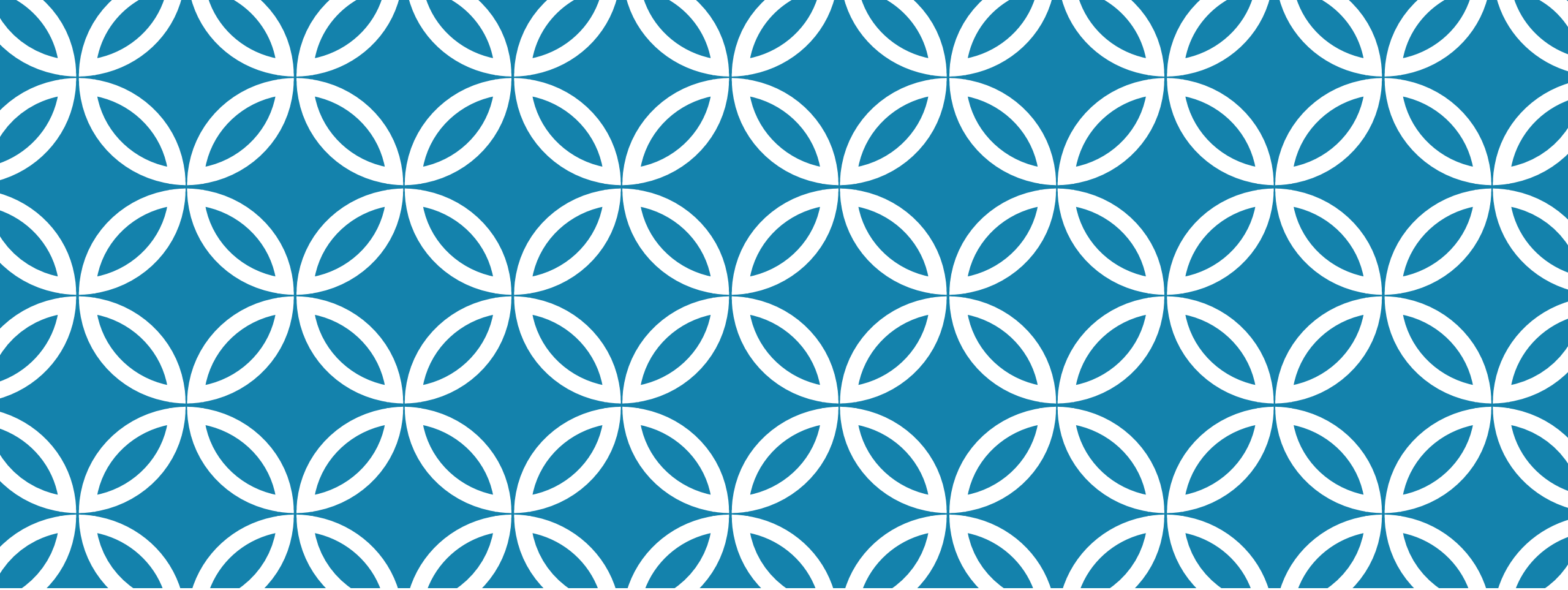


MEMORY ORGANIZATION

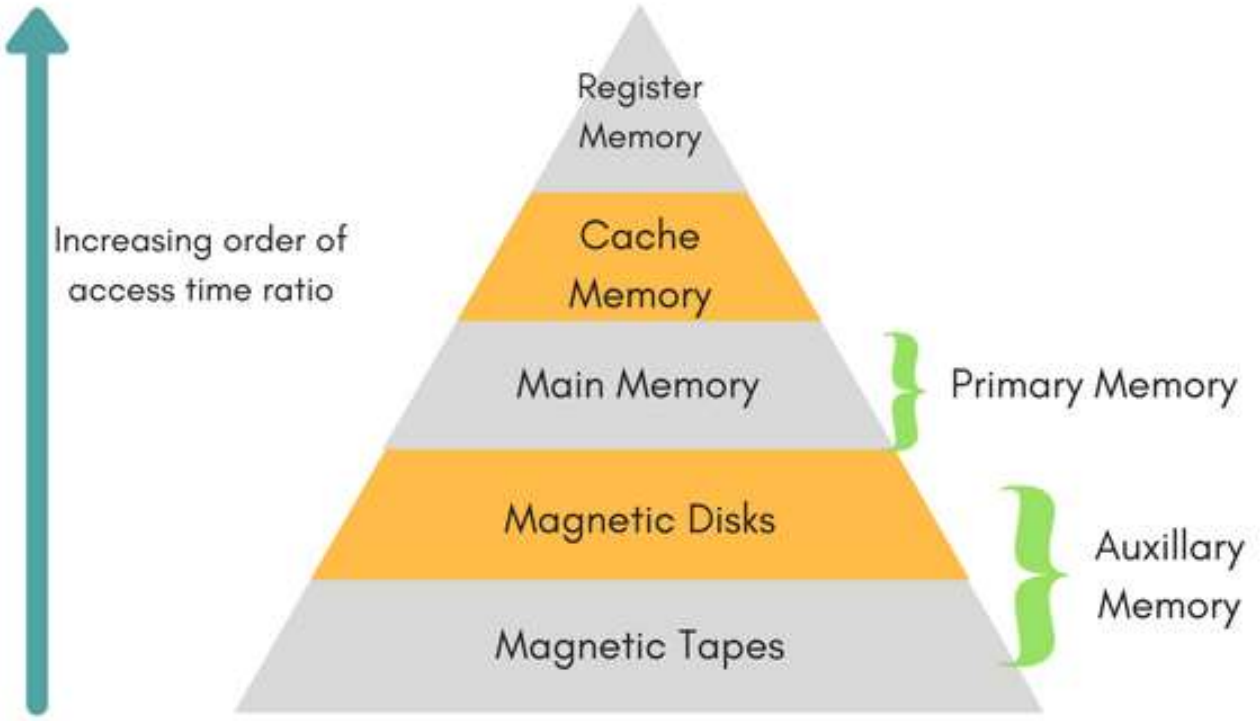
12

WHAT IS A MEMORY?

A memory unit is the collection of storage units or devices together. The memory unit stores the binary information in the form of bits. Generally, memory/storage is classified into 2 categories:

- **Volatile Memory**: This loses its data, when power is switched off.
- **Non-Volatile Memory**: This is a permanent storage and does not lose any data when power is switched off.

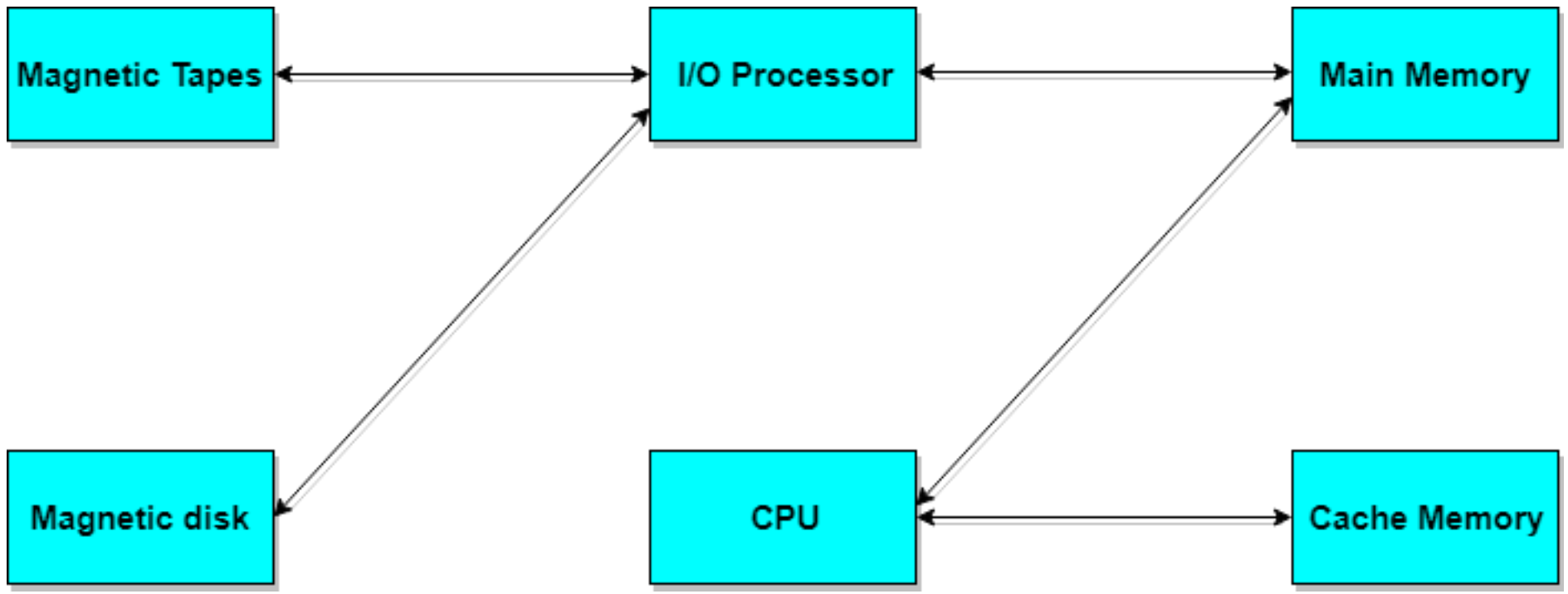
MEMORY HIERARCHY



Auxillary memory access time is generally **1000 times** that of the main memory, hence it is at the bottom of the hierarchy.

The **main memory** occupies the central position because it is equipped to communicate directly with the CPU and with auxiliary memory devices through Input/output processor (I/O).

The **cache memory** is used to store program data which is currently being executed in the CPU. Approximate access time ratio between cache memory and main memory is about **1 to 7~10**


Magnetic Tapes
I/O Processor
Main Memory
Magnetic disk
CPU
Cache Memory

AUXILIARY MEMORY

Devices that provide backup storage are called auxiliary memory. **For example:** Magnetic disks and tapes are commonly used auxiliary devices. Other devices used as auxiliary memory are magnetic drums, magnetic bubble memory and optical disks.

It is not directly accessible to the CPU, and is accessed using the Input/Output channels.

CACHE MEMORY

The data or contents of the main memory that are used again and again by CPU, are stored in the cache memory so that we can easily access that data in shorter time.

Whenever the CPU needs to access memory, it first checks the cache memory. If the data is not found in cache memory then the CPU moves onto the main memory. It also transfers block of recent data into the cache and keeps on deleting the old data in cache to accommodate the new one.

ASSOCIATIVE MEMORY

It is also known as **content addressable memory (CAM)**.

It is a memory chip in which each bit position can be compared. In this the content is compared in each bit cell which allows very fast table lookup.

Since the entire chip can be compared, contents are randomly stored without considering addressing scheme. These chips have less storage capacity than regular memory chips.

MEMORY ACCESS METHODS

Each memory type, is a collection of numerous memory locations. To access data from any memory, first it must be located and then the data is read from the memory location.

- **Random Access:** Main memories are random access memories, in which each memory location has a unique address. Using this unique address any memory location can be reached in the same amount of time in any order.
- **Sequential Access:** This methods allows memory access in a sequence or in order.
- **Direct Access:** In this mode, information is stored in tracks, with each track having a separate read/write head.

MAIN MEMORY

It is the central storage unit of the computer system. It is a large and fast memory used to store data during computer operations. Main memory is made up of **RAM** and **ROM**, with RAM integrated circuit chips holing the major share.

RAM: Random Access Memory

- **DRAM**: Dynamic RAM, is made of capacitors and transistors, and must be refreshed every 10~100 ms. It is slower and cheaper than SRAM.
- **SRAM**: Static RAM, has a six transistor circuit in each cell and retains data, until powered off.
- **NVRAM**: Non-Volatile RAM, retains its data, even when turned off. Example: Flash memory.

MAIN MEMORY

Static RAM (SRAM)

- a bit of data is stored using the state of a flip-flop.
- Retains value indefinitely, as long as it is kept powered.
- Mostly uses to create cache memory of CPU.
- Faster and more expensive than DRAM.

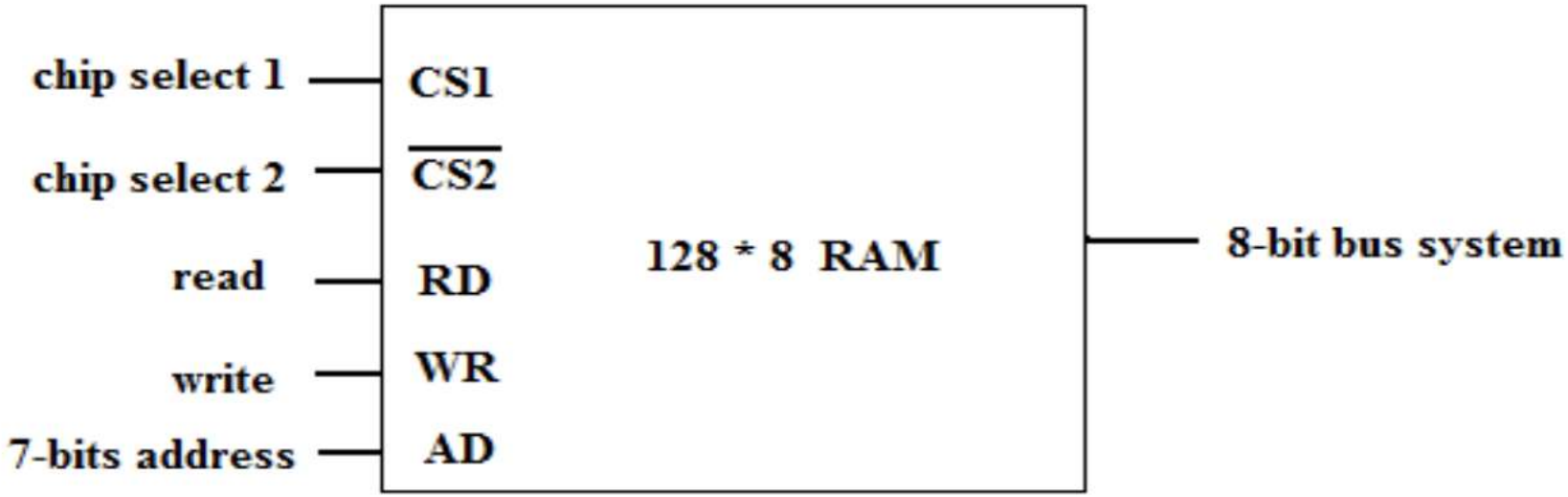
MAIN MEMORY

Dynamic RAM (DRAM)

- Each cell stores bit with a capacitor and transistor.
- Large storage capacity
- Needs to be refreshed frequently.
- Used to create main memory.
- Slower and cheaper than SRAM.

LET'S SEARCH AROUND

Go search in google and find the different products which can be considered as SRAM and DRAM



RAM : the capacity of the memory is 128 words of 8 bits (one byte) per word

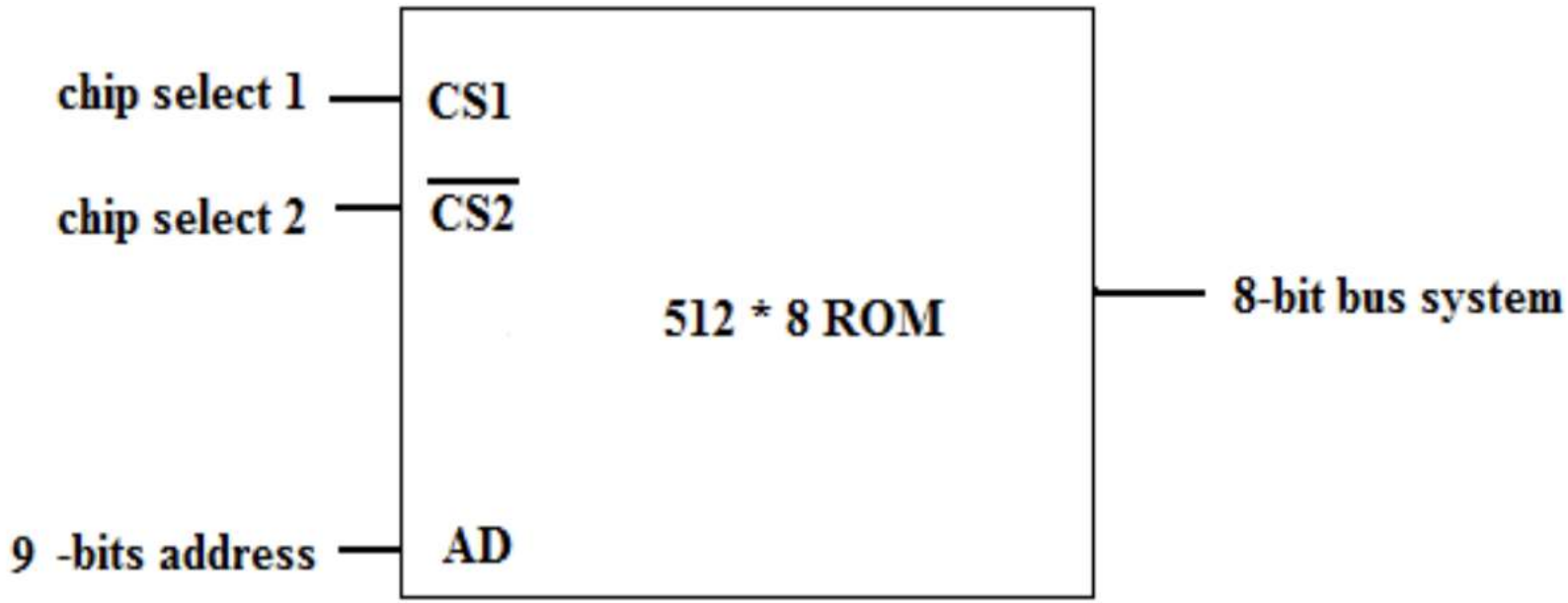
CS1	$\overline{CS2}$	RD	WD	Memory Function	State of data bus
0	0	*	*	Inhibit	High-impedance
0	1	*	*	Inhibit	High-impedance
1	0	0	0	Inhibit	High-impedance
1	0	0	1	Write	Input data to RAM
1	0	1	*	Read	Output data from RAM
1	1	*	*	Inhibit	High-impedance

MAIN MEMORY

ROM: Read Only Memory, is non-volatile and is more like a permanent storage for information. It also stores the **bootstrap loader** program, to load and start the operating system when computer is turned on.

Some commonly used ROMs :

- **PROM**(Programmable ROM)
- **EPROM**(Erasable PROM)
- **EEPROM**(Electrically Erasable PROM)


chip select 1
CS1
chip select 2
$\overline{CS2}$
512 * 8 ROM
8-bit bus system
9 -bits address
AD

ROM

HIT RATIO

The performance of cache memory is measured in terms of a quantity called **hit ratio**. When the CPU refers to memory and finds the word in cache it is said to produce a **hit**. If the word is not found in cache, it is in main memory then it counts as a **miss**.

The ratio of the number of hits to the total CPU references to memory is called hit ratio.

```
Hit Ratio = Hit/(Hit + Miss)
```

MEMORY ADDRESS MAP

- Memory Address Map is a pictorial representation of assigned address space for each chip in the system
- To demonstrate an example, assume that a computer system needs 512 bytes of RAM and 512 bytes of ROM
- The RAM have 128 byte and need seven address lines, where the ROM have 512 bytes and need 9 address lines

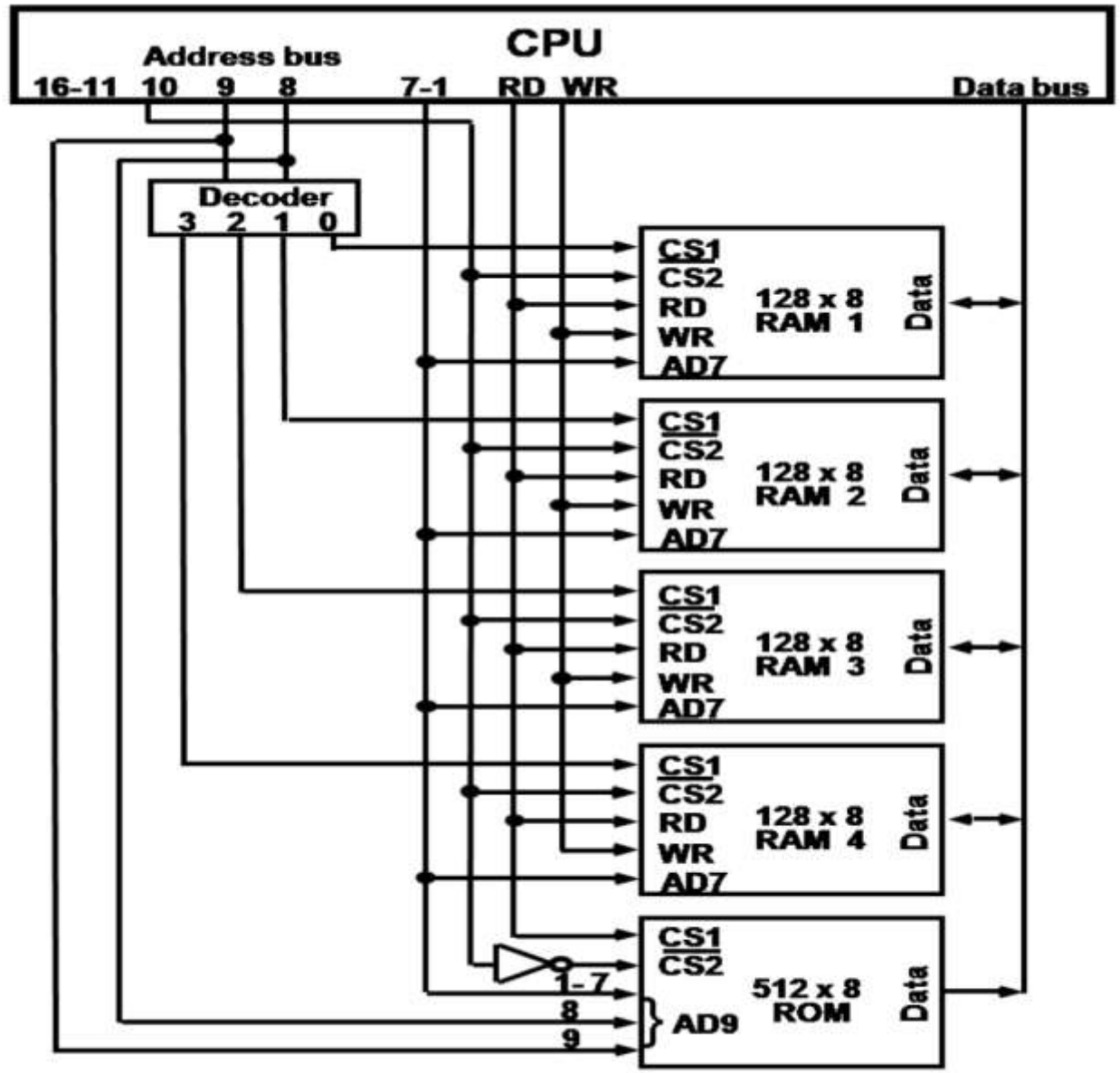
Component	Hexadecimal Address	10	9	8	7	6	5	4	3	2	1
RAM1	0000-007F	0	0	0	*	*	*	*	*	*	*
RAM2	0080-00FF	0	0	1	*	*	*	*	*	*	*
RAM3	0100-017F	0	1	0	*	*	*	*	*	*	*
RAM4	0180-01FF	0	1	1	*	*	*	*	*	*	*
ROM	0200-03FF	1	*	*	*	*	*	*	*	*	*

MEMORY ADDRESS MAP

The hexadecimal address assigns a range of hexadecimal equivalent address for each chip

Line 8 and 9 represent four distinct binary combination to specify which RAM we chose

When line 10 is 0, CPU selects a RAM. And when it's 1, it selects the ROM



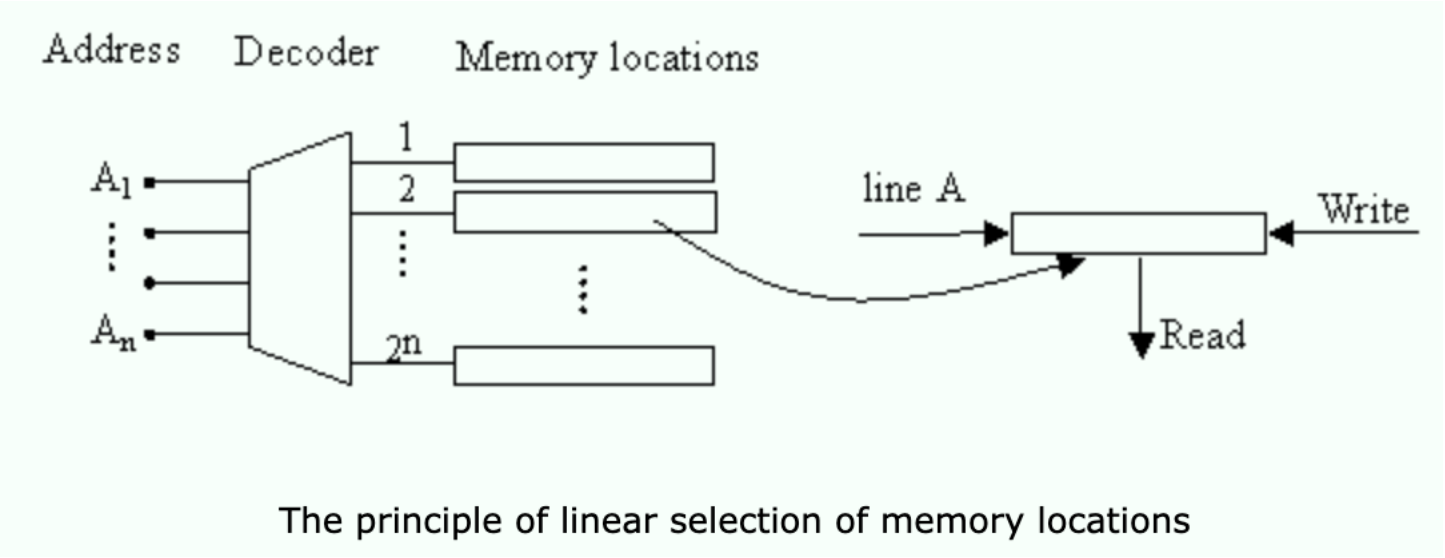
Memory connection to the CPU

MAIN MEMORY SELECTION

- A main memory can be built of a single or many **memory modules**.
- A main memory module is built of an address decoder and a set of memory locations.
- The locations store words of bits of data assigned to consecutive addresses.
- Organization structures of main memories can be divided, according to the circuit that selects memory locations, into the following types:
  - Main memory with linear selection (with a single address decoder)
  - Main memory with two-dimensional selection (with two address decoders)
  - Main memory with linear selection of multiple words (with a single address decoder and a selector)

MAIN MEMORY SELECTION

Main memory with linear selection

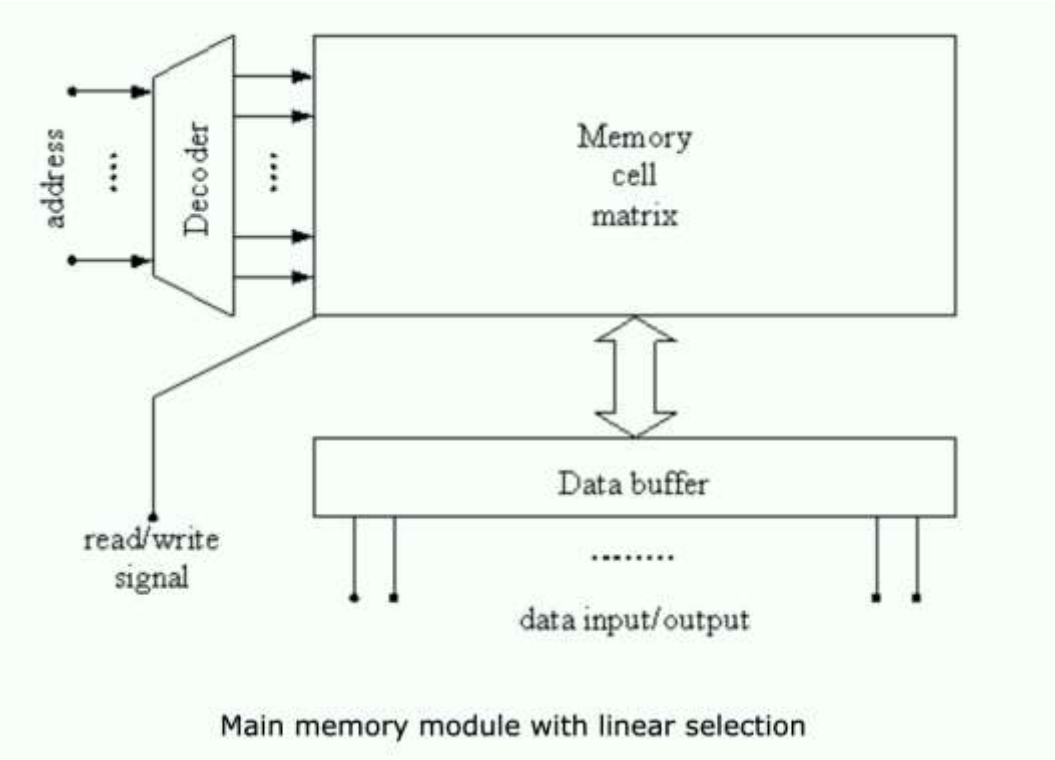


The principle of linear selection of memory locations

A block diagram of the main memory module with linear selection is shown at the left. Information is stored in a matrix of data word locations. The matrix can be represented as a two-dimensional set of elementary bit locations. That is why this type of memory is called sometime a **2D main memory.**

MAIN MEMORY SELECTION

Main memory with linear selection



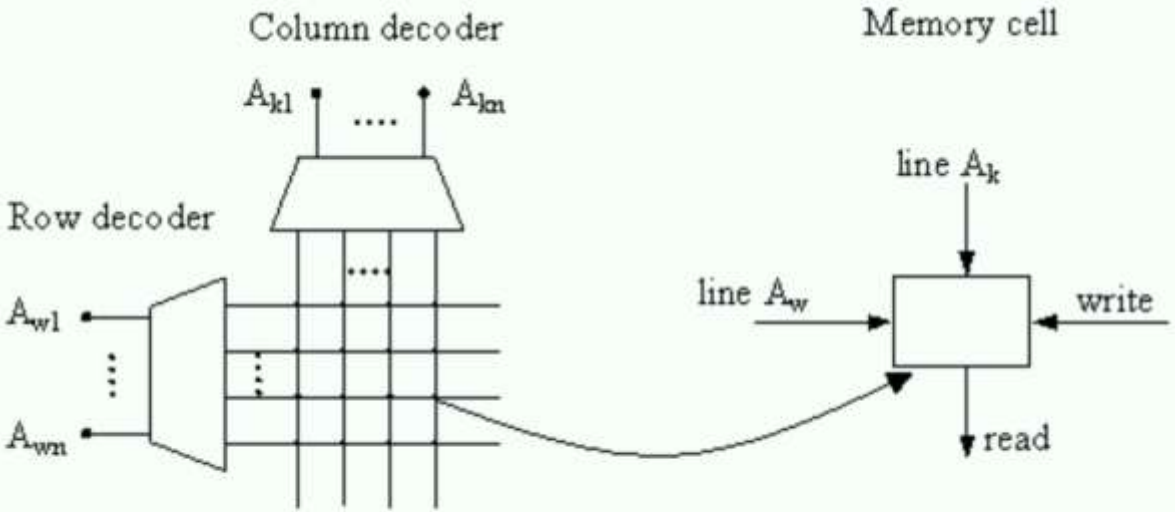
Main memory module with linear selection

To the inputs of the address decoder, the address bus of a processor is connected. The read/write signal is provided by the processor through the control bus.

At the output of the memory cell matrix, a buffer register is placed connected to the external data bus of the processor. It stores the data read from the memory or those which are to be there written.

MAIN MEMORY SELECTION

Main memory with two-dimensional selection



The principle of two-dimensional selection of memory locations

The name of this memory comes from the fact that each address of memory location is divided into the row address and the column address in a memory matrix, in which a given location can be found (at the intersection of the row and column)

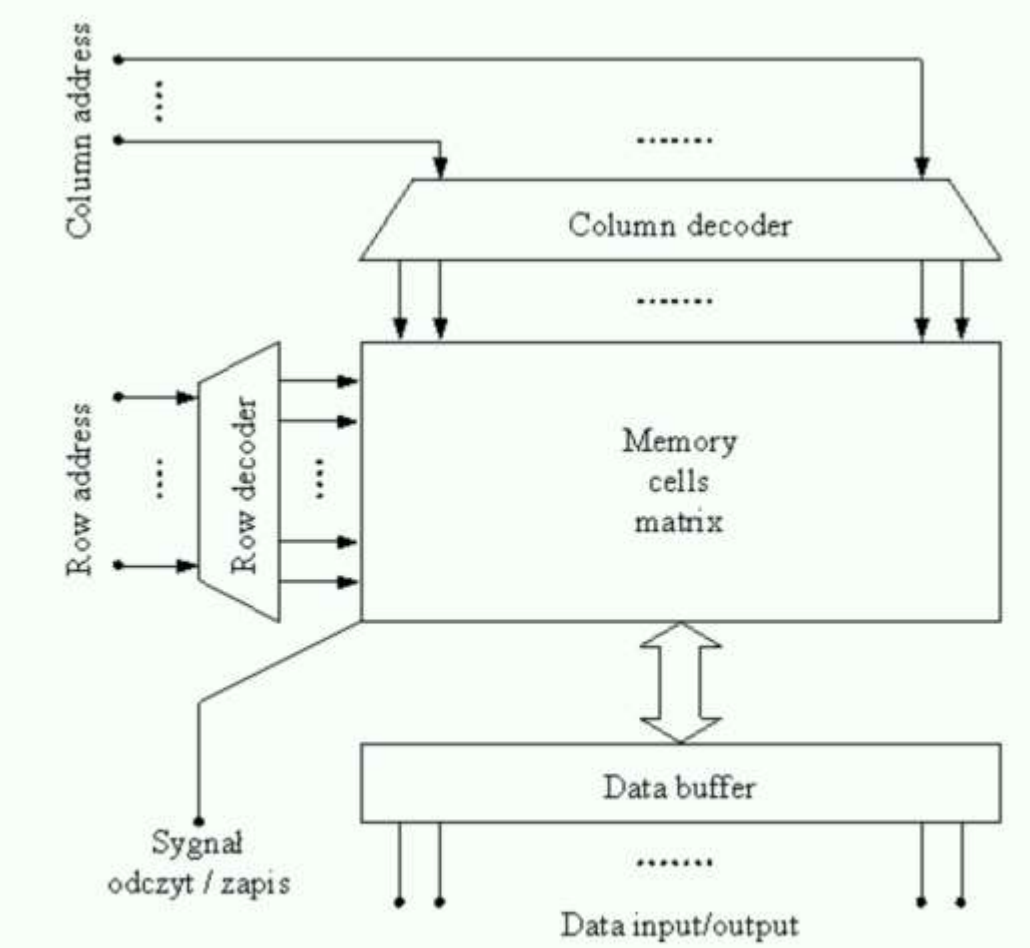
two address decoders are used: one for the row address and another for the column address.

After decoding of the address two lines coming from the decoders are activated. Bits of all words at the same bit position are stored in bit cells placed in the same rectangular bit cell matrix called a bit plane.

MAIN MEMORY SELECTION

Main memory with two-dimensional selection

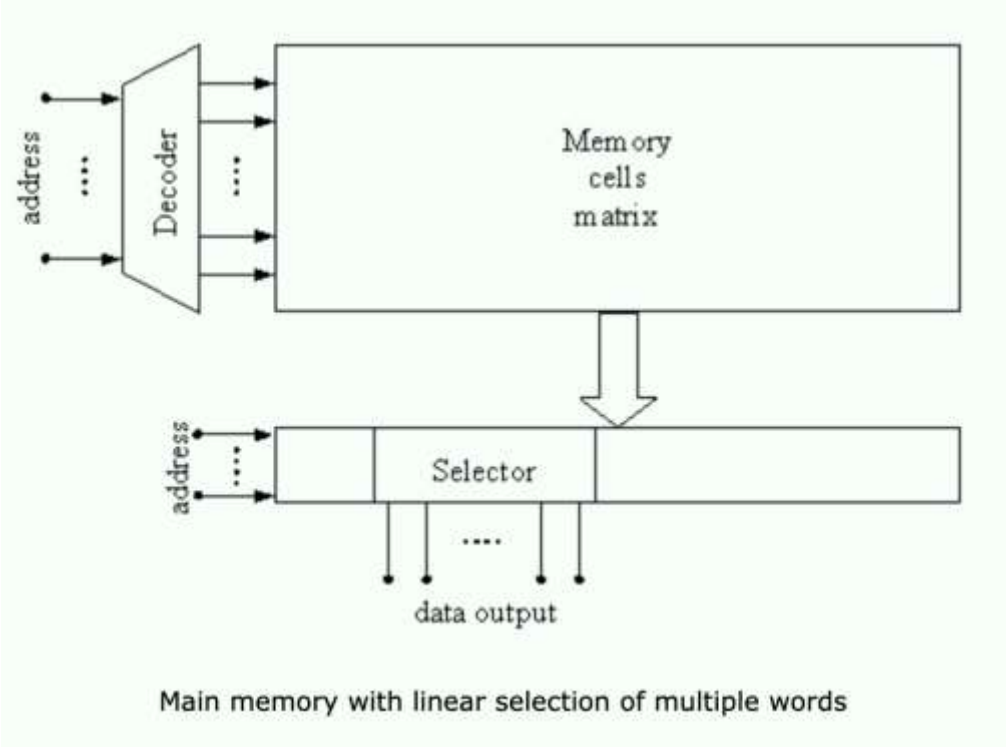
Each pair of output lines from the row and column decoders selects (activates) a sequence of bit cells that belong to planes of consecutive bits in a memory word. In such memory module, the full matrix of bit cells has three-dimensional structure. That is why such type of main memory is called a **3D memory**. The general block diagram of such memory is shown below. As in the memory with linear selection, at data outputs of the 3D memory data buffer register is placed.



Memory module with a two-dimensional selection of memory locations

MAIN MEMORY SELECTION

Main memory with linear selection of multiple words

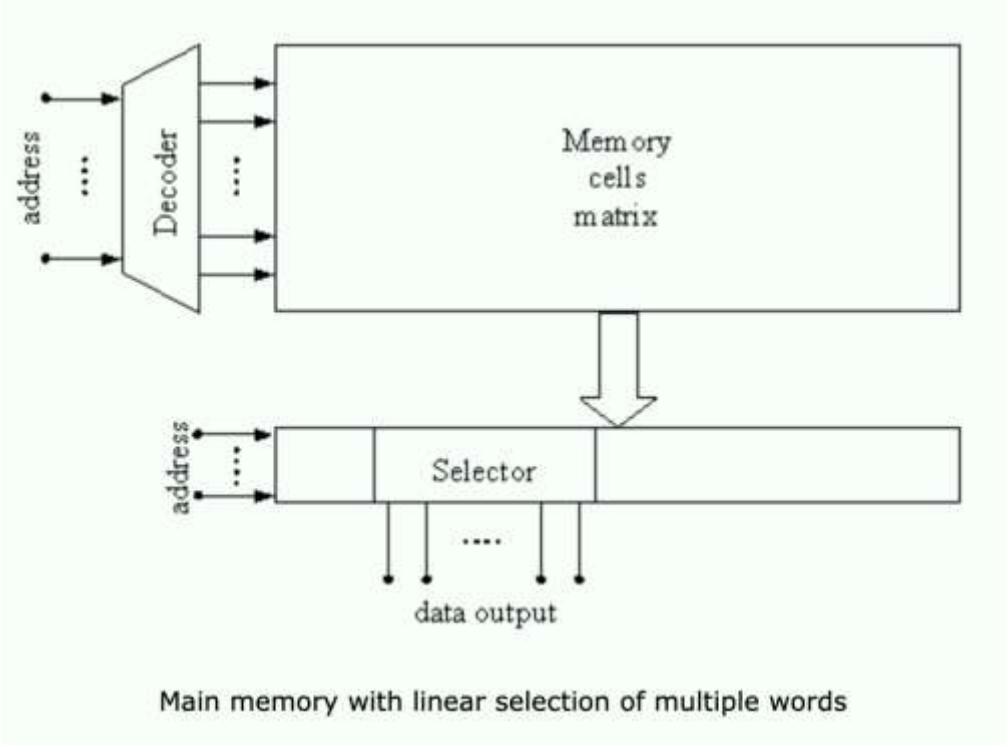


Main memory with linear selection of multiple words

The structure of this type of main memory module is based on the memory with a single address decoder (linear selection). However, in this type of memory, each output line of address decoder activates not a single memory word cell but a sequence of memory word cells. After the readout from the memory matrix, these word sequences are introduced at inputs of a selector, i.e. a next selection unit that selects a single word out of them.

MAIN MEMORY SELECTION

Main memory with linear selection of multiple words



Main memory with linear selection of multiple words

During memory write, data from the data bus are directed to the proper word cell in the sequence selected by the address decoder. The address is divided into two parts: one selects a given sequence of locations and the other is supplied to the selector circuit. In the selector, the second part of address is decoded and the output lines from the decoding select the memory locations to be used in the current memory operation.

In this type of main memory, the bit cells matrix is basically two-dimensional. However the information bits read from the matrix are subject to further selection, which is a half way in the direction towards a three-dimensional structure. For this reason this type of main memory is called **2.5 D memory**.