A Multi-Layer Topologically Reconfigurable Broadcast-and-Weight Photonic Neural Network

Joshua C. Lederman[†*], Yusuf Jimoh[†], Simon Bilodeau[†], Weipeng Zhang[†], Eric C. Blow[†], Thomas Ferreira de Lima[‡†], Bhavin J. Shastri[§], Paul R. Prucnal[†]
[†] Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA
[‡] NEC Laboratories America, Princeton, NJ 08540, USA
[§] Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada
[*] joshuacl@princeton.edu

2023 IEEE Photonics Conference (IPC) | 979-8-3503-4722-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IPC57732.2023.10360794

***Abstract*—Broadcast-and-weight (BaW) photonic neural networks can process high-bandwidth signals with limited chip area, but they traditionally lack topological reconfigurability. We propose using a fully-connected *recurrent* BaW system as a topologically reconfigurable network and demonstrate a multi-layer feedforward network implemented on such a system.**

I. Introduction

In a broadcast-and-weight (BaW) photonic neural network, neurons receive all inputs on a single waveguide, each encoded on a distinct wavelength of light—wavelength-division multiplexing [1]. Tunable micro-ring resonators (MRRs) transfer a portion of each wavelength onto a second waveguide, the "drop" waveguide, leaving the remainder on the "through" waveguide. The two waveguides terminate at two identical photodetectors (PDs), which are connected such that their photocurrents subtract. The output is a current corresponding to the weighted addition of all input signals. The current drives an output micro-ring modulator with a nonlinear transfer function, implementing neuron nonlinearity while also re-encoding the signal on an optical carrier for future propagation.

BaW networks are limited by their lack of full reconfigurability: the network topology is, in general, hard-coded during device fabrication. This can significantly extend the length of the design cycle and limit the breadth of applicability of such networks.

To address this limitation, we propose to apply a fully-recurrently-connected BaW network as a topologically reconfigurable feedforward or recurrent network. With wavelength-division multiplexing, fully-recurrent networks become highly area-efficient, as all connections may be implemented with a single waveguide. All neurons encode their outputs on carriers within this waveguide, and the waveguide then splits to provide all signals to all neurons, as shown in Fig. 2a. Any other network can be created from a fully-recurrent network by disabling connections, assuming enough neurons are available. However, ensuring the appropriate connections remain disabled even as other weights are tuned is made challenging by micro-ring cross-talk and other non-idealities. Nevertheless, we successfully demonstrate a multi-layer two-neuron feedforward BaW network implemented on top of a three-neuron fully-recurrently connected integrated silicon photonic system. To the authors' knowledge, this work represents the first experimental demonstration of feedforward cascaded BaW neurons.

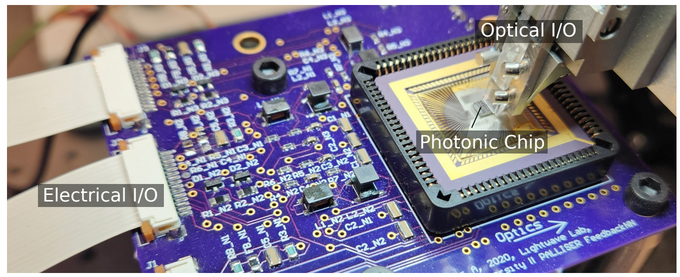

Fig. 1. Experimental Setup. The photonic chip is interfaced electrically through a custom PCB and wirebonds and interfaced optically via a fiber V-groove array.

II. Methods

Fig. 1 shows the experimental setup. An integrated silicon photonic chip containing the network under test is wirebonded to enable connection to a printed circuit board (PCB) with control devices. A number of Keithley 2606B source-measurement units provide analog electrical input/output (I/O) through the PCB. External lasers provide optical signals that are coupled onto the chip via a fiber V-groove array and grating couplers, with one signal modulated in intensity using an HP 8157A variable optical attenuator (VOA). All other optical processing and detection occurs on-chip. Each of the three integrated neurons include four MRRs with resonances tuned by N-doped heaters. On chip germanium-on-silicon photodetectors enable optoelectric conversion. Each neuron's output current drives a PN-junction micro-ring modulator, which may also be tuned thermally using embedded resistive heaters. Optical intensity in a "through" waveguide corresponds to negative weighting current and drives the associated PN-junction micro-ring in forward bias, resulting in a strong modulation effect. Output modulators are thermally tuned such that in the absence of PN-junction current they are at maximum resonance and thereby allow minimum intensity to pass. Any forward-bias current results in increased output intensity.

We implement the two-neuron network shown in Fig. 2b. A single external input signal encoded onto an optical carrier is

979-8-3503-4722-7/23/$31.00 ©2023 IEEE

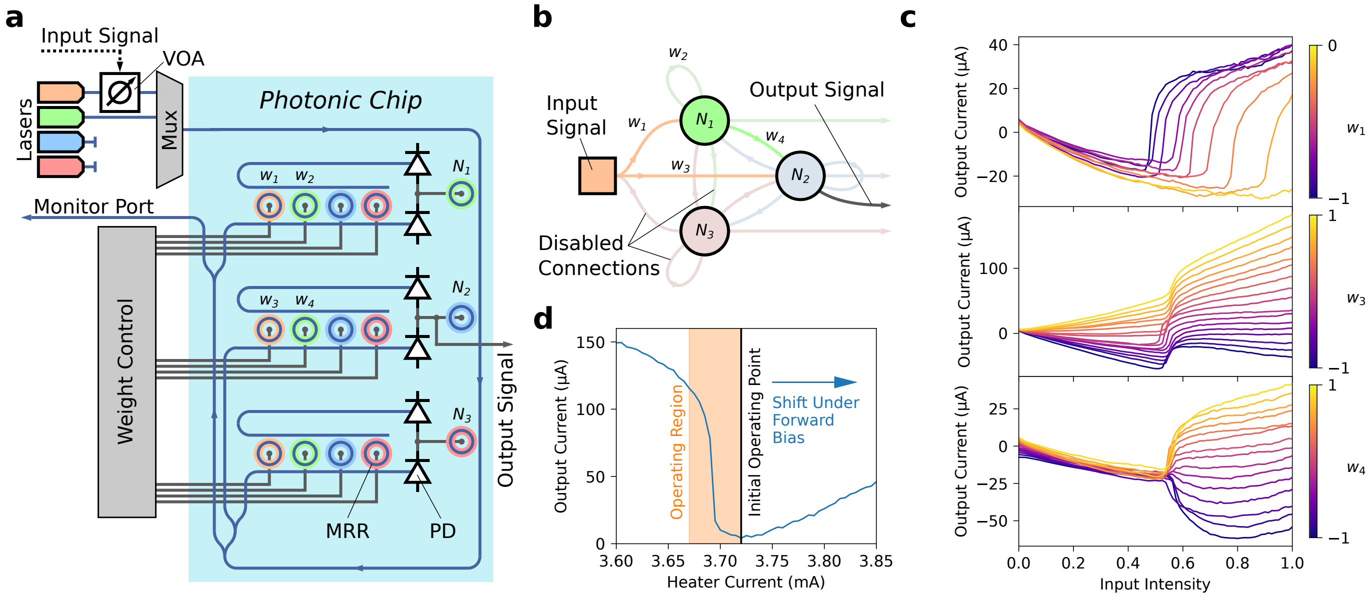

Fig. 2. Experimental setup and data. (a) Diagram of the experimental setup and integrated chip devices. (b) Diagram of the network topology during the experiment. (c) Neural network transfer function while sweeping each activated weight. Default weights: $w_1 = -0.8$, $w_3 = -0.3$, $w_4 = 1$. (d) Nonlinear activation function of the photonic neurons, collected by measuring total photocurrent while sweeping the level of heating of one output modulator.

received by both neuron 1 (N_1) and neuron 2 (N_2), after being weighted by the appropriate MRR—w_1 and w_3, respectively. The third neuron is available on-chip but disabled. A second optical carrier is modulated by the output modulator of N_1 and loops back to both active neurons. The recurrent connection, weighted by w_2, must be actively maintained at zero to prevent hysteresis or other recurrent artifacts. w_4 controls the connection between the neurons.

MRR calibration is implemented as a two-step process. Each MRR is first current-swept individually to identify its approximate zero-weight current (i.e. the applied thermal current resulting in minimum neuron output current). Once the full set of zero-weight currents has been approximated, all currents are set, and neuron output currents are further minimized. The Nelder-Mead derivative-free gradient descent algorithm is applied to simultaneously optimize all weights, similar to the approach of [2]. Online optimization of the weights ensures the effect of MRR crosstalk is mitigated. With accurate zero-weight points identified, positive or negative weights may be set with deviations to either side of the identified currents.

To optimize the N_1 activation function, the associated optical carrier's power is raised sufficiently to induce optical nonlinearity in the output modulator, though not so high as to result in hysteresis. The nonlinearity sharpens the transition of the modulator away from resonance under forward-bias current, resulting in a well-defined neuron nonlinearity, as shown in Fig. 2d. The N_2 output nonlinearity is disabled, with the output current directly measured.

III. Results and Discussion

Fig 2c shows the transfer function of the neural network as each weight is swept. The single-input, single-output network can be independently controlled along three distinct dimensions, behavior that cannot be replicated without multi-layer operation and neuron nonlinearity. These data validate the BaW architecture for multi-layer neural network operation and indicate that a recurrent architecture represents a feasible method of achieving topological reconfigurability for these systems. Though such a fully-connected network is limited to dozens of neurons by the micro-rings' free spectral range, small photonic neural networks have shown significant promise toward applications that require high bandwidth and low latency processing, including model-predictive control [3] and submarine fiber nonlinearity compensation [4]. Achieving topological reconfigurability overcomes a significant barrier to maturation and broader application of BaW networks.

Acknowledgment

J.C.L. acknowledges support from the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

References

[1] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *J. Lightwave Technol.*, vol. 32, no. 21, pp. 3427–3439, Nov 2014.

[2] J. C. Lederman, W. Zhang, T. F. de Lima, E. C. Blow, S. Bilodeau, B. J. Shastri, and P. R. Prucnal, "Real-Time Blind Photonic Interference Cancellation for mmWave MIMO," May 2023. [Online]. Available: http://arxiv.org/abs/2305.05528

[3] T. F. de Lima, H.-T. Peng, A. N. Tait, M. A. Nahmias, H. B. Miller, B. J. Shastri, and P. R. Prucnal, "Machine learning with neuromorphic photonics," *Journal of Lightwave Technology*, vol. 37, no. 5, pp. 1515–1534, 2019.

[4] C. Huang, S. Fujisawa, T. F. de Lima, A. N. Tait, E. C. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H.-T. Peng, H. G. Batshon, B. J. Shastri, Y. Inada, T. Wang, and P. R. Prucnal, "A silicon photonic–electronic neural network for fibre nonlinearity compensation," *Nature Electronics*, vol. 4, no. 11, pp. 837–844, Nov 2021.