Linköpings Universitet
Institutionen för datavetenskap (IDA)
Zebo Peng

Tentamen i kursen
TDTS08 Datorarkitektur

Examination of the course
TDTS08 Advanced Computer Architecture

2019-01-07, 8:00-12:00

Hjälpmedel:

Engelsk ordbok.

Supporting material:

English dictionary.

Poänggränser:

Maximal poäng är 40.
För godkänt krävs 21 poäng;
för betyg 4 krävs 27 poäng; och
för betyg 5 krävs 33 poäng.

Points:

Maximum points: 40.
You need 21 points to pass the exam;
for grade 4, 27 points are needed; and
for grade 5, 33 points are needed.

Jourhavande lärare (Teacher on duty)
Zebo Peng, tel. 013-28 2067

Note: You can give the answers in English or Swedish.

1. a) What is the basic idea of associative mapping for cache organization? What are the advantages and disadvantages of the associative mapping organization, as compared with the direct mapping organization?

    b) Why is the fully associative cache organization seldom used in practical computers? Which cache organization is commonly used? Why?

    (3p)

2. a) What does it mean by a memory hierarchy? Why it is useful to build a memory hierarchy?

    b) What is the fundamental assumption that makes a memory hierarchy work efficiently?

    c) Is the micro-memory used to store microprograms a part of the memory hierarchy? Why?

    (3p)

3. a) Discuss how the bimodal prediction technique work for branch prediction.

    b) Why does the bimodal prediction technique give better performance than the one-bit prediction method? Give a concrete example to support your argument.

    (3p)

4. A computer has an instruction pipeline with five pipeline stages. As a designer, you are asked to consider the possibility of increasing the number of pipeline stages of this computer in order to improve its performance. It turns out that one of the stages can't be divided into two or several shorter stages. Does it make sense to divide the other three stages so that the number of stages increased? Why?

    (3p)

5. We have discussed a technique called forwarding (also bypassing) in the context of instruction pipelining.

    a) Describe the basic idea of this technique.

    b) What problem does this technique solve? Why is it important to solve this problem?

    (3p)

Note: You can give the answers in English or Swedish.

6. a) What are the most essential characteristics of a superscalar architecture?

b) What does it means by register renaming? What is the main purpose of using such a technique?

c) Give a concrete example to illustrate how register renaming is used.

(3p)

7. a) Define the concepts of instruction-level parallelism and machine parallelism, in the context of a superscalar architecture. What are the differences between them?

b) Is the instruction-level parallelism completely determined by an application? Why?

(3p)

8. a) A VLIW architecture is said to support explicit parallel instruction execution. Define the concept of explicit parallelism. What are the advantages of exploiting explicit parallelism?

b) What is the main problem of a traditional VLIW computer? How is this problem addressed by the IA-64 architecture?

(3p)

9. How are general-purpose microprocessors usually extended to handle multimedia applications with little extra hardware cost?

(2p)

10. a) Describe the different multithreading approaches and discuss how they are applied in the context of a superscalar architecture. What are the advantages and disadvantages of these different approaches, respectively?

b) Why does multithreading improve system performance even in the case when there is only a single scalar processor in your computer?

(3p)

11. a) Define the concept of a cluster computer. What are the main features of such a computer?

b) What are the advantages of using a cluster computer?

(2p)

12. When we have an L1 cache that does not connect to the bus, and would still like to use the MESI cache coherence protocol, we can use the write-though policy for the L1 cache. This forces any modification to an L1 cache line out to the L2 cache, and therefore makes it visible to other L2 caches.

a) In order for this scheme to work, it is required that the L1 content must be a subset of the L2 content. This means that the associativity of the L2 cache should be equal to or greater than that of the L1 associativity. Why?

b) Give a concrete example (i.e., a cache access sequence) to support your answer to (a).

(3p)

13. a) What are the main features of a graphics processing unit (GPU)?

b) Which of the GPU features have contributed to its high performance? How?

c) Discuss the concept of divergent execution in a GPU processor. What is the main impact of such divergent execution?

(3p)

14. a) The dynamic power consumption for CMOS circuits is determined by three parameters, and therefore we have three degrees of freedom in the space for low-power design. What are these three parameters?

b) For each of these parameters, describe one technique which is most efficient, in your opinion, to achieve low-power design. Explain why you think the described technique is most efficient. (Note: You should present only one technique for each parameter; discussing more than one technique for a give parameter will receive no point).

(3p)