Credit where credit is due:
Slides contain original artwork (© Jacob, Wang 2005)
System Controller

Heavy demand placed on memory system
Heavier still in SMP/SMT/CMP system
System Controller == System traffic cop
Memory Request Overview

**Part A: Searching on-chip for data**
- Fetch
- Decode
- Exec
- Mem
- WB

**Part B: Going off-chip for data**
- I/O to memory traffic
- physical to memory address translation
- memory request scheduling
- read data buffer
- RAM
- DRAM

**Steps not required for some processor/system controllers, protocol dependant.**

Progression of a Memory Read Transaction Request Through Memory System
**“Memory Latency”**

A: Transaction request may be delayed in Queue
B: Transaction request sent to Memory Controller
C: Transaction converted to Command Sequences (may be queued)
D: Command/s Sent to DRAM
E₁: Requires only a **CAS** or
E₂: Requires **RAS** + **CAS** or
E₃: Requires **PRE** + **RAS** + **CAS**
F: Transaction sent back to CPU

“DRAM Latency” = A + B + C + D + E₁ + E₂/E₃
Small System Topologies

Classic small system topology (Lots of systems)

Point-to-point processor-controller system topology (AMD Athlon/Alpha EV6/PPC 970)

Integrated system controller system topology (AMD Opteron/Alpha EV7 etc.)

● represents point of synchronization* (for local access)
System Controller: Athlon

MRO: Memory Request Organizer
APC: AGP PCI Controller block
MCT: Memory Controller (SDRAM/DDR/DRDRAM)
**MRO:** Memory Request Organizer

- Request crossbar responsible for scheduling memory read and write requests from BIU, PCI, AGP
- Serves as the coherence point
- Requests are reordered to minimize page conflict and maximize page hits
- Anti-starvation mechanism by aging of entries
- Arbitration bypassed during idle conditions to improve latency
### AMD Athlon Controller:

<table>
<thead>
<tr>
<th>Chip Version</th>
<th>Tech &amp; Voltage</th>
<th>Max Core Speed</th>
<th>Die Size (pad limited)</th>
<th>No. of pins</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDRAM 1P, 2xAGP</td>
<td>0.35um, 3.3V</td>
<td>100 MHz</td>
<td>107 mm²</td>
<td>492</td>
</tr>
<tr>
<td>SDRAM, 2P, 2xAGP</td>
<td>0.35um, 3.3V</td>
<td>100 MHz</td>
<td>130 mm²</td>
<td>656</td>
</tr>
<tr>
<td>DDR, 1P, 4xAGP</td>
<td>0.25um, 2.5V</td>
<td>133 MHz</td>
<td>133 mm²</td>
<td>553</td>
</tr>
<tr>
<td>DDR, 2P, 4xAGP</td>
<td>0.25um, 2.5V</td>
<td>133 MHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDRAM, 1P, 4xAGP</td>
<td>0.25um, 2.5V</td>
<td>133 MHz</td>
<td>107 mm²</td>
<td>492</td>
</tr>
</tbody>
</table>
Cache Coherency I

Request: I would like data for cacheline 0x001CA980
Cache Coherency II

Snoop Request: Do you have cacheline 0x001CA980?

Memory Fetch: Give me data for 0x001CA980.
Cache Coherency IIIa

Snoop Response: No
SDRAM MCT: RAS to rank 2, bank 0, row 0x00842
SDRAM MCT: CAS to rank 2, bank 0, col 0x0C3
SDRAM MCT: Here’s the data.
Snoop Response: Yes, I have this cache line
SDRAM MCT: RAS to rank 2, bank 0, row 0x00842
SDRAM MCT: CAS to rank 2, bank 0, col 0x0C3
MRO: Here’s the data.
Why worry about CC? Part 1

What if distance to DRAM is shorter than distance to cache (in another CPU)?
Why worry about CC? Part 2

Data returned by system to CPU is assumed to be the “most current copy”. Read transaction latency is bound by max(snoop_latency, dram_latency)

Intel P6 system bus read transaction latency breakdown

Processors can grab request address off of shared bus in shared multi-drop topology

System controller rebroadcast request address to aid in snoop for point-to-point topology
Multiple Clock Domains I

Most clock domains are integer multiples of each other.
Multiple Clock Domains II

What if clock domains are not integer multiples of each other?
Multiple Clock Domains III

- FastClk (100MHz)
- SlowClk (66MHz)
- Data
- LatchEn
- Latched Data
- CtlMask

Vertical alignment of signals:
- D0
- D1
- D2

Horizontal alignment:
- Data signals are synchronized with LatchEn.
- Latched Data signals are delayed by CtlMask.

Diagram components:
- Cmb Logic
- Dff

Clock domains:
- Slow clock domain
- Fast clock domain

Chipset logic:
- AMD Athlon
- Gearbox Logic

University of Maryland
ECE Dept.
Multiple Clock Domains IV

Data transfer from 100 MHz clock domain to 133 MHz clock domain (Latency Optimal)

Data transfer from 100 MHz clock domain to 133 MHz clock domain (Bandwidth Optimal)
Multiple Clock Domains V

Data transfer from 400 MHz clock domain to 800 MHz clock domain (Latency Optimal)

Data transfer from 400 MHz clock domain to 800 MHz clock domain (Bandwidth Optimal)
Multiple Clock Domains VI

Processor to Processor Bus Interface

Fractional multipliers could impact performance, but we may not have a choice
Multiple Clock Domains VII

AMD Athlon SPEC CPU FP Completion Time

- Non-Harmonic Node
- Harmonic Node

Harmonic Node:
- 1733 MHz
- 1600 MHz
- 1533 MHz
- 1466 MHz

Non-Harmonic Node:
- 1666 MHz

Total Task Time (seconds) vs. Cycle Time (ns)
AMD Opteron

1. Opteron Processor Core
2. L2 Cache
3. L2 Request
4. L2 Tag
5. L2 Data
6. Addr to NB
7. Clock Boundary
8. SRQ
9. GART/Addr Decode
10. Memory Controller
11. Req to DRAM Pins
12. DRAM Access
13. Data to MCT
14. NB Route
15. Clock Boundary
16. CPU route/mux/ECC
17. Write L1 D$ & FWD

"memory latency"

NB = northbridge
Summary

- System Controller is a “traffic cop”
- Traffic cop may have to deal with clock domain synchronization issue
- Handles Cache Coherency for small scale SMP configuration
- “Memory Latency” depends on lots of little things, not just speed of DRAM.