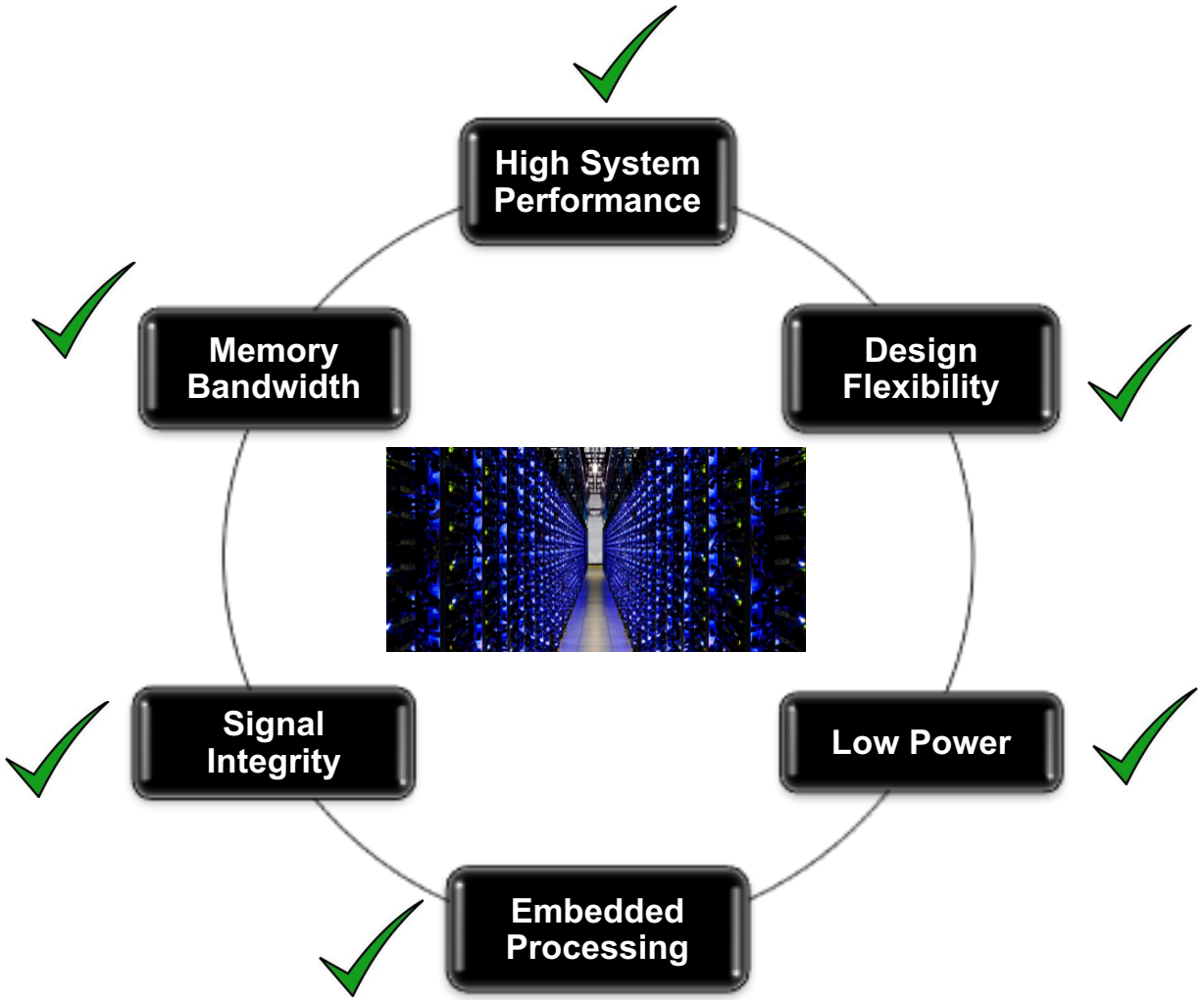
FPGAs in Flash Controller Applications

David McIntyre

DSMcIntyreConsulting@gmail.com

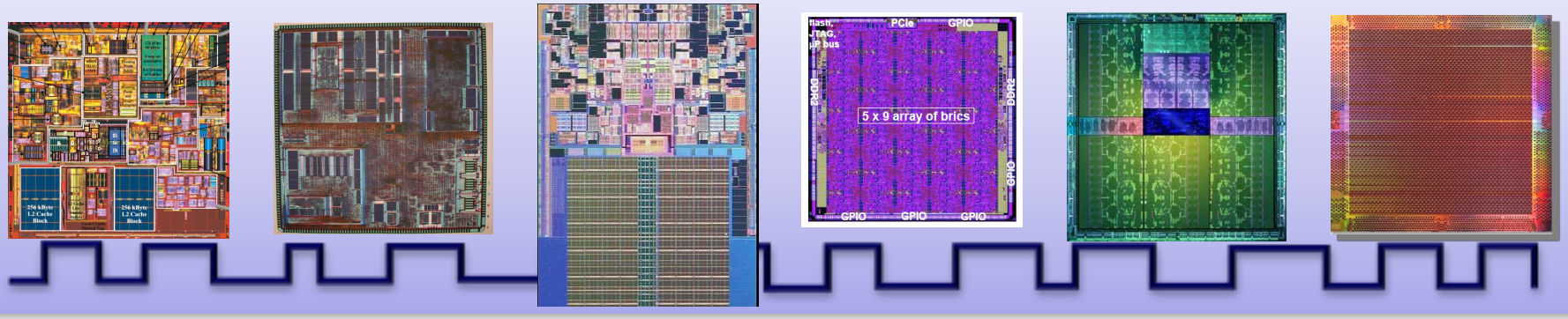
FPGA Then…

FPGA Now- Data Centers!



Processing Options

Technology scaling favors programmability and parallelism



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

Single Cores

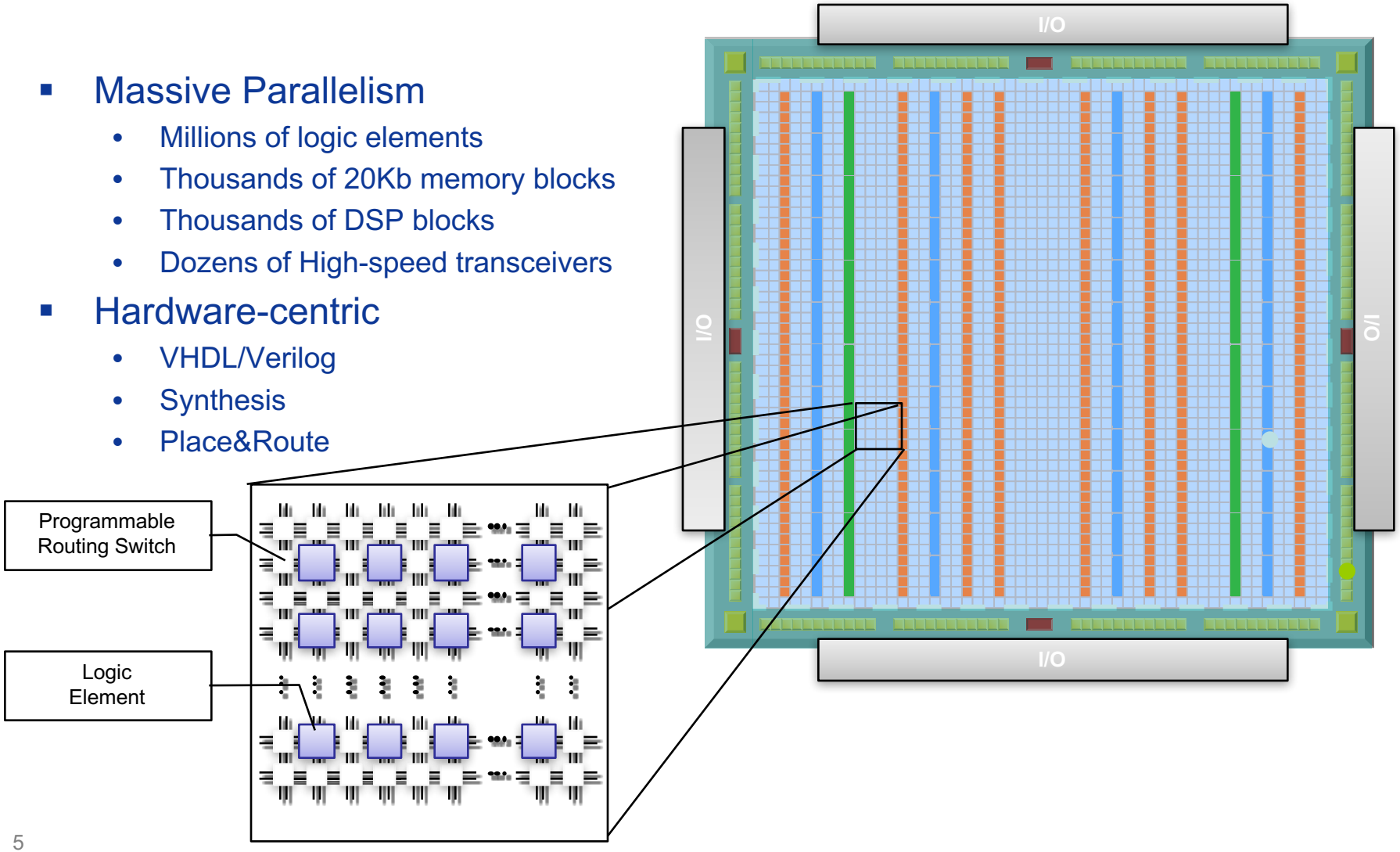
Multi-Cores
Coarse-Grained
CPUs and DSPs

Coarse-Grained
Massively
Parallel
Processor
Arrays

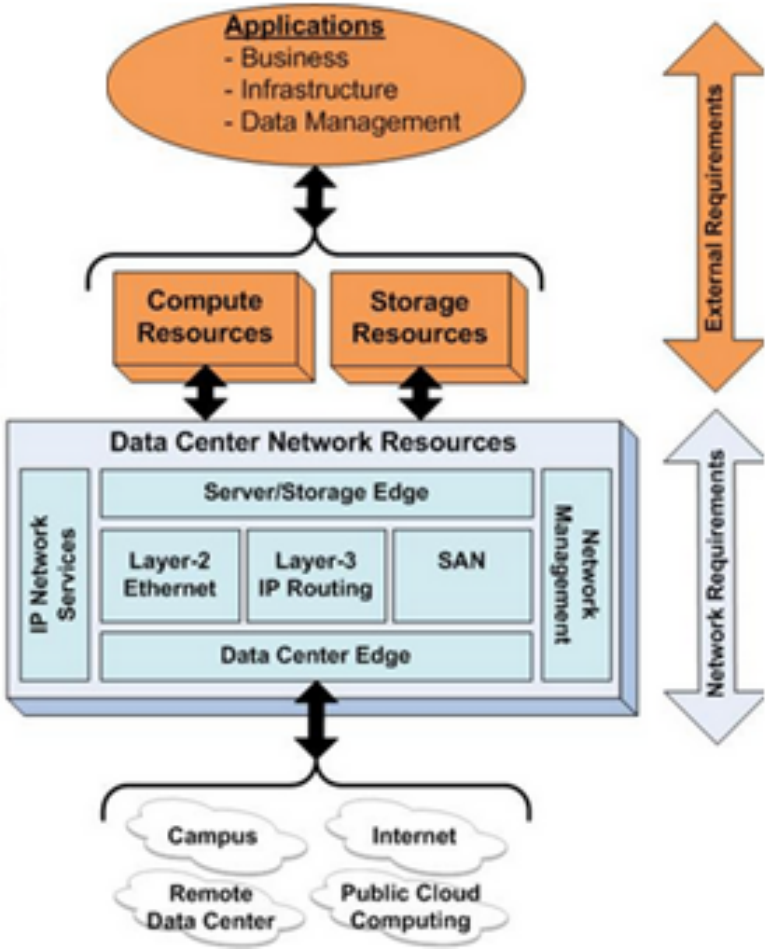
Fine-Grained
Massively
Parallel
Arrays

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



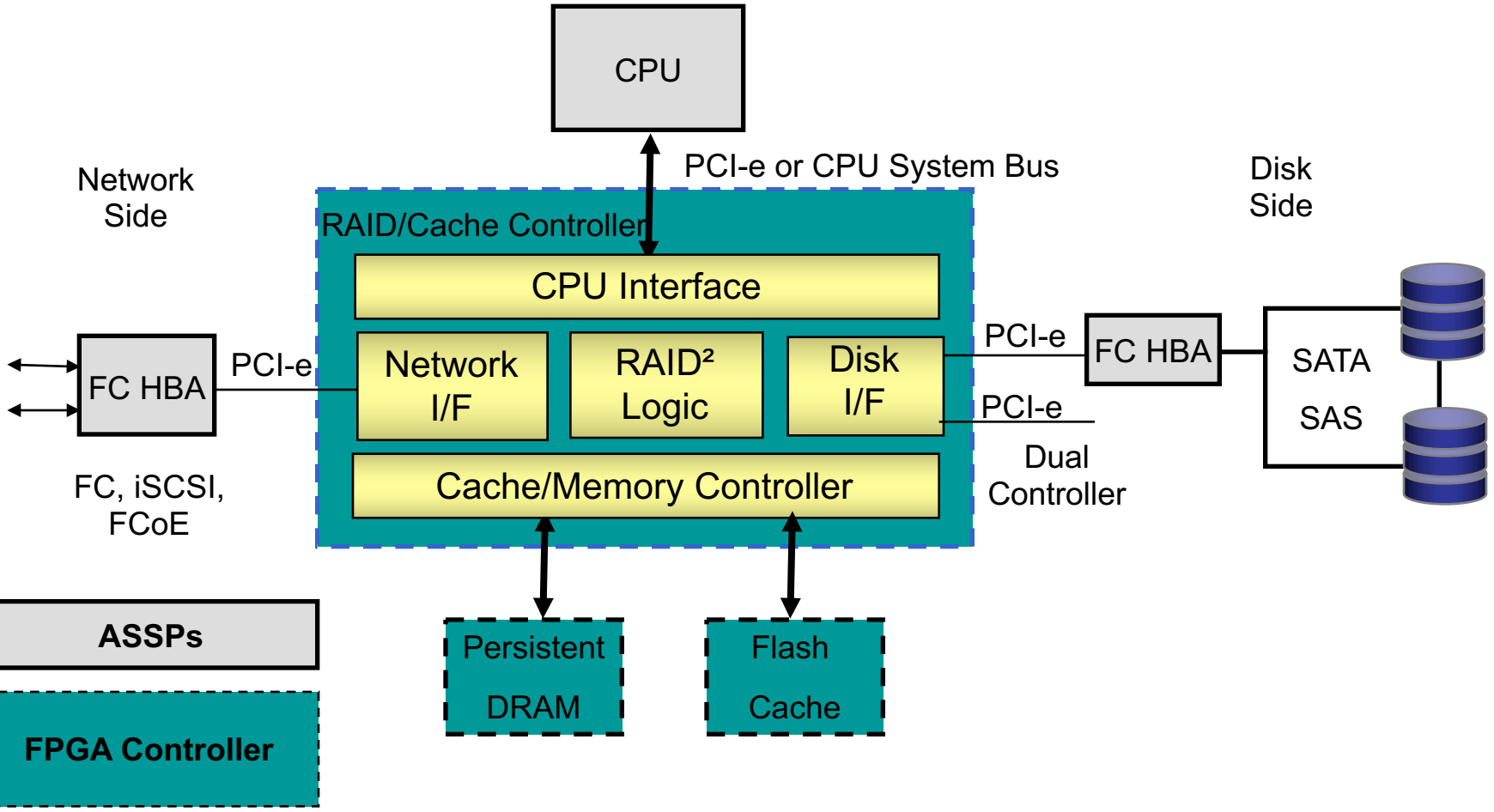
FPGA Utilization across Data Centers



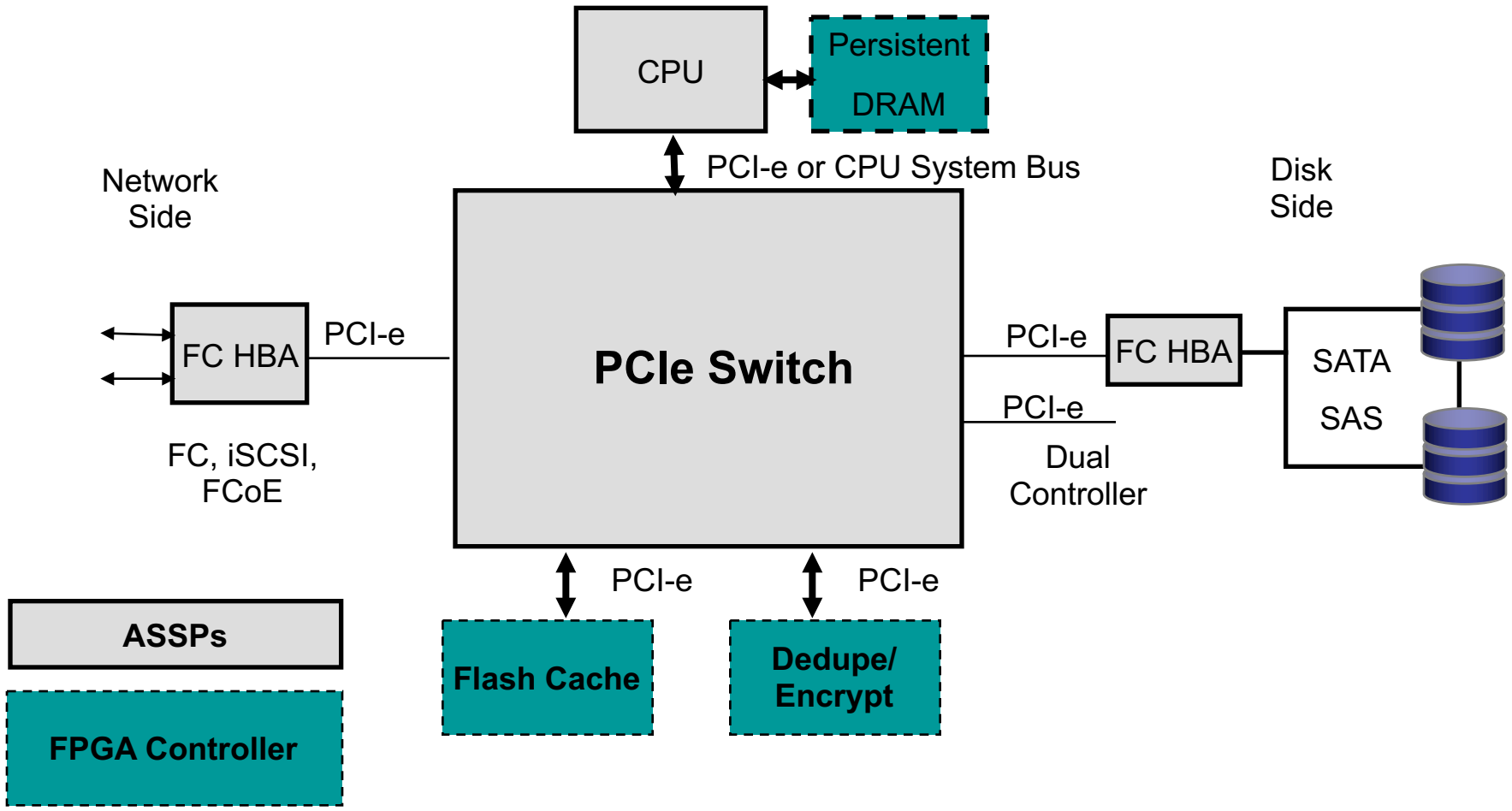
Point and SOC Solutions

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

Hybrid RAID System
- Persistent DRAM and Flash Caches

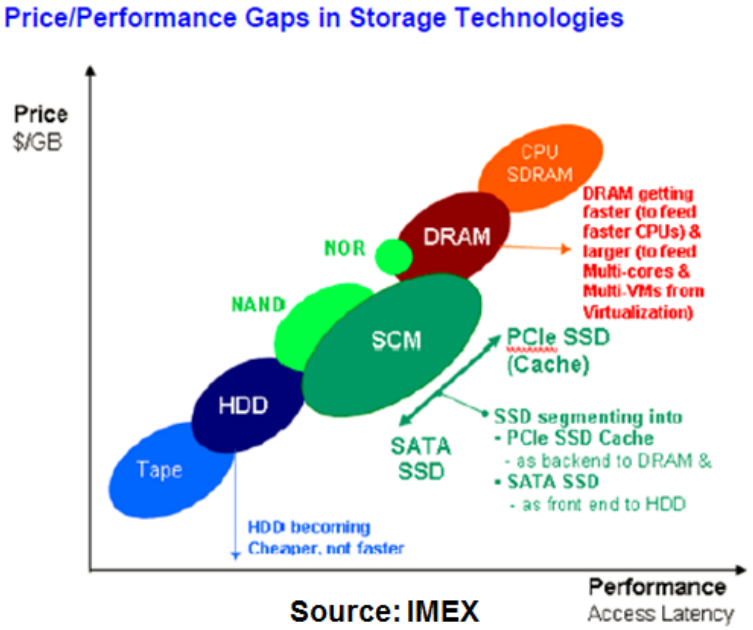


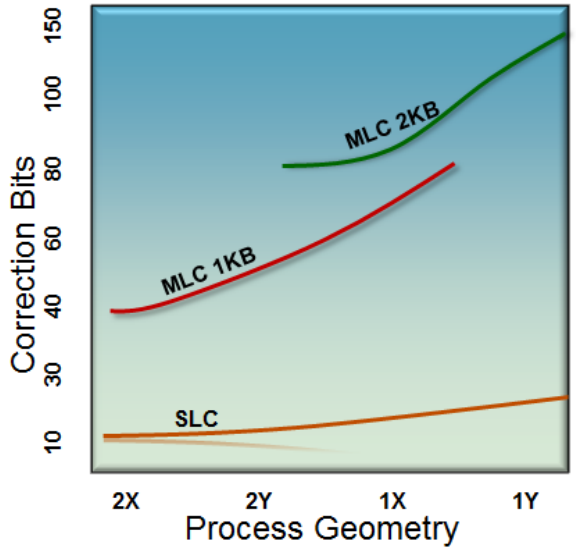
Hybrid RAID System
- PCIe Switch Centric



Flash Cache Challenges & Evolution

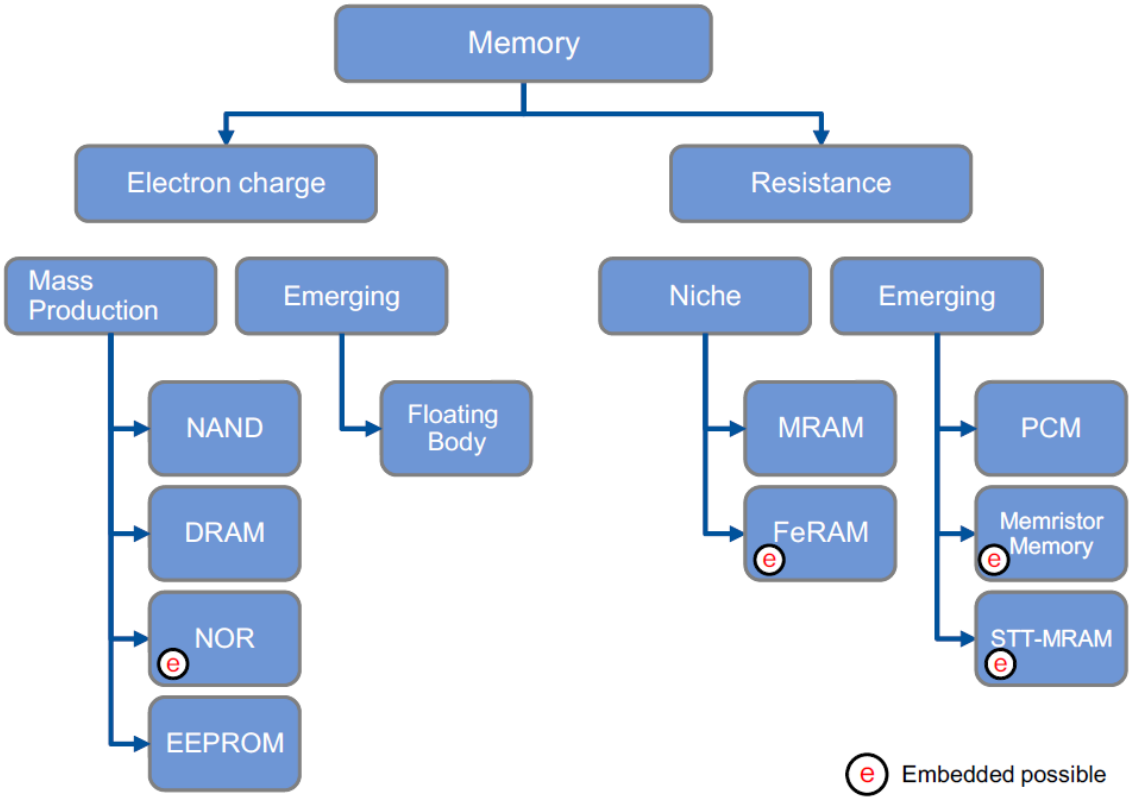
- Ongoing Challenges
  - Error correction costs increasing
  - Limited endurance (lifetime writes)
  - Slow write speed
  - SATA/SAS SSD interface is slow
- Storage over PCIe
  - Faster BW projections
  - SATA Express
  - NVM Express
  - SCSI Express
- NVMe over Fabrics
- Emerging flash technologies
  - MRAM (Magneto Resistive)
  - PCM (Phase Change)
  - RRAM (Resistive)
  - NRAM (Carbon Nanotube)…





Memory Categories

Figure 1. Categories of Memory (Charge Versus Resistivity)

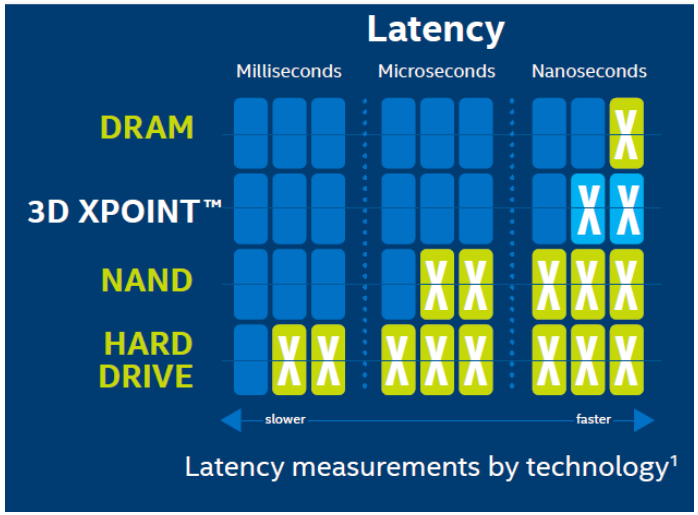


Key:
DRAM = dynamic RAM
EEPROM = electrically erasable programmable ROM
EPROM = erasable programmable ROM
FeRAM = ferroelectric RAM

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

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

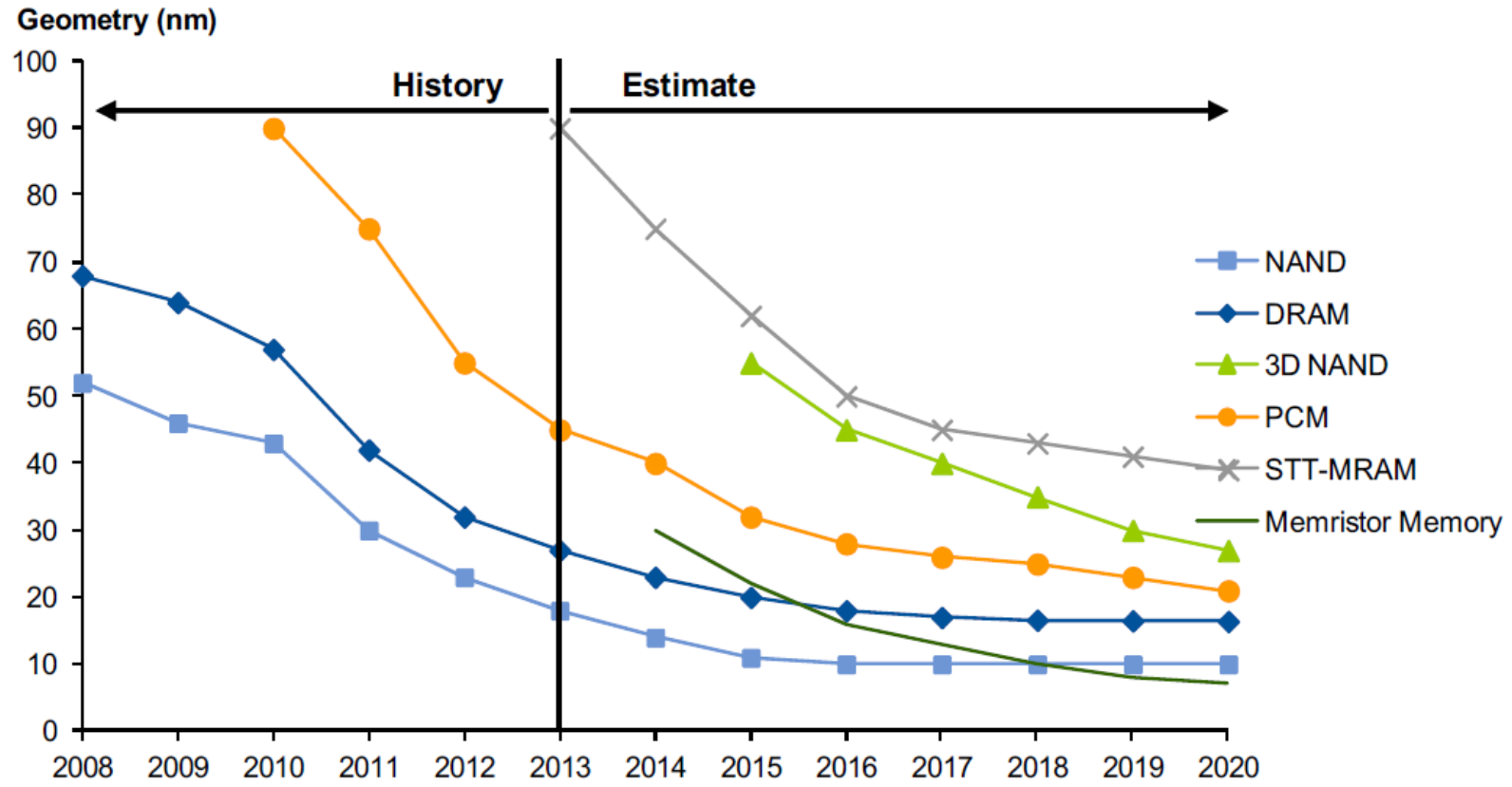


Source: Anantech.com

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

Migration Timeline- Cost

Figure 6. Migration Timeline for Emerging Memory Technologies



Source: Gartner

Flash Memory SUMMIT

5MB in Flight!

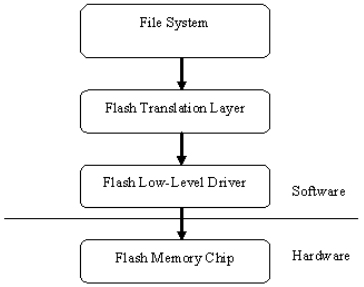
Flash Controller Design Considerations

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

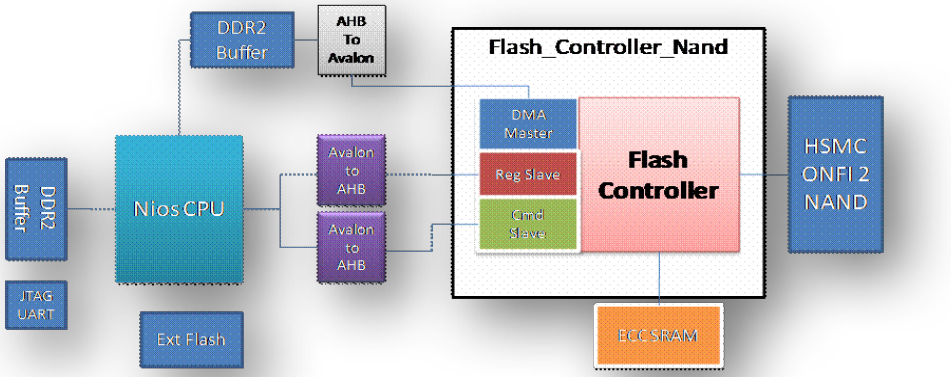


Flash Controller Support

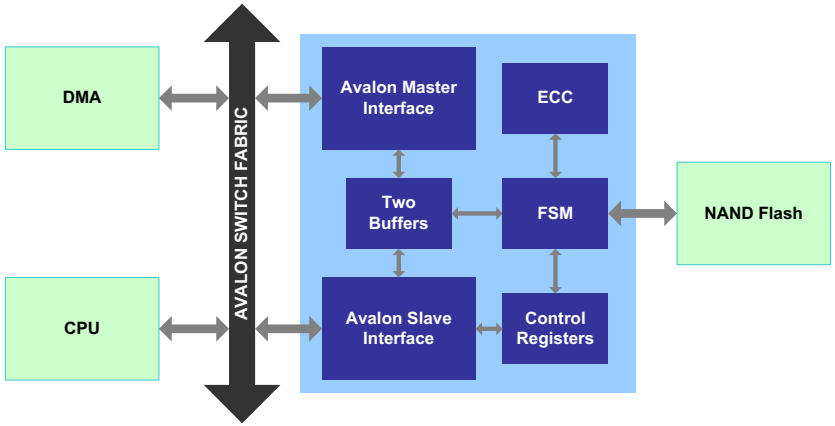
IP	IO	Speed	Logic Density	Comments
ONFI 3.x	40 pins/ch	400 MTps	5KLE/ch	NAND flash control, wear leveling, garbage collection
Toggle Mode 2.x	40 pins/ch	400 MTps	5KLE/ch	Same
DDR3	72 bit	1066 MHz	10KLE	Flash control modes available for NVDIMM
PCM			5KLE	PCM- Pending production $
MRAM			5KLE	MRAM- Persistent memory controller
BCH			<10KLE	Reference design
PCIe	G3x8	64Gbps	HIP	Flash Cache

Flash Cache Controller Examples

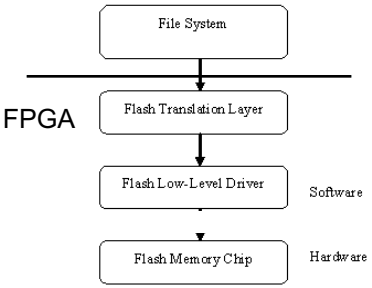
- **Multi Channel Controller**
  - Single to multi Flash channel capability
  - Basic NAND development platform
  - Provides High Speed ONFI & Toggle NAND PHY
  - ECC of 8 and 15 bits of error correction

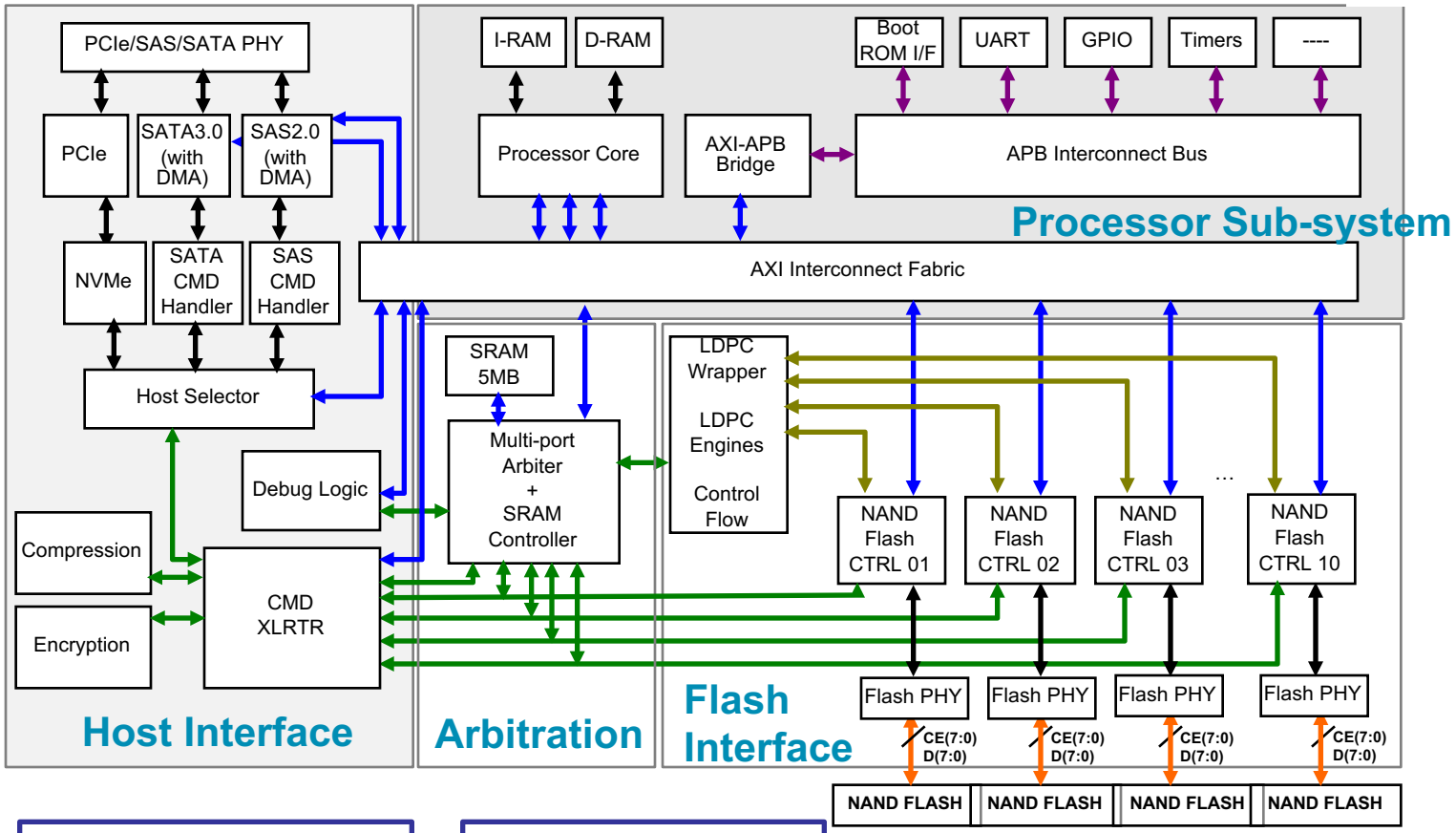


- **Single Channel Controller**



Typical SSD Controller Architecture





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

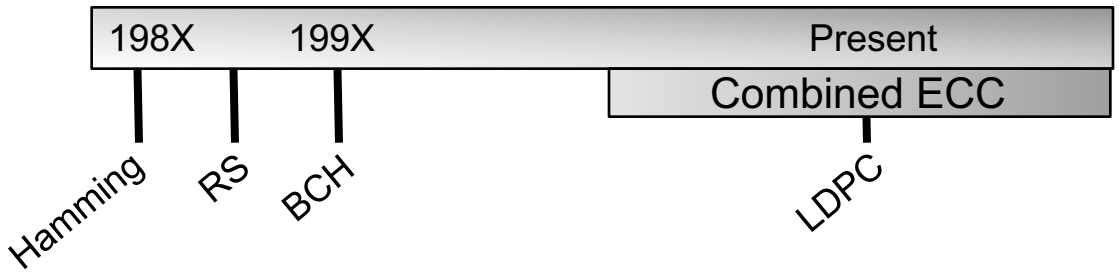
Error Correction Overview

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

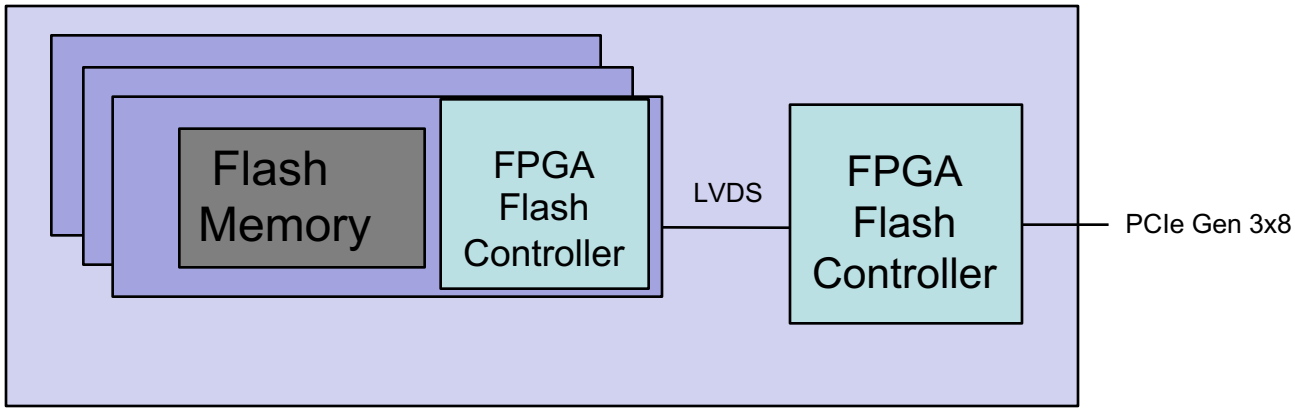
LDPC and Programmable Logic

- 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 Storage Arrays

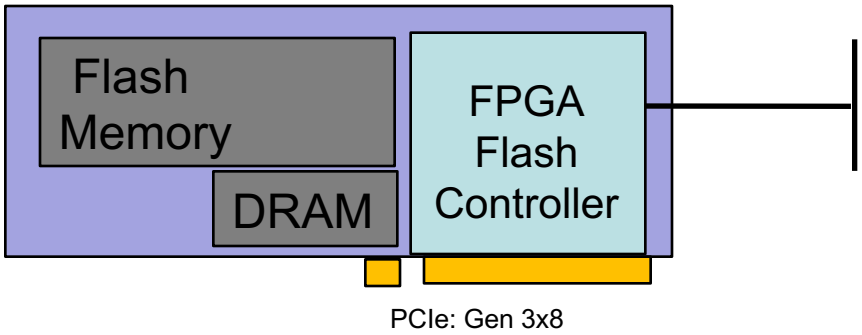
Target Application: Enterprise Tier-1 Storage: Databases and Virtualization



Function	Solution Rqts	IP Rqts
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.

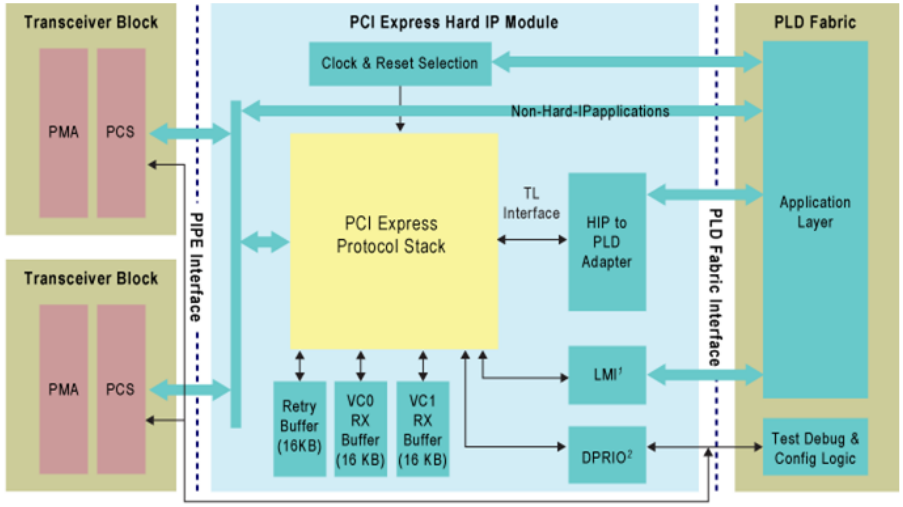
Function	Solution Rqts	IP Rqts
Flash Control	-ONFI 2.X/3.0 -Toggle Mode 2.0 - Multi flash load/ch -40 GPIO/ch -PCIe Gen 3 x8 -Low power & cooling	- Flash Controller (bad block mgt and wear leveling) -Flash RAID -Cache controller - BCH core -PCIe config < 100msec -Host interface/APIs

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

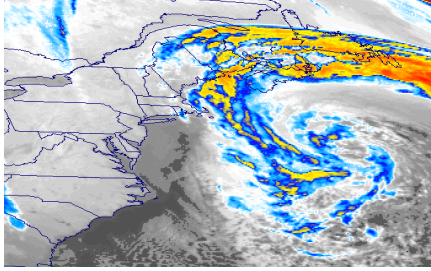
- Scalable host controller interface for PCIe-based solid state drives
- Optimized command issue and completion path
- Benefits
- Software driver standardization
- Direct access to flash
- Higher IOPS and MB/s
- Lower latency
- Reduced Power Consumption

DRAM Cache Backup

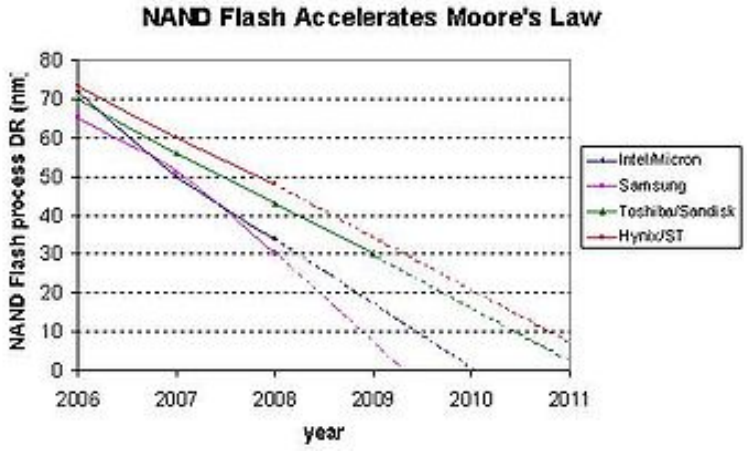
- Data Center server power outages continue
- Read/Write Consequences
  - Data Loss
  - Undetected errors in host application
- NVDIMM designs protect system integrity but…

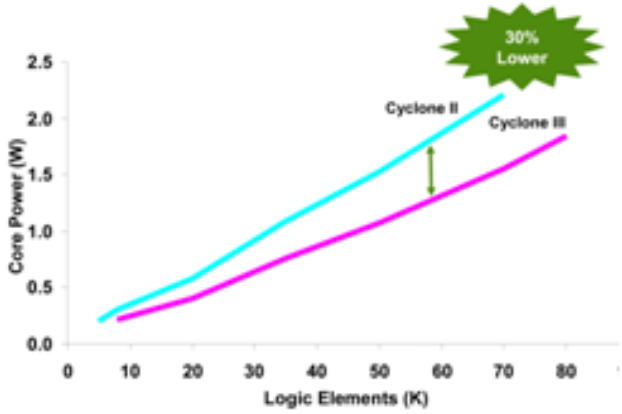
Battery Limitations	Issue
Shelf Life	One year max or 500 cycles
Disposal and Handling	Hazardous Waste Management
Data Storage Capacity	Up to 72 hours
Down Time	Charge Time up to 6 hours
Replacement Cost	Field Time and Materials

The Perfect Storm

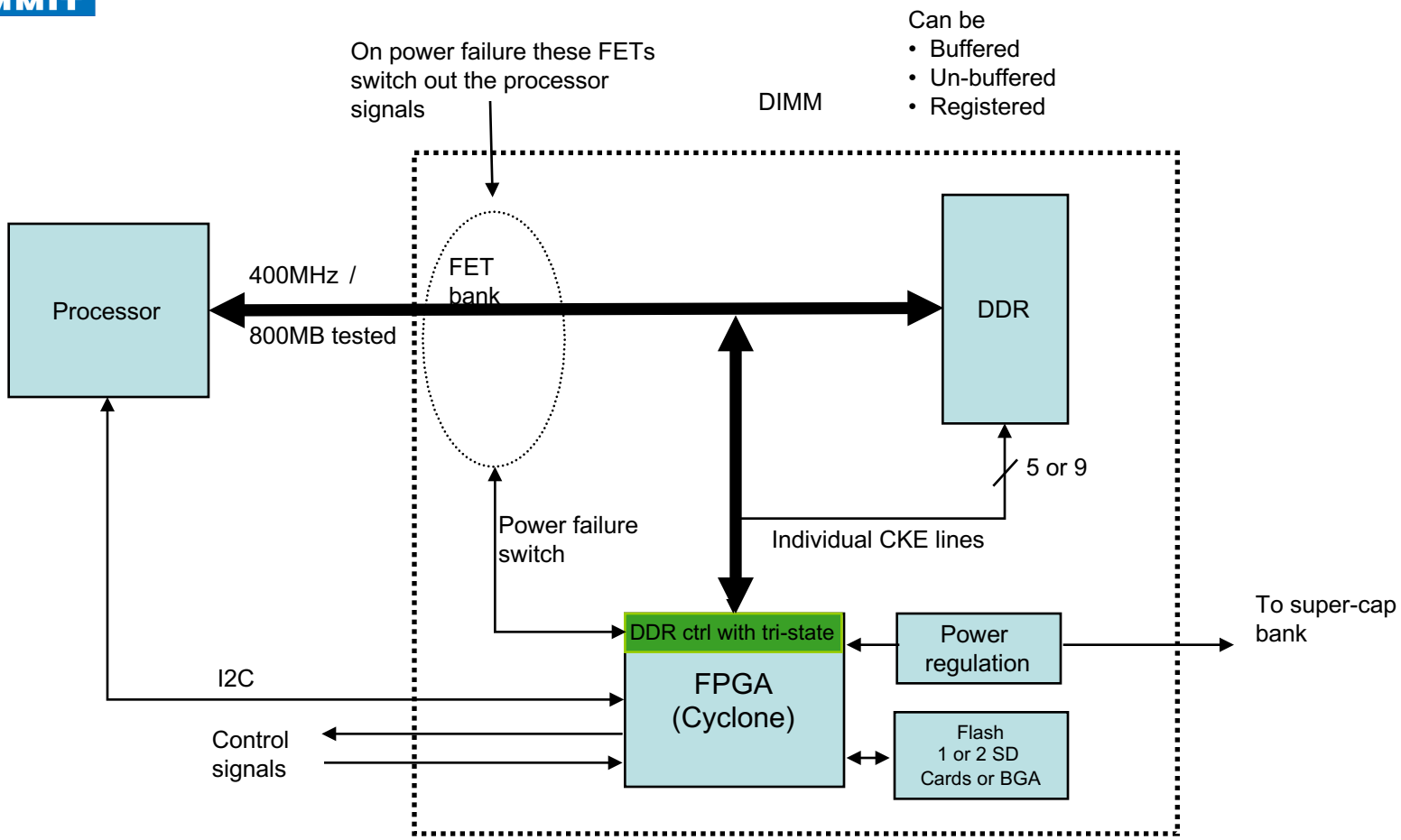


- **Technology Enablers**
  - **Super Capacitors** are production worthy
  - **Flash memory** costs continue to decline
  - **FPGA** technology meeting power/performance/cost





NVDIMM Controller Architecture



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