NAND Flash-based Storage

Jin-Soo Kim (jinsookim@skku.edu)
Computer Systems Laboratory
Sungkyunkwan University
http://csl.skku.edu
Today’s Topics

- NAND flash memory
- Flash Translation Layer (FTL)
- OS implications
Flash Memory Characteristics

- **Flash memory**
  - Non-volatile, Updateable, High-density
  - Low cost, Low power consumption, High reliability

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

- **Read faster than write/erase**

- **Bulk erase**
  - Erase unit: block
  - Program unit: byte or word (NOR), page (NAND)
NAND Flash Architecture

- **2Gb NAND flash device organization**

  Serial input (x8 or x16): 30ns (MAX CLK)

  PROGRAM: ~ 300μs/page

  NAND Flash Memory Array

  NAND Flash Page 2,112 bytes

  64 pages per block

  NAND Flash Block

  Data area: 2,048 bytes

  Spare area (ECC, etc.): 64 bytes

  Serial output (x8 or x16): 30ns (MAX CLK)

  READ (page load): ~ 25μs

  BLOCK ERASE: ~ 2ms

  2,048 blocks (2Gb SLC device)

Source: Micron Technology, Inc.
NAND Flash Types

- **SLC NAND Flash**
  - Small block ($\leq$ 1Gb)
  - Large block ($\geq$ 1Gb)

- **MLC NAND Flash**

- **TLC NAND Flash**

Source: Micron Technology, Inc.
NAND Applications

- Universal Flash Drives (UFDs)
- Flash cards
  - CompactFlash, MMC, SD, Memory stick, ...
- Embedded devices
  - Cell phones, MP3 players, PMPs, PDAs, Digital TVs, Set-top boxes, Car navigators, ...
- Hybrid HDDs
- Intel Turbo Memory
- SSDs (Solid-State Disks)
SSDs vs. HDDs (1)

Source: http://www.anandtech.com
## SSDs vs. HDDs (2)

<table>
<thead>
<tr>
<th>Feature</th>
<th>SSD (Samsung)</th>
<th>HDD (Seagate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Samsung SSD 840 EVO</td>
<td>ST3000DM001 (Barracuda 7200.14)</td>
</tr>
<tr>
<td>Capacity</td>
<td>1TB (19nm 128Gb TLC x 64, 3 cores)</td>
<td>3TB (3 Discs, 6 Heads, 7200RPM)</td>
</tr>
<tr>
<td>Form factor</td>
<td>2.5” Weight: 53g</td>
<td>3.5” Weight: 626g</td>
</tr>
<tr>
<td>Host interface</td>
<td>Serial ATA-3 (6.0 Gbps) Host transfer rate: 600MB</td>
<td>Serial ATA-3 (6.0 Gbps) Host transfer rate: 600MB</td>
</tr>
<tr>
<td>Power consumption</td>
<td>Active: 0.1W Idle/Standby/Sleep: 0.045W</td>
<td>Active: 8.0W Idle: 5.8W, Standby/Sleep: 0.75W</td>
</tr>
<tr>
<td>Performance</td>
<td>Sequential read: Up to 540 MB/s</td>
<td>Power-on to ready: 15 sec (typ)</td>
</tr>
<tr>
<td></td>
<td>Sequential write: Up to 520 MB/s</td>
<td>Average latency: 4.16 msec</td>
</tr>
<tr>
<td>Measured performance¹</td>
<td>Sequential read: 428.7 MB/s</td>
<td>Sequential read: 145.7 MB/s</td>
</tr>
<tr>
<td></td>
<td>Sequential write: 374.0 MB/s</td>
<td>Sequential write: 138.8 MB/s</td>
</tr>
<tr>
<td></td>
<td>Random read: 96.0 MB/s</td>
<td>Random read: ? MB/s</td>
</tr>
<tr>
<td></td>
<td>Random write: 366.3 MB/s</td>
<td>Random write: 1.09 MB/s</td>
</tr>
<tr>
<td>Price²</td>
<td>762,900 won (763 won/GB)</td>
<td>121,000 won (43 won/GB)</td>
</tr>
</tbody>
</table>

¹ Source: [http://www.anandtech.com](http://www.anandtech.com)
² Source: [http://www.danawa.com](http://www.danawa.com) (As of Nov. 24, 2013)
NAND Constraints (1)

- **No in-place update**
  - Require sector remapping (or address translation)

- **Bit errors**
  - Require the use of error correction codes (ECC)

- **Bad blocks**
  - Factory-marked & run-time bad blocks
  - Require bad block remapping

- **Limited program/erase cycles**
  - < 100K for SLCs
  - < 5K for MLCs
  - Require wear-leveling
NAND Constraints (2)

- **Limited NOP (Number of Programming)**
  - 1 / sector for most SLCs (4 for 2KB page)
  - 1 / page for most MLCs

- **Sequential page programming**
  - For large block SLCs and MLCs

- **Pair-page programming in MLCs**
  - Two pages inside a block are linked together
  - Performance difference
  - Interference
FTL (1)

What is FTL?

- A software layer to make NAND flash fully emulate traditional block devices (e.g., disks).

Source: Zeen Info. Tech.
FTL (2)

- Flash cards internals
FTL (3)

- SSD internals

Source: Indilinx
FTL (4)

- For performance
  - Address translation
  - Garbage collection
  - Hot/cold data identification/separation
  - Interleaving over multiple channels & flash chips
  - Request scheduling
  - Buffer management
  - ...

...
FTL (5)

- **For reliability**
  - Bad block management
  - Wear-leveling
  - Power-off recovery
  - Error correction code (ECC)
  - ...

- **Other features**
  - Encryption
  - Compression
  - Deduplication
  - ...
OS Implications (1)

- NAND flash has different characteristics compared to disks
  - No seek time
  - Asymmetric read/write access times
  - No in-place-update
  - Good sequential read/sequential write/random read performance, but bad random write performance
  - Wear-leveling
  - ...

- Traditional operating systems have been optimized for disks. What should be changed?
OS Implications (2)

- SSD support in Microsoft Windows 7
  - Turn off “defragmentation” for SSDs
  - New “TRIM” command
    - Remove-on-delete
  - Align file system partition with SSD layout
  - Larger block size proposal (4KB)
Beauty and the Beast

- NAND Flash memory is 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 (10K~100K program/erase cycles)
  - Bad blocks, ...

- Software support for NAND flash memory is very important for performance & reliability.
Beyond Flash

- Resistance-based memory technologies

<table>
<thead>
<tr>
<th></th>
<th>DRAM</th>
<th>NAND</th>
<th>MRAM</th>
<th>PRAM</th>
<th>ReRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell size</strong></td>
<td>~8F²</td>
<td>~5F² (SLC)</td>
<td>~8F²</td>
<td>~4F²</td>
<td>~8F² (transistor) &lt;1 F² (stack)</td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>xGigabit</td>
<td>xxGigabit</td>
<td>xxMegabit</td>
<td>xGigabit</td>
<td>--</td>
</tr>
<tr>
<td><strong>Latency</strong></td>
<td>~50 ns</td>
<td>20 to ~200 μs</td>
<td>~50 ns</td>
<td>~200 ns</td>
<td>~50 ns to ~1 μs</td>
</tr>
<tr>
<td><strong>Bandwidth</strong></td>
<td>~1 GBps</td>
<td>~100 MBps</td>
<td>~1 GBps</td>
<td>~100 MB/s</td>
<td>~100 MBps</td>
</tr>
<tr>
<td><strong>Vollatility</strong></td>
<td>Volatile</td>
<td>Nonvolatile</td>
<td>Nonvolatile</td>
<td>Nonvolatile</td>
<td>Nonvolatile</td>
</tr>
<tr>
<td><strong>Endurance</strong></td>
<td>&gt;10¹⁵</td>
<td>&gt;10⁵</td>
<td>&gt;10¹⁵</td>
<td>&gt;10⁶ to 10⁻¹²</td>
<td>&gt;10⁵</td>
</tr>
<tr>
<td><strong>Retention</strong></td>
<td>&gt;64 ms</td>
<td>&gt;10 years</td>
<td>&gt;10 years</td>
<td>&gt;10 years</td>
<td>&gt;10 years</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>Working memory</td>
<td>Data storage memory</td>
<td>Working nonvolatile RAM</td>
<td>Code memory and buffer memory</td>
<td>Data storage memory</td>
</tr>
</tbody>
</table>