



Non-Volatile Memory Technology Overview

Ugo Russo, Andrea Redaelli, Roberto Bez

► To cite this version:

Ugo Russo, Andrea Redaelli, Roberto Bez. Non-Volatile Memory Technology Overview. Norm Jouppi and Yuan Xie and Eren Kursun. WTAI: Workshop on Technology Architecture Interaction, Jun 2010, Saint-Malo, France. inria-00514448

HAL Id: inria-00514448

<https://inria.hal.science/inria-00514448>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Non-Volatile Memory Technology Overview

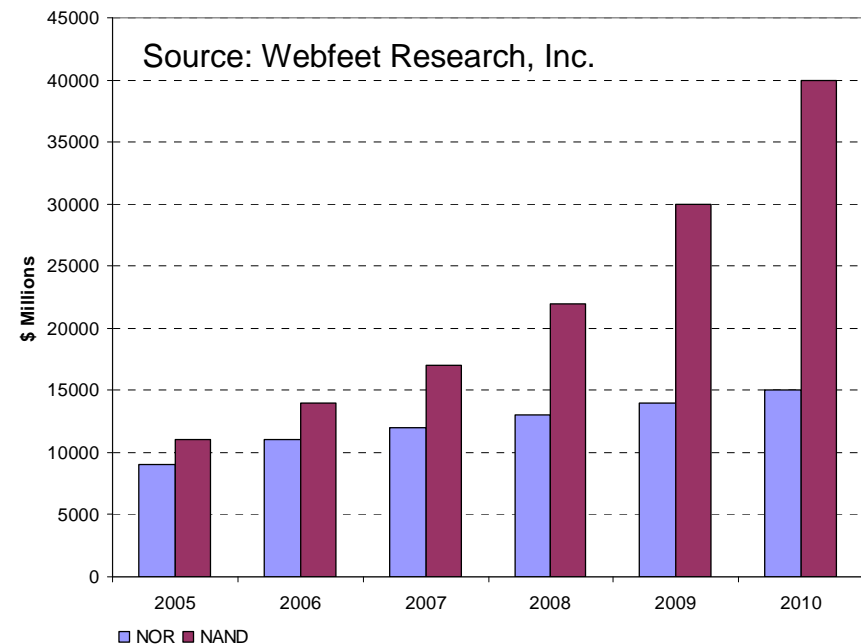
Ugo Russo, Andrea Redaelli and Roberto Bez
R&D Technology Development
Numonyx

Outline

- Non-Volatile Memory (NVM): from applications to technology scaling
- NVM technologies: evolutions and revolutions
 - Flash memory and its evolution
 - Ferroelectric memory (FeRAM)
 - Magnetic memory (MRAM)
 - Phase Change Memory (PCM)
 - Resistive-switching memory (RRAM)
- Summary

Non-volatile Memory: demand and market

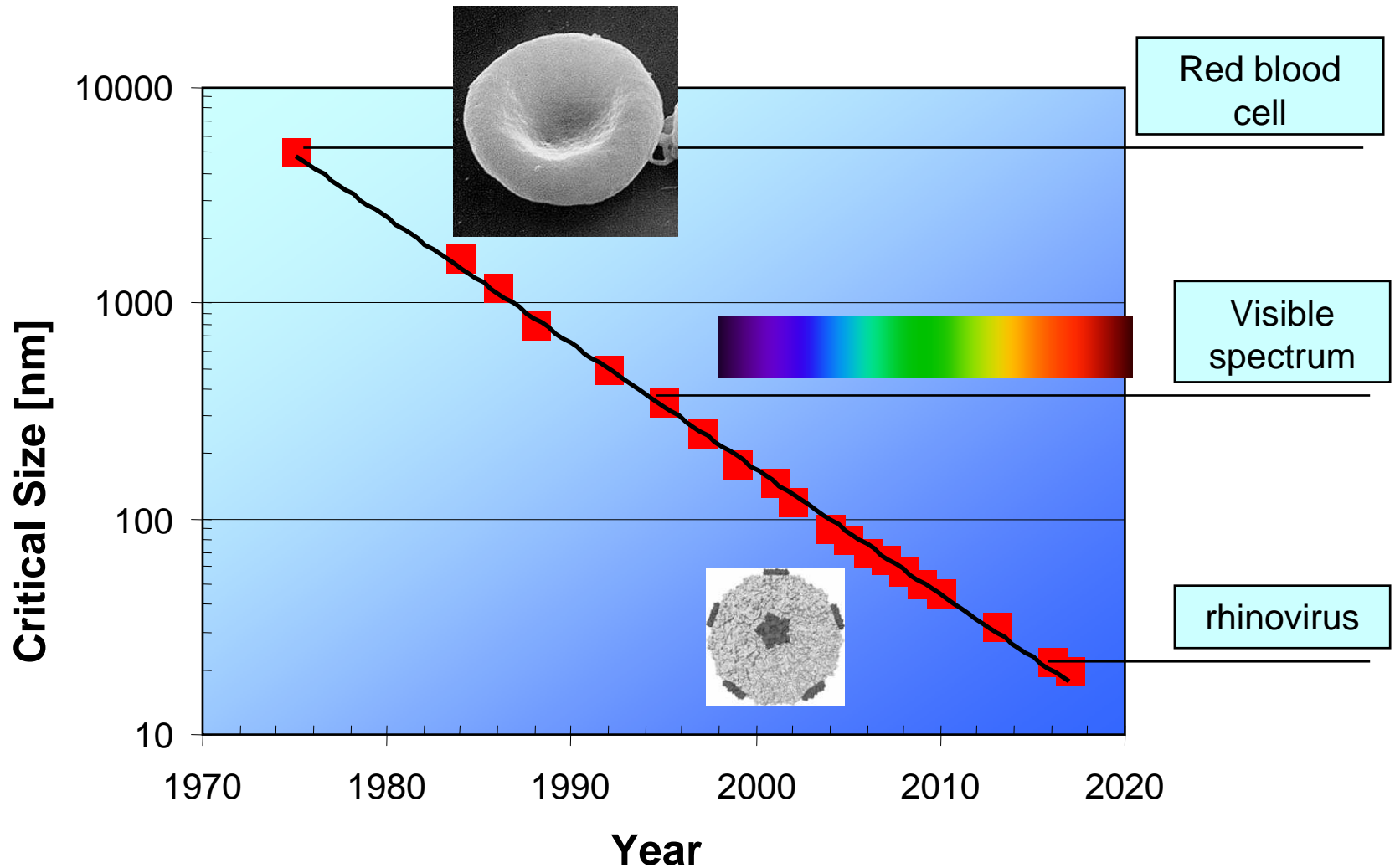
Enabling technology for consumer portable goods



Coming up: Solid State Discs

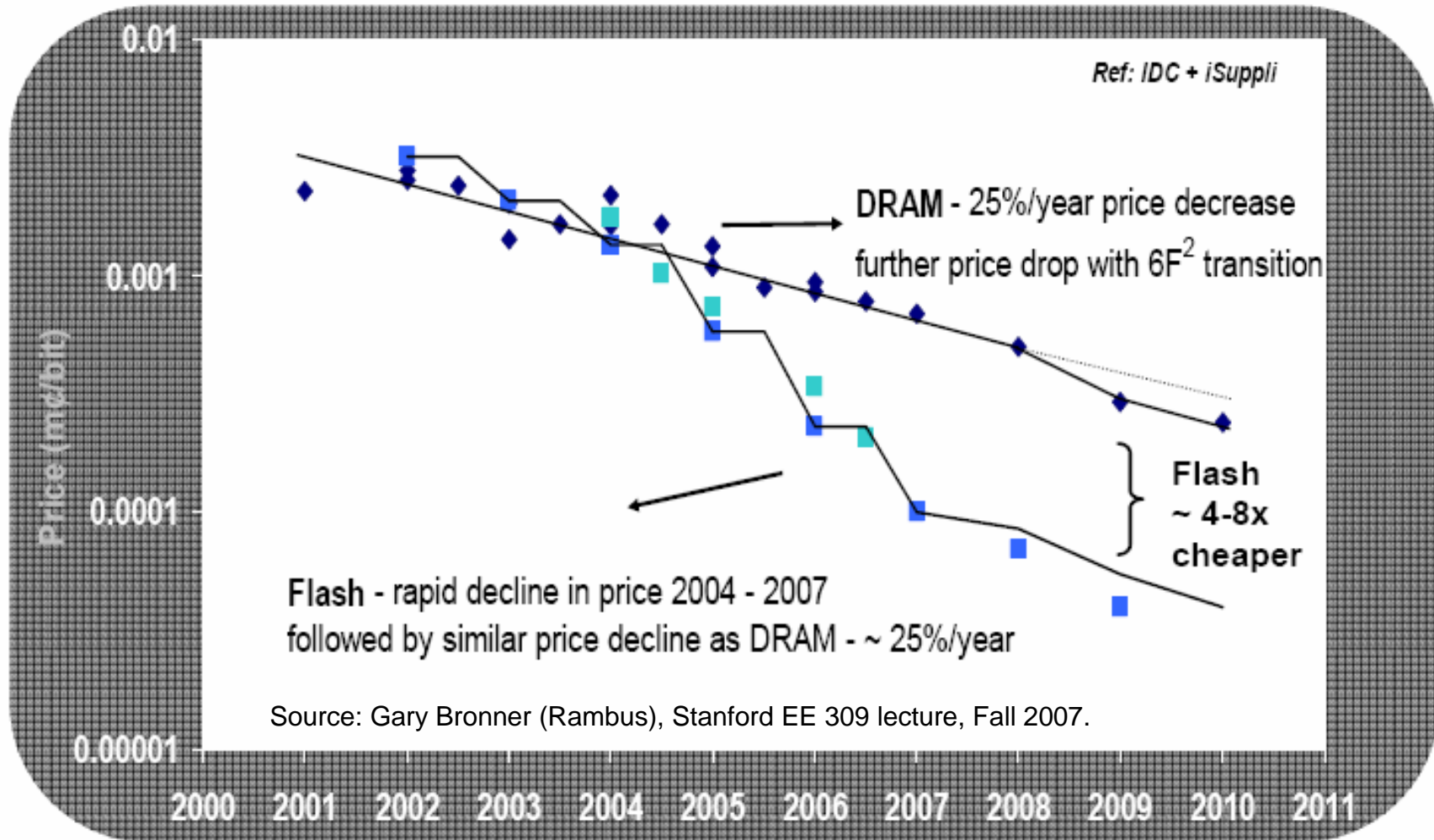
- Promising for low power, high performance, fast boot up
- Still need system-level optimization and to reduce costs

The driving force for multi-Gb NVM: A combination of size reduction...

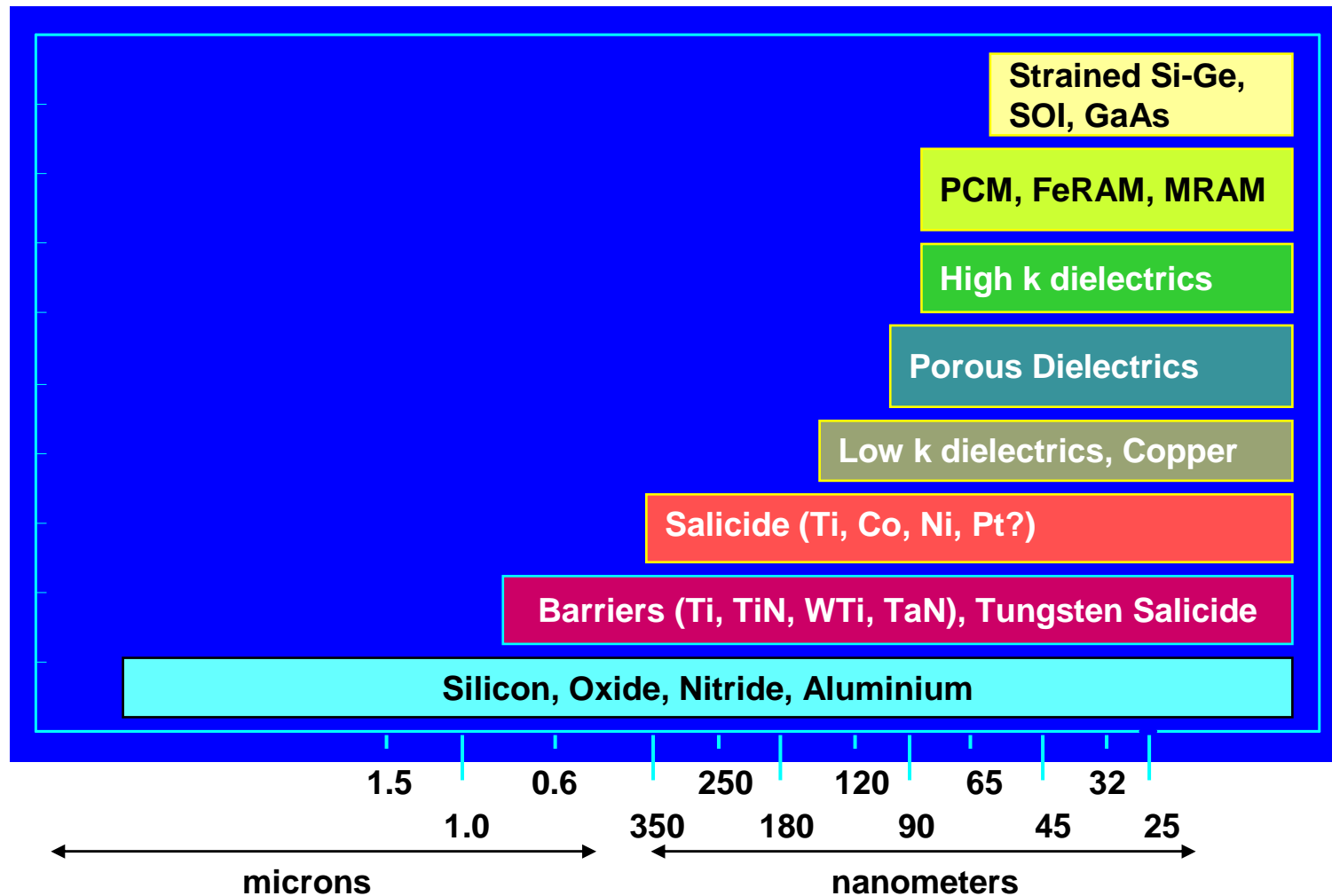


... and price reduction

DRAM and Flash Price Outlook



A broadening palette of materials



Floating Gate NVM History

1967 → First Floating Gate Structure

1977 → EPROM

1980 → EEPROM

1985 → 1T EEPROM (Flash)

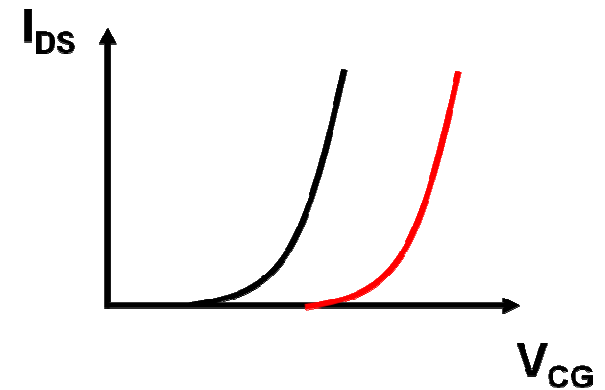
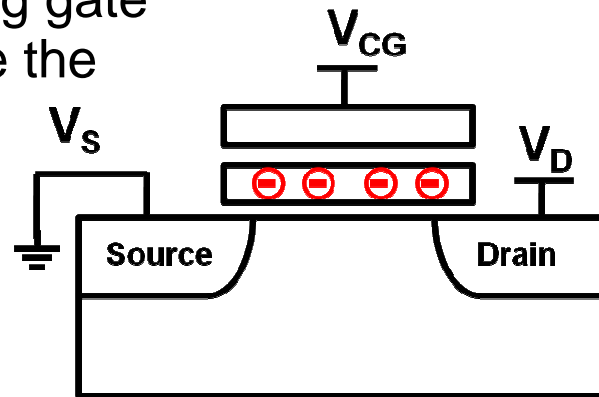
1988 → NOR Flash

1989 → NAND Flash

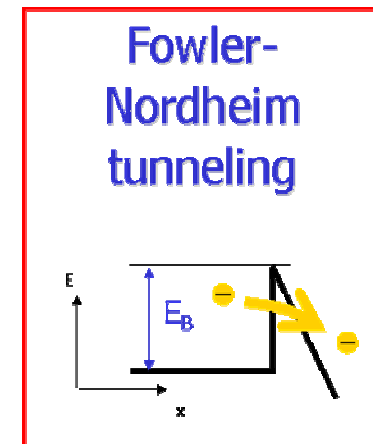
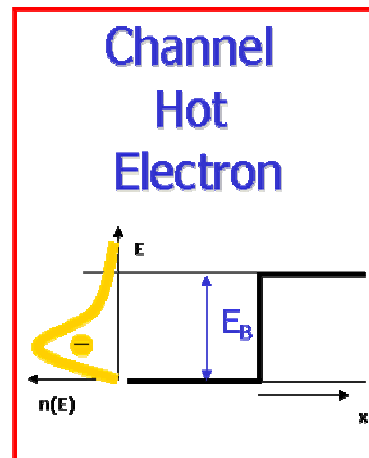
1995 → MLC Flash

Flash memory working principle

- A chargeable floating gate is introduced to tune the transistor threshold voltage



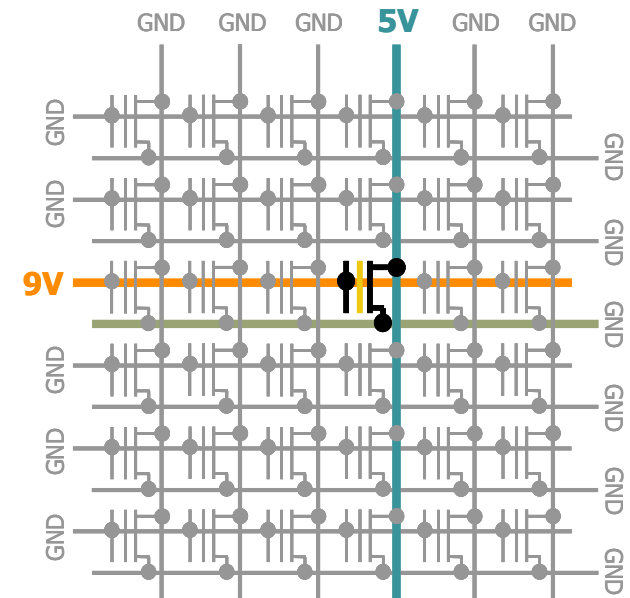
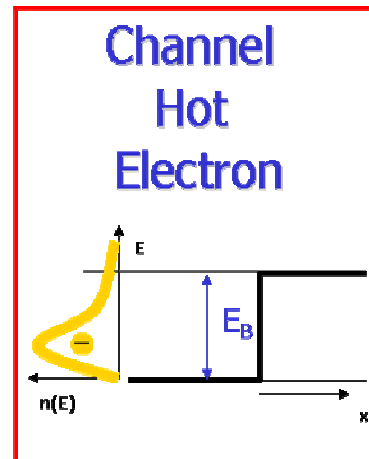
- Different physical mechanisms are used to charge/uncharge the floating gate, depending on memory architecture



NOR flash

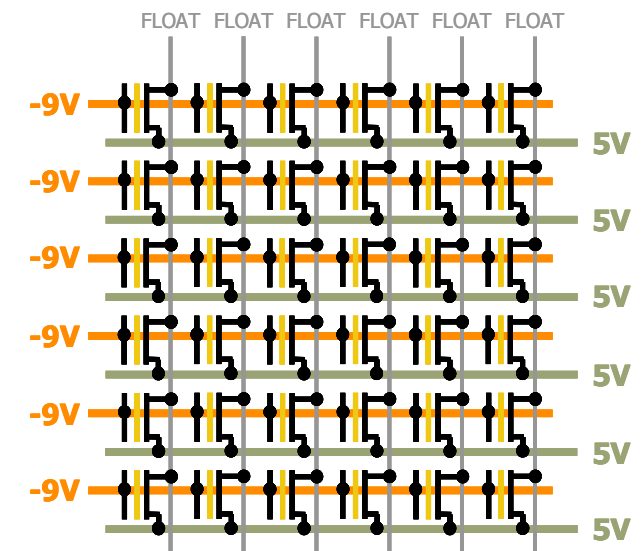
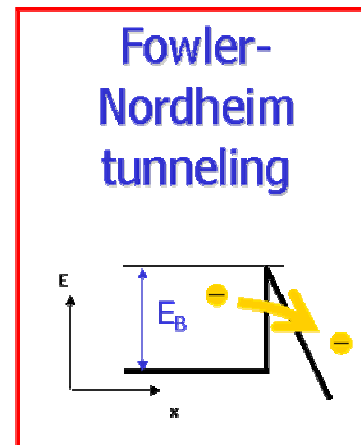
Programming:

- Single bit
- Random access
- High programming current



Erase:

- Block erase
- Low programming current



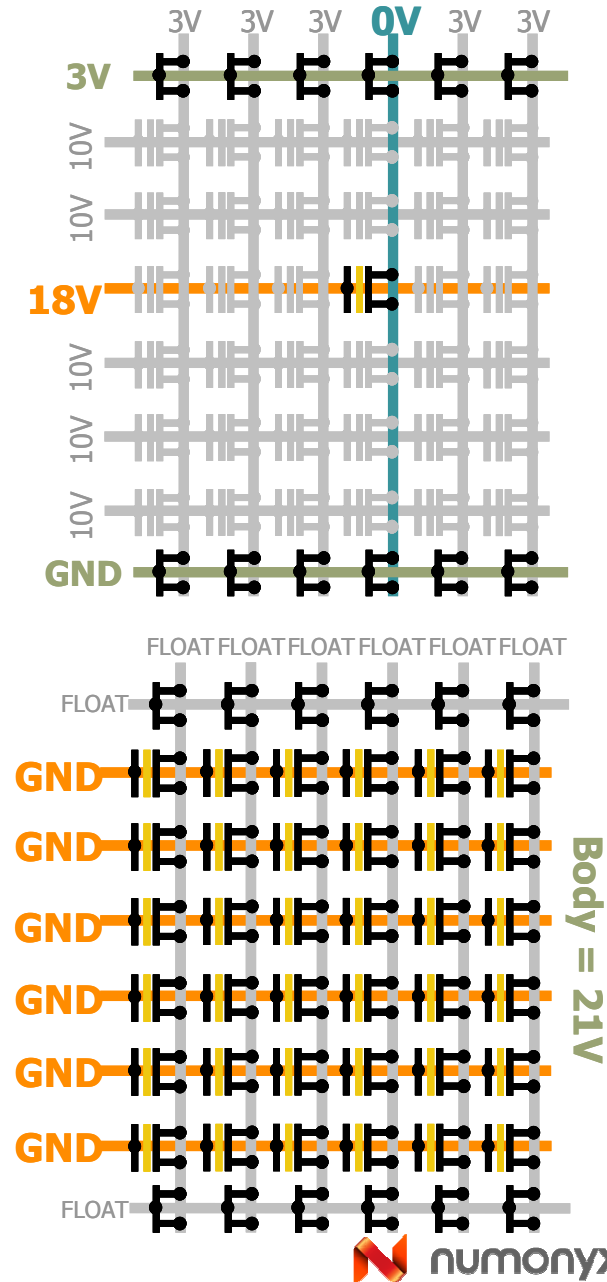
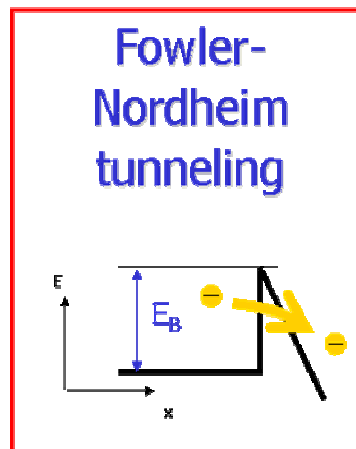
NAND flash

Programming:

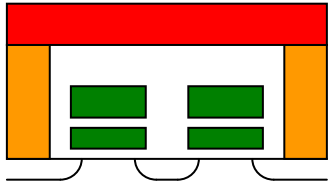
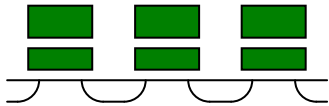
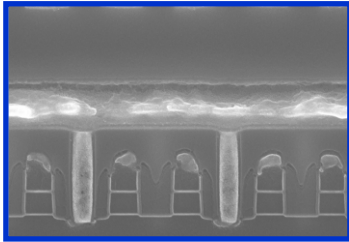
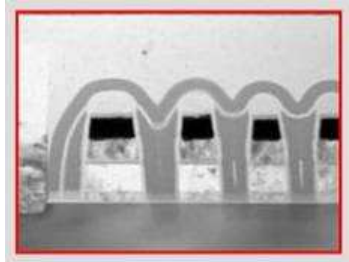
- Serial access
- Low programming current

Erase:

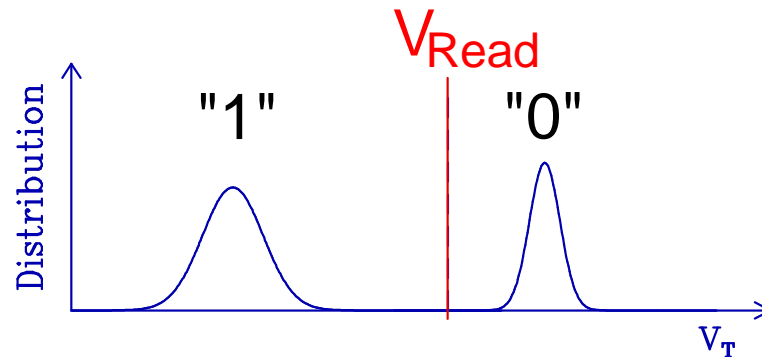
- Block erase
- Low programming current



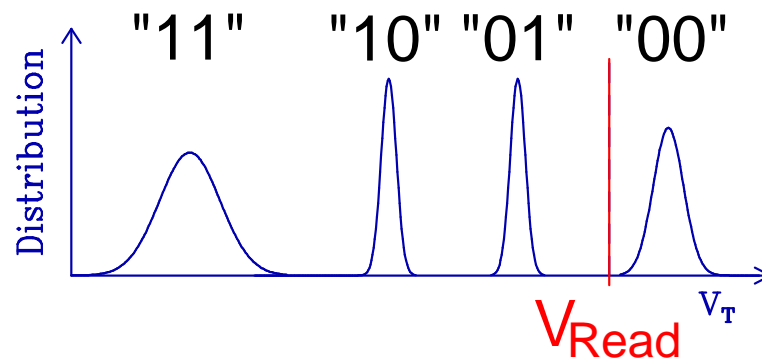
NOR and NAND Stacked Gate Flash

	NOR	NAND
		
Cell size (F ²)	10	5
Read access	Random (fast ~50ns)	Serial
Progr. Mechanism/ Throughput	CHE / 0.5 MB/s	FN / 7-10MB/s
SEM Cross-section (BL direction)		

The Multilevel Cell Concept



□ 1 bit/cell => 2 levels



□ 2 bit/cell => 4 levels

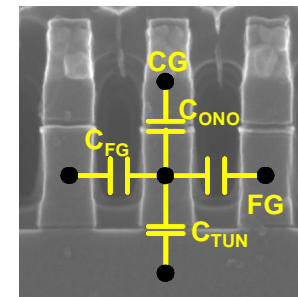
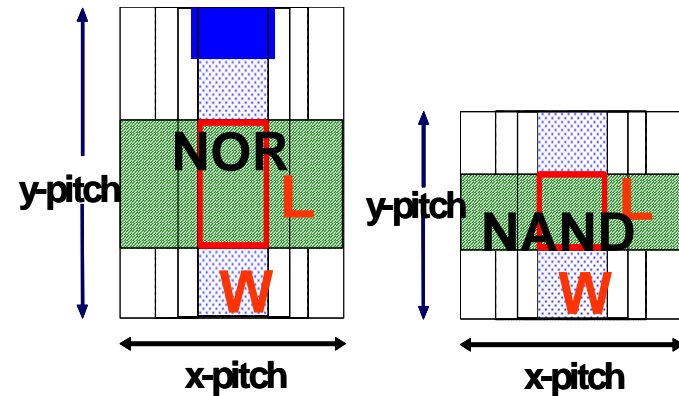
Multilevel storage is widely adopted in high density memories.

Trade-offs:

- + double density with the same technology
- Slower reading and programming
- more sensitive to parasitics

Flash Cell Scaling Challenges

- ❑ Cell basic structure unchanged through the different generations
- ❑ Cell area scaling through:
 1. Active device scaling (W/L)
 2. Passive elements scaling
- ❑ Main scaling issues:
 - Tunnel oxide thickness
 - Interpoly dielectric thickness
 - Cell gate length
 - Contact dimension/ isolation spacing
 - Cell proximity (FG-FG parasitic coupling cross talk)
 - Statistical effects

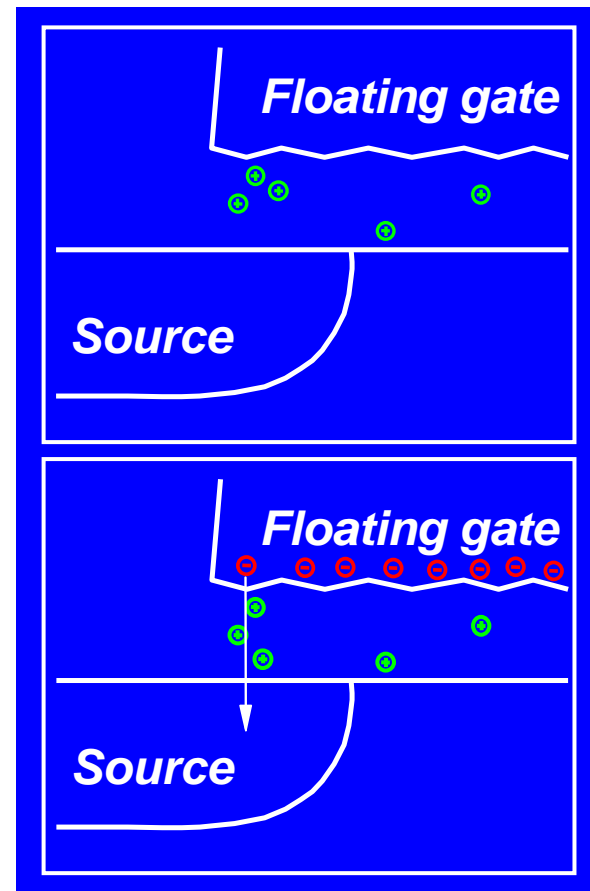


Statistical effects

The low number of electrons stored in the floating gate makes it sensitive also to the effect of single charges trapped in the tunnel dielectric.

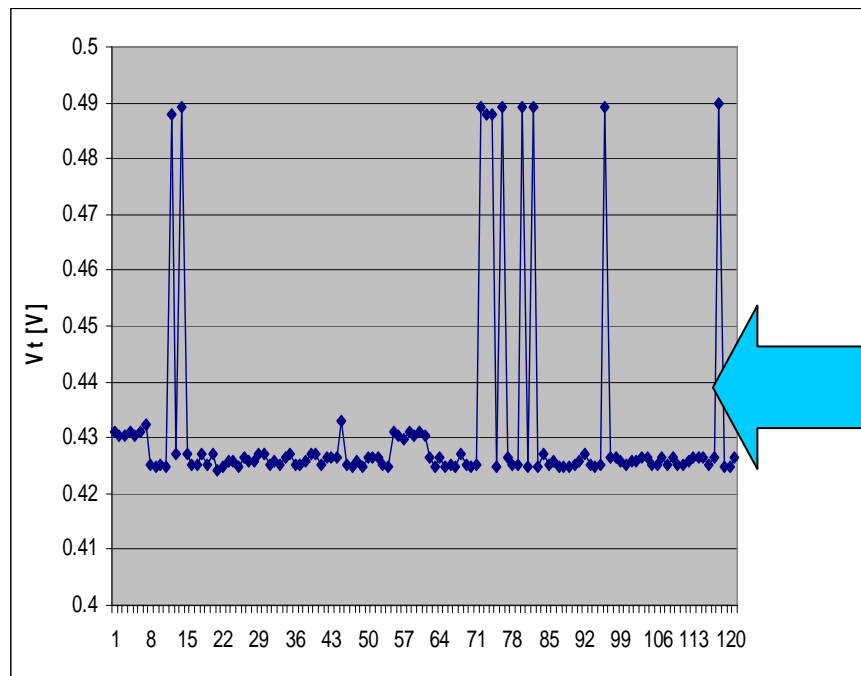
Several effects:

- ❑ Erratic bits
- ❑ Moving bits (SILC)
- ❑ Random Telegraph Signal
- ❑ Fast erasing bits

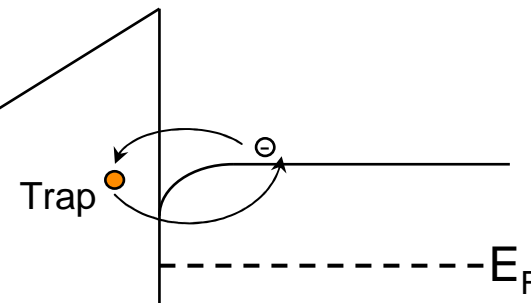


Random Telegraph Signal: when one electron makes the difference.

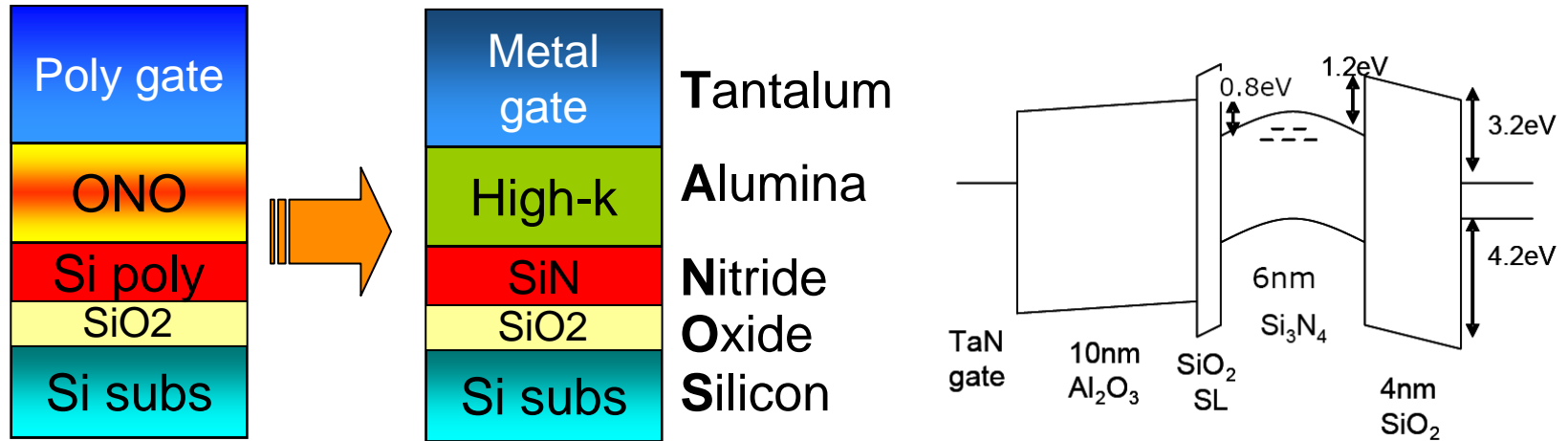
Flash Cell V_t instability



- Read/verify Cell threshold instability due to Random Telegraph Signal (RTS)



Evolution: TANOS NAND Flash



❑ From continuous floating gate to charge trapping layer

❑ Based on well known SONOS concept, but

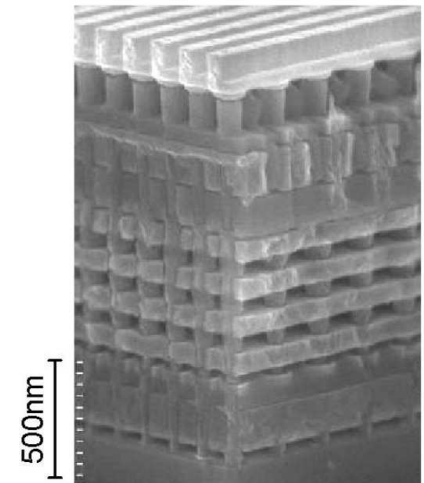
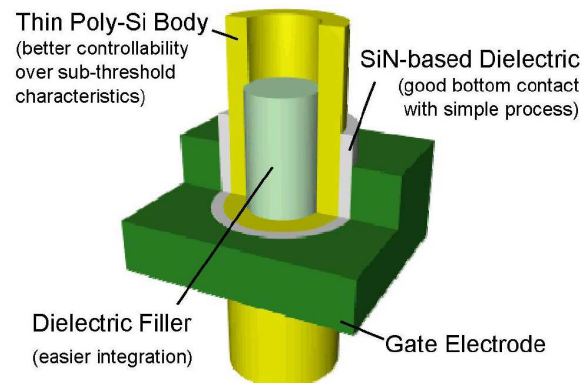
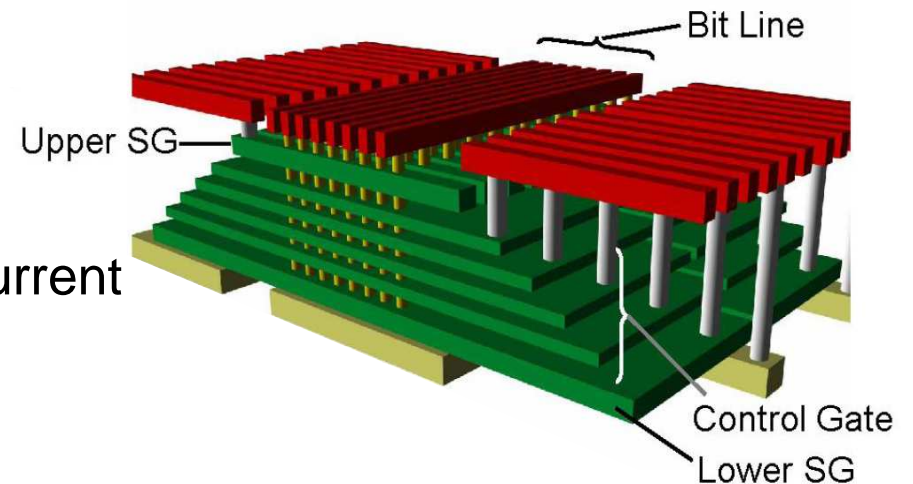
- o thicker tunnel oxide for retention
- o high-k (Al₂O₃) upper dielectric for capacitive coupling
- o metal (Tantalum) gate to prevent erase saturation
- o easy scalability, no capacitive coupling
- o requires introduction of new materials

3D architecture

- Bit-Cost Scalable (BiCS) flash memory
 - Increased memory density
 - Better control of channel charge/current
 - ➔ programmed V_{th} , subthreshold current
 - Integration complexity
 - Early development stage

Y. Fukuzumi et al., Toshiba Corp., IEEE IEDM, 2007

Y. Komori et al., Toshiba Corp., IEEE IEDM, 2008



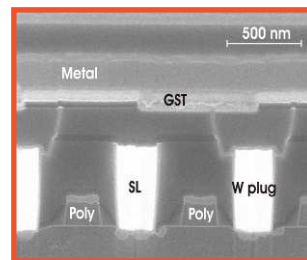
Near-Term and Long-Term Alternatives

More than 35 NVM alternatives have been so far proposed...

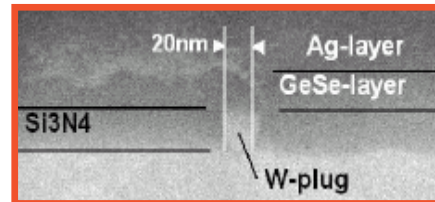
FERAM



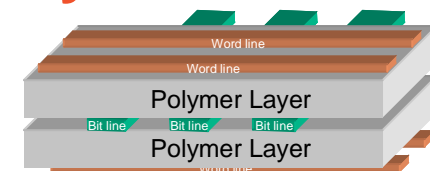
PCM



PMC RRAM



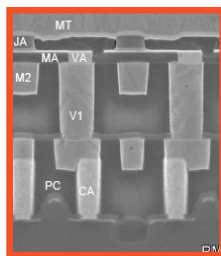
Polymer FeRAM



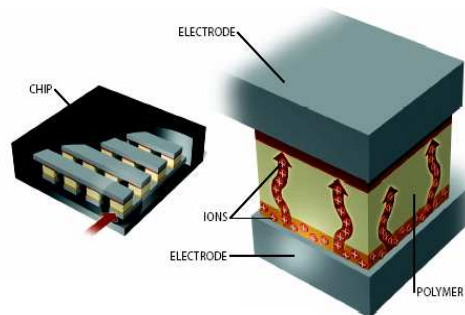
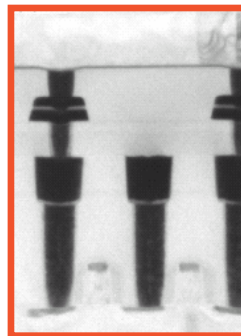
CNT



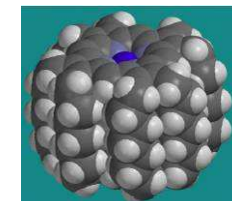
MRAM



MOx-RRAM Polymer RRAM

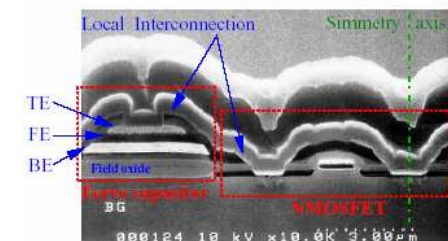
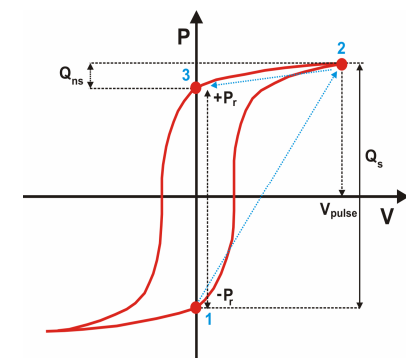
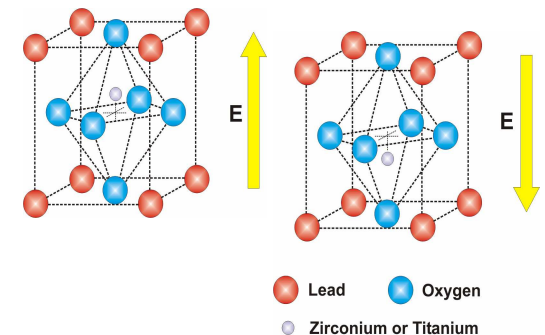


Molecular



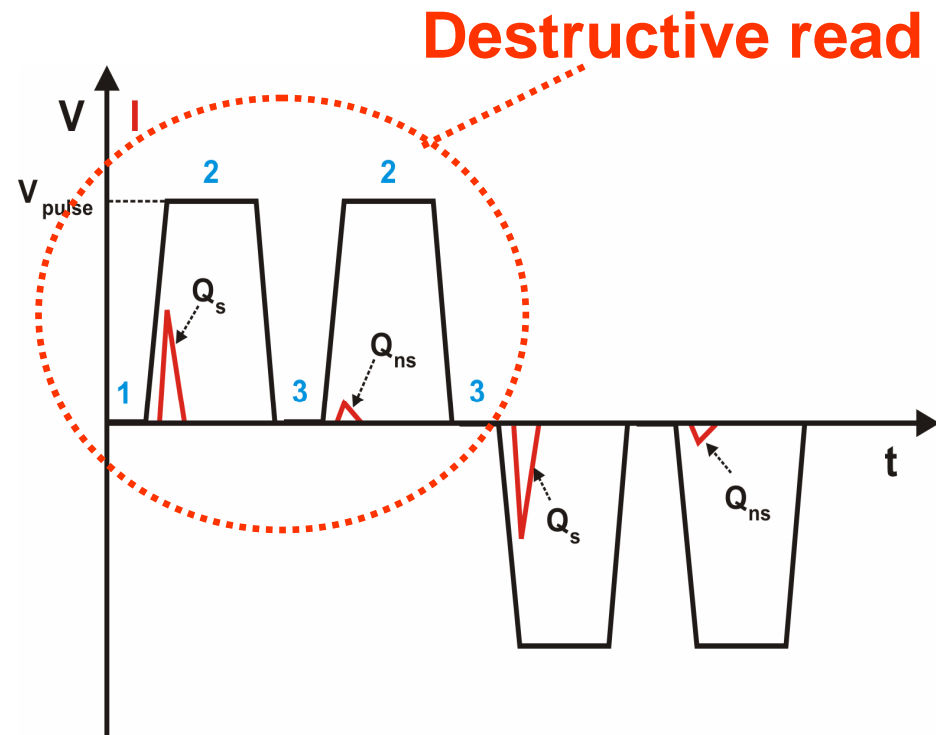
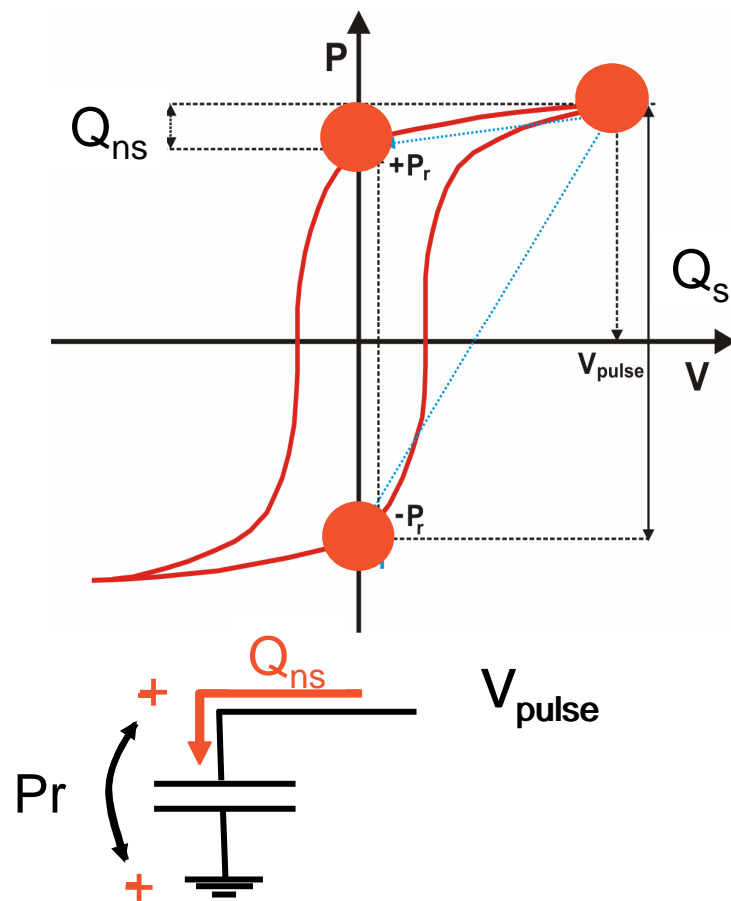
Ferroelectric RAM (FeRAM)

- **Storing mechanism**
 - Permanent polarization of a ferroelectric material
- **Writing mechanism**
 - Electric field produced in the ferroelectric layer by the voltage applied to the capacitor plates
- **Sensing mechanism**
 - Displacement current associated to the polarization switch
- **Cell structure**
 - DRAM-like: 1 transistor, 1 capacitor (1T/1C)



Ferroelectric RAM (FeRAM)

The basic memory element is a ferroelectric capacitor (FeCAP)



Read out signal: $Q_s - Q_{ns} = 2P_r$

FeRAM Scaling

Planar FeCAP area scaling

Minimum capacitance for sensing: 30 fF

Operating voltage: 1 V

Minimum charge for sensing: 30 fC

Equivalent electron number: 200 000

Technology node: 90 nm

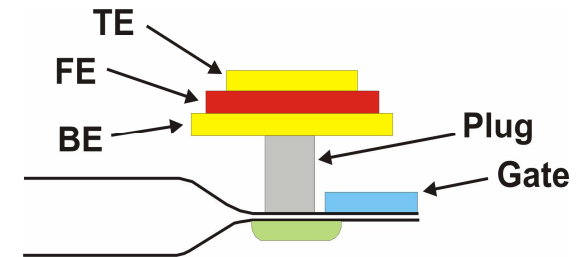
FeCAP area: 100 x 100 nm²

P_r: 20 μC/cm²

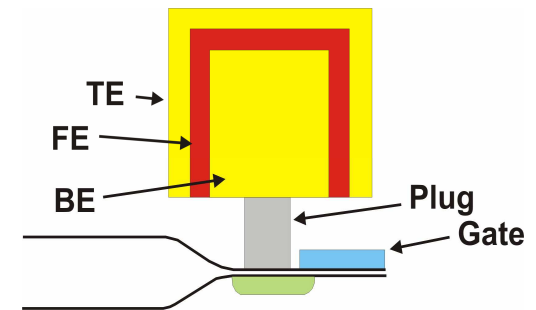
Available electron number: 25 000



3D approach necessary!

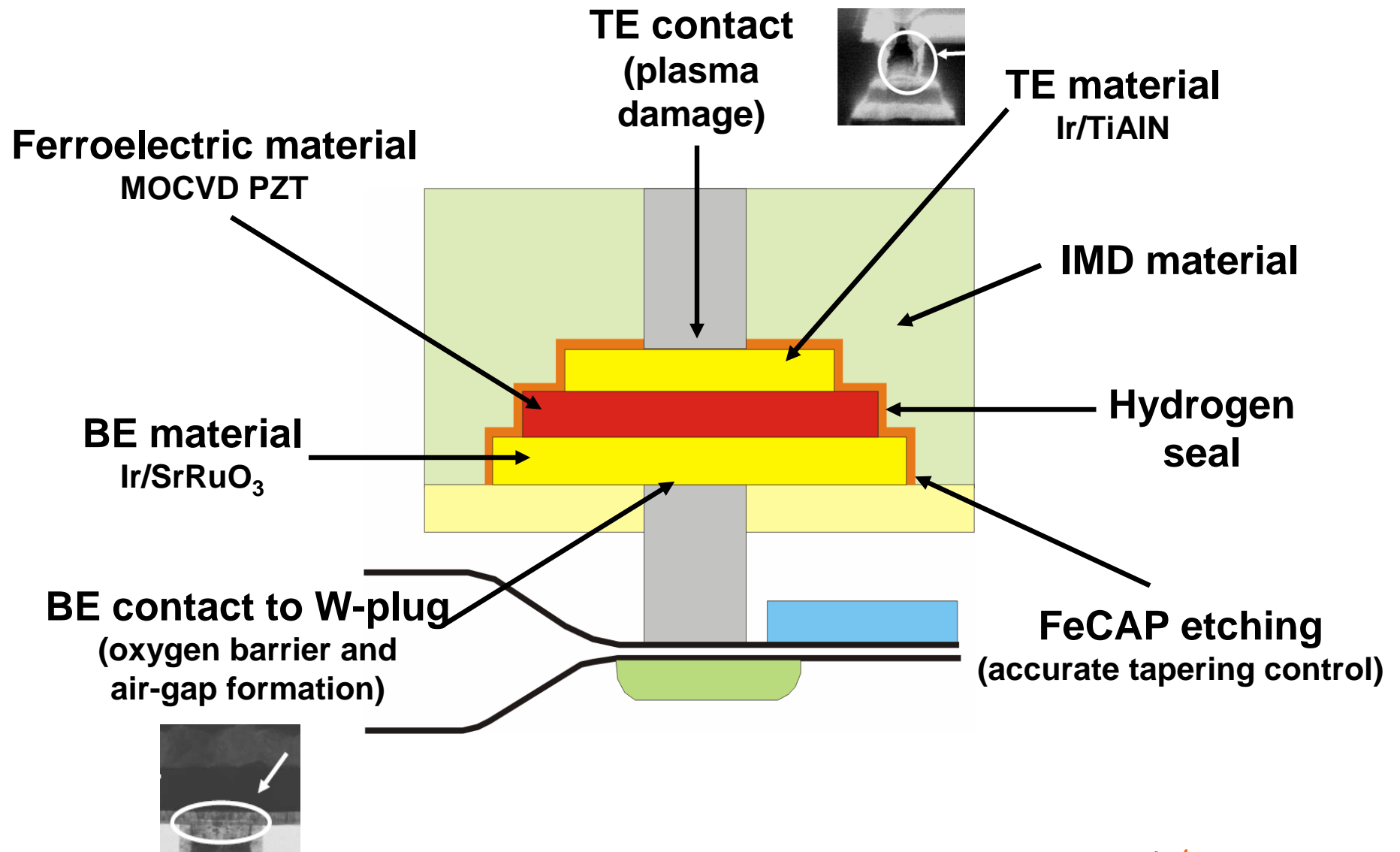


2D FeCAP



3D FeCAP

CMOS Integration



FeRAM: Advantages and Issues

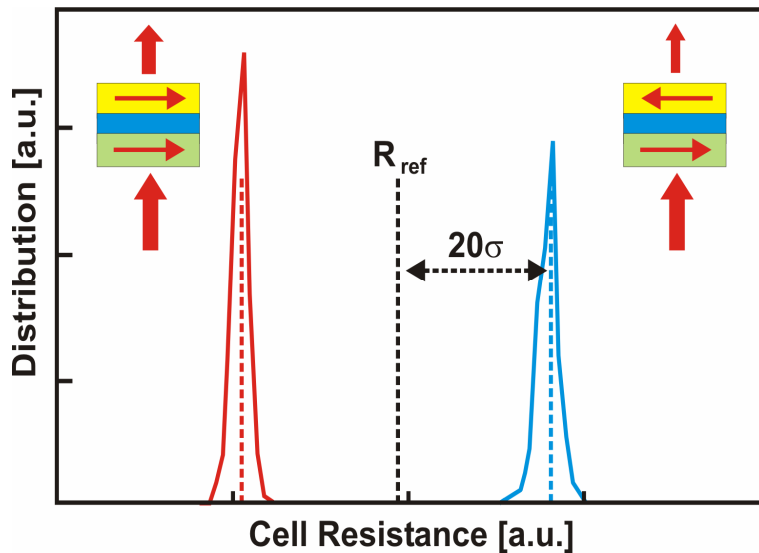
- Main advantages

- Fast ($<100\text{ns}$) read and write operations with no intrinsic limitation ($<100\text{ps}$)
- High write endurance ($>10^{12}$)
- Low voltage and low power operation

- Main issues

- Limited read endurance, destructive read-out (apart FeFET)
- Difficult process integration
- Large cell size vs. Flash and DRAM ($>15F^2$)
- Scaling limits and 3D capacitor required to go beyond the 90nm technological node

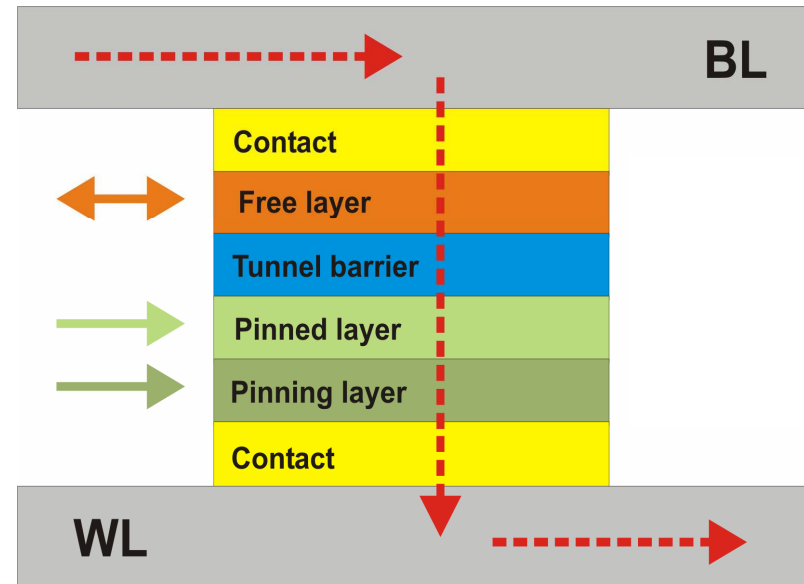
Magnetoresistive RAM (MRAM)



CoFe or NiFe/CoFe
 Al_2O_3 or MgO
 CoFe or NiFe/CoFe
 NiMn or IrMn or CrPtMn

Tunable magnetization of the free layer:

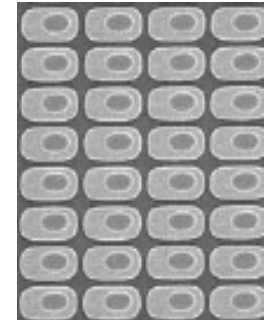
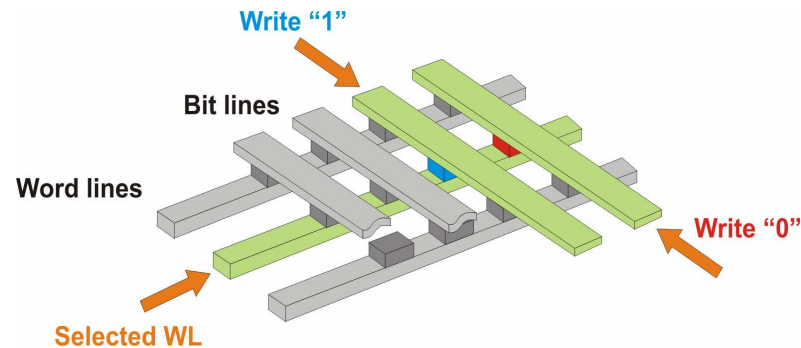
- Parallel to the pinned layer → low BL to WL resistance
- Antiparallel to the pinned layer → low BL to WL resistance



Sensing current

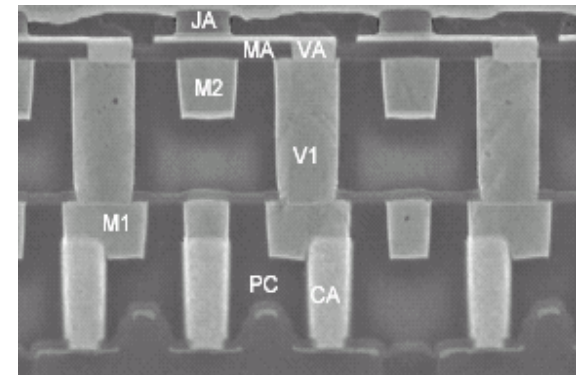
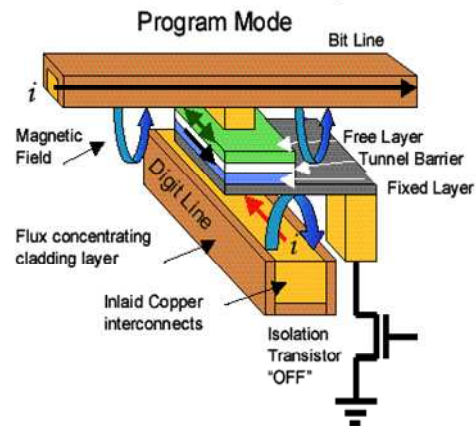
MRAM Cell Architectures

Cross-point architecture



N. Sakimura *et al.*, ISSCC 2003

MOSFET-selected architecture



W. J. Gallagher, Taiwan NVM Workshop, 2005

MRAM Scaling

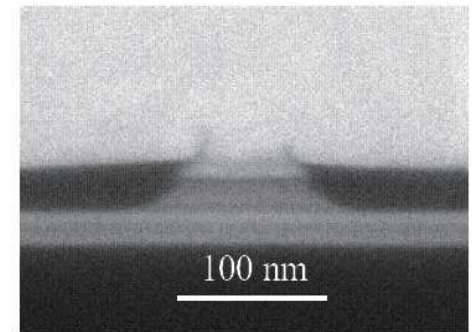
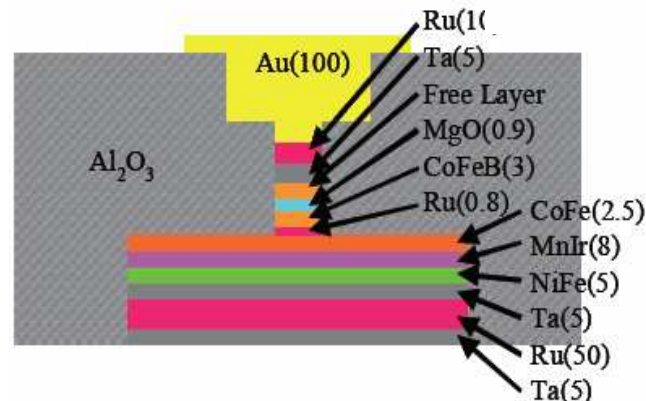
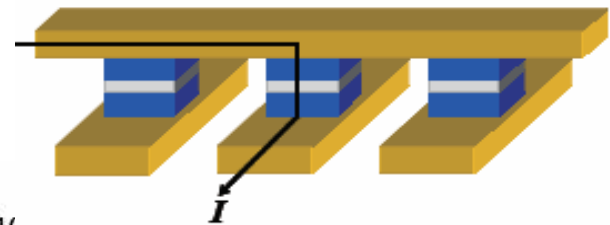
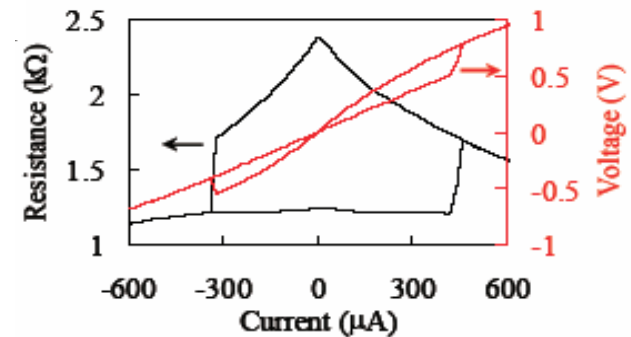
- Thermal stability
 - Smaller cell volume → lower thermal stability
 - Lower thermal stability → high-coercivity materials/ non-isotropic cell shape
 - high-coercivity materials/ non-isotropic cell shape → higher programming current
- Power consumption
 - Large magnetic fields for switching → large power consumption (~4mA in standard MRAM, ~7mA for Toggle MRAM)
 - Tradeoff with stability → increasing power consumption with scaling

**1st generation not expected to be viable
beyond the 90 nm technology node**

SPRAM (Spin-transfer torque RAM)

- Programming current flows through (not close to) the MTJ stack → spin torque mechanism
- Advantages
 - Lower programming current
 - Better scaling perspectives (retention/programming tradeoff)
- Issues
 - Self-read disturbance
 - Writing time depends on the device area
 - Integration

Proposed for 65nm
node and beyond



K. Miura *et al.*, VLSI Symp. on Tech. 2007

MRAM: Advantages and Issues

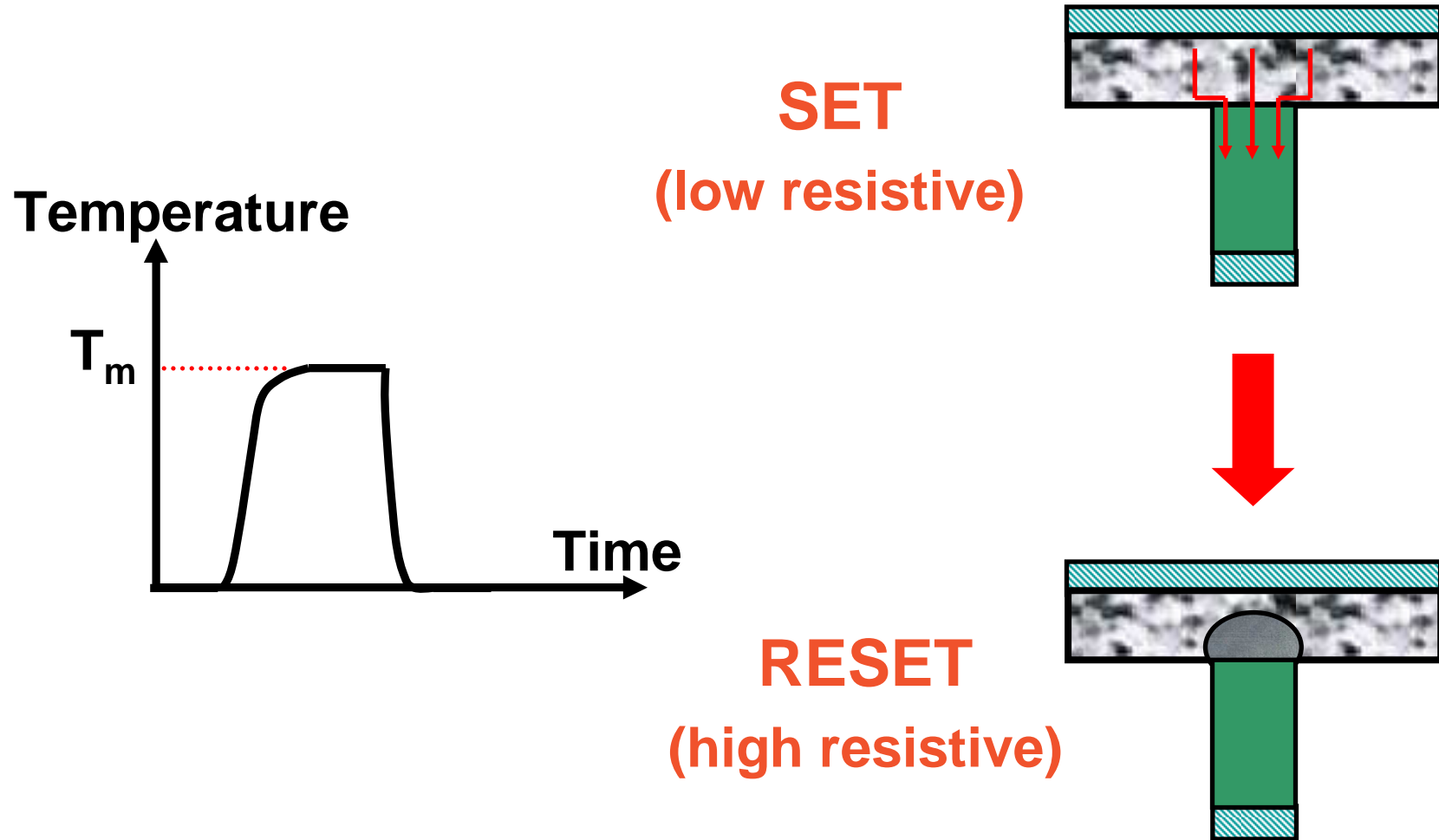
- Main advantages

- Fast write (<100 ns)
- High write endurance
- Low voltage write

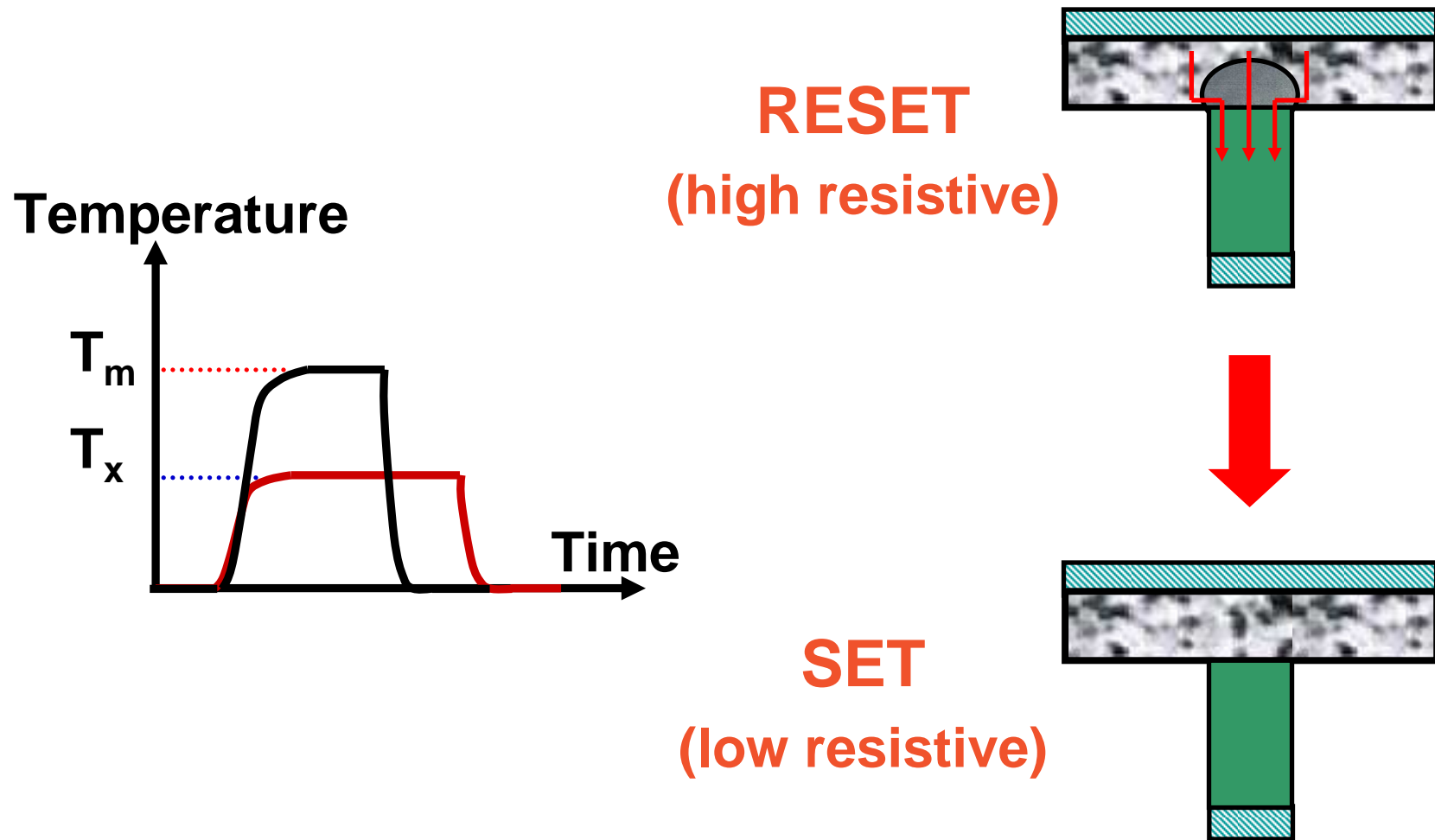
- Main issues

- Difficult process integration
- Large cell size vs. Flash and DRAM
- Large write current (> 10 mA/B)
- Small read signal
- Scaling limits

PCM Operating Principles - Set to Reset

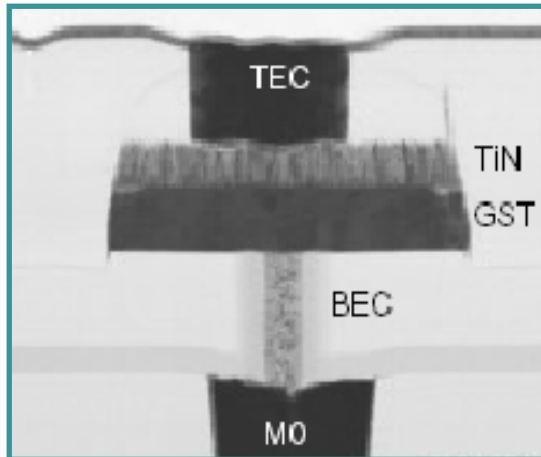


PCM Operating Principles - Reset to Set



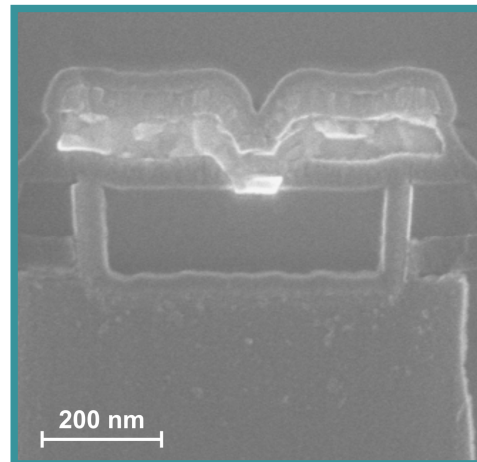
From PCM Cell Architectures...

Pore and lance structure



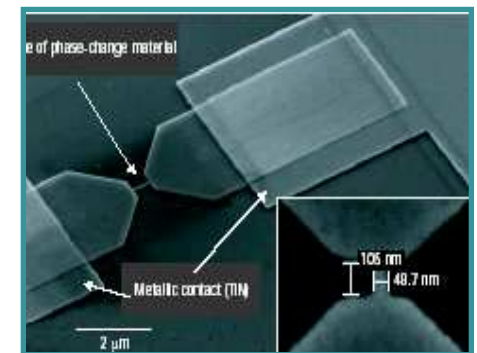
H. Horii *et al.*,
VLSI Symp. on Tech. 2003

μ trench structure



F. Pellizzer *et al.*,
VLSI Symp. on Tech. 2004

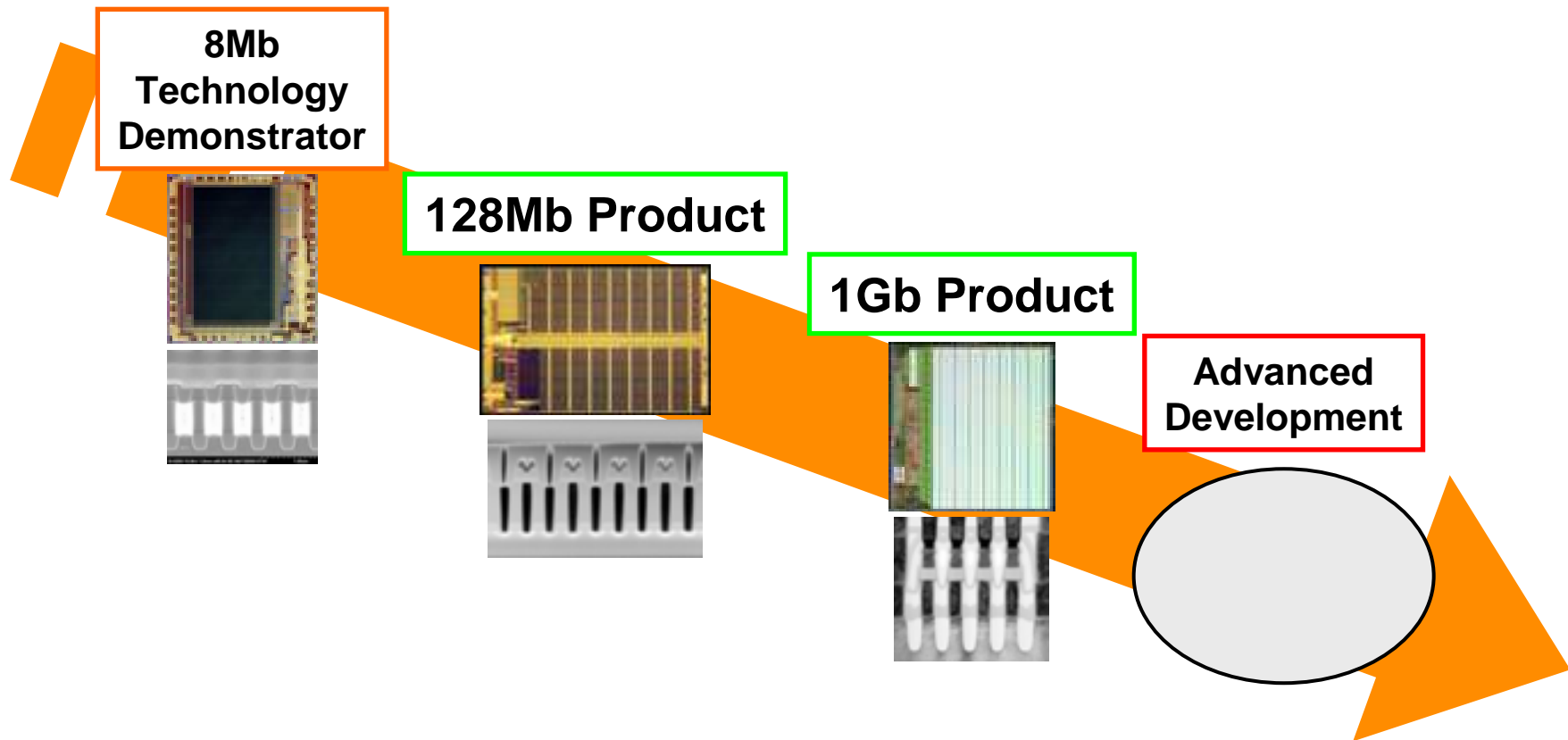
Planar structure



M. Lankhorst *et al.*,
Nature Material, April 2005

Efforts to reduce the programming current
still preserving a good controllability

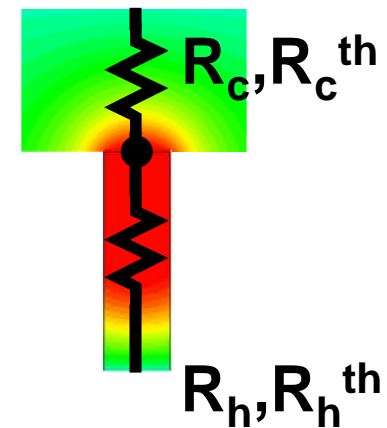
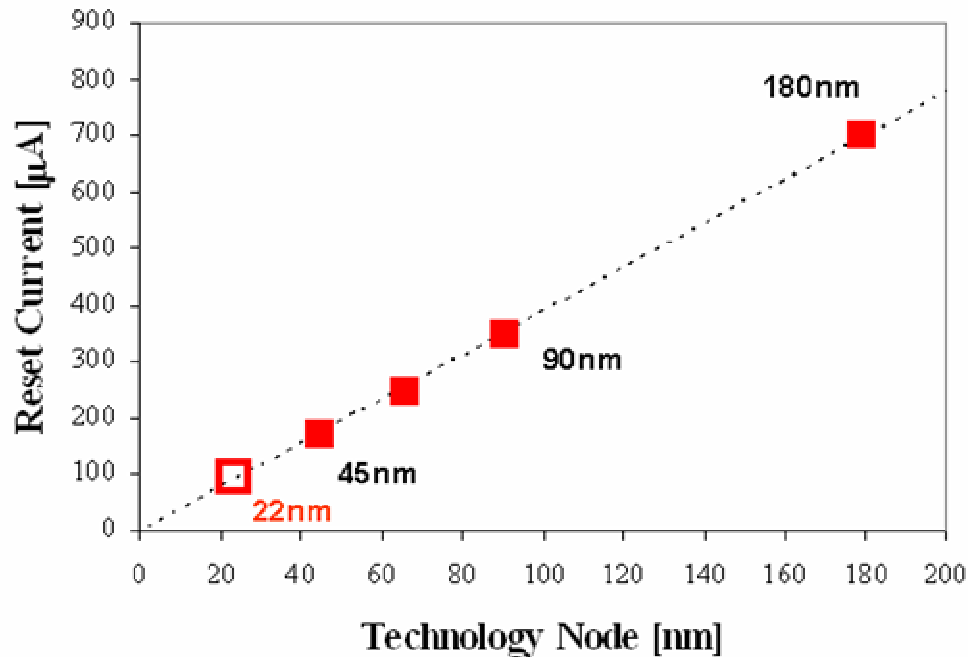
...To Technology Evolution



Technology Node (nm)

180	90	45	32	2X
------------	-----------	-----------	-----------	-----------

Scaling of programming current



$$\Delta T_M = P_M \cdot R^{th}$$



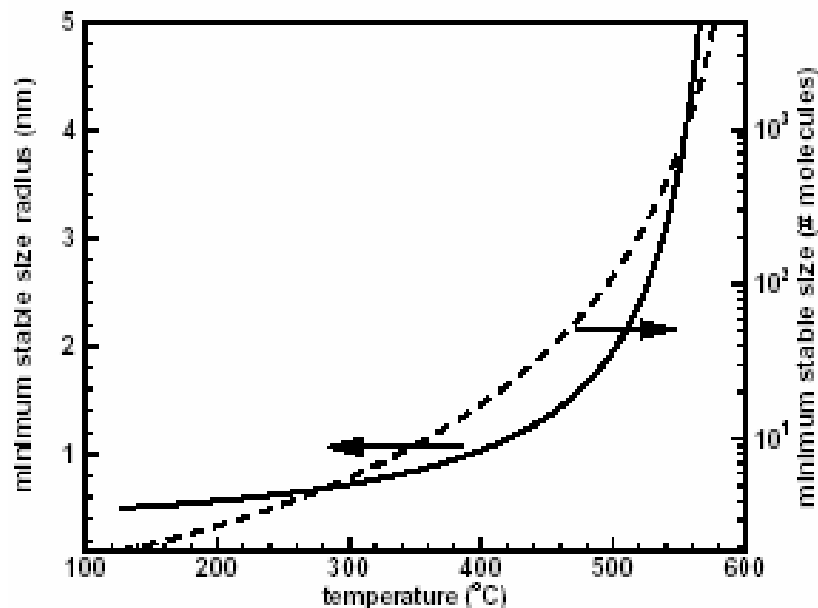
$$P_M = V_H \cdot I + \eta R I^2$$

PCM Stability

Phase change mechanism appears scalable to at least ~5nm

Can We Reach Tbit/sq.in. Storage Densities With Phase-Change Media?

C D Wright, M M Aziz, M Armand, S Senkader and W Yu
Department of Engineering, University of Exeter, UK



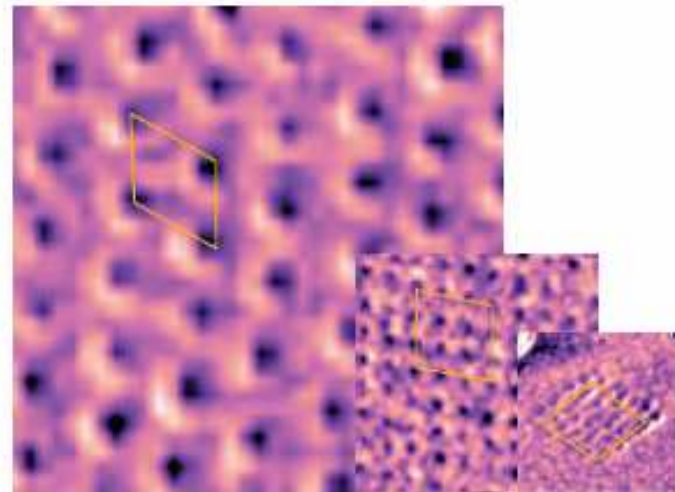
Min. stable cluster size for $\text{Ge}_2\text{Sb}_2\text{Te}_5$ vs. Temp.

D. Wright *et al.*, EPOCS 2004

14nm pitch 3.3Tb/in²

20nm pitch 1.6Tb/in²

40nm pitch 0.4Tb/in²



Crystalline Bits in Amorphous Matrix.

C. Lam, SRC NVM Forum 2004

PCM: Advantages and Issues

- Main advantages

- Fast write (~100ns)
- Good read signal window (factor ten in resistance)
- Medium/low voltage write
- Long endurance
- Cell size comparable to Flash and DRAM
- Good scalability
- MLC capabilities

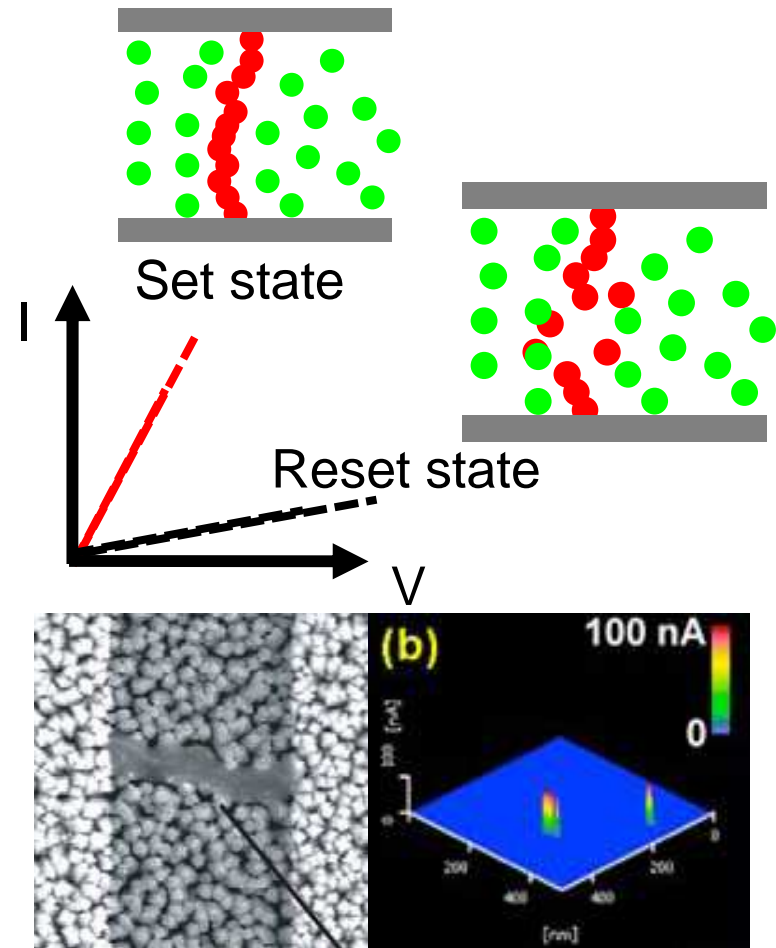
- Main issues

- Process integration for GST
- Heater-GST interface optimization
- Writing current reduction
- Retention at very high temperature (150°C)

Numonyx: 128Mbit Omneo™ PCM Products qualified, 1Gbit product presented at ISSCC 2010

Resistive RAM (RRAM)

- Storing mechanism
 - Resistive switching of a storage layer
- Writing mechanism
 - Current or voltage-induced conductance switching
- Sensing mechanism
 - Resistance change
- Cell structure
 - 1 transistor, 1 resistor (1T/1R) or cross-point (1R)

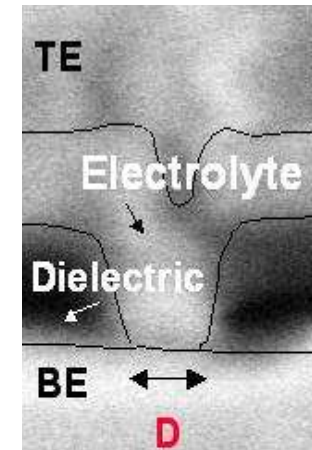


A. Sawa, *Materials Today* 2008

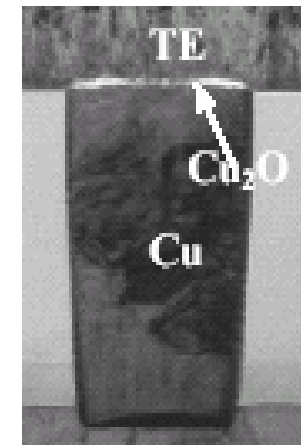
Yun et al., *Phys. Stat. Sol.* 2007

RRAM Proposed Alternatives

- Chalcogenide
 - GST and other phase-change alloys
 - AgGeSe, AgGeS, WO_3 and SiO_2 solid electrolyte
- Binary oxide
 - Nb_2O_5 , Al_2O_3 , Ta_2O_5 , TiO_2 , ZrO_x , Cu_xO and NiO
- Oxides with perovskite structure
 - SrZrO_3 , doped- SrTiO_3 , $\text{Pb}(\text{Zr}_x\text{Ti}_{1-x})\text{O}_3$ and $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$
- Conductive polymers
 - Bengala Rose, AlQ_3Ag , Cu-TCNQ



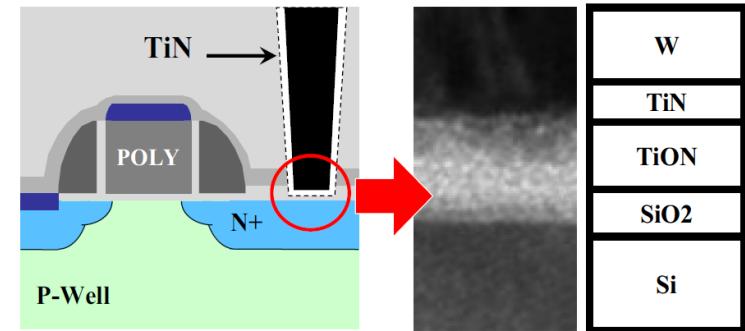
M. Kozicki, EPCOS 2006



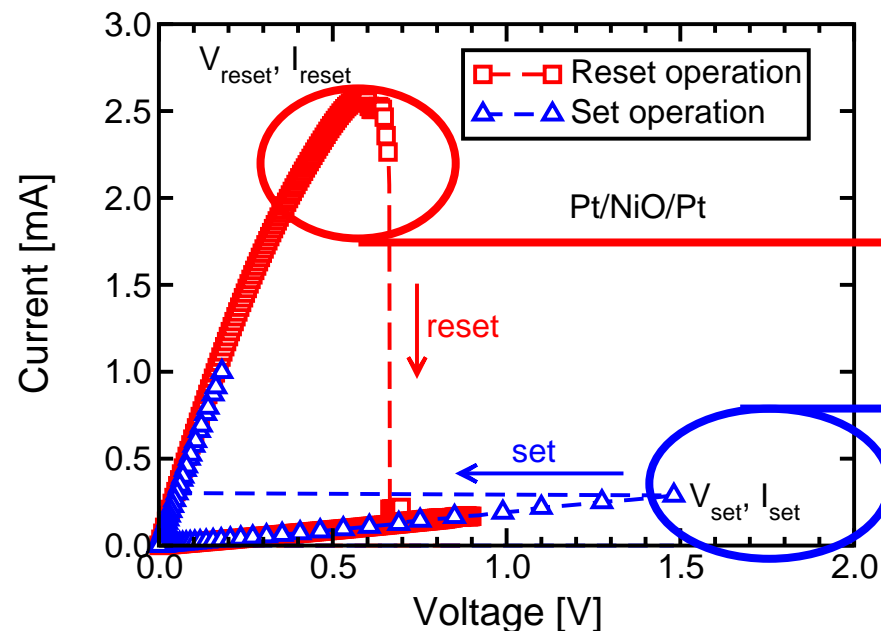
A. Chen *et al.*,
IEDM Tech. Dig. 2005

Resistive Oxides Memories

- **Unipolar** resistive switching in a binary oxide layer
- Feasible small cell size (4-8 F²), very low-cost solution



2Mbit 1T-1R 90nm test chip
(Y.H. Tseng, *et al.*,
IEDM Tech. Dig. 2005)

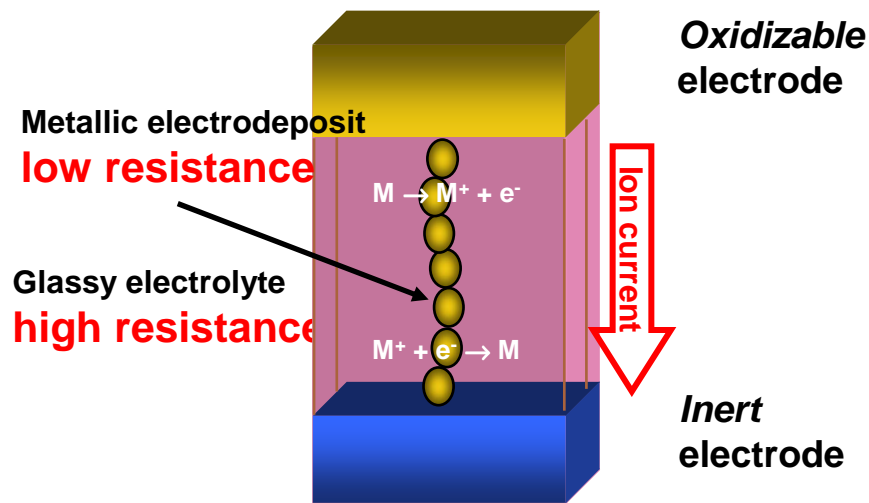


Joule heating + Ni thermal oxidation
(U. Russo *et al.*, IEEE T-ED, 2009)

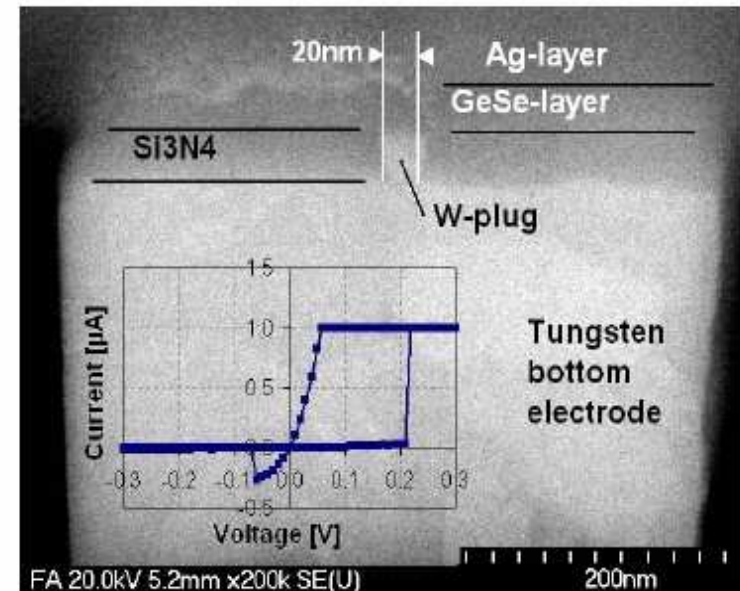
Threshold switching +
NiO reduction
(D. Ielmini *et al.*, APL 2009)

Programmable Metallization Cells – bipolar mode

Programmable metallization cell (PMC): a conductive filament of silver is created by diffusion into a chalcogenide (solid electrolyte) by applying an electric field



M. Kozicki, EPCOS 2006



M. Kund *et al.*, IEDM Tech. Dig., 2005

RRAM: Advantages and Issues

- Main advantages

- Good read signal window (factor ten in resistance)
- Medium/low voltage write
- Low programming current and energy (bipolar RRAM)
- Cross-point solutions available
- Good scalability

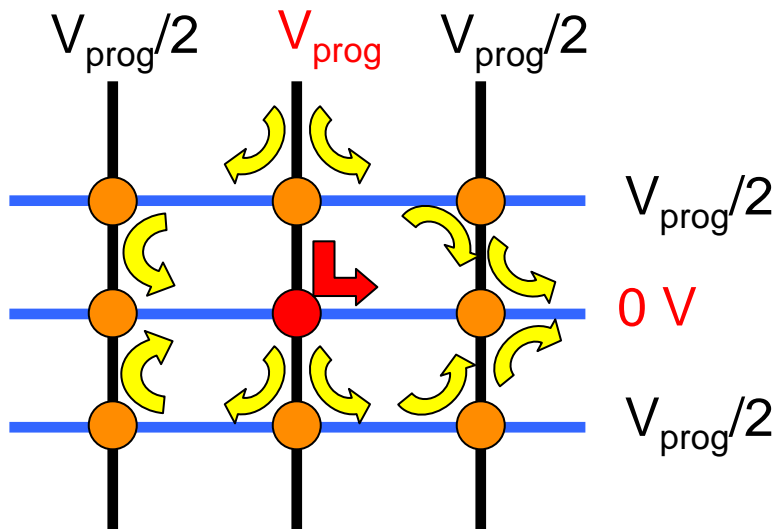
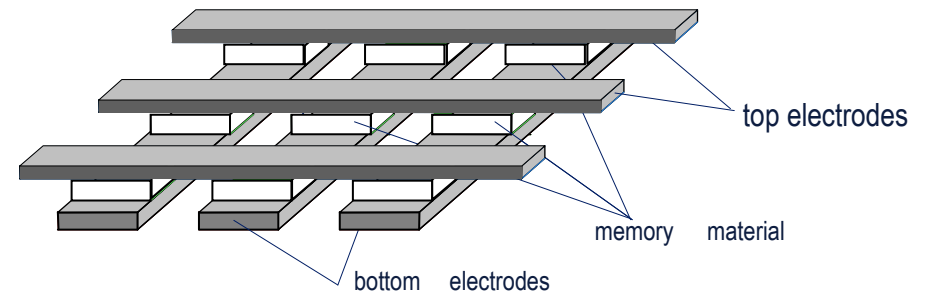
- Main issues

- Statistical variations of programming voltage/current and programmed resistance
- Difficult process integration (low thermal budget required)
- Retention capabilities for 10years at 85°C must be demonstrated

Huge variety of proposed materials and solutions, but still at research level

The crossbar integration

- “simple” structure → low cost
- reduced cell size: $4F^2$
- suitable for 3D stacking → cell size $(4/n)F^2$



- Parasitic paths exist through neighbouring cells
 - Programming (and also reading) can perturb the array
 - Some selector is needed to stop parasitic paths
- ➔ low-temperature diode selectors needed for process compatibility

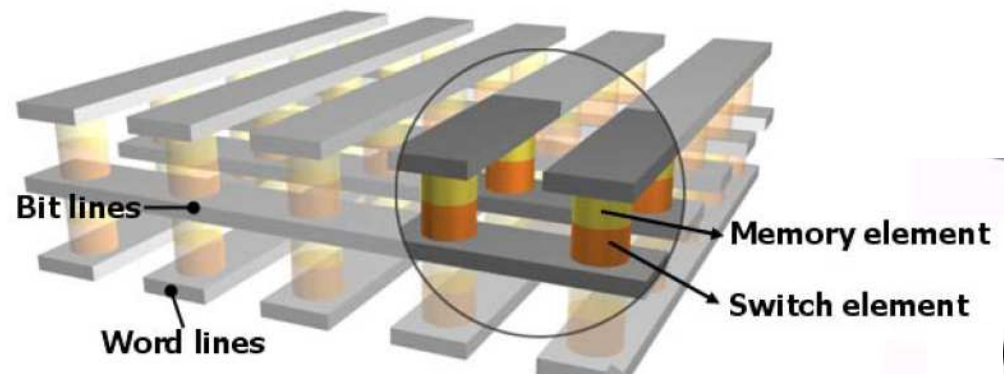
A wide range of material choices

Selector device

- Homojunctions → polySi p/n junctions
- Heterojunctions → p-CuO/n-InZnO
- Schottky diode → Ag/n-ZnO
- Chalcogenide Ovonic Threshold Switching (OTS) materials

Storing device

- RRAM → NiO, Nb₂O₅, Al₂O₃, Ta₂O₅, TiO₂, ZrO_x, Cu_xO
- PCM → chalcogenide material



Summary

- An impressive growth of the portable systems market has been enabled by the NVM Flash technology and it is expected that Flash will dominate the NVM production well beyond the 30nm technology node
- Semiconductor industries plan to exploit new physical mechanisms and new materials in nano-scale NVM
 - provide valuable solutions to cover specific application requirements
 - interesting challenges to be solved (integration into CMOS process, reliability assessment, scaling feasibility)
- Among these alternative NVM, Phase Change Memory is a most promising candidate to take over mainstream Flash technology
 - better performances
 - longer-term scalability