The AMD Opteron™ CMP NorthBridge Architecture: Now and in the Future

Pat Conway & Bill Hughes

August, 2006
AMD Opteron™ – The Industry’s First Native Dual-Core 64-bit x86 Processor

Integration:
- Two 64-bit CPU cores
- 2MB L2 cache
- On-chip Router & Memory Controller

Bandwidth:
- Dual channel DDR (128-bit) memory bus
- 3 HyperTransport™ (HT) links (16-bit each x 2 GT/sec x 2)

Usability and Scalability:
- Socket compatible: Platform and TDP!
- Glueless SMP up to 8 sockets
- Memory capacity & BW scale w/ CPUs

Power Efficiency:
- AMD PowerNow!™ Technology with optimized power management
- Industry-leading system level power efficiency
AMD Opteron™ – The Industry’s First Native Dual-Core 64-bit x86 Processor
A Clean Break with the Past

Legacy x86 Architecture
- 20-year old traditional front-side bus (FSB) architecture
- CPUs, Memory, I/O all share a bus
- Major bottleneck to performance
- Faster CPUs or more cores ≠ performance

AMD64’s Direct Connect Architecture
- Industry-standard technology
- Direct Connect Architecture reduces FSB bottlenecks
- HyperTransport™ interconnect offers scalable high bandwidth and low latency
- 4 memory controllers – increases memory capacity and bandwidth
4P System — Board Layout
System Overview

12.8 GB/s
128-bit

ncHT

12.8 GB/s per direction @ 2GT/s Data Rate

“NorthBridge”
Northbridge Microarchitecture Overview

### Virtual Channel Use

<table>
<thead>
<tr>
<th>Virtual Channel</th>
<th>Use</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request</td>
<td>• Read</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>• Non-posted Write</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Cache Block Commands</td>
<td></td>
</tr>
<tr>
<td>Posted Request</td>
<td>Posted Writes</td>
<td>Y</td>
</tr>
<tr>
<td>Probe</td>
<td>Broadcast probes</td>
<td>N</td>
</tr>
<tr>
<td>Response</td>
<td>• Read Response</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>• Probe response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Completion</td>
<td></td>
</tr>
</tbody>
</table>

2 DDR2 channels
Northbridge Command Flow

Core 0
- Victim Buffer (8-entry)
- Write Buffer (4-entry)
- Instruction MAB (2-entry)
- Data MAB (8-entry)

Core 1

System Request Queue 24-entry

Address MAP & GART
- HT0 Input
- HT1 Input
- HT2 Input

Memory Command Queue 20-entry

All buffers are 64-bit command/address
Northbridge Data Flow

All buffers are 64-byte cache lines

- Victim Buffer (8-entry)
- Write Buffer (4-entry)
- 5-entry Buffer
- 8-entry Buffer
- 8-entry Buffer
- 8-entry Buffer
- System Request Data Queue (12-entry)
- Memory Data Queue (8-entry)

XBAR

Core 0
- Victim Buffer (8-entry)
- Write Buffer (4-entry)
- from DCT
- from Host Bridge

Core 1
- 5-entry Buffer
- 8-entry Buffer
- from Host Bridge

HT0 input
- 8-entry Buffer
- HT0 output

HT1 input
- 8-entry Buffer
- HT1 output

HT2 input
- 8-entry Buffer
- HT2 output

System Request Data Queue (12-entry)
- to Core

Memory Data Queue (8-entry)
- to Host Bridge
- to DCT
**Lessons Learned #1**

*Allocation of XBAR Command buffer across Virtual Channels can have big impact on performance*

MP traffic analysis gives the best allocation

*e.g. Opteron Read Transaction*

- **Request (2 visits)**
- **Probe (3 visits)**
- **Response (8 visits)**
Lessons Learned #2
Memory Latency is the Key to Application Performance!

Performance vs Average Memory Latency
(single 2.8GHz core, 400MHz DDR2 PC3200, 2GT/s HT with 1MB cache in MP system)

- OLTP1
- OLTP2
- SW99
- SSL
- JBB

- 1N
- 2N
- 4N (SQ)
- 8N (TL)
- 8N (L)

System Performance:
- AvgD
- Latency
  - 0 hops: x * 0ns
  - 1 hops: x * 44ns (124 cpuclk)
  - 1.8 hops: x * 105ns (234 cpuclk)
  - 0.5 hops: x * 17ns (47 cpuclk)
  - 1.5 hops: x * 76ns (214 cpuclk)
Looking Forward
HyperTransport™-based Accelerators

Imagine it, Build it

- Open platform for system builders (“Torrenza”)
  - 3rd Party Accelerators
  - Media
  - FLOPs
  - XML
  - SOA

- AMD Opteron™ Socket or HTX slot

- HyperTransport interface is an open standard see hypertransport.org

- Coherent HyperTransport interface available if the accelerator caches system memory (under license)
AMD’s Next Generation Processor Technology

Native quad core die

Ideal for 65nm SOI and beyond

Expandable shared L3 cache

Enhanced Direct Connect Architecture and Northbridge

IPC enhanced CPU cores

- 32B instruction fetch
- Improved branch prediction
- Out-of-order load execution
- Up to 4 DP FLOPS/cycle
- Dual 128-bit SSE dataflow
- Dual 128-bit loads per cycle
- Bit Manipulation extensions (LZCNT/POPCNT)
- SSE extensions (EXTRQ/INSERTQ, MOVNTSD/MOVNTSS)

Four ungangable x16 HyperTransport™ links (up to 5.2GT/sec)

Enhanced crossbar

Next-generation memory support

FBDIMM when appropriate

Enhanced power management and RAS
Balanced, Highly Efficient Cache Structure

Efficient memory handling reduces the need for “brute force” cache sizes

**Dedicated L1**
- Locality keeps most critical data in the L1 cache
- Low latency
- 2 128 bit data paths
- 2 loads per cycle
Balanced, Highly Efficient Cache Structure

Efficient memory handling reduces the need for "brute force" cache sizes

Dedicated L2
- Sized to accommodate the majority of working sets today
- Dedicated to help eliminate conflicts common in shared caches
Balanced, Highly Efficient Cache Structure

Efficient memory handling reduces the need for “brute force” cache sizes

Shared L3 – Coming Soon

- Allocation policy which optimizes movement, placement and replication of data for multi-core
- Ready for expansion
Additional HyperTransport™ Ports

- Enable Fully Connected 4 Node (four x16 HT) and 8 Node (eight x8 HT)
- Reduced network diameter
  - Fewer hops to memory
- Increased Coherent Bandwidth
  - more links
  - cHT packets visit fewer links
  - HyperTransport3
- Benefits
  - Low latency because of lower diameter
  - Evenly balanced utilization of HyperTransport links
  - Low queuing delays

Low latency under load
4 Node Performance

**4N SQ (2GT/s HyperTransport)**
- Diam 2 Avg Diam 1.00
- XFIRE BW 14.9GB/s

**4N FC (2GT/s HyperTransport)**
- Diam 1 Avg Diam 0.75
- XFIRE BW 29.9GB/s

**4N FC (4.4GT/s HyperTransport3)**
- Diam 1 Avg Diam 0.75
- XFIRE BW 65.8GB/s

**W/ HYPERTRANSPORT3**
- (2X)

**+ 2 EXTRA LINKS**

XFIRE ("crossfire") BW is the **link-limited** all-to-all communication bandwidth (data only)
8 Node Performance

8N Twisted Ladder

- P0
- P1
- P2
- P3
- P4
- P5
- P6
- P7

I/O

8N TL (2GT/s HyperTransport)

- Diam 3 Avg Diam 1.62
- XFIRE BW: 15.2GB/s

8 Node 6HT 2x4

- P0
- P1
- P2
- P3
- P4
- P5
- P6
- P7

I/O

8N 2x4 (4.4GT/s HyperTransport3)

- Diam 2 Avg Diam 1.12
- XFIRE BW: 72.2GB/s

8 Node Fully Connected

- P0
- P1
- P2
- P3
- P4
- P5
- P6
- P7

I/O

8N FC (4.4GT/s HyperTransport3)

- Diam 1 Avg Diam 0.88
- XFIRE BW: 94.4GB/s

(5X)

(6X)

OR
Why Quad-Core?

Baseline is 2 Node x 2 Core blade running OLTP
Increasing Frequency

Baseline is 2 Node x 2 Core blade running OLTP
Decreasing Frequency

Baseline is 2 Node x 2
Core blade running OLTP
Quad-Core
Higher Performance within a Fixed Power Budget

Baseline is 2 Node x 2 Core blade running OLTP
Clock and Power Planes

Clock and Power Planes

VRM

VDD

DIMMs

SVI

Misc

VDDIO, VTT

PLLs

Northbridge

Core 0

PLL

Core 2

PLL

Core 1

PLL

Core 3

PLL

VDDA

VDDNB

VHT

PLLs

HyperTransport Links

21 August 2006 The Opteron CMP NorthBridge Architecture, Now and in the Future
DICE: Dynamic Independent Core Engagement

Ability to dynamically and individually adjust core frequencies to improve power efficiency
DICE: Dynamic Independent Core Engagement

Ability to dynamically and individually adjust core frequencies to improve power efficiency

100% Workload
33% Workload
33% Workload
33% Workload

60% Power State
DICE: Dynamic Independent Core Engagement

Ability to dynamically and individually adjust core frequencies for improved power efficiency

100% Workload

50% Workload

Halted

Halted

45% Power State
Enjoy the rest of the conference!

www.amd.com/power
Trademark Attribution

AMD, the AMD Arrow, AMD Opteron, AMD PowerNow logo and combinations thereof are trademarks of Advanced Micro Devices, Inc. HyperTransport and HTX are trademarks of the HyperTransport Consortium. PCIe is a trademark of the PCI-SIG. Other names used in this presentation are for informational purposes only and may be trademarks of their respective owners.