

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

5

Tentamen i kursen
TDTS08 Datorarkitektur
(Examination on TDTS 08 Advanced Computer Architecture)
2015-08-28, kl. 8-12

Hjälpmedel:

Engelsk ordbok.

Supporting material:

English dictionary.

Poänggränser:

Maximal poäng är 40.

För godkänt krävs 21 poäng.

Points:

Maximum points: 40.

You need 21 points to pass the exam.

Jourhavande lärare (Teacher on duty):

Zebo Peng, tel. 013-28 2067

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

1. a) Locality of reference is an important feature of programs, in the context of memory hierarchies. Explain what locality of reference means and why it is important.
b) In the IBM 360 Model 65 computers, memory addresses are assigned to two separate main memory units in such a way that all even-numbered words are in one unit and all odd-numbered words in the other. What might be the purpose of this particular arrangement? Explain also why this memory organization is useful.

(3p)

2. a) What does it 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 register file serves as a small and fast buffer for holding the variables that are currently manipulated by the CPU. From this point of view, a register file acts like a cache memory.
a) Can we then draw the conclusion that we can replace the cache with a large register file? Why?
b) Discuss the different characteristics of a large register file and a cache, including the way they are accessed and their usual contents, respectively.

(3p)

4. 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).
b) How can this penalty be reduced with the forwarding (bypassing) technique? Draw figures to illustrate the pipelined executions without and with forwarding.

(3p)

5. a) What does it mean by dynamic branch prediction? What is the basic principle that is used to implement dynamic branch prediction?
b) What are the advantages and disadvantages of dynamic branch prediction, as compared to static branch prediction?

(3p)

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

6. a) What is a CISC architecture? Describe the main features of a typical CISC architecture.
 b) What are the arguments for the CISC computers?
 c) What are the main problems with CISC computers?

(3p)

7. 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)

8. a) Identify all the true data dependencies, output dependencies and anti-dependencies on the following code. Provide the reasons for your answers.

```
L2: move r3,r9      Note: r3 <- r9
    load r8,(r3)    Note: r8 <- memory location pointed by r3
    add r3,r3,4     Note: r3 <- r3 + 4
    load r9,(r3)    Note: r9 <- memory location pointed by r3
    ble r8,r9,L3    Note: branch to L3 if r8 less than/equal r9
```

- b) Which of the identified dependencies can be eliminated? How?

(3p)

9. a) Even though both superscalar architecture and VLIW architecture exploit instruction level parallelism, a VLIW architecture can usually achieve a larger degree of parallelism than a superscalar architecture. Why?
 b) What are the main problems of a VLIW computer? How can these problems be addressed?

(3p)

10. a) Why is the placement of the “load from memory” operations an important issue for performance?
 b) Describe the speculative loading technique. What are the advantages of this technique?
 c) Illustrate the speculative loading technique with a simple example.

(3p)

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

11. a) There are two basic approaches to implement a snoopy protocol: write-invalidate and write-update. How do they work, respectively?
b) Describe the situation when the write-invalidate approach works better, and the situation when the write-update works better, respectively.
c) Both these approaches suffer from false sharing overheads. What does it mean by false sharing here?

(4p)

12. a) What is a thread? What does it mean by thread-level parallelism?
b) Describe the different multithreading approaches and discuss how they are applied in the context of superscalar architectures. What are the advantages and disadvantages of these different approaches, respectively?

(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)

