溫故之新: 하드디스크와 플래시메모리

민상렬 (symin@snu.ac.kr)
(+ Mobile Embedded Systems Lab 연구원들)
서울대학교 공과대학
컴퓨터공학부
Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
HDD internals

- Electronic components
  - Disk controller
  - Disk cache
- Mechanical components
Mechanical components

Arm Assembly

Arm

Head

Spindle

Cylinder

Platter

Track

Source: “ABCs of Disk Drives,” Sudhanva Gurumurthi
Data layout

- Rotating disks consist of **platters**, each with two **surfaces**
- Each surface consists of concentric rings called **tracks**
- Each track consists of **sectors** separated by gaps

Source: "http://camars.kaist.ac.kr/~joon/course/sep562_2006_1/notes/10_11%20Memory_Hierarchy.ppt"
Disk operation

The disk surface spins at a fixed rotational rate

The head is attached to the end of the arm and flies over the disk surface on a thin cushion of air

By moving radially, the arm can position the head over any track

Source: "http://camars.kaist.ac.kr/~joon/course/sep562_2006_1/notes/10_11%20Memory_Hierarchy.ppt"
Disk operation details

Source: "http://www.cs.duke.edu/~chase/cps110/slides/files1.ppt"
Disk operation details

Disk operation details

Disk operation details

Disk access time

- Disk access time
  - Seek time + Rotational latency + Transfer time

- Seek time
  - Time to position heads over cylinder containing target sector
  - 0 ~ 25 ms

- Rotational latency
  - Time waiting for first bit of target sector to pass under r/w head
  - Full rotation: 4 ~ 12 ms (15000 ~ 5400 RPM)

- Transfer time
  - Time to read the bits in the target sector
  - 1 sector transfer: 1.3 ~ 12.8 us (380 ~ 40 MB/s transfer rate)
Electronic components

- Presenting a simple abstract view of the complex sector geometry

![Diagram of electronic components]

- Host interface
- Disk controller
- Disk cache

(cylinder, surface, sector)
Electronic components

- Disk controller
  - Controlling the overall system
  - Major functions
    - Host interface
    - Request translation (LBA \[\leftrightarrow\] [cylinder, surface, sector])
    - Reliability mechanism (e.g. ECC, bad sector handling)
    - Performance improvement (e.g. request scheduling and disk caching)
    - Power management (e.g. spin down of spindle motor)
  - Typically, embedded processor (such as ARM) + logic circuits
Outline

- HDD Basics and Demo
  - Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
Demo HDD Specification

Model Name: SAMSUNG MP0402H (2.5 in)

- Size:
  - total 78,236,550 sectors
  - 40,057,113,600 bytes ≈ 37.30 GB

- Interface: ATA-6 (supports UDMA100)

- Buffer: 8MB DRAM

- Performance brief:
  - Avg. Seek time: 12 ms
  - Avg. Rotational Latency: 5.6 ms (5400 RPM)

- reference url: http://www.samsung.com/Products/HardDiskDrive/SpinPointMSeries/HardDiskDrive_SpinpointMseries_MP0402H_sp.htm
Demo I – Power-on sequence
Demo II – Sequential read/write

- Access pattern
  - read/write data whose address increases continuously
Demo III – Read/Write with a stride

- Access pattern
  - read/write data whose address increases with a regular interval
Demo IV – Read/Write in a convergent manner

- Access pattern
  - read/write data whose address is not overlapped and is in a convergent manner

Flash memory and Advanced Storage Technology group
Demo V - Random read/write

Access pattern
- read/write random addresses

78,236,550
Demo VI - Effect of read caching/write buffering

Access pattern
- Access on the fixed addresses repeatedly

78,236,550
Demo VII – Windows XP start-up
HDD performance trends (1)

- HDD access time trends are fairly flat due to mechanical nature of device.
A workload that was 5% disk bound in ‘96 would be 55% disk bound in ‘05
HDD density trends

Source: Hitachi Global Storage Technologies
HDD Summary

The Good
- High capacity
- Low cost

The Bad
- High latency
- High power consumption
- Low reliability
- Large form factor
- Limited parallelism

The Ugly
- Latent sector errors
Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
Conventional MOS Transistor

\[ \begin{align*}
&n^+ \text{ source (S)} & \quad p\text{-substrate} & \quad n^+ \text{ drain (D)} \\
\end{align*} \]

Schematic symbol
Conventional MOS Transistor: A Constant-Threshold Transistor

\[ I_d \quad V_{gs} \]

\[ V_{th} \]

\[ V_{gs} > V_{th} \]

\[ S \quad R_{on} \quad D \]
Flash Memory

Control gate

Floating gate

Thin tunneling oxide

erasure

programming

n$^+$ source

p-substrate

n$^+$ drain

Schematic symbol
Flash Memory

Erased Cell

Programmed Cell

Control gate

$\text{n}^+\text{ source}$ $p$-substrate $\text{n}^+\text{ drain}$

$\text{n}^+\text{ source}$ $p$-substrate $\text{n}^+\text{ drain}$
Flash Memory: A “Programmable-Threshold” Transistor

- Erased state
- “1” state
- “0” state
- Programmed state

Graph showing the relationship between $I_d$, $V_{th-0}$, $V_{th-1}$, and $V_{gs}$. The graph illustrates the transition between erased and programmed states with respect to gate voltage ($V_{th}$) and drain current ($I_d$). The figure highlights the concept of a programmable-threshold transistor, which can be programmed to different states by applying specific voltages to the control gate. The n+ source, p-substrate, and n+ drain are labeled in the diagram, indicating the different regions of the transistor.
More Bits Per Transistor

NAND Flash Memory Interface

- Read physical page
  - (chip #, block #, page #)
  - ~ 20 us
- Write physical page
  - (chip #, block #, page #)
  - ~ 200 us
- Erase block
  - (chip #, block #)
  - ~ 2 ms
Why (NAND) Flash Memory?

- Advantages of Flash Memory over HDD
  - Low latency
  - Low power consumption
  - Tolerant to shock & vibration
  - Silent operation
  - Small size
  - Abundant parallelism
  - ...

- Single NAND Flash Memory Chip Density Trends

Source: Samsung Electronics
(More) NAND Flash Memory Trends

<table>
<thead>
<tr>
<th>Year</th>
<th>DRAM</th>
<th>NAND Flash</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>$0.97</td>
<td>$1.35</td>
</tr>
<tr>
<td>2001</td>
<td>0.22</td>
<td>0.43</td>
</tr>
<tr>
<td>2002</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>2003</td>
<td>0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>2004</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>2005</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>2006</td>
<td>0.096</td>
<td>0.021</td>
</tr>
<tr>
<td>2007</td>
<td>0.057</td>
<td>0.012</td>
</tr>
<tr>
<td>2008</td>
<td>~0.025</td>
<td>&lt;0.005</td>
</tr>
</tbody>
</table>

CAGR: -32.1% / yr, -50.0% / yr

Source: Lane Mason (Denali Software), “NAND FlashPoint Platform”
## (More) NAND Flash Memory Trends

<table>
<thead>
<tr>
<th>Millions GB</th>
<th>DRAM</th>
<th>NAND Flash</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>30</td>
<td>1.1</td>
</tr>
<tr>
<td>2001</td>
<td>50</td>
<td>1.6</td>
</tr>
<tr>
<td>2002</td>
<td>71</td>
<td>4.6</td>
</tr>
<tr>
<td>2003</td>
<td>98</td>
<td>14.6</td>
</tr>
<tr>
<td>2004</td>
<td>158</td>
<td>68</td>
</tr>
<tr>
<td>2005</td>
<td>240</td>
<td>200</td>
</tr>
<tr>
<td>2006</td>
<td>340</td>
<td>600</td>
</tr>
<tr>
<td>2007</td>
<td>645</td>
<td>1600</td>
</tr>
<tr>
<td>2008</td>
<td>1000</td>
<td>4000</td>
</tr>
</tbody>
</table>

CAGR: +60.0% / yr  
NAND Flash: +150% / yr

Source: Lane Mason (Denali Software), “NAND FlashPoint Platform”
Solid State Disk

- Provides an interface identical to a hard disk, but uses flash memory as a storage medium
Solid State Disk: Form Factor Agnostic

<table>
<thead>
<tr>
<th>Density</th>
<th>Standard FF</th>
<th>Special FF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8”</td>
<td>4~64GB</td>
<td>4~64GB</td>
</tr>
<tr>
<td>2.5”</td>
<td>4~64GB</td>
<td>4~16GB</td>
</tr>
<tr>
<td>1.0”</td>
<td>30x40x4.0</td>
<td>70.6x53.6x:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0: 16/32GB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5: 4~8GB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53.6x70.6x3.0</td>
</tr>
<tr>
<td>Dimension (H x W x T)</td>
<td>78.5x54x8.0</td>
<td>100.2x70x9.5</td>
</tr>
<tr>
<td>Connector</td>
<td>ZIF/IDE 50pin</td>
<td>IDE 44pin</td>
</tr>
<tr>
<td></td>
<td>ZIF 35pin</td>
<td>ZIF 40pin</td>
</tr>
<tr>
<td>Weight</td>
<td>44g</td>
<td>46g</td>
</tr>
<tr>
<td></td>
<td>TBD</td>
<td>20g</td>
</tr>
<tr>
<td>Market</td>
<td>Notebook</td>
<td>Sub-Note / Tablet</td>
</tr>
<tr>
<td></td>
<td>DVC/GPS/ UMPC</td>
<td>UMPC</td>
</tr>
<tr>
<td></td>
<td>Custom</td>
<td></td>
</tr>
</tbody>
</table>

Flash memory summary

The Good
- Low latency
- Low power consumption
- High Reliability
- Small form factor
- Massive parallelism

The Bad
- No in-place updating
- Limited endurance
- Bad blocks
- Write disturbance
- Read disturbance

The Ugly
- Retention errors
- Paired page problem

FROM THE DARK NIGHT
Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
Storage Trends

Tape Is Dead
Disk Is Tape

- 1 TB disks are available
- 10+ TB disks are predicted in 5 years
- But: ~5..15 hours to read (sequential)
  ~15..150 days to read (random)
- Need to treat most of disk as Cold-storage archive

Source: Jim Gray (Microsoft), “Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King”
Storage Trends

Disk Is Tape
Flash Is Disk

- 1995 16 Mb NAND flash chips
  - 2005 16 Gb NAND flash chips
- 2012 1 Tb NAND flash chips
  - == 128 GB chip
  - == 1 TB or 2 TB solid state disk for ~$400
  - or 128 GB solid state disk for ~$40
  - or 32 GB solid state disk for ~$5

Source: Jim Gray (Microsoft), “Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King”
Disk is Tape / Flash is Disk

Poor Reliability
Carnegie Mellon & Google study show up to 8.6% annual failure rate for HDD in controlled environment

Low Performance
Low IOPS performance → High redundancy to compensate for low performance per drive

Heat
Rotating platters & moving heads need power → produces heat

High TCO
Initial purchase cost low, but maintenance, space, cooling & replacement will increase TCO substantially

Disk is Tape / Flash is Disk

- Performance

Source: Jim Gray (Microsoft), “Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King”
Disk is Tape / Flash is Disk

- Power Consumption

Future Outlook

Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
Conclusions

- In the animal world
  - Survival of the fittest

- In the memory world
  - Survival of the fastest or cheapest

<table>
<thead>
<tr>
<th></th>
<th>Volatile</th>
<th>Non-volatile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fastest</td>
<td>SRAM</td>
<td>FRAM?</td>
</tr>
<tr>
<td>Cheapest</td>
<td>DRAM</td>
<td>NAND Flash</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HDD</td>
</tr>
</tbody>
</table>
Conclusions

From the history

<table>
<thead>
<tr>
<th></th>
<th>IBM 360/85</th>
<th>IBM 360/91</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock Rate</td>
<td>80 ns</td>
<td>60 ns</td>
</tr>
<tr>
<td>Memory Speed</td>
<td>1040 ns</td>
<td>750 ns</td>
</tr>
<tr>
<td>Memory Interleaving</td>
<td>4 way</td>
<td>8 way</td>
</tr>
<tr>
<td>Additional Features</td>
<td>Cache Memory</td>
<td>Register Renaming, Out-of-order Execution, etc</td>
</tr>
</tbody>
</table>

But, IBM 360/85 faster on 8 of 11 programs!

The Ultimate Limit – HDD

Boeing 747

2,000,000 Miles Per Hour

1/100” Flying Height


Source: B. Parhami, Dependable Computing: A Multilevel Approach
The Ultimate Limit – Flash Memory

Scanning tunneling microscope image of a silicon surface showing 10 nm is ~20 atoms across

Outline

- HDD Basics and Demo
- Flash Memory Basics and Demo
- Storage Trends
- Conclusions
- (More Demos)
Flash Memory Software Development Platforms

Embedded Platform

Embedded Flash Memory (스프트웨에 솔루션 개발용)

SSD Platform

Solid State Disk (소프트웨어 개발용)

Flash / NV-RAM Modules

- Samsung SLC NAND
- RAMTRON FRAM (serial)
- Samsung MLC NAND
- RAMTRON FRAM (parallel)
- Samsung OneNAND
- FREESCALE MRAM (parallel)
- Hynix MLC NAND
- Samsung Phase-change RAM
Embedded Platform

Features

- For embedded Flash memory software development
- FPGA-based
- NAND slot x 2
- DAQ (Data Acquisition) interface
SSD (Solid State Disk) Platform

Features
- For SSD development
- FPGA-based
- SSD interface (P-ATA, S-ATA)
- NAND slot x 4