Semiconductor Memory

Seong-Ook Jung
2011. 4. 1.
sjung@yonsei.ac.kr

VLSI SYSTEM LAB, YONSEI University
School of Electrical & Electronic Engineering
Contents

1. Current Memory
2. Future of NAND Flash
3. Universal memory
   1. PRAM
   2. STT-MRAM
Memory Hierarchy

MEMORY

Volatile memory
- DRAM/SRAM
- FRAM

Non-volatile memory
- charge based device
  - Flash (NAND/NOR)
- resistance based device
  - MRAM
  - RRAM (ReRAM/PRAM)

by Samsung electronics
Volatile vs. Non-Volatile

◆ Volatile memory
  ● DRAM: fast speed, high density
     → Main memory
  ● SRAM: very fast speed, very low density
     → Cache memory

◆ Non-volatile memory
  ● NOR: very slow speed, low density
     → Program memory
  ● Flash: very slow speed, very high density
     → Storage memory
Charge Based vs. Resistance Based

- Charge based device (Current memory)
  - DRAM
  - SRAM
  - Flash

- Resistance base memory (Future memory → Universal memory)
  - PRAM
  - RRAM
SRAM

Cache hierarchy in Lynnfield
L1 ; 32KB (1core)
L2 ; 256KB (1core)
L3 ; 8MB (shared)

Intel processor ; Lynnfield

Layout of Lynnfield

SRAM cell
DRAM

Samsung DDR3 4GB DRAM

DRAM cell
DRAM Cell

Trench Cell

Stacked-capacitor Cell
Flash Memory

Samsung 256GB SSD

Flash memory cell

Samsung 32GB USB memory
## Comparison

<table>
<thead>
<tr>
<th></th>
<th>DRAM</th>
<th>FLASH</th>
<th>SRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speed</strong></td>
<td>fast</td>
<td>very slow</td>
<td>very fast</td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>high</td>
<td>very high</td>
<td>low</td>
</tr>
<tr>
<td><strong>Endurability</strong></td>
<td>better</td>
<td>poor</td>
<td>better</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>high</td>
<td>very low</td>
<td>low</td>
</tr>
<tr>
<td><strong>Refresh</strong></td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><strong>Retention</strong></td>
<td>volatile</td>
<td>non-volatile</td>
<td>volatile</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>bad</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td><strong>Data Storage Mechanism</strong></td>
<td>capacitor</td>
<td>FN tunneling, HCI to floating gate</td>
<td>Bi-stable flip-flop</td>
</tr>
</tbody>
</table>

By Korea Institute of Science & Technology Information (KISTI)
Improve **performance** and capacity of DRAM and SRAM
- Technology scaling
- Design technique

Function and role of DRAM and SRAM are not changed.
- SRAM ; cache memory in processor
- DRAM ; main memory unit in system

---

**DRAM and SRAM Trend**

<table>
<thead>
<tr>
<th>Year</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>SDRAM</td>
</tr>
<tr>
<td>2000</td>
<td>DDR1</td>
</tr>
<tr>
<td>2004</td>
<td>DDR2</td>
</tr>
<tr>
<td>2008</td>
<td>DDR3</td>
</tr>
<tr>
<td>2012</td>
<td>DDR4</td>
</tr>
</tbody>
</table>

**SRAM Cell Area**
- 0.5x every 2 years

**Contacted Gate Pitch**
- 0.7x every 2 years

*by Intel Technology Journal*
NAND Flash Trend

◆ Improve **capacity** and performance of NAND flash memory
  - Technology scaling
  - Design technique

◆ Positioning of NAND flash have been changed.
  - Past ; digital camera, MP3, USB memory.
  - Recent ; solid state drive (SSD) for replacing HDD

![Graph showing NAND Flash Trend](image)

- **SRAM**
- **DRAM**
- **NOR**
- **NAND**

**Storage Class Memory (SCM)**

- **IBM STT-MRAM** by Samsung electronics
- **3D ReRAM**
- **PRAM**
- **Universal Memory**

**Table:**

<table>
<thead>
<tr>
<th>Storage Class Memory (SCM)</th>
<th>IBM STT-MRAM</th>
<th>3D ReRAM</th>
<th>PRAM</th>
<th>Universal Memory</th>
</tr>
</thead>
</table>

**Figure:**

- **Write Speed**
  - **SRAM**
  - **DRAM**
  - **NOR**
  - **NAND**

**Endurance**

- **~10^15**
- **~10^5**

**Graph:**

- **Cell Size [um²/bit]**
  - **64M**
  - **128M**
  - **256M**
  - **512M**
  - **1G**
  - **2G**
  - **4G**
  - **8G**
  - **16G**
  - **SLC**
  - **MLC**

- **Bit Cost**
  - **Self-boosting ISPP etc.**
Future of NAND Flash
Invention of NAND Flash

ABSTRACT

A NAND structure cell is most promising as an ultra high density EEPROM to replace magnetic memories. This paper describes a new device technology to realize a high performance 4Mb EEPROM with NAND structure cell. The main features of the technology are a new NAND structure cell to realize wide threshold window and high reliability, and a high voltage CMOS process to realize program and erase operations, which require high voltage pulses such as 22V.

By using 1.0um design rules, the unit cell area per bit is 12.9um², which is small enough to realize a 4Mb EEPROM.

INTRODUCTION

In order to replace magnetic memories by solid state devices, a high density EEPROM has been required. Conventional 5V-only EEPROMs, utilizing Fowler-Nordheim tunneling, occupy a very large

by Toshiba, Flash handbook, 1992

by Toshiba, IEDM, 1988
Application I / Flash Cards

◆ Flash card is used for mobile devices with memory slot

- Digital camera
- Portable Video game
- Cellular phone
- PC
- Car navigation
Application II / Embedded Application

MP3 player

E-Book

8GB Flash

TOSHIBA TH58G6D1DTG80 8GB Flash
MARVELL 88W8686 and CSR Bluetooth
Software Distribution

- Distribute Software as a read-only device
  - No Viruses
- O/S and Bios already support optical drives
  - 100% emulation of CD/DVD drive
- Less packaging
- Perfect Application for 3/4bit MLC
- <$5 unit cost
Application IV / SSD

Samsung 256GB SSD

2.5 inch SSD

SATA interface compatibility
Evolution of NAND Flash Technology

- Self-boosting ISPP etc.
- SLC (양상)
- MLC (개발)
- Super-MLC (개발)

Start of Mass Production

YONSEI Univ.
School of EEE

by Samsung electronics
Limitation of NAND

- Beyond 30nm, Uncertainty of EUV Availability → Limit of Patterning
- Beyond 20nm, Uncertainty of Conv. Linear Scaling → Limit of Device
  → Super-MLC (3-bit, 4-bit, etc.), High Speed I/F, 3D Technology

![Graph showing NAND Technology Node (nm) over years from 1970 to 2030, with conventional linear scaling and litho tool limitations highlighted.](image-url)
High Speed Interface (DDR) NAND

- ONFI (Open NAND Flash Interface) : Intel, Micron, Hynix, etc.
- Toggle-mode NAND : Samsung, Toshiba

<table>
<thead>
<tr>
<th>Technology</th>
<th>50nm p-sub CMOS triple-well 3metal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell size</td>
<td>0.0111µm² (effective)</td>
</tr>
<tr>
<td>Chip size</td>
<td>169.5mm²</td>
</tr>
<tr>
<td>Organization</td>
<td>4096 x 128 pages x 512 blocks x 4 planes x 8</td>
</tr>
<tr>
<td>Power supply: Vcc</td>
<td>2.7V-3.6V</td>
</tr>
<tr>
<td>Power supply: VccQ</td>
<td>1.7V-1.95V or 2.7V-3.6V</td>
</tr>
<tr>
<td>Read time</td>
<td>30ns</td>
</tr>
<tr>
<td>Program time</td>
<td>160µs (typical)</td>
</tr>
<tr>
<td>Erase time</td>
<td>3ms (typical)</td>
</tr>
<tr>
<td>Clock cycle time</td>
<td>10ns</td>
</tr>
<tr>
<td>I/O width</td>
<td>x8</td>
</tr>
</tbody>
</table>
3D NAND for Tera-bit Storage

- 3D Vertical NAND
- High Density Oriented, CTF, MLC

by Toshiba, IEDM, 2007
by Toshiba, VLSI, 2007
### SDD vs. HDD

<table>
<thead>
<tr>
<th></th>
<th>2.5&quot; SATA 3.0Gbps SSD</th>
<th>2.5&quot; SATA 3.0Gbps HDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanism type</td>
<td>Solid NAND flash based</td>
<td>Magnetic rotating platters</td>
</tr>
<tr>
<td>Density</td>
<td>64GB</td>
<td>80GB</td>
</tr>
<tr>
<td>Weight</td>
<td>73g</td>
<td>365g</td>
</tr>
<tr>
<td>Performance</td>
<td>Read: 100MB/s, Write: 80MB/s</td>
<td>Read: 59MB/s, Write: 60MB/s</td>
</tr>
<tr>
<td>Active Power consumption</td>
<td>1W</td>
<td>3.86W</td>
</tr>
<tr>
<td>Operating Vibration</td>
<td>20G (10~2000Hz)</td>
<td>0.5G (22~350Hz)</td>
</tr>
<tr>
<td>Shock resistance</td>
<td>1,500G for 0.5ms</td>
<td>170G for 0.5ms</td>
</tr>
<tr>
<td>Operating temperature</td>
<td>0°C to 70°C</td>
<td>5°C to 55°C</td>
</tr>
<tr>
<td>Acoustic Noise</td>
<td>None</td>
<td>0.3 dB</td>
</tr>
<tr>
<td>Endurance</td>
<td>MTBF &gt;2M hours</td>
<td>MTBF &lt; 0.7M hours</td>
</tr>
</tbody>
</table>

However, high cost per capacity, now by Samsung electronics.
Component of SSD

- Performance = f(CPU, DRAM, Flash, Host Interface, HW automation)

**DRAM**
- 32 → 128 → 1024 MB
- Mobile SDRAM → DDR → DDR2
- PRAM / FRAM (Future)

**Host Interface**
- PATA → SATA 3G → SATA 6G (Note PC)
- SAS 3G → SAS 6G (Server / Storage)

**Power Supply**
- Power-Failure Monitor
- Super Cap

**NAND Flash**
- SLC/MLC (50 → 40 → 30 nm)
- SDR I/F (40 MB/s/chip)
- DDR I/F (133 → 266, → ... MB/s)

**Controller**
- ARM CPU (133 → 300 MHz)
- Single / Duo Core CPU
- Hardware Acceleration
- 4 / 8 / 16 ch. (NAND I/F)

**Firmware**
- w/o OS → Real Time OS
- w/o Queue → Queued CMD

**SSD Capacity**
- 32 / 64 / 128 / 256 / 512 GB

by Samsung electronics
SSD / Solving the I/O Bottleneck

◆ Bridge Performance Gap between CPU and HDD
SSD / Solving the Power Bottleneck

- HDD: Higher RPM = Higher Power + Generates more Heat
- SSD: Less Power / No Heat saves lifetime Energy Costs…

Watts used in Operation Mode

Watts used in Idle Mode

by Samsung electronics

YONSEI Univ.
School of EEE
Wear-leveling for Optimization

**Dynamic Wear-leveling**

- Red: Static data such as system data.
- Blue: Dynamic data such as user data

**Static Wear-leveling**

- Red: Static data such as system data.
- Blue: Dynamic data such as user data

- Wear-leveling by FTL (Flash Controller)

- Dynamic Wear-leveling
- Static Wear-leveling

- SSD

- NAND Controller
  - Bad Block Management
  - Wear-leveling
  - ECC

- Application Software
- File System (OS)

- SSD

- MP3, USB, DSC etc.

by Samsung electronics
Sector Size for Optimization

- Existing OS is for HDD! → OS should be optimized for SSD!!!

**Sector size optimization**

- Minimum write/read unit of NAND is a page.
- Typical page size is 4-8KByte.
- A page is written only ONCE to avoid the program disturbance.
- With current OS having 512Byte sector, one sector write wastes >80% of data in a page.

- LBD(Long Block Data) sector standard (Windows Vista): 4KByte sector size fits better with SSD.
- As the page size increases as NAND is shrinking, larger sector size such as 64KByte or 128KByte is required.
Self-monitoring, Analysis and Reporting Technology

SMART (Self-Monitoring, Analysis and Reporting Technology)

Monitor the storage and report/predict the failure.
SMART for HDD is NOT smart because it is very difficult to predict the mechanical failure.

- SMART for SSD can be really smart.
- Product lifetime can be predicted because the failure rate is highly correlated with the write/erase cycles.
- Predict the SSD lifetime by monitoring the write/erase cycles and replace SSD before the fatal failure occurs.
Future Memory
(Universal Memory)

VLSI
SYSTEM LAB.
Universal Memory

◆ Universal memory is desired for next-generation memory.
  ● Nonvolatile memory
  ● High speed
  ● High density
  ● High endurability
  ● Low power

◆ Some candidates
  ● PRAM
  ● STT-MRAM
  ● FeRAM
  ● ReRAM
  ● .......

By Samsung electronics
YONSEI Univ.
School of EEE
Power Dissipation of IT Equipments

◆ In 2025, Power dissipation will reach 5 times larger.
Normally OFF, Instant ON

- Read operation; sensing resistance of GST
- Voltage biased to GST must be limited under $V_{th}$ to prevent disturb.
- Current sensing scheme
  - Applying read voltage to cell converts from resistance to current
  - Load device converts from current to voltage
  - Sense amplifier converts from analog voltage value to digital output

by T. Kawahara, IEEE ED/SSC Mini-Colloquium, 2009
Change Memory Configuration

- Nonvolatile RAM enhances user’s convenience.
- Instant ON. Quickly software changing.

by T. Kawahara, IEEE ED/SSC Mini-Colloquium, 2009
Nonvolatile (NV) RAM Application

- Innovation for low power system including hardware, software, and architecture
Impact on Performance

- Power ON time is improved to 1/9.
- Nonvolatility achieves low power also.
Impact on Power

- High-end cellular phone with large memory with low standby power (1/10).
Universal Memory I
PRAM
Structure of PRAM Cell

RESET state = High resistance

SET state = Low resistance
Write Operation of PRAM

- **Reset pulse** (strong & short) ; Amorphous state (~100kΩ)
- **Set pulse** (weak & long) ; Crystalline state (kΩ)

by KIST, 대한전자공학회, 2005
Read Operation of PRAM

◆ Read operation; sensing resistance of GST
◆ Voltage biased to GST must be limited under $V_{\text{th}}$ to prevent disturb.
◆ Current sensing scheme
  - Applying read voltage to cell converts from resistance to current
  - Load device converts from current to voltage
  - Sense amplifier converts from analog voltage value to digital output

by KIST, 대한전자공학회, 2005
Recent Technical Issues
Reducing Required RESET Current

◆ Reducing BEC

Increasing heat by increasing current density → reducing $I_{\text{RESET}}$

◆ Confined contact

Increasing heat by increasing current density → reducing $I_{\text{RESET}}$
Recent Technical Issues
Reducing Required RESET Current

◆ Impurity doping

Increasing GST resistance
→ increasing heat
→ reducing $I_{\text{RESET}}$

Reset Current Regime

by Samsung Electronics, Sym. VLSI Tech., 2007
Recent Technical Issues
Obtaining Larger RESET Current

- Enhancing current driving capability
  - Vertical BJT

<table>
<thead>
<tr>
<th>Unit Cell Structure</th>
<th>256 Mb PRAM Cell</th>
<th>512Mb PRAM Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ISSCC2006</td>
<td>ISSCC2007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Switch Device</th>
<th>(3-terminal) MOS Switch</th>
<th>(2-terminal) Diode Switch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Tech.</td>
<td>100nm</td>
<td>90nm</td>
</tr>
<tr>
<td>Cell Size</td>
<td>0.166mm² (16.6F²)</td>
<td>0.0467mm² (5.8F²)</td>
</tr>
<tr>
<td>Switch Device Capability</td>
<td>6 mA/mm² @ Vₜ₉=1.5V, Vₚ₉=3.0V</td>
<td>21mA/mm² @ Vₐ₉=1.5V</td>
</tr>
<tr>
<td>Integration Density (Chip Size)</td>
<td>256 Mb (79.2 mm²)</td>
<td>512Mb (91.5 mm²)</td>
</tr>
</tbody>
</table>

by Samsung Electronics
**History of PRAM**

- Production as NOR flash replacement in recent years.

<table>
<thead>
<tr>
<th>Year</th>
<th>70's</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>Stand-alone</td>
<td>256b (ECD, Intel)</td>
<td>4Mb (STMicro.)</td>
<td>8Mb (Samsung)</td>
<td>256Mb (Samsung)</td>
<td>512Mb (Samsung)</td>
<td>128Mb (Numonyx/Micron)</td>
<td>512Mb (Samsung)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Embedded</td>
<td>4Mb (STMicro.)</td>
<td>4Mb (Hitachi/Renesas)</td>
<td>4Mb (STMicro.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

by T. Kawahara, ASP-DAC Non-volatile Memory, 2011
Future of PRAM

◆ When ?? 2011 ?
  ● Samsung and Micron expect to possess technology to mass-produce 512Mb~1Gb PRAM.
  ● Absence of alternative market is obstacle of commercialization of PRAM.
  ● Samsung or Micron may launch PRAM in 2011.

◆ Target !! NOR Flash !!
  ● Recently, improvement of NOR flash disturbs commercialization of PRAM.
  ● Main product of NOR flash is 256, 512Mb and uses for embedded system.

◆ Future !!
  ● SSD drive
  ● System that don’t need booting
Universal Memory II
STT-MRAM

VLSI
SYSTEM LAB.
MTJ Device Structure

- MTJ
  - Two magnetic layer (Free and pinned layer)
  - Insulating layer (Tunnel barrier)

- $R_{\text{MTJ}}$ ; depends on the state of free layer

<table>
<thead>
<tr>
<th>State</th>
<th>Effective resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Low ($R_0$)</td>
</tr>
<tr>
<td>Anti-Parallel</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>High ($R_1$ = $R_0*(1+\text{MR})$)</td>
</tr>
</tbody>
</table>

(MR ; magnetoresistance)

Reading operation $\rightarrow$ Reading resistance of MTJ
Writing operation $\rightarrow$ Switching free layer of MTJ
Write Operation of Conventional MRAM

◆ Applied Magnetic Fields

Selected BL
Selected Cell
Selected WL

< Applied magnetic field >

◆ State Change

< State change >
Write Operation of STT-MRAM

◆ Spin-polarized electron

(a) "0" Write (Parallelizing)
(b) "1" Write (Anti-Parallelizing)

(c) R-I Characteristics

< Parallelizing and Anti-Parallelizing Current>

by T. Kawahara, et al., ISSCC, 2007
**Conventional MRAM vs. STT-MRAM**

◆ STT-MRAM has good potential in scalability due to Write current.
Recent Technical Issues
Reducing Required Critical Current

◆ Perpendicular MTJ
  - TMR element with perpendicular magnetic anisotropy (P-TMR)
  - Lower critical current than I-TMR

<table>
<thead>
<tr>
<th>Research Group</th>
<th>Conference</th>
<th>MTJ size (diameter)</th>
<th>Critical current</th>
<th>Switching time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toshiba</td>
<td>IEDM2008</td>
<td>55nm</td>
<td>49uA (AP-P)</td>
<td>4ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100uA (P-AP)</td>
<td></td>
</tr>
</tbody>
</table>

Critical current ($I_c$) of P-ST is dependent on MTJ size.
(opposite result compared with page 2)

by T. Kishi, et al., IEDM, 2008
To minimize the cell area, the isolation area between a cell and an adjacent cell is replaced by the adjacent cell’s transistor; the “off” state of the channel region of the adjacent cell can insulate electrically among memory cells.

Fig. 1 2T1R Memory Cell

Larger write current with same area

by R. Takemura, et al., Symposium on VLSI circuits, 2009
NEW Application using MTJ
Spintronics Logic

- Recently, nonvolatile logic using MTJ had been studied.
- Highly expected for LSI advancement and innovation.

Full adder with nonvolatile input B
- Dynamic logic type
- Dynamic power can reduced to 23%

Nonvolatile Flip-flop
- Cross-coupled inverter latch type
- Standby power can reduced to 0%

by Noboru Sakimura, et al., CICC, 2008
Power Reduction by STT-MRAM and Spintronics Logic

◆ Keep normally the equipment turned off.
◆ Power consumption adjusted to one third.
History of STT-MRAM

- R&D moved to SPRAM.

<table>
<thead>
<tr>
<th>year</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conf.</td>
<td>512b (Motorola)</td>
<td>256kb (IBM)</td>
<td>1Mb (IBM/Infineon</td>
<td>128kb (IBM/Infineon)</td>
<td>16kb (Samsung)</td>
<td>4Mb (Freescale)</td>
<td>512kb (NEC)</td>
<td>16Mb (Toshiba/NEC)</td>
<td>16Mb (Toshiba)</td>
<td>64Mb (Hynix)</td>
<td></td>
</tr>
<tr>
<td>SPRAM</td>
<td>1Mb (Renesas)</td>
<td>4Mb (Freescale)</td>
<td>16Mb (Toshiba/NEC)</td>
<td>4Mb (TDK)</td>
<td>256kb (Grandis)</td>
<td>2Mb (Hitachi/Tohoku Univ.)</td>
<td>4Mb (TDK)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

by T. Kawahara, ASP-DAC Non-volatile Memory, 2011

YONSEI Univ.
School of EEE
Conclusion

- The density and performance of flash memory have been improved by improvement of process and design technology.
- Solid state drive (SSD) is remarkable as alternate storage device. As cost per capacity decreases, SDD will replace HDD, gradually.
- **Advantage of SSD**
  - Low power
  - High performance
  - No noise & vibration
  - Low heat
- Next generation memory had been studied for overcoming current memory. (PRAM, STT-MRAM, …)
- **Requirement for universal memory**
  - Nonvolatile
  - Low power
  - High performance
  - High density