Reliability issues of current and emerging NVMs

A. S. Spinelli, C. Monzio Compagnoni and D. Ielmini

Dip. di Elettronica e Informazione
Politecnico di Milano, Milano, Italy and IU.NET

alessandro.spinelli@polimi.it
Purpose

• Discuss the operation and some reliability issues of NVMs, with emphasis on
  – Floating-gate
  – Phase-change
  – Charge-trap

• Highlight the main problems and their current understanding, without going into many details
Acknowledgments

• A. L. Lacaita, C. Miccoli, A. Maconi, S. Amoroso, Politecnico di Milano

• A. Visconti, R. Bez and P. Tessariol, Micron, Agrate Brianza

• All authors whose works contributed to this tutorial
Outline

• Floating-gate technology
• Phase-change technology
• Charge-trap technology

• For each technology, the cell operation and reliability are discussed
  – Only single-cell mechanisms – no disturbs, interference, parameter spread,...
Outline - FG

- Flash memory cell concept and operation
- Flash reliability
  - Endurance
  - Retention (Detrapping and SILC)
  - Random telegraph noise
  - Charge injection statistics
Flash memory cell concept

- MOS device with a floating gate (FG)
- FG is electrically isolated by means of:
  - Tunnel oxide (bottom)
  - Interpoly dielectric (top)

- A charge $Q$ into the FG gives $\Delta V_T = -Q/C_{pp}$
- Stored charge does not leak from the FG $\Rightarrow$ NVM
Program/erase operation (concept)

- Electron charge is moved to/from the FG
- Verify levels PV and EV control the final $V_T$
Read operation (concept)

A reference voltage \( R \) is applied and the cell current is sensed
\( \textit{V}_T \) distributions

### NAND

- **Bit Distribution**
  - "1"
  - "0"

### NOR

- **Bit Distribution**
  - "1"
  - "0"

---

Voltage

- **Pass**
- **PV**
- **Ev**
- **DV**

---

"NAND" and "NOR" logic gates.
Multi-level cells

NAND

Bit Distribution

“11”
EV
R1
PV1
“10”
PV2
“00”
PV3
“01”
OP
PASS

NOR

Bit distribution

“11”
DV
EV
R1
PV1
“10”
PV2
“00”
PV3
“01”
R3
PV3
Outline - FG

• Flash memory cell concept and cell operation
• Flash reliability
  – Endurance
  – Retention (Detrapping and SILC)
  – Random telegraph noise
  – Charge injection statistics
Flash reliability

• A charge flow through the oxide is required for cell operation
• Very high oxide field is routinely applied during Flash P/E operations
• Flash devices work almost constantly under stress
• Oxide reliability is the key to a successful Flash technology
Endurance

From Lee et al., IRPS (2009)
NOR cell

From Lee et al., IRPS (2003)
NAND cell

- The $V_T$ levels move with cycling
- Asymmetric behavior of programmed and erased levels
1 – Oxide charge and FN tunneling

- FN tunneling generates traps in the oxide, mainly filled by electrons
- Electrons reduce the electric field and the tunneling current
- Less charge is transferred into the FG (even if CHE is used)

From Mielke, IRPS tutorial (2007)
A narrowing of the $V_T$ window is expected

From Shirota, IRPS tutorial (2005) - modified
2 – Oxide charge and $V_T$

- Trapped oxide charge shifts $V_T$ ($\Delta V_T = -Q/C$)
- Asymmetric effect results

From Shirota, IRPS tutorial (2005) - modified
3 – Interface states

From Lee et al., IRPS (2003)

- Interface states are generated after FN stress
- Device gm and SS are degraded $\Rightarrow V_T$ is further increased
Results

FN field reduction

Oxide charge

Interface states

All can be seen, depending on the leading mechanism
Non-uniform charge trapping at STI edges leads to different behavior for scaled cells

From Fayrushin et al., IEDM (2009)
Retention

• Retention = ability to maintain the stored information over time

• Intrinsic retention is limited by tunneling through the oxide (plus thermionic emission over the 3.1 eV barrier)
Intrinsic retention

An oxide thickness around 4.5 nm is enough for the intrinsic case.

From Ielmini et al., ME (2005)
Retention after cycling

- Main $V_T$ shift due to detrapping
- A small % of tail cells appear, losing charge much faster than the average
Detrapping

Program/erase

Retention
Detrapping in capacitors

From Scott et al., TED (1996)

From Spinelli et al., EDL (1999)

- A “transient SILC” is measured on capacitors after stress
- A $t^{-1}$ dependence is usually shown $\Rightarrow V_T \sim \ln(t)$
Detrapping features

• Dependence on cycling:
  – Increases with cycle number
  – Dependent on cycling pattern – see Mielke et al., IRPS (2006)

• High activation energy (1.1 eV) for retention (Ghidini et al, IRPS (2002)) and cycling (Mielke et al., IRPS (2006))

• Low field acceleration
Cycling pattern dependence

- Charge detrapping and/or trap annihilation take place between cycles
- Fast cycling is a worst-case condition
Relations

• Assuming that
  – Damage creation is not dependent on cycling pattern and temperature

• Then
  – Distributed cycling \((t_c, T_c)\) is equivalent to fast cycling + bake time \((t_b, T_b)\):

\[
t_b = t_0 + A t_c \exp \left[ E_A \left( \frac{1}{k_B T_b} - \frac{1}{k_B T_c} \right) \right]
\]

\[
\Delta V_T = \alpha \ln \left( 1 + \frac{t}{t_b} \right) = \alpha \text{ UDM}
\]
Comparison

From Monzio Compagnoni et al., IRPS (2010)

\[
t_b = t_0 + A t_c \exp \left[ E_A \left( \frac{1}{k_B T_b} - \frac{1}{k_B T_c} \right) \right]
\]
Universal damage metric

From Mielke et al., IRPS (2006)

\[ \Delta V_T = \alpha \ln \left( 1 + \frac{t}{t_b} \right) = \alpha UDM \]

From Monzio Compagnoni et al., IRPS (2010)

60nm technology
\[ p = 5 \times 10^{-5} \]

\[ t_0 = 0.8h \]
\[ A = 0.022 \]
\[ E_A = 0.52eV \]
SILC in capacitors

- Steady-state current following detrapping (or transient SILC)
- Modeled via trap-assisted tunneling (TAT) of carriers through oxide traps (Ielmini et al., TED (2000))
SILC in tail cells – 1

From Cappelletti et al., IEDM (2004)

From Spinelli, IRPS tutorial (2005)

Tail cells show a much larger leakage current than average (at the same stress level)
• A tunneling process assisted by two traps (2TAT) is held responsible for SILC in tail cells

• Successful models have been developed for SILC statistics (Ielmini et al., TED (2002) and Schuler et al., IRPS (2002))
Array distribution
Adapted from Ielmini et al., TED (2002)

Tail cells are those with an unlucky trap distribution, enhancing the leakage
SILC features

- Tail cells increase with cycle number
- Erratic behavior:
  - Cells can suddenly stop/start leaking (trap annihilation/reactivation)
  - Tail cells may be different from cycle to cycle
- Low activation energy (0.1 – 0.3 eV)
- Trap annealing above 85 °C
- Strong field acceleration
Over-erase in NOR cells
From Cappelletti et al., IEDM (2004)

- Some cells show enhanced tunneling at high fields \( \Rightarrow \) their \( V_T \) is lower than typical ones during erase
- Read errors may arise if \( V_T \) becomes \( < 0 \)
Origin of fast-erasing cells

- Positive charges trapped in the oxide
  - Three-charge cluster was assumed (Ong et al., VLSI (1993))
  - Generated by anode-hole injection (Chimenton et al., IRPS (2004))
- FG morphology
  - Barrier lowering/field enhancement (Muramatsu et al., IEDM (1994))
  - Field enhancement at polySi grain asperities (Nkansah et al., SSE (2000))

From Chimenton et al., PIEEE (2003)
Over-program in NAND cells

• Fast-program cells are expected in NAND arrays (opposite polarity with respect to NOR)
• Cells are programmed via ISPP ⇒ it is the $V_T$ fluctuation within a single step that matters
• If $V_T$ is programmed above $V_{pass}$, the cell will always remain OFF and the entire string can no longer be accessed (temporary failure)
Outline - FG

• Flash memory cell concept and cell operation
• Flash reliability
  – Endurance
  – Retention (Detrapping and SILC)
  – Random telegraph noise
  – Charge injection statistics
Random telegraph noise (RTN)

- Capture/emission of single electrons by oxide traps, with time constants $\tau_c$ and $\tau_e$
- Fluctuations in the drain current (or threshold voltage)
RTN in FG NVMs

• $\Delta V_T$ is the important parameter

• Due to the gate coupling coefficient $\alpha_G$, the RTN $V_T$ shift becomes (neglecting mobility fluctuations)

$$\Delta V_T = \frac{q}{C_{ox}WL\alpha_G}$$

• For $W = L = 45$ nm, $t_{ox} = 7$ nm, $\alpha_G = 0.65$ we have $\Delta V_T \approx 25$ mV
Anomalous RTN in FG NVMs

From Kurata et al., VLSI (2006)

\[ \frac{\Delta I_D}{I_D} \approx 15\% \text{ (90 nm technology)} \]

From Fantini et al., EDL (2007)

\[ \frac{\Delta I_D}{I_D} \approx 50\% \text{ (65 nm technology)} \]
Dynamic characterization of RTN

\[ \Delta V_T > 0 \]

\[ \Delta V_T = 0 \]

\[ \Delta V_T < 0 \]
Results

From Monzio Compagnoni et al., TED (2008)
Interpretation

From Muller et al., JAP (1996)

- Large conductance modulation is interpreted via a trap closing a percolation path
- Nonuniform conduction determined by fixed charge (and interface traps) distribution
Random dopant fluctuation and RTN

- Atomistic doping induces percolation paths
- Additional contribution to RTN amplitude

From Asenov et al., TED (2003)
Numerical simulation

From Ghetti et al., IRPS (2008)

- Exponential distribution confirmed for the single-trap RTN
- Multi-trap model in Monzio Compagnoni et al., TED (2008)
The drift phenomenon

From Monzio Compagnoni et al., TED (2008)

From Kurata et al., JSSC (2007)

From Fukuda et al., IEDM (2007)
• Slower and slower traps are observed as time elapses (RTN leading to 1/f noise)
• Increasingly large $\Delta V_T$ can be reached

From Spinelli et al., JJAP (2008)
Cycling dependence

From Kurata et al., JSSC (2007)

From Fukuda et al., IEDM (2007)
Effects on tail height and slope

From Monzio Compagnoni et al., EDL (2008)

60 nm NAND

65 nm NOR
RTN scaling trends

From Ghetti et al., TED (2009)

Scaling rule for the slope

\[ \lambda = \frac{K}{\alpha_G} \frac{t_{ox}^\alpha \sqrt{N_a}}{W \sqrt{L}} \]
Impact of STI edges


- Traps near STI edges have large impact on RTN
- Sharper STI edges degrade RTN
Electron injection statistics (EIS)

• During programming, an average number $\bar{n}$ of electrons is injected into the FG at each step

$$\bar{n} = \frac{J_{A} t_{step}}{q}$$

• Fluctuations in $\bar{n}$ (shot noise of the gate current) generate fluctuations in $\Delta V_T$
$\Delta V_T$ distribution

- Gaussian distribution of $\Delta V_T$
- Independent of cycling and temperature (Monzio Compagnoni et al., TED (2008))

From Monzio Compagnoni et al., TED (2010)
Experimental data

From Monzio Compagnoni et al., TED (2008) and TED (2010)

$\sigma_{\Delta V_T}$ saturates at high $\Delta V_T$

$\sigma_{\Delta V_T} \propto \sqrt{\Delta V_T}$ at low $\Delta V_T$
Low-$\Delta V_T$ regime

- Since $\Delta V_T = \frac{q n}{C_{pp}}$, we have $\sigma_{\Delta V_T} = \frac{q}{C_{pp}} \sigma_n$

- For small $n$, injection events are rare, hence not correlated $\Rightarrow n$ is ruled by a Poisson statistics, having $\sigma_n = \sqrt{n}$

$$\sigma_{\Delta V_T} = \frac{q}{C_{pp}} \sqrt{n} = \sqrt{\frac{q}{C_{pp}} \Delta V_T}$$
High-$\Delta V_T$ regime

- Electron injection modifies the FG potential, reducing the oxide field
- The tunneling current is reduced $\Rightarrow$ a correlation arises between the injection events
- The final result is $\sigma_{\Delta V_T} = \sqrt{\frac{q}{\gamma C_{pp}}}$ (Monzio Compagnoni et al., TED (2008))
Effect on program accuracy

EIS controls the ultimate limit to the $V_T$ accuracy

From Monzio Compagnoni et al., TED (2008)
Scaling trends

\[ \sigma_{\Delta V_T} [V] \]

From Monzio Compagnoni et al., TED (2008)

\[ C_{pp} \] is the main parameter driving the increase in \( \sigma_{\Delta V_T} \)
Outline

• Floating-gate technology
• Phase-change technology
• Charge-trap technology
Outline - PCM

• PCM cell concept and operation
• PCM reliability
  – Cycling wearout
  – Data retention
  – Resistance drift
PCM cell concept

Temperature

Time

T_m
PCM logic states

Set state

```
R = V/I

Ge$_2$Sb$_2$Te$_5$
```

![Diagram showing the relationship between voltage and current with two states: Set and Reset.](image)

Set state

![Image of GST material with labels for top and bottom contacts.](image)
Set/reset in PCM devices

- \( I < I_x \) \( \Rightarrow \) no change
- \( I_x < I < I_m \) \( \Rightarrow \) crystallization
- \( I > I_m \) \( \Rightarrow \) melting + quenching

From Redaelli et al., EDL (2004)
Outline - PCM

• PCM cell concept and operation

• PCM reliability
  – Cycling wearout
  – Data retention
  – Resistance drift
Cycling endurance

From Kim et al., IRPS (2005)
Stuck-set mechanism

From Rajendran et al., VLSI (2008)
Crystallization
Evidence for nucleation

From Lee et al., NL (2008)
Evidence for growth

From Shih et al., IEDM (2008)
From Redaelli et al., TED (2006)
**t_X extrapolation**

From Russo et al., TED (2007)

\[ \tau = \tau_{x0} e^{\frac{E_x}{kT}} \]

\[ E_x = 2.5 \text{ eV} \]

\[ \tau_{x0} = 10^{-23} \text{ s} \]
R drift – atomic relaxation

From Ielmini et al., TED (2009)

\[ R = R_0 \left( \frac{t}{t_0} \right)^\nu \]

\[ \tau = \tau_0 e^{\frac{E_A}{kT}} \]

\( T = 25^\circ C \)
Kinetic model for SR

From Ielmini et al., IEDM (2007)

Monomolecular dynamics:

\[ \frac{dN(t)}{dt} = - \frac{N_T}{(t/E_A)} \]

Distributed \( E_A \):

\[ N_T \]

\[ E_A \]

\[ \tau = \tau_0 e^{E_A/kT} \]
Why distributed $E_A$?

- Single $E_A$ level

\[ \tau = \tau_0 e^{E_A/kT} \]
Simulation results

From Lavizzari et al., TED (2009)
Temperature dependence

![Graph showing temperature dependence of resistance](image)

- Power-law extrapolations
- Crystallization

Legend:
- 90°C
- 110°C
- 130°C
- 170°C
- 180°C
Arrhenius plot

From Ielmini et al., APL (2009)
Interpretation: Meyer-Neldel rule

- Arrhenius law with distributed $E_A$
  \[ \tau_{SR} = \tau_0 \exp\left( \frac{E_A}{kT} \right) \]

- Crossing at $T_{MN} \Rightarrow$ pre-exponential $\tau_0$ is not constant, but is an exponential function of $E_A$
  \[ \Rightarrow \text{Meyer-Neldel (MN) rule (Yelon et al., PRB (1992))} \]
  \[ \tau_0 = \tau_{00} \exp\left( -\frac{E_A}{kT_{MN}} \right) N \]
SR and crystallization are described by the same Arrhenius + Meyer-Neldel rules.

\[ \tau = \tau_0 e^{\frac{E_A}{kT}} \]

\[ \tau_0 = \tau_{00} e^{\frac{E_A}{kT_{MN}}} \]

From Ielmini et al., APL (2009)
T-acceleration model

From Ielmini et al., ME (2009)

To normalize SR time to a given T:

\[ t_1 = \tau_0 \left( 1 - \beta \frac{T_2}{T_1} \right) \frac{T_2}{T_1} \]

\[ \beta = \frac{T_{MN} - T_1}{T_{MN} - T_2} \]
Outline

• Floating-gate technology
• Phase-change technology
• Charge-trap technology
Outline - CTM

• CTM cell concept and operation
• CTM reliability
  – Endurance
  – Retention (Detrapping and SILC)
  – Random telegraph noise
  – Charge injection statistics
CT memory cell concept

- Still a charge-based approach – same concept and operation as FG technology
- The polysilicon FG is replaced by a defective dielectric (usually silicon nitride)

FN tunneling is used for P/E $\Rightarrow$ NAND technology
Erase saturation in SONOS cells

Electron injection from CG limits the $V_T$ window
• Scaling of tunnel oxide leads to DT and holes injection, increasing $\Delta V_T$
• Retention is compromised
From SONOS to TANOS

From Lee et al., IEDM (2003)
CTM reliability

• Same issues as FG memories (endurance, retention, RTN,...)
• Understanding complicated by nitride and blocking dielectric layers
• Contradictory results exist in the literature (dependence on technology, cell size,...)

• Only standard TANOS are discussed for brevity
Endurance – capacitors

From Van den Bosch, IMW (2009)

In large-area capacitors, $\Delta V_T$ is mainly controlled by interface states
Endurance – cells

In scaled cells, a behavior similar to FG ones is obtained ⇒ same physical mechanisms are involved (i.e., edge effects)

From Lee et al., NVSMW (2006)
Retention

• **Intrinsic** retention is still the issue
• Both oxide, nitride and alumina affect CT retention
  – Thin oxide layer (4 – 5 nm)
  – Trappy nitride layer (1.2 – 1.8 eV from CB; actual values can vary)
  – Trappy alumina layer
A few mechanisms
Activation energy – 1

From Kim et al., DRC (2010)

- High activation energy (1 – 2 eV)
- Dependent on T range, but unusual (higher for low T)
Activation energy – 2

• Samples with 3.5/3, 5, 10/16 nm are investigated
• Low activation energy (≈ 0.4 – 0.6 eV)

From Bocquet et al., IMW (2009)
Role of nitride layer

- Thin nitride layer degrades retention ⇒ faster leakage of charge at the nitride edge
- Small dependence on oxide layer thickness ⇒ main leakage path is through alumina

From Melde et al., NVSMW (2008)

2 hours at 150°C

Different oxide thickness

Different nitride thickness
Role of alumina layer

Main leakage path is through alumina

From Amoroso et al., IRPS (2010)
Tail cells

- Tail cells have been observed also in CTM
- Still interpreted as TAT through oxide traps

From Lue et al., NVMTS (2009)
RTN

From Gu et al., IEDM (2006)

RTN is a reliability issue also in CTM
RTN and nitride charges

From Chiu et al., IEDM (2009)

Localized charge affect RTN amplitude, changing the percolation paths
Random dopant fluctuations remain the main responsible for RTN in CTM

From Monzio Compagnoni et al., TED (2010)
A Poissonian behavior for charge injection is recovered in CTM

From Lue et al., IMW (2010)
Conclusions – 1

• FG reliability is determined by different physical mechanisms, more and more connected in huge, scaled arrays

• Scaling brings on ever new phenomena, challenging our understanding of the “basic issues” and raising new ones

• Moreover...
Conclusions – 2

• Emerging memories raise totally new issues, related to their unique operating characteristics

• A good comprehension of the “reliability physics” of NVMs is needed to understand the nature of the new failure mechanisms and learn how to live with them
References – FG

- R. Shirota, IRPS tutorial (2005)
- N. Mielke, IRPS tutorial (2007)
- D. Ielmini et al., Microel. Eng. 80, 321 (2005)
- A. S. Spinelli et al., IRPS tutorial (2005)
- R.-V. Wang et al., TED, 2107 (2009)
- A. Ghetti et al., TED, 1746 (2009)
References – PCM and CTM

• K. Kim et al., Proc IRPS, 157 (2005)
• B. Rajendran et al., VLSI Tech. Dig., 96 (2008)
• S.-H. Lee et al., Nano Lett. 8, 3303 (2008)
• D. Ielmini et al., IEDM Tech. Dig., 939 (2007)
• D. Ielmini et al., Microelectron. Eng. 86, 1942 (2009)
• G. Van den Bosch, Proc. IMW (2009)
• M. Bocquet at al., Proc. IMW, (2009)
• S. Amoroso et al., Proc. IRPS, 966 (2010)
• H. T. Lue et al., Proc. NVMTS, 58 (2009)
• S. H. Gu et al., IEDM Tech. Dig., 487 (2006)
• J. P. Chiu et al., IEDM Tech. Dig., 843 (2009)
• H. T. Lue et al., Proc. IMW, 92 (2010)