

Roadmap on material-function mapping for photonic-electronic hybrid neural networks

EP

Cite as: APL Mater. 7, 100903 (2019); <https://doi.org/10.1063/1.5109689>

Submitted: 10 May 2019 . Accepted: 19 September 2019 . Published Online: 10 October 2019

Mario Miscuglio , Gina C. Adam , Duygu Kuzum , and Volker J. Sorger 

COLLECTIONS

Paper published as part of the special topic on [Emerging Materials in Neuromorphic Computing](#)

EP This paper was selected as an Editor's Pick



View Online



Export Citation



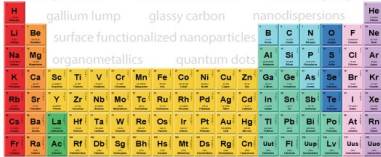
CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Fourier computed tomographic imaging of two dimensional fluorescent objects](#)APL Photonics 4, 106102 (2019); <https://doi.org/10.1063/1.5100525>[ITO-based electro-absorption modulator for photonic neural activation function](#)APL Materials 7, 081112 (2019); <https://doi.org/10.1063/1.5109039>[Hybrid plasmonic metasurfaces](#)Journal of Applied Physics 126, 140901 (2019); <https://doi.org/10.1063/1.5116885>



THE ADVANCED MATERIALS MANUFACTURER®



additive manufacturing epitaxial crystal growth cerium oxide polishing powder silver nanoparticles sputtering targets III-IV semiconductors CVD precursors europium phosphors

deposition slugs OLED Lighting spintronics solar energy osmium nanoribbons thin films chalcogenides AuNPs GDC Li-ion battery electrolytes 99.999% ruthenium spheres

endohedral fullerenes copper nanoparticles diamond micropowder CIGS MBE grade materials palladium catalysts flexible electronics beta-barium borate borosilicate glass dysprosium pellets YBCO pyrolytic graphite 3d graphene foam indium tin oxide mesoporous silica raman substrates sapphire windows tungsten carbide InGaAs barium fluoride carbon nanotubes lithium niobate scandium powder

gallium lump glassy carbon nanodispersions InAs wafers laser crystals ultra high purity materials MOFs surface functionalized nanoparticles organometallics quantum dot rare earth metals photovoltaics refractory metals MOCVD superconductors transparent ceramics ultra high purity silicon

American Elements opens up a world of possibilities so you can **Now Invent!**

Over 15,000 certified high purity laboratory chemicals, metals, & advanced materials and a state-of-the-art Research Center. Printable GHS-compliant Safety Data Sheets. Thousands of new products. And much more. All on a secure multi-language "Mobile Responsive" platform.

perovskite crystals yttrium iron garnet alternative energy h-BN gold nanocubes graphene oxide macromolecules photonics rhodium sponge fiber optics beamsplitters infrared dyes zeolites fused quartz metallocenes platinum ink buckyballs Ti-6Al-4V

Now Invent.™
The Next Generation of Material Science Catalogs

www.americanelements.com



Roadmap on material-function mapping for photonic-electronic hybrid neural networks

Cite as: APL Mater. 7, 100903 (2019); doi: 10.1063/1.5109689

Submitted: 10 May 2019 • Accepted: 19 September 2019 •

Published Online: 10 October 2019



Mario Miscuglio,¹  Gina C. Adam,¹  Duygu Kuzum,²  and Volker J. Sorger^{1,a)} 

AFFILIATIONS

¹George Washington University, 800 22nd Street NW, Washington, DC 20052, USA

²University of California, San Diego, La Jolla, California 92093, USA

Note: This paper is part of the Special Topic on Emerging Materials in Neuromorphic Computing.

^{a)}Email: sorger@gwu.edu

ABSTRACT

The state-of-the-art hardware in artificial neural networks is still affected by the same capacitive challenges known from electronic integrated circuits. Unlike other emerging electronic technologies, photonics provides low-delay interconnectivity suitable for node-distributed non-von Neumann architectures, relying on dense node-to-node communication. Here, we provide a roadmap to pave the way for emerging hybridized photonic-electronic neural networks by taking a detailed look into a single node perceptron. We discuss how it can be realized in hybrid photonic-electronic heterogeneous technologies. Furthermore, we assess that electro-optic devices based on phase change or strong carrier dispersive effects could provide a viable path for both the perceptron “weights” and the nonlinear activation function in trained neural networks, while simultaneously being foundry process-near materials. This study also assesses the advantages of using nonlinear optical materials as efficient and instantaneous activation functions. We finally identify several challenges that, if solved, could accelerate the adoption of such heterogeneous integration strategies of emerging memory materials into integrated photonics platforms for near real-time responsive neural networks.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5109689>

I. INTRODUCTION

While the scientific community still lacks a full understanding of the operation of the human brain, we yet can draw some parallels to compute-systems with respect to operating efficiency. Machine learning (ML) tasks performed by neural networks (NNs) can be used for such computer-vs-brain comparison since both are examples of nonlinear (NL) hypothesis systems that can be trained to classify patterns. Interestingly, one of the fastest computers¹ is only able to simulate ~1% of human brain activity in requiring about one hour and demanding massive compute infrastructure overheads (e.g., 82 000 processors and 1.73×10^9 virtual nerve cells connected by 10.4×10^{12} synapses) consuming about 10 MW, while the human brain operates on tens of Watts.²

In the aim to compensate this disparity and to match some of the brain’s computational and energy efficiency, the recent efforts to develop non-von Neumann neuromorphic hardware are capable

of efficiently implementing artificial NN operations, in particular, vector matrix multiplication (VMM) and backpropagation.^{3–5}

Conventional (von Neumann) compute architectures utilize centralized processors and rely on logic technologies, while executing stored programs sequentially. On the other hand, emerging non-von Neumann systems are inherently decentralized and, therefore, require significant communication between these many nodes of the network, yet process information in parallel and are trainable for applications such as in machine learning.

Electronic implementations of such brain-mimicked systems show significant 3–4 orders of magnitude^{6,7} energy-per-compute reductions defined by the underlying mathematical interconnection function of NNs, namely, multiply-and-accumulate (MAC). Thus, the efficiency and performance in the number of processed MAC’s per unit time and unit energy (i.e., MAC/s and MAC/J) are two key metrics to be considered when new hardware technology and emerging materials are explored for NN technology.

While electronic NN-based processors have shown improvement with respect to energy efficiency, the time delay to obtain a classification result from the NN (e.g., of an inference ML task) is not improved as compared to regular von Neumann systems due to the high-RC delay of electronic circuits.

In this work, we consider the NN being trained offline and hence only discuss the performance under inference operation. This approach, yet, requires mapping the ML-task to the NN hardware, which is not straightforward for photonics networks at the time this paper is written. However, initial approaches of modifying graph-simulators (e.g., TensorFlow) to include physical effects from the device layer (e.g., impact of noise, cascability) show higher inference accuracy using electro-optic (EO) hybrid approaches for NNs.

Emerging memory technologies, such as phase-change memory⁸ (PCM), conductive bridge random access memory⁹ (CBRAM), and resistive switching memory¹⁰ (RRAM), have been demonstrated

to be compelling candidates as synaptic devices for weight storage and matrix vector multiplication (VMM) in purely electronic neuromorphic circuitry. These devices are characterized by exceptional characteristics such as high footprint scalability, multibit storage capability, and long retention-time nonvolatility as well as an overall higher technological maturity compared to integrated photonics. Therefore, few efforts^{11–15} have targeted integrating non-volatile materials with waveguides, in order to explore novel ways of tuning their refractive index, either plasmonically or through phase changes in the crystallinity of the material. From this perspective, here we argue that integrated photonics and memristive hybrid systems (Fig. 1) could present significant improvements to the established digital implementations of VMM-tasks using graphic and tensor process units (GPU and TPU). However, their usage aiming to replace the current technology is still hindered by the above-mentioned hurdles and the state of the community of integrating complementary metal oxide semiconductor

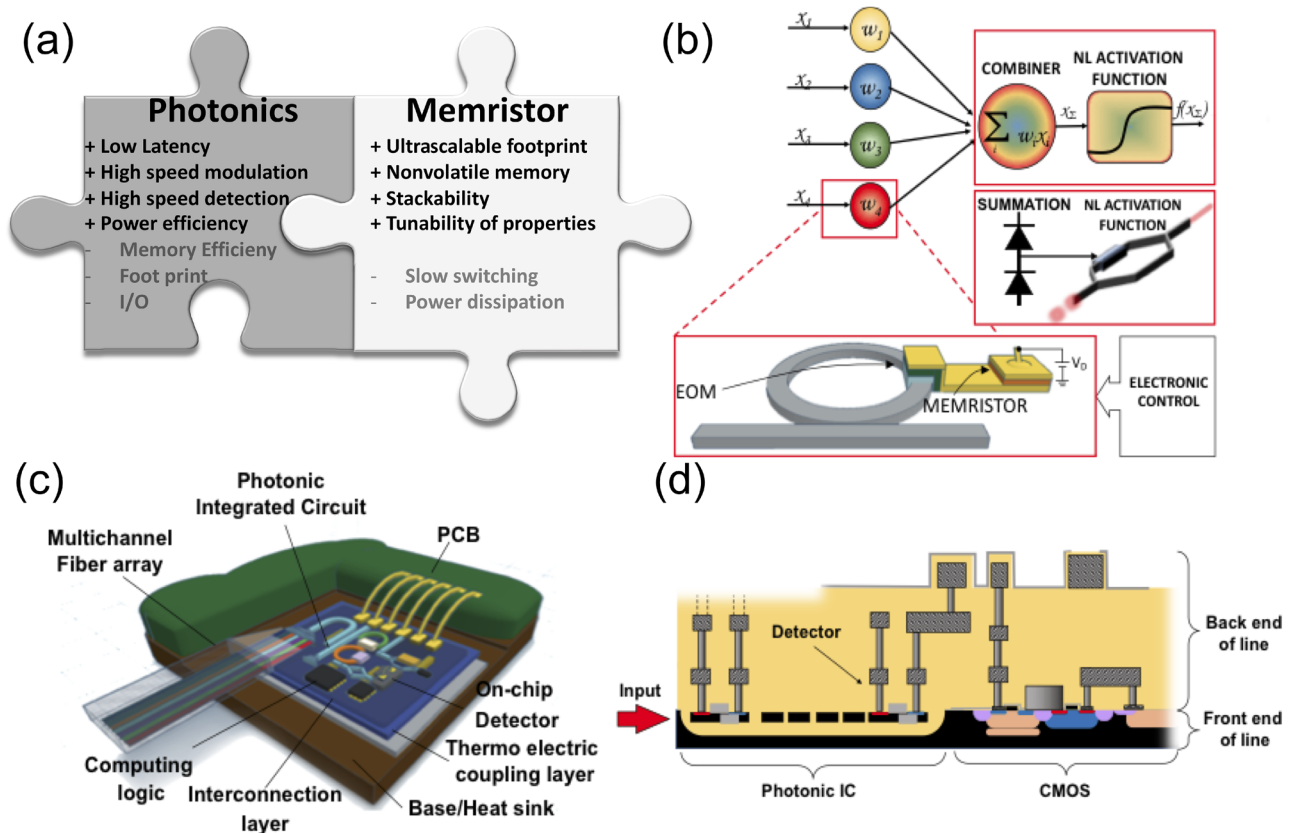


FIG. 1. Hybrid memristive photonic systems toward building dual-technology neural network (NN) architectures. (a) Specific pros and cons which characterize the two technologies, integrated photonics (dark) and memristor (light). (b) The perceptron model representing a single node of a NN can be mapped onto photonic hardware components. Thus, once the NN is trained and the weights are set in the electro-optic weighting modulators, the delay inference for the inference task is given by the photons time-of-flight through the photonic integrated circuit (PIC), which can be taken to be “real-time” (i.e., 1–10 ps) compared to electronic NN solutions. However, the opto-electronic weights require a constant voltage bias taking a toll on the power consumption. Thus, real-time fast and power efficient photonic NNs can be designed by integrating nonvolatile memory elements near the photonic weight components. This is synergistic to offline trained “weights,” which are updated infrequently, if ever. (c) Schematic of a back and front end-of-line of hybrid integration for nonvolatile weights. In addition to technological performance gains, the material compatibilities with fabrication processes of foundries need to be considered as well. (d) Schematic of a purely illustrative monolithic silicon or Si_3N_4 photonic integration, which comprises planar integration of Si photonics with complementary metal oxide semiconductor (CMOS) circuits on the same layer.

(CMOS) compatible phase change materials,¹⁶ such as germanium-antimony-tellurides (GST) (Ge₂Sb₂Te₅), with Photonic Integrated Circuit (PIC) is relatively at its infancy.

Nevertheless, the integration of the two technologies seems to be an interesting direction to pursuit to complement their strengths [Fig. 1(a)]. Yet, if successful, such hybrid integration would be particularly appealing since it could enable retention of the optical information, digitalization of the output, and introduction of nonlinearities at the same time, while building on the existing process know-how and made capital investments (i.e., material-foundry compatibility). The underlying mechanism for achieving weighting functionality in integrated photonics using memristors relies on specific active materials^{12,14,16–20} either alone (weighting + storing) or in combination with electro-optic modulators (storing), aiming to alter their optical response in a nonvolatile reversible way, by means of an activation potential (i.e., thermal, electrical or optical).

With the aforementioned electronic-NN based ML successes, the question becomes why perform NNs in the optical domain? Before we answer this, one can more generally ask what the rationale might be for performing information processing in optics and in integrated photonics? Here, we argue that the main reason for photonic NN hardware is found in the short delay,²¹ which could be important for applications such as in ranging, synthetic aperture radar, automated target recognition, or nonlinear predictive control²² (Table I). Once the dot-product, providing the “weights” of the photonic perceptron, is set (i.e., trained offline), the entire system only depends on the time-of-flight of the photon through which the NN plus the response time of the electro-optic components of the perceptron;^{5,23–25} here, several implementation options exist, and for the discussion of this roadmap, we focus on the dot-product photonic weights and the nonlinear activation function (NLAF). That is, the entire NNs’ delay (once trained) can be hundreds of picoseconds short using integrated photonics and high-speed (tens of gigahertz) photodetectors. In addition to delay, there are other benefits using photonics for NNs or well-established information processing paradigms such as parallelism from Wavelength Division Multiplexing (WDM) and vanishing capacitive “wire” delay (Table I). However, challenges exist too such as in function-per-wafer-footprint, electro-optic (EO) conversion efficiencies, packaging, and—most relevant for this roadmap paper—efficiency and number of states of photonic memories, whose development is still in its beginning.

It is this very reason that we turn our attention to a hybridization strategy by combining the best-of-both-worlds [Fig. 1(a)]; that is, using the nonvolatility from electronics yet the interconnectivity from optics combined with the compactness from integrated photonics which also aims to leverage semiconductor process economy of scale, through foundry services such as AIM, IMEC, or IME.

While a number of both fundamental and practical strengths exist, there are realistic “pain points” as well that need to be overcome in the long term in order for photonics (i.e., integrated optics) to become competitive. However, we note also that details on hardware competitiveness depend strongly on the applications sought after.

While volatile memory options in photonics could be implemented for the weighting function using, for example, electro-optic modulators,^{26–33} their high optical material index drifts, due to thermal and electrical noise, are suboptimal for the typical longevity of NN weights while performing inference of unseen data, which change only when a new training set is obtained (if ever), which can be particularly time consuming due to the enormous amount of training data employed to ensure correct training in a high dimensionality problem, namely, the “curse-of-dimensionality.”³⁴ Therefore, we firmly believe that the community should explore emerging materials that feature favorable performances for photonic, electro-optic material performance while keeping chip integration synergies in mind, as discussed below.

Neuromorphic photonics and memristor-based systems provide solutions that can deliver particularly high performances in terms of either throughput or energy efficiency by comparison with the current purely CMOS based execution.^{25,35} The main advantage of exploiting PIC for NN tasks is intrinsic to the wave-nature of the signal that carries information [Fig. 1(a)]. The fundamental operation in NN-tasks are multiplications and accumulations (MAC), which can be performed without any additional energy because a MAC is achieved by simple interference of phase-shifted electromagnetic waves traveling within waveguides.^{5,24,25} Moreover, integrated photonics is naturally predisposed to ultrahigh speed characterized by short latency due to (a) the time-of-flight of the photon in the chip (few picoseconds for large chips) and (b) to a considerable extent by its electro-optic components (few tens of picoseconds), such as detectors and modulators. Furthermore, exploiting Wavelength Division Multiplexing (WDM),⁵ PICs can perform

TABLE I. Rationale to perform information processing such as exploited in artificial neural networks for machine-learning tasks in the optical domain.

Rational for photonic information processing	
Strengths	Pain points
Nanosecond delay → “real time”	Efficient memory
Bosonification → parallelism (e.g., WDM)	functionality
Low loss communication → “wire” is free	(low, IL, efficiency, multistate)
→ distributed non-von-Neumann architectures	Footprint
“One-shot” execution (noniterative)	E-O conversion
Fabrication in progress → aim photonics, IMEC, IME	Packaging/alignment

selective weighting operation on multiple inputs mapped on different wavelengths, using the same physical channel, supporting higher parallelism and remarkable throughput.^{5,36} Furthermore, engineering light-matter interactions enable advances in efficient modulators, which can realize hundreds of attojoule-per-operation energy dissipation.^{37–39}

Nevertheless, the advancement of integrated photonics in neural networks is hindered by major challenges; beside the larger footprint with respect to microelectronics, which could be neglected considering the higher throughput and efficiency, the advancement of PICs at a larger scale is primarily held back by difficulties in the integration with a CMOS interface, local laser sources, and packaging (Table I). This is because the materials employed require dedicated industrial process recipes that have not reached enough maturity yet to provide cost effective, consistent, or scalable results. Just to clarify this point further: PICs-based components and sub-systems are commercially sold in the millions, namely, for data-centers, but using PICs for NN is a technology direction which is still at its infancy despite great progress having been made.^{5,7,22,23,25} It is also worth mentioning that, except for recent efforts,²³ NNs in integrated photonics are still lacking a straightforward implementation of an optical NLAF to mimic the action potential firing in the neuron or an analog tuning such as favored by the computer science community (e.g., rectifying linear unit ReLU). Also, the absence of a straightforward nonvolatile memory realized in photonics that can be written, erased, and read optically still limits the realization of all-photonics chip-scale information processing for NN tasks, yet initial design concepts exist.²⁴ What we find is that the required power consumption and bandwidth could both be improved when nonvolatile memory functions could be integrated directly in the optic domain, which is particularly advantageous in trained networks when the weights are fixed (or changing seldom).

Our aim in this work is to explore a roadmap of nonvolatile materials and approaches that can be integrated with photonics to enable dot-product multiplications and thresholding in the optical domain. Here, we review, discuss, and project challenges and opportunities for device type and material choices for these memristor-based electronic-photonics hybrid NNs from the perspective of key metrics such as power dissipation, electrical vs optical readout, multistate tunability to mention a few, thus laying a corner stone toward establishing a material roadmap for photonic NNs. We discuss the advantages and issues related to a hybrid integration of photonics and memristors in a complementary way, highlighting material systems, which are jointly shared between the two technologies and that could allow a seamless integration. CMOS compatibility and large-scale integration pose additional material-related constraints, limiting the choice of materials available for electrodes, device layers, and isolation.

II. INTEGRATED PHOTONICS FOR NEUROMORPHICS

For hybrid memristor-photonics NN systems, the fast and efficient response of the transfer function of an electro-optic modulator is mostly suitable for mimicking the NLAF of the neuron rather than the weighting since the network is mostly offline trained and performing only the inference tasks.^{5,23,40} On the contrary, the ability of the memory element to retain information for a comparable

long time can be exploited for the purpose of efficiently storing the weights and modulating the signal traveling in the waveguide accordingly. In this section (Sec. II B), we discuss recent work on integrated photonics platform, optical modulators, and their integration with CMOS electronics, before we further discuss the key metrics of the state-of-the-art modulators.

A. Passive photonic neural network interconnectivity: Waveguide platform options

For PIC operating at telecommunication wavelengths (1550 nm), the main platform for planar light-wave circuit is Silicon on Insulator (SOI). The crystalline silicon layer atop the insulator is used to create optical waveguides (through optical index contrasts) and can be extended to include both passive and active devices used to deliver NN functionality. The SOI chip can be realized by either smart cut or separation by ion implantation of oxygen (SIMOX) processes. The buried insulator enables propagation and strong confinement of infrared light in the silicon layer on the basis of total internal reflection, with low propagation losses (<1 dB/cm)⁴¹ and small bending radii (<5 μm)⁴² enabling PIC with compact footprint. This SOI platform also features monolithic electro-optic modulators, i.e., without the addition of any other material, either by thermally or electrostatically changing silicon's optical refractive index.^{43–46} Silicon's optical (i.e., all-optical) nonlinearities arise at few tens of milliwatts, which is a value that could limit the depth (numbers of layers) of the NN, and contemporary sets the limits to the insertion losses (IL) admissible.

The integration with CMOS electronics for logic circuitry with SOI, while technologically feasible,⁴⁷ is challenging due to technical and economic mismatches; therefore, hybrid integration is usually required. Layer-stacking and integration with light sources could be feasible by SOI wafer bonding techniques such as coupling the evanescent optical mode of a III-V laser to SOI waveguide,^{48,49} high-performance quantum dot (QD) lasers monolithically grown on Si,^{50,51} or other electrically driven solution.⁵² While silicon's bandgap is transparent at telecommunication frequencies, there is no conceptual rule why PIC-based NN could not operate at visible or also at mid-IR wavelengths. In fact, there are convincing reasons to consider small wavelengths: (a) the smaller wavelength enables denser PICs, (b) the higher bandgap can deliver lower optical losses saving chip power consumption, and (c) extending the pump-power range before NL become parasitic,²⁰ thus enabling NN cascability. Such a visible-photonics platform material is Si_3N_4 , which presents a bandgap at higher energy (0.4 μm). Interestingly, the process recipes for silicon nitride are favorable for large-scale photonic networks; when deposited by LPCVD, Si_3N_4 is characterized by both high material stability and refractive index regularity, as well as nanoscale etch-resolution and lower surface roughness, which leads to reduced scattering and hence low optical losses per unit length (<10 dB/m) for a 0.5 μm bending radii, which is 10–100 $\times$ lower compared to SOI.⁵³ In contrast to SOI, due to the fundamental flexibility of the deposition techniques,⁵⁴ Si_3N_4 also allows for easy integration and 3D stacking flexibility with either SOI or CMOS platforms.⁵⁵ Regarding other material integration into this silicon nitride platform, several options have been successfully embedded such as a number of metals or colloidal quantum dots.⁵⁶ This platform also allows creating more complex photonic

structures such as distributed feedback reflectors (DFB)⁵⁷ and has shown to deliver PICs with multiple photonic layers. Regarding electro-optic (EO) phase tunability, silicon nitride is less responsive than even silicon and hence is not a suitable material for monolithic weights of photonic NNs. Nonetheless, given its promising passive properties, it appears an ideal material for heterogeneous integration in addition to co-integration with SOI and CMOS logic circuitry.

B. Electro-optic “weights” and nonlinear activation function: Materials for high-speed and efficient modulators

To modulate the light-wave information traveling in the waveguides, efficient electro-optic or absorptive modulators must be designed and integrated, either monolithically or heterogeneously in the SOI or Si₃N₄ platform. Modulators are active photonic components that induce an optical absorption (EA, electroabsorption) or change the optical path length or refractive index (EO, electrorefraction) of a material.³⁹ The choice of electro-optic modulation over electroabsorption and vice versa is merely related to the device type and its use in the network. Using either electrorefractive or electroabsorptive modulation ultimately translates into an intensity modulation. Phase modulators require interferometric scheme to induce an amplitude change, such as the Mach-Zehnder (MZ) configuration or Microring Modulators (MRMs), whereas absorption modulators achieve it in a linear waveguide geometry. MRMs and MZM operate at specific wavelength, while electro-absorption modulation is principally broadband. MRMs can be employed in architectures that exploit the available bandwidth using Wavelength Division Multiplexing (WDM), in which WDM inputs are weighed through tunable MicroRing resonators (MRRs) multiplexed on the same physical channel. In the end, a modulator is a multi-trade-off device, and a straightforward comparison without application context lacks relevance. For instance, the material system LiNbO₃ has shown lower power and high-speed performance^{26,65} but comes with very large footprints (~centimeter size) due to the lack of any resonance effect. The other extreme is, for example, a graphene-plasmon modulator^{37,66} deploying both optical and material resonance allowing for micrometer small devices, yet at a (much) higher per-unit length device insertion loss. In fact, we have explored such material-mode-performance options for a variety of EO-active materials (Si, ITO, Graphene, III-V, QD, TMDs) and multiple optical modes (photonic, plasmonic, hybrids) in Refs. 28 and 67.

Nevertheless, the performance of the electrorefractive/absorptive modulators depends on the physical modulation mechanism such as Pockels effect (LiNbO₃), Kerr effect,^{68,69} quantum confined stark effect⁷⁰ (Ge-Si/Ge), and free carrier modulation (Si⁴⁴ and ITO²⁸).

For neuromorphic applications, the optical modulator is a key component having a significant impact on the overall metric of the NN.⁷¹ Indeed EA/EO modulation can perform two of the three functionalities required in NNs rely on a high interconnectivity function between nodes; hence, scaling questions rooted in component “performance-per-overhead” should be considered. A NL with a limited modulation range can translate into a non-efficient type of NonLinear Activation Function (NLAf), which does not allow to discriminate the data, hindering the accuracy of the network in the classification task, something we have shown in a recent paper,²³

where we investigated the NN inference accuracy as a function of the actual EO modulator material, and the modulator and photodetector. Taken together, both interconnectivity density and modulation performance are key for the core of the perceptron mechanism. The modulator can also be used for online learning by actively adjusting the weights of a to-be-trained NN.^{11,51} That is, when the network is trained offline, the weighting can be realized completely passive without energy consumption using EO or EA modulators that are gated via a nonvolatile storage element or by photonic memories [Fig. 1(b)]. Whereas the NLAf would mostly still rely on EO or EA modulators in combination with photodiodes. As reported by Miller,³⁹ while providing a signal modulation depth (extinction ratio, ER, >3 dB) at a sustained speed (>25 Gbit/s), the target power consumption for modulators has to be within few femtojoule/MAC (or better) for global on-chip connections to become technologically competitive (i.e., BER < 10⁻⁸). Materials such as silicon,^{43,72} lithium niobate (LiNbO₃),⁷³ germanium,^{30,63} and hybrid III-V^{60,74} are regularly employed as active component in modulators and have showcased reliable and, to a certain extent, scalable results. However, their use in densely integrated photonic circuit is not a viable approach due to their vastly (~10⁶×) larger footprint compared with electronic switches (e.g., MOSFETs). Usually, for improving the extinction ratio while enabling relatively restrained footprint, these materials are embedded on resonant structures such as microring resonators, thus limiting the optical response. Considering Silicon as the dominant material in PIC-platform for NNs, the most obvious material choice for implementing modulators for weights and/or NL activation module would be silicon. Although, as displayed in Fig. 2, Si-modulators based on carrier injection display low dynamic modulation (i.e., low ER) and if doped are particularly lossy (high insertion losses, IL), and additionally if thermally driven modulators are slow (1–10 kHz). On the other hand, even though characterized by competitive performances, III-V modulators are difficult to integrate into SOI platform as well as with CMOS circuitry due to process and material incompatibility, and the wafer sources are costly (10s times compared to SOI). Instead, LiNbO₃ based modulators require a 1–20 cm-large device lengths for inducing a π -shift if CMOS-compatible voltage biases (1–2 V) are desired,²⁶ thus hindering integration density.

In the view of providing modest energy reductions when performing NN tasks compared to electrical approaches, while still preserving a modest footprint, several engineering design and material choices must be made. As a main point, evidently, the selection of the active material is strongly impacted by its ability for voltage-efficient optical index tuning for the NL activation function and the weights; ideally, the absolute difference in the transmission of the modulator is in its two states (ON/OFF), avoiding modulators with significant background loss, therefore not using those active materials which present high intrinsic losses in the OFF-state. Although this might not be sufficient to achieve a sufficiently high modulation performance and low energy-per-compute for surpassing electronic efficiency, an enhanced modulation can be reached by using either quantum-confined system or subdiffraction limited plasmonic structures, either with monolithic or heterogeneous integration of other materials or structures. This results in a low-energy consumption of few attojoules-per-bit,^{37,39} up to the highest speeds,^{27,33,43,75} which corresponds to a compounded merit improvement of 10⁵ times compared to electronic switches. Hence, many

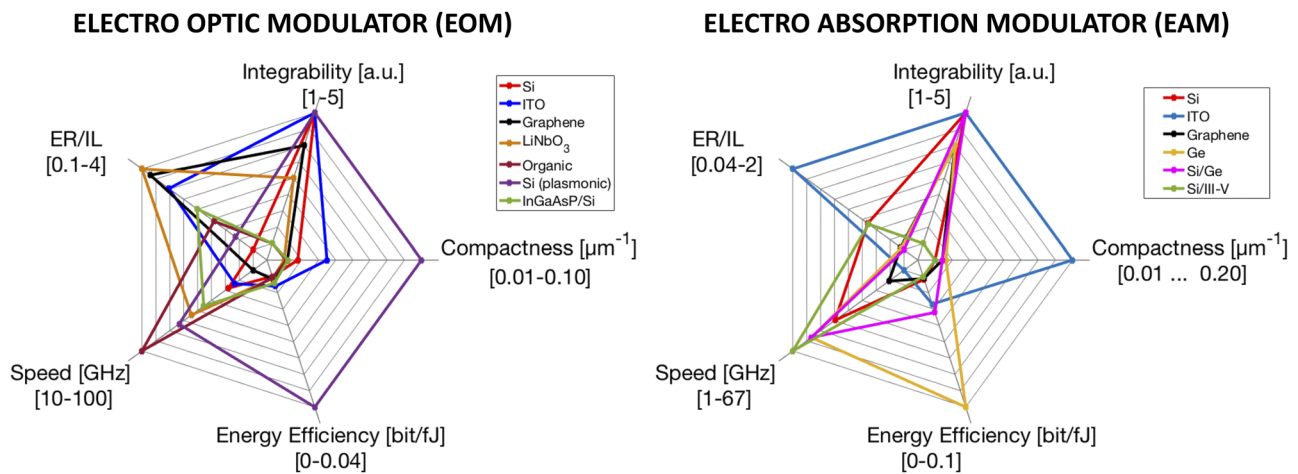


FIG. 2. Electro-optic modulators can be used not only as efficient tunable “weights” of the perceptron but also as the nonlinear (NL) activation function “threshold.” For the latter, the “weighted-addition” signal can be provided by a photodetector, whose photogenerated voltage becomes the input “gate” bias of the modulator. Refer to Refs. 7, 22, and 23 for details on possible perceptron architectures or to Fig. 1(b). A comparison of key metrics for electro-optic and absorption modulators for different material systems. The figure of merit considered is the integrability (qualitative number 1 low to 5 high), compactness (1/Length expressed in μm), energy efficiency (bit/fJ), speed (GHz), and modulation dynamics vs losses (ER/IL refers to the ratio between extinction ratio and insertion loss). The values of the key metrics increase moving further away from the center of the radar chart. Silicon,⁵⁸ Indium Tin Oxide (ITO), graphene,²⁷ LiNbO₃,²⁶ hybrid-organic,³³ Si-plasmonic,²⁹ InGaAsP/Si⁶⁰ electro-optic modulators are reported (left). Silicon, ITO,³² graphene,⁶² germanium,⁶³ Si/germanium,⁶⁴ III-V/Si⁶¹ electro absorption modulators (right). Integrability is an arbitrary value (1–5), given based on the “standardization” of the fabrication process as well as compatibility of the material with CMOS technology. CMOS compatible processes such as Si and (now recently also) ITO based modulators received 5 points, foundry-close materials (Ge, Si/Ge) and processes received 4 points, materials that show R&D-level evidences defined as successful material co-integration into pretaped-out chips, such as integration with SOI, received 3 points (graphene, LiNbO₃), 2 points are given to materials that are still relegated to the research level and 1 star to materials that are not compatible with standard CMOS processes (InP and polymers) and require complex integration or limited by small scale integration capability.

research groups have strived for engineering on-chip modulators beyond the solutions offered by heterogeneous integrated photonic-foundries to date. Recent developments of monolithically and CMOS compatible integrated emerging EO or EA materials such as Indium Tin Oxide (ITO),^{28,32} graphene,^{27,31,76,77} quantum-confined structures,⁷⁸ and TMDs into Si-photonics with specific device configuration aiming to enhance mode overlap allowed energy efficient,^{79,80} compact silicon photonic based modulators. The important performance metric for EOMs include high ER (>3 dB), low IL (<1 dB), modulation speed (>25 GHz), low energy consumption per bit (<10 fJ/bit), and compact footprint area (possibly 3D volume). ITO can be particularly suitable for heterogeneous integration in Si exhibiting formidable electro-optic effect characterized by unity-order index change at telecommunication frequencies. ITO carrier-based electroabsorption models, which are implemented via capacitive gating,³² have shown a sharp dynamic range, compact footprint, and potential for gigahertz-fast modulation.

In recent works, Amin *et al.*²⁸ demonstrated a monolithically SOI-integrated ITO electro-optic modulator based on a Mach Zehnder interferometer (MZI) featuring a high-performance half-wave voltage and active device length product of $V\pi L = 0.52 \text{ V mm}$. This device demonstrates a unity-strong index change in the active ITO layer enabling a 30 μm -short π -phase shifter, while purposefully operating ITO in the index-dominated region away from the epsilon-near-zero (ENZ) point, hence reducing optical losses. Moreover, some major electronics manufacturers have recently declared to integrate ITO into their foundry processes; hence, it can now

be termed a CMOS compatible material as it can be monolithically integrated in the photonic frameworks and directly interfaced with on-chip logic circuitry and memories (Fig. 2).

ITO has the advantage over silicon of higher tunable absorption [in electro absorption modulation (EAM) scheme in proximity of ENZ^{32,81}] and unity variation of the refractive index [in Electro-optic modulator (EM) devices^{28,82} away from ENZ]. The inherent low tunability of Si under electrical bias causes inadequate performances with respect to the ITO modulator given a fixed modulator length.⁸³ Moreover, by engineering the ITO process, one can tailor its optical response, thus increasing the modulation dynamic and lowering insertion losses, without necessitating plasmonic or dielectric cavities or ring resonators.

Incidentally, GST, discussed in Sec. III, is a second candidate that will enter foundry processes soon, while graphene has yet to receive such “permission” due to limitation in the substrates used for the growth and the inconvenience of the required transfer. In addition to ITO-based modulators,^{28,32,77} other active optoelectronic components are also demonstrated; Kim and Kim,⁵⁹ for instance, experimentally demonstrated efficient and potentially high-speed directional coupler-based on ITO that can also be employed as intensity modulation for NN weighting. Still at a research level due to their nonscalable integration, other emerging materials such as graphene,^{84,85} transition metal dichalcogenide flakes,^{85,86} organic³³ modulators, and detectors, have been reported, demonstrating, in some cases, striking performances, yet process maturity is far-away from foundry standards. As aforementioned, integrated photonics

necessitate materials could provide memory functionality. Recent work on different device configurations has showcased capabilities of optically write/read/reset functionality, as described in Sec. III.

Next, we proceed in depicting the hybridization of photonic NN, discussing details of novel memory devices embedded in photonic framework, used for storage and weighting functionalities.

III. NOVEL MEMORY DEVICES FOR PHOTONIC INTEGRATION

In a hybrid photonic system, memristor devices would be used mainly as weights, which do not require high-frequency updating. Therefore, the efficiency of programming them during training (either directly on the photonic hardware or uploading offline-trained weights) is of importance in terms of chip power budget and compatibility of available voltage ranges (i.e., signal dynamic range). Recent studies^{87–89} have highlighted the difficulties of training large-scale memristor networks acting as weights for neuromorphic computing. An alternative approach is to alleviate these challenges and to use both long- and short-term memory elements to implement these different temporal weight classes.

A. Efficient and long-term weights: Heterogeneous integration of nonvolatile memory

There are several long-term nonvolatile memory technologies that show promise for integration with photonics systems (Fig. 3). Given their technological maturity, flash memories (i.e., floating gate transistors) have the potential for gating the above-discussed modulators that perform the weighting or NL thresholding [Fig. 3(a)]. However, in a typical configuration for regular memory

applications, these transistors use polysilicon for the floating gate. For photonic integration, however, this option is not very desirable since the low-number of trapped charges can only impose slight changes in the modulators’ refractive index (low dynamic range). Weight banks or NL activation functions based on such a technology option will have limited neuron bit-density limiting inference accuracy and NN cascability. This essentially rules out online-learning options, where these modulators had to be adjusted according to a gradient-descent algorithm via back-propagation, which relies on differentiability of the modulators transfer function. For the steeper dynamic range, it is possible to achieve higher bit-density and consequently higher training efficiency, i.e., lower power. Also, the additional electrical capacitance from poly-depletion increases both the devices’ RC-delay and the energy-per-MAC similar to arguments in transistor technologies and hence high-k dielectrics and metal-based floating gates may be a more viable path. Recent work⁹⁰ proposed a floating gate utilizing a nonvolatile optical switch, where a graphene sheet is used as floating gate material⁹¹ with much larger refractive index variation³¹ [Fig. 3(a)]. However, electron-based memory storage can be problematic from the perspective of increased optical loss in semiconductor waveguides, reduced nonvolatility, and limited analog programmability. A more desirable approach could be integrating memories heterogeneously with photonic waveguide-modulators. Such micrometer-compact device-to-device integration was recently pointed out by Miller³⁹ due to improved capacitive loading; since every micrometer of metallic wire has a capacitance of 0.2 fF, just a few micrometer of wire would bring the power budget above 1 fJ-per-operation (i.e., bit or MAC) just for the wiring alone (excluding the functional devices) for a V_{DD} of just 1 V.

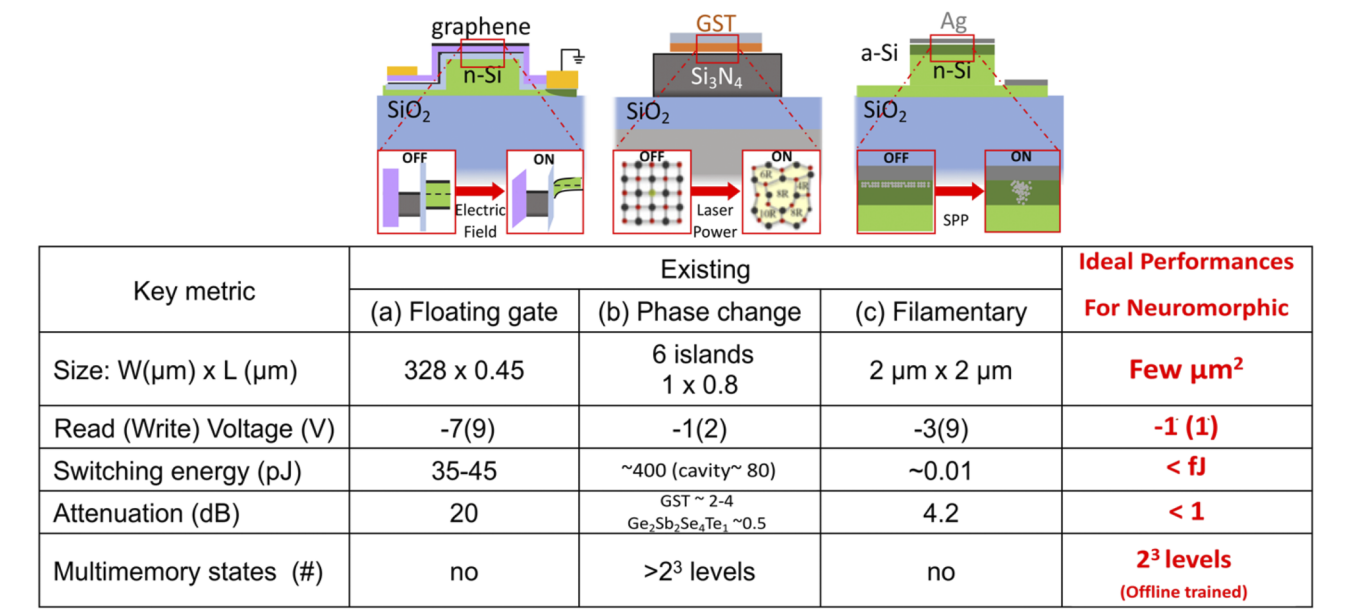


FIG. 3. Designs and key metrics of memory technologies monolithically integrated with waveguides. (a) Graphene-based floating gate,⁹⁰ (b) phase change memory synaptic device with optical readout⁹² (number of levels can exceed 3,⁹⁴ attenuation can be minimized using specific alloys,⁹⁵ and switching energy reduced when integrated in cavities⁹⁶), and (c) plasmonic memristor with silver filaments.⁹⁷

For such “tight” integration, phase change memory (PCM) materials could be a viable option. Chalcogenide phase change materials based on germanium-antimony-tellurides (GST-PCMs) displayed remarkable properties in nonvolatile memory technologies due to their high write and read speeds (ranging from few tens of nanoseconds to 1 ns¹²), high degree of integrability, CMOS compatibility, contained power consumption, long data retention, and multilevel storage capability. In photonics, the GST-PCM can significantly change the effective refractive index of the waveguide by local amorphization or crystallization, which produces unity-strong index changes (i.e., 10^3 – 10^4 times stronger than the free-carrier modulation of silicon, for example).

Furthermore, PCMs have a state retention time of (estimated) >10 years and analog programmability up to 3–4 bits during the “write” transition. Recent experimental demonstration⁹² showed a photonic synaptic behavior of a chalcogenide GST phase-change film integrated with a Si_3N_4 waveguide [Fig. 3(b)]. To achieve precise control of weight-programming using fixed-pulse characteristics, an innovative tapered waveguide structure with multiple discrete PCM islands was used, demonstrating 3 bit (8 level) operation using ~400 pJ for single pulse weighting. However, for such a power, if this GST-based synapse in a photonic multilayered perceptron framework are used for performing training, it would not provide any significant advantage with respect to the state-of-the-art electronic neuromorphic systems or hybrid analog Non-volatile memory (NVM)-based approaches,⁸⁷ thereby reducing the photonic advantages.

However, the short NN delay would still be the main rationale for photonic PCM-hybrid NN systems, when performing inference. To further clarify this concept, when the network is trained off-chip (i.e., in electronics, e.g., GPU), the weights are set and will (typically) not be updated often depending on the application (e.g., possibly weekly, monthly, yearly). On a device level, this means that the PCM-based memory must be written only once. This means that the optical phase-control of a portion of the PCM film must be altered only when the NN is set and should retain its phase throughout the inference process, in order to effortlessly modulate the quickly varying input signals, according to the value of the weights obtained by the off-chip training. In that instance, MAC operations are performed in a complete passive fashion with the obvious energy benefits.

On the other hand, if the network is performing training on-chip, the limiting factor would be the update speed of the weights, which need to be constantly and ideally rapidly updated. Large and continuous amount of data fed to such a NN will be used for updating the weights through algorithms based on backpropagation and gradient descent. If the weight functionality is implemented with current photonic memories based on PCM, the network would require a rather long learning time. Also, in terms of overall energy consumption when performing training, switching continuously the phase of the GST demands substantial amount of energy consumption (tens of picojoule per weight update), which scales superlinearly with the number of nodes and layers. Therefore, during the training process, having fast and energy-efficient switching materials is not just desirable but compulsory. Current PCM switching energy (tens of picojoule/bit) and delay (tens of nanoseconds) are at least 2 orders of magnitude larger than current EOM and would not provide an advantage compared to electronic

implementation when performing training on-chip. Replacement of GST with engineered phase-change materials that have low switching powers and fast response is therefore desirable from a material roadmap perspective.⁹³

Another intriguing alternative is plasmonic memristive devices that exploit the modulation of electron transmission in optical systems through filamentary switching [Fig. 3(c)]. Memristive devices based on metallic or highly reduced oxide filaments can be used for this purpose, for example. Plasmonic memristors based on silver filaments and integrated with silicon waveguides have shown a hysteretic behavior and optical readout functionality.⁹⁷ However, the poor ON/OFF signal ratio limits higher-bit densities (i.e., multilevel switching), but a lower power compared to PCMs highlights the potential for femtojoule operation. On the other hand, once the NN is trained, a power-optimized bit-density is about 3-bits only. In this context, photonics memory, in particular, those based on phase change materials and filamentary switching, can help reducing the overall computational cost and overall losses of a multilayer perceptron NN implemented in photonics when performing inference task, by adopting quantization schemes (lower number of bits up to 1), especially in regard of the weights quantization, which could lead toward photonic Quantized NN (QNN).^{98,99} Additionally, filamentary memristive devices can potentially be employed as electric perceptron mechanism triggered by a plasmonic response.¹⁰⁰

That is, the inference accuracy does only marginally improve for higher bit densities. This is in contrast to online training on the photonic hardware, where higher bit-density reduces the energy cost and training time during gradient descent back-propagation algorithm, while resulting in higher accuracy during inference operations.²⁵ Nonetheless, Li *et al.*⁹⁴ through opportunely shaped pulse, reported a 5 level GST cell, additionally quasicontinuous tuning of GST has been achieved¹⁴ when integrated on a ring resonator.

B. Neural network nonlinear thresholds and online training: Short-term memory options

Since memristors are challenged by device-to-device variability and require trains of tens of voltage pulses for precise tuning, short-term memory technologies are needed to speed-up both the training and to reduce the energy consumption for on-the-hardware training (offline training, i.e., using GPUs, and closed-loop transfer would be an alternative).⁸⁷

Capacitive-driven devices can be a useful short-term memory solution for integration with photonic platforms since footprint is not as significant of a concern as in ultrascaled purely electronic systems. Here, micron-scale capacitors can be fabricated in a stacked configuration atop waveguides, utilizing long-term memristor devices as photonic weights, for example [Fig. 1(b)]. Unlike capacitive-challenged arrays of electronic memories [either emerging PCMs or classical static random access memory (SRAMs)], EO components such as the NL thresholding modulators discussed above are capacitive stand-alone devices, just connected to their respective drivers (e.g., the summation providing photodetector). These modulator-activation functions can respond at <10s of picoseconds short delays thus do not slow down the (already fast) optical NN. If longer retention is needed, short-term memory elements can be used to bias the modulator's gate temporarily. Such a

complex cell could benefit from the advantages of each technology and minimize the drawbacks regarding training.

A novel technology for short-term memory, for instance, is the diffusive memristor;¹⁰¹ while designed as a selector device compatible with memristive crossbars, the diffusive memristor has the advantage of a metallic filament that can be programmed using low voltages and its memory-state dissolves rapidly (hundreds of microseconds to milliseconds) afterward forming nanoparticles. This behavior resembles the Ca^{2+} dynamics in a biological synapse thus providing the short-term memory capabilities required during training. Experiments in plasmonic metasurfaces using silver nanofilaments have shown short-term memory, and few volts switching voltage was found similar to that of a diffusive memristors.¹⁰²

The efficient diffusive memristor's dynamic can potentially mimic both short- and long-term plasticity of biological synapses when electrically driven; however, when triggered by plasmonic modes, this control might not be as straightforward as in the electronic counterpart; also integration of multilevel diffusive memristor in photonics has not been demonstrated yet. Ultimately, noble metals such as silver are well-known to exhibit plasmonic performances, but their use is restricted in a CMOS foundry as they are contaminants for circuitry, hindering the cointegration process.

C. Challenges in design of memristor devices for photonic integration

While the integration of nonvolatile memristive devices in photonic circuits would be an enabling step toward energy-efficient and NNNs capable of nanosecond-short inference tasks, here we address some of the challenges that should be considered in such a roadmap; a typical issue with novel material-chip integration is often a limited reproducibility for larger-scaled PICs.

Unlike NN approaches from the computer science community, which often follow a strategy of adding more neurons to increase inference accuracy and whose NNs approach millions of neurons, recent results on photonic neuromorphic show that (a) hundreds of neurons are sufficient to perform smaller inference tasks equally well to electronics,⁸ and (b) the inherent noise of the analog photonic system can be advantageous during training with respect to accuracy. More in detail, additional noise in the training phase might be beneficial.¹⁰³

A network trained with small amount of noise will be more robust and tolerant to noise during the inference task, and we argue that could potentially reduce the impact of the quantization error because of discrete weights. Nonetheless, reproducibility and reliability at the device-level is fundamental to ensure performance guarantees at the circuit and system levels during inference, and too high noise will, trivially, impact accuracy adversely. This problem is related to material engineering, interface control, and optimization of nanofabrication processes. Engineering the material stoichiometry to improve device performance is desirable for a broad range of materials of interest in both memristor and photonics fields. For example, ITO composition can be explored in a holistic fashion using reactive sputtering¹⁰⁴ to carefully control its electrical and optical properties. Similarly, phase-change materials can be tuned to lower the switching energy and improve reliability of amorphization/crystallization. For filamentary memristors, the issue of

filament robustness and controllability is driven also by the choice of materials; for example, the electrode material influences the filament shape based on the different free energy of oxide formation or the metal electromigration.¹⁰⁵ Such instabilities lead to noise of the NN. Interestingly, depending on the amount of noise,¹⁰⁶ training the NN with the actual system noise of these photonic analog networks results in higher inference accuracy than performing the training “signal-clean” digitally (i.e., GPUs). Indeed, highlighting material systems that are jointly shared between memory and photonic technologies could enable a seamless integration pathway (Fig. 4).

Here, CMOS compatibility and large-scale integration poses additional material-related constraints, limiting the choice of materials available for electrode, device layer, and isolation. For conductive bridge devices, short term transient memory effects on a metasurface can be obtained at voltages as low as 5 mV, yet the thermal noise floor of 26 meV at room temperature would dictate a signal-to-noise ratio (SNR) < 1 at the back-end photodetector of the photonic NN.⁸⁰ However, long-term memories require often high programming voltages, which reduce the endurance (number of memory's programmable cycles) leading to undesirable electric shorts as the failure mode, which could become extremely critical at high speed. In fact, if the endurance performance of a short-term memory element had a value of $< 10^{11}$ cycles and this memory would be used as the charge-storage that controls the NL threshold of the photonic NN, then this memory would fail after just 5 s when the inference data input is clocked at 20 GHz. This shows that some performance parameters do not translate well across these different applications.

However, for usage as NN “weights,” any state-of-the-art memory is already overperforming, given the infrequent updates (which, naturally, depend on the application). Oxide-based memristors have a more tunable filament at the expense of increased optical insertion loss. Devices based on electromigration are typically bipolar,

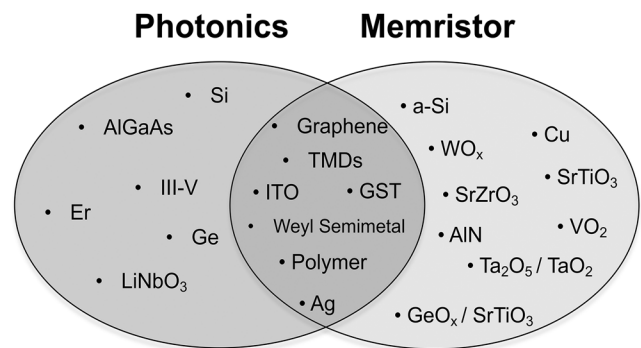


FIG. 4. Representative materials and material systems for photonics and memristors, highlighting the significant overlap. Here, we argue that for a multitude of reasons, synergistic opportunities exist when heterogeneously integration memristor materials to act as long and short-term photonic memory elements for neural network weights and nonlinear thresholding, respectively. Such synergies include process integration, reduced device-to-device capacitance, tailorable bit-density, and programmability (read/write speeds). Prominent options could be ITO and GST as active electro-optic materials for weighting and thresholding, which have both gained semiconductor foundry access recently.

TABLE II. Qualitative mapping of both the functions of the neural network (NN) and the mathematical model of the perceptron onto photonic-electronic hybrid neurons. The memory is most relevant in providing the charge to store the weights to bias the electro-optic modulators performing weighting. However, the same modulator can also be used as a nonlinear activation function.

Neural network function	Mathematical (perceptron) function	Hybrid photonic-electronic neuron			
		Photonic implementation ^a	Electronic memory		
			Retention time	Cycle endurance	Leakage current ^b
Weighting	Dot-product	{Electro-optic modulator or tunable ring-filters} plus memory, ^c or, nonvolatile material phase-transition Photodetector or Mx1 fan-in waveguide configuration (coherent system only)	Long-term (dep. on application)	>1–10 ³	Important
Summation	Summation			NA	
Nonlinear activation	Threshold	Electro-optic modulator (optional memory), or, all-optical (e.g., photonic bistability and phase change material) ^b	Short-term to no memory needed	>10 ¹⁸	Minor importance

^a Assumes analog amplitude signal control (perceptron). Other options such as spiking neurons could also be explored for higher-order neuron signal-shaping using a multitude of partial-differential equation.²²
^b Leakage current impacts power consumption of the NN. Electro-optic bistability possible to Ref. 108.
^c If WDM is used.

requiring both polarities of voltage for programming: one polarity for SET (switching from OFF to ON) and the other polarity for RESET (switching from ON to OFF). The bipolar programming is challenging for all-optical photonic integration since it is difficult to realize plasmonic optical programming.

A bipolar optical readout would require rectification of the light thus an optical rectenna would be required, yet, so far, optical rectenna technology is still in its infancy exhibiting limited performance.¹⁰⁷ On the other hand, phase-change memories are unipolar; thus, all-optical programmability can be achieved, as shown in recent work on all optical spike timing dependent plasticity (STDP) plasticity.⁹² For long-term memory elements, such neuromorphic technology can interface seamlessly with an optical compute system, thus reducing the overall delay by keeping the signal “longer” in the optical domain before eventually converting back to electronics via a photodetector. Nevertheless, electro-optic memristors offer advantages due to the integration synergies with circuitry in existing CMOS technology. Therefore, integrated memristor devices exhibiting electrical programmability might be desirable in some integrated photonic systems for offline weight training. These discussion points are qualitatively summarized in Table II.

IV. ALL-OPTICAL NONLINEAR ACTIVATION FUNCTION

The above-discussed mechanism for thresholding in a multi-layered perceptron relies on photodetection of the optical power (proportional to the square of the electric field), at the output of the weighted additions for each node to node connection, which is hence send as a gate to an EO modulator. This scheme, however, requires additional laser sources (possibly embedded on-chip), whose

intensity is controlled by the electro-optic modulation driven by the photogenerated electrical signal. This represents the main bottleneck of the implementation of the proposed perceptron scheme for achieving NL thresholding, which despites its relatively straightforward implementation and controllability, demands an O-to-E-to-O conversion step at every perceptron; this brings substantial trade-offs in terms of speed and power efficiency of the otherwise intrinsically instantaneous and effortless transmission of the signal through the PIC.

For this reason, an optical nonlinear device with picoseconds response, low insertion losses, high modulation range, which does not deteriorate, or quench is highly desirable and becomes a prerogative to the implementation of an all optical neuron.

Recently, several steps have been taken in this direction. In reservoir optical computing,¹⁰⁹ one of the most accredited optical nonlinearity consists of saturable absorber films, such as graphene layers¹¹⁰ or based on 2-photon absorption.¹¹¹ Other mechanisms are instead based on the nonlinearities of bistable switches,¹¹² and ring resonators¹¹³ have also been investigated. Approaches based on single graphene excitable laser¹¹⁴ have recently shown significant progress in the field of all optical spiking neural networks. However, the main criticality is still related to the enhancement of the modulation strength, operational speed, and the integrability and the reproducibility of such devices in photonic circuit both at the device and system level which represent still an open challenge. Therefore, a combined effort of material engineering and time resolved spectroscopy^{117,115–117} aiming to obtain materials with optical nonlinearities which concurrently provides steep modulation range and characterized by fast dynamics could substantially help the advancement of the field of photonics neural networks. According

to our predictions,²⁴ a 3-layer fully connected (AO-NN) would have a delay of about few picoseconds, or 10^{12} MAC/s, and up to 10^{17} MAC/J efficiency, when performing a standard classification task.

V. CONCLUSIONS

In conclusion, the heterogeneous integration of memristive materials and devices in photonic platforms suggests promising performance advantages for photonic-electronic hybrid artificial neural networks since these two technologies have complementary strengths. The advantage of photonics, for instance, lies in providing a picosecond-short delay between the various network nodes, hence enabling highly efficient “interconnectivity” of such non-von-Neumann compute and information processing architectures. However, photons are challenged to store states, which is where memories come in, such as to provide for the long-term neural network “weights” (nonvolatile) and the nonlinear thresholding or activation function (relatively volatile). In combining the “best-of-both-words,” we see a viable path forward in densely integrating memristive materials particularly with the active photonic components that determine the neural networks governing functions and thus performance.

We find that electro-optic modulators are unique candidates to perform both dot-product “weighting” as well as “thresholding,” yet vastly different time-scales are required for each, while both share the aim to execute the respective functionality with lowest power consumption. The advantage is strengthened by the fact that the two technologies share a broad range of materials, thus enabling seamless integration and stacking of integrated devices. Indeed, hybrid photonic memristive neuromorphic circuitry could enable systems capable of (sub)nanosecond-fast and energy-efficient inference tasks in trained networks.

On the other hand, promising improvements in the field of nonvolatile photonic memory pertain to the fabrication of more effective and efficient photonic memories in which a concurrent minimization of the losses and maximization of the modulation keep the information longer in the optical-domain, i.e., avoids cumbersome O-to-E-to-O conversions. This, however, requires wisely engineering the material process, e.g., interfacial phase change materials ($\text{GeTe/Sb}_2\text{Te}_3$)⁹³ and optimized alloy, such as $\text{Ge}_2\text{Sb}_2\text{Se}_4\text{Te}_1$.⁹⁵ Moreover, the fabrication of frequency selective memories based on phase change materials can dramatically increase the parallelism and consequently opening new frontiers in optical computing and communication. This vibrant field would also demand for strategies and device configurations which would enable an efficient non-volatile all-optical control in photonic multilevel perceptron, possibly integrated, e.g., photonic crystal cavities⁹⁶ or plasmonic slot waveguide.^{84,118}

Moreover, such platforms could be used to realize new neuromorphic architectures which could rely on hybrid devices based on single photon detection¹⁰⁰ or modulation. Finally, regarding the target applications for photonic neural networks, unlike GPUs which are suitable for big-data and high throughput tasks, photonic neural networks would be quite suitable for those specific tasks that rely on real-time (<microsecond) responses to inference machine learning tasks such as those found in military applications of ranging, synthetic aperture radar, or automated target recognition, to name a few, and a proper task-to-hardware mapping including trade-offs

in network complexity and size against performance (i.e., accuracy, delay) is yet outstanding.

REFERENCES

- ¹V. G. Kopytoff, *The New York Times* (The New York Times Company, 2011).
- ²L. Greenemeier, *Sci. Am.* (2013), see <https://www.scientificamerican.com/article/will-ibms-watson-usher-in-cognitive-computing/> for Will IBM's Watson Usher in a New Era of Cognitive Computing?
- ³M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, J. J. Yang, and R. S. Williams, in *2016 53rd ACM/EDAC/IEEE Design Automation Conference (DAC)* (IEEE, 2016), pp. 1–6.
- ⁴R. A. Heinz, J. O. Artman, and S. H. Lee, *Appl. Opt.* **9**, 2161 (1970).
- ⁵A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, *Sci. Rep.* **7**, 7430 (2017).
- ⁶P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, *Science* **345**, 668 (2014).
- ⁷P. R. Prucnal and B. J. Shastri, *Neuromorphic Photonics* (CRC Press, 2017).
- ⁸I. Boybat, M. L. Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, and E. Eleftheriou, *Nat. Commun.* **9**, 2514 (2018).
- ⁹M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. De Salvo, *IEEE Trans. Electron Devices* **60**, 2402 (2013).
- ¹⁰B. Li, P. Gu, Y. Shan, Y. Wang, Y. Chen, and H. Yang, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **34**, 1905 (2015).
- ¹¹Y. Lu, M. Stegmaier, P. Nukala, M. A. Giambra, S. Ferrari, A. Busacca, W. H. P. Pernice, and R. Agarwal, *Nano Lett.* **17**, 150 (2017).
- ¹²C. Ríos, M. Stegmaier, P. Hosseini, D. Wang, T. Scherer, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, *Nat. Photonics* **9**, 725 (2015).
- ¹³C. Ríos, N. Youngblood, Z. Cheng, M. L. Gallo, W. H. P. Pernice, C. D. Wright, A. Sebastian, and H. Bhaskaran, *Sci. Adv.* **5**, eaau5759 (2019).
- ¹⁴J. Zheng, A. Khanolkar, P. Xu, S. Colburn, S. Deshmukh, J. Myers, J. Frantz, E. Pop, J. Hendrickson, J. Doylend, N. Boechler, and A. Majumdar, *Opt. Mater. Express* **8**, 1551 (2018).
- ¹⁵J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, *Nature* **569**, 208 (2019).
- ¹⁶M. Rudé, J. Pello, R. E. Simpson, J. Osmond, G. Roelkens, J. J. G. M. van der Tol, and V. Pruneri, *Appl. Phys. Lett.* **103**, 141119 (2013).
- ¹⁷L. Waldecker, T. A. Miller, M. Rudé, R. Bertoni, J. Osmond, V. Pruneri, R. E. Simpson, R. Ernstorfer, and S. Wall, *Nat. Mater.* **14**, 991 (2015).
- ¹⁸M. Wuttig, H. Bhaskaran, and T. Taubner, *Nat. Photonics* **11**, 465 (2017).
- ¹⁹K. J. Miller, R. F. Haglund, and S. M. Weiss, *Opt. Mater. Express* **8**, 2415 (2018).
- ²⁰P. Xu, J. Zheng, J. Doylend, and A. Majumdar, *ACS Photonics* **6**, 553 (2019).
- ²¹V. K. Narayana, S. Sun, A.-H. A. Badawy, V. J. Sorger, and T. El-Ghazawi, *Microprocessors Microsyst.* **50**, 113 (2017).
- ²²M. P. Fok, Y. Tian, D. Rosenbluth, and P. R. Prucnal, *Opt. Lett.* **37**, 3309 (2012).
- ²³J. K. George, A. Mehrabian, R. Amin, J. Meng, T. F. de Lima, A. N. Tait, B. J. Shastri, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, *Opt. Express* **27**, 5181 (2019).
- ²⁴M. Miscuglio, A. Mehrabian, Z. Hu, S. I. Azzam, J. George, A. V. Kildishev, M. Pelton, and V. J. Sorger, *Opt. Mater. Express* **8**, 3851 (2018).
- ²⁵Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, *Nat. Photonics* **11**, 441 (2017).
- ²⁶C. Wang, M. Zhang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer, and M. Lončar, *Nature* **562**, 101 (2018).
- ²⁷V. Sorianello, M. Midrio, G. Contestabile, I. Asselberghs, J. V. Campenhout, C. Huyghebaert, I. Goykhman, A. K. Ott, A. C. Ferrari, and M. Romagnoli, *Nat. Photonics* **12**, 40 (2018).
- ²⁸R. Amin, R. Maiti, C. Carfano, Z. Ma, M. H. Tahersima, Y. Lilach, D. Ratnayake, H. Dalir, and V. J. Sorger, *APL Photonics* **3**, 126104 (2018).

- ²⁹C. Haffner, D. Chelladurai, Y. Fedoryshyn, A. Josten, B. Baeuerle, W. Heni, T. Watanabe, T. Cui, B. Cheng, S. Saha, D. L. Elder, L. R. Dalton, A. Boltasseva, V. M. Shalae, N. Kinsey, and J. Leuthold, *Nature* **556**, 483 (2018).
- ³⁰S. A. Srinivasan, M. Pantouvaki, S. Gupta, H. T. Chen, P. Verheyen, G. Lepage, G. Roelkens, K. Saraswat, D. V. Thourhout, P. Absil, and J. V. Campenhout, *J. Lightwave Technol.* **34**, 419 (2016).
- ³¹M. Liu, X. Yin, and X. Zhang, *Nano Lett.* **12**, 1482 (2012).
- ³²V. J. Sorger, N. D. Lanzillotti-Kimura, R.-M. Ma, and X. Zhang, *Nanophotonics* **1**, 17 (2012).
- ³³L. Alloatti, R. Palmer, S. Diebold, K. P. Pahl, B. Chen, R. Dinu, M. Fournier, J.-M. Fedeli, T. Zwick, W. Freude, C. Koos, and J. Leuthold, *Light: Sci. Appl.* **3**, e173 (2014).
- ³⁴R. Bellman, *Dynamic Programming*, Reprint ed. (Dover Publications, Mineola, NY, 2003).
- ³⁵L. Ceze, J. Hasler, K. K. Likharev, J.-s. Seo, T. Sherwood, D. Strukov, Y. Xie, and S. Yu, in *2016 74th Annual Device Research Conference (DRC)* (IEEE Device Research Conference, 2016), pp. 1–2.
- ³⁶A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, *J. Lightwave Technol.* **32**, 4029 (2014).
- ³⁷V. J. Sorger, R. Amin, J. B. Khurgin, Z. Ma, H. Dalir, and S. Khan, *J. Opt.* **20**, 014012 (2017).
- ³⁸W. Heni, Y. Fedoryshyn, B. Baeuerle, A. Josten, C. B. Hoessbacher, A. Messner, C. Haffner, T. Watanabe, Y. Salamin, U. Koch, D. L. Elder, L. R. Dalton, and J. Leuthold, *Nat. Commun.* **10**, 1694 (2019).
- ³⁹D. A. B. Miller, *J. Lightwave Technol.* **35**, 346 (2017).
- ⁴⁰I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, “Reprogrammable electro-optic nonlinear activation functions for optical neural networks,” *IEEE J. Sel. Top. Quantum Electron.* **26**, 1 (to be published).
- ⁴¹S. A. Miller, M. Yu, X. Ji, A. G. Griffith, J. Cardenas, A. L. Gaeta, and M. Lipson, *Optica* **4**, 707 (2017).
- ⁴²Y. A. Vlasov and S. J. McNab, *Opt. Express* **12**, 1622 (2004).
- ⁴³C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, and J. Leuthold, *Nat. Photonics* **9**, 525 (2015).
- ⁴⁴M. Ziebell, D. Marris-Morini, G. Rasigade, P. Crozat, J.-M. Fédéli, P. Grosse, E. Cassan, and L. Vivien, *Opt. Express* **19**, 14690 (2011).
- ⁴⁵F. Y. Gardes, A. Brimont, P. Sanchis, G. Rasigade, D. Marris-Morini, L. O’Faolain, F. Dong, J. M. Fedeli, P. Dumon, L. Vivien, T. F. Krauss, G. T. Reed, and J. Martí, *Opt. Express* **17**, 21986 (2009).
- ⁴⁶M. Pantouvaki, P. Verheyen, J. De Coster, G. Lepage, P. Absil, and J. Van Campenhout, in *2015 European Conference on Optical Communication (ECOC)* (IEEE, Valencia, Spain, 2015), pp. 1–3.
- ⁴⁷A. H. Atabaki, S. Moazeni, F. Pavanello, H. Gevorgyan, J. Notaros, L. Alloatti, M. T. Wade, C. Sun, S. A. Kruger, H. Meng, K. A. Qubaisi, I. Wang, B. Zhang, A. Khilo, C. V. Baiocco, M. A. Popović, V. M. Stojanović, and R. J. Ram, *Nature* **556**, 349 (2018).
- ⁴⁸M. R. Billah, M. Blaicher, T. Hoose, P.-I. Dietrich, P. Marin-Palomo, N. Lindenmann, A. Nesic, A. Hofmann, U. Troppenz, M. Moehrl, S. Randel, W. Freude, and C. Koos, *Optica* **5**, 876 (2018).
- ⁴⁹J. Heck, H. Bar, R. Jones, and H. Park, U.S. patent US20120189317A1 (26 July 2012).
- ⁵⁰A. Lee, Q. Jiang, M. Tang, A. Seeds, and H. Liu, *Opt. Express* **20**, 22181 (2012).
- ⁵¹R. E. Camacho-Aguilera, Y. Cai, N. Patel, J. T. Bessette, M. Romagnoli, L. C. Kimerling, and J. Michel, *Opt. Express* **20**, 11316 (2012).
- ⁵²K. Liu and V. J. Sorger, *Opt. Mater. Express* **5**, 1910 (2015).
- ⁵³J. F. Bauters, M. J. R. Heck, D. John, D. Dai, M.-C. Tien, J. S. Barton, A. Leinse, R. G. Heideman, D. J. Blumenthal, and J. E. Bowers, *Opt. Express* **19**, 3163 (2011).
- ⁵⁴P. Muñoz, G. Micó, L. A. Bru, D. Pastor, D. Pérez, J. D. Doménech, J. Fernández, R. Baños, B. Gargallo, R. Alemany, A. M. Sánchez, J. M. Cirera, R. Mas, and C. Domínguez, *Sensors* **17**, 2088 (2017).
- ⁵⁵R. Baets, A. Z. Subramanian, S. Clemmen, B. Kuyken, P. Bienstman, N. Le Thomas, G. Roelkens, D. Van Thourhout, P. Helin, and S. Severi, in *Optical Fiber Communication Conference* (OSA, Anaheim, California, 2016), p. Th3J.1.
- ⁵⁶W. Xie, T. Stöferle, G. Rainò, T. Aubert, S. Bisschop, Y. Zhu, R. F. Mahrt, P. Geiregat, E. Brainis, Z. Hens, and D. Van Thourhout, *Adv. Mater.* **29**, 1604866 (2017).
- ⁵⁷Y. Zhu, W. Xie, P. Geiregat, S. Bisschop, T. Aubert, E. Brainis, Z. Hens, and D. van Thourhout, in *2017 Conference on Lasers and Electro-Optics (CLEO)* (OSA Publishing, 2017), pp. 1–2.
- ⁵⁸A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, *Nature* **427**, 615 (2004).
- ⁵⁹J.-S. Kim and J. T. Kim, *J. Phys. D: Appl. Phys.* **49**, 075101 (2016).
- ⁶⁰T. Hiraki, T. Aihara, K. Hasebe, T. Fujii, K. Takeda, T. Kakitsuka, T. Tsuchizawa, H. Fukuda, and S. Matsuo, in *2017 European Conference on Optical Communication (ECOC)* (IEEE Proceedings, 2017), pp. 1–3.
- ⁶¹Y. Tang, J. D. Peters, and J. E. Bowers, *Opt. Express* **20**, 11529 (2012).
- ⁶²Y. T. Hu, M. Pantouvaki, S. Brems, I. Asselberghs, C. Huyghebaert, M. Geisler, C. Alessandri, R. Baets, P. Absil, D. V. Thourhout, and J. V. Campenhout, in *2014 IEEE International Electron Devices Meeting* (IEEE, 2014), pp. 5.6.1–5.6.4.
- ⁶³D. Marris-Morini, V. Vakarin, J. M. Ramirez, Q. Liu, A. Ballabio, J. Frigerio, M. Montesinos, C. Alonso-Ramos, R. X. Le, S. Serna, D. Benedikovic, D. Christina, L. Vivien, and G. Isella, *Nanophotonics* **7**, 1781 (2018).
- ⁶⁴L. Mastronardi, M. Banakar, A. Z. Khokhar, N. Hattasan, T. Rutirawut, T. D. Bucio, K. M. Grabska, C. Littlejohns, A. Bazin, G. Mashanovich, and F. Y. Gardes, *Opt. Express* **26**, 6663 (2018).
- ⁶⁵A. Honardoost, F. A. Juneghani, R. Safian, and S. Fathpour, *Opt. Express* **27**, 6495 (2019).
- ⁶⁶R. Amin, S. Khan, C. J. Lee, H. Dalir, and V. J. Sorger, in *Conference on Lasers and Electro-Optics* (Optical Society of America, 2018), p. SM11.5.
- ⁶⁷R. Amin, J. George, S. Sun, T. F. de Lima, A. N. Tait, J. Khurgin, M. Miscuglio, B. J. Shastri, P. R. Prucnal, T. El-Ghazawi, and V. J. Sorger, *APL Mater.* **7**, 081112 (2019).
- ⁶⁸C. Wang, M. Zhang, M. Yu, R. Zhu, H. Hu, and M. Loncar, *Nat. Commun.* **10**, 978 (2019).
- ⁶⁹I. Bar-Joseph, C. Klingshirn, D. A. B. Miller, D. S. Chemla, U. Koren, and B. I. Miller, *Appl. Phys. Lett.* **50**, 1010 (1987).
- ⁷⁰Y. Kuo, Y. K. Lee, Y. Ge, S. Ren, J. E. Roth, T. I. Kamins, D. A. B. Miller, and J. S. Harris, *IEEE J. Sel. Top. Quantum Electron.* **12**, 1503 (2006).
- ⁷¹H.-T. Peng, M. A. Nahmias, T. F. de Lima, A. N. Tait, and B. J. Shastri, *IEEE J. Sel. Top. Quantum Electron.* **24**, 1 (2018).
- ⁷²L. Alloatti, D. Korn, R. Palmer, D. Hillerkuss, J. Li, A. Barklund, R. Dinu, J. Wieland, M. Fournier, J. Fedeli, H. Yu, W. Bogaerts, P. Dumon, R. Baets, C. Koos, W. Freude, and J. Leuthold, *Opt. Express* **19**, 11841 (2011).
- ⁷³P. O. Weigel, J. Zhao, K. Fang, H. Al-Rubaye, D. Trotter, D. Hood, J. Mudrick, C. Dallo, A. T. Pomerene, A. L. Starbuck, C. T. DeRose, A. L. Lentine, G. Rebeiz, and S. Mookherjee, *Opt. Express* **26**, 23728 (2018).
- ⁷⁴J.-K. Park, S. Takagi, and M. Takenaka, *Opt. Express* **26**, 4842 (2018).
- ⁷⁵M. Ayata, Y. Fedoryshyn, W. Heni, B. Baeuerle, A. Josten, M. Zahner, U. Koch, Y. Salamin, C. Hoessbacher, C. Haffner, D. L. Elder, L. R. Dalton, and J. Leuthold, *Science* **358**, 630 (2017).
- ⁷⁶H. Dalir, Y. Xia, Y. Wang, and X. Zhang, *ACS Photonics* **3**, 1564 (2016).
- ⁷⁷C. Ye, S. Khan, Z. Ran Li, E. Simsek, and V. J. Sorger, *IEEE J. Sel. Top. Quantum Electron.* **20**, 40 (2014).
- ⁷⁸D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, *Phys. Rev. B* **32**, 1043 (1985).
- ⁷⁹R. Maiti, C. Patil, R. A. Hemnani, M. Miscuglio, R. Amin, Z. Ma, R. Chaudhary, A. T. C. Johnson, L. Bartels, R. Agarwal, and V. J. Sorger, *Opt. Mater. Express* **9**, 751 (2019).
- ⁸⁰R. Maiti, R. A. Hemnani, R. Amin, Z. Ma, M. H. Tahersima, T. A. Empante, H. Dalir, R. Agarwal, L. Bartels, and V. J. Sorger, *Nanophotonics* **8**, 435 (2019).
- ⁸¹H. W. Lee, G. Papadakis, S. P. Burgos, K. Chander, A. Kriesch, R. Pala, U. Peschel, and H. A. Atwater, *Nano Lett.* **14**, 6463 (2014).
- ⁸²E. Feigenbaum, K. Diest, and H. A. Atwater, *Nano Lett.* **10**, 2111 (2010).

- ⁸³R. Amin, C. Suer, Z. Ma, I. Sarpkaya, J. B. Khurgin, R. Agarwal, and V. J. Sorger, *Solid-State Electron.* **136**, 92 (2017).
- ⁸⁴Z. Ma, K. Kikunage, H. Wang, S. Sun, R. Amin, M. Tahersima, R. Maiti, M. Miscuglio, H. Dalir, and V. J. Sorger, e-print [arXiv:1812.00894](https://arxiv.org/abs/1812.00894) [Physics] (2018).
- ⁸⁵P. Ma, Y. Salamin, B. Baeuerle, A. Josten, W. Heni, A. Emboras, and J. Leuthold, *ACS Photonics* **6**, 154 (2019).
- ⁸⁶Y.-Q. Bie, G. Grosso, M. Heuck, M. M. Furchi, Y. Cao, J. Zheng, D. Bunandar, E. Navarro-Moratalla, L. Zhou, D. K. Efetov, T. Taniguchi, K. Watanabe, J. Kong, D. Englund, and P. Jarillo-Herrero, *Nat. Nanotechnol.* **12**, 1124 (2017).
- ⁸⁷S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bordini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, and G. W. Burr, *Nature* **558**, 60 (2018).
- ⁸⁸B. D. Hoskins, M. W. Daniels, S. Huang, A. Madhavan, G. C. Adam, N. Zhitenov, J. J. McClelland, and M. D. Stiles, "Streaming batch eigenupdates for hardware neural networks," *Front. Neurosci.* **13**, 793 (2019).
- ⁸⁹I. Kataeva, F. Merrikh-Bayat, E. Zamanidoost, and D. Strukov, in *2015 International Joint Conference on Neural Networks (IJCNN)* (IEEE Proceedings, 2015), pp. 1–8.
- ⁹⁰Y. Li, H. Yu, T. Dai, J. Jiang, G. Wang, L. Yang, W. Wang, J. Yang, and X. Jiang, *IEEE Photonics Technol. Lett.* **28**, 284 (2016).
- ⁹¹A. J. Hong, E. B. Song, H. S. Yu, M. J. Allen, J. Kim, J. D. Fowler, J. K. Wassei, Y. Park, Y. Wang, J. Zou, R. B. Kaner, B. H. Weiller, and K. L. Wang, *ACS Nano* **5**, 7812 (2011).
- ⁹²Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, *Sci. Adv.* **3**, e1700160 (2017).
- ⁹³R. E. Simpson, P. Fons, A. V. Kolobov, T. Fukaya, M. Krbal, T. Yagi, and J. Tominaga, *Nat. Nanotechnol.* **6**, 501 (2011).
- ⁹⁴X. Li, N. Youngblood, C. Ríos, Z. Cheng, C. D. Wright, W. H. Pernice, and H. Bhaskaran, *Optica* **6**, 1 (2019).
- ⁹⁵Y. Zhang, J. B. Chou, J. Li, H. Li, Q. Du, A. Yadav, S. Zhou, M. Y. Shalaginov, Z. Fang, H. Zhong, C. Roberts, P. Robinson, B. Bohlin, C. Ríos, H. Lin, M. Kang, T. Gu, J. Warner, V. Liberman, K. Richardson, and J. Hu, *Nat. Commun.* **10**, 1 (2019).
- ⁹⁶J. von Keitz, J. Feldmann, N. Gruhler, C. Ríos, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, *ACS Photonics* **5**, 4644 (2018).
- ⁹⁷A. Emboras, I. Goykhman, B. Desiatov, N. Mazurski, L. Stern, J. Shappir, and U. Levy, *Nano Lett.* **13**, 6151 (2013).
- ⁹⁸I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, e-print [arXiv:1609.07061](https://arxiv.org/abs/1609.07061) [Cs] (2016).
- ⁹⁹M. Courbariaux, Y. Bengio, and J.-P. David, in *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., 2015), pp. 3123–3131.
- ¹⁰⁰A. Emboras, A. Alabastri, F. Ducry, B. Cheng, Y. Salamin, P. Ma, S. Andermatt, B. Baeuerle, A. Josten, C. Hafner, M. Luisier, P. Nordlander, and J. Leuthold, *ACS Nano* **12**, 6706 (2018).
- ¹⁰¹Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, and J. J. Yang, *Nat. Mater.* **16**, 101 (2017).
- ¹⁰²Z. L. Sámson, K. F. MacDonald, F. De Angelis, B. Gholipour, K. Knight, C. C. Huang, E. Di Fabrizio, D. W. Hewak, and N. I. Zheludev, *Appl. Phys. Lett.* **96**, 143105 (2010).
- ¹⁰³Y. Luo and F. N. Yang (2014), see <http://www.andrew.cmu.edu/user/fanyang1/DLWN.pdf> for Deep learning with noise.
- ¹⁰⁴Y. Gui, M. Miscuglio, Z. Ma, M. H. Tahersima, S. Sun, R. Amin, H. Dalir, and V. J. Sorger, *Sci. Rep.* **9**, 1 (2019).
- ¹⁰⁵C. B. Lee, B. S. Kang, A. Benayad, M. J. Lee, S.-E. Ahn, K. H. Kim, G. Stefanovich, Y. Park, and I. K. Yoo, *Appl. Phys. Lett.* **93**, 042115 (2008).
- ¹⁰⁶A. Mehrabian, M. Miscuglio, Y. Alkabani, V. J. Sorger, and T. El-Ghazawi, e-print [arXiv:1906.10487](https://arxiv.org/abs/1906.10487) [Cs, Eess] (2019).
- ¹⁰⁷A. Sharma, V. Singh, T. L. Bougher, and B. A. Cola, *Nat. Nanotechnol.* **10**, 1027 (2015).
- ¹⁰⁸A. Majumdar and A. Rundquist, *Opt. Lett.* **39**, 3864 (2014).
- ¹⁰⁹A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, *Opt. Express* **22**, 10868 (2014).
- ¹¹⁰Q. Bao, H. Zhang, Z. Ni, Y. Wang, L. Polavarapu, Z. Shen, Q.-H. Xu, D. Tang, and K. P. Loh, *Nano Res.* **4**, 297 (2011).
- ¹¹¹R. W. Schirmer and A. L. Gaeta, *J. Opt. Soc. Am. B* **14**, 2865 (1997).
- ¹¹²M. Soljačić, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, *Phys. Rev. E* **66**, 055601 (2002).
- ¹¹³F. D. Coarer, M. Sciamanna, A. Katumba, M. Freiburger, J. Dambre, P. Bienstman, and D. Rontani, *IEEE J. Sel. Top. Quantum Electron.* **24**, 1 (2018).
- ¹¹⁴B. J. Shastri, M. A. Nahmias, A. N. Tait, A. W. Rodriguez, B. Wu, and P. R. Prucnal, *Sci. Rep.* **6**, 19126 (2016).
- ¹¹⁵S. I. Azzam and A. V. Kildishev, *Opt. Mater. Express* **8**, 3829 (2018).
- ¹¹⁶H. Leng, B. Szychowski, M.-C. Daniel, and M. Pelton, *Nat. Commun.* **9**, 4012 (2018).
- ¹¹⁷M. Z. Alam, I. D. Leon, and R. W. Boyd, *Science* **352**, 795 (2016).
- ¹¹⁸Y. Salamin, P. Ma, B. Baeuerle, A. Emboras, Y. Fedoryshyn, W. Heni, B. Cheng, A. Josten, and J. Leuthold, *ACS Photonics* **5**, 3291 (2018).