

---

CSEN702-Microprocessors  
Winter Semester 2021

Midterm Exam

---

Bar Code

Instructions: **Read Carefully Before Proceeding.**

- 1- Non-programmable calculators are allowed
- 2- Write your solutions in the space provided
- 3- The exam consists of **(5) questions**
- 4- This exam booklet contains **(13) pages** including this page
- 5- Total time allowed for this exam is **(120) minutes**
- 6- When you are told that time is up, stop working on the test

Good Luck!

---

Question	1	2	3	4	5	$\Sigma$
Possible Marks	10	20	10	30	20	90
Final Marks						

### Question 1 (10 pts)

Mark each of the following statement with T or F. Ambiguous letters will not be considered.

#	Statement	T/F
1	The first access to any array element $A[i]$ will result in a compulsory miss.	
2	Using multi-level caches helps in reducing the miss penalty.	
3	Way-prediction lowers the average memory access time by lowering the miss rate.	
4	In the blocking optimization, the total number of bytes occupied by one matrix block, should be always less or equal than the number of bytes in one cache block.	
5	The only limitation of the loop unrolling is the growing code size.	
6	An output dependence is a WAW hazard.	
7	For a program with no loops at all, branch prediction will increase the execution time compared with not using prediction at all.	
8	Leakage in microprocessor chips increases with lower voltages.	
9	Thermal Design power (TDP) is the maximum heat generated by the chip.	
10	The global miss rate of L2 cache should, in theory, be always less than that of the L1 cache.	

## Question 2 (20 pts)

Answer all the following short questions.

A) Consider a pipelined processor with 50% ALU operations, 30% branches, and 20% memory operations. It uses a branch prediction scheme for its branch instructions with a 2 cycles penalty. Given that we have 1 stall cycle every 5 instructions due to data hazards and that no structural hazards are present, what's the minimum **accuracy** of the branch prediction scheme to keep the CPI less than or equal to 1.3?

**Answer:**

B) Consider the following code:

```
for (i=0; i<9; i++)
    for (j=0; j<9; j++)
    {
        A[j][i]++;
    }
```

Given that A is a double-precision floating point array of size 10x10, and that the cache is fully associative, starts empty, and it has a total size of 192 KB:

*What effect does loop interchange on the overall performance compared with the original code performance shown above if the cache block size is 8 bytes? Discuss.*

**Answer:**

C) What's the difference between power gating and clock gating? Which one(s) affect the dynamic power and/or the static power?

**Answer:**

D) Two processors A and B exhibit the same average power and the same clock frequency. They have a different instruction set and architecture but manage to get the same CPI. Both have no stalls. Will they have the same energy efficiency? Explain why or why not.

**Answer:**

### Exercise 3 (10 pts)

Consider a cache system with a miss penalty of  $C$  cycles for a cache block of size  $C$  words. This is due to copying each word into the cache requiring 1 cycle.

Assume the following memory content organized in words. (A word is 32 bits).

Word 0
Word 1
Word 2
Word 3
Word 4
Word 5
Word 6
Word 7
Word 8
Word 9
Word 10
And so on...

i) For a cache that starts empty, with block sizes of 16 bytes, and a ***critical word first*** approach employed, compute the miss penalty cycles when word 2 is requested from memory. Also compute the number of cycles reduced compared the conventional approach.

ii) For a cache that starts empty, with block sizes of 16 bytes, and an ***early restart approach*** employed, compute the miss penalty cycles when word 2 is requested from memory. Also compute the number of cycles reduced compared the conventional approach.

iii) For which words, do both approaches perform the same? Discuss.

iv) For which words, does early restart perform its worst? Discuss

#### Exercise 4 (30 pts)

Consider two memory hierarchy systems M1 and M2.

M1 has two-level caches with the following measurements:

- In 100 memory references, 80 hit in L1 and 16 hit in L2.
- Hit time of L1 is 1 cycle and doubles in L2.
- Penalty of L2 is 10 cycles.

M2 has three-level caches with the following measurements:

- In 150 memory references, 120 hit in L1.
- Global miss rate of L2 is 6%, while it's 3% in L3.
- Penalty of L3 is 8 clock cycles.
- Hit time in L1 is 1 cycle, 3 cycles in L2 and 4 cycles in L3.

A) Compute the **local and global** miss rates of all caches in M1 and local miss rates of L1, L2 and L3 in M2. Show your work.

B) Compare both systems in terms of average memory access time. Show your work.

C) Given that 60% of the instructions are ALU and branch operations and the remaining are loads and stores, Compute the misses per instruction in all caches in both systems M1 and M2.



D) compare both systems in terms of average memory stalls per instruction. Show your work.

### Exercise 5 (20 pts)

Consider the following MIPS code and answer all the parts.

```
LOOP: L.D      F0, 800(R1)
      L.D      F1, 800(R2)
      ADD.D    F0, F0, F1
      MUL.D    F0, F0, F7
      S.D      F0, 800(R1)
      ADDUI    R1, R1, -8
      ADDUI    R2, R2, -8
      BEQ      R2, R3 LOOP
```

Assume the following latencies and that forwarding is applied and also assume branches execute in the decode stage normally and they don't use a delayed slot.

<i>Instruction producing result</i>	<i>Instruction using result</i>	<i>Latency in clock cycles</i>
FP load	FP ALU op	1
FP add	FP store	2
FP multiply	FP store	4
FP add	FP multiply	3

3.1) Compute the number of cycles needed for 1 iteration with no scheduling or unrolling.

Show your work.

The code is re-given here for your reference.

<b>LOOP:</b>	<b>L.D</b>	<b>F0, 800(R1)</b>
	<b>L.D</b>	<b>F1, 800(R2)</b>
	<b>ADD.D</b>	<b>F0, F0, F1</b>
	<b>MUL.D</b>	<b>F0, F0, F7</b>
	<b>S.D</b>	<b>F0, 800(R1)</b>
	<b>ADDUI</b>	<b>R1, R1, -8</b>
	<b>ADDUI</b>	<b>R2, R2, -8</b>
	<b>BEQ</b>	<b>R2, R3 LOOP</b>

3.2) Unroll the loop 2 times and schedule the loop to reduce the number of stalls to minimum.

Re-compute the number of cycles needed per iteration. (Use the registers F2 and F3, in addition to F0 and F1).

Show your work

Extra sheet

## Formula sheet

Pipelining
CPI pipelined = Ideal CPI + Pipeline stall clock cycles per instruction
$\text{Speedup} = \frac{\text{Pipeline depth}}{1 + \text{Pipeline stall cycles per instruction}}$ (can be used with ideal/balanced pipelining conditions)
Power and Performance
Energy <sub>workload</sub> = average power x execution time for the workload
$\text{Energy}_{\text{dynamic}} \propto 1/2 \times \text{Capacitive load} \times \text{Voltage}^2$
$\text{Power}_{\text{dynamic}} \propto 1/2 \times \text{Capacitive load} \times \text{Voltage}^2 \times \text{Frequency switched}$
$\text{Power}_{\text{static}} \propto \text{Current}_{\text{static}} \times \text{Voltage}$
$\text{Geometric mean} = \sqrt[n]{\prod_{i=1}^n \text{sample}_i}$
Memory hierarchy
CPU execution time = (CPU clock cycles + Memory stall cycles) x Clock cycle time
$\begin{aligned} \text{Memory stall cycles} &= \text{Number of misses} \times \text{Miss penalty} \\ &= \text{IC} \times \frac{\text{Misses}}{\text{Instruction}} \times \text{Miss penalty} \\ &= \text{IC} \times \frac{\text{Memory accesses}}{\text{Instruction}} \times \text{Miss rate} \times \text{Miss penalty} \end{aligned}$
$\frac{\text{Misses}}{\text{Instruction}} = \frac{\text{Miss rate} \times \text{Memory accesses}}{\text{Instruction count}} = \text{Miss rate} \times \frac{\text{Memory accesses}}{\text{Instruction}}$
Average memory access time = Hit time + Miss rate x Miss penalty
$\begin{aligned} \text{Average memory access time} &= \text{Hit time}_{L1} + \text{Miss rate}_{L1} \\ &\quad \times (\text{Hit time}_{L2} + \text{Miss rate}_{L2} \times \text{Miss penalty}_{L2}) \end{aligned}$
$\begin{aligned} \text{Average memory stalls per instruction} &= \text{Misses per instruction}_{L1} \times \text{Hit time}_{L2} \\ &\quad + \text{Misses per instruction}_{L2} \times \text{Miss penalty}_{L2} \end{aligned}$