NC State University

Department of Electrical and Computer Engineering

ECE 463/521: Fall 2015 (Rotenberg)

Project #1: Cache Design, Memory Hierarchy Design

by

ARJUN AUGUSTINE

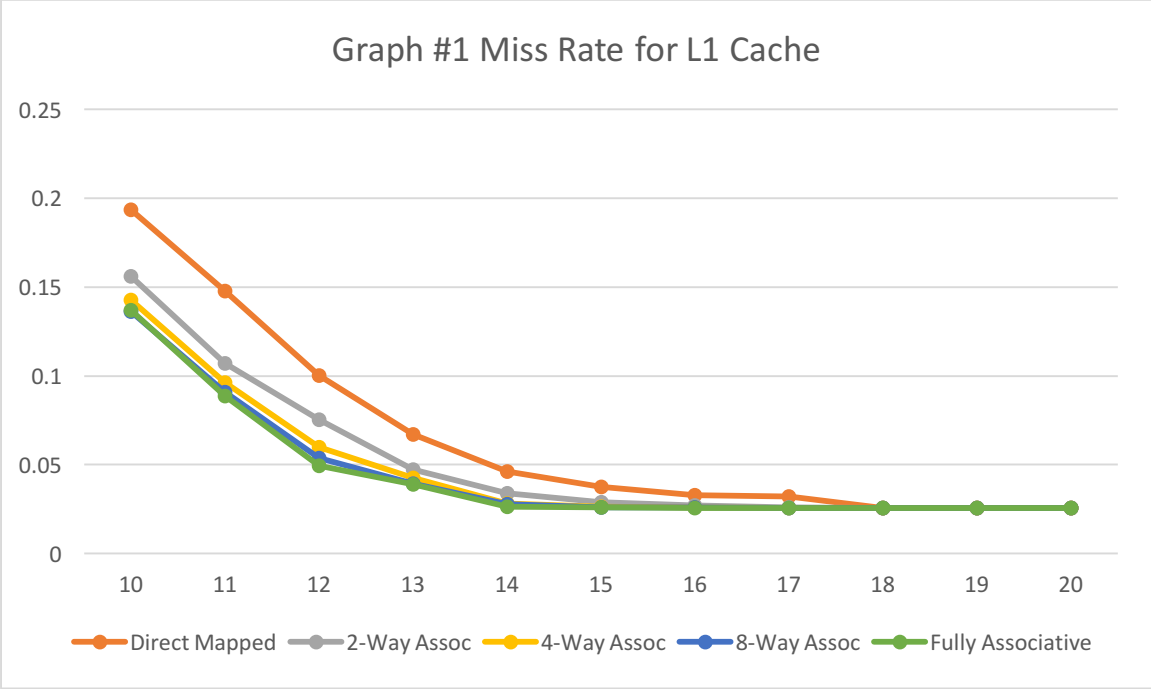
NCSU Honor Pledge: "I have neither given nor received unauthorized aid on this test or assignment."

Student's electronic signature: _____Arjun Augustine_____
(sign by typing your name)

Course number: _____521_____
(463 or 521 ?)

Graph #1:
For this experiment:
• L1 cache: SIZE is varied, ASSOC is varied, BLOCKSIZE = 32.
• Victim Cache: None.
• L2 cache: None.

Plot L1 miss rate on the y-axis versus $\log_2$(SIZE) on the x-axis for cache sizes: 1 KB, 2KB, … , 1MB.



Discussion to include in your report:

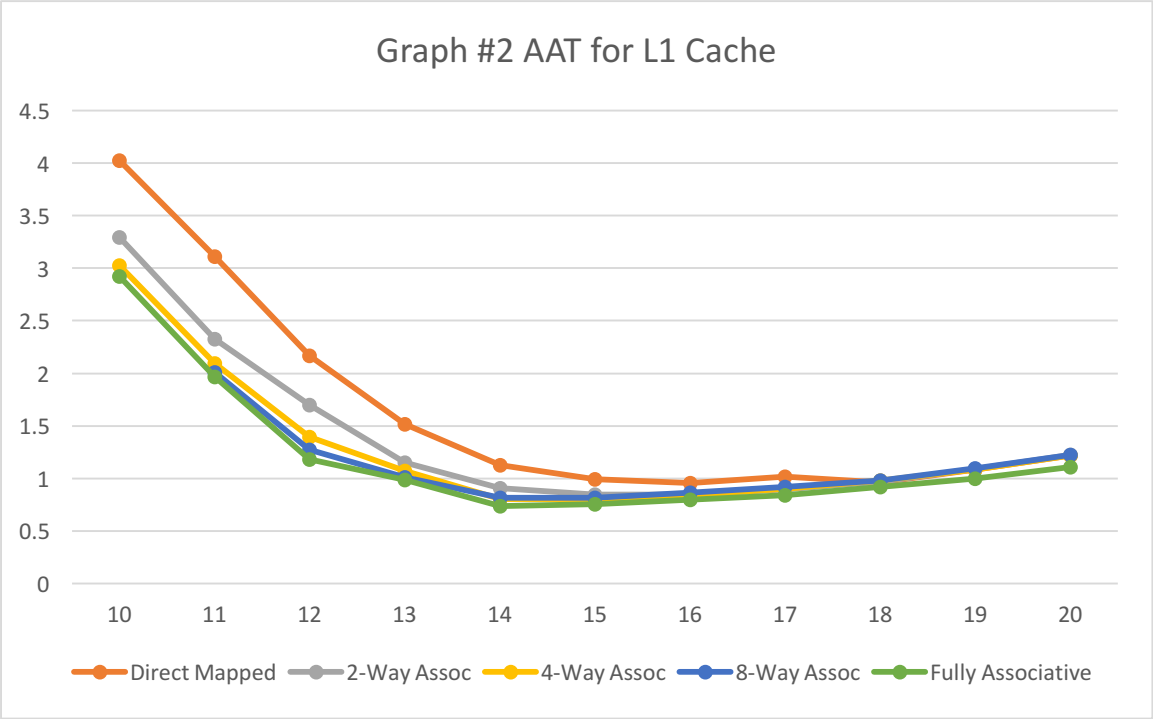
1. Discuss trends in the graph. For a given associativity, how does increasing cache size affect miss rate? For a given cache size, what is the effect of increasing associativity?
   ANSWER: Increasing cache size for a given associativity reduces miss rate proportionately initially. The curve then straightens off indicating that the miss rates are approaching compulsory miss rate values. Increasing associativity for a cache size also decreases miss rate but the effect wanes off when increasing it beyond 4 or 8 way set associative cache.

2. Estimate the *compulsory miss rate* from the graph.
   ANSWER: Compulsory miss rate is around 0.025 which is the value which all curves tend to close in to for high values of cache sizes.

3. For each associativity, estimate the *conflict miss rate* from the graph.
   ANSWER: Conflict miss rate is zero in fully associative cache. It is close to 25% of total misses in direct mapped cache at lower cache sizes. 2-way assoc: Close to 10%, 4-way assoc: Close to 5% and 8-way assoc: close to zero. Conflict miss percentage decreases to approach zero with increase in cache sizes.

Graph #2:

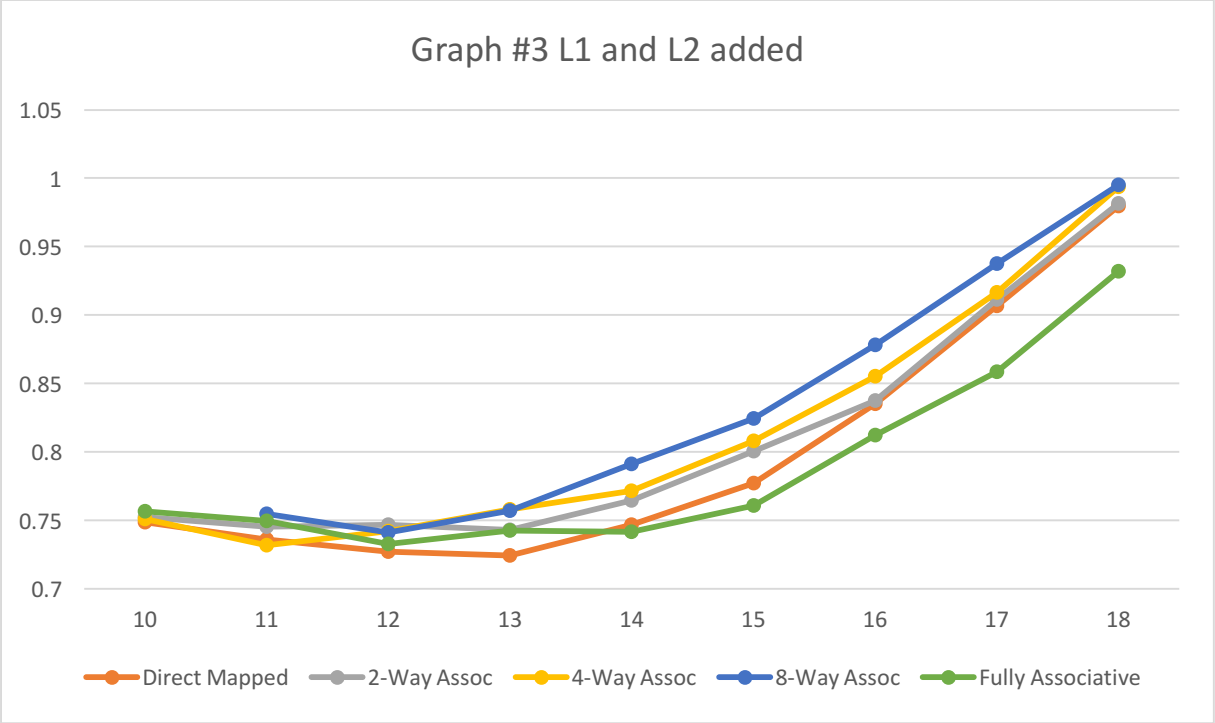
Plot AAT on the y-axis versus $\log_2$(SIZE) on the x-axis for cache sizes: 1KB, 2KB, …, 1MB.



Discussion to include in your report:
1. For a memory hierarchy with only an L1 cache and BLOCKSIZE = 32, which configuration yields the best (*i.e.*, lowest) AAT?

Lowest AAT is in case of Fully Associative cache of size 16KB which is 0.736868ns.

Graph #3
Plot AAT on the y-axis versus $\log_2$(SIZE) on the x-axis for cache sizes: 1KB, 2KB, …, 256KB with an L2 cache of 512KB, 8 way set associative cache.



Discussion to include in your report:

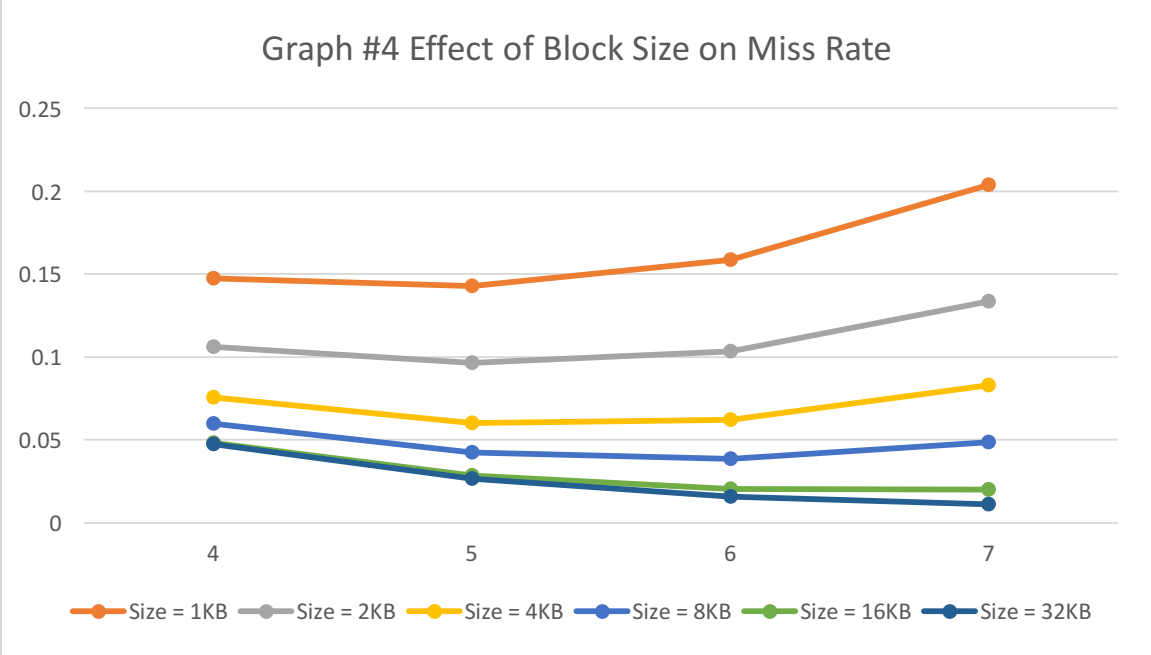
1. With the L2 cache added to the system, which L1 cache configurations result in AATs close to the best AAT observed in GRAPH #2 (*e.g.*, within 5%)?
   ANSWER: A direct mapped L1 cache with size 8KB gives lowest AAT. Most configurations with size between 1KB and 16KB gives AAT within 5% of the AAT in graph #2. The exceptions are in when associativity is 8 where only 4KB, 8KB sizes fall between accepted values, and in Fully associative cache whose 32KB configuration falls in the said range.

2. With the L2 cache added to the system, which L1 cache configuration yields the best (*i.e.*, lowest) AAT? How much lower is this optimal AAT compared to the optimal AAT in GRAPH #2?
   ANSWER: A direct mapped L1 cache with size 8KB gives lowest AAT which is 0.724168219ns. This is 0.012ns lesser than first one.

3. Compare the *total area* required for the optimal-AAT configurations with L2 cache (GRAPH #3) versus without L2 cache (GRAPH #2).
   ANSWER: Area with L2 Cache 2.69 sq.mm. whereas the one without it has an area of 0.06344 sq.mm. The former is larger than the latter by about 2.63 sq.mm.

Graph #4

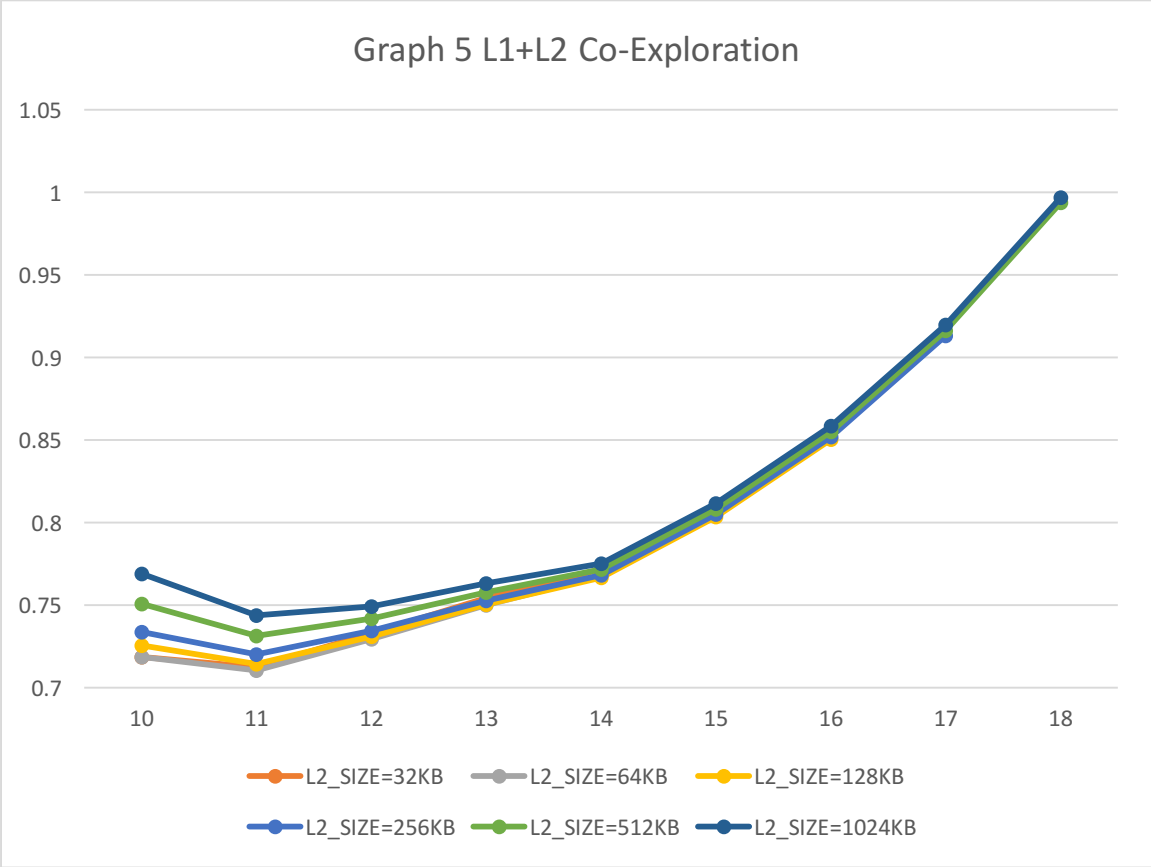
GRAPH #4 (*total number of simulations: 24*)
For this experiment:
• _L1 cache: SIZE is varied, BLOCKSIZE is varied, ASSOC = 4.
• _Victim Cache: None.
• _L2 cache: None.



Discussion to include in your report:
1. Discuss trends in the graph. Do smaller caches prefer smaller or larger block sizes? Do larger caches prefer smaller or larger block sizes? Why? As block size is increased from 16 to 128, is the tradeoff between *exploiting more spatial locality* versus *increasing cache pollution* evident in the graph, and does the balance between these two factors shift with different cache sizes?

Smaller cache sizes prefer smaller block sizes and vice-versa. As block size increases, the miss rate initially improves until it reaches a tipping point from where it starts to worsen. This tipping point comes early for smaller caches. And the further you increase block size after this, the higher miss rate goes. However, for size = 32KB, cache pollution effect is not so evident. This is because that the block size has not reached the tipping point yet for this size and thus the highest block size still gives a better spatial locality exploitation. Thus, the balance between the two factors have changed.

Graph #5
For this experiment:
• _L1 cache: SIZE is varied, BLOCKSIZE = 32, ASSOC = 4.
• _Victim Cache: None.
• _L2 cache: SIZE is varied, BLOCKSIZE = 32, ASSOC = 8.



Discussion to include in your report:

1. Which memory hierarchy configuration yields the best (*i.e.*, lowest) AAT?

ANSWER: L2_SIZE = 64 KB and L1_SIZE of 2KB gives optimum AAT of 0.710246575ns

2. Which memory hierarchy configuration has the smallest total area, that yields an AAT within 5% of the best AAT?

ANSWER: L1 Cache size of 1KB and L2 Cache Size of 32KB gives minimal total area of 0.2572 sq.mm.

Graph #6

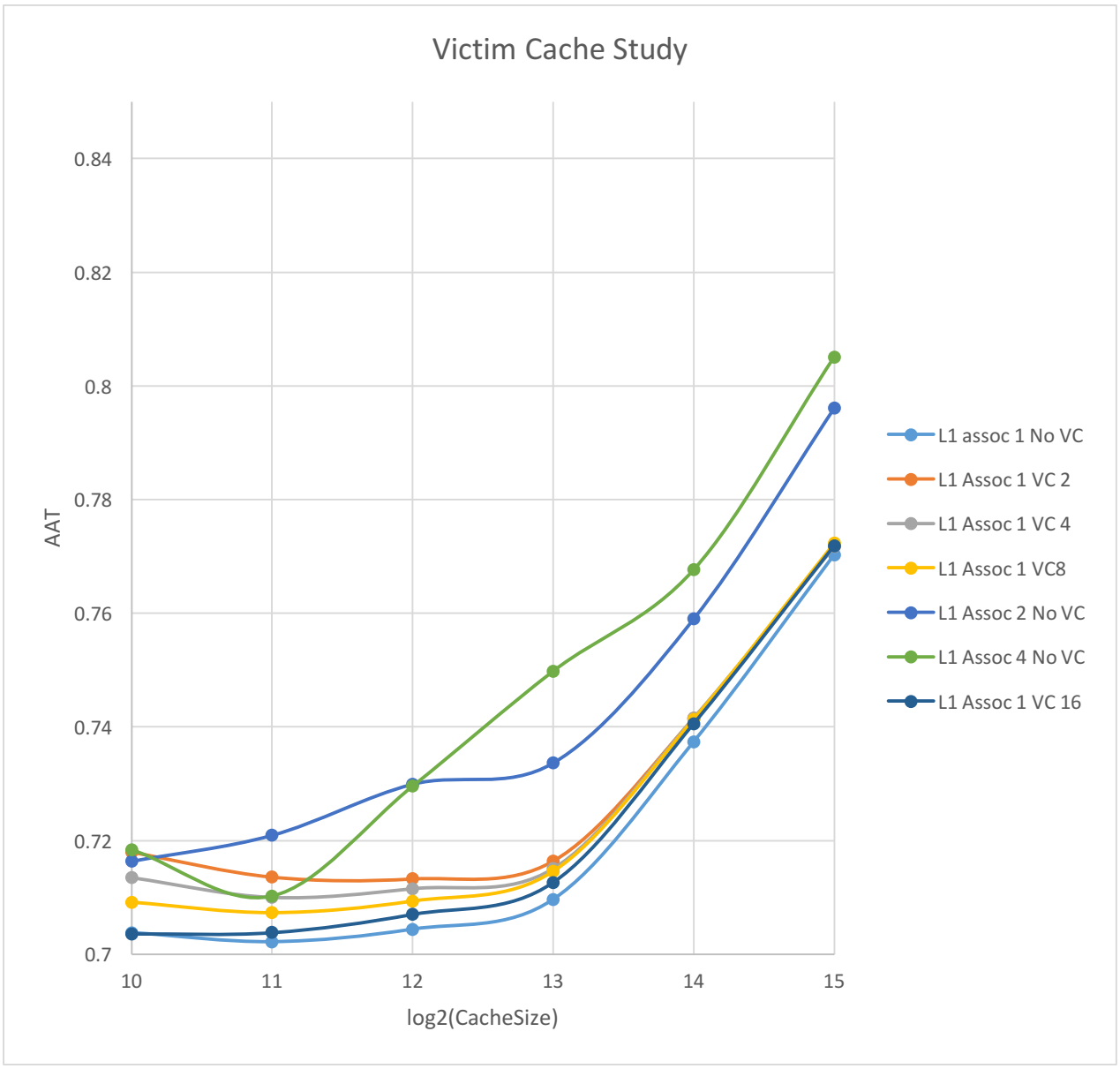
For this experiment:
• _L1 cache: SIZE is varied, ASSOC is varied, BLOCKSIZE = 32.
• _Victim Cache: # entries is varied.
• _L2 cache: SIZE = 64KB, ASSOC = 8, BLOCKSIZE = 32.

Plot AAT on the y-axis versus $\log_2$(L1 SIZE) on the x-axis, for six different L1 cache sizes: L1 SIZE = 1KB, 2KB, … , 32KB, in powers-of-two. (That is, $\log_2$(L1 SIZE) = 10, 11, …, 15.) The graph should contain seven separate curves (*i.e.*, lines connecting points), one for each of the following scenarios:
• _Direct-mapped L1 cache with no Victim Cache.
• _Direct-mapped L1 cache with 2-entry Victim Cache.
• _Direct-mapped L1 cache with 4-entry Victim Cache.
• _Direct-mapped L1 cache with 8-entry Victim Cache.
• _Direct-mapped L1 cache with 16-entry Victim Cache.
• _2-way set-associative L1 cache with no Victim Cache.
• _4-way set-associative L1 cache with no Victim Cache.
Discussion to include in your report:

1. Discuss trends in the graph. Does adding a Victim Cache to a direct-mapped L1 cache yield performance comparable to a 2-way set-associative L1 cache of the same size? …for which L1 cache sizes? …for how many Victim Cache entries?
   ANSWER: For small cache sizes, the AAT turns out higher in configuration, but has comparable AAT, that has Victim cache compared to another, but as cache size increases, the the difference between them increases and is not comparable to L1 cache not supported with a Victim Cache. As victim cache goes higher, the access time decreases except for the one with no victim cache. Increasing associativity also decreases AAT.

2. Which memory hierarchy configuration yields the best (*i.e.*, lowest) AAT?
   ANSWER: 2KB L1 Cache with direct mapping with no VC gives best AAT of 0.702ns.

3. Which memory hierarchy configuration has the smallest total area, that yields an AAT within 5% of the best AAT?\
   ANSWER: L1 size of 1KB Associativity of 2 with no VC gives best area of 0.369 sq.mm.


Victim Cache Study
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
AAT
10
11
12
13
14
15
log2(CacheSize)
L1 assoc 1 No VC
L1 Assoc 1 VC 2
L1 Assoc 1 VC 4
L1 Assoc 1 VC8
L1 Assoc 2 No VC
L1 Assoc 4 No VC
L1 Assoc 1 VC 16