Neuromorphic Photonic Networks

Bhavin J. Shastri[1,2], Simon Bilodeau[2], Bicky A. Marquez[1], Alexander N. Tait[2], Thomas Ferreira de Lima[2], Chaoran Huang[2], Lukas Chrostowski[3], Sudip Shekhar[3], and Paul R. Prucnal[2]

[1]Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada
[2]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
[3]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
shastri@ieee.org

Abstract: Neuromorphic photonics exploit optical device physics for neuron models, and optical interconnects for distributed, parallel, and analog processing for high-bandwidth, low-latency and low switching energy applications in artificial intelligence and neuromorphic computing.

OCIS codes: (200.3050) Information processing; (200.4700) Optical neural systems; (250.5300) Photonic integrated circuits.

Neural networks have enjoyed renewed popularity over the last decade under the appellation of "deep learning"[1], [2]. The idea of mimicking the brain to process information, however, can be traced back half a century prior to Rosenblatt's perceptron [3], and the first experimental models of biological neurons to Hodgkin and Huxley a few years prior [4]. The artificial neurons that make up neural networks take many forms, some more closely related to this biological inspiration. Yet all neural networks take the form of simple nodes that (a) perform a linear operation on multiple other neurons' outputs, (b) integrate the resulting signals, and (c) perform a nonlinear transformation on the summed, weighted inputs. Various interconnection topologies – feedforward, feedback (recurrent), close-neighbor translationally-invariant (convolutional), etc. – endow the network with different computational properties.

Such an asynchronous, parallel framework is at odds with the digital von Neumann architecture that electronic microprocessors often employ for their emulation. This mismatch was recognized early on, leading to pioneering work by VLSI engineers starting in the 1980's to map the physics of transistors to neuronal models for gains in computational density, energy efficiency, and speed [5]. However, Moore's Law and Dennard scaling kept such "neuromorphic" architecture outside of the limelight in favor of general-purpose digital processors. Today, this scaling nears its end, and researchers turn to ever more specialized hardware such as graphical processing units [6], tensor processing units [7] and specially configured field-programmable gate arrays [8] to run demanding neural network models. This is renewing interest in neuromorphic application-specific integrated circuits (ASICs), the extrapolated conclusion of this trend.

Since the requirements of neuromorphic hardware differ from von Neumann digital computing, it is not obvious that silicon microelectronics must provide the best substrate for neuromorphic ASICs [9], [10]. The reliance of neural networks on simple networked nodes suggests that a platform suited for communications, such as photonics, might have an advantage. This was recognized in the 80's [11], yet the lack of integrability limited investigations at the time. The commercial silicon photonic platforms that have arisen over the last few years, however, now offer high index contrast, low-loss waveguides integrated with high bandwidth optoelectronics for signal modulation and detection [12]. Furthermore, the reuse of materials and processes from microelectronics allows the platform to enjoy its economies of scale. This, combined with the intrinsic appeal of photonics to emulate neural models, is one of the reasons that the newly termed field of *neuromorphic photonics* has attracted considerable attention [13]-[22].

In this talk, we will provide an overview of current neuromorphic silicon photonics architectures including coherent approaches based on Mach-Zehnder interferometers [15], [23] and incoherent (multiwavelength) optoelectronic approaches based on microring resonators [16], [17], [24]. We also describe several real-world applications for control and deep learning inference. Lastly, we will discuss scalability in the context of designing a full-scale neuromorphic photonic processing system, considering aspects such as signal integrity, noise, and hardware fabrication platforms [3].

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, Art. no. 7553, May 2015.
[2] J. Schmidhuber, 'Deep learning in neural networks: An overview', Neural Netw., vol. 61, pp. 85–117, Jan. 2015.
[3] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," Psychol. Rev., vol. 65, no. 6, pp. 386–408, 1958

[4] A. L. Hodgkin and A. F. Huxley, ‘A quantitative description of membrane current and its application to conduction and excitation in nerve’, J. Physiol., vol. 117, no. 4, pp. 500–544, 1952.
[5] C. Mead and M. Ismail, Eds., Analog VLSI Implementation of Neural Systems. Springer US, 1989.
[6] V. K. Pallipuram, M. Bhuiyan, and M. C. Smith, ‘A comparative study of GPU programming models and architectures using neural networks’, J. Supercomput., vol. 61, no. 3, pp. 673–718, Sep. 2012
[7] N. P. Jouppi et al., ‘In-Datacenter Performance Analysis of a Tensor Processing Unit’, ArXiv170404760 Cs, Apr. 2017, Accessed: Oct. 24, 2020. [Online]. Available: http://arxiv.org/abs/1704.04760.
[8] J. Duarte et al., ‘Fast inference of deep neural networks in FPGAs for particle physics’, J. Instrum., vol. 13, no. 07, pp. P07027–P07027, Jul. 2018.
[9] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, ‘Physics for neuromorphic computing’, Nat. Rev. Phys., vol. 2, no. 9, Art. no. 9, Sep. 2020.
[10] K. Berggren et al., ‘Roadmap on emerging hardware and technology for machine learning’, Nanotechnology, vol. 32, no. 1, p. 012002, Jan. 2021.
[11] D. Psaltis and N. Farhat, ‘Optical information processing based on an associative-memory model of neural nets with thresholding and feedback’, Opt. Lett., vol. 10, no. 2, pp. 98–100, Feb. 1985.
[12] L. Chrostowski and M. Hochberg, Silicon Photonics Design: From Devices to Systems. Cambridge: Cambridge University Press, 2015.
[13] P. R. Prucnal and B. J. Shastri, Neuromorphic Photonics. Boca Raton, FL: CRC Press, 2017.
[14] Shastri, B.J., Tait, A.N., Ferreira de Lima, T. et al. Photonics for artificial intelligence and neuromorphic computing. Nat. Photonics 15, 102–114 (2021).
[15] Y. Shen, N. C. Harris, S. Skirlo, et al., “Deep learning with coherent nanophotonic circuits,” Nat. Photon., vol. 11, no. 7, pp. 441–446, 2017.
[16] A. N. Tait, T. F. de Lima, E. Zhou, et al., “Neuromorphic photonic networks using silicon photonic weight banks,” Sci. Rep., vol. 7, no. 1, p. 7430, 2017 [Online].
[17] A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, et al., “Silicon photonic modulator neuron,” Phys. Rev. Appl., vol. 11, no. 6, p. 064043, 2019 [Online]. https://doi.org/10.1103/ PhysRevApplied.11.064043.
[18] Feldmann, J., Youngblood, N., Karpov, M. et al. Parallel convolutional processing using an integrated photonic tensor core. Nature 589, 52–58 (2021).
[19] Xu, X., Tan, M., Corcoran, B. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. Nature 589, 44–51 (2021).
[20] 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., vol. 4, p. 1364, 2013.
[21] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, et al., “Experimental demonstration of reservoir computing on a silicon photonics chip,” Nat. Commun., vol. 5, p. 3541, 2014.
[22] P. R. Prucnal, B. J. Shastri, T. Ferreira de Lima, M. A. Nahmias, and A. N. Tait, “Recent progress in semiconductor excitable lasers for photonic spike processing,” Adv. Opt. Photon., vol. 8, no. 2, pp. 228–299, 2016
[23] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, “Training of photonic neural networks through in situ backpropagation and gradient measurement,” Optica, vol. 5, no. 7, pp. 864–871, 2018
[24] 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., vol. 32, no. 21, pp. 4029–4041, 2014.
[25] M. A. Nahmias, T. F. D. Lima, A. N. Tait, H. Peng, B. J. Shastri, and P. R. Prucnal, “Photonic multiply-accumulate operations for neural networks,” IEEE J. Sel. Top. Quantum Electron., vol. 26, no. 1, pp. 1–18, 2019.