	Prism ID:
Name:	GTID#: 9

Problem	(Points, Min)	Lost	Gained	Running Total	TA
1	(1, Omin)				
2	(15, 10min)				
3	(14, 10min)				
4	(15, 10min)				
5	(14, 10min)				
6	(16, 10min)				
7	(12, 15min)				
8	(13, 15min)				
Total	(100, 80min)				

- You may ask for clarification but you are ultimately responsible for the answer you write on the paper.
- Illegible answers are wrong answers.
- Please do not discuss this test by any means (until 5 pm today)
- Please look through the entire test before starting. WE MEAN IT!!!

#### Illegible answers are wrong answers.

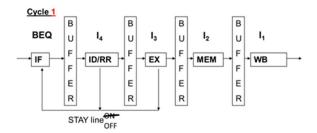
Good luck!

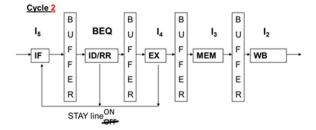
- 1. (1 point, 0 min) (circle one; you get a point regardless of  $correct/incorrect\ answer)$
- "It's tough to make predictions, especially about the future."
- (a) Said Gubernatorial candidate Jason Carter on his chances of winning the election
- (b) Said a famous computer scientist about choosing a victim page for replacement
- (c) Said President Obama about Obamacare surviving after the elections
- (d) Said Mahatma Gandhi about British rule in India prior to Independence
- (e) Said Yogi Berra about life in general

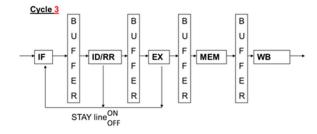
	Prism ID:
Name:	GTID#: 9

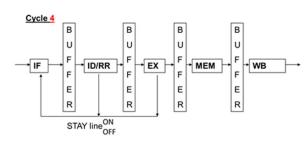
#### Pipelining

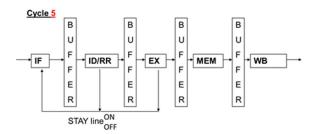
- 2. (15 points, 10 min)
- (a) (5 points)consider a pipelined processor that uses **conservative handling of branch instruction**. For this question, assume there are NO data or control or structural hazards among any of the instructions shown in flight or that would follow the branch instruction. Let  $\mathbf{I_t}, \mathbf{I_{t+1}}, \dots$  be the sequence of instructions from the target of the BEQ instruction. The "STAY line" can be turned ON or OFF by the ID/RR or the EX stage, to signal to the IF stage to NOT send new instructions down the pipeline (in this case, the IF stage will send NOP instructions to the next stage). We have shown what happens in the first two cycles. Assuming that the branch is taken, for EACH subsequent cycle show
  - Which instructions are in the different stages of the pipeline
  - The state (ON/OFF) of the "STAY line"







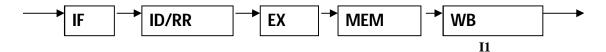




	Prism ID:
Name:	GTID#: 9

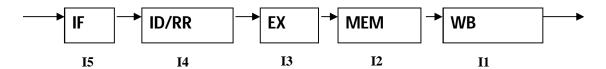
- (b) Consider the following three consecutive instructions in a program in flight in a pipelined processor with **register forwarding**:
- I1: R1 ← R2 + R3
- I2: R1 ← R1 + R4
- I3: R5 ← R1 + R6

The state of the pipeline is:



- (i) (2 points) Show in the above figure, where I2 and I3 are.
- (ii) (3 points) Why are I2 and I3 where you have shown them to be?

(c) (5 points) There are 5 instructions in flight as shown below. An external interrupt can occur asynchronously at any time.



Suggest a scheme for dealing with the external interrupt. Your answer should clearly state what happens to the instructions already in the above pipeline, and where the original program will be restarted when it is resumed.

	Prism ID:
Name:	GTID#: 9
Process Scheduling	
3. (14 points, 10 min)	
(a) (2 points) (Select one correct choice)	
One of the following is NOT part of the state of	of a running program
1. General Purpose Registers that are visib	le to the instruction set
2. Program counter and the register that rep	presents the stack pointer
3. Layout of the program in memory	
4. Priority information	
5. Internal registers in the datapath of the	e processor

### (b) (2 points) (**True/False with Justification**) "The role of the dispatcher is to select a

"The role of the dispatcher is to select a new process for running on the processor."

#### (c) (2 points) (True/False with Justification)

"A non-preemptive scheduling algorithm will schedule a different process to run **ONLY** when the current one terminates."

#### (c) (2 points) (True/False with Justification)

"Upon context switch, the scheduler saves the volatile state of the current process in the system stack."

#### (d) (2 points) (True/False with Justification)

"If you are the system administrator at a google data center, your raise most likely depends on the how good the CPU utilization of the servers are."

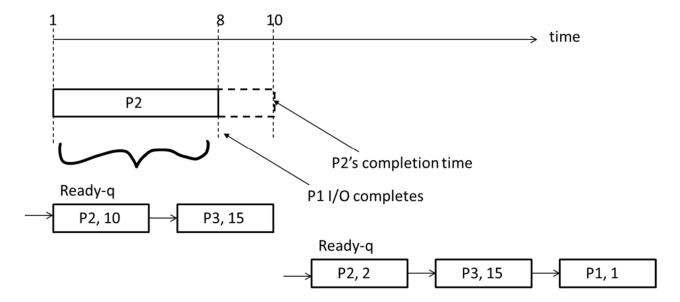
#### (e) (2 points) (True/False with Justification)

"SJF will never result in starvation."

#### (f) (2 points) (True/False with Justification)

	Prism ID:
Name:	GTID#: 9

- 4. (15 points, 10 min)
- (a) (5 points) The scheduling discipline in use is Shortest Remaining Time First (SRTF). Given the timeline below:

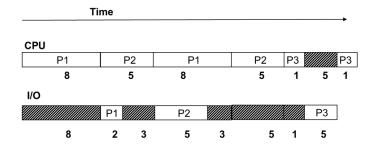


- P2 is executing currently; (Ready queue as shown on left)
- P1 completes I/O at time = 8;
- Ready queue changes as shown on the right.

Explain what (if anything) will happen as a result of this change in the ready-queue?

	Prism ID:
Name:	GTID#: 9

(b) There are three processes all of which arrive in the order P1, P2, P3. All of them are ready to be scheduled at **time 0**. Given the schedule below:

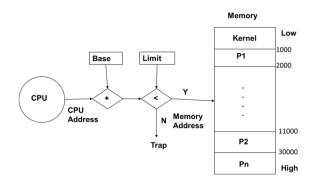


- (i) (2 points) What type of scheduler will result in the above schedule?
- (ii) (2 points) What is the turnaround time of P1?
- (ii) (2 points) What is the wait time of P2?
- (iii) (2 points) What is the observed throughput of the system?
- (c) (2 points) (**True/False with Justification**) "Linux scheduler does its work in O(1) time."

	Prism ID:
Name:	GTID#: 9

#### Memory Management and Virtual Memory (Note: K = 1024)

- 5. (14 points, 5 min)
- (a) (2 points) A given architecture uses "Base" and "Limit" registers for memory protection of individual processes. P1 is currently executing on the processor



What are the contents of Base and Limit registers?

(b) Consider the following allocation table with fixed-size partition memory allocation.

Memory

# Allocation Table Occupied bit Partition Size Process 1 6K P2 (need 3K) 1 10K P1 (need 6K)

- (i) (2 points) How much is the total internal fragmentation?
- (ii) (2 points) How much is the external fragmentation?

	Prism ID:
Name:	GTID#: 9

#### (c) (2 points)(True/False with Justification)

"Variable size partition" memory allocation suffers from internal fragmentation."

#### (d) (3 points)

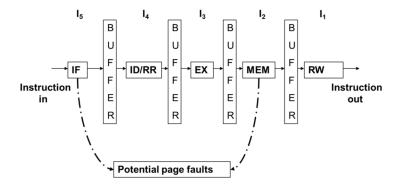
Virtual address is 64 bits; pagesize 8 Kbytes; How many entries are there in the page table? (show your work)

#### (e) (3 points)

For the same memory system as in (d), the physical address is 48 bits. How many physical page frames does the memory system have? (show your work)

#### Demand paging, Working set, page replacement

- 6. (16 points, 15 min)
- (a) Consider a 5-stage piplelined processor, which uses demand-paged virtual memory management. Instruction I2 incurs a page fault.



(i) (2 points) What will happen to the instructions in flight in the processor?

	Prism ID:
Name:	GTID#: 9

(ii) (2 points) At which instruction will the program be resumed?

(iii) (2 points) Given VPN = 0x501 maps to PFN = 0x11E0, what is the physical address corresponding to the virtual address 0x5010F8E0

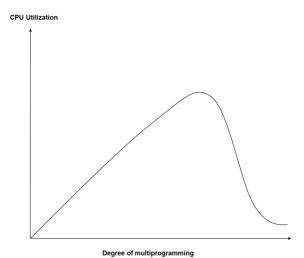
```
(b) (2 points) Working set of Process1 = \{p1, p22, p53\}; working set of Process2 = \{p1, p2, p53, p104\}, where pi refers to virtual page "i" of a given process.
```

What is the current memory pressure of the system?

#### (c) (2 points) (True/False with Justification)

"Clock sweep is just another name for FIFO page replacement algorithm."

#### (d) (2 points) (True/False with Justification)

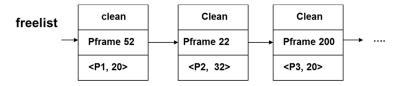


"In the picture above, the reason for CPU utilization decrease as we increase the degree of multiprogramming beyond a certain point is due to all processes alternating between CPU bursts and I/O bursts."

	Prism ID:
Name:	GTID#: 9

#### (e) (4 points)

The freelist (as shown below) is a data structure of the memory manager that contains the pool of page frames that are available for allocation to satisfy page faults.



Each entry in the freelist, shows the page frame number that is available for allocation as well as the reverse mapping, i.e., the process-ID and the VPN of that process that was hosted in this page frame.

Process P3 is currently executing on the processor. It incurs a page fault and the faulting VPN is 20.

Explain what the memory manager would do to service this page fault.

#### TLB, Cache/Memory, Execution time, Cache Design (Note: K = 1024)

- 7. (12 points, 15 min)
- (a) (2 points) (True/False with Justification)

"The principle of **Spatial locality** implies that once brought into the cache, we should keep the data around as long as possible."

#### (b) (2 points)(True/False with Justification)

#### (c) (2 points) (True/False with Justification)

"The motivation for **virtually indexed physically tagged cache** is to parallelize the lookup of the TLB and the first-level cache."

	Prism ID:
Name:	GTID#: 9

(d) (6 points)

In a pipelined processor

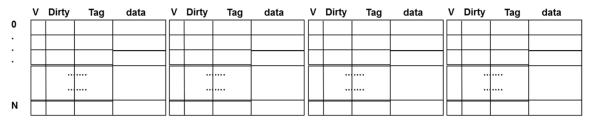
- average CPI = 1.8 without accounting for memory stalls.
- I-Cache has a **hit rate** of 99%
- D-Cache has a hit rate of 99%
- Memory reference instructions account for 25% of all the instructions executed
- Out of these memory reference instructions 80% are loads and 20% are stores.
- Read-miss penalty (for instruction or data) is 25 cycles
- Write-miss penalty is 4 cycles.

Compute the effective CPI of the processor accounting for the memory stalls.

	Prism ID:
Name:	GTID#: 9

8. (13 points, 15 min)

Consider a 4-way set associative cache with the following parameters:



- Cache size (i.e, the amount of actual data it can hold) of 256 Kbytes
- 64-bit byte-addressable memory
- Each memory word contains 8 bytes
- cache block size is **64 bytes**.
- write-back policy. One dirty bit per word.
- one valid bit per block.

(i) (4 points) How big is N in the above picture (show your work)?

(ii) (6 points) The 62-bit memory address is split into block offset, tag, and index as shown below:



What are the values of t, n, and b (show your work)?

(iii) (3 points) How many meta-data bits in  $\mathbf{each}$  of the four parallel caches shown in the above picture? (show your work)

(Note: Remember that meta-data refers to all the other stuff (besides the data itself) that goes into cache.)

	Prism ID:
Name:	GTID#: 9

#### Acronyms

IF - instruction fetch (fetch into IR and increment PC)
ID/RR - instruction decode/register read (read register contents)
EX - execute (perform arithmetic/logic/address computation - maybe)
MEM - memory (fetch/store memory operand - maybe)
WB - write back to register file (may be)
TLB - Translation Look-aside Buffer
PTE - Page Table Entry
PFN - Physical Frame Number
VPN - Virtual Page Number
RLT - Reverse Lookup Table
PCB - Process Control Block
FCFS - First Come First Served
SJF - Shortest Job First
SRTF - Shortest Remaining Time First

Name	Notation	Units	Description
CPU Utilization	-	%	Percentage of time the CPU is busy
Throughput	n/T	Jobs/s	System-centric metric quantifying the number of jobs <i>n</i> executed in time interval <i>T</i>
Avg. Turnaround time (t <sub>avg</sub> )	$(t_1+t_2+\dots+t_n)/n$	Secs	System-centric metric quantifying the average time it takes for a job to complete
time (w <sub>avg</sub> )	(w <sub>1</sub> +w <sub>2</sub> + +w <sub>n</sub> )/ n	Secs	System-centric metric quantifying the average waiting time that a job experiences
Response time	$\mathbf{t}_i$	Secs	User-centric metric quantifying the turnaround time for a specific job <i>I</i>
Variance in Response time	$\begin{bmatrix} \mathbf{E}[(\mathbf{t}_i - \mathbf{t}_{avg})^2] \end{bmatrix}$	Secs <sup>2</sup>	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 $(t_{avg})$
Starvation	-	-	User-centric qualitative metric that signifies denial of service to a particular process or a set of processes due to some intrinsic property of the scheduler
<b>Convoy</b> effect	-	-	User-centric qualitative metric that results in a detrimental effect to some set of processes due to some intrinsic property of the scheduler