Silicon Photonics for Training Deep Neural Networks

Bhavin J. Shastri[1,2], Matthew J. Filipovich[1], Zhimu Guo[1], Paul R. Prucnal[2], Sudip Shekhar[3], and Volker J. Sorger[4]

[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
[4]Department of Electrical and Computer Engineering, George Washington University, Washington, DC V6T 1Z4, USA
shastri@ieee.org

Abstract: Analog photonic networks as deep learning hardware accelerators are trained on standard digital electronics. We propose an on-chip training of neural networks enabled by a silicon photonic architecture for parallel, efficient, and fast data operations. © 2022 The Author(s)

Photonic systems for information processing have gathered significant interest as an alternative to conventional electronic computer architectures [1]. The emerging field of neuromorphic (i.e., neuron-isomorphic) photonics [2] proposes implementing high-performance neural networks and related machine learning algorithms using electro-optic circuits. These applications require high bandwidth, low latency, and low energy consumption. As such, the last decade has seen a rise of photonic neural networks [1], [3] that can be divided into feedforward and recurrent, including random recurrent, i.e., reservoir computing [4]–[6]), or coherent (single-wavelength) [7–9] and multiwavelength [10]–[19] approaches, or continuous-time networks and spiking networks, or integrated approaches and free-space.

An area of machine learning that would benefit from the low power consumption and high information processing bandwidth enabled by photonics is the training of large neural networks. Several photonic architectures have been proposed for executing in-memory computation of neural network inference [7], [10], [13]. However, for the neural network to perform a practical task, the optimal network parameters (weights and biases) must first be determined using deep learning training algorithms. These algorithms have high computation and memory costs that challenge the current hardware platforms executing them [20]. The substantial energy required to train large neural networks using standard von Neumann architectures presents a high financial and environmental cost [21].

The recently proposed direct feedback alignment (DFA) supervised learning algorithm [22] has gathered interest as a bio-plausible alternative to the popular backpropagation training algorithm [23]. The DFA algorithm is a supervised learning algorithm that propagates the error through fixed random feedback connections directly from the output layer to the hidden layers during the backward pass [23]. Unlike backpropagation, the DFA algorithm does not require the network layers to be updated sequentially during the backward pass, enabling the algorithm to be a suitable candidate for efficient parallelization using photonics. The training algorithm has been used to train neural networks using the MNIST, CIFAR-10, and CIFAR-100 datasets and yields comparable performance to backpropagation [23]. The DFA algorithm has also been shown to obtain performances similar to fine-tuned backpropagation in applications requiring state-of-the-art deep learning networks, including natural language processing and neural view synthesis [24]. A recent theory suggests that training shallow networks with the DFA algorithm occurs in two steps: the first step is an alignment phase where the weights are modified to align the approximate gradient with the actual gradient of the loss function, which is followed by a memorization phase where the model focuses on fitting the data [25].

This talk will summarize our recently proposed silicon photonic architecture [26] that uses an electro-optic circuit to calculate the gradient vector of each neural network layer in situ, the most computationally expensive operation performed during the backward pass. The proposed architecture exploits the speed (10s of GHz range in photonics but only 100s of MHz in electronics) and energy advantages of photonics to determine the gradient vector of each neural network layer in a single operational cycle.

While practical neuromorphic processors may be years away, we have outlined in [10], [27] some scientific and technological advances necessary to meet the challenges.

References

[1] Shastri, Bhavin J., et al. "Photonics for artificial intelligence and neuromorphic computing." Nat. Photon. 15.2 (2021): 102-114.
[2] Prucnal, Paul R., and Bhavin J. Shastri. Neuromorphic photonics. CRC Press, 2017.
[3] Huang, Chaoran, et al. "Prospects and applications of photonic neural networks." *Advances in Physics: X* 7.1 (2022): 1981155.
[4] Brunner, Daniel, et al. "Parallel photonic information processing at gigabyte per second data rates using transient states." Nat. Commun. 4.1 (2013): 1-7.
[5] Vandoorne, Kristof, et al. "Experimental demonstration of reservoir computing on a silicon photonics chip." Nat. Commun. 5.1 (2014): 1-6.

[6]   Larger, Laurent, et al. "Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing." Opt. Express 20.3 (2012): 3241-3249.

[7]   Shen, Yichen, et al. "Deep learning with coherent nanophotonic circuits." Nat. Photon. 11.7 (2017): 441-446.

[8]   Lin, Xing, et al. "All-optical machine learning using diffractive deep neural networks." Science 361.6406 (2018): 1004-1008.

[9]   Hughes, Tyler W., et al. "Training of photonic neural networks through in situ backpropagation and gradient measurement." Optica 5.7 (2018): 864-871.

[10]   Feldmann, Johannes, et al. "Parallel convolutional processing using an integrated photonic tensor core." *Nature* 589.7840 (2021): 52-58.

[11]   Miscuglio, Mario, and Volker J. Sorger. "Photonic tensor cores for machine learning." *Applied Physics Reviews* 7.3 (2020): 031404.

[12]   Bangari, Viraj, et al. "Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs)." *IEEE Journal of Selected Topics in Quantum Electronics* 26.1 (2019): 1-13.

[13]   Tait, Alexander N., et al. "Neuromorphic photonic networks using silicon photonic weight banks." Sci. Rep. 7.1 (2017): 1-10.

[14]   Tait, Alexander N., et al. "Silicon photonic modulator neuron." Phys. Rev. Appl. 11.6 (2019): 064043.

[15]   Tait, Alexander N., et al. "Broadcast and weight: an integrated network for scalable photonic spike processing." J. Lightwave Technol. 32.21 (2014): 4029-4041.

[16]   Huang, Chaoran, et al. "A silicon photonic–electronic neural network for fibre nonlinearity compensation." Nature Electronics 4.11 (2021): 837-844.

[17]   Shi, B., Calabretta, N. & Stabile, R. Deep neural network through an InP SOA-based photonic integrated cross-connect. IEEE J. Sel. Top. Quantum Electron. 26, 7701111 (2020).

[18]   Peng, H. T. et al. Neuromorphic photonic integrated circuits. IEEE J. Sel. Top. Quant. Electron. 24, 6101715 (2018).

[19]   Mourgias-Alexandris, G. et al. An all-optical neuron with sigmoid activation function. Opt. Express 27, 9620–9630 (2019).

[20]   Esser, S., et al. "Convolutional networks for fast, energy-efficient neuromorphic computing. arXiv 2016." arXiv preprint arXiv:1603.08270.

[21]   Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and policy considerations for deep learning in NLP." arXiv preprint arXiv:1906.02243 (2019).

[22]   Nøkland, Arild. "Direct feedback alignment provides learning in deep neural networks." Advances in neural information processing systems 29 (2016).

[23]   Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533-536.

[24]   Launay, Julien, et al. "Direct feedback alignment scales to modern deep learning tasks and architectures." Advances in neural information processing systems 33 (2020): 9346-9360.

[25]   M. Refinetti, S. d'Ascoli, R. Ohana, and S. Goldt, "The dynamics of learning with feedback alignment," arXiv:2011.12428 [cond-mat, stat], Nov 2020.

[26]   Filipovich, Matthew J., et al. "Monolithic Silicon Photonic Architecture for Training Deep Neural Networks with Direct Feedback Alignment." arXiv preprint arXiv:2111.06862 (2021).

[27]   Berggren, Karl, et al. "Roadmap on emerging hardware and technology for machine learning." Nanotechnology 32.1 (2020): 012002.