National Institute of Technology, Tiruchirappalli

Department of Computer Science and Engineering

ASSIGNMENT
CSPC31 – Computer Architecture

---

1. Consider the sequence of machine instructions given below:

   MUL R5, R0, R1
   DIV R6, R2, R3
   ADD R7, R5, R6
   SUB R8, R7, R4

   In the above sequence, R0 to R8 are general purpose registers. In the instructions shown, the first register stores the result of the operation performed on the second and the third registers. This sequence of instructions is to be executed in a pipelined instruction processor with the following 4 stages: (1) Instruction Fetch and Decode (IF), (2) Operand Fetch (OF), (3) Perform Operation (PO) and (4) Write back the Result (WB). The IF, OF and WB stages take 1 clock cycle each for any instruction. The PO stage takes 1 clock cycle for ADD or SUB instruction, 3 clock cycles for MUL instruction and 5 clock cycles for DIV instruction. The pipelined processor uses operand forwarding from the PO stage to the OF stage. How many clock cycles will the above sequence of instructions take?

2. Suppose that in 1000 memory references there are 40 misses in the first-level cache and 20 misses in the second-level cache. What are the various miss rates? Assume the miss penalty from the L2 cache to memory is 200 clock cycles, the hit time of the L2 cache is 10 clock cycles, the hit time of L1 is 1 clock cycle, and there are 1.5 memory references per instruction. What is the average memory access time and average stall cycles per instruction? Ignore the impact of writes.

3. Suppose you have a machine which executes a program consisting of 50% floating point multiply, 20% floating point divide, and the remaining 30% are from other instructions.
   a) Management wants the machine to run 4 times faster. You can make the divide run at most 3 times faster and the multiply run at most 8 times faster. Can you meet management's goal by making only one improvement, and which one?
   b) Dogbert has now taken over the company removing all the previous managers. If you make both the multiply and divide improvements, what is the speed of the improved machine relative to the original machine?
4. Given an application that has two phases of execution on a single processor: Phase 1 is serial and runs for 4 minutes, and Phase 2 is parallelizable and runs for 8 minutes. What is the speedup when this application is run on a parallel computer with 4 CPUs?
5. You find yourself in a game show presented with 2 machines (running different compilers). You are supposed to pick the fastest one to win an awesome prize! You are given the following information about the two machines A and B. Machine A has a clock rate of 2 GHz with the following measurements.

Instruction Class	CPI for the Class	Instruction Count (in Billions)
A	4	9
B	1	15
C	5	12
D	2	24

Machine B has a clock rate of 2.5 GHz with the following measurements.

Instruction Class	CPI for the Class	Instruction Count (in Billions)
A	1	16
B	3	24
C	1	12
D	5	28

(a) What is the average CPI of machine A and B?
(b) On which machine is the program faster with respect to
I. Execution time
II. MIPS rating

6. The largest configuration of a Cray T90 (Cray T932) has 32 processors, each capable of generating 4 loads and 2 stores per clock cycle. The processor clock cycle is 2.167 ns, while the cycle time of the SRAMs used in the memory system is 15 ns. Calculate the minimum number of memory banks required to allow all processors to run at full memory bandwidth.

7. Consider a loop like this one:
```
for (i=0; i<100; i=i+1) {
A[i+1] = A[i] + C[i]; /* S1 */
B[i+1] = B[i] + A[i+1]; /* S2 */
}
```
Assume that A, B, and C are distinct, non-overlapping arrays. (In practice, the arrays may sometimes be the same or may overlap. Because the arrays may be passed as parameters to a procedure that includes this loop, determining whether arrays overlap or are identical often requires sophisticated, inter procedural analysis of the program.) What are the data dependences among the statements S1 and S2 in the loop?

8. Show the code for MIPS and VMIPS for the following DAXPY loop. Assume that the starting addresses of X and Y are in Rx and Ry, respectively

```
        DADDIU   R4,R1,#800  ;  R1 = upper bound for X
foo:    L.D      F2,0(R1)    ; (F2) = X(i)
        MUL.D    F4,F2,F0    ; (F4) = a*X(i)
        L.D      F6,0(R2)    ; (F6) = Y(i)
        ADD.D    F6,F4,F6    ; (F6) = a*X(i) + Y(i)
        S.D      F6,0(R2)    ; Y(i) = a*X(i) + Y(i)
        DADDIU   R1,R1,#8    ; increment X index
        DADDIU   R2,R2,#8    ; increment Y index
        DSLTU    R3,R1,R4    ; test: continue loop?
        BNEZ     R3,foo      ; loop if needed
```

9. Assume a hypothetical GPU with the following characteristics:
   ■ Clock rate 1.5 GHz
   ■ Contains 16 SIMD processors, each containing 16 single-precision floating point units
   ■ Has 100 GB/sec off-chip memory bandwidth Without considering memory bandwidth, what is the peak single-precision floating-point throughput for this GPU in GLFOP/sec, assuming that all memory latencies can be hidden? Is this throughput sustainable given the memory bandwidth limitation?
10. Show how the following code sequence lays out in convoys, assuming a single copy of each vector functional unit:

```
LV          V1,Rx        ;load vector X
MULVS.D     V2,V1,F0     ;vector-scalar multiply
LV          V3,Ry        ;load vector Y
ADDVV.D     V4,V2,V3     ;add two vectors
SV          V4,Ry        ;store the sum
```

How many chimes will this vector sequence take? How many cycles per FLOP (floating-point operation) are needed, ignoring vector instruction issue overhead?

******