

# Computer Architecture Ph. D. Qualification Exam. 2025

[calculator allowed; mobile phone is prohibited]

Name:

ID#

1. (20% total) Assume a 32-bit, byte-addressed machine with virtual addressing. The two high-order bits are **11** is treated as **unmapped**. These addresses are only accessible by the operating system and bypass virtual address translation. The answers need to express as a multiple of a power of 2, or in terms of KB, MB, GB, or TB as appropriate.
  - 1.1 (4%) What is the maximum amount of **physical memory** this system can address?
  - 1.2 (4%) What is the maximum amount of **virtual memory** any single process on this system can address?
  - 1.3 (4%) How many **virtual pages** are available to each process, assuming 4KB pages?
  - 1.4 (4%) Assume each page table entry is 4 bytes; how much memory would a single-level page table require?
  - 1.5 (4%) Assuming a two-level page table where half the bits of the virtual page number are used to index the first level, and the other half are used to index the second level, how many second-level page tables can a process use if its total page table size is limited to 400KB?

2. (20% total) The instruction latency for a given CPU is shown in Table 1.

Table 1.

| Instructions | Breakdown | Latency  |
|--------------|-----------|----------|
| load         | 5%        | 3 cycles |
| add          | 10%       | 5 cycles |
| divide       | 10%       | 8 cycles |
| branch       | 50%       | 2 cycles |
| shift left   | 15%       | 5 cycles |
| shift right  | 10%       | 1 cycles |

- 2.1 (5%) Calculate the **CPI** of the given CPU?
- 2.2 (5%) **Variation 1:** All add instructions are replaced with corresponding subtract instructions that take same amount of time. Calculate the **CPI** of the **Variation 1**?
- 2.3 (5%) **Variation 2:** Remove the highest-latency instruction, and replace all those instructions with additional latency of three cycles for the lowest latency instruction in the mix. Calculate the **CPI** of the **Variation 2**?
- 2.4 (5%) **Variation 3:** Reduce the latency for divides by a factor of four, but increase the latencies of branches by 50%. Calculate the **CPI** of the **Variation 3**?

3. (20%) Suppose that (after optimization) a typical **n**-instruction program requires an additional **4\*n NOP** instructions to correctly handle data hazards.

3.1 (6%) Suppose that the cycle time of this pipeline without forwarding is 250 ps. Suppose also that adding forwarding hardware will reduce the number of **NOPs** from **0.4\*n** to **0.05\*n**, but increase the cycle time to 300 ps. What is the speedup of this new pipeline compared to the one without forwarding?

3.2 (6%) Can a program with only **0.075\*n NOPs** possibly run faster on the pipeline with forwarding? Explain why or why not.

3.3 (8%) At a minimum, how many **NOPs** (as a percentage of code instructions) must a program have before it can possibly run faster on the pipeline with forwarding?

- 4 (20%) A memory system has four channels, and 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. A bank contains 32K rows. Each row in one bank is 8KB. 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 the memory controller which occupies the command bus on the associated memory channel for 5 ns and the associated bank for 40 ns.

Let us consider a 1.024 second span of time. We define 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. (Note: For each calculation in this question, you may leave your answer in simplified form in terms of powers of 2 and powers of 10.)

- 4.1 (8%) How many refreshes are performed by the memory controllers during the 1.024 second period in total across all four memory channels?
  - 4.2 (6%) What command bus utilization, across all memory channels, is directly caused by DRAM refreshes?
  - 4.3 (6%) What bank utilization (on average across all banks) is directly caused by DRAM refreshes?
- 
5. (20%) Processor microarchitecture: VLIW vs. Superscalar vs. Array Processor
    - 5.1 (4%) What do VLIW, superscalar execution, and array processing concepts have in common?
    - 5.2 (6%) Provide two reasons why a VLIW microarchitecture is simpler than a "same-width" superscalar microarchitecture.
    - 5.3 (4%) Provide a reason why a superscalar microarchitecture could provide higher performance than a "same-width" VLIW microarchitecture.
    - 5.4 (3%) Provide a reason why a VLIW processor is more flexible than an array processor.
    - 5.5 (3%) Provide a reason why an array processor is simpler than a "same-width" VLIW processor.