## HW7.4. Cache Programming

Q1: The following C program is run (with no optimization) on a processor with a direct-mapped data cache (**the cache starts out empty**) with a <u>size of 1 KiB and a block size of 32 bytes</u>:

```
int i, j, array[256*256];

/* ... */

for (i = 1 ; i < 256 ; i++) {
    for (j = 0 ; j < 256 ; j++) {
        array[256*j] += array[256*j+i];
    }
}</pre>
```

Assume sizeof(int) == 4 and array ==  $0 \times 0000 \times 4000$ .

The sizeof(int) is highly relevant for the following questions. <u>How many ints can fit in a cache block?</u>

Q1.1: For the first iteration of the outer loop (i = 1), what is the hit rate of this code? Hint: How would you expand the expression array[x] += array[y]? How many memory accesses are in there? Which of those would miss?



Q1.2: After the first n iterations of the outer loop, the hit rate changes. What is n? Hint: Try writing out the array indices for every iteration of the loop. Recall the sizeof(int). At what point will array[256\*j+i] exceed a cache block?



Q1.3: After the first n iterations of the outer loop, the hit rate changes. What is the new hit rate of each iteration of the outer loop?

Hint: From 1.2: if array[256\*j+i] exceeds a cache block, how many misses would we have?



Q1.4: What is the overall hit rate of this code?

Hint: How many total memory accesses do we have for one iteration of the outer loop? How would you incorporate the pattern you observed from the previous questions?



Q2: We decide to rewrite our code to be more cache-efficient, while maintaining the same behavior. Fill in the blanks of this code (write your answers in the same format as the original code):

```
int i, j, array[256*256];

/* ... */

for ([CODE A]; i < 256 ; i++) {
    for ([CODE B]; j < 256 ; j++) {
        array[[CODE C]] += array[[CODE D]];
    }
}</pre>
```

Hint: Coming from the previous questions, we now want to have more consecutive memory accesses on the inner loop so that we get more cache hits.





