Neuromorphic Photonics: Current Status and Challenges

Paul R. Prucnal[(1)], Thomas Ferrieira de Lima[(1)], Chaoran Huang[(1)], Bicky A. Marquez[(2)], Bhavin J. Shastri[(1,2)]

[(1)] Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, prucnal@princeton.edu
[(2)] Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada, shastri@ieee.org

Abstract *Artificial intelligence and neuromorphic computing driven by neural networks has enabled many applications. Software implementations of neural networks on electronic platforms are limited in speed and energy efficiency. Neuromorphic photonics aims to build processors in which optical hardware mimic neural networks in the brain.*

Introduction

The field of neuromorphic (i.e., neuron-isomorphic) computing aims to bridge the gap between the energy efficiency of von Neumann computers and the human brain [1], [2]. The rise of neuromorphic computing can be attributed the widening gap between current computing capabilities and current computing needs [3], [4]. Consequently, this has spawned research into novel brain-inspired algorithms and applications uniquely suited to neuromorphic processors. These algorithms attempt to solve artificial intelligence (AI) tasks in real-time while using less energy. We posit that we can make use of the high parallelism and speed of photonics to bring the same neuromorphic algorithms to applications requiring multiple channels of multi-gigahertz analog signals, which digital processing struggles to process in real-time.

By combining the high bandwidth and parallelism of photonic devices with the adaptability and complexity attained by methods similar to those seen in the brain, photonic neural networks have the potential to be at least ten thousand times faster than state-of-the-art electronic processors while consuming less energy per computation [5]. An example of such an application is nonlinear feedback control; a very challenging task that involves computing the solution of a constrained quadratic optimization problem in real time. Neuromorphic photonics can enable new applications because there is no general-purpose hardware capable of dealing with microsecond environmental variations [6].

Neuromorphic photonics approaches

Neuromorphic photonic [7] approaches can be divided into two main categories: coherent (single wavelength) and incoherent (multiwavelength) approaches. Neuromorphic systems based on reservoir computing [8]–[10] and Mach-Zehnder interferometers [11], [12] are example of coherent approaches. In reservoir computing the predefined random weights of their hidden layers cannot be modified. An alternative approach uses silicon photonics to design fully programmable neural networks [6], with a so-called broadcast-and-weight protocol [13],[14]. In this architecture, photonic neurons output optical signals with unique wavelengths. These are multiplexed into a single waveguide and broadcast to all others, weighted, and photodetected. Each connection between a pair of neurons is configured independently by one microring resonator (MRR) weight, and the wavelength division multiplexed (WDM) carriers do not mutually interfere when detected by a single photodetector. Consequently, the physics governing the neural computation is fully analog and does not require any logic operation or sampling, which would involve serialization and sampling. Thus, they exhibit distinct, favorable trends in terms of energy dissipation, latency, crosstalk and bandwidth when compared to electronic neuromorphic circuits [5]. The advantage of this approach over the aforementioned approaches is that it has already demonstrated fan-in, inhibition, time-resolved processing, and autaptic cascadability [15].

However, the same physics also introduce new challenges, especially reconfigurability, integration, and scalability. Information carried by photons is harder to manipulate compared to electronic signals, especially nonlinear operations and memory storage. Photonic neurons described here solve that problem by using optoelectronic components (O/E/O), which can be mated with standard electronics providing reconfigurability. However, neuromorphic photonic circuits are challenging to scale up because they do not benefit from digital information, memory units and a serial processor, and therefore requires a physical unit for each element in a neural network, increasing size, area and power consumption. Here, integration costs

2020 European Conference on Optical Communications (ECOC) | 978-1-7281-7361-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/ECOC48923.2020.9333362

must also be considered, since the advantages of using analog photonics (high parallelism and high bandwidth) must outweigh the costs of interfacing it with digital electronics (requiring both O/E and analog/digital conversion).

Vision of a neuromorphic processor

Recently, in our tutorial, Ref. [16], we proposed a vision for a neuromorphic processor. We discussed how such a neuromorphic chip could potentially be interfaced with a general-purpose computer (Fig. 1), i.e. a CPU, as a coprocessor to target specific applications. Broadly speaking, there are two levels of complexity associated with co-integrating a general-purpose electronic processor with an application-specific optical processor. Firstly, a CPU processes a series of computation instructions in an undecided amount of time and is not guaranteed to be completed. Neural networks, on the other hand, can process data in parallel and in a deterministic amount of time. CPUs have a concept of a 'fixed' instruction set on top of which computer software can be developed. However, a neuromorphic processor would require a hardware description language (HDL) because it describes the intended behavior of a hardware in real-time. Secondly, seamlessly interfacing a photonic integrated circuit with an electronic integrated circuit will take several advances in science and technology including on-chip lasers and amplifiers, co-integration of CMOS with silicon photonics, system packaging, high-bandwidth digital-to-analog converters (DAC) and analog-to-digital converters (ADCs).

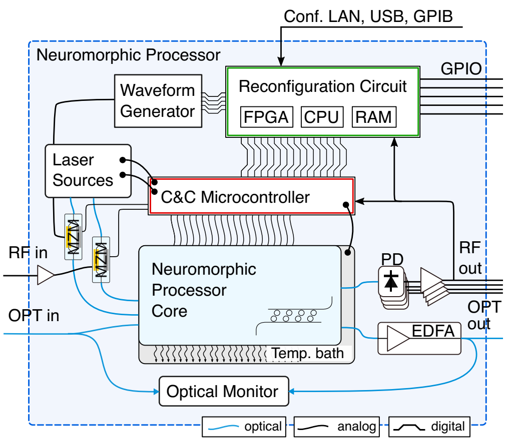

Fig. 1: Simplified schematics of a neuromorphic processor. Thanks to integrated laser sources and photodetectors, it can input and output RF signals directly as an option to optically modulated signals. The waveform generator allows for programming arbitrary stimulus that can be used as part of a machine learning task. Reproduced from [6].

Application example: fiber nonlinearity impairment compensation

Neuromorphic photonic processors are well suited for applications in which signals are in the analog and/or optical domain. This alleviates some of the I/O challenges associated with DACs. One such application is in fiber nonlinearity compensation in long-haul transmission systems.

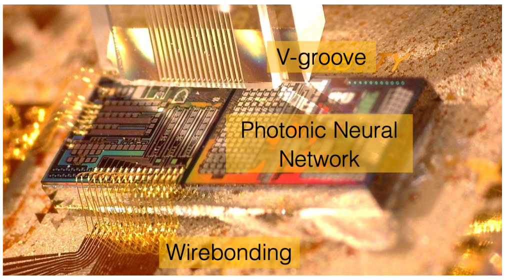

Fig. 2: Micrograph image of a wirebonded photonic neural network chip from Princeton University's Lightwave Lab.

Artificial neural networks (ANNs) have been demonstrated for optical fiber communication, such as fiber nonlinearity compensation (NLC) in long-haul transmission systems [17]. Benefiting from the training and execution procedures of ANNs, ANN-NLC algorithms can create effective fiber transmission models from the received symbols without needing prior knowledge of transmission link parameters. Compared with the deterministic NLC approaches, such as digital back propagation [18], ANN-NLC provides comparable system performance with lower computational complexity. However, despite the reduced complexity with ANN-NLC, the hardware implementation of real-time ANN-NLC for high-speed optical transmission systems is still a challenge with conventional electronics (e.g., ASIC), considering the required computation speed and associated power consumption. The challenges in high circuit complexity, together with tight power budget, have prohibited implementing high-performance but computationally intensive DSP algorithms like ANNs in real time. So far, most efforts have been focused on developing new algorithms that requires compromises between transmission link performance and DSP complexity.

Applications such as ANN-NLC for optical communications demand for low-power and high-speed neural network implementation, and therefore necessitates the investigation of new hardware beyond purely electronic physics. Neuromorphic photonic processors are ideal for processing high-speed optical communication signals. Our prior research on PNN has revealed the analogy between the neural networks and WDM photonic hardware and demonstrated underlying on-chip devices that allow practical implementation on silicon photonic platforms [14,15]. The advances of silicon photonics enable integrations of optical devices and interconnects with sufficient density to perform computing tasks driven by real-world applications [6].

In OFC 2020 [19] demonstrated the experimental demonstration of a neuromorphic photonic processor (Fig. 2) to compensate for fiber nonlinearity over a 10,080 km trans-Pacific transmission link of 32 Gbaud PM-16QAM signals. By utilizing this photonic processor, we have achieved Q-factor improvement of 0.51 dB, which is only 0.06 dB lower than implementing the ANN with numerical simulation. The superior precision of photonic processor demonstrates the feasibility of using it for optical fiber transmission applications. Although the bandwidth of our current chip is limited, caused by the low extinction ratio of the modulator on chip, it can be realistically increased to accommodate the high-speed communication signals in future iterations [11]. Given such bandwidths, neuromorphic photonic processors could allow real-time ANN-enabled signal processing for high-speed communication signals with a single pipeline.

Advances in Science and Technology to Meet Challenges

In the recent roadmap article [21], we outlined some scientific and technological advances necessary to meet the challenges to envision a neuromorphic processor outlines in the previous section.

Photonic processors have light sources, passive and active devices. Currently, there is no single commercial fabrication platform that can simultaneously offer devices for light generation, wavelength multiplexing, photodetection, and transistors on a single die; state-of-the-art devices in each of these categories use different photonic materials (SiN, Ge, InP, GaAs, 2D materials, etc) with incongruous fabrication processes (silicon-on-insulator, CMOS, FinFETs). Silicon photonics is becoming an ideal platform for integrating these devices while offering a combination of foundry compatibility, device compactness, and cost that enables the creation of scalable photonic systems on chip.

Materials: Energy efficient and fast switching optical and electro-optical materials are needed for non-volatile photonic storage and weighting, as well as high-speed optical switching and routing, with low power consumption. Neural non-linearities are already possible on mainstream platforms using electrooptic transfer functions [15], but new materials promise significant performance opportunities. Phase change materials (PCMs), and graphene and ITO-based modulators can also be utilized for implementing non-linearities. Plasmonic PCMs can bridge the optical and electrical signals, through the dual operation modes [22]. A general material design method is in urgent need to develop appropriate photonic materials for different photonic components [23].

Lasers and amplifiers: On-chip optical gain and power will require co-integration with active InP lasers and semiconductor optical amplifiers. Current approaches involve either III–V to silicon wafer bonding (heterogeneous integration) or co-packaging with precise assembly (hybrid approach) [24]. Quantum dot lasers are another promising approach as they can be grown directly onto silicon, but fabrication reliability does not currently reach commercial standards [25].

Electrical control: Co-integrating CMOS controller chips with silicon photonics to provide electrical tuning control/stabilization will be critical. Candidates include wire-bonding, flip-chip bonding, 2.5D integration (interposers), 3D stacking (through-silicon-vias), and monolithic integration. Each has performance and design tradeoffs [26].

System packaging: A photonic processor must be inter- faced with a computer. It would need to be self-contained, robust to temperature fluctuations, and with electrical inputs/outputs [6]. Currently, manufacturers do not assemble electrical/thermal elements and chip-to-fiber interconnects.

Algorithms: Significant advances will be required to map abstract neural algorithms to photonic processor to usher these platforms into the commercial space. So far, only individual devices and small control circuits are described in the literature. The goal is to enable neural network programming tools (TensorFlow) to directly reconfigure a neuromorphic photonic processor [6].

Conclusion

Neuromorphic photonics has reached an inflection point, benefiting from great opportunities as the world looks for alternative processor architectures. The physical limits of Dennard scaling is galvanizing the community to put forward candidates for next generation computing, from bio- to quantum computers. Photonics and in particular neuromorphic photonics are a formidable candidate for analog reconfigurable processing. We expect the development of this field to accelerate as neuroscience makes further leaps towards our understanding of the nature of cognition and artificial intelligence demands more computational resources for machine learning. As photonics technology matures and becomes more accessible to academic groups and small companies, we expect this acceleration to continue.

References

[1] B. Marr et al. IEEE Trans. Very Large Scale Integr. Syst. 21, 147 (2013).

[2] J. Hasler and H. B. Marr, Front. Neurosci. 7 (2013).

[3] P. A. Merolla et al. Science 345, 668 (2014).

[4] M. Davies et al. IEEE Micro 38, 82 (2018).

[5] T. Ferreira de Lima et al. Nanophotonics 6, 577 (2017).

[6] T. Ferreira de Lima et al. J. Lightw. Technol. 37, 1515 (2019).

[7] P. R. Prucnal and B. J. Shastri. Neuromorphic Photonics (CRC Press, 2017).

[8] D. Brunner et al. Nat. Commun. 4, 1364 (2013).

[9] K. Vandoorne et al. Nat. Commun. 5, 3541 (2014).

[10] L. Larger et al. Opt. Express 20, 3241 (2012).

[11] Y. Shen et al. Nat. Photon. 11, 441 (2017).

[12] T. W. Hughes et al. Optica 5, 864 (2018).

[13] A. N. Tait et al. J. Lightwave Technol. 32, 4029 (2014).

[14] A. N. Tait et al. Sci. Rep. 7, 7430 (2017).

[15] A. N. Tait et al. Phys. Rev. Appl. 11, 064043 (2019).

[16] T. Ferreira de Lima et al. Nanophotonics 9, 4055 (2020).

[17] S. Zhang et al. Nat. Commun. 10, 3033 (2019).

[18] X. Li et al., Opt. Express 16, pp. 880-888 (2008).

[19] C. Huang et al. in Opt. Fiber Commun. (OFC) Conf. Postdeadline Papers (2020), paper Th4C.6.

[20] K. Nozaki et al. Nat. Photon. 13, 454 (2019).

[21] K. Berggren et al 2020 Nanotechnology 32, 012002 (2020).

[22] N. Farmakidis et al. Sci. Adv. 5, eaaw2687 (2019).

[23] W. Zhang et al. Nat. Rev. Mater. 4, 150 (2019).

[24] D Liang and J. E. Bowers, Nat. Photon. 4, 511 (2010).

[25] S. Chen et al. Nat. Photon. 10, 307 (2016).

[26] A.H. Atabaki et al. Nature 556, 349 (2018).