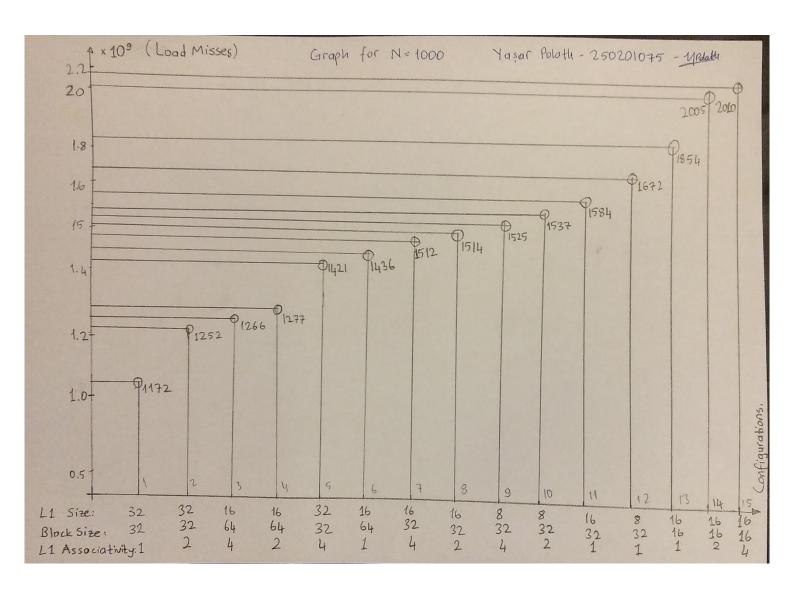
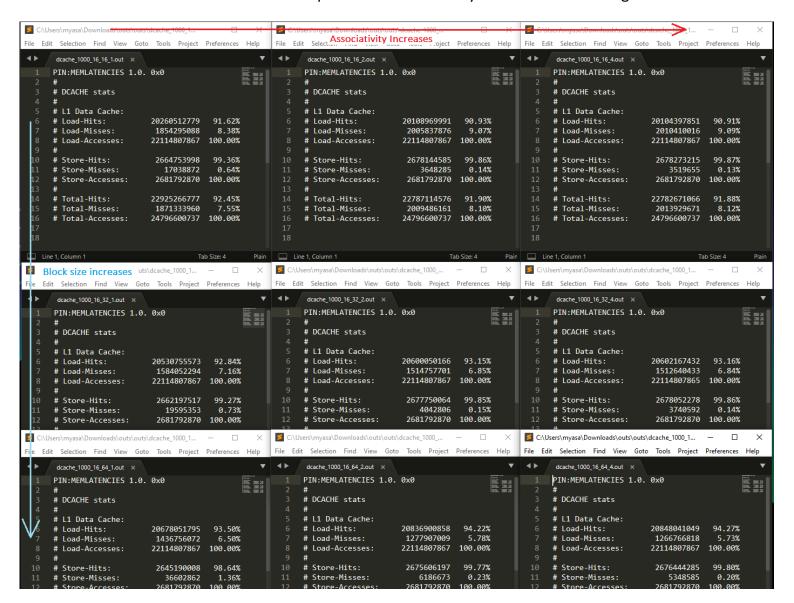
CENG311 Fall 2020

REPORT FOR PROGRAMMING ASSIGNMENT 4

Firstly, let us start the report with default matrix size (N=1000). Here is the graph that shows number of load-misses for L1 data cache:

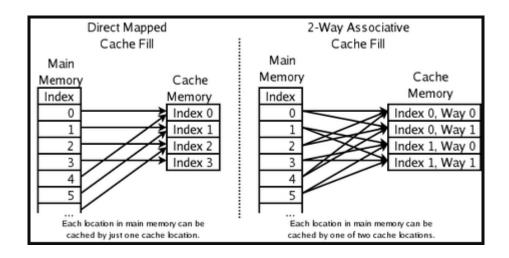


In order to comment the cache parameters differently let's look at this image:



As you can see when the associativity becomes larger (red line), miss rate getting smaller. Because the associativity means more room for data (2 associativity means, 2 blocks for 1 data) to be placed. And it increases the hit rate, decreases the misses naturally.

Some pictures from lecturer materials that show what I try to explain:





Associativity allows a set of blocks to go in cache entries. When the one of block is full, data can be located another block which is element of the same set.

There is a coincidence to must indicate is that when the cache size and block size is equal (when they both are 16 or 32), miss rate is getting bigger with associativity. I am not sure, but this might happen because block size becomes a significant portion of the cache size (actually their size is equal) and we cached lots of values (N=1000) that we use just once.

When the cache size and block size are both small there is high miss rate is inevitable. To lower these misses associativity can be increased.

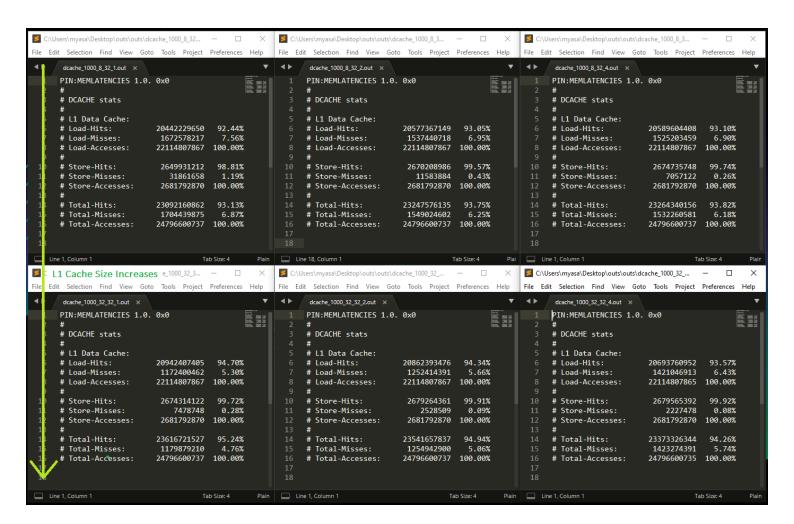
Second part of my commenting, I will explain block size differences.

When the block size is small, <u>we cannot use spatial locality</u>. <u>This situation occurs more misses.</u>

As you can see (blue line) when the block size increases, misses getting smaller.

And moreover, the block size matters a lot because the row-majority of matrix implementation. We can access continuous elements of matrix easily when the block size is bigger, due to spatial locality.

When we consider the different cache sizes about load misses let us look at this image:



As you can see when the cache size increases (green line), number of load-misses getting decrease. Because bigger cache size can <u>exploit temporal locality better</u>. There is a pros about miss rates, cons about latency for larger cache size but we are just focusing number of load-misses.

Secondly, let us check the results for another matrix size (N=10).

In this scenario there is not much difference that should be indicated.

But some observations about like this:

- Since number of operations decreases miss rate is smaller than 1000. When the operation is data intensive (N=1000) there is lots of deletion operation occurs and this affects overall rate. But when the N=10 there is not much deletion needed so miss-rate (%) is smaller than before.
- There is no coincidence happening here since data intense is small.

Other than these, observations and comment would be the same with as N=1000.

Here is the values and the graph for N=10:

