Homework 5: 100 Points

This Homework contains examples on caches, memory hierarchy design, and cache optimization.

Please submit your answers in a document and upload it in Canvas by the due date.

Present your work neat, clear, organized, logical, and make your case; don’t leave it to our interpretation of your answer. There is 10% penalty per day for delay in submitting the homework.

1. (16 Points) Exercise B.1 Parts a and b
2. (28 Points) Exercise B.5 Appendix B. Assume L1 cache Write-through with no write-allocate and L2 cache Write-back with write-allocate.

Hint: A useful tool for solving this type of problem is to extract all of the available information from the problem description. It is possible that not all of the information will be necessary to solve the problem, but having it in summary form makes it easier to think about.

1. (16 Points) Consider the code segment:

fld f0, 0(x0) ; load f0 from address 0+x0

Loop: fld f2, 0(x2) ; load f2 from address 0+x2

fmult f2, f2, f0 ; f2 = f2 \* f0

fsd f2, 0(x2) ; store f2 at address 0+x2

addwi x2, x2, 8 ; x2 = x2 +8

subwi x4, x3, x2 ; x4 = x3 – x2

bne x4, x0, loop ; branch to loop if x4!= 0

Assume the initial value of x3 is x2 + 64. Assume x0 =0 and x2=16, and the memory contains:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M[0] | M[8] | M[16] | M[24] | M[32] | M[40] | M[48] | M[56] | M[64] | M[72] | M[80] | M[88] |
| 4.0 | 1.0 | 7.0 | 9.0 | 5.0 | 3.0 | 1.0 | 2.0 | 5.0 | 8.0 | 7.0 | 3.0 |

0 8 16 24 32 40 48 56 64 72 80

The cache size is 64 bytes and the block size is 16 bytes. Make table similar to Example 5 Lecture Slides.

1. Display the content of memory, cache hit or miss status, and the content of cache as the loop executes, assuming the cache is direct mapped, with write through and no write allocate.
2. Repeat Part (a) assuming cache is a 2-way set associative, with FIFO replacement policy, and write back, and write allocate.
3. (8 Points) A computer has a L1 data cache and a L1 instruction cache. The latencies of the different kinds of accesses are as follows: cache hit 1cycle, cache miss 16 cycles. If the miss rate for instructions is 5% and the miss rate for data is 6%, what is the average memory access time in cycles?
4. (16 Points) Assume we have a 512 byte cache with 64 byte blocks. We also assume that the main memory is 2KB large. We can regard the memory as an array of 64-byte memory blocks M0, M1, M2, …M31. The table below displays the memory blocks that can reside in different cache blocks if the cache was fully associative.

|  |  |  |
| --- | --- | --- |
| Cache Block | Set | Memory blocks that can reside in cache blocks |
| 0 | 0 | M0, M1, M2, …. M31 |
| 1 | 0 | M0, M1, M2, …. M31 |
| 2 | 0 | M0, M1, M2, …. M31 |
| 3 | 0 | M0, M1, M2, …. M31 |
| 4 | 0 | M0, M1, M2, …. M31 |
| 5 | 0 | M0, M1, M2, …. M31 |
| 6 | 0 | M0, M1, M2, …. M31 |
| 7 | 0 | M0, M1, M2, …. M31 |

1. Show the elements of the table if cache is organized as a direct mapped cache.
2. Show the elements of the table if cache is organized as a two-way set associative cache.
3. (8 Points) Assume we have a computer where the CPI=1 when all memory accesses hit in cache. Only load and store access the data cache, and these total 20% of instructions. If the cache miss rate is 2% and the miss penalty is 25 cycles, how faster would the computer be if all instructions were cache hits?
4. (8 Points) Let’s assume a computer with L2 cache that has a 256-byte cache block. It takes 10 clock cycles to transfer the first 8 bytes to L1, and then 1 clock cycle per 8 bytes to transfer the rest of the block. Compare the miss penalty times with and without Critical Word First?