Physical systems for thermodynamic computing (Some examples)

J. Joshua Yang

The Department of Electrical and Computer Engineering

University of Massachusetts, Amherst

outline

1. Memristor introduction

2. Simple computing without thermodynamics
   (supervised learning)

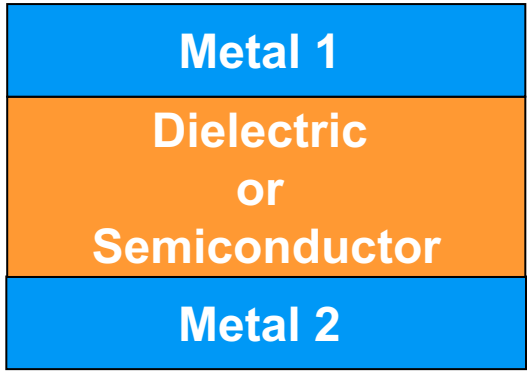
3. Advanced computing with thermodynamics
   (unsupervised learning)

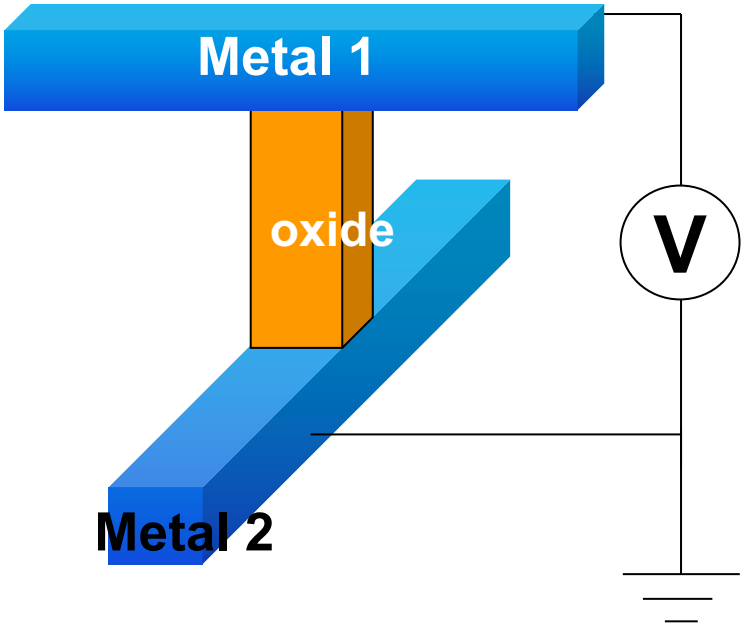
4. Other physical systems

Memristor introduction

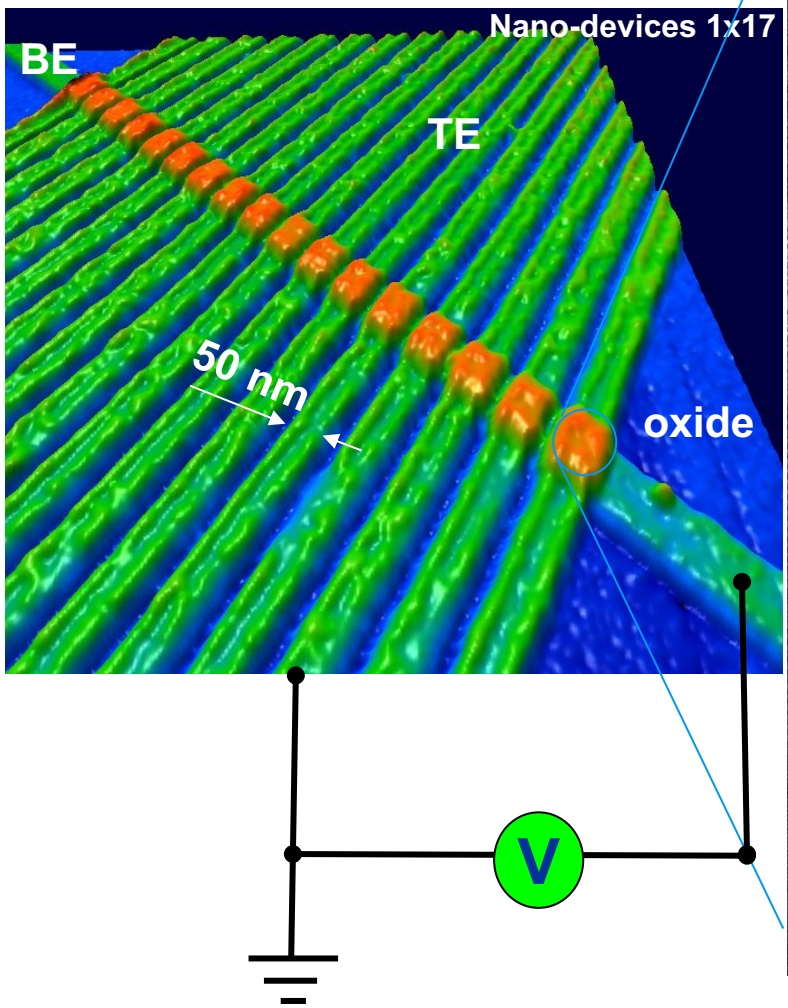
Simple but complicated materials and device

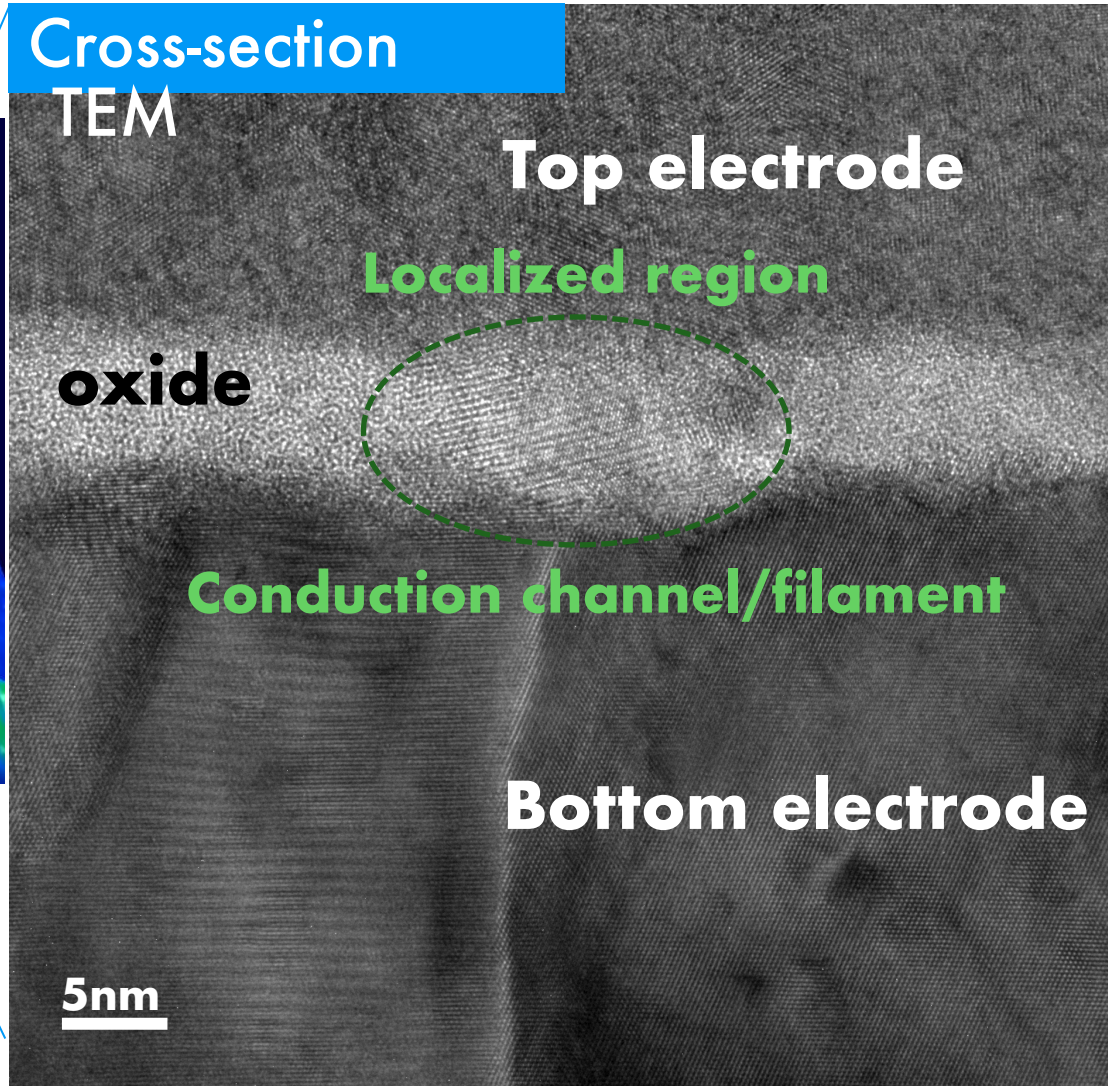
Complicated physics, chemistry, materials, i.e., thermodynamics, issues





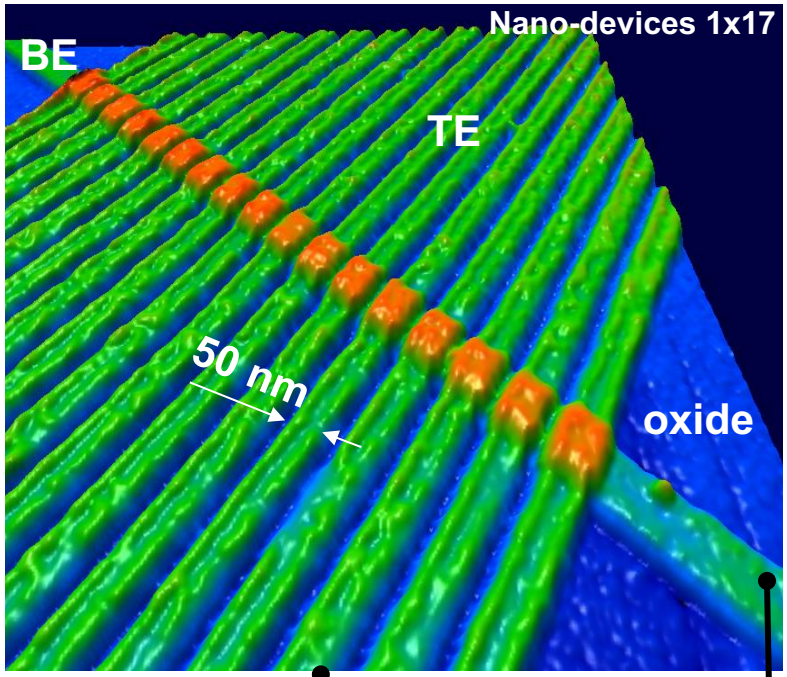
The first oxide nanodevice

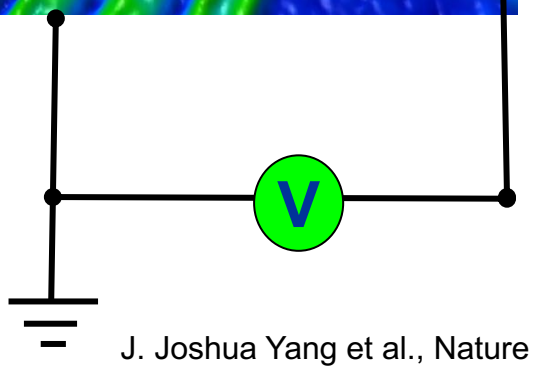


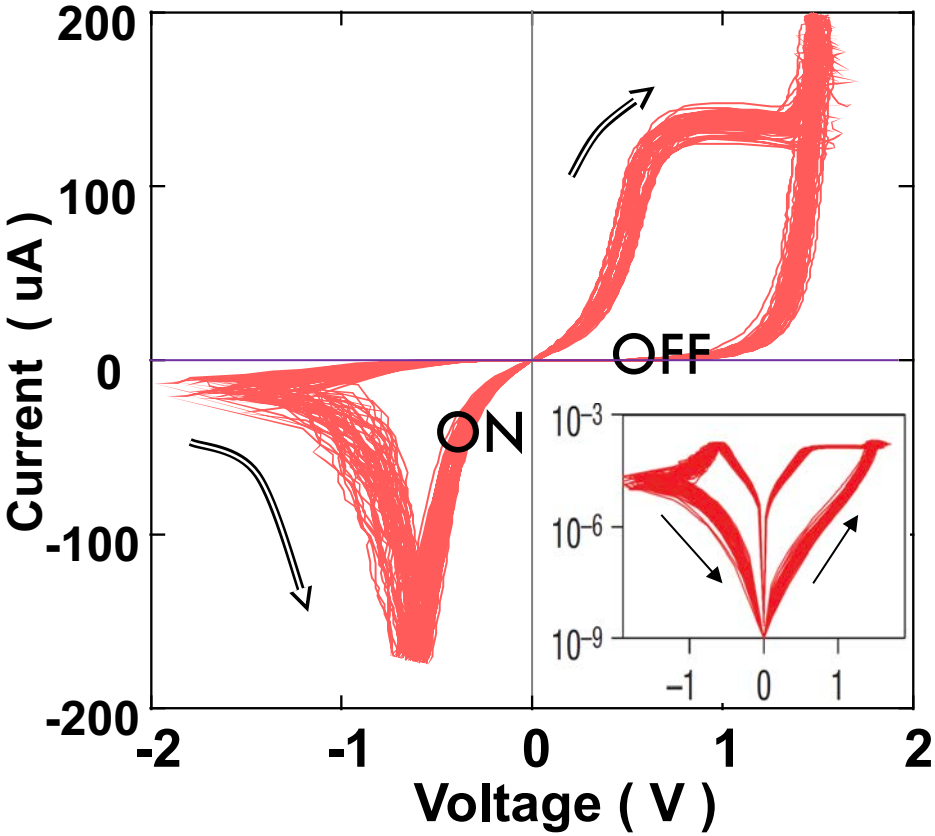


J. Joshua Yang et al., Nature Nanotechnology **3**, 429 (2008)

The signature of memristive devices







J. Joshua Yang et al., Nature Nanotechnol. **3**, 429 (2008)

Promises and Challenges

Promises

1. *Speed (85ps)*
2. *Scalability (2nm)*
3. *Multilevel (>64)*
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

Challenges

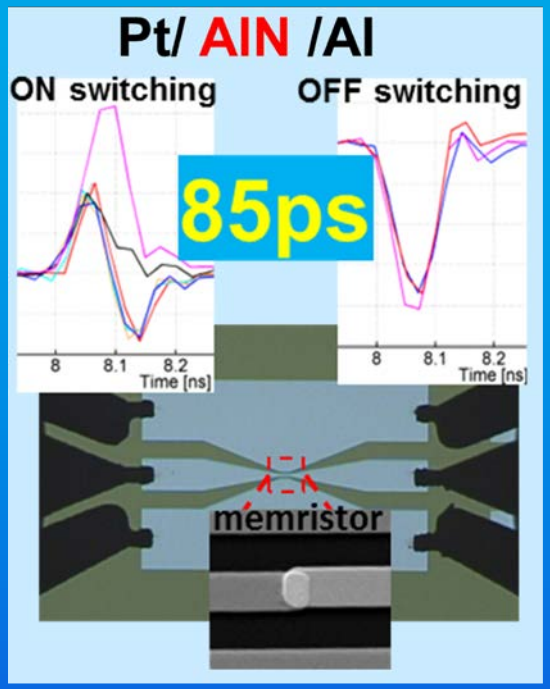
- ***Mechanism***
- ***Access device in large array***
  (Transistor, selector)
- ***Desirable dynamics***
  (Diffusive memristor)

Promises

Promises

1. **Speed (85ps)**
2. *Scalability (2nm)*
3. *Multilevel (>64)*
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

85ps ON/OFF switching



B. J. Choi, et al, *Adv. Funct. Mater. 26, 5290*, (2016).

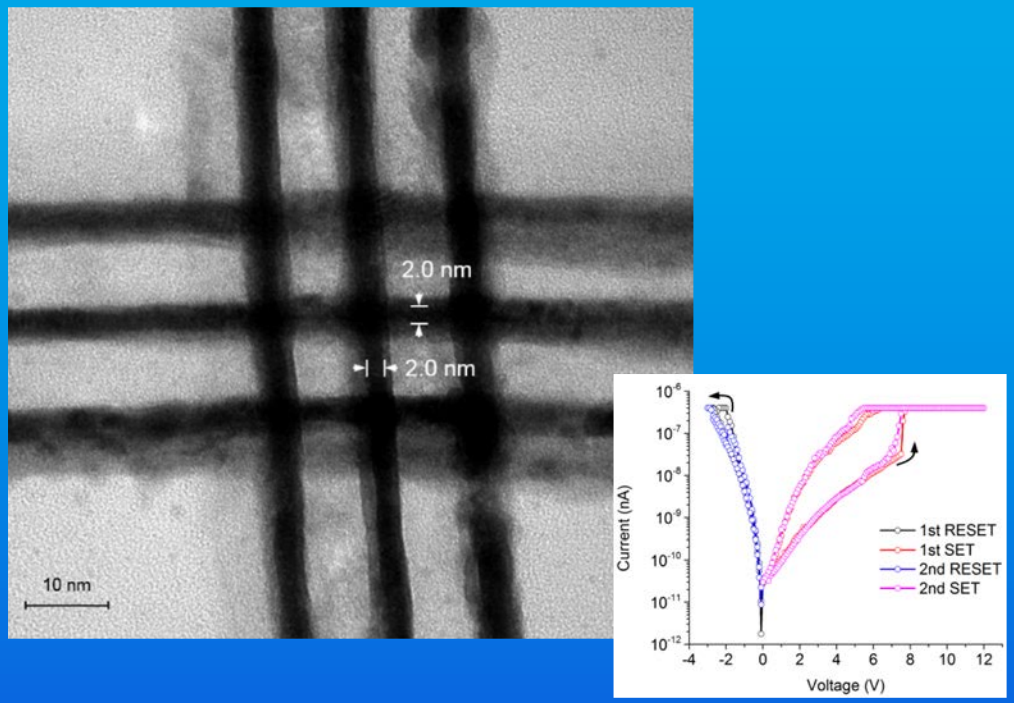
T., Antonio C., John Paul Strachan, G. Medeiros-Ribeiro, and R. Stanley Williams. "Sub-nanosecond switching of a tantalum oxide memristor." *Nanotechnology* 22, no. 48 (2011): 485203.

Promises

Promises

1. *Speed (85ps)*
2. **Scalability (2nm)**
3. *Multilevel (>64)*
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

2nm x 2nm memristor crossbar



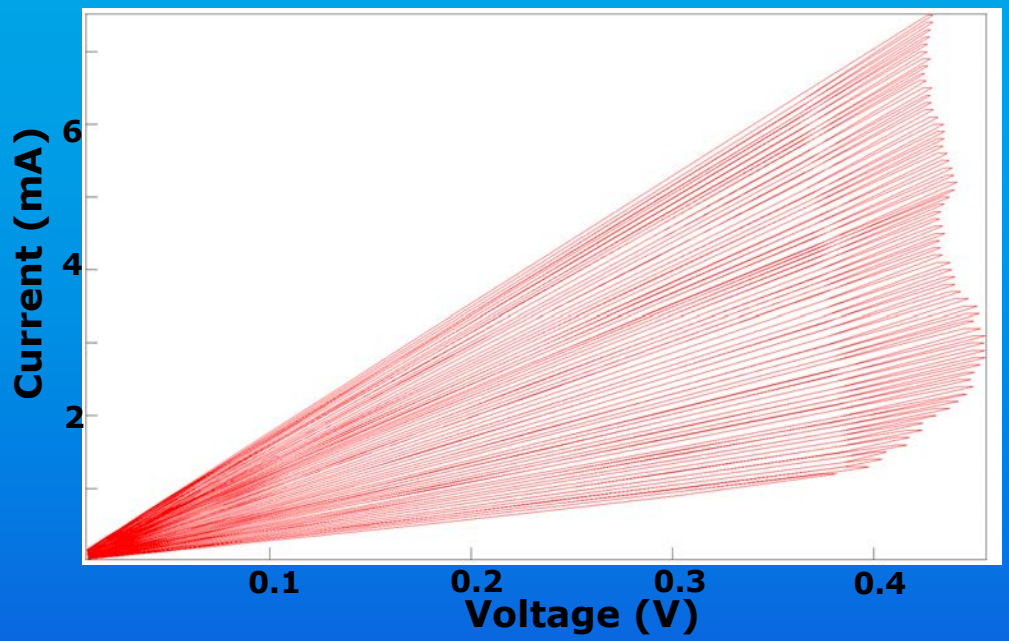
S. Pi et al., *Nature Nanotechnology* ***14****, 35* (2018). (Q. Xia)

Promises

Promises

1. *Speed (85ps)*
2. *Scalability (2nm)*
3. **Multilevel (>64)**
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

64 resistance levels



N. Ge et al., unpublished (2018).

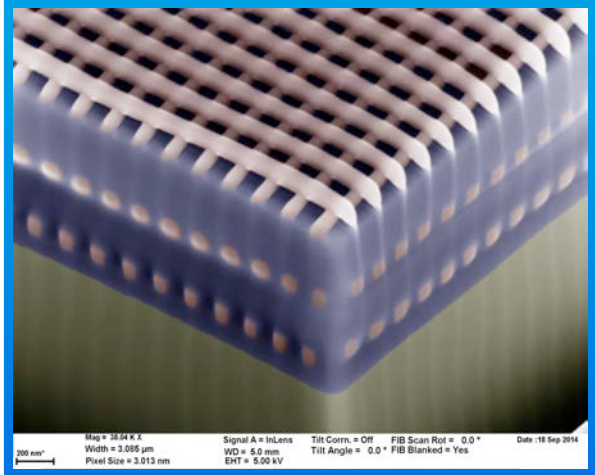
Promises

Promises

1. *Speed (85ps)*
2. *Scalability (2nm)*
3. *Multilevel (>64)*
4. **Stackability (8 layers)**
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

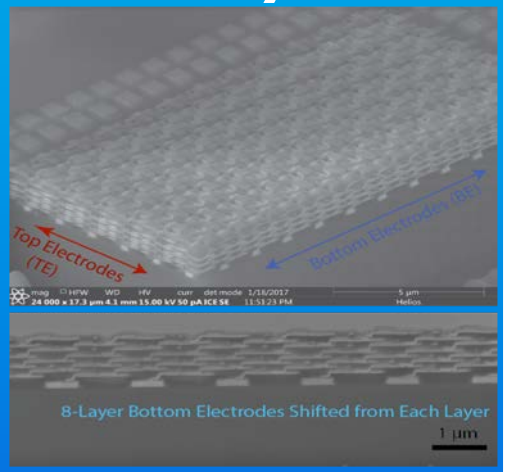
3D integration of memristor on CMOS

5 Layers



C. Li et al., Nature Comm. 8, 15666 (2017) (Q. Xia)

8 Layers



P. Lin et al., under review (2018) (Q. Xia)

Promises and Challenges

Promises

1. *Speed (85ps)*
2. *Scalability (2nm)*
3. *Multilevel (>64)*
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

Challenges

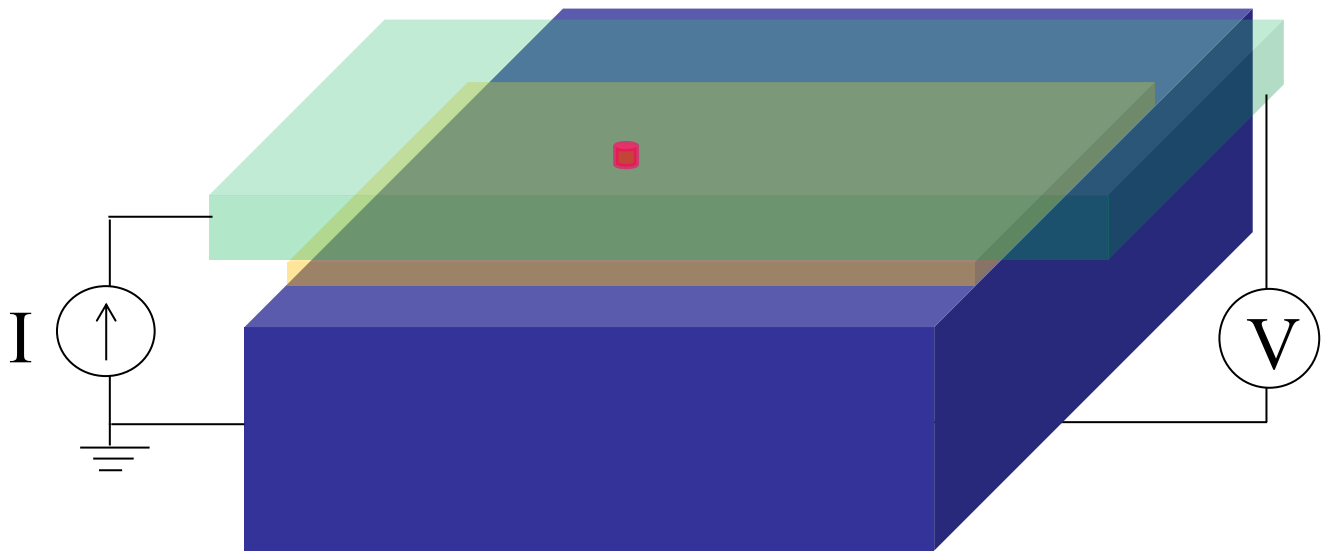
(priority of research directions at device level)

- **Mechanism**
- ***Access device in large array***
  (Transistor, selector)
- ***Desirable dynamics***
  (Diffusive memristor)

Difficulty in understanding the mechanism

Whatever changes is:

- **Buried under the top electrode (invisible)**
- **Localized both laterally and vertically (very small:1~100nm)**
- **Random location (elusive)**
- **Take place very fast (~0.1ns)**



A simplified Mechanism in a model system: TiOx devices

Where:

Bulk (oxide only) or <u>interface</u> (oxide/electrode)?

What:

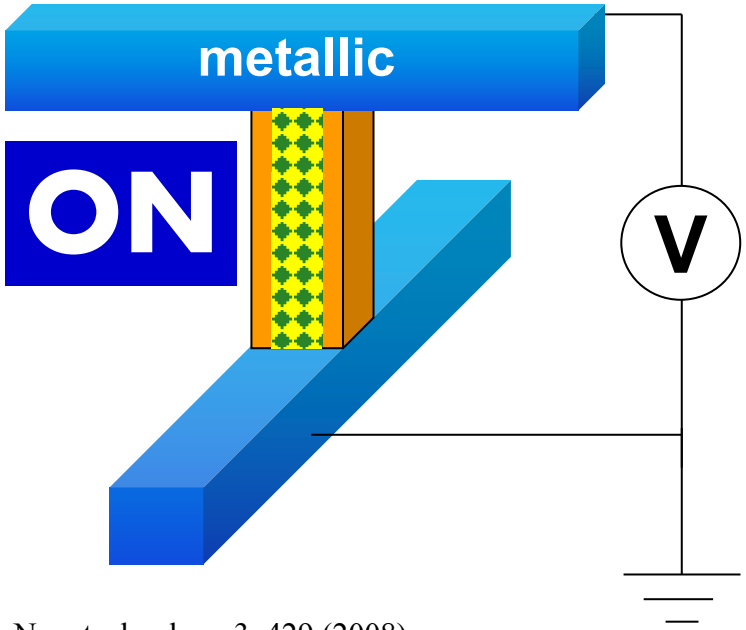
Active species (<u>V_o^{2+}</u>) responsible for the switching?

How :

<u>localized</u> switching channel grows (ON) and retracts (OFF)

There are other types of mobile ions e.g. $Ag:SiN_yO_x$

Diffusive memristor

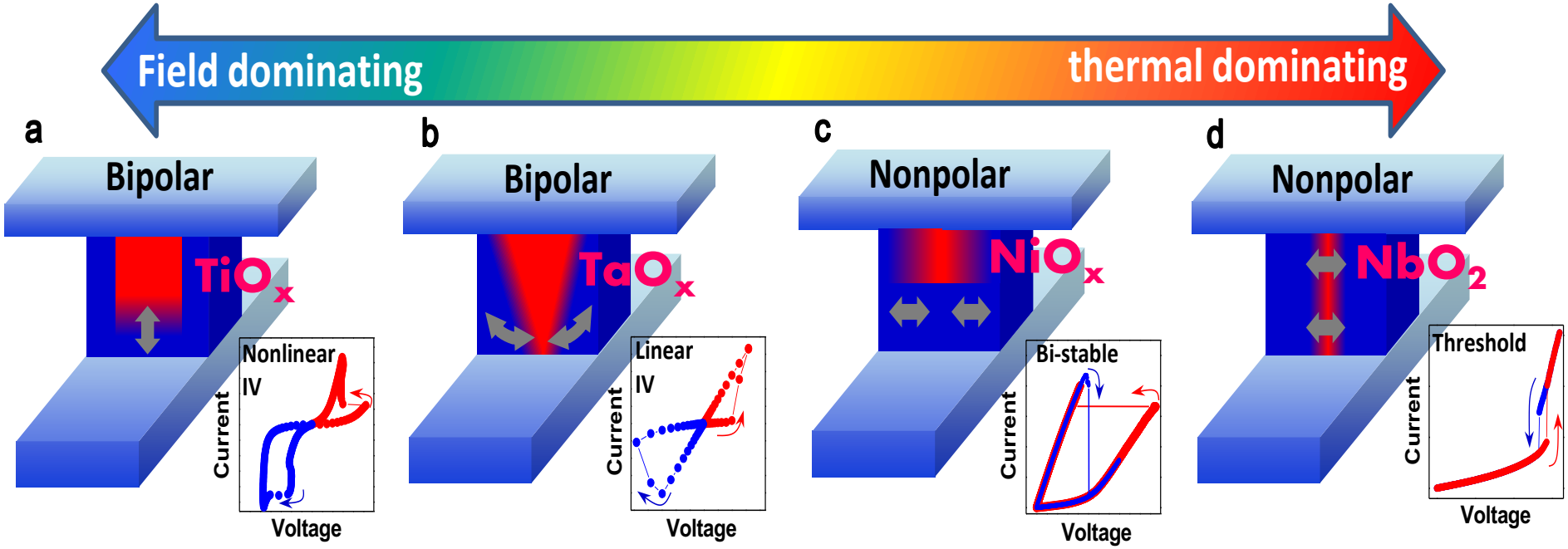


J. Joshua Yang et al., Nature Nanotechnology 3, 429 (2008)

J. Joshua Yang et al., Advanced Materials, **21**, 3754 (2009)

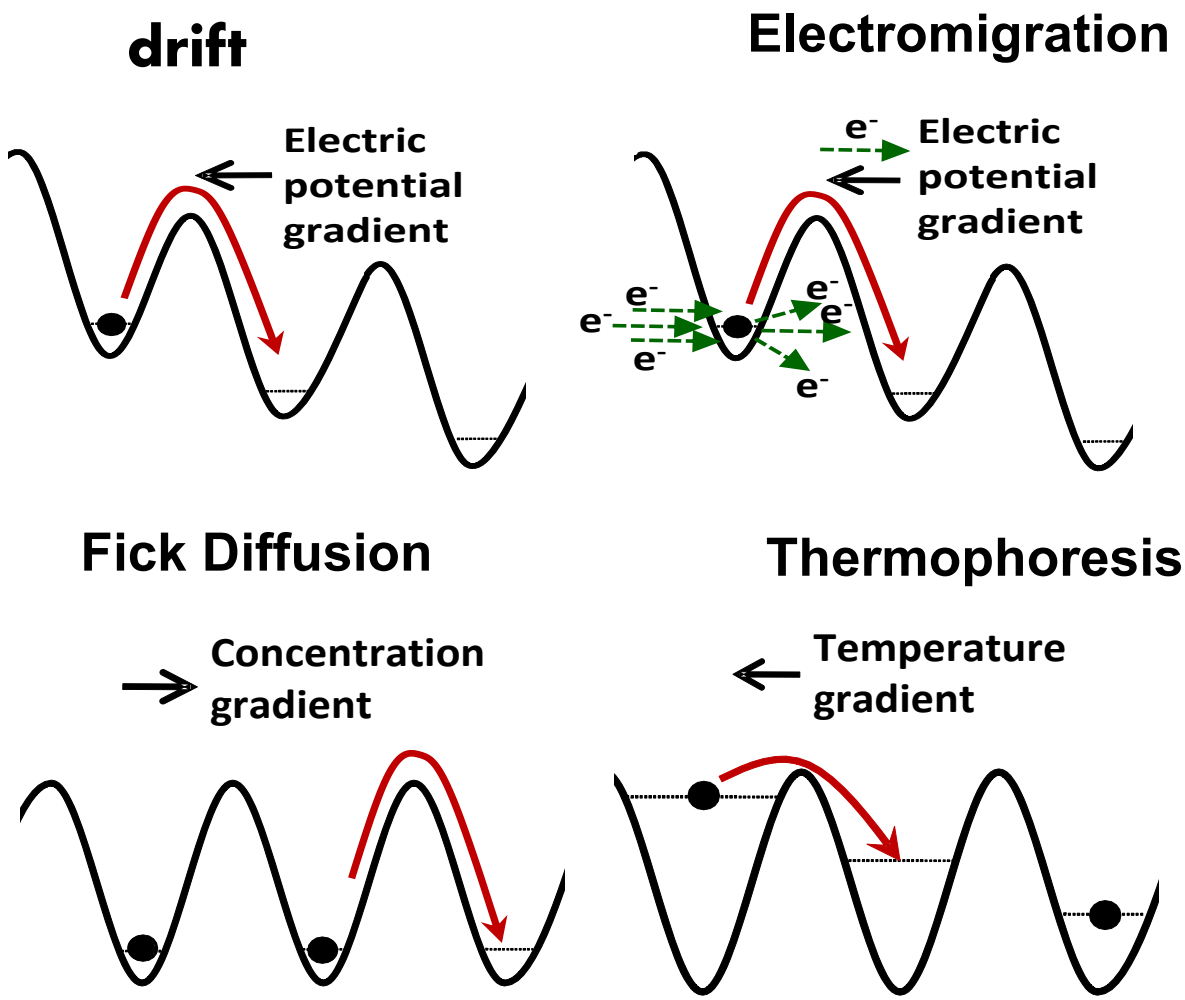
Mechanisms: Driving forces and switching types

Switching is a result of **ionic motion** under the combined effect of **electric field and Joule heating**



J. Joshua Yang et al., Nature Nanotechnology **8,** 13 (2013)

Mechanisms: Ionic transport mechanisms



J. Joshua Yang et al., Nature Nanotechnology **8,** 13 (2013)

Promises and Challenges

Promises

1. *Speed (85ps)*
2. *Scalability (2nm)*
3. *Multilevel (>64)*
4. *Stackability (>8 layers)*
5. *CMOS compatibility*
6. *Non-volatility (>10years)*
7. *Non-destructive reading*
8. *Low cost*
9. *……*

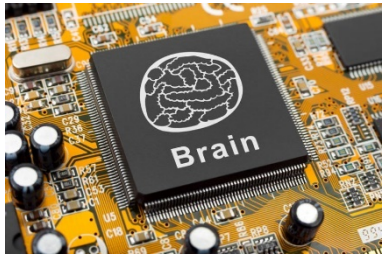
Challenges

(priority of research directions at device level)

- ***Mechanism***
- ***Access device in large array***
  (Transistor, selector)
- ***Desirable dynamics***
  (Diffusive memristor)

Simple computing without thermodynamics

Computing with memristive devices at UMass



Source: Institute of Neuroinformatics (INI)

Bio-inspired approach
(important thermodynamics)

New building blocks

Artificial Synapse

Nature Materials **16**, 101 (2017).
Nature Comm. **8**, 752 (2017).

Artificial Neuron

Nature Electronics, **1**, 137 (2018).
Nature Comm. **9**, 417 (2018).
Nature Comm. **9**, *3208* (2018).

Memristor

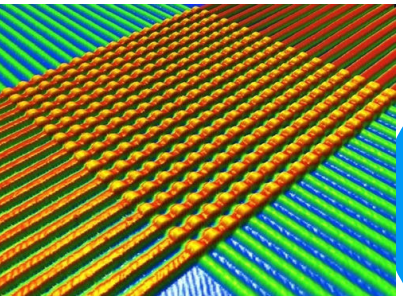
Dot Product Engine

Nature Electronics **1**, 52 (2018).
Adv. Mater. **29**, 1705914 (2018).
Nature Communications **9**, *2385 (2018)*
Nature Machine Intelligence **1**, Accepted (LSTM, 2018)

Accelerate Deep Neural Network

Source: http://www.google.com/selfdrivingcar/

Traditional AI approaches
(little dynamics)



Source: HP Lab

Novel Ultra Large Array Operation

Memory, Robotics, Security…

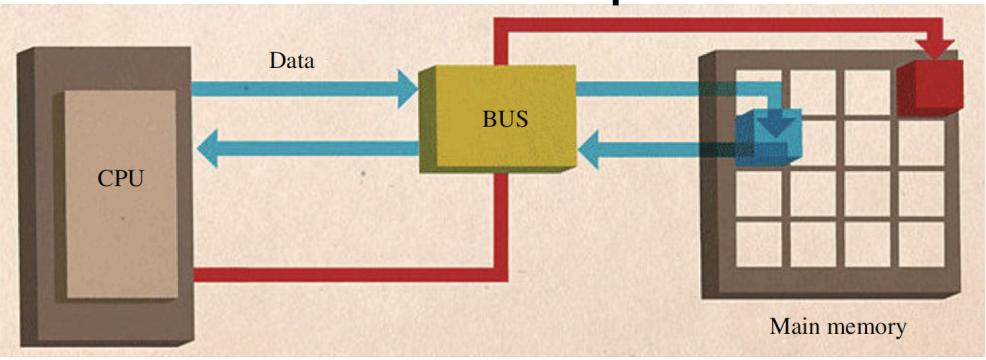
Nature Materials **16**, 396 (2017).
Nature Comm. **8**, 882 (2017).
Nature Comm. **8**, 15666 (2017).
Nature Electronics **1**, 548 (2018).

Nature Electronics **1**, 130 (2018).
Adv. Func. Mater. **27**, 1704862 (2017).
Adv. Mater. **29**, 1604457 (2017).

Other applications
(some dynamics)

Issues with classical computer hardware in big-data era

Classical computer

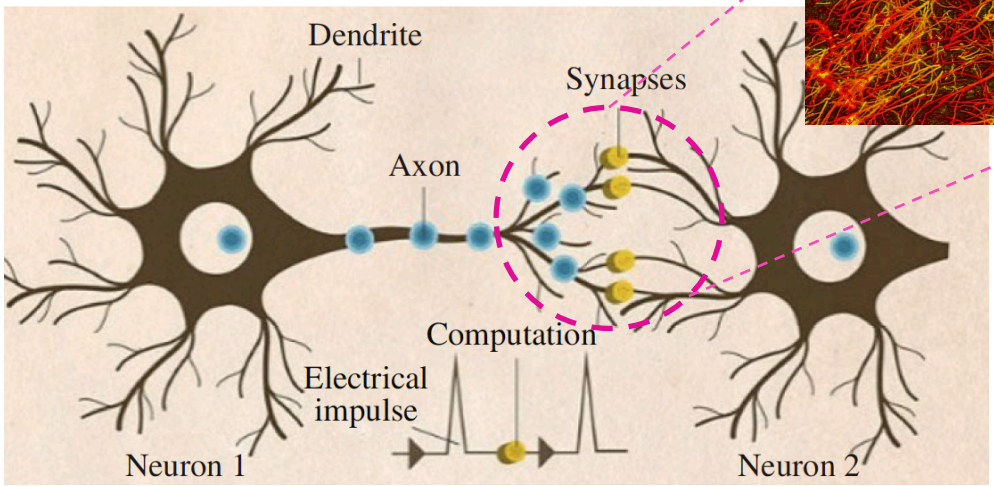


Devices:
reaching physical limit for scaling down (Moore's law is over)

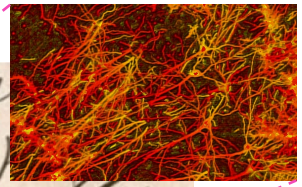
Architecture:

- separate processor and storage (Von Neumann bottleneck)
- Sequential process
- Analog/Digital conversion

Brain



networks



In Brain:
Neurons, Synapses and networks

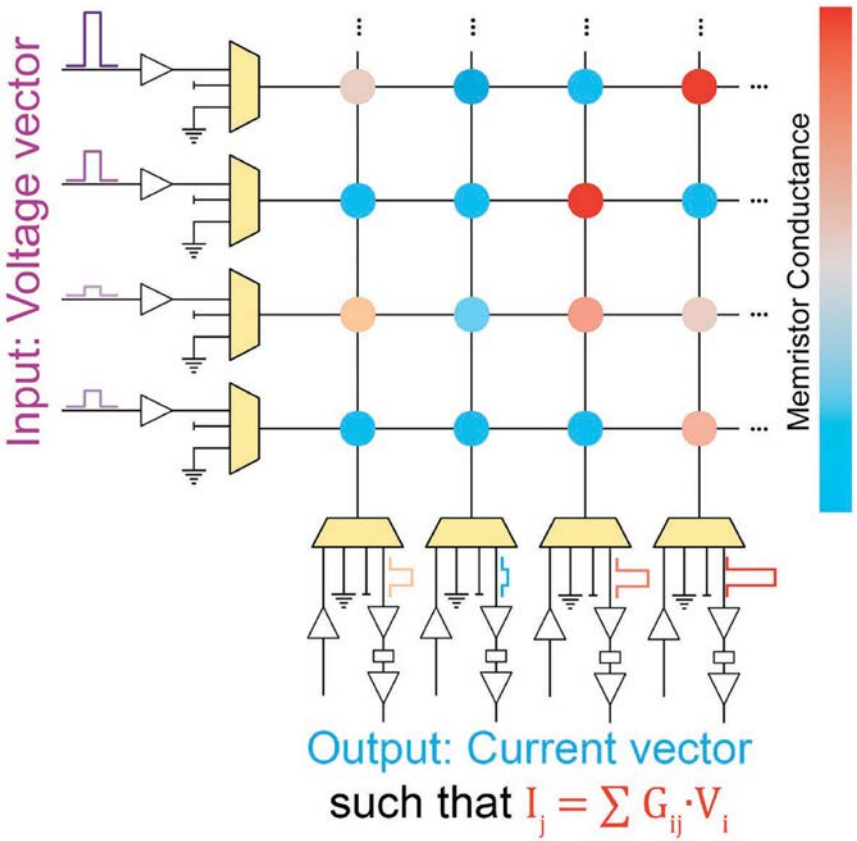
- In-memory computing
- Parallel comping (instead of sequential)
- Analog computing

U., Navnidhi, S. Joshi, J. Joshua Yang. "Synaptic electronics and neuromorphic computing." *Science China Information Sciences* 59, 061404 (2016)

Physical computing application: Dot-product Engine (for Deep Neural Network)

- **Vector x Matrix** multiplications
- Computation intensive tasks (many multiplication and addition steps)

Only one step in dot-product engine

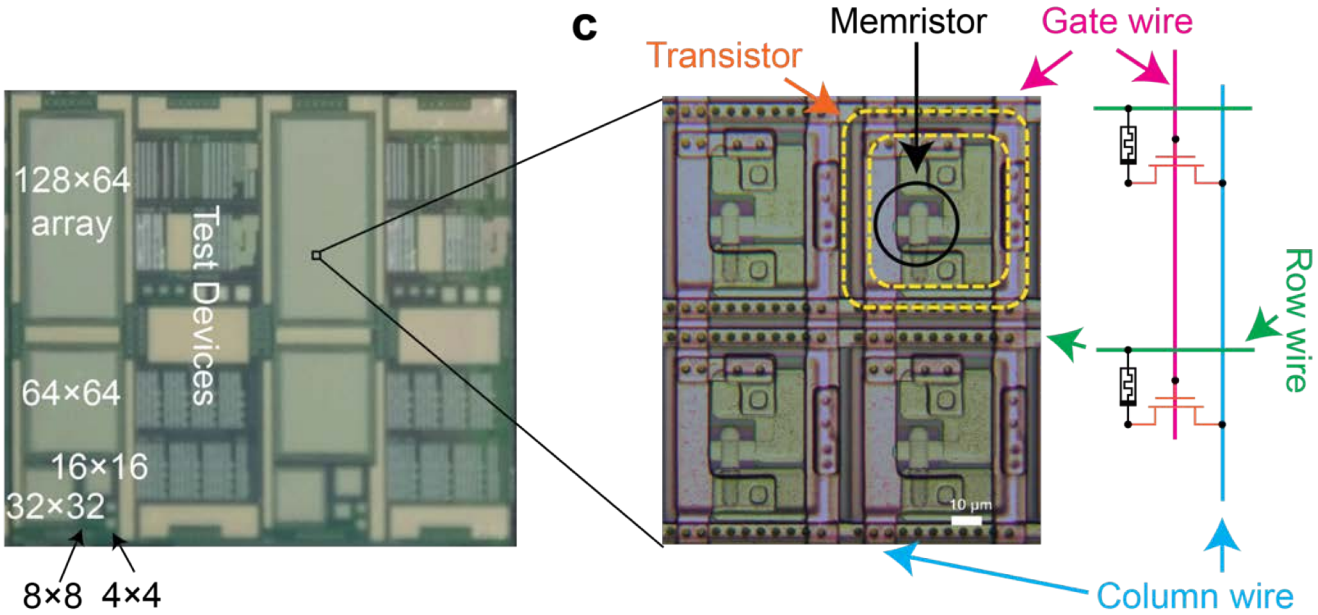


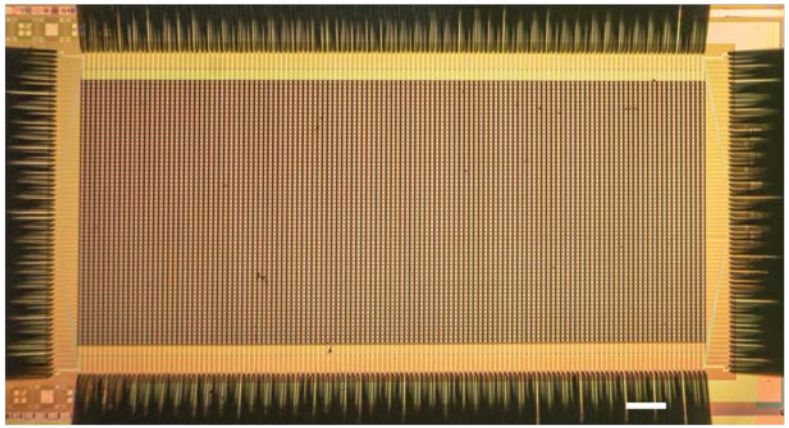
With orders of magnitude Improvement in speed and power efficiency!

"Dot-Product Engine for Neuromorphic Computing: Programming 1T1M Crossbar to Accelerate Vector-Matrix Multiplication", ***the 53rd Design Automation Conference (DAC), 2016***.

M. Hu et al., Adv. Mater. ***29***, 1705914 (2018).

Deep Neural Network: Dot-Product Engine (DPE)

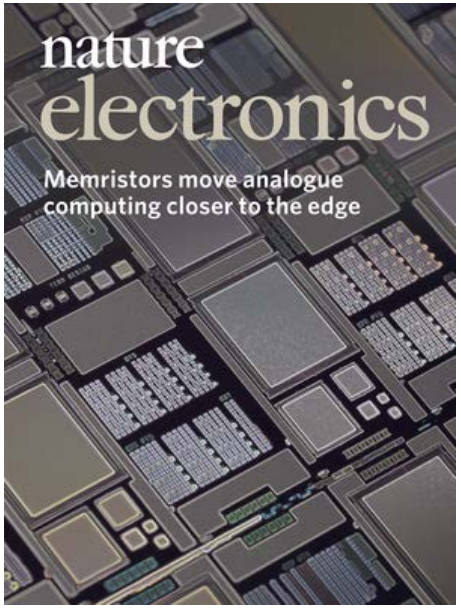




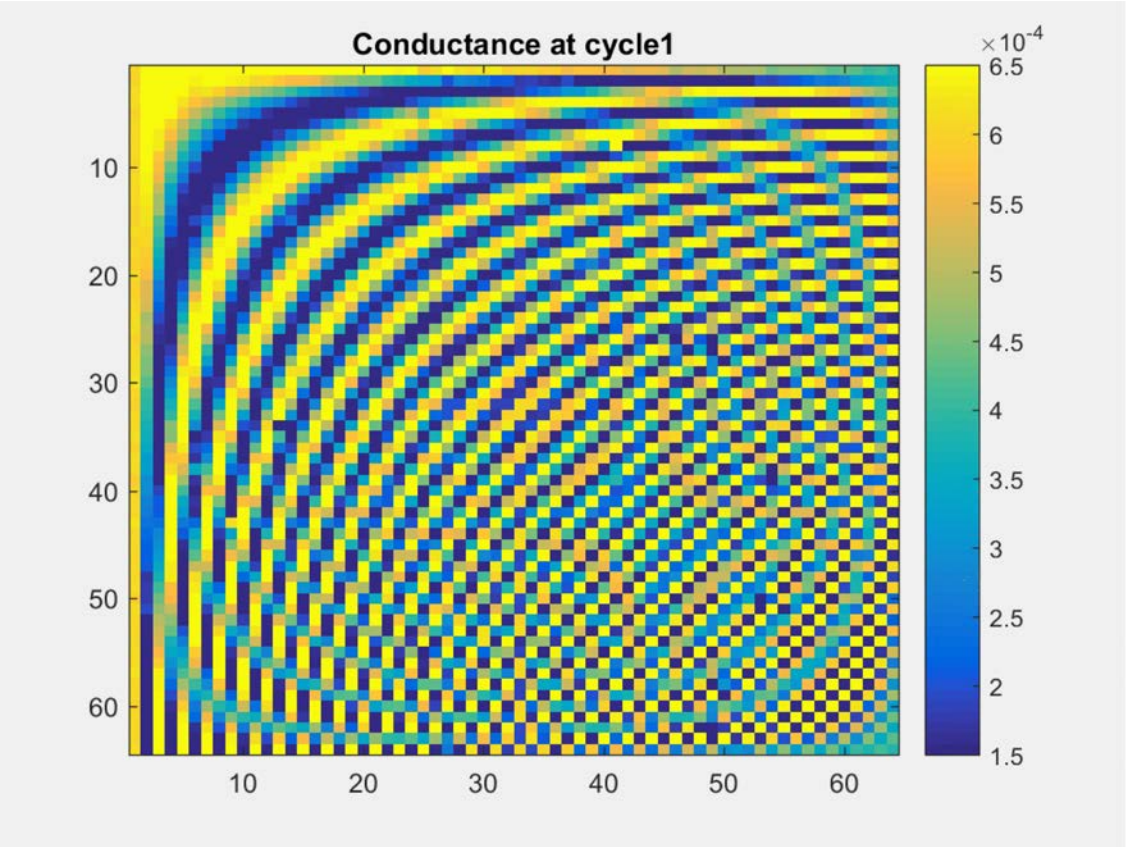
C. Li et al, Nature Electronics 1, 52 (2018)

(collaborated with Stan Williams & John Paul Strachan)

Accurate programming of resistance states in the array



(Cover article of inaugural issue)



C. Li et al., Nature Electronics 1, 52 (2018)

Image Processing (experimental demo)

Image Compression Using Analog Memristor Array (DCT)

(20:3 compression ratio)

Original image

Software processed

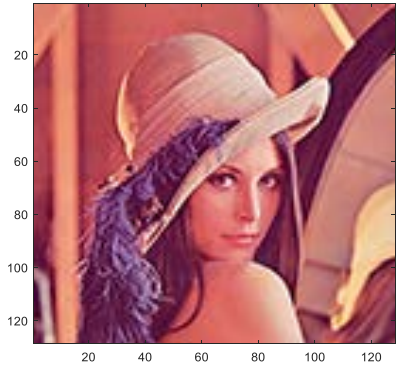
Memristor array processed (Experimental)

C. Li et al., Nature Electronics 1, 52 (2018)

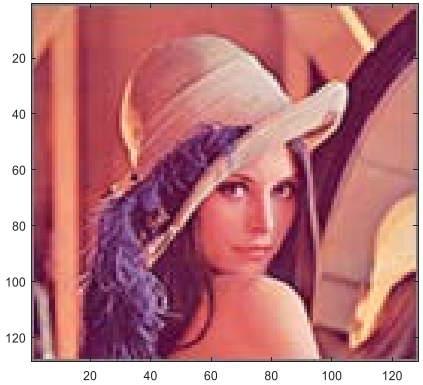
Image filtering (experimental demo)

7 Different Filters Applied at the Same Time

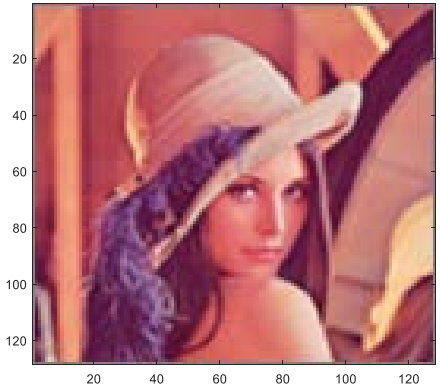
Original



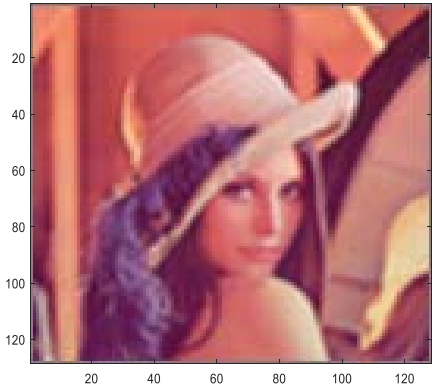
Gaussian



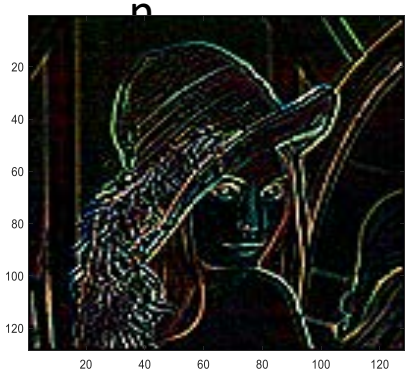
Disk



Average

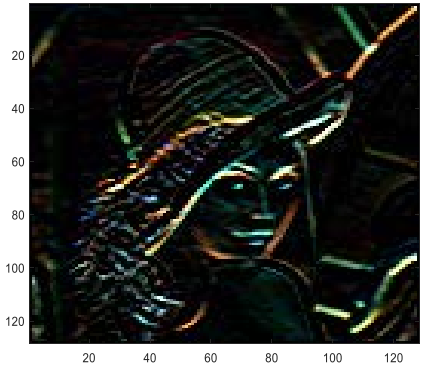


Laplacian

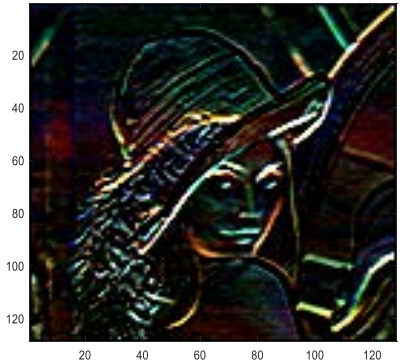


LoG

Prewitt (Horizontal)

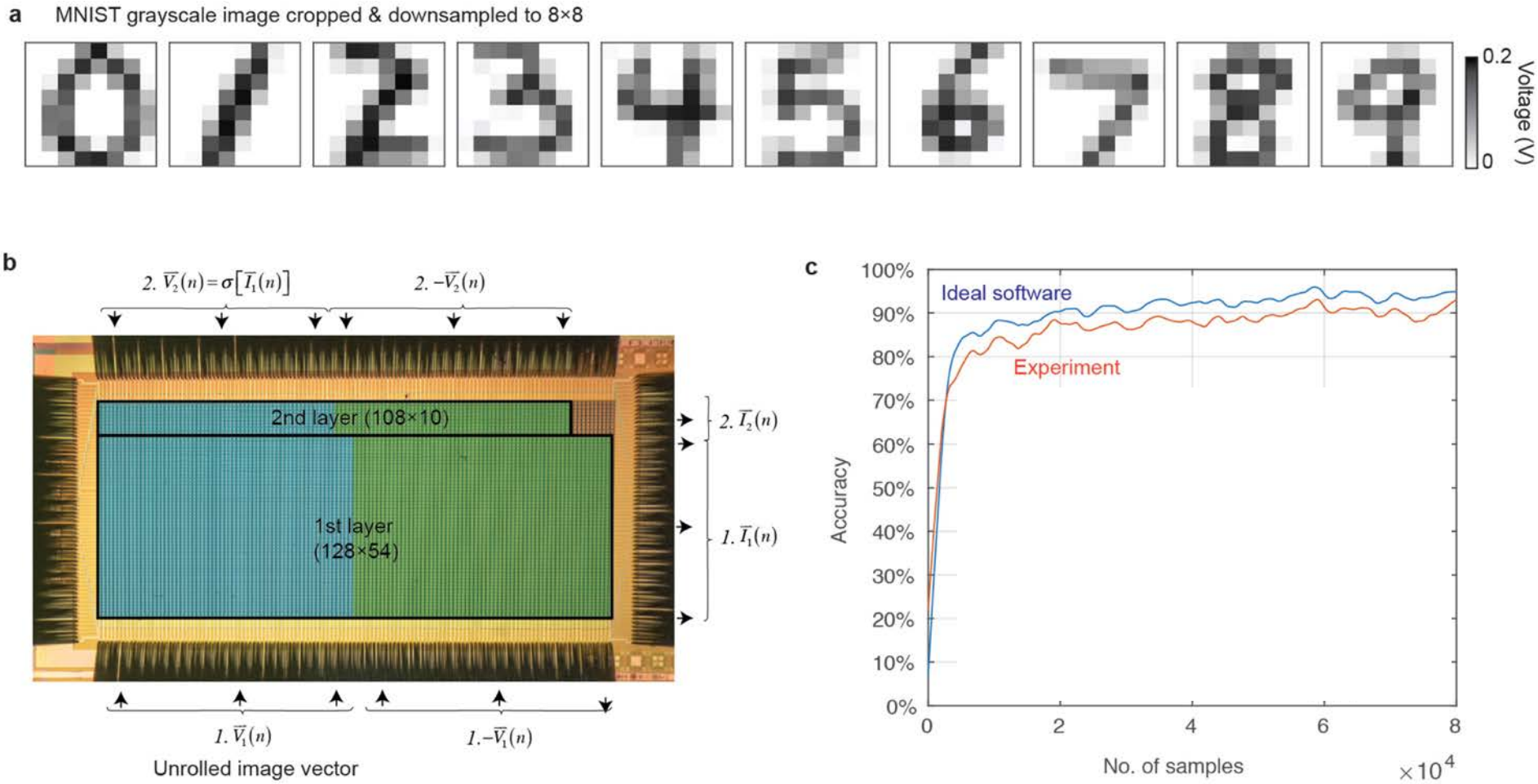


Sobel (Horizontal)



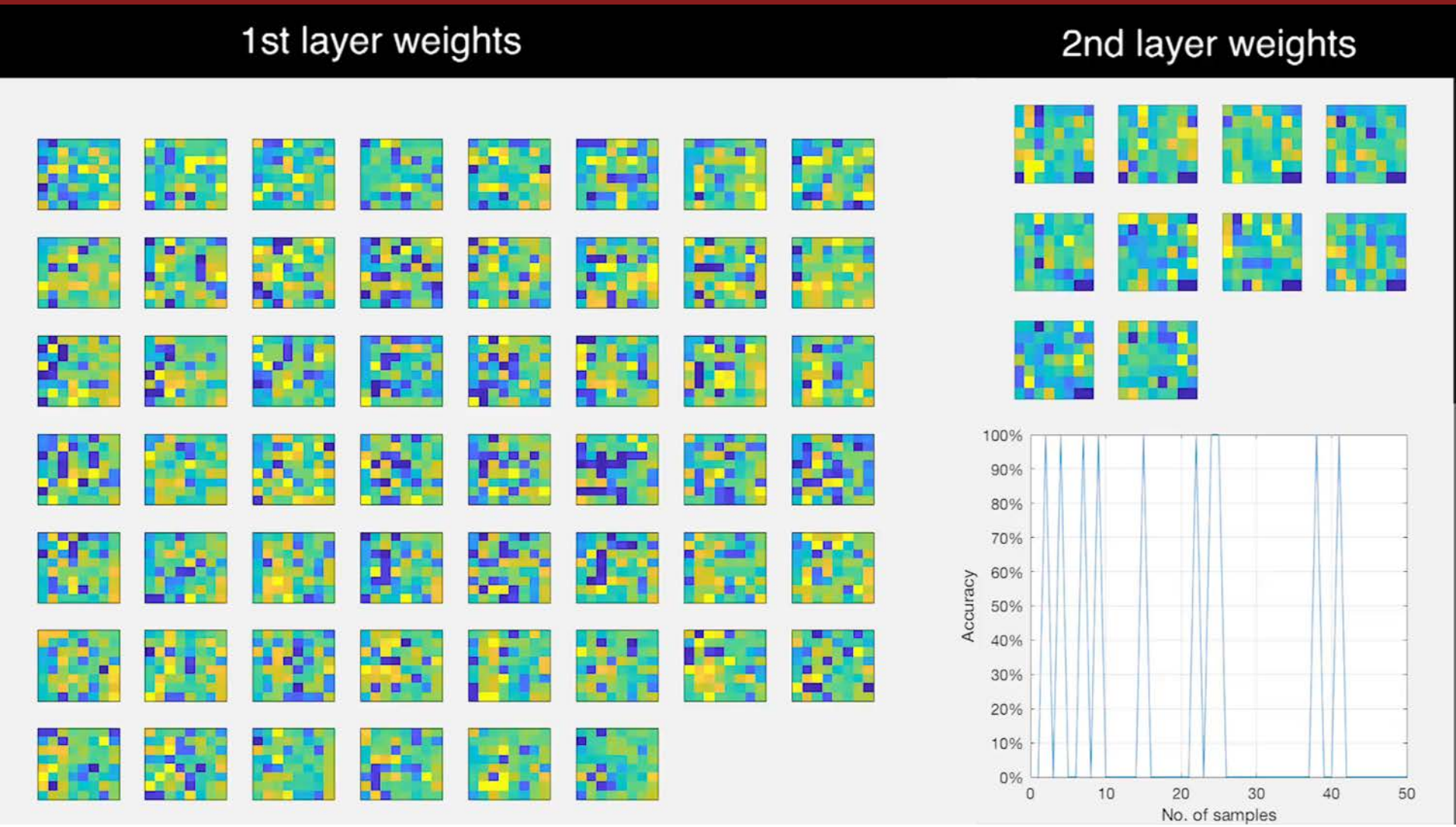
C. Li et al., Nature Electronics 1, 52 (2018)

2-layer Neural Network with supervised learning



Li et al., Nature Communications 9, 2385 (2018).

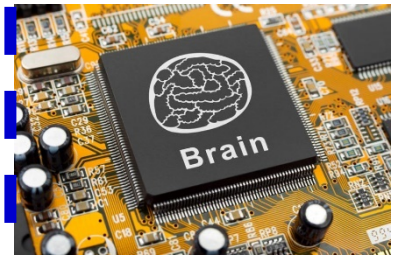
2-layer Neural Network with supervised learning



Li et al., Nature Communications 9, 2385 (2018).

More advanced computing (with thermodynamics)

Computing with memristive devices at UMass



ource: Institute of Neuroinformatics (INI)

Bio-inspired approach
(important thermodynamics)

New building blocks

Artificial Synapse

Nature Materials **16**, 101 (2017).
Nature Comm. **8**, 752 (2017).

Artificial Neuron

Nature Electronics, **1**, 137 (2018).
Nature Comm. **9**, 417 (2018).
Nature Comm. **9**, *3208* (2018).

Memristor

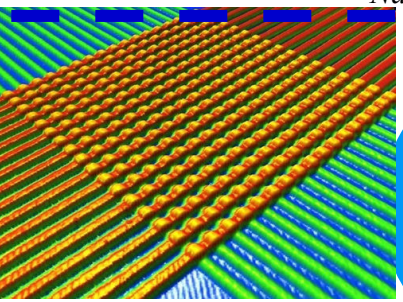
Dot Product Engine

Nature Electronics **1**, 52 (2018).
Adv. Mater. **29**, 1705914 (2018).
Nature Communications **9**, *2385 (2018)*
Nature Machine Intelligence **1**, Accepted (LSTM, 2018)

Traditional AI approaches
(little dynamics)

Accelerate Deep Neural Network

Source: http://www.google.com/selfdrivingcar/



Source: HP Lab

Novel Ultra Large Array Operation

Memory, Robotics, Security…

Nature Materials **16**, 396 (2017).
Nature Comm. **8**, 882 (2017).
Nature Comm. **8**, 15666 (2017).
Nature Electronics **1**, 548 (2018).

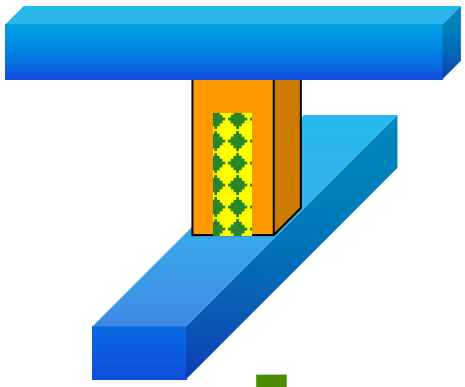
Nature Electronics **1**, 130 (2018).
Adv. Func. Mater. **27**, 1704862 (2017).
Adv. Mater. **29**, 1604457 (2017).

Other applications
(some dynamics)

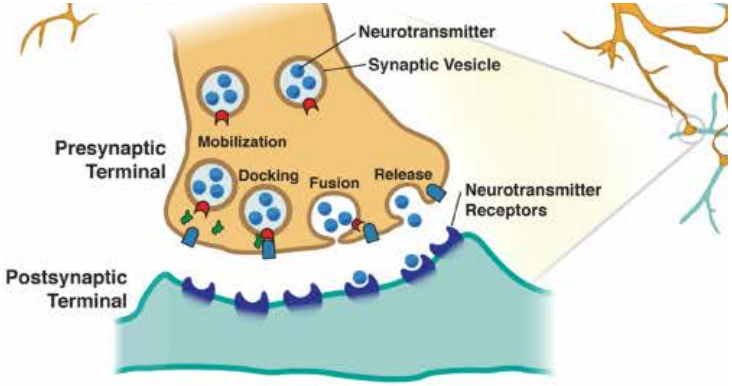
Problems with traditional memristors for synapse: lack of synaptic dynamics

Spike-timing dependent plasticity (STDP)

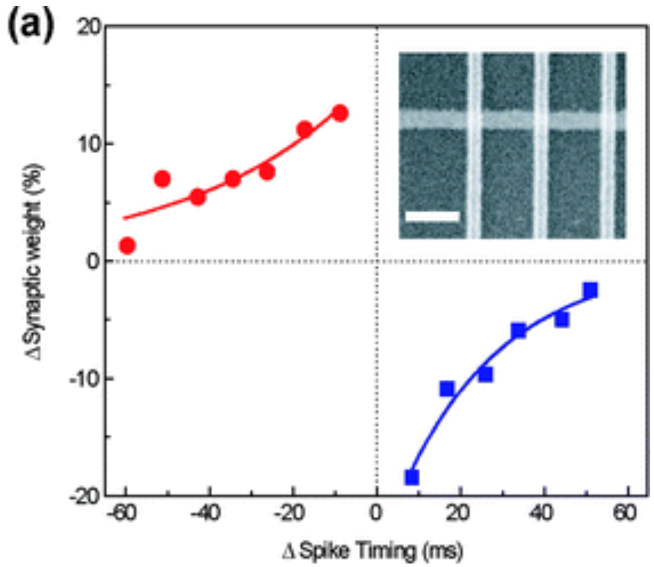
Memristor

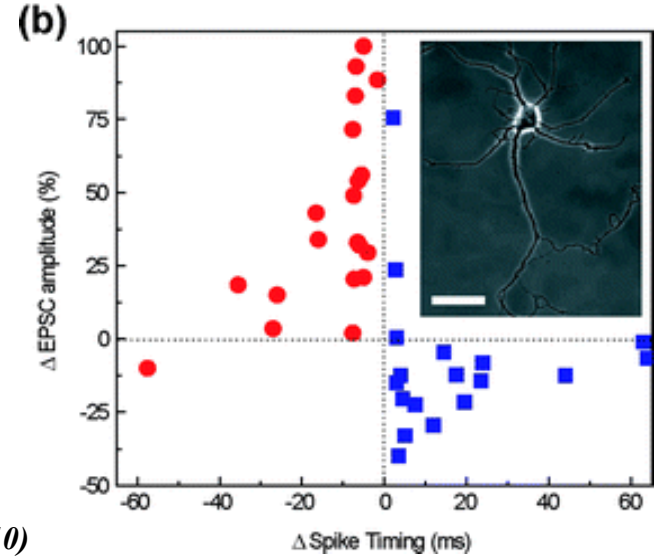


Synapse



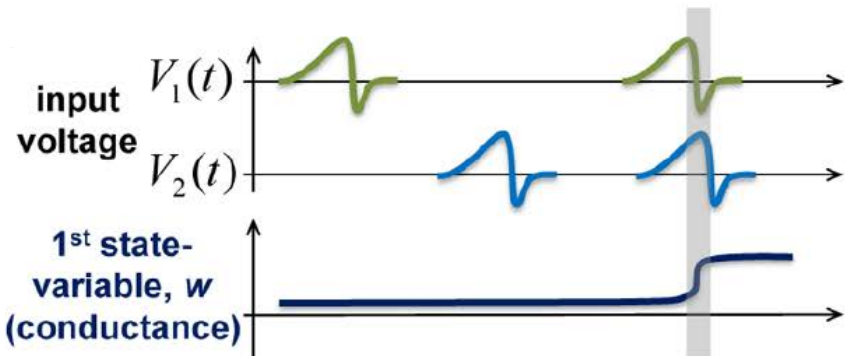
Indirect STDP

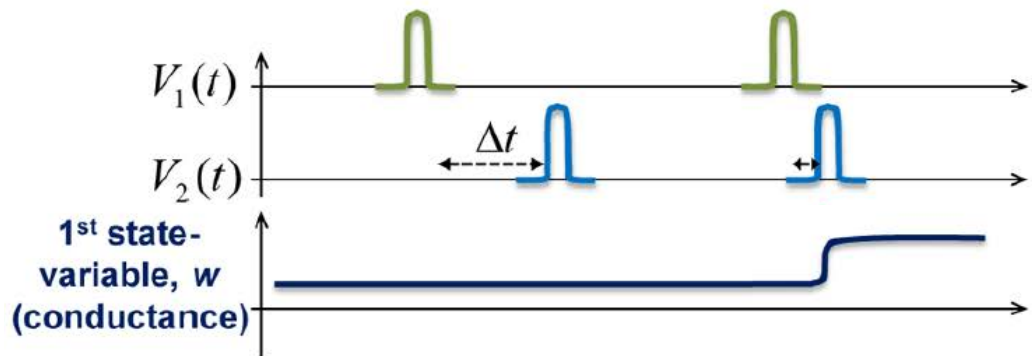




Nano Lett. 10, 1297 (2010)

Synapse: spike timing carries the information!





(Umich) Nano Lett. 15, 2203 (2015), Adv. Funct. Mater.25, 4290 (2015)

- Higher energy (not sparse spikes)
- More complicated circuits and algorithms
- Only phenomenological emulation of one or two functions and missing most others

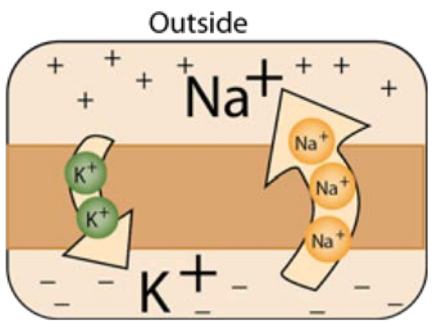
- Dynamics is critical for synaptic functions
- With dynamics, many synaptic functions come out naturally and simultaneously!!

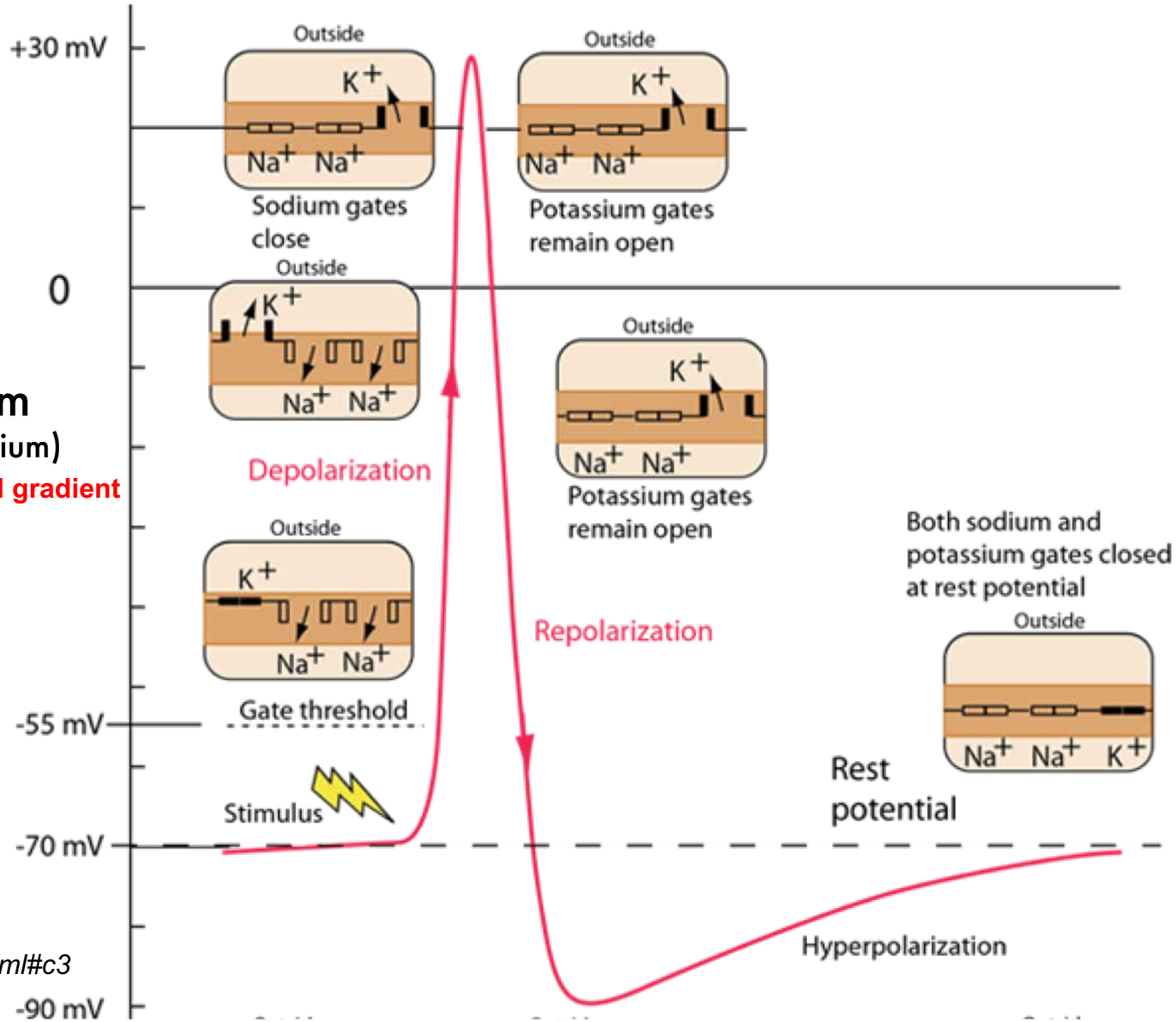
Amplitude is less important, timing carries the information→ synaptic dynamics!

Thermodynamics in neuronal computing: action potential

ATP pump to keep neuron in equilibrium
(out of thermodynamic equilibrium)

Energy is stored in electrochemical gradient

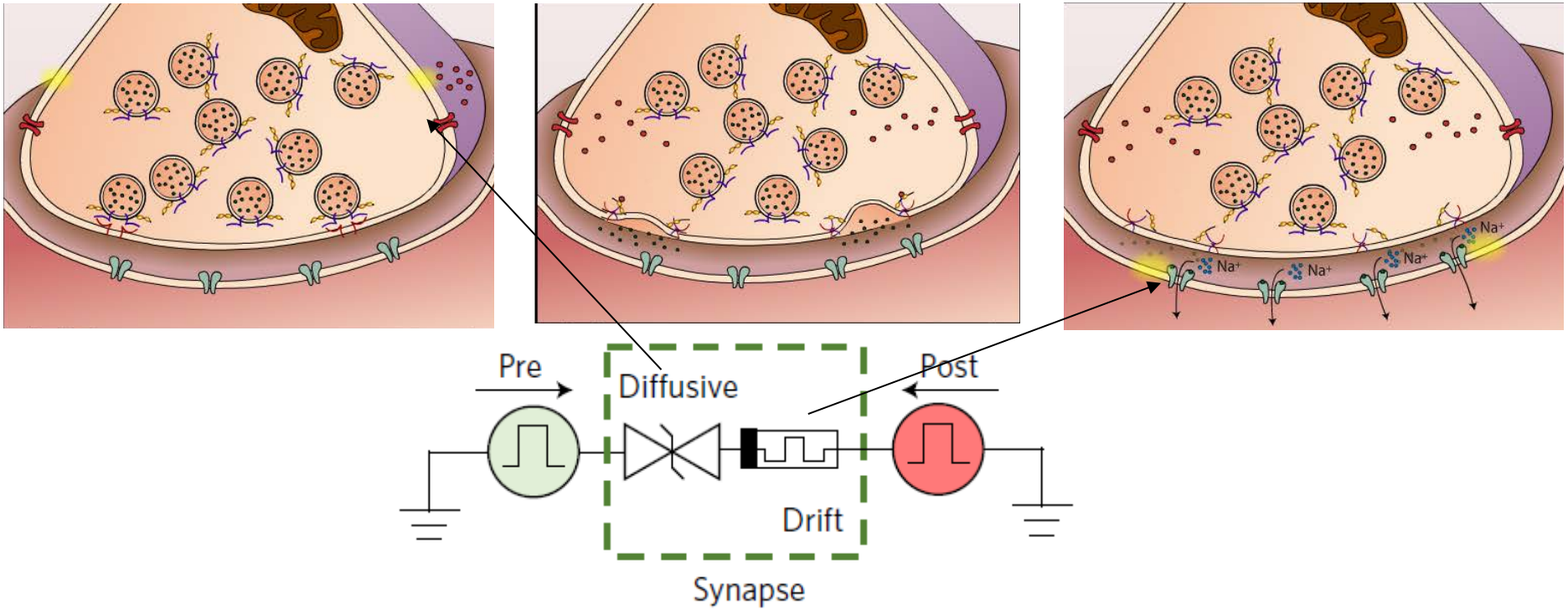




http://hyperphysics.phy-astr.gsu.edu/hbase/Biology/actpot.html#c3

Reading process: similar role of Ca and Ag

Impulse from upper axon terminal → **if <u>large enough</u>,** it opens the **<u>voltage-controlled calcium channels</u>** in pre-neuron, Ca **<u>diffuses</u>** through →neurotransmitter release → open receptors in post-neuron →ions like Na+ go into post-neuron, generating impulse, passing on. The postsynaptic response (impulse) is weighted by for example the number of receptors

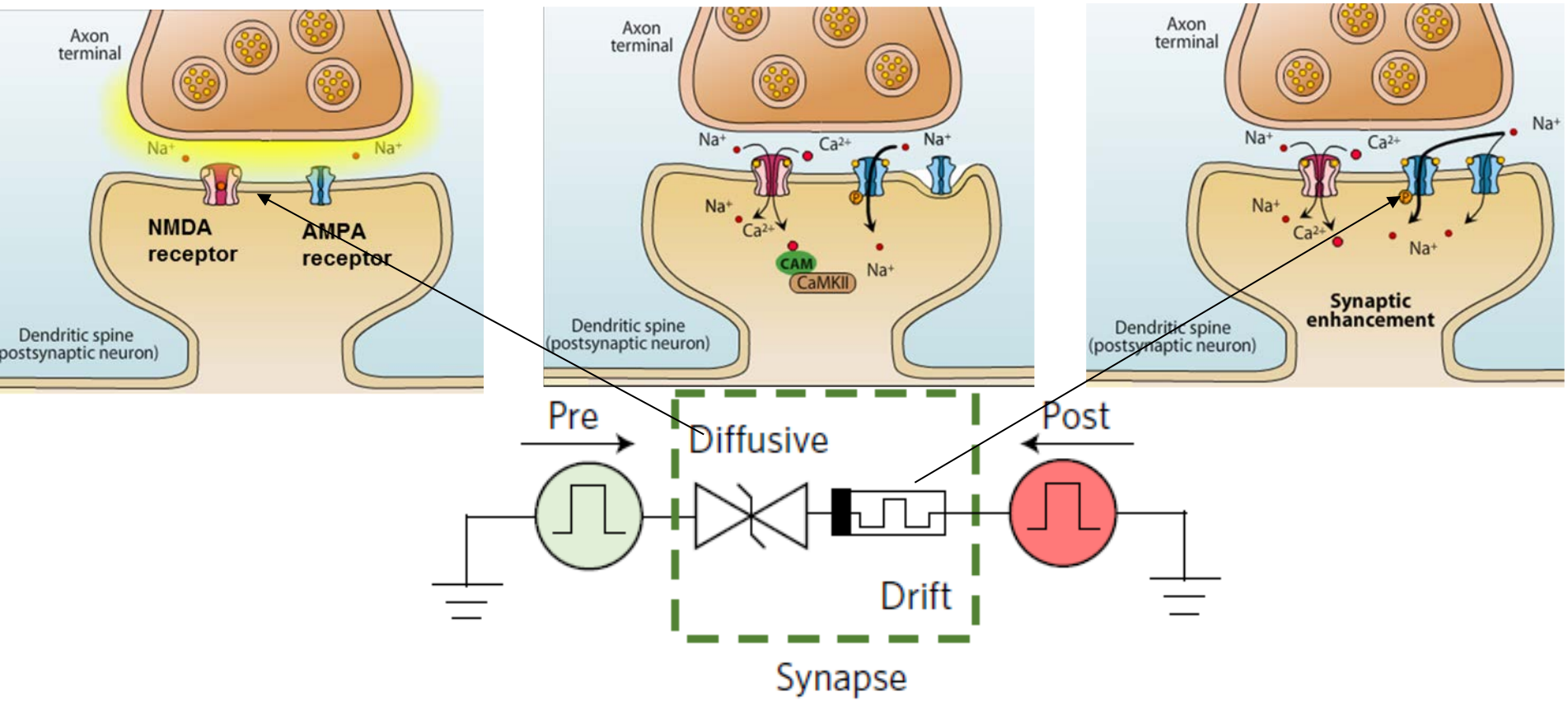


The Ca diffusion dynamics control the process!

http://sites.sinauer.com/neuroscience5e/animations05.01.html

Diffusion dynamics is critical for neuromorphic functions

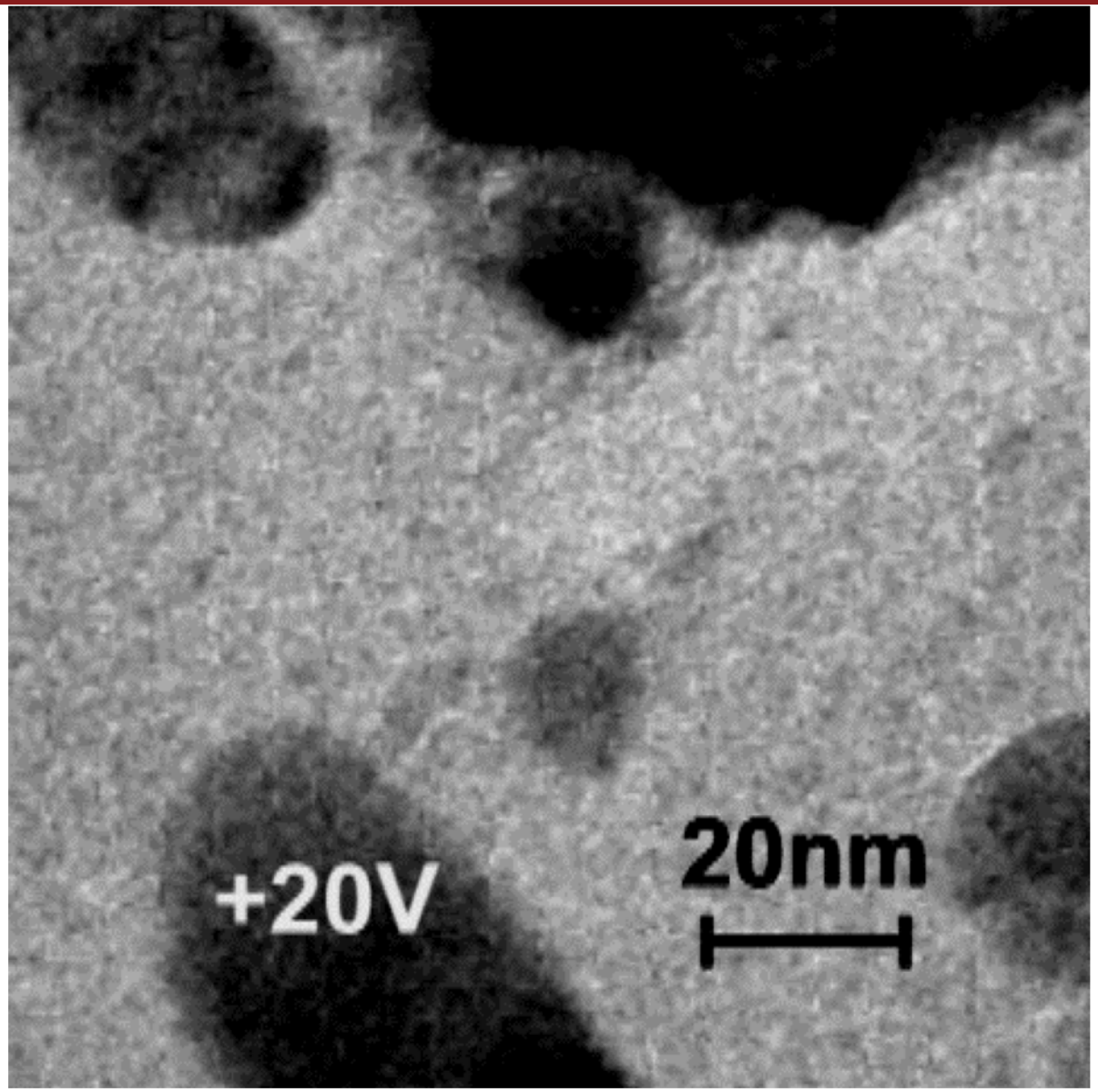
Very Large Impulse from upper axon terminal → open NMDA receptor, Ca^{2+} diffuses through → activate some bio-activities → increasing both the number and conductance of AMPA receptors→ synapse enhanced – long term plasticity.



The Ca^{2+} dynamics controls the synaptic weight changes!

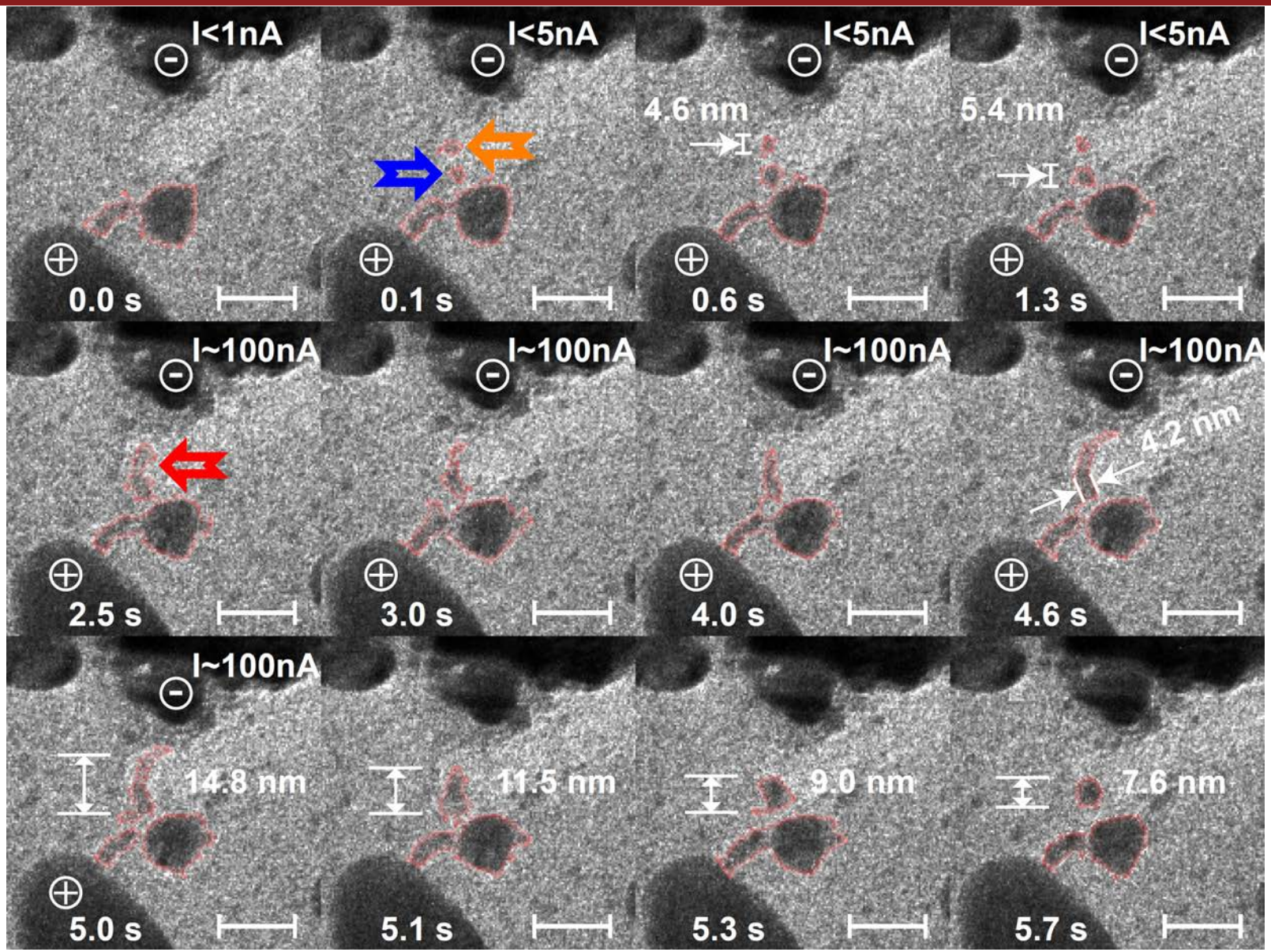
http://sites.sinauer.com/neuroscience5e/animations08.02.html

Diffusion dynamics in Diffusive memristors

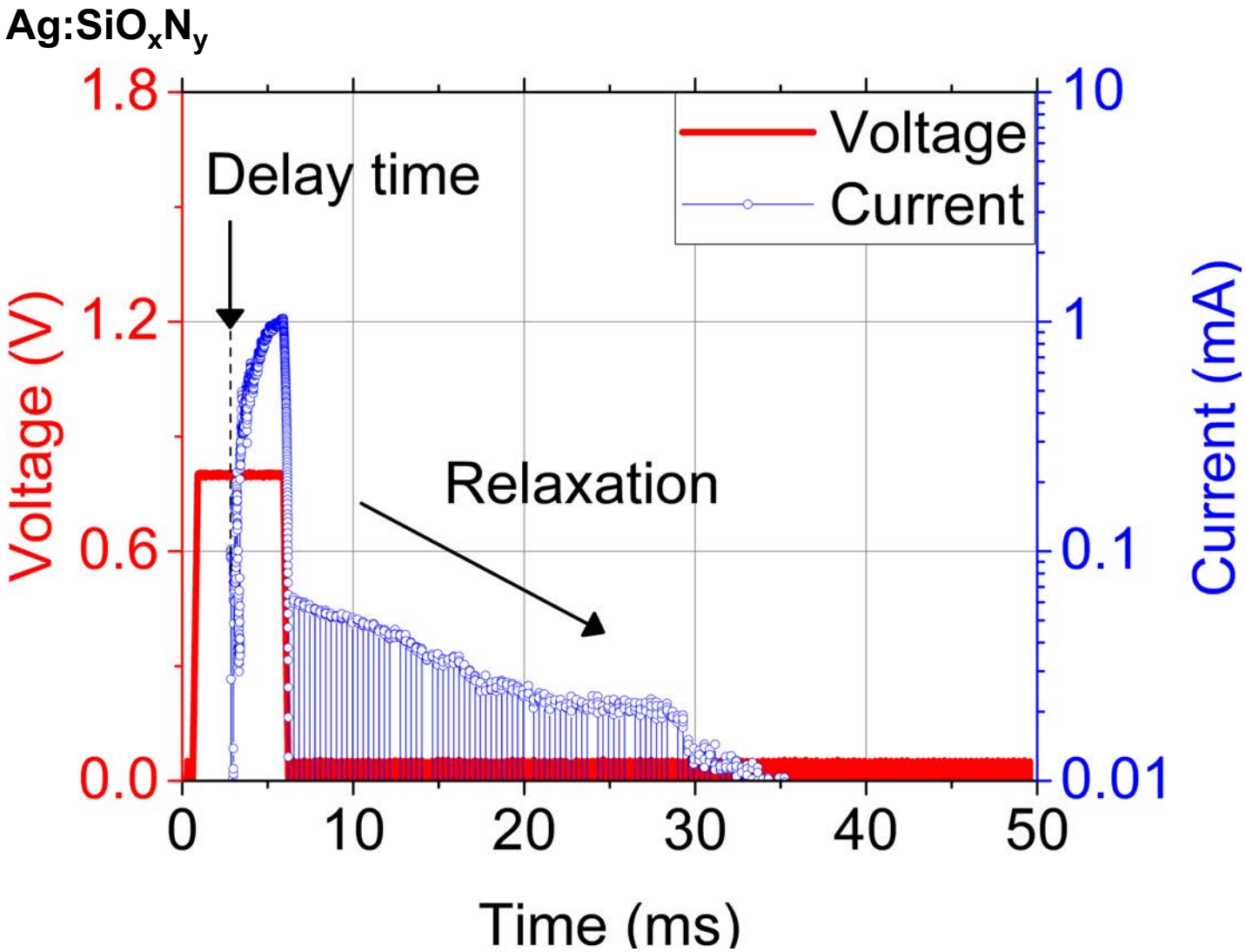


Z. Wang *et al.*, Nature Materials 16, 101 (2017)

Thermal diffusion and Interfacial energy minimization



Diffusive memristors: dynamics governed by thermodynamics



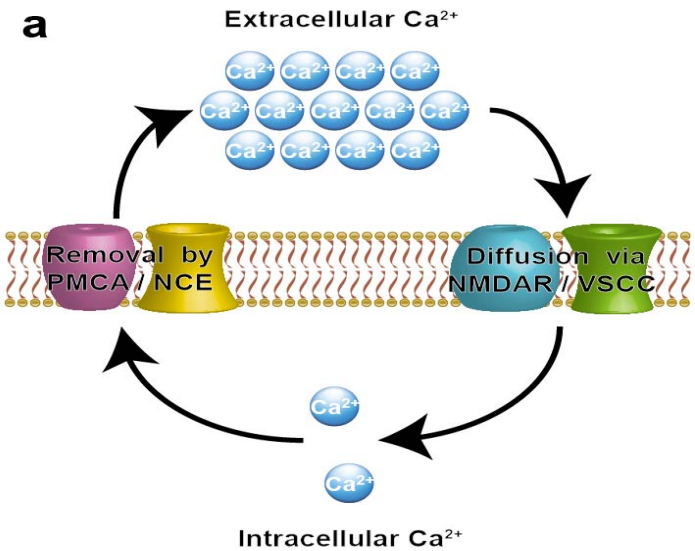
Z. Wang et al., Nature Materials 16, 101 (2017)

Realization of critical Dynamics in memristors: (Ca/Ag diffusion analogy)

How to realize the critical dynamics:
Physical similarity leads to mechanism and function similarities!!

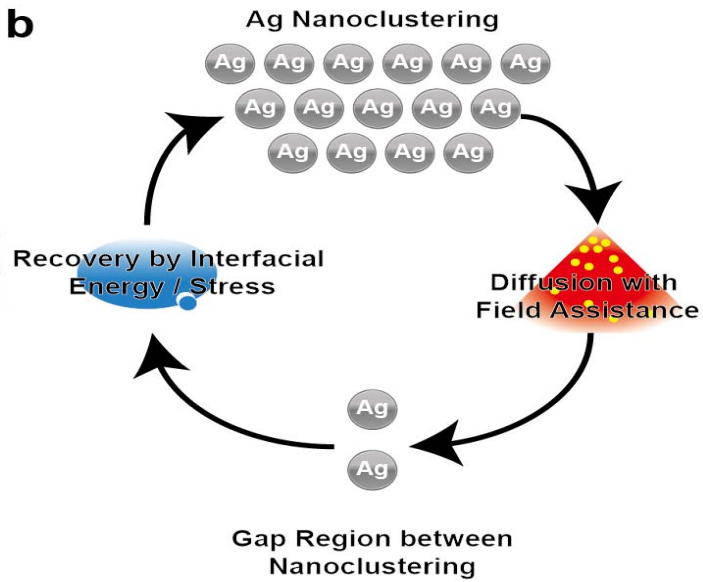
bio-system:

Ca^{2+} diffusion dynamics

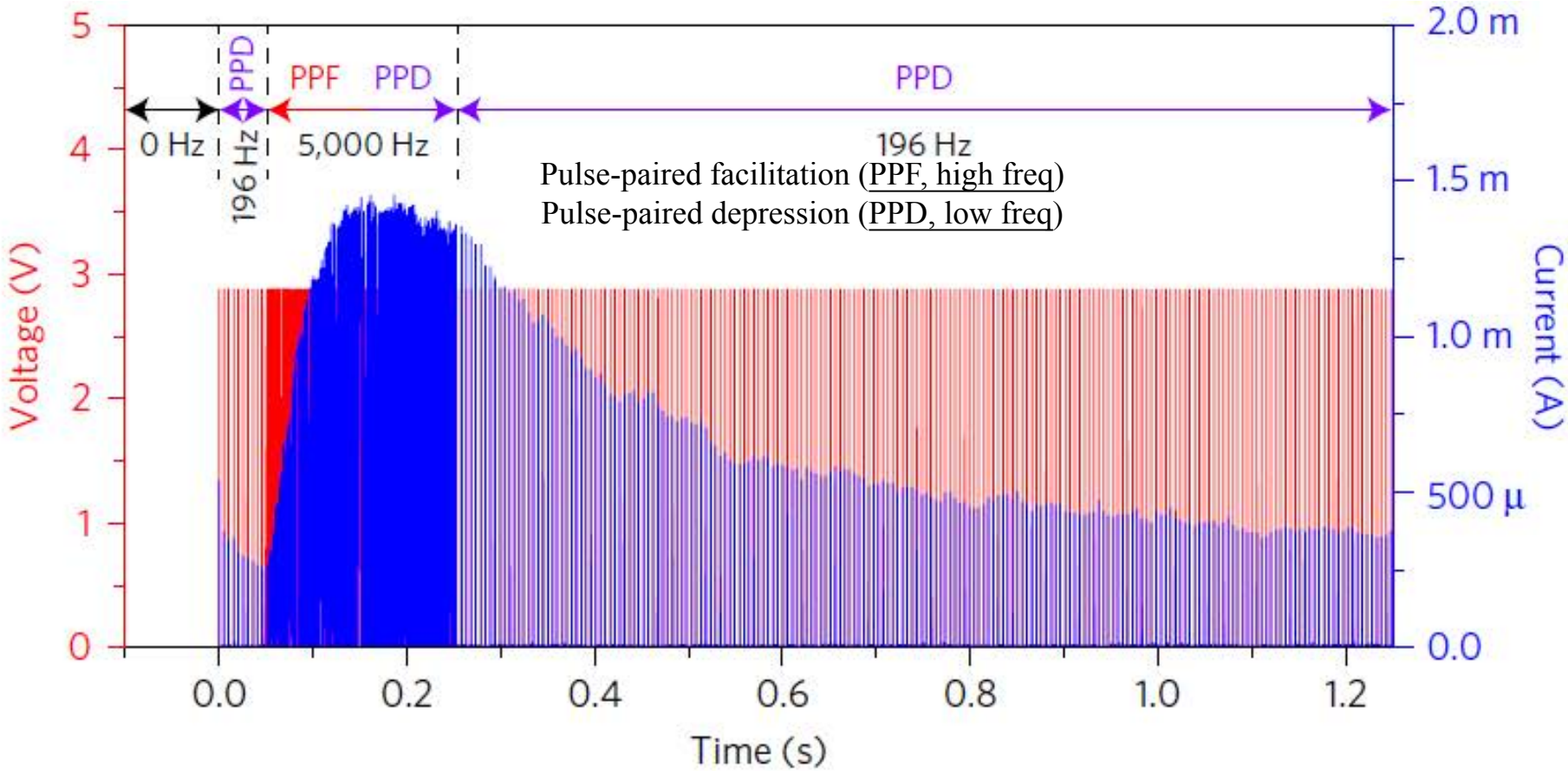


diffusive memristor:

Ag diffusion dynamics



Artificial synapse:
Realizing Short-term plasticity behavior faithfully

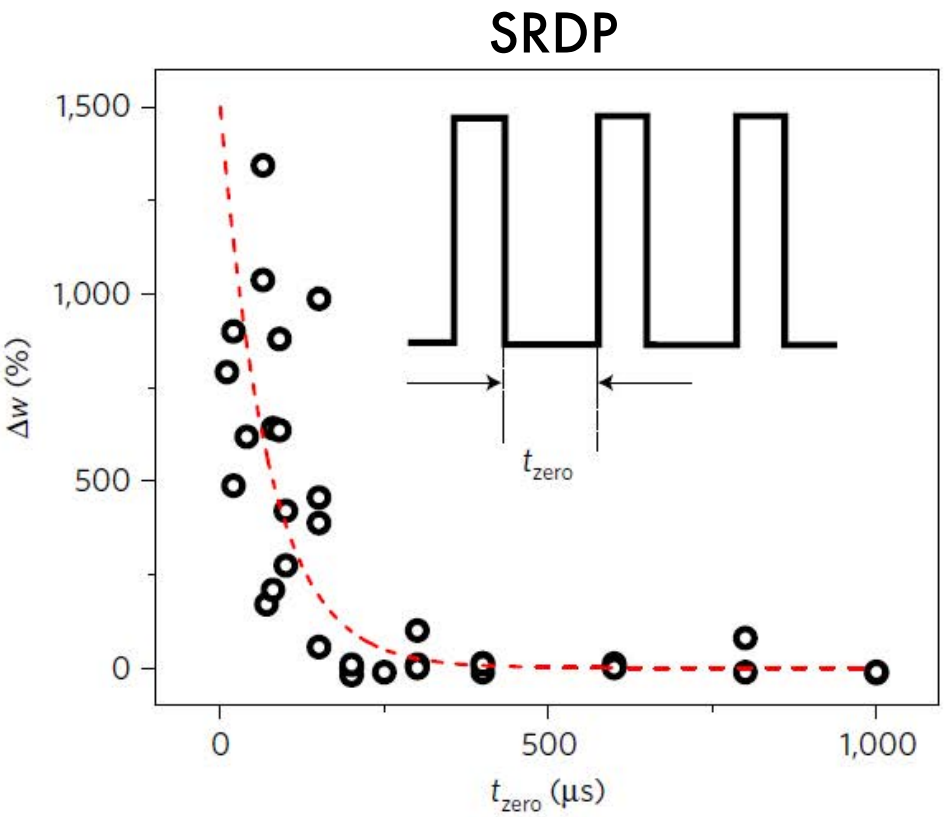


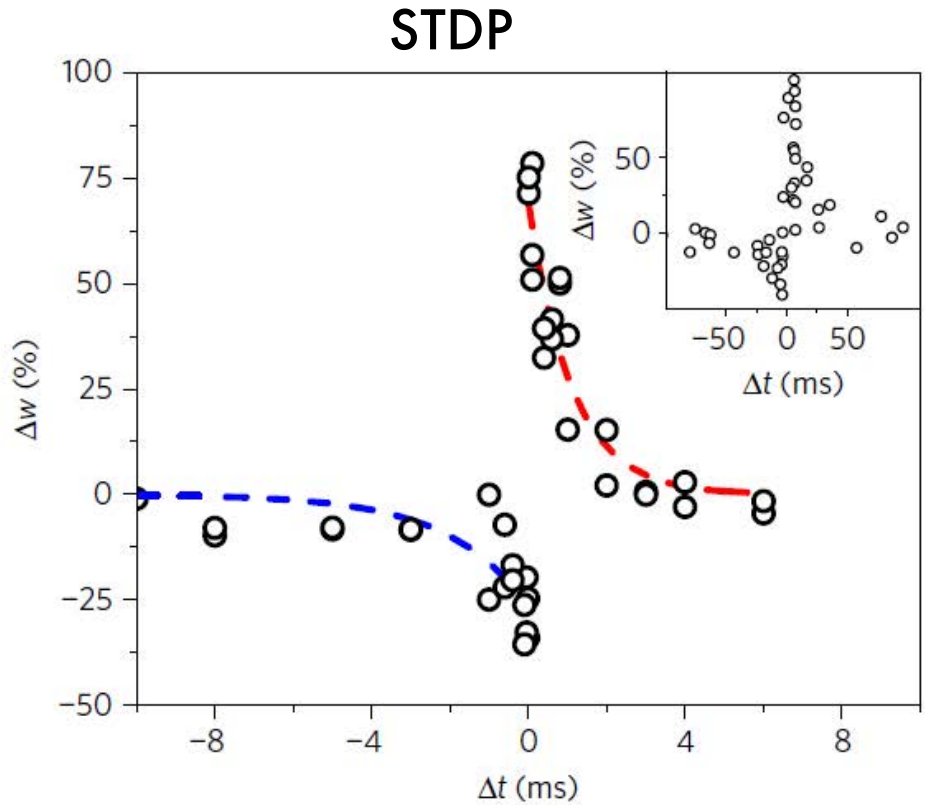
Naturally Reproduce synaptic behaviors long-observed in rat hippocampus!

(Dunwiddie,1977)

Z. Wang *et al.*, Nature Materials 16, 101 (2017)

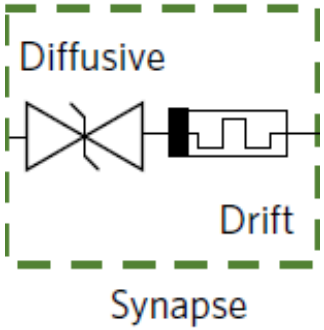
Artificial synapse:
Realizing Long-term plasticity behavior faithfully



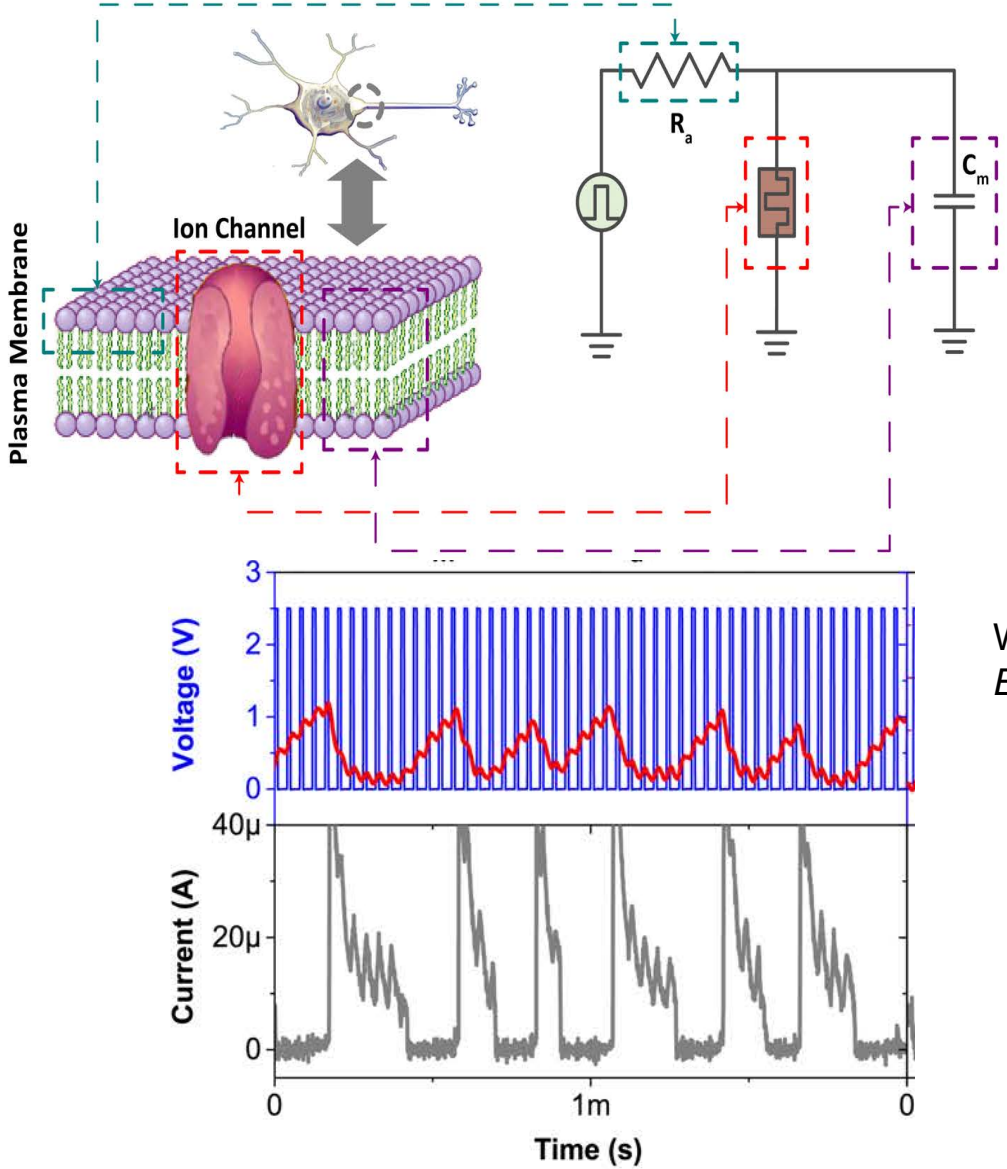


No pulse overlapping or stimulation engineering
Emulation with physical similarity (emulate many functions all together) instead of a pure phenomenological simulation (simulate one function a time)

Z. Wang *et al.*, Nature Materials 16, 101 (2017)



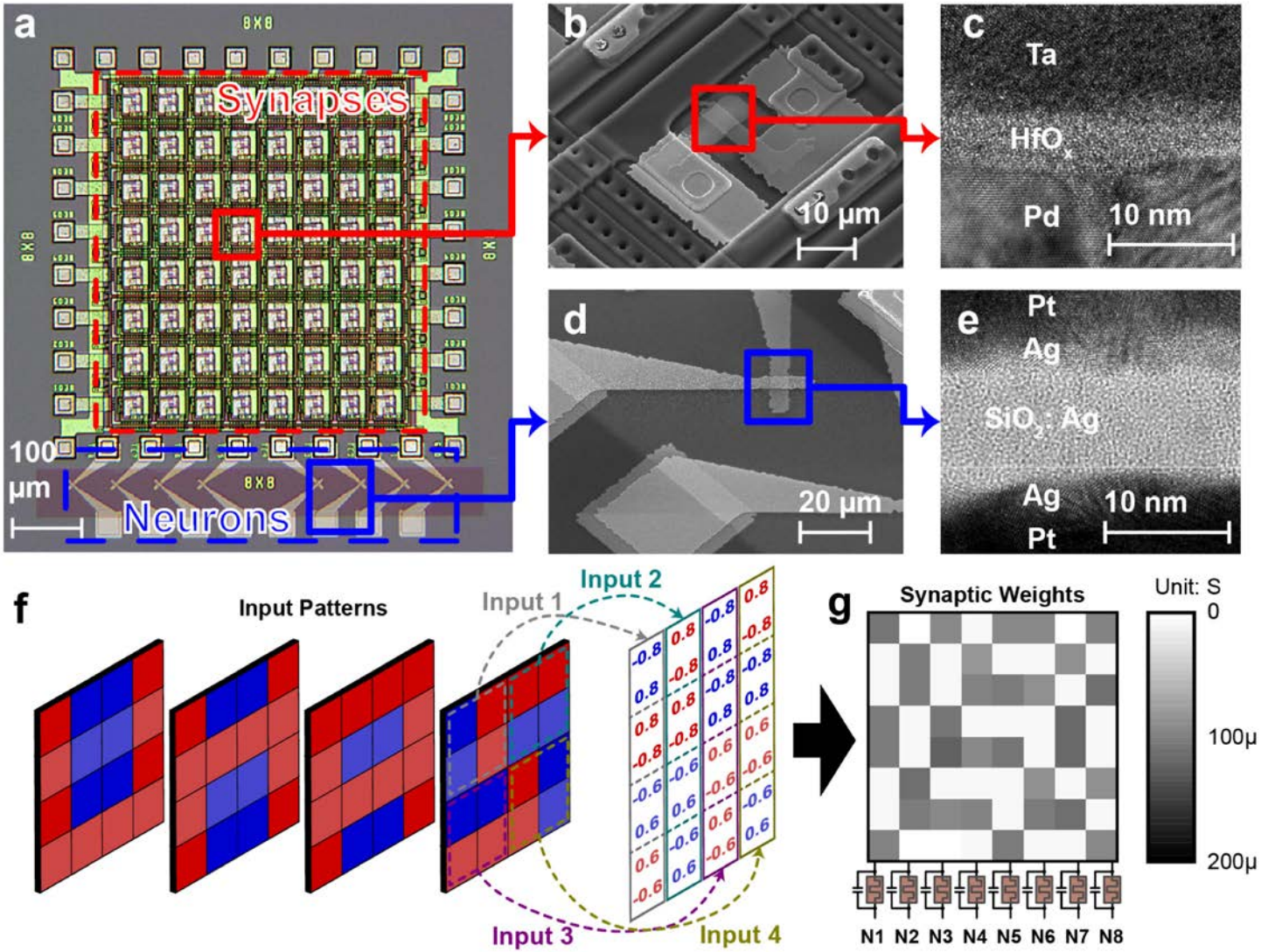
Artificial neuron: leaky integration and fire with dynamics



Wang et al., *Nature Electronics* 1,137 (2018).

From device to neural networks (with thermodynamics)

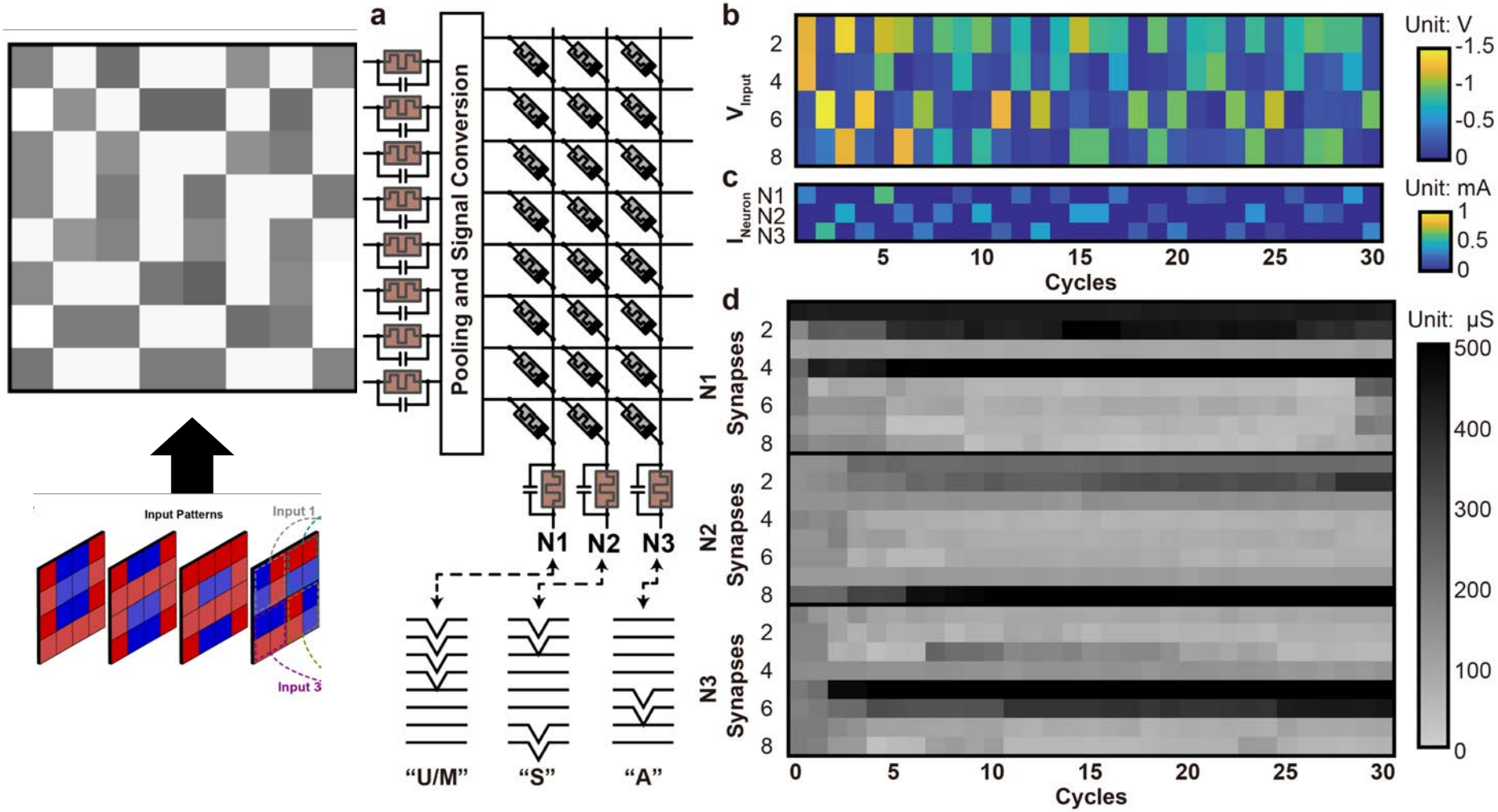
1st Integrated Fully Memristive Neural Network: Pattern Classification with Unsupervised Learning



"UMAS" Pattern classification

Wang et al., *Nature Electronics* 1,137-145 (2018).

1st Integrated Fully Memristive Neural Network: Unsupervised Learning



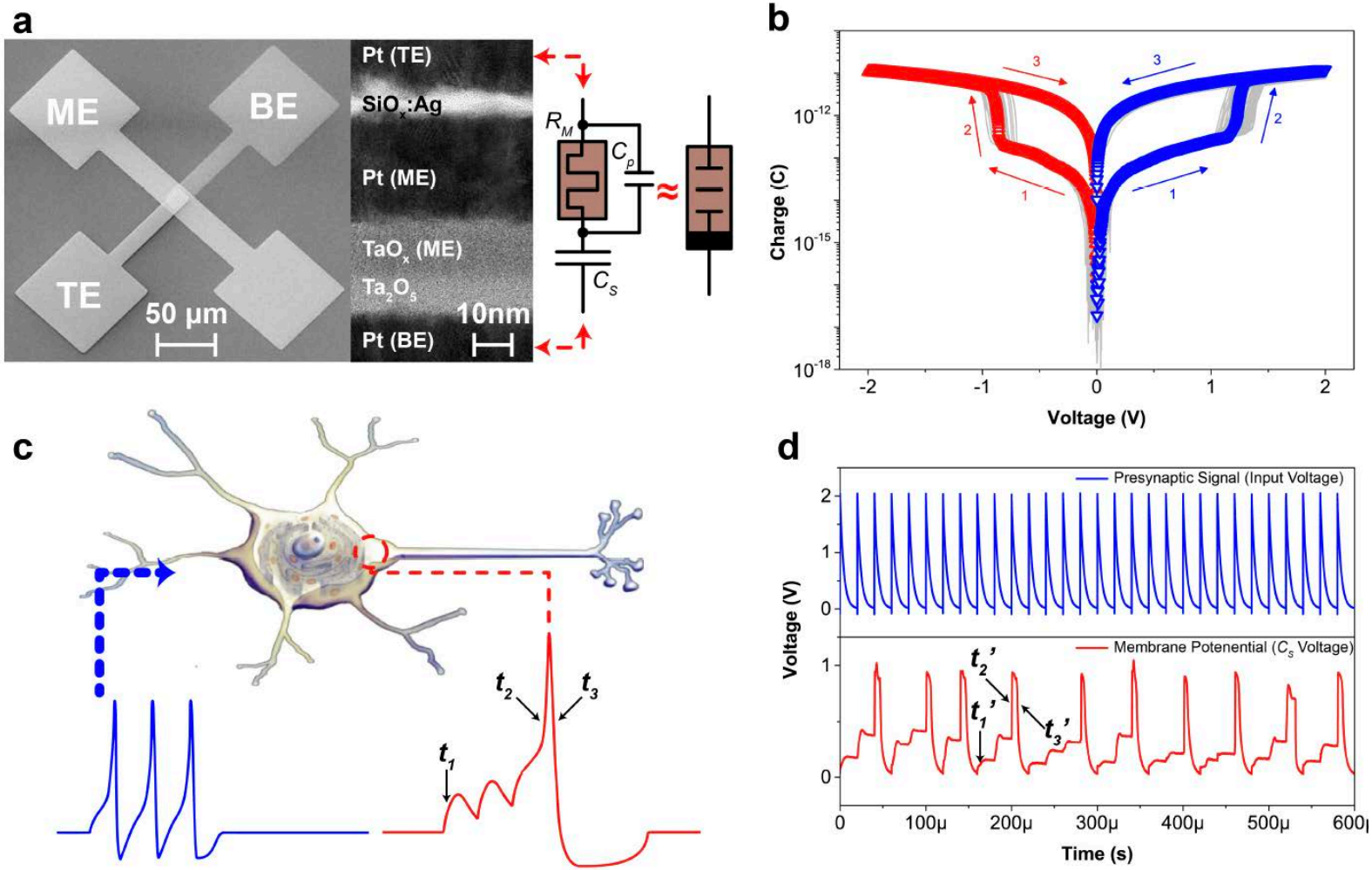
Example of Self-organization

Wang et al., Nature Electronics 1,137-145 (2018).

Capacitive neural network?

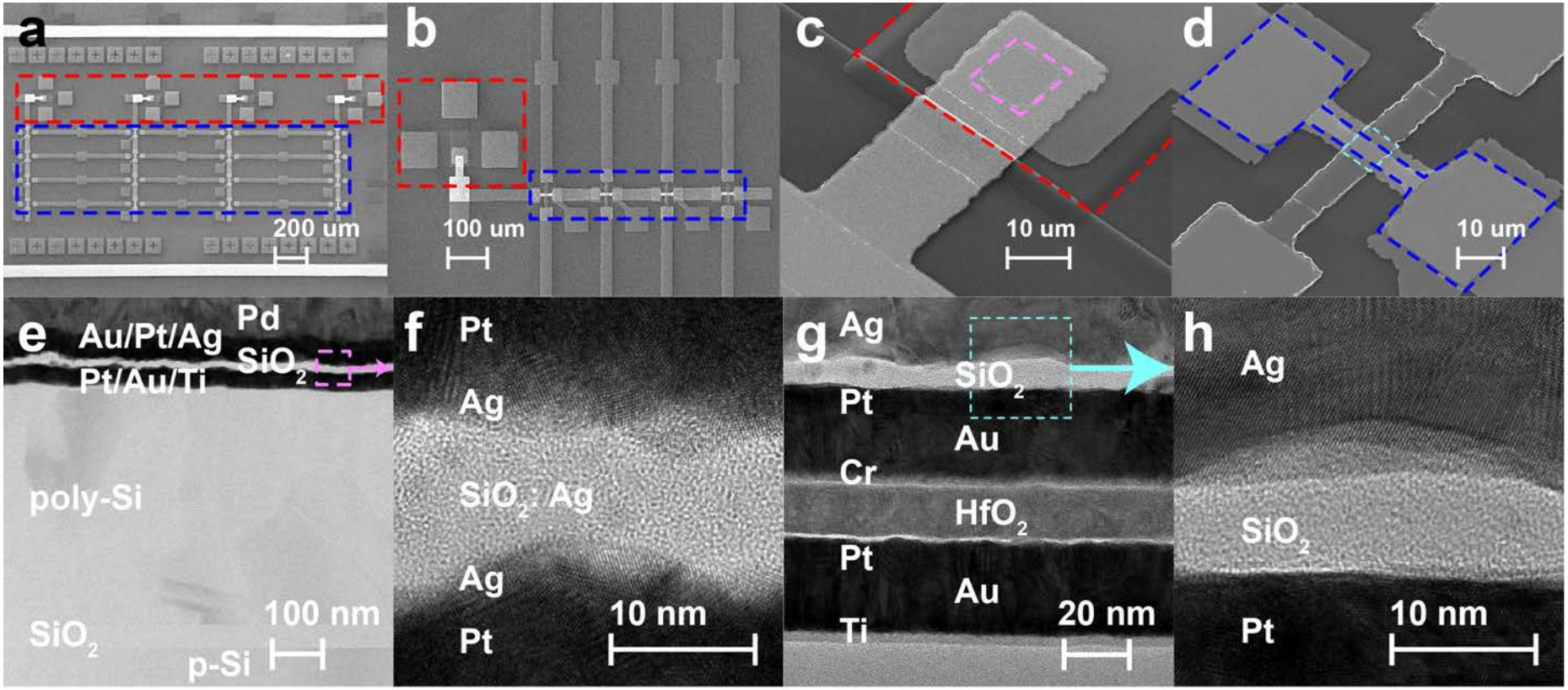
- No static current – less energy into heat;
- Recycle energy of electrical signals;
- Convenient for Spiking neural networks
- …

Pseudo – memcapacitor and neuro-transistor



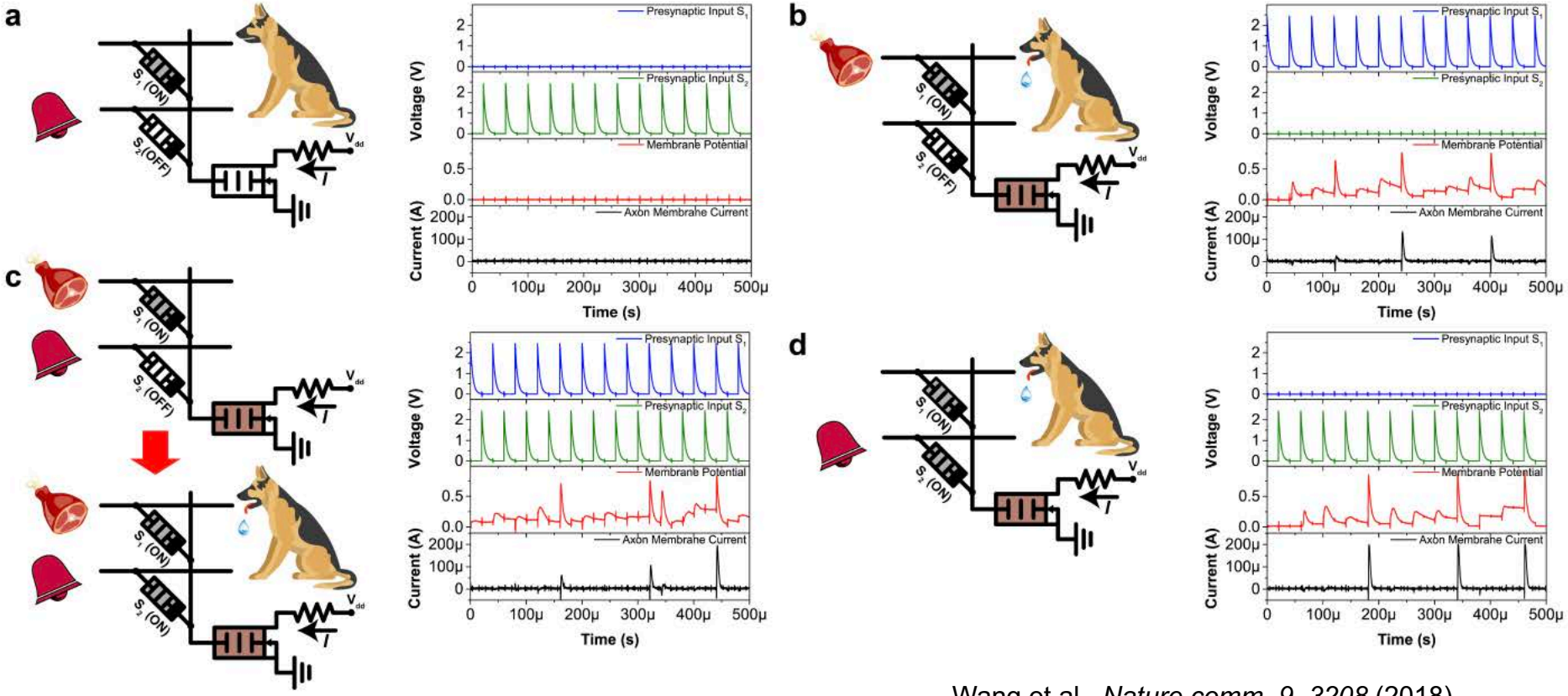
Wang et al., *Nature comm. 9, 3208* (2018).

a fully integrated capacitive neural network



Wang et al., *Nature comm. 9, 3208* (2018).

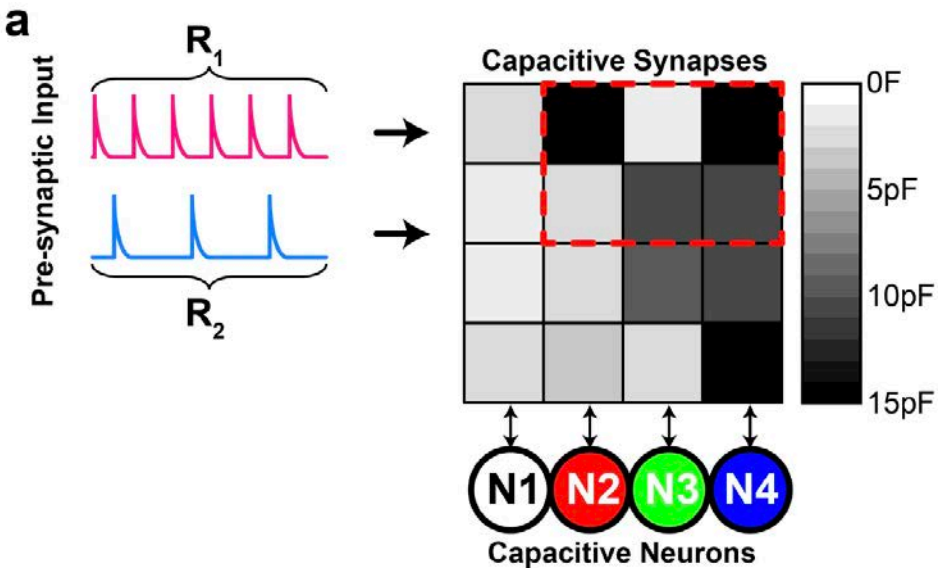
Learning demonstration with capacitive neural network:
Associative learning based on Hebbian-like mechanism

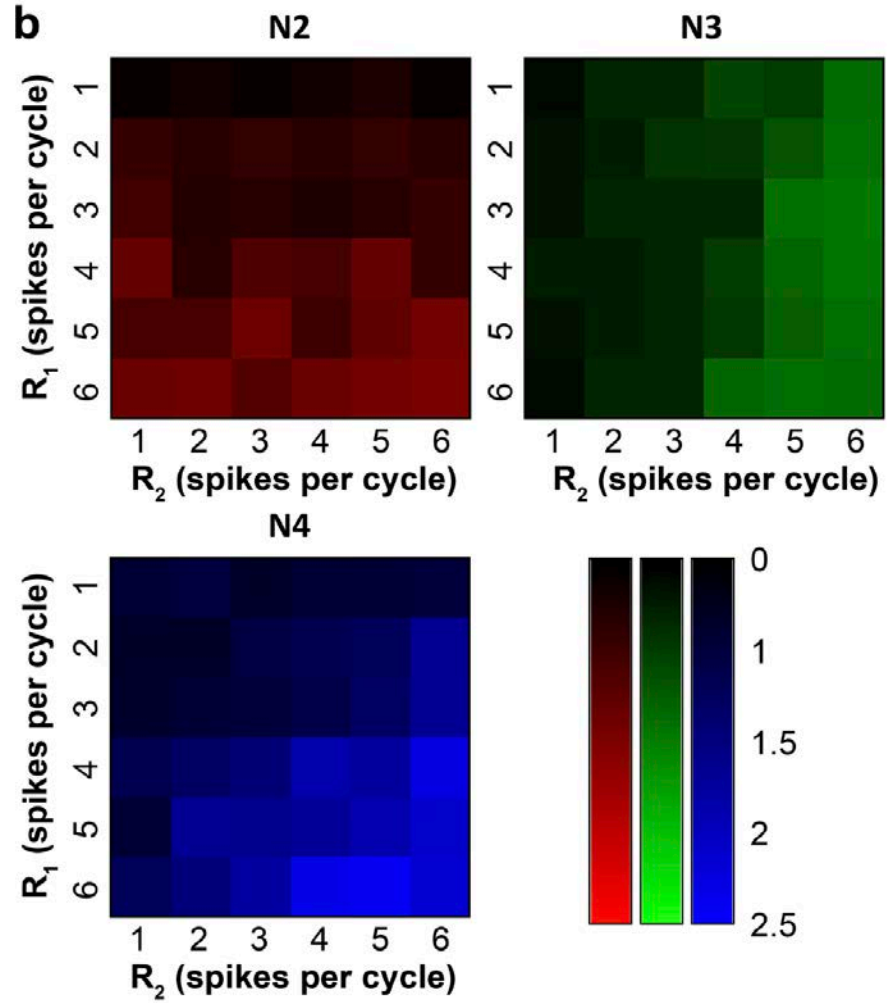


Wang et al., *Nature comm. 9, 3208* (2018).

Inference demonstration with capacitive neural network:
Classification of spiking signals with different frequency

Capacitive version of dot-product



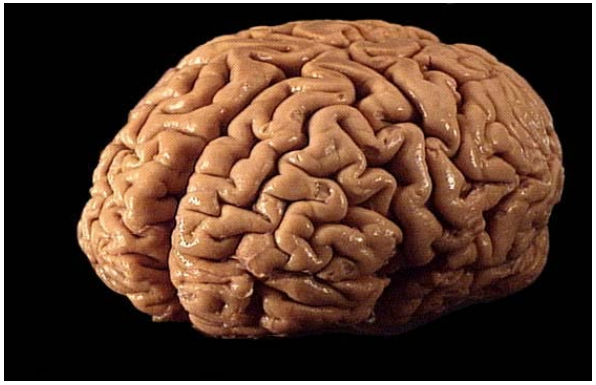


Wang et al., *Nature comm. 9, 3208* (2018).

Other potential physical systems

Implement Neural network with CMOS devices

The last computing frontier

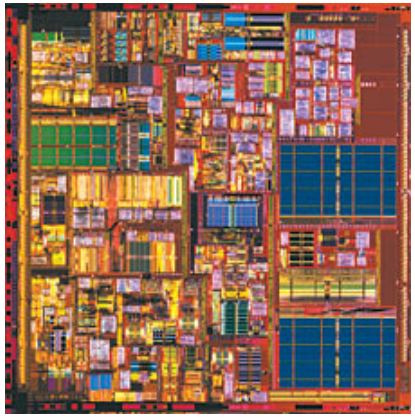


100 W

10^{11} neurons
$10^{14\text{-}15}$ synapses

1 Mimristor = neuron
1 Mimristor = synapse

Possible with nano memristors



CMOS

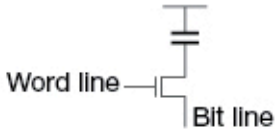
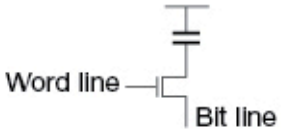
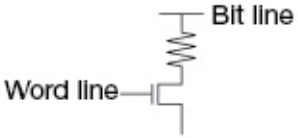
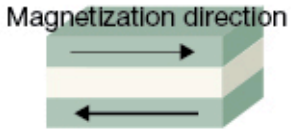
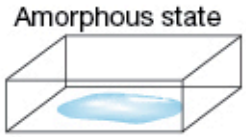
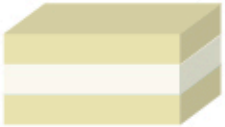
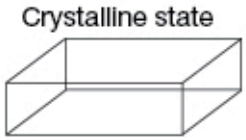
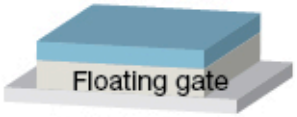
100 W

10^{9} transistors
10 transistors / synapse
10^{8} synapses
10^{5} neurons

Need ~ **10^{7}** chips!!!

Impossible with CMOS chips

Other physical systems: analog memories

	DRAM	FeRAM	MRAM	PRAM	Flash memory
Storage	Capacitor	Ferroelectric capacitor	TMR device	Phase-change device	Floating gate
Cell equivalent circuit					
"1" storage device					
"0" storage device					
Advantages	• High-speed write/read • Low cost	• Non-volatile • Low current consumption • Medium-speed write/read	• Non-volatile • Effectively infinite number of rewrites • High-speed write/read	• Non-volatile • Easy to manufacture	• Low cost • Non-volatile
Disadvantafes	• High standby current • Volatile	• Scalability (not easy to port to smaller line widths) • Destructive read	• High write current	• High write current • Fewer rewrites than MRAMs, FeRAMs, etc • Rewrite slower than MRAMs, FeRAMs, etc	• Slow write • 10^5 or fewer rewrites

Take home messages

- **Bio-intelligent systems** may have a significant component of thermodynamic computing.

- Simple computing, such as inference after supervised learning, can be accelerated by neural networks of memristors **without much thermodynamics** involved. (**just a programmable network**)

- More advanced computing, such as pattern classification with unsupervised learning, can be achieved in neural networks of memristors functioning **based on thermodynamics**. (possible to generate **real intelligence**)

- Traditional Si devices were **not** created or optimized for thermodynamic computing; emerging devices, e.g. memristor, MRAM, FeRAM, PCM, FLASH are promising.

- Higher **energy efficiency** and **throughput** as well as **real intelligence** might be expected in those new computing paradigms.

Acknowledgements:

Postdoc and students: Zhongrui Wang, Saumil Joshi, Jungho Yoon, Rivu Midya, Mingyi Rao, Wenhao Song, Navnidhi Upadhyay

Special acknowledgement: Prof. Qiangfei Xia, UMass, Amherst

Collaborators:

- **John Paul Strachan, Z. Li, R. Stan Williams, Hewlett Packard Labs**
- Matthew J. Marinella, Sandia National Laboratories
- Sergey E. Savel'ev, Loughborough University, UK
- Huolin L. Xin, Brookhaven National Laboratory
- Hai (Helen) Li, Duke University

Acknowledgements:

Sponsors:

IARPA

AFRL (Rome)

Hewlett Packard Enterprise

HPE

DARPA