NAND Flash Memory

Dongkun Shin (dongkun@skku.edu)
Embedded Software Laboratory
Sungkyunkwan University
http://nyx.skku.ac.kr/
Flash Memory
Modern Storage Media

Timeline: Data Backup Storage

- **Blu-ray disk & HD-DVD**
  - 1994: Network Backup Storage
  - 2003: Online Backup Services

- **Flash Drive**
  - The timeline illustrates the evolution of storage strategies for backups.
  - It shows when this or that storage device or technology began/stopped to be used for backup purposes.

- **DVD**
  - 1995

- **CD - R/RW**
  - 1984

- **Floppy Disk**
  - 1971

- **Hard Disk**
  - 1956

- **Magnetic Tape**
  - 1890

- **Punch Card**
  - 1950

Year:
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000
- 2005

Categories:
- **Optical**
- **Electrical**
- **Optical**
- **Optical**
- **Magnetic**
- **Magnetic**
- **Magnetic**
- **Mechanical**

http://www.backuphistory.com
Storing Data

• Change the state of something
• Read the current state
• Maintain the state without any power (non-volatility)
• Better if we can change the state multiple times (overwrite)

• Having just two states (0 or 1) is simplest and most reliable
NAND Flash Categories

NAND Flash

Raw NAND
- SLC
- MLC
- TLC

Managed NAND
- Card
- Embedded
- SSD

Raw NAND uProcessor F/W (FTL)
Flash Memory Basics

• Two states based on the presence of electrons

- 0 = Electrons present
- 1 = No electrons

• Challenges
  – How to attract or expel electrons?
  – How to find whether there are electrons or not?
  – How to keep electrons without any power?
Flash Memory Cell

• Transistor with floating gate
  – The floating gate is insulated all around with an oxide layer
  – Electrons trapped in the floating gate can remain for up to years

http://www.thenandflash.com
Flash Memory Operations

• **Write (or program)**
  – Apply a high voltage at the CG
  – Trap electrons inside the FG
  – Once programmed, the cell can not be reprogrammed until it is erased

• **Erase**
  – Apply a large voltage in the opposite direction
  – Pull the electrons away from the FG
Flash Memory Operations

• Read
  – Electrons in the FG partially cancel the electric field from the CG, increasing the threshold voltage of the cell
  – A higher voltage must be applied to the CG to make the channel conductive
Flash Memory Characteristics

• Erase-before-write
  – Read
  – Write or Program: $1 \rightarrow 0$
  – Erase: $0 \rightarrow 1$

• Bulk erase
  – Program unit:
    • NOR: byte or word
    • NAND: page
  – Erase unit: block
NAND vs. NOR Flash Memory

<table>
<thead>
<tr>
<th></th>
<th>NOR flash</th>
<th>NAND flash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access mode:</td>
<td>Linear random access</td>
<td>Page access</td>
</tr>
<tr>
<td>Replaces:</td>
<td>ROM</td>
<td>Mass Storage</td>
</tr>
<tr>
<td>Cost:</td>
<td>Expensive</td>
<td>Cheap</td>
</tr>
<tr>
<td>Device Density:</td>
<td>Low (64MB)</td>
<td>High (1GB)</td>
</tr>
<tr>
<td>Erase block size:</td>
<td>8k to 128K typical</td>
<td>32x512b / 64x2K pages</td>
</tr>
<tr>
<td>Endurance:</td>
<td>100k to 1M erasures</td>
<td>10k to 100k erasures</td>
</tr>
<tr>
<td>Erase time:</td>
<td>1 second</td>
<td>2ms</td>
</tr>
<tr>
<td>Programming:</td>
<td>Byte by Byte, no limit on writes</td>
<td>Page programming, must be erased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>before re-writing</td>
</tr>
<tr>
<td>Data sense:</td>
<td>Program byte to change 1s to 0s.</td>
<td>Program page to change 1s to 0s.</td>
</tr>
<tr>
<td></td>
<td>Erase block to change 0s to 1s</td>
<td>Erase to change 0s to 1s</td>
</tr>
<tr>
<td>Write Ordering:</td>
<td>Random access programming</td>
<td>Pages must be written sequentially</td>
</tr>
<tr>
<td></td>
<td></td>
<td>within block</td>
</tr>
<tr>
<td>Bad blocks:</td>
<td>None when delivered, but will wear</td>
<td>Bad blocks expected when delivered.</td>
</tr>
<tr>
<td></td>
<td>out so filesystems should be fault</td>
<td>More will appear with use.</td>
</tr>
<tr>
<td></td>
<td>tolerant</td>
<td>Thus fault tolerance is a necessity.</td>
</tr>
<tr>
<td>OOB data:</td>
<td>No</td>
<td>Yes (16 bytes)</td>
</tr>
</tbody>
</table>

| Cell size              | 4F²                                    | 10F²                                 |

For Code Storage  
For Mass Storage
Trees of MOS memory

**Volatile**
- Random Access Memory (RAM)
  - Static RAM (SRAM)
    - 1970 by Intel
  - Dynamic RAM (DRAM)
    - 1970 by Intel

**Nonvolatile (NVM)**
- Read Only Memory (ROM)
- Programmable ROM (PROM)
  - UV erase
  - elec. erase
- Mask ROM
  - fixed
- EEPROM
  - 1971 by Intel
  - bit-wise
- Flash
  - Conventional
    - 1979 by Intel
  - 1984 by Toshiba
NAND Flash Memory
Making it Smaller

• Hwang’s law
  – The density of the top-of-the-line flash memory chips will double every 12 months
Density Growth

Source: IBM, Flash Summit 2018
Cost Trends

Source: IEEE Computer, 2011
Recent Cost Trends

https://www.anandtech.com/show/9799/best-ssds
NAND Global Market Share

Source: DRAMeXchange & Statista, 2015.
### NAND Technology by Company

<table>
<thead>
<tr>
<th>Makers</th>
<th>20 nm Class (2D)</th>
<th>10 nm Class (2D)</th>
<th>3D NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samsung</td>
<td>27nm</td>
<td>19nm</td>
<td>24L</td>
</tr>
<tr>
<td></td>
<td>21nm</td>
<td>16nm</td>
<td>32L MLC</td>
</tr>
<tr>
<td>Toshiba/SanDisk</td>
<td>24nm</td>
<td>19nm</td>
<td>32L TLC</td>
</tr>
<tr>
<td>Micron</td>
<td>25nm</td>
<td>15nm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20nm</td>
<td>A-19nm</td>
<td></td>
</tr>
<tr>
<td>SK Hynix</td>
<td>26nm</td>
<td>16nm</td>
<td></td>
</tr>
</tbody>
</table>

J. Choe, Comparison of 20nm & 10nm-class 2D Planar NAND and 3D V-NAND Architecture, FMS, 2015.
NAND by Applications

Source: Samsung Electronics, 2014.
NAND Flash Architecture
Logical View of NAND Flash

- A collection of blocks
- Each block has a number of pages
- The size of a block or a page depends on the technology (but, it’s getting larger)
NAND Flash Example

- 2Gb NAND flash device organization

Source: Micron Technology, Inc.
Figure 1. K9GAG08X0C Functional Block Diagram
# Command

<table>
<thead>
<tr>
<th>Command</th>
<th>Command Cycle 1</th>
<th>Number of Address Cycles</th>
<th>Data Cycles Required</th>
<th>Command Cycle 2</th>
<th>Valid During Busy</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ PAGE</td>
<td>00h</td>
<td>5</td>
<td>No</td>
<td>30h</td>
<td>No</td>
</tr>
<tr>
<td>READ PAGE CACHE SEQUENTIAL</td>
<td>31h</td>
<td>–</td>
<td>No</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>READ PAGE CACHE SEQUENTIAL LAST</td>
<td>3Fh</td>
<td>–</td>
<td>No</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>READ for INTERNAL DATA MOVE</td>
<td>00h</td>
<td>5</td>
<td>No</td>
<td>35h</td>
<td>No</td>
</tr>
<tr>
<td>RANDOM DATA READ</td>
<td>05h</td>
<td>2</td>
<td>No</td>
<td>E0h</td>
<td>No</td>
</tr>
<tr>
<td>READ ID</td>
<td>90h</td>
<td>1</td>
<td>No</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>READ STATUS</td>
<td>70h</td>
<td>–</td>
<td>No</td>
<td>–</td>
<td>Yes</td>
</tr>
<tr>
<td>PROGRAM PAGE</td>
<td>80h</td>
<td>5</td>
<td>Yes</td>
<td>10h</td>
<td>No</td>
</tr>
<tr>
<td>PROGRAM PAGE CACHE</td>
<td>80h</td>
<td>5</td>
<td>Yes</td>
<td>15h</td>
<td>No</td>
</tr>
<tr>
<td>PROGRAM for INTERNAL DATA MOVE</td>
<td>85h</td>
<td>5</td>
<td>Optional</td>
<td>10h</td>
<td>No</td>
</tr>
<tr>
<td>RANDOM DATA INPUT</td>
<td>85h</td>
<td>2</td>
<td>Yes</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>ERASE BLOCK</td>
<td>60h</td>
<td>3</td>
<td>No</td>
<td>D0h</td>
<td>No</td>
</tr>
<tr>
<td>RESET</td>
<td>FFh</td>
<td>–</td>
<td>No</td>
<td>–</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Enhanced Techniques

- Cache programming
  - Double buffering
  - Overlap the programming of current data and the transfer of next data
- Copy-back operation
- Multi-plane operation
- Chip level interleaved operations
Plane

- Each plane has its own page register and cache register
- Pages can be programmed or read at once
- Optional feature: 1, 2, 4, 8, … planes

1 page = (2K + 64 bytes)
1 block = (2K + 64) bytes x 64 pages = (128K + 4K) bytes
1 plane = (128K + 4K) bytes x 2,048 blocks = 2,112Mb
1 device = 2,112Mb x 2 planes = 4,224Mb
Die / Chip

- Each chip has multiple dies (can be stacked)
- + extra circuits, chip enable signal, ready/busy signal
Multi-bit Flash Memory

- Store multi-bit data into a cell
  - increase memory density without chip size increase
  - difficult to achieve in terms of process, reliability, etc…

SLC

- 1 bit

MLC

- 2 bit

TLC

- 3 bit
Samsung V-NAND
Characteristics of NAND Flash
Erase Before Write

- In-place update (overwrite) is not allowed
- Pages must be erased before new data is programmed
- The erase unit is much larger than the read/write unit
  - Read/write unit: page (4KB, 8KB, 16KB, …)
  - Erase unit: block (64-512 pages)

- What if there are live pages in the block we wish to erase?
Limited Lifetime

• The number of times NAND flash blocks can be reliably programmed and erased (P/E cycle) is limited
  – SLCs: 50,000 ~ 100,000
  – MLCs: 1,500 ~ 5,000
  – eMLCs (Enterprise MLCs): 10,000 ~ 30,000
  – TLCs: < 1,000

• High voltage applied to cell degrades oxide
  – Electrons are trapped in oxide
  – Break down of the oxide structure

• Requires wear leveling
Writing Letters and Erasing Paper

“Because thin paper wears down more easily than thick paper.”

Old technology
Thick Paper

Thin Paper
Advanced Semiconductor technology

2000

Flash Endurance

E. Grochowski et al., Future Technology Challenges for NAND Flash and HDD Products, FMS, 2012.
Asymmetric Read/Write Latency

• Reading a page is faster than programming it
• Usually more than 10x
  – e.g. 1ynm MLC\(^1\): Read 45\(\mu\)s, Program 1350\(\mu\)s, Erase 4ms
• Programming a page should go through multiple steps of Program & Verify phases

• As the technology shrinks, read/write latency tends to increase
• MLC and TLC make it even worse

\(^1\) D. Sharma, System Design for Mainstream TLC SSD, FMS, 2014.
MLC Programming

• LSB programmed first
  – Cell cannot move to the lower voltage before erase

LSB Program: 1) Erase → Erase, 2) Erase → LSB
MSB Program: 1) Erase → Erase, 2) Erase → PV1, 3) LSB → PV2, 4) LSB → PV3
TLC Programming

- Three step programming

![Diagram showing 3 steps of TLC Programming]

Program order to minimize WL-to-WL coupling
Three pages need to be kept for reprogram WL(n)

A New 3-bit Programming Algorithm using SLC-to-TLC Migration for 8MB/s High Performance TLC NAND Flash Memory, 2012 VLSI Symposium
Paired Pages in MLC/TLC

• One cell represents two or three bits in paired pages
  – LSB: low voltage, fast program, less error
  – MSB: high voltage, slow program, more error

• Performance difference

• LSB page can be corrupted when MSB page programming is interrupted

L. M. Grupp et al., The Harey Tortoise: Managing Heterogeneous Write Performance in SSDs, USENIX ATC, 2013.
### MLC vs. TLC

<table>
<thead>
<tr>
<th></th>
<th>MLC NAND 1ynm 128Gb</th>
<th>TLC NAND 1ynm 128Gb</th>
<th>TLC NAND (SLC mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t&lt;sub&gt;R&lt;/sub&gt; (read)</td>
<td>45 μs</td>
<td>80 μs (1st)</td>
<td>50 μs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>105 μs (2nd)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80 μs (3rd)</td>
<td></td>
</tr>
<tr>
<td>t&lt;sub&gt;PROG&lt;/sub&gt; (program)</td>
<td>1350 μs</td>
<td>550 μs (1st)</td>
<td>350 μs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1700 μs (2nd)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4650 μs (3rd)</td>
<td></td>
</tr>
<tr>
<td>t&lt;sub&gt;BERS&lt;/sub&gt; (erase)</td>
<td>4 ms</td>
<td>10 ms</td>
<td>10 ms</td>
</tr>
</tbody>
</table>

Bit Errors

- Bits are flipping frequently
- Error Correction Code (ECC) in spare area

### Bits Required in the NAND Flash Spare Area

<table>
<thead>
<tr>
<th>Error Correction Level</th>
<th>Bits Required in the NAND Flash Spare Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hamming</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Source: Micron Technology, Inc.
ECC Requirements

• Endurance continues to deteriorate
• Stronger ECCs are required: RS, BCH, LDPC

Sources of Error

- **Write disturbance**
  - When a page is programmed, adjacent calls receive elevated voltage stress

- **Read disturbance**
  - Repeated reading from one page can alter the values stored in other unread pages

- **Retention error**
  - Threshold voltage shifts down due to charge leakage from the floating gate
Bad Blocks

• Initial bad blocks
  – Due to production yield constraints and the pressure to keep costs low
  – SLCs: up to 2%
  – MLCs: up to 5%

• Run-time bad blocks
  – Read, write, or erase failure
  – Permanent shift in the voltage levels of the cells due to trapped electrons

• Requires run-time bad block management
Page Programming Constraints

• NOP
  – The number of partial-page programming is limited
  – 1 / sector for most SLCs (4 for 2KB page)
  – 1 / page for most MLCs and TLCs

• Sequential page programming
  – Pages should be programmed sequentially inside a block
  – For large block SLCs, MLCs, and TLCs

• SLC mode
  – Possible to use only LSB pages in MLCs and TLCs
  – Faster and more reliable, higher P/E cycles
# Comparisons

<table>
<thead>
<tr>
<th></th>
<th>SLC</th>
<th>MLC</th>
<th>TLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bits per cell</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Performance</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Endurance</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Capacity</td>
<td>★</td>
<td>★★</td>
<td>★★★</td>
</tr>
<tr>
<td>Reliability</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Cost / GB</td>
<td>$$$</td>
<td>$</td>
<td>$</td>
</tr>
<tr>
<td>Applications</td>
<td>Enterprise</td>
<td>Enterprise / Consumer</td>
<td>Consumer</td>
</tr>
</tbody>
</table>
NAND Read/Write Time

- Total time required for a NAND read or write
  - $t_R/t_{PROG}$: NAND array access time
  - $t_{RC}/t_{WC}$: NAND data transfer time between the host and the data register
  - As page sizes have increased from 2kB to 8kB, the data transfer time has approximately quadrupled from 53µs to 219µs at the conventional transfer rate of 40 Mbps.

- How to reduce data transfer time
  - DDR devices have been introduced in which two bytes are transferred per cycle. (133 Mbps)
Toggle-mode NAND

Growing Need for Higher Performance

Slide form Yoon (Samsung), Flash Summit 2009
Toggle-mode NAND

- High speed “Toggle-Mode” operation
  - No clock – Asynchronous Double Data Rate
  - High performance by using the asynchronous

- Interface for backward compatibility
  - Bidirectional DQS for read and write operations
  - DQS signal is driven by the host when it is writing data to the NAND and is driven by the NAND when the NAND is sending to the host.
  - Each rising and falling edge of the DQS signal is associated with a data transfer.

Slide form Yoon (Samsung), Flash Summit 2009
Why Toggle-Mode NAND?

- High performance
  - Supports 133Mbps and higher
- No free-running clock
  - Less power consumption
  - Consumes power only during a read or write operation
  - Free from IP issues
- Flexibility of operating frequency
  - 15ns~25ns in toggle-mode NAND vs. 25ns in legacy NAND
  - No additional mode change required
- Easy migration from legacy NAND
  - Same signal functionality as legacy NAND

Slide from Yoon (Samsung), Flash Summit 2009
Two major DDR NAND Designs

• ONFI NAND
  – Open NAND Flash Interface (ONFI) Working Group, supported by several NAND manufacturers including Hynix, Intel, Micron, SanDisk and Spansion

• DDR NAND
  – Samsung and Toshiba
  – Asynchronous double data rate NAND without a separate clock signal to enable a lower power solution
NAND Programming (ISPP)

- For precise programming at NAND flash memory
- Repeats the program-and-verify pulses with a stair case program voltage until all the memory cells in the target word-line are programmed.
- $\Delta V_{PP}$ determines the width of $V_{th}$ distribution

S. Park and D. Shin, “Adaptive Program Verify Scheme for Improving NAND Flash Memory Performance and Lifespan,” A-SSCC’12
NAND Flash Error

- Two parameters affect the threshold voltage ($V_{th}$) of memory cells
  - Random telegraph noise (RTN) over PE cycling
    - Stress: Increase $V_{th}$ due to charge trapping
    - Recovery: Decrease $V_{th}$ due to charge detrapping at retention mode
  - Cell-to-cell interference

- Flash memory cells gradually wear out with P/E cycling
  - More RTN
  - Must fabricate enough number of ECC that can tolerate the worst-case storage reliability at the end of memory lifetime.
Exploiting Wear-Out Dynamics

• Existing worst-case oriented ECC redundancy largely under-utilized over the entire lifetime of memory, especially at its early lifetime.

• Adaptively trade such underutilized ECC redundancy for improving performance ➔ PE cycling-aware adaptive scheme

• Early PE cycles (Low RTN): Large $\Delta V_{pp}$ (fast) ➔ More error

• Late PE cycles (high RTN): Small $\Delta V_{pp}$ (slow) ➔ Less error

"Exploiting Memory Device Wear-Out Dynamics to Improve NAND Flash Memory System Performance," FAST’11
**Beauty and the Beast**

- **NAND Flash memory is a beauty**
  - Small, light-weight, robust, low-cost, low-power non-volatile device

- **NAND Flash memory is a beast**
  - Much slower program/erase operations
  - No in-place-update
  - Erase unit > write unit
  - Limited lifetime
  - Bit errors, bad blocks, …

- **Software support is essential for performance and reliability!**