

國立中山大學 107 學年度碩士暨碩士專班招生考試試題

科目名稱：計算機結構【資工系碩士班甲組、乙組】

題號：434001

※本科目依簡章規定「不可以」使用計算機(問答申論題)

共 2 頁第 1 頁

1. (10% total) You are designing a processor for an IoT embedded system. Based on the analysis of the monitoring software, the instructions listed in Table 1 show the specified latencies.

Table 1.

Instructions	Frequency	Latency
load	10%	7 cycles
store	15%	10 cycles
branch	15%	4 cycles
add	55%	4 cycles
divide	5%	5 cycles

- 1.1 (2%) What is the CPI of your processor on this mix of instructions?
- 1.2 (2%) Based on your design, if you could halve the cycle latency of any single category of instruction, but at the cost of increasing the cycle time by 20%. Should you make this change, and if so, what category of instruction should you speed up? (load, store, branch, add, or divide)
- 1.3 (3%) What is the CPI of your new design?
- 1.4 (3%) What is the speedup of your revised design over the original one, rounded to the nearest tenth?
2. (15% total) The standard pipeline has five stages: fetch (F), decode (D), execute (X), memory (M), and writeback (W). Consider the following **four-stage** pipelines, formed by combining two stages of the classic five-stage pipeline.
- 2.1 (3%) What specific **positive** impact would combining the “D” and “X” stages have on **CPI**?
- 2.2 (3%) What specific **positive** impact would combining the “X” and “M” stages have on **CPI**?
- 2.3 (3%) What specific **positive** impact would combining the “M” and “W” stages have on the **clock frequency** of the pipeline?
- 2.4 (3%) What specific impact would combining the “F” and “D” stages have on the implementation of branch prediction?
- 2.5 (3%) If a single-cycle datapath with delay of n nanoseconds is converted into a pipelined datapath with p stages, the clock period will be longer (slower) than n/p nanoseconds. Give two reasons for this.
3. (10% total) Please answer the following questions:
- 3.1 (2%) What do VLIW, superscalar execution, and array processing concepts have in common?
- 3.2 (4%) Provide two reasons why a VLIW microarchitecture is simpler than a “same-width” superscalar microarchitecture?
- 3.3 (4%) Provide two reasons why a superscalar microarchitecture could provide higher performance than a “same-width” VLIW microarchitecture?
4. (15% total) In Table 2, we have given you four different sequences of addresses generated by a program running on a processor with a data cache. Cache hit ratio for each sequence is also shown in Table 2. Assuming that the cache is initially empty at the beginning of each sequence, find out the following parameters of the processors’ data cache:
- 4.1 (3%) Associativity? (1, 2, or 4 ways)
- 4.2 (4%) Block size? (1, 2, 4, 8, 16, or 32 bytes)
- 4.3 (4%) Total cache size? (256 bytes, or 512 bytes)
- 4.4 (4%) Replacement policy? (LRU, or FIFO)

Table 2.

Sequence No.	Address Sequence	Hit Ratio
1	0, 2, 4, 8, 16, 32	0.33
2	0, 512, 1024, 1536, 2048, 1536, 1024, 512, 0	0.33
3	0, 64, 128, 256, 512, 256, 128, 64, 0	0.33
4	0, 512, 1024, 0, 1536, 0, 2048, 512	0.25

國立中山大學 107 學年度碩士暨碩士專班招生考試試題

科目名稱：計算機結構【資工系碩士班甲組、乙組】

題號：434001

※本科目依簡章規定「不可以」使用計算機(問答申論題)

共 2 頁第 2 頁

5. (10% total) Please answer the following questions:
- 5.1 (6%) Suppose you would like to connect 625 processors, and you are considering three different topologies: bus, point-to-point network, and mesh. Describe one disadvantage of each : (i) A single Bus(2%) (ii) A point-to-point network(2%) (iii) A 25×25 Mesh(2%).
- 5.2 (4%) Which one of these three topologies would you choose to connect 625 processors? Why?
6. (15% total) The Tomasulo's algorithm requires tag broadcast and comparison to enable wake-up of dependent instructions. To implement Tomasulo's algorithm in a machine that has the following properties:
- 8 functional units where each functional unit has a dedicated separate tag and data broadcast bus.
 - 32 64-bit architectural registers.
 - 16 reservation station entries per functional unit.
 - Each reservation station entry can have two source registers.
- 6.1 (3%) What is the number of tag comparators per reservation station entry?
- 6.2 (3%) What is the total number of tag comparators in the entire machine?
- 6.3 (3%) What is the (minimum possible) size of the tag?
- 6.4 (3%) What is the (minimum possible) size of the register alias table (or, frontend register file) in bits?
- 6.5 (3%) What is the total (minimum possible) size of the tag storage in the entire machine in bits?
7. (10% total) You are requested to parallelize an old program so that it could run faster on modern multicore processors. Please answer the following questions.
- 7.1 (5%) If your program can perform 90% of its work (measured as processor-seconds) in the parallel portion and 10% of its work in the serial portion. The parallel portion is perfectly parallelizable. What is the maximum speedup of the program if the multicore processor had an infinite number of cores?
- 7.2 (5%) How many processors would be required to attain a speedup of 4?
8. (15% total) A memory system is configured as:
- There are two DRAM channels.
 - Each channel has two ranks of DRAM chips.
 - Each memory channel is controlled by a separate memory controller.
 - Each rank of DRAM contains eight banks.
 - Each bank contains 32K rows.
 - Each row in one bank is 8KB.
- Assume the minimum retention time among all DRAM rows in the system is 64 ms. In order to ensure that no data is lost, every DRAM row is refreshed once per 64 ms. Every DRAM row refresh is initiated by a command from memory controller which occupies the command bus on the associated memory channel. Consider a 1.024 second span of time.
- Define the *utilization* (of a resource such as a bus or a memory bank) as the fraction of total time for which a resource is occupied by a refresh command.
- For each calculation in this question, you may leave your answer in simplified form in terms of powers of 2 and powers of 10.
- 8.1 (4%) How long does each refresh command occupy the command bus (in ns) such that across all memory channels, the command bus utilization due to refreshes is 8.192%? (Hint $8.192 = 2^{13}/1000$)
- 8.2 (4%) How long does each refresh command occupy the DRAM banks (in ns) such that across all the banks, the banks utilization due to refreshes is 8.192%?
- 8.3 (3%) What data bus utilization, across all memory channels, is directly caused by DRAM refreshes?
- 8.4 (4%) How many refreshes are performed by the memory controllers during the 1.024 second period in total across both memory channels combined?