IIT Guwahati - Department of Computer Science & Engineering

CS 222- Computer Organization & Architecture – Tutorial #1 (06.04.2019)

1. Given a non-pipelined architecture running at 1.5 GHz, that takes 5 cycles to finish an instruction. You want to make it pipelined with 5 stages. Due to hardware overhead the pipelined design will operate only at 1 GHz. 5% of memory instructions cause a stall of 50 cycles, 30% of branch instruction cause a stall of 2 cycles and load-ALU combinations cause a stall of 1 cycle. Assume that in a given program, there exist 20% of branch instructions and 30% of memory instructions. 10% of instructions are load-ALU combinations. What is the speedup of pipelined design over the non-pipelined design?
2. A program has 2000 instructions in the sequence L.D, ADD.D, L.D, ADD.D,..... L.D, ADD.D. The ADD.D instruction depends on the L.D instruction right before it. The L.D instruction depends on the ADD.D instruction right before it. If the program is executed on the 5-stage pipeline what would be the actual CPI with and without operand forwarding technique?
3. Consider a 4-way associative cache that can be operated in one among the two modes at a time. In mode-1 it uses pseudo LRU block replacement policy and in mode-2 it uses Last In First Out 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-3. Consider the following 14 block numbers all mapped to a particular set n given in the order of arrival.
   A, B, C, D, A, B, E, F, A, B, F, C, D, A.
   (a) Find the number of cache misses (excluding compulsory misses) in mode-1.
   (b) Draw the pseudo LRU tree for set n after processing these requests in mode-1.
   (c) Find the number of cache misses (including compulsory misses) in mode-2.
   (d) Draw the content of set n (way-0 to way-3) after processing these requests in mode-2.
4. Consider a computer system based on 32 bit processor with 8KB direct mapped on-chip I-cache and 16KB 2-way set associative on-chip D-cache. The off-chip unified cache is 128 KB 4-way set associative. Block size for on-chip cache is 8 words and off-chip cache is 16 words. Four fixed length instruction starting at main memory word address 22 (in decimal) are executed. These instructions refer to data at main memory word address 260, 261 and 275 (all in decimal). Assuming caches are initially empty; indicate the non empty blocks on all the caches after execution of the instruction.
5. Consider a cache system with miss rate of an I-cache is 2% and that of D-cache is 4%. The processor CPI=2 without memory stalls and miss penalty =100 cycles for all misses. Determine how much faster the processor would run with a perfect cache that never missed. Assume frequency of all loads and store is 36 %.
6. Assume a 2-level cache system with the following specifications. L1 Hit Time = 1 cycle, L1 Miss Rate = 2.5%, L2 Hit Time = 6 cycles, L2 Miss Rate = 17% (% L1 misses that miss), L2 Miss Penalty = 120 cycles. Compute the average memory access time.
7. 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)?
8. 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 reduces AMAT?