COMPUTER SCIENCE TRIPOS Part II – 2017 – Paper 8

3 Comparative Architectures (RDM)

(*a*) What features of a bus does a snoopy cache coherence protocol depend on? [5 marks]

(*b*) It is now possible to integrate general-purpose processor cores and a GPU on a single die and for them to share the same address space. Why might supporting hardware coherence between multiple general-purpose cores and a GPU be problematic? [5 marks]

(*c*) Consider a chip multiprocessor constructed from identical building blocks or tiles. Each tile contains a processor and private L1 caches, an on-chip network router and a slice (or fraction) of the shared L2 cache. Addresses are interleaved across these L2 slices at the granularity of cache lines. Coherence is maintained in the shared L2 by adding directory bits to each L2 line to track sharers (i.e. those L1 caches that hold a copy of the same block).

(*i*) What would be the advantages and disadvantages of adopting private L2 caches rather than a single shared L2 cache? [5 marks]

(*ii*) It is suggested that the performance of the shared cache could be improved by also keeping copies of evicted L1 cache lines in the tile's local L2 slice. Copies are made when L1 cache lines are evicted because of conflict or capacity misses. What extra actions are now required on a L1 cache miss and when a tile receives an invalidation request? [5 marks]