University of Toronto
Faculty of Applied Science and Engineering

ECE 552F

COMPUTER ARCHITECTURE

Spring 2001

Final Exam

Examiner: T.S. Abdelrahman

Duration: Two and half hours

This exam is open textbook(s) and open notes. Use of computing but non-communicating devices is permitted.

Do not remove any sheets from this test book. Answer all questions in the space provided. No additional sheets are permitted.

Work independently. The value of each part of each question is indicated. The total value of all questions is 100.

Write your name and student number in the space below. Do the same on the top of each sheet of this exam book.

Name: _____
(Underline last name)

Student Number: _____

Q1. _____	Q5. _____
Q2. _____	Q6. _____
Q3. _____	Q7. _____
Q4. _____	Q8. _____

Total

Question 1. (17 marks). *General.*

Answer the following questions by providing a **very brief and direct** answer. Answers that are either not brief or not direct will be penalized.

(a) **(1 mark).** What is the main advantage of VLIW processors over superscalar processors?

(b) **(1 mark).** What is the main advantage of superscalar processors over VLIW processors?

(c) **(1 mark).** True or False? The purpose of a Reorder Buffer (ROB) in a speculative out-of-order execution processor is to ensure that instructions are issued in the same order in which they appear in the program.

(d) **(1 mark).** How do registers get re-named in Tomasulo's scheme for out-of-order execution?

(e) **(1 mark).** How does Tomasulo's scheme for out-of-order execution avoid WAW hazards?

(f) **(1 mark).** Write the formula for the average access time for a 4-level cache system. Assume the hit time for the L1, L2, L3, and L4 levels of caches are H_{L1}, H_{L2}, H_{L3}, and H_{L4} respectively and that the miss penalty to memory is M.

(g) **(1 mark).** Why is there a need to disable some processors during execution on a SIMD multiprocessor?

(h) **(1 mark).** True or False? Multistage interconnection networks deliver the same performance as crossbar switches, but do so with less complex hardware.

(i) **(1 mark)**. What is the main advantage of adding a "clean and private" state to the cache coherence protocol of Figure 8.12 in your textbook?

(j) **(1 mark)**. True or False? False sharing in shared memory multiprocessors results when the unit of coherence is larger than the unit of access?

(k) **(2 marks)**. A distributed shared-memory multiprocessor has 32 processors. Each has a 512-Kbyte cache with 64 bytes cache lines, and a 4-Mbyte portion of shared memory.

How many entries are in the directory of each processor?

How many bits are in each entry?

(l) **(2 marks)**. Consider and the following code for two processors P_1 and P_2 on a multiprocessor system. Both A and B are assumed to be initialized to 0.

P_1	P_2
A = 1	print B
B = 2	print A

Which of the following outputs of the program are **not** possible on a multiprocessor system that enforces sequential consistency? Circle your answer(s).

1. B=2, A=1
2. B=2, A=0
3. B=0, A=1
4. B=0, A=0

(m) **(3 marks)**. In Assignment 3, you simulated the operation of a cache to determine the number of cache misses, and hence, the miss rate of the cache. How would you use your simulator to determine the percentage of these misses that are compulsory, capacity, and conflict?

(n) **(0 marks)**. What is the name of the music band that walked into class on December 5th, 2001?

Question 2. (15 marks). *Performance.*

Consider a pipelined processor with a 6-stage pipeline. For simplicity, assume that all instructions take 6 cycles. The following instruction mix is assumed.

Instruction Type	Frequency
Load	20%
Store	10%
ALU	50%
Branch	20%

(a) **(3 marks).** What is the ideal speedup of pipelining for this processor?

(b) **(3 marks).** Stalls due to data hazards occur only under two scenarios. A stall of one cycle occurs when a load instruction is followed by an ALU instruction that uses the result of the load. This scenario exists for 40% of load instructions. A stall of two cycles occurs when a branch instruction is preceded by an ALU operation whose result is used as the branch condition. This scenario exists for 50% of branch instructions.

What is the decrease in the ideal speedup of pipelining due **only** to data hazards?

The ideal speedup of pipelining is [ ] **% less due to data hazards.**

(c) **(3 marks).** Assume that branches resolve in the 5th stage of the pipeline. A branch target buffer is used to reduce the penalty of branch misprediction. Nonetheless, branches are mispredicted at the rate of 10%.

What is the decrease in the ideal speedup of pipelining due **only** to branch mispredictions?

The ideal speedup of pipelining is [ ] **% less due to branch mispredictions.**

(d) **(3 marks)**. Assume that the processor is connected to a memory system that consists of a 64-Kbyte first-level (L1) cache, a 512-Kbyte second-level (L2) cache, and a 128-Mbyte main memory. The hit rate for the L1 cache is 95%, while the hit rate for the L2 cache is 98%. The hit time for the L1 cache is 1 cycle and for the L2 cache is 4 cycles. The miss penalty for the L2 cache is 50 cycles.

What is the decrease in the ideal speedup of pipelining due **only** to the memory system?

The ideal speedup of pipelining is ☐ **% less due to the memory system.**

(e) **(3 marks)**. What is the decrease in the ideal speedup of pipelining due to data hazards, branch mispredictions and the memory system?

The ideal speedup of pipelining is ☐ **% less due to data hazards, branch mispredictions and the memory system.**

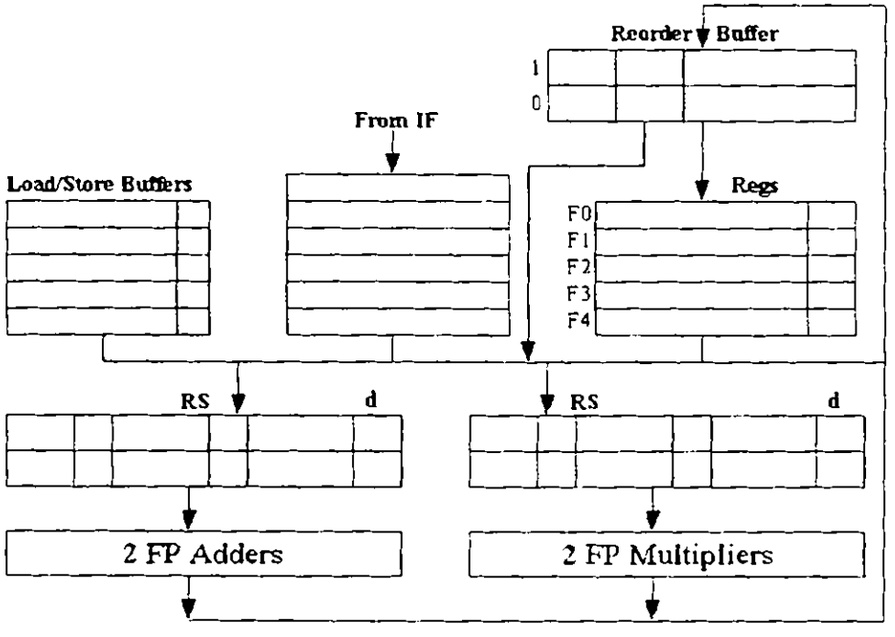
Question 3. (13 marks). *Out-Of-Order Speculative Execution.*

Simulate the execution of the following code fragment using the speculative Tomasulo's algorithm, as described in the lectures and in your textbook. Use the diagrams at the end of the question to aid your simulation, and then answer the questions below.

```
MULTF   F0, F1, F2
MULTF   F3, F2, F0
ADDF    F4, F2, F4
ADDF    F3, F4, F0
```

Make the following assumptions. A single instruction is issued per cycle. There are four function-units: two floating-point adders and two floating-point multipliers. An addition operation takes 2 cycles in the floating-point adder. A multiply operation takes 6 cycles in the floating-point multiplier. There are four reservation stations, one for each function unit. There is a 2-enrty re-order buffer. If more than one function unit requires the use of the CDB at the same time, the adder unit gets the CDB. At the beginning of cycle 1, all the instructions in the code fragment are in the instruction fetch buffer. All control bits in the RS and ROB are initialized to 0.

(a) **(4 marks).** Using the diagram shown below, show the state of the machine at the end of cycle 4.

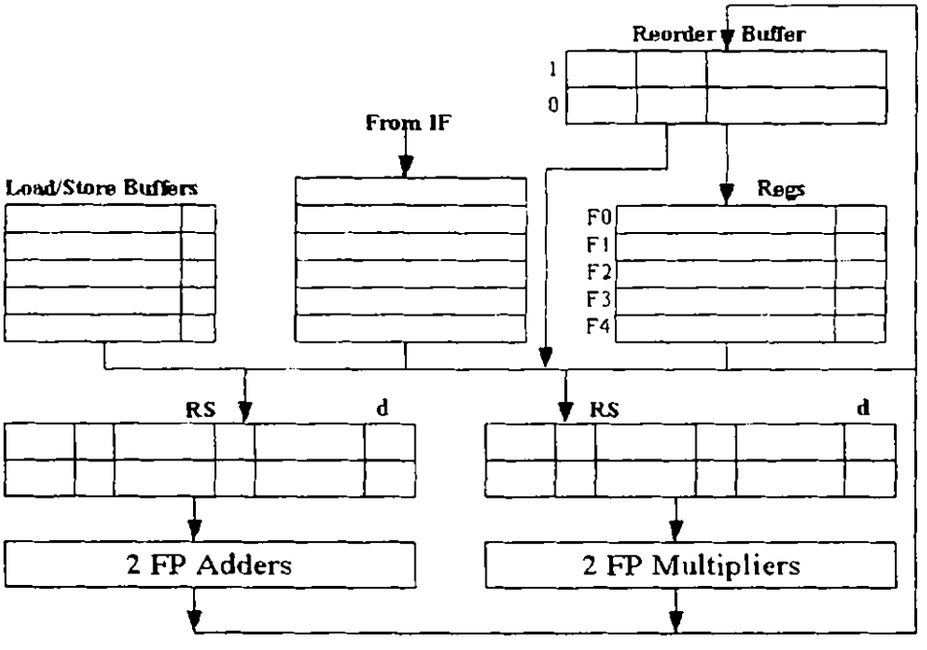


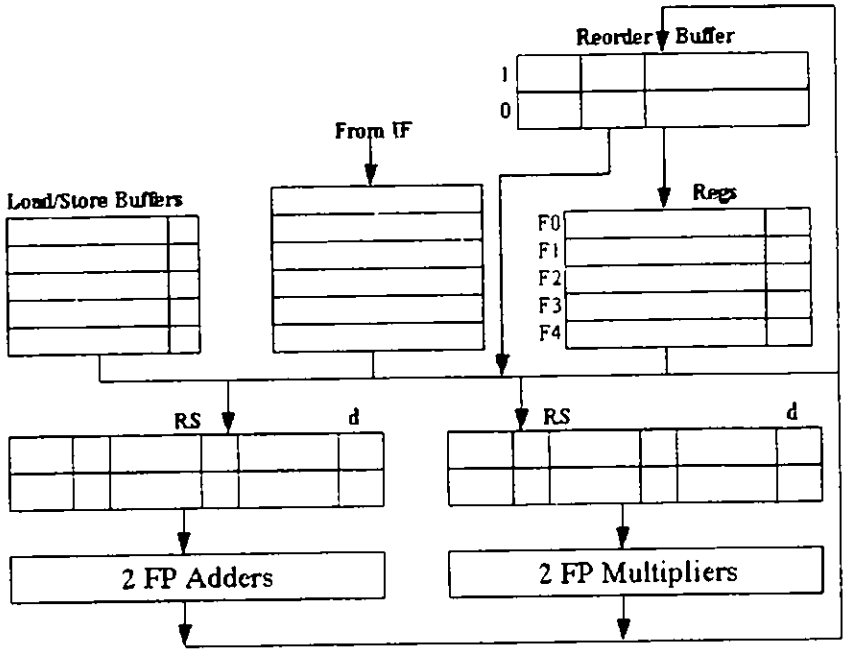
(b) **(4 marks)**. At the end of which cycle does the code fragment complete its execution? Assume the fragment completes its execution when all its instructions commit their results.

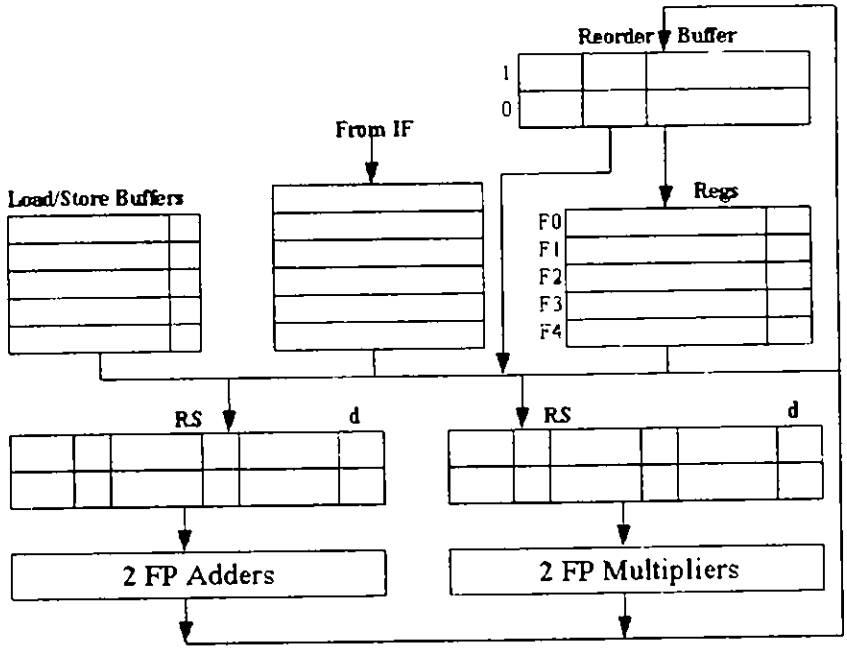
(c) **(2 marks)**. How many times was an operand forwarded to a reservation station using the CDB?

(d) **(3 marks)**. Is it possible for the above code fragment to benefit from compiler-based instruction scheduling in spite of the fact that the hardware supports out-of-order speculative execution? If no, explain why. If yes, show scheduled code.

Use the following diagrams to aid your simulation of Tomasulo's algorithm for the given code fragment.




Reorder Buffer
1
0
From IF
Load/Store Buffers
Regs
F0
F1
F2
F3
F4
RS
d
RS
d
2 FP Adders
2 FP Multipliers


Reorder Buffer
1
0
From IF
Load/Store Buffers
Regs
F0
F1
F2
F3
F4
RS
d
RS
d
2 FP Adders
2 FP Multipliers

Question 4. (16 marks). *Dynamic Branch Prediction.*

Consider the following code segment within a loop body:

```
if (x is even) {                // branch b1
  ++a;                          // b1 taken
}

if (x is multiple of 10) {      // branch b2
   ++b;                         // b2 taken
}
```

Assume that the following 8 values of x are processed by successive 9 iterations of this loop: 8, 9, 10, 11, 12, 20, 29, 30, and 31.

(a) **(6 marks).** Assume that a 1-bit branch history table predictor is used. Indicate in the table below the predicted and actual branch directions (i.e., taken (T) or not-taken (NT)) for b1 and b2 for each iteration of the loop.

	Values of x (i.e., iterations of loop)								
	8	9	10	11	12	20	29	30	31
b1 predicted									
b1 actual									
b2 predicted									
b2 actual									

Assume that each dynamic branch updates predictor entries before the next dynamic branch accesses the predictor. Also ignore the effect of branches in the code other than b1 and b2.

Assume that all predictor bits are initialized to 0; i.e., to NT.

(b) **(2 marks).** What is the misprediction rate for the 1-bit predictor above?

(c) **(6 marks)**. Assume a (1,1) correlating predictor is used. Indicate in the table below the predicted and actual branch directions (i.e., taken (T) or not-taken (NT)) for b1 and b2 for each iteration of the loop.

	Values of x (i.e., iterations of loop)								
	8	9	10	11	12	20	29	30	31
b1 predicted									
b1 actual									
b2 predicted									
b2 actual									

Assume that each dynamic branch updates predictor entries before the next dynamic branch accesses the predictor. Also ignore the effect of branches in the code other than b1 and b2.

Assume that all predictor bits are initialized to 0; i.e., to NT.

(d) **(2 marks)**. What is the misprediction rate for the (1,1) predictor above?

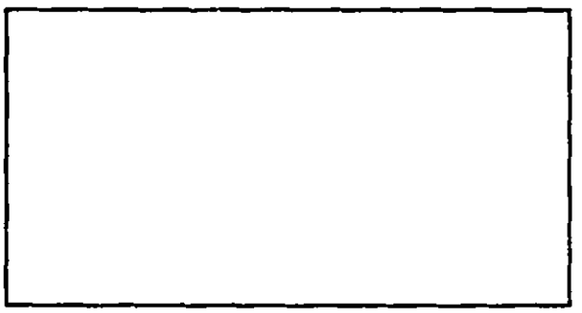
Question 5. (8 marks). *Predication.*

A processor allows the predication of any instruction on the value of a register being equal to, or not equal to zero. Hence, for example, an instruction MULT R1, R2, R3 can be predicated on R6 in one of two forms: CMULT R1, R2, R3 R6=0 or CMULT R1, R2, R3 R6!=0.

Predicate as many instructions as possible in the following code fragment.

```
Loop:   ADD   R1, R2, R3
        SGT   R9, R1, R11
        BEQZ  R9, Next
        SUB   R5, R6, R7
Next:   ADD   R8, R5, R12
        SLE   R10, R8, R13
        BNEZ  R10, Loop
        ADD   R7, R5, R12
```

Write predicated code here

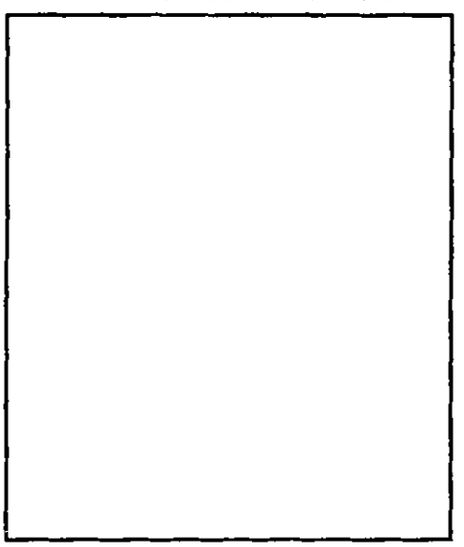


Question 6. (8 marks). *Instruction Scheduling.*

A processor has a 1-cycle branch delay slot. The following code fragment is part of a larger program. In the absence of code scheduling, "nop" instructions will be used to fill branch delays. Schedule the instructions in the code fragment to eliminate the use of as many "nop" instructions as possible. Assume that in the rest of the program, there are no upward exposed uses of registers R1, R6 and R9, but there are such uses of registers R7, R8 and R10.

```
        :
        :
Loop:   ADD   R1, R2, R3
        ADD   R6, R1, R7
        SGT   R9, R1, R11
        BEQZ  R9, Next
        SUB   R6, R6, R7
Next:   ADD   R8, R6, R12
        SLE   R10, R8, R13
        BNEZ  R10, Loop
        ADD   R7, R5, R12
        :
        :
```

Write scheduled code here



Question 7. (9 marks). *Page Tables.*

Consider a virtual memory system with the following parameters: a 36-bit address space, a page size of 4 Kbytes, and a physical memory size of 256 Mbytes.

(a) **(3 marks).** Assume the operating system uses a regular page table.

How many entries are in the page table?

How big is each entry in the page table in bits? Assume 8 bits for status.

How big is the page table (in Kbytes) in this case?

(b) **(3 marks).** Assume the operating system uses an inverted page table.

How many entries are in the page table?

How big is each entry in the page table in bits? Assume 8 bits for status.

How big is the page table (in Kbytes) in this case?

(c) **(3 marks).** Assume that the operating system limits the size of virtual memory used by a process to 4 times that of physical memory.

How many entries are in the page table?

How big is each entry in the page table in bits? Assume 8 bits for status.

How big is the page table (in Kbytes) in this case?

Question 8. (14 marks). *Cache Coherence.*

Consider the simple cache coherence protocol of Figure 8.12 on page 665 in your textbook. Assume three processors only, P_1, P_2, and P_3, each with a direct-mapped write-back cache C_1, C_2, and C_3 respectively. Each cache is 256 Kbytes with a cache line size of 64 bits, which holds two words of 32-bits each. The processors access memory using a shared bus. If more than one processor attempts to access the bus at the same time, the smallest numbered processor acquires the bus.

The three processors access three words X, Y, and Z in memory. The words are located in memory such that X and Y are the first and second words in one block B_1. The word Z is in a different block B_2. Because of the direct-mapped organization of the cache, B_1 and B_2 map into the same location L in each of C_1, C_2, and C_3.

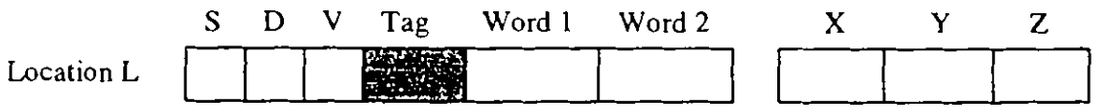
Assume the values of X, Y, and Z are initially 0 and that the memory blocks B_1 and B_2 do not initially exist in any cache.

Now consider the following sequence of operations performed by the two processors, shown in the table below. For simplicity, it is assumed that the processors are synchronized in cycles as indicated in the table. A blank entry for a processor in a given cycle implies that the processor does not accesses to X, Y, or Z in that cycle. A processor unable to acquire the bus in a cycle delays all its accesses until the bus is acquired.

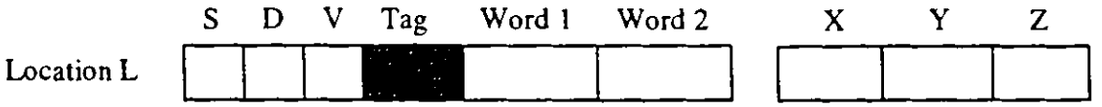
	P_1's Accesses	P_2's Accesses	P_3's Accesses
Cycle 1	Read X		
Cycle 2	Read X	Read Z	
Cycle 3		Read Z	Write 5 to X
Cycle 4	Write 4 to Y		
Cycle 5	Read Z		
Cycle 6	Read X		
Cycle 7	Write 8 to Z		
Cycle 8			Read X
Cycle 9		Read X	

In the following questions, use an "x" to indicate a "don't care" value. **Please note carefully which cache each question refers to.**

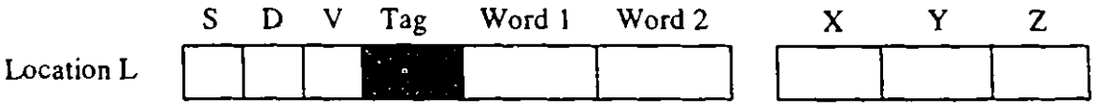
(a) **(2 marks).** Indicate below what the state and contents of Location L in C_1 and the values of X, Y, and Z in memory at the end of cycle 1.



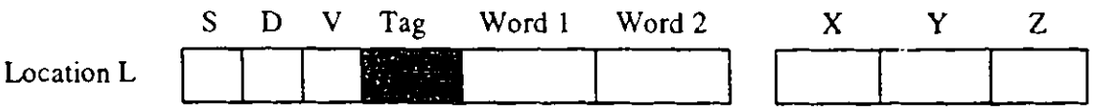
(b) **(2 marks)**. Indicate below what the state and contents of Location L in C_2 and the values of X, Y, and Z in memory at the end of cycle 2.



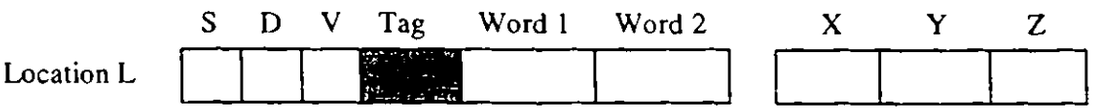
(c) **(2 marks)**. Indicate below what the state and contents of Location L in C_2 and the values of X, Y, and Z in memory at the end of cycle 3.



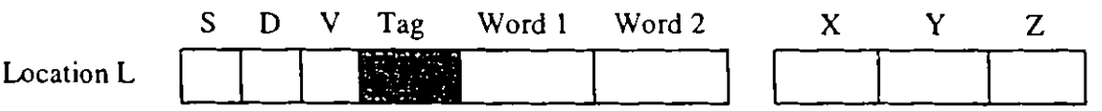
(d) **(2 marks)**. Indicate below what the state and contents of Location L in C_3 and the values of X, Y, and Z in memory at the end of cycle 3.



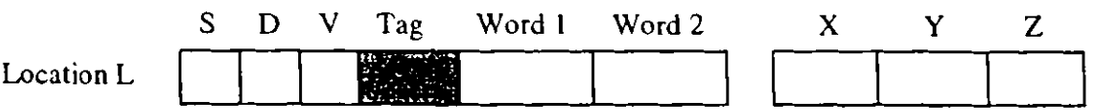
(e) **(2 marks)**. Indicate below what the state and contents of Location L in C_1 and the values of X, Y, and Z in memory at the end of cycle 5.



(f) **(2 marks)**. Indicate below what the state and contents of Location L in C_3 and the values of X, Y, and Z in memory at the end of cycle 5.



(g) **(2 marks)**. Indicate below what the state and contents of Location L in C_1 and the values of X, Y, and Z in memory at the end of cycle 9.



THIS PAGE IS INTENTIONALLY BLANK FOR ANSWER OVERFLOWS

THIS PAGE IS INTENTIONALLY BLANK FOR ANSWER OVERFLOWS