UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

COMPUTER SCIENCE FOUR

PARALLEL ARCHITECTURES

Thursday 18 April 2002

09:30 to 11:00

Examiners:

Prof. K. Turner (External)

Prof. C. Stirling (Chairman)

INSTRUCTIONS TO CANDIDATES

Answer any TWO questions.

All questions carry equal weight.

1. (a) Identify and explain the various types of inter-instruction hazard that can occur in a processor such as the DLX with a set of parallel arithmetic execution units. What mechanisms can be used to ensure that these hazards do not cause erroneous operation? [*7 marks*]

   (b) Explain briefly the principles of operation of each of the following:

      i. SIMD extension instructions
      ii. Superscalar architectures
      iii. VLIW architectures

   [*9 marks*]

   (c) You have been asked to advise a start-up company on the choice of processor to be used in a new high-performance workstation. In order to remain competitive in the future, the company plans to take advantage of the increases in the degree of parallelism available within the chosen processor architecture. For each of the architecture types identified above in part (b), explain:

      i. how easy or difficult it will be for the processor manufacturer to increase the degree of parallelism
      ii. how easy or difficult it will be for the company to exploit these increases in the degree of parallelism

   [*9 marks*]

2. (a) Explain what is meant by a *routing pattern* and give examples of the different kinds of pattern which can occur in a multiprocessor system. Explain what is meant by a *hot spot* and identify which of the routing patterns is likely to give rise to hot spots. [*6 marks*]

   (b) Compare and contrast buses and interconnection networks as mechanisms for joining together processors and memories in parallel computing systems, with particular reference to *cache coherence.* [*7 marks*]

   (c) Figure 1 on the next page shows the state transition diagrams for the Illinois cache coherence protocol and Figure 2 shows the states of the lines in the caches of 3 processors in a multiprocessor system which uses this protocol (I=Invalid, E=Exclusive, S=Shared, P=Private Dirty). State what happens and what the new contents and states of the lines are as each of the following occurs in sequence:

      i. P0 reads a word in line 2 using address tag 1
      ii. P1 writes to a word in line 0 using address tag 2
      iii. P0 reads a word in line 3 using address tag 2
      iv. P2 writes to a word in line 1 using address tag 3
      v. P0 reads a word in line 1 using address tag 3
      vi. P1 writes to a word in line 4 using address tag 1

   [*12 marks*]

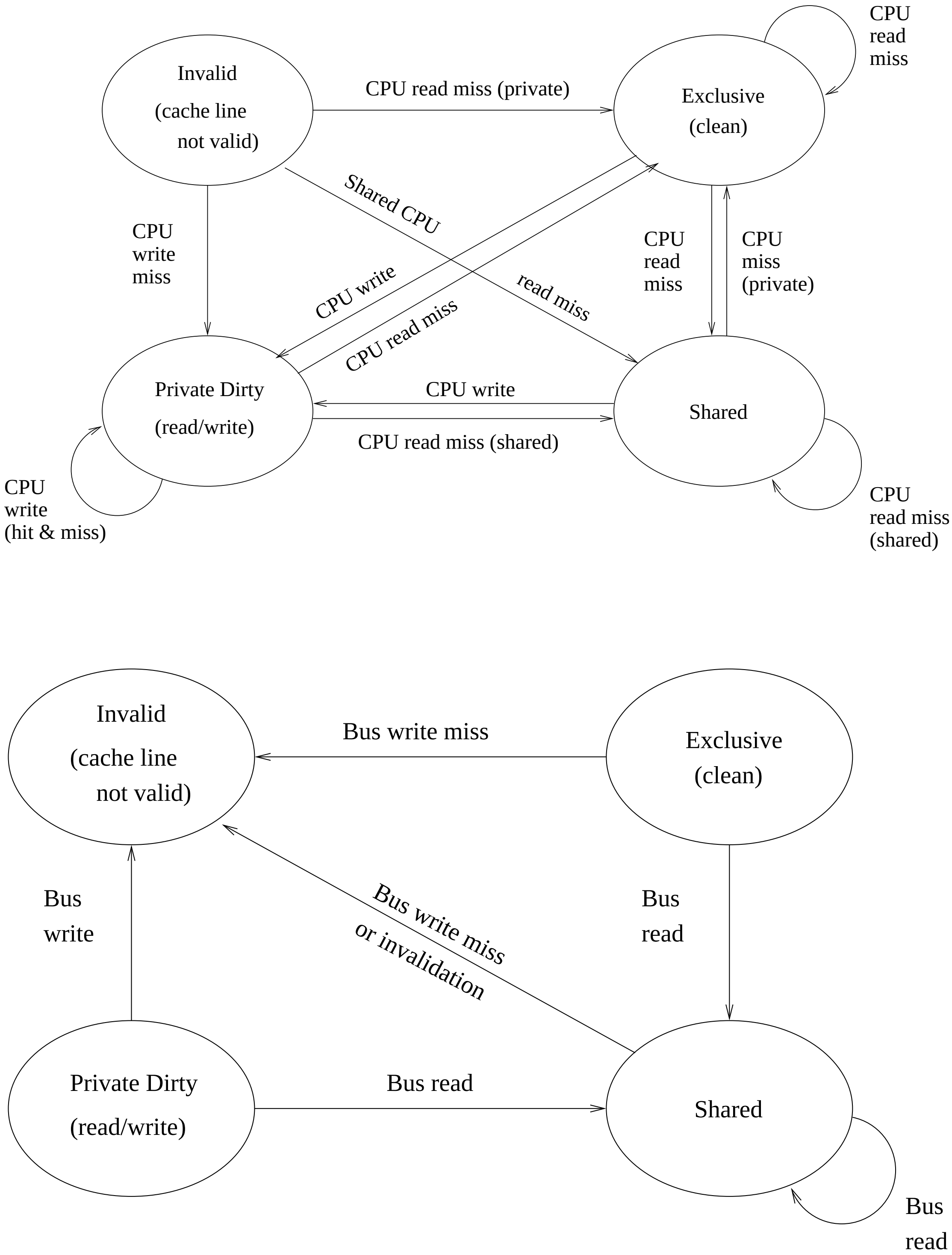

Figure 1 Illinois Cache Coherence Protocol

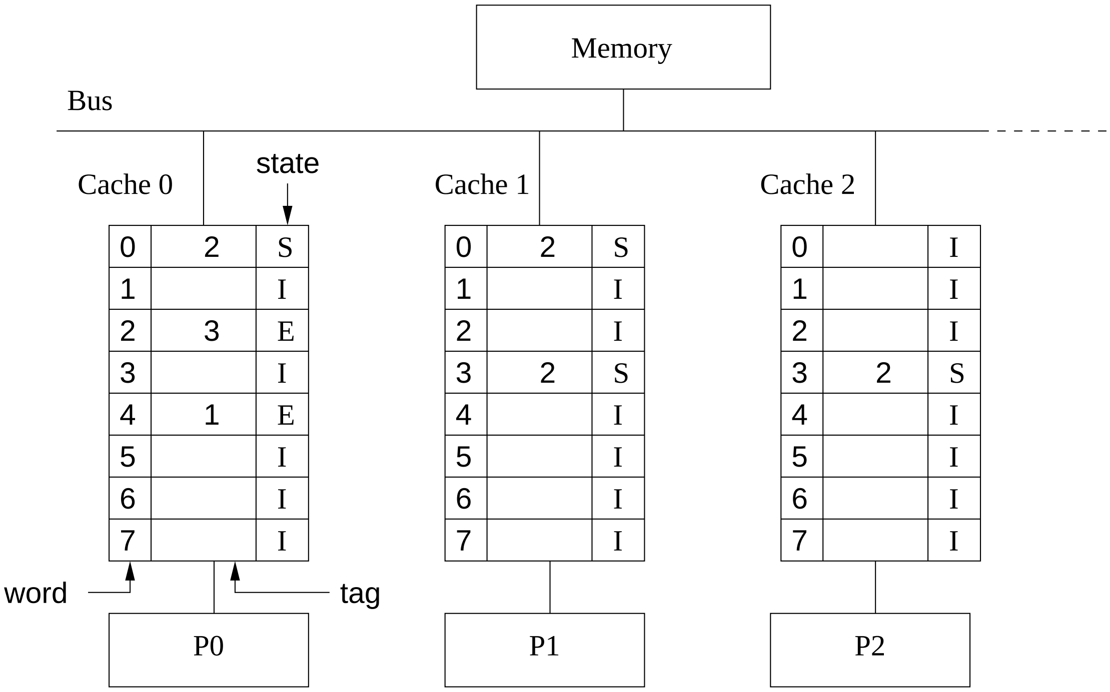

Figure 2 Cache line contents

3. (a) What do you understand by the terms *granularity* and *degree of parallelism*? How are these two related?

   Explain the term *granularity efficiency* and derive an expression for granularity efficiency in terms of grain size, context switching time and synchronisation time. [*8 marks*]

   (b) In a certain parallel system, context switching takes 40 clock periods and synchronisation takes $8 + \sqrt{N}$ clock periods, where N is the number of processors in use. The granularity efficiency of a program used to solve a particular problem on this system is 90% when using 64 processors. Assuming that the grain size remains constant as the problem size increases, how many processors can be used before the granularity efficiency falls below 80%? [*7 marks*]

   (c) Derive an equation which demonstrates why the average degree of parallelism in a program is an upper bound on the speed-up of a parallel system with an unlimited number of processors. Explain what happens when a program is run on a system with a limited number of processors. What are the factors which influence system performance in these circumstances? [*10 marks*]