Emergence of Neuromorphic Photonics

B. J. Shastri, T. Ferreira de Lima, A. N. Tait, M. A. Nahmias, and P. R. Prucnal

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

shastri@ieee.org, prucnal@princeton.edu

Abstract: Digital electronics are becoming the limiting factor in signal processing areas such as radio. Photonic integration offers new potential for moderate-scale photonic systems. Neuromorphic Photonics bridges integrated photonic physics and neural models of information processing.

OCIS codes: (070.1170) Analog optical signal processing; (200.4700) Optical neural systems; (250.5300) Photonic integrated circuits.

The success of digital electronics has created a data-hungry consumer society, which in turn, reinvested in more capable, faster and cheaper machines. For decades, the transistor count of CPUs doubled every two years, a trend which became known as Moore's law. Microprocessor clock rates also increased exponentially, but current leakage in nanometric nodes became prevalent, causing a halt to this growth at about 4 GHz [1]. At the same time, the past decade has seen the breakdown of Dennard scaling [2]—the power density of microelectronic chips no longer stays constant as they get denser, i.e. smaller transistors do not consume less power. The recent shift to multi-core scaling alleviated these constraints, but the breakdown of Dennard scaling has limited the number of cores than can simultaneously be powered on with a fixed power budget and heat extraction rate—giving rise to the *dark silicon* phenomenon [3]. Projections for the 8 nm node indicate that over 50% of the chip will be *dark* [3]. Fundamentally, these issues can be traced to two primary physical bottlenecks: the bandwidth limitations of metal interconnects, and the energy consumption–and subsequently, heat generation of digital switching [4]. In summary, operating speed and power efficiency of CPUs have reached physical barriers that cannot be addressed through Dennard scaling. Consequently, this has opened up new opportunities in unconventional information processing architectures, which include an array of different processing modalities [5].

Respecting power budgets is now a top priority for digital processors. Data centers, Wi-Fi routers and Internet traffic represent a tremendous electric energy consumption. Current trends indicate a shift of electricity usage from consumer device use to network and data centers [6, 7]. In the worst case scenario, at the rate at which societal consumption and production of data is growing, it is predicted that fixed-access networks (Wi-Fi and LAN) and data centers will consume up to 33% of world's energy use [7].

To counter that trend, power-aware large-scale integration techniques in photonics are just emerging, being pushed forward by data communication applications and a market need for increased information flow between processors, both on the macro and micro scale [6, 8]. This has led to an explosion in photonic integrated circuit (PICs), which are already finding their way into fast ethernet switches for servers and supercomputers, and will likely emerge in more traditional processor architectures as electronic interconnects fail to keep up with data volume. The average energy efficiency of the world's fastest supercomputers lies in the order of 1 nJ/FLOP [9], where FLOP stands for floating-point operation, a standard computing unit. In green data centers and high-performance computers, there is an urgent need for unconventional, special-purpose co-processors with efficiencies beyond 1 nJ/FLOP, with a caveat: these co-processors must operate at the same throughput handled by the high-speed digital and analog circuits it interfaces with, so they do not become a bottleneck.

This efficiency level is not fundamentally impossible. In fact, the human brain is estimated to being able to compute an amazing 10^{20} MAC/s using only 20 W of power [10] (MAC: multiply and accumulate operation, similar to FLOP but more appropriate for digital signal processors). It does this with 10^{11} neurons with spike firing rate of ~1 Hz but with a large number of interconnects per neuron (10^4), highlighting the importance of distributed processing. The calculated computational efficiency for the brain is therefore 9 orders of magnitude beyond that of current supercomputers ($<$ aJ/MAC). *Neuromorphic computing* offers hope to building large-scale "bio-inspired" hardware for specialized processing while attempting computational efficiencies toward those of a human brain.

We will review our recent progress in neuromorphic photonics research [11–16], focusing especially on integrated photonic devices. An elegant parallel between neural networks and optoelectronic devices such as excitable lasers

can be established and exploited for processing. We will introduce the concept of a *photonic neuron*, followed by a discussion on its feasibility. Finally, we will also present a scalable neurmorphic networking architecture that efficiently channelizes the spectrum of an integrated waveguide.

Photonics has revolutionized information transmission (communication and interconnects), while electronics, in parallel, has dominated information transformation (computation). This leads naturally to the following question: how can we unify the advantages of the two as effectively as possible? [17]. CMOS gates only draw energy from the rail when and where called upon; however, the energy required to driving an interconnect from one gate to the next dominates CMOS circuit energy use. Relaying a signal from gate to gate, especially using a clocked scheme, induces penalties in latency and bandwidth compared to an optical waveguide passively carrying multiplexed signals.

This suggests that starting up a new architecture from a photonic interconnection fabric supporting nonlinear opto-electronic devices can be uniquely advantageous in terms of energy efficiency, bandwidth, and latency, sidestepping many of the fundamental tradeoffs in digital and analog electronics. It may be one of the few practical ways to achieve ultrafast, complex on-chip processing without consuming impractical amounts of power [18].

Complex photonic systems have been largely unexplored due to the absence of a robust photonic integration industry. Recently, however, the landscape for manufacturable photonic chips has been changing rapidly and now promises to achieve economies of scale previously enjoyed solely by microelectronics. In particular, a new photonic manufacturing hybrid platform that combines in the same chip both active (e.g. lasers and detectors), and passive elements (e.g. waveguides, resonators, modulators) is emerging [20]. A neuromorphic photonic approach based on this platform could potentially operate 6–8 orders of magnitude faster than neuromorphic electronics[1] when accounting for the bandwidth reduction of virtualizing interconnects [19] (cf. Fig. 1).

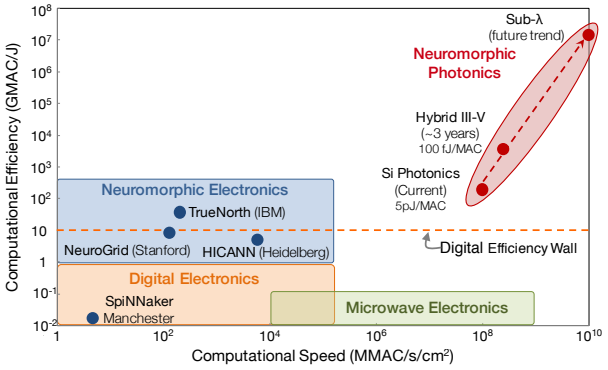

Fig. 1: Speed and efficiency metrics that are accessible by various neuromorphic hardware platforms. On the top-right: the photonic neuron platforms studied in Ref. [18]. Hybrid III-V/Si stands for III-V/Silicon hybrid platform spiking neural network photonic integrated circuit. Sub-λ stands for sub-wavelength photonics. The other points refer to recent electronic neuromorphic hardware, discussed in Ref. [19]. The regions highlighted in the graph are approximate, based on qualitative tradeoffs of each technology.

Key criteria for nonlinear elements to enable a scalable computing platform include [17, 25–27]: thresholding, fan-in, and cascadability. Past approaches to optical computing have met challenges realizing these criteria, and, so far, no optical logic device satisfying all of them has been proposed. More recent investigations, introduced in the following sections, have concluded that a photonic neuromorphic processor could satisfy them by implementing a model of a neuron, as opposed to the model of a logic gate.

Neuromorphism implies a strict isomorphism between artificial neural networks and optoelectronic devices. There are two research challenges necessary to establish this isomorphism: the nonlinearity (equivalent to thresholding) in individual neurons, and the synaptic interconnection (related to fan-in and cascadability) between different neurons. Once the isomorphism is established and large networks are fabricated, we anticipate that the computational neuroscience and software engineering will have a new optimized processor for which they can adapt their methods and algorithms.

The emerging field of neuromorphic photonics has received tremendous interest and continues to receive further developments as PICs increase in performance and scale. As novel applications requiring real-time, ultrafast processing—such as the exploitation of the RF spectrum—become more critical, we expect that these systems will find use in a variety of high performance, time-critical environments.

References

1. N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, V. Narayanan, and R. Elmquist, "Leakage current: Moore's law meets static power," Computer **36**, 68–75 (2003).

[1]Neurogrid (Stanford University's Brains in Silicon program) [21]; IBM's TrueNorth (DARPA's SyNAPSE program [22]); HICANN (U. of Heidelberg's FACETS/BrainScaleS project) [23]; and U. of Manchester's neuromorphic chip (SpiNNaker project) [24].

2. R. Dennard, F. Gaensslen, W.-N. Yu, L. Rideout, E. Bassous, and A. Le Blanc, "Design of Ion-Implanted Small MOSFET ' S Dimensions with Very," IEEE Journal of Solid State Circuits **9**, 257–268 (1974).
3. H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, "Dark Silicon and the End of Multicore Scaling," IEEE Micro **32**, 122–134 (2012).
4. D. A. B. Miller, "Attojoule Optoelectronics for Low-Energy Information Processing and Communications: a Tutorial Review," (2016).
5. M. B. Taylor, "Is dark silicon useful? Harnessing the four horsemen of the coming dark silicon apocalypse," Proc. of Design Automation Conference pp. 1131–1136 (2012).
6. C. Kachris and I. Tomkos, "A Survey on Optical Interconnects for Data Centers," IEEE Communications Surveys & Tutorials **14**, 1021–1036 (2012).
7. A. Andrae and T. Edler, "On Global Electricity Usage of Communication Technology: Trends to 2030," Challenges **6**, 117–157 (2015).
8. M. Hochberg, N. C. Harris, R. Ding, Y. Zhang, A. Novack, Z. Xuan, and T. Baehr-Jones, "Silicon photonics: The next fabless semiconductor industry," IEEE Solid-State Circuits Magazine **5**, 48–58 (2013).
9. "Green500 list," https://www.top500.org/green500/. June 2016.
10. J. Hasler and B. Marr, "Finding a roadmap to achieve large neuromorphic hardware systems," Frontiers in Neuroscience **7**, 118 (2013).
11. M. A. Nahmias, B. J. Shastri, A. N. Tait, and P. R. Prucnal, "A Leaky Integrate-and-Fire Laser Neuron for Ultrafast Cognitive Computing," IEEE Journal of Selected Topics in Quantum Electronics **19** (2013).
12. 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. Lightw. Technol. **32**, 3427–3439 (2014).
13. B. J. Shastri, M. A. Nahmias, A. N. Tait, A. W. Rodriguez, B. Wu, and P. R. Prucnal, "Spike processing with a graphene excitable laser," Scientific Reports **6**, 19,126 (2016).
14. P. R. Prucnal, B. J. Shastri, T. Ferreira de Lima, M. A. Nahmias, and A. N. Tait, "Recent progress in semiconductor excitable lasers for photonic spike processing," Advances in Optics and Photonics **8**, 228 (2016).
15. A. N. Tait, A. X. Wu, T. F. de Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Microring weight banks," IEEE Journal of Selected Topics in Quantum Electronics **22**, 1–14 (2016).
16. A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic silicon photonics: Demonstration and benchmarking," Scientific Reports **arXiv:1611.02272** (2017).
17. R. W. Keyes, "Optical logic-in the light of computer technology," Optica Acta: International Journal of Optics **32**, 525–535 (1985).
18. P. R. Prucnal, B. J. Shastri, A. N. Tait, M. A. Nahmias, and T. Ferreira de Lima, *Neuromorphic Photonics* (CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2017).
19. T. Ferreira de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Progress in neuromorphic photonics," Nanophotonics **10.1515/nanoph-2016-0139** (2017).
20. D. Liang, G. Roelkens, R. Baets, and J. E. Bowers, "Hybrid integrated platforms for silicon photonics," Materials **3**, 1782 (2010).
21. B. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. Arthur, P. Merolla, and K. Boahen, "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," Proceedings of the IEEE **102**, 699–716 (2014).
22. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," Science **345**, 668–673 (2014).
23. J. Schemmel, D. Briiderle, A. Griibl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural modeling," in "Proceedings of 2010 IEEE International Symposium on Circuits and Systems," (IEEE, 2010), pp. 1947–1950.
24. S. Furber, F. Galluppi, S. Temple, and L. Plana, "The spinnaker project," Proceedings of the IEEE **102**, 652–665 (2014).
25. D. A. B. Miller, "Are optical transistors the logical next step?" Nat Photon **4**, 3–5 (2010).
26. H. J. Caulfield and S. Dolev, "Why future supercomputing requires optics," Nat Photon **4**, 261–263 (2010).
27. R. S. Tucker, "The role of optics in computing," Nat Photon **4**, 405–405 (2010).