# York University <br> lassonde School of Engineering <br> Dept. of Electrical Engineering and Computer Science Fall 2015 

| EECS4201 | Final | Computer Architecture |
| :--- | :---: | :---: |
| Dec. 10, 2015 | $2: 00-4: 00 \mathrm{pm}$ |  |


ID $\qquad$
Instructions to students:
Answer all questions.
Marks are shown in front of each question number.
Show your work

This examination consists of 7 questions

| Problem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Points | $/ 6$ | $/ 14$ | $/ 6$ | $/ 6$ | $/ 6$ | $/ 8$ | $/ 6$ | $/ 52$ |

## Question 1 (6 points)

Below is a C loop and its assembly language implementation.

```
Source Code:
for (i=1000; i>0; i--) {
w[i] = x[i] + y[i] + z[i];
}
Assembly Code:
Loop:
L.D F1, 0(R2) // Get x[i]
L.D F2, 0(R3) // Get y[i]
L.D F3, 0(R4) // Get z[i]
ADD.D F4, F2, F1 // Add two numbers
ADD.D F5, F3, F4 // Add the third number
S.D F5, 0(R5) // Store the result into w[i]
DADDUI R2, R2, #-8 // Decrement R2
DADDUI R3, R3, #-8 // Decrement R3
DADDUI R4, R4, #-8 // Decrement R4
DADDUI R5, R5, #-8 // Decrement R5
BNE R2, R1, Loop // Check if we've reached end of the loop
NOP
```

Assume floating point addition takes 3 cycles in the execution stage, the rest of the instructions are 1-cycle execution, and a standard 5-stage MIPS-like pipeline.
a) How many stall cycles per iteration without any rearranging or loop unrolling?
b) Without unrolling the loop, shuffle the instructions to minimize the stall cycles. How many stall cycle after shuffling?

Problem 2 (6+2+6 points)
a) A CPU architect is trying to come up with a good solution for average memory access time (AMAT). A number of caches are available
o L1: $32 \mathrm{~KB} \quad$ 1-cycle Miss rate 5\%
o L2: $\quad 256 \mathrm{~KB} \quad 10$-cycle Miss rate $10 \%$
o L3: $2 \mathrm{MB} \quad 40$-cycle Miss rate $20 \%$
o L4: $32 \mathrm{MB} \quad$ 80-cycle Miss rate $40 \%$
o Memory: 200-cycles
Estimate the extra stall cycles per memory access for these cache architectures i. L1-L2-L3-L4
ii. L2-L3-L4
iii. L1-L2-L4
b) A 16 MB L3 cache has a 64 byte line (block) size and is 16 -way set-associative. How many sets does the cache have?
i. How many bits are used for the offset, index, and tag, assuming that the CPU provides 48-bit addresses?
ii. How large is the tag array?
c) For the following access pattern, indicate if each access is a hit or miss. What is the hit rate? Assume that the cache has 2 sets and is 2-way set-associative. Assume the following mapping from cache block to sets Also assume LRU replacement policy.

| Cache block | A | B | C | D | E | F |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Set Number | 0 | 1 | 0 | 1 | 0 | 1 |

Access pattern: ACCBAECDBAFDCEA.
$\begin{array}{lllllllllll}\text { A } & \text { C } & \text { C } & \text { B } & \text { A } & \text { E } & \text { C } & \text { D } & \text { B } & \text { A } & \text { F }\end{array}$
D
C E
A

## Question 3(6 points)

Consider the following pipeline.

|  |  | ALUFP | ALUFP | WB |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Fetch | Decode | ALU Int | WB |  |  |

After instruction fetch, the instruction goes through "Decode" stage where dependences are analyzed. After this, an instruction takes one of three paths
o Integer, branch, jump instructions $\boldsymbol{\rightarrow}$ the middle path
o FP operation instructions $\boldsymbol{\rightarrow}$ the upper path
o Load/Store instructions $\rightarrow$ the lower path.
Assume that the register file have an infinite number of input ports, which means we can write any number of registers in the same cycle. Assume also that we can implement forwarding from any stage to any other stage

How many stalls cycles are introduced in the following cases?
a) Int-add followed by a dependent int add
b) FP add followed by a dependent store
c) Load followed by a FP operation that use the loaded value
d) FP operation followed by a store that stores the calculated value?

## Question 4 (6 points)

Assume that we have a very fast processor that takes all the area of a rather large chip. We can use the same area to design 2 processors (on the same chip), but each one will have half the performance of the original fast (and big) processor.
Now, calculate the relative speed of the 2-processor chip assuming that the speed of the fast (big) processor is 1 under the following conditions
a) The parallelizable part of the code is $30 \%$
b) The parallelizable part of the code is $60 \%$
c) The parallelizable part of the code is $90 \%$

## Problem 5 (6 Points)

Consider the following access sequence to cache blocks by 3 processors using a snoopy cache coherence protocol. Assume that blocks X and Y are mapped to the same block in the cache.
Write the state of cache blocks X and Y in every processor cache and the memory using the following notations.
$S(X)$ : Block $X$ is shared
$\operatorname{EX}(\mathrm{X})$ :means block X is exclusive
$I(X)$ : Block $X$ is invalid

|  | A | B | C | MEMORY |
| :--- | :--- | :--- | :--- | :--- |
| A: Rd X |  |  |  |  |
| B: Rd X |  |  |  |  |
| C: Rd X |  |  |  |  |
| A: Wr X |  |  |  |  |
| A: Wr X |  |  |  |  |
| C: Wr X |  |  |  |  |
| B: Rd X |  |  |  |  |
| A: Rd X |  |  |  |  |
| A: Rd Y |  |  |  |  |
| B: Wr X |  |  |  |  |
| B: Rd Y |  |  |  |  |
| B: Wr X |  |  |  |  |
| B: Wr Y |  |  |  |  |

## Problem 6 (8 points)

Consider a byte-addressable memory system that uses virtual memory, but no cache where addresses are 32-bit long.
a) For a physical address in a system that has 2048 physical pages, where the page size is 1024 words (remember that each word contains 4 bytes): compute the length of a byte address and draw a diagram to show how the bits in the physical address are divided into physical page number, page offset, and byte offset.
b) When the virtual memory system is implemented to augment the physical memory space, how many bits of the 32-bit byte address can be used for virtual page number? What will be the length of the page table for such a system?

## Problem 7 (6 Points)

Consider a DRAM where it takes 20ns to read a row into a row buffer, 20 ns to send the row from the row buffer to the ship's pins, and 20ns for pre-charge.
Write the time at which the following memory requests are completed. By completed we mean sent to the chip pins. Consider both open page and close page policy. Note that X , $\mathrm{X}+1, \mathrm{X}+2$ maps to the same row, $\mathrm{Y}, \mathrm{Y}+1, \mathrm{Y}+\ldots$ maps to a different row in the same bank

| Request | Time of arrival | Completion time |  |
| :---: | :---: | :---: | :---: |
|  |  | Open page | Closed page |
| $X$ | 0 ns |  |  |
| Y | 10 ns |  |  |
| $\mathrm{X}+1$ | 100 |  |  |
| $\mathrm{X}+2$ | 200 |  |  |
| $\mathrm{Y}+1$ | 250 |  |  |
| $\mathrm{X}+3$ | 300 |  |  |

