NC State University

Department of Electrical and Computer Engineering

ECE 463/563: Fall 2018 (Rotenberg)

Project #1: Cache Design, Memory Hierarchy Design

by

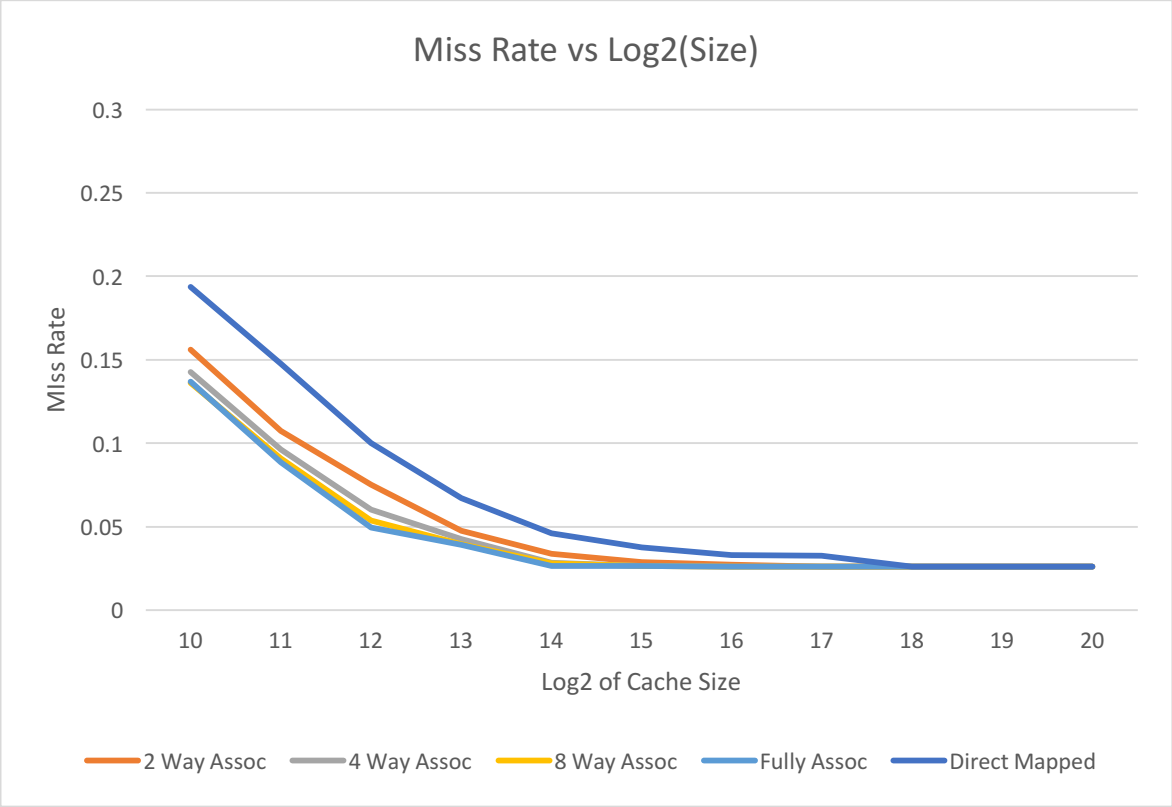
Priyank Kashyap

NCSU Honor Pledge: "I have neither given nor received unauthorized aid on this test or assignment."

Student's electronic signature: Priyank Kashyap

(sign by typing your name)

Course number: ECE-563



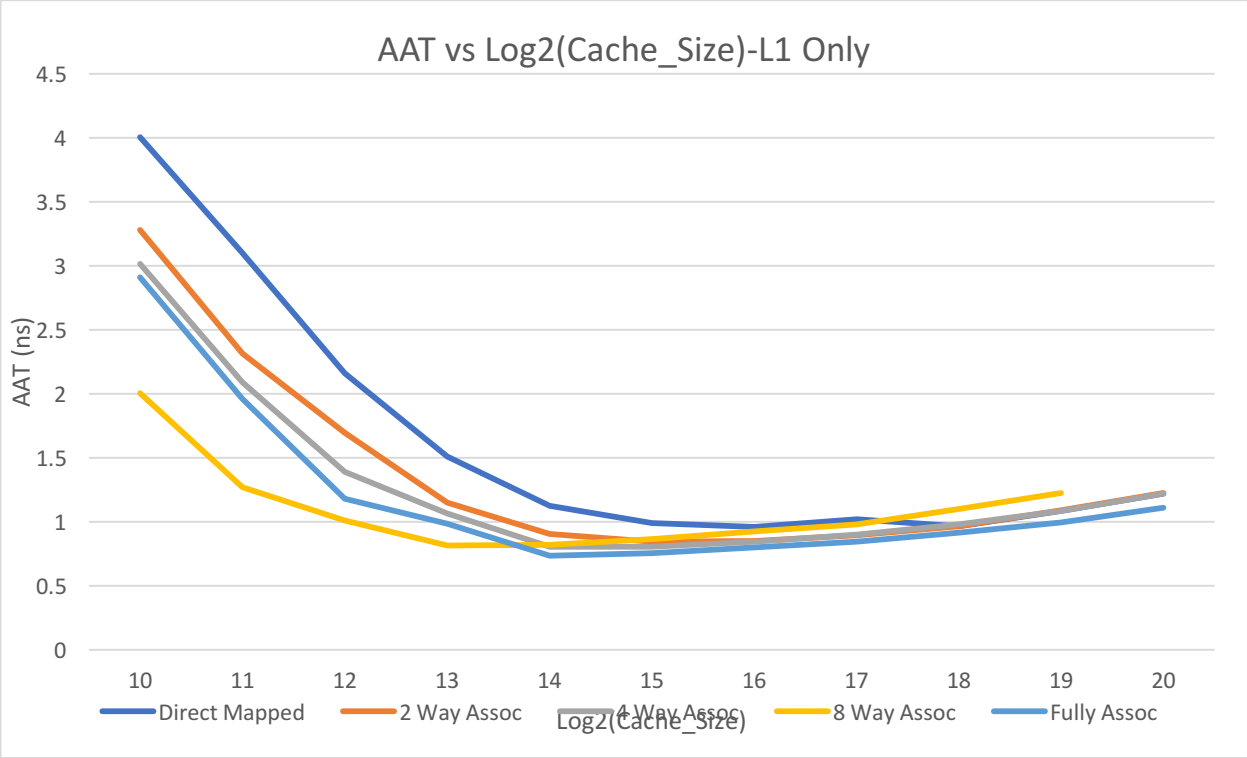
As the cache size is increased the miss rate goes down tremendously up until 32KB after which the decrease in miss rate is minimal. In addition to increasing the cache size, by increasing the associativity the miss rate does down as unused blocks aren't evicted thus increasing the chance of a cache miss.

The compulsory miss rate for this trace would be the miss rate of around 0.025 as it the miss rate seen by all the configurations towards the end of the graph, with 2582 misses.

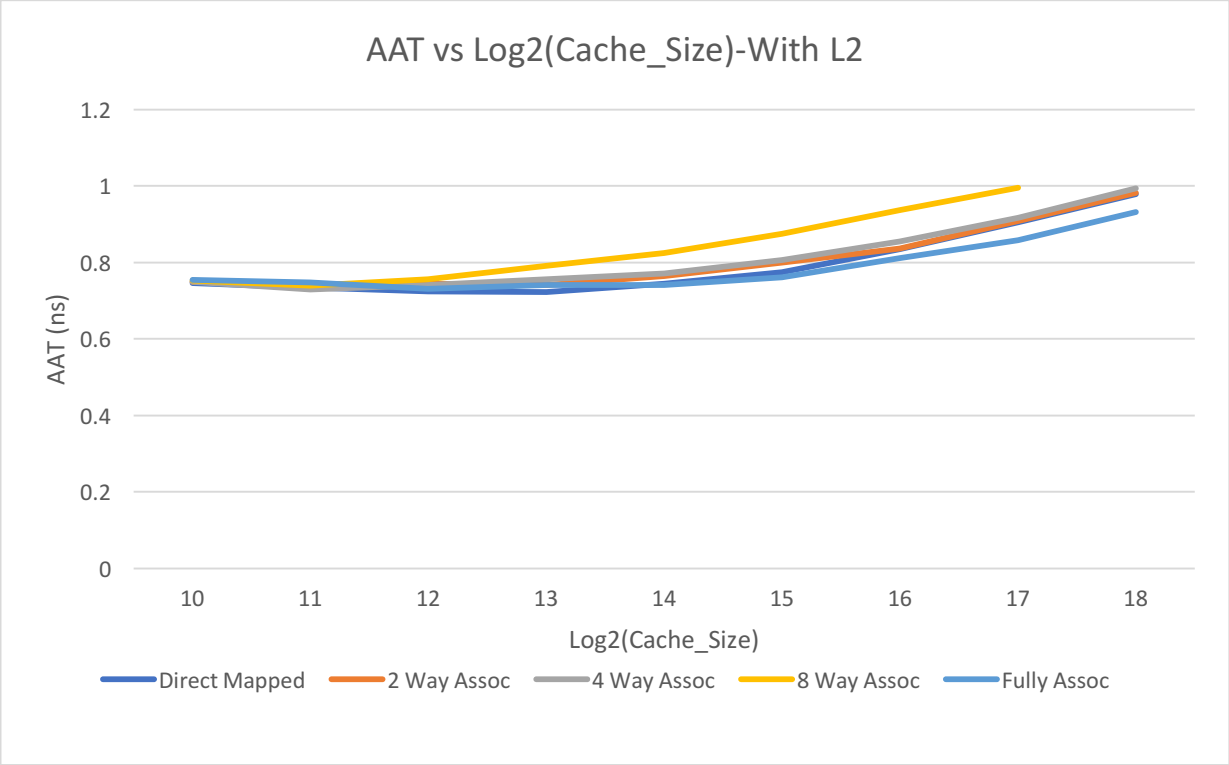
For direct mapped the conflict miss rate is the highest at the smaller sizes at approximately 0.04 however large sizes direct mapped conflict miss rate drops to under 0.01 and for sizes greater than 256KB, there is no conflict misses, just compulsory misses.

For 2 Way set associative the conflict miss rate for smaller configurations is high, at approximately 0.015. However, with increasing cache sizes it drops to under 0.01 at 32KB after which it goes to roughly 0 as it experiences only compulsory misses.

4 Way set associative and 8 way set associative conflict miss rates are both around 0.01 and eventually taper down to 0 as they experience only compulsory misses.



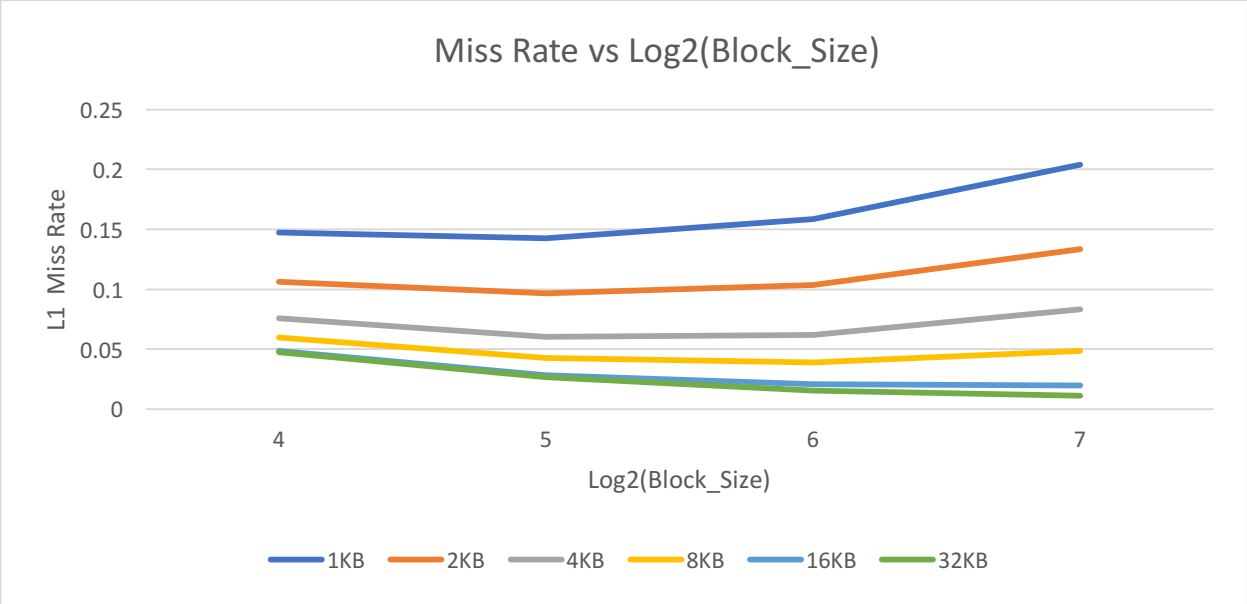
The fully associative 16KB L1 cache yields the best AAT at 0.736248ns.



2KB direct mapped comes very close the best AAT observed in the Figure 2.

8KB direct mapped with the L2 cache attached performed the best amongst all the Configurations tested here. Moreover, it performed better 1.5% better than the best configuration in the previous test.

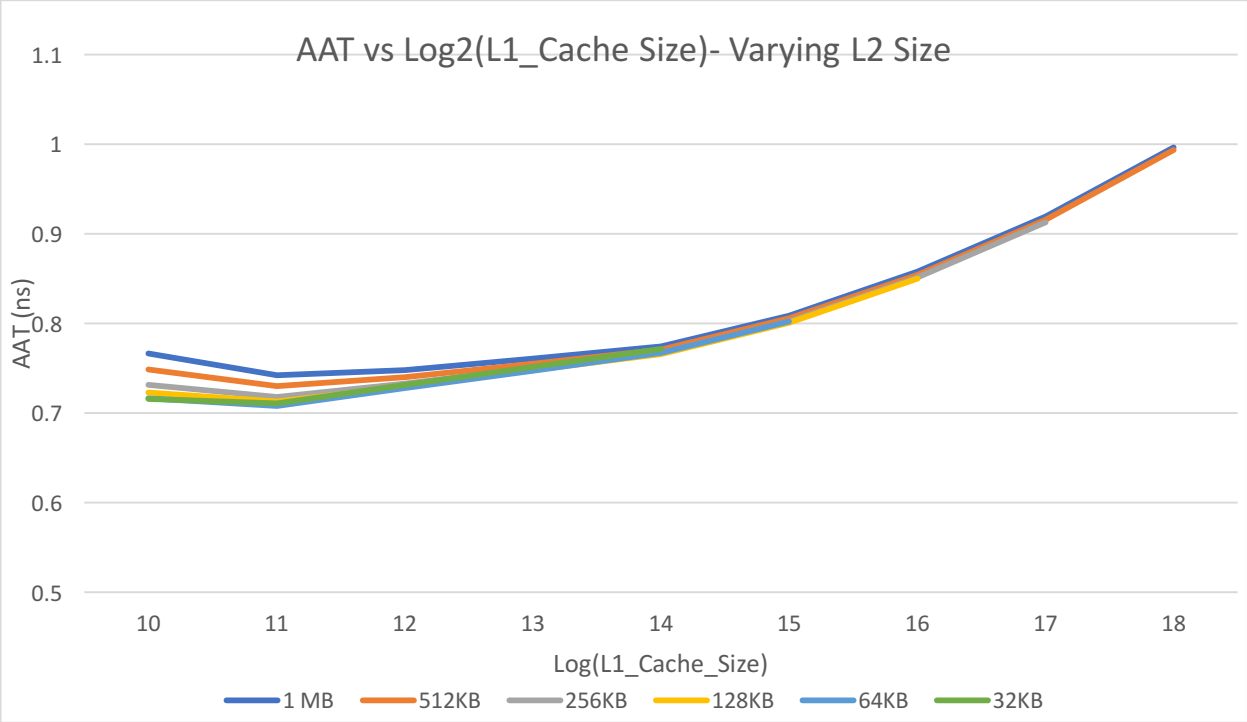
The total area in the case with the L2 enabled was 3.361728mm whereas in the case with no L2 in the pervious experiment it was 0.198417mm. The large difference is largely due to the presence of a large L2 cache, however the difference is large enough that performance increase doesn't justify it.



We can also draw form this that smaller caches such as the 1KB and 2KB work better with smaller block sizes as they fill up faster and have more cache misses.

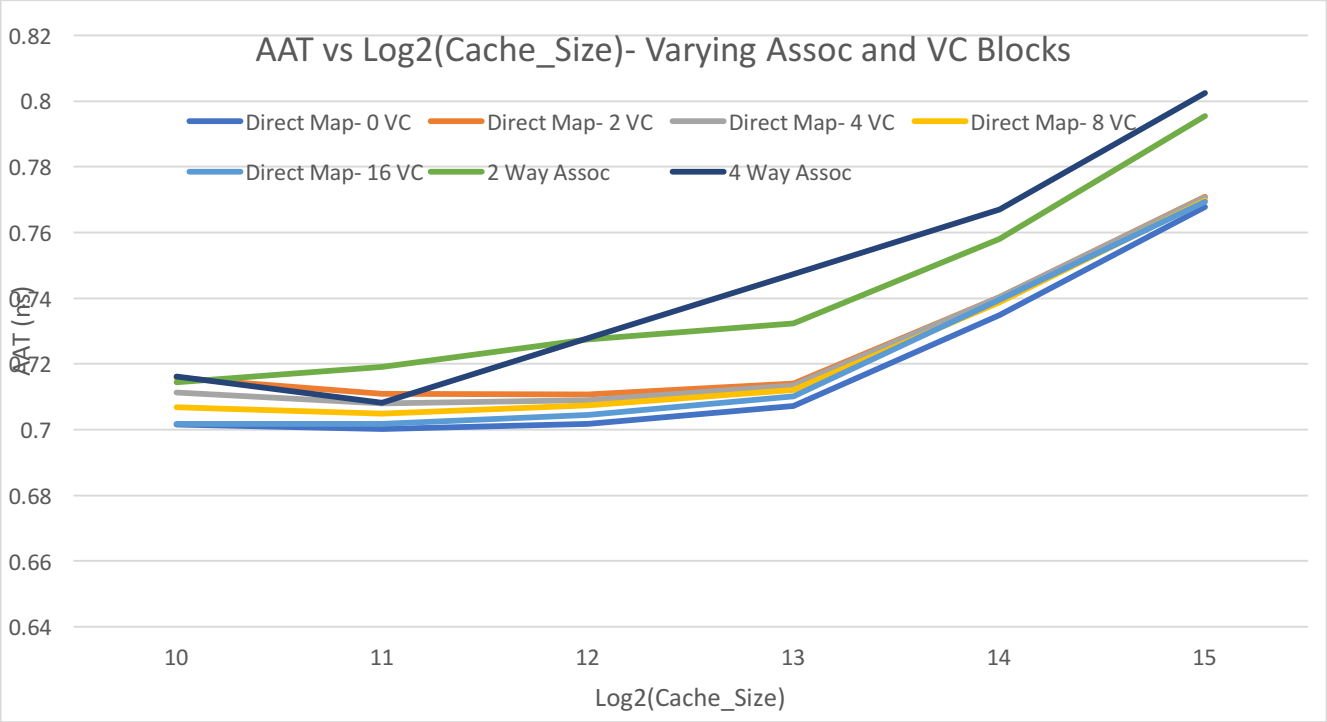
From figure 4, it is clear that the miss rate first decreases and then at particular point, around a block size of 64, the miss rate goes higher. This is to be expected due to the fact that by bringing in larger blocks there is data there doesn't fit our need and thus we get more cache misses.

Though we do get more spatial locality as a result of larger block sizes, there might be instances this does come at a cost for smaller caches as blocks might not have data for a different program and would otherwise would have been in the set, would now have a cache miss.



32KB L2 cache added to a 2KB L1 cache has the lowest AAT and thus performs the best amongst all of the cache configuration. This might simply be down to the low hit times for both those configurations of caches.

The memory configuration with the smallest area would be the 32KB L2 added to the 1KB L1, this has a AAT that is comparable to the lowest AAT.



There is a significant performance increase by adding a victim cache to a 2-Way or 4 Way set associative cache. However, the same can't be said for adding a victim cache to a direct mapped cache, this is probably explained by the fact the time for swaps negates the time for a miss penalty to L2. In addition to this, it interesting to note that 16 block victim cache added to a 1KB or 2KB has a very similar performance to a direct mapped cache. Also, we see that the 4-way set associative cache performs slightly better the 2-Block victim cache added to a 2KB cache, which may be due to the expanded memory offered by the 4-way cache.

We see from Figure 6 that best configuration was comparable amongst all the direct mapped L1 caches as the access times were less with the direct mapped 2KB cache performing the best. However, amongst the configurations with the victim cache enabled 1KB with16 blocks yield the lowest AAT, thus would be a best configuration to use.

The 2-Way set associative cache no victim cache enabled had the smallest area within a 5% margin of the best AAT. However, amongst those configurations with the victim cache enabled, 1KB with 2 blocks had the lowest area and was within the 5% margin of the fastest AAT.