Comparing Ethernet and RapidIO®

Greg Shippen
System Architect
Freescale Semiconductor

The Embedded Fabric Choice
Agenda

- Interconnect Trends
- Technical Overview
- Comparison
- Applications
- Conclusion
Market Trends

- **Bandwidth**: GB/s
- **Cost**: $$$ OPEX, $$$ NRE, $$$ CAPEX
- **Modularity, Reuse**
- **Connected Devices**
- **Protocols**: CSIX, TDM, SPI4.2, TCP/IP, PCI Express, ATM

© Copyright 2006 RapidIO® Trade Association
Interconnect Trends

1. **1st Generation Point-to-Point**
   - Packet switched
   - PHY: Source-sync differential
   - Lower pin count
   - Example: HT/P-RapidIO

2. **Hierarchical Bus**
   - Bridged Hierarchy
   - Broadcast
   - PHY: Single-ended
   - Example: PCI/PCI-X/SCSI

3. **Shared Bus**
   - Single segment
   - Broadcast
   - PHY: Single-ended
   - Highest pin count
   - Example: VME

4. **2nd Generation Point-to-Point**
   - Packet switched
   - PHY: SERDES differential
   - Lowest pin count
   - ≥ 10 GHz
   - Ex: PCIe, Serial RapidIO, SATA, SAS
**Interconnect Roles**

- Chip-to-chip
- Board-to-Device
- Board-to-board
- Chassis-to-chassis
### Interconnect Clarifying Terminology

<table>
<thead>
<tr>
<th>Serial Interconnect</th>
<th>Switched Interconnect</th>
<th>Switched Fabric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Protocol</td>
<td>Serial Protocol</td>
<td>O and Packets Protocols</td>
</tr>
<tr>
<td>Point-to-point, no System Addressing</td>
<td>System Addressing</td>
<td>Robust Link Layer</td>
</tr>
<tr>
<td><strong>Aurora</strong></td>
<td>I/O Load &amp; Store</td>
<td>Traffic Managed</td>
</tr>
<tr>
<td>Lightweight</td>
<td>Datagram (Packets)</td>
<td>CPU offload</td>
</tr>
</tbody>
</table>

- **Ethernet**
Agenda

• Interconnect Trends
• Technical Overview
• Comparison
• Applications
• Conclusion
Ethernet Overview

- **WAN scale interconnect**
  - Box-to-box, board-to-board, chip-to-chip, backplane
  - Connect thousands to millions of endpoints
  - Physical layer defined for LAN-scale interconnection
    - Closet to computer
    - 100+ m distance

- **Target market**
  - Initially used in aggregation settings
    - High performance switches, routers and LAN backbones
  - Now WAN to workstations, PCs and laptops

- **Gigabit Ethernet ubiquitous now**
  - Gigabit Specification standard completed in 1998
    - Gigabit Copper (1000Base-T) in 1999
      - No 10G Copper PHY defined yet
    - Various MAC-to-PHY Standards defined

- **Extensible Layered Specification**
  - Point-to-point packetized architecture
    - Variable packet size
    - High header overhead
    - 46-1500 byte packet L2 PDU
    - Up to 9000 byte jumbo frames
Ethernet Layer 2

Layer 2 Packet Type: 1500 Byte Max Packet PDU

Total = 294 Bytes
(256 Byte PDU)
Ethernet + UDP

**Preamble/SFD**: 8 Bytes

**L2 Header**: 14 Bytes

**IP Header**: 20 Bytes

**UDP Header**: 8 Bytes

**User PDU**: 256 Bytes

**FCS**: 4 Bytes

**IFG**: 12 Bytes

**322 Bytes**: (256 Byte User PDU)

**UDP Packet Type**: 1472 byte User PDU
Ethernet + TCP/IP

TCP/IP Packet Type: 1460 Byte Max User PDU
RapidIO Overview

- Chassis scale interconnect
  - Chip-to-chip, Board-to-board via connector or cabling
  - Physical layer defined for backplane interconnection
    - ~80-100 cm + 2 connectors (Serial)

- Target market
  - Embedded systems
  - Compute, defense, networking & telecom line cards
  - CPU I/O, Line-card aggregation, backplane
  - Serial PHY allowed expansion to data plane
    - Flow control, encapsulation, streams

- Initially a processor interconnect
  - First spec a Motorola & Mercury collaboration

- First revision standard completed in 1999
  - Rolled out with processors, bridges and switches
  - Parallel 8-bit RapidIO @ 500 MHz applied clock
  - 5-6G PHY, 2, 8 and 16x lanes + Virtual Channels nearing completion

- Extensible Architecture
  - Layered architecture

- Point-to-point packetized architecture
  - Low overhead
  - Variable packet size
  - Maximum 256 byte PDU
  - SAR support for 4 K-byte messages
RapidIO Packet Format: SWRITE

<table>
<thead>
<tr>
<th>AckID</th>
<th>Prior</th>
<th>tt</th>
<th>FTYPE (0 1 1 0)</th>
<th>Target ID</th>
<th>Source ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Address

Packet PDU

Early CRC

Packet PDU

CRC

266 Bytes (256 Byte PDU)

SWRITE Packet Type: 256 Byte Max Packet PDU
RapidIO Packet Format: Message

Message Packet Type: 256 Byte Max Packet PDU, 4K w/SAR
Agenda

- Interconnect Trends
- Technical Overview
- Comparison
- Applications
- Conclusion
## Logical Layer Comparison

<table>
<thead>
<tr>
<th>Payload Size (Bytes)</th>
<th>Ethernet</th>
<th>RapidIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 2</td>
<td>Layer 3+</td>
<td>Logical Layer</td>
</tr>
<tr>
<td>46-1500 (802.3)</td>
<td>26-1460 (802.3)</td>
<td>1-256</td>
</tr>
<tr>
<td>46-9192(^a) (Jumbo)</td>
<td>26-9172(^b) (Jumbo)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory-mapped R/W</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>RDMA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Write w/Response</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address Size</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>34, 50, 66-bits</td>
</tr>
<tr>
<td>64-bits (RDMA)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Messaging Support</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>TCP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Datagram Support</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>64KB user payloads(^c)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Globally Shared Memory</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deadlock Avoidance</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>L3+ must address</td>
</tr>
</tbody>
</table>

\(^a\)Largest common jumbo frame size  \(^b\)Assumes 20-byte IP Header  \(^c\)Dataplane Extensions feature
## Transport Layer Comparison

<table>
<thead>
<tr>
<th></th>
<th>Ethernet</th>
<th>RapidIO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Layer 2</td>
<td>Layer 3+</td>
</tr>
<tr>
<td><strong>Topologies</strong></td>
<td>Any</td>
<td>Any</td>
</tr>
<tr>
<td><strong>Delivery Service</strong></td>
<td>Best Effort</td>
<td>Guaranteed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(TCP, SCTP, others)</td>
</tr>
<tr>
<td><strong>Routing</strong></td>
<td>MAC Address</td>
<td>IP Address</td>
</tr>
<tr>
<td><strong>Maximum Endpoints</strong></td>
<td>$2^{48}$</td>
<td>$2^{32}$ (IPv4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2^{128}$ (IPv6)</td>
</tr>
<tr>
<td><strong>Header Fields which change link-to-link</strong></td>
<td>None</td>
<td>TTL, MAC Addr, FCS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Redundant Link Support</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

$^a$Data Streaming Spec
## Physical Layer Comparison

<table>
<thead>
<tr>
<th></th>
<th>Gigabit Ethernet</th>
<th>RapidIO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1000Base-T</td>
<td>LP-LVDS</td>
</tr>
<tr>
<td></td>
<td>SERDES&lt;sup&gt;a&lt;/sup&gt;</td>
<td>LP-Serial</td>
</tr>
<tr>
<td><strong>Channel</strong></td>
<td>100m Cat5 Cable</td>
<td>~50cm FR4</td>
</tr>
<tr>
<td></td>
<td>~50cm FR4</td>
<td>~50-80cm FR4 + 2 connectors</td>
</tr>
<tr>
<td><strong>Data rate (Per pair)</strong></td>
<td>1 Gbps</td>
<td>1 Gbps</td>
</tr>
<tr>
<td></td>
<td>500-2000 Mbps</td>
<td>1, 2, 2.5 Gbps</td>
</tr>
<tr>
<td><strong>Symbol rate (Per pair)</strong></td>
<td>125 Mbaud</td>
<td>250-1000 Mbaud</td>
</tr>
<tr>
<td></td>
<td>1.25 Gbaud</td>
<td>1.25, 2.5, 3.125 Gbaud</td>
</tr>
<tr>
<td><strong>Encoding</strong></td>
<td>8b → 4 quinary symbols</td>
<td>DDR</td>
</tr>
<tr>
<td></td>
<td>8b → 10b symbols</td>
<td>8b → 10b symbols</td>
</tr>
<tr>
<td><strong>Signaling</strong></td>
<td>5 Layer PAM Code</td>
<td>NRZ</td>
</tr>
<tr>
<td></td>
<td>(Multilevel Signaling)</td>
<td>NRZ</td>
</tr>
<tr>
<td></td>
<td>NRZ</td>
<td>NRZ</td>
</tr>
<tr>
<td><strong>Signal Pairs (Per Direction)</strong></td>
<td>4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10, 19</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1, 2, 4, 8, 16</td>
</tr>
<tr>
<td><strong>Electricals</strong></td>
<td>Custom</td>
<td>XAUI</td>
</tr>
<tr>
<td></td>
<td>XAUI</td>
<td>LVDS (w/Long &amp; Short Reach)</td>
</tr>
<tr>
<td><strong>Clocking</strong></td>
<td>Embedded</td>
<td>Source Sync.</td>
</tr>
<tr>
<td></td>
<td>Embedded</td>
<td>Embedded</td>
</tr>
</tbody>
</table>

<sup>a</sup>Defacto standard. Some using 1000Base-CX electricals over backplanes  
<sup>b</sup>Each pair carries both Tx and Rx
Protocol Efficiency

- RapidIO NWrite
- Ethernet L2
- Ethernet UDP

Efficiency vs. PDU Size (Bytes)
Effective Bandwidth

- SRIO 4x 3.125G
- SRIO 4x 2.5G
- SRIO 4x 1.125G
- 10G Ethernet: UDP
- 1G Ethernet: UDP
Quality-of-Service (QoS) Dependencies

QoS depends on proper hooks across the interconnect fabric

- Hierarchical Flow Control
  - Addresses short, medium and long-term congestion events
  - Link and end-to-end

- Ability to define many streams of traffic
  - Often defined as a logical sequence of transactions between two endpoints

- Ability to differentiate classes of traffic among streams

- Ability to reserve and allocate bandwidth to streams and classes

Overall Interconnect Traffic → Streams → Classes
QoS Comparison: Ethernet

- No universal QoS standard
- Some Layer 2+ switches support Priority Tagging (802.1d/q)
  - Eight classes
- Increasing number of routers support MPLS at L3

UDP Packet Type: 1472 byte User PDU
QoS Comparison: RapidIO

- All implementations must support 3 prioritized flows
  - No ordering between flows
  - Allows shared buffer pool across flows
- Switches required to provide some improved service
  - Extent of improvement is implementation dependant
- Dataplane Extensions adds carrier-grade QoS
  - Support for thousands of flows, hundreds of traffic classes
  - End-to-end traffic management
Flow Control Comparison

Ethernet
- Link-to-link flow control
  - PAUSE frames
- L3+ end-to-end flow control
  - ECN, TCP windowing, others

RapidIO
- Link-to-link flow control
- Congestion control
  - XON, XOFF
- Fine-grained end-to-end flow control
- Data Streaming
  - Logical Layer
High-bandwidth interconnects require low CPU overhead usage model
  - Hardware support for logical, transport and link layer
  - Low overhead DMA with QoS support
Ethernet Performance

- Microsecond+ fall through latencies (~100us?)
  - Not just the hardware, data has to traverse the SW stack
- High CPU overhead
  - Rule of thumb appears to be borne out in data for TCP/IP SW overhead
    - 1 Hz of CPU per bit of throughput (per direction)
  - Wire speed achievable with GHz class processors
    - Some CPU will be left but how much depends on
      - Protocol being terminated
      - Offload features of GigE interfaces
  - Too often advanced off-load features cannot be leveraged
    - OS & SW stack support issues
- UDP or MAC/Layer 2 solutions sometimes also use proprietary higher layer protocols
  - Can defeat the value of off-the-shelf standards-based solution
- Error correction at endpoint stacks introduce latency jitter and determinism issues
- Works well for application bandwidth < ~300Mbps
  - Lack of flow control problematic for systems that can't significantly overprovision
RapidIO Performance

ÂLatency
   • Sub-microsecond switch latencies

ÂEnd-to-end latency
   • Lower latency than Ethernet since latency does not include a SW stack

ÂArchitecture
   • RapidIO switches straightforward and orthogonal in architecture
     • Strict peer-to-peer
     • Packet headers architected to reduce logic
     • No need to recalculate CRC
Some Economics

- RapidIO and Ethernet with modest TCP/IP offload have similar underlying silicon costs
  - Aggressive TCP/IP Offload engine larger than a RapidIO endpoints
  - GigE Copper PHY is very large (~20mm^2 in 130nm)

- Leveraging Ethernet volume economics not always a reality
  - L2+ Ethernet switches suitable for aggregation and backplanes are not high volume
    - 16-24 ports, QoS features and SERDES PHYs for backplane
    - 12-16 ports, QoS features for aggregation
  - Terminating TCP/IP demands significant processor overhead
    - Dedicate processor or reduce performance and/or application features
Agenda

• Interconnect Trends
• Technical Overview
• Comparison
• Applications
• Conclusion
Summary

Ethernet widely used in low bandwidth embedded applications
- Undisputed standard for wide area networks (WAN)
  - 10G Ethernet role in backplane is yet to be determined
- Broad endpoint silicon and software support
- Flexible and adaptable software protocol stack
- High overhead & latency
- Significant cost jump for bandwidth above 1Gbps
- No standard Off-load or backplane SERDES PHY

RapidIO has captured DSP and line card aggregation role
- Looking to expand role onto the backplane
- Low overhead protocol supporting both control and data plane
- Superior Quality-of-Service
- Variety of PHY speeds
- Cost competitive against 1G and 10G Ethernet
- Growing ecosystem
# Application Fit By Bandwidth

<table>
<thead>
<tr>
<th>Application</th>
<th>Interconnect Bandwidth &lt; 300 Mbps</th>
<th>300 Mbps &lt; Bandwidth &lt; 1 Gbps</th>
<th>1 Gbps &lt; Bandwidth &lt; 10 Gbps</th>
<th>Bandwidth &gt; 10 Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Plane (Low latency, Reliable)</td>
<td>Stack Latency</td>
<td>Bonding, 10Ge</td>
<td>Bonding, 10Ge</td>
<td>10Ge Bonding</td>
</tr>
<tr>
<td>Data Plane (QoS, Streams, HA)</td>
<td></td>
<td>Bonding, 10Ge</td>
<td>Bonding, 10Ge</td>
<td>10Ge Bonding</td>
</tr>
<tr>
<td><strong>Good Fit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marginal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Poor Fit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Application Fit By Attribute

<table>
<thead>
<tr>
<th>System Requirement</th>
<th>Ethernet</th>
<th>RapidIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>High backplane bandwidth (3i 10 Gbps+)</td>
<td>Requires CPUs</td>
<td></td>
</tr>
<tr>
<td>Low latency</td>
<td>Stack Traversal</td>
<td></td>
</tr>
<tr>
<td>Low CPU overhead</td>
<td>SW Stack</td>
<td></td>
</tr>
<tr>
<td>High Availability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High QoS (low latency jitter, many streams, flow control)</td>
<td>Flow Control</td>
<td></td>
</tr>
<tr>
<td>Distributed computing</td>
<td>High Latency</td>
<td></td>
</tr>
<tr>
<td>5+ endpoints in a network</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Good Fit**
- **Marginal**
- **Poor Fit**

© Copyright 2006 RapidIO® Trade Association
Multiservice Switch

- SONET MultiPHY Line Card:
  - OC-12/STM-4:
    - Optics
    - Optics
    - Optics
    - Optics
    - Optics
    - MPHY Framer
    - Control Processor
    - Switch

- Gigabit Ethernet Line Card:
  - GbE:
    - PHY
    - PHY
    - Packet Processing
    - Control Processor
    - Switch

- Voice Line Card:
  - T1/E1:
    - LIU
    - LIU
    - LIU
    - LIU
    - Control Processor
    - Switch
    - DSP
    - DSP
    - DSP
    - DSP
    - DSP
    - DSP
    - DSP
    - DSP
    - DSP
Mobile Switching Center

- Host Processor
  - Initialization
  - Software Download
  - Configuration
  - Signaling Support
  - Network Management

Host Interface:

- From BSC/RNC
- Ethernet/ATM
- Serial RapidIO Switch
- Aggregator Logic
- SDRAM
- Flash
- DSP
- DSP
- DSP
- DSP
- DSP
- DSP
- TDM Bus
- To PSTN
- Backplane

© Copyright 2006 RapidIO® Trade Association
Enterprise Storage Switch

To Media: Fibre Channel, SCSI

To Servers: (FibreChannel, GbE) or mainframes (ESCON, FICON)
Signal and Image Processing
Conclusion

ÅGigabit Ethernet will serve a limited role as a system interconnect
  ï Works in low performance settings where significant over provisioning is possible

ÅRapidIO will expand its existing role as the standard system fabric
  ï Available now
  ï Efficient protocol supporting both control and data plane
  ï Variety of PHY speeds
  ï Cost competitive underlying economics