Demonstration of a silicon photonic neural network

Alexander N. Tait, Ellen Zhou, Allie X. Wu, Mitchell A. Nahmias,
Thomas Ferreira de Lima, Bhavin J. Shastri, and Paul R. Prucnal
Princeton University, Princeton, NJ, 08544 USA
atait@princeton.edu

***Abstract*—Silicon photonic integration could enable high-performance brain-inspired photonic processors. We demonstrate a 3 node recurrent photonic neural network. Cusp and Hopf bifurcations induced by synaptic reconfiguration are shown as proof-of-concept. The prototype represents an early step towards network-based models of physical computing with integrated photonics.**

Research in optics and computing spans generations of scientific pursuit. Some fruitful forays into this intersection exploit analog opto-electronic physics, instead of abtracting physics beneath a logical framework. The most successful known information processors based on physical dynamics are found in the animal nervous system. Neural network models provide a broad and well-researched theoretical bridge between low-level physics and high-level computational tasks, so it is natural that they have played a role in optics and computing. Past optical neural networks attained moderate scale in free-space, yet their incompatibility with technology manufacturing ultimately led to a dormancy in this field.

The recent adoption of silicon photonic integration potentially endow large-scale optical systems with the manufacturing economies typically enjoyed by microelectronics. Integrated photonic neural networks could address novel computational domains that demand extreme speed, energy efficiency, significant complexity, and full programmability. Interest in laser devices with neuron-like behavior has flourished over the past several years [1]. Neural networks of N nodes contain up to N^2 connections, or synapses, with programmable, continuous-valued weights. In [2], we proposed a photonic neural networking approach based on mainstream integrated photonic devices. In a so-called broadcast-and-weight network, optical signals are wavelength-division multiplexed (WDM), distributed to tunable microring resonator (MRR) weight banks, and summed by balanced photodetectors (PD). Synaptic operations implemented by configurable transmission elements and incoherent photodetection maintain efficiency at extremely high bandwidths, unlike the multiply-accumulate implementations used in electronics.

Here, we report the first experimental demonstration of an analog neural network integrated in a silicon photonic chip. As proof-of-concept, we show simple bifurcations – qualitative changes in dynamical behavior – that are induced by network weight reconfiguration, and that correspond to the theory of small-scale neural networks [3]. Neural networks are described by a set of ordinary differential equations coupled through a matrix of reconfigurable weights, $\mathbf{W}$. Depending on the type of neural model, each neuron is represented in these equations either by a nonlinear function, σ, (e.g. continuous-time) or a dynamical subsystem (e.g. spiking) [4]. In the continuous-time case,

$$\frac{d\vec{s}(t)}{dt} = \mathbf{W}\vec{y}(t) + \vec{w}_{in}x(t) \quad (1)$$

$$\vec{y}(t) = \sigma\left[\vec{s}(t)\right], \quad (2)$$

where $s(t)$ are synaptic variables, $y(t)$ are neuron outputs, w_{in} are input weights, and $x(t)$ is the input.

Samples were fabricated on silicon-on-insulator (SOI) wafers at the Washington Nanofabrication Facility through the UBC SiEPIC rapid prototyping group [5]. Silicon thickness is 220 nm, and buried oxide thickness is 3 µm. 500 nm wide WGs were patterned by Ebeam lithography and fully etched to the buried oxide [6]. After a cladding oxide (3 µm) is deposited, two metal layers are deposited: Ti/W for heating elements, and Al for routing (Applied Nanotools, Alberta). Ohmic heating in these contacts causes thermo-optic resonant wavelength shift in the MRR filters. The sample is mounted on a temperature-controlled alignment stage, and coupled to a 9-fiber array using focusing subwavelength grating couplers [7].

The experimental setup and an image of the integrated network are shown in Fig. 1. The network in a broadcast STAR configuration consists of splitting Y-junctions and 4 MRR weight banks with 4 weights each. The photodetected weighted sums constitute synaptic signals, s, which drive Mach-Zehnder modulators (MZMs) modulating distinct wavelengths λ_1–λ_3. MZMs have a sinusoidal electro-optic transfer function which serves as the saturating nonlinearity, σ, associated with the continuous-time neuron. A fourth weight bank is unused, and λ_4 is used as an external input, x. The three neuron outputs, y, and input signal, x are multiplexed off-chip. Each MRR weight bank is calibrated offline using the method introduced in Ref. [8].

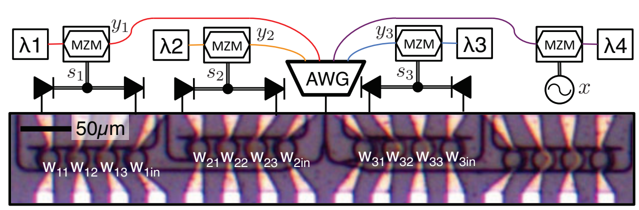

Fig. 1. Experimental setup with three MZM neurons, wavelength-multiplexed in an AWG and coupled into an on-chip broadcast-and-weight neural network. λ_4 carries an external input derived from a signal generator. The 3x3 recurrent network state is configured by MRR weight banks.

When all weights are held zero, and one autaptic connection is swept, the transition from stable to bistable is observed in Fig. 2. Transitions in this regime are known as the cusp bifurcation because of the shape of the bistable area when viewed in the $y = 0$ plane. The input voltage is a 3kHz triangle wave with 50% duty. The opening of an area between rising (blue) and falling (red) outputs is indicative of a bistable regime, whose shape corresponds well with theoretical predictions. The bistability measurements also serve as a control showing that no spurious oscillations due to time-delayed dynamics are present.

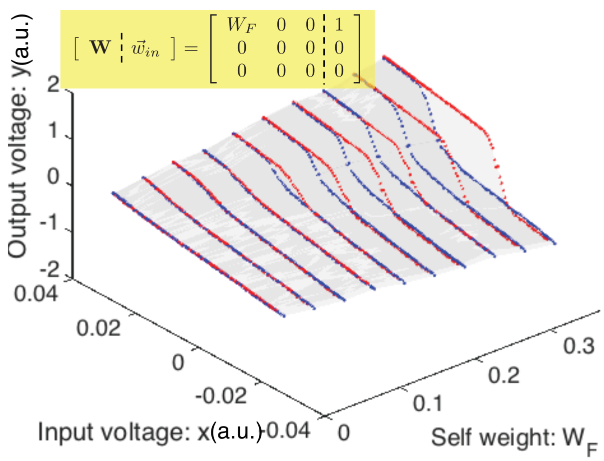

Fig. 2. A cusp bifurcation with a single node with self-connection and an external input. A sweep of self-weight strength, W_F, is controlled by a MRR weight. Blue points: increasing input voltage. Red points: decreasing input voltage. The transition from stable to bistable is parameterized by the weight.

When two neurons are connected with asymmetric off-diagonal weights, oscillation can occur. Varying both feedback weights together, the system undergoes a supercritical Hopf bifurcation between stable and limit cycle behavior, as shown in Fig. 3. Oscillation frequency varied continuously from 2 – 5kHz, depending on feedback strength. Other regimes of behavior observable in two-node systems (not shown) include cooperative bistability, competetive bistability, and quadstability. The Hopf bifurcation occurs in systems of more than one dimension, thus confirming the observation of a small integrated photonic neural network.

Large feedback delay (∼100ns) through fiber optic components forced slow operation in order to spoil time-delayed dynamics, as seen in Ref. [9]. Time-delayed dynamical systems resist simple analysis applied to continuous-time recurrent neural networks and their sensitivity to environmental perturbations can confound dynamical observations. By limiting operational bandwidth of the synaptic detector-driver circuitry to 10kHz, all time-delayed dynamics were eliminated, as evidenced by the expected stability observed in Fig. 2. With on-chip modulators or lasers, time-of-flight delays less than one picosecond allow usable operational bandwidths in the GHz range. While other kinds of electro-optic oscillators, including those based on time-delayed dynamics, have been shown to possess a rich repertoire of behaviors, this route to complexity lacks the theoretical scaffolding of neural networks, although reservoir techniques have been shown [10]. A key advantage of hardware conforming to network-based models is the tremendous breadth of existing research on programming neural networks to accomplish desired computing tasks. While this demonstration used only two nodes, adding neurons increases the repertoire of complex behaviors in a way that is modular and amenable to known programming methods.

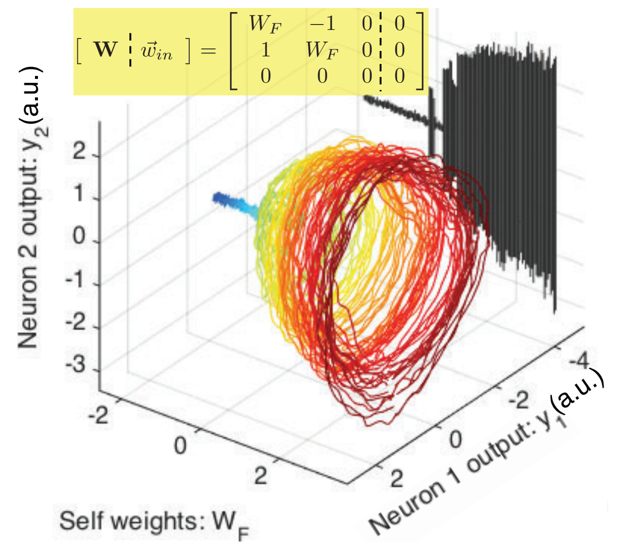

Fig. 3. A supercritical Hopf bifurcation involving two nodes and no input. Off-diagonal weights are asymmetric, and diagonal self weights are swept in concert. A transition between stable and autonomous oscillation is observed. Color represents self weight value.

We have demonstrated a reconfigurable analog neural network in a silicon photonic integrated circuit. Network-mediated cusp and Hopf bifurcations were observed as a first proof-of-concept of an integrated broadcast-and-weight system [2]. Neural network abstractions are powerful tools for bridging the gap between physical dynamics and useful application, and silicon photonic manufacturing introduces opportunities for large-scale photonic systems. At increased scale, silicon photonic neural networks could be applied to unaddressed computational areas requiring ultrafast, reconfigurable, and efficient hardware processors.

References

[1] M. A. Nahmias *et al.*, in *Proc. IEEE Photonics Conf. (IPC)*. Seattle, WA, USA: paper ME3.4, Sep. 2013, pp. 93–94.
[2] A. N. Tait *et al.*, *J. Lightwave Technol.*, vol. 32, no. 21, pp. 3427–3439, Nov 2014.
[3] R. D. Beer, *Adaptive Behavior*, vol. 3, no. 4, pp. 469–509, 1995.
[4] B. J. Shastri *et al.*, *Sci. Rep.*, vol. 5, p. 19126, Dec. 2015.
[5] L. Chrostowski *et al.* Cambridge University Press, 2015.
[6] R. J. Bojko *et al.*, *J. Vac. Sci. Technol., B*, vol. 29, no. 6, 2011.
[7] Y. Wang *et al.*, *Opt. Express*, vol. 22, no. 17, pp. 20 652–20 662, 2014.
[8] A. N. Tait *et al.*, *Opt. Express*, no. peer review, 2016.
[9] E. Zhou *et al.*, in *Optical Interconnects Conference, 2016 IEEE*. IEEE, May 2016, p. TuP9.
[10] D. Brunner *et al.*, *Nat Commun*, vol. 4, p. 1364, 01 2013.