CMOS-Compatible Logic Embedded High-K Charge-Trap Multi-Time-Programmable Memory

Janakiraman Viraraghavan, Derek Leu, Balaji Jayaraman, Alberto Cestero, Ming Yin, John Golz, Rajesh R. Tummuru, Ramesh Raghavan, Dan Moy, Thejas Kempanna, Faraz Khan, Toshiaki Kirihata, Subramanian Iyer
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm Bulk FinFET
• Future scope
• Summary
Introduction – Computer Architecture

CPU
- Cache Memory
  - L1: SRAM
  - L2: SRAM/eDRAM
  - L3: eDRAM
- Chip ID: OTPM

Ext. Cache Memory
- L4: eDRAM

Memory
- DRAM
- Firmware (BIOS) ROM
- OTPM / MTPM

Input / Output Controllers

Hard Disk
- NVRAM
- CDROM
- TAPE

Keyboard
- Mouse
- Monitor
- Printer

Network
Introduction – Computer Architecture

CPU
- Cache Memory
  - L1: SRAM
  - L2: SRAM/eDRAM
  - L3: eDRAM
- Chip ID: OTPM

Ext. Cache Memory
- L4: eDRAM

Memory
- DRAM
- Firmware (BIOS)
- ROM
- OTPM / MTPM

Input / Output Controllers

Hard Disk
- NVRAM
- CDROM
- TAPE

Keyboard
- Mouse
- Monitor
- Printer

Network
Some integrations are beneficial for mobile applications. Low cost with process & test adders??

- Why not interposer or 3D solution?
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Emerging Memory (MRAM)
- Flash (SSD/Media)

- Small System Space
- New Applications
- Embedded System
- Firmware

- Improve Performance
- Reduce Power
- Improve Security
- OTPM ID, Redundancies

L3 cache
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Embedded System
- L3 cache
- Emerging Memory (MRAM)
- OTPM ID, Redundancies
- Flash (SSD/Media)
- Firmware

- Improve Performance
- Reduce Power
- Improve Security
- Small System Space
- New Applications
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Embedded System
- Emerging Memory (MRAM)
- Flash (SSD/Media)
- Firmware

- Improve Performance
- Reduce Power
- Improve Security
- Small System Space
- New Applications
- OTPM ID, Redundancies
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Embedded System
- Emerging Memory (MRAM)
- Flash (SSD/Media)

- Improve Performance
- Reduce Power
- Improve Security
- Small System Space
- New Applications

Firmware
MTPM – Rewritable
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Embedded System
- Emerging Memory (MRAM)
- Flash (SSD/Media)

- Improve Performance
- Reduce Power
- Improve Security
- Small System Space
- New Applications

OTPM ID, Redundancies
L3 cache
Embedded Memories

- DRAM (Main Memory)
- SRAM Cache
- Embedded System
- Emerging Memory (MRAM)
- L3 cache
- Improving System Space
- New Applications
- Improve Performance
- Reduce Power
- Improve Security
- OTPM ID, Redundancies
- Flash (SSD/Media)
- Firmware MTPM - Rewritable
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm Bulk FinFET
• Future scope
• Summary
Embedded Nonvolatile Memories

- Integrating Firmware on chip is beneficial
  - Cost reduction $\rightarrow$ If done with zero-mask adder to logic process.
  - Enhanced security.

<table>
<thead>
<tr>
<th>Type</th>
<th>Technique</th>
<th>Density</th>
<th>Re-writability</th>
<th>ID</th>
<th>Security</th>
<th>2D Red.</th>
<th>3D Red.</th>
<th>Firmware</th>
<th>Data</th>
<th>Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTPM</td>
<td>eFUSE</td>
<td>Low</td>
<td>One Time</td>
<td>●</td>
<td>▲</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Dense OTPM</td>
<td>Anti-fuse</td>
<td>Medium</td>
<td>Emulation</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>● (green)</td>
<td>●</td>
<td>▲</td>
<td>●</td>
</tr>
<tr>
<td>MTPM</td>
<td>Charge Trap</td>
<td>Medium</td>
<td>Medium</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>● (green)</td>
<td>●</td>
<td>▲</td>
<td>●</td>
</tr>
<tr>
<td>Flash</td>
<td>FG</td>
<td>High</td>
<td>High</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>● (green)</td>
<td>●</td>
<td>▲</td>
<td>●</td>
</tr>
<tr>
<td>MRAM</td>
<td>Magnetic</td>
<td>High</td>
<td>High</td>
<td>●</td>
<td>●</td>
<td>● (yellow)</td>
<td>●</td>
<td>●</td>
<td>● (green)</td>
<td>●</td>
</tr>
</tbody>
</table>

- Ideally suited
- Fairly suited
- Somewhat suited
- Not suited
eNVM For Redundancy

3D memory requires significantly more OTPMs

P. Klim et. al, VLSI 2008

4096b Fuse Cell Core (64 WL x 64BL)

eNVM for Automotive application

- 4MB code flash + 64KB data flash integrated on-chip.
- Targeted for automotive MCUs for engine control and driver assistance applications.

Y. Taito et al., ISSCC 2015

<table>
<thead>
<tr>
<th></th>
<th>Code Flash</th>
<th>Data Flash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>28nm SG-MONOS</td>
<td></td>
</tr>
<tr>
<td>Memory Cell Size</td>
<td>0.053μm²</td>
<td></td>
</tr>
<tr>
<td>Memory Capacity</td>
<td>2MB X 2</td>
<td>64KB</td>
</tr>
<tr>
<td>Power Supply</td>
<td>Core(VDD) 1.1V ±0.1V, I/O(VCC) 2.7-5.5V</td>
<td></td>
</tr>
<tr>
<td>Operating Temp.</td>
<td>-40~170°C (Tj)</td>
<td></td>
</tr>
<tr>
<td>Read I/O number</td>
<td>(128bit + 10bit) X 2</td>
<td>32bit + 7bit</td>
</tr>
<tr>
<td>Random Read Freq.</td>
<td>200MHz</td>
<td>10MHz</td>
</tr>
<tr>
<td>Program Speed</td>
<td>2.0MB/s</td>
<td>150us/4B</td>
</tr>
<tr>
<td>Erase Speed</td>
<td>0.91MB/s</td>
<td>1.5ms/64B</td>
</tr>
<tr>
<td>P/E endurance</td>
<td>10k cycles</td>
<td>&gt;1M cycles</td>
</tr>
<tr>
<td>Maximum Capacity</td>
<td>32MB</td>
<td>512KB</td>
</tr>
</tbody>
</table>
eNVM for Media application

Embedded Flash memory in 0.5um CMOS for voice-storage application:

- 32Mb, 4-level embedded flash to store 64 minutes of voice.
- On-chip Memory BIST is included.

M. Borgatti et al., IEEE JSSC 2001
eNVM for other applications

STT-MRAM based cache memory in 65nm node,
H. Noguchi et al., ISSCC 2016

Memory cell: 2T 2MTJ

Nonvolatile logic-in-memory array processor using MTJ/MOS in 90nm node,
M. Natsui et al., ISSCC 2013
This talk focuses on a CTT based eNVM targeted for:

- Secure OTPM ID
- 2D and 3D Redundancies
- Firmware applications
Comparison of Embedded NVRAM Solutions

- **Floating Gate**
  - Dual poly [1]
  - Process Overhead
  - High Voltage

- **BEOL NVM**
  - ReRAM [3]
  - STT MRAM[4]
  - Process Overhead

- **OTPM**
  - eFuse [5]
  - Anti Fuse [6]
  - One Time Programmable

- **CTT** [This Work]
  - CHE in HiK HfO₂
  - Voltage ~ 1.5-2V

**Charge Trap NVM**

- **Split Gate SGMONOS** [2]
  - CHE in Si₃N₄ in Split Gate
  - Voltage ~ 7-10V

- **OTP from NSCore** [9]
  - CHE in Si₃N₄ Side wall spacer
  - Voltage ~ 5-7V

CTT – Dense, No mask adder, Bulk/ SOI FIN scalable, logic voltage compatible MTPM
CTT MTPM Applications

- Embedded Flash
  - Additional process
  - FIN: Very difficult

- CTT MTPM
- Anti-Fuse (MTP Emulation)

- External Flash

- # Writes
- 1K

- OTP

- ID
- Repair
- Code
- Data

- 1KB
- 1GB
- 1MB

- 2D
- 3D

- Few Multi-Writes
- Multi-Writes

- >> 1M Rewrites
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm Bulk FinFET
• Future scope
• Summary
Charge Trap Transistor (CTT) Technology

Charge trapping mechanism:
Intrinsic Oxygen vacancies in HfO$_2$ [7]

Intrinsic Oxygen vacancies in HfO$_2$ HiK dielectrics

Charge Trapping done at logic process compatible voltages
OTPM / MTPM Twin Cell

MTPM Twin Cell

WL

<table>
<thead>
<tr>
<th>e^-</th>
<th>+DVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLt</td>
<td>SL</td>
</tr>
</tbody>
</table>

0.108$\text{mm}^2$ in 32nm node

6T SRAM 0.169$\text{mm}^2$ in 32nm node

1T1C DRAM 0.039$\text{mm}^2$

in 32nm node
### 1/4 Standby – Initial State Before Programming

<table>
<thead>
<tr>
<th>Mode</th>
<th>WL</th>
<th>BL</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standby</td>
<td>0V</td>
<td>float</td>
<td>0V</td>
</tr>
<tr>
<td>Program</td>
<td>2V</td>
<td>0V</td>
<td>~1.5V</td>
</tr>
<tr>
<td>Read</td>
<td>1V</td>
<td>Signal</td>
<td>1V</td>
</tr>
<tr>
<td>Erase</td>
<td>-1V</td>
<td>float</td>
<td>~2V</td>
</tr>
</tbody>
</table>

![Diagram showing WL, BLt, SL, BLc connections]
## 2/4 Program

<table>
<thead>
<tr>
<th>Mode</th>
<th>WL</th>
<th>BL</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standby</td>
<td>0V</td>
<td>float</td>
<td>0V</td>
</tr>
<tr>
<td>Program</td>
<td>2V</td>
<td>0V</td>
<td>~1.5V</td>
</tr>
<tr>
<td>Read</td>
<td>1V</td>
<td>Signal</td>
<td>1V</td>
</tr>
<tr>
<td>Erase</td>
<td>-1V</td>
<td>float</td>
<td>~2V</td>
</tr>
</tbody>
</table>

### WL=0V

### WL=2V

### BLt=0V SL=1.5V BLc
3/4 Read

<table>
<thead>
<tr>
<th>Mode</th>
<th>WL</th>
<th>BL</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standby</td>
<td>0V</td>
<td>float</td>
<td>0V</td>
</tr>
<tr>
<td>Program</td>
<td>2V</td>
<td>0V</td>
<td>~1.5V</td>
</tr>
<tr>
<td>Read</td>
<td>1V</td>
<td>Signal</td>
<td>1V</td>
</tr>
<tr>
<td>Erase</td>
<td>-1V</td>
<td>float</td>
<td>~2V</td>
</tr>
</tbody>
</table>

\[ DSA = V_{BLt} - V_{BLc} \quad \text{OR} \quad I_{BLt} - I_{BLc} \]
4/4 Erase

<table>
<thead>
<tr>
<th>Mode</th>
<th>WL</th>
<th>BL</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand-by</td>
<td>0V</td>
<td>float</td>
<td>0V</td>
</tr>
<tr>
<td>Program</td>
<td>2V</td>
<td>0V</td>
<td>~1.5V</td>
</tr>
<tr>
<td>Read</td>
<td>1V</td>
<td>Signal</td>
<td>1V</td>
</tr>
<tr>
<td>Erase</td>
<td>-1V</td>
<td>float</td>
<td>~2V</td>
</tr>
</tbody>
</table>

WL = -1V

SL = 2V
CTT : Operating zones

Circuit Assist Techniques needed to operate in the "Safe" Zone
HiK Charge Trap Transistor Enables Process Free Multi-time Programmability

<table>
<thead>
<tr>
<th></th>
<th>eFUSE</th>
<th>Anti Fuse</th>
<th>MTPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mb/(mm$^2$)</td>
<td>&lt;0.1</td>
<td>~1</td>
<td>~1</td>
</tr>
<tr>
<td>Rewritable</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

1st Write

2nd Write

Bitmap Image
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm Bulk FinFET
• Future scope
• Summary
80Kb Twin Cell Macro Architecture

80 Data Line (DL) Read Out (80x4 = 320 Bit Lines)

Detect DIFVT
DIFVT = VTT - VTC

Protection devices not shown
Overwrite Protection

Before Programming

After Programming

Overwrite Protection

Overwrite: Too much VT Shift (TDDB Risk)

Insufficient write Too small VT Shift (Retention Risk)

DIFVT Distribution (Normalized)

DIFVT = VTt - Vtc
## Circuit Assist – Over Write Protection

<table>
<thead>
<tr>
<th></th>
<th>Program Cycle 1</th>
<th>Read Cycle 1</th>
<th>Program Cycle 2</th>
<th>Read Cycle 2</th>
<th>Margin Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell a</td>
<td>DIN=1</td>
<td>1</td>
<td>MATCHES DIN</td>
<td>1</td>
<td>MATCH</td>
</tr>
<tr>
<td>Cell b</td>
<td>DIN = 1</td>
<td>0</td>
<td>MISMATCH with DIN</td>
<td>1</td>
<td>MATCH</td>
</tr>
</tbody>
</table>

 comparable

\[ V_{TH} \text{ difference} = (V_{Tt} - V_{Tc}) \]

**Write in multi-steps with OWP**
Circuit Assist – Block Write

Step 1: Initial write (w)

Step 2 & 3: read+verify+write (rvw) (rewrite if not verified correctly)

Step 4: margin write (mw)

Cell a plane  |  Cell b plane
----------- | -----------
Step 1: Initial write (w)

Step 2 & 3: read+verify+write (rvw)

Step 4: margin write (mw)
Circuit Assist – Slew Sense Amplifier

- Turn on column switch
- Connect cell to SA
- Ramp the WL slowly to VDD
- Rise time ~500ps
- Differential charging of SA<sub>t</sub> and SA<sub>c</sub>
- Self timed Sense Amp
- Capacitance mismatch
- Fixable using reduced slew
Circuit Assist – Slew Sense Amplifier

NMOSc is still OFF

NMOSt turns ON in Saturation
Circuit Assist – Slew Sense Amplifier

SSA can sense 10% less VT shift compared to CCSA
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm Bulk FinFET
• Future scope
• Summary
SOI Hardware Results - OWP

Over-Write Protection ensures operation in the “safe” zone

J. Viraraghavan et al., IEEE VLSI Symp. 2016
Bake Tests Indicate 30% $V_{TH}$ Degradation in 10 years @ 125°C
SOI Hardware Results – Multi-Time Programming

a: Multiple Write (4X) with OWP

<table>
<thead>
<tr>
<th>PRE</th>
<th>1W</th>
<th>1ER</th>
<th>2W</th>
<th>2ER</th>
<th>3W</th>
<th>3ER</th>
<th>4W</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 1 Data
- 0 Data

PRE – Initial State
W – Write/ Program
ER - Erase

J. Viraraghavan et al., IEEE VLSI Symp. 2016
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm bulk FinFET
• Future scope
• Summary
14nm bulk FinFET Hardware Results

40Kb Checker Board Programming

256b Multi Programming

Pre  1\textsuperscript{st} Write  2\textsuperscript{nd} Write  3\textsuperscript{rd} Write  4\textsuperscript{th} Write
Prototype is functional using VDD = 0.7V
Projected Charge loss for 10 year product life: <35%
Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm bulk FinFET
- Future scope
- Summary
Future : Endurance Improvement Concept

Initially few vacancies

**Higher** gate voltage & long time to create more vacancies

Low gate voltage & short time to avoid new vacancies

Process Phase

Manufacturing Test Phase

User Phase

Initially few vacancies

**Higher** gate voltage & long time to create more vacancies

Low gate voltage & short time to avoid new vacancies

Initially few vacancies
**Future : Hardware-Based Security**

- Programmed bits physically invisible

- Enable Physically Unclonable Fuse (PUF) for a secure product source solution.

- Create a unified memory array, using these three modes of operation
  - OTPM Mode: Used for stable non-volatile bits
  - MTPM Mode: Self Destructive Re-writable Memory non-volatile bits
  - Intrinsic ID Mode: Physically Unclonable Fuse (PUF).

![Twi Cell Memory Cell](image)
Future: Unified memory using CTT MTPM

OTPM / MTPM / PUF

Error Correction Code Logic

Sense Amps

OTPM

SL

Re-write

Self-Destructive

MTPM

ECC

w/ OTPM bits

PUF

Rewritable

Random VT

Bitline Drivers
Outline

• Introduction
• Embedded Nonvolatile Memory Applications
• Charge Trap Transistor (CTT) Technology
• Macro Overview
• Hardware Results
  • SOI
  • 14nm bulk FinFET
• Future scope
• Summary
Summary

• MTPM demonstrated in both SOI and Bulk FIN
• Scalable standard Hi-K logic process
  • Process as it is (ex. No mask adder)
  • Operated in logic compatible voltages
  • ~30X more dense than eFUSE
• Future: MTPM fundamental concepts
  • Endurance improvement, using forming approach
  • Unified memory, using OTPM, self-destructive MTPM, PUF.

<table>
<thead>
<tr>
<th>Technology</th>
<th>32nm SOI</th>
<th>22nm SOI</th>
<th>14nm FIN Bulk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell</td>
<td>0.109μm² with 1.4nm Gox NMOS</td>
<td>0.144μm² with 1.2nm Gox NMOS</td>
<td>0.1411μm² with FIN NMOS</td>
</tr>
<tr>
<td>Macro Density</td>
<td>80Kb</td>
<td>64Kb</td>
<td>40Kb</td>
</tr>
<tr>
<td>Density/mm²</td>
<td>~2Mb/mm²</td>
<td>~2.5Mb/mm²</td>
<td>~1.3*Mb/mm²</td>
</tr>
<tr>
<td>Activation Energy</td>
<td>~1.4eV</td>
<td>-</td>
<td>~1.6eV</td>
</tr>
</tbody>
</table>
References

Acknowledgements

Darren Anand
John Fifield
Sami Rosenblatt
Xiang Chen
Krishnan Rengarajan
Giuseppe Larosa
Norman Robson
Jim Pape
Daniel Berger
Dan Moy
Robert Katz
Yoann Mamy Randriamihaja
Zakariae Chbili
Andreas Kerber
Raman Kodhandaraman

This work at UCLA is partially supported by the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings contained in this article are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.
Thank You