Multicore System Performance and Power Optimization through Content-Aware Routing

Copyright HyperTransport Consortium 2008
Guest Member Company

In-Server Content-Aware Routing

Ze’ev Rivlin
Vice President, Marketing

Copyright HyperTransport Consortium 2008
Snapshot on HTX™

How and Why HyperTransport HTX Proves Best Choice for Compute-Intensive Applications

Copyright HyperTransport Consortium 2008
Latency is **The** HPC Market Driver

“Peak computing power is no longer the metric to consider. Feeding processing units with data is the real concern, and flow control dependencies in applications make latency the key performance metric.”

Prof. Jose Duato
Polytechnic University of Valencia, Spain
World renowned HPC expert

“1 mS of Latency Worth $100M in Stock Trading Business Value”
NYSE Executive
AMD Analysts Day, July 26, 2007
HTX™ Delivers Latency Edge

- That PCI-Class Peripheral Interconnects Cannot Deliver
- Via Direct Connect Between CPU and Compute-Intensive Subsystems
  - No Performance Bottlenecks by Intermediate Control Logic
- Leverages HyperTransport’s Array of Low Latency, High Bandwidth Features
- Implemented Using Low Cost PCIe Connector
  - Reverse-Installed to Prevent Wrong Card Insertion
HTX3™ Specification Released
Jun 25, 2008

• 3x the Bandwidth of Original HTX Specification
  • 2.6 GHz Clock Rate – 20.8 GB/s Aggregate Bandwidth
• Added Power Management Features
• Full Backward Compatibility
• Powerful Link-Splitting Capability

Copyright HyperTransport Consortium 2008
No – HTX Complements and Coexists with PCIe to Provide the Capability that PCIe Cannot Deliver
**HTX™ Key Values**

**Enables**
- HPC Product Performance that PCI-Class Interconnects Cannot Deliver
- Easy Integration of System Functionality Too New/Complex/Costly for MB Integration

**Empowers**
- HPC Solution Providers with a Competitive Edge
  - No Risks of Premature MB Integration
  - Shortest Time-to-Market
  - One MB Fits Multiple Markets/Applications
  - Up-Sell Factor
Compute Intensive

• Low Latency + High Bandwidth
• Multi-Processing, Co-Processing

Target Markets

• Database Analytics
• High Traffic Web Services
• Stock Trading Acceleration
• Server Clustering and SMP
• Streaming Media Servers
• Financial Modeling
Unique HTX™ Performance

Aggregate Latency Advantages Compared to PCIe

• 20% Better Physical Layer Latency and Bandwidth due to Absence of 8B/10B Clock Recovery Overhead
  – No SerDes

• 55% Lower Latency Per Transaction due to Absence of Intermediate Control Logic Overhead
  – 95nS of PCIe Gen2’s Estimated Round Trip Penalty out of 170nS Total on Short, Open Page DRAM Reads

• Vastly Leaner Protocol (Packet Payload)
  – 12 Less Bytes of Overhead per Packet Compared to PCIe

• 20nS Better Per-Transaction Latency in Heavy Traffic Environments due to HT’s Priority Request Interleaving™
### Superior Bandwidth

<table>
<thead>
<tr>
<th>Feature</th>
<th>PCIe Gen1</th>
<th>PCIe Gen2</th>
<th>HTX</th>
<th>HTX3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Clock Rate</td>
<td>2.5 GHz</td>
<td>5.0 GHz</td>
<td>800 MHz</td>
<td>2.6 GHz</td>
</tr>
<tr>
<td>Double Data Rate</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Max Bandwidth x Lane</td>
<td>2.5 Gbps</td>
<td>5.0 Gbps</td>
<td>1.6 GT/s (*)</td>
<td>5.2 GT/s (*)</td>
</tr>
<tr>
<td>8B/10B Penalty</td>
<td>-20%</td>
<td>-20%</td>
<td>No Penalty</td>
<td>No Penalty</td>
</tr>
<tr>
<td>Net Bandwidth x Lane</td>
<td>2.0 Gbps</td>
<td>4.0 Gbps</td>
<td>1.6 GT/s (*)</td>
<td>5.2 GT/s (*)</td>
</tr>
<tr>
<td>Net Bandwidth 16-Bit - Aggregate</td>
<td>8 Gbytes/s</td>
<td>16 Gbytes/s</td>
<td>6.4 GBytes/s</td>
<td>20.8 GBytes/s</td>
</tr>
</tbody>
</table>

(*) HyperTransport supports Double Data Rate (DDR), transferring data on both the leading and trailing edge of the clock. Therefore HyperTransport’s bandwidth is more appropriately represented by the term “Transfers/second” than the term “Bits/second.”
Tangible Time-to-Result Savings

Compute-Intensive Tasks Require 100Ks to Billions of Packet Transactions

**HTX3™ Time-to-Result Savings vs. PCIe Gen2**

<table>
<thead>
<tr>
<th>Number of Packets Transferred</th>
<th>100,000 Per Task</th>
<th>1 Million Per Task</th>
<th>1 Billion Per Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bytes per Packet Transferred</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.78 mS</td>
<td>7.8 mS</td>
<td>7.8 Sec</td>
</tr>
<tr>
<td>16</td>
<td>4 mS</td>
<td>40 mS</td>
<td>40 Sec</td>
</tr>
<tr>
<td>256</td>
<td>0.32 Sec</td>
<td>3.20 Sec</td>
<td>53 Min</td>
</tr>
<tr>
<td>512</td>
<td>1.16 Sec</td>
<td>11.62 Sec</td>
<td>3.23 Hrs</td>
</tr>
</tbody>
</table>

The results take into account PCIe’s 20% clock recovery, packet payload and 55% chipset overhead penalties. HTX’s Priority Request Interleaving™, if applicable, will add to HTX’s total time-to-result latency advantage.
Unique HTX™ Performance (cont.)

Example: Celoxica Accelerator Company's Benchmark Results

1.4 uS Latency (*)

<10 uS Latency (*)

(*) Latency Access to Network Data Regardless of Packet Size
Agenda

• Commex Technologies
• Problem & Solution
• Sample Application Benefits with Thunder
• Commex Competitive Landscape
• Summary
Commex Technologies

• Fabless chip company based in Tel Aviv
• Established in 2005
• Backed by Tier 1 Israeli VCs & the RAD* Group
  *20 independent companies >$750M income in ‘07
• Innovative x86 platform technology
• AMD Torrenza partner
Challenge: X86 Platform Traffic Jam

Inefficient Data Flows with current chipsets

Efficient Data Flow Based on Commex In-server Content-Aware Routing
Solution: Commex Thunder™ In-Server Content-Aware Router

Commex **Vulcan** add-in card (powered by Thunder™)

*Patent-pending technology*
Configuring Thunder™

- Thunder solution includes Thunder device driver and CMX Tool (Thunder configuration utility)

- CMX Tool enables Thunder configuration via script of Linux shell command lines
e.g.,
```
>> cmxtool –ce eth2 channel 0 // show channel 0 coalesce values
```

- Configuration script provided by Commex, either for particular app type or tailored to customer needs
Web App: Parallelized TX

Action: Incoming request classified and forwarded to specific core using RX DMA channel, then response is transmitted to Commex parallelized TX mechanism

Benefit: Network traffic processing performance boost, more concurrent sessions per server
Performance Results – SPECweb “Support”

<table>
<thead>
<tr>
<th>2P/Dual core System</th>
<th>SPECweb Results</th>
<th>Power</th>
<th>Perf / Watt</th>
</tr>
</thead>
<tbody>
<tr>
<td>OEM server 1</td>
<td>14240</td>
<td>490</td>
<td>29.06</td>
</tr>
<tr>
<td>OEM server 2 with Commex in-server content-aware routing</td>
<td>15800</td>
<td>300</td>
<td>52.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1U System</th>
<th>SPECweb Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>OEM server 3</td>
<td>9750</td>
</tr>
<tr>
<td>OEM server 4</td>
<td>9760</td>
</tr>
<tr>
<td>OEM server 2 with Commex in-server content-aware routing</td>
<td>15800</td>
</tr>
</tbody>
</table>

Increased perf / watt by 81%

#users/Watt

Increased perf / 1u by 62%

#Concurrent users
**Telecom App: Multicore Scalability**

**Action:** Incoming network data classified based on affinity, format, and/or pattern match & forwarded to specific core using RX DMA channel

**Benefit:** Network traffic processing performance scales with number of cores

![Diagram of multicore scalability](image)

**SCTP: K msgs/sec vs. number of CPU cores**

- 1 core: 65
- 2 cores: 130
- 3 cores: 200
- 4 cores: 265

**North Bridge**

2 x 10/1GbE
QoS App: Differential Services

**Action:** Incoming network data classified based on priority and forwarded to specific core via QoS queue & DMA channel configured for proper level of priority

**Benefit:** Multiple QoS schemes – e.g. simultaneous fixed low latency and high throughput

![Latency vs. Load Graph](image)

2 x 10/1GbE North Bridge
VoD App: Availability

**Action:** Incoming traffic classified based on traffic type/platform load and handled accordingly

**Variations:**
1. Bypass traffic overload to another server
2. Drop/filter irrelevant or “hostile” traffic
3. Send all “hostile” or overload traffic to fixed core
4. Continue processing relevant service traffic within server throughput capability

**Server A – 2P/quad core**
- 8 cores 2.6Ghz
- 1200 KPPS
- 0.54¢/PPS

**Server B – 2P/dual core**
- 4 cores 2.6Ghz
- 850 KPPS
- 0.41¢/PPS

**Server C – Same as B with Thunder™**
- 4 cores 2.6Ghz
- 1500 KPPS
- 0.24¢/PPS

**Reduced TCO by ~40%**
- **Direct HT Host Interface** – Commex is the only company that delivers 10G solution using Direct HyperTransport™ interface utilizing up to 19.2Gb/s (VS actual 16Gbs on PCIe bus)

- **L7 Classification**
  - Pattern match – Identify patterns in header & body of packet
  - L7 format id – micro-parser allows identification and classification of particular packet format (e.g., HTTPS, TCP SYN, SCTP, SIP, …)

- **Content-Aware Bypass** – content-aware filtering of packets

- **Multicore Mode** – Inter-core DMA data transfer

<table>
<thead>
<tr>
<th>Feature</th>
<th>Commex</th>
<th>Vendor #1</th>
<th>Vendor #2</th>
<th>Vendor #3</th>
<th>Vendor #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Interface</td>
<td>Dual port 10/1GbE</td>
<td>Dual port 10/1GbE</td>
<td>Dual port 10/2.5/1GbE</td>
<td>2 x10GbE + 2x 1GbE</td>
<td>2 x10GbE or 4 x 1GbE</td>
</tr>
<tr>
<td>Direct HT Host Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multicore I/F – e.g., DMAs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1-L4 classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L7 classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content-aware bypass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multicore mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Conventional x86 multicore platforms do not efficiently leverage all processing cores for data traffic processing
- Commex Thunder™ - Content-Aware Routing Processor Delivers:
  1. Reduced TCO via more efficient server CPU resource utilization (Multicore Scalability)
  2. Multicore Differential Services
  3. Improved Availability

Commex Thunder™ Available Now for Evaluation
For More on Commex Technology

Visit the HyperTransport Web Portal
www.hypertransport.org and click on Products, HTX Products and on Commex

Visit Commex’s web site: www.commextech.com

Contact Commex Directly: zeevr@commextech.com

Copyright HyperTransport Consortium 2008
For More on HyperTransport HTX™

Visit the HyperTransport Web Portal
www.hypertransport.org
And click on “Technology”, “HTX Slot Connector” and HTX Reference Material” Buttons

Or contact us directly: info@hypertransport.org - 925-968-0220

Copyright HyperTransport Consortium 2008
Thank You!