Annual Flash Controller Update

David McIntyre

DSMcIntyreConsulting@gmail.com

Text FMS to (408) 772-7044
Overview

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

➢ Hyper Converged Infrastructure
  – Integrated Compute/Storage/Networking
  – Massive interconnectivity
  – Good for Exchange, Oracle, SQL databases
  – Software managed virtualized resources

➢ Hyper Scale
  – Independent scaling of compute and storage resources
  – Good for elastic workloads, e.g. Hadoop, NoSQL
  – Also software managed
Data Center Trends

➢ Storage
  • Convergence of RAM/cache and SCM.
  • All flash and hybrid arrays
  • Persistent memory cache

➢ Compute
  • GPU, TPU and FPGA accelerators

➢ Networking
  • Low latency, high performance RDMA networks

➢ Hybrid Cloud
  • For lease and on premises-equipment
  • Deployment Options, e.g. OpenStack and Docker
Hyperscaler Priority

Hyperscale in 2020

<table>
<thead>
<tr>
<th>Requirement</th>
<th>By 2020</th>
<th>Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>of all data center servers</td>
<td>47%</td>
<td>21%</td>
</tr>
<tr>
<td>of all data center processing power</td>
<td>68%</td>
<td>39%</td>
</tr>
<tr>
<td>of all data stored in data centers</td>
<td>57%</td>
<td>49%</td>
</tr>
<tr>
<td>of all data center traffic</td>
<td>53%</td>
<td>34%</td>
</tr>
</tbody>
</table>

Flash controllers must support hyperscale requirements (latency, performance/watt, endurance, reliability)
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 L3 cache 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)
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
Efficient ECC with NVMe Performance for 3D NAND TLC/QLC
## 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>400 MTps</td>
<td>5KLE/ch</td>
<td>NAND flash control, wear leveling, garbage collection</td>
</tr>
<tr>
<td>Toggle Mode 2.x</td>
<td>40 pins/ch</td>
<td>400 MTps</td>
<td>5KLE/ch</td>
<td>Same</td>
</tr>
<tr>
<td>DDR4</td>
<td></td>
<td>3.2Gbps</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></td>
<td>&lt;10KLE</td>
<td>Reference design</td>
</tr>
<tr>
<td>PCIe</td>
<td>G4x8</td>
<td>128Gbps</td>
<td>HIP</td>
<td>Flash Cache</td>
</tr>
</tbody>
</table>
Persistent Friend or Foe

- Intel/Micron Xpoint Claimed Attributes vs. NAND
  - Performance (10X)
  - Endurance (1000X)
  - Latency (1/1000X)
  - Byte addressable

- Est Cost (2X+)

- Opportunity for NAND to support load/store-driven data center applications (NVDIMM-F and NVDIMM-P)
Flash Controller Challenges: Now

➢ **Host Interface IO**
  - Gen Z, CCIX, OpenCAPI
  - PCIe Gen 4
  - Open Channel

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

➢ **Hybrid Control**
  - 3D NAND, 2D NAND
  - Cache: 3DXpoint, MRAM
Coherent Networks Roadmap

➢ Cache coherency will continue to expand into SCM into SSD caches

Re: OFA.org
NVDIMM Controller Architecture

On power failure these FETs switch out the processor signals

Can be
- Buffered
- Un-buffered
- Registered

FPGA

I2C

Control signals

FET bank

Power failure switch

400MHz / 800MB tested

DIMM

To super-cap bank

DDR ctrl with tri-state

Individual CKE lines

Power regulation

Flash 1 or 2 SD Cards or BGA

Processor

DDR

400MHz / 800MB tested

5 or 9

Flash Memory Summit 2018
Santa Clara, CA
Typical SSD Controller Architecture

- **Host Interface**
  - PCIe/SAS/SATA PHY
  - SATA3.0 (with DMA)
  - SAS2.0 (with DMA)
  - NVMe
  - SATA CMD Handler
  - SAS CMD Handler
  - Host Selector
  - Compression
  - Encryption
  - CMD XLRTR

- **Processor Sub-system**
  - Processor Core
  - AXI-APB Bridge
  - APB Interconnect Bus
  - AXI Interconnect Fabric
  - SRAM 5MB
  - Multi-port Arbiter + SRAM Controller
  - LDPC Wrapper
  - LDPC Engines
  - Control Flow

- **Arbitration**
  - Boot ROM I/F
  - UART
  - GPIO
  - Timers

- **Flash Interface**
  - NAND Flash CTRL 01
  - NAND Flash CTRL 02
  - NAND Flash CTRL 03
  - NAND Flash CTRL 10
  - NAND Flash PHY
  - NAND Flash PHY
  - NAND Flash PHY
  - NAND Flash PHY
  - Flash PHY
  - Flash PHY

- **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

---

Flash Memory Summit 2018
Santa Clara, CA
NVMe Roadmap - a NVMeF Precursor

**YESTERDAY**
- DMA to/from host memory
- CPU handles command transfer

**TODAY**
- Peer-to-peer DMA through PCI bus
- No host memory use

**TOMORROW**
- Network integrated NVMe storage
- No CPU or memory involvement
Encapsulated protocol transmitted over fabrics
RDMA (iWARP, RoCE)
FC, Infiniband, L1 Tunneling

The goal is to enable next-gen technologies to deliver a 4KB I/O in less than 10μs - about one thousandth of the latency of a 7200 RPM SATA drive
SSD Scale Out over Fabrics

➢ NVMeF breaks through local NVMe barrier and supports low latency
Open Channel Pros and Challenges

➢ I/O Isolation and Determinism
➢ Software managed resources
➢ Application-centric
➢ Linux kernel support required
➢ Vendor-specific attributes

Re: CNEX
Differentiation Paradox

➢ Hyperscaler standardization
➢ Vendor competitive selling features
➢ Application requirements
  • I/O
  • Power Consumption
  • Capacity
  • Latency
Controller Options

Technology scaling favors programmability and parallelism

Single Cores
Multi-Cores
Coarse-Grained CPUs and DSPs
Coarse-Grained Massively Parallel Processor Arrays
Fine-Grained Massively Parallel Arrays

CPUs
DSPs
Multi-Cores
Array
GPGPUs
FPGAs
➢ 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 |
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

DS McIntyre Consulting@gmail.com

Text FMS to (408) 772-7044
Flash Controller Challenges: Then

- **Emerging memory types**
  - ONFI 4.0, Toggle Mode 2.x
  - PCM, MRAM
  - DDR4

- **Controller Performance Options**
  - Write back cache, queuing, interleaving, striping

- **ECC levels**
  - BCH, LDPC, Hybrid

- **FTL location - Host or companion**

- **Data transfer interface support**
  - PCI Express, SAS/SATA, FC, IB
Flash Controller Challenges: Now

- **I/O Performance**
  - Interchip coherency
  - Host Communications
  - Network
- **Latency**
  - HPC network latencies
- **Density**
  - 3D, HBM2
- **Heterogeneous flash memory types**
- **Reliability and Endurance**