The Long-Term Future of Solid State Storage

Jim Handy
Objective Analysis
Agenda

- How did we get here?
  - Why it’s suboptimal
  - How we move ahead
- Why now?
  - DRAM speed scaling
  - Changing role of NVM in computing
- The computer of tomorrow
  - Storage or Memory? Memory or Storage?
  - Capital Cost vs. Performance
Agenda

- How did we get here?
  - Why it’s suboptimal
  - How we move ahead

- Why now?
  - DRAM speed scaling
  - Changing role of NVM in computing

- The computer of tomorrow
  - Storage or Memory? Memory or Storage?
  - Capital Cost vs. Performance
Establishing a Schism

Memory

(Bytes)

DRAM
Cache

Flash?

Storage

(Blocks)

Disk
Tape
DVD
SAN/NAS
Cloud

Flash
Is It Really About Blocks?

- Block: 512 Bytes – 4KB
- NAND Page: 512 Bytes – 4KB
- DRAM: 32-64 Bytes
- CPU Cache Line: 64 Bytes

Almost nothing happens at the byte level!
Is There Another Differentiator?

**Memory**
- (Volatile)
  - DRAM
  - Cache

**Storage**
- (Persistent)
  - Disk
  - Tape
  - DVD
  - SAN/NAS
  - Cloud

**Storage Class Memory (SCM)**

- Flash

X
That’s Confusing!
What Is an SSD?

- An SSD is memory masquerading as storage
SSDs Are Nothing New

1978: StorageTek 4305
45MB, 600μs Access, $400K

1989: EMC Orion
4MB, 500 IOPS,
100μs Access, $34K

1998: Texas Memory Systems SAM-520
16GB, 50K IOPS/channel, $50K

1997: Quantum Rushmore
134MB-3.2GB, 9K IOPS
50μs Access, <$55K
DEC* RC-11 and RF-11 fixed-head disc...and Data General Novadisc® users:

Replace Fixed-Head Disc with Dataram

**BULK CORE**

Now, all the remarkable features of Dataram’s BULK CORE memory module are available to you in a unique storage peripheral with complete interfaces to emulate DEC and Data General fixed-head discs.

Basic building block of this dramatic, new peripheral is Dataram’s BULK CORE module, which provides 256 kilobytes of storage on a single board. Eight of these modules can be packaged in a standard 19” chassis to provide two megabytes of storage.

To give you more of what you can’t get from fixed-head discs, BULK CORE gives you microsecond-range access time, high reliability; and greatly improved maintainability. And at a price unheard of for core or semiconductor memory. Until now.

Until Dataram made its BULK CORE memory system plug-compatible with PDP®-11 and Nova® minicomputers, To provide:

- Access time 1/10,000 of FHD
- High Throughput
- Zero Error Rate
- Self-Test for Fault Isolation
- Hardware & Software Transparent
- LED-spotlighted Fault Isolation
- 216 KI Modularity
- Non-Volatile
- Non-Mechanical
- High MTBF/Low MTTR
- Low Power
- Partial Check

Reasons enough to find out more about BULK CORE. If you use a DEC or Data General minicomputer—or any kind—and want to move ahead in performance, move a BULK CORE into your system.

PDP and DEC are registered trademarks of Digital Equipment Corporation. Nova and Novadisc are registered trademarks of Data General Corporation.
Disk Interfaces Create Delays
Where Do Delays Come From?

<table>
<thead>
<tr>
<th>Time (ms)</th>
<th>HDD</th>
<th>SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time scale in milliseconds

Note tiny band of rainbow colors
Looking Closer at SSD Delays

Expand time scale to microseconds
3 Ways to Reduce Delays:
Chip Interface, System Interface, & Media

- MLC NAND
  - SATA 3
  - ONFi 2
- MLC NAND
  - SATA 3
  - ONFi 3
- MLC NAND
  - PCIe x4 G3
  - ONFi 3
- Future NVM
  - PCIe x4 G3

Transfer Time
ONFi 3 Shrinks Transfer Time
PCIe Shrinks Link Transfer & Platform & Adapter Time
New NVM Shrinks Read, Transfer, & Misc SSD Times

Time (μs):
0 μs, 20 μs, 40 μs, 60 μs, 80 μs, 100 μs, 120 μs
Focus Areas For Standards Groups

![Diagram showing focus areas for standards groups]

- Future NVM PCIe x4 G3
- NVMe & STA
- SNIA
- Link Transfer
- Platform & Adapter
- Software

Time scale expanded again
How to Move Ahead?

- Free Flash from disk interfaces
  - PCIe, NVMe have made much progress
- Lay groundwork for newer technologies
  - More on this later
- Design for cost, not persistence
Agenda

- How did we get here?
  - Why it’s suboptimal
  - How we move ahead
- Why now?
  - DRAM speed scaling
  - Changing role of NVM in computing
- The computer of tomorrow
  - Storage or Memory? Memory or Storage?
  - Capital Cost vs. Performance
DRAM Transfer Rates

- SDRAM
- DDR
- DDR2
- DDR3
- DDR4

Transfer Rate (GB/s)
Is DRAM Running Out of Speed?

- No clear successor to DDR4
- DDR4 made several concessions:
  - Very low signaling voltages
  - Point-to-point signals
- HMC or HBM a likely next step
  - Will this be a “Final Level Cache?”
- Everything points to fixed memory sizes
NVM to the Rescue

- DRAM’s not the only upgrade path
  - Dollar for dollar NAND is a better option
- NV Memories aim to fit between DRAM & NAND
  - Intel/Micron 3D XPoint – “Optane”
- Future memory systems will include everything:
  - DRAM, NVM, NAND, HDD
- One won’t kill off the others
New Memories Are Faster than NAND

Smith et al: *Through the Looking Glass II – Trend Tracking for ISSCC 2013*  
Commemorative Supplement to the Digest of Technical Papers, ISSCC 2013
Merging Storage and Memory

Role-sharing Positioning of Emerging Research Memories

Clock Frequency (Hz)

Cell Area ($F^2$)

- High-Speed Graphics: Games
- High-density work memory: PC
- High-speed MCU: Car
- Parameter storage: IC tag / smart-card
- Program storage: Mobile phone, PDA
- Data storage: DSC, SSD, USB memory
- Low-power work memory: Mobile phone
- Front-end SoC: ASSP, ASIC
- High-speed MPU: Supercomputer

From Ed Grochowski, 2014 Report on New Storage Technologies
Agenda

- How did we get here?
  - Why it’s suboptimal
  - How we move ahead
- Why now?
  - DRAM speed scaling
  - Changing role of NVM in computing
- The computer of tomorrow
  - Storage or Memory? Memory or Storage?
  - Capital Cost vs. Performance
Memory or Storage?
What is Storage Class Memory?

Storage-class memory (SCM) combines the benefits of a solid-state memory, such as high performance and robustness, with the archival capabilities and low cost of conventional hard-disk magnetic storage.

*IBM Almaden Research Labs*
New NVM Has Disruptively Low Latency

Latency Budgets

Context Switch

NUMA

Latency (nS)

HDD | SATA SSD | NVMe | Persistent

1ns | 10ns | 100ns | 1µs | 10µs | 100µs | 1ms | 10ms | 100ms | 1s
Non-Blocking I/O

- Software overheads are being driven to keep pace with devices.
- NUMA latencies up to 200ns have historically been tolerated.
  - Anything above 2-3μs will probably need to context switch.
  - Latencies below these thresholds cause disruption.
New Memory-Mapped Files Eliminate File System Latency

Today

Kernel

User

Application

File System

Disk Driver

HW

Disk

New

Application

Memory Mapped Files

User

Persistent Memory

Load/Store
The Computer of Tomorrow

- Fixed DRAM Size
  - Stacked packaging
- Upgradeable NVM
  - Tomorrow’s version of a DIMM
- Both flash and disk
  - Flash on PCIe or its own bus
  - No foreseeable $/GB crossover
- Slowly sneaking up on SCM software
  - Very much work needed here
Questions?