1. (10%) What are the advantages and challenges of longer pipelines (larger number of stages)?
What are possible solutions of the challenges?

2. (10%) Explain the concept of Delayed Branch and Cancelling Branch. What are the
common ways to get instructions to fill branch delay slot?

3. (15%) What is Loop Unrolling when exploiting instruction level parallelism? What are the
advantages? Explain it clearly by an example.

4. (15%) Suppose we would like to achieve a speedup of 25 with 60 processors. What fraction
of the original computation can be sequential?

5. (24%) A 32-bit cache memory address is decomposed into 3 fields as shown as below:

<table>
<thead>
<tr>
<th>31</th>
<th>16</th>
<th>15</th>
<th>4</th>
<th>3</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag</td>
<td>Set</td>
<td>Offset</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1. 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? (8 pts.)

5.2 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. 0x13AB1A34
b. 0x13AB2810
c. 0x13AC1A34
d. 0x13AB1A3A
e. 0x13AC2810
f. 0x13AC1A35
g. 0x13AA1A30
h. 0x13AC1A34
i. 0x13AA1A34
j. 0x13AB1A30
k. 0x13AB280B
l. 0x13AB2812

If the block replacement policy is LRU (Least Recently Used). Please indicate which
memory accesses (a~j) are “miss” and which (a~j) are “hit”? (8 pts.)

5.3 Please describe how TLB, page table, processor cache, main memory and disk function
together for a memory access. (8 pts.)
6. (18%) You are comparing two different computers M1 and M2. M1 has a clock rate of 1GHz and M2 has a clock rate of 500MHz.

6.1 (4 pts.) The following shows the CPI of 3 instruction classes:

<table>
<thead>
<tr>
<th>Instruction Class</th>
<th>CPI for M1</th>
<th>CPI for M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

Compare the two processors' PEAK performance in terms of MIPS (i.e. peak MIPS).

6.2 (6 pts.) Following (a). There are three compilers: C1, C2 and C3 that produce the instruction mixes as follows:

<table>
<thead>
<tr>
<th>Instruction Class</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30%</td>
<td>30%</td>
<td>50%</td>
</tr>
<tr>
<td>B</td>
<td>50%</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>C</td>
<td>20%</td>
<td>50%</td>
<td>20%</td>
</tr>
</tbody>
</table>

For C1, compare M1 and M2. Which processor is faster and by how much?

6.3 (8 pts.) Following (a) and (b). Assume that any of the three compilers (C1, C2, C3) can be used on M1 and M2. In addition, assume that the numbers of instructions from C1 and C2 are the same for both M1 and M2 and are equal to 1000, while the number of instructions from C3 is 1100 (for both M1 and M2). Please calculate the execution time to determine which "computer+compiler" you will choose. (That is, you have to calculate the execution time in each combination to decide what you will choose.)

7. (8%) About IEEE 754

7.1 (4 pts.) Write the IEEE 754 representation in the single precision for -11.125.

7.2 (4 pts.) What decimal number is represented by this word in IEEE 754?

01000000111100000000000000000000