

Q2 Branch prediction

10 Points

We have modified the LC-2200 ISA to contain a new conditional branch instruction, **BNEQ**. When this instruction is executed, we branch to a target location if the two register operands are not equal. Assume we have a classical five-stage pipeline with all possible data forwarding paths. The pipeline is not equipped with any form of branch prediction and the ISA does not support branch delay slots. Given the following code:

```
lea $t1, var
    addi $t0, $zero, 4
loop: addi $t0, $t0, -1
    sw $t0,0($t1)
    bneq $t0, $zero, loop
    addi $v0, $zero, 0xf
var: .fill 0
```

Q2.1

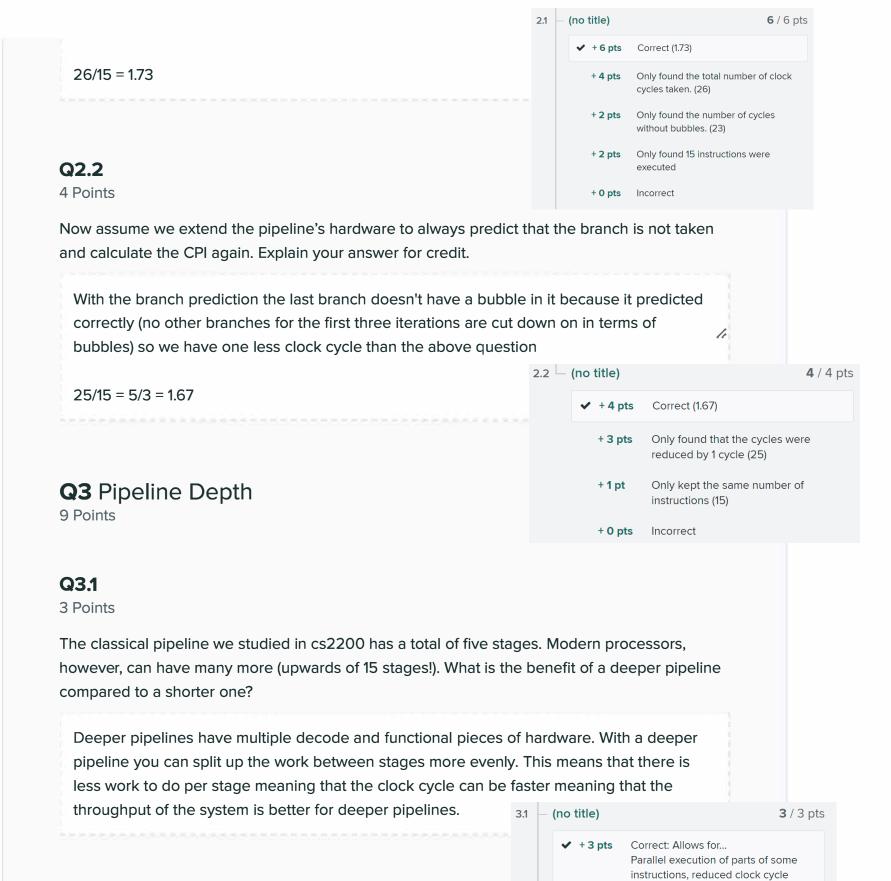
6 Points

Calculate the Cycles Per Instruction (CPI) achieved by the pipeline without any branch prediction. Explain your answer for credit.

```
CPI = Num cycles / num instruction cycles.

The number of instructions can be calculated as: 2 instructions before the loop PLUS 4 instructions in the loop TIMES 4 iterations of the loop - three instructions. i.e. 2 + 4(4) - 3 = 2 + 16 = 15

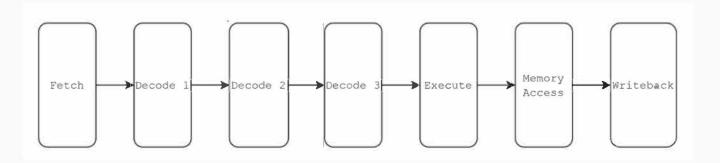
Num cycles = num instruction cycles + 4 (for all to complete) + num bubbles = 15 + 4 + 2(3) + 1(1) = 15 + 4 + 6 + 1 = 26
```



Q3.2

width, higher throughput. - Overall improved efficiency. 3 Points + 0 pts Incorrect

Say that we modified the LC-2200 pipeline to look like the following figure. Assume the pipeline uses branch prediction and a branch that is not taken is mispredicted as taken.



How many bubbles would result in the pipeline? Explain your answer.

There would be four bubbles since there are 4 stages before the EXECUTE stage. (no title) **3** / 3 pts ✓ + 3 pts Correct (4 bubble generated) + 0 pts Incorrect

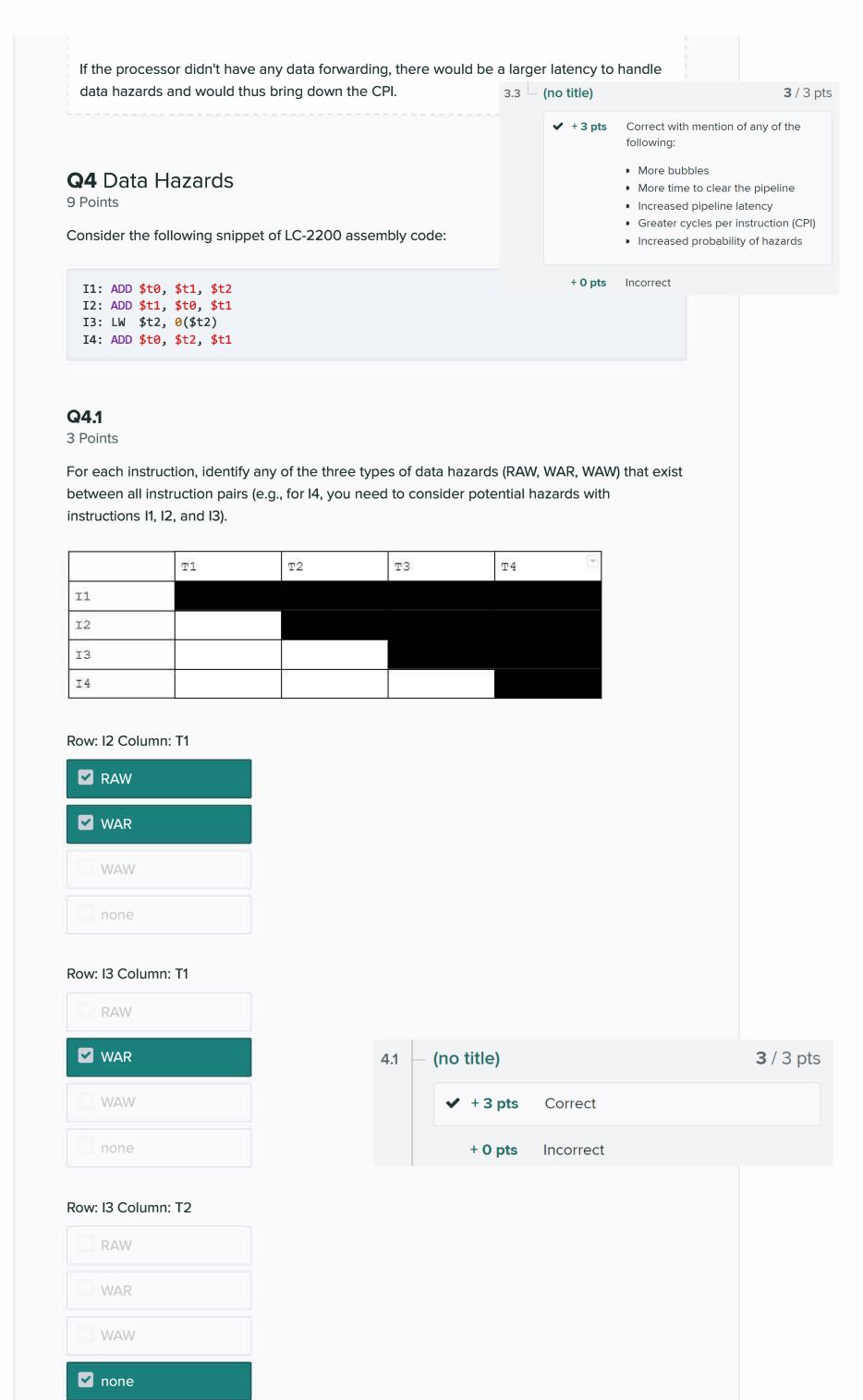
Q3.3

3 Points

Comment on a potential disadvantage of a longer pipeline on the performance of the currently executing program, other than an increased branch penalty.

It is harder to manage shared resources like registers.

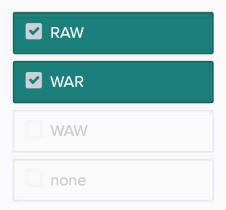
It is more costly since there is more hardware (and complexity) involved.



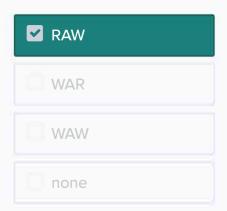
Row: I4 Column: T1



Row: I4 Column: T2



Row: I4 Column: T3



Q4.2

4 Points

Assuming the classical five-stage pipeline without any data forwarding support, count the number of bubbles that will be present upon the execution of the program. Explain your answer for credit.

Bubbles are only a consequence of RAW data hazards. Due to the RAW of I2 trying to read to before it is written in I1, this RAW hazard will result in 3 nops to correct since I1 and I2 are back to back and I2 will be in the ID/RR stage and I1 will be in the EX stage so I2 will have to wait for I1 in EX to go through both MEM and WB (3 cycles).

The other RAW hazards come as a consequence of I4. Because I4 tries to read t2 before I3 can write to it, this would also cause 3 nops since I3 and I4 are right next to each other.

There is another RAW hazard with I4 trying to read t1 which is written in I2 and would cause 2 nops already, but since the other raw hazard mentioned earlier in this paragraph is already longer than 2 nops we don't have to consider it.



Q4.5

2 Points

Count the number of bubbles with full data forwarding support. Explain your answer for credit.

With data forwarding, the only bubble we have is when we have a RAW data hazard where the writing instruction is a load instruction. The only RAW data hazard like this is due to I4 trying to read t2 before I3 can load it into t2. Because of this, when I3 is in the EX stage and I4 is in the ID/RR stage, I4 needs to value of t2 but it has not been loaded yet. Thus, a nop is

introduced to the EX stange when I4 goes to the MEM stage. Before this clock cycle is over, I4 will forward the data to I3 in the ID/RR stage so it can continue. Thus, we only need one nop.

4.3 (no title)

2 / 2 pts

We only have 1 bubble with data forwarding.

+ 2 pts Correct (1 bubble from I3)

+ 0 pts Incorrect

Q5 Branch Table Buffer

9 Points

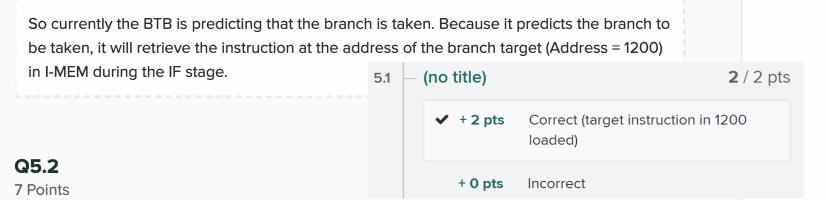
Consider a 5-stage pipelined processor (same as the one in the textbook) equipped with a branch target buffer (BTB which has only 1 bit of branch history). Assume a 32-bit word-addressable architecture. Upon misprediction, there will be a 2-cycle penalty (i.e., two NOPs). A BEQ instruction is fetched from memory location 1000 by the "IF" stage. The BTB currently has an entry for this BEQ instruction:

```
PC=1000 | branch predicted TAKEN;
PC of BEQ target = 1200;
PC of next sequential instruction = 1001
```

Q5.1

2 Points

What will be the action in IF stage in the next clock cycle?



When BEQ is in the EX stage, the outcome of the branch turns out to be NOT TAKEN. What are the actions that will ensue as a result of this outcome?

Because the branch was not taken, there are currently two instructions in our pipeline that are bad. BEQ is in the EX stage meaning that the processor predicted two bad instructions (1 is currently in the IF stage and another is in the ID/RR stage). Upon noticing that the branch is not taken, the EX stage will assert its feedback line to the ID/RR stage to tell it that it must "flush" its instruction. The ID/RR stage will relay this assertion on its feedback line to the IF stage to "flush" the IF stage's instruction. The bad instructions will then be replaced with nops (bubbles).

The BTB will be updated with the misprediction.



Q6 Scheduling

12 Points

Shown below are the duration of the CPU burst and I/O burst times for the three processes.

	P1	P2	P3
CPU	2	6	3
1/•	4	1	3

Assume processes P1, P2, and P3 are ready to run at the beginning of time. Assume each process does:

- CPU burst; first CPU burst
- I/O burst; one I/O burst
- CPU burst; second CPU burst
- Done; process execution complete**

Priorities:

- P3 highest priority
- P2 next higher priority
- P1 lowest priority

Process arrival:

- P1, P2, P3 all arrive within milliseconds of each other in this order
- all are ready to be scheduled at time 0

Shown below are the timelines of activities on the CPU and I/O for the three processes with some scheduling discipline.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
CPU	Р	1		Р3		P2	P	1	P2	P3 P2				P2									
Ю				Р	1			Р3	23						P2								

Q6.1

3 Points

What scheduling algorithm is implemented here? Explain your choice.

Because at time t = 7, P1 is given priority over P2 (which was being executed in t = 6) and, at t = 10 P3 is given priority over P2, we have the SRTF since at t = 7 P1 needs only two clock cycles to complete its CPU burst (whilst P2 needs five more) and at t = 10 P3 only needs three more clock cycles to complete its CPU burst (whilst P2 needs four more).

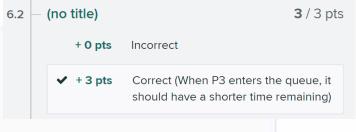


Why does the scheduler pick P3 to run on the processor at time 15?

Because we are running SRTF scheduling and at t = 10 P3 only needs three more clock cycles to complete its CPU burst (whilst P2 needs four more). Thus there are fewer clock

cycles to complete P3's CPU burst than P2's.

Correction: This question should be at time t = 10



Q6.3

3 Points

What is the wait time for P2? Explain your answer.

wait time = time not executing

wait time for P2 = 5 + 2 + 3 = 10



Q6.4

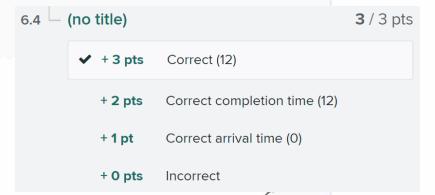
3 Points

What is the turnaround time for P3? Explain your answer.

Turnaround time = time it takes to complete a process

Turnaround time for P3 = 12

12 clock cycles

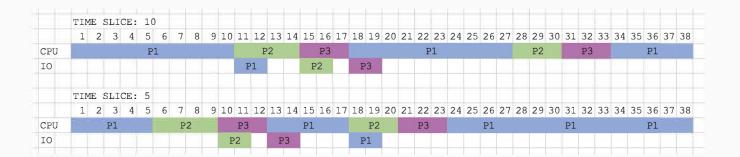


Q7 Round Robin Scheduler

17 Points

We are writing a new scheduler that will implement a round-robin algorithm in order to schedule processes on the processor. The ready queue contains three processes P1, P2, P3 in that order. The execution characteristics of the three processes are as shown in the table below.

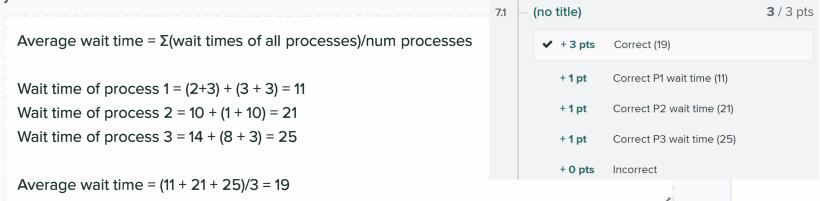
Process ID	CPU Burst	I/O Burst	CPU Burst		
P1	10	2	15		
P2	4	2	3		
Р3	3	2	3		



Q7.1

3 Points

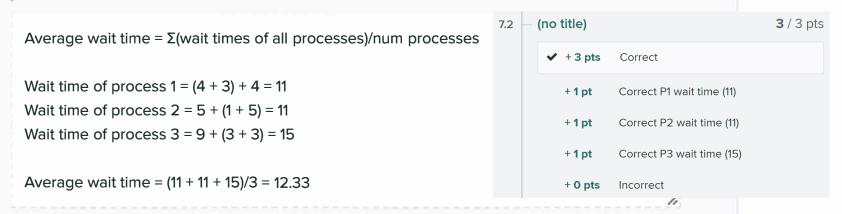
Given the above timeline with a timeslice of ten (10) time units, and assuming no scheduling or context-switching overhead, compute the average waiting time for the three processes. Show your work for credit.



Q7.2

3 Points

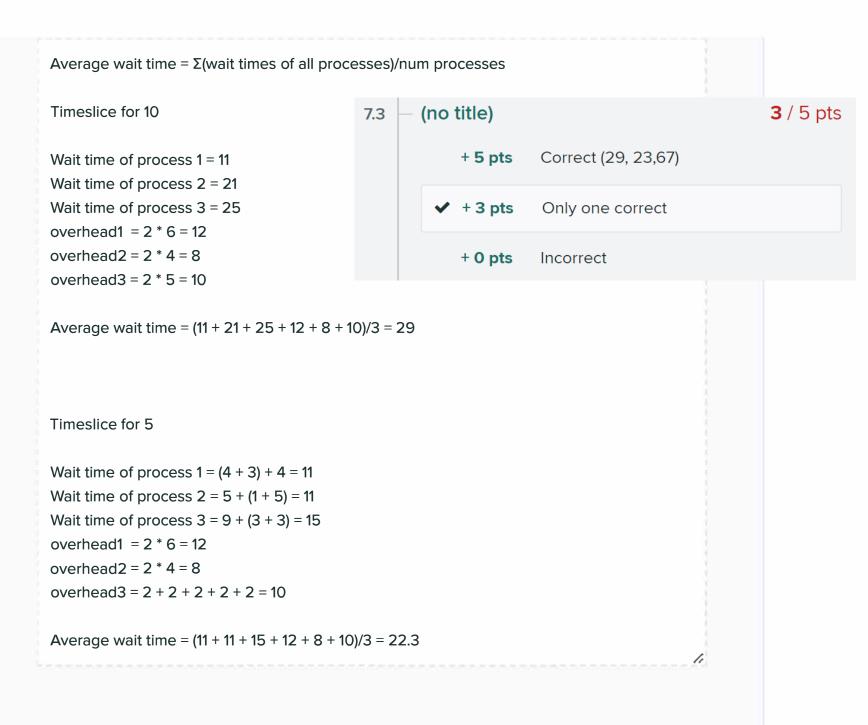
Given the above timeline with a timeslice of five (5) time units, and assuming no scheduling or context-switching overhead, compute the average waiting time for the three processes. Show your work for credit.



Q7.3

5 Points

Now assume that the scheduling and context-switching overhead (i.e., the time to pick the next process to run and dispatch it on the processor) is two (2) time units and I/O completions do not introduce any additional overheads. Compute the average waiting time for the above processes, for each of the *two timeslice* values (5 and 10). Show your work for credit.



Q7.4

3 Points

Qualitatively, state your intuition on the effect of the choice of timeslice on the waiting time for processes with short CPU bursts.

For longer processes, it doesn't seem to change the waiting time (P1 waiting time for all questions above was 11) but it seems to really drive up the time for shorter jobs when it is longer. Having a shorter timeslice give a fairer share of CPU time thus decreasing the average wait time.

7.4 (no title)

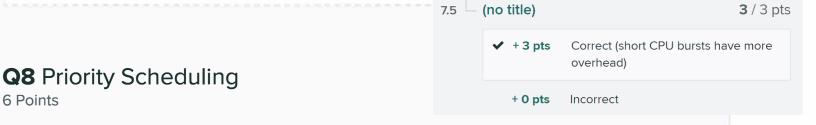
3 / 3 pts

4 + 3 pts Correct (less wait time)

3 Points

Qualitatively, state your intuition on the effect of the scheduling and context-switching overhead on the waiting time for processes with short CPU bursts.

It drives up wait time since they have to deal with switching between all the different processes. If the number of switches can be minimized we can minimize overhead.



Assume an operating system with a preemptive priority scheduler that runs two application classes A and B. When processes of type A start, they are assigned a high priority; when processes of type B start, they are assigned a low priority.

Q8.1

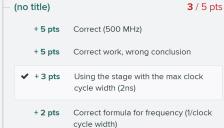
3 Points

Describe a scenario in which certain processes in our system could experience starvation.



What is the maximum clock cycle frequency this pipeline can operate at?

The clock must run at 2 ns since that is the slowest time in any stage. Each stage must be 1 clock cycle long, so despite every other stage but the MEM stage less than 2 ns, we must make the clock cycle 2 ns so that the MEM stage can finish what its doing before the next clock cycle.



+ 0 pts Incorrect

Q10.2

5 Points

We now decide to modify the pipeline to only have four stages. We must combine two stages to accomplish this. Which two stages should we combine to maintain maximum performance? Explain your answer.

We should combine the ID/RR and EX stages because they won't increase our maximum cycle time (1.75 ns vs 2 ns) and we can get the output of the REG file and then do combinational logic in EX to finish EX.



Q11 Appendix A: Useful Formulas

0 Points

Name	Notation	Units	Description
CPU Utilization	-	%	Percentage of time the CPU is busy
Throughput	n/T	Jobs/sec	System-centric metric quantifying the number of jobs <i>n</i> executed in time interval <i>T</i>
Avg. Turnaround time (t _{avg})	$(t_1+t_2+\ldots+t_n)/n$	Seconds	System-centric metric quantifying the average time it takes for a job to complete
Avg. Waiting time (w _{avg})	$((t_1-e_1) + (t_2-e_2) + \dots + (t_n-e_n))/n$ or $(w_1+w_2+\dots+w_n)/n$	Seconds	System-centric metric quantifying the average waiting time that a job experiences
Response time/ turnaround time	$ t_i $	Seconds	User-centric metric quantifying the turnaround time for a specific job <i>i</i>
Variance in Response time	$E[(t_i - e_i)^2]$	Seconds ²	User-centric metric that quantifies the statistical variance of the actual response time (t_i) experienced by a process (P_i) from the expected value (t_{avg})

Q12 Appendix B: LC-2200 ISA

0 Points

Mnemonic	Format	Opcode	Action
Example		.00	Register Transfer Language
add add Sv0, Sa0, Sa1	R	0 0000 ₂	Add contents of reg Y with contents of reg Z, store results in reg X. RTL: \$v0 ← \$a0 + \$a1
nand sv0, \$a0, \$a1	R	1 0001 ₂	Nand contents of reg Y with contents of reg Z, store results in reg X. RTL: \$v0 ← ~(\$a0 && \$a1)
addi addi \$v0, \$a0, 25	1	2 0010 ₂	Add Immediate value to the contents of reg Y and store the result in reg X. RTL: \$v0 ← \$a0 + 25
lw lw \$v0, 0x42(\$fp)	1	3 0011 ₂	Load reg X from memory. The memory address is formed by adding OFFSET to the contents of reg Y. RTL: \$v0 ← MEM[\$fp + 0x42]
sw sa0, 0x42(\$fp)	1	4 0100 ₂	Store reg X into memory. The memory address is formed by adding OFFSET to the contents of reg Y. RTL: MEM[Sfp + 0x42] ← \$a0
beq beq \$a0, \$a1, done	1	5 0101 ₂	Compare the contents of reg X and reg Y. If they are the same, then branch to the address PC+1+OFFSET, where PC is the address of the beq instruction. RTL: if(\$a0 == \$a1) PC ← PC+1+OFFSET
the OFFSET value from	m the numbe example, the	er or symb	lementer confusion), the assembler computes of given in the instruction and the assembler's stores done-(PC+1) in OFFSET so that the me.
jalr jalr Sat, Sra	J	6 01102	First store PC+1 into reg Y, where PC is the address of the jalr instruction. Then branch to the address now contained in reg X. Note that if reg X is the same as reg Y, the processor will first store PC+1 into that register, then end up branching to PC+1. RTL: \$ra ← PC+1; PC ← \$at Note that an unconditional jump can be realized using jalr \$ra, \$t0, and discarding the value stored in \$t0 by the instruction. This is why there is no separate jump instruction in LC-2200.
nop	n.a.	n.a.	Actually a pseudo instruction (i.e. the assembler will emit: add \$zero, \$zero, \$zero
halt halt	0	7 0111 ₂	assembler will cline and szero, szero, szero

Exam 2 GRADED STUDENT Eric Anders Gustafson TOTAL POINTS 85 / 103 pts QUESTION 1 Pipeline Buffer **9** / 11 pts 1.1 Instruction **1**/1pt 1.2 — OPCODE **0** / 1 pt 1.3 - RX **1** / 1 pt 1.4 RY **1**/1pt **1.5** – RZ **1**/1 pt **1.6** — Decoded RX **1**/1 pt 1.7 — Decoded RY **1**/1 pt 1.8 — Decoded RZ **1**/1 pt 1.9 — SEXT Offset **0** / 1 pt 1.10 ALU output (RY + Offset) **1** / 1 pt 1.11 MEM[address] **1**/1 pt

QUESTION 2

Branch prediction	10 / 10 pts
2.1 — (no title)	6 / 6 pts
2.2 (no title)	4 / 4 pts
QUESTION 3	
Pipeline Depth	9 / 9 pts
3.1 — (no title)	3 / 3 pts
3 2 — (no title)	3 / 3 pts
3.3 (no title)	3 / 3 pts
QUESTION 4	
Data Hazards	9 / 9 pts
41 — (no title)	3 / 3 pts
42 — (no title)	4 / 4 pts
4.3 — (no title)	2 / 2 pts
QUESTION 5	- 10 ·
Branch Table Buffer	7 / 9 pts
51 (no title)	2 / 2 pts
5.2 (no title)	5 / 7 pts
QUESTION 6 Scheduling	12 / 12 pts
6,1 (no title)	3 / 3 pts
6,2 (no title)	3 / 3 pts
6.3 — (no title)	3 / 3 pts
6.4 (no title)	3 / 3 pts
QUESTION 7 Round Robin Scheduler	15 / 17 pts
7.1 (no title)	3 / 3 pts
7.2 (no title)	3 / 3 pts
7.3 (no title)	3 / 5 pts
7.4 (no title)	3 / 3 pts
7.5 (no title)	3 / 3 pts
QUESTION 8	
Priority Scheduling	6 / 6 pts
81 — (no title)	3 / 3 pts
8.2 (no title)	3 / 3 pts
QUESTION 9	
Hidden Question	0 / 10 pts
9.1 (no title)	0 / 5 pts
9.2 — (no title)	0 / 5 pts
QUESTION 10	
Hidden Question	8 / 10 pts
10.1 — (no title)	3 / 5 pts
+ 5 pts Correct (500 MHz) + 5 pts Correct work, wrong conclusion	
 ✓ +3 pts Using the stage with the max clock cycle width (2ns) +3 pts Correct formula for frequency (1/clock cycle width) 	
+ 2 pts Correct formula for frequency (1/clock cycle width) + 0 pts Incorrect	
10.2 (no title)	5 / 5 pts
	3 / 5 pts
Appendix A: Useful Formulas	0 / 0 pts
	0 / 0 pts

Appendix B: LC-2200 ISA 0 / 0 pts