Advanced Computer Architecture
Quiz 4

Name: _____ Roll #: _____

Question 1: Multiple Choice Questions

1. Lower-level memories are
   a. Lower in size
   b. Fastest in speed
   c. Lowest hit time
   d. None of these (larger in size, slowest in speed, and highest hit time)
2. If a book is referenced, then it again be referenced soon, the statement states a. Spatial locality
   b. Spectral locality
   c. Temporal locality
   d. Temporary locality
3. The principle of locality used for implementing memory of computers as
   a. Temporal locality
   b. Spatial locality
   c. Memory hierarchy
   d. All of these
4. Block address 16 has to place in directly mapped cache of 8 blocks. The block 16 goes which block no. of the cache?
   a. 0
   b. 8
   c. Anywhere
   d. None of the
5. If the total cache size is kept the same, how will increasing associativity affect the following?
   a. The number of blocks per set will decrease
   b. Index size will increase
   c. The tag size will decrease
   d. All of these
   e. None of these
6. For Set-associative and Fully associative which is not a block replacement strategy?
   a. Random
   b. Least recently used
   c. First in, First out
   d. None of these … (Collison)
7. Find the AMAT for a processor with a 1 ns clock cycle time, a miss penalty of 10 clock

cycles, a miss rate of 0.5 misses per instruction, and a cache access time (including hit detection) of 2 clock cycle. Assume that the read and write miss penalties are the same and ignore write stalls.

a. 1ns
b. 5ns
c. 7ns
d. None of these

Question 2:

A four-processor shared-memory system implements the MESI protocol for the cache coherence. For the following sequence of memory references, show the state of the line containing the variable **a** in each processor's cache after each reference is resolved. Each processors start out with the line containing **a** invalid in their cache.

	State of P0's cache	State of P1's cache	State of P2's cache	*State of P3's cache*
P0 reads a	*E*	*I*	*I*	*I*
P1 reads a	*S*	*S*	*I*	*I*
P2 reads a	*S*	*S*	*S*	*I*
P3 writes a	*I*	*I*	*I*	*M*
P0 reads a	*S*	*I*	*I*	*S*

Question 3:

"Critical Word First" and "Early Restart" are two impatience strategies to reduce "Miss Penalty". What is the difference between them?

- Critical Word First
  - Request the missed word first from memory and send it to the processor as soon as it arrives
  - Let the processor continue execution while filling the rest of the words in the block.

Early Restart

- Fetch the words in normal order
- But as soon as the requested word of the block arrives send it to the processor and let the processor continue execution.

What are Non-blocking Caches and how they impact Cache Bandwidth?

4th Optimization: Non-blocking Caches to increase Cache Bandwidth

- For computers allowing out-of-order execution, CPU does not need to stall on a data miss
- A non-blocking cache/lock-up free cache continues to supply cache hits during a miss
  - The processor could continue fetching instructions from the instruction cache while waiting for the data cache to return the missing data
- This "hit under miss" optimization reduces the effective miss penalty by being helpful during a miss

How modern compiler techniques can exploit the spatial & temporal locality of data in the following loop. (8th Optimization: Compiler Optimizations to Reduce Miss Rate)

```
for (j = 0; j < 100; j = j+1)     for
(i = 0; i < 5000; i = i+1)
       x[i][j] = 2 * [i][j];
```

By loop interchange