1. Suppose a computer using direct mapped cache has 2^{32} words of word-addressable main memory, and a cache of 1024 blocks, where each cache block contains 32 words.
a. How many blocks of main memory are there?
b. What is the format of a memory address as seen by the cache, i.e., what are the sizes of the tag, block, and offset fields?
c. To which cache block will the memory address $000063FA_{16}$ map?

Ans.
a. $2^{32}/2^5 = 2^{27}$
b. 32 bit addresses with 17 bits in the tag field, 10 in the block field, and 5 in the offset field
c. 000063FA = 00000000000000000 1100011111 11010, which implies Block 31F

2. Suppose a word-addressable computer using set associative cache has 2^{16} words of main memory and a cache of 32 blocks, and each cache block contains 8 words.
a. If this cache is 2-way set associative, what is the format of a memory address as seen by the cache, that is, what are the sizes of the tag, set, and offset fields?
b. If this cache is 4-way set associative, what is the format of a memory address as seen by the cache?

Ans.
a. A total of 2^{16} words of main memory implies we have 16 bits in an address. Cache contains 2^5 blocks, but each set must have 2 blocks, so we have $2^5/2=2^4$ sets. Therefore our 16-bit address is divided into 9 bits for the tag field, 4 bits for the set field, and 3 bits for the offset field.
b. The 32 blocks in cache must now be divided into sets with 4 blocks each, implying we have only 8 sets. The tag field would now have 10 bits, the set field 3 bits, and the offset field 3 bits.

3. Consider a byte-addressable computer with 24-bit addresses, a cache capable of storing a total of 64K bytes of data and blocks of 32 bytes. Show the format of a 24-bit memory address for:
a. direct mapped
b. associative
c. 4-way set associative

Ans.
a. $64K = 2^6 2^{10} = 2^{16}$; $2^{16}/2^5 = 2^{11}$ blocks in cache, so 11 bits are needed for the block field. 5 bits are needed for the offset field, leaving 8 for the tag.
b. Again, 5 bits are needed for the offset field, leaving 19 for the tag.
c. There are $2^{11}/2^2 = 2^9$ sets in cache, so 9 bits are needed for the set field. We still need 5 bits for the offset field, leaving 10 for the tag field.

4. A direct-mapped cache consists of eight blocks. A byte-addressable main memory contains 4K blocks of eight bytes each. Access time for the cache is 22 ns and the time required to fill a cache slot from main memory is 300 ns (this time will allow us to determine the block is missing and bring it into cache). Assume a request is always started in parallel to both cache and to main memory (so if it is not found in cache, we do not have to add this cache search time to the memory access). If a block is missing from cache, the entire block is brought into the cache and the access is restarted. Initially, the cache is empty.

a. Show the main memory address format that allows us to map addresses from main memory to cache. Be sure to include the fields as well as their sizes.
b. Compute the hit ratio for a program that loops 4 times from locations 0 to 67_{10} in memory.
c. Compute the effective access time for this program.

Ans.
a. 4K = 212 blocks * 23 words each implies 215 words of main memory, so we have 15 bits in an address. Cache contains 23 blocks, so the block field requires 3 bits. Each block has 23 words, so the offset field requires 3 bits. That leaves 9 bits for the tag field.
Therefore our 15-bit address is divided into 9 bits for the tag field, 3 bits for the block field, and 3 bits for the offset field.

b. Block 0 of main memory (addresses 0 - 7) and Block 8 of main memory (addresses 64 - 67) must share a cache block. The remaining blocks are brought in and are not replaced. So for each access to Block 0, there is one miss and 7 hits. For each access to Block 8, there is one miss and 3 hits. The remaining blocks have one miss each, with all other accesses being hits. If we loop 4 times, we have:
Block 0: 4 misses, 28 hits
Block 1: 1 miss, 31 hits
Block 2: 1 miss, 31 hits
Block 3: 1 miss, 31 hits
Block 4: 1 miss, 31 hits
Block 5: 1 miss, 31 hits
Block 6: 1 miss, 31 hits
Block 7: 1 miss, 31 hits
Block 8: 4 misses, 12 hits for a total of 15 misses, 257 hits, or a hit ratio of 94.49%.

c. The effective access time is .9449(22ns) + .0551(300ns + 22ns)