Examination for Advances in Computer Architecture

15th December 2008 10:00-13:00

Instructions

You are required to answer question 1 plus any 2 of the remaining 3 questions. You may not bring books or lecture notes into the examination and you have three hours to complete the examination. Each question carries an equal weight in the assessment. Questions parts are labeled with percentage marks for that part in square brackets e.g. [x%]. The percentage is of the final assessment, i.e. a total of 60% for the examination or 20% for each question answered.

Question 1.

(a) Memory is slow compared to processor speed; how can this difference in speed impact a processor's performance? Describe how poor performance can be avoided using the following techniques: (i) a cache hierarchy (ii) out-of-order instruction issue (iii) multi-threading. [4%].

(b) A pipeline has L stages and a clock cycle of t seconds for each stage. Derive formulae for the following parameters of the pipeline: the peak performance rate, $r_{infinity}$, the half-performance vector length, $n_{1/2}$, and a performance rate r(n) where n is the number of operations processed by the pipeline without bubbles. Explain why most computer systems may be characterised by these pipeline formulae [4%]

(c) Concurrent instruction issue can increase processor throughput and avoid dependencies on long latency operations. Describe two non-threaded instruction issue techniques (one static one dynamic) that can be used to schedule instructions concurrently. List the advantages and disadvantages of each. [4%]

(d) Define the following measures used to characterise multi-computer networks, what are the units of these measures[4%]:

(i) node degree
(ii) channel bandwidth
(iii) diameter
(iv) bisection bandwidth

(e) Compute values of the measures in part (d) for a 2-dimensional torroidal

network of 1024 nodes, where the channel cycle time is 1ns and the channel width is 1 Byte. [4%]

Question 2.

(a) The register file provides rapid access to data in a microprocessor, it can usually be read in a single cycle and provides operands to the execution units. Describe how the demands placed on a register file will change with increasing concurrency of instruction issue and what effect this will have on the silicon area of the register file implementation [5%].

(b) Draw a block diagram of a generic superscalar architecture that issues instructions out-of-order from an instruction window, which can issue and retire up to 4 instructions per cycle. This should show details of the register file and execution units and identify all data paths and ports required. Briefly describe its operation. [5%].

(c) Draw a block diagram of a generic VLIW architecture that can issue 4 instructions per cycle. This should show details of the register file and execution units and identify all data paths and ports required. Briefly describe its operation. [3%].

(d) Draw a diagram that illustrates the Niagara I architecture with 8 4-way threaded cores. This should show details of the register file and execution units and identify all data paths and ports required. Briefly describe its operation. [5%]

(e) Given that each of the above requires the same number of registers per instruction issued, compare the physical sizes of the register files for each of these architectures. [2%]

Question 3.

(a) Define the following dependencies that may occur in the concurrent execution of operations. For each, give a fragment of assembly code or pseudo assembly code to illustrate the dependency. Describe how you would detect such dependencies when issuing instructions with operands taken from a register file.

(i) Data dependency [3%]
(ii) Output dependency [3%]
(iii) Anti dependency [3%]

(Assembly format: `op rega <- regb, regc;` rega is the target register.)

(b) Which of the dependencies described above is introduced by: (i) allowing instructions to complete out of order; and (ii) allowing instructions to be issued

out of order? [2%]

(c) What technique can be used to eliminate the dependencies you have identified in part (b) above. Describe one way in which this may be implemented in a superscalar architecture. [6%]

(d) You wish to execute the loop bodies in the following pseudo code in parallel using multiple threads (one per index value from 1 to n):

```
forall i in{1..n}
    a[i] = a[i] + rho*a[i+k];
```

What can you say about the dependencies introduced in executing this code concurrently. If k>0 rewrite this code to eliminate any dependencies. [3%]

Question 4.

Describe the following mapping strategies in relation to cache memory. Make use of diagrams in your answers where appropriate.

(i) direct mapped [2%]
(ii) 4-way set associative [2%]
(iii) fully associative [2%]

A Level-1 Instruction cache of 32Kbytes with a cache line of 128 bits is addressed by a 30-bit address to access a 4-byte instruction word. Given these parameters, then:

(a) For each mapping strategy, identify how many different cache lines an arbitrary address in memory can be mapped to and how many tag matches are involved in the access of that address? [2%]

(b) For each of the mapping strategies, divide the 30-bit address into fields and describe how each field is used in accessing an instruction in the cache. Use diagrams of the components that make up the cache in your answer. [4% for each strategy]