

How does CPU Cache work? What is L1, L2 and L3 Cache?

Computer processors have advanced quite a bit in recent years. Transistors get smaller and smaller with each passing year, and advances reach the point where Moore's law is becoming redundant.

When it comes to microprocessors, what matters is not only the number of transistors and frequency but also the cache. You may have heard of cache memory when the CPU (Central Processing Unit) was discussed. So how important exactly is CPU cache and how does it work?

What is CPU memory cache?

Simply put, CPU memory cache is just a really fast type of memory. In the early days of computers, processor speeds and memory speeds were very low. However, in the 1980s, processor speeds began to increase rapidly. System memory at the time (RAM) could not keep up or match the increasing CPU speed, and so a new type of ultrafast memory was born: CPU memory cache.

Now, your computer has many types of memory inside it. Having primary storage, like a hard drive or SSD, stores most of the data - the operating system and programs.

Next, we have Random Access Memory, commonly known as RAM. This memory is much faster than main memory but is only a short-term storage medium. Computers and the programs on them use RAM to store frequently accessed data, making computer operations smooth and fast.

Finally, the CPU also has faster memory units within itself, called CPU memory caches.

Computer memory has a hierarchy based on how fast it works. The CPU cache is at the top of this hierarchy and is the fastest memory. It is also the closest place to where the central processing takes place, part of the CPU itself.

Computer memory also comes in different types. Cache memory is a form of static RAM (SRAM), while your regular system RAM is called dynamic RAM (DRAM). Static RAM can hold data without constant refresh, unlike DRAM, which makes SRAM ideal for cache memory.

How does CPU Cache work?

Programs and applications on a computer are designed as a set of instructions that the CPU interprets and runs. When you run a program, instructions go from main memory (hard drive) to the CPU. This is where the memory hierarchy comes into play.

First, the data is loaded into RAM and then sent to the CPU. Today's CPUs are capable of executing a large number of instructions per second. To take full advantage of it, the CPU needs access to super-fast memory, which is where CPU cache comes in.

The memory driver takes data from RAM and sends it to the CPU cache. Depending on your CPU, the driver is found on the CPU or the Northbridge chipset (northbridge chipset) is found on the motherboard.

The cache memory then performs the data transfer back and forth within the CPU. The memory hierarchy also exists in the CPU cache.

CPU cache memory levels: L1, L2 and L3

CPU cache memory is divided into 3 "levels": L1, L2 and L3. The memory hierarchy is back to speed and, therefore, to the size of the cache.

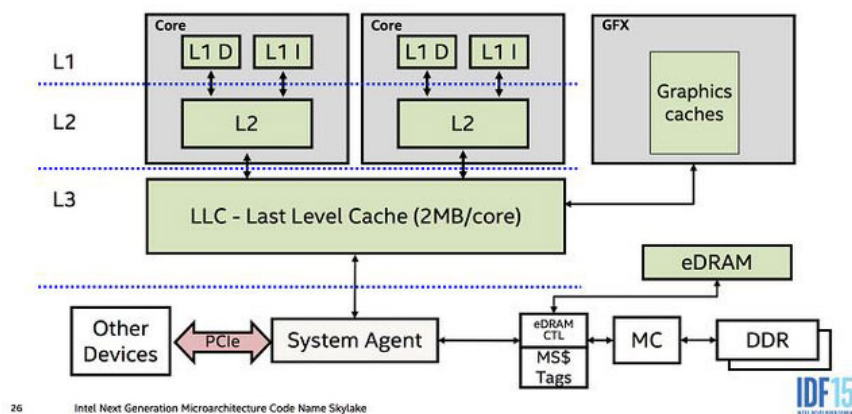
So does CPU cache size make a difference to performance?

L1 Cache

L1 Cache is the fastest memory available in a computer system. In terms of access priority, the L1 Cache owns the data the CPU is most likely to need while completing a certain task.

The size of the L1 Cache depends on the CPU. Some of today's top consumer CPUs have 1MB L1 Cache, like the Intel i9-9980XE, but they cost a lot of money and the number of options is still small. Some server chipsets, like Intel's Xeon series, also have 1 - 2MB of L1 cache.

There is no "standard" L1 Cache size, so you must check your CPU's specs to determine the exact L1 cache size before purchasing.



L1 cache is usually divided into two parts: instruction cache and data cache. The instruction cache processes information about the operation that the CPU has to perform, while the data cache stores the data on which the operation is to be performed.

L2 Cache

L2 Cache is slower than L1 Cache but larger in size. Where L1 Cache can be measured in kilobytes, while modern L2 Cache is measured in megabytes. For example, AMD's highly rated Ryzen 5 5600X has a 384KB L1 Cache and a 3MB L2 Cache (plus 32MB L3 Cache).

L2 Cache size varies depending on CPU, but its size is usually from 256KB to 8MB. Most modern CPUs will have more than 256KB L2 Cache and this size is currently considered small. Furthermore, some of the most powerful modern CPUs have larger L2 caches that exceed 8MB.

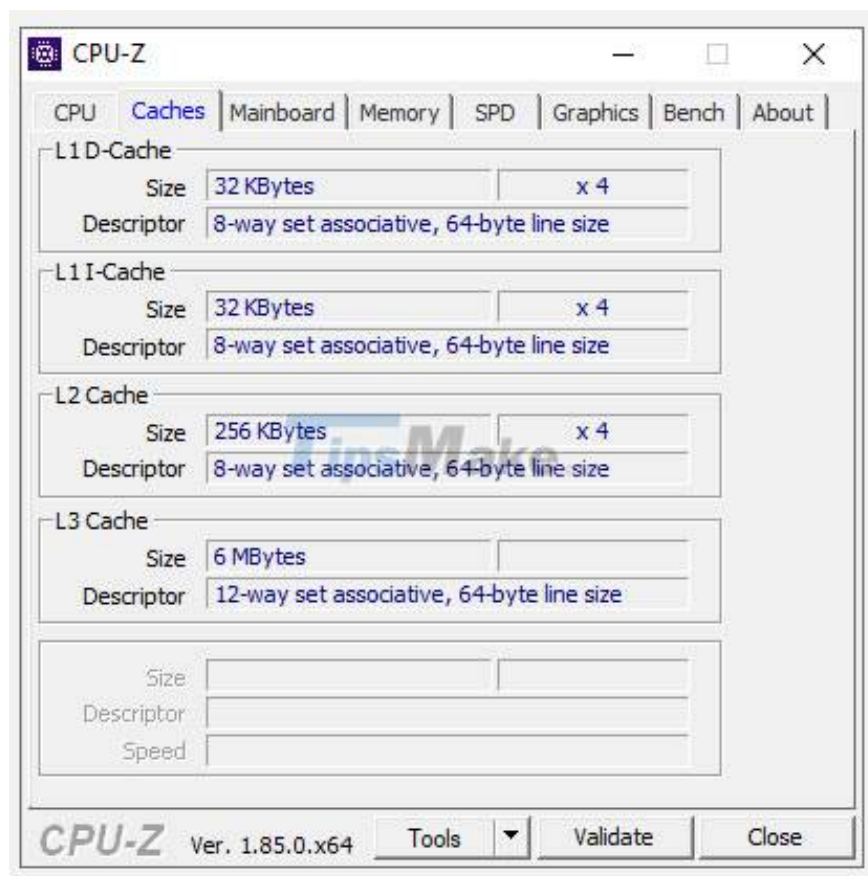
When it comes to speed, L2 Cache loses to L1 Cache but is still much faster than system RAM. L1 cache is typically 100 times faster than RAM, while L2 Cache is about 25 times faster.

L3 Cache

In the early days, the L3 cache was actually found on motherboards. This happened a long time ago, when most CPUs were just single core processors. Now, L3 Cache in CPUs can be huge, with top consumer CPUs having L3 Cache up to 32MB. Some server L3 Cache CPUs can exceed this, with capacities up to 64MB.

L3 Cache is the largest cache but also the slowest. Modern CPUs include the L3 Cache on top of the CPU itself. But while the L1 and L2 Cache exist for each core on the chip itself, the L3 Cache resembles a common pool of memory that the entire chip can use.

The following image shows the CPU cache memory levels for the Intel Core i5-3570K CPU:



Notice how the L1 Cache is split into two, while L2 and L3 are respectively larger.

You finished reading the article "**How does CPU Cache work? What is L1, L2 and L3 Cache?**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.
