CS110 sp24 HW5 Solution

Due: TBD

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).* (5 pts)

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.* (18 pts, 2 for each)

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. (22 pts, 2 for each)

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	Conflict
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. (20 pts, 10 for each)
Note: Since miss type clarification is not in the exam scope, you will receive full credit as long as you finish this section.

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? (7 pts)
Solution: $\frac{2}{15}$
There are total three iterations of the outer loop, two iterations of the inner loop and two memory access in each inner loop: reading `b[j]` and `a[i + j]`. There are also three memory access in each outer loop: writing to `results[i]`. Since the blocks of `a` and `b` have the same index, the accesses to a and b will always be misses (first two are compulsory and the rest are conflict). The accesses to `results` do not cause thrashing, there is a compulsory miss at first and then two hits, giving the total hit rate $\frac{2}{15}$.

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? (7 pts)
Solution: $\frac{4}{5}$
Since now we have two cache lines in the same index, the accesses to `a` and `b` will not thrash. We only have three compulsory misses (one for each array) and the rest are hits, giving us a total hit rate $\frac{4}{5}$.

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? (5 pts)
Solution: $\frac{4}{5}$
Similar to question4.2.

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? (5 pts)
Solution: 40%
Local miss rate = $\frac{\text{L2 gloabal miss rate}}{\text{L2 access times}} = \frac{20\%}{50\%} = 40\%$.

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? (5 pts)
Solution: 10 cycles
AMAT = $2 + 20\% \times (15 + 25\% \times 100)$= 10 cycles.

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? (6 pts)
Solution: 10 cycles
AMAT $= 2 + 20\% \times (15 + 25\% \times (H + 25\% \times 100)) \leq 8$
Solving for H, we find that $H \leq 35$, so the largest hit time is 35 cycles.