

## CPU Cache Memory

### What is CPU Cache memory?

A CPU Cache(pronounced cash) is a piece of memory located on the CPU chip(the computer's brain that does all the arithmetic and logic calculations). A Cache is a place that temporarily stores information on some recently used or recently accessed locations or tasks. It is far quicker to access data from the Cache memory than from your hard drive or ram(two other forms of storage in a computer).

### How does Cache memory work?

The Cache works by using a system of hit or miss. When the CPU looks to complete an instruction, it checks the Cache to see if the same information or task was accessed recently. If the information the instruction requires is in the Cache, it's a 'hit,' and the information will be returned to the CPU; this is the fastest way to access data. If the information required is not located in the Cache, it's a 'miss,' and your pc will have to go and search for it in slower and further away locations such as your ram or your hard drive.

### How fast is Cache memory?

To understand how fast Cache memory is, we must first talk about RAM(random access memory). RAM is similar to Cache memory but, it is much further away in comparison to Cache memory. We store the information the processor needs to run programs and execute tasks here. For example, if you are playing a video game, the processor will load all the data and information necessary to run the game into the RAM to access it quickly. The Cache is a much smaller but much faster place and, in general, can be 10-100x faster than RAM. (Lutkevich, 2020`)

### Types of Cache Memory

There are commonly at least two types of Cache memory on most CPUs. Cache memory is categorized into levels based on how close they are to the CPU or how fast the CPU can access them. Level 1 or L1 Cache is the fastest and most likely will be built directly into the CPU's chip. Level 2 or L2 Cache is also high-speed but is located a bit further away, so not quite as fast as L1 Cache.

### How do we find things in Cache memory?

Unlike RAM, which uses linear storage, where everything is stored one piece of data after another piece of data, Cache memory uses a storage system more like a pool of memories seemingly less organized. An advanced algorithm is used to find, use and store all these memories. All entries into the Cache are given a tag calculated from their original address. If the CPU ever needs this address again, it will search the Cache for the corresponding tag and return a 'hit.'

### Are there different types of Cache?

There are many different types of Cache; everything that requires the access of the same information over and over again will use a kind of Cache to help speed up this process. Web browsers and other apps also use a cache. For example, when you view a web page, your web browser will load it into its Cache. The next time you want to view the same page, the web browser will use the copy it has saved in the Cache to save time instead of downloading the whole page again. Many other apps will work similarly to this, keeping data in a place that can be accessed quickly instead of going back to the source.

## Are there any downsides to using a Cache?

In most cases, the benefits of using a cache outway the negatives, but it's essential to understand that there can be some issues with using a Cache. A Cache can take up a lot of data if a program saves data to the Cache. A data set in the Cache can become corrupted, it can cause the program to misbehave. In rare cases, the Cache can show a piece of data that is not the most up-to-date version. Issues are fixed by clearing the Cache, meaning to wipe all the data stored there and start fresh. (Johnson, 2020)

## Data Writing Policies

There are two main methods of writing data to Cache memory in modern computer architecture. Write-through and Write-back. Write-through means that RAM and Cache are updated simultaneously, so they both contain the same version of data and are always synchronized; this can be slow as the RAM has to be assessed every time there is an update. Write-back means that the data is only stored firstly in the Cache, and this means that it can be assessed several times before it ever needs to be written to the RAM. (TutorialsInHand, 2020)

## Replacement Policies

Due to the limited size of the Cache, there inevitably needs to be a policy in place for replacing old data with new data. There are a few ways to do this, although none are perfect. A common strategy used is to remove the least recently used data(LRU). (Wikipedia, 2021)

## References

Johnson, D., 2020. *What is a cache? A complete guide to caches and their important uses on your computer, phone, and other devices.* [Online]

Available at: <https://www.businessinsider.com/what-is-cache?r=US&IR=T>

[Accessed 23 10 2021].

Lutkevich, B., 2020. *What is Cache Memory? Cache Memory in computers, Explained.* [Online]

Available at: <https://searchstorage.techtarget.com/definition/cache-memory>

[Accessed 23 10 2021].

TutorialsInHand, 2020. *Cache write policies in computer Architecture.* [Online]

Available at: <https://tutorialsinhand.com/tutorials/computer-organization-and-architecture-coa-tutorial/cache-memory/cache-write-policy.aspx>

[Accessed 24 10 2021].

Wikipedia, 2021. *CPU Cache.* [Online]

Available at: [https://en.wikipedia.org/wiki/CPU\\_cache#Replacement\\_policies](https://en.wikipedia.org/wiki/CPU_cache#Replacement_policies)

[Accessed 25 10 2021].