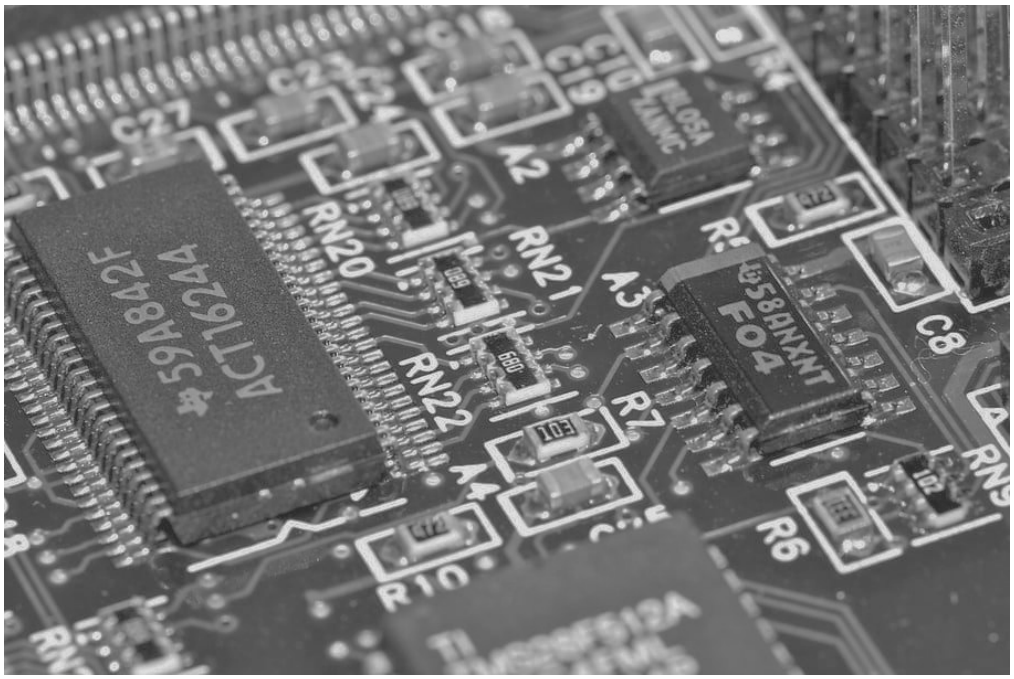
Cache Memory



Jayaweera W.M.P.N
TG/2018/392
Department of ICT
University of Ruhuna

Table of Contents

What is Cache Memory?

Cache memory is a very high speed memory, which is used for speed up and synchronization with the CPU. This memory is more expensive than any other memory but not as expensive as CPU registers.

This memory is located between the RAM and the CPU. It holds the frequently accessed data to provide to the CPU as soon as the CPU is requested. Because of the cache memory, it reduced the time delay between the CPU and the RAM.

Levels of Memory

- Level 1 - Registers
- Level 2 - Cache Memory
- Level 3 - Main Memory
- Level 4 - Secondary Memory

Types of Cache

1. Primary Cache

Primary Cache is located in processor chip. The size of this memory is small and has equal speed as registers. Known as L1 cache.

2. Secondary Cache

This cache memory is placed between the primary cache and the rest of the memory. Known as the L2 cache. This is also located in the processor chip.

The L3 cache is also under the secondary cache category. This is the largest among the all the cache, even though it is slower, it's still faster than the RAM.

Cache Mapping

There are three different types of mapping used for the purpose of cache memory which are as follows.

1. Direct Mapping

Assign each main memory block in a single possible cache line. or
In direct allocation, assign each memory block to a specific line in the cache. If a line is previously occupied by a memory block when a new block needs to be loaded, the previous block is discarded. An address space is divided into a two-part index field and a label field. The cache is used to store the label field, while the rest is stored in main memory.

2. Associative Mapping

Associative memory is used to store content and addresses of the memory word. Any block can go to any line in the cache. This means that the word identification bits are used to identify which block word is needed, but the tag is converted to all remaining bits. This allows the placement of any word anywhere in the cache. It is considered the fastest and most flexible way of mapping.

3. Set-associative Mapping

This form of mapping is an improved form of direct mapping where the inconveniences of direct mapping are eliminated. Establish associative associations to the problem of a possible beating in the direct mapping method. It does this by saying that instead of having exactly one line that a block can allocate in the cache, we will group some lines to create a set. Then, a block in memory can be mapped to any of the lines of a specific set. Associative set allocation allows each word that is present in the cache to have two or more words in the main memory for the same index address.

Set associative cache mapping combines the best of direct and associative cache mapping techniques.

Application of Cache Memory

1. Typically, the cache can store a reasonable number of blocks at any given time, but this number is small compared to the total number of blocks in the main memory.
2. The correspondence between the main memory blocks and those in the cache memory is specified by a mapping function.

-END-