Caches and programming

Alexandru Duțu

based on Prof. Lebeck slides
Caches

Memory address

<table>
<thead>
<tr>
<th>tag</th>
<th>index</th>
<th>offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct mapped caches

- frames
- \# frames \(\Rightarrow\) \# index bits
- 1 frame holds 1 block
- block or cache line
- block size \(\Rightarrow\) \# offset bits

Friday, November 18, 11
Caches

Memory address

<table>
<thead>
<tr>
<th>tag</th>
<th>index</th>
<th>offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.....</td>
<td>256</td>
</tr>
<tr>
<td>1</td>
<td>.....</td>
<td>257</td>
</tr>
<tr>
<td>2</td>
<td>.....</td>
<td>258</td>
</tr>
<tr>
<td>253</td>
<td>.....</td>
<td>510</td>
</tr>
<tr>
<td>254</td>
<td>.....</td>
<td>511</td>
</tr>
<tr>
<td>255</td>
<td>.....</td>
<td>512</td>
</tr>
</tbody>
</table>

Set associative caches

- set - multiple frames
- # sets => # index bits
- 1 frame holds 1 block
- block or cache line
- block size => # offset bits

Friday, November 18, 11
Misses

- Compulsory - first time the address is accessed
- Capacity - cache is too small
- Conflict - cache entry maps multiple memory addresses
- Invalidation - external sources update memory (e.g. DMA controller)
Take-aways

• capacity - decreases capacity misses
• associativity - decreases conflict misses
• block size - decreases compulsory/capacity misses
Caches and programming

- Caches optimize the locality available in your code
- Locality
  - spatial - access stride
  - temporal - accessed dataset size
Exercise

6.45 Download the *mountain* program from textbook’s website and run it on any teer system. Use the results to estimate the size of the cache.
Memory mountain

Spatial locality

Memory

Temporal locality

L1

L2

L3
Memory mountain

![3D graph showing memory throughput versus size and strides for Core i7-2](image)
Conclusions

- Spatial locality - decreases as access stride increases
- Temporal locality - decreases as data set size increases