforms.

### Parallelized IPNN

MRRs possess characteristic filtering behavior, enabling

selective transmission or suppression of specific optical frequencies and reallocation of optical power at selected resonances. Based on this, WDM systems composed of MRRs can realize weight distribution and summation, and through wavelength-channel filtering and weighting, perform the matrix operations required for inference in PNNs directly in the optical domain. As a key technology in optical communications to expand capacity, WDM likewise accelerates photonic neural computing via wavelength-parallel operations. The photonic "Broadcast-and-Weight"

scheme is naturally compatible with WDM<sup>37</sup>, input data and weight matrices are loaded onto different wavelength channels via a WDM modulation array and MRR weight bank, thereby enabling highly parallel computation. If combined with PCMs, the programmable absorption contrast between crystalline and amorphous states can be exploited for weight tuning, endowing the network with nonvolatility and self-learning ability. Compared with coherent MZI-based architectures, MRR networks perform matrix computation via intensity modulation with a simpler structure, avoiding complex phase control and programming algorithms and offering higher compactness and energy efficiency in large-scale networks.

In 2023, Yin RY et al. proposed an on-chip parallel computing scheme combining mode-division multiplexing (MDM) with WDM<sup>64</sup>. As shown in Fig. 5(a), three TE modes (TE<sub>0</sub>–TE<sub>2</sub>) and four wavelengths share the same bus waveguide, and a  $3 \times 4$  microring array performs mode demultiplexing, elementwise multiplication, and weighted accumulation. MRRs write weights by thermal tuning and achieve continuously tunable 0–1 weight via on/off-resonance control. By stacking orthogonal MDM and WDM degrees of freedom and combining negative-weight mapping with MRR weight programming, the system achieved high parallelism, low energy consumption, and easy on-chip integration for convolution and mode-detection tasks. In 2025, Hong SH et al. further proposed and demonstrated a scalable parallel silicon-photonics signal-processing chip<sup>65</sup>, as shown in Fig. 5(b). The chip integrates two WDM-based 16-channel (de)multiplexers and 6-bit tunable delay lines (TDLs), realizing wavelength-time interleaved multiplexing. On the same hardware, it supports multiple functions including microwave signal reception and beamforming, reconfigurable microwave photonic filtering, ultra-wideband arbitrary waveform generation, and high-speed parallel optical computing, demonstrating excellent generality and scalability.

In recent years, research on parallel PNNs has advanced not only in structural design but also in application scope. Beyond common digit, speech, and gesture recognition, researchers have explored novel tasks such as emotion recognition. In 2023, Zhao BH et al. built an on-chip integrated optical neural network using a microcomb source providing dozens of evenly spaced wavelengths together with a thermally tuned MRR array<sup>66</sup>, as shown in Fig. 5(c). The system realized high-accuracy optical tensor convolution and 6-class human emotion recognition. While this scheme demonstrated the feasibility of optical convolution for multi-class emotion recognition, its components were still implemented discretely, which is unfavorable for further on-chip integration and miniaturization. In the same year, Bai et al. for the first time realized a fully integrated system combining a DFB-pumped microcomb<sup>67</sup>, electro-optic modulators (EOMs), an MRR weight array, and on-chip delay lines, as shown in Fig. 5(d). By combin-

ing multi-wavelength parallelism from the microcomb with programmable weight/delay control, this approach achieved high-accuracy, calibratable WDM convolution on chip, with recognition accuracy approaching digital implementations and setting a new record in optical core compute density. Also in 2023, Jiang XY et al. proposed a triangular waveguide-mesh structure<sup>68</sup>. With thermo-tuned ring nodes, the structure solves subset-sum problems (SSPs) by setting rings to splitting or through states and emulates two-state discrete-time stochastic processes (SPSs) by controlling transmission. The triangular mesh is more compact, facilitating scale-out and programmable photonic computing. In 2025, Yu X et al. proposed a triangular-like parallel optical computing system driven by a soliton microcomb source and an MZI network<sup>69</sup>, as shown in Fig. 5(e). The scheme supports over 100 wavelength inputs, greatly improving optical computing parallelism and data throughput.

Previously, in 2019, Feldmann J et al. proposed an all-optical pulsed synaptic network architecture based on an MRR-WDM system combined with PCMs<sup>41</sup>, as shown in Fig. 6(a). The system uses PCMs to control synaptic weights and achieves optical nonlinear responses through reversible amorphous-crystalline switching, enabling both supervised and unsupervised learning tasks. Compared with traditional electronic approaches, optical neural networks significantly improve computational speed and energy efficiency and are particularly suitable for high-speed, large-bandwidth data processing scenarios. Subsequently, in 2021, the Feldmann team further built an integrated photonic tensor core on a Si<sub>3</sub>N<sub>4</sub> platform<sup>30</sup>, as shown in Fig. 6(b). By combining MRRs, WDM, and PCMs, the device performs parallel optical convolutions on chip, achieving up to 10<sup>12</sup> multiply-accumulate (MAC) operations per second. Leveraging microcombs together with PCMs, multiple MVM operations can be executed within a single physical device, greatly improving computational efficiency. In 2023, Dong BW et al. proposed a photonic radio frequency (RF) modulation scheme based on PCMs and WDM<sup>70</sup>, as shown in Fig. 6(c), introducing an additional degree of freedom to increase parallelism. The system achieved approximately 100× more parallel computing capacity, an improvement of two orders of magnitude over systems relying solely on spatial and wavelength degrees of freedom. In the same period, Zhou W et al. developed nonvolatile, electrically programmable multi-level PCM waveguide memory cells<sup>71</sup>, as shown in Fig. 6(d). Using WDM for multi-channel parallel multiplication, they constructed a  $2 \times 2$  optical convolution kernel composed of four PCM cells and demonstrated on-chip handwritten-digit and image recognition.

In summary, parallel PNNs significantly increase on-chip neural-compute throughput and energy efficiency by multiplexing across multiple degrees of freedom such as wavelength, mode, and time. MRR-based WDM architectures, thanks to their spectral selectivity and structural tunability, can realize high-precision weight loading and matrix

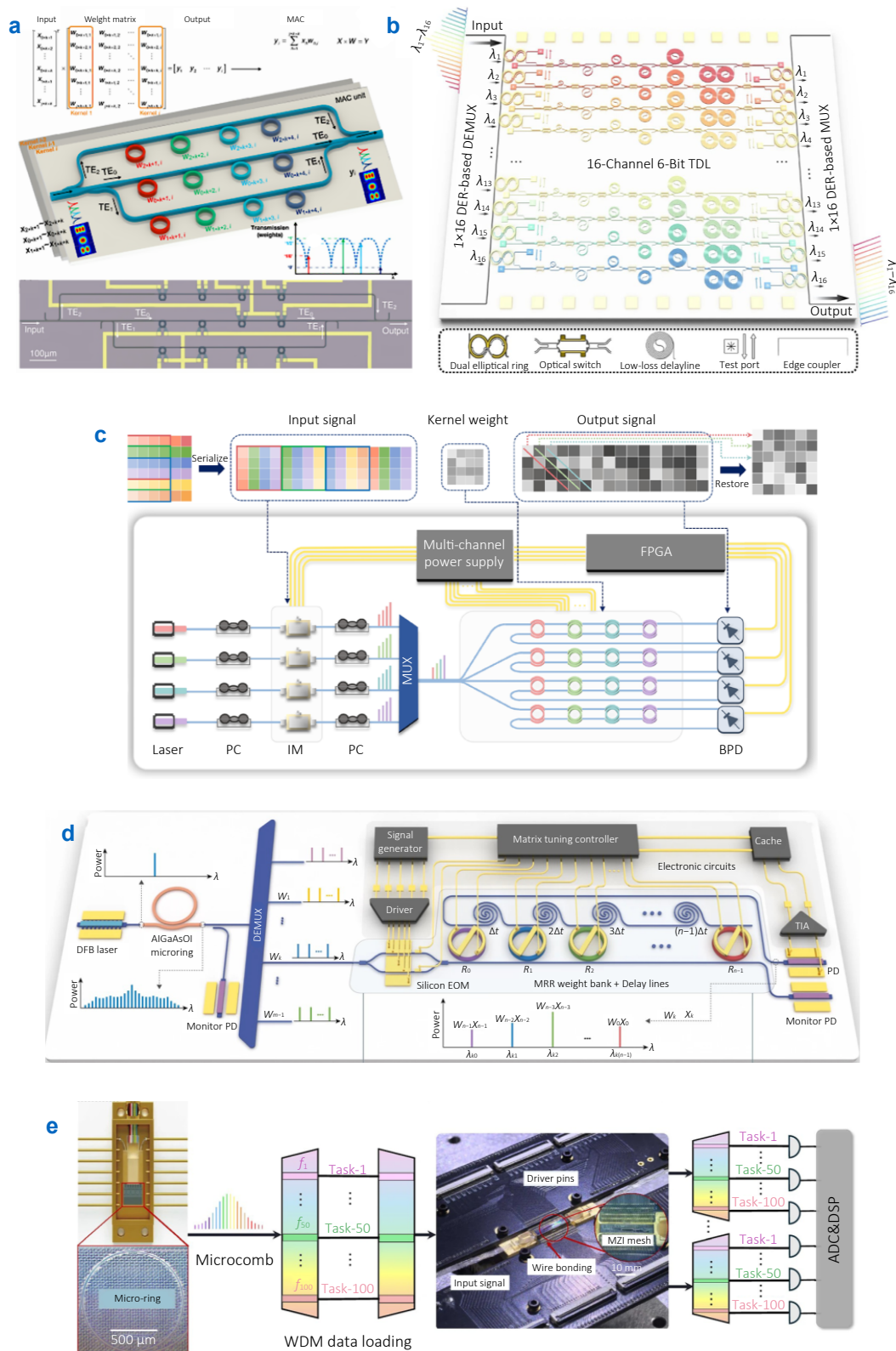


Fig. 5 | parallelized photonic neural network based on (a) MDM and WDM, (b) TDLs, (c) human emotion recognition, (d) fully integrated system, and (e) triangular-like system driven by a soliton microcomb source and an MZI network. Figure reproduced with permission from: (a) ref.<sup>64</sup>, (b) ref.<sup>65</sup>, (c) ref.<sup>66</sup>, (d) ref.<sup>67</sup>, (e) ref.<sup>69</sup>, under a Creative Commons Attribution License.



computation; with the introduction of PCMs, they further gain reconfigurability and nonvolatility, enabling on-chip self-learning. With the fusion of technologies such as MDM, TDLs, and microcomb sources, parallel PNNs continue to make progress in compactness, energy optimization, and multi-task adaptability, achieving accuracy close to electronic neural networks with much higher speed in tasks such as speech recognition, emotion recognition, image classification, and combinatorial optimization. Although most current systems still depend on discrete components, overall integration and interconnect efficiency are steadily improving, and the trend toward fully on-chip integration and cross-domain co-design is clear. Continued development in this direction will provide a solid physical and engineering foundation for the realization of genuinely practical photonic intelligent computing.

### Integrated diffractive optical neural networks

Diffractive optical neural networks (DONNs) are generally composed of discrete diffractive optical elements containing thousands to millions of photonic neurons and interconnections, featuring extremely strong parallel computing capability and ultra-low energy consumption. Three-dimensional diffractive neural networks (3D DONNs) stack multiple diffractive layers to realize deeper and wider network structures, thereby enhancing model expressivity and computational complexity. In 2018, Lin X et al. first proposed an optical-diffraction-based deep neural network that classified handwritten-digit images<sup>72</sup>, demonstrating the potential of diffractive optical computing for image recognition. However, 3D structures involve higher design and fabrication complexity and usually entail higher energy and volume costs. In recent years, researchers have begun exploring DONN migration to PIC platforms to realize more compact and controllable on-chip integrated architec-

tures. Integrated DONNs combine the advantages of optical diffraction and deep learning and offer a viable route to efficient, low-power photonic computing.

In the integrated diffractive systems, phase control of the optical wave is usually implemented by dielectric metasurfaces formed by arrays of subwavelength structures. Such metasurfaces impose local phase shifts on fields propagating in a wide slab waveguide. In 2020, Zarei S et al. designed an integrated DONN based on a high-contrast transmission metasurface<sup>73</sup>. The system used a subwavelength diffractive array with a 500 nm period and 140 nm cell width to realize one-dimensional diffraction control at a 1550 nm wavelength. In the study, the 1D metasurface was defined as a hidden layer of the DONN, and each diffractive cell was treated as a neuron. By changing the lengths of diffractive cells under fixed width and period, precise wavefront control and spatial modulation of the optical field were achieved.

In 2021, the group of Chen HW at Tsinghua University proposed an integrated diffractive network based on a 2D spatial electromagnetic propagation model on chip<sup>74</sup>, as shown in Fig. 7(a). This model accurately approximates light propagation in slab waveguides and serves as the forward model during DONN parameter training, allowing fast optimization of structural parameters via gradient descent. The team also proposed a "weight-mapping model", defining three same-size subwavelength diffractive structures as a basic diffractive unit and treating it as the minimum computational unit of the DONN. In 2023, building on earlier theory, the group designed and fabricated the first integrated DONN<sup>75</sup>. By combining particle-swarm optimization with experimental measurements, they implemented online error compensation after chip packaging, significantly improving system robustness and computational accuracy. In the same year, based on subwavelength diffractive units, they designed an on-chip

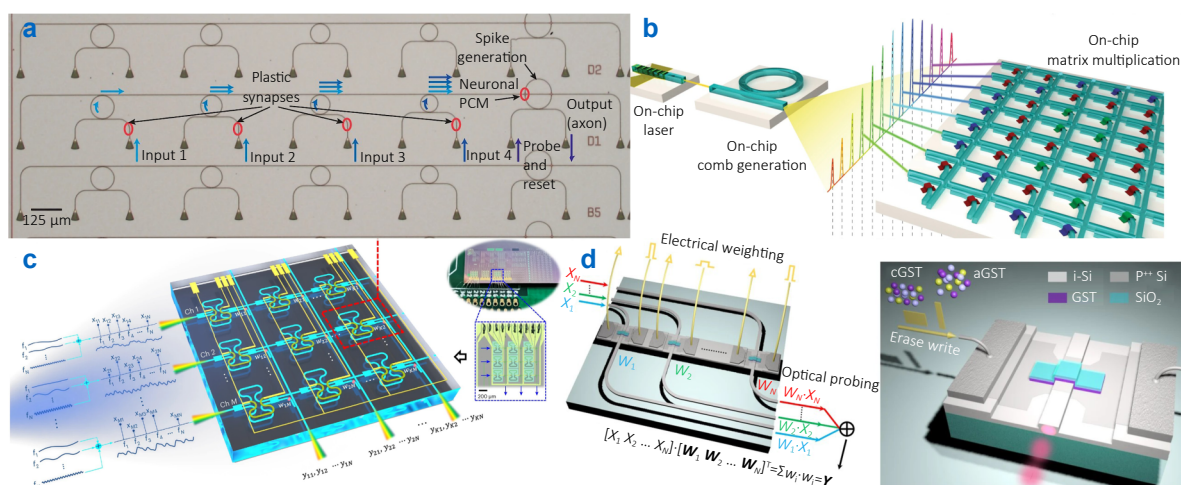


Fig. 6 | Parallelized photonic neural network based on PCM with (a) WDM, (b) Si<sub>3</sub>N<sub>4</sub>, (c) photonic RF modulation scheme, (d) multi-channel parallel multiplication. Figure reproduced with permission from: (a) ref.<sup>41</sup>, (b) ref.<sup>30</sup>, Springer Nature; (c) ref.<sup>70</sup>, (d) ref.<sup>71</sup>, under a Creative Commons Attribution License.

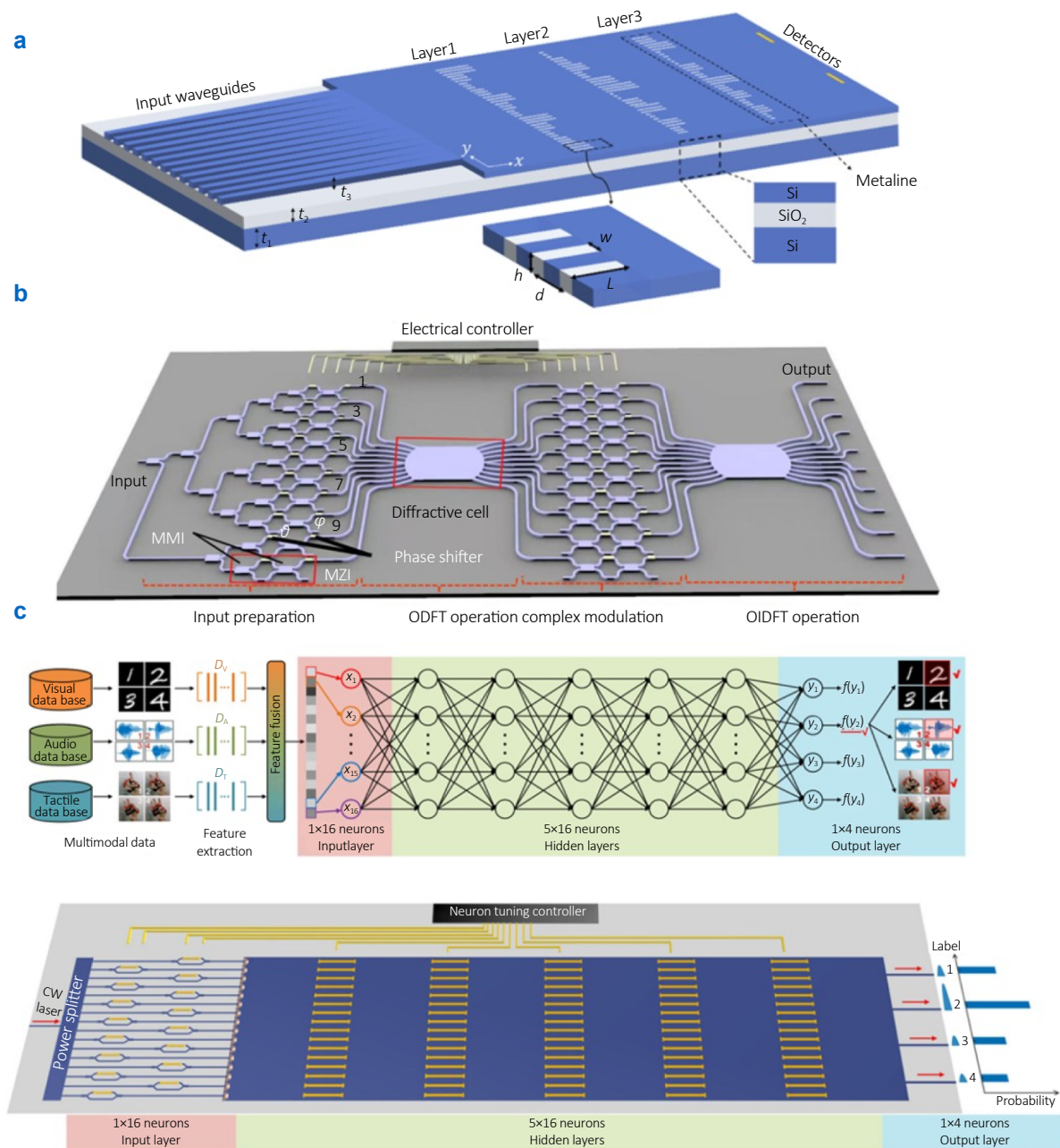


Fig. 7 | Diffractive optical neural networks based on (a) 2D spatial electromagnetic propagation model, (b) ultra-compact diffraction units and multiple MZIs, and (c) trainable. Figure reproduced with permission from: (a) ref.<sup>74</sup>, (b) ref.<sup>78</sup>, (c) ref.<sup>79</sup>, under a Creative Commons Attribution License.

diffraction optical convolution kernel<sup>76</sup>, and further constructed an on-chip optical convolutional neural network, validating excellent performance in image classification and image denoising. In addition, they proposed using a fitting network to map learned parameters to physical structures<sup>77</sup>; compared with traditional weight-mapping models, this approach avoids restrictive approximations and greatly increases the integration density of integrated DONNs. Under the new mapping rule, an integrated DONN can theoretically integrate about  $6 \times 10^4$  computational units per square millimeter.

Besides implementations based on dielectric metasurfaces, integrated diffraction layers can also be realized using phase-shifter arrays and diffraction elements such as multimode interferometers (MMIs). Compared with 1D metasurface structures, multi-port directional couplers achieve smaller footprints between closely spaced parallel waveguides, while in-layer phase-shifter arrays provide reconfigurability at the cost of reducing the number of available photonic neurons. In 2022, Zhu HH et al. demonstrated a silicon-based scalable, fully integrated diffraction neural network<sup>78</sup>, as shown in Fig. 7(b). The network used

two ultra-compact diffractive units and multiple MZIs to realize parallel Fourier transforms, convolutions, and optical computation. The compact diffractive units reduce area and power scaling from quadratic (typical of MZI-based PNNs) to linear with input dimension, thereby significantly increasing on-chip compute density and energy efficiency and offering a new solution for energy-efficient, large-scale optical computing chips.

In 2024, Cheng JW et al. proposed and demonstrated a trainable diffractive optical neural network (TDONN) chip based on integrated diffractive optical devices<sup>79</sup>, as shown in Fig. 7(c), addressing the difficulty of training integrated DONNs. The TDONN chip contains an input layer, five hidden layers, and an output layer, and requires only a single forward pass for inference, without frequent O/E conversions. The researchers designed customized stochastic gradient descent and dropout mechanisms for photonic neurons, enabling in-situ optical training with fast convergence. The chip demonstrated high compute density (447.7 tera operations per second (TOPS)/mm<sup>2</sup>), high energy efficiency (7.28 TOPS/W), and extremely low optical latency (30.2 ps), and achieved four-class classification across multimodal tasks (vision, audio, touch) with 85.7% accuracy.

Unlike traditional interference computing, diffractive computing relies on diffraction through multiple transmissive or reflective layers for information processing. Each neuron performs computation by modulating the phase or amplitude of the optical wave, and the entire process completes at the speed of light. This gives integrated DONNs significant advantages in processing speed and energy efficiency and makes them particularly suitable for low-latency, highly integrated application scenarios. However, the computational capability of integrated DONNs is still constrained by current photonic integration technology, and trade-offs among integration density, power consumption, and compute performance remain an important direction for future research.

### Photonic reservoir computing

Photonic reservoir computing (RC) is a time-series processing method inspired by biological neural networks and is especially suitable for dynamic signals and time-dependent tasks. Compared with traditional recurrent neural networks, RC trains only the output layer while keeping the reservoir fixed, which significantly simplifies training and reduces computational overhead. In 2014, Vandoorne K et al. performed the first experimental validation of reservoir computing on a passive silicon-photonics chip<sup>80</sup>, marking the birth of photonic RC. An RC system maps input signals into a high-dimensional nonlinear space via a dynamic, fixed-topology "reservoir" to capture temporal dependencies and complex patterns. Nonlinearity in the reservoir is typically produced by physical nodes that rely on optical nonlinear effects such as two-photon absorption and free-

carrier dispersion. In photonic implementations, silicon-photonics devices such as MRRs are often used as nonlinear nodes, simulating neuronal dynamics by modulating optical phase and intensity. In 2018, Coarer FDL et al. proposed an all-optical RC architecture using silicon MRRs and successfully demonstrated the delayed-XOR task at 20 Gb/s with power consumption below 2.5 mW<sup>81</sup>, verifying the high energy efficiency of silicon photonic RC. To further improve scalability and compute density, the introduction of "virtual nodes" has become a key path. Through time-multiplexing and related techniques, multiple virtual nodes can be implemented in a single physical node, reducing hardware complexity while increasing processing speed. In 2021, Borghi et al. first proposed and validated an optical RC scheme based on a single silicon microring and time-multiplexing<sup>82</sup>, as shown in Fig. 8(a). The system uses carrier nonlinearities (two-photon absorption and free-carrier dispersion) to realize inter-node coupling and nonlinear responses, requiring no photodetectors or active materials and greatly simplifying the architecture.

In 2025, Donati G et al. proposed an optical-pulse processing and reservoir-computing framework based on silicon MRRs<sup>83</sup>, realizing a sparse and efficient spiking-style RC. The system uses two negatively detuned modes of the microring to inject time-series signals in parallel; when input power exceeds a threshold, self-pulsation occurs. Features of the Iris dataset are encoded separately in wavelength and time, and 92% classification accuracy was achieved. Although currently suitable mainly for relatively simple tasks, the work provides a new idea for temporal feature encoding in photonic RC. In the same year, Lugnan A et al. combined the volatile nonlinear dynamics of MRRs with the nonvolatile storage characteristics of PCMs to construct an adaptive photonic dynamical system<sup>84</sup>, as shown in Fig. 8(b). Using a hierarchical MRR-PCM array together with a lightweight logistic-regression-based training algorithm (no backpropagation), the system reached 98.2% accuracy on MNIST handwritten-digit recognition. The system is highly compact and robust and can be scaled to larger problems via multi-wavelength expansion. The team subsequently proposed a multi-node optical reservoir with 64 silicon MRRs<sup>85</sup>. This architecture can learn and infer the timing and rate of single-pulse inputs and can still respond correctly long after the input excitation has ended, significantly extending the memory timescale of photonic RC and showing potential for complex dynamic signal processing.

In 2024, Wang DL et al. proposed and experimentally demonstrated a silicon-photonics RC chip based on next-generation reservoir computing (NG-RC)<sup>86</sup>, as shown in Fig. 8(c). The chip uses delay lines and star couplers to build a feed-forward reservoir layer and trains the output layer via linear regression, requiring neither the recurrent feedback nor strong nonlinear nodes on which traditional RC relies. Experiments showed an information-processing



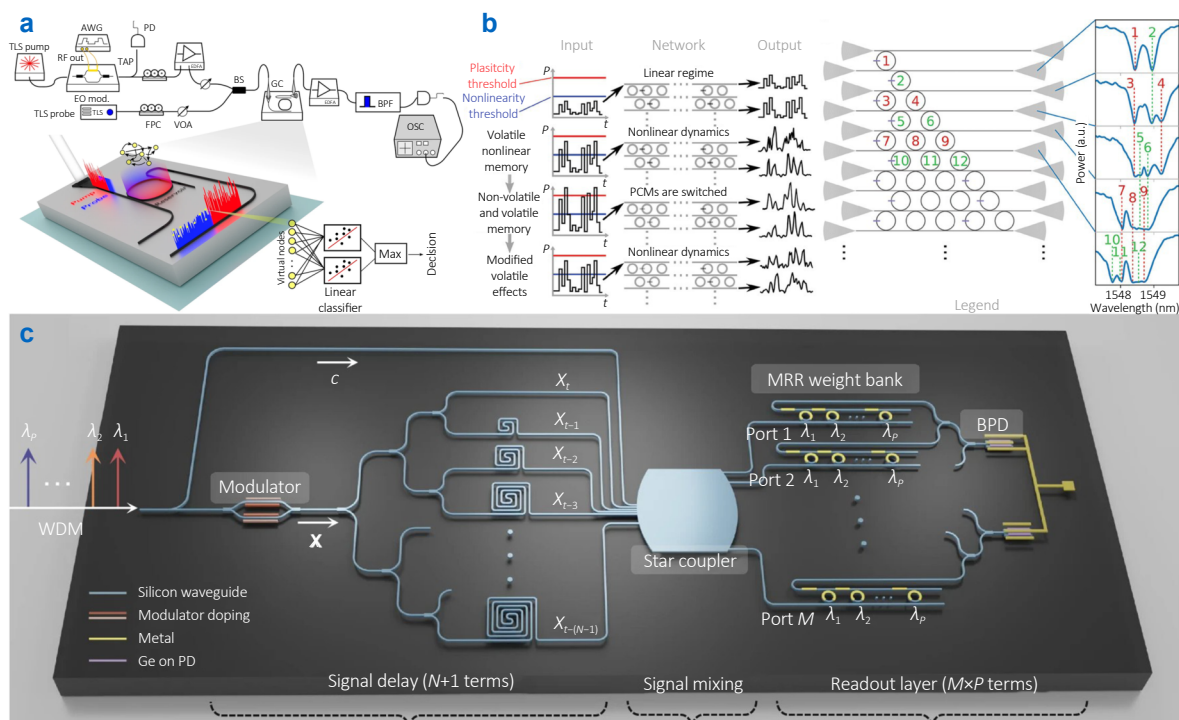


Fig. 8 | Photonic reservoir computing based on (a) time-multiplexing, (b) PCMs, and (c) delay lines and star couplers. Figure reproduced with permission from: (a) ref.<sup>82</sup>, under a Creative Commons Attribution License; (b) ref.<sup>84</sup>, John Wiley and Sons; (c) ref.<sup>86</sup>, under a Creative Commons Attribution License.

rate exceeding 60 GHz, an equivalent compute throughput of about 211 TOPS, a total power of 5.1 W, and an energy efficiency of about 41 TOPS/W, significantly breaking bandwidth and scalability limits.

Overall, photonic RC has demonstrated excellent performance across a variety of temporal tasks. Applications include time-series classification, signal prediction, and nonlinear equalization in optical communications. In classification, RC effectively handles complex time-series data such as speech recognition and handwritten-digit recognition; in prediction, it models temporal dependencies for high-accuracy forecasting; and in optical communications, RC compensates nonlinear distortions caused by transmission, significantly improving system performance. These results demonstrate the great potential of photonic reservoir computing for high-speed, energy-efficient intelligent processing and lay a solid foundation for future ultrafast, low-power photonic neural-computing architectures.

## Prospect

As AI compute demand continues to grow, traditional electronic architectures are increasingly constrained by bandwidth, latency, and power consumption. Photonic neural networks, especially IPNNs, with ultrahigh processing speed, low power, and high-bandwidth parallelism, are becoming a key direction for next-generation intelligent computing. Leveraging the massless, high-speed, and charge-free nature of light, IPNNs exhibit significant

advantages for large-scale matrix multiplication. By optimizing key photonic structures such as MRRs, MZIs, and PCMs, high-density, low-power neural computing can be achieved. IPNNs have already shown strong performance in tasks such as image recognition, speech processing, natural language processing, and scientific computing; several studies have realized optical convolutions, image filtering, and eigenvalue computation, indicating the potential to transition from proof-of-concept to practical deployment. In particular, for inference tasks such as classification and recognition, their low latency and high throughput provide real-time capability suitable for edge computing, autonomous driving, and intelligent sensing.

Optoelectronic hybrid integration is regarded as a key technological route to practical neuromorphic photonic computing. While all-optical schemes offer advantages in energy efficiency and parallelism, they still rely on electrical functions for nonlinear activation, storage, and detection. In 2021, Shastri BJ et al. proposed a system-level packaged neuromorphic photonic processor<sup>11</sup>, as shown in Fig. 9(a). The processor combines advanced PICs, co-integrated electronics, and commercial packaging technology. A reconfigurable photonic neural network is integrated on a silicon PIC, and nonvolatile optical storage uses PCM devices for analog weight tuning. External light sources can be coupled to silicon photonic circuits via fiber coupling, photonic bonding, or modular connections to laser chips. In virtually all brain-inspired photonic architectures, electronic circuitry remains indispensable for network calibration,

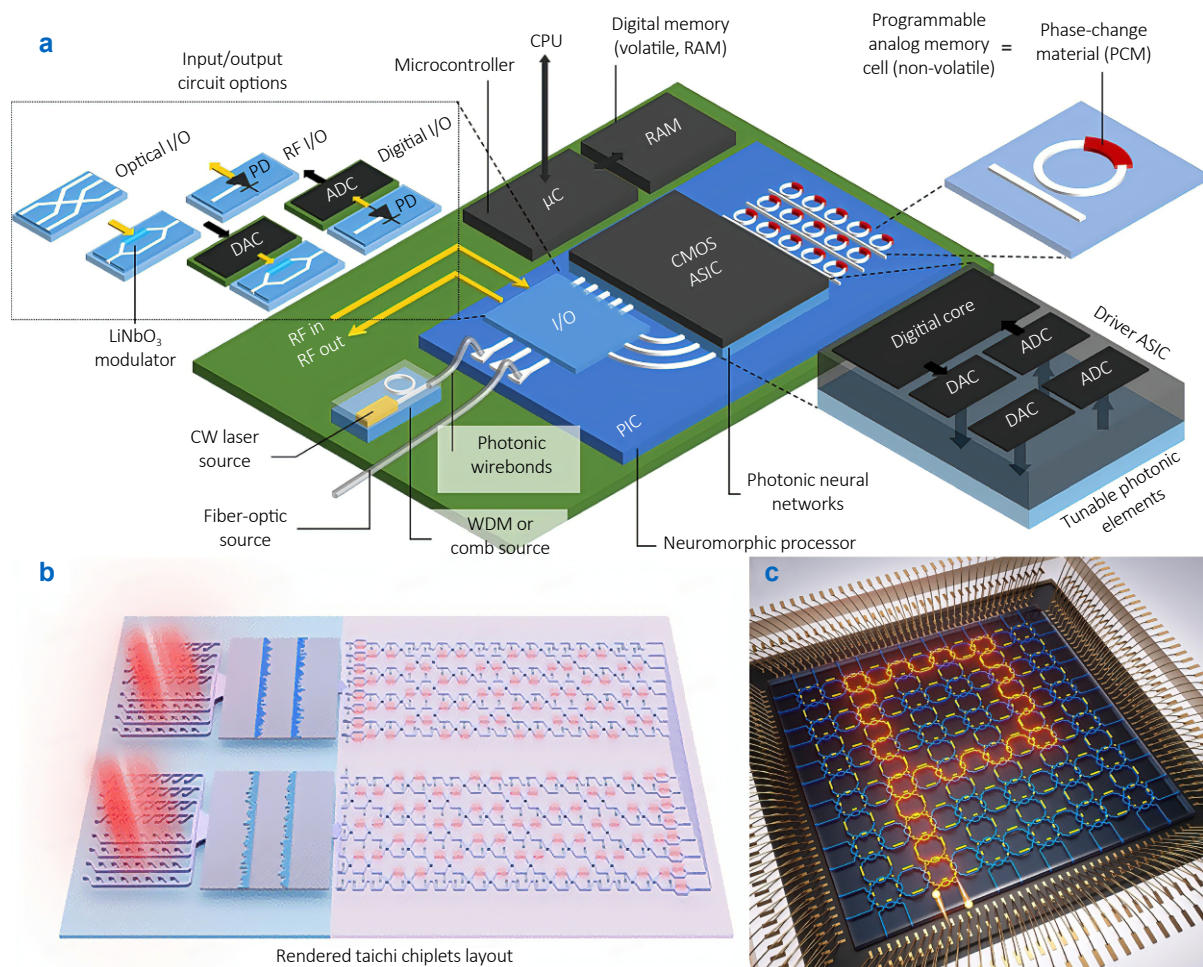


Fig. 9 | The next generation of optical neural networks based on (a) optoelectronic hybrid integration, (b) diffractive-interferometric hybrid architecture, and (c) fully programmable topological photonic chip. Figure reproduced with permission from: (a) ref.<sup>11</sup>, Springer Nature; (b) ref.<sup>87</sup>, (c) ref.<sup>88</sup>, under a Creative Commons Attribution License.

signal modulation, and feedback control. Therefore, deeply fusing photonics and electronics to construct hybrid photonic processors is considered the optimal path to high-performance neuromorphic computing. In this architecture, a CMOS application-specific integrated circuit (ASIC) is flip-chip bonded to the silicon PIC to drive EO devices, configure the photonic network, and provide digital-to-analogue converter (DAC)/analogue-to-digital converter (ADC) interfaces for instruction execution and signal feedback control.

Overall, integrated neuromorphic photonic processors represent a new form of hardware realization for neural networks. This progress has been enabled by high-performance photonic integrated chips, advanced packaging, and algorithmic methodologies tailored for training photonic networks. After decades of sustained research, multiple photonic neuron models, photonic memristors, network architectures, and training strategies have emerged, some of which show significant potential for building neuromorphic processors with high throughput, high density, and high energy efficiency. For practical deployment, achieving

co-packaging and modular real-time integration of optics, electronics, and sources remains a key challenge for commercialization of IPNNs.

In addition to optoelectronic hybrid integration, diffractive-interferometric hybrid architectures have also become an emerging direction for IPNNs. In 2024, Xu ZH et al. proposed a large-scale photonic chipset system named Taichi<sup>87</sup>, as shown in Fig. 9(b), designed for energy-efficient artificial general intelligence tasks. Taichi adopts a diffractive-interferometric hybrid architecture with distributed shallow structures, modular chipset design, and task decomposition/aggregation. The system consists of multiple photonic computing units (tunable execution units, TEUs), each including a diffractive encoder, an interferometric embedding layer (MZI), and a diffractive decoder. Complex tasks are decomposed into multiple binary subtasks; each path completes a local decision, and the results are then encoded and combined to form the final output. Experiments show that Taichi can execute classification with up to 1000 classes and achieves energy efficiency up to 160 TOPS/W for content-generation tasks.

In the future, by integrating lasers, modulators, and photodetectors, a fully packaged system can be realized; by combining direct laser writing (DLW) and PCM technology, fully programmable weight control can be achieved. Taichi's distributed approach is compatible with existing PIC architectures and is expected to further enhance system flexibility and general-purpose computing capability. This architecture breaks through current bottlenecks of photonic neural networks in scale, energy efficiency, and functionality, achieving accuracy close to electronic systems for classification and content generation while delivering order-of-magnitude gains in energy efficiency, thereby demonstrating the practical potential of optical computing beyond the laboratory.

Furthermore, compared to the two-dimensional diffraction neural network, the three-dimensional diffraction neural network offers higher network density and computational power. Our research group, Yu HY et al., has already developed a 3D diffraction optical neural network using three-dimensional two-photon nanolithography technology<sup>89</sup>. In the future, the 3D diffraction optical neural network can be vertically integrated onto PIC to form a 3D hybrid integrated architecture. Compared to the 2D diffractive-interferometric hybrid architecture like Taichi, the 3D-integrated hybrid architecture further improves the network density of diffractive computing, push the IPNN towards high density, high efficiency and high spatial parallelism, without sacrificing its advantages in high speed and programmability.

At present, most photonic neural chips are still designed for specific tasks and lack generality. Looking forward, developing low-cost, highly programmable, general-purpose photonic chips is of great significance. In 2024, Dai TX et al. developed a fully programmable topological photonic chip on a silicon-photonics platform<sup>88</sup>, as shown in Fig. 9(c). The chip treats microring resonators as "artificial atoms" and uses tunable MZIs as "inter-atomic interactions", enabling independent control of coupling strength and phase for each "atom". As a result, various topological optical lattices and operating points can be rapidly reconfigured on the same chip, allowing direct observation and verification of different topological phenomena.

Compared with traditional "one-off custom" samples, this type of chip functions as a reprogrammable hardware carrier, supporting phase-transition control, statistical-physics studies, and prototyping of different topological lattices on a single device. With advances in large-scale optoelectronic packaging and silicon-based heterogeneous integration, such platforms are expected to support more complex fundamental studies (e.g., disordered, inhomogeneous, and anisotropic systems) and to serve as robust optical hardware modules for classical and quantum information processing.

In summary, integrated optical neural networks represent a cutting-edge path to overcoming the bottlenecks of

traditional electronic computing and show great potential in computational density, energy efficiency, and multi-task processing capability. Although current systems still face challenges in nonlinear devices, memory mechanisms, and system integration complexity, continued developments in optoelectronic hybrid integration, diffractive-interferometric hybrid architectures, and general-purpose programmable photonic platforms are expected to enable IPNNs with large scale, low power, and high programmability, providing a solid foundation for the next generation of intelligent computing hardware.

## References

1. Sun C, Wade MT, Lee Y et al. Single-chip microprocessor that communicates directly using light. *Nature* **528**, 534–538 (2015).
2. Atabaki AH, Moazeni S, Pavanello F et al. Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip. *Nature* **556**, 349–354 (2018).
3. Marpaung D, Yao JP, Capmany J. Integrated microwave photonics. *Nat Photonics* **13**, 80–90 (2019).
4. Wan NH, Lu TJ, Chen KC et al. Large-scale integration of artificial atoms in hybrid photonic circuits. *Nature* **583**, 226–231 (2020).
5. Lundstrom M. Moore's law forever. *Science* **299**, 210–211 (2003).
6. Thompson SE, Parthasarathy S. Moore's law: the future of Si microelectronics. *Mater Today* **9**, 20–25 (2006).
7. Cavin RK, Lugli P, Zhirnov VV. Science and engineering beyond Moore's Law. *Proc IEEE* **100**, 1720–1749 (2012).
8. Davies M, Srinivasa N, Lin TH et al. Loihi: a neuromorphic many-core processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
9. Merolla PA, Arthur JV, Alvarez-Icaza R et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
10. Furber SB, Galluppi F, Temple S et al. The SpiNNaker project. *Proc IEEE* **102**, 652–665 (2014).
11. Shastri BJ, Tait AN, Ferreira de Lima T et al. Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics* **15**, 102–114 (2021).
12. Xu XY, Tan MX, Corcoran B et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
13. Willner AE, Khaleghi S, Chitgarha MR et al. All-optical signal processing. *J Lightwave Technol* **32**, 660–680 (2014).
14. Zhu SK, Zheng ZH, Meng WJ et al. Harnessing disordered photonics via multi-task learning towards intelligent four-dimensional light field sensors. *Photonix* **4**, 26 (2023).
15. Xu ZW, Tian H, Zeng Z et al. Harnessing nonlinear optoelectronic oscillator for speeding up reinforcement learning. *Photonix* **6**, 5 (2025).
16. Meng XY, Zhang GJ, Shi NN et al. Compact optical convolution processing unit based on multimode interference. *Nat Commun* **14**, 3000 (2023).
17. Nahmias MA, de Lima TF, Tait AN et al. Photonic multiply-accumulate operations for neural networks. *IEEE J Sel Top Quantum Electron* **26**, 7701518 (2020).
18. Shu HW, Chang L, Tao YS et al. Microcomb-driven silicon photonic systems. *Nature* **605**, 457–463 (2022).
19. Feldmann J, Youngblood N, Karpov M et al. Publisher Correction: parallel convolutional processing using an integrated photonic tensor core. *Nature* **591**, E13–E13 (2021).
20. Wang J, Yang JY, Fazal IM et al. Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat Photonics* **6**, 488–496 (2012).
21. van Uden RGH, Amezcua Correa R, Antonio Lopez E et al. Ultra-high-density spatial division multiplexing with a few-mode multicore fibre. *Nat Photonics* **8**, 865–870 (2014).
22. Li RJ, Gong YH, Huang H et al. Photonics for neuromorphic



- computing: fundamentals, devices, and opportunities. *Adv Mater* **37**, 2312825 (2025).
23. Kutluyarov RV, Zakoyan AG, Voronkov GS et al. Neuromorphic photonics circuits: contemporary review. *Nanomaterials* **13**, 3139 (2023).
  24. Miscuglio M, Mehrabian A, Hu ZB et al. All-optical nonlinear activation function for photonic neural networks [Invited]. *Opt Mater Express* **8**, 3851–3863 (2018).
  25. Chen C, Yang Z, Wang T et al. Ultra-broadband all-optical nonlinear activation function enabled by MoTe<sub>2</sub>/optical waveguide integrated devices. *Nat Commun* **15**, 9047 (2024).
  26. Chen R, Fang ZR, Perez C et al. Non-volatile electrically programmable integrated photonics with a 5-bit operation. *Nat Commun* **14**, 3465 (2023).
  27. Yuan Y, Peng YW, Cheung S et al. All-silicon non-volatile optical memory based on photon avalanche-induced trapping. *Commun Phys* **8**, 39 (2025).
  28. Shen YC, Harris NC, Skirlo S et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).
  29. Tait AN, de Lima TF, Zhou E et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci Rep* **7**, 7430 (2017).
  30. Feldmann J, Youngblood N, Karpov M et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
  31. Hua SY, Divita E, Yu SS et al. An integrated large-scale photonic accelerator with ultralow latency. *Nature* **640**, 361–367 (2025).
  32. Goi E, Zhang QM, Chen X et al. Perspective on photonic memristive neuromorphic computing. *Photonix* **1**, 3 (2020).
  33. Ashtiani F, Geers AJ, Aflatouni F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
  34. Cheng ZG, Ríos C, Pernice WHP et al. On-chip photonic synapse. *Sci Adv* **3**, e1700160 (2017).
  35. Li RZ, Dong YB, Qian FS et al. CsPbBr<sub>3</sub>/graphene nanowall artificial optoelectronic synapses for controllable perceptual learning. *Photonix* **4**, 4 (2023).
  36. Shi B, Calabretta N, Stabile R. Deep neural network through an InP SOA-based photonic integrated cross-connect. *IEEE J Sel Top Quantum Electron* **26**, 7701111 (2020).
  37. Bangari V, Marquez BA, Miller H et al. Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs). *IEEE J Sel Top Quantum Electron* **26**, 7701213 (2020).
  38. Reck M, Zeilinger A, Bernstein HJ et al. Experimental realization of any discrete unitary operator. *Phys Rev Lett* **73**, 58–61 (1994).
  39. Clements WR, Humphreys PC, Metcalf BJ et al. Optimal design for universal multiport interferometers. *Optica* **3**, 1460–1465 (2016).
  40. Tait AN, Nahmias MA, Shastri BJ et al. Broadcast and weight: an integrated network for scalable photonic spike processing. *J Light-wave Technol* **32**, 4029–4041 (2014).
  41. Feldmann J, Youngblood N, Wright CD et al. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
  42. Williamson IAD, Hughes TW, Minkov M et al. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J Sel Top Quantum Electron* **26**, 7700412 (2020).
  43. Tait AN, de Lima TF, Nahmias MA et al. Silicon photonic modulator neuron. *Phys Rev Appl* **11**, 064043 (2019).
  44. Amin R, George JK, Sun S et al. ITO-based electro-absorption modulator for photonic neural activation function. *APL Mater* **7**, 081112 (2019).
  45. Shi Y, Ren JY, Chen GY et al. Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks. *Nat Commun* **13**, 6048 (2022).
  46. Nahmias MA, Shastri BJ, Tait AN et al. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. *IEEE J Sel Top Quantum Electron* **19**, 1800212 (2013).
  47. Peng HT, Nahmias MA, de Lima TF et al. Neuromorphic photonic integrated circuits. *IEEE J Sel Top Quantum Electron* **24**, 6101715 (2018).
  48. Xiang SY, Shi YC, Guo XX et al. Hardware-algorithm collaborative computing with photonic spiking neuron chip based on an integrated Fabry-Perot laser with a saturable absorber. *Optica* **10**, 162–171 (2023).
  49. Xiang SY, Shi YC, Zhang YH et al. Photonic integrated neuro-synaptic core for convolutional spiking neural network. *Opto-Electron Adv* **6**, 230140 (2023).
  50. Wu B, Li HK, Tong WY et al. Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator. *Opt Mater Express* **12**, 970–980 (2022).
  51. Wu CM, Yu HS, Lee S et al. Programmable phase-change meta-surfaces on waveguides for multimode photonic convolutional neural network. *Nat Commun* **12**, 96 (2021).
  52. Strukov DB, Snider GS, Stewart DR et al. Erratum: the missing memristor found. *Nature* **459**, 1154–1154 (2009).
  53. Youngblood N, Ríos Ocampo CA, Pernice WHP et al. Integrated optical memristors. *Nat Photonics* **17**, 561–572 (2023).
  54. Li CL, Yu HY, Shu T et al. PZT optical memristors. *Nat Commun* **16**, 6340 (2025).
  55. Kumar S, Wang XX, Strachan JP et al. Dynamical memristors for higher-complexity neuromorphic computing. *Nat Rev Mater* **7**, 575–591 (2022).
  56. Spagnolo M, Morris J, Piacentini S et al. Experimental photonic quantum memristor. *Nat Photonics* **16**, 318–323 (2022).
  57. Cai FX, Kumar S, van Vaerenbergh T et al. Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks. *Nat Electron* **3**, 409–418 (2020).
  58. Pai S, Sun ZH, Hughes TW et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* **380**, 398–404 (2023).
  59. Bandyopadhyay S, Sludds A, Krastanov S et al. Single-chip photonic deep neural network with forward-only training. *Nat Photonics* **18**, 1335–1343 (2024).
  60. Liao K, Lian YX, Yu MT et al. Hetero-integrated perovskite/Si<sub>3</sub>N<sub>4</sub> on-chip photonic system. *Nat Photonics* **19**, 358–368 (2025).
  61. SeyedinNavadeh S, Milanizadeh M, Zanetto F et al. Determining the optimal communication channels of arbitrary optical systems using integrated photonic processors. *Nat Photonics* **18**, 149–155 (2024).
  62. Li XK, Ma JX, Li XY et al. High-efficiency reinforcement learning with hybrid architecture photonic integrated circuit. *Nat Commun* **15**, 1044 (2024).
  63. Dong BW, Brücknerhoff-Plückelmann F, Meyer L et al. Partial coherence enhances parallelized photonic computing. *Nature* **632**, 55–62 (2024).
  64. Yin RY, Xiao HF, Jiang YH et al. Integrated WDM-compatible optical mode division multiplexing neural network accelerator. *Optica* **10**, 1709–1718 (2023).
  65. Hong SH, Wu JC, Xie YW et al. Versatile parallel signal processing with a scalable silicon photonic chip. *Nat Commun* **16**, 288 (2025).
  66. Zhao BH, Cheng JW, Wu B et al. Integrated photonic convolution acceleration core for wearable devices. *Opto-Electron Sci* **2**, 230017 (2023).
  67. Bai BW, Yang QP, Shu HW et al. Microcomb-based integrated photonic processing unit. *Nat Commun* **14**, 66 (2023).
  68. Jiang XY, He ZM, Wu B et al. Programmable photonic solver for computationally complex problems. *ACS Photonics* **10**, 4340–4348 (2023).
  69. Yu X, Wei ZQ, Sha FY et al. Parallel optical computing capable of 100-wavelength multiplexing. *eLight* **5**, 10 (2025).
  70. Dong BW, Aggarwal S, Zhou W et al. Higher-dimensional processing using a photonic tensor core with continuous-time data. *Nat Photonics* **17**, 1080–1088 (2023).
  71. Zhou W, Dong BW, Farmakidis N et al. In-memory photonic dot-product engine with electrically programmable weight banks. *Nat Commun* **14**, 2887 (2023).
  72. Lin X, Rivenson Y, Yardimci NT et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
  73. Zarei S, Marzban MR, Khavasi A. Integrated photonic neural network based on silicon metalines. *Opt Express* **28**, 36668–36684 (2020).

- (2020).
74. Fu TZ, Zang YB, Huang HH et al. On-chip photonic diffractive optical neural network based on a spatial domain electromagnetic propagation model. *Opt Express* **29**, 31924–31940 (2021).
  75. Fu TZ, Zang YB, Huang YY et al. Photonic machine learning with on-chip diffractive optics. *Nat Commun* **14**, 70 (2023).
  76. Huang YY, Fu TZ, Huang HH et al. Sophisticated deep learning with on-chip optical diffractive tensor processing. *Photonics Res* **11**, 1125–1138 (2023).
  77. Liu WC, Fu TZ, Huang YY et al. C-DONN: compact diffractive optical neural network with deep learning regression. *Opt Express* **31**, 22127–22143 (2023).
  78. Zhu HH, Zou J, Zhang H et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat Commun* **13**, 1044 (2022).
  79. Cheng JW, Huang CR, Zhang JL et al. Multimodal deep learning using on-chip diffractive optics with in situ training capability. *Nat Commun* **15**, 6189 (2024).
  80. Vandoorne K, Mechet P, van Vaerenbergh T et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat Commun* **5**, 3541 (2014).
  81. Coarer FDL, Sciamanna M, Katumba A et al. All-optical reservoir computing on a photonic chip using silicon-based ring resonators. *IEEE J Sel Top Quantum Electron* **24**, 7600108 (2018).
  82. Borghi M, Biasi S, Pavesi L. Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations. *Sci Rep* **11**, 15642 (2021).
  83. Donati G, Biasi S, Pavesi L et al. All-optical spiking processing and reservoir computing with a passive silicon microring and wavelength-time division multiplexing. *Photonics Res* **13**, 2641–2653 (2025).
  84. Lugnan A, Aggarwal S, Brücknerhoff-Plückelmann F et al. Emergent self-adaptation in an integrated photonic neural network for back-propagation-free learning. *Adv Sci* **12**, 2404920 (2025).
  85. Lugnan A, Biasi S, Foradori A et al. Reservoir computing with all-optical non-fading memory in a self-pulsing microresonator network. *Adv Opt Mater* **13**, 2403133 (2025).
  86. Wang DL, Nie YK, Hu GL et al. Ultrafast silicon photonic reservoir computing engine delivering over 200 TOPS. *Nat Commun* **15**, 10841 (2024).
  87. Xu ZH, Zhou TK, Ma MZ et al. Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence. *Science* **384**, 202–209 (2024).
  88. Dai TX, Ma AQ, Mao J et al. A programmable topological photonic chip. *Nat Mater* **23**, 928–936 (2024).
  89. Yu HY, Huang ZH, Lamon S et al. All-optical image transportation through a multimode fibre using a miniaturized diffractive neural network on the distal facet. *Nat Photonics* **19**, 486–493 (2025).

## Acknowledgements

This work was supported by the Shanghai Municipal Science and Technology Major Project, the Science and Technology Commission of Shanghai Municipality (No. 21DZ1100500), the Shanghai Frontiers Science Center Program (2021–2025 No. 20), the National Key Research and Development Program of China (No. 2021YFB2802000), the National Natural Science Foundation of China (No. 61975123, No. 62305217), the National Natural Science Foundation of China (No. 52075504), and the Shanghai Pujiang Programme.

## Competing interests

The authors declare no competing financial interests.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

©The Author(s) 2025.  
Published by Editorial Office of *Opto-Electronic Technology*, Institute of Optics and Electronics, Chinese Academy of Sciences.

