

Unit 6 : Memory Organization

Lesson 1 : Memory Terminology

1.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ understand and correctly use the terminology associated with memory systems.

1.2. Introduction

A major advantage of digital systems over analog systems is the ability to easily store large quantities of digital information and data for short or long periods. This memory capability is what makes digital systems so versatile and adaptable to many situations.

We have already become very familiar with the flip-flop, which is an electronic memory device. We have also seen how groups of FFs called registers can be used to store information. FF registers are high-speed memory elements which are used extensively. Advances in Large scale Integration (LSI) and Very Large Scale Integration (VLSI) technology have made it possible to obtain large numbers of FFs on a single chip arranged in various memory-array formats. These bipolar and Metal oxide semiconductor (MOS) semiconductor memories are the fastest memory devices available, and their cost has been continuously decreasing as LSI technology improves.

Semiconductor memories are used as the internal memory of a computer where fast operation is important.

Semiconductor memories are used as the internal memory of a computer where fast operation is important. A computer's internal memory is also called its main memory or working memory. A program and any data used by program reside in the internal memory while the computer works on that program. RAM and ROM (to be defined shortly) make up internal memory.

Another form of storage in a computer is performed by auxiliary memory which is separate from the internal working memory. Auxiliary memory is also called mass storage has the capacity to retain massive amounts of data without the need for electrical power. Auxiliary memory operates at a much slower speed than internal memory. The information is transferred to the internal memory when the

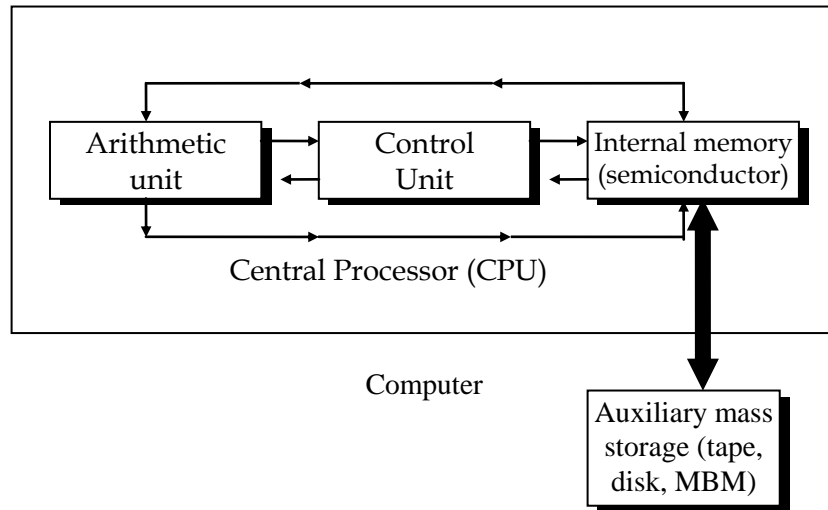


Fig. 6.1 : A computer system with high-speed internal memory and slower external memory.

computer needs them. Common auxiliary memory devices are magnetic disk, magnetic tape, and magnetic bubble memory (MBM).

1.3. Memory Terminology

Memory Cell : A device or electrical circuit used to store a single bit (0 or 1). Example of memory cells are a flip-flop, a charged capacitor, and a single spot on magnetic or disk.

Memory terminology.

Memory Word : A group of bits (cells) in a memory that represents instructions or data of some type. For example, a register consisting of eight FFs can be considered to be a memory that is storing an 8-bit word.

Byte : A special term used for a group of 8-bits. A byte always consists of 8-bits. Word sizes can be expressed in byte as well as in bits. For example, a word size of 8-bits is also a word of one byte; a word size of 16-bits is two bytes word, and so on.

Capacity : A way of specifying how many bits can be stored in a particular memory device or complete memory system. To illustrate, suppose that memory which can store 4096 twenty-bit words. This represents a total capacity of 81,920 bits. We could also express this memory's capacity as 4096 X 20. When expressed this way, the first number (4096) is the number of words and the second number (20) is the number of bits per word. Thus, a memory that has a storage capacity of 4K X 20 is actually a 4096 X 20 memory. The development of larger memories has brought about the designation "1M" or "1 Mega" to

Memory Organization

represent $2^{20} = 1,048,576$. Thus, a memory that has a capacity of $2M \times 8$ is actually the one with a capacity of $2,097,152 \times 8$.

Problem 1

A certain semiconductor memory chip is specified as $4K \times 8$. How many words can be stored on this chip? What is the word size? How many total bits can this chip store?

$$4K = 4 \times 1024 = 4096 \text{ words}$$

Each word is 8 bits (one byte). The total number of bits is therefore $4096 \times 8 = 32,768$ bits

Problem 2

Which of the memories store more bits : a $5M \times 8$ memory, or a memory that stores $1M$ words of word size of 16 bits?

Solution

$$5M \times 8 = 5 \times 1,048,576 \times 8 = 41,943,040 \text{ bits}$$

$$1M \times 16 = 1,048,576 \times 16 = 16,777,216 \text{ bits}$$

The $5M \times 8$ memory stores more bits.

5M X 8 memory stores more bits.

Density : Another term for capacity. When we say that one memory device has a greater density than another, we mean that memory of greater densities it can store more bits in the same amount of space. It is more dense.

Address : A number that identifies the location of a word in memory. Each word stored in a memory device or system has a unique address. Addresses are always expressed as a binary number, although octal, hexadecimal, and decimal numbers are often used for convenience.

Read Operation : The operation whereby the binary word stored in a specific memory location (address) is sensed and then transferred to another device. For example, if we want to use word at an address of the memory for some purpose, we have to perform a read operation on address 100. The read operation is often called a fetch operation, since a word is being fetched from memory.

Write operation.

Write Operation : The operation whereby a new word is placed into a particular memory location. It is also referred to as store

Addresses	
000	Word 0
001	Word 1
010	Word 2
011	Word 3
100	Word 4
101	Word 5
110	Word 6
111	Word 7

Fig. 6.2 : Each word location as a specific binary address.

operation. Whenever a new word is written into a memory location, it replaces the word that was previously stored there.

Access Time : A measure of a memory device's operating speed. It is the amount of time required to perform a read operation. More specifically, it is the time between the memory receiving a new address input, and the data becoming available at the memory output. The symbol T_{ACC} is used for access time.

Access Time
Volatile Memory

Volatile Memory : Memory requires electrical power to store information. If the electrical power is removed, all information stored in the memory will be lost. Many semiconductor memories are volatile, while all magnetic memories are nonvolatile, which means they can retain information without electrical power.

Random-Access Memory (RAM) : Memory in which the actual physical location of a memory word has no effect on how long it takes to read from or write into that location.

Random-Access Memory
Read/Write Memory

Read/Write Memory (RWM) : Any memory that can be read from or written into with equal ease.

Read-Only Memory (ROM) : A ROM can be written into (programmed) only once, and this operation is normally performed at the factory. Thereafter information can only be read from the memory. The various types of ROM will be discussed later. All ROM is nonvolatile and will store data when electrical power is removed.

Read-Only Memory
Static Memory Devices

Static Memory Devices : Semiconductor memory devices in which the stored data will remain permanently stored as long as power is applied, without the need for refreshing periodically and will store data when electrical power is removed.

Memory Organization

Dynamic Memory Devices : Semiconductor memory devices in which the stored data will not remain permanently stored, even with power applied, unless the data are periodically rewritten into memory. The latter operation is called a refresh operation.

Dynamic Memory Devices
Internal Memory
Auxiliary Memory

Internal Memory : Also referred to as the computer's main or working memory. It stores instructions and data the CPU is currently working on. It is the highest speed memory in the computer and is always a semiconductor memory.

Auxiliary Memory : Also referred to as mass storage because it stores massive amounts of information external to the internal memory. It is slower in speed than internal memory and is always nonvolatile. Magnetic tape and disk are common auxiliary memory devices.

1.4. Exercise

1.4.1. Questions for short answers

- a) Define the following terms : memory cell, memory word, address, byte, access time.
- b) A certain memory has a capacity of $8K \times 16$. How many bits are in each word? How many words can be stored? How many memory cells does this memory contain?

1.4.2. Analytical questions

- a) Explain the difference between the read (fetch) and write (store) operations.
- b) Explain the difference between RWM and ROM.

Lesson 2 : Memory Operation

2.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ describe the difference between read/write memory and read-only memory
- ◆ outline the steps that occur when the CPU reads from or writes to memory.

Memory is one of the five fundamental parts of a computer. Memory is the storage device of computer.

Memory is one of the five fundamental parts of a computer. Memory is the storage device of computer. It can store program and data. Although each type of memory is different in its internal operation, certain basic operating principle are the same for all memory systems.

Every memory system requires several different types of input and output lines to perform the different functions.

Fig. 6.3(a) illustrates the basic functions in a simplified diagram of a 32×4 memory that stores thirty-two 4-bit words. Since the word size is 4 bits, there are four data input lines I_0 to I_3 and four data output lines O_0 to O_3 . During a write operation the data to be stored into memory have to be applied to the data input lines. During a read operation the word being read from memory appears at the data output lines.

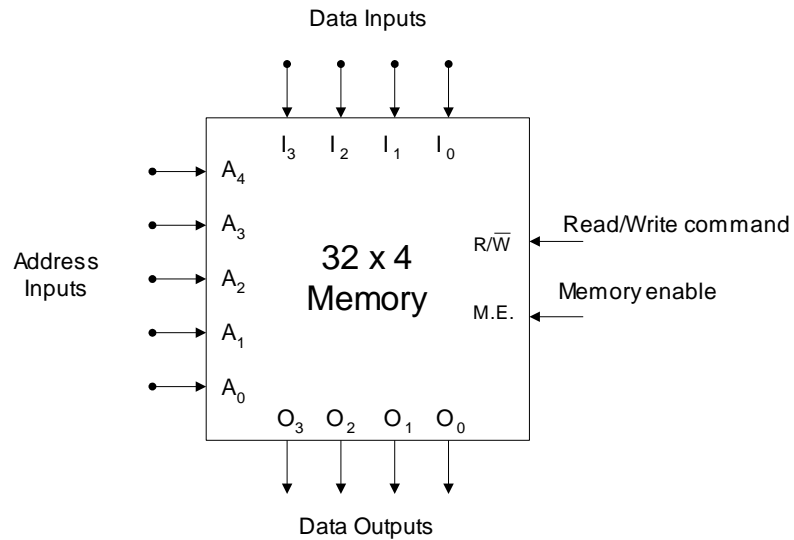
Address Inputs : Since this memory stores 32 words, it has 32 different storage locations and therefore 32 different binary addresses are ranging from 00000 to 11111 (0 to 31 in decimal). Thus there are five address inputs, A_0 to A_4 . To access one of the memory locations for a read or write operation, the 5-bit address code for that particular location is applied to the address inputs. In general, N address inputs are required for a memory that has a capacity of 2^N words.

As illustrated in Fig. 6.3(b), each address location is shown containing four memory cells that hold 1's and 0's that make up the data word stored at that location. For example, the data word 0110 is stored at address 00000, the data 1001 is stored at address 00001, and so on.

The R/ \bar{W} Input : This input controls which memory operation is to take place: read (R) or write (W). The input is labeled R/W; since there is no bar over the R, this indicates that the read operation occurs when $R/W = 1$. The bar over the W indicates that the write operation takes place when $R/\bar{W} = 0$.

Memory Organization

A simplified illustration of the read and write operations is shown in Fig. 6.3(a). Part (a) shows the data word 0100 being written into the memory register at address location 00011.



0	1	1	0	00000	0	1	1	0
1	0	0	1	00001	1	0	0	1
1	1	1	1	00010	1	1	1	1
0	1	0	0	00011	0	1	0	0
0	0	0	1	00100	0	0	0	1
0	0	0	0	00101	0	0	0	0
				⋮				
				⋮				
				⋮				
				⋮				
1	1	0	1	11101	1	1	0	1
1	1	0	1	11110	1	1	0	1
0	1	1	1	11111	0	1	1	1

(a) WRITING the data
word 0100 into memory
location 00011.

(b) READING the data
word 1101 from
memory location 11110.

Fig. 6.3 : (a) Diagram of a 32×4 memory : (b) virtual arrangement of memory cells into thirty-two 4-bit words.

Memory Enable : Many systems have some means for completely disabling all or part of the memory so that it will not respond to the other inputs. It enables the memory to operate normally when it is kept HIGH. A low on this disables the memory so that it will not respond to the address and R/W inputs. This type of input is useful when several memory modules are combined to form a larger memory.

Problem 3

Problem 3

Describe the conditions at each input and output when the contents of address location 00100 are to be read.

Solution

Address inputs : 00100
Data inputs : xxxx (not used)
R/ \bar{W} : HIGH
MEMORY ENABLE : HIGH
Data outputs : 0001

Problem 4

Describe the conditions at each input and output when the data word 1110 is to be written into address location 01101.

Solution

Solution

Address input : 01101
Data inputs : 1110
R/ \bar{W} : Low
MEMORY ENABLE : HIGH
Data outputs : xxxx (not used; usually Hi-Z)

Problem 5

A certain memory has a capacity of 4K X 8.

- (a) How many data input and data output lines does it have?
- (b) How many address lines does it have?
- (c) What is its capacity in bytes?

Solution

- (a) Eight of each, since the word size is eight.
- (b) The memory stores $4K = 4 \times 1024 = 4096$ words. Thus, there are 4096 memory addresses. Since $4096 = 2^{12}$, it requires a 12-bit address code to specify one of 4096 addresses.
- (c) A byte is 8 bits. This memory has a capacity of 4096 bytes.

Semiconductor memory, makes up the internal memory of most modern computers. This internal memory is in constant communication with the CPU (central processing unit).

A computer's internal memory is made up to RAM and ROM ICs that are interfaced to the CPU over three groups of signal or buses. These are

Memory Organization

shown in Fig. 6.4 the address lines or address bus, data lines or data bus, and the control lines or control bus. Each of these buses consists of several lines (note how they are represented by a single line with a slash), and the number of lines in each bus will vary from one computer to the next. The three buses play a necessary part in allowing the CPU to write data into memory and to read data from memory.

When a computer is executing a program of instructions, the CPU continually fetches (reads) information from those locations in memory that contain (1) the program codes representing the operations to be performed and (2) the data to be operated upon. The CPU will also store (write) data into memory locations as dictated by the program instructions.

2.2. Interaction of Computer CPU with memory

Fig. 6.4 depicts a simple diagram in which the connection between CPU and two memory ICs are shown. Address is send only from CPU to memory. It is represented by an arrow from address bus to memory. Similarly, an arrow from control bus to memory represents that control signals are provided from CPU to memory. An arrow from data bus to memory and another arrow from memory to data bus indicate that data can be transmitted from CPU to memory and from memory to CPU.

Interaction between CPU and memory can be identified by two operations.

- Write operation.
- Read operation.

Interaction between CPU and memory can be identified by two operations.

- Write operation.
- Read operation.

Write operation

Whenever the CPU wants to write data to a particular memory location, the following steps must occur :

1. The CPU supplies the binary address of the memory location where the data are to be stored. It places this address on the address bus lines.
2. The CPU places the data to be stored on the data bus.
3. The CPU activates the appropriate control signal lines for the memory write operation.
4. The memory ICs decode the binary address to determine which location is being selected for the store operation.
5. The data on the data bus are transferred to the selected memory location.

Read Operation

To transfer a byte of data from the memory to the CPU, a read operation must be performed. This address from address buffer is put onto the address bus.

The read signal causes the contents of the selected address to be put on the data bus.

Once the address code is on the bus, the CPU sends a read signal to the memory. At the memory, the address bits are decoded and the desired memory location is selected. The read signal causes the contents of the selected address to be put on the data bus. The data byte is then loaded into the data buffer. This completes read operation. Whenever the CPU wants to read data from a specific memory location, the following steps must occur :

1. The CPU supplies the binary address of the memory location from which data are to be retrieved. It places this address of the address bus lines.
2. The CPU activates the appropriate control signal for read operation .

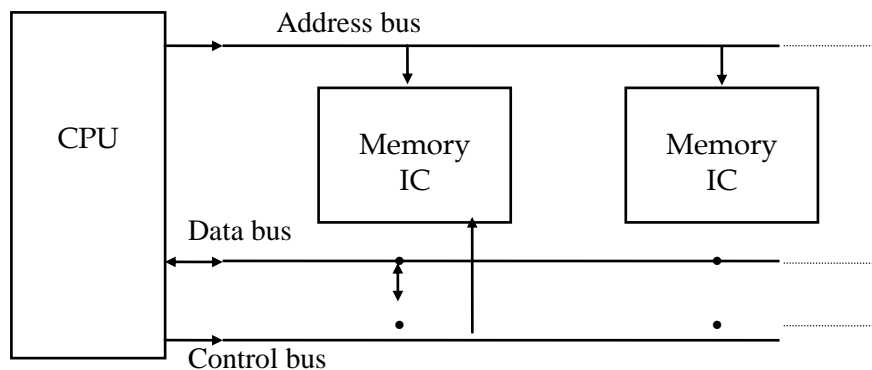


Fig. 6.4 : Three groups of lines (buses) connect the internal memory ICs to the CPU.

3. The memory ICs decode the binary address to determine which location is being selected for the read operation.
4. The memory ICs place data from the selected memory location on to the data bus, from where they are transferred to the CPU.

Bus is the lines through which data, address and control signals transmit from one device to another.

CPU, memory and other devices communicate with each other through bus. Bus is the lines through which data, address and control signals transmit from one device to another. Bus are classified into three categories according to the information passes through it.

- Address Bus
- Data Bus
- Control Bus.

Memory Organization

Address Bus

Address Bus : This is a unidirectional bus that carries the binary address, output from the CPU to the memory ICs to select one memory location. When CPU wants to read data from memory, it must send the address of the location of data. Now, some numbers represent location of memory. This number is called address. During reading or writing data from or to memory, CPU sends address in binary form through address bus.

*Data Bus
Control Bus*

Data Bus : This is a bi-directional bus that carries data between the CPU and the memory ICs.

Control Bus : This bus carries control signals from the CPU to other devices. When CPU instructs some devices to perform some actions, the instruction or orders are provided by some control signals. For example, CPU activate memory by providing 0 (zero) at \overline{CS} control line and provide order for read operation by sending 1 at R/ \overline{W} control line. So, control bus transmits signals for controlling purposes.

2.3. Exercise

2.3.1. Questions for short answers

- How many address inputs, data inputs, and data outputs are required for a 16K X 12 memory?
- What is the function of the R/W input?
- What is the function of the MEMORY ENABLE input?

2.3.2. Analytical questions

- Name the three groups of lines that connect the CPU and internal memory.
- Outline the steps that take place when the CPU reads from memory.
- Outline the steps that occur when the CPU writes to memory.

Lesson 3 : Read-Only Memory (ROM)

3.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ discuss the difference between volatile and nonvolatile memory
- ◆ distinguish among the various types of ROMs and cite common applications
- ◆ compare the relative advantages and disadvantages of EPROM, EEPROM, and flash memory.

The read-only memory is a type of semiconductor memory that is designed to hold data permanently.

The read-only memory is a type of semiconductor memory that is designed to hold data permanently. During normal operation, new data can not be written into a ROM but data can be read from ROM. For some ROMs the data have to be stored during the manufacturing process; for other ROMs the data can be entered electrically. The process of entering data is called programming or burning in the ROM. Some ROMs can be erased and reprogrammed as often as desire. We will take a detailed look later at various types of ROMs.

ROMs are used to store data and information that are not to change be during the normal operation of a system. A major use for ROMs is the storage of programs in microcomputers. Since all ROMs are nonvolatile, these programs are not lost when electrical power is turned off. When the microcomputer is turned on, it can immediately begin executing the program stored in ROM. ROMs are also used for program and data storage in microprocessor controlled equipment such as electronic cash registers, appliances, and security systems.

A $n \times b$ bit ROM can store n words of b bits permanently or semi-permanently (for erasable ROM). Depending on the technology employed and the programming technique, different types of ROMs are realized. While bipolar ROMs have higher speeds of operation having access time around 60 nano second, the MOS ROMs have typical speeds around 500 nano seconds.

3.2. Types of ROMs

Types of ROMs

Now that we have a general understanding of the internal architecture and external operation of ROM devices, we will look at the various types of ROMs.

3.2.1. Mask-Programmed ROM

The mask-programmed ROM has its storage locations written into (programmed) by the manufacturer according to the customer's specifications. A photographic negative called mask is used to control the electrical interconnections on the chip. A special mask is required for each of different set of information to be stored in the ROM. A major disadvantage of this type of ROM is that it cannot be reprogrammed. Mask-programmed ROMs, however, still represent the most economical approach when a large quantity of identically programmed the ROMs are needed. The acronym MROM is used refer to mask-programmed ROMs.

*Mask-Programmed ROM
Programmable ROMs
(PROMS)*

3.2.2. Programmable ROMs (PROMS)

A mask-programmable ROM is very expensive. For lower-volume applications, manufacturers have developed fusible link PROMs that are user-programmable; that is, they are not programmed during the manufacturing process but are custom-programmed by the user. Once programmed, however, a PROM is like an MROM in that it cannot be erased and reprogrammed. Thus, if the program in the PROM is faulty or has to be changed, the PROM has to be thrown away. For this reason, these devices are often referred to as "one time programmable" ROMs.

The user can selectively blow any of these fuse links to produce the desired stored memory data. Typically, data is programmed or "burned" into an address location by : applying the address to the address inputs, placing the desired data at the data pins, and then applying a high-voltage pulse (10-30 V) to a special programming link, on the IC. This causes a large current to flow through each selected fuse link, burning it open and permanently storing a logic 0 in that cell. Once all address locations have been programmed in this manner, the data are permanently stored in the PROM and can be read over and over again by accessing the appropriate address. The data will not change when power is removed from the PROM chip, because nothing will cause an open fuse link to become closed again.

The process of programming a PROM and verifying that the stored data are correct is done automatically by a special apparatus called a PROM programmer.

3.2.3. Erasable Programmable ROM (EPROM)

An EPROM can be programmed by the user and it can also be erased and reprogrammed as often as desire. Once programmed, the EPROM is a nonvolatile memory that will hold stored data permanently. The process for programming an EPROM involves the application of special voltage levels (typically in the 10-to 25-V range) to the appropriate chip

*EPROM is a nonvolatile
memory that will hold
stored data permanently.*

inputs for a specified amount of time (typically 50 ms per address location). The programming process is usually performed by a special programming circuit. The complete programming process can take up to several minutes for one EPROM chip.

The storage cells in an EPROM are MOS transistors. In its normal state, each transistor is off and each cell stores a logic 1. A transistor can be turned on by the application of a high-voltage programming pulse. This keeps the transistor on permanently even when power is removed from the device, and the cell is now storing a logic 0.

Once an EPROM cell has been programmed, it can be erased by exposing it to ultraviolet (UV) light applied through a window on the chip package. This process turns transistor off, and restoring the cell to the logic 1 state. This erasing process typically requires 15 to 20 minutes of exposure to UV rays. Unfortunately, there is no way to erase only selected cells; the UV light erases all cells at the same time so that an erased EPROM stores all 1s. Once erased, the EPROM can be reprogrammed. OE is the output enable V_{pp} is the special programming voltage required during the programming process.

OE is the output enable V_{pp} is the special programming voltage required during the programming process.

3.2.4. Electrically Erasable PROM (EEPROM)

Electrically Erasable PROM

The disadvantages of the EPROM were overcome by the development of the electrically erasable PROM (EEPROM) as an improvement over the EPROM. The modification produce the EEPROM's major characteristic- its electrical reusability. Since this mechanism requires very low currents, the erasing and programming of an EEPROM can be done in circuit (i.e., without a UV light source and special PROM programmer unit).

Advantage of the EEPROM over the EPROM is the ability to electrically erase and rewrite individual bytes (8-bit words) in the memory array.

Another advantage of the EEPROM over the EPROM is the ability to electrically erase and rewrite individual bytes (8-bit words) in the memory array. During a write operation, internal circuitry automatically erases all the cells at an address location prior to writing in the new data. This byte erasability makes it much easier to make changes in the data stored in an EEPROM. Additionally, an EEPROM can be programmed more rapidly than many EPROMs; typically, it takes 5 ms to write into an EEPROM location, compared to 50 ms for an EPROM, though newer EPROMs are much faster (100us).

3.3. Flash Memory

EPROMs are nonvolatile, offer fast access times (typically 120 ns), and have high density and low cost per bit. EEPROMs are nonvolatile, offer fast read access, and allow rapid in-circuit erasure and reprogramming of individual bytes. They are from lower density and much higher cost than EPROMs.

The challenge for semiconductor engineers was to fabricate a nonvolatile memory with the EEPROM's in-circuit electrical erasability, but with densities and costs much closer to EPROMs, while retaining the high-speed read access of both. The response to this challenge was the production of the flash memory.

Flash memories are so-called because of their rapid erase and write times.

Flash memories are so-called because of their rapid erase and write times. Cells on the chip are erased simultaneously; this bulk erase process typically requires hundreds of milliseconds. Some newer flash memories offer a sector erase mode, where specific sectors of the memory array can be erased at one time.

3.4. ROM Applications

The term ROM can be used to include EPROMs, EEPROMs, and flash memory, because during normal operation the stored contents of these devices are not changed. So ROMs are taken to include all semiconductor nonvolatile memory devices, and they are used in applications where nonvolatile storage of information, data, or program codes is needed, and where the stored data rarely or never change. Here are some of the most common application areas.

ROMs are taken to include all semiconductor nonvolatile memory devices

3.4.1. Firmware

The most widespread application of ROMs is in the storage of data and program codes that must be available on power-up in microprocessor-based systems. These data and program codes are called firmware because they are firmly stored in hardware (i.e., ROM chips) and are not subject to change during normal system operation. Some PCs, business computers, and laptop computers store their operating system programs and language interpreters (e.g., BASIC, PASCAL) in ROM firmware so that the computer can be used immediately after power is turned on.

Data and program codes are called firmware because they are firmly stored in hardware.

3.4.2. Bootstrap Memory

Many microcomputers and most larger computers do not have their operating system programs stored in ROM. Instead, these programs are stored in external mass memory, usually magnetic disk. A relatively small program, called a bootstrap program, is stored in ROM. When the

A relatively small program, called a bootstrap program, is stored in ROM.

computer is powered on, it will execute the instructions that are in this bootstrap program. These instructions typically cause the CPU to load the system programs from mass storage (disk) into its main internal memory. At that point the computer is ready to respond to the user commands. This startup process is often called "booting up the system".

3.4.3. Data Tables

ROMs are often used to store tables of data that do not change. Some examples are the trigonometric tables (i.e., sine, cosine, etc.) and code-conversion tables.

3.5. Exercise

3.5.1. Questions for short answers

- a) How does a PROM differ from an MROM? Can it be erased and reprogrammed?
- b) How EPROM can be erased?
- c) What EPROM shortcoming are overcome by EEPROM?
- d) What are the major drawbacks of the EEPROM?
- e) What is the main advantage of flash memory?

3.5.2. Analytical questions

- a) Describe how a computer uses a bootstrap program.
- b) Illustrate some of the ROM applications.

Lesson 4 : ROM Structure and Addressing

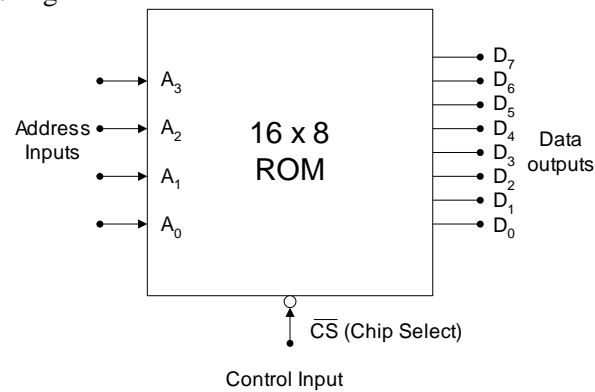
4.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ describe the read and write operation of read-only memory
- ◆ determine the capacity of a memory device from its inputs and outputs
- ◆ understand the structure of read-only memory.

4.2. ROM Block Diagram

A typical block diagram for a ROM is shown in Fig. 6.5. It has three sets of signals : address inputs, control inputs(s), and data outputs. This ROM is storing



(a)

Word	Address				Data								Word	Address				Data D ₇ -D ₀
	A ₃	A ₂	A ₁	A ₀	D ₇	D ₆	D ₅	D ₄	D ₃	D ₂	D ₁	D ₀		A ₃	A ₂	A ₁	A ₀	
0	0	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	DE	
1	0	0	0	1	0	0	1	1	1	0	1	0	1	1	1	3A		
2	0	0	1	0	1	0	0	0	0	1	0	1	2	2	2	85		
3	0	0	1	1	1	0	1	0	1	1	1	1	3	3	3	AF		
4	0	1	0	0	0	0	0	1	1	0	0	1	4	4	4	19		
5	0	1	0	1	0	1	1	1	1	0	1	1	5	5	5	7B		
6	0	1	1	0	0	0	0	0	0	0	0	0	6	6	6	00		
7	0	1	1	1	1	1	1	0	1	1	0	1	7	7	7	ED		
8	1	0	0	0	0	0	1	1	1	1	0	0	8	8	8	3C		
9	1	0	0	1	1	1	1	1	1	1	1	1	9	9	9	FF		
10	1	0	1	0	1	0	1	1	1	0	0	0	10	A	A	B8		
11	1	0	1	1	1	1	0	0	0	1	1	1	11	B	B	C7		
12	1	1	0	0	0	0	1	0	0	1	1	1	12	C	C	27		
13	1	1	0	1	0	1	1	0	1	0	1	0	13	D	D	6A		
14	1	1	1	0	1	1	0	1	0	0	1	0	14	E	E	D2		
15	1	1	1	1	0	1	0	1	1	0	1	1	15	F	F	5B		

(b)

(c)

Typical block diagram for a ROM.

Fig. 6.5 : (a) Typical ROM block symbol; (b) table showing binary data at each address location; (c) same table in hex.

16 words, since it has $2^4 = 16$ possible addresses, and each word contains 8 bits, since there are eight data outputs. Thus, this is a 16 X 8 ROM. Another way to describe this ROM's capacity is to say that it stores 16 bytes of data. The most common numbers of data outputs for ROMs are 4, 8, and 16 bits, 8-bit words being the most common.

The control input CS stands for chip select. This is essentially an enable input that enables or disables the ROM outputs. Therefore it must be in the LOW state to enable the ROM data to appear at the data outputs.

The Read Operation

The Read Operation : Sixteen different data words are stored at the 16 different address locations. For example, the data word stored at location 0011 is 10101111. Of course, the data are stored in binary inside the ROM, but very often we use hexadecimal notation to efficiently show the programmed data.

4.3. ROM Architecture

The internal architecture (structure) of a ROM IC is very complex. There are four basic part : register array, row decoder, column decoder, output buffers.

Register Array : The register array stores the data that have been programmed into the ROM. Each register contains a number of memory cells equal to the word size. We can specify the position of each register as being in a specific row and specific column.

The data outputs of each register are corrected to an internal data bus that runs through the entire circuit. Each register has enable inputs.

Address Decoders : The address code determines which register in the array will be enabled to place its 8-bit data word onto the bus. Address bits are fed to a decoder which activates one row-select line, and address bits

*ROM Architecture
Register Array
Address Decoders
Output Buffers*

Only one register will be in both the row and the column selected by the address inputs, and this one will be enabled.

Output Buffers : The register that is enabled by the address inputs will place its data on the data bus. These data feed into the output buffers, which will pass the data to the external data outputs, provided that CS is LOW.

Problem

Problem

Describe the internal architecture of a ROM that stores 4K bytes and uses a square register array.

Solution

Solution

4K is actually $4 \times 1024 = 4096$, and so this ROM holds 4096 eight-bit words. Each word can be thought of as being stored in an 8-bit register, and there are 4096 registers connected to a common data bus internal to the chip. Since $4096 = 64^2$, the registers are arranged in a 64-by-64 array; that is, there are 64 rows and 64 columns. This requires a 1-of-64 decoder to decode six address inputs for the row select, and a second 1-of-64 decoder to decode six other address inputs for the column select. Thus, a total of 12 address inputs is required. This makes sense, since $2^{12} = 4096$, and there are 4096 different address.

4.4. Exercise

4.4.1. Questions for short answer

a) Draw the block diagram of a 16×4 ROM.

4.4.2. Analytical question

a) Describe the function of the row-select decoder, column-select decoder, and output buffers in the ROM architecture.

Lesson 5 : Random Access Memory (RAM)

5.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ describe the read and write operation of random access memory
- ◆ determine the capacity of random access memory device from its inputs and outputs.

5.2. Semiconductor RAMs

RAM stands for random-access memory, meaning that any memory address location is easily accessible.

RAM stands for random-access memory, meaning that any memory address location is easily accessible. The term RAM is used with semiconductor memories which is usually taken to mean read/write memory (RWM).

RAM is used in computers for the temporary storage of programs and data. The contents of many RAM address locations will be read from and written to as the computer executes a program. This requires fast read and write cycle times for the RAM so as not to slow down the computer operation.

A major disadvantage of RAM is that it is volatile and will lose all stored information if power is interrupted or turned off. Some CMOS RAMs, however, use such small amounts of power in the standby mode (no read or write operations taking place) that they can be powered from batteries whenever the main power is interrupted. Of course, the main advantage of RAM is that it can be written into and read from rapidly with equal ease.

5.3. RAM Architecture

The RAM consisting of a number of registers, each storing a single data word, and each having a unique address. RAMs typically come with word capacities of 1K, 4K, 8K, 16K, 64K, 128K, 256K, and 1024K, and word sizes of 1, 4 or 8 bits. The word capacity and word size can be expanded by combining memory chips.

The simplified architecture of a RAM stores 64 words of 4 bits each (i. e., a 64×4 memory).

The simplified architecture of a RAM stores 64 words of 4 bits each (i. e., a 64×4 memory). These words have addresses ranging from 0 to 63_{10} . In order to select one of the 64 address locations for reading or writing, a binary address code is applied to a decoder circuit. Since $64 = 2^6$, the decoder requires a 6-bit input code. Each address code activates one particular decoder output, which, in turn, enables its corresponding register. For example, assume an applied address code of

Memory Organization

$$A_5A_4A_3A_2A_1A_0 = 011010$$

Since $011010_2 = 26_{10}$, decoder output 26 will go high, selecting register 26 for either a read or write operation.

Read Operation : The address code picks out a register in the memory chip for reading or writing. In order to read the contents of the selected register, the READ/ WRITE (R/W)* input must be a 1. In addition, the CHIP SELECT (CS) input must be activated (a 0 in this case). The combination of R/W = 1 and CS = 0 enables the output buffers so that the contents of the selected register will appear at the four data output. R/W = 1 also disables the input buffers so that the data inputs do not affect the memory during a read operation.

Read Operation
Write Operation
Chip Select

Write Operation : To write a new 4-bit word into the selected register requires R/W = 0 and CS = 0. This combination enables the input buffers so that the 4-bit word applied to the data inputs will be loaded into the selected register. The R/W = 0 also disables the output buffers, which are tristate, so that the data outputs are in their Hi-Z state during a write operation. The write operation, of course, destroys the word that was previously stored at that address.

Chip Select : Most memory chips have one or more CS inputs which are used to enable the entire chip or disable it completely. In the disabled mode all data inputs and data outputs are disabled (Hi-Z) so that neither a read nor a write operation can take place. In this mode the contents of the memory are unaffected. The reason for having CS inputs will become clear when we combine memory chips to obtain larger memories. These inputs are called CHIP ENABLE (CE). When the CS or CE inputs are in their active state, the memory chip is said to be selected; otherwise it is said to be deselected. Many memory ICs are designed to consume much lower power when they are deselected. In large memory systems, for a given memory operation, one or more memory chips will be selected while all others are deselected.

Common Input/ Output Pins : The data input and data output functions use common input/ output pins. The R/W input controls the function of these I/O pins. During a read operation, the I/O pins act as data outputs which reproduce the contents of the selected address location. During a write operation, the I/O pins act as data inputs to which the data to be written are applied.

5.4. Exercise

5.4.1. Questions for short answers

- a) What do you understand by RAM?
- b) What is RAM structure?

5.4.2. Analytical question

- a) Give a brief a description of read and write operations using RAM.

Lesson 6 : Static and Dynamic RAM

6.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ describe the structure and operation of SRAM and DRAM.

Prior to the availability of MOS technology, memory chips used to be fabricated with cells built with bipolar junction transistors. Due to inherent advantages of higher density and lesser power consumption, MOS technology has subsequently become de facto standard for main storage module of a computer system. Basically, MOS semiconductor storage cells can be of two types :

1. Static cell, where binary information is stored due to the state of conduction of a transistor.
2. Dynamic cell, where information is stored due to small capacitance which holds binary information in the form of presence or absence of electric charge. While a static cell can store information indefinitely, the charge in a dynamic device gets depleted with time due to leakage and hence it needs periodic refreshing. However, a dynamic cell needs lesser power and silicon area compared to a static cell.

6.2. Static RAM (SRAM)

Static Memory Devices are semiconductor memory devices in which the stored data will remain permanently stored as long as power is applied, without the need for periodically and will store data when electrical power is removed.

Static Memory Devices are semiconductor memory devices in which the stored data will remain permanently.

The RAM operation that have been discussing up to this point applies to a static RAM - one that can store data as long as power is applied to the chip. Static RAM memory cells are essentially flip-flops that will stay in a given state (store a bit) indefinitely, provided that power to the circuit is not interrupted. Dynamic RAMs, which store data as charges on capacitors. With dynamic RAMs the stored data will gradually disappear because of capacitor discharge, so that it is necessary to periodically refresh the data (i. E., recharge the capacitors).

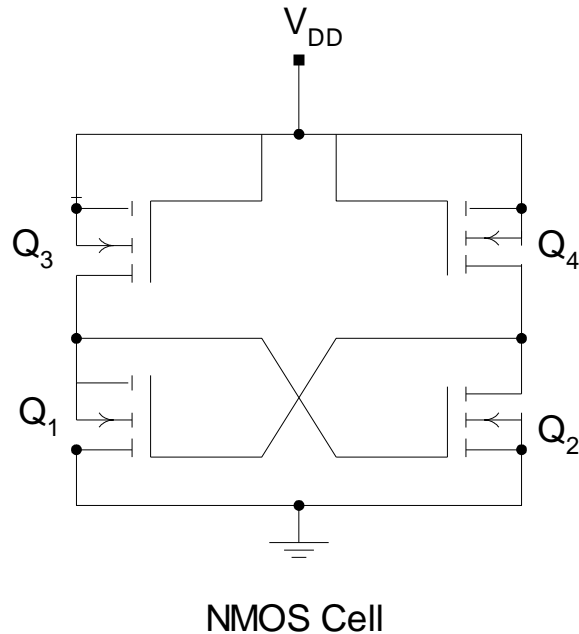


Fig. 6.7 : Typical bipolar and NMOS static RAM cells.

Static RAMs (SRAMs) are available in bipolar, MOS and BICMOS technologies, the majority of applications use NMOS or CMOS RAMs. As stated earlier, the bipolar have the advantage in speed and MOS devices have much greater capacities and lower power consumption. Fig. 6.7 shows for comparison a typical bipolar cell contains two bipolar transistors and two resistors. The bipolar cell requires more chip area than the MOS cell because a bipolar transistor is more complex than a MOSFET, and because the bipolar cell requires separate resistors while the MOS cell uses MOSFETS as resistors (Q_3 and Q_4). A CMOS memory cell.

The main advantages of static RAM are that these are non-destructive, simple to operate, need no refreshing and can be operated at higher speed. However, such cells need large dc power and more support circuit for decoding and control. This leads to larger area resulting in reduced packing density. Use of dynamic memory cells, have played a crucial role to increase the packaging density of memory chips.

6.3. Dynamic RAM (DRAM)

Dynamic Memory Devices are semiconductor memory devices in which the stored data will not remain permanently stored, even with power applied, unless the data are periodically rewritten into memory. The latter operation is called a refresh operation.

Dynamic Memory Devices are semiconductor memory devices in which the stored data will not remain permanently stored.

Memory Organization

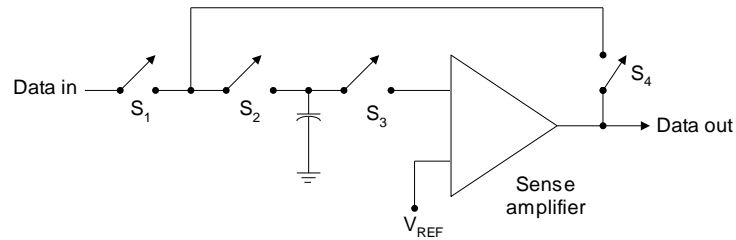


Fig. 6.8 : Symbolic representation of a dynamic memory cell. During a WRITE operation, semiconductor switches S₁ and S₂ are closed. During a read operation, all switches are closed excepts S₁.

Dynamic RAMs are fabricated using MOS technology and are noted for their high capacity, low power requirement, and moderate operating speed. Unlike static RAMs, which store information in FFs, dynamic RAMs store 1's and 0's as charges on a small MOS capacitor. Because of the tendency for these charges to leak off after a period of time, dynamic RAMs require periodic recharging of the memory cells; this is called refreshing the dynamic RAM. In modern DRAM chips, each memory cell must be refreshed typically every, 2, 4, or 8 ms, or its data will be lost.

The need for refreshing is a drawback of dynamic RAM as compared to static RAM because it may require external support circuitry. Some DRAM chips have built-in refresh control that does not require extra external hardware but does require special timing of the chips input control signals. Designing with and using DRAM in a system is more complex than SRAM. However, their much larger capacities and much lower power consumption make DRAM the memory of choice in systems where the most important design considerations are keeping down of size, cost, and power.

For applications where speed and reduced complexity are critical than cost, space, and power considerations, static RAMs are still the best. They are generally faster than dynamic RAMs and require no refresh operation. They are simpler to design with, but they cannot compete with the higher capacity and lower power requirement of dynamic RAMs.

Because of their simple cell structure, DRAMs typically have four times the density of SRAMs. This allows four times as much memory capacity to be placed on a single board. The cost per bit of dynamic RAM storage is typically one-fifth to one-fourth that of static RAMs. Power requirements of a dynamic RAM, typically one-sixth to one-half those of a static RAM.

The main applications of SRAMs are in areas where only amounts of memory are needed (up to 64K), or where high speed is required.

The main applications of SRAMs are in areas where only amounts of memory are needed (up to 64K), or where high speed is required. Many microprocessor-controlled instruments and appliances have very small memory capacity requirements. Some instruments, such as digital storage oscilloscopes and logic analyzers, require, very high-speed memory and SRAM is normally used.

The main internal memory of most personal microcomputers uses DRAM because of its high capacity and low power consumption. These computers, however, sometimes use some small amounts of SRAM for functions requiring maximum speed such as video graphics and look-up tables.

6.4. Exercise

6.4.1. Questions for short answers

- a) Which memory technology generally uses the least power?
- b) What are the main drawbacks of dynamic RAM compared with static?
- c) List the advantages of dynamic RAM compared with static RAM.

6.4.2. Analytical questions

- a) Describe the input conditions needed to read a word from a specific RAM address location.
- b) Describe the structure and operation of 16×1 Dynamic RAM.

Lesson 7 : Expanding Word Size and Capacity Expansion

7.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ combine memory ICs to form memory modules with larger word size and capacity.

Several memory chips have to be combined to provide the capacity and / or word size.

In many memory applications the required RAM or ROM memory capacity or word size cannot be satisfied by one memory chip. Instead, several memory chips have to be combined to provide the capacity and / or word size.

7.2. Expanding Word Size

Suppose we need a memory that can store sixteen 8-bit words and all we have are RAM chips which are arranged as 16×4 chips to produce the desired memory.

Since each chip can store sixteen 4-bit words and we want to store sixteen 8-bit words we are using each chip to store half of each word. In other words, RAM-0 stores the four higher-order bits of each of the 16 words, and RAM-1 stores the four lower-order bits of each of the 16 words. A full 8-bit word is available at the RAM outputs connected to the data bus.

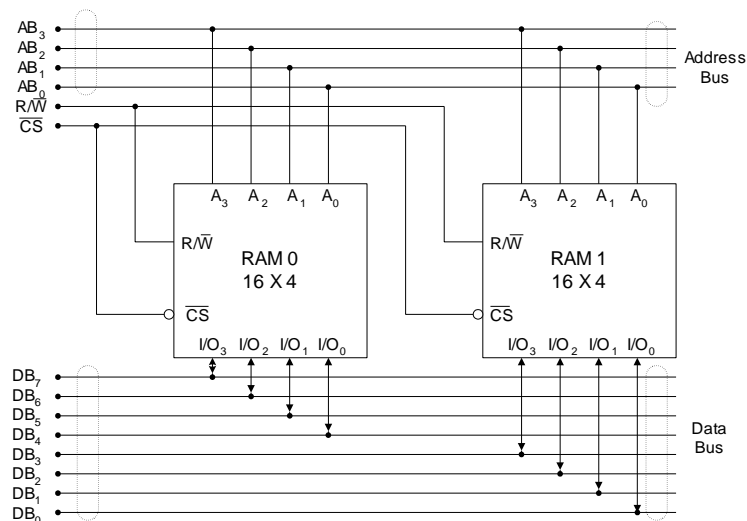


Fig. 6.9 : Combining two 16×4 RAMs for a 16×8 module.

Any one of the 16 words is selected by applying the appropriate address code to the four-line address bus (AB_3, AB_2, AB_1, AB_0).

Each, address bus line is connected to the corresponding address input of each chip. This means that once a address code is places on the address bus, this same address code is applied to both chips so that the same location is each chip is accessed at the same time.

Once the address is selected, we can read or write at this address under control of the common R/W and CS line. To read, R/W must be high and CS must be low. This causes the RAM I/O lines at act as outputs. RAM-0 places its selected 4-bit word on the upper four data bus lines and RAM-1 places its selected 4-bit word on the lower four data bus lines. The data bus then contains the full selected 8-bit word.

To write, $R/W = 0$ and $CS = 0$ causes the RAM I/O lines to act as inputs. The 8-bit word to be written is places on the data bus. The higher 4 bits will be written into the selected location of RAM-0 and the lower 4 bits will be written into RAM-1.

The combination of the two RAM chips acts like a single 16×8 memory chip. We would refer to this combination as a 16×8 memory module.

The same basic idea for expanding word size will work for many different situations.

7.3. Expanding Capacity

Suppose we need a memory that can store thirty-two 4-bit words and all we have are the 16×4 chips. By combining two 16×4 chips we can produce the desired memory.

By combining two 16×4 chips we can produce the desired memory.

Each RAM is used to store sixteen 4-bit words. The data four I/O pins of each RAM are connected to a common four-line data bus. Only one of the RAM chips can be selected at one time. This is ensured by driving the respective CS inputs from different logic signals.

Memory Organization

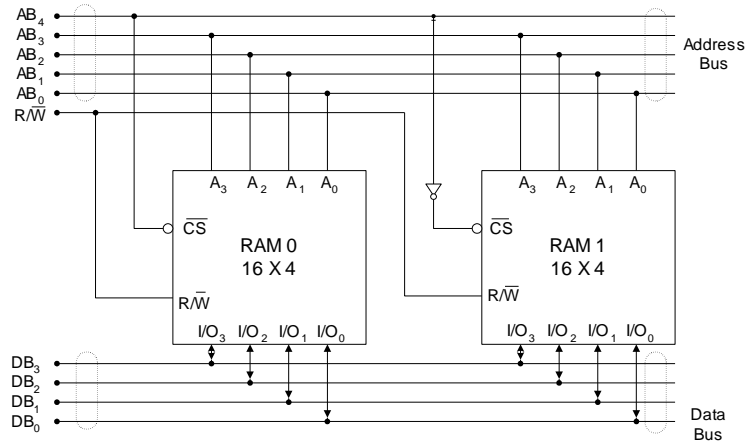


Fig. 6.10 : Combining two 16×4 chips for a 32×4 memory.

Since the total capacity of this memory module is 32×4 , there have to be 32 different addresses. This requires few address bus lines. The upper address line AB_4 is used to select one RAM or the other (via the CS inputs) as the one that will be read from or written into. The other four address lines AB_0 to AB_3 are used to select the one memory location out of 16 from the selected RAM chip.

When $AB_4 = 0$, the CS of RAM-0 enables this chip for read or write. Then, any address location in RAM-0 can be accessed by AB_3 through AB_0 . The latter four address lines can range from 0000 to 1111 select the desired location. Thus, the range of addresses representing locations in RAM-0 is

$$AB_3AB_2AB_1AB_0 = 00000 \text{ to } 01111$$

When $AB_4 = 0$, the CS of RAM-1 is high, so that its I/O lines are disabled (Hi-Z) and cannot communicate (give or take data) with the data bus.

When $AB_4 = 1$, the roles of RAM-0 and RAM-1 are reversed. RAM-1 is now enabled and the lines AB_3 to AB_0 select one of its locations. Thus, the range of addresses location in RAM-1 is

$$AB_3AB_2AB_1AB_0 = 10000 \text{ to } 11111$$

Problem

Problem and Solution

To combine several $2K \times 8$ PROMs to produce a total capacity of $8K \times 8$. How many PROM chips are needed? How many address bus lines are required?

Solution

For PROM chips are required, with each one storing 2K of the 8K words. Since $8K = 8 \times 1024 = 8192 = 2^{13}$, thirteen address lines are needed.

The configuration for the memory of example 11-13 is similar to the 32×4 memory. However, it is slightly more complex, because it requires a decoder circuit for generating the CS input signals.

7.4. Exercise

7.4.1. Analytical questions

- a) Describe briefly, how you can expand a word size.
- b) Explain the expanding capacity of a computer.

Lesson 8 : Memory Mapping and Other Memory Devices

8.1. Learning Objectives

On completion of this lesson you will be able to :

- ◆ understand the memory mapping techniques
- ◆ understand the construction and operation of Charge-Couple Devices, magnetic bubble memory and other memory devices.

8.2. Memory

Memory is an essential components of a microcomputer system. It stores binary in structions and data for the microprocessor.

Memory is an essential components of a microcomputer system. It stores binary in structions and data for the microprocessor. Two types of memory were identified in the last chapter : Read/ Write Memory (R/WM) and Read Only Memory (ROM). The R/W Memory is made of registers, and each register has a group of flip-flops that stores bits of information. The number of bits stored in a register is called a memory word; memory devices (chips) are available in various word sizes. The MPU can read from or write into this memory. The second type of memory, the ROM, stores information permanently in the form of diodes; a group of diodes can be viewed as a register. The MPU can only read information form the ROM; it cannot write into this memory.

In a memory chip, all registers are arranged in a sequence and identified by binary numbers called memory addresses. The assignment of memory addresses to various registers in a memory chip and the process of communication between the MPU and the memory are described below. The following discussion is equally applicable to R/WM and ROM except for slight differences in Read/ Write control signals.

8.3. Memory Map

Memory map is defined as the assignment of addresses to memory registers in various memory chips in a system.

Memory map is defined as the assignment of addresses to memory registers in various memory chips in a system. In a system based on the 8085/8080A microprocessor, the entire memory map can range from 0000H to FFFFH ($2^{16} = 65,536$). This memory map can be illustrated with an analogy of identical house built in a sequence and their postal addresses or numbers.

Let us assume that houses are given four-digit decimal numbers, which will enable us to number ten thousand houses from 0000 to 9999. Since it is cumbersome to direct someone to house with large numbers, the numbering scheme can be devised with the concept of a row or block. Each block will have a hundred houses to be numbered with the last two

digits from 00 to 99. Similarly, the blocks are also identified by the first two decimal digits. For example, a house with the number 0247 is house number 47 in block 2. With this scheme, all the houses in block 0 will be identified from 0000 to 0099, in block 20 from 2000 to 2099, and in block 99 from 9900 to 9999. This numbering scheme with four decimal digits is capable of giving addresses to ten thousand houses from 0000 to 9999 (100 blocks of 100 houses each). A new area under development may have only two blocks completed - block 0 and block 20 - the houses on these blocks can have addresses 0000 to 0099 and 2000 to 2099, even if other blocks are still empty. Let us also assume that all houses are identical, and have eight rooms.

Memory Mapping

The example of numbering the houses is directly applicable to assigning addresses to memory registers. In the binary number system, sixteen binary digits can have 65,536 (2^{16}) different combinations. In the hexadecimal number system, sixteen binary bits are equivalent to four Hex digits that can be used to assign addresses to 65,536 memory registers from 0000H to FFFFH; however, it is easier to think in terms of two Hex digits than four Hex digits. Therefore, the concept of memory page, similar to the concept of block, can be devised, as explained below.

In a memory system, memory registers can be conceptually organized in groups to be numbered with low-order two hexadecimal digits, similar to the last two digits of the house address. With two Hex digits, 256 registers can be numbered from 00H to FFH; this is known as a page with 256 lines to read from or write on. Although the number FFH is equal to 255, the total number of lines is equal to 256 because the first line is numbered 0. Similarly, the high-order (first) two Hex digits can be used to number the pages from 00H to FFH. For example, the memory address 020FH represents line 15 (register) on page 2, the address 07FFH represents line 255 on page 7, and the address 1064H represents line 100 on page 16 ($64_{10} = 100_{10}$ and $10_{10} = 16_{10}$). The total memory address will range from 0000H to FFFFH - 256 pages with 256 lines each ($256 \times 256 = 65,536$, known as 65K). To complete the analogy, a line (register) is equivalent to a house, a page is equivalent to a block, and eight flip-flops in a register are equivalent to eight rooms in a house. These concepts are further illustrated by the following example of a memory chip with 256 bytes of memory.

8.4. Charge-Couple Devices

A charge-couple device (CCD) resembles a shift register except.

A charge-couple device (CCD) resembles a shift register except. It is based on the presence or absence of electric charge to represent stored information rather than the state of a flip-flop. Charge, in the form of holes or electrons, is made to move along a line from one location to another with the use of electrodes placed at each location. It consists of a

Memory Organization

substrate of n-type semiconductor material covered with a layer of insulation upon which are placed many electrodes. The electrodes are connected to three lines such that every third electrode is connected to the same line. Each such line is driven by one phase of a three-phase clock.

The apparent simplicity in the structure of charge-coupled devices leads to the potential of high density and low cost per bit, making them quite attractive for use in applications of computers, specially microcomputer.

8.5. Magnetic-Bubble Devices

Magnetic-bubble memory devices store information by the means of magnetic domains that move within a sheet of magnetic material.

Magnetic-bubble memory devices store information by the means of magnetic domains that move within a sheet of magnetic material. Magnetic domains are localized regions of magnetic flux that normally occur in any ferromagnetic material owing to alignment of the magnetized molecules. The presence of such domains is what gives rise to the magnetic field of a permanent magnet. The domains in a thin sheet of ferromagnetic material can be shaped into so-called bubbles by applying an external magnetic field in a direction perpendicular to the sheet. These bubbles are easily moved about with the use of other magnetic fields.

Magnetic-bubble memory devices are only economical if they are made to store a relatively large number of bits. However, this would normally lead to an excessively long average access time.

8.6. Queue Memories

The transfer information between components of a microcomputer or other digital system is an important concern. Situations often arise in which information is sent at varying rates from one component to another component that can only accept such information at or below a prescribed rate. To reconcile such differences in data rate for the two components, a storage device known as a queue memory may be used. A queue memory is a form of sequential-access memory in which words are written and read in the same order but at rates independent of each other. They are sometimes commercially referred to as first-in first-out (FIFO) memories.

A queue memory is often useful as a storage buffer between a computer and a peripheral device. In the case of an output device the computer can then output information as it is determined without concern for the rate in which the device can accept the information, provided that the capacity of the queue is not exceeded.

8.7. Exercise

8.7.1. Questions for short answers

- a) What is the memory mapping technique?
- b) How information is stored in the Charge-Couple Devices?

8.7.2. Analytical questions

- a) Describe the construction of Charge-Couple Devices and Magnetic-Bubble Devices.
- b) Explain the operation of the queue Memories.