

Two hours

**UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE**

System Architecture

Date: Wednesday 30th May 2012

Time: 09:45 - 11:45

Please answer any THREE Questions from the FOUR questions provided

Use a SEPARATE answerbook for each SECTION

**For full marks your answers should be concise as well as accurate.
Marks will be awarded for reasoning and method as well as being correct.**

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are
not programmable and do not store text.

[PTO]

Section A

1. Caches

- a) Why does cache memory give rise to functional problems in computer systems that implement Direct Memory Access (DMA) facilities for peripheral devices? Give examples of two problems. (5 marks)
- b) Explain the meaning of a “Compulsory” cache miss and give an example of when it might arise. (3 marks)
- c) Explain one way of resolving the problems you describe in part (a) that would increase the number of Compulsory cache misses. (3 marks)
- d) Explain one way of resolving the problems you describe in part (a) that would not alter the number of cache misses. (3 marks)
- e) A CPU runs with a combined Instruction and Data cache with an access time of 1 clock cycle, and a main memory access time of 25 cycles. For a particular program, the cache hit rate is 95%. By considering the total time to access 100 memory locations, or otherwise, and ignoring write cycles, what is the average memory access time seen by the program? (3 marks)
- f) Give an expression for the average memory access time for part (e) if a second level cache with access time T cycles and hit rate $H\%$ is added to the system. (3 marks)

2. Storage

- a) Explain what “Seek time”, “Search time”, and “Transfer time” mean when used to describe hard disk operation. (3 x 2 marks)

A modern desktop disk drive is specified as having a capacity of 3 terabytes, a transfer rate of 90 megabytes per second, a rotation rate of 7200 revolutions per minute, and a mean seek time of 10 milliseconds.

- b) How long on average would a transfer of 4 kilobytes take from a random position on the disk? (3 marks)
- c) How long would it take to read the entire disk entirely sequentially? (3 marks)
- d) How long would it take to read the entire disk if all reads were of 4 kilobytes and reads were completely random? (4 marks)
- e) What are the implications of the results of parts (c) and (d) for the architecture and usage of RAID arrays? (4 marks)

Section B

3. Pipelining

- a) Explain how pipelining can reduce the execution time of a stream of instructions in a processor. (2 Marks)
- b) With the aid of a diagram explain the operation of each stage of a classic 5 stage pipeline. (4 Marks)
- c) Explain what data hazards and control hazards are and how they can be eliminated with NOP instructions. Use the sequence of instructions below to illustrate this for data hazards. (6 Marks)

```

MUL R2 R2 R2
LDR R1 X
ADD R1 R1 R2
MUL R4 R4 R4
LDR R3 X
ADD R3 R3 R4

```

(Assume arithmetic instructions are of the form Operation, Output, Input1, Input2)

- d) With the aid of a diagram explain how forwarding can be used to reduce or eliminate the need to insert NOP instructions for arithmetic operations and loads. Demonstrate this with the sequence of instructions in part c. (4 Marks)
- e) Using the set of instructions from part c (or part d if you have them) show how instruction reordering can be used to reduce or eliminate the need to insert NOP instructions. What are the advantages and disadvantage of doing reordering in a compiler vs in the chip? (4 Marks)

[PTO]

4. Multithreading

- a) What is “Multithreading” in the context of CPU design? (NB: NOT in the design of software!) What is “Multi-core”? (4 Marks)

What does each of the following terms mean:

- b) Course-grained multithreading? (2 Marks)
- c) Fine-grained multithreading? (2 Marks)
- d) Simultaneous multithreading? (2 Marks)
- e) What is cache coherency? Why is it important? Give an example of what may happen without it. (4 Marks)

A RISC CPU is running at a clock rate of 1GHz, and can issue one instruction per clock cycle, with no pipeline delays in the best case (100% cache hits; 100% branch predict success etc). There is only one level of data cache provided, and the CPU implements 2 threads, with a switch only on a cache miss.

- f) In a particular program, one instruction in 5 reads a memory location and there are no writes. If the main memory has an access time of 75 nanoseconds, what is the pipeline throughput with 0.1% data cache misses? (3 Marks)
- g) What is the pipeline throughput at a 10% data cache miss rate, why might this be worse in a real system? (3 Marks)

END OF EXAMINATION