BOISE STATE UNIVERSITY

Addressing Challenges in Neuromorphic Computing with Memristive Synapses

Vishal Saxena[1], Xinyu Wu[1] and Maria Mitkova[2]

[1]Analog Mixed-Signal and Photonic IC (AMPIC) Lab

[2]Nanoionic Materials and Devices Lab

ECE Department, Boise State University, ID

http://lumerink.com

Agenda

- Introduction
- Synapse Devices and Challenges
- Mixed-Signal Neural Learning Architecture
- Realizing Multibit Synapses
- Conclusion

Reinventing Computing

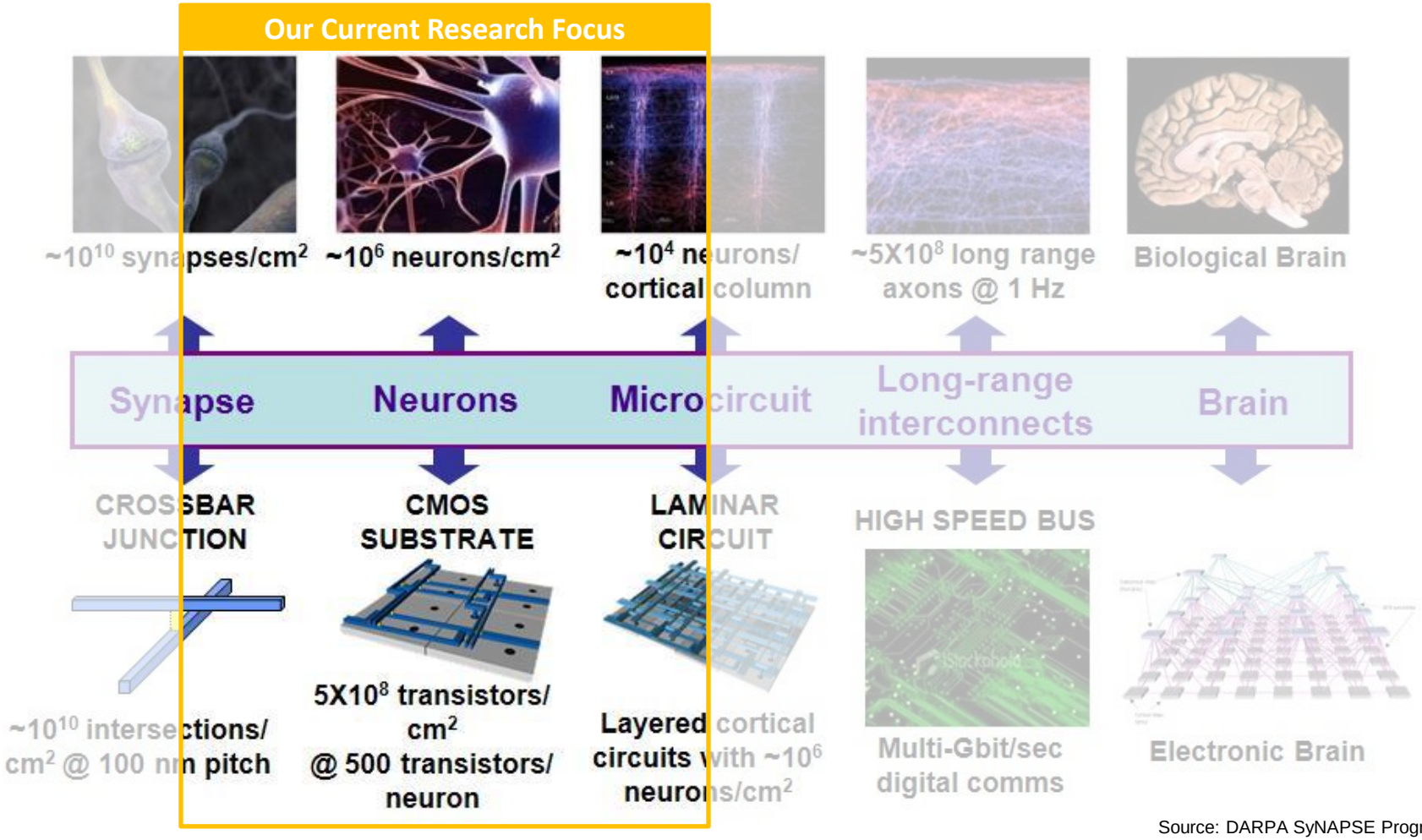
- Von Neumann computing in the nano-CMOS regime
  - + Massive transistor count, but….
  - – Device variability
  - – Dark silicon issue
  - – Not well suited for processing massively parallel data streams and pattern recognition tasks that come easy to a brain.
- US Office of Science and Technology (OSTP) RFI, June 2015:

"Create a new type of computer that can proactively interpret and learn from data, solve unfamiliar problems using what it has learned, and operate with the energy efficiency of the human brain"

DOE, Brain Initiative, AFOSR, DARPA, SRC

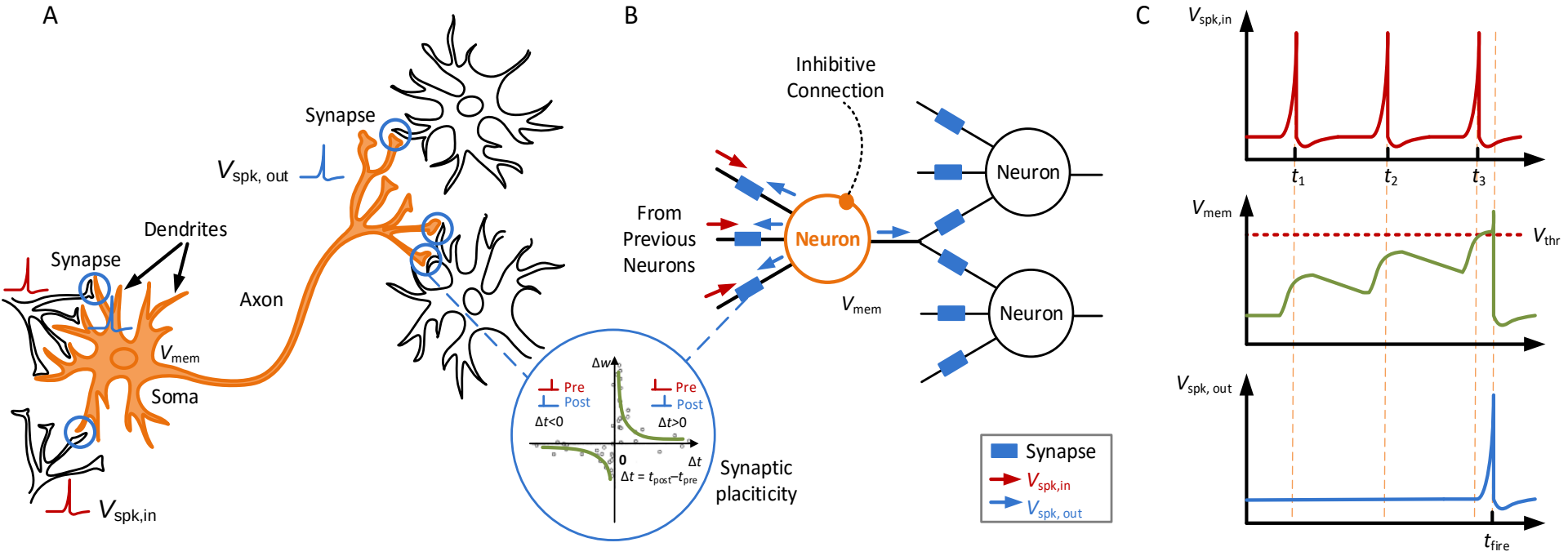
IEEE rebooting COMPUTING

Brain-Inspired Neuromorphic System



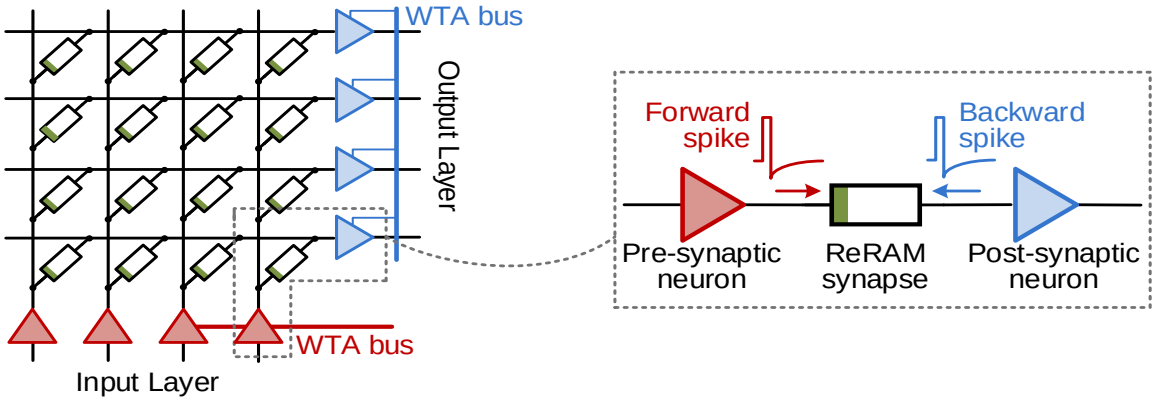
Source: DARPA SyNAPSE Program

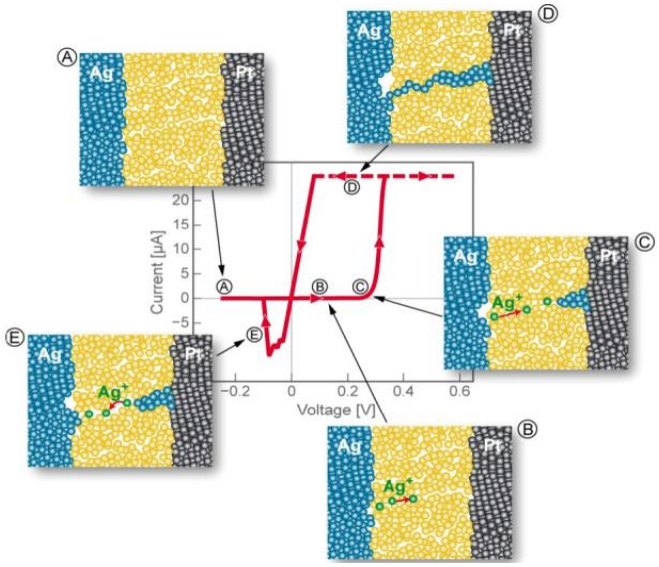
Revisiting Biological Inspiration



- Emerging understanding of neural computation from Computational Neuroscience
  - Hierarchical spiking neural networks with localized learning
- Synapses strengthen or weaken by relative firing times of the pre- and post-neurons
  - Plasticity rules: Spike-Timing Dependent Plasticity (STDP)

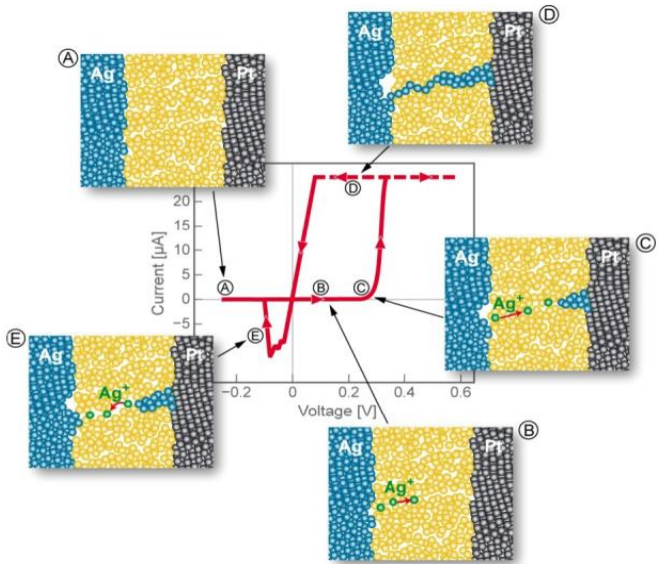
Neuromorphic Computing with RRAMs



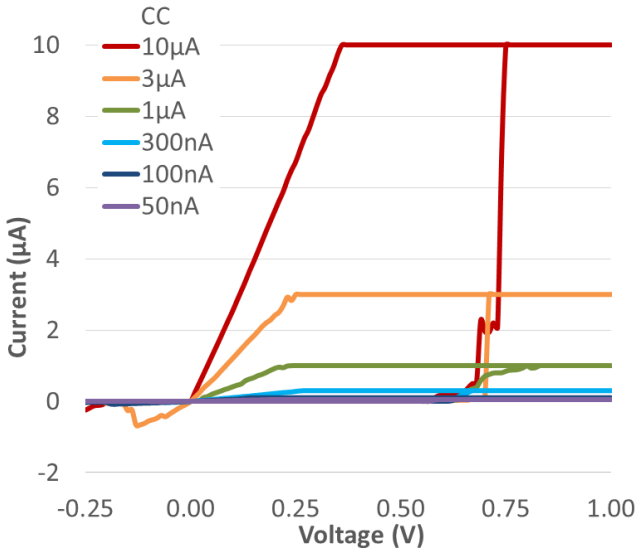


- To achieve biology-like synapse density, emerging nanoscale devices are under consideration
  - Resistive RAM (RRAM) or Memristors
  - Chalcogenide-based Conductive Bridge RAM (CBRAM) devices from Mitkova Group at Boise State
- Three necessary criterions for the realization of neuromorphic hardware capable of deep learning: **(1) *non-volatility of the trained weights,* (2) *massive synaptic density, and* (3) *ultra-low-power operation.***

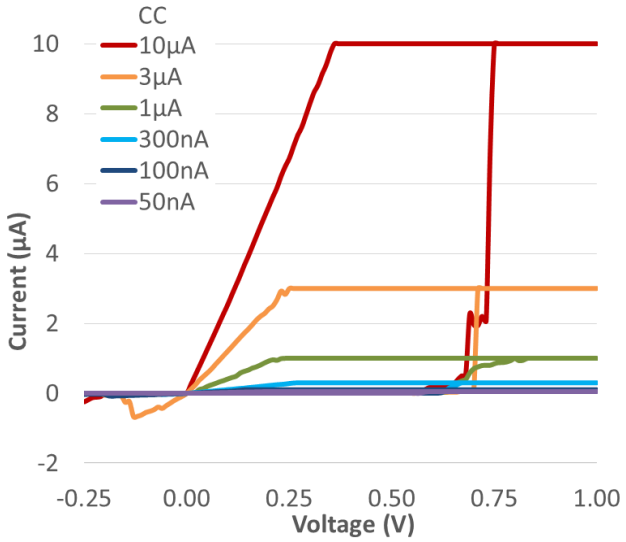
CBRAM I-V Characteristics



I-V Sweep

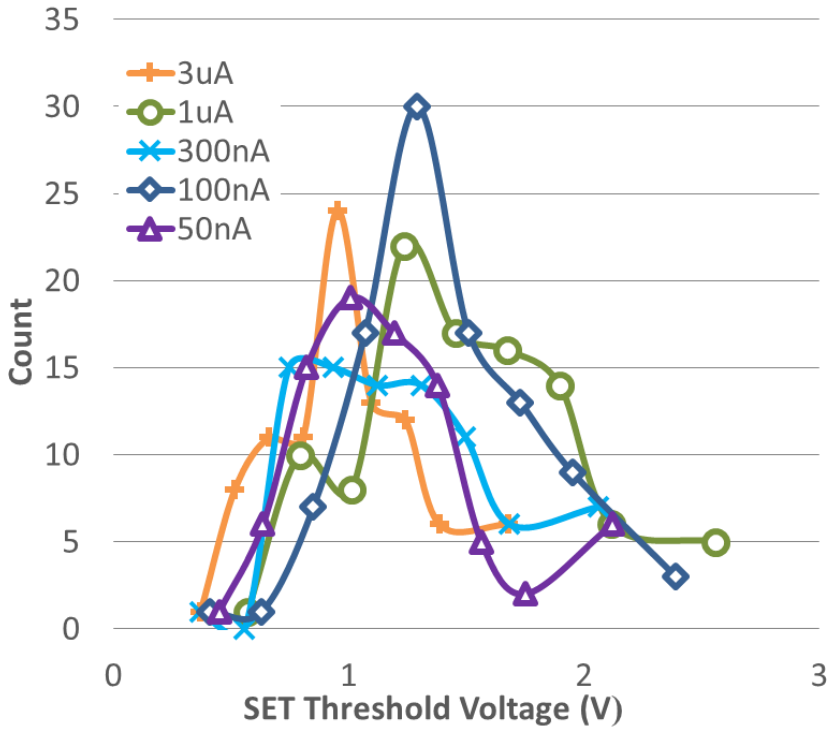


Resistance

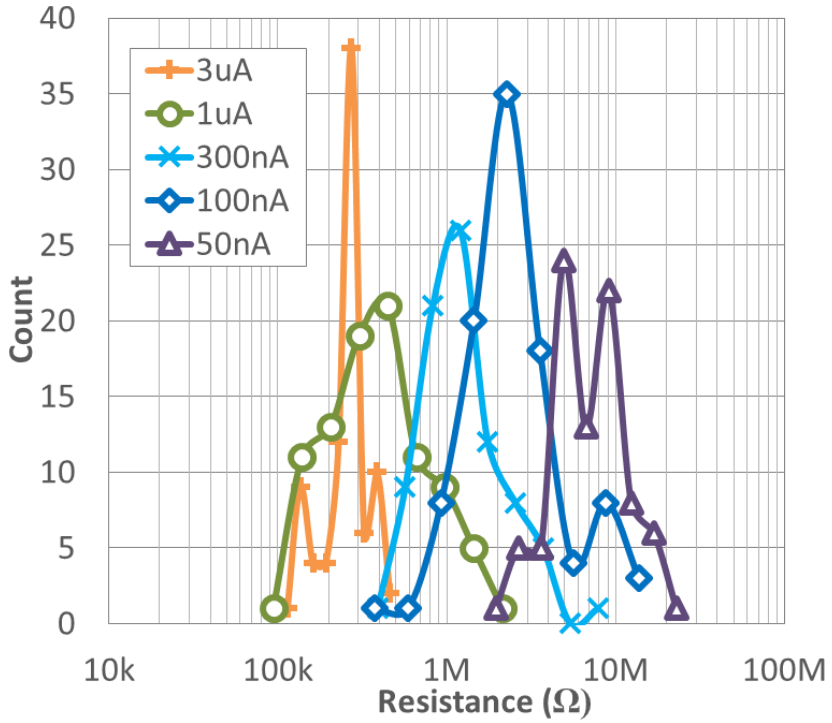


Stochastic CBRAM Behavior

V_{TH_Set} Distributions

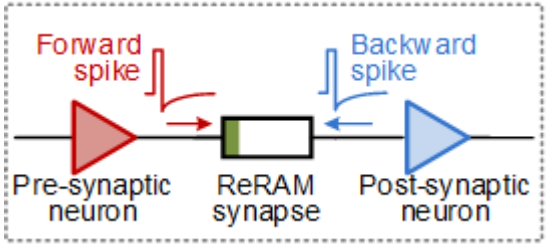


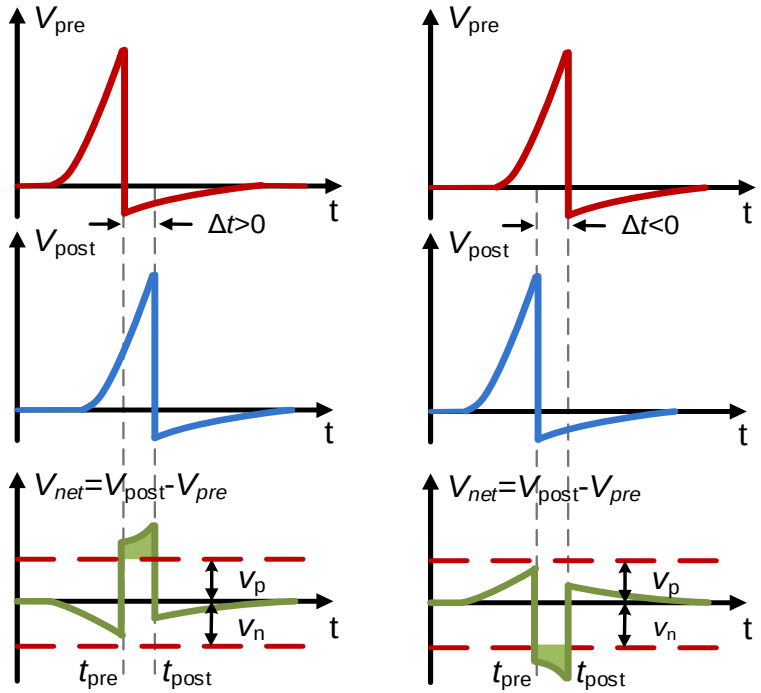
R_{ON} Distributions

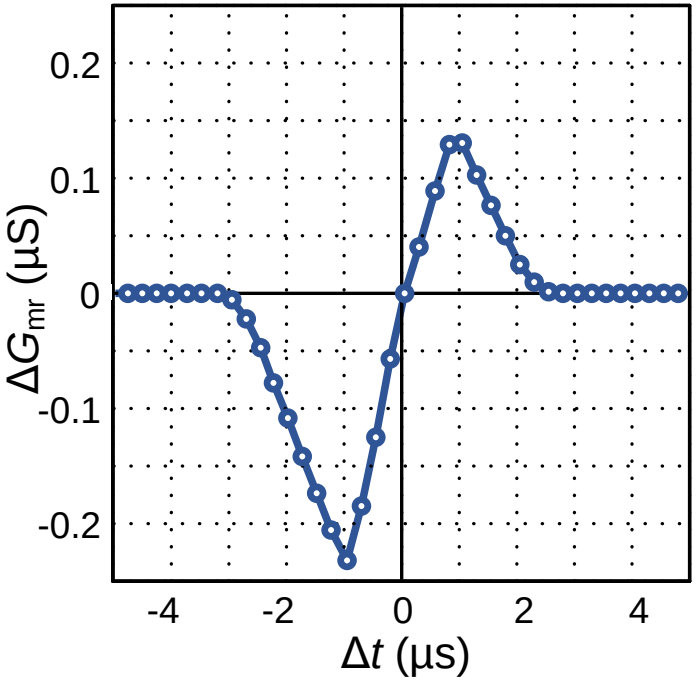


- The CBRAM switching behavior is stochastic
- The 'analog' synapse value depends upon the electroforming process
  - Weak bridge-> more intermediate resistance values
  - The analog resistance value relaxes in few minutes to hours

Spike Timing Dependent Plasticity (STDP)





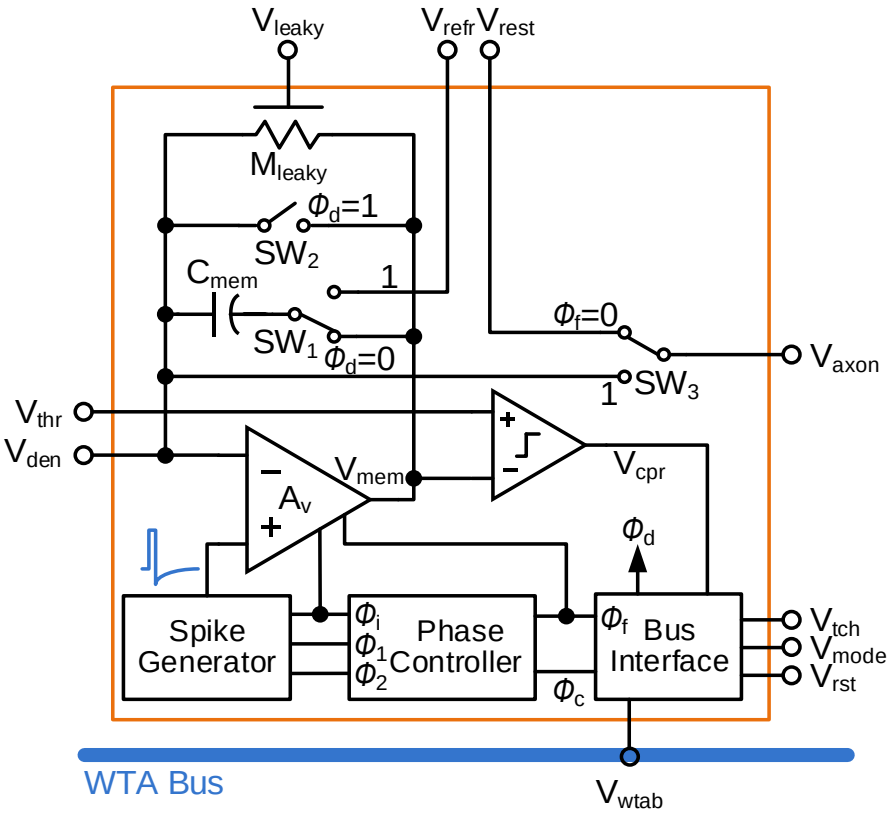


Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Accepted.

CBRAM Synapse Challenges

- ***Variability:*** Device characteristics such as the switching threshold voltages are variable from device to device
  - Switching characteristics depend upon the initial 'forming' step and compliance current
- ***Resistive Load:*** Thousands of devices in parallel present a difficult (low resistance) load to drive by the neuron circuit
- ***Resolution:*** Challenging to realize long-term persistence of weights for more than 1-bit resolution in CBRAM devices

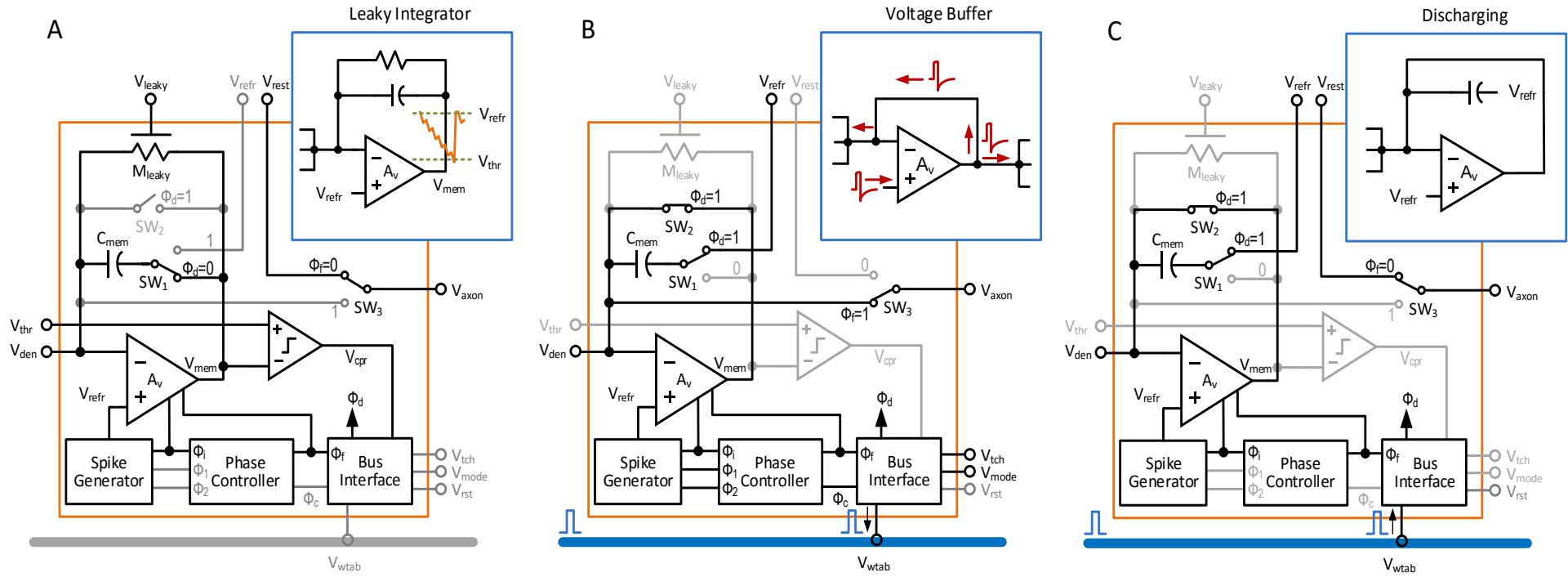
CMOS Integrate and Fire Neuron



- **Asynchronous integrated and fire operation**
- **Accommodates RRAM**
  - Provides current summing node
  - Sustain a fixed voltage for synapses
  - Dynamic biasing for large current drive
- **STDP compatible**
  - STDP-compatible spike generator
  - In-situ STDP learning
- **Winner-takes-all bus for effective local decision making**
- **Single contact point to a synapse**

Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015.

Novel Tri-mode Operation

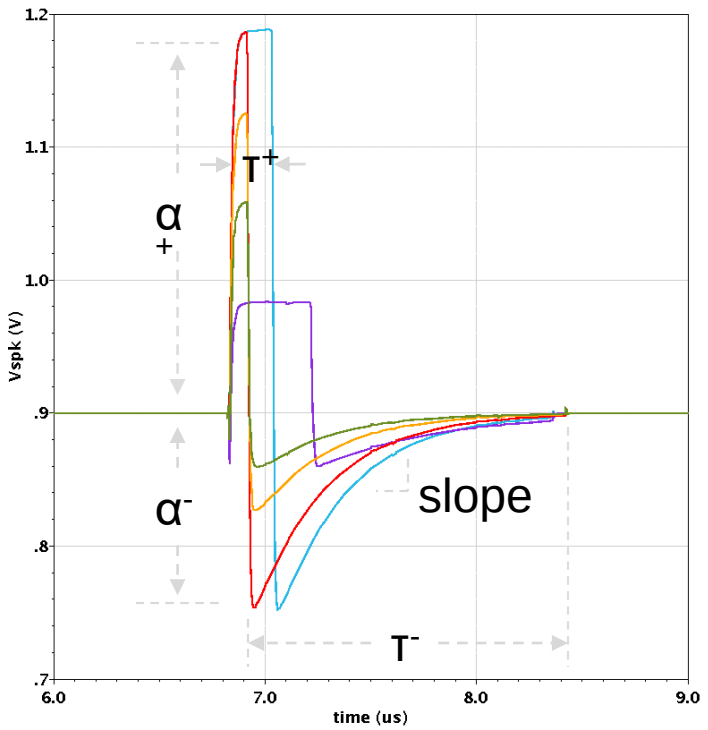


- A first CMOS neuron that is compatible with RRAMs and drives a large load
  - Several resistive devices in parallel load the neuron

Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015.

Simulation: Output Spikes

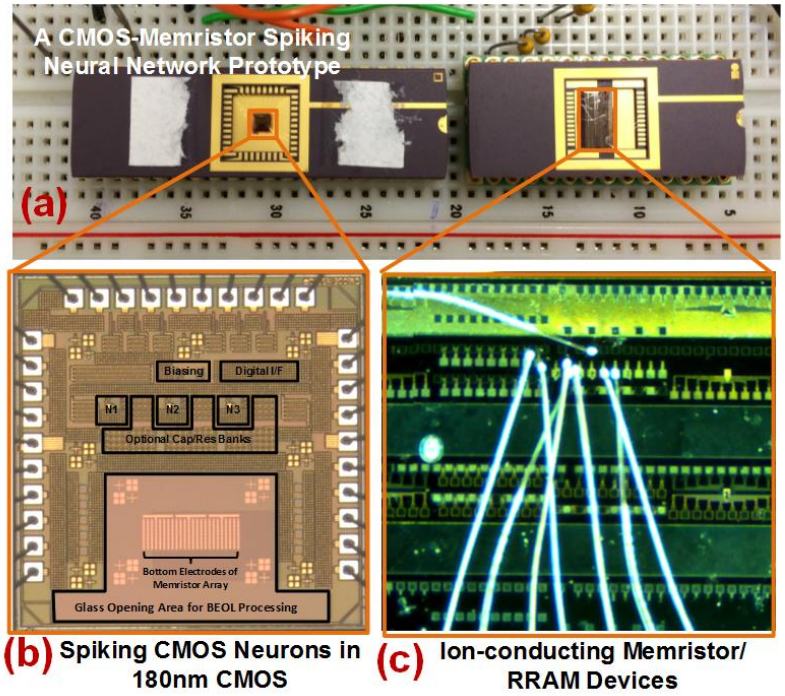
Output Spikes

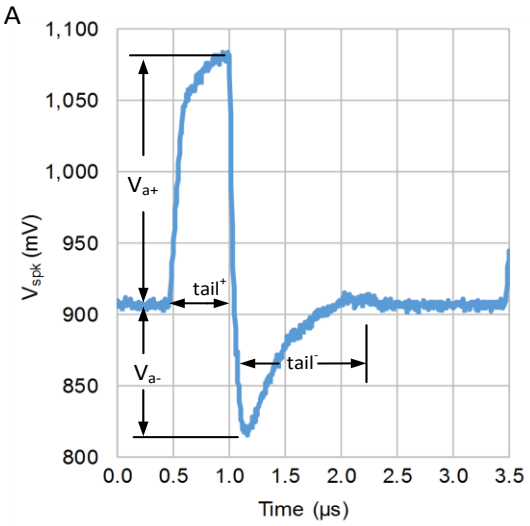


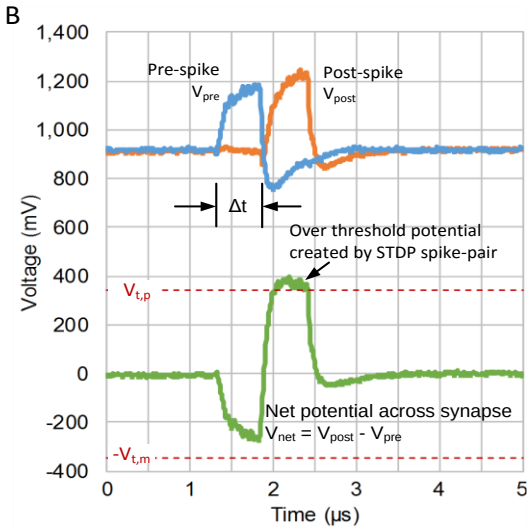
Parameter	Symbol	Unit	Min	Max	Step
Positive amplitude	α^+	mV	0	360	24
Negative amplitude	α^-	mV	0	360	24
Positive pulse width	τ^+	ns	48	396	23
Negative pulse width	τ^-	ns	282	1125	56
Negative pulse slope	s	mV/ns	0.011	0.52	0.034

Source: X. Wu, V. Saxena and K. Zhu, "A CMOS Spiking Neuron for Dense Memristor-Synapse Connectivity for Brain-Inspired Computing", *International Joint Conference on Neural Networks (IJCNN)*, 2015.

Test Chip interfaced with CBRAM Devices



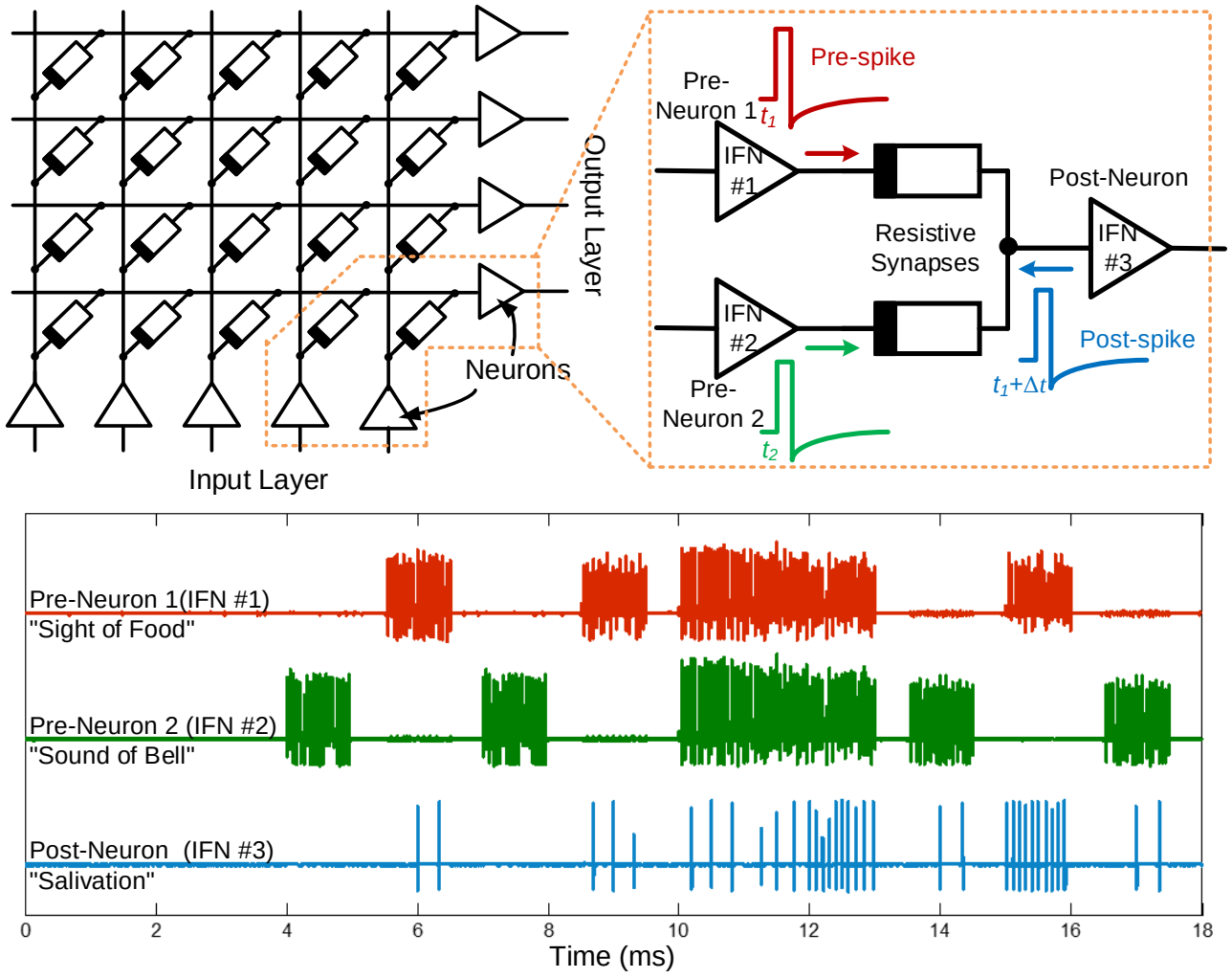




- Mixed-signal IC with 180-nm CMOS Neurons connected with CBRAM Synapses
- A first RRAM compatible STDP learning Neuron
- Power consumed in the neuron is 9.3pJ/spike/synapse (in terms of 10,000 synapses each having around 1MΩ resistance).

Source: X. Wu, **V. Saxena**, K. Zhu and S. Balagopal, "A CMOS Spiking Neuron for Brain-Inspired Neural Networks With Resistive Synapses and In Situ Learning", *IEEE TCAS-II*, *62*(11), pp.1088-1092.

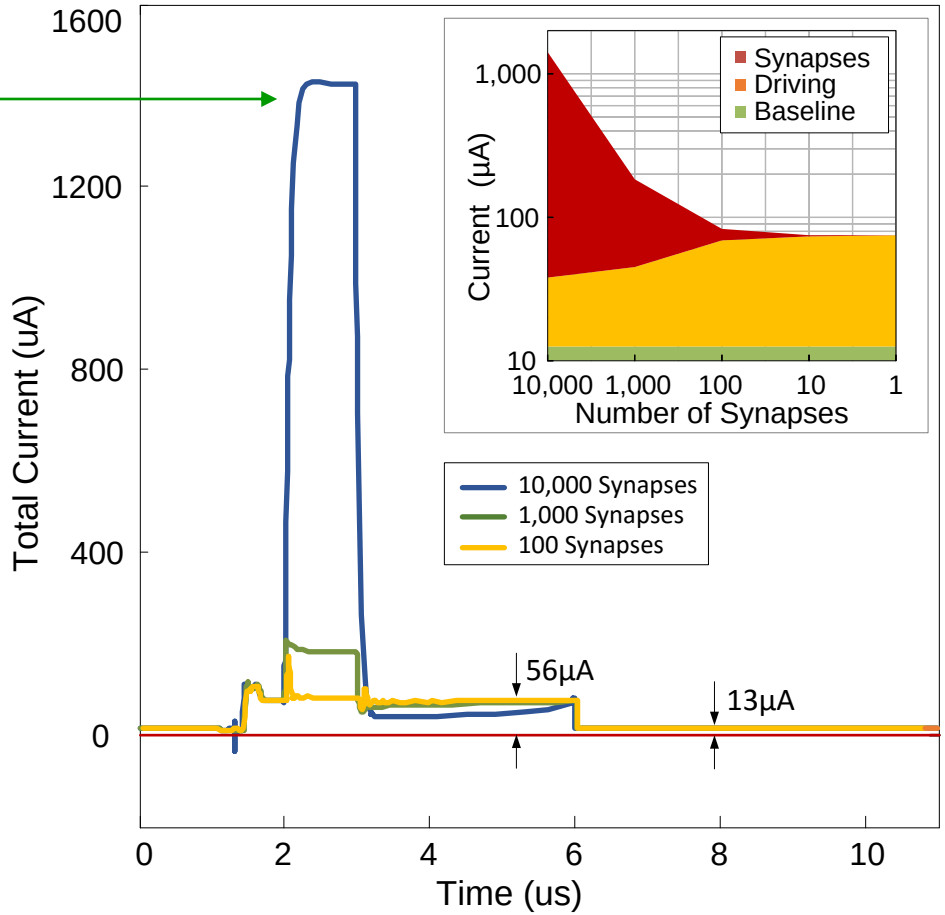
Associative Learning Experiment



Source: X. Wu, V. Saxena, K. Zhu and S. Balagopal, "A CMOS Spiking Neuron for Brain-Inspired Neural Networks With Resistive Synapses and In Situ Learning", *IEEE TCAS-II*, *62*(11), pp.1088-1092.

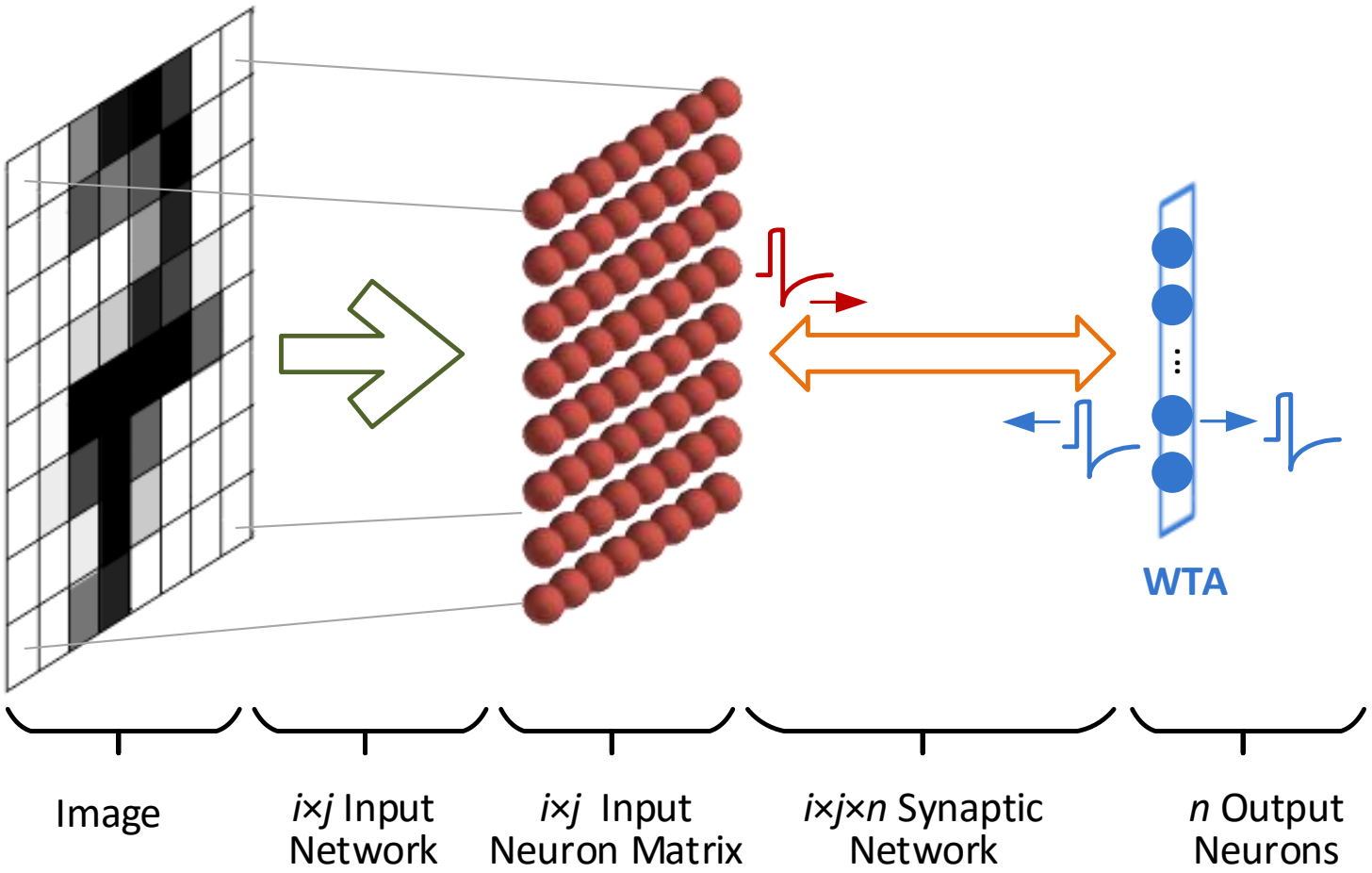
Dynamic Resistive Load Driving

97% Efficiency
@ driving 10,000 x 1MΩ synapses



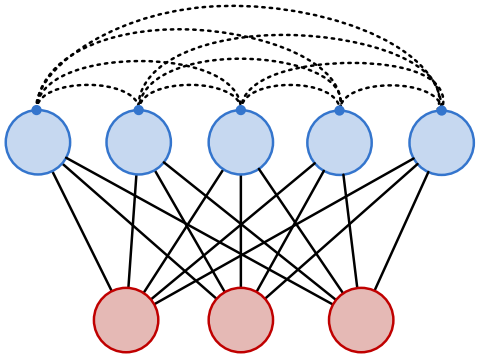
Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015.

Spike-based Pattern Recognition Circuit



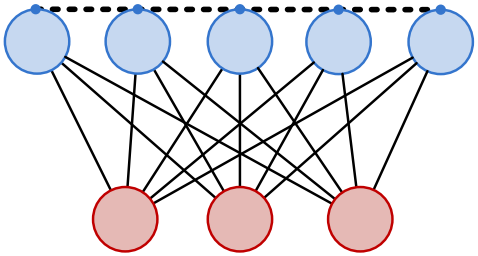
Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015, in press.

Asynchronous WTA Bus Interface

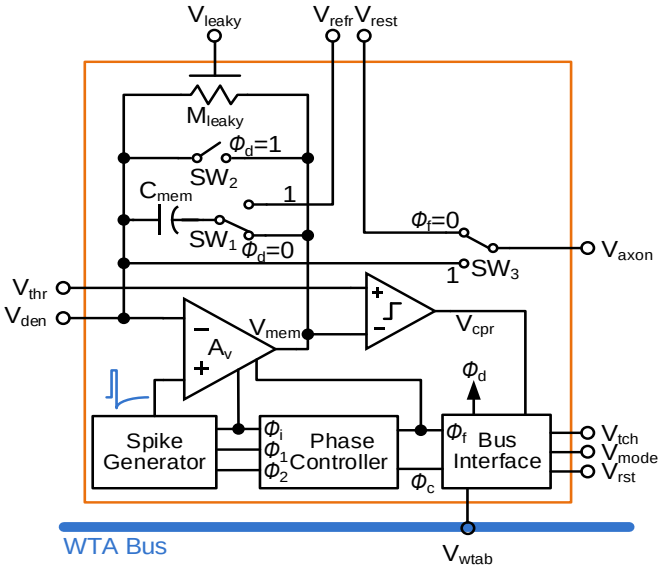


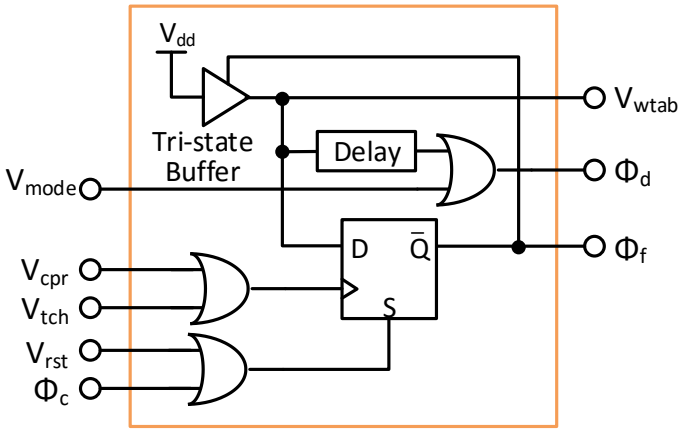
Explicit one-one inhibitory connections

Inhibitory connection
Excitatory connection



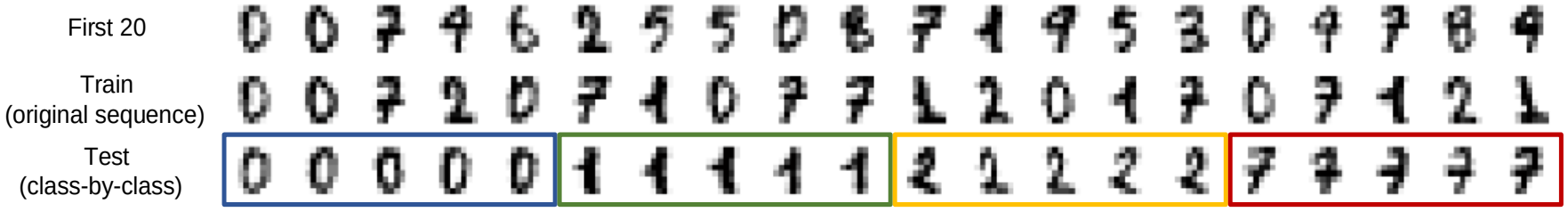
Implicit inhibition via shared WTA bus

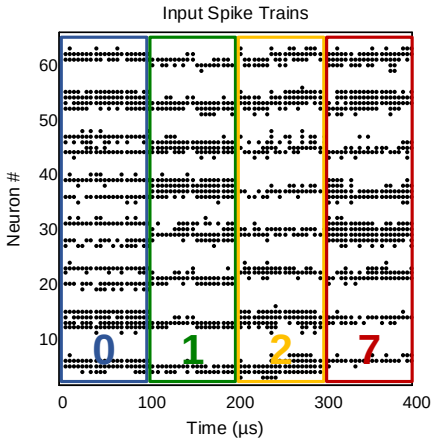


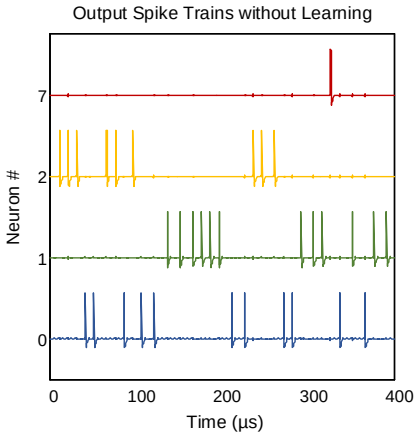


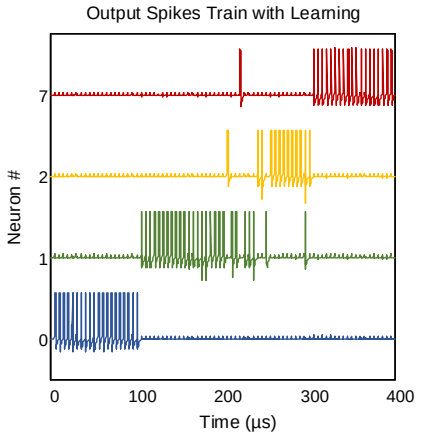
Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Accepted.

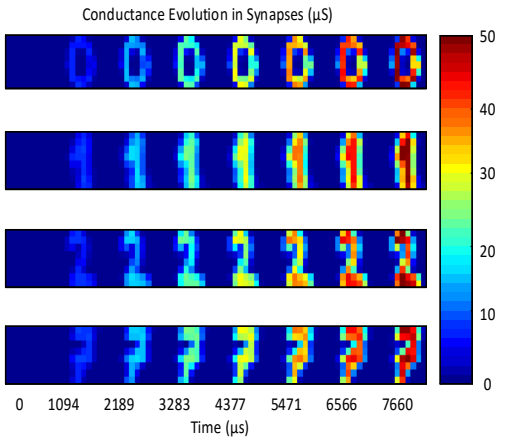
Demo: Handwrite Digits Classification







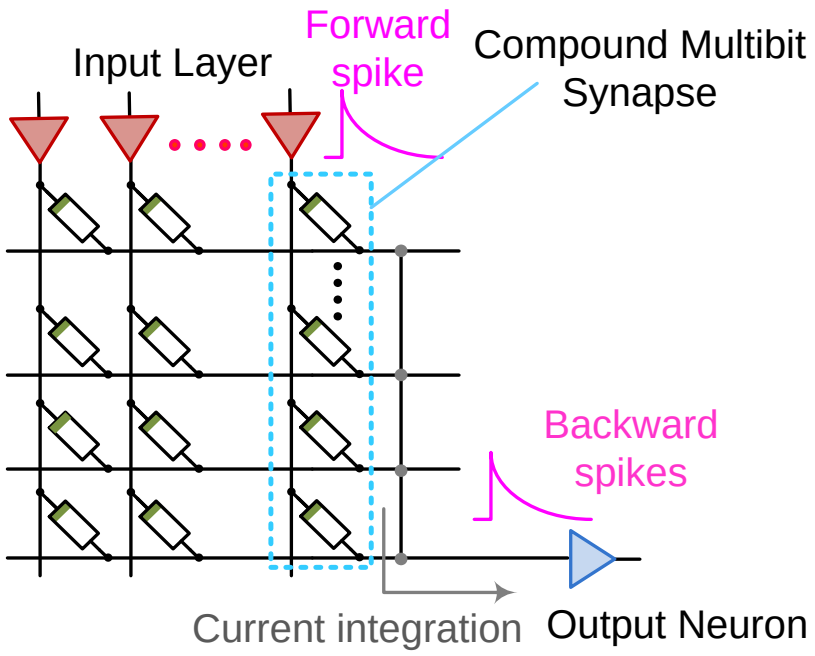




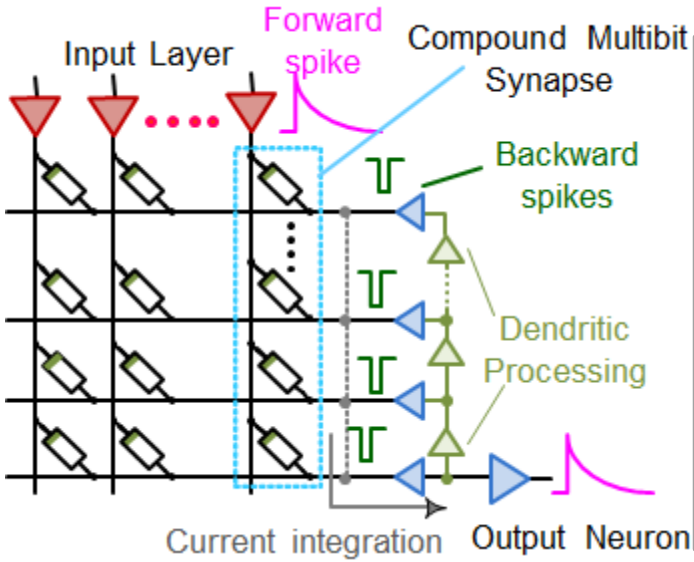
Source: X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015.

Parallel Multibit Synapses

- Emerging spike-based deep learning architectures require at least **4-bits** of synaptic resolution.
- The devices fundamentally have bistable nonvolatile states with stochastic switching behavior
- ***Recent approach:*** Use several of these device in parallel
  - Potentially multiple resistance states due to variation in individual threshold voltages



Multibit Synapses: Dendritic Processing



- Dendritic processing to include sensitivity to time-overlap between the pre- and post-synaptic spikes
- Shift in time or pulse voltage in the back-propagating spikes
- Simulations exhibit remarkable STDP learning behavior; experiments underway with physical devices

X. Wu and V. Saxena, "Realizing Multibit Synapses with Bistable RRAM Devices", *Unpublished result*.

Conclusion

- Nanoscale emerging memory devices present opportunity for realizing non von Neumann computing
- Mixed-Signal architectures for spiking neural networks can realize a versatile neuromorphic computing motif
- Further need to characterize device capabilities and limitations for large scale integration
- Algorithm/architecture development to be driven ground up by emerging device capabilities

References

1. X. Wu, V. Saxena and K. Zhu, "Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3), 2015, in press.
2. X. Wu, V. Saxena and K. Zhu, "A CMOS Spiking Neuron for Dense Memristor-Synapse Connectivity for Brain-Inspired Computing", *International Joint Conference on Neural Networks* (*IJCNN)*, 2015, Accepted.
3. X. Wu, V. Saxena and K. A. Campbell, "Energy-efficient STDP-based learning circuits with memristor synapses", *SPIE Machine Intelligence and Bio-inspired Computation: Theory and Applications VIII*, 2014.
4. X. Wu, V. Saxena and K. Zhu, "A CMOS Spiking Neuron for Brain-Inspired Neural Networks with Resistive Synapses and In-Situ Learning", *IEEE Transactions on Circuits and Systems II: Express Briefs*, In final review.
5. X. Wu and V. Saxena, "Associative Learning on CMOS-Memristor Spiking Neuromorphic Chip", *IEEE Transactions on Nanotechnology,* In 2nd round review.

Questions?