The Directory-Based Cache Coherence Protocol for the DASH Multiprocessor

Lenoski, Laudon, Gharachorloo, Gupta, Hennessy
ISCA 1990

DASH Priorities
- Correctness (issues?)
- Performance
  - Latency
  - Bandwidth
- Distributed Control and Complexity

Overview
- Distributed directory
- Illinois protocol???
- Off-the-shelf SMP
- Split transaction
  - Def?
  - How support?
Directory Controller Board

Memory Consistency
- Stay tuned…
- Release Consistency – one of several proposed weak consistency models.
- Requires explicit fence operations.

DASH Coherence
- Local, home, remote nodes
- Invalidation-based
- Nodes are clusters, not processors.
- Memory states
  - Uncached
  - Shared-remote
  - Dirty-remote

DASH Coherence
- Processor Cache states
  - D, S, I
- Read
  - Local L2, cache-cache, RAC, home node, …
  - At home node, causes read or forward to remote for read
- RAC – looks like another processor cache on the local bus
- Read of dirty line held in local cluster????
- Three hop protocol
  - Implications?
  - What happens with two reads close together (of remotely held data?)
DASH Read Latency

Read-Exclusive
- Local L2, cache-cache, RAC, home node
- Home node sends invalidations
- May need to forward

Other protocol features
- Queued locks
- Read prefetch, read-excl prefetch
- Explicit write-update

Bit-vector directory
- Not scalable?
- Other Solutions?
Performance

Details

<table>
<thead>
<tr>
<th>Execution Attribute</th>
<th>1 PE</th>
<th>2 PE</th>
<th>4 PE</th>
<th>8 PE</th>
<th>16 PE</th>
<th>24 PE</th>
<th>32 PE</th>
<th>40 PE</th>
<th>48 PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speedup (times)</td>
<td>1.00</td>
<td>1.60</td>
<td>2.48</td>
<td>3.33</td>
<td>4.12</td>
<td>4.93</td>
<td>5.85</td>
<td>6.00</td>
<td>6.42</td>
</tr>
<tr>
<td>Busy Pitfalls between Proc. Stalls</td>
<td>82.5</td>
<td>46.1</td>
<td>41.6</td>
<td>44.2</td>
<td>38.6</td>
<td>40.7</td>
<td>56.0</td>
<td>58.0</td>
<td>61.5</td>
</tr>
<tr>
<td>Task Processor Utilization (%)</td>
<td>15.8</td>
<td>19.3</td>
<td>21.0</td>
<td>28.8</td>
<td>24.8</td>
<td>21.2</td>
<td>20.5</td>
<td>26.2</td>
<td>20.4</td>
</tr>
<tr>
<td>Cache Read (%)</td>
<td>98.3</td>
<td>72.9</td>
<td>67.3</td>
<td>57.7</td>
<td>56.2</td>
<td>54.8</td>
<td>51.8</td>
<td>52.2</td>
<td>53.4</td>
</tr>
<tr>
<td>Cache Read Exclusive (%)</td>
<td>0.0</td>
<td>26.0</td>
<td>32.3</td>
<td>42.0</td>
<td>42.7</td>
<td>45.5</td>
<td>47.3</td>
<td>46.6</td>
<td>45.2</td>
</tr>
<tr>
<td>Cache Lock (%)</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Cache Unlock (%)</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>References to Local Memory (%)</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Remote Ref's satisfied Locally (%)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Remote Ref's satisfied in Home (%)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Unloaded Remote Cache Fill (Percent)</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
<td>101.0</td>
</tr>
<tr>
<td>Measured Remote Cache Fill (Percent)</td>
<td>102.5</td>
<td>117.5</td>
<td>130.4</td>
<td>158.7</td>
<td>169.7</td>
<td>172.2</td>
<td>186.8</td>
<td>191.3</td>
<td>192.7</td>
</tr>
<tr>
<td>Bus Utilization (%)</td>
<td>16.8</td>
<td>42.9</td>
<td>78.5</td>
<td>77.3</td>
<td>78.6</td>
<td>75.6</td>
<td>71.0</td>
<td>69.6</td>
<td>67.6</td>
</tr>
<tr>
<td>Req. Net Rejection Util. (%)</td>
<td>0.5</td>
<td>0.4</td>
<td>0.9</td>
<td>0.7</td>
<td>15.0</td>
<td>30.2</td>
<td>32.5</td>
<td>33.1</td>
<td>32.5</td>
</tr>
<tr>
<td>Reply Net Rejection Util. (%)</td>
<td>0.4</td>
<td>0.5</td>
<td>0.9</td>
<td>13.7</td>
<td>15.6</td>
<td>30.1</td>
<td>29.6</td>
<td>29.3</td>
<td>28.7</td>
</tr>
</tbody>
</table>

Token Coherence: Decoupling Performance and Correctness

Martin, Hill, Wood
ISCA 2003
Figure 1. (a) 16-processor meshed tree interconnect and (b) 16-processor (4x4) bi-directional torus interconnect. The boxes marked "P" represent highly aggregated nodes that include processor, caches, memory controllers, and other necessary controllers. The meshed broadcast tree uses discrete switches, while the torus is a directly connected interconnect. In this example, the torus has lower latency (two vs. four chip crossings on average) and does not require any glue chips. However, unlike the indirect tree, the torus provides no sequential order and is unsuitable for traditional coherency.

Figure 2. Example Race. A request for shared (ReqP) racing with a request for modified (ReqM).

Figure 3. Correctness substrate state transitions for the (a) processor, (b) memory, and (c) persistent request arbiter. As a simplification, the figure shows only tokens and with data. The symbol represents the current token count, and the "P" represents all the tokens. Solid arcs are transitions in response to incoming messages. Dashed arcs are transitions a persistent request arbiter (Section 3) can invoke at any time (e.g., when receiving a transmitted request). The "P" state occurs when a node receives a another processor's persistent request from the arbiter. Each processor must also remember its own persistent request, not explicitly shown in this figure. The initial states are emphasized with thick borders.

Table 2. Overhead due to released requests

<table>
<thead>
<tr>
<th>Workload</th>
<th>Non-Related</th>
<th>Related 0</th>
<th>Related 1</th>
<th>Persistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache</td>
<td>95.75%</td>
<td>2.23%</td>
<td>0.11%</td>
<td>0.29%</td>
</tr>
<tr>
<td>OLTP</td>
<td>97.57%</td>
<td>1.70%</td>
<td>0.43%</td>
<td>0.23%</td>
</tr>
<tr>
<td>SPECweb</td>
<td>97.64%</td>
<td>5.07%</td>
<td>0.30%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Average</td>
<td>96.57%</td>
<td>2.38%</td>
<td>0.48%</td>
<td>0.19%</td>
</tr>
</tbody>
</table>