Non-Volatile Memory Technology Overview
Ugo Russo, Andrea Redaelli, Roberto Bez

To cite this version:

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

Comming 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

DRAM - 25%/year price decrease
followed by similar price decline as DRAM - ~ 25%/year

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

Further price drop with 6F² transition

Flash ~ 4-8x cheaper

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

Fowler-Nordheim tunneling

Copyright © 2008 Numonyx B.V.
# NOR and NAND Stacked Gate Flash

<table>
<thead>
<tr>
<th></th>
<th>NOR</th>
<th>NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell size (F^2)</strong></td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td><strong>Read access</strong></td>
<td>Random</td>
<td>Serial</td>
</tr>
<tr>
<td></td>
<td>(fast ~50ns)</td>
<td></td>
</tr>
<tr>
<td><strong>Progr. Mechanism/ Throughput</strong></td>
<td>CHE / 0.5 MB/s</td>
<td>FN / 7-10MB/s</td>
</tr>
<tr>
<td><strong>SEM Cross-section (BL direction)</strong></td>
<td><img src="image1" alt="SEM Image" /></td>
<td><img src="image2" alt="SEM Image" /></td>
</tr>
</tbody>
</table>

![NOR NAND Schematic](schematic.png)

Copyright © 2008 Numonyx B.V.
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.

- 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
  - thicker tunnel oxide for retention
  - high-k (Al₂O₃) upper dielectric for capacitive coupling
  - metal (Tantalum) gate to prevent erase saturation
  - easy scalability, no capacitive coupling
  - 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**
- **MRAM**
- **MOx-RRAM**
- **Polymer FeRAM**
- **Polymer RRAM**
- **CNT**
- **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

Available electron number: 25,000

3D approach necessary!
CMOS Integration

- **Ferroelectric material**
  - MOCVD PZT

- **BE material**
  - Ir/SrRuO₃

- **BE contact to W-plug**
  - (oxygen barrier and air-gap formation)

- **TE contact**
  - (plasma damage)

- **TE material**
  - Ir/TiAlN

- **IMD material**

- **Hydrogen seal**

- **FeCAP etching**
  - (accurate tapering control)
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
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
\( \text{Al}_2\text{O}_3 \) or MgO
CoFe or NiFe/CoFe
NiMn or IrMn or CrPtMn

---

Copyright © 2008 Numonyx B.V.
MRAM Cell Architectures

Cross-point architecture

MOSFET-selected architecture

N. Sakimura et al., ISSCC 2003

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

1\textsuperscript{st} 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 \( \rightarrow \) 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

<table>
<thead>
<tr>
<th>Main advantages</th>
<th>Main issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast write (&lt;100 ns)</td>
<td>Difficult process integration</td>
</tr>
<tr>
<td>High write endurance</td>
<td>Large cell size vs. Flash and DRAM</td>
</tr>
<tr>
<td>Low voltage write</td>
<td>Large write current (&gt; 10 mA/B)</td>
</tr>
<tr>
<td></td>
<td>Small read signal</td>
</tr>
<tr>
<td></td>
<td>Scaling limits</td>
</tr>
</tbody>
</table>
PCM Operating Principles - Set to Reset

Temperature

\[ T_m \]

Time

SET

(low resistive)

RESET

(high resistive)
PCM Operating Principles - Reset to Set

RESET
(high resistive)

SET
(low resistive)
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
...To Technology Evolution

<table>
<thead>
<tr>
<th>Technology Node (nm)</th>
<th>180</th>
<th>90</th>
<th>45</th>
<th>32</th>
<th>2X</th>
</tr>
</thead>
</table>

Copyright © 2008 Numonyx B.V.
Scaling of programming current

\[ \Delta T_M = P_M \cdot R_{th} \]

\[ P_M = V_H \cdot I + \eta R \cdot I^2 \]
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$_2$Sb$_2$Te$_5$ 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$^2$

C. Lam, SRC NVM Forum 2004

Crystalline Bits in Amorphous Matrix.

Copyright © 2008 Numonyx B.V.
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
RRAM Proposed Alternatives

- **Chalcogenide**
  - GST and other phase-change alloys
  - AgGeSe, AgGeS, WO$_3$ and SiO$_2$ solid electrolyte

- **Binary oxide**
  - Nb$_2$O$_5$, Al$_2$O$_3$, Ta$_2$O$_5$, TiO$_2$, ZrO$_x$, Cu$_x$O and NiO

- **Oxides with perovskite structure**
  - SrZrO$_3$, doped SrTiO$_3$, Pb(Zr$_x$Ti$_{1-x}$)O$_3$ and Pr$_{0.7}$Ca$_{0.3}$MnO$_3$

- **Conductive polymers**
  - Bengal Rose, AlQ$_3$Ag, Cu-TCNQ
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

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 10 years 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 $\rightarrow$ polySi p/n junctions
- Heterojunctions $\rightarrow$ p-CuO/n-InZnO
- Schottky diode $\rightarrow$ Ag/n-ZnO
- Chalcogenide Ovonic Threshold Switching (OTS) materials

Storing device

- RRAM $\rightarrow$ NiO, Nb$_2$O$_5$, Al$_2$O$_3$, Ta$_2$O$_5$, TiO$_2$, ZrO$_x$, Cu$_x$O
- PCM $\rightarrow$ 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