Computer Architecture HW #3
## Due: Friday, Feb. 10 (5 PM in ITT 305 mailbox or under my door, ITT 313)

Chapter 6. Exercises: 11, 13, 14, **and the following problems:**

Question A. Consider the following two sections of C code that both sum the elements of a 10,000 x 10,000 two-dimensional array M which contains floating points.

Code A	Code B

sum = 0.0;
for (r = 0; r < 10000; r++)
   for (c = 0; c < 10000; c++)
      sum = sum + M[r][c];
	sum = 0.0;
	for (c = 0; c < 10000; c++)
	   for (r = 0; r < 10000; r++)
	      sum = sum + M[r][c];

Explain why Code A takes 1.27 seconds while Code B takes 2.89 seconds. Hint: C uses row-major ordering to store two-dimensional arrays i.e.,
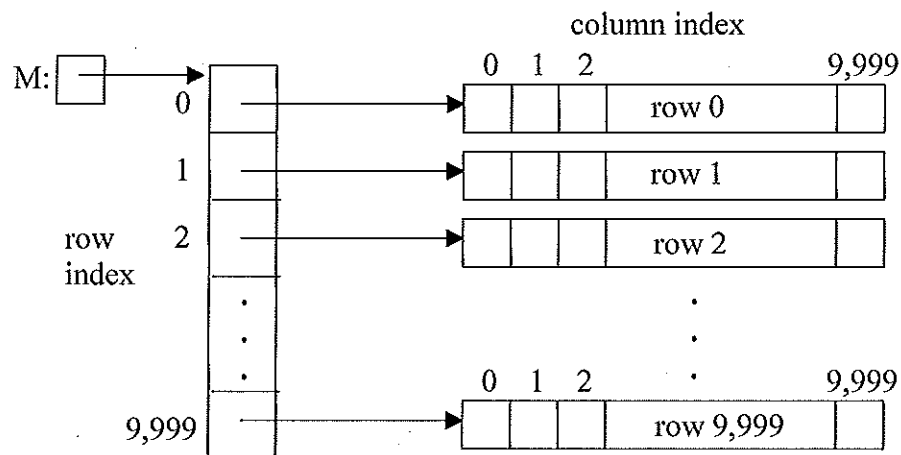
*11. Suppose we have a computer that uses a memory address word size of 8 bits. This computer has a 16-byte cache with 4 bytes per block. The computer accesses a number of memory locations throughout the course of running a program.

Suppose this computer uses direct-mapped cache. The format of a memory address as seen by the cache is shown here:

Tag 4 bits	Block 2 bits	Offset 2 bits

The system accesses memory addresses (in hex) in this exact order: 6E, B9, 17, E0, 4E, 4F, 50, 91, A8, A9, AB, AD, 93, and 94. The memory addresses of the first four accesses have been loaded into the cache blocks as shown below. (The contents of the tag are shown in binary and the cache "contents" are simply the address stored at that cache location.)

Block 0 — Tag Contents: 1110 — Cache Contents (represented by address): E0, E1, E2, E3

Block 1 — Tag Contents: 0001 — Cache Contents (represented by address): 14, 15, 16, 17

Block 2 — Tag Contents: 1011 — Cache Contents (represented by address): B8, B9, BA, BB

Block 3 — Tag Contents: 0110 — Cache Contents (represented by address): 6C, 6D, 6E, 6F

a) What is the hit ratio for the entire memory reference sequence given above, assuming we count the first four accesses as misses?

b) What memory blocks will be in the cache after the last address has been accessed?

13. A direct-mapped cache consists of eight blocks. Byte-addressable main memory contains 4K blocks of eight bytes each. Access time for the cache is 22ns and the time required to fill a cache slot from main memory is 300ns. (This time allows us to determine the block is missing and bring it into cache.) Assume a request is always started in parallel to both cache and to main memory (so if it is not found in cache, we do not have to add this cache search time to the memory access). If a block is missing from cache, the entire block is brought into the cache and the access is restarted. Initially, the cache is empty.

a) Show the main memory address format that allows us to map addresses from main memory to cache. Be sure to include the fields as well as their sizes.

b) Compute the hit ratio for a program that loops 4 times from locations 0 to 67_{10} in memory.

c) Compute the effective access time for this program.

14. Consider a byte-addressable computer with 24-bit addresses, a cache capable of storing a total of 64KB of data, and blocks of 32 bytes. Show the format of a 24-bit memory address for:

a) direct mapped

b) associative

c) 4-way set associative