4 Memory Hierarchy, Cache, Main Memory and Secondary Storage

1. Usage of cache memory can be justified by:
   (A) To increase the speed of cpu clock.
   (B) To improve memory cycle time.
   (C) To hold as many blocks of data as possible.
   (D) To match the speed of processor.

2. A set associative cache memory consists of 128 blocks divided into four block sets. The main memory consists of 16384 blocks and each block contains 256 eight bit words. Number of bits required for addressing the main memory, TAG, SET and WORD fields respectively are?
   (A) 24, 11, 5, 8
   (B) 22, 9, 5, 8
   (C) 25, 11, 6, 8
   (D) 25, 11, 5, 9

3. The access times of the main memory and the Cache memory, in a computer system, are 500 n sec and 50 n sec, respectively. It is estimated that 80% of the main memory request are for read the rest for write. The cache-hit-rate for the read access only is 90% and control can access cache and memory independently. If two different organizations are realized, one with a write back policy and second one with a write-through policy, the difference between the average times of the two policies is:
   (A) 0
   (B) 81
   (C) 85
   (D) 109

4. A computer system has a 16 K word cache organized in block-set-associative manner with 8 blocks per set, 64 words per block. The number of bits in the SET and WORD fields of the main memory address format is:
   (A) 15, 40
   (B) 6, 4
   (C) 7, 5
   (D) 5, 6

5. Consider a direct mapped cache of size 128 KB with block size 32 bytes. The CPU generates 40 bit addresses. The number of bits needed for cache tags and the number of bits for indexing are respectively,
   (A) 10, 17
   (B) 11, 24

(C) 23, 12
(D) 13, 22

6. Consider a computer with a 4-ways set-associative mapped cache of the following characteristics: a total of 1 MB of main memory, a word size of 1 byte, a block size of 128 words and a cache size of 8 KB. While accessing the memory location 0C795H by the CPU, the contents of the TAG field of the corresponding cache line is
(A) 000011000
(B) 110001111
(C) 00011000
(D) 110010101

7. If the associativity of a processor cache is doubled while keeping the capacity and block size unchanged, which one of the following is guaranteed to be NOT affected?
(A) Width of tag comparator
(B) Width of set index decoder
(C) Width of way selection multiplexer
(D) Width of processor to main memory data bus

8. The main memory of a computer has 2*c*m blocks while the cache has 2*c blocks. If the cache uses the set associative mapping scheme with 2 blocks per set, then block k of the main memory maps to the set
(A) (k mod m) of the cache
(B) (k mod c) of the cache
(C) (k mod 2c) of the cache
(D) (k mod 2 cm) of the cache

9. In designing a computer's cache system, the cache block (or cache line) size is an important parameter. Which one of the following statements is correct in this context?
(A) A smaller block size implies better spatial locality
(B) A smaller block size implies a smaller cache tag and hence lower cache tag overhead
(C) A smaller block size implies a larger cache tag and hence lower cache hit time
(D) A smaller block size incurs a lower cache miss penalty

10. More than one word are put in one cache block to:
(A) exploit the temporal locality of reference in a program
(B) exploit the spatial locality of reference in a program
(C) reduce the miss penalty

(D) None of the above

11. Consider a two-level cache hierarchy with L_1 and L_2 caches. An application incurs 2 memory accesses per instruction on average. For this application, the miss rate of L_1 cache is 0.2; the L_2 cache experiences, on average, 35 misses per 1000 instructions. The miss rate of L_2 expressed correct to two decimal places is :
(A) 0.0875
(B) 0.0667
(C) 0.175
(D) 0.0625

12. A CPU has 32-bit memory address and a 256 KB cache memory. The cache is organized as a 4-way set associative cache with cache block size of 16 bytes. What is the number of bits needed for TAG, SET and OFFSET information in the address?
(A) 16,12,4
(B) 14,14,4
(C) 15,13,4
(D) 16,11,5

13. Consider a 8-way set associative cache consisting of 1024 lines with a line size of 32 words. The CPU generates a 20-bit address of a word in main memory. The number of bits in the TAG, LINE and WORD fields are respectively:
(A) 9, 6, 5
(B) 8, 7, 5
(C) 7, 5, 8
(D) 9, 5, 6

14. Consider a 4-way set associative cache memory with 8 sets and total 32 cache blocks and a main memory with 256 blocks. What memory blocks will not be present in the cache after the following sequence of memory block references if MRU policy is used for cache block replacement. Assuming that initially the cache did not have any memory block from the current job?
16,48,74,135,199,231,56,224,31,12,247,74,231,112,255
(A) 31, 112, 224, 231.
(B) 199, 231, 247.
(C) 74, 112, 224, 231
(D) 31, 224, 231

15. A CPU has 32-bit memory address and a 1024 KB cache memory. The cache is organized as a 16-way set associative cache with cache block size of 32 bytes. What is the number and size of comparators required for tag matching?
(A) 4 comparators of size 16 bits.
(B) 8 comparators of size 32 bits.
(C) 16 comparators of size 16 bits.
(D) 32 comparators of size 32 bits.

16. A CPU has 32-bit memory address and a 2048 KB cache memory. The cache is organized in such a way that a set contains eight blocks with cache block size of 64 bytes. If there are 2 control bits required while storing tag bits in tag directory, what is the total amount of extra memory (in Kbytes) required for the tag bits?
(A) 16
(B) 24
(C) 28
(D) 32
(E) 64

17. Assume that for a certain processor, a read request takes 50 nanoseconds on a cache miss and 5 nanoseconds on a cache hit. Suppose while running a program, it was observed that 80% of the processor's read requests result in a cache hit. The average read access time in nanoseconds is:
(A) 12
(B) 14
(C) 18
(D) 22

18. Consider two cache organizations. First one is 32 kbytes 4-way set associative with 32 byte block size, the second is of same size but direct mapped. The size of an address is 32 bits in both cases . One valid/invalid bit and two control bits are also stored along with tag information. A 2-to-1 multiplexer has latency of 0.5ns while a K-bit comparator has latency of 0.5Kns The hit latency of the set associative organization is h_1 while that of direct mapped is h_2. The value of h_1 is:
(A) 8.5
(B) 8.9
(C) 9
(D) 10.5
(E) 20

19. In above question what is the value of h_2?
(A) 8.5

(B) 8.9
(C) 9.5
(D) 10.5
(E) 20

20. The memory access time is 1 nanosecond for a read operation with a hit in cache, 5 nanoseconds for a read operation with a miss in cache, 2 nanoseconds for a write operation with a hit in cache and 10 nanoseconds for a write operation with a miss in cache. Execution of a sequence of instructions involves 100 instruction fetch operations, 60 memory operand read operations and 40 memory operand write operations. The cache hit-ratio is 0.9. The average memory access time (in nanoseconds) in executing the sequence of instructions is:
(A) 1.54
(B) 1.68
(C) 2.25
(D) 3.36

21. For inclusion to hold between two cache levels L1 and L2 in a multi-level cache hierarchy, which of the following are necessary?
I. L1 must be write-through cache
II. L2 must be a write-through cache
III. The associativity of L2 must be greater than that of L1
IV. The L2 cache must be at least as large as the L1 cache
(A) IV only
(B) I and IV only
(C) I, II and IV only
(D) I, II, III and IV

22. A cache line is 64 bytes. The main memory has latency 10ns per word while the word is of size four bytes and bandwidth 8Gbits/s. The time required to fetch the entire cache line from the main memory is
(A) 64 ns
(B) 128 ns
(C) 96 ns
(D) 224 ns
(E) 288 ns

23. In a C program, an array is declared as: 'float A[8000]'. Each array element is 4 Bytes in size, and the starting address of the array is 0x00000000. This program is run on a computer that has a direct mapped data cache of size 8 Kbytes, with block (line) size of 16 Bytes. Which of the following elements of the array conflict with element A[111] in the data cache?

(A) A[4203]
(B) A[2159]
(C) A[6259]
(D) A[4303]

24. In a C program, an array is declared as: 'float A[8000]'. Each array element is 4 Bytes in size, and the starting address of the array is 0x00000000. This program is run on a computer that has a direct mapped data cache of size 8 Kbytes, with block (line) size of 16 Bytes. If the program accesses the elements of this array one by one in reverse order i.e., starting with the last element and ending with the first element, how many data cache misses would occur? Assume that array is already accessed once in normal order and no other data is fetched before the reverse access of array.
(A) 2048
(B) 1984
(C) 1023
(D) 2020
(E) 1536

25. In a C program, an array is declared as: 'long int A[6000]'. Each array element is 8 Bytes in size, and the starting address of the array is 0x00000000. This program is run on a computer that has a direct mapped data cache of size 8 Kbytes, with block (line) size of 64 Bytes. If the program accesses the elements of this array one by one till the the 2500^{th} of the array and then it starts accessing the array in reverse order from the current element till the first element followed by accessing the whole array, how many total misses took place in doing so?
(A) 1024
(B) 1180
(C) 1046
(D) 1248
(E) 1120

26. If the same sequence of events occurred in a 2-way set associative cache memory while remaining organization is kept intact, how many total misses occurred?
(A) 1024
(B) 1180
(C) 1046
(D) 1248
(E) 1120

27. Consider a small two-way set-associative cache memory, consisting of 8 blocks. For choosing the block to be replaced, use the least recently used

(LRU) scheme. The percentage of cache misses for the following sequence of block addresses is
8,10,12,14,6,14,8,0,5,13,5,9
(A) 50
(B) 45
(C) 75
(D) 95

28. Consider a small two-way set-associative cache memory, consisting of eight blocks. For choosing the block to be replaced, use the most recently used (MRU) scheme. The number of conflict misses for the following sequence of block addresses is
9,20,28,13,17,8,24,20,23,28,13,15,9,24
(A) 2
(B) 3
(C) 4
(D) 5
(E) 6

29. An 8KB direct-mapped write-back cache is organized as multiple blocks, each size of 32-bytes. The processor generates 32-bit addresses. The cache controller contains the tag information for each cache block comprising of the following.
1 valid bit
1 modified bit
As many bits as the minimum needed to identify the memory block mapped in the cache.
What is the total size of memory needed at the cache controller to store meta-data (tags) for the cache?
(A) 4864 bits
(B) 6144 bits
(C) 6656 bits
(D) 5376 bits

30. Consider a machine with a byte addressable main memory of 512kbytes. Assume that a direct mapped data cache consisting of 16 lines of 128 bytes each is used in the system. A 50 x 50 two-dimensional array of bytes is stored in the main memory starting from memory location 1000H. Assume that the data cache is initially empty. The complete array is accessed twice. Assume that the contents of the data cache do not change in between the two accesses. How many data misses will occur in total?
(A) 28
(B) 40
(C) 48

(D) 56

31. As per the data provided in above question, which of the following lines of the data cache will be replaced by new blocks in accessing the array for the second time?
(A) line 4 to line 11
(B) line 4 to line 7
(C) line 0 to line 3
(D) line 2 to line 5

32. A CPU has a 256 KB direct mapped cache with 256 byte-block size. Suppose A is two dimensional array of size as specified in the code below, with elements that occupy 8-bytes each. Consider the following two C code segments, P1 and P2.
P1:
for (i=0; i<=1023; i++)

for (j=0; j<2048; j++)

x +=A[i] [j];

P2:
for (i=0; i<2048; i++)

for (j=0; j<=1023; j++)

x +=A[j] [i];

P1 and P2 are executed independently with the same initial state, namely, the array A is not in the cache and i, j, x are in registers. Let the number of cache misses experienced by P1 be M1and that for P2 be M2.
The value of M1 is:
(A) 0
(B) 2048
(C) 16384
(D) 65536
(E) 2097152

33. This question is based on the same context as of above question. The value of the ratio $\frac{M_1}{M_2}$ when multiplied by sixteen is:

(A) 0.03125
(B) 0.125
(C) 0.25
(D) 0.5
(E) 1

34. An access sequence of cache block addresses is of length N and contains n unique block addresses. The number of unique block addresses between two consecutive accesses to the same block address is bounded above by k. What is the miss ratio if the access sequence is passed through a cache of associativity A≥k exercising least-recently-used replacement policy?
(A) $\frac{n}{N}$
(B) $\frac{1}{N}$
(C) $\frac{1}{A}$
(D) $\frac{k}{n}$

35. Consider a fully associative cache with 8 cache blocks (numbered 0-7) and the following sequence of memory block requests: 4, 3, 25, 8, 19, 6, 25, 8, 16, 35, 45, 22, 8, 3, 16, 25, 7
If LRU replacement policy is used, which cache block will have memory block 7?
(A) 2
(B) 3
(C) 5
(D) 6
(E) 7

36. Consider a fully associative cache with 8 cache blocks (numbered 0-7) and the following sequence of memory block requests: 4, 3, 25, 8, 19, 6, 25, 8, 16, 35, 45, 22, 8, 3, 16, 25, 7
If MRU replacement policy is used, which cache block will have memory block 7?
(A) 2
(B) 3
(C) 5
(D) 6
(E) 7

37. Consider a fully associative cache with 8 cache blocks (numbered 0-7) and the following sequence of memory block requests: 4, 3, 25, 8, 19, 6, 25, 8, 16, 35, 45, 22, 8, 3, 16, 25, 7
Which of the following represents the difference between the cache misses

which occurred when LRU and MRU replacement policies are used, respectively, for the above sequence?
(A) 0
(B) 1
(C) 2
(D) 3
(E) 4

38. A computer has a 256kBytes, 4-way set associative, write back data cache with block size of 64 bytes. The processor sends 40 bit addresses to the cache controller. Each cache tag directory entry contains, in addition to address tag, 2 valid bits, a modified bit and a replacement bit. The size of the cache tag directory is
(A) 28 Kbits
(B) 56 Kbits
(C) 112 Kbits
(D) 224 Kbits
(E) 512 Kbits

39. Consider a system with 2 level cache. Access times of Level 1 cache, Level 2 cache and main memory are 1 ns, 10 ns, and 500 ns, respectively. The hit rates of Level 1 and Level 2 caches are 0.8 and 0.9, respectively. What is the average access time of the system ignoring the search time within the cache?
(A) 13.0
(B) 12.8
(C) 12.6
(D) 12.4

40. Consider a system with three level cache organization. Access times of Level 1 cache, Level 2 cache, level 3 cache and main memory are 2 ns, 30 ns, 80 ns and 800 ns, respectively. The hit rates of Level 1, Level 2 and Level 3 caches are 0.85 and 0.9 and 0.98 respectively. What is the average access time of the system ignoring the search time within the cache?
(A) 5.5
(B) 6.25
(C) 6.8
(D) 7.2

41. Consider a machine with a 4-way set associative data cache of size 64 Kbytes and block size 32 bytes. The cache is managed using 44 bit virtual addresses and the page size is 64 Kbytes and 4 control bits are also stored with each tag entry. A program to be run on this machine begins as follows:

```
double ARR[1024][2048];
int i, j;
/*Initialize array ARR to 0.0 */
for(i = 0; i ¡ 1024; i++)
for(j = 0; j ¡ 2048; j++)
ARR[i][j] = 0.0;
```

The size of double is 8 bytes. Array ARR is located in memory starting at the beginning of virtual page 0xFFAAA000000 and stored in row major order. The cache is initially empty and no pre-fetching is done. The only data memory references made by the program are those to array ARR.
The total size of the tags in the cache directory is:
(A) 52 Kbits
(B) 54 Kbits
(C) 60 Kbits
(D) 68 Kbits

42. Consider a machine with a 4-way set associative data cache of size 64 Kbytes and block size 32 bytes. The cache is managed using 44 bit virtual addresses and the page size is 64 Kbytes and 4 control bits are also stored with each tag entry. A program to be run on this machine begins as follows:

```
double ARR[1024][2048];
int i, j;
/*Initialize array ARR to 0.0 */
for(i = 0; i ¡ 1024; i++)
for(j = 0; j ¡ 2048; j++)
ARR[i][j] = 0.0;
```

The size of double is 8 bytes. Array ARR is located in memory starting at the beginning of virtual page 0xFFAAA000000 and stored in row major order. The cache is initially empty and no pre-fetching is done. The only data memory references made by the program are those to array ARR.
Which of the following array elements have the same cache index as ARR[5][0]?
(A) ARR[10][4]
(B) ARR[4][0]
(C) ARR[25][0]
(D) ARR[27][0]
(E) ARR[26][0]

43. Consider a machine with a 4-way set associative data cache of size 64

Kbytes and block size 32 bytes. The cache is managed using 44 bit virtual addresses and the page size is 64 Kbytes and 4 control bits are also stored with each tag entry. A program to be run on this machine begins as follows:

```
double ARR[1024][2048];
int i, j;
/*Initialize array ARR to 0.0 */
for(i = 0; i ¡ 1024; i++)
for(j = 0; j ¡ 2048; j++)
ARR[i][j] = 0.0;
```

The size of double is 8 bytes. Array ARR is located in memory starting at the beginning of virtual page 0xFFAAA000000 and stored in row major order. The cache is initially empty and no pre-fetching is done. The only data memory references made by the program are those to array ARR.
The cache hit ratio for this initialization loop is
(A) 0%
(B) 25%
(C) 50%
(D) 75%