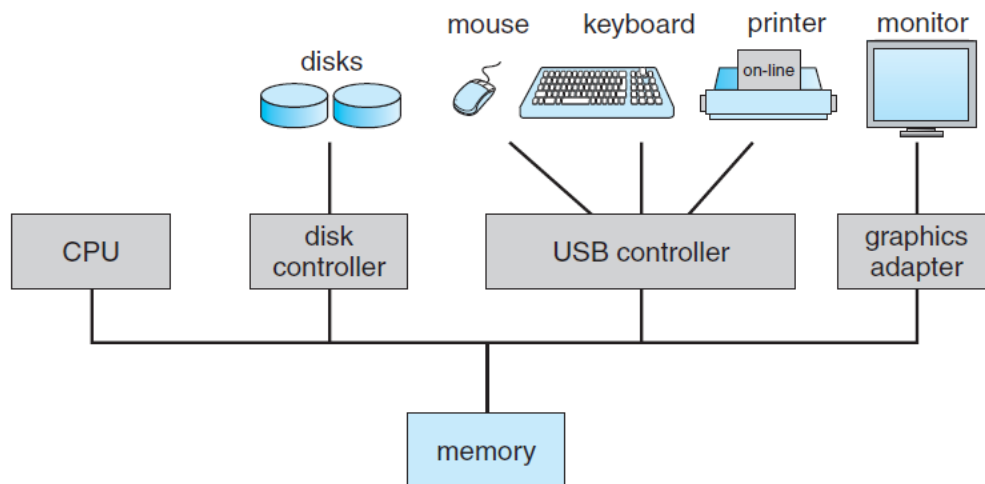




## Computer-System Organization

A **modern general-purpose computer system** consists of one or more CPUs and a number of device controllers connected through a **common bus that provides access to shared memory** as shown in figure 1. Each device controller is in charge of a specific type of device (for example, disk drives, audio devices, or video displays).

The CPU and the device controllers **can execute in parallel**, with a memory controller ensuring synchronized memory access. When powered on or rebooted, the computer runs a simple **bootstrap program** stored in ROM or EEPROM (firmware). This program initializes the CPU, device controllers, and memory, then locates and **loads the operating system kernel** to start the system.



**Figure 1** A modern computer system.

### ➤ STORAGE DEFINITIONS AND NOTATION

The basic unit of computer storage is the **bit**. **A bit can contain one of two values, 0 and 1**. All other storage in a computer is based on collections of bits. A **byte** is 8



bits, and on most computers it is the smallest convenient chunk of storage. A less common term is **word**, which is a given computer architecture's unit of data. A word is made up of one or more bytes. For example, a computer that has **64-bit registers** and 64-bit memory addressing typically has 64-bit (8-byte) words.

- **Computer storage**, along with most computer throughput, is generally measured and manipulated **in bytes** and collections of bytes. A **kilobyte**, or **KB**, is 1,024 bytes; a **megabyte**, or **MB**, is 1,024<sup>2</sup> kilobytes bytes; a **gigabyte**, or **GB**, is 1,024<sup>3</sup> bytes; a Storage Structure

The CPU can **load instructions** only from memory, so any programs **to run must be stored there**. **General-purpose computers** run most of their programs from rewritable memory, called main memory (also called **random-access memory**, or **RAM**). Main memory commonly is implemented in a semiconductor technology called **dynamic random-access memory (DRAM)** **terabyte**, or **TB**, is 1,024<sup>4</sup> bytes; and a **petabyte**, or **PB**, is 1,024<sup>5</sup> bytes. Here is the storage size sequence, in a table.

Order	Unit Name	Abbreviation	Equal
1	Byte	B	8 bits
2	Kilobyte	KB	1,024 Bytes
3	Megabyte	MB	1,024 KB
4	Gigabyte	GB	1,024 MB
5	Terabyte	TB	1,024 GB
6	Petabyte	PB	1,024 TB



### ➤ Storage Structure

The CPU can **load instructions only from memory**, so **any programs to run must be stored there**. General-purpose computers run most of their programs from rewritable memory, called main memory (also called random-access memory, or **RAM**). Main memory commonly is implemented in a semiconductor technology called dynamic random-access memory (DRAM).

All forms of memory provide an array of bytes. Each byte has its own address. Interaction is achieved through a sequence of load or store instructions to specific memory addresses.

The **load instruction** moves a byte or word from main memory to an internal register within the CPU, whereas the **store instruction** moves the content of a register to main memory. In addition to explicit loads and stores, the **CPU automatically loads instructions** from main memory for execution. **Perfectly**, we want the programs and data to reside in main memory permanently. This arrangement usually is not possible for the following two reasons:

1. Main memory is usually too small to store all needed programs and data permanently.
2. Main memory is a **volatile** storage device that loses its contents when power is turned off or otherwise lost.



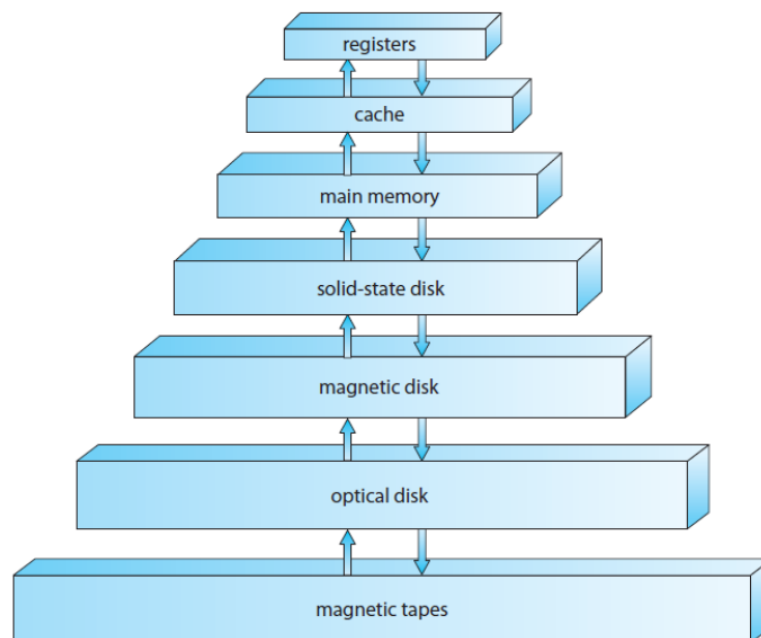
Most computer systems provide **secondary storage** as an extension of main memory. The main requirement for secondary storage is that it be able to hold large quantities of data permanently.

The most common secondary-storage device is a **magnetic disk**, which provides storage for both programs and data. Most programs (system and application) are stored on a disk until they are loaded into memory.

The wide variety of storage systems can be organized in a hierarchy Figure 2 according to speed and cost. The higher levels are expensive, but they are fast.

In addition to differing in speed and cost, the various storage systems are either volatile or nonvolatile. As mentioned earlier, **volatile storage** loses its contents when the power to the device is removed, but **nonvolatile storage** It is not affected by power outages.

Computers use other forms of memory as well. We have already mentioned read-only memory, ROM) and electrically erasable programmable read-only memory, EEPROM). All forms of memory provide an array of bytes. Each byte has its own address.



**Figure 2** Storage-device hierarchy

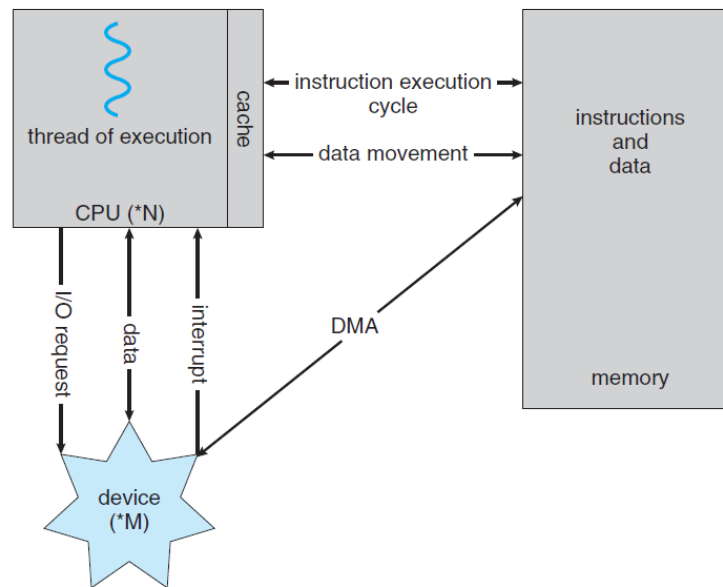


## ➤ I/O Structure

Storage is only one of many types of I/O devices within a computer. A general-purpose computer system consists of CPUs and multiple device controllers that are connected through a common bus. Each device controller is in control of a specific type of device. **The device controller** is responsible for moving the data between the peripheral devices that it controls and its local buffer storage.

**The controller starts** the transfer of data from the device to its local buffer. Once the transfer of data is complete, the device controller informs the device driver via an interrupt that it has finished its operation.

This form of **interrupt-driven I/O** is fine for moving small amounts of data but can produce high overhead when used for bulk data movement such as disk I/O. **To solve this problem, direct memory access (DMA)** is used. After setting up buffers, pointers, and counters for the I/O device, **the device controller transfers an entire block of data directly to or from its own buffer storage to memory, with no intervention by the CPU.** **Figure 3** How a modern computer system works.



**Figure 3** How a modern computer system works.