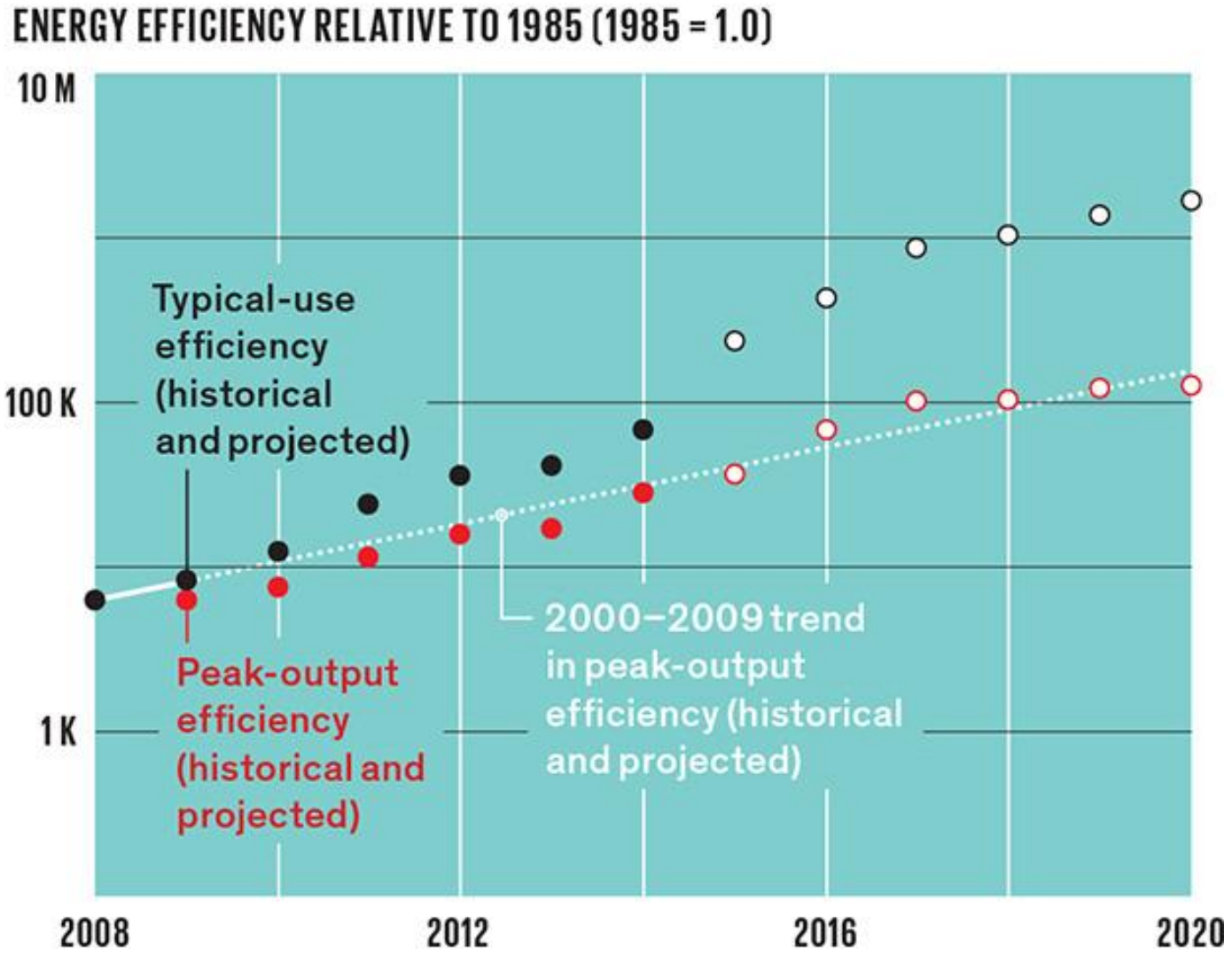
Emerging Non-volatile Memory: *Energy-efficient* Design with Carbon Nano-materials

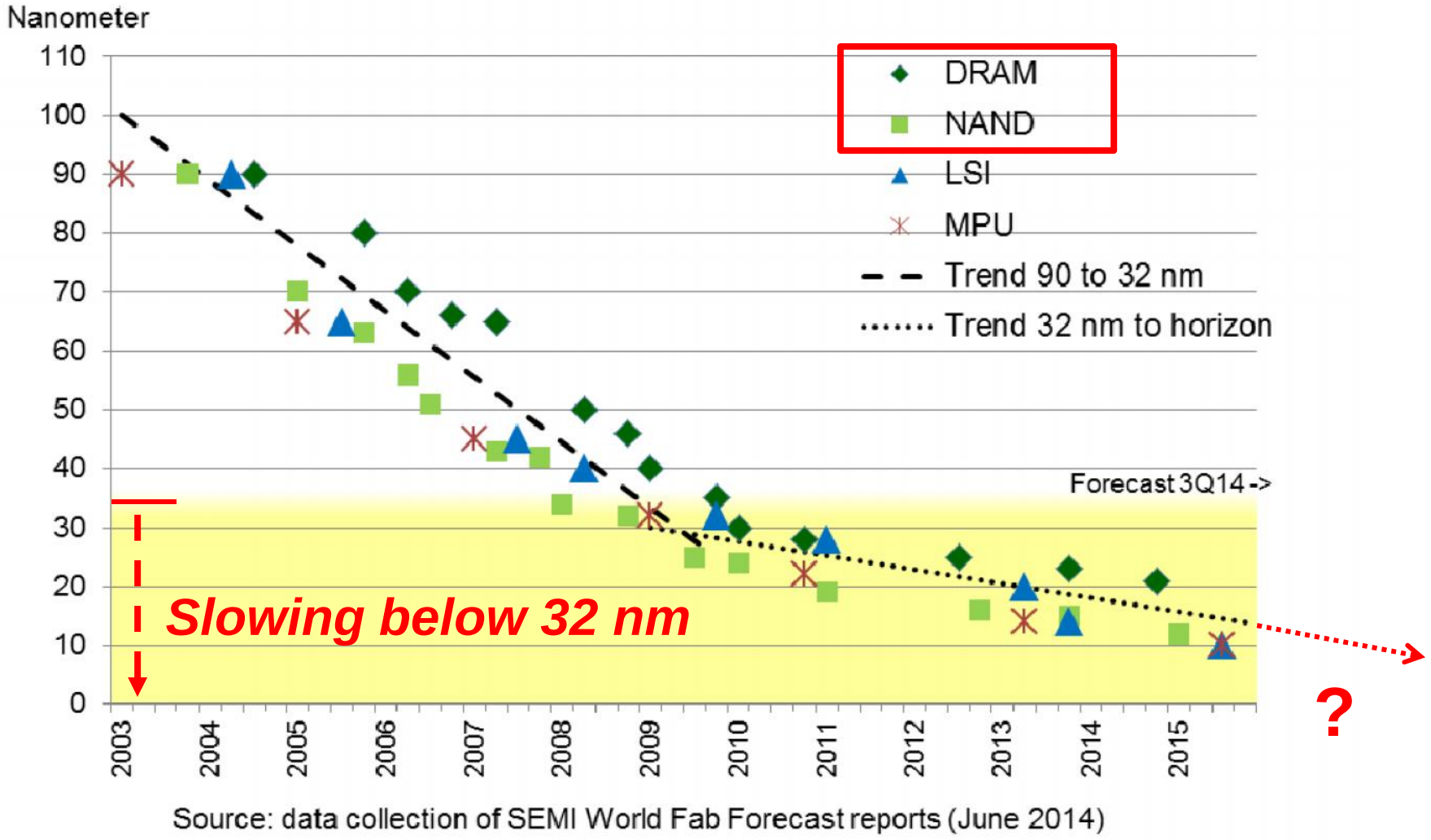
Ethan C. Ahn, Ph.D.
Electrical Engineering, Stanford University, CA, U.S.A.

50 years of Moore's law, "*how about energy?*"



Data sources: AMD, Koomey et al. (2011)

"Technology node transitions (volume production)"



“What’s the difference?”

vs.

$ 299

$ 399

“Adding more NAND”

vs.

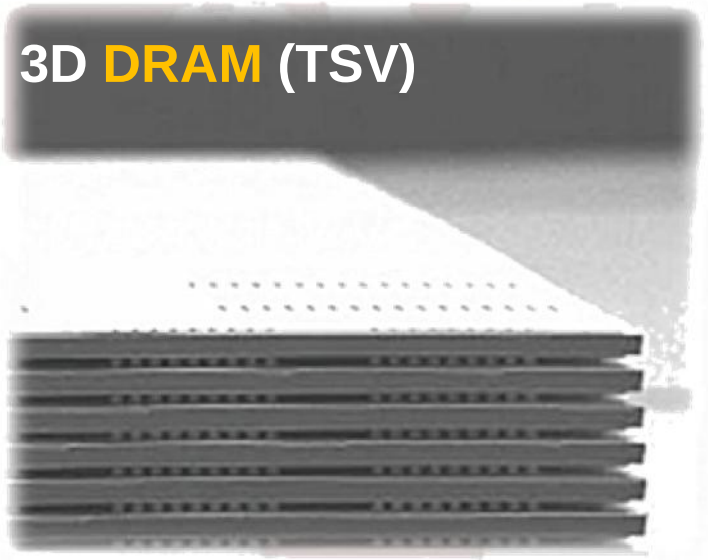
16 GB

(DRAM + NAND)

64 GB

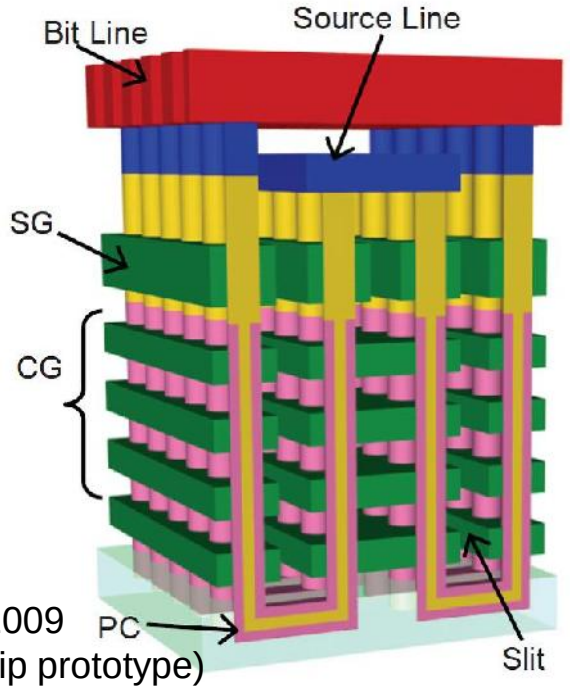
(DRAM + ***MORE NAND***)

"New tricks to further increase density"



SK Hynix, 2013 (product)

3D NAND
- Tohiba (BiCS)
- Sandisk (BiCS)
- Samsung (TCAT)
- SK-Hynix
- Micron
- …



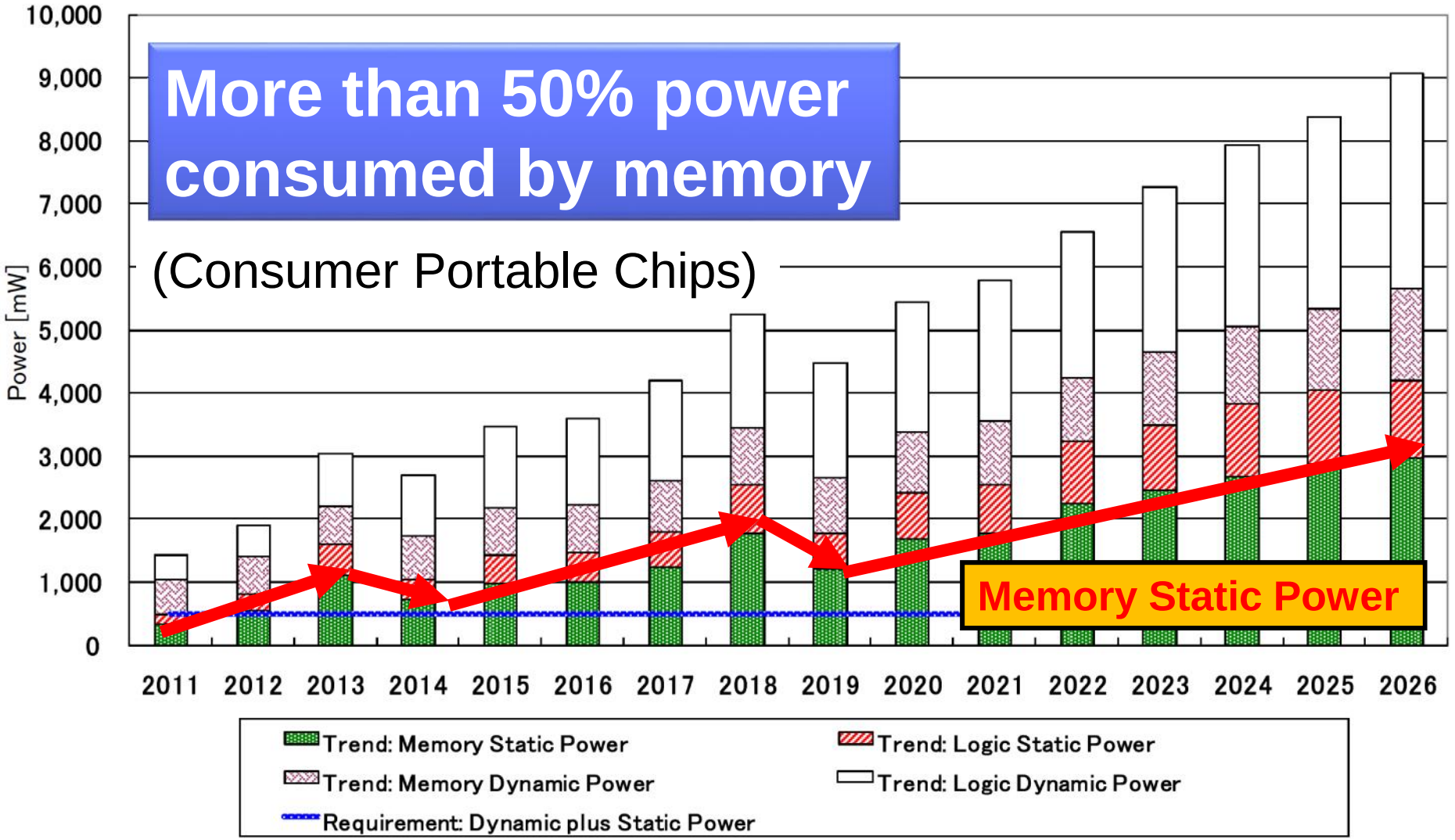
Toshiba, 2009
(32 Gb-chip prototype)

"Fundamental solution"

Emerging non-volatile memory (NVM)

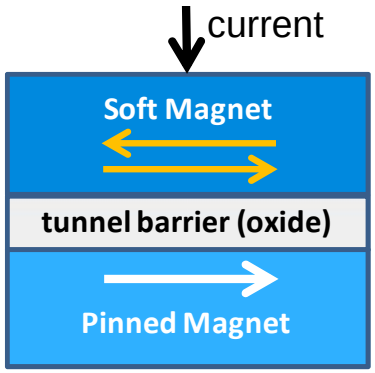
; sub-10 nm scalability & low cost (<0.1$/GB)

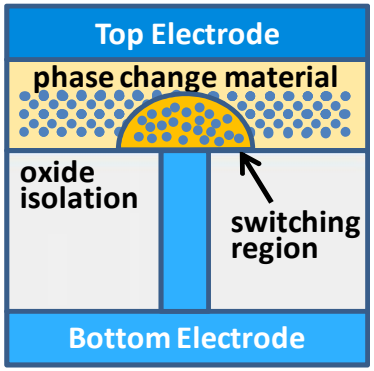
Memory: **Limiting factor for energy-efficient system**

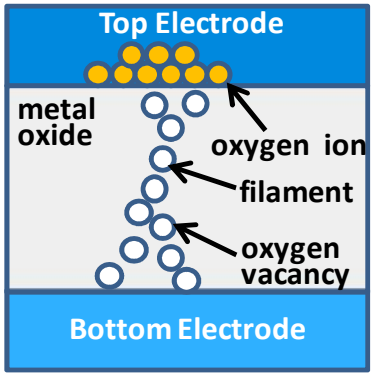


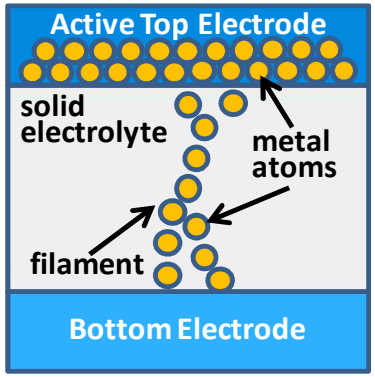
Ref: ITRS 2011 Edition (System Drivers)

“New” Players in NVM









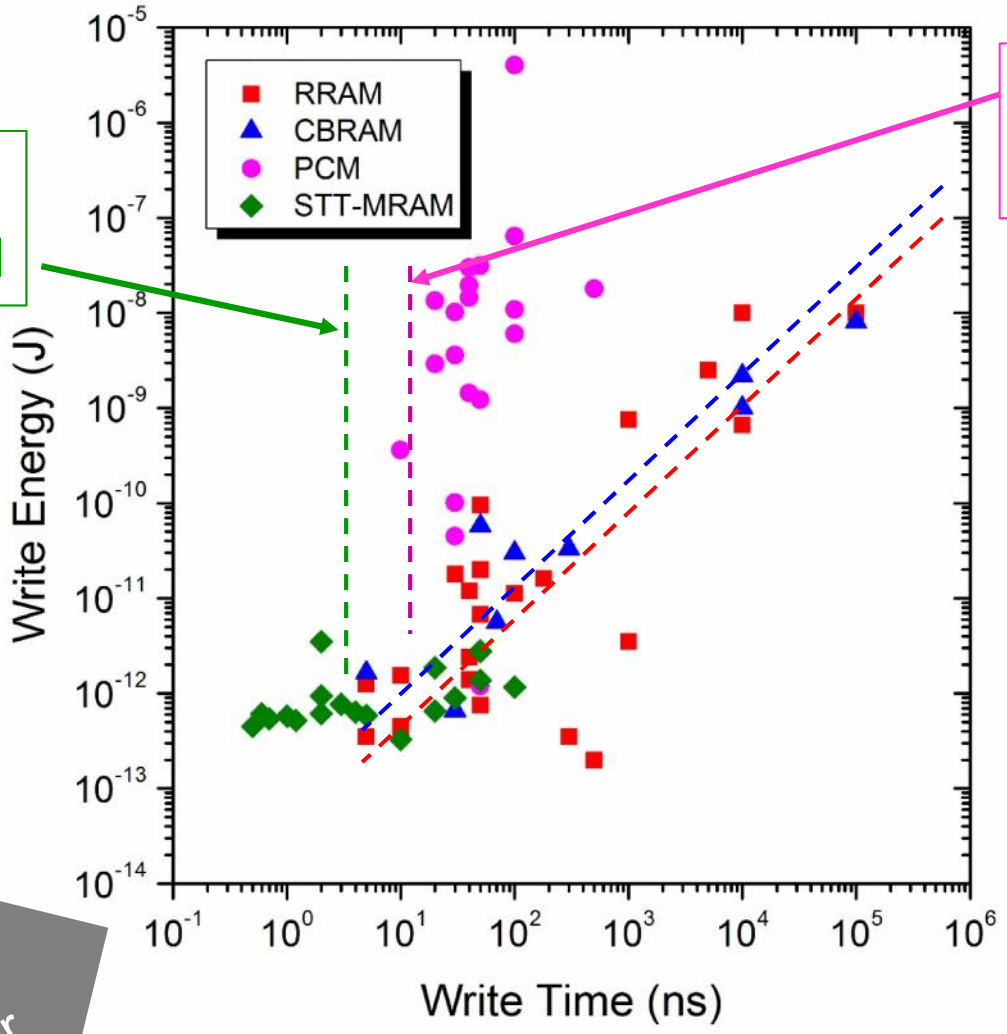
STT-MRAM	PCM	RRAM	CBRAM
Spin torque transfer magnetic random access memory	Phase change memory	Resistive switching random access memory	Conductive bridge random access memory

Random access, non-volatile, no erase before write

Energy vs Speed Trade-Off @ Device Level

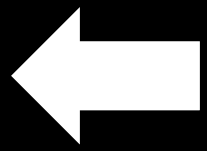
Nothing faster than STT-MRAM

Speed limited by physics



Good corner

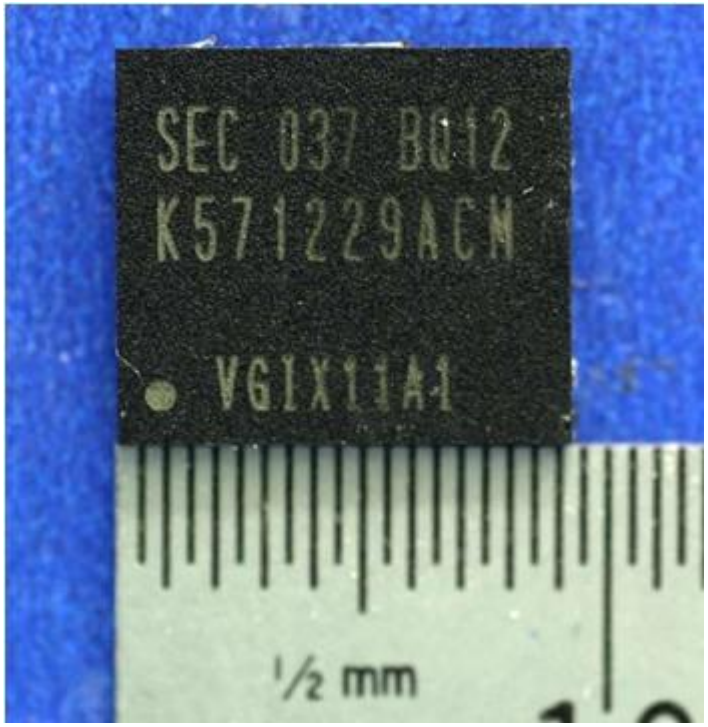
Wong and Ahn et al. “Stanford Memory Trends”, https://nano.stanford.edu

1. Energy-efficient *Cell* design 

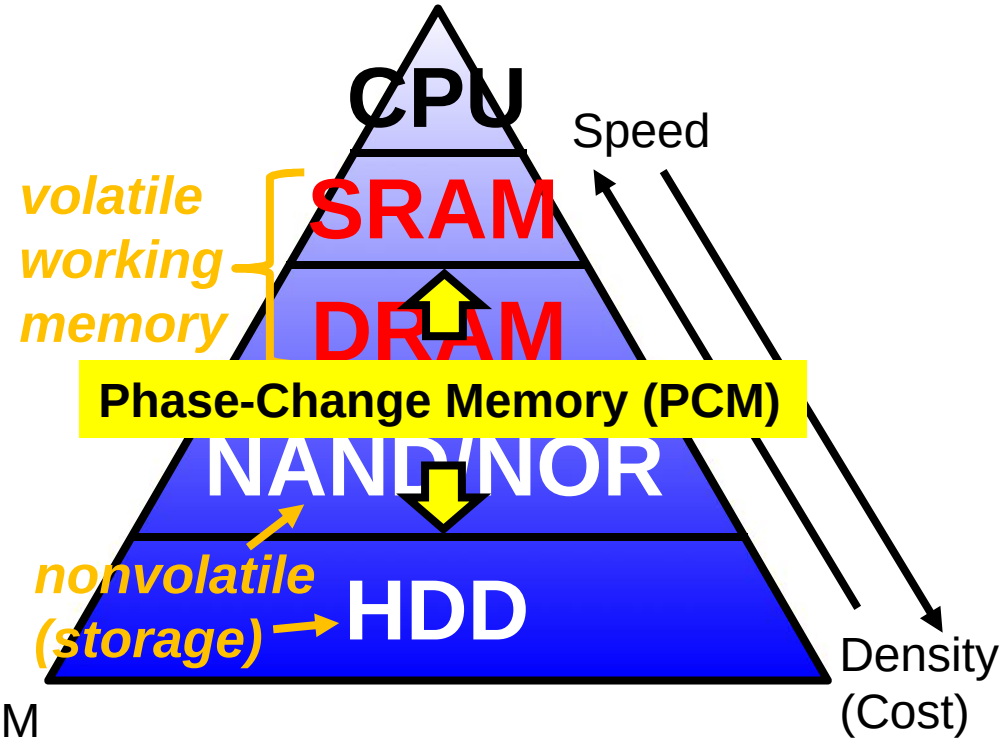
2. Energy-efficient *Architecture* design

PCM

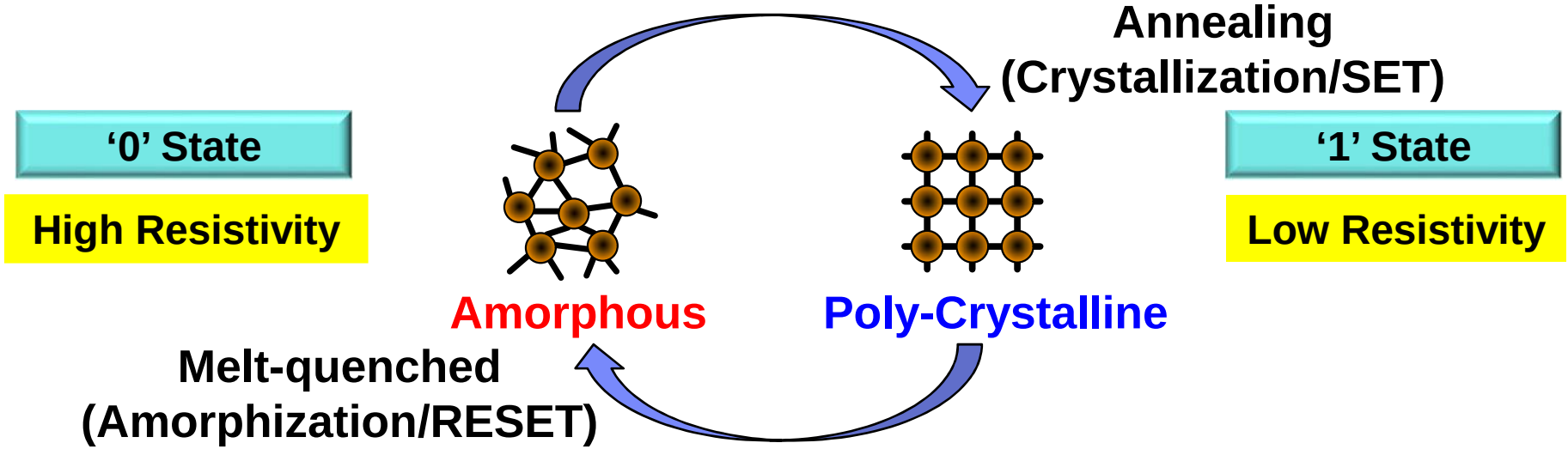
: Phase-change memory (PRAM/PCRAM)

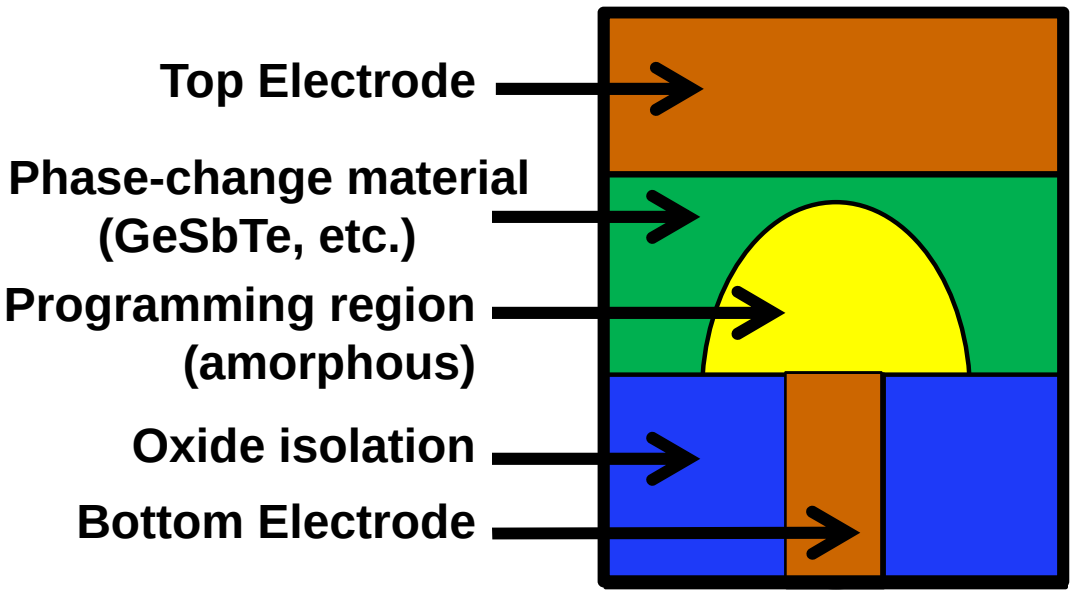


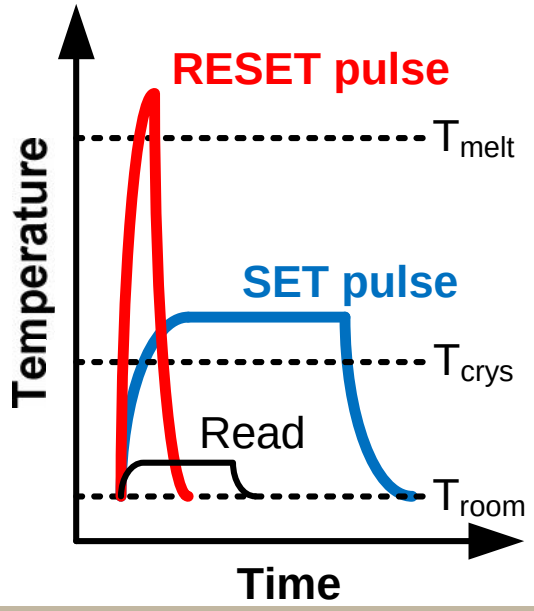
Packaged MCP that includes 512 Mb PCM
(NOR-compatible, **Samsung, 2010**)



PCM at a glance: **Based on Joule-heating**







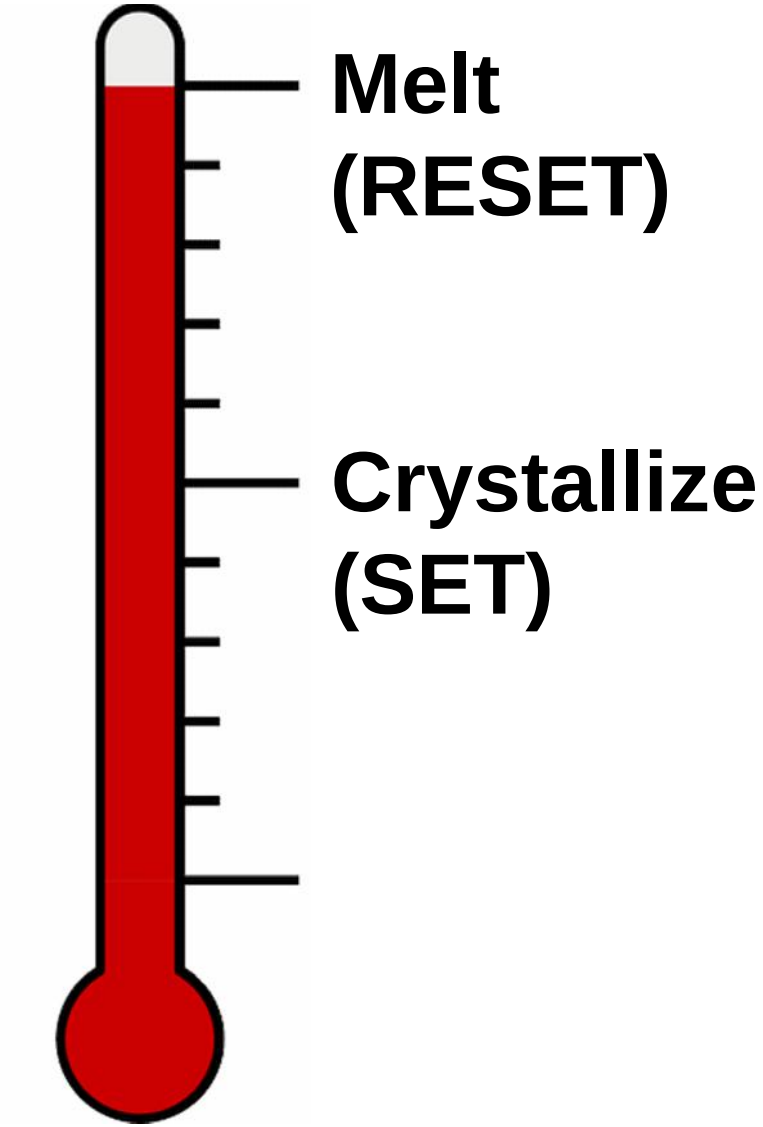
PCM: **Key challenges and issues**

- Physics of threshold switching, crystallization, etc.
- High write latency (due to long crystallization time)
- Cost-effective array architecture
- High programming (RESET) current
- Resistance drift (difficulty in MLC)
- Cross-talk (thermal disturbance)
- etc.

PCM: **Key challenges and issues**

- Physics of threshold switching, crystallization, etc.
- High write latency (due to long crystallization time)
- Cost-effective array architecture
- **High programming (RESET) current** ⇦
- Resistance drift (difficulty in MLC)
- Cross-talk (thermal disturbance)
- etc.

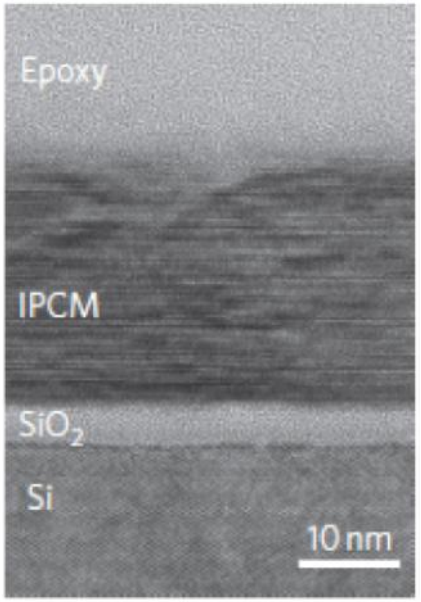
“How hot is $Ge_2Sb_2Te_5$?”



Toward lower I_{RESET}: **Materials** Engineering

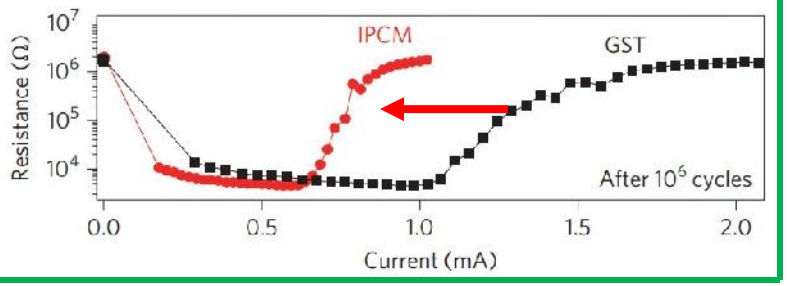
Example:

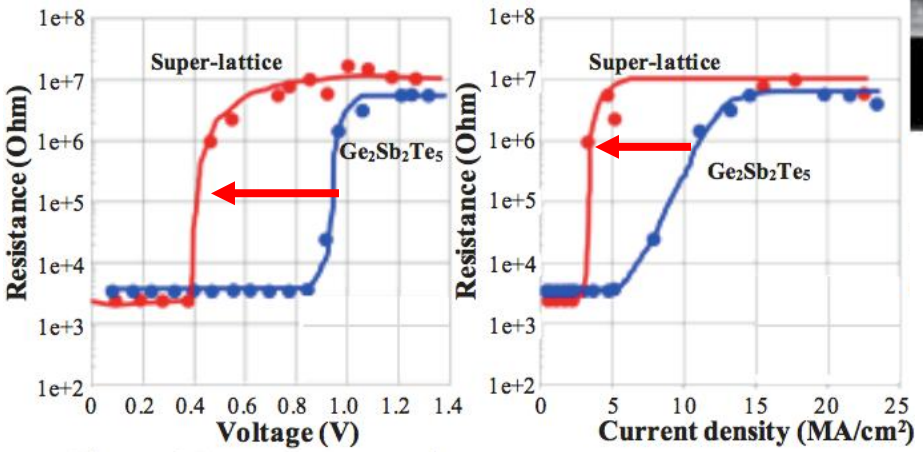
Recent studies with **$GeTe/Sb_2Te_3$ super-lattice** structure



IPCM (Interfacial PCM)

(R.E. Simpson, Nature Nanotech. **6**, 2011)



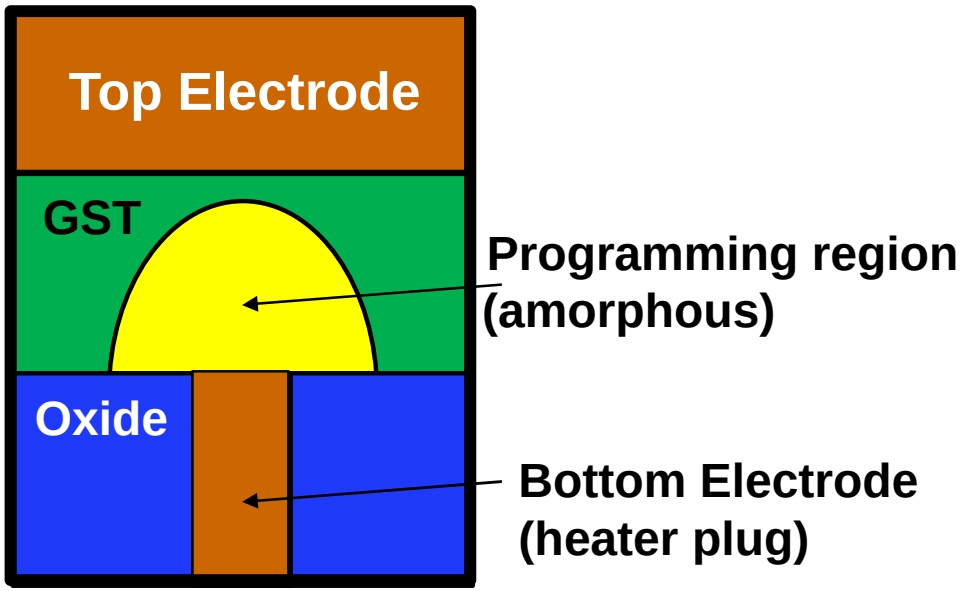


Charge-injection Super-lattice PCM

(N. Takaura, VLSI 2013)

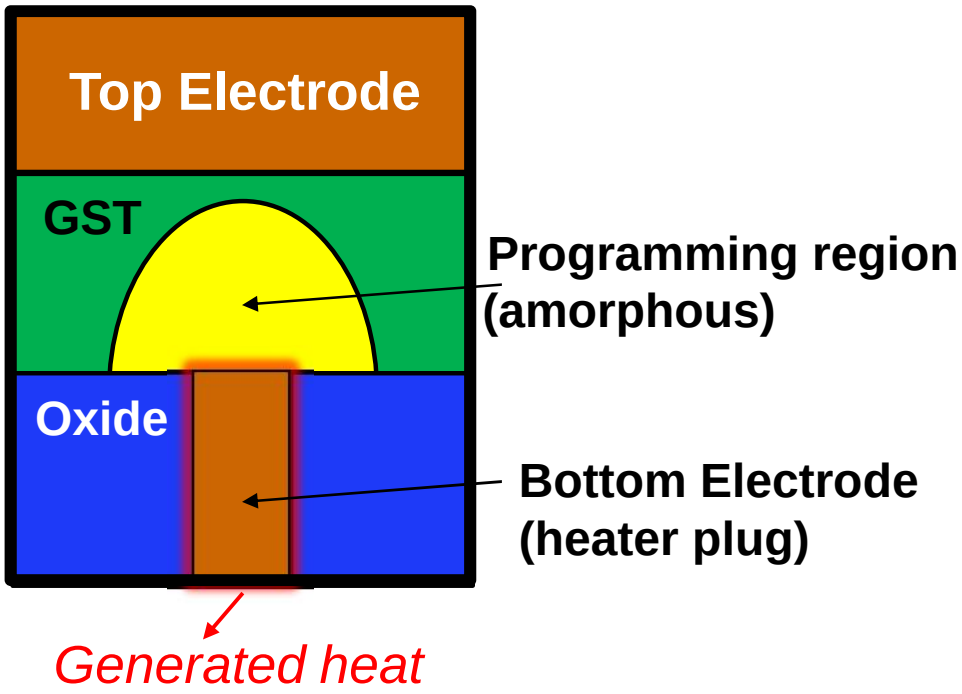
Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on **"<u>Joule Heating,</u>"**



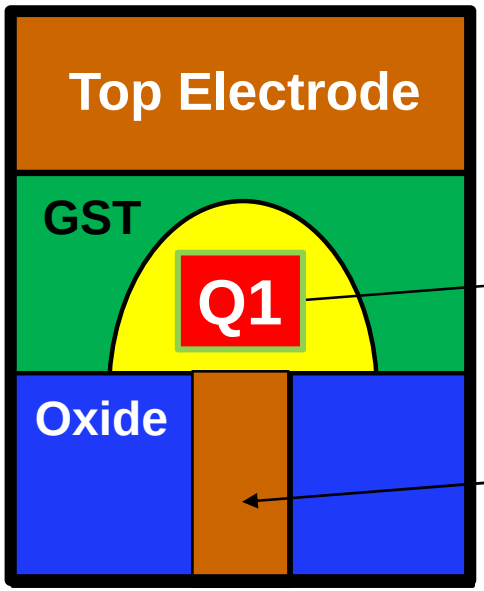
Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Programming region (amorphous)

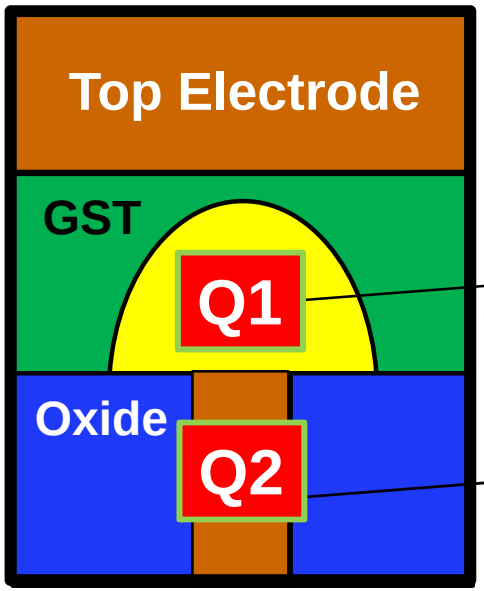
Bottom Electrode (heater plug)

Q1 Used for switching

Heat dissipated during RESET
in a typical mushroom PCM cell

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Programming region (amorphous)

Bottom Electrode (heater plug)

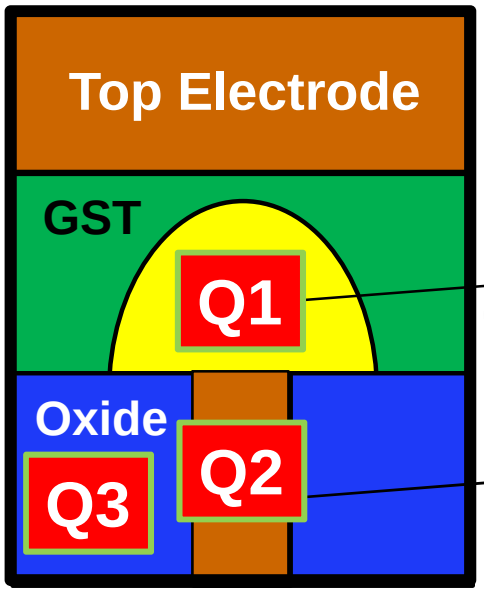
Q1 Used for switching

Q2 Stored in the heater

Heat dissipated during RESET in a typical mushroom PCM cell

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Programming region (amorphous)

Bottom Electrode (heater plug)

Q1 Used for switching

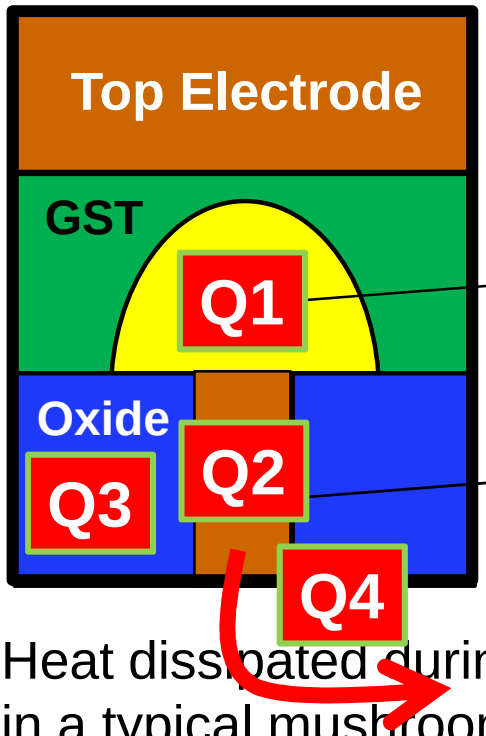
Q2 Stored in the heater

Q3 Diffused into oxide

Heat dissipated during RESET
in a typical mushroom PCM cell

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Programming region (amorphous)

Bottom Electrode (heater plug)

Heat dissipated during RESET in a typical mushroom PCM cell

Q1 Used for switching

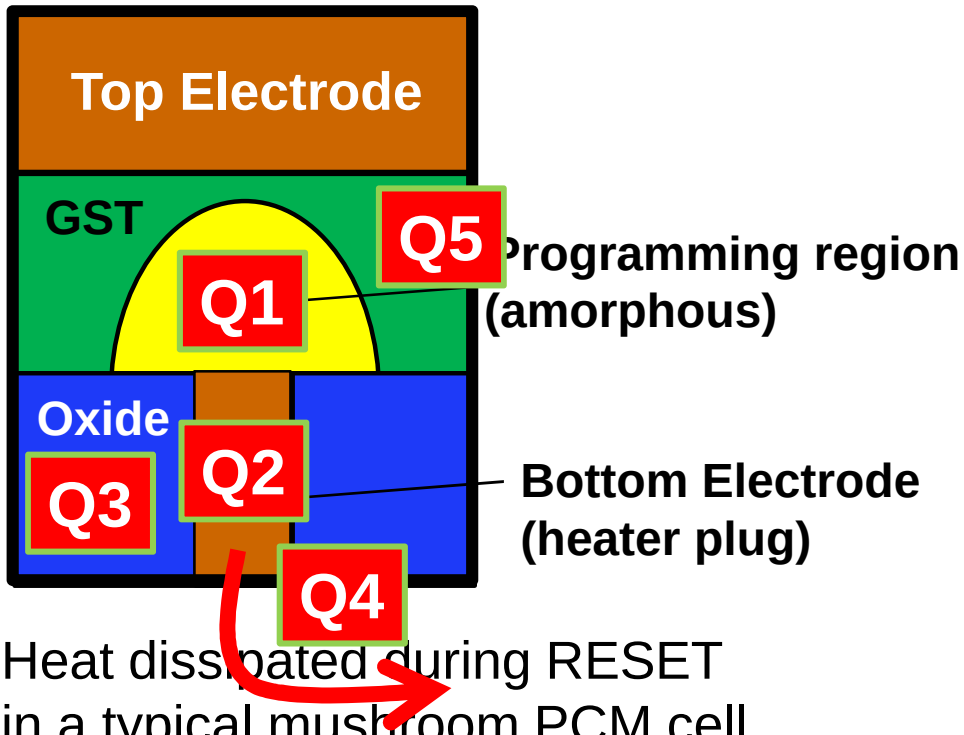
Q2 Stored in the heater

Q3 Diffused into oxide

Q4 Flows into the metal (at the bottom)

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Heat dissipated during RESET
in a typical mushroom PCM cell

Q1 Used for switching

Q2 Stored in the heater

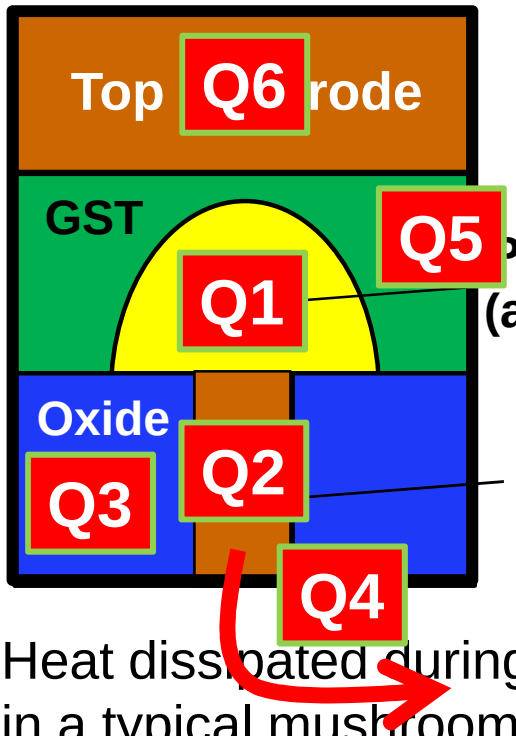
Q3 Diffused into oxide

Q4 Flows into the metal (at the bottom)

Q5 Diffused into surrounding GST (crystalline)

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"



Heat dissipated during RESET
in a typical mushroom PCM cell

Q1 Used for switching

Q2 Stored in the heater

Q3 Diffused into oxide

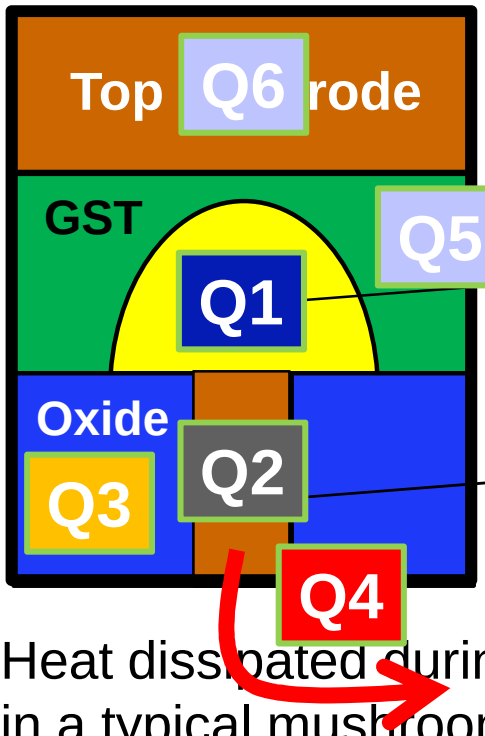
Q4 Flows into the metal (at the bottom)

Q5 Diffused into surrounding GST (crystalline)

Q6 Flows into the metal (at the top)

Toward lower I_{RESET}: **Thermal** Engineering

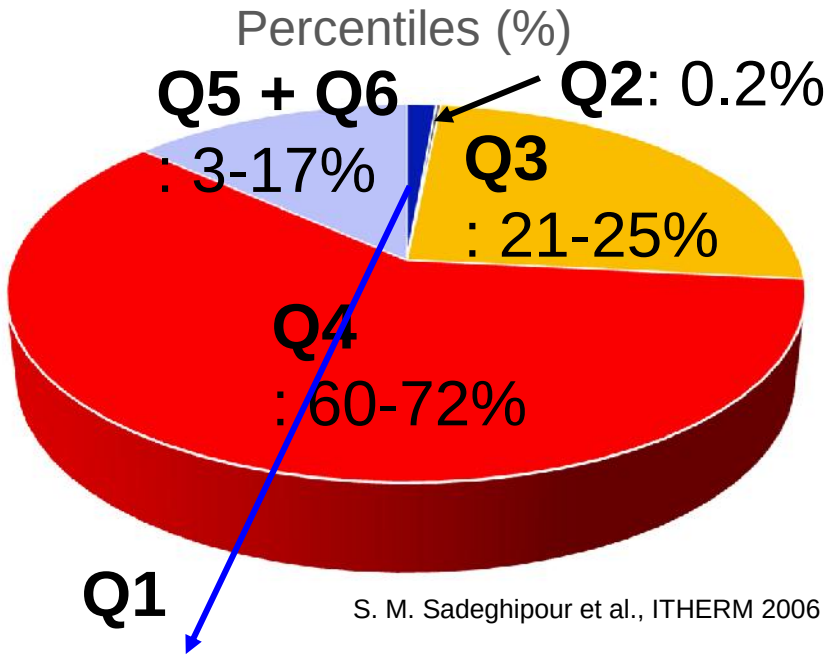
Remembering that PCM operation is based on "**Joule Heating,**"



Programming region (amorphous)

Bottom Electrode (heater plug)

Heat dissipated during RESET in a typical mushroom PCM cell

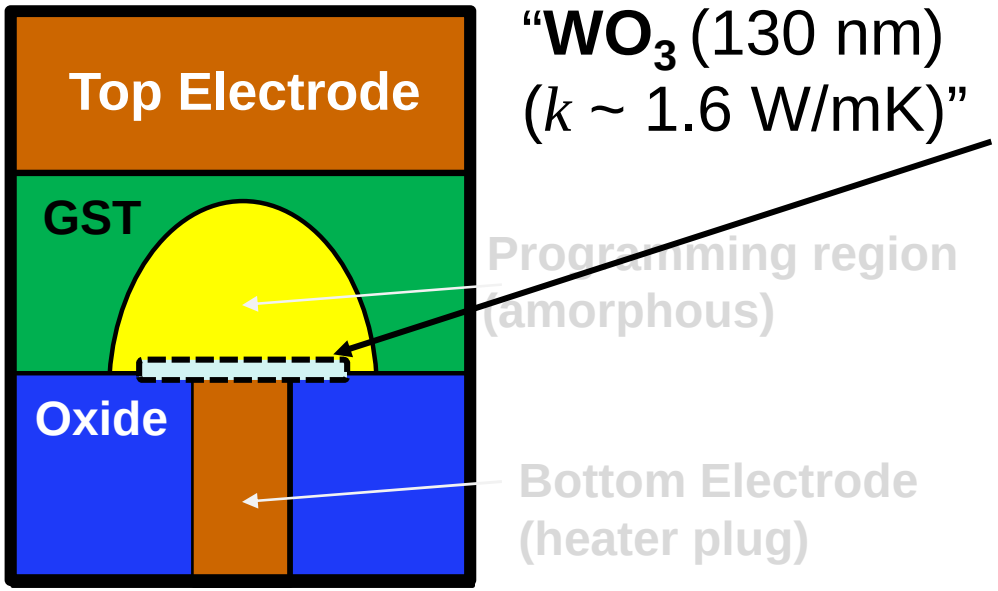


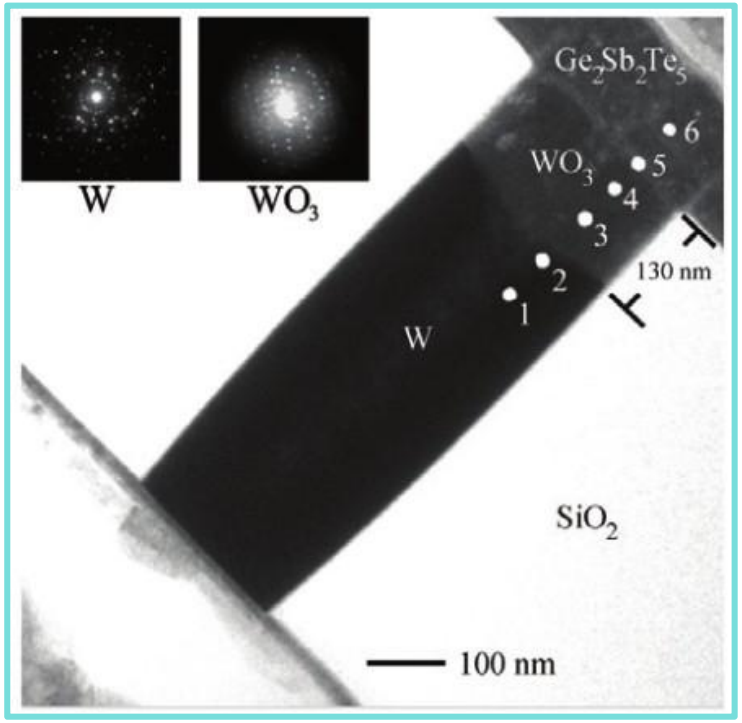
S. M. Sadeghipour et al., ITHERM 2006

"Only a very small fraction (< 1 %) of the generated heat is actually used in the active region"

Toward lower I_{RESET}: **Thermal** Engineering

Remembering that PCM operation is based on "**Joule Heating,**"





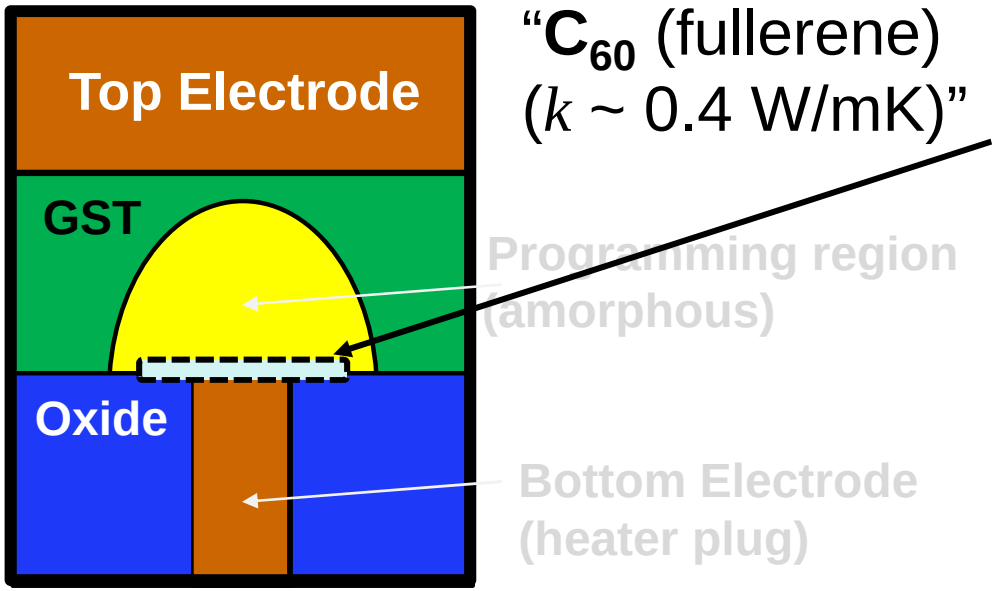
F. Rao, Nanotech. **19**, 2008

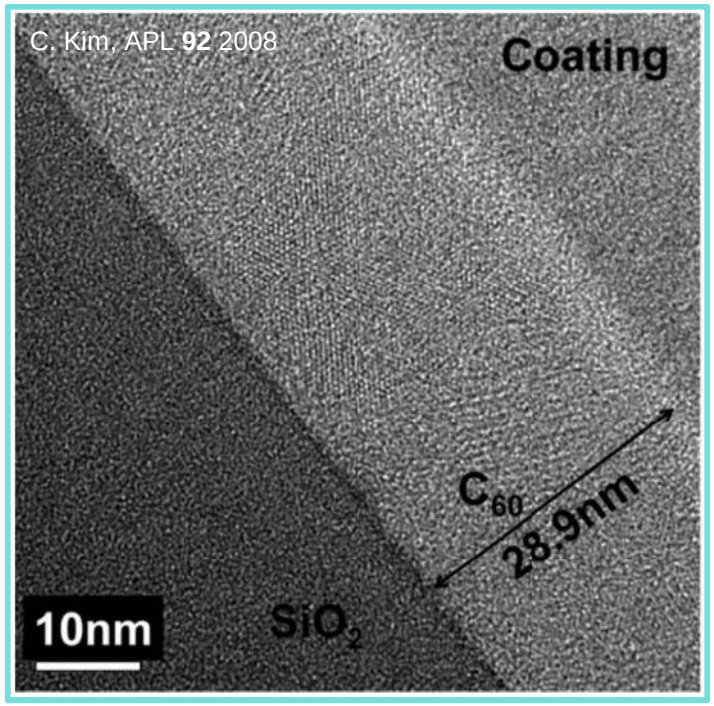
Heat dissipated during RESET
in a typical mushroom PCM cell

*"Only a very small fraction (< 1 %)
of the generated heat is actually
used in the active region"*

Toward lower I_{RESET}: **Thermal** Engineering

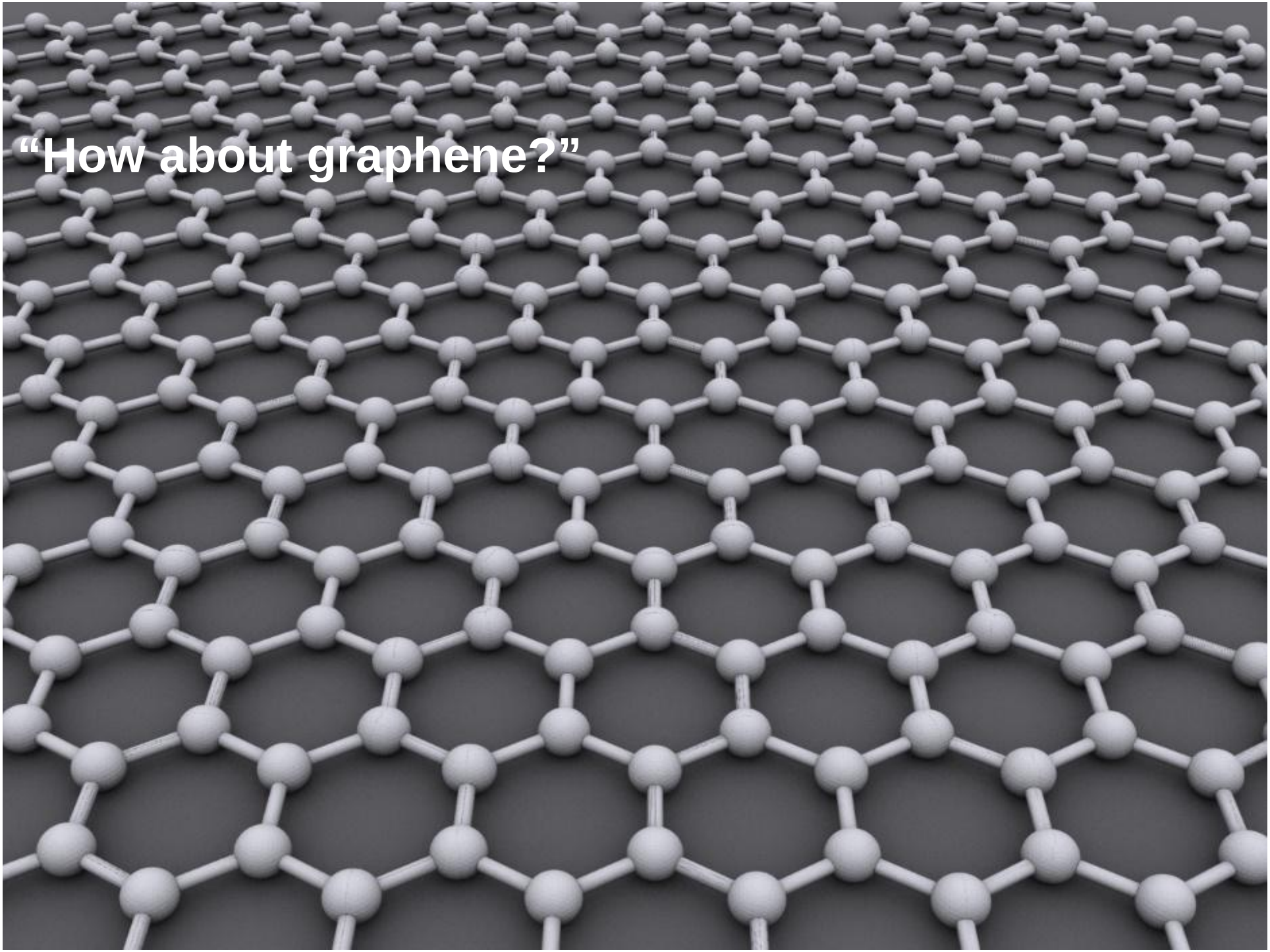
Remembering that PCM operation is based on "**Joule Heating,**"





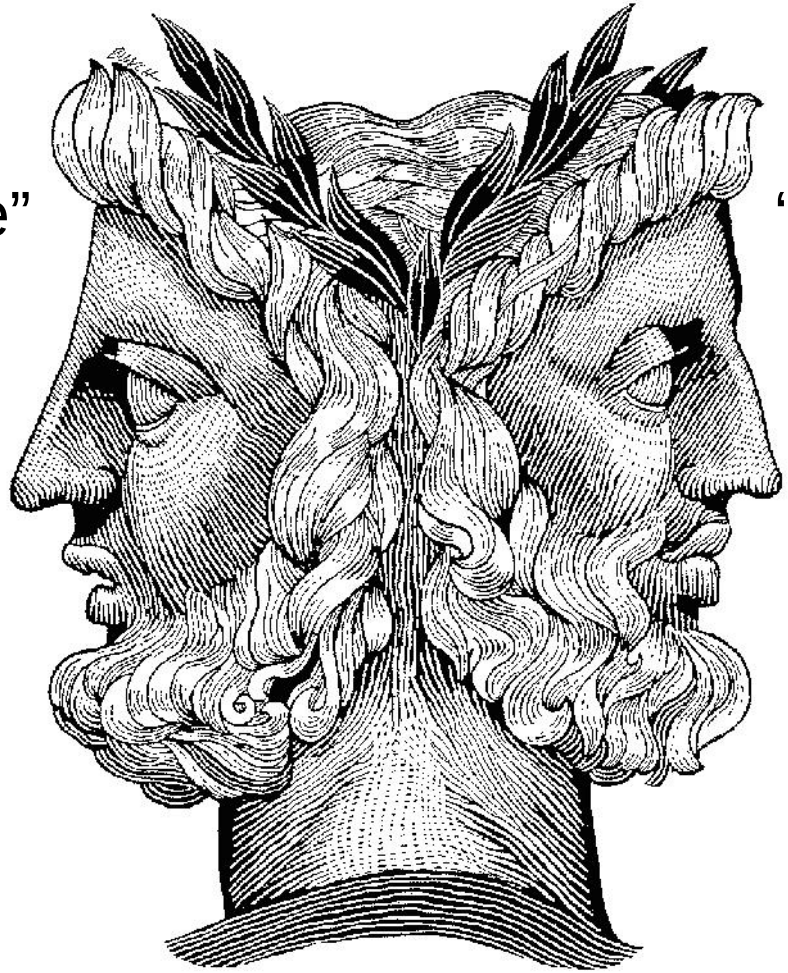
Heat dissipated during RESET
in a typical mushroom PCM cell

"Only a very small fraction (< 1 %) of the generated heat is actually used in the active region"


“How about graphene?”

The Janus faces of graphene

“In-plane”



“Out-of-plane”

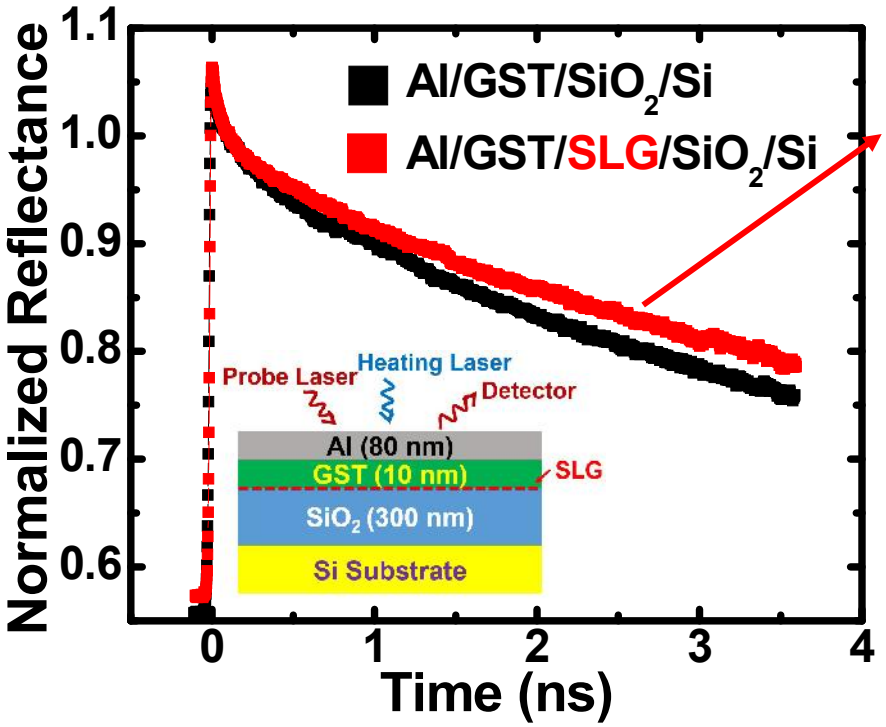
See references

[1] Pop et al. MRS Bull. 2012
[2] Guzman et al. ITherm. 2014
[3] Koh et al. Nano Lett. 2010
[4] Mak et al. APL 2010

Graphene as a thermal barrier: **TDTR measurements**

Q1: Is a single-layer graphene good as a thermal barrier?
(Is a single-layer graphene **thermally-resistive** well enough?)

"**TDTR** (**T**ime **D**omain **T**hermo**R**eflectance)"



Slower thermal decay with SLG

↓

Larger **T**hermal **B**oundary **R**esistance

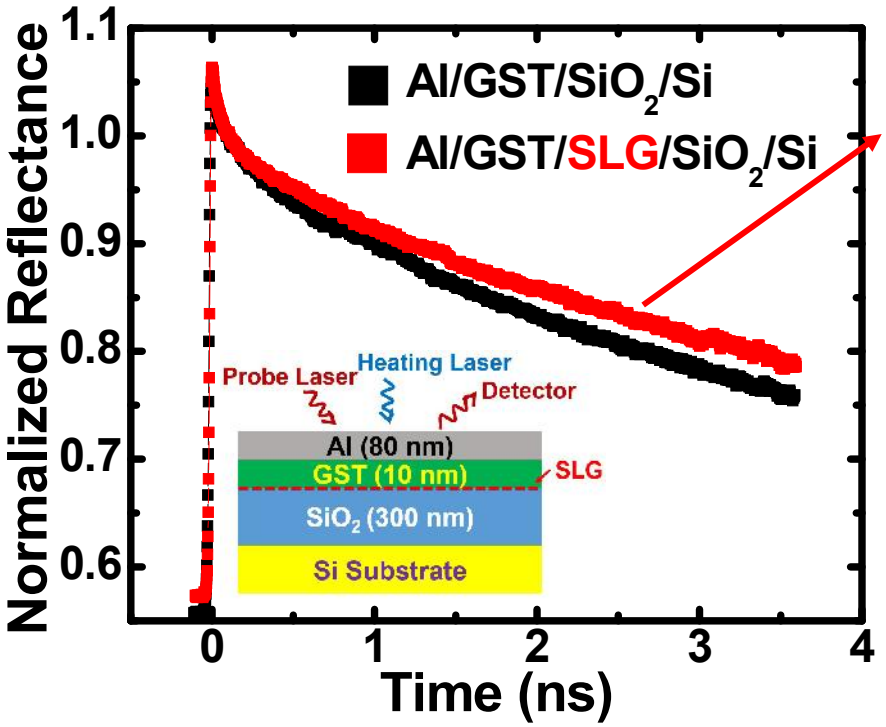
A1: YES

Ahn et al. Nano Letters 2015

Graphene as a thermal barrier: **TDTR measurements**

Q1: Is a single-layer graphene good as a thermal barrier?
(Is a single-layer graphene **thermally-resistive** well enough?)

"**TDTR** (**T**ime **D**omain **T**hermo**R**eflectance)"



Slower thermal decay with SLG

↓

Larger **T**hermal **B**oundary **R**esistance

A1: YES

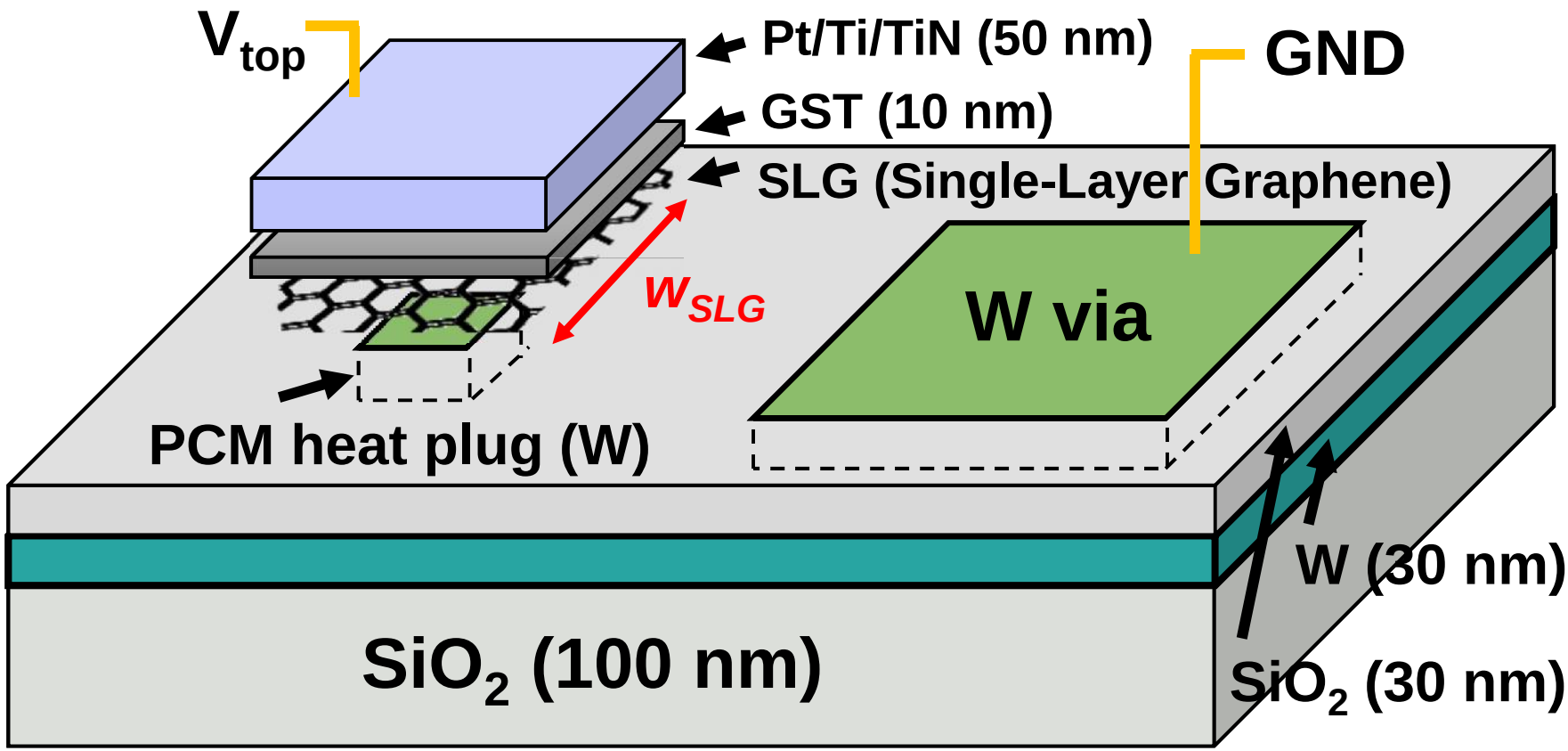
Q2: *HOW LARGE?*

A2:

ΔTBR (m^2K/GW)
= 32 +/- 10 for as-dep. GST
44 +/- 3 for crystalline GST

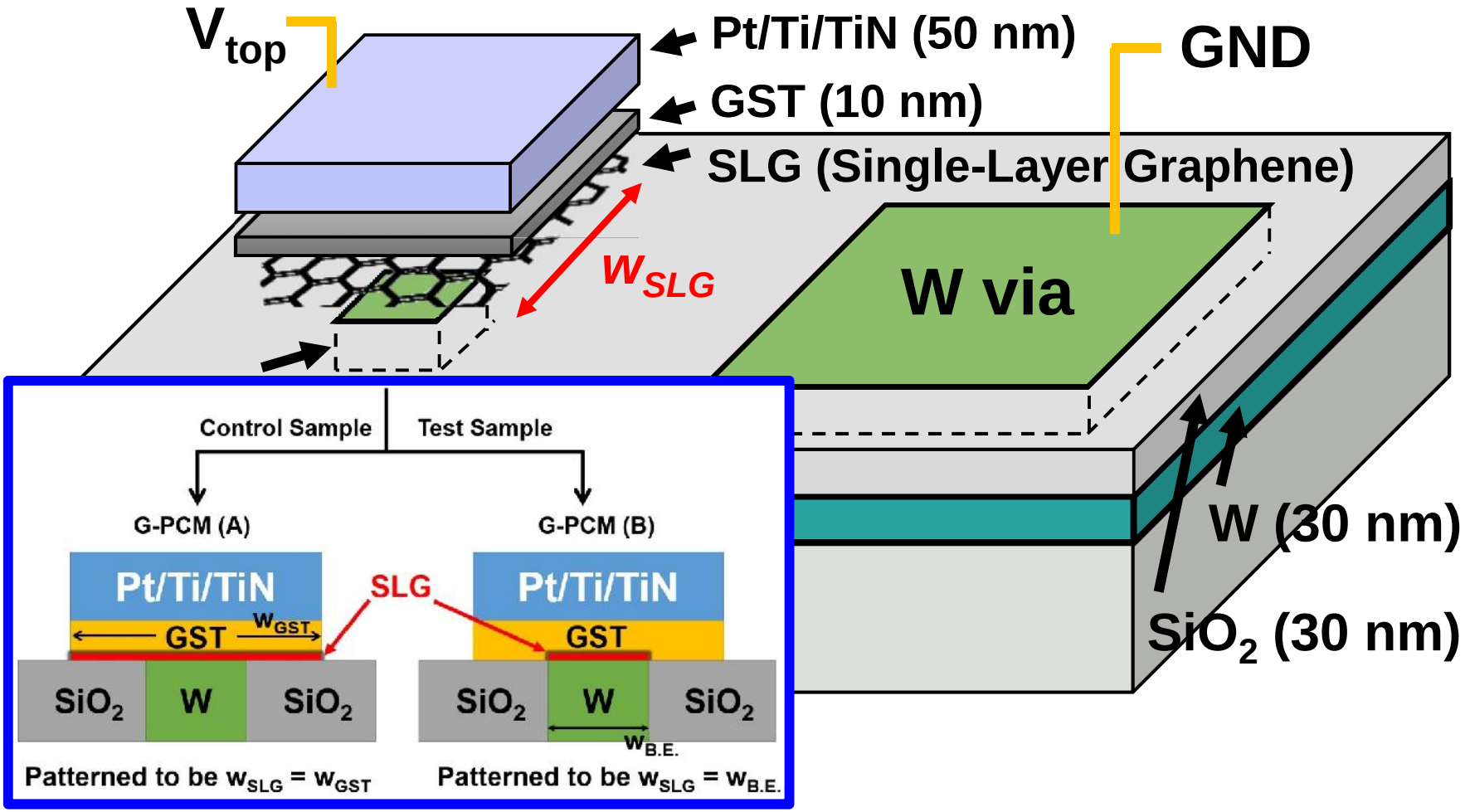
Ahn et al. Nano Letters 2015

G-PCM: **Device structure**



Ahn et al. Nano Letters 2015

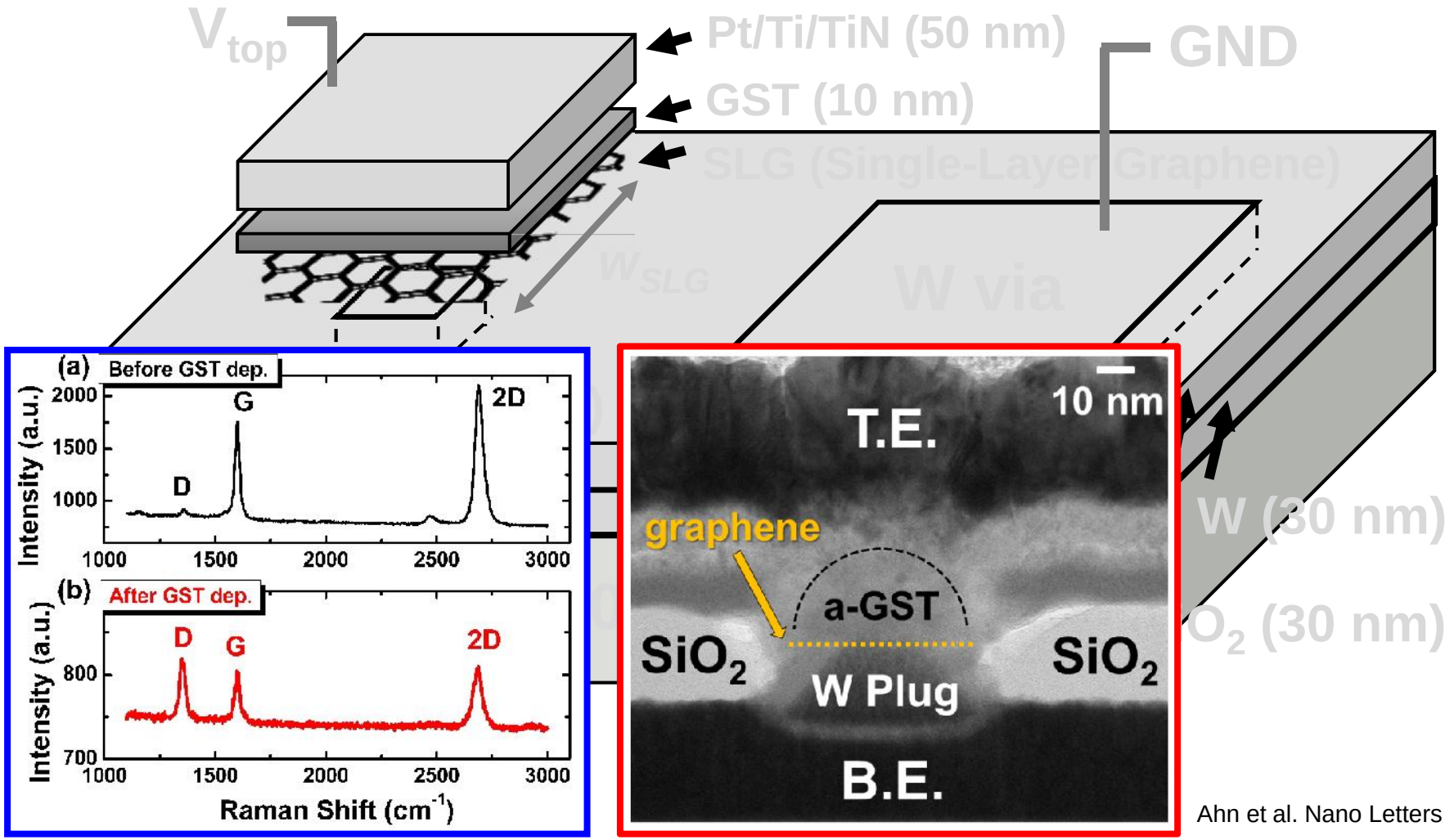
G-PCM: Device structure



Active device region

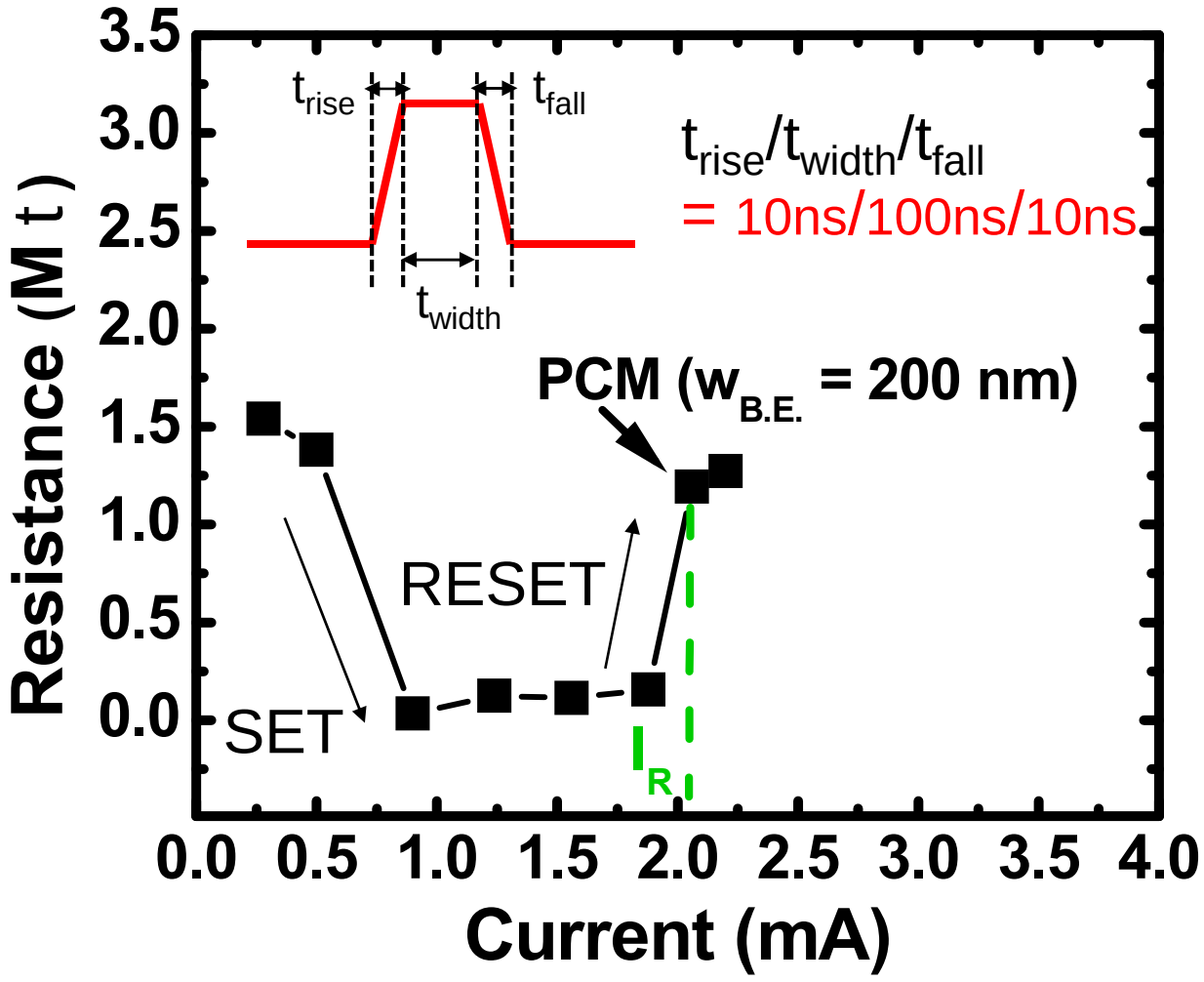
Ahn et al. Nano Letters 2015

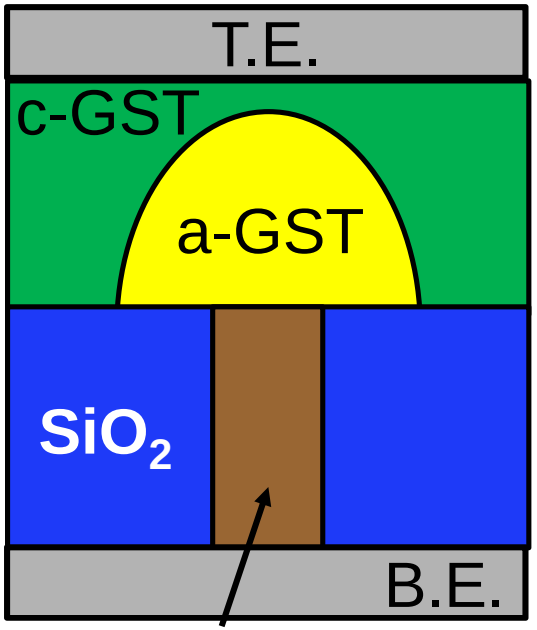
G-PCM: **Raman & TEM studies**



Ahn et al. Nano Letters 2015

G-PCM: I_{RESET} of traditional PCM



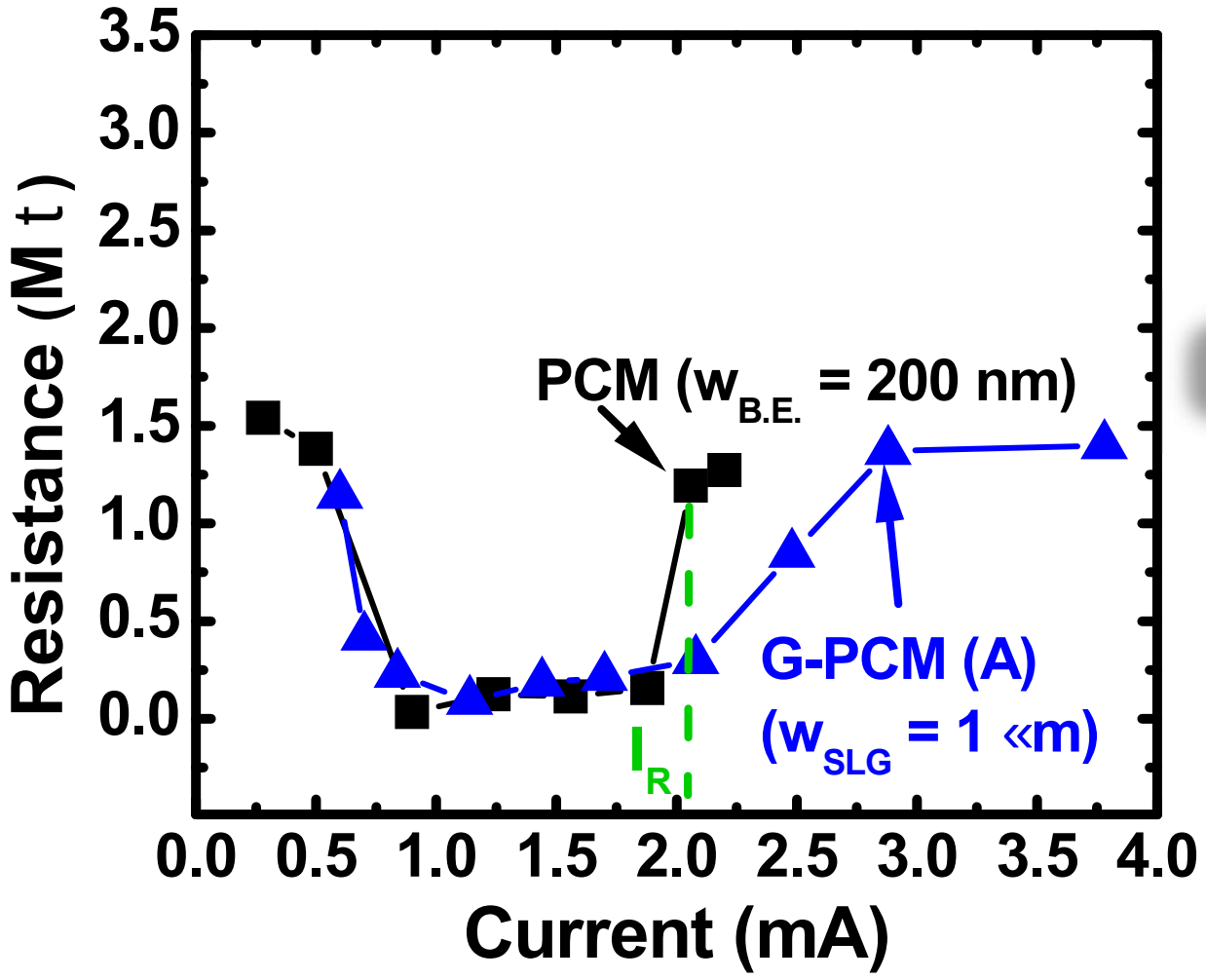


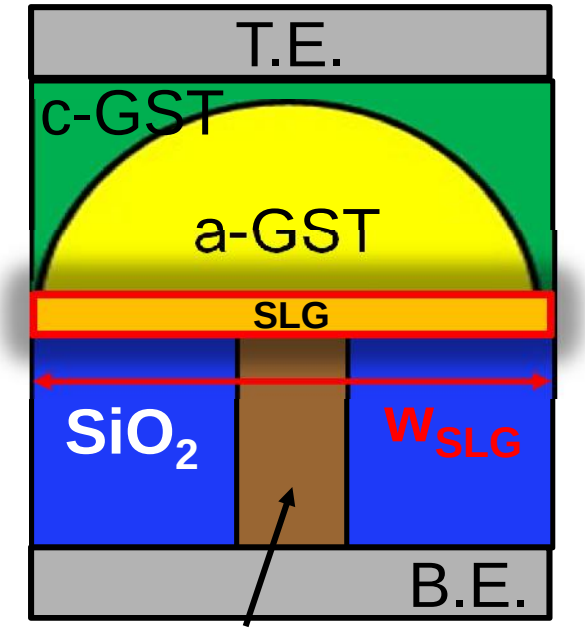
Heater (W plug)

"PCM"

Ahn et al. Nano Letters 2015

G-PCM: I_{RESET} of GPCM (A) – Control sample



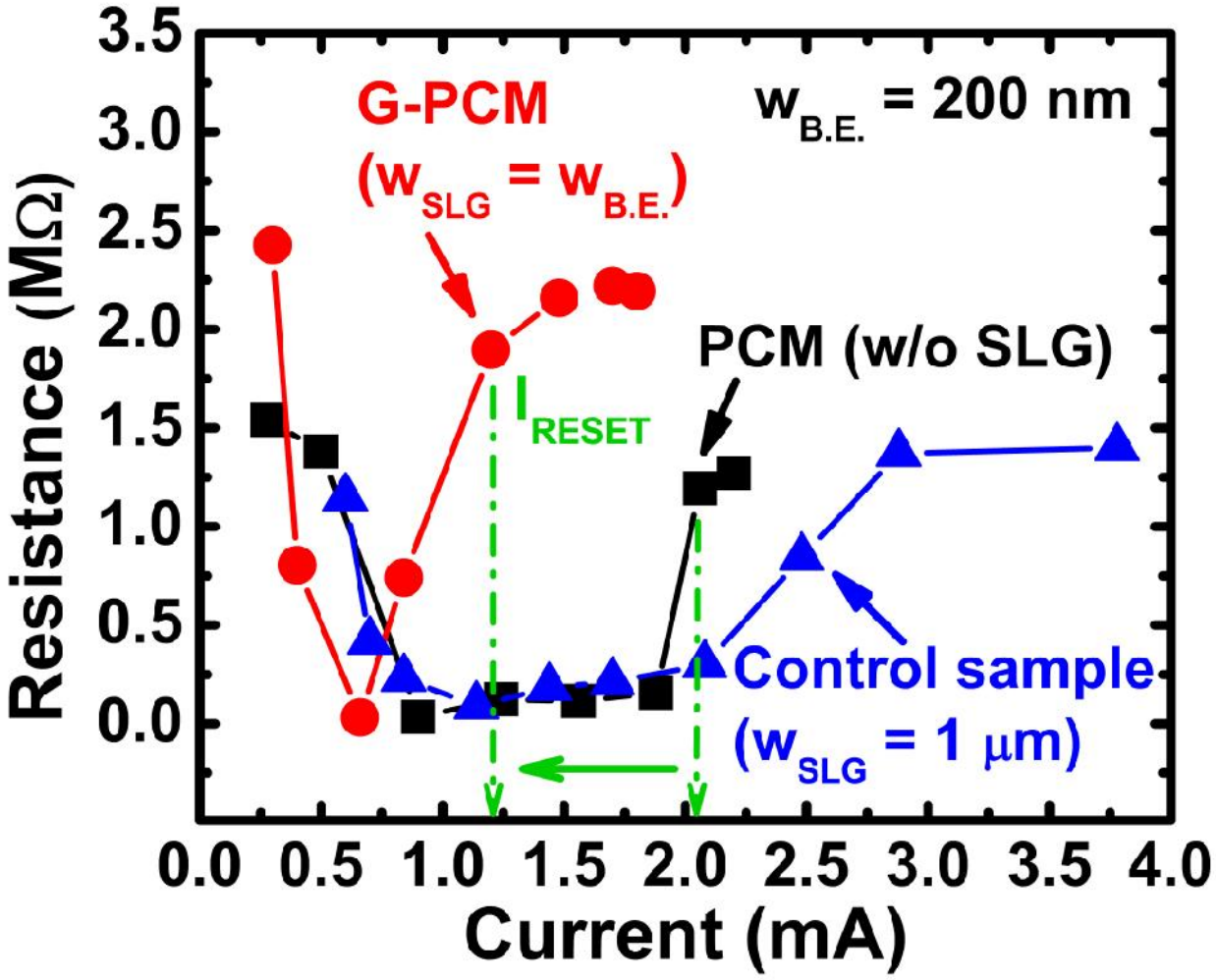


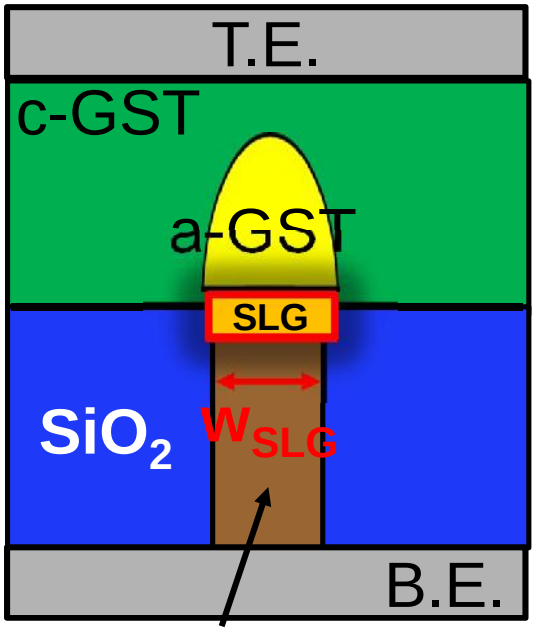
Heater (W plug)

"G-PCM (A)"

Ahn et al. Nano Letters 2015

G-PCM: I_{RESET} of GPCM (B) – Optimal design



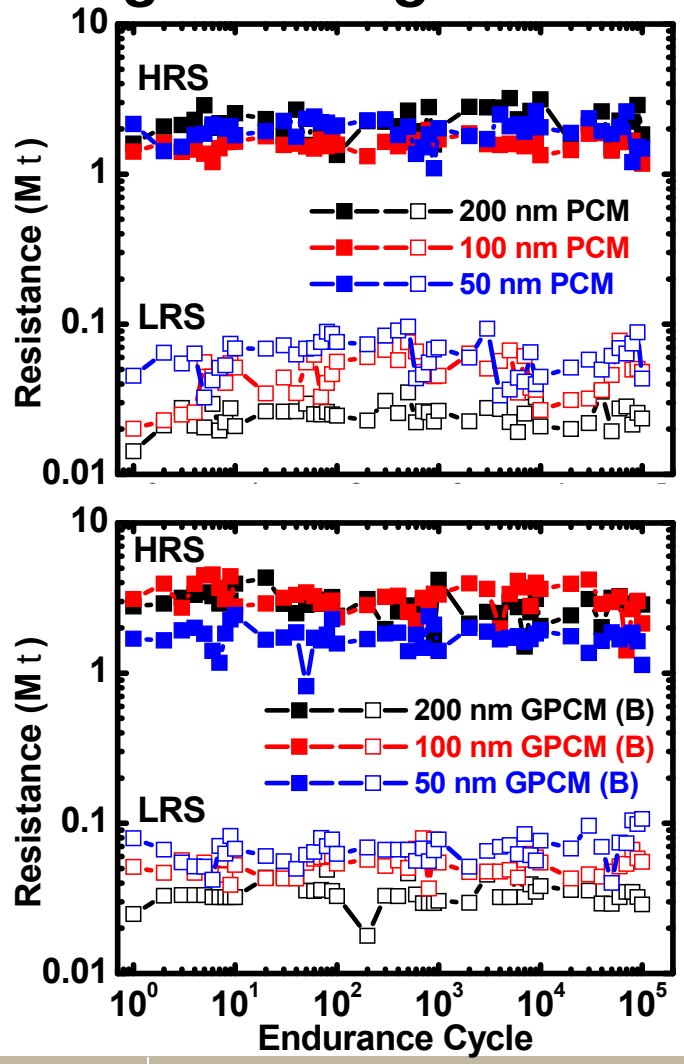


"G-PCM (B)"

Ahn et al. Nano Letters 2015

G-PCM: Verifying endurance is of great importance

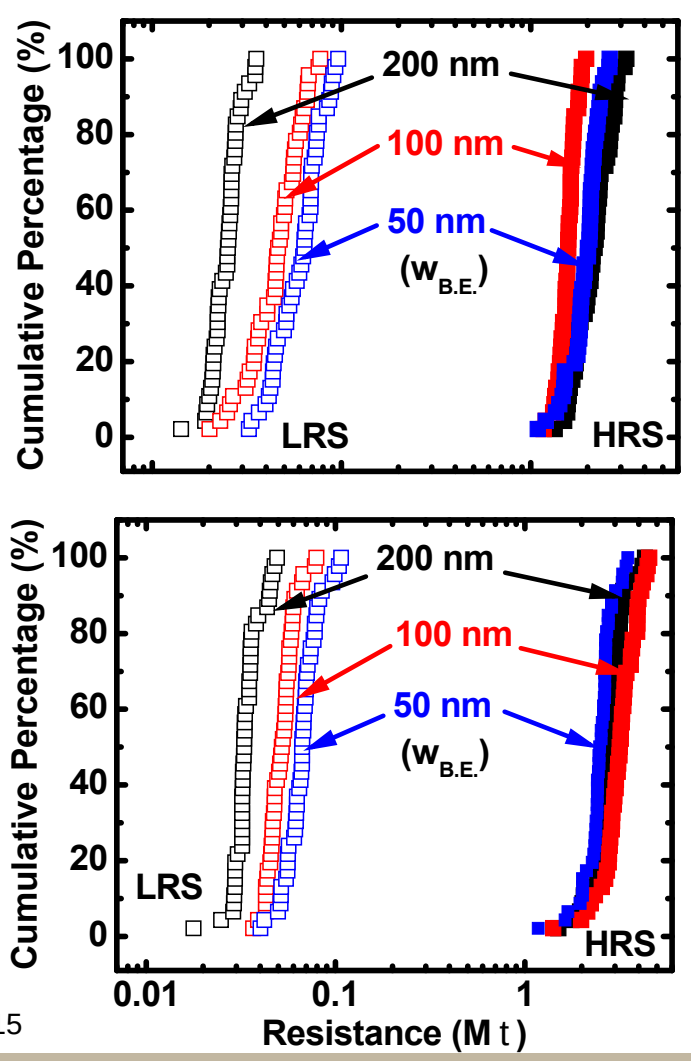
"Programming endurance"



PCM

G-PCM (B)

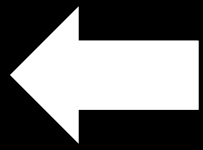
"Resistance distribution"



Ahn et al. Nano Letters 2015

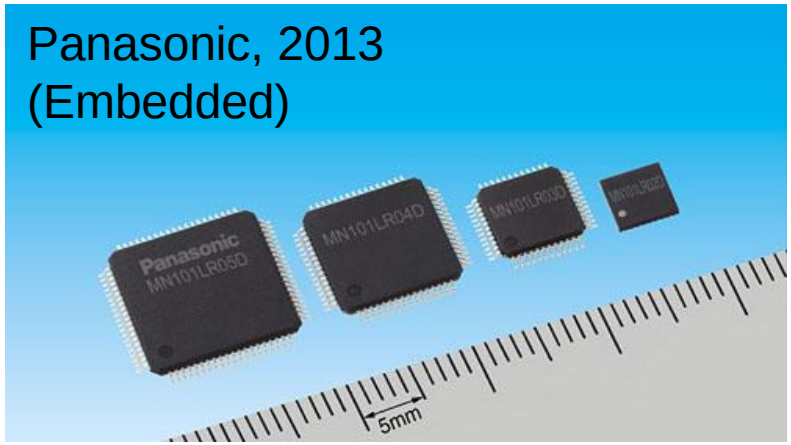
1. Energy-efficient *Cell* design

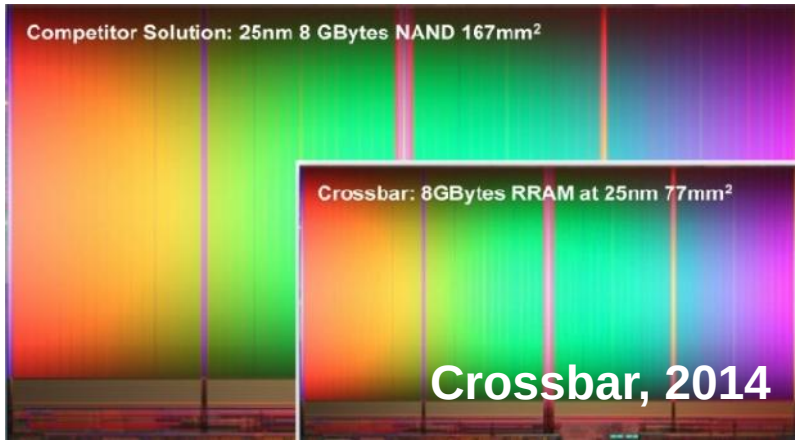
2. Energy-efficient *Architecture* design

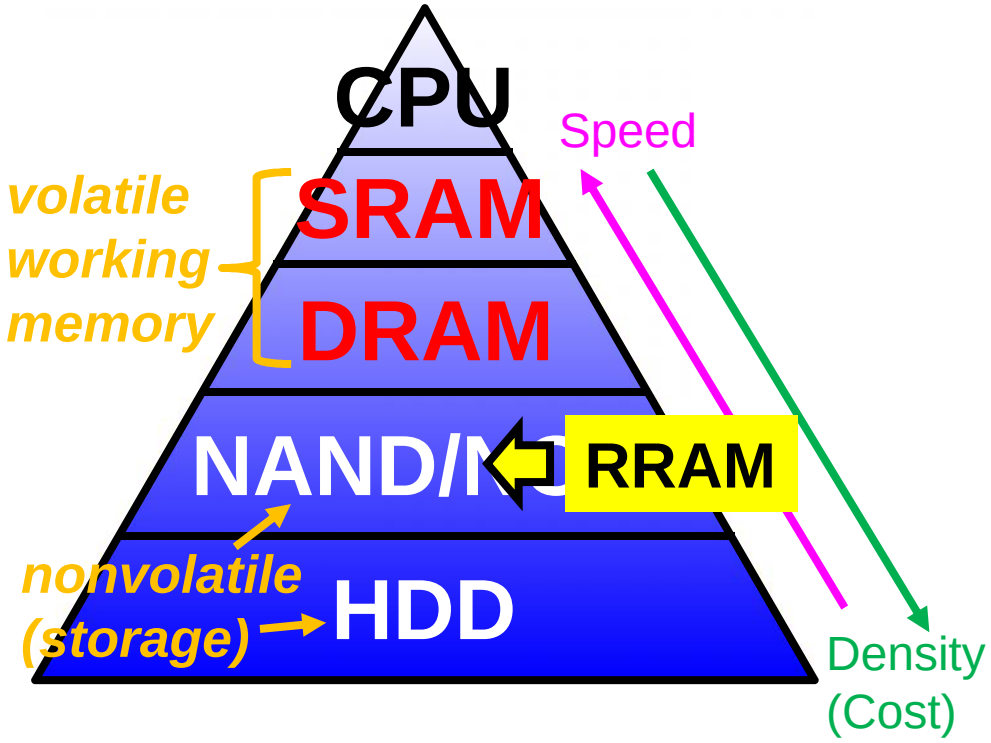


RRAM

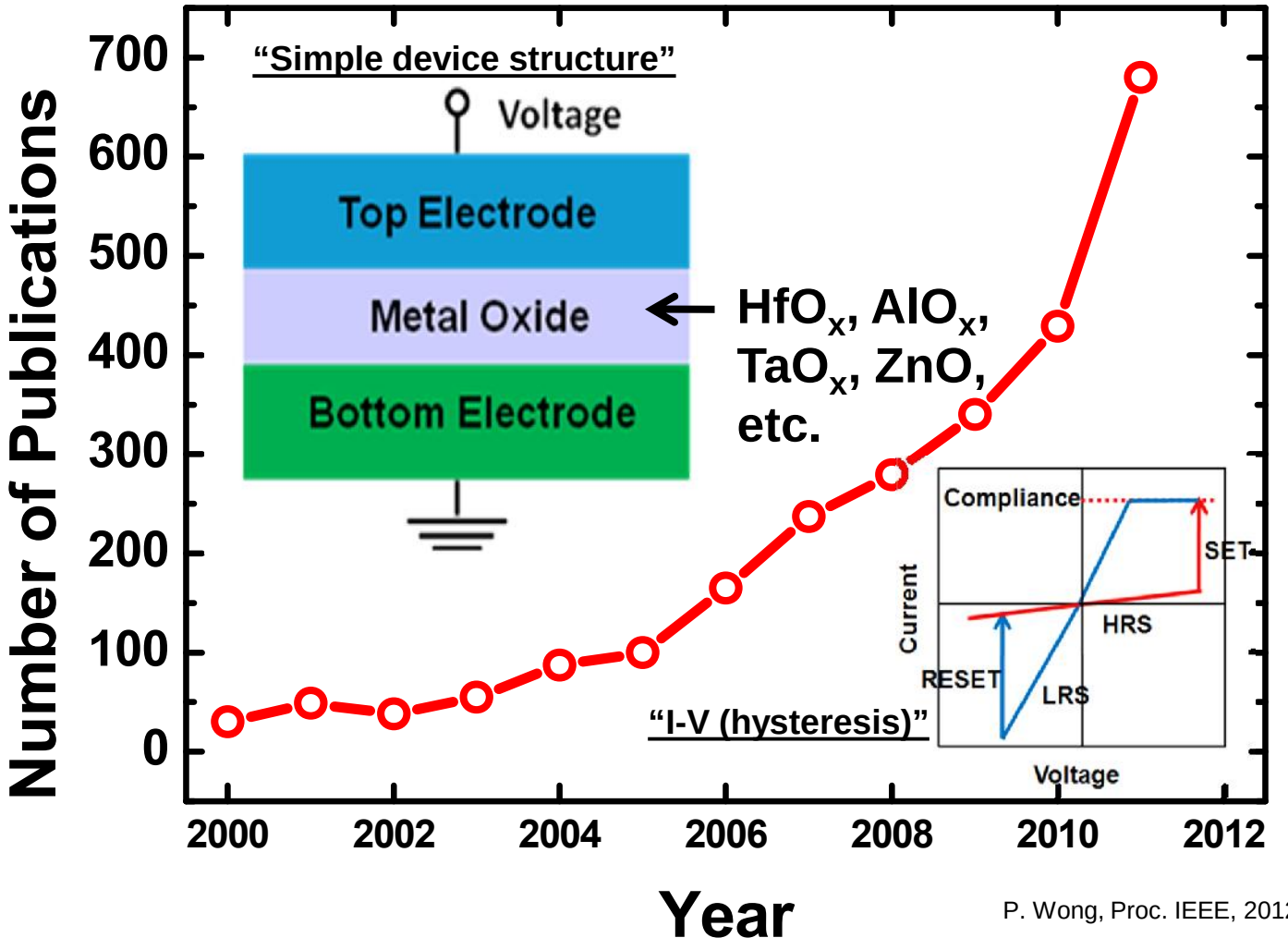
: Resistive (switching) RAM or Metal-oxide RAM





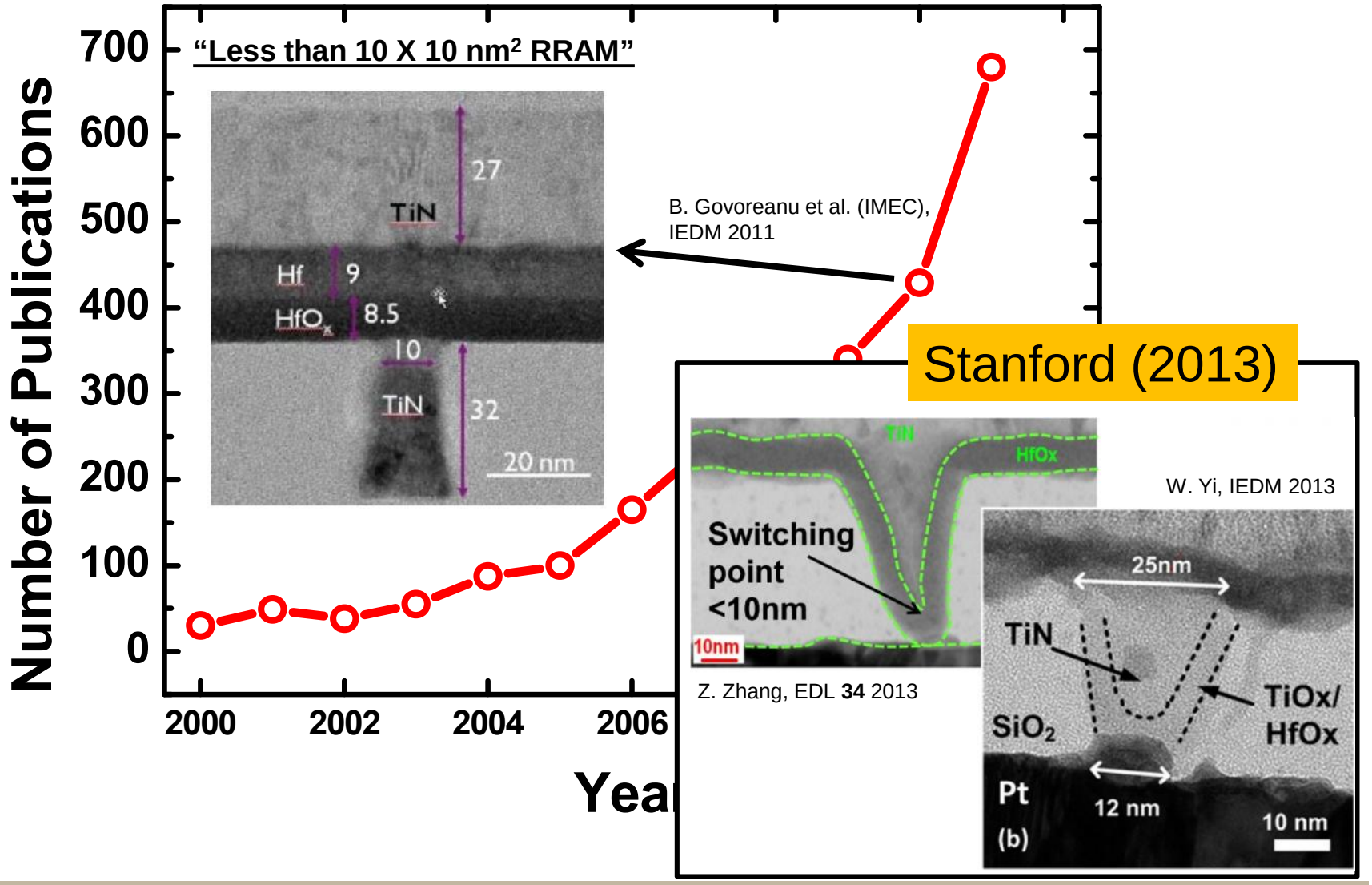


RRAM: **Emerging candidate** for sub-10 nm NVM



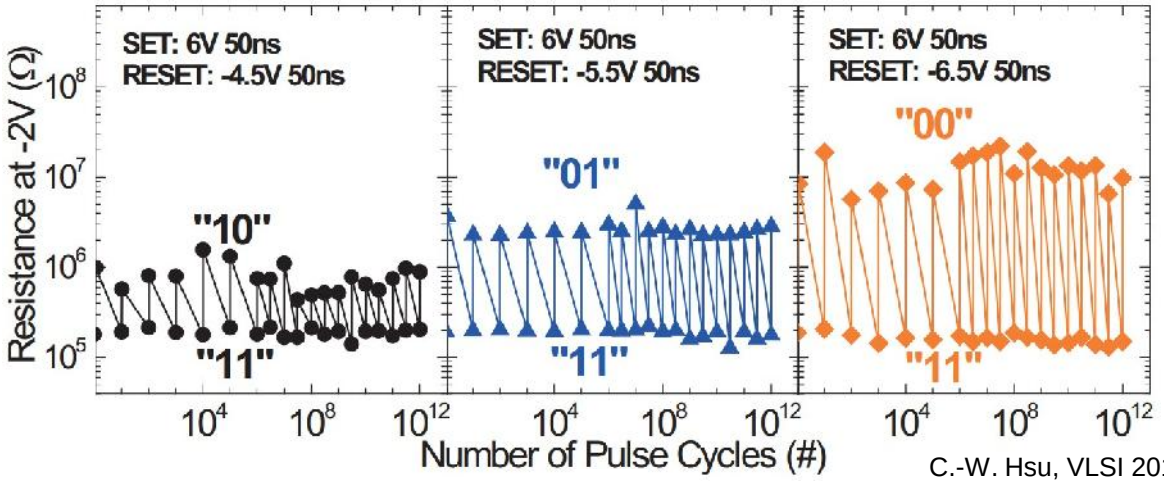
P. Wong, Proc. IEEE, 2012

RRAM: **Emerging candidate** for sub-10 nm NVM



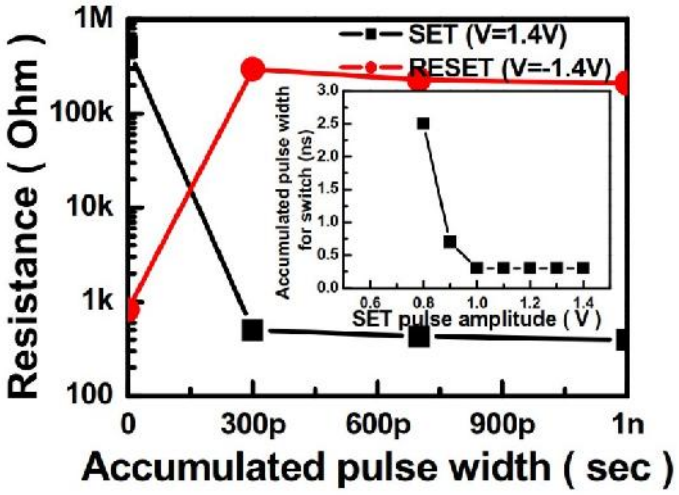
RRAM: **Key attributes** (electrical performance)

- **High endurance (> 10^{12} cycles) with MLC**



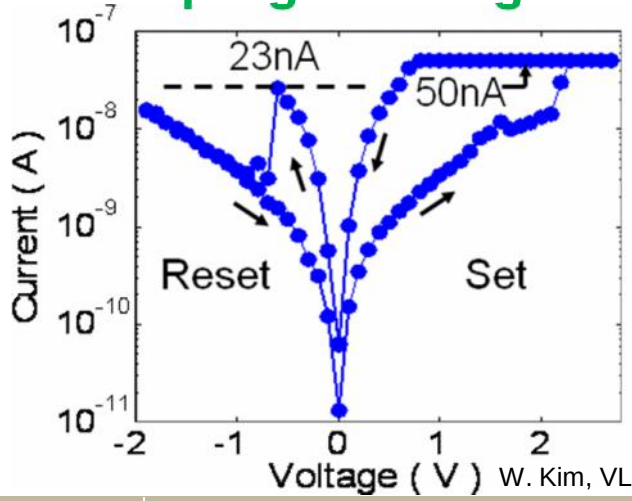
C.-W. Hsu, VLSI 2013

- **High speed (< 1 ns)**

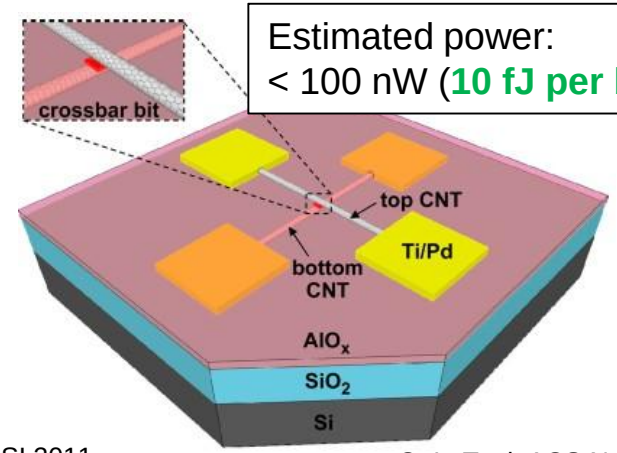


H. Y. Lee, IEDM 2010

- **Low programming voltage, current, and power**



W. Kim, VLSI 2011



Estimated power:
< 100 nW (**10 fJ per bit**)

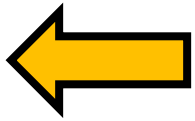
C.-L. Tsai, ACS Nano **7**, 2013

- CMOS compatible
- Low temperature
- High degree of freedom (engineering design)
- …

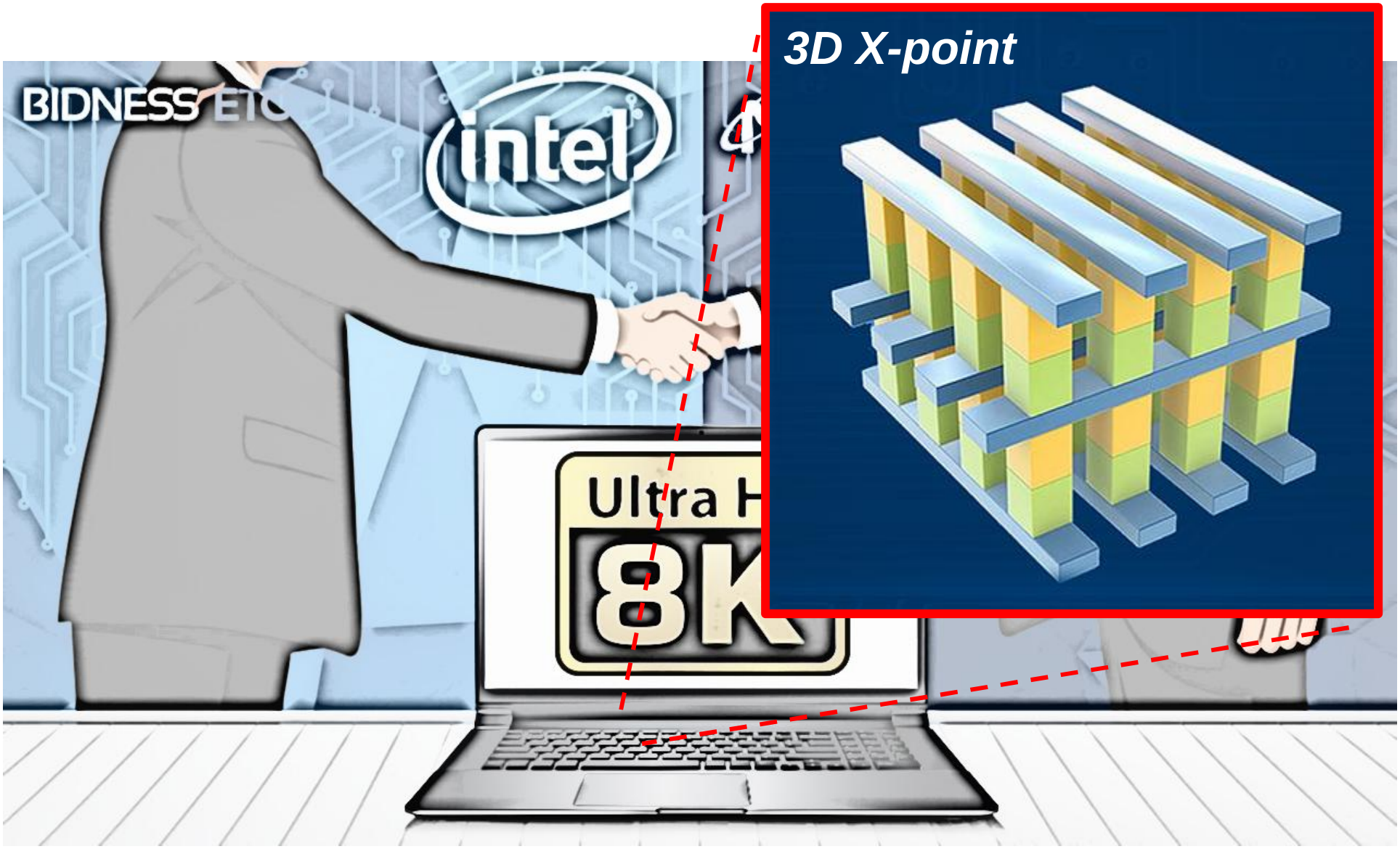
RRAM: **Key challenges and issues**

- Physics of resistive switching and conduction
- Array architecture
- Killer application

RRAM: **Key challenges and issues**

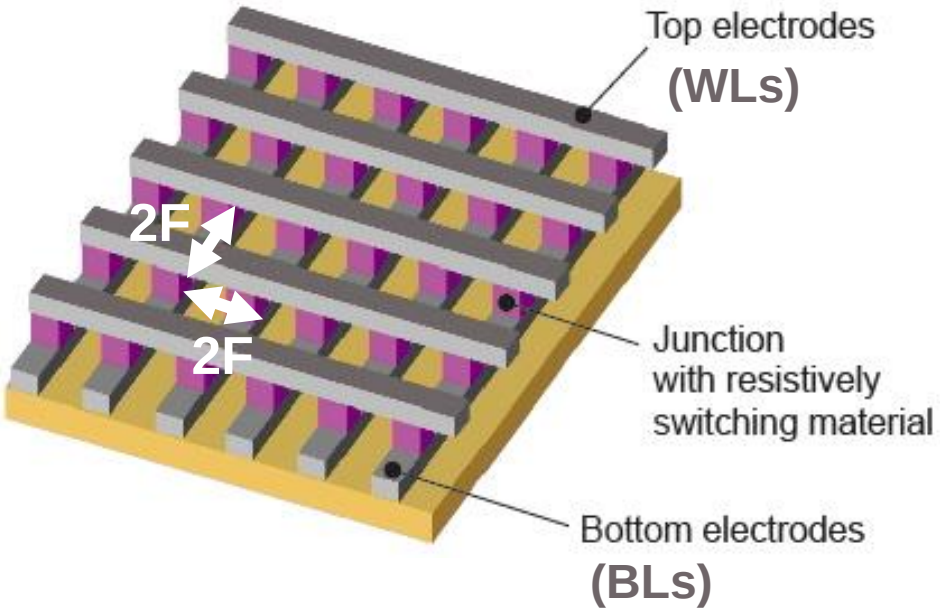
- Physics of resistive switching and conduction
- **Array architecture** 
- Killer application

July 2015



July 2015

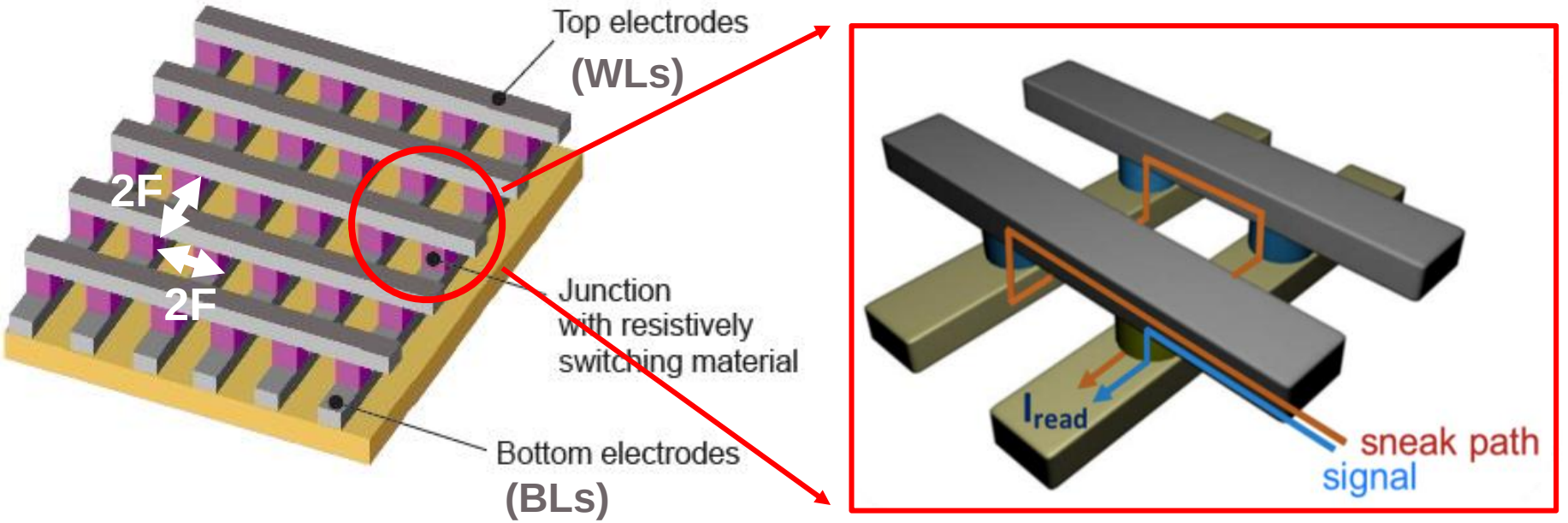
RRAM array: **Cross-point structure**



"attractive," due to

- ❑ Easy to fabricate
- ❑ Small cell size: **$4F^2$**
- ❑ Potential for 3D stacking
  ($4F^2/N$, N = number of layers)

RRAM array: **Cross-point structure**



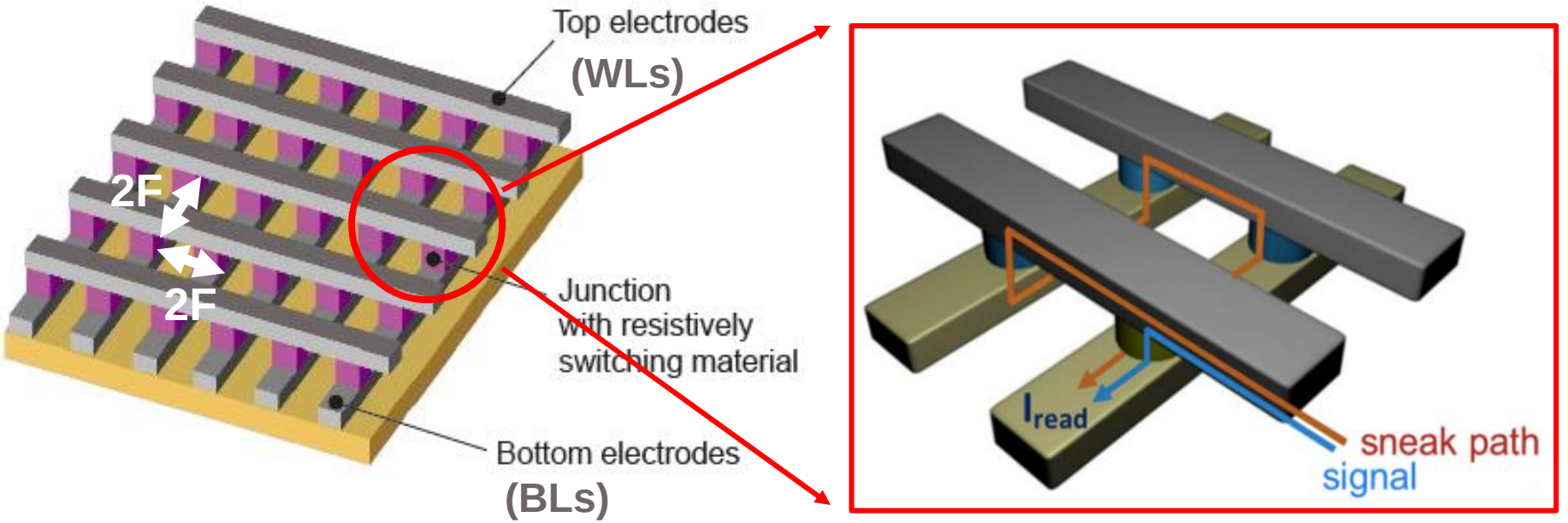
"attractive," due to

- ❑ Easy to fabricate
- ❑ Small cell size: **$4F^2$**
- ❑ Potential for 3D stacking ($4F^2/N$, N = number of layers)

"problematic," due to

- ❑ **Sneak path problem**
  - Increased power consumption
  - Reduced write/read margin (limiting maximum allowable array size)

RRAM array: **Cross-point structure**



"attractive," due to

- East to fabricate
- Small cell size: $4F^2$
- Potential for 3D stacking ($4F^2/N$, N = number of layers)

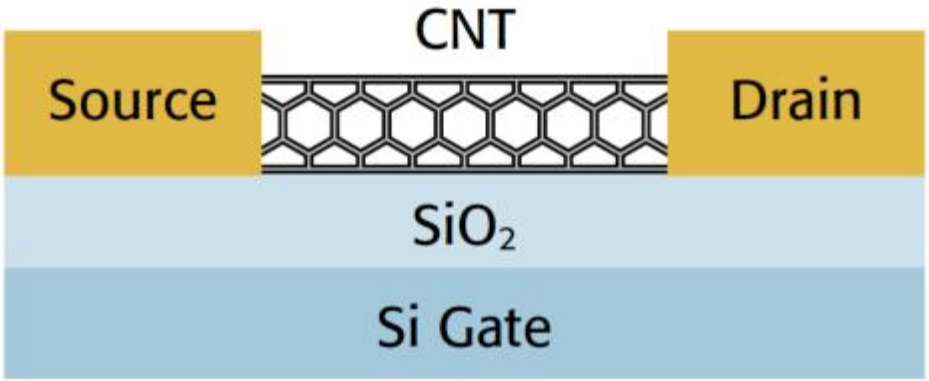
"problematic," due to

We need "Selection Device" to cut-off sneak leakage current

(limiting maximum allowable array size)

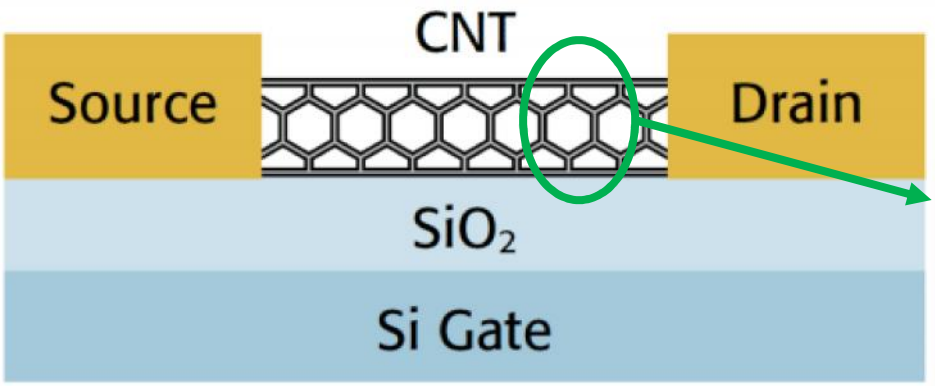
“What selector will be the best choice for you?”

Carbon nanotube field-effect transistors (CNFETs)



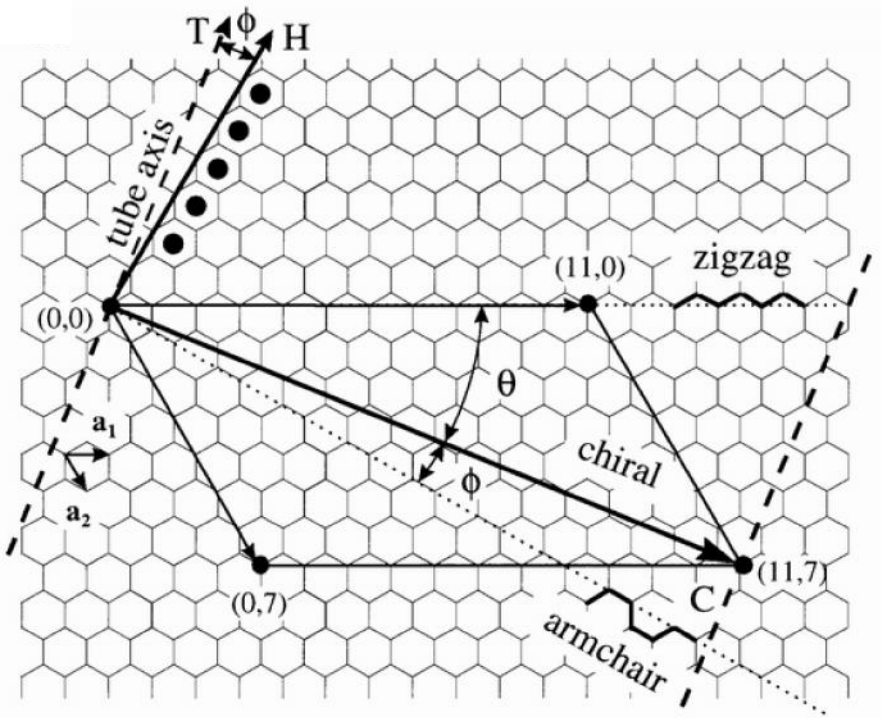
Schematic representation of CNFET

Carbon nanotube field-effect transistors (CNFETs)



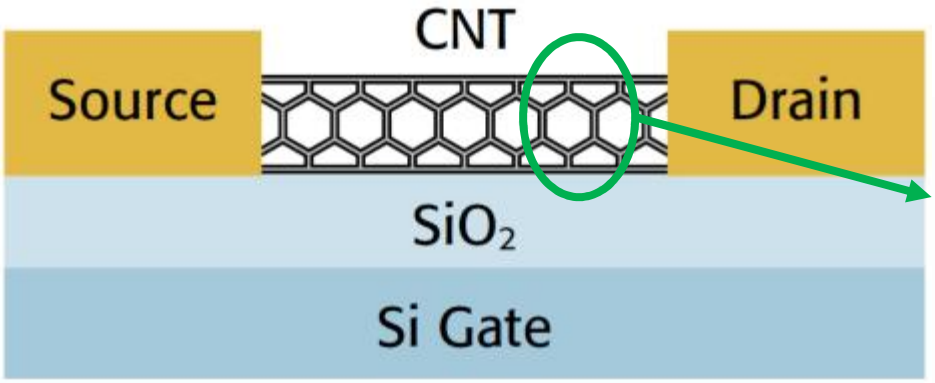
Schematic representation of CNFET

Rolling up a sheet of graphene



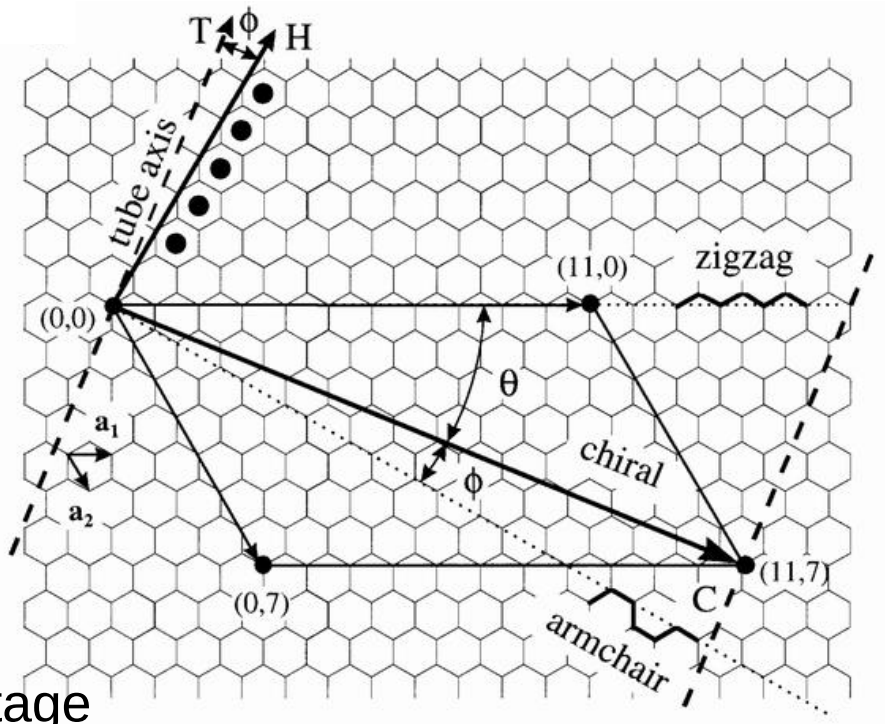
Ref: J. Wilder et al, Nature **391**, 1998

Carbon nanotube field-effect transistors (CNFETs)



Schematic representation of CNFET

Rolling up a sheet of graphene



Ref: J. Wilder et al, Nature **391**, 1998

"Why CNTs to replace Si?"

Ballistic transport → Higher on-current
Lower operating voltage

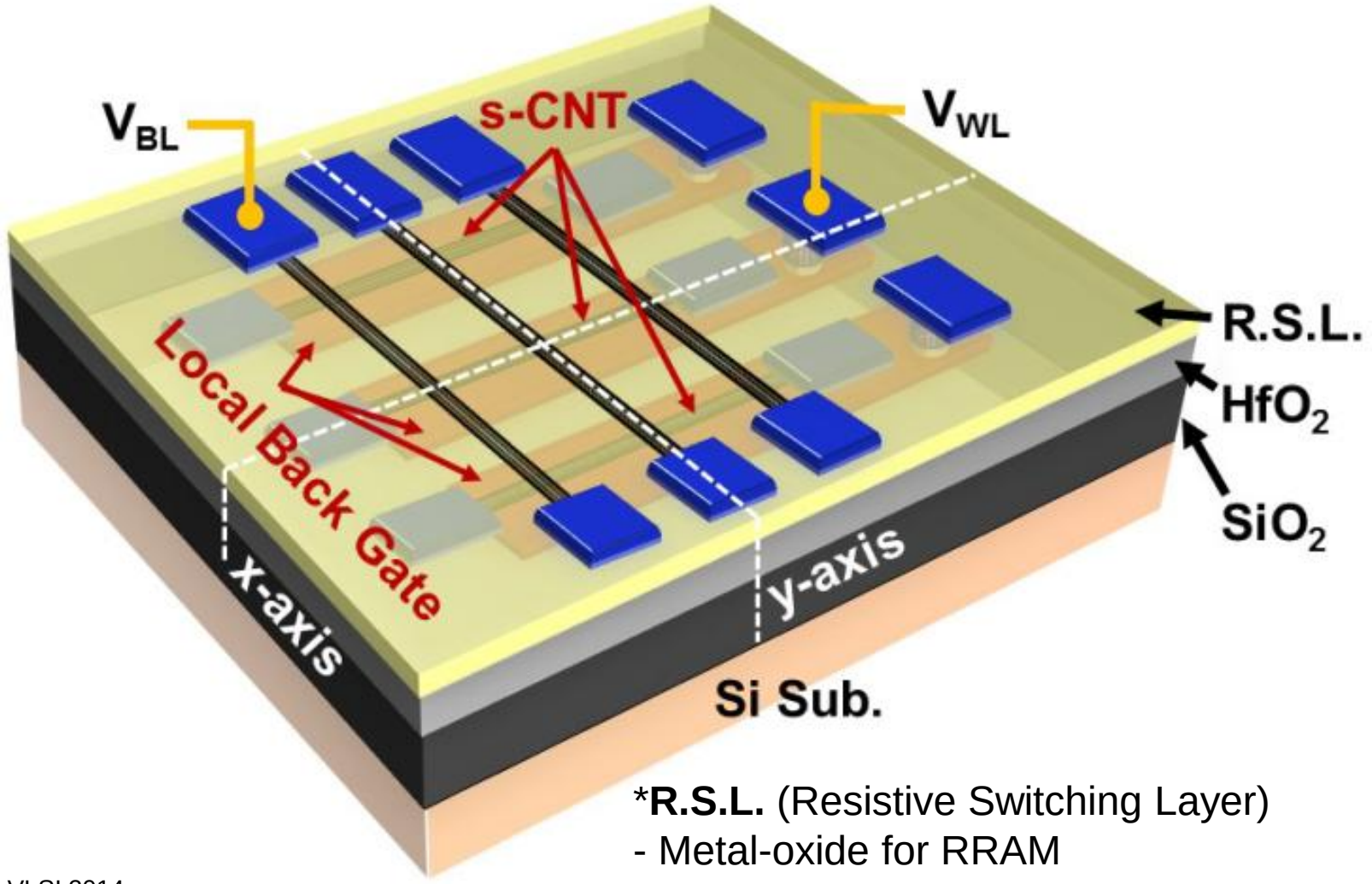
Ultra-thin body → Aggressive scaling
Excellent electrostatic control

CNFET: Ideal selection device for memory array

- **High forward-current (I_{on}) densities (J_{on})**
  to program aggressively scaled memory device
  J_{on} > 10 MA/cm^2
- **High On/off ratio (I_{on}/I_{off})**
  to have high selectivity of memory bits
  I_{on}/I_{off} > 10^6
- **Low off-current (I_{off})**
  to accommodate un- and half-selected cells
  I_{off} < 10 pA
- **Low processing temperature (T)**
  to allow 3D stacking
  T < 300 °C
- **Bipolar operation**
  to allow for best-of-breed RRAM
  Symmetric I-V at both polarities

+ "small device area"

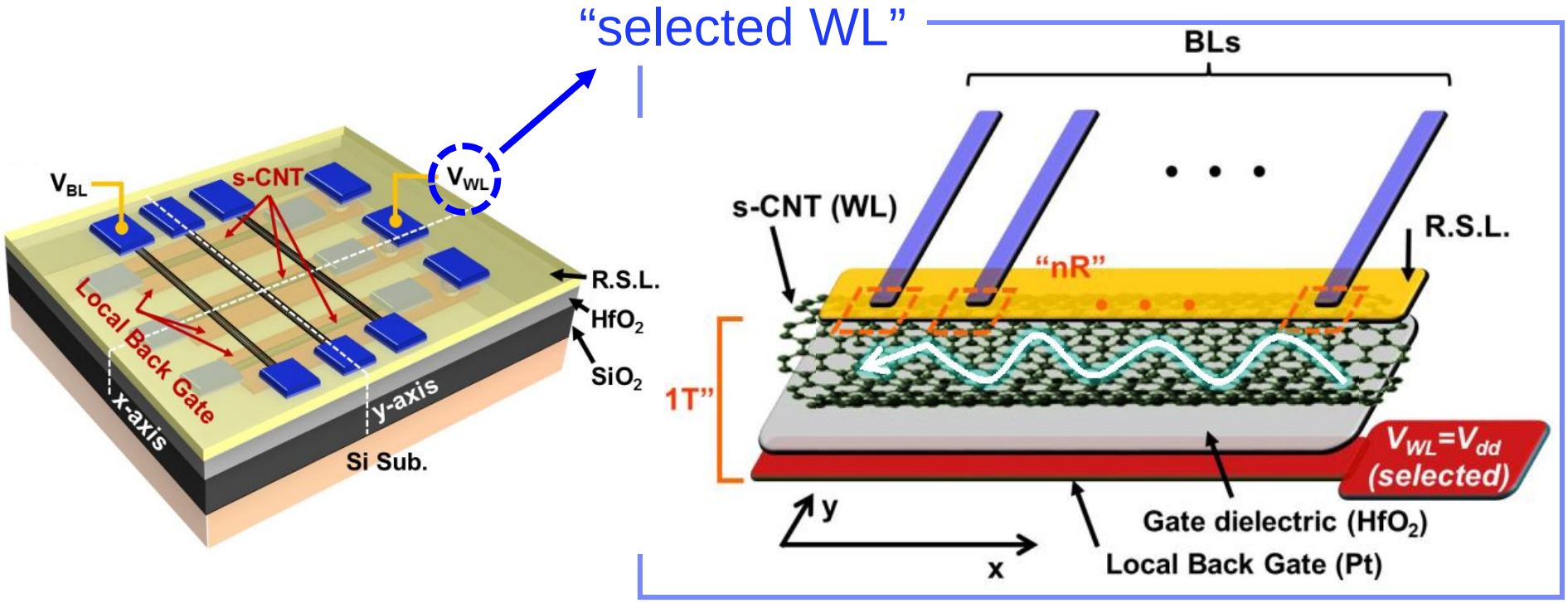
1TnR array: **Based on CNFET selection device**



*__R.S.L.__ (Resistive Switching Layer)
- Metal-oxide for RRAM
- Phase-change material for PCM

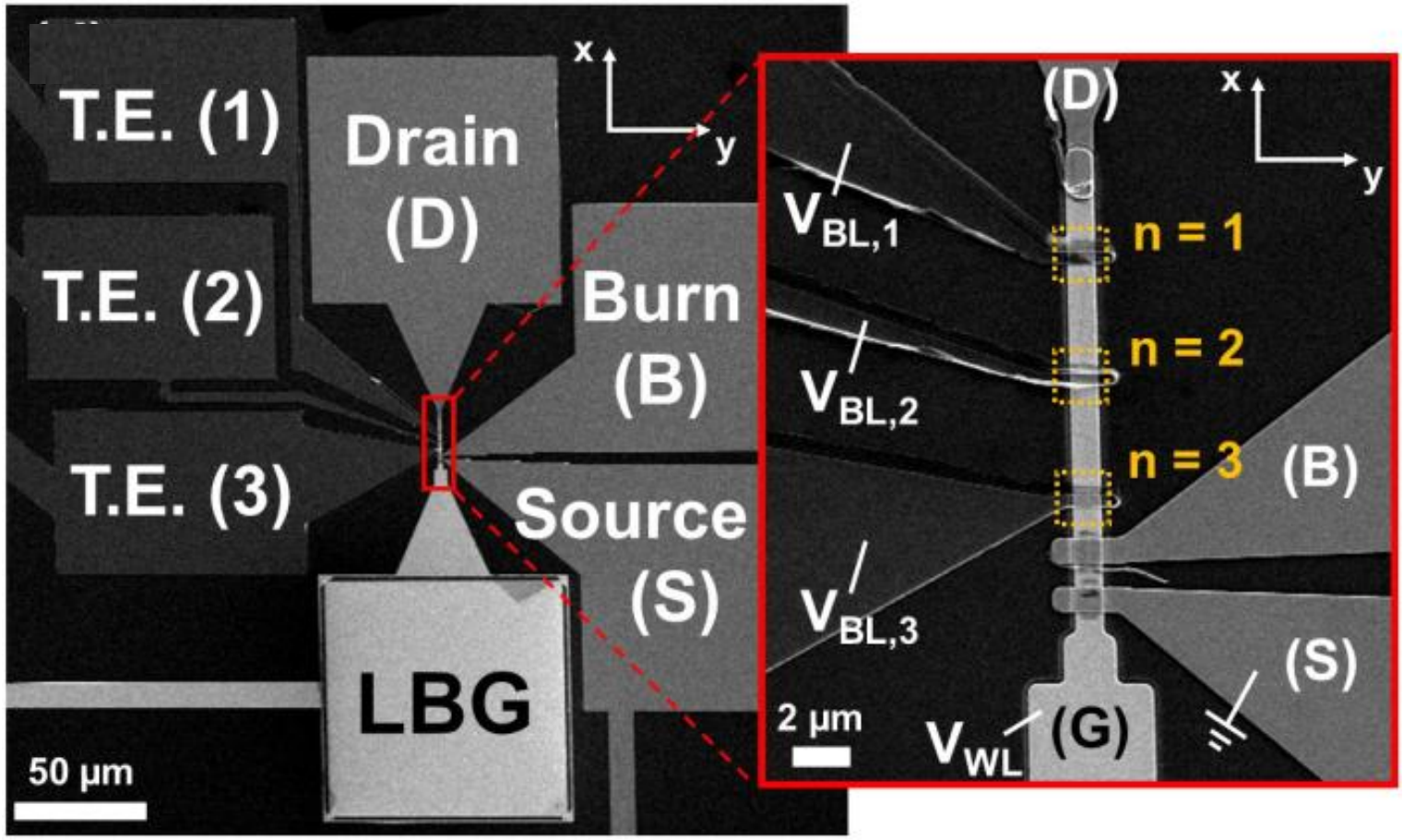
Ahn et al. VLSI 2014
Ahn et al. IEEE TED 2015

1TnR array: **Reduced sneak leakage**



"Sneak leakage is much reduced from 2D to 1D,"
as it is confined within the 1D CNT channel

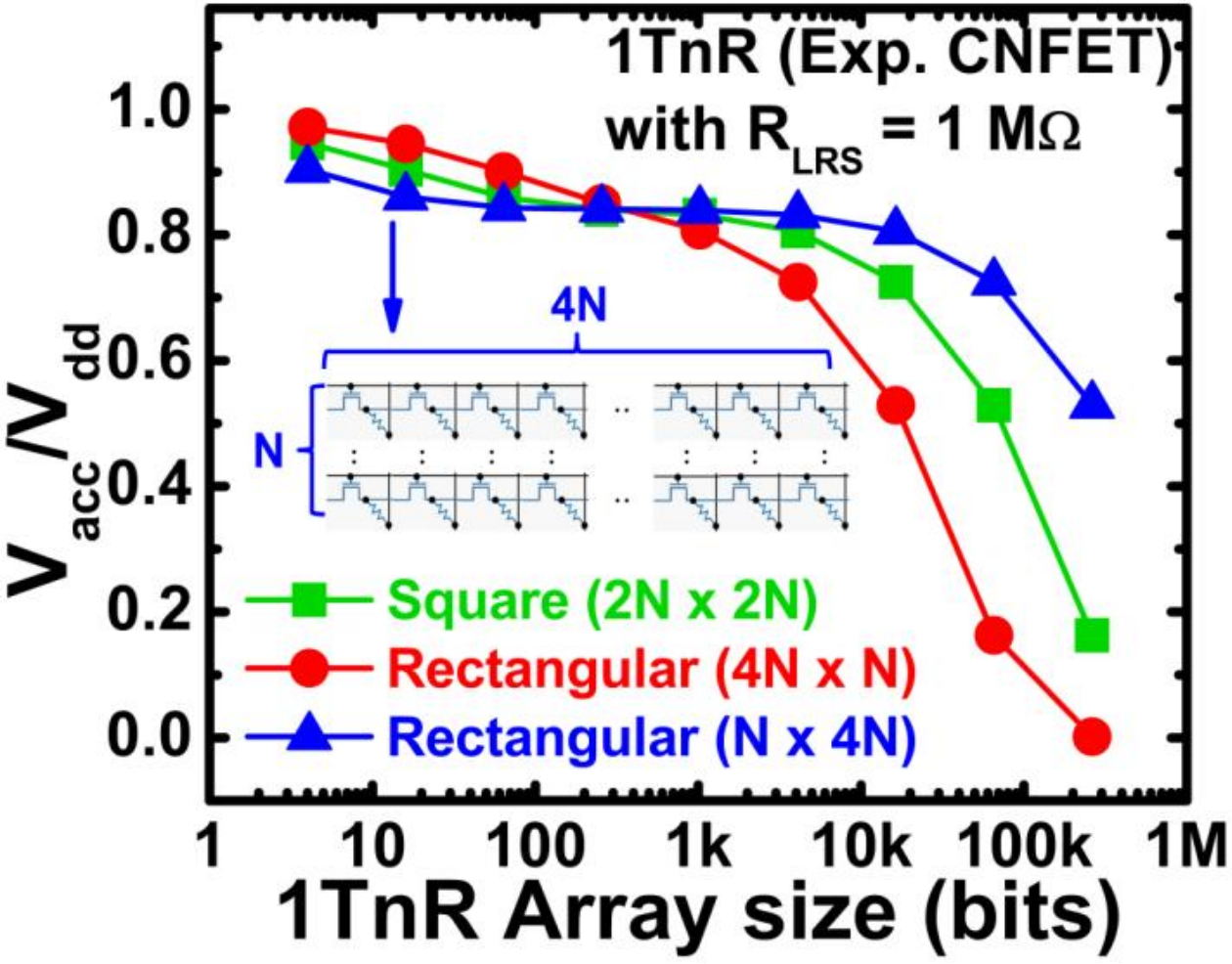
1TnR: (1) Requires NO additional contacts/wiring



"CNFET selector is tightly integrated, with CNT as B.E."

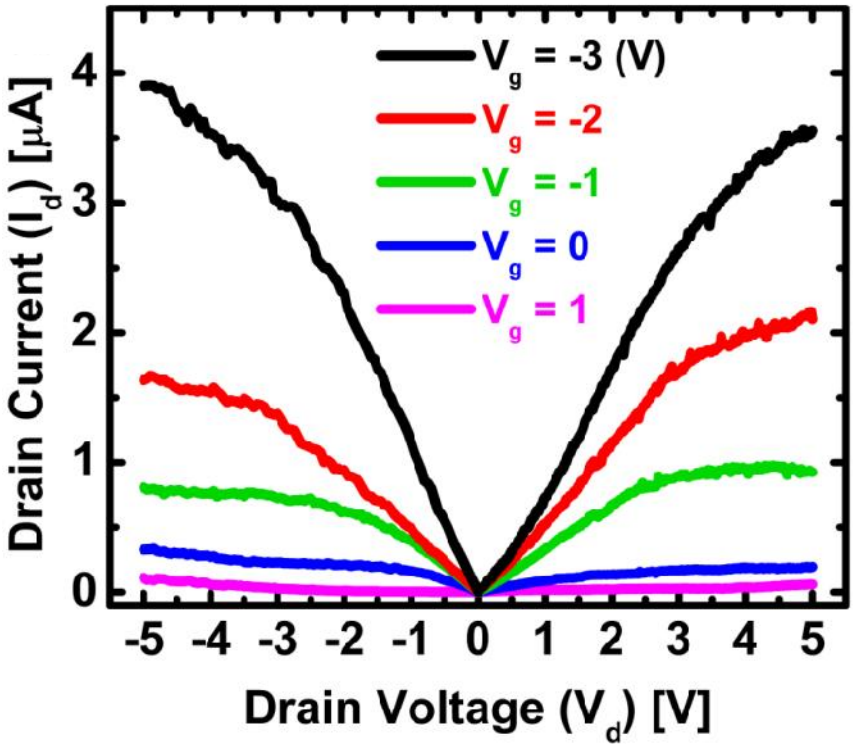
Ahn et al. VLSI 2014
Ahn et al. IEEE TED 2015

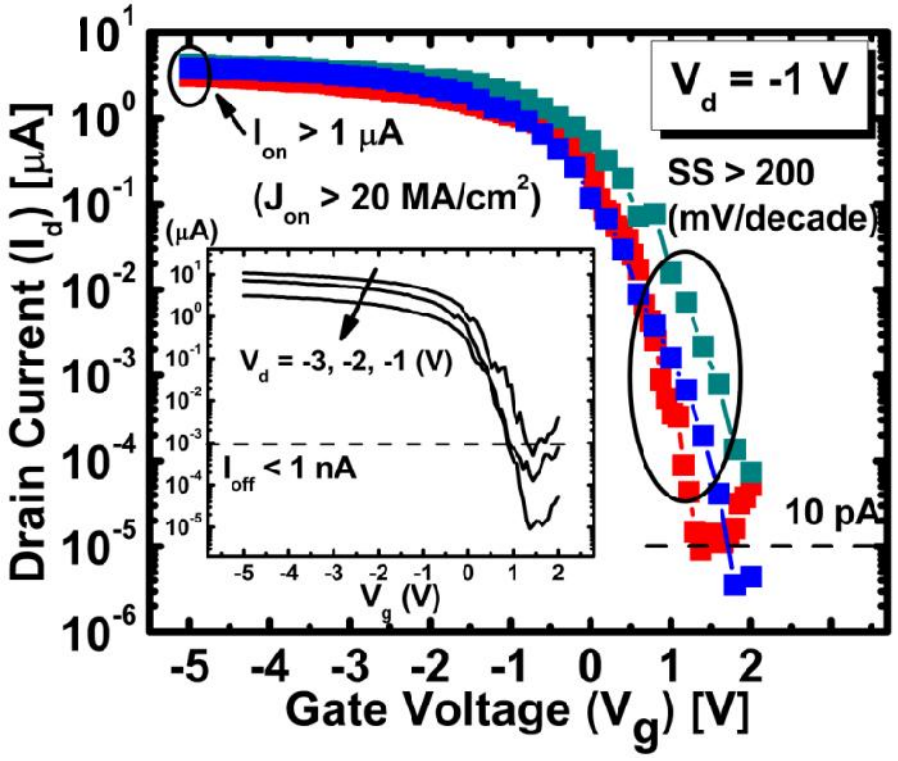
1TnR: (2) Rectangular array preferred



Ahn et al. VLSI 2014
Ahn et al. IEEE TED 2015

Electrical results: **IVs of fabricated CNFETs**





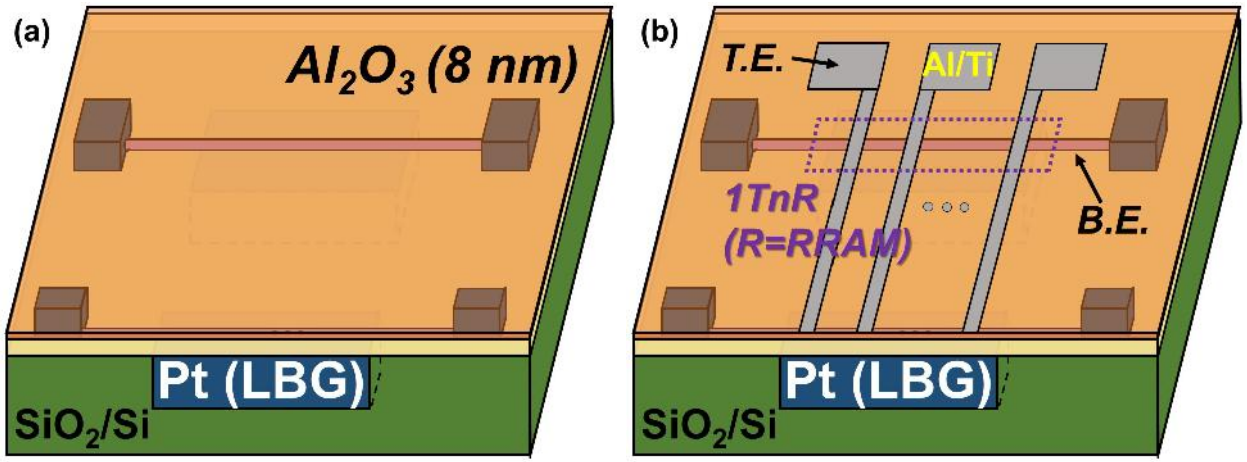
High current-drive (> 20 MA/cm^2)
Good electrostatic control by gate
Near-symmetric I-V (bipolar)

High on/off ratio (10^5 ~ 10^7)
Low leakage current < 10 pA (even at high V_d)

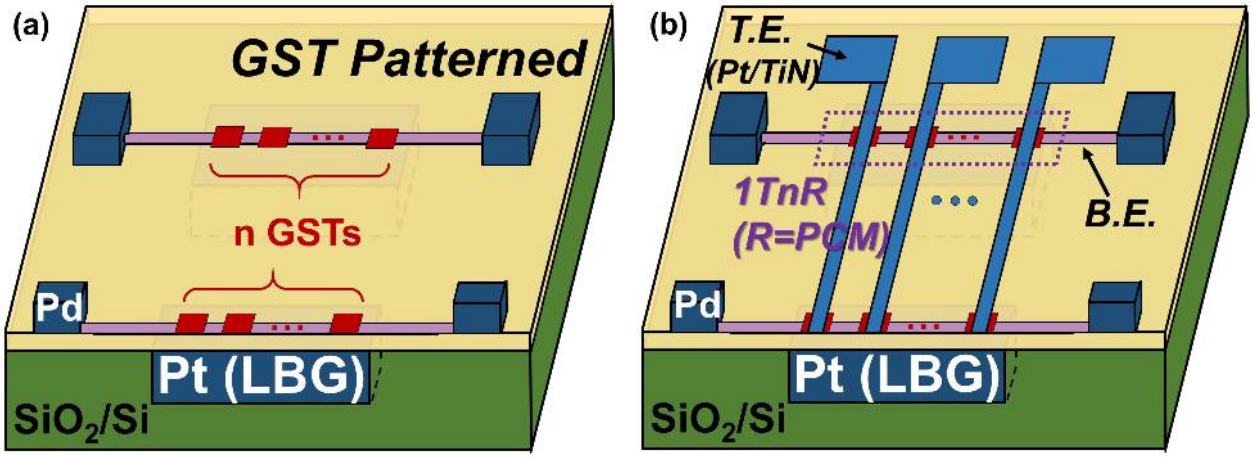
Ahn et al. IEEE TED 2015

Integration: CNFET + Memory = 1TnR

"Integrating with Al_2O_3-based RRAM"



"Integrating with PCM (GST)"

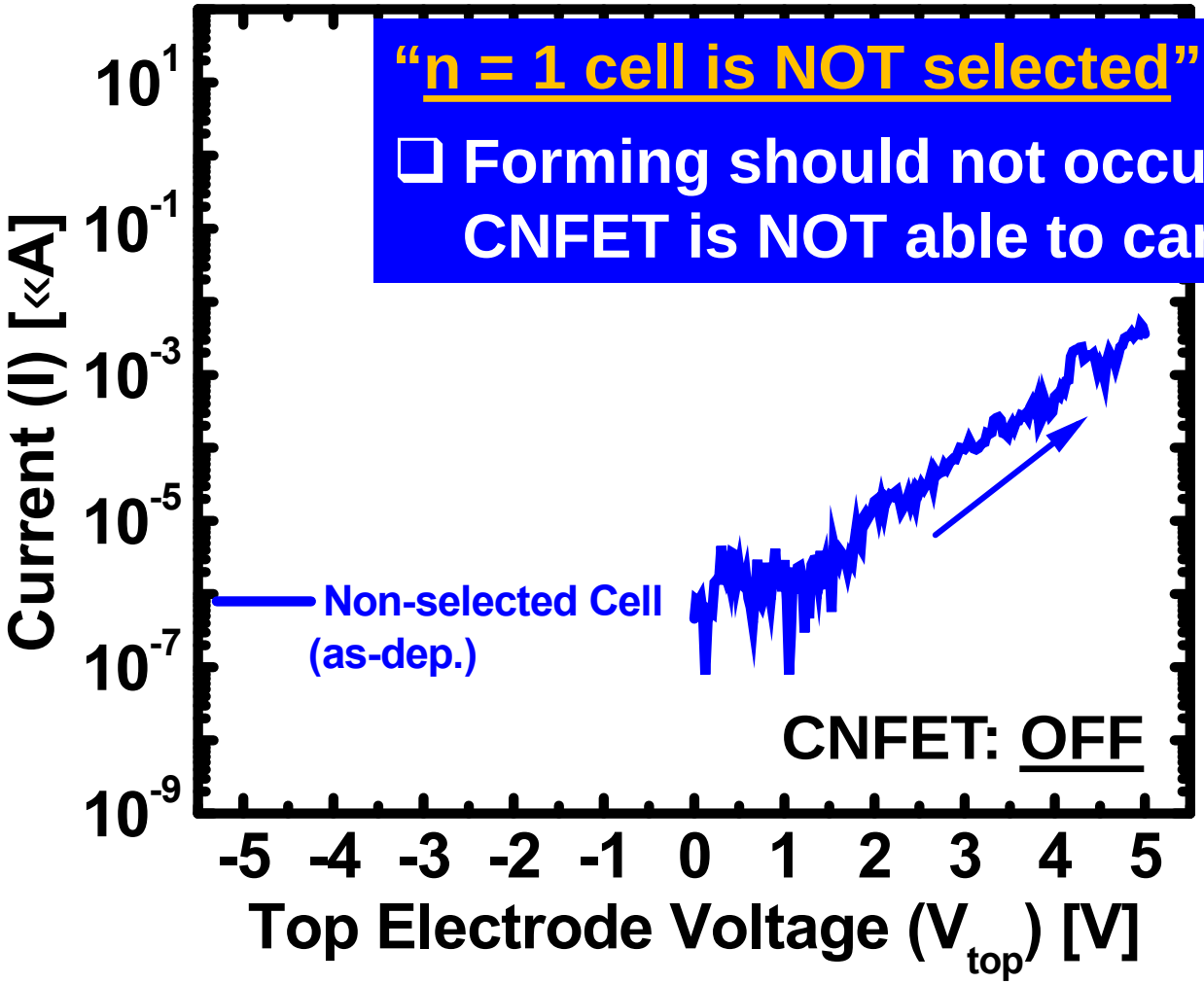


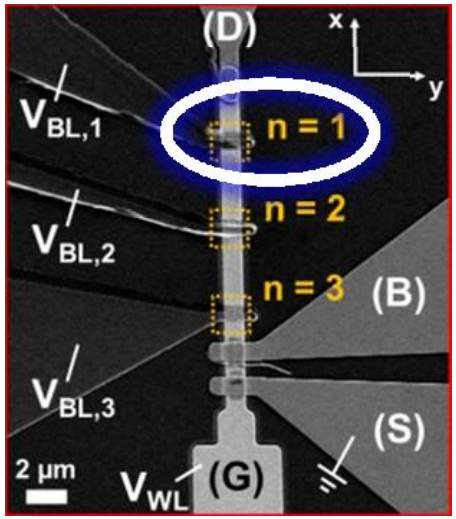
Ahn et al. IEEE TED 2015

Electrical results: 1TnR RRAM – Selective switching (DC)

"n = 1 cell is NOT selected"

- **Forming should not occur, as the turned-off CNFET is NOT able to carry high currents**





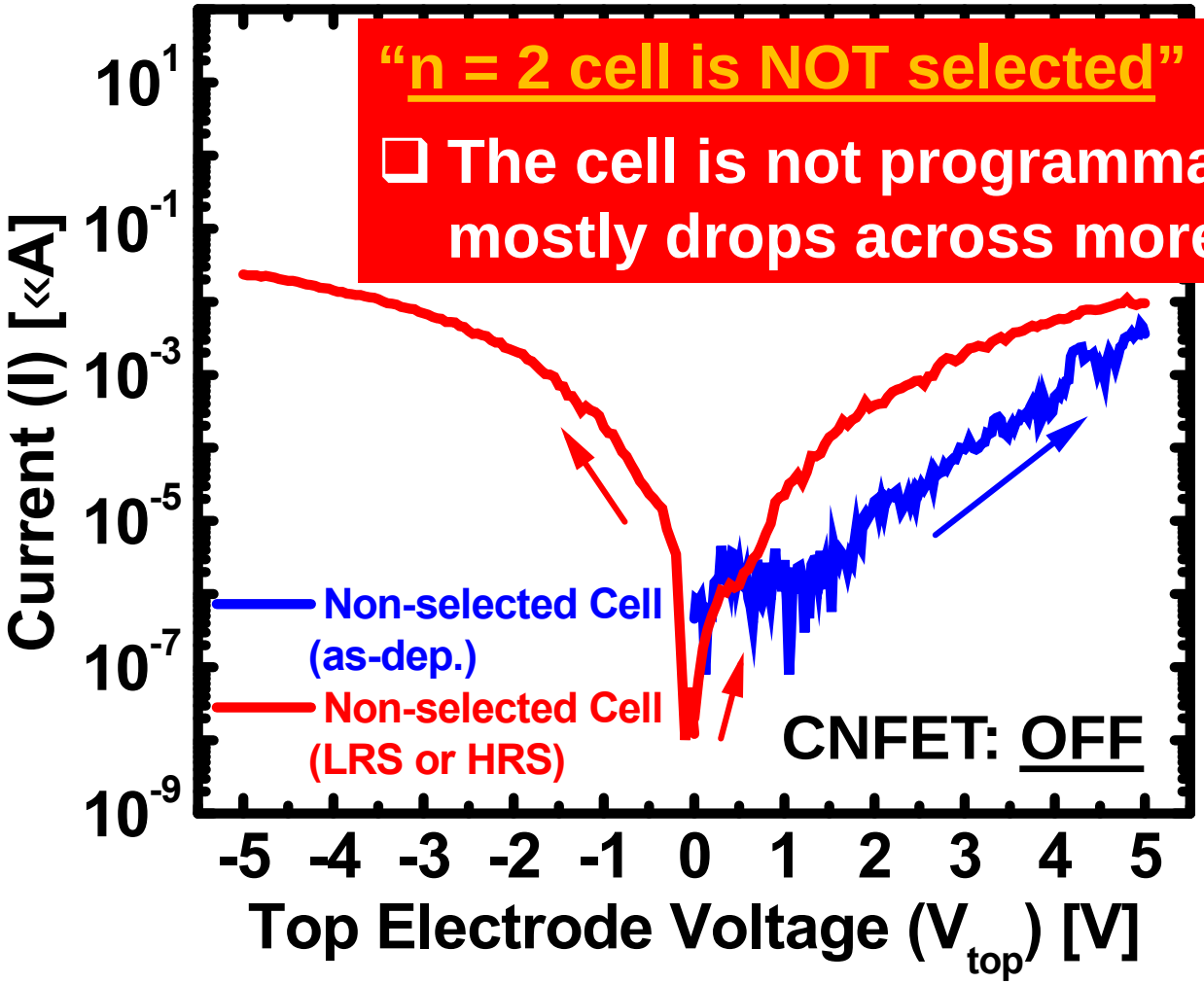
(V_{WL} = + 2V)

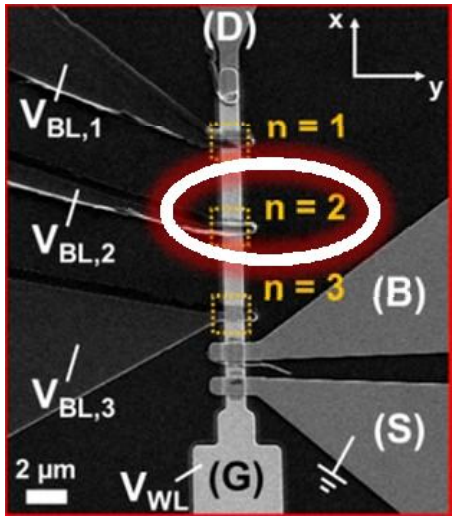
Ahn et al. IEEE TED 2015

Electrical results: 1TnR RRAM – Selective switching (DC)

"n = 2 cell is NOT selected"

- **The cell is not programmable, as the voltage mostly drops across more-resistive CNFET**

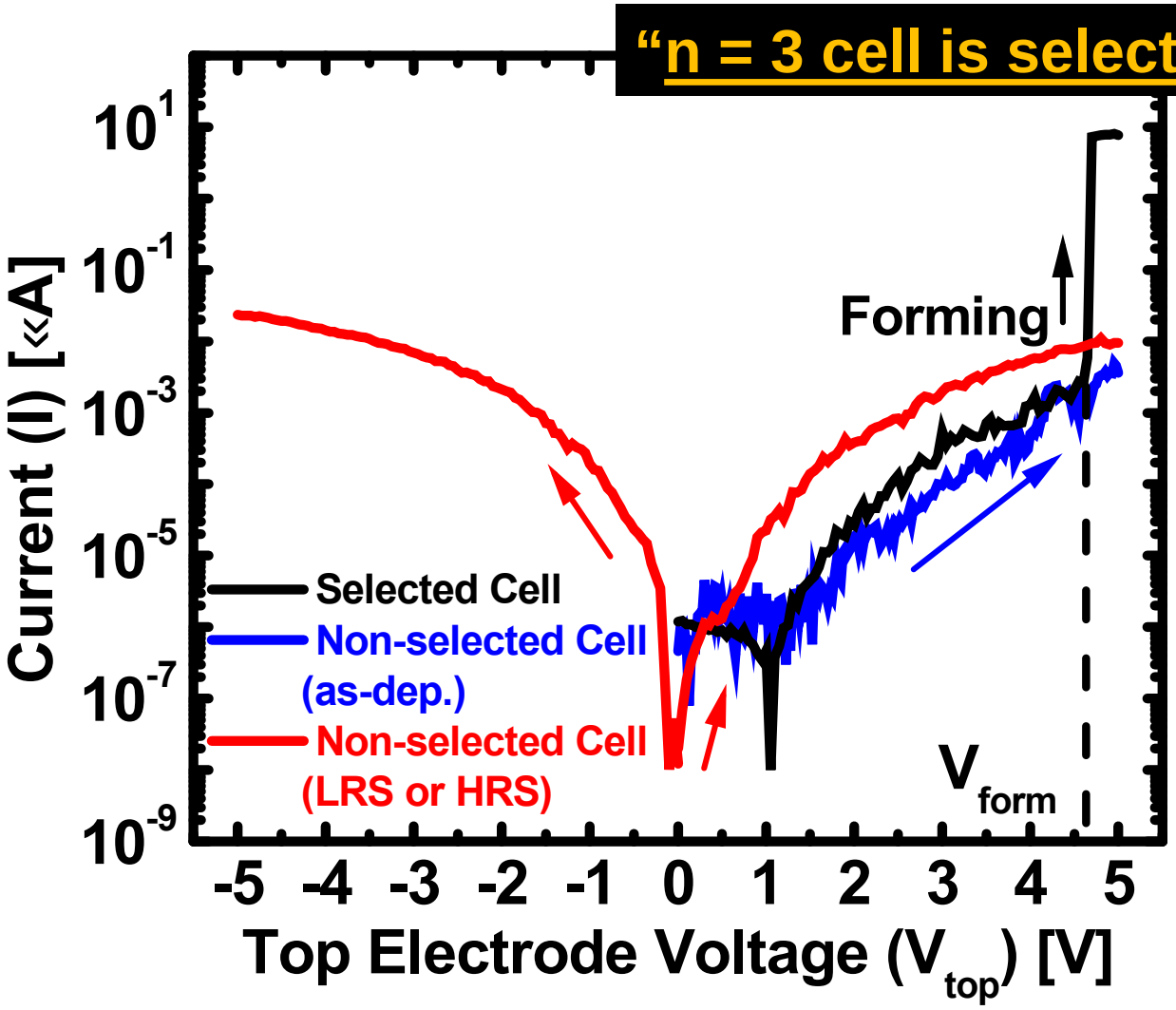


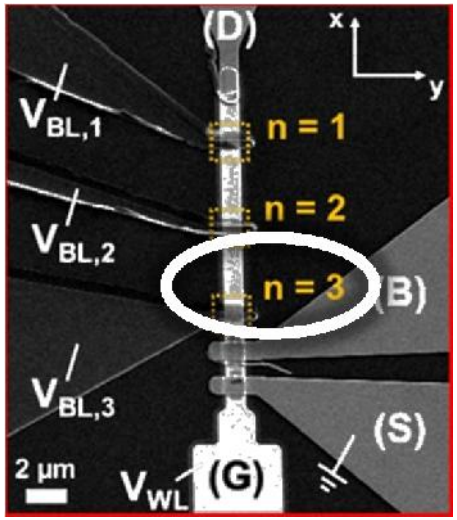


(V_{WL} = + 2V)

Ahn et al. IEEE TED 2015

Electrical results: 1TnR RRAM – Selective switching (DC)



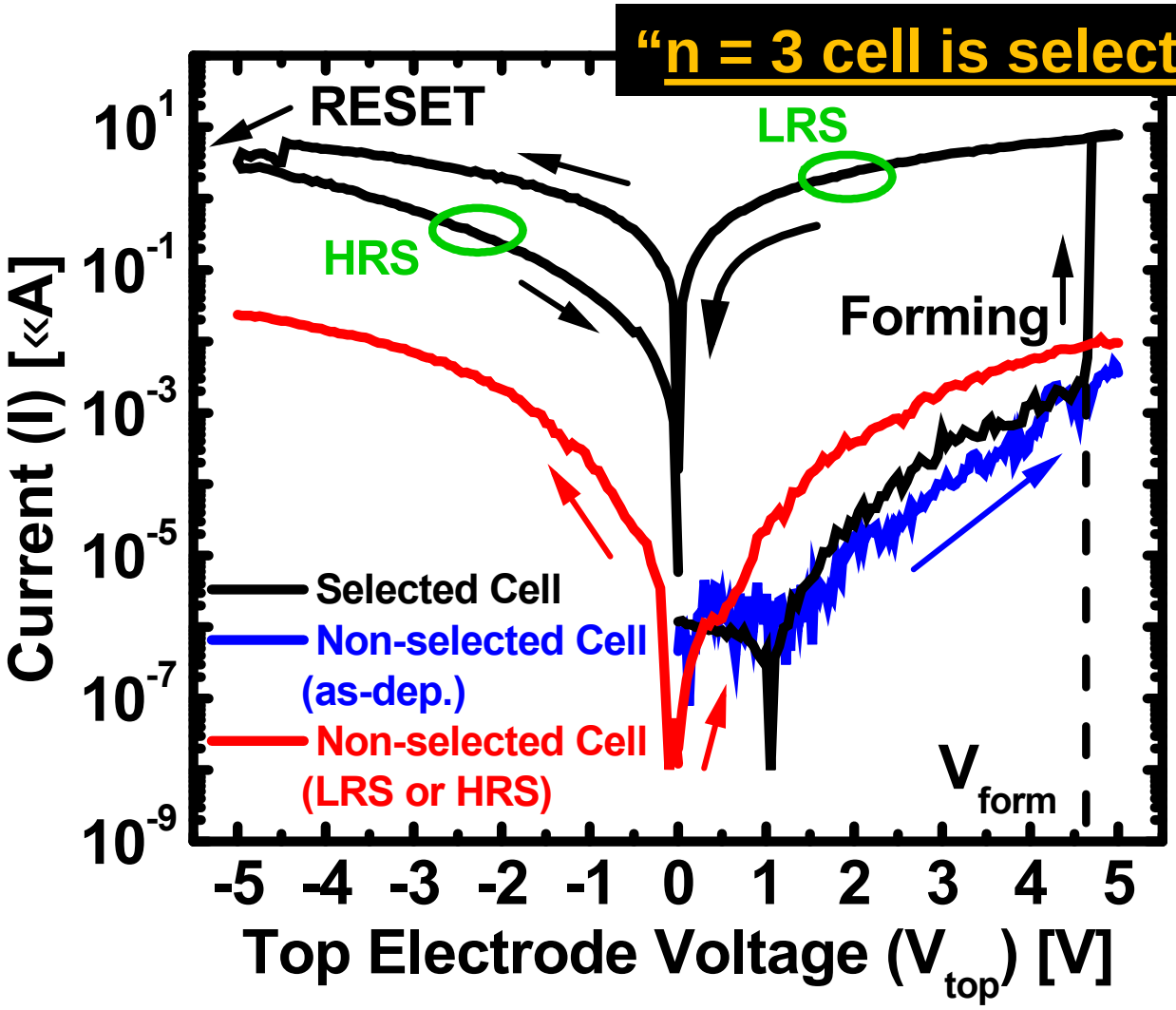


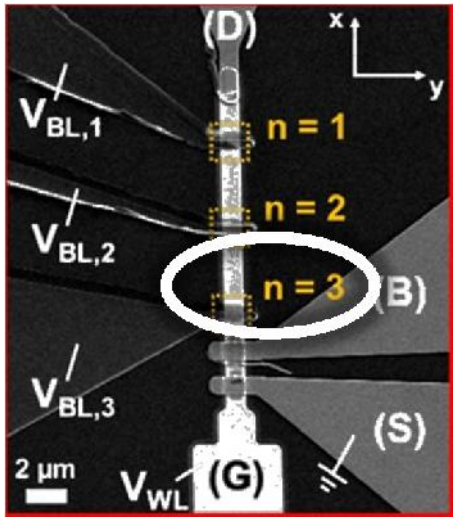
(V_{WL} = - 5V)

"CNFET: ON"

Ahn et al. IEEE TED 2015

Electrical results: 1TnR RRAM – Selective switching (DC)



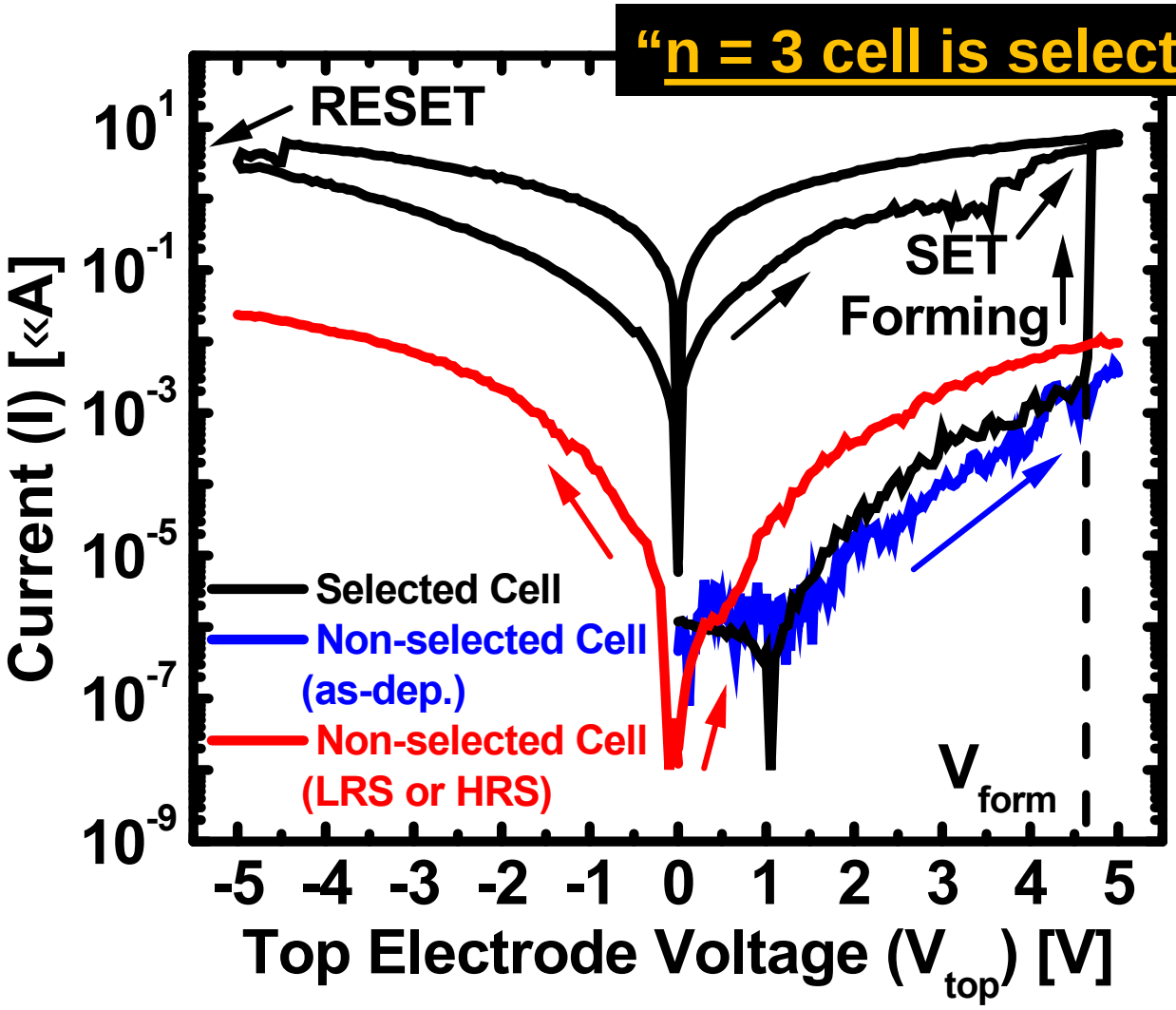


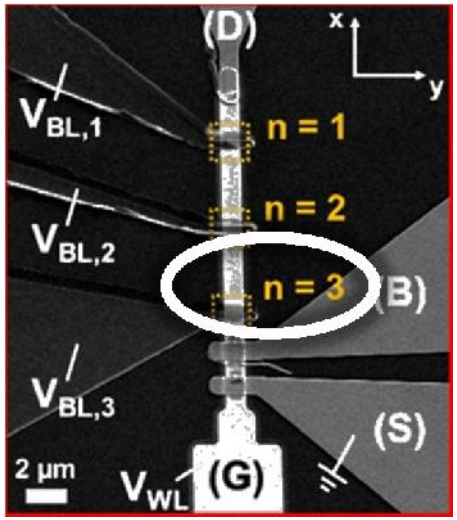
(V_{WL} = - 5V)

"CNFET: ON"

Ahn et al. IEEE TED 2015

Electrical results: 1TnR RRAM – Selective switching (DC)



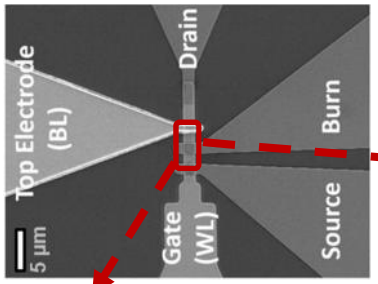


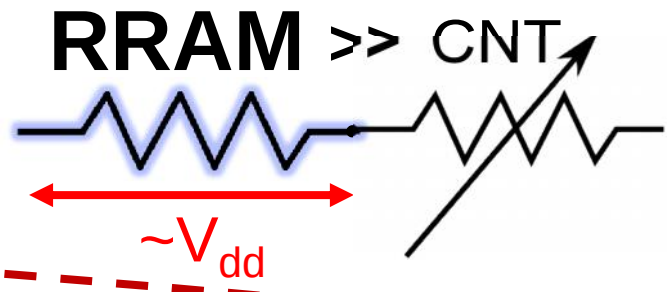
(V_{WL} = - 5V)

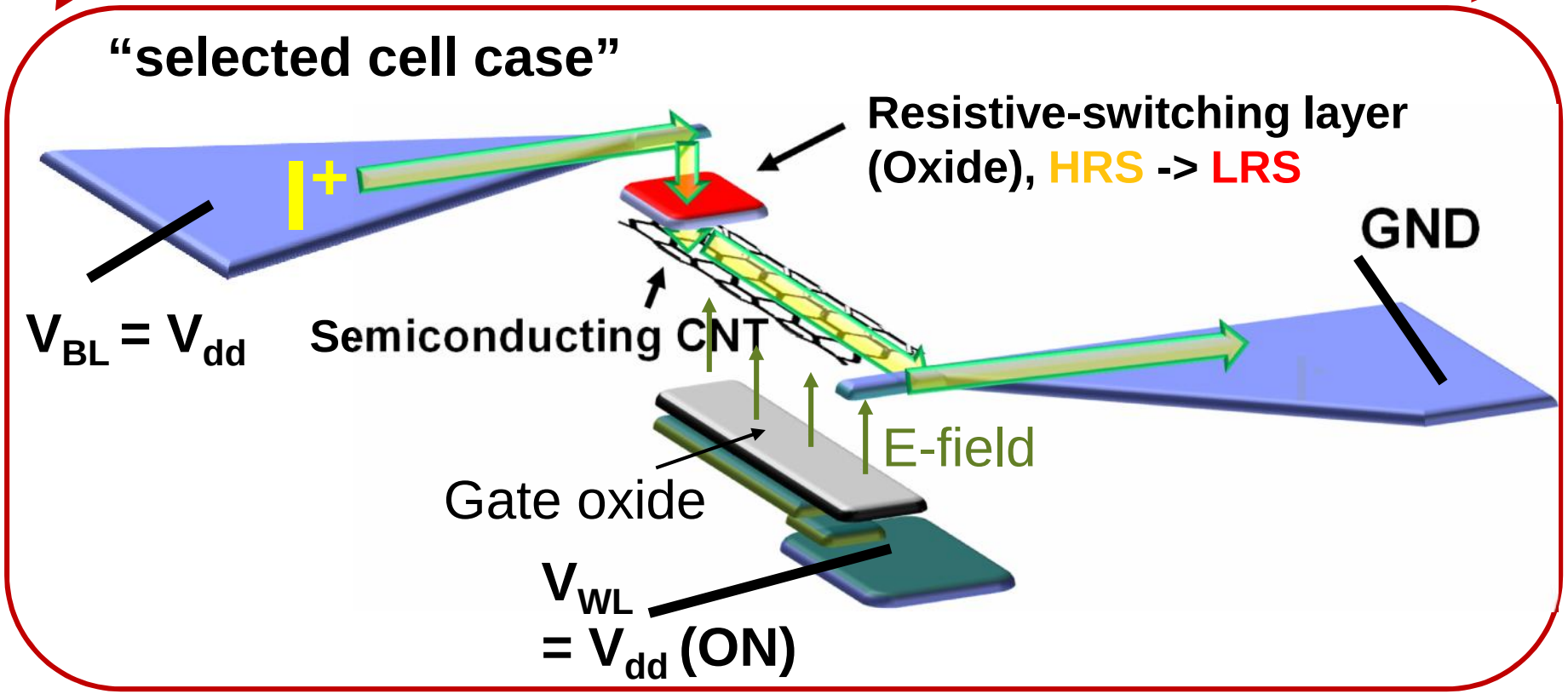
"CNFET: ON"

Ahn et al. IEEE TED 2015

Critical roles of CNTs

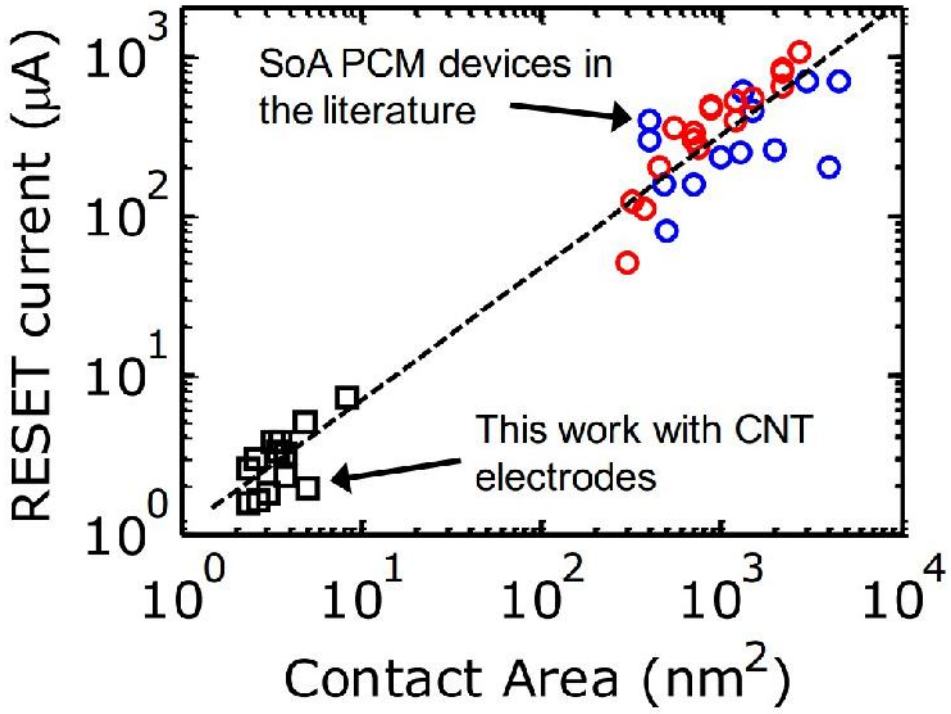






Critical dimension: **< 5 nm²**

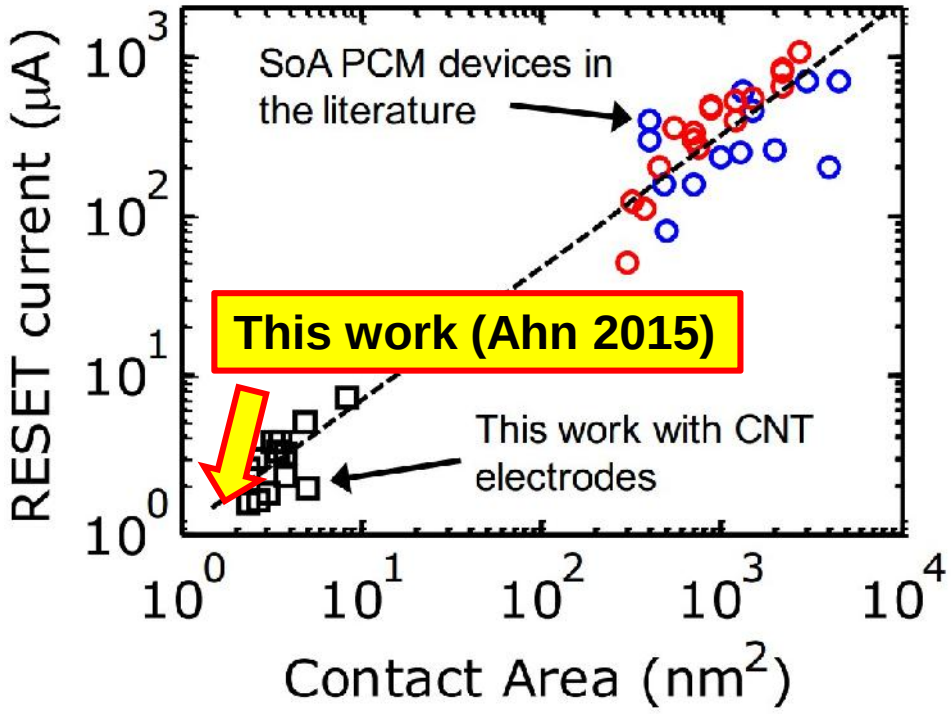
"PCM reset current scales with the effective contact area"



Ref: F. Xiong et al, Nano Letters **13**, 2013

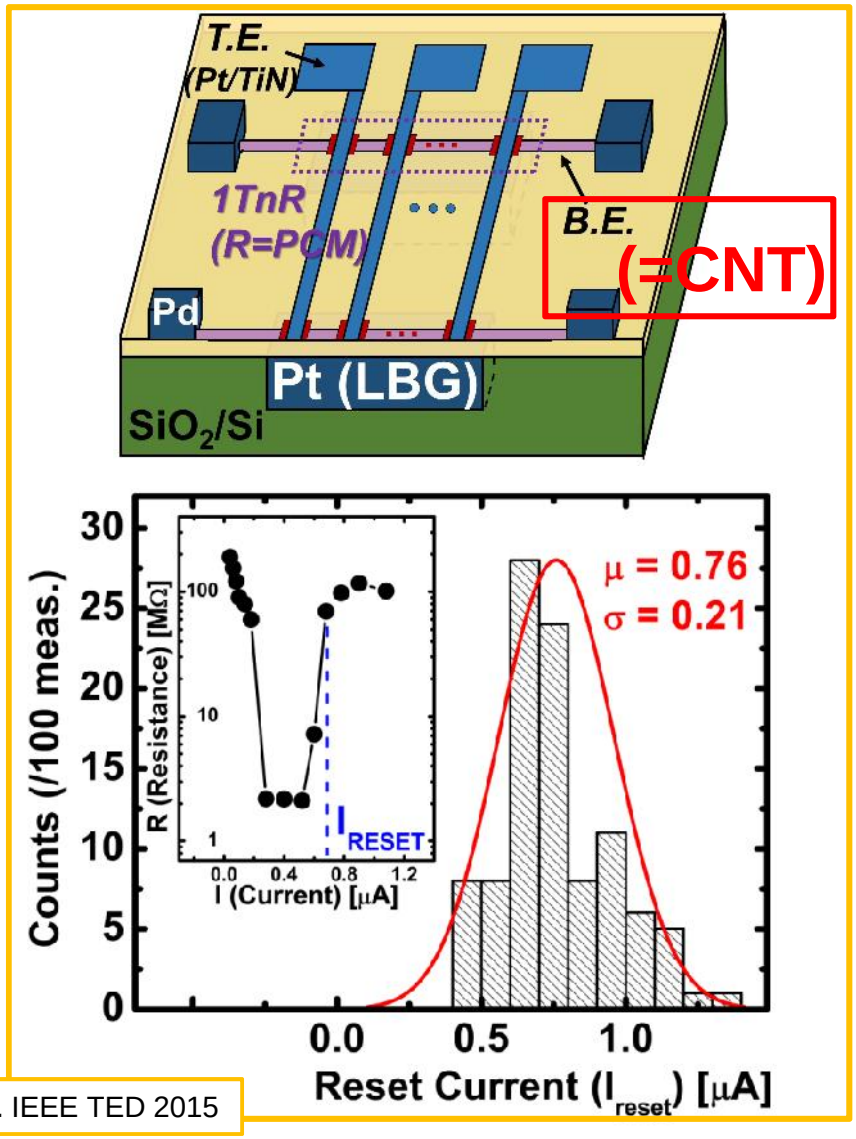
Critical dimension: < 5 nm²

"Sub-1 μA RESET current PCM integrated into the 1TnR array"



Ref: F. Xiong et al, Nano Letters **13**, 2013

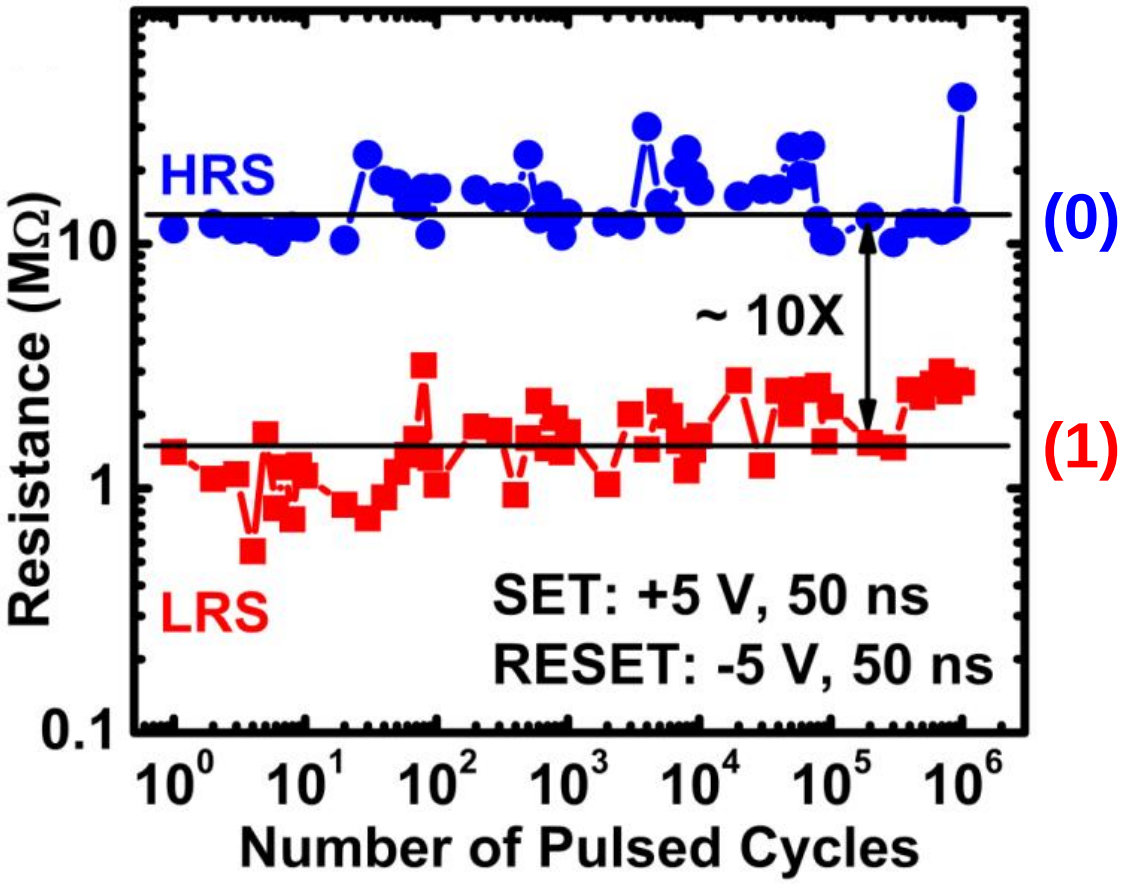
1TnR PCM



Ahn et al. IEEE TED 2015

Electrical results: **1TnR RRAM – Pulsed endurance**

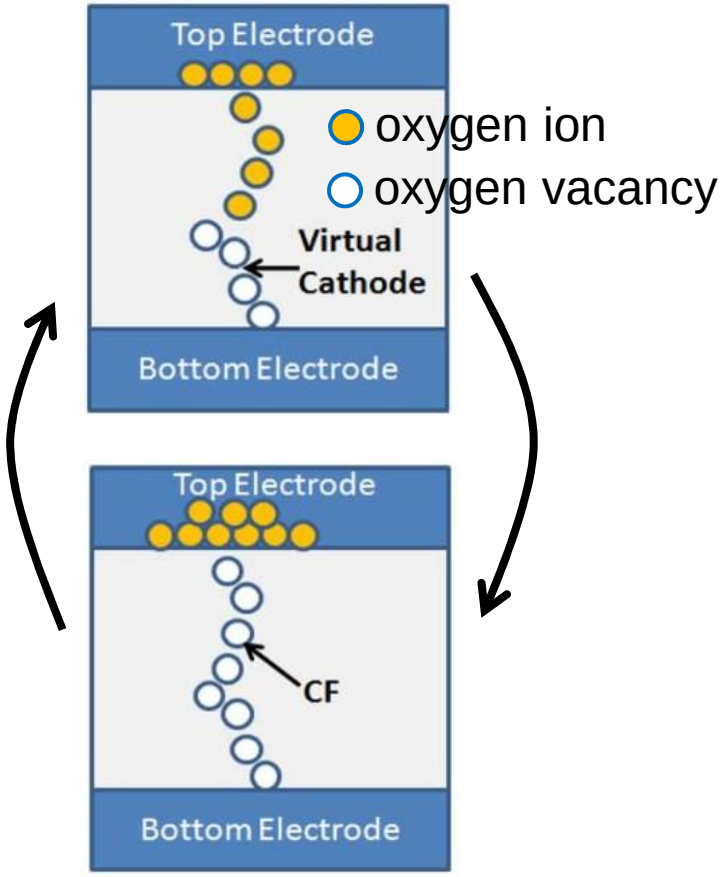
Al_2O_3 RRAM → Low programming power → Size of CF: small



Ahn et al. IEEE TED 2015

(0)

(1)

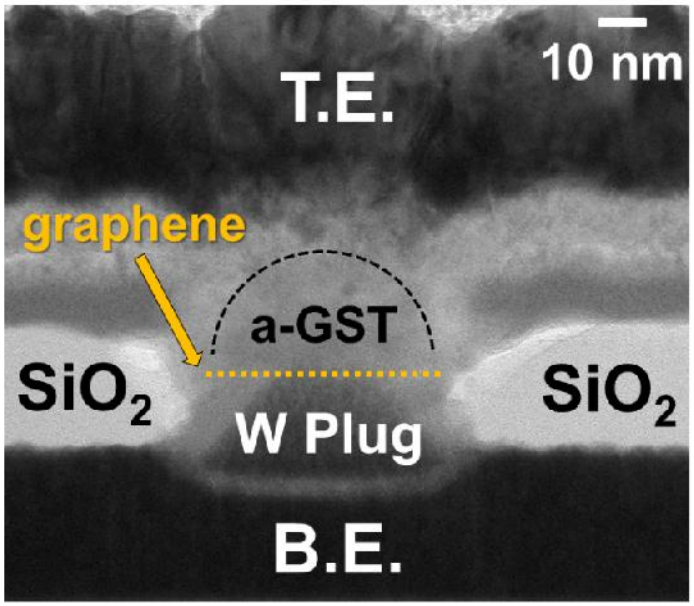


Wong, Proc. IEEE, 2012

"Summary"

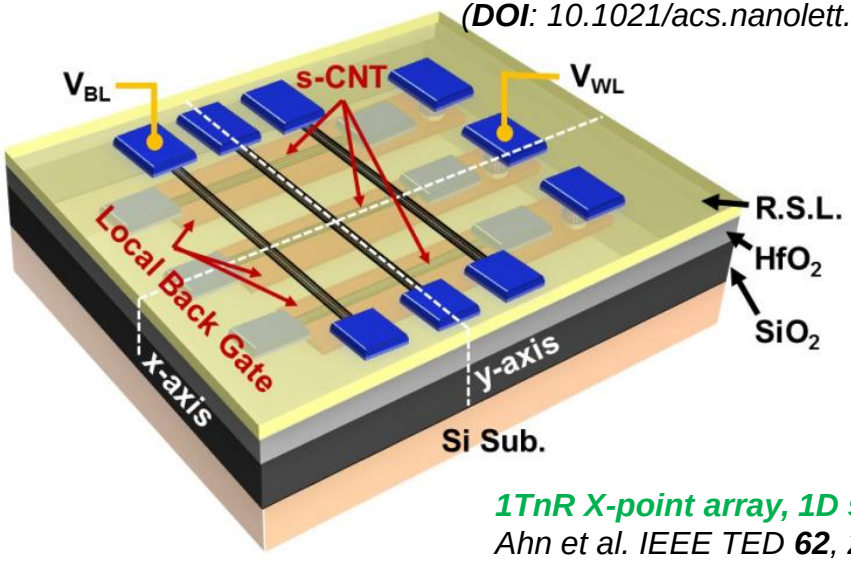
Energy-efficient NVM design

Cell



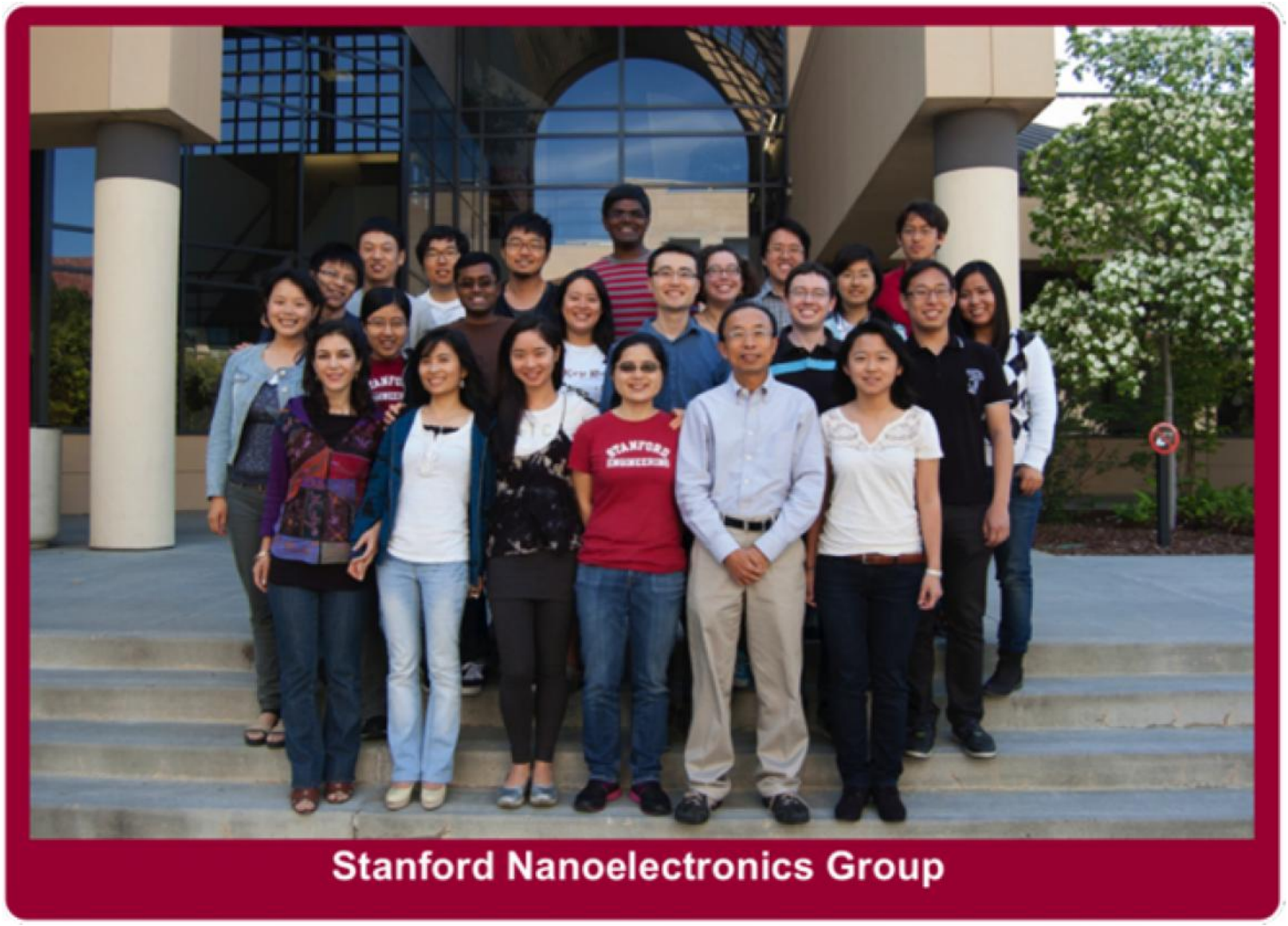
Graphene-PCM, Low-power PCM
*Ahn et al. Nano Letters, **in press**, 2015*
*(**DOI**: 10.1021/acs.nanolett.5b02661)*

Architecture

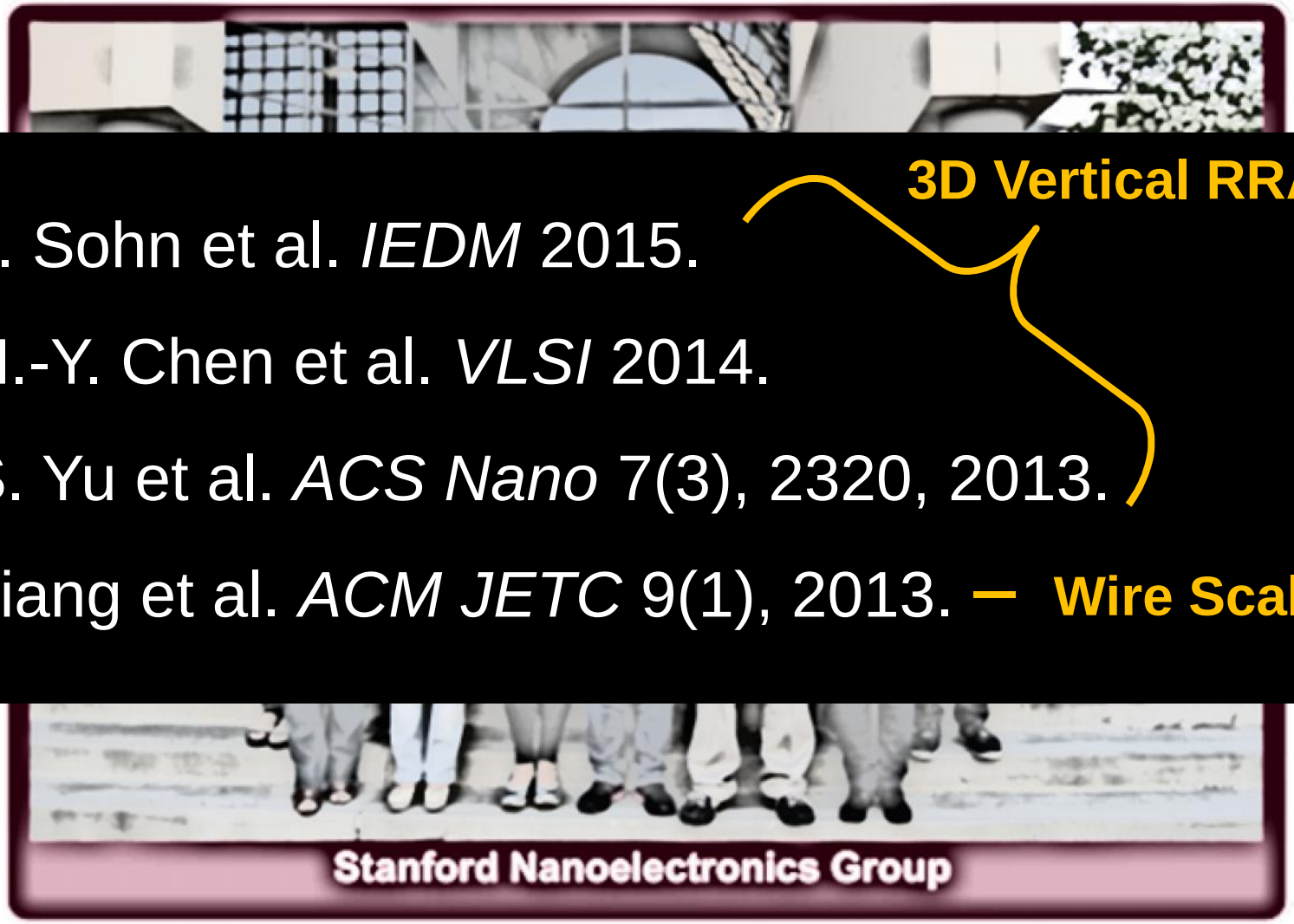


1TnR X-point array, 1D selector
*Ahn et al. IEEE TED **62**, 2197, 2015*

“Acknowledgement”



“Closely-related publications”



1. J. Sohn et al. *IEDM* 2015.
2. H.-Y. Chen et al. *VLSI* 2014.
3. S. Yu et al. *ACS Nano* 7(3), 2320, 2013.
4. Liang et al. *ACM JETC* 9(1), 2013. – **Wire Scaling**

3D Vertical RRAM

"Which memory?"

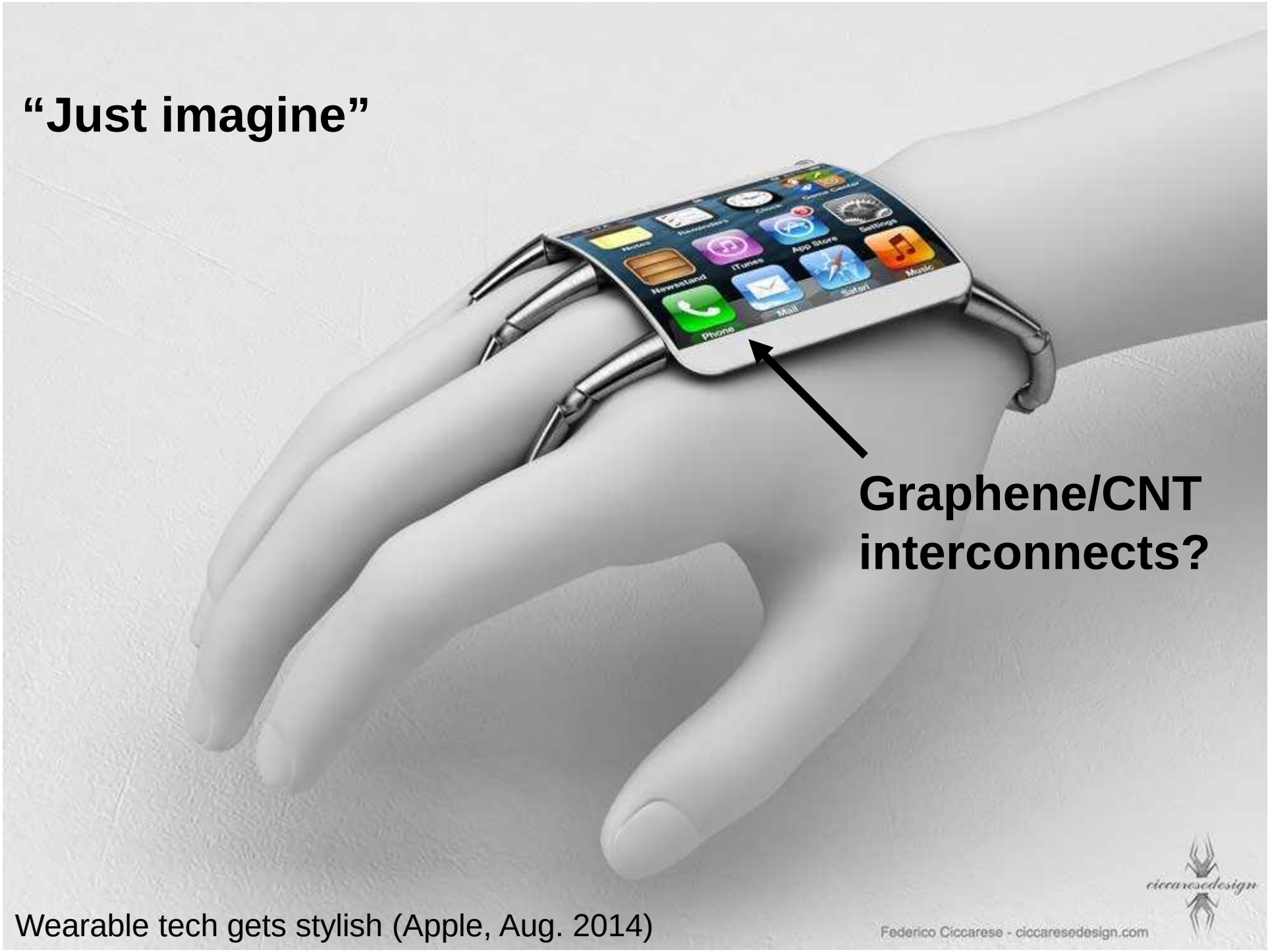
	RRAM	PCM	STT-MRAM	SRAM	DRAM	NAND Flash
Memory type	Emerging			Established		
Non-volatility	Yes	Yes	Yes	No	No	Yes
Cell size (in F^2)	4 (< 4 if 3D)	4 (< 4 if 3D)	6-12	> 120	> 6	4 (< 4 if 3D)
MLC	4-bit	> 2-bit	1-bit	1-bit	1-bit	3-bit
Read latency	< 50 ns	< 50 ns	< 10 ns	< 10 ns	50-60 ns	> 10 µs
Write time	< 20 ns	> 50 ns	< 10 ns	< 1 ns	~10 ns	> 100 µs
Endurance (cycles)	~ 10^{12}	> 10^{9}	~ 10^{15}	> 10^{16}	> 10^{16}	10^{4}~10^{5}
Energy (per bit)	< 1 pJ	< 20 pJ	< 0.1 pJ	~ 0.5 fJ	~ 5 fJ	> 1 fJ

Ethan Ahn, Stanford University Ph.D. Dissertation
(available online at http://purl.stanford.edu/fp960bw6447)

“Just imagine”

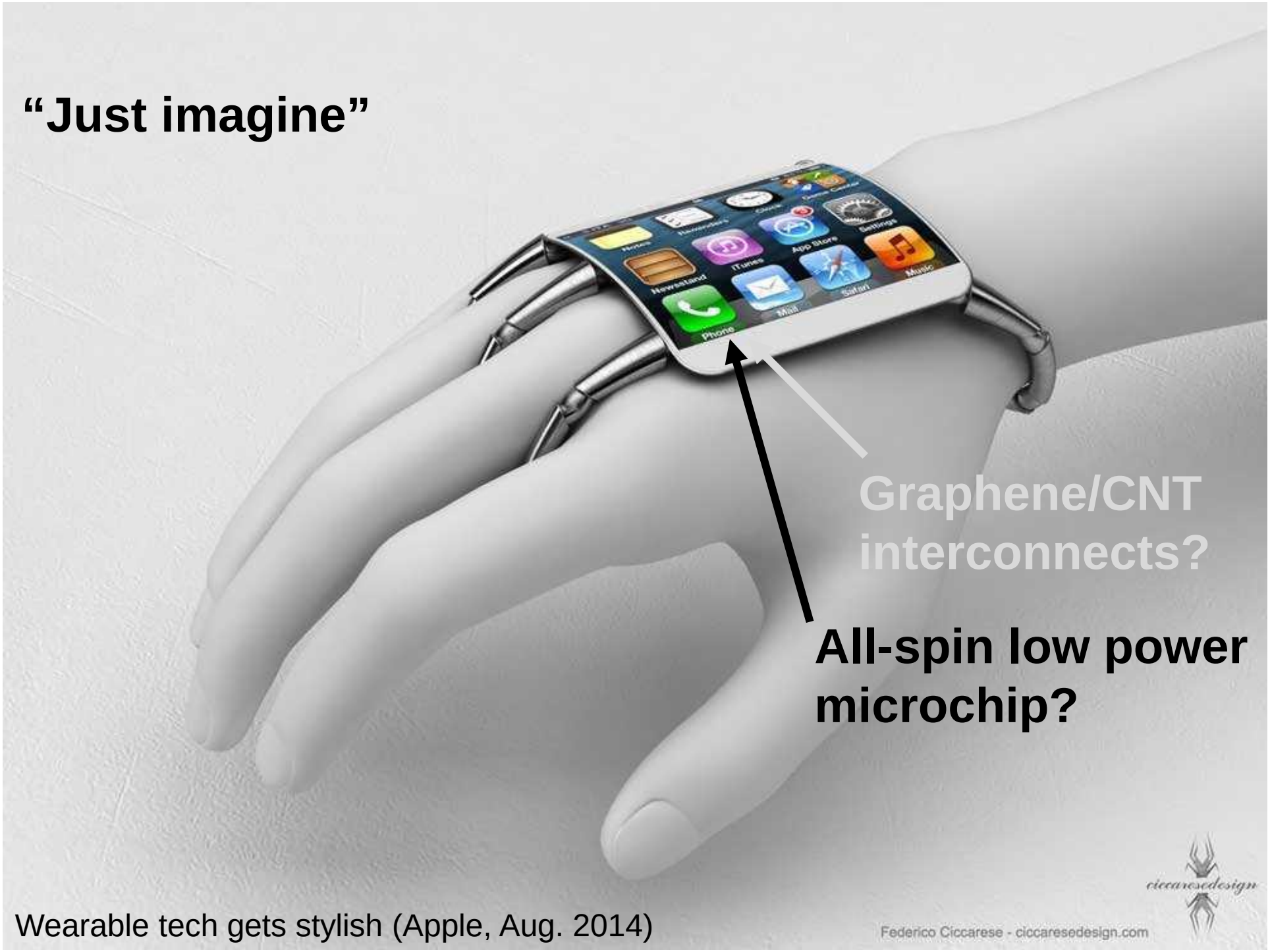
Wearable tech gets stylish (Apple, Aug. 2014)

“Just imagine”



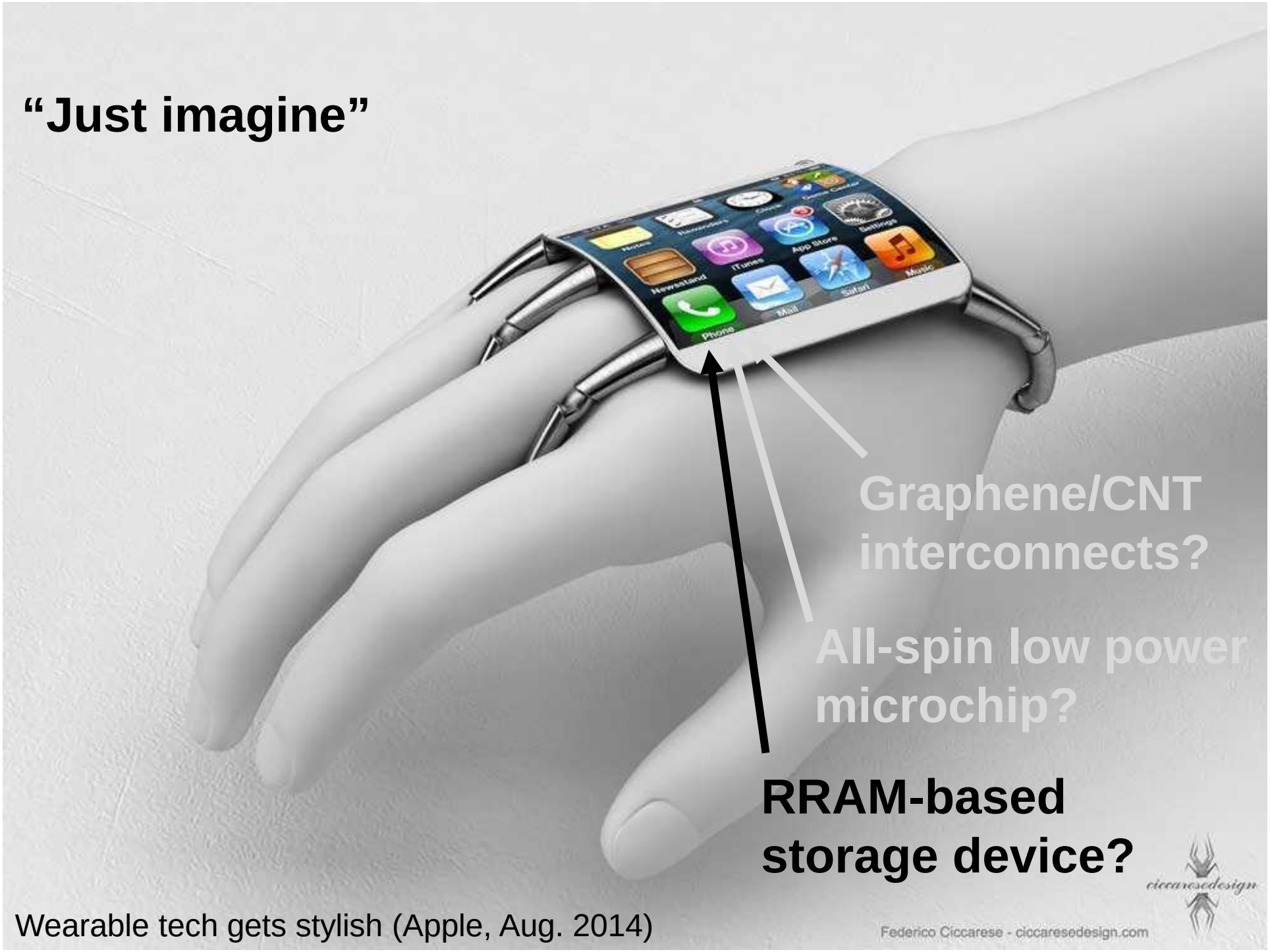
Wearable tech gets stylish (Apple, Aug. 2014)

“Just imagine”



Wearable tech gets stylish (Apple, Aug. 2014)

“Just imagine”



Wearable tech gets stylish (Apple, Aug. 2014)