“An Evaluation of Directory Schemes for Cache Coherence”
Presented by Scott Weber

Outline
- Motivation and goals for directory schemes
- Directory schemes
- Schemes evaluated
  - Directory-based
  - Snoopy
- Insights from the evaluation
- Directory scheme alternatives
- Conclusions and Retrospective

Motivation and Goals
- Snooping does not scale past ~20 processors
  - Protocol depends on low-latency broadcasts
- Snooping interferes with the processor-cache connection
- Avoid broadcast nature of snooping
- Directory-based protocols should be competitive with snoopy protocols
- Access to a directory cannot be a bottleneck

Directory Schemes
- Tang scheme (DirₙNB)
  - Multiple clean blocks, one dirty block
  - Copy of tags, dirty bits for each cache in directory
  
  **Read miss**
  - check directory,
  - if dirty then write dirty back,
  - supply the data,
  - update directory.

  **Write miss**
  - check directory,
  - if dirty then flush dirty back,
  - invalidate clean copies,
  - perform the write,
  - update directory.

  **Write hit (dirty)**
  - if dirty bit set then write.

  **Write hit (clean)**
  - if dirty bit not set,
  - notify directory,
  - invalidate clean copies,
  - update directory, update dirty bit.
## Directory Schemes

- **Modifications to Tang’s scheme**
  - Censier and Feautrier (DirₙNB)
    - Vector of valid bits for each cache and dirty bit
    - Use the address of the data to access directory
  - Yen and Fu (DirₙNB) refines C & F
    - Single bit in each cache to indicate only copy
    - When set, do not have to access directory
    - Requires more bandwidth to update single bits

- **Archibald and Baer (Dir₀B)**
  - Four states:
    - block not cached
    - block clean in exactly one cache
    - block clean in an unknown number of caches
    - block dirty in exactly one cache
  - Requires broadcasts to do invalidations and write backs
  - Organization is still centralized
  - Easy to add more caches to the systems

## Schemes Evaluated

- **Classification**
  - Dir(cache pointers)[Broadcast|No Broadcast]
- **Dir₁NB** – Tang (with n = 1) and variants
- **Dir₀B** – Archibald and Baer
- **Alternatives attempt based on results**
  - Dir₀NB, DirₙNB, Dir₁B, Dir₀B
- **Write-Through-With-Invalidation (WTI)**
- **Dragon Update Protocol**

## Evaluation Methodology

- **Trace-driven simulation**
- **Interested in memory traffic for CC (use ∞ cache)**
- **Machine independent metric**
  - Communication cost/memory reference
- **Assume bus for comparison**
- **Measure event frequencies for various types of memory accesses (differ for each protocol)**
- **Weight event frequencies in terms of bus cycles**
  - Non-pipelined shared bus model
  - Pipelined split address/data bus model
### Evaluation of the Protocols

- Dir\(_1\)NB has a high read miss rate (5.18%)
  - Caused by read sharing among processes
  - Limitation of data only being in one cache
  - Dir\(_0\)B has a low read miss rate (0.62%)
- Dir\(_0\)B and WTI have same rates since they have same state changes on data in cache
- Dragon is dominated by write hits (updates)
- 36% of misses are coherency-related misses

### Evaluation of the Protocols

- >85% writes to previously clean blocks cause invalidations in 0 or 1 caches
  - Motivates Dir\(_n\)NB, Dir\(_n\)NB, Dir\(_1\)B, Dir\(_r\)B
- Directory bandwidth similar to memory
  - Can distribute directory and memory to scale
- Estimating average memory access makes protocol bus cycles more equal
- Spin-locks on shared variables hurt Dir\(_1\)NB

### Directory Scheme Alternatives

- Schemes introduced to decrease broadcasts
  - Dir\(_n\)NB Performs sequential invalidations
  - Dir\(_1\)B performs a single invalidation (common case) if broadcast bit is clear, otherwise broadcast
  - Dir\(_r\)NB and Dir\(_r\)B use limited number of ptrs
  - Limited broadcasts invalidate to cache subsets
    - 01XX01 encoding indicate subsets
- Schemes like these scale since new directory bits do not necessarily have to be added when scaling

### Conclusions

- Bandwidth to directory is similar to bandwidth to memory
  - Distribute the directory and memory
  - Allows to scale with the number of processors
- Eliminates the inefficiency of broadcasts
  - Most blocks shared by 0 or 1 caches
  - Only need a few pointers in each directory entry
- Snoopy and broadcast protocols are competitive
  - Need to keep the number of spin-locks to a minimum
Retrospective

- Paper led to the development of DASH (Dir$_n$NB) prototype
- Concern at paper time was if snoopy and directory-based protocols were competitive
- Real issues
  - Scalability of coherence scheme
  - Implementation complexity
  - Overhead of coherence protocol
  - Performance with many processors
  - Implementing distributed directory coherence

Event Frequencies

<table>
<thead>
<tr>
<th>Event</th>
<th>Dir$_n$NB</th>
<th>W11</th>
<th>Dir$_n$B</th>
<th>Dragon</th>
</tr>
</thead>
<tbody>
<tr>
<td>instr</td>
<td>49.72</td>
<td>49.72</td>
<td>49.72</td>
<td>49.72</td>
</tr>
<tr>
<td>read</td>
<td>39.82</td>
<td>39.82</td>
<td>39.82</td>
<td>39.82</td>
</tr>
<tr>
<td>rd-hit</td>
<td>34.32</td>
<td>38.88</td>
<td>38.88</td>
<td>39.20</td>
</tr>
<tr>
<td>rd-miss(rm)</td>
<td>5.18</td>
<td>0.62</td>
<td>0.62</td>
<td>0.30</td>
</tr>
<tr>
<td>rm-blk-cln</td>
<td>4.78</td>
<td>-</td>
<td>0.23</td>
<td>0.14</td>
</tr>
<tr>
<td>rm-blk-dirty</td>
<td>0.40</td>
<td>-</td>
<td>0.40</td>
<td>0.17</td>
</tr>
<tr>
<td>rm-first-ref</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>write</td>
<td>10.46</td>
<td>10.46</td>
<td>10.46</td>
<td>10.46</td>
</tr>
<tr>
<td>wrt-hit(w)</td>
<td>10.19</td>
<td>10.25</td>
<td>10.25</td>
<td>10.36</td>
</tr>
<tr>
<td>wh-blk-cln</td>
<td>-</td>
<td>-</td>
<td>0.41</td>
<td>-</td>
</tr>
<tr>
<td>wh-blk-dirty</td>
<td>-</td>
<td>-</td>
<td>9.84</td>
<td>-</td>
</tr>
<tr>
<td>wh-distrib</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.74</td>
</tr>
<tr>
<td>wh-local</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.82</td>
</tr>
<tr>
<td>wrs-miss(wm)</td>
<td>0.17</td>
<td>0.12</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>wrm-blk-cln</td>
<td>0.08</td>
<td>-</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>wrm-blk-dirty</td>
<td>0.09</td>
<td>-</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>wrm-first-ref</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Bus Cycle Breakdown