Second Trimester Examination

Name: _____

Student Id: _____

This is an open book examination. You are allowed to make use of the Internet, your class notes, etc. But you are not allowed to consult any human being, either in person or remotely. Attempt all questions.

Exam is graded out of 20, maximum grade points is 20.

Question 1 (5 points)

Consider the following code:

```
{
        double array[2^20];
        array[0] = (double) 0;
        array[1] = (double) 1;

        for(i = 0; i < 2^20 – 3; i++)
                array[i +2] = array[i] + array[i + 1];

}
```

a. If the cache line size is 64 bytes, estimate the hit ratio at the data L1 cache. Assume a cold start. State your assumptions clearly.

Question 2 (5 points)

a. For problem 1, assume a page size of 4KB and a fully associative data TLB with 8 entries. Compute the TLB hit ratio. State your assumptions clearly.

b. If the page table consists of a 6-level page table, would this change the answer to question 1? If so, how? If not, why?

Question 3 (5 points)

Consider a system where the base address of the page table is stored in a privileged register. The page table is a 6-level tree structured page table. When the operating system runs a user process, it loads the register with the starting address of the process. Loading and setting the value in the register are two privileged instructions that run only when the processor is in supervisor mode (the mode that allows access to privileged instructions). Now, it is desired to run a virtualized environment with a hypervisor and several operating systems. Answer the following questions:

a. Is the value in the register described above a virtual address or a real one? Justify your answer.

b. How can the hypervisor allow multiple operating systems to run on the system? How can it organize access to the privileged register? Assume no hardware change is possible.

Question 4 (5 points)

You have been tasked to design *a system* that works well for both cloud workloads and high-performance computing workloads. The processor chip you have is contains eight core, where each core is multithreaded among four symmetric hardware threads (symmetric hardware threads are identical threads in terms of priority and access to the processor resources). Each thread has a direct-mapped instruction and data L1 caches, whereas they share the L2 cache which is 4-way set-associative.

The cloud workloads consist of threads that run independently and belong to different contexts (or processes). The HPC workloads consists of threads that run cooperatively and belong to a single context (or a process).

For cloud workloads, we would like to get a level of performance isolation between the threads, whereas in HPC workloads we would like the threads to cooperate. How would you solve this problem?