NC State University

Department of Electrical and Computer Engineering

ECE 463/521: Fall 2013 (Rotenberg)

Project #1: Cache Design, Memory Hierarchy Design

by

<< Neal G O'Hara >>

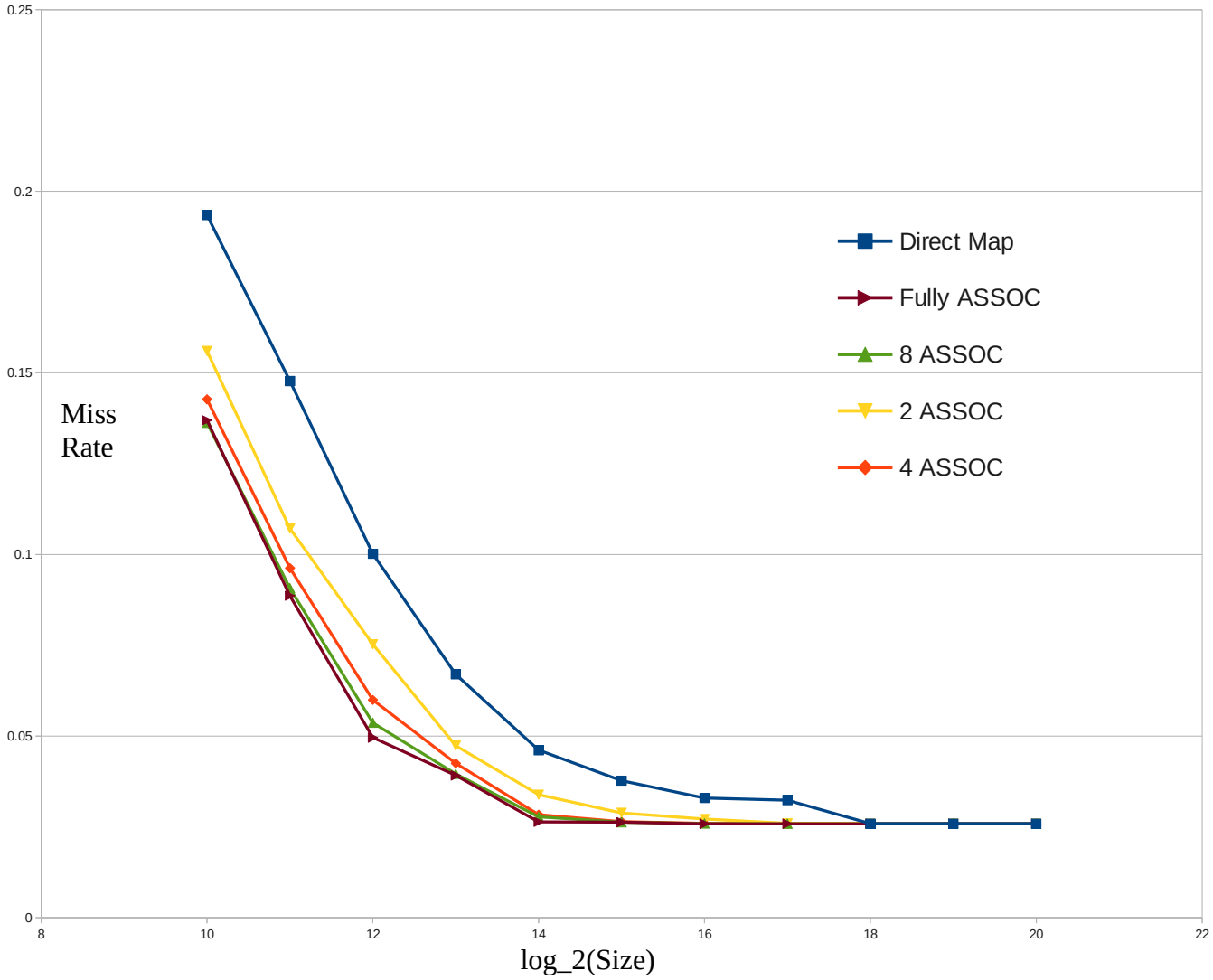
NCSU Honor Pledge: "I have neither given nor received unauthorized aid on this test or assignment."

Student's electronic signature: ____Neal O'Hara_____
(sign by typing your name)

Course number: _____ECE 521_____
(463 or 521 ?)

Section 9.1. L1 Cache exploration: SIZE and ASSOC

Graph #1
L1 cache: SIZE is varied, ASSOC is varied, BLOCKSIZE = 32.
L2 cache: None
Prefetching: None



Discussion

1) For each associativity, increasing the cache size decreases the miss rate exponentially.
   For each cache size, increasing the associtivity decreases the miss rate with diminishing returns, ie. 2 ASSOC is much better than Direct Map, but Fully ASSOC is not much better than 8 ASSOC.

2) From the graph and data, the compulsory miss rate appears to be around 0.02582, ie. where the graphs bottom out.

3) Conflict miss rate can be approximated by Direct Map Miss Rate – Fully ASSOC Miss Rate. This calculates out to:

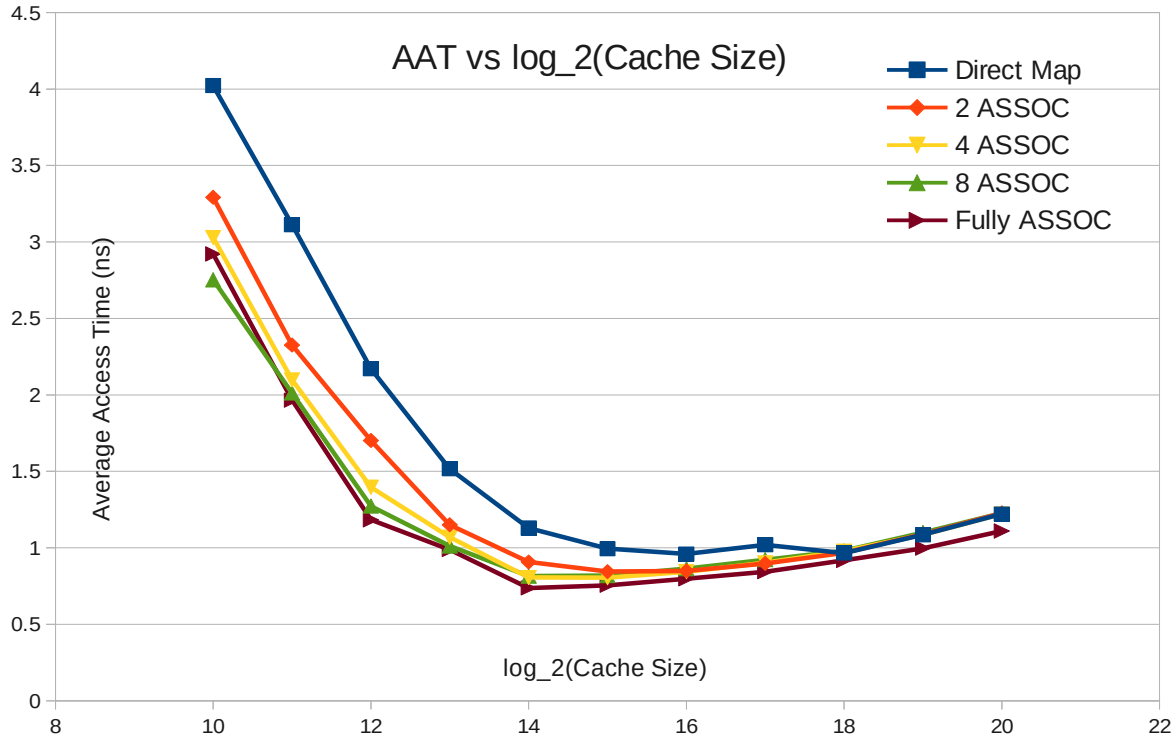
Size	Conflict Miss Rate
1024	0.0565
2048	0.05914
4096	0.05063
8192	0.02788
16384	0.01975
32768	0.01144
65536	0.00709
131072	0.00651
262144	0.00002
524288	0.00002
1048576	0.00002

GRAPH #2
Same as GRAPH #1, but the y-axis should be AAT instead of L1 miss rate.

Formula Used:

Miss_Penalty (in ns) = 20 ns + 0.1*(BLOCKSIZE / 16 B/ns)
= 20 + 0.1*(32/16) = 20.2 ns

Average access time (AAT) = HT_L1 + MR_L1* Miss_Penalty



Discussion

1) By graph and data, one can see that a fully associative, 16 kb cache has the best AAT, at 0.737676 ns .

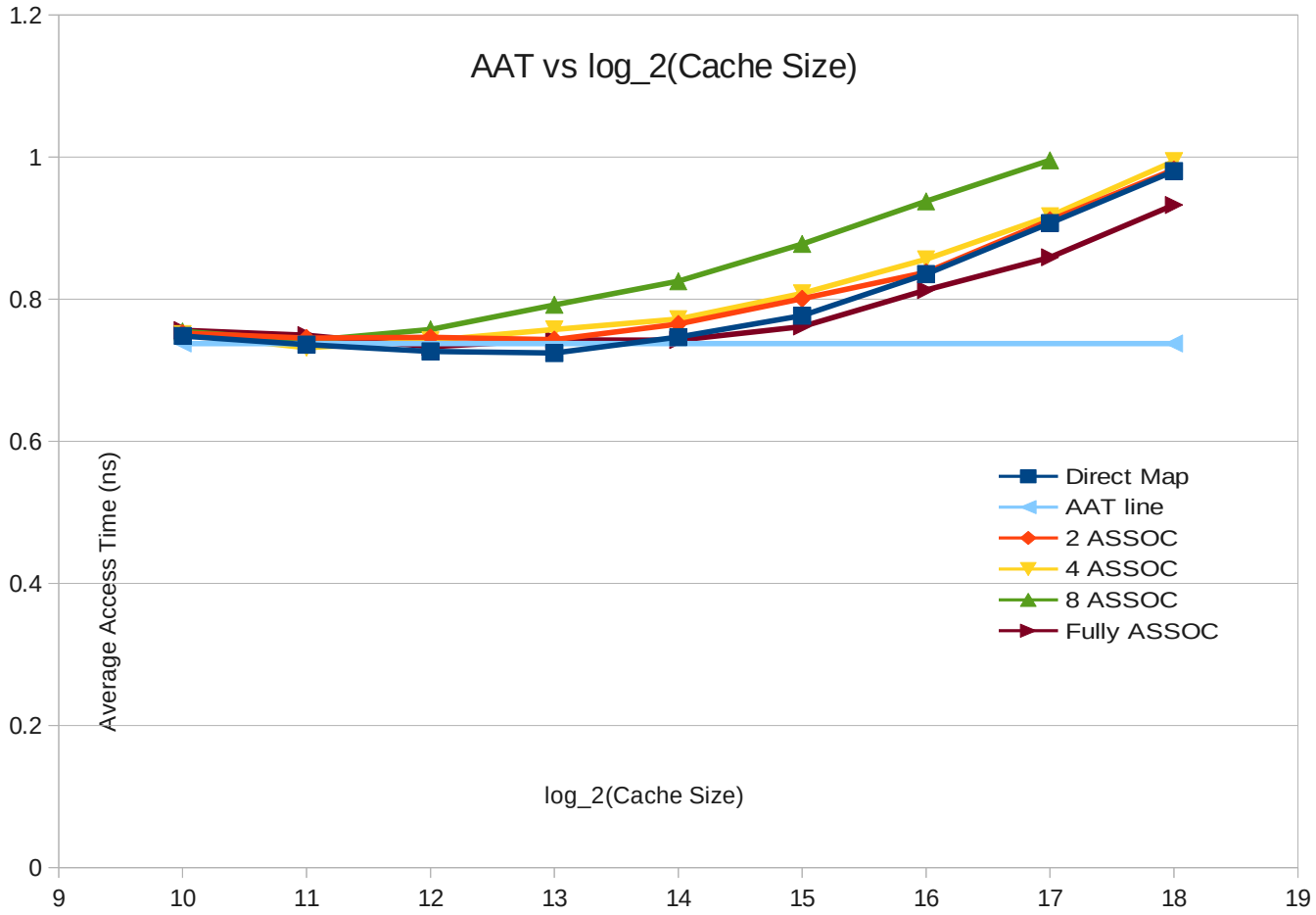
GRAPH #3
Add the following L2 cache to the memory hierarchy: 512KB, 8-way set-associative, same block size as L1 cache.
Vary the L1 cache size only between 1KB and 256KB (since L2 cache is 512KB).

Formula Used:

Miss_Penalty (in ns) = 20 ns + 0.1*(BLOCKSIZE / 16 B/ns)
= 20 + 0.1*(32/16) = 20.2 ns

Average access time (AAT) = HT_L1 + MR_L1* (HT_L2 + MR_L2*Miss_Penalty)

HT_L2 for 512KB, 8 ASSOC cache = 0.578177 ns



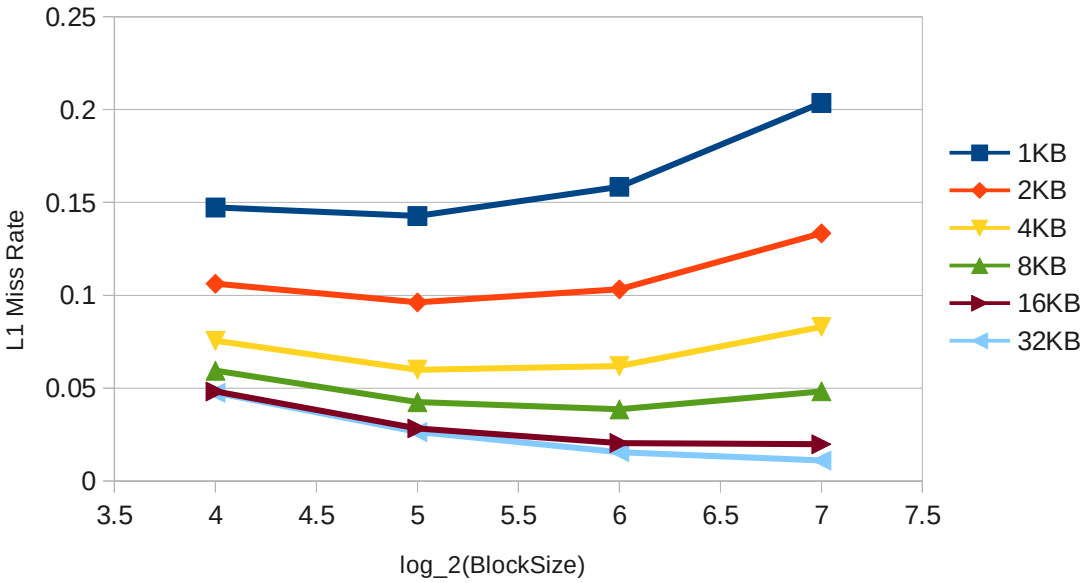
Discussion

1) The following Caches (table below), with their respective L1 Size, L1 ASSOCC, and AAT, fell within 5% of the best AAT time for the previous section. The arrows indicate the AATs less than the best from the earlier section.
2) The best configuration is the 8192 B Direct Map L1 cache, at 0.724131 ns AAT. This is 0.013544 ns faster.

L1 Size	ASSOC	AAT	
1024	1	0.7482140203	
2048	1	0.7360724325	<<<<
4096	1	0.7264853848	<<<<
8192	1	0.7241316772	<<<<
16384	1	0.7466289097	
1024	2	0.752105966	
2048	2	0.7452006822	
4096	2	0.7462198834	
8192	2	0.7431297667	
16384	2	0.7650462608	
1024	4	0.750889763	
2048	4	0.7316914528	<<<<
4096	4	0.7418935145	
8192	4	0.7572925555	
16384	4	0.7718738789	
2048	8	0.7546846124	
4096	8	0.741647682	
8192	8	0.7573364297	
1024	FA	0.7562345789	
2048	FA	0.749305264	
4096	FA	0.7331548946	<<<<
8192	FA	0.7427629287	
16384	FA	0.7424010957	
32768	FA	0.7614754162	

Section 9.2. L1 cache exploration: SIZE and BLOCKSIZE

GRAPH #4
L1 cache: SIZE is varied, BLOCKSIZE is varied, ASSOC = 4.
L2 cache: None.
Prefetching: None.



Discussion

1) The 32KB cache has the lowest miss rate at the high end of the Block Sizes, while the 1KB cache has the lowest miss rate at the low end of Block Sizes. This indicates the smaller caches prefer smaller blocks, while larger caches prefer larger blocks. In these scales, 1, 2, and 4 KB count as small, and 16 and 32 KB count as large. 8KB is right in the middle, about the same miss rate on either end of the spectrum.

Section 9.3. L1 + L2 co-exploration

GRAPH #5
L1 cache: SIZE is varied, BLOCKSIZE = 32, ASSOC = 4.
L2 cache: SIZE is varied, BLOCKSIZE = 32, ASSOC = 8.
Prefetching: None.

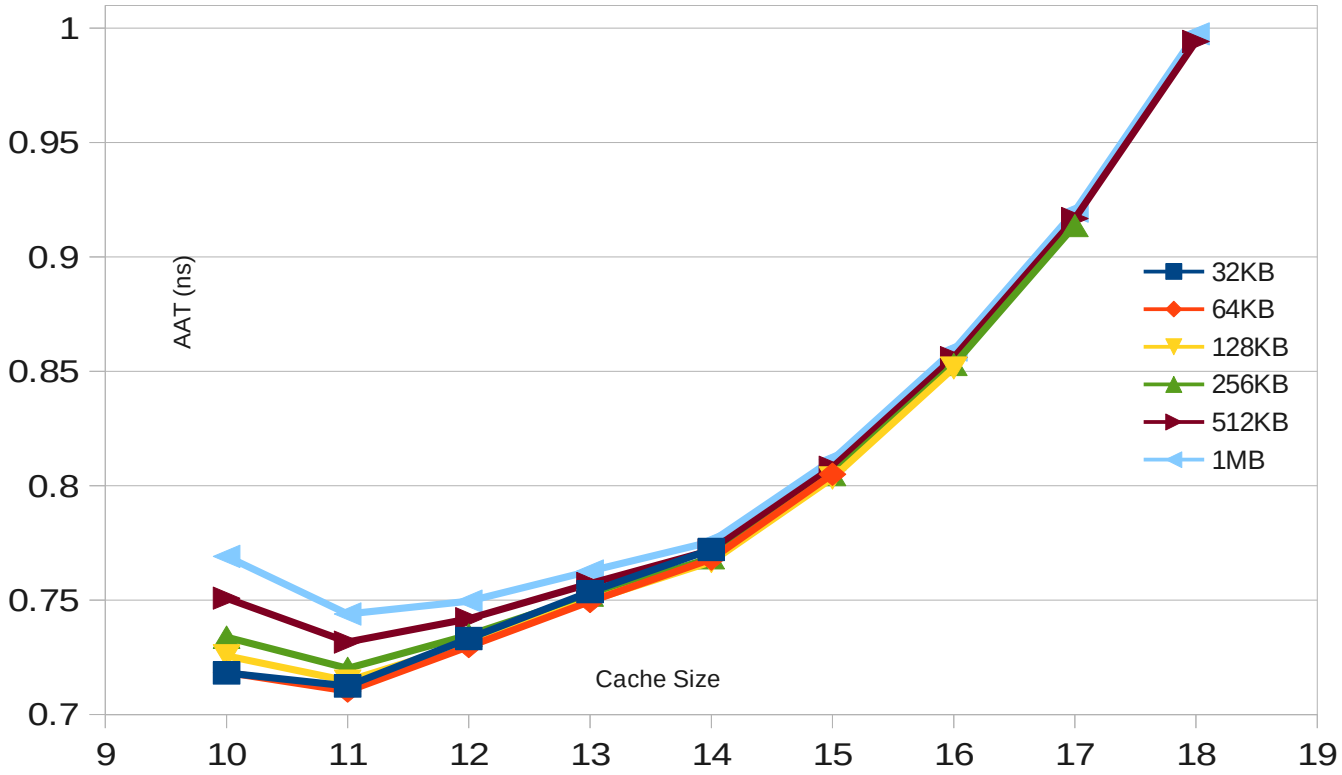
Formula Used:

Miss_Penalty (in ns) = 20 ns + 0.1*(BLOCKSIZE / 16 B/ns)
= 20 + 0.1*(32/16) = 20.2 ns

Average access time (AAT) = HT_L1 + MR_L1* (HT_L2 + MR_L2*Miss_Penalty)

L1 Size	L1 Hit
1024	0.14682
2048	0.154496
4096	0.185685
8192	0.211173
16384	0.233936
32768	0.27125
65536	0.319481
131072	0.38028
262144	0.457685

L2 Size (KB)	L2 Hit (ns)
32768	0.288511
65536	0.341213
131072	0.401236
262144	0.458925
524288	0.578177
1048576	0.705819



Formula:

Area = A1 + A2

A1(2KB L1 with Blocksize = 32, ASSOC = 4) = 0.018662359 mm^2

A2(64KB with Blocksize = 32, ASSOC =8) = 0.360317611 mm^2

Discussion

1) The 64KB L2 cache with L1 cache of 2KB has the lowest AAT of 0.710306 ns.
2) Best_AAT_Total_Area = 0.397897997 mm^2
   The only memory with 5% of the above's AAT, with a smaller area, is the 64KB L2 cache with L1 cache of 1KB. The smallest five L1 caches of the 32KB L2 configuration all have smaller areas, but are not within 5% of the area. The first three are within 5% of the AAT however.

Section 9.4 Stream Buffer Study

L1 cache: SIZE = 1KB - 16KB, ASSOC = 4, BLOCKSIZE = 16 - 64.
L2 cache: SIZE = 64KB, ASSOC = 8, BLOCKSIZE = same as L1's block size.
L1 prefetcher: number of stream buffers = 1 to 4, stream buffer depth = 1 to 4 blocks.
L2 prefetcher: none.

Formula Used:

Miss_Penalty (in ns) = 20 ns + 0.1*(BLOCKSIZE / 16 B/ns)

Average access time (AAT) = HT_L1 + MR_L1* (HT_L2 + MR_L2*Miss_Penalty)

1) Using the formula above, the calculated minimum AAT is 0.2303341 ns, for a Blocksize of 64b, L1 cache size of 4KB, and 4 way associativity, with a Stream Buffer 4 deep and 4 wide (all with the inluded L2 cache of 64KB, 8 way associativity).
This configuration has a total area of 0.393085372 mm^2, and a total energy of 0.02315942 nJ.

2) After calculating the within %5 AATs of the other configurations, 1 setting comes out top in terms of minimum energy, area, and AAT. The 2KB L1 cache with 64b blocksize costs 0.37409918 mm^2 and 0.0226543 nJ, and ranks at 0.2343843195 ns for AAT. This makes it the best overall cache setting of the variety we simulated, because it combines cheaper area and energy consumption with a very similar AAT to the minimum.