# Assignment 2 - Cache Evaluation

## Abstract

In this assignment, we are aiming to achieve a simulator to evaluate the effects of different structures of cache on a system’s performance. Various combinations of cache size, associativity and words per block affect are considered. Other perspectives such as hit time, RAS latency, CAS latency, DRAM structure are chosen in sense of more realistic simulation of the real world situation.

## Simulation specification

Cache size: 8kB, 32kB, 512kB.

Associativity: direct mapping (1 way), 4 way set associative.

Words per block: 2 or 8

Hit time is chosen according to various cache size.

Miss time: 2 clock cycles plus DRAM access latency.

Combinations as follow:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Simulation No. | Associativity | Cache size (kB) | Words per block | Hit time |
| 1 | Direct mapping | 8 | 2 | 1 |
| 2 | Direct mapping | 32 | 2 | 2 |
| 3 | Direct mapping | 512 | 2 | 3 |
| 4 | Direct mapping | 8 | 8 | 1 |
| 5 | Direct mapping | 32 | 8 | 2 |
| 6 | Direct mapping | 512 | 8 | 3 |
| 7 | 4 way set associative | 8 | 2 | 1 |
| 8 | 4 way set associative | 32 | 2 | 2 |
| 9 | 4 way set associative | 512 | 2 | 3 |
| 10 | 4 way set associative | 8 | 8 | 1 |
| 11 | 4 way set associative | 32 | 8 | 2 |
| 12 | 4 way set associative | 512 | 8 | 3 |

## Latency source

Comparing to short latency (1-3 clock cycle) by accessing data in cache, the main source of latency is access data in DRAM. There are two type of DRAM access latency, RAS and CAS. RAS latency as 72 clock cycles is much higher than a CAS latency of 24 clock cycles. The occurrence of RAS is less than CAS, because of temporal and/or spatial locality. DRAM is specified to have a 64bit wide bus, 2 words per column, and 8 columns per row. In this simulation cache write through is adapted.

Here is considering situation when cache miss, needs access to main memory. For a reading activity, if it is accessing data on same raw as previous activity, it will be a RAS only. Otherwise, if it is access data on different row, clock delay will be both one RAS and one CAS. For a writing activity, because of cache write through, two CAS miss if access same row of previous activity, otherwise two CAS plus one RAS for accessing different row.

## Simulation result analysis