Reconfigurable Materials and Devices for Neuromorphic Computing Focus Topic
Room 203A - Session RM+EM+NS-TuA
IoT Session: Reconfigurable Materials and Devices for Neuromorphic Computing

Moderator: Brian Hoskins, National Institute of Standards and Technology (NIST)

2:20pm **RM+EM+NS-TuA1 Non-volatile Memories for Neuromorphic Computing**, ***Alec Talin,*** Sandia National Laboratories **INVITED**

Inspired by the efficiency of the brain, CMOS-based neural architectures and memristors are being developed for pattern recognition and machine learning. However, the volatility, design complexity and high supply voltages for CMOS architectures, and the stochastic and energy-costly switching of memristors complicate the path to achieve the interconnectivity, information density, and energy efficiency of the brain using either approach. In my talk, I will review the latest advances in neuromorphic computing architectures based on deep neural networks implemented using CMOS and memristors and describe the challenges in achieving both high accuracy and energy efficiency using these devices. I will then discuss an alternative approach based on the non-volatile redox memory (NVRM): a device with a resistance switching mechanism fundamentally different from existing memristors, involving the reversible, electrochemical reduction/oxidation of a material to tune its electronic conductivity. The first type of NVRM that I will describe is based upon the intercalation of Li-ion dopants into a channel of $Li_{1-x}CoO_2$. This Li-ion synaptic transistor for analog computing (LISTA) switches at low voltage (mVs) and energy, displays hundreds of distinct, non-volatile conductance states within a 1V range, and achieves high classification accuracy when implemented in neural network simulations[1]. The second type of NVRM I will describe operates on a similar principle but is based on the polymer system PEDOT:PSS, and which we call the electrochemical neuromorphic organic device (ENODe) [2]. Plastic ENODes are fabricated on flexible substrates enabling the integration of neuromorphic functionality in stretchable electronic systems. Mechanical flexibility makes ENODes compatible with three-dimensional architectures, opening a path towards extreme interconnectivity comparable to the human brain.

1) E. J. Fuller et al., Advanced Materials 29, 1604310 2017.

(2) Y. B. van de Burgt et al., Nature Materials 16, 414 2017.

(3) S. Agarwal et al., IEEE 2017 Symposium on VLSI Technology Digest of Technical Papers, DOI: 10.23919/VLSIT.2017.7998164.

4:20pm **RM+EM+NS-TuA7 Memristor Neural Networks for Brain-Inspired Computing**, ***Qiangfei Xia,*** University of Massachusetts Amherst **INVITED**

As CMOS scaling approaches its limits, it becomes more difficult to keep improving the speed-energy efficiency of traditional digital processors. To address this issue, computing systems augmented with emerging devices particularly memristors, offer an attractive solution. Memristors use conductance to represent analog or digital information. The dynamic nature of memristor with both long-term and short-term memories, together with its small effective size contributes to the energy efficiency in weight updating (training). The in-memory computing scheme in a crossbar breaks the 'von Neumann bottleneck' as the weights are stored locally in each device during computing. The read out (inference) is finished in one clock cycle regardless of the array size, offering massive parallelism and hence high throughput. The capability of using physical laws for computing in a crossbar enables direct interfacing with analog signals from sensors without energy- hungry analog/digital conversions.

We developed a Ta/hafnium oxide memristor with stable multilevel resistance, linear current voltage characteristics in chosen conductance ranges, in addition to high endurance and long retention. We further integrated the memristors with foundry-made transistors into large arrays. We demonstrated that the reconfigurable memristor networks are capable of analog vector matrix multiplication, and successfully implemented a number of important applications including signal processing, image compression and convolutional filtering. We also built a multilayer memristor neural network, with which we demonstrated in-situ and self-adaptive learning capability with the MNIST handwritten digit dataset. The successful demonstration of analog computing and in-situ online training suggests that the memristor neural network is a promising hardware technology for future computing.

5:00pm **RM+EM+NS-TuA9 Indium Phosphide Synaptic Device on Silicon for Scalable Neuromorphic Computing**, ***Jun Tao,*** *D. Sarkar, R. Kapadia,* University of Southern California

Inspired by the superior capability of the brain, neuronal spiking, and synaptic behavior have been mimicked by the CMOS-based neuronal cell in hardware, which contains 6-12 transistors depending on specific functionality and the robustness of the design. However, the higher energy consumption and physical area have led researchers to look for architectures based on single device and novel materials.

In our work, utilizing thin-film vapor-liquid-solid growth method, we fabricated scalable Indium phosphide (InP) channel transistors directly on Si/SiO_2 wafer, which can emulate significant synaptic characteristics such as elasticity, short- and long-term plasticity, metaplasticity, spike number dependent plasticity and spike timing dependent plasticity, by modeling gate electrode as the pre-synaptic axon terminal, the drain electrode as the post-synaptic dendrite, and the gate oxide-semiconductor channel as the synapse junction, in which we also interpreted the FET channel conductance as the synaptic weight.

We also demonstrated that by controlling the charging and discharging of interfacial traps at the gate oxide-semiconductor stack, we can essentially engineer hysteresis of the synaptic device to customize the synapse behavior and modify the synapse weight non-linearly. It underpins optimal selectivity of signal transduction and satisfies the key neuromorphic architecture characteristic—training and learn. Tuning hysteresis in a family of transfer characteristics in spike timing dependent plasticity (STDP) emulation, we attain maximum potentiation (depression) for the minimum positive (negative) interval time, which gradually decays down to elasticity, as we expected, indicating the scalable InP channel transistors on silicon as promising devices and platform for neuromorphic computation.

5:20pm **RM+EM+NS-TuA10 Ultra-low Power Microwave Oscillators based on Phase Change Oxides as Solid-State Neurons**, ***Boyang Zhao,*** *J. Ravichandran,* University of Southern California

Voltage or current controlled oscillators are well-established candidates for solid-state implementations of neurons. Metal to insulator transition (MIT) based phase change electrical oscillators are one of the many candidates for solid-state neurons, but current implementations are far from the ideal performance limits of energy and time necessary to induce the transition. We propose the use of nanoscale, epitaxial heterostructures of phase change oxides such as VO_2, NbO_2 and oxides with metallic conductivity as a fundamental unit of a low power electrical oscillator, capable of operating as neurons for neuromorphic computing architectures. Our simulations such that such oscillators can operate in the microwave regime and overcome many of the power consumption issues plagued by phase change electrical oscillators.

5:40pm **RM+EM+NS-TuA11 Leveraging Nanodevice Volatility for Low Energy Computing Inspired from Nature**, ***Alice Mizrahi,*** NIST/University of Maryland*; T. Hirtzlin,* Centre de Nanosciences et Nanotechnologies*; B. Hoskins,* NIST Center for Nanoscale Science and Technology*; A. Fukushima,* AIST*; A. Madhavan,* NIST Center for Nanoscale Science and Technology*; H. Kubota, S. Yuasa,* AIST*; N.B. Zhitenev, J. McClelland, M.D. Stiles,* NIST Center for Nanoscale Science and Technology*; D. Querlioz,* Centre de Nanosciences et Nanotechnologies, France*; J. Grollier,* UMR CNRS/Thales **INVITED**

Artificial neural networks are performing tasks, such as image recognition and classification, that were thought only accessible to the brain. However, these algorithms run on traditional computers and consume orders of magnitude more energy more than the brain does at the same task. One promising path to reduce the energy consumption is to build dedicated hardware to perform cognitive tasks. Nanodevices are particularly interesting because they allow for complex functionality with low energy consumption and small size. I discuss two nanodevices. First, I focus on stochastic magnetic tunnel junctions, which can emulate the spike trains emitted by neurons with a switching rate that can be controlled by an input. Networks of these tunnel junctions can be combined with CMOS circuitry to implement population coding to build low power computing systems capable of processing sensory input and controlling output behavior. Second, I turn to different nanodevices, memristors, to implement a different type of computation occurring in nature: swarm intelligence. A broad class of algorithms inspired by the behavior of swarms have been proven successful at solving optimization problems (for example an ant colony can solve a maze). Networks of memristors combined with CMOS circuitry can perform swarm intelligence and find the shortest paths in mazes. These results are striking illustrations of how matching the functionalities of nanodevices with relevant properties of natural systems

open the way to low power hardware implementations of difficult computing problems.

Author Index

Bold page numbers indicate presenter