The 2\textsuperscript{nd} Era of Flash-based Storage Device (SSD): Trends, Opportunities and Technical Challenges

2011.11.9

Kyung Ho Kim (Senior Engineer)
Flash S/W Development
Memory Business / Device Solution
Samsung Electronics
Contents

I Introduction: IT Trends

II The Value of 1st Era Flash Storage (SSD)

III New Challenges of 2nd Flash Storage (SSD)
Exponential data traffic/storage growth in our digital universe via new IT devices / networking applications

- Mobile Traffics: 1,500 PB/M (@’11) → 25,000 PB/M (@’15) (x17 ↑)
- IT Storages: 1.68ZB (@’11) → 35.2ZB (@’20) (x21 ↑)

[Source: Cisco, Feb’11]

1 ZB = 1 Trillion Gigabytes

[Source: IDC Digital Universe Study, May’10]
Expanding conventional infra brings increasing on three types of major concerns.
  - This way should be considered later because of budget concern

1. New investment for new equipment
   - Enterprise system consists of server, storage, network switch, UPS, etc.
   - System or other instruments price is too high to expand on plan
   - Now Cloud Computing is mega trend

2. Operating cost (TCO) revisited
   - Power dissipation for various instrument, cooling and UPS
   - Fixing failed system and device (Sanitization, Recovery Services..)
   - Electricity cost, Maintenance cost

3. Space
   - Rental fee of datacenter is not less than system operating fee
   - More reduced equipment, less spent money (likely Google …)
   - $750/month rental fee for one rack space
- Upgrading system requires less investment than buying system
  - Comparing three major component affecting system performance, easily can know HDD is the weakest point of those components

- In real, the perf. gap between “CPU & Memory” and “HDD” has been getting worse → Now HDD is Performance Bottleneck.
IT Trends – Getting Smaller and New FFs

Series 9 Notebook

- Duralumin, the New Definition of Lightweight
- See Vivid Content, Wherever You Go
- Powerful Processing
- Power Back up in 3 Seconds with FastStart

$1649.99

Specifications

- Processor: Intel® Core™ i5 Processor 2537M
- Operating System: Genuine Windows® 7 Home Premium (64b)
- Graphics: Intel® HD Graphics 3000
- Storage: 128 GB Solid State Drive
- Multimedia: 3 W Stereo Speaker (1.5 W x 2) with HD Audio
- Multimedia: 1.3 MP HD Webcam
- Dimensions: 12.9" (W) x 8.9" (D) x 0.62" – 0.64" (H)

MacBook Air

The new MacBook Air

The next generation of MacBooks.

11-inch : 128GB

- 1.4GHz Intel Core 2 Duo processor
- 2GB memory
- 128GB flash storage¹

NVIDIA GeForce 320M graphics

All-Flash Storage

Designed exclusively around flash storage, MacBook Air is fast, reliable, and snaps to life in an instant.
The most radical innovation occurs on storage system because other components are already at acme.

Adopting SSD is the best choice to upgrade system performance as resolving performance bottleneck.
I  Introduction: IT Trends

II  The Value of 1st Era Flash Storage (SSD)

III  New Challenges of 2nd Flash Storage (SSD)
SSD is Solid State Drive

- Solid State Drive is a large capacity of Storage using NAND Flash Memory as its media

- For interface wise, SSD is using S-ATA/SAS/FC interface for compatibility to conventional industry and also considering PCIe interface for lower latency
SSD is growing steadily in its all application fields

- At 2014, Set Shipment is 78M pcs and Total Revenue is $8.5B
- 15% of notebooks and 10% of servers plan to adopt SSD in 2014

Source: Gartner Q4 ‘10
SSD Values: Reliability

- **High Reliability/ High Roughness / No Acoustic** due to simple and strong composition of SSD
  - Maintenance cost will be down because of lower probability of failure

### Environmental Spec

**HDD vs. SSD**

- **MTBF**
  - 1.5Mhrs vs 2Mhrs

- **Shock**
  - 60G vs 1500G

- **Vibration**
  - 1.2G vs 20G

- **Acoustics**
  - 3.1Bels vs 0Bels

*Source: Available datasheet
  *G: gravity
  *Bel: sound power
SSD Values : Power Dissipation

- SSD saves more power than 15Krpm SAS HDD as a result of no-moving part
  - In case of active mode, 55% power saving
  - In case of idle mode, 76% power saving

[Source : Storage review.com]
- SSD shows max. 110 times higher random performance comparing to conventional 15Krpm SAS HDD
SSD Values: Efficiency - IOPS/Watt

- SSD grants max. 240 time higher random performance to power efficiency

[Source: Storage review.com]
Replacing 15Krpm SAS HDD with SSD, system shows minimum 2 times benefit.

The benefit is getting wider as number of user is increased.

**TPC-C Benchmark Test**

<table>
<thead>
<tr>
<th>Performance [tpmC]</th>
</tr>
</thead>
<tbody>
<tr>
<td>150000</td>
</tr>
<tr>
<td>100000</td>
</tr>
<tr>
<td>50000</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

- **8ea 15K SAS HDD**
- **2ea 10GB SSD**

**System Configuration**
- HP DL380G6
- OS: MS Winserver 2008 Enterprise 64bit
- DBMS: SQL Server 2008 R2
- 8GB Mem

- 15Krpm SAS HDD 8ea (RAID50)
- Samsung SS1605 SSD2 2ea (RAID1)
SSD Values in System: Web Service

- SSD based server can enhance work efficiency up to 60%.

SSD Benefit from SPEC Benchmark Test

HDD vs. SSD: Power efficiency comparison on web server

SSD-based system opens 326 page views per 1 watt for e-commerce and e-banking, while
HDD-based system can open 203 page views per 1 watt under 24/7 operation.

- With HDD: 204 Pages/Watt
- Power efficiency: 60% up

- With SSD: 327 Pages/Watt

Source: Standard Performance Evaluation Corporation (SPEC.org)
Assume: E-commerce and e-banking server are accessed for 24hrs.
SSD based server saves job completion time up to 90% faster

Test Condition:
Model: HP DL385 G7
Storage Option: 1) 120GB Samsung SSD 2) 146GB 15K rpm Enterprise HDD
Test Process: 1) Executing 4 of virtual machine on one HP DL385 G7 2) Executing SQL workload for 290Gb database on each virtual machine 3) Comparing job completion time between SSD based server and HDD based server
SSD TCO: “Instantaneous Break-even”

15K SAS HDD 2.5"
- Average IOPS: 410
- Active Power: 8.3W

SSD SATA 2.5"
- Average IOPS: 39,000
- Active Power: 3.7W

- IOPS per W: 49
- TPC-C: 18503
- 1 SSD can Replace up to 20 15K Hard Drives...

Source: Storagereview.com, March, 2009, IOPS calculated by IOMeter File Server average 1 I/O to 128 I/O w/ RR70% and RW30%
Source: HP website / www.hp.com/go/thermallogic)
Contents

I  Introduction: IT Trends

II  The Value of 1st Era Flash Storage (SSD)

III New Challenges of 2nd Flash Storage (SSD) - Internal Issues
1. Endurance(1) : De-duplication

- Traditionally, the de-duplication is widely adopted in the server storage layer.
- Currently, CAFTL uses Fingerprint\[^1\][^2\] (SHA-1, MD5 and etc) for De-duplication.

**Issues**
- How to manage the duplicated data? How about dealing de-dup. when GC?
- Is Duplicated data similar to cold data? Is Fingerprint(SHA-256) safe?

\[^1\] Feng Chen, etc., “CAFTL: A Content-Aware Flash Translation Layer Enhancing the Lifespan of Flash Memory based Solid State Drives”, FAST 2011
\[^2\] Jonghwa Kim, etc., “Deduplication in SSD for Reducing Write Amplification Factor”, FAST 2011
There are few researches about mixing H/W Comp. and Dedup.

De-duplication algorithms & Compression algorithms have the similar mechanism

- EX : Broder’s Delta-encoding (dedup.) is similar to the mechanism of Deflate Encoding (LZ77 - comp.). The intermediate code-words of the Deflate Encoding can be used for the delta-encoding.
1. Endurance(3) : Research Tip (Hot/Cold Separation)

- **The limitation of the previous Hot/Cold Separation Method**
  - They depend on Time Stamp\(^1\), Counter, bloom filter \(^2\)\(^3\)
  - They are weak in Adaptability on workload.

- **Issues**
  - How to separate hot/cold pages adaptively depending on the workload?
  - What is the best off-line method?

---

The distribution of the invalid pages depends on the workload.

---

The # of Invalid Pages vs. Creation Time of blocks

- Random Write
- Random With Hot data

WAF = 3.338
WAF = 2.706539

---

[7] “Janus–FTL finding the optimal point on the spectrum between page and block mapping schemes”, 2010
Contents

I  Introduction: IT Trends

II  The Value of 1st Era Flash Storage

III New Challenges of 2nd Flash Storage (SSD) - Applications
System-Wide View of Cloud Computing Services

1. Storage Class Memory (SCM)
   - Tech. for Mass Production
   - System & OS architecture

2. NVM Aware S/W Stack

3. Directly Attached SSD

4. NVM Based Storage Architecture

5. Management S/W
   - Cooperative De-duplication/Compression

6. Enhanced NAND Cell Characteristics
   - Endurance
   - Performance
   - Power

User Devices
- Tablets
- Laptops

Storage as a Service (SaaS)
- Object Storage Service
- Block Storage Service
- Database Service

Computing Server
- Application/Middle-Ware
- OS DBMS
- VFSL (Virtualized Flash Storage Layer)

Storage Server
- NVM Array Controller
- NVM Module Array
- SSD Array

References:
[2] Jay Prass, “Block Storage as a Service (BSaaS) within the Cloud”, Flash Memory Summit 2011
[3] Violin Memory

Align with your imagination
SSD software stack is evolving from HDD emulation to vertically optimized one for pursuing system balance.

(a) Traditional block interface
(b) Block interface with SSD extensions
(c) Simplified SSD interface
(d) Native flash interface
(e) Object storage device interface

- Trim command in Windows 7
- PCIe SSD from Fusion IO
- ClearNAND from Micron
- Future arch. under research
By virtue of NAND flash memory, the storage I/O is not any more the bottleneck of the system.

Storage Device can take part of the some work, which was done in the server side CPU conventionally.

1. Acceleration of DBMS operation

2. Acceleration of Data Mining for Hadoop distributed file system
ISP can be a very promising scale-out solution for the next generation data-intensive computing paradigm in terms of performance, cost and power.

- In-Host Processing
- CPU: Device CPU (ARM)
- hw: Hardware Acceleration per Channel

**Modeling in SoC designer**

![Diagram showing Embedded CPU, DRAM, Scan HW, Join SW, 8-channel FMCs, 8-channel flash chip array]

**Throughput Comparison: SCAN**

Throughput (# of scanned records/s)

- IHP-ch8
- IHP-ch16
- cpu-ch8
- cpu-ch16
- hw-ch8
- hw-ch16

**Energy Consumption Comparison**

<table>
<thead>
<tr>
<th></th>
<th>String search</th>
<th>Nested block loop join</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP (modified firmware)</td>
<td>0.142</td>
<td>0.134</td>
</tr>
<tr>
<td>IHP (conventional)</td>
<td>1.00</td>
<td>7~8x Reduction 1.00</td>
</tr>
</tbody>
</table>

- Project with Samsung
Q&A
Thank you


* Kyung Ho Kim, etc., “ System–Wide Issues for Efficient use of enterprise SSD”, NVRAMOS 2011 Spring