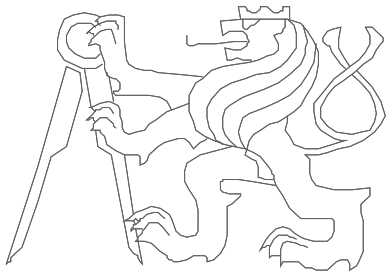
Pokročilé architektury počítačů

Realizace moderního IO podsystému (NUMA architektury)

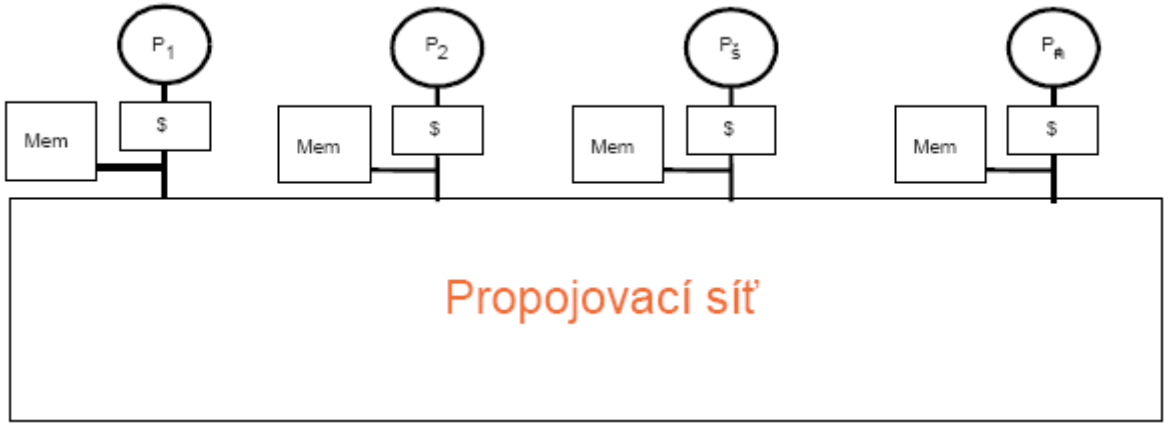


České vysoké učení technické, Fakulta elektrotechnická

Co je úkolem?

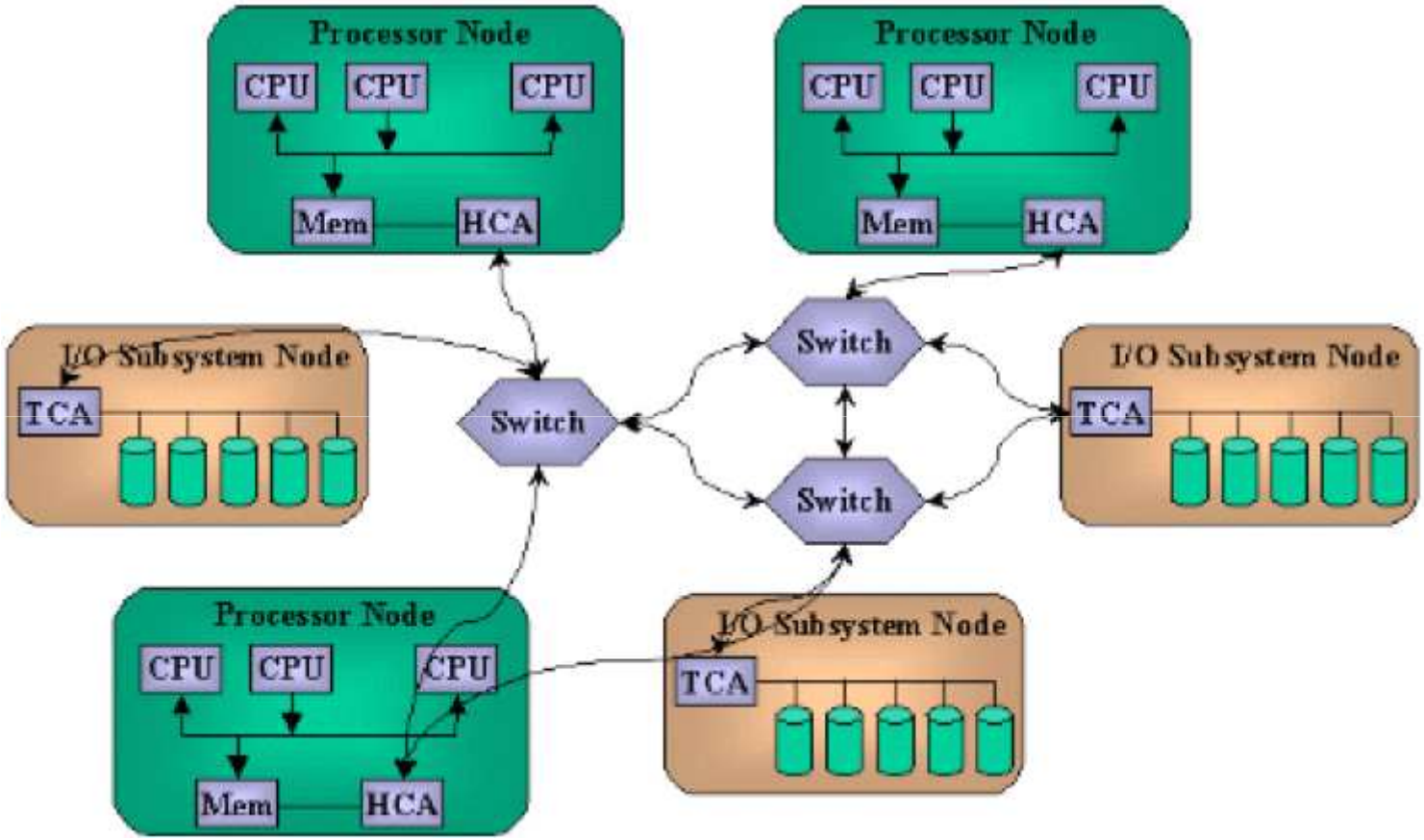
- Propojit jednotlivé části výpočetního systému
- Požadavky:
  - Vytvořit optimální datové cesty hlavně pro nejdůležitější periferie (vnější paměti),
  - zajistit koherenci paměťového (nezapomeňte NUMA!) podsystému.
- Možnosti řešení:
- S ohledem na závislost cena/výkon existuje hranice výkonnosti, kdy
  - datové cesty už není možné sdílet.

Nástin systému:



- Hlavní paměť je distribuovaná, části jsou u jednotlivých procesorů (systém NUMA) ale globálně fyzicky adresovaná a SP (cache) je koherentní
- Jak taková „Propojovací síť“ vypadá?

Propojovací síť může, například, vypadat takhle:



Řešení: InfiniBand!

- Kanálově orientovaná přepínaná struktura se sériovou komunikací.
- 1Gb/s – 30Gb/s.
- Podobnost se sítí? – ona to síť je!
- Základní prvky – HCA, TCA, Switch/Router.

InfiniBand - terminologie

- HCA - Host Channel adapter. Připojuje procesor/y,
- TCA – Target Channel adapter. Připojuje koncová zařízení.
- Switch/Router – Přepínač/Směrovač.
- Komunikace = zasílání zpráv.
- Zpráva = pakety.

InfiniBand

- Switch = 16b adresa, 48k adres zařízení, 16k address multicast,
- Router ~ IPv6
- Spojení – 1 servisní vždy, 15 aplikačních přes jeden kabel.
- Zaručená přenosová šířka (podobnost s ATM).

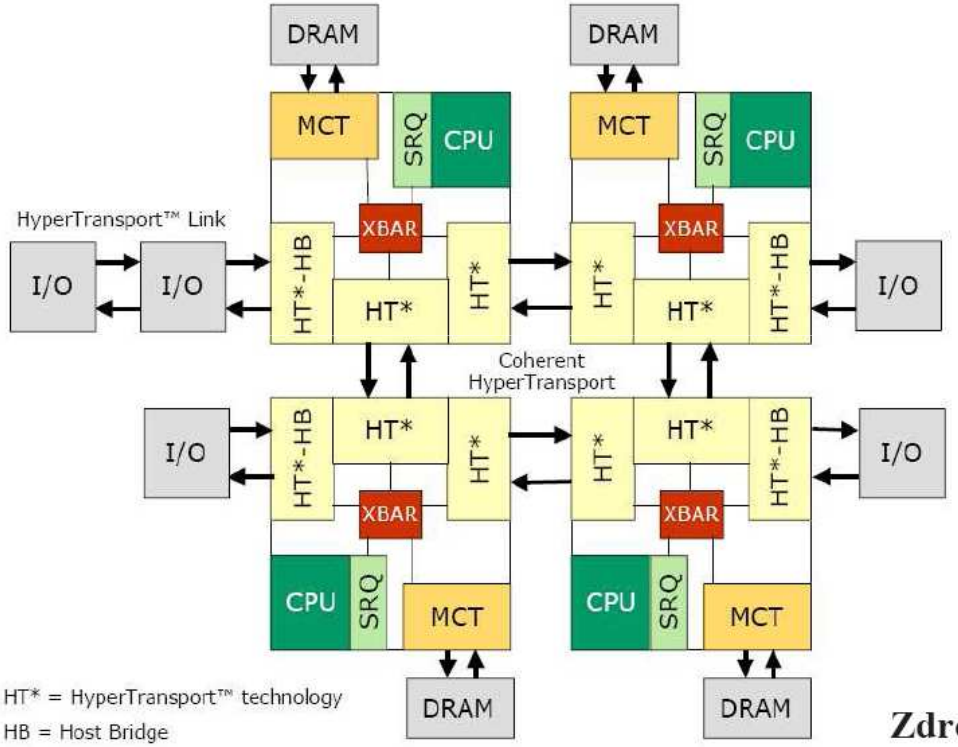
Infiniband – technické parametry

	InfiniBand
data signaling speed	1,25 GHz
bandwidth	500MB/s - 6 GB/s
link width	2, 8, or 24
Protocol used	IPv6
Package size	68-4096 bytes
Memory model	non-coherent
Environment	network
Plug & Play	Yes

Infiniband – konektor kabelu:

Jediné řešení našeho problému?

- Ne!
- Existuje řešení automaticky zajišťující koherenci. Jaké?
- HyperTransport

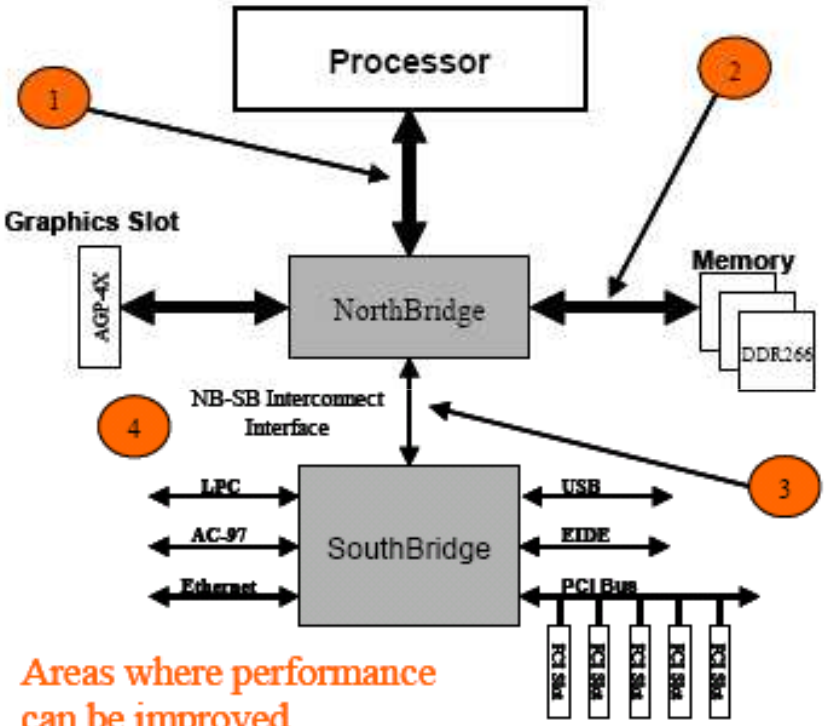


Zdroj: AMD

Konkrétně?

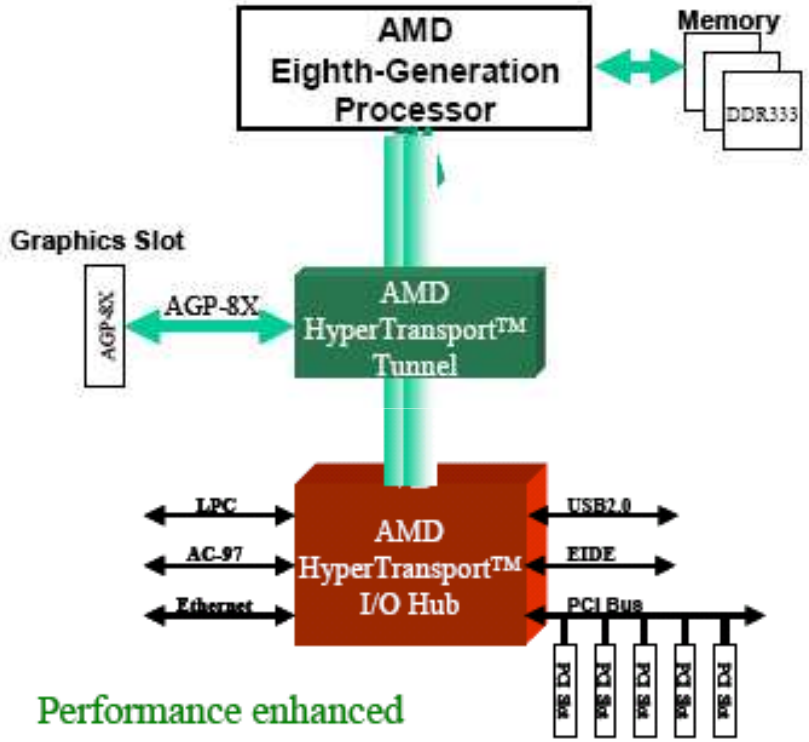
- HyperTransport (HT) technologie je rychlý dvoubodový spoj s nízkou latencí (zpožděním) navržený s ohledem na vysokou komunikační rychlost. Propojuje integrované obvody v počítačích, serverech, vestavěných systémech a síťových a telekomunikačních zařízeních až 48x rychleji, než je to běžné v jiných technologiích. Je obvykle integrován přímo v procesoru.

Existing System Architecture



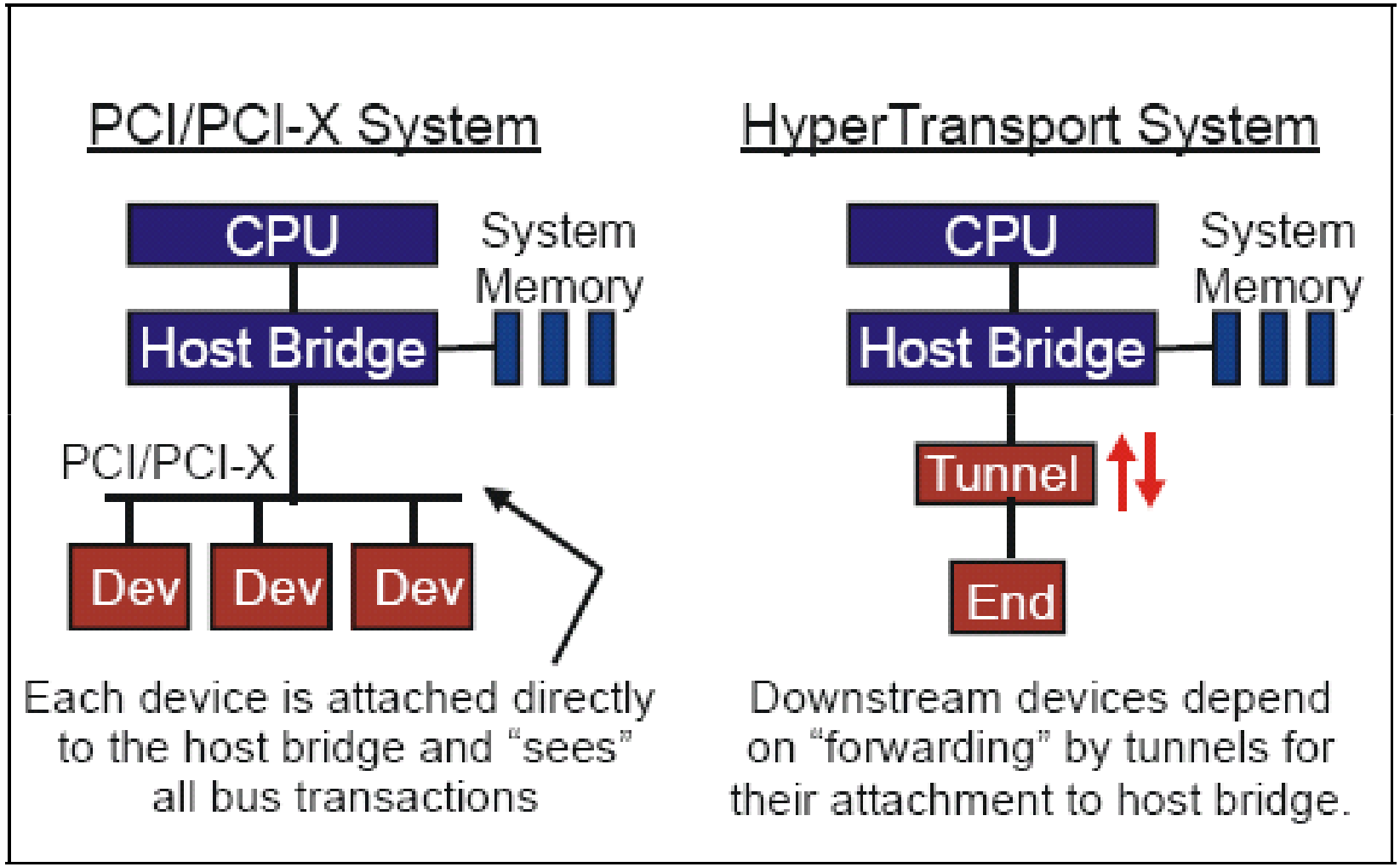
Areas where performance can be improved

HyperTransport™ System Architecture

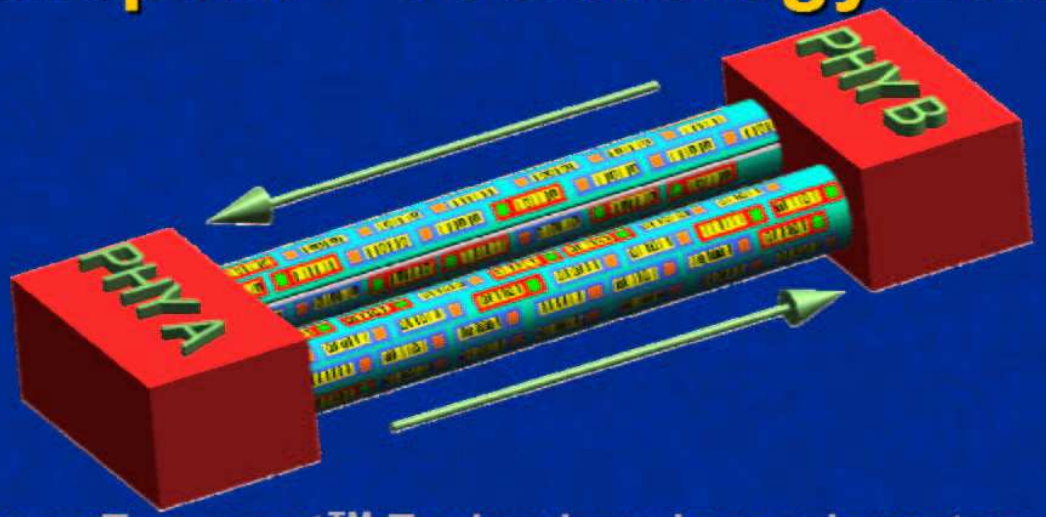


Performance enhanced

Figure 11-1: Routing: Shared Bus vs. HyperTransport Point-Point



HyperTransport™ Technology Basics



- HyperTransport™ Technology buses have two unidirectional point-to-point links
  - The links can be 2-, 4-, 8-, 16-, or 32-bits wide in each direction
  - HyperTransport links have a data rate up to 1600 Megabits/second per pin-pair (800 MHz clock)
    - E.g., 4 bits each way give up to 1.6 GB/sec total bandwidth
    - E.g., 8 bits each way give up to 3.2 GB/sec total bandwidth
    - E.g., 16 bits each way give up to 6.4 GB/sec total bandwidth
    - E.g., 32 bits each way give up to 12.8 GB/sec total bandwidth
- Packets are multiples of 4-bytes in length
- Serial link with commands, addresses and data use the same bits

**HyperTransport is a trademark of the HyperTransport Technology Consortium. AMD and AMD Athlon are trademarks of Advanced Micro Devices, Inc. All the other TM belong to their respective owners*

Figure 3-1: HyperTransport Signal Groups

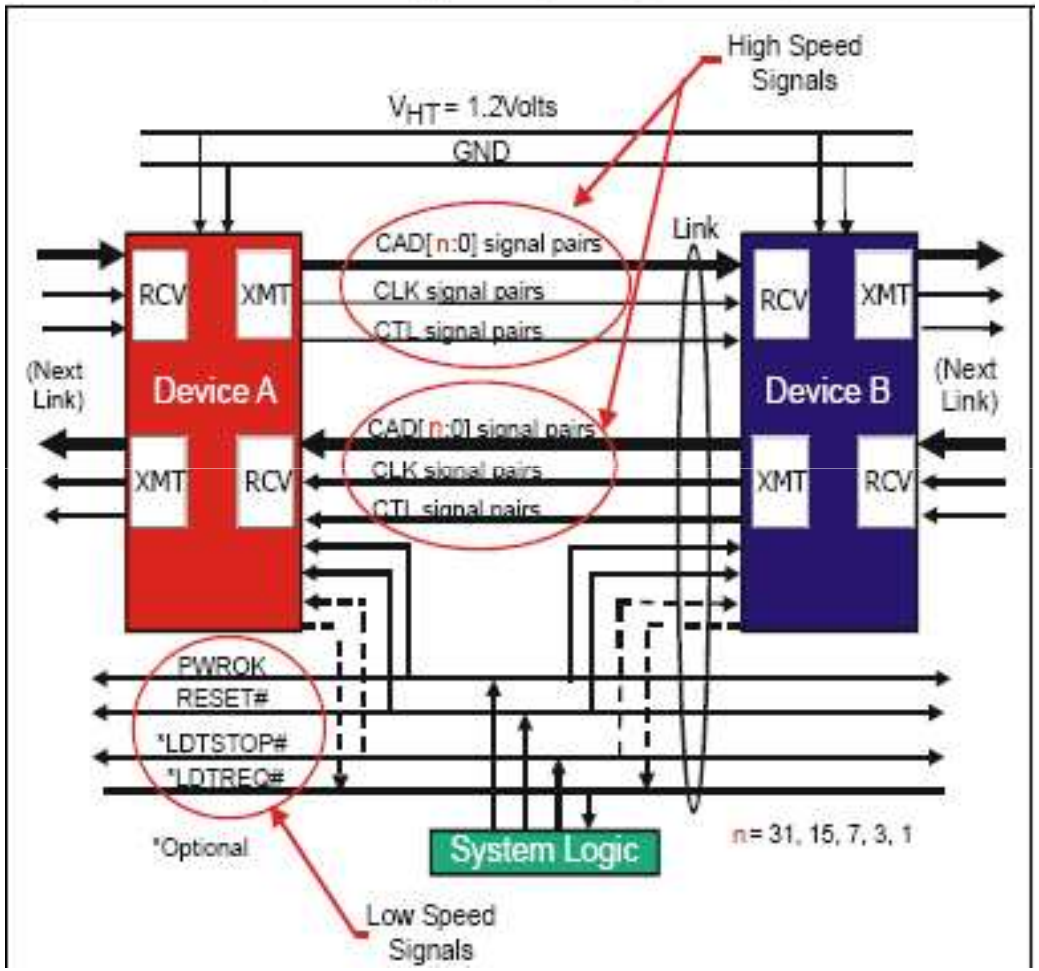


Figure 15-1: Simple Synchronous Clocking Interface

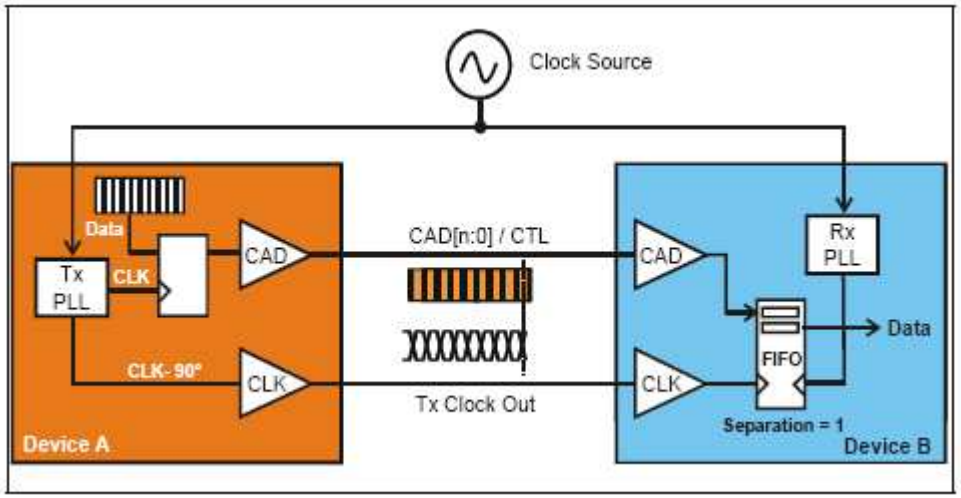
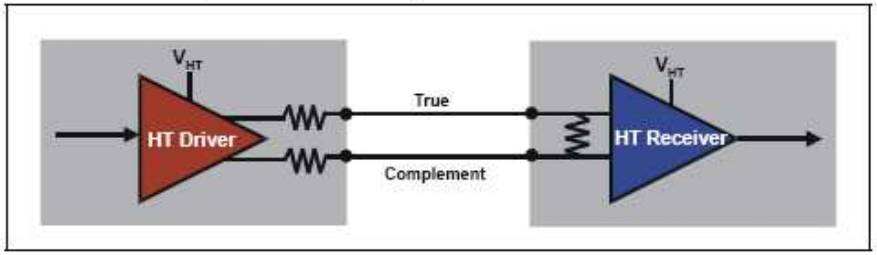
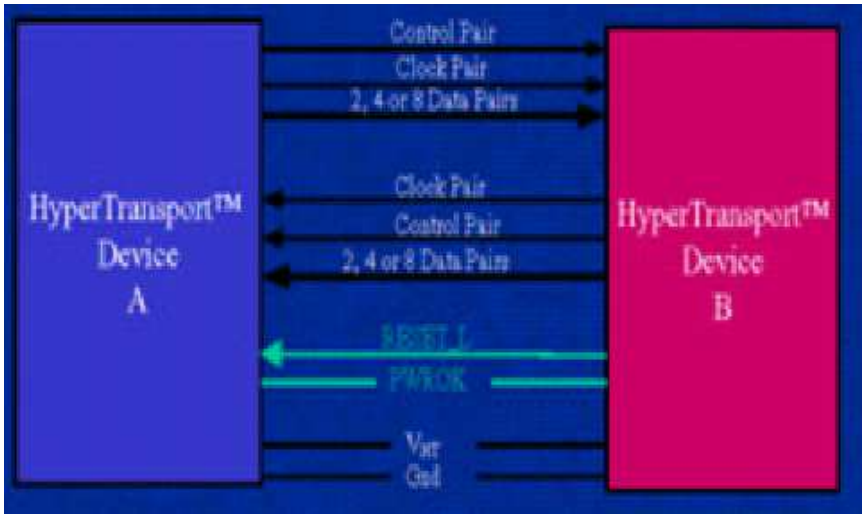
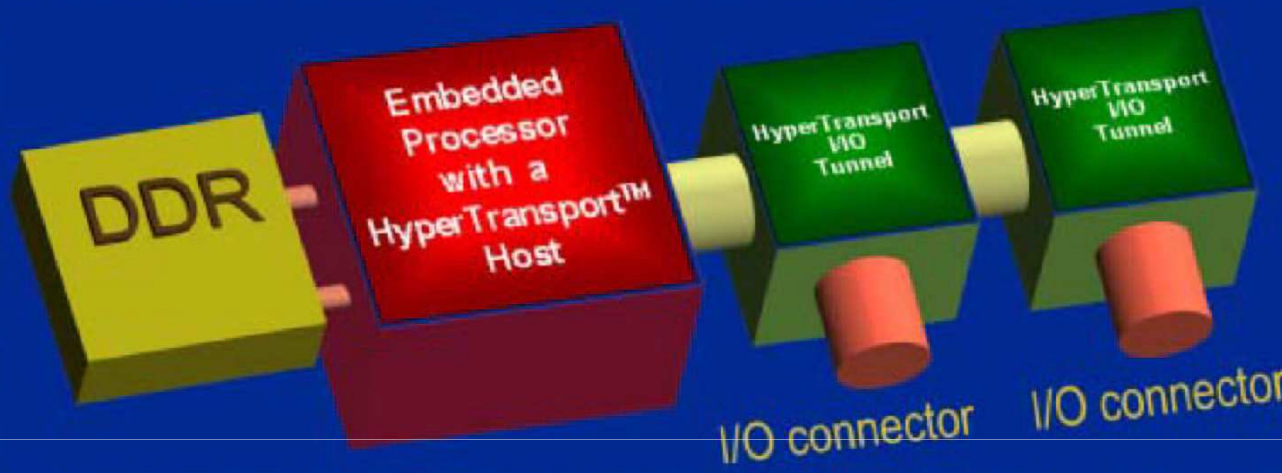


Figure 14-2: HT Link Differential Driver and Receiver





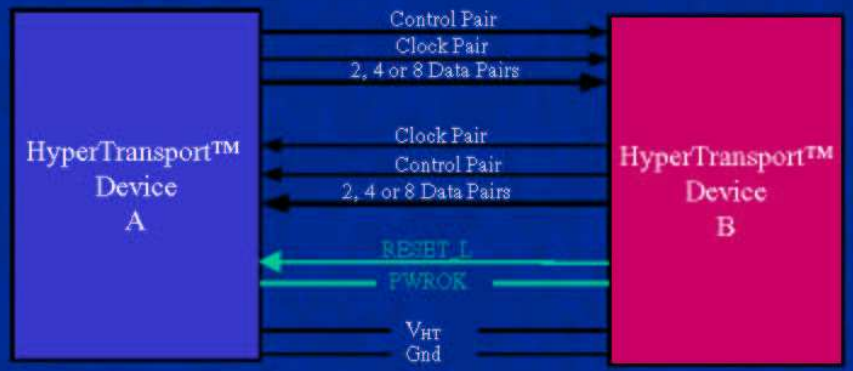
Embedded Applications and I/O Tunnels



- **For the 1st time in the industry:**
  - I/O devices shared among computation and communication industry
  - Unique "TUNNELING" capability gives almost unlimited I/O expandability
  - Fundamentally different microprocessor and memory controllers may be designed to use the very same I/O components
  - Pin count adjustable for the necessary Bandwidth
  - Cost reduced due the the cumulative volume
  - Extended component life

HyperTransport™ Device Pin Count

- Additional HyperTransport™ Device signals
  - Power OK (PWROK)
  - Reset HyperTransport Device (RESET_L)
- 55-pin HyperTransport device bus provides 12X the bandwidth of PCI-32/33 with fewer pins
- Signal to ground ratio is designed to be 4:1
- Optional link power down signals for mobile systems
  - HyperTransport Device Stop_L
  - DevReq_L
- Power per pin-pair is nil when in HyperTransport Device Stop mode



PWROK, RESET_L required for proper reset & init
$V_{HyperTransport}$ routed between devices is required for proper common mode range

Bus Width (Each Way)	2	4	8	16	32
Data Pins (total)	8	16	32	64	128
Clock Pins (total)	4	4	4	8	16
Control Pins (total)	4	4	4	4	4
Subtotal (high speed)	**16**	**24**	**40**	**76**	**148**
VLDT	2	2	3	6	10
GND	4	6	10	19	37
PWROK	1	1	1	1	1
RESET_L	1	1	1	1	1
Total Pins	**24**	**34**	**55**	**103**	**197**
Total Max BW GB/s	0.8	1.6	3.2	6.4	12.8

DC Power per Pin-Pair: 4 - 9 mW, 6 $mW_{Typical}$
Signal to V_{HT}/Gnd Ratio: 4:1

Figure 4-1: Four Byte Packet On An 8-Bit Interface

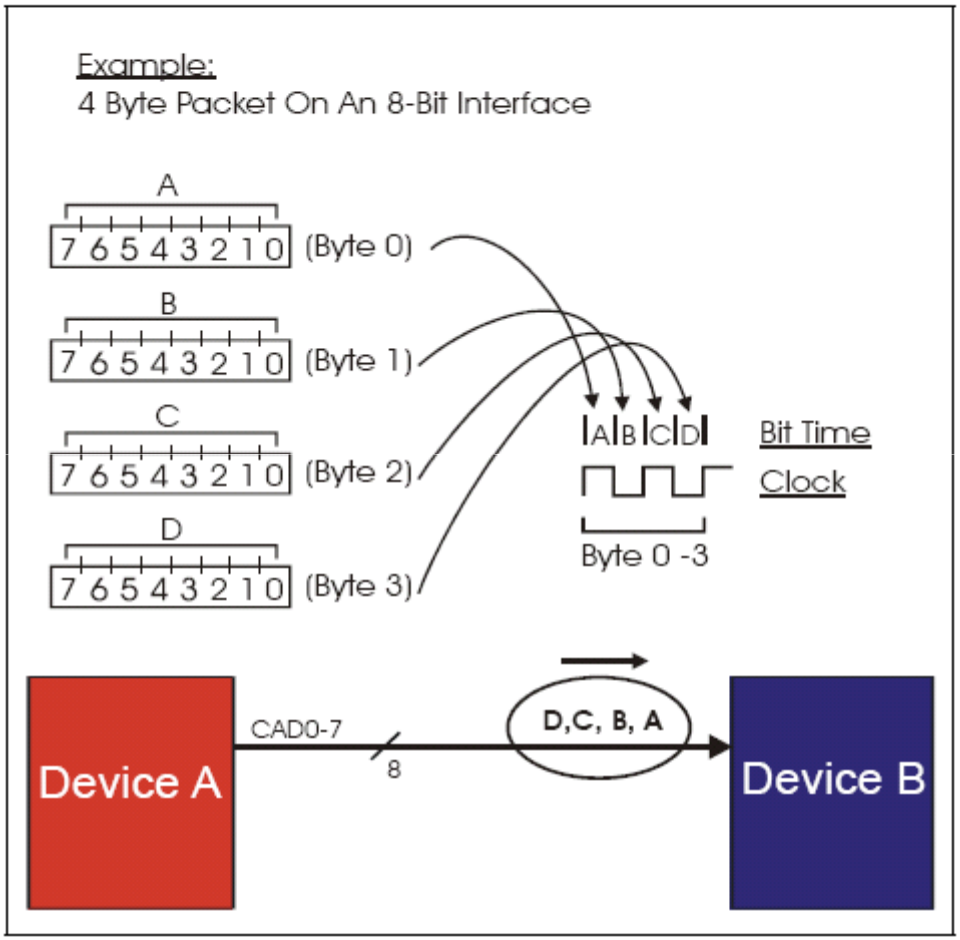


Figure 4-2: Four Byte Packet On A 2-Bit Interface

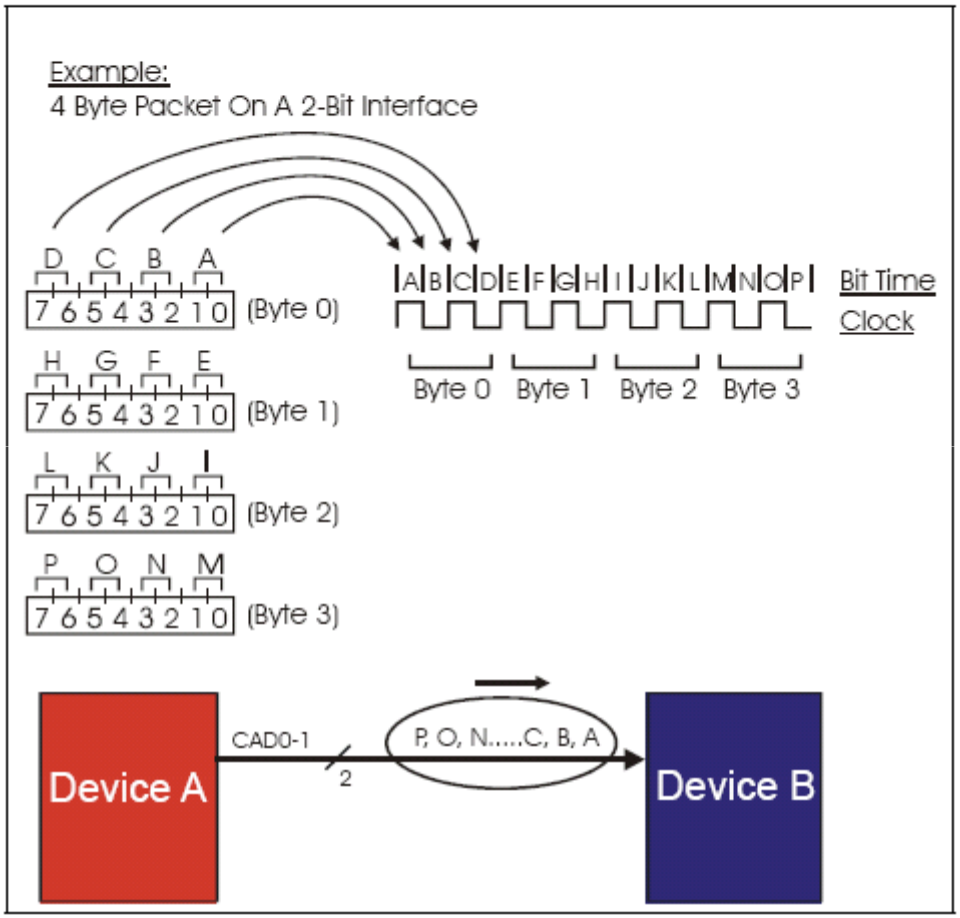
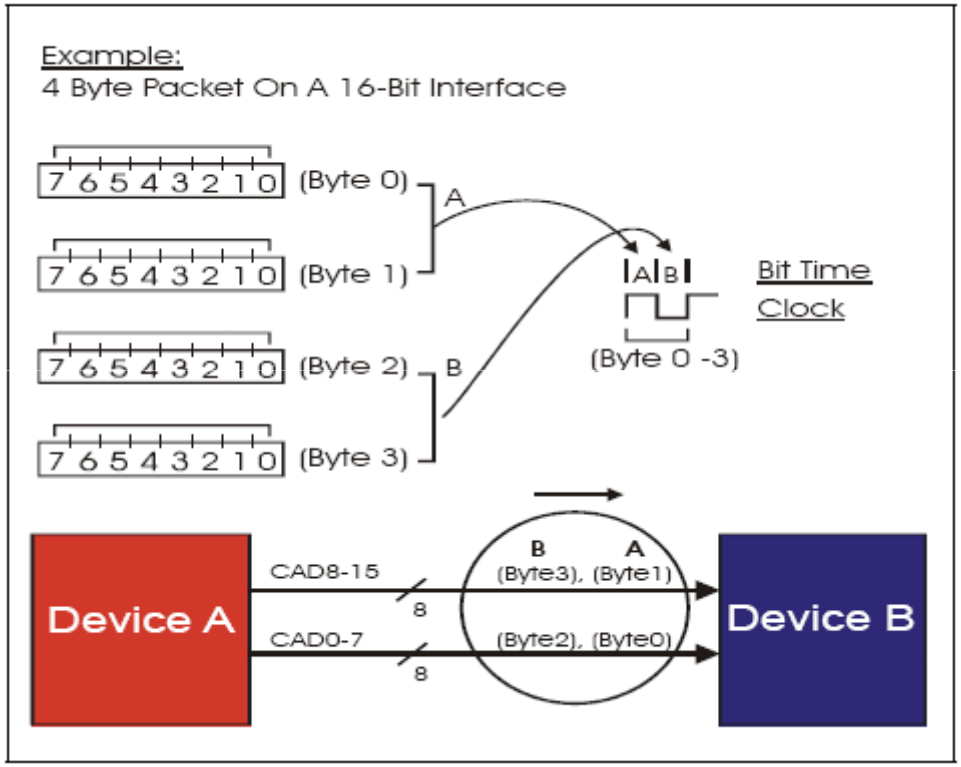
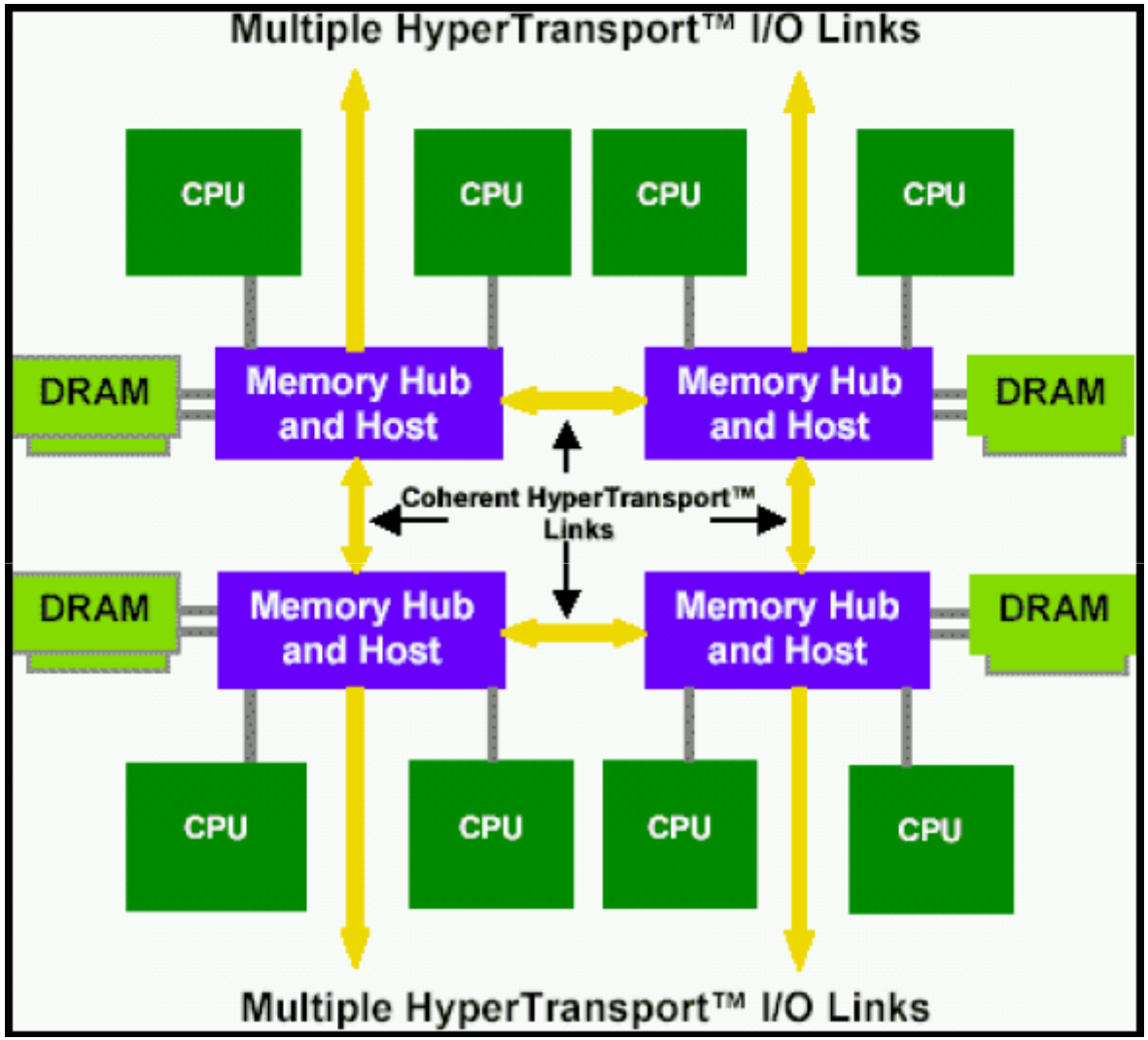
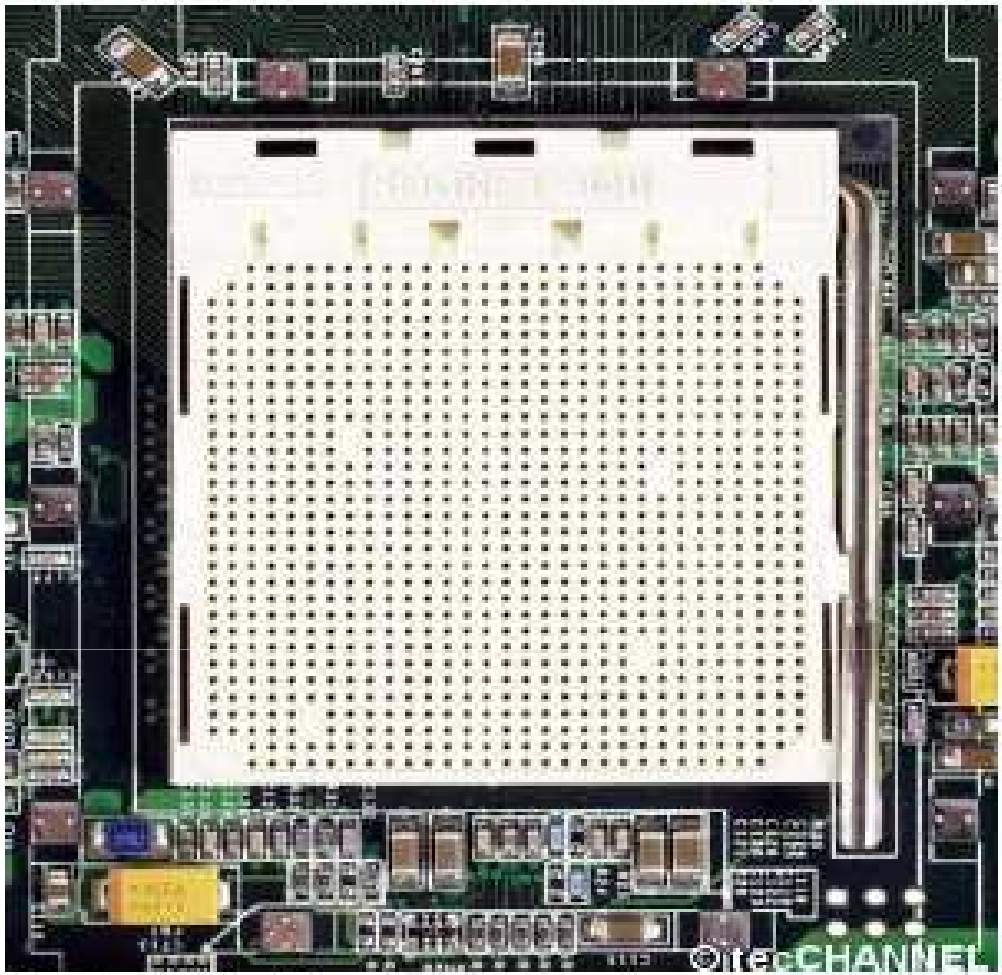


Figure 4-3: Four Byte Packet On A 16-Bit Interface

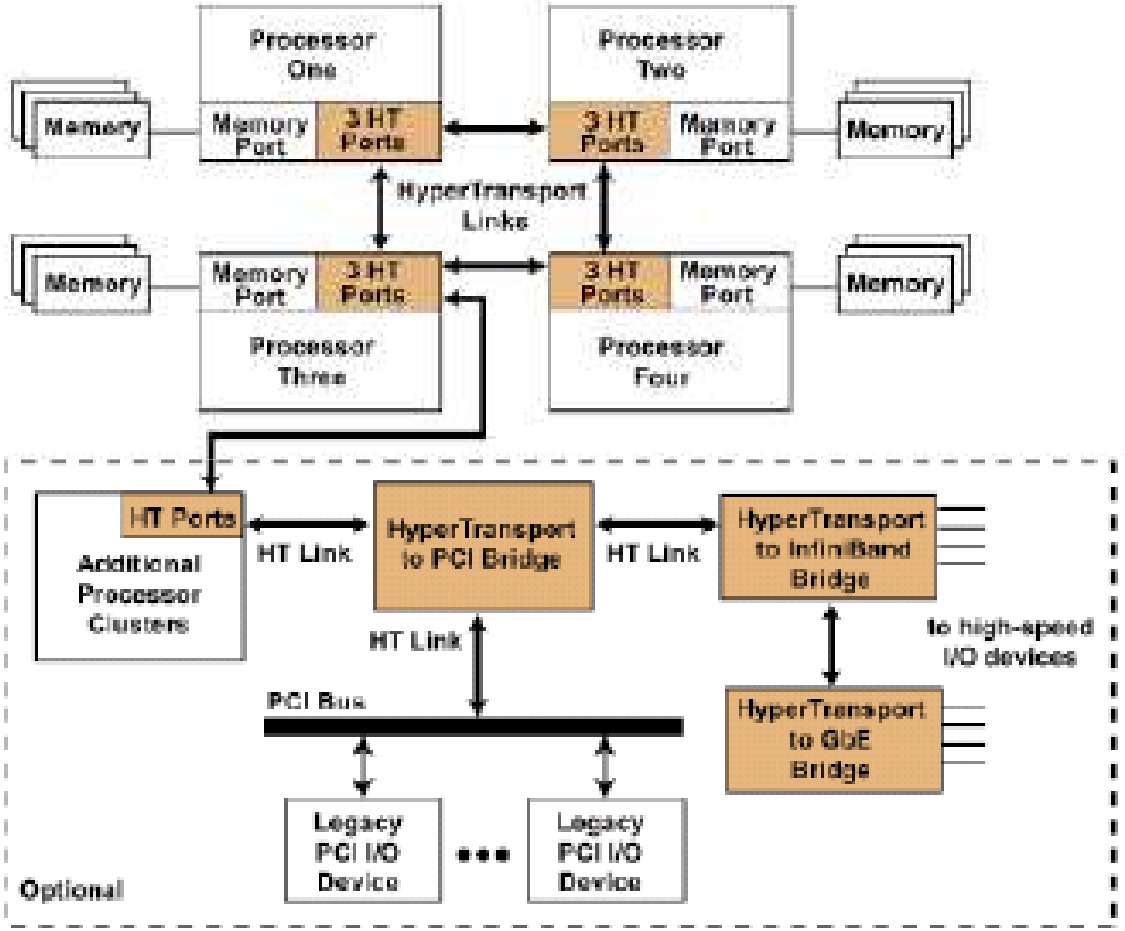



Multiple HyperTransport™ I/O Links
CPU
CPU
CPU
CPU
DRAM
Memory Hub and Host
Memory Hub and Host
DRAM
Coherent HyperTransport™ Links
DRAM
Memory Hub and Host
Memory Hub and Host
DRAM
CPU
CPU
CPU
CPU
Multiple HyperTransport™ I/O Links

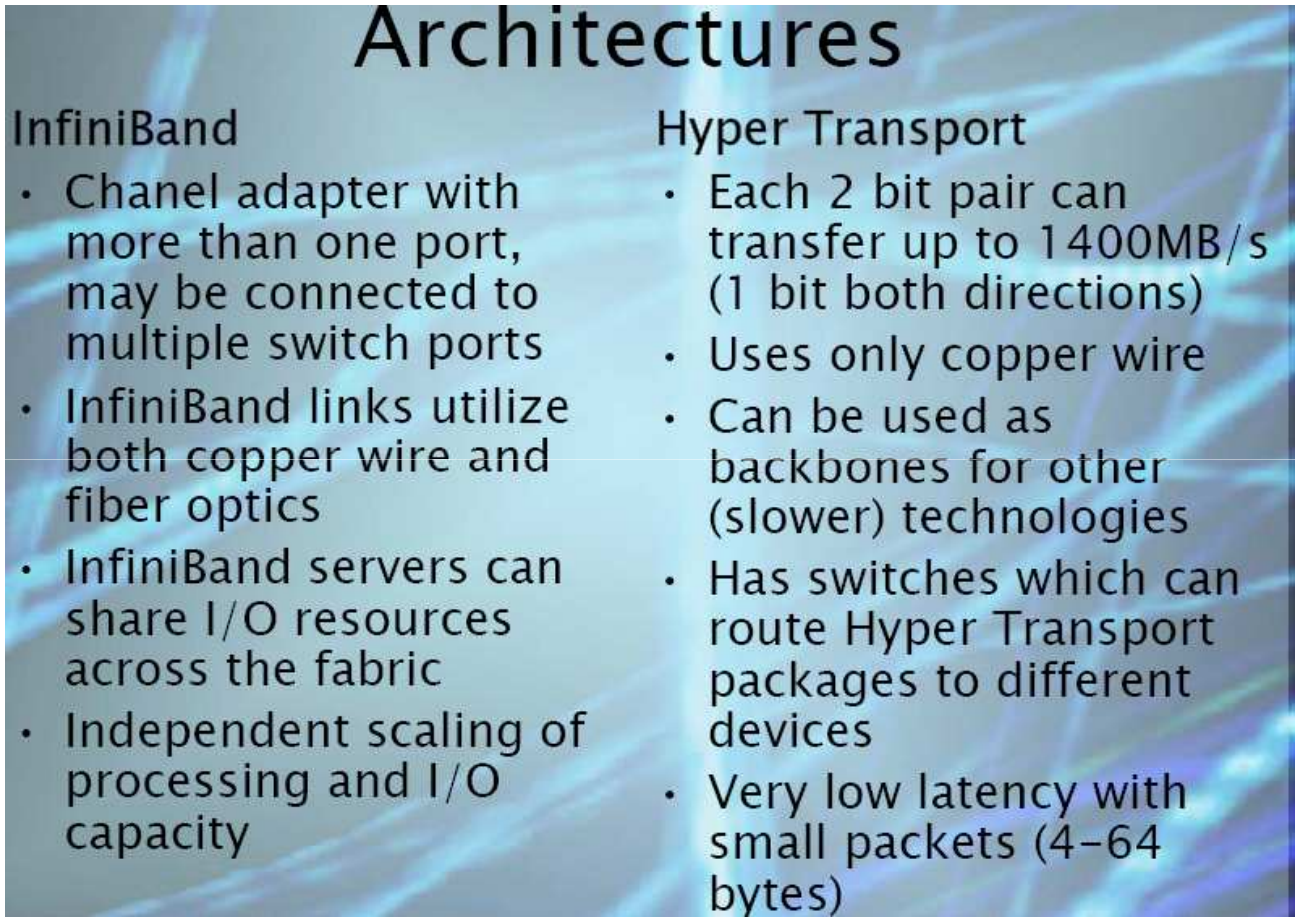


Socket 940: Der Opteron verlangt nach einem neuen Steckplatz und wird mit einer Core-Spannung von 1,55 V versorgt.

System Architecture



HT vs. IB - srovnání



Srovnání technických parametrů

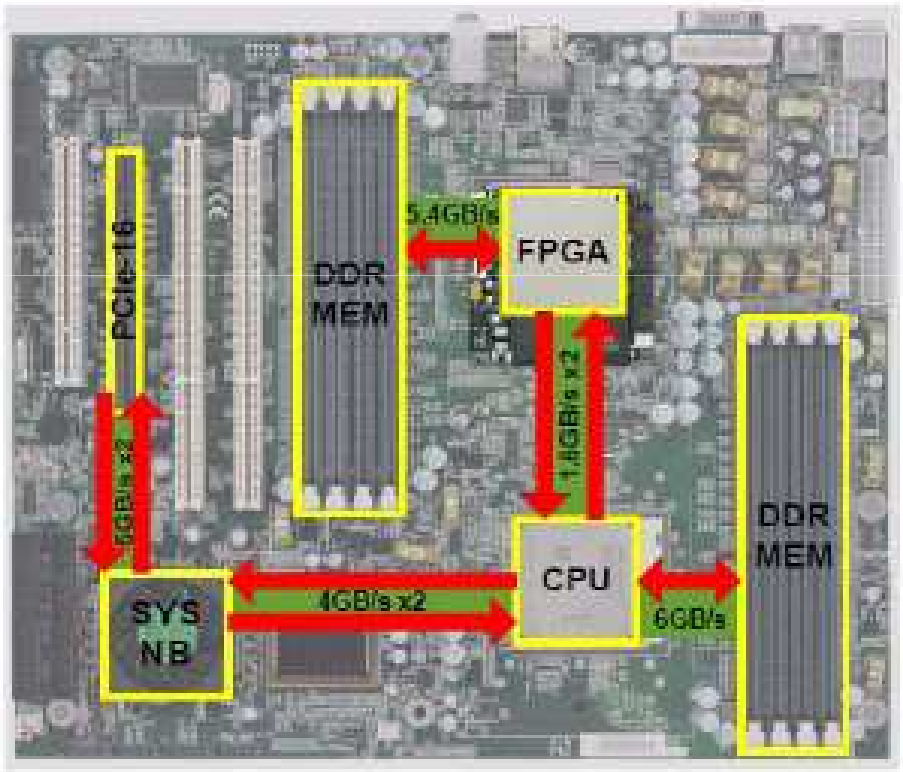
Technical differences

	InfiniBand	Hyper Transport
data signaling speed	1,25 GHz	400Mhz to 2.8Ghz
bandwidth	500MB/s – 6 GB/s	100MB–22.4 GB/s
link width	2, 8, or 24	2,4,8,16 or 32 bits
Protocol used	IPv6	Custom
Package size	68–4096 bytes	4–64 bytes
Memory model	non–coherent	Coherent and non–coherent
Environment	network	Inside the box
Plug & Play	Yes	No (devices are)

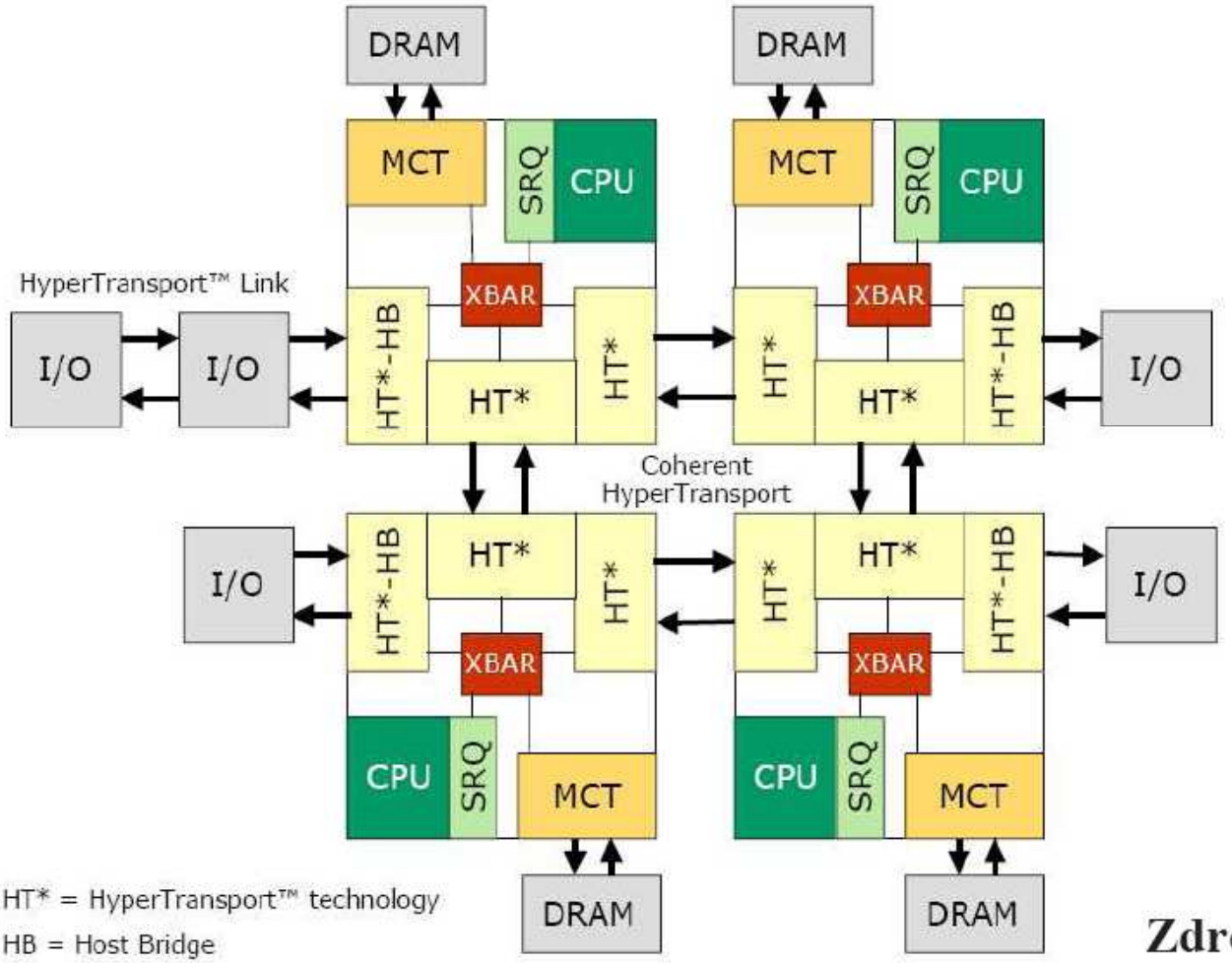
Suma sumarum

- Both techniques are already in use
- Both techniques provide much better bandwidth than the previous ones
- Allows faster connectivity for parallel computers, network connections, inside clusters and common computers
- Hyper Transport allows fast internal connections, InfiniBand fast internal and external connections

Hypertransport-based Co-processors (cont.)



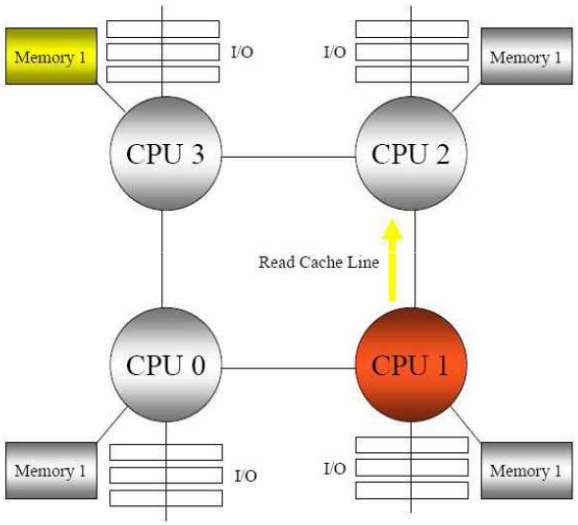
Takhle vypadá AMD Quad – Coherent HyperTransport



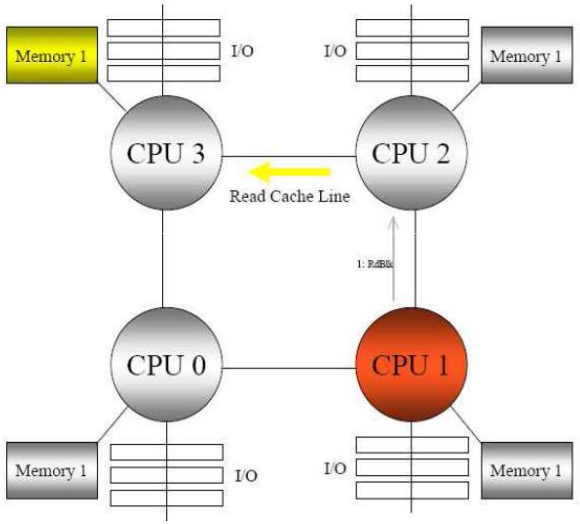
Zdroj: AMD

Příklad: CPU1 čtení položku, která je domovská v CPU3

Step 1

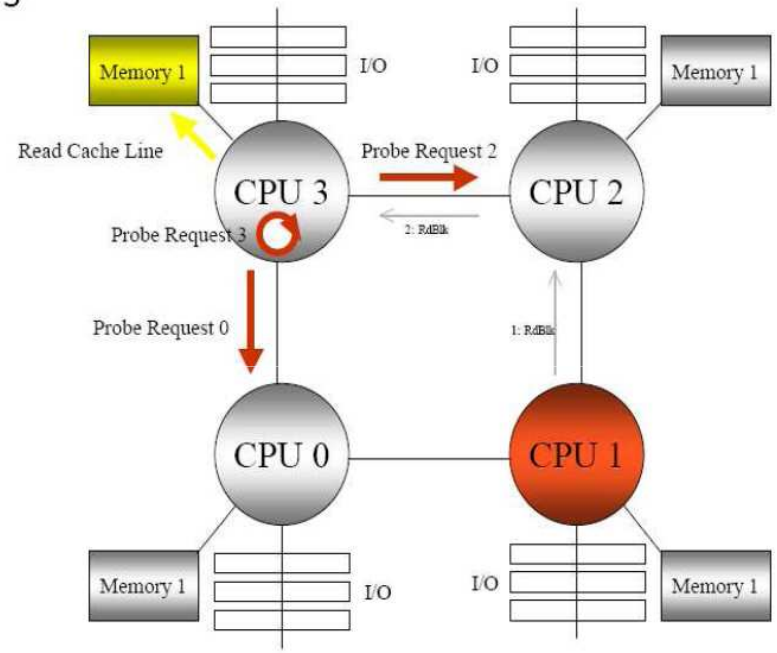


Step 2

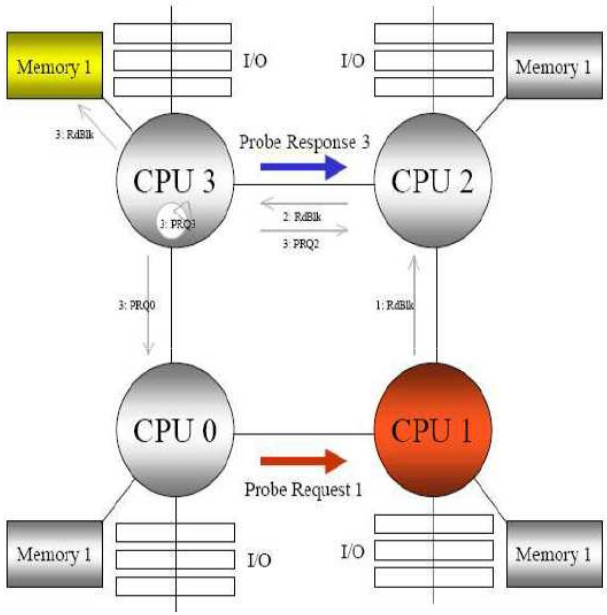


CPU1 čte položku, která je domovská v CPU3 - pokračování

Step 3

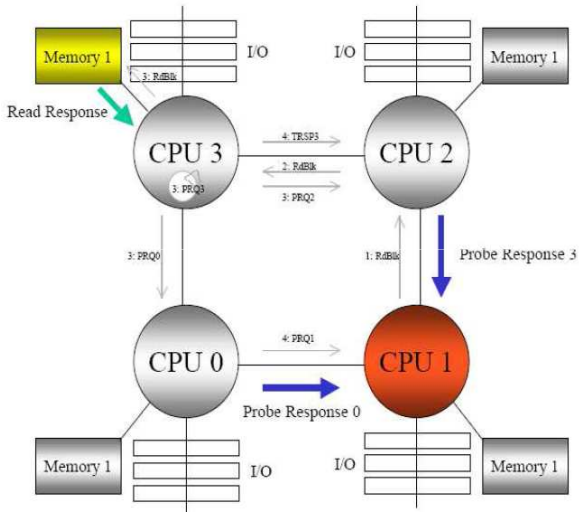


Step 4

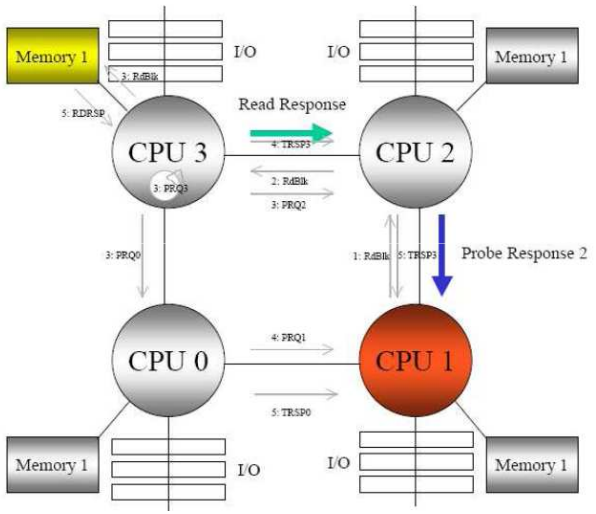


CPU1 čte položku, která je domovská v CPU3 - pokračování

Step 5

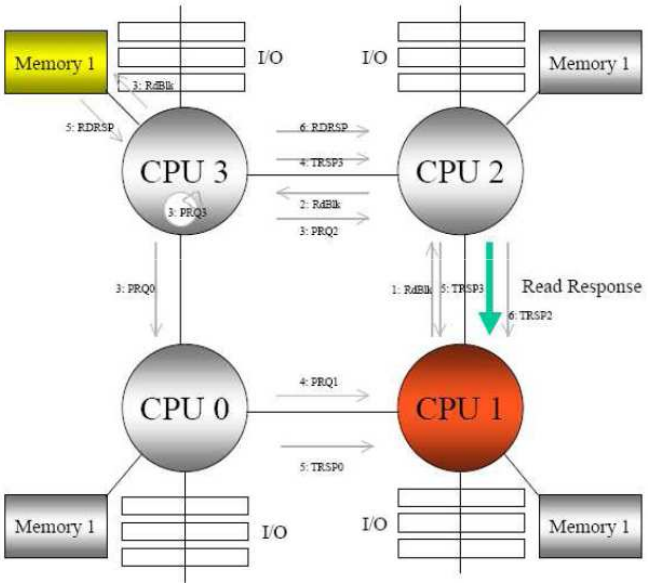


Step 6

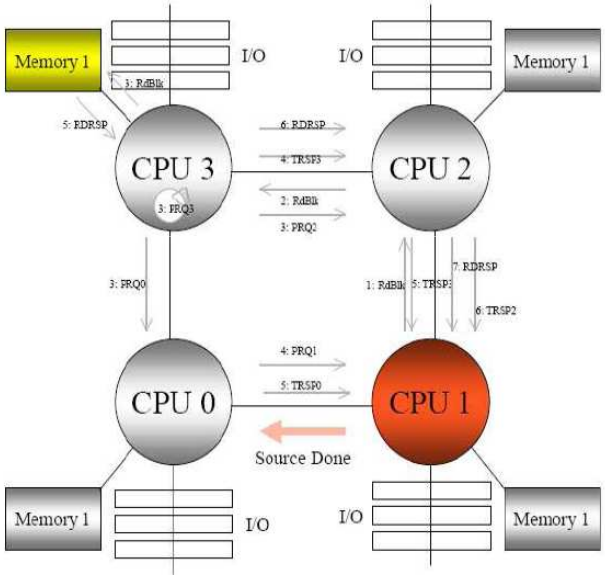


CPU1 čte položku, která je domovská v CPU3 - pokračování

Step 7

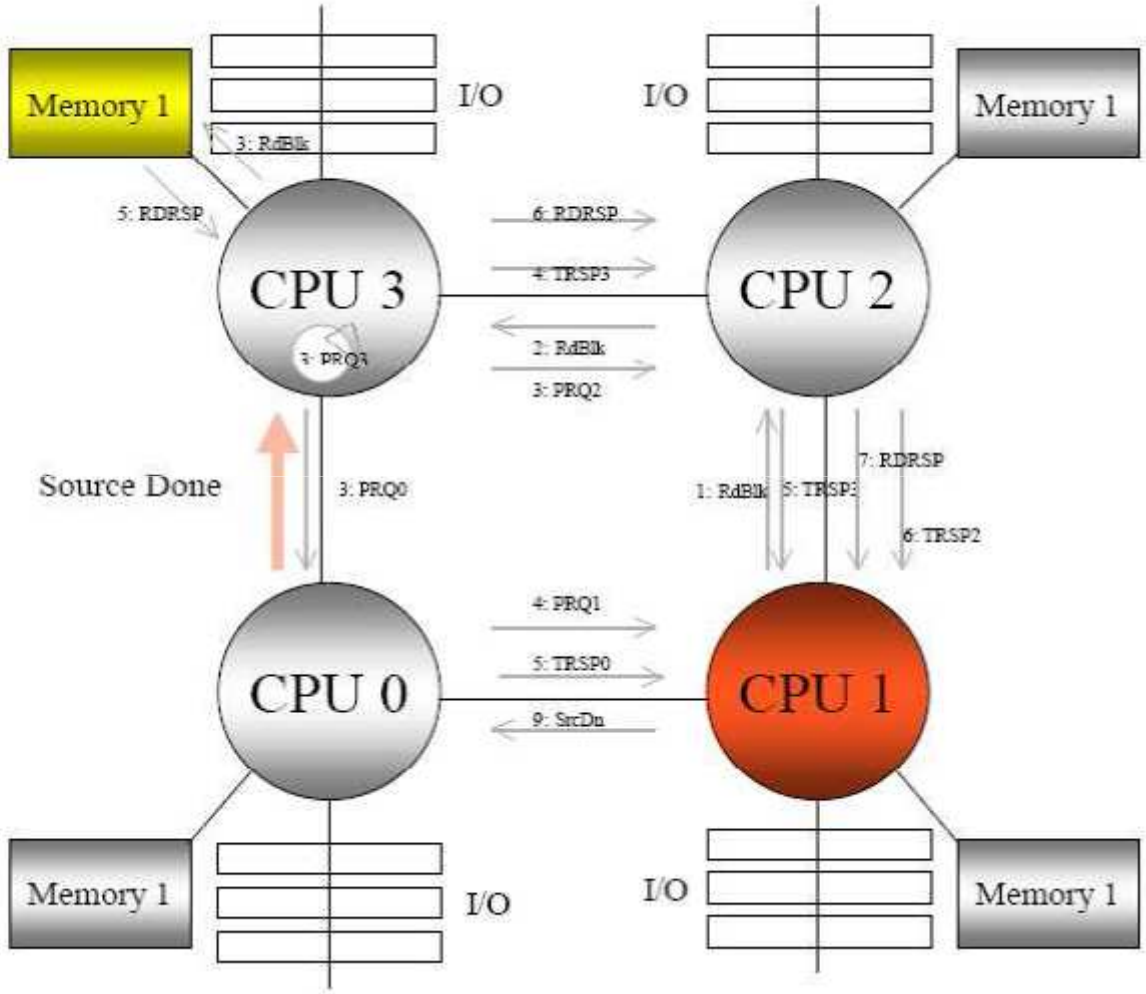


Step 8



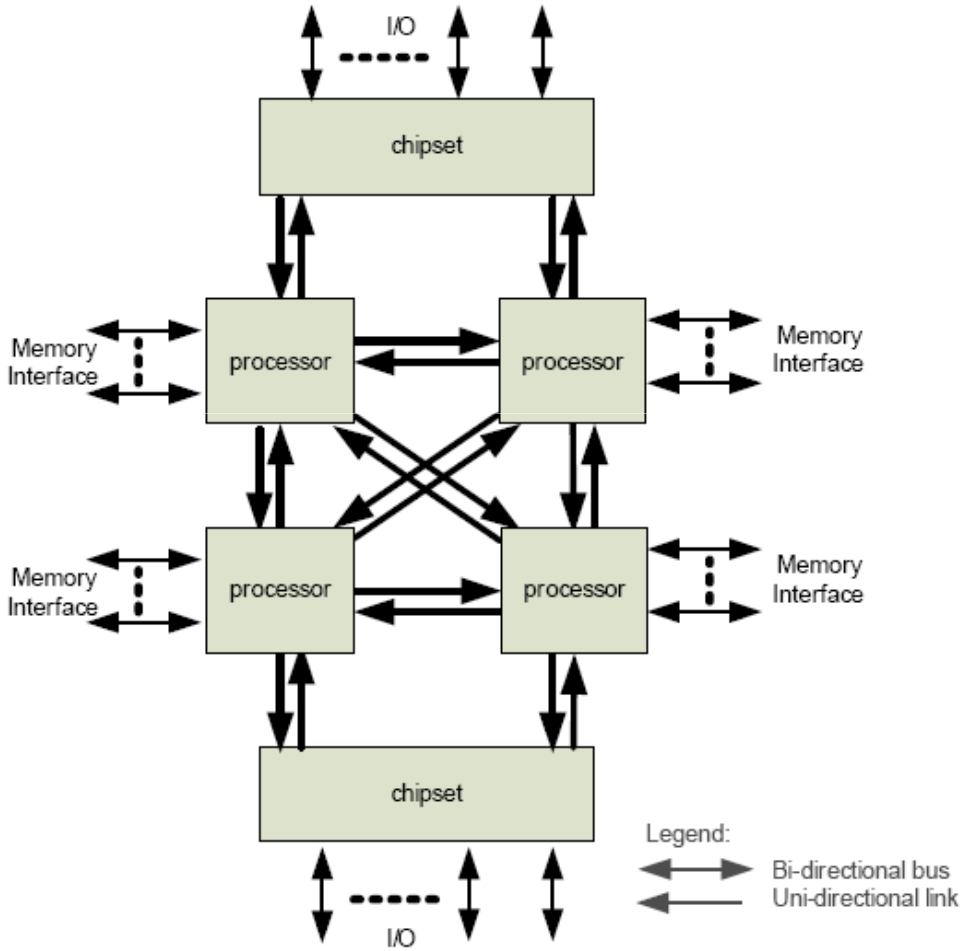
CPU1 čte položku, která je domovská v CPU3 - dokončení

Step 9



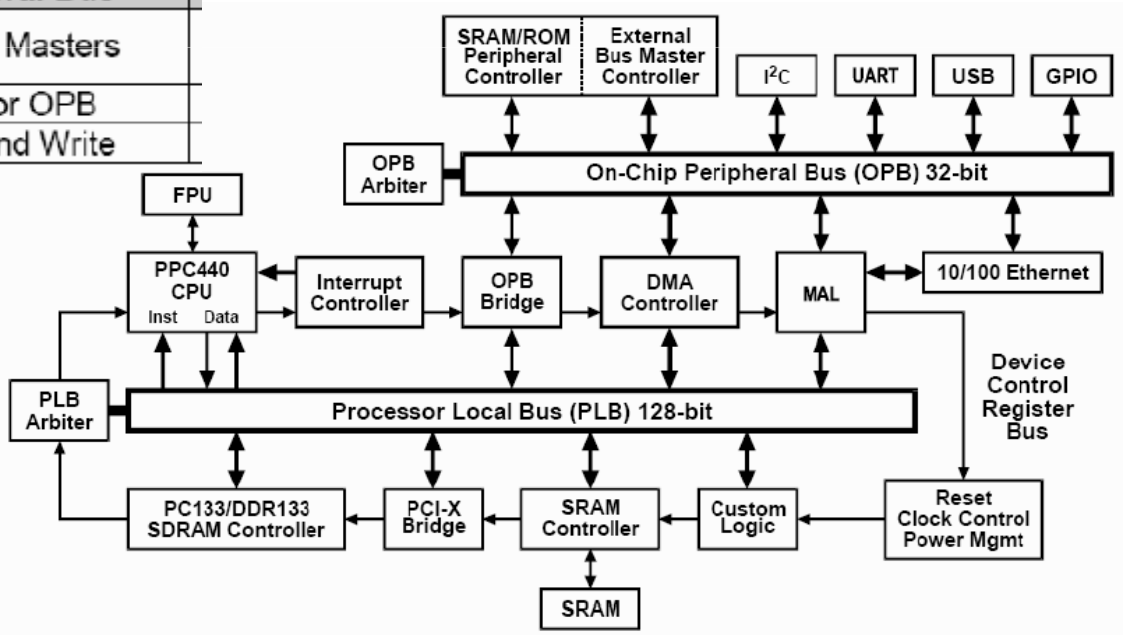
Jsou i jiná možná řešení SoC sběrnic?

Intelovské QPI?

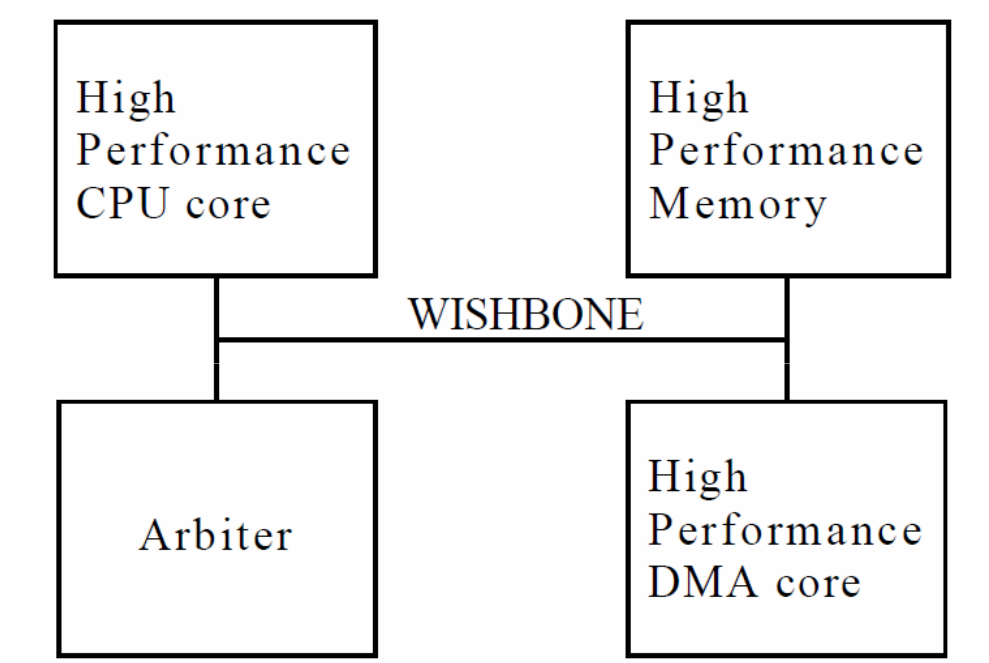


IBM CoreConnect Bus

	IBM CoreConnect Processor Local Bus
Bus Architecture	32-, 64-, and 128-bits Extendable to 256-bits
Data Buses	Separate Read and Write
Key Capabilities	Multiple Bus Masters 4 Deep Read Pipelining 2 Deep Write Pipelining Split Transactions Burst Transfers Line Transfers
	On-Chip Peripheral Bus
Masters Supported	Supports Multiple Masters
Bridge Function	Master on PLB or OPB
Data Buses	Separate Read and Write



Wishbone



http://cdn.opencores.org/downloads/soc_bus_comparison.pdf