12-1 Memory Hierarchy

The memory unit is an essential component in any digital computer since it is needed for storing programs and data. A very small computer with a limited application may be able to fulfill its intended task without the need of additional storage capacity. Most general-purpose computers would run more efficiently if they were equipped with additional storage beyond the capacity of the main memory.
There is just not enough space in one memory unit to accommodate all the programs used in a typical computer. Moreover, most computer users communicate and continue to store large amounts of data-processing software. Not all accumulated information is needed by the processor at the same time.
Therefore, it is more economical to use low-cost storage devices to serve as a backup for storing the information that is not currently used by the CPU. The memory unit that communicates directly with the CPU is called the **main memory**. Devices that provide backup storage are called **auxiliary memory**.
The most common auxiliary memory devices used in computer systems are magnetic disks and tapes. They are used for storing system programs, large data files, and other backup information. Only programs and data currently needed by the processor reside in main memory.

12-2 Main Memory

The main memory is the central storage unit in a computer system. It is a relatively large and fast memory used to store programs and data during the computer operation. The principal technology used for the main memory is based on semiconductor integrated circuits.

Random-Access Memory (RAM):
RAM consists of memory cells capable of storing and retrieving binary information. Dynamic RAM (DRAM) stores binary information in the form of electric charges that are applied to capacitors. These capacitors need periodic refreshing due to charge leakage. Static RAM (SRAM), on the other hand, uses flip-flops and does not need refreshing.

Read-Only Memory (ROM):
ROM stores programs that are permanently in memory and not subject to change. Used for startup programs (bootstrap loader), ROM contents are not lost when power is turned off.

Bootstrap Loader:
A small portion of ROM contains the initial program required to start a computer once it is powered on.

Computer Startup:
The startup of a computer involves turning the power on and executing the bootstrap loader from ROM.

12-3 Auxiliary Memory

Total memory capacity can be visualized as a **hierarchy** — from slow, cheap auxiliary memory to faster main memory, to cache memory. Auxiliary memory includes magnetic disks and optical storage. These devices serve as the **backup storage**.

Cache Memory:
Cache memory is very fast, small, and expensive. It is used to store **frequently used instructions and data** for fast CPU access. Cache helps overcome the mismatch between CPU speed and main memory access speed.

12-4 Memory Hierarchy (Continued)

The hierarchy helps to balance cost and speed. For example, auxiliary memory access times are usually 1000 times that of main memory. Block sizes in cache: 1 to 16 words; block sizes in auxiliary memory: 256 to 2048 words.

Multiprogramming

Allows the CPU to handle more than one program at a time. It switches rapidly between programs to keep all parts of the system busy. For instance, while one program waits for input, the CPU can execute another.