Försättsblad till skriftlig tentamen vid Linköpings universitet

Datum för tentamen	2019-10-22
Sal (1)	T2(27)
Tid	8-12
Utb. kod	TDTS08
Modul	TEN1
Utb. kodnamn/benämning **Modulnamn/benämning**	Datorarkitektur Skriftlig tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	13
Jour/Kursansvarig Ange vem som besöker salen	Zebo Peng
Telefon under skrivtiden	013-282067
Besöker salen ca klockan	10.00
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Veronica Kindeland Gunnarsson, tel 28 5634, veronica.kindeland.gunnarsson@liu.se
Tillåtna hjälpmedel	Engelsk ordbok
Övrigt	
Antal exemplar i påsen	

Linköpings Universitet
Institutionen för datavetenskap (IDA)
Zebo Peng

Tentamen i kursen
TDTS08 Datorarkitektur

Examination of the course
TDTS08 Advanced Computer Architecture

2019-10-22, 8:00-12:00

Hjälpmedel:

Engelsk ordbok.

Poänggränser:

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

Supporting material:

English dictionary.

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 are the differences between direct mapping and associative mapping for a cache?

   b) For a set-associative mapping cache, a main memory address is viewed as consisting of three fields. Define these three fields, and explain how they are used.

   c) Assume that a 4-way set-associate cache has line size of 32 bytes and a total cache size of 4 kB; the 32 MB main memory is byte addressable. Show the format of the main memory addresses (i.e., how they are divided into the three fields).

   (4p)

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) What does it mean by a page fault, in the context of a virtual memory?

   b) Describe all things that need to be done with a page fault occurs.

   c) Is it the operating system or is it the hardware that takes control when a page fault occurs? Why?

   (3p)

4. 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)

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) Prior to the early 1980s, computers were built with more and more complex instruction sets. Why has there been a move to RISC machines away from complex instruction machines?

   b) Describe all features of a RISC machine that make it better than a CISC machine with respect to performance and power efficiency.

   (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) 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)

9. a) Multi-level caches can be designed with different organizations. In this context, define the two organizations: inclusive caches and exclusive caches, and describe their respective features?

   b) Consider an example of a two level cache hierarchy where the size of L2 is only 2 times of the size of L1. Which of the above two organizations is better? Why?

   (3 p)

10. 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)

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

11. 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)

12. How are x86 instructions translated into VLIW instructions by the Crusoe processor? Describe the features of this translation process and the main mechanism used in Crusoe to speed up the execution of the translated code.

    (3p)

13. a) In a multi-core computer, if the power supply to a core is maintained, but its clock is stopped in order to save power, the modified L1 data-cache lines of this core should be written back to the L2 cache shared with other cores or to the main memory. Explain why this is required.

    b) Explain also why we don't need to do the same thing for the L1 instruction-cache.

    (3p)