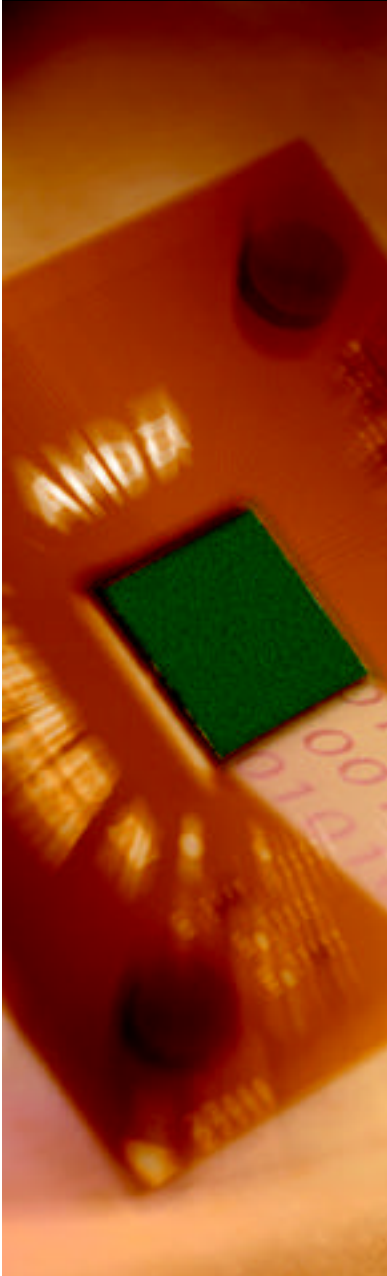


AMD's Next Generation Microprocessor Architecture

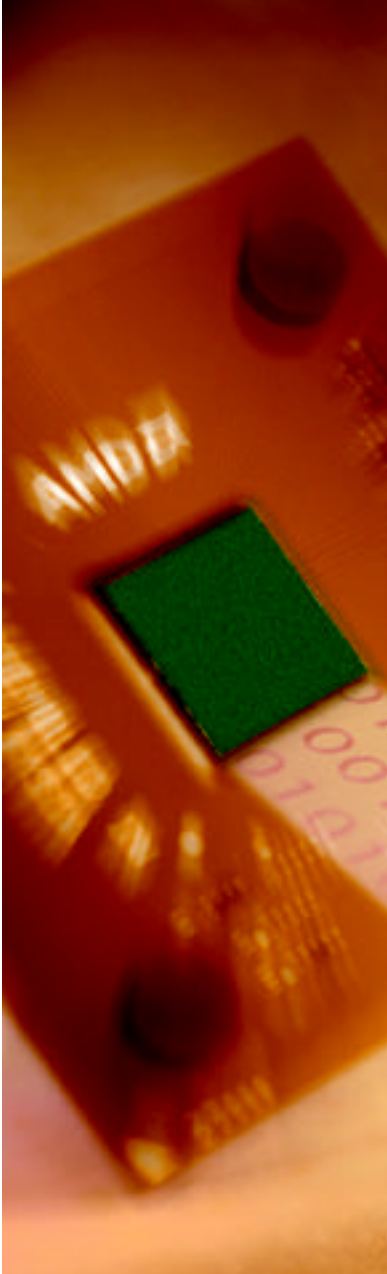
Fred Weber

October 2001



- **Build a next-generation system architecture which serves as the foundation for future processor platforms**
- **Enable a full line of server and workstation products**
 - **Leading edge x86 (32-bit) performance and compatibility**
 - **Native 64-bit support**
 - **Establish x86-64 Instruction Set Architecture**
 - **Extensive Multiprocessor support**
 - **RAS features**
- **Provide top-to-bottom desktop and mobile processors**

- **x86-64™ Technology**
- **"Hammer" Architecture**
- **"Hammer" System Architecture**



x86-64™ Technology

Why 64-Bit Computing?



- **Required for large memory programs**
 - Large databases
 - Scientific and Engineering Problems
 - Designing CPUs ☺
- **But,**
 - **Limited Demand for Applications which require 64 bits**
 - Most applications can remain 32-bit x86 instructions, if the processor continues to deliver leading edge x86 performance
- **And,**
 - **Software is a huge investment (tool chains, applications, certifications)**
 - **Instruction set is first and foremost a vehicle for compatibility**
 - Binary compatibility
 - Interpreter/JIT support is increasingly important

- **x86-64 mode built on x86**
 - Similar to the previous extension from 16-bit to 32-bit
 - Vast majority of opcodes and features unchanged
 - Integer/Address register files and datapaths are native 64-bit
 - 48-Bit Virtual Address Space, 40-Bit Physical Address Space
- **Enhancements**
 - Add 8 new integer registers
 - Add PC relative addressing
 - Add full support for SSE/SSEII based Floating Point Application Binary Interface (ABI)
 - including 16 registers
 - Additional Registers and Data Size added through reclaim of one byte increment/decrement opcodes (0x40-0x4F) for use as a single optional prefix
- **Public specification**
 - www.x86-64.org

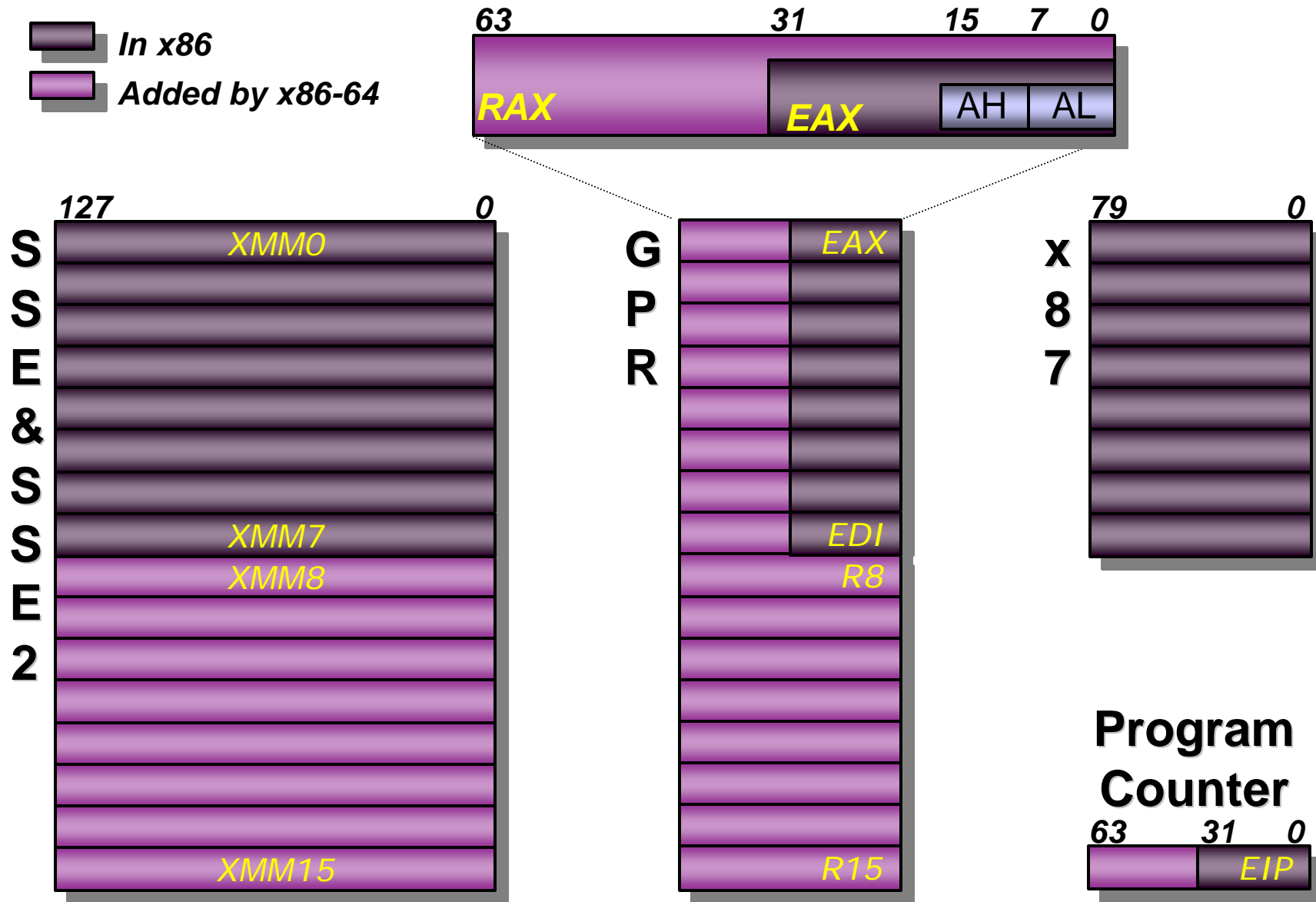
x86-64 Programmer's Model



In x86



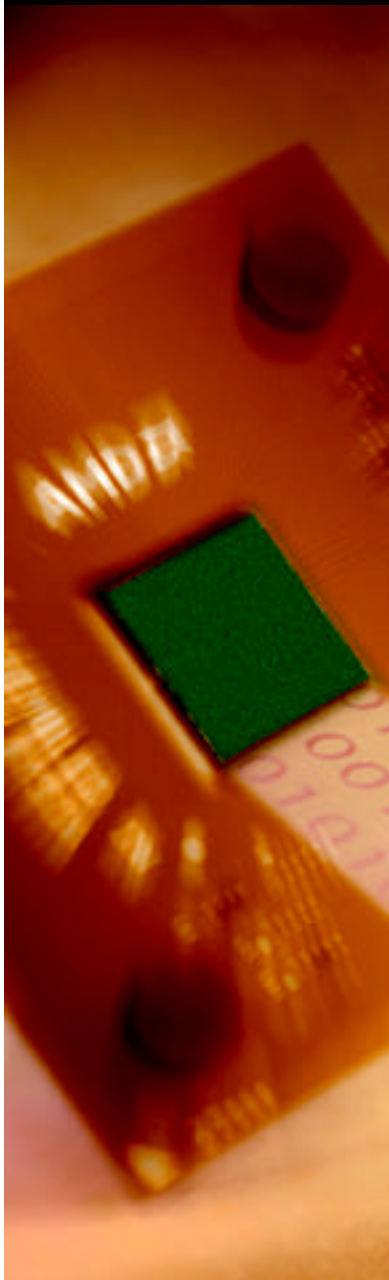
Added by x86-64



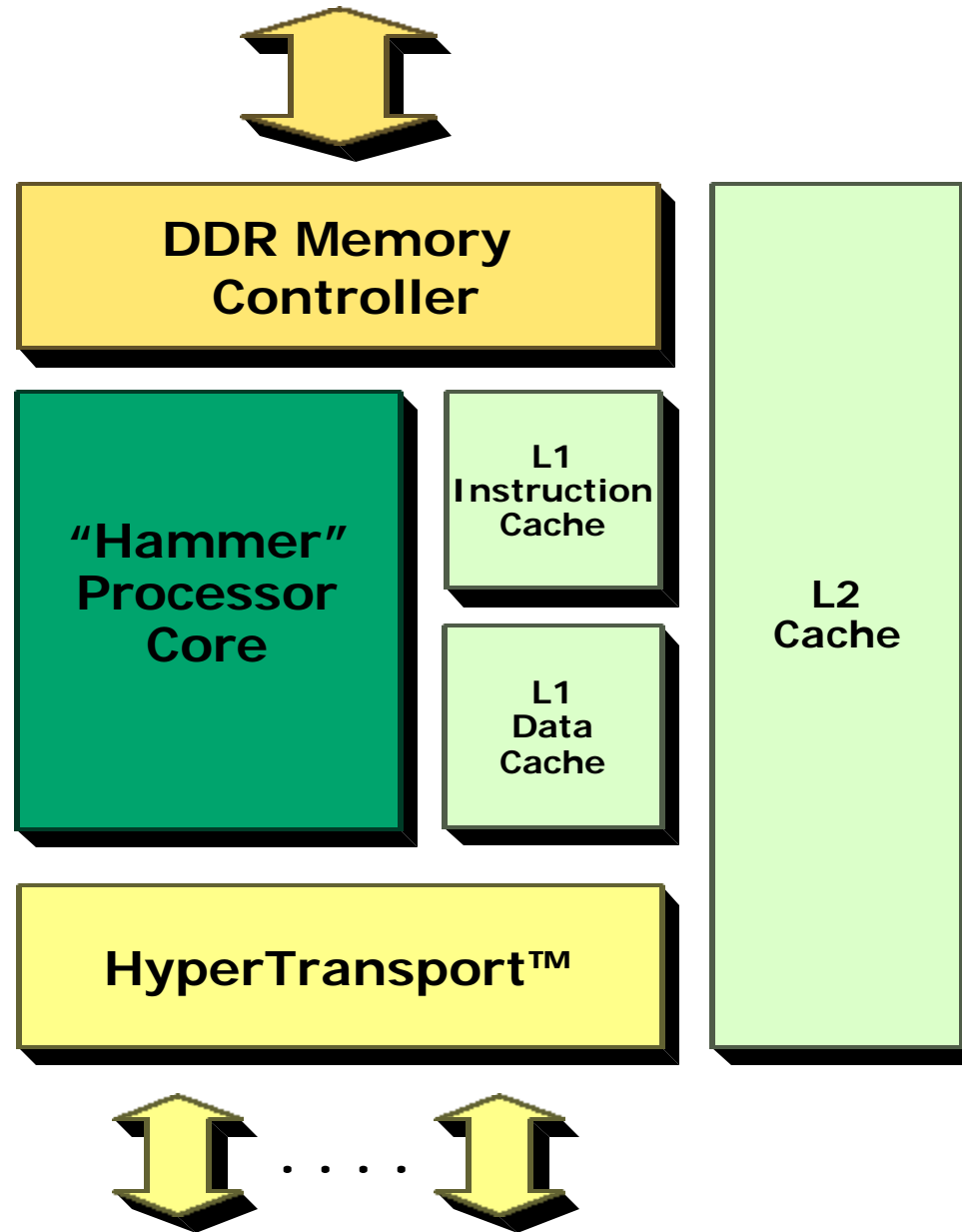
- **Compiler and Tool Chain is a straight forward port**
- **Instruction set is designed to offer all the advantages of CISC and RISC**
 - Code density of CISC
 - Register usage and ABI models of RISC
 - Enables easy application of standard compiler optimizations
- **SpecInt2000 Code Generation (compared to 32 bit x86)**
 - Code size grows <10%
 - Due mostly to instruction prefixes
 - Static Instruction Count **SHRINKS** by 10%
 - Dynamic Instruction Count **SHRINKS** by at least 5%
 - Dynamic Load/Store Count **SHRINKS** by 20%
 - All without any specific code optimizations

- **Processor is fully x86 capable**
 - Full native performance with 32-bit applications and OS
 - Full compatibility (BIOS, OS, Drivers)
- **Flexible deployment**
 - Best-in-class 32-bit, x86 performance
 - Excellent 64-bit, x86-64 instruction execution when needed
- **Server, Workstation, Desktop, and Mobile share same architecture**
 - OS, Drivers and Applications can be the same
 - CPU vendors focus not split, ISV focus not split
 - Support, optimization, etc. all designed to be the same

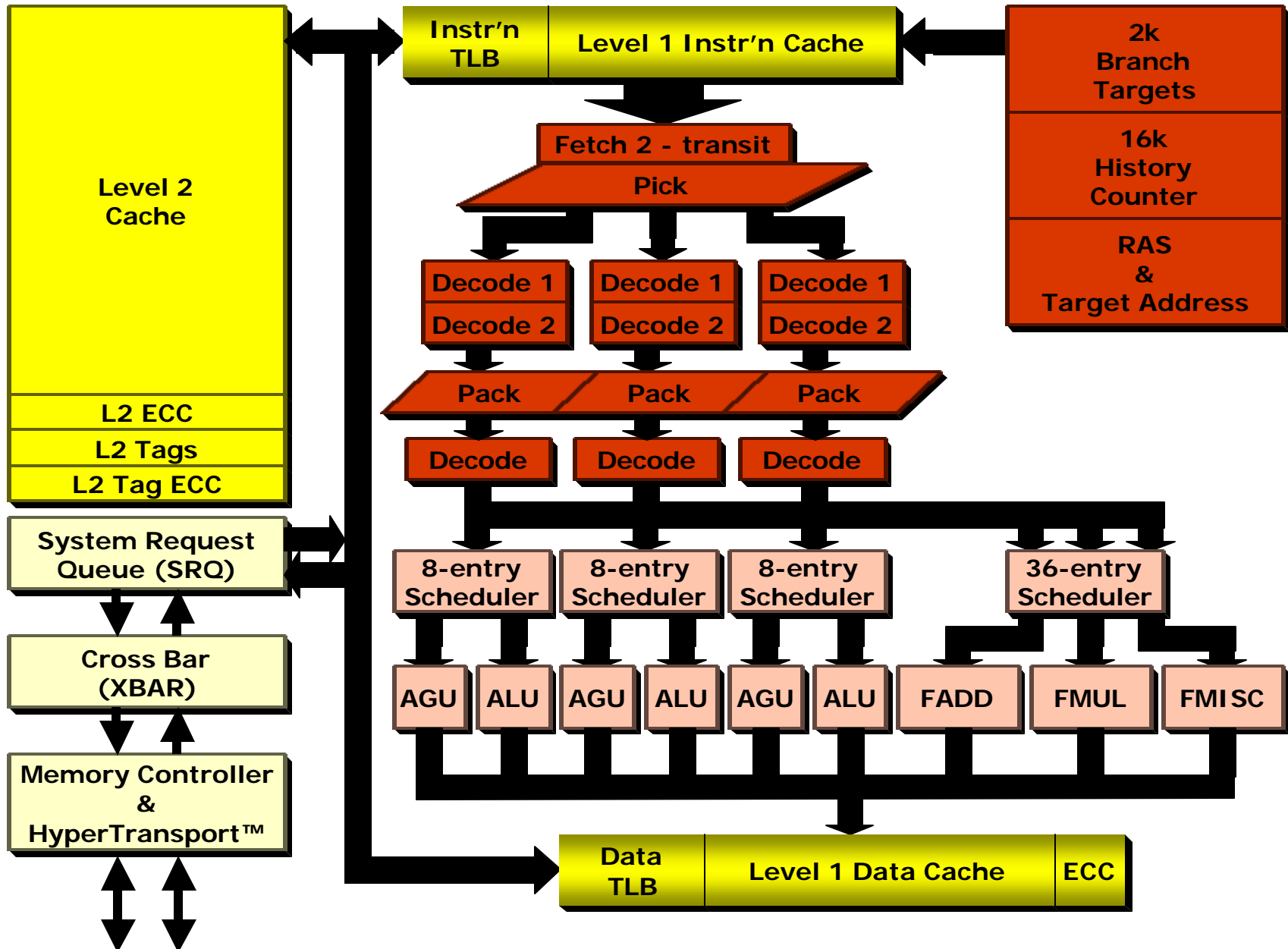
The "Hammer" Architecture



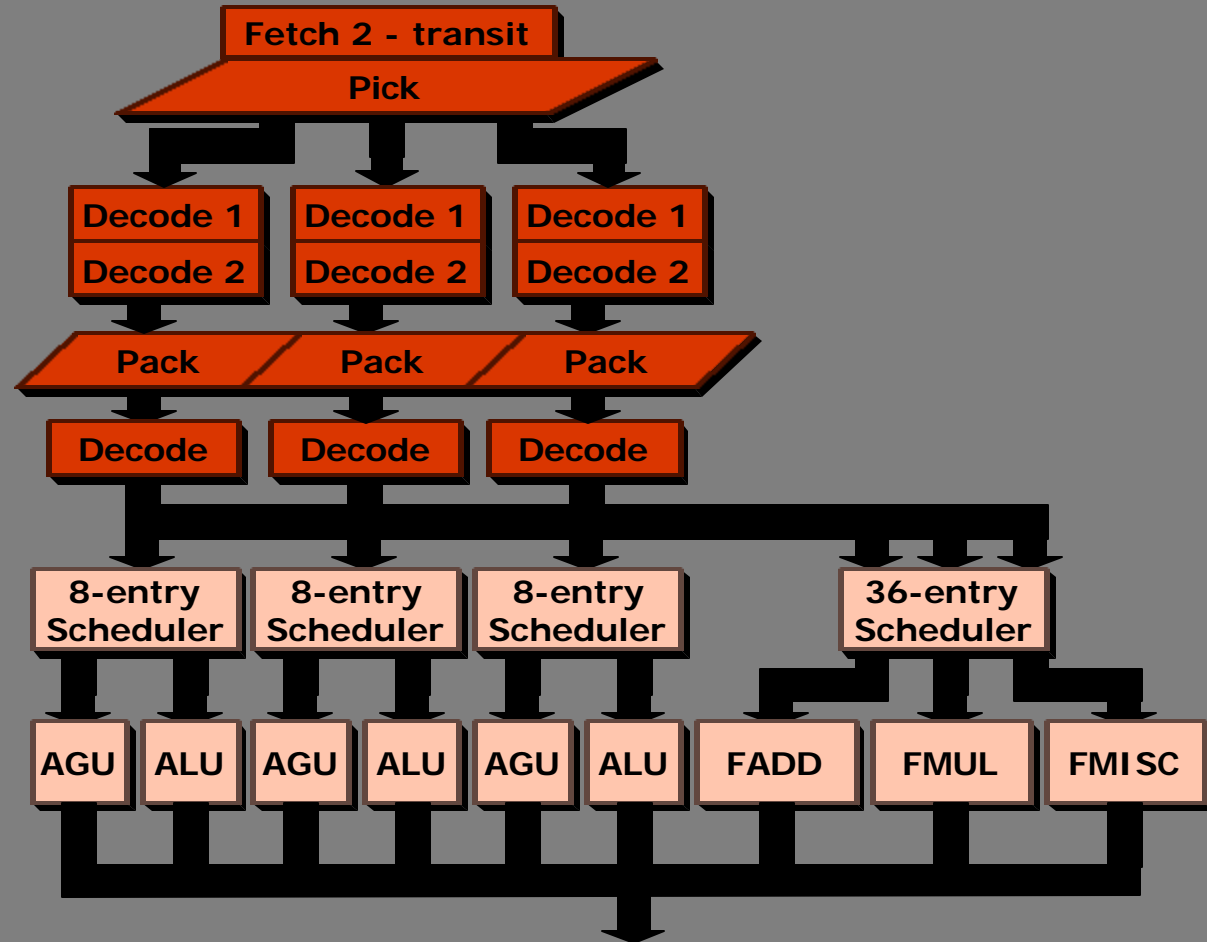
The "Hammer" Architecture



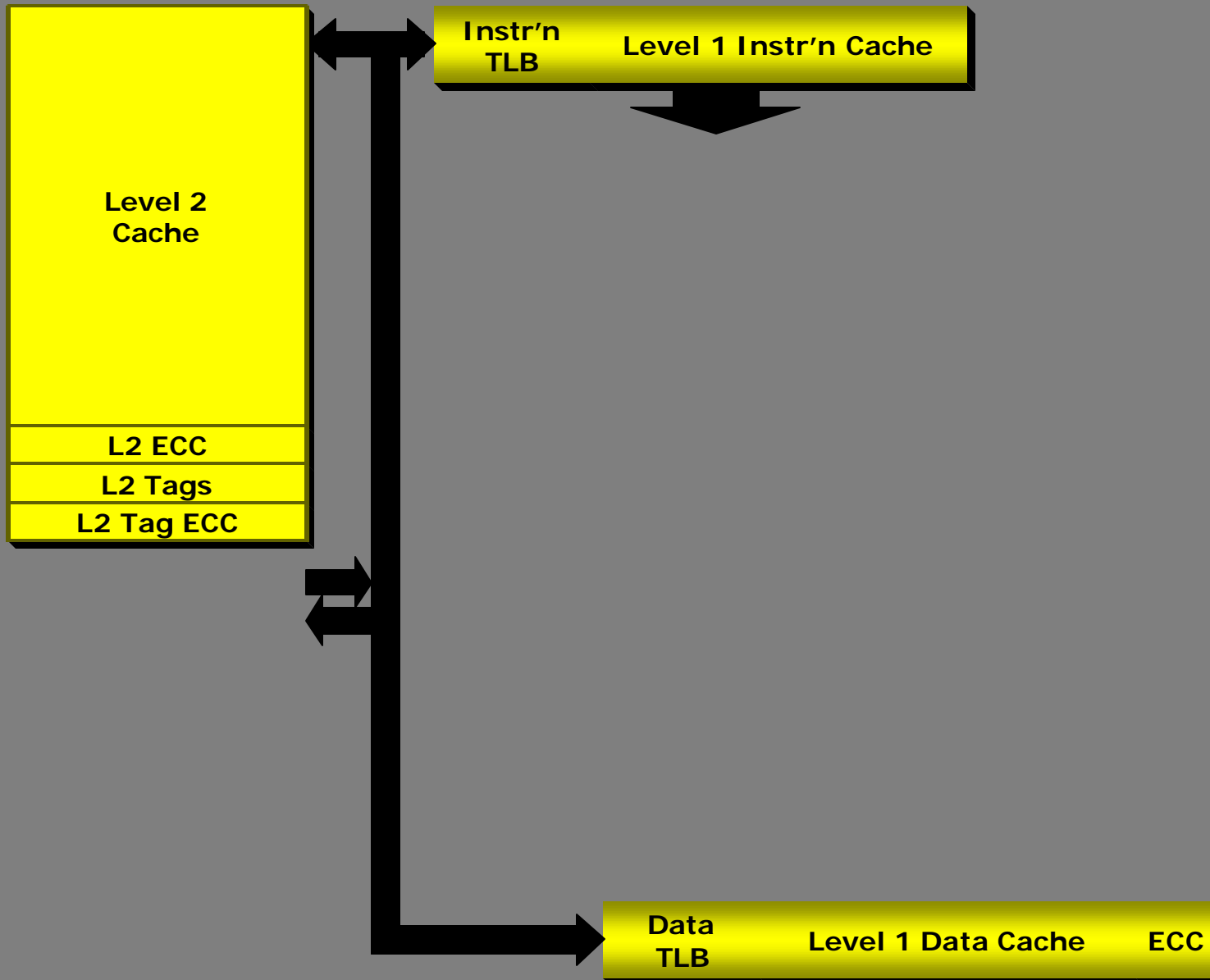
Processor Core Overview



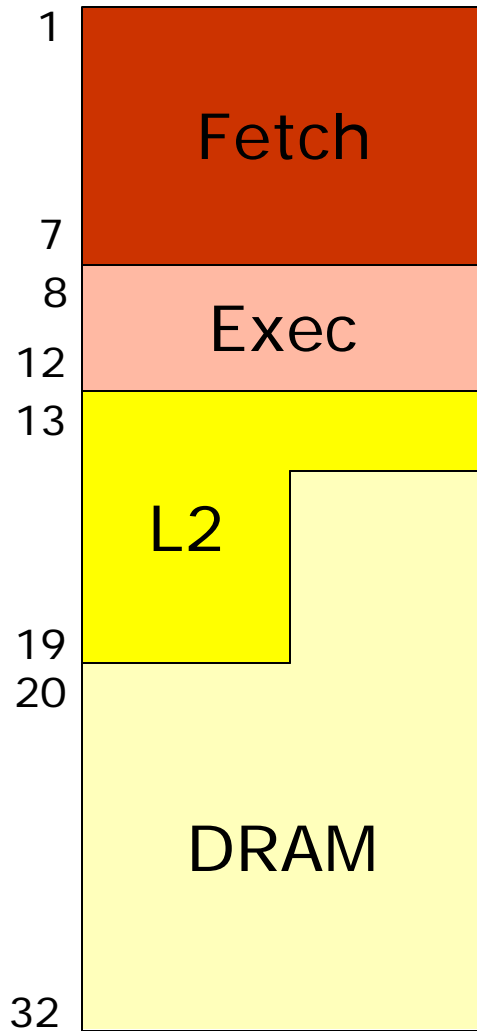
Processor Core Overview



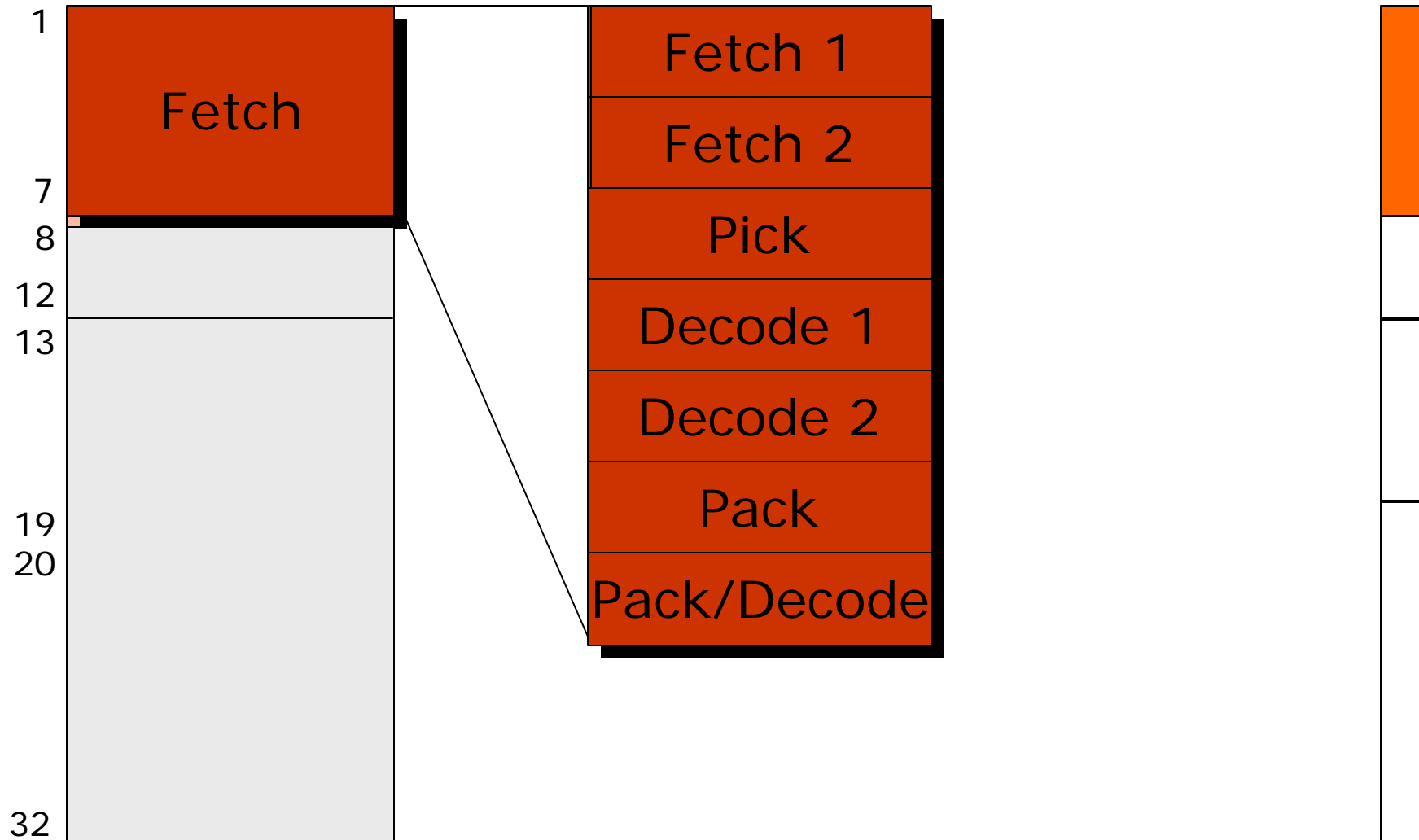
Processor Core Overview



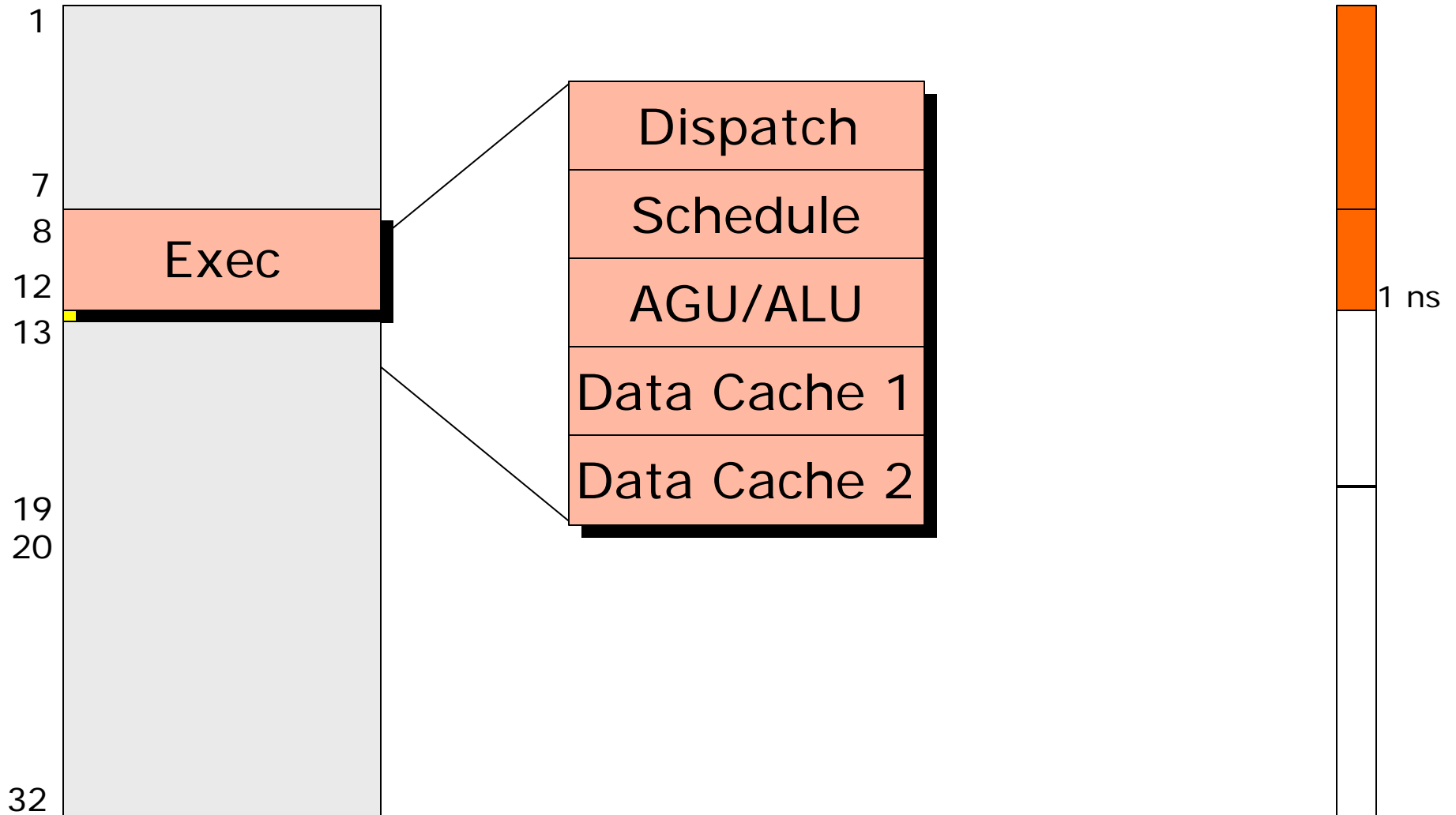
"Hammer" Pipeline



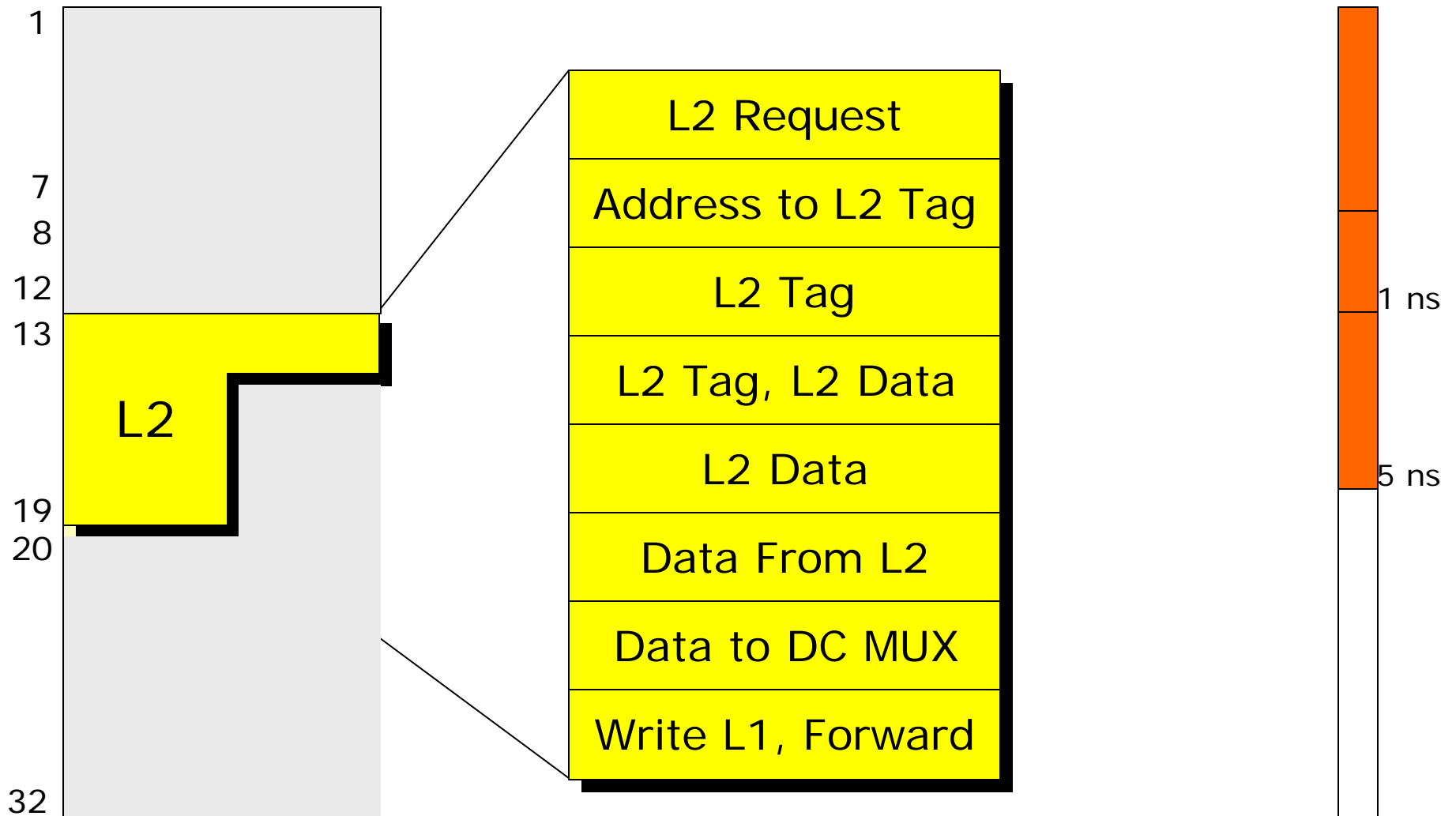
Fetch/Decode Pipeline



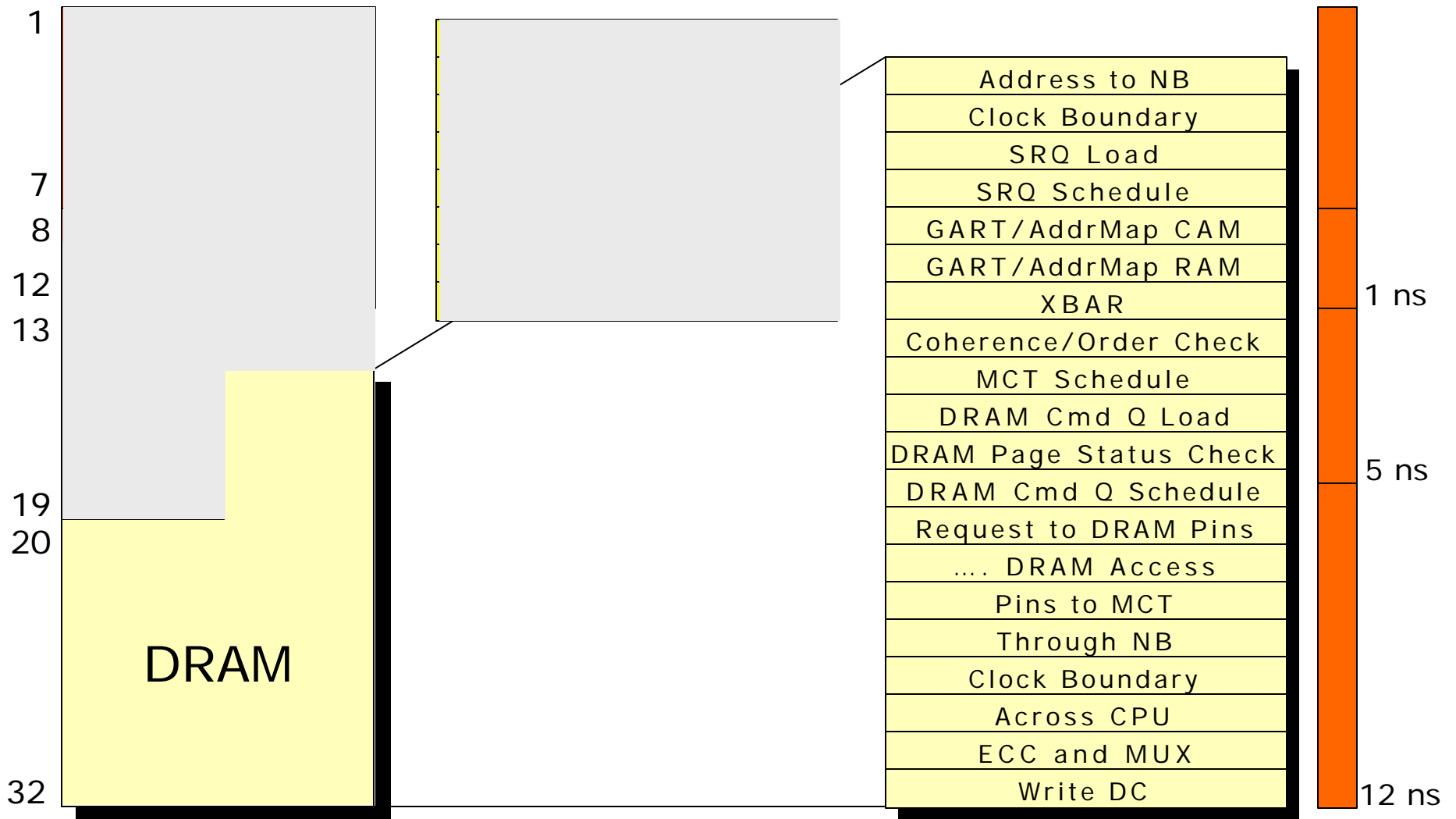
Execute Pipeline



L2 Pipeline



DRAM Pipeline



Large Workload Branch Prediction



- Sequential Fetch



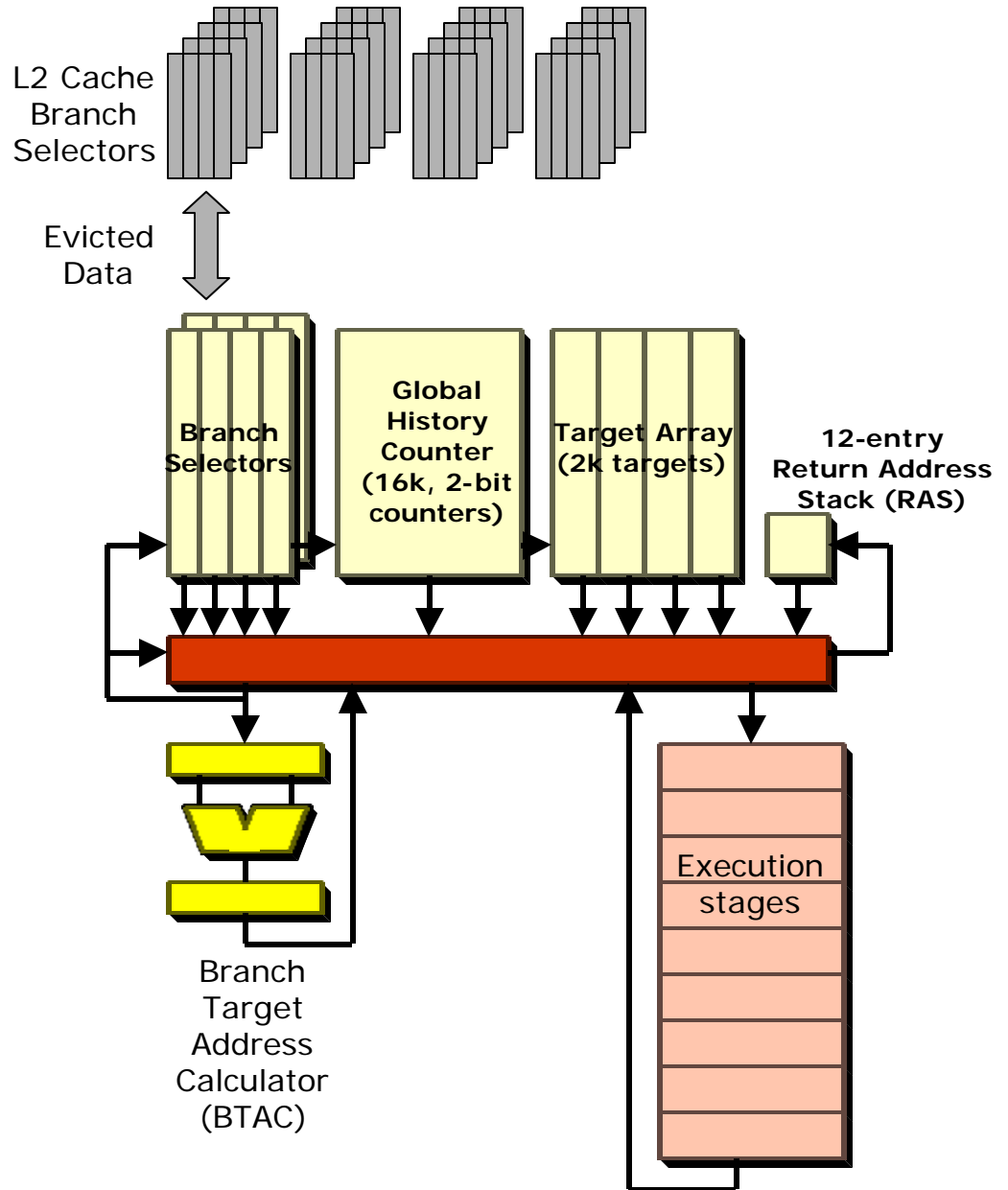
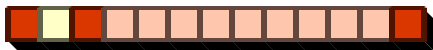
- Predicted Fetch



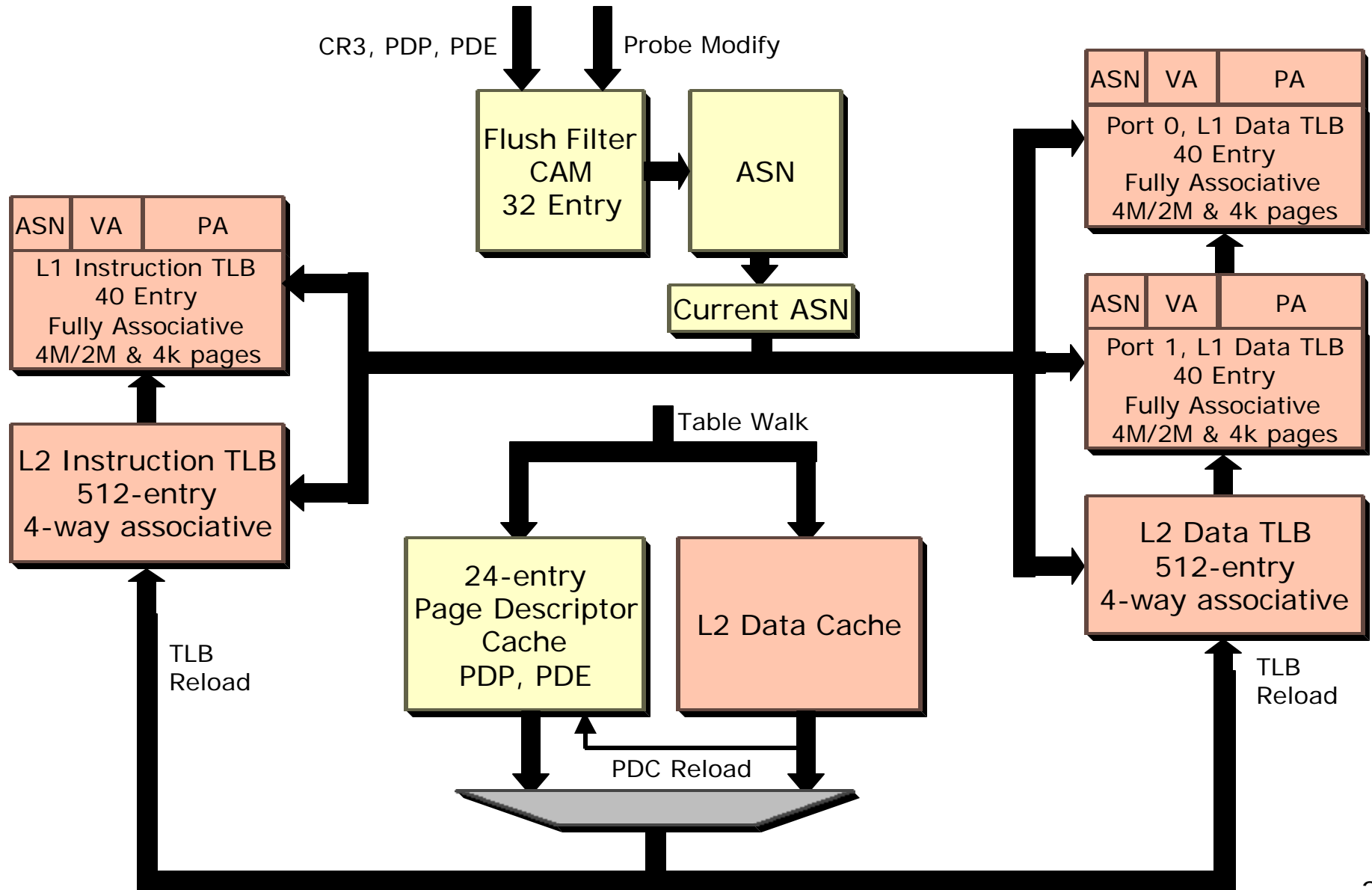
- Branch Target Address Calculator Fetch



- Mispredicted Fetch



Large Workload TLBs



- **Integrated Memory Controller Details**
 - **Memory controller details**
 - 8 or 16-byte interface
 - 16-Byte interface supports
 - **Direct connection to 8 registered DIMMs**
 - **Chipkill ECC**
 - Unbuffered or Registered DIMMs
 - PC1600, PC2100, and PC2700 DDR memory
- **Integrated Memory Controller Benefits**
 - **Significantly reduces DRAM latency**
 - **Memory latency improves**
 - as CPU and HyperTransport™ link speed improves
 - **Bandwidth and capacity grows with number of CPUs**
 - **Snoop probe throughput scales with CPU frequency**

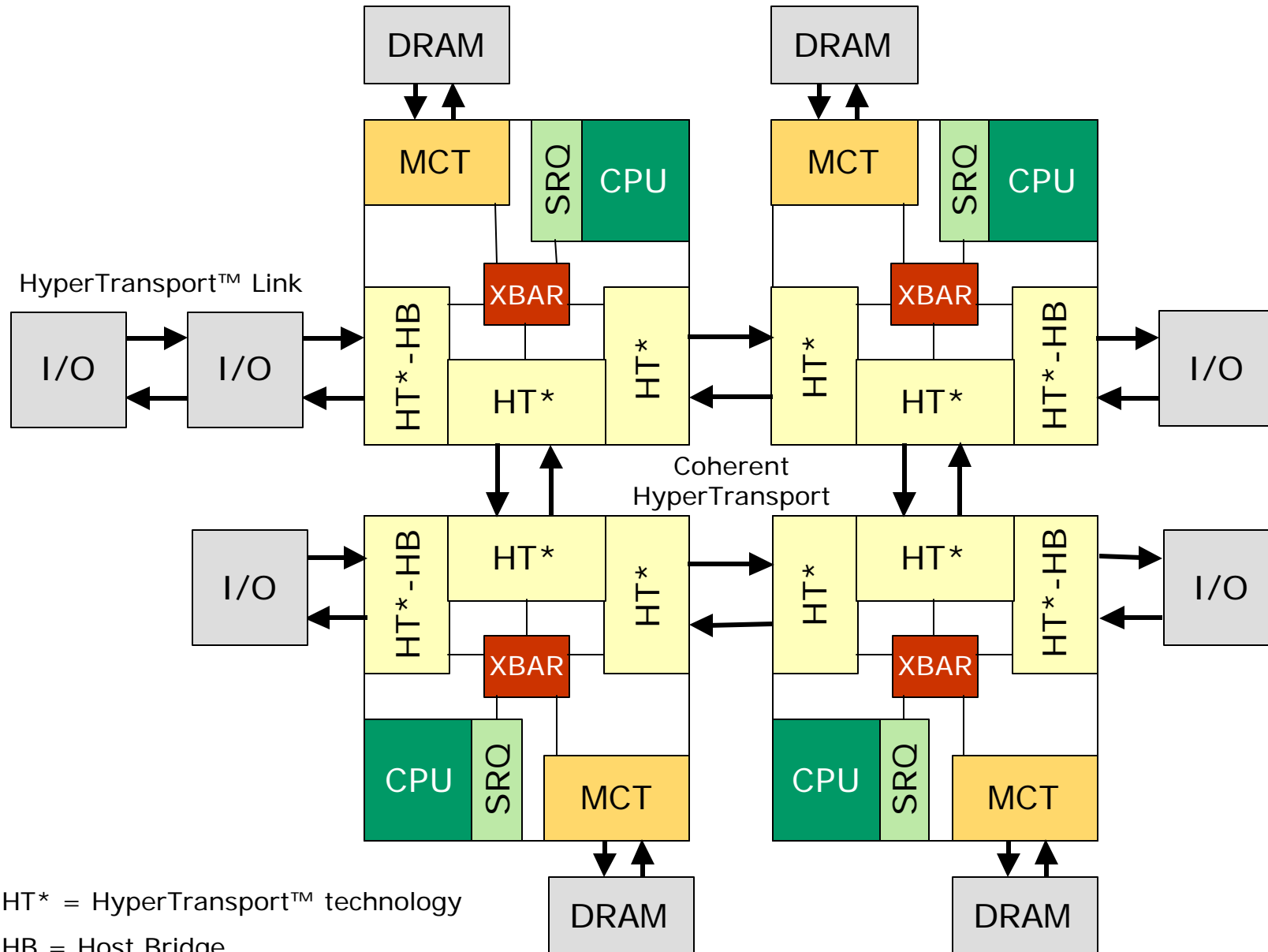
- **L1 Data Cache ECC Protected**
- **L2 Cache AND Cache Tags ECC Protected**
- **DRAM ECC Protected**
 - With Chipkill ECC support
- **On Chip and off Chip ECC Protected Arrays include background hardware scrubbers**
- **Remaining arrays parity protected**
 - L1 Instruction Cache, TLBs, Tags
 - Generally read only data which can be recovered
- **Machine Check Architecture**
 - Report failures and predictive failure results
 - Mechanism for hardware/software error containment and recovery

HyperTransport™ Technology



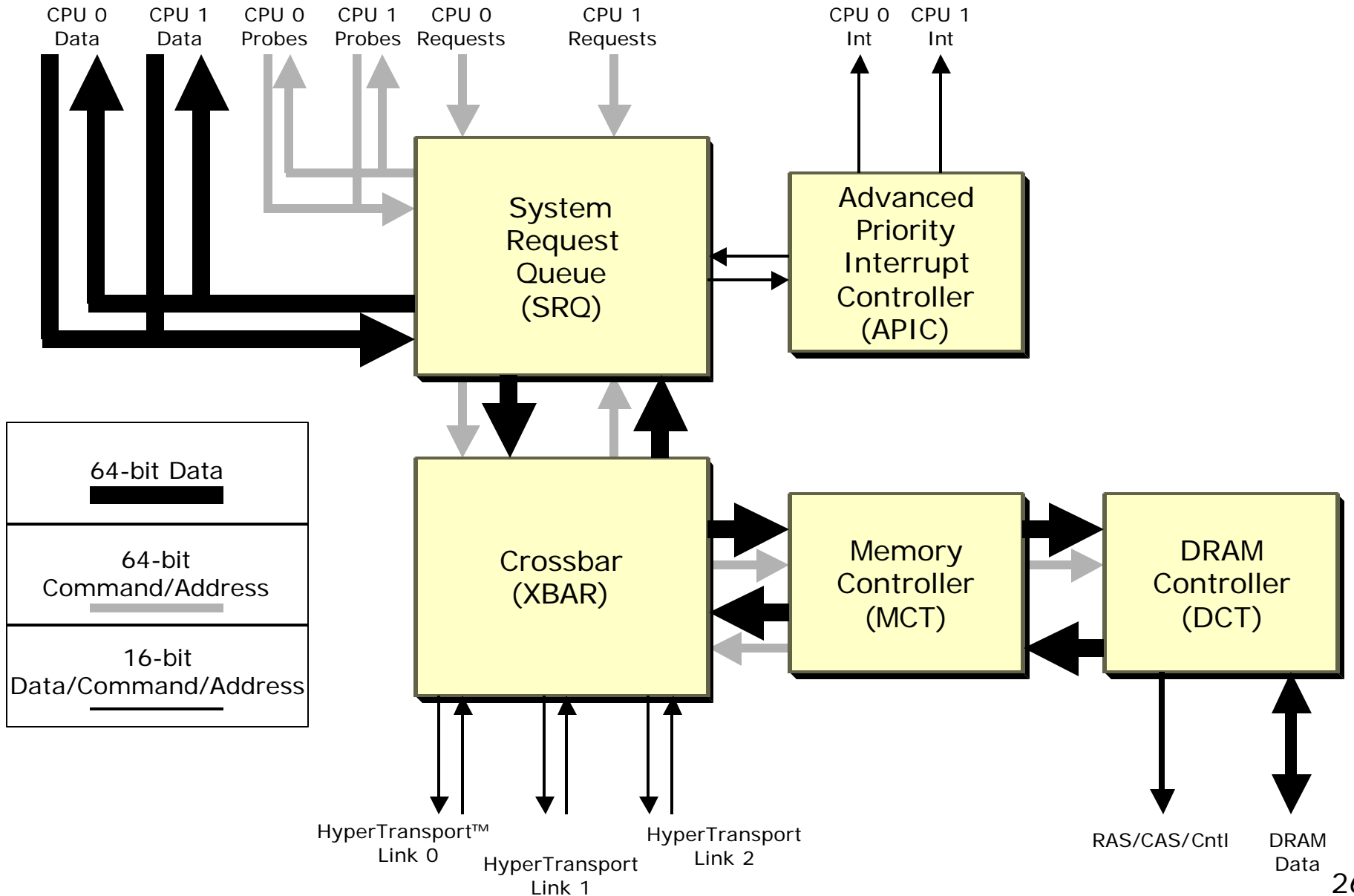
- **Next-generation computing performance goes beyond the microprocessor**
- **Screaming I/O for chip-to-chip communication**
 - High bandwidth
 - Reduced pin count
 - Point-to-point links
 - Split transaction and full duplex
- **Open standard**
 - Industry enabler for building high bandwidth I/O subsystems
 - I/O subsystems: PCI-X, G-bit Ethernet, Infiniband, etc.
- **Strong Industry Acceptance**
 - 100+ companies evaluating specification & several licensing technologies through AMD (2000)
 - First HyperTransport technology-based south bridge announced by nVIDIA (June 2001)
- **Enables scalable 2-8 processor SMP systems**
 - Glueless MP

CPU With Integrated Northbridge

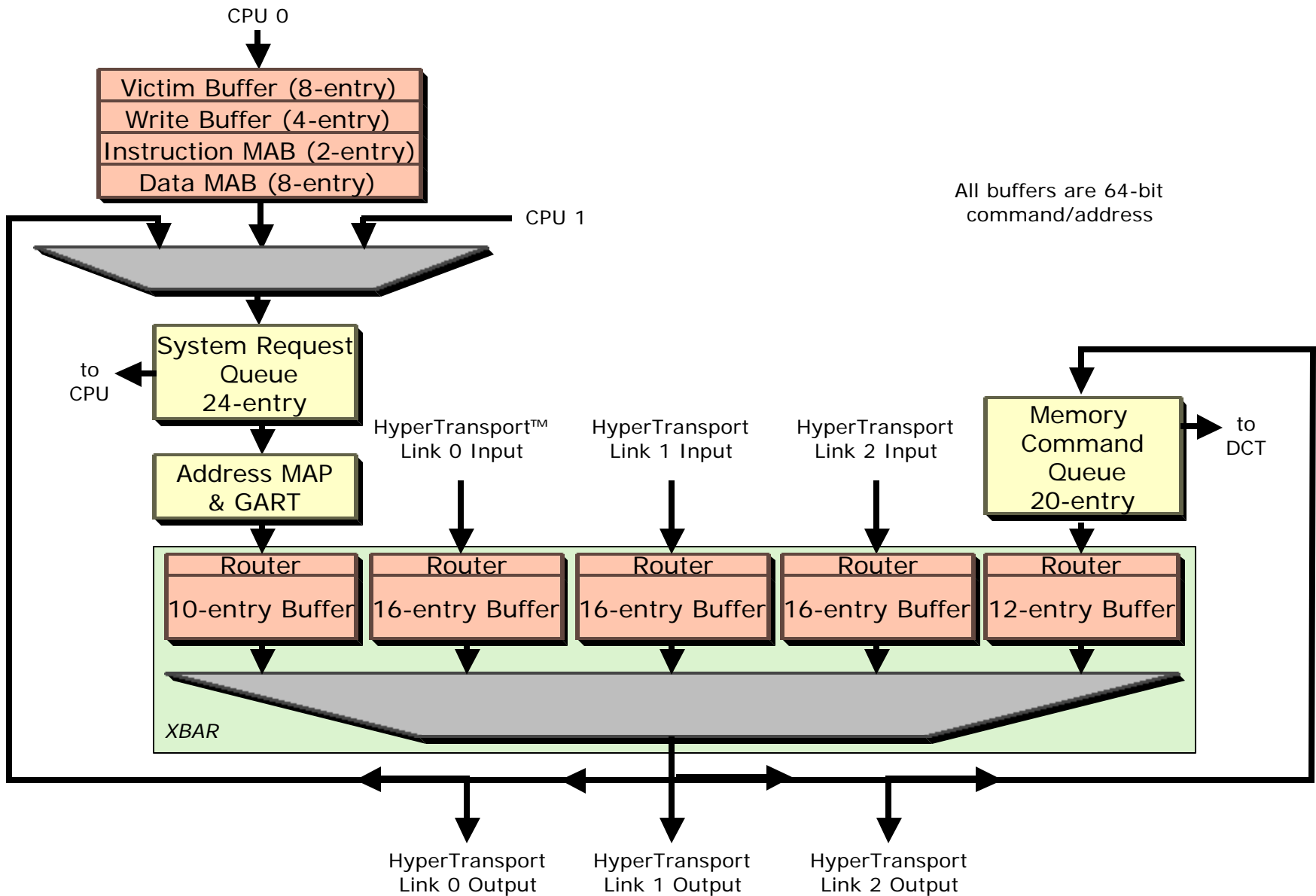


HT* = HyperTransport™ technology
 HB = Host Bridge

Northbridge Overview



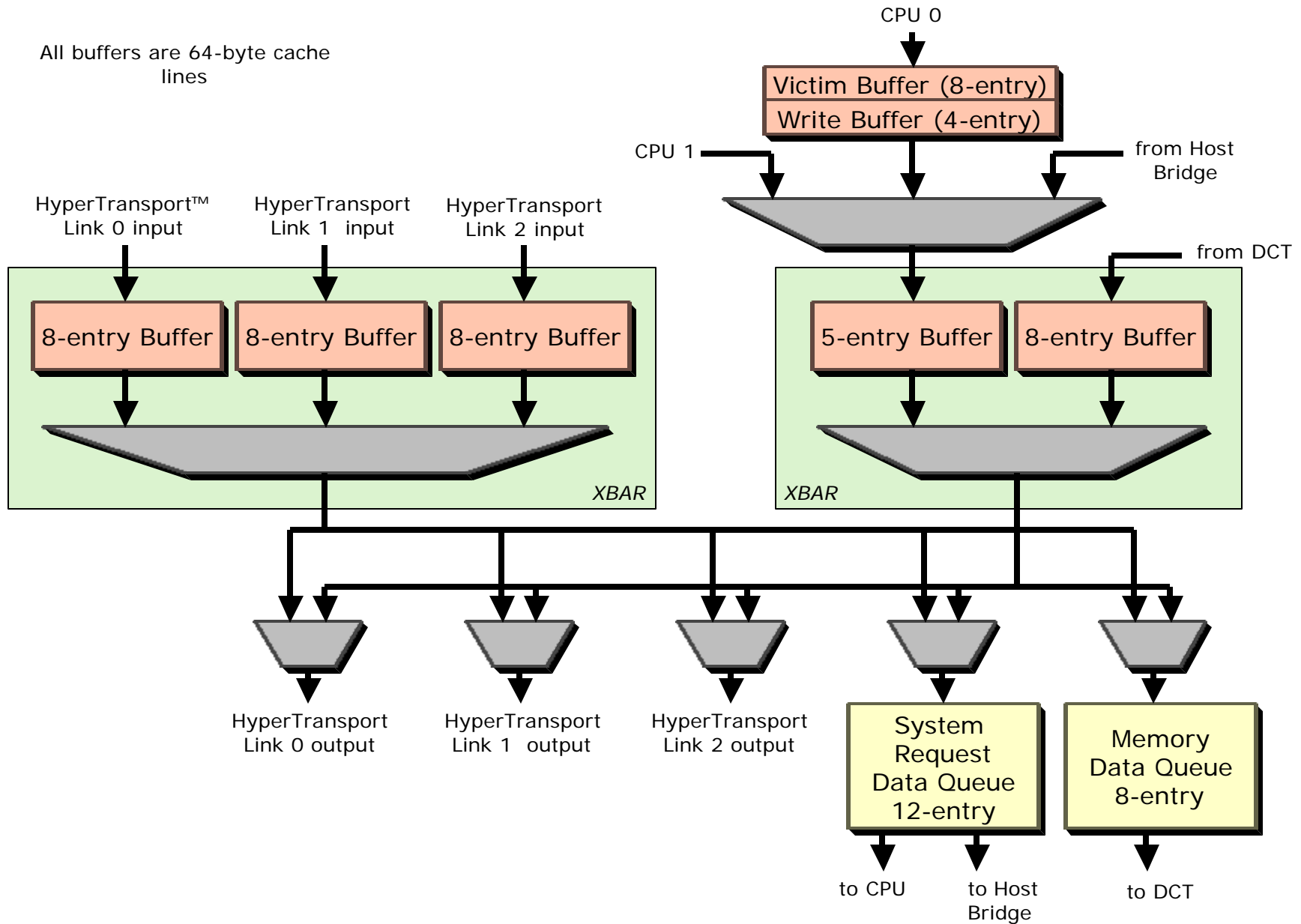
Northbridge Command Flow



Northbridge Data Flow



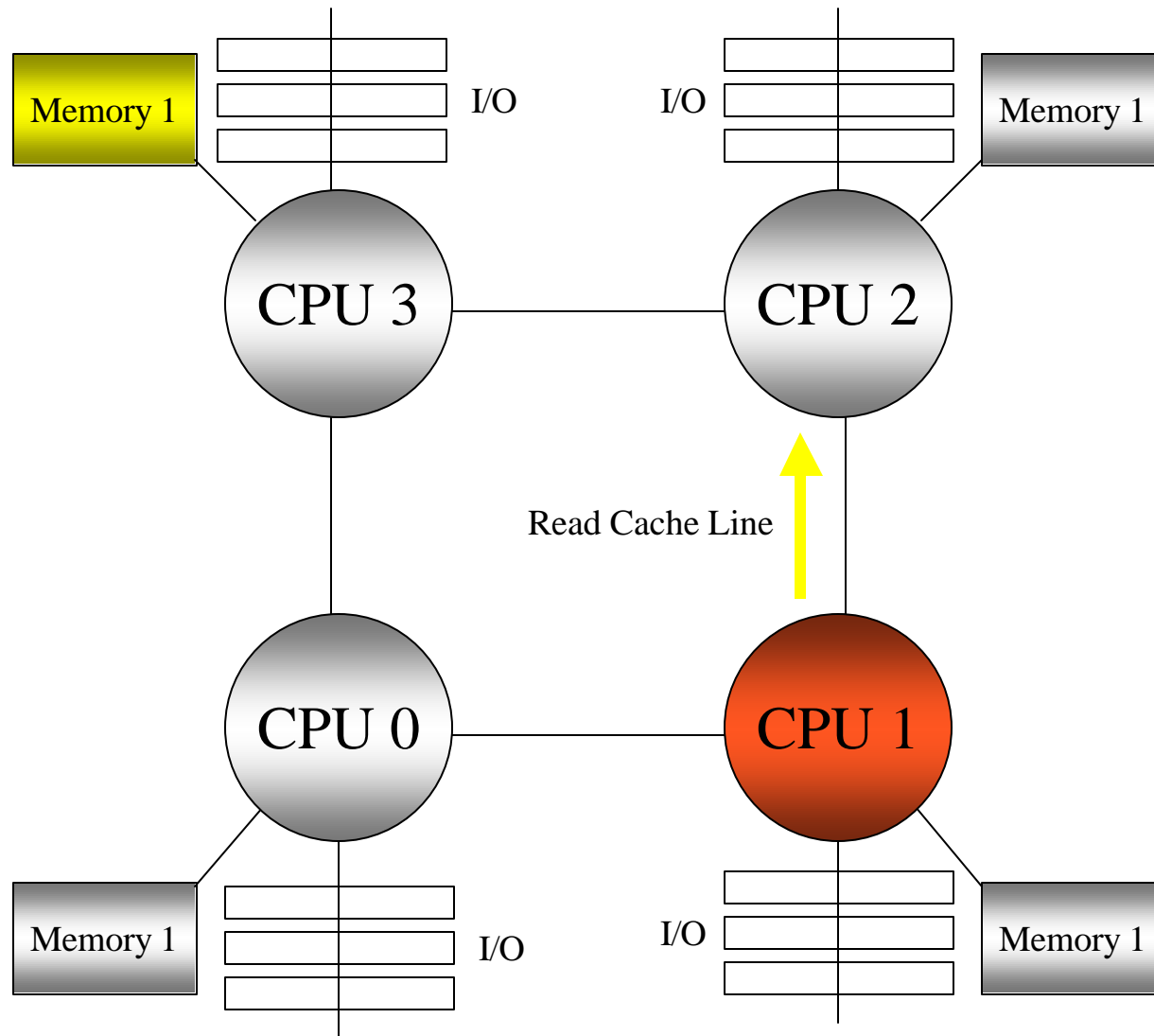
All buffers are 64-byte cache lines



Coherent HyperTransport™ Read Request



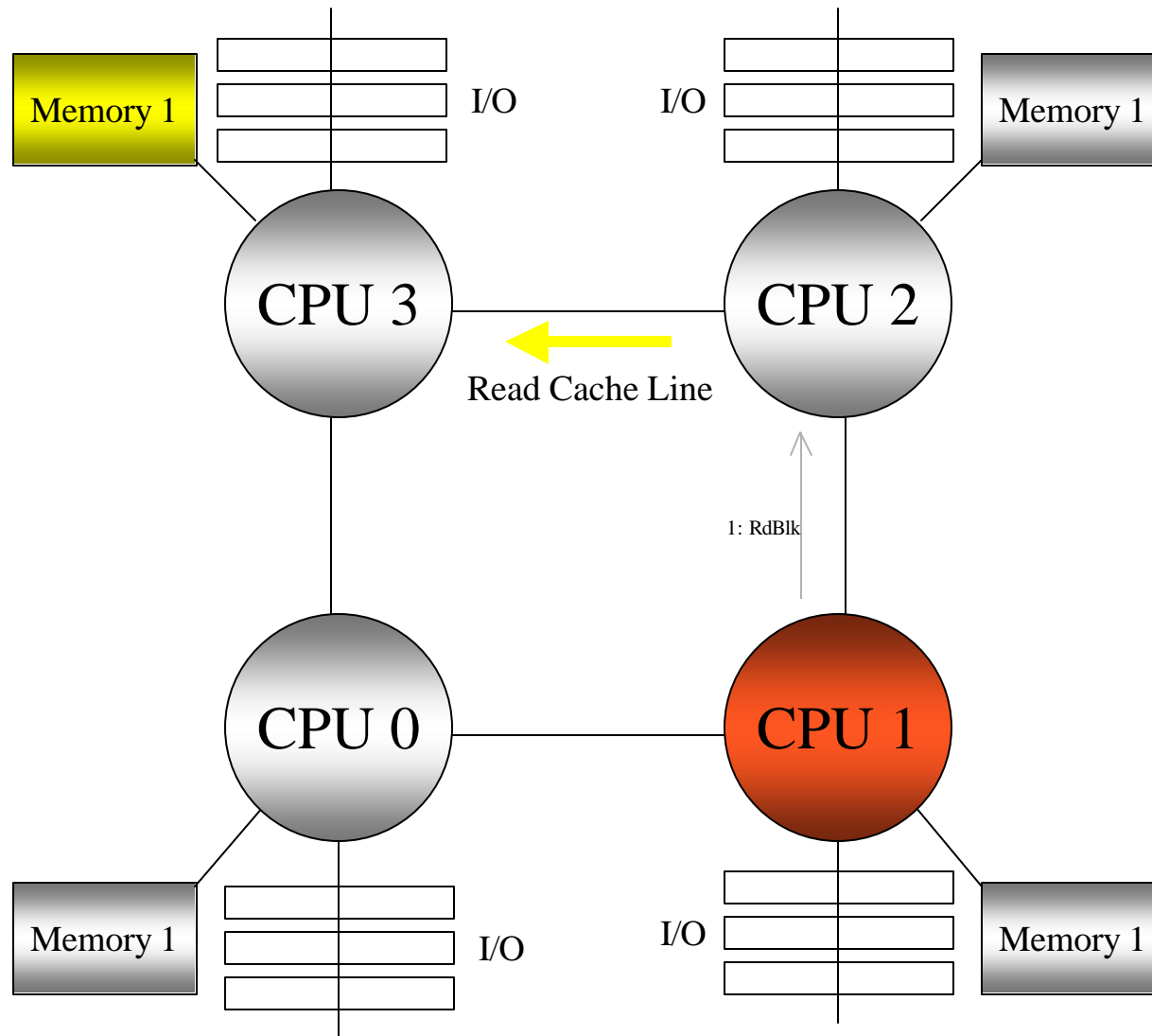
Step 1



Coherent HyperTransport™ Read Request



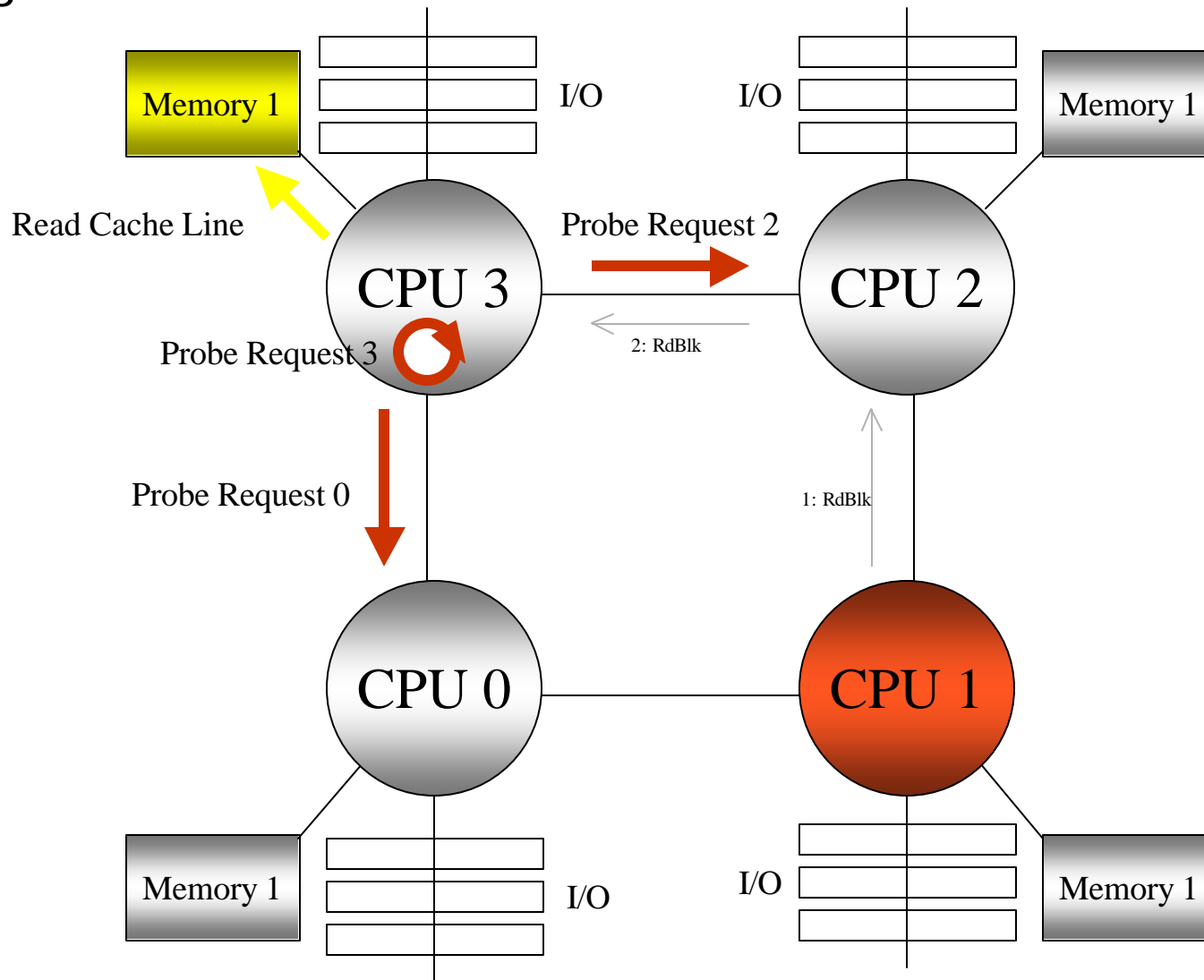
Step 2



Coherent HyperTransport™ Read Request



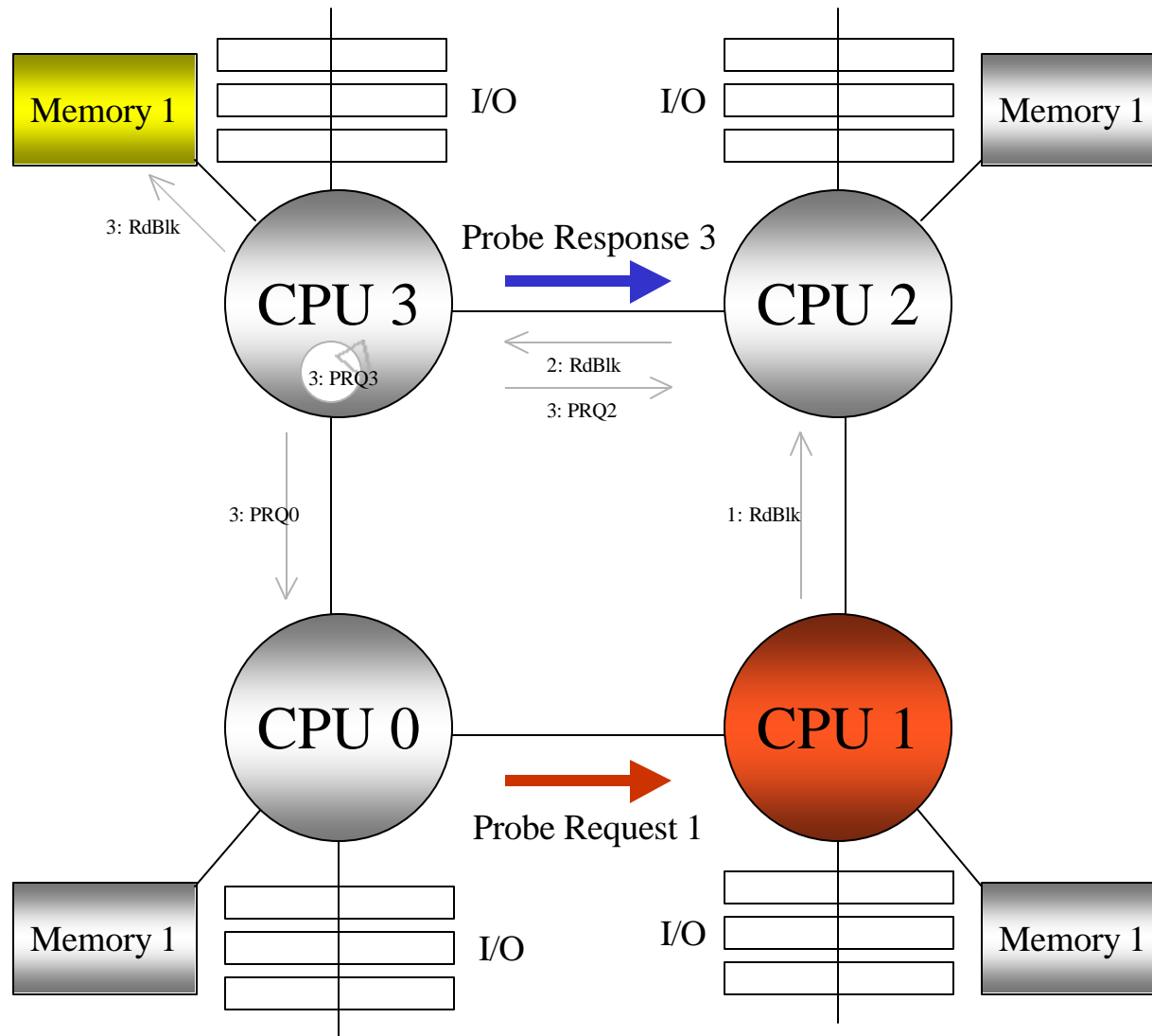
Step 3



Coherent HyperTransport™ Read Request



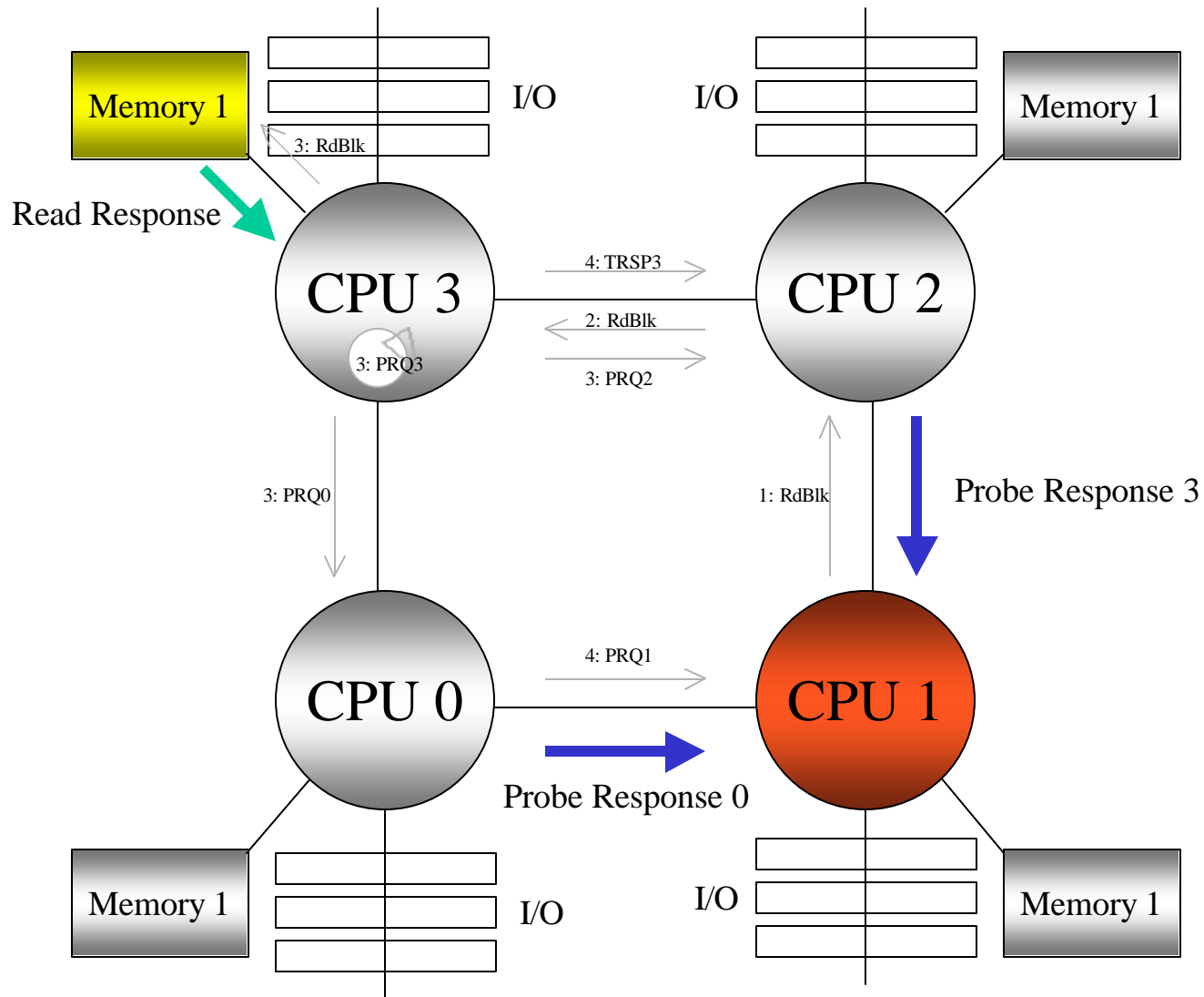
Step 4



Coherent HyperTransport™ Read Request



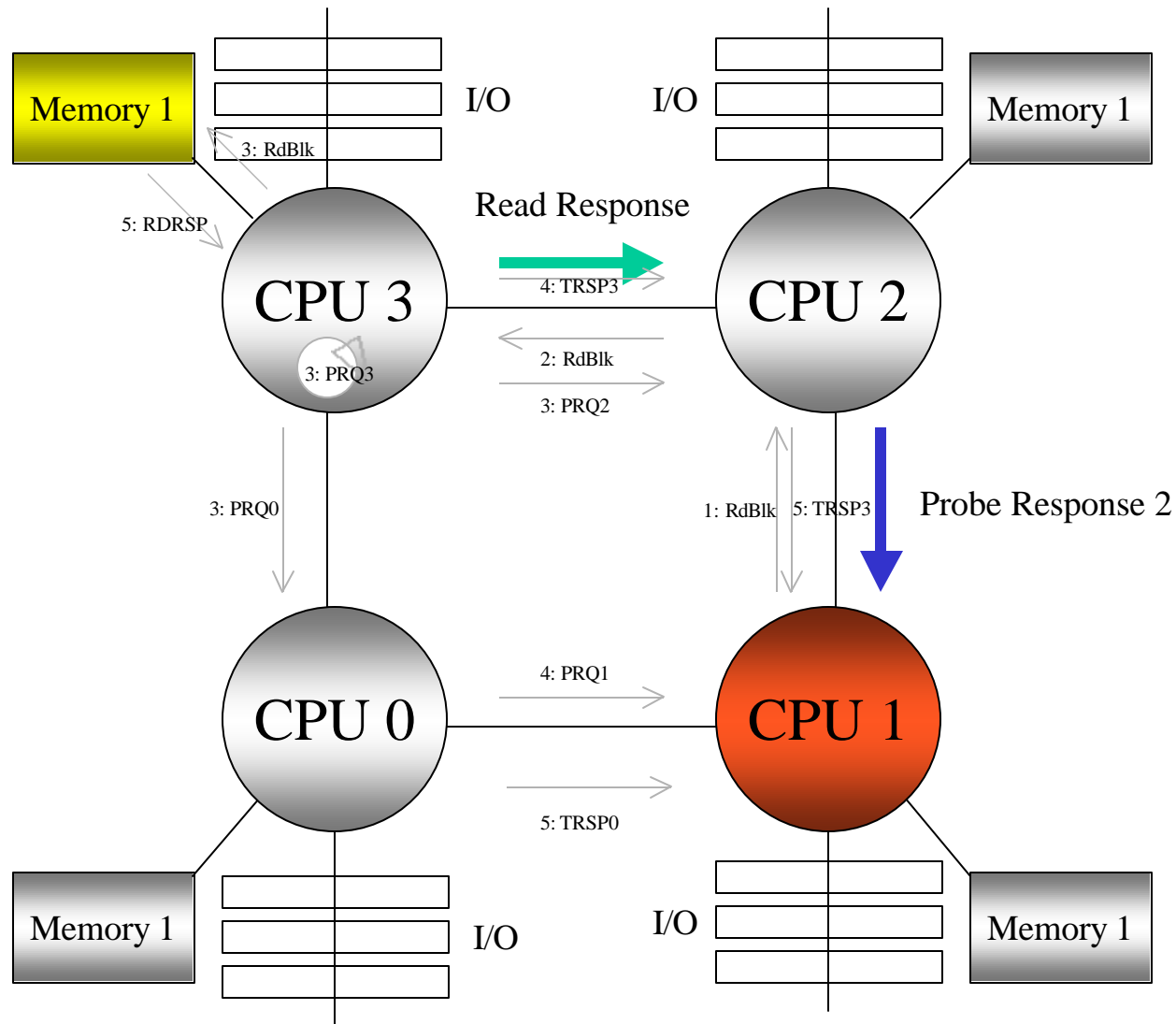
Step 5



Coherent HyperTransport™ Read Request



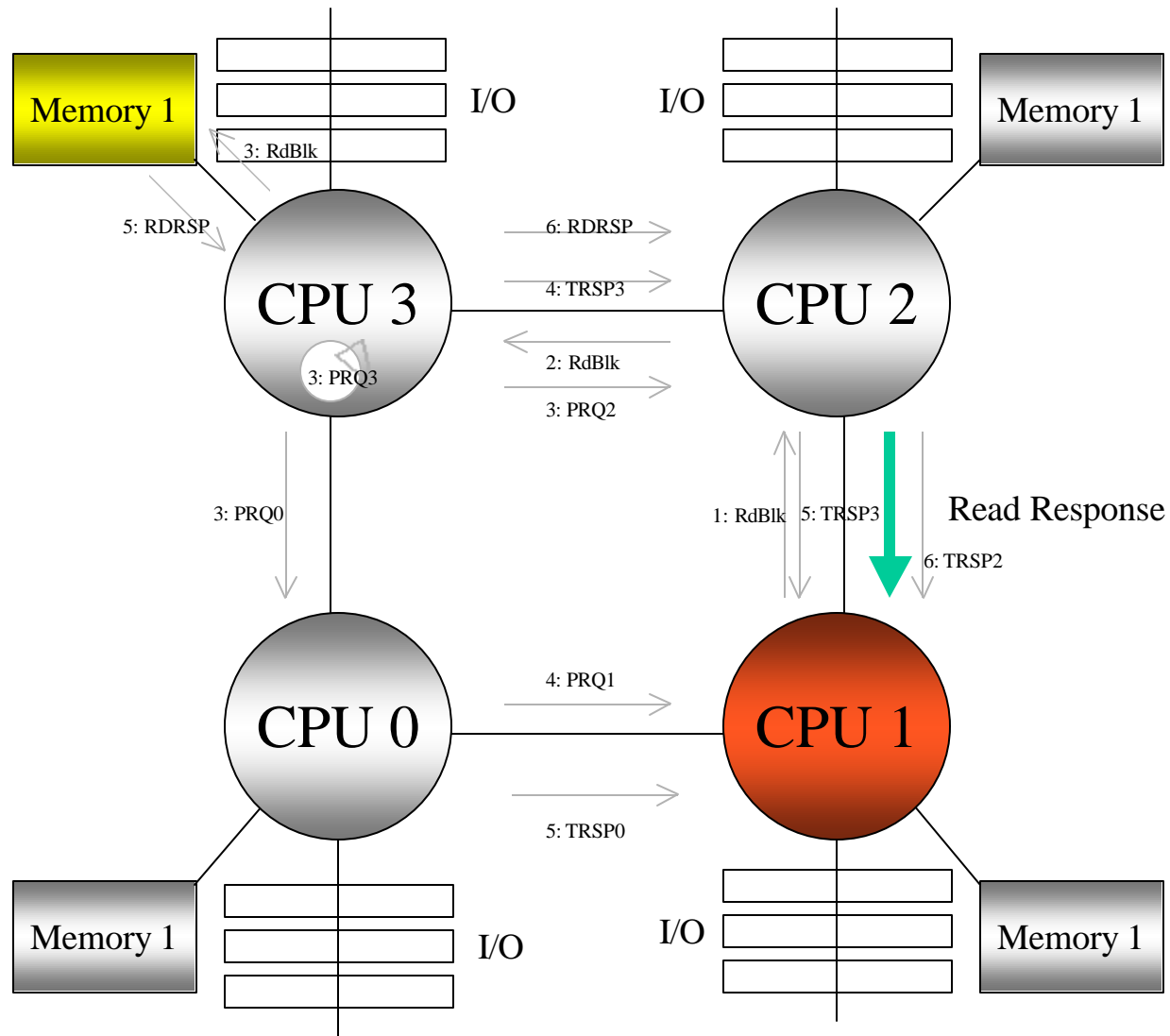
Step 6



Coherent HyperTransport™ Read Request



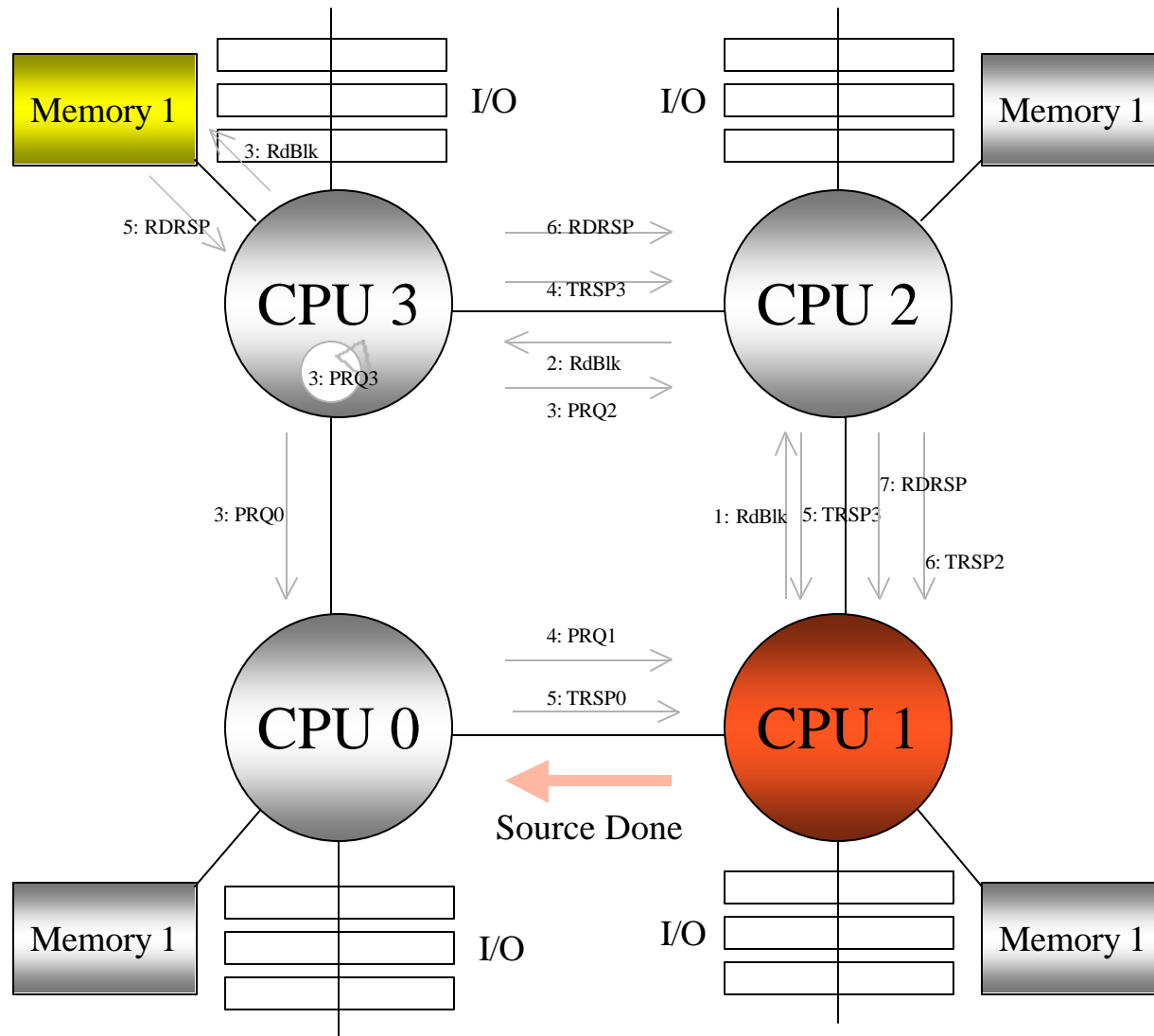
Step 7



Coherent HyperTransport™ Read Request



Step 8

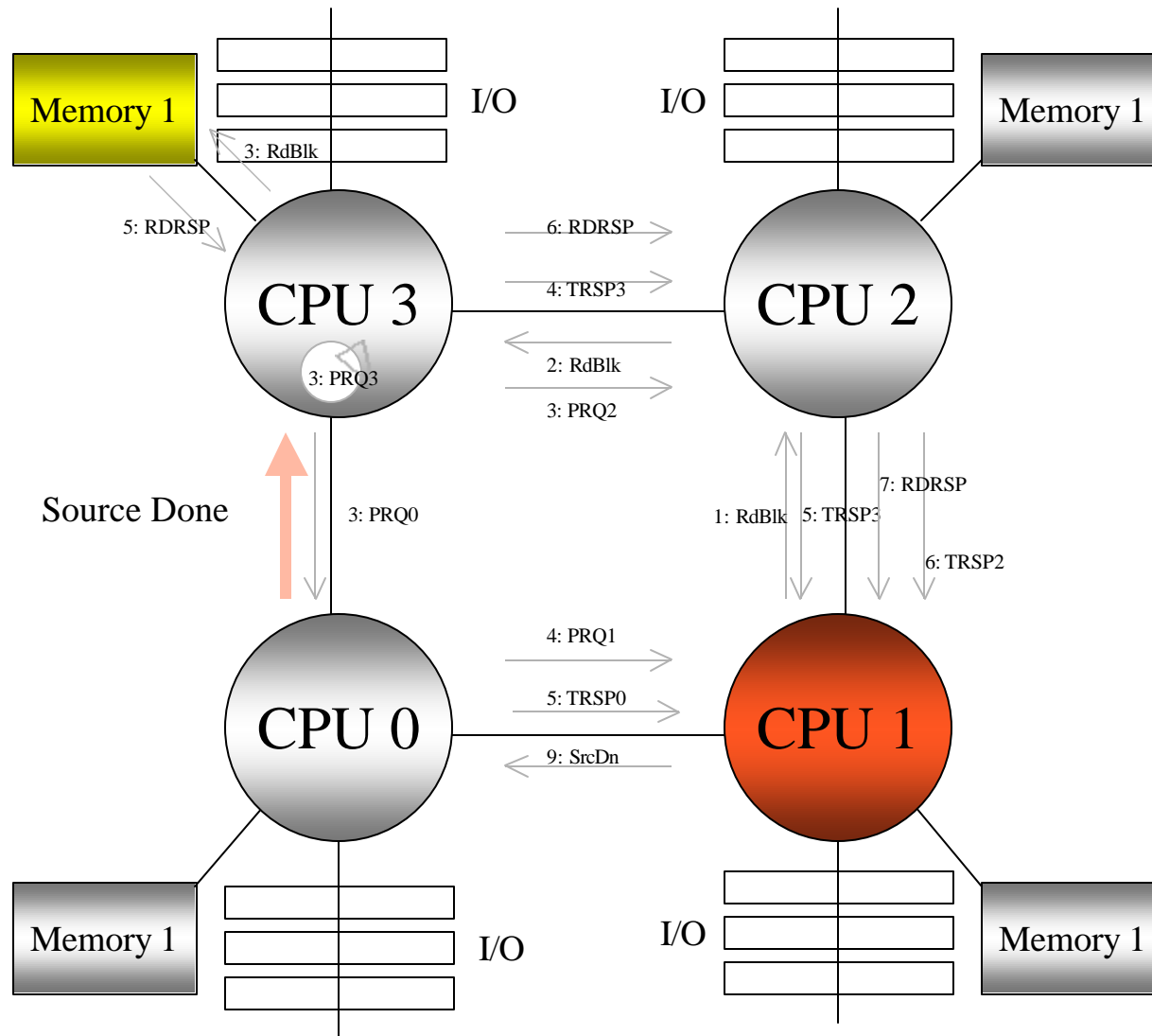


Coherent HyperTransport™

Read Request

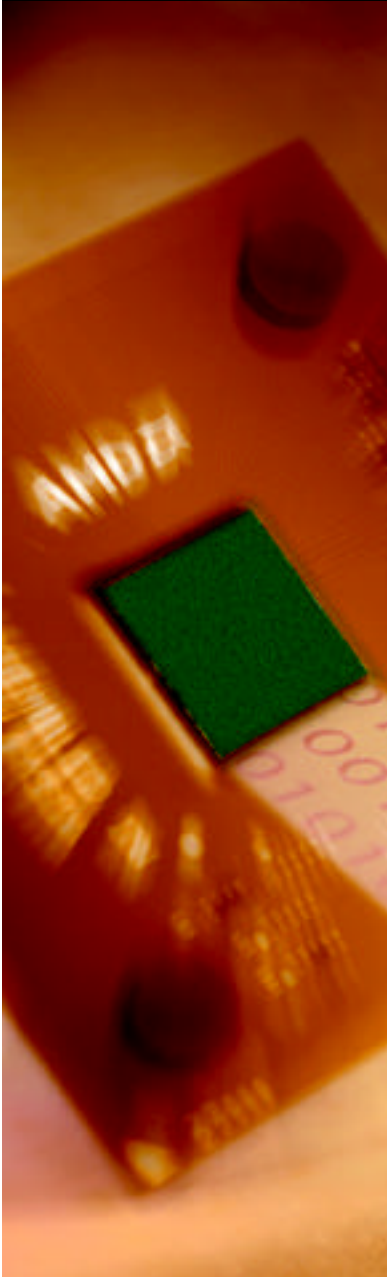


Step 9



"Hammer" Architecture Summary

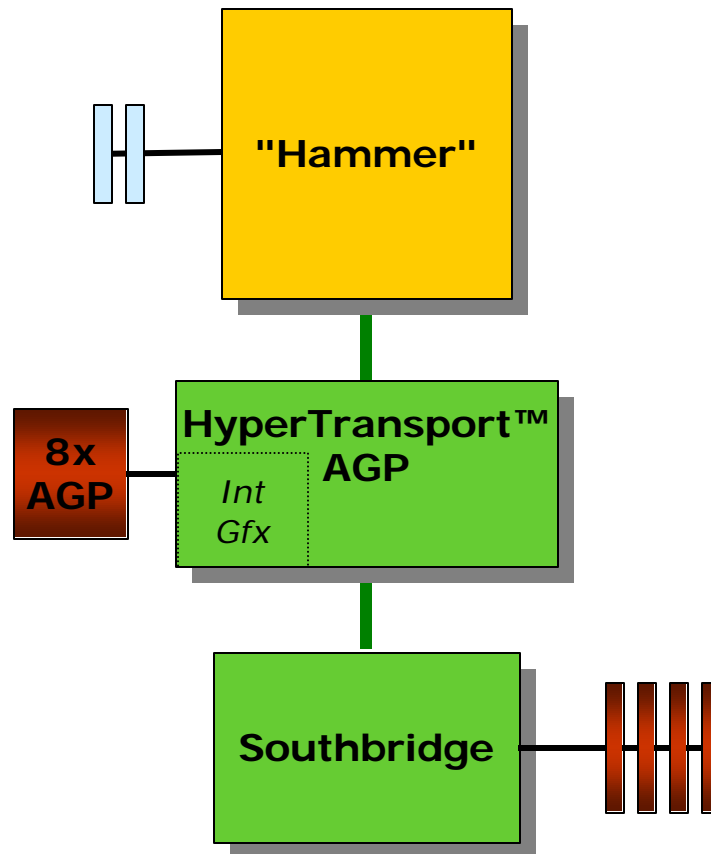
- **8th Generation microprocessor core**
 - Improved IPC and operating frequency
 - Support for large workloads
- **Cache subsystem**
 - Enhanced TLB structures
 - Improved branch prediction
- **Integrated DDR memory controller**
 - Reduced DRAM latency
- **HyperTransport™ technology**
 - Screaming I/O for chip-to-chip communication
 - Enables glueless MP



"Hammer" System Architecture

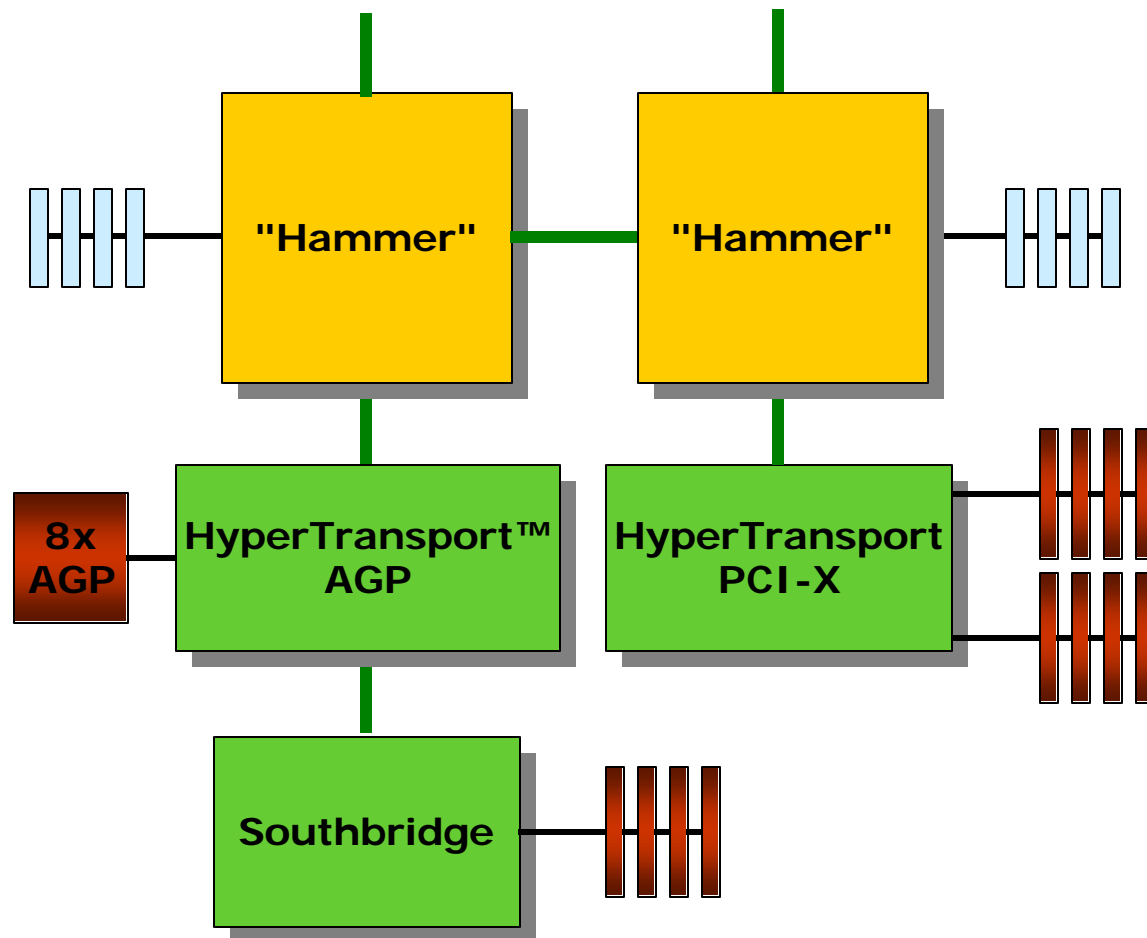
"Hammer" System Architecture

1-way



"Hammer" System Architecture

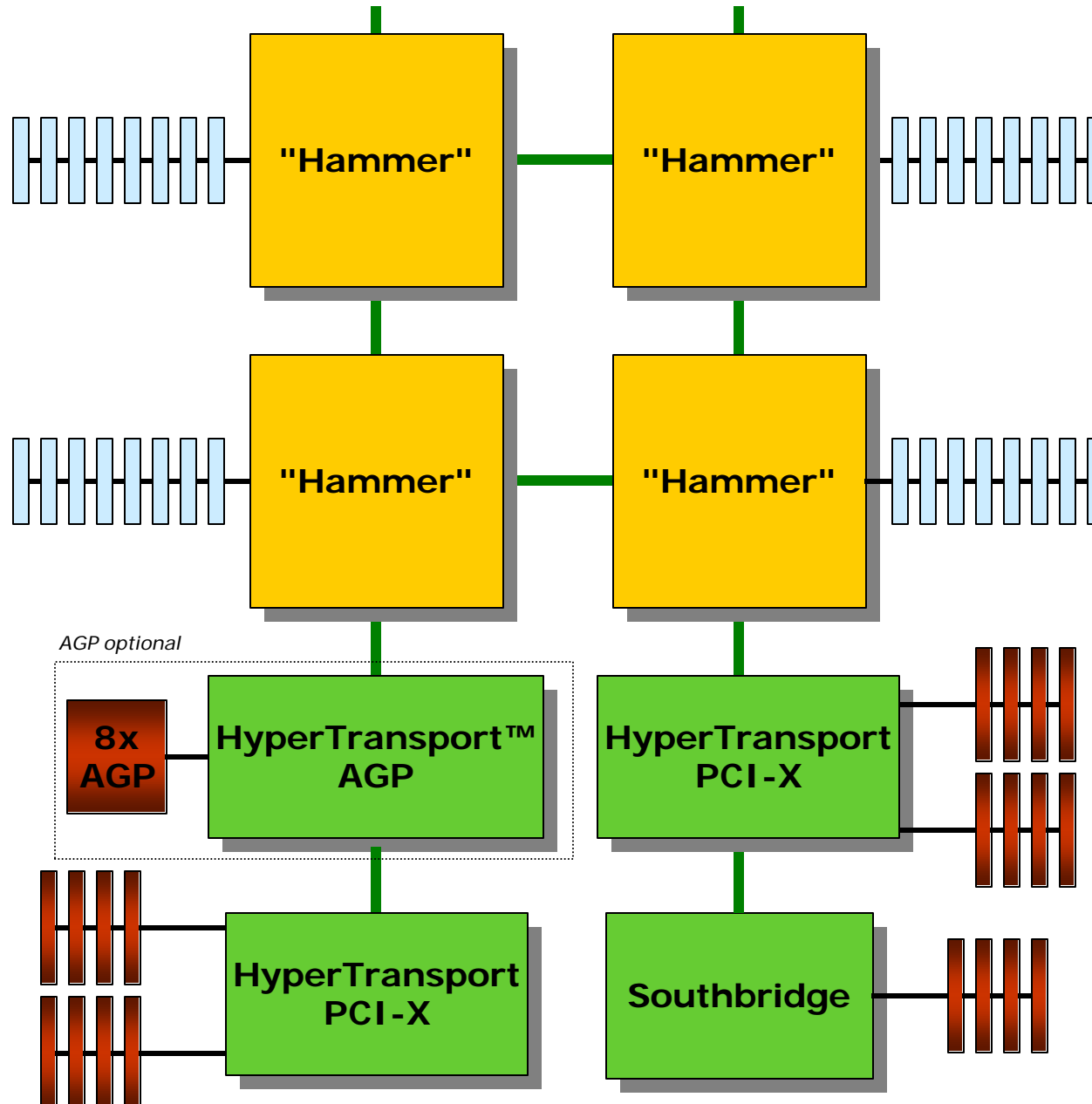
Glueless Multiprocessing: 2-way



"Hammer" System Architecture



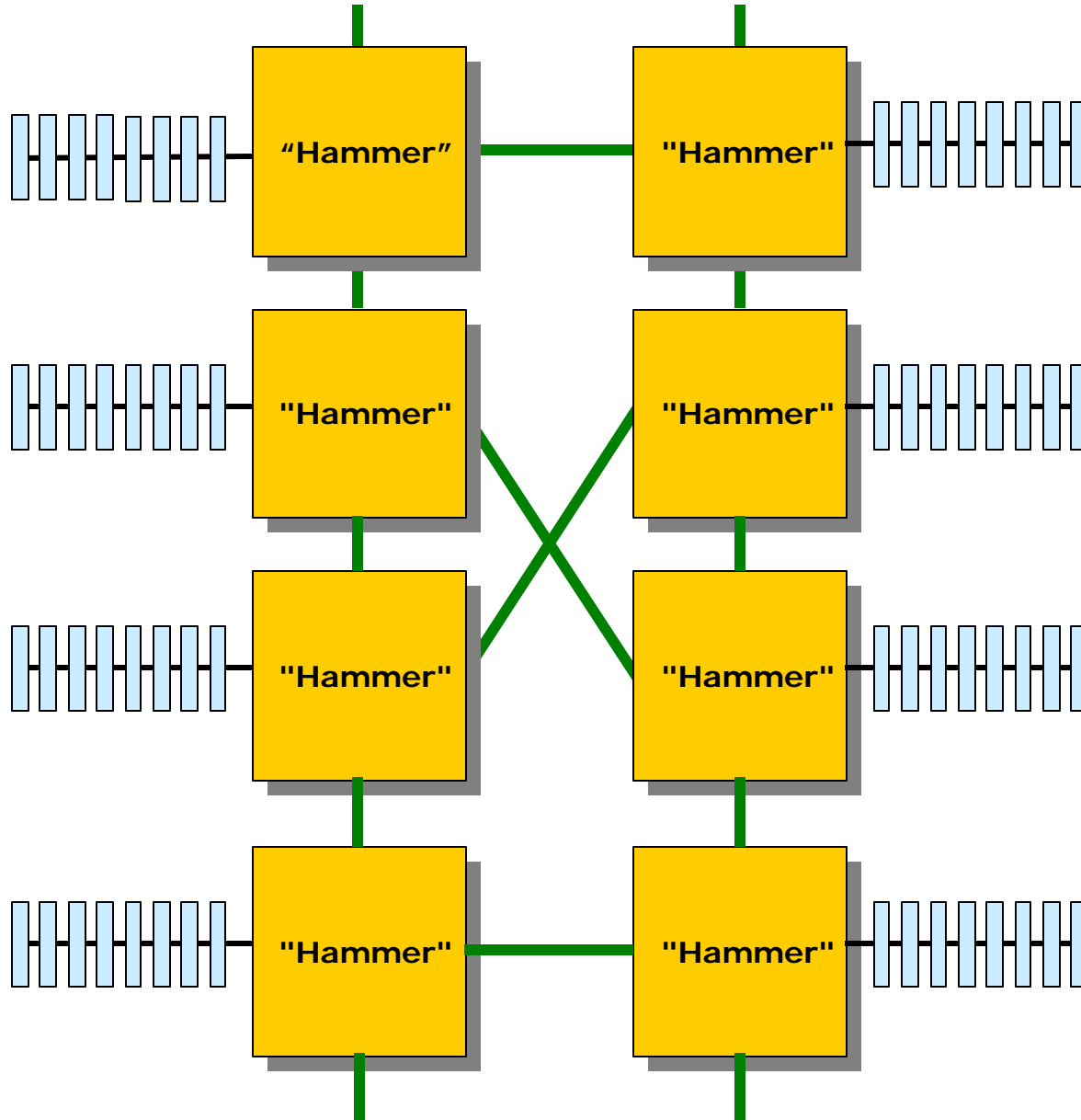
Glueless Multiprocessing: 4-way



"Hammer" System Architecture



Glueless Multiprocessing: 8-way



- **Software view of memory is SMP**
 - Physical address space is flat and fully coherent
 - Latency difference between local and remote memory in an 8P system is comparable to the difference between a DRAM page hit and DRAM page conflict
 - DRAM location can be contiguous or interleaved
- **Multiprocessor support designed in from the beginning**
 - Lower overall chip count
 - All MP system functions use CPU technology and frequency
- **8P System parameters**
 - 64 DIMMs (up to 128GB) directly connected
 - 4 HyperTransport links available for IO (25GB/s)

- **Bandwidth**

- 4P system designed to achieve 8GB/s aggregate memory copy bandwidth
 - With data spread throughout system
- Leading edge bus based systems limited to about 2.1GB/s aggregate bandwidth (3.2GB/s theoretical peak)

- **Latency**

- Average unloaded latency in 4P system (page miss) is designed to be 140ns
- Average unloaded latency in 8P system (page miss) is designed to be 160ns
- Latency under load planned to increase much more slowly than bus based systems due to available bandwidth
- Latency shrinks quickly with increasing CPU clock speed and HyperTransport link speed

"Hammer" Summary



- **8th generation CPU core**
 - Delivering high-performance through an optimum balance of IPC and operating frequency
- **x86-64™ technology**
 - Compelling 64-bit migration strategy without any significant sacrifice of existing code base
 - Full speed support for x86 code base
 - Unified architecture from notebook through server
- **DDR memory controller**
 - Significantly reduces DRAM latency
- **HyperTransport™ technology**
 - High-bandwidth I/O
 - Glueless MP
- **Foundation for future portfolio of processors**
 - Top-to-bottom desktop and mobile processors
 - High-performance 1-, 2-, 4-, and 8-way servers and workstations

©2001 Advanced Micro Devices, Inc.

AMD, the AMD Arrow logo, 3DNow! And combinations thereof are trademarks of Advanced Micro Devices. HyperTransport is a trademark of the HyperTransport Technology Consortium. Other product names are for informational purposes only and may be trademarks of their respective companies.