TEKNISKA HÖGSKOLAN I LINKÖPING
Institutionen för datavetenskap (IDA)
Zebo Peng

Tentamen i kursen
TDTS08 Datorarkitektur

Examination of the course
TDTS08 Advanced Computer Architecture

2017-04-20, 14:00-18:00

Hjälpmedel:

Engelsk ordbok.

Poänggränser:

Maximal poäng är 40.
För godkänt krävs 21 poäng.

Supporting material:

English dictionary.

Points:

Maximum points: 40.
You need 21 points to pass the exam.

Jourhavande lärare (Teacher on duty)

Arian Maghazeh, tel. 013-28 1407

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

1. a) What are the motivations for using a virtual memory?

   b) How is a virtual memory organized?

   c) How does a virtual memory ensure that two different processes will not write into each other's address space?

   (3p)

2. a) What does it by mean by a memory of sequential access type?

   b) Give an example of a sequential access memory.

   c) Can a memory of sequential access type be used as the main memory of a computer system? 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 an concrete example to support your argument.

   (3p)

4. The design of RISC architectures is based on certain characteristics of program execution.

   a) What are the characteristics related to procedure calls and returns?

   b) What is the mechanism used in a RISC architecture to make procedure calls and returns efficient? How does this mechanism work?

   (3p)

5. a) What is a data hazard in a pipelined unit? Illustrate this problem by an example and show how penalties are produced (consider a 6-stage pipeline as an example).

   b) How can this penalty be reduced with the forwarding (bypassing) technique? Draw figures to illustrate the pipelined executions without and with forwarding.

   (3p)

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

   b) Explain the following two policies for instruction execution:

   - in-order issue with out-of-order completion, and
   - out-of-order issue with out-of-order completion.

c) Why is the window of execution an important mechanism for a superscalar architecture?

(3p)

7. a) Draw a picture to show a typical VLIW architecture.

b) 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?

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

(4p)

8. a) What is the main function of a vector unit? Why can a vector unit speed up the execution of vector programs dramatically?

b) What is the role of the mask register in a vector unit? Give an example to illustrate the use of the mask register.

(3p)

9. With the development of silicon technologies, a huge number of transistors can now be integrated in a single chip. We can implement therefore either a superscalar architecture or a multi-core architecture in a single chip.

a) Discuss the main features of a superscalar architecture and a multi-core architecture, respectively.

b) Compare these two different architectures to each other. When implemented with the same amount of silicon area, which one of these two architectures will usually give better performance? Why?

(3p)

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. When we have an L1 cache that does not connect to the bus, and would still like to use the MESI protocol, we can use the write-though policy in the L1 cache. This forces any modification to an L1 line out to the L2 cache, and therefore makes it visible to other L2 caches. 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? Give an example to illustrate that if the L1 associativity is greater than the L2 associativity, the content of L1 might not be a subset of the L2 content, even if L2 is much bigger than L1 in size.

(3p)

12. 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)

13. 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)