

Department of Computer Science and Engineering
National Sun Yat-sen University
First Semester of 2024 PhD Qualifying Exam

Subject : Computer Architecture

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 have a single-level page table require?
 - 1.5 (4%) Assuming a two-level page table where half the hits 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%) **Variant 1:** All add instructions are replaced with corresponding subtract instructions that take same amount of time. Calculate the **CPI** of the **Variant 1**?
 - 2.3 (5%) **Variant 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 **Variant 2**?
 - 2.4 (5%) **Variant 3:** Reduce the latency for divides by a factor of four, but increase the latencies of branches by 50%. Calculate the **CPI** of the **Variant 3**?
3. (20% total) An 8-bits byte-addressable virtual address space and the physical memory has 128 bytes. Each page contains 16 bytes. A simple, one level translation scheme is used and the page table resides on physical memory. The initial contents of the frames of physical memory are shown Table 2.

Table 2.

Frame Number	Frame Contents
0	Empty
1	Page 13
2	Page 5
3	Page 2

4	Empty
5	Page 0
6	Empty
7	Page Table

A three-entry translation lookaside buffer that uses the Least Recently-Used (LRU) replacement is added to this system. Initially, this TLB contains the entries for pages 0, 2, and 13. For the following sequence of reference, put a circle around those that generate a TLB hit and put a rectangle around them. (Note: LRU is used to select pages for replacement in physical memory.)
References (to pages): 0, 13, 5, 2, 14, 14, 13, 6, 6, 13, 15, 14, 15, 13, 4, 3.

- 3.1 (6%) What is the hit rate of the TLB for this sequence of reference?
- 3.2 (6%) At the end of this sequence, what three entries are contained in the TLB?
- 3.3 (8%) What are the contents of the 8 physical frames?

4. (20% total) A memory system has four channels, each with 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% total) Considering a system which applies the Sequential Consistency (SC). Suppose that three processes, P1, P2 and P3 are applied with different processors on a system (the values of RA, RB, RC were all zeros before the execution):

P1	P2	P3
P1.1: ST (A), 1	P2.1: ST (B), 1	P3.1: ST (C), 1
P1.2: LD RC, (C)	P2.2: LD RA, (A)	P3.2: LD RB, (B)

After all processes are executed, it is possible for the system to have multiple machine states. For example, $\{RA, RB, RC\} = \{1,1,1\}$ is possible if the execution sequence of instructions is $P1.1 \rightarrow P2.1 \rightarrow P3.1 \rightarrow P1.2 \rightarrow P2.2 \rightarrow P3.2$. Also, $\{RA, RB, RC\} = \{1,1,0\}$ is possible if the sequence is $P1.1 \rightarrow P1.2 \rightarrow P2.1 \rightarrow P3.1 \rightarrow P2.2 \rightarrow P3.2$. For each state of $\{RA, RB, RC\}$: $\{0,0,0\}$ $\{0,1,0\}$ $\{1,0,0\}$ $\{0,0,1\}$, specify the execution sequence of instructions that results in the corresponding state. If the state is NOT possible with SC, just put X.