A Complete System Approach to NAND in Computing

Dean Klein

Micron Technology, Inc.
Issues with Storage (aka: Memory)

**Performance**: The Memory Wall

**Power**: Exabytes and Megawatts

**Reliability**: Redundancy and Resiliency

**Scaling**: Density and Cost
Memory Hierarchy Gaps

Relative Latency

Relative Bandwidth

HDD

DRAM

L1

L2
System Architecture

CPU
  L1
  L2

IOH

Peripheral & Bus Control

DRAM
  DDR2

GFx

SATA2

FC

NIC/IB

PCIe

Link

©2010 Micron Technology, Inc.
System Architecture

Benefit:
Latency reduction by elimination of controller “hops”.

- CPU L1
- CPU L1
- CPU L1
- CPU L1
- L2
- GFx
- Peripheral & Bus Control
- PCIe
- PCIe
- PCIe
- PCIe
- SAS
- NIC/IB
- SATA2
DDR4: Memory Architecture Improvement

DDR4 is coming – High Bandwidth, Low Power, High Density

**Major Advantages** of DDR4:

- High Bandwidth: *Up to 3.2 Gbps*
- Power Consumption: *1.2V* is 10-15% improvement over DDR3 & DDR3L
- High Density: 2Gb - 16Gb; **up to 8H** (128Gb stack); **single load**
- DDR4 *cost overhead* appears better than previous 1st gen technologies
- **Power, density and bandwidth:** Optimized for HPC and servers
Memory Hierarchy Gaps

Relative Latency

Relative Bandwidth

DRAM & System Architecture:
• Increased parallelism
• Reduced Latency
Memory Hierarchy Gaps

DRAM & System Architecture:
- Increased parallelism
- Reduced Latency

Relative Bandwidth

Relative Latency
Requirements

• Low random latency:
  ▶ As the number of CPU cores (and thus the parallel tasks) increase the traffic becomes increasingly random.
  ▶ As the use of VM’s increase the traffic becomes increasingly random.

• High endurance:
  ▶ Accepted DRAM endurance is >1E15 cycles.
  ▶ Accepted NAND endurance is <1E5 cycles or 3E3 cycles.

• Intelligent management
Closing the Storage to DRAM Gap

- Technical requirements:
  - Low latency
  - High endurance
  - Symmetric read and write
  - Block-based architecture
  - Able to use existing interfaces

- Economic requirements:
  - Lower cost/bit than DRAM
  - Lower power than DRAM
NAND Endurance

- NAND Process and Design
  - + Improved processes: new materials and structures
  - - Smaller geometries: fewer electrons and more interference

- SSD Architecture and Algorithms
  - + Improved error handling and efficiency
  - ++ NAND optimizations
Fast Forward: Future System Architecture

Peripheral & Bus Control

IOH

DRAM DDR4 L2 DRAM DDR4 L2 DRAM DDR4 L2 DRAM DDR4 L2

Link

PCIe

Storage-Class Memory

High-Density Local SS Storage

NIC/IB SATA3

Gen3

SAS
Solid State Storage as the Ultimate Cache

- Pinnable, secure, high performance.
- System software can:
  - Mitigate wear
  - Allow graceful degradation
  - Accelerate slower storage, both local and remote
- Fits with SSD today and PCM tomorrow.
- Can take advantage of lower latency, higher bandwidth connections.
- Optimized solution is integrated solution.
The Holistic Approach to Storage

- The gap between CPU and main memory is being closed by architecture and DRAM advanced.
- The gap between DRAM and storage is widening.
- Increasing randomization of storage traffic demands a storage system with low random latency.
- Endurance requirements will dictate advanced NAND management, and must include system software as a part of the solution.
- NAND, and ultimately PCM, will close the storage gap.