Microsystems
Science, Technology & Components

SAND2015-4984C

Characterizing Switching Variability in TaO_x Resistive Memories

D. R. Hughart, P. R. Mickel, R. T. Apodaca,
G. S. Haase, F. El Gabaly, A. A. Talin,
C. D. James, and M. J. Marinella

Sandia National Labs

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation,
a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's
National Nuclear Security Administration under contract DE-AC04-94AL85000.

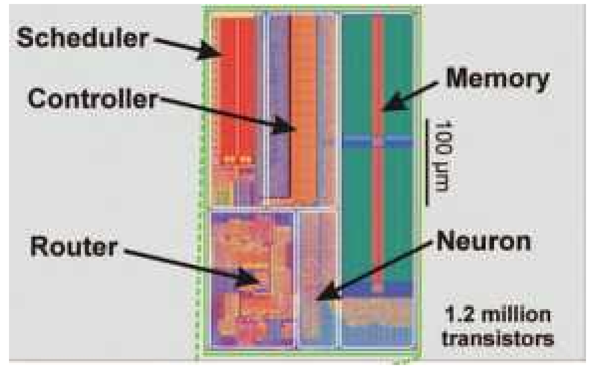
Sandia National Laboratories

Neural-inspired Computing Hardware

Mark I Perceptron (Rosenblatt 1960):

Arvin Calspan Advanced Technology Center; Hecht-Nielsen, R. *Neurocomputing* (Reading, Mass.: Addison-Wesley, 1990); Cornell Library;

DARPA , IBM TrueNorth (2014):

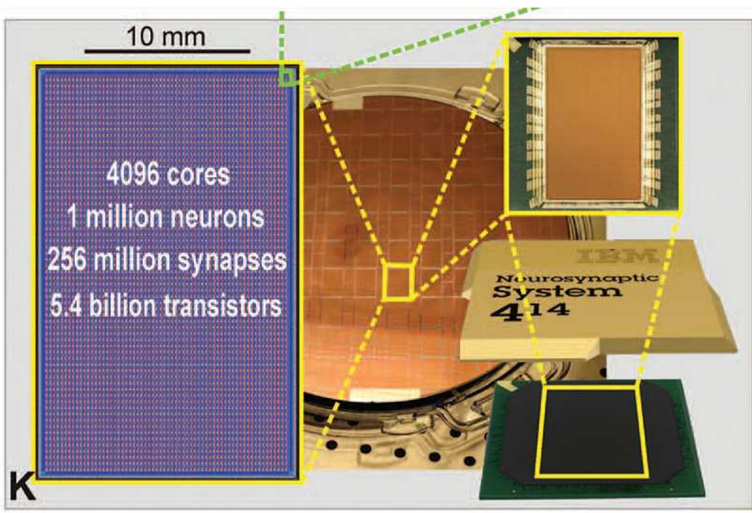


EU HBP, SpiNNaker (2014):

?

Current State of the Art in Neural Algorithm Computing

- CPU/GPU
  - Most general; common programming languages
  - Lowest power efficiency and performance
  - Memory separate from chip
  - Example: Google deep learning study (CPU→GPU)
- FPGA
  - General; requires hardware design language
  - Moderate performance and efficiency
- Custom IC (Truenorth, Spinnaker)
  - Specific: ex. executes STDP
  - Highest performance and efficiency
  - Expensive, 40MB local memory
  - Example: IBM Truenorth
- Moore law has provided enormous benefits to all of these technologies



Why Do We Need an HW Accelerator?

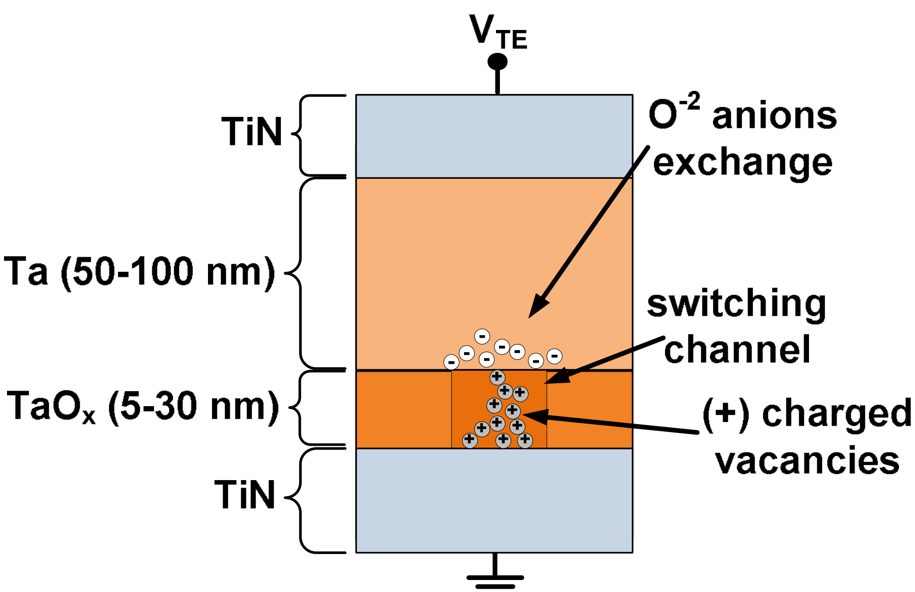
- **Problem: perceptron network training slow and extremely computationally intensive**
- Use simulation results for similar algorithm as example
- Significant power savings using a memristor-based HW accelerator :
- 16x reduction in power over SRAM ASIC
- 6x reduction in chip area over SRAM ASIC
  - Equivalent to 6x improvement in performance/area

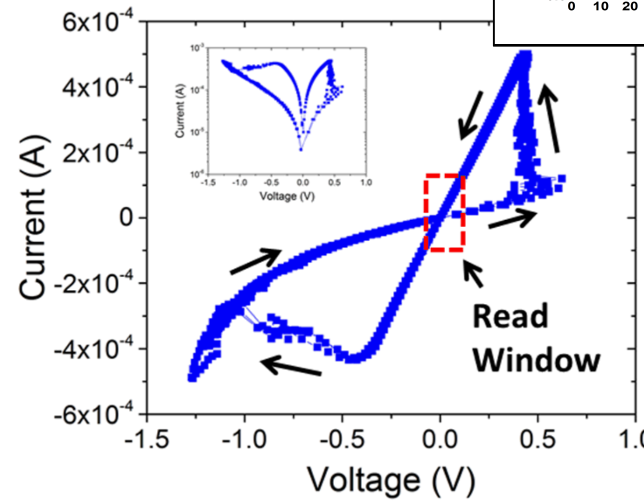
	Example 1: 25,600 neurons 100,000 iterations/s				
Configuration	# of chips	Chip area (mm^2)	% active	Power (W)	**Power eff. over Xeon**
Memristor Analog (config 4)	1	5.9	38.6%	0.07	**234,859**
Memristor Digital (config 5)	1	18.2	89.6%	0.62	**16,968**
SRAM (config 6)	1	29.1	89.6%	1.13	**8,215**
NVIDIA M2070	12	529.0	99.2%	2700.00	**6**
Intel Xeon X5650	179	240.0	99.9%	17005.00	**1**

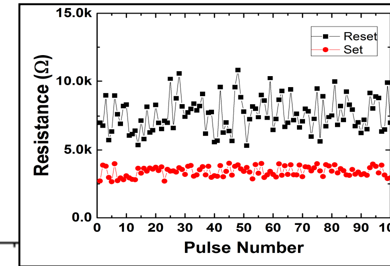
T. Taha, et al. IEEE IJCNN 2013.

Memristor I-V Characteristics

- Resistive RAM stores state in the form of resistance
- Applied current and voltage can change resistance state
  - Hysteresis loop
- Low voltages can read state
  - Read window





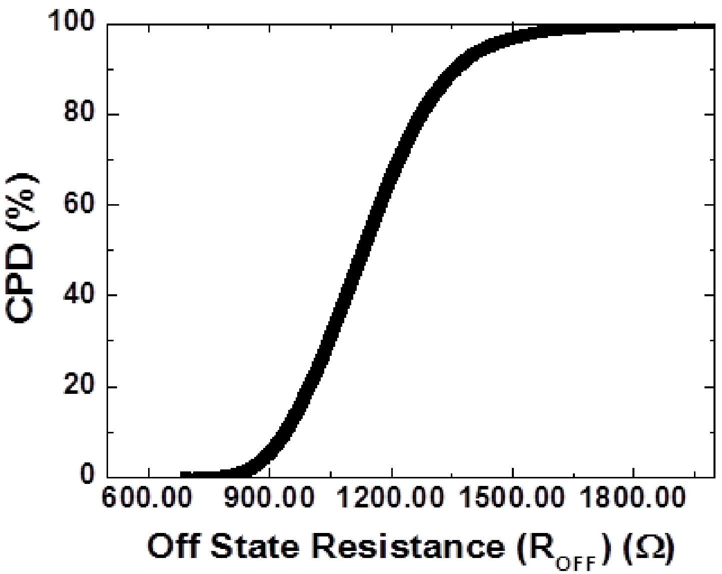


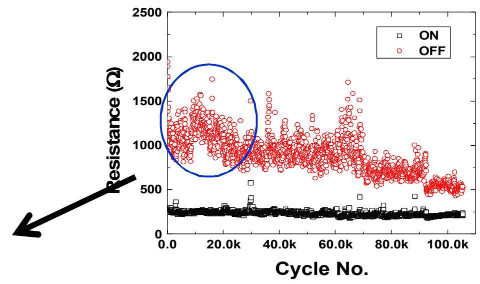
- Resistive switching
  - Oxygen vacancies
- TaO_x
  - Oxygen anions

Variability and Noise

- **Interdevice variability**: device to device, can be >10x
  - Variations in film thickness, topography
- **Intradevice variability**: cycle to cycle, can be >2x
  - Fundamental physical attributes
- **Random telegraph noise:**
  - Affects read current

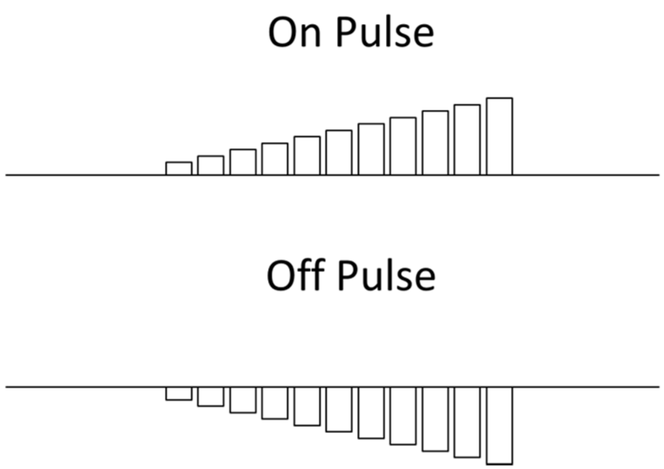
- Characterizing cycle to cycle variability is essential for future neuromorphic applications





Testing Algorithms

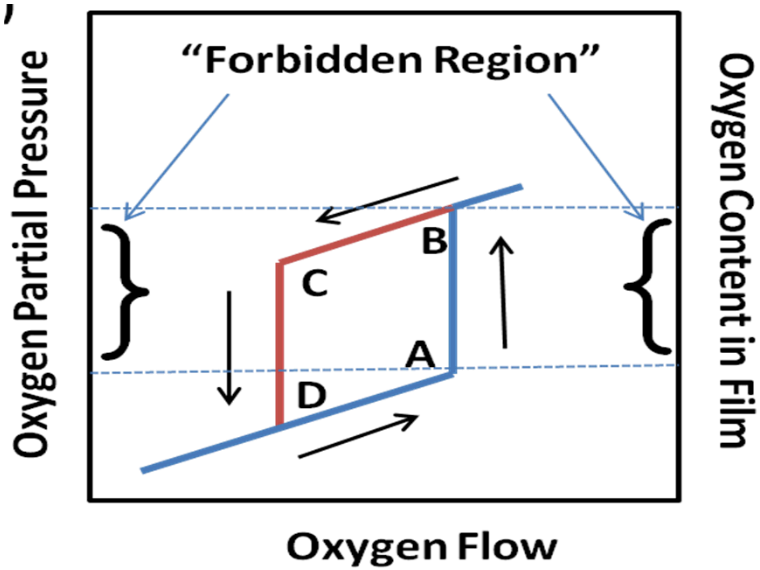
- An automated test algorithm has been developed to form, reset, and cycle devices with gradually increasing pulse heights
- The algorithm can be programmed to target specific resistances, stopping once a particular resistance threshold is reached



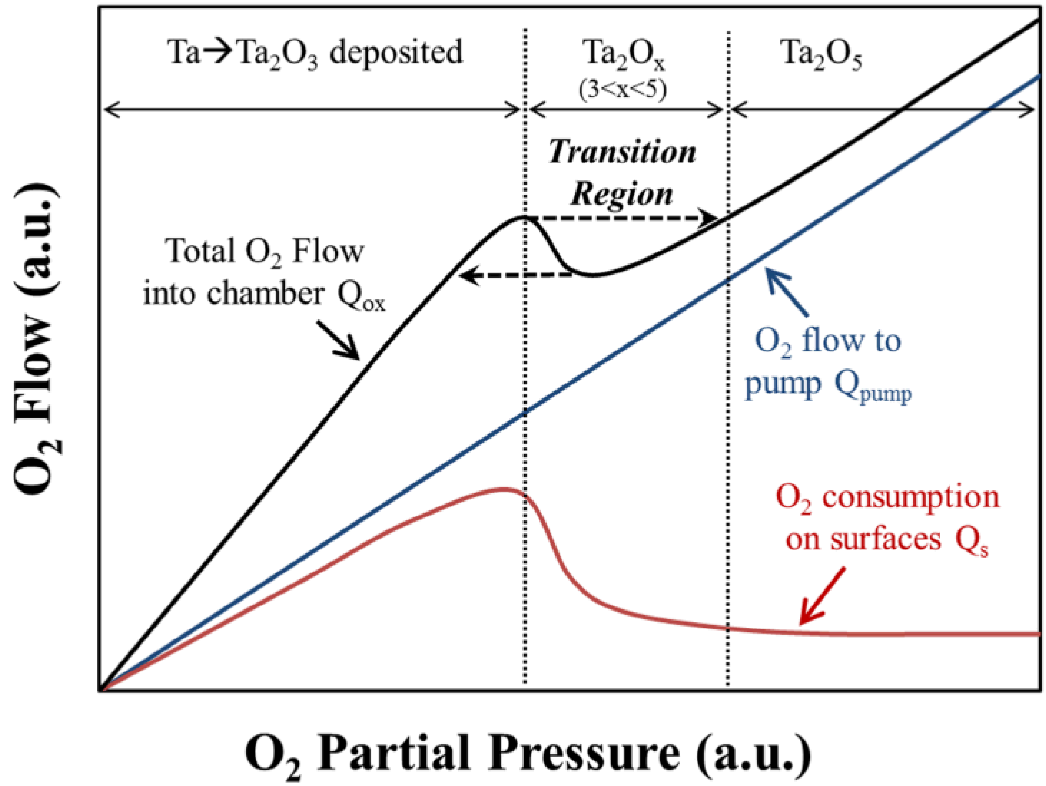
- Evaluate the effects of process changes like thickness and stoichiometry to optimize future devices for the targeted characteristics
  - Multi-state suitability

Switching Film Development: Overcoming the "Forbidden Region"

- One of the parameters that we vary is oxygen content
- Forbidden oxygen flow-pressure region occurs due to target poisoning
  - This is the region we need to be in to get ideal ReRAM stoichiometry



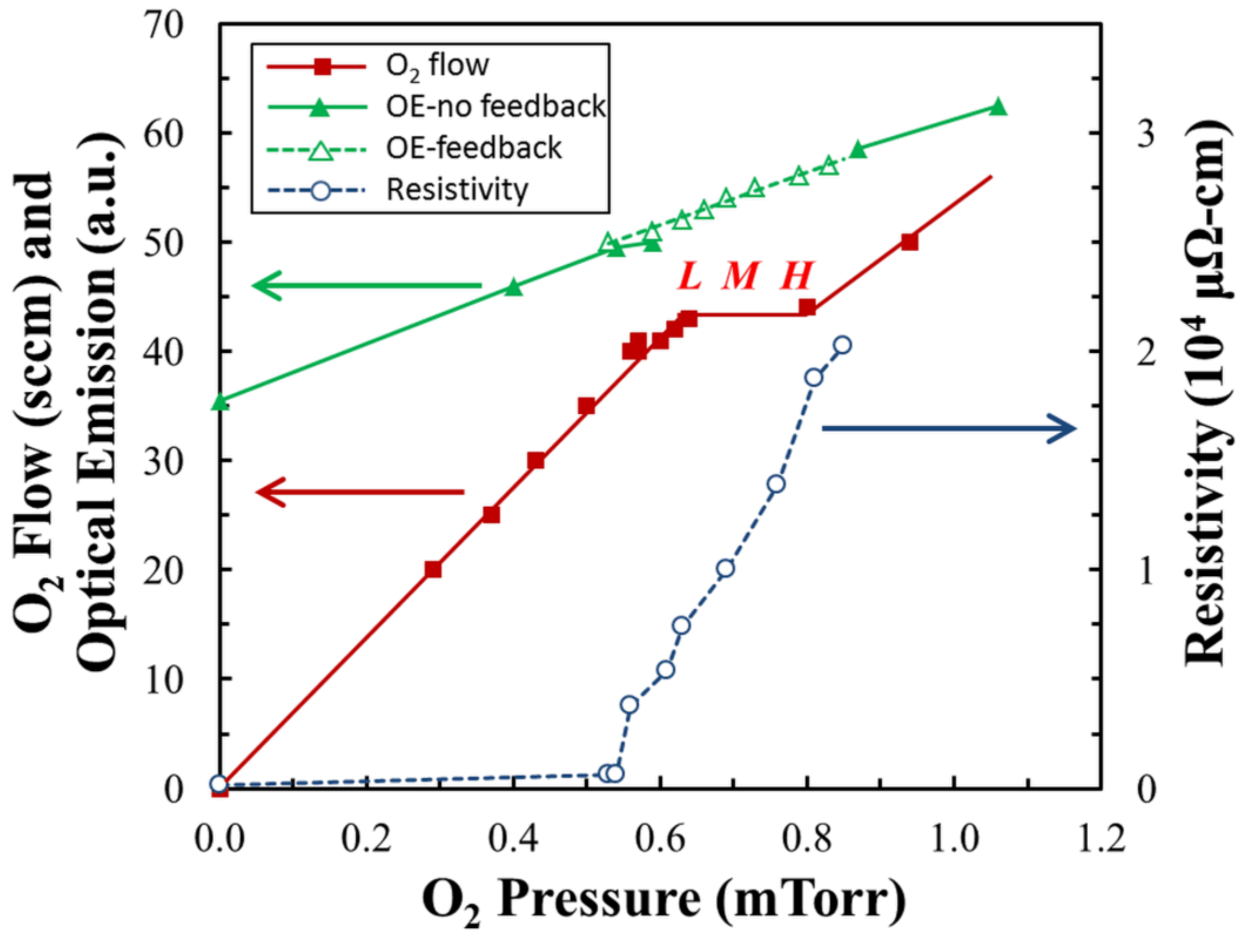
A.J. Lohn et al APL 103, 063502 (2013)



J.E. Stevens et al, J. Vac. Sci. Technol. A 32, 021501 (2014)

Solution: Feedback System

- **Penning gauge measures oxygen partial pressure**
- **Real time feedback to oxygen flow control**
- **Elimination of Forbidden Region!**



J.E. Stevens et al, J. Vac. Sci. Technol. A 32, 021501 (2014)

Effect on Device Behavior

Fraction of "forbidden region"	Partial pressure set #1	x: Ta_2O_x set #1	Yield (%)	Endurance	Partial pressure set #2	X: Ta_2O_x set #2
0	50.0	1.9 + 0.5	20	~3 cycles	43.0	2.1 + 0.5
1/3	52.5	3.3 + 0.5	80	$>10^4$ cycles	44.6	3.5 + 0.5
2/3	55.0	4.2 + 0.5	0	…	46.2	4.0 + 0.5
1	57.5	4.6 + 0.5	0	…	48.0	4.4 + 0.5
4/3	60.0	5.0 + 0.5	0	…	49.5	5.0 + 0.5

A.J. Lohn et al APL 103, 063502 (2013)

- **Thickness and stoichiometry splits tested for multi-level suitability**
- **Another lot had a range of partial pressures that spanned from 50% to 100% of the forbidden region and a thickness of 8 nm**
  - **No multi-level switching seen**
- **Another lot had a range of oxygen-flow that spanned from 60% to 150% of the forbidden region and thicknesses of 5, 8, and 11 nm**
  - **Only the 11 nm samples showed multi-level switching**
  - **May be due to higher resistance ranges**

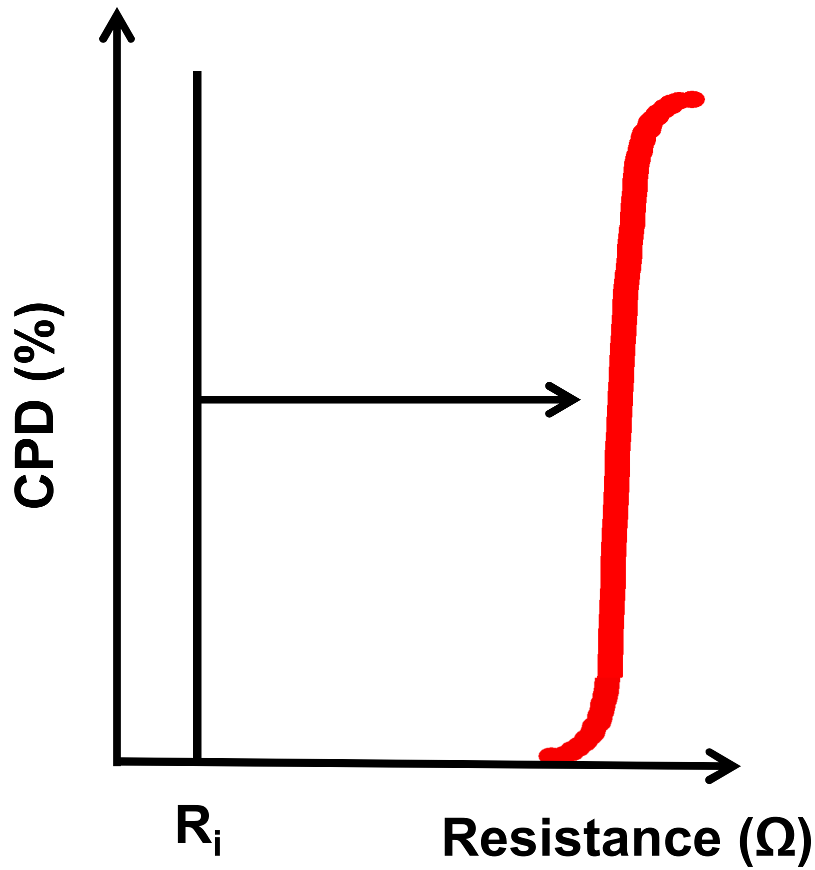
Modeling Variation

- Characterizing cycle to cycle variation can inform models to help predict whether the variation present will be tolerable by neuromorphic algorithms

- Analytical model

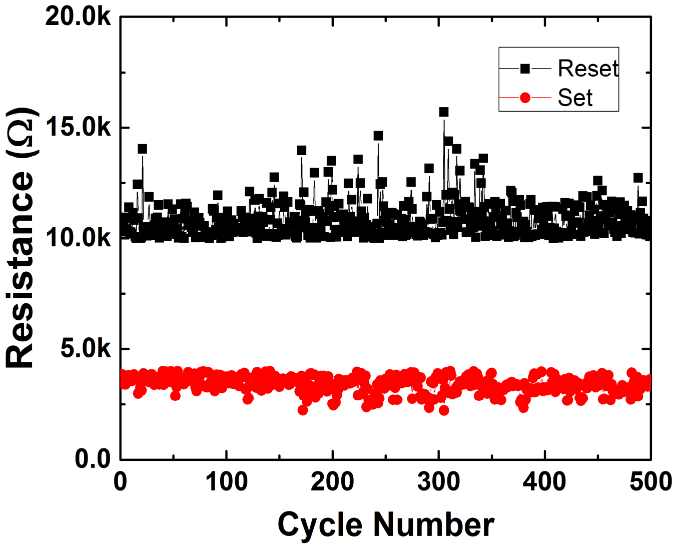
$$R_{final}(V,t) = R_i + R_{avg}(V,t) \pm R_{\sigma}(V,t)$$

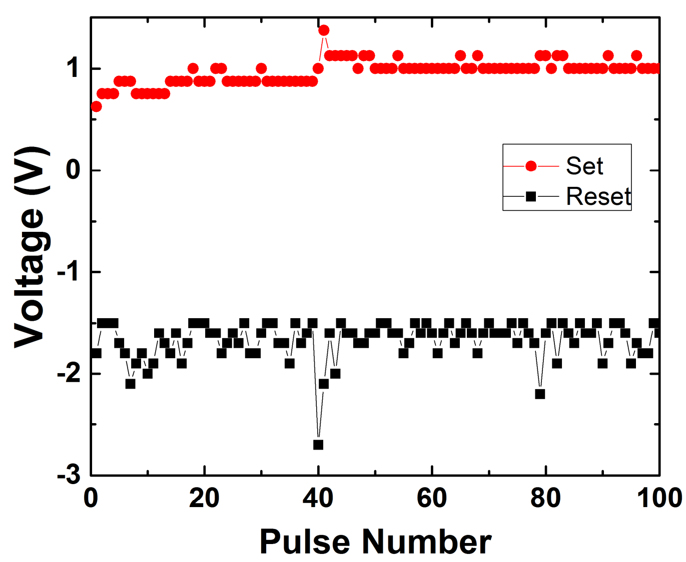
  – Feedback helps to achieve a more consistent R_i



Programming Resistance with Feedback

- Initial tests are run with feedback designed to target specific resistances
  – Useful when characterizing a specific transition
- Voltages are recorded when the resistance threshold is reached
  – Can be used to determine programming pulses for operation without feedback

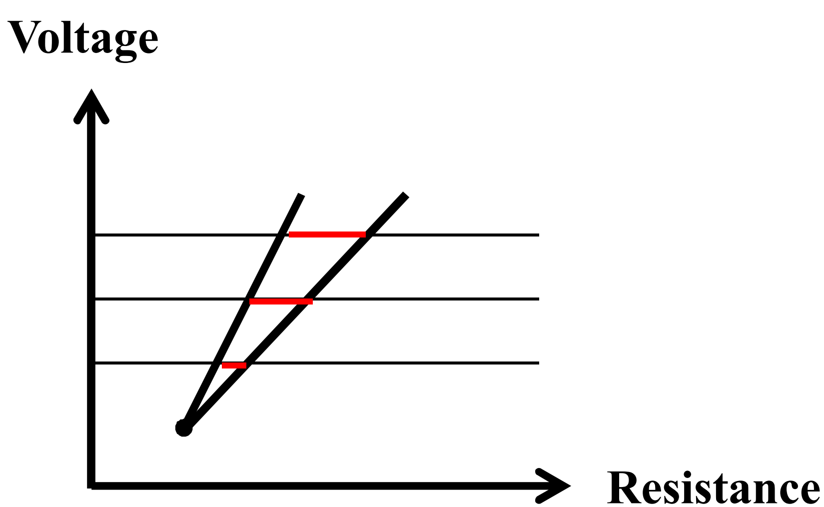




Sandia National Laboratories

Switching Variation for Multiple States

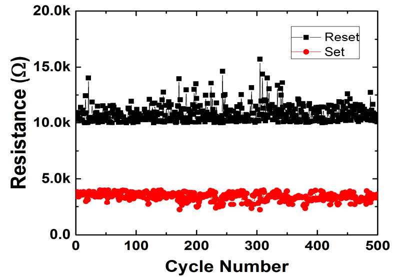
- One concern is how the variation changes when switching between multiple potential states
  - Does it vary based on the magnitude of the switch?



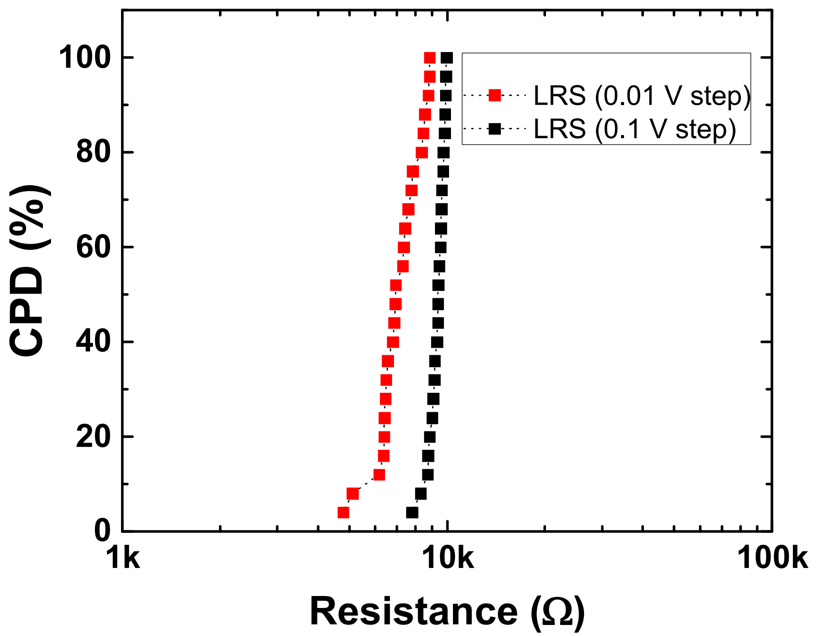
- To investigate, devices were programmed to a specific on-state resistance and programmed with a single reset pulse to multiple levels

Voltage Step Size

- The voltage step size used to program the on-state for the results in this talk was 0.1 V



- A finer step size was tested to see if the resistance distribution would be better



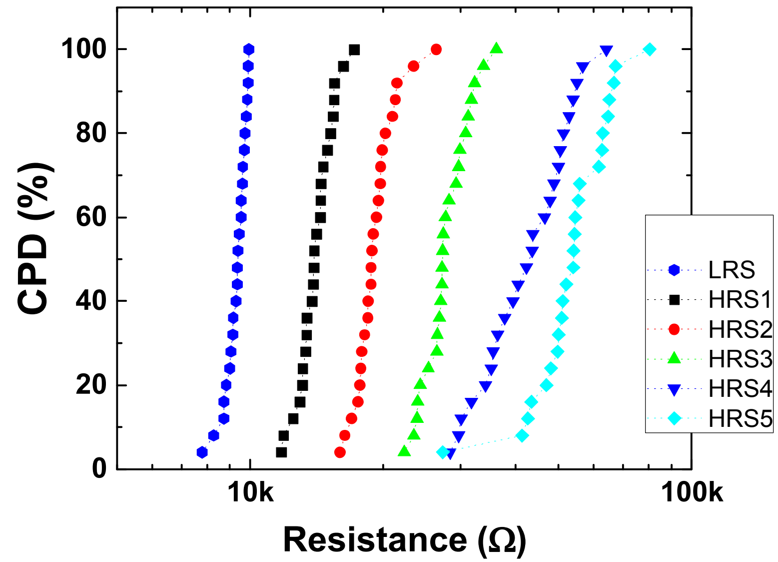
- The larger step size actually showed a larger distribution
  - Device quality likely a larger factor in this regime than voltage step size

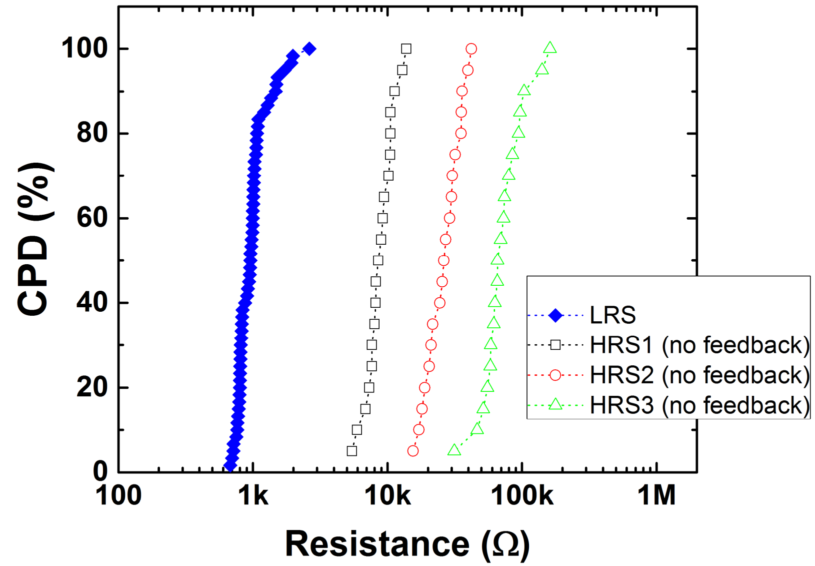
Multi-State Operation

- Devices were cycled with feedback to get the initial reset voltage, which was increased by -0.25 V for each test

- Targeted resistances for all levels also possible

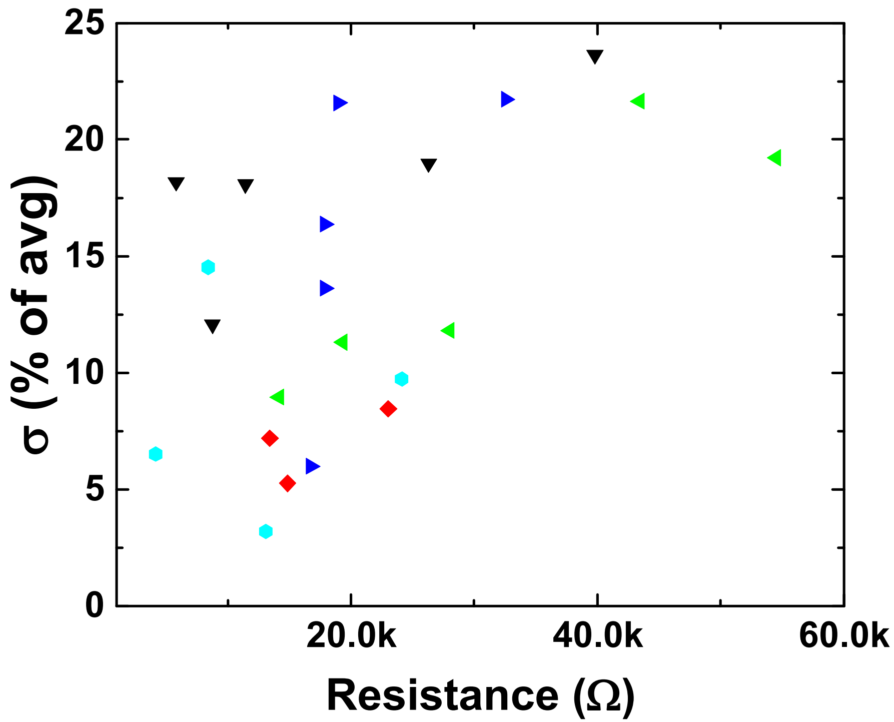
- Future tests will continue to investigate devices with different combinations of thickness and oxygen content to optimize multi-level switching





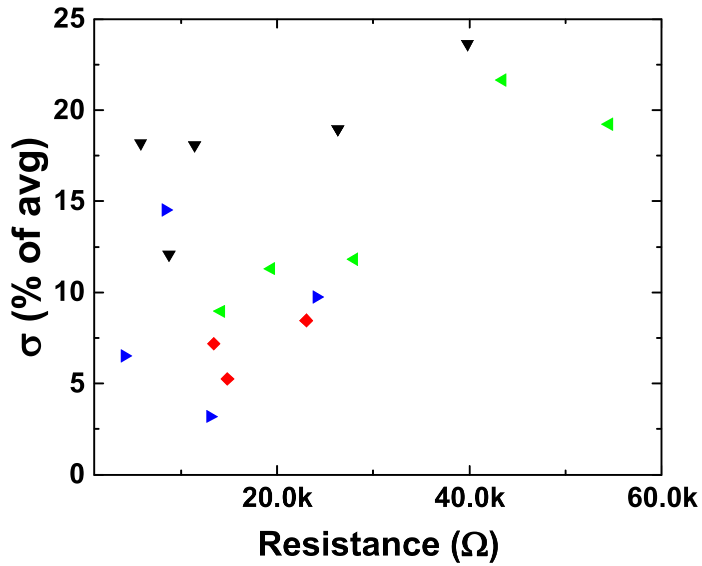
Variability at Different Resistances

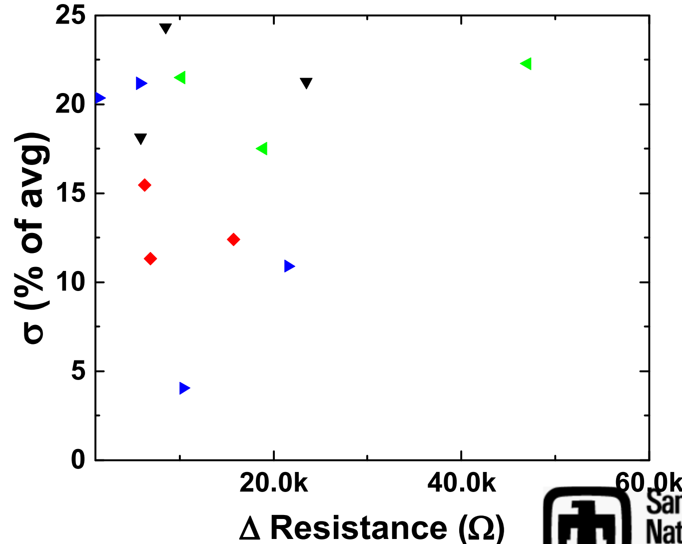
- The standard deviation as a percentage of the average resistance has been plotted at various resistances
- Initial data suggests that the variability in switching increases as the resistance increases
  - Fewer critical defects?
  - It may be easier to create more discrete states at lower resistances



Variability at Different Resistances

- The standard deviation as a percentage of the average resistance increases with increasing resistance
  - Not with magnitude of resistance change
- Initial data suggests that the variability in switching increases as the resistance increases
  - Fewer critical defects?
  - It may be easier to create more discrete states at lower resistances





Summary

- Feedback algorithms have been used to gather data on cycle to cycle variability
  - Essential for modeling realistic devices for neural algorithms
  - Data collection is ongoing, expanding to include more process variables and programming pulses
- Multiple high resistance states have been demonstrated on TaO_x devices
  - Thicker films show more promise for multi-state operation
  - Higher resistance range will be investigated
  - Multiple methods of multi-state programing
- Cycle to cycle variation tends to increase at higher off-state resistances
  - The number of discrete states may increase when targeting lower resistances