2008 Paper 7 Question 5

Comparative Architectures

(*a*) Why are multi-level caches often used in preference to a single larger cache? How might the parameters of an L1 and L2 data cache typically differ? [8 marks]

(*b*) What does it mean if a cache memory hierarchy adopts a multi-level inclusion policy? What might influence a decision on whether to adopt a multi-level inclusion or exclusion policy? [6 marks]

(*c*) A processor's multi-level cache hierarchy consists of L1 and L2 caches with the following characteristics: L1 miss rate is 2%, L1 hit time is 2 cycles, local L2 miss rate is 20%, L2 hit time is 10 cycles, L2 miss penalty is 200 cycles. It is suggested that reducing the size of the L1 cache will improve overall performance by reducing the L1 cache's hit time to a single cycle. The reduction in L1 cache size will increase the L1's miss rate to 3%. Will average memory access time actually be improved? Clearly state any formulae you use and show your calculations. [6 marks]