CS110 sp24 HW5

Due: 14^{th} May

You should finish this homework either by writing it **neatly** by hand or using LaTeX (**highly recommended!!!**). You can find the `.tex` file on Piazza.

1 T/I/O Breakdown

1. Given that we have a direct-mapped byte-addressed cache with capacity 32B and block size of 8B. Of the 32 bits in each address, which bits are offset bits? Which bits are index bits? What about tag? *Note: Please provide your answer in the format [n:m] to denote the range from the m^{th} bit to the n^{th} bit (e.g., [1:0] represents the two lowest bits).*

Tag	Index	Offset
$[31:5]$	$[4:3]$	$[2:0]$

2. Given the cache in question 1.1, assuming that we will access memory addresses in the following order, classify each of the accesses as a cache hit (H), cache miss (M) or cache miss with replacement (R). Ignore miss types for now. *Note: The distinction of M and R here is just for your understanding, and that the cache doesn't behave differently for these cases.*

Address	Hit, Miss, Replace	Miss Type
0x00000004	M	Compulsory
0x00000005	H	
0x00000068	M	Compulsory
0x000000C8	R	Compulsory
0x00000068	R	Conflict
0x000000DD	M	Compulsory
0x00000045	R	Compulsory
0x000000CF	R	Conflict
0x000000F3	M	Compulsory

2 Set-Associative Caches

Given that we have a 2-way set associative cache. This time we have an 8-bit address space, 8B blocks, and a cache size of 32B. Classify each of the accesses as a cache hit (H), cache miss (M) or cache miss with replacement (R). Assume that we have an LRU replacement policy. Ignore miss types for now.

Address	Hit, Miss, Replace	Miss Type
0b0000 0100	M	Compulsory
0b0000 0101	H	
0b0110 1000	M	Compulsory
0b1100 1000	M	Compulsory
0b0110 1000	H	
0b1101 1101	R	Compulsory
0b0100 0101	M	Compulsory
0b0000 0100	H	
0b0011 0000	R	Compulsory
0b1100 1011	R	Capacity
0b0100 0010	R	Capacity

3 The 3C's Cache Misses

Go back to question 1 and 2 and classify each miss as one of the three types of misses.

4 Code Analysis

Consider the following function that takes in two integer arrays, a (of length a_len) and b (of length b_len), and returns the 1D convolution of a and b. Assume results is properly allocated. Let a=0x1000, b=0x2000, results=0x3030, a_len=4, and b_len=2. *Note: The register keyword in C provides a hint to the compiler to consider storing a variable in a processor register.*

```
void convolve_1d(int* a, int a_len, int* b, int b_len, int*
    results) {
     for (int i = 0; i < a_len - b_len + 1; i++) {
          register int sum = 0;
          for (int j = 0; j < b_len; j++) {
               sum += b[j] * a[i + j];
          }
          results[i] = sum;
     }
}
```

1. Given that we have a single-level, direct-mapped 64B cache with 16B blocks and 16-bit addresses. What is the overall hit rate for a call to convolve_1d?

Answer:
hit rate $= \frac{2}{15}$

- #Cache line $= \frac{\text{Cache Size}}{\text{Block Size}} = \frac{64B}{16B} = 4$,
  direct-mapped $\implies$ Size of index $= \log_2(\#\text{Cache line}) = 2$
- Size of offset $= \log_2(\#\text{Block Size}) = 4$
- Size of tag = 16 - Size of index - Size of offset = 10
- Elements in array a and b have the same index 0. Because the elements in the a.b array are accessed alternately in the inner loop, all access to the array a.b require reloading the block into the cache.

  $$3(\#\text{outer loop}) * 2(\#\text{inner loop}) * 2(\text{access to b and a}) = 12$$

  12 Misses/12 Accesses
- Index of result is 3, only access to result[0] is Miss.
  1 Miss/3 Accesses

  $$\text{hit rate} = \frac{(12+3)-(12+1)}{12+3} = \frac{2}{15}$$

2. Given that we have a 2-way set associative cache of the same size with a LRU replacement policy. What is the overall hit rate for a call to convolve_1d?

Answer:
hit rate $= \frac{12}{15} = 80\%$

- 2-way set associative $\implies$ Size of index $= \log_2(\frac{1}{2}\#\text{Cache line}) = 1$
- Size of offset $= \log_2(\#\text{Block Size}) = 4$
- Size of tag = 16 - Size of index - Size of offset = 11
- Due to 2-way set associative cache, a and b's block have no conflict when loadING in the cache.
- Block Size is 16B, so elements in one array can be obtained within one block. There are 3 Misses that only happen in the first access to each array.

  $$\text{hit rate} = \frac{15-3}{15} = \frac{12}{15}$$

3. Given that we have a fully associative cache of the same size with a LRU replacement policy. What is the overall hit rate for a call to convolve_1d?

Answer:
hit rate $= \frac{12}{15} = 80\%$

- Since there's no miss with replacement in 2-way set associative, result of fully associative is same with it.

5 AMAT

1. In a 2-level cache system, if L1 has a local miss rate of 50% and the global miss rate of L2 is 20%, what is the local miss rate of L2?

Answer:

$$\begin{aligned}\text{global miss rate} &= \frac{\#\text{Misses in all levels}}{\#\text{Total Accesses}} \\ &= \frac{\#\text{Misses in L2}}{\#\text{Accesses in L2}} \times \frac{\#\text{Misses in L1}}{\text{Total Accesses}} \\ &= \text{L2's local miss rate} \times \text{L1's local miss rate} \\ \text{L2's local miss rate} &= \frac{20\%}{50\%} = 40\%\end{aligned}$$

Suppose your system consists of:

1. An L1 that has a hit time of 2 cycles and has a local miss rate of 20%.
2. An L2 that has a hit time of 15 cycles and has a global miss rate of 5%.
3. Main memory where accesses take 100 cycles.

2.What is the AMAT of the system?

Answer:

$$\begin{aligned}\text{L2's local miss rate} &= \frac{\text{global miss rate}}{\text{L1's local miss rate}} = 25\% \\ \text{AMAT} &= \text{L1's hit time} + \\ &\quad \text{L1's local miss rate} \times (\text{L2's hit time} + \text{L2's local miss rate} \times \text{miss penalty}) \\ &= 2 + 20\% * (15 + 25\% * 100) \\ &= 10 \text{ cycles}\end{aligned}$$

3. Suppose we want to reduce the AMAT of the system to 8 cycles or lower by adding in a L3. If the L3 has a local miss rate of 25%, what is the largest hit time that the L3 can have?

Answer:

$$\begin{aligned}\text{AMAT} &= \text{L1's hit time} + \\ &\quad \text{L1's local miss rate} \times (\text{L2's hit time} + \text{L2's local miss rate} \times \\ &\quad (\text{L3's hit time} + \text{L3's local miss rate} \times \text{miss penalty})) \\ &\quad \leq 8 \text{ cycles} \\ \text{AMAT} &= 2 + 20\% * (15 + 25\% * (X + 25\% * 100) \\ &= 6.25 + 0.05X \leq 8 \text{ cycles} \\ X &\leq 35 \text{ cycles}\end{aligned}$$