Ultra-Low Power Neuromorphic Computing With Spin-Torque Devices

Mrigank Sharad, Deliang Fan, Karthik Yogendra

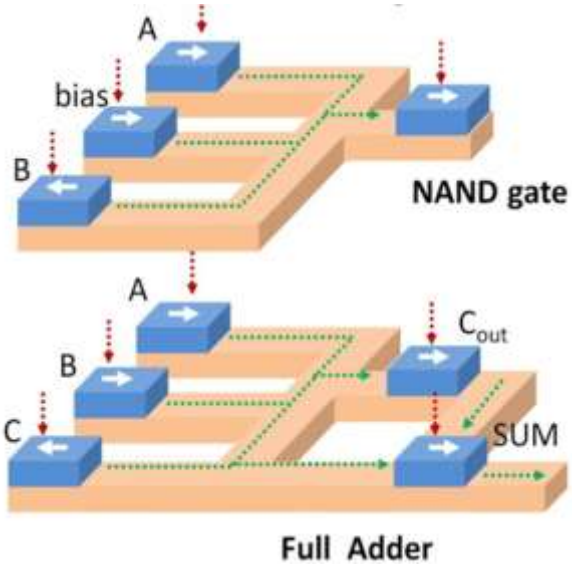
Prof. Kaushik Roy

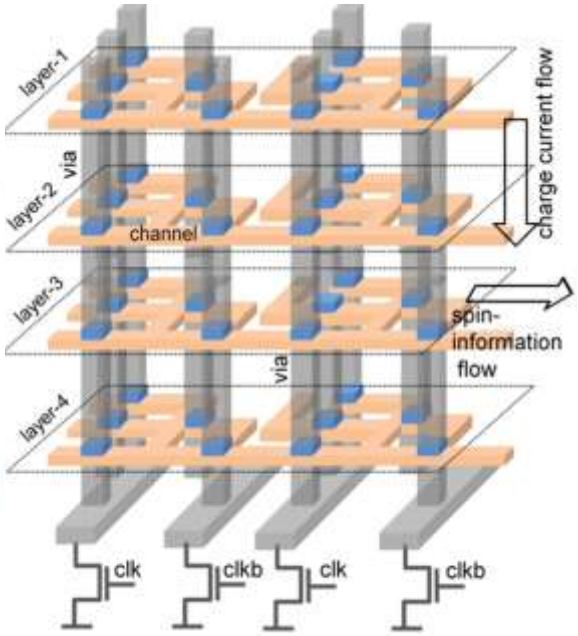
School of Electrical and Computer Engineering

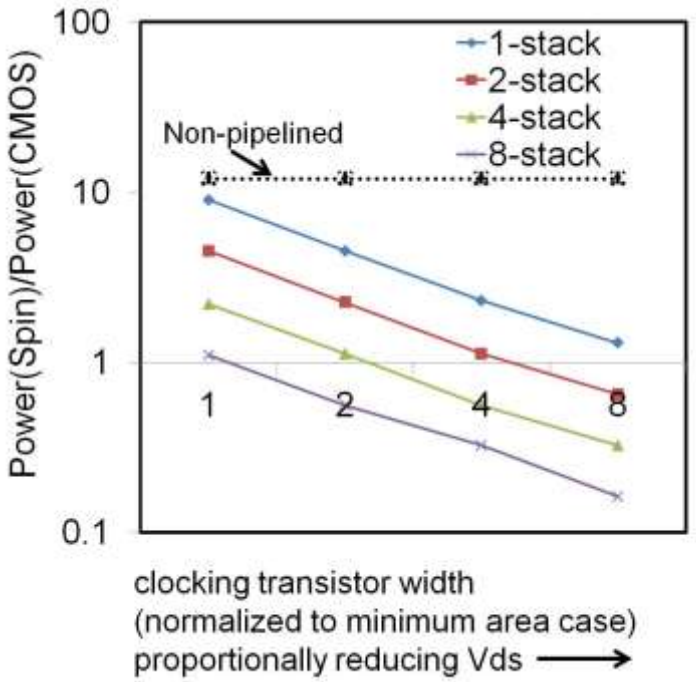
Purdue University

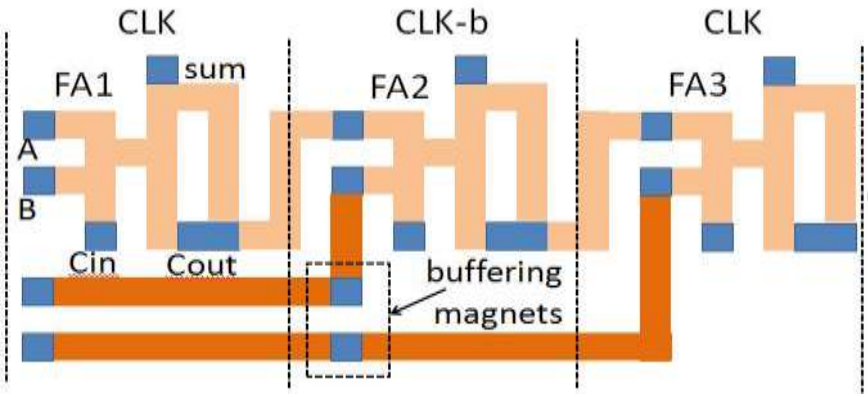
https://engineering.purdue.edu/NRL/index.html

Boolean Computation using Spin Torque Device





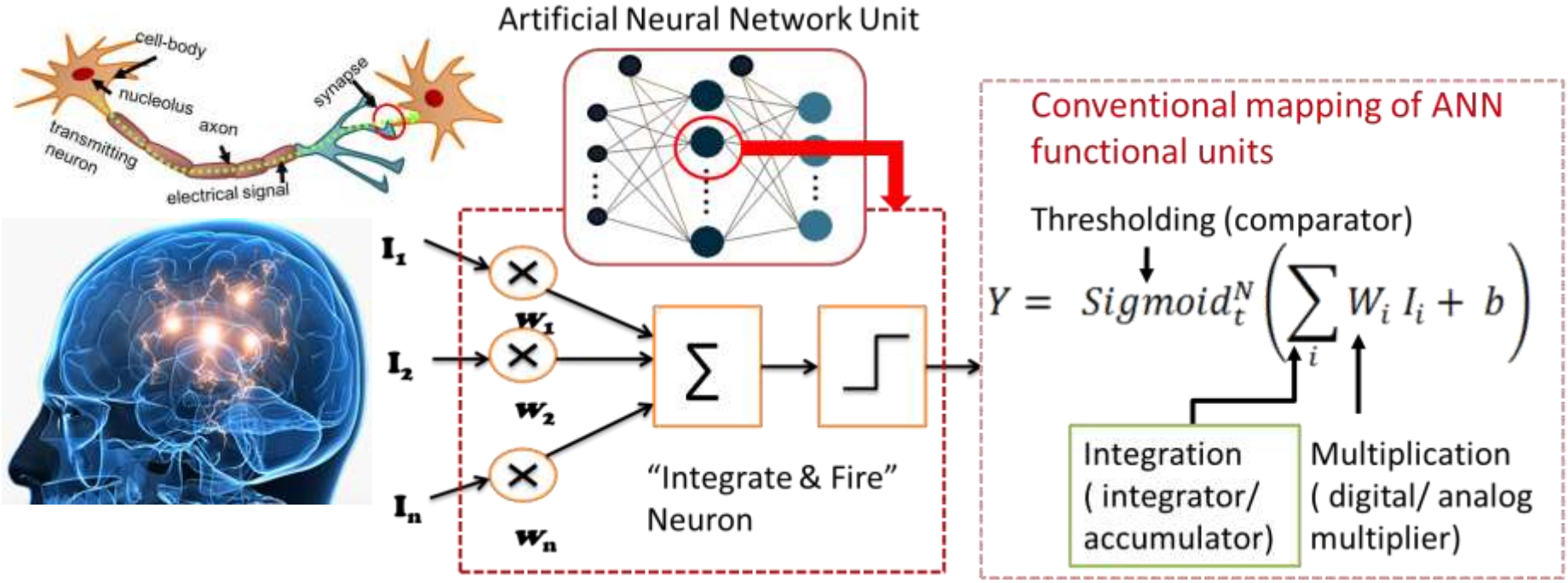




- Three orders of magnitude higher density as compared to 15 nm CMOS design
- By proper pipelining, 3D stacking and clocking transistor sizing, we can get comparable power consumption and performance to state of art CMOS technology

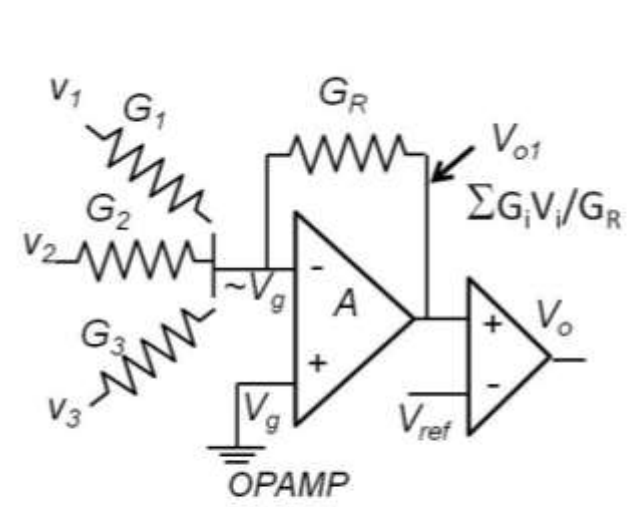
Non-Boolean Computing

- Traditional computing models (Boolean logic, von Neumann architectures) are highly inefficient at performing tasks that humans routinely perform, such as visual recognition, semantic analysis, and reasoning.
- Bio-inspired computation can outperform Von-Neumann designs in many such data Processing applications

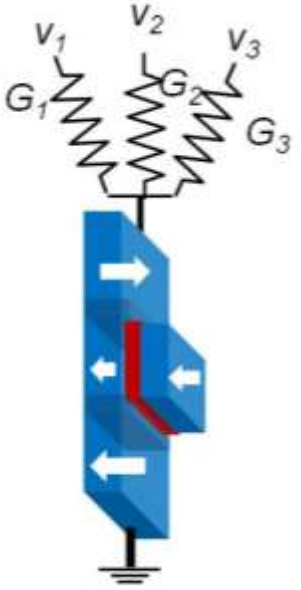


Non-Boolean Computing with STT Devices

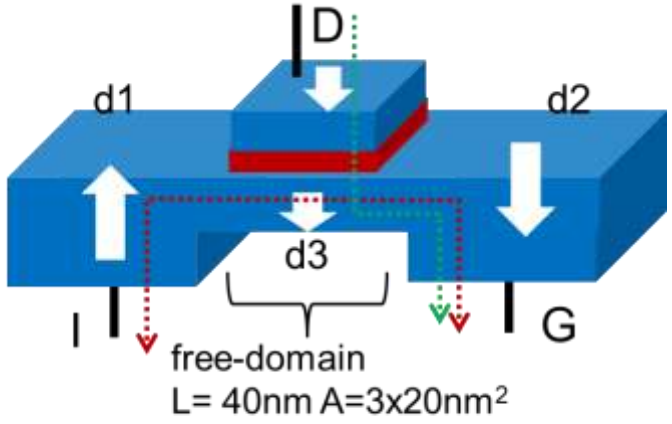
CMOS circuits for artificial neurons can be too complex and power hungry for designing large scale non-Boolean/neuromorphic hardware



analog CMOS circuit for artificial neuron



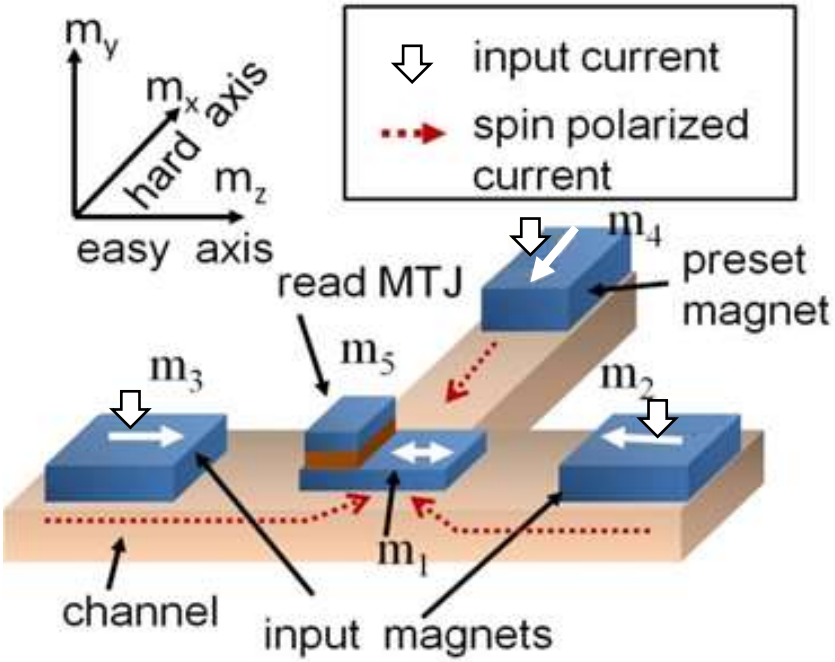
compact, energy-efficient spin-neuron



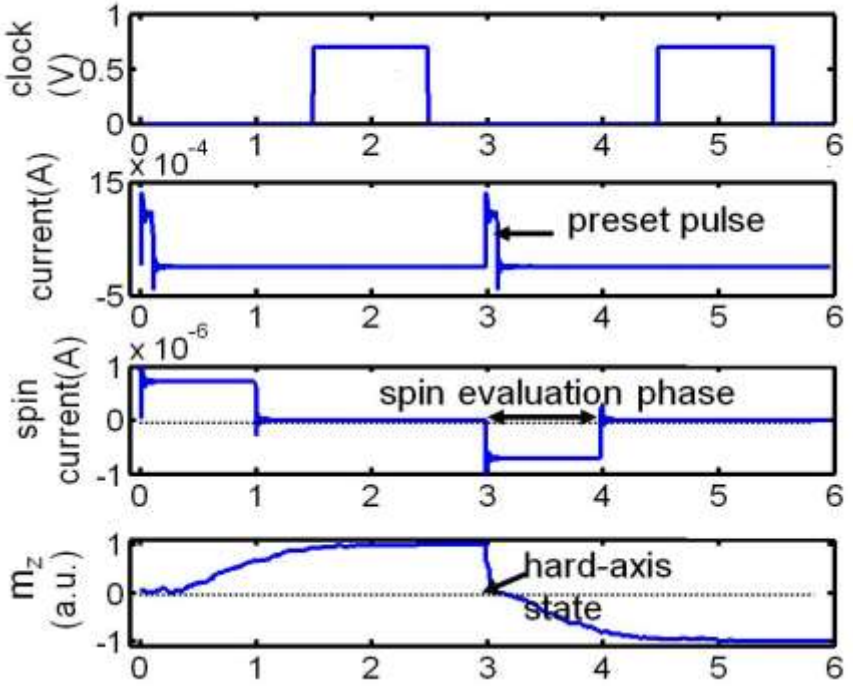
- Designed ultra-low voltage and high-speed current-mode switches using spin-torque devices that can mimic 'neuron'
- Developed device-models for spin-neurons and explored their application in the design of ultra-low power neuromorphic circuits and architectures

Bipolar Spin Neuron

Device Structure

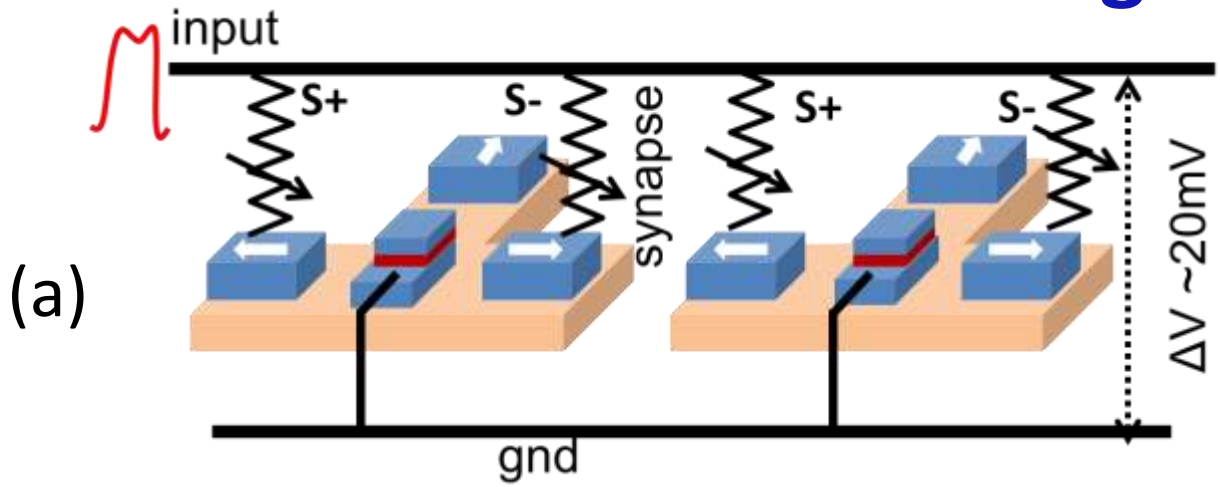


Device operation

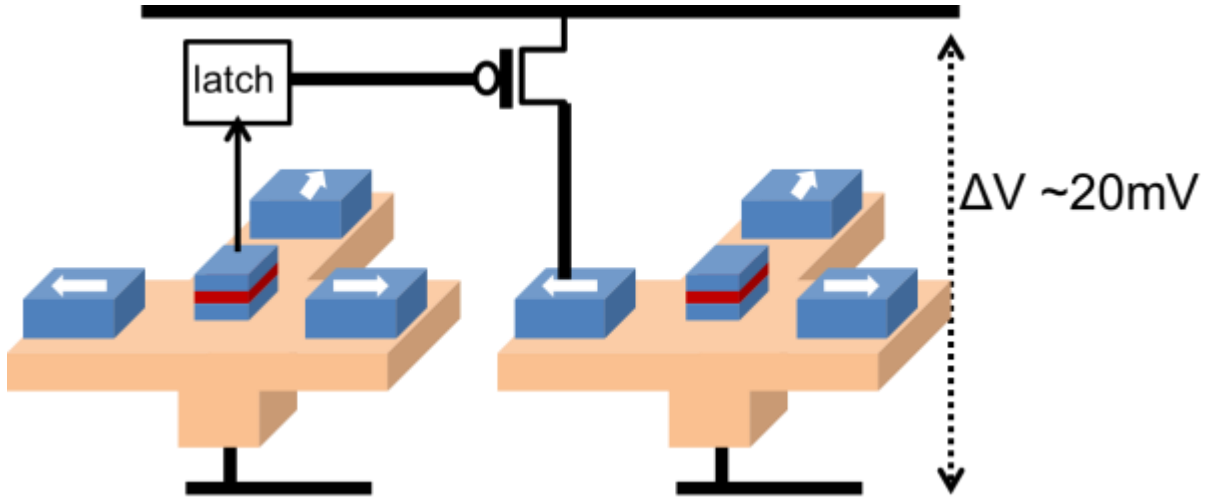


- The neuron device essentially acts as an ultra low voltage current comparator and can be employed to perform analog-mode computation

Communication Through Synapses



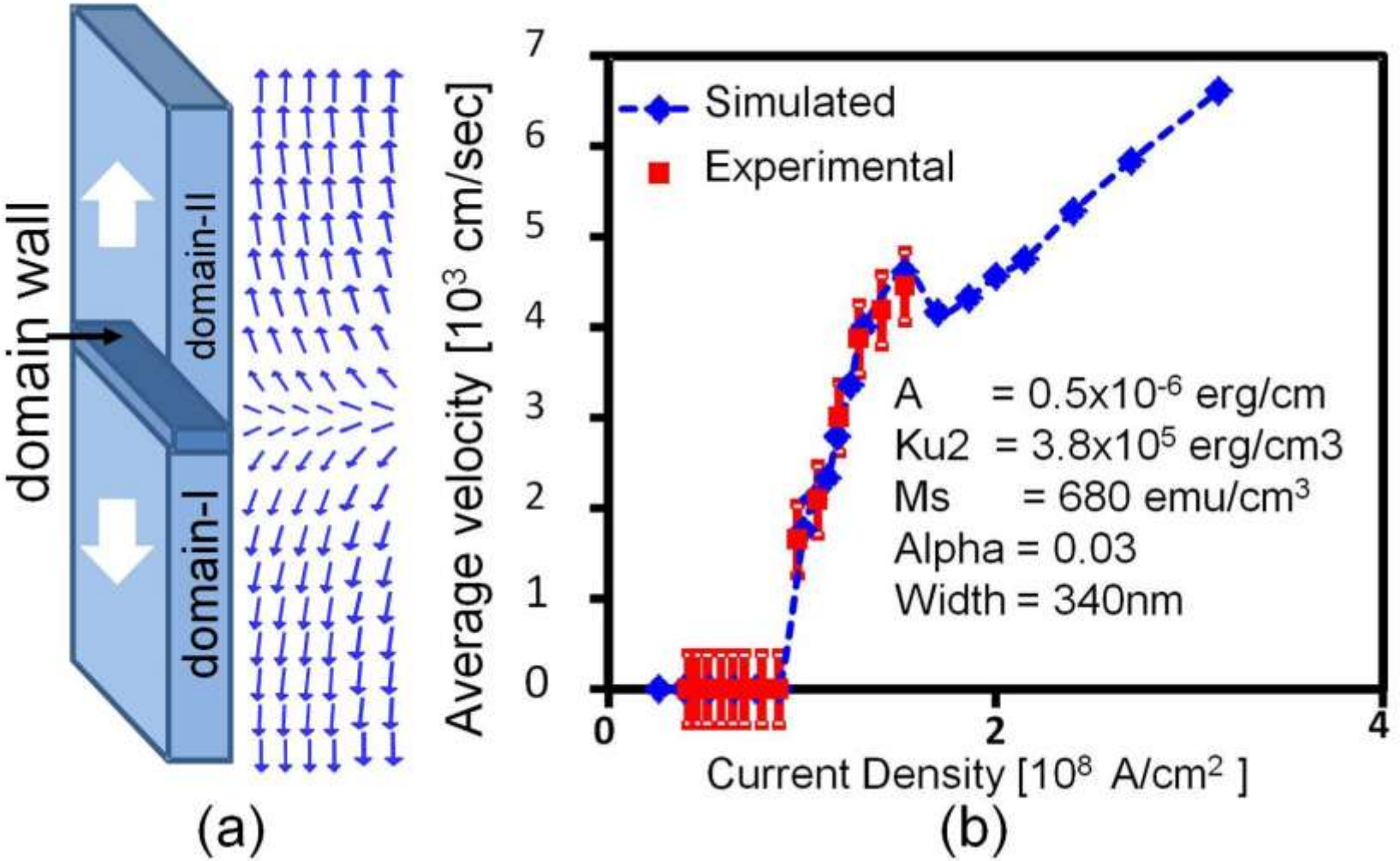
(a)



(b)

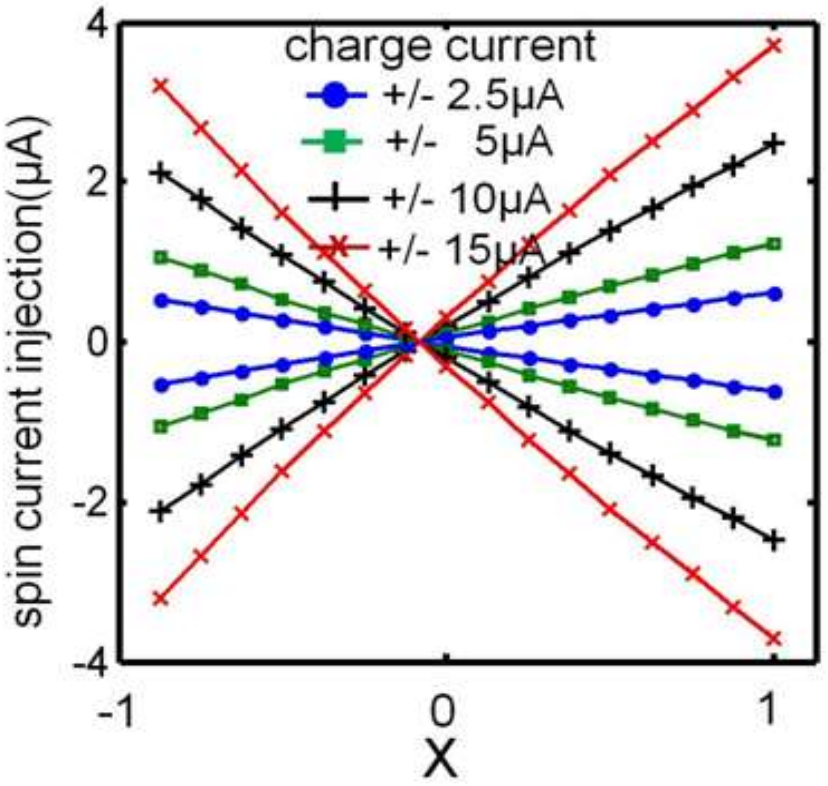
- Memristors/DWM/DTCS can be used for realizing low power neuromorphic computation array using bipolar spin neuron
- The magneto-metallic neurons facilitate input voltage levels of ~20mV resulting in low computation power

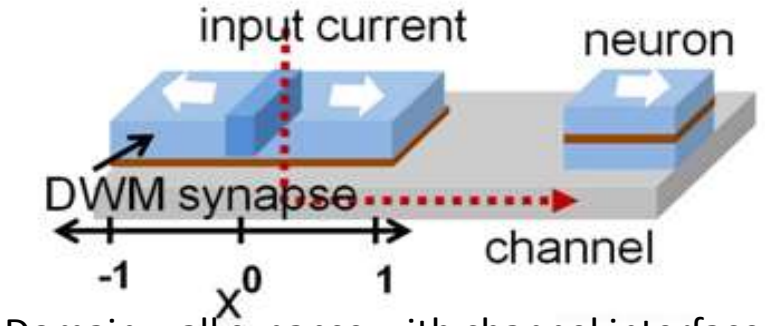
Example Synapse: Domain Wall Magnet



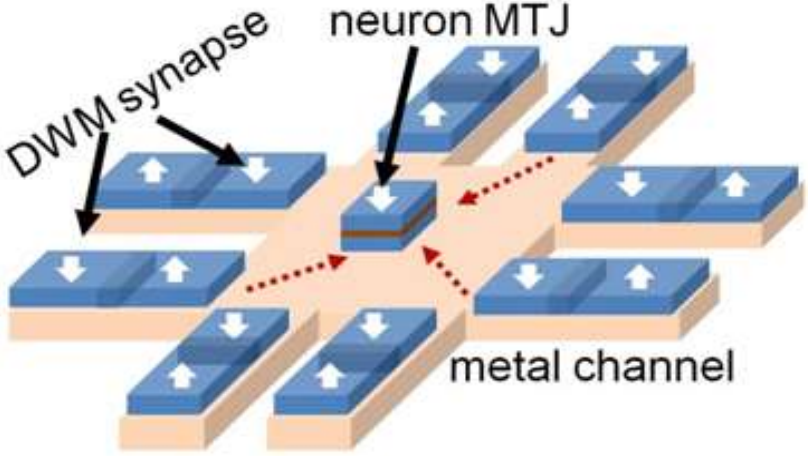
- A DWM consists of opposite polarity domains separated by a Non-magnetic region call the domain wall which can be moved by charge injection/ magnetic field

“Spin Neuron” with Domain Wall Magnets as Synapses



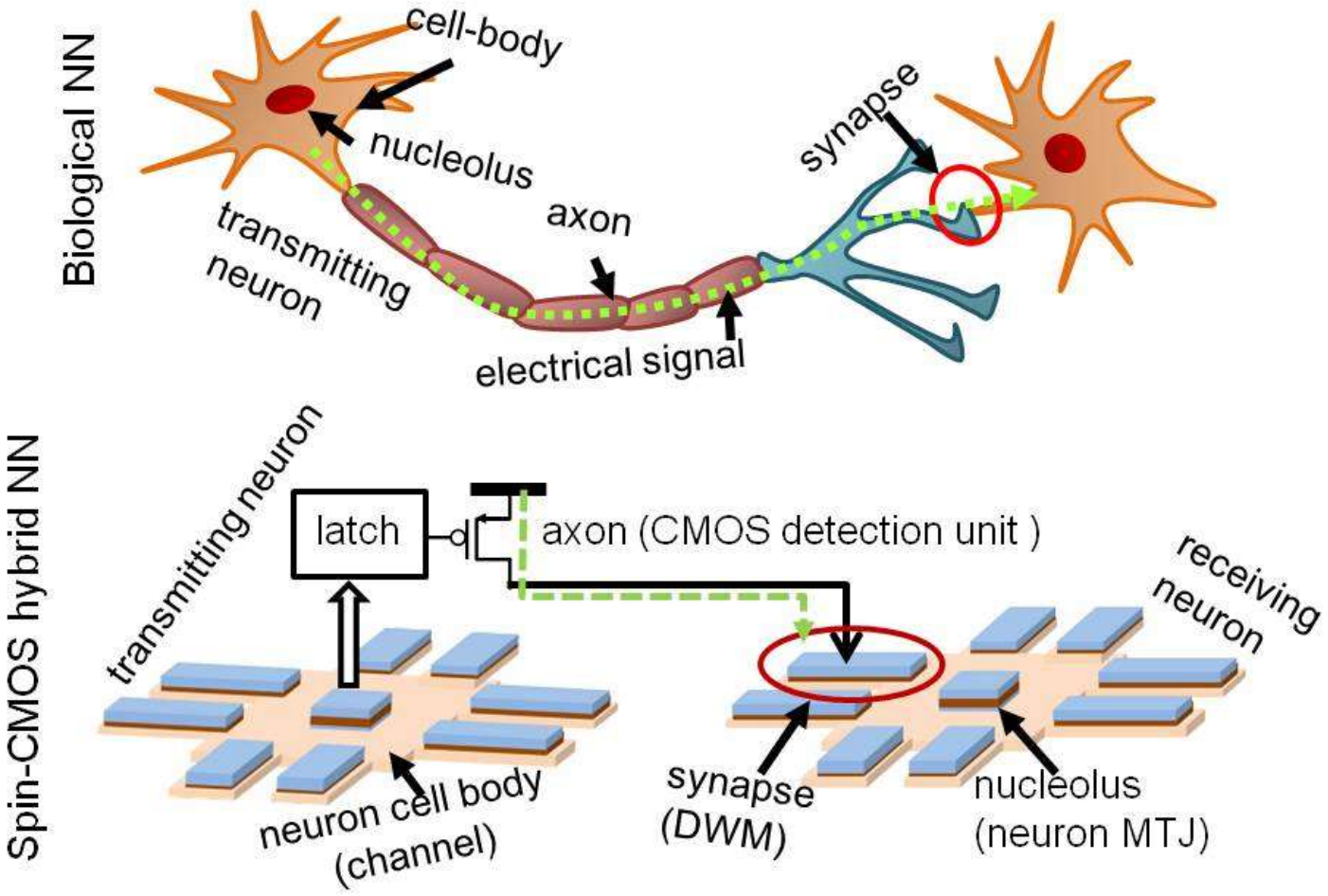


Domain wall synapse with channel interface

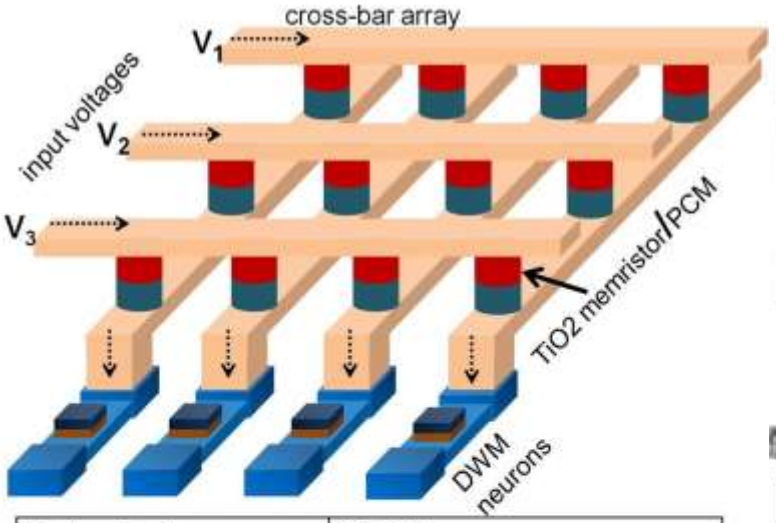


- Neuron with small number of programmable DWM inputs can be employed to realize configurable data processing array of cellular neurons

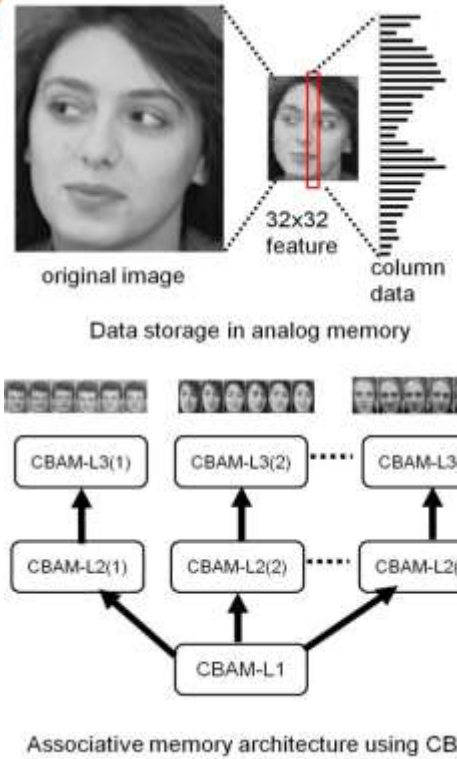
Drawing Analogy with Biological Neural Network



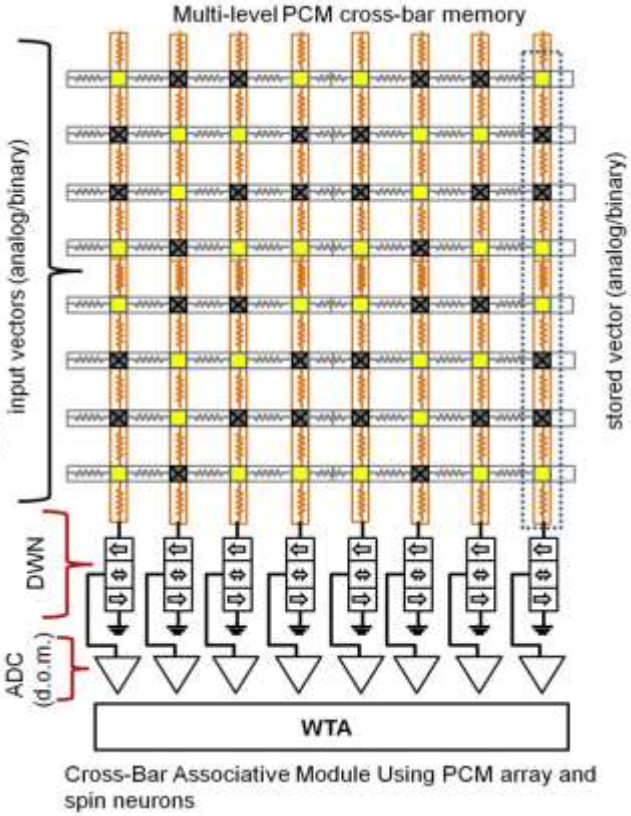
Spin Based Neuron for Cross-Bar Neural Network



Network size	32x64	
Weight levels	16 (4 bits)	
PCM conductance	5K to 100K	
frequency	500MHz	
32 element analog associative memory using PCM cross-bar	Spin neuron:	0.06mW
	CMOS neuron	9mW



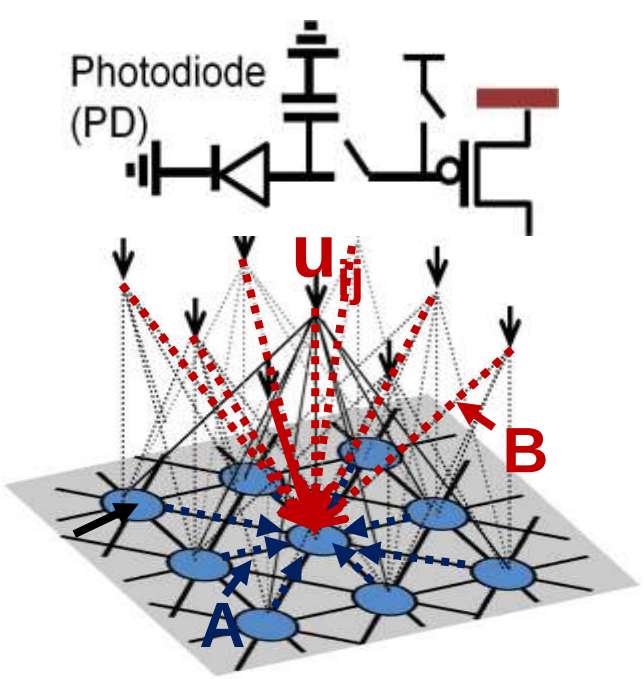
Associative memory architecture using CBAM

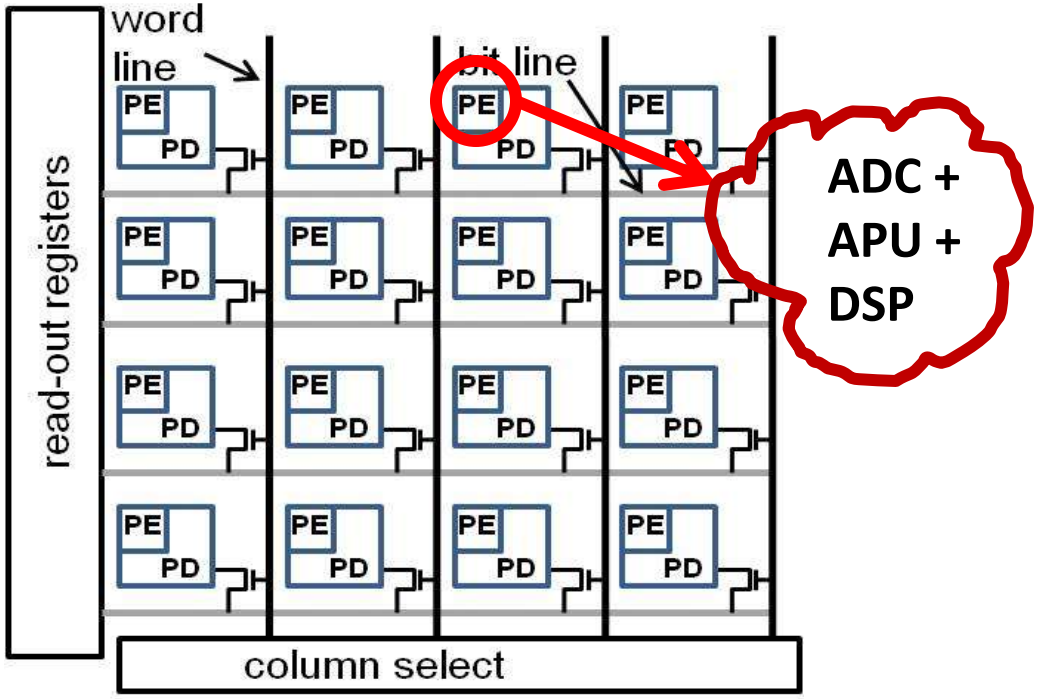


Cross-Bar Associative Module Using PCM array and spin neurons

- The spin based neuron unit achieves ~100X improvement in power consumption for cross bar ANN architecture Based on memristor/PCM

Ultra Low Energy Analog Signal Acquisition and Processing





- Hardware based on Cellular neural networks can be employed in several image processing applications
- Each pixel in a CMOS sensor array may contain analog processing units along with DSP
- Analog units lead to large power consumption even for simple image processing applications
- We employ spin-CMOS hybrid PE to achieve ultra low energy analog computation

Results for Image Processing using CNN

Results show the possibility of more than ~100x energy-improvement over state-of the art CMOS for neuromorphic computing:

Table-I CMOS vs. spin based feature extraction IC

Ref	CMOS Tech	E(CMOS) /E(spin)
[23]	0.35μ	253
[24]	0.35μ	560
[25]	0.25μ	470

Table-II CMOS vs. spin based ADC

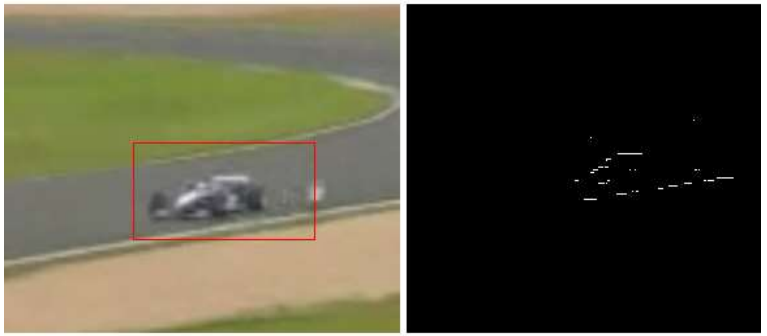
Ref	CMOS Tech	E(CMOS) /E(spin)
[26]	0.18μ	133
[27]	0.90μ	70
[28]	0.90μ	72

$E = (\text{total power}) / (S^2 \times \#\text{Pixels} \times \text{Fps})$

S : technology ratio

fps : frames/sec

Halftone sensing and compression



- Low voltage, compact and fast switching spin-neurons can provide the essential neuron functionality, leading to ultra-low energy and high-density neuromorphic/non-Boolean computing systems.

Summary

- We explored the possibility of non-Boolean computing using Spin Torque Devices
- We noted that current-mode switching of STT-devices can be employed in ultra low power analog computing
- Spin-'neurons' can lead to ultra low power non-Boolean architectures based on resistive memory , due to ultra low voltage, low current operation. Such design can be applied to numerous applications like associative computing , programmable threshold logic and neuromorphic hardware design.
- We plan to explore the application of STT devices in global interconnect design
- Design challenges related to precise voltage supply generation and distribution will be explored.

Funded by: DARPA, MARCO, StarNet, SRC and NSF