University of Central Florida

Department of Computer Science

CDA 5106: Fall 2020

Machine Problem 1: Cache Design, Memory Hierarchy Design

by

<< Nabila Shahnaz Khan >>

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

Student's electronic signature: _____Nabila Shahnaz Khan_____
(sign by typing your name)

L1 cache exploration: SIZE and ASSOC

Input for this experiment:

- BLOCKSIZE = 32
- L1 cache SIZE: varied (11 different cache sizes: 1KB, 2KB, … , 1MB)
- L1 Associativity: varied (5 different associativity: direct-mapped, 2-way set-associative, 4-way set-associative, 8-way set associative, and fully- associative)
- L2 cache: None
- Replacement policy: LRU
- Inclusion property: non-inclusive
- Total number of simulations: 55

GRAPH #1

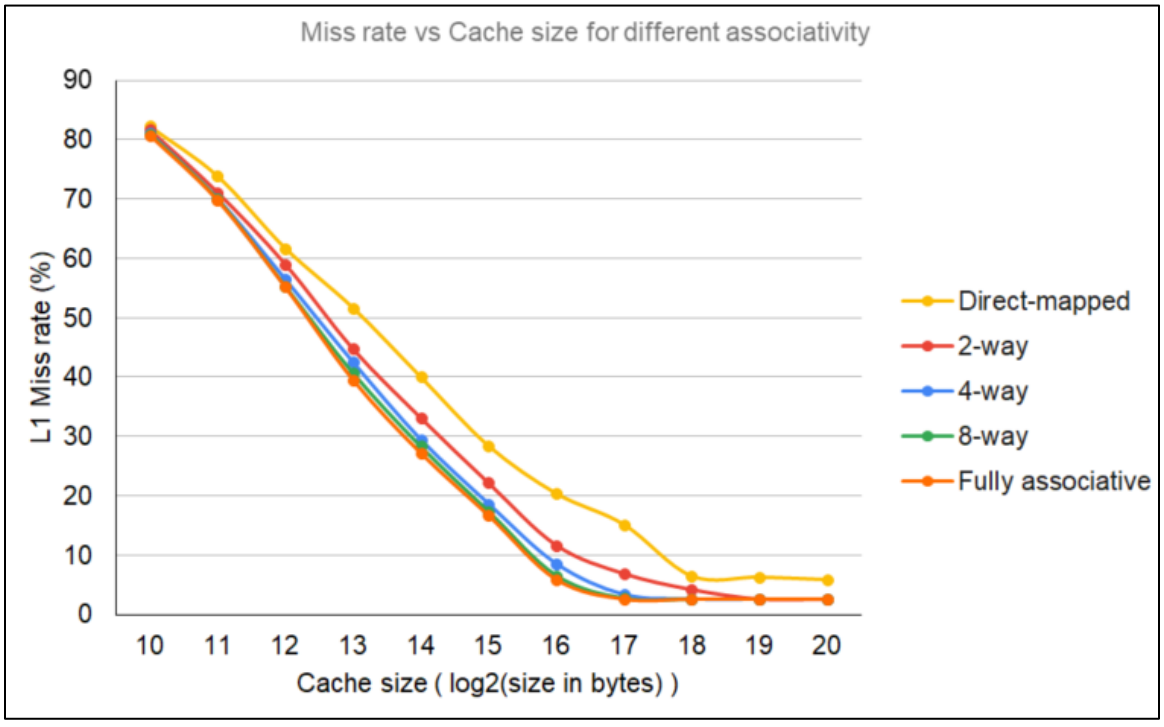

Figure 1: Miss rate versus log2(SIZE) of L1 cache for different associativity

Discussion:

Topic 1:

Relation between miss rate and cache size:

From Figure 1, it is clearly visible that for a given associativity, miss rate gradually decreases with increasing cache size. Initially, the decrease rate is pretty high, but as the cache size keeps increasing, the difference in miss rate becomes less visible (i.e. cache size 256 KB, 512 KB, 1 MB). This finding is reasonable because larger cache will be able to store more data which will eventually result in smaller miss rate.

Relation between miss rate and associativity:

From Figure 1, we can see that for a given cache size, as the associativity increases, the miss rate decreases slightly. Higher associativity increases the possible slot location for a single data, which results in lower miss rate. Though there is an inverse relation between miss-rate and associativity, still the decrease in the miss rate isn't significantly high.

Topic 2:

Compulsory Miss Rate:

As we can see from the figure, after a certain cache size, the miss rate becomes nearly constant. This is because even for a large cache, some cache misses will occur as the cache initially remains empty. This is known as compulsory miss rate. Here, the compulsory miss rate is 2.582%.

Topic 3:

Conflict Miss Rate:

Conflict miss occurs when several blocks are mapped to the same set. It occurs more in direct-mapped cache and less in fully associative cache. After 64 kb, miss rate becomes constant (2.582%) for fully associative cache. As conflict miss rate is 0 for fully-associative cache, so all the cache misses except for this compulsory cache miss can be considered as conflict miss rate for the direct-mapped and set-associative caches.

GRAPH #2

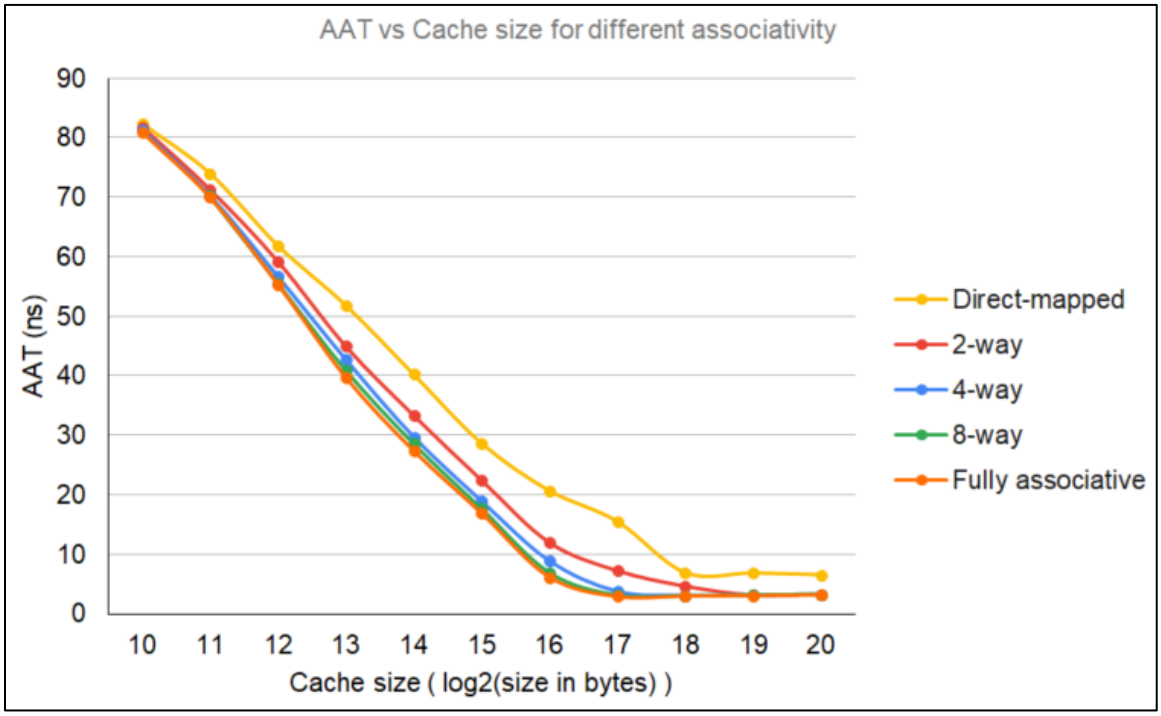

Figure 2: AAT (ns) versus log2(SIZE) of L1 cache for different associativity

Discussion:

From Figure 2, we can see that fully-associative cache has the lowest AAT for a memory hierarchy consisting of an L1 cache with a BLOCKSIZE of 32 bits. So, we can say that the fully-associative cache yields best for this configuration of memory hierarchy.

Replacement policy study

Input for this experiment:

- BLOCKSIZE = 32
- L1 cache SIZE: varied (9 different cache sizes: 1KB, 2KB, … , 256KB)
- L1 Associativity: 4
- L2 cache: None

- Replacement policy: varied (LRU, Pseudo-LRU, Optimal)
- Inclusion property: non-inclusive
- Total number of simulations: 27

GRAPH #3

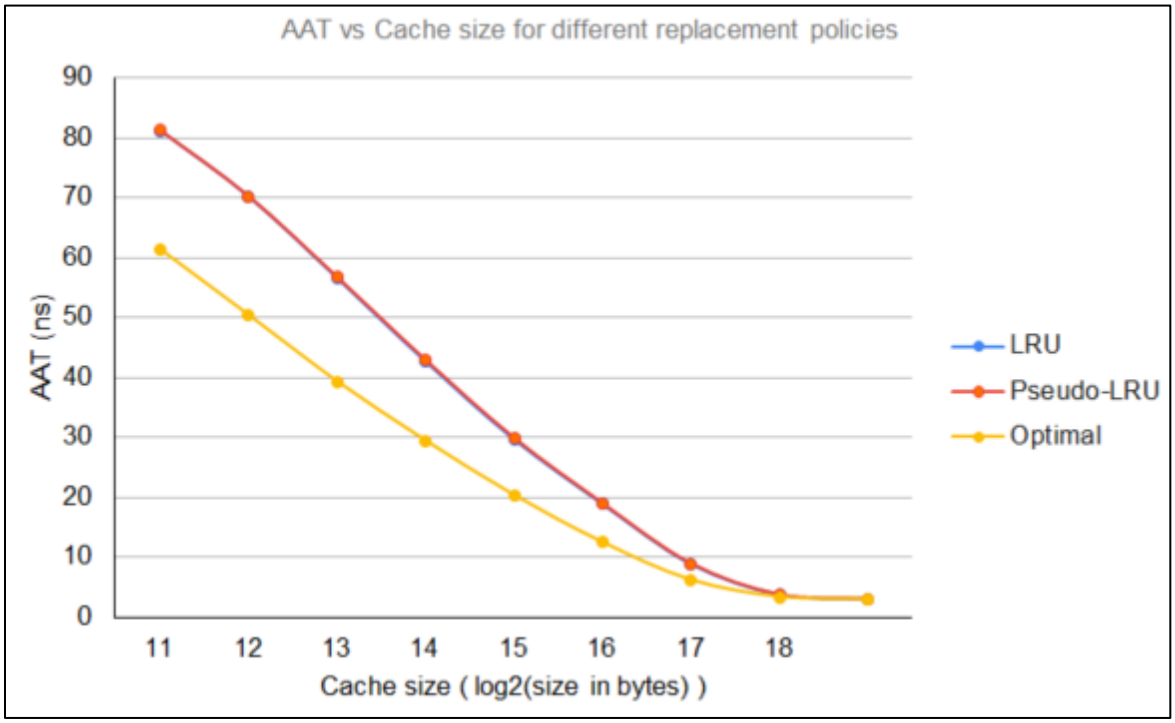

Figure 3: AAT (ns) versus log2(SIZE) of L1 cache for different replacement policies

Discussion:

Here, for 3 replacement policies (LRU, Pseudo-LRU, Optimal), Figure 3 shows the average access time of cache L1 with respect to different cache size. As we can see, for 'Optimal replacement policy', the AAT is lowest for all cache sizes. This means optimal replacement policy works best for these cache configurations. But unfortunately, it is not possible to implement optimal replacement policy in real-time. LRU and Pseudo-LRU performs almost similar as they both have almost similar AAT for all cache sizes. If we look closely, we will see that the AAT of LRU is slightly less compared to Pseudo-LRU. So, LRU performs slightly better than Pseudo-LRU.

Inclusion property study

Input for this experiment:

- BLOCKSIZE = 32
- L1 cache SIZE: 1 KB
- L1 Associativity: 4
- L2 cache SIZE: varied (6 different cache sizes: 2KB – 64KB)
- L2 Associativity: 8
- Replacement policy: LRU
- Inclusion property: varied (non-inclusive, inclusive)
- Total number of simulations: 12

GRAPH #4

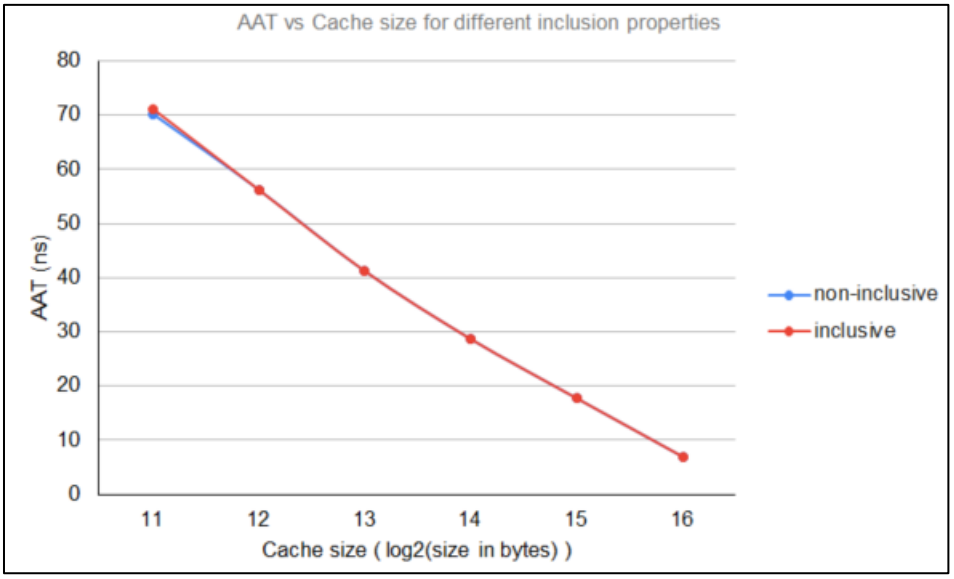

Figure 4: AAT (ns) versus log2(SIZE) of L1 cache for different inclusion policies

Discussion:

As we can see from Figure 4, for different sizes of cache L2, the AAT is almost similar for inclusion policy “inclusive” and “non-inclusive”. However, non-inclusive performs slightly better for 2 KB cache size. So, for the given configurations of cache, non-inclusive policy yields better and might be a better choice.