IIT Guwahati - Department of Computer Science & Engineering

CS 223- Computer Organization & Architecture

Practice Tutorial Sheet #3 [21.05.2020]

1. A cache has access time (hit latency) of 10 ns and miss rate of 5%. An optimization was made to reduce the miss rate to 3% but the hit latency was increased to 15 ns. Under what condition this change will result in better performance (Lower AMAT)?
2. A cache has hit rate of 95%, block size of 128B, cache hit latency of 5ns. Main memory takes 50 ns to return first word (32 bits) of a block and 10 ns for each subsequent word.
   (a) What is the miss latency of the cache?
   (b) If doubling the cache block size reduces the miss rate to 3%, does it reduce AMAT?
3. A 16KB direct mapped 256B block unified cache is attached to a 16MB main memory system. The word length as well as instruction length of the processor is 16 bits. Consider a program that consists of a main routine M which in turn calls a subroutine S. M consists of 12 instruction words which are loaded in the main memory from the address 0x4230FA onwards. The last five instructions of M is a loop that is iterated 10 times. The second instruction in the loop is a call to subroutine S. S consists of 4 instruction words loaded in the main memory from the address 0x70F168. The last instruction of S is a subroutine return back to M. The only two data words that are used by M and S are at addresses 0x748074 and 0x846064. Assume the caches are initially empty. Ignore OS level interruption and subsequent cache impact on context switching.
   (a) Find the number of cache misses occurred during the execution of the program.
   (b) How many cache block evictions happened during the execution of the program?
   (c) List out the block numbers (in decimal) in the cache that are non-empty after the execution of the program.
4. Consider an 8-way set associative cache that uses pseudo LRU block replacement policy. Assume all the cache blocks are initially empty and filling up of empty blocks in a given cache set happens from way 0 to way-7. Consider the following 14 block numbers all mapped to a particular set n given in the order of arrival. Show the organization of set n (way number & the block number residing in the way) after servicing these requests.
   A, B, C, A, D, E, B, F, G, C, H, A, E, Q.
5. There are two variants (alpha and beta) of cache memory organization that is possible in a processor. Alpha is a split cache organization with 16 KB instruction cache and a 16 KB data cache. Both the I and D caches have its own data port. Beta is a 32 KB unified cache organization that has only one cache port, so that a load or store hit takes 1 extra clock cycle at MEM stage due to structural hazard with IF stage. Assume a hit takes 1 clock cycle and the miss penalty is 50 clock cycles. For a program that has 40% load-store instructions, the MPKI of I-cache and D-cache of alpha are 40 and 120, respectively, whereas the MPKI on unified cache of beta is 200. Which one beta or alpha has the higher AMAT? Explain with proper justifications.