SEMINAR ANNOUNCEMENT

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
Faculty of Engineering
Website: https://www.eng.nus.edu.sg/ece/

Area: Microelectronic Technologies and Devices

Host: Dr Evgeny Zamburg

TOPIC	:	**High Throughput, Area-Efficient, and Variation-Tolerant 3D In-memory Compute System for Deep Convolutional Neural Networks**
SPEAKER	:	**Ms Hasita Veluri** **Graduate student, ECE Dept, NUS**
DATE	:	**Friday, 27 November 2020**
TIME	:	**10.00AM to 11.00AM**
WEBINAR	:	**Join Zoom Meeting:** **https://nus-sg.zoom.us/j/86801139910?pwd=cGJ4THFJTE5ncWZsT1U4RzJEeWRnZz09** **Meeting Id: 868 0113 9910** **Password: 608872**

ABSTRACT

Untethered computing using Deep Convolutional Neural Networks at the edge of IoT with limited resources requires systems that are exceedingly power and area efficient. Analog in-memory matrix-matrix multiplications enabled by emerging memories can significantly reduce the energy budget of such systems and result in compact accelerators. In this work, we report a high-throughput RRAM-based DCNN processor that boasts 7.12× area-efficiency (AE) and 6.52× power-efficiency (PE) enhancements over state-of-the-art accelerators. We achieve this by coupling a novel in-memory computing methodology with a staggered-3D memristor array. Our variation-tolerant in-memory compute method, which performs operations on signed floating-point numbers within a single array, leverages charge domain operations and conductance discretization to reduce peripheral overheads. Voltage pulses applied at the staggered bottom electrodes of the 3D-array generate a concurrent input shift and parallelize convolution operations to boost throughput. The high density and low footprint of the 3D-array, along with the modified in-memory M2M execution, improve peak AE to 9.1TOPsmm-2 while the elimination of input regeneration improves PE to 10.6TOPsW-1. This work provides a path towards infallible RRAM-based hardware accelerators that are fast, low-power, and low-area.

BIOGRAPHY

Hasita Veluri is currently pursuing her Ph.D. in Electrical and Computer Engineering since 2018. Her research interests include developing hardware-aware neural network accelerators with advanced memories for power efficient systems to realize visions of IoT.

https://www.eng.nus.edu.sg/ece/highlights/events/