Annual Flash Controller Update

David McIntyre

@DavidSMcIntyre
#flashmem

1:1 Mtgs Text FMS to (408) 772-7044
Overview

- Data Center Drivers
- Memory Hierarchy Drivers
- Flash Controller Challenges
- Supporting Technologies
Data Center Trends

➢ Hyper Converged Infrastructure
  • Integrated Compute/Storage/Networking
  • Massive interconnectivity (25Gb to 100Gb)
  • Software managed virtualized resources

➢ Hyper Scale
  • Independent scaling of compute and storage resources
  • Good for elastic workloads, e.g. Hadoop, NoSQL
  • Acceleration As a Service
Flash controllers must support hyperscale requirements (deterministic latency, performance/watt, endurance, reliability)
Data Center Trends

➢ **Storage**
  • Computational Storage
  • Convergence of RAM/cache and SCM
  • NVDIMM N and P
  • NVMe-oF, NVMe/TCP

➢ **Compute**
  • GPU, TPU and FPGA accelerators

➢ **Networking**
  • Low latency, high performance RDMA networks
  • 100Gbps+

➢ **Hybrid Cloud**
  • For lease and on premises-equipment
Data Center Applications

- Hyperscale Public Cloud
- Enterprise Private Cloud
- Telco Cloud / Edge
- High Performance Computing

- Financial Analytics
- Machine Learning
- Security
- Video Transcoding
- Genomics
- Big Data Analytics

Compute
Storage
Network
Supporting Infrastructure

- Compute Acceleration
- Memory & Storage Tiering
- Low Latency Networks

Application Features
Memory and Storage Tiering

- L0: CPU registers (optimized by compiler)
- L1: on-chip L1 cache (SRAM) - 8-way associative in Intel Core i7
- L2: off-chip L2 cache (SRAM) - 8-way associative in Intel Core i7
- L3: off-chip cache L3 shared by multiple cores (SRAM) - 16-way associative in Intel Core i7
- L4: main memory (DRAM)
- L5: local secondary storage (local disks)
- L6: remote secondary storage (distributed file systems, web servers)
Flash System Challenges

➢ Error correction costs increasing
➢ Endurance limits
➢ Slow write speeds continue
➢ IO bottlenecking
➢ Emerging NV technologies (MRAM, PCM, RRAM)
➢ Form Factors (M.2, EDSFF (E1.x, E3.x))

Source: IMEX
Controller Trends - Intel Optane

➢ Separate Media Controllers
  – Silicon Motion SM2263: 1TB QLC NAND
  – Intel SSL3D: 32 GB Optane

+ Intel® Rapid Storage Technology Software
Computational Storage

➢ Tightlly coupled CSSD
  ▪ Embedded CPU Cores
  ▪ Hardware Accelerators
  ▪ Memory
  ▪ NAND Flash

➢ Purpose-built data paths
  ▪ Any to any connectivity
  ▪ 10X-100X Internal Bandwidth

➢ Distributed, Scale-out model

Adaptive Storage Acceleration

- Encryption
- Compression
- Data Dedupe
- RAID & Erasure codes
- Key-Value Offloads
- Database ETL & Query Offloads
- Spark-SQL / Map-Reduce
- Video / Image Transcoding, Processing and Delivery
- Search - Text, Image, Video etc.
- Stats / Counters
- Machine Learning
Computational Storage- Smart SSD

Introducing SmartSSD

- Scalable and accelerated performance with SmartSSD
- Greater integration

FPGA Accelerator
Data Movement through CPU and Memory

Host Memory Transfer
PCIe Bus

CPU
DRAM

Next Gen SmartSSD

FPGA Accelerator
SSD

SAMSUNG
The Heart of Everything
Computational Storage Controller Options

Source: NVMexpress.org (both graphics)
Flash Controller Opportunity

- SOC Integrated solution
- Hybrid Controller- SCM caching
- Deterministic Latency
- Flash Density and Performance (3D QLC)
- Byte addressable
- Opportunity for NAND to support load/store-driven data center applications (e.g. NVDIMM-P)
NVDIMM-N Controller Architecture

On power failure these FETs switch out the processor signals.

Power failure switch

Individual CKE lines

To super-cap bank

Backup and Restore Solution
Courtesy of Agigatech
NVDIMM-P

- High capacity, transactional access DDR4/DDR5
- Persistent memory

Supported applications
- Database caching
- Enterprise storage
- High Performance computing
Controller Options

Technology scaling favors programmability and parallelism

- **CPUs**
- **DSPs**
- **Multi-Cores**
- **ASICs**
- **GPGPUs**
- **FPGAs**

**Single Cores**

- Multi-Cores
- Coarse-Grained CPUs and DSPs

**Fixed Design**
- Efficient Performance

**Massively Parallel**
- Processor Elements

**Massively Parallel**
- Programmable Logic and SOC attributes
Flash Controller Technology Options

- Data center metric is performance/watt
- Performance, power efficiency and flexibility is required to support data center applications
## Technology Comparison

<table>
<thead>
<tr>
<th>Technology</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| CPU        | Well established products | • Limited cores for parallel processing  
• Power consumption |
| FPGA       | Heterogeneous parallel processing  
Performance/Watt  
Flexibility | • Rudimentary development environment  
• Inefficient per unit costing |
| GPU        | Same task parallel processing  
Developer ecosystem | • Power consumption  
• Leading variable types |
| ASIC       | Highest Performance | • High NRE  
• Custom design |
| ASSP       | Custom Performance | • Limited functionality |
Error Correction Overview

Driving Factors for New ECC
- Increasing Bit errors in NAND Flash
- Soft error occurrences
- Decrease in write cycles
- RS, BCH overhead for data and spare area
- Increase use of Metadata in file systems
- Correction Overhead
- Gate count
- Requirement for no data loss

Comparing ECC Solutions

<table>
<thead>
<tr>
<th>Features</th>
<th>BCH</th>
<th>LDPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate Count</td>
<td>Low</td>
<td>Mid</td>
</tr>
<tr>
<td>Latency</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Tuneablity</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Soft Data</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

198X | 199X | Present
Hamming | RS | BCH | LDPC

Combined ECC
## Flash Controller Support

<table>
<thead>
<tr>
<th>IP</th>
<th>IO</th>
<th>Speed</th>
<th>Logic Density</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONFI 4.1</td>
<td>40 pins/ch</td>
<td>1200MT/s</td>
<td>5KLE/ch</td>
<td>NAND flash control, wear leveling, garbage collection</td>
</tr>
<tr>
<td>DDR4/5</td>
<td></td>
<td>6.4 Gbps</td>
<td>10KLE</td>
<td>Flash control modes available for NVDIMM</td>
</tr>
<tr>
<td>PCM</td>
<td></td>
<td></td>
<td>5KLE</td>
<td>PCM- Pending production $</td>
</tr>
<tr>
<td>MRAM</td>
<td></td>
<td></td>
<td>5KLE</td>
<td>MRAM- Persistent memory controller</td>
</tr>
<tr>
<td>BCH</td>
<td></td>
<td>&lt;10KLE</td>
<td></td>
<td>Baseline ECC standard</td>
</tr>
<tr>
<td>LDPC</td>
<td></td>
<td>50KLE+</td>
<td></td>
<td>Increased performance for FPGAs</td>
</tr>
<tr>
<td>PCIe</td>
<td>Gen 4x8</td>
<td>16 GT/s</td>
<td>HIP</td>
<td>Flash Cache</td>
</tr>
</tbody>
</table>
Typical SSD Controller Architecture

Typical Attributes
- Number of Ports 8 to 32
- Pin Count 250 to 1000+
- Power 1 to 3.5 Watts
- Internal, External RAM

Variations
- Number of CPU’s
- Error Correction
- Interfaces
- Memory Type and Size
Cache coherency will continue to expand into SCM into SSD caches
Changing of the Guard

Capacity Storage Revenue Flash vs. HDD ($M)

Units

Re: Wikibon

Re: NVMeExpress.org
Controller Challenges Summary

➢ **Host Interface IO**
  - Gen Z, CCIX, OpenCAPI
  - PCIe Gen 5
  - NVMe-oF and NVMe/TCP

➢ **Application Requirements**
  - Deterministic latencies
  - Load/Store vs Block
  - Performance
  - Endurance

➢ **Hybrid Control**
  - 3D NAND, 2D NAND
  - Cache: 3DXpoint, MRAM
Flash Control has extended into tiered subsystem management

- Caching has extended into SCM, necessitating hybrid control
- IO interfaces need to support fabric
- Advancing geometries and process technologies require more and advanced error correction
- Hyperscaler applications demand load/store performance with deterministic latency
Annual Flash Controller Update

David McIntyre

Text FMS to (408) 772-7044