SOLUTION NOTES

Computer Design 2003 Paper 5 Question 2 (IAP)

(*a*) Name and describe three reasons for a cache miss. [6 marks]

Compulsory: memory location has never been accessed before.

Capacity: memory location has been in the cache before, but would have been displaced (even if the cash was fully associative).

Conflict: memory location would have still been present if the cache was fully associative, but was actually displaced due to aliasing of multiple locations to the same line.

(*b*) For each reason, suggest a technique for reducing the number of misses. [6 marks]

Compulsory: larger cache line size causes fewer compulsory misses due to spatial locality.

Capacity: larger cache. Also, reducing line size may help by providing more flexibility. Modifying s/w to reduce cache footprint will also help.

Conflict: add more associativity. Alternatively, modify software to try to reduce aliasing through careful memory placement.

(*c*) Why might it be advantageous to use a set-associative cache instead of a fully associative one? [4 marks]

Fully associative caches tend to be slower than set-associative ones due to the need to compare the target address against a large list of potential matches. Thought this can be mostly be done in parallel, the substantial extra logic can slow things down (e.g. combining the result from all the comparators), and reduces the storage capacity of the cache for a given die area.

(*d*) Describe two techniques for reducing miss penalty. [4 marks]

Adding another (larger and typically slightly slower) level of caching. Using burst-mode reads from DRAMS to rapidly return the line, possibly using critical-word-first fetch order.