The material contained in this presentation is copyrighted by the SNIA and is a compilation of slides extracted out of several SNIA tutorials.

Member companies and individuals may use this material in presentations and literature under the following conditions:

- Any slide or slides used must be reproduced without modification
- The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.

Neither the Author nor the Presenter is an attorney and nothing in this presentation is intended to be nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.

The information presented herein represents the Author's personal opinion and current understanding of the issues involved. The Author, the Presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

**NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.**
Agenda

- Computer and Storage Evolution
- What is Solid State Storage?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues
- Q&A
### Machine Specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>8088</td>
</tr>
<tr>
<td>Speed</td>
<td>4.77 MHz</td>
</tr>
<tr>
<td>RAM</td>
<td>16KB</td>
</tr>
<tr>
<td>Storage</td>
<td>Cassette Tape, optionally 5.25&quot; 160KB floppy drives</td>
</tr>
<tr>
<td>Expansion</td>
<td>5 expansion slots</td>
</tr>
<tr>
<td>Bus</td>
<td>Industry Standard Architecture (ISA)</td>
</tr>
<tr>
<td>Video</td>
<td>Initially CGA (320x200x16 color, 640x200x2 color) or monochrome (80x25 text only))</td>
</tr>
<tr>
<td>I/O</td>
<td>Parallel, Serial</td>
</tr>
<tr>
<td>OS</td>
<td>Basic 1 (ROM)</td>
</tr>
<tr>
<td>Killer App</td>
<td>VisiCalc</td>
</tr>
</tbody>
</table>
Fast Forward – to Today

- Today, we have CPUs which are 1,000x faster
  - Instead of MHz, we have GHz
  - Instead of one core, we have multi-core
- Today, we have RAM which is 1,000,000x larger
  - Instead of KB, we have GB
  - Some machines are approaching TB (!)
- Today, we have storage which is 1,000,000x deeper
  - Instead of MB, we have TB
- So what’s the problem?
In a perfect world, I/O would not be necessary
- 1st level store would hold everything, forever

Access Density – IOPS/GB
- Getting WORSE over time for rotating magnetic
- Will it get worse over time for non-rotating SSD?

Access Density Example (more is better):
- August 1981 – 625 KB/s, 8.33 ms, 3,600 RPM, 20 IOPS
  - \(\text{IOPS/GB} = \frac{20}{0.001} = 20,000\)
- Today – 170,000 KB/s, 2.9 ms, 15,000 RPM, 250 IOPS
  - \(\text{IOPS/GB} = \frac{250}{300} = 0.833\)
Why Solid State Disks?
Little Change in HDD RPM

- 1956 RAMAC
  - the first disk drive
  - 5 MB storage
  - 1,200 RPM

- 2007 SCSI Hard Drive
  - 400 GB storage
  - 15,000 RPM

- From 1956 to 2007:
  - 12.5 times increase in RPM
  - 80,000 times increase in capacity
Agenda

- Computer and Storage Evolution
- What is Solid State Storage
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues
- Q&A
Solid State Storage – What is it?

- Storage devices constructed from chips instead of rotating platters or streaming tape
- Various form factors
  - HDD forms using HDD interfaces (e.g. FC, SATA)
  - Non-HDD forms such as PCI-Express cards
- Two basic technologies:
  - DRAM
    - Usually accompanied by battery/persistent media
  - NAND Flash
    - Other Flash types are becoming less viable
What is a Solid State Disk?

- **Flash Memory Based**
  - Same class of memory used in consumer electronics
  - Inherently non-volatile
  - Best known for ruggedness and good random read performance.

- **DDR Memory Based**
  - Same memory used in enterprise servers
  - Requires batteries and backup hard disks for non-volatility
  - Best known for outstanding performance and high cost.

- **Cached Flash**
  - Mix of DDR RAM and NAND Flash.
SSS vs. HDD

Images of HDD and Representative SSS

Photo provided by Intel Corp 2008
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues
- Q&A
Flash SSD Characteristics

- **Consistent Characteristics**
  - Small block random read performance is excellent versus HDD
  - Read / write performance is asymmetric
    - Writes slower than reads
  - Sequential performance today not much better than enterprise HDD
  - Power less than HDD
  - More expensive per capacity than HDD / less expensive per IOPS
The process to write data to flash follows these steps:

- Determine “block” to update. NAND flash is typically divided into 128KB blocks and further subdivided into 2KB pages.
- Copy data from the existing “block” (if necessary)
- Erase the “block” and reset cell to all “1’s”
- Rewrite the “block”, only 0’s can be written to a flash cell.
Flash Memory Characteristics

Two types of NAND Flash Memory

- SLC – single layer
  - 100,000 writes per cell
  - Primarily used in industrial and military applications
  - Higher cost
  - 1.5 millisecond erase times; 200 microsecond write times; 25 microsecond read times
  - Maximum density – 16Gbit with 32Gbit on the way
  - SLC memory is best suited for the enterprise

- MLC – multi layer
  - 10,000 writes per cell
  - Primarily used in consumer electronics
  - Lower cost
  - Half the performance of SLC
  - Maximum density – 32Gbit with 64Gbit on the way.

- Expect mixed SLC – MLC flash SSDs in the near future.
DDR RAM Characteristics

- Unlimited writes per cell
- Primarily used as computer memory
- Higher cost and lower density than NAND flash
- Inherently volatile
- Highly reliable chip design
- 10-15 nanoseconds read and write times
- Maximum density 1Gbit (2Gbit is being sampled)
- DDR RAM is well suited for the enterprise.
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues
- Q&A
Why Solid State Disks? 
Latency Matters

“Money can buy bandwidth, but latency is forever”

John R. Mashey, Chief Scientist SGI, 
“Big Data and the Next Wave of InfraStress”, USENIX, 1999
Why Solid State Disks?
Data Access Times

Data Access Times
(assumes a cache-miss)

Milliseconds

Read
Write

HDD
4

Flash
2

0.2

0.02

DDR RAM

(4)
Why Solid State Disks?
Random I/O’s

Random I/O's Per Second
(assumes a cache-miss)

I/Os Per Second

<table>
<thead>
<tr>
<th></th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDD</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>Flash</td>
<td>100,000</td>
<td>20,000</td>
</tr>
<tr>
<td>DDR RAM</td>
<td>400,000</td>
<td>400,000</td>
</tr>
</tbody>
</table>
Why Solid State Disks?
Low Power

Relative Comparison of Storage Mediums (Tiers) For Similar Capacities

Why Solid State Disks? Dropping Prices

Average Price per GB Comparison - SSD and HDD by Form Factor

- Flash SSD
- 1.8in HDD
- 2.5in HDD
- DRAM SSD

Note: DRAM SSD include additional components (i.e. batteries, etc)

Why Solid State Disks?
Low Price for Performance

![Price / IOPS Chart]

- DDR: $0.20
- Flash: $0.84
- HDD: $2.00

Price per Random IOPS

Solid State Storage Overview
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM characteristics
- Why SSS/SSD?
- SSS Market
  - System Architectures using SSS
  - SSS Reliability and Endurance
  - SSS Performance
  - Some SSS Performance Issues
- Q&A
Growth in the Enterprise SSD Market


Units (kU) vs. Revenue ($M)

- 80% CAGR
- 217% CAGR

2007 2008 2009 2010 2011 2012

Units
Revenue

Solid State Storage Overview
Impact of Solid State Storage

- Server CPUs today are multicore, GHz
- Millions of CPU cycles during just one HDD seek
- SSS can potentially eliminate waste:
  - Server infrastructure – reduce CPU load/wait time
  - Storage infrastructure – reduce short-stroking HDDs
  - Applications – reduce I/O bounded-ness
  - Humans – reduce wait time, screen refreshes, queries
  - Data Centers – reduce power, cooling load, rackspace
- Reliability – a whole other topic …
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- **System Architectures using SSS**
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues
- Q&A
Architectural Approaches in SSS

- Use of internal techniques such as striping across multiple chips – parallel access – even more IOPS

![Diagram showing the architecture with DRAM, Controller, 32 NAND chips, and I/O Bus with many potential data paths.]
Architectural Approaches in SSS

- Servers – take advantage of existing busses/interconnect
- Brings SSS closer to the CPU/RAM, architecturally
Architectural Approaches in SSS

- Standalone – take advantage of specifically isolating SSS
- Brings SSS to bear on a single server/compute job
Storage Performance Pyramid

- SSD
- Cached Flash
- Flash
- Cached RAID
- RAID

Performance
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- **SSS Reliability and Endurance**
- SSS Performance
- Some SSS Performance Issues
- Q&A
The GOOD:
- No moving parts
- Catastrophic device failures are rare (post infant mortality)

The BAD:
- Relatively high bit error rate, increasing with wear
  - MLC wear rate (higher capacity density) worse than SLC
  - Higher density NAND Flash will increase bit error rate
- Program and Read Disturbs

The UGLY:
- Partial Page Programming
- Data retention is poor at high temperature
- Infant mortality is high (large number of parts…)
Controller Reliability Management

- Wear leveling & Spare Capacity (e.g. Spare Blocks)
- Read & Program Disturb Controls
- Data & Index Protection
  - ECC Correction
  - Internal RAID
  - Data Integrity Field (DIF)
- Management

Poor Media + Great Controller → Great SSS Solution
Disc (Drive) Virtualization

Physical disc drive

Disk Virtualization

Logical data layout
- Logical Block Addresses (LBA)
- ‘Defect-Free’

LBA
000
001
002
003
004
005
006
.. 
nnn
Memory Cell Wear Out

Higher BER

< Data Integrity

< Life
Error Rate Decreases with Increase in Temperature but Data Retention Decreases with Temperature
Erase Cycle Effect on Read Disturbs

![Graph showing Read Disturbs Bit Errors](Image)

- Read BER after 0 Cycles
- Read BER after 2K Cycles
- Read BER after 4K Cycles
- Read BER after 6K Cycles

Bit Error Rate (BER) vs. Read Cycles
SSS Life Extension Mechanisms

- Robust ECC
- RAID (Internal)
- Wear leveling – Reduce hot spots
- Wear-out prediction – Planned maintenance
- Write amplification avoidance
- Garbage collector efficiency
ECC Extends the PEB Life (MLC)
Probability of Exceeding ECC (MLC)
Flash SSD Wear Life

Factors affecting wear life

- **Flash Technology**
  - SLC – 100,000 P/E cycle
  - MLC – 10,000 P/E cycles

- **Controller Design**
  - Average Flash Writes Per Host Write
  - Efficiency of Wear Leveling
  - SSD Rated Capacity
  - SSD Flash Capacity Above Rated Capacity

- **Use**
  - Write rate at IO profiles
  - Duty Cycle
<table>
<thead>
<tr>
<th></th>
<th>Life Expectancy</th>
<th>Average Rate</th>
<th>Used per Day</th>
<th>Consumed per Year</th>
<th>Life Expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consumer Vehicle</strong></td>
<td>200K miles</td>
<td>45 MPH</td>
<td>1.5 Hrs</td>
<td>25K Miles</td>
<td>8.1 years</td>
</tr>
<tr>
<td><strong>Commercial Vehicle</strong></td>
<td>5,000K miles</td>
<td>55 MPH</td>
<td>11 Hrs</td>
<td>221K Miles</td>
<td>22.6 years</td>
</tr>
<tr>
<td><strong>Consumer (MLC)</strong></td>
<td>32,000 TBW</td>
<td>100 MB/s</td>
<td>7 Hrs</td>
<td>920 TB</td>
<td>34.8 years</td>
</tr>
<tr>
<td><strong>Enterprise (SLC)</strong></td>
<td>160,000 TBW</td>
<td>250 MB/s</td>
<td>24 Hrs</td>
<td>7,884 TB</td>
<td>20.3 years</td>
</tr>
</tbody>
</table>

*Your Mileage May Vary*
## Life Calculation

<table>
<thead>
<tr>
<th></th>
<th>&quot;Enterprise&quot;</th>
<th>&quot;Consumer&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program-Erase Cycles</td>
<td>1,000,000</td>
<td>100,000</td>
</tr>
<tr>
<td>Capacity (B)</td>
<td>160,000,000,000</td>
<td>320,000,000,000</td>
</tr>
<tr>
<td>Total Bytes Written</td>
<td>160,000,000,000,000</td>
<td>32,000,000,000,000,000</td>
</tr>
<tr>
<td>Hours used per day</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>Seconds used per day</td>
<td>86,400</td>
<td>25,200</td>
</tr>
<tr>
<td>Write Rate (B/s)</td>
<td>250,000,000</td>
<td>100,000,000</td>
</tr>
<tr>
<td>Bytes Written per day</td>
<td>21,600,000,000,000</td>
<td>2,520,000,000,000</td>
</tr>
<tr>
<td>Bytes Written per yr (B/yr)</td>
<td>7,884,000,000,000,000</td>
<td>919,800,000,000,000</td>
</tr>
<tr>
<td>Life in years</td>
<td>20.3</td>
<td>34.8</td>
</tr>
</tbody>
</table>
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- **SSS Performance**
- Some SSS Performance Issues
- Q&A
Media Performance

❖ The GOOD:
   ❖ Performance is great (wrt HDDs)
   ❖ High performance/power (IOPS/Watt)
   ❖ Low pin count: shared command / data bus \(\rightarrow\) good balance

❖ The BAD:
   ❖ Not really a random access device
      › Block oriented
      › Slow effective write (erase/transfer/program) latency
      › R/W access speed imbalance
   ❖ Performance changes with wear
   ❖ Some controllers use inefficient garbage collection

❖ The UGLY:
   ❖ Some controllers do read/erase/modify/write
Performance Drivers – SSS Design

- Interconnect
- Number of NAND Flash Chips (Die)
- Number of Buses (Real / Pipelined)
- Data Protection (internal/external RAID; DIF; ECC…)
- SLC / MLC
- Effective Block (LBA; Sector) Size
- Write Amplification
- GC Efficiency
- Bandwidth Throttling
- Buffer Capacity & Mgmt
Performance Drivers - External Cond

- Transfer Size
- Read/Write Ratios
- Temporal Randomness of Access
- Reserve Capacity Setting (% of used capacity)
- System Limitations (especially wrt scalability)
  - External Controller (#, Type, Performance); # Threads
  - CPU (#Cores, GHz)
  - System Bandwidth
  - Software Stack; Interrupt Handler
- External RAID
- Life of device (change in device affects t-Erase & t-Program)
Flash SSD vs. HDD Performance

- Be careful about performance assumptions when dealing with asymmetrical read and write performance
Performance v R/W Ratio

IOPS @ 512 B

Bandwidth (MB/s) @ 128 KB

Read/Write Collisions → Drop in Mixed Performance
Scalability

- Following Slides Show Scalability of \{1, 2, 4, 8\} units
  - Only 1 SATA controller is used – limiting scalability
    - Only 1 thread running

- Measurements taken at Read/Write Ratios of
  - \{100/0, 75/25, 50/50, 25/70, 0/100\}
  - RMS value is the “root mean square” of these five values

- IOPS measurement taken at 512 Byte Transfers

- Bandwidth taken at 128K Byte Transfers
  - Unless shown differently
  - Linux has a 128K limit
Scalability v R/W Ratio

IOPS @ 512 B

Bandwidth (MB/s) @ 128 KB

R/W Ratio (# Units in Parallel)
Performance v Block Size (75/25)

75/25 R/W IOPS

Block Size

75/25 R/W Bandwidth (MB/s)

Block Size

SATA-A
SATA-B
PCI-C
Flash SSD vs. HDD Performance

- HDD Performance
  - Roughly correlated to rotational speed

- SSD Performance
  - Varies widely between drives
  - Varies with work load in surprising ways
  - Varies with previous access patterns
## Flash SSD vs. HDD Performance

<table>
<thead>
<tr>
<th>SSD Type</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD A</td>
<td>220</td>
<td>115</td>
</tr>
<tr>
<td>SSD B</td>
<td>130</td>
<td>120</td>
</tr>
<tr>
<td>SSD C</td>
<td>57</td>
<td>38</td>
</tr>
<tr>
<td>SSD D</td>
<td>100</td>
<td>80</td>
</tr>
<tr>
<td>15K rpm HDD</td>
<td>171</td>
<td>171</td>
</tr>
<tr>
<td>7.2K rpm HDD</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>5.4K rpm HDD</td>
<td>61</td>
<td>61</td>
</tr>
</tbody>
</table>
## Flash SSD vs. HDD Performance

<table>
<thead>
<tr>
<th>SSD Type</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD A</td>
<td>45000</td>
<td>16000</td>
</tr>
<tr>
<td>SSD B</td>
<td>19000</td>
<td>130</td>
</tr>
<tr>
<td>SSD C</td>
<td>7000</td>
<td>15</td>
</tr>
<tr>
<td>SSD D</td>
<td>6300</td>
<td>926</td>
</tr>
<tr>
<td>15K rpm HDD*</td>
<td>185</td>
<td>170</td>
</tr>
<tr>
<td>7.2K rpm HDD*</td>
<td>79</td>
<td>73</td>
</tr>
<tr>
<td>5.4K rpm HDD*</td>
<td>60</td>
<td>57</td>
</tr>
</tbody>
</table>

*calculated from data sheet seek time*
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance

- Some SSS Performance Issues

- Q&A
SSD Performance Surprises

- Mixing reads & write slower than writes
SSD Performance Surprises

- Previous access patterns can dramatically affect current performance
Questions to Ask: Things to Know

- Design impacts on data integrity; life; failures & performance
  - ECC robustness
  - Write amplification / GC efficiency
  - Internal RAID
  - Bandwidth throttling
  - Partial Page Programming
- Test Conditions
  - RAID On/Off during testing?
  - Caching On/Off during testing?
  - Workload
  - Temporal Relationships
  - User capacity / reserve capacity
GC: Pathological Write Conditions

IF high percentage of total storage capacity utilized

\[ \text{AND} \]

High percentage of data has no correlation-in-time

\[ \text{AND} \]

Continuous writing (no recovery time for GC)

\[ \text{THEN} \]

Efficiency of GC greatly diminished
Pathological Write Condition

User Capacity
Formatted of Total

- 30GiB of 80G PCI-C
- 40GiB of 80G PCI-C
- 60GiB of 80G PCI-C
- 70GiB of 80G PCI-C
- 74GiB of 80G PCI-C
- 28GiB of 30G SATA-B
Final Thoughts on Solid State Storage

What is the impact on application workloads?
- SSS may be better → ideal for virtualized workloads
- Multiple workloads tend to randomize I/O streams
  - E.g. two simultaneous sequential workloads = one large random

What is the impact on server design?
- Potential to entertain swap of RAM for SSS
- Potential to use bus extension instead of channels
- Potential to improve in-box reliability (no moving parts)

Best practice – use all layers of the storage pyramid
Application mix may change – layers remain the same
Many thanks to the following individuals for their contributions to these slides.

Woody Hutsell, Texas Memory Systems
Brian McKean, LSI
Phil Mills, IBM
Rob Peglar, Xiotech
Jonathan Thatcher, Fusion-io
References

(1) **Overview and Current Topics in Solid State Storage**
by Rob Peglar, Xiotech Corporation
(http://www.snia.org/forums/ssi/knowledge/education/Overview_and_Current_Topics_Solid_State_Storage.pdf)

(2) **NAND Flash-based Solid State Storage Performance – An In-depth Look**
by Jonathan Thatcher, Fusion-io
(http://www.snia.org/forums/ssi/knowledge/education/NAND_Flash-based_Solid_State_Storage_Performance--An_In-depth_Look.pdf)

(3) **Solid State Storage Reliability and Data Integrity – An In-depth Look**
by Jonathan Thatcher, Fusion-io
(http://www.snia.org/forums/ssi/knowledge/education/Solid_State_Storage_Reliability_and_Data_Integrity--An_In-depth_Look.pdf)
References ...

(4) Solid State Storage for the Enterprise
by Woody Hutsell, Texas Memory Systems, Inc.
and Brian McKean, LSI Corporation – ESG
(http://www.snia.org/forums/sssi/knowledge/education/Solid_State_Storage__Reliability_and_Data_Integrity--An_In-depth_Look.pdf)
Thank you!

Khaled Amer
E-mail: khaledamer23@gmail.com
Agenda

- Computer and Storage Evolution
- What is Solid State Storage (SSS)?
- Flash Memory & DDR RAM Characteristics
- Why SSS/SSD?
- SSS Market
- System Architectures using SSS
- SSS Reliability and Endurance
- SSS Performance
- Some SSS Performance Issues

Q&A