Multiwavelength Neuromorphic Silicon Photonics

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

Artificial Intelligence (AI) is transforming our lives in the same way as the advent of the Internet and cellular phones has done. AI is revolutionizing the healthcare industry with complex medical data analysis, actualizing self-driving cars, and beating humans at strategy games such as Go. However, it takes thousands of CPUs and GPUs, and many weeks to train the neural networks in AI hardware. Over the last six years, this compute power has doubled every 3.5 months. Traditional CPUs, GPUs and even neuromorphic electronics (IBM TrueNorth [1] and Google TPU [2]) 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. Furthermore, moving data electronically on metal wires has fundamental bandwidth and energy efficiency limitations, thus remaining a critical challenge facing deep learning hardware accelerators [3].

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 and, in the near future, attojoules per bit [4]. Aggregate bandwidths continue to increase with 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 systems may be possible with comb generators in the near future [5].

Recently, there has been much work on neuromorphic photonic [6] processors to accelerate information processing and reduce power consumption using: artificial neural networks [7–10], spiking neural networks [11–13], and reservoir computing [14]. 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 [15,16].

In neuromorphic photonics, 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 [17], 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. A disadvantage of silicon photonics is the reliance on external lasers, typically built in III–V platforms, which require difficult and expensive co-packaging solutions. Many applications are driving the research community to find an industry-compatible solution for lasers-on-silicon, with candidates such as III–V/Si hybrid fabrications, or quantum dot lasers grown directly on silicon. Innovations over the next 5 years could allow neuromorphic photonic processors to be fabricated in a single die.

We 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. 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] A. Graves et al. Nature **538**, 471 (2016).
[3] Y. Chen et al. IEEE J. Solid-State Circuits **52**, 127 (2017).
[4] V. J. Sorger et al. J. Opt. **20**, 014012 (2018).
[5] B. Stern et al. Nature **562**, 401 (2018).
[6] P. R. Prucnal and B. J. Shastri, *Neuromorphic Photonics* (CRC Press, 2017).
[7] A. N. Tait et al. arXiv preprint arXiv:1812.11898 (2018).
[8] Y. Shen et al. Nat. Photon. **11**, 441 (2017).
[9] J. M. Shainline et al. Phys. Rev. Appl. **7**, 034013 (2017).
[10] A. N. Tait et al. Sci. Rep. **7**, 7430 (2017).
[11] H. T. Peng et al. IEEE J. Sel. Top. Quantum Electron. **24**, 6101715 (2018).
[12] B. J. Shastri et al. Sci. Rep. **5**, 19126 (2016).
[13] M. A. Nahmias et al. IEEE J. Sel. Top. Quantum Electron. **19**, 1800212 (2013).
[14] G. Van der Sande et al. Nanophotonics, **6**, 561 (2017).
[15] A. N. Tait et al. J. Lightwave Technol. **32**, 4029 (2014).
[16] T. Ferreira de Lima et al. Nanophotonics **6,** 577 (2017).
[17] A. Rahim et al. Proc. IEEE **106**, 2313 (2018).

978-1-7281-0469-0/19/$31.00 ©2019 IEEE