NC State University

Department of Electrical and Computer Engineering

ECE 463/521: Spring 2012 (Rotenberg)

Project #1: Cache Design, Memory Hierarchy Design

by

<< Munawira Kotyad >>

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

Student's electronic signature: Munawira Kotyad

(sign by typing your name)

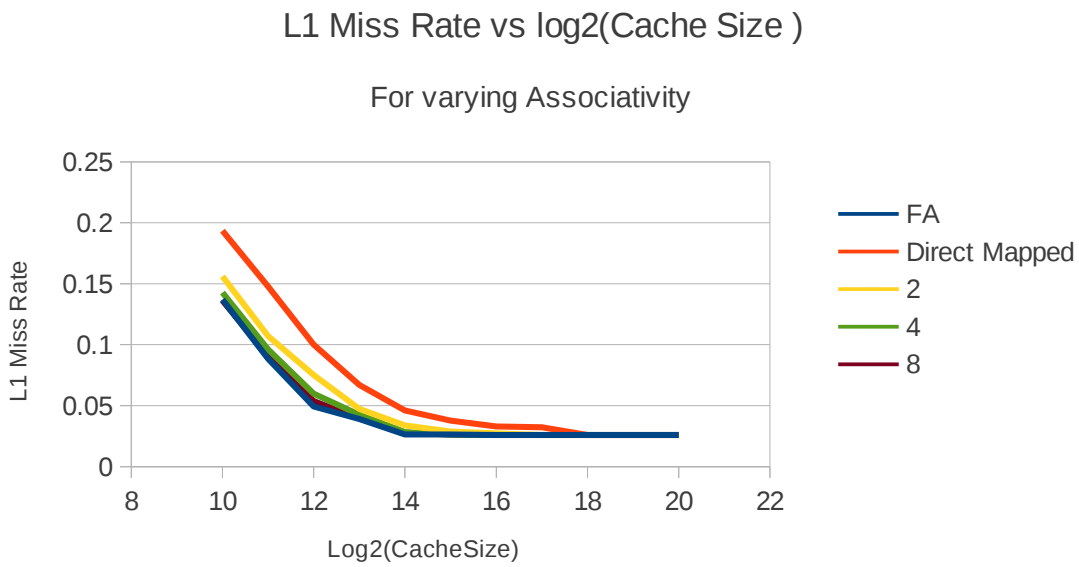
Course number: 521

(463 or 521 ?)

Graph # 1

1. L1 Cache: Size varied, Assoc varied, Block Size = 32
2. L2 Cache: None.
3. Prefetching: None.

The Graph shows the L1 miss rate versus log2(Size) for different values of Associativity.



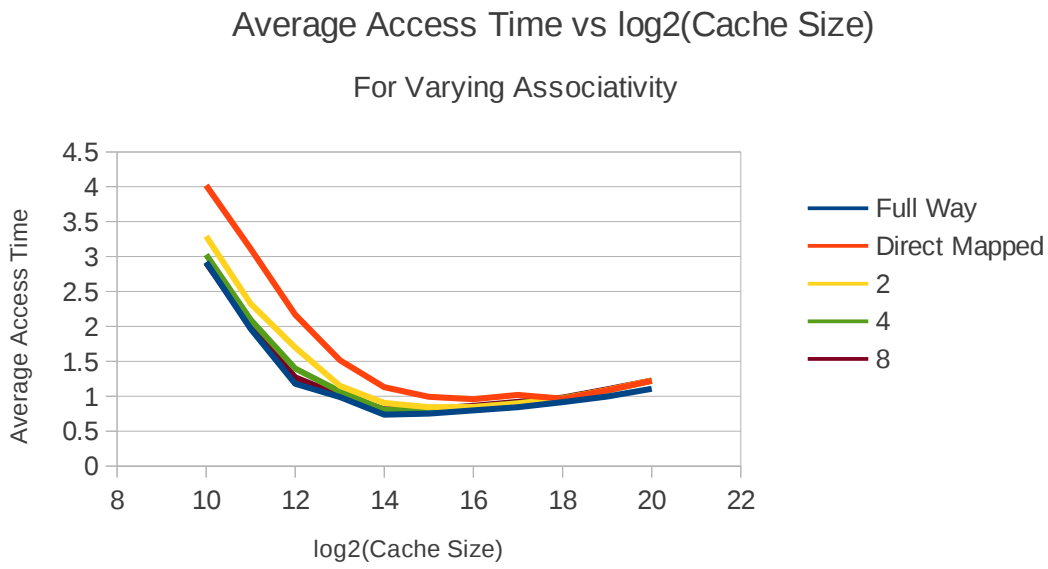
From the graph we observe the following

1) For a given Associativity, as we increase the Cache Size, the L1 miss rate goes on decreasing. This is because as the cache size increases, the capacity misses become negligible, and the misses only comprise of capacity missses.
   As we increase the associativity for a given cache size, the miss rate decreases as the number of conflict misses reduces and becomes constant until only the capacity and compulsory misses remain.
2) The compulsory miss rate from the graph and table can be estimated as 0.2582.
3) The conflict miss rate for each associativity can be estimated as follows:
   - Direct Mapped : 0.0565
   - 2-way : 0.0197
   - 4-way : 0.0574
   - 8-way: 0
   - Fully Associative: 0

Graph #2

1. L1 Cache: Size varied, Assoc varied, Block Size = 32
2. L2 Cache: None.
3. Prefetching: None.

The Graph shows the varying Average Access Time vs log2(Size) for varying values of associativity.

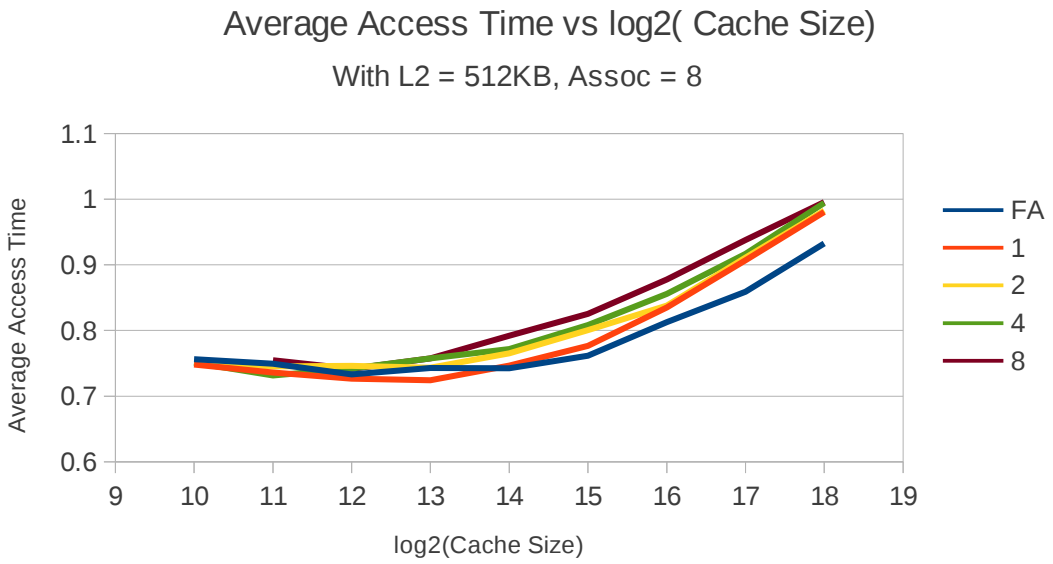


1. For the given memory heirarchy, a cache size of 16Kb with full way associativity, gives the lowest Average Access Time of 0.737676.

Graph# 3

1. L1 Cache: Size varied, Assoc varied, Block Size = 32
2. L2 Cache: Size:512KB, 8-way Set Associativity, Block Size = 32
3. Prefetching: None.

The Graph shows the varying Average Access Time vs log2(Size) for varying values of associativity.



1. The following configurations give results close to the best AAT in graph# 2.

- Fully Associative cache from 1KB to 32 KB
- Direct Mapped cache from 1KB to 16KB
- 2-way set associative cache from 1KB to 16KB
- 4-way set associative cache from 1KB to 16KB
- 8-way set associative cache from 2KB to 8KB

2. The configuration that gives the best AAT is Direct Mapped cache of 8KB with an access time of 0.724138 ns.
Compared to the optimal AAT of graph# 2 i.e. 0.737676 , the optimal AAT with L2 cache is lesser by 0.013544 ns.
3. The total area for the optimal configuration in graph #3 is the sum of
Area of direct mapped 8KB cache with blocksize 32B = 0.053293238424 mm^2
Area of 8-way set associative 512KB cache with blocksize 32B = 2.640142073492

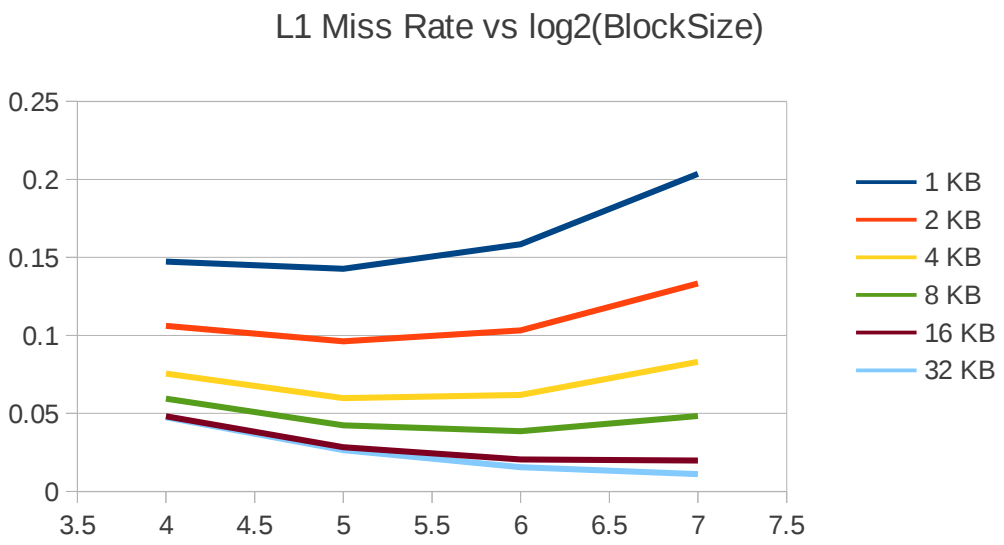
The Total Area = 2.693435312mm^2
Area of the optimal configuration of graph#2 = 0.063446019 mm^2

Thus with an L2 cache the area increases almost 43 times, compared to that of graph #2.

Graph# 4

1. L1 Cache: Size varied, Assoc = 4, Block Size varied
2. L2 Cache: None.
3. Prefetching: None.

The graph shows the L1 miss rate vs log2(BlockSize) for varying values of Cache Size



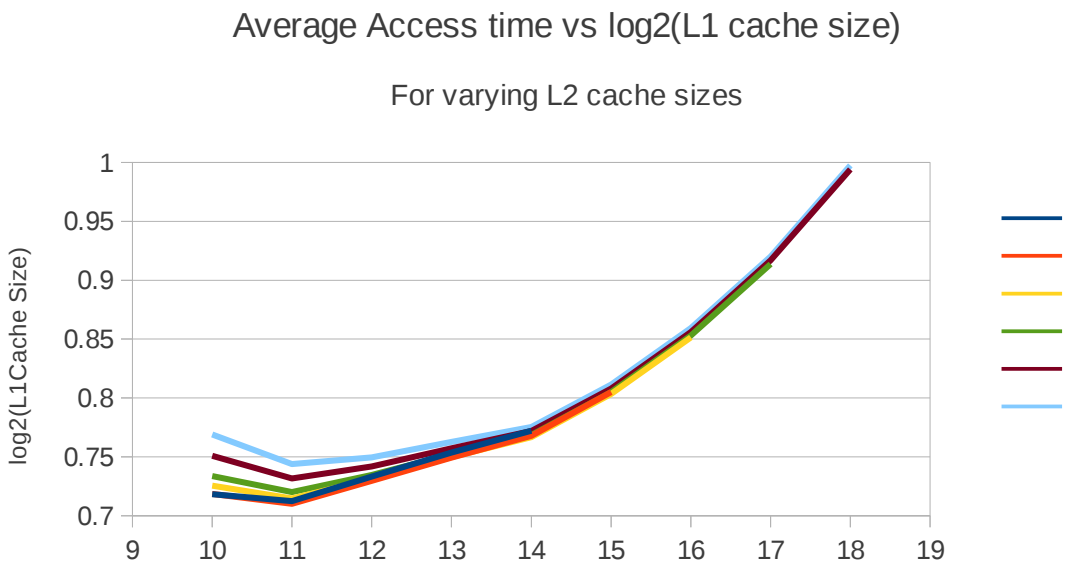
- As we increase the block size, it helps exploit ,more spatial locality, on the other hand it also leads to cache pollution due to fetching of unwanted data from the memory.
- For smaller Cache sizes, the miss rate increases as we increase the block size. Thus for smaller cache sizes, smaller block sizes are preferable. This is because for smaller cache sizes, bringing in larger block sizes leads to cache pollution, and thus higher capacity misses.
- Larger cache sizes prefer larger block sizes as it helps exploit more spatial locality. Larger block sizes do not cause cache pollution for larger cache sizes, as the size of the cache is enough to accommodate a large number of blocks even with larger block sizes.

- From the graph it is evident that, increasing the block size reduces the miss rate uptil a certain point and then the miss rate increases again for smaller cache sizes due to cache pollution.

Graph# 5

1. L1 Cache: Size varied, Assoc = 4, Block Size = 32
2. L2 Cache: Size varied, Assoc = 8, Block Size = 32
3. Prefetching: None.

In this graph, L1 and L2 cache sizes are varied with respect to each other keeping the associativity and block size constant for both.



1. The following configuration gives the lowest AAT amongst all the configurations.
   - L1 cache size 2KB , Assoc = 4, BlockSize = 32
   - L2 cache size 64KB, Assoc = 8, BlockSize = 32
   - The AAT for the above configuration is = 0.7103057496 ns.

2. The configuration that has the smallest area is
   - L1 cache size 1KB , Assoc = 4, BlockSize = 32
   - L2 cache size 32KB, Assoc = 8, BlockSize = 32

The total area is = 0.242170635 + 0.015114948 = 0.25728558 mm^2

STREAM BUFFER STUDY

In this experiment we vary the following parameters:

- 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.

1. The lowest AAT is observed for the following configuration:

- L1 cache: SIZE = 4KB, ASSOC = 4, BLOCKSIZE = 64.
- L2 cache: SIZE = 64KB, ASSOC = 8, BLOCKSIZE = 64
- L1 prefetcher: number of stream buffers = 4, stream buffer depth = 4 blocks.

The AAT for this configuration is 0.2303337842 ns.

2. The total Area for the configuration above is
   - Area of L1 = 0.037648551369
   - Area of L2 = 0.355436821278

     The total area is 0.393085373

As the die area increases , the cost of manufacturing also increases. Thus we should try to find another configuration, that gives an AAT close to the optimal one, but with a lower die area.

The two configurations could be
1.
- L1 cache: SIZE = 2KB, ASSOC = 4, BLOCKSIZE = 64.
- L2 cache: SIZE = 64KB, ASSOC = 8, BLOCKSIZE = 64
- L1 prefetcher: number of stream buffers = 2, stream buffer depth = 4 blocks.

AAT = .234073 ns

2.

- L1 cache: SIZE = 4KB, ASSOC = 4, BLOCKSIZE = 64.

- L2 cache: SIZE = 64KB, ASSOC = 8, BLOCKSIZE = 64
- L1 prefetcher: number of stream buffers = 3, stream buffer depth = 1 blocks.
  AAT = .230989 ns

These two cache configurations can be explored, to lessen the Area and also in effect reduce the energy consumed. The Average Access Time for both the configurations does not vary more than 5% from that of the optimal configuration.