OCP SUMMIT

March 20-21
2018
San Jose, CA

OPEN. FOR BUSINESS.

Accelerating Flash Memory with the High Performance, Low Latency, OpenCAPI Interface

Allan Cantle, CTO & Founder, Nallatech/Molex

Marcy Byers, Processor Development, IBM

Nallatech at a Glance

Server qualified accelerator cards featuring FPGAs, network I/O and an open architecture software/firmware framework. Design Services/Application Optimisation

- **Nallatech** – a **Molex** company
- 25 years of FPGA heritage
- Energy-Efficient High Performance Heterogeneous Computing
- Real-time, low latency network and I/O processing
- **Intel** PSG (Altera) OpenCL partner
- **Xilinx** Alliance partner
- Server partners: Cray, DELL, HPE, IBM, Lenovo
- Application porting & optimization services
- Successfully deployed high volumes of FPGA accelerators

Nallatech
a molex company

Hyperconvergence vs Disaggregation – An unavoidable Oxymoron?

Can we hyperconverge & disaggregate Flash Memory at the same time?

Hyperconverged Architectures

- CPU Centric Playbook
- Best Single Threaded Performance
- Tightest of CPU/Accelerator Coupling
  - Holy Grail = ∞ Bandwidth & 0 Latency
- Easier acceleration of Legacy Code

- PCIe is today's convergence bus
  - E.g. NVMe SSDs

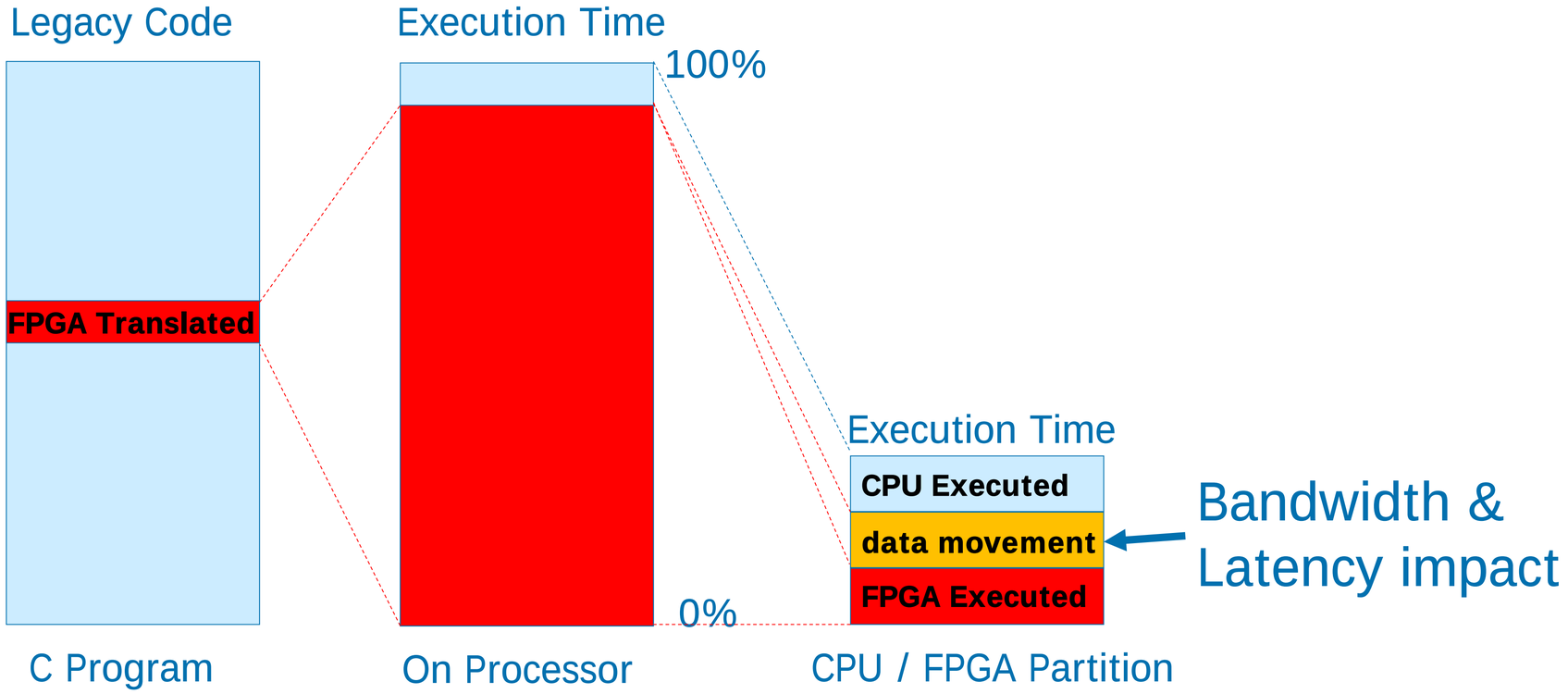
Disaggregated Architectures

- Data Centric Playbook
- Heterogeneous & Distributed compute
- Prioritize Application Dataflow needs
- Can put congestion back into the Network
- Latency managed, compute => data

- Ethernet is today's disaggregation fabric
  - E.g. NVMe-oF

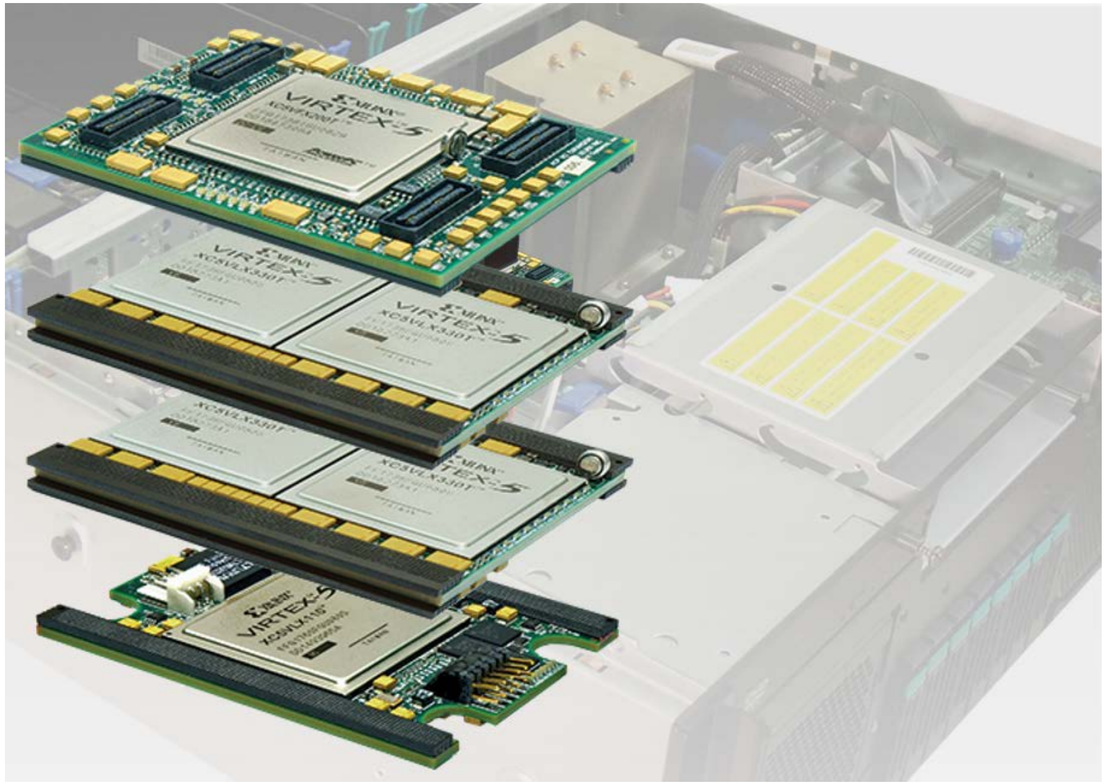
Hyperconverged accelerations quest for ∞ Bandwidth & 0 Latency

- Moving a single thread of data from the CPU to an accelerator can negate the acceleration benefit
  - therefore Acceleration > overhead of data movement to/from accelerator
- Partition code for minimum data movement and maximum acceleration



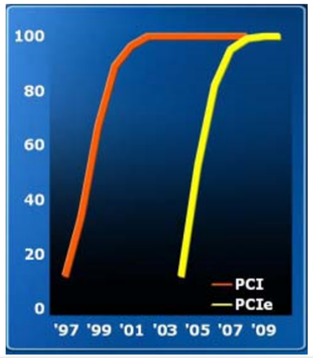
Hyperconverged tightly Coupled FPGA Acceleration is not new.....

- Intel, Xilinx, Nallatech & ISI collaborated on FSB & QPI attached Accelerators
- Started Circa 2007
- Despite a decade of ongoing efforts, commercial reality has been elusive



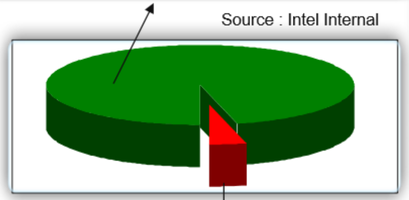
Accelerators – Open Attach Strategy

Source : Pat Gelsinger Keynote Fall IDF 2006



Geneseo – PCI Express*

Source : Intel Internal



Tightly Coupled

1. **Open Ubiquitous Standards Based Approach PCIe* Gen1, PCIe* Gen2, and Geneseo – (Extend PCI Express* Gen 2 - Joint Intel/IBM Proposal in PCI-SIG)**
2. **Enable third party FSB-FPGA Modules – targeted for FSI, Oil and Gas, Life Sciences, Digital Health, etc**

   FSB-FPGA Modules Targeted 4Q07/1Q08

3. **Intel® QuickAssist Technology Accelerator Abstraction Layer that seamlessly allows the SW to access acceleration across various technologies.**

Open Standards Based Attach Strategy

SOURCE : http://rssi.ncsa.uiuc.edu/2007/docs/industry/Intel_presentation.pdf

Why are Tightly Coupled FPGA Accelerators so challenging?

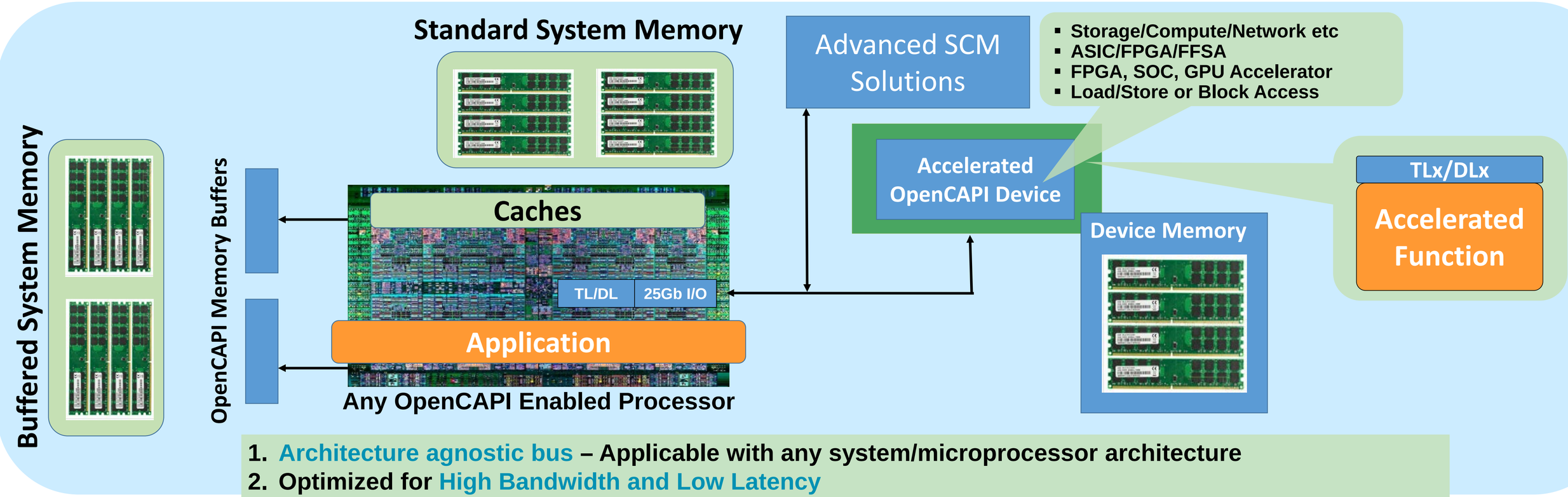
- Tied to complex proprietary coherent busses
  - Rapid cadence of bus standards
  - Limited interface documentation
  - Onerous licensing terms
- Coherent busses not natively designed with FPGAs in mind
  - Pushes limits of FPGA's capabilities
- Heavy burden on FPGA resources for interface IP
  - Impacts performance, in particular latency
  - Reduces resources available for acceleration
- Can drag down the performance of native CPUs using the same bus

OpenCAPI addresses these issues

OpenCAPI™ Overview

Open Compute Project 2018

OpenCAPI Key Attributes

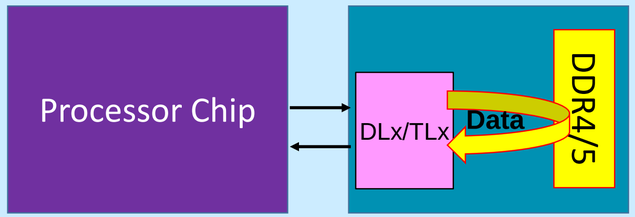


1. Architecture agnostic bus – Applicable with any system/microprocessor architecture
2. Optimized for High Bandwidth and Low Latency
3. High performance 25G interface design with zero ‘overhead’
4. Coherency - Attached devices operate natively within application’s user space and coherently with host microprocessor
5. Virtual addressing enables low overhead with no Kernel, hypervisor or firmware involvement
6. Wide range of Use Cases and access semantics
7. CPU coherent device memory (Home Agent Memory)
8. Architected for both Classic Memory and emerging Advanced Storage Class Memory

Comparison of Memory Paradigms

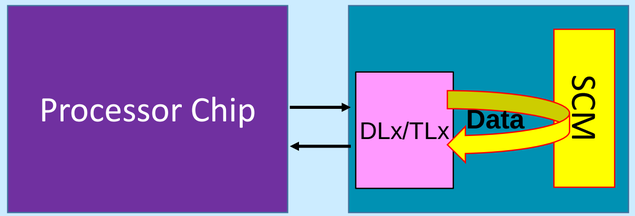
Main Memory

Example: Basic DDR attach



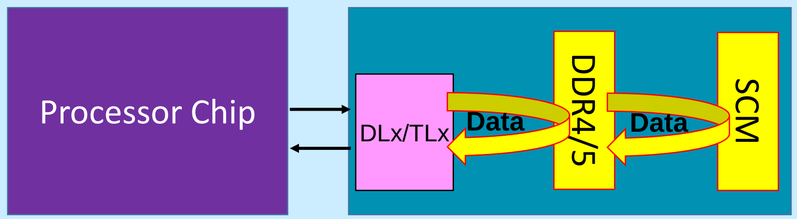
OpenCAPI 3.1 Architecture
Ultra Low Latency ASIC buffer chip adding +5ns on top of native DDR direct connect!!

Emerging Storage Class Memory



Storage Class Memories have the potential to be the next disruptive technology…..
Examples include ReRAM, MRAM, Z-NAND……
All are racing to become the defacto

Tiered Memory

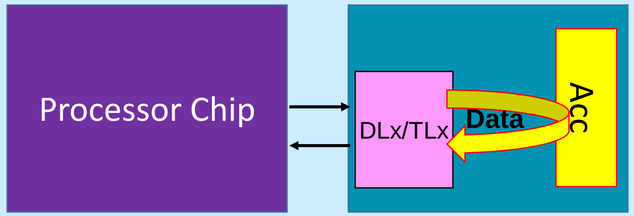


Storage Class Memory tiered with traditional DDR Memory all built upon OpenCAPI 3.1 & 3.0 architecture.
Still have the ability to use Load/Store Semantics

Acceleration Paradigms with Great Performance

Main Transform

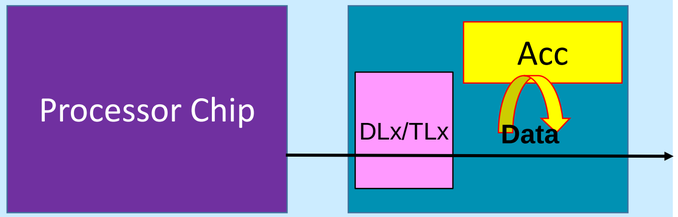
Example: Basic work offload



OpenCAPI WINS due to Bandwidth to/from accelerators, best of breed latency, and flexibility of an Open architecture

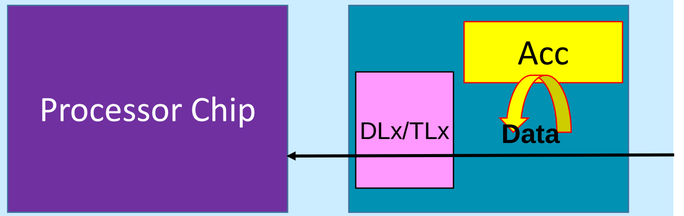
Examples: Machine or Deep Learning such as Natural Language processing, sentiment analysis or other Actionable Intelligence using OpenCAPI attached memory

Egress Transform



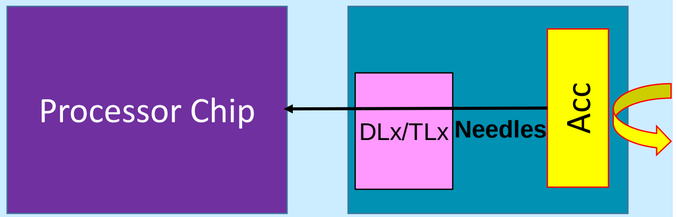
Examples: Encryption, Compression, Erasure prior to delivering data to the network or storage

Ingress Transform



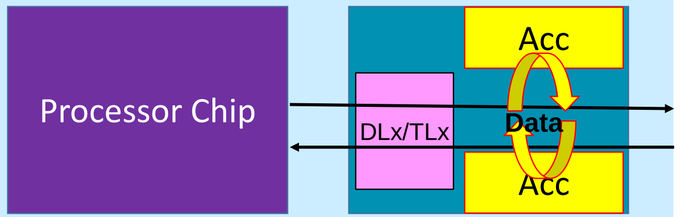
Examples: Video Analytics, Network Security, Deep Packet Inspection, Data Plane Accelerator, Video Encoding (H.265), High Frequency Trading, etc

Needle-In-A-Haystack



Examples: Database searches, joins, intersections, merges
Only the Needles are sent to the processor

Bi-Directional Transform



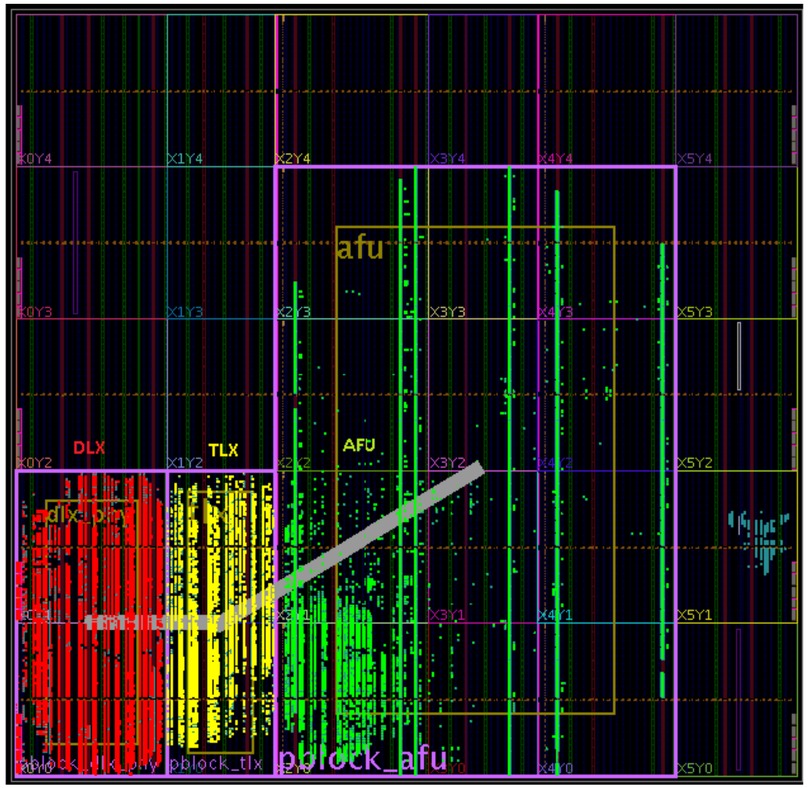
Examples: NoSQL such as Neo4J with Graph Node Traversals, etc

TLx and DLx Reference Designs in an FPGA

- TLx and DLx will be provided as reference designs to OpenCAPI consortium members
  - Associated reference design specifications for TLx and DLx will also be delivered along with RTL
- TLx and DLx are not symmetric with OTL and ODL that are on the host processor
- Designed to operate at 400MHz
- Xilinx Vivado 2017.1 TLx and DLx Statistics on **VU3P** Device

VU3P Resources	CLB FlipFlops	LUT as Logic	LUT Memory	Block Ram Tile
DLx	9392/788160 (1.19%)	19026/394080 (4.82%)	0/197280 (0%)	7.5/720 (1.0%)
TLx	13806/788160 (1.75%)	8463/394080 (2.14%)	2156/197280 (1.09%)	0/720 (0%)

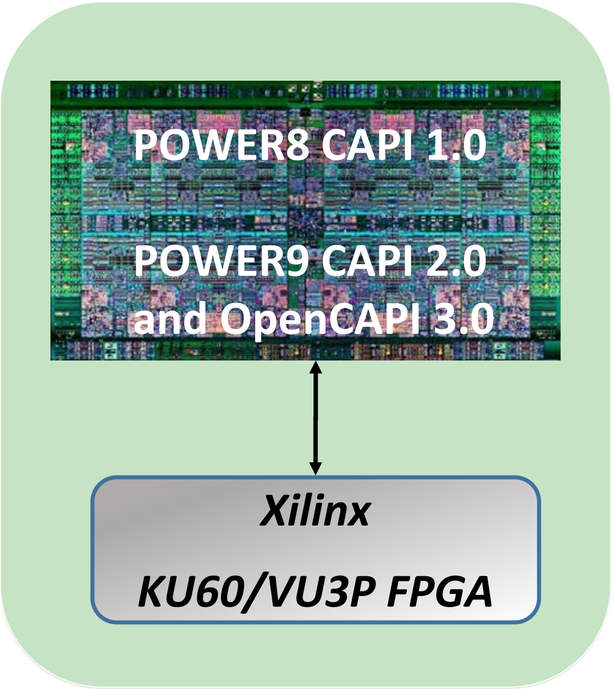
FPGA	Total kLUTs	Fabric Utilization
VU3P	394	8.1%
KU15P	523	6.1%
VU9P	1182	2.7%



OpenCAPI IP Floorplan on a VU3P

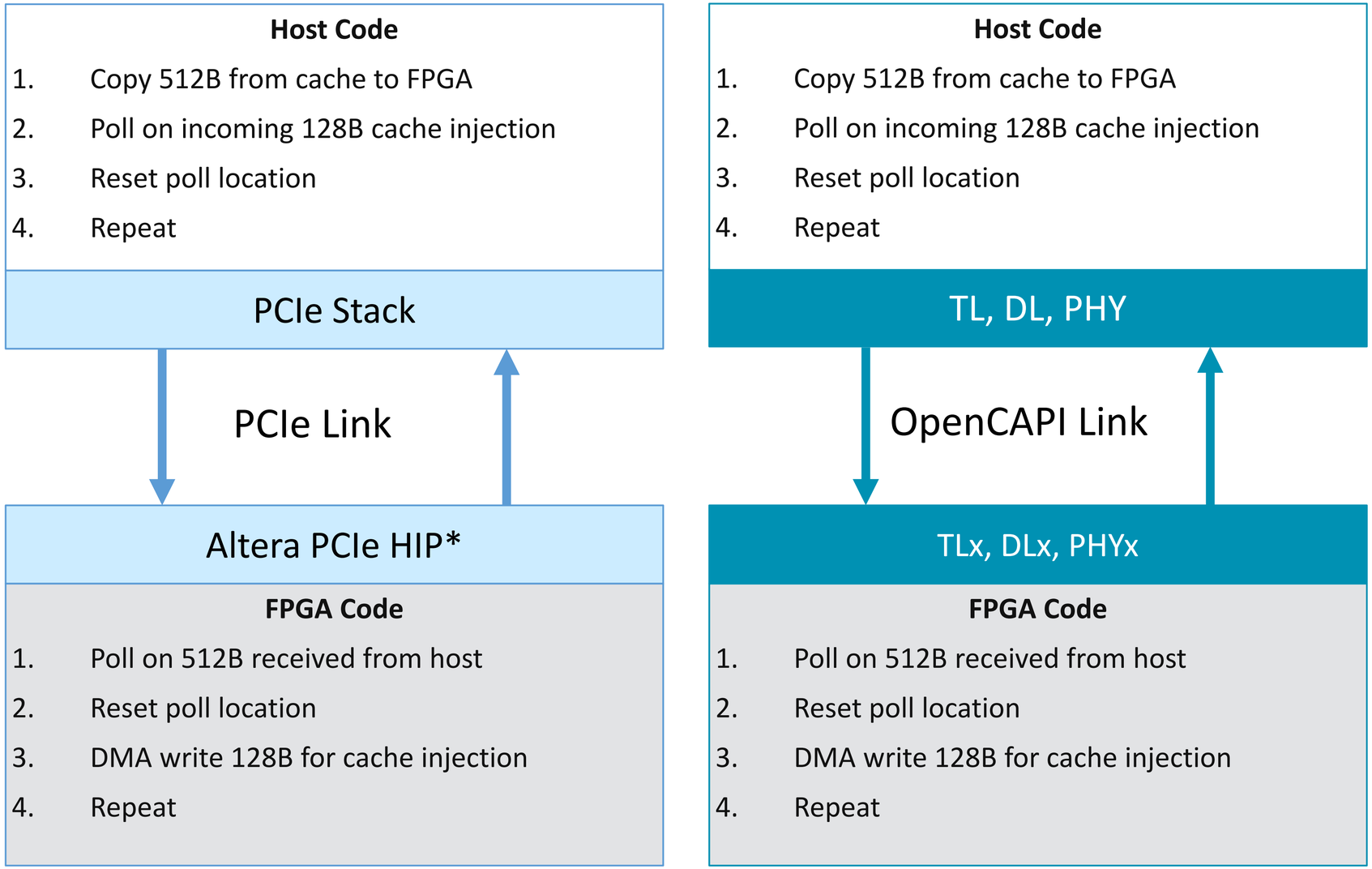
CAPI and OpenCAPI Performance

	CAPI 1.0 PCIE Gen3 x8 Measured Bandwidth @8Gb/s	CAPI 2.0 PCIE Gen4 x8 Measured Bandwidth @16Gb/s	OpenCAPI 3.0 25 Gb/s x8 Measured Bandwidth @25Gb/s
128B DMA Read	3.81 GB/s	12.57 GB/s	**22.1 GB/s**
128B DMA Write	4.16 GB/s	11.85 GB/s	**21.6 GB/s**
256B DMA Read	N/A	13.94 GB/s	**22.1 GB/s**
256B DMA Write	N/A	14.04 GB/s	**22.0 GB/s**
	POWER8 Introduction in 2013	***POWER9 2nd Generation***	***POWER9 Open Architecture with a Clean Slate Focused on Bandwidth and Latency***



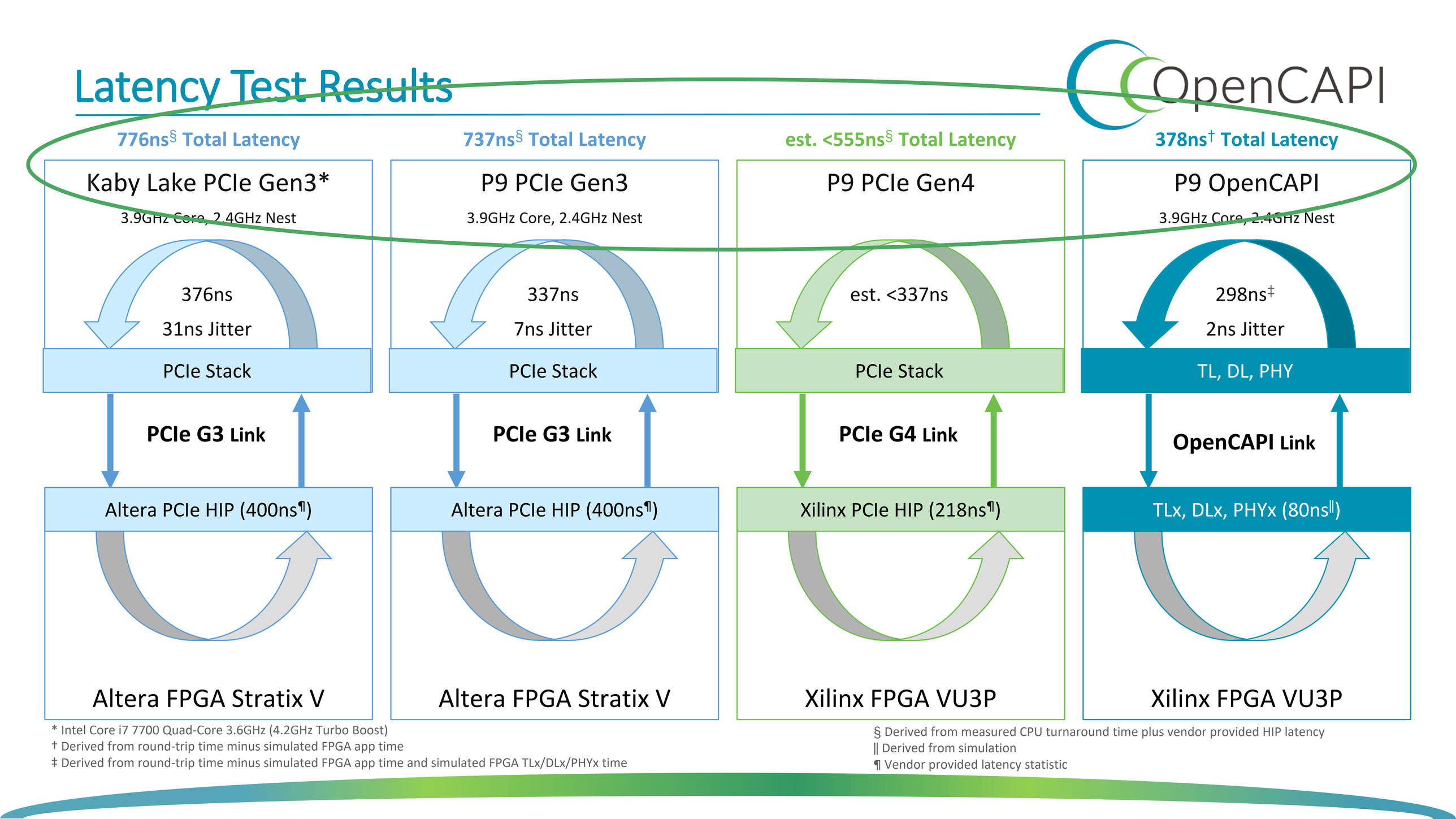
Latency Ping-Pong Test

- Simple workload created to simulate communication between system and attached FPGA
- Bus traffic recorded with protocol analyzer and PowerBus traces
- Response times and statistics calculated



* HIP refers to hardened IP

Latency Test Results



* Intel Core i7 7700 Quad-Core 3.6GHz (4.2GHz Turbo Boost)
† Derived from round-trip time minus simulated FPGA app time
‡ Derived from round-trip time minus simulated FPGA app time and simulated FPGA TLx/DLx/PHYx time

§ Derived from measured CPU turnaround time plus vendor provided HIP latency
‖ Derived from simulation
¶ Vendor provided latency statistic

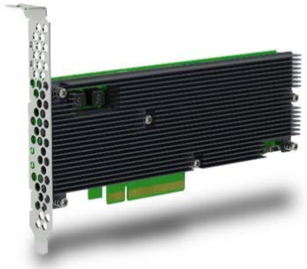
Nallatech / Molex ASG – CAPI Flash Acceleration Product Timeline

CAPI 1.0 Bridge to IBM's Flash Drawer Power8 PCIe Gen 3

Altera 5SGXA7 FPGA to Fiber Channel Interface

Introduced 2014

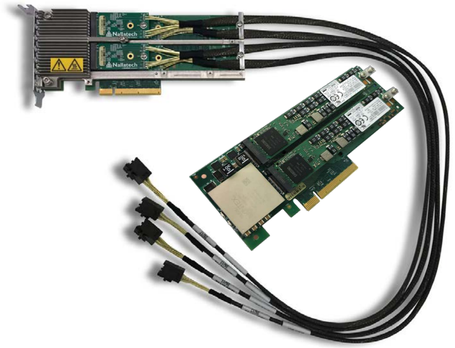
CAPI 1.0 FlashGT Power8 PCIe Gen 3 Storage Acceleration



Xilinx KU060 FPGA + 2x 1TByte M.2 NVMe SSDs

Introduced 2016

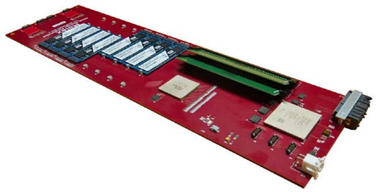
CAPI 2.0 FlashGT+ Power9 PCIe Gen 4 Storage Acceleration



Xilinx KU15P FPGA + 4x 1TByte M.2 NVMe SSDs Or 4x cabled U.2 NVMe SSDs

Introduced 2017

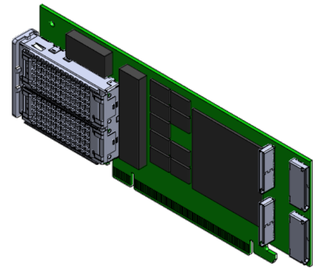
OpenCAPI Hyperconverged & Disaggregatable Flash Storage Accelerator for Zaius/Barreleye-G2 OCP Power9 platforms



Xilinx ZU19P MPSoC FPGA + 8x 2TByte M.2 NVMe SSDs + 50GB/s Dataplane Fabric IO

Introduced 2018

OpenCAPI , CAPI 2.0, PCIe 250-SoC Generic Storage Acceleration



Xilinx ZU19P MPSoC FPGA + 4x PCIe G4x8 cabled Storage IO 50GB/s Dataplane Fabric IO

Introduced 2018

Leveraging OpenCAPI as a Bridge to a Data Centric World

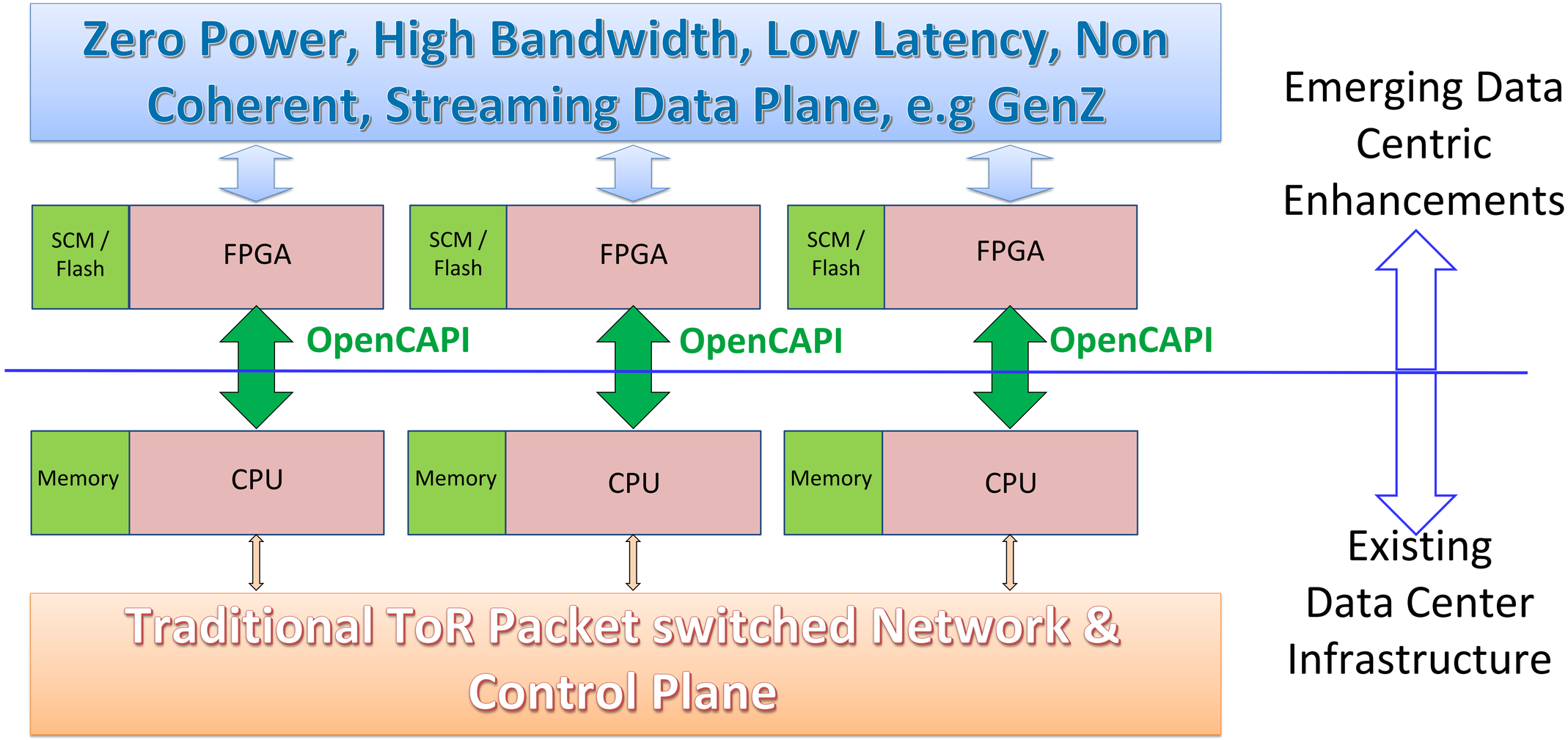
Data Centric Architectures - Fundamental Principles

1. Consume Zero Power when Data is Idle
2. Don’t Move the Data unless you absolutely have to
3. When Data has to Move, Move it as efficiently as possible

Which Translates to : -

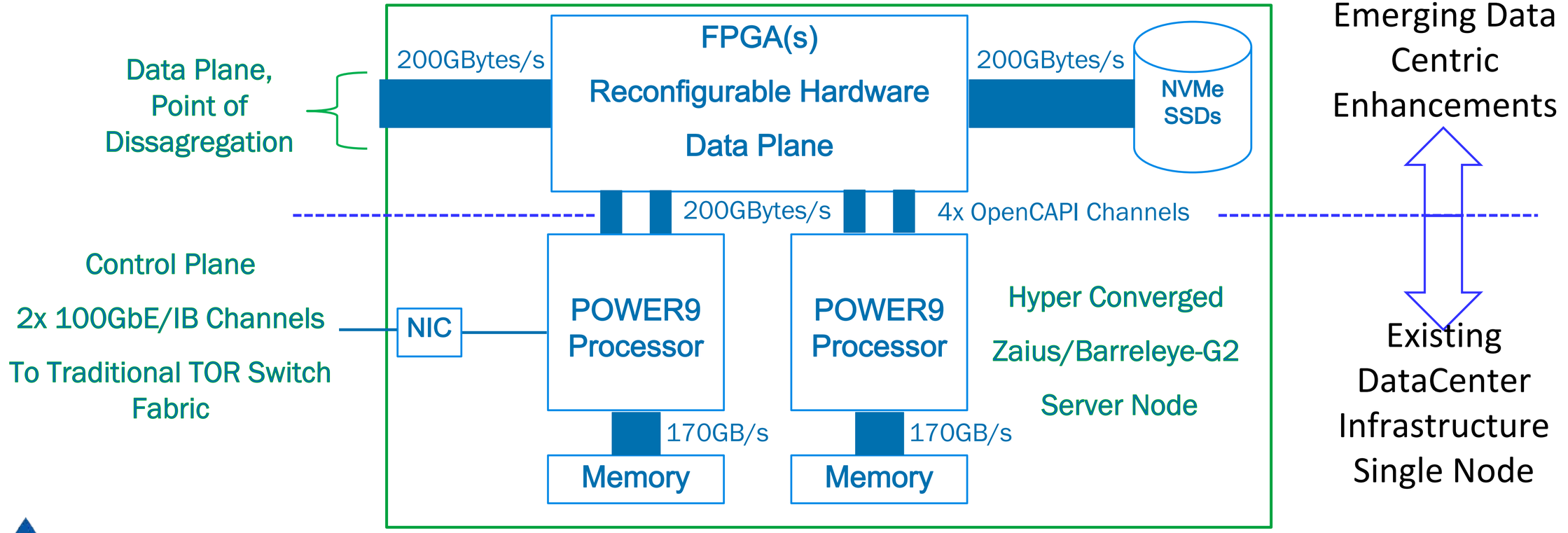
1. Use Non Volatile Memory where possible
2. Move the compute to the data
3. Leverage independent power efficient Dataplanes

Data Center Architectures, blending evolutionary with revolutionary



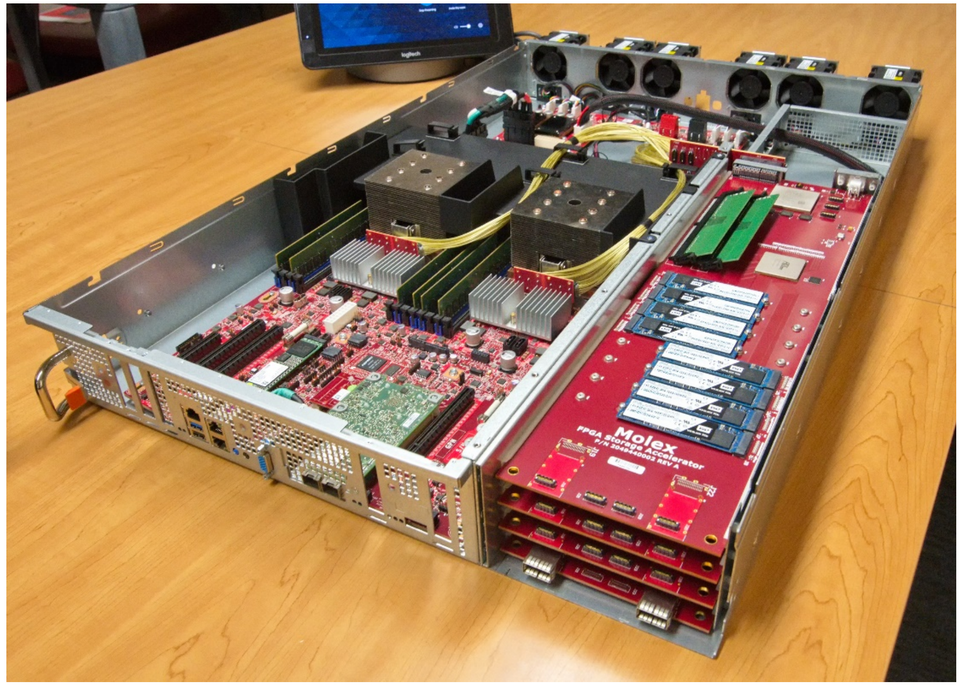
Molex ASG HyperConverged & Disaggregatable Server

- Leverage Google & Rackspace's OCP Zaius/Barreleye G2 platform
- Reconfigurable FPGA Fabric with Balanced Bandwidth to CPU, Storage & Data Plane Network
- OpenCAPI provides Low Latency & coherent Accelerator / Processor Interface
- GenZ Memory-Semantic Fabric provides Addressable shared memory up to 32 Zetabytes

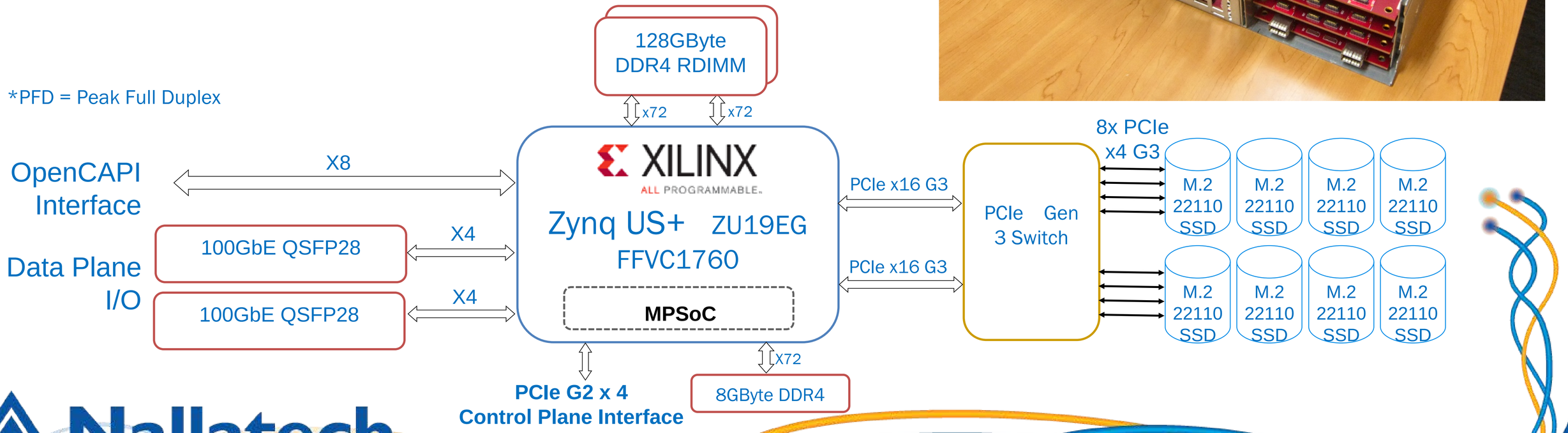


Molex ASG Flash Storage Accelerator, FSA, in Barreleye-G2 OCP Server

- Xilinx Zynq US+ 0.5OU High Storage Accelerator Blade
- 4 FSAs in 2OU Barreleye-G2 OCP Storage drawer deliver :-
  - 152 GByte/s PFD* Bandwidth to 1TB of DDR4 Memory
  - 256 GByte/s PFD* Bandwidth to 64TB of Flash
  - 200 GByte/s PFD* Bandwidth through the OpenCAPI channels
  - 200 GByte/s PFD* Bandwidth through the GenZ Fabric IO
- Open Architecture software/firmware framework



*PFD = Peak Full Duplex



Summary

- The OpenCAPI interface standard is a perfect compliment to the OCP Initiative bringing best in class features including :-
  - Coherency
  - Lowest Latency
  - Highest Bandwidth
  - Open Standard
  - Perfect Bridge to blend CPU Centric & Data Centric Architectures
- Simultaneous Hyperconverged & Dissagregatable Flash Memory solutions can be built without performance compromise
- OCP, OpenCAPI & FPGA Acceleration are now bringing highly optimized Data Centric server architectures closer to reality

JOIN TODAY!

www.opencapi.org

Come see us in the Expo Hall

OpenCAPI Booth C5

OCP SUMMIT