UNIT IV

5. THE MEMORY SYSTEM

5.1 Some Basic Concepts

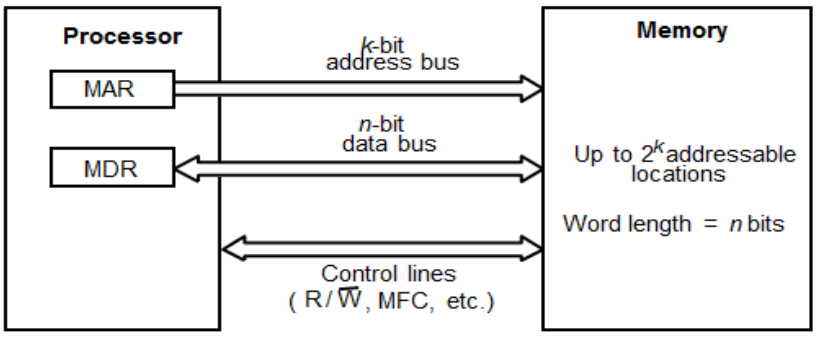
5.1.1 Size of the memory

Maximum size of the memory that can be used is dependent on the number of address lines that is present in the system. For example a computer with 16 bit address lines is capable of addressing 2^{16} locations which is equal to 64K locations. In general a computer with n bit address lines is capable of addressing 2^n locations of the memory.

5.1.2 Word Length

Number of bits present in a word is called as **word length.** Data from the memory system is accessed based on the word length. As an example a system with word length of 32 bit can access 32 bits of data on a single access.

5.1.3 Connection between memory and processor



One of the major functionality for the instruction execution is the communication between the memory unit and processor. Communication between memory unit and the processor is done through processor bus.

Processor bus consists of three different lines:

1. Address lines: Set of wires responsible for carrying the address information to be accessed from the memory unit.
2. Data lines: Set of wires that carry data information between the processor and the memory unit.
3. Control lines: Set of wires that carry the control signals from the processor required to control the memory unit.

When doing the communication, address is loaded every time into the register MAR and data to be sent or received from the memory unit is loaded into the register MDR. While doing the read operation, R/W signal is set to 1 and while doing the write operation, it is set to 0.

5.1.4 Memory access time

It is defined as the time between the initiation of the operation and the completion of the operation.

5.1.5 Memory Cycle Time

It is defined as the minimum time delay required between the initiation of two successive operations.

5.1.6 RAM (Random Access Memory)

It is the memory units were any location for read/write operation can be accessed within a fixed amount of time independent of the location address.

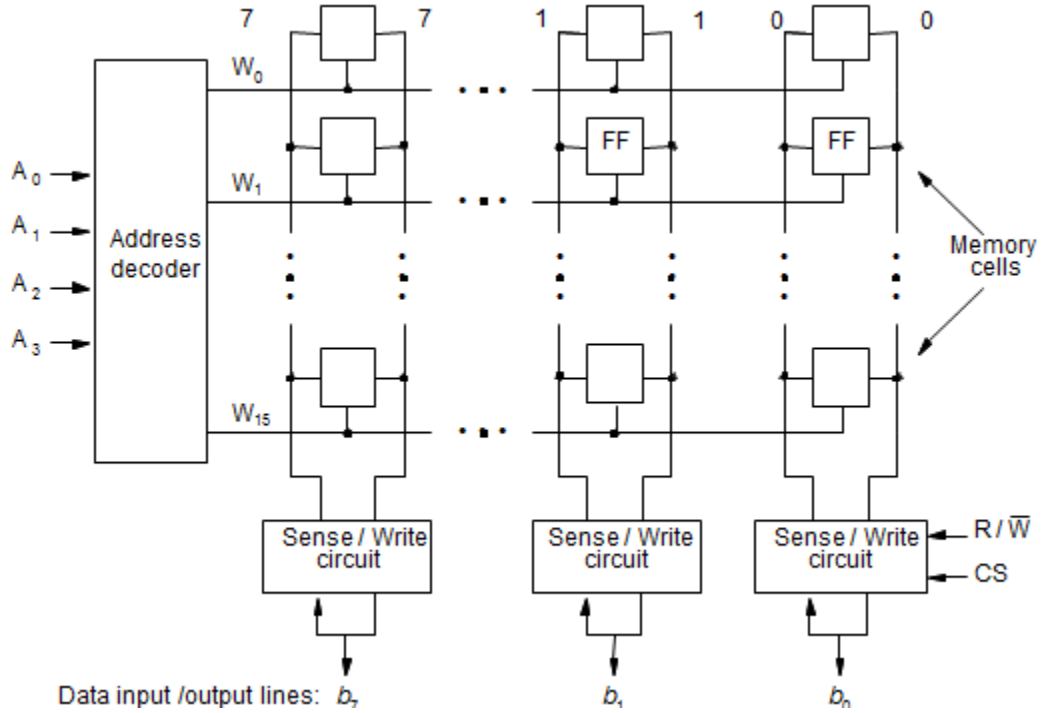
5.1.7 Cache Memory

It is the faster version of the RAM that is present on the processor chip.

5.2 Semiconductor RAM Memories

Semiconductor RAM memories are available in a very wide range of speeds. The cycle time for those memories range from 10ns to 100ns. First introduces around late 1960's and proved to be very expensive. Because of the invention of VLSI technology their cost dropped very dramatically.

5.2.1 Internal Organization of Memory Chips



→ Memory cells of the semiconductor RAMs are organized in the form of a 2D array were each cell has the capacity of storing 1bit of data information.

→ Above shown figure is the organization of 16X8 memory cells were in there are 16 rows and 8 columns with each row having 8 memory cells arranged in a column structure. The organization can store 16 words with each word has the word length of 8 bits

→ When accessed, the organization accesses one complete row with all 8 cells arranged across it.

→ All the cells arranged across the row are connected to one common line called as **word line** which activates all the bits of the row when a particular row is accessed at a given time. For the circuit shown above, there are 16 word lines.

→ Word lines are controlled by address decoder which selects the required word from the address ranging between 0 to 15. First word line is given the address 0 and last word is given the address 15. To select required word, there is a 4 bit address input lines which ranges from 0000 to 1111. According to the input value, one out of 15 lines is activated at a given time.

→ Cells arranged across the column are again connected to a sense/write circuit in common by two bit lines. There are 8 columns and altogether there 8 sense/write circuits connected in common.

→ Each sense/write circuit has one data input/output line connected across them which is a bidirectional line connected to a data bus and has one common control line R/W to select read or write operation and one common line CS which selects a particular chip.

→ Read operation: To do a read operation, a particular word line is activated by the 4 bit address and R/W is set to 1. All the cells in a specified row is activated and data from the row is sensed from each of the cell.

→ Write operation: To do a write operation, a particular word line is activated by the 4 bit address and R/W is set to 0 and the required data value is loaded on the data lines. All the cells in a specified row is activated and data from the data line is written into the cell.

→ Thus the above shown circuit has:

4 address lines

8 data lines

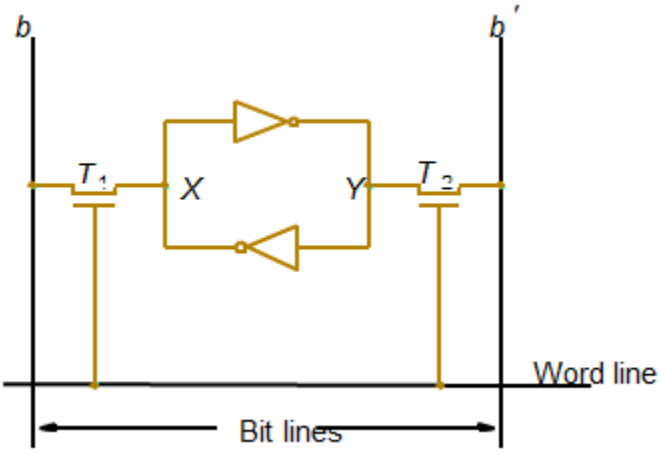
2 control lines

5.2.2 Static Memories

A memory that has the circuits which are capable of retaining their state as long as power is applied is called as static memories.

Even if there is a small interruption to continues power, it will wipe out all the data. That is why it is called as **volatile memory**. Type of cells used in this are called as static RAM cells

Organization of SRAM cell



→ To implement each RAM cell, two transistor inverters are cross connected to implement a basic flip-flop (latch).

→ This latch is connected to transistors T1 and T2 which acts as switch and these transistor lines are in turn connected to word line and a bit line b and b$^{|}$. Bit line b and b$^{|}$ are connected to each of the sense/write circuit. When word line is at ground level, the transistors are turned off and the latch retains its state.

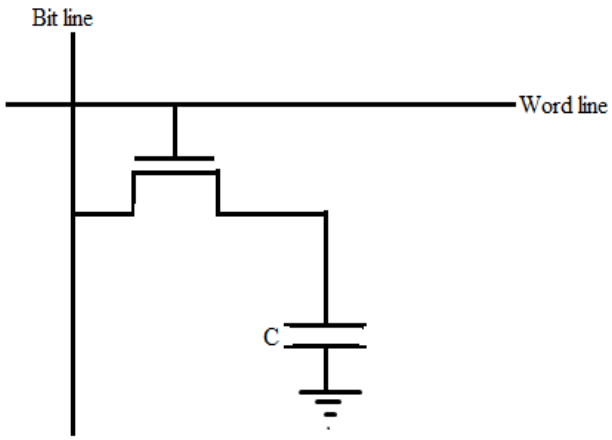
→ **Read operation:** Read operation is performed by activating a particular word line to close the switches T1 and T2. Sense/Write circuits at the bottom monitor the state of b and b$^{|}$.

→ **Write Operation:** It is performed by placing the required on the line b and its compliment on line b$^{|}$ and particular word line is activated. Data from the sense/write circuit is stored inside the cell.

5.2.3 DRAMS (Dynamic RAM)

Static RAMs are fast but the major disadvantage of them is that they come at a very high cost because of too much involvement of the transistors in a single cell. It can be made to implement in a more simple way using another type of cells called as **Dynamic RAM** cells.

The cells are called as dynamic RAMs because they do not have the capacity to retain the state indefinitely for a long period of time when compared to static RAMs. Below figure shows the organization of DRAM cell.



Here the capacitor is used store the data in the form of a charge. The capacitor is connected to transistor and the transistor is connected to word line and the bit line. When word line is activated the switch is closed and required operation is done.

In the DRAM cells, the capacitor gradually loses the charge with respect to time when switch is opened. So to retain the state, it is very much necessary to restore the charge periodically. The method of restoring the charge periodically in the DRAM cell is called as **Refresh Process.**

Refresh Process: After T is turned off, capacitor starts to discharge the charge. So the required cell is read time to time by closing T and applying the required voltage to retain the bit value. When closed, sense amplifier senses if the voltage of the cell is above or below the thresh hold voltage value. If above applies the charge and sets 1. If below pulls down to the ground voltage.

There are two types of DRAMs:

1. Asynchronous DRAM
2. Synchronous DRAM

We will discuss each one of them in detail in the next sections.

5.2.4 Asynchronous DRAMs

One of the types of DRAM chips were the timing signals of the RAM circuit are controlled asynchronously. I.e. there is no centralized timing signal and the circuit derives timing signal by its own while doing any kind of operation.

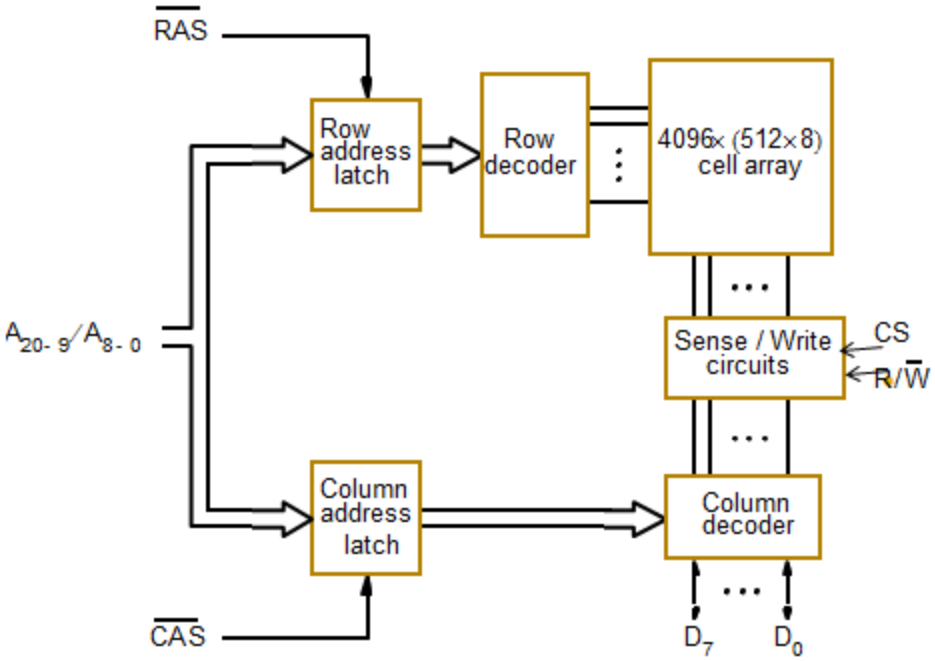
We will look into the detailed organization of asynchronous DRAM with a 16 MB DRAM. A 16MB DRAM is arranged as 4KX4K cells were in there are 4096 rows and each row has 4096 cells arranged in column wise fashion.

4096 cells arranged in each row are again divided into group of 8 cells forming each word. Thus 4096 cells are divided into 512 groups with each group is a word having 8 cells. Arrangement of the cell is shown below:

4KX4K = 4096X4096/8 = 4096X512X8

According to the above equation, it means there are 4096 rows and there are 512 columns and each column has 8 cells in it which forms a single word. When accessed one out of 512 columns is accessed.

Therefore this configuration is called as 16MB DRAM chip configured as 2MX8 which means there are 2M different locations arranged in column wise. Circuit arrangement for asynchronous DRAM is shown below:



In the above circuit there are 4096 rows and each row there are 512 groups having 8 bits each which means each row has altogether 512 bytes of data. Therefor to access a data it requires both row address and column address information. There are 4096 rows and to access row it requires 12 bits of information and there are 512 columns and to access a column it requires 9 bits of information. Altogether to access a particular byte in a column, it requires 21 bits of address information. The higher order 12 bits are row address and low order 9 bits are column address. For a single access, it retrieves 8 bits of information. So there must be 8 data lines to access data from or to any location.

Following are the different units of the circuit and their functionality:

1. Row Address latch: This latch is used to strobe the row address to be accessed. The latch is controlled via signal RAS (Row Address Strobe) which is active low and when activated sends the row address to be accessed.

2. Column Address latch: This latch is used to strobe the column address to be accessed. The latch is controlled via signal CAS (Column Address Strobe) which is active low and when activated sends the column address to be accessed.

3. Row decoder: This unit receives the output of Row Address latch as its input and takes the responsibility of selecting one row out of 4096 rows.

4. Column decoder: Row decoder: This unit receives the output of Column Address latch as its input and takes the responsibility of selecting one column out of 512 columns. It also has bidirectional data lines which retrieves or sends the data information into a particular column.

5. Sense/Write Circuits: These are the circuits connected to columns of the DRAM chip in common. There are altogether 512 sense/write circuits which takes the responsibility of reading/ writing operation.

Following are the important operations performed by the asynchronous DRAM chip:

1. Address multiplexing: One of the important features implemented in this DRAM chip. We know that 21 bits of address information are required to access any of the location. But at the same it is even known that Row address and Column address are accessed separately one after the other. So to reduce the number of pins row and column address are multiplexed into same pins and the number of pins is chosen to be the highest valued information. In this case row address is 12 bits and column address is 9 bits. So column address is highest and number of pins is 12.

2. Read/Write Operation:

→ During Read/Write operation, row address is applied first and is loaded onto the row address latch with signal $\overline{\mathbf{RAS}}$. A complete row is selected

→ Next shorty after this, column address is loaded into column address latch with the signal $\overline{CAS}$.

→ Row address and column address are decoded and appropriate Read/Write signal is activated in the sense/Write ckt. If it is read operation data flows from the cell into the line and if write, it flows from the line into the cell.

3. Refresh Process: Required cells are read time to time to ensure the retaining of the data. Each and every row is read time to time and charge on them is ensured. This functionality in the circuit is controlled by a special circuit called memory controller circuit.