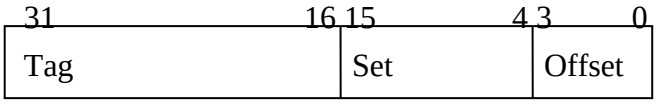
1. (30 %) A 32-bit cache memory address is decomposed into 3 fields as shown as below:



1.1 What is the total size (including tag, data and valid bits) of the cache memory in ***bits*** if the cache is (a) direct-mapped and (b) 2-way set associative? (10 pts.)

1.2 Assume that the cache memory is 2-way set associative and is originally empty. Then the CPU accesses the memory with the following address sequence:

A.0x24AB1A34
B.0x24AB2810
C.0x24AC1A34
D.0x24AB1A3A
E.0x24AC2810
F.0x24AC1A35
G.0x24AA1A30
H.0x24AC1A34
I.0x24AA1A34
J.0x24AB1A30
K.0x24AB280B
L.0x24AB2812

If the block replacement policy is LRU (Least Recently Used). Please indicate which memory accesses (A~L) are “miss” and which (A~L) are “hit”? (10 pts.)

1.3 Please describe how TLB, page table, processor cache, main memory and disk function together for a memory access. (10 pts.)

2. (20%) Consider two different implementations of the same instruction set architecture. There are five classes of instructions, A, B, C, D and E. The clock rate and the CPI of each implementation are given in the following table.

	ClockRate	CPIA	CPIB	CPIC	CPID	CPIE
P1	2.0GHz	1	2	3	5	4
P2	1.8GHz	3	3	3	3	3

2.1 Given a program with 10^5 instructions divided into classes as follows: 10% Class A, 15% Class B, 40% Class C, 20% Class D and 15% Class E. Which one is fast by their MIPS? Is MIPS a reasonably reliable metric in this case? (10 pts.)

2.2 Following (2.1), you are trying to improve P1 by saving 20% in the execution time. However, this leads to an increase of 20% of the average CPI. What clock rate should we have to set to achieve this time reduction? (10 pts.)

3. (10%) Please use a figure to illustrate a (2,2) predictor selected by the lower order 11 bits of branch address and explain how it works in details. How many bits are required in this predictor?

4. (10%) Consider the following code sequence. F1- F10 are registers. Assume the following execution time for different operations: MUL takes 8 cycles, DIV takes 20 cycles, and ADD and SUB both take 1 cycle to complete. The instructions are issued into the pipeline in order, but out of order execution and completion is allowed. What hazards will be encountered when executing this code sequence?
   MUL F6, F5, F6
   SUB F2, F7, F8
   DIV F10, F6, F5
   ADD F5, F8, F2

5. (10%) Suppose we would like to achieve a speedup of 20 with 40 processors. What fraction of the original computation can be sequential?

6. (10%) What is dynamic scheduling? Why is dynamic scheduling important?

7. (10%) What is a loop-carried dependence? Please give me an example of a loop code sequence that cannot be unrolled.