Embedded Systems

Cache Memory Organizations and Replacement Policy Examples

Element (Word) Hit ratio

- Assume two adjacent levels, numbered \( M_1 \) and \( M_2 \), in a memory hierarchy
- The word hit ratio, \( H \), for the lower level in the hierarchy is the number of address references that are found in the lower level \( (N_1) \) divided by the total number of references \( (N_1+N_2) \)

\[
H = \frac{N_1}{N_1 + N_2}
\]

- These references can be to instructions or data
Block Address Reference Streams

- The performance of a replacement policy in a given memory organization can be analyzed using a block reference stream generated by a set of computations.
- The block reference stream is a list of block addresses that are referenced in a particular program execution.
- For example:

```
2 3 2 1 5 2 4 5 3 2 5 2
```

would imply that the first block referenced is B₂, the second is B₃, etc.
- There may be many consecutive word references within a single block reference.

Block Hit Ratio

- Let $N_1^*$ and $N_2^*$ represent the number of block references to $M_1$ and $M_2$ respectively.
- For our purposes $M_1$ is a cache memory and $M_2$ is a main memory.
- The block hit ratio $H^*$ is defined as

$$H^* = \frac{N_1^*}{N_1^* + N_2^*}$$

- This is analogous to the word hit ratio $H$.
- Let $n^*$ represent the average number of consecutive word address references within each block.
- $H$ can be estimated from $H^*$ as

$$H = 1 - \frac{1 - H^*}{n^*}$$
**Comparison of Replacement Policies**

- Consider a memory hierarchy with a fully associative cache with a capacity of three blocks.
- Assume the execution of a program $Q$ requires access to five distinct main memory blocks $B_i$, where $i=1,2,3,4,5$ and $i$ is the block address.
- A block address stream formed by executing $Q$ is $232152453252$.
- What are the actions of the FIFO, LRU, and OPT replacement policies given this address stream?

---

**Example FIFO Replacement**

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Address Stream</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2*</th>
<th>2*</th>
<th>2*</th>
<th>2*</th>
<th>5</th>
<th>5</th>
<th>5*</th>
<th>5*</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4*</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- * represents a block to be replaced if necessary.
- $H^*=3/12=0.25$.
- If $n^*=100$ then $H=.9925$ (99.25%).
### Example LRU Replacement

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Address Stream</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2*</th>
<th>2*</th>
<th>2</th>
<th>2</th>
<th>2*</th>
<th>2</th>
<th>2</th>
<th>2*</th>
<th>3</th>
<th>3</th>
<th>3*</th>
<th>3*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>1</td>
<td>1</td>
<td>1*</td>
<td>4</td>
<td>4</td>
<td>4*</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- * represents a block to be replaced if necessary
- \( H^* = \frac{5}{12} = 0.416 \)
- If \( n^* = 100 \) then \( H = 0.9941 \) (99.41%)

### Example OPT Replacement

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Address Stream</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2*</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2*</th>
<th>2</th>
<th>2</th>
<th>2*</th>
<th>4*</th>
<th>4*</th>
<th>4*</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>1*</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

- * represents a block to be replaced if necessary
- \( H^* = \frac{6}{12} = 0.5 \)
- If \( n^* = 100 \) then \( H = 0.995 \) (99.5%)
Other Considerations for Block Replacement

• If the cache is direct mapped there is only one specific cache block into which a main memory block can be mapped
  – The block hit ratio for a direct mapped cache cannot be greater than that for a fully associative cache given a particular block address reference stream

• If the cache is set associative there are only a few cache blocks into which a main memory block can be mapped
  – The block hit ratio for a set associative cache cannot be greater than that for a fully associative cache given a particular block address reference stream