Quiz 3 - Solution

Open book and note

Fall 2023

Notes:

- *You know these already but let's review it again!*
- You are allowed to access any software, tools, online material, notes, and books during the quiz. But you are not allowed to text, chat, or post anything during the quiz. You are also not allowed to use AI generation tools (LLMs, etc.).
- If you have any questions, the preferred method of communication is to DM me on Campuswire, but you can also raise your hand.
- You have 75 minutes to complete the quiz (until 1:30 PM).
- You can write your answers on a laptop, tablet, or paper.
- Once you are finished, please upload your answers to Gradescope. You have an additional 10 minutes to upload your files. If you upload it after 1:30 PM, on Gradescope, it will show it as late, but you won't lose any points. Note that you won't be allowed to upload anything after 1:40 PM.
- If your internet doesn't work after a few tries, you can hand in your paper answers. If you wrote down your results in PDF, just take a screenshot and timestamp your file. We will let you upload it later.
- Good luck! *It was a pleasure!*

Caption: Another mandatory cat photo!
[or you after week 10 :D]

Q1- **Cache Coherency:** *'There is nothing private between us! or is it?'*

Suppose we have a private 2-entry fully associative L1 cache with **pLRU** replacement policy and MOESIF coherence protocol for three cores as shown below. For each core, the initial state of each line is shown in the table. Assume that the pLRU flag bit for all lines is initially zero.

P1

Line	Coherency State	Tag
0	I	100
1	S	120

P2

Line	Coherency State	Tag
0	I	100
1	S	120

P3

Line	Coherency State	Tag
0	M	100
1	S	120

a) If we have the following sequence of accesses, complete the table after each access. (You can also show the rows that have changed).

```
P2: read <100>
P2: write <120>
P2: read <140>
P1: write <120>
P1: read <140>
P3: write <140>
P3: read <120>
```

Solution:

`P2: read <100>`

P1

Line	Coherency State	Tag
0	I	100
1	S	120

P2

Line	Coherency State	Tag
0	S	100 (1)
1	S	120

P3

Line	Coherency State	Tag
0	O	100
1	S	120

`P2: write <120>`

P1

Line	Coherency State	Tag
0	I	100
1	I	120

P2

Line	Coherency State	Tag
0	S	100 (1)
1	M	120 (1)

P3

Line	Coherency State	Tag
0	O	100
1	I	120

`P2: read <140>`

P1

Line	Coherency State	Tag
0	I	100
1	I	120

P2

Line	Coherency State	Tag
0	E	140 (1)
1	M	120

P3

Line	Coherency State	Tag
0	O	100
1	I	120

`P1: write <120>`

P1

Line	Coherency State	Tag
0	I	100
1	M	120 (1)

P2

Line	Coherency State	Tag
0	E	140 (1)
1	I	120

P3

Line	Coherency State	Tag
0	O	100
1	I	120

`P1: read <140>`

P1

Line	Coherency State	Tag
0	S	140 (0)
1	M	120 (0)

P2

Line	Coherency State	Tag
0	F	140 (1)
1	I	120

P3

Line	Coherency State	Tag
0	O	100
1	I	120

`P3: write <140>`

P1

Line	Coherency State	Tag
0	I	140 (0)
1	M	120 (0)

P2

Line	Coherency State	Tag
0	I	140 (1)
1	I	120

P3

Line	Coherency State	Tag
0	O	100
1*	M	140 (1)

`P3: read <120>`

P1

Line	Coherency State	Tag
0	I	140 (0)
1	O	120 (0)

P2

Line	Coherency State	Tag
0	I	140 (1)
1	I	120

P3

Line	Coherency State	Tag
0	S	120 (1)
1	M	140 (1)

* Invalid line is evicted.

b) Assume that every miss causes 100 stall cycles and every write-back causes 5 stall cycles (due to snooping). What is the total number of stall cycles for P3? (show your work)

Solution: (we assume cache-to-cache transfer is not a cache miss)

P3 stall cycles = 0×100 (no miss since P1 forwards) + 1×5 (write-back <100>) = 5

(Considering 1 miss has partial credit. More than one miss has no partial credit.)

c) If we use the TSO consistency model, what are the possible values for <140> in P1? Assume that <140> memory is initially zero and write <140>, writes x=1 to address 140.

Solution:

T1.1: write <120> **T1.2: read <140>**	**T2.1: read <100>** **T2.2: write <120>** **T2.3: read <140>**	**T3.1: write <140>** **T3.2: read <120>**

T1.2 < T3.1 => 0

T3.1 < T1.2 => 1

Possible values: 0, 1

Q2- **Virtual Memory:** *'I believe this quiz is simple, or could it be merely an illusion?!'*

Assume we have a design where we have an 8-way set-associative L1 cache with 64 sets and a TLB with 8 entries (no L2 cache). The TLB is fully associative. The virtual address space is 4GB, the page size is 4KB, and the cache block size is 4 bytes. Each way contains one block. The physical address space is 256MB. Assume that we store no flags and/or metadata for each page.

Also, assume that the hit time and miss rate for the memory hierarchy are as follows.

Unit	Hit Time (cycles)	Miss Rate
TLB	1	.01%
Cache	1	3%
MM	150	.0001%
Disk	10,000	0%

Answer the following questions (*show your work for each part*).

a) What is the size of TLB?

Solution:

Virtual memory size: 4 GB = 2^{32} bytes => Virtual address: 32 bits

Physical memory size: 256 MB = 2^{28} bytes => Physical address: 28 bits

Page size: 4 KB = 2^{12} bytes => Page offset: 12 bits

 # Virtual pages = $2^{32}/2^{12} = 2^{20}$ (VPN = 20 bits)

 # Physical pages = $2^{28}/2^{12} = 2^{16}$ (PPN = 16 bits)

TLB size = entries×(VPN+PPN) = 8×(20+16) = 288 b = 36 B

b) If we have a flat page table that is stored in the main memory (MM), what is the average access time (AAT) for this system? Assume that the cache is physically indexed physically tagged (PIPT).

Solution:

AAT = HitTime + MissRate×MissPenalty

Address: TLB -> MM AAT = 1 + 0.01%×150 = 1.015 (not Disk)

Data: Cache -> MM -> Disk
AAT = 1 + 3%×150 + 3%×0.0001%×10000 = 5.5003

AAT = DATA + ADDR = 6.5153 cycles

c) If the cache is VIPT, what is the new AAT?

Solution:

We consider three scenarios:

1) **L1 hit: AAT = 1**
2) **L1 miss, TLB hit: Penalty = 150 + 0.0001%×10000 = 150.01 (data only)**
3) **L1 miss, TLB miss: Penalty = [150 + 0.0001%×10000] + [150] = 300.01 (data, address)**

In case of a cache miss, we have to fetch the data from the main memory or the disk; and we already know the result of TLB miss/hit.
In case of a TLB hit we have the address but not the data.
In case of a TLB miss, we need to fetch the address from the main memory as well.

Final AAT is the weighted average of these three cases:

AAT = 1 + 3%×[99.99%×150.01 + 0.01%×300.01] = 5.50075 cycles

Q3. ***Memory Consistency:*** *'This is it! Can you find them all? It's OK if you didn't! What is important is that you consistently do your best!'*

Assume that we have a multi-core system with the following instruction sequences:

Initial Values: `Per Thread: x1=0` `Global: Money = 1`

THREAD 1

```
x1 = Money;
fence;
Money = Money+x1;
acquire (lock);
x1= Money;
release (lock);
```

THREAD 2

```
acquire (lock)
x1 = Money + x1;
release (lock);
Money = 2;
fence;
```

THREAD 3

```
x1 = Money;
fence;
acquire (lock);
Money = Money + 2;
release (lock);
x1 = x1 + Money;
```

Assume that we have ***release consistency*** memory model. What are the possible values for x1 in T3? (Show the sequence by denoting numbers as T1.1, T1.2, etc. where Tx.y means the Yth instruction in Thread X. For each unique answer, show only one possible sequence that would create that value.)

Solution:

Thread 1	Thread 2	Thread 3
T1.1: x1= Money;	**T2.1: acquire (lock);**	**T3.1: x1 = Money;**
T1.2: fence;	**T2.2: x1 = Money + x1;**	**T3.2: fence;**
T1.3: Money = Money + x1;	**T2.3: release (lock);**	**T3.3: acquire (lock);**
T1.4: acquire (lock);	**T2.4: Money = 2;**	**T3.4: Money = Money + 2;**
T1.5: x1= Money;	**T2.5: fence;**	**T3.5: release (lock);**
T1.6: release (lock);		**T3.6: x1 = x1 + Money;**

(release consistency) code above acquire < acquire < code between acquire and release < release < code after release
T1.1, T1.2, T1.3 < T1.4 < T1.5 < T1.6
T2.1 < T2.2 < T2.3 < T2.4, T2.5
T3.1, T3.2 < T3.3 < T3.4 < T3.5 < T3.6

(fence)
T1.1 < T1.2, T1.3, T1.4, T1.5, T1.6
T2.1, T2.2, T2.3, T2.4 < T2.5
T3.1< T3.2, T3.3, T3.4, T3.5, T3.6

x1 in T3	Sequence *
3	T3.1 < T3.2 < T3.3 < T3.4 < T3.5 < T1 or T2 < T3.6
4	T3 < T1, T2
5	T3.1 < T1 < T2 < T3.2 < T3.3 < T3.4 < T3.5 < T3.6
6	T1 < T2 < T3
7	T3.1 < T3.2 < T2 < T1 < T3.3 < T3.4 < T3.5 < T3.6
8	T1.1 < T2 < T1.2 < T1.3 < T3
9	T3.1 < T3.2 < T2 < T3.3 < T3.4 < T3.5 < T1 < T3.6
10	T2 < T3.1 < T3.2 < T3.3 < T3.4 < T3.5 < T1 < T3.6

*** The answers for sequences are not unique.**
There are other correct sequences as long as x1 in T3 has the above values (3-10).

Q4- ***Bonus:*** *Have a restful Winter break ;)*

a) If you can date with anyone in the world (past or present), who would that person be and why? (happy to forward this to your secret crush if you want :D)

Nader: There is only one correct answer for me! And the reason is because she is the best ;)

b) Don't forget the course evaluation! Do it NOW!

We got it! Thank you all for participating!