Neuromorphic Photonic Processor Applications

Bhavin J. Shastri[1,3], Alexander N. Tait[2], Mitchell A. Nahmias[3], Thomas Ferreira de Lima[3], Hsuan-Tung Peng[3], and Paul R. Prucnal[3]

[1]Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada
[2] National Institute of Standards and Technology, Boulder, CO 80305, USA
[3]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
shastri@ieee.org

Abstract: Reconfigurable photonic processors promise orders of magnitude improvements in both speed and energy efficiency over digital electronics for applications in neuromorphic computing and machine learning. We will provide an overview of neuromorphic photonic systems and their application to optimization and machine learning problems.

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 [6-9], and the inundation of machine learning (ML) with neural models [10]. Neural networks have held some role in ML (e.g. image and voice recognition, language translation, pattern detection, and others) since the 1950s [11, 12]. They fell out of favor in the 90's because they are difficult to train.

Over the past decade, neural network models have decisively retaken the helm of ML under the alias of "deep networks". There are three main reasons: 1) major algorithmic innovations [13, 14], 2) the Internet: an inexhaustible source of millions of training examples, and 3) new hardware, specifically graphical processing units (GPUs) [15]. Central processing units (CPUs) are woefully inefficient 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 [16].

Today's demand for evaluating neural network models necessitates new hardware. High-tech juggernauts and research agencies have heavily invested in massively parallel application-specific integrated circuits (ASICs) for evaluating neural network models more efficiently, notably IBM [6], HP [17], Intel [9], Google [18,19], the Human Brain Project [20], and DARPA SyNAPSE [21]. Some of these architectures aim to be ML number crunchers [18,22], and others have enabled novel neuroscientific tools [23,24] and previously unforeseen low-power mobile applications [25].

The primary performance driver for the neuromorphic electronics community is computational power efficiency; speed is a secondary consideration. Neuromorphic electronics have largely focused on biological-timescale neural networks: kHz (with one 10MHz exception [23]). They universally rely on digital time-and event-multiplexing, which means they cannot simply run faster by turning up the clock. Nevertheless, there are compelling applications for neural networks with nanosecond latency. Some applications could be offline (i.e. number crunching) such as accelerators for deep network training and inference; others could be online (i.e. real-time) such as pattern detectors for wideband radio frequency (RF) signals and feedback controllers for systems subject to short-timeconstant instabilities. Moving beyond the nanosecond will require moving beyond purely electronic physics.

Photonic processors can outperform electronic systems that fundamentally depend on interconnects. Silicon photonic waveguides bus data at the speed of light. The associated energy costs are currently on the order of femtojoules per bit [26] and, in the near future, attojoules per bit [27]. Aggregate bandwidths continue to increase by combining multiple wavelengths of light (i.e., wavelength-division multiplexing (WDM)), theoretically topping out at 10 Tb/s per single-mode waveguides using 100 Gb/s per channel and up to 100 channels. On-chip scaling of many-channel dense WDM (DWDM) systems may be possible with comb generators in the near future [28].

Recently, there has been much work on photonics processors to accelerate information processing and reduce power consumption using: artificial neural networks [29–34], spiking neural networks [35–42], and reservoir computing [43-46]. By combining the high bandwidth and efficiency of photonic devices with the adaptive, parallelism and complexity attained by methods similar to those seen in the brain, photonic processors have the potential to be at least ten thousand times faster than state-of-the-art electronic processors while consuming less energy per computation [35,36].

In neuromorphic photonics [47,48], there is an isomorphism between the analog artificial neural networks and the underlying photonic hardware, which allows continuous functions to be fully represented in an analog way. An analog representation of information avoids overhead energy consumption and speed reduction caused by sampling and digitization into binary streams processed by clocked logic gates. But because of this analog representation, we cannot dissociate the information that flows through the neural network from the photonic physics that impacts distortion, noise and loss. Integration platforms for photonics also dictate how practical and how efficient neuromorphic photonic circuits can be. The most mature technology is silicon photonics [50], whose high-volume manufacturing allows for

978-1-7281-0597-0/19/$31.00 ©2019 IEEE

the most repeatable and robust platform for photonic circuits. Using silicon as a substrate also enables greater compatibility with digital electronic technology, allowing more compact solutions for neuromorphic hardware [51]. A great disadvantage of silicon photonics is the reliance on external lasers, typically built in III–V platforms, which require difficult and expensive co-packaging solutions. There are many applications driving the research community to find an industry-compatible solution for lasers-on-silicon, with good candidates such as III–V/Si hybrid fabrications, or quantum dot lasers grown directly on silicon. Industrial experts predict enabling innovations in the next five years that will allow neuromorphic photonic processors to be fabricated in a single die.

This talk will provide an overview of neuromorphic photonic systems and their application to optimization and machine learning problems. We will discuss the physical advantages of photonic processing systems, and we will describe underlying device models that allow practical systems to be constructed. 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.

References

[1] J. S. Orcutt et al. *Opt. Express* **20**, 12222 (2012).
[2] A.-J. Lim et al. *IEEE J. Sel. Top. Quantum Electron.* **20**, 405 (2014).
[3] D. Thomson et al. *J. Opt.* **18**, 073003 (2016).
[4] B. Marr et al. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **21**, 147 (2013).
[5] J. Hasler and H. B. Marr, *Front. Neurosci.* **7**, 118 (2013).
[6] P. A. Merolla et al. *Science* **345**, 668 (2014).
[7] S. B. Furber et al. *Proc. IEEE* **102**, 652 (2014).
[8] B. Benjamin et al. *Proc. IEEE* **102**, 699 (2014).
[9] M. Davies et al. *IEEE Micro.* **38**, 82 (2018).
[10] J. Schmidhuber, *Neural Netw.* **61**, 85 (2015).
[11] J. Von Neumann, *Automata Studies* **34**, 43 (1956).
[12] F. Rosenblatt, *Psychol. Rev.* **65**, 386 (1958).
[13] Y. Lecun et al. *Proc. IEEE* **86**, 2278 (1998).
[14] Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**, 436 (2015).
[15] V. K. Pallipuram, M. Bhuiyan, and M. C. Smith, *J. Supercomput.* **61**, 673 (2012).
[16] A. Diamond, T. Nowotny, and M. Schmuker, *Front. Neurosci.* **9**, 491 (2016),
[17] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, *Nat. Mater.* **12**, 114 (2013).
[18] N. P. Jouppi et al. *2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA)* (2017, June).
[19] A. Graves et al. *Nature* **538**, 471 (2016).
[20] H. Markram, *Sci. Am.* **306**, 50 (2012).
[21] A. S. Cassidy et al. *The 2013 International Joint Conference on Neural Networks (IJCNN)* (2013, Aug.).
[22] D. Miyashita et al. *IEEE J. Solid-State Circuits* **52**, 2679 (2017).
[23] T.Pfeil et al. *Front. Neurosci.* **7**, 11 (2013).
[24] S. Friedmann et al. *Front. Neurosci.* **7**, 160 (2013).
[25] W. Y. Tsai et al. *IEEE Trans. Comput.* **66**, 996-1007(2017).
[26] E. Timurdogan et al. *Nat. Commun.* **5**, 4008 (2014).
[27] V. J. Sorger et al. *J. Opt.* **20**, 014012 (2018).
[28] B. Stern et al. *Nature* **562**, 401 (2018).
[29] A. N. Tait et al. arXiv preprint arXiv:1812.11898 (2018).
[30] T. W. Hughes et al. *Optica* **5**, 864 (2018).
[31] A. N. Tait et al. *Sci. Rep.* **7**, 7430 (2017).
[32] Y. Shen et al. *Nat. Photon.* **11**, 441 (2017).
[33] J. M. Shainline et al. *Phys. Rev. Appl.* **7**, 034013 (2017).
[34] A. N. Tait et al. J. *Lightwave Technol.* **32**, 4029 (2014).
[35] P. R. Prucnal et al. *Adv. Opt. Photon.* **8**, 228 (2016).
[36] H. T. Peng et al. *IEEE J. Sel. Top. Quantum Electron.* **24**, 6101715 (2018).
[37] Y. Zhang et al. *Sci. Rep.* **8**, 16095 (2018).
[38] T. Deng et al. *IEEE J. Sel. Top. Quantum Electron.* **23**, (2017).
[39] B. Romeira et al. *Sci. Rep.* **6**, 19510 (2016).
[40] B. J. Shastri et al. *Sci. Rep.* **5**, 19126 (2016).
[41] A. Aragoneses et al. *Sci. Rep.* **4**, 4696 (2014).
[42] M. A. Nahmias et al. *IEEE J. Sel. Top. Quantum Electron.* **19**, 1800212 (2013)
[43] G. Van der Sande et al. *Nanophotonics*, **6**, 561 (2017).
[44] D. Brunner et al. *Nat. Commun.* **4**, 1364 (2013).
[45] K. Vandoorne et al. *Nat. Commun.* **5**, 3541 (2014).
[46] L. Larger et al. *Opt. Express* **20**, 3241 (2012).
[47] P. R. Prucnal and B. J. Shastri. *Neuromorphic Photonics* (CRC Press, 2017).
[48] T. Ferreira de Lima et al. *Nanophotonics* 6**,** 577 (2017).
[49] M. A. Nahmias et al. *Opt. Photon. News* **29**, 34 (2018).
[50] A. Rahim et al. *Proc. IEEE* **106**, 2313 (2018).
[51] A. N. Tait et al. *Opt. Express* **26**, 26422 (2018).