CLEO 2018 © OSA 2018

Microring Weight Banks for Neuromorphic Silicon Photonics

A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal
Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA
atait@ieee.org

***Abstract*—Research in photonic information processing has experienced a recent resurgence. Microring weight banks enable reconfigurable neural network approaches on silicon photonics. We discuss the latest results in neuromorphic silicon photonic networks and microring weight banks.**

As compared to electronics, optics lacks key physical requirements for implementing digital logic; however, these same physical qualities make it superior at analog signal processing. For example, radio frequency (RF) and microwave photonics (MWP) exploit the bandwidth, linearity, and tunability of optics to address next-generation wireless system needs [1], but their processing repertoire is usually constrained to single-input linear operations. Extending this performance to much more complex tasks would require scalable models and photonic implementations of network-based information processing. Optical neural networks have been explored using free-space optics, but were effectively considered untenable about two decades ago because they were large, expensive, sub-GHz, and severely limited in scalability.

Neuromorphic photonics [2] aims to map physical models of optoelectronic systems to abstract models of neural networks. Today, rapid advances in photonic manufacturing could revolutionize large-scale photonic systems in terms of size, cost, device performance, reliability, and scalability. The processing repertoire of single neurons is relatively simple while that of overall neural networks can be incredibly complex and varied. Network connection strengths (a.k.a. weights) are closely tied to computational function, not just data communication. Neural networks owe their name to biology, but are in fact a class of well-studied mathematical models. This extensive knowledge of how to relate weight profiles to function can be leveraged by hardware that can be made isomorphic to a neural network framework – neuromorphic systems. Neuromorphic electronic architectures utilizing this strategy have recently attracted tremendous research interest [3]. Here, we discuss recent advances in integrated neuromorphic photonics and silicon photonic weight banks.

Interest in integrated lasers with neuron-like spiking behavior has flourished over the past several years [4]. These lasers exhibit excitable (a.k.a spiking) dynamics where the outputs are pulses of fixed amplitude and continuous in time. The differential equations describing a laser cavity with embedded saturable absorber can be designed to map exactly to those of a common neuron model called leaky integrate-and-fire [5], except on timescales roughly $10^7\times$ faster than their biological

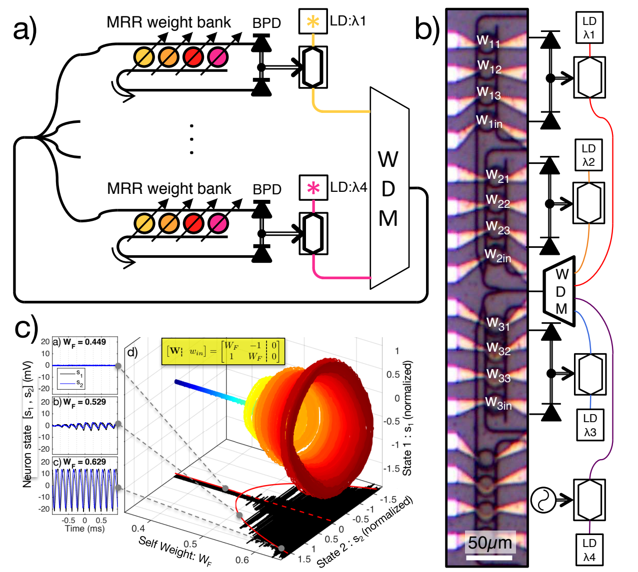

Fig. 1. Silicon Photonic Neural Networks. a) Broadcast-and-weight architecture. b) Demonstrated broadcast-and-weight network based on microring weight banks. c) Observed oscillatory bifurcation in the experiment in (b).

counterparts. Experimental work on excitable lasers has so far focused on isolated neurons [6] and fixed, cascadable chains [7], [8]. The shortage of research on networks of these lasers might be explained by the challenges of implementing low-loss, compact, and tunable filters in the active III/V platforms required for laser gain.

The "broadcast-and-weight" architecture (Fig. 1a) was proposed as a protocol for implementing networks of photonic neurons using integrated photonic devices [9]. Neuron outputs are wavelength division multiplexed (WDM) and broadcast to other neurons. At each neuron input, these WDM signals are weighted by reconfigurable, continuous-valued filters called microring (MRR) weight banks and then summed by total power detection. This electrical weighted sum then modulates the corresponding WDM carrier through a nonlinear electro-optic device [10]. The first demonstration of a broadcast-and-weight network [11] established a dynamical correspondence with a continuous-time recurrent neural network (CTRNN) model. Evidence of this dynamical correspondence was found by reproducing bifurcations expected in the CTRNN. This correspondence allows for the application of neural network design tools to analog photonic networks, making them not merely reconfigurable, but programmable. For example, a simulated continuous-time recurrent neural network with modulator-class neurons can be programmed to solve nonlinear differential equations with a 294× speedup compared to a CPU baseline [11].

MRR weight banks are the seat of reconfigurability in integrated analog photonic networks. Their performance is therefore closely tied to the potential of these overall systems. In a MRR weight bank, the transmission seen by a WDM channel is configured by thermally tuning that filter on and off resonance [12]. A bank of these filters coupled to two bus waveguides independently weights all WDM channels. A complementary –1 to +1 weight range is achieved by balanced photodetection of the multiplexed bus WGs. Techniques for extracting weight vectors from time-domain WDM measurements and for precise control of MRR weights were introduced in Ref. [13]. The demonstrated weight resolution of 4.1 bits plus a sign bit (i.e. 34 distinguishable levels) is on par with the weight resolution of digital neuromorphic electronics.

In a MRR-based WDM device, the channel count limit is determined by the finesse of the resonators and the channel spacing normalized to the filter linewidth. In MRR demultiplexers, this channel density parameter trades off with inter-channel crosstalk. In a weight bank, all WDM signals exit the same pair of waveguides, so the concept of inter-channel crosstalk breaks down. Instead, the channel density is limited by the ability to weight neighboring signals independently [14]. A unique property of the weight bank is the presence of two bus waveguides between filters that act upon neighboring WDM channels. Coherent paths can be formed for wavelengths that partially couple through the neighboring filter.

The nature of inter-filter interference is fundmentally different for MRR filters that are odd-pole vs. even-pole. In a odd-pole bank, a channel partially coupled through a neighboring filter returns through the opposite bus WG to complete a resonator-like feedback path; in an even-pole bank, the partially dropped channel continues in the same direction to instead complete an interferometer-like feedforward path. Interferometer-like interference depends on a path length difference, rather than a sum, so changes that affect both bus WGs equally should not change the interferometric phase condition. Two-pole designs investigated in [15] can exploit inter-filter interference to achieve a WDM density improvement of 3.4×. Furthermore, the inter-channel phase condition is tolerant to dynamic tuning only in 2-pole weight banks.

Neuromorphic silicon photonics [2] combines physical models of optoelectronic systems with computational models of neural networks. It represents a new opportunity for machine information processing on sub-nanosecond timescales. The strategy of neuromorphic engineering is to externalize the risk of developing computational theory alongside hardware. The strategy of remaining compatible with silicon photonics externalizes the risk of platform development. Silicon photonic manufacturing introduces unprecedented opportunities for large-scale, analog photonic systems with wide reconfigurability. By applying neural abstractions for programming

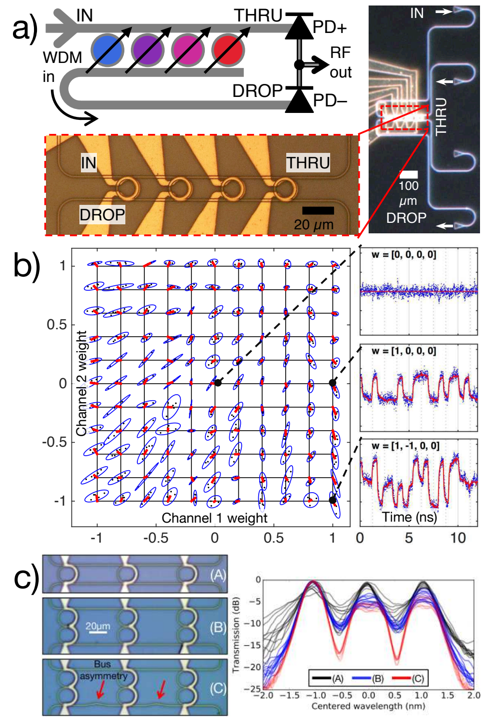

Fig. 2. Microring Weight Banks. a) Concept and device with tunable MRRs – IN, DROP, and THRU are all multiplexed. b) Precise two-channel control results – black grid is target; red is mean error; blue is variance. c) One-pole (A) vs. Two-pole (B, C) designs. Two-pole designs are robust to tuning with the asymmetric device (C) exhibiting the deepest isolation between peaks.

and learning interconnects, these systems could find application in new regimes of information processing where speed, adaptibility, and energy efficiency are paramount. Neuromorphic photonics is likely to first impact microwave signal processing and scientific computing. It might also be applied to ultrafast control and machine learning [16]. Since this ultrafast computational regime is largely unexplored, it could furthermore enable applications that are as of yet unknown.

References

[1] D. Pérez *et al.*, *Nature Communications*, vol. 8, no. 1, p. 636, 2017.
[2] P. R. Prucnal and B. J. Shastri. CRC Press, May 2017.
[3] G. Indiveri and S. C. Liu, *Proceedings of the IEEE*, vol. 103, no. 8, pp. 1379–1397, Aug 2015.
[4] P. R. Prucnal *et al.*, *Adv. Opt. Photon.*, vol. 8, no. 2, pp. 228–299, Jun 2016.
[5] M. A. Nahmias *et al.*, *IEEE J. Sel. Top. Quantum Electron.*, vol. 19, no. 5, pp. 1–12, Sept 2013.
[6] F. Selmi *et al.*, *Phys. Rev. Lett.*, vol. 112, p. 183902, May 2014.
[7] T. V. Vaerenbergh *et al.*, *Opt. Express*, vol. 20, no. 18, pp. 20 292–20 308, Aug 2012.
[8] B. J. Shastri *et al.*, *Sci. Rep.*, vol. 5, p. 19126, Dec. 2015.
[9] A. N. Tait *et al.*, *J. Lightwave Technol.*, vol. 32, no. 21, pp. 4029–4041, Nov 2014.
[10] M. A. Nahmias *et al.*, *Applied Physics Letters*, vol. 108, no. 15, 2016.
[11] A. N. Tait *et al.*, *Scientific Reports*, vol. 7, no. 1, p. 7430, 2017.
[12] A. Tait *et al.*, *Photonics Technol. Lett.*, vol. 28, no. 8, pp. 887–890, April 2016.
[13] A. N. Tait *et al.*, *Opt. Express*, vol. 24, no. 8, pp. 8895–8906, Apr 2016.
[14] ——, *IEEE J. Sel. Top. Quantum Electron.*, vol. 22, no. 6, 2016.
[15] ——, *Opt. Lett.*, vol. (peer review), 2018.
[16] T. Ferreira de Lima *et al.*, *Nanophotonics*, vol. 6, no. 3, 2017.