Photonic Long-Short Term Memory Neural Networks with Analog Memory

Emma R. Howard, Bicky A. Marquez, Bhavin J. Shastri
Department. of Physics, Engineerignng Physics & Astronomy, Queen's University, Kingston, Canada
emma.howard@queensu.ca

***Abstract*—A Photonic implementation is proposed for the Long-Short Term Memory neural network, offering fundamental speed and bandwidth advantages over digital electronic implementations. Integrated analog memory for photonics is designed as a component of this network.**

I. Introduction and Motivation

Photonic hardware is favorable for applications requiring high-bandwidth transmission, such as the global internet. Recent innovations in silicon photonic fabrication have enabled on-chip implementation of photonic circuits. This has opened a low-cost, high-precision, and scalable avenue for the development of photonic computing. Advances in photonic computing have demonstrated suitability for applications requiring high-bandwidth parallel processing, especially neural networks [1]. So-called 'Neuromorphic Photonic' designs offer greater speed and less energy consumption than equivalent networks implemented in digital electronics [1].

The Long-Short Term Memory (LSTM) neural network is a recurrent architecture that offers advantages for time-series processing [2]. Neuromorphic photonic LSTMs offer a solution to growing demand for high-speed, high-bandwidth neural networks in time-series applications, including video processing, autonomous driving and optical communications [3][4][5]. Here, we present a photonic architecture for implementing LSTM networks on an integrated silicon photonics platform with analog electronics. The network is designed to exploit fundamental speed, bandwidth, and efficiency advantages offered by photonic computing designs.

As part of this design, an analog electronic memory system for photonics is introduced. Analog memory provides many advantages in photonic processing, eliminating speed limitations introduced by ADCs and DACs [6]. Optical memory research has only recently shifted focus from elementary memory units to functionally complete memory cells, and many optical memory technologies still rely on advanced materials and fabrication processes [7]. Co-integrated analog electronic memories with photonics take advantage of the benefits of analog memory while avoiding the challenges of working with experimental optical memory devices.

II. Photonic LSTM Architecture

A photonic network architecture was developed such that the governing equations describing the physical circuit are isomorphic to the equations describing the LSTM node [8]. An annotated photonic circuit design is seen in Figure 1c, alongside a schematic of an LSTM node in Figure 1a. This circuit has a similar recurrent architecture to the broadcast-weight photonic recurrent neural network, a neuromorphic photonic design also based on isomorphism [9]. Simulation and in-lab performance of the broadcast-weight network validates several design choices in the proposed LSTM, including the microring resonator (MRR) weight banks, used to implement the multiple weight-and-sum operations in the neurons, and the use of wavelength-division multiplexing to carry vector data in the system [9]. Compared to the proven multiwavelength (i.e broadcast-and-weight) network [10], additional photonic devices in the LSTM enable analog memory and element-wise multiplication as required. Addition is accomplished by the superposition of coherent intensity-encoded signals, concatenation is accomplished by WDM.

An analog memory circuit was designed to store the LSTM 'cell state' vector. This memory system eliminates the latency introduced by Analog-To-Digital conversion which can be a limiting factor in high-bandwidth systems [6], and takes advantage of electronic technology that can be fabricated on-chip. The memory cell is an analog electronic circuit, but it is designed with integrated photonic edge devices: A photodiode detects incoming laser pulses and a microring modulator (MRM) writes stored values to the photonic domain. The system is designed to trigger on the rising edge of a photonic pulse and write the analog pulse intensity value to a two-transistor electronic memory cell with write and read capability. The memory system can be seen in figure 1b. The design was simulated in SPICE using equivalent circuit models for the photodiode and MRM [11]. The memory circuit was shown to accurately capture pulses as short as 0.3 μs, and store values without refresh in a 16-level analog memory scheme for up to 130 μs. Tunable amplifiers within the design make it possible to optimize the memory for a variety of system speeds, wavelengths, and intensities, and ensure operation in the linear region of amplification.

A novel MRM layout is introduced to implement analog multiplication at the electronic/optical barrier. The approach consists of two linearly controlled microring modulators for the same wavelength in the all-pass configuration,

978-1-7281-5891-4/20/$31.00 ©2020 IEEE

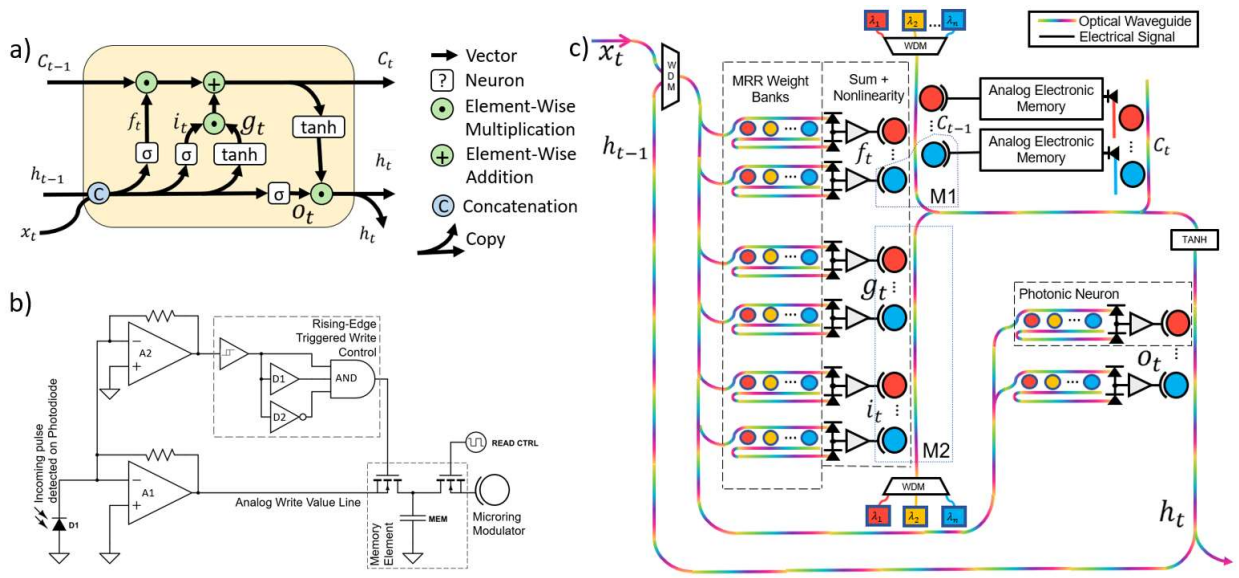

Figure 1: a) Schematic of operations in a recurrent LSTM, based on equations presented in Hochreiter et. al. [8] b) Overview of an analog electronic memory circuit, showing devices that cross the electro-optic barrier (photodiode and MRM) c) High-level overview of photonic long-short term memory design. The photonic neuron and its constituent elements (MRR Weight Banks and sum/nonlinearity) are called out, and two examples of MRM multiplication are highlighted. Ellipses within vectors and weight banks are included to indicate that the design can be expanded to include larger vectors, by including more wavelengths.

arranged in series along a single waveguide, for example the multiplication of vector elements of f_t and c_t (Box M1) in Figure 1c. This design can be scaled to accommodate multiple wavelengths on the same waveguide, as seen in Figure 1c (Box M2). Simulations demonstrated modulator multiplication to be within 10% of ideal multiplication on approximately one quarter of the dynamic range of a photonic computing system, making it an effective solution when coupled with amplifiers.

III. Discussion and Conclusion

A transfer-level simulation of the full design was constructed in Python. Training and optimization at the network level are ongoing, simulations will be used to create a physical layout for the network. The LSTM is designed using devices that can be fabricated with standard silicon foundry process, making this design relatively inexpensive and scalable [12]. Either flip-chip or wire-bond connections are planned to connect silicon photonic and electronic components. Fabrication is anticipated in November 2020.

The photonic LSTM design proposed presents an advance in neuromorphic photonics by enabling the effective handling of time-series data. The analog memory for photonics proposed as a part of this design provides a novel and scalable approach to memory in photonic circuits. Throughout the design process, ease of fabrication and scalability for this system have been emphasized, to enable rapid development and market readiness. With enormous opportunities in the field of high-speed high-bandwidth real-time neural network applications, the photonic LSTM with analog memory has great research and market potential.

References

[1] T. F. de Lima *et al.*, "Progress in neuromorphic photonics," *Nanophotonics*, vol. 6, no. 3, pp. 577–599, 2017.
[2] K. Greff *et al.*, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[3] A. Alahi *et al.*, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 961–971.
[4] F. Altché and A. de La Fortelle, "An LSTM network for highway trajectory prediction," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct. 2017, pp. 353–359.
[5] F. N. Khan *et al.*, "An Optical Communications Perspective on Machine Learning and Its Applications," *J. Light. Technol.*, vol. 37, no. 2, pp. 493–516, Jan. 2019.
[6] J. Meng *et al.*, "Electronic bottleneck suppression in next-generation networks with integrated photonic digital-to-analog converters," arXiv: 1911.02511, Dec. 2019.
[7] T. Alexoudi *et al.*, "Optical RAM and integrated optical memories: a survey," *Light Sci. Appl.*, vol. 9, no. 1, p. 91, May 2020.
[8] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[9] A. N. Tait *et al.*, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Aug. 2017.
[10] A. N. Tait *et al.*, "Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing," *J. Light. Technol.*, vol. 32, no. 21, pp. 3427–3439, Nov. 2014.
[11] T. F. de Lima *et al.*, "Noise Analysis of Photonic Modulator Neurons," *IEEE J. Sel. Top. Quantum Elec.*, vol. 26, no. 1, pp. 1–9, Jan. 2020.
[12] W. Bogaerts *et al.*, "Silicon Photonics Circuit Design: Methods, Tools and Challenges," *Laser Photonics Rev.*, vol. 12, no. 4, 2018.