APL Materials

ITO-based electro-absorption modulator for photonic neural activation function

Cite as: APL Mater. 7, 081112 (2019); https://doi.org/10.1063/1.5109039
Submitted: 05 May 2019 . Accepted: 28 July 2019 . Published Online: 15 August 2019

R. Amin, J. K. George, S. Sun, T. Ferreira de Lima, A. N. Tait, J. B. Khurgin, M. Miscuglio, B. J. Shastri, P. R. Prucnal, T. El-Ghazawi, and V. J. Sorger

openaccess

View Online Export Citation CrossMark

AMERICAN ELEMENTS
THE ADVANCED MATERIALS MANUFACTURER®

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.

Now Invent.™
The Next Generation of Material Science Catalogs

www.americanelements.com

AIP Publishing

© 2019 Author(s).

ITO-based electro-absorption modulator for photonic neural activation function

Cite as: APL Mater. 7, 081112 (2019); doi: 10.1063/1.5109039
Submitted: 5 May 2019 • Accepted: 28 July 2019 •
Published Online: 15 August 2019

View Online Export Citation CrossMark

R. Amin,[1] J. K. George,[1] S. Sun,[1] T. Ferreira de Lima,[2] A. N. Tait,[2] J. B. Khurgin,[3] M. Miscuglio,[1] B. J. Shastri,[2,4] P. R. Prucnal,[2] T. El-Ghazawi,[1] and V. J. Sorger[1,a)]

AFFILIATIONS

[1]Department of Electrical and Computer Engineering, George Washington University, Washington, District of Columbia 20052, USA
[2]Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA
[3]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA
[4]Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, Ontario KL7 3N6, Canada

Note: This paper is part of the special topic on Emerging Materials in Neuromorphic Computing.
[a)]**Electronic mail:** sorger@gwu.edu

ABSTRACT

Recently, integrated optics has become a functional platform for implementing machine learning algorithms and, in particular, neural networks. Photonic integrated circuits can straightforwardly perform vector-matrix multiplications with high efficiency and low power consumption by using weighting mechanism through linear optics. However, this cannot be said for the activation function, i.e., "threshold," which requires either nonlinear optics or an electro-optic module with an appropriate dynamic range. Even though all-optical nonlinear optics is potentially faster, its current integration is challenging and is rather inefficient. Here, we demonstrate an electroabsorption modulator based on an indium tin oxide layer monolithically integrated into silicon photonic waveguides, whose dynamic range is used as a nonlinear activation function of a photonic neuron. The thresholding mechanism is based on a photodiode, which integrates the weighed products, and whose photovoltage drives the electroabsorption modulator. The synapse and neuron circuit is then constructed to execute a 200-node MNIST classification neural network used for benchmarking the nonlinear activation function and compared with an equivalent electronic module.

© 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/10.1063/1.5109039

I. INTRODUCTION

With the ongoing advancement of neural network systems, there is a pressing demand for new technological paradigms that can perform advanced artificial intelligence tasks without trading off throughput (operations/s) and power dissipation. Photonics based artificial neurons can potentially provide the archetypal solution for this specific challenge.[1–3] The main advantage of photonics over current digital electronics implementations is that distinct signals due to their wave-nature can be straightforwardly and efficiently combined exploiting attojoule efficient electro-optic (EO) modulators,[4–7] phase shifters, and combiners, simplifying essential operations such as weighted sum or addition, vector matrix multiplications, or convolutions.[8] Moreover, photonics enables high parallelism,[9] hence a higher baud-rate, since multiple wavelengths can travel in the same physical channel by exploiting wavelength-division multiplexing (WDM).[8,10–13] Additionally, photonics is marked by a further degree of freedom in modulating the information carried by the optical waves since a signal can be modulated by altering its phase, amplitude, or polarization.

In a photonic neural network, a photonic neuron performs the following operations: receive multiple input signals (fan-in), weight each by the coefficient and sum them (weighted sum or addition), utilize a nonlinear process (nonlinear activation function), and send its output to other neurons or nodes (fan-out). The Non-Linear Activation Function (NL AF) has a key role in the network pivotal to the convergence into final states by discriminating data and suppressing noise. Different NL AFs with significantly different ranges and

© Author(s) 2019

trends[14] have been proposed and extensively investigated, providing suitable advantages according to the different applications. More in detail, interesting experimental and numerical all-optical nonlinear modules based on saturable and reverse absorption,[15,16] graphene excitable lasers,[17,18] two-section distributed-feedback (DFB) lasers,[19] quantum dots,[20] disks lasers,[21,22] induced transparency in quantum assembly,[15] etc., have been recently reported and showed promising results in terms of efficiency and throughput for different kinds of neural network and applications, ranging from convolutional neural network, spiking neural network, and reservoir computing. However, a more straightforward implementation is currently attained by exploiting electro-optic tuned nonlinear materials[23,24] or absorptive modulator directly connected to a photodiode, as shown in Refs. 14 and 25–27. In this case, the photogenerated current, proportional to the detected optical power at the weighted addition, alters the voltage drop on the active material, thus changing its carrier concentration and consequently the effective modal index of the propagating waveguide mode. This approach is affected by RC-latency and by the electro-optic conversion; therefore, it trades off baud-rate with energy efficiency.[28] Nonetheless, it still provides higher controllability, reconfigurability, reliability, and easy-integration with respect to more exotic design choices. In this case, the adoption of a specific modulator type will directly impact the shape of the nonlinear activation function and, subsequently, the inference mechanism and ultimately the performance of the neural network. Indeed, the performance of a modulator is directly related to the underlying physics of the entirety of the optoelectronic devices, including active material and device configuration. Therefore, for achieving efficient modulation and, consequently, promising performance for the neural network, one should concurrently engineer both. On the device engineering side, maximizing the fraction of the optical field in the mode overlapping with the active material, namely, intensifying the modal confinement factor, is a suitable and accessible strategy for enhancing the modulation efficiency and the dynamic range of an electro-optic modulator embedded in a photonic integrated circuit. For achieving that, current schemes consist of using subdiffraction limited plasmonic structures[29–36] or photonic cavities[35] aiming to maximize the light matter interaction in order to achieve a rather high modulation performance and low energy-per-compute surpassing electronic efficiency, while compensating in terms of insertion losses (optical mode hybridization) due to the plasmonic nature of the mode. Another characteristic to consider when engineering an electroabsorption (EA) modulator is the modulation bandwidth, which is the result of material choices and device configuration and, if well engineered, can enable high-throughput communication links. Though necessary, a thorough device configuration might not be enough for achieving an efficient complex index modulation due to inefficient active material choices. The main aspect to consider when designing and engineering an effective electroabsorption modulator (EAM) is, in fact, the variation of the complex refractive index due to applied bias (i.e., carrier tunability), which is inherent to the selected active material.[26] Silicon (Si) is the conventional material choice usually as fabrication facilities can benefit tremendously from the mature Si process, but the inherent low tunability of Si under electrical bias forces inadequate performances at higher scaling as increased modulator lengths need to be employed to achieve the desired dynamic range for the NL AF.

Moreover, in order to respond to the demand of densely integrated neural networks, the material of choice for the EA modulator needs to be easily and consistently integrated in a photonic platform, and preferably, the selected active material should be CMOS compatible, thus allowing integration with on-chip memory and EO converter such as the digital-to-analog converter (DAC) and analog-to-digital converter (ADC) enabling large-scale functional photonic network-on-chip. Undoubtedly, one class of materials which adapts to these requirements is transparent conductive-oxides (TCOs). Indium Tin Oxide (ITO) belongs to this class of material. The advantages of using ITO as an active material in EA modulators are manifolds; ITO films are able to deliver unity-strong index modulation when placed in a plasmonic cavity configuration[30,37–39] and also show strong modulation in epsilon-near-zero (ENZ) behavior[40] in the telecommunication frequency band,[41] supporting both strong index modulation and slow light effects.[42] In recent years, the massive demand of ITO for industrial purposes[43] and research-level advancements favored the rapid development of controllable and tunable processes for ITO deposition, by Radio Frequency (RF),[6,44–47] DC sputtering,[48] and evaporation,[49] which enabled a very high yield and reliable process and monolithic hybrid integration.[50]

In this paper, we start our study by analyzing the free carrier absorption dynamics in ITO (Sec. II A), only to then build a model of an amplifier coupled electro-optic neuron for the electroabsorption modulator (Sec. II B). Subsequently, we derive its cascaded signal-to-noise ratio (SNR) as a function of its length and laser power. We compare the model obtained with a silicon-based modulator, being still largely employed with successful results[33,51–55] (Sec. III A). As a subcase of our study, we also experimentally demonstrate an ITO-based absorption modulator which, due to intrinsic nonideality such as initial doping and asymmetric behavior, outperforms the theoretical model initially described. Ultimately, we evaluate and benchmark the performance of the experimentally fabricated ITO modulator as a NL AF on toy networks such as MNIST feed-forward neural network, designed with two layers of 100 nodes each, using optimized parameters from the SNR analysis. The value of accuracy for the trained network, including noise, reached a remarkable 98%, exploiting also nonideal and asymmetric behavior in the transfer function, discussed in Sec. III B. In conclusion, our results show that ITO-based electroabsorption modulators offer a significantly higher modulation dynamic range and efficiency with respect to similar silicon-based devices and hence are suitable for implementing the nonlinear activation functions in electro-optic neural networks as they require only a small fraction of the power and footprint. We believe that this work could constitute a viable approach not only toward the implementation of a photonic perceptron mechanism based on ITO modulators, both for weighting the inputs and implementing nonlinear activation functions, but also could represent a comprehensive guide for other material integration to avail future neural network trends.

II. METHODS

Initially, we study the electro-optical absorption dynamics in ITO dictated by the free carrier mechanism, which is essential for determining the model for the absorption as a function of the injected carriers and bias voltage. We compare these results with

© Author(s) 2019

silicon carrier dynamics as the underlying modulation mechanism for both Si and ITO are the same—free carrier absorption arising from accumulation/depletion of the carriers.

A. Free carrier absorption dynamics

Let us consider a free electron gas in the conduction band of a semiconductor such as in Si, ITO, or any other similar material with the carrier density of $N_c(x, y)$. The schematics of a generic waveguide modulator case is shown in Fig. 1(a) employing such active materials. Here, we derive the expression for the absorption cross section of free electrons in these akin materials. The relative permittivity of aforementioned free electron plasma in the Drude approximation can be written as

$$\varepsilon_r(\omega, x, y) = \varepsilon_\infty - \frac{N_c(x,y)q^2}{\varepsilon_0 m_{\text{eff}}} \frac{1}{\omega^2 + i\omega\gamma}, \tag{1}$$

where ε_∞ is the dielectric constant of the undoped semiconductor also referred to as the high-frequency or "background" permittivity, q is the electronic charge, m_{eff} is the conduction effective mass, ω is the angular frequency, and γ is the collision frequency of the free carriers quantified by the detuning from any interband resonances thereof. The rate of absorption of electromagnetic energy is proportional to the imaginary part of the dielectric constant as

$$\frac{dU(x,y)}{dt} = \omega\varepsilon_0\varepsilon_{r,\text{im}}(\omega, x)\frac{E_{\text{eff}}^2(x)}{2} = \frac{N_c(x,y)q^2}{m_{\text{eff}}}\frac{\gamma}{\omega^2+\gamma^2}\frac{E^2(x,y)}{2}, \tag{2}$$

where the effective electric field, E_{eff}, is the same as the total field, E. Now, let us consider an electromagnetic wave in a waveguide with width W, propagating along the z-direction [Fig. 1(a)], with the propagation constant β expressed as $E = E(x, y)e^{i(\beta z - \omega t)}$. We can evaluate the Poynting vector as $S_z(x) = E_y H_x$. According to Maxwell equations, $\frac{\partial H_x}{\partial z} = \beta H_x = \omega n^2(x)\varepsilon_0 E_y$. Thus, the time-averaged Poynting vector magnitude is $S_z(x) = \omega n^2(x)\varepsilon_0 E_y^2/2\beta = n^2(x)E_y^2/2n_{\text{eff}}\eta_0$, where η_0 is the impedance of free space ($\simeq$377 Ω). The effective index has been introduced as $n_{\text{eff}} = \beta c/\omega$. Integrating over the waveguide cross section [Fig. 1(a)], we obtain

$$\frac{dP}{dz} = -\int_{\text{active}}\int_{-W/2}^{W/2}\frac{dU}{dt}\,dx\,dy = -\frac{q^2}{m_{\text{eff}}}\frac{\gamma}{\omega^2+\gamma^2}\frac{W}{2}\int_{-\infty}^{\infty}N_c(x)E^2(x)\,dx. \tag{3}$$

The total power flow then becomes

$$P = W\int_{-\infty}^{\infty}S_z(x)\,dx = \frac{W}{2n_{\text{eff}}\eta_0}\int_{-\infty}^{\infty}n^2(x)E_y^2\,dx = \frac{Wt_{\text{eff}}n_{eff}}{2\eta_0}\frac{\int_{-\infty}^{\infty}N_c(x)E^2(x)\,dx}{N_{2D}}. \tag{4}$$

Here, N_{2D} is defined as the 2-dimensional planar carrier concentration and the effective thickness is

$$t_{eff} = \frac{1}{n_{eff}^2}\frac{\int_{-\infty}^{\infty}N_c(x)\,dx\int_{-\infty}^{\infty}n^2(x)E_y^2\,dx}{\int_{-\infty}^{\infty}N_c(x)E^2(x)\,dx} \approx \frac{1}{n_{eff}^2}\int_{-\infty}^{\infty}n^2(x)E_x^2\,dx/E_{a0}^2. \tag{5}$$

For a very narrow active layer, this approximation can be inferred—otherwise, it is slightly smaller. This definition of the effective thickness may differ from many other in the literature, but the difference is not significant. Usually, it is just in the arrangement of all kinds of indices. Note, this definition includes the fact that active layer is not necessary at the center of the waveguide, i.e., E_{a0} need not be the peak electric field. The effective thickness can also be thought of as a relative measure of the confinement of the optical mode to the active layer as confinement factor, $\Gamma \simeq d_a/t_{\text{eff}}$.

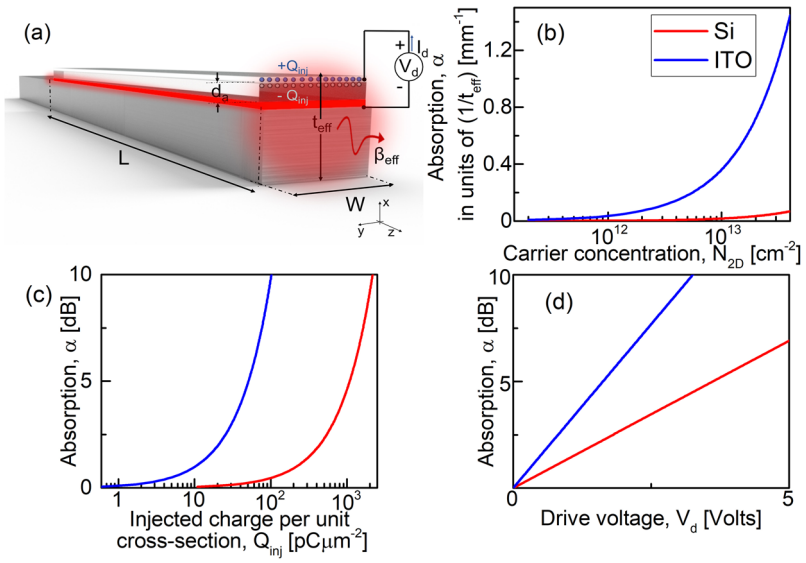

FIG. 1. Free carrier based absorption modulators in silicon (Si) and indium tin oxide (ITO). (a) Schematic of a generic waveguide modulator showcasing the cross-sectional structure (width, W, and active layer thickness, d_a) and modal effective thickness, t_{eff}. The mode inside the structure is propagating in the z-direction with propagation constant, β_{eff}, L is the length of the modulator, V_d and I_d are the drive voltage and drive current, respectively, responsible for the injected charge accumulation/depletion ($\pm Q_{\text{inj}}$) in the active layer/gate interface. (b) Absorption, α, in units of $1/t_{\text{eff}}$ (mm^{-1}) of our cross-sectional geometry in (a) vs 2-dimensional planar carrier concentration, N_{2D} (cm^{-2}). (c) Absorption, α, (dB) vs charge injected per unit cross-sectional area, Q_{inj} (pCμm^{-2}). A 10 dB ceiling for the attainable extinction ratio (ER) was chosen arbitrarily. (d) Absorption, α, (dB) vs drive voltage, V_d, (Volts) corresponding to electrostatic gating in the presence of a gate dielectric with $\varepsilon_{\text{eff}} = 10$ and thickness, $d_{\text{spacer}} = 5$ nm.

© Author(s) 2019

It is worthwhile to mention that the only waveguide geometrical parameter distinguishing the modulator's performance characteristics under various optical waveguide modes is the effective thickness, t_{eff}; i.e., different modes can have different t_{eff} values, but their corresponding performance metrics also follow this important parameter. As such, we carry out our entire performance analysis in units of t_{eff} to ascertain only the material contributions aside from any modal geometric engineering advances thereof. Now, we can obtain from (3) as

$$\frac{dP(z)}{dz} = -\frac{q^2\eta_0}{m_{eff}n_{eff}}\frac{N_{2D}}{t_{eff}}\frac{\gamma}{\omega^2+\gamma^2}P(z) = -\alpha(\omega)P(z). \tag{6}$$

The absorption coefficient can be found as $\alpha(\omega) = \sigma_{abs}(\omega)N_{2D}/t_{eff}$, where the absorption cross section is expressed as

$$\begin{aligned}\sigma_{abs}(\omega) &= \frac{q^2\eta_0}{m_{eff}n_{eff}\gamma}L(\omega) = \frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar}{m_{eff}}\frac{\gamma}{\omega^2+\gamma^2}\\ &= \frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar^2}{m_0(\hbar\gamma_{eff})}\frac{m_0}{m_{eff}}\\ &\approx \frac{7.02\times10^{-17}\ \text{cm}^2}{\hbar\gamma_{eff}}\times\frac{m_0}{m_{eff}}.\end{aligned} \tag{7}$$

Here, α_0 is the fine structure constant, and $L(\omega) = \gamma^2/(\omega^2+\gamma^2)$ represents a Lorentzian line shape that has maximum value at resonance equal to unity. The effective detuning then becomes $\gamma_{eff} = \gamma + \omega^2/\gamma$. Note that the maximum absorption is achieved at $\gamma = \omega$ and when $\gamma_{eff} = 2\omega$ and

$$\begin{aligned}\sigma_{abs}(\omega)|_{max} &= \frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar}{m_{eff}\omega}\times\frac{1}{2}\\ &\approx 3.53\times10^{-17}\ \text{cm}^2\frac{1}{\hbar\omega n_{eff}}\frac{m_0}{m_{eff}}.\end{aligned} \tag{8}$$

Evidently, this expression is similar to the absorption cross section of semiconductor quantum dots (QDs) with one major difference—due to nonresonant character of the free carrier absorption, the cross section for the free carriers is orders of magnitude lower than that for the resonant QDs.[28] As one can see, the absorption cross section does not depend on too many things—at the resonance, it depends only on the resonance width, or, better $Q = \omega/\gamma$ and the optical dipole transition rate, r_{12}. We can invoke the oscillator sum rule (assuming all the dipoles along the axis parallel to the electric field are lined up) which states $2m_{eff}\omega r_{12}^2 \le \hbar$. Therefore, we can introduce the oscillator strength as $f_{12} = 2m_{eff}\omega r_{12}^2/\hbar \le 1$ and consequentially get

$$\sigma_{abs} = 2\sigma_{max}\frac{\gamma^2}{\omega^2+\gamma^2} \approx 2\sigma_{max}\frac{\gamma^2}{\omega^2}. \tag{9}$$

If we now assume that the material used is a semiconductor with a bandgap, E_g, and the momentum matrix element quantified by P_{cv}, we can actually relate the effective mass of electron to the wavelength

$$\frac{m_0}{m_{eff}} = 1 + \frac{2P_{cv}^2}{m_oE_g} \approx \frac{E_p}{E_g}, \tag{10}$$

where $E_p = 2P_{cv}^2/m_0 \approx 16-24$ eV, for a wide range of semiconductors. Therefore, we obtain (assuming that the bandgap is close to the photon energy)

$$\sigma(\omega) \approx 9.6\times10^{-15}\ \text{cm}^2\frac{f_{12}}{(\hbar\gamma)(\hbar\omega)n_{eff}}. \tag{11}$$

Thus, while modulators based on two-level systems do inherently benefit from the small γ, the condition for the free carrier absorption based modulators is quite opposite as they benefit from for large γ. Also, we see that for the wavelength in the telecom range the cross section actually scales with broadening caused by scattering γ. Therefore, ITO leads intrinsically to more modulation per unit charge than high-quality silicon. We found in our previous work that for all materials operating with Pauli blocking (saturable absorption schemes) show comparable values of absorption cross section ($\sigma \sim 10^{-14}\ \text{cm}^2$ at room temperature),[28] while the free carriers offer a worse performance due to nonresonant character of absorption since according to (8) for free carriers $\gamma_{eff} = \gamma + \omega^2/\gamma > 2\omega$. We further analyze our perturbative results by calculating the dependence of the absorption coefficient on the injected (induced by the gate) carriers, N_{2D}, as the elemental free carriers responsible for absorption tuning in the electrostatic modulators as

$$\alpha(\omega) = \frac{4\pi\alpha_0}{n_{eff}t_{eff}}\frac{\hbar}{m_{eff}}\frac{\gamma}{\omega^2+\gamma^2}N_{2D}. \tag{12}$$

Here, $\gamma = 1/\tau$ is the carrier scattering rate, i.e., collision frequency with τ being the mean scattering time corresponding to the mean free path between collisions between the free electrons in the plasma. The electron mobility μ and τ are related by $\mu = |q|\tau/m_{eff}$. We compare two free carrier absorption based materials for our absorption modulators to employ in the following neuromorphic study, namely, silicon and ITO. For Si, the conductivity effective mass, m_{eff}, is taken as $0.26m_0$, where m_0 is the free electron rest mass.[56] μ is taken as $1100\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ at $10^{16}\ \text{cm}^{-3}$ carrier concentration level (i.e., electrons for silicon).[57] We used the dielectric constant of nondoped silicon $\varepsilon_\infty = 11.66$.[58] Unlike the doped silicon, the chemical composition of ITO is usually given as In_2O_3: SnO_2 and can be considered an alloy as concentration of Sn relative to In can be as high as 10%. Several previous studies have calculated the permittivity of ITO using the experimentally measured reflectance and transmittance, and we chose a fitting result of Michelotti *et al.*, whereas γ depends on the deposition conditions, defect states, and film thicknesses, etc.; in our analysis, we have taken $\gamma = 1.8\times10^{14}\ \text{rads}^{-1}$, background permittivity, $\varepsilon_\infty = 3.9$, and $m_{eff} = 0.35m_0$.[38,59–61] In the near-infrared, ITO is quasimetallic since its free electrons dictate its optical response. In fact, ITO and related transparent conducting oxides have recently been explored as plasmonic materials in the optical frequency range.[60,62]

The inherent low broadening parameter γ of Si limits the attainable absorption with a carrier density level variation even at higher concentrations. More importantly, the slope of absorption vs carrier density characteristics is responsible for the modulator performance [i.e., slope steepness ~ Extinction ratio (ER)/V_{bias}], and as such ITO exhibits prominent features as the ITO curve is several orders per decade higher in steepness than the Si one [Fig. 1(b)]. Additionally, we evaluate the total absorption as a function of the injected charge per unit waveguide cross-sectional area setting a 10 dB ER ceiling, $Q_{inj} = qN_{2D}L/t_{eff}$ [Fig. 1(c)]. In accordance to our perturbative estimation, the injected charge follows, as expected, $Q_{inj} \sim 2.2q/\sigma$.[28] Furthermore, the drive voltage can be obtained via $V_d = Q_{inj}/C_g$, where C_g is the gate capacitance [Fig. 1(d)]. A dielectric spacer

© Author(s) 2019

such as an oxide layer for gating facilitation is considered in these analytical formalism. A gate oxide with a relative dielectric constant ε_{eff} of 10 and gate spacer thickness d_{spacer} of 5 nm was chosen for acquiring these results. As the drive voltage is increased, free carriers leading to dispersive effects are induced and a corresponding net increase occurs in the carrier concentration. Our analytic results reveal device scaling for the aforementioned modulation dynamic range of 10 dB ceiling as 33 905 for Si, and 1768 for ITO in units of t_{eff}.

While both ITO and Si are capable of tuning the index via the free carrier dependent Drude model, nevertheless, ITO exhibits disparate significant advantages over Si with respect to the permittivity variation; first, the carrier concentration in ITO can exceed that of Si by at least a couple of orders of magnitude; the concentration of indium atoms in ITO reaches a few percent, which is significantly beyond the attainable donor concentration in doped Si.[63] Second, the effect of inducing the carrier concentration change in ITO on its refractive index is more dramatic than in Si. This can be attributed to the higher bandgap and consequently lower refractive index of ITO compared to that of Si. The higher bandgap allows the density of states (DoS) near band edge to be strongly confined leading to a higher free carrier concentration necessary to change the index. For a discrete spectrum, the DoS consists of a number of delta peaks at the energy levels of the system with weighted degeneracy of that level. If the change of the carrier concentration ∂N_{2D} (e.g., due to an applied bias) causes a change in the relative permittivity (dielectric constant) $\partial\varepsilon$, the corresponding change in the refractive index can be written as $\partial n = \partial\varepsilon^{1/2} \sim \partial\varepsilon/2\varepsilon^{1/2}$, and hence, the refractive index change is greatly enhanced when the permittivity ε is small.[49,62] Third, the presence of an epsilon-near-zero (ENZ) region in the tolerable carrier concentration range within electrostatic gating constraints makes ITO a promising substitute to conventional Si. Operation in the vicinity of the $\varepsilon \sim 0$ (ENZ) condition can understandably result in stronger modulation effects as the optical mode experiences gradually intensifying slow-light effects closer to the ENZ region as a result of increasing light matter interaction thereof. The inherently low background permittivity of ITO allows the ENZ carrier density condition at the telecommunication relevant wavelength of 1550 nm resulting in experimentally obtainable effects ($\simeq 6 - 7 \times 10^{20}$ cm^{-3}).[49,62,63] Interestingly, the Si ENZ condition can be approached at a carrier concentration level of about 3.38×10^{19} cm^{-3}, operating at a wavelength of 10 μm in the IR region away from the region of our interest.

B. Neural network

The model of the ITO absorption mechanism is then incorporated in the NL AF of a neural network. A standard broadcast-and-weight feed-forward network architecture in which connections are configured by microring weight banks,[19] characterized by two layers and 100 nodes per layer, is considered in Fig. 2. The broadcast-and-weight network shown in Fig. 2(b) is used to implement a fully connected neural network whose topology is shown in Fig. 2(a). The network and its functioning have been previously discussed in Refs. 13 and 64 and present similarities with the fiber networking broadcasting techniques. It consists of a group of nodes

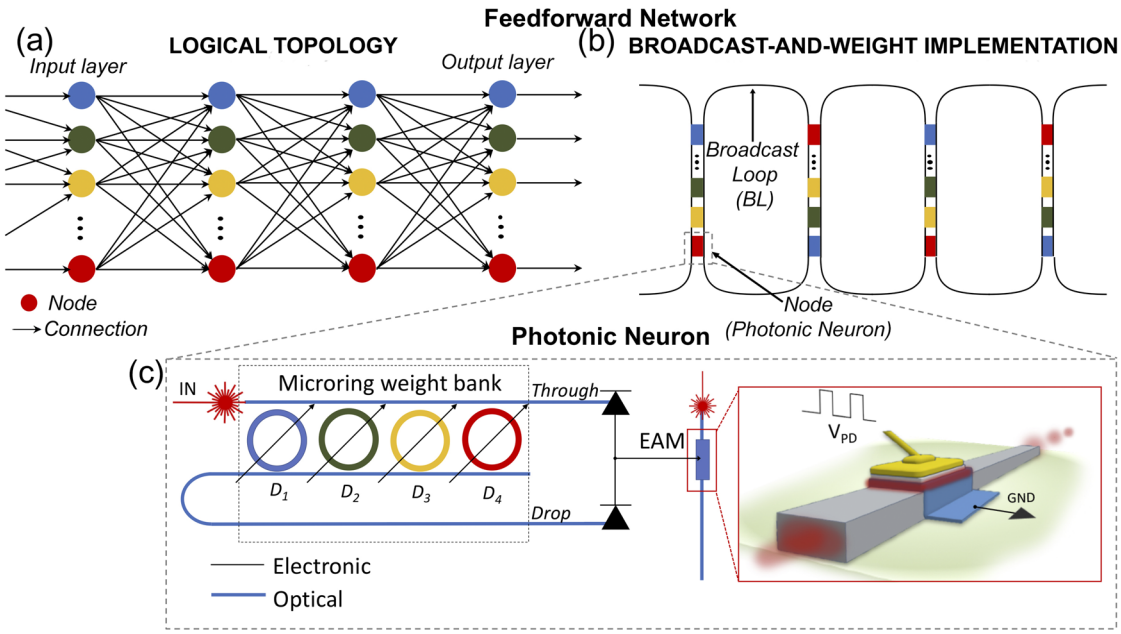

FIG. 2. (a) A topological representation of a fully connected neural network and (b) a photonic feedforward neural network implementation using the broadcast-and-weight protocol. Here, feedforward networks can be constructed using interfacial photonic neurons that connect between broadcast loop (BL) (waveguides).[19] (c) Photonic neuron and perceptron mechanism. WDM Inputs are weighed through tunable microring resonators (MRRs). The optical power is accumulated and detected by a balanced photodiode. The photovoltage drives the electroabsorption modulator, which nonlinearly modulates the laser power mimicking an activation function.

© Author(s) 2019

(N = 100) sharing a common medium, namely, the broadcast loop (BL). The node is the fundamental unit cell of the network which performs both physical and logical functionalities; it is required for both broadcast-and-weight networking and neuromorphic processing, respectively. Each node comprises a tunable spectral filter bank,[65] which can be implemented by a series of microring resonators. The resonance of the filters can be thermally or electronically controlled, allowing for a continuous drop of portion of its corresponding wavelength channel, enabling neural weighting functionalities. The broadcast loop, which could be implemented by a circular fully multiplexed waveguide, implements an all-to-all interconnection, supporting (not necessarily) all N^2 potential connections between participating units. The network can be constructed [Fig. 2(b)] using interfacial photonic neurons that connect between broadcast loops (waveguides).The network architecture footprint can be computed based on metrics of the underlying building blocks, fabricated by accredited foundries and recent literature. On a first approximation, the overall footprint of a neuron bank is estimated to be 1600 μm^2 ($\pi r_{ring}^2 \cdot N$), N being the number of neurons. Assuming that every connection has a dedicated tunable microring resonator (MRR) filter, all critically coupled to the bus waveguide, and that they are single-pole, the N-to-N network footprint is $100 \times 1600\,\mu m^2$ = 0.16 mm^2, which is still on a chip-scale. In our adaptive architecture, which considers the reuse of the bandwidth, the broadcast loop can be of at least 4 cm ($4\,\mu m \times 100^2$). On the other hand, the power budget and noise analysis of the network cannot be mapped trivially (cross talk if thermal tuning is used to control the weights) and the restrictions might be application-dependent and generically loose due to the statistical and intrinsically noisy nature of neuromorphic algorithms. The network is trained to classify 10 individual digits in a set of images of handwritten digits in a grayscale 28 × 28 pixel (MNIST dataset[66]) as a toy example. It is worth noticing that here, each pixel is encoded in a unique wavelength carrier in a wavelength division multiplexed (WDM) and broadcast scheme. Subsequently, the network inference performance are benchmarked. The simulation environment was developed in Python using Keras[67] and TensorFlow.[68] The activation function was replaced with a custom designed nonlinear transfer function resulting from our *ab initio* ITO based electroabsorption modulator model. The weighting mechanism relies on microring resonator banks as proposed by Tait *et al.*[13,25] which can potentially support more than 100 channels.[13] In this case, the weights were bound between minus one and one to simulate input optical weighting by rings in a push-pull configuration.

Noise is modeled in Keras as additive Gaussian noise with the standard deviation proportional to the root-mean square (rms) noise power. Shot noise at the photodiodes is modeled as $\sqrt{2q(I_s + I_d)\Delta f}$, where I_s is the signal current and I_d is the dark current taken as 0.05 nA. In addition, thermal noise is modeled as $\sqrt{4k_B T \Delta f / R_{eq}}$, where T is taken as room temperature of 300 K and R_{eq} is the equivalent resistance of 500 Ω.

1. Training

The neural network models are trained in Keras[67] with the Adagrad[69] method using a categorical cross-entropy loss function, a learning rate of 0.005, zero decay, and with 500 training cycles (epochs) of a 1024 batch size. The amplitude of the noise power model is reduced during training to 10% of the final inference noise power, to allow for training convergence while keeping the model from overfitting to a noiseless activation function.

2. Parameter optimization

The photonic neural network differs from the digital neural network in three ways: First, the noise of the photonic neural network is analog, with sources from shot and thermal noise, and unlike the digital neural network has no quantization noise. Second, the photonic neural network, as an analog system, cascades noise. However, unlike an analog repeater, the analog neural network acts as a signal regenerator when the activation function is pushed into saturation, making the analog neural network a quasidigital system in certain configurations. Finally, the photonic neural network without parameter optimization will shift the bias and dynamic range moving with depth into the network.

To counteract the shifting bias and dynamic range, an optimization procedure was designed to locate the center and range of the activation function at each layer of the network as follows: the activation function is swept over 200 input voltage points with no noise, and the point with the greatest slope is found, a range is built by adding the points to the left and right of the starting point until the slope has dropped below 0.05. After identifying the center of the dynamic range, a DC bias is applied to the modulator to move the minimum input to the center of the identified range.

III. RESULTS AND DISCUSSION

A. ITO electroabsorption activation function: Inference results

An activation function derived from Eq. (12) for both ITO and Si based absorption modulators coupled to a photodiode [Fig. 2(c)] through an ideal transimpedance amplifier was trained in Keras[67] at a simulated bandwidth of 1 GHz. The length for both modulators was taken to be the 10 dB dynamic range point for ITO at 1768 in units of the effective thickness t_{eff} to demonstrate reduced footprint operation in both materials. We keep the scaling

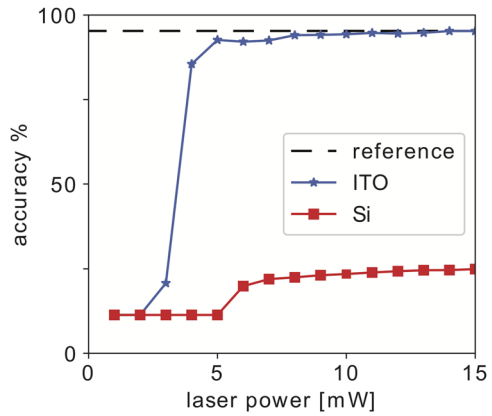

FIG. 3. Simulated MNIST accuracy results for ITO and Si modulator activation functions swept over laser power for a fully connected neural network of two layers, each with 100 nodes.

© Author(s) 2019

independent of the modal structure choice (only material dependent, i.e., Si/ITO) by using the effective thickness as the underlying unit for the device length as t_{eff} can be different for different modes but the physical length required for modulation relates back to t_{eff} based on the modal choice as L/t_{eff} depends only on material constraints (i.e., maximum attainable absorption).[28] The simulation resulted in accuracy greater than 90% for laser power greater than 5 mW (Fig. 3) for the ITO modulator, while the Si modulator never reached accuracy greater than 30% even at the final sweep power of 15 mW. It is worth noticing that here, we compare Si and ITO modulator with the same physical footprint exemplified by the modulator length in effective thickness ($\sim L/t_{eff}$) to achieve 10 dB modulation for the ITO device. Moreover, as aforementioned in Sec. II A, this effect is due to the lower modulation range (ER/V_{bias}) in Si compared to ITO ascribable to the inherent low effective broadening parameter, γ_{eff} of Si which drastically limits the attainable absorption even at higher carrier density levels. As such, a similar accuracy level in executing neural network tasks can be obtained for Si based modulators with substantially larger linear footprint compared to ITO-based electroabsorption modulators, when used as nonlinear activation functions. Even with the enhanced footprint, Si based devices could be susceptible to accuracy challenges as opposed to their highly scalable ITO counterparts because of the steeper dynamic response of the latter [Fig. 1(d)].

B. Experimental characterization of an ITO based electroabsorption modulator device

Using the same approach presented in Ref. 30, we fabricate two different EA modulators with different lengths (5 and 20 μm). The device configuration consists of a Si waveguide (800 nm × 340 nm) and stack, placed on top, which comprises 10 nm ITO layer, gate oxide (SiO_2, t_{ox} = 20 nm), and a metallic gold pad (Fig. 4). The stack was fabricated using electron-beam lithography for defining the pattern, electron-beam evaporation, and lift off processing. In this capacitor configuration, according to the potential applied, we are able to modulate the carrier density of the ITO film, thus tuning the portion of the electric field absorbed by the thin layer. A 1550 nm TM mode traveling in the waveguide is subjected to a shift in the mode profile due to the presence of the plasmonic stack, and accordingly, enhances the mode overlap with the active ITO layer. Increasing the carrier concentration level with active electrostatic gating enhances the modulation effect due to the increased free carrier absorption of the optical mode dynamics as previously discussed. The performances of the characterized electroabsorption modulators display extinction ratios of 5 dB and more than 20 dB for corresponding device lengths of 5 and 20 μm, respectively; when a voltage bias between the ITO and the metal contact is applied. The insertion losses are relatively low and for a device of 5 μm are approximately 1 dB. Going beyond the comparison between the proposed device

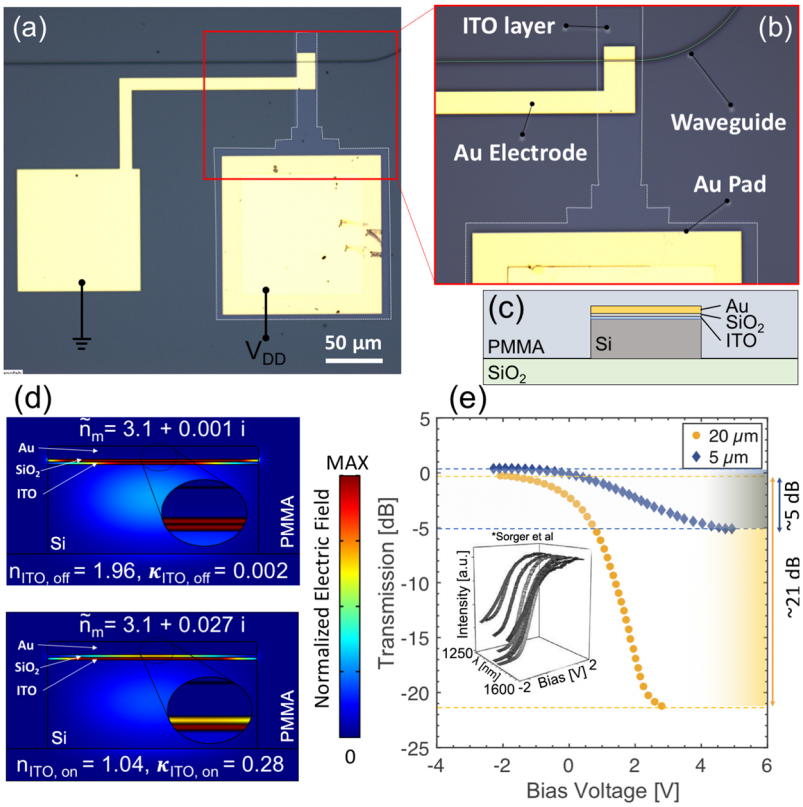

FIG. 4. [(a) and (b)] Optical image of the electroabsorption device and schematic of the cross section (c). A 10 nm ITO film is deposited on a silicon waveguide (800 nm × 340 nm) covered by a thin SiO_2 spacer and a 40 nm Au metal contact. The scale bar is 50 μm. (d) Numerical simulation (FEM) of the normalized electric field distribution for the OFF and ON state, considering a variation of refractive index of the ITO active layer from 1.96 + 0.002i to 1.04 + 0.28i, respectively. (e) Experimentally measured transmission as function of bias voltage (−4 V to 4 V) of a 5 (blue, rhombus) and 20 (yellow, circles) μm ITO based EA modulator. The inset was taken from Ref. 30 and shows the broadband response of similar EAM for different bias voltages between −2 and 2 V.

© Author(s) 2019

TABLE I. Comparison to state-of-the-art silicon and $LiNbO_3$ based traveling-wave modulators, currently employed. The devices are compared in terms of the main figure of merit: Insertion Losses (IL) (dB), linear footprint (μm), Speed (Gbit/s), energy consumption (fJ/bit), Extinction Ratio (ER) (dB). Considering the capacitive effect of our device we found around 300 GHz for a resistance of R = 500 Ω and a 5 μm long device.

Material	Device type	IL (dB)	Linear footprint (μm)	Speed (Gbit/s)	Energy (fJ/bit)	(ER) (dB)
$LiNbO_3$[70]	MZM	2	7000	60	1700	8
Si[71]	MZM	10	10 500	4	...	...
Si[72]	MZM	2	3000	50	450	3
Si[73]	MRR	3	10	40	80	7
Si[74]	MRR	3	10	56	45	4
Si[75]	MRR	1	5	40	4	8
This work	EAM	1	5	Lit. 25[76]	60	5

and free carrier absorption in Si modulators, in Table I, we relate our ITO based EAM to the state-of-the-art silicon and $LiNbO_3$ based traveling-wave modulators, currently employed in integrated photonic circuits. The demonstrated EAM modulator exhibits a competitive figure of merit (FOM); it reaches a modulation range in a rather compact (linear) footprint (~1 dB/μm) and considerably low insertion losses (1 dB). Therefore, this ITO based EAM represents a practical option to other, less performing, Si and $LiNbO_3$ state-of-the-art EO modulators, without necessitating of cavity feedback (e.g., microring resonators) and is, therefore, characterized by a broadband response. In our view, considering the advantageous ER/IL ratio ($\frac{\Delta\alpha}{\alpha}$ = 5) and compact footprint, this EAM is indeed particularly suitable for implementing optical module of a neuron activation function especially for those whose weighting scheme relies on WDM because these EAMs are spectrally broadband (no resonance used).

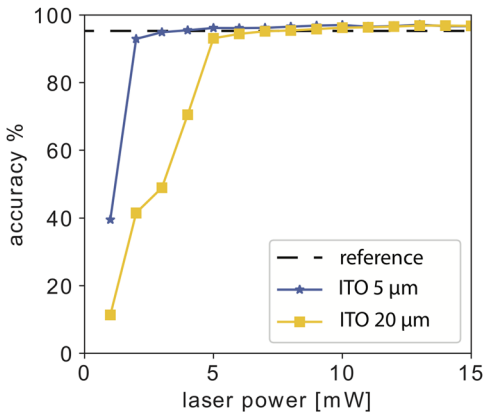

FIG. 5. Simulated MNIST accuracy for activation functions fit to experimental voltage absorption data from two ITO modulators, 5 μm and 20 μm.[30] A 10° clipped polynomial was used to fit the transmittance vs voltage data from each modulator and used as nonlinear activation function in the simulation environment built in tensor flow. Laser power was swept from 1 mW to 15 mW over the fully connected neural network of two layers each with 100 nodes with both modulator types achieving greater than 90% accuracy for laser power ≥5 mW.

Proceeding further with our analysis, we introduce the experimental EA modulator performance as component of the NL activation function in the MNIST classifier of handwritten digits as previously done for our theoretical study. We fit a 10° polynomial with saturation to the experimental data (Fig. 5). In this case with the greater performance of the plasmonic mode, we modeled an activation function with the modulators coupled to the photodiodes through TIAs.[14] The results show greater performance of the 5 μm ITO modulator over the 20 μm ITO modulator for laser power less than 5 mW, while both modulators asymptotically converge to an accuracy greater of 97% as laser power is increased to 15 mW.

IV. CONCLUSIONS

In conclusion, in this work, we described the free carrier absorption dynamics in ITO. The model of the ITO absorption mechanism is then incorporated as a nonlinear activation function of a neural network which comprises 2 layer with 100 nodes each. After an accurate multiparameter optimization, the proposed noise-material capacitor-circuit model shows that the ITO based modulator used as a component of the NL activation mechanism can provide high accuracy (up to 97%) in the inference results in neural networks, which implements handwritten classification prediction tasks characterized by low latency and a power budget in the order of less than 10 W. For the optical platform alone, in a trained network, considering a capacitance of just 7 fF for a 5 μm long device ($\varepsilon\frac{A}{t_{ox}} = \varepsilon_0\varepsilon_{ox}\frac{5\ \mu m \times 0.8\ \mu m}{10\ nm}$) and applied bias change of 4 V, the consumption of one electroabsorption module during the inference phase is assumed approximately to be 60 fJ/bit ($\frac{1}{2}CV^2$). At the system level, neglecting I/O and laser power, this translates into 12 pJ/bit (200 NL AF × 60 fJ/bit) consumption for the NL activation functions. Moreover, we incorporate, in our photonic model of tensor-flow, experimental data of a fabricated ITO based EA modulator showing even higher performance than the ones predicted by the model, which does not account for device asymmetries and material nonidealities. In our vision, these results make ITO a very competitive material for electro-optic neural network, particularly for implementing nonlinear activation function for low latency and power demanding applications such as communications and adaptive control of multiantenna systems in LiDAR.

© Author(s) 2019

SUPPLEMENTARY MATERIAL

See supplementary material for details on the optical properties of ITO and figures of merit of the neural network.

ACKNOWLEDGMENTS

V.S., T.E., and P.P. are supported via the E2CDA program: V.S., T.E., and P.P. under NSF Grant No. 1740262, V.S. and T.E. under SRC nCORE Grant No. 1740235, and P.P. under SRC nCORE Grant No. 2018-NC-2763-A.

NOMENCLATURE

ADC	analog to digital converter
AF	activation function
BL	broadcast loop
DAC	digital to analog converter
DFB	distributed feedback laser
EA	electroabsorption
EAM	electroabsorption modulator
ENZ	epsilon-near-zero
EO	electro-optic
EOM	electro-optic modulator
ER	extinction ratio
DoS	density of states
FEM	finite element modeling
FOM	figure of merit
I/O	input/output
IL	insertion losses
ITO	indium tin oxide
LiDAR	light detection and ranging
MRR	microring resonator
MZM	Mach-Zehnder modulator
NL	nonlinear
QD	quantum dot
RF	Radio frequency
rms	root mean square
SNR	signal to noise ratio
TCO	transparent conductive-oxide
WDM	wavelength division multiplexing

REFERENCES

[1] *Neuromorphic Photonics*, 1st ed., edited by P. R. Prucnal and B. J. Shastri (CRC Press, Boca Raton, FL, 2017).
[2] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Principles of neuromorphic photonics," in *Encyclopedia of Complexity and Systems Science*, edited by R. Meyers (Springer, Berlin, Heidelberg, 2018).
[3] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
[4] W. 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, "Plasmonic IQ modulators with attojoule per bit electrical energy consumption," Nat. Commun. **10**, 1694 (2019).
[5] R. Amin, Z. Ma, R. Maiti, S. Khan, J. B. Khurgin, H. Dalir, and V. J. Sorger, "Attojoule-efficient graphene optical modulators," Appl. Opt. **57**, D130–D140 (2018).
[6] K. Nozaki, A. Shakoor, S. Matsuo, T. Fujii, K. Takeda, A. Shinya, E. Kuramochi, and M. Notomi, "Ultralow-energy electro-absorption modulator consisting of InGaAsP-embedded photonic-crystal waveguide," APL Photonics **2**, 056105 (2017).
[7] Y. Chiu, "Design and fabrication of optical electroabsorption modulator for high speed and high efficiency," in *2015 14th International Conference on Optical Communications and Networks (ICOCN)* (IEEE, Nanjing, 2015), pp. 1–4.
[8] T. F. de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Progress in neuromorphic photonics," Nanophotonics **6**, 577–599 (2017).
[9] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," Nat. Commun. **4**, 1364 (2013).
[10] D. J. Blumenthal, "Integrated combs drive extreme data rates," Nat. Photonics **12**, 447 (2018).
[11] A. Moscoso-Mártir, J. Müller, E. Islamova, F. Merget, and J. Witzens, "Calibrated link budget of a silicon photonics WDM transceiver with SOA and semiconductor mode-locked laser," Sci. Rep. **7**, 12004 (2017).
[12] A. Mehrabian, Y. Al-Kabani, V. J. Sorger, and T. El-Ghazawi, "PCNNA: A photonic convolutional neural network accelerator," in *31st IEEE International System-on-Chip Conference (SOCC), Arlington, VA* (IEEE, 2018), pp. 169–173.
[13] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," J. Lightwave Technol. **32**, 4029–4041 (2014).
[14] J. 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, "Neuromorphic photonics with electro-absorption modulators," Opt. Express **27**, 5181–5191 (2019).
[15] M. Miscuglio, A. Mehrabian, Z. Hu, S. I. Azzam, J. George, A. V. Kildishev, M. Pelton, and V. J. Sorger, "All-optical nonlinear activation function for photonic neural networks," Opt. Mater. Express **8**, 3851–3863 (2018).
[16] A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," Opt. Express **22**, 010868 (2014).
[17] P. Y. Ma, B. J. Shastri, T. F. de Lima, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "All-optical digital-to-spike conversion using a graphene excitable laser," Opt. Express **25**, 033504 (2017).
[18] B. J. Shastri, M. A. Nahmias, A. N. Tait, A. W. Rodriguez, B. Wu, and P. R. Prucnal, "Spike processing with a graphene excitable laser," Sci. Rep. **6**, 19126 (2016).
[19] H. Peng, M. A. Nahmias, T. F. de Lima, A. N. Tait, and B. J. Shastri, "Neuromorphic photonic integrated circuits," IEEE J. Sel. Top. Quantum Electron. **24**, 1–15 (2018).
[20] C. Mesaritakis, A. Kapsalis, A. Bogris, and D. Syvridis, "Artificial neuron based on integrated semiconductor quantum dot mode-locked lasers," Sci. Rep. **6**, 39317 (2016).
[21] K. Alexander, T. V. Vaerenbergh, M. Fiers, P. Mechet, J. Dambre, and P. Bienstman, "Excitability in optically injected microdisk lasers with phase controlled excitatory and inhibitory response," Opt. Express **21**, 26182–26191 (2013).
[22] F. D. Coarer, M. Sciamanna, A. Katumba, M. Freiberger, J. Dambre, P. Bienstman, and D. Rontani, "All-optical reservoir computing on a photonic chip using silicon-based ring resonators," IEEE J. Sel. Top. Quantum Electron. **24**, 1–8 (2018).
[23] W. Cai, A. P. Vasudev, and M. L. Brongersma, "Electrically controlled nonlinear generation of light with plasmonics," Science **333**, 1720–1723 (2011).
[24] M. Kauranen and A. V. Zayats, "Nonlinear plasmonics," Nat. Photonics **6**, 737–748 (2012).
[25] A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," Sci. Rep. **7**, 7430 (2017).
[26] J. George, R. Amin, A. Mehrabian, J. Khurgin, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks," in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)* (Optical Society of America, 2018), paper SpW4G.3.
[27] S. Yoo, R. Kim, J.-H. Park, and Y. Nam, "Electro-optical neural platform integrated with nanoplasmonic inhibition interface," ACS Nano **10**, 4274–4281 (2016).

© Author(s) 2019

[28]R. Amin, J. B. Khurgin, and V. J. Sorger, "Waveguide-based electro-absorption modulator performance: Comparative analysis," Opt. Express **26**, 15445–15470 (2018).
[29]R. F. Oulton, V. J. Sorger, D. A. Genov, D. F. P. Pile, and X. Zhang, "A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation," Nat. Photonics **2**, 496–500 (2008).
[30]V. J. Sorger, N. D. Lanzillotti-Kimura, R.-M. Ma, and X. Zhang, "Ultra-compact silicon nanophotonic modulator with broadband response," Nanophotonics **1**, 17–22 (2012).
[31]C. 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. Shalaev, N. Kinsey, and J. Leuthold, "Low-loss plasmon-assisted electro-optic modulator," Nature **556**, 483 (2018).
[32]G. A. Keeler, K. M. Geib, D. K. Serkland, S. Parameswaran, T. S. Luk, T. S. Luk, A. J. Griñe, J. Ihlefeld, S. Campione, and J. R. Wendt, "Multi-gigabit operation of a compact, broadband modulator based on ENZ confinement in indium oxide," in *Optical Fiber Communication Conference* (Optical Society of America, 2017), paper Th3I.1.
[33]S. Zhu, G. Q. Lo, and D. L. Kwong, "Phase modulation in horizontal metal-insulator-silicon-insulator-metal plasmonic waveguides," Opt. Express **21**, 8320–8330 (2013).
[34]Z. Wu, R. L. Nelson, J. W. Haus, and Q. Zhan, "Plasmonic electro-optic modulator design using a resonant metal grating," Opt. Lett. **33**, 551–553 (2008).
[35]X. Gan, R.-J. Shiue, Y. Gao, K. F. Mak, X. Yao, L. Li, A. Szep, D. Walker, J. Hone, T. F. Heinz, and D. Englund, "High-contrast electrooptic modulation of a photonic crystal nanocavity by electrical gating of graphene," Nano Lett. **13**, 691–696 (2013).
[36]R. M. Briggs, J. Grandidier, S. P. Burgos, E. Feigenbaum, and H. A. Atwater, "Efficient coupling between dielectric-loaded plasmonic and silicon photonic waveguides," Nano Lett. **10**, 4851–4857 (2010).
[37]H. W. Lee, G. Papadakis, S. P. Burgos, K. Chander, A. Kriesch, R. Pala, U. Peschel, and H. A. Atwater, "Nanoscale conducting oxide PlasMOStor," Nano Lett. **14**, 6463–6468 (2014).
[38]A. P. Vasudev, J.-H. Kang, J. Park, X. Liu, and M. L. Brongersma, ""Electro-optical modulation of a silicon waveguide with an 'epsilon-near-zero' material," Opt. Express **21**, 26387–26397 (2013).
[39]V. E. Babicheva, A. Boltasseva, and A. V. Lavrinenko, "Transparent conducting oxides for electro-optical plasmonic modulators," Nanophotonics **4**, 165–185 (2015).
[40]M. Y. Abdelatty, M. M. Badr, and M. A. Swillam, "Hybrid plasmonic electro-optical absorption modulator based on epsilon-near-zero characteristics of ITO," in *Integrated Optics: Devices, Materials, and Technologies XXII* (International Society for Optics and Photonics, 2018), Vol. 10535, p. 105351T.
[41]M. Z. Alam, I. D. Leon, and R. W. Boyd, "Large optical nonlinearity of indium tin oxide in its epsilon-near-zero region," Science **352**, 795–797 (2016).
[42]H. Zhao, Y. Wang, A. Capretti, L. D. Negro, and J. Klamkin, "Silicon photonic modulators based on epsilon-near-zero indium tin oxide materials," in *2014 IEEE Photonics Conference* (IEEE, 2014), pp. 334–335.
[43]L. Huang and J. Z. Zhong, "Indium tin oxide (ITO) layer forming," U.S. patent 8049862b2 (2011).
[44]L. Kerkache, A. Layadi, E. Dogheche, and D. Rémiens, "Physical properties of RF sputtered ITO thin films and annealing effect," J. Phys. D: Appl. Phys. **39**, 184 (2006).
[45]C. Chityuttakan, P. Chinvetkitvanich, S. Chatraphorn, and S. Chatraphorn, "Influence of deposition parameters on the quality of ITO films for photovoltaic application," AIP Conf. Proc. **2091**, 3 (2019).
[46]Y. Hu, X. Diao, C. Wang, W. Hao, and T. Wang, "Effects of heat treatment on properties of ITO films prepared by rf magnetron sputtering," Vacuum **75**, 183–188 (2004).
[47]M. T. Bhatti, A. M. Rana, and A. F. Khan, "Characterization of rf-sputtered indium tin oxide thin films," Mater. Chem. Phys. **84**, 126–130 (2004).
[48]H.-C. Lee and O. Ok Park, "Behaviors of carrier concentrations and mobilities in indium–tin oxide thin films by DC magnetron sputtering at various oxygen flow rates," Vacuum **77**, 69–77 (2004).
[49]R. Amin, R. Maiti, C. Carfano, Z. Ma, M. H. Tahersima, Y. Lilach, D. Ratnayake, H. Dalir, and V. J. Sorger, "0.52 V mm ITO-based Mach-Zehnder modulator in silicon photonics," APL Photonics **3**, 126104 (2018).
[50]Y. Gui, M. Miscuglio, Z. Ma, M. H. Tahersima, S. Sun, R. Amin, H. Dalir, and V. J. Sorger, "Towards integrated metatronics: a holistic approach on precise optical and electrical properties of indium tin oxide," Sci. Rep. **9**, 11279 (2019).
[51]Y. Tang, J. D. Peters, and J. E. Bowers, "Over 67 GHz bandwidth hybrid silicon electroabsorption modulator with asymmetric segmented electrode for 1.3 μm transmission," Opt. Express **20**, 11529–11535 (2012).
[52]X. Tu, T.-Y. Liow, J. Song, M. Yu, and G. Q. Lo, "Fabrication of low loss and high speed silicon optical modulator using doping compensation method," Opt. Express **19**, 18029–18035 (2011).
[53]R. Soref and B. Bennett, "Electrooptical effects in silicon," IEEE J. Quantum Electron. **23**, 123–129 (1987).
[54]G. T. Reed, G. Z. Mashanovich, F. Y. Gardes, M. Nedeljkovic, Y. Hu, D. J. Thomson, K. Li, P. R. Wilson, S.-W. Chen, and S. S. Hsu, "Recent breakthroughs in carrier depletion based silicon optical modulators," Nanophotonics **3**, 229–245 (2013).
[55]C. 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, "All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale," Nat. Photonics **9**, 525–528 (2015).
[56]S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices* (John Wiley & Sons, 2006), google-Books-ID: o4unkmHBHb8C.
[57]S. Nashima, O. Morikawa, K. Takata, and M. Hangyo, "Measurement of optical properties of highly doped silicon by terahertz time domain reflection spectroscopy," Appl. Phys. Lett. **79**, 3923–3925 (2001).
[58]J. D. Holm and K. S. Champlin, "Microwave conductivity of silicon and germanium," J. Appl. Phys. **39**, 275–284 (1968).
[59]F. Michelotti, L. Dominici, E. Descrovi, N. Danz, and F. Menchini, "Thickness dependence of surface plasmon polariton dispersion in transparent conducting oxide films at 1.55 μm," Opt. Lett. **34**, 839–841 (2009).
[60]M. A. Noginov, L. Gu, J. Livenere, G. Zhu, A. K. Pradhan, R. Mundle, M. Bahoura, Y. A. Barnakov, and V. A. Podolskiy, "Transparent conductive oxides: Plasmonic materials for telecom wavelengths," Appl. Phys. Lett. **99**, 021101 (2011).
[61]G. V. Naik, V. M. Shalaev, and A. Boltasseva, "Alternative plasmonic materials: Beyond gold and silver," Adv. Mater. **25**, 3264–3294 (2013).
[62]R. Amin, M. H. Tahersima, Z. Ma, C. Suer, K. Liu, H. Dalir, and V. J. Sorger, "Low-loss tunable 1D ITO-slot photonic crystal nanobeam cavity," J. Opt. **20**, 054003 (2018).
[63]R. Amin, C. Suer, Z. Ma, I. Sarpkaya, J. B. Khurgin, R. Agarwal, and V. J. Sorger, "A deterministic guide for material and mode dependence of on-chip electro-optic modulator performance," Solid-State Electron. **136**, 92–101 (2017).
[64]A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H. B. Miller, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Silicon photonic modulator neuron," Phys. Rev. Appl. **11**, 064043 (2019).
[65]A. N. Tait, A. X. Wu, T. F. de Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Microring weight banks," IEEE J. Sel. Top. Quantum Electron. **22**, 312–325 (2016).
[66]Y. LeCun and C. Cortes, MNIST handwritten digit database, 2010, http://yann.lecun.com/exdb/mnist/.
[67]F. Chollet *et al.*, Keras, 2015, https://keras.io.
[68]M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, software available from tensorflow.org.
[69]J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," Technical Report UCB/EECS-2010-24, EECS Department, University of California, Berkeley, 2010.

© Author(s) 2019

[70]C. Wang, M. Zhang, B. Stern, M. Lipson, and M. Loncar, "Nanophotonic lithium niobate electro-optic modulators," Opt. Express **26**, 1547–1555 (2018).
[71]D. Patel, M. Parvizi, N. Ben-Hamida, C. Rolland, and D. V. Plant, "Frequency response of dual-drive silicon photonic modulators with coupling between electrodes," Opt. Express **26**, 8904 (2018).
[72]M. Streshinsky, R. Ding, Y. Liu, A. Novack, Y. Yang, Y. Ma, X. Tu, E. K. S. Chee, A. E.-J. Lim, P. G.-Q. Lo, T. Baehr-Jones, and M. Hochberg, "Low power 50 Gb/s silicon traveling wave Mach-Zehnder modulator near 1300 nm," Opt. Express **21**, 030350 (2013).
[73]G. Li, X. Zheng, H. Thacker, J. Yao, Y. Luo, I. Shubin, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, "40 Gb/s thermally tunable CMOS ring modulator," in *The 9th International Conference on Group IV Photonics GFP* (IEEE, San Diego, CA, USA, 2012), pp. 1–3.
[74]M. Pantouvaki, P. Verheyen, J. De Coster, G. Lepage, P. Absil, and J. Van Campenhout, "56gb/s ring modulator on a 300mm silicon photonics platform," in *2015 European Conference on Optical Communication (ECOC)* (IEEE, Valencia, Spain, 2015), pp. 1–3.
[75]E. Timurdogan, C. M. Sorace-Agaskar, J. Sun, E. Shah Hosseini, A. Biberman, and M. R. Watts, "An ultralow power athermal silicon modulator," Nat. Commun. **5**, 4008 (2014).
[76]J.-S. Kim and J. T. Kim, "Silicon electro-optic modulator based on an ITO-integrated tunable directional coupler," J. Phys. D: Appl. Phys. **49**, 075101 (2016).

© Author(s) 2019