1. Cache performance is a factor of several parameters. For each of these, describe the issues that arise if their value is either too small or too large:
   A. Cache size
   B. Line size
   C. Associativity
   D. Victim cache size
   E. Replacement policy (LRU vs MRU)

2. Consider the matrix_add function shown below:
   ```c
   int matrix_add(int a[128][128], int b[128][128], int c[128][128])
   {
       int i, j;
       for(i = 0; i < 128; i++)
           for(j = 0; j < 128; j++)
               c[i][j] = a[i][j] + b[i][j];
   return 0;
   }
   ```
   In each iteration, the compiled code will load a[i][j] first, and then load b[i][j]. After performing the addition of a[i][j] and b[i][j], the result will be stored to c[i][j]. The processor has a 64KB, 2-way, 64Byte-block L1 data cache, and the cache uses LRU policy once a set if full. The L1 data cache is write-back and write-allocate. If the addresses of array a, b, c are 0x10000, 0x20000, 0x30000 and the cache is currently empty, please answer the following questions:
   A. What is the L1 D-cache miss rate of the matrix_add function? How many misses are contributed by compulsory miss? How many misses are conflict misses?
   B. If the L1 hit time is 1 cycle, and the L1 miss penalty is 20 cycles. What is the average memory access time?

3. For caches of small size, a direct-mapped instruction cache can sometimes outperform a fully associative instruction cache using LRU replacement.
   A. Explain how this would be possible with an example access pattern.
   B. Come up with a replacement policy that makes a fully associative cache outperform a direct mapped cache for your access pattern in part A.
   C. Where does replacement policy fit into the three C’s model?

4. The processor design that you are working on currently has 95% instruction cache hit ratio, 80% data cache hit ratio, and 98% L2 cache hit ratio. Given a workload of 1 memory access in every 3 instructions, which portion of the hardware should you focus on to improve overall
performance: instruction cache's hit ratio, data cache's hit ratio, or L2 cache hit ratio? Assume the penalty for L1 cache miss is 10 cycles and L2 cache miss is 40 cycles.

5. For the following code, would a stream buffer or a victim buffer be more effective? Why?

   ```c
   void foo(List *head) {
       List * cur = head;
       while(cur->next) {
           cur = cur->next;
       }
   }
   ```

6. You are given a cache that has 16 byte blocks, 512 rows, and is 2-way set associative. Integers are 4 bytes. Give the C code for a loop that has a 100% miss rate in this cache but whose hit rate rises to almost 100% if you double the size of the cache.