SOLUTION NOTES FOR EXAMINERS AND SUPERVISORS ONLY

Computer Design 2000 Paper 3 Question 6 (SWM)

(a) Gordon Moore's law is actually an observation made in 1965 that memory chip capacity doubles every year. In the late 1960's the law was revised to match the trend at the time: memory chip capacity doubles every 18 months. More recently the law was applied to the improvement in processor performance.

(b) Memory chips are planar devices. Their surface area has changed little in the last 30 years since increased device density has been used to increase the amount of storage rather than reduce the physical size of the chips.

A simple DRAM has row select wires running horizontally and vertical column wires both of which run across the chip. Accessing a row of the memory is initiated by a row select wire being enabled, which results in one row of data being sent down the column wires. In the worst case the row select signal followed by the column data signal has to travel nearly both sides of the memory chip. Since memory area has changed little, these row and column signals still travel the same distance. In practice it is very difficult to improve signal propagation time which is ultimately limited by the speed of light. Thus, memory access latency has improved little.

(c) A hierarchy of caches are used to hide poor main memory access times. Caches are smaller, faster memories which store data which is likely to been needed in the near future with the hope that the processor can be supplied with data from a cache rather than slow main memory.

Caches exploit temporal and spatial locality of data and program memory accesses. Temporal locality is a simple prediction that if data has been accessed once then it is likely to be accessed again in the near future. Spatial locality predicts that if data has been accessed in one location then its neighbors are likely to be accessed.

Temporal and spatial locality predictors only work because many applications exhibit these forms of locality.