



Integrated Device Technology

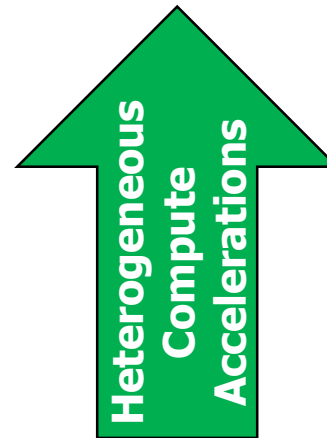
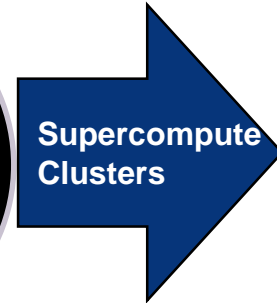
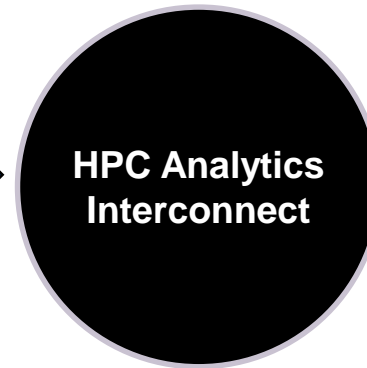
RapidIO based Low Latency Heterogeneous Supercomputing

Devashish Paul, Director Strategic Marketing, Systems Solutions
devashish.paul@idt.com

CERN Openlab Day 2015

© Integrated Device Technology

Agenda



- RapidIO Interconnect Technology Overview and attributes
- Heterogeneous Accelerators
- Open Compute HPC
- RapidIO at CERN OpenLabV

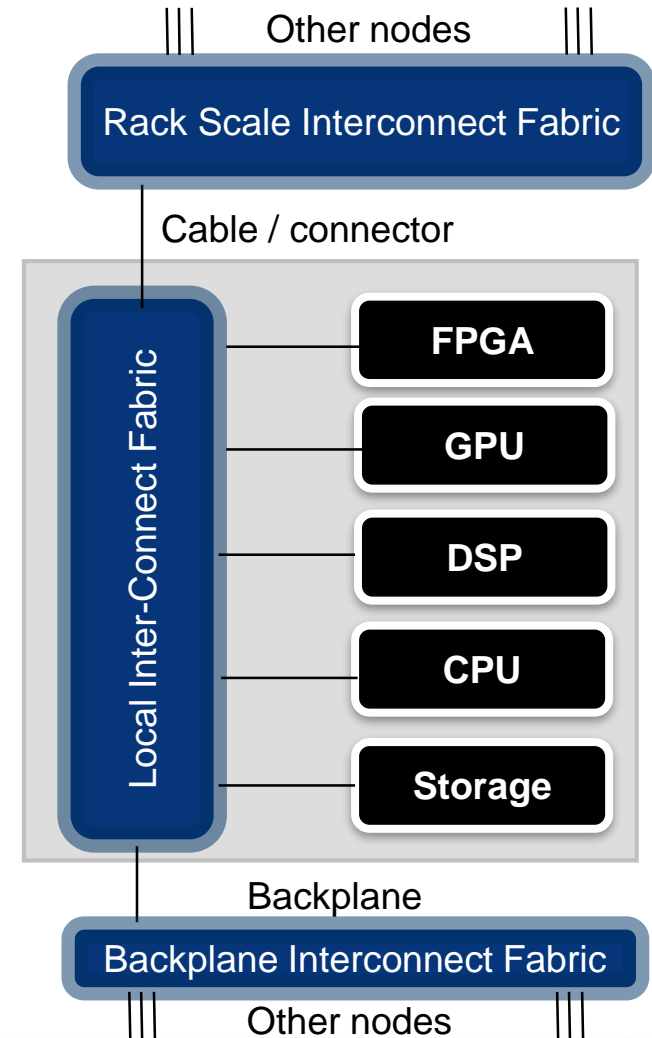
Low Latency | Reliable | Scalable | Fault-tolerant | Energy Efficient

Supercomputing Needs Heterogeneous Accelerators



- Chip to Chip
- Board to Board across backplanes
- Chassis to Chassis
- Over Cable
- Top of Rack
- Heterogeneous Computing

**Rack Scale Fabric
For any to any compute**



RapidIO

Multi-Processor
Embedded Interconnect

Switched | Scalable | Low Latency | Reliable

10 Gbps

20 Gbps

40 Gbps

100+ Gbps

ANY TOPOLOGY
ANY PROCESSOR
OPEN STANDARD



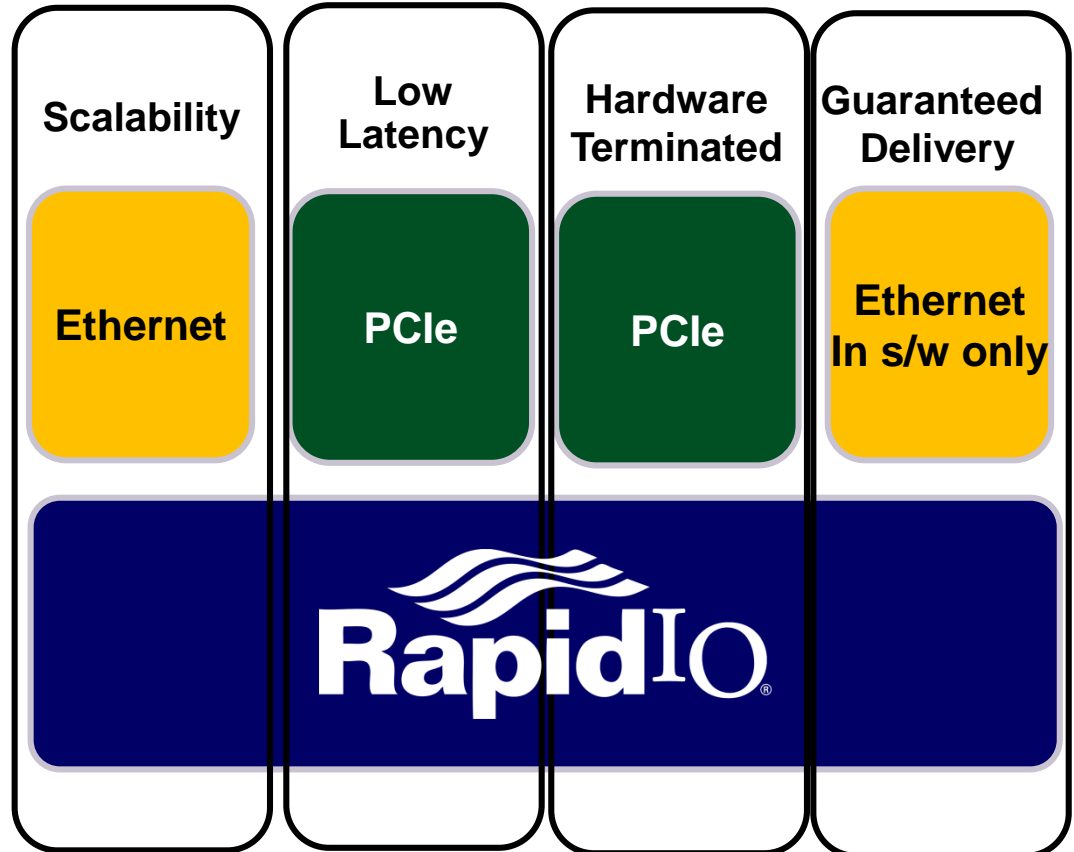
WIRELESS INFRASTRUCTURE | SERVER | HPEC | IMAGING | AEROSPACE | INDUSTRIAL

- **20 Gbps per port / 6.25Gbps/lane** in productions
- **40Gbps per port /10 Gps lane** in development
Embedded RapidIO NIC on processors, DSPs, FPGA and ASICs.
- Hardware termination at PHY layer: 3 layer protocol
- Lowest Latency Interconnect ~ 100 ns
- Inherently scales to large system with 1000's of nodes

- **Over 13 million RapidIO switches shipped**
- **> 2xEthernet (10GbE)**
Over 70 million 10-20 Gbps ports shipped
- **100% 4G interconnect market share**
- 60% 3G, 100% China 3G market share

Clustering Fabric Needs

- **Lowest Deterministic System Latency**
- **Scalability**
- **Peer to Peer / Any Topology**
- **Embedded Endpoints**
- **Energy Efficiency**
- **Cost per performance**
- **HW Reliability and Determinism**



RapidIO Interconnect combines the best attributes of PCIe and Ethernet in a multi-processor fabric

RapidIO Interoperable Eco-system

Low Latency | Reliable | Scalable | Fault-tolerant | Energy Efficient

RapidIO

ALTERA
FPGA: Arria and Stratix Family

TEXAS INSTRUMENTS
DSP: several products In TCI64xx family

LSI
Axxia Communications Processor

freescale™
semiconductor
DSP, PowerQUICC & QorIQ multicore

BROADCOM
XLS416 family Multicore Processor

IDT
Switches, Bridges & IP CPS and Tsi Family

XILINX
FPGA: Virtex 4/5/6 families

Lattice
Semiconductor Corporation
FPGA

octasic
DSP Oct22xx

ARM

HUAWEI

MNDSPEED™
Wireless Baseband Processor

AMD

CAVIUM NETWORKS
Network Processor Octeon 2 family

applied micro
POWERING YOUR...
PowerPC based processors 460GT

Wintegra™
Network Processor WinPath3

intel

mobiveil
Investors in Innovation™

Source - http://www.rapidio.org/files/RapidIO_Asia_Summit_Intro.pdf

Open Interoperable Ecosystem
No vendor lock in

RapidIO.org ARM64 bit Scale Out Group



Latency

Scalability

Hardware Termination

Energy Efficiency



The slide titled "Task Group Contributors" features the RapidIO logo at the top left. Below the title, logos for AMD, ARM, CAVIUM, freescale, IDT, MERCURY SYSTEMS, mobiveil, TEXAS INSTRUMENTS, and XILINX are arranged in a grid. At the bottom, it includes the date "22-Oct-2014", the tagline "RapidIO - the Unified Fabric for Performance Critical Computing", and the page number "4".

- 10s to 100s cores & Sockets
- ARM AMBA® protocol mapping to RapidIO protocols
 - AMBA 4 AXI4/ACE mapping to RapidIO protocols
 - AMBA 5 CHI mapping to RapidIO protocols
- Migration path from AXI4/ACE to CHI and future ARM protocols
- Supports Heterogeneous computing
- Support Wireless/HPC/Data Center Applications

Source – www.rapidio.org, Linley Processor conf

NASA Space Interconnect Standard

Next Generation Spacecraft Interconnect Standard



Key Driving Differentiators

- Serial RapidIO has the following salient features among four protocols:
 - Transparent compatibility with wired and fiber-optic
 - Applicable to chip-to-chip, board-to-board, and box-to-box
 - Light-weight and modular (features are configurable)
 - Low power with less than 192 mW per node
 - Scalable fault tolerance with link-level error detection
 - Scalable bandwidth up to 3.125 Gbps per lane
 - Real-time with sub-microsecond latency and jitter
 - Switch-based flexible topology
 - Built-in shared-memory support with low S/W overhead
 - Embedded provisions allow backward-compatible protocol extension



**RapidIO selected from
Infiniband / Ethernet
/FiberChannel / PCIe**

NGSIS members: BAE,
Honeywell, Boeing, Lockheed-
Martin, Sandia
Cisco, Northrup-Grumman,
Loral, LGS, Orbital Sciences,
JPL, Raytheon, AFRL

Hyperscale Data Center Real Time Analytics Example

PayPal solves real-time analytics problems with HP Moonshot

Global online payment service analyzes complex event streams in real time using HP ProLiant m800 Server Cartridges



“The ProLiant m800’s combination of ARM and multicore DSPs with high-speed, low-latency networking and tiered memory management creates a very energy-efficient, extremely capable parallel processing platform with a familiar Linux interface. It’s a truly new approach to bringing scale-out design “inside the box,” and breaks barriers between HPC and enterprise technology.”

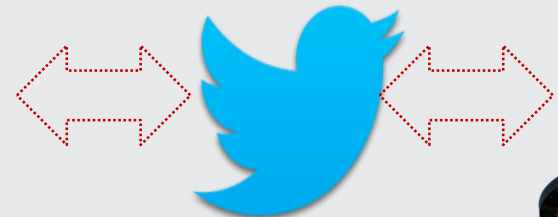
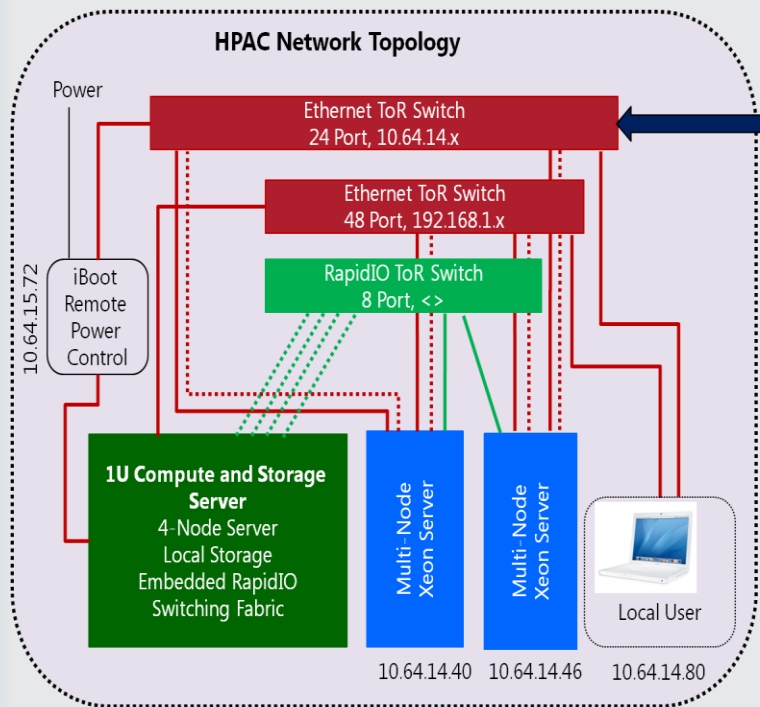
– S. Ryan Quick, Principal Architect, Advanced Technology Group, PayPal

<http://www.enterprisetech.com/2014/09/29/hp-arms-moonshot-servers-datacenters/>

Market emerges for real time compute analytics in data center

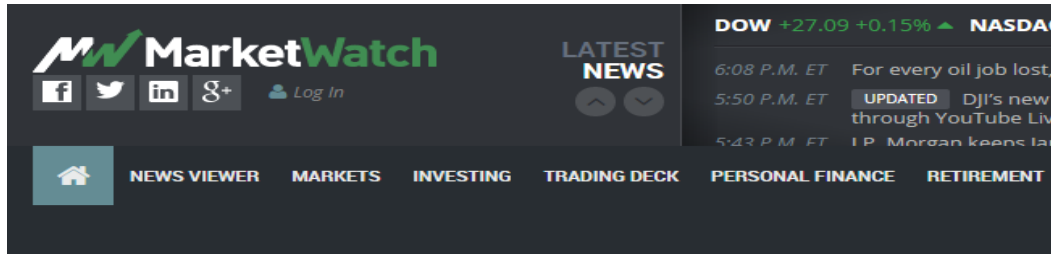
Social Media Real Time Analytics – FIFA World Cup 2014

Analyze User Impressions on World Cup 2014



Future Analytics acceleration using GPU

RapidIO at CERN LHC and Data Center



PRESS RELEASE

IDT Collaborates With CERN to Speed and Improve Data Analytics at Large Hadron Collider and Data Center

- RapidIO Low latency interconnect fabric
- Heterogeneous computing
- Large scalable multi processor systems
- Desire to leverage multi core x86, ARM, GPU, FPGA, DSP with uniform fabric
- Desire programmable upgrades during operations before shut downs



Why RapidIO for Low Latency

Bandwidth and Latency Summary	
System Requirement	RapidIO
Switch per-port performance raw data rate	20 Gbps – 40 Gbps
Switch latency	100 ns
End to end packet termination	~1-2 us
Fault Recovery	2 us
NIC Latency (Tsi721 PCIe2 to S-RIO)	300 ns
Messaging performance	Excellent

Peer to Peer & Independent Memory System

- Routing is easy: Target ID based
- Every endpoint has a separate memory system
- All layers terminated in hardware

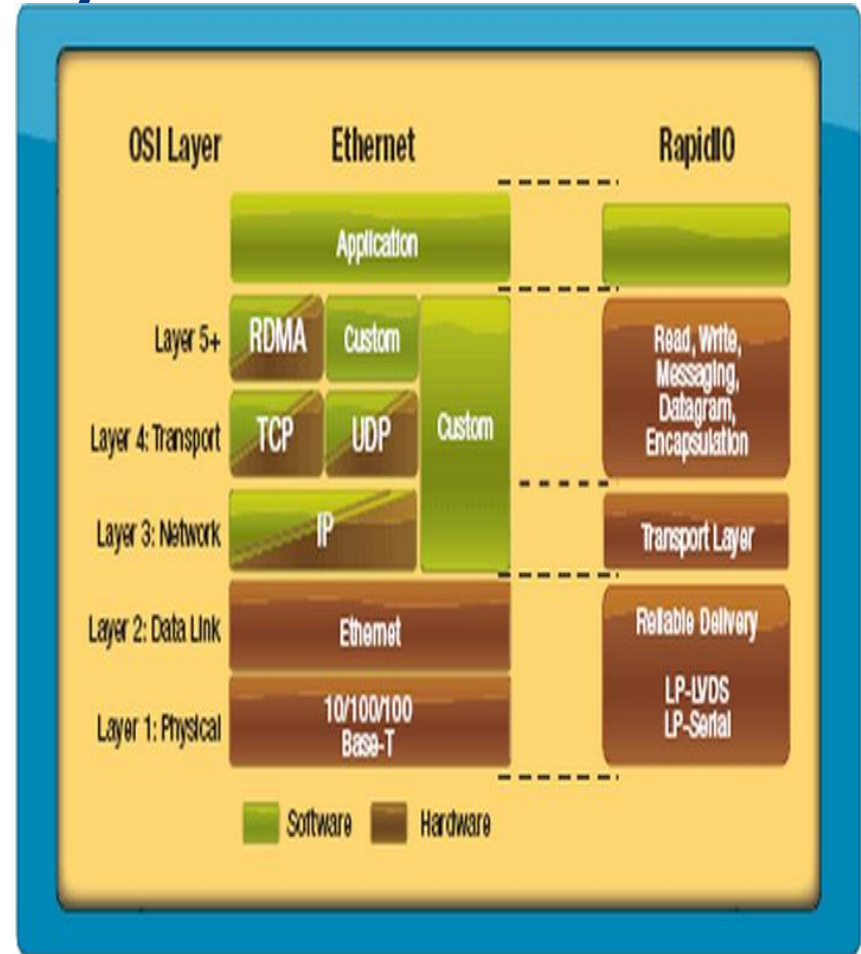
	1	3	1	1	2	2
Prev Packet	S	AckID	Rsrv	S	Rsrv	Prio

2	8 or 16	8 or 16	4
TT	Target Address	Source Address	Transaction





































4	4	8	32 or 48 or 64
Ftype	Size	Source TID	Device Offset Address

8 to 256 Bytes	16	
Optional Data Payload	CRC	Next Packet

Physical
 Transport
 Logical



HPC/Supercomputing Interconnect 'Check In'

Interconnect Requirements	RapidIO	Infiniband	Ethernet	PCIe	Intel Omni Path	The Meaning of 
Low Latency						<ul style="list-style-type: none"> Switch silicon: ~100 nsec Memory to memory : < 1 usec
Scalability						Ecosystem supports any topology, 1000' s of nodes
Integrated HW Termination						Available integrated into SoCs AND Implement guaranteed, in order delivery without software
Power Efficient						3 layers terminated in hardware, Integrated into SoC' s
Fault Tolerant						Ecosystem supports hot swap Ecosystem supports fault tolerance
Deterministic						Guaranteed, in order delivery Deterministic flow control
Top Line Bandwidth						Ecosystem supports > 8 Gbps/lane

Heterogeneous HPC with RapidIO based Accelerators and Open Compute HPC

HPC Goals and Interconnect Relationship

Project Charter Items

Fully open **heterogeneous computing** , networking and fabric platform

Optimized for multi-node **processor agnostic** any to any computing using x86, ARM, PowerPC, FPGA, ASICs, DSP, and GPU silicon on hardware platform

Enables rapid innovation in low latency high Performance Computing and Big Data analytics through open non-lock-in computing, interconnect, and software stack.

Energy efficient compute density

Distributed and central **storage** for large data manipulation (non spinning disk) with low latency

Operating System – Linux based operating systems and developer tools and **open APIs**

Path to **Open Silicon** and Open APIs, initially leveraging existing industry standards, later developing its own silicon

Re use developments from OCP Server group and Open Rack where appropriate

Leverage **industry standard interconnects**, no proprietary interconnects for main fabric and networking

Open interconnect should support HPC charter

RapidIO in Open Compute Interconnect

Twitter, Inc. [US] <https://twitter.com/OpenComputePrj>

Search Twitter Have an account? Log in

Open Compute Project
@OpenComputePrj


OCP is an initiative started by Facebook that aims to accelerate data center and server innovation while increasing computing efficiency.

opencompute.org

TWEETS 248 FOLLOWING 105 FOLLOWERS 2,980 FAVORITES 6 [Follow](#)

Tweets **Tweets & replies** Photos & videos

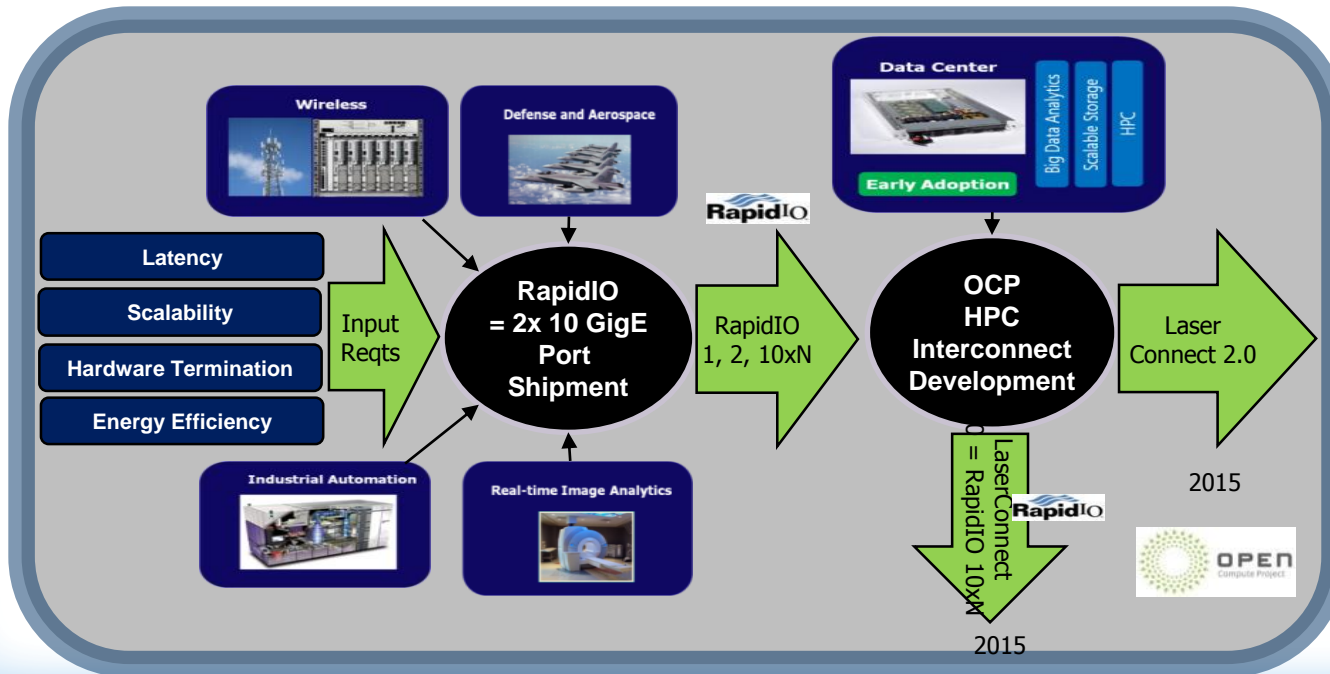
Open Compute Project @OpenComputePrj · Dec 10
IDT Launches RapidIO 40-100 Gbps Interface Portfolio - nasdaq.com/press-release/...



Supporting Opencompute.org HPC

Computing: Open Compute Project HPC and RapidIO

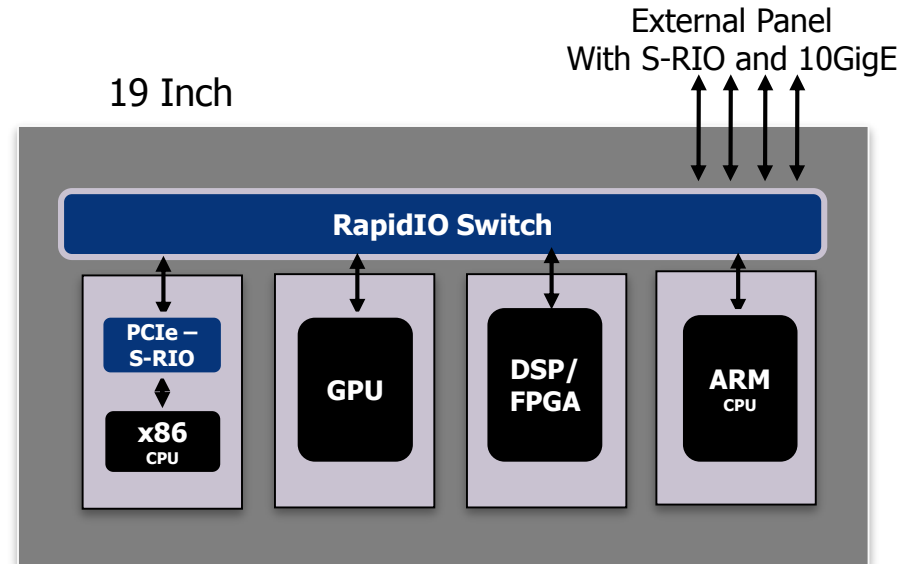
- **Low latency analytics** becoming more important not just in HPC and Supercomputing but in other computing applications
- mandate to create open latency sensitive, energy efficient board and silicon level solutions
- **Low Latency Interconnect RapidIO submission**



4x 25 Gbps
Multi Vendor
Collaboration

RapidIO Heterogenous switch + server

- 4x 20 Gbps RapidIO external ports
- 4x 10 GigE external Ports
- 4 processing mezzanine cards
- In chassis 320 Gbps of switching with 3 ports to each processing mezzanine
- Compute Nodes with x86 use PCIe to S-RIO NIC
- Compute Node with ARM/PPC/DSP/FPGA are native RapidIO connected with small switching option on card
- 20Gbps RapidIO links to backplane and front panel for cabling
- Co located Storage over SATA
- 10 GbE added for ease of migration

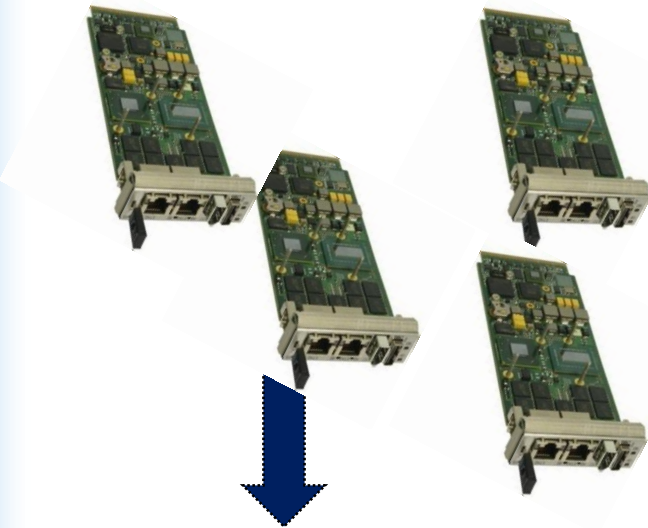


X86 + GPU Analytics Server + Switch

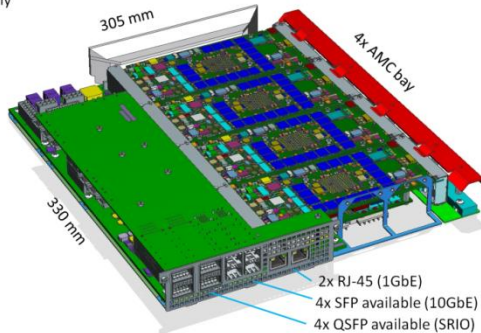
- Energy Efficient Nvidia K1 based Mobile GPU cluster acceleration

- 300 Gb/s RapidIO Switching
- 19 Inch Server
- 8 – 12 nodes per 1U
- 12 - 18 Teraflops per 1U

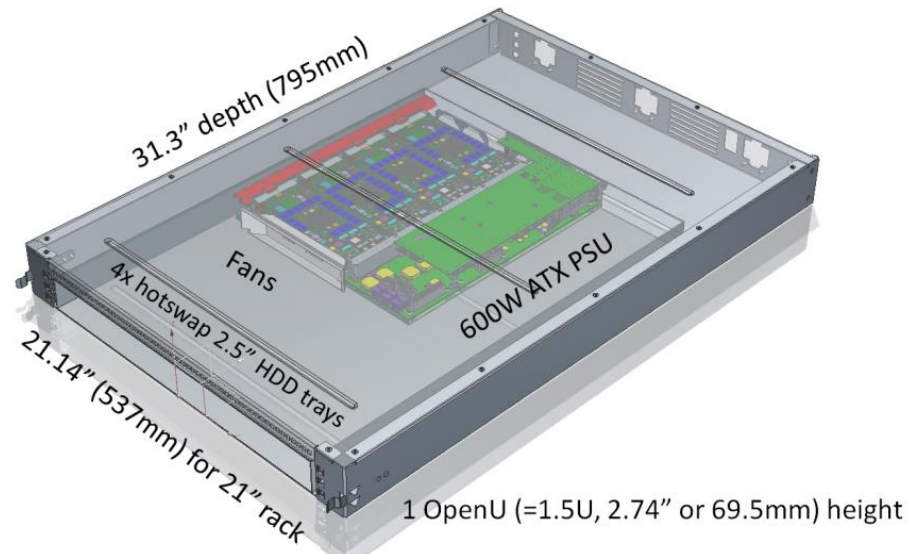
DCCN PCB fitted in 19" rack-mount enclosure & OCP 21" shelf



PCB only



- 2x RJ-45 (1GbE)
- 4x SFP available (10GbE)
- 4x QSFP available (SRIO)

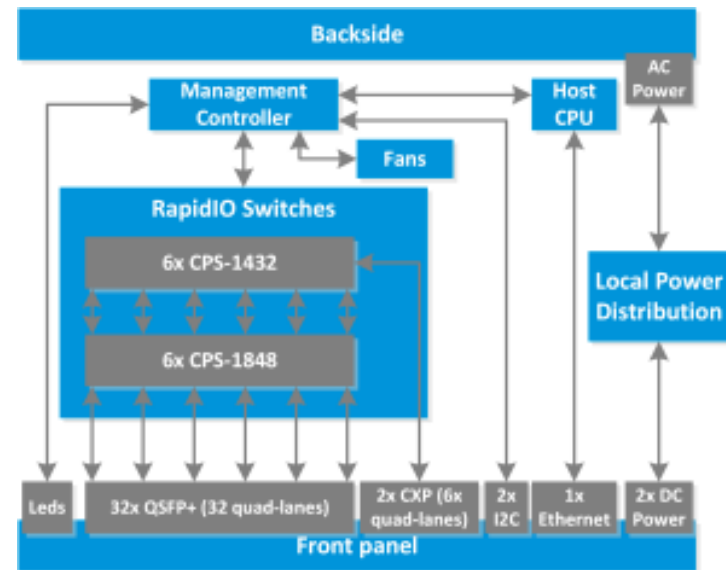


38x 20 Gbps Low Latency ToR Switching



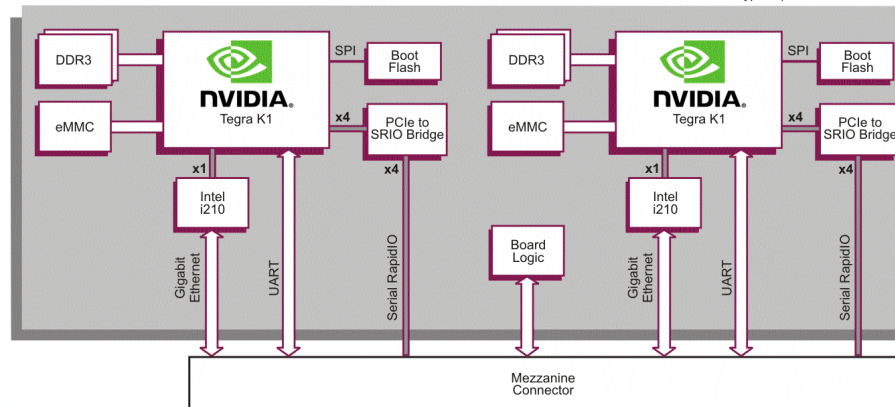
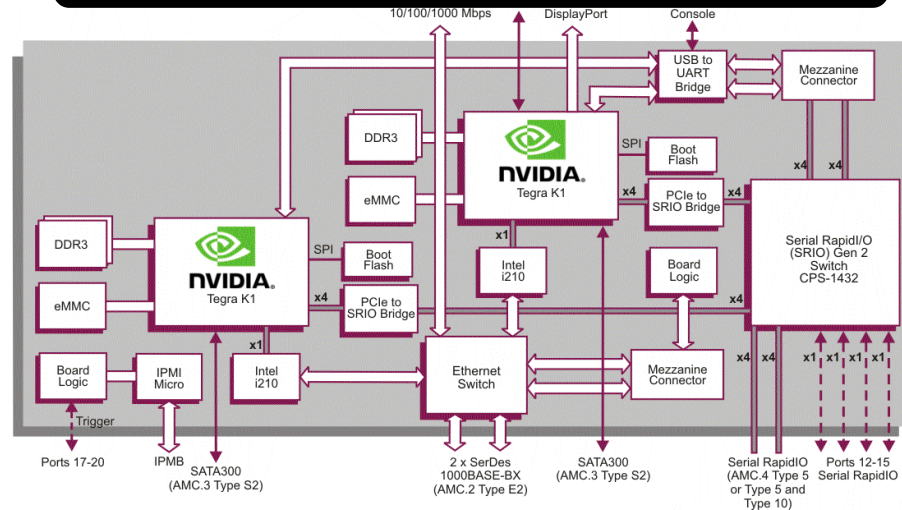
Making your solutions more competitive

- Switching at board, chassis, rack and top of rack level with Scalability to 64K nodes, roadmap to 4 billion
- 760Gbps full-duplex bandwidth with 100ns – 300ns typical latency
- 32x (QSFP+ front) RapidIO 20Gbps full-duplex ports downlink
- 6x (CXP front) RapidIO 20Gbps ports downlink
- 2x (RJ.5 front) management I²C
- 1x (RJ.5 front) 10/100/1000 BASE-T



RapidIO with Mobile GPU Compute Node

- 4 x Tegra K1 GPU
- RapidIO network 140 Gbps embedded RapidIO switching
- 4x PCIe2 to RapidIO NIC silicon
- 384 Gigaflops per GPU
- >1.5 Tflops per AMC
- 12 Teraflops per 1U
- 0.5 Petaflops per rack



Project Caldey Island Mezzanine Block Diagram rev 01

RapidIO 10xN in Development

3rd Generation

Scalable embedded

peer to peer

Multi processor

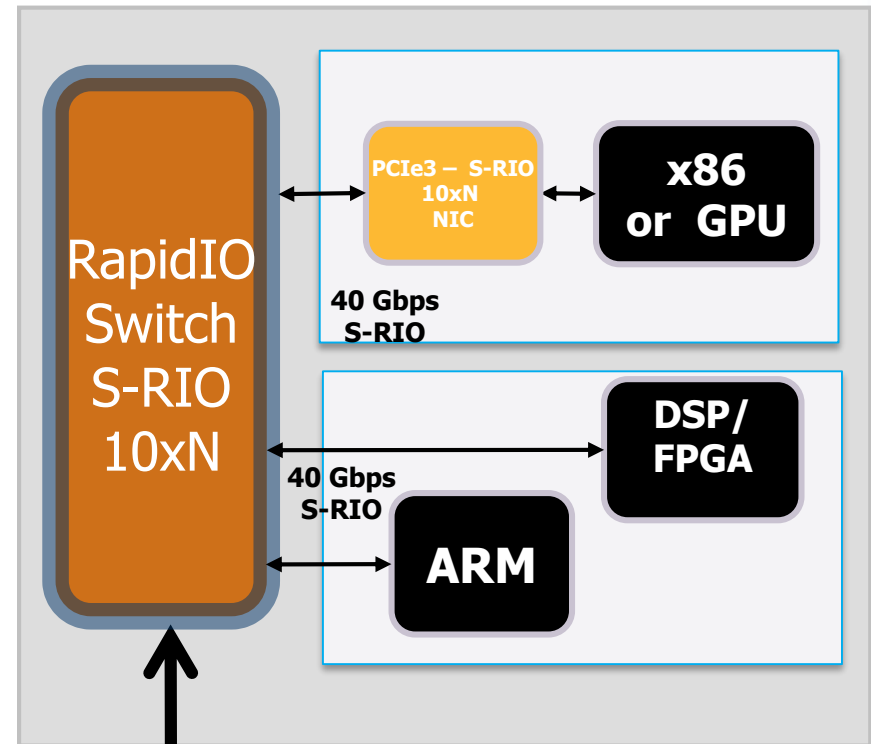
Interconnect

On board, board-to-board and

Chassis to Chassis



- S-RIO 10xN: data rate of 40-160 Gbps per port
- 100 Gbps in definition
- 10.3125 Gbaud per serial lane with option of going to 12.5 Gbaud in future
- Long-reach support (100 cm through two connectors), Short Reach 20 cm 1 connector, 30 cm no connector
- Backward compatibility with RapidIO Gen2 switches (5 & 6.25 Gbps) and endpoints
- Lane widths of x1, x2, x4, x8, x16
- Speed granularity from 1, 2, 4, 5, 10, 20, 40 Gbps



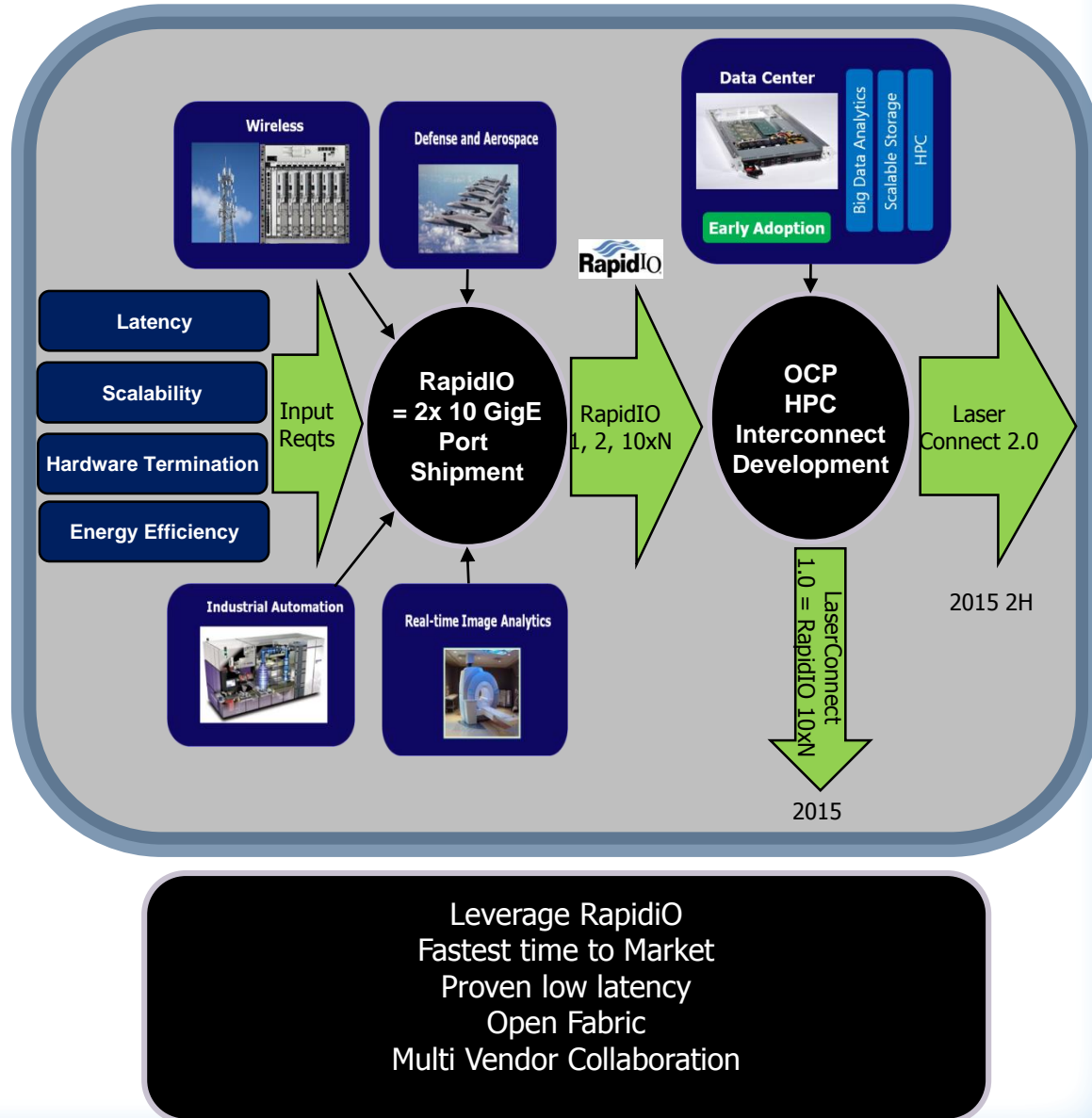
40 Gbps
S-RIO

Key Additional Features

- 10 Gbps per lane
- 10 to 160 Gbps per port
- Power management
- Time Distribution

Getting to OCP Open Interconnect Fabric

- “LaserConnect 1.0” can be based on RapidIO 10xN
 - Variety of SoC, ARM, Switches and other silicon already in development
 - Open Ecosystem, no vendor lock in
 - Faster time to market for an OCP Open Fabric
- “LaserConnect 2.0 ” builds on LaserConnect 1.0
 - In corporate additional needs of HPC and low latency storage
 - Increase top line speeds leveraging industry standard SerDes
 - Keep standard Open



Summary OCP HPC Open Fabric: LaserConnect

- 100% 4G market share, all 4G calls worldwide go through RapidIO switches
- 2x market size of 10 GbE (60 million ports S-RIO)
- 20 Gbps per port in production, 40 Gbps per port silicon in development now
- Low 100 ns latency, scalable, energy efficient interconnect
- Wide semiconductor and wireless ecosystem to accelerate LaserConnect 1.0
- Server Platforms developed for HPC and analytics market in Open community
- Software support in Linux community



**Rack Scale Fabric
For any to any compute**

Fast time to market
OCP HPC Open Fabric path