

6 Comparative Architectures (RDM)

A large Last Level Cache (LLC) is necessary to achieve good performance in many applications. Recent server class processors have included LLCs with capacities of 40 MBytes or more. Large caches such as this are constructed from numerous smaller SRAM banks.

- (a) Describe an appropriate on-chip network to interconnect 32 SRAM banks to create a large LLC. The delay to access a bank should increase as we move further away from the cache controller and bus interface. The SRAM banks are square and the time taken for a signal to travel along the edge of a SRAM bank is much less than your network's clock cycle time. [5 marks]
- (b) To implement a set-associative LLC we may spread each set across multiple banks, i.e. each "way" of the set will be in a different bank. The different associative ways will have different access latencies depending on their distance from the cache controller. How might we optimise the placement of lines in particular banks (or ways) to minimise the cache's average access latency? Remember to consider the cost of moving lines. [6 marks]
- (c) How might the SRAM banks be efficiently interconnected so that the cache's access time is constant regardless of which bank is accessed? [4 marks]
- (d) Why might it be advantageous to be able to manage the amount of LLC used by each co-scheduled thread in a chip multiprocessor? [5 marks]