RRAM for Future Memory and Computing Applications

Ming Liu

Key Lab. of Microelectronic Devices &Integrated Technology, (CAS)

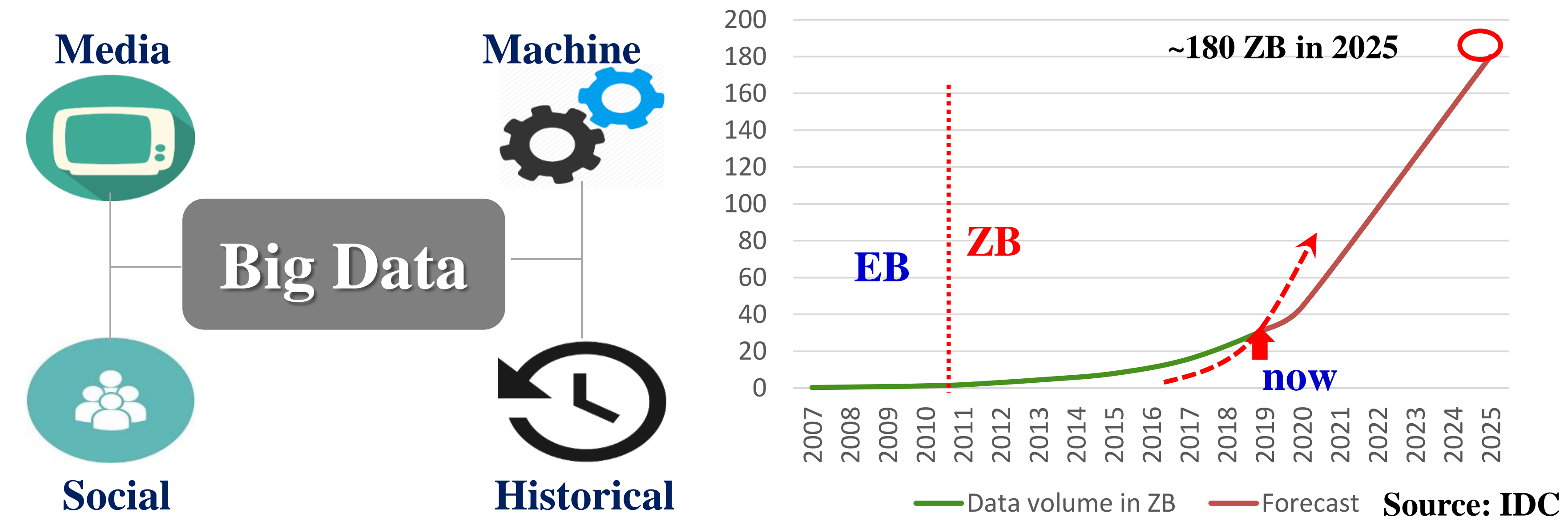
Institute of Microelectronics, CAS

Macao University, July7.2018

Outline

- Computing System's Challenge
- RRAM: Memory/Storage Convergence
- RRAM: Memory/Computing Convergence
- Summary

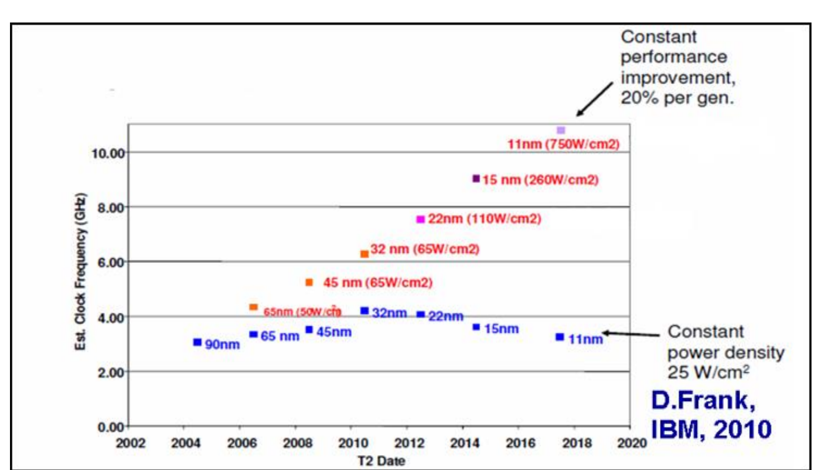
Big Data comes

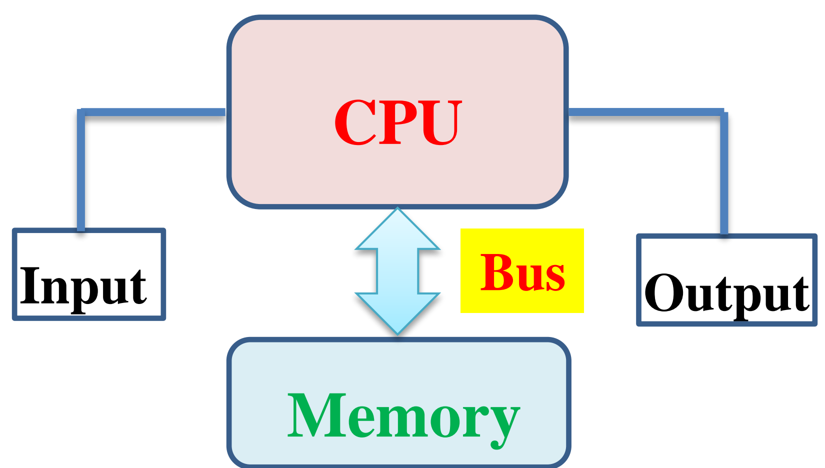


- From 2013, data nearly **doubles every two years**
- In 2025, it's expected that the data volume will reach **~180 ZB**

Powerful and Energy Efficiency Computing to Process Big Data!

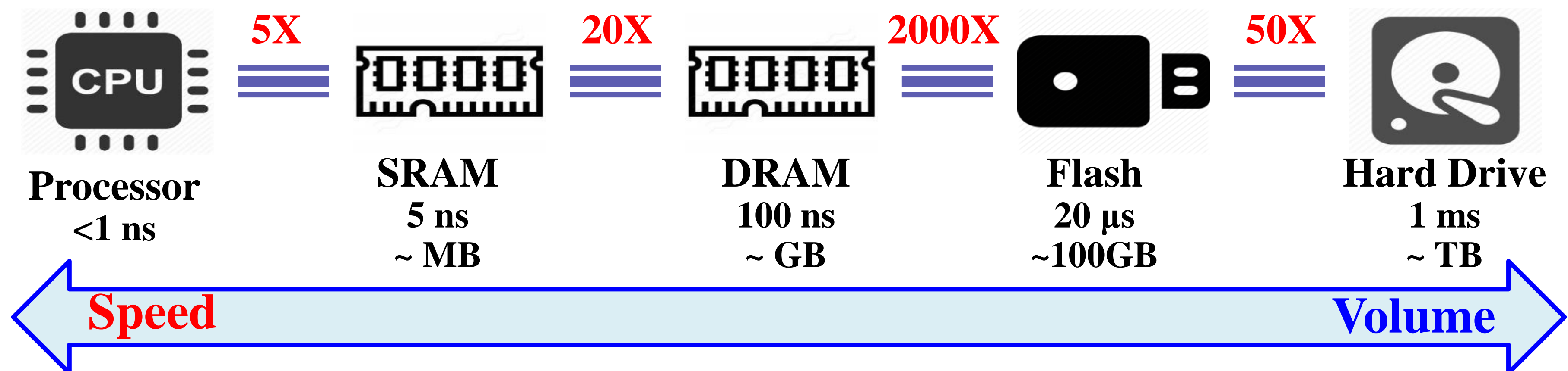
Computing challenge in Big Data era





- Computing performance improved 10^{10} times in past 60 years.
- **Device:** energy efficiency slow down due to power constraints at 22nm;
- **Architecture:** CPU and memory was physically separated.
- An increasing performance gap between CPU and memory, which is known as the memory wall.

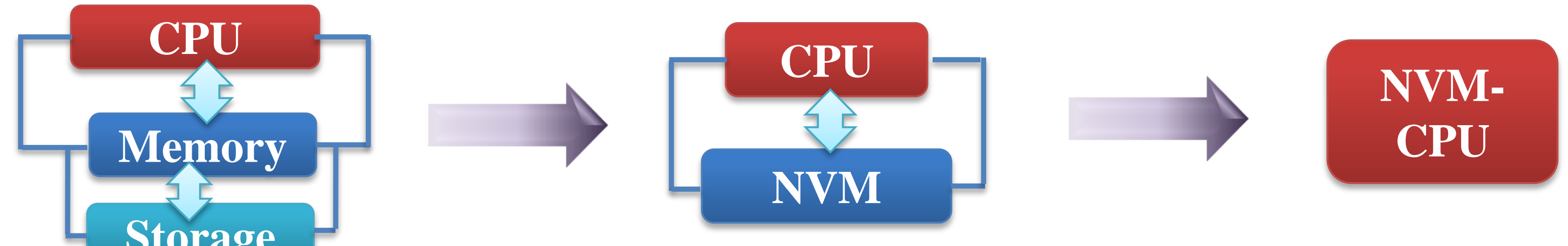
Memory hierarchy



- Memory Hierarchy: tradeoff between speed and density, bottleneck to limit the computing performance.
- Universe memory: blurs distinction between memory (fast, volatile and low density) and storage (slow, non-volatile and high density).

NVM play more important role in future computing!

NVM: a solution to future computing



Memory Hierarchy: more date movement

M/S Convergence: less data movement

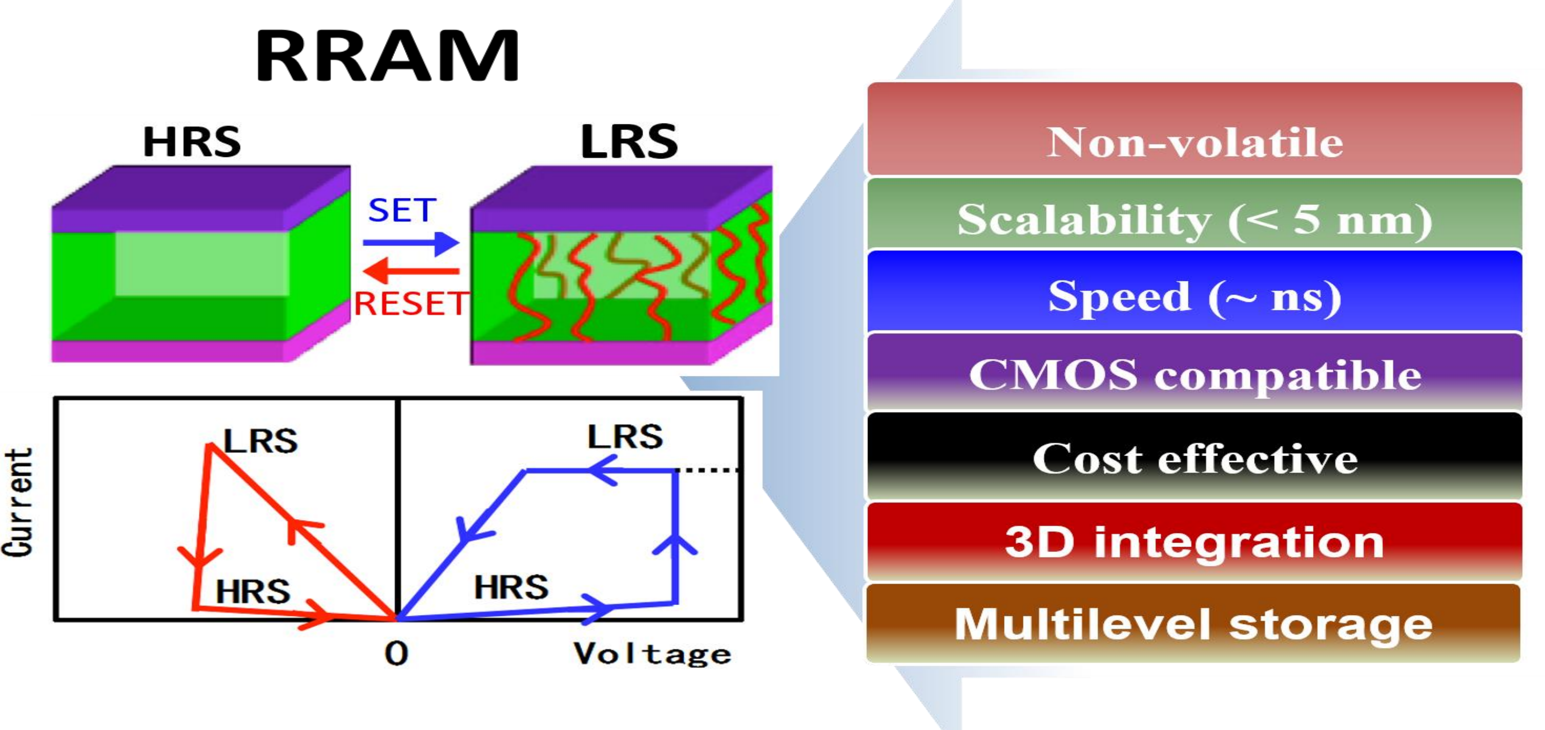
M/C Convergence: no data movement

- **Near term: M/S convergence by new NVM, simplifying memory hierarchy, less data movement, high performance;**
- **Long term: M/C convergence by integrating memory and computation in one device, "Memory Wall" problem can be solved.**

Outline

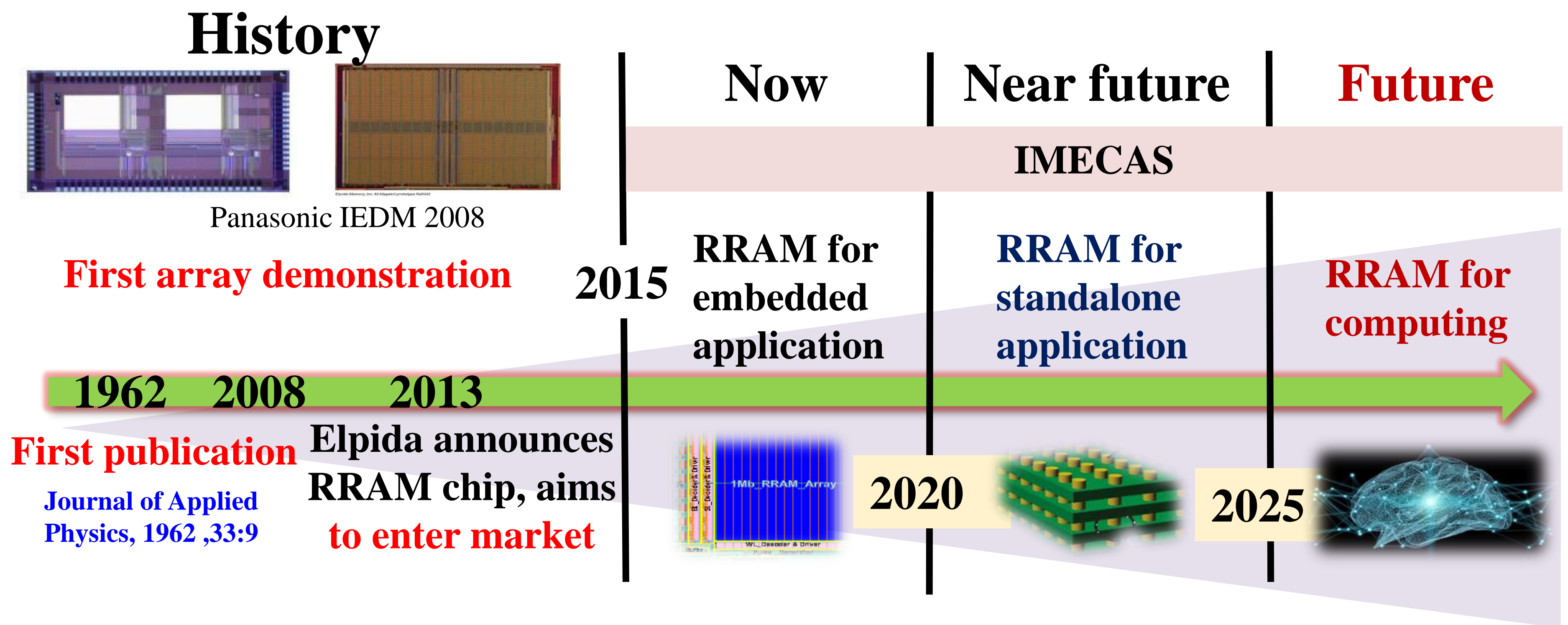
- Computing System's Challenge
- RRAM: Memory/Storage Convergence
- RRAM: Memory/Computing Convergence
- Summary

RRAM: a promising candidate



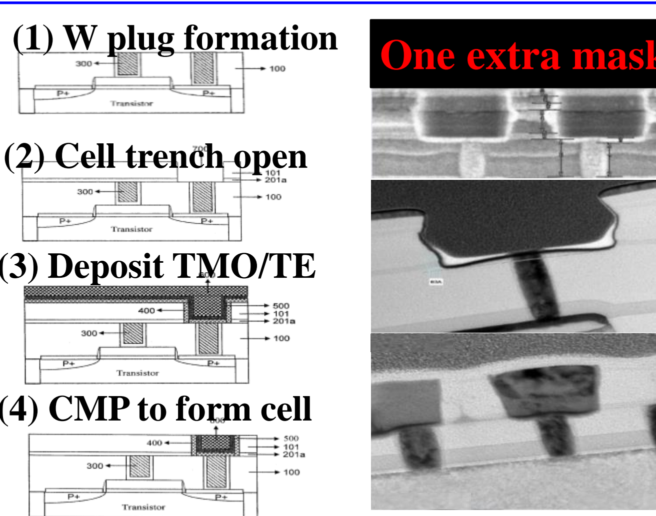
RRAM: Promising Emerging Memory Technology !

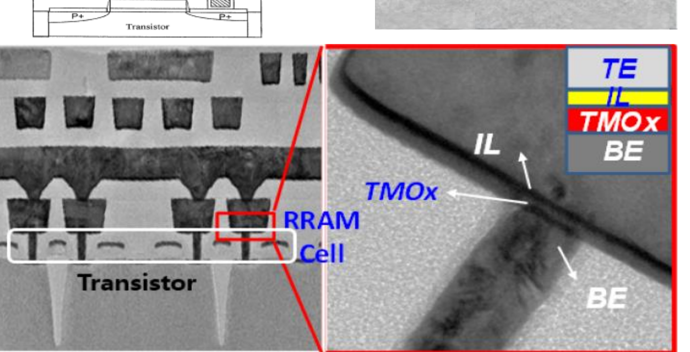
RRAM's history and future



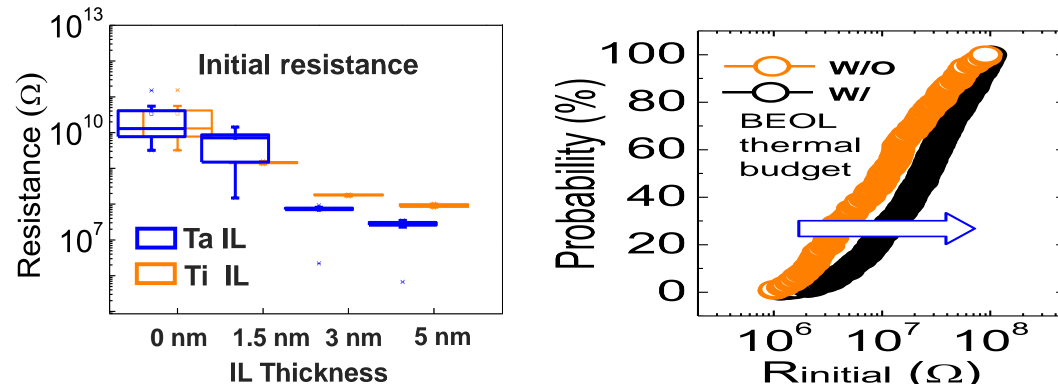
Our group started the joint development of RRAM in embedded and stand-alone applications with industry from 2015.

28 nm RRAM integration





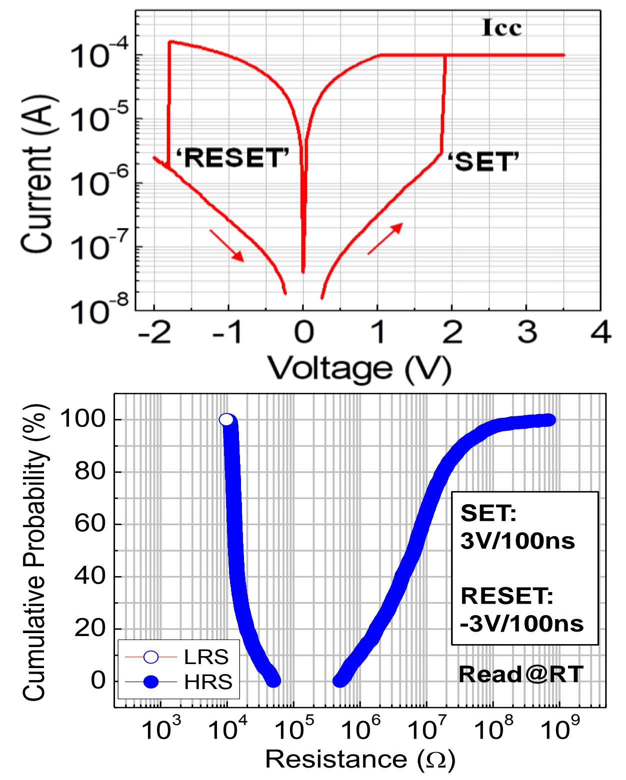
US Patent, 8735245



- RRAM built between CT and Metal 1, W as BE, TMO as switching layer and M1 as TE
- Interfacial layer between TMO and TE:
  - Initial Resistance more uniform and sensitive to the thickness of Ti or Ta ;
  - Block TE migration at BEOL thermal process. PECVD, High T Q time, Ashing, alloy. More than 90 min annealing at 400°C .

IEDM 2017, 36-39

Electrical test on the 1T1R array

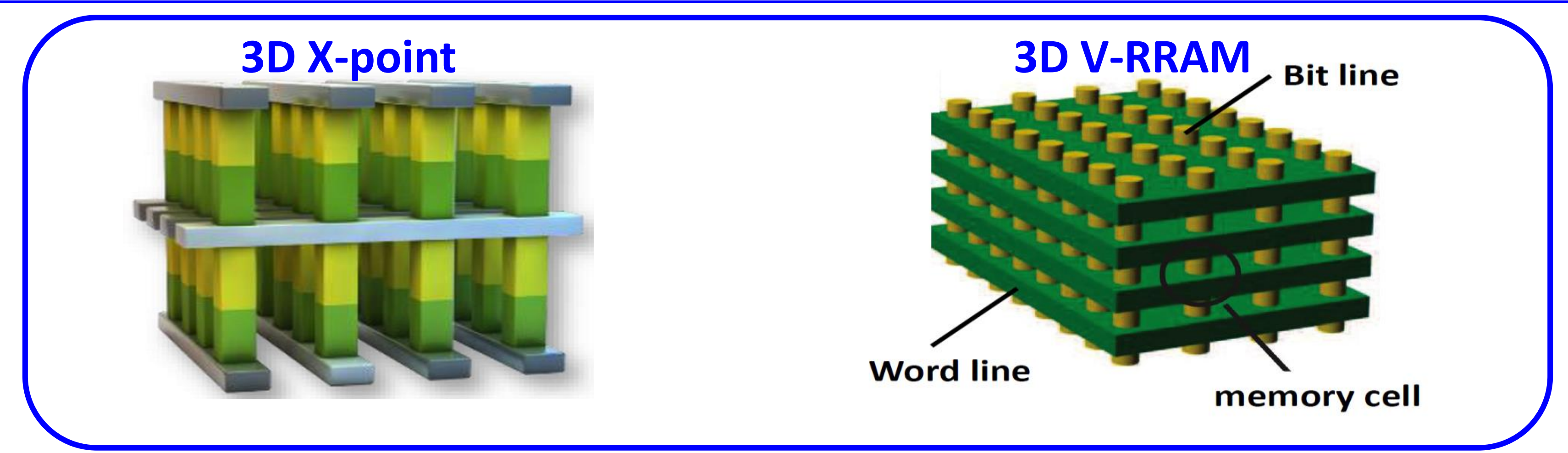


HRS and LRS distribution in 1Mb array by 3V/100ns SET and -3V/100ns RESET pulse.

Category		1T1R RRAM
Device structure	Switch layer Material	TaOx
	Electrode Material(BE/TE)	Cu/W
Forming		1.5~3V
VSet(V)		0.8 V~1.5V
VReset(V)		-0.5 V~-1.5V
R_HRS/R_LRS		>100
Retention		10y@85C
Cycling		1 M
Cell Size		40nmx40nm
Technology node		28nm
Memory array size		1kb, 1Mb
Processing temperature		<400C
Drop-out Cause		Stuck at LRS

IEDM 2017, 36-39

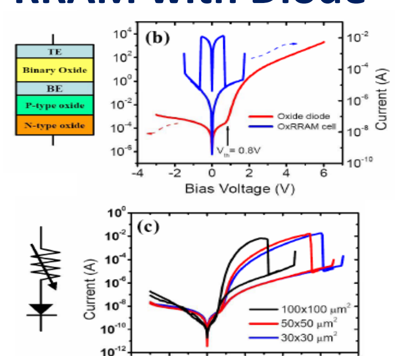
RRAM: a good choice for 3D stacking



- Suitable for 3D integration; either in 3D X-point or BiCS 3D NAND like vertical array (VRRAM).
- RRAM devices linear I-V in LRS, unselected cells in LRS, sneaking current could be generated.
- A high performance nonlinear selector or self-selective RRAM cell.

Solution for the sneak current issues

RRAM with Diode



- p-n type diodes, Schottky diodes, Hetero-junction…
- Generally, applying to the unipolar RRAM
- rectifying ratio is defined as R_{-V}/R_V

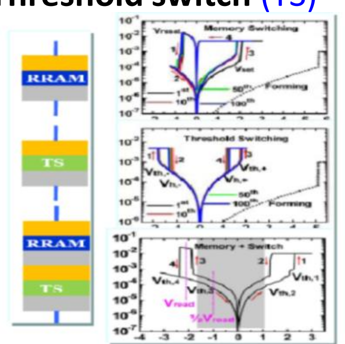
W/TiO_x/Ni diode with self-compliance to integrate bipolar Cu/HfO_2/Pt

Nanoscle, 2013, 5:4785

RRAM with switch- based selector

Mater insulator transition (MIT)

Threshold switch (TS)



- threshold switch be as volatile switch
- applying to the unipolar or bipolar RRAM
- rectifying ratio is defined as $R_{V/2}/R_V$

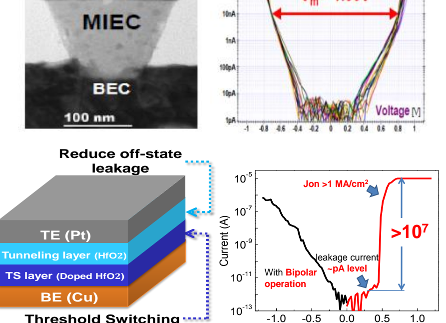
Pd/TaO_x/Ta/Pd with non-linearity of 5×10^3

Nanoscle, 2015, 7:4964

Resistive switch:

Mixed ionic electronic (MIEC)

Complementary resistive switch structure

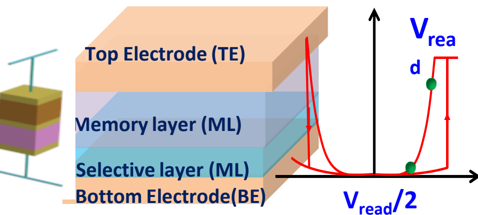


- applying to the unipolar or bipolar RRAM
- rectifying ratio is defined as $R_{V/2}/R_V$

Cu doped HfO_x with non-linearity >10^7, Jon>1MA/cm²

IEDM 2015, 245-248

Self-Selective Cell



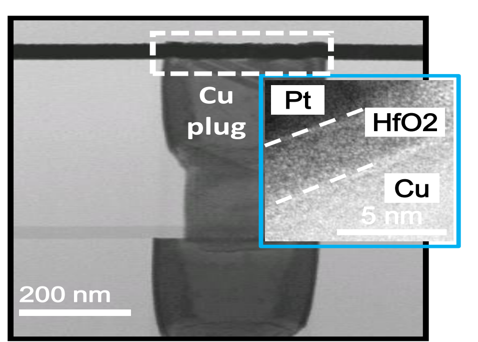
- Hybrid selective layer and memory layer
- Nonlinearity ratio is defined as I@Vread/Vread/2
- The only choice for 3D Vertical RRAM.

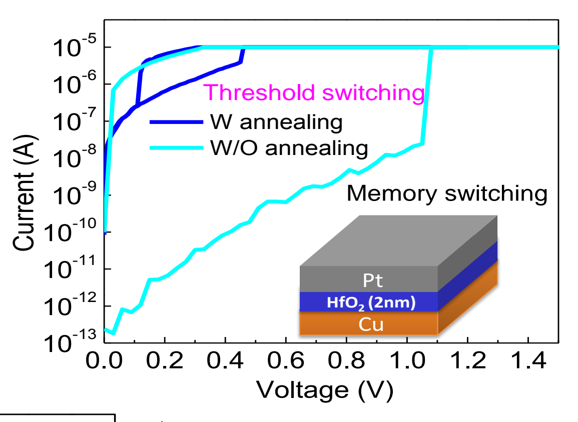
Self-rectifying RRAM: Pt/WO_3/a-Si/Cu

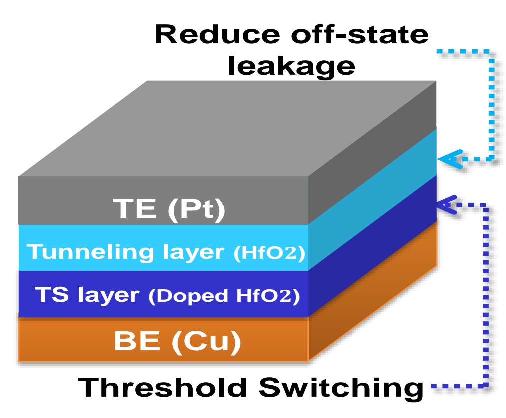
Self-rectifying Au/ZrO_2:nc-Au/n+-Si

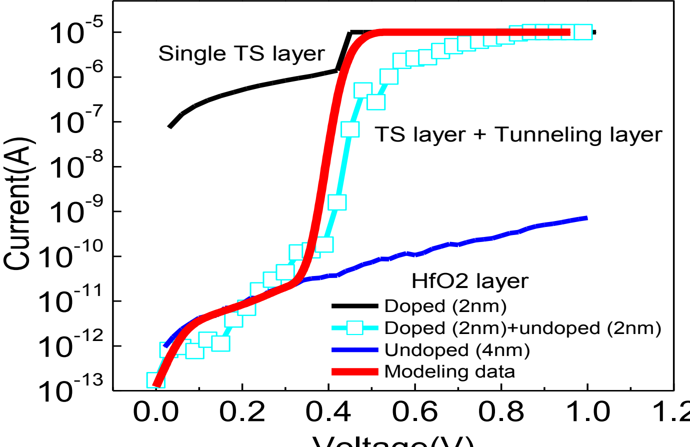
JAP, 2009, 106:073724; IEEE EDL, 2010, 31:344; IEEE EDL, 2013, 34:229

Threshold Switching in Cu doped HfO_x





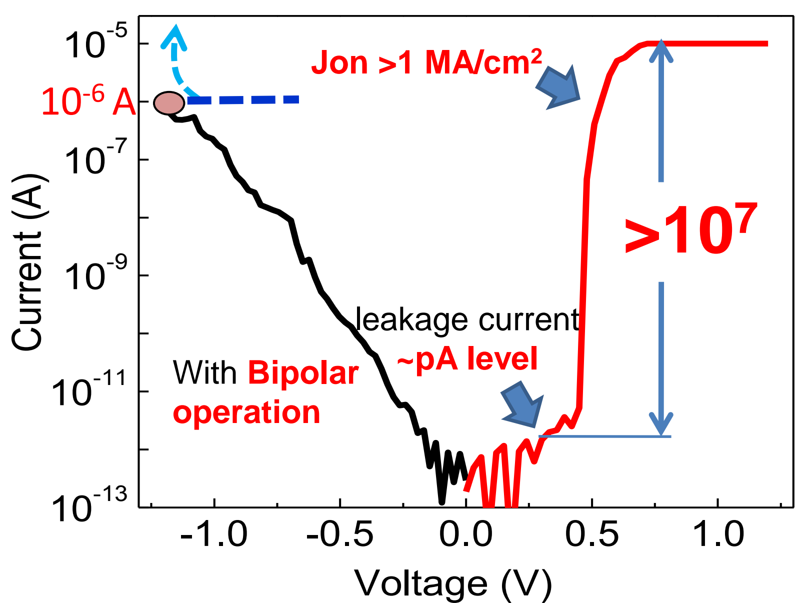


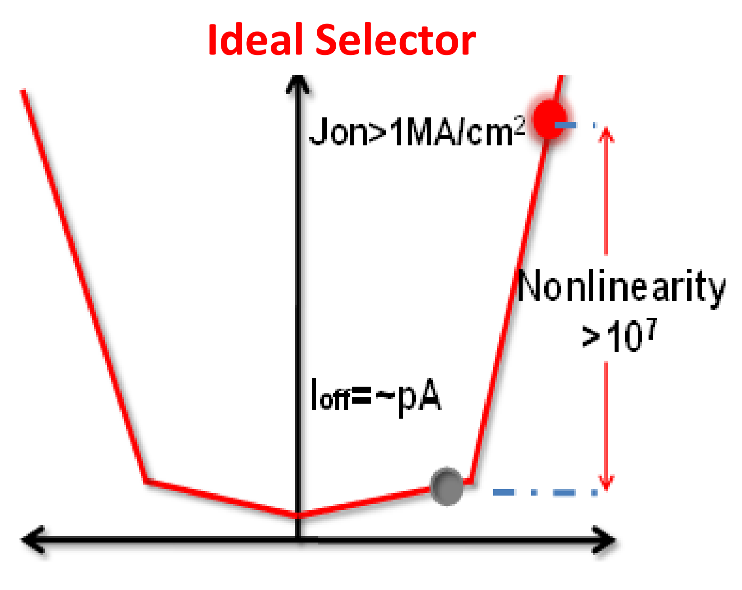


- **Cu doped HfO_x RRAM fabricated in 130nm BEOL. TS observed after annealing 30min at 125^0C.**
- **Introducing the 2nd tunneling layer, the leakage current was reduced by 5 orders.**

IEDM 2015, 245-248

Bilayer Selector Device

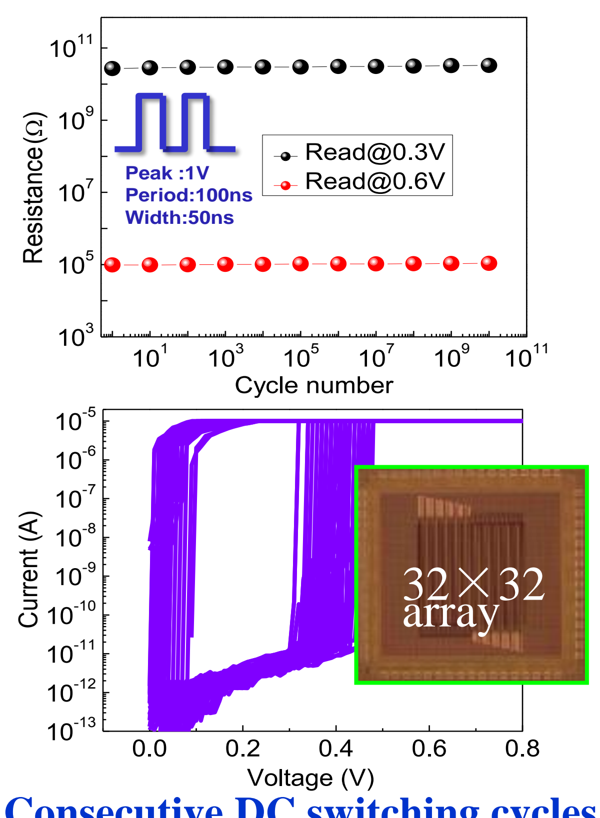




- ➢ Non-linearity $>10^7$, Jon>1MA/cm^2, Leakage current: pA level.
- ➢ Asymmetrical I-V curve might be resulted from the barrier height between top electrode and the tunneling layer.

IEDM 2015, 245-248

Selector Array



Consecutive DC switching cycles

@85°C

read@0.3V
read@0.6V

Stress time(s)

Current(A)

Distribution of on and off states current

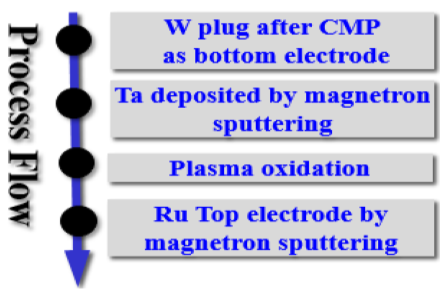
Cumulative Probability

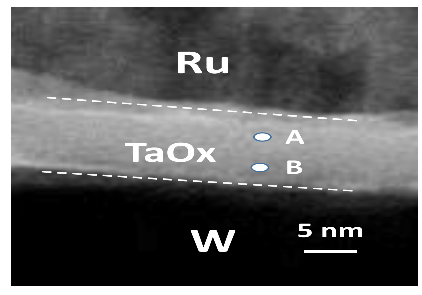
Current(A)

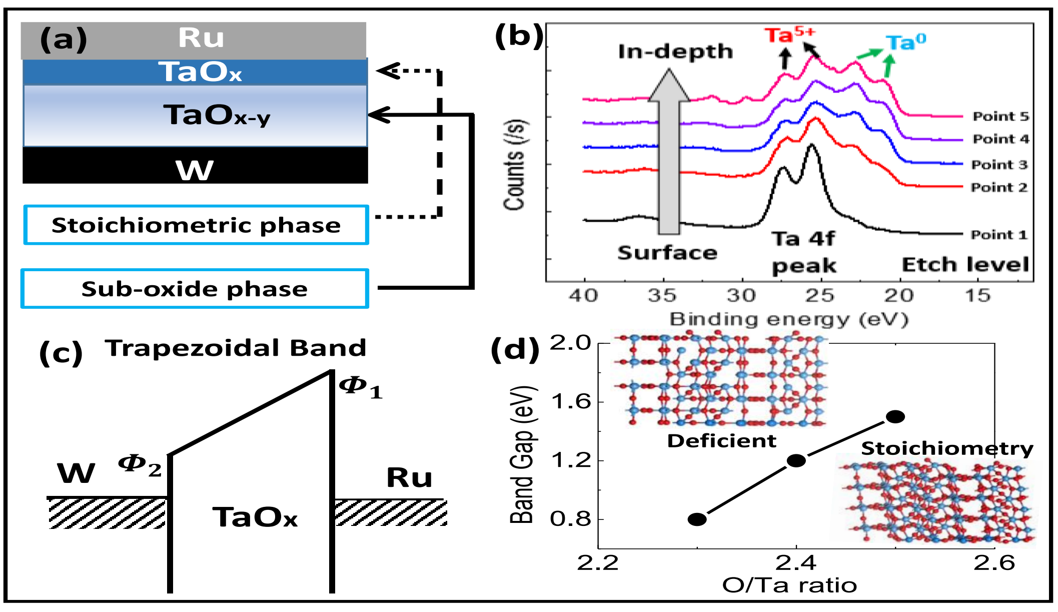
- **Endurance: 10^{10}, high temperature without degradation.**
- **1 kb selector array with 1T-1S: High nonlinearity, High on-current density, tight distributions on and off current.**
- **Switching voltage variation, limited voltage window for reading.**

IEDM 2015, 245-248

Interface type selector

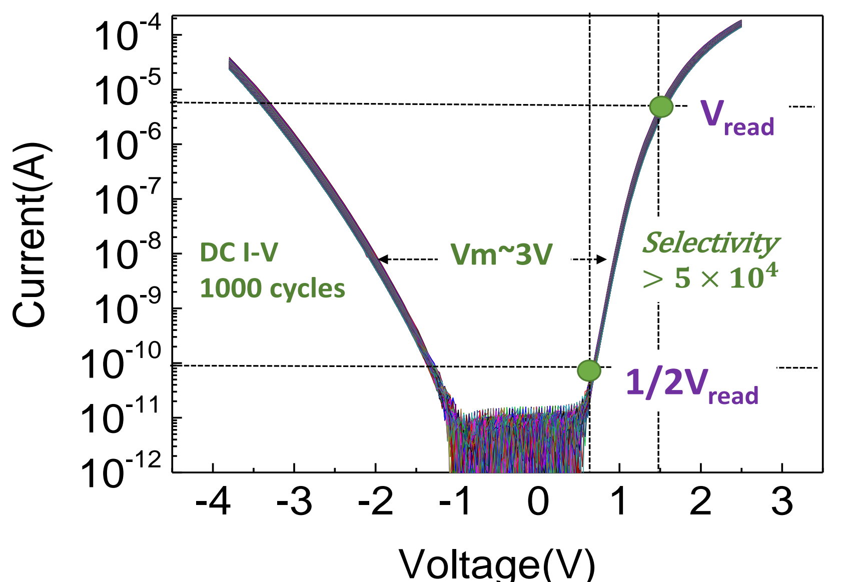


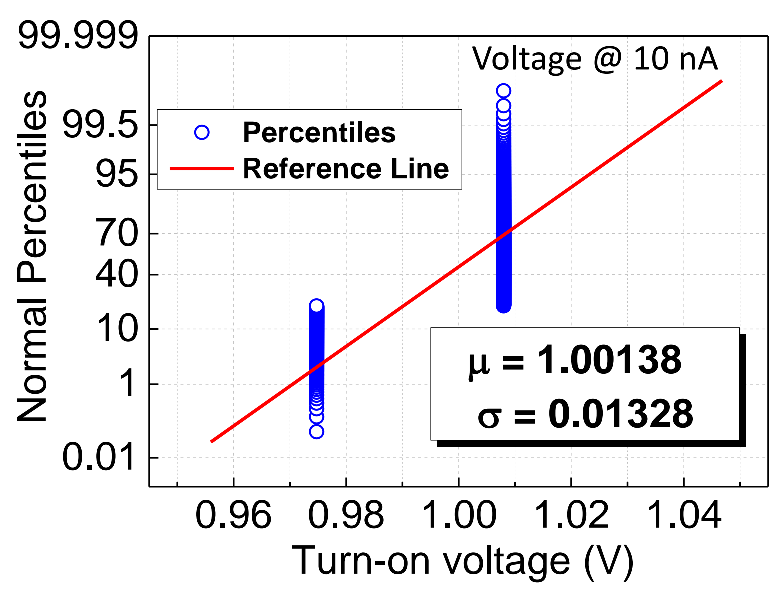




- Good uniform and larger R window , Lower nonlinear and on-current.
- Trapezoidal band shape: high nonlinear and on-current compared to uniform or crested barrier.
- O- gradually changed TaO_x layer: in surface, Ta was fully oxidative, oxygen component decreased as depth increased.

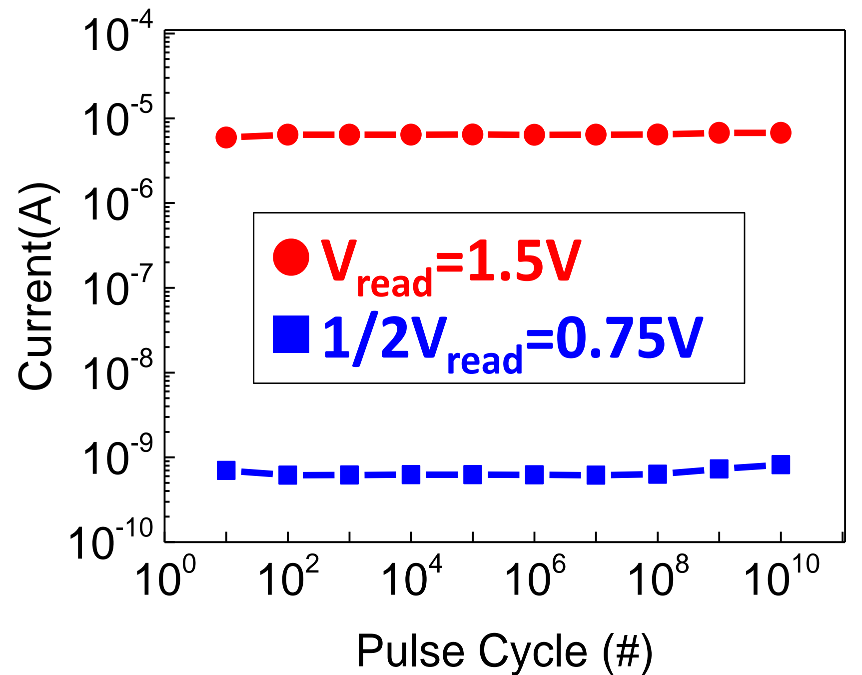
Device performance

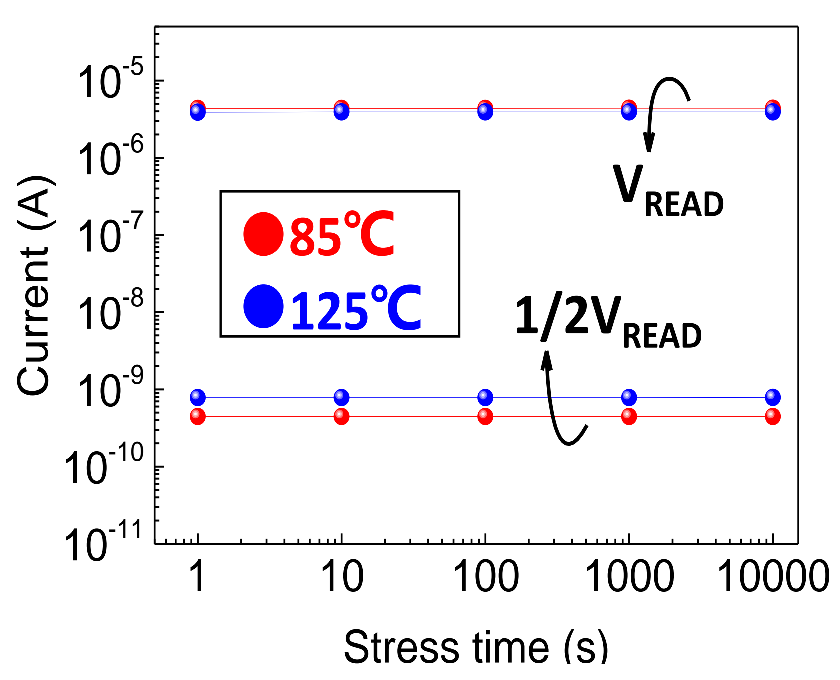




- **A higher current density (~1 MA), high selectivity (~5×10^4), lager voltage margin V_M (3V) achieved.**
- **After 10^3 successive DC cycles, each I-V curve is almost overlapped, standard deviation is negligible, showing excellent uniformity.**

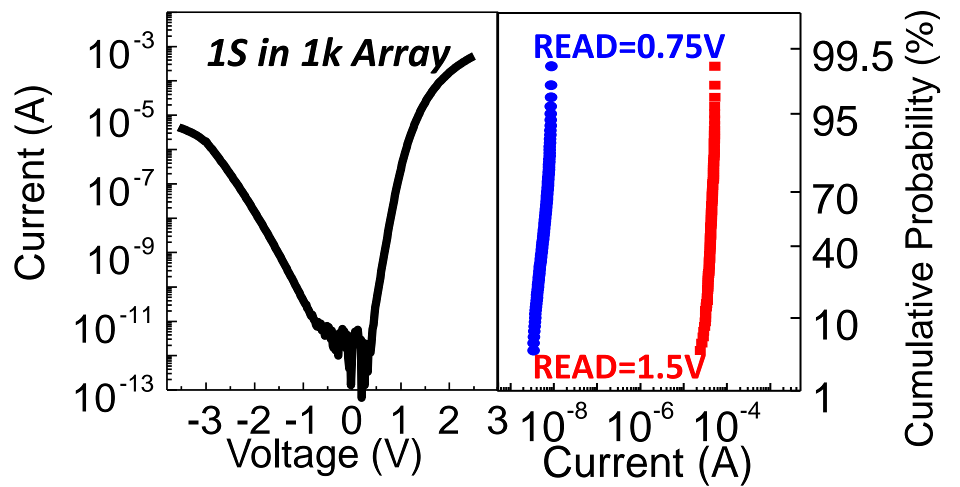
Endurance and High T Operation

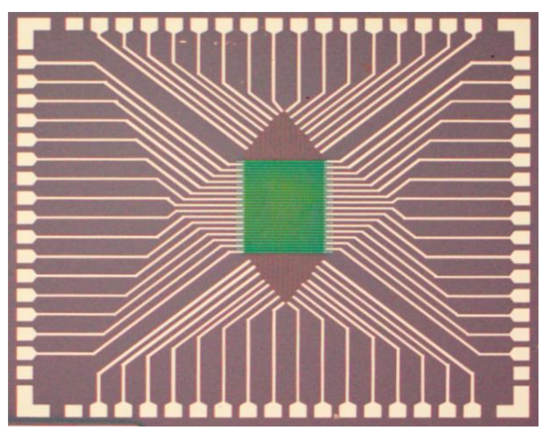


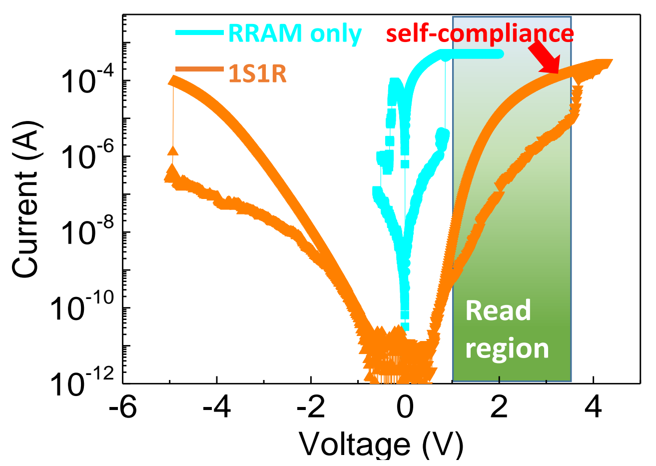


- Endurance as high as 10^{10} has been achieved.
- High temperature without degradation is allowed.

1S 1R integration in 1kb Array

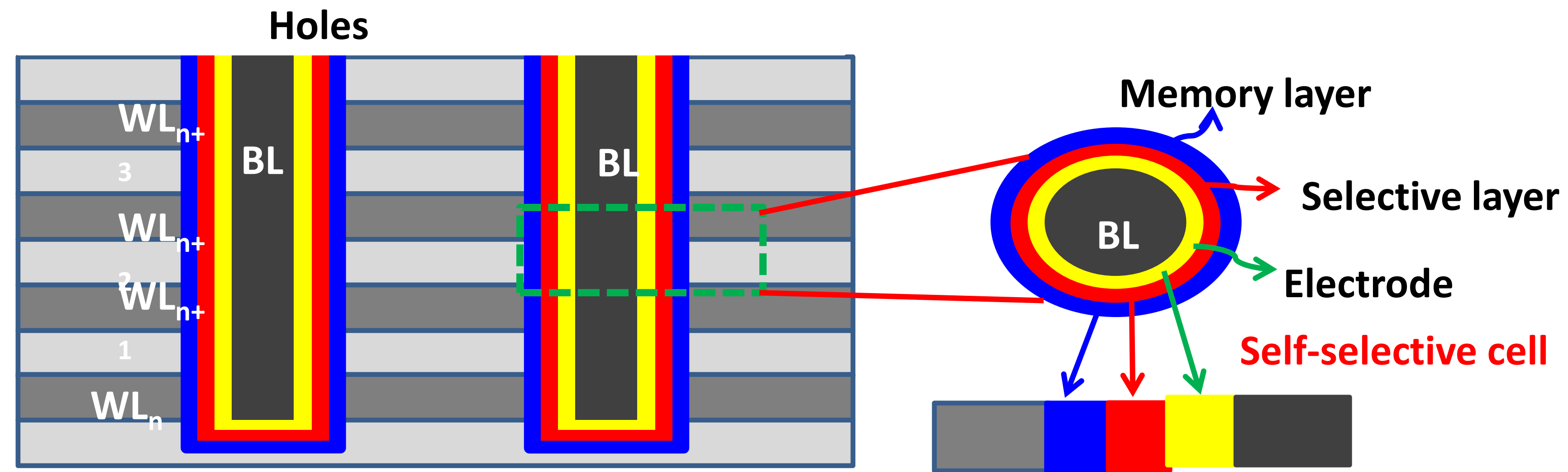






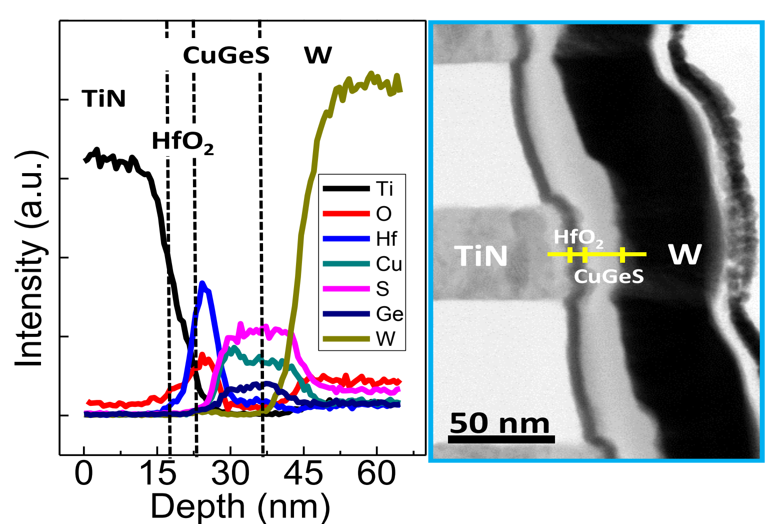
- 10^4 nonlinearity was achieved in 1S1R with excellent uniformity.
- The read region is from 1.2V to 3.8V.
- The read region with nonlinearity higher than 10^3 is from 1.2V to 2.4V.

Self-selective Cell (SSC) for VRRAM

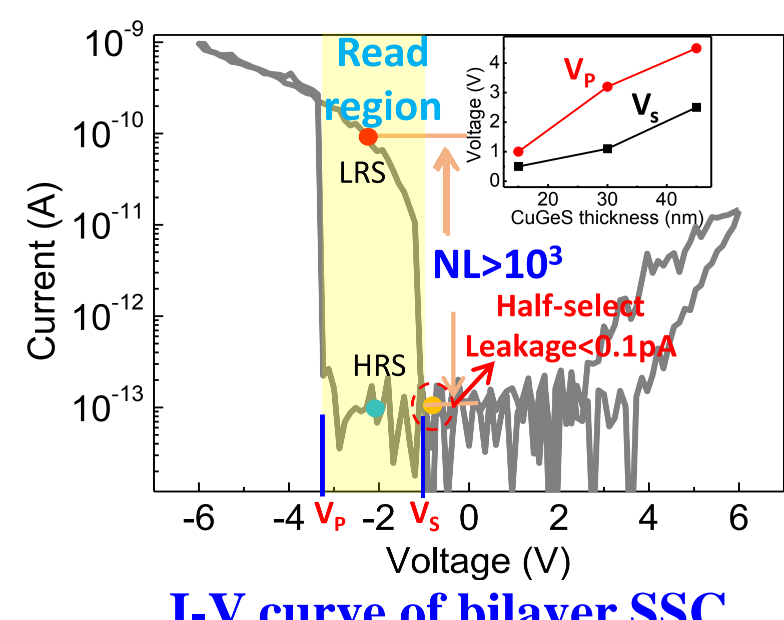


- In 3D VRRAM, intermediate electrode is not allowed, memory cells on the same BL will be shorted, connecting with the same selector.
- The self-selective memory cell with rectifying or build-in nonlinearity is the only choice for 3D VRRAM.

Typical I-V curve of bilayer SSC



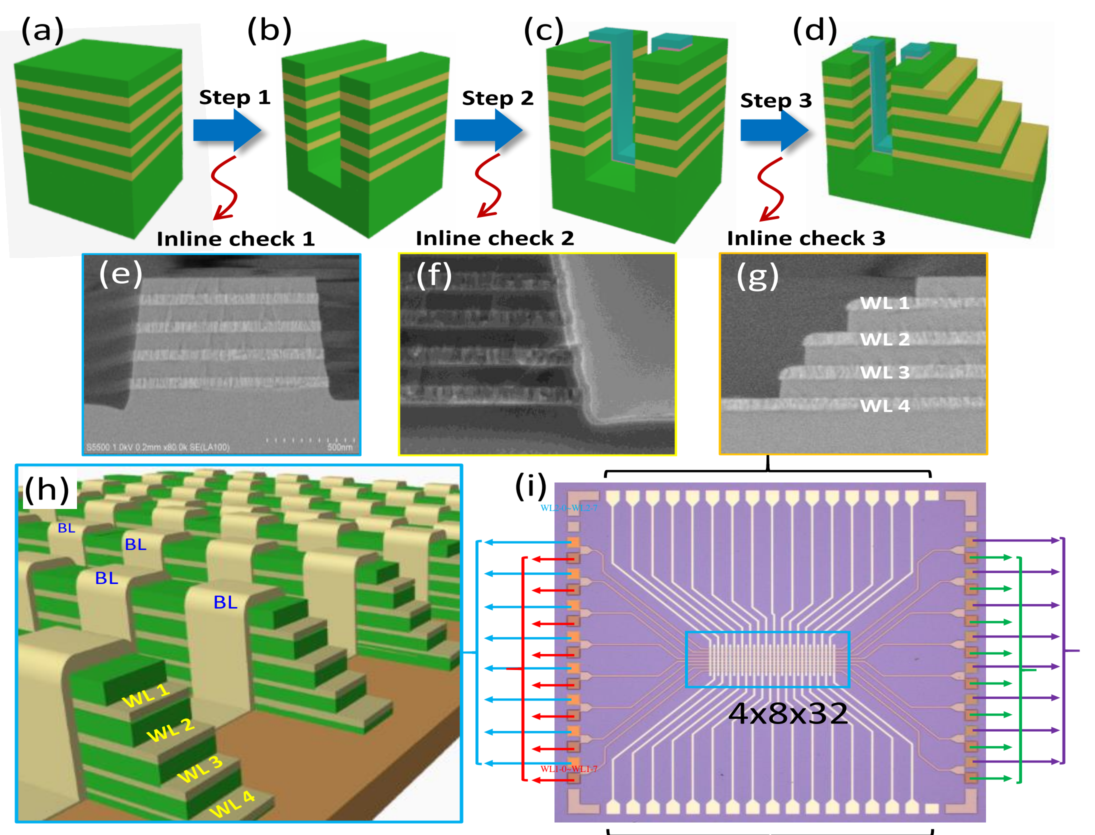
TEM and EDS of bilayer device



I-V curve of bilayer SSC

- Ultra-low half-select leakage (<0.1 pA)
- Very high nonlinearity (>10^3)
- Low operation current (below nA)

3D VRRAM Integration of SSC

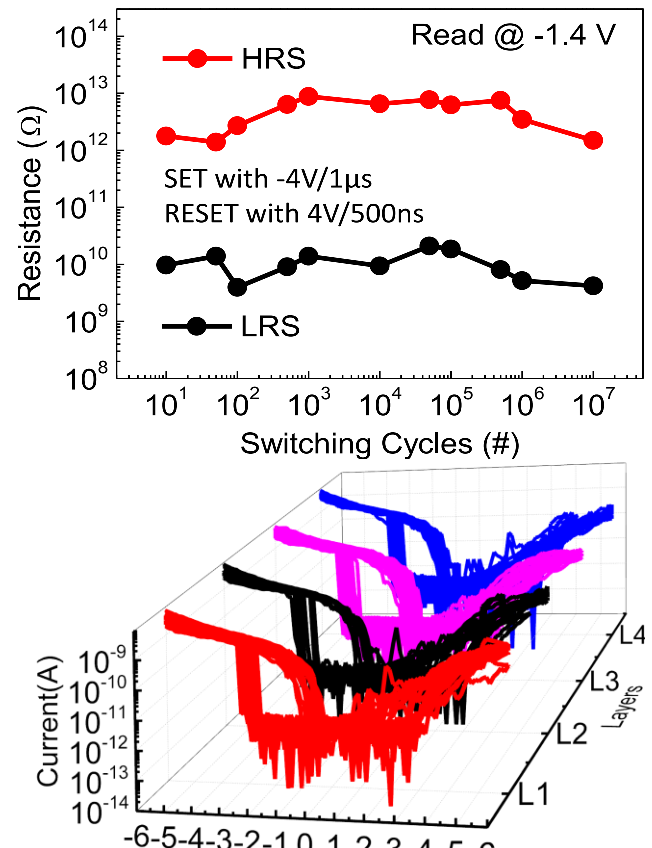


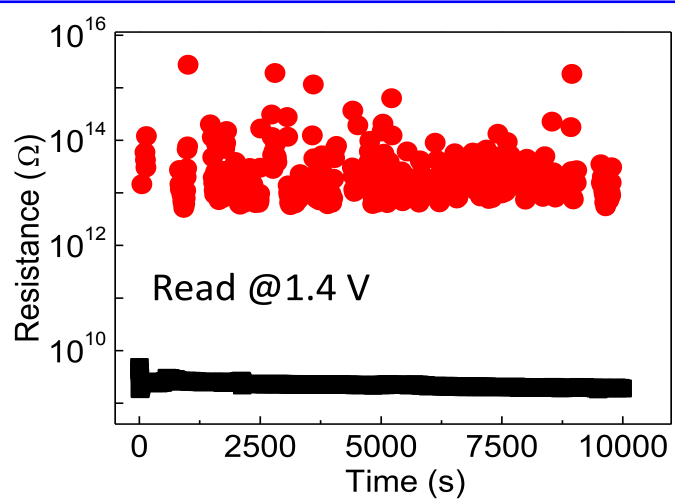
4 layer 8×32 3D VRRAM array

- HfO_2/CuGeS bi-layer SSC with TE deposited on sidewall by sputtering.
- Each horizontal WL was opened by selective etching.
- Staircase WL contacts on each layer are formed.

IEDM 2015, 245-248

Reliability test of SSC

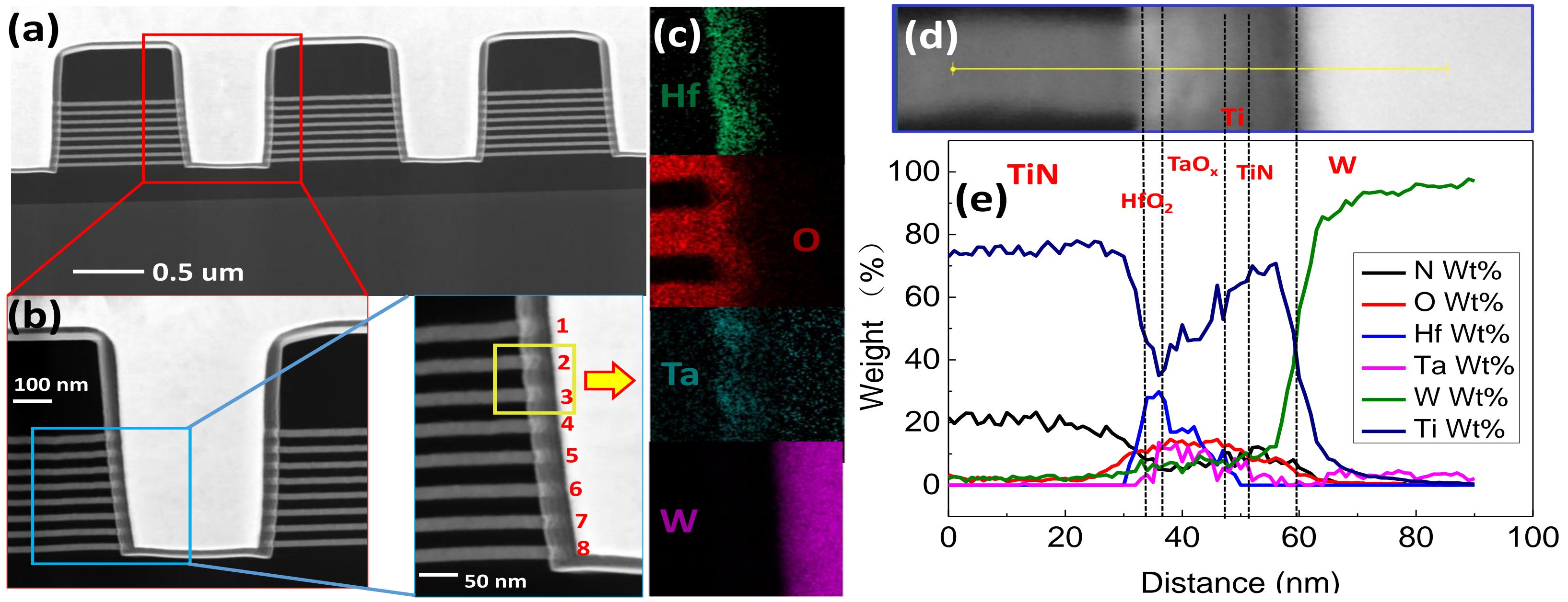




- Endurance of SSC with 10^7.
- Retention of SSC for 10000s.
- Each layer devices exhibit stable and uniform characteristics.

IEDM 2015, 245-248

8 layer integration of 3D VRRAM



- An 8-layer integration of 3D VRRAM achieved.
- High uniformity with on/off ratio (≈100 times) and 100x nonlinearity.

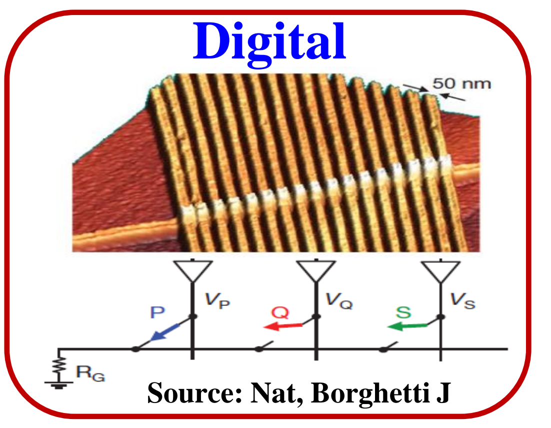
IEDM 2017, 48-51

Outline

- Computing System’s Challenge
- RRAM: Memory/Storage Convergence
- RRAM: Memory/Computing Convergence
- Summary

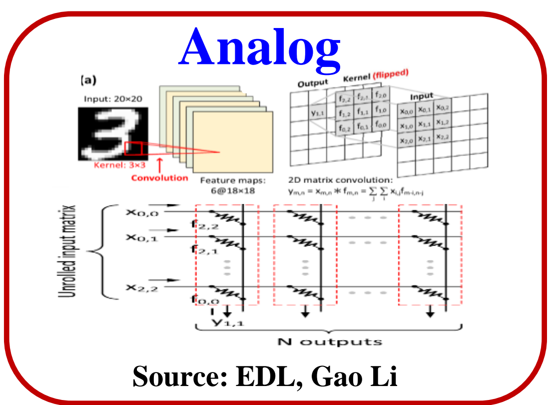
Ways of Memory/computing convergence

Digital



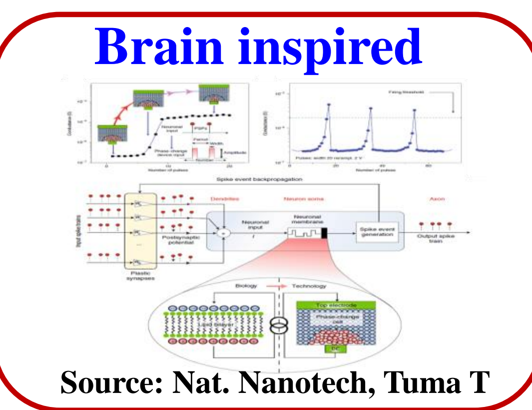
Source: Nat, Borghetti J

Analog



Source: EDL, Gao Li

Brain inspired

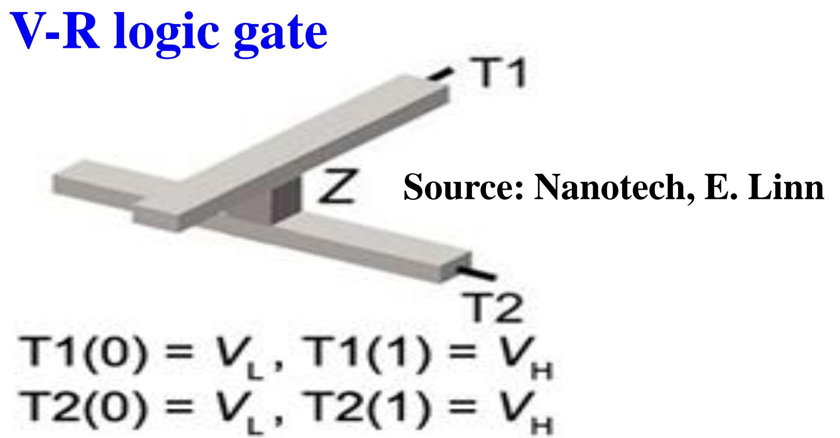


Source: Nat. Nanotech, Tuma T

- **In memory computing, to eliminate the energy-intensive and time-consuming data movement.**
- **Focused on identifying novel logic gate concepts with lower energy and area consumption.**
- **RRAM's advantages, as direct access by interconnect lines, capability to electrically reconfigure device, and nanoscale miniaturization.**

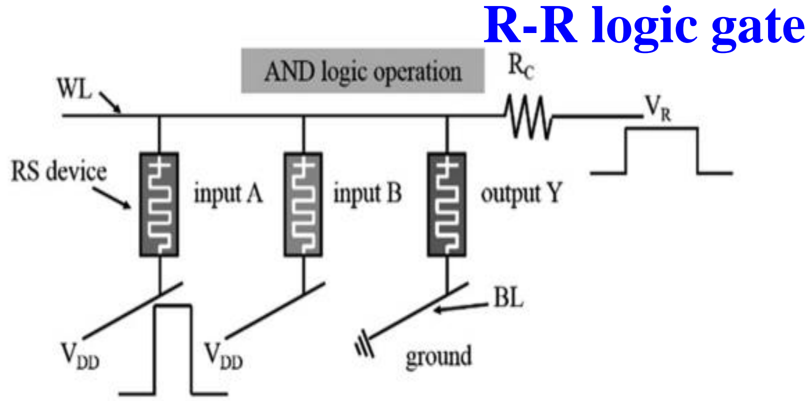
RRAM-based logic unit

V-R logic gate



Source: Nanotech, E. Linn

R-R logic gate

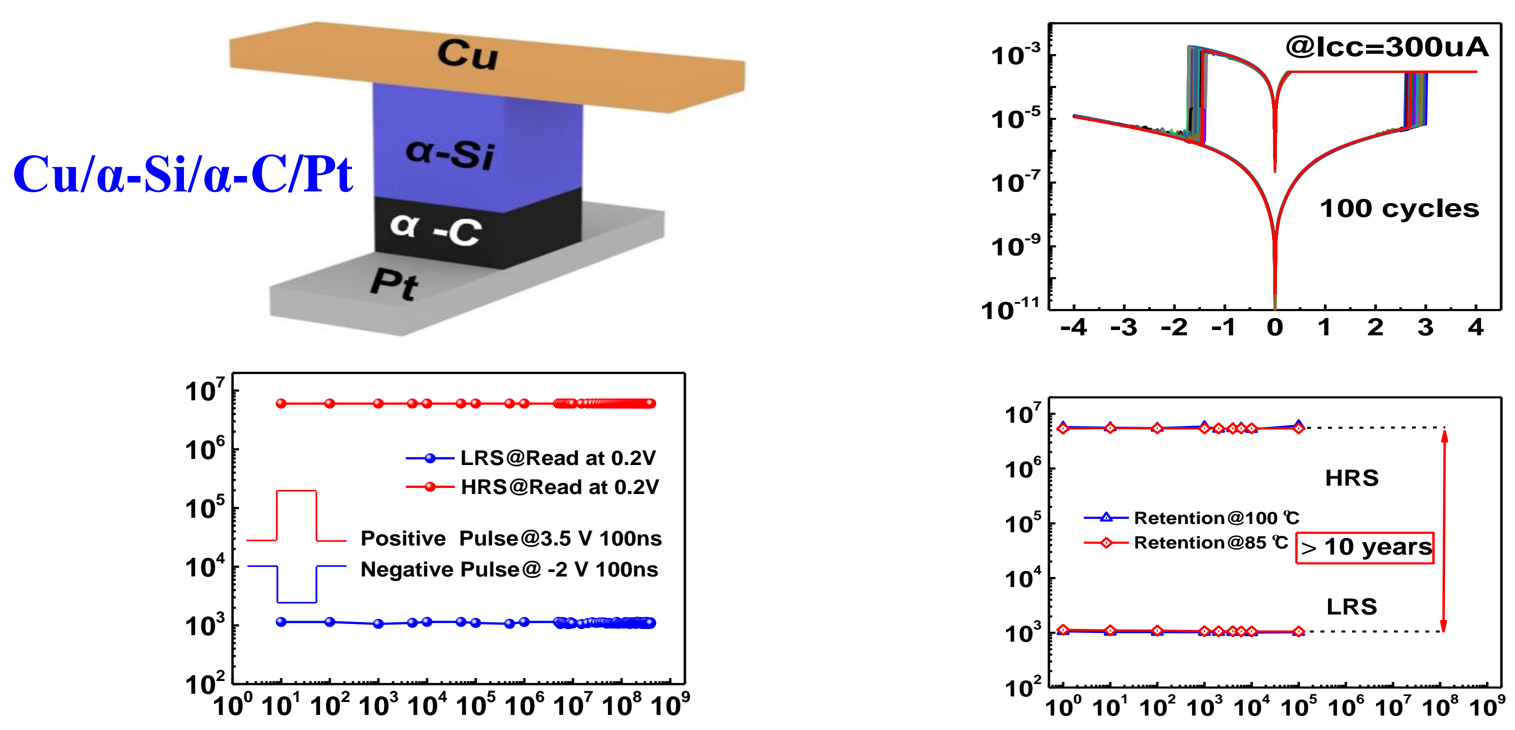


Source: Adv. Mater, P. Huang

- ✓ A considerable saving of static power
- ✓ Low requirement of device characteristics
- Input (voltage) and output (resistance) signals are physically different. Additional hardware burden, time and power dissipation will be cost.

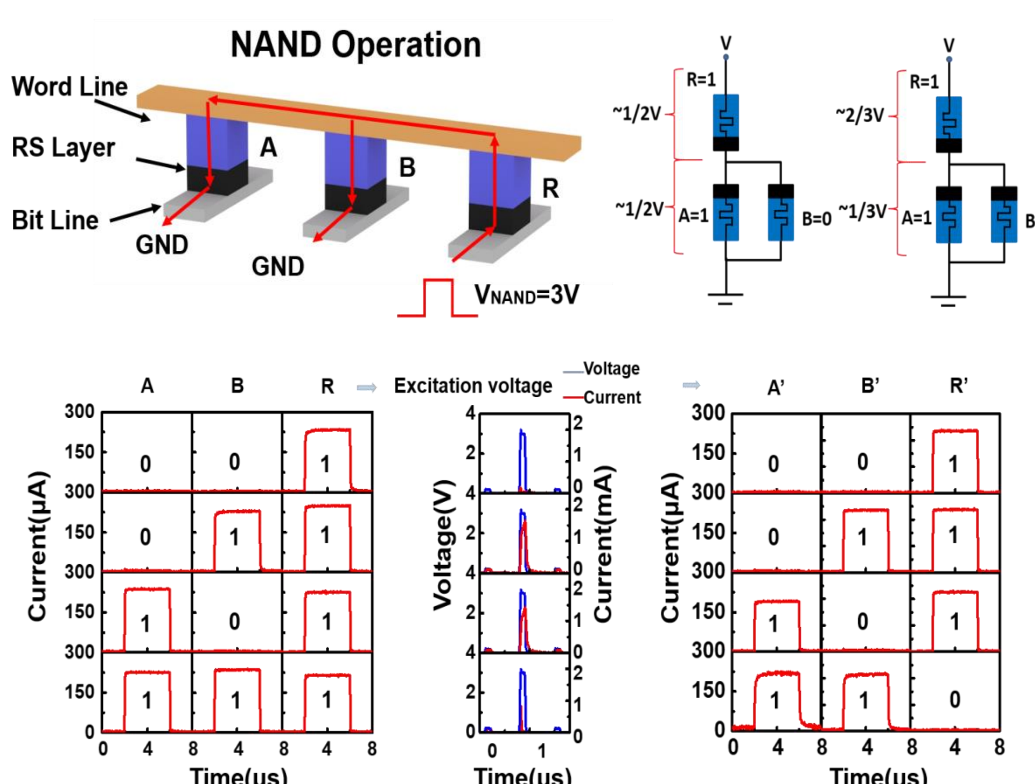
- ✓ Only physical variable-resistance.
- ✓ gate cascading can be achieved easily
- Devices with high uniform characteristics are necessary

High uniformity of RRAM



Cu/α-Si/α-C/Pt shows good endurance, retention and uniformity.

NAND logic gate



- Based on principle of resistance interaction, NAND operation was realized.
- Device A and B hold input signal and device R store operation result.
- NAND is basic operation of all the Boolean logic, other logics can achieve by proper cascading.

Implementation of 16 Boolean logic

Logic Name	TRUE	FALSE	A	B	NOT A	NOT B	OR	NAND
Operation Steps	Set operation	Reset operation	V, A	V, B	V, A, 1, R=1	V, 1, B, R=1	V, R=0, A, B	V, A, B, R=1
Function And Step number	1 1 step	0 1 step	A 1 step	B 1 step	$\overline{A}$ 1 step	$\overline{B}$ 1 step	A+B 1 step	$\overline{AB}$ 1 step

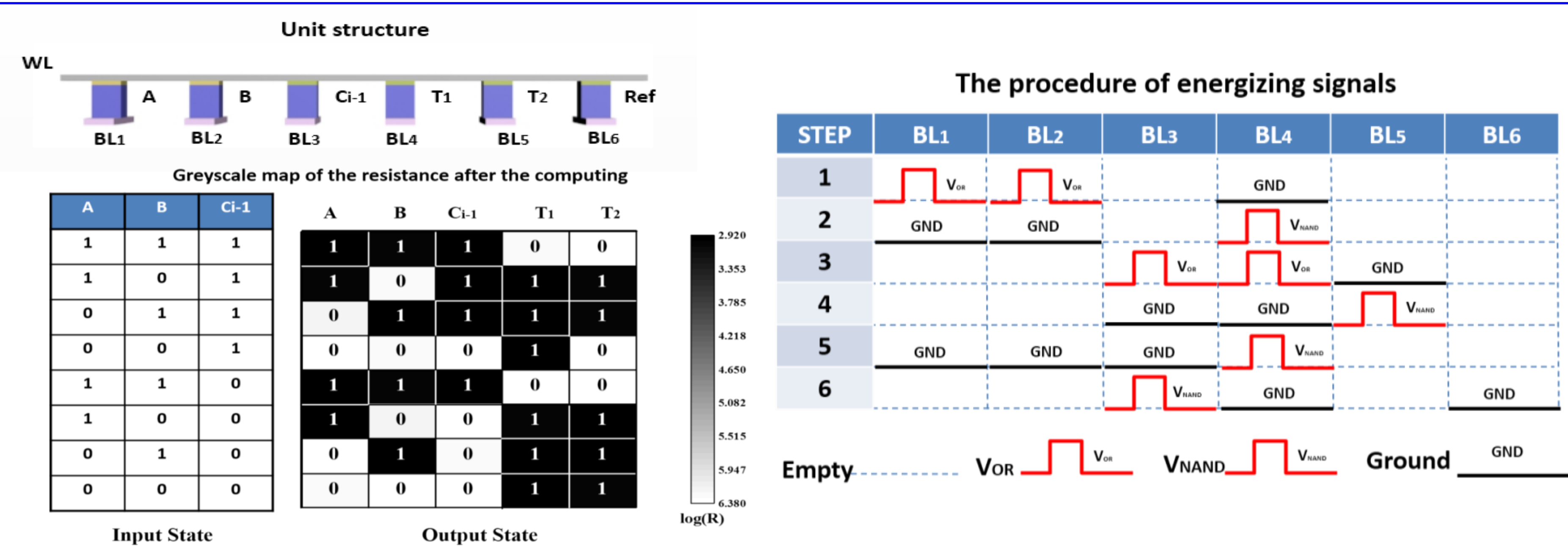
- Resistance states of RRAM for representation of logic "0" and "1";
- Via cascade of logic units, 16 Boolean logic can be implemented;
- 10 logic can be accomplished in 1 step.

Implementation of 16 Boolean logic

Logic Name	AND	IMP	NIMP	RNIMP	RIMP	NOR	NXOR	XOR
Operation Steps								
Function And Step number	$AB = \overline{\overline{AB}}$ 2 step	$\overline{A}+B$ 2 step	$A\overline{B}$ 1 step	$\overline{A}B$ 1 step	$\overline{B}+A$ 2 step	$\overline{A+B}$ 2 step	$\overline{A+B}+AB = \overline{(A+B)(\overline{AB})}$ 3 step	$A\overline{B}+\overline{A}B = (A+B)(\overline{AB})$ 2 step

Implementation of NXOR is the most complex one, it needs 5 devices in 3 steps.

Realization of 1 bit full adder



A	B	Ci-1
1	1	1
1	0	1
0	1	1
0	0	1
1	1	0
1	0	0
0	1	0
0	0	0

Input State

A	B	Ci-1	T1	T2
1	1	1	0	0
1	0	1	1	1
0	1	1	1	1
0	0	0	1	0
1	1	1	0	0
1	0	0	1	1
0	1	0	1	1
0	0	0	1	1

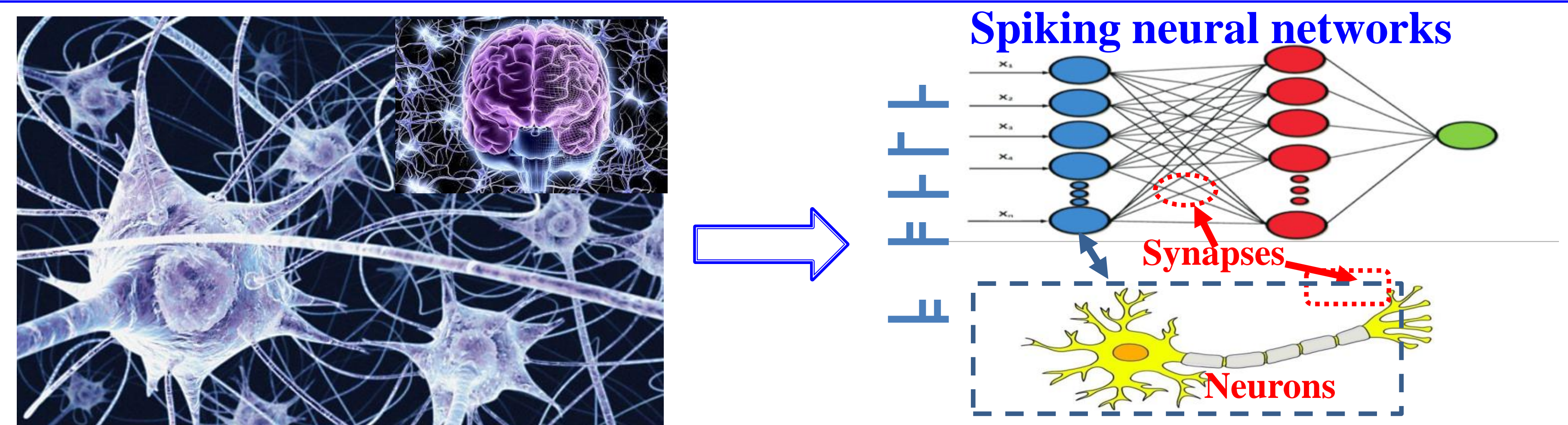
Output State

STEP	BL1	BL2	BL3	BL4	BL5	BL6
1	V_{OR}	V_{OR}		GND		
2	GND	GND		V_{NAND}		
3			V_{OR}	V_{OR}	GND	
4			GND	GND	V_{NAND}	
5	GND	GND	GND	V_{NAND}		
6			V_{NAND}	GND		GND

- 1 bit full adder needs **5** devices and 1 reference resistor, operation is finished in **6** steps.

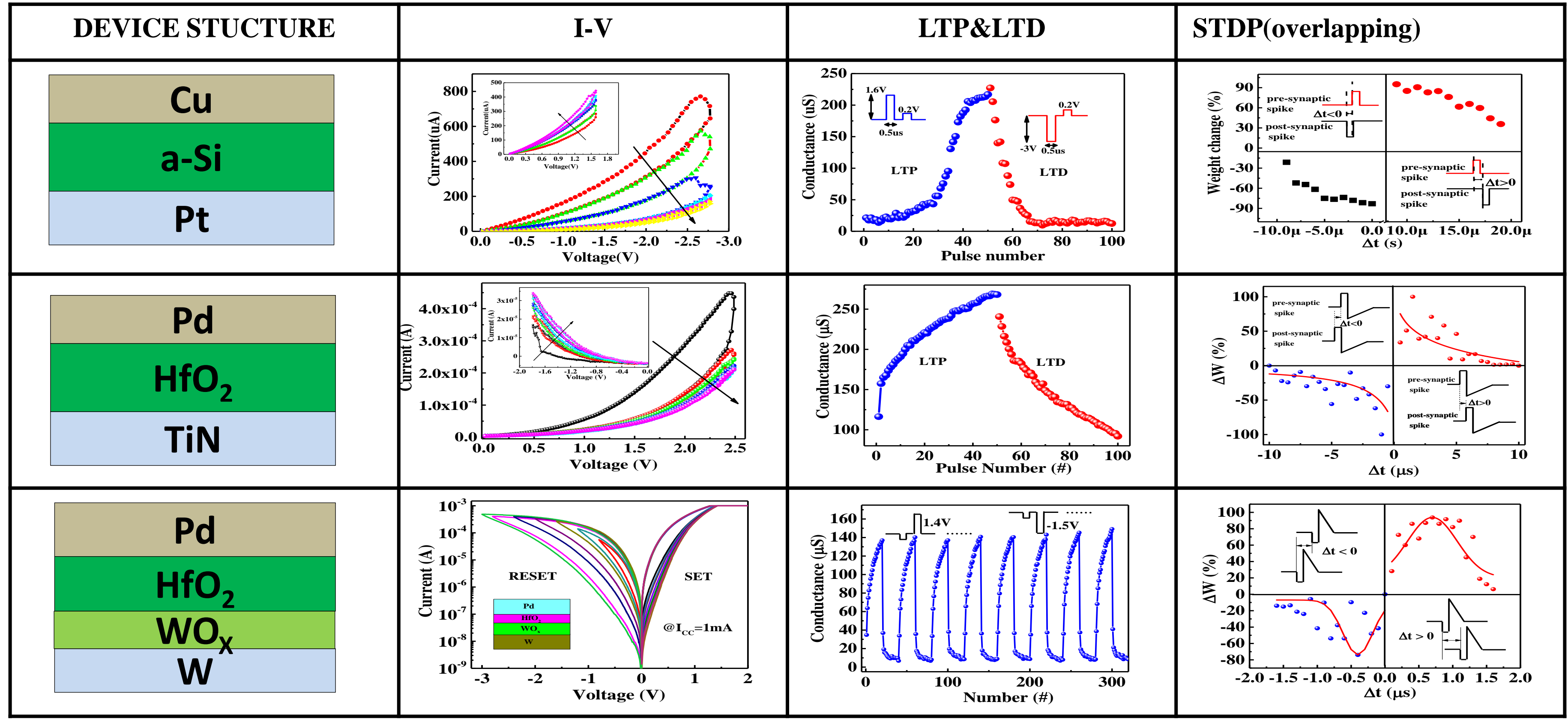
Unpublished

Brain-inspired computing with RRAM



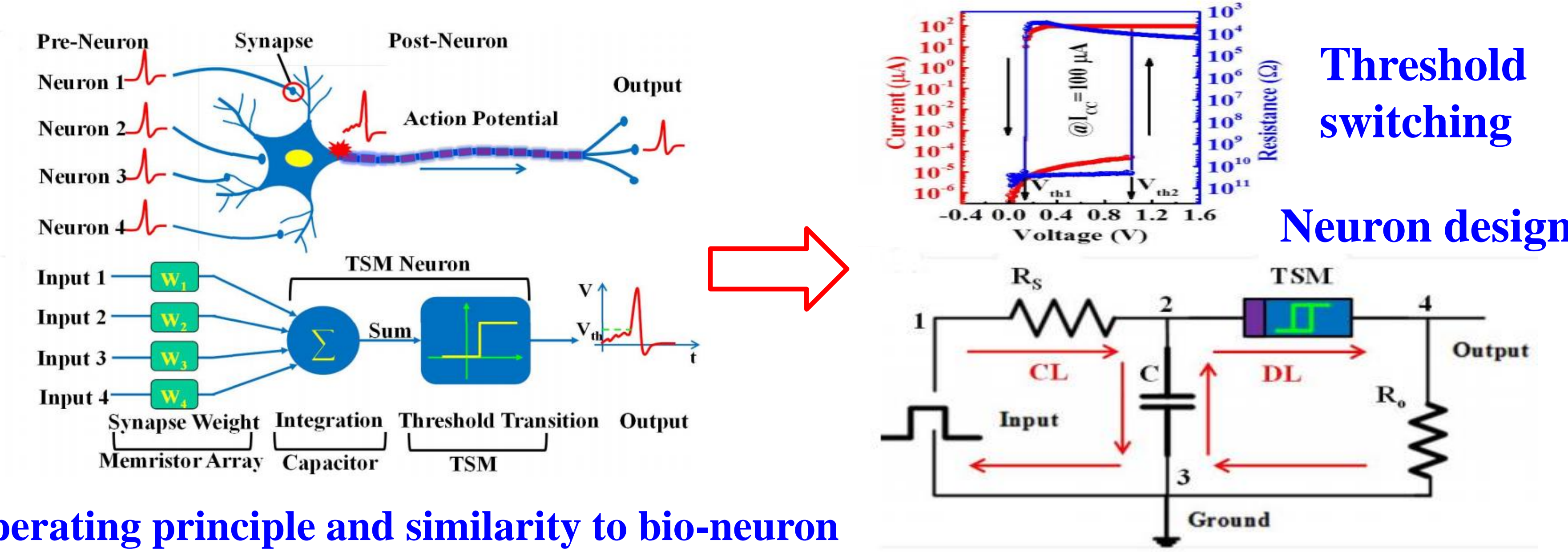
- The human neural system is inherently memory/computation convergence.
- Basic elements: neurons(receives, processes, stores and transmits information via its synapses), and synapses (connections between neurons).
- In spiking neural networks (SNN), the neurons integrate inputs from neurons in the previous layer and fires when a threshold value is reached, while synapses are connections between neurons.

Synaptic functions



IEEE EDL, 2017, 38:1208; Adv. Funct. Mater., 2018, 28:1705320; Nanoscale, 2017, 9:14442.

Artificial neuron circuit

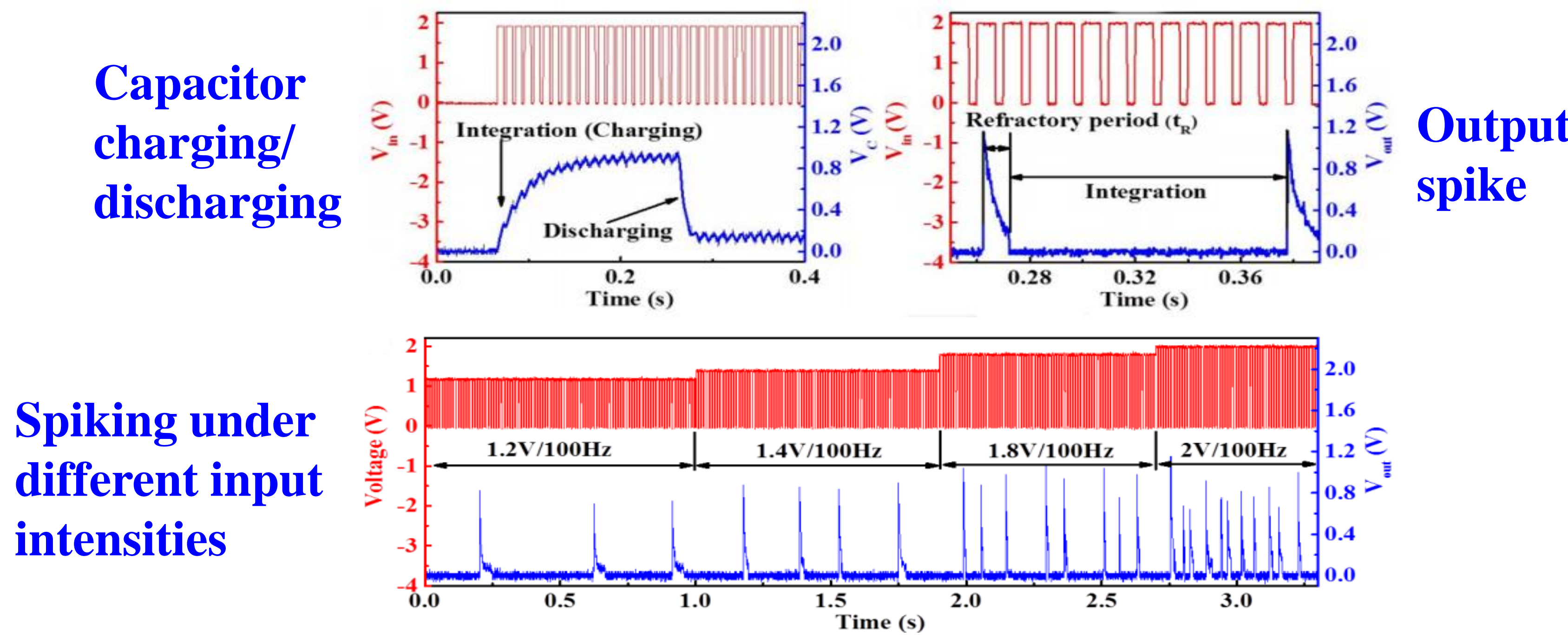


Operating principle and similarity to bio-neuron

➢ According to the integration and fire model, a simply neuron circuit is constructed with 1 TSM, 1 capacitor and 1 resistor.

IEEE EDL, 2018, 39:308

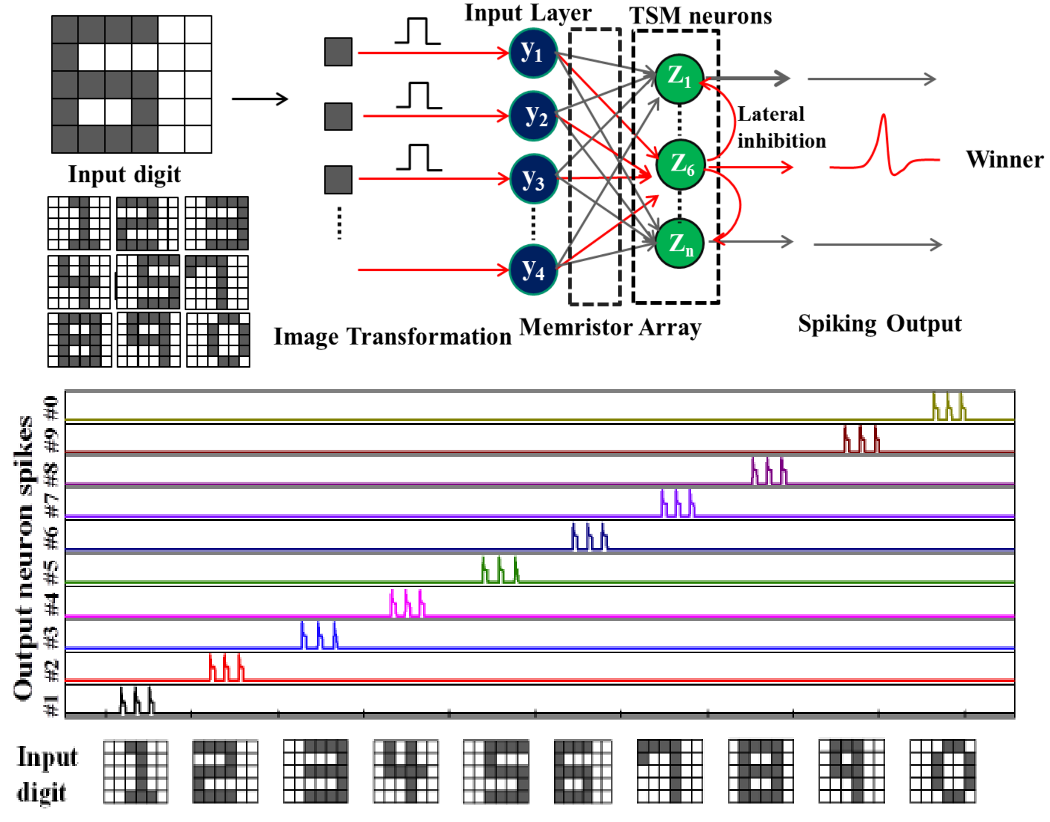
Functions of artificial neuron



The TSM neuron successfully achieved four key behaviors of bio-neurons: the all-or-nothing spiking, threshold-driven spiking, a refractory period, and a strength-modulated frequency response.

IEEE EDL, 2018, 39:308

Digit recognition simulation



- **Digit recognition** simulation with TSM neurons demo. on an array with 1 input layer (30 synapses) and 1 output layer (10 neurons);
- Lateral inhibition is implemented with the winner-take-all rule.

IEEE EDL, 2018, 39:308

Outline

- Computing System's Challenge
- RRAM: Memory/Storage Convergence
- RRAM: Memory/Computing Convergence
- Summary

Summary

- Computing technology improved 10^{10} times in past 60 years, face big challenge:
  - ✓ Moore law slow-down (trade off between performance and power density),
  - ✓ Limitation of traditional memories (fast, high density, cheap, non-volatile)
  - ✓ Von Neumann architecture(performance gap between memory and CPU)
- M/S convergence to reduce memory hierarchy and M/C convergence to realize brain-like high efficiency computing.
- RRAM as a new Memory technology, has already entered niche market for embedded application.
- Highly promising, significant efforts are still needed to address the interdisciplinary challenges of device optimization, circuit design, and system management.

Acknowledgement

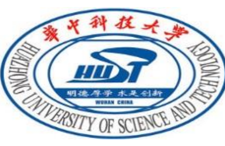
Group Members:

Prof. shibing Long, Qi Liu, Hangbing Lv Dr. Jinshun Bi, Feng Zhang, Jing Liu, Yuanlu Xie, Xiaoxin Xu, Qing Luo, Tuo Shi, etc.

Graduated Students:

Weihua Guan, Yingtao Li, Sen Zhang, Ming Wang, Qingyun Zuo, Meiyun Zhang, Sen Liu, Xiaolong Zhao, Tiancheng Gong, Xumeng Zhang, Rongrong Cao, Rui Wang, etc.

Collaborator:



Funding:

National Natural Science Foundation of China (NSFC)

National Basic Research Program of China (973 Program)

Hi-Tech Research And Development Program Of China (863 Program)

National Key Project (NKP)

Thanks for your attention!