Lowest Latency Scalable Switching Fabric in Data Center Computing and Networking

Mohammad Akhter
Principal Architect
mohammad.akhter@idt.com
Integrated Device Technology Inc.

January 29, 2014

RapidIO Trade Association and Working Group

http://www.rapidio.org/home

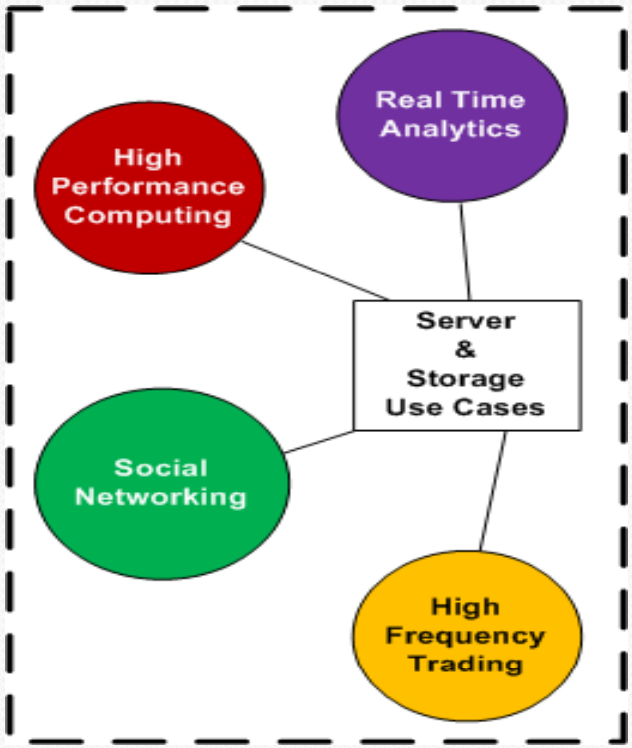
Data Center Computing and Networking (DCCN)
Technical Working Group

Silicon
Software
Systems
Partner Companies

Server
Storage
Networking
OCP

DCCN Example Use Cases

Data Center Workload = *f*(Use Cases)



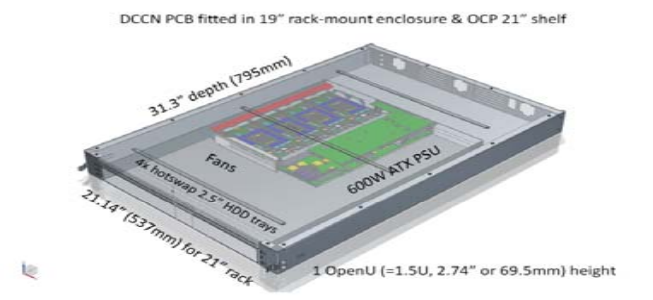
1. Time constrained Processing
2. Diverse Workload
3. Ever-changing Traffic
4. Large Data set
5. Structured/Unstructured Data

DCCN Plan and Target

Phase 1

- Specification Development
- DCCN Hardware Platform

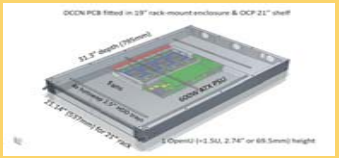
~6 Months



Phase 2

- System Software/Apps
- Submission to OCP
- Hardware Optimization

Server/Storage/Analytics **Apps**



Q3 2013 | Q1 2014 | Q2 2014 | Q3 2014

Leveraging RapidIO Interoperable Eco-system

Industry Leading Solution

Interoperable | Reliable| Best Latency | Fault-tolerant | Lower TCO

RapidIO

TEXAS INSTRUMENTS

DSP: several products
In TCI64xx family

LSI

Axxia Communications
Processor

freescale semiconductor

DSP, PowerQUICC & QorIQ multicore

ALTERA

FPGA: Arria
and Stratix Family

XILINX

FPGA: Virtex 4/5/6
families

Lattice Semiconductor Corporation

FPGA

BROADCOM

XLS416 family
Multicore
Processor

IDT

Switches, Bridges & IP
CPS and Tsi Family

MINDSPEED

Wireless Baseband Processor

octasic

DSP Oct22xx

ARM

HUAWEI

HISILICON

AMD

CAVIUM NETWORKS

Network Processor
Octeon 2 family

applied micro POWERING YOUR C

PowerPC based processors
460GT

Wintegra

Network Processor
WinPath3

intel

mobiveil

Overview

Industry standard form factor (Phase 1)

- 19” 1U rack-mount enclosure & OCP 21” shelf

High Performance Compute Agnostic Platform

- Up-to 4 Processing Modules
- x86 + ARM + DSP + FPGA + Power
- Superior FLOPS/Watt density

Work-load optimized Fabric Technology

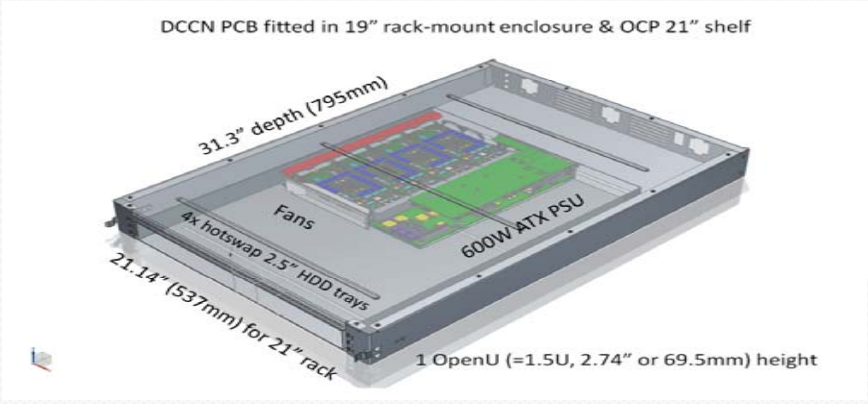
- 20G RapidIO
- 10G Ethernet
- 6G SATA

Power and Cooling

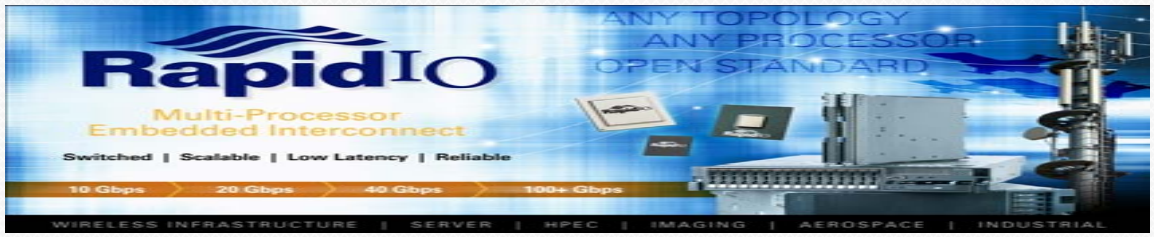
- Back to Front Cooling
- ATX PSU, Up-to 600 W total
- Internal power supply +12V to EATX motherboard

Familiar Software Model

- Service and management console access (TCP/IP)
- Linux Software Stack (TCP/IP)
- Drivers and Data-path Library



RapidIO: *big-data* Fabric Technology



Best-in-class Switch Devices reduces TCO

- Superior performance/watt/$ switch fabric silicon
- 30G/Watt (10W Typical 48 lane switching)

Lowest Latency Fault-tolerant

- 100 nsec switching latency
- Guarantee of delivery and hardware reliability
- High Performance Messaging and DMA

Scalable and Modular Fabric Technology

- Scales to 1000s of nodes
- Layered Protocol Architecture
- Supports copper and optical interconnects

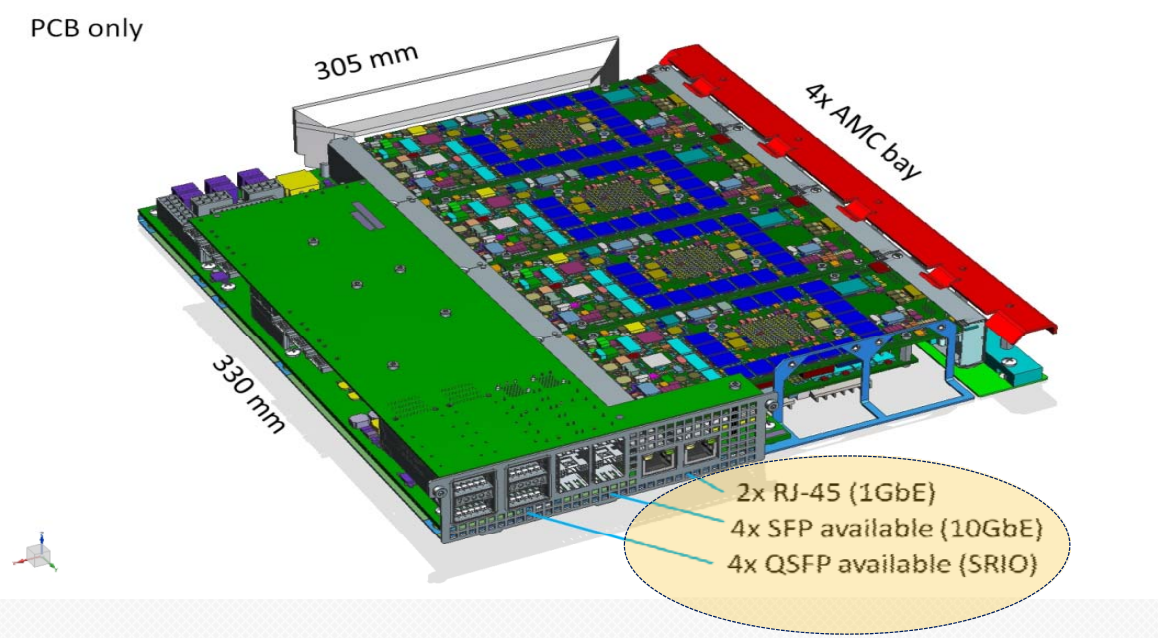
High Speed Interconnects

- Integrated NIC for ARM/PowerPC (20Gb/s)
- Low power NIC for x86 – (PCIe-RapidIO 16Gb/s)

Fabric and I/O

Work-load optimized Fabric Technology

- Computing: 4x QSFP 20G RapidIO
- Networking: 4x SFP 10G Ethernet & 2x 1G Ethernet
- Storage: 6G SATA



Computing I/O

S-RIO Low latency Switching Fabric

- Up-to 300 Gb/s on-board switching
- < 100 nsec switching latency
- Low Latency Messaging and remote DMA
- 4x QSFP (S-RIO) external connections

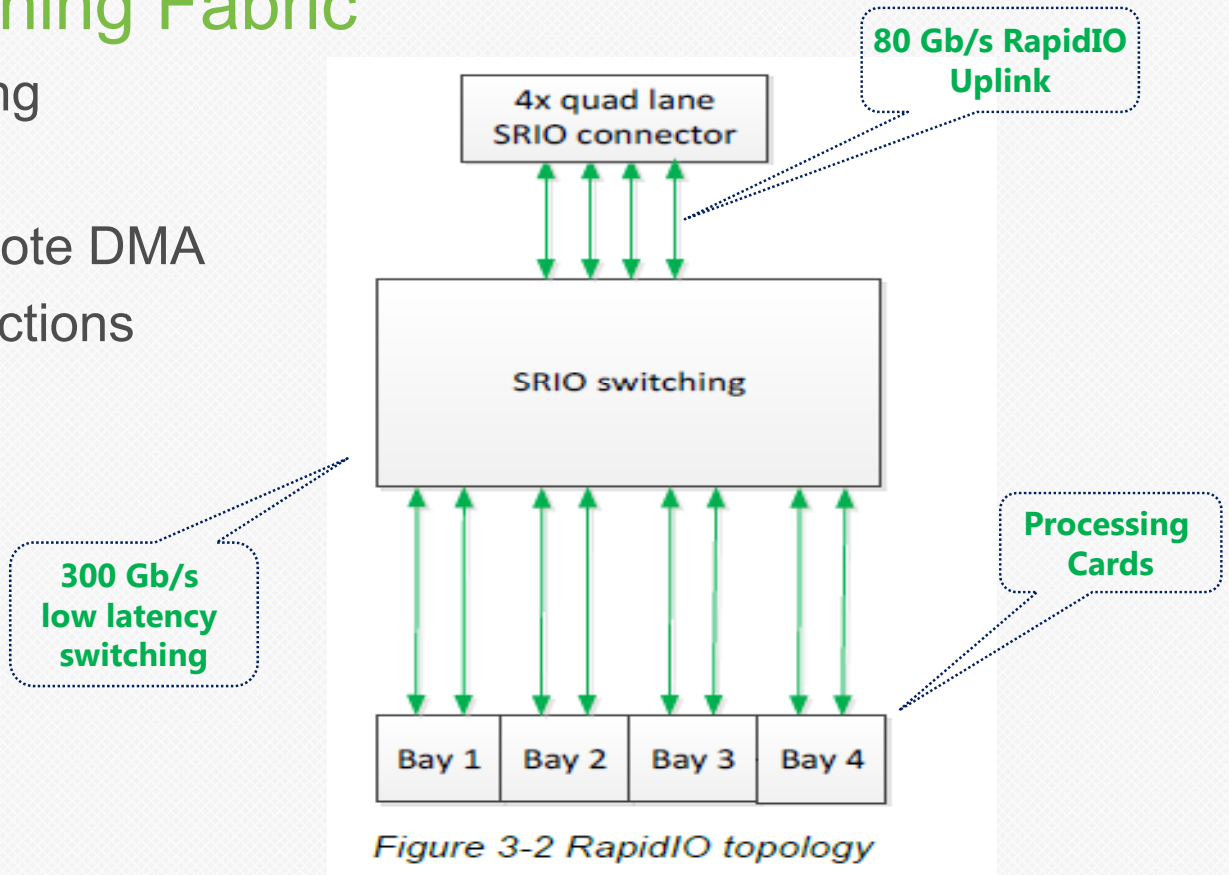


Figure 3-2 RapidIO topology

Networking I/O

In-chassis Storage Fabric

- 4x SFP (10GbE) external connections
- 2x RJ-45 (1GbE)
- TCP/IP and management

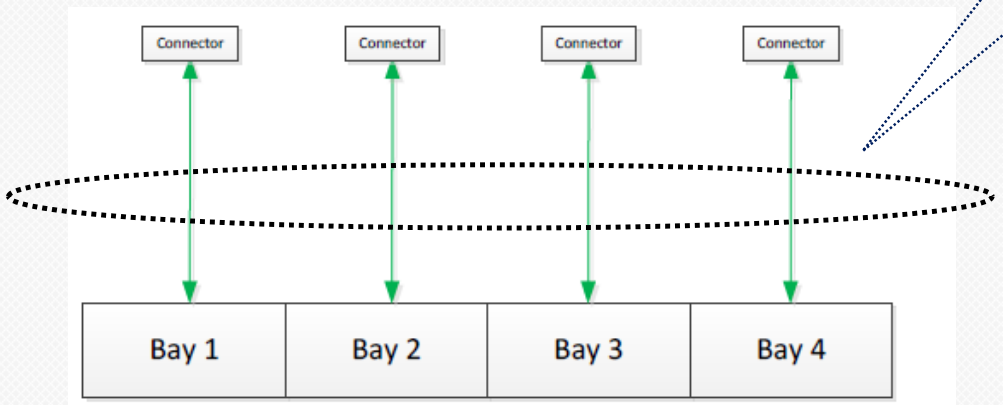


Figure 3-4 10GbE interconnections

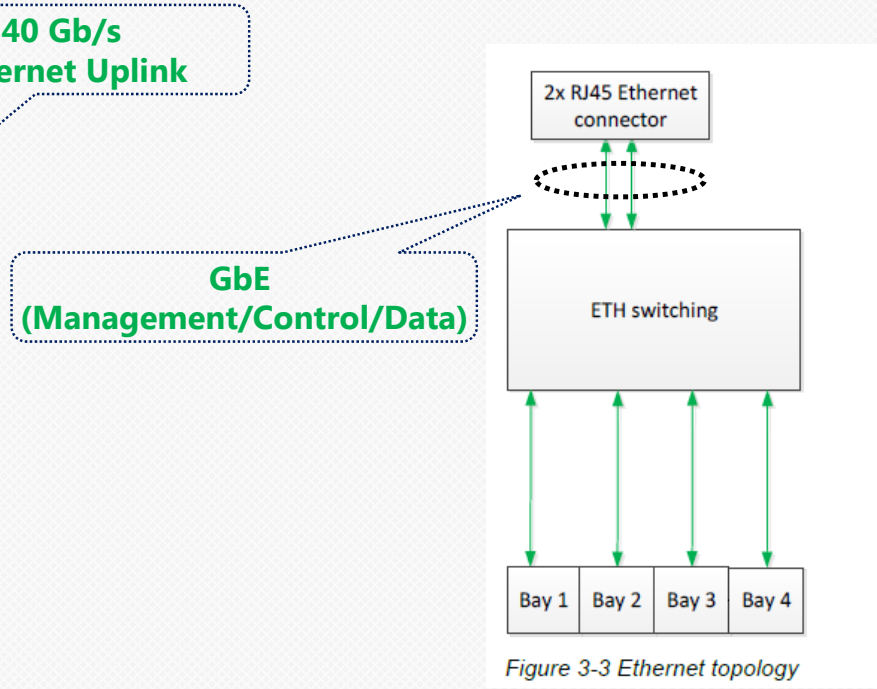


Figure 3-3 Ethernet topology

Storage I/O

In-chassis Storage Fabric

- 8 SATA 3.1 Connections (12G per Module interface)
- Up-to 8 hot-swap 2.5” 6G SATA Drives

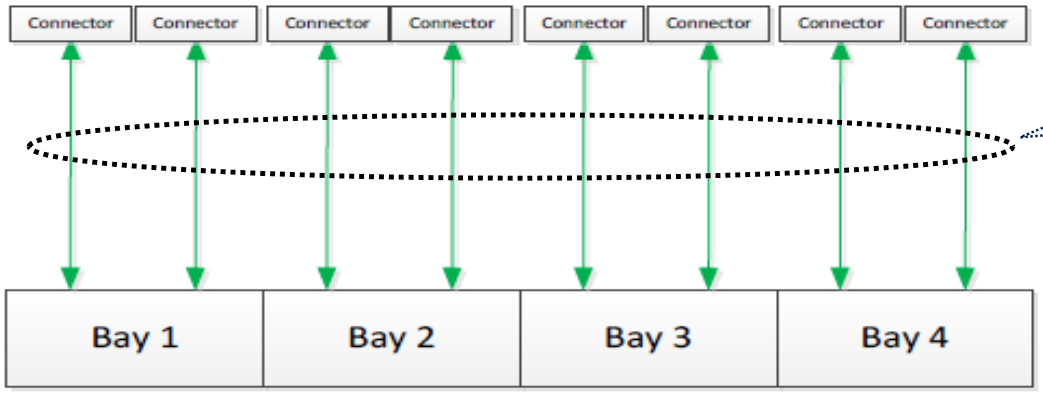
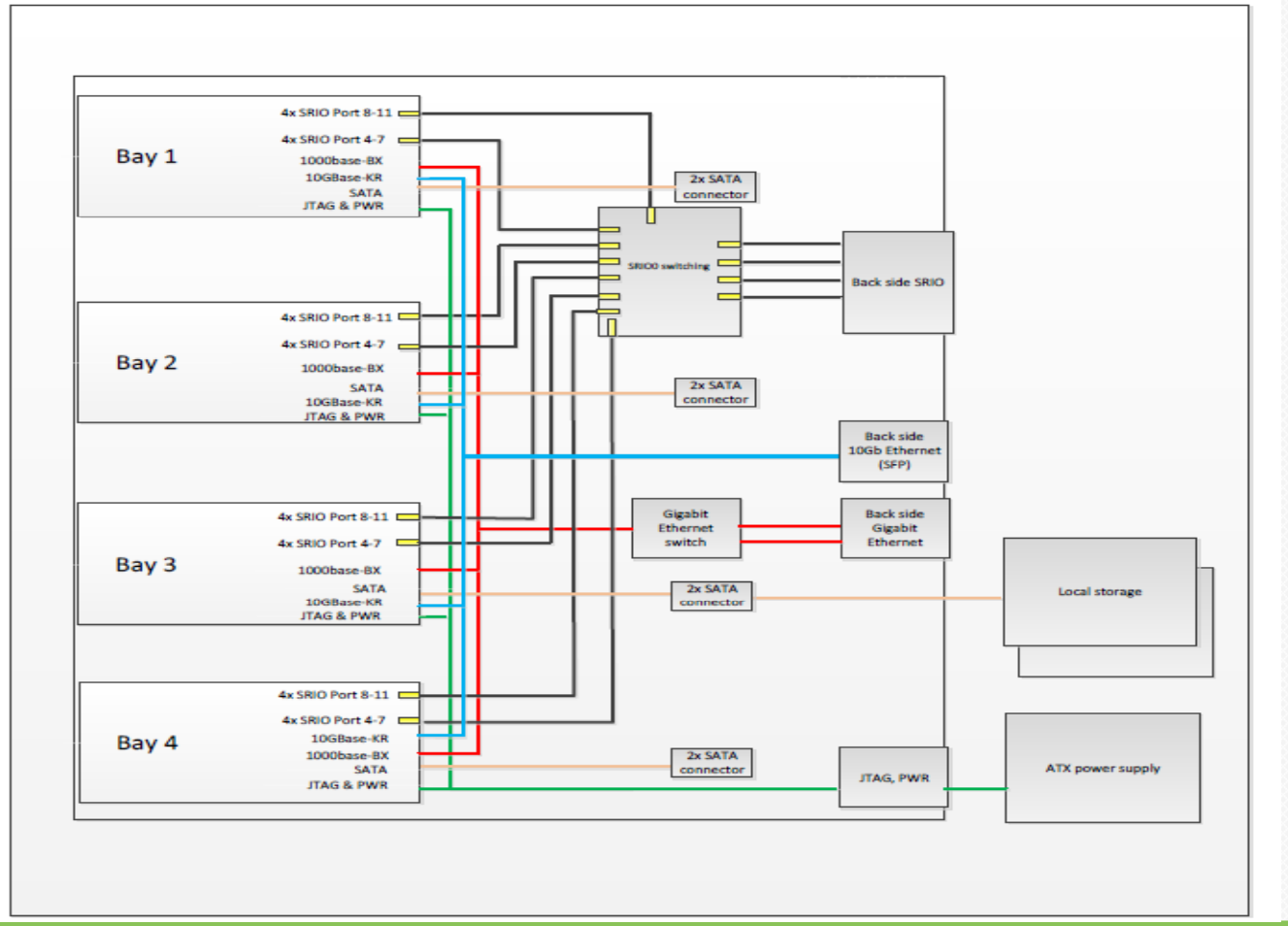


Figure 3-5 SATA storage solutions interconnections

Fabric and I/O Connectivity



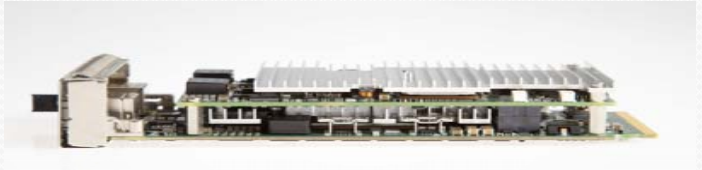
Heterogeneous Computing

Proven Interoperable hardware and Industry support (Phase 1)

- Leverage industry standard AMC form-factor
- Optimize computation based on the workload

Prodrive AMC TI ARM + DSP

- ARM + DSP Processing
- Superior performance/watt Computing



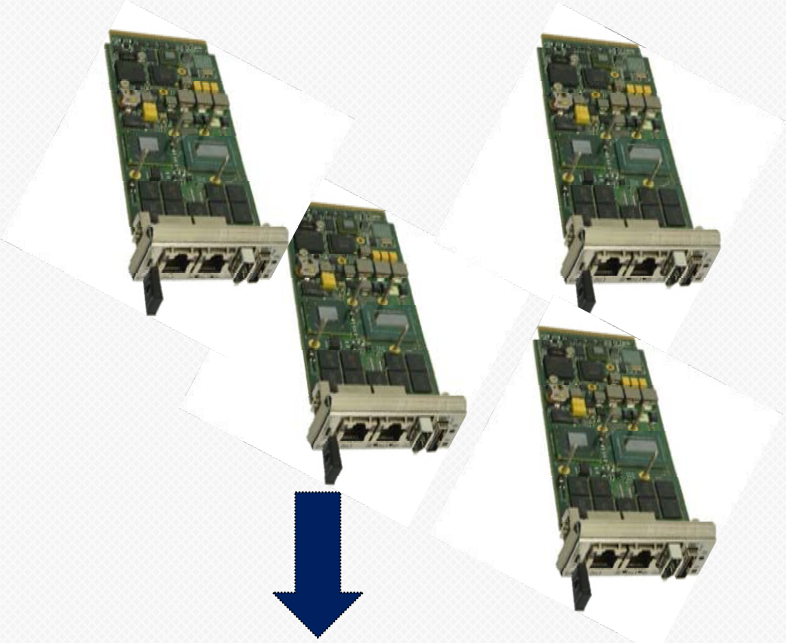
NAT AMC Freescale QorIQ P4080 + Xilinx V6

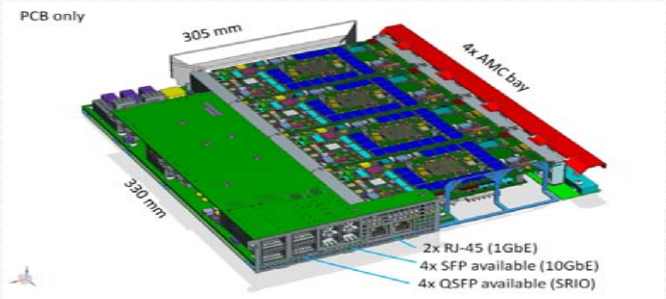
- FPGA + DSP Acceleration
- High performance storage/compute

Concurrent AMC Intel CPU

- x86 Computation and Eco-system
- High Performance Analytics and Storage Solution

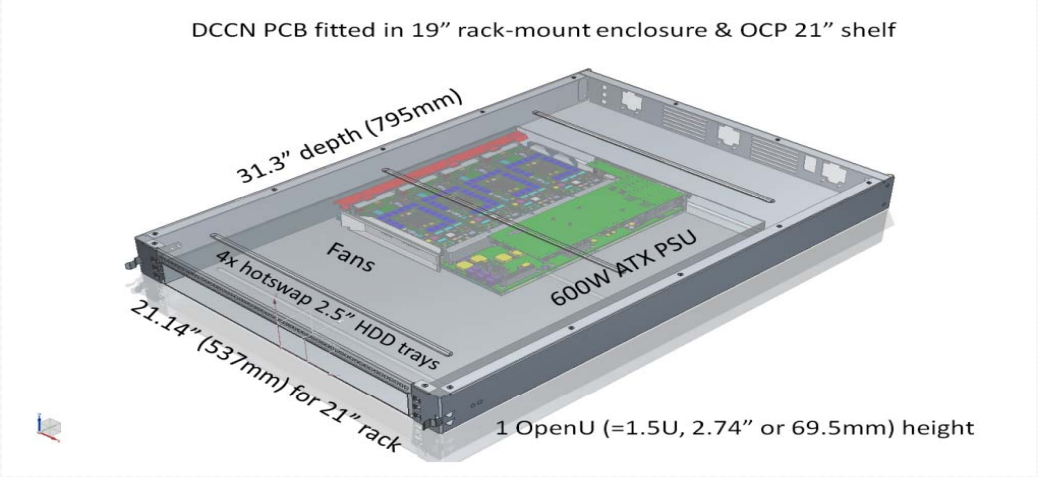
Phase 1 DCCN Example Configuration





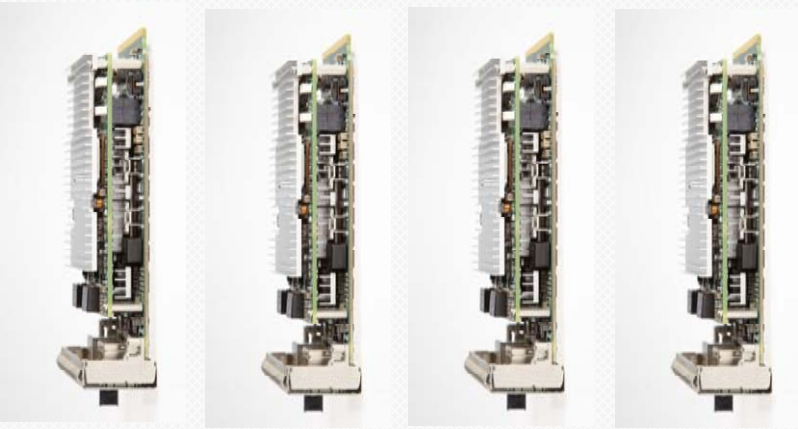
Phase 1 Example Configuration (x86)

- 300 Gb/s RapidIO Switching
- 268.8 GFlops/1U*
  - 4-core Intel i7-3612QE 2.1 GHz Processor
- 64 GB of DDR3-1600 memory
  - 667 MHz 1333 MT/s



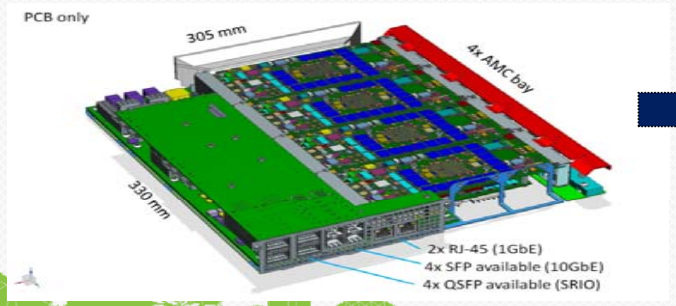
*core_i7-3600_m.pdf – Intel Mobile Processor Series

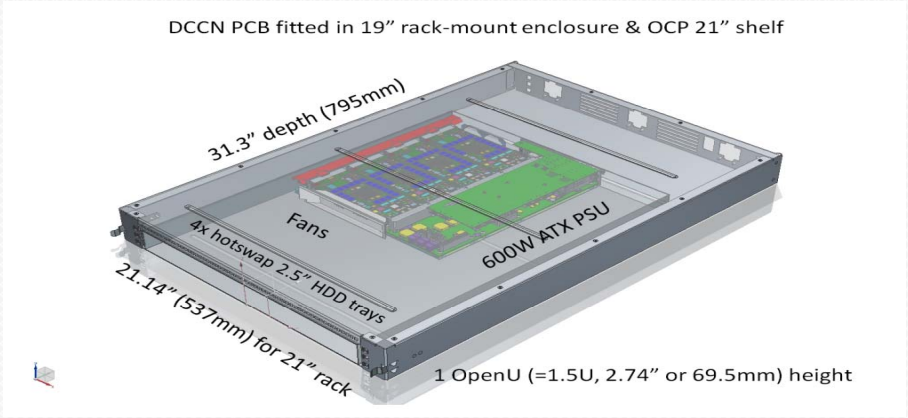
Phase 1 DCCN Example Configuration



Phase 1 Example Configuration (ARM+DSP)

- 300 Gb/s RapidIO Switching
- 1.84 TFLOPS DSP/1U
  - 96 DSP TMS320C66x Cores, 19.2 Gflops/Core (1.2 GHz)
- 16 ARM Cores ARM
  - Cortex-A15 Cores – 1.4 GHz
- 104 GB of DDR3 memory
  - 667 MHz 1333 MT/s





Chassis Maintenance/Boot-up/Monitoring

Maintenance

- Maintenance and management using TCP/IP over Ethernet
- One IP address fixed for management
- Status LED

Monitoring

- Current and Voltage monitoring per module
- Temperature sensors in the OMS (Baseboard)

Booting

- Linux booting from one of the on-board SSDs
- Supports S-RIO-SATA bridging using one of the Compute Modules
- TFTP, NFS, DHCP etc. services over RapidIO

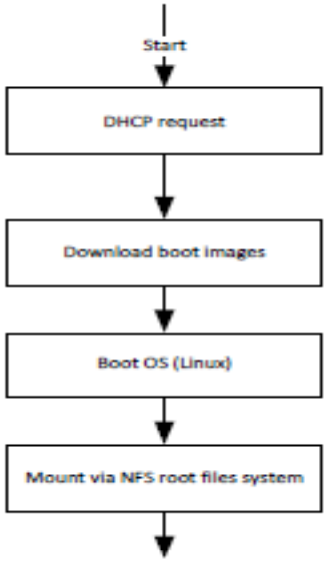


Figure 6-1 Boot sequencing diagram

Phase 1 DCCN Platform