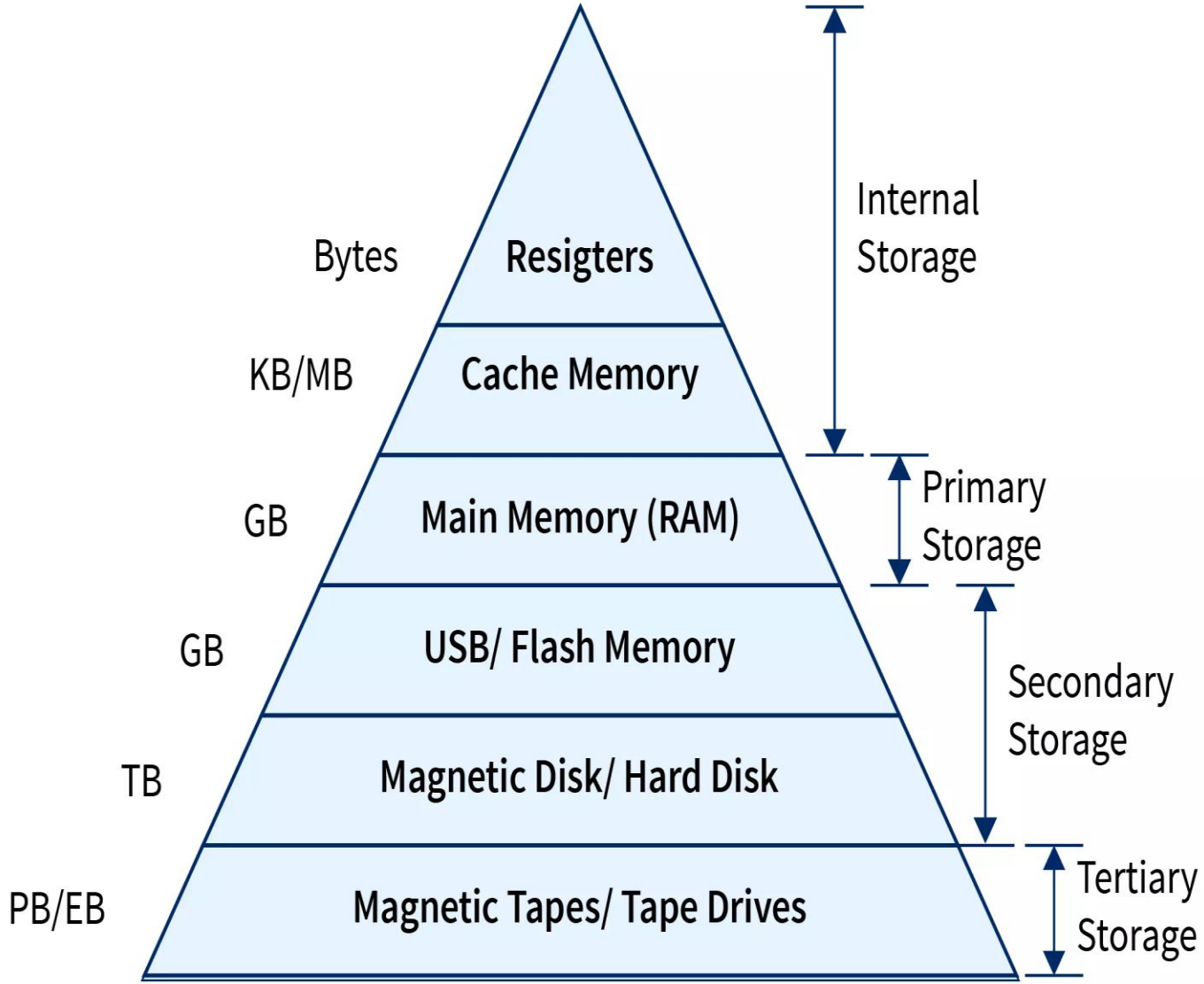
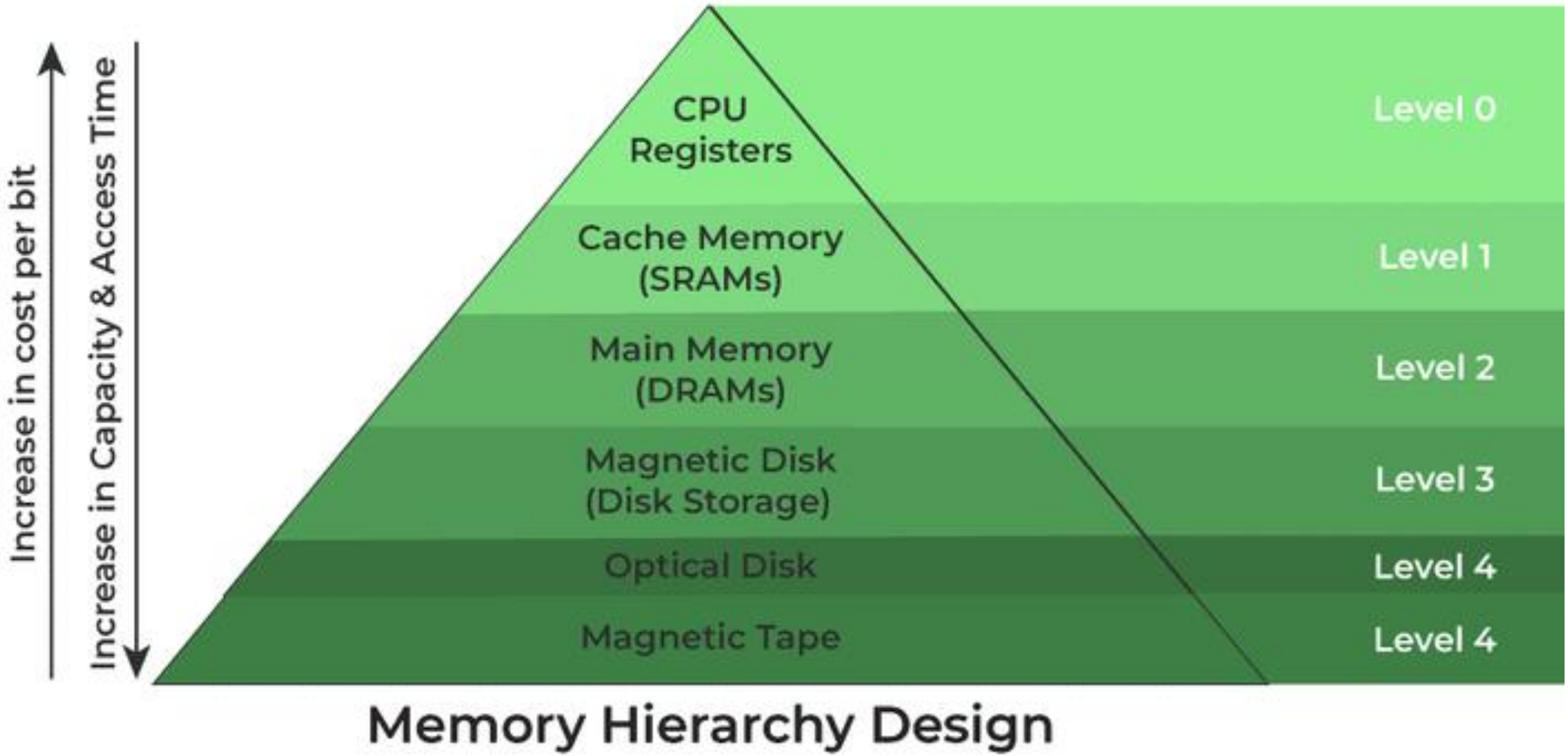
Cache memory


Bytes
Resigters
KB/MB
Cache Memory
GB
Main Memory (RAM)
GB
USB/ Flash Memory
TB
Magnetic Disk/ Hard Disk
PB/EB
Magnetic Tapes/ Tape Drives
Internal Storage
Primary Storage
Secondary Storage
Tertiary Storage



Memory Hierarchy Design

2.2.6	compare the three types of cache memory, i.e. Level-1 (L1) cache, Level-2 (L2) cache and Level-3 (L3) cache;

Cache Memory

- Cache memory is a high-speed, small-sized type of **volatile** computer memory that plays a crucial role in improving CPU (Central Processing Unit) performance. It serves as a bridge between the CPU and the slower, larger main memory (RAM). The CPU registers form part of the cache memory.
- The primary purpose of cache memory is to reduce the time it takes for the CPU to access frequently used data and instructions, thereby enhancing overall system performance.
- Cache memory is static ram (SRAM) that stores its data in flop flop circuits.

Cache hits and cache miss

- A cache hit occurs when the CPU requests a piece of data or instruction, and that data is found in the cache memory. In other words, the data needed by the CPU is already present in the cache.

- A cache miss occurs when the CPU requests data or instructions that are not present in the cache memory. In this case, the cache does not contain the required data, and the CPU must fetch it from a slower memory hierarchy level, typically from the main memory (RAM).
- Cache hits are desirable because they result in **very fast memory access times**. The CPU can retrieve the required data directly from the cache, avoiding the longer latency associated with accessing data from main memory (RAM) or even slower storage devices like hard drives or SSDs.

Quiz

- Cache memory operates at a higher _____compared to the main memory.

Pros & Cons of Cache Memory

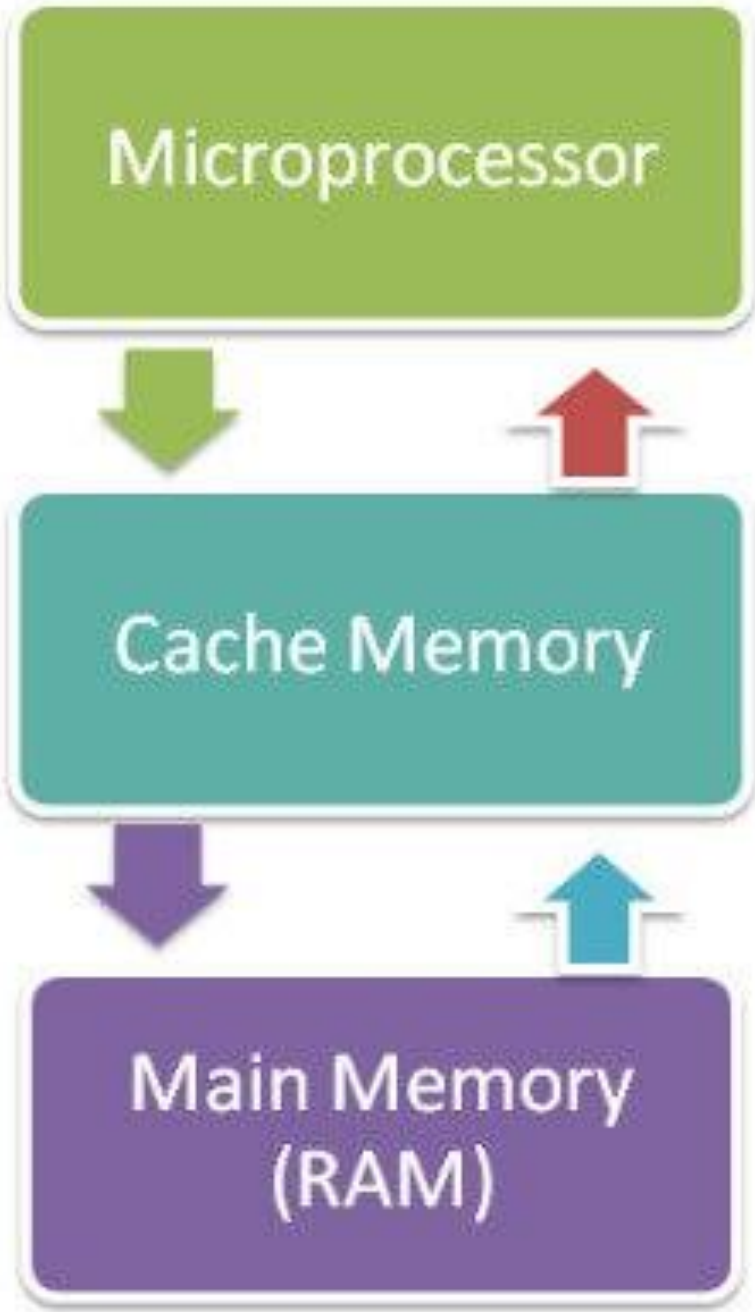
- **Advantages**
- Cache memory is faster than RAM memory because it is located directly on the CPU using flip-flop circuits.
- **Disadvantages**
- Cache memory is more expensive than RAM because the circuitry is more complex than RAM.
- It is also more energy intensive, due to the flip flop circuits being constantly active.

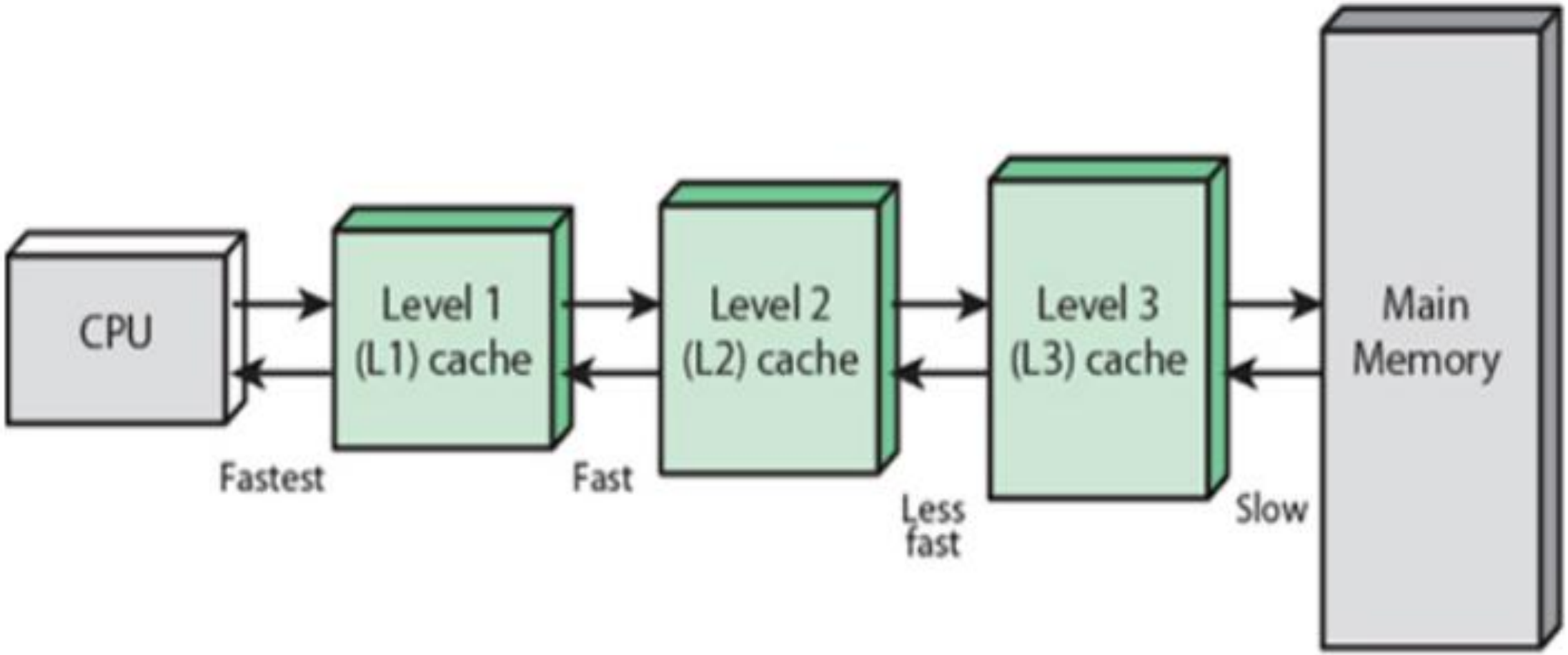
How does cache memory improve computer performance?

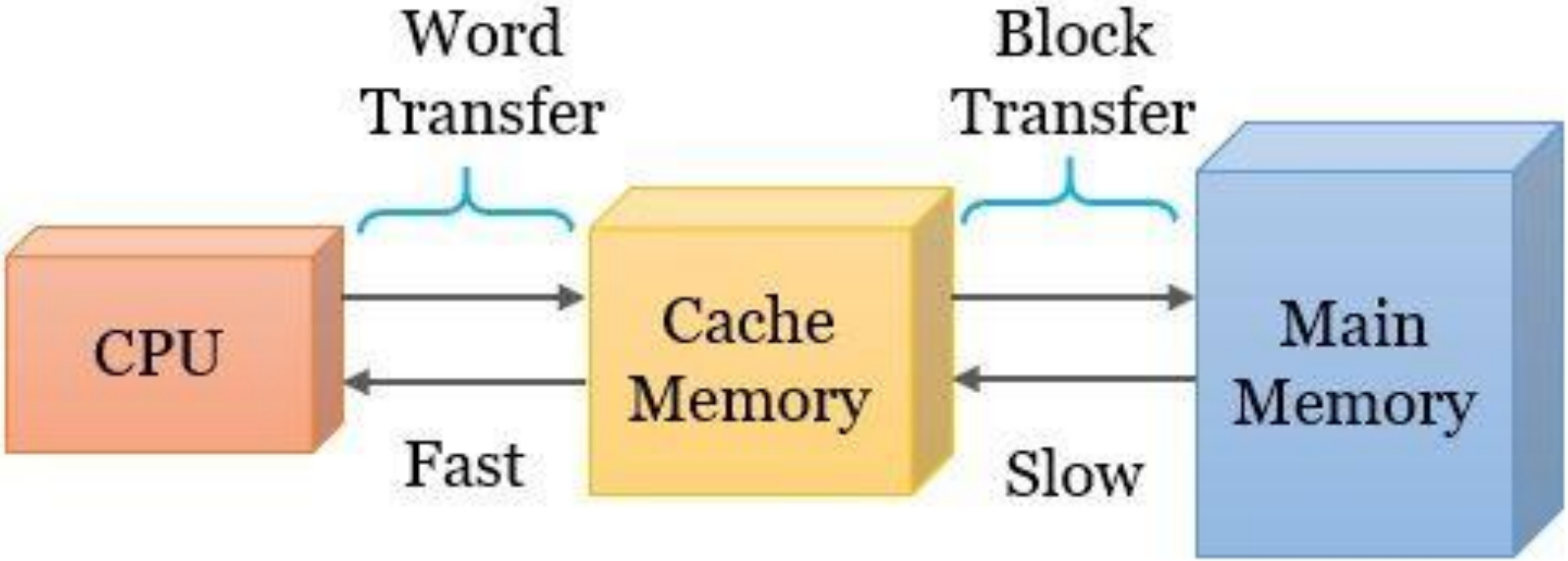
- By expanding the main memory capacity.
- By increasing the CPU speed.
- By reducing the average time to access data.
- None of the above.

How does cache memory improve computer performance?

- By expanding the main memory capacity.
- By increasing the CPU speed.
- By reducing the average time to access data.
- None of the above.


Microprocessor
Cache Memory
Main Memory
(RAM)


CPU
Level 1
(L1) cache
Level 2
(L2) cache
Level 3
(L3) cache
Main
Memory
Fastest
Fast
Less
fast
Slow


Word
Transfer
Block
Transfer
CPU
Cache
Memory
Main
Memory
Fast
Slow

Types of cache memory

- Cache memory is fast and expensive. Traditionally, it is categorized as "levels" that describe its closeness and accessibility to the microprocessor. There are three general cache levels:
- **L1 cache**, or primary cache, is extremely fast but relatively small, and is usually embedded in the processor chip as CPU cache.
- **L2 cache**, or secondary cache, is often more capacious than L1. L2 cache may be embedded on the CPU, or it can be on a separate chip or coprocessor and have a high-speed alternative system bus connecting the cache and CPU. That way it doesn't get slowed by traffic on the main system bus.
- **Level 3 (L3) cache** is specialized memory developed to improve the performance of L1 and L2. L1 or L2 can be significantly faster than L3, though L3 is usually double the speed of DRAM. With multicore processors, each core can have dedicated L1 and L2 cache, but they can share an L3 cache.

References

- https://revise.learnlearn.uk/app/section/576/216