NC State University

Department of Electrical and Computer Engineering

ECE 521: Fall 2015 (Rotenberg)

Project #1: Cache Design, Memory Hierarchy Design

By

LAXMAN SOLE
Student Id - 200112088

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

Student's electronic signature: Laxman Sole

Course number: 521

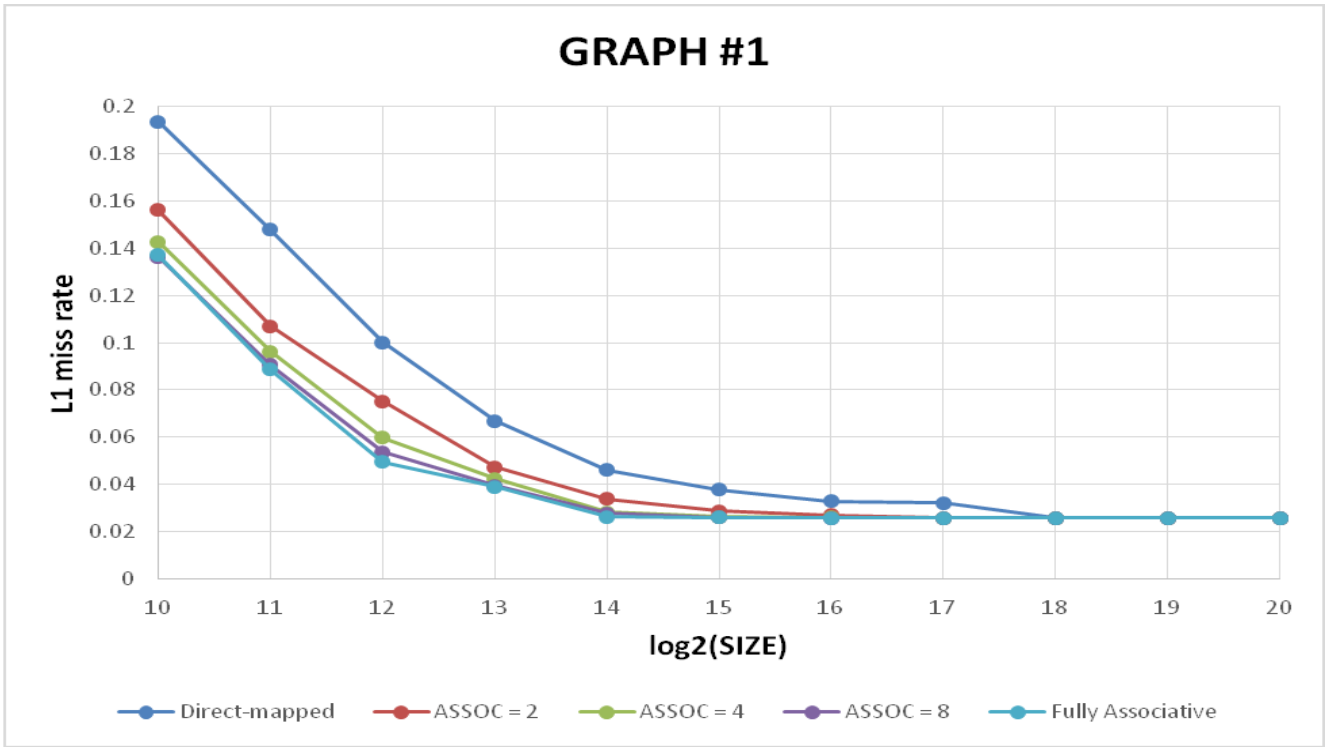
Contents

GRAPH #1

Data

Size	log2(Size)	ASSOC -1	ASSOC -2	ASSOC -4	ASSOC -8	Fully Associative
1024	10	0.1935	0.156	0.1427	0.1363	0.137
2048	11	0.1477	0.1071	0.0962	0.0907	0.0886
4096	12	0.1002	0.0753	0.0599	0.0537	0.0495
8192	13	0.067	0.0473	0.0425	0.0395	0.0391
16384	14	0.0461	0.0338	0.0283	0.0277	0.0263
32768	15	0.0377	0.0288	0.0264	0.0262	0.0262
65536	16	0.0329	0.0271	0.026	0.0259	0.0258
131072	17	0.0323	0.0259	0.0258	0.0258	0.0258
262144	18	0.0258	0.0258	0.0258	0.0258	0.0258
524288	19	0.0258	0.0258	0.0258	0.0258	0.0258
1048576	20	0.0258	0.0258	0.0258	0.0258	0.0258

Plot



Discussion

1) From the graph, it is clear that for a given associativity, miss rate decreases with increasing cache size. Also for a given cache size, increasing associativity, miss rate decreases. In both these cases, miss rate decreases up to certain point and then it remains constant.
2) As discussed in point 1, miss rate, even for fully associative cache, remains constant at particular level. This indicates that all capacity and conflict misses, which we could avoid by increasing capacity and associativity, reduced to zero and there are only compulsory misses. So compulsory miss rate is 0.0258
3) As fully associative cache does not suffer the conflict misses, for a given set associative cache, conflict miss rate can be simply estimated by subtracting miss rate of fully associative cache from the miss rate of given set associative cache. For cache size of 2 KB, conflict miss rates are as follows:
   a) Direct-mapped cache: 0.1477-0.0886 = 0.0591
   b) 2-way set associative cache: 0.1071 - 0.0886 = 0.0185
   c) 4-way set associative cache: 0.0962 - 0.0886 = 0.0076
   d) 8-way set associative cache: 0.0907 - 0.0886 = 0.0021
   e) Fully associative cache: 0

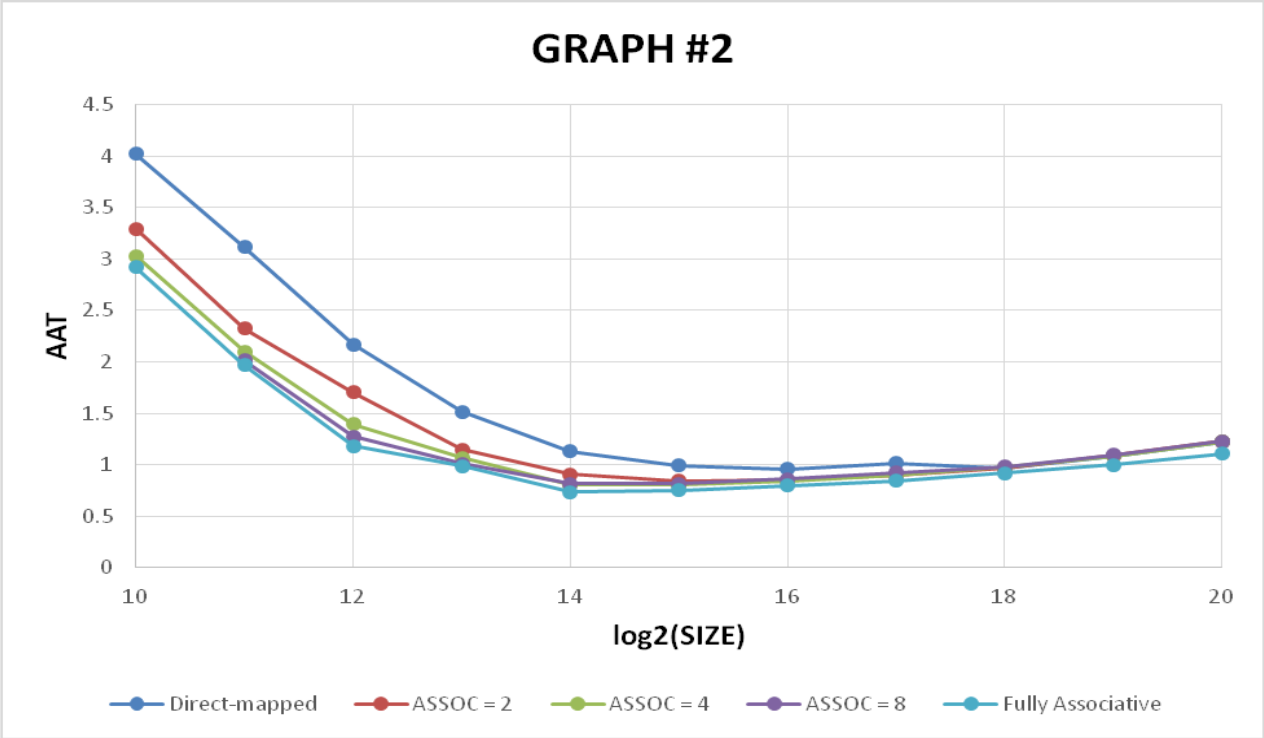
   For 1 KB size 8-way associative cache performs better than fully associative cache.

GRAPH #2

Data

Size	log2(Size)	AAT -1	AAT -2	AAT -4	AAT -8	AAT -F
1024	10	4.023497	3.291529	3.02936		2.922884
2048	11	3.11263	2.325111	2.097736	2.012826	1.966235
4096	12	2.171045	1.702191	1.395665	1.273805	1.182848
8192	13	1.51723	1.149655	1.069673	1.010811	0.988401
16384	14	1.129637	0.906677	0.805596	0.813894	0.736868
32768	15	0.994893	0.844206	0.80453	0.817751	0.75398
65536	16	0.959207	0.848147	0.844681	0.864393	0.797441
131072	17	1.01926	0.897783	0.90144	0.922396	0.843646
262144	18	0.964972	0.967089	0.978845	0.980085	0.917169
524288	19	1.084611	1.088904	1.085578	1.099337	0.996888
1048576	20	1.22054	1.227206	1.220767	1.226979	1.109634

Plot



Discussion

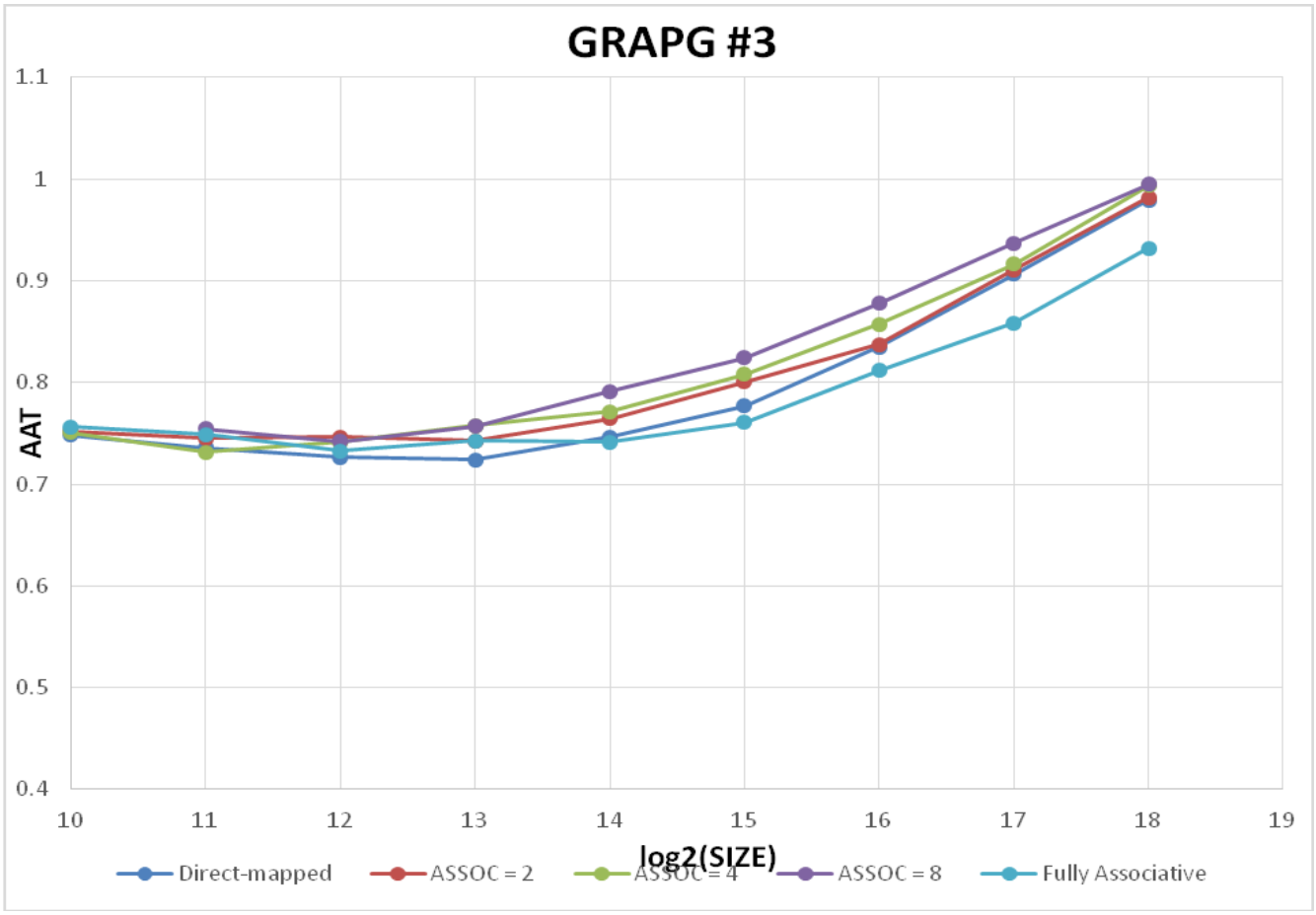
For a memory hierarchy with only L1 cache and BLOCKSIZE = 32; 16KB fully associative cache yields the best AAT of 0.736868 ns.

GRAPH #3

Data

log2(Size)	AAT-1	AAT-2	AAT-4	AAT-8	AAT-F
10	0.7484857	0.752048212	0.750777344		0.756349149
11	0.736009535	0.744997977	0.731487919	0.754736912	0.74926589
12	0.726735847	0.746391308	0.741698184	0.742198467	0.732715642
13	0.724168219	0.742650656	0.757713523	0.756777692	0.742468921
14	0.746740404	0.764405263	0.771480831	0.791189335	0.741608233
15	0.776957481	0.80047081	0.808061713	0.824219701	0.760660397
16	0.834879117	0.837375211	0.857087602	0.877955198	0.812149503
17	0.906529673	0.911135926	0.916356967	0.937312967	0.858562967
18	0.979472039	0.981589039	0.993761967	0.995001967	0.932085967

Plot



Discussion

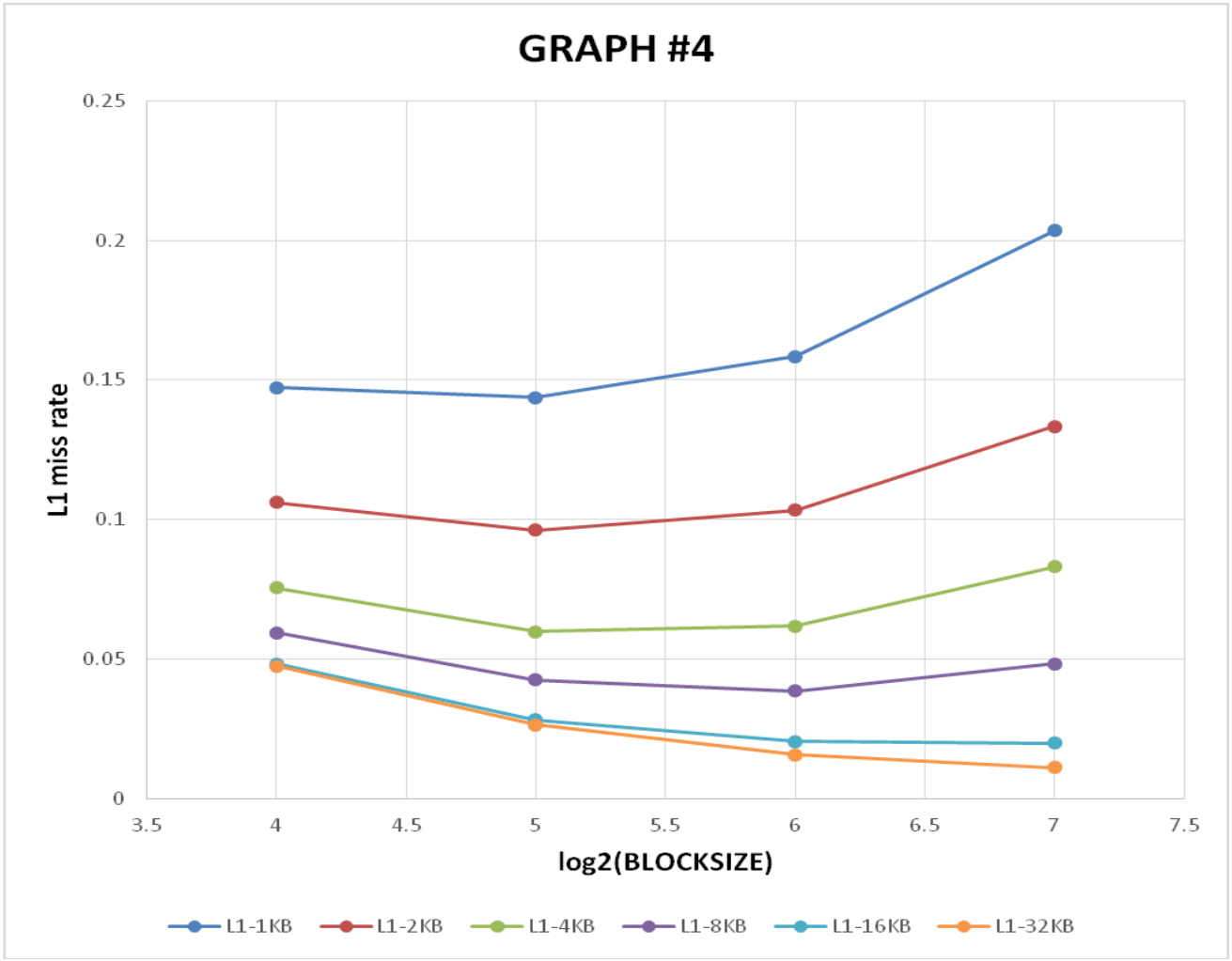
1) 2KB direct-mapped cache results in AATs close to the best AAT (0.736868 ns) observed in GRAPH #2 with AAT of 0.736009535 ns.
2) 8KB direct-mapped cache yields the best AAT of 0.724168219 ns. This is 1.7235% lower than the previous value.
3) Total area required for the optimal-AAT configurations without L2 cache configuration in GRAPH # 2 is 0.0634 millimeter squared.
   Total area required for the optimal-AAT configurations with L2 cache configuration in GRAPH #3 is 0.0532 + 2.6401 = 2.6933 millimeter squared. This is 42.48 times more than the area required for optimal cache configuration in graph 2.

GRAPH #4

Data

Blocksize	log2(Blocksize)	1KB	2KB	4KB	8KB	16KB	32KB
16	4	0.1473	0.1062	0.0755	0.0595	0.0482	0.0475
32	5	0.1437	0.0962	0.0599	0.0425	0.0283	0.0264
64	6	0.1584	0.1033	0.0619	0.0386	0.0204	0.0156
128	7	0.2036	0.1334	0.083	0.0483	0.0198	0.0111

Plot



Discussion

From the graph, we can conclude that smaller cache prefers smaller block size and larger cache prefers larger block size. If smaller cache has larger block, it achieves the spatial locality but it also reduces the no of sets and increases the cache pollution. Hence there are more conflict misses.

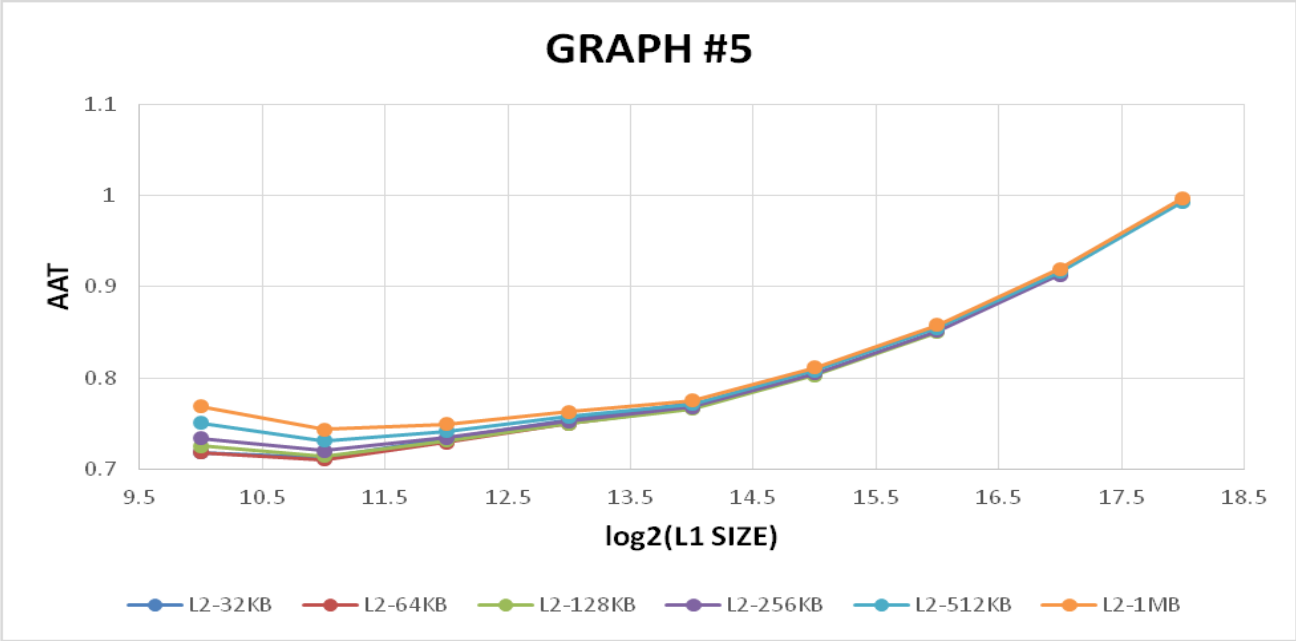
As block size increased from 16 to 128, tradeoff between exploiting more spatial locality versus increasing cache pollution is evident in the graph. For same size cache with fixed associativity if block size increased, miss rate reduces up to certain point and it increases further that point. Initial decrease in miss rate is due to exploiting more spatial locality. After certain point increasing cache pollution becomes dominant and miss rate increases. From the graph, balance point between these two factors is at Block size of 32B for L1 cache of size 1KB, 2KB and 4KB. It shifts to 64B for 8KB L1 cache and more further for larger caches.

GRAPH #5

Data

log2(Size)	AAT-32	AAT-64	AAT-128	AAT-256	AAT-512	AAT-1024
10	0.71837788	0.718403851	0.725527863	0.733760084	0.75077734	0.7689919
11	0.71236663	0.710246575	0.714466195	0.720015877	0.73148792	0.7437671
12	0.733059047	0.729561007	0.731099418	0.73455499	0.74169818	0.7493439
13	0.754073568	0.749788803	0.75019353	0.752645313	0.75771352	0.7631383
14	0.771801017	0.76763305	0.766473401	0.768106	0.77148083	0.7750931
15		0.805058871	0.80339047	0.80491346	0.80806171	0.8114315
16			0.850437112	0.851931258	0.85501988	0.8583258
17				0.913280265	0.91635697	0.9196501
18					0.99376197	0.9970551

Plot



Discussion

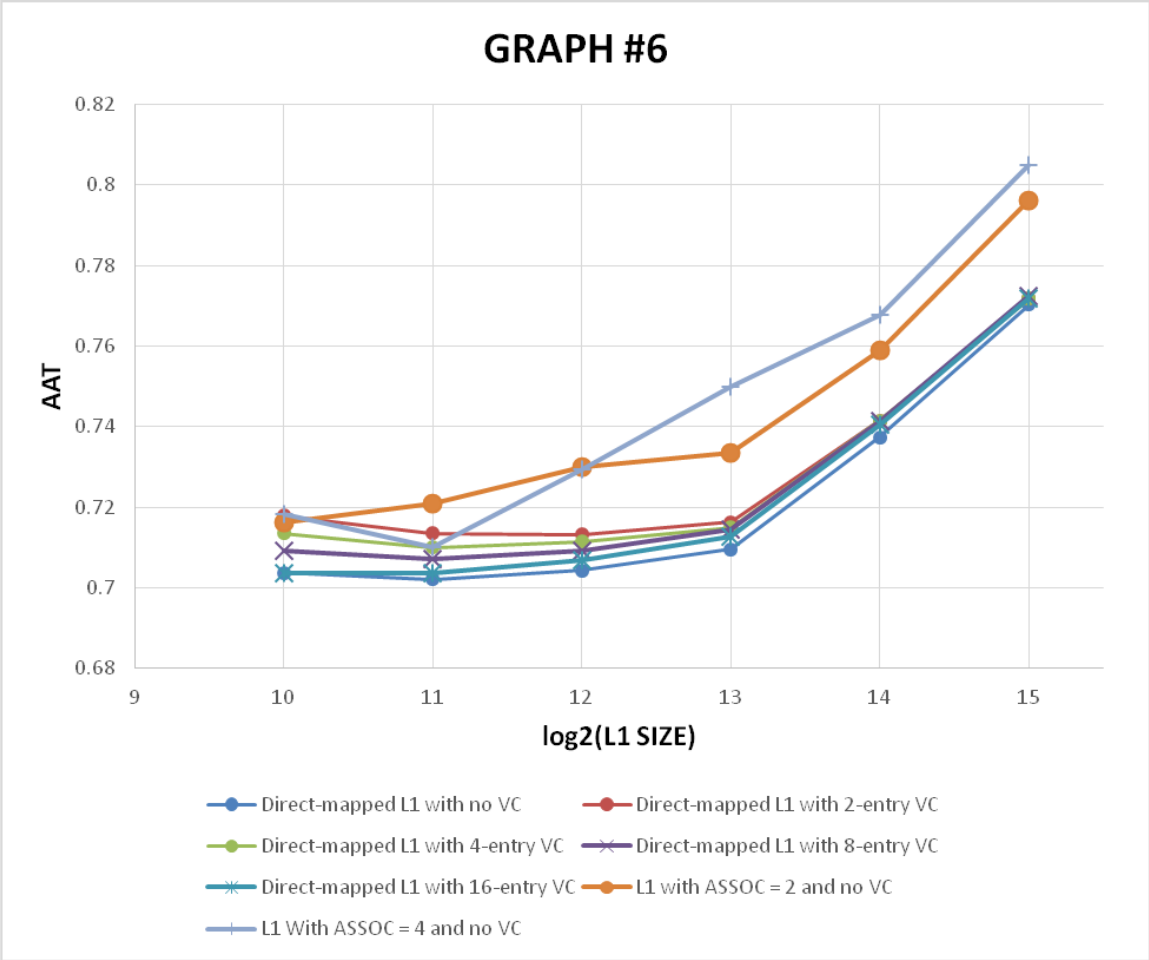
1) 2KB L1 cache and 64 KB L2 cache give the lowest average access time of 0.710246575 ns for the given block size and associativity values.
2) 1KB L1 cache and 32 KB L2 cache have the smallest total area of 0.2572 millimeter squares and give an AAT value of 0.7183 which falls within 5% of the optimal AAT.

GRAPH #6

Data

log2(L1 Size)	L1 - No VC	L1-2 VC blocks	L1-4 VC blocks	L1-8 VC blocks	L1-16 VC blocks	2ASSOC L1-No VC	4ASSOC L1-No VC
10	0.7038058	0.71797281	0.7134985	0.70914647	0.70361077	0.716342308	0.718403851
11	0.7022034	0.71359313	0.7100232	0.70730303	0.70382761	0.720917184	0.710246575
12	0.7044089	0.71326944	0.7115531	0.70934557	0.70703705	0.729916873	0.729561007
13	0.709645	0.71639977	0.7150789	0.71457384	0.71265537	0.733639817	0.749788803
14	0.7373994	0.74151433	0.7415661	0.74136853	0.74056712	0.758990367	0.76763305
15	0.7703086	0.77188547	0.7717189	0.77233973	0.77186076	0.796089638	0.805058871

Plot



Discussion

1) Adding Victim Cache to a direct mapped L1 cache increases overall performance of L1 cache. For 1KB direct-mapped L1 cache size adding a 2 victim cache entries yields performance comparable to a 2-way set-associative cache. But overall 2-way set associative cache has more access time than L1 cache with victim cache.
2) 2KB direct- mapped L1 cache gives the lowest AAT of 0.7022 ns.
3) 1 KB direct-mapped L1 cache gives an AAT within 5% of the optimal AAT of 0.7038058 ns and the smallest area 0.1029 millimeter squared.