



Memory System Design

Lecture 5

MSc. Riyam K. Marjan

2022-2023



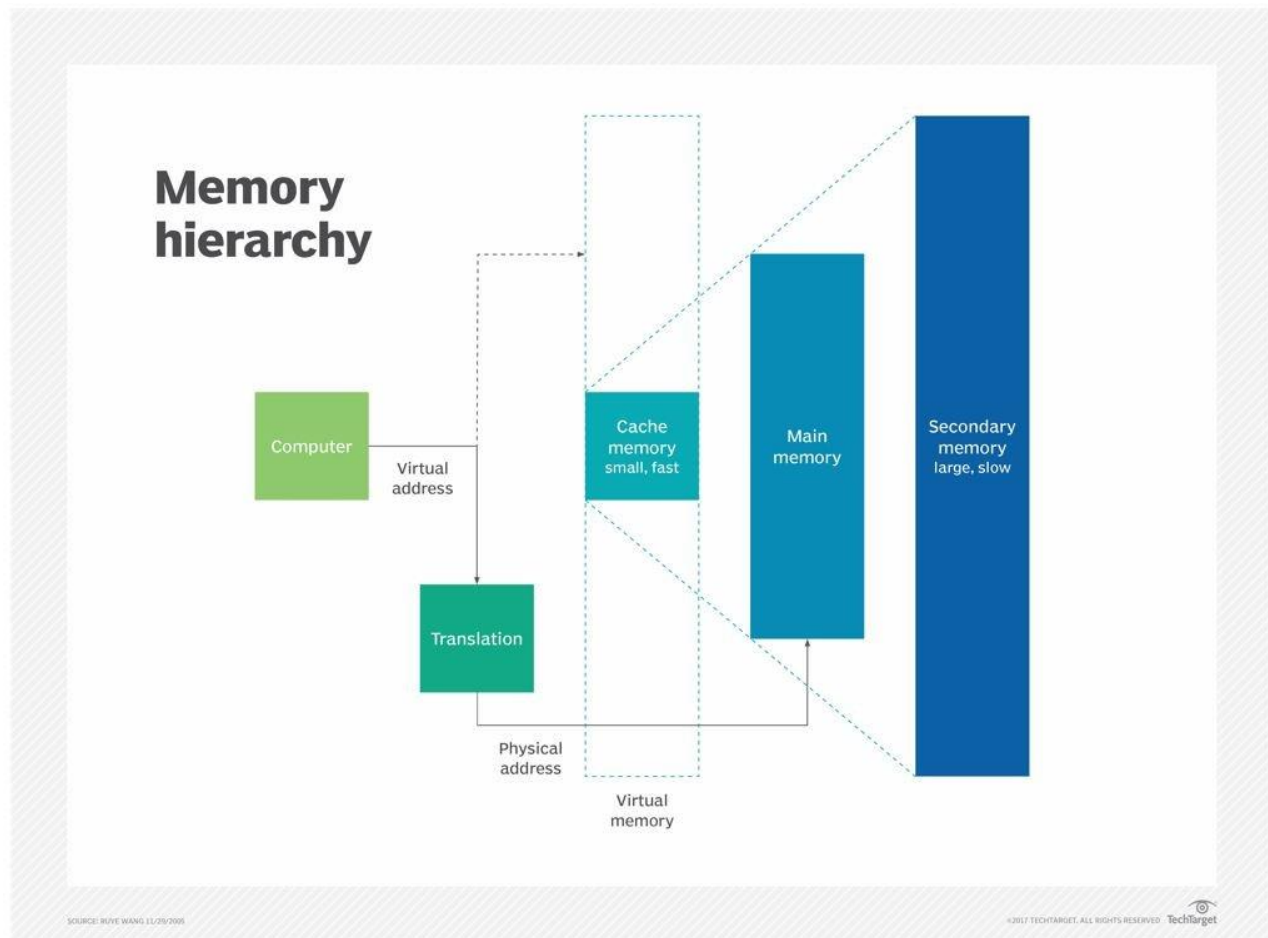
I. Basic Concepts

Memory is essential to the operation of a computer system, and nothing is more important to the development of the modern memory system than the concept of the memory hierarchy.

- Each memory cell can hold one bit of information.
- Memory cells are organized in the form of an array.
- One row is one memory word.

As a result of the trade-off between capacity and performance, a modern memory system typically consists of two components

- Cache : a component that is small but relatively fast-to-access
- Main memory, a component that is large but relatively slow-to-access.

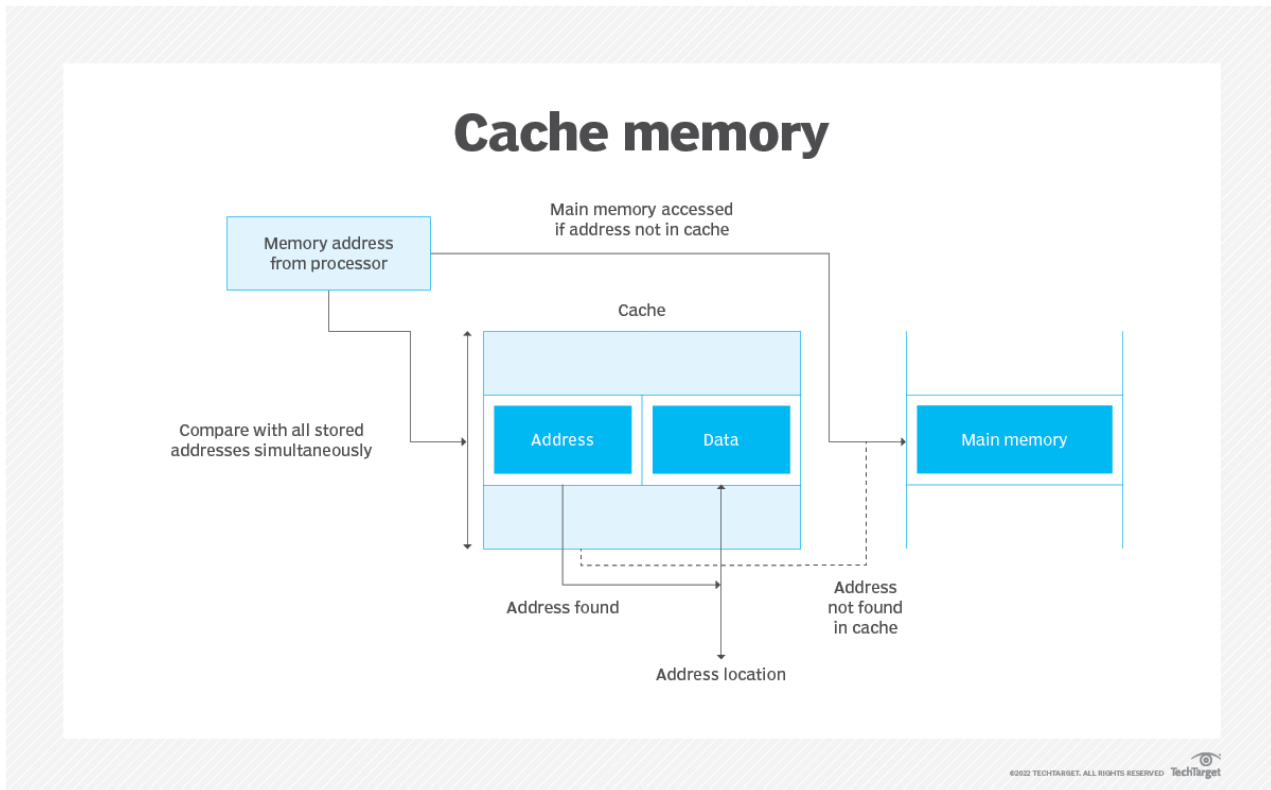


Levels of memory:

- **Level 1 or Register** – It is a type of memory in which data is stored and accepted that are immediately stored in CPU. Most commonly used register is accumulator, Program counter, address register etc.
- **Level 2 or Cache memory** – It is the fastest memory which has faster access time where data is temporarily stored for faster access.
- **Level 3 or Main Memory** – It is memory on which computer works currently. It is small in size and once power is off data no longer stays in this memory.
- **Level 4 or Secondary Memory** – It is external memory which is not as fast as main memory but data stays permanently in this memory.

II. Cache Memory

- Cache memory is a small-sized type of volatile computer memory that provides high-speed data access to a processor and stores frequently used computer programs, applications and data.
- A temporary storage of memory, cache makes data retrieving easier and more efficient. It is the fastest memory in a computer, and is typically integrated onto the motherboard and directly embedded in the processor.



III. Main Memory

The main memory is the fundamental storage unit in a computer system. It is associatively large and quick memory and saves programs and information during computer operations. The technology that makes the main memory work is based on semiconductor integrated circuits.

RAM is the main memory. Integrated circuit Random Access Memory (RAM) chips are applicable in two possible operating modes are as follows –



- **Static** – It consists of internal flip-flops, which store the binary information. The stored data remains solid considering power is provided to the unit. The static RAM is simple to use and has smaller read and write cycles.
- **Dynamic** – It saves the binary data in the structure of electric charges that are used to capacitors. The capacitors are made available inside the chip by Metal Oxide Semiconductor (MOS) transistors. The stored value on the capacitors contributes to discharge with time and thus, the capacitors should be regularly recharged through stimulating the dynamic memory.
- **Random Access Memory**
RAM is the main memory of a computer. Its objective is to store data and applications that are currently in use. The operating system controls the usage of this memory. It gives instructions like when the items are to be loaded into RAM, where they are to be located in RAM, and when they need to be removed from RAM.
- **Read-Only Memory**
In each computer system, there should be a segment of memory that is fixed and unaffected by power failure. This type of memory is known as Read-Only Memory or ROM.
- **SRAM**
RAMs that are made up of circuits and can preserve the information as long as power is supplied are referred to as Static Random Access Memories (SRAM). Flip-flops form the basic memory elements in an SRAM device. An SRAM consists of an array of flip-flops, one for each bit.
- **DRAM**
SRAMs are faster but their cost is high because their cells require many transistors. RAMs can be obtained at a lower cost if simpler cells are used. A MOS storage cell based on capacitors can be used to replace the SRAM cells.

IV. Virtual Memory

- ✚ Virtual memory is an area of a computer system's secondary memory storage space (such as a hard disk or solid state drive) which acts as if it were a part of the system's RAM or primary memory.
- ✚ Ideally, the data needed to run applications is stored in RAM, where they can be accessed quickly by the CPU. But when large applications are being run, or when many applications are running at once, the system's RAM may become full.



- ✚ To get around this problem, some data stored in RAM that is not actively being used can be temporarily moved to virtual memory (which is physically located on a hard drive or other storage device). This frees up space in RAM, which can then be used to accommodate data which the system needs to access imminently.

- ✚ By swapping data between RAM and virtual memory when it is not needed and back from virtual memory to RAM when it is needed, a system can continue to work smoothly with far less physical RAM than it would otherwise require.

- ✚ Virtual memory enables a system to run larger applications or run more applications at the same time without running out of RAM. Specifically, the system can operate as if its total RAM resources were equal to the amount of physical RAM, plus the amount of virtual RAM.