Architectural Considerations for Optimizing SSDs

Michael Abraham (mabraham@micron.com)
Advanced Engineering Architect
Micron Technology, Inc.
Abstract

• NAND Flash continues to shrink and is making the transition to 3D memories.
• What are some of the challenges that SSD designers face?
• This talk focuses on the NAND and other technologies that are bridging these gaps to improve SSD architecture, performance, reliability, and energy.
The Beginnings of Flash

- 1984 – Toshiba announces “Flash” (IEDM)
- 1987 – Toshiba announces NAND Flash (IEDM)
- 1994 – Compact Flash: Starts with NOR and later moves to NAND Flash
- 1995 – SmartMedia
- 1999 – USB Flash Drive
- 1999 – Flash-based Solid State Disk

FREMONT, Calif.--(BUSINESS WIRE)--Sept. 14, 1999--Addressing the growing demand for higher capacity solid state storage solutions, BiTMicro today announced the availability of the enhanced flash-based solid state E-Disk SNX35 product line with up to unprecedented 13,312 MBytes capacity record in a half-height 3.5-inch form factor.

With sustained read/write transfer rate of over 4.5 MBytes/sec, the E-Disk SNX35 SCSI Normal solid state flash disk drive is 2 times faster than the closest competitor. In addition, SNX35 system access time has been enhanced to less than 0.7 msec, which is over 7 times faster than the fastest SCSI flash disk solid state storage competitor.
Early Storage Optimizations

- Faster
- Upgradeable
- Slightly more expensive

- Slow
- Early obsolescence
- Relatively inexpensive
SSDs: Optimize for Performance

- Fastest use of Flash memory
- More expensive ($/bit) than previous solutions
SSD Product Priorities

- Lower Cost ($/bit)
- Higher Performance (MB/s, IOPS)
- Sufficient Reliability (3-5 years)
- Smaller Form Factors (area/volume)
- Lower Energy (pJ/bit)
Lowering SSD Costs: Shrink the NAND

- Industry is continuing transition to 3D NAND technologies

Data based on publicly available information
Lowering SSD Costs:
Add More NAND Bits per Cell

Source: iSuppli 2Q14

SLC is less than 1.5% of the market!
Less Expensive SSDs

Mission Accomplished?
### SSD Performance Trends

<table>
<thead>
<tr>
<th>Trend</th>
<th>Throughput</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read</td>
<td>Saturated</td>
<td>Increasing</td>
</tr>
<tr>
<td>Write</td>
<td>“Optimized”</td>
<td>Increasing</td>
</tr>
</tbody>
</table>

- **Read Throughput**
  - NAND Flash interface speed
- **Read/Write Latency**
  - Bits per cell
  - # of die in the SSD
  - ECC correction time
- **Write Throughput**
  - Sequential: Bits per cell, # of die in the SSD
  - Random: Block size
NAND Interface Speed: Going the Right Direction

- Read operations are still interface limited when two or more NAND die share the same I/O bus.
- Almost all SSDs today use NAND interface speeds of 200MT/s or faster.
- ONFI 4.0 is published:
  - Up to 800MT/s throughput
  - Reduces energy per bit with 1.2V interface,
Latency: Programming Times Also Affect Read Latencies

- Most applications favor read operations over write operations
- Most read operations are 4KB data sectors
- As monolithic NAND density increases, less NAND die are being used for a fixed system density
- As tPROG increases, the latency of random 4KB sector reads becomes more variable in mixed-operation environments as the probability of needing to read from a die that is busy increases

**Random 4KB Read Latency**

Read Latency (µs) vs. tR / tPROG (µs)

- Min Latency
- Max Latency
Sequential Write Performance

- For a fixed page size across process nodes, write throughput decreases as the NAND process shrinks.
- NAND vendors increase the page size to compensate for slowing array performance.
- Write throughput decreases with more bits per cell.

Data Bytes per Operation (Page Size * # of Planes)

Santa Clara, CA
August 2014
NAND Block Size Is Increasing

- The block is the smallest erasable unit made up of programmable and readable pages
- Block size impacts garbage collection times and random write throughput
Performance Optimizations

- Overprovision the amount of NAND die (slightly higher cost)
  - Use smaller NAND die on the latest litho
  - Use largest NAND die on the latest litho and add a few
- Write data sequentially to the NAND
- Improve garbage collection algorithms (e.g. TRIM) and scheduling

<table>
<thead>
<tr>
<th>Node</th>
<th>NAND Density</th>
<th>Data size</th>
<th># of Die</th>
<th>Seq Write (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen 1</td>
<td>32Gb (4GB)</td>
<td>8192</td>
<td>32</td>
<td>230</td>
</tr>
<tr>
<td>Gen 2</td>
<td>64Gb (8GB)</td>
<td>16,384</td>
<td>16</td>
<td>184</td>
</tr>
<tr>
<td>Gen 3</td>
<td>128Gb (16GB)</td>
<td>32,768</td>
<td>8</td>
<td>133</td>
</tr>
<tr>
<td>Gen 3</td>
<td>64Gb (8GB)</td>
<td>32,768</td>
<td>16</td>
<td>324</td>
</tr>
</tbody>
</table>
Reliability: Endurance and Data Retention

- Process shrinks lead to less electrons per floating gate
- ECC improves data retention and endurance
- ECC algorithms are transitioning from BCH to LDPC and codeword sizes are increasing
- Data retention is measured at max endurance cycles
- Possible to optimize endurance for data retention (examples below)
  - Consumer MLC NAND: 3K cycles, 1 year retention
  - Enterprise MLC NAND: 10K cycles, 3 months data retention
- Overprovisioning also helps: More die or a few more blocks per die
SSD Form Factors Are Shrinking

- 2.5” form factor for SSDs is becoming obsolete
- M.2 form factor enables Ultrabooks, tablets
- Fewer NAND placements drives 8- and 16-die package development
  - Can increase component cost because of reduced package yield

Source: PCIe M.2 Electromechanical Spec Rev1.0
Energy was previously not a high priority
Becoming very important in battery-powered systems
Rule of thumb: “Hurry up and wait”
Ways to reduce energy
  • Queue and burst operations
  • Reduce write amplification
  • Increase interface speed and lower its voltage
Some MLC and TLC NAND Flash have the ability to write and read data in SLC mode

- Smaller block size; mode is determined block by block and managed by the SSD controller
  - MLC (X pages per block) → SLC Mode (1/2 X)
  - TLC (Y pages per block) → SLC Mode (1/3 Y)

- Array times decrease, particularly for tPROG to 300-500µs resulting in higher burst performance

- Excellent enabler for TLC-based SSDs for lower SSD costs

- Initial write uses less energy as tPROG is shorter

- Best improvement with bursty workloads (e.g. consumer applications)
## SSD Optimization Summary

<table>
<thead>
<tr>
<th><strong>Desired Features</strong></th>
<th><strong>NAND Lithography, MLC→TLC</strong></th>
<th><strong>DRAM cache</strong></th>
<th><strong>Faster interfaces (SSD, NAND)</strong></th>
<th><strong>Overprovisioning</strong></th>
<th><strong>Block Management Optimizations</strong></th>
<th><strong>Die Stacking</strong></th>
<th><strong>SLC Mode</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Cost</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Higher Performance</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Sufficient Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Smaller Form Factors</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lower Energy</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Questions?
About Michael Abraham

• Advanced Engineering Architect in the Storage Business Unit at Micron

• Covers advanced NAND Flash and emerging memories

• IEEE Senior Member

• BS degree in Computer Engineering from Brigham Young University