

Bicky A. Marquez¹, Zhimu Guo¹, Hugh Morison¹, Sudip Shekhar², Lukas Chrostowski², Paul Prucnal³ and Bhavin J. Shastri¹

²*Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.*

³Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA.

The Author(s)

One of the primary bottlenecks of digital implementations of perceptron-based neural networks is efficiently computing the matrix multiplications required for training and inference [1]. Perceptrons encompass the most general functions of biological neurons, which can be summarized as weighted-additions nonlinearly transformed by activation functions [2]. Weighted-additions also represent the core operation for dot products between matrices. Therefore, dedicated hardware accelerators for perceptron-based neural networks would be designed to perform weighted-additions through operations between matrices.

Dedicated analog hardware for artificial neural networks (ANNs) is expensive if we consider that modern deep networks sizes scale up to thousands (or millions) of neurons to solve AI related tasks. In order to overcome such challenge, the high speed and parallelism that analog photonic systems can achieve makes them natural candidates for efficient brain-inspired computing [3].

In this work, we present a photonic integrated circuit able to perform online training and testing based around matrix multiplications. A perceptron-based Hopfield ANN will be trained and tested for a pattern reconstruction task [4]. The task consists in training and testing this recurrent network three times to recognize three different corrupted patterns. These patterns are represented as 4×4 matrices that model the image of numbers 0, 1 and 2 with some defects. Due to their speed, energy efficiency and reconfigurability, such matrix multiplications will be performed using photonic devices. In particular, we consider the use of the broadcast-and-weight protocol (BWP) which uses micro ring-resonators (MRRs) to directly encode different matrix elements as amplitude values in parallel optical channels [5].

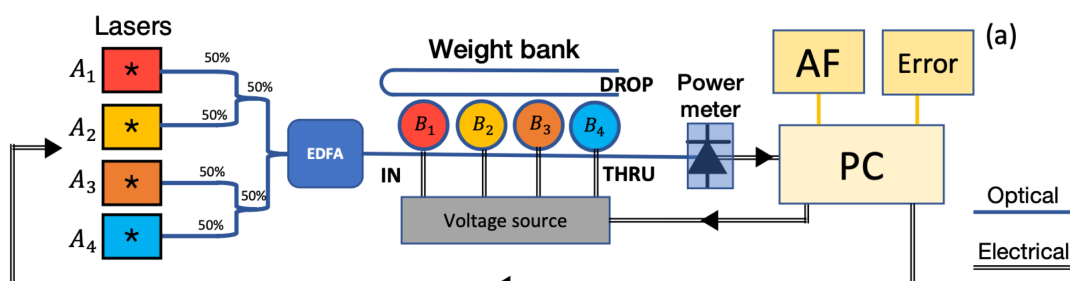


Fig. 1. Schematic illustration of the photonic vector dot product architecture between vectors \mathbb{A} and \mathbb{B} .

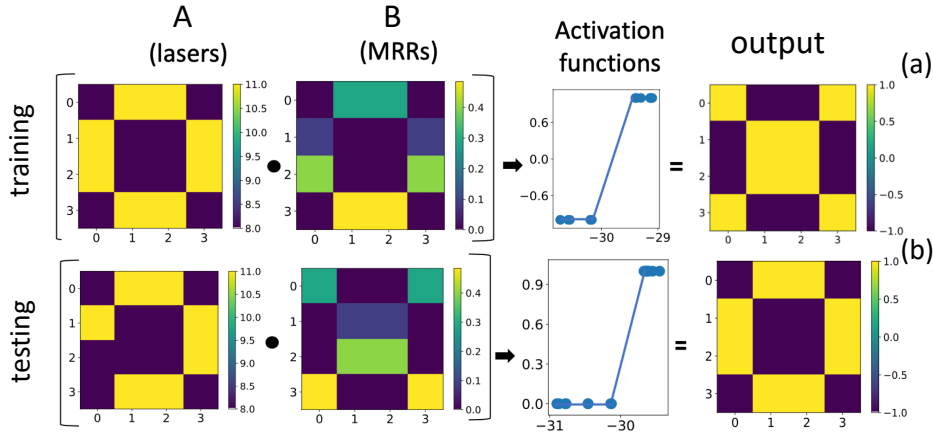


Fig. 2. (a) Training and (b) testing stages of the Hopfield network to memorize and reconstruct image 0, respectively.

2. Photonic Hopfield neural network

A hopfield architecture is a recurrent neural network typically used as an associative memory [6]. Hopfield's network memory property allows for pattern reconstruction of faulty datasets. In this work, only one pattern will be stored in memory. Therefore, for this particular task a weight matrix W can be estimated by multiplying every input pattern X with itself ($W = x^T \cdot x$, where T is the transpose function). In the inference stage we multiply a corrupted image y by the weight matrix W , and apply a sign activation function to that result, obtaining $O = \text{sign}(y \cdot W)$. The results of such experiments are determined by the estimation of the mean absolute error (MAE).

As the operations described above are based on dot products between vectors containing four elements each, we will demonstrate that such operations can be performed using a bank of four on-chip silicon MRRs. Let us define two vectors $\{\mathbb{A}, \mathbb{B}\}$ that will represent any theoretical set of vectors considered in the forthcoming experimental dot products ($\mathbb{A} \cdot \mathbb{B}$). In Fig. 1, the elements of such vectors are experimentally represented. To encode elements of the vector \mathbb{A} , we vary the power intensity P_i (with $i = 1, 2, 3, 4$) of four tunable lasers. The elements of the second vector \mathbb{B} are implemented by four on-chip add-drop MRRs, when the waveguide refractive index by means of an applied voltage V_i is tuned.

Training and testing the Hopfield network to recognize one pattern x at a time requires the estimation of a weight matrix W through the dot product between input matrices $x^T \cdot x$. In Fig. 2(a), we show an example of how to estimate the weight matrix (output) for an image of number zero, using lasers encoding x^T and MRRs encoding x . Figure 2(b), shows the result of testing (output) for an image of number zero, using lasers encoding y and MRRs encoding W . Post-processed weight and output matrices reached accuracies for over 85.6% not only for image zero, but also for image of numbers one and two. In the case of number zero, the accuracies were 100%.

3. Conclusion

We have shown that MRR-based photonic integrated circuits can implement training and inference stages of a Hopfield network for pattern reconstruction, based on experimental vector operations. The reconfigurability feature of our photonic circuits allowed for the design of a special-purpose analog Hopfield network.

References

1. B. J. Shastri, et al., "Photonics for artificial intelligence and neuromorphic computing," arXiv:2011.00111 (2020).
2. R. Rojas, *Neural Networks: A Systematic Introduction* (Springer, 1996).
3. P. Prucnal and B. J. Shastri, *Neuromorphic photonics* (CRC Press, 2017).
4. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," PNAS **79**, 2554–2558 (1982).
5. H. Chaoran et al., "Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits," APL Photonics **5**, 040803 (2020).
6. K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey," IEEE Journal of Solid-State Circuits **41**, 712–727 (2006).