Agenda

• Glueless MP systems
• MP system configurations
• Cache coherence protocol
• 2-, 4-, and 8-way MP system topologies
• Beyond 8-way MP systems
AMD Opteron™ Processor Architecture

HT = HyperTransport™ technology

3.2 GB/s per direction @ 1600 MHz Data Rate

3.2 GB/s per direction @ 1600 MHz Data Rate

3.2 GB/s per direction @ 1600 MHz Data Rate

5.3 GB/s 128-bit

HT = HyperTransport™ technology
Glueless MP System

HT = HyperTransport™ technology
MP Architecture

• Programming model of memory is effectively SMP
  – Physical address space is flat and fully coherent
  – Far to near memory latency ratio in a 4P system is designed to be < 1.4
  – Latency difference between remote and local memory is comparable to the difference between a DRAM page hit and a DRAM page conflict
  – DRAM locations can be contiguous or interleaved
  – No processor affinity or NUMA tuning required

• MP support designed in from the beginning
  – Lower overall chip count results in outstanding system reliability
  – Memory Controller and XBAR operate at the processor frequency
  – Memory subsystem scale with frequency improvements
MP Architecture (contd.)

• **Integrated Memory Controller**
  – 333 MHz 128-bit DRAM interface with up to 8 registered DIMMs
  – High-bandwidth (5.3 GB/s peak) and low-latency memory access
  – Snoop throughput scales with Processor frequency
  – Broadcast cache coherence protocol
    • Avoids serialization delay of directory based systems
    • Snooping the processors caches is overlapped with DRAM access
HyperTransport™ Technology

• **Screaming I/O for chip-to-chip communication**
  – High bandwidth
  – Point-to-point links
  – Split transaction and full duplex
  – Differential Signaling
  – Tunneling capability

• **Enables scalable 2-8 processor Cache-Coherent MP systems**
  – Glueless MP

• **HyperTransport™ Links**
  – Up to three 16-bit links (3.2 GB/s per direction)
  – Reduced pin count compared to the typical Bus based systems
  – Compatible with high-volume PC board infrastructure
  – Each can be:
    • cHT: coherent (Processor-to-Processor) link or,
    • HT: non-coherent (Processor-to-I/O) link
  – For more info see:  http://www.HyperTransport.org/
High–Performance Workstation Implementation

AMD-8151™ HyperTransport™ AGP3.0 Graphics Tunnel

AMD Opteron™

6.4GB/s HyperTransport

32bits @ 533Mhz

AGP 3.0

AMD-8131™ HyperTransport™ PCI-X Tunnel

AMD Opteron™

6.4GB/s coherent HyperTransport

64bits @ 133Mhz

PCI-X Hot Plug

1000 BaseT

U320 SCSI

Ethernet

Gbit Ethernet

PCI-Express

Legacy PCI

32bits @ 33Mhz

800MB/s HyperTransport

144-Bit Reg DDR

200-333MHz

PCI-X

64bits @ 133Mhz

144-Bit Reg DDR

200-333MHz

6.4GB/s HyperTransport

Ethernet

USB 2.0

AC’97

SIO

LPC

SM Bus

EIDE

FLASH

Gbit Ethernet

1000 BaseT

U320 SCSI

SM Bus

AC’97

USB 2.0

Ethernet

EIDE

FLASH

LPC

SIO
4-Way Server Implementation

- AMD Opteron™
- AMD Opteron
- AMD-8131™ HyperTransport™ PCI-X Tunnel
- AMD-8111™ HyperTransport™ I/O Hub
- PCI-X
- 1000 BaseT
- U320 SCSI
- Ethernet
- Gbit Ethernet
- AC’97
- USB 2.0
- Ethernet
- EIDE
- SIO
- LPC
- SM Bus

6.4GB/s coherent HyperTransport™
6.4GB/s coherent HyperTransport
6.4GB/s coherent HyperTransport
6.4GB/s coherent HyperTransport

200-333MHz 144-Bit Reg DDR
200-333MHz 144-Bit Reg DDR
200-333MHz 144-Bit Reg DDR
4P System — Board Layout
8-Way Implementation
Local vs. Remote memory access

- Local Memory Access (0-hop)
- Remote1 Memory Access (1-hop)
- Remote2 Memory Access (2-hops)
Cache Coherence Protocol
Read Transaction Example

Step 1

Read Cache Line
Cache Coherence Protocol
Read Transaction Example

Step 2

Memory 0 ➔ P0 ➔ P1 ➔ Memory 1

Memory 2 ➔ P2 ➔ P3 ➔ Memory 3

Read Cache Line
Cache Coherence Protocol
Read Transaction Example

Step 3

Read Cache Line
Snoop Request P0
Snoop Request P2

P0
P1
P2
P3

Memory 0
Memory 1
Memory 2
Memory 3
Cache Coherence Protocol
Read Transaction Example

Step 4

Memory 0

P0

Snoop Response P0

P1

Memory 1

P2

Snoop Request P3

P3

Memory 2

Memory 3

P0

P1

P2

P3
Cache Coherence Protocol
Read Transaction Example

Step 5

Memory 0
Read Response M0

P0

Memory 1

Snoop Response P0

P1

P2

Snoop Response P2

P3

Memory 2

Memory 3
Cache Coherence Protocol
Read Transaction Example

Step 6

P0

Memory 0

P1

Memory 1

Read Response M0

P2

Snoop Response 1

P3

Memory 2

Memory 3
Cache Coherence Protocol
Read Transaction Example

Step 7

Memory 0

P0

Memory 2

P2

Memory 3

P3

Memory 1

P1

Read Response M0
Cache Coherence Protocol
Read Transaction Example

Step 8
Cache Coherence Protocol
Read Transaction Example

Step 9

Memory 0

P0

Memory 2

P2

Source Done to M0

Memory 3

P3

Memory 1

P1
Cache Coherence Protocol
Read Transaction Example

Step 10

P0

P1

P3

P2

Memory 0

Memory 1

Memory 2

Memory 3

Source Done to M0
2-way System Topology

- System parameters
  - 16 DIMMs (up to 32 GB using 256Mb DRAM)
  - 2 HyperTransport links available for I/O
  - Bisection-bandwidth = 6.4 GB/s
  - Diameter = 1 hop
4-way System Topology

- **System parameters**
  - 32 DIMMs (up to 64 GB using 256Mb DRAM)
  - 4 HyperTransport links available for I/O
  - Bisection-bandwidth = 12.8 GB/s
  - Average-diameter = 1.33 Hops
4-way System Topology (contd.)

- System parameters
  - 32 DIMMs (up to 64 GB using 256Mb DRAM)
  - 2 HyperTransport links available for I/O
  - Bisection-bandwidth = 19.2 GB/s
  - Average-diameter = 1.17 Hops
8-way System Topology

- System parameters
  - 64 DIMMs (up to 128GB using 256Mb DRAM)
  - 4 HyperTransport links available for I/O
  - Bisection-bandwidth = 25.6 GB/s
  - Average-diameter = 1.71 hops
8-way System Topology (contd.)

- System parameters
  - 64 DIMMs (up to 128GB using 256Mb DRAM)
  - 2 HyperTransport links available for I/O
  - Bisection-bandwidth = 32 GB/s
  - Average-diameter = 1.64 hops
Scalability Beyond 8P

- Scaling beyond 8P is enabled
  - External HyperTransport switch

- Coherent Interconnect
  - Snoop filter
  - Data caching
The Rewards of Good Plumbing

• **High Bandwidth**
  – 2P system is designed to achieve 7 GB/s aggregate memory Read bandwidth
  – 4P system is designed to achieve 10 GB/s aggregate memory Read bandwidth
    • With data spread uniformly across the nodes

• **Low Latency**
  – Average 2P unloaded latency (page hit) is designed to be < 120 ns
  – Average 4P unloaded latency (page hit) is designed to be < 140 ns
  – Latency under load increases slowly due to excess Interconnect Bandwidth
  – Latency shrinks quickly with increasing CPU clock speed and HyperTransport link speed
Trademark Attribution

AMD, the AMD Arrow Logo, AMD Opteron and combinations thereof are trademarks of Advanced Micro Devices, Inc. HyperTransport is a licensed trademark of the HyperTransport Consortium. Other product names used in this presentation are for identification purposes only and may be trademarks of their respective companies.