Linköping University
Dept. of Computer and Information Science (IDA)
Zebo Peng 2024-12-09

TDTS10 Computer Architecture

- Reading Instructions -

These instructions are compiled to guide your reading of the course book by William Stallings. They specify those chapters/sections in the book that will directly be covered by the written examination. They also include some review questions (quizzes) which you can be used to check whether you have understood the course materials or not.

Most of the listed review questions are typical for the course and may appear in the examination. However, you should not consider these as the *only* possible examination questions. To pass the examination and get good grades, you must understand the course materials and, at the same time, know how to apply the knowledge to solve problems.

The chapter numbers listed in this document refer to Stallings' book, "*Computer Organization and Architecture*," 11th Edition. Note that the questions are grouped into eight main areas, corresponding roughly to the eight lectures. You are also recommended to read all the lecture notes, available on the course website, in detail.

Lecture 1: Introduction, CPU, and Instruction Execution

Readings: Chapters 1.1-1.3, 2.4, 3.1, 3.2

1. What are the main components of a computer system? Briefly explain the basic function of each component.
2. What are the von Neumann architecture principles?
3. What are the advantages of the von Neumann architecture?
4. Why do we say that a computer is a general-purpose machine?
5. Describe the instruction execution cycle (machine cycle). What are the two main phases of instruction execution?
6. What are the main components of a CPU? What are the different components used for?
7. Explain briefly how the CPU components work together to execute instructions.
8. What is the function of each of the following registers in the CPU?
   - instruction register (IR)
   - program counter (PC)
   - accumulator register (AR)
   - general-purpose registers
9. Define the computer performance measurement units MIPS and FLOPS.
10. Explain Moore's law and discuss its implications.

Lecture 2: Memory System I

Readings: Chapters 4.1-4.4, 6,1, 7.1, 7.4, 7.5

1. What is the function of the main memory?
2. Why is memory access the bottleneck of a computer?
3. How do you define and compute the memory bandwidth of a memory?
4. How to increase the bandwidth of the main memory?
5. What does it mean by interleaving placement of program and data? Why is this placement approach useful?
6. What are the main features of a semiconductor main memory?
7. What are the differences among sequential access, direct access, and random access?
8. What is an associative memory? What is the advantage of using an associative memory?
9. What can a read-only memory (ROM) be used for? Why?
10. What are the main purposes of secondary memories?
11. Briefly explain how the following secondary storage devices work and discuss their main features:
    - Magnetic tape
    - Hard disk
    - Optical memory
    - USB flash drive
12. Give a short definition of seek time, rotational delay, read/write time, and data/transfer rate for a disk-based device.
13. How is a secondary memory accessed by the CPU?
14. What does it mean by a memory hierarchy? Why it is useful to build a memory hierarchy?
15. What is the fundamental assumption that makes a memory hierarchy work efficiently?
16. Give the definitions of locality of reference, temporal locality, and spatial locality.
17. What is the general relationship among access time, storage capacity, and cost of a given memory technology?

Lecture 3: Cache and Virtual Memory

Readings: Chapters 5.1-5.3, 9.3

1. What is a cache memory? How does it work? What are the main features of a cache?
2. What are the advantages of having a cache?
3. Is there any disadvantage of having a cache?
4. How is the average access time calculated for a given combined cache/memory system?
5. What are the main cache design issues/parameters?
6. Define the Harvard architecture. Why is it useful to use the Harvard architecture?

7. What are the advantages of using a unified cache, instead of two separate caches?
8. Describe the different cache mapping functions. Briefly describe the main features of each of the different mapping functions.
9. For a direct-mapped cache, a main memory address is viewed as consisting of three fields. List and define the three fields. Describe how these three fields are used to locate the content stored in the address in a step-by-step manner.
10. For an associative cache, a main memory address is viewed as consisting of two fields. List and define the two fields.
11. For a set-associative cache, a main memory address is viewed as consisting of three fields. List and define the three fields. Describe how these three fields are used to locate the content stored in the address in a step-by-step manner.
12. What are the different cache replacement algorithms?
13. Describe the two different write policies that are used to keep the cache contents and the contents of the main memory consistent.
14. What does it mean by virtual memory? Describe how a virtual memory works.
15. Why is it useful to have a virtual memory?
16. Is it necessary for all of the pages of a program to be in the main memory while the program is being executed?
17. How is a logical (virtual) address converted into a physical address of the main memory?
18. What does it mean by page fault? How is a page fault dealt with by the computer?
19. Describe the main principles for memory page replacement in a virtual memory.

Lecture 4: Instruction Set and Languages

Readings: Chapters 13.1, 13.3, 13.5, 14.1, 14,3, 15.1, 15.2

1. What is an instruction set? Explain why we say that the instruction set serves as the interface between computer hardware and software.
2. What are the main issues to be considered when designing an instruction set?
3. What information is usually specified in a machine instruction? Why?
4. List and briefly discuss the four main types of instructions.
5. What is the instruction format in an instruction set?
6. If an instruction contains one/two/three/four address(es), what can be the purpose of each address?
7. Describe the four basic addressing modes: direct addressing, indirect addressing, index addressing, and relative addressing. Give examples to show how they are used.
8. What does it mean by immediate addressing? Why is it useful to have this addressing mode?
9. What are machine codes?
10. What are assembly programs? What is the basic function of an assembler?
11. What are the differences between a compiler and an interpreter?

Lecture 5: Instruction Pipelining

Readings: Chapters 16.1-16.5.

1. What is the basic principle of an instruction pipeline?
2. Discuss one way to divide the instruction execution cycle into several stages.
3. In general, a larger number of pipeline stages gives better performance. Why doesn't this assumption lead to the situation that we have a huge number of pipeline stages?
4. Briefly discuss the different pipeline hazards that limit the performance of an instruction pipeline and the different solutions to address them.
5. Which pipeline hazard has the largest impact on the performance of a computer? Why?
6. What does it mean by the forwarding (bypassing) technique in the context of the instruction pipeline? Which problem does this technique solve?
7. What does it mean by delayed branch? Which problem does this technique solve?
8. What is speculative execution?
9. Briefly describe the static branch prediction methods.
10. What are the dynamic branch prediction methods? Discuss briefly the one-bit predictor, the bimodal prediction method, and the two-level prediction techniques.
11. What is the advantage of the bimodal prediction method, as compared with the one-bit predictor?

Lecture 6: I/O and Control

Readings: Chapters 8.1-8.5, 19.1-19.4

1. What are the main features and types of I/O devices and operations?
2. What are the main functions of an I/O module?
3. Explain the following ways of controlling I/O devices. What are the advantages and disadvantages of each technique?
   - Programmed I/O
   - Interrupt-driven I/O
   - Direct memory access (DMA)
4. Why do we need different control strategies for the I/O devices?
5. Describe the interrupt mechanism. What can it be used for?
6. What are the purpose and function of an Interrupt Service Routine (ISR)? Describe the main steps of the execution of ISR.
7. Discuss the two different ways to handle multiple interrupts.
8. What is the main purpose of the control unit inside a computer?
9. What are the purposes of the control signals generated by the control unit?
10. What are the two main techniques for control unit implementation?
11. Describe how the microprogramming technique works.
12. What is the purpose of a control memory?

13. Is the control memory part of the memory hierarchy of a computer system? Why?
14. What are the advantages of using the microprogramming technique?
15. Why is a ROM used for storing microprograms rather than a RAM?

Lecture 7: Superscalar and VLIW Processors

Readings: Chapters 18.1, 18.2

1. What is a superscalar architecture? What are the main features of such an architecture?
2. Define the concept of instruction-level parallelism.
3. Draw a diagram of a typical superscalar architecture design.
4. Briefly define the following terms in a superscalar architecture:
   - Resource conflict
   - Control (Procedural) dependency
   - Data conflict
   - True data dependency
   - Output dependency
   - Anti-dependency
5. Why is data dependency an important issue for superscalar architecture?
6. Which data dependencies can be eliminated? What is the technique used to eliminate them?
7. Why do we have a "commit" mechanism in a superscalar architecture? How does this mechanism work?
8. What is the purpose of an instruction window?
9. Discuss the impact of the instruction-window size with respect to performance and cost.
10. What are the main features of a VLIW processor?
11. Draw a diagram of a typical VLIW architecture design.
12. What does it mean by explicit parallelism? What are the advantages of an explicit parallel processor?
13. Why can VLIW architecture lead to a higher degree of parallelism than superscalar machines, in general?
14. Which architecture is more power efficient, superscalar or VLIW? Why?
15. How do you compare superscalar machines with VLIW computers? What are their advantages and disadvantages, respectively?

Lecture 8: RISC and Parallel Computers

Readings: Chapters 17.1, 17.2, 17.4, 17.5, 20.1, 20.2, 20.5, 20.6, 21.1

1. What are the main features of RISC computers? What are the advantages of RISC computers?
2. What are the main features of CICS computers?
3. Discuss the differences and arguments for RISC and CISC machines, respectively.
4. What are the main characteristics of program execution that have led to the development of the RISC architecture?

5. Why is it useful to have many registers in a CPU? Is there any disadvantage of having many registers?
6. Describe the concept of overlapping register windows. What are the main advantages of having such a mechanism?
7. Define Flynn's taxonomy for classification of computers. List and briefly define the different types of computer system organization according to Flynn.
8. What are the main features of an array processor?
9. Discuss the difference between a multi-processor system and a multi-computer system.
10. What is a symmetric multiprocessor system? What are its main characteristics?
10. What are the main features of a multi-core computer?
11. Discuss the main features of a NUMA architecture.
12. Define and discuss the main features of a cluster computer system. What are the advantages of having such a system?