FPGAs in Flash Controller Applications

David McIntyre

DSMcIntyreConsulting@gmail.com

Flash Memory Summit

Data Center Trends

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

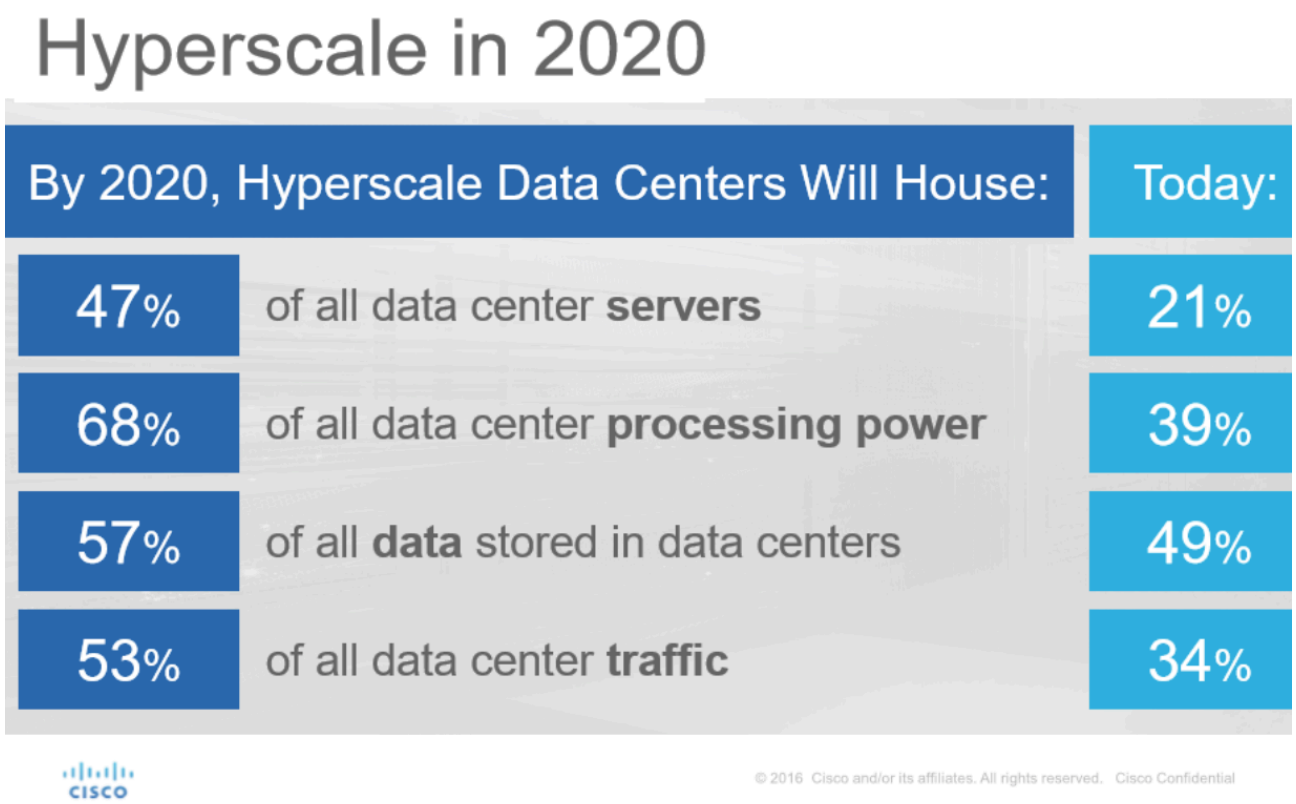
- **Hyper Scale**
  - Independent scaling of compute and storage resources
  - Good for elastic workloads, e.g. Hadoop, Cassandra, noSQL
  - Also software managed

Flash Memory Summit

Data Center Trends

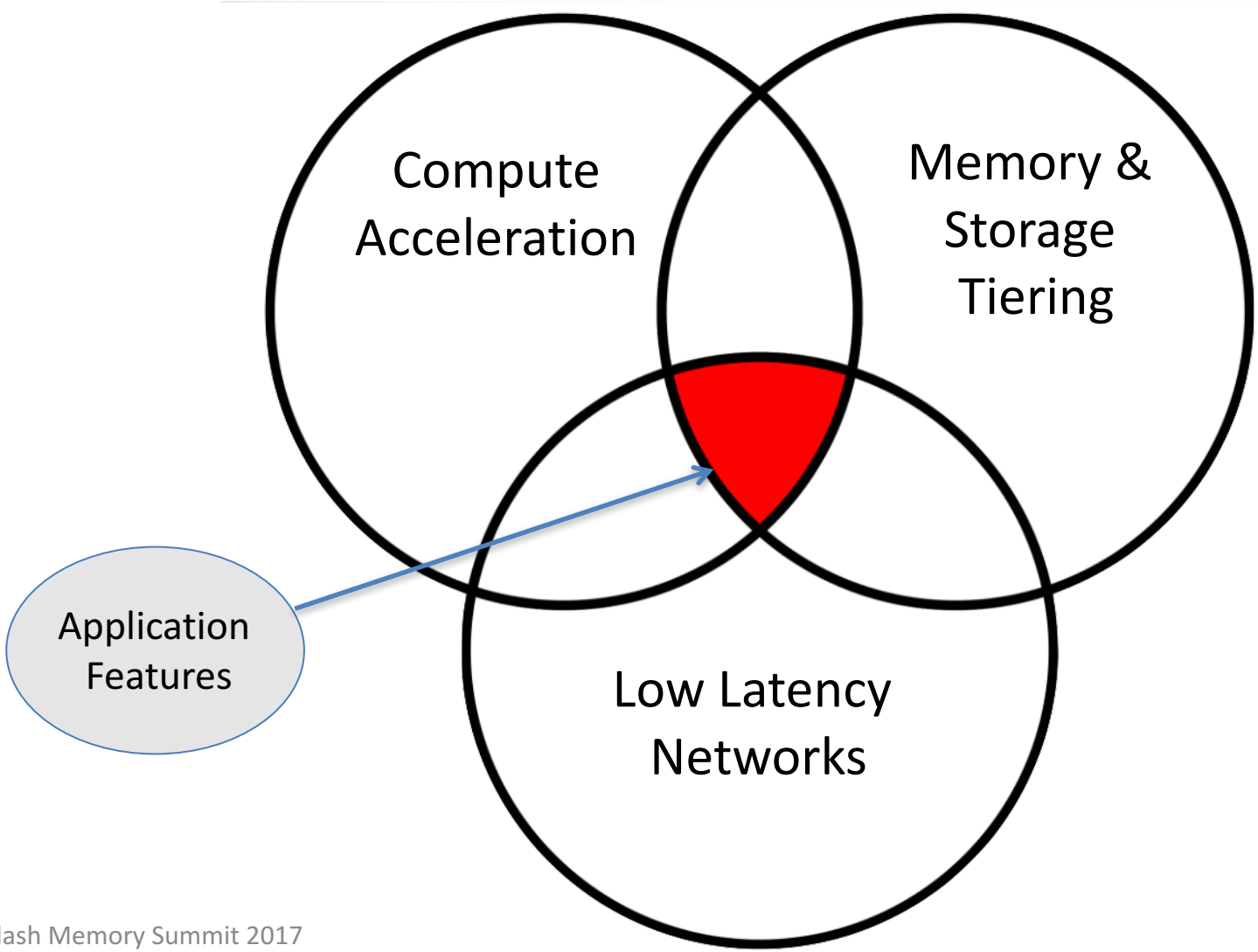
- **Storage**
  - Convergence of RAM/cache and SCM
  - All flash and hybrid arrays
  - Persistent memory cache
- **Compute**
  - GPU, TPU and FPGA accelerators
- **Networking**
  - Low latency, high performance RDMA networks
- **Hybrid Cloud**
  - For lease and on premises-equipment
  - Deployment Options, e.g. OpenStack and Docker

Hyperscaler Priority

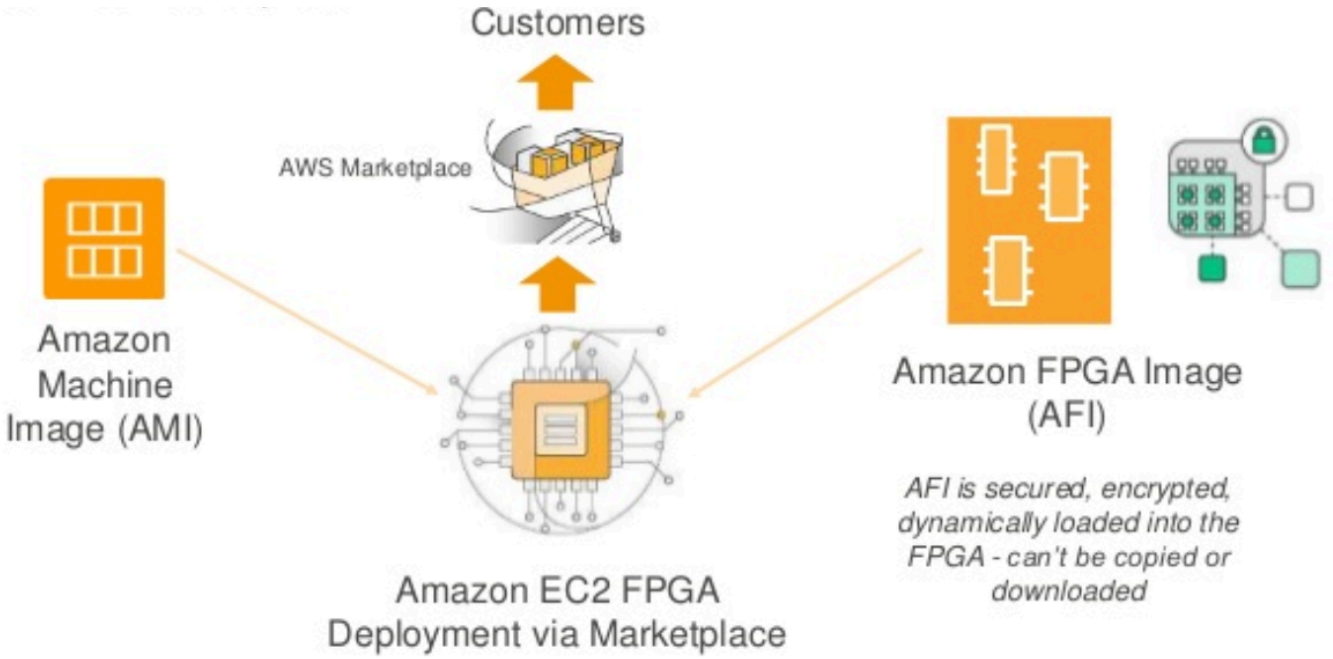


- Flash controllers must support hyperscale requirements (latency, performance/watt, endurance, reliability)

Supporting Infrastructure

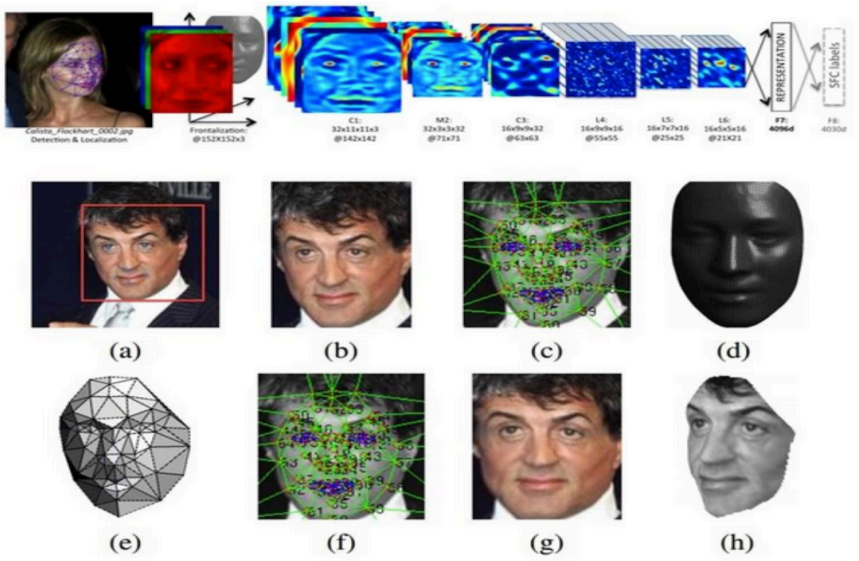


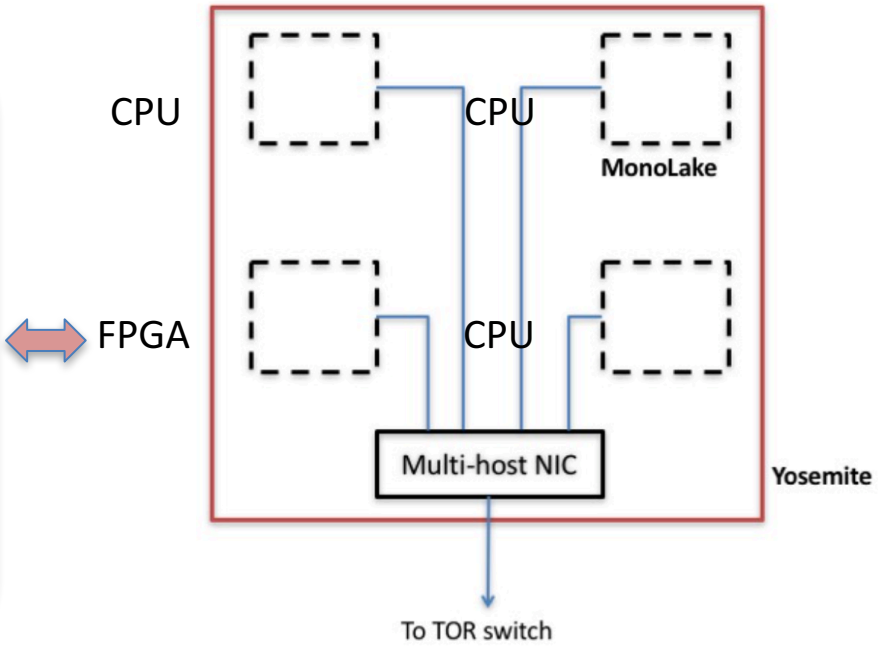
Amazon



- AWS cloud service upgraded with FPGA leasing option to support HPC vertical markets (Big Data, Video analytics, Financial, Genomics, Security)
- Combines machine images with FPGA images to accelerate specific routines call out by the customer
- Development tools and resources with be critical for success

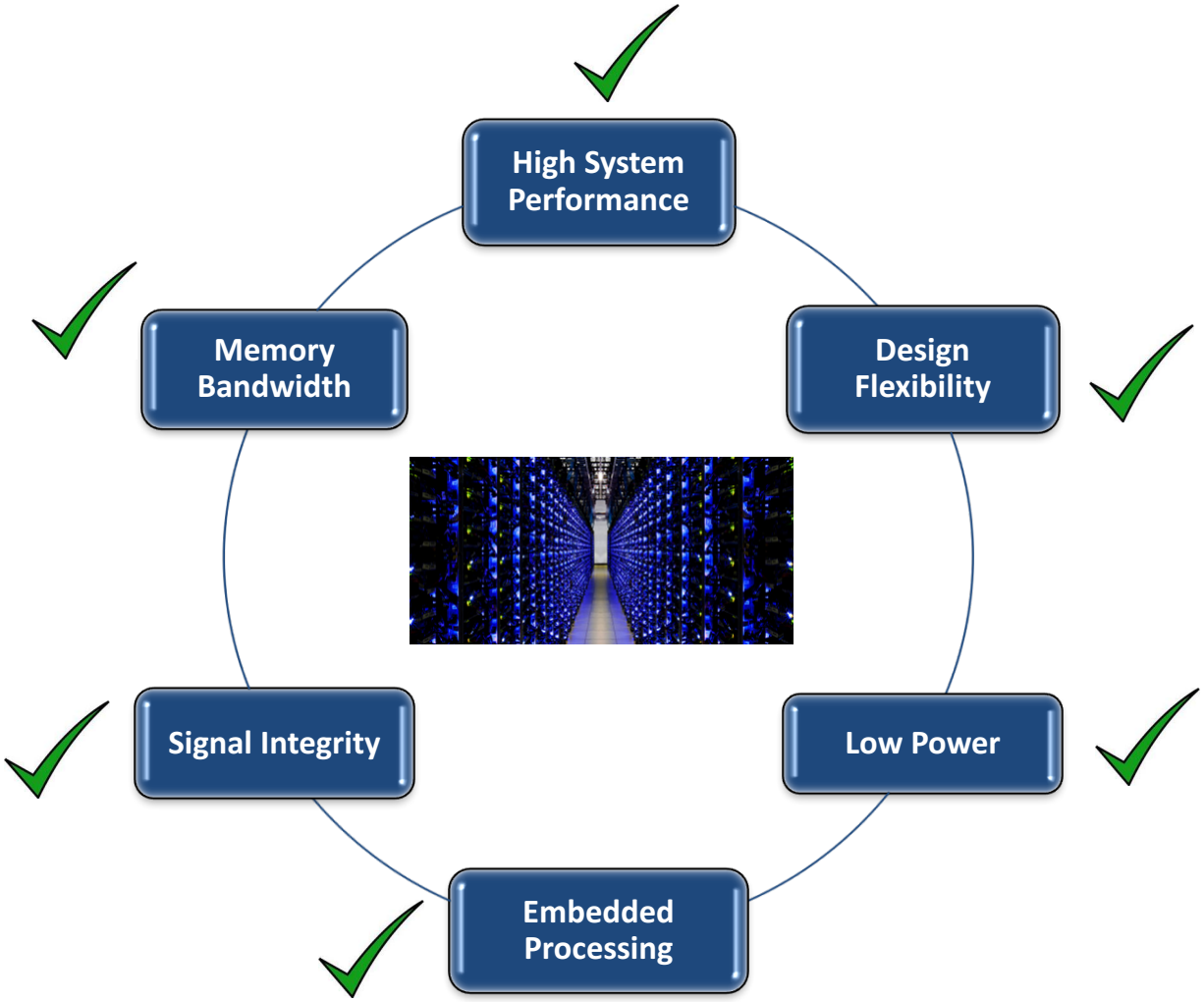
Facebook





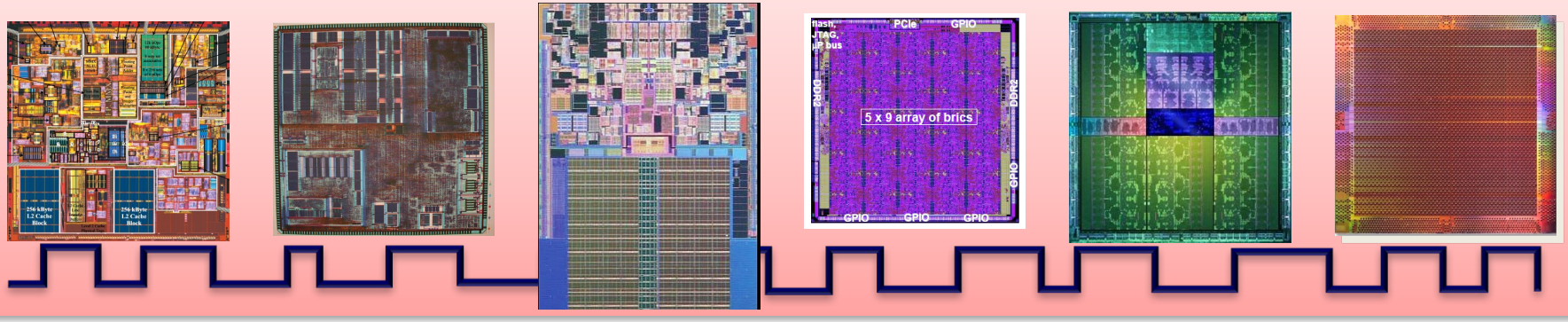
- Particular focus on video analytics to search for human attributes (facial characteristics, unique signatures)
- Nearly one billion searches in <5 sec
- Close to human recognition performance on frontal views
- Yosemite sleds support single socket Xeon-D and FPGA/GPU
- Optimized performance/Watt rack density

FPGA Benefits for Data Centers



Processing Options

Technology scaling favors programmability and parallelism



CPUs — DSPs — Multi-Core — Gate Array — GPGPUs — FPGAs

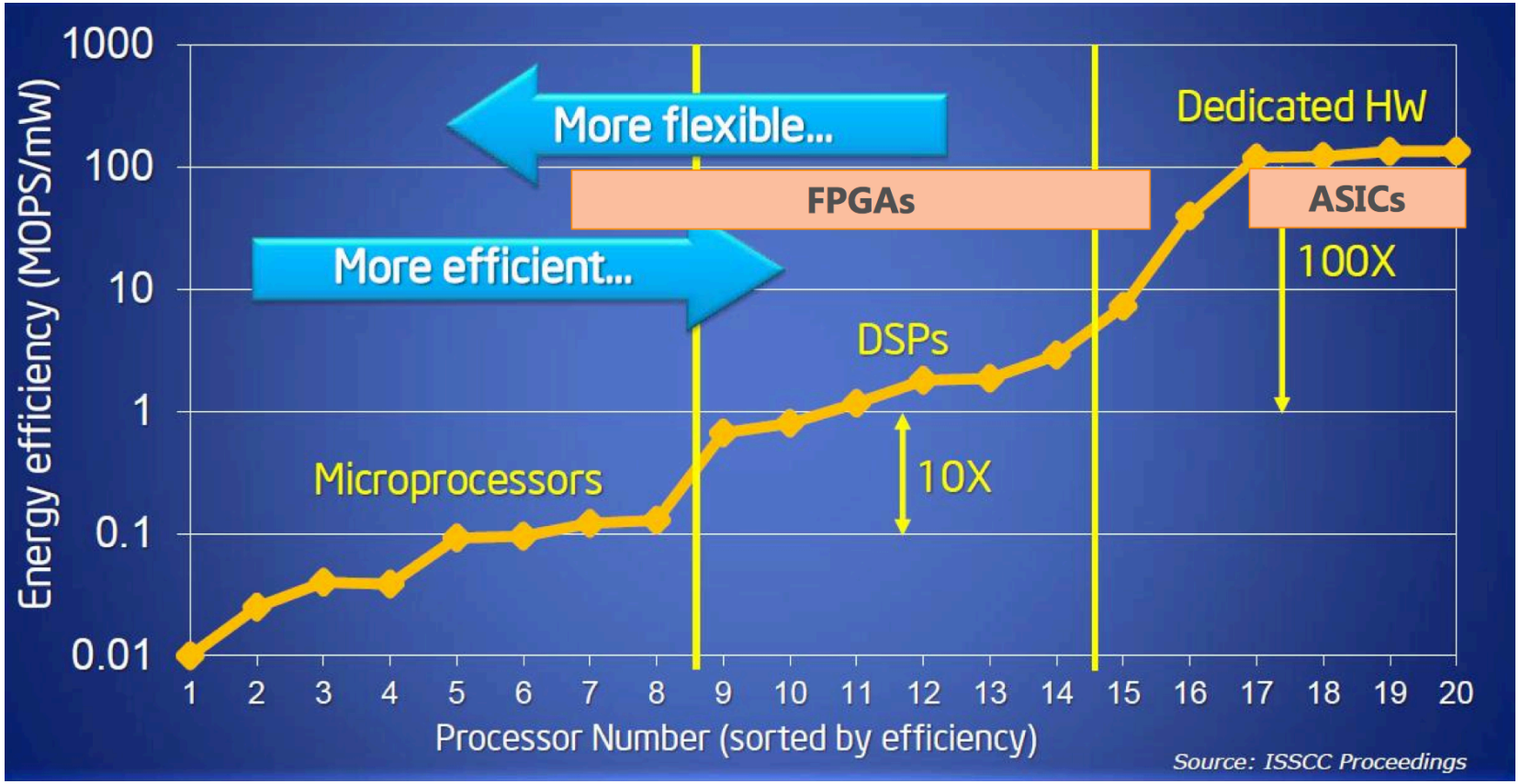
Single Cores

Multi-Cores Coarse-Grained CPUs and DSPs

Coarse-Grained Massively Parallel Processor Arrays

Fine-Grained Massively Parallel Arrays

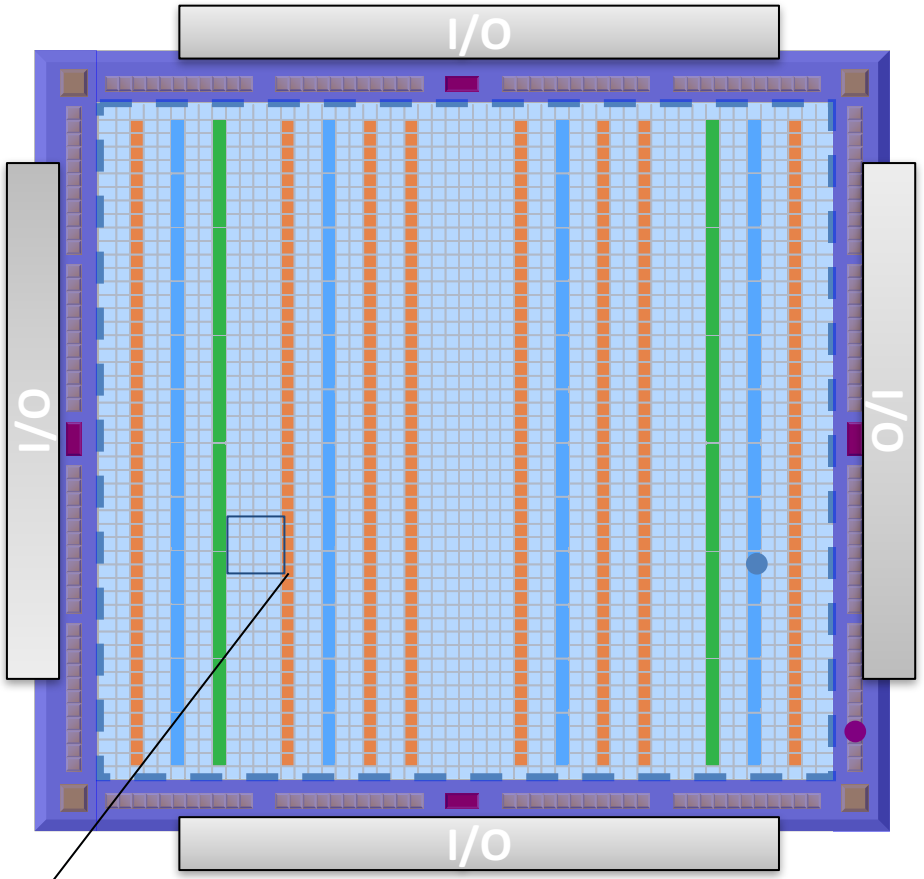
Technology Tradeoffs

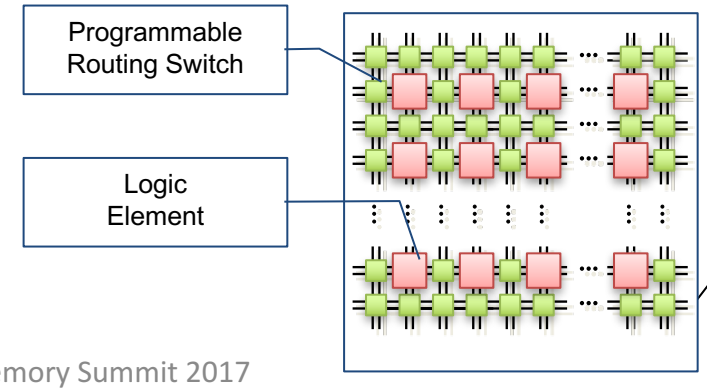


- Data center metric is performance/watt
- The proper balance of performance and efficiency while maintaining flexibility to support data center applications

FPGA Architecture

- Massive Parallelism
  - Millions of logic elements
  - Thousands of 20Kb memory blocks
  - Thousands of DSP blocks
  - Dozens of High-speed transceivers
  - SOC processing
  - Optics
- Hardware-centric
  - VHDL/Verilog
  - Synthesis
  - Place&Route



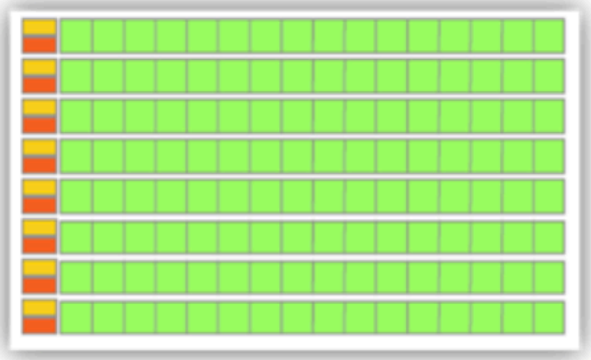


GPU Offload Advantage

CPU

* Low compute density
* Complex control logic
* Large caches (L1$/L2$, etc.)
* Optimized for serial operations
  - Fewer execution units (ALUs)
  - Higher clock speeds
* Shallow pipelines (<30 stages)
* Low Latency Tolerance
* Newer CPUs have more parallelism

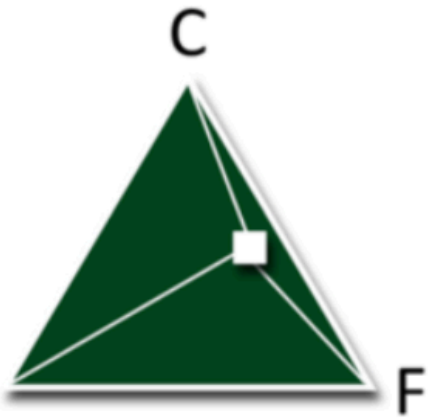
GPU



* High compute density
* High Computations per Memory Access
* Built for parallel operations
  - Many parallel execution units (ALUs)
  - Graphics is the best known case of parallelism
* Deep pipelines (hundreds of stages)
* High Throughput
* High Latency Tolerance
* Newer GPUs:
  - Better flow control logic (becoming more CPU-like)
  - Scatter/Gather Memory Access
  - Don't have one-way pipelines anymore

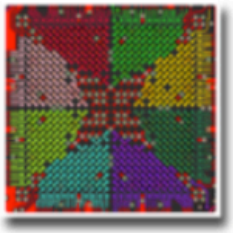
FPGA Comparison Continued

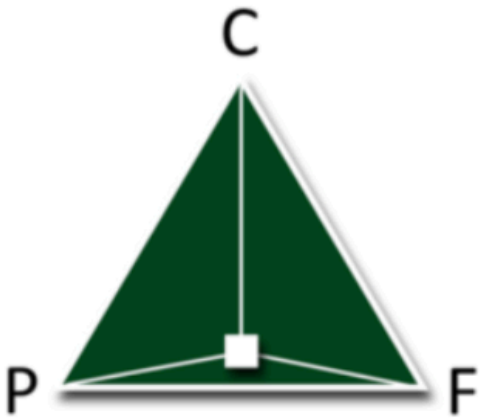
Traditional CPU



COST ~
PERFORMANCE
FLEXIBILITY ✓

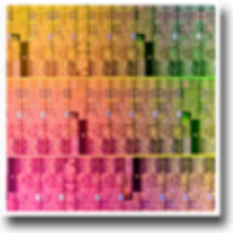
FPGA

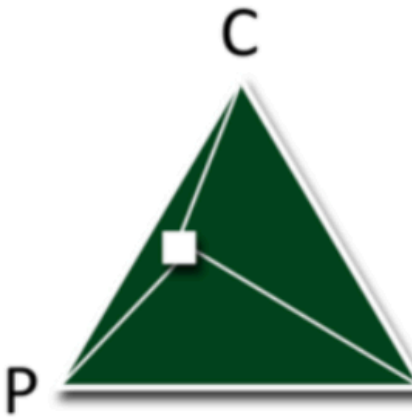




COST
PERFORMANCE ✓
FLEXIBILITY ✓

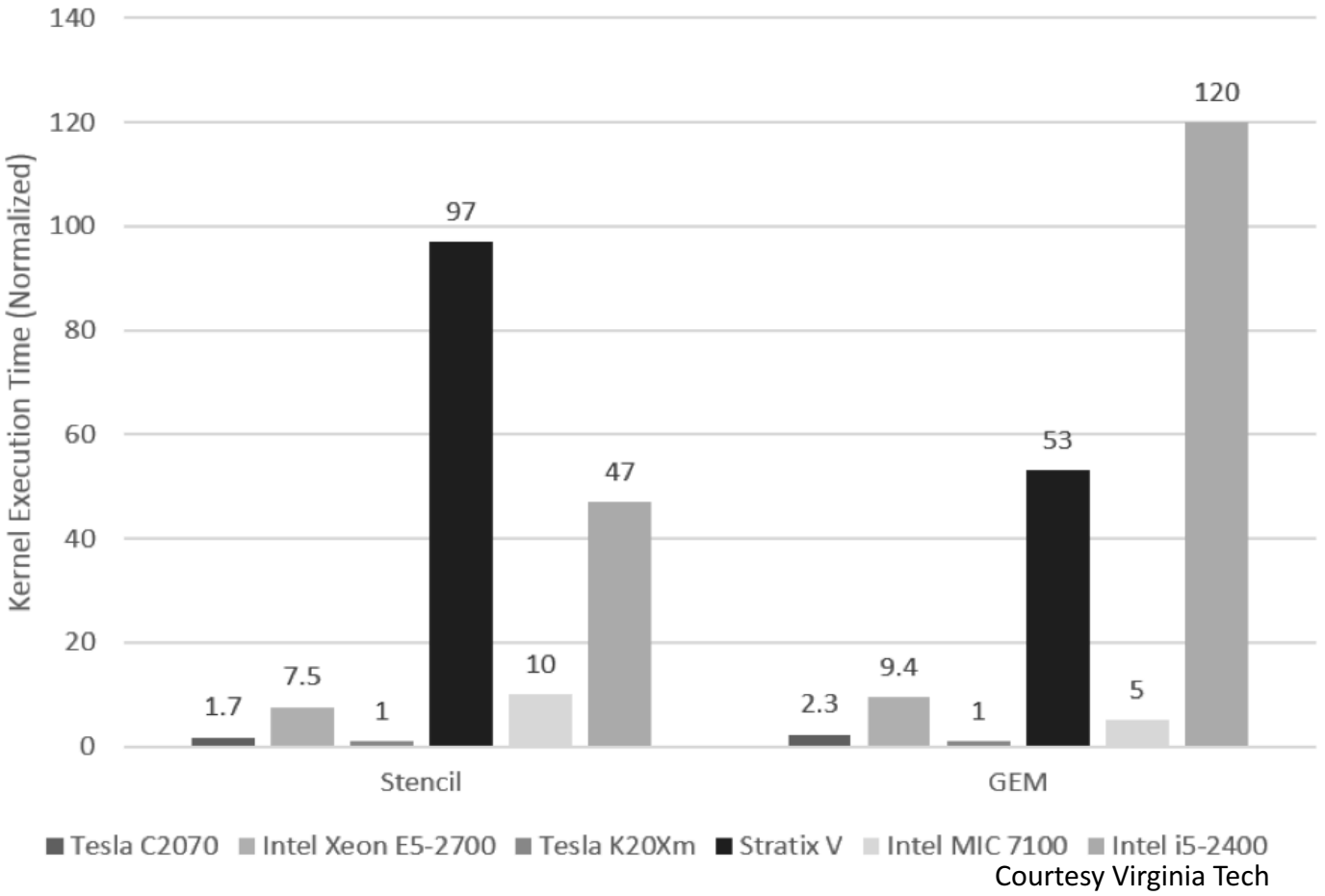
Network ASIC





COST ✓
PERFORMANCE
FLEXIBILITY

Inference Workloads



Courtesy Virginia Tech

- **Stencil:** Memory bound, simple and synchronized compute blocks
- **GEM:** Matrix multiplication

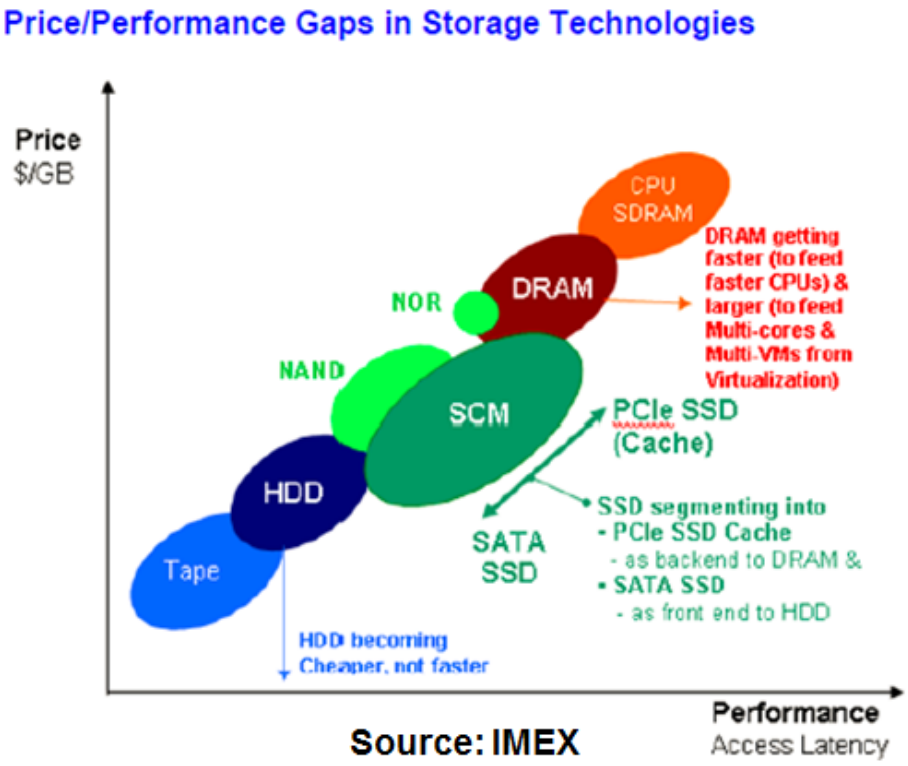
Flash Memory Summit

Technology Comparison

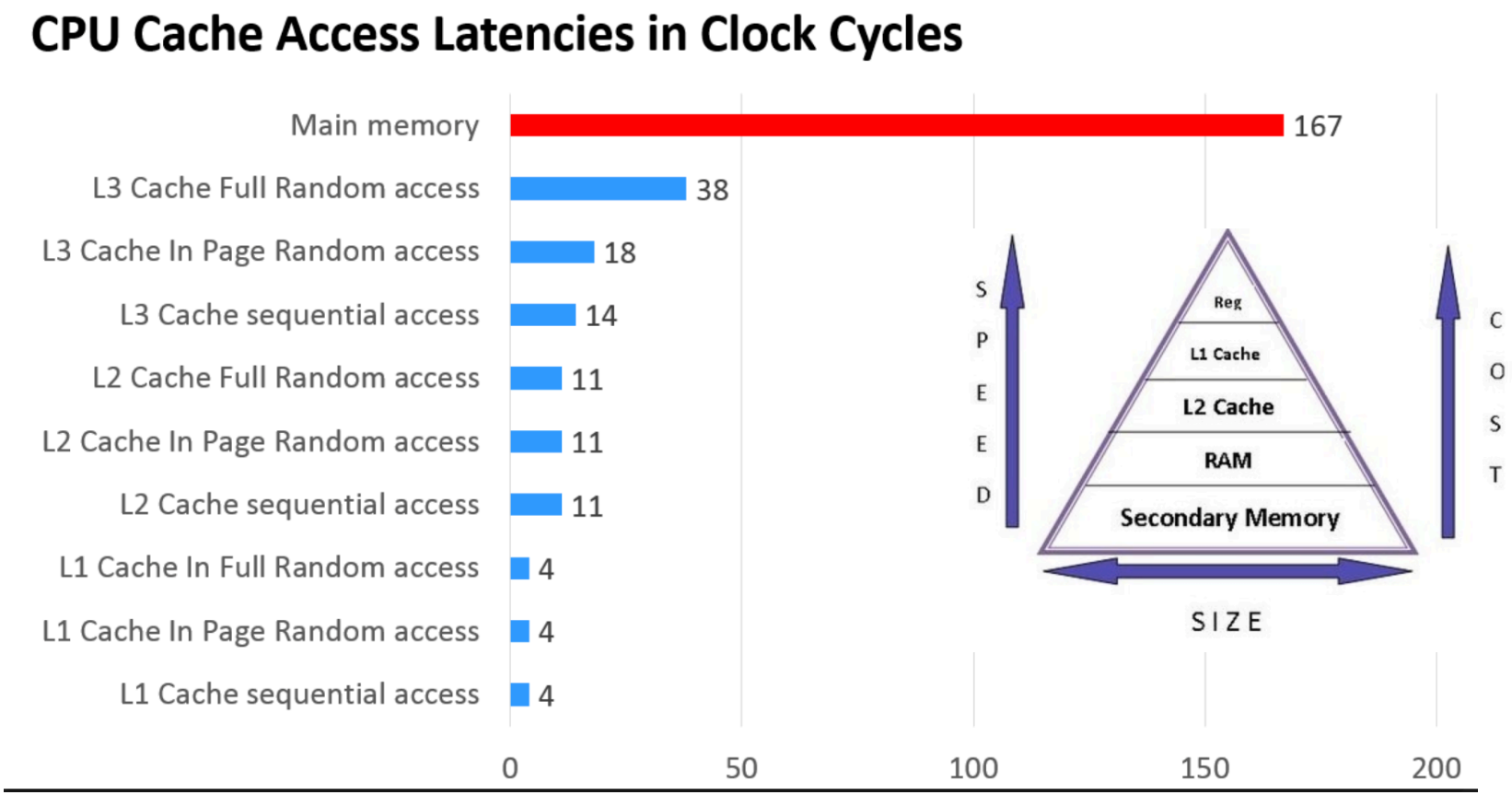
Technology	Pros	Cons
CPU	Well established products	• Limited cores for parallel processing • Power consumption
FPGA	Heterogeneous parallel processing Performance/Watt Flexibility	• Rudimentary development environment • Inefficient per unit costing
GPU	Same task parallel processing Developer ecosystem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

Flash Controller Challenges

- Error Correction
  - LDPC, BCH
  - Error correction costs increasing
  - Limited endurance (lifetime writes)
  - Hybrid correction schemes
- Storage over PCIe
  - NVMe and NVMeF
  - PCIe Gen 4
- Cache technologies
  - MRAM (Magneto Resistive)
  - PCM (Phase Change)
  - ReRAM (Resistive)
  - NRAM (Carbon Nanotube)

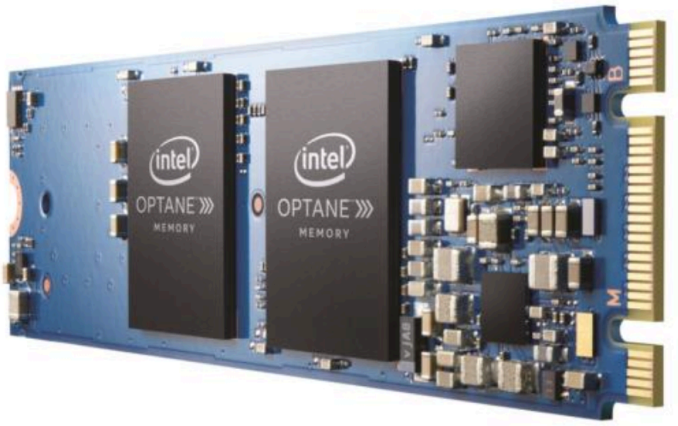


Memory Hierarchy



Persistent Friend or Foe

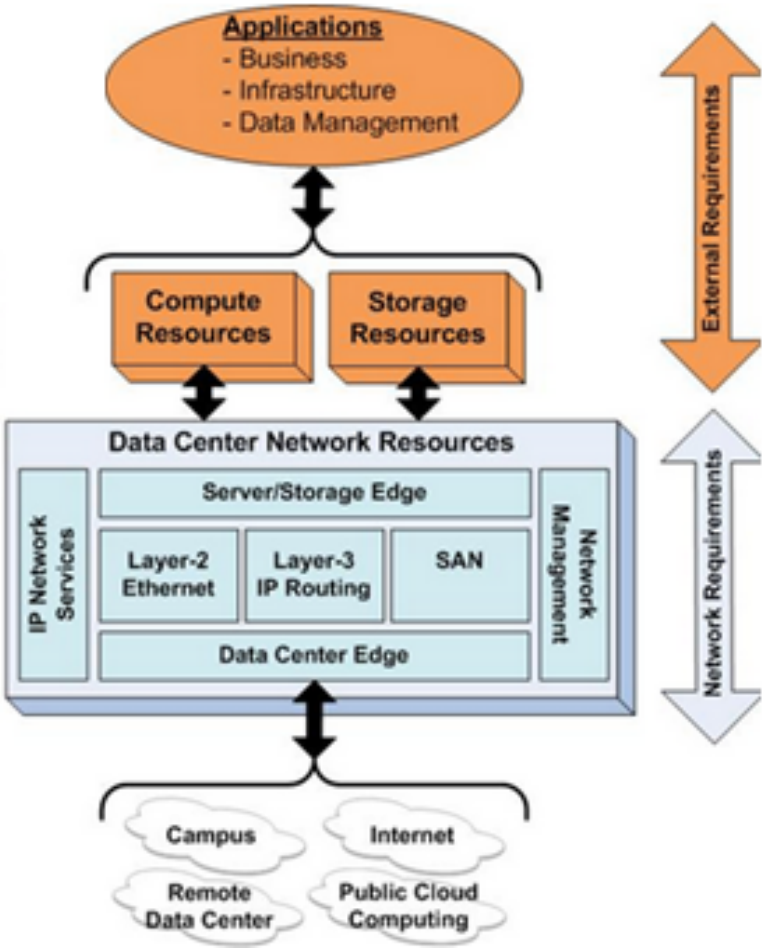
- Intel/Micron Xpoint Claimed Attributes vs. NAND
  - Performance (10X)
  - Endurance (1000X)
  - Latency (1/1000X)
  - Byte addressable



- Cost (5X)

- Opportunity for NAND to support load/store-driven data center applications (NVDIMM-F and NVDIMM-P)

FPGA Utilization across Data Centers



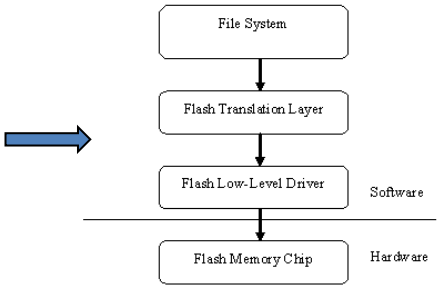
Point and SOC Solutions

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

Flash Memory Summit

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**
  - NVMe, NVMeF, PCIe Gen 3/Gen 4, optics
  - Open Channel



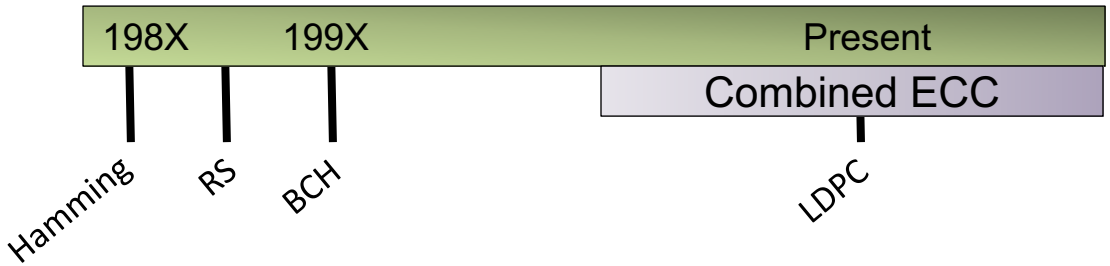
Error Correction Overview

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

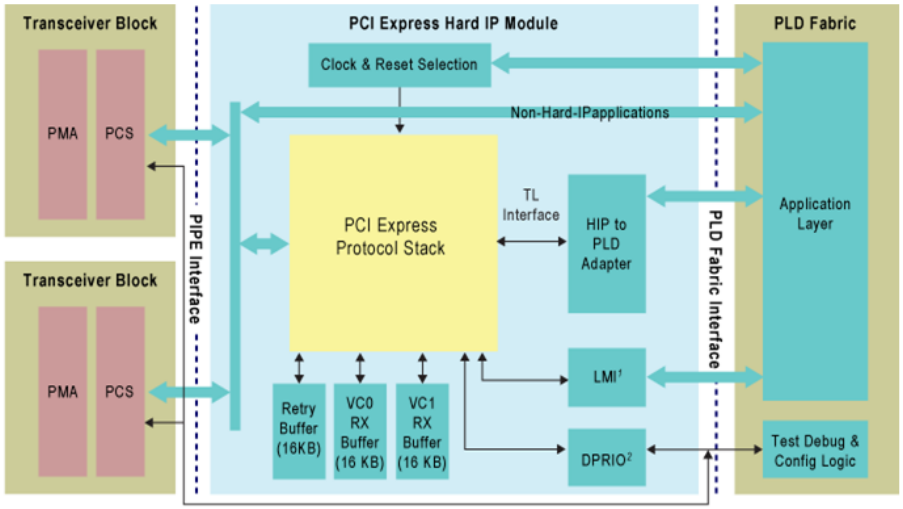
Features	BCH	LDPC
Gate Count	High	Mid
Latency	Low	Medium
Tuneablity	low	high
Soft Data	no	high
Data Overhead	high	low



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 MRAM			5KLE 5KLE	PCM- Pending production $ MRAM- Persistent memory controller
BCH			<10KLE	Reference design
PCIe	G3x8	64Gbps	HIP	Flash Cache

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	2018

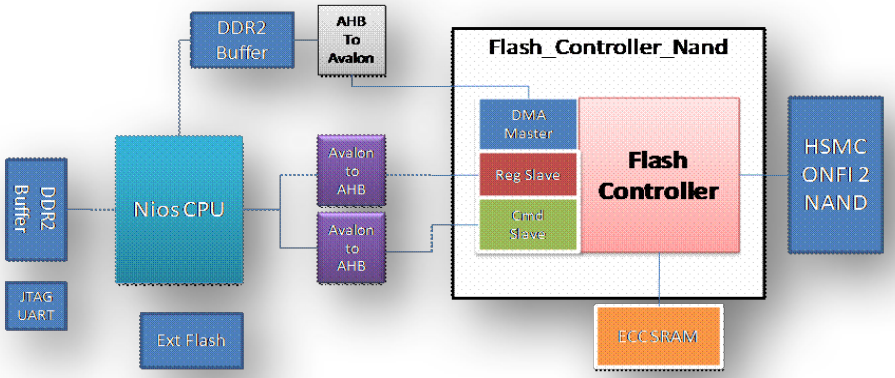
Hardened IP (HIP) Advantages

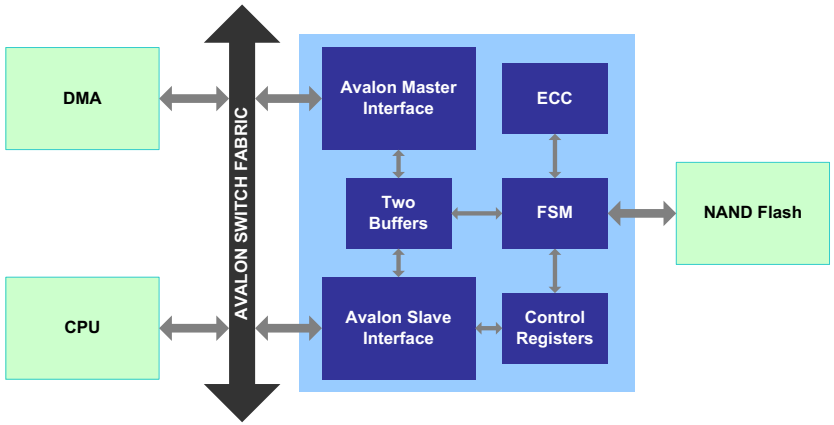
- Resource savings of 8K to 30K logic elements (LEs) per hard IP instance, depending on the initial core configuration mode
- Embedded memory buffers included in the hard IP
- Pre-verified, protocol-compliant complex IP
- Shorter design and compile times with timing closed block
- Substantial power savings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

Flash Cache Controller Examples

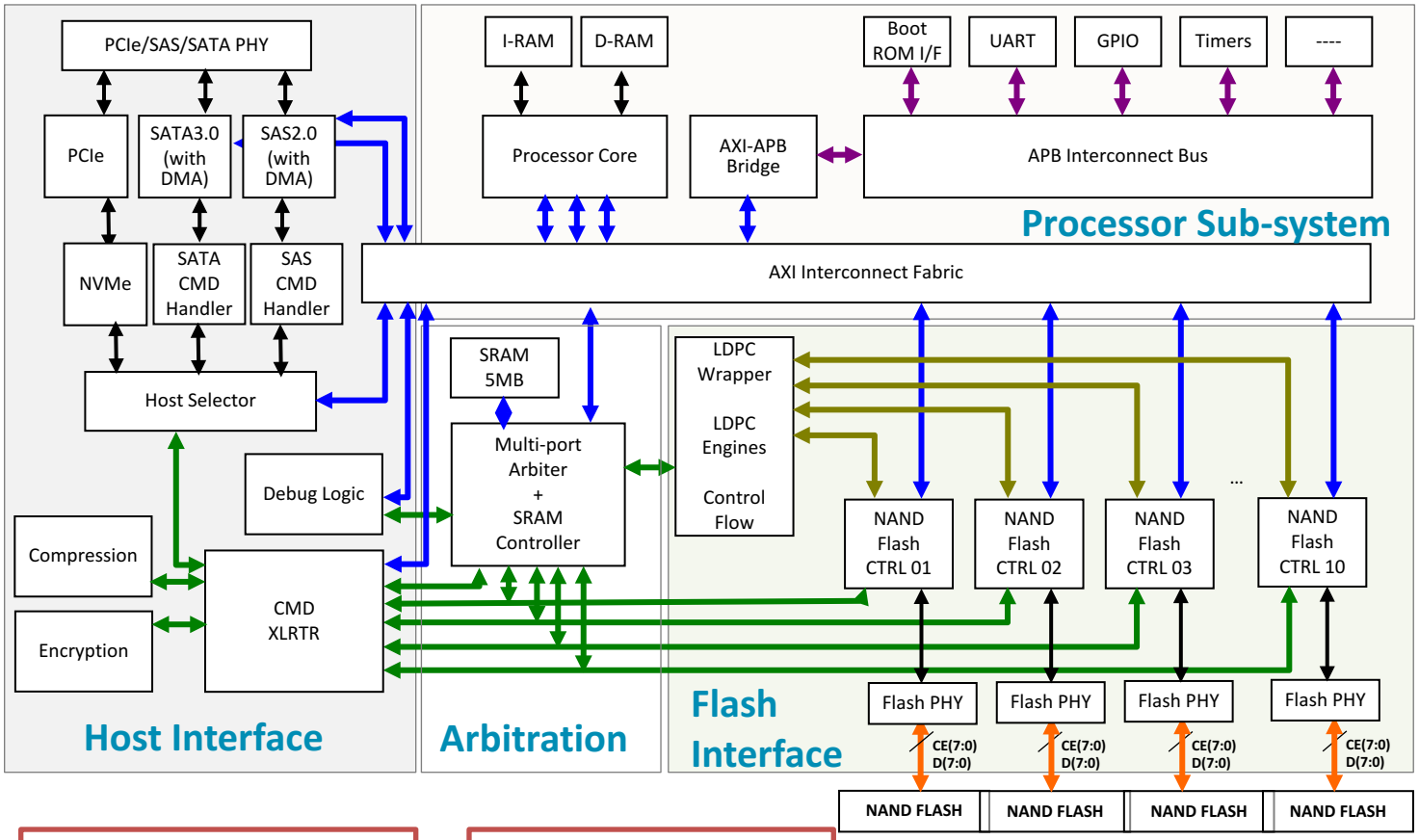
- **Multi Channel Controller**
  - Single to multi Flash channel capability
  - Basic NAND development platform
  - Provides High Speed ONFI & Toggle NAND PHY
  - ECC of 8 and 15 bits of error correction

- **Single Channel Controller**





SSD Controller in FPGA



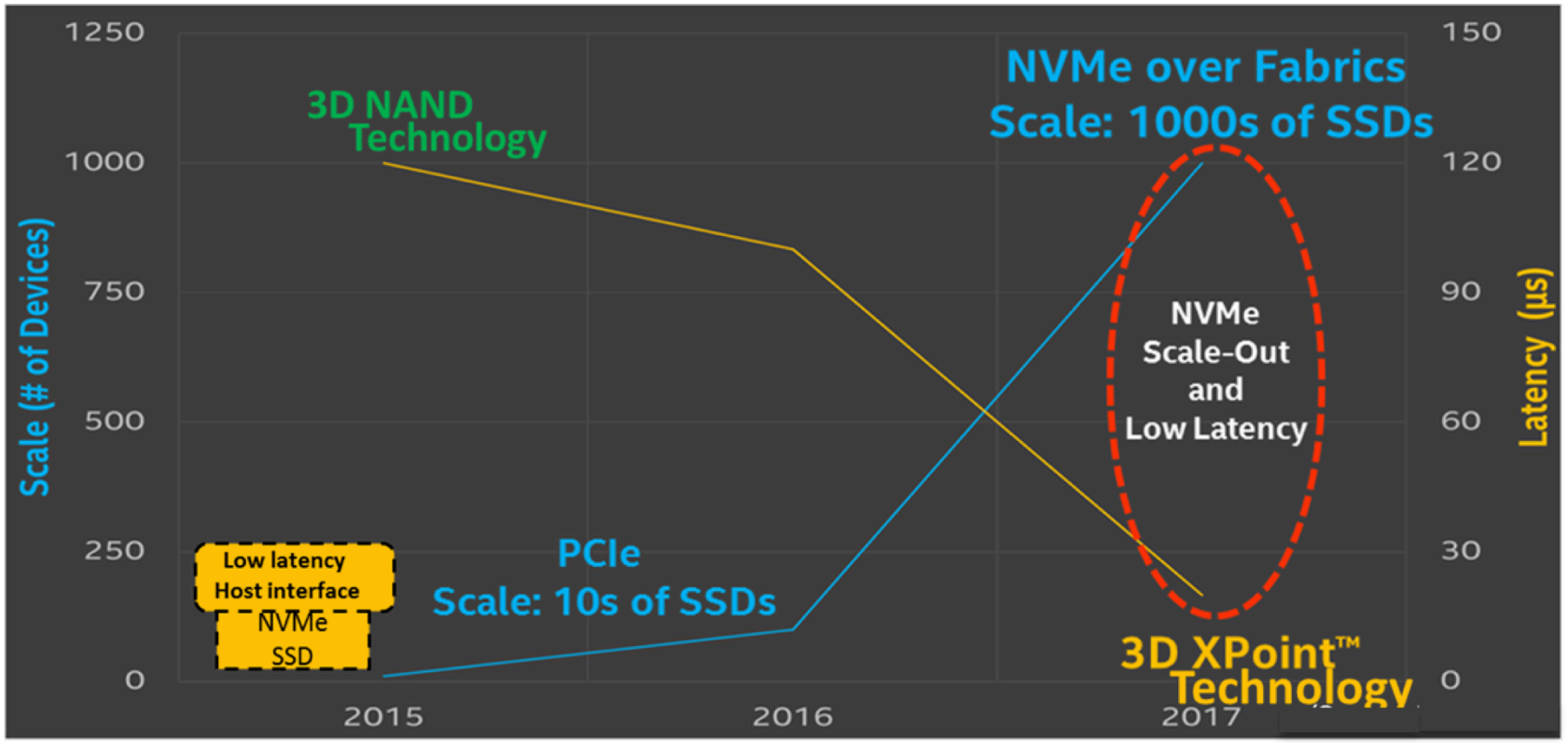
Typical Attributes
- Number of Ports 8 to 32
- Pin Count 250 to 1000+
- Power 1 to 3.5 Watts
- Internal, External RAM

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

AXI Interconnect
APB Bus
AXI for Memory

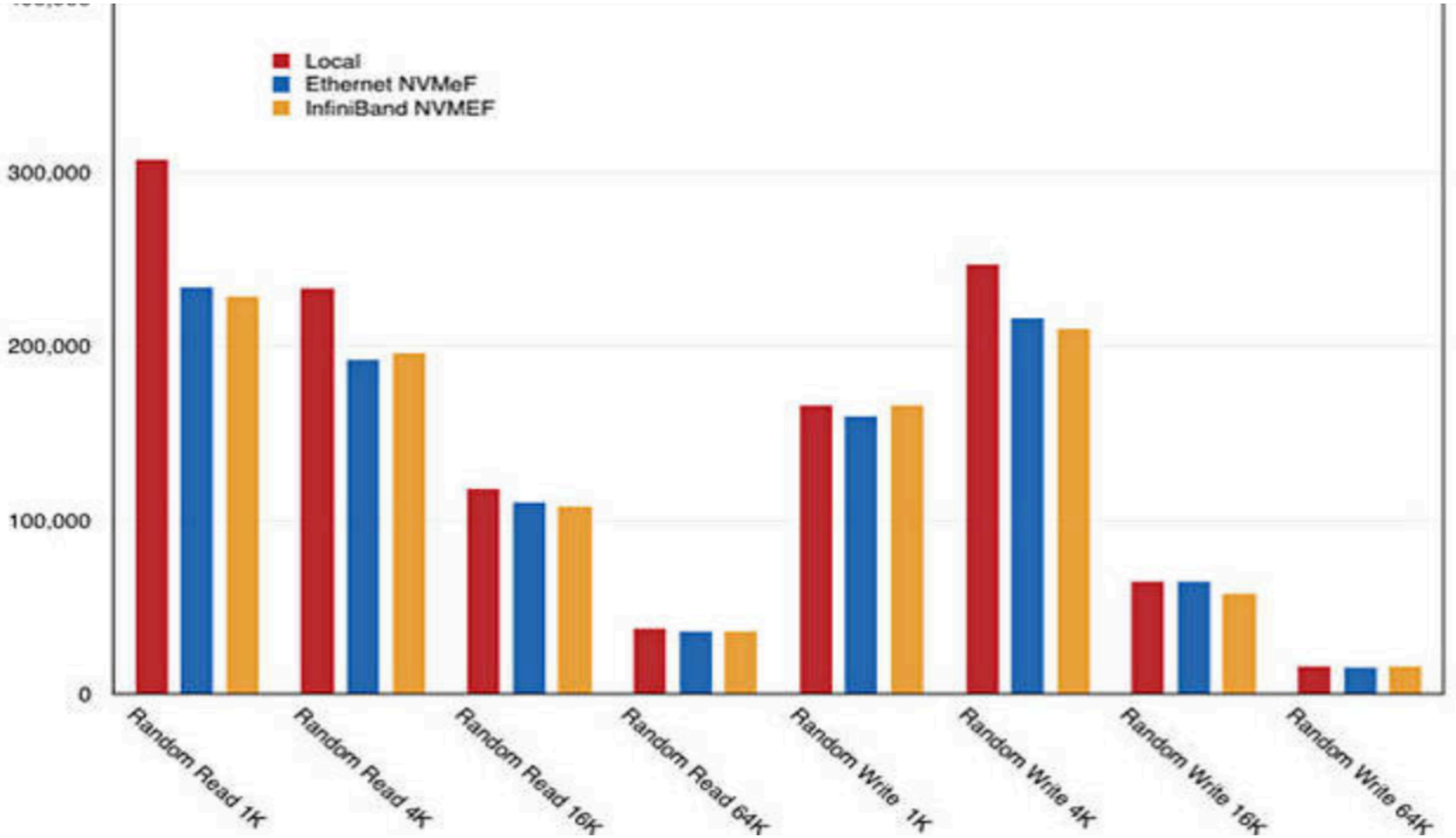
Flash Memory Summit

SSD Scale Out over Fabrics

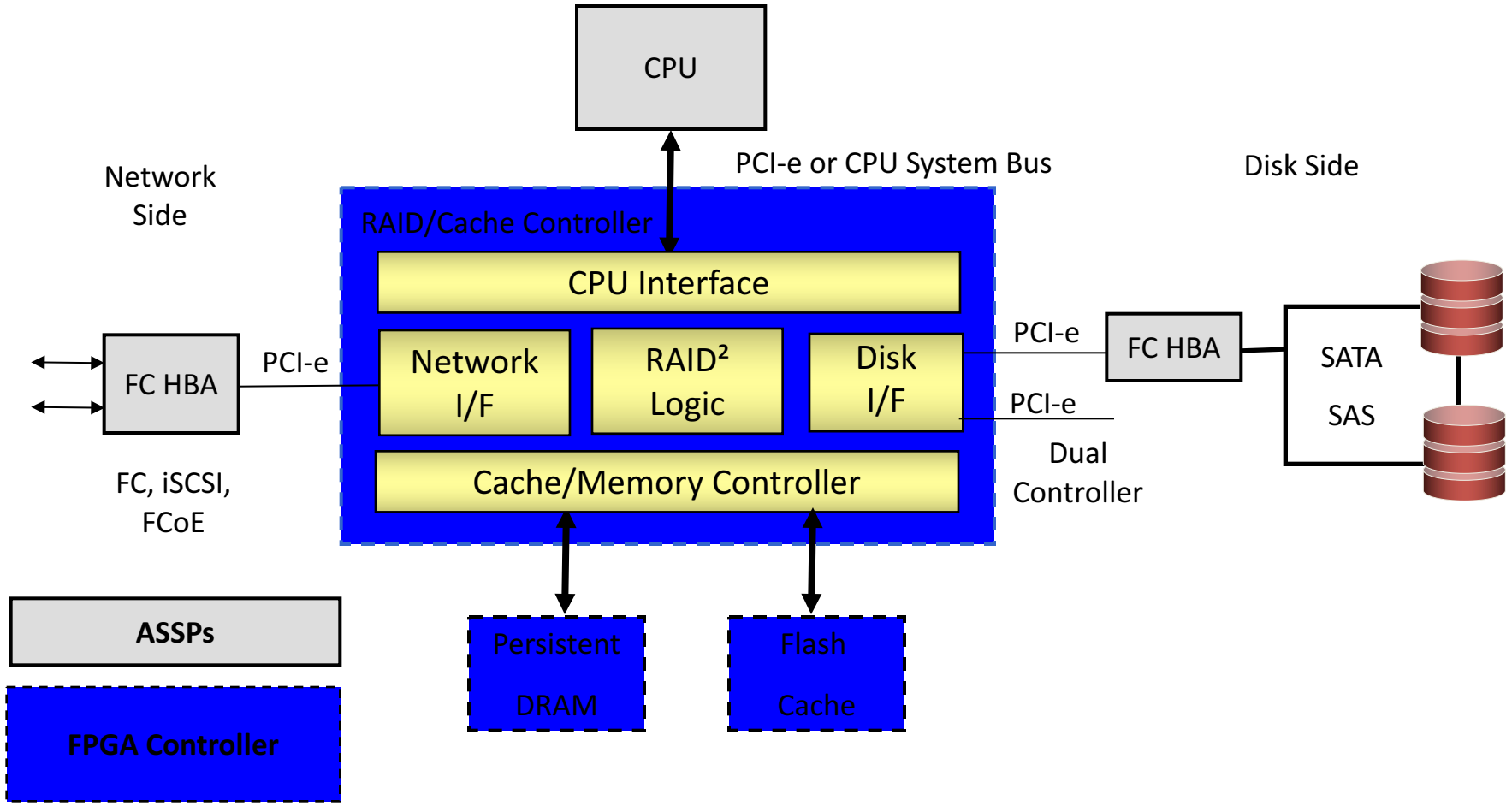


- NVMeF breaks through local NVMe barrier and supports low latency

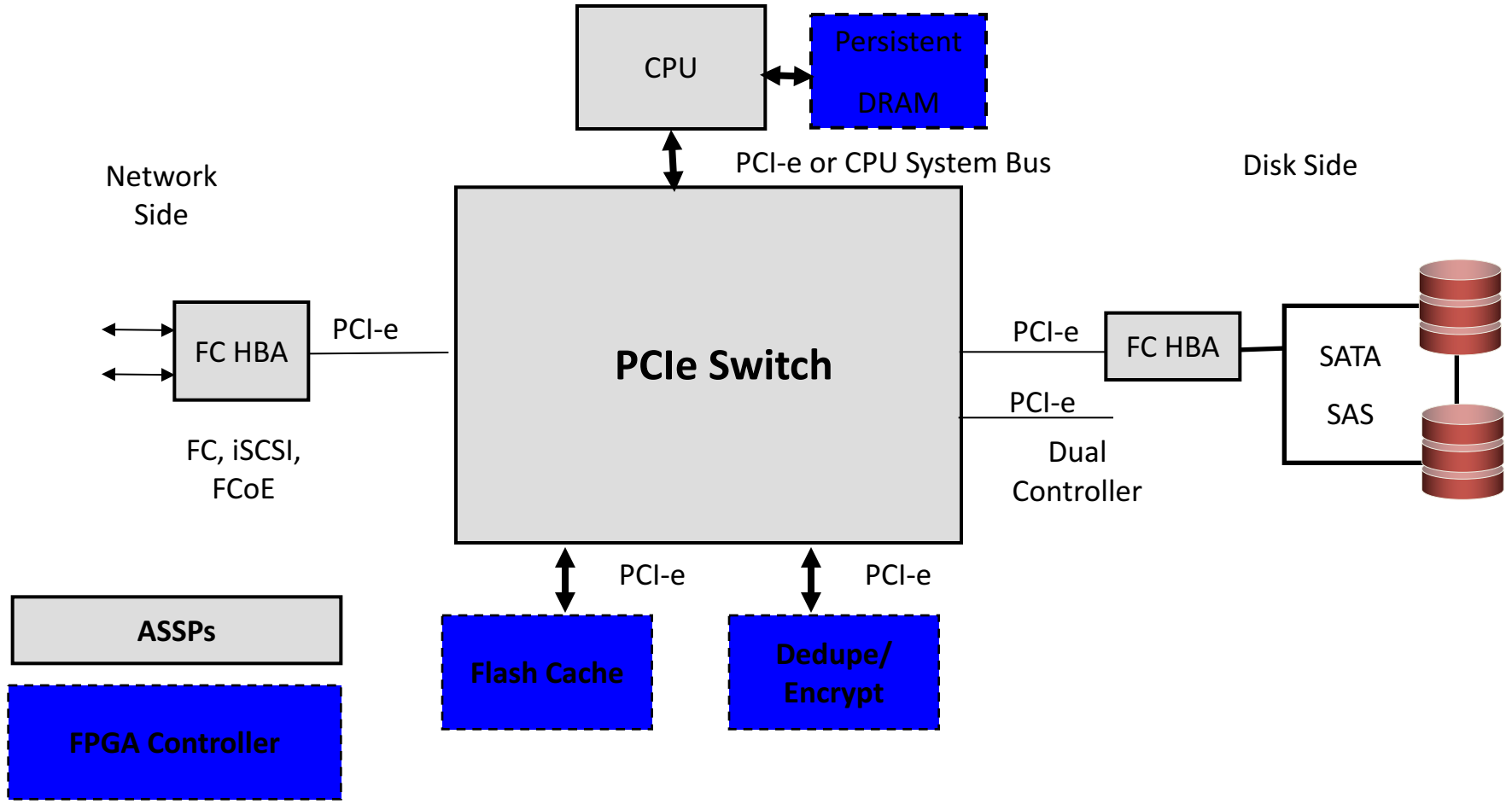
NVMeF approaches NVMe Performance



Hybrid RAID System
- Persistent DRAM and Flash Caches

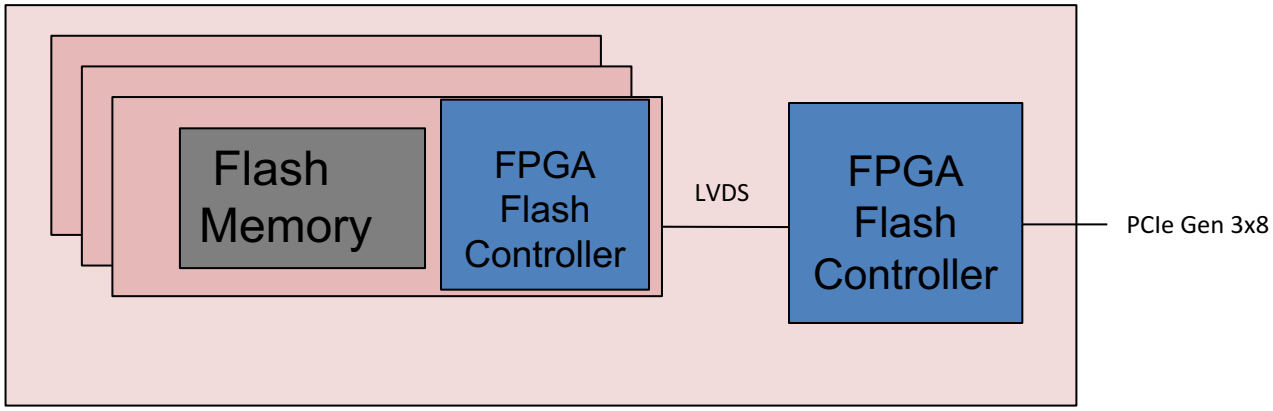


Hybrid RAID System
- PCIe Switch Centric



Flash Storage Arrays

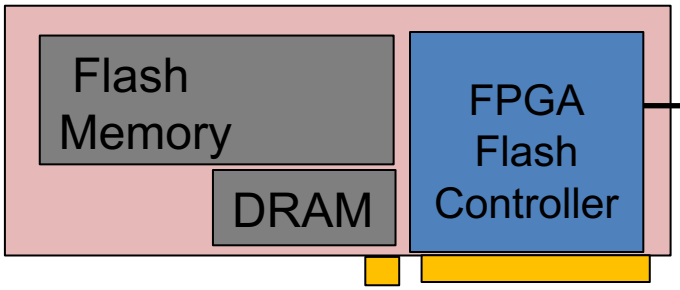
Target Application: Enterprise Tier-1 Storage: Databases and Virtualization



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

Flash Memory Summit

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