CHAPTER 4

COMPUTER MEMORY

4.0 Computer memory

Computer memory can be defined as the long-term or short-term change in one or more of the physical properties of matter caused by some event. That is after the event has taken place, the change remains. **Computer memory** – refers to the interrelated set of devices served for storage and issuing of sequences of software instructions (i.e. programs) and transient information (data) used during processing. Any member of this interrelated set is called the memory element or device. To create a computer memory, we need to find some property of matter that can be modified and at a later time, the modification be detected. Computer memories are organized into locations. Each location consists of the same number of cells. A memory cell can store only one of two different values: the binary digits of 0 or 1 – called a bit. Hence, the quantity contained in a location, whether this is an instruction or a piece of data, is in binary form.

The same is true about information that flows over buses or information stored in other parts of a computer. Usually, a memory location stores one byte (or one character) that is, a group of eight bits. Both human and computer memories share a common property known as forgetfulness. Human memory may gradually fade unless it is refreshed by repetition or a reminder. A type of computer memory called dynamic memory also forgets its stored data unless it is periodically refreshed.

Each location of the memory has a unique address which itself is a binary number. Figure 4.1 shows the functional units of a computer memory.

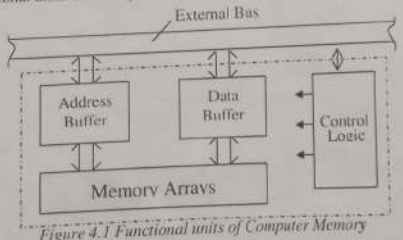

Figure 4.1 Functional units of Computer Memory

Like the CPU, the memory includes an address buffer and a data buffer. Buffers – (or registers) are high-speed memory locations used to store important (temporary) information during CPU operations. The control circuit receives read/write signals sent by the CPU or I/O over the external bus. On the basis of these inputs, the control circuit generates other signals

which control internal memory operations. The basic memory operations are **memory read**and **memory write**. Both operations are known as memory access – operations and are usually executed by the CPU.

Computer memory can be divided into primary (main) and secondary (or auxiliary or external) memory.

4.1 Primary Memory

Typical examples of primary memories are Random Access Memory (RAM) and Read Only Memory (ROM).

RAM: Random Access Memory is also called Read/Write memory. In RAM, the access time (t_a) of a memory word is independent of which location was accessed previously that is, locations can be accessed randomly and independently of each other. Access time is therefore the same for all locations in contrast to other types of memories such as sequential ones in which the time needed to access a location depends on its distance from the currently accessed location. It is for this reason that semiconductor memory devices are called random access memories.

RAM makes up the memory in which user-written or externally loaded programs and data are stored. In other words, the RAM is the program (user) work space.

Because of its speed and consistency, RAM is used as 'main memory' or primary storage. In most personal computers the RAM is not an integral part of the motherboard or CPU. It comes in the easily upgradeable form of modules called *memory sticks or RAM sticks*, which is about the size of a few sticks of chewing gum. These can quickly be removed and replaced should they become damaged or too small for current purpose. A smaller amount of random-access memory is also integrated with the CPU, but this is usually referred to as "cache" memory, rather than RAM. Figures 4.2a and 4.2b show DRAM and SRAM sticks.

Figure 4.2a: 512 MB DRAM Modules

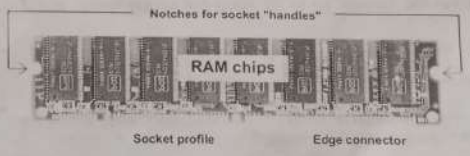

Figure 4.2b: Typical layout of the SD RAM module.

Modern RAM generally stores a bit of data as either a charge in a capacitor, as in dynamic RAM (DRAM), or the state of a flip-flop, as in static RAM (SRAM). Some types of RAM

can detect or correct random faults called memory errors in the stored data by using RAM parity and error correction codes.
Many types of RAM are volatile, which means that unlike some other forms of computer storage such as disk storage and tape storage, they lose all data when power is switched off. For these reasons, nearly all PC's use disks as "secondary storage". Small PDAs (personal digital assistances) and music players may dispense with disks, but rely on flash memory to maintain data between sessions of use.

RAM Disk
Software can "partition" a portion of a computer's RAM, allowing it to act as a much faster hard drive that is called a RAM disk. Unless the memory used is non-volatile, a RAM disk loses the stored data when the computer is shut down. However, volatile memory can retain its data when the computer is shut down if it has a separate backup power source, usually a battery.
If a computer's speed becomes slow on RAM during intensive application cycles, the computer can resort to so-called **virtual memory**. In this case, computer temporarily uses hard drive space as additional memory. Constantly relying on this type of backup memory is called **thrashing**, which is generally undesirable, as virtual memory lacks the advantages of RAM. In order to reduce the dependency on virtual memory more RAM can be installed.
RAM is usually writable as well as readable, so "RAM" is often used interchangeably with "read-write memory". The alternative to this is "ROM", or Read Only Memory.

Flash Memory
Flash memory is a ROM/RAM hybrid that can be written to, but which does not require power to maintain its contents. RAM is not strictly the opposite of ROM, however. The word random indicates a contrast with serial access or sequential access memory.

Many CPU-based designs actually have a memory hierarchy consisting of registers, on-die SRAM caches, DRAM, paging systems, and virtual memory of swap space on a hard-drive. This entire pool of memory may be referred to as "RAM" by many developers, even though the various subsystems can have very different access times, violating the original concept behind the "random access" term in RAM. Even within a hierarchy level such as DRAM, the specific row/column/bank/rank/channel/interleave organizations of the components make the access time variable, although not to the extent that rotating storage media or a tape is variable.

4.2 Types of RAM

Figure 4.3 shows both different types of PC RAM modules, while table 4.1shows various types of RAM in the market at the moment.

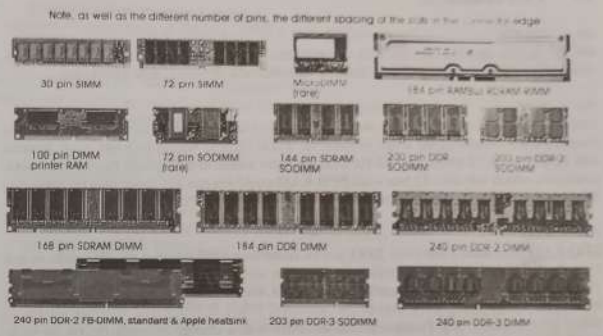

Figure 4.3: PC RAM types

Table 4.1 RAM types

RAM type	Pins	Width	Usage
SD RAM	168	64 bit	Older and slower type. No use.
Rambus RAM	184	16 bit	Advanced RAM. Only used for very few Pentium 4's with certain Intel chipsets.
DDR RAM	184	64 bit	A faster version of SD RAM. Used both for Athlon and Pentium 4's. 2,5 Volt.
DDR2 RAM	240	64 bit	New version of DDR RAM with higher clock frequencies. 1,8 Volt.

SDRAM is an old and proven type, which is used in the majority of existing PC's. DDR RAM is a refinement of SDRAM, which is in reality double clocked. Rambus RAM is an advanced technology which in principle is superior to DDR RAM in many ways. However, Rambus has had a difficult birth. The technology has been patented by Rambus Inc., which has been involved in many legal suits. A number of important manufacturers (such as VIA) have opted out of Rambus, and only develop products which use DDR RAM. With the new DDR2 standard, there is no obvious need for Rambus RAM.

4.3 RAM Technologies

(a) DRAM

DRAM was used in the 1980's. This was dynamic RAM, which was relatively slow. It was replaced by FPM (*Fast Page Mode*) RAM which was also dynamic, only a bit faster.

(b) FPM

Originally, loose RAM chips were installed directly in large *banks* on the motherboard. Later people started combining the chips in modules. These came in widths of 8 bits (with 30 pins) and 32 bits (with 72 pins). The 32-bit modules were suited to the system bus for the 80486 processor, which was also 32 bits wide.FPM RAM could not run any faster than 66 MHz, but that was fine for the system bus clock frequency in the original Pentium processors.

(c) EDO

EDO RAM (*Extended Data Out*). EDO is a bit faster than FPM because the data paths to and from the RAM cells have been optimized. The gain was a 3-5 % improvement in bandwidth. The clock frequency could be increased to 75 MHz, but basically, EDO is not very different from FPM RAM.

When Intel launched the Pentium processor, there was a change in using the 64 bit wide RAM modules (with 168 pins, as in SDRAM).

4.4 RAM Bandwidth

We can start by looking at the theoretical maximum bandwidth for the various systems. This is easy to calculate by *multiplying the clock frequency by the bus width*. Table 4.2 shows the RAM bandwidth.

Table 4.2: RAM bandwidth.

Module type	Max. transfer.
SD RAM, PC100	800 Mb/sec
SD RAM, PC133	1064 MB/sec
Rambus, PC800	1600 MB/sec
Rambus, Dual PC800	3200 MB/sec
DDR 266 (PC2100)	2128 MB/sec
DDR 333 (PC2700)	2664 MB/sec
DDR 400 (PC3200)	3200 MB/sec
DUAL DDR PC3200	6400 MB/sec
DUAL DDR2-400	8600 MB/sec
DUAL DDR2-533	10600 MB/sec