Learning Accuracy Analysis of Memristor-based Nonlinear Computing Module on Long Short-term Memory

Hongyu An, Mohammad Shah Al-Mamun, Marius K. Orlowski, Yang Yi

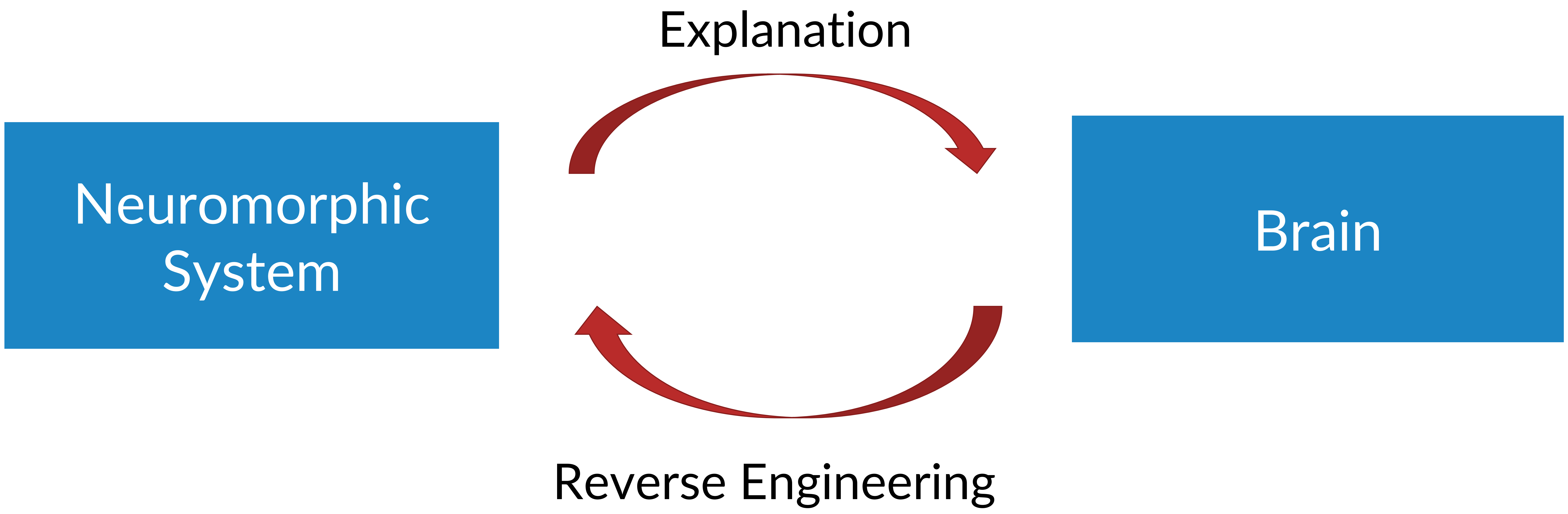
Multifunctional Integrated Circuits and Systems (MICS) Group
The Bradley Department of Electrical and Computer Engineering
Virginia Tech, Blacksburg, VA, USA

July 25, 2018

Outline

- Backgrounds and Motivations
  - Von Neumann Computing Architecture Revisit
  - Emerging Neuromorphic Computing Architectures
- Memristor-based Nonlinear Computing Module
- Learning Accuracy Analysis of Memristor-based Nonlinear Computing Module on Long Short-term Memory
- Conclusions

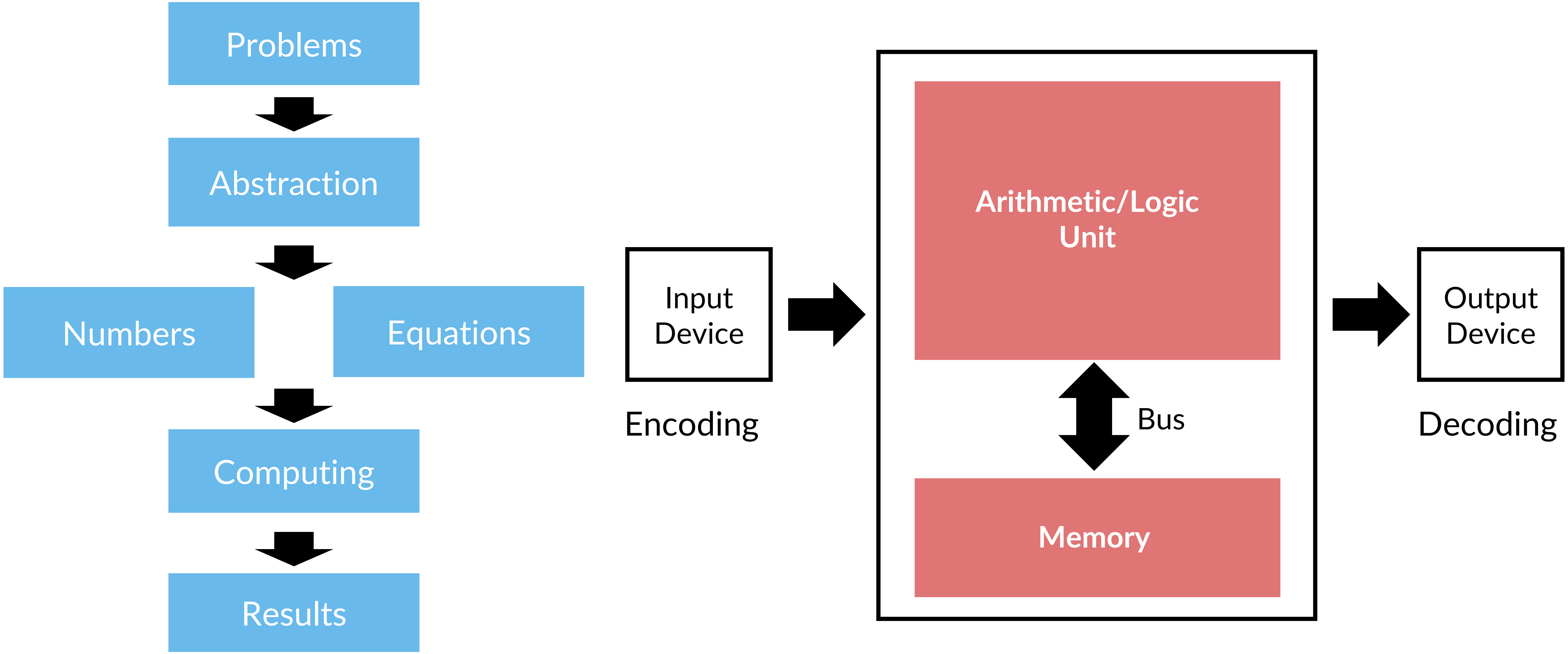
Neuromorphic Computing



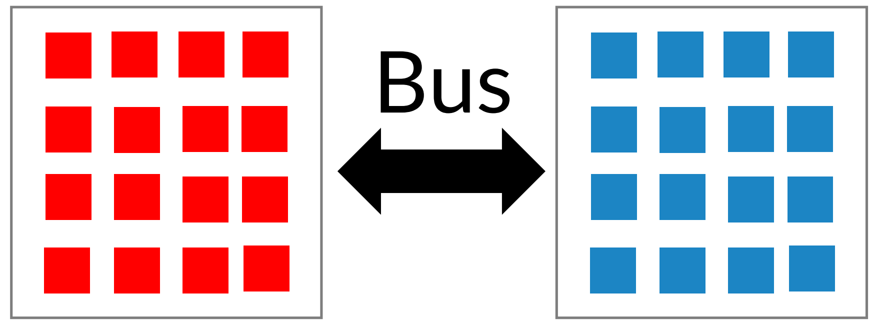
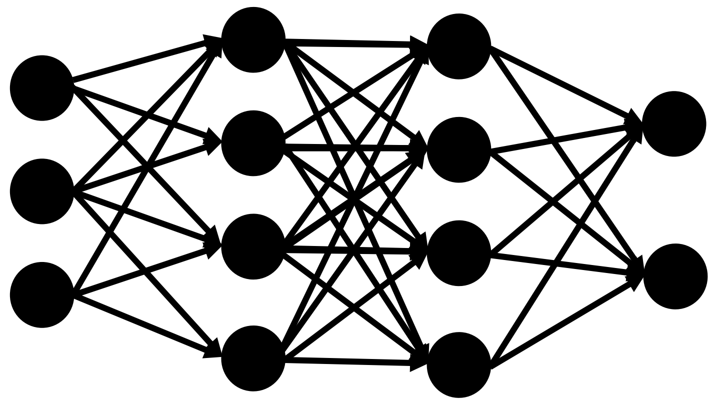
Engineering Contributions:
- More power efficiency system
- Neuromorphic learning system;

Scientific Contributions:
- Optical illusion
- Mechanism of Memory
- Cognition

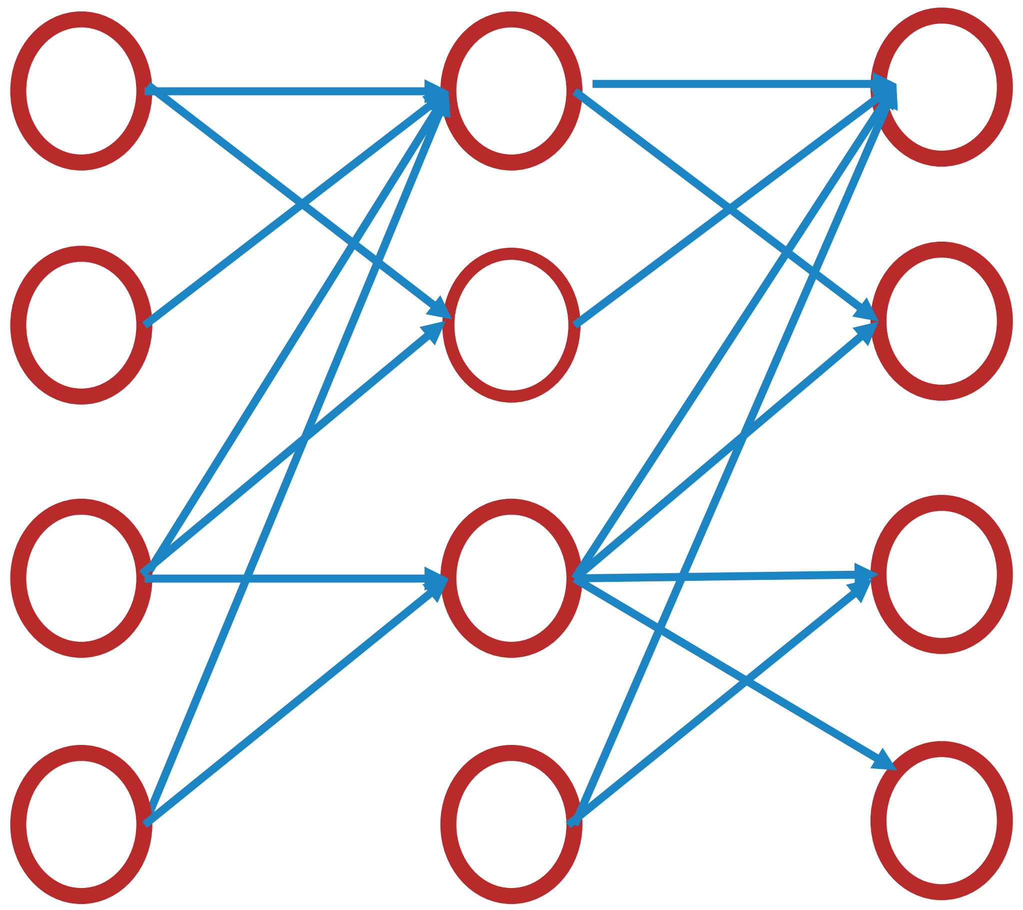
Von Neumann Architecture: Design for Computing

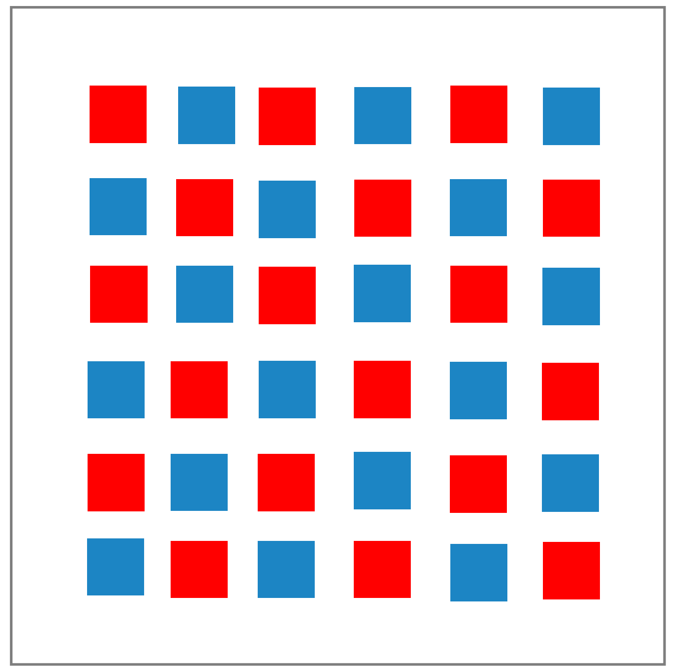


Neuromorphic System: Design for Learning

Digital Computer		Neuromorphic System
Programs/Logic	Algorithm	Learning Algorithms
Binary signals	Encoding Scheme	Analog signals
 CPU Memory Von Neumann Architecture	Architecture	 Neural Networks
CPUs(Logic Gates, etc.), Memory(SRAM, etc.)	Devices	Neurons and Synapses

Emerging Neuromorphic Computing Architectures: Distributed Neuromorphic Computing Architecture





Distributed Neuromorphic Computing Architecture

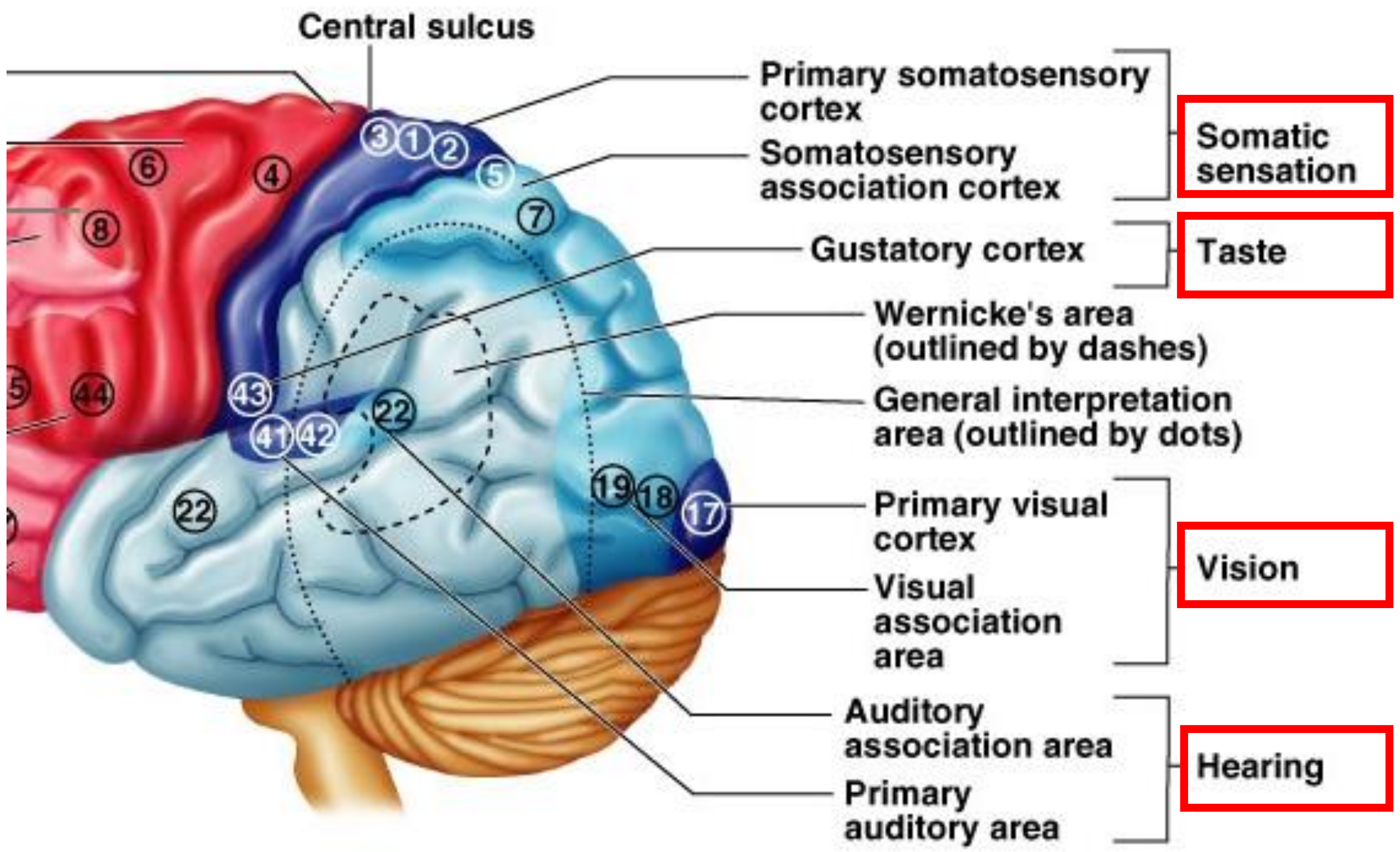
Neurons
Synapses

[An, H., et al. (2017)]

Analog Signals + Nonlinear Computing Module + Matrix Computing Module + Training Module

Basic Building Modules

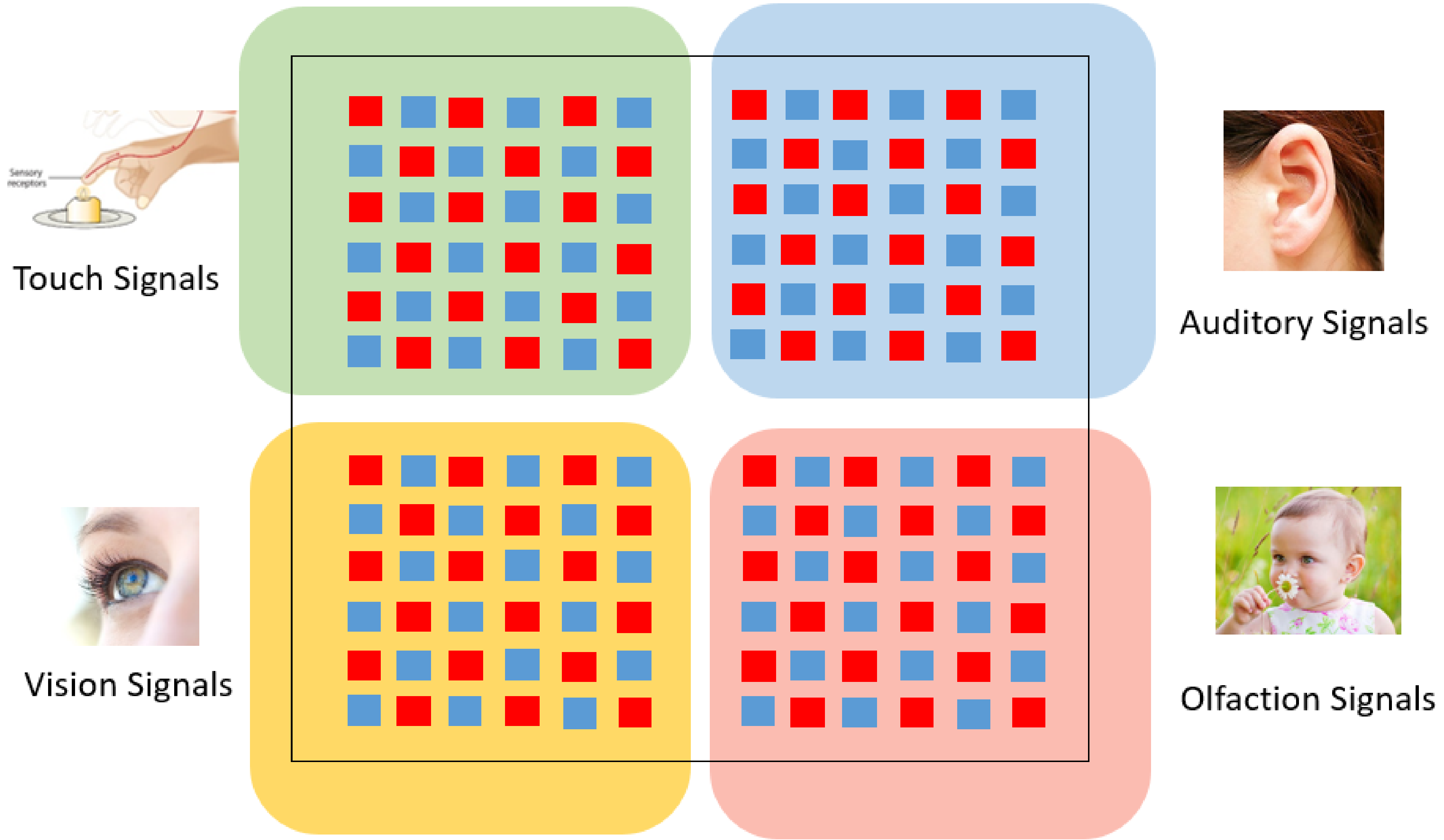
Emerging Neuromorphic Computing Architectures: Cluster Neuromorphic Computing Architecture



[Kandel, Eric R, et al. *Principles of Neural Science.* (2000)]

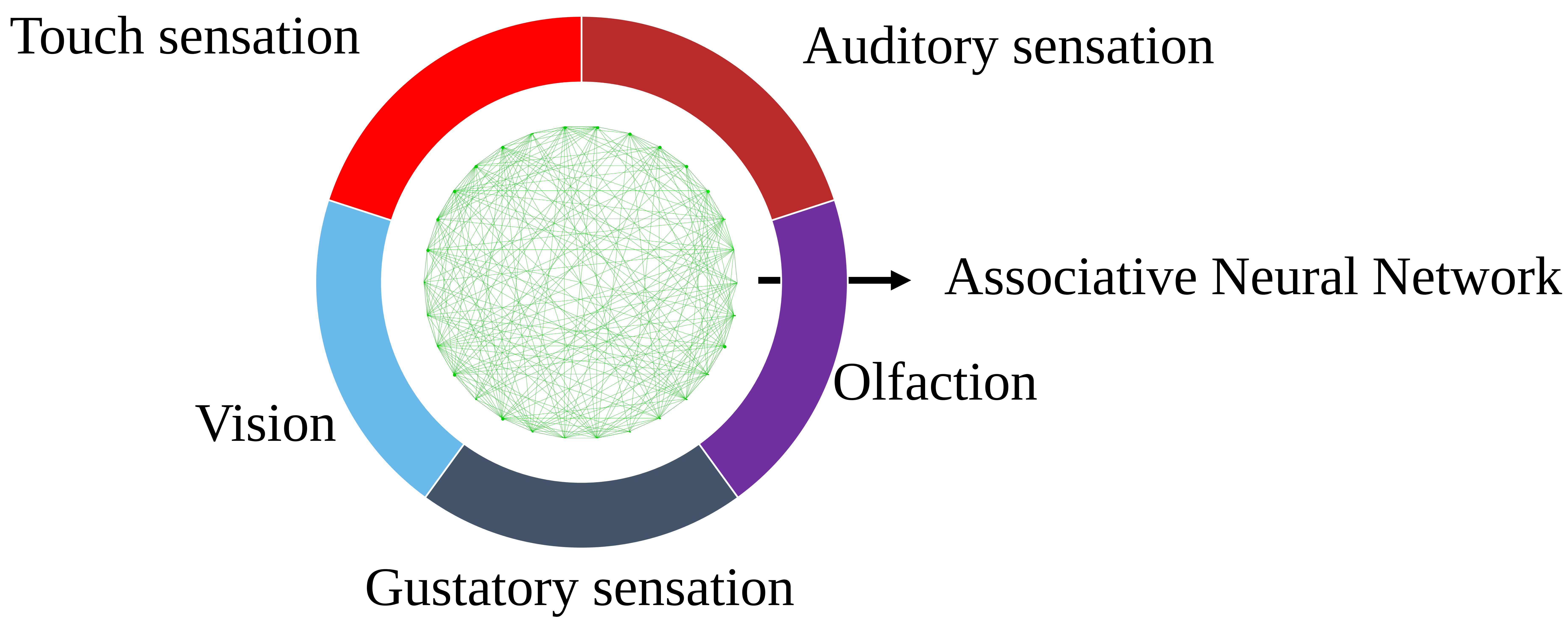
- Different sensory signals are processed in different regions;

Cluster Neuromorphic Computing Architecture



[**H. An**. et al. "Opportunities and challenges on nanoscale 3D neuromorphic computing system,", 2017.]
[**H. An**. et al. "The Roadmap to Realize Memristive Three-dimensional Neuromorphic Computing System,", 2018.]

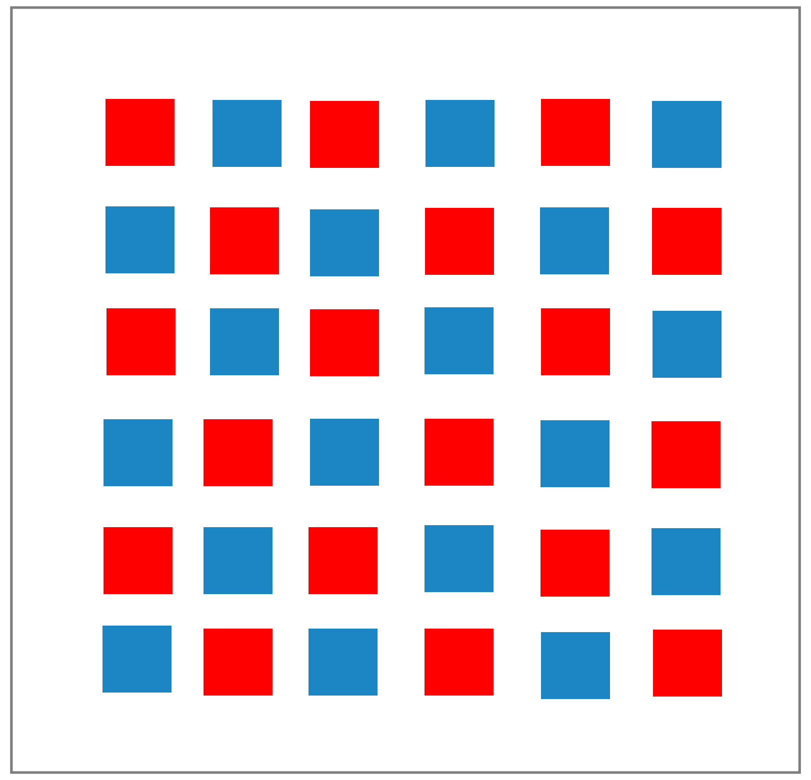
Emerging Neuromorphic Computing Architectures: Associative Neuromorphic Computing Architecture



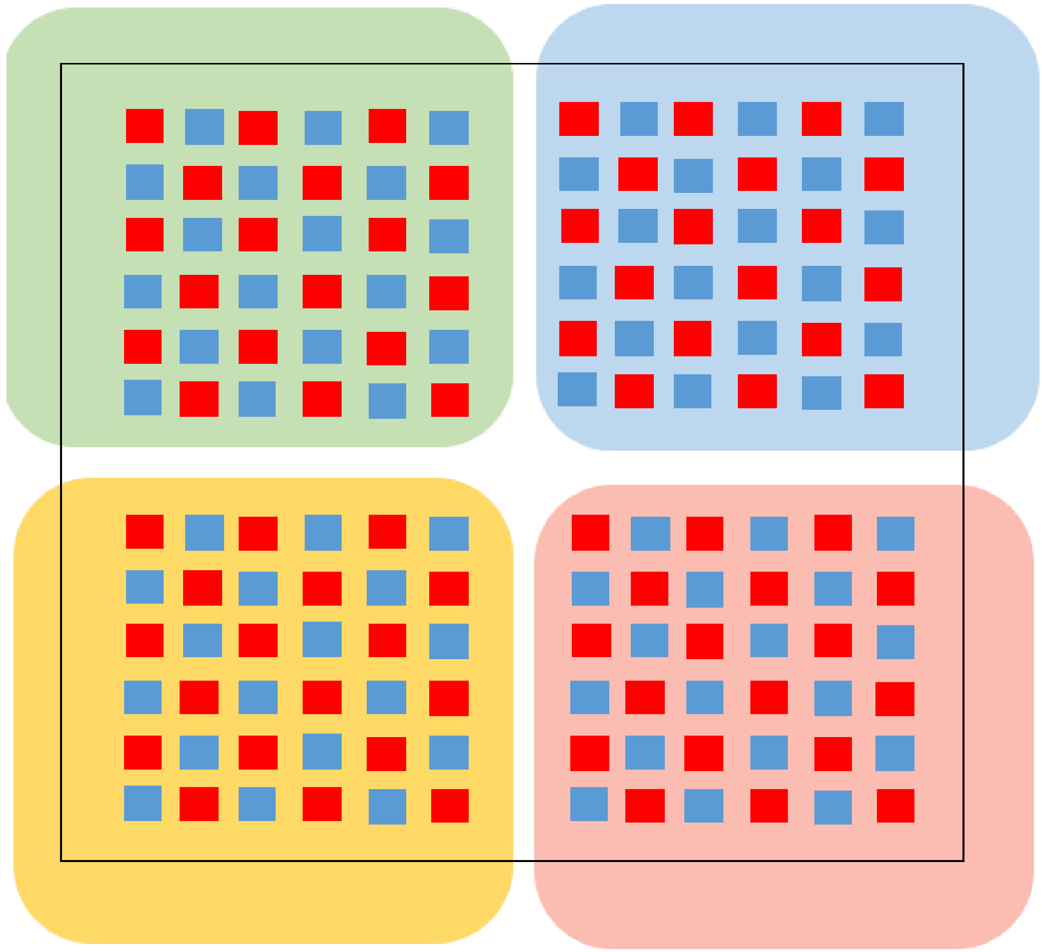
Associative Neuromorphic Computing Architecture

[**H. An**. et al. "Opportunities and challenges on nanoscale 3D neuromorphic computing system,", 2017.]
[**H. An**. et al. "The Roadmap to Realize Memristive Three-dimensional Neuromorphic Computing System,", 2018.]

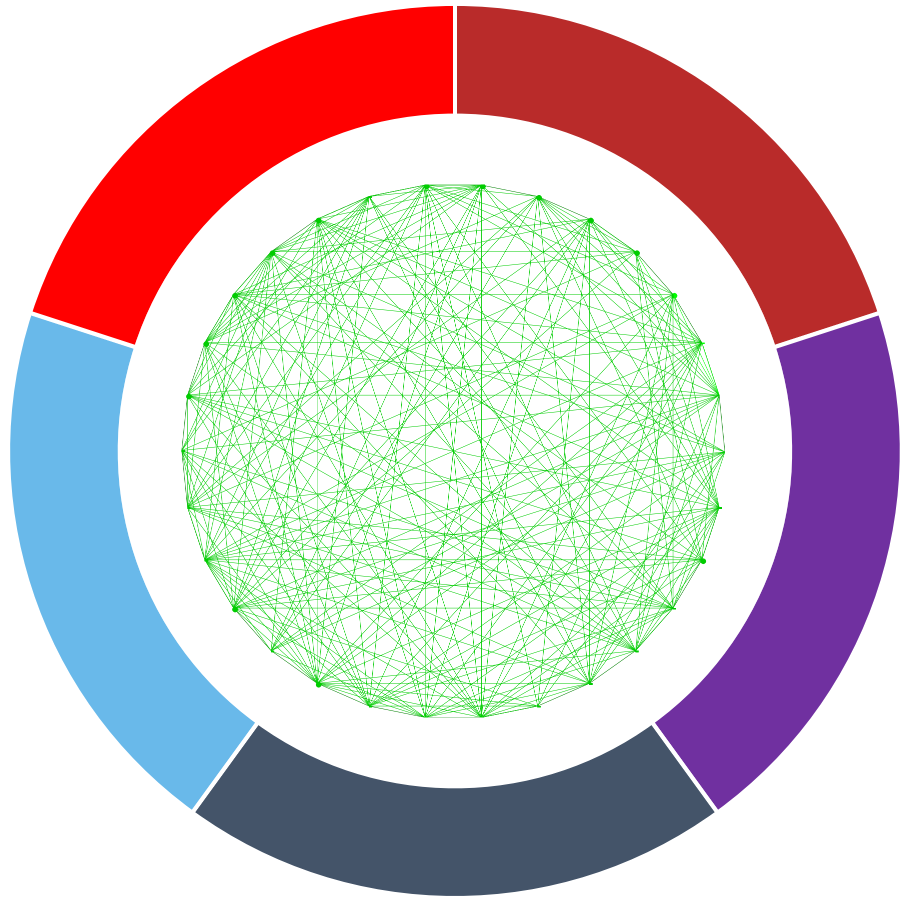
Hardware Implementation: Nonlinear Computing Module



Distributed Neuromorphic Computing Architecture



Cluster Neuromorphic Computing Architecture



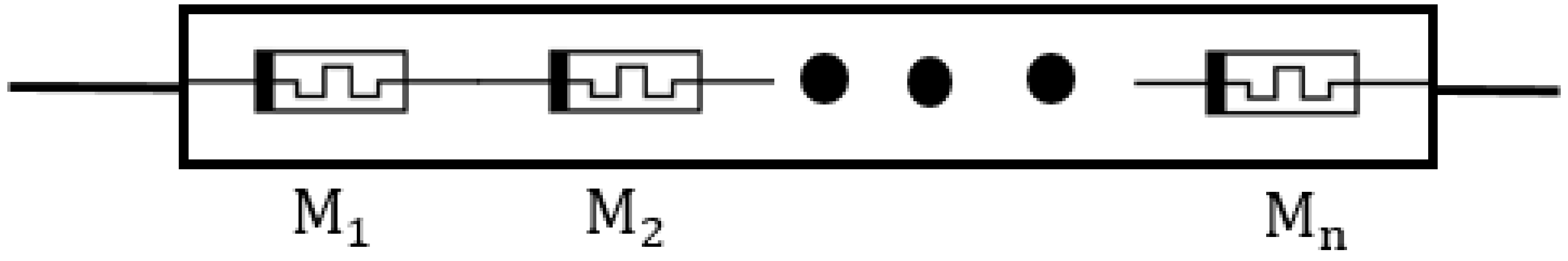
Associative Neuromorphic Computing Architecture

Analog Signals + Nonlinear Computing Module + Matrix Computing Module + Training Module

Physical Implementations

[**H. An.** et al. "Opportunities and challenges on nanoscale 3D neuromorphic computing system,", 2017.]
[**H. An.** et al. "The Roadmap to Realize Memristive Three-dimensional Neuromorphic Computing System,", 2018.]

Memristor-based Nonlinear Computing Module



$$R^{(1)} = R_1^H + R_2^H + R_3^H + \ldots + R_i^H + \ldots R_n^H$$

$$R^{(2)} = \boxed{R_1^L} + R_2^H + R_3^H + \ldots + R_i^H + \ldots R_n^H$$

$$R^{(3)} = \boxed{R_1^L + R_2^L} + R_3^H + \ldots + R_i^H + \ldots R_n^H$$

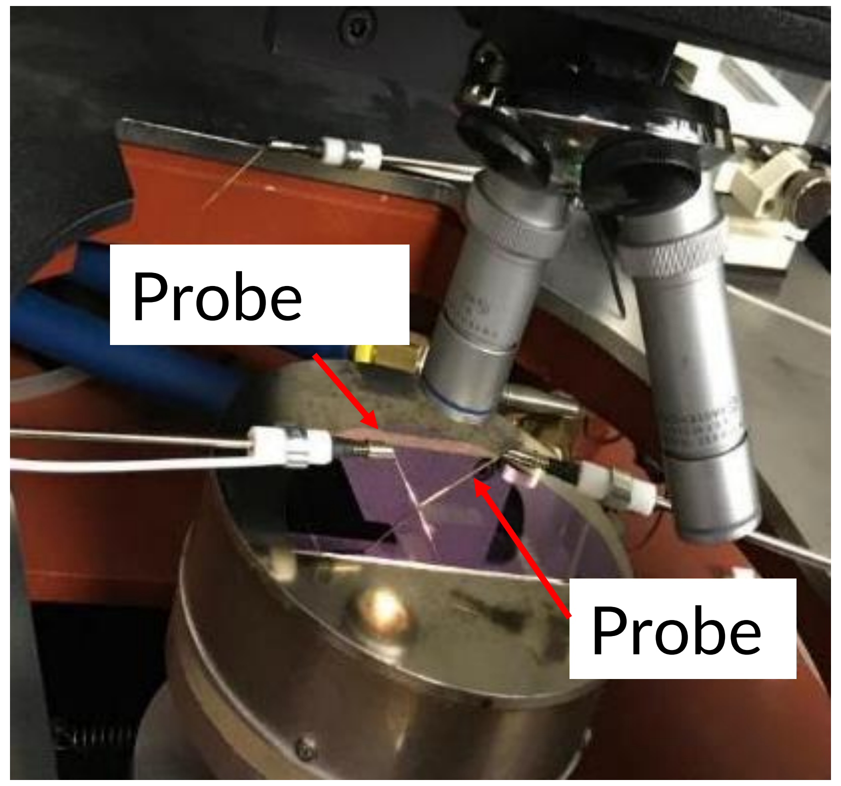
$$R^{(i+1)} = \boxed{R_1^L + R_2^L + \ldots + R_i^L +} \ldots R_n^L$$

R_H is the high resistance value of each memristor;
R_L is the low resistance value of each memristor;

- Prerequisites:
  - The cascaded memristors can switches
  - The controllable set voltage
  - The controllable high resistance state/low resistance state (HRS/LRS)

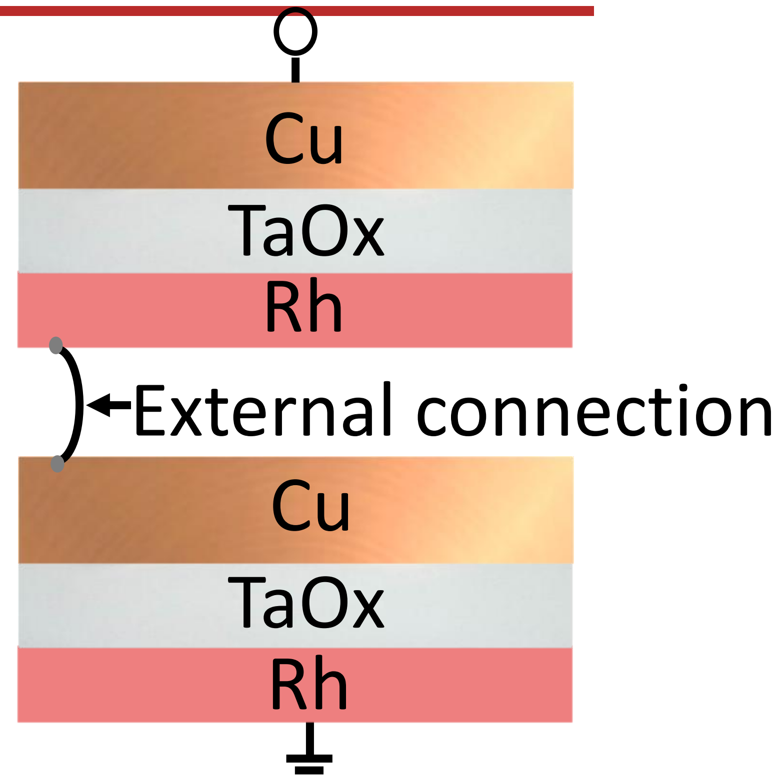
The switching Behavior Investigation of the Cascaded Memristors

Keithley 4200-SCS Semiconductor Parameter Analyzer

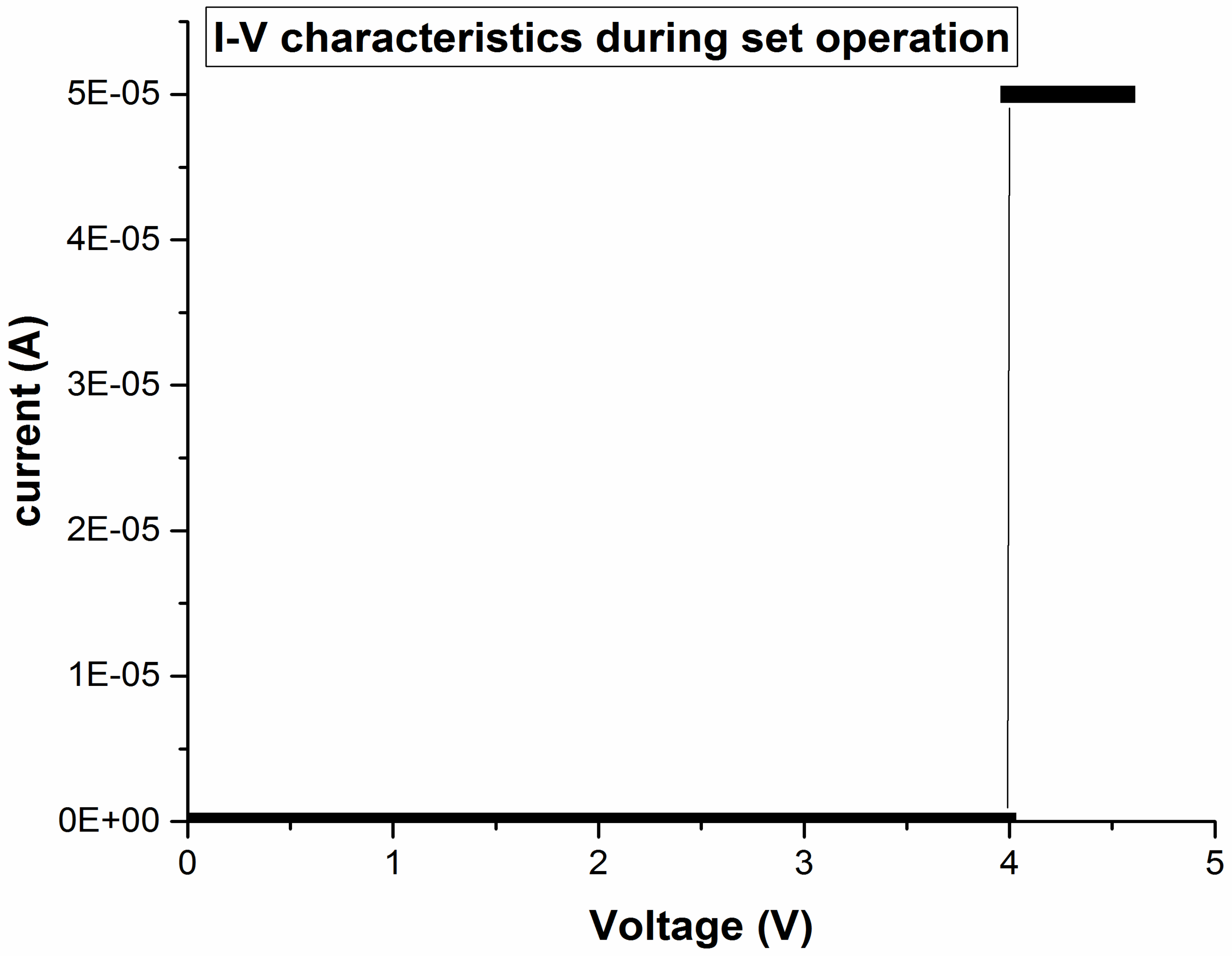


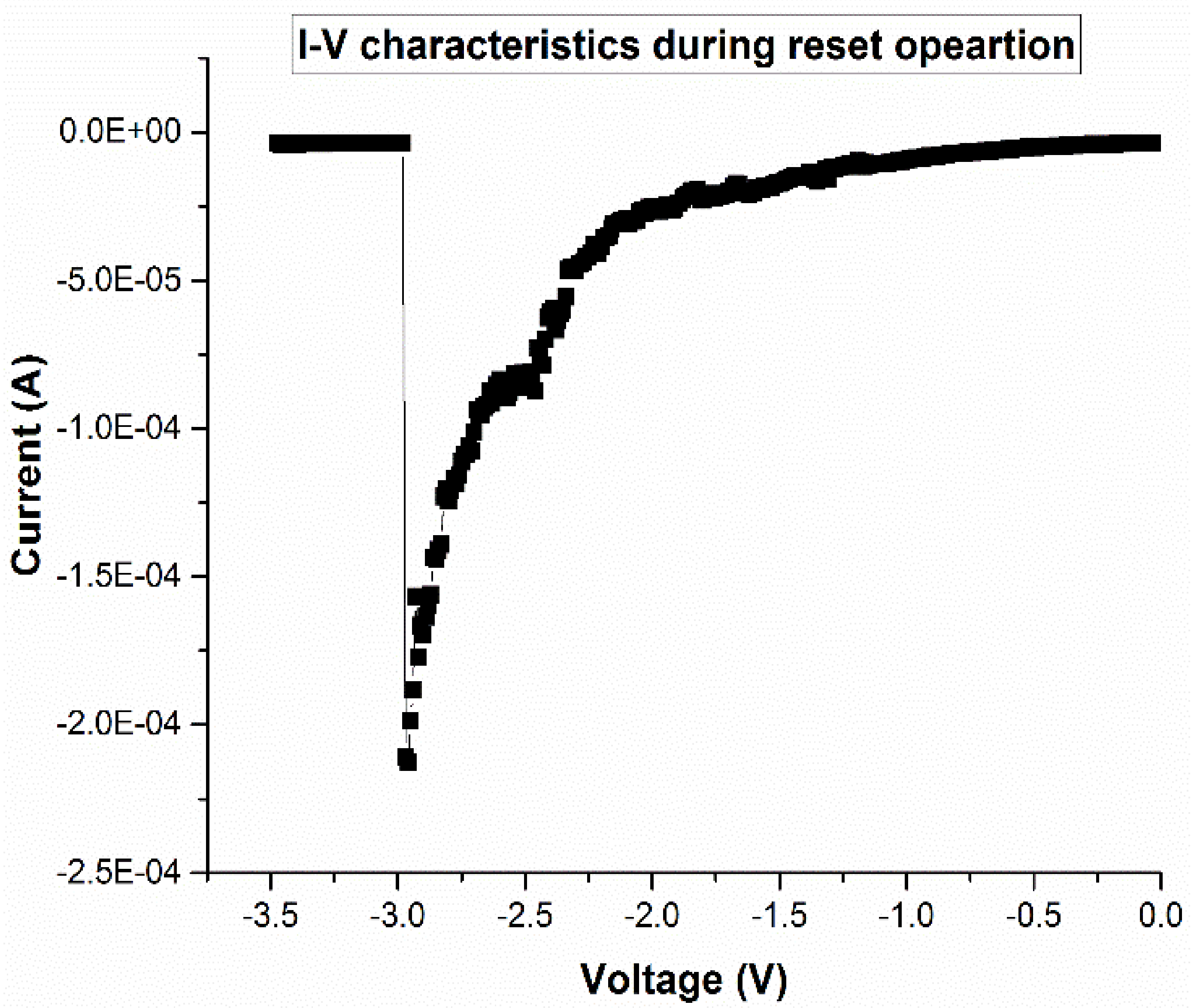
Microscope view

Memristor cell



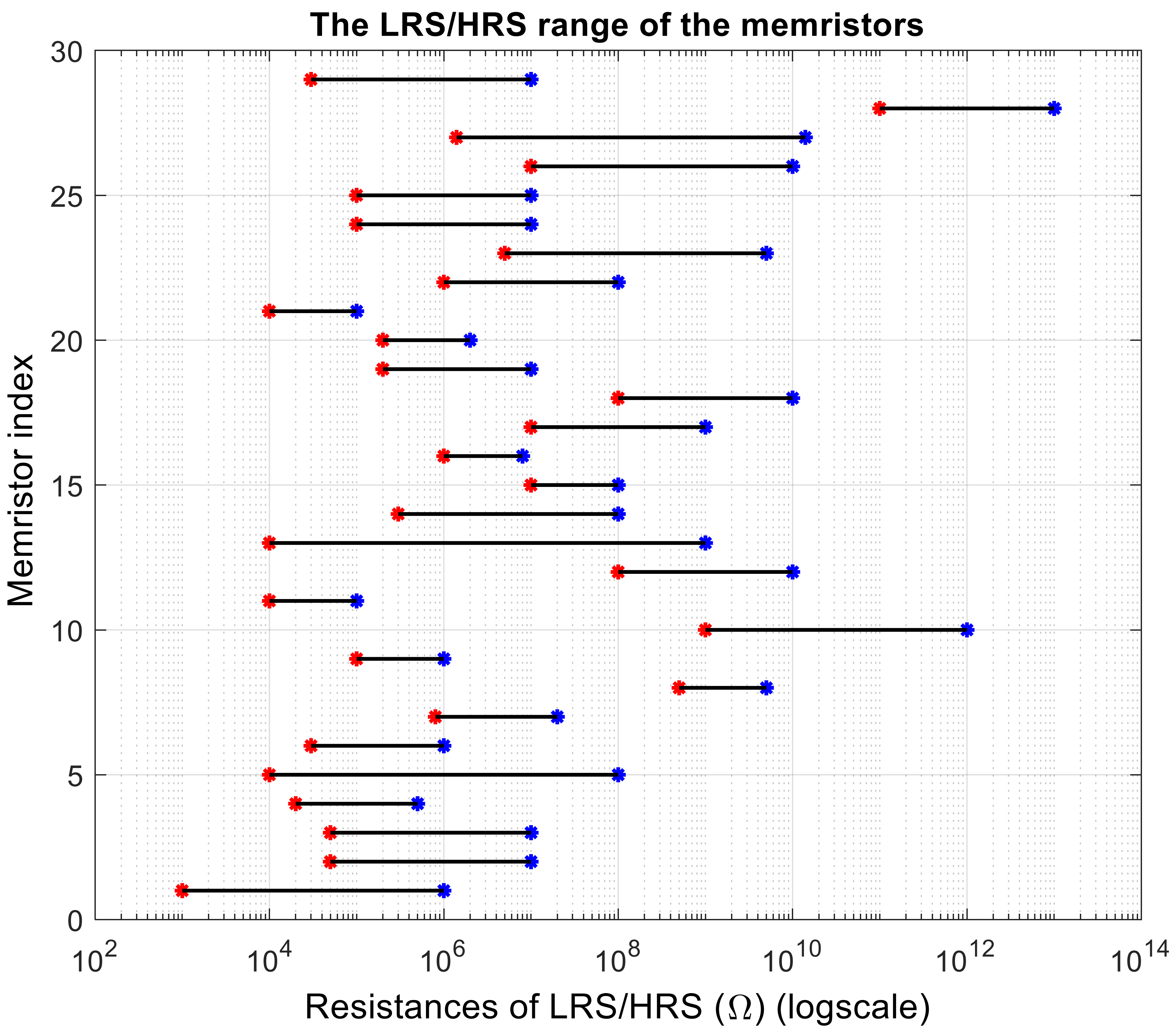
Cu (150nm)/TaOx(25 nm)/Rh(50 nm)



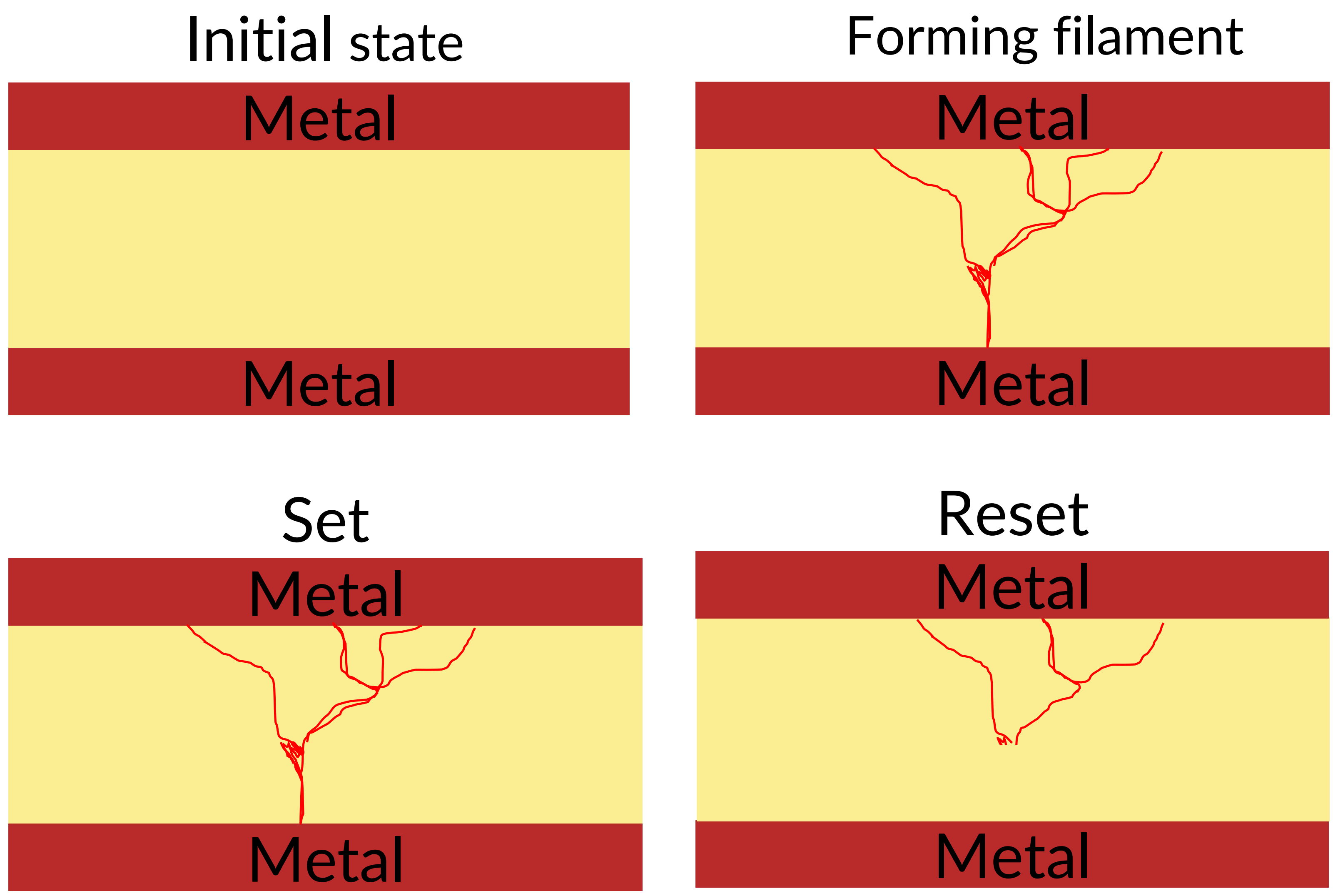


H. An.

High Resistance State & Low Resistance State Distribution



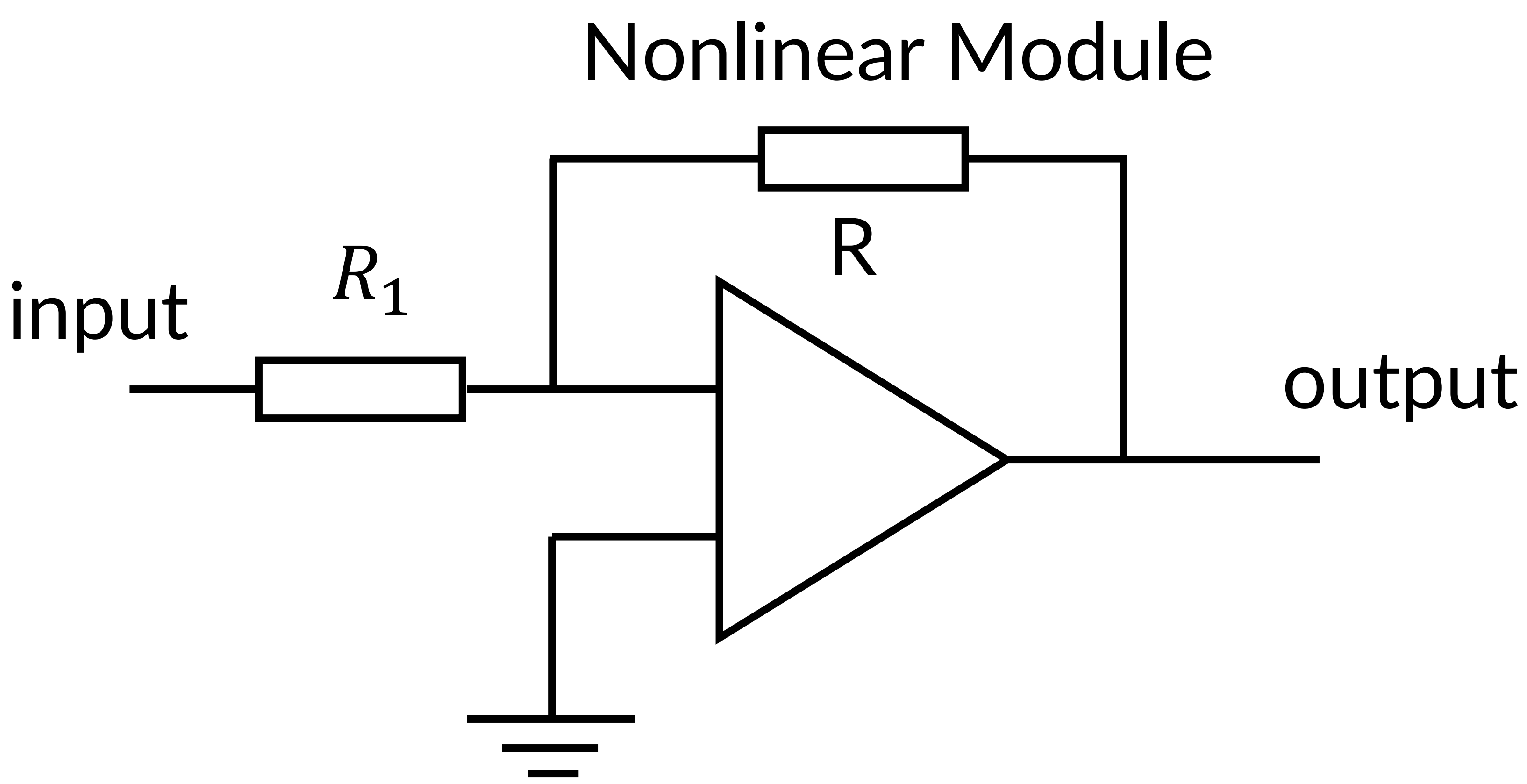
https://nano.stanford.edu/stanford-memory-trends

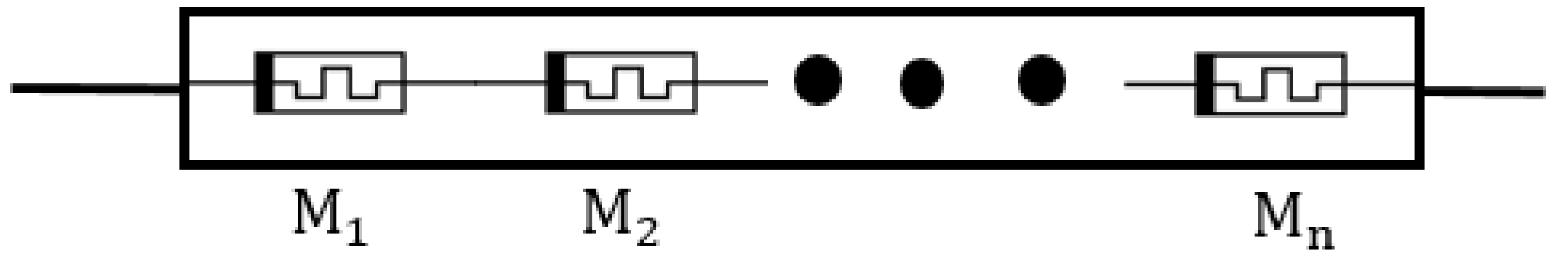


The set/reset voltage and HRS/LRS are determined by

- Materials;
- Physical geometry;
- Temperature;

The Mathematical Model of the Cascade Memristor-based Nonlinear Computing Module





Nonlinear Module

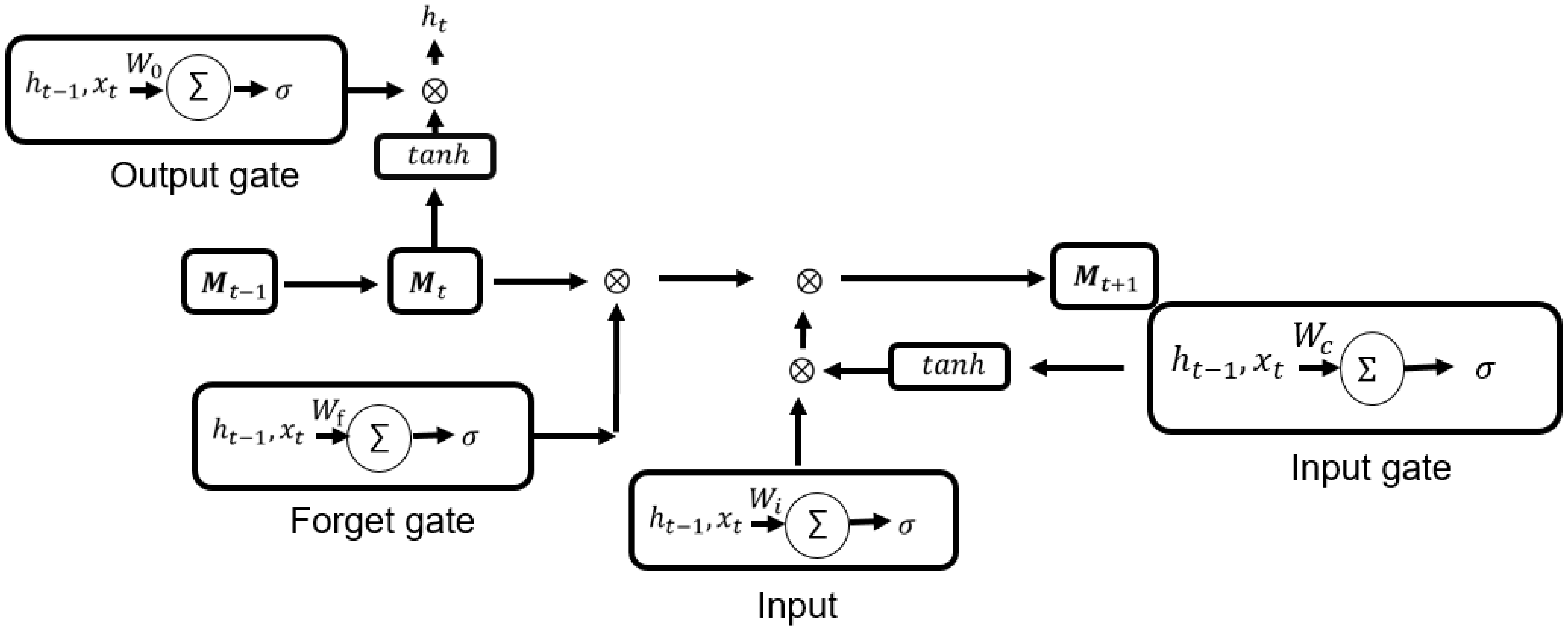
$$\begin{cases} R = (n-k)R_{\mathrm{H}} + kR_{\mathrm{L}} \\ I = k \times \Delta I_{th} \end{cases}$$

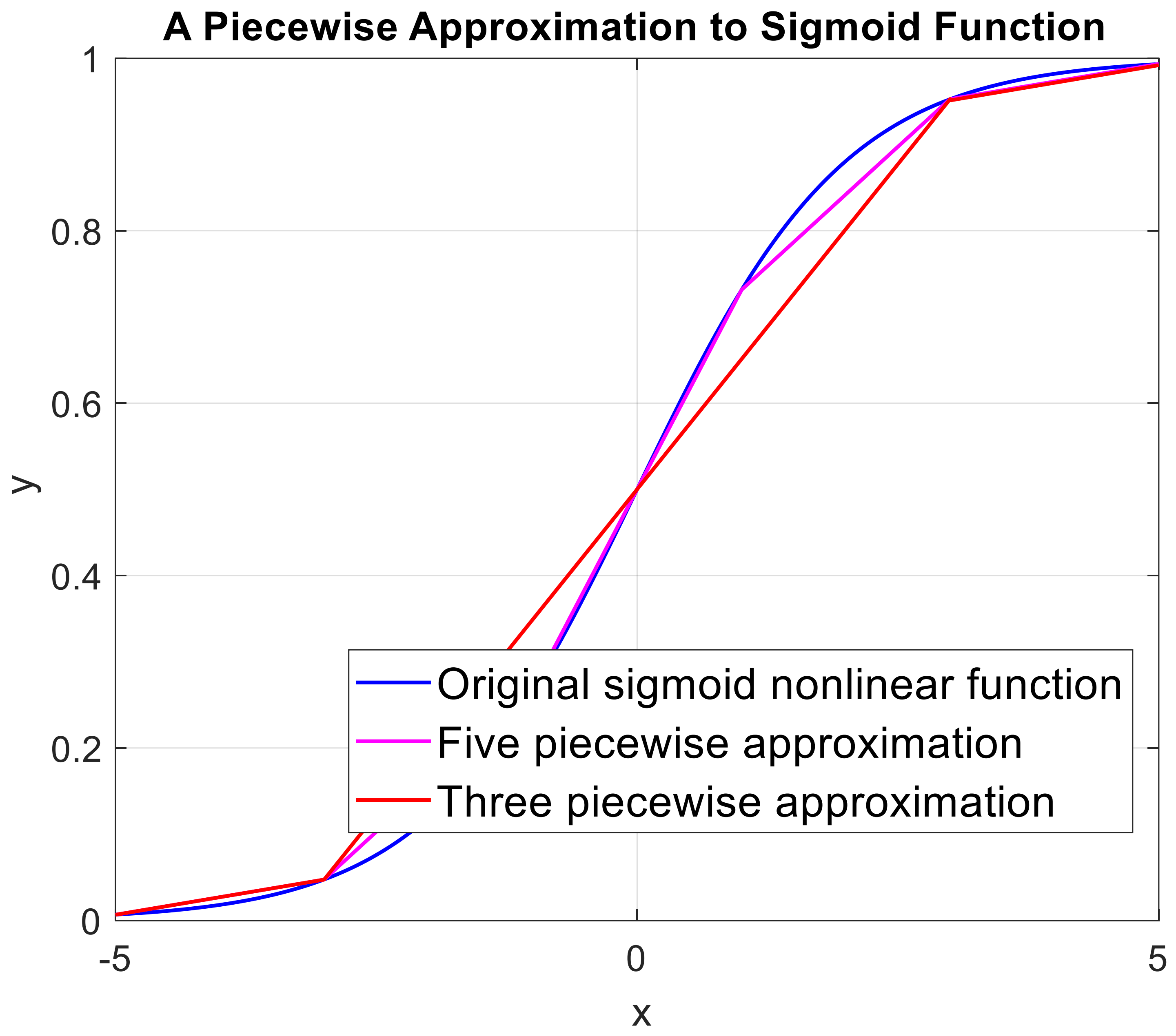
$$R = \left(n - \frac{I}{\Delta I_{th}}\right) R_{\mathrm{H}} + \frac{I}{\Delta I_{th}} R_{\mathrm{L}}$$

$$V_{out} = -\left(\frac{R}{R_1}\right) V_{in}$$

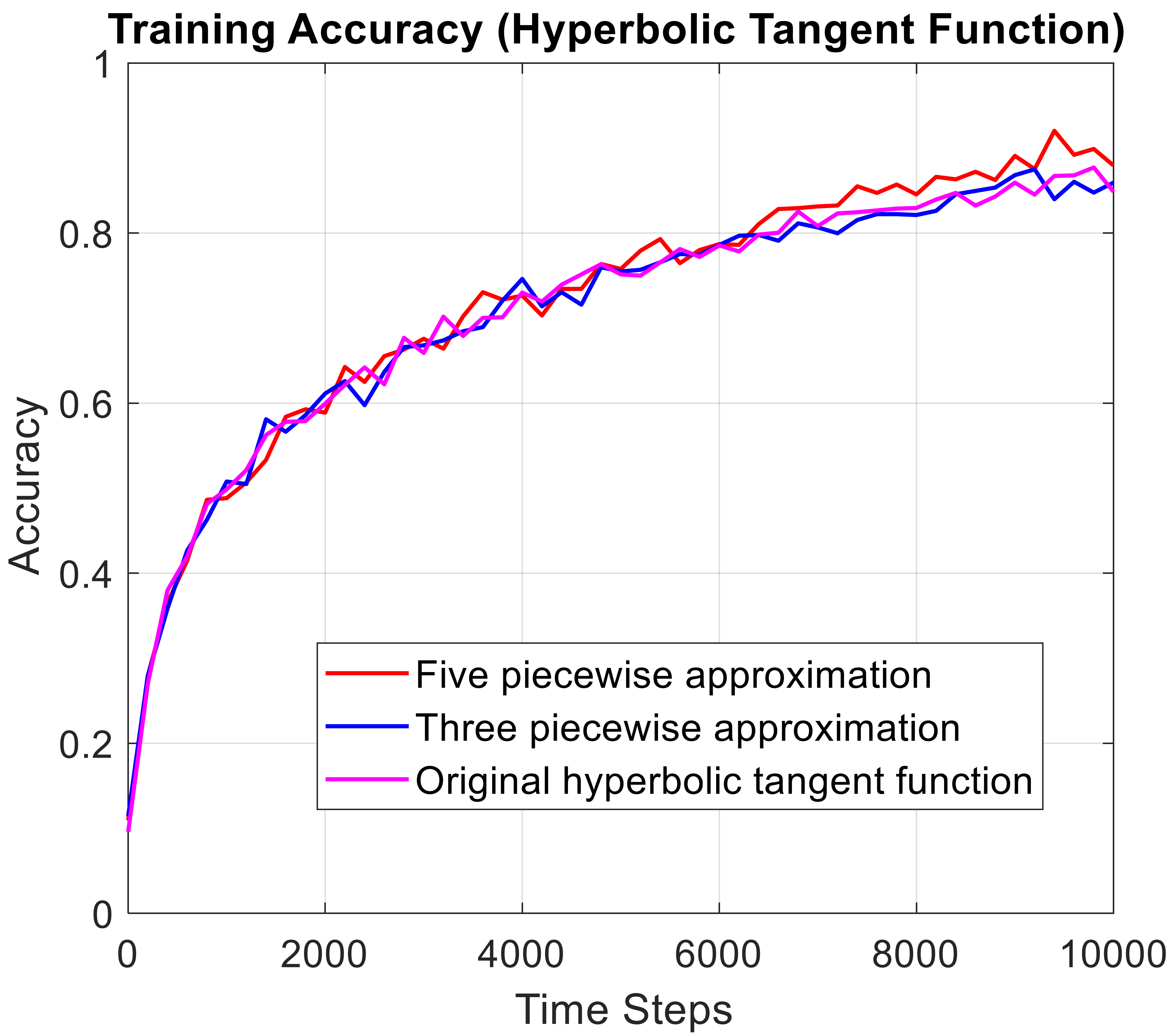
Where
R is the total resistance of nonlinear module;
R_H is the high resistance value of each memristor;
R_L is the low resistance value of each memristor;
n is the total number of memristors in the module;
k is the step index whose value is from 0 to n;
ΔI_{th} is an interval of threshold current values between two consecutive memristor ($I_{th_k} - I_{th_{k-1}} = \Delta I_{th}$), where I_{th_k} is the threshold value of *k*th memristor.

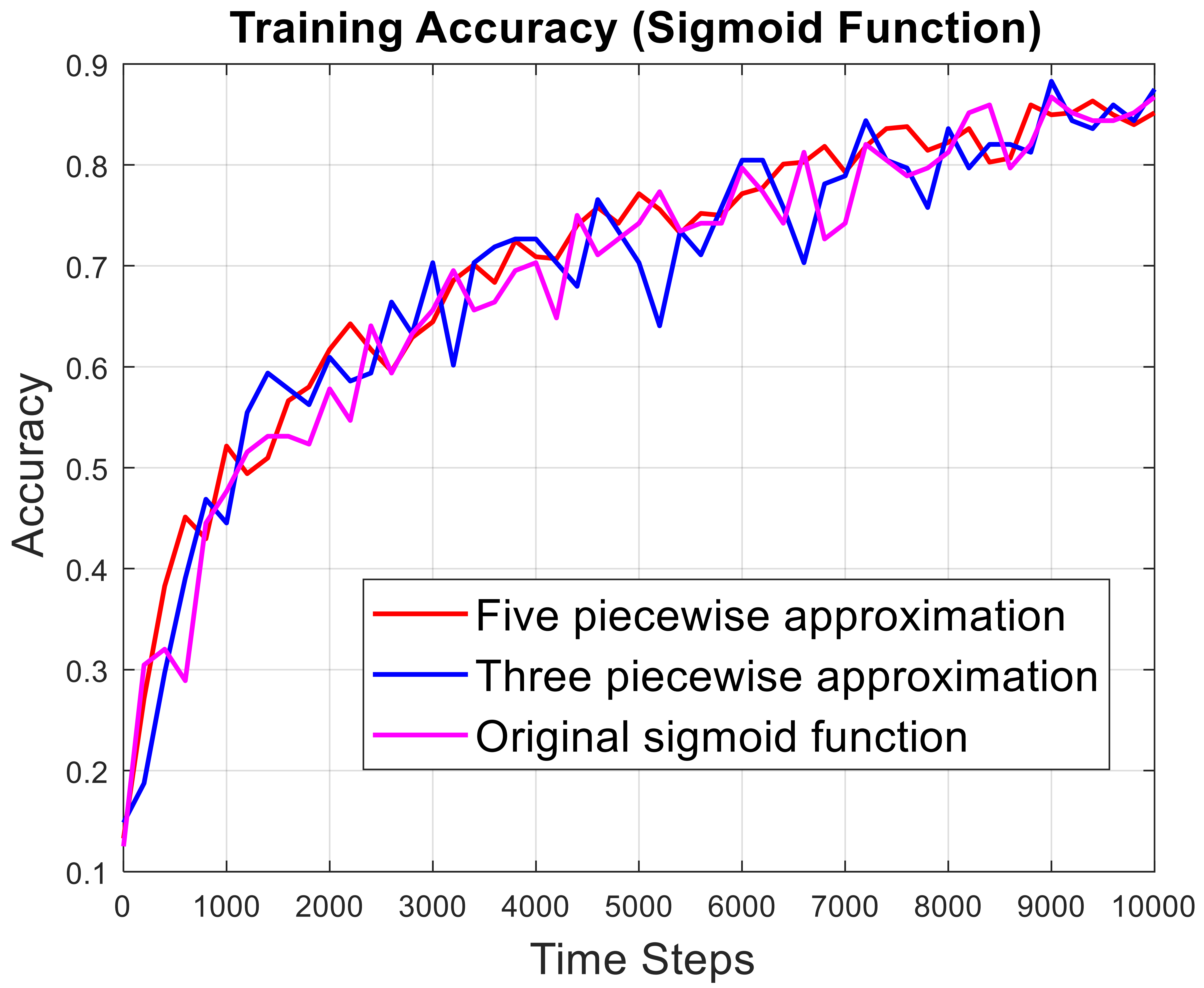
Application to Digit Recognition with Long-short Term Memory





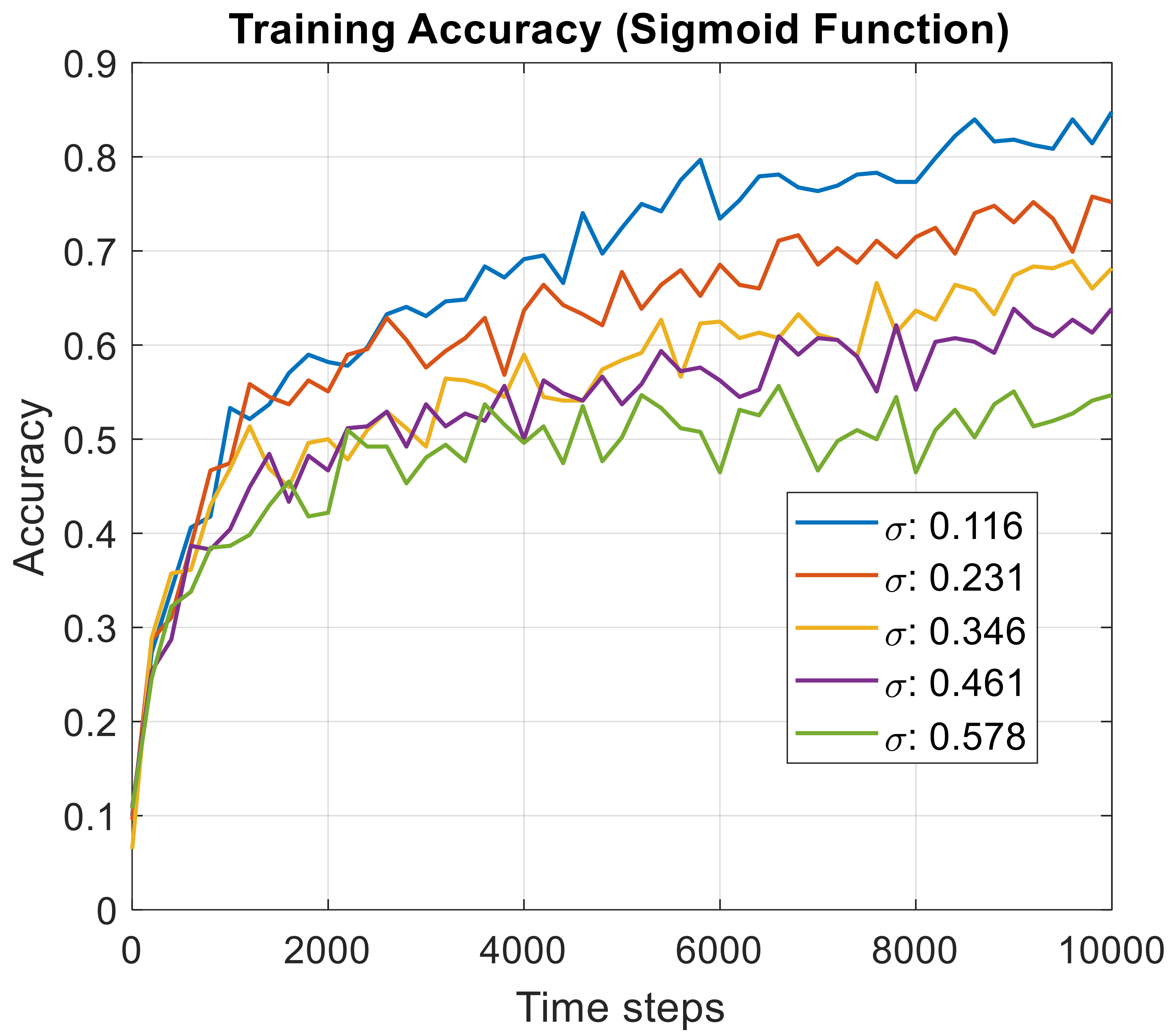
Training Accuracy of LSTM with Piecewise Approximation

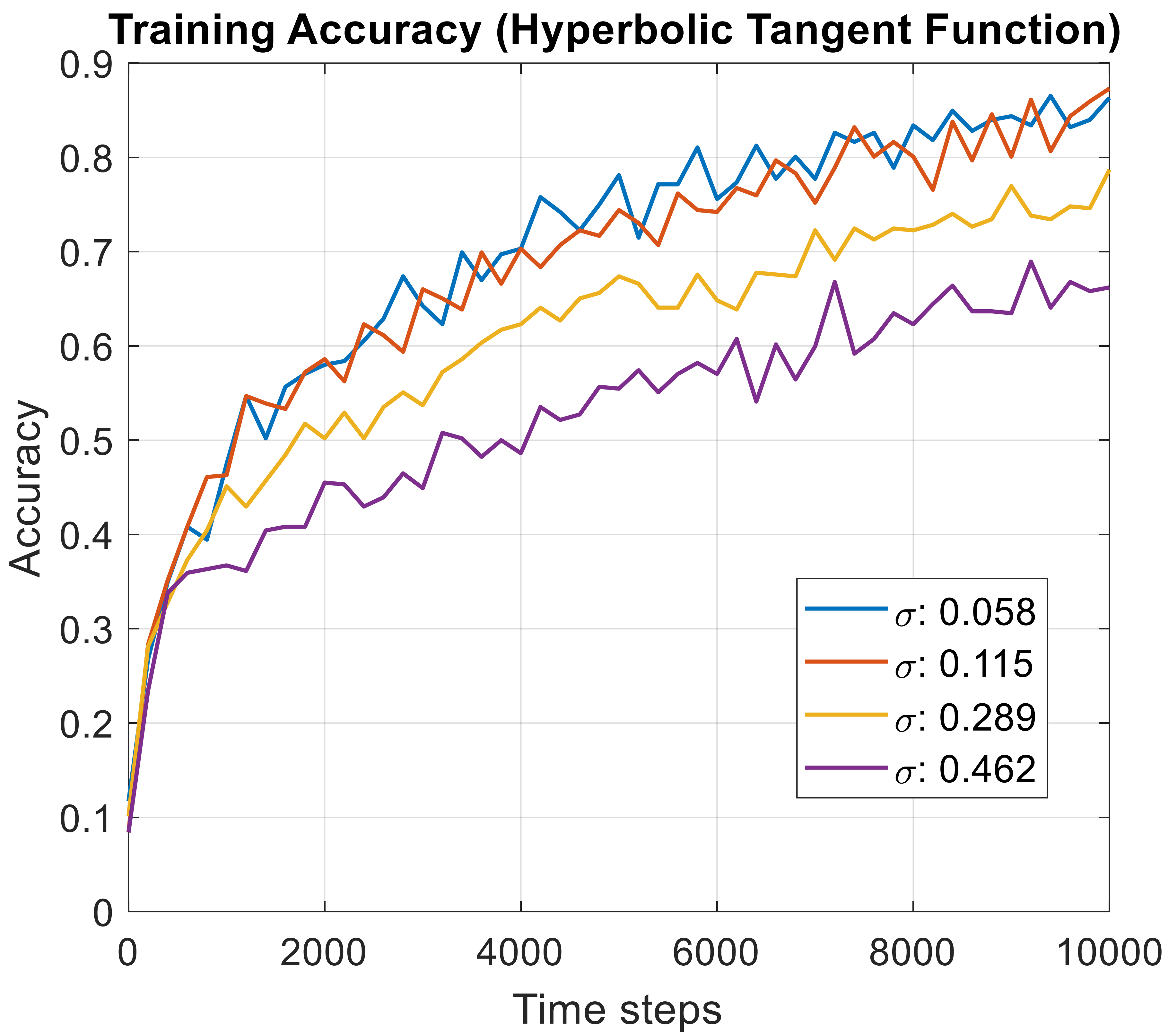




- The training accuracies do not degrade by replacing the nonlinear function with piecewise approximation

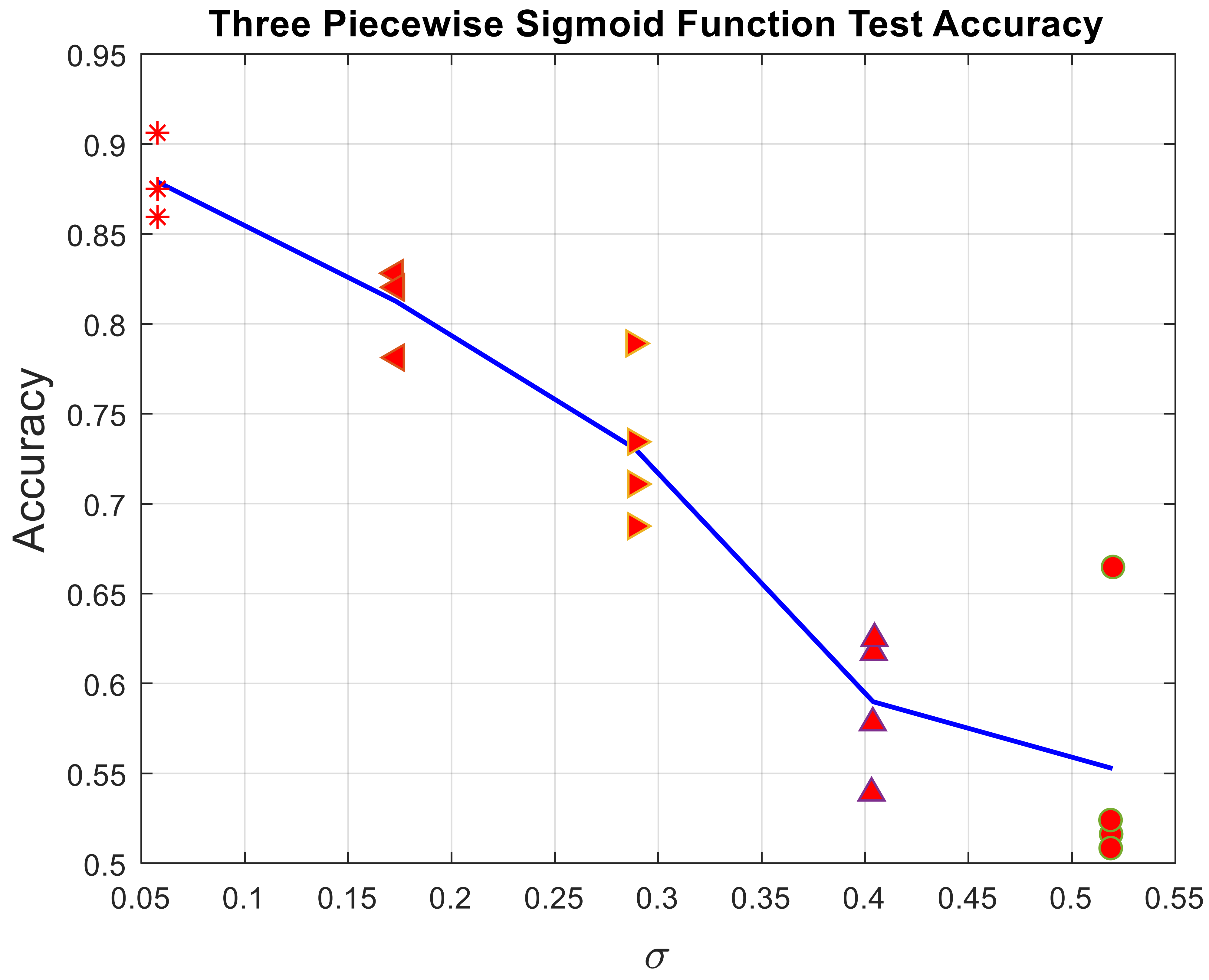
Training Accuracy with Switching Resistance Variations of Memristor

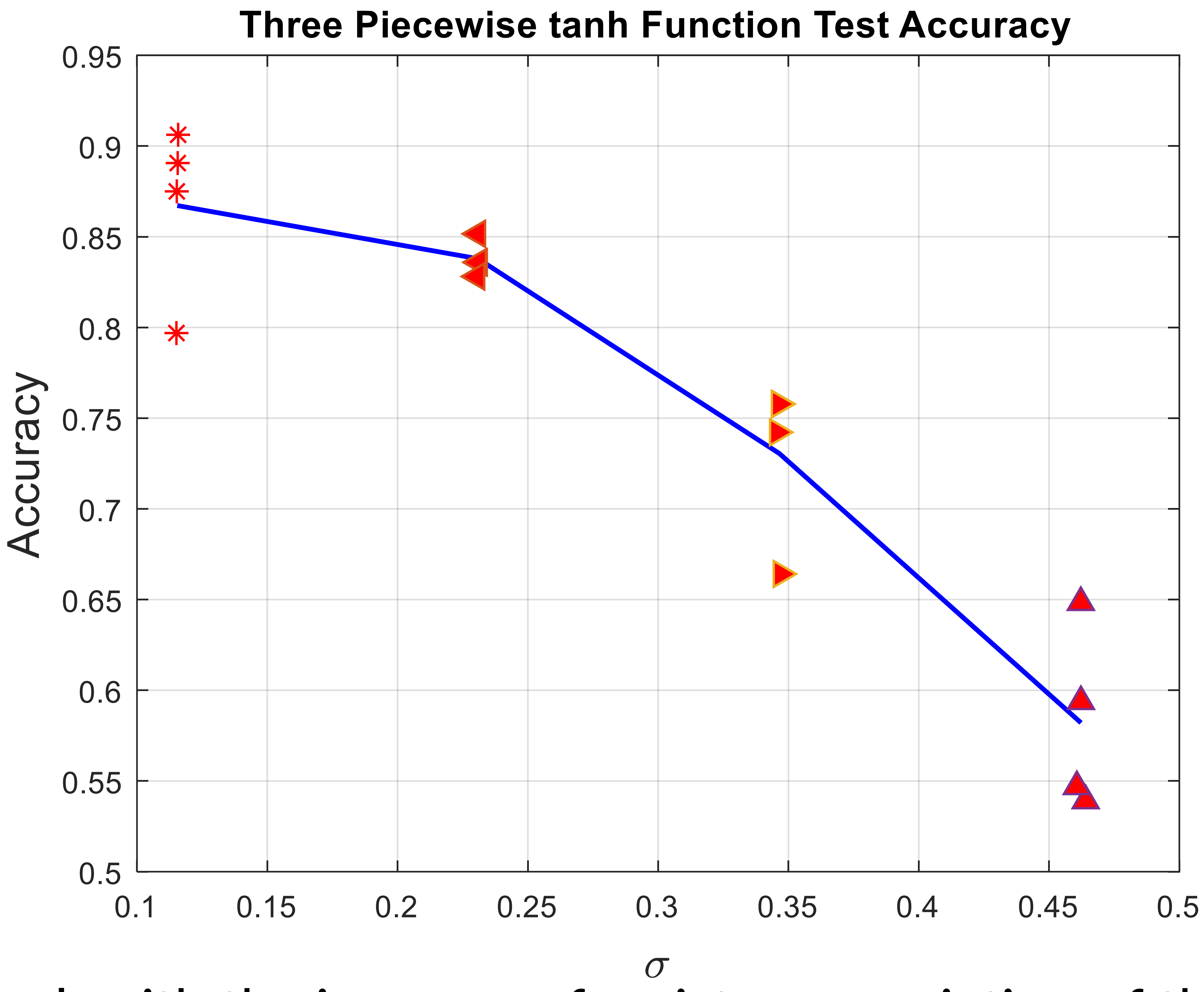




The testing accuracies decrease by the impact of the large resistance switching variation of memristor.

Test Accuracy with Switching Resistance Variations of Memristor





The testing accuracies decrease almost proportional with the increase of resistance variation of the memristor.

Conclusions

- Introduce three emerging neuromorphic architectures: Distributed Neuromorphic computing architecture; Cluster Neuromorphic Neuromorphic Computing Architecture; Associative Neuromorphic Computing Architecture
- We designed and evaluated a memristor-based nonlinear computing module in the Long Short-term Memory with the application on digit number recognition
- The training accuracy would not be degraded by using the proposed nonlinear computing module with ideal memristor
- The large resistance switching variation of memristor would significantly reduce the learning and testing accuracy, and the accuracies decrease is almost proportional to the increase of resistance variation of the memristor

Acknowledgement

Dr. Yang (Cindy) Yi
Assistant Professor

Virginia Tech
cindy_yangyi@vt.edu

Mohammad Shah Al-Mamun
Ph.D. Candidate

Virginia Tech
samamun@vt.edu

Dr. Marius K. Orlowski
Professor

Virginia Tech
marius@vt.edu

References

- H. An, Z. Zhou, and Y. Yi, "Opportunities and challenges on nanoscale 3D neuromorphic computing system," in *Electromagnetic Compatibility & Signal/Power Integrity (EMCSI), 2017 IEEE International Symposium on, 2017, pp. 416-421*.
- H. An, K. Bai, and Y. Yi, "The Roadmap to Realize Memristive Three-dimensional Neuromorphic Computing System," chapter in *"Memristive Neural Networks,"* Intech publishing in 2018 (ISBN 978-953-51-6803-4). In press
- Kandel, Eric R, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and AJ Hudspeth. *Principles of Neural Science. Vol. 4: McGraw-hill New York, 2000.*

Q & A

Hongyu An
Ph.D. Candidate

Virginia Tech
hongyu51@vt.edu