1. (15 pts) The 32 bit address generated by a certain processor is divided up as shown to access its cache:

<table>
<thead>
<tr>
<th>Tag</th>
<th>Index</th>
<th>Offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 bits</td>
<td>11 bits</td>
<td>6 bits</td>
</tr>
</tbody>
</table>

If this address is used to access a 2-way set associative cache, answer the following questions. If you cannot determine the answer to a question from the information given, then so state. Assume the memory is byte addressable.

- What is the cache line (block) size in bytes?

- How many sets are in the cache?

- How many blocks are in the cache?

- Is the cache a virtually addressed or physically addressed cache?

- What is the total cache size in bytes?
2. (9 pts) When designing a cache memory, a designer can choose several different parameters for that cache - total cache size, cache line (block) size, and associativity. For a given set of these parameters, explain what you would need to do to reduce the following types of cache misses:

a) Compulsory misses

b) Capacity misses

c) Conflict misses

3. (6 pts) With virtual memory, it is not necessary for the virtual address space to be the same as the physical address space. For example, the larger models of the 16-bit PDP11 processor had a 4 mB (22 bit) physical address space. Conversely, several models of the 64-bit Alpha processor have a 44-bit physical address. Explain how this is done.
4. (20 pts) You currently have a processor with a direct-mapped, L1 cache. Its hit rate is 90%, and the miss penalty to main memory is 80 cycles. You are considering adding an off-chip L2 cache. This cache will be sized so that its hit rate is 96%. The hit time in this cache is 8 cycles, and its miss penalty to main memory is 90 cycles. Determine the speedup you can expect to gain (if any) with this cache, if the base CPI of the processor is 1.5, with an instruction mix that includes 20% loads and 10% stores.
5. (12 pts) Explain briefly how the following techniques can improve the performance of a cache memory system.

- Merging write buffers
- Victim buffer
- Compiler-based prefetching

6. (10 pts) Strip-mining is a technique to efficiently use vector processors.

- (5 pts) Explain what strip-mining is.
- (5 pts) Explain why it is used.
7. (8 pts) Explain why an atomic memory transaction is difficult with a shared bus multiprocessor. Describe how the RISC-V processor deals with this problem.

8. (10 pts) In the cache coherence protocol that the book describes (e.g., in Figure 5.6), it assumes that a write hit and a write miss are treated the same. The MESI protocol adds an additional state (Modified) to the protocol. What advantage does this protocol have compared to the one in the book?

9. (10 pts) Sequential consistency is similar to processor consistency in multiprocessor systems. Show an example of a transaction where the results would be different for a sequentially consistent processor and a processor consistent one.