Architecture

A1. (25 pts) A typical user runs 3 programs - P1, P2 and P3 - with some specified frequency. A program Px takes time Tx(A) to run on architecture A and Tx(B) to run on architecture B.

a) (4 pts) Give the equation for the relative performance of A and B if we know that a user runs P1 once for every 3 runs of P2 and for every 5 runs of P3.

b) (4 pts) Give the equation for the relative performance of A and B if we know that a user runs P1 1/9th of time, P2 3/9ths of time and P3 the rest of the time, on a reference machine R. Assume that T1(R) = 1, T2(R)=3 and T3(R)=10.

c) (5 pts) On architecture A, we can apply an enhancement for integer instructions that shortens integer calculation by 30%. If integer instructions are performed 80% of the time, how large is the overall speedup?

d) (6 pts) Now repeat part c) if integer instructions are performed 50% of the time **after the enhancement is applied.**

e) (6 pts) Repeat part c) if every integer instruction takes 2 clock cycles, any other instruction takes 5 clock cycles, and integer instructions make up 20% of all instructions in the program mix.

Architecture

A2. (25 pts) For the following loop:

```
Loop:   L.D F0, 100(R1)
        ADD.D F4, F0, F2
        ADD.D F6, F4, F8
        S.D F6, 100(R2)
        DADDUI R1, R1, #4
        DADDUI R2, R2, #-4
        DADDUI R3, R3, #-1
        BNEZ R3, Loop
```

a) (7 pts) How many cycles does it take to execute one iteration of this loop (without any instruction shuffling) on the following architecture:
Integer pipeline: IF ID EX MEM WB
Floating point pipeline: IF ID EX1 EX2 EX3 EX4 EX5 MEM WB
Assume that there is no structural hazard during the execution (pipeline resources are duplicated as needed), the processor is performing a single instruction issue per clock cycle and the only slow-down factor is the existence of RAW hazards. Also assume that branches are completely resolved in EX stage. Do not forget to account for branch delay in total cycles needed to execute one loop iteration.

b) (8 pts) Shuffle instructions around to get rid of as many stalls as you can (including the branch delay). You may change offsets if necessary. Show the transformed loop and list how many cycles does it take to execute one loop iteration now.

c) (10 pts) Apply software pipelining to the original loop and show the transformed code. Do not forget to show the startup and cleanup code. Make sure you don't have unnecessary instructions in the startup and cleanup code but you needn't optimize this code. You may also assume that registers R1, R2 and R3 are not used in the code following the loop. How many cycles does it take to execute one iteration of the transformed loop?

Architecture

A3. (25pts) For each of the three following dynamic scheduling techniques, describe how they work in detail and how they handle: RAW hazards, WAR hazards, WAW hazards and control hazards. For each hazard only discuss how it is detected/handled if it occurs on data stored in registers. You must list all the resources and actions used to handle hazards by each technique to receive full credit.

a) (7 pts) Scoreboarding (you may skip the discussion of control hazards for this part)
b) (9 pts) Tomasulo's algorithm without speculative execution
c) (9 pts) Tomasulo's algorithm with speculative execution

Architecture

A4. (25 pts) You are given a system with 2-level unified cache. L1 cache is no-write-allocate, write-through cache, with size 64KB and a block size of 64B. L2 cache is write-allocate, write-back cache with size 1MB and a block size of 128B. On the average 20% of blocks in L2 are dirty. For every 1000 cache system accesses, there are 50 misses in L1 cache and 8 misses in L2 cache. L1 access time is 1ns, L2 access time is 10ns and memory access time is 100ns. CPU reads and writes 32B words. The following table gives the cost of transferring 32B of data between CPU, L1, L2 and memory.

32B between CPU and L1	32B between L1 and L2	32B between L2 and MEM
0.5ns	5ns	10ns

a) (3 pts) List local and global miss rates for L1 and L2
b) (3 pts) How large is L2 read miss penalty?
c) (3 pts) How large is L2 write miss penalty?
d) (3 pts) How large is L1 read miss penalty?
e) (3 pts) How large is L1 write miss penalty?
f) (3 pts) How much do we pay for L1 read hit?
g) (3 pts) How much do we pay for L1 write hit?
h) (2 pts) What is average memory access time for reads?
i) (2 pts) What is average memory access time for writes?