Memory Unit

Unit: 4

Computer Organization & Architecture

B Tech 3rd Sem

Khushboo

Assistant Professor

ECE Department

Content

Memory:

- Basic concept and hierarchy
- semiconductor RAM memories
- 2D & 2 1/2D memory organization
- ROM memories
- Cache memories: concept and design issues & performance address mapping and replacement
- Auxiliary memories: magnetic disk, magnetic tape and optical disks
- Virtual memory: concept implementation

.

Course Objective

- Study the various memory used by computer like hierarchical memory system, RAM, ROM, 2D & 2 1/2D memory organization.
- Study of Cache memories and virtual memory , Auxiliary memories: magnetic disk, magnetic tape and optical disks .

Course Outcome

Understanding the hierarchical memory system, cache memories and virtual memory

CO-PO Mapping

COMPUTER ORGANIZATION AND ARCHITECTURE (KCS-302)												
CO.K	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
KCS-302.4	3	2	2	2	2	1	1	1	1	1	1	2

CO- PSO Mapping

COMPUTER ORGANIZATION AND ARCHITECTURE (KCS-302)				
CO.K	**PSO1**	**PSO2**	**PSO3**	**PSO4**
KCS-302.2	2	3	3	2

Prerequisite and Recap

- Basics of Computer Organization & architecture
- Functional unit and their interconnection

Memory Unit

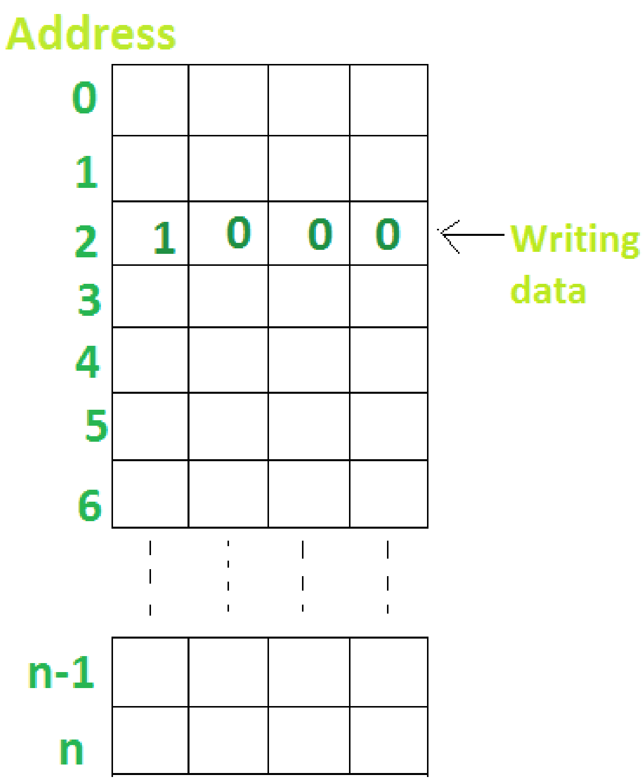
- A **Memory Unit** is a collection of storage cells together with associated circuits needed to transfer information in and out of storage.
- Memories are made up of registers. Each register in the memory is one storage location. The storage location is also called as memory location. Memory locations are identified using **Address**.
- The total number of bit a memory can store is its **capacity**.
- The memory stores **binary information(1's and 0's)** in groups of **bits** called **words**. A storage element is called a **Cell**.

Memory Unit

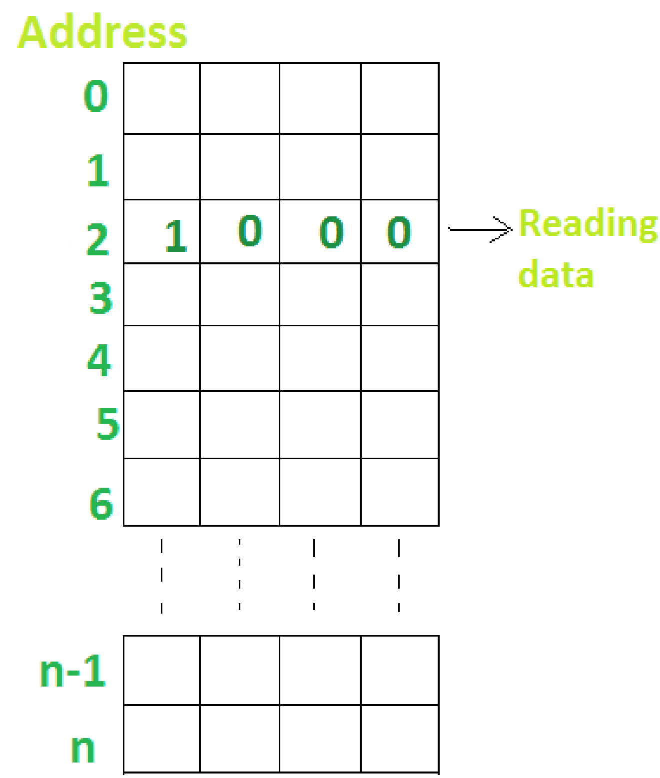
- A group of **eight bits** is called a **byte**. Most computer memories use words whose number of bits is a multiple of 8.

- The capacity of memories in commercial computers is usually stated as the total number of bytes that can be stored.

Memory Unit



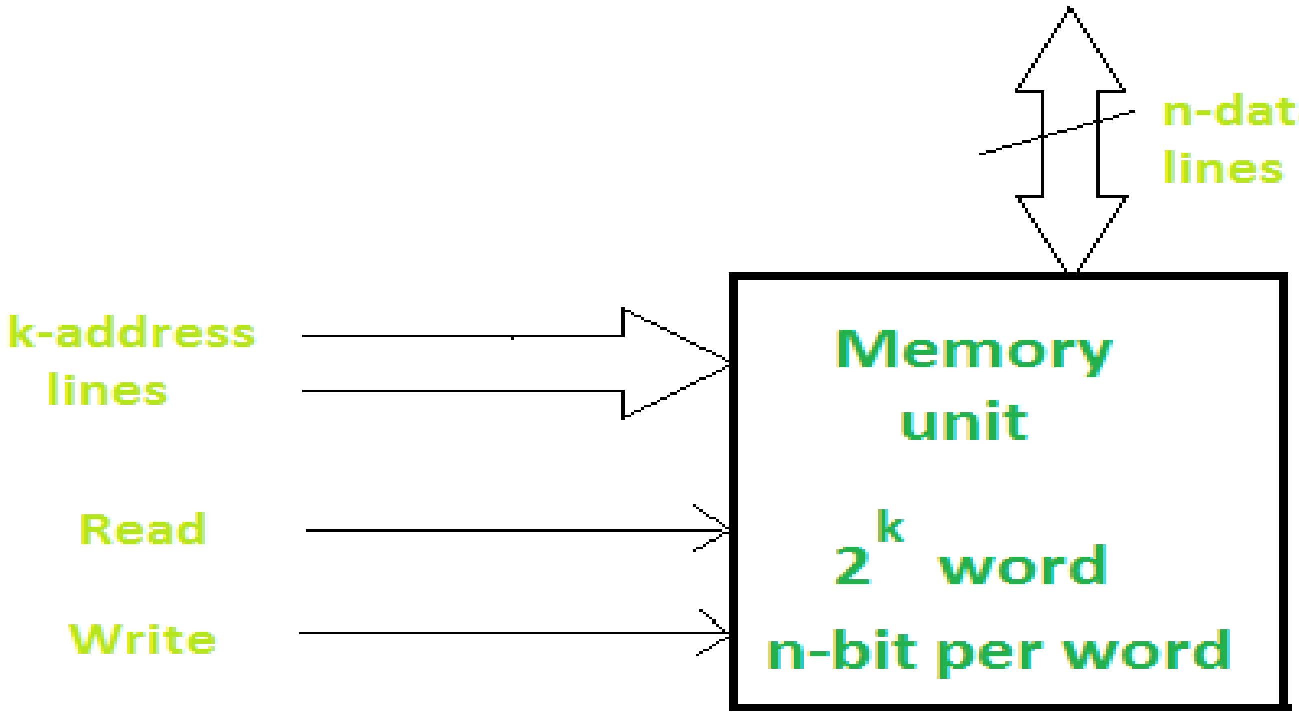
A)Write operation



B)Read Operation

Memory Unit

- A memory unit consists of data lines, address selection lines, and control lines that specify the direction of transfer. The block diagram of a memory unit is shown below:



Memory Unit

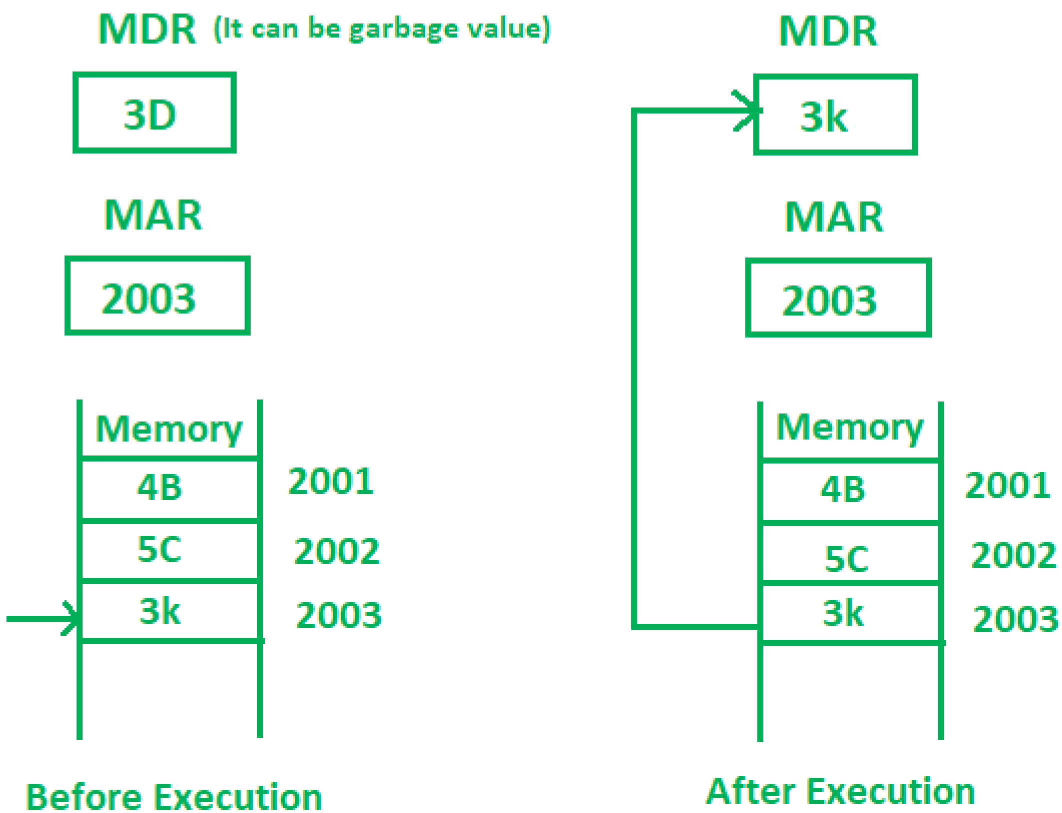
- Data lines provide the information to be stored in memory. The control inputs specify the direction transfer.

- **The k-address lines specify the word chosen. When there are k address lines, 2^k memory word can be accessed.**

- Memory read operation transfers the desired word to address lines and activates the read control line.

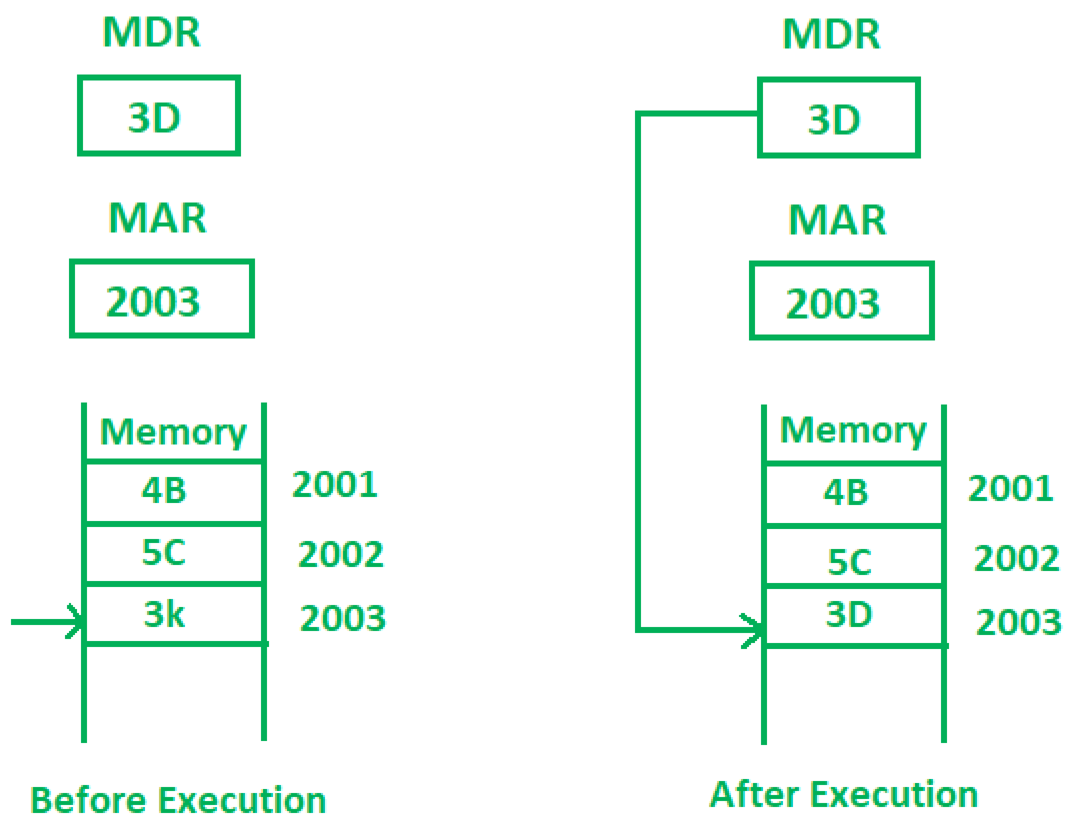
- Memory write operation transfers the address of the desired word to the address lines, transfers the data bits to be stored in memory to the data input lines. Then it activates the write control line.

Memory Unit



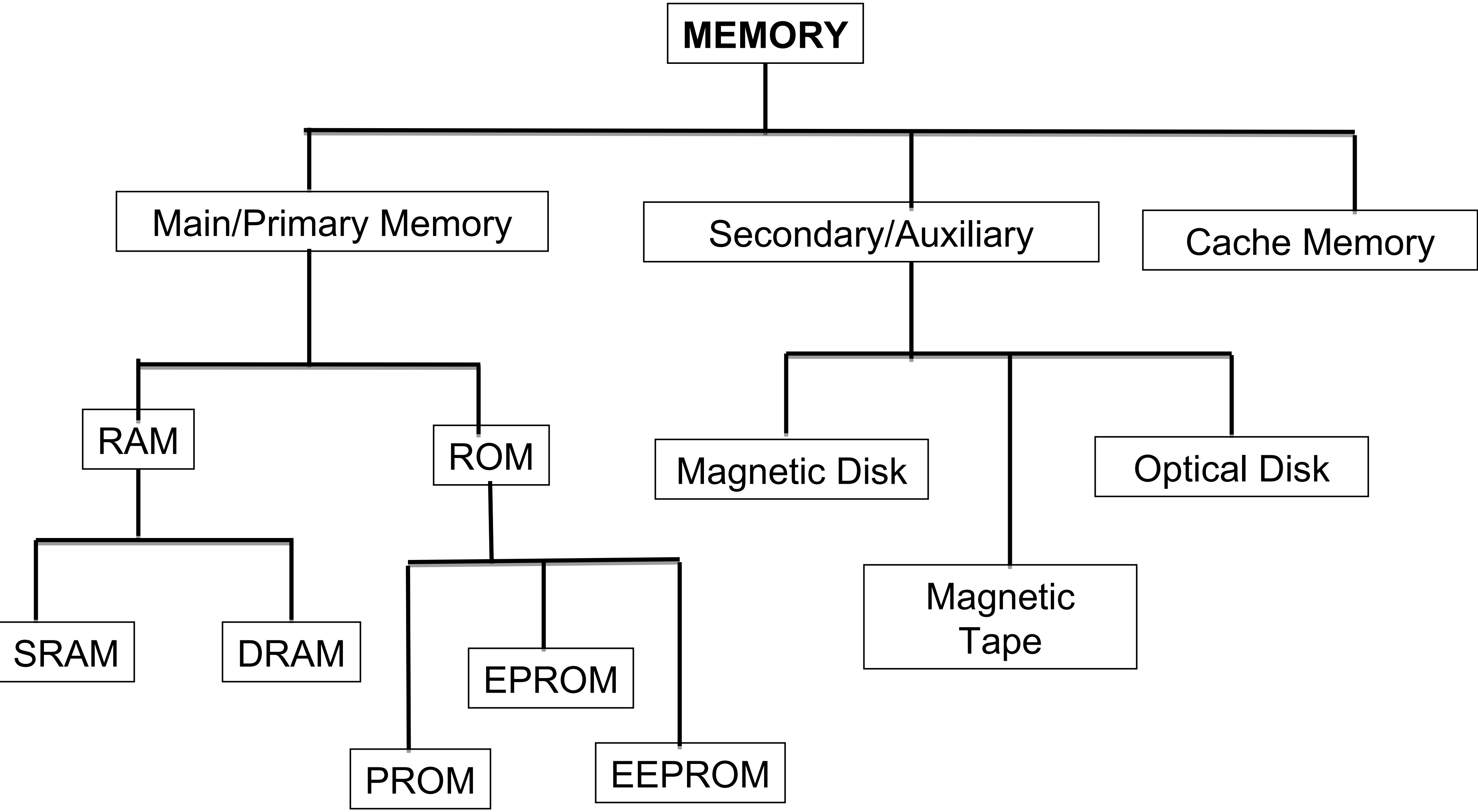
Memory Read Operation

Memory Unit



Memory Write Operation

Classification of Memory



Classification of Memory

Primary memory	Secondary memory
Primary memory is directly accessible by Processor/CPU.	Secondary memory is not directly accessible by the CPU.
Nature of Parts of Primary memory varies, RAM- volatile in nature. ROM- Non-volatile.	It's always Non-volatile in nature.
Primary memory devices are more expensive than secondary storage devices.	Secondary memory devices are less expensive when compared to primary memory devices.
The memory devices used for primary memory are semiconductor memories.	The secondary memory devices are magnetic and optical memories.
Main memory or Internal memory.	External memory or Auxiliary memory.
Examples: RAM, ROM, Cache memory, PROM, EPROM, Registers, etc.	Examples: Hard Disk, Floppy Disk, Magnetic Tapes, etc.

Classification of Memory

Classification of Primary Memory

Primary memory can be broadly classified into two parts:

1.Read-**O**nly **M**emory (ROM)

2.Random **A**ccess **M**emory (RAM)

1.Read-Only Memory

•Any data which need not be altered are stored in ROM.

•Most of the main memory in a general-purpose computer is made up of RAM integrated circuit chips, but a portion of the memory may be constructed with ROM chips.

•ROM includes those programs which run on booting of the system *(know as* a *bootstrap program* that initializes *OS*-The bootstrap loader is a program whose function is to start the computer software operating when power is turned on.*)* along with data like algorithm required by OS.

Classification of Memory

Types of ROM:

•**PROM:** *Programmable ROM* can be modified once by the user. The user buys a blank PROM and writes the desired content but once written content cannot be altered.

•**EPROM:** *Erasable and Programmable ROM* Content can be changed by erasing the initial content which can be done by exposing EPROM to UV radiation. This exposure to ultra-violet light dissipates the charge on ROM and content can be rewritten on it.

•**EEPROM:** *Electrically Erasable and Programmable ROM* Content can be changed by erasing the initial content which could be easily erased electrically. However, one byte can be erased at a time instead of deleting in one go. Hence, reprogramming of EEPROM is a slow process.

Classification of Memory

2. Random Access Memory

- Any process in the system which needs to be executed is loaded in RAM which is processed by the CPU as per Instructions in the program. Like if we click on applications like Browser, firstly browser code will be loaded by the Operating system into the RAM after which the CPU will execute and open up the Browser.

Types of RAM:

- RAM can be broadly classified into SRAM (Static RAM) and DRAM (Dynamic RAM) based on their behavior:

Classification of Memory

1. **Static Random Access Memory (SRAM) :**

- Data is stored in transistors and requires a constant power flow. Because of the continuous power, SRAM doesn't need to be refreshed to remember the data being stored.
- SRAM is called static as no change or action i.e. refreshing is not needed to keep the data intact.
- It is used in cache memories.

Classification of Memory

2. Dynamic Random Access Memory (DRAM) :

- Data is stored in capacitors.
- Capacitors that store data in DRAM gradually discharge energy, no energy means the data has been lost.
- So, a periodic refresh of power is required in order to function. DRAM is called dynamic as constant change or action i.e. refreshing is needed to keep the data intact.
- It is used to implement main memory.

Classification of Memory

SRAM	DRAM
Transistors are used to store information in SRAM.	Capacitors are used to store data in DRAM.
Capacitors are not used hence no refreshing is required.	To store information for a longer time, contents of the capacitor needs to be refreshed periodically
SRAM is faster as compared to DRAM.	DRAM provides slow access speeds
These are expensive.	These are cheaper.
SRAMs are low density package devices.	DRAMs are high density package devices.
These are used in cache memories.	These are used in main memories.
Consumes more power	Relatively less power consumption

Classification of Memory

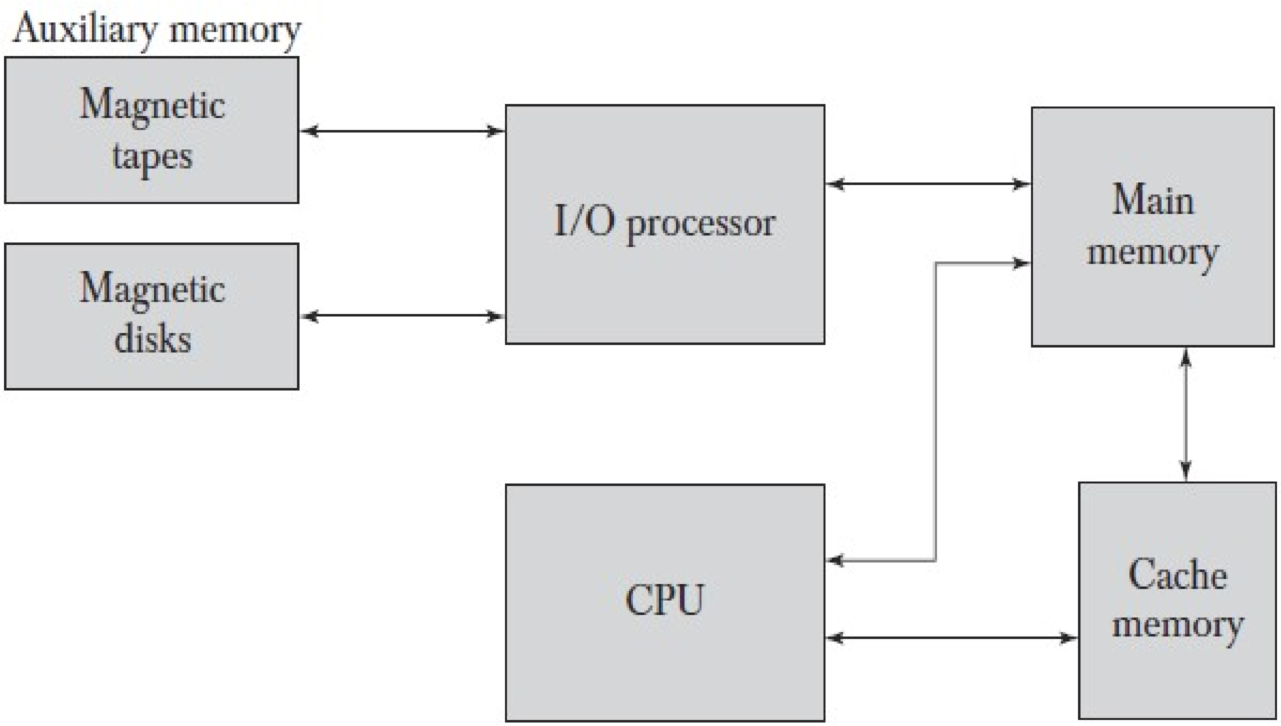
SRAM	DRAM
Interfacing is simpler	Interfacing is complex
More chip area	Less chip area

Memory Hierarchy

- The total memory capacity of a computer can be visualized as being a hierarchy of components.
- The memory hierarchy system consists of all storage devices employed in a computer system from the slow but high-capacity auxiliary memory to a relatively faster main memory, to an even smaller and faster cache memory accessible to the high-speed processing logic.
- Figure 12-1 illustrates the components in a typical memory hierarchy.
- At the bottom of the hierarchy are the relatively slow magnetic tapes used to store removable files.
- Next are the magnetic disks used as backup storage.

Memory Hierarchy

Figure 12-1 Memory hierarchy in a computer system.

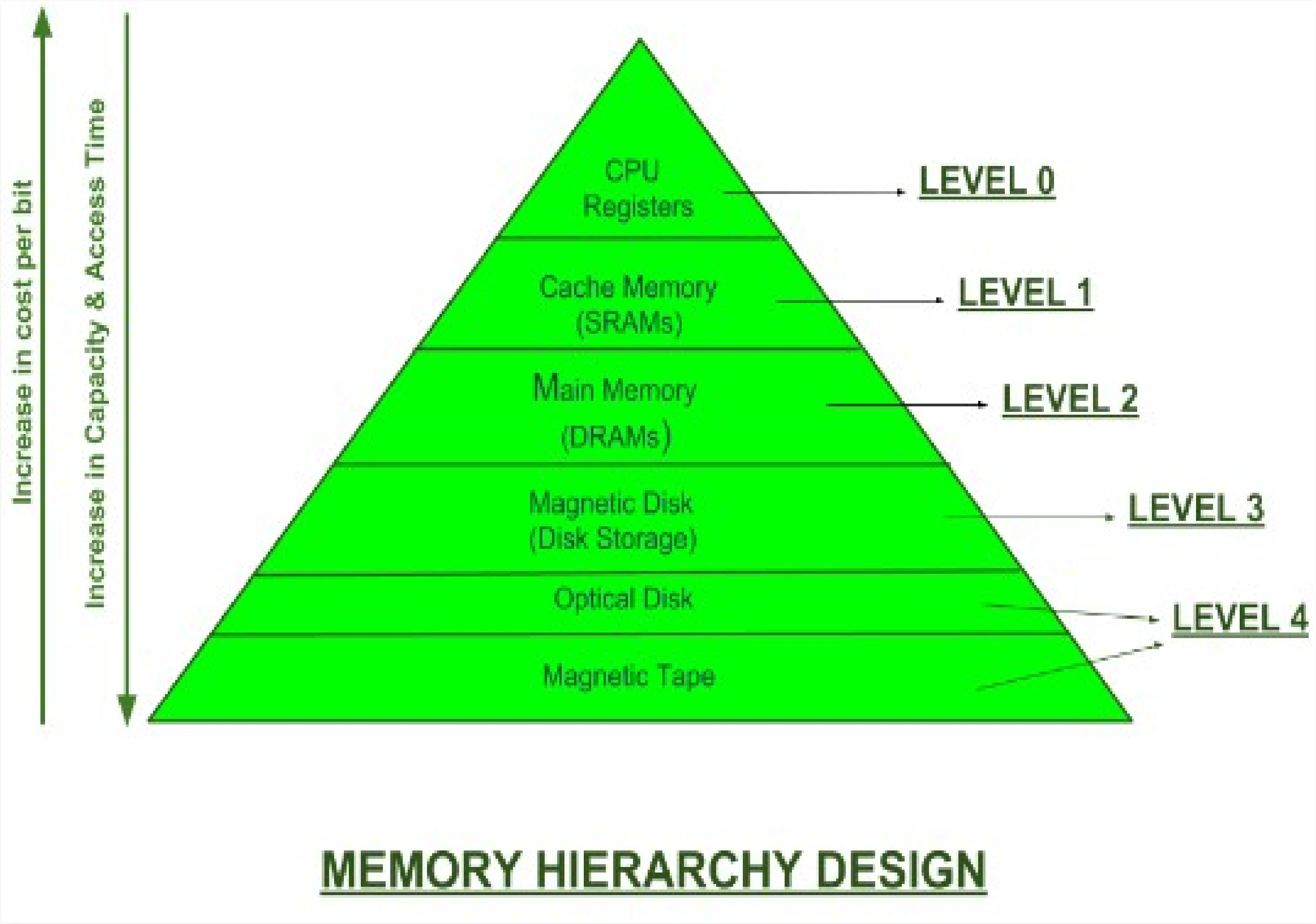


Memory Hierarchy

- In the Computer System Design, **Memory Hierarchy is an enhancement to organize the memory such that it can minimize the access time.**
- The Memory Hierarchy was developed based on a program behaviour known as locality of references.

- The figure below clearly demonstrates the different levels of memory hierarchy :

Memory Hierarchy



Semiconductor RAM

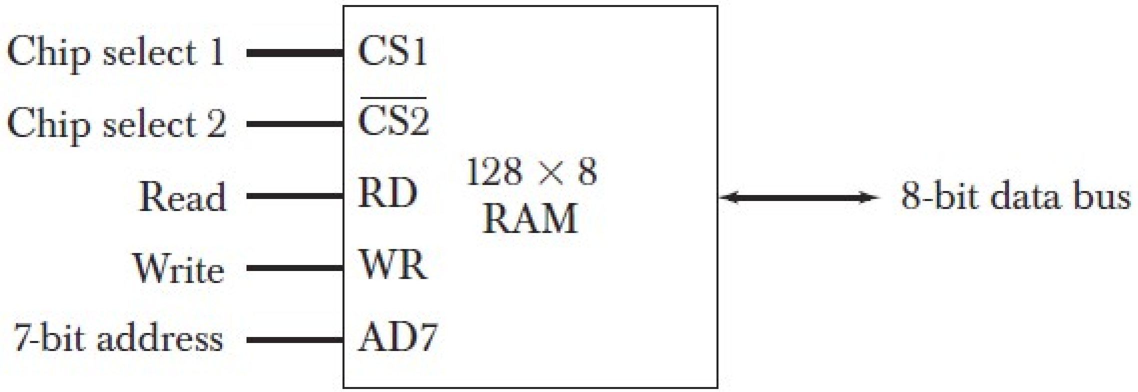
- The principal technology used for the main memory is based on semiconductor integrated circuits.
- RAM and ROM chips are available in a variety of sizes.
- **If the memory needed for the computer is larger than the capacity of one chip, it is necessary to combine a number of chips to form the required memory size.**

- To demonstrate the chip interconnection, we will show an example of a **1024 X 8 memory** constructed with

➢ **128 X 8 RAM chips and**
➢ **512 X 8 ROM chips.**

Semiconductor RAM

Figure 12-2 Typical RAM chip.



(a) Block diagram

CS1	$\overline{CS2}$	RD	WR	Memory function	State of data bus
0	0	×	×	Inhibit	High-impedance
0	1	×	×	Inhibit	High-impedance
1	0	0	0	Inhibit	High-impedance
1	0	0	1	Write	Input data to RAM
1	0	1	×	Read	Output data from RAM
1	1	×	×	Inhibit	High-impedance

(b) Function table

Semiconductor ROM

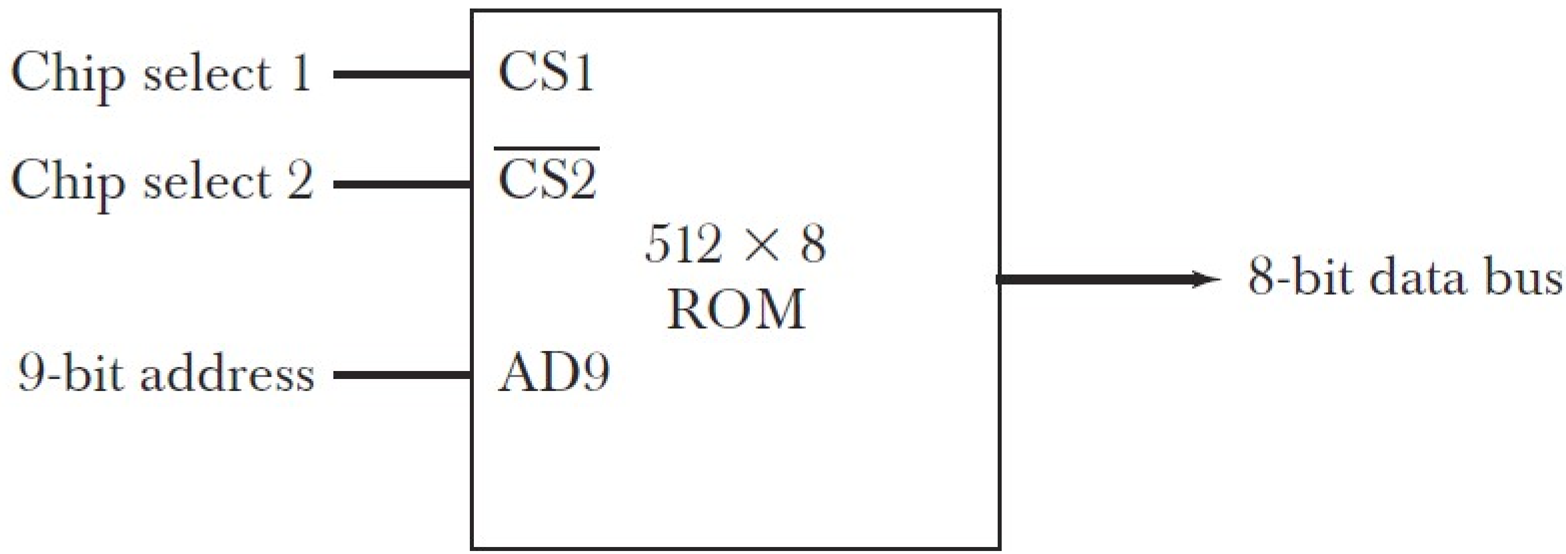


Figure 12-3 Typical ROM chip.

Semiconductor RAM

Memory Address Map

•To demonstrate with a particular example, assume that a computer system needs 512 bytes of RAM and 512 bytes of ROM.

•The memory address map for this configuration is shown in Table 12-1.

TABLE 12-1 Memory Address Map for Microprocomputer

Component	Hexadecimal address	Address bus 10	9	8	7	6	5	4	3	2	1
RAM 1	0000–007F	0	0	0	x	x	x	x	x	x	x
RAM 2	0080–00FF	0	0	1	x	x	x	x	x	x	x
RAM 3	0100–017F	0	1	0	x	x	x	x	x	x	x
RAM 4	0180–01FF	0	1	1	x	x	x	x	x	x	x
ROM	0200–03FF	1	x	x	x	x	x	x	x	x	x

Memory Connection to CPU

- **Memory Connection to CPU**
- RAM and ROM chips are connected to a CPU through the data and address buses.
- The low-order lines in the address bus select the byte within the chips and other lines in the address bus select a particular chip through its chip select inputs.
- The connection of memory chips to the CPU is shown in Fig. 12-4.
- This configuration gives a memory capacity of 512 bytes of RAM and 512 bytes of ROM.
- Each RAM receives the seven low-order bits of the address bus to select one of 128 possible bytes.
- The particular RAM chip selected is determined from lines 8 and 9 in the address bus. This is done through a 2 X 4 decoder whose outputs go to the CS1 inputs in each RAM chip.

Memory Connection to CPU

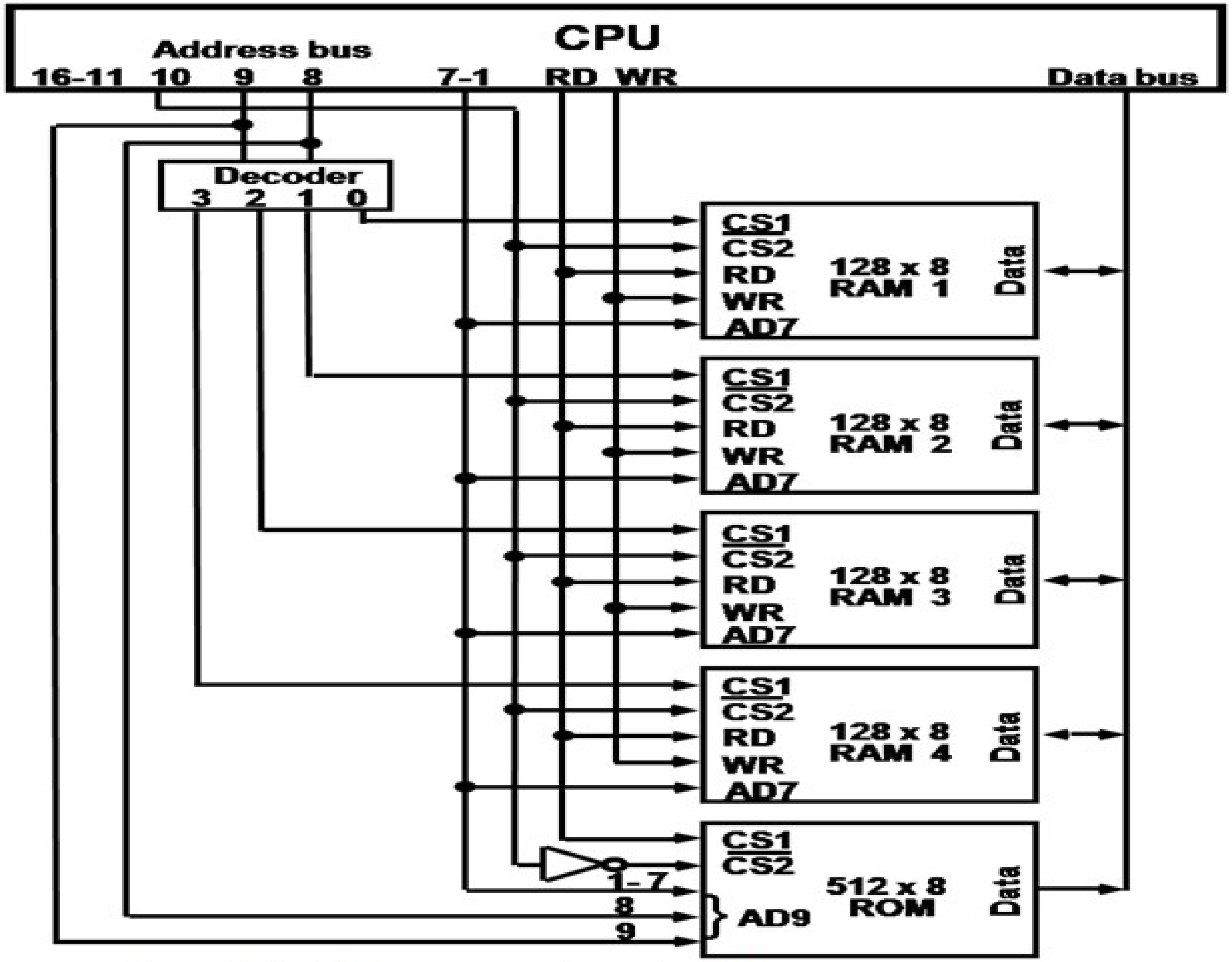
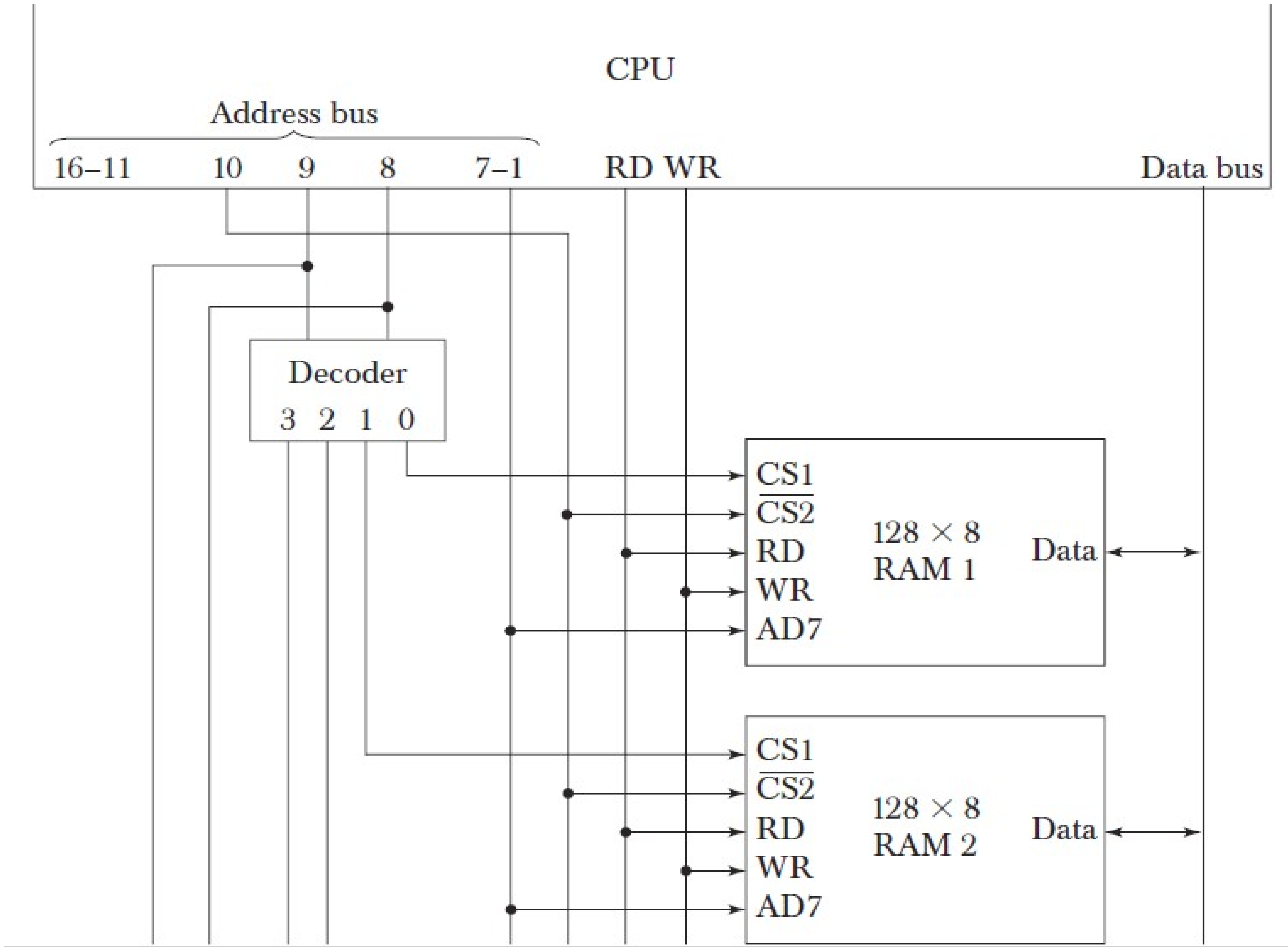


Figure 12.4 Memory connection to the CPU

Memory Connection to CPU



Memory Connection to CPU

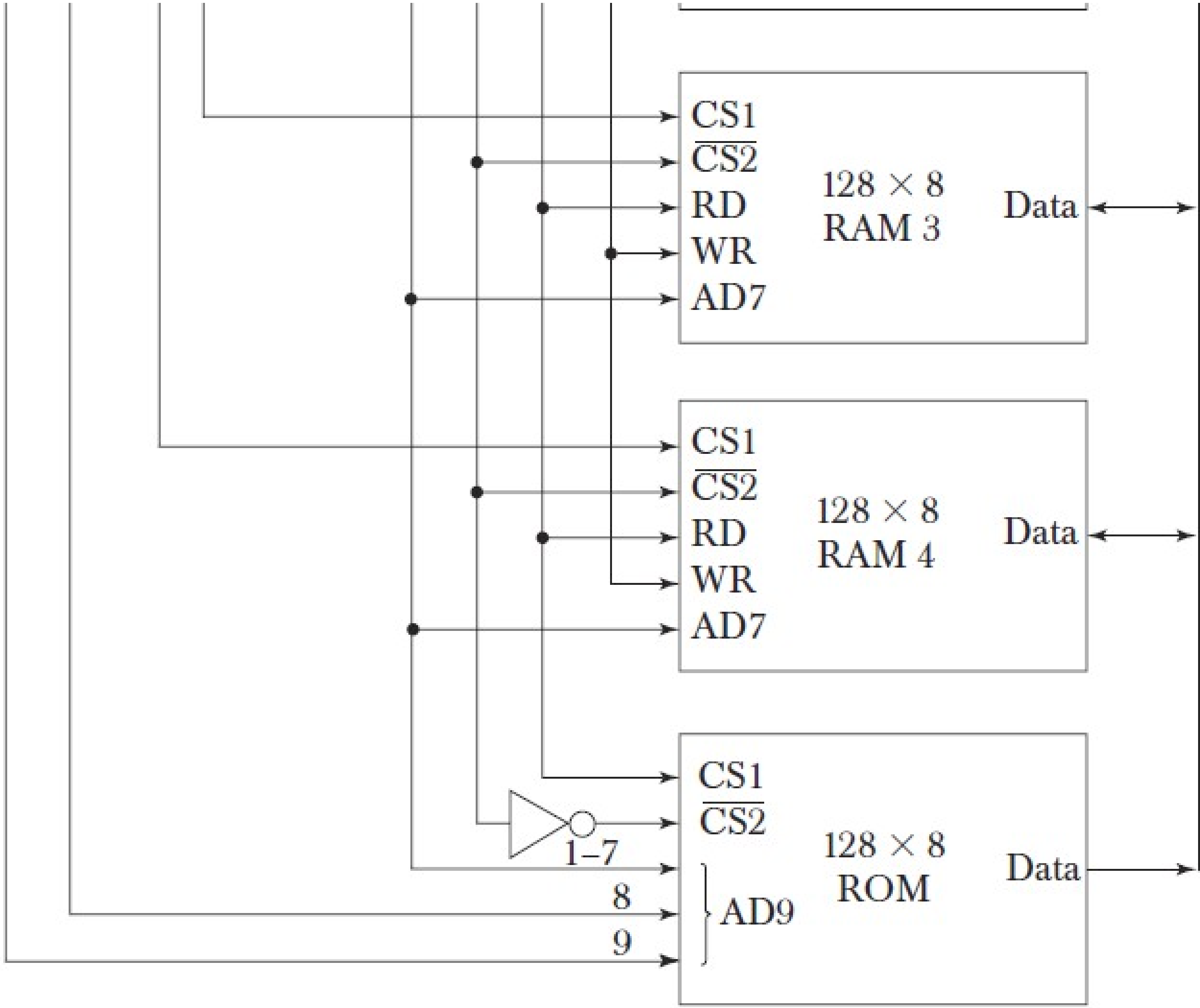


Figure 12-4 Memory connection to the CPU.

Cache Memory

Locality of reference-

- Analysis of a large number of typical programs has shown that the references to memory at any given interval of time tend to be confined within a few localized areas in memory.
- This phenomenon is known as the property of ***locality of reference.***

- Locality of reference property states that over a short interval of time, the addresses generated by a typical program refer to a few localized areas of memory repeatedly, while the remainder of memory is accessed relatively infrequently.

Cache Memory

- If the **active portions of the program and data are placed in a fast small memory, the average memory access time can be reduced,** thus reducing the total execution time of the program.
- Such a fast small memory is referred to as a ***cache memory.***
- The cache memory access time is less than the access time of main memory by a factor of 5 to 10.

- It is placed between the CPU and main memory as illustrated in Fig. 12-1.

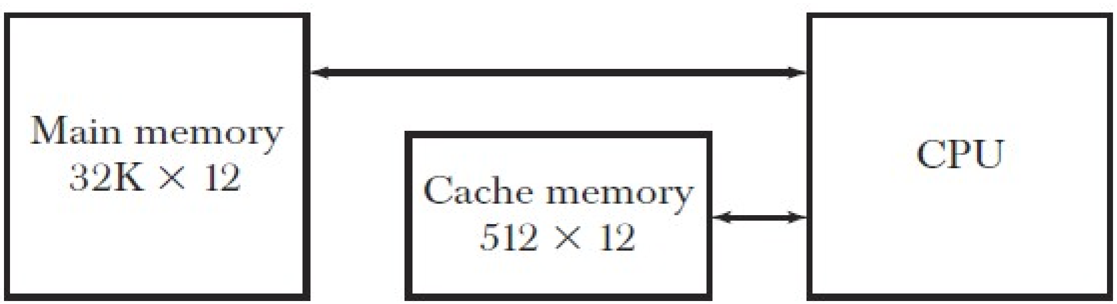


Figure 12-10 Example of cache memory.

Cache Memory

Hit Ratio

- The performance of cache memory is frequently measured in terms of a quantity called ***hit ratio.***

1. When the CPU refers to memory and finds the word in cache, it is said to produce a ***hit.***
2. If the word is not found in cache, it is in main memory and it counts as a ***miss.***

- **The ratio of the number of hits divided by the total CPU references to memory (hits plus misses) is the hit ratio.**

Hit ratio = hit / (hit + miss)

- Hit ratios of 0.9 and higher have been reported.
- This high ratio verifies the validity of the locality of reference property.

Cache Memory

- The basic characteristic of cache memory is its fast access time.

- **The transformation of data from main memory to cache memory is referred to as a *mapping* process.**

- Three types of mapping procedures are of practical interest when considering the organization of cache memory:

1. Associative mapping
2. Direct mapping
3. Set-associative mapping

Cache Memory

- Example of a memory organization as shown in Fig. 12-10.
- The main memory can store 32K words of 12 bits each.
- The cache is capable of storing 512 of these words at any given time.
- For every word stored in cache, there is a duplicate copy in main memory.
- The CPU communicates with both memories.
- It first sends a 15-bit address to cache.

1. If there is a hit, the CPU accepts the 12-bit data from cache.
2. If there is a miss, the CPU reads the word from main memory and the word is then transferred to cache.

Cache Memory

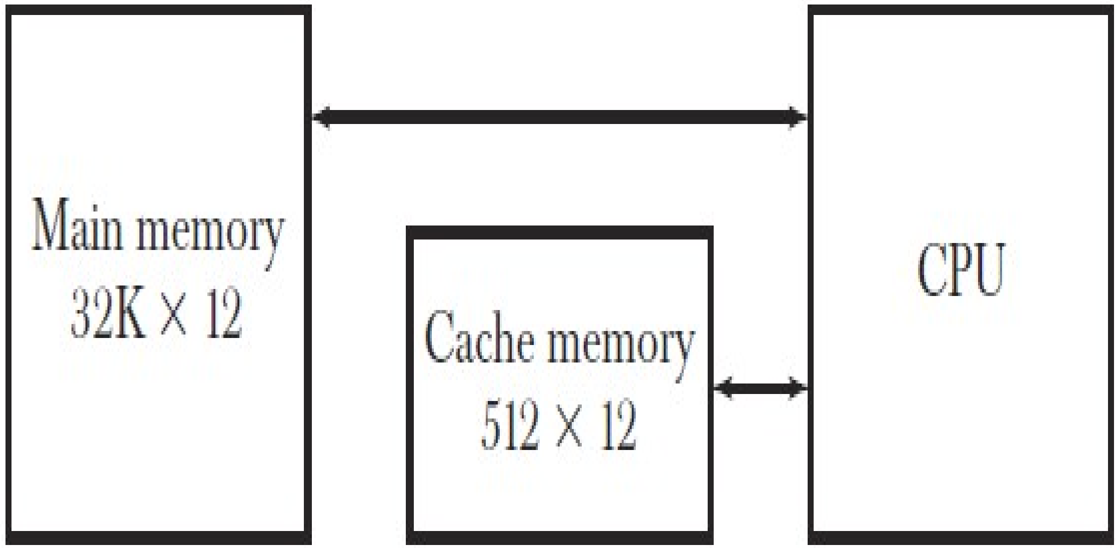
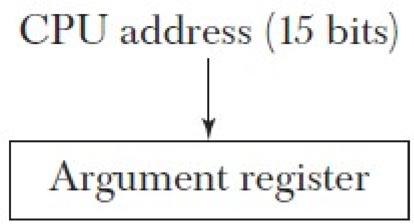


Figure 12-10 Example of cache memory.

Associative Mapping

Figure 12-11 Associative mapping cache (all numbers in octal).



Address	Data
0 1 0 0 0	3 4 5 0
0 2 7 7 7	6 7 1 0
2 2 3 4 5	1 2 3 4

Associative Mapping

- If the address is found, the corresponding 12-bit data is read and sent to the CPU.
- If no match occurs, the main memory is accessed for the word.
- The address—data pair is then transferred to the associative cache memory.
- If the cache is full, an address—data pair must be displaced to make room for a pair that is needed and not presently in the cache.
- The decision as to what pair is replaced is determined from the replacement algorithm that the designer chooses for the cache.
- This constitutes a first-in first-out (FIFO) replacement policy.

- **Disadvantage of associative memories-**

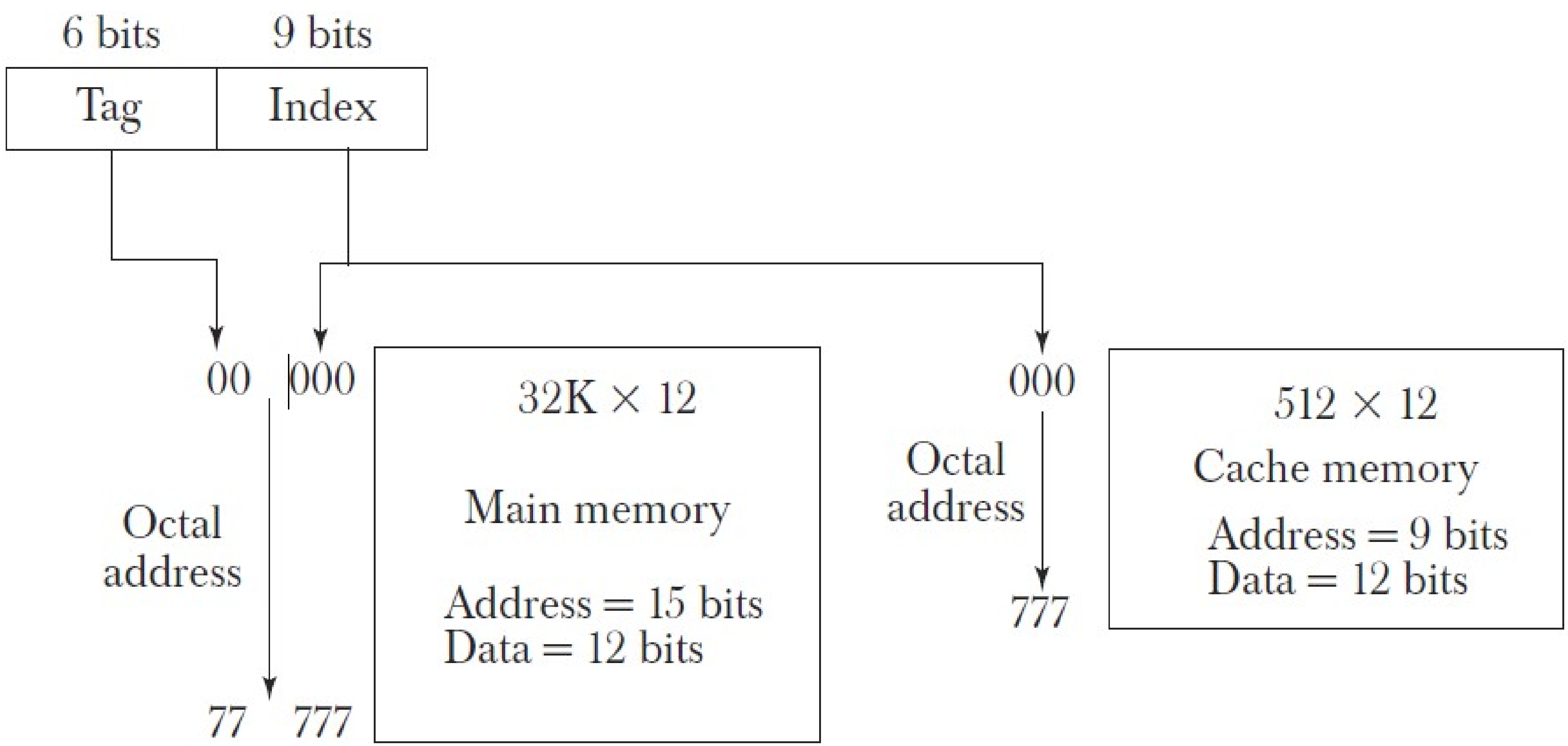
Associative memories are expensive compared to random-access memories because of the added logic associated with each cell.

Direct Mapping

- The possibility of using a random-access memory for the cache is investigated in Fig. 12-12.
- The CPU address of 15 bits is divided into two fields.
- **The nine least significant bits constitute the *index* field** and the **remaining six bits form the *tag* field**.
- The figure shows that main memory needs an address that includes both the tag and the index bits.
- The number of bits in the index field is equal to the number of address bits required to access the cache memory.

Direct Mapping

Figure 12-12 Addressing relationships between main and cache memories.



Direct Mapping

- The internal organization of the words in the cache memory is as shown in Fig. 12-13(b).
- Each word in cache consists of the data word and its associated tag.
- When a new word is first brought into the cache, the tag bits are stored alongside the data bits.
- **When the CPU generates a memory request, the index field is used for the address to access the cache.**
- **The tag field of the CPU address is compared with the tag in the word read from the cache.**
- **If the two tags match, there is a hit and the desired data word is in cache. If there is no match, there is a miss and the required word is read from main memory.**
- It is then stored in the cache together with the new tag, replacing the previous value.

Direct Mapping

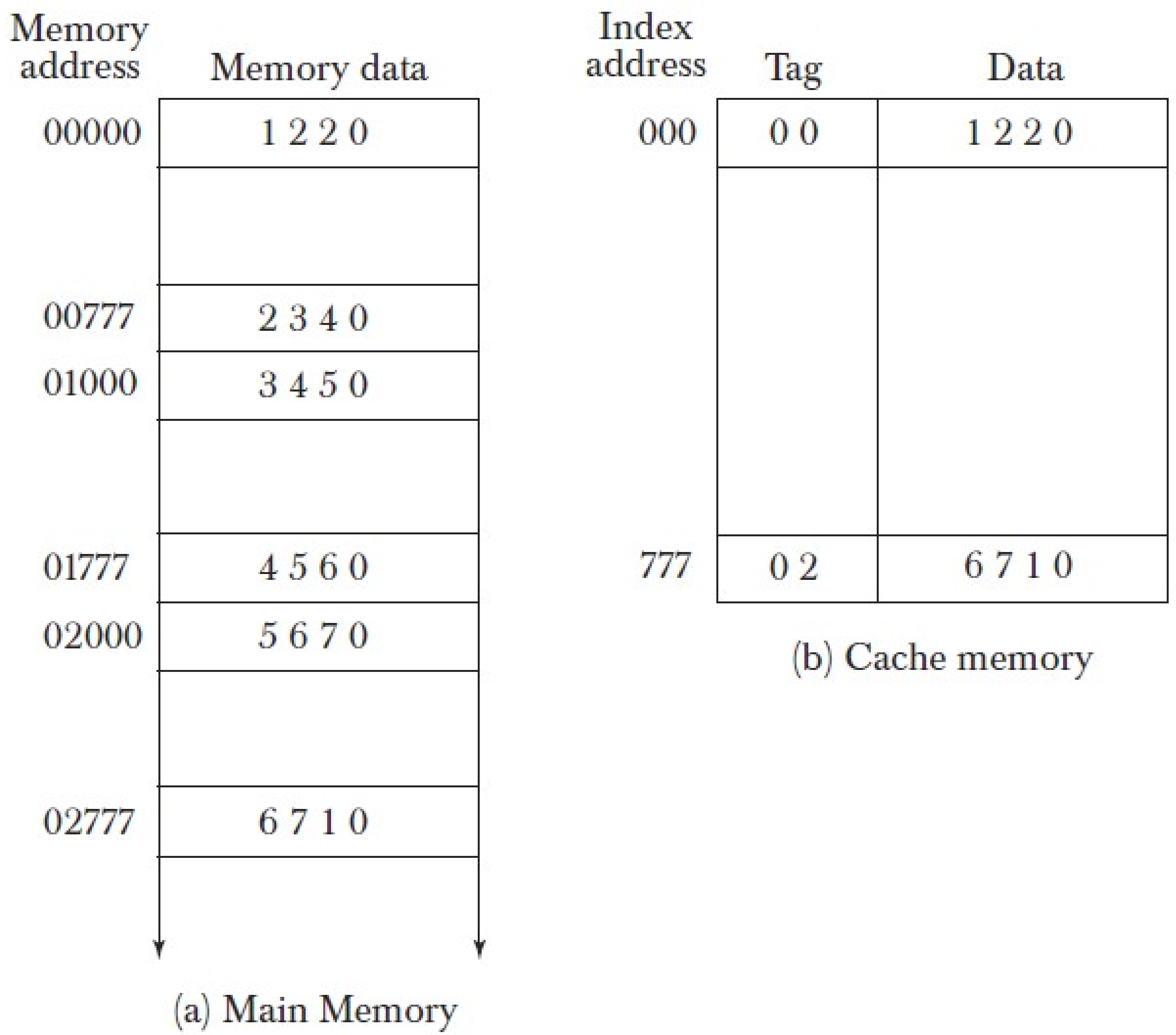


Figure 12-13 Direct mapping cache organization.

Direct Mapping

Disadvantage of direct mapping -

•The hit ratio can drop considerably if two or more words whose addresses have the same index but different tags are accessed repeatedly.

•Two words with the same index in their address but with different tag values cannot reside in cache memory at the same time.

Direct Mapping

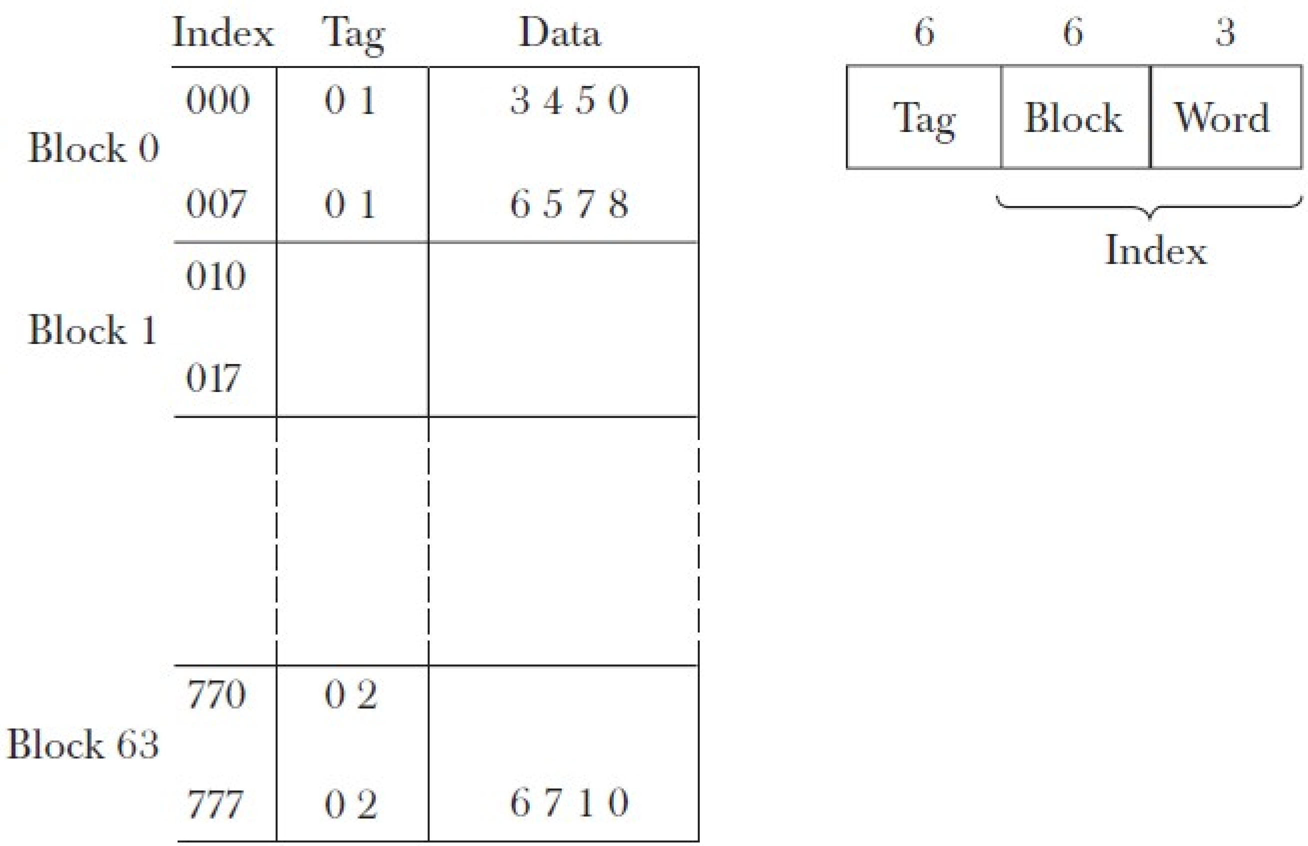


Figure 12-14 Direct mapping cache with block size of 8 words.

Set-Associative Mapping

- Third type of cache organization, called set-associative mapping, is an improvement over the direct mapping organization in that each word of cache can store two or more words of memory under the same index address.
- **Each data word is stored together with its tag and the number of tag—data items in one word of cache is said to form a set**.
- An example of a set-associative cache organization for a set size of two is shown in Fig. 12-15.

Set-Associative Mapping

Index	Tag	Data	Tag	Data
000	0 1	3 4 5 0	0 2	5 6 7 0
777	0 2	6 7 1 0	0 0	2 3 4 0

Figure 12-15 Two-way set-associative mapping cache.

Replacement Algorithms

- When a miss occurs in a set-associative cache and the set is full, it is necessary to replace one of the tag-data items with a new value.

- **The most common replacement algorithms used are: random replacement, first-in, first-out (FIFO), and least recently used (LRU).**

1. With the random replacement policy the control chooses one tag—data item for replacement at random.
2. The FIFO procedure selects for replacement the item that has been in the set the longest.
3. The LRU algorithm selects for replacement the item that has been least recently used by the CPU.

- Both FIFO and LRU can be implemented by adding a few extra bits in each word of cache.

Writing into Cache

Write-Through

•The simplest and most commonly used procedure is to update main memory with every memory write operation, with cache memory being updated in parallel if it contains the word at the specified address.

•This is called the *write-through* method.

•This method has the **advantage that main memory always contains the same data as the cache.**

•This characteristic is important in systems with direct memory access transfers.

•It ensures that the data residing in main memory are valid at all times so that an I/O device communicating through DMA would receive the most recent updated data.

Writing into Cache

Write-Back

•**In this method only the cache location is updated during a write operation**.

•The location is then marked by a flag so that later when the word is removed from the cache it is copied into main memory.

•The reason for the write-back method is that during the time a word resides in the cache, it may be updated several times; however, as long as the word remains in the cache, it does not matter whether the copy in main memory is out of date, since requests from the word are filled from the cache.

•It is only when the word is displaced from the cache that an accurate copy need be rewritten into main memory.

Virtual Memory

- **Virtual memory is a storage allocation scheme in which secondary memory can be addressed as though it were part of the main memory.**
- Virtual memory is a concept used in some large computer systems that permit the user to construct programs as though a large memory space were available, equal to the totality of auxiliary memory.
- The addresses a program may use to reference memory are distinguished from the addresses the memory system uses to identify physical storage.
- Each address that is referenced by the CPU goes through an address mapping from the so-called virtual address to a physical address in main memory.

Virtual Memory- Address Space and Memory Space

- **An address used by a programmer will be called a *virtual address.***
- The set of such addresses the ***address space.***
- **An address in main memory is called a *location* or *physical address*.**
- The set of such locations is called the ***memory space.***

- The address space is allowed to be larger than the memory space in computers with virtual memory

Address Space and Memory Space

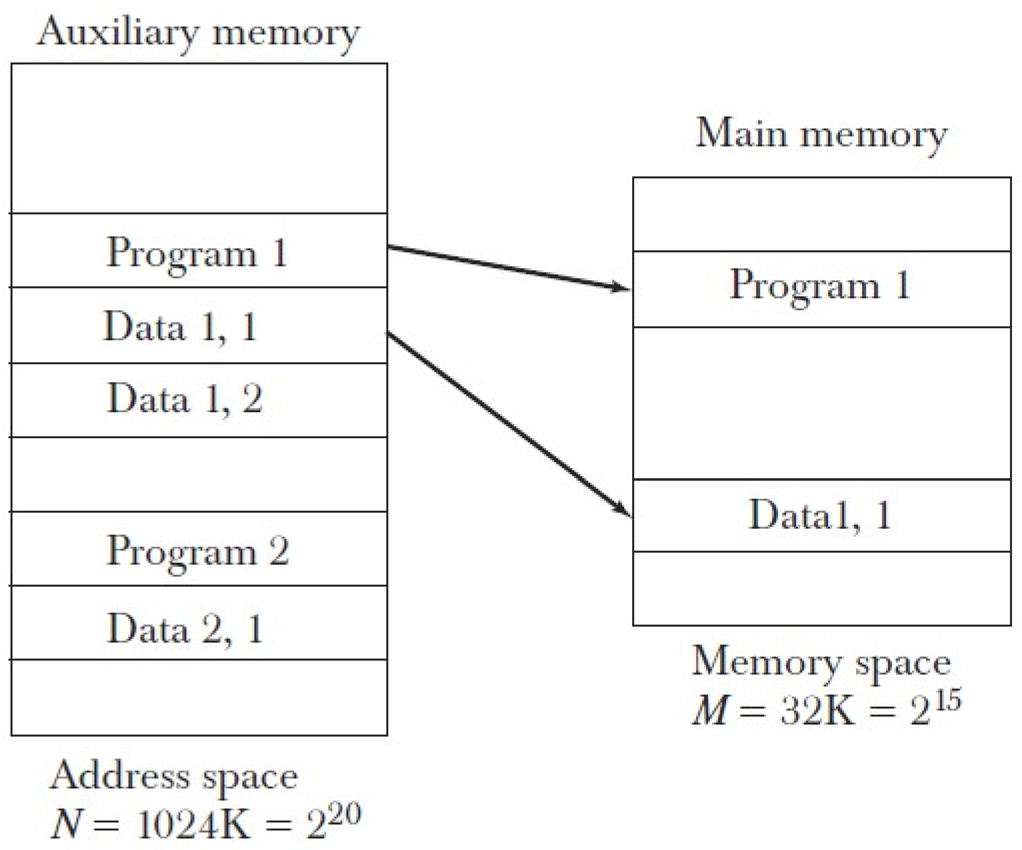
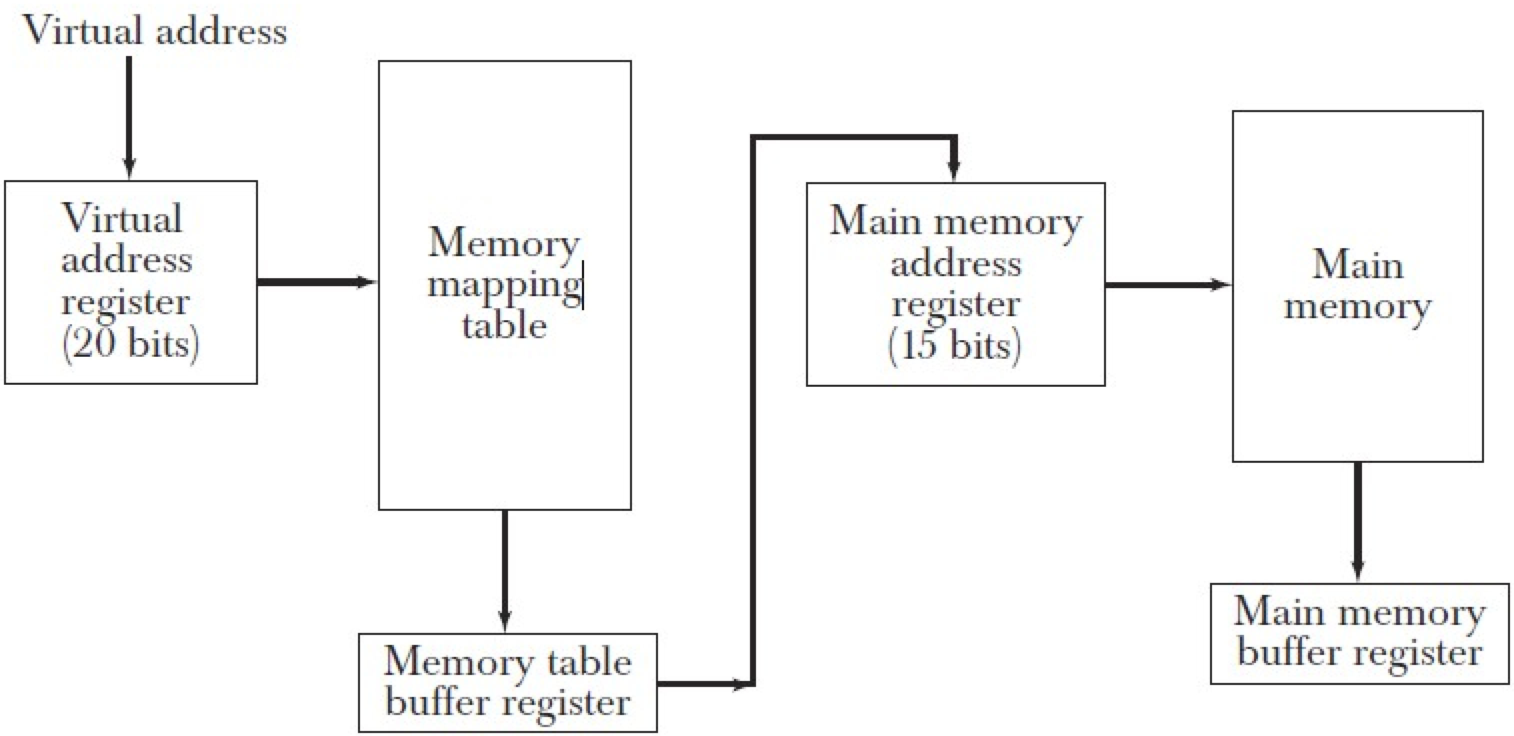


Figure 12-16 Relation between address and memory space in a virtual memory system.

Virtual Memory-Mapping

Figure 12-17 Memory table for mapping a virtual address.



Address Mapping Using Pages

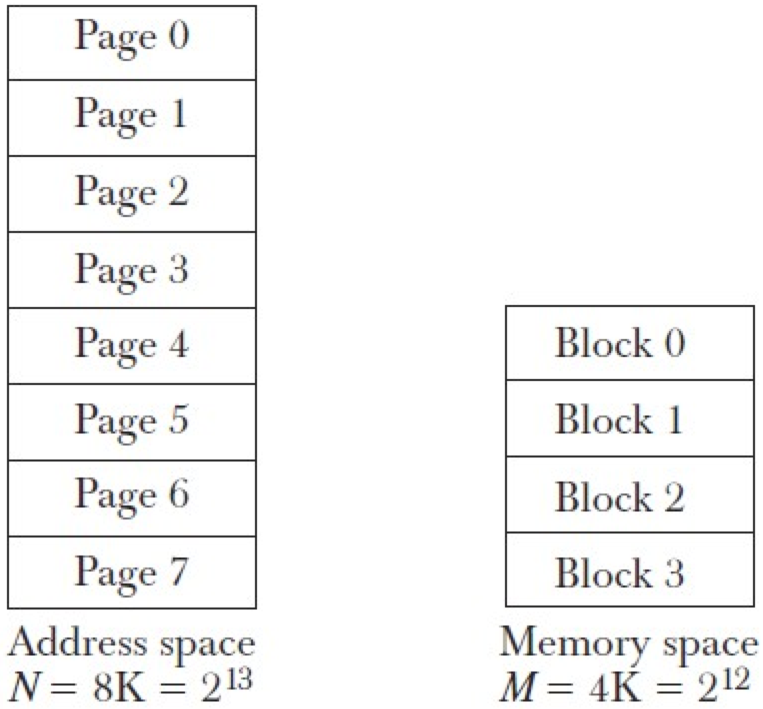
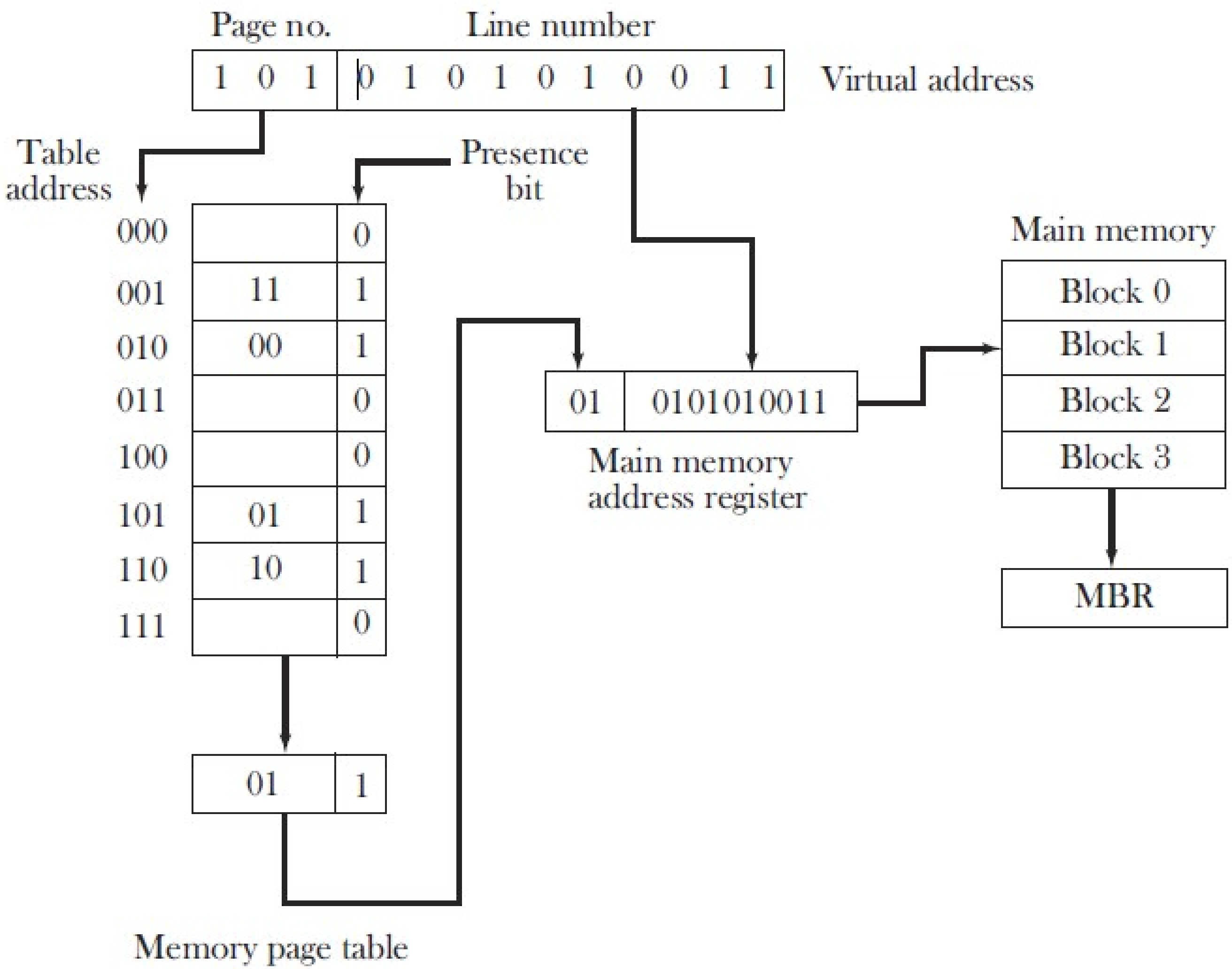


Figure 12-18 Address space and memory space split into groups of 1K words.

Address Mapping Using Pages

Figure 12-19 Memory table in a paged system.



Virtual Memory-Page fault

Page fault

•When a program starts execution, one or more pages are transferred into main memory and the page table is set to indicate their position.

•The program is executed from main memory until it attempts to reference a page that is still in auxiliary memory.

•This condition is called *page fault.*

•When page fault occurs, the execution of the present program is suspended until the required page is brought into main memory

Virtual Memory-Page Replacement

- When a page fault occurs in a virtual memory system, it signifies that the page referenced by the CPU is not in main memory.

- A new page is then transferred from auxiliary memory to main memory. If main memory is full, it would be necessary to remove a page from a memory block to make room for the new page.

- Two of the most common replacement algorithms used are-

1. ***First-in first-out* (FIFO) and**
2. **L*east recently used* (LRU)**

Organization of Memory Chips

- The **internal structure** of Memory either RAM or ROM is made up of memory cells that contain a memory bit.
- A group of 8 bits makes a byte.
- The memory is in the form of a multidimensional array of rows and columns.
- In which, each cell stores a bit and a complete row contains a word.
- A memory simply can be divided into this below form.

$$2^n = N$$

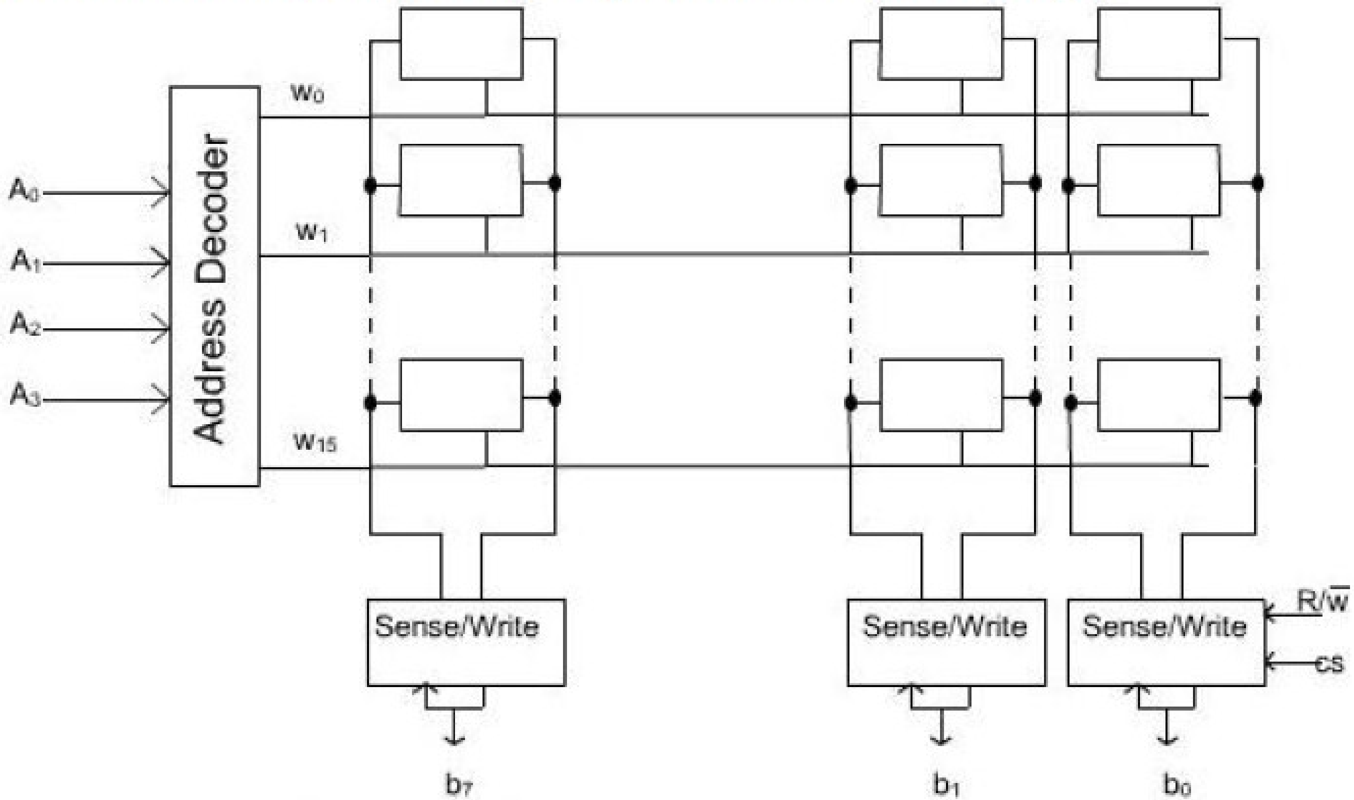
- Where , n is the no. of address lines and N is the total memory in bytes.

  There will be 2^n words.

Organization of Memory Chips

Internal Organization of Memory Chips

A memory cell is capable of storing 1-bit of information. A number of memory cells are organized in the form of a matrix to form the memory chip. One such organization is shown in the Figure.

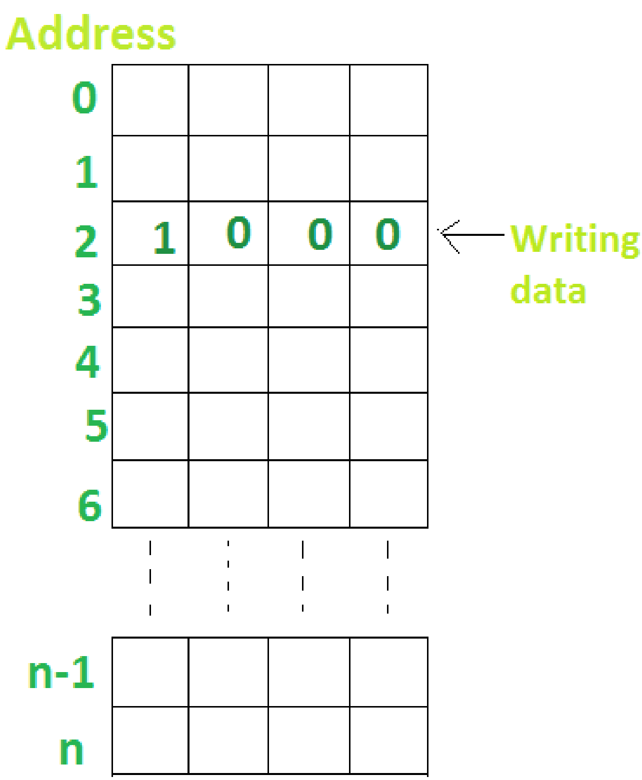


16 memory location w_0, w_1,..................w_{15}

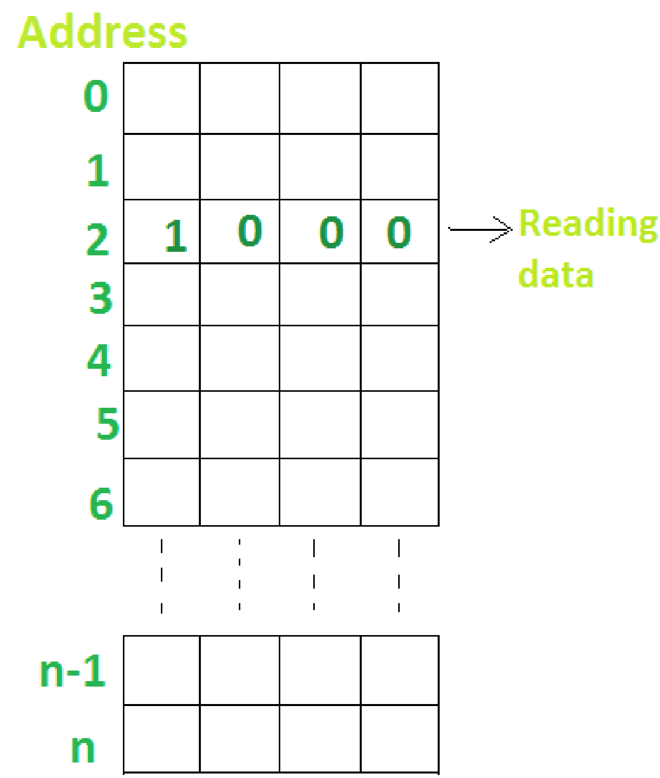
8 bits in each location b_0, b_1, ...b_7

Figure : 16 X 8 Memory Organization

Organization of Memory Chips



A)Write operation



B)Read Operation

2D and 2.5D Memory organization

2D Memory organization

•In 2D organization memory is divides in the form of rows and columns. Each row contains a word now in this memory organization there is a decoder.

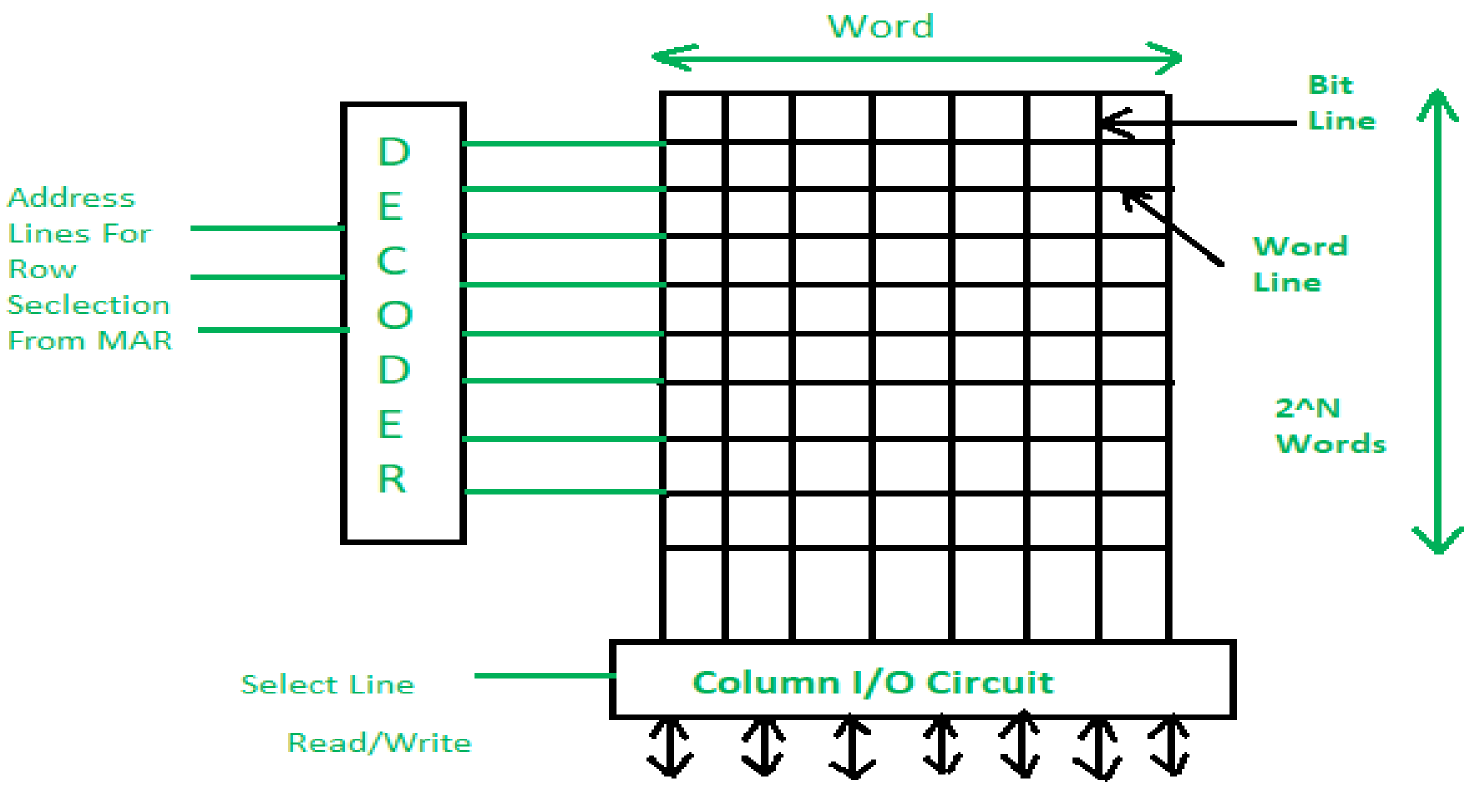
•A decoder is a combinational circuit which contains n input lines and 2^n output lines.

•One of the output line will select the row which address is contained in the MAR.

•The word which is represented by the row that will get selected and either read or write through the data lines.

2D and 2.5D Memory organization

2D Memory organization



2D Memory Organization

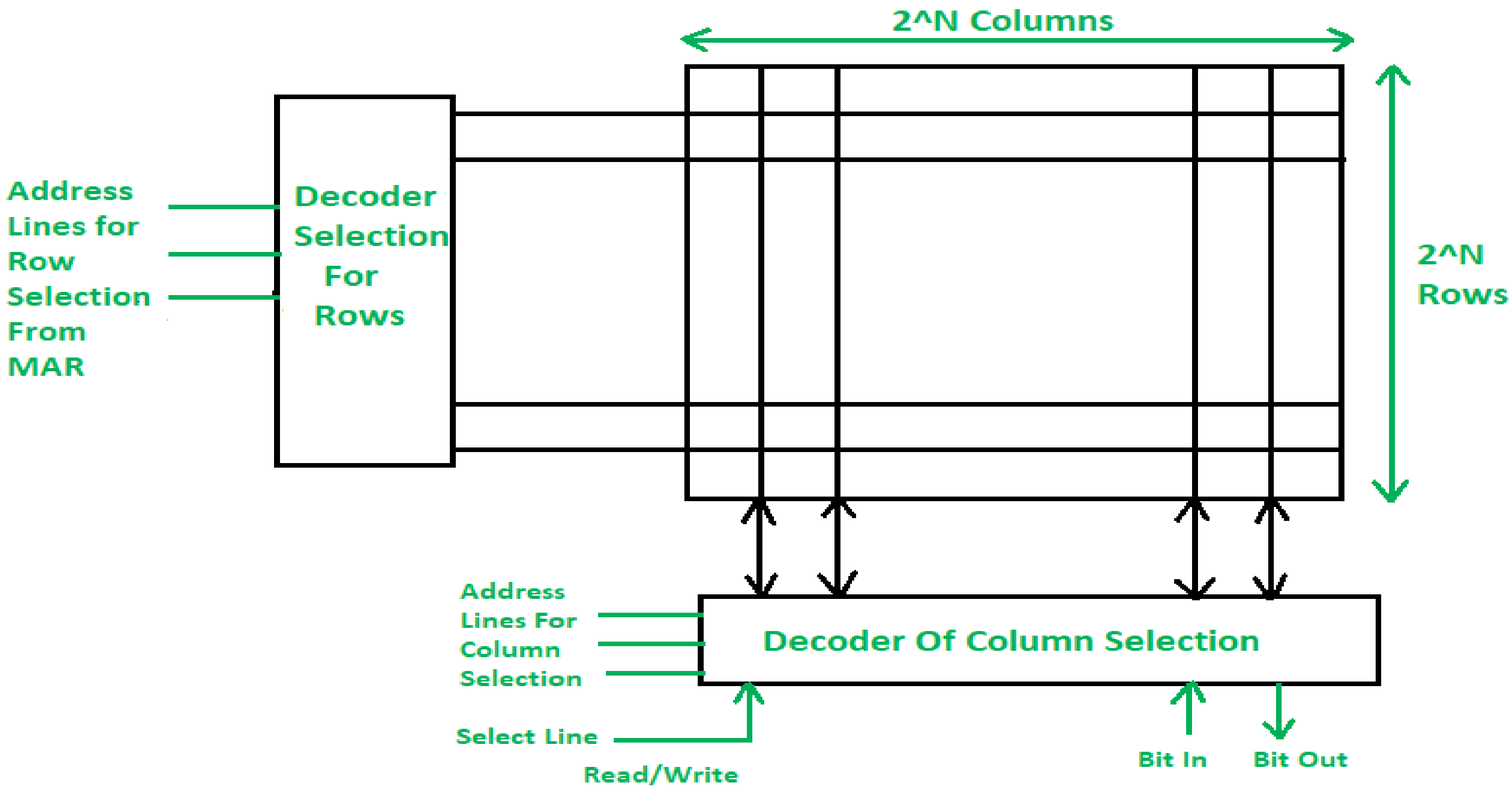
2.5 D Memory organization

- In 2.5D Organization the scenario is the same but we have two different decoders one is column decoder and another is row decoder.

- Column decoder used to select the column and row decoder is used to select the row. Address from the MAR will go in decoders' input.

- Decoders will select the respective cell. Through the bit outline, the data from that location will be read or through the bit in line data will be written at that memory location.

2.5D Memory organization



2.5D Memory Organization

2D and 2.5D Memory organization

Read and Write Operations

•If the select line is in Read mode then the Word/bit which is represented by the MAR that will be coming out to the data lines and get read.

•If the select line is in write mode then the data from memory data register (MDR) will go to the respective cell which is addressed by the memory address register (MAR).

•With the help of the select line the data will get selected where the read and write operations will take place.

2D and 2.5D Memory organization

Comparison between 2D & 2.5D Organizations

- In 2D organization hardware is fixed but in 2.5D hardware changes.
- 2D is more complex in comparison to the 2.5D Organization.
- Error correction is not possible in the 2D organization but In 2.5D error correction is easy.
- 2D is more difficult to fabricate in comparison to the 2.5D organization.

Auxiliary Memories

Auxiliary memory (also referred to as *secondary storage*)

•It is the non-volatile memory lowest-cost, highest-capacity, and slowest-access storage in a computer system.

•It is where programs and data kept for long-term storage or when not in immediate use. It is not directly accessible by the CPU.

For example:

•Magnetic disks and tapes, Optical Disk.

Secondary Storage Media

There are the following main types of storage media:

1.Magnetic Storage media

- **a) Magnetic Disks**
  - **i. Floppy Disk**
  - **ii. Hard Disk**
- **b) Magnetic Tape**

2.Optical Storage media

Auxiliary Memories

- The physical properties of these storage devices can be quite complex, their logical properties can be characterized and compared by a few parameters.
- The important characteristics of any device are its access mode, access time, transfer rate, capacity, and cost.

1. Access Time-

- The **average time required to reach a storage location in memory and obtain its contents is called the access time.**

2. Seek time –

- In electromechanical devices with moving parts such as disks and tapes, the access time consists of a *seek* time.
- It is **time required to position the read-write head to a location**.

Auxiliary Memories

3. Transfer Time-

Time required to transfer data to or from the device.

- Because the seek time is usually much longer than the transfer time, auxiliary storage is organized in records or blocks. A record is a specified number of characters or words.
- Reading or writing is always done on entire records.

4. Transfer Rate-

It is the number of characters or words that the device can transfer per second, after it has been positioned at the beginning of the record.

Auxiliary Memories

- Magnetic drums and disks are quite similar in operation. Both consist of high-speed rotating surfaces coated with a magnetic recording medium.
- The recording surface rotates at uniform speed and is not started or stopped during access operations.
- **Bits are recorded as magnetic spots on the surface as it passes a stationary mechanism called a *write head.***
- **Stored bits are detected by a change in magnetic field produced by a recorded spot on the surface as it passes through a *read head.***
- The amount of surface available for recording in a disk is greater than in a drum of equal physical size.
- For this reason, disks have replaced drums in more recent computers.

Auxiliary Memories

Magnetic Disks

- A magnetic disk is a circular plate constructed of metal or plastic coated with magnetized material.
- Often both sides of the disk are used and several disks may be stacked on one spindle with read/write heads available on each surface.
- All disks rotate together at high speed and are not stopped or started for access purposes.
- **Bits are stored in the magnetized surface in spots along concentric circles called tracks**.
- **The tracks are commonly divided into sections called sectors.**
- **The minimum quantity of information which can be transferred is a sector**.
- The subdivision of one disk surface into tracks and sectors is shown in Fig. 12-5.

Auxiliary Memories

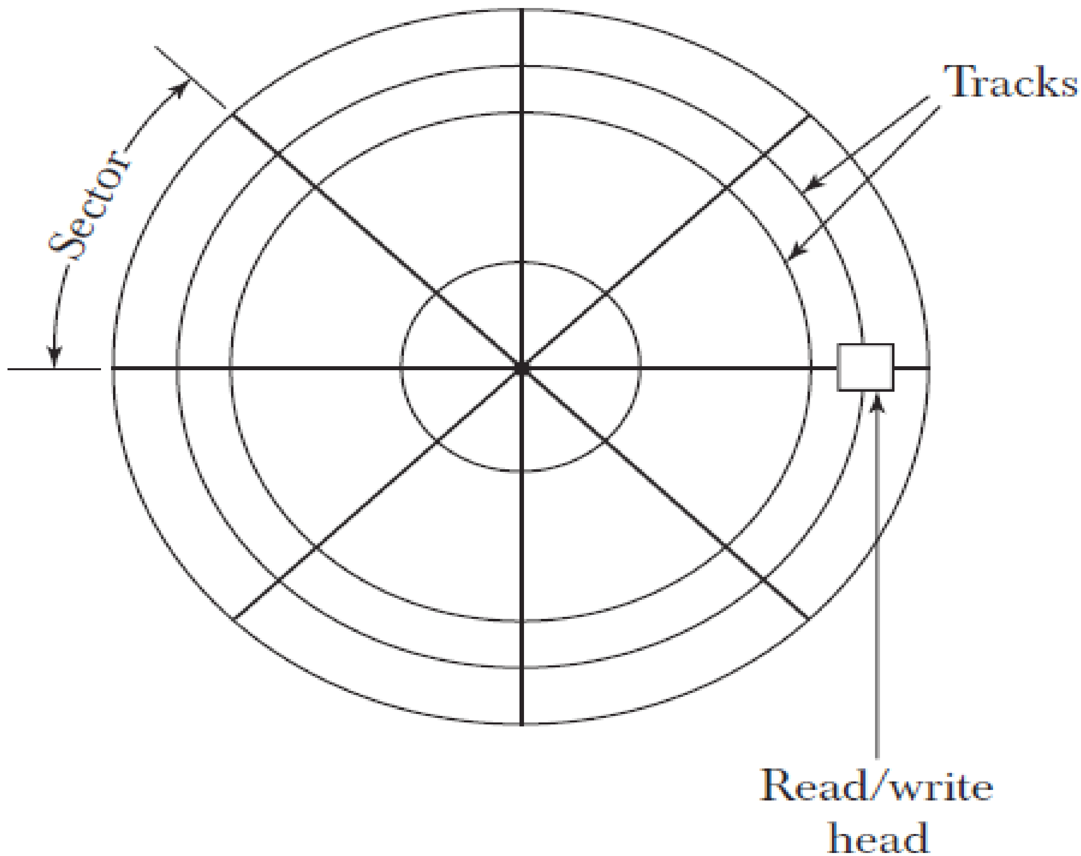


Figure 12-5 Magnetic disk.

Auxiliary Memories

- **A disk system is addressed by address bits that specify the disk number, the disk surface, the sector number and the track within the sector.**
- After the read/write heads are positioned in the specified track, the system has to wait until the rotating disk reaches the specified sector under the read/write head.
- Information transfer is very fast once the beginning of a sector has been reached.
- Disks may have multiple heads and simultaneous transfer of bits from several tracks at the same time.

Auxiliary Memories

- **Disks that are permanently attached to the unit assembly and cannot be removed by the occasional user are called *hard disks*.**
- **A disk drive with removable disks is called a *floppy disk.***
- The disks used with a floppy disk drive are small removable disks made of plastic coated with magnetic recording material.
- There are two sizes commonly used, with diameters of 5.25 and 3.5 inches.
- The 3.5-inch disks are smaller and can store more data than can the 5.25-inch disks.

Auxiliary Memories

MAGNETIC DISK	MAGNETIC TAPE
In magnetic disk for data recording, magnetic material is coated on only both side of the platters.	In magnetic tape for data recording, magnetic material is coated on only one side of the tape.
Magnetic tape contains reels of tape which is in form of strip of plastic.	Magnetic disk contains round platters which is made up of plastic or metaL
The cost of magnetic tape is less.	The cost of magnetic disk is high
Reliability of magnetic tape is less.	Reliability of magnetic disk is more.

Auxiliary Memories

MAGNETIC DISK	MAGNETIC TAPE
Access time for magnetic tape is more.	Access time for magnetic disk is less
Data transfer rate for magnetic tape is comparatively less.	Data transfer rate for magnetic disk is more
Magnetic tape is used for backups.	Magnetic disk is used as a secondary storage.

Auxiliary Memories

Magnetic Tape

- A magnetic tape transport consists of the electrical, mechanical, and electronic components to provide the parts and control mechanism for a magnetic-tape unit.
- **The tape itself is a strip of plastic coated with a magnetic recording medium.**
- **Bits are recorded as magnetic spots on the tape along several tracks.**
- Usually, seven or nine bits are recorded simultaneously to form a character together with a parity bit.
- **Read/write heads are mounted one in each track so that data can be recorded and read as a sequence of characters.**
- Magnetic tape units can be stopped, started to move forward or in reverse, or can be rewound.
- However, they cannot be started or stopped fast enough between individual characters.
- For this reason, information is recorded in blocks referred to as records.

Auxiliary Memories

- **Each record on tape has an identification bit pattern at the beginning and end.**
- By reading the bit pattern at the beginning, the tape control identifies the record number.
- By reading the bit pattern at the end of the record, the control recognizes the beginning of a gap.
- A tape unit is addressed by specifying the record number and the number of characters in the record.
- Records may be of fixed or variable length.

Auxiliary Memories

Optical Disk

- An optical disk is any computer disk that uses optical storage techniques and technology to read and write data.
- It is a storage device in which optical (light) energy is used.
- It is a computer storage disk that stores data digitally and uses laser beams to read and write data.
- It uses the optical technology in which laser light is centred to the spinning disks

Auxiliary Memories

MAGNETIC DISK	OPTICAL DISK
Media type used is Multiple fixed disk	Media type used is Single removable disk
Intermediate signal to noise ratio	Excellent signal to noise ratio
Sample rate is Low	Sample rate is High
Implemented where data is randomly accessed.	Implemented in streaming files.
Only one disk can be used at a time	Mass replication is possible
Tracks in the magnetic disk are generally circular	In optical disk the tracks are constructed spirally.

Memory System Performance

- Memory system, and not processor speed, is often the bottleneck for many applications.
- Memory system performance is largely captured by two parameters, latency and bandwidth.
- Latency is the time from the issue of a memory request to the time the data is available at the processor.
- Bandwidth is the rate at which data can be pumped to the processor by the memory system.

Memory System Performance

Memory bandwidth -

- **Memory bandwidth is the rate at which data can be read from or stored into a semiconductor memory by a processor.**

- Memory bandwidth is usually **expressed in units of bytes/second.**

- Memory bandwidth that is advertised for a given memory or system is usually the maximum theoretical bandwidth.

- In practice the observed memory bandwidth will be less than (and is guaranteed not to exceed) the advertised bandwidth.

- A variety of computer benchmarks exist to measure sustained memory bandwidth using a variety of access patterns.

Memory System Performance

Memory Latency -

- **Memory latency is the time (the latency) between initiating a request for a byte or word in memory until it is retrieved by a processor.**
- If the data are not in the processor's cache, it takes longer to obtain them, as the processor will have to communicate with the external memory cells.
- Latency is therefore a fundamental measure of the speed of memory.
- Less the latency, the faster the reading operation.
- Latency should not be confused with memory bandwidth, which measures the throughput of memory.
- **Latency can be expressed in clock cycles** or in time measured in nanoseconds.

Auxiliary Memories

Auxiliary memory (also referred to as *secondary storage*) is the non-volatile memory lowest-cost, highest-capacity, and slowest-access storage in a computer system.
It is where programs and data kept for long-term storage or when not in immediate use. It is not directly accessible by the CPU. **For example:** Magnetic disks and tapes, Optical Disk.

Secondary Storage Media
There are the following main types of storage media:

1. Magnetic storage media:

Magnetic media is coated with a magnetic layer which is magnetized in clockwise or anticlockwise directions. When the disk moves, the head interprets the data stored at a specific location in binary 1s and 0s at reading.
Examples: hard disks, floppy disks and magnetic tapes.

Auxiliary Memories

Floppy Disk: A floppy disk is a flexible disk with a magnetic coating on it. It is packaged inside a protective plastic envelope.

Hard disk: A hard disk consists of one or more circular disks called platters which are mounted on a common spindle. Each surface of a platter is coated with a magnetic material.

2. Optical storage media

In optical storage media information is stored and read using a laser beam. The data is stored as a spiral pattern of pits and ridges denoting binary 0 and binary 1.

Examples: CDs and DVDs

Compact Disk: A Compact Disc drive(CDD) is a device that a computer uses to read data that is encoded digitally on a compact disc(CD). A compact disk or CD can store approximately 650 to 700 megabytes of data.

Auxiliary Memories

DVD:

It stands for Digital Versatile Disk or Digital Video Disk. It looks just like a CD and use a similar technology as that of the CDs but allows tracks to be spaced closely enough to store data that is more than six times the CD's capacity.

It is a significant advancement in portable storage technology. A DVD holds 4.7 GB to 17 GB of data.

Blue Ray Disk:

This is the latest optical storage media to store high definition audio and video. It is similar to a CD or DVD but can store up to 27 GB of data on a single layer disk and up to 54 GB of data on a dual layer disk.

While CDs or DVDs use red laser beam, the blue ray disk uses a blue laser to read/write data on a disk.

Auxiliary Memories

Example of Auxiliary Memories

DVD

CD

Hard Disk

Faculty Video Links, You tube Courses Details

You tube/other Video Links

- https://www.youtube.com/watch?v=NVUWlO5zsk0
- https://www.youtube.com/watch?v=zwovvWfkuSg
- https://www.youtube.com/watch?v=m1dA7D6c3C0
- https://www.youtube.com/watch?v=o2_iCzS9-ZQ
- https://www.youtube.com/watch?v=pJ6qrCB8pDw&list=PLIY8eNdw5tW-BxRY0yK3fYTYVqytw8qhp

MCQ

1. In how many categories memory/storage is classified – three (3)
2. When power is switched off which memory loses its data - Volatile Memory
3. What is the formula for Hit Ratio - Hit/(Hit + Miss)
4. The fastest data access is provided using – Registers
5. Property of locality of reference may fail, if a program has –many operand.
6. memory refreshing may be done -by an external refresh controller
7. The BOOT sector files of the system are stored in- ROM
8. 1k* 8 ROM chips which are required to built a 16K*8 memory system are of number – 16
9. Computer memory is made up of – cell.
10. Time for replacing the block from memory, is referred to as – miss penalty

MCQ

11. In how many categories memory/storage is classified – three (3)
12. The main memory of a computer can act as a – Virtual memory
13. Main purpose of computer secondary storage device is to calculate data – False
14. The spatial locality is also known as - locality in space
15. In DRAM, state of R/W during read operation will be - high

Daily Quiz

- Sketch memory hierarchy .
- Sketch 2.5D RAM organization.
- Write some differences between SRAM and DRAM.
- What do mean by virtual memory in COA.
- Which page replacement algorithm is good & why ?

Weekly Assignment

- Explain memory hierarchy with suitable diagram.
- Analysis different mapping scheme of cache memory.
- Discuss 2 D RAM and 2+1/2D RAM with suitable diagram.
- Write short notes on-
  a) Cache Memory b) Associative Memory c) Auxiliary Memory
- Explain the method to improve the performance of cache memory.

MCQ

1. Size of the ____memory mainly depends on the size of the address bus.
   a) Main b) Virtual c) Secondary d) Cache
2. What is the location of the internal registers of CPU?
   a) Internal b) On-chip c) External d) Motherboard
3. MAR stands for ___________
4. Which of the following is non-volatile storage?
   a) Backup b) Secondary c) Primary d) Cache
5. Which of the following is used in main memory?
   a) SRAM b) DRAM c) PRAM d) DDR
6. RAID stands for __________

Solution a , b, Memory address register., b , b , Redundant array of independent disks

Summary

In previous slides we discuss in details
Memory Unit:

- Basic concept and hierarchy
- semiconductor RAM memories
- 2D & 2 1/2D memory organization
- ROM memories
- Cache memories: concept and design issues & performance address mapping and replacement
- Auxiliary memories: magnetic disk, magnetic tape and optical disks
- Virtual memory: concept implementation