

Brain-Derived 3D NanoPhotonic-NanoElectronic Neuromorphic Computing

S. J. Ben Yoo
Department of Electrical and Computer Engineering
University of California
Davis, California, USA
sbyoo@ucdavis.edu

Abstract—We propose a brain-derived—rather than a brain-inspired—Neuromorphic Computing architecture for flexible learning intelligent systems capable of handling complex tasks in unpredictable environments. We will discuss 3D Nanophotonic-Nanoelectronic integrated circuits that can realize energy-efficient, high-throughput, and scalable realization of brain-derived Neuromorphic Computing.

Keywords—neuromorphic computing, 3D integration, photonic integrated circuits, electronic integration, neuroscience.

I. INTRODUCTION

For many decades, there have been world-wide efforts to design and realize a brain-like flexible learning system of similar capability, comparable power consumption, and compact size as the human brain. While the software and algorithms for machine learning (ML) and artificial intelligence (AI) have advanced remarkably, the actual ML and AI hardware systems lagged significantly compared to the brain in terms of its flexible learning capability and its size, weight, and power. Current approaches in ML and AI are based on computational architectures that are loosely inspired by the brain. While the AI agents have been able to outperform humans at complex games like “Go” [1] and “Chess” when trained on a relatively narrow range of tasks, they often require a considerable degree of training, and they are brittle in the face of unexpected changes to task characteristics. The human brain, in contrast, is capable of remarkably fast learning in a manner that is flexible and enables generalization to new situations and tasks, and it does so with a remarkably low level of energy consumption relative to traditional computational hardware. For example, the natural-language model, Generative Pre-trained Transformer 3 (GPT-3), consists of 175 billion parameters and requires 9.8E29 FLOP for a single training on Tesla V100 GPUs [2], corresponding to ~\$10M for the energy cost alone, far surpassing the estimated energy consumption for a human brain. The popularity of the deep learning AI systems is currently driving AI-related energy-consumptions to double every 3.4 months [3], [4]. On the other hand, neuromorphic computing hardware efforts (e.g., IBM’s TrueNorth [2], Intel’s Loihi [3], University of Manchester’s SPINNAKER) have not been able to reproduce the uniquely flexible and adaptive nature of human intelligence, and the scalability, throughput, and energy-efficiency have also been lagging.

II. FUNDAMENTAL SCIENTIFIC AND TECHNOLOGICAL GAPS IN THE CURRENT BRAIN-INSPIRED NEUROMORPHIC APPROACHES

The recently rising consensus is that this disparity is due to the fundamental scientific and technological gaps listed [5] below:

Table 1. The four fundamental scientific and technological gaps between the biological and the typical neuromorphic computing systems.

Gap 1: Methods to realize bio-realistic algorithms and models. The algorithms and models used in biological neural networks are likely why the brain is so efficient in performing many tasks such as learning, which cannot be efficiently realized with high fidelity using CMOS hardware.
Gap 2: Next generation bio-inspired materials with neuromorphic dynamics. CMOS technologies were created for logic operation and arithmetic computation, and thus their dynamics are not well matched with the dynamics that are critical for neuromorphic computing. The dynamics of neurons, synapses and dendrites come from ion diffusion, which is not present in CMOS devices.
Gap 3: Innovative circuits to serve as required computing elements. Different from the traditional linear circuits built with CMOS devices, bio-realistic circuit designs should be highly nonlinear and rich in dynamics, which present great challenges in unconventional circuit designs. There is a lack of research to take advantage of the intrinsic nonlinear dynamics in designing novel circuit elements to efficiently and faithfully emulate synapses, neurons, and dendrites.
Gap 4: Scalable computing architecture. Even with more capable and suitable devices and circuits, there is a lack of research and existing knowledge on how to construct a scalable architecture for real-world problems. It requires co-designs and co-optimizations across different system abstract layers to incorporate the desired properties of the new materials and circuits to realize an efficient, reliable, and scalable computing system.

III. NEW APPROACHES: BRAIN-DERIVED NEUROMORPHIC COMPUTING BY 3D NANOPHOTONIC-ELECTRONIC INTEGRATED CIRCUITS

We propose to pursue Brain-Derived rather than Brain-Inspired neuromorphic computing that addresses the four Gaps of Table 1 and exploiting the new direction depicted in Figure 1. The new 3D Nanoscale Photonic-Electronic neuromorphic computing pursues new material, device, circuit, and system capabilities of co-designed 3D photonic and electronic integrated circuits (3D EPICs) designed for hierarchical learning. As Figure 2 illustrates, the proposed 3D nanoscale photonic-electronic integrated circuits for

hierarchical neuromorphic computing consisting of photonic neuromorphic computing circuits and electronic/ionic neuromorphic integrated circuits [6], [7].

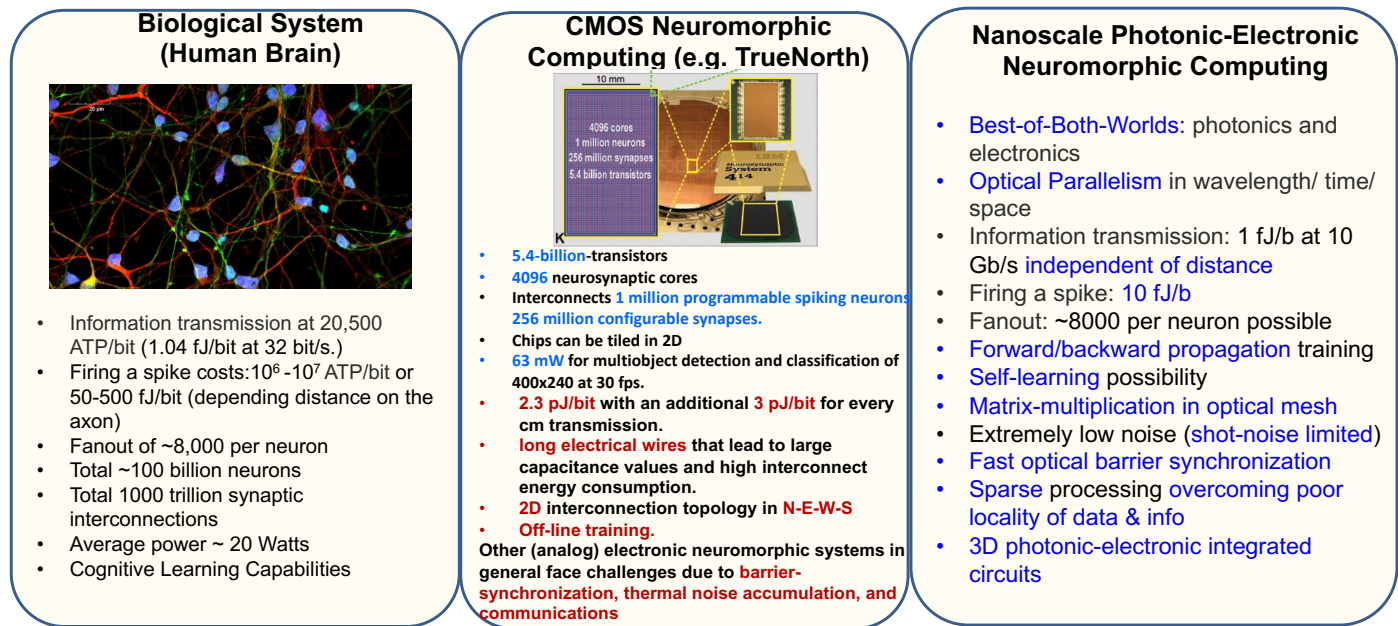


Figure 1. comparisons between the biological system (human brain), the CMOS Neuromorphic Computing (e.g. TrueNorth), and the proposed 3D Nanoscale photonic-electronic neuromorphic computing.

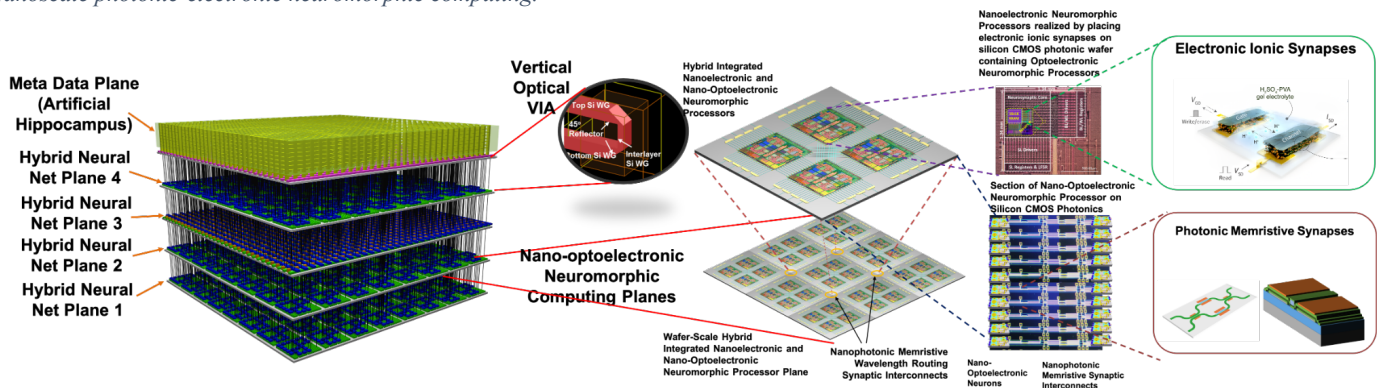


Figure 2. The proposed 3D nanoscale photonic-electronic integrated circuits for hierarchical neuromorphic computing consisting of photonic neuromorphic computing circuits and electronic/ionic neuromorphic integrated circuits.

ACKNOWLEDGMENT

The author is indebted to the ExPlor team members consisting of Profs. D. Cox, R. Chaudhuri, R.C O'Reilly of UC Davis; Profs. H.-S. P. Wong, A. Salleo, D.A. B. Miller of Stanford University; Prof. K. Bouchard of UC Berkeley; Prof. V. Sorger of George Washington University, and Profs. L. C. Kimerling and J.J. Hu of MIT.

REFERENCES

- [1] D. Silver *et al.*, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017, doi: 10.1038/nature24270 <http://www.nature.com/nature/journal/v550/n7676/abs/nature24270.html#supplementary-information>.
- [2] M. Taddeo, A. Tsamados, J. Cowls, and L. Floridi, "Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations," *One Earth*, vol. 4, no. 6, 2021, doi: 10.1016/j.oneear.2021.05.018.
- [3] D. Amodei, D. Hernandez, G. Sastry, J. Clark, G. Brockman, and I. Sutskever, "AI and Compute," *Open AI*, May 16, 2018.
- [4] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for modern deep learning research," in *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020. doi: 10.1609/aaai.v34i09.7123.
- [5] HAL S. GREENWALD, QING WU, and SABYASACHI GANGULI, "FOA-AFRL-AFOSR-2021-0005".
- [6] W. Wan *et al.*, "A Voltage-Mode Sensing Scheme with Differential-Row Weight Mapping for Energy-Efficient RRAM-Based In-Memory Computing," in *2020 IEEE Symposium on VLSI Technology*, 2020, pp. 1–2. doi: 10.1109/VLSITechnology18217.2020.9265066.
- [7] Y. van de Burgt *et al.*, "A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing," *Nat Mater*, vol. 16, no. 4, pp. 414–418, 2017, doi: 10.1038/nmat4856.