FPGAs in Flash Controller Applications and More

David McIntyre

Altera Corp.

dmcintyr@altera.com

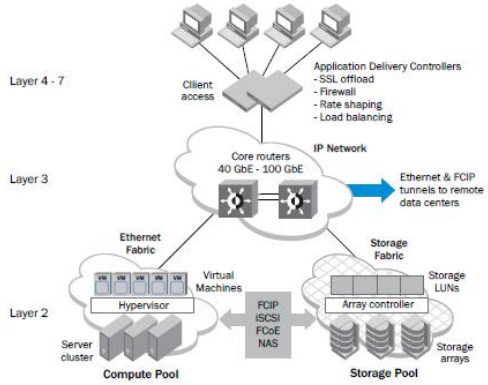
Flash Memory Summit 2014
Session 102-C
Tues August 5, 9:45am

Agenda

- FPGAs and Data Centers
- Emerging Memory Types
- Applications I
- Channel I/O Considerations
- Applications II
- Flashing Forward!

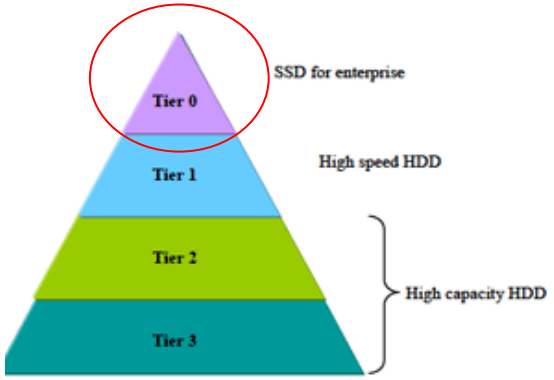
Data Center Trends

- **Network Convergence**
  - PCIe in the rack
  - Low latency 10GbE between racks
  - Switch aggregation

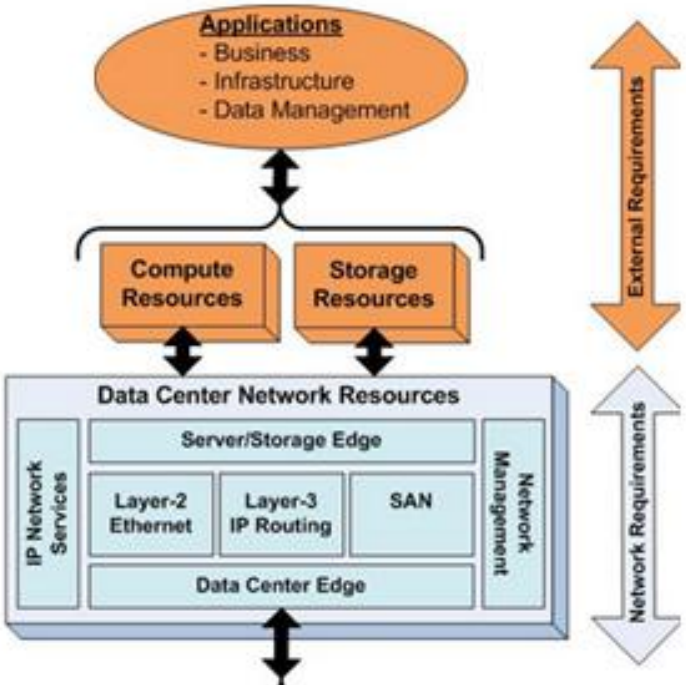


- **Algorithmic Acceleration**
  - Big Data
  - Cloud computing applications
    - Financial, Government, Scientific

- **Tiered Storage- Server Caching**
  - Application specific caching
    - Performance (speed and width, latency)
  - Non volatile memory types beyond Flash



FPGA Utilization across Data Centers



Point and SOC Solutions

- Application Acceleration
- Embedded Processing
- I/O Protocol Support
- Memory Control
- Compression
- Security
- Port Aggregation & Provisioning

TEST & MEASUREMENT

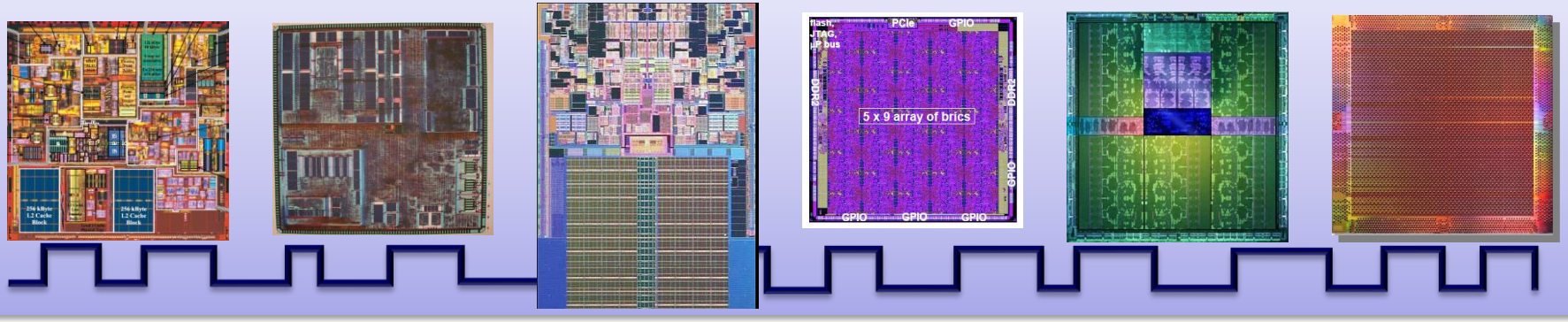
R&D, Integration	Protocol Conformance Test	Application Performance Trials	Deployment	Monitoring & Upgrades



Application Acceleration

Processing Options

Technology scaling favors programmability and parallelism



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

Single Cores

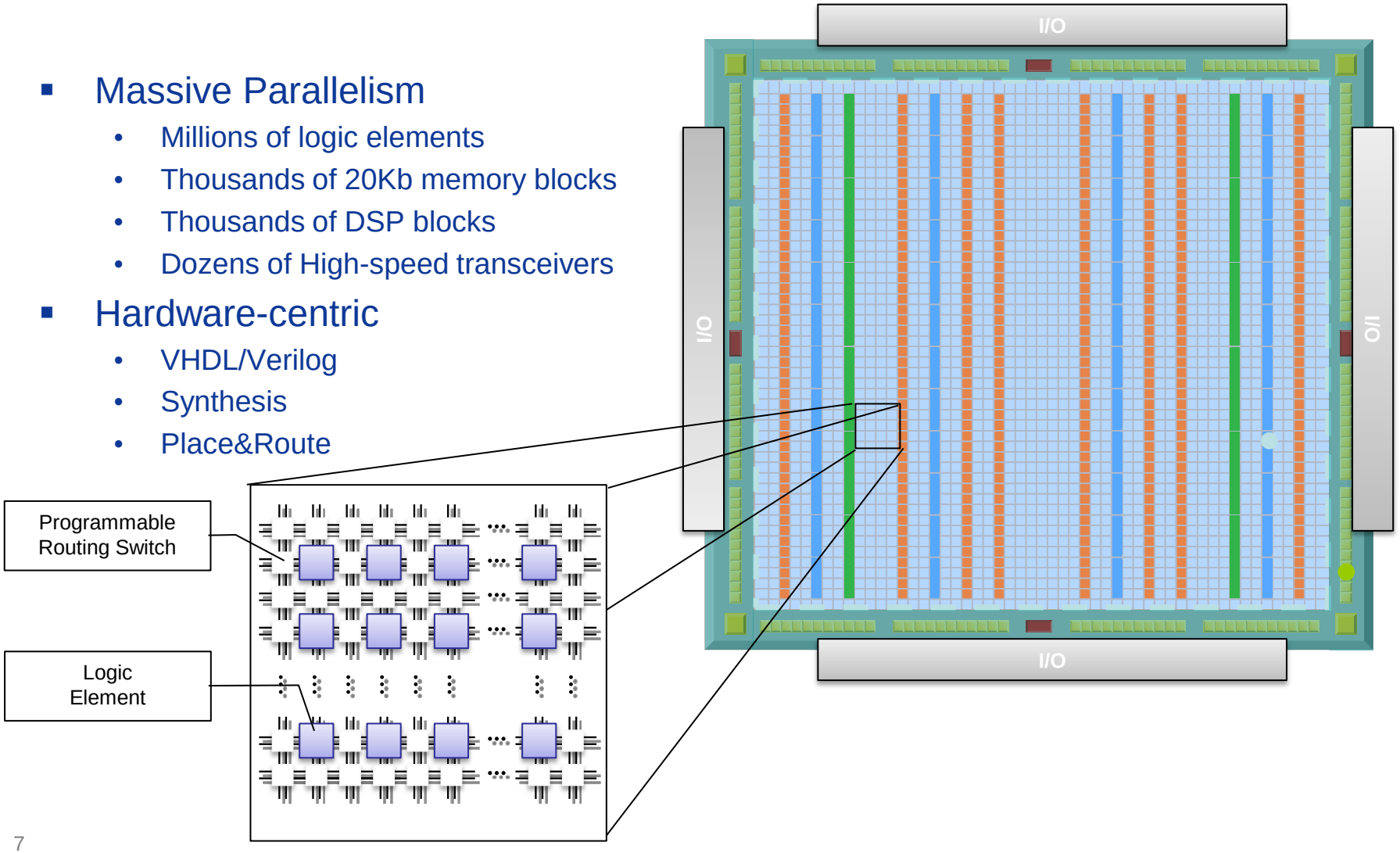
Multi-Cores
Coarse-Grained
CPUs and DSPs

Coarse-Grained
Massively
Parallel
Processor
Arrays

Fine-Grained
Massively
Parallel
Arrays

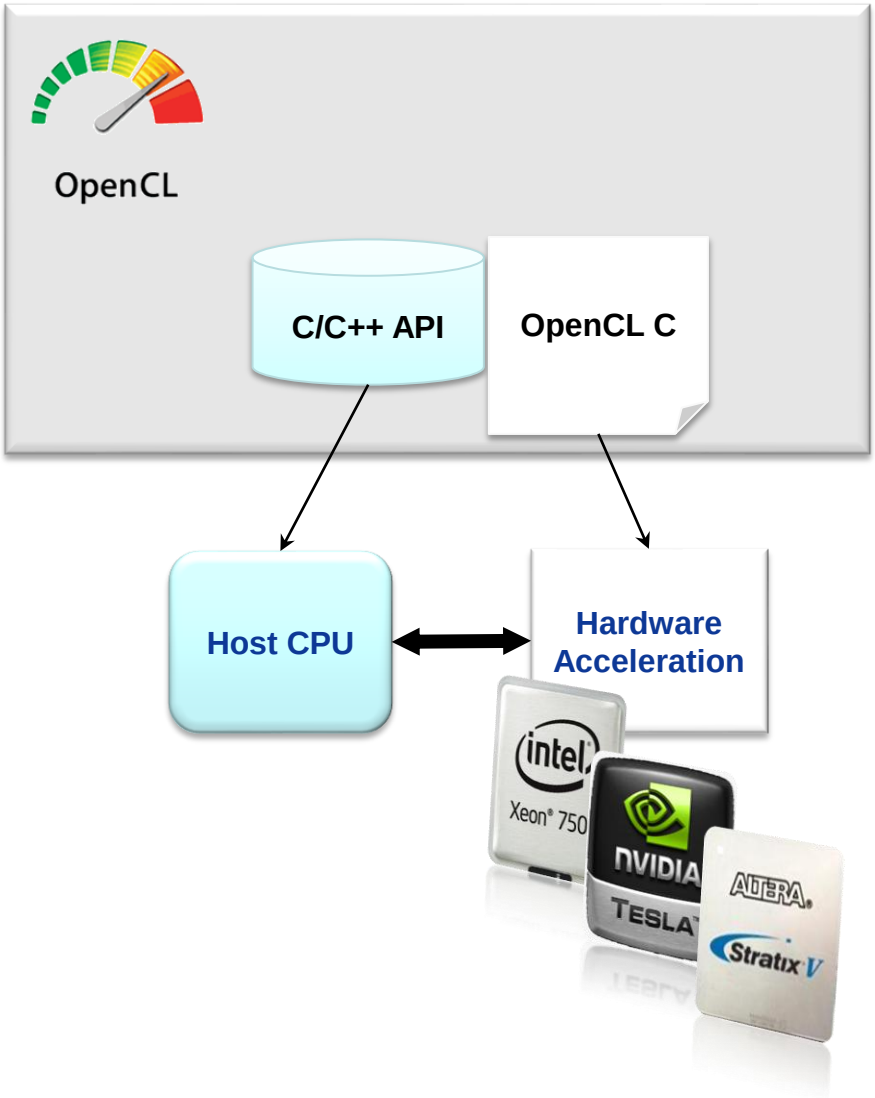
Altera FPGA Technology – Hardware Programming

- 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



OpenCL Overview

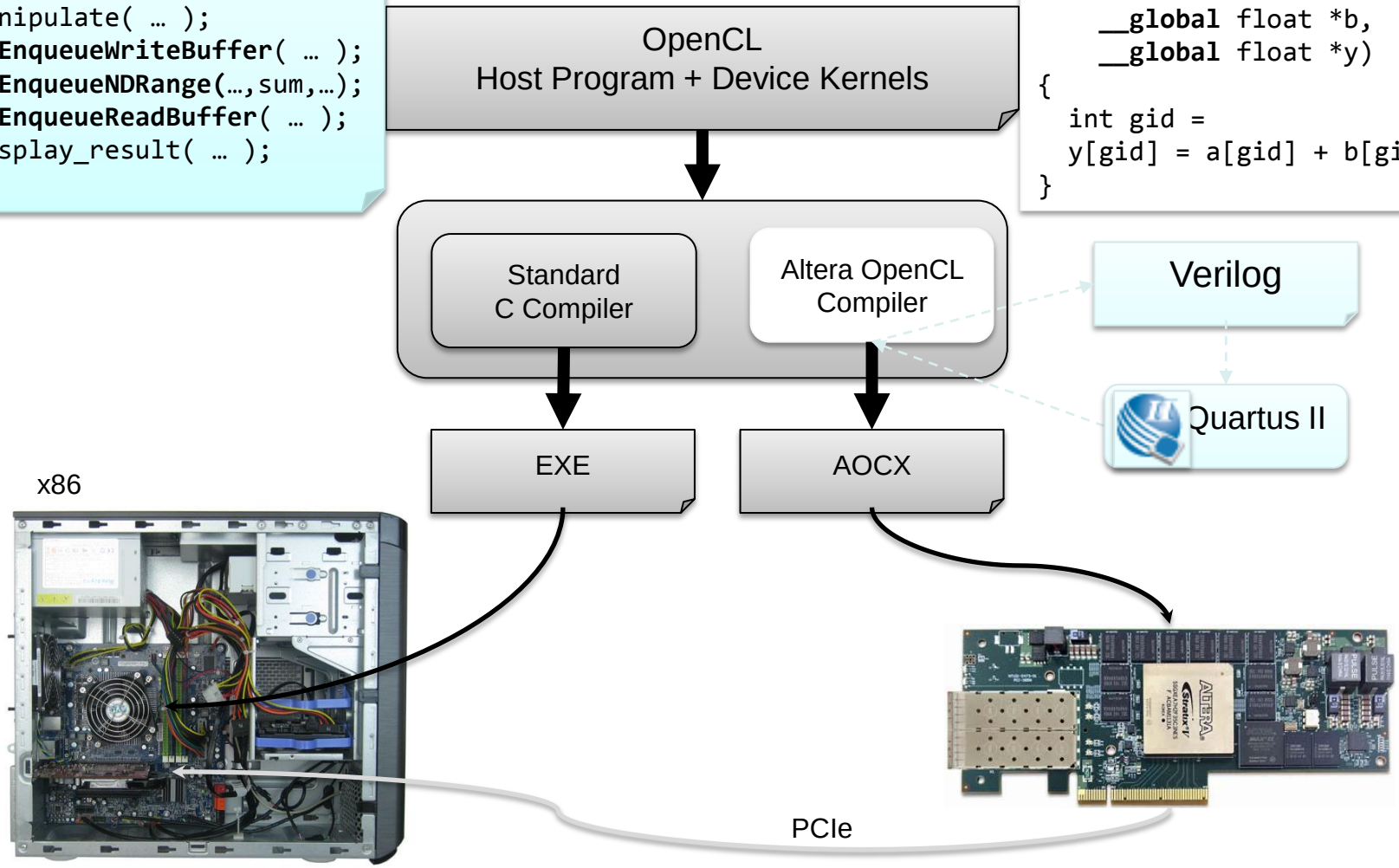
- Open Computing Language
  - Software-centric
    - C/C++ API for host program
    - OpenCL C (C99-based) for acceleration device
  - Unified design methodology
    - CPU offload
      - Memory Access
      - Parallelism
      - Vectorizaton
  - Developed and published by
    - http://www.khronos.org



OpenCL Process

```
main() {
   read_data( … );
   manipulate( … );
   clEnqueueWriteBuffer( … );
   clEnqueueNDRange(…,sum,…);
   clEnqueueReadBuffer( … );
   display_result( … );
}
```

```
__kernel void
sum(__global float *a,
    __global float *b,
    __global float *y)
{
  int gid =                (0);
  y[gid] = a[gid] + b[gid];
}
```



Multi-Asset Barrier Option Pricing

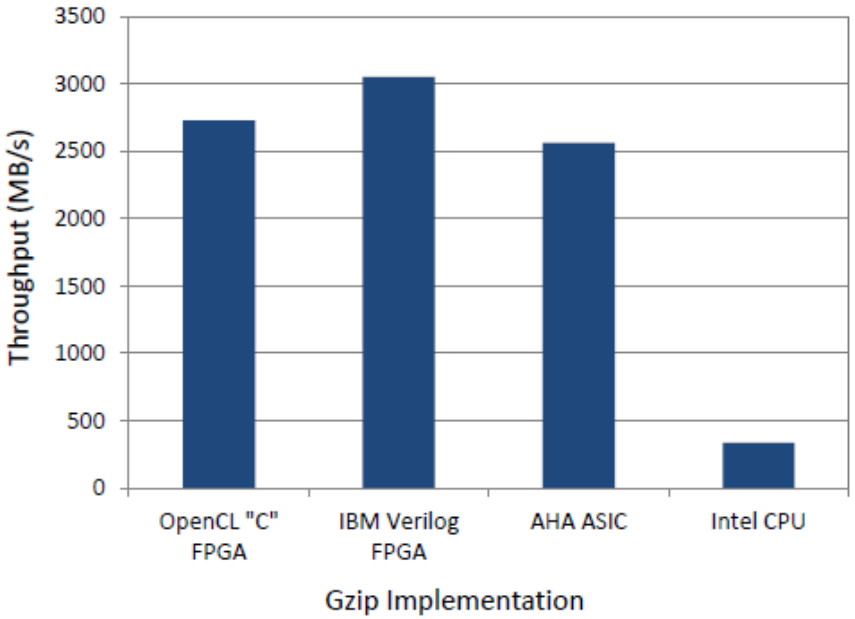
- Monte-Carlo simulation
  - Heston Model

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^S$$
$$dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t} dW_t^v$$

  - ND Range
    - Assets x Paths (64x1000000)
- Advantage FPGA
  - Complex Control Flow
- Results

	Power (W)	Performance (Msims/s)	Msims/W
W3690 Xeon Processor	130	32	0.25
nVidia Tesla C2075	225	63	0.28
PCIe385 D5 Accelerator	23	170	7.40

GZIP Comparison



Same compression ratio

12X better performance/Watt

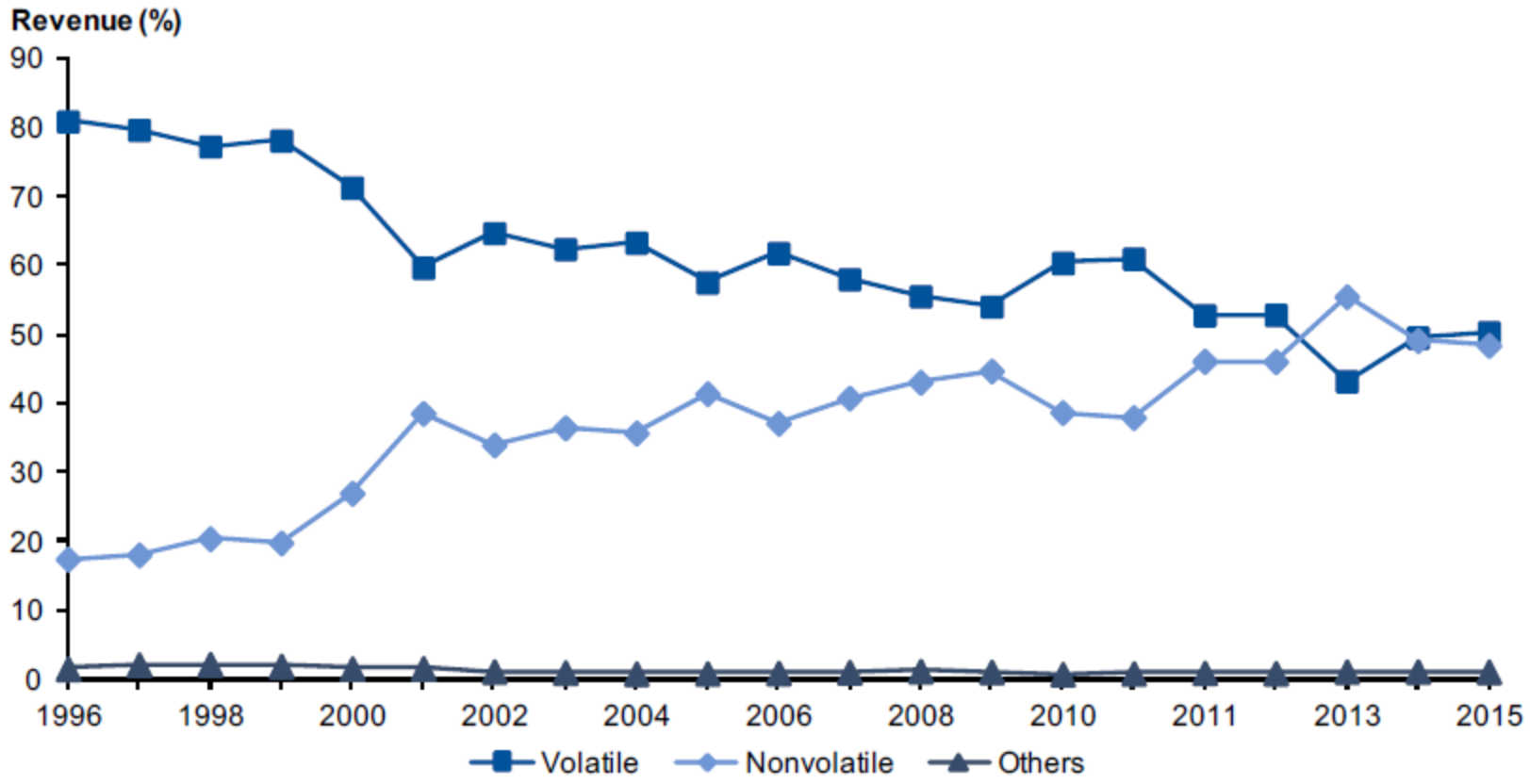
TABLE I: Comparison between our OpenCL FPGA and the best CPU implementation of Gzip.

	Performance	Performance/Watt	Compression Ratio
OpenCL FPGA	2.71 GB/s	111 MB/J	2.17×
Intel Gzip	338 MB/s	9.26 MB/J	2.18×
Gap	8.2× faster	12× better	on par

Storage Trends

Volatile vs NVM Market Share

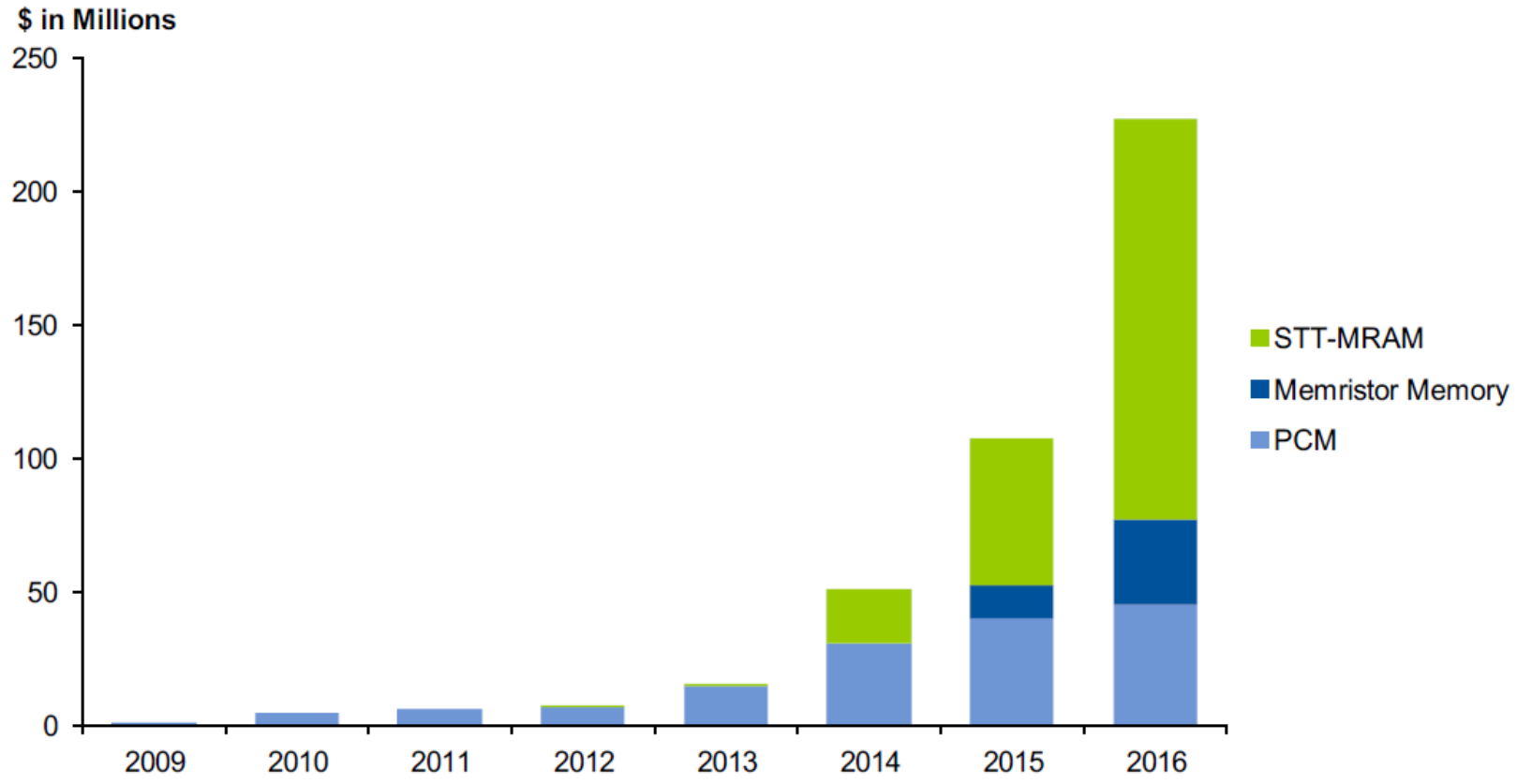
Figure 2. Total Memory Revenue Share by Memory Type, Worldwide, 1996-2015



Source: Gartner

Revenue Forecast

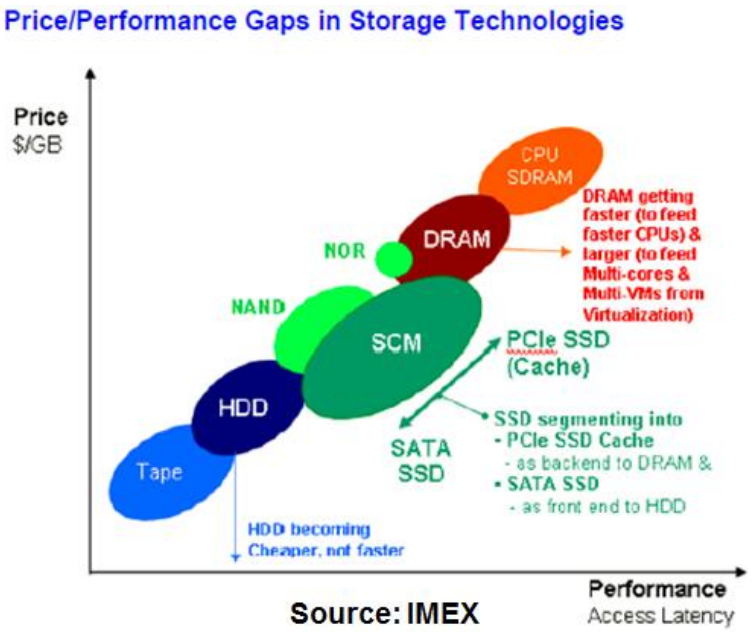
Figure 7. Revenue Forecast for Emerging Memory Technologies

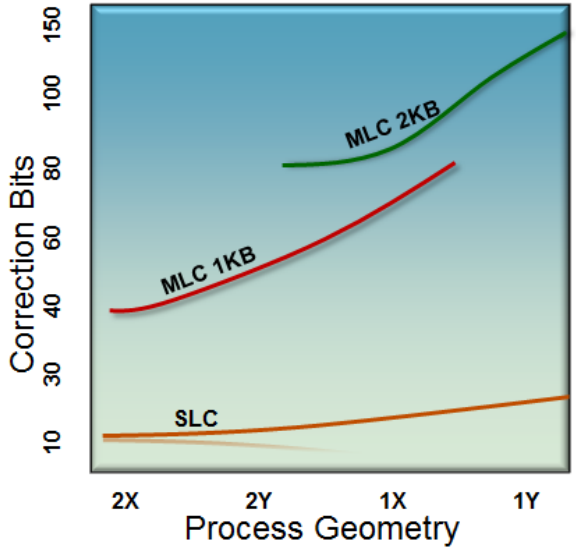


Source: Gartner

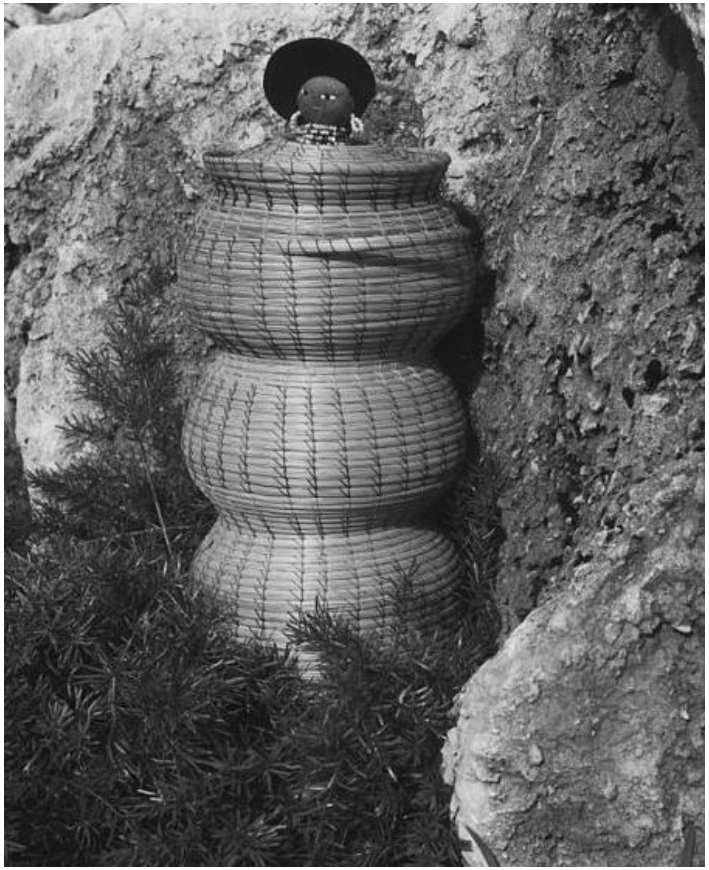
Flash Cache Challenges & Evolution

- 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
- Emerging flash technologies
  - MRAM (Magneto Resistive)
  - PCM (Phase Change)
  - RRAM (Resistive)
  - NRAM (Carbon Nanotube)



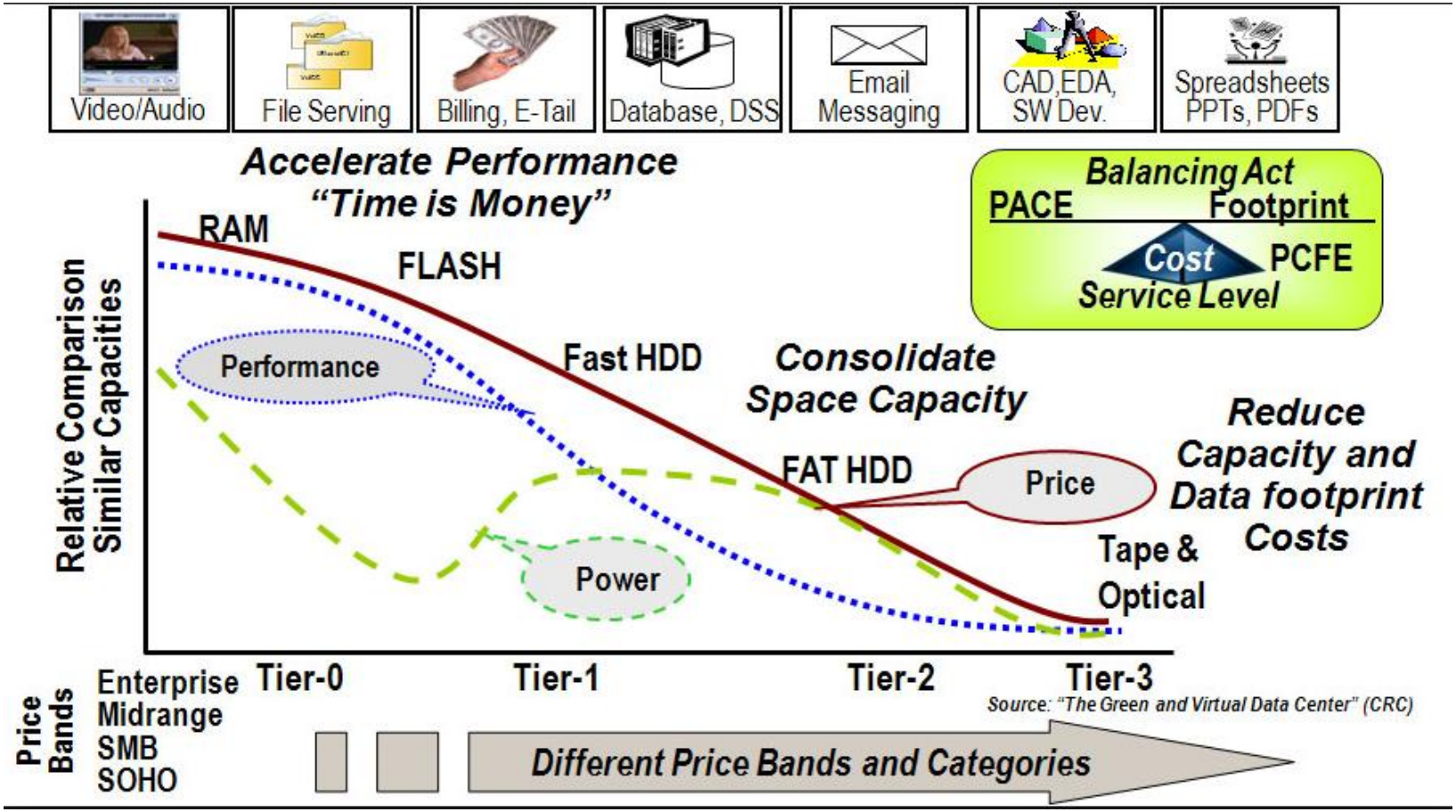


Tiered Storage, Then

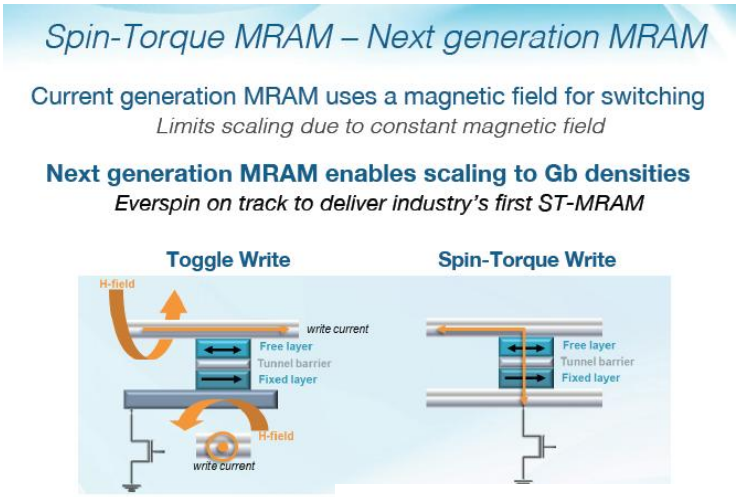


Seminole Indian Storage

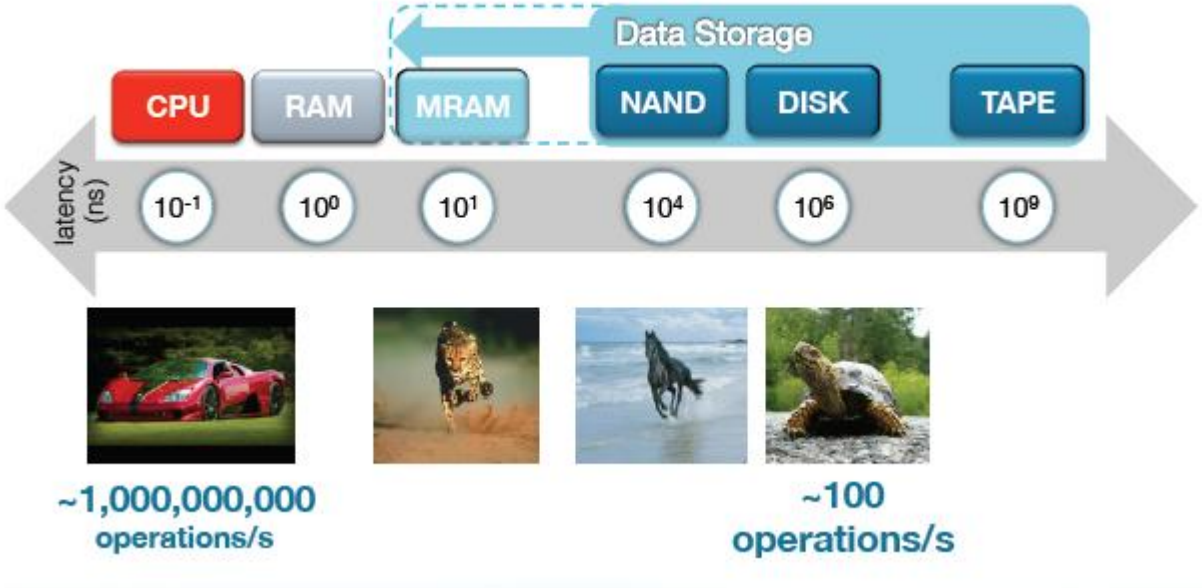
Tiered Storage, Now



Flexibility Required to Support Emerging Memory Technologies

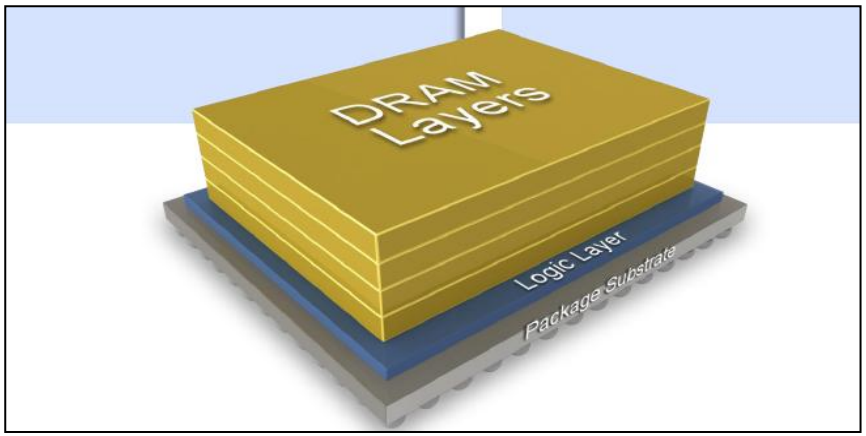


- HDD leveraged as capacity optimized data storage
  - *Benefits : Lowest cost per GB/TB for data storage*
  - *Challenges: Random access, active power & power fail*
- NAND SSD leveraged as performance optimized storage
  - *Benefits : More IOPS, reduced latency & less overall power*
  - *Challenges: Write latency & variability, endurance, power fail*
- ST-MRAM leveraged as non-volatile buffer/cache for storage
  - *Benefits : DRAM like access, unlimited endurance & power fail*
  - *Challenges: New storage architecture, density & cost scaling*



Hybrid Memory Cube (HMC)

- **HMC technology basics**
  - Ultra high performance, multi-bank DRAM memory
  - DRAM die stacked using state-of-the-art 3D process
  - Built-in memory controller with logic base die

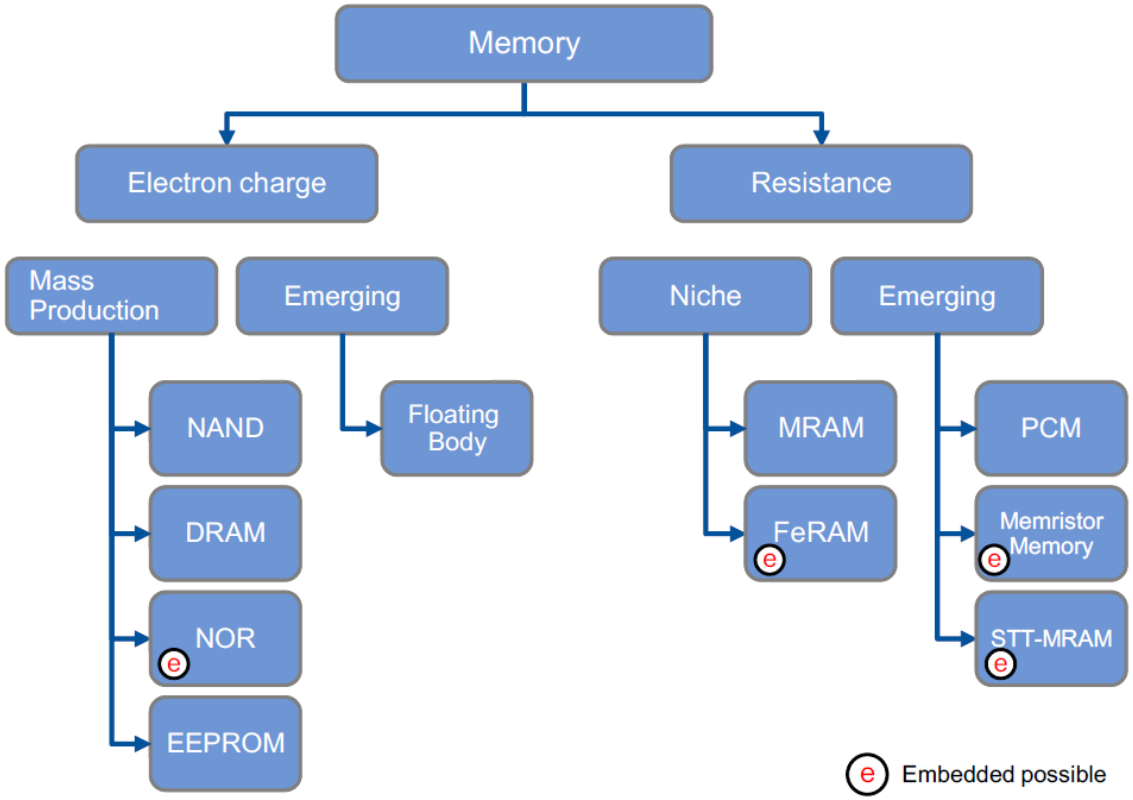


Parameters	Specification
Link	4
Speed	10,12.5,15 Gbps
Density	2GB, 4GB
Vaults	16
Banks	128, 256
DRAM B/W	160 GB/s (1.2 Tbps)
Vault B/W	10 GB/s (80 Gb/s)

- **Unparalleled gains with HMC**
  - Maximum DRAM bandwidth of up to 160 GB/s
  - Four links running at 15 Gbps offering nearly 1 T bps raw interface bandwidth
  - Up to 4GB density (storage) capacity, low PHY power (pj/bit)
  - Best in class RAS feature set

Memory Categories

Figure 1. Categories of Memory (Charge Versus Resistivity)



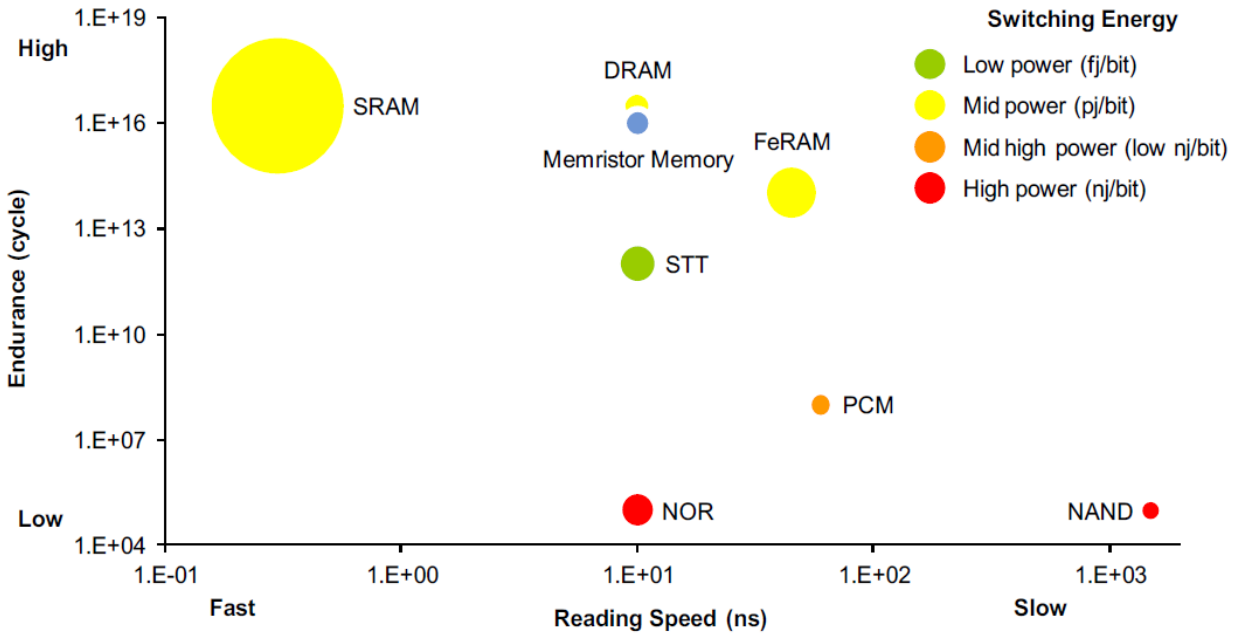
Key:
DRAM = dynamic RAM
EEPROM = electrically erasable programmable ROM
EPROM = erasable programmable ROM
FeRAM = ferroelectric RAM

MRAM = magnetoresistive RAM
PRAM = phase-change RAM
PSRAM = pseudostatic RAM
SRAM = static RAM

Memory Comparison

Figure 2 shows a comparison of these emerging memory technologies. We compare the properties of most memory technologies in reading speed (x-axis), endurance (y-axis), cell size (size of the ball) and switching energy (color of the ball).

Figure 2. Comparison of Emerging Memory Technologies



Source: Gartner

We estimate the comparative capability of these emerging memories in four areas:

- Scalability: Memristor memory > PCM > STT-MRAM
- Read Speed: Memristor memory ~ STT-MRAM > PCM
- Feature size (cost): STT-MRAM > PCM > Memristor memory

Memory Comparisons- Performance

Figure 3. Memory Comparison

	SST-MRAM	DRAM	NAND	SRAM	PCM	FRAM	MRAM
Status	Prototype	Product	Product	Product	Product	Product	Product
Cell Element	1T1MTJ	1T1C	1T	6T	1T1R	1T1C	1(2)T1R
Non-Volatile	Yes	No	Yes	No	Yes	Yes	Yes
Read Speed	★★★	★★★	★	★★★	★★	★	★★★
Write Speed	★★	★★★	★★	★★★	★	★★	★★★
Power Consumption	★★★	★	★★	★	★★	★★★	★
Endurance	★★★	★★★	★	★★★	★★	★★★	★★★
Scalability	★★★	★	★★	★	★★	★	★
MLC Capability	Yes	No	Yes	No	Yes	No	No
New Material	Yes	No	No	No	Yes	Yes	Yes
Vendors	Hynix IBM Renesas Samsung TDK Toshiba	Elpida Memory Hynix Micron Nanya Samsung	Intel Micron Samsung SanDisk Toshiba	Cypress GSI Renesas Samsung	Micron Samsung	Ramtron Texas Instruments	Crocus Technology Everspin Technologies

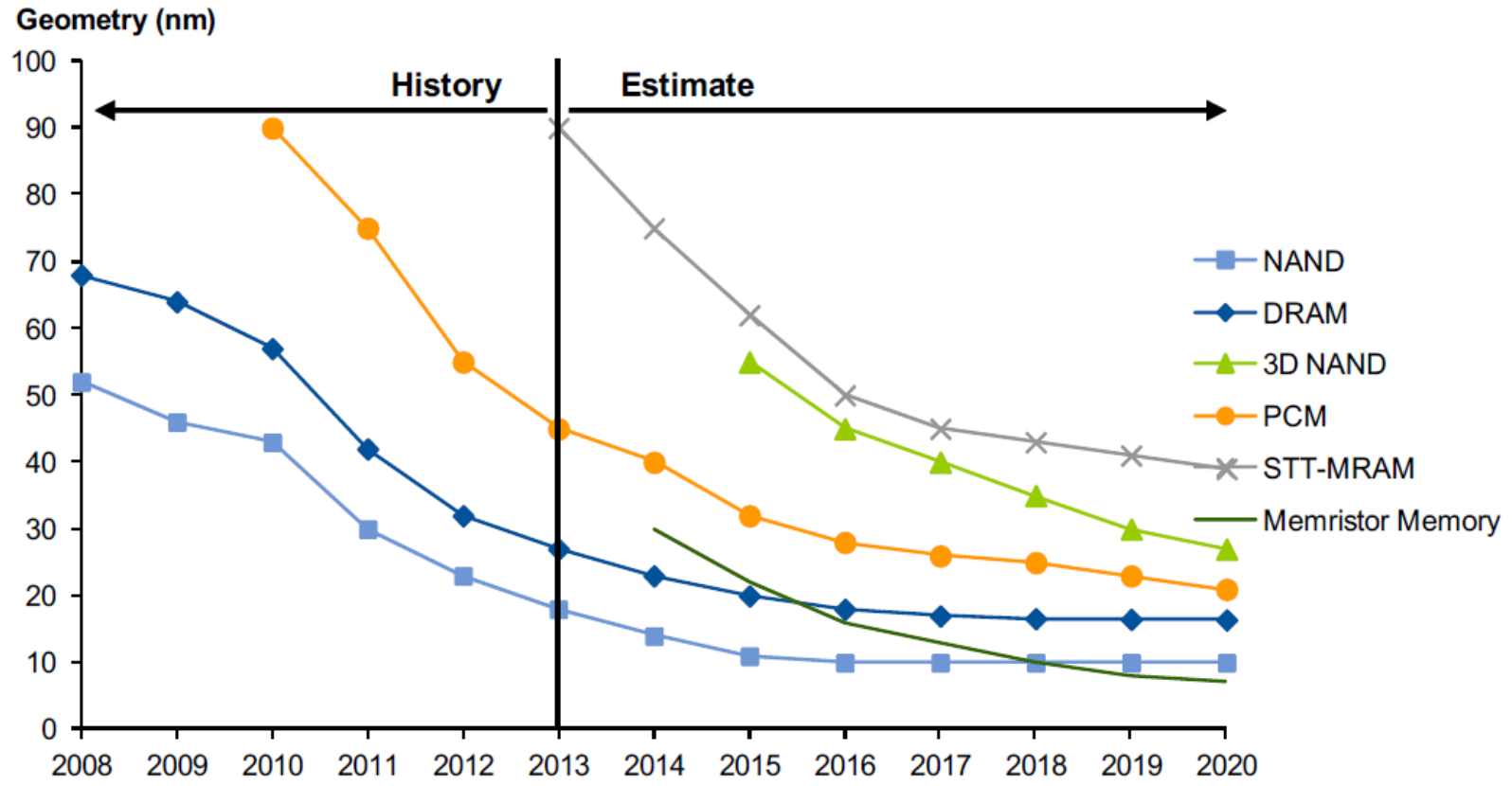
C = capacitor, FRAM = ferroelectric RAM, MLC = multilevel cell, MRAM = magnetoresistive RAM, MTJ = magnetic tunnel junction, PCM = phase-change memory, R = resistor, SRAM = static RAM, STT-MRAM = spin-transfer torque magnetic random-access memory, T = transistor

Key: 3 stars = Excellent; 2 stars = Good; 1 star = Reasonable. Red coloring signifies the best in each category.

Source: Gartner

Migration Timeline- Cost

Figure 6. Migration Timeline for Emerging Memory Technologies



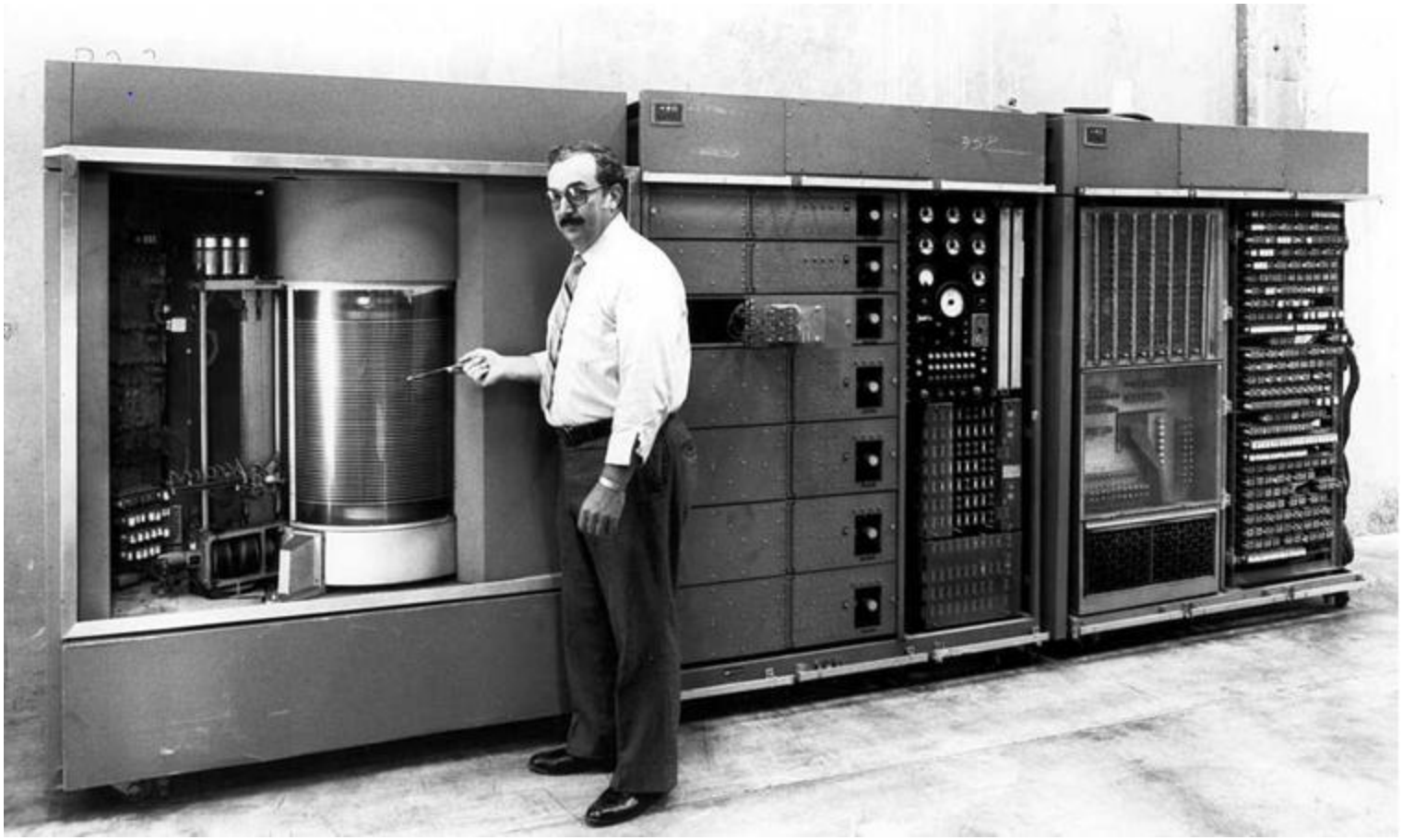
Source: Gartner

Flash Memory SUMMIT

5MB in Flight!

Flash Memory SUMMIT

Convergence.



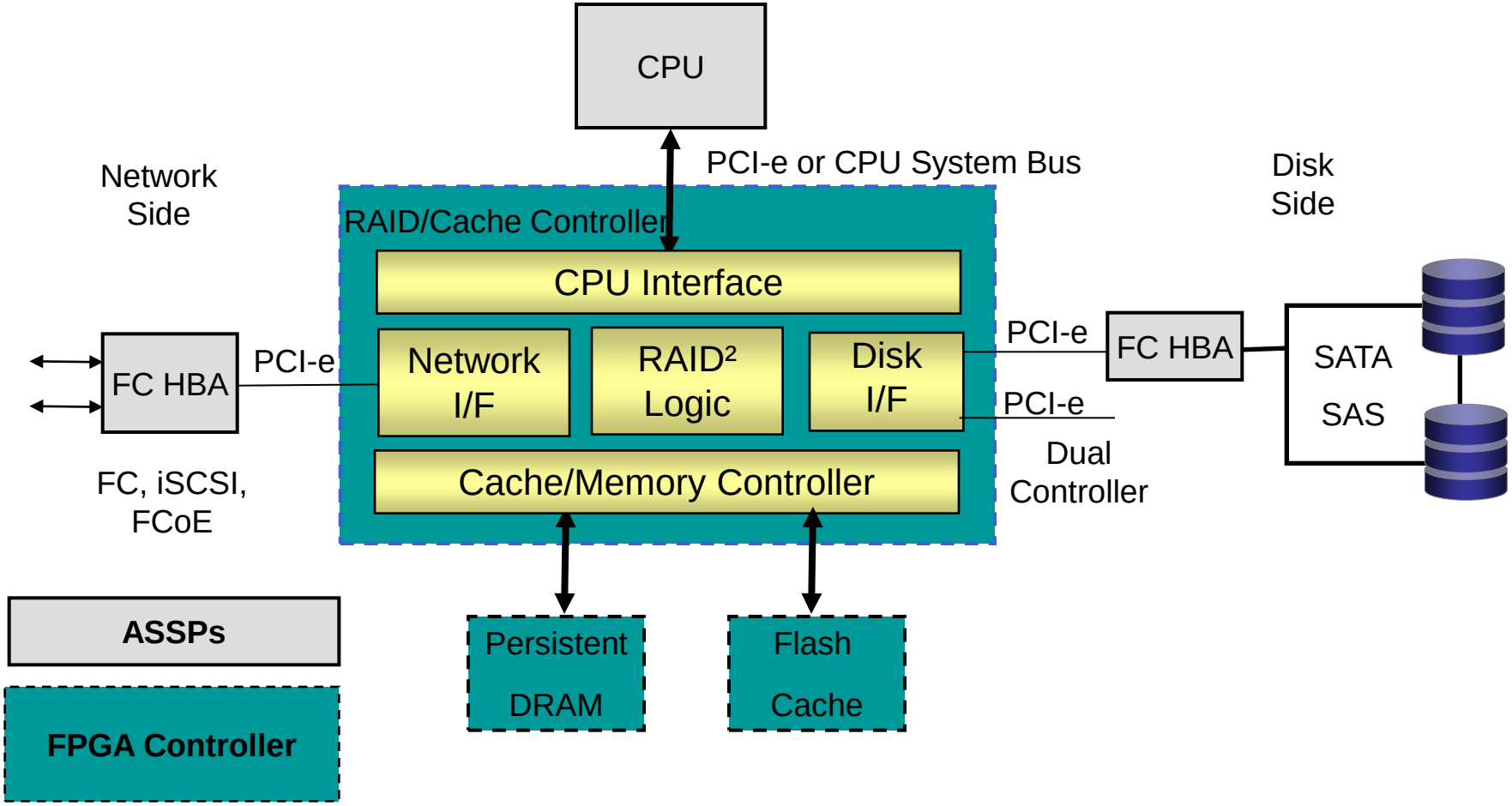
Data Center Applications- Servers

Application	Usage Examples
Flash SSD  Flash Controller	PCIe to ONFI bridging, Flash Control
Acceleration  Accelerator Card	Algorithm acceleration for vertical markets
Bridge Plus	Interface bridging with IP function, e.g. compression and encryption, Dedupe
I/O Virtualization (10GbE and PCIe) 	ASIC alternative; low cost with flexibility
Co-ASIC  Mainframe	Features enhancement
Management (BMC, KVM)  Blade Server	IP Flexibility supported with low power

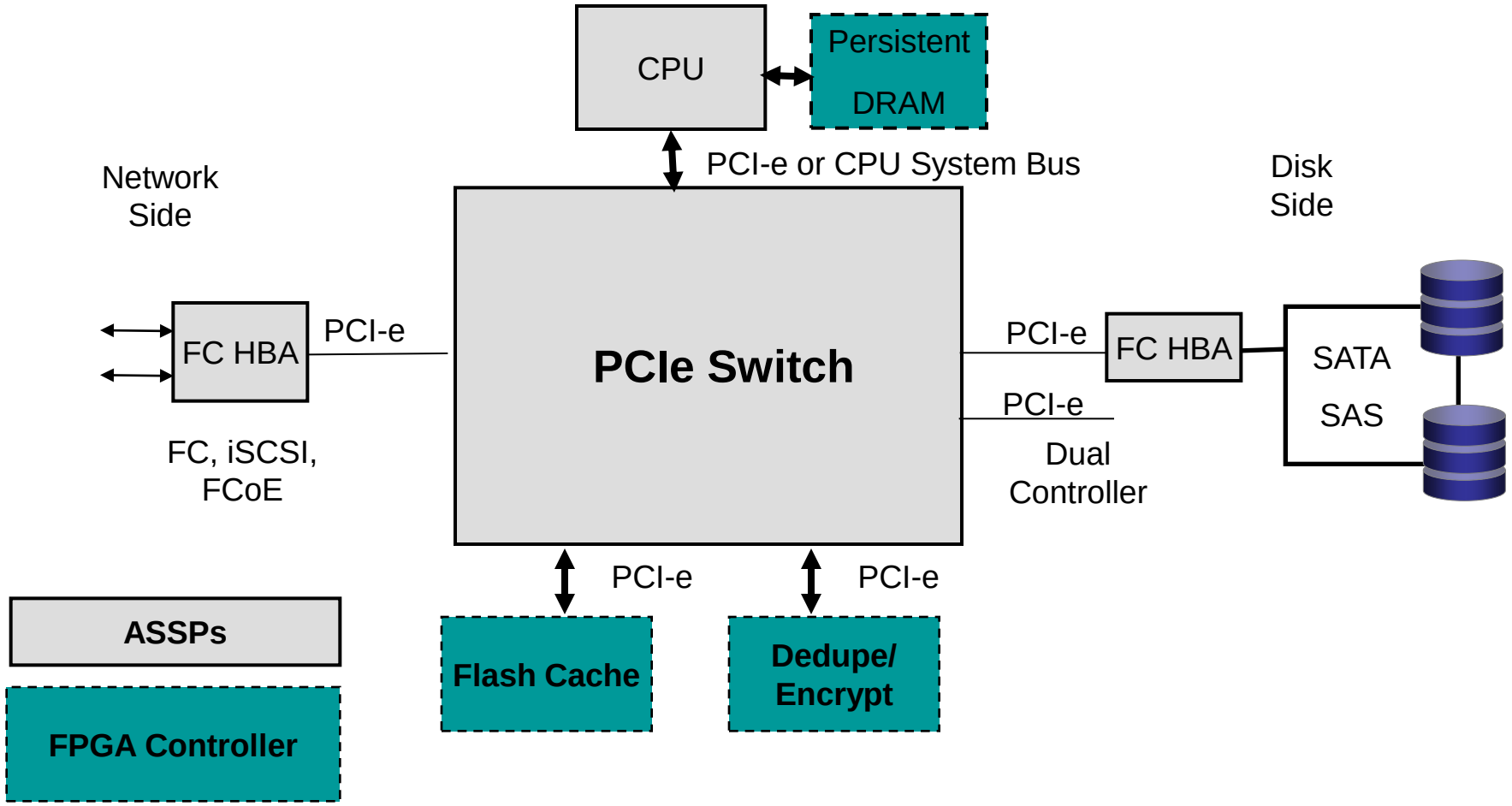
Data Center Applications- Storage

Application	Usage Examples
Flash Cache/SSD  Memory BackUp/Restore	ONFI bridging and RAID adaptor NV DIMM backup, RAID for Flash
RAID Bridging	PCIe Gen 3 x8 best of class signal integrity
Bridge Plus	Interface bridging with IP function
ASIC Replacement  Tape	Lower cost development with flexibility

Hybrid RAID System
- Persistent DRAM and Flash Caches



Hybrid RAID System
- PCIe Switch Centric



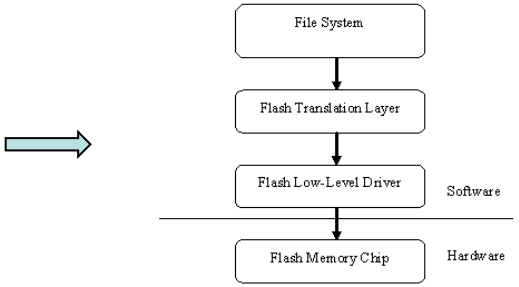
Flash Controller Design Considerations

Flash Controller Requirements

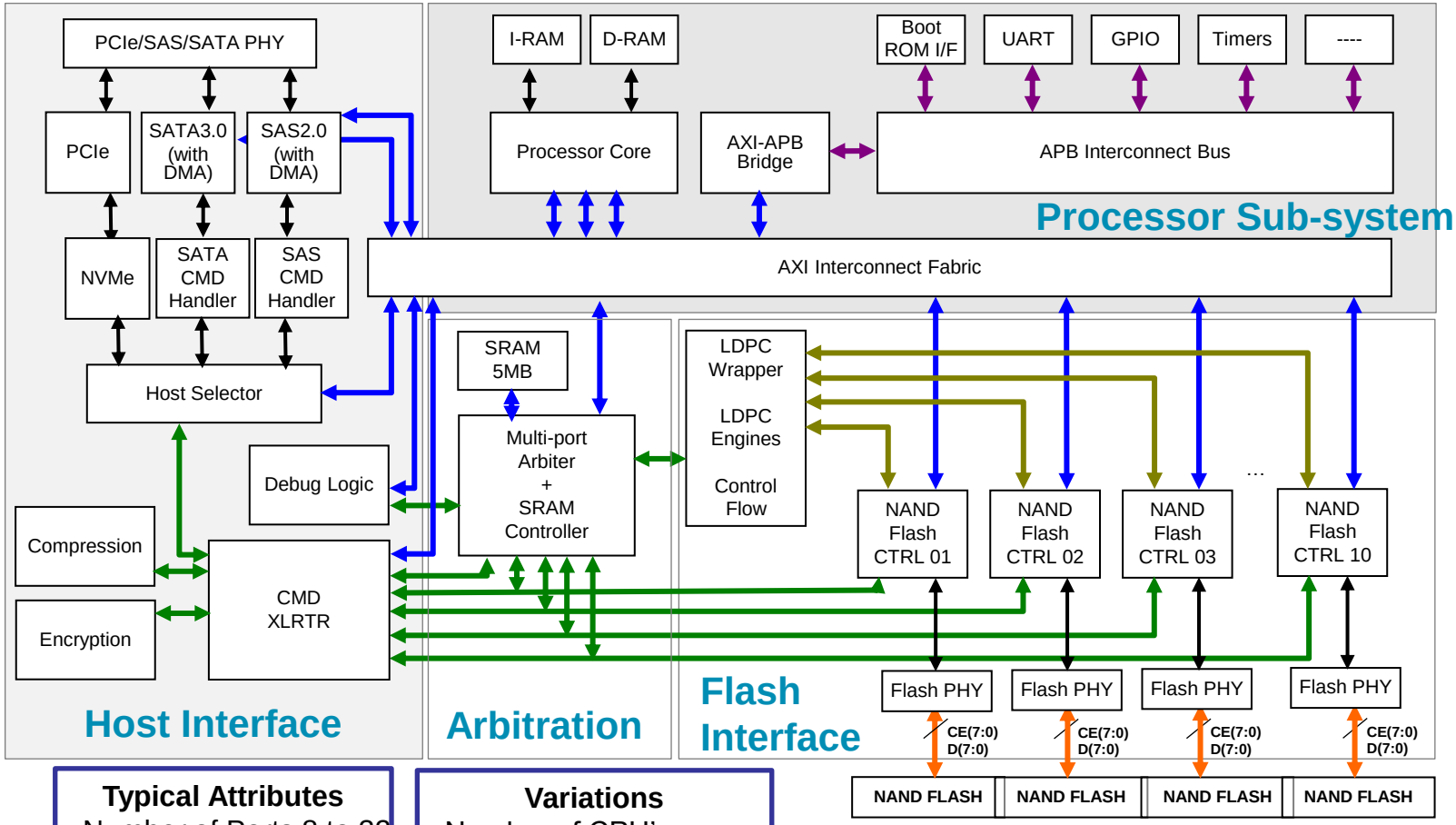
- Uncertainty Favors PLDs for Flash Control Solutions
- Flash Challenges Continue
  - Data loss, slow writes, wear leveling, write amplification, RAID
- Many Performance Options
  - Write back cache, queuing, interleaving, striping, over provisioning
- Many Flash Cache Opportunities
  - Server, blade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



Typical SSD Controller Architecture

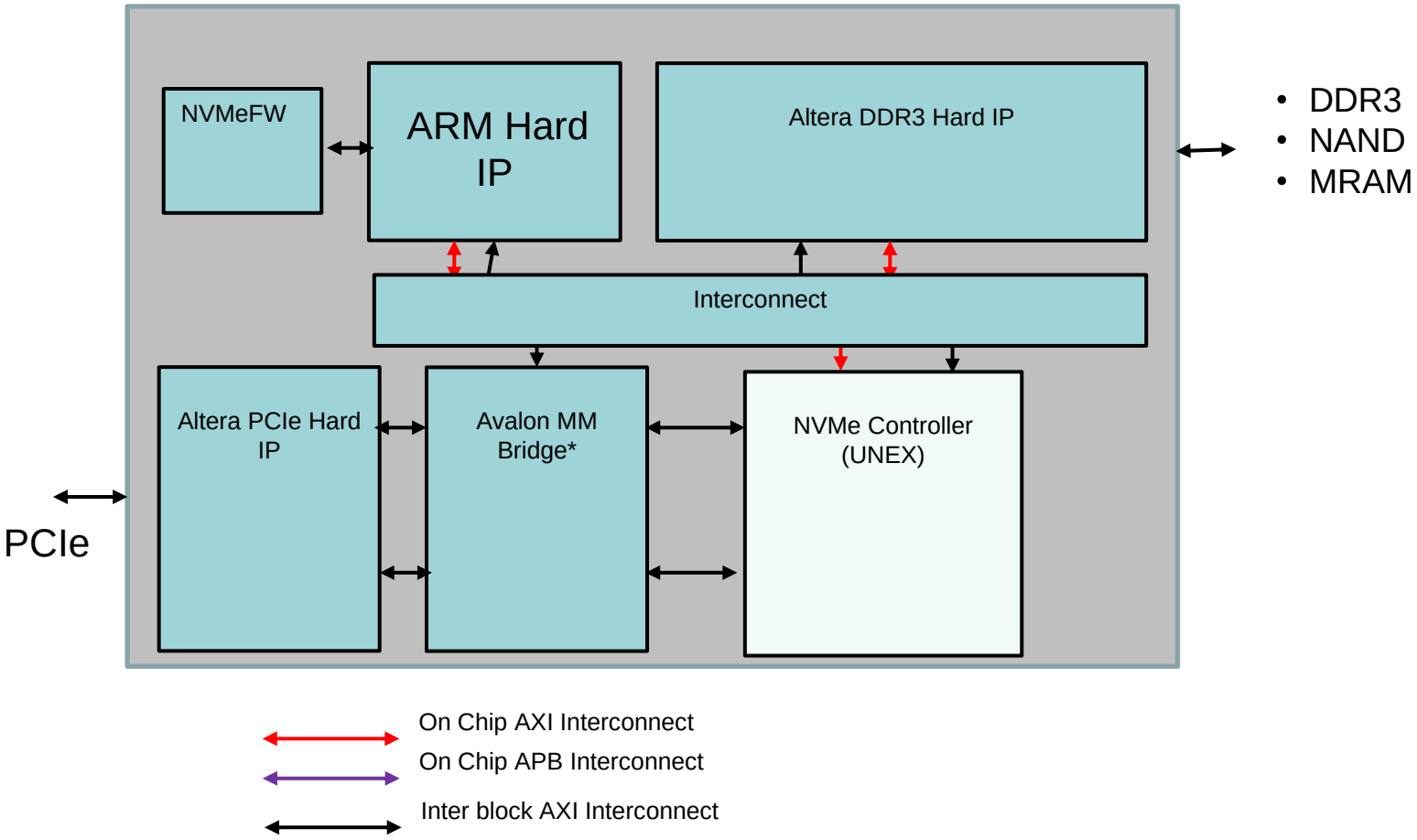


Typical Attributes
- Number of Ports 8 to 32
- Pin Count 250 to 1000+
- Power 1 to 3.5 Watts
- Internal, External RAM

Variations
- Number of CPU's
- Error Correction
- Interfaces
- Memory Type and Size

AXI Interconnect
APB Bus
AXI for Memory

Mobiveil NVMe SSDC Solution



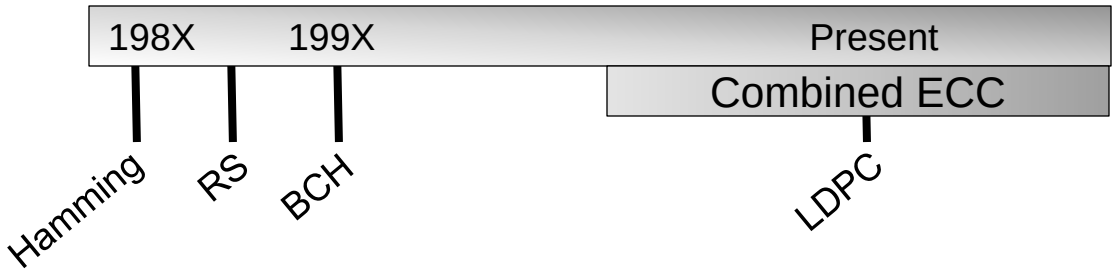
Error Correction Overview

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

Features	BCH	LDPC
Gate Count	High	Mid
Latency	Low	Medium
Tuneablity	low	high
Soft Data	no	high
Data Overhead	high	low



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

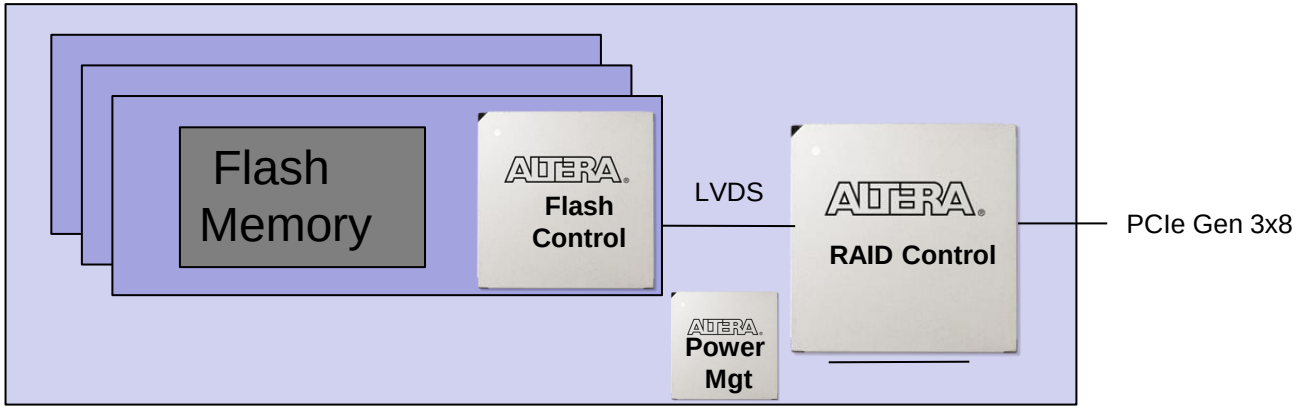
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	
0	0	1	0	0	1	1	1	0	0	0	0	$c_3 \oplus c_6 \oplus c_7 \oplus c_8 = 0$
1	1	0	0	1	0	0	0	0	0	0	1	$c_1 \oplus c_2 \oplus c_5 \oplus c_{12} = 0$
0	0	0	1	0	0	0	0	1	1	1	0	$c_4 \oplus c_9 \oplus c_{10} \oplus c_{11} = 0$
0	1	0	0	0	1	1	0	0	1	0	0	$c_2 \oplus c_6 \oplus c_7 \oplus c_{10} = 0$
1	0	1	0	0	0	0	1	0	0	1	0	$c_1 \oplus c_3 \oplus c_8 \oplus c_{11} = 0$
0	0	0	1	1	0	0	0	1	0	0	1	$c_4 \oplus c_5 \oplus c_9 \oplus c_{12} = 0$
1	0	0	1	1	0	1	0	0	0	0	0	$c_1 \oplus c_4 \oplus c_5 \oplus c_7 = 0$
0	0	0	0	0	1	0	1	0	0	1	1	$c_6 \oplus c_8 \oplus c_{11} \oplus c_{12} = 0$
0	1	1	0	0	0	0	0	1	1	0	0	$c_2 \oplus c_3 \oplus c_9 \oplus c_{10} = 0$

Flash Controller Support

IP	IO	Speed	Logic Density	Comments
ONFI 3.0	40 pins/ch	400 MTps	5KLE/ch	NAND flash control, wear leveling, garbage collection
Toggle Mode 2.0	40 pins/ch	400 MTps	5KLE	Same
DDR3	72 bit	1066 MHz	10KLE	Flash control modes available for NVDIMM
PCM			5KLE	PCM- Pending production $
MRAM			5KLE	MRAM- Persistent memory controller (Altera based)
BCH			<10KLE	Reference design
PCIe	G3x8	64Gbps	HIP	Flash Cache

Flash Storage Arrays

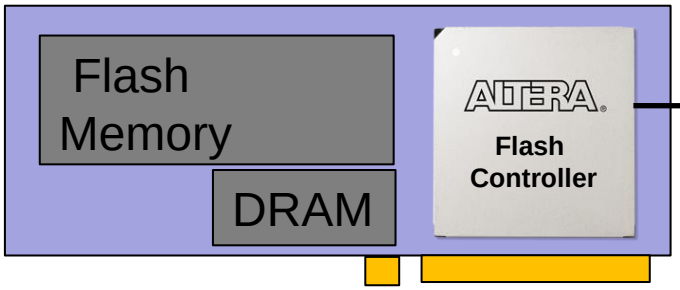
Target Application: Enterprise Tier-1 Storage: Databases and Virtualization



Function	Solution Rqts	Target Product	IP Rqts
Flash Control	-ONFI 2.X/3.0 -Toggle Mode 2.0 - Multi flash load/ch - 40 GPIO/ch	Arria V (28nm)	- Flash Controller (bad block mgt and wear leveling) - Metadata & caching - ECC BCH core
RAID Control	PCIe Gen 3	Arria 10 (20nm)	- Flash-specific RAID - Switching and aggregation
Power Mgt		Enpirion	

Flash PCIe Cards

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



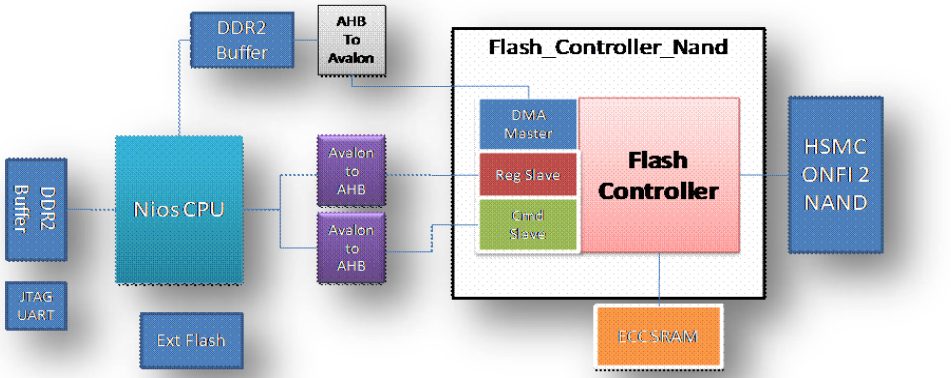
PCIe: Gen 3x8

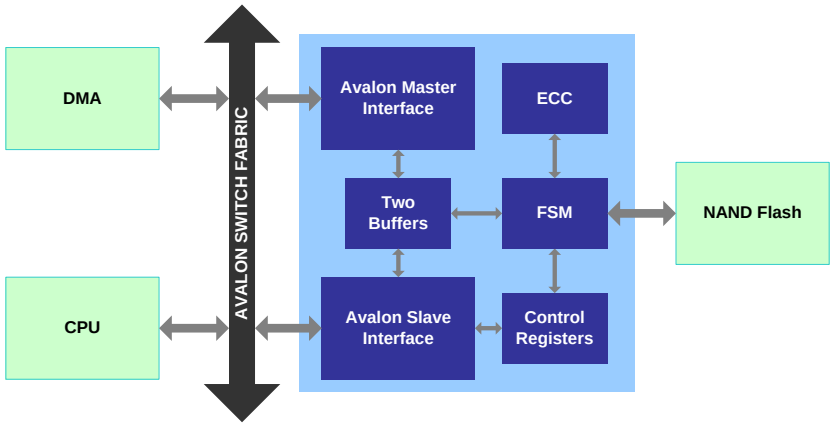
FPGA controller provides flexiblity to integrate multiple complex functions and adapt to changing interfaces & APIs.

Function	Solution Rqts	Target Product	IP Rqts
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	Arria V Arria 10	- Flash Controller (bad block mgt and wear leveling) -Flash RAID -Cache controller - BCH core -PCIe config < 100msec -Host interface/APIs

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**

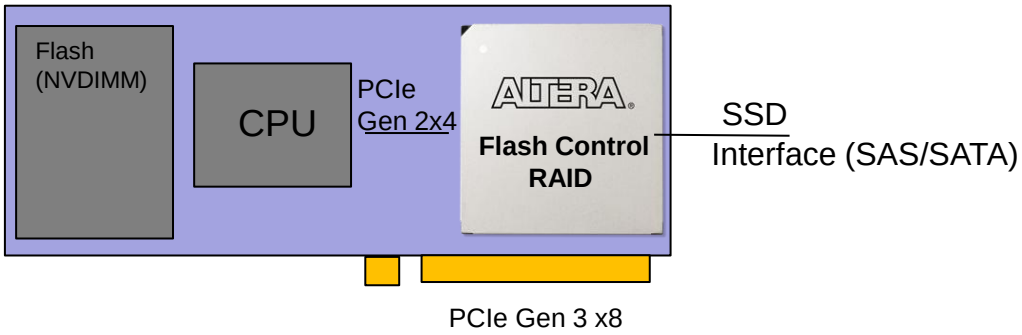




PCIe to RAID controller

Target Application

Embedded storage for flash cache and high performance computing

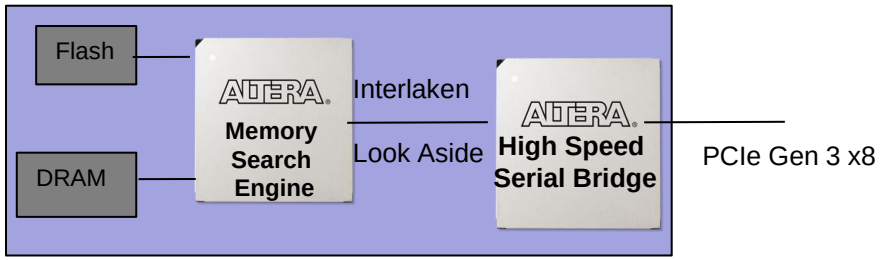


Function	Solution Rqts	Target Product	IP Rqts	IP Partner
RAID Control	-ONFI 2.X/3.0 -Toggle Mode 2.0 - Multi flash load/ch -40 GPIO/ch -PCIe Gen 3 -6Gb SAS/SATA	Arria 10	Flash Controller RAID PCIe config < 100msec 6Gb SAS/SATA	SLS CAST Altera Reference Design (3Q12) CEVA/Inteliprop

Memory Lookup and Cache

Target Application

Embedded Memory search and cache for high performance computing



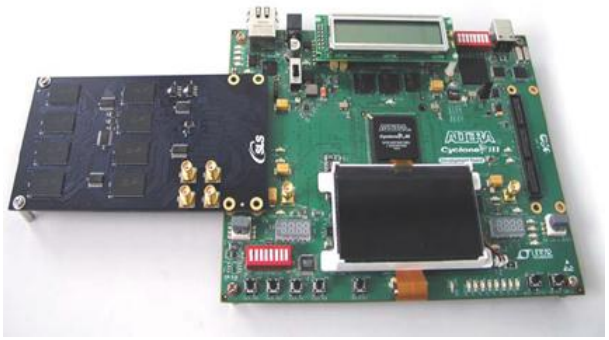
Function	Solution Rqts	Target Product	IP Rqts	IP Partner
Memory Control	-ONFI 2.X/3.0 -Toggle Mode 2.0 - Multi flash load/ch -40 GPIO/ch -PCIe Gen 2 x8/Gen3 x8 -Interlaken	Arria V Arria 10	Flash Controller PCIe config < 100msec 6Gb SAS/SATA “Look-aside variant”	SLS CAST

Development Platforms

Stratrix/Arria FPGA

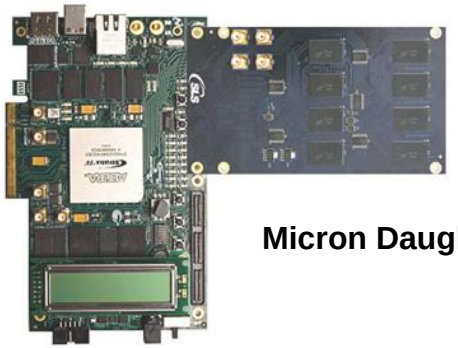
Everspin
64MB DIMM

Introspect
I/O Tester



Cyclone Development Board

Gidel ProceV Board



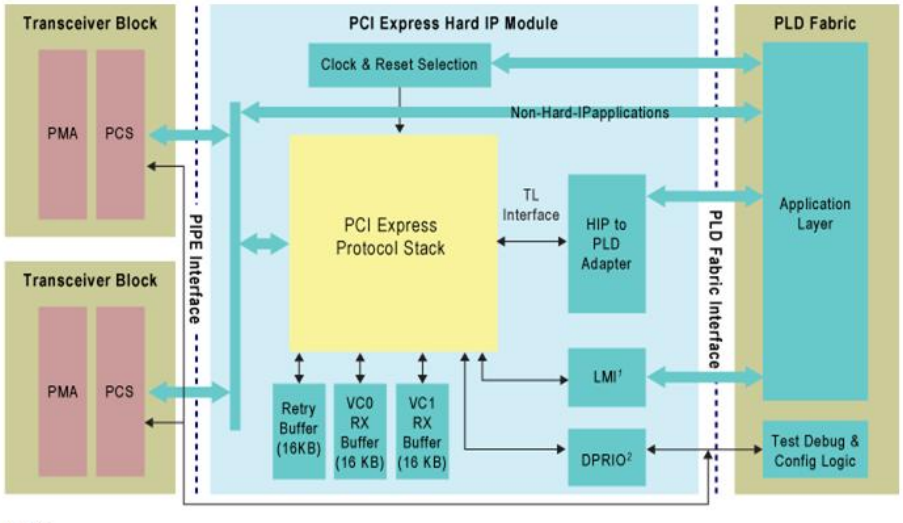
Micron Daughter Card

System IO Considerations

System IO

- System Application Requirements
  - Performance- bandwidth
  - IO network
  - Memory
  - Latency

PCI Express Support



Note:

1. LMI: Local Management Interface
2. DPRIO: Dynamic Partial Reconfigurable Input/Output

PCIe Mode	Thruput (GT/s per lane)	Production
Gen 2	5.0	Now
Gen 3	8.0	Now
Gen 4	16.0	2016

Hardened IP (HIP) Advantages

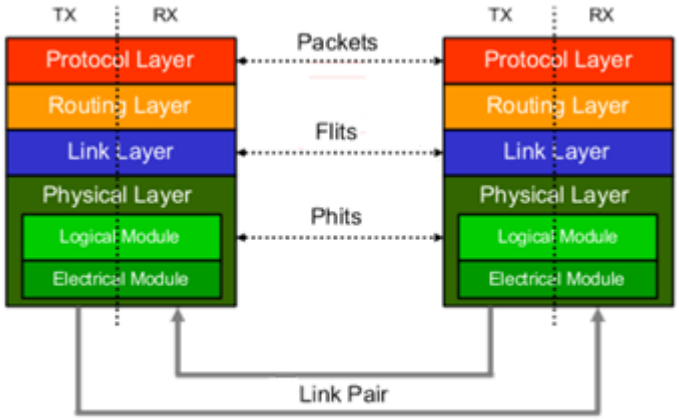
- Resource savings of 8K to 30K logic elements (LEs) per hard IP instance, depending on the initial core configuration mode
- Embedded memory buffers included in the hard IP
- Pre-verified, protocol-compliant complex IP
- Shorter design and compile times with timing closed block
- Substantial power savings 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

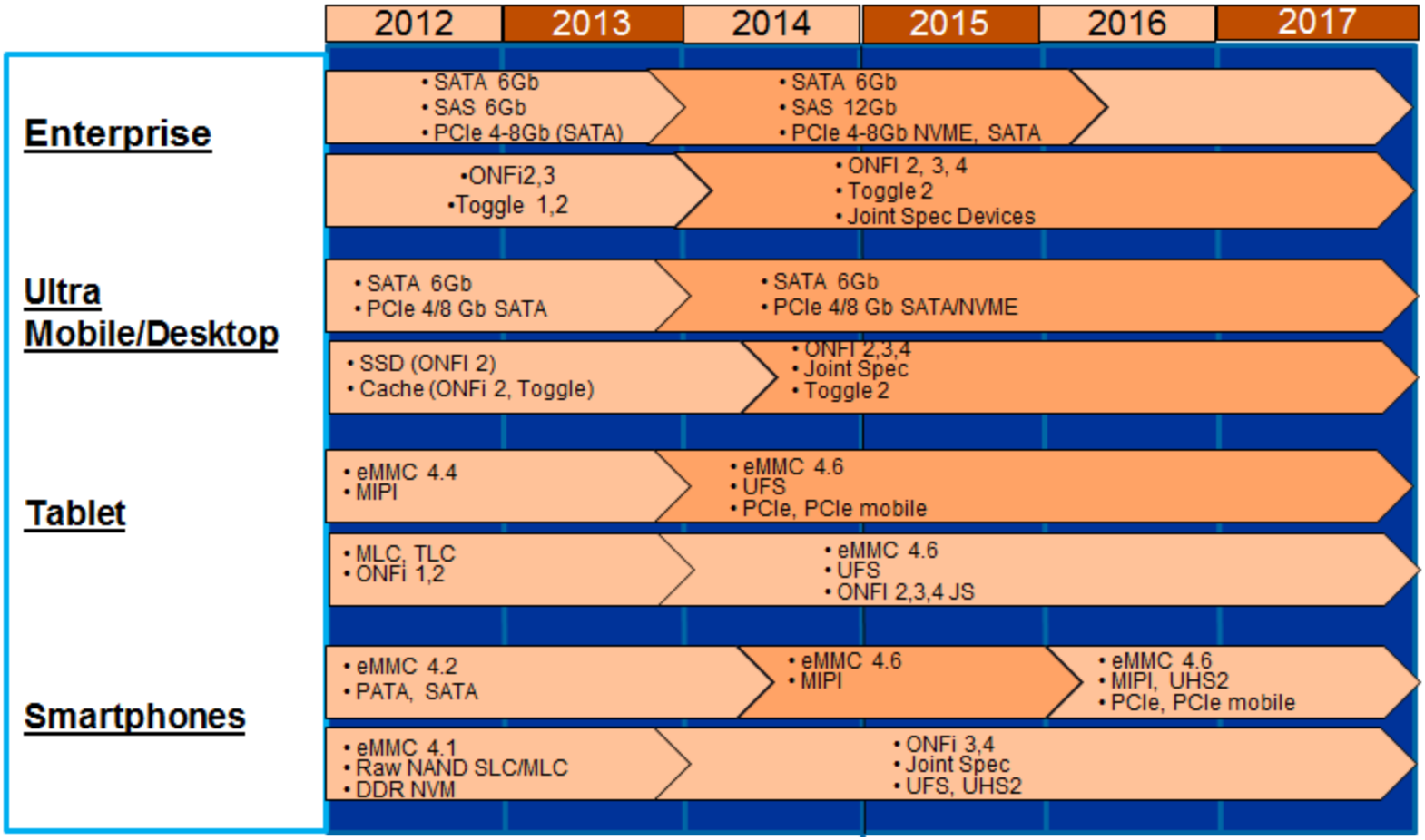
Intel QPI

- Up to (2) full width QPI 1.1
  - 6.4, 8.0 Gbps
- Designed for QPI Electricals
  - Common Mode Voltage
  - Lane detection on die
- **Chip to chip interconnectivity**
- Fast HP 28nm. process
- Hard PHY + Upper Layers Option
  - Decrease latency, power, & fpga logic
  - Use “Embedded Hardcopy Blocks”
  - Improve throughput by 75+%
    - 6-8GB/s data payload each way

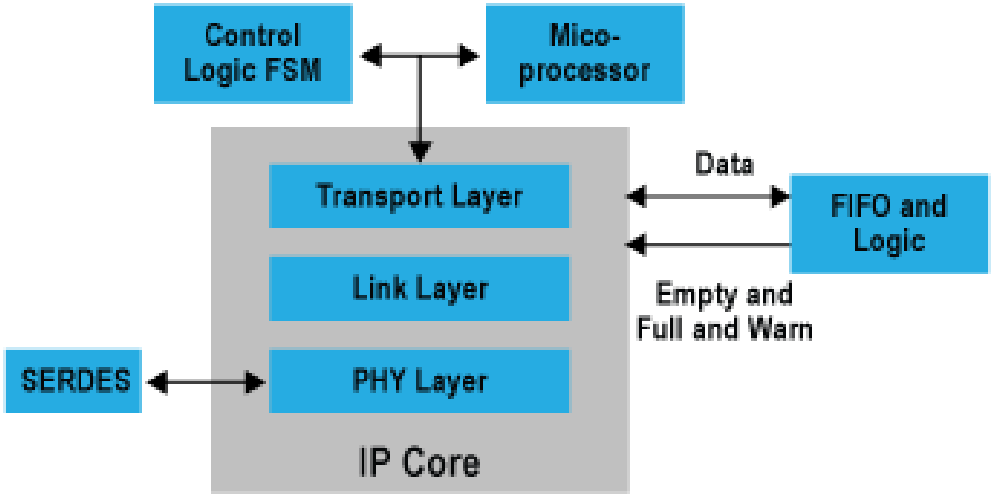


Note: Routing Layer not used

Industry Interface Convergence



SAS/SATA

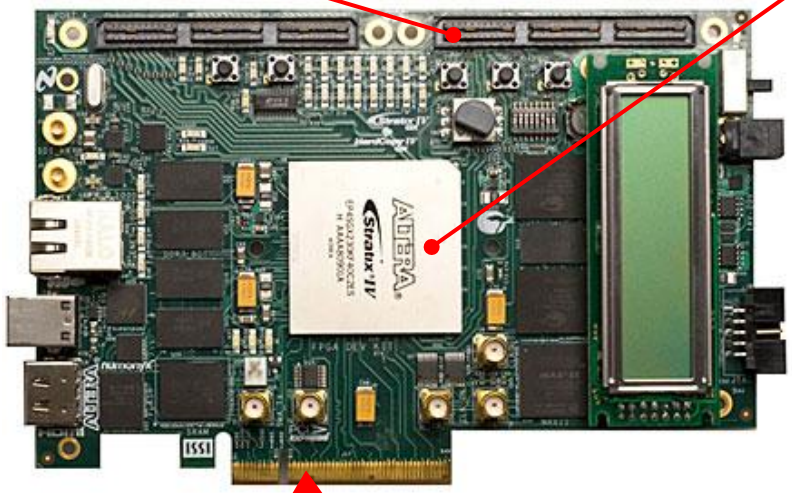


- Flash Controllers manage SAS/SATA SSD interfaces
- 12Gbps SAS support required for enterprise drives
- FPGA transceivers need to support electrical performance and OOB signaling

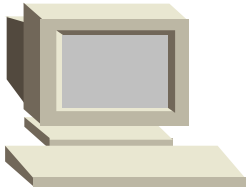
6G SAS Demo

SAS Cable Connection to SAS peer (via connector mezzanine board)

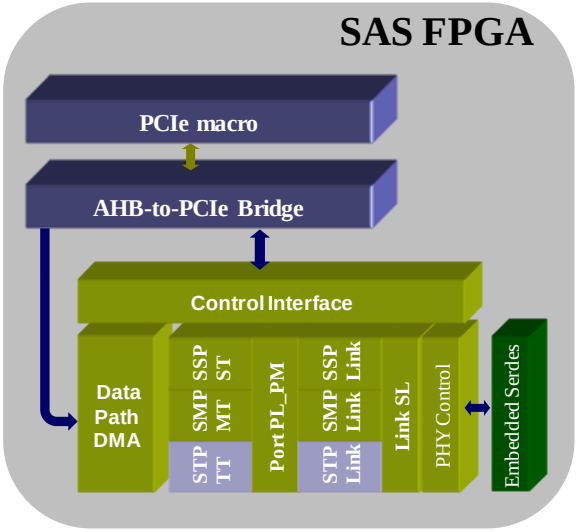
SAS peer is Lecroy SAS Sierra M6-2 or various SAS HDD



PCIe Interface



PC running SSD Emulation Windows Driver or CEVA SASTool test software (as if embedded processor)



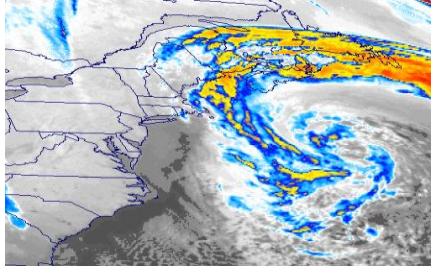
FPGA Flash Applications- Part Two

DRAM Cache Backup

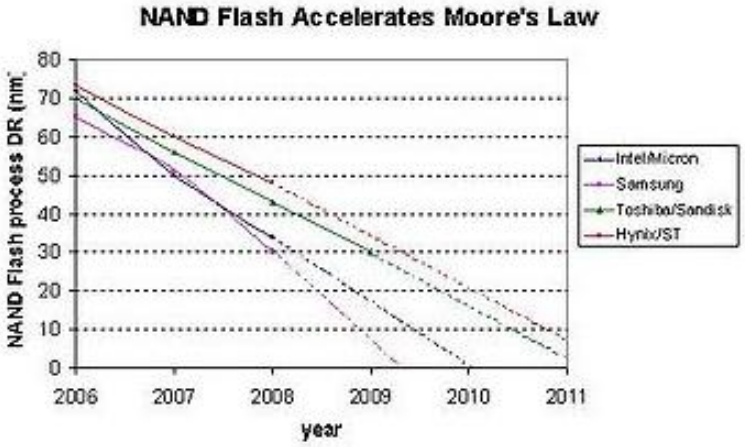
- Data Center server power outages continue
- Read/Write Consequences
  - Data Loss
  - Undetected errors in host application
- NVDIMM designs protect system integrity but…

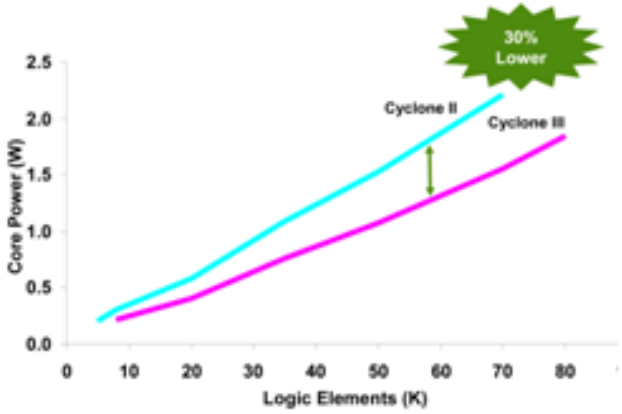
Battery Limitations	Issue
Shelf Life	One year max or 500 cycles
Disposal and Handling	Hazardous Waste Management
Data Storage Capacity	Up to 72 hours
Down Time	Charge Time up to 6 hours
Replacement Cost	Field Time and Materials

The Perfect Storm



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





Memory Backup

Target Application

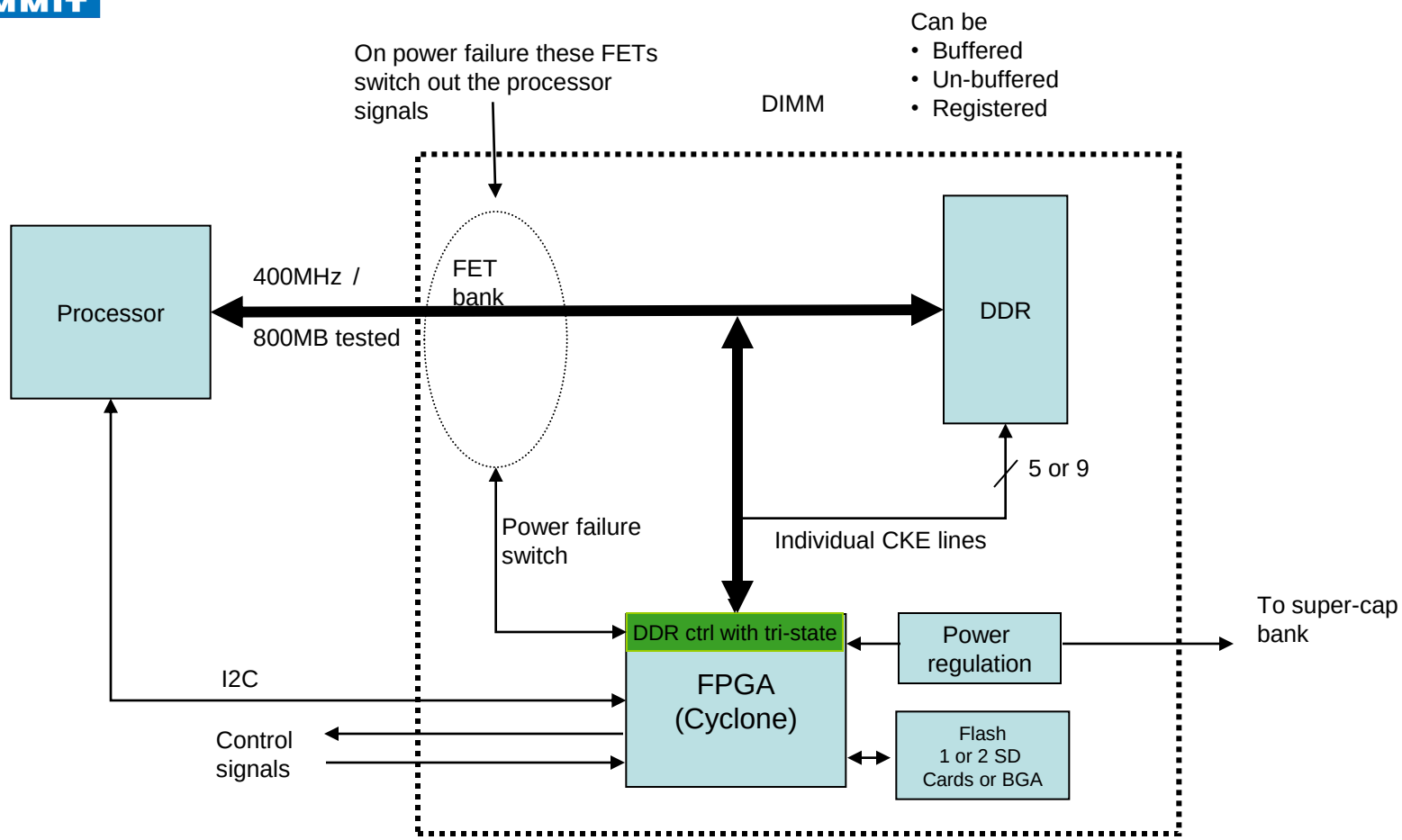
Nonvolatile DIMM for data recovery backup and restore

Benefits vs Batteries

Attribute	Ultracapacitors	Batteries
Environmental	Green	Hazardous disposal
Shelf Life	Years	Months
Charge Time	Seconds	Hours
Conditioning	None	Initial and periodic
Weight	Lighter	Heavier
Operating Temp	Up to 70°C	60°C max
Operating Life	Up to 10 years	1 to 3 years
Maintenance	None	Replace very 1-2 years

Component/Circuit	Function	Component Benefit
Cyclone V FPGA	Circuit management and control	Low power (28nm), DDR3 memory, and ONFI Flash control
High Speed Switching FETs	DDR3 Switch Signaling	Switching with noise suppression
ONFI NAND Flash	Memory Cache	2-8GB, matches DRAM density
DDR3 DRAM	Memory Source	2-8GB
Charging Circuit	Low Power, Green Energy Supply	Ultra Capacitor Bank with optional battery supply
Processor	Power Good Signal	Circuit toggle

NVDIMM Controller Architecture



MRAM

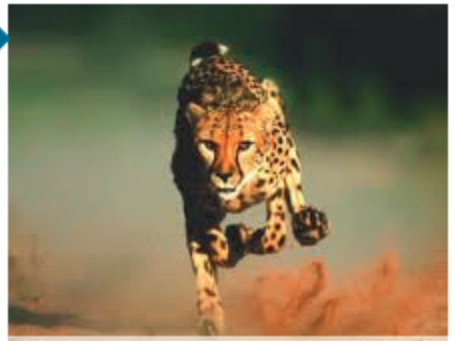
Nanosecond-class MRAM Storage

NAND SSD

500x Performance…

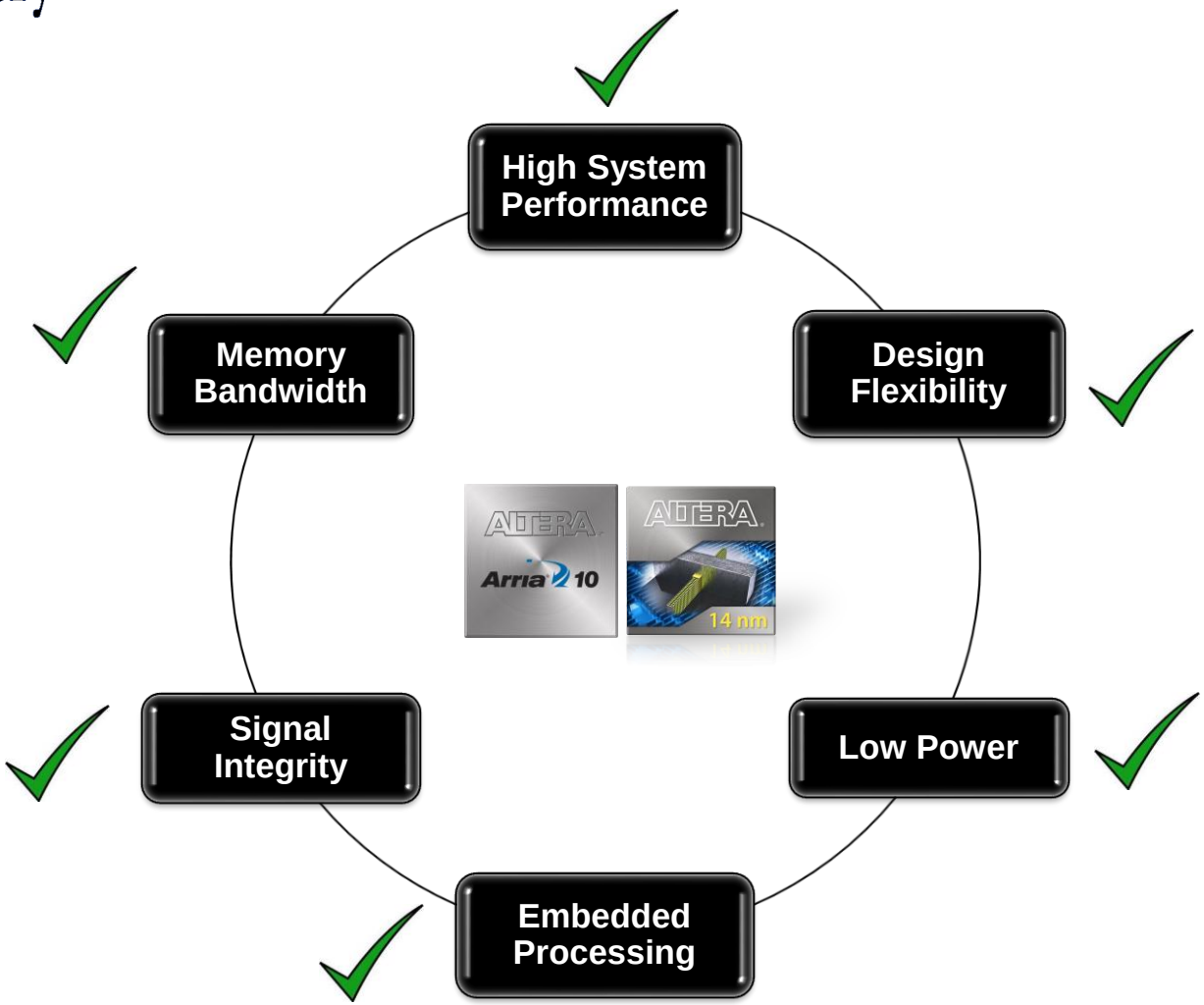
	NAND	MRAM
Density	64Gb	1Gb
Latency	50us	45ns
4k Write IOPS	800	400k
Cost/GB	1	50

…at only 50x Cost/GB



MRAM SSD

FPGA Benefits Summary



Flashing Foward

- FPGAs are a great technology option for Data Centers
  - Networking: Port aggregation
  - Compute: Application Acceleration
  - Storage: Controllers for non volatile memories
- All development phases supported
  - Prototyping
  - Production
  - Test Validation
  - Upgrades

Innovation Leader Across the Board

Flash Memory SUMMIT

ALTERA®

Exhibit Booth 320

MAX® Series	Cyclone® Series	Arria® Series	Stratix® Series	ENPIRION
CPLDs Lowest Cost, Lowest Power	*FPGAs Cost/Power Balance SoC & Transceivers*	*FPGAs Mid-range FPGAs SoC & Transceivers*	*FPGAs Optimized for High Bandwidth*	*PowerSoCs High-efficiency Power Management*

RESOURCES

Embedded Soft and Hard Processors

Nios® II

ARM®

Design Software

QUARTUS® II

DSP Builder

Development Kits

Intellectual Property (IP)

- Industrial
- Computing
- Enterprise