

1. (a) Identify all data dependencies in the following sequence of instructions, when executed on the pipelined MicroMIPS implementation (with no forwarding).

```
addi $9, $zero, 0      [1]
addi $12, $zero, 5000  [2]
loop: addi $12, $12, 4   [3]
       lw    $8, 40($12)  [4]
       add   $9, $9, $8    [5]
       addi $11, $11, -1   [6]
       bne  $11, $zero, loop [7]
```

- (b) Determine the number and places of bubbles that must be inserted to correctly execute the instruction sequence in part (a). Explain your reasoning in each step.
(c) Can you suggest any reordering of instructions that would reduce the number of bubbles ?
2. [5 marks] (*) Data hazards occur if the sequence of read and write (to/from memory or registers) in the instruction sequence are such that the end result is inconsistent. There are four combinations of the read/write sequences and we will name them, “X after Y”, where X and Y could be *Read* or *Write* (RAW, RAR, WAW, WAR).
- Which of these four are possible hazards in general pipeline implementations? Explain why the other(s) are not hazards.
 - Which of the potential data hazards that you enumerated above, if any, is not a hazard in the 5-stage pipeline implementation that we have considered in class. Why?
 - Consider the following instruction sequence. What type of a data hazard does this sequence contain.

```
add $r2, $r3 $r4
sub $r5, $r6 $r2
```

- (d) Consider the instruction sequence below

```
sw $r0, 100($r2)
lw $r1, 100($r2)
```

Compare the sequence of operations in the above sequence and the one in the previous question and explain why there will be or will not be a “read after write” hazard in the 5-stage implementation that we consider in class.

- (e) Consider the following code sequence to be implemented on the basic 5-stage pipeline that we have considered in class (with no throughput enhancement technique employed)

```
add $r5, $r6 $r7
lw $r6, 100($r7)
sub $r7, $r6 $r8
```

Assuming that the pipeline is flushed before this segment begins execution, how many clock cycles does this segment take. Show the instruction in each stage of the pipeline in each clock cycle.

3. [3 marks] (*) A computer architect needs to design the pipeline of a new microprocessor. An example workload program core has 10^6 instructions, and each instruction takes 100 ps to finish.

- How long does it take to execute the workload program on a nonpipelined processor ?
- The current state-of-the-art microprocessor has about 20 pipeline stages. Assume it is perfectly pipelined. How much speedup will it achieve compared to the nonpipelined processor ?
- Real pipelining isn't perfect, since implementing pipelining introduces some overhead per pipeline stage. Will this overhead affect instruction latency, instruction throughput, or both ?

4. (a) Delayed branching can help in the handling of control hazards. For all delayed conditional branch instructions, irrespective of whether the condition evaluates to true or false,

- the instruction following the conditional branch instruction in memory is executed

- B. the first instruction in the fall through path is executed
 C. the first instruction in the taken path is executed
 D. the branch takes longer to execute than any other instruction
- (b) The following code is to run on a pipelined processor with one branch delay slot:

```
I1: add $r2, $r7, $r8
I2: sub $r4, $r5, $r6
I3: add $r1, $r2, $r3
I4: sw $r1, 100($r4)

...
bltz $r1, Label
    <-- Delay slot
...
```

Which of the instructions I1, I2, I3 or I4 can legitimately occupy the delay slot without any other program modification?

5. Consider three machines with the same processor but different cache configurations. $C1$ is direct mapped with one-word blocks, $C2$ is direct mapped with four-word blocks and $C3$ is two-way set associative with four-word blocks. $C1$ has an instruction miss rate of 4% and a data miss rate of 8%, while the corresponding numbers for $C2$ are 2% and 5% respectively and for $C3$ are 2% and 4% respectively. Assume half of the instructions have a data reference, and the cache miss penalty is $6 + \text{Block size in words}$. The CPI for this workload on a machine with $C1$ was measured to be 2.0. Determine the time spent by each machine on cache misses. If the clock period is 2.0 ns for the first and second configurations and 2.4 ns for the third, determine the MIPS (million instructions per second) rating of the three machines.
6. [4 marks] (*) The following C program is run (with no optimizations) on a machine with a cache that has four-word (16-byte) blocks and holds 256 bytes of data

```
int i, j, c, stride, array[256]
. . .
for (i=0; i<10000; i++)
    for (j=0; j<256; j=j+stride)
        c=array[j]+5;
```

If we consider only the cache activity generated by references to the array and we assume that integers are words, what is the expected miss rate when the cache is direct mapped and $\text{stride} = 132$? How about $\text{stride} = 131$? Would either of these change if the cache were to be two-way set-associative?

7. A processor with two levels of caches had a CPI of 1 when there is no level-1 cache miss. At level-1, the hit rate is 95 % and a miss incurs a 10-cycle penalty. For the two-level caches as a whole, the hit rate is 98 % (meaning that 2% of the time the main memory must be accessed) and the miss penalty is 60 cycles.
 - (a) What is the effective CPI after cache misses are factored in ?
 - (b) If a single-level cache were to be used in lieu of this two-level cache system, what hit rate and miss penalty would be needed to provide the same performance ?
8. The following sequence of number represents memory addresses in a 64-word main memory: 0, 1, 2, 3, 4, 15, 14, 13, 12, 11, 10, 9, 0, 1, 2, 3, 4, 56, 28, 32, 15, 14, 13, 12, 0, 1, 2, 3. Classify each of the accesses as a cache hit, compulsory miss, capacity miss, or conflict miss, given the following cache parameters. In each case, also provide the final cache contents.
 - (a) Direct-mapped, 2-word block, 4-block capacity
 - (b) Two-way set associative, 4-word blocks, 2-set capacity, LRU replacement

9. [4 marks] (*) A computer system has 4 GB of byte-addressable main memory and a 256 KB unified cache memory with 32-byte blocks.
 - (a) Draw a diagram showing each of the components of a main memory address (i.e., how many bits for tag, set index, and byte offset) for a four-way set-associative cache.
 - (b) The performance of the computer system with four-way set-associative cache architecture proves unsatisfactory. Two redesign options are being considered, implying roughly the same additional design and production costs. Option A is to increase the size of the cache to 512 KB. Option B is to increase the associativity of the 256 KB cache to 16-way. In your judgement, which option is more likely to result in greater overall performance and why ?
10. An address for a byte-addressable memory presented to the cache unit is divided as follows: 13-bit tag, 14-bit block index, 5-bit byte offset.
 - (a) What is the cache size in bytes ?
 - (b) What is the cache mapping scheme ?
 - (c) For a given byte in cache, how many different bytes in the 2^{32} -byte main memory can occupy it ?