1. In a k-way set associative cache, the cache is divided into v sets, each of which consists of k lines. The lines of a set are placed in sequence one after another. The lines in set s are sequenced before the lines in set (s+1). The main memory blocks are numbered 0 onwards. The main memory block numbered j must be mapped to any one of the cache lines from.
   a. **(j mod v) * k to (j mod v) * k + (k-1)**
   b. (j mod v) to (j mod v) + (k-1)
   c. (j mod k) to (j mod k) + (v-1)
   d. (j mod k) * v to (j mod k) * v + (v-1)

   Sol: Number of sets in cache = v. So, main memory block j will be mapped to set (j mod v), which will be any one of the cache lines from (j mod v) * k to (j mod v) * k + (k-1).
   (Associativity plays no role in mapping- k-way associativity means there are k spaces for a block and hence reduces the chances of replacement.)

2. A computer system has an L1 cache, an L2 cache, and a main memory unit connected as shown below. The block size in L1 cache is 4 words. The block size in L2 cache is 16 words. The memory access times are 2 nanoseconds. 20 nanoseconds and 200 nanoseconds for L1 cache, L2 cache
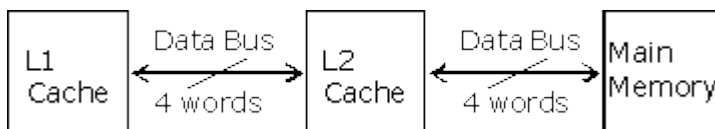
   and main memory unit respectively. When there is a miss in L1 cache and a hit in L2 cache, a block is transferred from L2 cache to L1 cache. What is the time taken for this transfer?
   a. 2ns   b. 20ns   c. 22 ns   d.88ns

   Answer: 20 + 2 = 22 ns

3. A CPU has a cache with block size 64 bytes. The main memory has k banks, each bank being c bytes wide. Consecutive c − byte chunks are mapped on consecutive banks with wrap-around. All the k banks can be accessed in parallel, but two accesses to the same bank must be serialized. A cache block access may involve multiple iterations of parallel bank accesses depending on the amount of data obtained by accessing all the k banks in parallel. **Each iteration requires decoding the bank numbers to be accessed in parallel and this takes. k/2 ns** The latency of one bank access is 80 ns. If c = 2 and k = 24, the latency of retrieving a cache block starting at address zero from main memory is:
   a. 92ns   b. 104ns   c. 172ns   d. 184ns

   Sol : Size of cache block=64 B
   No. of main memory banks K=24
   Size of each bank C=2 bytes
   i.e each bank in memory is 2 bytes and there are 24 such banks. So, in one iteration we can get 2*24 = 48 bytes and getting 64 bytes requires 2 iterations.
   So time taken for parallel access T=decoding time +latency time.
   T = (K/2)+latency
   = 12+80 = 92 ns
   But C=2 for accesses
   =2*92=184ns
   (since in each iteration we need to select the banks and the bank decoding time (k/2) is independent of the number of banks we are going to access)

4. If a processor generates a request for an invalid cache block then, it is treated as
   a. Normal read miss, placed on the bus
   b. Abnormal read miss, placed on the bus
   c. Normal read hit, and the requested word is placed on the bus
   d. Abnormal read hit, and the requested word is placed on the data bus

   Answer a

5. GPU consists of
   a. SISD processors   b. MISD processors   c. MIMD processors   SIMD processors

6. A CPU has a five-stage pipeline and runs at 1 GHz frequency. Instruction fetch happens in the first stage of the pipeline. A conditional branch instruction computes the target address and evaluates the condition in the third stage of the pipeline. The processor stops fetching new instructions following a conditional branch until the branch outcome is known. A program executes 10^9 instructions out of which 20% are conditional branches. If each instruction takes one cycle to complete on average, the total execution time of the program is:

   a. 1 Second   b. 1.2 Second   c. 1.4 Second   d. 1.6 Second

   Sol: In the 3rd stage of pipeline, there will be 2 stall cycles i.e. 2 delay slots.
   Total number of instructions = 10^9
   20% out of 10^9 are conditional branches.
   Therefore, Cycle penalty = $0.2 * 2 * 10^9 = 4 * 10^9$
   Clock speed is 1 GHz and each instruction on average takes 1 cycle.
   Total execution time = $(10^9 / 10^9) + 4 * (10^8 / 10^9)$
   = 1.4 seconds

7. A computer has a 256 KByte, 4-way set associative, write back data cache with block size of 32 Bytes. The processor sends 32 bit addresses to the cache controller. Each cache tag directory entry contains, in addition to address tag, 2 valid bits, 1 modified bit and 1 replacement bit.
   The number of bits in the tag field of an address is
   a. 11   b. 14   c. 16   d. 27

8. In X = (M + N x O)/(P x Q), how many one-address instructions are required to evaluate it?
   a. 10
   b. 12
   c. 9
   d. None of these
   Ans: 8

9. The minimum time delay between the initiation of two independent memory operations is called
   a. Latency time   b. Cycle Time   c. Access Time   d. Rotational Time
   Ans: Cycle Time

10. A pipeline P operating at 400 MHz has a speedup factor of 6 and operating at 70% efficiency. How many stages are there in the pipeline?
    a. 5   b. 6   c. 8   d.9
    Sol: Efficiency = Speedup factor/ Number of stages
    0.7 = 6 / Number of stages
    Number of stages = 8.56 = 9; So, option (D) is correct.