Multiwavelength Neuromorphic Photonics

Paul R. Prucnal[1], Alexander N. Tait[1], Mitchell A. Nahmias[1], Thomas Ferreira de Lima[1], Hsuan-Tung Peng[1] and Bhavin J. Shastri[1,2]
[1]*Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA*
[2]*Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada*
prucnal@princeton.edu

Abstract: Neuromorphic photonics promises orders of magnitude improvements in both speed and energy efficiency over digital electronics. We will give an overview of neuromorphic photonic systems and their application to optimization and machine learning problems.
OCIS codes: (200.3050) Information processing; (200.4700) Optical neural systems; (250.5300) Photonic integrated circuits.

Artificial Intelligence (AI) has always captured our imagination. AI has the potential to drastically change almost every aspect of our lives through new medical treatments, new assistive robots, intelligent modes of transportation, and much more. Inspired by the human brain and spurred by the advances in deep learning, the past six years has seen a renaissance in AI. IBM [1], HP [2], Intel [3], and Google [4], have all shifted their core technological strategies from "mobile first" to "AI first". Deep learning with artificial neural networks (ANNs) [5] has expanded from image recognition [6] to translating languages [7] and beating humans at highly complex strategy games like Go [8]. The general consensus amongst the scientific and private sector community is that three factors will drive the future advance of AI: better algorithms, more training data, and the amount of compute power available for training. While there has been no shortage of innovative architecture variants for these neural networks nor data to train them, the most pressing bottleneck for AI is now processing power (Fig. 1). Over the last six years, the amount of compute power required to train state-of-the- art AI has been doubling every 3.5 months [9]. For instance, Google's AlphaGo AI requires 1920 CPUs and 280 GPUs, which translates into massive power consumption, reaching around $3000 USD in electric bill per game. Training neural networks also takes a considerable amount of computational time. For example, image classification tasks with residual neural networks (ResNet-200) requires 8 GPUs and takes more than three weeks of training to achieve classification error rates at around 20.7% [10]. Traditional CPUs, GPUs and even neuromorphic electronics (IBM TrueNorth [1] and Google TPU [4]) have improved both energy efficiency and speed enhancement for learning (inference) tasks. However, electronic architectures face fundamental limits as Moore's law is slowing down [11]. Furthermore, moving data electronically on metal wires has fundamental bandwidth and energy efficiency limitations [12], thus remaining a critical challenge facing deep learning hardware accelerators [13].

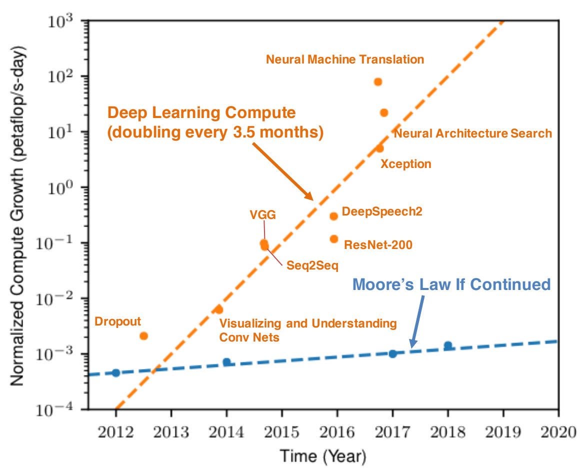

Fig. 1. High-performance computing is dominated by deep learning which is quickly saturating available compute growth. The orange dots show the total amount of compute, normalized to petaflop/s-day, that was used to train each of selected neural network architectures. The blue dots show the trend of Moore's law. (A petaflop/s-day is the number of operations of performing 1015 operations per second for one day, which in total is 8.64 × 1019 operations).

Photonic processors can significantly 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 [14] and, in the near future, attojoules per bit [15]. 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 [16].

Recently, there has been much work on photonics processors to accelerate information processing and reduce power consumption using: artificial neural networks [17–22], spiking neural networks [23–30], and reservoir computing [31–34]. 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 [35,37], 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 [40], whose high-volume manufacturing allows for 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[41,42]. 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] S. K. Esser et al. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11441 (2016).
[2] M. D. Pickett et al. *Nat. Mater.* **12**, 114 (2013).
[3] M. Davies et al. *IEEE Micro* **38**, 82 (2018).
[4] A. Graves et al. *Nature* **538**, 471 (2016).
[5] Y. LeCun et al. *Nature* **521**, 436 (2015).
[6] K. Simonyan and A. Zisserman, CoRR abs/1409.1556 (2014).
[7] Y. Wu et al. CoRR abs/1609.08144 (2016).
[8] D. Silver et al. *Nature* **529**, 484 (2016).
[9] D. Amodei and D. Hernandez, "AI and Compute." https://blog.openai.com/ai-and-compute/
[10] K. He et al. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016, June).
[11] M. M. Waldrop, *Nature News* **530**, 144 (2016).
[12] D. A. B. Miller, *Proc. IEEE* **97**, 1166 (2009).
[13] Y. Chen et al. *IEEE J. Solid-State Circuits* **52**, 127 (2017).
[14] E. Timurdogan et al. *Nat. Commun.* **5**, 4008 (2014).
[15] V. J. Sorger et al. *J. Opt.* **20**, 014012 (2018).
[16] B. Stern et al. *Nature* **562**, 401 (2018).
[17] A. N. Tait et al. arXiv preprint arXiv:1812.11898 (2018).
[18] T. W. Hughes et al. *Optica* **5**, 864 (2018).
[19] A. N. Tait et al. *Sci. Rep.* **7**, 7430 (2017).
[20] Y. Shen et al. *Nat. Photon.* **11**, 441 (2017).
[21] J. M. Shainline et al. *Phys. Rev. Appl.* **7**, 034013 (2017).
[22] A. N. Tait et al. J. *Lightwave Technol.* **32**, 4029 (2014).
[23] P. R. Prucnal et al. *Adv. Opt. Photon.* **8**, 228 (2016).
[24] H. T. Peng et al. *IEEE J. Sel. Top. Quantum Electron.* **24**, 6101715 (2018).
[25] Y. Zhang et al. *Sci. Rep.* **8**, 16095 (2018).
[26] T. Deng et al. *IEEE J. Sel. Top. Quantum Electron.* **23**, (2017).
[27] B. Romeira et al. *Sci. Rep.* **6**, 19510 (2016).
[28] B. J. Shastri et al. *Sci. Rep.* **5**, 19126 (2016).
[29] A. Aragoneses et al. *Sci. Rep.* **4**, 4696 (2014).
[30] M. A. Nahmias et al. *IEEE J. Sel. Top. Quantum Electron.* **19**, 1800212 (2013)
[31] G. Van der Sande et al. *Nanophotonics*, **6**, 561 (2017).
[32] D. Brunner et al. *Nat. Commun.* **4**, 1364 (2013).
[33] K. Vandoorne et al. *Nat. Commun.* **5**, 3541 (2014).
[34] L. Larger et al. *Opt. Express* **20**, 3241 (2012).
[35] P. R. Prucnal and B. J. Shastri. *Neuromorphic Photonics* (CRC Press, 2017).
[36] T. Ferreira de Lima et al. *Nanophotonics* 6**,** 577 (2017).
[37] M. A. Nahmias et al. *Opt. Photon. News* **29**, 34 (2018).
[40] A. Rahim et al. *Proc. IEEE* **106**, 2313 (2018).
[41] A. N. Tait et al. *IEEE J. Sel. Top. Quantum Electron.* **22**, 312 (2016)
[42] A. N. Tait et al. *Opt. Express* **26**, 26422 (2018)