

Toward Scalable Energy Efficient Learning Machines

H.-S. Philip Wong

Department of Electrical Engineering and Stanford SystemX Alliance
Stanford University, Stanford, California 94305, USA

E-mail: hspwong@stanford.edu

The key elements of a scalable, fast, and energy-efficient computation platform for learning machines are: massive on-chip memory co-located with highly energy-efficient computation, enabled by monolithic 3D integration using ultra-dense and fine-grained massive connectivity. There will be multiple layers of analog and digital memories [1] interleaved with computing logic, sensors, and application-specific devices. We call this technology platform N3XT – Nanoengineered Computing Systems Technology [2]. N3XT will support computing architectures that embrace sparsity, stochasticity, and device variability, including those that are neuromorphic and learning-based.

In this talk, I will give an overview of nanoscale memory and logic technologies for implementing N3XT. I will describe ENIGMA, a class of robust, energy-efficient learning machines using hyperdimensional computing [3], to be implemented [4] using the N3XT technology platform.

This work is a collaboration of UC Berkeley (Sayeef Salahuddin, Jan Rabaey, Pentti Kanerva, Bruno Olshausen) and Stanford University (Subhasish Mitra, H.-S. Philip Wong) in a project funded by the NSF (Award #1640060, https://nsf.gov/awardsearch/showAward?AWD_ID=1640060).

References:

- [1] H.-S. P. Wong and S. Salahuddin, “Memory Leads the Way to Better Computing,” invited paper, *Nature Nanotechnology*, Vol. 10, pp. 191 – 194 (2015)
- [2] M. Aly, “Energy-Efficient Abundant-Data Computing: The N3XT 1,000X,” *IEEE Computer*, pp. 24 – 33, December 2015.
- [3] A. Rahimi, P. Kanerva, J. M. Rabaey, “A Robust and Energy-Efficient Classifier Using Brain-Inspired Hyperdimensional Computing,” [ISLPED '16](#), Proceedings of the 2016 International Symposium on Low Power Electronics and Design, pp. 64 – 69 (2016).
- [4] H. Li, T.F. Wu, A. Rahimi, K.-S. Li, M. Rusch, C.-H. Lin, J.-L. Hsu, M.M. Sabry, S.B. Eryilmaz, J. Sohn, W.-C. Chiu, M.-C. Chen, T.-T. Wu, J.-M. Shieh, W.-K. Yeh, J. M. Rabaey, S. Mitra, and H.-S. P. Wong, “Hyperdimensional Computing with 3D VRRAM In-Memory Kernels: Device-Architecture Co-Design for Energy-Efficient, Error-Resilient Language Recognition,” IEEE International Electron Devices Meeting (IEDM), paper 16.6, December 5 – 7, San Francisco, 2016