Memory Performance Features of the 64-bit PA-8000

Bob Naas
Hewlett-Packard Co.
Ft. Collins, CO
ren@fc.hp.com

PA-8000 - Overview

• Leadership Performance
  > 360 SPECint92
  > 550 SPECfp92
  > 700 TPS (est.) Uniprocessor

• 56 Entry Instruction Reorder Buffer
• 8 Computational Units
• 2 Load / Store Units
• 4 Instructions Executed / Cycle
PA-8000 - Memory Performance

Three Major Design Considerations

- How to Maximize Memory Performance When Data is in Cache
- How to Maximize Memory Performance When Data is Not in Cache
- How to Maximize Performance Scaling in MP Systems
PA-8000 - Memory Subsystem

PA-8000 - Data Cache

- Large, Single-Level Off-Chip Cache
- 2 Cycle Latency
- Industry Standard Synchronous SRAMs
- Two Way Interleaving of Odd / Even Banks
- Dual Copies of Tags
PA-8000 - Data Cache Detail

PA-8000 - Data Cache: Opportunities for Parallelism

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Cache Activity Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load - Hit</td>
<td>1 Tag Read + 1 Data Read</td>
</tr>
</tbody>
</table>
| Store - Hit - Non-Dirty | 1 Tag Read +  
|                      | 2 Tag Writes + 1 Data Write                |
| Store - Hit - Dirty  | 1 Tag Read +  
|                      | 1 Data Write                               |
| Coherency Check - Miss | 1 Tag Read                                      |
PA-8000 - Data Cache
Access Sequence Example

Read Cycles:
- 3 Data Loads to A, B, C
- 1 Coherency Tag Check to D
- 2 Store Tag Checks to E,F

Write Cycles:
- 5 Data Writes to G,H,I,J,K
- 2 Dirty Bit Tag Updates to G,K

<table>
<thead>
<tr>
<th>CYCLE</th>
<th>TAG</th>
<th>DATA</th>
<th>DATA</th>
<th>TAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>STORE E</td>
<td>STORE G</td>
<td>STORE H</td>
<td>STORE G</td>
</tr>
<tr>
<td></td>
<td>(Tag Check)</td>
<td></td>
<td>(Tag Update)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>STORE F</td>
<td>STORE I</td>
<td>STORE J</td>
<td>STORE K</td>
</tr>
<tr>
<td></td>
<td>(Tag Check)</td>
<td></td>
<td>(Tag Update)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LOAD A</td>
<td>LOAD A</td>
<td>STORE K</td>
<td>COHERENCY D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Tag Check)</td>
</tr>
<tr>
<td>3</td>
<td>LOAD B</td>
<td>LOAD B</td>
<td>LOAD C</td>
<td>LOAD C</td>
</tr>
</tbody>
</table>

[Diagram showing data cache bus activity with even and odd bank activities]

PA-8000 Memory Subsystem:
Cache Misses to Main Memory

- Out Of Order Execution Hides Latency
- Load to GR0 Prefetch
- Efficient Path to Runway Bus
PA-8000 - Cache Miss Path

PA-8000 - Memory Subsystem: Scalable Multiprocessing

- MP "Knowledge" Kept Out Of Data Cache Subsystem
- Minimal Overhead For Processing Coherency Checks
Summary

- PA-8000 Memory Subsystem Provides Superb Memory Performance

  - Aggressive Data Cache:
    Flexibility + Parallelism → Bandwidth

  - Optimized Accesses to Main Memory:
    Parallelism + Low Latency

  - Scalable MP:
    Simple, Low Overhead Coherency