CS2323 Computer Architecture 2018

Homework 1

- Your submission should be named as RollNumber_CA_HW1.pdf. For example, if your roll number is cs16mtech11075, then your submission should be cs16mtech11075_CA_HW1.pdf. Except pdf, no other format is acceptable. **10 marks will be deducted for not following these instructions or if you submit a zipped file.**
- If you submit hand-written solution after scanning, make sure all the text is legible.
- The reasoning for obtaining the answer should be clearly shown to obtain full marks. At the same time, be concise.
- There are 8 bonus marks for writing your solution in Latex (in HW2, there will no bonus for writing your solution in Latex). You need to include the following line at the beginning of your solution (i.e., in your *.tex file) to get the bonus:
  This document is generated by \LaTeX
- Late Submission Penalty: 20% for each late day (including weekend)

==========

Q1. (2 marks) Consider an L2 cache which gets 50000000 accesses per second. It is designed with a memory which has leakage power of 0.07W and dynamic energy of 0.8nJ/access. If an application executes for one second, how much percentage of total cache energy comes from leakage energy?

Q2. (1 mark) Find out the reach of this DTLB:

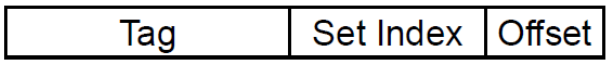
PageSize	Entries	associativity
4KB	64	4-way
2MB	32	8-way
1GB	8	fully-associative

Q3. (10 marks) Assume that a processor uses 8-bit address space. Assume that the address pattern that accesses the cache is:

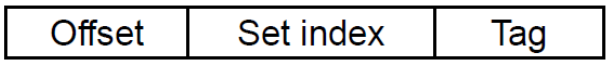
Sequence1: 0, 63, 1, 62, 2, 61, 3, 60, 4, 59, 5, 58, 6, 57, 7, 56, 8, 55, 9, 54, 10, 53, 11, 52

Assume two different caches use the following two different address subdivision methods (figure is not drawn to scale).

Address subdivision method used by Cache 1



Address subdivision method used by Cache 2



Both the caches are direct-mapped, with a block size of 4 and have 8 sets each. In other words, the architectures are identical, except that they use different subdivision methods.

(a) Compute the tag and set for each address for both subdivision methods (hint: you can write a small C program to do that). You need not show this in your submission. For each address, show whether it leads to a hit or a miss and finally, what is the hit ratio (hits/accesses) for each cache?
(b) Repeat (a) but with the following sequence:

Sequence2: 0, 64, 128, 192, 1, 65, 129, 193, 11, 75, 139, 203, 9, 137, 201, 73

Q4. (4 marks) Consider two processors (P1 and P2) which run the same instruction set architecture (ISA). The frequency of P1 and P2 are 2.2GHz and 1.6GHz, respectively.

In this ISA, there are four classes of instructions A, B, C, and D. The CPI of each of these classes are given in the following table.

	A	B	C	D
P1	1	2	3	4
P2	2	2	2	2

There is a program which has 10^6 instructions divided into classes as follows: 20% class A, 25% class B, 45% class C, and 10% class D. Which processor is faster for this program?

Q5. (4 marks) Assume that a system has 4 processors (P=4). Assume that directory-based coherence protocol is used. Show the state of (P+1) bit directory for a cache block after each of these operations to that block.

i. P0 has read miss
ii. P1 has write miss
iii. P3 has write miss
iv. P2 has read miss
v. P2 has write miss
vi. P0 has read miss

Q6. (3 marks) Two applications are running on a processor which has shared L2 cache.

For application1: L2 cache misses with 2 and 6 ways (of last level cache) are 1000 and 400, respectively.

For application2: L2 cache misses with 2 and 6 ways (of last level cache) are 2000 and 1800, respectively

Assume that in between 2 and 6 ways, number of misses scale linearly (i.e., use linear interpolation).

Assume the cache has 8 ways, then which application should get how many ways for improving performance. An application needs to get at least two ways.

Q7. (2 marks) Three applications P, Q, R have a transactions rate of 34 per minute, 58 per minute and 81 per minute respectively. They run one after another. If each of them make 500 transactions, find the correct average value of transactions per minute. Also write which mean would you use to get the average.

Q8. (2 marks) An application spends 25% of time in initialization, 37% of time in vision-processing function and remaining time in signal-processing function. In system0, all the tasks are run on a single-core CPU. However, system1 has an signal-processing accelerator and a vision-processing accelerator which give a speedup of 10X and 6X, respectively over the single-core CPU execution. Find the speedup of system1 over system0 assuming that both the accelerators are used on system1.

Q9. (5 marks) Consider a processor that runs at 3 GHz and 1 Volt. The processor is capable of executing safely at voltages between 0.8 V to 1.2 V. Voltage and frequency follow a linear relationship (i.e., if voltage doubles, frequency doubles as well). When running a given CPU-bound program, the processor consumes 110 W, of which 30 W is leakage. The program takes 30 seconds to execute. Compute the following values (and also show at what frequency/voltage they are obtained): (i) The smallest time it takes to execute the program (1 mark). (ii) The lowest power to execute the program (2 mark). (iii) The lowest energy to execute the program (2 mark).

Q10. (2 marks) A processor uses 48 bit virtual address. It has 2GB physical memory and the memory addresses are defined at the level of each byte. Page size is 2KB. A processor has only one level of TLB which has 32 entries. Find out the size of TLB (excluding valid bits, etc).

Q11. (2 marks) Consider a TLB which has 4 ports. The processor has a frame size of 1KB. In a given cycle, the addresses coming to the four ports of TLB are: 0x4795BA21, 0x4795BB21, 0x5795BA21 and 0x4785BA21. Find out how many

unique accesses can be sent to TLB if it uses intra-cycle compaction technique to save energy.

Q12. (3 mark) We have a small 2-entry, 2-way cache and the block size is 1B. Consider an access stream with addresses

A, B, C, D, A, B, C, D, A, B, C, D.

Show whether each of the access is a hit or a miss with (a) LRU (least recently used) replacement policy (1 mark) and (b) MRU (most recently used) replacement policy. (you don't need to show the state of the cache. Just show the hit/miss decision for each access and total number of misses).