HyperTransport™ Technology Tutorial

Prof. José Duato
Technical University of Valencia, Spain
Simula Research Laboratory, Oslo, Norway
HyperTransport Technology Consortium

Hot Chips Symposium
August 23, 2009

With us Today and Happy to Address Your Questions

Brian Holden
VP and Chair, Technical Working Group
HyperTransport Technology Consortium

brian.holden@hypertransport.org
408-472-6310

Topics:
HyperTransport Technology

Mario Cavalli
General Manager
HyperTransport Technology Consortium

mario.cavalli@hypertransport.org
925-968-0220

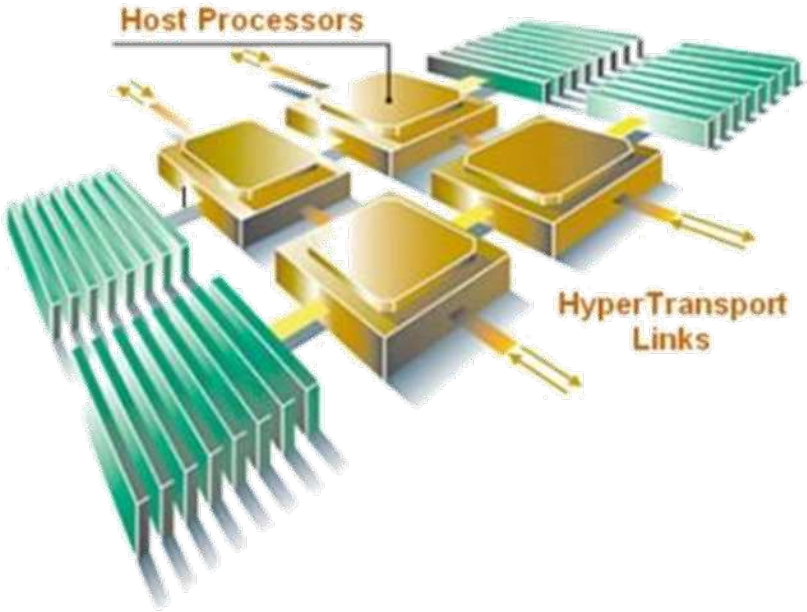
Topics:
HyperTransport Market Positioning
HyperTransport Consortium

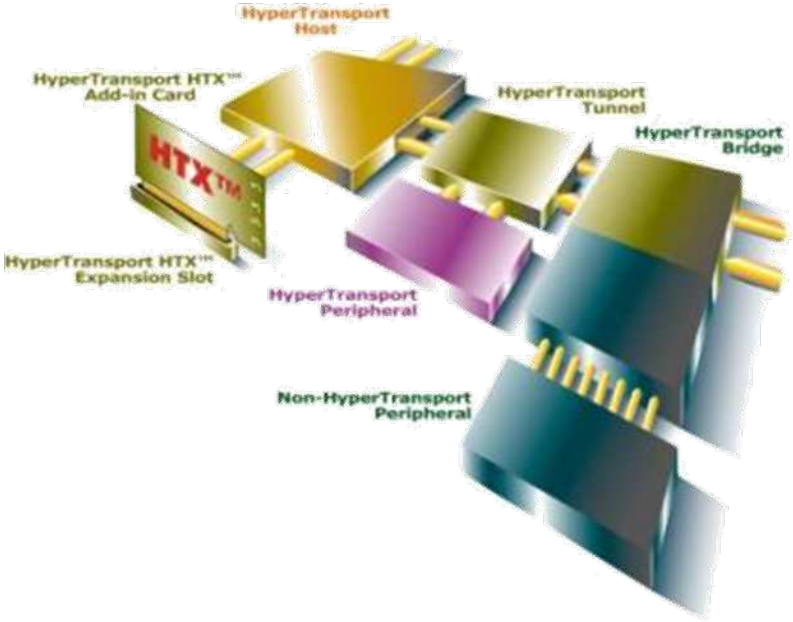
Topics

- Scope and Design Goals
- HyperTransport Defined
  - Host Interface
  - Connecting Device to Host
  - Connecting Multiple Devices to Host
  - Interconnecting Multiple Hosts
- AMD Cache Coherence Support
- Beyond Motherboards
- New in HT3
- Beyond HT3
- Beyond Conventional
- HyperTransport Technology Consortium

Scope and Design Goals

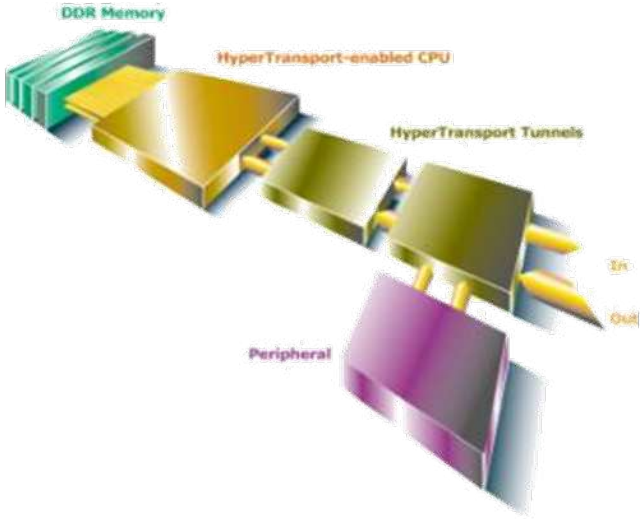
- System Area Network Supporting Cache-Coherent Shared-Memory Multiprocessors and I/O Devices
  - High-Performance Replacement for Processor Front Side Bus (Point-to-Point Links vs. Bus)





Scope and Design Goals (cont.)

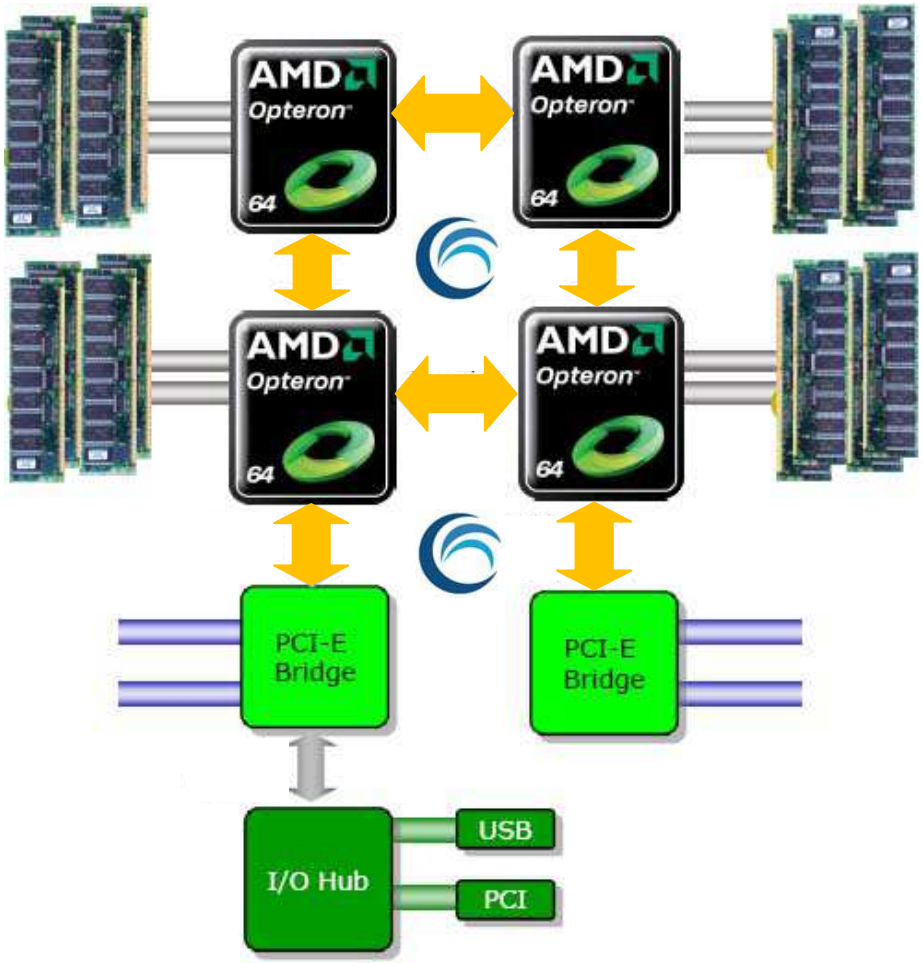
- System Area Network Supports Cache-Coherent Shared-Memory Multiprocessors and I/O Devices
  - High-Performance Replacement for Processor Front Side Bus (Point-to-Point Links vs. Bus)
- HyperTransport's Distinction: Processor-Native
  - Integrated in Processor Architectures



Scope and Design Goals (cont.)

- Lowest-Latency, High Bandwidth, Cost-Effective, Reliable Motherboard-Level Interconnect
  - SMP Programming Model
- Unified Interface For Local and Remote Memory
- Self-Configuring Topology and Link Speed
- HT3 Enhancements
  - Increased Bandwidth and Reliability
  - Link Splitting
  - Dynamic Power Management
  - AC Mode
  - Hot Plugging

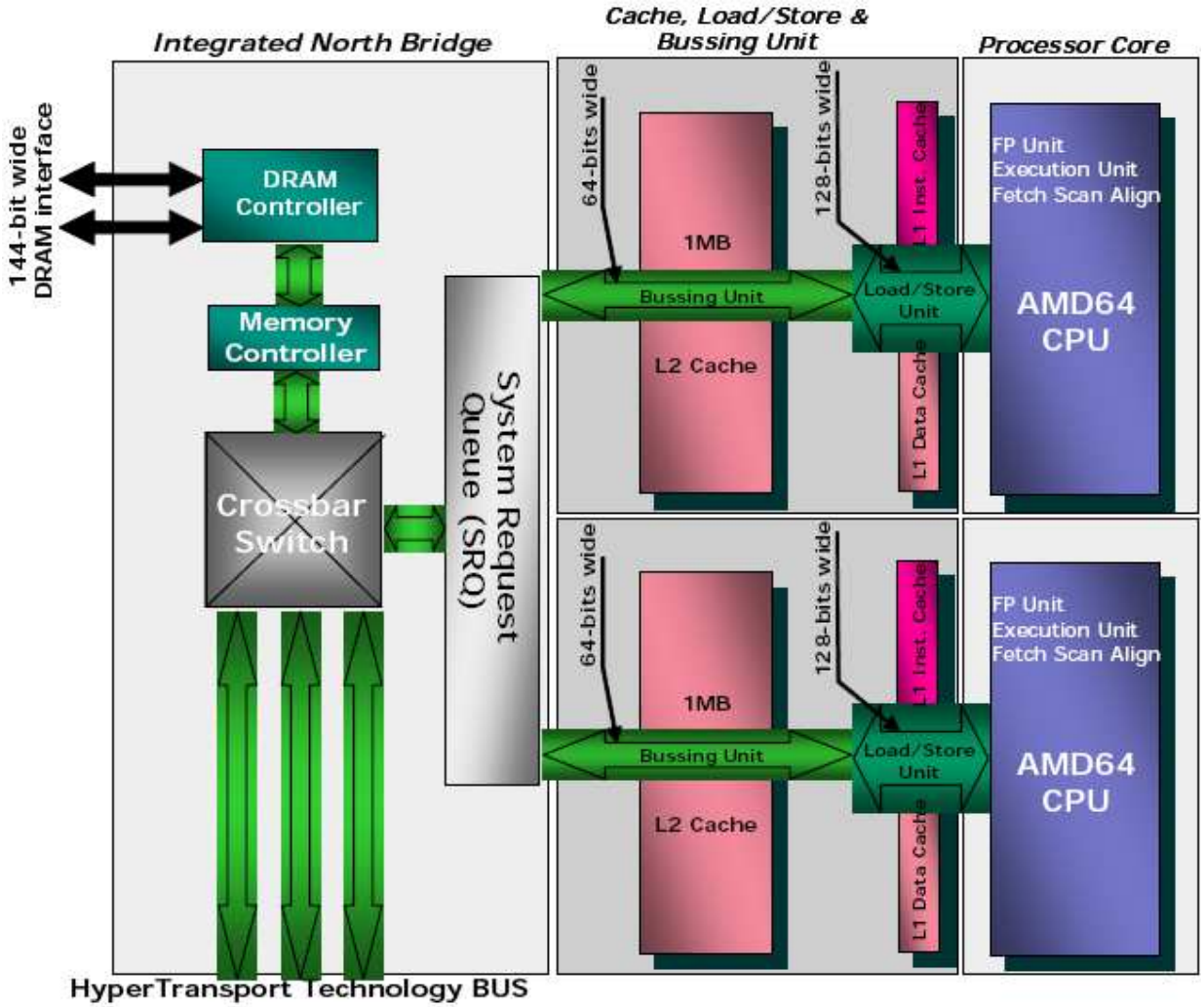
Typical Server Architecture



HyperTransport™ Defined

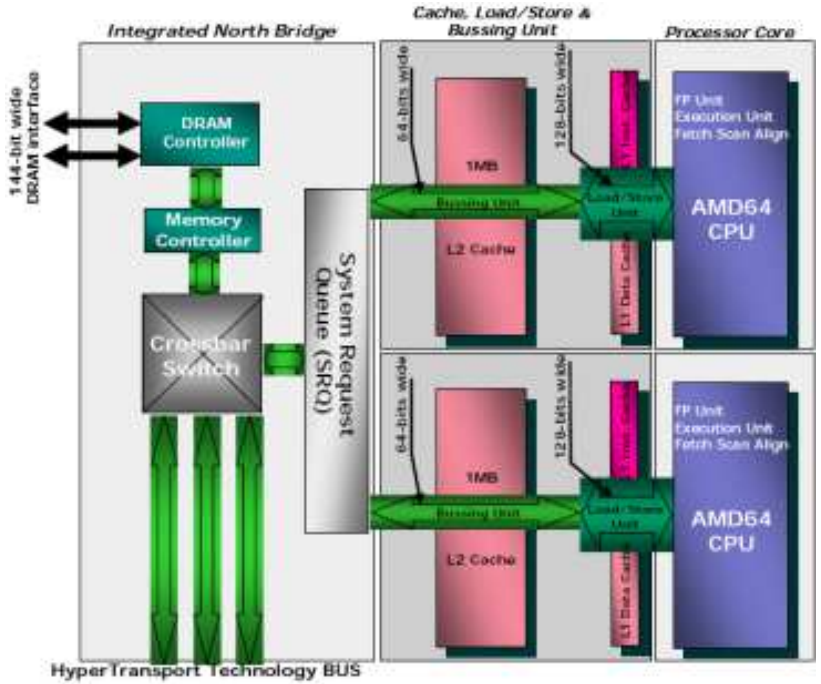
Nine Years of Fine Tuning, Perfecting, Polishing

Host Interface

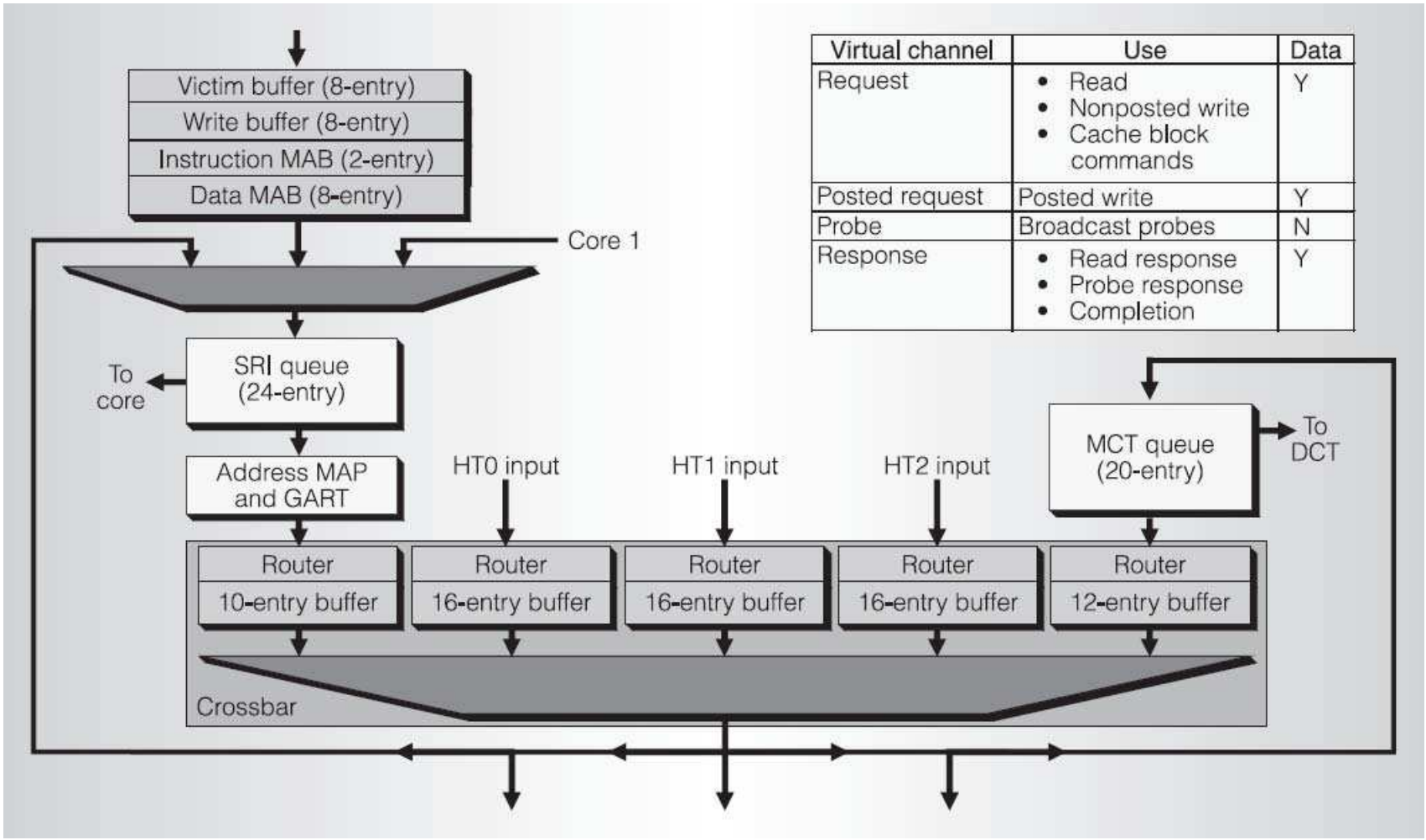


Host Interface (cont.)

- Single Interface for All Cores (SRQ/SRI)
- On-Chip Crossbar and Routing
- Host Bridge for I/O Device Chain



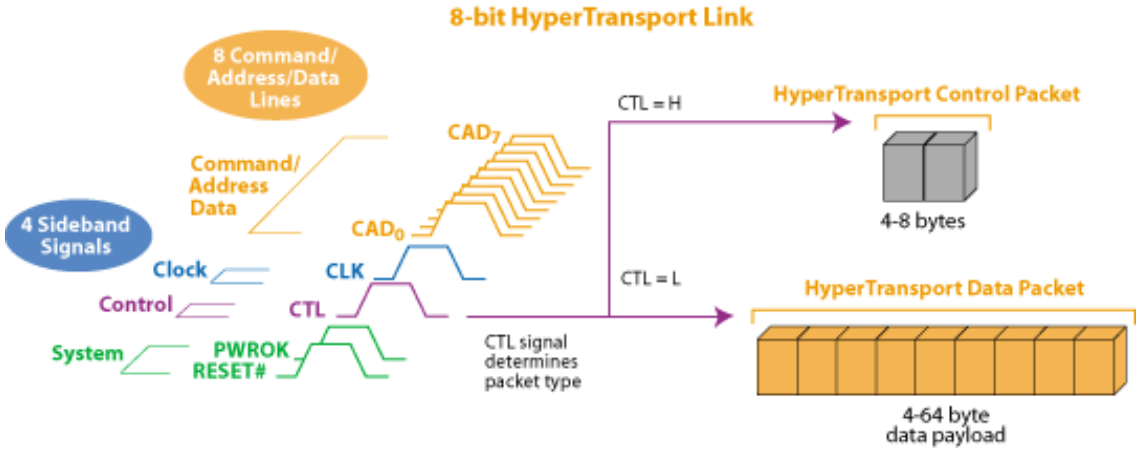
Northbridge Architecture



Virtual channel	Use	Data
Request	• Read • Nonposted write • Cache block commands	Y
Posted request	Posted write	Y
Probe	Broadcast probes	N
Response	• Read response • Probe response • Completion	Y

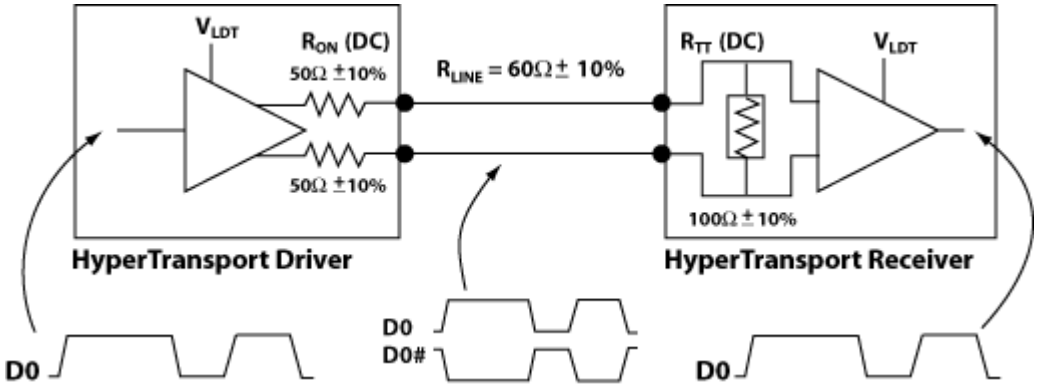
Connecting Device to Host

- **Approach: High Speed Point-to-Point Parallel Link**
  - **Point-to-Point Link Minimizes Parasitic Capacitance**
  - **Clock Forwarding Removes Clock Recovery Overhead**
  - **Parallel Link Delivers High Bandwidth, Low Latency**
- **Control and Data Packets Interleaved on Each Link**
  - **CTL Signal Distinguishes Between Control and Data**
  - **Two Additional System Signals: PWROK and RESET**

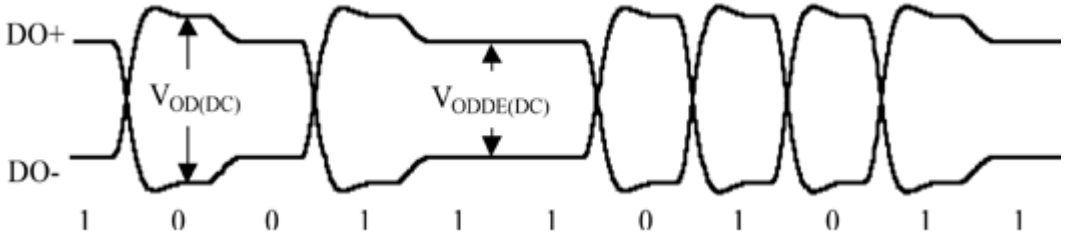


HT Physical Layer

- **Low-Voltage Differential Signalling (LVDS)**

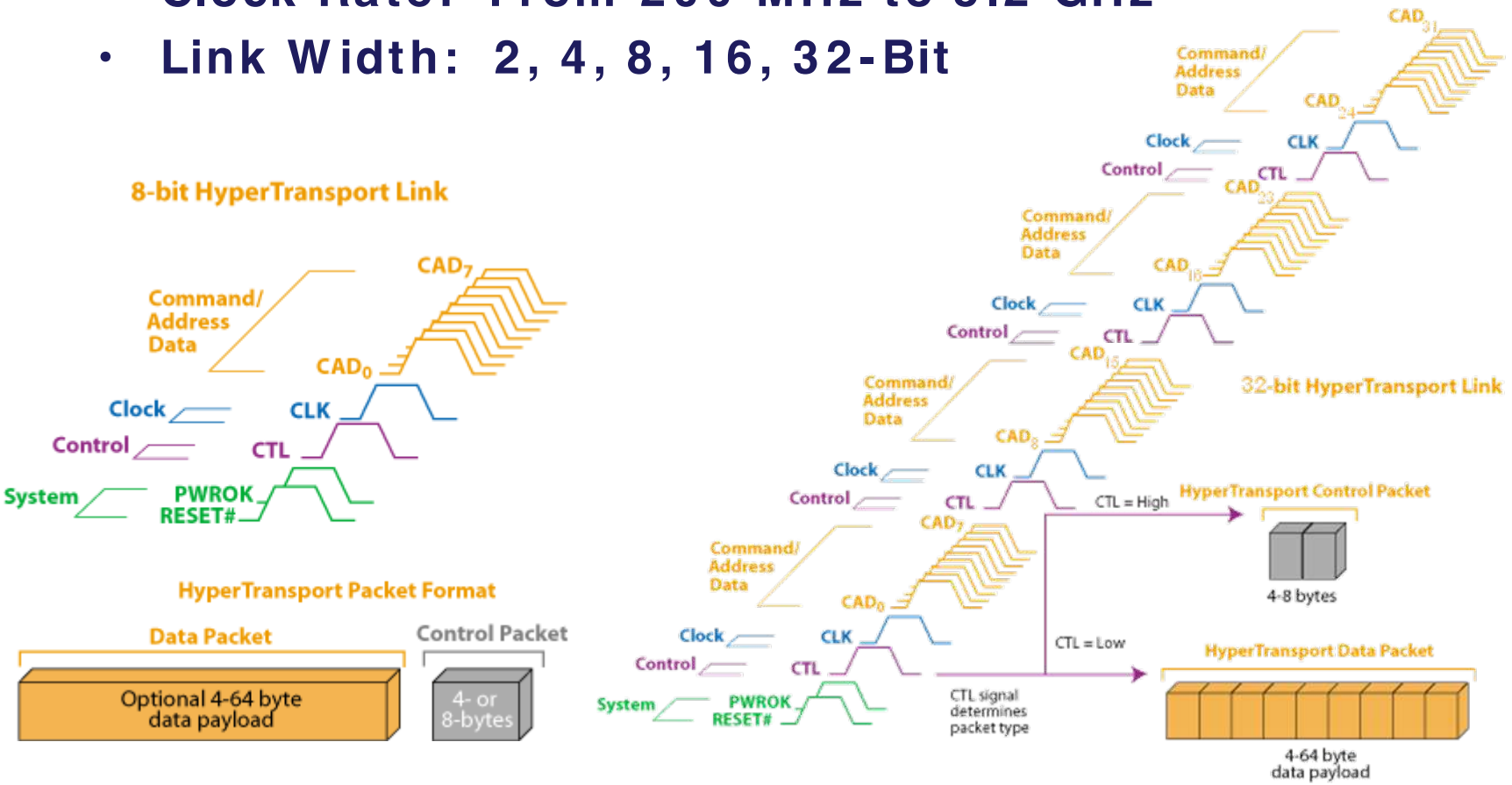


- **Pre-Emphasis Supports Higher Clock Rates**



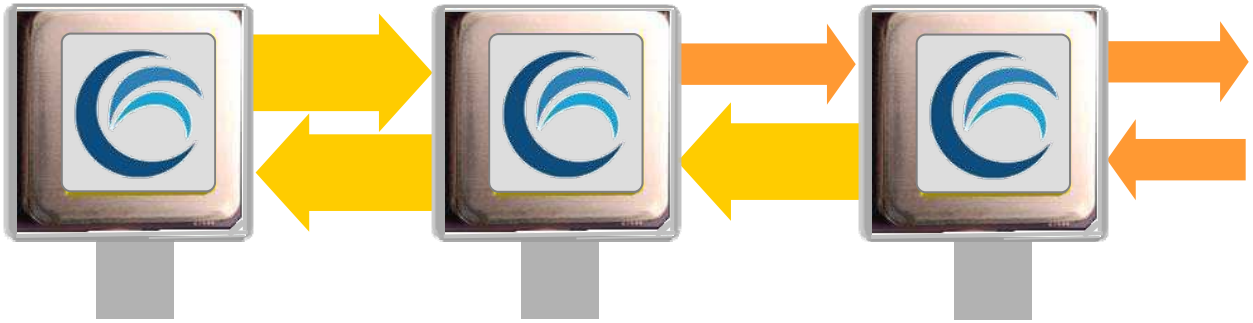
HT Physical Layer (cont.)

- Clock Rate: From 200 MHz to 3.2 GHz
- Link Width: 2, 4, 8, 16, 32-Bit

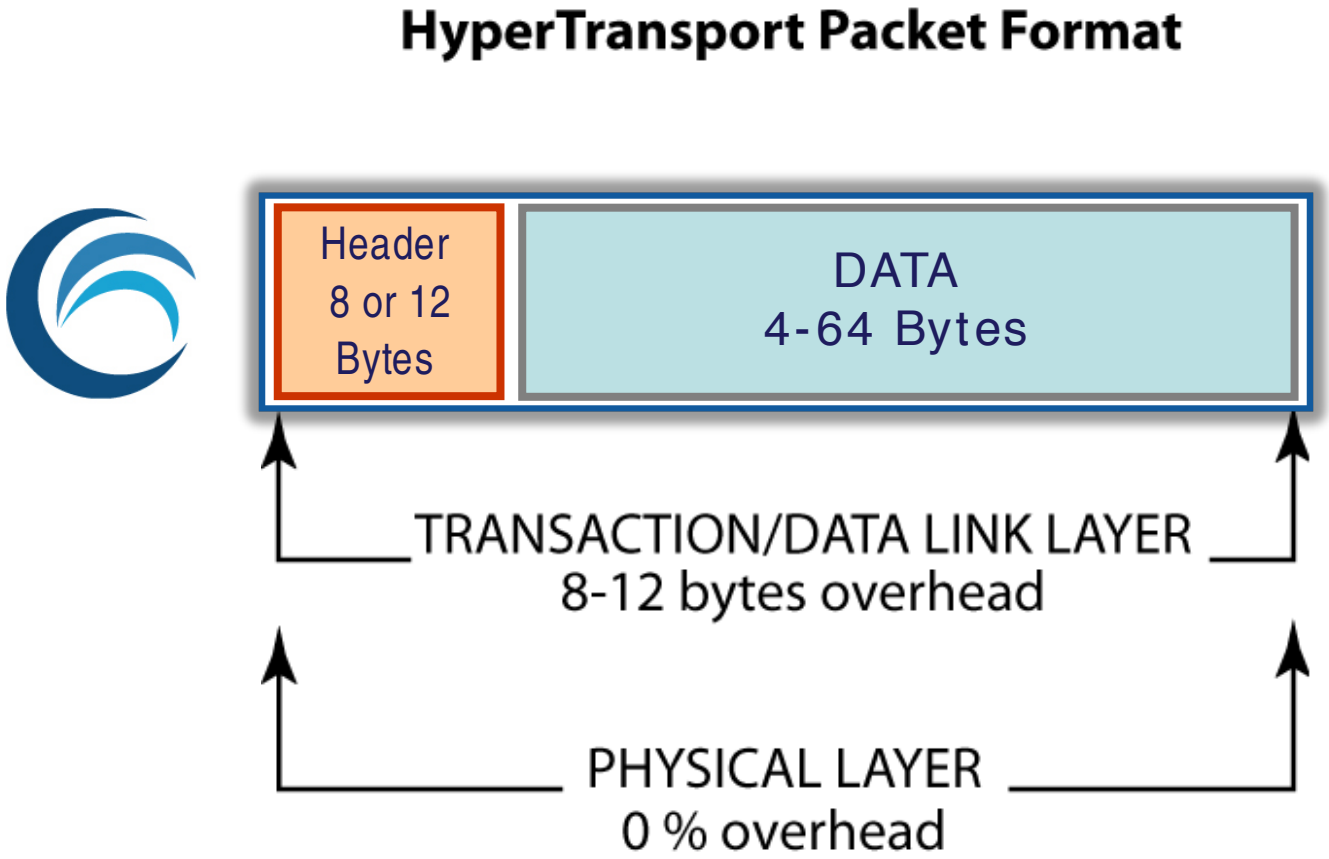


HT Physical Layer (cont.)

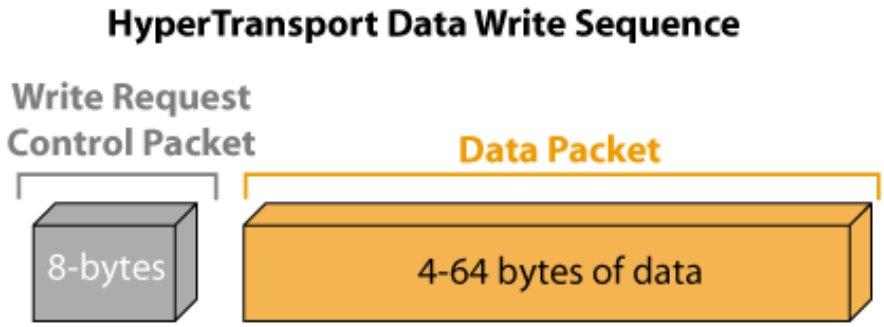
- **Support for Asymmetric and Mixed Link Width**

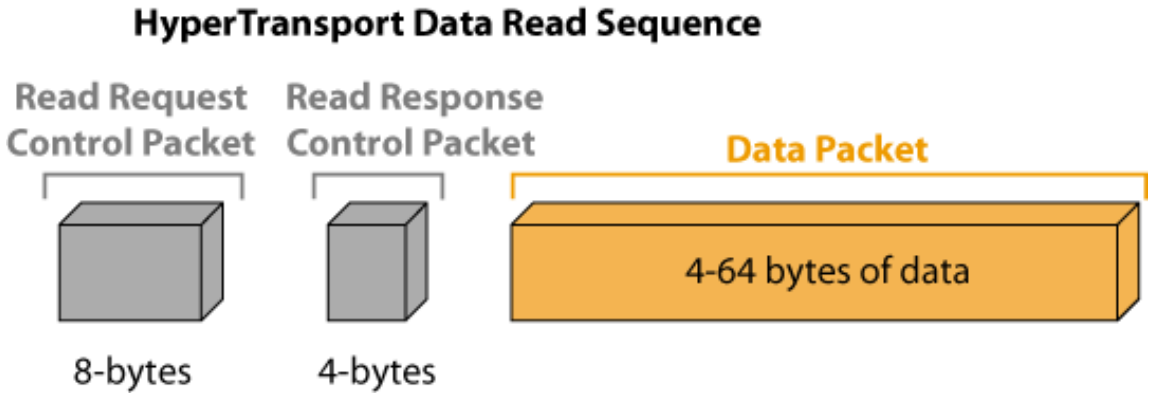


HT Transaction/ Data Link Layer



HT Basic Read/ Write Sequences





HT Request Packet Format

HyperTransport Request Packet Format With Address

Bit-Time	7	6	5	4	3	2	1	0
0	SeqID[3:2]		Cmd[5:0]=type of command					
1	Pass PW	SeqID[1:0]		UnitID[4:0]				
2	Command-Specific							
3	Command-Specific							
4	Addr[15:8]							
5	Addr[23:16]							
6	Addr[31:24]							
7	Addr[39:32]							

HT Request Packet Format (cont.)

Command Field

Cmd[5:0] =

HyperTransport Request Packet Format With Address

Bit-Time	7	6	5	4	3	2	1	0
0	SeqID[3:2]		Cmd[5:0]=type of command					
1	Pass PW	SeqID[1:0]		UnitID[4:0]				
2	Command-Specific							
3	Command-Specific							
4	Addr[15:8]							
5	Addr[23:16]							
6	Addr[31:24]							
7	Addr[39:32]							

Typical Request Types

Command Type	Packet Type/ (size in bytes)
Write Request - Nonposted/posted - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Addr/Data (8)
Write Request with Extended Address	Req/Addr/Data (12)
Read Request - Ordering - pass/no pass - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Address (8)
Broadcast Message	Req/Address (8)
Atomic Read-Modify-Write	Req/Addr/Data (8)

HT Request Packet Format (cont.)

Typical Request Types

Command Field		Command Type	Packet Type/ (size in bytes)
Cmd[5:0]	=	Write Request - Nonposted/posted - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Addr/Data (8)
		Write Request with Extended Address	Req/Addr/Data (12)
		Read Request - Ordering - pass/no pass - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Address (8)
		Broadcast Message	Req/Address (8)
		Atomic Read-Modify-Write	Req/Addr/Data (8)

HT Read Response Packet Format

HyperTransport Read Response Packet Format

Bit-Time	7	6	5	4	3	2	1	0
0	Isoc	Rsv	Cmd[5:0]: 110000					
1	Pass PW	Bridge	Rsv	UnitID[4:0]				
2	Count[1:0]		Error0	SrcTag[4:0]				
3	Rsv/RqUID		Error1	Rsv/RspVCSet			Count[3:2]	

Important Read Response Packet Fields

Cmd[5:0]=110000	Command Type
Count[1:0]	Byte count
Count[3:2]	Byte count
Isoc	Isochronous/ Normal bandwidth

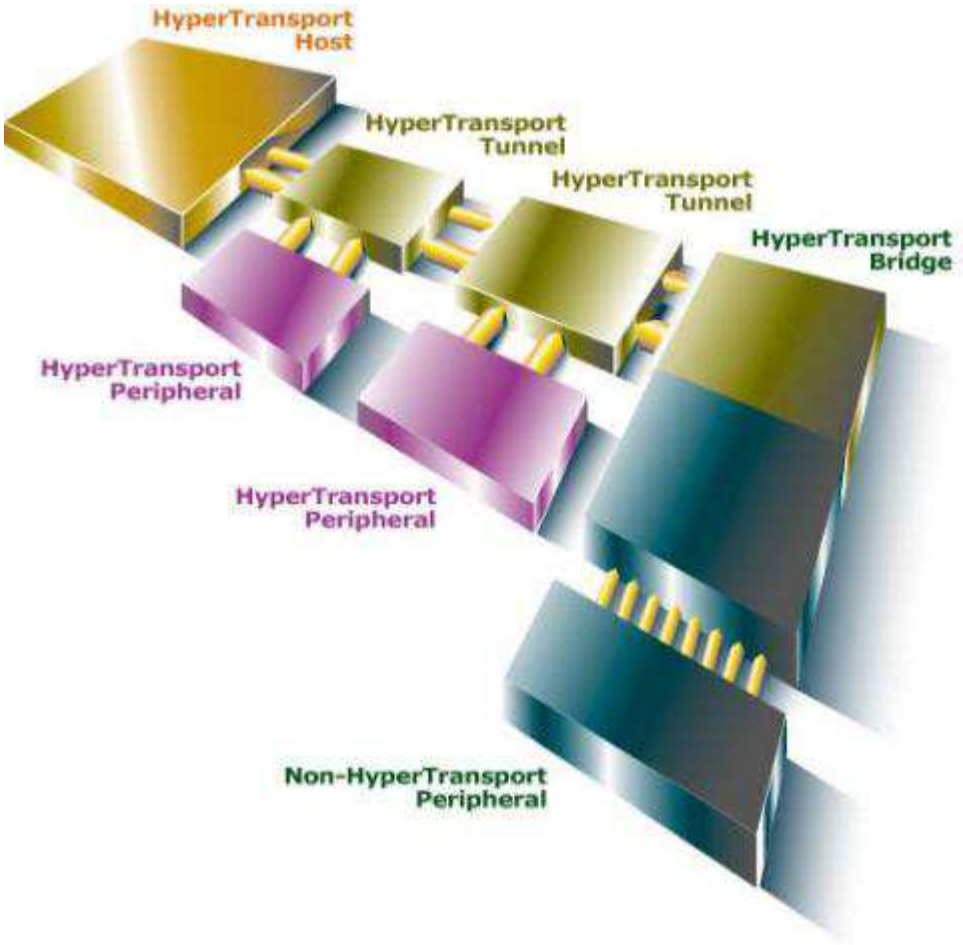
Error0 | Error1 = Error Status

Error Codes		
E_0	E_1	Meaning
0	0	Normal Completion
0	1	Target Abort Request
1	0	Data Error
1	1	Master Abort

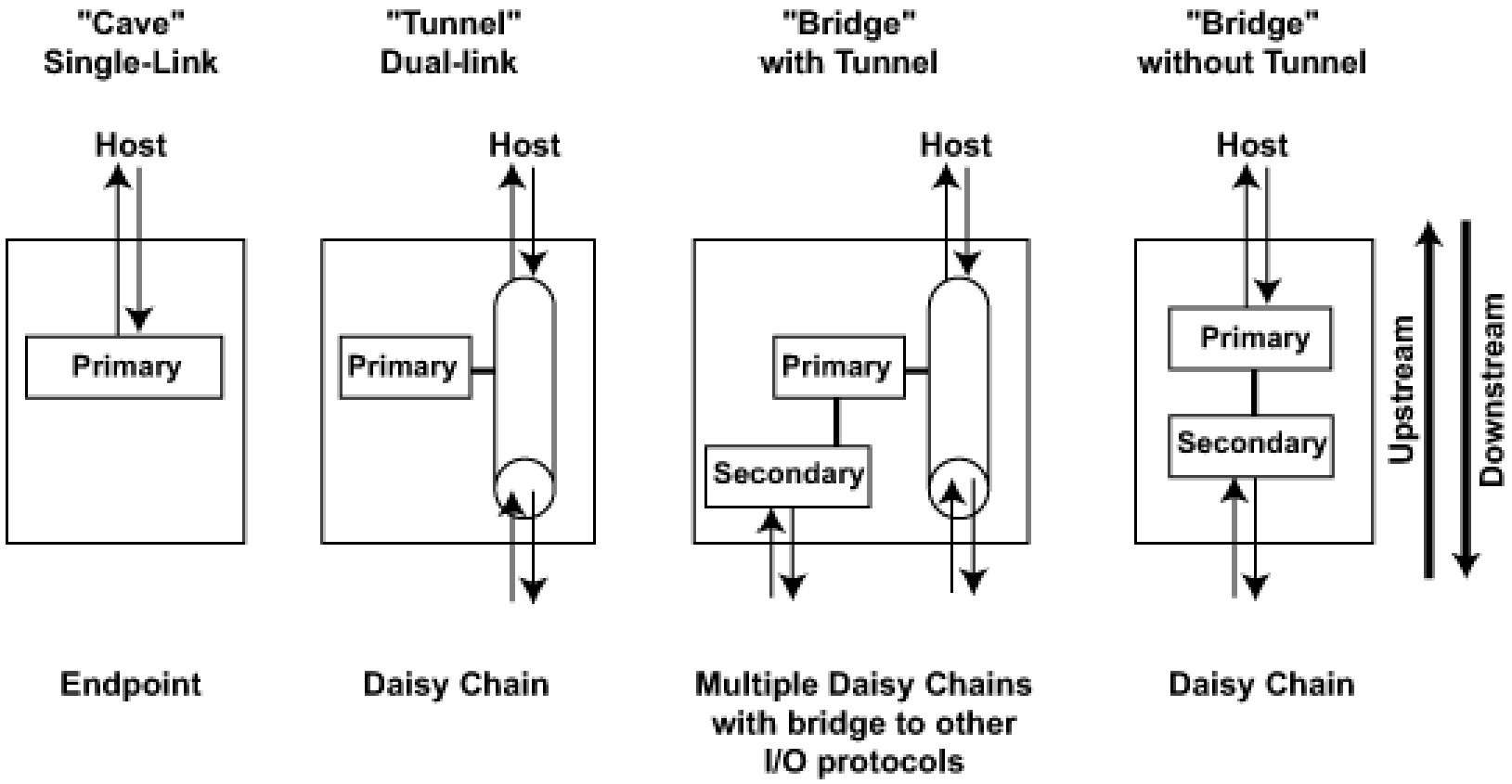
Common HT Command Types

Command Field		Command Type	Packet Type/ (size in bytes)
Cmd[5:0]	=	Write Request - Nonposted/posted - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Addr/Data (8)
		Write Request with Extended Address	Req/Addr/Data (12)
		Read Request - Ordering - pass/no pass - data length = byte/doubleword - normal/isochronous - noncoherent/coherent	Req/Address (8)
		Read Response	Resp/Data (4)
		Target Done	Response (4)
		Broadcast Message	Req/Address (8)
		Atomic Read-Modify-Write	Req/Addr/Data (8)
		Address Extension	Address (12)
		Flush posted writes	Request (4)
		Fence for posted requests	Request (4)
		Extended Flow Control for VCSets 0-7	Info (4)
		Link Synchronization and Error Packet	Info (4)

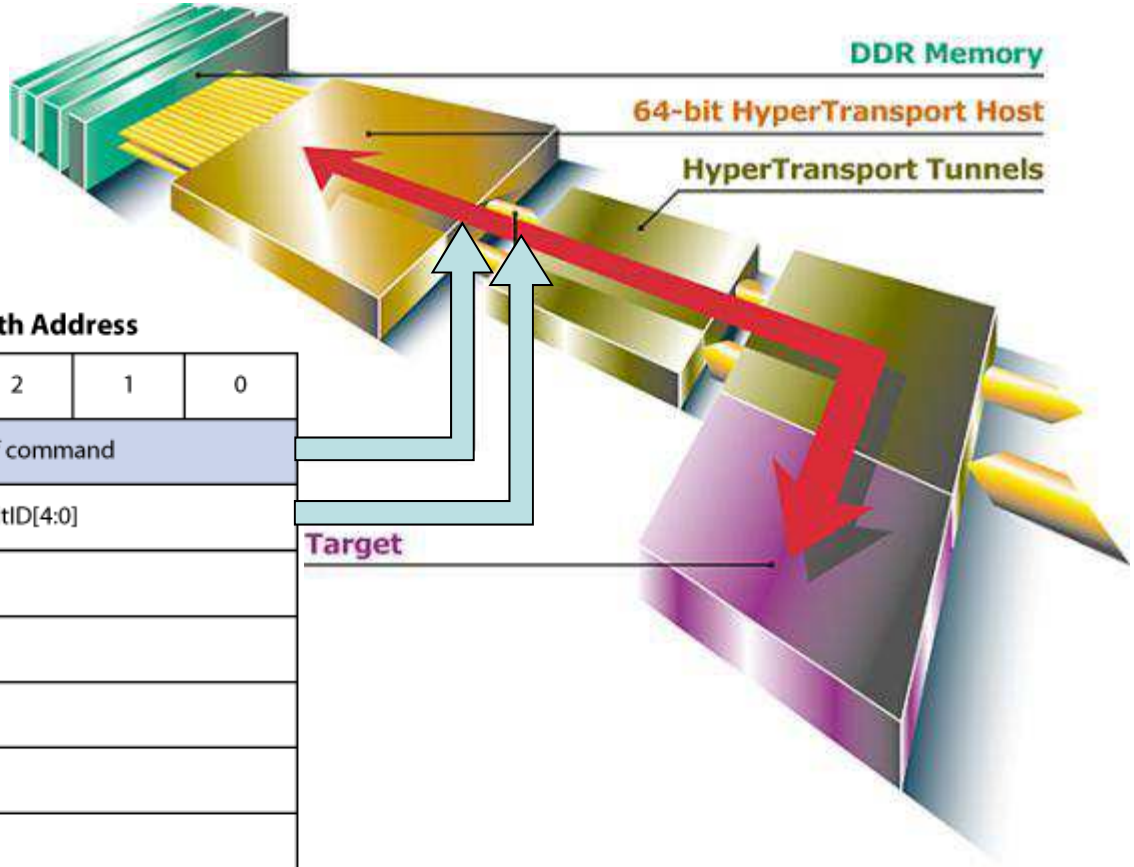
Connecting Multiple Devices to Host



HyperTransport I/O Device Configurations



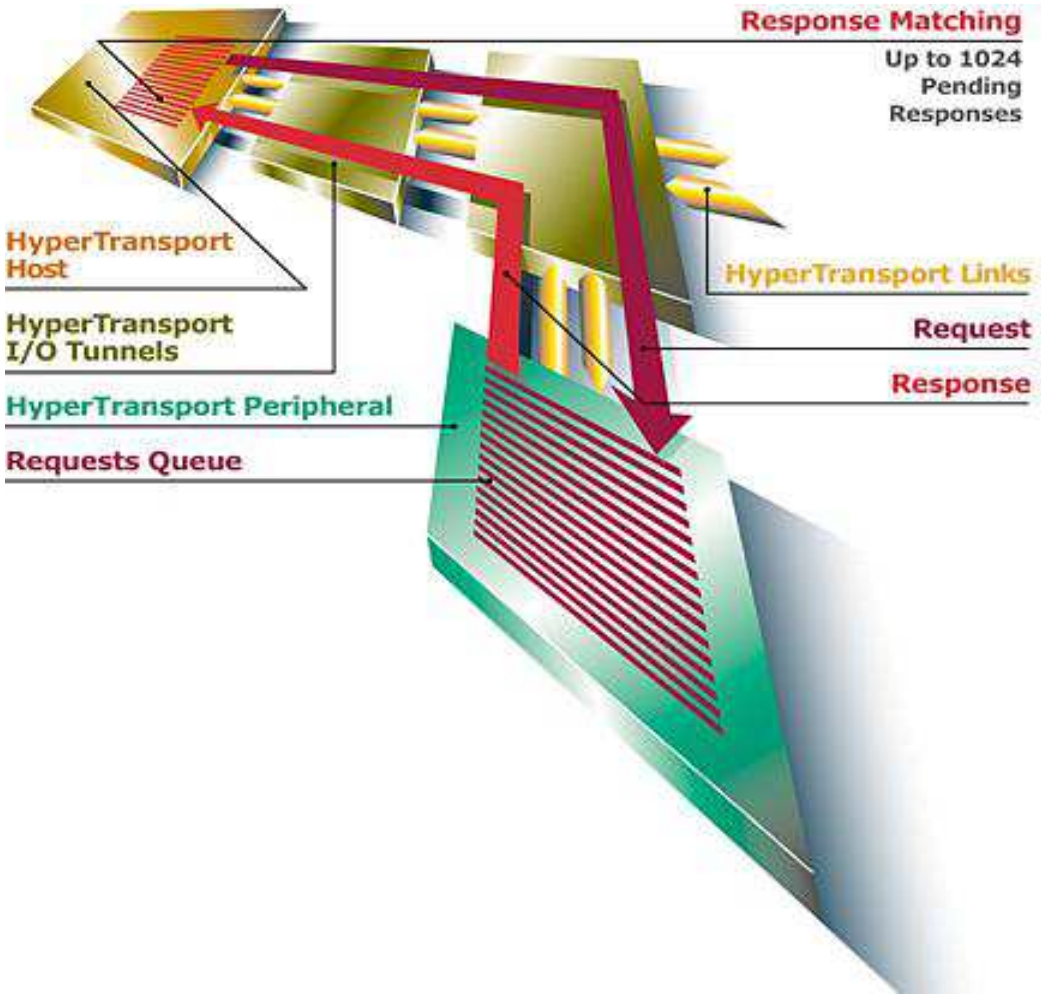
Routing to Target Device



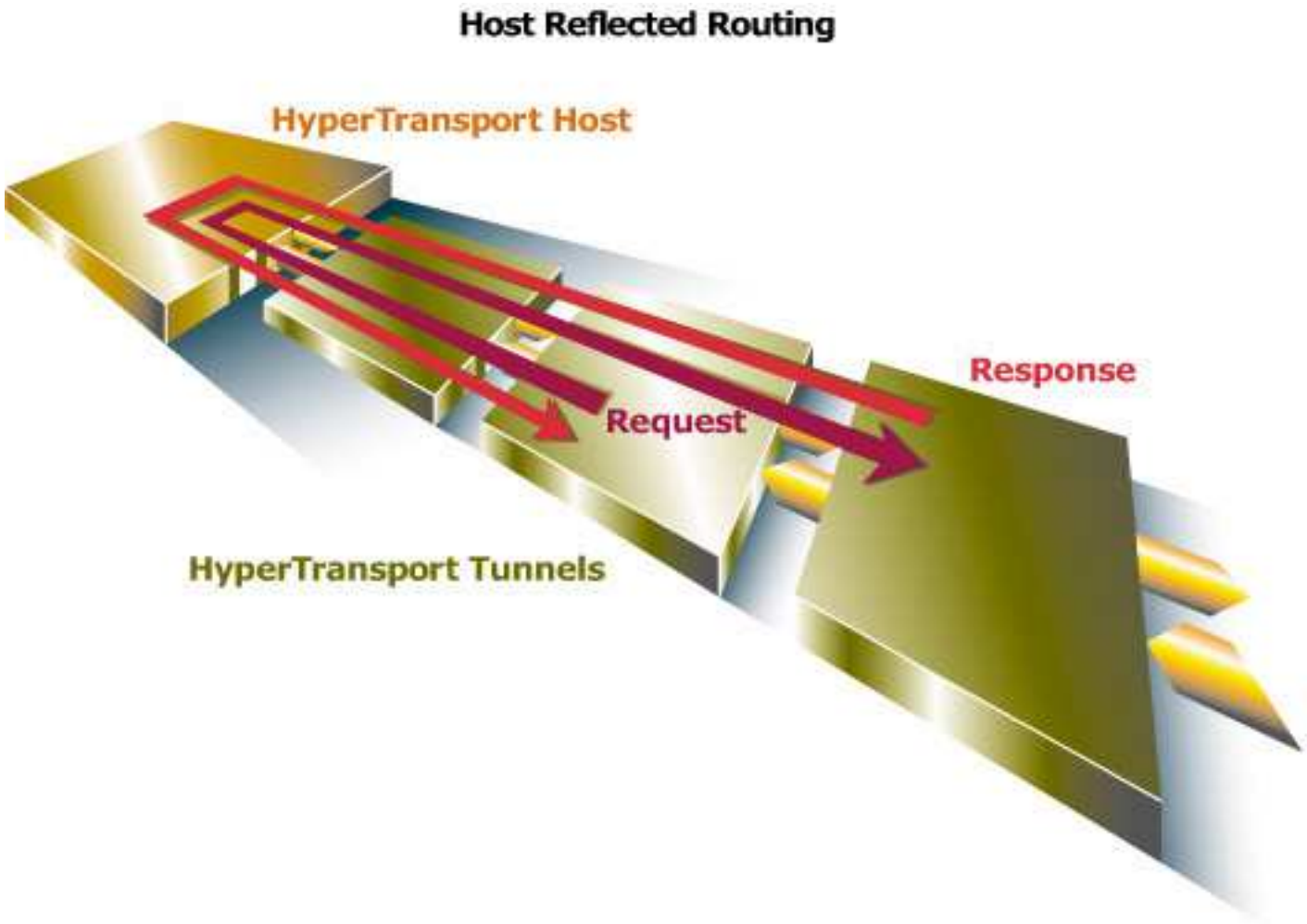
HyperTransport Request Packet Format With Address

Bit-Time	7	6	5	4	3	2	1	0
0	SeqID[3:2]		Cmd[5:0]=type of command					
1	Pass PW	SeqID[1:0]	UnitID[4:0]					
2	Command-Specific							
3	Command-Specific							
4	Addr[15:8]							
5	Addr[23:16]							
6	Addr[31:24]							
7	Addr[39:32]							

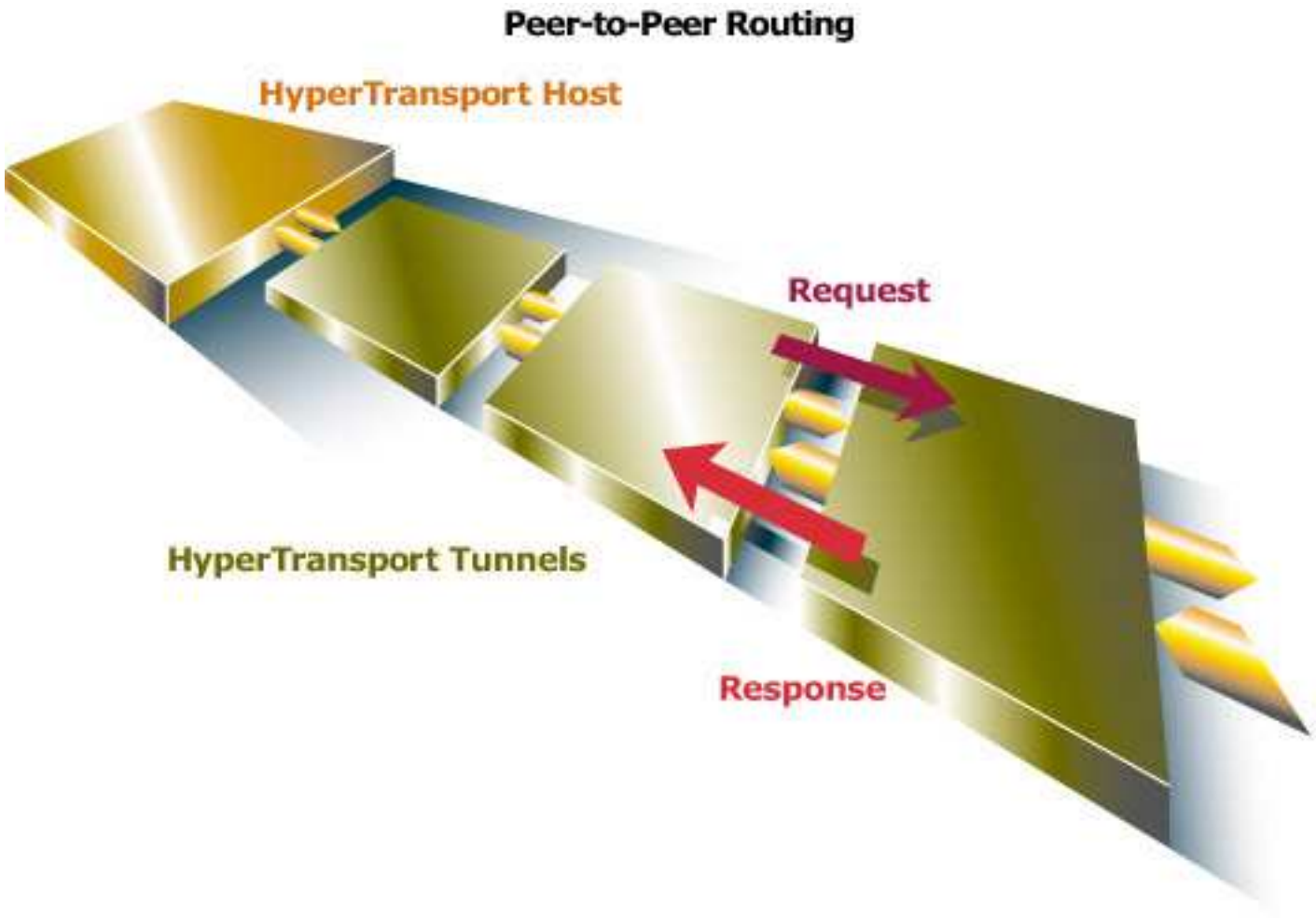
Pipelining Multiple Requests



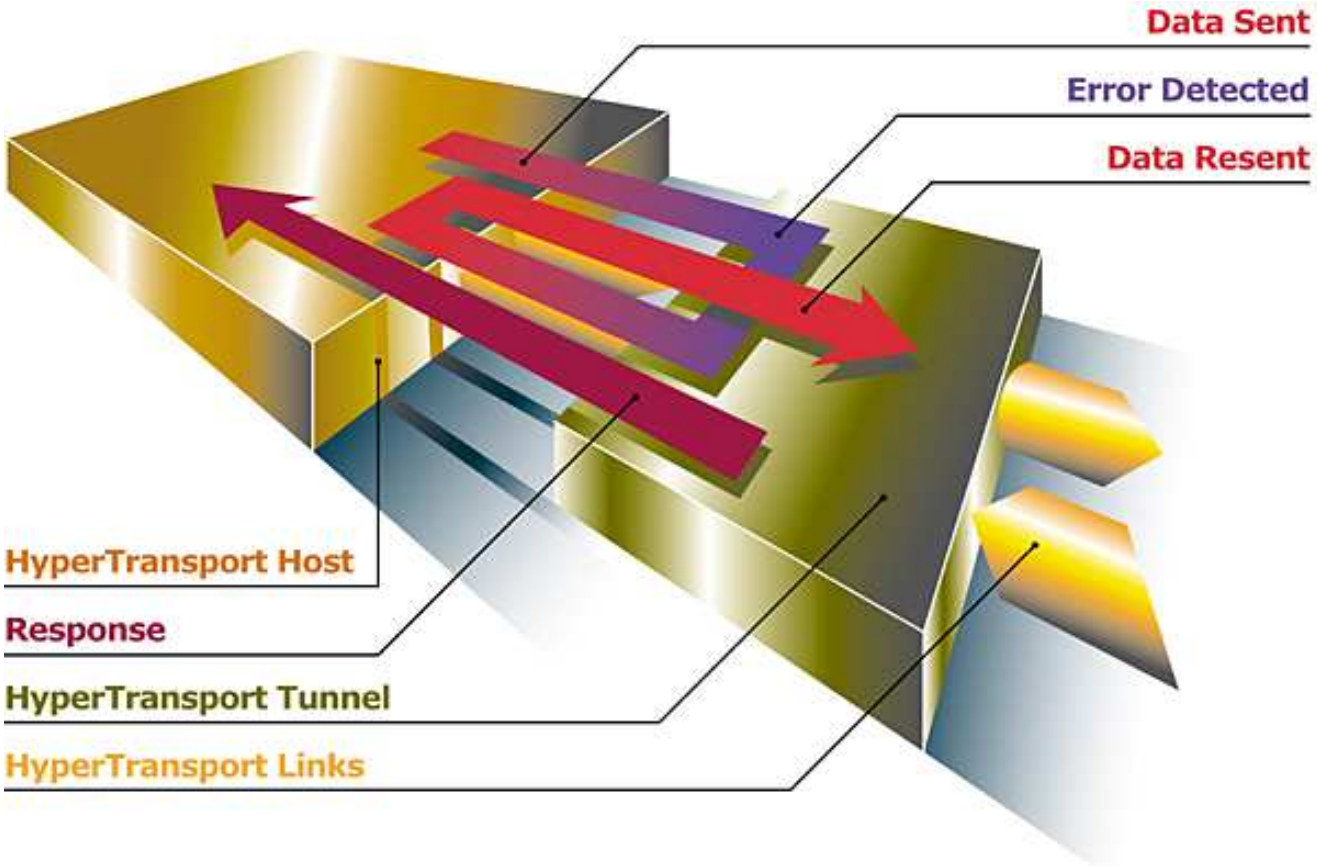
Communication Between Two I/O Devices



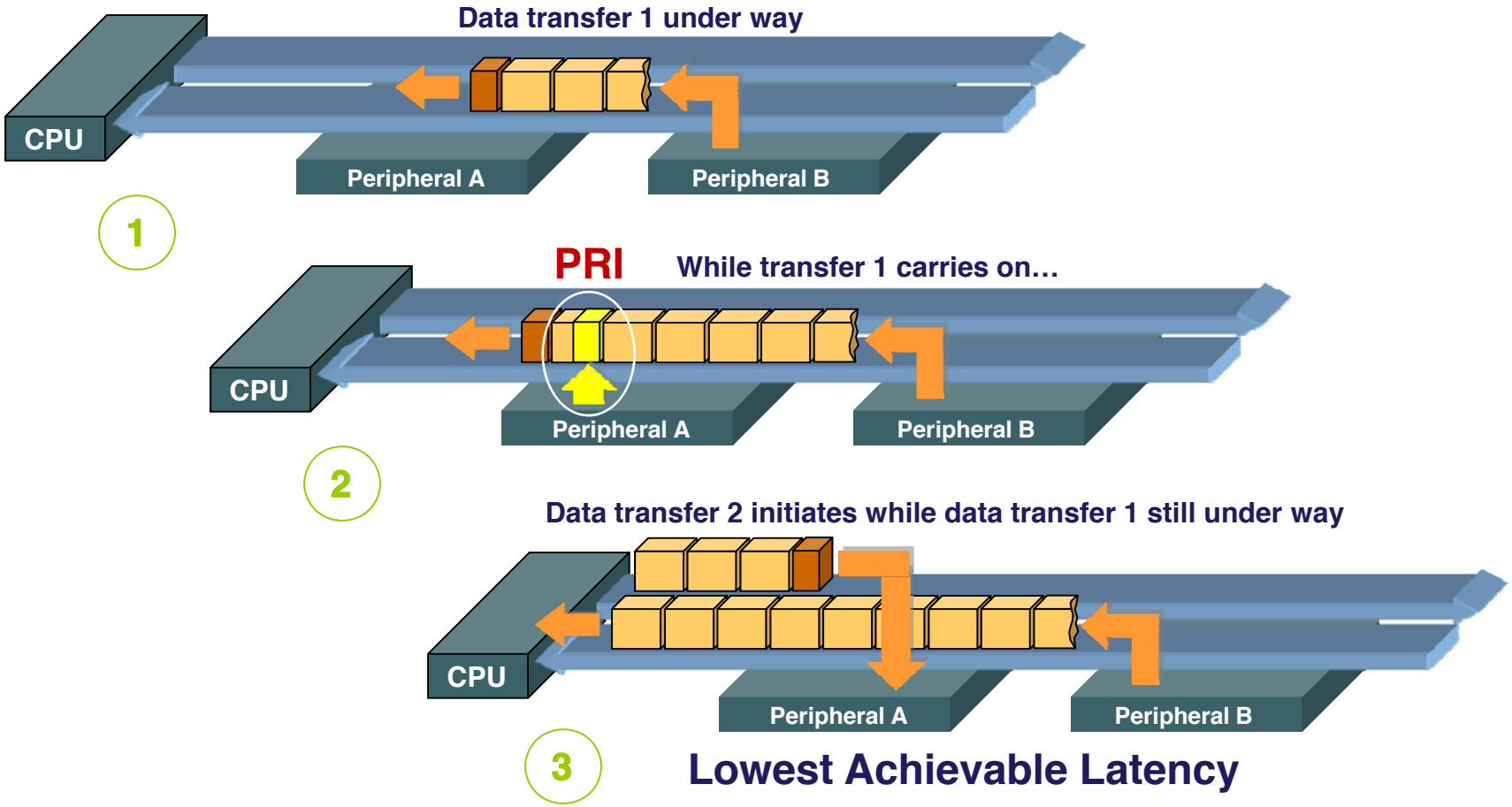
Communication Between Two I/O Devices (cont.)



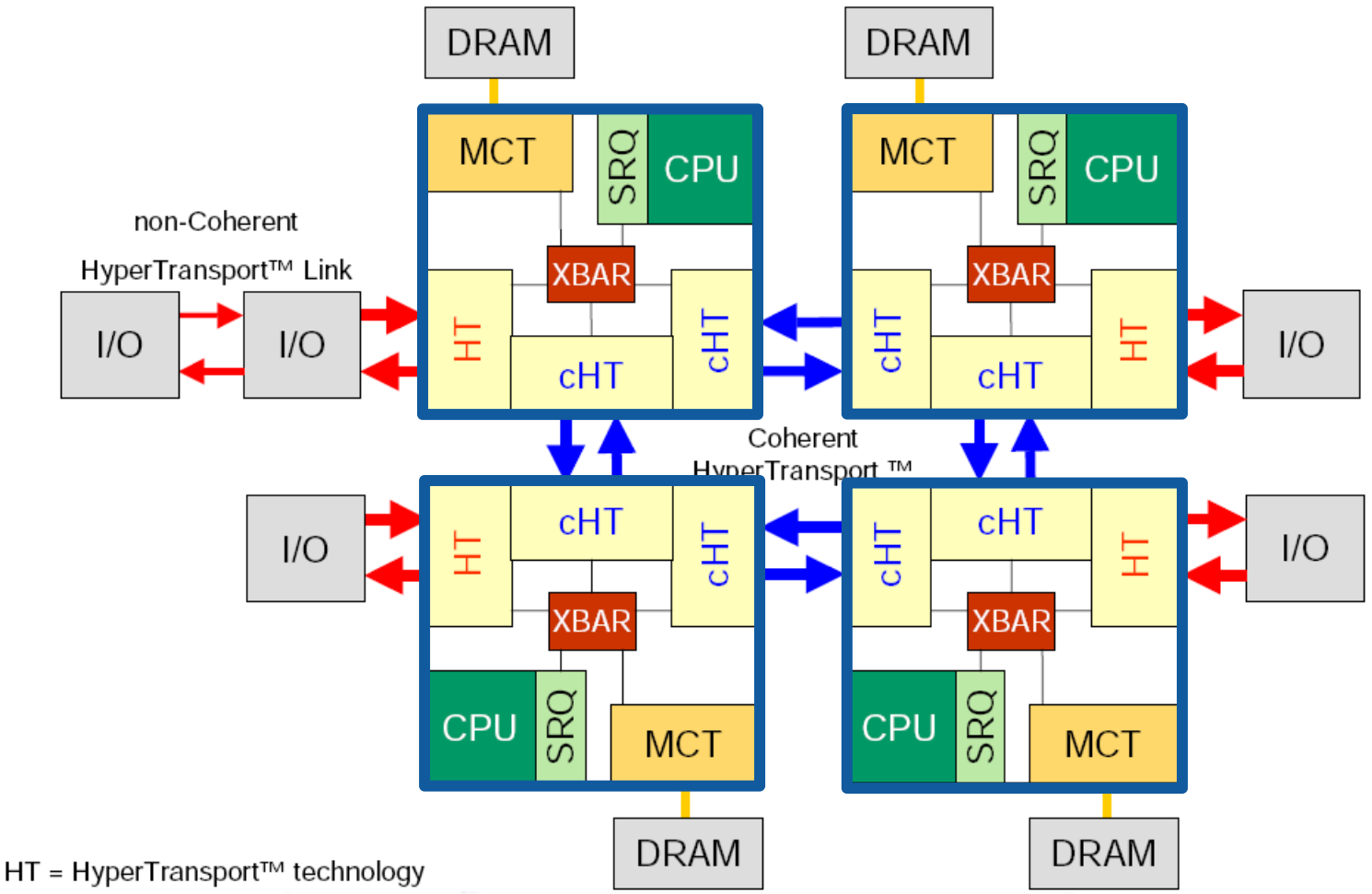
Transmission Error Handling



Priority Request Interleaving™



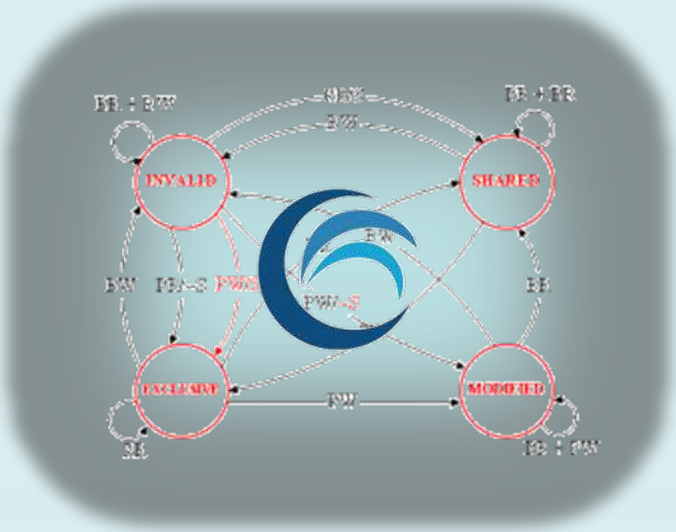
Interconnecting Multiple Hosts



AMD

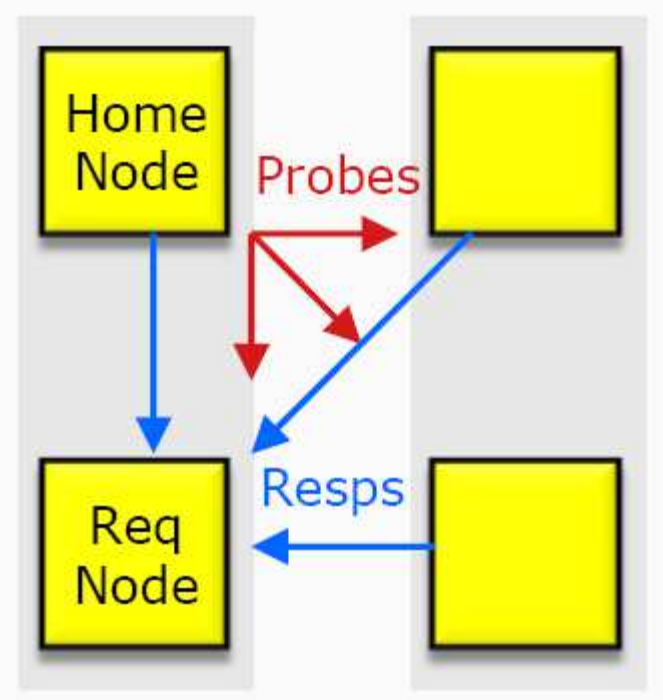
Cache Coherence Support

Proprietary Technology



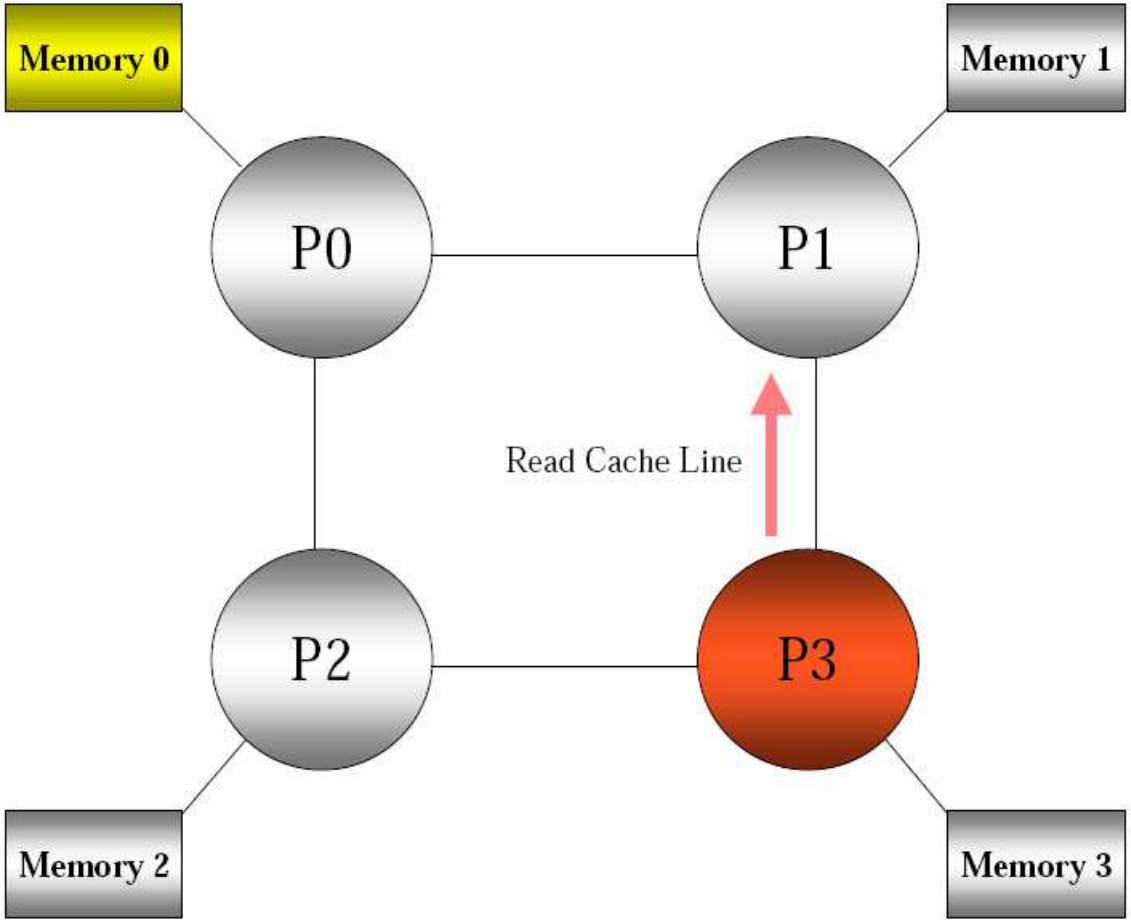
AMD cHT Basics

- On-Chip Support for Up to 8 CPUs
- Broadcast-Based 3-Hop Invalidation Cache Coherence Protocol



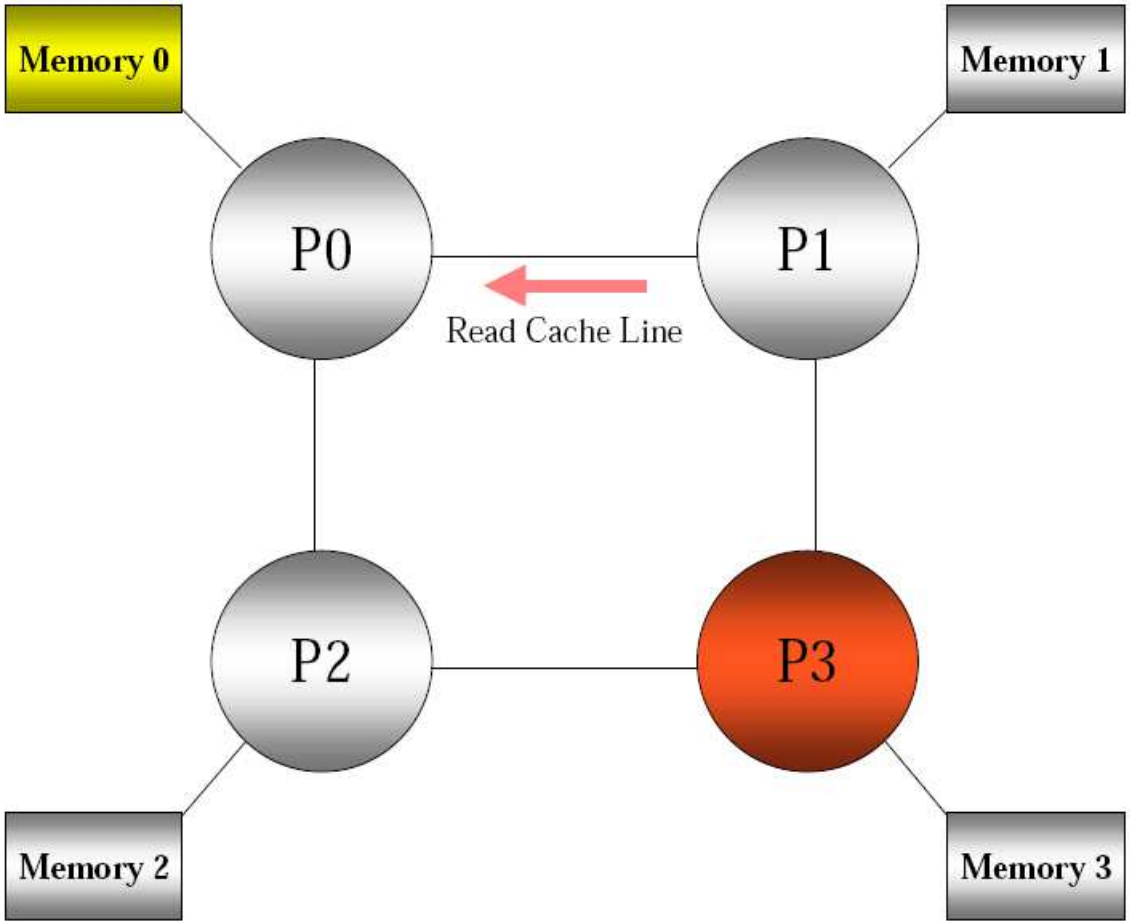
AMD cHT Read Request Example

Step 1

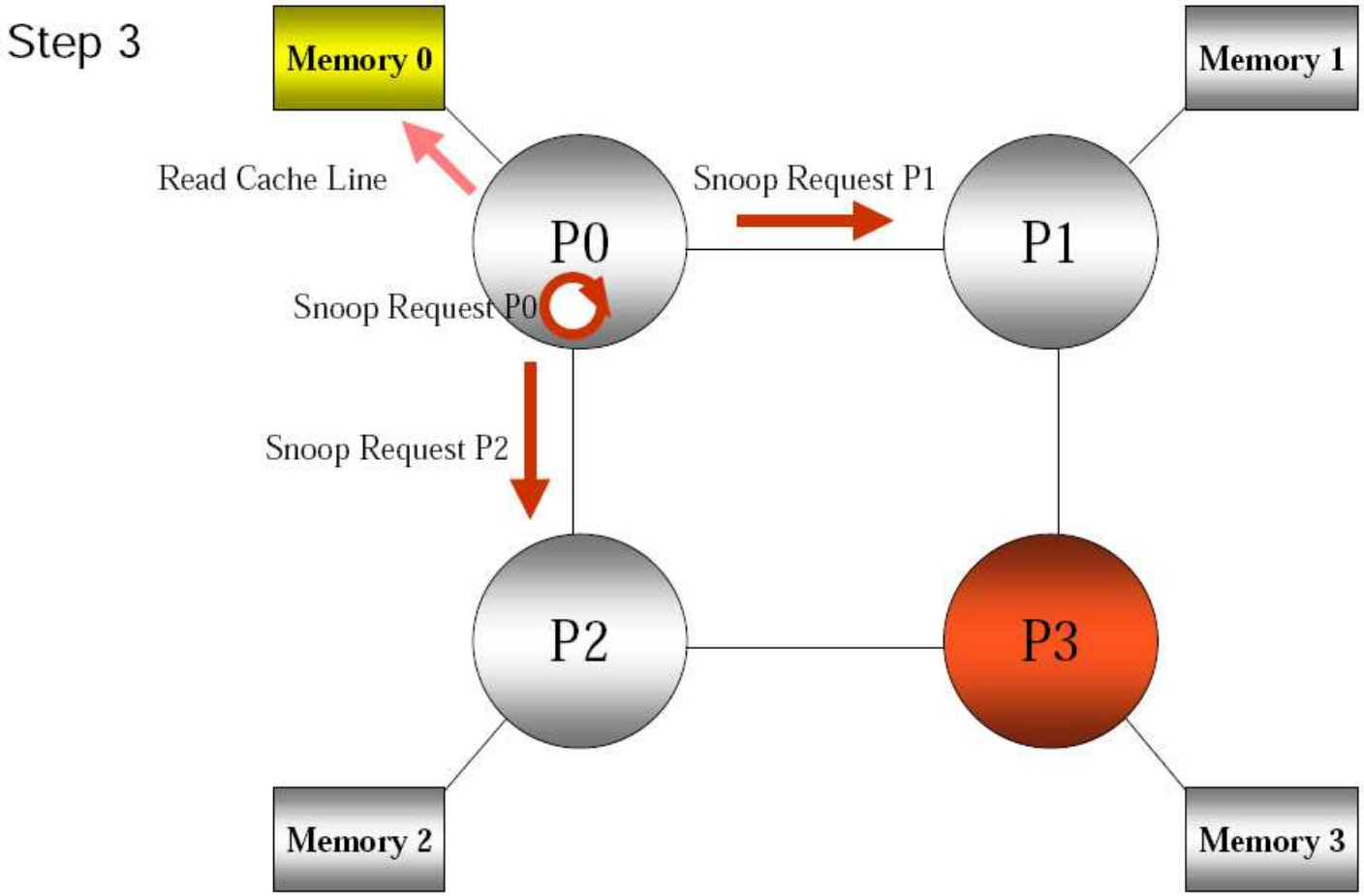


AMD cHT Read Request Example (cont.)

Step 2

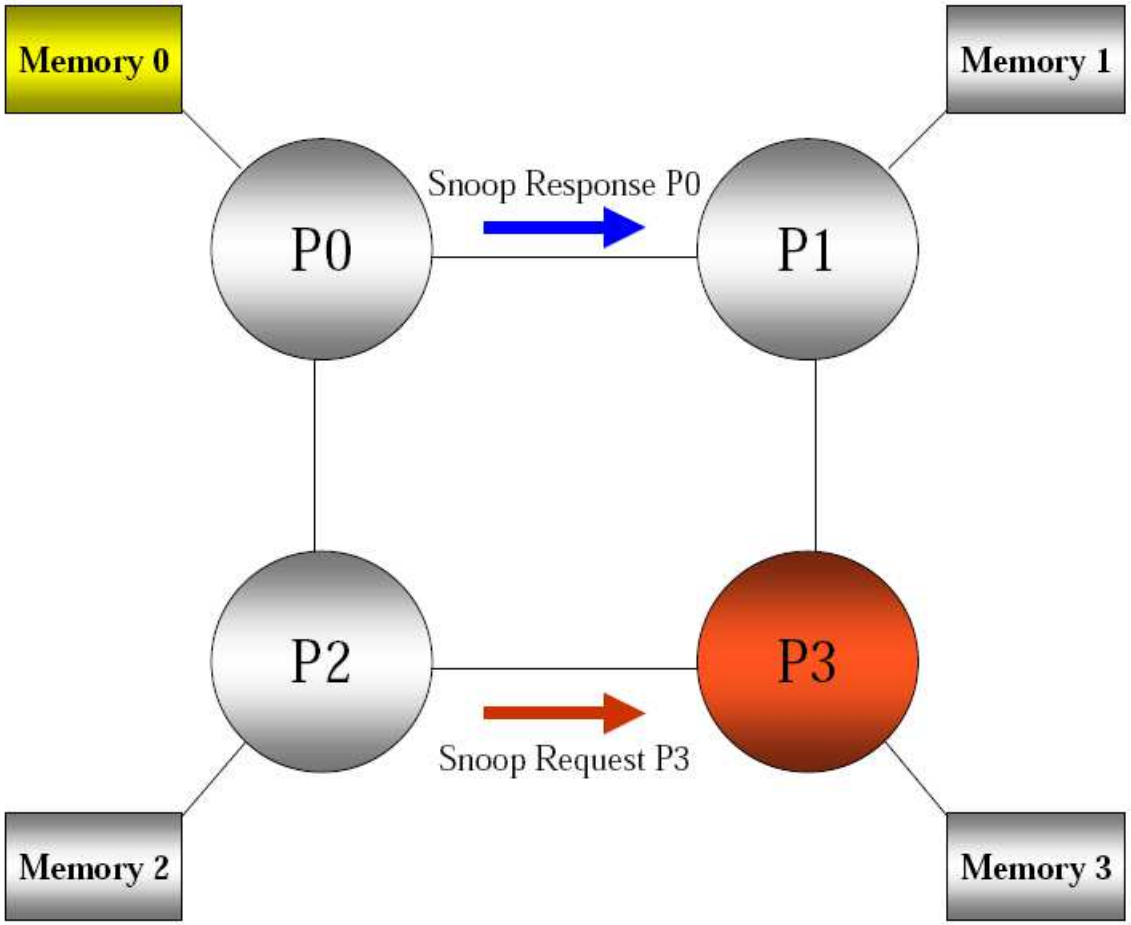


AMD cHT Read Request Example (cont.)

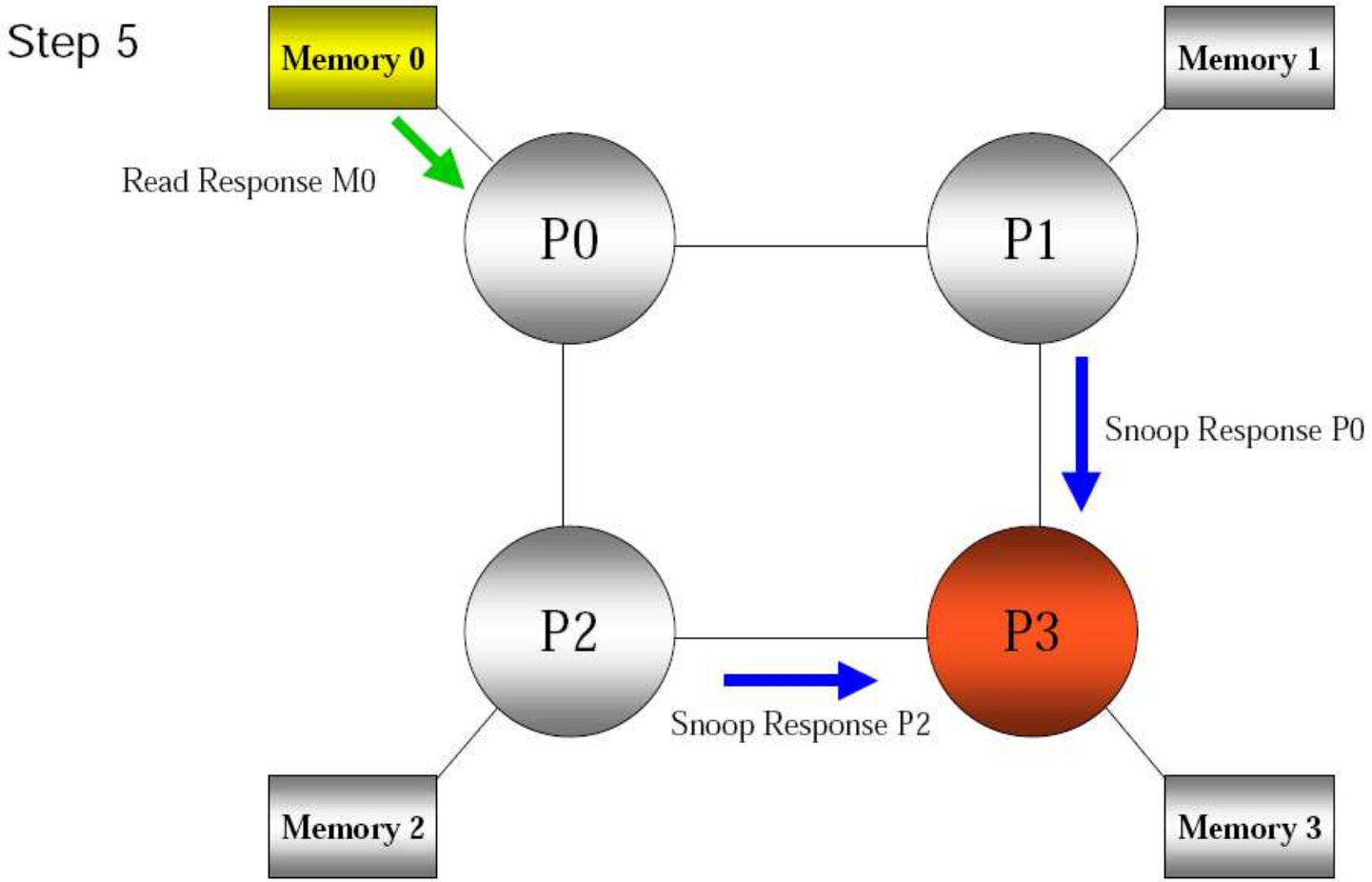


AMD cHT Read Request Example (cont.)

Step 4

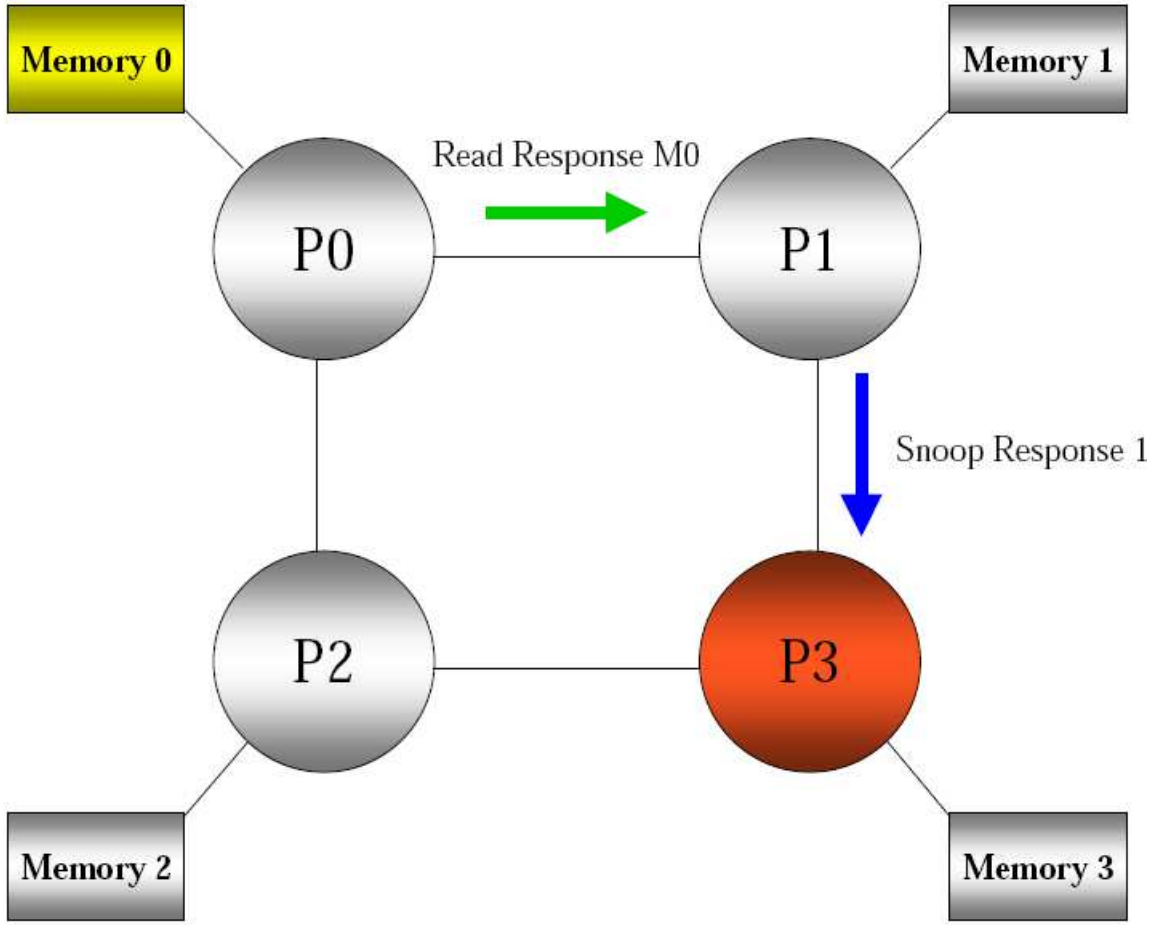


AMD cHT Read Request Example (cont.)



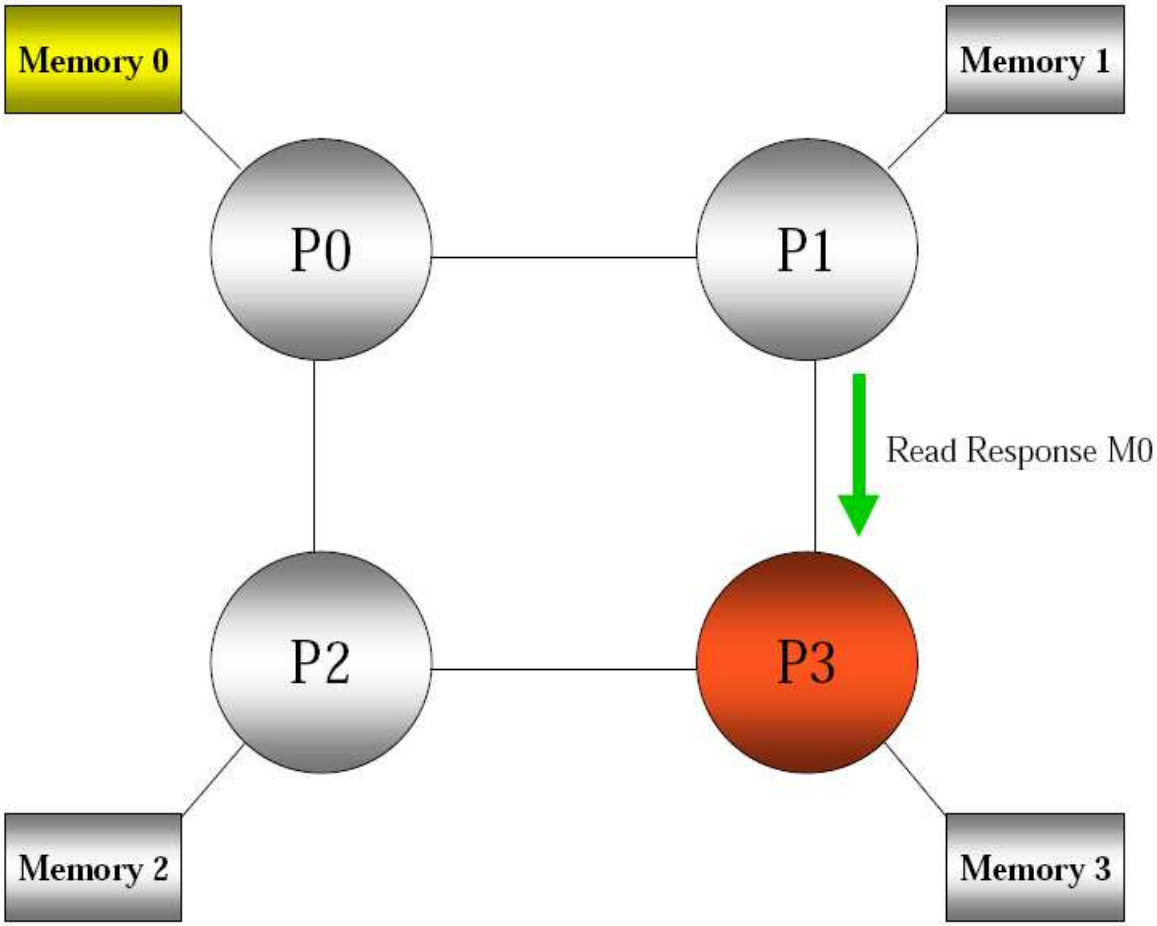
AMD cHT Read Request Example (cont.)

Step 6



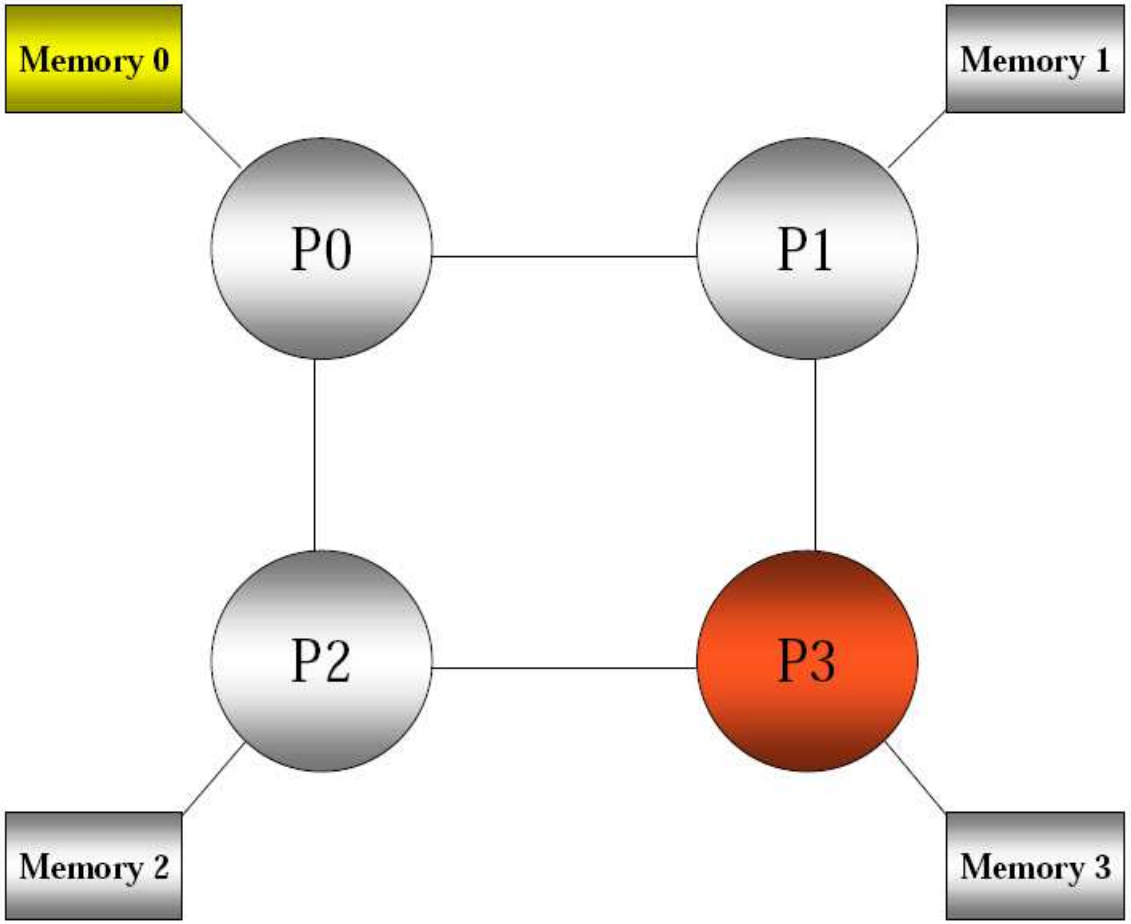
AMD cHT Read Request Example (cont.)

Step 7



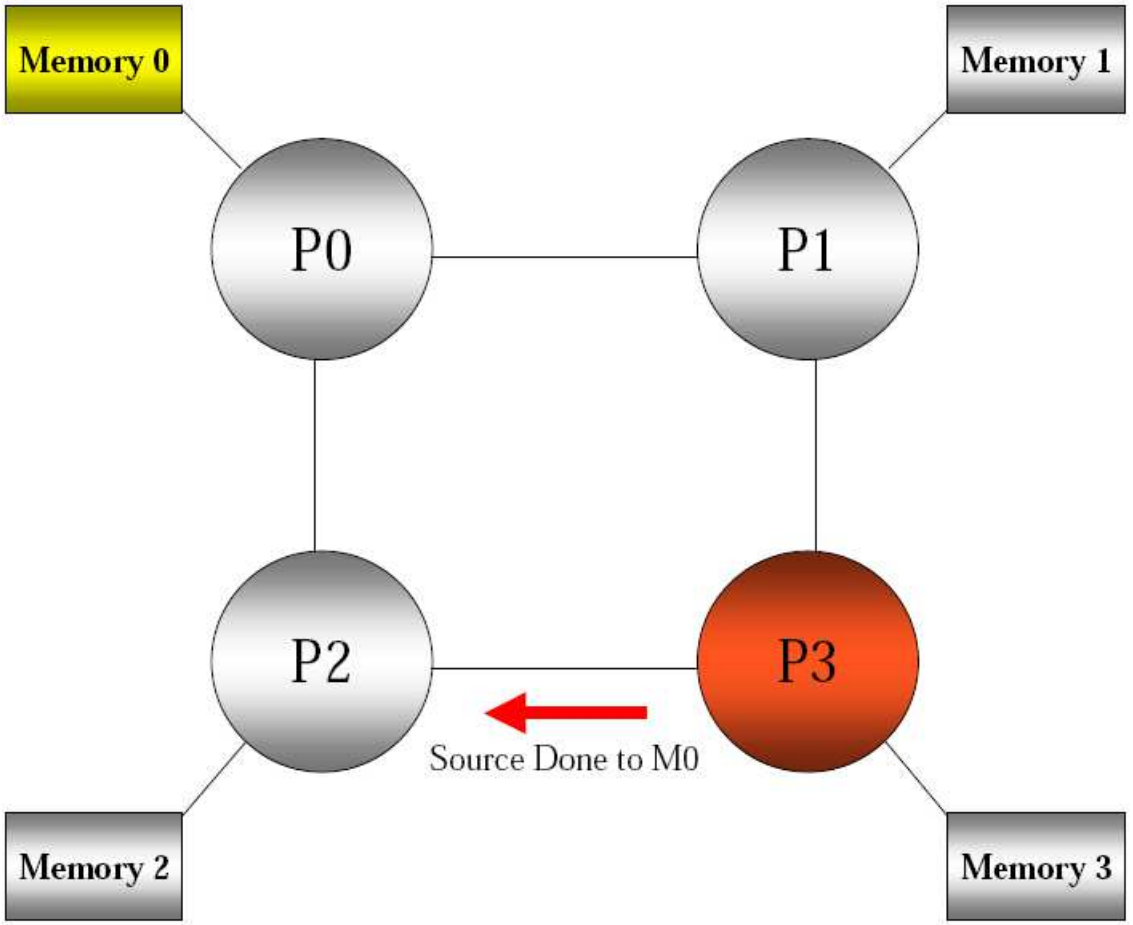
AMD cHT Read Request Example (cont.)

Step 8



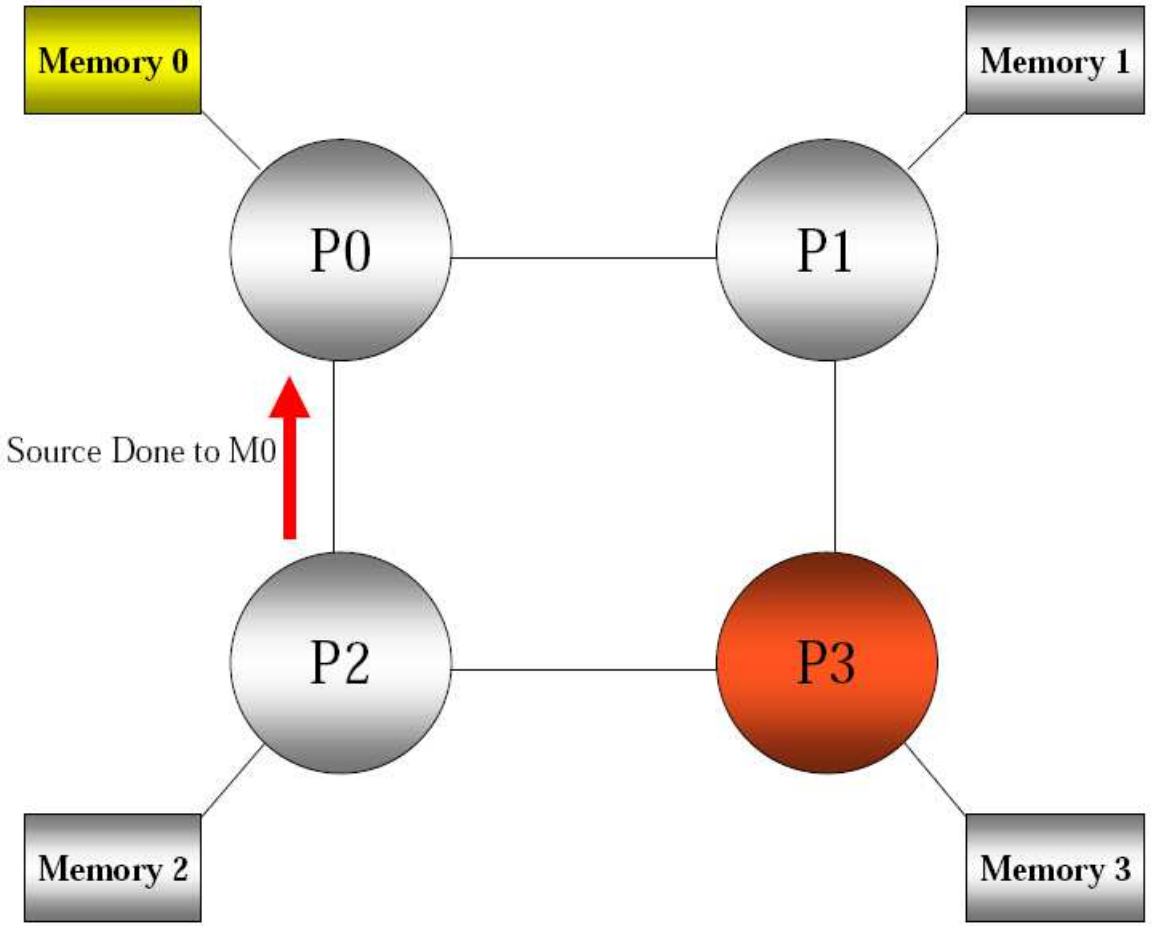
AMD cHT Read Request Example (cont.)

Step 9



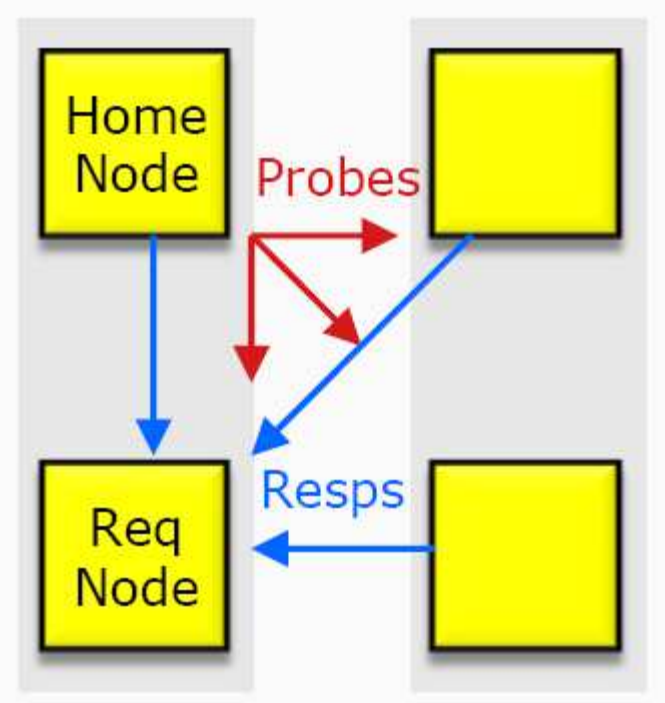
AMD cHT Read Request Example (cont.)

Step 10



AMD cHT HT-Assist (Probe Filter)

- Old cHT Broadcast Protocol Broadcasts Probes to Invalidate Copies even if Memory Line is Clean



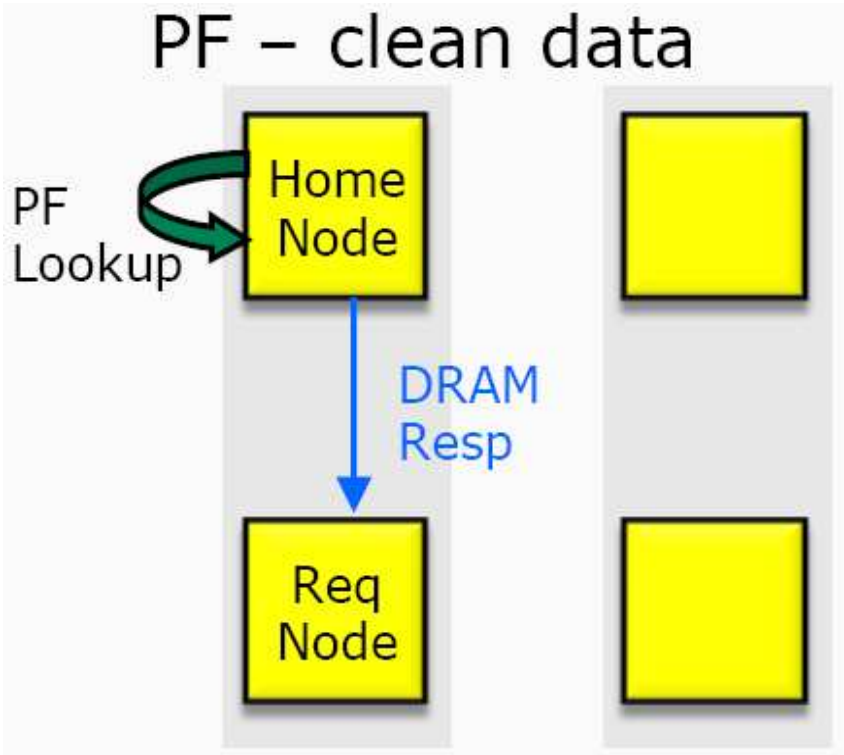
AMD cHT HT-Assist (Probe Filter) (cont.)

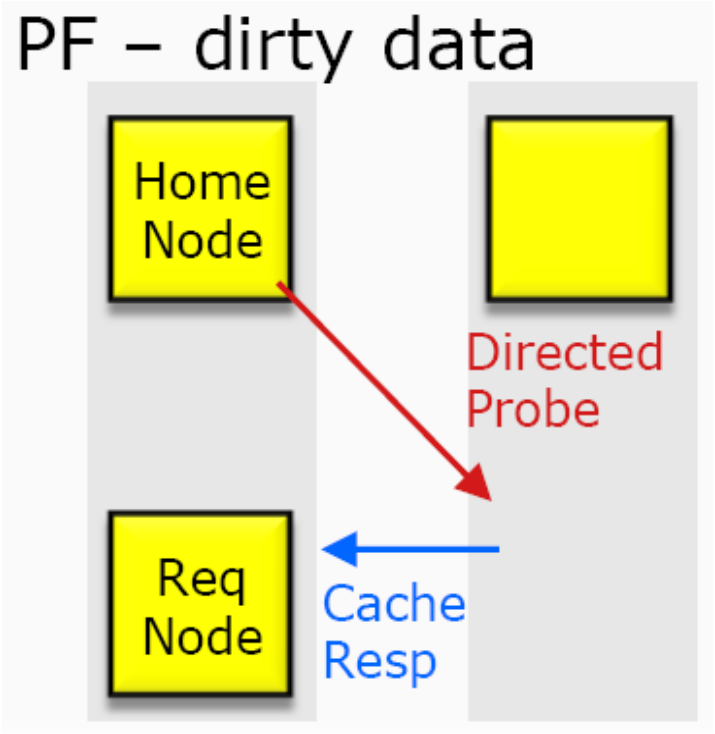
- Sparse Directory Cache Next to Memory Controller
- Rule: If a Line is Cached, it has an Entry in PF
  - Replacement Policy Makes Room for New Lines
- Enhanced behavior:
  - No Probing for Uncached Lines
  - Directed Probe to Request Copy of Cached Line
- Benefits:
  - Significantly Less Bandwidth Use
  - Shorter Access Latency, Mainly for Uncached Lines

For More Details:

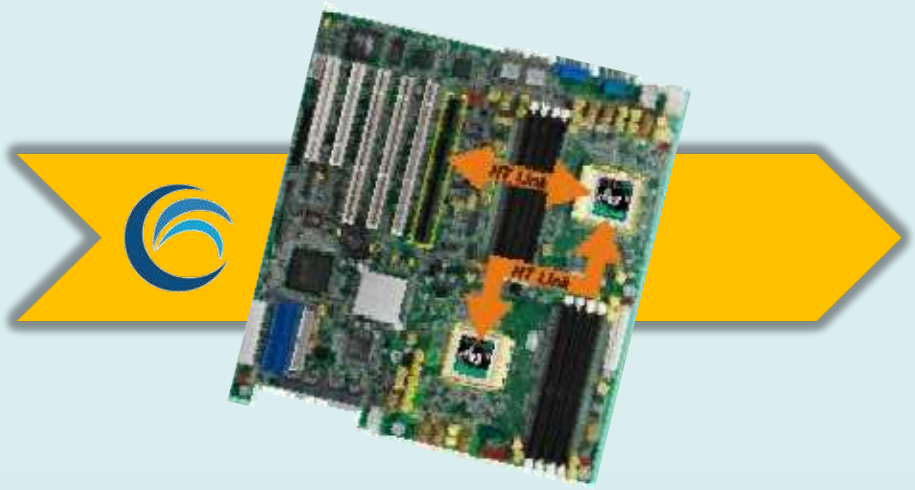
"***Blade Computing with The AMD Magny-Cours Processor***"
Presented by AMD - Pat Conway, Hot Chips 2009

AMD cHT HT-Assist (Probe Filter) (cont.)



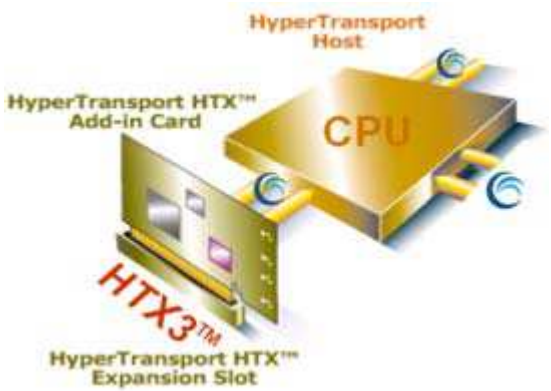


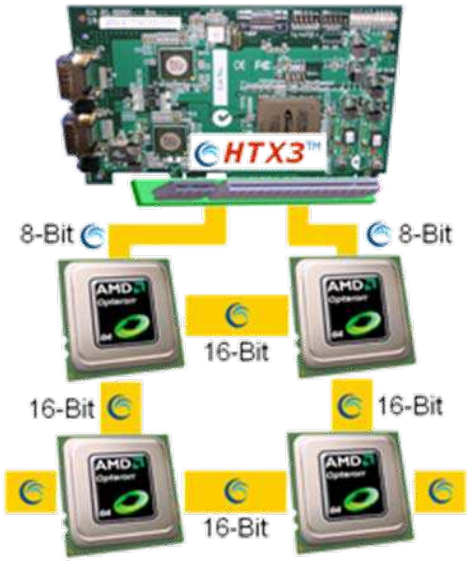
Beyond Motherboards



HTX Connector

- Low Latency CPU-to-High-Perf. Subsystem Direct Connect
- Removes Performance Bottlenecks in Compute-Intensive Data Processing and Acceleration Functions
- Complements PCI-Class Interconnects
- Link Splitting Capability

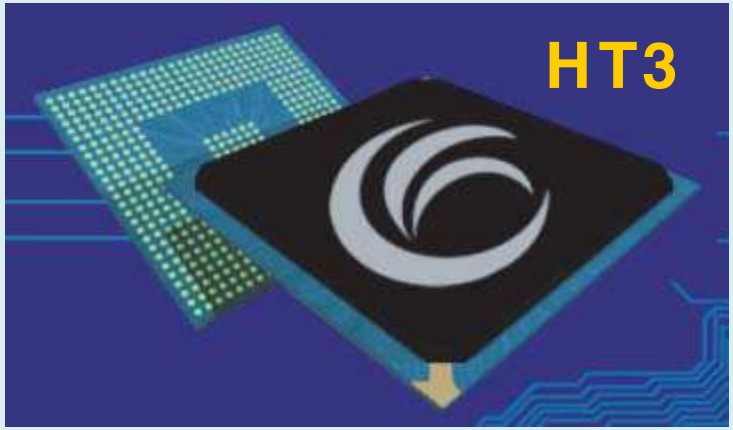




HTX Specification Evolution

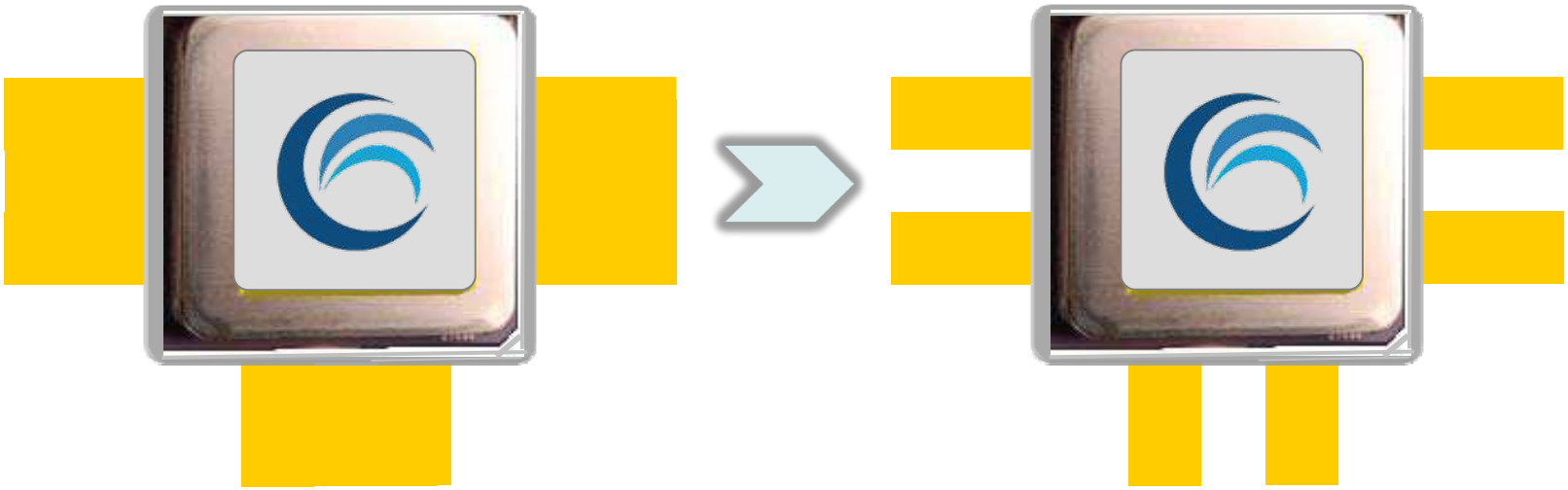
Feature	HTX	HTX3	Notes
Max Clock Rate	800 MHz	**2.6 GHz**	12" Trace length
Max Bandwidth x Lane	1.6 GT/s	**5.2 GT/s**	Bi-directional
Max Bandwidth Aggregate	6.4 GB/s	**20.8 GB/s**	Bi-directional 16-Bit HT link
HT3 Link Splitting Support	NO	**YES**	HT link can be 1x 16-Bit or 2x 8-Bit for mutli-CPU support
HT3 Extended Power Management	NO	**YES**	LDTREQ# Signal Added to participate in x86 power states
Extended FPGA Guidelines	NO	**YES**	Incorporated field-proven recommendations
Full Backward Compatibility	--	**YES**	Level shifters and signal allocation

New in HyperTransport™ 3



HT3 - Link Splitting

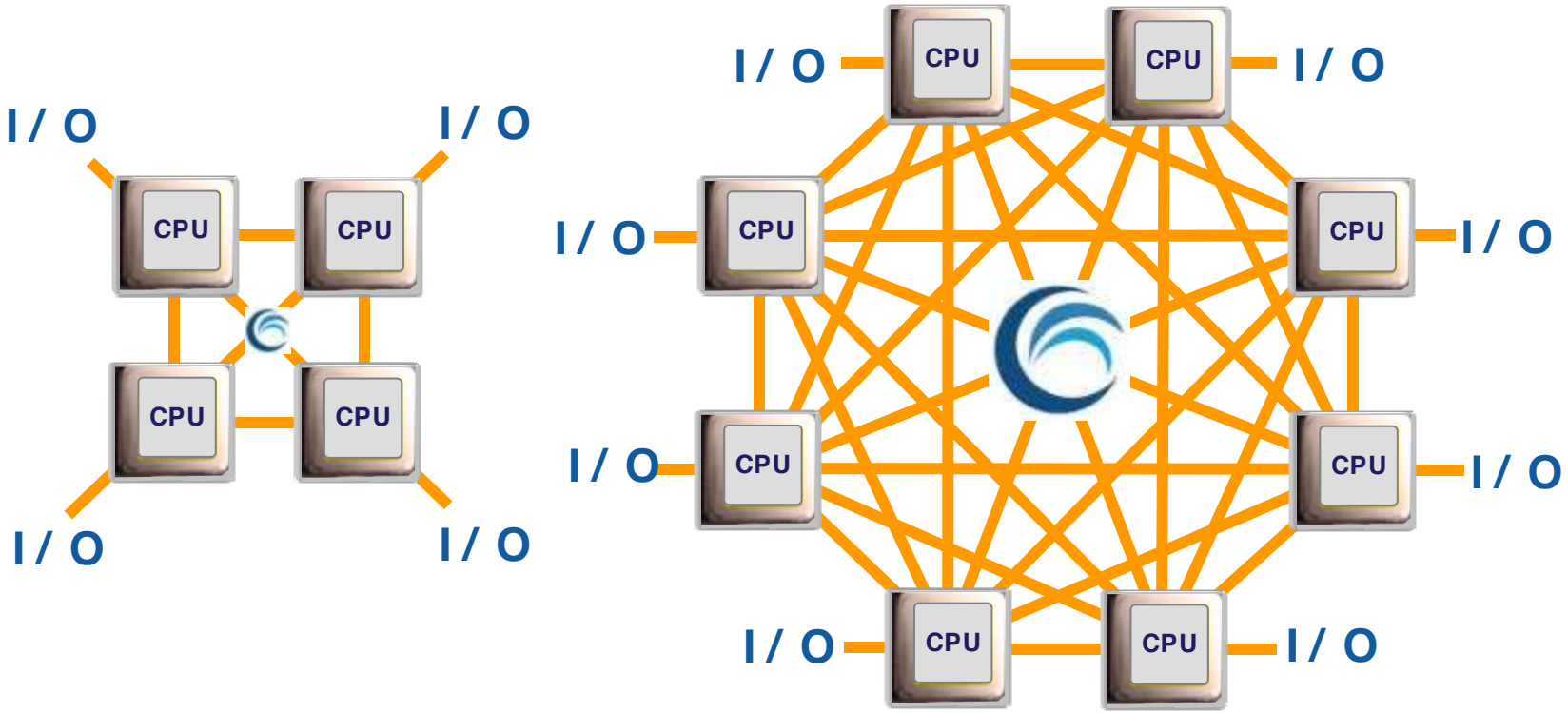
All Links or Individual Links



1x 4-Bit → 2x 2-Bit
1x 8-Bit → 2x 4-Bit
1x 16-Bit → 2x 8-Bit
1x 32-Bit → 2x 16-Bit

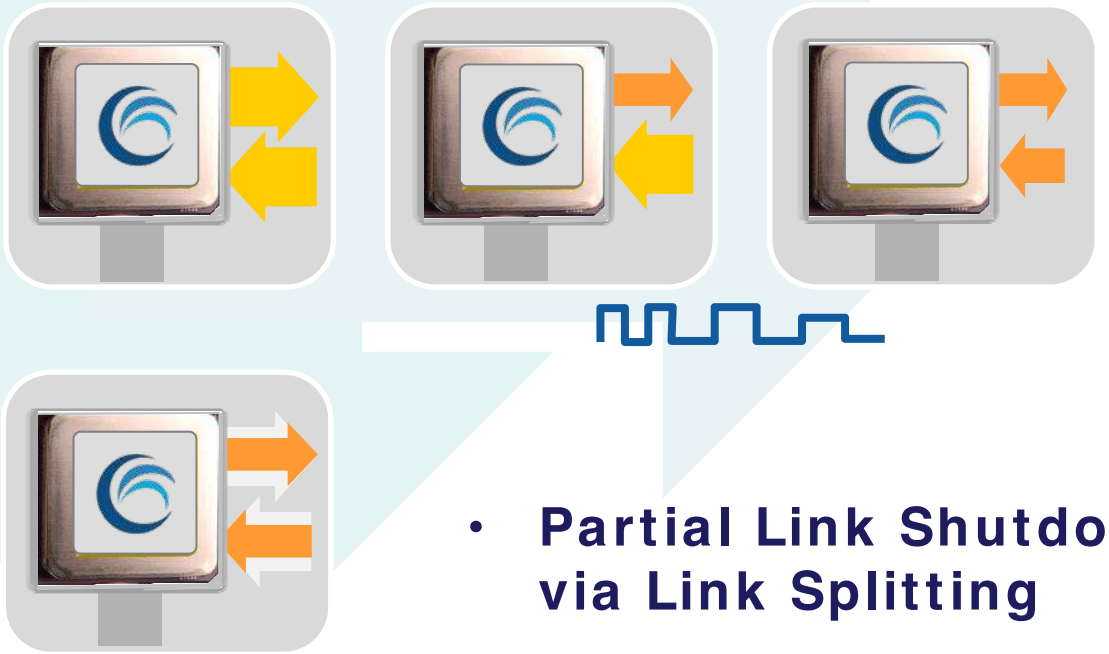
HT3 - Link Splitting (cont.)

Extended SMP Topologies Enabled



HT3 – Dynamic Power Management

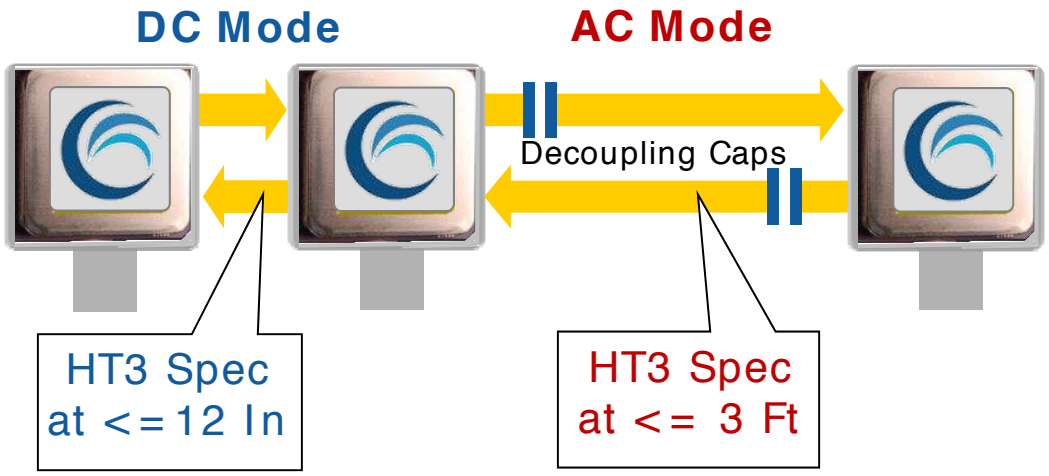
- Dynamic Link Width, Clock Rate and Voltage Scaling

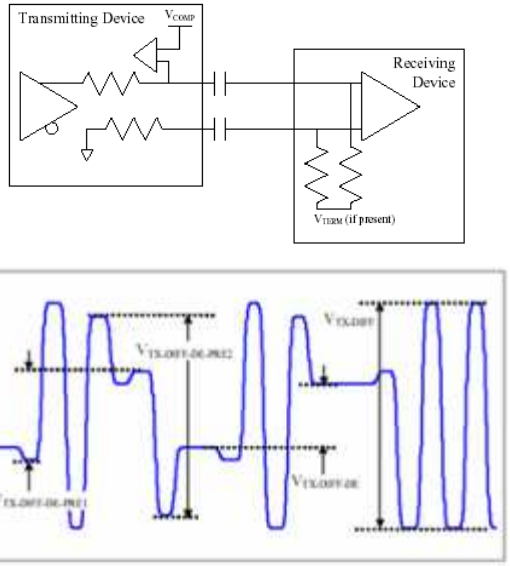


- Partial Link Shutdown via Link Splitting

HT3 – AC Mode (Optional – Enabled if Needed)

- 8b/ 10b Encoding
  - Lower Bandwith, Higher Latency than DC Mode
- DC/ AC Autoconfiguration
- TX Equalization

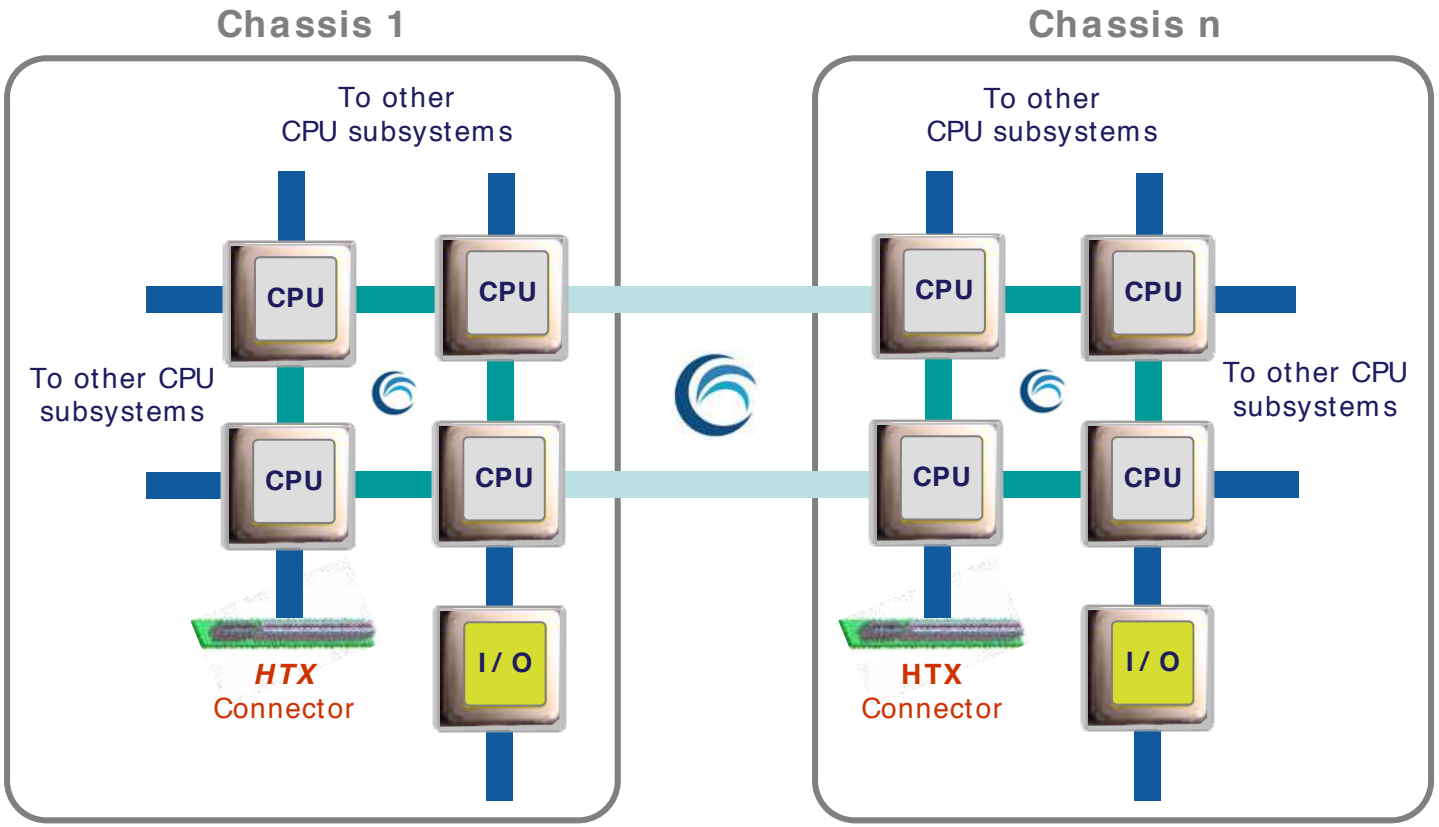




Transmit with Pre- and Post-Cursor De-Emphasis

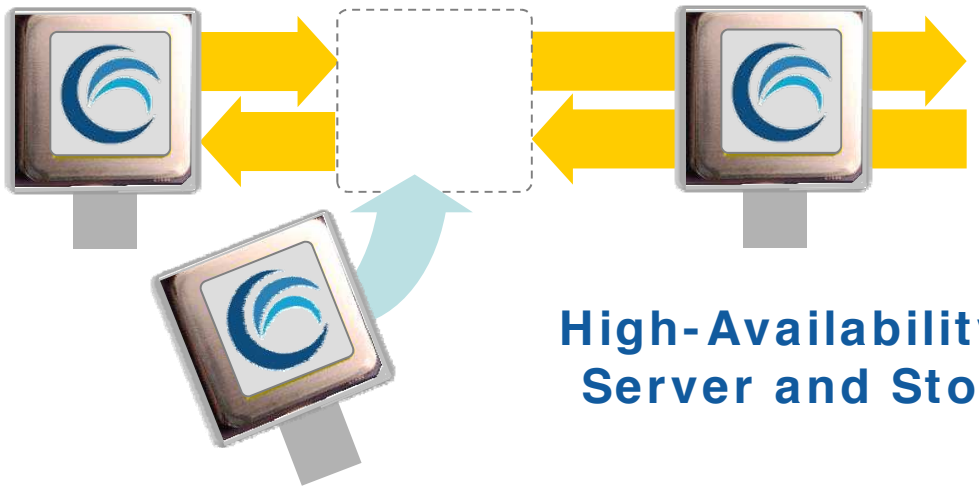
HT3 - AC Mode + Link Splitting

Maximize Multi-Processor Expansion Capability



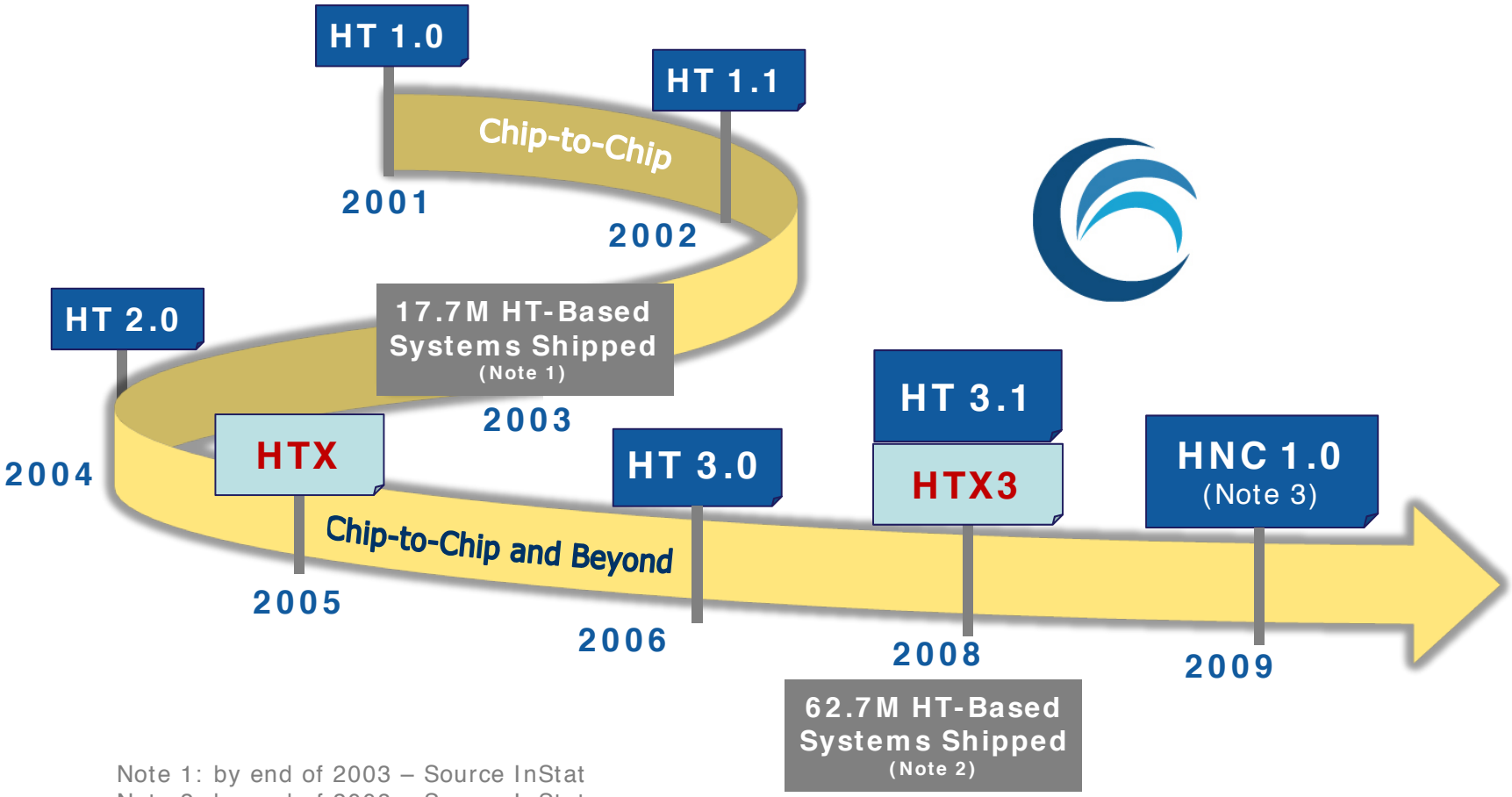
HT3 – Hot Plugging

- Add/ Remove Devices from HT Fabric Without Disrupting Other Operations
  - Defined Link Termination Methods
  - Transaction Termination Behaviors
  - Sync Flood Isolation
  - Link Training Times
- Parameter Configuration Mechanism



High-Availability Applications
Server and Storage Markets

HT Specifications Evolution

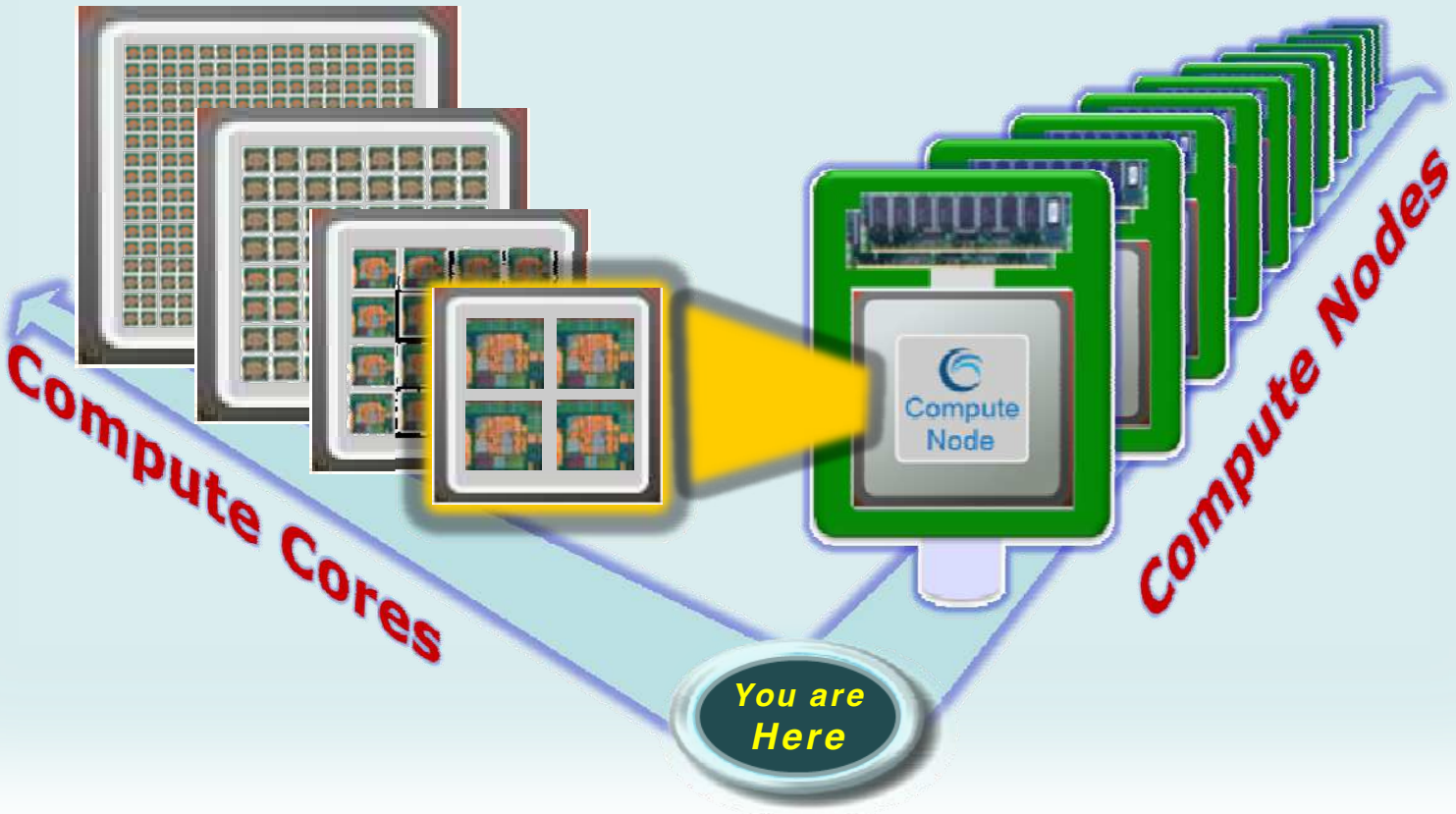


Note 1: by end of 2003 – Source InStat
Note 2: by end of 2008 – Source InStat
Note 3: High Node Count Specification 1.0 - Accessible/Useable by Promoter and Contributor Members Only

HT Specifications Evolution (cont.)

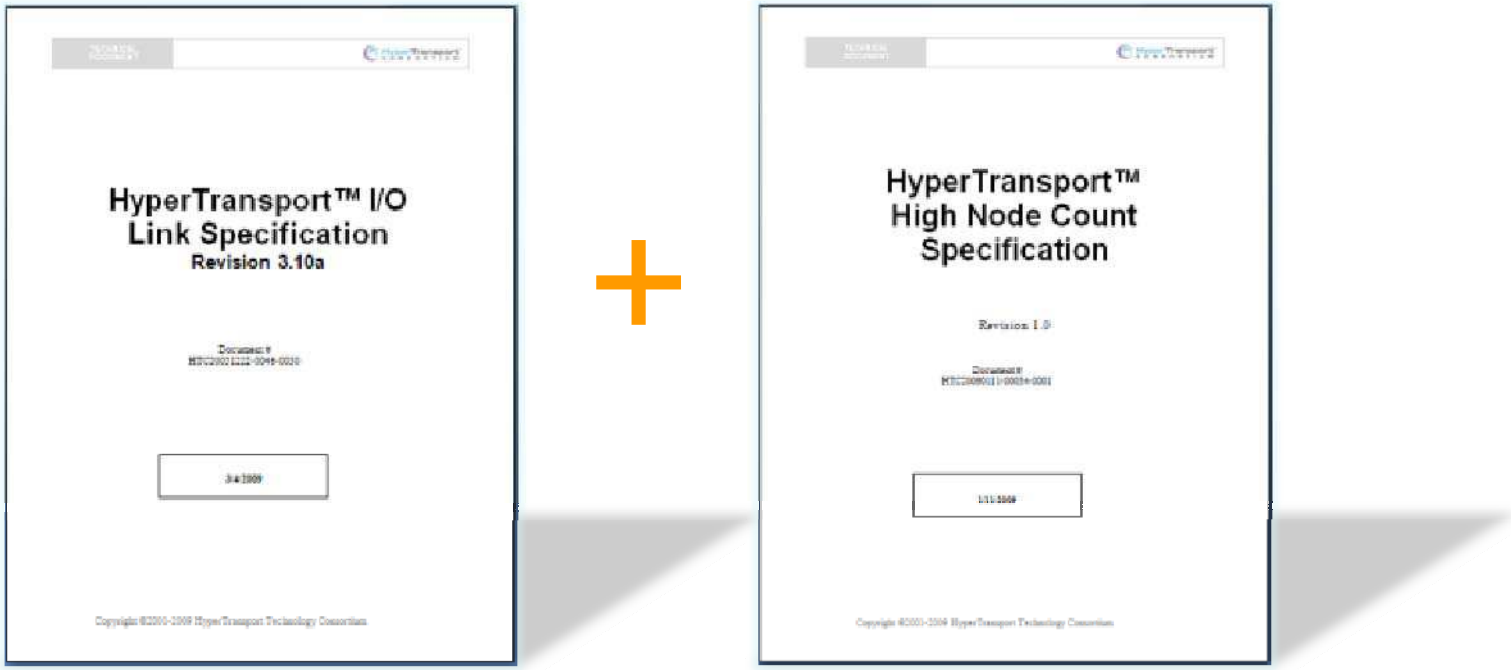
Special Features	HT 1.x	HT 2.0	HT 3.0	HT 3.1
Max Clock Speed	800 MHz	1.4 GHz	2.6 GHz	3.2 GHz
Max Aggregate Bandwidth (32-bit links)	12.8 GB/s	22.4 GB/s	41.6 GB/s	51.2 GB/s
AC Operation – Capacitive Coupling (optional) with AC/DC Autosensing, Autoconfiguration	No	No	Yes	Yes
Link Splitting (un-ganging) Each HT Link Split into 2x Half-Width Links	No	No	Yes	Yes
Hot-Plugging	No	No	Yes	Yes
Dynamic Link Clock/Width Adjustment	No	No	Yes	Yes
DirectPackets™ Data Streaming - +16 Virtual Channels (22 total), Peer-to-Peer Support, Native Packet Handling	HT 1.1 only	Yes	Yes	Yes
PCI Express Mapping	No	Yes	Yes	Yes
Common Features				
2 x Unidirectional, Low-Voltage-Differential-Signaling Links per HT Bus				
Scalable Link Width (2-bit to 32-bit)				
Asymetric Link Support (different link widths)				
Asynchronous Link Operation				
Clock Forwarding (no SerDes latency penalty)				
DC Operation (direct signal coupling)				
PCI, PCI-X Mapping				
Priority Request Interleaving™ (lowest I/O latency)				
Virtual Channels Support (6)				
Error Retry				

Beyond HT3



High Node Count (HNC) Specification Complements HT3

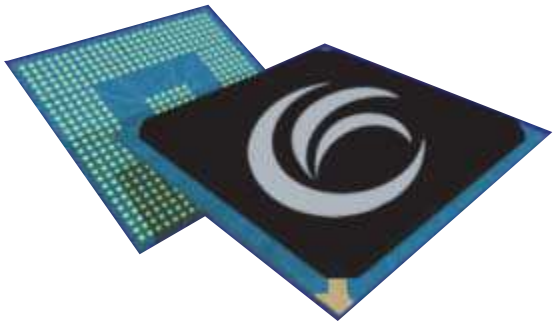
HyperTransport Link Specifications with HNC Extended Addressing Features as Options

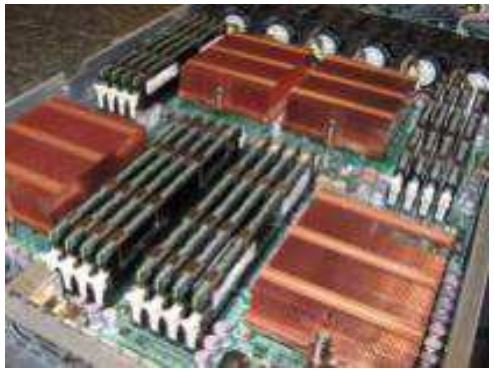


100% Backward Compatible

HNC Model

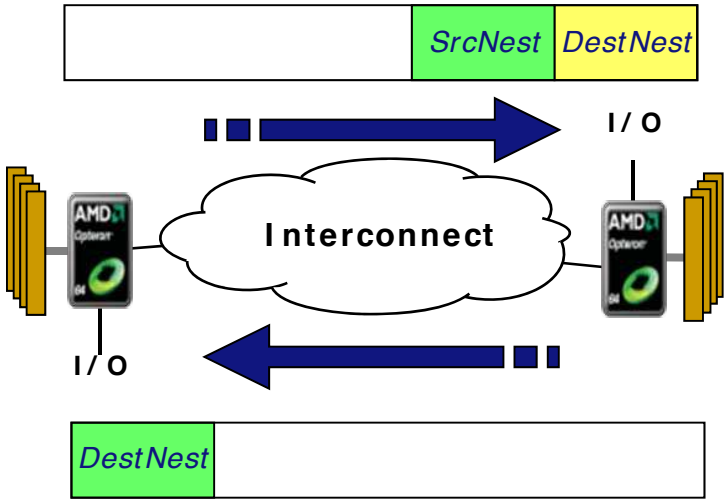
- Physically Distributed, Logically Shared Memory System (PGAS)
- Cache Coherence Not Mandatory, but Multiple Coherent Domains Can be Part of it
- Nest: Each HNC-Clustered Component





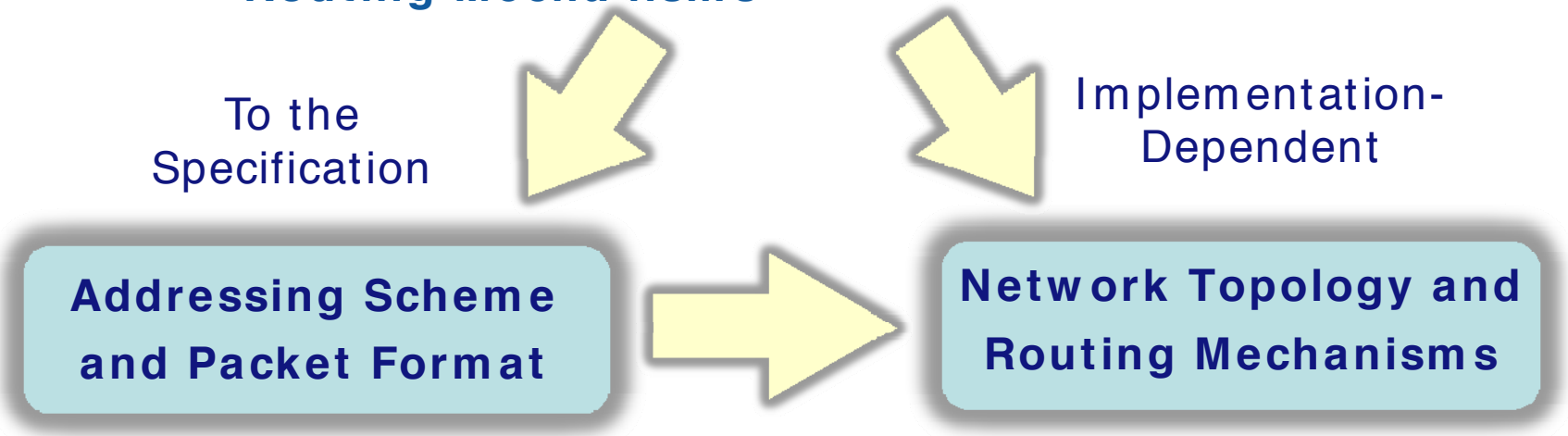
HNC Model (cont.)

- DestNest Identifier Location in Request and Response Packets was Carefully Designed to Minimize Routing Time
- Optimized for Most Common Cases
- Minimal Extra Latency and Bandwidth Use



HNC Extensions

- **Interconnecting Large Numbers of Hosts Requires:**
  - **Addressing Scheme: Ability to Address Large Number of Devices. Affects Packet Format**
  - **Network Topology: Support for High Connectivity Topologies that Enable More Concurrent Transmissions**
  - **Routing Mechanisms**



HNC Specification Removes

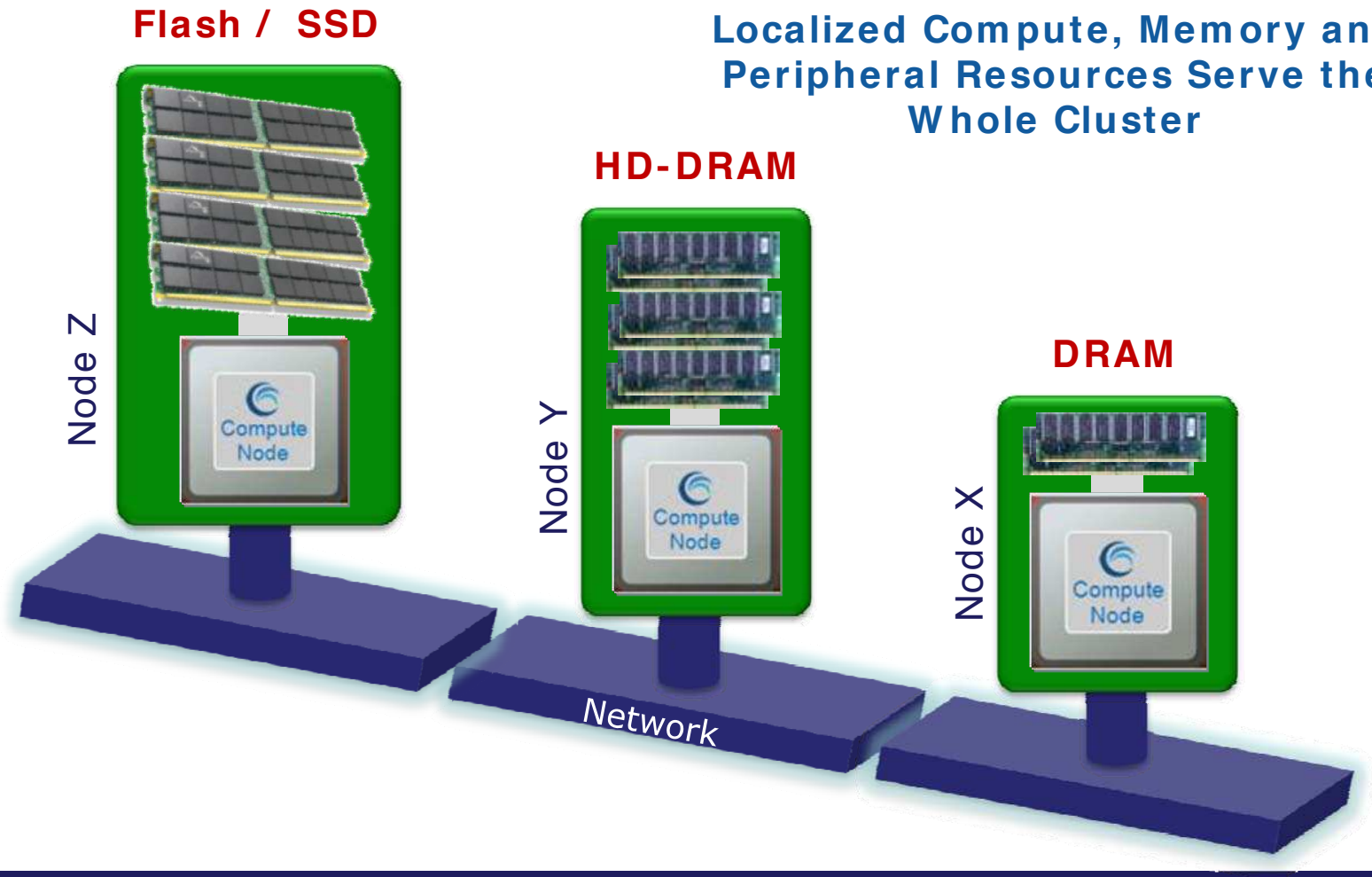
Latency Overhead of Message-Passing Protocols

Through HyperTransport Protocol Encapsulation

Using System-Wide HNC Addressing Scheme

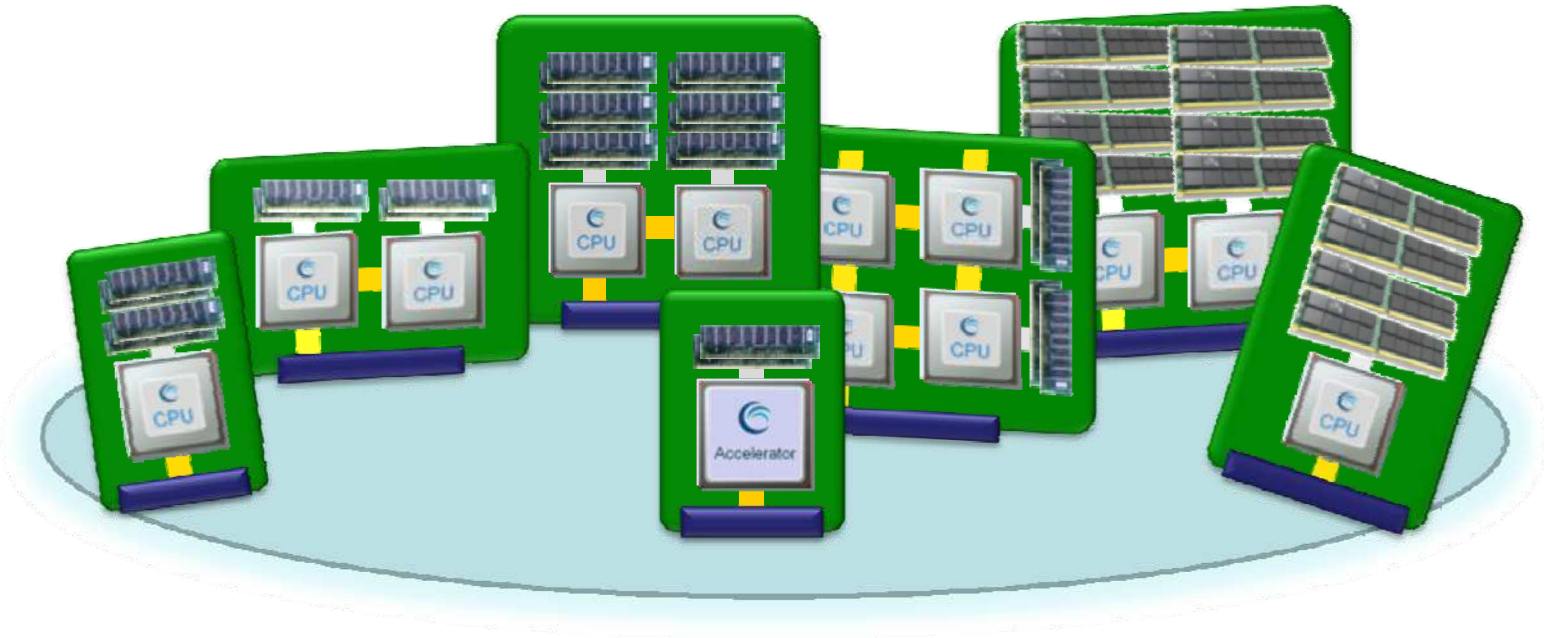
HNC Specification Enables

Scalable Global Resource-Sharing & Partitioning



HNC Specification Delivers

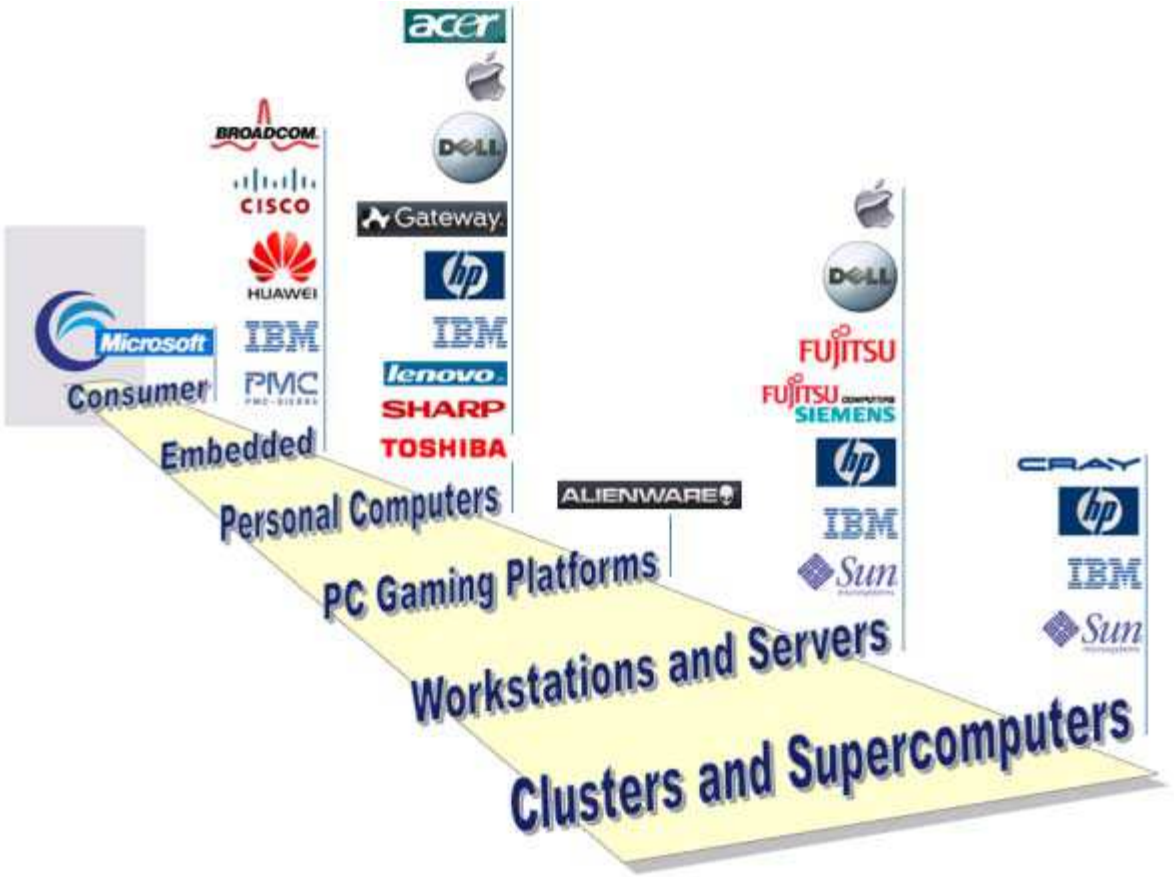
Resource-Sharing & Modularity Resulting in
Less Over-Provisioning
Best Resource Utilization
Maximum Energy Efficiency



Beyond Conventional

HT Everywhere

HyperTransport Empowers More Processor Architectures and Market Segments than its AMD Roots May Suggest

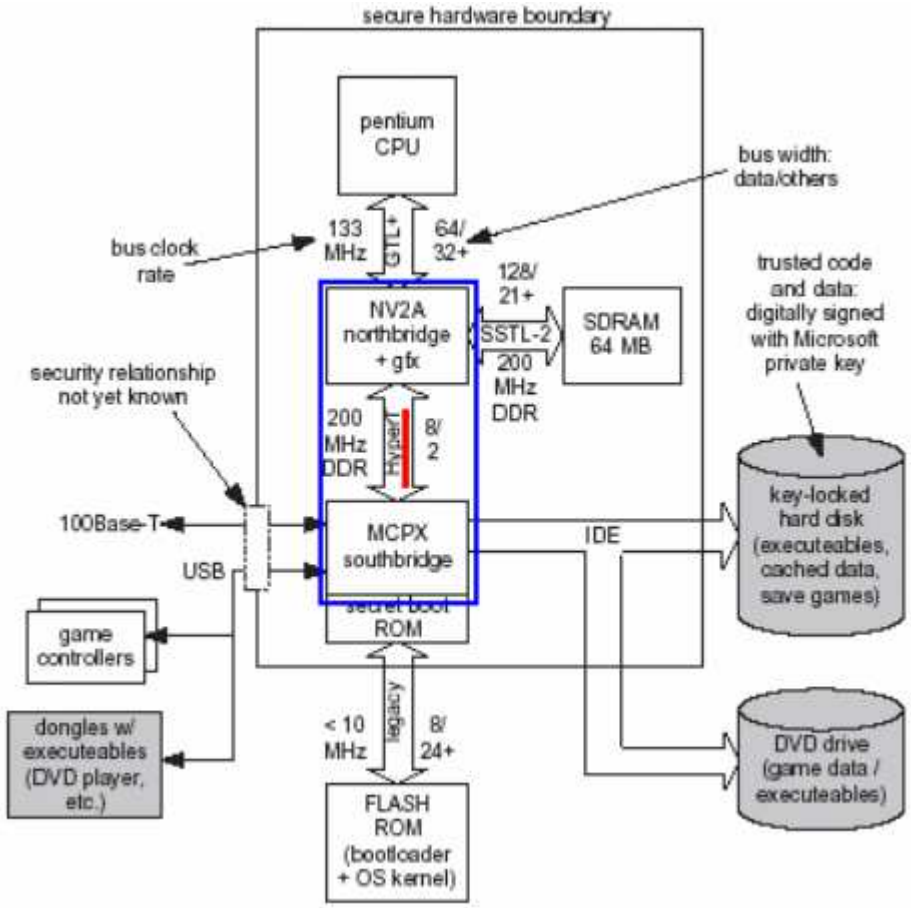


In Video Gaming

Microsoft®

XBox

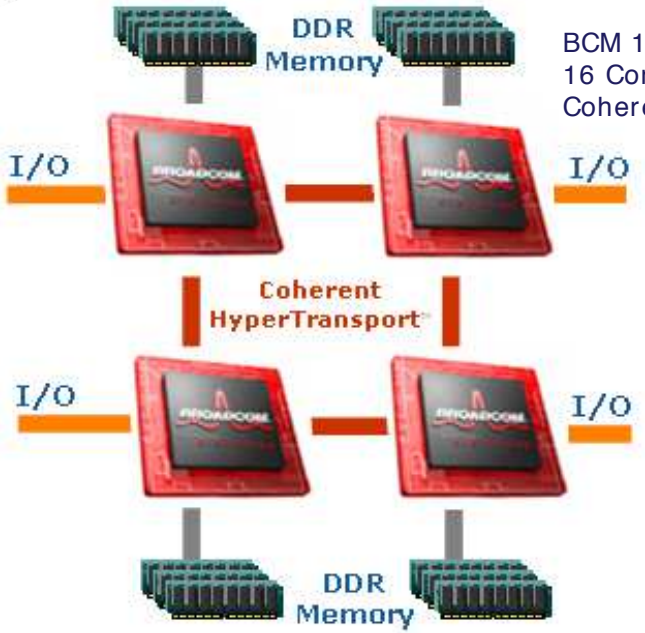
"If the HyperTransport link in the Xbox was a pipe carrying water, and every bit of information equalled one gallon, the pipe would fill up the Pacific Ocean every second!"
NVIDIA Web SIte



In Communications Processors

BROADCOM.

BCM1125H 64-bit MIPS Processor
BCM 1250 64-bit MIPS Processor
BCM 1280 64-bit MIPS Processor
BCM1480 4-Core 64-bit MIPS Processor



BCM 1480 x 4
16 Cores
Coherent HT SoC Cluster

PMC
PMC-SIERRA

RM9150 64-bit MIPS Processor

RM9200/9100 64-bit MIPS Dual/Single Processor

RM11200 64-bit MIPS Multiprocessor

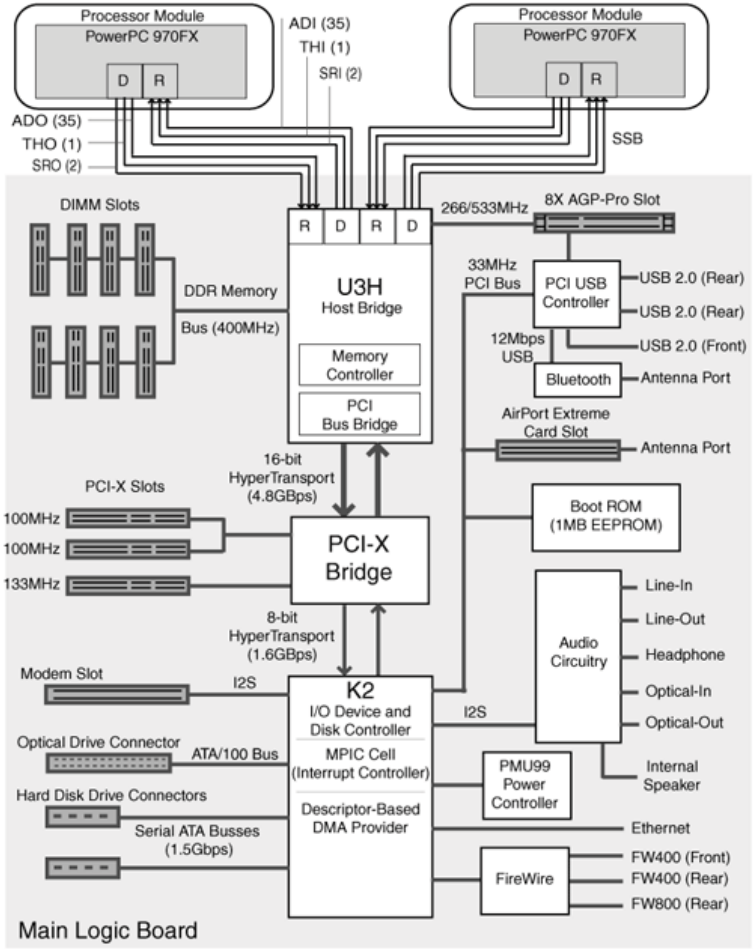
RM9220/9224 64-bit MIPS Multiprocessor

In PowerPC Computing

iMac G5

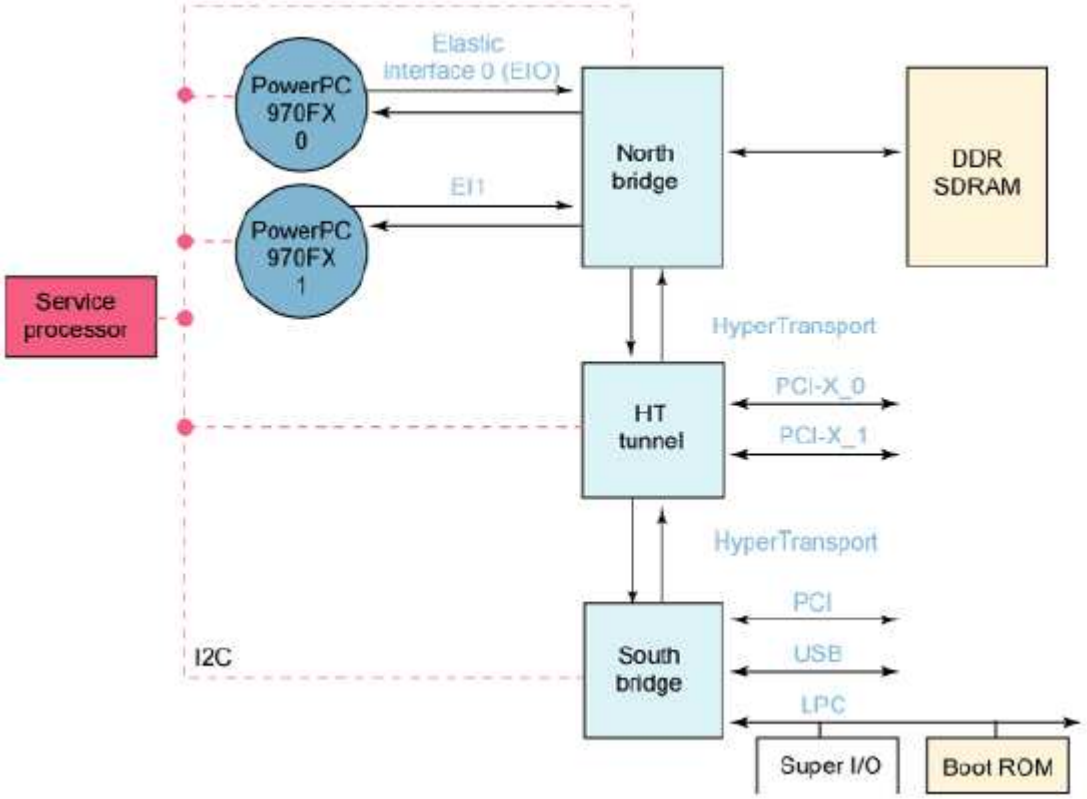
Power Mac G5

Xserve G5

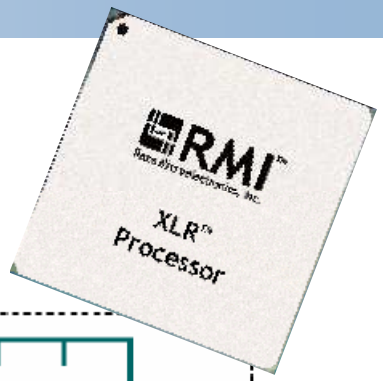


In PowerPC Embedded Systems

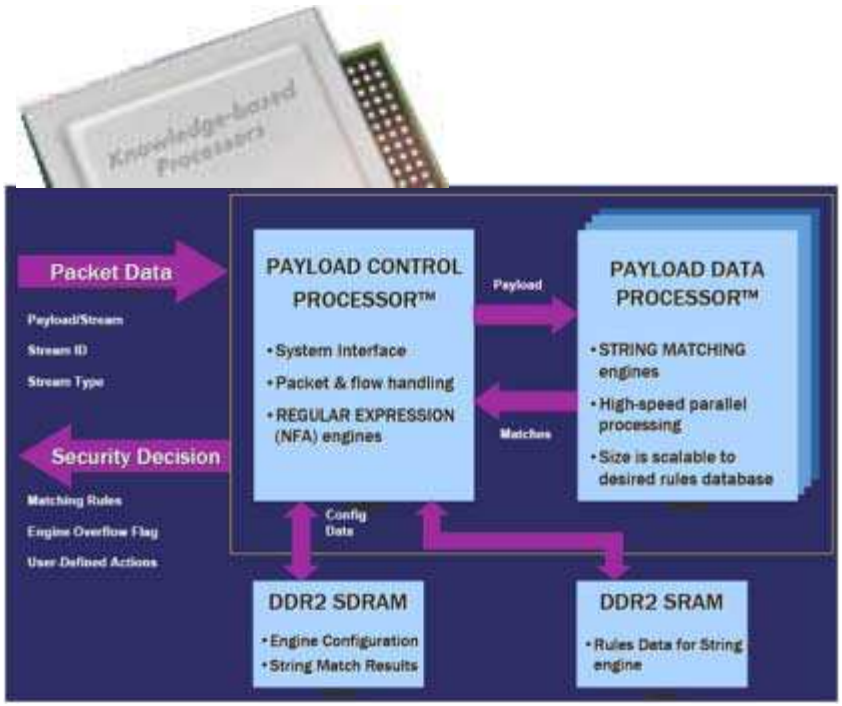
IBM



In Knowledge-Based Processors

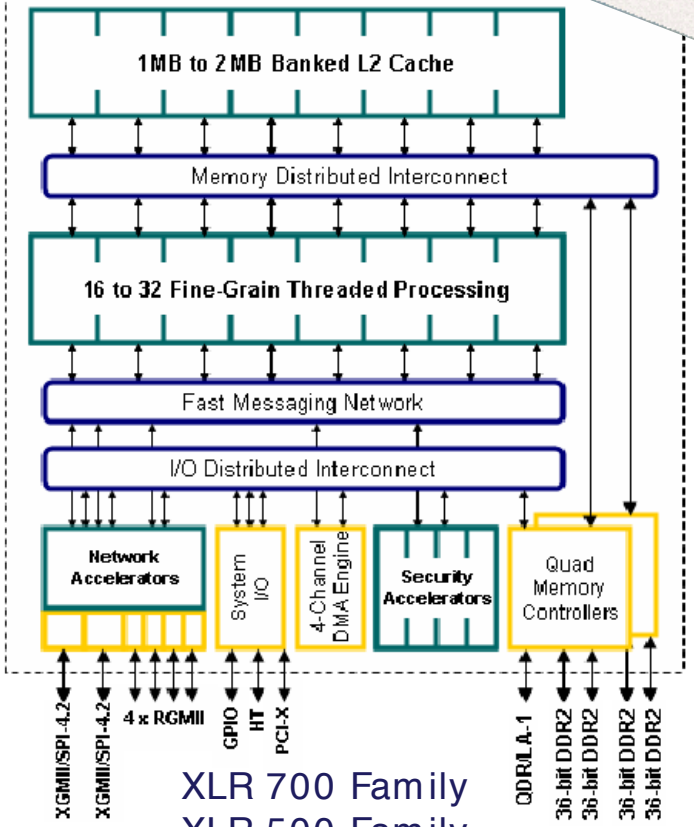


NETLOGIC MICROSYSTEMS



NETL7

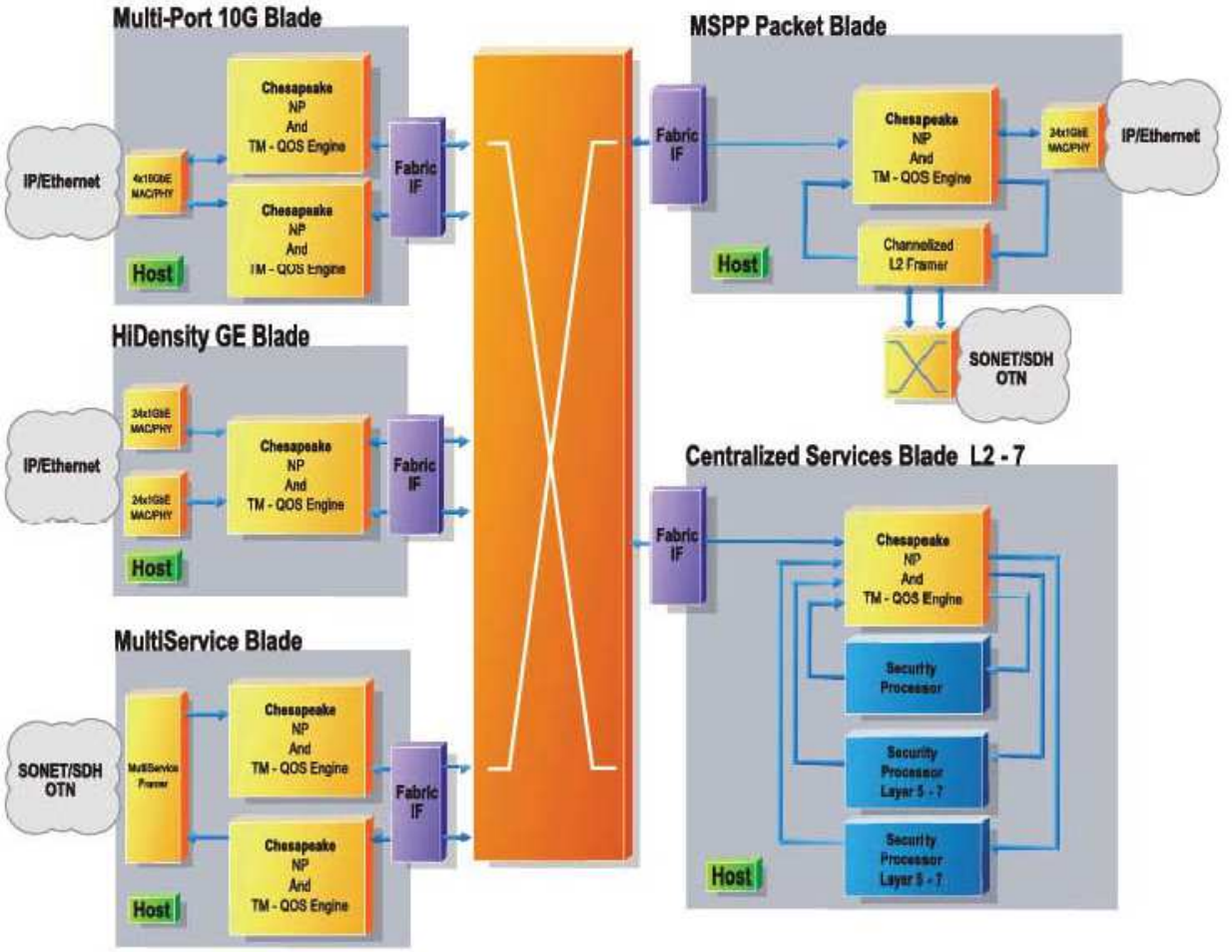
RMI
Raza Microelectronics, Inc.



XLR 700 Family
XLR 500 Family

In Network/ Media Processors

Chesapeake



In Reconfigurable Accelerators

DRC Computer

RPU 110
RPU 100
Opteron Socket Plug-In

Celoxica
pure acceleration

RCHTX
Accelerator Board
HTX Slot Connector

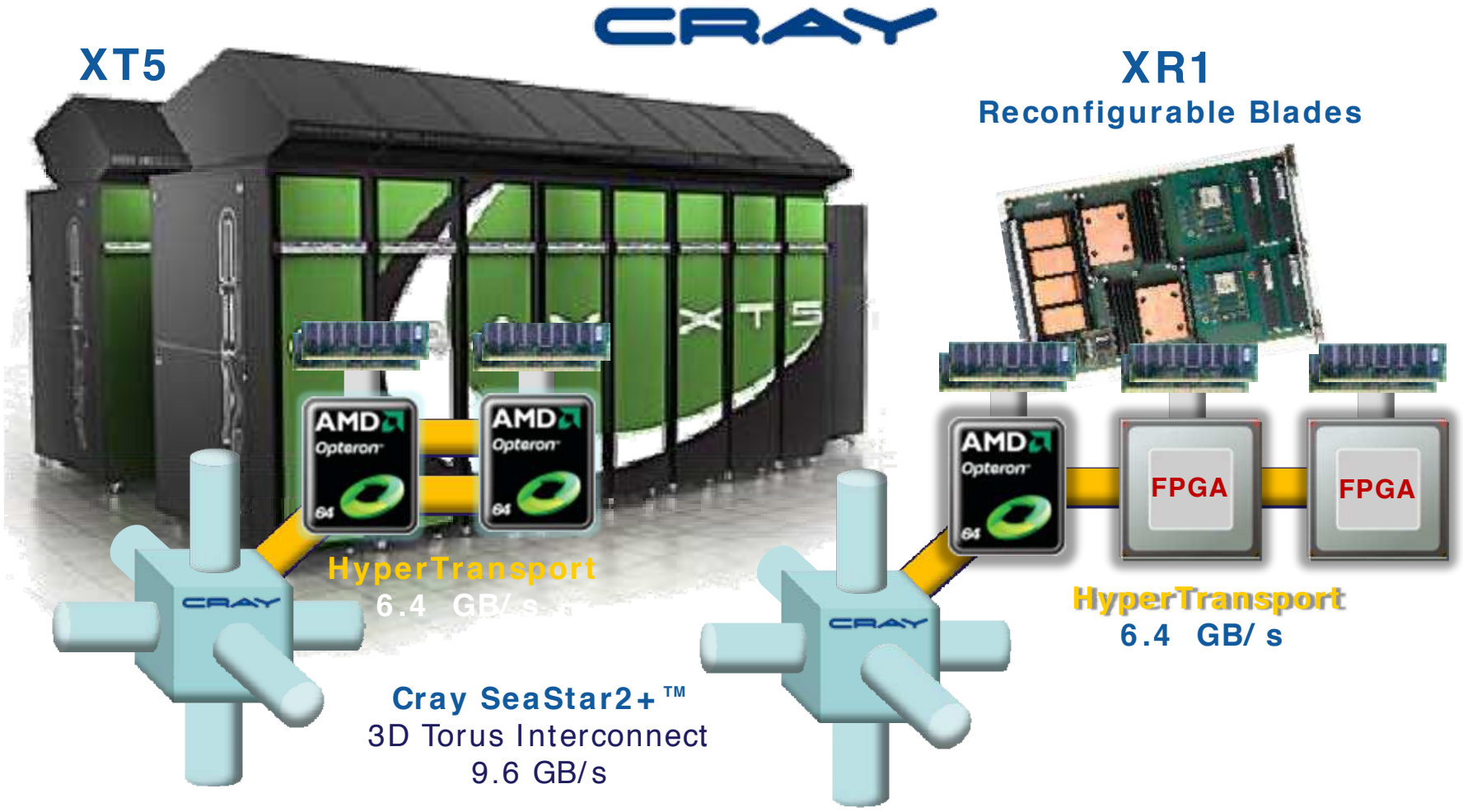
XtremeData, Inc.
COMPUTING REDEFINED

XD 2000F
XD 1000
Opteron Socket Plug-In

In Routing Processors

Classification and action engine capable of consistent association between CPU cores and their data streams, resulting in efficient multi-core system load balancing of network traffic processing and lower power consumption

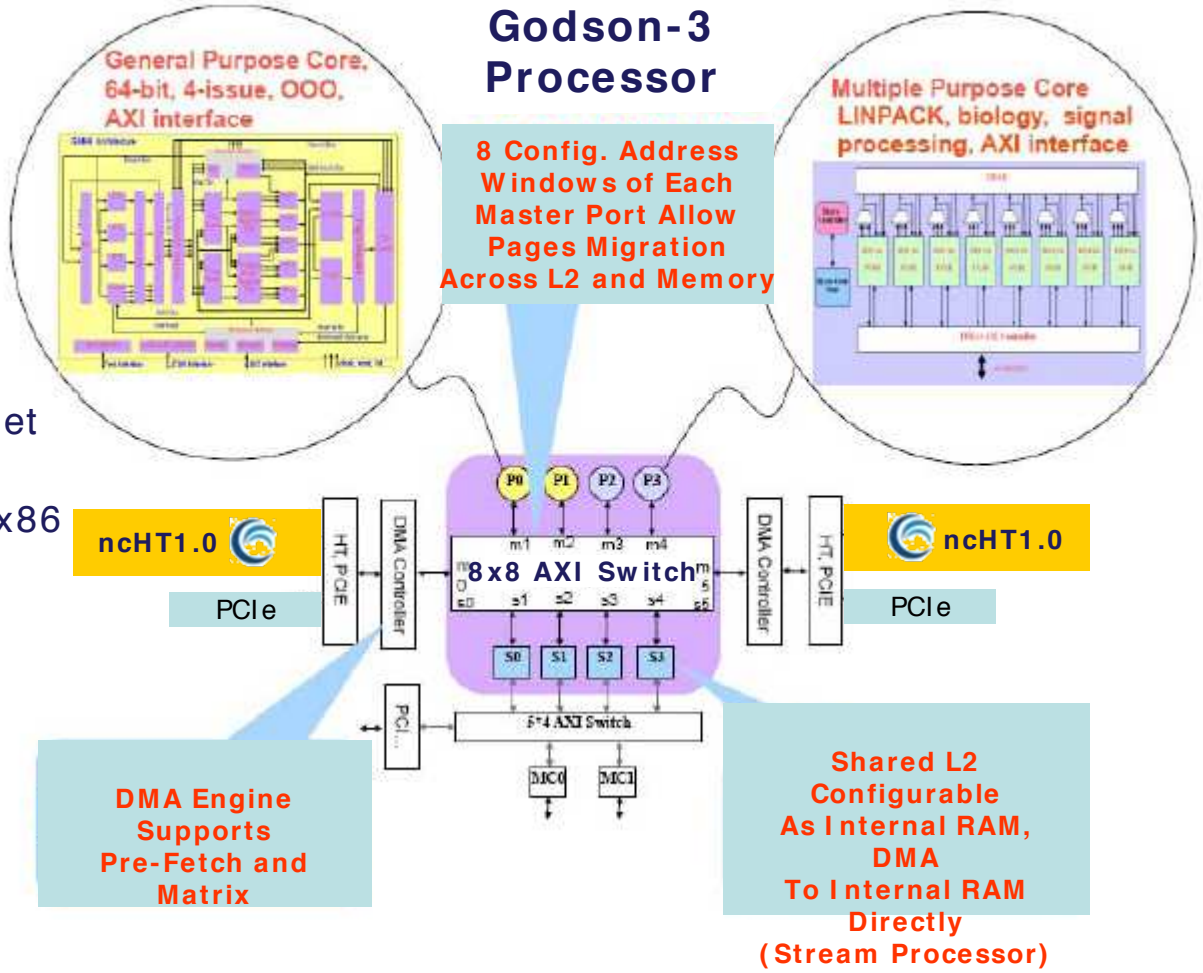
In State-of-The-Art Supercomputers



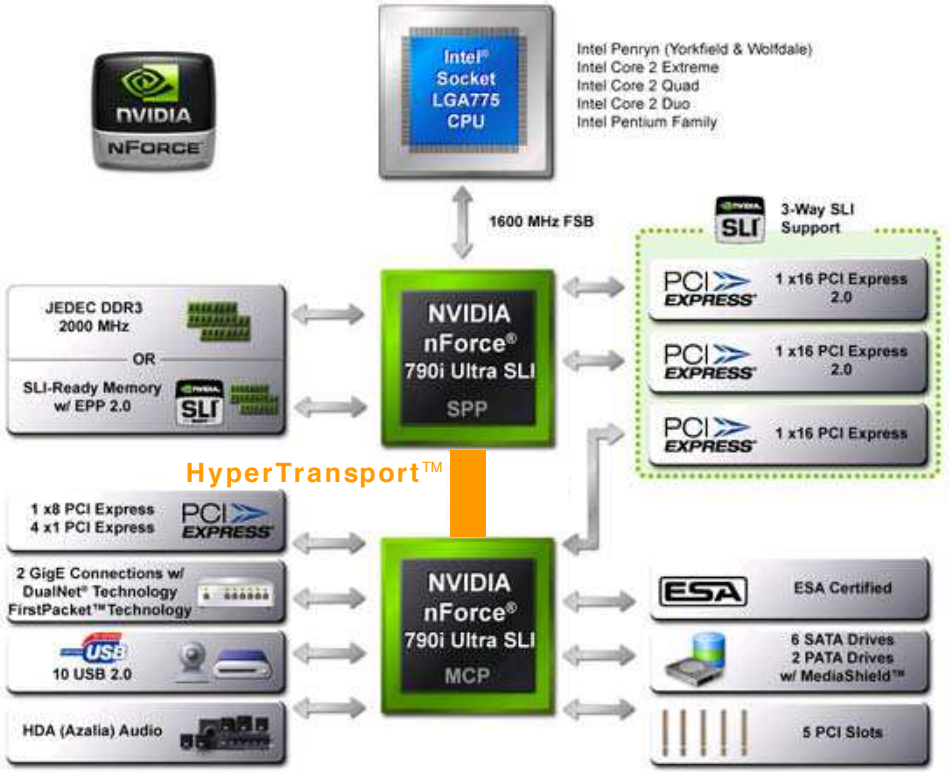
In x86 Emulation Processors

中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

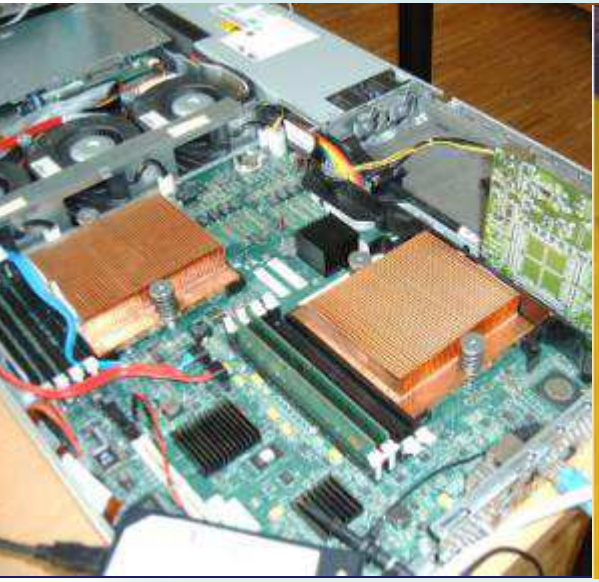
- Petascale Performance Target
- 4-Core MIPS64-Based with 200+ More Instructions for x86 Translation and Acceleration
- 16 GFLOPS at 1GHz
- 10W of Power

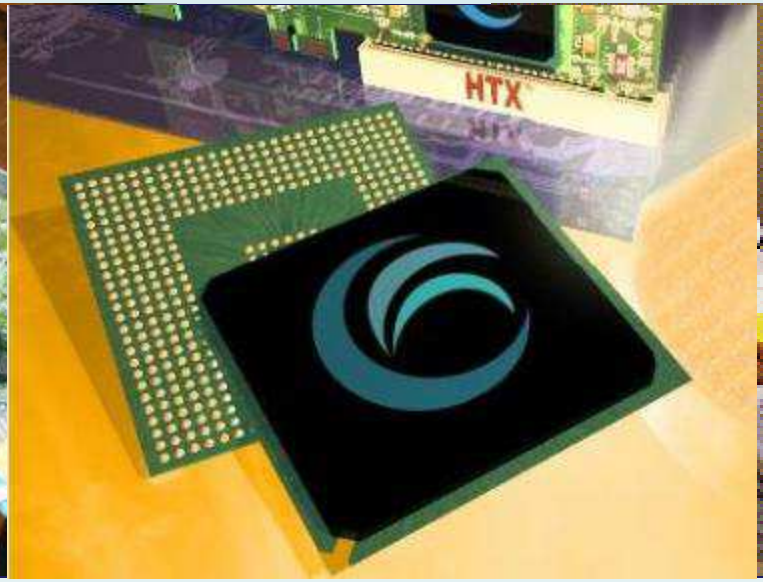


Also in Computing Platforms in which…
One Would Not Expect to Find HyperTransport!



HyperTransport™ CONSORTIUM



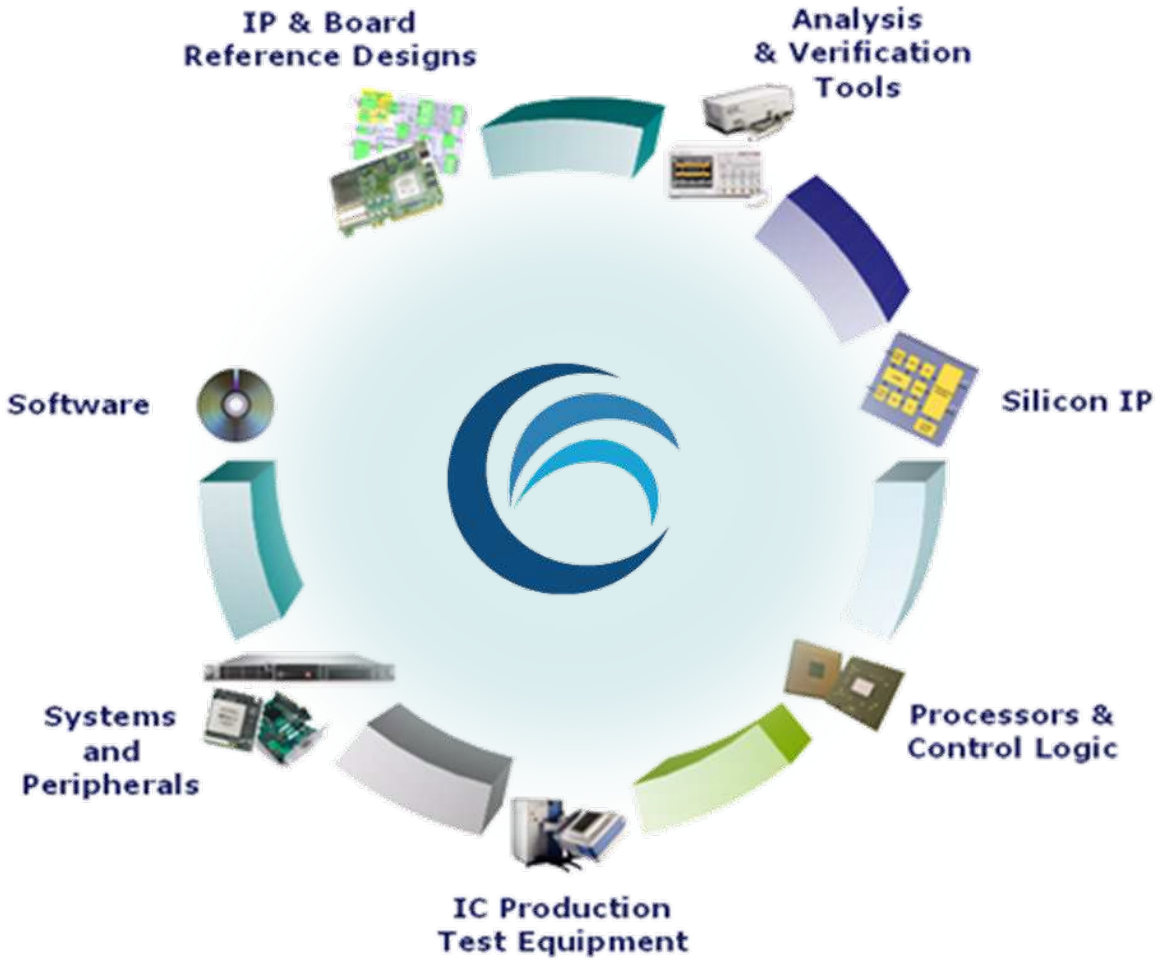


Vibrant Open Standard Consortium

- Founded 2001
- Controls, Promotes and Licenses HyperTransport Technology to Global Industry
- 60 Members-Strong, Including Technology Leaders

Vast HT Products Ecosystem Fosters Technology Strength

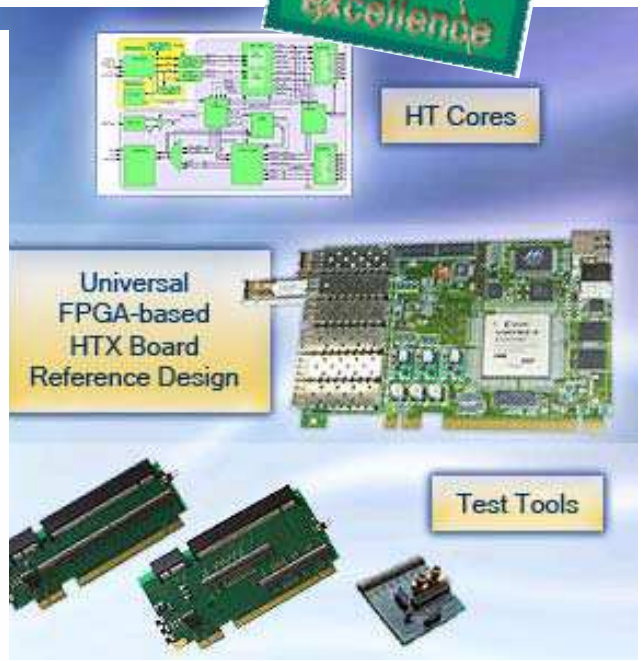
From IP to Software



From Technology Licensor to Members Business Partner

- HT Enablement
- Product Validation Services

Online Design, Simulation and Test Support



All About HT on Consortium Web Portal

www.hypertransport.org

Technology

Member Login

Members General & Extended
HT Technical Support Database

Public Web Portal

HyperTransport Specifications

Design Support

White Papers

Members Material

HT Link

HT High Node Count

HTX and DUT Interface Connectors

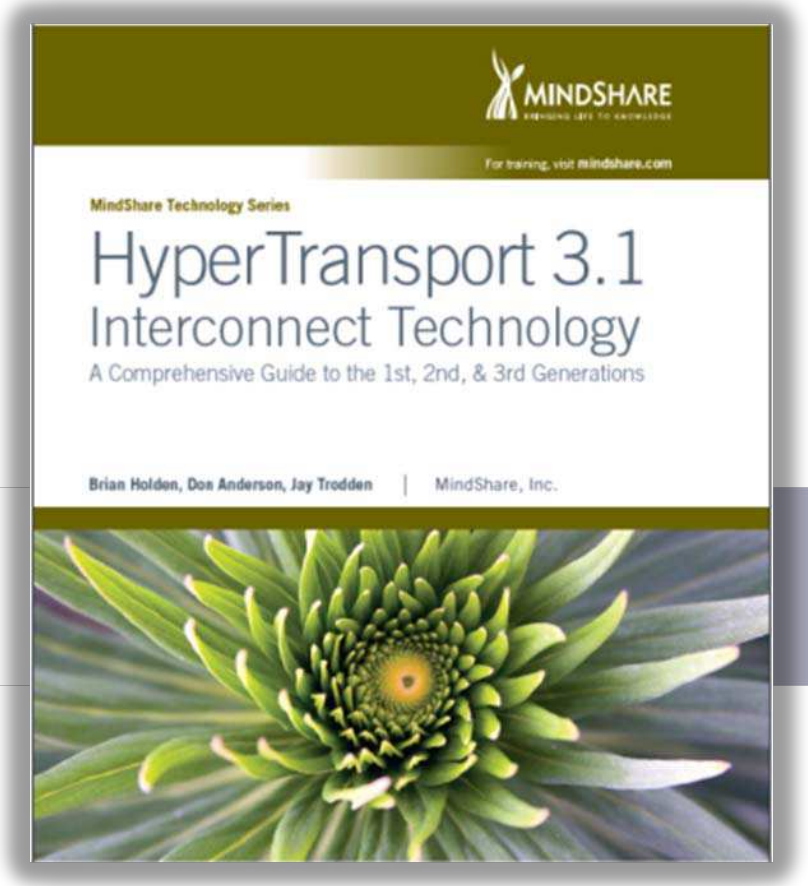
New HyperTransport Book

700-Page Must Have HT Tutorial

Covers All HT Link and HTX Specification Releases

Available Online from MindShare in Paper and eBook Formats

www.mindshare.com



Additional Information

- HyperTransport™ I/O Link Specification 3.10a
- HyperTransport HTX™ Slot Connector Specifications
- HyperTransport™ High Node Count Specification 1.0
- White papers
- Hennessy & Patterson, "Computer Architecture: A Quantitative Approach", 4th Edition, Appendix E
- Dally & Towles, "Principles and Practices of Interconnection Networks"
- Duato, Yalamanchili & Ni, "Interconnection Networks: An Engineering Approach"
  - All three books published by Morgan Kauffman

Thank You!