REVL WALTER MBOGO NJUKI – SCT212-0053/2020

Computer Architecture - tutorial 5 [SOLUTIONS]

Context, Objectives and Organization

The goal of the quantitative exercise in this tutorial is to explore qualitatively and quantitatively some hardware and software optimizations to improve cache performance.

E1: CAR September 2003 exam P2 - groups of 2 - 35 min

Problem

Consider a computer system with a first-level data cache with the following characteristics: size: 16KBytes; associativity: direct-mapped; line size: 64Bytes; addressing: physical.

The system has a separate instruction cache and you can ignore instruction misses in this problem. This system is used to run the following code:

```
for (i=0; i<4096; i++)
   X[i] = X[i] * Y[i] + C
```

Assume that both X and Y have 4096 elements, each consisting of 4 bytes (single precision floating point). These arrays are allocated consecutively in physical memory. The assembly code generated by a naive compiler is the following:

```
loop: lw f2, 0(r1)          # load X[i]
        lw f4, 0(r2)        # load Y[i]
      multd f2, f2, f4      # perform the multiplication
      addd f2, f2, f0       # add C (in f0)
       sw 0(r1), f2         # store the new value of X[i]
      addi r1, r1, 4        # update address of X
      addi r2, r2, 4        # update address of Y
      addi r3, r3, 1        # increment loop counter
```

bne r3, 4096, loop # branch back if not done

a. How many data cache misses will this code generate? Breakdown your answer into the three types of misses. What is the data cache miss rate?

b. Provide a software solution that significantly reduces the number of data cache misses. How many data cache misses will your code generate? Breakdown the cache misses into the three types of misses. What is the data cache miss rate?

c. Provide a hardware solution that significantly reduces the number of data cache misses. You are free to alter the cache organization and/or the processor. How many data cache misses will your code generate? Breakdown the cache misses into the three types of misses. What is the data cache miss rate?

Solution

A.

Cache size: 16 KB, Line size: 64 bytes, Direct-mapped.
16 KB / 64 B = 256 lines
Address breakdown: 6 bits for block offset, 8 bits for index, 18 bits for the tag.
For each iteration of the loop, three memory accesses occur: Load X[i], Load Y[i], Store X[i].

X and Y each contain 4096 elements (16 KB total for each array), and they are mapped to the same cache location.

Cold misses for X: 4096 / 16 = 256
Cold misses for Y: 4096 / 16 = 256
Conflict misses for X: 4096
Conflict misses for Y: 4096

Total misses = 256 + 256 + 4096 + 4096 = 8704 misses.
Miss Rate = 8704 / 12288 = 0.71 or 71%.

B.

By processing elements in blocks, we can improve spatial locality, reducing conflict misses. The following code snippet implements loop blocking, where 16 elements of both X and Y are processed in a block:

```
for (i = 0; i < 4096; i += 16)
    for (j = 0; j < 16; j++)
        X[i+j] = X[i+j] * Y[i+j] + C;
```

Miss Breakdown:

Cold misses for X: 256
Cold misses for Y: 256
Conflict misses for X: 256
Conflict misses for Y: 256

Total misses = 256 (cold misses for X) + 256 (cold misses for Y) + 256 (conflict misses for X) + 256 (conflict misses for Y) = 1024 misses.

Miss Rate = 1024 / 12288 = 0.083 or 8.3%.

C.

Alternative 1: Double the cache size to 32KB

With the cache size doubled to 32KB and cache lines remaining at 64 bytes, the cold misses for X and Y are:

- Cold Misses: 2 × (4096 / 16) = 512 total cold misses.
- Conflict Misses: No conflict misses because the increased cache size allows X and Y to coexist without evictions.

Total Misses = 512 (cold) + 0 (conflict) = **512 misses**
Miss Rate = 512 / 12288 = **0.0417** or **4.17%**

Alternative 2: Make the cache set associative

In a 2-way set-associative cache:

- Cold Misses: Remains the same as the direct-mapped cache, 512 total cold misses.
- Conflict Misses: No conflict misses due to the cache being set-associative.

Total Misses = 512 (cold) + 0 (conflict) = **512 misses**
Miss Rate = 512 / 12288 = **0.0417** or **4.17%**

If the cache block size is increased by a factor of zzz:

- Cold Misses: Cold misses decrease by a factor of zzz, so the cold misses are 4096 / (16 × zzz).
  For example, if z=2z = 2z=2, cold misses = 128.
- Conflict Misses: Conflict misses increase due to larger blocks causing evictions. With z=2z = 2z=2, conflict misses = 8192.

Total Misses = Cold Misses + Conflict Misses = 128 + 8192 = **8320 misses** (for z=2z = 2z=2)
Miss Rate = 8320 / 12288 = **0.676** or **67.6%**

Boris Grot 2018. Thanks to Vijay Nagarajan, Nigel Topham and Marcelo Cintra.