Silicon Photonic Modulator Neuron

Alexander N. Tait,*,† Thomas Ferreira de Lima, Mitchell A. Nahmias, Heidi B. Miller,‡
Hsuan-Tung Peng, Bhavin J. Shastri,‡ and Paul R. Prucnal
Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA

(Received 9 December 2018; revised manuscript received 3 February 2019; published 18 June 2019)

There has been recent interest in neuromorphic photonics, a field with the promise to access pivotal and unexplored regimes of machine intelligence. Progress has been made on isolated neurons and analog interconnects; nevertheless, this renewal of interest has yet to produce a demonstration of a silicon photonic neuron capable of interacting with other like neurons. We report a modulator-class photonic neuron fabricated in a conventional silicon photonic process line. We demonstrate the behaviors of transfer-function configurability, fan-in, inhibition, time-resolved pulse processing, and, crucially, autaptic cascadability—a su cient set of behaviors for a device to act as a neuron participating in a network of like neurons. The silicon photonic modulator neuron constitutes the final piece needed to make photonic neural networks fully integrated on currently available silicon photonic platforms.

DOI: 10.1103/PhysRevApplied.11.064043

I. INTRODUCTION

Renewed interest in neuromorphic photonics has been heralded by advances in photonic integration technology [1–3], roadblocks in conventional computing performance [4,5], the return of neuromorphic electronics [6–10], and the inundation of machine learning (ML) with neural models [11]. Neural networks have played some role in ML (e.g., image and voice recognition, language translation, pattern detection, and others) since the 1950s [12,13]. They fell out of favor in the 1990s because they are di cult to train.

Over the past decade, neural network models have decisively retaken the helm of ML under the alias of "deep networks" [14]. There are three main reasons: (1) major algorithmic innovations [15,16], (2) the Internet—an inexhaustible source of millions of training examples—and (3) new hardware, specifically graphical processing units (GPUs) [17]. Central processing units (CPUs) are woefully ine cient at evaluating these models because they are centralized and instruction based, whereas networks are distributed and capable of adaptation without a programmer. GPUs are more parallel but, today, even they have been pushed to their limits [18].

Today's demand for evaluating neural network models necessitates new hardware. High-tech behemoths and research agencies—notably IBM [6], HP [19], Intel [10], Google [20,21], the Human Brain Project [22], and DARPA SyNAPSE [23]—have invested heavily in massively parallel application-specific integrated circuits (ASICs) for evaluating neural network models more e ciently. Some of these architectures aim to be ML number crunchers [20,24] and others have enabled novel neuroscientific tools [25,26] and previously unforeseen low-power mobile applications [27].

Moving beyond the nanosecond will require beyond purely electronic physics.

Photonic physics exhibit properties distinct from those of electronics in terms of multiplexing, energy dissipation, and cross talk. These properties are favorable for dense, high-bandwidth interconnects [29] in addition to configurable analog signal processing [30–32]. Consequently, neuromorphic photonic systems could operate 6–8 orders of magnitude faster than neuromorphic electronics [33]

---

*atait@ieee.org
†Now at National Institute of Standards and Technology, Boulder, CO 80305, USA.
‡Also at Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston ON, K7L 3N6, Canada.

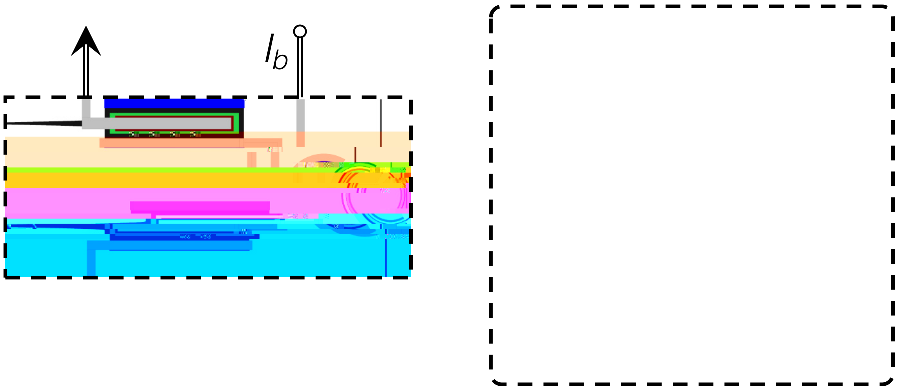
with potentially higher energy efficiencies [34]. Neural interconnects based on field evolution in free space [35, 36], holograms [37,38], and fiber [39] have been demonstrated but have not been widely adopted, in part because they cannot be integrated and thereby scaled robustly and manufactured cheaply. Analog interconnects integrated on

II. METHODS

A. Device description

1. T e f e a

The modulator neuron is an optical-to-electrical-to-optical (O-E-O) device consisting of two photodetectors


I_b

There are three signal generators used in the following experiments, two analog (a.k.a. synths) and one binary. A simple slow-wave-form generator (HP 8116A) is used to acquire the transfer functions (Sec. III A) and the autapse behavior (Sec. III E). The 8116A o ers control of saw-tooth wave forms that can be used to separate rising and falling aspects. Burst inputs are generated by a Rohde and Schwartz SMBV 100A VG (R&S), which is used in Secs. III A, B, and C. The R&S burst can also be viewed as trains of return-to-zero (RZ) pulses of varying amplitude. The binary-pulsed inputs used in Sec. III D are generated by a pulse-pattern generator (PPG) (Anritsu MP1761B). The PPG provides the highest instantaneous bandwidth but the least control over wave forms.

The neuron's output is coupled o -chip, detected, and observed in a sampling oscilloscope (Tektronix DSA8300). Between the output coupler and the oscilloscope, there is a signal-to-noise enhancement stage, not diagrammed, consisting of an erbium-doped fiber amplifier (EDFA), an optical band-pass filter at λ_n, a is widely used in feedforward-machine-learning networks today, i.e., in multilayer perceptrons (MLPs) and convolutional neural networks (CNNs) [20]. Positive and negative ReLUs are obtained by biasing slightly off resonance, either above or below the pump wavelength. A network that combines sigmoid and ReLU neurons is well suited to solving nonlinear optimization problems with constraints, some of which are reviewed in Ref. [69]. The peaked transfer functions of Figs. 5(e) and 5(f) are known as radial basis functions (RBFs). When biased on resonance, the RBF is centered at zero, resulting in a quadratic or rectifying transfer function. The off-centered RBF is obtained by setting the electrical bias to achieve the highest resonator Q

is central to the idea of network-based processing, so it is particularly important to demonstrate this feature directly

feedback signal to influence the same neuron. In other words, the experiment also verifies the presence of a fan-in mechanism.

There are several works that have successfully approached cascadability by avoiding all-optoelectronic signal pathways and instead using an O-E-O chain consisting of a photodetector connected to a laser [48,81,82] or modulator [53,71]. Wavelength constraints and phase sensitivity vanish because this information is lost in the electronic domain. In addition, the E-O conversion step can o er strong nonlinearity, as employed here, and the electronic domain itself o ers e cient mechanisms for nonlinearity and amplification. In Ref. [83], an O-E-O neuron based on cryogenic silicon LEDs, superconducting detectors, and superconducting amplifiers [84] was proposed. Its physical cascadability was demonstrated by the E-O-E LED-detector link shown in Ref. [85], and its gain cascadability has been addressed in more recent simulation works [81,82]. A potential downside of O-E-O is a vulnerability to electrical parasitics; however, these parasitics can remain small regardless of the network scale because O-E-O occurs entirely within a neuron, not between neurons.

When light combines, it interferes, posing a fundamental challenge to fan-in [75]. Optical fan-in results in either phase dependence, when coherent, or N-fold loss, when incoherent (e.g., 3 dB at $N = 2$). In some all-optical devices where the in-out wavelengths can be the same (cascadable), these wavelengths also must be the same, meaning they cannot have more than one input [44,79]. Fan-in with coherent signals can be achieved by exerting complete control of the optical phase in the interconnect [41] but then signal-dependent phase changes in a neuron profoundly a ect the behavior of the subsequent interconnect, precluding any cascadability. In Ref. [41], neuron calculations were implemented at low speed in a CPU; a neuron based on saturable absorption was contemplated, but it was not discussed how this element would regenerate a consistent optical phase.

Fan-in has also been achieved using inputs that are coherent but mutually incoherent, such as di erent spatial modes [86,87], di erent polarizations, or di erent wavelengths [66,88,89]. These signals do not interfere and, since they are individually coherent, can be multiplexed and routed and/or weighted independently by tunable resonators [58,90]. The total power is sensed by a photodetector (O-E), making this fan-in approach compatible with the O-E-O approach to cascadability. Multiwavelength-weighted addition was combined with O-E-O laser neurons in Ref. [48,50], wherein cascadability was also considered but not directly demonstrated. A downside of relying on multiple wavelengths is the need for a di erent laser source for each channel. The size of a single all-to-all subnetwork is capped by the available spectrum and the ability to distinguish adjacent channels, found in Ref. [91] to be less than 950 if using the resonators of Ref. [92]; however, multiples of these subnetworks could be interfaced on a single chip [55].

B. Nonspiking photonic neurons

The great majority of the work on photonic neurons has focused on lasers that implement spiking models similar to biological neurons [42–50], reviewed in Ref. [51]. To claim a nonspiking modulator as a photonic neuron r

There is no fundamental reason why photonics must

is met when the neuron’s di erential optical-to-optical gain, g, exceeds unity.

1. Theor

The gain g can be derived from the device properties. It is defined as follows:

$$g = \frac{dP_{\text{out}}}{dP_{\text{in}}} \tag{A1}$$

$$= \frac{dP_{\text{out}}}{dT_{\text{mod}}} \frac{dT_{\text{mod}}}{dV} \frac{dV}{dP_{\text{in}}}, \tag{A2}$$

where T_{mod} is modulator transmission, and V is the junction which crosses zero when

$$P_{\mathrm{pump}}|_{J=0} = \frac{2V}{R_{\mathrm{pd}} R_b}. \tag{A16}$$

Thus, the expression for the pump power where the autapse loses monostability corresponds exactly with that where

learning under hardware constraints—using a RISC processor embedded in a neuromorphic substrate, Front. Neurosci. **7**, 00 (2013).
[27] W. Y. Tsai, D. Barch, A. Cassidy, M. Debole, A. Andreopoulos, B. Jackson, M. Flickner, J. Arthur, D. Modha, J. Sampson, and V. Narayanan, Always-on speech recognition using TrueNorth, a reconfigurable, neurosynaptic processor, IEEE Trans. Comput. **66**, 996 (2016).
[28] Andrew Mundy, Ph.D. thesis, School University of Manchester, School of Computer Science, 2016.
[29] S. Rakheja and V. Kumar, in *2012 13th International Symposium on Quality Electronic Design (ISQED)* (IEEE, Santa Clara, CA, USA, 2012), p. 732.
[30] Andrew M. Weiner, Ultrafast optical pulse shaping: A tutorial review, Opt. Commun. **284**, 3669 (2011).
[31] Daniel Pérez, Ivana Gasulla, Lee Crudgington, David J. Thomson, Ali Z. Khokhar, Ke Li, Wei Cao, Goran Z. Mashanovich, and José Capmany, Multipurpose silicon photonics signal processor core, Nat. Commun. **8**, 636 (2017).
[32] Weilin Liu, Ming Li, Robert S. Guzzon, Erik J. Norberg, John S. Parker, Mingzhi Lu, Larry A. Coldren, and Jianping Yao, A fully reconfigurable photonic integrated signal processor, Nat. Photon. **10**, 190 (2016).
[33] Bhavin J. Shastri, Alexander N. Tait, Thomas Ferreira de Lima, Mitchell A. Nahmias, Hsuan-Tung Peng, and Paul R. Prucnal, Principles of Neuromorphic Photonics, arXiv:1801.00016 (2018).
[34] Thomas Ferreira de Lima, Bhavin J. Shastri, Alexander N. Tait, Mitchell A. Nahmias, and Paul R. Prucnal, Progress in neuromorphic photonics, Nanophotonics **6**, 00 (2017).
[35] Daniel Brunner and Ingo Fischer, Reconfigurable semiconductor laser networks based on di ractive coupling, Opt. Lett. **40**, 3854 (2015).
[36] Demitri Psaltis and Yong Quio, Optical neural networks, Opt. Photon. News **1**, 17 (1990).
[37] J. W. Goodman, A. R. Dias, and L. M. Woody, Fully parallel, high-speed incoherent optical method for performing discrete fourier transforms, Opt. Lett. **2**, 1 (1978).
[38] Praveen Asthana, Gregory P. Nordin, Armand R. Tanguay, Jr., and B. Keith Jenkins, Analysis of weighted fan-out/fan-in volume holographic optical interconnections, Appl. Opt. **32**, 1441 (1993).
[39] M. T. Hill, E. E. E. Frietman, H. de Waardt, G.-D. Khoe, and H. J. S. Dorren, All fiber-optic neural network using coupled SOA based ring lasers, IEEE Trans. Neural Netw. **13**, 1504 (2002).
[40] Jacques Carolan, Christopher Harrold, Chris Sparrow, Enrique Martń-López, Nicholas J. Russell, Joshua W. Silverstone, Peter J. Shadbolt, Nobuyuki Matsuda, Manabu Oguma, Mikitaka Itoh, Graham D. Marshall, Mark G. Thompson, Jonathan C. F. Matthews, Toshikazu Hashimoto, Jeremy L. O'Brien, and Anthony Laing, Universal linear optics, Science **349**, 711 (2015).
[41] Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljačić, Deep learning with coherent nanophotonic circuits, Nat. Photon. **11**, 441 (2017).
[42] M. A. Nahmias, B. J. Shastri, A. N. Tait, and P. R. Prucnal, A leaky integrate-and-fire laser neuron for ultrafast cognitive computing, IEEE J. Sel. Top. Quantum Electron. **19**, 1 (2013).
[43] W. Coomans, L. Gelens, S. Beri, J. Danckaert, and G. Van der Sande, Solitary and coupled semiconductor ring lasers as optical spiking neurons, Phys. Rev. E **84**, 036209 (2011).
[44] Thomas Van Vaerenbergh, Koen Alexander, Joni Dambre, and Peter Bienstman, Excitation transfer between optically injected microdisk lasers, Opt. Express **21**, 28922 (2013).
[45] Maia Brunstein, Alejandro M. Yacomotti, Isabel Sagnes, Fabrice Raineri, Laurent Bigot, and Ariel Levenson, Excitability and self-pulsing in a photonic crystal nanocavity, Phys. Rev. A **85**, 031803 (2012).
[46] F. Selmi, R. Braive, G. Beaudoin, I. Sagnes, R. Kuszelewicz, and S. Barbay, Relative Refractory Period in an Excitable Semiconductor Laser, Phys. Rev. Lett. **112**, 183902 (2014).
[47] B. Romeira, R. Avó, José M. L. Figueiredo, S. Barland, and J. Javaloyes, Regenerative memory in time-delayed neuromorphic photonic resonators, Sci. Rep. **6**, 19510 (2016).
[48] Mitchell A. Nahmias, Alexander N. Tait, Leonidas Tolias, Matthew P. Chang, Thomas Ferreira de Lima, Bhavin J. Shastri, and Paul R. Prucnal, An integrated analog O/E/O link for multi-channel laser neurons, Appl. Phys. Lett. **108**, 151106 (2016).
[49] T. Deng, J. Robertson, and A. Hurtado, Controlled propagation of spiking dynamics in vertical-cavity surface-emitting lasers: Towards neuromorphic photonic networks, IEEE J. Sel. Top. Quantum Electron. **23**, 1 (2017).
[50] H. T. Peng, M. A. Nahmias, T. Ferreira de Lima, A. N. Tait, B. J. Shastri, and P. Prucnal, Neuromorphic photonic integrated circuits, IEEE J. Sel. Top. Quantum Electron. **24**, 00 (2018).
[51] Paul R. Prucnal, Bhavin J. Shastri, Thomas Ferreira de Lima, Mitchell A. Nahmias, and Alexander N. Tait, Recent progress in semiconductor excitable lasers for photonic spike processing, Adv. Opt. Photon. **8**, 228 (2016).
[52] Bhavin J. Shastri, Mitchell A. Nahmias, Alexander N. Tait, Alejandro W. Rodriguez, Ben Wu, and Paul R. Prucnal, Spike processing with a graphene excitable laser, Sci. Rep. **6**, 19126 (2016).
[53] Alexander N. Tait, Thomas Ferreira de Lima, Ellen Zhou, Allie X. Wu, Mitchell A. Nahmias, Bhavin J. Shastri, and Paul R. Prucnal, Neuromorphic photonic networks using silicon photonic weight banks, Sci. Rep. **7**, 7430 (2017).
[54] Paul R. Prucnal and Bhavin J. Shastri, *Neuromorphic Photonics* (CRC Press, CRC Press. Boca Raton, FL, USA, 2017).
[55] 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 (2014).
[56] Alexander N. Tait, Hasitha Jayatilleka, Thomas Ferreira De Lima, Philip Y. Ma, Mitchell A. Nahmias, Bhavin J. Shastri, Sudip Shekhar, Lukas Chrostowski, and Paul R. Prucnal, Feedback control for microring weight banks, Opt. Express **26**, 26422 (2018).
[57] A. N. Tait, A. X. Wu, T. Ferreira de Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, Microring

weight banks, IEEE J. Sel. Top. Quantum Electron. **22**, 00 (2016).
[58] Alexander N. Tait, Thomas Ferreira de Lima, Mitchell A. Nahmias, Bhavin J. Shastri, and Paul R. Prucnal, Multichannel control for microring weight banks, Opt. Express **24**, 8895 (2016).
[59] R. Soref and B. Bennett, Electrooptical e ects in silicon, IEEE J. Quantum Electron. **23**, 123 (1987).
[60] Richard Soref, Silicon photonics: A review of recent literature, Silicon **2**, 1 (2010).
[61] Lukas Chrostowski and Michael Hochberg, *Silicon Photonics Design: From Devices to Systems* (Cambridge University Press, Cambridge, United Kingdom, 2015).
[62] Tom Baehr-Jones, Ran Ding, Ali Ayazi, Thierry Pinguet, Matt Streshinsky, Nick Harris, Jing Li, Li He, Mike Gould, Yi Zhang, Andy Eu-Jin Lim, Tsung-Yang Liow, Selin Hwee-Gee Teo, Guo-Qiang Lo, and Michael Hochberg, A 25 Gb/s Silicon Photonics Platform, arXiv:1203.0767 (2012).
[63] Monireh Moayedi Pour Fard, Glenn Cowan, and Odile Liboiron-Ladouceur, Responsivity optimization of a high-speed germanium-on-silicon photodetector, Opt. Express **24**, 27738 (2016).
[64] Qianfan Xu, Sasikanth Manipatruni, Brad Schmidt, Jagat Shakya, and Michal Lipson, 12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators, Opt. Express **15**, 430 (2007).
[65] A. Novack, Y. Liu, R. Ding, M. Gould, T. Baehr-Jones, Q. Li, Y. Yang, Y. Ma, Y. Zhang, K. Padmaraju, K. Bergmen, A. E. J. Lim, G. Q. Lo, and M. Hochberg, in *10th International Conference on Group IV Photonics* (IEEE, Seoul, South Korea, 2013), p. 7.
[66] Alexander N. Tait, John Chang, Bhavin J. Shastri, Mitchell A. Nahmias, and Paul R. Prucnal, Demonstration of WDM weighted addition for principal component analysis, Opt. Express **23**, 12758 (2015).
[67] Alex Tait, Thomas Ferreira de Lima, Philip Y. Ma, Aashu Jha, Hsuan-Tung Peng, Heidi Miller, and Paul R. Prucnal, lightwave-lab/LIGHTLAB: Version 1.0.5 (2018).
[68] J. J. Hopfield and D. W. Tank, “Neural” computation of decisions in optimization problems, Biol. Cybern. **52**, 141 (1985).