Trace Cache: A Low Latency Approach To High Bandwidth Instruction Fetch

Eric Rotenberg, Steve Bennet, James E. Smith
Context

- Modern processor need a lot of fetch bandwidth
  - Pipelines are deep
  - On a cache miss, we need to refill it as quickly as possible.
  - Issue widths are getting larger than basic blocks
- Basic block alignment, poor locality, and poor branch predictor performance limit bandwidth

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>taken %</th>
<th>avg basic block size</th>
<th># instr between taken branches</th>
</tr>
</thead>
<tbody>
<tr>
<td>eqntott</td>
<td>86.2%</td>
<td>4.20</td>
<td>4.87</td>
</tr>
<tr>
<td>espresso</td>
<td>63.8%</td>
<td>4.24</td>
<td>6.65</td>
</tr>
<tr>
<td>xlisp</td>
<td>64.7%</td>
<td>4.34</td>
<td>6.70</td>
</tr>
<tr>
<td>gcc</td>
<td>67.6%</td>
<td>4.65</td>
<td>6.88</td>
</tr>
<tr>
<td>sc</td>
<td>70.2%</td>
<td>4.71</td>
<td>6.71</td>
</tr>
<tr>
<td>compress</td>
<td>60.9%</td>
<td>5.39</td>
<td>8.85</td>
</tr>
</tbody>
</table>

Table 1. Branch and basic block statistics.
Key Idea

• Build a specialized cache that avoids these problems (to the extent possible) and fetch from there most of the time.
The Trace Cache

Figure 2. High level view of the trace cache.
More Detail

Figure 4. The trace cache fetch mechanism.
Figure 8. IPC results (fetch latency = 1 cycle).
Figure 10. Improvement over SEQ.3.
Alternatives to Trace Caches

- Align branch targets
  - Eliminates the problem of branching into the middle of a line.
- Aggressive branch prediction
- Next-line predictor -- Extra state in L1 I-Cache gives the likely next line that will be needed.
- A specialized form of BP. Builds a “linked list” of cache lines -- effectively a trace.
Trace Cache Success

- Trace Caches found success in x86 processors starting with Pentium Pro
- In the fetch stage, these processors broke x86 CISC instructions down into RISC-like “micro-ops”
- The micro-ops were stored in the trace cache.
- This optimization eliminated the micro-op conversion stages from the pipeline in the common case.