SSD Characterization:
From Energy Consumption's Perspective

Youjip Won

Hanyang University
Outline

- Motivation
- Related Works
- SSD Organization and Energy Consumption
- Channels, Ways and Clusters
- Trivia in Measurement Methodology
- Case Study
- Power Budget
- Conclusion
Motivation
Understanding of Internal Mechanism of Storage Device is very important!

- Hard disk
  - Sector layout: cylinder serpentine vs. surface serpentine vs. hybrid serpentine
  - Number of zones
  - Degree of track skew
  - Disk scheduling algorithm
Characterizing HDD: Sector Layout

- Jongmin Gim et al, ACM ToS 6, 2 (July 2010)

![Diagram showing different types of HDD sector layouts](chart.png)

Hitachi HDT7725032VLA360 for 2000 tracks

- Traditional
- Surface serpentine
- Cylinder serpentine
- Zone rewind

![Graph showing time vs track number](graph.png)
HDD Characterization is via measuring **Seek time** and **Rotational Latency**.

**Characterizing SSD... what do we use?...**
Motivation

How do we figure out the internals of SSD?

- What is available
  - the number of channels
  - the number of chips/packages per channels

- What is not available?
  - Sector placement, Garbage collection algorithm
Operations on NAND Flash Cell

- **Program**
  - Control Gate
  - Floating Gate
  - Oxide Layer
  - 18~35V
  - Source
  - Drain

- **Erase**
  - Control Gate
  - Floating Gate
  - Oxide Layer

**Programmed:** 1 → 0

**Erased:** 0 → 1
Operations on NAND Flash Cell
SSD Characterization

We will use “Energy Consumption”
Related Works
Related Works

- Dongkun Shin et al, NVRAMOS 2010 Spring
  
  Identifying the relationship between workload characteristics and aggregate power consumption for each workload
  
  Applied mixed workloads (random, sequential, etc) to SSDs with different request sizes and varied the file systems
  Measured the Power Consumption (measured voltage change)

- Laura M. Grupp et al, MICRO 2009

  Custom Board + Flash Memory

  Basic Operation

  Read
  Program
  Erase

  Power Consumption of Flash Memory Basic Operations

  Examine the average power consumption (W) and energy (J)
Related Works

- Euiseong Seo et al, HotPower’08
  
  Analyzed the power consumption patterns of the SSDs

- Vidyabhushan Mohan et al, Date ’10

  CACTI 5.3

  developed a detailed power model for the NAND flash chip itself with CACTI 5.3

  different hardware configurations to the various atomic operations and the combination of file systems and workloads.
SSD Organization and Energy Consumption
Power Consumption of Storage Devices: System Boot

Youjip Won  Hanyang University  NVRAMOS’11, Jeju, Korea

Active current of HDD
stand-by current
SSDs have various stand-by currents
Characteristics of SSD Behavior

- Yellow = Page Write
- Blue = Write Complete

[Diagram showing SSD behavior with page write and write complete states]
IO Size vs. Energy Consumption

![Graph showing IO Size vs. Energy Consumption for Intel X25M](image)

Legend:
- 4kbyte
- 8kbyte
- 16kbyte
- 32kbyte
- 64kbyte
- 128kbyte
- 256kbyte
- 512kbyte

X-axis: Time (ms)
Y-axis: mA

Way Depth:
- 1CH
- 2CH
- 4CH
- 8CH
- 10CH

Youjip Won
Hanyang University
NVRAMOS’11, Jeju, Korea
Power consumption vs. Channels, Ways, and Clusters
Read vs. Write

- **Read**: Page Operation
- **Write**: Operation Complete

**Read Operation**

- Current (mA)
- Time

**Write Operation**

- Current (mA)
- Time

---

Youjip Won  Hanyang University  NVRAMOS’11, Jeju, Korea  20/68
Writing 4 pages: 1 Channel X 1 Way

- Page Write
- Write Complete

Current (mA) vs. Time

C
D
P

Youjip Won
Hanyang University
NVRAMOS’11, Jeju, Korea
Way switch vs. Channel switch Delay

Way Switch vs. Channel Switch Delay

Current (mA) vs. time

Way Switch

Channel Switch

C

D

P

Way Switch

Channel Switch

22/68
Writing 4 pages: 1 Channel X 2 Way

- Yellow box = Page Write
- Blue box = Write Complete

Diagram shows a process timeline with time (in mA) and a series of steps labeled C, D, and P, indicating different stages of the writing process.
Writing 4 pages: 2 Channel X 1 Way

=Page Write
=Write Complete

\[ I(t) \]

\[ P \text{=Page Write} \]
\[ C \text{=Write Complete} \]

Current (mA)

Time

\[ 24/68 \]
Writing 4 pages: 2 Channel X 2 Way

- Page Write
- Write Complete

[Diagram showing writing process with channel and way switches]
Cluster

- Cluster: Write Unit of SSD

- Page Write

- 4KB Write

- 8KB Write
Trivia in Measurement Methodology
Trivia of Measurement Methodology

Sampling interval should be smaller than Read/Program Latency

- Smoothing the data to filter out measurement noise.
Trivia of Measurement Methodology

- Deactivate DRAM cache
  - SATA command 82h
    - Samsung MXP and Intel X-25M SSDs
  - Trigger DRAM flush
    - OCZ Vertex and Hanamicron Forte+
Case Studies
**Measurement**

<table>
<thead>
<tr>
<th>Model</th>
<th>Vendor</th>
<th>Size</th>
<th>Channels</th>
<th>DRAM Size</th>
<th>Package</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>X25M</td>
<td>Intel</td>
<td>80GB</td>
<td>10</td>
<td>16MB</td>
<td>20</td>
<td>MLC</td>
</tr>
<tr>
<td>MXP</td>
<td>SAMSUNG</td>
<td>128GB</td>
<td>8</td>
<td>128MB</td>
<td>16</td>
<td>MLC</td>
</tr>
<tr>
<td>Vertex</td>
<td>OCZ</td>
<td>60GB</td>
<td>8</td>
<td>64MB</td>
<td>16</td>
<td>MLC</td>
</tr>
<tr>
<td>Forte+</td>
<td>Hanamicron</td>
<td>32GB</td>
<td>8</td>
<td>32MB</td>
<td>8</td>
<td>MLC</td>
</tr>
</tbody>
</table>

- Oscilloscope (TDS3032)
- High resolution current probe (TCP202)
- Current probe to power line (Vdd) of the SSD
- Sampling interval (10 samples MA): 10usec
- Each request measured 10 times
Case Study: Intel X25M

Intel X25M

<table>
<thead>
<tr>
<th>Capacity</th>
<th>80GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Channels</td>
<td>10</td>
</tr>
<tr>
<td>Packages/Channel</td>
<td>2</td>
</tr>
<tr>
<td>Package</td>
<td>4 GB</td>
</tr>
</tbody>
</table>

NAND (4GB)
Case Study: Intel X25M Write

- IO Size: 4KB to 80KB

- Channel switch = 30 μsec
- NAND programming: 17 mA
- Simple Round Robin

- Increase 20 steps

Graph:
- mA vs. Time (ms)
- 600 μsec
- 17 mA
- 100 mA
Case Study: Intel X25M Write

- 4KB to 80KB Sequential Write

![Diagram showing page write and channel switch]

Channel Switch = 30 μsec

MUX

0 1 2 3 4 5 6 7 8 9
10 11 12 13 14 15 16 17 18 19
Case Study: Intel X25M Write

- IO Size: 80KB to 160KB Write

![Graph showing performance metrics for Intel X25M Write with specific points indicating 80 KByte and 84 KByte.]
Case Study: Intel X25M Write

- =1 Page Write
- =Write Complete

MUX

Flow

Current

Youjip Won  Hanyang University

NVRAMOS’11, Jeju, Korea
Case Study: Intel X25M Read

- IO Size: 4KB to 16KB Read

[Graph showing IO size and peak current]
Case Study: Intel X25M Read

- IO Size: 16KB to 28KB Read

![Graph showing IO size and power consumption over time for different byte sizes.]
Case Study: Intel X25M Read

- IO Size: 28KB to 48KB Read

![Graph showing IO Size and Time](image)

- Same Peak Current
Case Study: Intel X25M Read

- IO Size: 48KB to 52KB Read

![Graph showing current consumption over time for 48Kbyte and 52Kbyte reads. The peak current is the same for both sizes.]
Case Study: Intel X25M Read

- IO Size: 56KB to 72KB Read

![Graph showing read performance of different IO sizes](graph.jpg)

Same Peak Current
Case Study: Intel X25M Read

- IO Size: 76KB to 92KB Read

![Graph showing IO Size and Peak Current comparison for different data sizes.](image)

- Same Peak Current for different data sizes.
Case Study: Intel X25M Read

- IO Size: 96KB to 100KB Read

![Graph showing IO size comparison](image)

- Same Peak Current
## Case Study: Intel X25M Summary

<table>
<thead>
<tr>
<th>Cluster Size</th>
<th>Write</th>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>4KB</td>
<td>48KB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>17mA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Peak Current</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>500mA</td>
<td>180mA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel switch</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>30μsec</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Policy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>Round Robin</td>
</tr>
</tbody>
</table>

- Different from Write, peak current remains the same.
- Only duration changes.
Case Study: SAMSUNG MXP

Samsung MXP

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>128GB</td>
</tr>
<tr>
<td>No. of Channels</td>
<td>8</td>
</tr>
<tr>
<td>Packages/Channel</td>
<td>2</td>
</tr>
<tr>
<td>Package</td>
<td>8 GB</td>
</tr>
</tbody>
</table>

NAND (8GB)
Case Study: SAMSUNG MXP

- Write Operation: From 4Kb to 32KB

Peak Current

![Graph showing peak current variations with different data sizes](image)

- 50mA
- 100mA

Time (ms)
Case Study : SAMSUNG MXP

- Write Operation : From 36Kb to 64KB

![Graph showing same peak current for different write operations from 36Kbyte to 64Kbyte]
Case Study: SAMSUNG MXP

- Write Operation: From 68Kb to 96KB

Graph showing the peak current for different write operations in the range of 68Kb to 96KB. Each operation is represented by a line color indicating the respective byte size.
**Case Study : SAMSUNG MXP**

- **Write Operation**: 100Kb to 112KB

```
Cluster Size of MXP is 32KB
```

```
Same Peak Current
```

```
49\/68
```

```
Write Operation : 100Kbyte to 112Kbyte
```

```
Cluser Size of MXP is 32KB
```

```
mA
```

```
Time(ms)
```

```
0 0.5 1 1.5 2 2.5 3
```

```
100Kbyte   104Kbyte   108Kbyte   112Kbyte
```

```
0 50 100 150 200 250 300 350 400
```

```
Youjip Won Hanyang University NVRAMOS'11, Jeju, Korea 49/68
```
What we still do not know of

Peak of 128 Kbyte < Peak Current of 124 KByte
Case Study: OCZ Vertex

OCZ Vertex

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>60GB</td>
</tr>
<tr>
<td>No. of Channels</td>
<td>8</td>
</tr>
<tr>
<td>Packages/Channel</td>
<td>2</td>
</tr>
<tr>
<td>Package</td>
<td>4 GB</td>
</tr>
</tbody>
</table>

NAND (4GB)
Case Study: OCZ Vertex

- write-caching off, 1KB write, 10 counts

![Graphs showing current over time with different interval times: 200ms, 20ms, 100ms, 200ms, and 500ms intervals.](image)
Case Study: OCZ Vertex

64MB Write Cache is filled with 64MB write request

Test I/O Size

< 100ms

Cache Flush

1Mbyte

Test I/O Size

Time(s)

mA

0 0.2 0.4 0.6 0.8 1

50 100 150 200 250 300 350 400
Case Study : OCZ Vertex

- IO Size: 4KB to 32KB

- Peak Current Change: 160mA
- Same Peak Current: 120mA

Cluster Size of Vertex is 16KB
Case Study: OCZ Vertex

- write-caching off, 1counts, IO Size: 64k to 512k

![Graph showing power consumption over time for different data sizes: 64Kbyte, 128Kbyte, 256Kbyte, and 512Kbyte. The graph shows peaks at approximately 350mA and 220mA.]
Case Study : OCZ Vertex

- IO Size: 4KB to 4096KB Write

![Graph showing power consumption over time for different IO sizes.](image-url)
### Case Study: Hanamicron Forte+

**Hanamicron Forte+**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>32GB</td>
</tr>
<tr>
<td>No. of Channels</td>
<td>8</td>
</tr>
<tr>
<td>Packages/Channel</td>
<td>1</td>
</tr>
<tr>
<td>Package</td>
<td>4 GB</td>
</tr>
</tbody>
</table>

**NAND (4GB)**

![Diagram showing Hanamicron Forte+ structure]
Case Study: Hanamicron Forte+

- IO Size: 4KB to 32KB

- Peak Current Change
- Same Peak Current

Cluster Size of Forte+ is 16KB

Size of cluster same with OCZ

40mA

Time (ms)
Case Study: Hanamicron Forte+

- IO Size: 4KB to 512KB Write

Graph showing IO size from 4KB to 512KB and corresponding current consumption at different time intervals.
Comparison: Write Energy

- Energy Consumption

![Graph showing comparison of write energy consumption across different I/O sizes for various brands: X-25, MXP, OCZ, Hanamicron. The graph plots energy consumption in mJ against I/O size in KB.]
Comparison: Peak and Duration of 512K Write

- Samsung
- OCZ
- Hanamicron
- Intel

- 500mA
- 350mA
- 220mA
Forth Coming Problem in Multi-channel SSD

**Model Name** | **Release**
--- | ---
SATA3035 (Mtron) | 2008.01
Vertex (OCZ) | 2009.03
Vertex2 (OCZ) | 2010.07
REVO Drive X2 (OCZ) | 2011.01
Forth Coming Problem in Multi-channel SSD

- 10 channel: peak 500 mA
- 16 Channel: peak 800 – 900 mA → SSD is no long Green.
- Further, excessive peak current can cause...
  - supply voltage drop
  - ground bounce
  - signal noise
  - black-out
  - Etc...

Supply voltage drop
Ground bounce
Signal noise
Black-out
Etc...
**Power Budget**

- **8 Channels X 1 Way**
  - Power Budget
  - Feasible Solutions
- **4 Channels X 2 Ways**
  - Performance Bound
  - Feasible Region
- **2 Channels X 4 Ways**
- **1 Channel X 8 Ways**

Graph showing the relationship between I/O latency and power budget for different configurations.
## Summary

<table>
<thead>
<tr>
<th></th>
<th>X25M</th>
<th>MXP</th>
<th>OCZ</th>
<th>Hanamicron</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Size</td>
<td>4KB</td>
<td>32KB</td>
<td>16KB</td>
<td>16KB</td>
</tr>
<tr>
<td>programming</td>
<td>17 mA</td>
<td>35 mA</td>
<td>15~20 mA</td>
<td>15~20 mA</td>
</tr>
<tr>
<td>Peak Current</td>
<td>500 mA</td>
<td>350 mA</td>
<td>350 mA</td>
<td>220 mA</td>
</tr>
<tr>
<td>Channel switch</td>
<td>30 μsec</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Energy
- Performance
- Small Write
- Large write
- Standby current
Energy Consumption is a very good tool to characterize SSD.

For larger number of channels, peak current will soon be a significant issue.

We introduce the notion of Power Budget to resolve this issue.