REVIEW ARTICLE | FOCUS
https://doi.org/10.1038/s41566-020-00754-y

nature photonics

Photonics for artificial intelligence and neuromorphic computing

Bhavin J. Shastri[1,2,7]✉, Alexander N. Tait[2,3,7]✉, T. Ferreira de Lima[2], Wolfram H. P. Pernice[4], Harish Bhaskaran[5], C. D. Wright[6] and Paul R. Prucnal[2]

Research in photonic computing has flourished due to the proliferation of optoelectronic components on photonic integration platforms. Photonic integrated circuits have enabled ultrafast artificial neural networks, providing a framework for a new class of information processing machines. Algorithms running on such hardware have the potential to address the growing demand for machine learning and artificial intelligence in areas such as medical diagnosis, telecommunications, and high-performance and scientific computing. In parallel, the development of neuromorphic electronics has highlighted challenges in that domain, particularly related to processor latency. Neuromorphic photonics offers sub-nanosecond latencies, providing a complementary opportunity to extend the domain of artificial intelligence. Here, we review recent advances in integrated photonic neuromorphic systems, discuss current and future challenges, and outline the advances in science and technology needed to meet those challenges.

Conventional computers are organized around a centralized processing architecture (that is, with a central processor and separated memory), which is suited to running sequential, digital, procedure-based programs. Such an architecture is inefficient for computational models that are distributed, massively parallel and adaptive, most notably those used for neural networks in artificial intelligence. Artificial intelligence is an attempt to approach human-level accuracy in tasks that are challenging for traditional computers but easy for humans. Major achievements have been realized by machine learning algorithms based on neural networks[1], which process information in a distributed fashion and adapt to past inputs rather than being explicitly designed by a programmer. Machine learning has had an impact on many aspects of our lives, with applications ranging from translating languages[2] to cancer diagnosis[3]. Neuromorphic engineering is partly an attempt to move elements of machine learning and artificial intelligence algorithms to hardware that reflects their massively distributed nature. Matching hardware to algorithms will potentially lead to faster and more energy-efficient information processing. Neuromorphic hardware is also applied to problems outside machine learning, such as robot control, mathematical programming and neuroscientific hypothesis testing[4,5]. Massively distributed hardware relies heavily—more so than other computer architectures—on massively parallel interconnections between lumped elements (that is, neurons). Dedicated metal wiring for every connection is not practical. Current state-of-the-art neuromorphic electronics therefore use some form of shared digital communication bus that is time-division multiplexed, trading bandwidth for interconnectivity[4]. Optical interconnects could negate this trade-off and thus have the potential to accelerate machine learning and neuromorphic computing.

Light is established as the communication medium of telecoms and datacentres, but it has not yet found widespread use in information processing and computing. The same properties that allow optoelectronic components to excel at communication are at odds with the requirements of digital gates[6]. However, non-digital computing models, such as neural networks, could be more conducive to being implemented in photonics. The goal of neuromorphic photonic processors should not be to replace conventional computers, but to enable applications that are unreachable at present by conventional computing technology—those requiring low latency, high bandwidth and low energies[7]. Examples of applications for ultrafast neural networks include:

- Enabling fundamental physics breakthroughs: qubit read-out classification[8], high-energy-particle collision classification[9,10], fusion reactor plasma control[11]
- Nonlinear programming: solving nonlinear optimization problems (robotics, autonomous vehicles, predictive control)[12] and partial differential equations[13]
- Machine learning acceleration: vector–matrix multiplications[14], deep learning inference[15], ultrafast or online learning[16]
- Intelligent signal processing: wideband radio-frequency signal processing[17], fibre-optic communication[18,19]

Photonic circuits are well suited to high-performance implementations of neural networks for two predominant reasons: interconnectivity and linear operations. Connections between pairs of artificial neurons are described by a scalar synaptic weight (a primary memory element), so the layout of interconnections can be represented as a matrix–vector operation, where the input to each neuron is the dot product of the output from connected neurons attenuated by a weight vector. Optical signals can be multiplied by transmission through tunable waveguide elements, and they can be added through wavelength-division multiplexing (WDM) by accumulation of carriers in semiconductors[20,21], electronic currents[22,23] or changes in the crystal structure of a material induced by photons[24]. Neural networks require relatively long-range connections to perform non-trivial distributed information processing. When comparing metal wire connections with photonic waveguides, optical signals experience lower attenuation and generate less heat (the latter provided the light source if off-chip) as a function of distance.

[1]Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, Ontario, Canada. [2]Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. [3]Applied Physics Division, National Institute of Standards and Technology, Boulder, CO, USA. [4]Institute of Physics and Center for Soft Nanoscience (SoN), University of Münster, Münster, Germany. [5]Department of Materials, University of Oxford, Oxford, UK. [6]Department of Engineering, University of Exeter, Exeter, UK. [7]These authors contributed equally: Bhavin J. Shastri, Alexander N. Tait.
✉e-mail: shastri@ieee.org; alexander.tait@nist.gov

More importantly, waveguides have no inductance or skin effect, which means that frequency-dependent signal distortions are minimal for the long-range connections present in neural interconnects. Electronic point-to-point links today take advantage of transmission line and active buffering techniques; however, neural networks are not based on point-to-point links but instead involve massively parallel signal fan-out and fan-in. It is not practical to use state-of-the-art transmission line and active buffering techniques for each physical connection. Consequently, to avoid the trade-offs exhibited by electronic wiring, neuromorphic electronic architectures employ digital time-multiplexing[4] that allows for the construction of larger neural networks at the expense of bandwidth. For many applications, however, bandwidth and low latency are paramount, and these applications can be met only by direct, non-digital photonic broadcast (that is, many-to-many) interconnects.

Optics has long been recognized as a promising medium for matrix multiplication[25] and interconnects[26,27]. Optical approaches to neural networks were pioneered decades ago by Psaltis and Farhat[28]. Today, societal demands for computing have changed. It is this new demand as well as factors related to the maturity of enabling technologies that has created a renewed case for photonic neural networks. One factor is silicon photonics, which is a crucial advance over previous efforts. Silicon photonic platforms can host high-quality passive components combined with high-speed active optoelectronics, all available with competitive integration density[29]. In 2014, some of us introduced a proposal for a scalable silicon photonic neural network[21], which was demonstrated in 2017[30] concurrently with other silicon photonic neuromorphic architectures[14,22]. On-chip silicon electronics for calibration and control provide a route to overcome component sensitivity[31], and progress in on-chip optoelectronics provides a route to cascadability and nonlinearity[32]. The ability of neuromorphic photonic systems to provide a step change in our computing capabilities is moving ever closer, with potentially petaMAC (multiply-accumulate operations) per second per mm^2 processing speeds[33] and attojoule per MAC energy efficiencies[32]. Although photonics provides advantages in connectivity and linear operations over electronics, other aspects, such as storing and accessing neuron weights in on-chip memory, present new challenges. There have been major investigations of optical memories including 'in-memory' computing[34,35]; however, they cannot usually be written to and read from at high frequencies. Future scalable neuromorphic photonic processors will need to have a tight co-integration of electronics with potentially hybrid electronic and optical memory architectures, and take advantage of the memory type (volatile versus non-volatile) in either digital or analogue domains depending on the application and the computation been performed.

In this Review, we survey recent research and current challenges in neuromorphic photonics and highlight possible solutions. We discuss analogue interconnects (an area where optics conventionally excels) and approaches for implementing neuron nonlinearities (an area where optics conventionally faces difficulties). Neural networks fall into a number of general categories: layered or recurrent, spiking or continuous-time and so on. We survey these various categories and their implications for hardware implementation. The survey directs readers to previous reviews of photonic reservoir computing, a related area of photonic information processing. In the subsequent sections, we discuss key technologies needed for demonstrations of neuromorphic photonic hardware to scale to practical systems, including active on-chip electronics and light sources. Finally, we highlight some emerging research directions towards increasing functionality and efficiency, including non-volatile memory, photonic digital-to-analogue converters (DACs) and frequency comb sources.

Survey of photonic neural networks

Research in neuromorphic photonics encompasses a variety of hardware implementations, and, crucially, multiple neural network types, each with different application classes. In general, all types of neural network consist of nonlinear elements (also known as neurons) interconnected via configurable, linear weights (synapses) (Box 1). Each neural model has a different signal representation, training method and network topology. For example, artificial neurons with a continuous-variable nonlinear transfer function can be trained with backpropagation via gradient descent[1,36], whereas spiking neurons are better suited to different spike-timing-dependent update rules.

Implementations of weighted interconnects (synapses). Connections between a pair of neurons are weighted by their intervening synapse. These synaptic weights are scalar multipliers. Before being received by downstream neurons, the weighted signals from upstream neurons are summed. The weighted interconnects can therefore be represented by a matrix whose entries are the weight values, with each entry multiplying a particular synapse's input signal. One purpose of the photonic system is to perform that matrix multiplication.

Figure 1 shows demonstrations of various integrated photonic circuits for matrix multiplication and weighted interconnection. Implementations fall into two broad categories, one based on wavelength and the other on optical modes. In Fig. 1a, WDM signals are weighted in parallel by a bank of microring resonators (MRRs) used as tunable filters[37,38]. This approach forms an essential part of an integrated architecture called broadcast-and-weight, first proposed in ref. [21] and demonstrated in ref. [30]. Several other architectures for multiwavelength synapses and neural networks have been proposed. Most employ WDM fan-in for weighted addition[23,24,39,40], but they differ in terms of how the channels are weighted. The architectures in refs. [24,39] demultiplex the wavelengths, attenuate each channel and then remultiplex before WDM fan-in. In ref. [39] (Fig. 1b), the weighted attenuators are made from semiconductor optical amplifiers (SOAs); in ref. [24] (Fig. 2g), the tunable attenuators are composed of phase-change materials (PCMs).

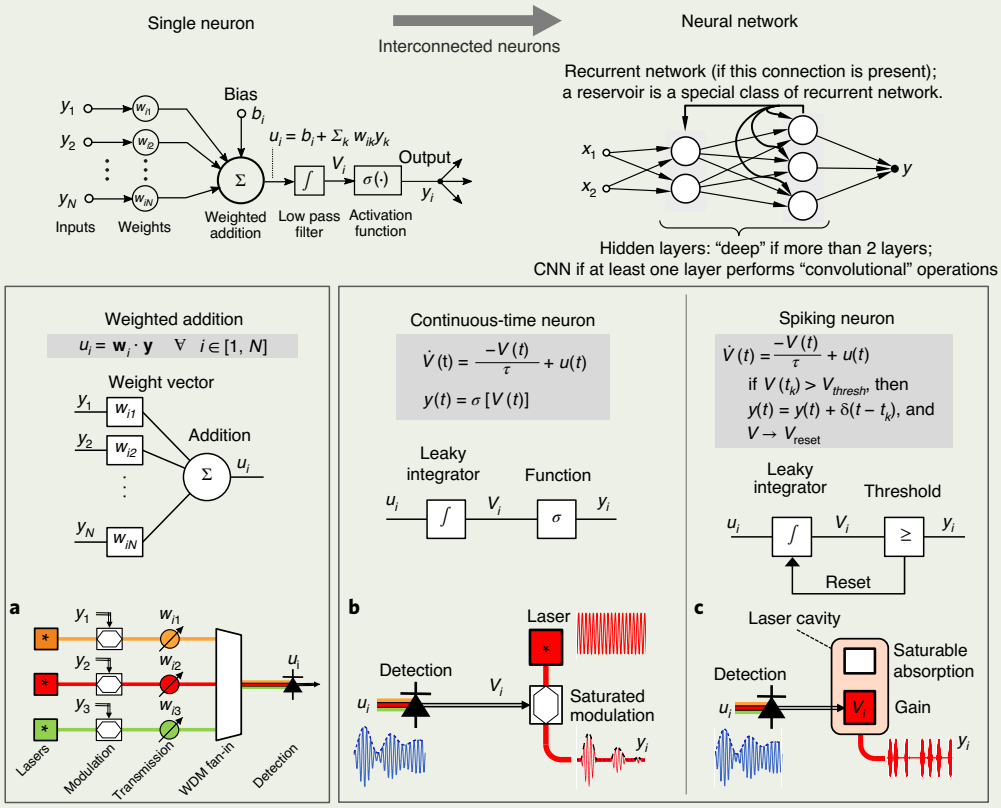
An array of beam splitters and phase shifters can implement unitary matrix transforms using interference between different paths of coherent input light[41], where inputs are assigned to different waveguides and power modulated. This principle underlies mode-based weighted interconnection for photonic neural networks, examples of which are shown in Fig. 1c. This unitary transformation architecture was implemented on an integrated platform using thermally tuned silicon waveguides and directional couplers arranged in a mesh of Mach–Zehnder interferometers (MZIs)[42]. Neural network interconnects can be any matrix—not just unitary. The needed non-unitary neural interconnect was shown to be possible by factoring the weight matrix into one unitary MZI mesh, one array of tunable attenuators and a second unitary MZI mesh[14] (Fig. 1c). The cryogenic architecture from refs. [22,43] uses different optical modes in multiple waveguide layers[44] (Fig. 1d). In contrast to the MZI mesh, signals are incoherent and produced by all-silicon integrated light sources[45]. In addition to the dense crossbar layout, the multilayered waveguide approach enables complex waveguide routing layouts[43].

Modulation of the effective refractive index of signal-carrying waveguides is another optical mode-based approach to weight configuration. A number of index tuning mechanisms have been developed. Thermal tuning with metal filament microheaters is perhaps the easiest way to induce large index changes, but it is slow and power inefficient[46]. Thermal tuning with waveguide-embedded heaters is similar[47], but provides a feedback signal for weight control[48]. In silicon, the strongest effects are the thermo-optic effect, free-carrier absorption and free-carrier dispersion (also known as plasma dispersion). One can directly manipulate carrier concentrations by selectively p- and n-doping the waveguide in a lateral junction[49]. Alternatively, hybrid waveguides can be made of a silicon core and other materials with favourable index modulation

Box 1 | Neuromorphic engineering

Neural network models used in engineering are much simpler than those describing biological neural networks, yet they offer a general computing framework for a wide range of problems. Each artificial neuron in a network can be considered as two functional blocks (top left): a weighted addition unit and a nonlinear unit. The weighted addition has multiple inputs that are 'fanned-in', and one output representing a linear combination of the inputs. The nonlinear unit applies an activation function to the weighted sum, yielding the output of the neuron. The output of the neuron is broadcast (or fanned-out) to many other neurons, possibly including itself. We identify a few possible ways of building a network. The first is a feed-forward neural network (as used in deep neural networks), meaning that signals travel from left to right. The second is a recurrent neural network, in which each neuron can receive outputs from previous, subsequent and the same layer—these are called recurrent connections. A reservoir can be constructed with neural networks containing random, but fixed, recurrent connections. Another important class of neural networks that form an essential part of deep learning is convolutional neural networks, which can be feed-forward or recurrent.

Although these models are constructed of simple elements, it has been shown that networks of neurons are sufficient to perform sophisticated computations and tasks, as exemplified by recent achievements in machine learning. Implementing neural network models directly in hardware, as research shows, leads to speeds and efficiencies unmatched by software implementations, while borrowing the immense body of knowledge about neural network modelling, design and training. Engineering this architecture requires hardware that is isomorphic to neurons and neural networks—hence neuromorphic—to be created. When this isomorphism is achieved, the governing physical dynamics of the hardware will carry out the neural computations in an analogue way. The two diagrams at the top show a schematic of all components a single neuron contains (left), and the multiple ways of connecting neurons to form a neural network (right). The figure shows three isomorphisms that have been explored in photonics: a weighted network[14,24,30] (a), a continuous-time neuron[57,61] (**b**) and a spiking neuron[67,70] (**c**). Each column matches up the model equations and diagram with some photonic devices whose behaviour approximates the equations. **y**, output values of all neurons; **x**, external input vector; w_{ik}, weight strength between neuron i and k; **u**, results of weighted addition; **V**, results of integration of **u**; τ, leaky integration time constant; **b**, adjustable bias offset; V_{thresh} and V_{reset}, threshold and reset voltages of spiking neurons; t, time variable; t_k, point in time at which a spike occurs, when V crosses V_{thresh}.



properties close enough to the core that interact with the evanescent field. Some examples include group III–V hybrid integration[50], lithium niobate[51] and graphene[52] modulation. These mechanisms are faster and require much less power than heaters, but typically provide a smaller tuning range before electrical damage. Tuning methods based on chalcogenide PCMs allow weights to retain their values without further holding power after being set[24,35].

Examples of non-volatile synapse implementations are shown in Fig. 1e,f. These approaches have been referred to as all-optical because they do not need electrical inputs for tuning. Both are based on the use of optically induced changes in chalcogenide materials to control the light propagation in waveguides (the former Si_3N_4 integrated waveguides[24], the latter metal sulphide fibres[53]). Weight configurations based on non-volatile optical materials could have

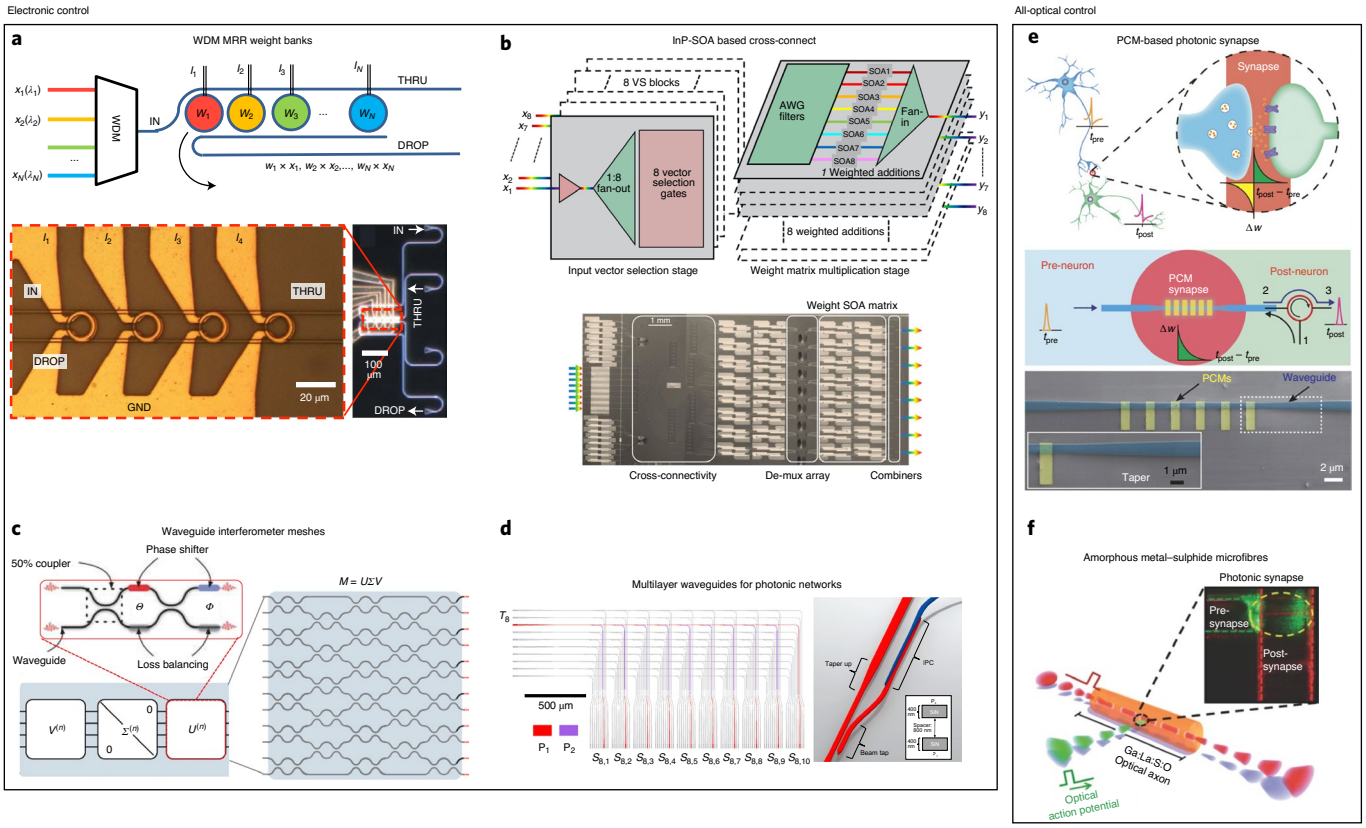


Fig. 1 | Implementations of weights or photonic synapses. a, Parallel weighting of WDM signals (x_i) with MRR weight banks as tunable filters[37,38] (top). The optical micrograph (bottom) shows a fabricated silicon MRR weight bank with metal heaters for thermo-optic tuning of weights (w_i) with currents (I_i). A balanced photodetector sums these signals (Σwx) and allows for positive and negative weights. **b**, Weighted addition with an SOA chip on an InP platform[39]. A schematic (top) and microscope image (bottom) of a chip co-integrating eight weighted additions for eight WDM input vectors, providing eight WDM outputs, are shown. **c**, An MZI composed of waveguides and directional couplers with phase shifters implements a unitary transform[14]. Representing a weight matrix $M = U\Sigma V^{\dagger}$ through singular value decomposition, unitary matrices U and $V^{\dagger}$ are implemented with MZIs and diagonal matrix Σ with a Mach–Zehnder modulator. **d**, Photonic routing and weighting scheme for all-to-all connectivity using two vertically integrated planes of silicon nitride waveguides with a beam tap and an interplanar coupler (IPC)[44]. **e**, Photonic synapse implemented with PCMs integrated on silicon nitride waveguides[116]. Synaptic weight is varied by the number of optical pulses sent down the waveguide. **f**, Photonic synapses demonstrated using metal sulphide microfibres[53]. The transmission of pulses along the fibre is altered through photodarkening as a result of exposure at a subbandgap wavelength. This photomodulation plays the role of either inhibitory or excitatory action potentials in the post-synaptic axon. Figure adapted with permission from ref. [37], IEEE (**a**) and ref. [14], SNL (**c**). Figure reproduced with permission from ref. [39], IEEE (**b**); ref. [53], Wiley (**f**); ref. [44], AIP (**d**) and ref. [116], Wiley (**f**). Distributed under Creative Commons license CC BY 4.0.

a notable impact on the challenges of electrical input/output (I/O) and heat dissipation.

Implementations of nonlinearities (neurons). In all neural network models, some form of nonlinearity is required in the primary signal pathway to implement the thresholding effect of the neuron. A multitude of photonic devices exhibit nonlinear transfer functions that resemble neuron-like or gate-like transfer functions; however, a nonlinear response alone is not sufficient for a photonic device to act as a neuron. Photonic neurons must be capable of reacting to multiple optical inputs (fan-in), applying a nonlinearity and producing an optical output suitable to drive other like photonic neurons (cascadability). Optical devices face fundamental challenges in satisfying these requirements in particular, as pointed out by Keyes[6] and Goodman[54] decades ago. Today, these challenges and requirements are being addressed with integrated photonic solutions, with some successful approaches shown in Fig. 2. These approaches fall into two major categories on the basis of the physical representation of signals within the neuron: optical/electrical/optical (O/E/O) versus all-optical.

O/E/O neurons, proposed in 2013 by Nahmias et al.[55] and Romeira et al.[56], involve a transduction of optical power into electrical current and back within the primary signal pathway. The primary signal pathway or drive chain refers to elements representing the rapidly changing neuron state variables akin to membrane voltage or synaptic state. Outside this primary pathway, one always finds electronics representing the relatively slowly varying neuron parameters and control logic. In O/E/O neuron signal pathways, nonlinearities occur in the electronic domain or in the E/O conversion stage that uses lasers or saturated modulators. The use of E/O nonlinearities for photonic neurons was demonstrated using modulators in refs. [57–59] (Fig. 2a,b) and using lasers in ref. [60] (Fig. 2c). A photodetector–modulator neuron for MZI meshes was proposed in ref. [61] (Fig. 2d). Other O/E/O approaches implement nonlinearity purely in the electronic domain; for example, it was proposed in refs. [22,62] that the nonlinear dynamics of spiking photonic neurons could be implemented with a superconducting electronic signal pathway (Fig. 2e).

In separating input light–matter interaction (O/E) from output light–matter interaction (E/O), O/E/O neurons can modulate their output signal onto a fresh optical carrier that is unconstrained by the optical characteristics (that is power, phase, mode, wavelength) of its inputs. Importantly, this means that the output can be

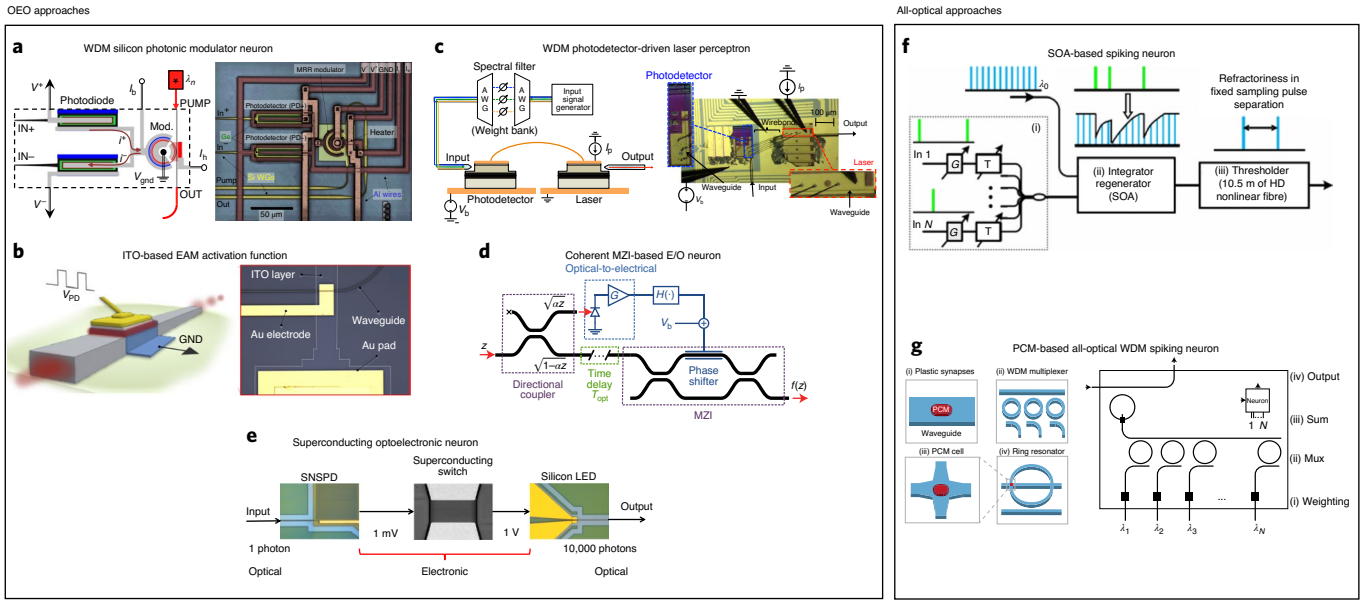


Fig. 2 | Photonic neurons incorporating weighting and nonlinearity. a, Silicon photonic modulator neuron[57]. A balanced photodiode (from Fig. 1a) sums multiple wavelengths and implements positive (excitatory) and negative (inhibitory) weights, and drives a ring modulator exploiting its electro-optic nonlinearity. **b**, Similar to **a**, but has an electro-absorption modulator (EAM) with an indium tin oxide (ITO) layer monolithically integrated into silicon photonic waveguides[58]. **c**, The device uses WDM to achieve multichannel fan-in, a photodetector to sum signals together and a laser cavity to perform a nonlinear operation[60]. In **a**–**c**, device diagrams are shown on the left next to micrographs of each device. **d**, A photodetector-driven MZI-based nonlinear activation function[61] is used for waveguide interferometer mesh weights (as in Fig. 1c). **e**, A superconducting optoelectronic spiking neuron based on a superconducting-nanowire single-photon detector (SNSPD) driving a superconducting switch (amplifier)[62] followed by a silicon LED[45]. **f**, An integrate-and-fire SOA spiking neuron[65]. Neuron inputs are weighted and delayed (with attenuators and delay lines; i), integrated with an SOA (ii) and thresholded with a highly Ge-doped fibre (iii). **g**, A PCM-based spiking neuron[24]. Inputs are weighted using PCM synapses (i; similar to Fig. 1e), summed using a WDM multiplexer (MUX) (ii) and thresholded with a PCM cell (iii) on a ring resonator (iv). Figure adapted with permission from ref. [58], AIP (**b**); ref. [62], SNL (**f**) and ref. [45], AIP (**e**). Figure reproduced with permission from ref. [57], APS (**a**); ref. [60], AIP (**c**); ref. [61], IEEE (**d**); ref. [65], OSA (**f**) and ref. [24], SNL (**g**). Distributed under Creative Commons license CC BY 4.0.

appreciably stronger than the input, which is not generally the case for all-optical neurons, as discussed below. A key distinction among O/E/O photonic neurons is whether the high-bandwidth nonlinear transfer function is imparted by analogue components (such as E/O responses, transistor circuits, single-photon detectors) or by a digital lookup circuit[14,39,61]. The transfer functions of analogue neurons can be configured by electrical biasing, but their shapes are constrained by the response of whatever device provides the nonlinearity. Digital counterparts provide the flexibility to implement arbitrary transfer functions. Several works have shown, however, that matching particular transfer function shapes is not necessary: neural training and programming techniques can be adapted to transfer functions naturally exhibited by analogue photonic devices[18,30,63].

All-optical neurons do not ever represent the neuron signal as an electrical current, but instead represent it as changes in material properties such as semiconductor carriers or optical susceptibility. Optical nonlinear susceptibilities are power inefficient—not just very weak—meaning that neuron output is necessarily, and often substantially, weaker than its input and thus incapable of driving even a single other neuron. Solutions to fan-out and cascadability have been demonstrated by combining nonlinear optical devices with optical carrier regeneration. Regeneration means that each neuron outputs a fresh carrier wave, which is power-modulated by that neuron's output signal, a function of the sum of its input signals. Carrier regeneration approaches were first adapted to photonic neurons in 2002 by Hill et al.[64] and have been used in all experimental demonstrations of photonic neurons to date, including all-optical neurons.

Carrier regeneration involves the control of output light with input light. All-optical neurons must provide this function in addition to a nonlinear function. Carrier regeneration enabling photonic neurons has been shown in a feed-forward fashion using semiconductor carrier populations: using cross-gain modulation[65] (Fig. 2f) or using cross-phase modulation in an interferometer[20]. It can also be achieved by changing a material state, such as via a structural phase transition[24,66] (Fig. 2g). All-optical carrier regeneration introduces a new challenge: differentiating the controller signal from the controlled signal. Both affect the material substrate, so the output must be weaker than the input. Optical amplifiers can be employed to boost the output such that it can drive downstream neurons. Like the continuous-time neurons above, spiking laser neurons can also be categorized into the two broad classes of O/E/O and all-optical. A range of implementations for both classes is summarized in Fig. 3. Spiking laser neurons achieve strong nonlinearity, carrier regeneration and neural dynamics all within a single device consisting of gain, a cavity and a saturable process. Spiking neurons have been demonstrated using saturable semiconductor media[55,67,68] (Fig. 3a,f), resonant tunnelling diodes[56,69] (Fig. 3b), graphene saturable absorbers[70] (Fig. 3d) and mode competition[71–73] (Fig. 3c,e).

A perceived advantage of all-optical neuron implementations is that they are inherently faster than O/E/O implementations due to relatively slow carrier drift and/or current flow stages in the latter. Indeed, for optical telecommunications, O/E/O for the purpose of digital regeneration is considered undesirable and inefficient; however, the bottleneck is not due to the transduction between light and current—it is caused by a need to demultiplex and digitize many different channels. The majority of proposed O/E/O neurons do not involve digitization, and thus rarely impose bandwidth bottlenecks. In fact, recent analogue O/E/O devices have exhibited bandwidth and energy performance on par with or better than all-optical

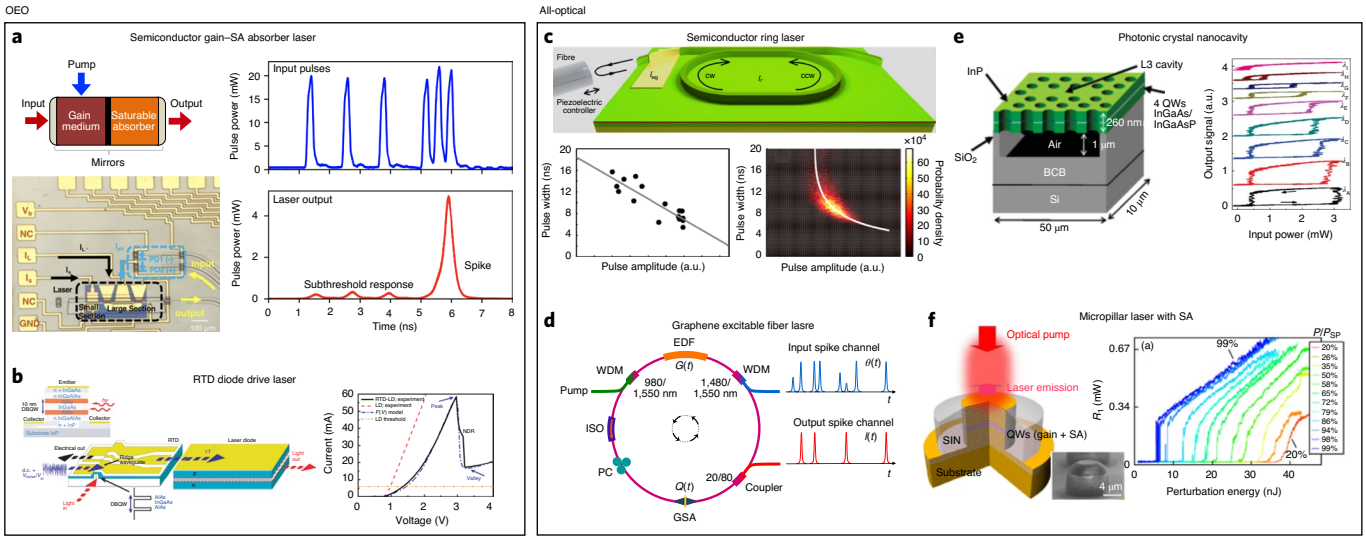


Fig. 3 | Excitable lasers and resonators for spiking. a, Left, a two-section gain (integrate) and saturable absorber (SA; thresholder) excitable laser as an integrate-and-fire neuron[55] (top) and micrograph of electrically injected excitable distributed feedback laser[68] driven by balanced photodetector pair (bottom). Right, measured excitable power of the input pulses (top) and output of the laser (bottom). **b**, Left, resonant-tunnelling diode (RTD) layer stack, photodetector, and laser diode[56]. Right, excitability is achieved by biasing a double-barrier quantum well (DBQW) within the RTD in the negative differential resistance (NDR) region of its d.c. current–voltage curve. **c**, A semiconductor ring laser consisting of an electrically pumped group III–V ring resonator coupled to a waveguide (top) and excitable behaviour (bottom)[71]. Two counterpropagating (CW and CCW) modes per frequency lead to bistability. Excitability arises when this symmetry is broken. **d**, Graphene SA excitable fibre laser[70]. An erbium-doped fibre (EDF) acts as gain medium that is optically injected and pumped. **e**, Left, InP-based 2D photonic crystal nanocavity with quantum wells (QWs)[72]. This device exploits fast third-order nonlinearity for excitability. Right, hysteresis cycles show bistability with different detuning values with respect to the cavity resonance. a.u., arbitrary units. **f**, Left, optically pumped group III–V micropillar laser with an SA[67]. Right, amplitude response to a single pulse perturbation versus perturbation energy for bias pump P relative to the self-pulsing threshold P_{SP}, demonstrating the distinction between excitable and self-pulsing thresholds. Figure adapted with permission from refs. [55,68], IEEE (**a**) and ref. [133], APS (**c**, bottom). Figure reproduced with permission from ref. [56], OSA (**b**); ref. [134], ScienceDirect (**c**, top); ref. [70], SNL (**d**); ref. [72], APS (**e**) and ref. [67], APS (**f**). Distributed under Creative Commons license CC BY 4.0.

components, as compellingly illustrated in the O/E/O work of Nozaki and others[32]. For neurons that apply nonlinearities in the digital domain—the most extreme example being in a CPU—the digital subsystem is often the determinant of the maximum system bandwidth.

Neuromorphic architectures (neural networks). Neuromorphic hardware architecture is governed by models of artificial neural networks; however, neural networks models have many subclasses. They can differ in terms of neuron signal representation, weight configuration and network topology. Weight configuration, in this context, refers to the approach of setting weights so that the network accomplishes a particular computational task. Configurations can be guided by supervised training, unsupervised learning (plasticity) or programmatic 'compilation'. Topology describes the graph structure of non-zero weights between neurons. In the most general case—the all-to-all recurrent topology—there are forward- and backward-directed connections between each pair of neurons. However, constraining the topology in particular ways can unlock powerful analytical tools to guide weight configuration. For example, feed-forward topologies yield to chain rule decomposition[36], and symmetric topologies yield to an energy surface formulation[74]. Configuration and topology are intertwined with signal representation and the behaviour of individual neurons. In feed-forward networks, the output is completely determined by the present inputs, meaning that neuron input signals can be scalar values as opposed to functions of time (although they change when new inputs are presented or when weights are updated). In recurrent networks, on the other hand, the outputs also depend on the history of inputs. This means that neurons must have internal states that evolve non-instantaneously through time. They are referred to as stateful neurons. Box 1 shows two types of stateful neurons, one with continuous-valued outputs and another with outputs consisting of temporal delta functions (that is, spikes).

Examples of experimentally demonstrated photonic architectures are shown in Fig. 4. The model in Fig. 4a (ref. [30]) is recurrent, continuous-time and programmed by compiler[75]. The model used in Fig. 4b (ref. [14]) is feed-forward, single-valued and externally trained. The model in Fig. 4c (ref. [24]) is feed-forward and spiking, with both external and local training. Figure 4d shows a feed-forward multilayer perceptron architecture that combines semiconducting few-photon light-emitting diodes (LEDs) with superconducting-nanowire single-photon detectors to behave as spiking neurons, connected by a network of optical waveguides[22]. Figure 4e,f shows free-space diffractive network implementations: a recurrent type[76] (Fig. 4e) and a feed-forward deep neural network implementation[77] (Fig. 4f). A free-space diffractive network with a nonlinear activation function has also been demonstrated[78]. The computational tasks of the above examples range over, respectively, audio classification, dynamical system emulation, image classification, nonlinear optimization and neuroscientific hypothesis testing.

Classification of unstructured signals is a key application of machine learning. Networks used for classification are often feed-forward: arranged in layers with all connections from a given layer projecting to the following layer. They are often called deep networks. Convolutional neural networks (CNNs or ConvNet)[36] are a type of deep network that consists of a series of convolutional layers that perform a sliding dot product on vectors from the previous layer for local feature extraction, followed by a pooling layer to merge similar features into one. In supervised training, the outputs

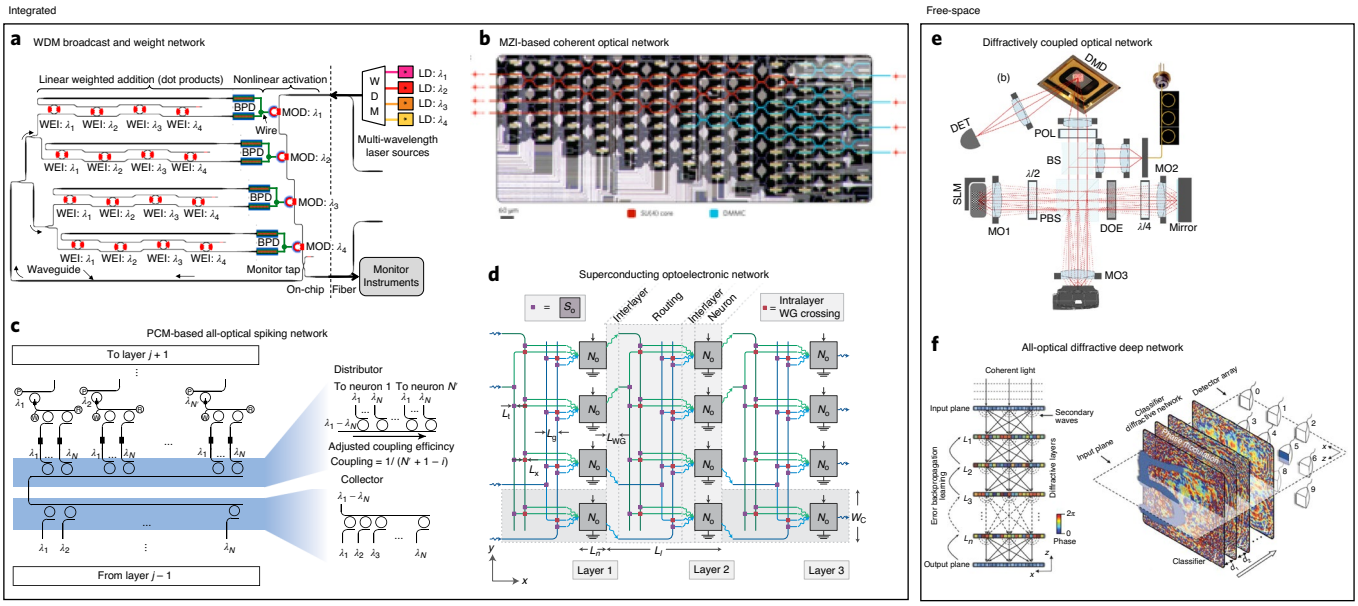


Fig. 4 | Photonic neural network implementations. a, A multiwavelength broadcast-and-weight photonic neural network[21,57] composed of MRR weight banks (WEI) (as in Fig. 1a) and balanced photodiodes (BPD) for linear weighting and summing, and MRR modulators (MOD) for nonlinear activation (as in Fig. 2a). **b**, A coherent all-optical feed-forward network composed of MZI meshes (as in Fig. 1c) configured to implement a matrix using singular value decomposition: red meshes implement a unitary matrix and blue meshes implement a diagonal matrix[14]. **c**, A WDM-based all-optical neural network using PCM[24]. A collector made of MRRs multiplexes optical pulses from the previous layer (bottom) and a distributor broadcasts the input signal equally to the PCM synapses of each neuron (top) (Fig. 2g). **d**, A multilayer perceptron implemented (as in Fig. 2e) with superconducting optoelectronic network platform[22] with N_o neurons. **e**, Diffractively coupled photonic nodes forming a large-scale recurrent neural network[76]. A spatial light modulator (SLM) encodes the networks' state and a digital micromirror device (DMD) creates a spatially modulated image of the SLM's state. The output is obtained via superimposing the detected modulated intensities. **f**, Left, diffractive deep neural network based on coherent waves[77]. Right, the network comprises multiple transmissive (or reflective) layers; each point on a given layer acts as a neuron, with a complex-valued transmission (or reflection) coefficient. Figure reproduced with permission from ref. [57], APS (**a**); ref. [14], SNL (**b**); ref. [24], SNL (**c**); ref. [22], APS (**d**); ref. [76], OSA (**e**) and ref. [47], AAAS (**f**).

are compared to the correct outputs for corresponding inputs. The classification error determines how the weights are updated. After this training phase, the network is able to generalize to inputs that have not been seen before. An important case of supervised training for deep networks, called backpropagation, provides rules to update all of the weights in the network[1]. Backpropagation training calculations can be performed by a computer and then applied back to weights and biases in photonic hardware, either coherent approaches in feed-forward networks[14,77,79] or WDM-based CNNs[23,80]. Convolutional operations require a delicate balance between processing and memory access—a challenge for photonics discussed later. Photonic hardware implementations of backpropagation have also been proposed[16].

In addition to machine learning and classification, neural networks can be programmed to solve challenging mathematical problems, specifically nonlinear differential equations and nonlinear optimization. Such methods have been applied to photonics to address predictive control for fast-moving bodies[12] and Ising machines for the study of many-body physics and molecular chemistry[81,82]. Networks with inhibitory connections (weights) have also been shown to solve combinatorial problems for compressed sensing[83]. The Neural Engineering Framework[75] is a promising tool that mixes compiling and reward-based adaptation to implement block-diagram-style control and decision systems. The Neural Engineering Framework represents variables with populations of neurons—as opposed to single neurons—and has been applied to photonic neural networks in ref. [30]. Further work is needed to identify benchmarks and applications of extreme computing with ultrafast neuromorphic hardware[84].

Other architectures under investigation typically involve a higher level of brain-inspired principles. Spiking neuron models represent signals as trains of Dirac delta functions, digital in amplitude but analogue in time. Temporal coding (that is, neural processing related to the timing of spikes) has been proposed as one of the keys to energy efficiency in biological nervous systems[85]. Extensive research in photonics has been directed to optoelectronic devices that approximate spiking dynamics[22,24,55,56,67,70–72,86–88]. Spike coding opens key new unsupervised plasticity methods—especially spike-timing-dependent plasticity, which is considered to be essential to a number of brain areas and has been explored in photonics in refs. [24,43,89–91].

Photonic reservoir computing. Photonic reservoir computing is a growing field of interest and importance in photonic information processing. Full coverage of this area falls outside the scope of this Review; the interested reader is directed towards the many excellent reports in the literature (for example, refs. [92–96], and recent extensive reviews[97,98]). Reservoir computers consist of a network of random connections and nonlinearities (the reservoir) followed by a read-out layer. The reservoir generates a large number of complex possible behaviours in response to an input, and the read-out layer is trained to select the behaviour that solves a particular computational problem of interest. Many types of physical substrates can be used as the reservoir, including a variety of photonic systems. Research in photonic reservoir computing has realized dynamical complexity by combining optical delays with one of the following: a photodetector and modulator[92], optical amplifiers[94], a semiconductor nonlinearity[95] or a recurrent interferometer mesh[96]. Although both reservoir and neuromorphic approaches share roots in neuroscience, they represent complementary approaches to information processing. All neuromorphic systems require a known isomorphism between the physical hardware and a neural network model on which training

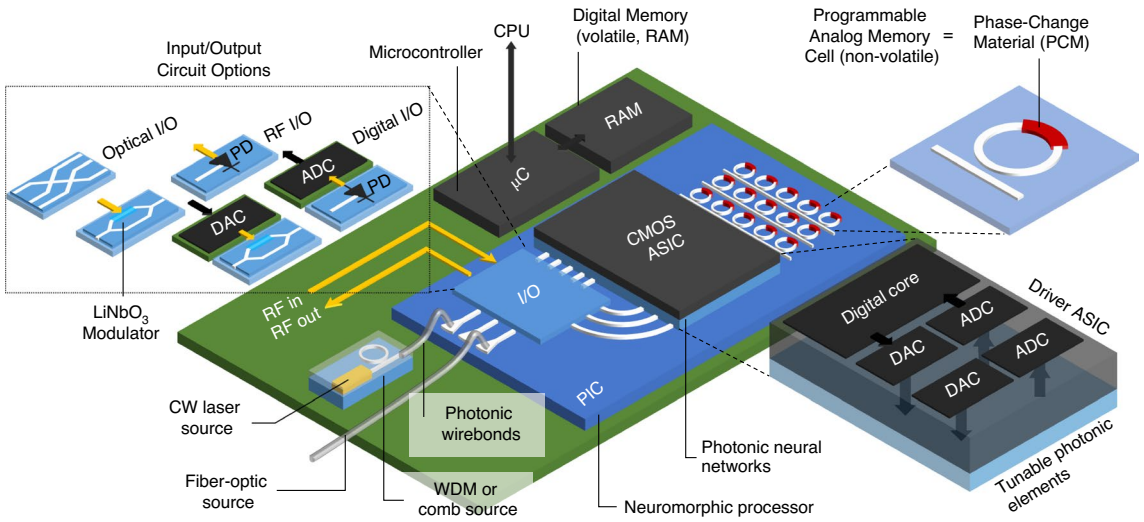


Fig. 5 | Neuromorphic photonic processor architecture. A conceptual system-in-package employing commercially available photonic packaging technology and some emerging ideas in the field of integrated photonics is shown. At a high level, the co-integration, packaging and I/O strategies are closely related to those in programmable photonics[135]. The photonic neural network with configurable optical elements resides on a silicon PIC die. Some elements can be configured in an analogue, non-volatile way by PCMs. White lines represent waveguide routing. One key challenge is getting optical power onto the silicon die. Optical power can be provided by an optically active die that is able to generate light, or, alternatively, an external fibre interface. For signal I/O (top left inset), E/O conversions and vice versa are performed by silicon photonic modulators (hexagonal structures) and photodetectors (triangles with crosses). This means that all package-level I/O can be electrical or optical, digital or analogue, depending on the chosen application. The other key challenge is controlling the photonic neural network. Black boxes represent CMOS dies satisfying various control, interface and programming roles. The digital programming interface consists of a microcontroller (μC) with co-located digital memory (RAM), both of which are standard components. A CMOS ASIC is flip-chip bonded to the PIC. The ASIC generates a large number of voltages to drive the electro-optic elements (for example, waveguide-embedded heaters) and thus configure a photonic neural network. The ASIC also provides DACs to set drive voltages on the basis of μC instructions and digital registers to maintain drive values when not being addressed by the μC. As it is outside the primary signal pathway, the ASIC need not be high-bandwidth and thus can be manufactured on modestly performant CMOS nodes using commercially available design blocks. RF, radio-frequency.

and programming are based; reservoir systems can accomplish processing goals without requiring this isomorphism, instead requiring supervised training to read out desired behaviours. Consequently, despite their behavioural complexity, photonic reservoirs can be simple to construct and rapid progress has been made in task-based demonstrations—for example, a single optoelectronic modulator and optical fibre were used to classify spoken digits[95] and a spatial light modulator-based reservoir computer[76] was used for computer vision tasks to recognize human actions[99].

Towards a neuromorphic photonic processor

Mainstream silicon photonic platforms offer a device library (modulators, waveguides, detectors) with which to implement the main signal pathways in at least some neuromorphic architectures. Modifications to standard manufacturing processes, including the introduction of PCMs and/or superconducting electronics, extend potential realizable architectures even further, as will be discussed in the Emerging ideas and outlook section. In all architectures, there is a need for complex on-chip electronic circuitry for calibration and control of the network parameters, and a need to generate light. Such components are not yet widely available on current commercial silicon photonic platforms. Here we discuss different routes to integrating the electronic circuits and light sources necessary for a neuromorphic photonic processor. A complete processor in a package is rendered in Fig. 5, illustrating the roles of the respective technologies.

Active on-chip electronics. Photonic chips require d.c. analogue signals (bias voltages/currents for example), control systems (such as feedback, algorithms and so on), interfaces with electronics (DACs and analogue-to-digital converters (ADCs)) and require stabilization (in terms of temperature, for example). Dedicated analogue electronic circuits are needed for these purposes—some low bandwidth (d.c. configuration) and some high bandwidth (DACs, ADCs, transimpedance amplifiers, feedback). As a result, neuromorphic photonic chips generally require considerably more electrical ports than optical ones, leading to a high electronic interconnect density. The required number of electrical ports usually scales quadratically with the number of optical ports. This challenge can be overcome by co-integration of complementary metal oxide semiconductor (CMOS) electronics with photonic chips. There are several technological routes with different trade-offs. We discuss three of these options: wirebonding, flip-chip bonding and monolithic fabrication.

CMOS chips with digitally controlled analogue devices can be connected with the photonic integrated circuit via lateral wirebonds. However, as the processors scale in numbers of elements it will (1) become physically impossible to have dedicated wirebonds, with an eventual limit to the electrical I/O and (2) be very expensive to have a large part of the chip's real estate dedicated to routing metal traces to its perimeter. In addition, wirebonding imposes a limit on signal bandwidth because of parasitic inductance in the wires. Wirebonding remains, however, a good approach to engineer small-scale prototype systems in a laboratory setting.

Flip-chip bonding involves fabricating two dies, one optimized for CMOS electronics and the other optimized for silicon photonics. The dies have matching electrical pads. They are soldered, or bonded, to one another to create a large number of electrical connections[31]. Figure 5 illustrates the flip-chip bonding approach.

A custom CMOS application-specific integrated circuit (ASIC) mates with the pads on a silicon photonic integrated circuit and provides the interface between the many distributed, continuously configurable optical devices and the predominantly serial digital realm. Each DAC for each tuning element must be capable of holding its value in a digital register—in other words, the memory for parameters (that is, weights, biases) is distributed and co-located with corresponding optical elements.

Flip-chip bonding brings several advantages over wirebonding. First, the connection number grows with the area of each chip. Second, the interconnections have reduced parasitic impedance and higher bandwidth. Finally, the multi-die approach can combine the best of electronic and photonic technologies, because each die can be optimized independently. From an economic standpoint, this type of integration does not require the development of new processes and can rely on existing designs for microelectronic microprocessors, memory and DACs. Multi-die approaches introduce new challenges in managing thermal fluctuations from the CMOS die. Through-silicon via, a vertical electrical connection that passes through the silicon wafer or die, is an alternative strategy to connect electronic and silicon dies. Through-silicon vias trade off high integration density and low interconnect parasitics with complexity in thermal management. A comprehensive review and comparison of these technologies is provided in ref. [31].

Monolithic fabrication entails the integration of electronics and photonics on the same substrate, and several approaches have been investigated to achieve this. So-called zero-change platforms earn their name by offering a monolithic fabrication process that is minimally changed from an industry-standard fabrication process (such as a 45 nm silicon-on-insulator CMOS process[100,101]). Another direction is to adapt silicon-on-insulator photonic processes to enable some active integrated electronic logic[102]. A dedicated homogeneous process called 9 WG, recently offered by GlobalFoundries[103], looks to balance photonic performance with electronic circuitry for analogue or digital control.

Light sources. A substantial challenge for silicon as a photonic platform is its inability to generate light on-chip. Current approaches in the silicon photonics data communications industry rely on fibre packaging with external light sources. For neuromorphic silicon photonic systems, fibre packaging is also the most straightforward way to get light into the chip. However, as photonic processing systems do not intrinsically need to send optical signals off-chip, ideally all optical signals would be confined within an integrated circuit package. Thus, co-packaged light sources will eventually be critical for the efficiency, stability and scalability of neuromorphic photonics. Figure 5 depicts both an external fibre-optic light source and a multi-die approach using a non-silicon die for light generation. Both are shown with photonic wirebond connections, although other techniques for coupling light between dies are available. Substantial research has been dedicated to integrating light sources directly onto the silicon waveguide layer[104]. Approaches include rare-earth-element doping, strain engineering of germanium and all-silicon emissive defects[45]. Each approach offers distinct features, and drawbacks, to thermal efficiency, integration compatibility, scalability and temperature stability.

Another approach involves the use of group III–V devices integrated with silicon. InGaAsP, for example, provides high gain and saturation powers, but is difficult to integrate directly on silicon due to the crystal lattice mismatch between the materials. Ways around this include either (1) using a separate group III–V die to couple light to the silicon die, or (2) integrating group III–V devices on silicon waveguides. The former has been achieved by pick-and-place bonding of group III–V and Si dies with finished gain ridges[105], or using laser micropackages[106]. Pick-and-place techniques can suffer from tight alignment tolerances. System-in-package approaches—where the laser die is placed next to the silicon die—can be enabled by photonic wirebonding[107]. Photonic wirebonding involves writing three-dimensional waveguides in a photosensitive polymer. As this step is performed after die placement, alignment requirements are relaxed. Integration of group III–V dies with Si has been achieved by bonding wafers with group III–V quantum wells to silicon waveguides[108]. Finally, group III–V quantum dots have been grown directly on silicon[109]. Unlike quantum layers, quantum dots can tolerate lattice mismatch without the loss of their optical gain properties.

The requirements for lasers for neuromorphic photonics depend mostly on the type of neuron. Multi-die techniques are well suited to neuromorphic systems with modulator-class neurons because the light source can be well outside the neural signal pathway. For laser-class neurons, the gain must be tightly integrated on waveguides, necessitating wafer bonding, quantum dot growth or silicon emissive defects. Sources for modulator neurons are faced with one of two non-trivial performance requirements. For multi-wavelength architectures, as in Fig. 1a, many different wavelengths would require many different sources. Another option would be a single source that produces multiple wavelengths (for example, a frequency comb source). Coherent architectures, as in Fig. 1c, face almost the opposite challenge: there can only be a single source laser because the network requires a single optical phase reference. A challenge is that this single laser must generate enough optical power for the entire system.

Emerging ideas and outlook

As pointed out in a recently published 'roadmap' on emerging hardware and technology for machine learning[110], neuromorphic photonics could provide an outstanding candidate. Some neuromorphic photonic architectures can feasibly be built on commercial silicon photonic platforms, given electronic and light-source integration; however, to transform early system demonstrations into practical and fully performant artificial intelligence processors, neuromorphic photonic systems must evolve to incorporate new technologies. For example, correcting fabrication variability can reduce heat dissipation and the amount of current needed for tuning. Memory circuits that are able to interact directly with light can enable more agile reconfiguration in the processor. Finally, information can be transferred between electrons and photons with reduced heat dissipation, which can be achieved with further technological improvements in optical sources, high-efficiency modulators and photonic analogue-to-digital interfaces. This section highlights key emerging technologies that could have great impact on the performance potential of neuromorphic photonic processors.

Memory in neuromorphic processors. So far, the approaches described here have relied on a combination of specialized photonic devices being driven by more generalized electronic circuits or micro-controllers. This is mostly because current photonic platforms lack some of the building blocks that are common in electronics, such as logic gates, high-level compilers and assemblers, analogue–digital–analogue conversion and, importantly, memory.

In some machine learning and neuromorphic applications (for example, deep learning inference) the synaptic weights, once trained, do not have to be updated often or at all. In these cases, non-volatile analogue memory would be beneficial, a potential candidate being 'in-memory' computing, which can be implemented optically or electronically with PCMs[34,66]. When controlled by digital electronic drivers running photonics-compatible firmware, one can implement a neural network running in real time, precompute the necessary weights and directly load an inference task onto hardware. Although weights may not need to be updated often, there are cases (for example in long short-term memory recurrent neural networks) where the output of the neurons need to be stored

temporarily (that is, written to and read from memory quickly). In these cases, PCMs are probably not the right technology to store that type of data. Digital—or perhaps short-term analogue[111,112]—electronic memory with electro-optic interfaces to analogue photonics would be best suited to store that particular type of data without displacing a role for PCMs in storing long-term weights. When compared with digital memory, analogue memory would be limited in precision and noise, but it has been shown[113] that deep and recurrent neural networks work well even in low precision.

All modern computers have heterogeneous memory technologies (registers, caches, flash, dynamic random-access memory (DRAM) and so on) within a single system. Neuromorphic photonic systems are expected to follow a similar principle, with tighter co-packaging of electronic memory. We may also witness the integration of novel photonic memory technologies. For example, progress beyond inference machines could benefit from non-volatile, yet reconfigurable, optical materials for memory. The concept of memory here is closely related to 'learning'. In applications such as online learning and training, synaptic weights need to be updated frequently. In these cases, fast memory (normally provided by DRAM, for example) is necessary and photonics DACs could play a role. In modern computers, memory is used to store both programs and data. The CPU juggles between reading instructions, executing them and manipulating data in and out of a memory unit. Non-volatile optical materials could serve a role beyond just data storage or instruction storage. They can be part of the computational algorithm running on the neuromorphic processor. Synaptic weights are often considered as effectively the long-term memory of a network, but physiological synapses are constantly changing according to neural activity and due to their molecular chemistry.

Non-volatile photonic memory can be implemented by cladding waveguides with PCMs, including chalcogenides, which are used at present for memory storage applications (such as rewritable optical disks)[35,114]. These materials feature multiple stable phases of matter that have distinct optical properties. The most commonly used PCM is the archetypal alloy $Ge_2Sb_2Te_5$, but other compositions are also being investigated (for example the 'low-loss' material GeSbSeTe; ref. [115]). Changing the state of the PCM (from amorphous to crystalline) varies the waveguide's effective refractive index (changing the optical-path length) and modulates the absorption. As devices such as silicon weight banks rely on the refractive index of a waveguide—as reviewed in Fig. 1 and rendered in Fig. 5—PCM-cladded waveguides offer non-volatile weights that can be reconfigured with optical or electrical signals[34,116]. The process of setting the weights is also reversible, which limits the need to read from and write to electronic memories with DACs and ADCs. Non-volatile weights for photonic neural networks can also be achieved in a cryogenic setting through the long-term storage of superconducting currents[43]. With a reversible weight-setting process, dynamic synaptic plasticity and online learning can be enabled by local feedback circuits implemented in each neuron.

Correcting for variability. Analogue circuits often need trimming to function correctly and to correct for manufacturing variabilities and environmental sensitivity. In integrated photonics, resonant devices (such as MRRs) can pose particular problems in this respect[117]. One approach to address this challenge is resonance trimming. Active trimming, for example inducing changes in refractive index by heating a waveguide, is useful to counteract environmental variability, such as temperature and vibration, but requires constant input power. Furthermore, it requires microsecond response speeds and a continuous, rather than discrete, response to an electrical signal. An alternative approach is to use permanent or non-volatile methods to trim the refractive index of a device (that is, passive trimming). This approach can be employed to correct for fabrication variation, or to preprogram a circuit to a default state. Permanent methods, typically employed during manufacture, include electron-beam-induced compaction and strain of oxide cladding[118], electron-beam bleaching of a polymer cladding[119] and, the most CMOS-foundry-compatible approach, germanium ion implantation and annealing[120]. Non-volatile methods include field-programmable PCMs, and can be reconfigured in place[35].

To map application tasks to photonic hardware, new analogue-aware compilers are necessary. Compilers on conventional computers change high-level code to machine instructions that differ depending on the computer. Likewise, photonic compilers should be able to abstract away the idiosyncrasies that occur when we represent signals in WDM lightwaves subject to nonlinear distortion, limited dynamic range, limited gain and cross-talk. This work is in progress in the academic community in the fields of neuromorphic photonics[7,12,14,30] and programmable photonics[121,122].

Frequency comb-based WDM sources. In combination with on-chip multiplexers and demultiplexers for WDM, tailored light sources that provide evenly spaced emission wavelengths aligned to standardized communication channels are desirable. One approach is to use an array of integrated WDM lasers, yet an attractive alternative is chip-scale frequency combs leveraging nonlinear optics, requiring only one laser.

Using soliton microcombs, frequency combs can now be built using CMOS-compatible photonic integrated circuits[123]. The resonators are designed such that the nonlinearity compensates for the dispersion and a single soliton can circulate within the cavity. A frequency comb then forms inside the resonator with a frequency spacing equal to the free spectral range of the resonator[124,125]. A single soliton state is prepared in a high-Q-factor ring resonator, leading to a very stable and broadband frequency comb with a wavelength spacing that can be aligned to the telecom ITU grid (100 GHz), determined by the free spectral range of the resonator. For photonic neuromorphic processing, chip-scale microcombs provide the means to generate tailored input signals in a highly parallel fashion from a single source.

Lithium niobate-on-insulator modulators. Lithium niobate ($LiNbO_3$) was one of the early enabling platforms for realizing 'planar lightwave circuits'; it offers attractive material properties, including a wide transparency window covering visible to mid-infrared wavelengths and a strong electro-optical coefficient[126]. It also provides high second-order optical nonlinearity, which allows for parametric wavelength conversion and nonlinear mixing. Thin-film $LiNbO_3$-on-insulator (LNOI) has emerged as a photonic substrate and become commercially available. The LNOI platform allows photonic circuits with dimensions compatible with silicon photonic devices[127] to be built and provides the ability to integrate fast electro-optical modulators and efficient nonlinear optical elements on the same chip.

LNOI-based modulators can achieve very high modulation frequencies (up to 110 GHz; ref. [128]) while maintaining a low voltage–length product ($V_\pi L$), which is beneficial in terms of energy consumption and device footprint. Recent demonstrations showed monolithically integrated $LiNbO_3$ electro-optic modulators that feature CMOS-compatible driving voltages, data rates up to 210 Gb s^{-1} and an on-chip optical loss of less than 0.5 dB (ref. [129]). In combination with the capability to electro-optically modulate on-chip, this approach enables the integration of nanophotonic waveguides, MRRs, filters and modulators on the same chip. Such devices could lead to large-scale ultra-low-loss photonic circuits that are reconfigurable on a picosecond timescale.

Photonic DACs. Interfacing analogue processors invariably requires ADC and DAC circuits operating at full data rates (tens of gigahertz). Going from digital–electronic to analogue–photonic signals

requires two rather costly conversion steps: digital-to-analogue conversion and electro-optic modulation. One possibility is to combine these two steps into one, using photonic DACs to achieve high sampling rates, high precision, and low distortion, while being less affected by jitter or electromagnetic noise than electronic counterparts. Neuromorphic photonic processors would benefit from photonic DACs compatible with silicon photonic integration for reduced footprint, high sampling rates and low power consumption.

One approach is based on optical intensity weighting of multiwavelength signals, modulated with silicon MRRs with depletion-mode PN junctions[130,131]. In this approach, the bit number is determined by the number of cascaded MRRs, which is further determined by the free spectral range of the MRR and the channel spacing each MRR occupies (which affects the cross-talk). The main challenge here is achieving high-speed operation with a high extinction ratio. A 2 bit photonic DAC was demonstrated in ref. [130], with a silicon MRR with a modulation data rate of up to 128 Gb s^{-1} (64 Gbaud). Another 2 bit DAC demonstration was based on a silicon-on-insulator travelling-wave multi-electrode Mach–Zehnder modulator operating at 100 Gb s^{-1} (50 Gbaud)[131]. Scaling to a higher number of bits may be achieved with a coherent parallel photonic DAC proposed in ref. [132].

Conclusion

Neuromorphic photonics is the creation of optoelectronic hardware that is isomorphic to neural networks. As a consequence of this isomorphism, photonic neural networks will have remarkable capabilities, they will have strong technological and societal demand, and will leverage existing algorithmic methodologies for programming and training, including the leaps and bounds occurring in the practice of deep learning.

Research in photonic neural networks has proliferated considerably in recent years. Several architectural concepts and implementations of a variety of neuron models, training techniques and topologies are already under investigation. This diversity implies that neuromorphic photonics research should not be expected to converge to a single winning implementation or a single application. Continuous research is required to identify applications where photonics will most excel over the continually advancing state of the art in electronic computing. Relatedly, there will be a persistent demand for benchmarks comparing emerging photonic and electronic technologies. Most promising will be real-time applications, where decisions must occur in a very short time. Moving forwards, it is now time to focus on scaling the number of neurons integrated in single networks. With key photonic libraries having been demonstrated on scalable (silicon) photonics platforms, the critical technological challenges are co-packing of control electronics and light sources. Abetted by modern integrated platforms for programmable photonics, new ideas and devices for on-chip cascadability and nonlinearity—seemingly insurmountable barriers to the bandwidth of neuromorphic electronics and an obvious societal demand for neural network processors—neuromorphic photonics has a high potential to extend the frontiers of machine learning and information processing.

Received: 23 April 2020; Accepted: 16 December 2020;
Published online: 29 January 2021

References

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Wu, Y. et al. Google's neural machine translation system: bridging the gap between human and machine translation. Preprint at https://arxiv.org/abs/1609.08144 (2016).
3. Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
4. Merolla, P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
5. Davies, M. et al. Loihi: a neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
6. Keyes, R. W. Optical logic-in the light of computer technology. *Opt. Acta* **32**, 525–535 (1985).
7. Prucnal, P. R. & Shastri, B. J. *Neuromorphic Photonics* (CRC, 2017).
8. Magesan, E., Gambetta, J. M., Corcoles, A. D. & Chow, J. M. Machine learning for discriminating quantum measurement trajectories and improving readout. *Phys. Rev. Lett.* **114**, 200501 (2015).
9. Radovic, A. et al. Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560**, 41–48 (2018).
10. Duarte, J. et al. Fast inference of deep neural networks in FPGAs for particle physics. *J. Instrum.* **13**, P07027 (2018).
11. Kates-Harbeck, J., Svyatkovskiy, A. & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568**, 526–531 (2019).
12. Ferreira de Lima, T. et al. Machine learning with neuromorphic photonics. *J. Lightwave Technol.* **37**, 1515–1534 (2019).
13. Han, J., Jentzen, A. & Weinan, E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl Acad. Sci. USA* **115**, 8505–8510 (2018).
14. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
15. Jouppi, N. P. et al. In-datacenter performance analysis of a tensor processing unit. In *Proc. 44th Annual International Symposium on Computer Architecture* 1–12 (Association for Computing Machinery, 2017).
16. Hughes, T. W., Minkov, M., Shi, Y. & Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **5**, 864–871 (2018).
17. Tait, A. N. et al. Demonstration of multivariate photonics: blind dimensionality reduction with integrated photonics. *J. Lightwave Technol.* **37**, 5996–6006 (2019).
18. Huang, C. et al. Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems. In *Optical Fiber Communication Conference* Th4C–6 (Optical Society of America, 2020).
19. Zhang, S. et al. Field and lab experimental demonstration of nonlinear impairment compensation using neural networks. *Nat. Commun.* **10**, 3033 (2019).
20. Kravtsov, K. S., Fok, M. P., Prucnal, P. R. & Rosenbluth, D. Ultrafast all-optical implementation of a leaky integrate-and-fire neuron. *Opt. Express* **19**, 2133–2147 (2011).
21. Tait, A. N., Nahmias, M. A., Shastri, B. J. & Prucnal, P. R. Broadcast and weight: an integrated network for scalable photonic spike processing. *J. Lightwave Technol.* **32**, 4029–4041 (2014).
22. Shainline, J. M., Buckley, S. M., Mirin, R. P. & Nam, S. W. Superconducting optoelectronic circuits for neuromorphic computing. *Phys. Rev. Appl.* **7**, 034013 (2017).
23. Bangari, V. et al. Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs). *IEEE J. Sel. Top. Quantum Electron.* **26**, 7701213 (2020).
24. 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).
25. Goodman, J. W., Dias, A. R. & Woody, L. M. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Opt. Lett.* **2**, 1–3 (1978).
26. Goodman, J. W., Leonberger, F. J., Kung, S.-Y. & Athale, R. A. Optical interconnections for VLSI systems. *Proc. IEEE* **72**, 850–866 (1984).
27. Miller, D. A. B. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–749 (2000).
28. Psaltis, D. & Farhat, N. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Opt. Lett.* **10**, 98–100 (1985).
29. Soref, R. & Bennett, B. Electrooptical effects in silicon. *IEEE J. Quantum Electron.* **23**, 123–129 (1987).
30. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
31. Bogaerts, W. & Chrostowski, L. Silicon photonics circuit design: methods, tools and challenges. *Laser Photon. Rev.* **12**, 1700237 (2018).
32. Nozaki, K. et al. Femtofarad optoelectronic integration demonstrating energy-saving signal conversion and nonlinear functions. *Nat. Photon.* **13**, 454–459 (2019).
33. Nahmias, M. A. et al. Photonic multiply-accumulate operations for neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 7701518 (2020).
34. Ríos, C. et al. In-memory computing on a photonic platform. *Sci. Adv.* **5**, eaau5759 (2019).
35. Ríos, C. et al. Integrated all-photonic non-volatile multi-level memory. *Nat. Photon.* **9**, 725–732 (2015).

36. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
37. Tait, A., Ferreira de Lima, T., Nahmias, M., Shastri, B. & Prucnal, P. Continuous calibration of microring weights for analog optical networks. *Photon. Technol. Lett.* **28**, 887–890 (2016).
38. Tait, A. N. et al. Microring weight banks. *IEEE J. Sel. Top. Quantum Electron.* **22**, 312–325 (2016).
39. 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).
40. Xu, X. et al. Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks. *Laser Photon. Rev.* **14**, 2000070 (2020).
41. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994).
42. Carolan, J. et al. Universal linear optics. *Science* **349**, 711–716 (2015).
43. Shainline, J. M. et al. Superconducting optoelectronic loop neurons. *J. Appl. Phys.* **126**, 044902 (2019).
44. Chiles, J., Buckley, S. M., Nam, S. W., Mirin, R. P. & Shainline, J. M. Design, fabrication, and metrology of 10 × 100 multi-planar integrated photonic routing manifolds for neural networks. *APL Photon.* **3**, 106101 (2018).
45. Buckley, S. et al. All-silicon light-emitting diodes waveguide-integrated with superconducting single-photon detectors. *Appl. Phys. Lett.* **111**, 141101 (2017).
46. Harris, N. C. et al. Efficient, compact and low loss thermo-optic phase shifter in silicon. *Opt. Express* **22**, 10487–10493 (2014).
47. Jayatilleka, H. et al. Wavelength tuning and stabilization of microring-based filters using silicon in-resonator photoconductive heaters. *Opt. Express* **23**, 25084–25097 (2015).
48. Tait, A. N. et al. Feedback control for microring weight banks. *Opt. Express* **26**, 26422–26443 (2018).
49. Patel, D. et al. Design, analysis, and transmission system performance of a 41 GHz silicon photonic modulator. *Opt. Express* **23**, 14263 (2015).
50. Komljenovic, T. et al. Heterogeneous silicon photonic integrated circuits. *J. Lightwave Technol.* **34**, 20–35 (2016).
51. 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).
52. Sorianello, V. et al. Graphene–silicon phase modulators with gigahertz bandwidth. *Nat. Photon.* **12**, 40–44 (2018).
53. Gholipour, B. et al. Amorphous metal-sulphide microfibers enable photonic synapses for brain-like computing. *Adv. Opt. Mater.* **3**, 635–641 (2015).
54. Goodman, J. W. Fan-in and fan-out with optical interconnections. *Opt. Acta* **32**, 1489–1496 (1985).
55. Nahmias, M. A., Shastri, B. J., Tait, A. N. & Prucnal, P. R. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. *IEEE J. Sel. Top. Quantum Electron.* **19**, 1800212 (2013).
56. Romeira, B. et al. Excitability and optical pulse generation in semiconductor lasers driven by resonant tunneling diode photo-detectors. *Opt. Express* **21**, 20931–20940 (2013).
57. Tait, A. N. et al. Silicon photonic modulator neuron. *Phys. Rev. Appl.* **11**, 064043 (2019).
58. Amin, R. et al. ITO-based electro-absorption modulator for photonic neural activation function. *APL Mater.* **7**, 081112 (2019).
59. George, J. K. et al. Neuromorphic photonics with electro-absorption modulators. *Opt. Express* **27**, 5181–5191 (2019).
60. Nahmias, M. A. et al. An integrated analog O/E/O link for multi-channel laser neurons. *Appl. Phys. Lett.* **108**, 151106 (2016).
61. Williamson, I. A. D. et al. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 7700412 (2020).
62. McCaughan, A. N. et al. A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors. *Nat. Electron.* **2**, 451–456 (2019).
63. Mourgias-Alexandris, G. et al. An all-optical neuron with sigmoid activation function. *Opt. Express* **27**, 9620–9630 (2019).
64. Hill, M., Frietman, E. E. E., de Waardt, H., Khoe, G.-D. & Dorren, H. All fiber-optic neural network using coupled SOA based ring lasers. *IEEE Trans. Neural Netw.* **13**, 1504–1513 (2002).
65. Rosenbluth, D., Kravtsov, K., Fok, M. P. & Prucnal, P. R. A high performance photonic pulse processing device. *Opt. Express* **17**, 22767–22772 (2009).
66. Sebastian, A. et al. Tutorial: brain-inspired computing using phase-change memory devices. *J. Appl. Phys.* **124**, 111101 (2018).
67. Selmi, F. et al. Relative refractory period in an excitable semiconductor laser. *Phys. Rev. Lett.* **112**, 183902 (2014).
68. Peng, H. T. et al. Neuromorphic photonic integrated circuits. *IEEE J. Sel. Top. Quant. Electron.* **24**, 6101715 (2018).
69. Romeira, B., Avo, R., Figueiredo, J. M. L., Barland, S. & Javaloyes, J. Regenerative memory in time-delayed neuromorphic photonic resonators. *Sci. Rep.* **6**, 19510 (2016).
70. Shastri, B. J. et al. Spike processing with a graphene excitable laser. *Sci. Rep.* **6**, 19126 (2016).
71. Coomans, W., Gelens, L., Beri, S., Danckaert, J. & Van der Sande, G. Solitary and coupled semiconductor ring lasers as optical spiking neurons. *Phys. Rev. E* **84**, 036209 (2011).
72. Brunstein, M. et al. Excitability and self-pulsing in a photonic crystal nanocavity. *Phys. Rev. A* **85**, 031803 (2012).
73. Robertson, J., Deng, T., Javaloyes, J. & Hurtado, A. Controlled inhibition of spiking dynamics in VCSELS for neuromorphic photonics: theory and experiments. *Opt. Lett.* **42**, 1560–1563 (2017).
74. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
75. Stewart, T. C. & Eliasmith, C. Large-scale synthesis of functional spiking neural circuits. *Proc. IEEE* **102**, 881–898 (2014).
76. Bueno, J. et al. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **5**, 756–760 (2018).
77. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
78. Zuo, Y. et al. All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).
79. Xu, S., Wang, J., Wang, R., Chen, J. & Zou, W. High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays. *Opt. Express* **27**, 19778–19787 (2019).
80. Mehrabian, A., Miscuglio, M., Alkabani, Y., Sorger, V. J. & El-Ghazawi, T. A Winograd-based integrated photonics accelerator for convolutional neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 6100312 (2020).
81. McMahon, P. L. et al. A fully programmable 100-spin coherent Ising machine with all-to-all connections. *Science* **354**, 614–617 (2016).
82. Roques-Carmes, C. et al. Heuristic recurrent algorithms for photonic Ising machines. *Nat. Commun.* **11**, 249 (2020).
83. Tang, P. T. P., Lin, T.-H. & Davies, M. Sparse coding by spiking neural networks: convergence theory and computational results. Preprint at https://arxiv.org/abs/1705.05475 (2017).
84. Davies, M. Benchmarks for progress in neuromorphic computing. *Nat. Mach. Intell.* **1**, 386–388 (2019).
85. Maass, W. Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* **10**, 1659–1671 (1997).
86. Peng, H. et al. Temporal information processing with an integrated laser neuron. *IEEE J. Sel. Top. Quantum Electron.* **26**, 5100209 (2020).
87. Robertson, J., Hejda, M., Bueno, J. & Hurtado, A. Ultrafast optical integration and pattern classification for neuromorphic photonics based on spiking VCSEL neurons. *Sci. Rep.* **10**, 6098 (2020).
88. Chakraborty, I., Saha, G., Sengupta, A. & Roy, K. Toward fast neural computing using all-photonic phase change spiking neurons. *Sci. Rep.* **8**, 12980 (2018).
89. Fok, M. P., Tian, Y., Rosenbluth, D. & Prucnal, P. R. Pulse lead/lag timing detection for adaptive feedback and control based on optical spike-timing-dependent plasticity. *Opt. Lett.* **38**, 419–421 (2013).
90. Toole, R. et al. Photonic implementation of spike-timing-dependent plasticity and learning algorithms of biological neural systems. *J. Lightwave Technol.* **34**, 470–476 (2016).
91. Xiang, S. et al. STDP-based unsupervised spike pattern learning in a photonic spiking neural network with VCSELs and VCSOAs. *IEEE J. Sel. Top. Quantum Electron.* **25**, 1700109 (2019).
92. Larger, L. et al. Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing. *Opt. Express* **20**, 3241–3249 (2012).
93. Paquot, Y. et al. Optoelectronic reservoir computing. *Sci. Rep.* **2**, 287 (2012).
94. Duport, F., Schneider, B., Smerieri, A., Haelterman, M. & Massar, S. All-optical reservoir computing. *Opt. Express* **20**, 22783–22795 (2012).
95. Brunner, D., Soriano, M. C., Mirasso, C. R. & Fischer, I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 1364 (2013).
96. Vandoorne, K. et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **5**, 3541 (2014).
97. Brunner, D. et al. Tutorial: Photonic neural networks in delay systems. *J. Appl. Phys.* **124**, 152004 (2018).
98. Brunner, D., Soriano, M. C. & der Sande, G. V. *Photonic Reservoir Computing* (De Gruyter, 2019).
99. Antonik, P., Marsal, N., Brunner, D. & Rontani, D. Human action recognition with a large-scale brain-inspired photonic computer. *Nat. Mach. Intell.* **1**, 530–537 (2019).
100. Sun, C. et al. Single-chip microprocessor that communicates directly using light. *Nature* **528**, 534–538 (2015).

101. Stojanovic, V. et al. Monolithic silicon-photonic platforms in state-of-the-art CMOS SOI processes [Invited]. *Opt. Express* **26**, 13106 (2018).
102. Jha, A. et al. Lateral bipolar junction transistor on a silicon photonics platform. *Opt. Express* **28**, 11692–11704 (2020).
103. Giewont, K. et al. 300-mm monolithic silicon photonics foundry technology. *IEEE J. Sel. Top. Quantum Electron.* **25**, 8200611 (2019).
104. Zhou, Z., Yin, B. & Michel, J. On-chip light sources for silicon photonics. *Light Sci. Appl.* **4**, e358 (2015).
105. Song, B., Stagarescu, C., Ristic, S., Behfar, A. & Klamkin, J. 3d integrated hybrid silicon laser. *Opt. Express* **24**, 10435–10444 (2020).
106. Mack, M. et al. Luxtera's silicon photonics platform for transceiver manufacturing. In *2014 International Conference on Solid State Devices and Materials* 506–507 (Luxtera, Inc., 2014).
107. Billah, M. R. et al. Hybrid integration of silicon photonics circuits and inp lasers by photonic wire bonding. *Optica* **5**, 876–883 (2018).
108. Liang, D. & Bowers, J. E. Recent progress in lasers on silicon. *Nat. Photon.* **4**, 511–517 (2010).
109. Chen, S. et al. Electrically pumped continuous-wave III–V quantum dot lasers on silicon. *Nat. Photon.* **10**, 307–311 (2016).
110. Berggren, K. et al. Roadmap on emerging hardware and technology for machine learning. *Nanotechnology* **32**, 012002 (2021).
111. Ambrogio, S. et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67 (2018).
112. Li, C. et al. Long short-term memory networks in memristor crossbar arrays. *Nat. Mach. Intell.* **1**, 49–57 (2019).
113. Sze, V., Chen, Y., Yang, T. & Emer, J. S. Efficient processing of deep neural networks: a tutorial and survey. *Proc. IEEE* **105**, 2295–2329 (2017).
114. Cheng, Z. et al. Device-level photonic memories and logic applications using phase-change materials. *Adv. Mater.* **30**, 1802435 (2018).
115. Zhang, Y. et al. Broadband transparent optical phase change materials for high-performance nonvolatile photonics. *Nat. Commun.* **10**, 4279 (2019).
116. Cheng, Z., Ríos, C., Pernice, W. H. P., Wright, C. D. & Bhaskaran, H. On-chip photonic synapse. *Sci. Adv.* **3**, e1700160 (2017).
117. Bogaerts, W. et al. Silicon microring resonators. *Laser Photon. Rev.* **6**, 47–73 (2012).
118. Schrauwen, J., Van Thourhout, D. & Baets, R. Trimming of silicon ring resonator by electron beam induced compaction and strain. *Opt. Express* **16**, 3738 (2008).
119. Prorok, S., Petrov, A. Y., Eich, M., Luo, J. & Jen, A. K.-Y. Trimming of high-Q-factor silicon ring resonators by electron beam bleaching. *Opt. Lett.* **37**, 3114 (2012).
120. Milosevic, M. M. et al. Ion implantation in silicon for trimming the operating wavelength of ring resonators. *IEEE J. Sel. Top. Quantum Electron.* **24**, 8200107 (2018).
121. Harris, N. C. et al. Linear programmable nanophotonic processors. *Optica* **5**, 1623–1631 (2018).
122. Perez, D., Gasulla, I., Mahapatra, P. D. & Capmany, J. Principles, fundamentals, and applications of programmable integrated photonics. *Adv. Opt. Photon.* **12**, 709–786 (2020).
123. Gaeta, A. L., Lipson, M. & Kippenberg, T. J. Photonic-chip-based frequency combs. *Nat. Photon.* **13**, 158–169 (2019).
124. Kippenberg, T. J., Holzwarth, R. & Diddams, S. A. Microresonator-based optical frequency combs. *Science* **332**, 555–559 (2011).
125. Del'Haye, P. et al. Optical frequency comb generation from a monolithic microresonator. *Nature* **450**, 1214–1217 (2007).
126. Turner, E. H. High-frequency electro-optic coefficients of lithium niobate. *Appl. Phys. Lett.* **8**, 303–304 (1966).
127. Wang, C., Zhang, M., Stern, B., Lipson, M. & Loncar, M. Nanophotonic lithium niobate electro-optic modulators. *Opt. Express* **26**, 1547–1555 (2018).
128. Mercante, A. J. et al. 110 GHz CMOS compatible thin film $LiNbO_3$ modulator on silicon. *Opt. Express* **24**, 15590–15595 (2016).
129. Wang, C. et al. Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* **562**, 101–104 (2018).
130. Sun, J. et al. A 128 Gb/s PAM4 silicon microring modulator with integrated thermo-optic resonance tuning. *J. Lightwave Technol.* **37**, 110–115 (2019).
131. Patel, D., Samani, A., Veerasubramanian, V., Ghosh, S. & Plant, D. V. Silicon photonic segmented modulator-based electro-optic dac for 100 gb/s pam-4 generation. *IEEE Photon. Technol. Lett.* **27**, 2433–2436 (2015).
132. Meng, J., Miscuglio, M., George, J. K., Babakhani, A. & Sorger, V. J. Electronic bottleneck suppression in next-generation networks with integrated photonic digital-to-analog converters. *Adv. Photon. Res.* https://doi.org/10.1002/adpr.202000033 (2021).
133. Gelens, L et al. Excitability in semiconductor microring lasers: experimental and theoretical pulse characterization. *Phys. Rev. A* **82**, 063841 (2010).
134. Beri, S. et al. Excitability in optical systems close to Z_2-symmetry. *Phys. Lett. A* **374**, 739–743 (2010).
135. Bogaerts, W. & Rahim, A. Programmable photonics: an opportunity for an accessible large-volume PIC ecosystem. *IEEE J. Sel. Top. Quantum Electron.* **26**, 8302517 (2020).

Acknowledgements

B.J.S. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). T.F.d.L. and P.R.P. acknowledge support from the Office of Naval Research (ONR), Defense Advanced Research Projects Agency (DARPA) and National Science Foundation (NSF). We thank J. Shainline, P. Kuo and N. Sanford for editorial contributions.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to B.J.S. or A.N.T.

Peer review information *Nature Photonics* thanks Yichen Shen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.