Solid State Storage: Overview and Opportunities

Dean Klein
VP of Memory System Development
von Neuman Needs NAND!

Dean Klein
VP of Memory System Development
Agenda

- The Memory Wall – The Storage Wall
- DRAM in Review
- Memory Technologies Compared
- NAND and the Computer Architecture
- Breaking Speed Barriers
- Interesting Differences Between HDD’s and SSD’s
- Summary
The Memory Wall and Storage Wall

- CPU
- Memory & Bus Control
- Graphics
- Peripheral & Bus Control
- I/O:

- DDR3 DRAM
- PCIe
- USB
- SATA HDD

3+GHZ*N Cores
(N=1,2,4,6,8,12,…64, 80,…)

Brick walls don’t move well…
The Power of Memory: On an Exabyte Level

Assumptions

- Cell Voltage 1.5 V
- Cell Capacitance 25 fF
- Bitline Capacitance 75 fF
- Memory System Bandwidth 1 EB/sec

Simplified Results:

- Energy/bit 56 fJ
- Total Memory Cell Power 450 KW
- With Bitline 1800 KW
- With 64X Over-Fetch 115 MW
DRAM Access Time Evolution...

- CPU Cycle Time
- Multi Core Effective Cycle Time
- Memory Access Time
## Comparative Memory Cells

<table>
<thead>
<tr>
<th>Cell Size (u²)</th>
<th>Tech Node (nm)</th>
<th>Cell Size(F²)</th>
<th>Endurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM/Infineon MRAM</td>
<td>1.42</td>
<td>180</td>
<td>44</td>
</tr>
<tr>
<td>Freescale 6T-SRAM</td>
<td>1.15</td>
<td>90</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>0.69</td>
<td>65</td>
<td>163</td>
</tr>
<tr>
<td>Intel 65nm process 6T-SRAM</td>
<td>0.57</td>
<td>65</td>
<td>135</td>
</tr>
<tr>
<td>Freescale eDRAM</td>
<td>0.12</td>
<td>65</td>
<td>28</td>
</tr>
<tr>
<td>Freescale TFS: Nanocrystalline</td>
<td>0.13</td>
<td>90</td>
<td>16</td>
</tr>
<tr>
<td>Samsung 512Mbit PRAM Device</td>
<td>0.050</td>
<td>95</td>
<td>5.5</td>
</tr>
<tr>
<td>Micron 40-series DRAM</td>
<td>0.037</td>
<td>78</td>
<td>6</td>
</tr>
<tr>
<td>Micron 50-series NAND</td>
<td>0.013</td>
<td>53</td>
<td>4.5</td>
</tr>
</tbody>
</table>
NAND: SLC vs. MLC

- **SLC: Single Level Cell**
  - One transistor stores one bit
  - Greatest endurance: $10^5$–$10^6$ cycles with ECC
  - Greatest resistance to disturb
  - Cell size: $4.5F^2$

- **MLC: Multi-Level Cell**
  - One transistor stores $N$ bits. $N=2, 3…$
  - Lower endurance: $10^4$ cycles with ECC
  - Effective cell size: $\frac{4.5}{N}F^2$
## The Storage Wall

<table>
<thead>
<tr>
<th>CPU</th>
<th>Relative Latency</th>
<th>Relative Cost/bit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1,800</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>1,400</td>
</tr>
<tr>
<td></td>
<td>1,200</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>25,000,000</td>
<td>1</td>
</tr>
</tbody>
</table>

Cost/bit Data as of November 2007
The Storage Wall

<table>
<thead>
<tr>
<th>CPU</th>
<th>Relative Latency</th>
<th>Relative Cost/bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Cache</td>
<td>1</td>
<td>1,800</td>
</tr>
<tr>
<td>L2 Cache</td>
<td>2.5</td>
<td>1,400</td>
</tr>
<tr>
<td>DRAM</td>
<td>1,200</td>
<td>10</td>
</tr>
<tr>
<td>SSD</td>
<td>25,000</td>
<td>3</td>
</tr>
<tr>
<td>HDD</td>
<td>25,000,000</td>
<td>1</td>
</tr>
</tbody>
</table>

NAND Flash Closes the Latency Gap

Cost/bit Data as of November 2007
NAND in Computer Architecture

- CPU
- Memory & Bus Control
- Graphics
- PCIe
- DDR3
- USB
- Flash Drive
- Peripheral & Bus Control
- Intel “Robson” Flash
- SATA
- HDD
- PCIe
- PCIe
- Link
- USB

Drum Diagram:
- DRAM
- DDR3
- USB to USB
- PCIe to PCIe
- SATA to SATA
NAND in Computer Architecture

- CPU
- Memory & Bus Control
- DDR3
- DRAM
- PCIe
- Graphics
- Link
- Peripheral & Bus Control
- USB
- USB Flash Drive
- PCIe
- NVM–HCI Flash
- SATA
- SSD
The Next EEPC

- CPU
- Memory & Bus Control
- Graphics
- DDR2 DRAM
- USB
- PCIe
- xHCI Flash
- Peripheral & Bus Control
NAND in Server Architecture

- CPU
- Memory & Bus Control
- Peripheral & Bus Control
- Storage Control
- DDR3
- PCIe
- SATA
- USB
- HDD
- Capability Storage
- Capacity Storage
NAND in Server Architecture

CPU

Memory & Bus Control

DDR3

DRAM

Peripheral & Bus Control

PCIe

USB

eUSB Drive

SATA

SSD

AHCI Flash

Storage Control

Cache

SSD

Capacity Storage
The “Mercury” Demonstration

• What did we learn?
  ‣ QA doesn’t like us naming projects after hazardous substances…
  ‣ SATA has some legs, but the CPU utilization is high
  ‣ There are many opportunities for better NAND management
  ‣ Write amplification can be managed
  ‣ We can wear out NAND – quickly.
  ‣ Controller architecture has huge impacts on drive performance, system performance and NAND endurance.

• The result: Our next controller will be better!
Interesting Differences: HDD to SSD

- Physical differences
- Logical differences
The World’s First Hard Drive

1956 – The RAMAC

- 5 MB capacity
- 55 bits per inch
- Over 2000 pounds
- 1200 rpm
- Lease price of $750 per month
The World’s First Mobile Hard Drive

1956 – The RAMAC

- 5 MB capacity
- 55 bits per inch
- Over 2000 pounds
- 1200 rpm
- Lease price of $750 per month
SSDs vs HDDs: The Promise

<table>
<thead>
<tr>
<th>Metric</th>
<th>SSD</th>
<th>HDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per bit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Based on recent advances in NAND lithography, SSD densities reach capacities for mass market appeal.
- SSD offers many features that lead to improved user experiences.
- Early shortcomings for reliability and endurance have been overcome.

Much of the bad press garnered by SSD’s has been erroneous and can be attributed to a poor controller used in the original EEPC.
Subtle Differences Between SSD and HDD

- SSD’s have a flexible Logical to Physical map
Logical to Physical Mapping: HDD

Data is always placed in the same physical location.
Logical to Physical Mapping: SSD

Logical Address Space

Data may be grouped in the order it is received.

Physical Address Space

Buffer

LUT

©2008 Micron Technology, Inc. All rights reserved.
Subtle Differences Between SSD and HDD

- How data is erased matters to NAND
  - Erasing data by modifying a FAT table entry still requires the SSD to retain the actual data in the NAND.
  - Wear–management algorithms and garbage–collection algorithms will not be able to free and use this “erased” area.
Static Wear Leveling

- Swap blocks containing static data with blocks that have high erase counts

Method

- Periodically scan the Block Table (in the background)
- Identify Static block candidate to wear level
- Identify blocks with high erase count
- Reclaim static block candidate at appropriate time
- Move valid data from static block to a block with high erase count at appropriate time
Dynamic Wear Leveling

- Dynamic wear leveling is necessary to free up blocks for reprogramming
- Reclaim blocks with the highest amount of invalid (obsolete) pages
  - ‘Invalid’ pages contain user data that has been re-written to a different page
  - Define a threshold of total invalid pages in a block to determine candidates to reclaim (programmable)
- Method
  - Periodically scan the Block Table
  - Identify block candidate that is above the invalid page threshold
  - Reclaim the block candidate
Micron’s Lineup

1.8–inch & 2.5–inch, 32GB & 64GB SSD for notebook and performance computing

Embedded USB, 1 to 8GB SSD for cost–effective industrial and blade server applications

Module with custom form factor, density and interface
Summary

- NAND is poised to make an entry into computing
  - Premium Notebooks
  - High-end Servers
  - Low cost PC’s
- There are a lot of opportunities for innovation
  - Controller innovation
  - Algorithm innovation
  - NAND innovation