Statistical Computing with Photonic Integrated Circuits

Yichen Shen,[1,3] Huaiyu Meng[2], Li Jing[1], Dirk Englund,[2] and Marin Soljačić[1]

[1] *Department of Physics, Massachusetts Institute of Technology,Cambridge, MA 02139, USA*

[2] *Research Laboratory for Electronics, Massachusetts Institute of Technology,Cambridge, MA 02139, USA*

[3] *Lightelligence Inc.,Cambridge, MA 02139, USA*

Abstract: I will show that some conventional statistical processing algorithms can be equivalently carried out by nanoscaled optical systems, and such can be at least 3 orders of magnitude faster and power efficient in forwarding propagation than an electronic chip.

OCIS codes: 200.4260 (Neural Networks), 200.4700 (Optical Neural Systems), 130.3120 (Integrated Optical Device)

Computers that can learn, combine, and analyze vast amounts of information quickly, efficiently, and without the need for explicit instructions are a powerful tool for handling large datasets. Indeed, statistical computing algorithms have received an explosion of interest in both academia and industry for their utility in image recognition, language translation, decision making problems, and more [1]. Traditional central processing units (CPUs) are far suboptimal for implementing these algorithms [2]; and a growing effort in academia and industry has been put towards the development of new hardware architectures tailored towards applications in artificial neural networks and deep learning [3]. Graphical Processing Unit (GPUs), Application Specific Integrated Circuits (ASICs) and field-programmable gate arrays (FPGAs), have enabled both energy efficiency and speed enhancements for learning tasks. In parallel, hybrid optical-electronic systems that implement spike processing and reservoir computing have been shown. However, the computational speed and power efficiency achieved with these hardware architectures are still limited by electronic clock rates and Ohmic losses.

Optical based statistical processing offer a promising alternative approach to microelectronic and hybrid optical-electronic implementations. In fact, modern statistical computing algorithms are a promising fully-optical computing paradigm because (1) *they rely heavily on fixed matrix multiplications*: linear transformations (and certain non-linear transformations) can be performed at the speed of light and detected at rates exceeding 100 GHz [4] in photonic networks, and in some cases, with minimal power consumption. (2) *they have weak requirements on nonlinearities and bit-resolutions*: many inherent optical nonlinearities can be directly used to implement the nonlinear operations in modern machine learning algorithms, and the nature of statistical computing significantly reduced the accuracy requirement of each operation; and (3) *once the machine learning model is trained, the architecture can be passive*, the computation on the optical signals will be performed without additional energy input.

A typical statistical machine learning architecture contains an input layer, at least one hidden layers, and an output layer. In each layer, information propagate through linear combination (e.g. matrix multiplication) followed by some nonlinear activation function applied to the result from linear combination. In training a machine learning model, data are fed into the input layer, and output is calculated through the forward propagation step. Then the parameters are optimized through the back propagation procedure. The general architecture of our optical statistical processing (OSP) is depicted in Fig. 1A. It is mainly composed of three optical processing units:

1. **Optical Interference Unit.** Which is used to perform arbitrary *unitary* matrix multiplication on the input optical signal. The unitary matrix can be obtained using a network of Mach-Zehnder interferometers, as shown in Fig. 1D,E. Mathematically, it can be rigorously proved that any arbitrary unitary matrix can be represented by the network of Mach-Zehnder interferometers [5] .

2. **Optical Amplification Unit.** Which is used to generalize the unitary matrix to arbitrary matrix operation. In general, any arbitrary matrix can be generated using optical interference and linear amplification through SVD decomposition.

3. **Optical Nonlinearity Unit.** Which is used to apply the nonlinear activation function. Many materials respond to external light signals in a nonlinear way with respect to light intensity. One of the most commonly used optical nonlinearities – saturable absorption in our system. The saturable absorber nonlinear function can be modeled as $\sigma\tau_s I_0 = \frac{1}{2}\frac{\ln(T_m/T_0)}{1-T_m}$ [6], as shown in Fig. 1C.

With these three units, in principle, our OSP architecture can do computation in a way that is mathematically equivalent to electrical computers.

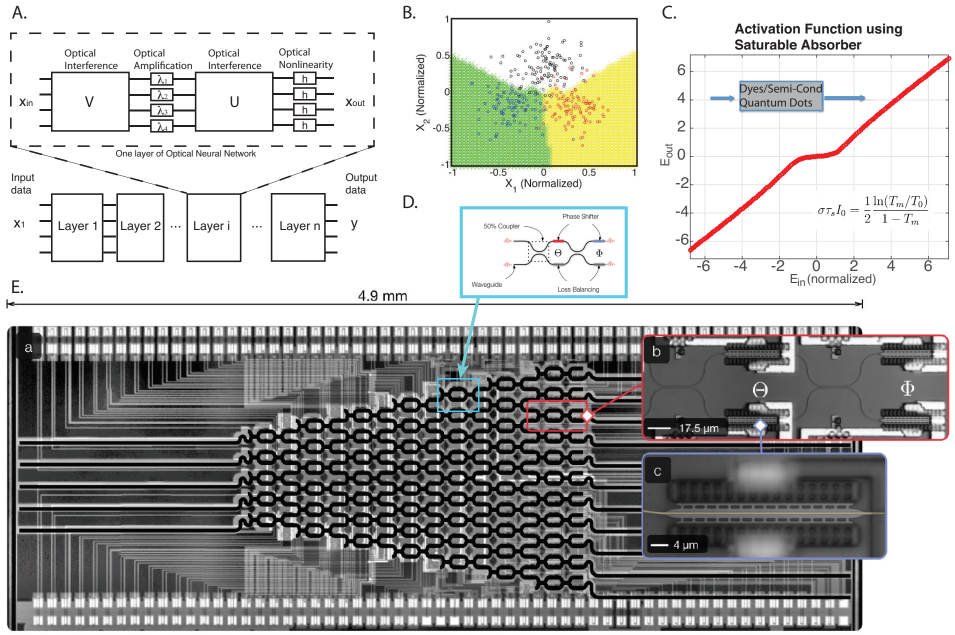

Fig. 1: (A) General Architecture of Optical Neural Network (ONN), one of the most popular example of statistical machine learning (B) Decision boundary for a simple 2 dimensional, 3 classes classification problem trained on our Optical Neural Network System. Three categories of data are labeled by different colors, areas are also labeled by different colors based on predictions (C) The optical response of a nonlinear optical system (saturable abosrption) that perform nonlinear calculation. Inset: Schematic illustration of the saturable absorption system. (D). Schematic illustration of a single Mach-Zehnder Interferometer. (E) Microscope image of an experimentally fabricated 5×5 unit on chip optical interference unit [7]

I will also discuss optical implementation of several other statistical processing algorithms, and analyze the pros and cons of using optics to implement those algorithms. Finally, I will discuss how to co-design optics processors and algorithms.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
2. S. K. Esser, P. A. Merolla, J. V. Arthur, A. S. Cassidy, R. Appuswamy, A. Andreopoulos, D. J. Berg, J. L. McKinstry, T. Melano, D. R. Barch, C. di Nolfo, P. Datta, A. Amir, B. Taba, M. D. Flickner, and D. S. Modha, "Convolutional networks for fast, energy-efficient neuromorphic computing," Proceedings of the National Academy of Sciences **113**, 11,441–11,446 (2016).
3. C. Mead, "Neuromorphic electronic systems," Proceedings of the IEEE **78**, 1629–1636 (1990).
4. L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J. M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, and J. M. Fédéli, "Zero-bias 40gbit/s germanium waveguide photodetector on silicon," Opt. Express **20**, 1096–1101 (2012).
5. M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," Phys. Rev. Lett. **73**, 58–61 (1994).
6. A. Selden, "Pulse transmission through a saturable absorber," British Journal of Applied Physics **18**, 743 (1967).
7. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund *et al.*, "Deep learning with coherent nanophotonic circuits," Nature Photonics **11**, 441 (2017).