# Non-Volatile Memory Technology Overview

Ugo Russo, Andrea Redaelli, Roberto Bez

# 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

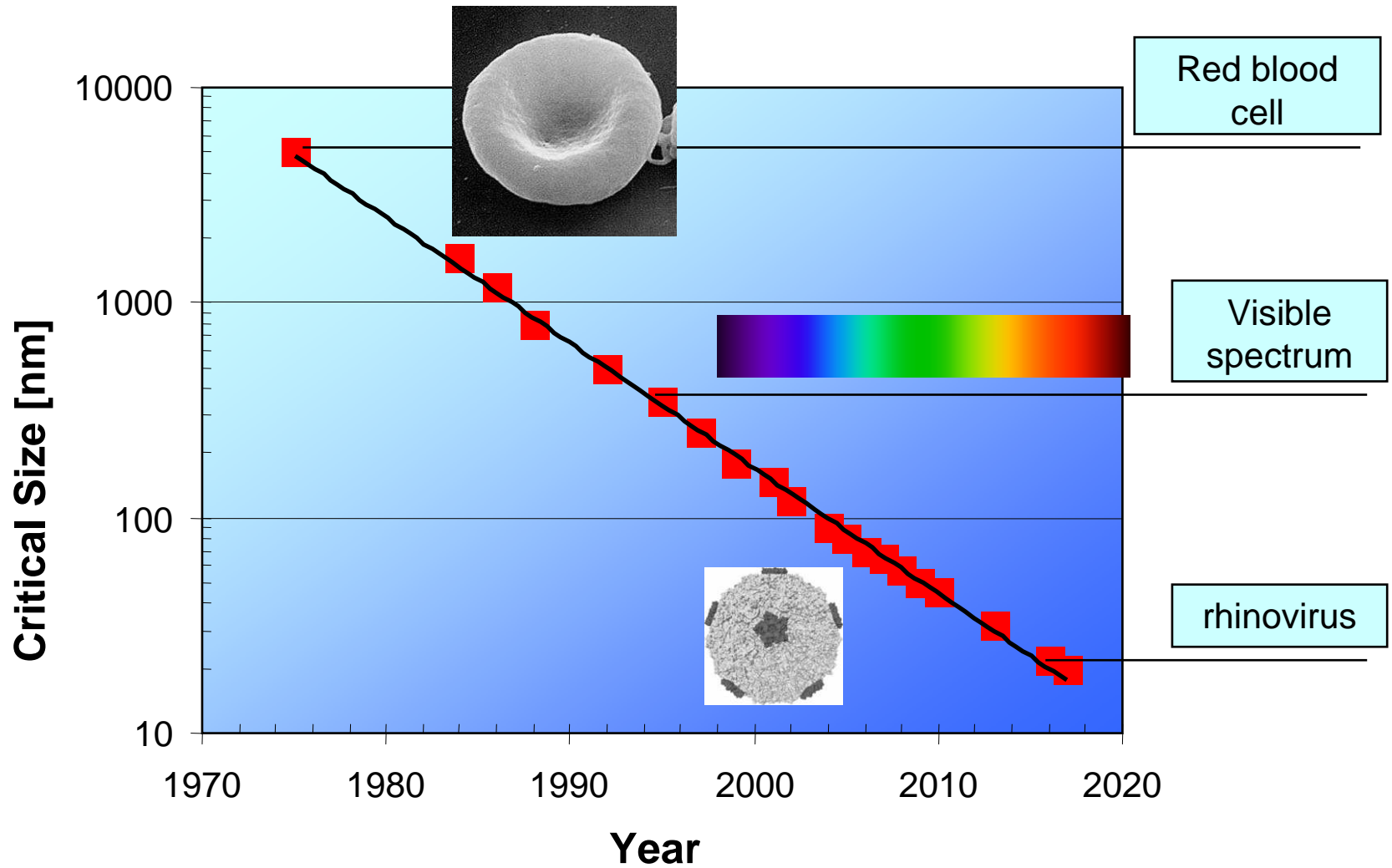# Non-volatile Memory: demand and market

Enabling technology for consumer portable goods



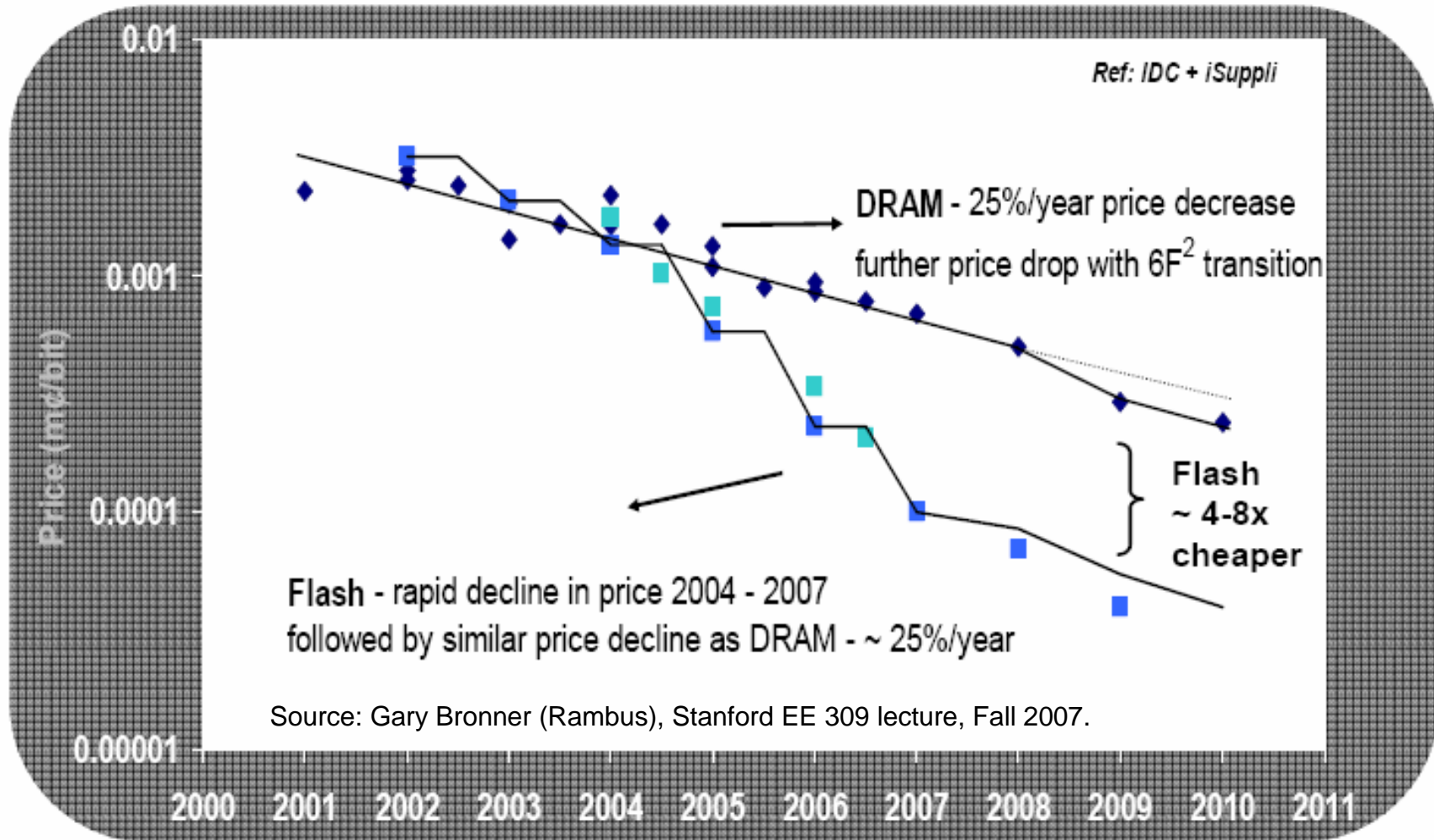Coming up: Solid State Discs

- 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



Ref: IDC + iSuppli

DRAM - 25%/year price decrease
further price drop with $6F^2$ transition

Flash ~ 4-8x cheaper

Flash - rapid decline in price 2004 - 2007
followed by similar price decline as DRAM - ~ 25%/year

Source: Gary Bronner (Rambus), Stanford EE 309 lecture, Fall 2007.

numonyx™

# A broadening palette of materials

| | |
|---|---|
| **Strained Si-Ge, SOI, GaAs** | |
| **PCM, FeRAM, MRAM** | |
| **High k dielectrics** | |
| **Porous Dielectrics** | |
| **Low k dielectrics, Copper** | |
| **Salicide (Ti, Co, Ni, Pt?)** | |
| **Barriers (Ti, TiN, WTi, TaN), Tungsten Salicide** | |
| **Silicon, Oxide, Nitride, Aluminium** | |

1.5    0.6    250    120    65    32

1.0    350    180    90    45    25

← microns →    nanometers

numonyx™
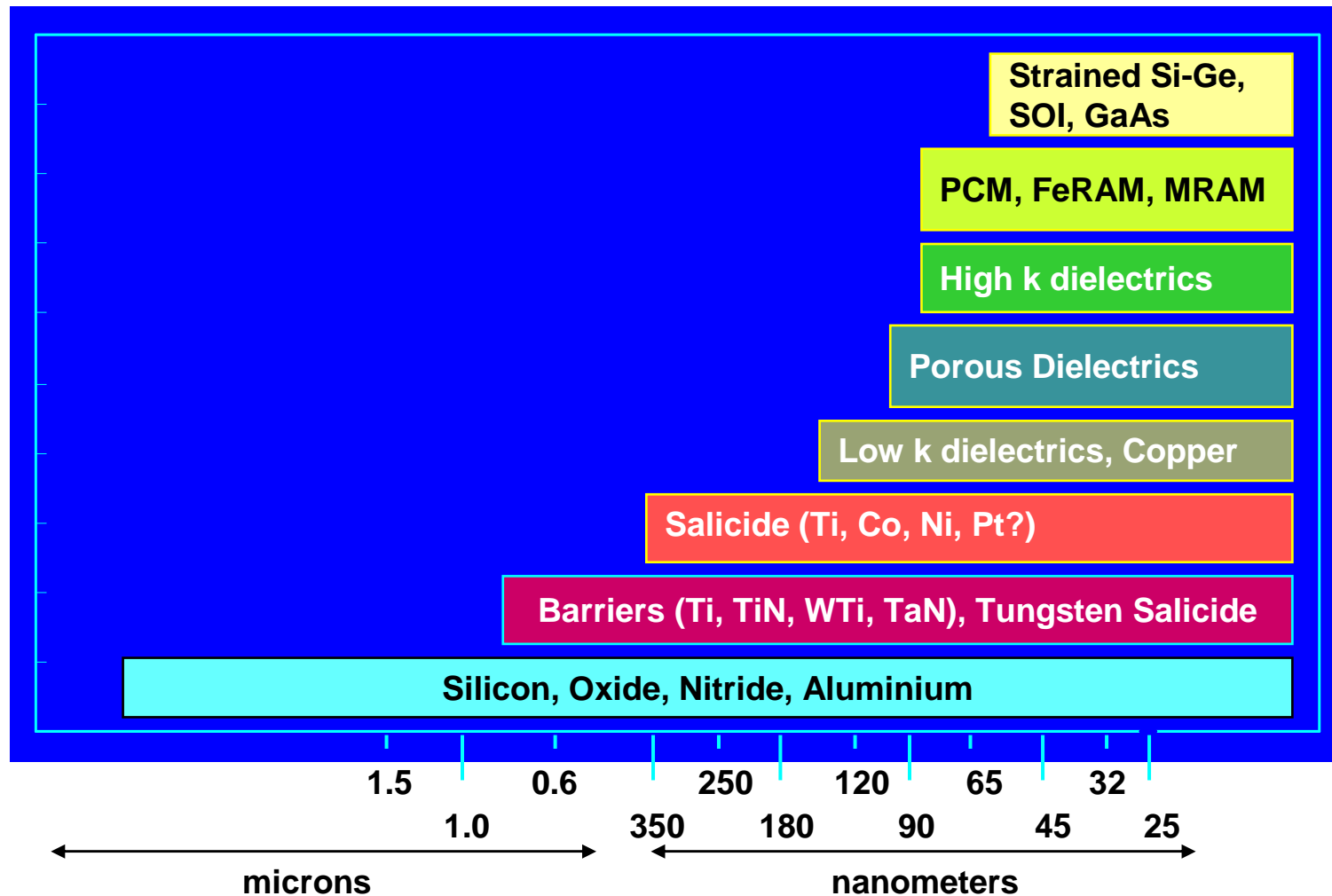
# Floating Gate NVM History

1967 → **First Floating Gate Structure**

1977 → **EPROM**

1980 → **EEPROM**

1985 → **1T EEPROM (Flash)**

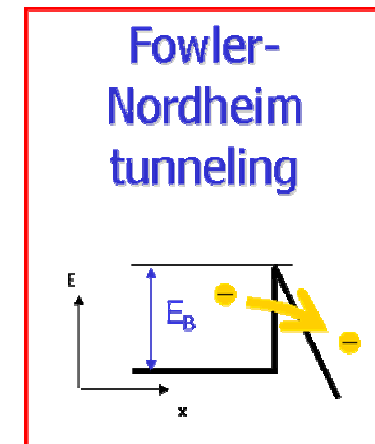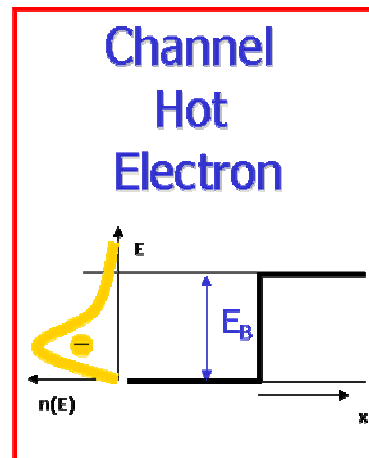1988 → **NOR Flash**

1989 → **NAND Flash**

1995 →**MLC Flash**

**numonyx™**

# Flash memory working principle

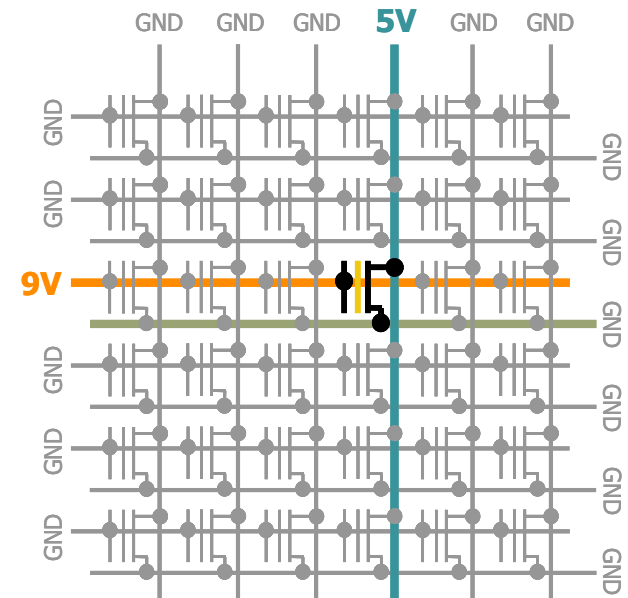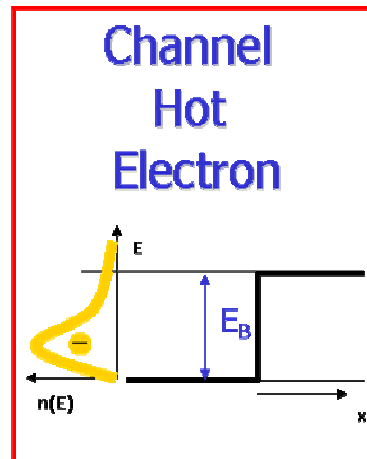- A chargeable floating gate is introduced to tune the transistor threshold voltage
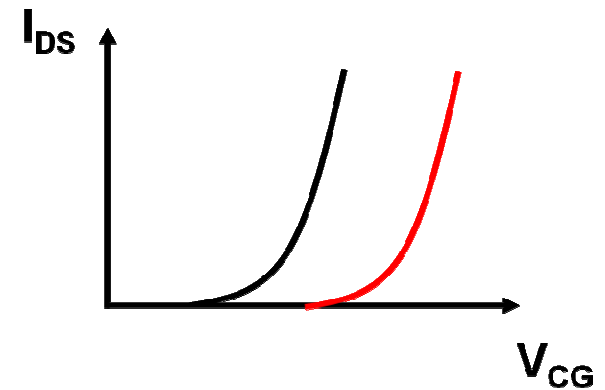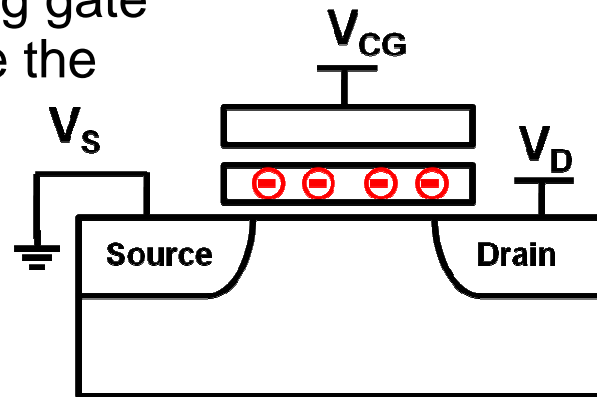


- Different physical mechanisms are used to charge/uncharge the floating gate, depending on memory architecture



Channel Hot Electron

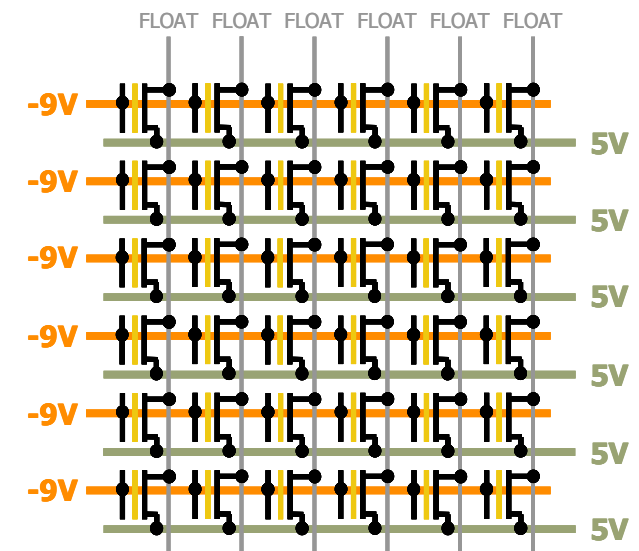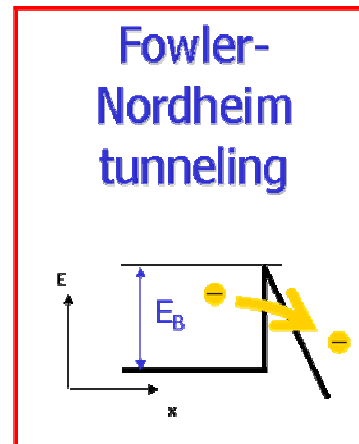

Fowler-Nordheim tunneling

numonyx™

# NOR flash

Programming:

- Single bit
- Random access
- High programming current

Erase:

- Block erase
- Low programming current

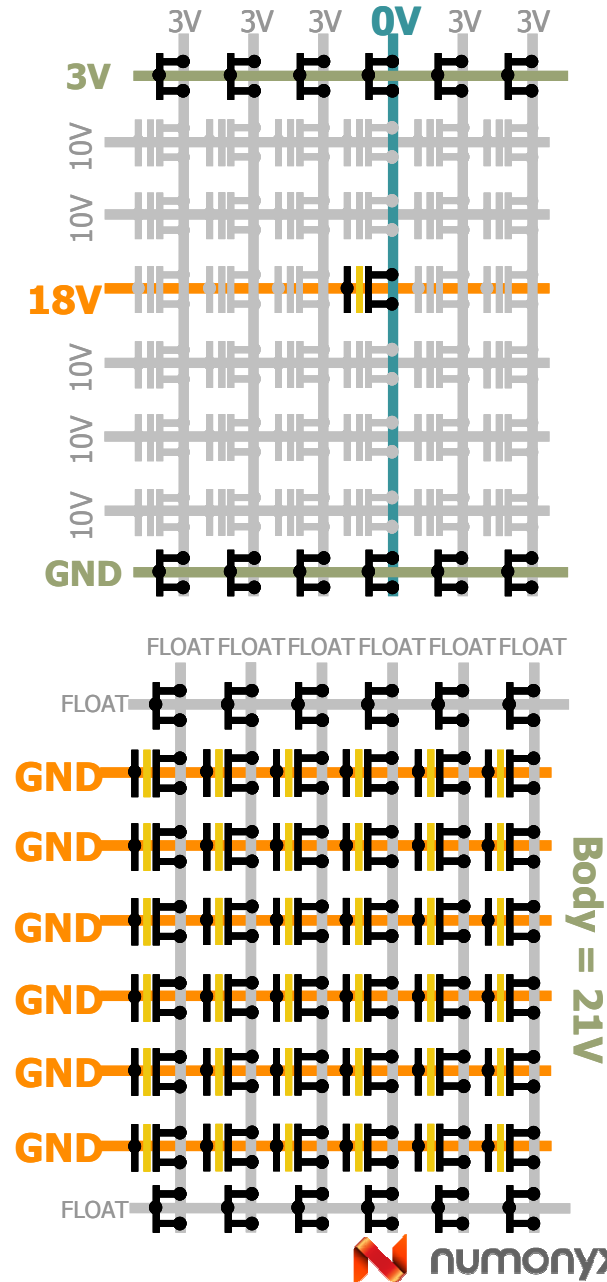Channel Hot Electron

Fowler-Nordheim tunneling

numonyx™

# NAND flash
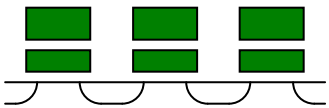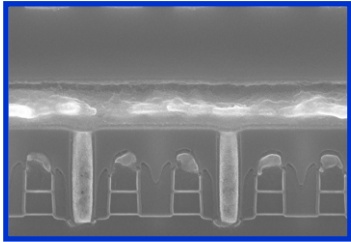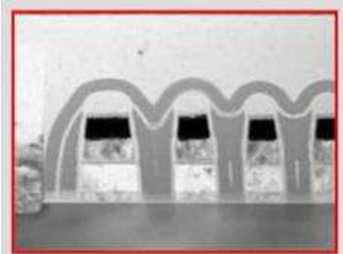
Programming:

- Serial access
- Low programming current

Fowler-Nordheim tunneling

$E_B$

Erase:

- Block erase
- Low programming current

3V 3V 3V 0V 3V 3V

3V

10V 10V

10V

18V

10V

10V

10V

GND

FLOAT FLOAT FLOAT FLOAT FLOAT FLOAT

FLOAT

GND

GND

GND

GND

GND

GND

FLOAT

Body = 21V

numonyx™
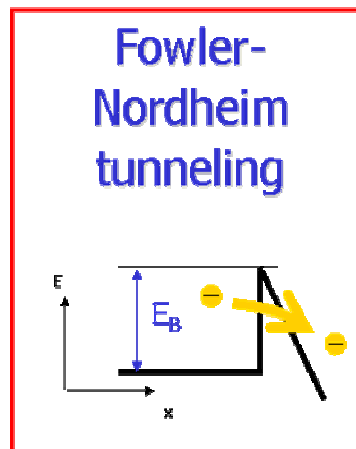
# NOR and NAND Stacked Gate Flash
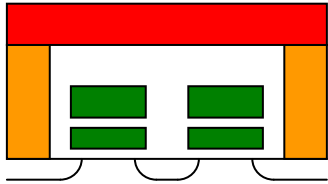
| | NOR | NAND |
|---|---|---|
| |  |  |
| Cell size ($F^2$) | 10 | 5 |
| Read access | Random (fast ~50ns) | Serial |
| Progr. Mechanism/ Troughput | CHE / 0.5 MB/s | FN / 7-10MB/s |
| SEM Cross-section (BL direction) |  |  |

numonyx™

# The Multilevel Cell Concept
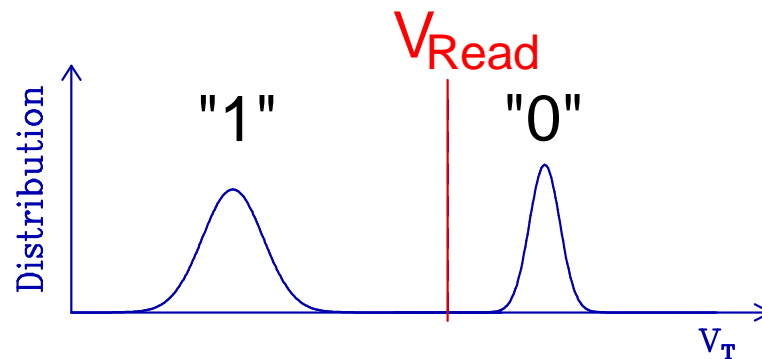


❑ 1 bit/cell => 2 levels
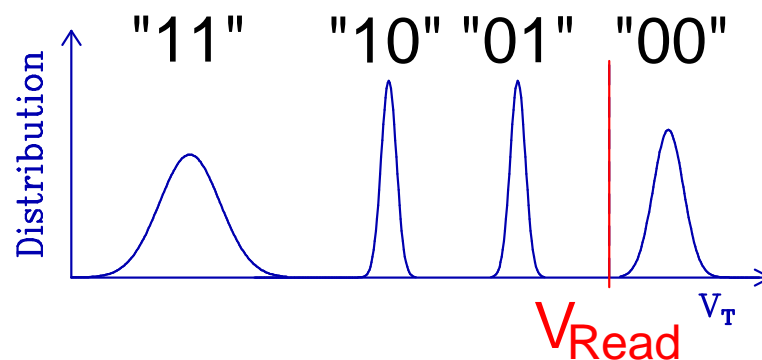
❑ 2 bit/cell => 4 levels

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

numonyx™

# Flash Cell Scaling Challenges

- **Cell basic structure unchanged through the different generations**

- **Cell area scaling through**:
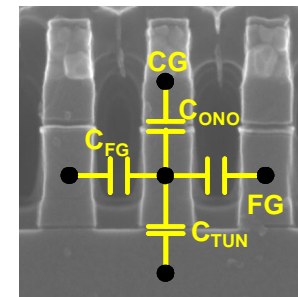    1. Active device scaling (W/L)
    2. Passive elements scaling

- **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

numonyx™
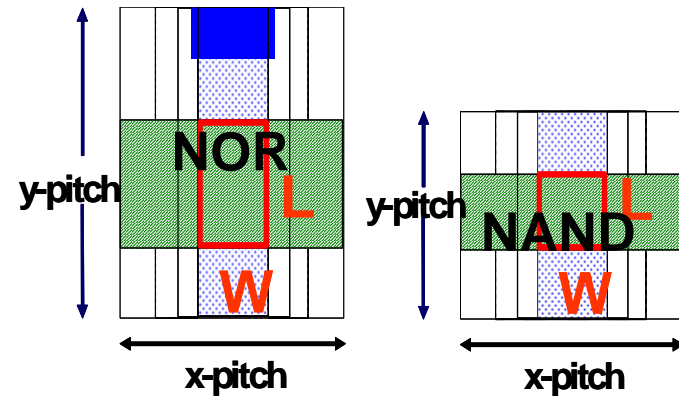
# Statistical effects

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
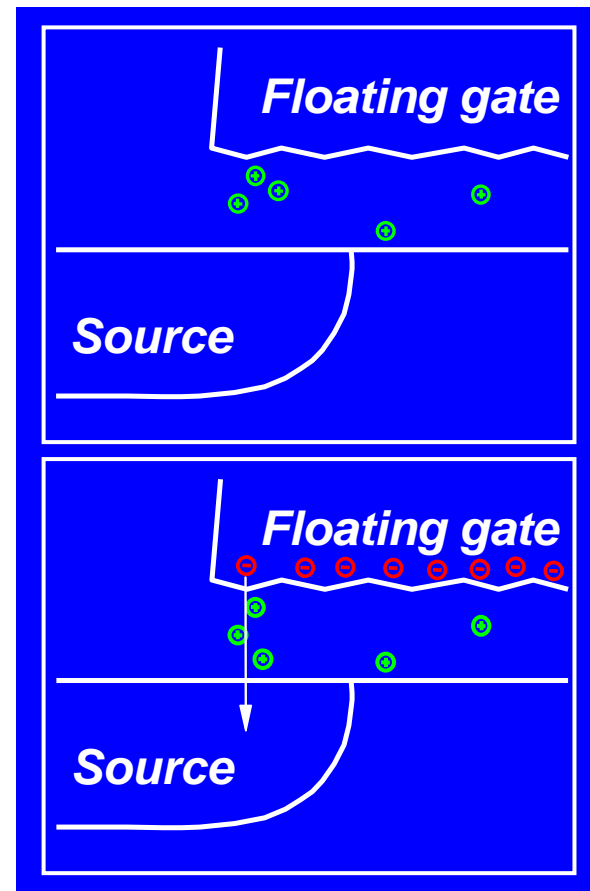
# Random Telegraph Signal:
## when one electron makes the difference.

### Flash Cell Vt instability



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



Trap

$E_F$

numonyx™
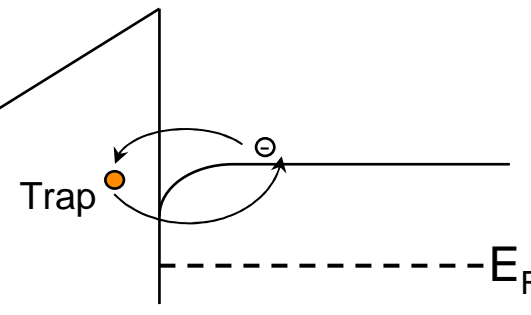
# Evolution: TANOS NAND Flash



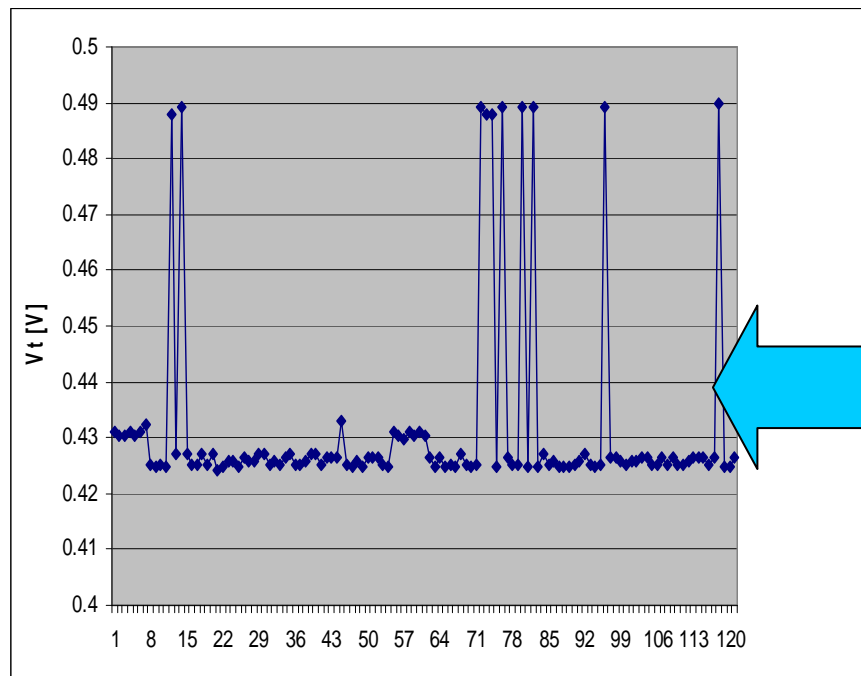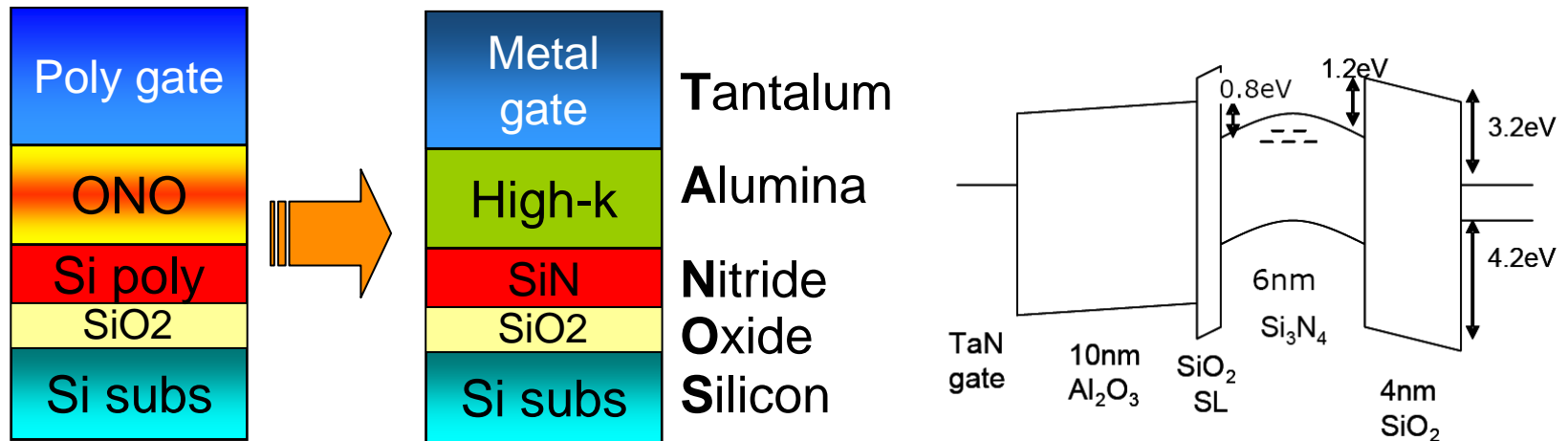☐ **From continuous floating gate to charge trapping layer**

☐ **Based on well known SONOS concept, but**

    o  thicker tunnel oxide for retention

    o  high-k ($Al_2O_3$) 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

**numonyx™**

# 3D architecture

- **Bit-Cost Scalable (BiCS) flash memory**

  – Increased memory density

  – Better control of channel charge/current ➔ programmed $V_{th}$, subthreshold current
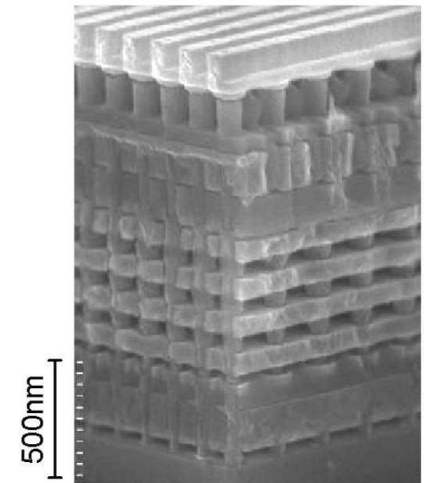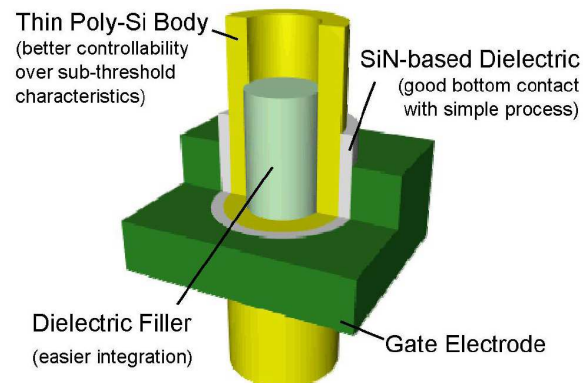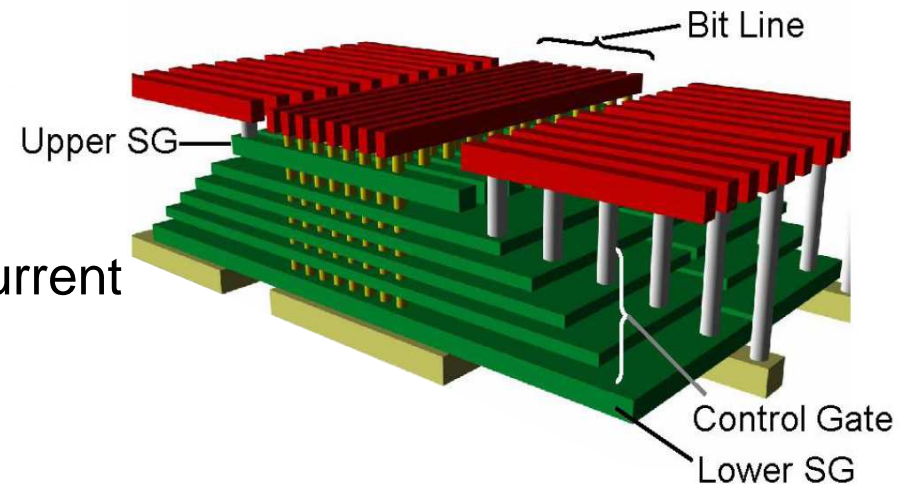
  – Integration complexity

  – Early development stage

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

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



Bit Line

Upper SG

Control Gate

Lower SG



Thin Poly-Si Body (better controllability over sub-threshold characteristics)

SiN-based Dielectric (good bottom contact with simple process)

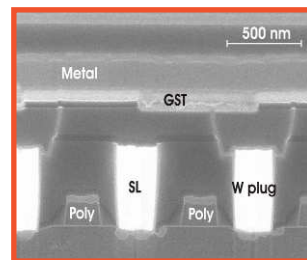Dielectric Filler (easier integration)

Gate Electrode

500nm

**numonyx™**

# Near-Term and Long-Term Alternatives

**More than 35 NVM alternatives have been so far proposed…**

**Polymer FeRAM**

**FERAM**

**PCM**
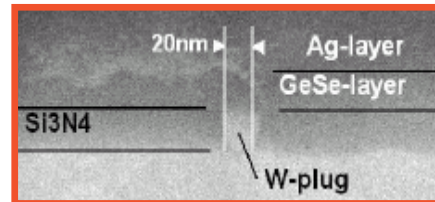
**PMC RRAM**
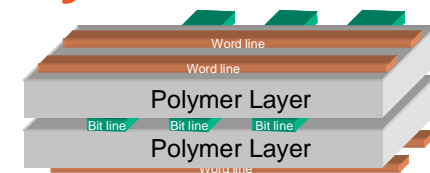
**CNT**

**MRAM**
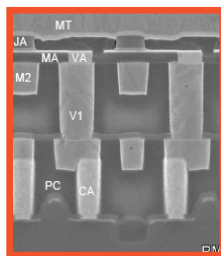
**MOx-RRAM**

**Polymer RRAM**

**Molecular**

numonyx™
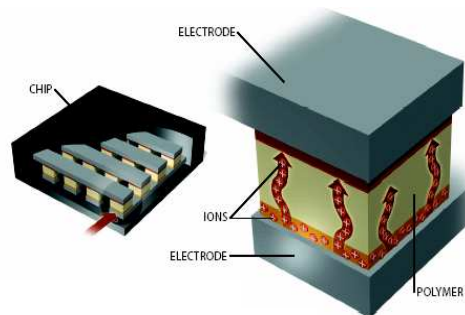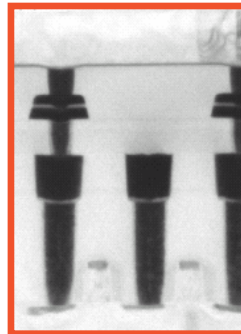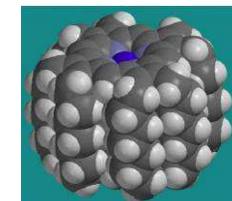
# Ferroelectric RAM (FeRAM)

- **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)
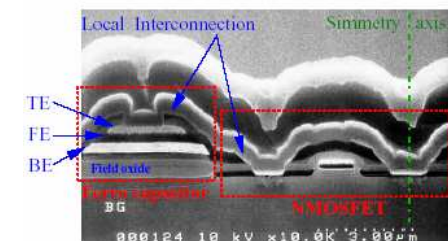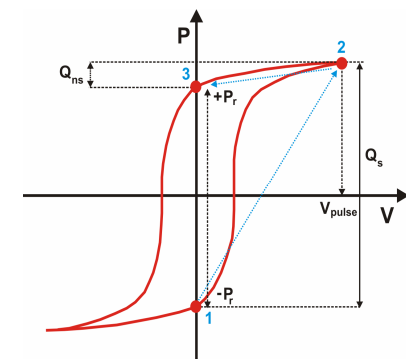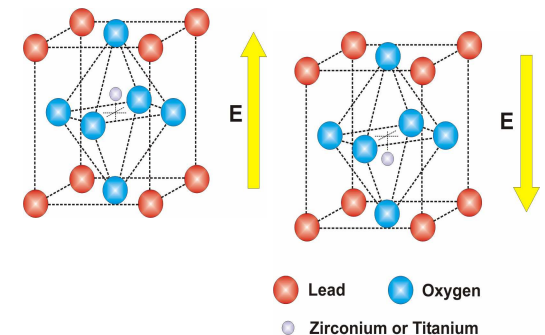
# Ferroelectric RAM (FeRAM)

The basic memory element is a ferroelectric capacitor (FeCAP)



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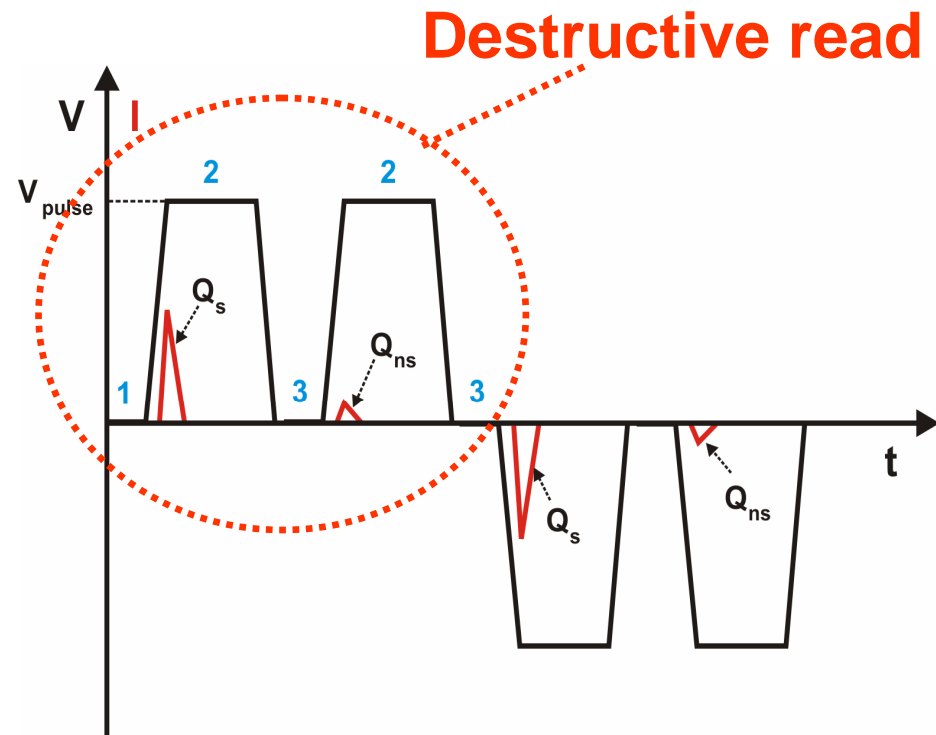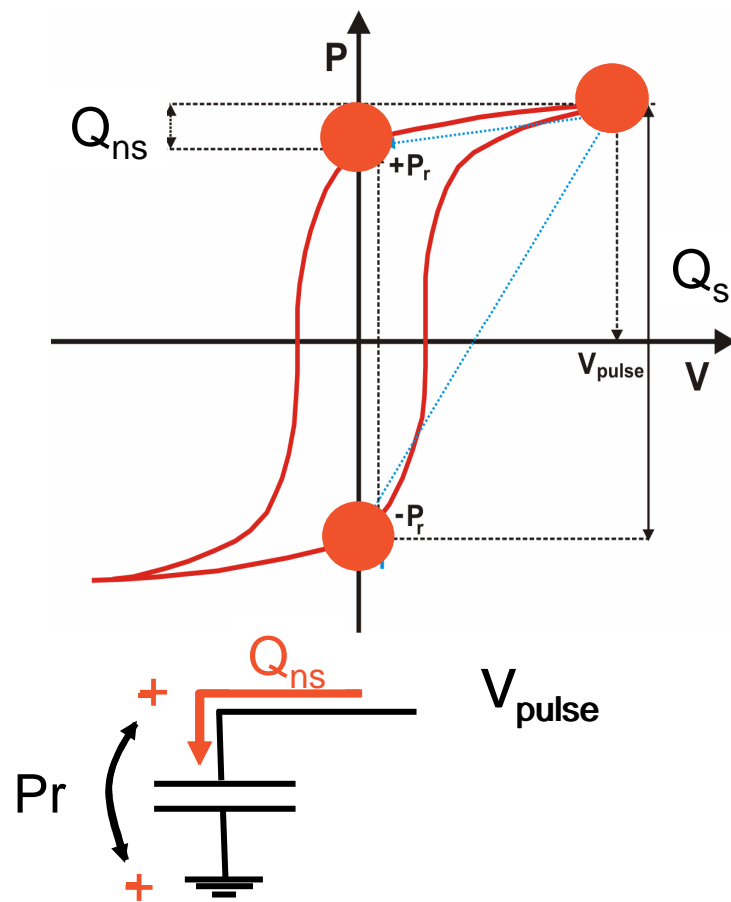
**Read out signal: $Q_s - Q_{ns} = 2P_r$**

numonyx™

# 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 nm$^2$**
**$P_r$:**   **20 μC/cm$^2$**

**Available electron number:**   **25˙000**

**3D approach necessary!**

**2D FeCAP**

**3D FeCAP**

numonyx™

# CMOS Integration



Labels on the diagram:

- **TE contact** (plasma damage)
- **TE material** Ir/TiAlN
- **Ferroelectric material** MOCVD PZT
- **IMD material**
- **BE material** Ir/SrRuO$_3$
- **Hydrogen seal**
- **BE contact to W-plug** (oxygen barrier and air-gap formation)
- **FeCAP etching** (accurate tapering control)

numonyx™
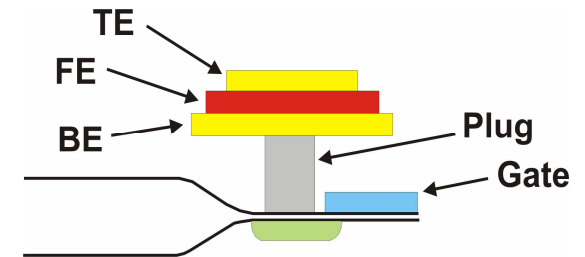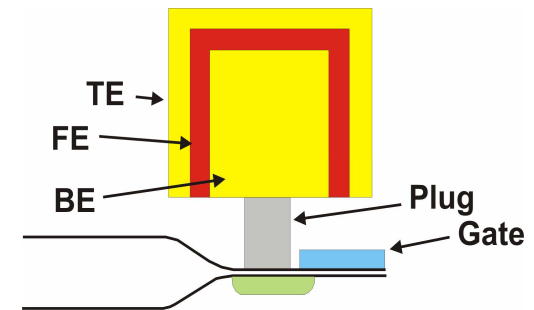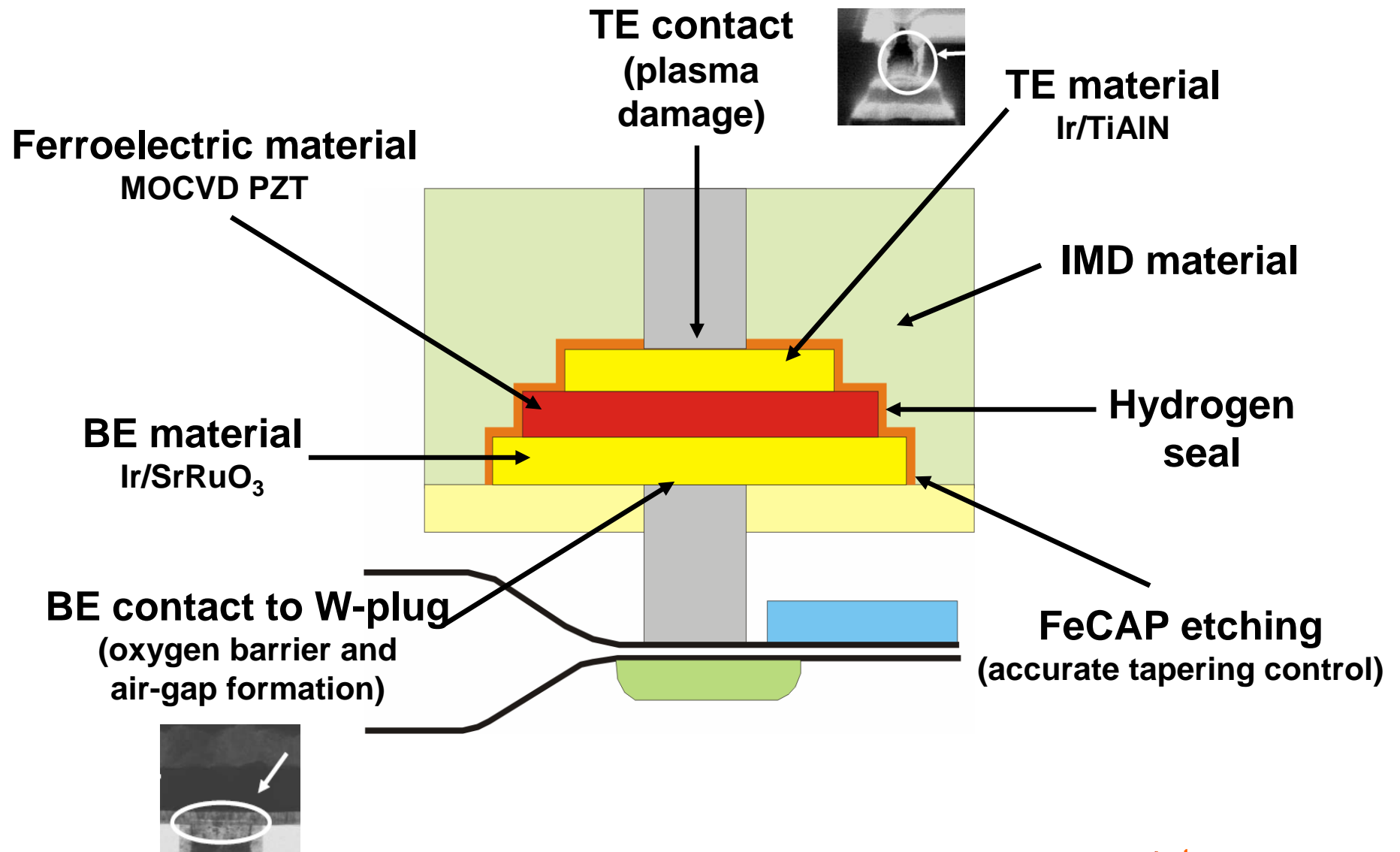
# FeRAM: Advantages and Issues

- Main advantages

  – Fast (<100ns) read and write operations with no intrinsic limitation (<100ps)

  – 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

numonyx™

# Magnetoresistive RAM (MRAM)

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
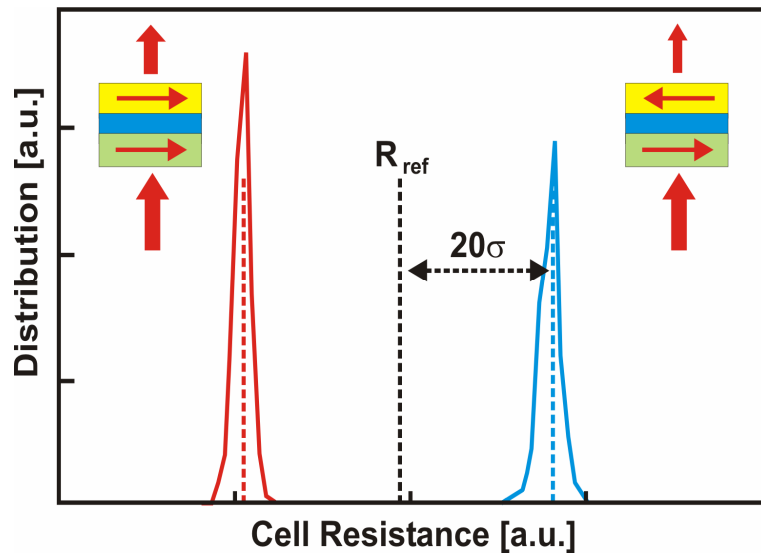


CoFe or NiFe/CoFe

$Al_2O_3$ or MgO

CoFe or NiFe/CoFe

NiMn or IrMn or CrPtMn



BL

Contact
Free layer
Tunnel barrier
Pinned layer
Pinning layer
Contact

WL

Sensing current

numonyx™
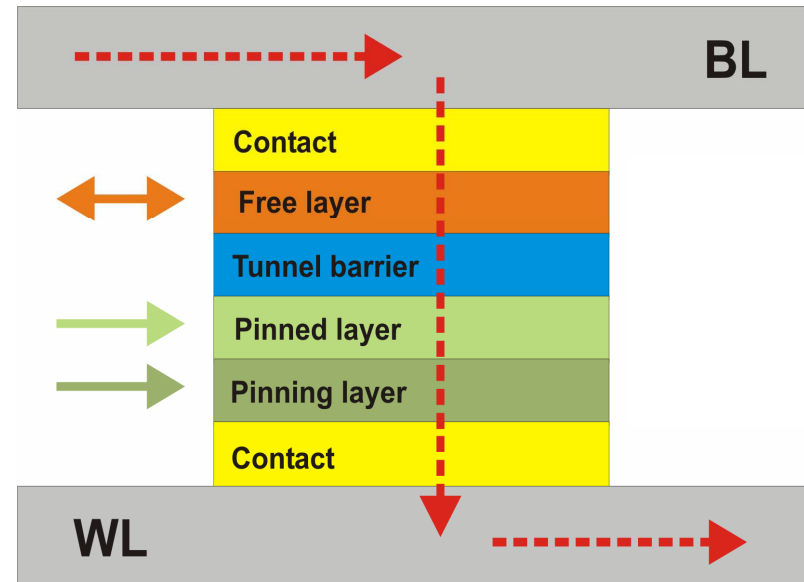
# MRAM Cell Architectures

**Cross-point architecture**



Write "1"

Bit lines

Word lines

Write "0"
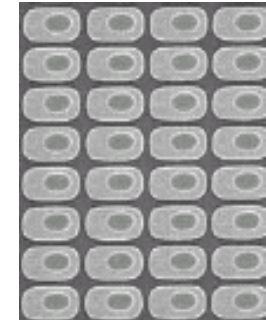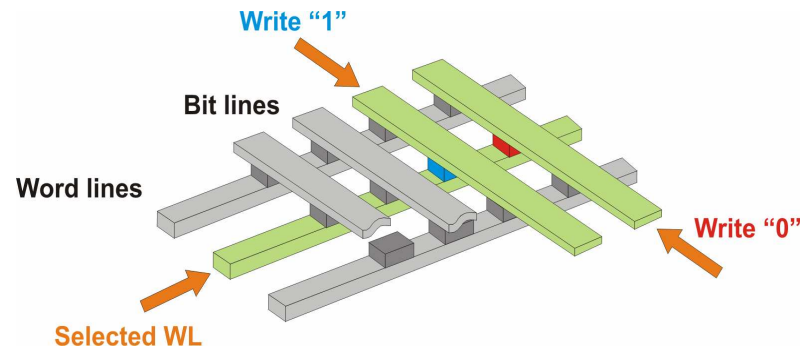
Selected WL



N. Sakimura *et al.*, ISSCC 2003

**MOSFET-selected architecture**



Program Mode

Bit Line

i

Magnetic Field

Free Layer

Tunnel Barrier

Fixed Layer

Digit Line

Flux concentrating cladding layer

Inlaid Copper interconnects

Isolation Transistor "OFF"



JA

MA  VA

M2

V1

M1

PC  CA
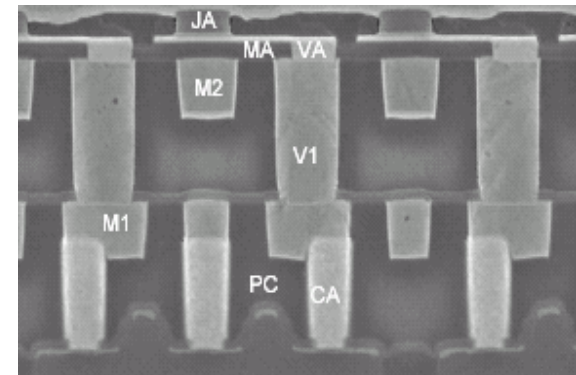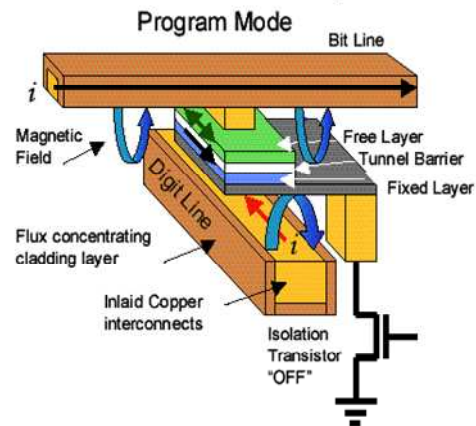
W. J. Gallagher, Taiwan NVM Workshop, 2005

numonyx™

# MRAM Scaling

- Thermal stability
  - Smaller cell volume ➔ lower <u>thermal stability</u>
  - Lower thermal stability ➔ high-coercivity materials/ non-isotropic cell shape
  - high-coercivity materials/ non-isotropic cell shape ➔ higher <u>programming current</u>

- 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**

numonyx™

# 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
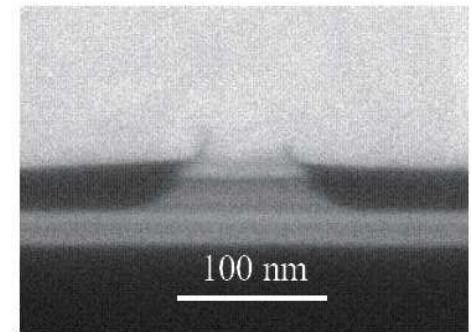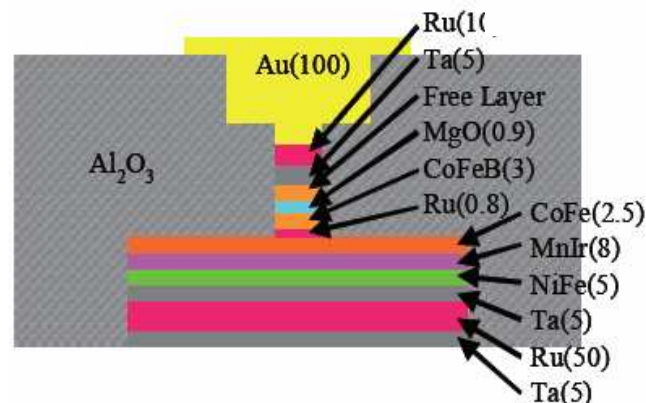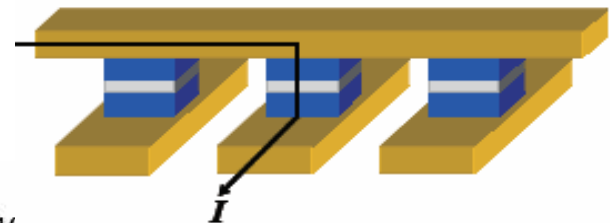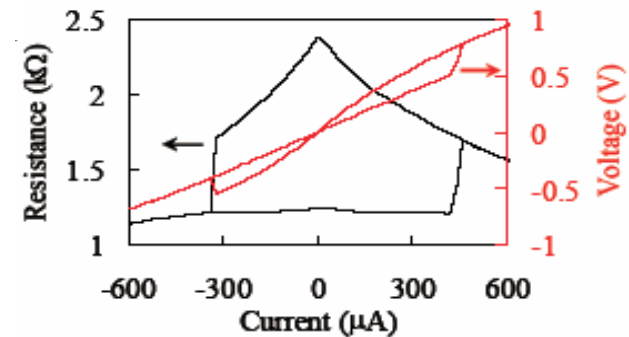


**K. Miura et al., VLSI Symp. on Tech. 2007**

# 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

numonyx™

# PCM Operating Principles - Set to Reset



Temperature

$T_m$

Time

SET
(low resistive)

RESET
(high resistive)

numonyx™
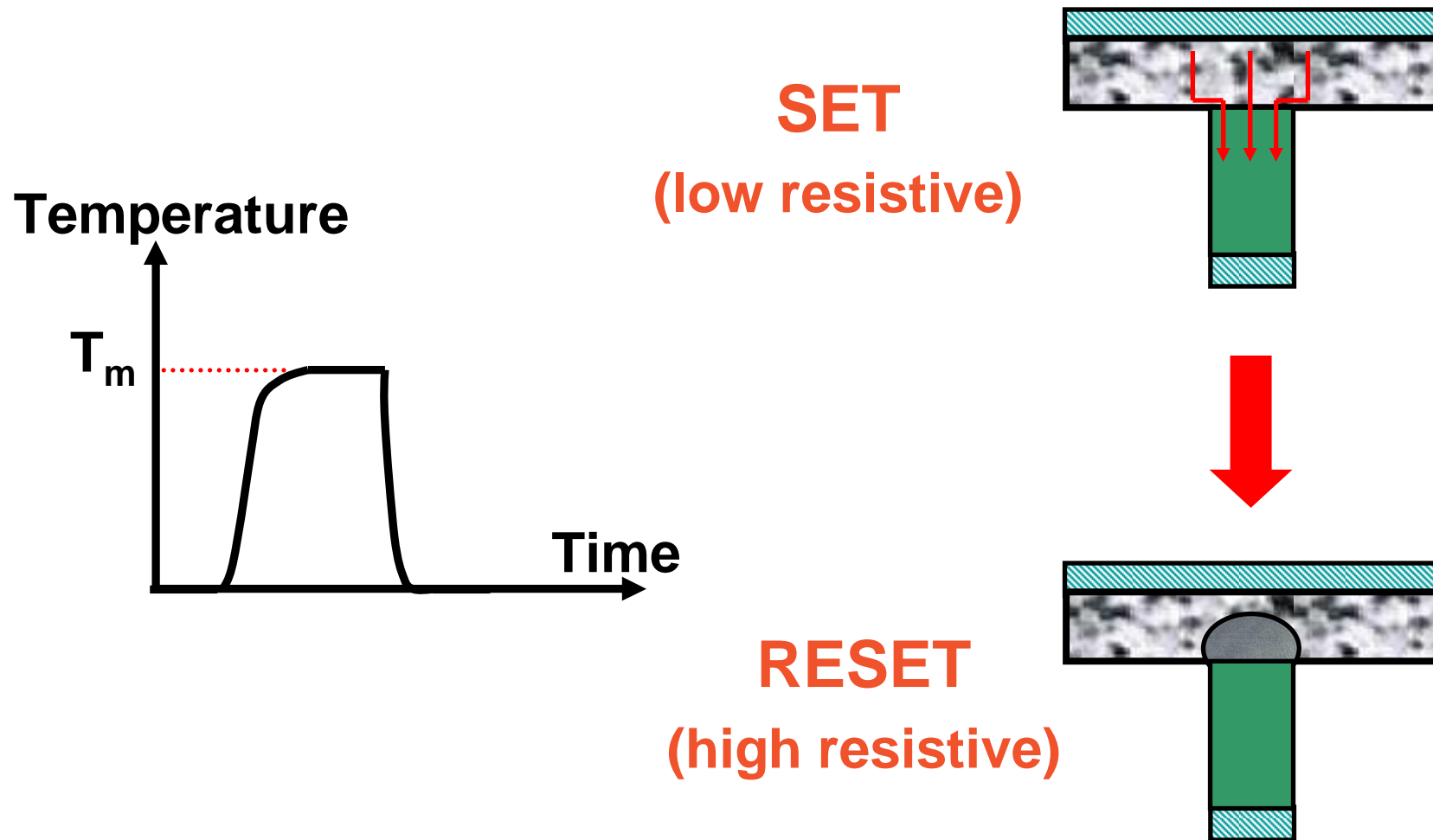
# PCM Operating Principles - Reset to Set



**Temperature**

$T_m$

$T_x$

**Time**

**RESET**
(high resistive)

**SET**
(low resistive)

**numonyx**™
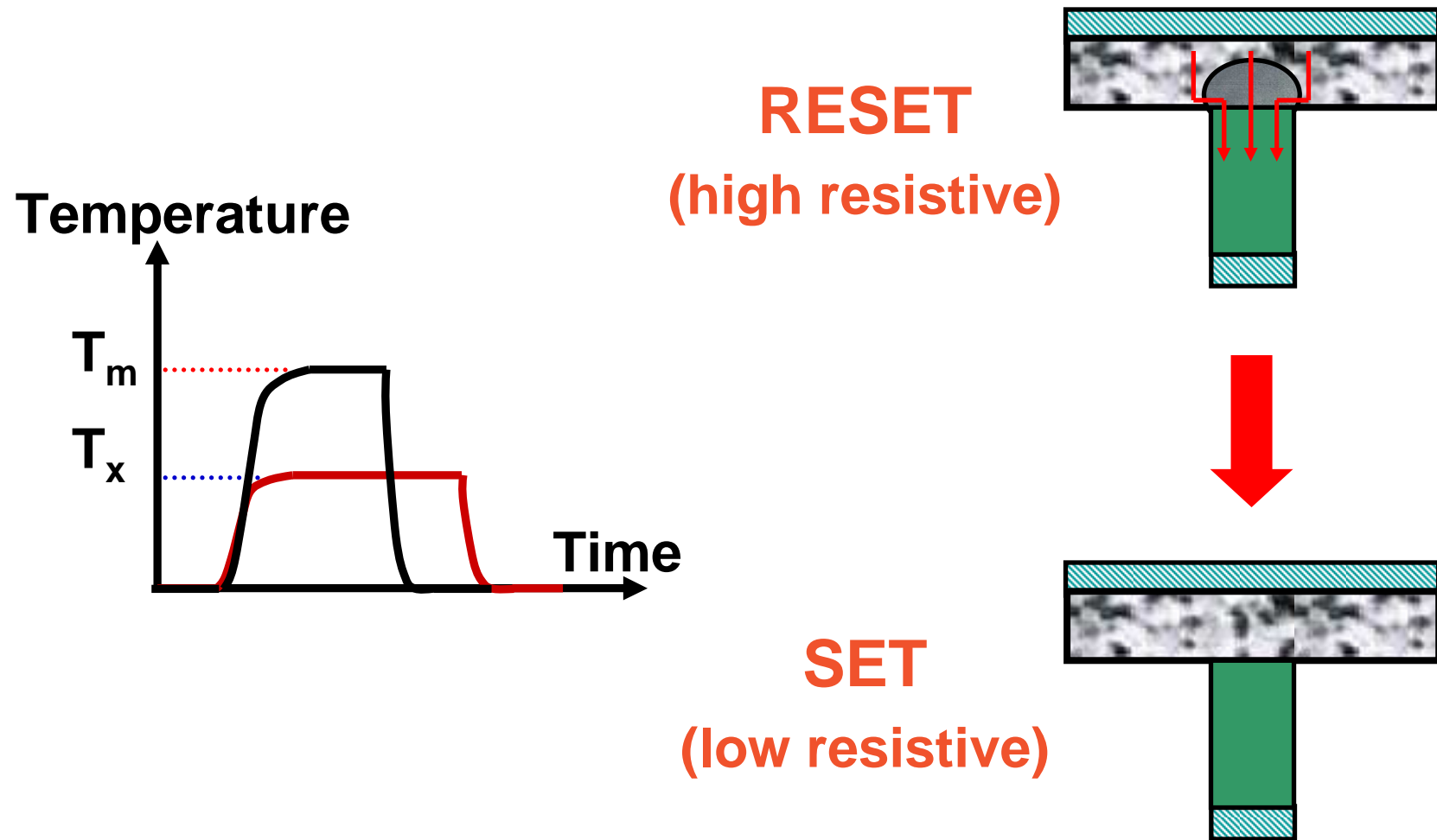
# From PCM Cell Architectures…

**Pore and lance structure**

**μtrench structure**
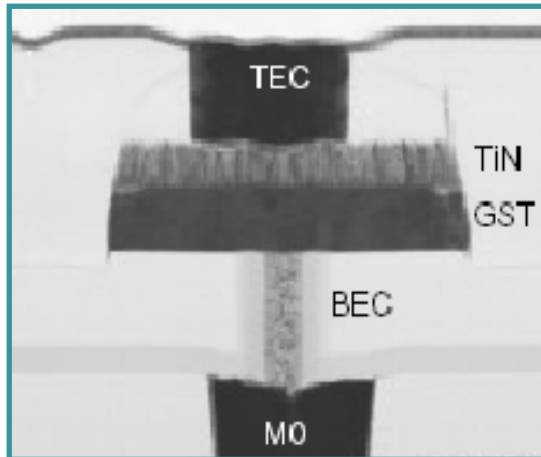
**Planar structure**



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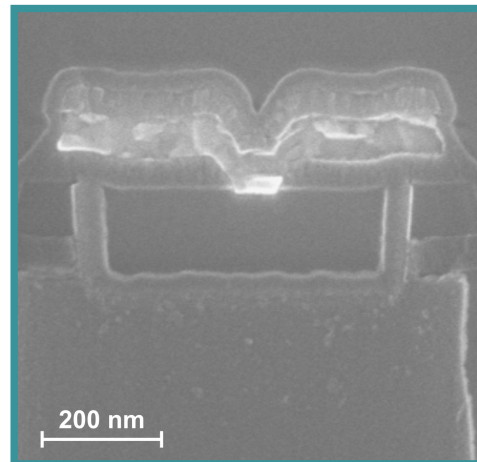


F. Pellizzer *et al.*,
VLSI Symp. on Tech. 2004
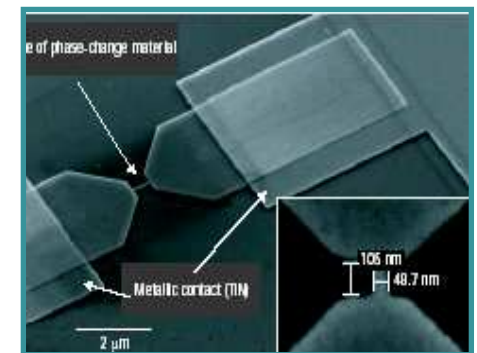


M. Lankhorst *et al.*,
*Nature Material, April 2005*

Efforts to reduce the programming current
still preserving a good controllability

**numonyx**™

# …To Technology Evolution

**8Mb Technology Demonstrator**

**128Mb Product**

**1Gb Product**

**Advanced Development**
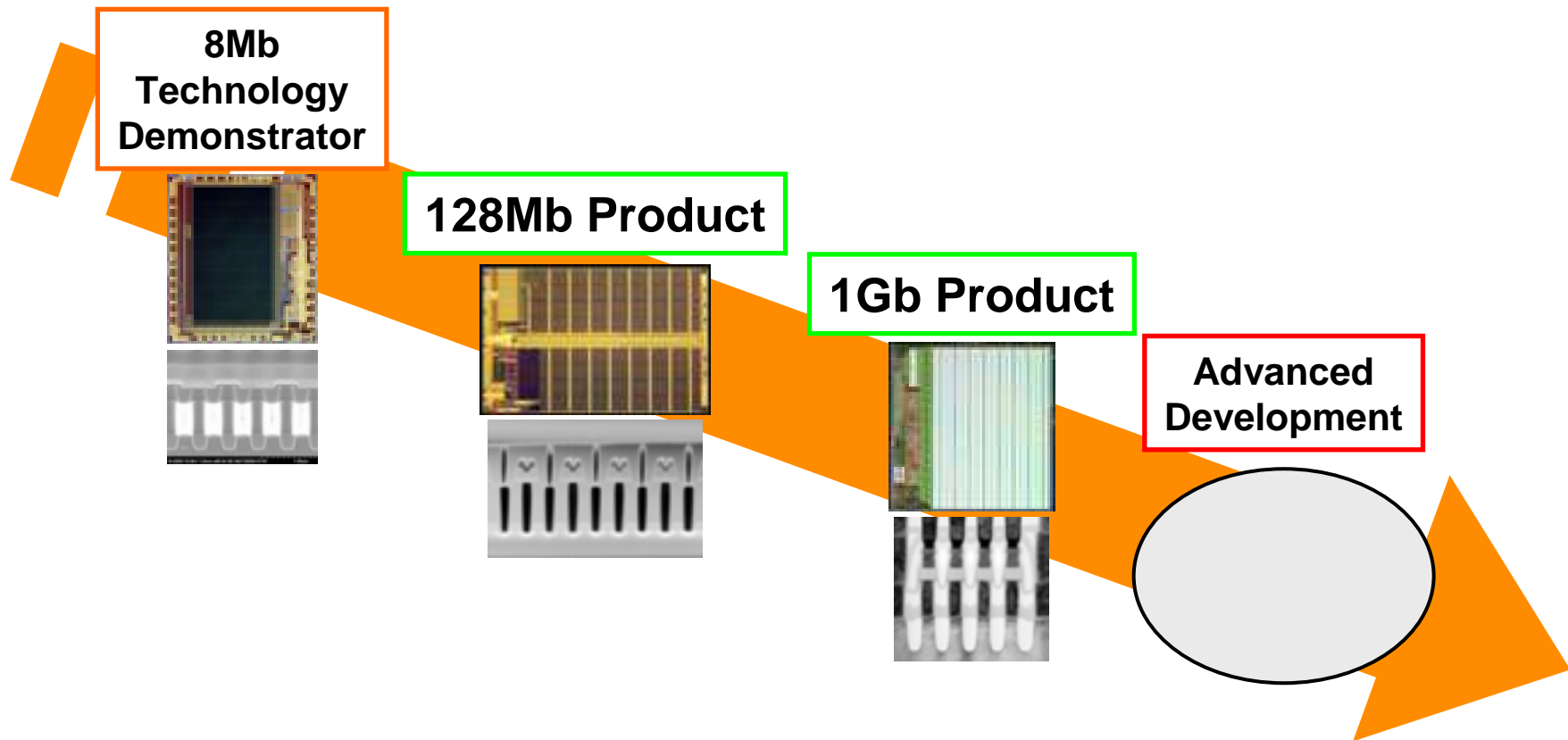
**Technology Node (nm)**

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

numonyx™

# Scaling of programming current
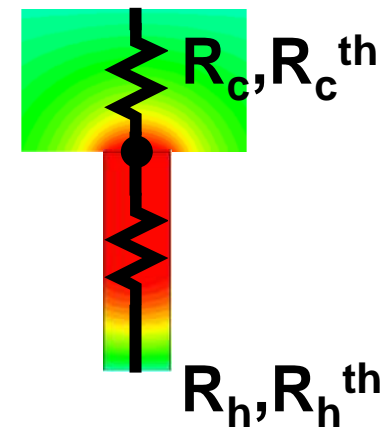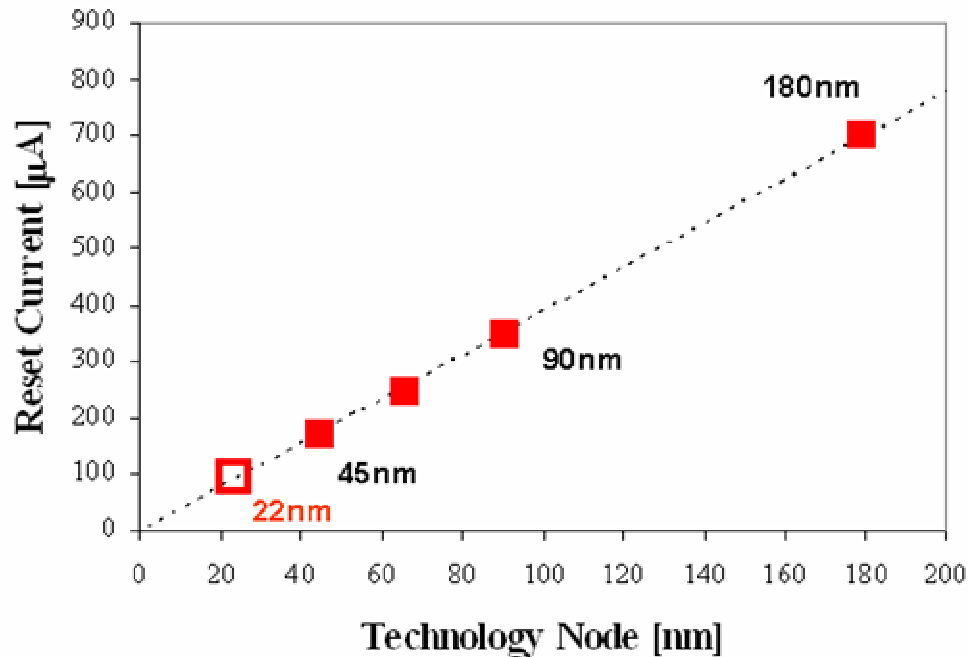


$$\Delta T_M = P_M \cdot R^{th}$$

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

# 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 $Ge_2Sb_2Te_5$ vs. Temp.**
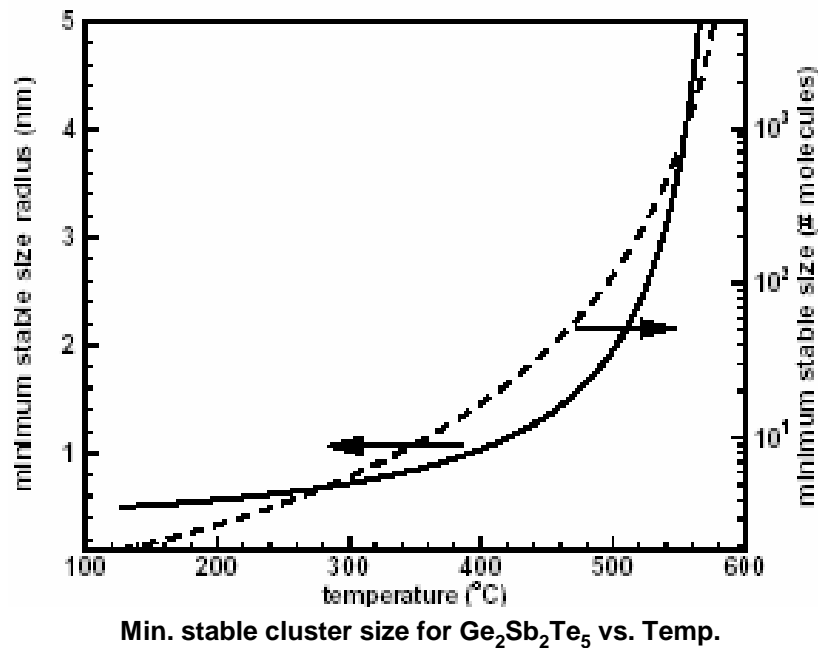
**D. Wright *et al.*, EPOCS 2004**



14nm pitch 3.3Tb/in$^2$
20nm pitch 1.6Tb/in$^2$
40nm pitch 0.4Tb/in$^2$
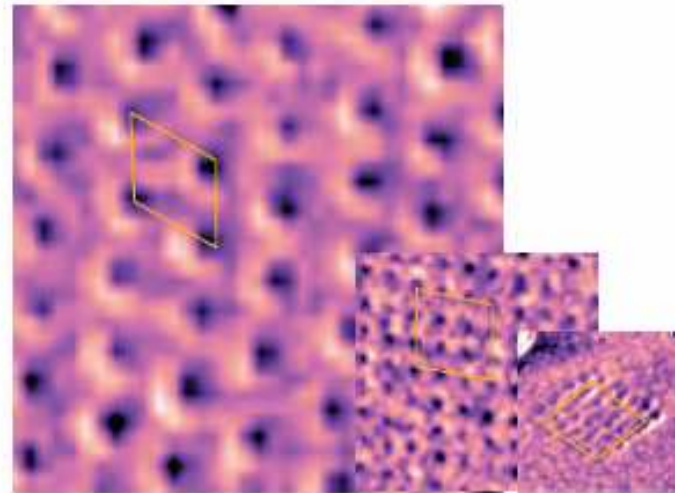
**Crystalline Bits in Amorphous Matrix.**

**C. Lam, SRC NVM Forum 2004**

**numonyx™**

# 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

numonyx™

# 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)



Set state

Reset state
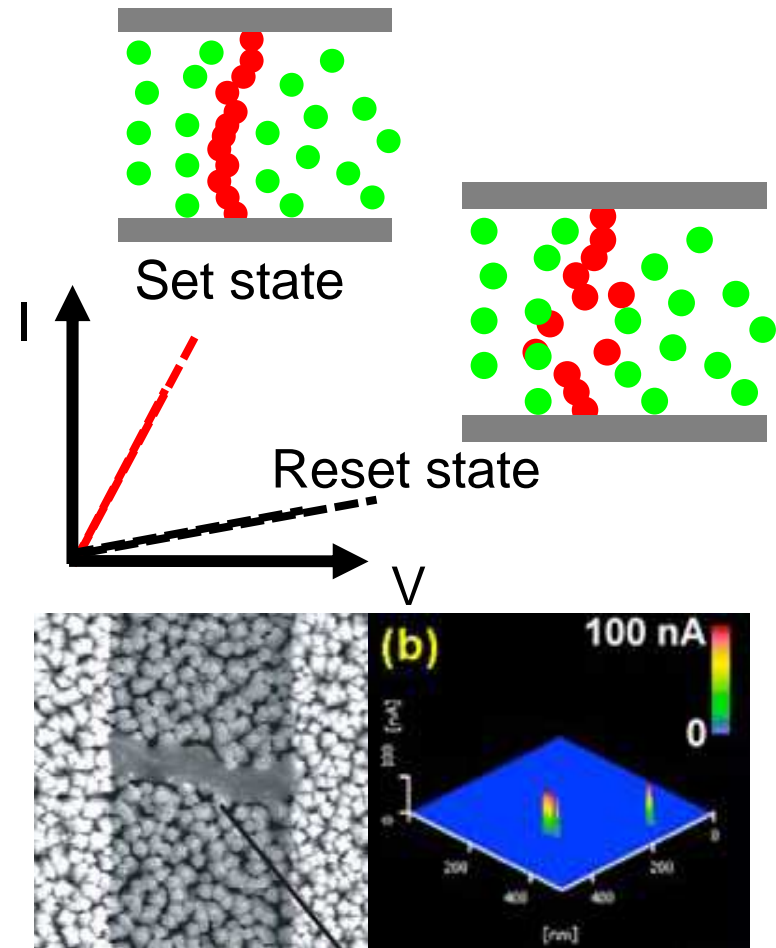
**A. Sawa, Materials Today 2008**

**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- $SrTiO_3$, $Pb(Zr_xTi_{1-x})O_3$ and $Pr_{0.7}Ca_{0.3.}MnO_3$
- ## Conductive polymers
  - Bengala Rose, $AlQ_3Ag$, Cu-TCNQ
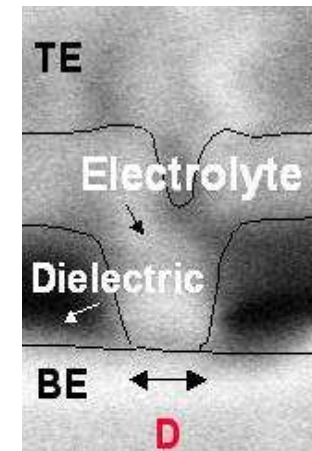


**M. Kozicki, EPCOS 2006**
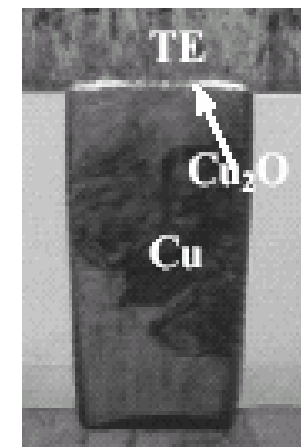


**A. Chen *et al.*,
IEDM Tech. Dig. 2005**

numonyx™

# Resistive Oxides Memories

- **Unipolar** resistive switching in a binary oxide layer
- Feasible small cell size (4-8 $F^2$), very low-cost solution
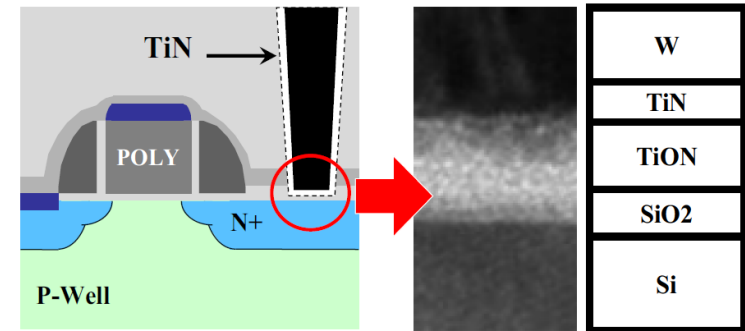


2Mbit 1T-1R 90nm test chip
**(Y.H. Tseng, *et al.*,
IEDM Tech. Dig. 2005)**



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

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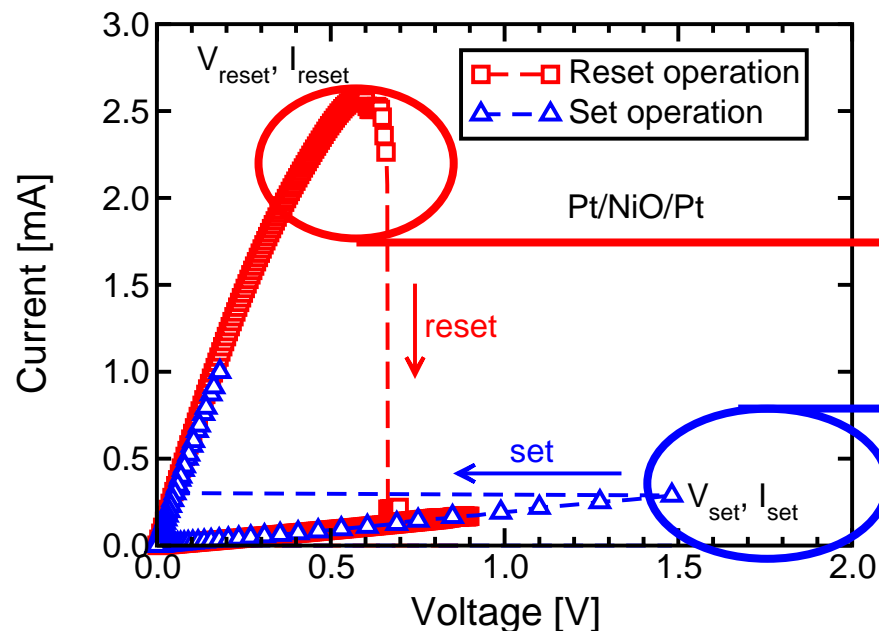**(D. Ielmini et al., APL 2009)**

# 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**



Oxidizable electrode

Metallic electrodeposit
**low resistance**

$M \rightarrow M^+ + e^-$

Glassy electrolyte
**high resistance**

$M^+ + e^- \rightarrow M$

Ion current

Inert electrode
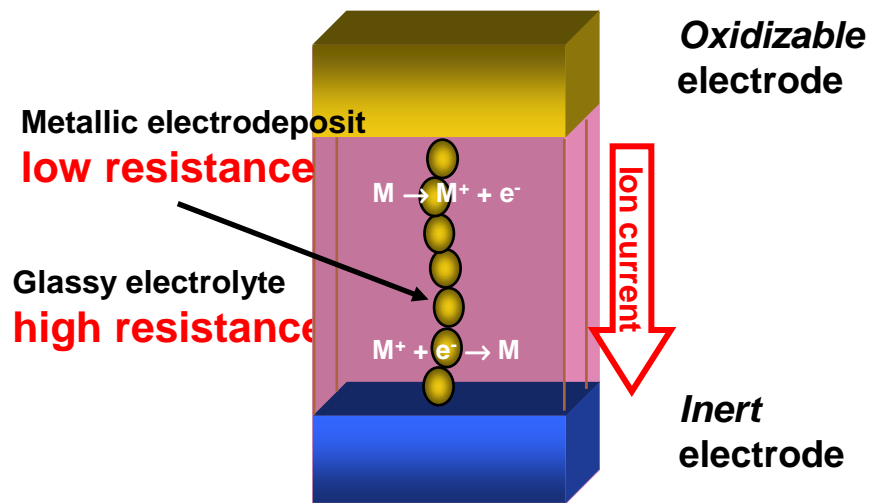
**M. Kozicki, EPCOS 2006**
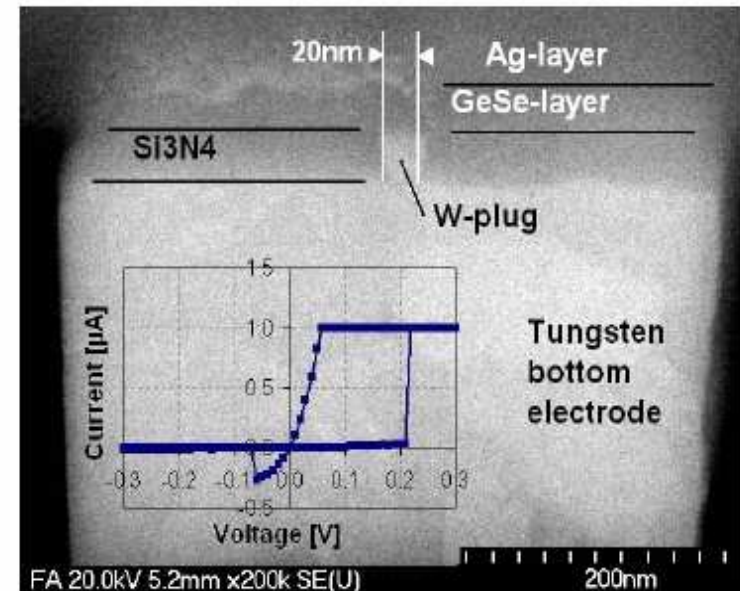


**M. Kund _et al._, IEDM Tech. Dig., 2005**

# 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)
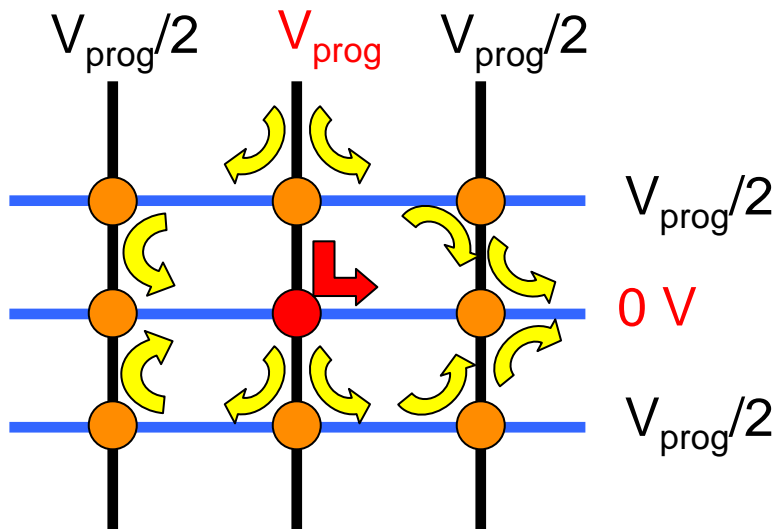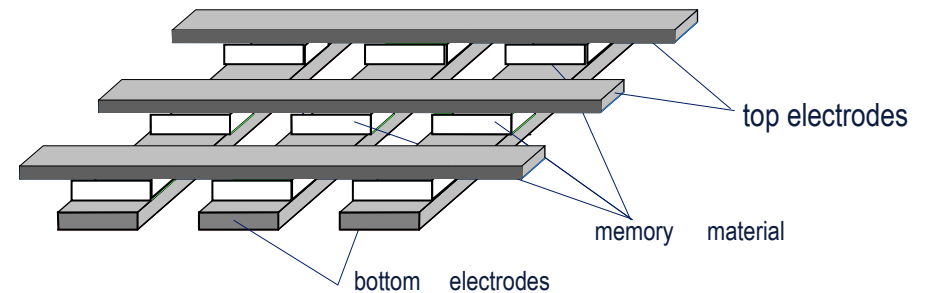
  – Retention capabilities for 10years at 85°C must be demonstrated

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

numonyx™

# The crossbar integration

- "simple" structure → low cost

- reduced cell size: $4F^2$

- suitable for 3D stacking → cell size $(4/n)F^2$



top electrodes

memory    material

bottom    electrodes



$V_{prog}/2$    $V_{prog}$    $V_{prog}/2$

$V_{prog}/2$

0 V

$V_{prog}/2$

- 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

numonyx™

# A wide range of material choises

### 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, $Nb_2O_5$, $Al_2O_3$, $Ta_2O_5$, $TiO_2$, $ZrO_x$, $Cu_xO$

- PCM → chalcogenide material



Bit lines

Word lines

Memory element

Switch element

numonyx™
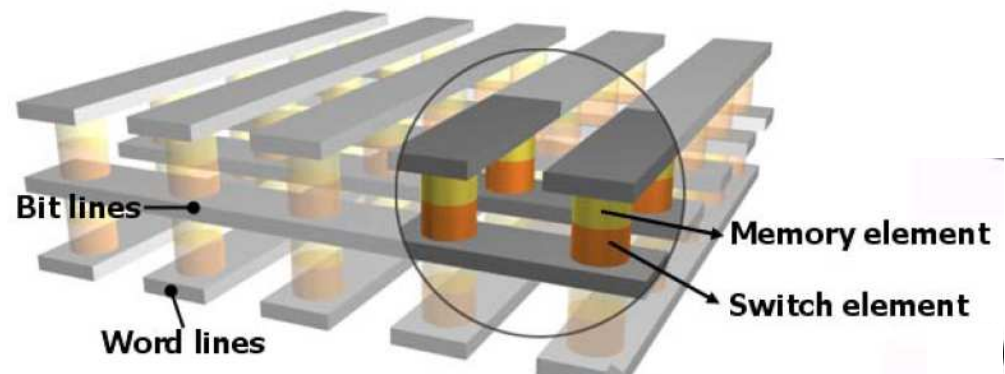
# Summary

- An impressive growth of the portable systems market has been enabled by the <u>NVM Flash technology</u> and it is expected that Flash will dominate the NVM production well <u>beyond the 30nm technology node</u>

- Semiconductor industries plan to exploit <u>new physical mechanisms and new materials</u> 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

numonyx™