

**CSCI-GA.3033-016**  
**Multicore Processors: Architecture & Programming**  
Homework Assignment # 2

1. Suppose that you have an array of structure. The declaration looks as follows:

```
struct info{  
    float a;  
    float b;  
    float c;  
    int d;  
}A[8];
```

8 threads will be accessing the different 8 elements of the array simultaneously (i.e. thread 1 accesses A[0], thread 2 accesses A[1], and so on).

a. [1] Draw a simple figure that shows how this array structure is stored in memory. Do not worry about exact addresses.

b. [1] Based on your figure, will there be a lot of cache misses or not when the 8 threads access the array (assume we have one shared cache only)? Justify

c. [2] If you say that the number of cache misses will be small, justify. If you say that we will have a lot of cache misses, how do we deal with that (from a programmer perspective, so don't mention a hardware technique)?

2. [3] Having too many threads (i.e. way more than the available cores) in an application may not be a good idea. State three reasons for that.

3.[3] Having too few threads in an application is not a good idea either. State three reasons explaining why. Assume the problem size is big enough.

4. Assume we have  $p$  hardware threads (i.e.  $p$  cores with one-way hyperthreading or  $n$  cores with  $v$ -way hyperthreading where  $v \cdot n = p$ ). For each of the following problems, specify how you are going to divide the problem among threads. Do not write code.

a) [3] Find all prime numbers between 1 and  $n$ .

b) [3] Find whether a number  $x$  is a prime number.

5. [2] A sequential application with a 20% part that must be executed sequentially, is required to be accelerated three-fold. How many CPUs are required for this task? How about five-fold speedup?

6. [2] Suppose we have a system with three level of caches: L1 is close to the processor, level 2 is below it, and level 3 is the last level before accessing the main memory. We know that two main characteristics of a cache performance are: cache access latency (How long does the cache take before responding with hit or miss?) and cache hit rate (how many of the cache accesses are hits?). As we go from L1 to L2 to L3, which of the two characteristics become more important? and why?