RapidIO for Low Latency Servers and Wireless Base Station

Devashish Paul
Senior Product Manager IDT
Chairman RapidIO Trade Association: Marketing Council

Oct 2013
IDT’s Presentation Agenda

- Server Topologies with RapidIO
- Wireless Network Evolution
- RapidIO Gen2 building blocks in Production

Connecting Wireless Today Converging with Server tomorrow
• Over 6 million RapidIO switches shipped
• Over 30 million 10-20 Gbps ports shipped > Ethernet (10GbE)
  • 100% 4G interconnect market share
  • 60% 3G, 100% China 3G market share

Today: 6.25Gbps/lane - 20 Gbps/port embedded RapidIO i/f on processors, DSPs, FPGA and ASICs.
• 10Gbps/lane in development (10xN)
• 25Gbps/lane (25xN) next generation
• No NIC needed with embedded fabric interface
• Hardware termination at PHY layer: 3 layer protocol
• Lowest Latency Interconnect ~ 100 ns
• Inherently scales to large system with 1000’s of nodes
Wireless topologies: moving to Data Center and HPC

Computing Topologies: 20 - 100 Gbps
- Micro Server
- Super computers
- Blade server
- Storage

Embedded Topologies
Up to 20 Gbps
- Wireless Base Station
- Video Conferencing
- Imaging
- Mil / Aero

Wireless, Video, Military, Imaging
Features

Compute = 6.4 Gflops/ W
Switching = 2.2 Tbps/ Chassis

- “Green 500” #1: 3.2 GFlops/Watt (June 2013)

Successful Industry Collaboration

- Open Standard: Interoperable
• Today: 80% North-South traffic in bulk server

• Compute and Analytics = East West “any node to any node traffic”

• Desire for high bandwidth fat pipe interconnect fabric inside the box (already done in wireless)

• Reduce energy, power and latency associated with NIC’s and adaptors

100 ns 1 hop switching
What Is/Was a Server

**Yesterday**
- Workstation (Processor) + LAN Connection (Ethernet)
- Scale with cabling and Ethernet / IB Switching Box

**Today**
- Multiple Processors + NIC + Ethernet
- Scale across backplane (Blade)
- Large scale out with Ethernet / IB top of rack and cabling
- Inter-rack latency: 100+ us
- Scale: 10s of racks

• 20 years, no change in architecture
• We just took LAN architecture and cramped them down
• More processing in smaller form factors, but not great
Dell: Pushing Ethernet Outside of Rack

Evolution of the converged infrastructure fabric
Future converged infrastructure platform will remove the barriers between the separate internal fabrics

TODAY’s separate fabrics in CI system

TOMORROW’s rack-based fabric in CI system

Disparate Fabrics

Ethernet In and Outside rack

Embedded Interconnect inside Rack
For Processor to Processor

Unified Interconnect
No NICs

Ethernet is Outside only

Rack-level Fabric
A single switching function
RapidIO vs. Ethernet for Processor Access

**Long Latency Path**

<table>
<thead>
<tr>
<th>Interconnect</th>
<th>Latency (ns)</th>
<th>Board Density factor</th>
<th>Incremental Discrete Devices</th>
<th>Interconnect Power (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet</td>
<td>&gt;1000</td>
<td>10x</td>
<td>2</td>
<td>~10</td>
</tr>
<tr>
<td>RapidIO</td>
<td>&lt;100</td>
<td>1x</td>
<td>0</td>
<td>&gt;0.25</td>
</tr>
</tbody>
</table>

Short Latency Path:
- RapidIO Direct Processor Interconnect
- Form factor 1x
- 20 Gbps RapidIO

This is why Dell is pushing Ethernet out of the Box:
Too many adaptors, too much latency, too much power.
3 Layers Terminated in Hardware vs. TCP Offload

- RapidIO 3 layers are terminated in hardware, offloading the processor of protocol termination tasks.

- The end result is:
  - Reduction in latency
  - Better Throughput
  - Lower Power
  - Frees up processor cycles
  - Higher performances systems

- TCP offload can use multiple cores in an ARM or Xeon class processors when terminating multiple TCP sessions.

RapidIO has zero Protocol Offload penalty. Ethernet TCP Offload can use multiple cores.
## Why RapidIO for Low Latency

### Bandwidth and Latency Summary

<table>
<thead>
<tr>
<th>System Requirement</th>
<th>RapidIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch per-port performance raw data rate</td>
<td>20 Gbps – 40 Gbps</td>
</tr>
<tr>
<td>Switch latency</td>
<td>100 ns</td>
</tr>
<tr>
<td>End to end packet termination</td>
<td>Sub 1 us</td>
</tr>
<tr>
<td>Hardware Failure Recovery</td>
<td>2 us</td>
</tr>
<tr>
<td>NIC Latency (Tsi721 PCIe2 to S-RIO)</td>
<td>300 ns</td>
</tr>
<tr>
<td>Messaging performance</td>
<td>Excellent</td>
</tr>
</tbody>
</table>
Scaling with RapidIO Switching

- Proven architecture from embedded systems
- Scalable to 64k nodes
- Scalable to 100s of racks
- Inter-rack latency < 5us
- 20 Gbps in production bandwidth per link today
- No NIC, low latency path with embedded fabric interface

Server Blade with native RapidIO Processors

RapidIO Switch

RapidIO Backplane Switch

20-80 Gbps Per port

latency ➔ scalability ➔ Low power
- X86 processors lead market for performance in terms of Gflops, but lack easy scale out for clusters

- Performance is key in: High Performance Computing, Server, Imaging, Wireless, Aerospace and other embedded applications

- The same applications need the performance of RapidIO Interconnect for:
  - Peer to peer networks with scalability
  - Lowest system power with protocol terminated in Hardware
  - Lowest end to end packet latency

**PCIe to RapidIO NIC Attributes**
13x13mm, 2W, hard terminated, 20 Gbaud per port, $49

6U Compute Node with 2x Intel I7 in production
Scaling PCIe with RapidIO: Atom based x86 Server

- Easily scale PCIe by using S-RIO switching and PCIe to S-RIO NIC devices
- Small form factor, low power NIC. 13x13 mm with 2 Watts
- Total power for 8 nodes = 23W (typ.)
- NIC latency 600 ns, switch latency 100 ns, superior to Enet and Infiniband NIC based scaling

![Diagram of PCIe and RapidIO connections]

Server Blade with x86 Processor and RapidIO

Low latency, High Density, x86 Computer nodes
48 card/ 96 processing nodes (Two x86 per card):
Power, Cost, Latency

<table>
<thead>
<tr>
<th></th>
<th>RapidIO (x86)</th>
<th>10G Ethernet</th>
<th>PCIe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(compute switch + NIC+ central)</td>
<td>(central +NIC)</td>
<td>(compute node switch + HBA)</td>
</tr>
<tr>
<td>Switch Only Power</td>
<td>0.33W / 10G</td>
<td>1.25W / 10G</td>
<td>1.25W / 10G</td>
</tr>
<tr>
<td>Aggregate Power 96 Nodes</td>
<td>0.384KW</td>
<td>0.424KW</td>
<td>0.834 KW</td>
</tr>
<tr>
<td>(NICs, HBAs, Local Switch, Central)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interconnect Bandwidth per card</td>
<td>160 Gbps</td>
<td>20 Gbps</td>
<td>84 Gbps</td>
</tr>
<tr>
<td>Interconnect Gbps per Watt</td>
<td>20 Gbps/W</td>
<td>2.26 Gbps/W</td>
<td>4.84 Gbps/W</td>
</tr>
<tr>
<td>Cost per node</td>
<td>$39</td>
<td>$257</td>
<td>$122</td>
</tr>
<tr>
<td>(NIC/ HBA + % switching) @ 10k volume (public pricing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ per Gbps</td>
<td>$0.65/ Gbps</td>
<td>$25.62/ Gbps</td>
<td>$2.88/ Gbps</td>
</tr>
<tr>
<td>Latency (End to End, Any to Any)</td>
<td>Sub 1 us</td>
<td>&gt;10-30 us</td>
<td>&gt;5-10 us</td>
</tr>
</tbody>
</table>

Power <25%  Interconnect 4-8x  Cost <75%
Applications: Wireless and Server Convergence
Evolution of Bandwidth

More processing load per base station

GSM

Smart Phone

Tablet

x1

x40

x400

2G → 3G → 4G
Why RapidIO 10xN

Convergence of Wireless and Cloud

More Bandwidth, More Users,

Moore’s Law Does not Scale with Network

We need faster interconnect: RapidIO Embedded 40G

2G  →  3G  →  4G
Base Station with Caching Micro Server Gen2

Base Station

- RapidIO switch
- FPGA enabled with S-RIO
- Processor with S-RIO
- Ethernet to backhaul
- 20 Gbps S-RIO Per port to front panel

Micro Server

- RapidIO switch
- Processor with S-RIO
- Ethernet to backhaul
- 20 Gbps S-RIO Per port to front panel

CPRI

x86 CPU

PCIe2 - S-RIO2 NIC

20 Gbps Per port S-RIO

DSP with S-RIO

FPGA enabled with S-RIO

20 Gbps Per port S-RIO

20 Gbps Per port S-RIO

20 Gbps S-RIO Per port to front panel
Wireless Micro Server Attributes

- Seamless Integration with Base Station via RapidIO front Panel
- Processors can snoop packet up to layer 7 in look aside mode and cache repeated content locally
- Eliminates redundant traffic from backhaul
- Reduces overall system power consumption
- Allows for ease of data intensive traffic to WiFi co located with 4G base stations. This reduces the overall load on licensed spectrum
- Scalable Solutions for and base station. Simply stack more pizza boxes with S-RIO interconnect. Option for adding S-RIO local switch box
- Support Green Initiatives, with overall reduced energy footprint in datacenter with S-RIO
- Superior end to end packet latency, throughput and fault tolerance

Base Station Micro Server

- 20 Gbps S-RIO Per port to front panel

![Diagram of Base Station Micro Server with x86 CPU, PCIe2 - S-RIO2 NIC, and Micro Server connected via RapidIO switch.](image-url)
RapidIO with x86 or ARM Server Architectures

- Supports both x86 and ARM based Architectures
  - Higher performance Servers x86 CPU with RapidIO
  - Data movement - ARM CPU with power efficient low cost direct RapidIO Interconnect

- Best-in-class end-to-end latency
  - Switch latency around 100 ns

- Supports Secured Virtualization

- Supports any kind of topology
  - Start, Mesh, Dual-star, Hypercube, Torus etc..

Superior to PCIe to Infiniband and PCIe to Ethernet NIC options which are Not useable in wireless
• RapidIO connects WCDMA and 4G - LTE base stations
  • Every call, app download, email, & web page
• 20G RapidIO is in production now
• LTE and data usage causing drive for more interconnect bandwidth
• Moore’s Law cannot keep up. Need more processors in systems.
• Need more intra processor interconnect speed at low latency
• Backplane/Inter chassis scaling: RapidIO @ 40 Gbps
• Baseband Subsystem: @ 20-40 Gbps
• ASICs for WCDMA and CDMA when doing multi mode
Building Blocks with RapidIO Gen2
IDT RapidIO Gen2 Portfolio Highlights

- RapidIO Gen2 supports 1.25, 2.5, 3.125, 5, 6.25 Gbaud
- 100 cm^2 connector reach for backplane support
- Backward compatible with 1.3 switches and endpoints
- 240 Gbps highest performance backplane switches in embedded industry
- PCIe2 to S-RIO2 protocol conversion bridge in development to expand RapidIO ecosystem options
- RapidIO 2 Endpoint IP available
CPS Gen 2 Overview

- Designed to S-RIO v1.3 and 2.1
- Up to 48 lanes - 12x4, 18x2, 18x1
- **Up to Full duplex 240 Gbps non blocking bandwidth**
  - Supports all RapidIO speeds: 1.25, 2.5, 3.125, 5, 6.25 Gbaud
- **Cut through latency 100 ns**
- 40% per 10 Gbps power reduction vs S-RIO 1.3
- **Several switch fabric related patents filed**

**Detailed Features**
- High-Performance SerDes
  - Long reach 100 cm 2 connector with DFE support
  - Transmit Pre Emphasis and receive Equalization
- As low as 300 mW per 10 Gbps of data
- Dynamic ingress and egress buffers management:
  - 40 multicast groups per port
  - Supports cut-through and store & forward
  - Error management extension support
  - Error Log: sequence of events in time
- Packet mirror, trace, filter support
- Receiver and transmitter based flow control
- Per port reset mode robust support for hot swap
- Multicast event control symbol generation input pin

CPS-1848
- 12x20 Gbps
- 18x 10 Gbps
- 18x 5 Gbps
- 29x29 FCBGA

CPS-1432
- 8x20 Gbps
- 14x 10 Gbps
- 14x 5 Gbps
- 25x25 FCBGA

CPS-1616
- 4x20 Gbps
- 8x 10 Gbps
- 16x 5 Gbps
- 21x21 FCBGA
**FEATURES**

- Gen 1 and Gen 2 support
  - PCIe v2.1
  - S-RIO v2.1
- PCIe to S-RIO bridging
  - Non-transparent for transaction mapping
  - 8 DMA and messaging engines
- Single port, x1/x2 or x4 @ 1.25, 2.5, 3.125, 5 Gbaud
  - S-RIO 1.3 and 2.1 compliant
  - PCI-express 1.1 and 2.1 compliant (End-Point)
  - Can buffer up to 32 S-RIO max size packets
  - Full line rate throughput for 64 byte and > packets
- Low Power ~ 1.5-2 W typical
- Power down unused lanes, when used in x1 or x2
- Lane swap and polarity inversion support
- Reach support: 60 cm over 2 connectors
- S-RIO and PCIe endpoint compatible clocking options 100 MHz, 125 MHz 156.25 MHz
- Forward Bridge (must have microprocessor on PCIe side of the network)
- JTAG 1149.1 and 1149.6
- 13x13mm FCBGA package
- Commercial and industrial variants

Connect PCIe Processors To S-RIO Networks For Superior Performance Over 10 GbE and Infiniband
IDT RapidIO Roadmap

Switch Products

- **CPS-1848**
  - 18x1, 18x2, 12x4

- **CPS-1432**
  - 14x1, 14x2, 8x4

- **CPS-1616**
  - 16x1, 8x2, 4x4

- **SPS-1616**
  - 16x1, 8x2, 4x4

Bridge Products

- **Tsi721**
  - PCIe2 to S-RIO 10xN Bridge

Ecosystem

- **S-RIO Gen2 IP**
- **SRDP2 1848/1616**
- **Commagility Xilinx V6 AMC CPS-1848**
- **Tsi721 PCIe to S-RIO Eval Board**
- **TI Gen 2 DSP EVM AMC CPS-1848**

Specifications

- **S-RIO 10xN IP**
  - 40 G per Port

- **S-RIO 10xN**
  - 10 / 20 / 40 / 80 / 160G Specification

Legend:

- Production
- Development
- Concept
Today 100% of the 4G OEMs use RapidIO for baseband interconnect with over 6 million switches shipped

4G technologies are driving need for interprocessor communication

Wireless and Server worlds are converging

Today RapidIO leads the market with 20 Gbps embedded fabric interconnect

Tier 1 customers worldwide pushing IDT for 40 Gbps

IDT is developing S-RIO 10xN Switches, bridges and endpoint IP @ 40 Gbps per port

Peer to Peer Scalable Interconnect for Wireless,

- 10.3125 Gbaud per lane, 40 Gbps per port embedded interconnect
- 100 ns Latency with 5x effective bandwidth of 10 GigE for embedded systems
- <300mW per 10 Gbps of data
Backup Slides

Devashish Paul  
Senior Product Manager IDT  
Chairman RapidIO Trade Association: Marketing Council

Oct 2013

©2013 Integrated Device Technology, Inc.
RapidIO vs Ethernet Latency Path

**RapidIO Direct Processor Interconnect**

**ETHERNET**
- Long Latency
- NIC Required High Footprint
- Higher Power
- Software Intervention

**RAPIDIO**
- Short Latency
- NO NIC embedded Fabric
- Low Power
- Native Hardware Termination

**Short Latency Path**

**Long Latency Path**

PCI Express (I/O Interconnect)
Computer Node with PCIe to RapidIO NIC and switching

- 13x13 NIC, 25x25 32 Lane Switch
- 2W per NIC, 4W per Switch
- Total interconnect power 8W (typ.)
- $49/NIC, $80/Switch
- Total Interconnect per Compute Node $65
Compare to PCIe and Ethernet 2 Node Compute

Micro Server Building Block with PCIe only

Micro Server Building Block with Ethernet only

Scale out limitation with PCIe, Bandwidth and latency Limits with Ethernet