Data Provided: None

The University Of Sheffield.

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

Autumn Semester 2006-2007 (2 hours)

Advanced Computer Architectures 4

Answer **THREE** questions. **No marks will be awarded for solutions to a fourth question.** Solutions will be considered in the order that they are presented in the answer book. Trial answers will be ignored if they are clearly crossed out. **The numbers given after each section of a question indicate the relative weighting of that section.**

1. a. i. Draw a basic schematic for a Banyan Network where *s*=2, *f*=3, and *l*=2. **(4)**

ii. What approaches can be used to route data inputs to output and what advantages/disadvantages are associated with each approach? **(2)**

iii. Draw a schematic showing how a switching network could be used in a tightly-coupled multiprocessor system to connect between the processors and memories. **(1)**

iv. Estimate the complexity of the Banyan network in **i.** relative to the correspondingly sized *cross-point switch*. **(4)**

b. A closely-coupled processor system consists of a number of processor modules connected by a cross-point switch to a number of memory modules. It can be shown that the probability, p_A, that a memory access generated by a processor proceeds without delay is:

$$p_A = \frac{M}{NRT_{acc}}\left(1-\left(1-\frac{RT_{acc}}{M}\right)^N\right)$$

Identify the meaning of the terms in this equation. **(1)**

A particular processing system is being designed with the following characteristics:

- There are eight identical processors connected via a cross-point switch to four identical memory modules.
- Each core processor performs $4x10^8$ memory accesses per second and have internal, write-back caches put between the core processors and their connection to the cross-point switch. Studies show that for three different possible line lengths, the caches that could be used have the properties shown in **Table 1**, where 'Dirty Lines' is the proportion of the lines in the cache that get written to, on average.

Line Length (words)	Hit Rate (%)	Dirty Lines (%)
16	97	15
32	99	10
64	98	8

Table 1

- When a line is transferred to/from memory, it is done as a single block transfer and the time taken to perform a block transfer across the connection network is 30+2*n ns, where n is the number of words transferred in a single block.

Identify the cache line length that should be chosen, for best performance – ensure that your answer is supported by an analysis. **(8)**

2. A pipelined system, as shown in **Figure 2**, consists of four, interconnected processing blocks which are all of approximately equal complexity and cost.

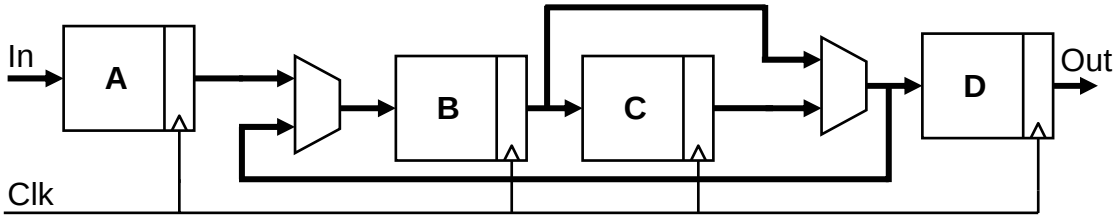

Figure 2

The algorithm applied to the sequence of input data is:

$$In \rightarrow A \rightarrow B \rightarrow C \rightarrow B \rightarrow B \rightarrow C \rightarrow B \rightarrow D \rightarrow Out$$

a. For this set of operations:

i. Draw the reservation table; **(4)**

ii. Determine the value of the collision vector; **(2)**

iii. Calculate the throughput of the pipeline in datum/clock cycle; **(2)**

iv. Calculate the utilisation of each of the processing blocks. **(2)**

b. How do you recognise that there are no simple *cosmetic* changes (e.g. only adding a register) that might be made to the design to improve performance? **(2)**

c. You recognise that there is a simple *functional* change to the architecture of the pipeline that you can make to improve performance.

i. What is this change? **(4)**

ii. How does this change affect throughput? **(2)**

iii. What is the *cost/benefit* associated with the change? **(2)**

3. a. Draw a schematic diagram for a 4-way set associative cache system that uses LRU as the replacement policy, identifying how it operates. **(6)**

b. The cache in part **a.** is 16Kwords in size and is organised as lines that are 16 words long. The memory address (for accessing words) issued by the processor is 32 bits long. Estimate the number of bits of memory used in the controller. **(2)**

c. A set-associative cache is controlled synchronously using a 1GHz clock and the basic access time for the memory used to construct the cache is one clock cycle. On a miss the memory access is posted to a reorder buffer (connected to the external memory) with *read-around-write* capability. The probability of a memory cycle being a memory read is 0.7, the hit rate of the cache is 0.85 and the penalty on a cache miss is 30ns (this figure subsumes *all* of the cost of replacing a block within the cache on a miss).

i. Estimate, simply, the *effective* access time of the cache. **(10)**

ii. Does it matter how big the reorder buffer is? **(2)**

4. a. i. What is *interleaving* in the context of a memory system and why does it help? **(2)**

ii. Distinguish between high-, low-, and mixed-order interleaving, explaining where they might be used and their advantages/disadvantages. **(2)**

iii. Can interleaving help the performance of a uni-processor system and, if so, how? **(2)**

iv. How can *home memory modules* be used to reduce the cost and/or traffic on a switching network in a multiprocessor? **(2)**

v. What problem might exist in a vector processor relating to interleaving (ensure that your answer describes *stride*)? **(2)**

vi. What is the difference between static and dynamic coherence in a multiprocessor system and what problem do the methods attempt to overcome? **(2)**

vii. What is *indivisibility* when applied to memory accessing? **(2)**

viii. What basic function within an operating system does indivisible memory accessing support? **(2)**

ix. Identify how a processor implements indivisible memory accessing in a uni-processor system. What basic part of a ‘normal’ processor’s behaviour leads to this requirement for indivisible memory accessing? **(2)**

x. Identify how this method must be enhanced if indivisible memory accessing is to be supported in a closely-coupled multi-processing system. **(2)**

NLS / NA