Computer Architecture

Q1. What does “RISC” stand for? List the features of RISC type processors. What are *pros and cons* of RISC type architecture compared to CISC type architecture?

Q2. What is “cache associativity”? Discuss how cache hit rates change by cache associativity levels. How many comparators are required for fully associative, 4-way set associative, and direct-mapped caches if the cache has total N blocks?

Q3. Explain how “polling-based” and “interrupt-based” methods work for handling I/O devices in typical computer systems. What are *pros and cons* of each approach?

Q4. What is “superscalar” architecture? What are advantages of the superscalar architecture? Discuss why typical superscalar processors do NOT exhibit ideal performance.

Q5. What is “superpipelining” architecture? Discuss *pros and cons* of the superpipelining architecture.

Q6. Write an example (pseudo) RISC-type assembly code incurring a “data hazard” and a “control hazard” in the typical 5-stage pipelined CPU. Explain why the hazards occur using *pipeline execution diagrams*. The pipeline stages of the typical 5-stage pipelined CPU are IF, ID, EX, ME, and WB.

Q7. What does DMA stand for? Explain how DMA works between a host processor and an I/O device. What are benefits from DMA?

Q8. Let us assume that a server computer equips two processors supporting MESI cache coherence protocol. Write an example code incurring each of the following state transition and explain why each transition occurs with your example code.

(a) I-stage → E-state:

(b) E-state → S-state:

(c) S-state → M-state:

(d) M-state → S-state:

Q9. What are the advantages of a pipelined processor compared to a single-cycle processor? List three types of hazards that can be observed in the pipelined processors. Explain why each type of hazards occurs. Discuss possible solutions to reduce the occurrence of each type of hazards. (Note that pipeline stall cannot be a solution as clock cycles are wasted.)

Q10. We can classify misses in a cache based on the source of misses. This is called "**3C of cache**". Namely we can categorize the misses in a cache as *compulsory (cold)* misses, *capacity* misses, and *conflict* misses. Describe the source of each cache miss class. Discuss how we can reduce the miss rate of each cache miss class.

Q11. What is "Amdahl's law"? Using Amdahl's law, explain why we cannot get the expected peak performance with multi-core architecture. You *should* explain your answer with an equation, where N is the number of cores and p is the fraction of parallelizable parts in a workload.

Q12. What is "precise exceptions"? Explain why original Tomasulo algorithm (i.e. out-of-order processor without reorder buffer) cannot support precise exceptions.

Q13. In out-of-order (OoO) processor domains, we should consider *data hazards* caused by data dependencies between instructions.

(a) List the **three** types of data hazards that should be considered in OoO processors. Write a full name (not abbreviated form) of each type of data hazards.

(b) Write a two-line code that can cause *each* type of data hazards. You should use only a RISC-type **ADD** instruction and **registers** (i.e., R1, R2, R3, …) in your codes. Your answer should look like the following example.

```
XXX hazard:  ADD R1, R2, R3;   ADD R1, R2, R3;
YYY hazard:  ADD R1, R1, R1;   ADD R1, R1, R1;
```

Q14. Let us assume that memory subsystem is configured as follows. Answer the following questions.

Size of virtual address = 40 bits
Size of physical address = 32 bits
Page size = 4KB

(a) What is the maximum size of the virtual memory supported by the system?

(b) What is the size of the physical memory?

(c) Assuming the size of a single entry in the single-level page table is 4 byte, calculate the size of a page table.

(d) The size of a single-level page table is large, thus page tables can waste the memory space. Discuss how we can reduce the memory space occupied by page tables. What is a disadvantage of the proposed method?

Q15. Categorize processor types using Flynn's taxonomy. List example processors that are included in each category.

Q16. What are two types of data locality observed in typical workloads? List example data accesses that exhibit each type of data locality.

Q17. What does “CPI” stand for? Explain why we can say the performance of a computer system can be improved if CPI is decreased. (You should exhibit the equation that calculates execution time.) Show counterexamples (i.e. cases that performance is degraded even if CPI is reduced).

Q18. With dynamic instruction scheduling (e.g., scoreboard and Tomasulo), processors can perform the out-of-order execution. However, using static scheduling, processors can only perform in-order execution that is known to be problematic. Discuss the major pitfalls of the in-order execution.

Q19. What are the advantages of “dynamic branch prediction” compared to “static branch prediction”? Describe limitations of simple 1-bit and 2-bit dynamic branch prediction mechanisms (along with examples).

Q20. When we increase the block size of a cache, the miss rate initially decreases. But if the block size becomes a large fraction of the cache size, the miss rate increases again. Explain the reasons.