FPGAs in Flash Controller Applications

David McIntyre
DSMcIntyreConsulting@gmail.com
FPGA Then...
FPGA Now- Data Centers!

- High System Performance
- Memory Bandwidth
- Design Flexibility
- Signal Integrity
- Low Power
- Embedded Processing
Technology scaling favors programmability and parallelism

**Processing Options**

- CPUs
- DSPs
- Multi-Cores
- Array
- GPGPUs
- FPGAs

- Single Cores
- Multi-Cores
  - Coarse-Grained
  - CPUs and DSPs
- Coarse-Grained
  - Massively Parallel
  - Processor Arrays
- Fine-Grained
  - Massively Parallel
  - Arrays
Altera FPGA Technology – Hardware Programming

- **Massive Parallelism**
  - Millions of logic elements
  - Thousands of 20Kb memory blocks
  - Thousands of DSP blocks
  - Dozens of High-speed transceivers

- **Hardware-centric**
  - VHDL/Verilog
  - Synthesis
  - Place&Route
FPGA Utilization across Data Centers

Point and SOC Solutions

- Application Acceleration
- Embedded Processing
- I/O Protocol Support
- Memory Control
- Compression
- Security
- Port Aggregation & Provisioning
Hybrid RAID System
- Persistent DRAM and Flash Caches

CPU

PCI-e or CPU System Bus

Network Side

FC HBA

PCI-e

FC, iSCSI, FCoE

ASSPs

FPGA Controller

RAID/Cache Controller

CPU Interface

Network I/F

RAID² Logic

Disk I/F

Cache/Memory Controller

Persistent DRAM

Flash Cache

Disk Side

FC HBA

SATA

SAS

PCI-e

Dual Controller
Hybrid RAID System
- PCIe Switch Centric

CPU

Persistent DRAM

PCI-e or CPU System Bus

FC HBA

FC, iSCSI, FCoE

PCle Switch

Network Side

ASSPs

FPGA Controller

Flash Cache

Dedupe/Encrypt

PCI-e

Disk Side

FC HBA

SATA

SAS

Dual Controller

PCI-e

PCI-e

PCI-e
Flash Cache Challenges & Evolution

- **Ongoing Challenges**
  - Error correction costs increasing
  - Limited endurance (lifetime writes)
  - Slow write speed
  - SATA/SAS SSD interface is slow

- **Storage over PCIe**
  - Faster BW projections
  - SATA Express
  - NVM Express
  - SCSI Express

- **NVMe over Fabrics**

- **Emerging flash technologies**
  - MRAM (Magneto Resistive)
  - PCM (Phase Change)
  - RRAM (Resistive)
  - NRAM (Carbon Nanotube)
Figure 1. Categories of Memory (Charge Versus Resistivity)

Key:
DRAM = dynamic RAM
EEPROM = electrically erasable programmable ROM
EPROM = erasable programmable ROM
FeRAM = ferroelectric RAM
MRAM = magnetoresistive RAM
PRAM = phase-change RAM
PSRAM = pseudostatic RAM
SRAM = static RAM

Embedded possible
A Cost Effective Bridge between DRAM and NAND?

- Intel/Micron Xpoint (NV Memory)
  - Vertical placement of floating gate cells
  - Vastly improved endurance and performance vs. NAND
  - 256GB 32-tier 3D TLC

- Sandisk/Toshiba
  - 256GB 48 layer 3D NAND (TLC)

Source: Anantech.com
Figure 6. Migration Timeline for Emerging Memory Technologies

Source: Gartner
5MB in Flight!
Flash Controller Design Considerations
Flash Controller Requirements

- Uncertainty Favors PLDs for Flash Control Solutions
- Flash Challenges Continue
  - Data loss, slow writes, wear leveling, write amplification, RAID
- Many Performance Options
  - Write back cache, queuing, interleaving, striping, over provisioning
- Many Flash Cache Opportunities
  - Server, blade and appliance
Flash Controller Design Challenges

- 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 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 3.x</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>DDR3</td>
<td>72 bit</td>
<td>1066 MHz</td>
<td>10KLE</td>
<td>Flash control modes available for NVDIMM</td>
</tr>
<tr>
<td>PCM</td>
<td></td>
<td>5KLE</td>
<td></td>
<td>PCM- Pending production $</td>
</tr>
<tr>
<td>MRAM</td>
<td></td>
<td>5KLE</td>
<td></td>
<td>MRAM- Persistent memory controller</td>
</tr>
<tr>
<td>BCH</td>
<td></td>
<td>&lt;10KLE</td>
<td></td>
<td>Reference design</td>
</tr>
<tr>
<td>PCIe</td>
<td>G3x8</td>
<td>64Gbps</td>
<td>HIP</td>
<td>Flash Cache</td>
</tr>
</tbody>
</table>
Flash Cache Controller Examples

- **Multi Channel Controller**
  - Single to multi Flash channel capability
  - Basic NAND development platform
  - Provides High Speed ONFI & Toggle NAND PHY
  - ECC of 8 and 15 bits of error correction

- **Single Channel Controller**
Typical SSD Controller Architecture

- AXI Interconnect
- APB Bus
- AXI for Memory
- SATA3.0 (with DMA)
- Compression
- Processor Core
- Multi-port Arbiter + SRAM Controller
- SDRAM 5MB
- LDPC Wrapper
- LDPC Engines
- Control Flow
- Storage Parallel for Memory

**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

---

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

**Arbitration**
- I-RAM
- D-RAM
- Processor Core
- AXI-APB Bridge
- Boot ROM I/F
- UART
- GPIO
- Timers
- PCIe/SAS/SATA PHY
- Host Selector
- Debug Logic
- Compression
- Encryption
- CMD XLRTR

**Flash Interface**
- Flash PHY
- NAND Flash
- NAND Flash
- NAND Flash
- NAND Flash

**Variations**
- Blue: AXI Interconnect
- Purple: APB Bus
- Green: AXI for Memory

---

**Processor Sub-system**
- Flash Translation Layer
- Flash Low-Level Driver
- Flash Memory Chip
- Hardware

---

**FPGA**
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>High</td>
<td>Mid</td>
</tr>
<tr>
<td>Latency</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Tuneability</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Soft Data</td>
<td>no</td>
<td>high</td>
</tr>
<tr>
<td>Data Overhead</td>
<td>high</td>
<td>low</td>
</tr>
</tbody>
</table>

198X | 199X | Present
------------------------
Hamming | RS | BCH | LDPC | Combined ECC
The Parade of Codes

ECC - Block Hamming
- DRAM variant
- Applicable to the flash page block sizes
- Smaller blocks used as error rates increased

Reed Solomon
- CD-ROM basis, stronger than Hamming
- Split correction blocks split into 9 bit symbols
- Good for clumped errors

BCH
- Better supports MLC >8bits correction block
- BCH ECC increasing with correction block sizes
LDPC and Programmable Logic

- Addresses higher BER across process node curve
- Good for TLC
- FPGA parallelism of Parity Matrix allows for faster processing of algorithm
Target Application: Enterprise Tier-1 Storage: Databases and Virtualization

<table>
<thead>
<tr>
<th>Function</th>
<th>Solution Rqts</th>
<th>IP Rqts</th>
</tr>
</thead>
</table>
| Flash Control | - ONFI 2.X/3.0  
- Toggle Mode 2.0  
- Multi flash load/ch  
- 40 GPIO/ch       | - Flash Controller  
(bad block mgt and wear leveling)  
- Metadata & caching  
- ECC BCH core       |
| RAID Control | PCIe Gen 3                                                                   | - Flash-specific RAID  
- Switching and aggregation                                                |
**Flash PCIe Cards**

**Target Application:** Embedded PCIe storage for flash cache and scale-out computing

FPGA controller provides flexibility to integrate multiple complex functions and adapt to changing interfaces & APIs.

<table>
<thead>
<tr>
<th>Function</th>
<th>Solution Rqts</th>
<th>IP Rqts</th>
</tr>
</thead>
</table>
| Flash Control     | - ONFI 2.X/3.0  
                   - Toggle Mode 2.0  
                   - Multi flash load/ch  
                   - 40 GPIO/ch  
                   - PCIe Gen 3 x8  
                   - Low power & cooling | - Flash Controller  
                   (bad block mgt and wear leveling)  
                   - Flash RAID  
                   - Cache controller  
                   - BCH core  
                   - PCIe config < 100msec  
                   - Host interface/APIs |
System IO Considerations
System Application Requirements
- Performance - bandwidth
- IO network
- Memory
- Latency
Hardened IP (HIP) Advantages
- Resource savings of 8K to 30K logic elements (LEs) per hard IP instance, depending on the initial core configuration mode
- Embedded memory buffers included in the hard IP
- Pre-verified, protocol-compliant complex IP
- Shorter design and compile times with timing closed block
- Substantial power savings relative to a soft IP core with equivalent functionality
PCI Express NVMe

- Scalable host controller interface for PCIe-based solid state drives
- Optimized command issue and completion path
- Benefits
  - Software driver standardization
  - Direct access to flash
  - Higher IOPS and MB/s
  - Lower latency
  - Reduced Power Consumption
DRAM Cache Backup

- Data Center server power outages continue
- Read/Write Consequences
  - Data Loss
  - Undetected errors in host application
- NVDIMM designs protect system integrity but…

<table>
<thead>
<tr>
<th>Battery Limitations</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shelf Life</td>
<td>One year max or 500 cycles</td>
</tr>
<tr>
<td>Disposal and Handling</td>
<td>Hazardous Waste Management</td>
</tr>
<tr>
<td>Data Storage Capacity</td>
<td>Up to 72 hours</td>
</tr>
<tr>
<td>Down Time</td>
<td>Charge Time up to 6 hours</td>
</tr>
<tr>
<td>Replacement Cost</td>
<td>Field Time and Materials</td>
</tr>
</tbody>
</table>
The Perfect Storm

- **Technology Enablers**
  - **Super Capacitors** are production worthy
  - **Flash memory** costs continue to decline
  - **FPGA** technology meeting power/performance/cost
NVDIMM Controller Architecture

On power failure these FETs switch out the processor signals.

DIMM

Can be
- Buffered
- Un-buffered
- Registered

400MHz / 800MB tested

I2C

Control signals

FPGA
(Cyclone)

DDR ctrl with tri-state

Power failure switch

Individual CKE lines

Power regulation

Flash
1 or 2 SD Cards or BGA

To super-cap bank
Flashing Forward

- FPGAs are a great technology option for Data Centers
  - Networking: Port aggregation
  - Compute: Application Acceleration
  - Storage: Persistent Memory Control

- All development phases supported
  - Prototyping
  - Production
  - Test Validation
  - Upgrades