1. Given a processor with 3-level hierarchical cache, please fill the following table and compute the access time.

There are some notes:

1. We use direct-mapped cache here, and the number of offset bits including byte offset

(that is, if we have 4 bits of word offset, we should fill 4+2=6 in the “Offset Bits” blank).

1. The access time of each level (from L1 to L3) is 2ns, 5ns, 25ns, respectively.

And the memory access time is 100ns (Suppose that all data has been loaded into memory.)

1. Then according to the statics, the local miss rate of each level (from L1 to L3) is 2%, 5%, 20%, respectively.
2. average memory access time (AMAT) = Hit Time + Miss Rate x Miss Penalty

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Address Bits | Cache Size | Block Size | Tag Bits | Index Bits | Offset Bits |
| L1 cache | 32 | 8KB | 2 words |  |  |  |
| L2 cache | 32 | 32KB | 4 words |  |  |  |
| L3 cache | 32 | 512KB |  |  |  | 5 |

1. Please compute average memory access time (AMAT).

(Since processor can access cache and memory directly, we don’t replace the block in the upper level. That is we only consider the access time.)

Ans

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Address Bits | Cache Size | Block Size | Tag Bits | Index Bits | Offset Bits |
| L1 cache | 32 | 8KB | 2 words | 19 | 10 | 2 |
| L2 cache | 32 | 32KB | 4 words | 17 | 11 | 4 |
| L3 cache | 32 | 512KB | 8 words | 13 | 14 | 5 |

average memory access time

= 2 + 0.02x(L1 miss penalty)

= 2 + 0.02x(5 + 0.05x(L2 miss penalty))

= 2 + 0.02x(5 + 0.05x(25 + 0.2x(L3 miss penalty)))

= 2 + 0.02x(5 + 0.05x(25 + 0.2x100)

= 2 + 0.02x5 + 0.02x0.05x25 + 0.02x0.05x0.2x100

= 2 + 0.1 + 0.025 + 0.02 = 2.145ns

1. Write down the questions for hierarchy design and give a brief explanation of their answers.
2. Where can a block be placed in the upper level?

block placement; Because the capacity in upper level is smaller, and the one in lower level is larger. If you want to move blocks to upper level from lower level, where you want to move it, that’s related to block placement.

1. How to access elements?

By block finding; if you know how to place the blocks, and you will know how to find it.

1. Which block should be replaced on a miss?

block replacement; Because the capacity of upper level is small, which blocks should be written back to lower level? It has something to do with block replacement.

1. What happens on a write?  
   write strategy; it depends on how you design, so the answer is write strategy.