**Notes**:

* Closed Text and Closed Notes Exam
* Time: 12:30 pm – 2:30 pm
* There are **8**questions. Answer all questions. Each question carries 10 pts.
* Maximum points: 80.
* Answer in clear and legible handwriting. Partial credit will be given.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **I** | **II** | **III** | **IV** | **Total** |
|  |  |  |  |  |
| **V** | **VI** | **VII** | **VIII** |
|  |  |  |  |

1. (10 pts.) Instruction Set Design Principles
   1. (2 pts.) What is **Computer Architecture?**
   2. (4 pts.) Decide whether each of the following is true or false. Add brief explanation (1-2 sentences) to get full credit.
      1. The performance of the system is limited by the slowest component even if some components are made 10X faster.
      2. You can afford not to pay attention to Amdahl’s law because it is not applicable anymore.
      3. MIPS rating of a processor is a good metric to measure its performance.
      4. The future of performance improvement will be mostly dependent on parallelization of programming rather than blindly adding multiple cores to a chip.
   3. (4 pts.) Some microprocessors today are designed to have adjustable voltage, so a 15% reduction in voltage may result in a 15% reduction in frequency. What would be the impact on dynamic energy and on dynamic power?
2. **Performance Measurement**

Suppose we made the following measurements:

Frequency of FP operations = 25%

Average CPI of FP operations = 4.0

Average CPI of other instructions = 1.33

Frequency of FSQRT = 2%

CPI of FSQRT = 20.

(4 pts) What is the CPI?

(6 pts) Assume that the two design alternatives are to decrease the CPI of FSQRT to 2 or to decrease the average CPI of all FP operations to 2.5. Which one is better? Provide quantitative justification.

1. **Memory Hierarchy**
   1. (6 pts.) For the following advanced cache optimizations, briefly the idea.
      1. Way predicting Cache
      2. Nonblocking Cache
      3. Critical Word first and early restart
   2. (4 pts.) Which is more important for floating point programs: **two-way set associativity** or **hit under one miss** for primary data caches? Assume the following average miss rates for 32 KB data caches: 5.2% for FP programs with direct mapped cache, 4.9% for programs with 2-way set associative cache. Assume the miss penalty to L2 is 10 cycles, and L2 misses and penalties are the same. Also assume hit under one miss reduces the average data cache latency for FP programs to 87.5% of a blocking cache.
2. **Basic Pipelining**
   1. (4 pts) The following series of branch outcomes occurs for a single branch in a program. (T means the branch is taken, N means the branch is not taken).

Index 1 2 3 4 5 6 7 8 9 10 11 12 13  
 T, T, N, T, N, T, T, T, T, N, T, T, N

Assume that we are trying to predict this sequence with a Branch History Table (BHT) using a 1-bit prediction. The counters of the BHT are initialized to the N state. Which of the branches would be mispredicted? Show their indices.

* 1. (6 pts) Use the following code segment:

Loop:ld x1, 0(x2) : load x1 from address 0+x2

Addi x1, x1, 1 ; x1=x1+1

Sd x1, 0, (x2); store x1 at address 0+x2

Addi x2, x2, 4 ; x2 =x2 + 4

Sub x4, x3, x2 ; x4 = x3 -x2

Bnez x4, Loop ; branch to Loop if x4!=0

Assume that the initial value of x3 is x2 + 396. List all of the data dependencies in the code above. Record the register, source instruction, and destination instruction.

1. **RISC- Architecture**

Answer the following RISC-V Base ISA architecture:

1. (4 pt.) Name any *four ISA* design principles that RISC-V implements.
2. (1 pt.) No. of integer registers \_\_\_\_\_\_\_

No. of FP registers\_\_\_\_\_\_\_\_\_\_

1. (1 pt.) Instruction word length\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. (1 pt.) No. of instruction formats \_\_\_\_\_\_\_\_\_\_\_\_\_
3. (1 pt.) Value of x0 is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. (2 pts) Why is it called a Load-store architecture?

1. **Instruction level Parallelism**
   1. (6 pts) Draw the block diagram of basic structure of a RIS-V FP Unit using Tomasulo’s algorithm. Briefly explain the purpose of each block. Assume FP adders, FP multipliers for arithmetic computation.
   2. (4 pts) Show what the information tables look like for the following code sequence when only the first load has complete and written its result.
      * 1. fld f6, 32(x2)
        2. fld f2, 44(x3)
        3. fmul.d f0, f2, f4
        4. fsub.d f8, f2, f6
        5. fdiv.d f0, f0, f6
        6. fadd.d f6, f8, f2
2. **Data level Parallelism**
   1. (2 pts.) What is the basic idea of a vector processor?
   2. (8 pts.) With an example, explain the basic idea of the following techniques:
      1. Vector chaining
      2. Strip Mining
      3. Vector Stride
      4. Gather and Scatter
3. **Thread Level Parallelism**
   1. (5 pts.) Explain how the following protocols work:
      1. Snooping coherence protocol
      2. Directory based coherence protocol.
   2. (2.5 pts.)Using a state-diagram like figure, show how a **read miss** from local node is handled by the home node and remote node. Briefly explain how it works.
   3. (2.5 pts.) Using a state-diagram like figure, show how a **write miss** from local node is handled by the home node and remote node. Briefly explain how it works.

**Formulas**

2. If X is ‘n’ times fast as Y then:
3. Amdahl’s Law:

1. For Multilevel Caches: