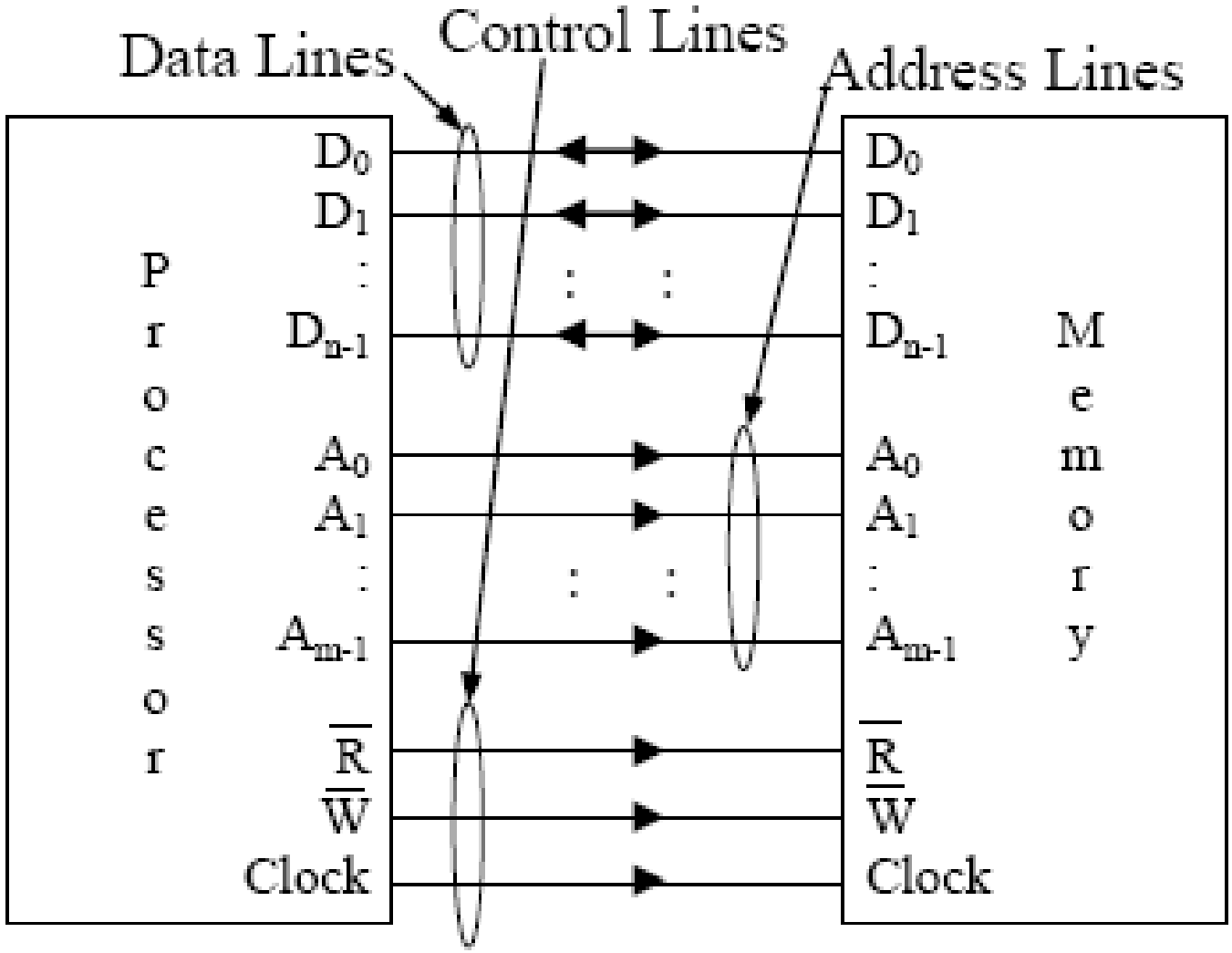
Computer Memory and Memory Types

Lecture 4

Objectives

- Memory interface with processor.
- Memory connection on bus.
- Memory Terminology.
- Memory types & Modules.
  - Main memory types.
  - Secondary memory types.



e 12-3 Basic Processor to Memory Device Interface

A method had to be developed to allow a processor to communicate to multiple memory devices across the same set of wires. If this wasn't done, the processor would need a separate set of data, address, and control lines for each device placing an enormous burden on circuit board designers for routing wires.

By using a bus, the processor can communicate with exactly one device at a time even though it is physically connected to many devices. If only one device on the bus is enabled at a time, the processor can perform a successful data transfer. If two devices tried to drive the data lines simultaneously, the result would be lost data in a condition called *bus contention*.

Figure 12-4 presents a situation where data is being read from memory device 1 while memory device 2 remains "*disconnected*" from the bus. Disconnected is in quotes because the physical connection is still present; it just doesn't have an electrical connection across which data can pass.

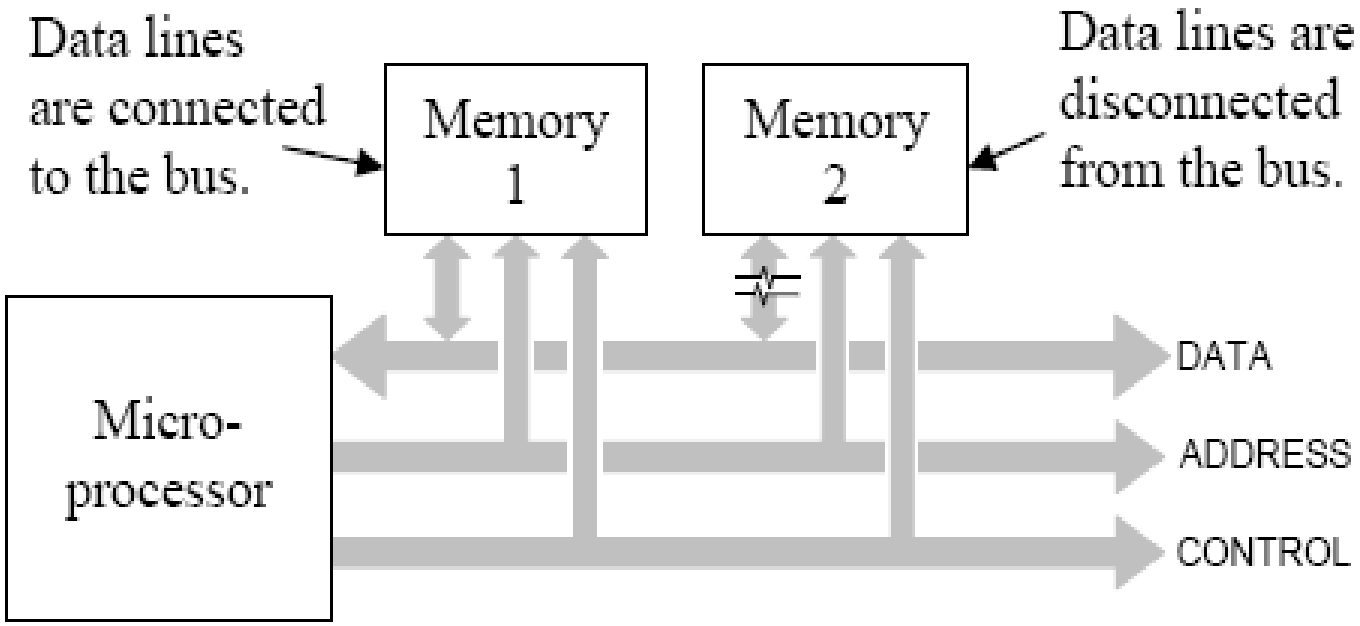


Figure 12-4 Two Memory Devices Sharing a Bus

- Notice that Figure 12-4 shows that the only lines disconnected from the bus are the *data lines*. This is because bus contention only occurs when multiple devices are trying to output to the same lines at the same time. Since only the microprocessor outputs to the address and control lines, they can remain connected.

- In order for this scheme to work, an additional control signal must be sent to each of the memory devices telling them when to be connected to the bus and when to be disconnected. This control signal is called a ***chip select*.**

- A *chip select* is an active low signal connected to the enable input of the memory device. If the chip select is high, the memory device remains **idle** and its data lines are **disconnected** from the bus. When the processor wants to communicate with the memory device, it pulls that device's chip select ***low*** thereby enabling it and connecting it to the bus.

Each memory device has its own chip select. For example, Table 12-1 shows the only possible values of the chip selects for a system with four memory devices.

Table 12-1 The Allowable Settings of Four Chip Selects

	CS_0	CS_1	CS_2	CS_3
Only memory device 0 connected	0	1	1	1
Only memory device 1 connected	1	0	1	1
Only memory device 2 connected	1	1	0	1
Only memory device 3 connected	1	1	1	0
All devices disconnected	1	1	1	1

The disconnection of the data lines is performed using ***tristate outputs for the data lines of the memory chips*****.** A tristate output is *digital output with a third state added to it*.

This output can be a logic 1, a logic 0, or a third state that acts as a high impedance or open circuit.

It is like someone opened a switch and nothing is connected.

This third state is controlled by the chip select. When the active low chip select equals 1, data lines are set to high impedance, sometimes called the ***Z state.*** A chip select equal to 0 causes the data lines to be active and allow input or output

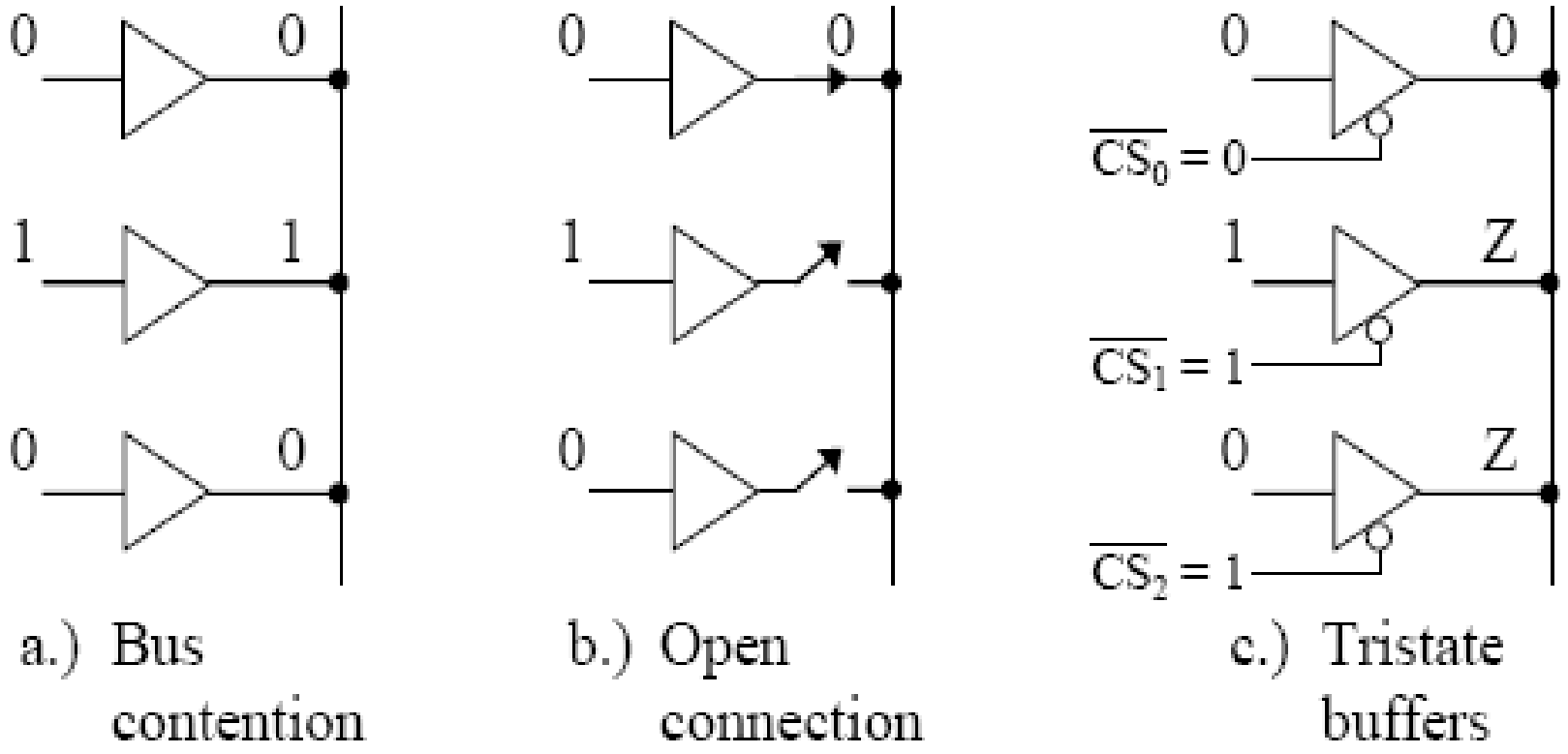


Figure 12-5 Three Buffers Trying to Drive the Same Output

Memory Terminology

There are many different purposes for memory in the operation of a computer. Some memory is meant to store data and programs only while the computer is turned on while other memory is meant to be permanent.

Some memory contains application code while other memory is meant to store the low-level driver code to control devices such as an IDE interface or a video card. Some memory may have a larger capacity while other memory may be faster

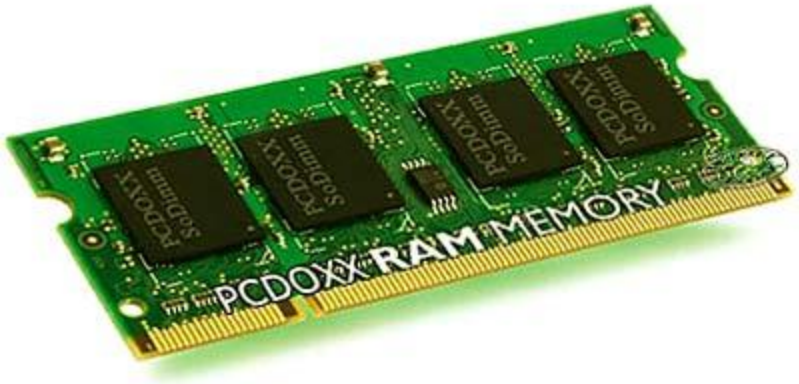
In order to understand what memory technologies to apply to which processor operation, we need to understand a little bit more about the technologies themselves. This section discusses some of the terminology used to describe memory.

Main "primary" storage

Random Access Memory (RAM):

The primary storage is referred to as random access memory (RAM) because it is possible to randomly select and use any location of the memory directly store and retrieve data. It takes same time to any address of the memory as the first address. It is also called read/write memory.

- The storage of data and instructions inside the primary storage is temporary. It disappears from RAM as soon as the power to the computer is **switched off**. The memories, which loose their content on failure of power supply, are known as **volatile** memories .So now we can say that RAM is volatile memor



Types of RAM

The following are some common types of RAM:

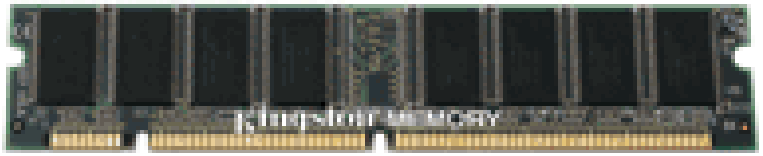
- **SRAM**: ***Static random access memory*** uses multiple transistors, typically four to six, for each memory cell but doesn't have a capacitor in each cell. It is used primarily for cache.

- **DRAM**: ***Dynamic random access memory*** has memory cells with a paired **transistor** and **capacitor** requiring "constant" periodic charge refreshing, to maintain data storage.

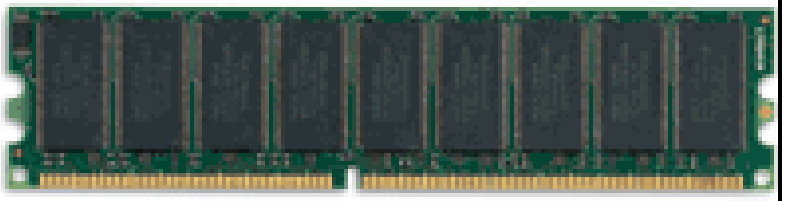
- **FPM DRAM**: ***Fast page mode dynamic random access memory*** was the original form of DRAM. It waits through the entire process of locating a bit of data by column and row and then reading the bit before it starts on the next bit. Maximum transfer rate to L2 cache is approximately 176 Mbps.

- **SDRAM**: ***Synchronous dynamic random access memory*** takes advantage of the burst mode concept to greatly improve performance. It does this by staying on the row containing the requested bit and moving rapidly through the columns, reading each bit as it goes. The idea is that most of the time the data needed by the CPU will be in sequence. SDRAM is the most common form in desktops today. Maximum transfer rate to L2 cache is approximately 528 Mbps.

- **DDR SDRAM**: ***Double data rate synchronous dynamic RAM*** is just like SDRAM except that is has higher bandwidth, meaning greater speed. Maximum transfer rate to L2 cache is approximately 1,056 Mbps.



SDRAM



DDR

RAMBUS

Memory Modules

DIMM dual in-line memory module

SIMM single in-line memory module

SODIMM small outline dual in-line memory module

From the top: SIMM, DIMM and SODIMM memory modules

Read Only Memory (ROM):

There is another memory in computer, which is called Read Only Memory (ROM). Again it is the ICs inside the PC that form the ROM. The storage of program and data in the ROM is *permanent*. The ROM stores some standard processing programs supplied by the manufacturers to operate the personal computer. The ROM can only be read by the CPU but it cannot be changed.

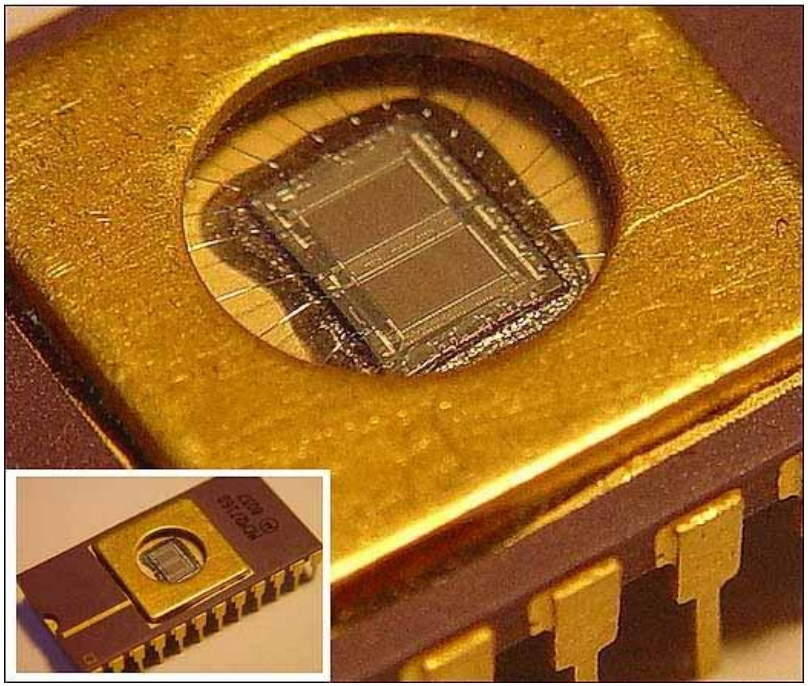
- The basic input/output program is stored in the **ROM** that examines and initializes various equipment attached to the PC when the switch is made ON. The memories, which do not loose their content on failure of power supply, are known as **non-volatile** memories. ROM is non-volatile memory.

- **PROM**

There is another type of primary memory in computer, which is called *Programmable Read Only Memory* (PROM). You know that it is **not possible** to *modify or erase programs* stored in ROM, but it is possible for you to store your program in PROM chip. Once the programs are written it **cannot be changed** and remain intact even if power is switched off. Therefore programs or instructions written in PROM or ROM cannot be erased or changed.

- **EPROM**

This stands for *Erasable Programmable Read Only Memory*, which over come the problem of PROM & ROM. EPROM chip can be programmed time and again by erasing the information stored earlier in it. Information stored in EPROM exposing the chip for some time ultraviolet light and it erases chip is reprogrammed using a special programming facility. When the EPROM is in use information can only be read.



EPROM

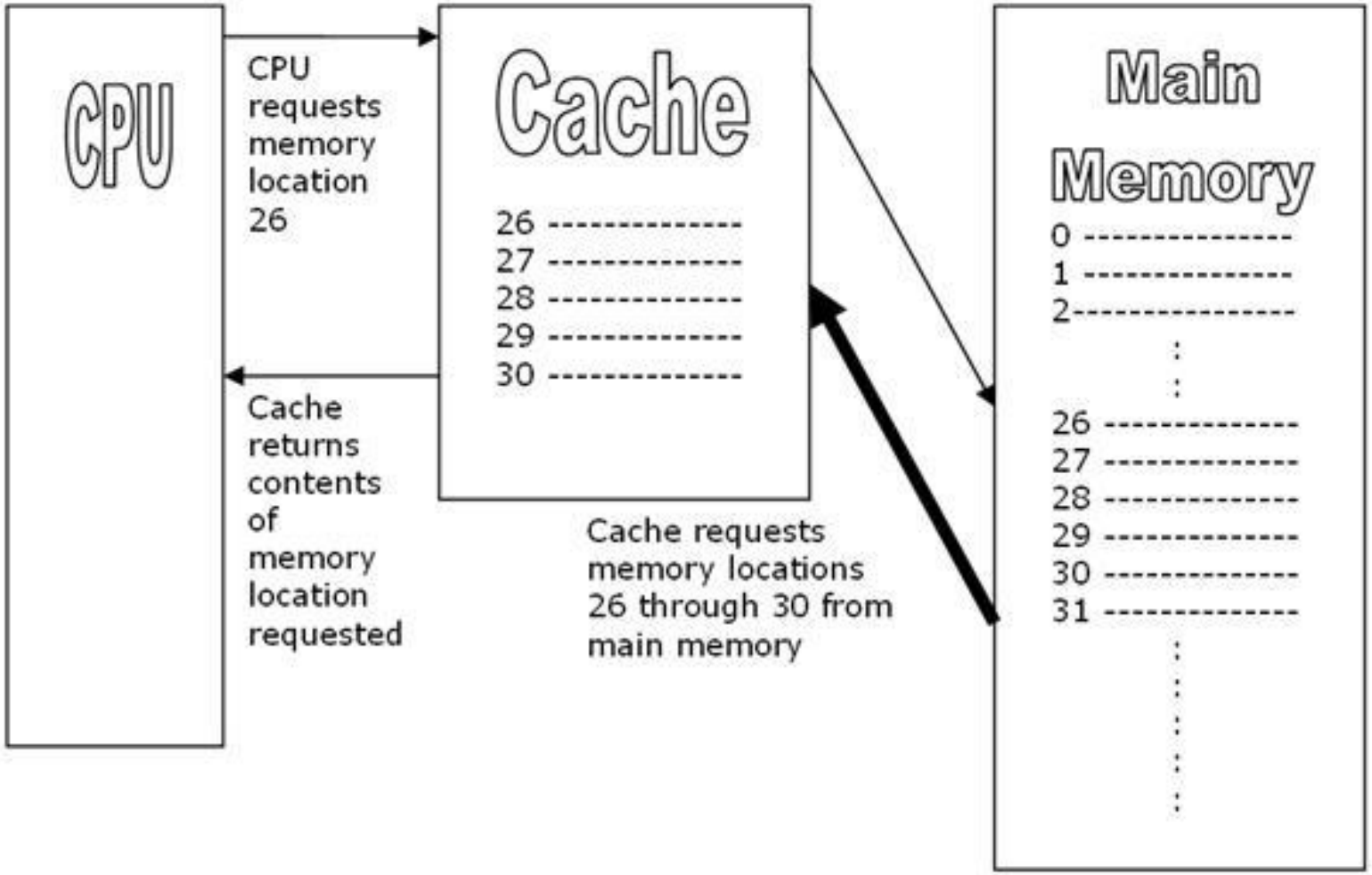
UV pacifier sterilizer erases chips

•EEPROM

EEPROM stands for *electronically erasable programmable read only memory*. In this memory, user can erase and write instructions with the help of *electrical pulses*. If there is any error in writing the instructions, the user can erase the contents electronically. The contents of EEPROM can be modified easily.

Cache Memory

The speed of CPU is extremely high compared to the access time of main memory. Therefore the performance of CPU decreases due to the slow speed of main memory. To decrease the mismatch in operating speed, a small memory chip is attached between CPU and Main memory whose access time is very close to the processing speed of CPU. It is called CACHE memory.



cache memory is built into the CPU

- **CACHE** memories are accessed *much faster* than conventional RAM. It is used to store programs or data *currently* being executed or *temporary* data frequently used by the CPU. So each memory makes main memory to be faster and larger than it really is. It is also very expensive to have bigger size of cache memory and its size is normally kept small comparing with RAM.

Registers

The CPU processes data and instructions with high speed, there is also movement of data between various units of computer. It is necessary to transfer the processed data with high speed. So the computer uses a number of special memory units called ***registers***. They are not part of the main memory but they store data or information **temporarily** and pass it on as directed by the control unit.

Secondary Storage

You are now clear that the operating speed of primary memory or main memory should be as fast as possible to cope up with the CPU speed. These high-speed storage devices are very expensive and hence the cost per bit of storage is also very high.

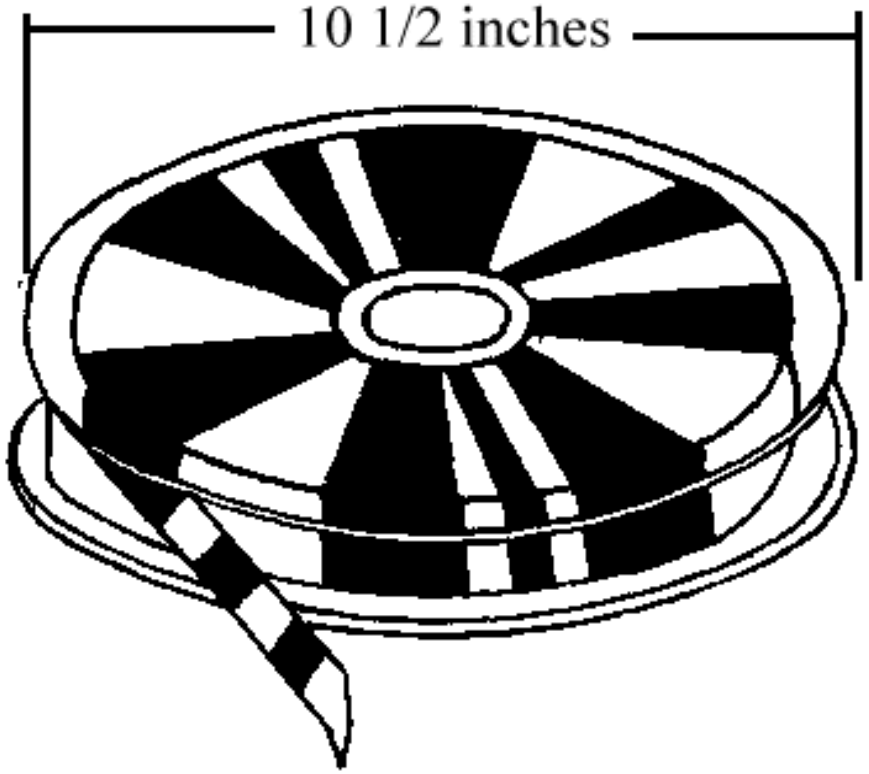
Again the storage capacity of the main memory is also very limited. Often it is necessary to store hundreds of millions of bytes of data for the CPU to process. Therefore additional memory is required in all the computer systems. This memory is called *auxiliary memory* or *secondary storage*.

In this type of memory the cost per bit of storage is low. However, the operating speed is slower than that of the primary storage. Huge volume of data are stored here on **permanent** basis and transferred to the primary storage as and when required. Most widely used secondary storage devices are *magnetic tapes*, *magnetic disk* and *Optical Disk*.

➢Magnetic Tape

Magnetic tapes are used for large computers like main-frame computers where large volume of data is stored for a longer time. In PC also you can use tapes in the form of cassettes. The cost of storing data in tapes is inexpensive.

Tapes consist of magnetic materials that store data permanently. It can be 12.5 mm to 25 mm wide plastic film-type and 500 meter to 1200 meter long which is coated with magnetic material. The deck is connected to the central processor and information is fed into or read from the tape through the processor. It similar to cassette tape recorder.



Magnetic Tape

Advantages of Magnetic Tape:

- **Compact:** A 10-inch diameter reel of tape is 2400 feet long and is able to hold 800, 1600 or 6250 characters in each inch of its length. The maximum capacity of such tape is 180 million characters. Thus data are stored much more compactly on tape.

- **Economical**: The cost of storing characters is very less as compared to other storage devices.
- **Fast**: Copying of data is easier and fast.
- **Long term Storage and Reusability**: Magnetic tapes can be used for long term storage and a tape can be used repeatedly with out loss of data.

➢Magnetic Disk

You might have seen the gramophone record, which is circular like a disk and coated with magnetic material. Magnetic disks used in computer are made on the same principle. It rotates with very high speed inside the computer drive. Data is stored on both the surface of the disk. Magnetic disks are most popular for *direct access* storage device. Each disk consists of a number of invisible *concentric circles* called *tracks*.

Gramophone record

Gramophone record disk

Information is recorded on tracks of a disk surface in the form of tiny magnetic spots. The presence of a magnetic spot represents *one bit* and its absence represents zero bit. The information stored in a disk can be read many times without affecting the stored data. So the reading operation is non-destructive. But if you want to write a new data, then the existing data is erased from the disk and new data is recorded.

➢ Floppy Disk

It is similar to magnetic disk discussed above. They are 5.25 inch or 3.5 inch in diameter. They come in single or double density and recorded on one or both surface of the diskette. The capacity of a 5.25-inch floppy is 1.2 mega bytes whereas for 3.5 inch floppy it is 1.44 mega bytes. It is cheaper than any other storage devices and is portable. The floppy is a low cost device particularly suit-able for personal computer system.

Floppy Disk

➢Optical Disk

With every new application and software there is greater demand for memory capac-ity. It is the necessity to store large volume of data that has led to the development of optical disk storage medium. Optical disks can be divided into the following categories:

Compact Disk/ Read Only Memory (CD-ROM): CD-ROM disks are made of reflective metals. CD-ROM is written during the process of manufacturing by high power *laser beam.*

Here the storage density is very high, storage cost is very low and access time is relatively fast. Each disk is approximately 4.75 inches in diameter and can hold over 600 MB of data. As the CD-ROM can be *read only* we cannot write or make changes into the data contained in it.

- ***Erasable Optical Disk***

These are optical disks where data can be written, erased and re-written. This also applies a laser beam to write and re-write the data. These disks may be used as alternatives to traditional disks.

Erasable optical disks are based on a technology known as *magnetic optical* (MO). To write a data bit on to the erasable optical disk the MO drive's laser beam heats a tiny, precisely defined point on the disk's surface and magnetizes it.

Storage	Speed	Capacity	Relative Cost ($)	Permanent ?
Registers	Fastest	Lowest	Highest	No
RAM	**Very Fast**	**Low/Moderate**	**High**	**No**
Floppy Disk	Very Slow	Low	Very Low	Yes
Hard Disk	Moderate	Very High	Low	Yes

The benefits of secondary storage can be summarized as follows:

- **Capacity.** Organizations may store the equivalent of a roomful of data on sets of disks that take up less space than a breadbox. A simple diskette for a personal computer holds the equivalent of 500 printed pages, or one book. An optical disk can hold the equivalent of approximately 400 books

❖**Reliability.** Data in secondary storage is basically safe, since secondary storage is physically reliable. Also, it is more difficult for unscrupulous people to tamper with data on disk than data stored on paper in a file cabinet.

❖**Convenience.** With the help of a computer, authorized people can locate and access data quickly.

❖**Cost.** Together the three previous benefits indicate significant savings in storage costs. It is less expensive to store data on tape or disk (the principal means of secondary storage) than to buy and house filing cabinets. Data that is reliable and safe is less expensive to maintain than data subject to errors. But the greatest savings can be found in the speed and convenience of filing and retrieving data.