SOLUTION NOTES

Computer Design 2002 Paper 5 Question 2 (SWM)

(a) I would expect a diagram indicating that first the upper parts of the address and control signals are setup, together with a row select signal (RAS). Then the lower bits of the address are transmitted and the column address signal (CAS) is set. The processor is then told to wait until the memory starts to return data. At this point the processor cycles through the lower address bits to specify the ordering of the burst read.

(b) A direct mapped cache locates data using a simple hash function of the address, typically a small range of address bits. On the other hand, an associative cache performs an associative lookup on the address in order to identify where the data is stored. Thus, a direct mapped cache can only store data for a particular address in one place in the cache where as an associate cache can store data anywhere in the cache.

(c) The cache line replacement policy for a direct mapped cache is trivial since data may only be stored in one place in the cache. However, for an associate cache there is a choice as to which cache line is flushed to make room for another. In this instance a number of policies are possible, for example: random, least recently used and not last used.

(d) Caches store data blocks called cache lines. A cache line fill (read) or spill (write) results in a burst transaction which may be exploited when accessing DRAM.

(e) Bus snooping is typically performed by a cache. The bus to the slower memory is snooped (monitored) by the cache control logic in order to identify write transactions which should be reflected in the contents of the cache. This is often used in small scale parallel processing machines to present a coherent shared memory model to the programmer.