

Homework 1

Due Tuesday, Sept 25, 2018 (11 pm) on T-square

This homework is meant for preparation of Midterm 1. It consists of midterm questions from previous three offerings of this course from the instructor (Fall 2012, Spring 2014, Spring 2015, and Fall 2017). This is a submit-only homework, which means as long as you make a reasonable effort in solving at least half of the questions you will get full points, regardless of the correctness of your solution. We will provide the solutions for these questions after the submission deadline, so you can check your answers and figure out where your concepts may be weak. Note that while this homework counts for only 1% of your grade, if you do it seriously it will help you prepare for the Midterm, which is worth 20%.

Good Luck!

Moin

Department of Electrical and Computer Engineering
Georgia Institute of Technology

ECE4100/ECE6100/CS4290/CS6290 (Section B, Q): Advanced Computer Architecture
Moinuddin K. Qureshi, Instructor
Prashant Nair and Junghee Lee, TAs

Exam 1, September 28, 2012

Name : _____

GT Account: _____

Problem 1 (24 points): _____

Problem 2 (16 points): _____

Problem 3 (20 points): _____

Problem 4 (20 points): _____

Problem 5 (20 points): _____

Total (100 points) : _____

This exam is given under the Georgia Tech Honor Code System. Anyone found to have submitted copied work instead of original work will be dealt with in full accordance with Institute policies.

Georgia Tech Honor Pledge: **“I have neither given nor received aid on this exam.”**

[MUST sign:] _____

Note: Where needed, show all your intermediate results to receive full credit. Do all your work in this examination handout. Use the back of the exam sheets if necessary.

Note: Please be sure your name is recorded on each sheet of the exam.

GOOD LUCK!

Name: _____

Problem 1 – “Potpourri” (24 points, there are four parts, each worth 6 points)

(I) Your team is considering a design change that affects the performance of four workloads. Comparing the performance of the new design with the existing design shows a speedup of 1.5, 1.25, 0.8, and 0.6 for these workloads.

The average speedup of the new design is: _____

Would you recommend the new design (yes/no)? _____

(II) Ignore (we have not covered Virtual Memory yet)

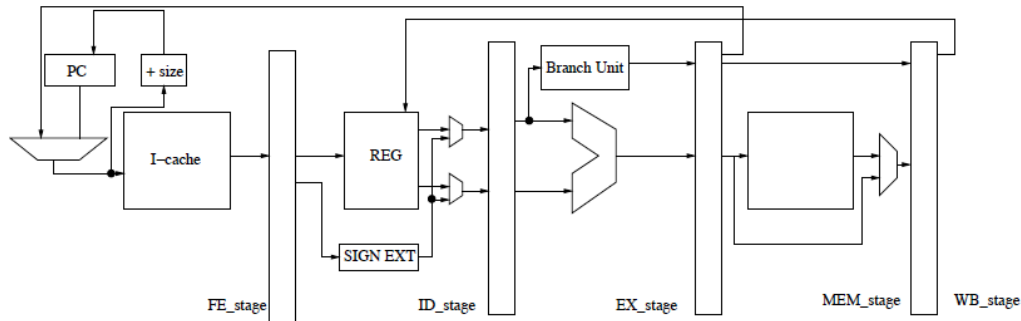
(III) Ignored (as we have not covered PCM yet)

(IV) We want to implement register renaming for a machine that has 16 architectural registers, each 8 byte wide. We are considering 256 physical registers. The total size of the Register Alias Table (RAT) for this machine will be _____ bytes. The size of the physical register file will be _____ bytes.

Name: _____

Problem 2 “pipeline operation” (16 points)

Consider the pipeline shown below, which is similar to the one you used for lab 1 and lab 2 (the PC update logic may be slightly different). A decoded branch instruction will stall the fetch stage until the branch direction is resolved. Let us define the cycle at which the instruction reaches the WB stage as the *completion cycle* of that instruction. For example, a sequence of four independent instructions A, B, C, D will have completion cycle of A at cycle 5, B at cycle 6, C at cycle 7, and D at cycle 8.



For the code snippets shown below what is the completion cycle for B, C, and D. The completion cycle for A remains cycle 5.

- I. A. ADD R3, R2, R1 ➔ cycle 5
 B. ADD R4, R3, R2 ➔ cycle ____
 C. ADD R5, R4, R3 ➔ cycle ____
 D. ADD R6, R5, R4 ➔ cycle ____

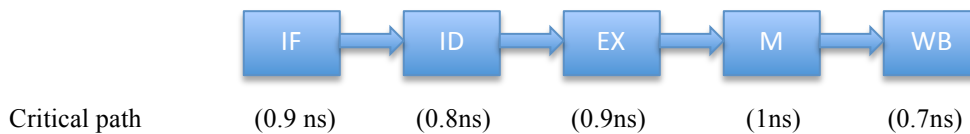
- II. A. ADD R3, R2, R1 ➔ cycle 5
 B. BR (Taken to C) ➔ cycle ____
 C. ADD R5, R4, R3 ➔ cycle ____
 D. ADD R6, R5, R4 ➔ cycle ____

- III. A. ADD R3, R2, R1 ➔ cycle 5
 B. BNEZ R3 (Taken to C) ➔ cycle ____
 C. ADD R5, R4, R3 ➔ cycle ____
 D. ADD R6, R5, R4 ➔ cycle ____

Name: _____

Problem 3 “To forward or not to forward that is the question” (20 points)

The stall time of a pipelined design can be reduced with data forwarding. We are considering data forwarding for the five-stage pipeline, similar to the one used for lab 1. Performance evaluation showed that implementing data-forwarding can reduce the Cycles Per Instruction (CPI) of the pipeline by 15% for our workloads. However, forwarding requires adding a multiplexer to either the ID stage or the EX stage. This multiplexer would add 0.4ns to the critical path of the stage. The critical path of different stages in our pipeline is shown below.



I. What is the highest frequency at which this pipeline can operate correctly?

II. If we implement data forwarding by adding the multiplexer to the ID stage, what will be the operating frequency, and overall speedup? Frequency _____ Speedup _____

III. If we implement data forwarding by adding the multiplexer to the EX stage, what will be the operating frequency, and overall speedup? Frequency _____ Speedup _____

IV. Select one recommendation from below, based on the answers above:

- A. Implement data-forwarding by adding multiplexer in ID stage
- B. Implement data-forwarding by adding multiplexer in EX stage
- C. Do not implement data forwarding

V. Will your recommendation change if the multiplexer delay was 0.2 ns? Why or why not?
The speedup of this system is: _____

Name: _____

Problem 4 “DRAM performance analysis” (20 points) --- IGNORE as we have not covered DRAM yet

Problem 5 “Demystifying the effectiveness of prefetching” --- IGNORE as we have not covered prefetching yet

Department of Electrical and Computer Engineering
Georgia Institute of Technology

ECE4100/ECE6100/CS4290/CS6290 (Section B): Advanced Computer Architecture
Moinuddin K. Qureshi, Instructor
Amit Karande and Vinson Young, TAs

Midterm 1, February 18, 2014

Name : _____

GT Account: _____

Problem 1 (25 points): _____

Problem 2 (15 points): _____

Problem 3 (20 points): _____

Problem 4 (20 points): _____

Problem 5 (20 points): _____

Total (100 points) : _____

This exam is given under the Georgia Tech Honor Code System. Anyone found to have submitted copied work instead of original work will be dealt with in full accordance with Institute policies.

Georgia Tech Honor Pledge: "I have neither given nor received aid on this exam."

[MUST sign:] _____

Note: Where needed, show all your intermediate results to receive full credit. Do all your work in this examination handout. Use the back of the exam sheets if necessary.

Note: Please be sure your name is recorded on each sheet of the exam.

GOOD LUCK!

Name: _____

Problem 1 – “Potpourri” (25 points, there are five parts, each worth 5 points)

(I) We have three workloads (X, Y, and Z) each containing the same number of instructions. The IPC for X is 1, for Y is 0.5, and for Z is 2. A new design option can change the IPC of X to 2, Y to 0.25, and Z to 4, while maintaining the same frequency as before. What is the average IPC before the design change? _____

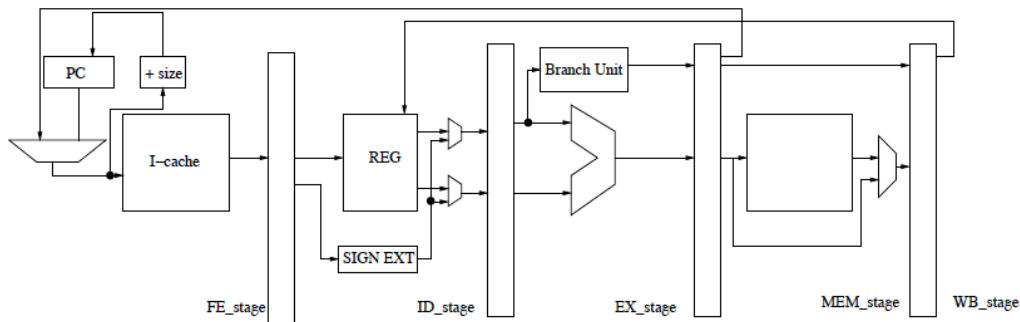
What is the average IPC after the design change? _____

Would you recommend this design change (yes/no)? _____

Name: _____

Problem 2 “pipeline operation” (15 points)

Consider the pipeline shown below, which is similar to the one you used for lab 1 and lab 2 (the PC update logic may be slightly different, and the writeback stage does not have a forwarding path to the decode stage so the decode stage must read the value only from the register file). A decoded branch instruction will stall the fetch stage until the branch direction is resolved. Let us define the cycle at which the instruction reaches the WB stage as the completion cycle of that instruction. For example, a sequence of four independent instructions A, B, C, D will have completion cycle of A at cycle 5, B at cycle 6, C at cycle 7, and D at cycle 8.



For the code snippets shown below what is the completion cycle for B, C, and D. The completion cycle for A is shown. Note that ADD/SUB takes one cycle to execute, MUL takes 2 cycles and DIV takes four cycles.

- IV. A. ADD R3, R2, R1 ➔ cycle 5
 B. MUL R4, R3, R2 ➔ cycle ____
 C. DIV R5, R4, R3 ➔ cycle ____
 D. SUB R6, R5, R4 ➔ cycle ____

- V. A. SUB R3, R2, R1 ➔ cycle 5
 B. BR (Taken to C) ➔ cycle ____
 C. ADD R5, R4, R3 ➔ cycle ____
 D. MUL R6, R5, R4 ➔ cycle ____

- VI. A. SUB R3, R2, R1 ➔ cycle 5
 B. BNEZ R3 (Taken to C) ➔ cycle ____
 C. DIV R5, R4, R3 ➔ cycle ____
 D. DIV R6, R5, R4 ➔ cycle ____

Name: _____

Problem 3 “Balanced Pipelines” (20 points)

The frequency of a pipelined machine can be increased by splitting the slowest stage. We are considering such an optimization for the five-stage pipeline, similar to the one used for lab 1. The critical path of different stages in our baseline machine M1 is shown below.

M1



Critical path (0.8 ns) (0.8ns) (1 ns) (0.6 ns) (0.7ns)

To increase the frequency of our pipeline we decide to split the EX stage into two EX1 and EX2. The critical path of each stage of this machine (M2) is shown below

M2



Critical path (0.8 ns) (0.8ns) (0.5 ns) (0.5 ns) (0.6ns) (0.7ns)

I. What is the highest frequency at which M1 can operate correctly? _____

II. What is the highest frequency at which M2 can operate correctly? _____

III. If the Cycles Per Instruction (CPI) of both M1 and M2 are similar, the speedup of M2 over M1? _____

IV. In reality, a pipeline with more stages typically has a higher CPI than a pipeline with fewer stages. Why?

V. Let the CPI of M1 be 1.25, and the CPI of M2 be 1.5. What is the speedup of M2 with respect of M1? _____

Name: _____

Problem 4 “Caching Insights” (20 points)

A tiny computer system has a 256 byte physically addressed cache, with a linesize of 16 bytes.

The following loop is executed on this machine:

```
sum=0;

for(i=0; i<256; i++){
    sum += (X[i] + Y[i]*Z[i]);
}
```

The starting addresses of the array in physical memory are as follows:

X: x2000

Y: x6180

Z: xC280

All 256 elements of each array are contiguous in physical memory. Each element in the each array is stored as a 16-bit word. Assume that **sum** and **i** stored in registers.

I) Calculate the cache hit ratio for the loop if the cache is direct mapped. _____

II) Calculate the cache hit ratio for the loop if the cache is two way set associative. _____

III) What is the minimum associativity of the cache that will yield an improvement in hit rate, assuming that the cache capacity remains at 256B? Please be brief and specific

IV) What is the improved hit rate with the new cache configuration? _____

Note: Problem 5 is omitted, as we have not covered DRAM systems yet.

School of Electrical and Computer Engineering
Georgia Institute of Technology

ECE4100/ECE6100 (Section A,Q): Advanced Computer Architecture
Moinuddin K. Qureshi, Instructor

Midterm 1, February 12, 2015

Name : _____

GT Account: _____

Problem 1 (25 points): _____

Problem 2 (15 points): _____

Problem 3 (15 points): _____

Problem 4 (15 points): _____

Problem 5 (15 points): _____

Problem 6 (15 points): _____

Total (100 points) : _____

This exam is given under the Georgia Tech Honor Code System. Anyone found to have submitted copied work instead of original work will be dealt with in full accordance with Institute policies.

Georgia Tech Honor Pledge: "I have neither given nor received aid on this exam."

[MUST sign:] _____

Note: Where needed, show all your intermediate results to receive full credit. Do all your work in this examination handout. Use the back of the exam sheets if necessary.

Note: Please be sure your name is recorded on each sheet of the exam.

GOOD LUCK!

Name: _____

Problem 1 – “Potpourri” (25 points, there are five parts, each worth 5 points)

(A) You are considering a micro-architectural change that will yield different speedups for four different benchmarks. The speedup for the four benchmarks is as follows: 1.1x, 1.2x, 1.3x, 0.5x.

i. What is the average speedup? _____

ii. Would you implement the technique (yes/no)? _____

(C) The accuracy of branch predictors can be increased by a hybrid approach, which combines multiple predictors. We have a base predictor B that has an accuracy of 80%. We have a choice of combining it with either predictor X or predictor Y.

Predictor X has an accuracy of 60%, however half of the time when B is incorrect X is correct.

Predictor Y has an accuracy of 80%, however quarter of the time when B is incorrect Y is correct.

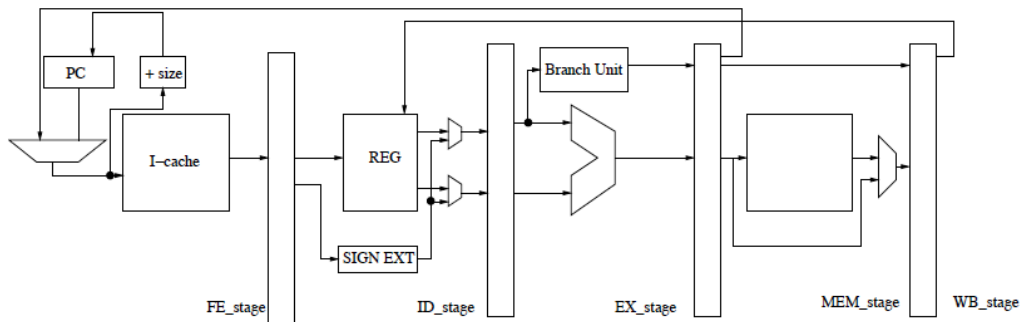
Which predictor (if any) should be combined with predictor B? ____

What is the branch prediction accuracy of this combination (assuming perfect meta predictor)? ____

Name: _____

Problem 2 “Pipeline Operation” (15 points)

Consider the pipeline shown below, which is similar to the one you used for Lab 2 (the PC update logic may be different). **The register file is written in the first half of the clock cycle and can be read in the second half of the clock cycle.** There is no forwarding from the MEM/EX stage to the ID stage. A decoded branch instruction will stall the fetch stage until the branch direction is resolved. Let us define the cycle at which the instruction reaches the WB stage as the completion cycle of that instruction. For example, a sequence of four independent instructions A, B, C, D will have completion cycle of A at cycle 5, B at cycle 6, C at cycle 7, and D at cycle 8.



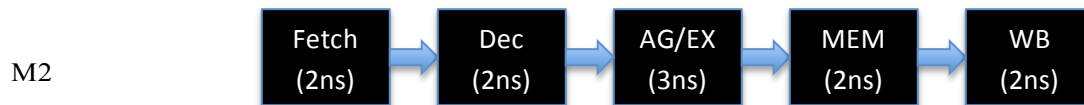
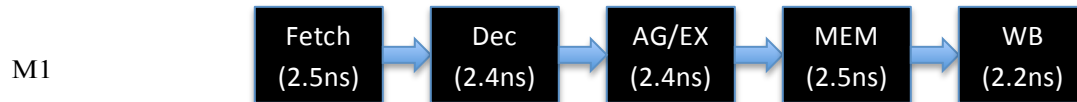
For the code snippet shown below what is the completion cycle for all the instructions. The completion cycle for A is already provided. Note that ADD/SUB takes one cycle to execute, and MUL takes 2 cycles. Note, BRnp instruction is a branch instruction, which reads the condition code (we assume that when this branch resolves, it gets taken to instruction F).

- | | | | |
|----|-------------------|---|------------|
| A. | ADD R3, R2, R1 | → | cycle 5 |
| B. | MUL R4, R3, R2 | → | cycle ____ |
| C. | SUB R6, R5, R4 | → | cycle ____ |
| D. | BRnp (Taken to F) | → | cycle ____ |
| E. | ADD R0, R0, R0 | → | cycle ____ |
| F. | SUB R0, R0, R1 | → | cycle ____ |

Name: _____

Problem 3 “Balanced Pipelines” (12 points)

Assume a single cycle non-pipelined implementation of a processor that operates at a clock frequency of 100 MHz (clock time of 10 ns). Your job is to convert this design into a five stage pipelined implementation. You are considering two designs: M1 and M2. Shown below are the critical paths for each of the five stages for these two designs.



A) What is the clock frequency at which M1 can operate?

B) What is the clock frequency at which M2 can operate?

C) Which design would you recommend?

D) For the recommended design, what is the expected speedup compared to the non pipelined implementation. Assume that the CPI of the pipelined machine is 1 for this example.

Name: _____

Problem 4 “Caching Insights” (15 points)

You are in charge of designing the cache hierarchy of an upcoming chip for PiePhone 6.0. Your design team is provided with two cache designs. First a small 32 KB cache that provides a hit rate of 90% **for the data stream generated by the processor**. Second a large 1 MB cache that provides a hit rate of 98% **for the data stream generated by the processor**.

The access latency for the 32KB cache is 1 cycle, for the 1MB cache is 10 cycles, and accessing the main memory is 100 cycles.

You are considering three design points:

- X. Use only one level of cache that is small and fast (32 KB)
- Y. Use only one level of cache that is large and slow (1MB)
- Z. Use a two level structure that has 32KB cache as L1 and 1MB cache as L2.

We want a cache structure that will provide the data with lowest latency on average, so we will use Average Memory Access Time (AMAT) as the figure of merit.

A. What is the AMAT for design X?

B. What is the AMAT for design Y?

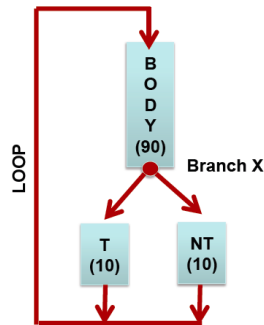
C. What is the AMAT for design Z?

D. Which design would you recommend? _____

Name: _____

Problem 5 “Predication versus Prediction” (15 points)

Figure below shows a control-flow diagram with a difficult-to-predict branch, “Branch X”. Before X, there is a part of the loop body containing 90 instructions. Furthermore, both the “Taken (T)” and “Not-Taken (NT)” path of X contains 10 instructions each, after which the control flow converges. This code gets executed a large number of times. To mitigate performance penalty of branch mispredictions, we are considering predicated execution for X. Recall that, with predicated execution, both T and NT paths will be executed regardless of the branch direction.



We will execute this code on a 4-wide processor with 25-stage deep pipeline, so the branch misprediction penalty is approximately 100 instructions.

A. If branch prediction accuracy for branch X is 80%, should we use predication or branch prediction? Why (show calculations)? [Hint: Estimate performance penalty of X as the expected number of wasted instructions per loop]

B. If we use state-of-the-art branch predictor that improves the branch prediction accuracy of Branch X to 95%, should we use predication or branch prediction? Why (show calculations)?

(Problem 6 is not covered yet)

School of Electrical and Computer Engineering
Georgia Institute of Technology

ECE4100/ECE6100 (Section A,Q) and CS4290/CS6290: Advanced Computer Architecture
Moinuddin K. Qureshi, Instructor

Midterm 1, October 3, 2017

Name : _____

GT Account: _____

Problem 1 (25 points): _____

Problem 2 (15 points): _____

Problem 3 (15 points): _____

Problem 4 (15 points): _____

Problem 5 (15 points): _____

Problem 6 (15 points): _____

Total (100 points) : _____

This exam is given under the Georgia Tech Honor Code System. Anyone found to have submitted copied work instead of original work will be dealt with in full accordance with Institute policies.

Georgia Tech Honor Pledge: "I have neither given nor received aid on this exam."

[MUST sign:] _____

Note: Where needed, show all your intermediate results to receive full credit. Do all your work in this examination handout. Use the back of the exam sheets if necessary.

Note: Please be sure your name is recorded on each sheet of the exam.

GOOD LUCK!

Name: _____

Problem 1 – “Potpourri” (25 points, there are five parts, each worth 5 points)

(A) We have three workloads (X, Y, and Z) each containing the same number of instructions. The IPC for X is 1, for Y is 2, and for Z is 3. A new design can change the IPC of X to 0.5, Y to 3, and Z to 6, while having the same frequency.

What is the average IPC before the design change? _____

What is the average IPC after the design change? _____

Would you recommend this design change (yes/no)? _____

(D) What impact does Loop Unrolling typically has on the following aspects of a program execution (circle one, increase or decrease)

Number of Static Instructions (Code Footprint):	Increase/Decrease
Number of Dynamic Instructions:	Increase/Decrease
Number of Times Branch Predictor is consulted:	Increase/Decrease
Hit Rate of the Instruction-Cache:	Increase/Decrease

(E) Write-back and Write-through are two policies for a cache.

Which policy requires larger tag-store overhead? _____ Why (less than five words)?

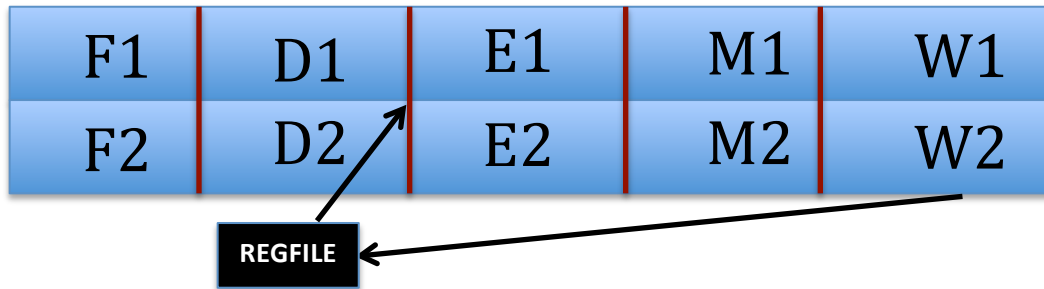
--

The disadvantage of write-through policy is that it requires higher _____

Name: _____

Problem 2 “Pipeline Operation” (15 points)

Consider the two-wide pipeline shown below, similar to the one you used for Lab 2. **The register file is written in the first half of the clock cycle and can be read in the second half of the clock cycle.** There is no forwarding from the MEM/EX stage to the ID stage. Let us define the cycle at which the instruction reaches the WB stage as the completion cycle of that instruction. For example, a sequence of four independent instructions A, B, C, D will have completion cycle for A&B at cycle 5, and C&D at cycle 6.



For the code snippet shown below what is the completion cycle for all the instructions. The completion cycle for A is already provided. Note that ADD/SUB takes one cycle to execute, and MUL takes 2 cycles.

- | | | | |
|----|----------------|---|------------|
| A. | ADD R3, R0, R0 | ➔ | cycle 5 |
| B. | SUB R4, R1, R1 | ➔ | cycle ____ |
| C. | MUL R5, R3, R4 | ➔ | cycle ____ |
| D. | ADD R6, R5, R0 | ➔ | cycle ____ |

CYCLE	FETCH F1, F2	DECODE D1, D2	EXECUTE E1, E2	MEMORY M1, M2	WRITEBACK W1, W2
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Name: _____

Problem 3 “Evaluating Performance” (15 points)

You are considering purchase of one of three machines P, Q, and R for doing neural-network simulations. To compare performance, you came up with a representative code kernel and compiled it for these three machines. Shown below are the number of instructions, Cycles Per Instruction (CPI) and frequency of these three machines for your kernel.

Metric	Machine P	Machine Q	Machine R
Instructions	110 Billion	100 Billion	90 Billion
CPI	1	2	4
Frequency	1 GHz	2 GHz	3 GHz

A) Compute the MIPS (Million Instructions per Second) for these machines

B) Compute the Execution time (in seconds) for these machines

C) Which machine would you recommend?

Name: _____

Problem 4 “Caching Insights” (15 points)

Two students A and B are asked to write a program to obtain the number of students who have scored more than K times the class average. The student record is stored as follows:

#CLASS_SIZE 2048

```
Struct StudentInfo{
    char[60]    name ; /* too bad if the student's name is more than 60 characters */
    unsigned int score ; /* a four byte value */
} student[CLASS_SIZE];
```

Coded by A

```
Num_Students_Greater_than_Avg(K) {
    sum=0;

    for(ii=0; ii<CLASS_SIZE; ii++){
        sum += student[ii].score
    }

    avg = sum/CLASS_SIZE;
    retval=0;

    for(ii=0; ii<CLASS_SIZE; ii++){
        if(student[ii].score > k*avg)
            retval++;
    }

    return retval;
}
```

Coded by B

```
Num_Students_Greater_than_Avg(K) {
    sum=0;

    for(ii=0; ii<CLASS_SIZE; ii++){
        sum += student[ii].score
    }

    avg = sum/CLASS_SIZE;
    retval=0;

    for(ii=CLASS_SIZE-1; ii>=0; ii--){
        if(student[ii].score > k*avg)
            retval++;
    }

    return retval;
}
```

The above codes are to be run on a machine with 64KB LRU-managed data cache with a line-size of 64 bytes.

1. What is the size of the total data structures student[CLASS_SIZE] ?
2. How many misses will be caused by one execution of code shown on left (coded by A)? For this exercise, assume that the cache is cold (no valid data) and that conflict misses are negligible.
3. How many misses will be caused by one execution of code shown on right (coded by B)? For this exercise again, assume that the cache is cold (no valid data) and that conflict misses are negligible.

Name: _____

Problem 5 “Predictor Scalability” (15 points)

In class, we studied a two-level predictor that uses a History Register (HR) that indexes a Pattern History Table (PHT) to obtain the prediction. The PHT typically stores a two-bit counter for making predictions. Both the HR and the PHT could be either global or per-branch (local). This question deals with the storage budget of different predictors.

Consider that our target machine is used for running a program that contains eight branches (B1-B8).

1. What would the total storage budget of the two-level predictor for **gAG (global HR and global PHT)** if we want to track 16 bits of history?

Bytes

2. What would the total storage budget of the two-level predictor for **pAG (local HR and global PHT)** if we want to track 16 bits of history?

Bytes

3. What would the total storage budget of the two-level predictor for **gAP (global HR and local PHT)** if we want to track 16 bits of history?

Bytes

4. What would the total storage budget of the two-level predictor for **pAP (local HR and local PHT)** if we want to track 16 bits of history?

Bytes

5. If we had a **perceptron predictor** instead of the two-level predictor, and still tracked 16-bits of history, what would be the storage budget for the **pAP (local HR and local PHT)** organization? Assume that the perceptron table stores the weights as 8-bit entities.

Bytes

(Problem 6 is omitted as we have not covered it yet)