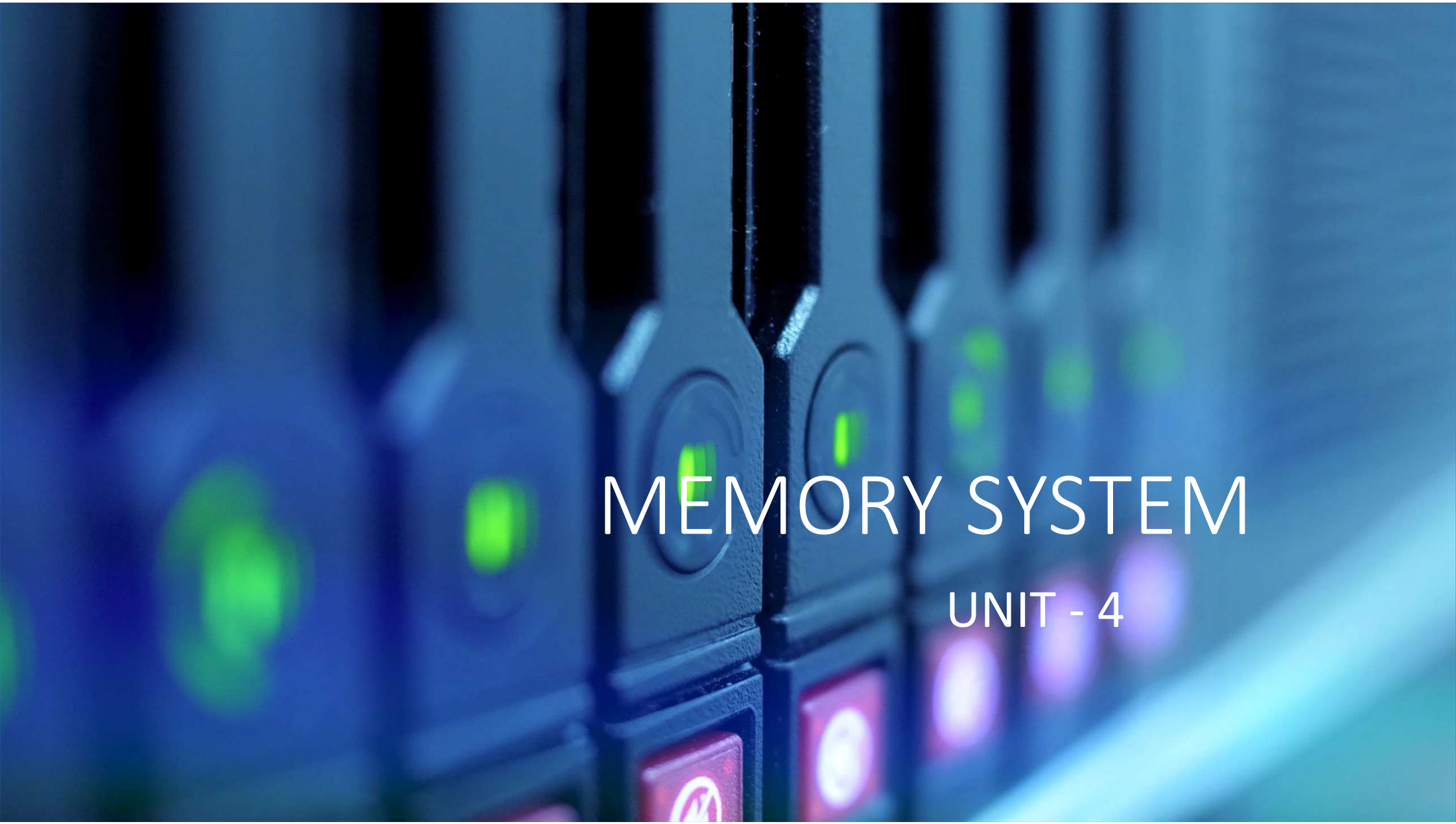


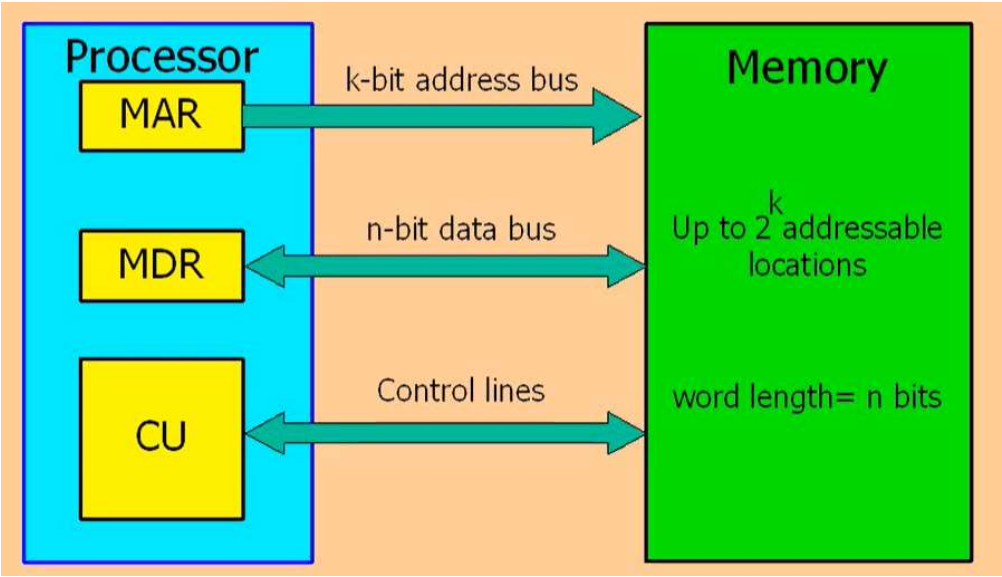
MEMORY SYSTEM

UNIT - 4

SYLLABUS

- BASIC CONCEPTS
- TYPES OF MEMORY
- SPEED, SIZE AND COST
- THE MEMORY Hierarchy
- Locality of reference
- CACHE MEMORY
  - MAPPING FUNCTION
  - REPLACEMENT ALGORITHMS
  - EFFECTIVE ACCESS TIME AND HIT RATIO
- VIRTUAL MEMORY
  - PAGING

BASIC CONCEPTS



The number of addressable location is defined by the size of address lines off the computer this is also referred as address space of the computer

Data is stored or retrieved from memory in word-length quantities.

- Data transfer between memory and processor takes place through the use of 2 processor registers usually called **MAR** (memory address register) and **MDR** (memory data register).
- During a memory cycle and **n** bits of data are transferred between the memory and processor.
- This transfer takes place over the processor bus which has **k** address lines and **n** data lines.
- Bus also contains the control lines such as read/write $(R/\overline{W})$ signal, memory function complete (MFC) etc. for coordinating data transfer.
- Processor reads data from memory by loading address of required memory location into MAR and setting $(R/\overline{W})$ line to 1.
- Memory responds by placing data from addressed location onto the data lines and confirms this action by asserting the MFC signal. Upon receipt of the MFC signal processor loads the data on the data lines into the MDR register

The maximum size of the memory that can be used in any computer is determined by the addressing scheme. For example, a 16-bit computer that generates 16-bit addresses is capable of addressing up to $2^{16} = 64K$ memory locations. Similarly, machines whose instructions generate 32-bit addresses can utilize a memory that contains up to $2^{32} = 4G$ (giga) memory locations, whereas machines with 40-bit addresses can access up to $2^{40} = 1T$ (tera) locations. The number of locations represents the size of the address space of the computer.

A useful measure of the speed of memory units is the time that elapses between the initiation of an operation and the completion of that operation, for example, the time between the Read and the MFC signals. This is referred to as the *memory access time.*

Another important measure is the *memory cycle time,* which is the minimum time delay required between the initiation of two successive memory operations, for example, the time between two successive Read operations. The cycle time is usually slightly longer than the access time, depending on the implementation details of the memory unit.

A memory unit is called *random-access memory* (RAM) if any location can be accessed for a Read or Write operation in some fixed amount of time that is independent of the location's address. This distinguishes such memory units from serial, or partly serial, access storage devices such as magnetic disks and tapes. Access time on the latter devices depends on the address or position of the data.

Comparison between Serial and Random Access Memories

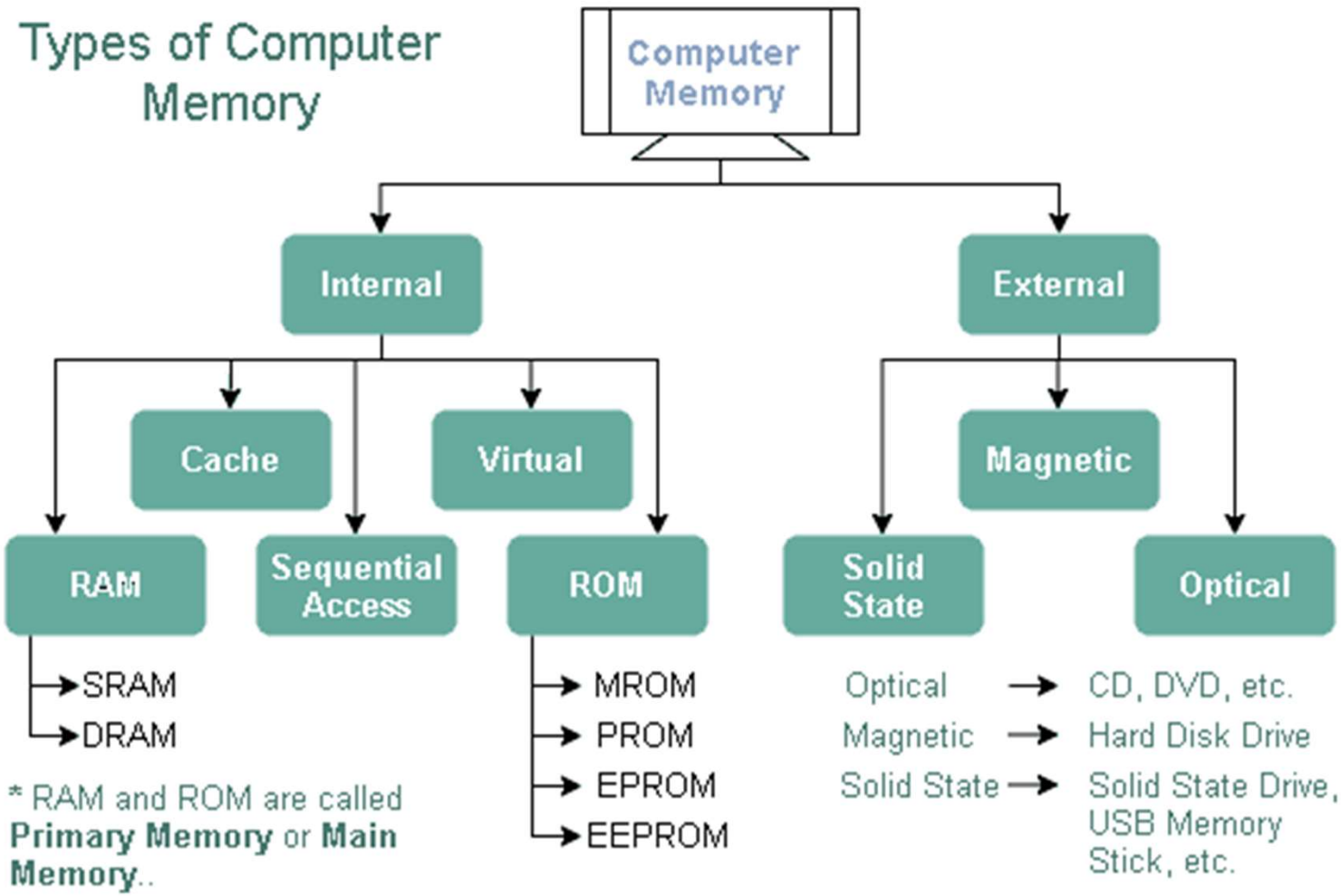
Sr. No.	Serial access memories	Random access memories
1.	Use serial access method.	Use random access method.
2.	Memory is organized into units of data, called **records**. Records are accessed sequentially. If current record is 1, then in order to read record N, it is necessary to read physical records 1 through N–1.	Each storage location in the memory has an unique address and it can be accessed independently of the other locations.
3.	Memory access time is dependent on the position of storage location.	Memory access time is independent of storage location being accessed.
4.	The time required to bring the desired location into correspondence with a read-write head increases effective access time, so serial access tends to be slower than random access.	Memory access time is less.
5.	Cheaper than random access memories.	Random access memories are comparatively costly.
6.	Nonvolatile memories.	May be volatile or nonvolatile depending on physical characteristics.
7.	Magnetic tape is an example of serial access memories.	Semiconductor memories are random access memories.

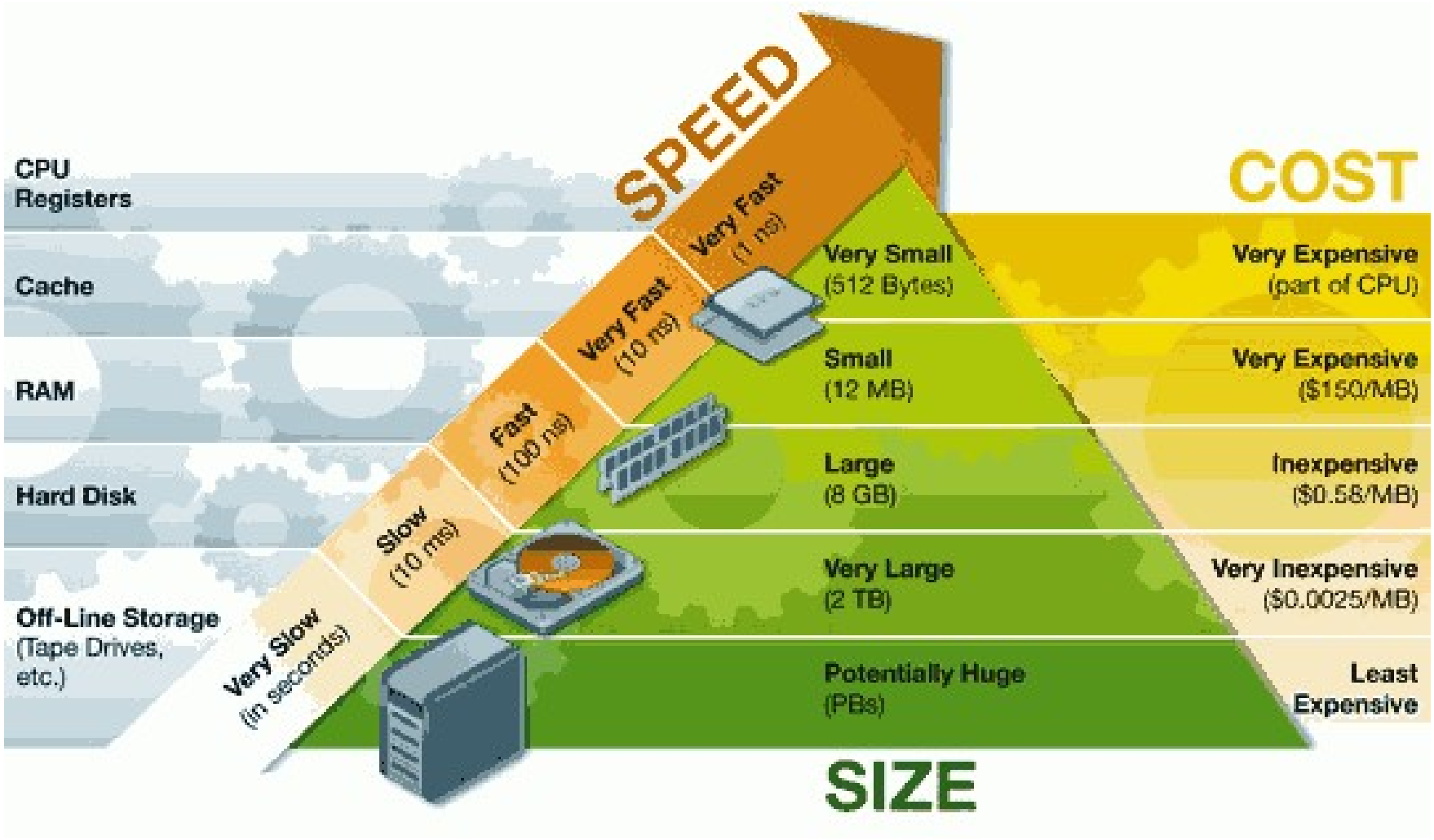
Volatile/Non-volatile	If memory can hold data even if power is turned off, it is called as non-volatile memory ; otherwise it is called as volatile memory.

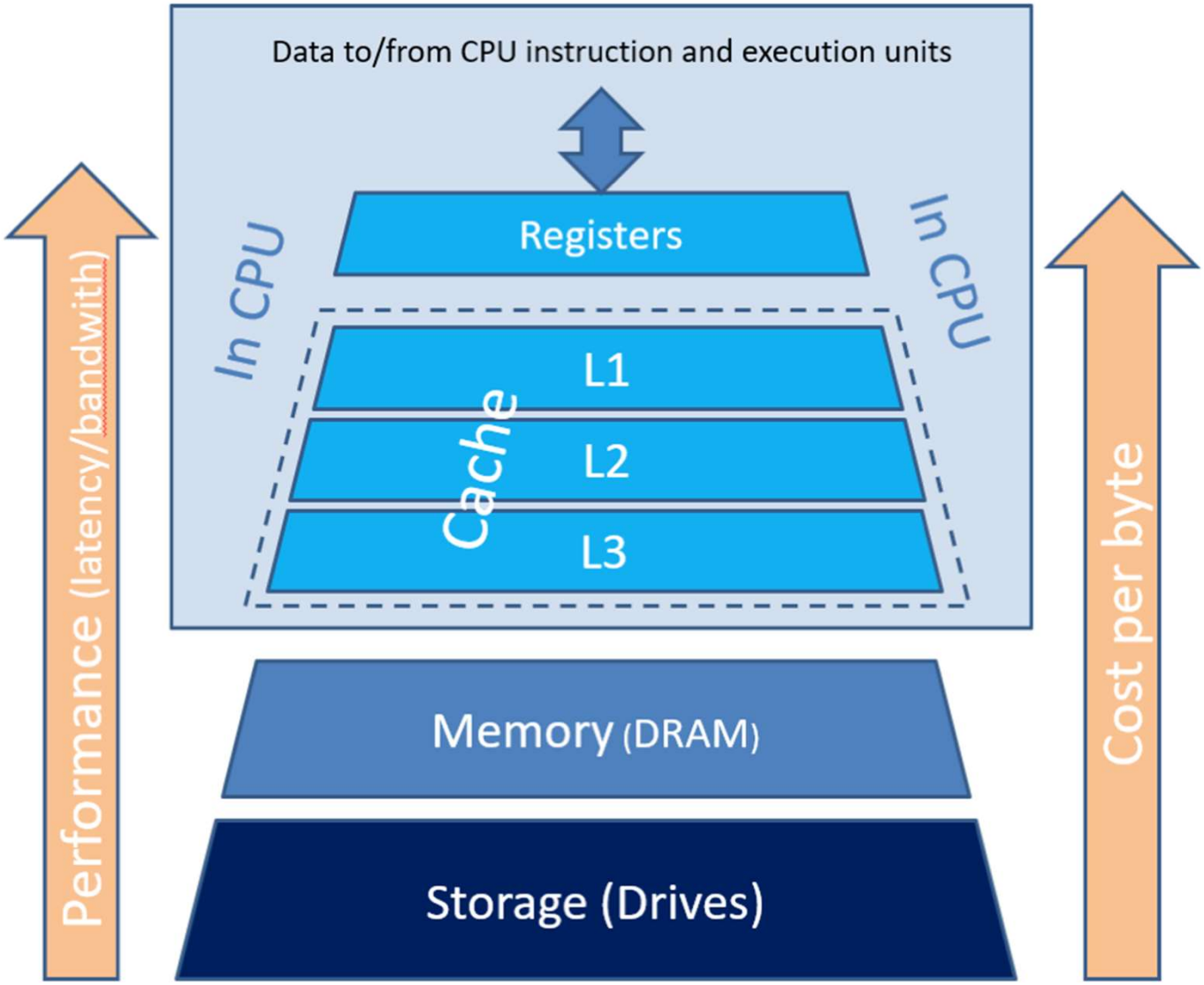
Number of word	This term specifies the number of words available in the particular memory device. For example, if memory capacity is 4 K × 8, then its word size is 8 and number of words are 4 K = 4096.

The processor of a computer can usually process instructions and data faster than they are fetched from the memory unit. The memory cycle time, then is the bottleneck in the system. One way to avoid this bottleneck is to use a **cache memory**. Cache memory is a small, fast memory that is inserted between the larger, slower main memory and the processor. It usually holds the currently active segments of a program and their data.

In most modern computers, the physical main memory is not as large as the address space spanned by an address issued by the processor. Here, the virtual memory technique is used to extend the apparent size of the physical memory. It uses secondary storage such as disks, to extend the apparent size of the physical memory.

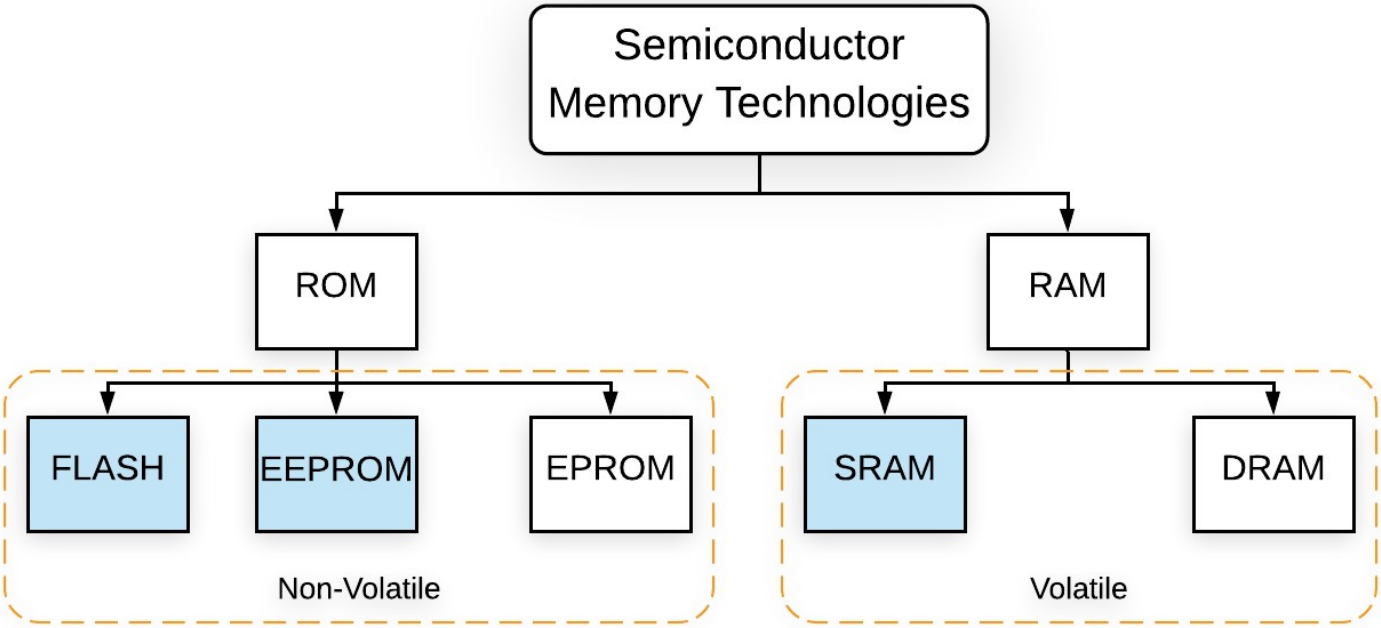

Types of Computer Memory
Computer Memory
Internal
External
Cache
Virtual
Magnetic
RAM
Sequential Access
ROM
Solid State
Optical
SRAM
DRAM
MROM
PROM
EPROM
EEPROM
* RAM and ROM are called Primary Memory or Main Memory..
Optical → CD, DVD, etc.
Magnetic → Hard Disk Drive
Solid State → Solid State Drive, USB Memory Stick, etc.


CPU
Registers
Cache
RAM
Hard Disk
Off-Line Storage
(Tape Drives,
etc.)
SPEED
Very Fast
(1 ns)
Very Fast
(10 ns)
Fast
(100 ns)
Slow
(10 ms)
Very Slow
(in seconds)
Very Small
(512 Bytes)
Small
(12 MB)
Large
(8 GB)
Very Large
(2 TB)
Potentially Huge
(PBs)
COST
Very Expensive
(part of CPU)
Very Expensive
($150/MB)
Inexpensive
(0.5/MB)
Very Inexpensive
($0.0025/MB)
Least
Expensive
SIZE

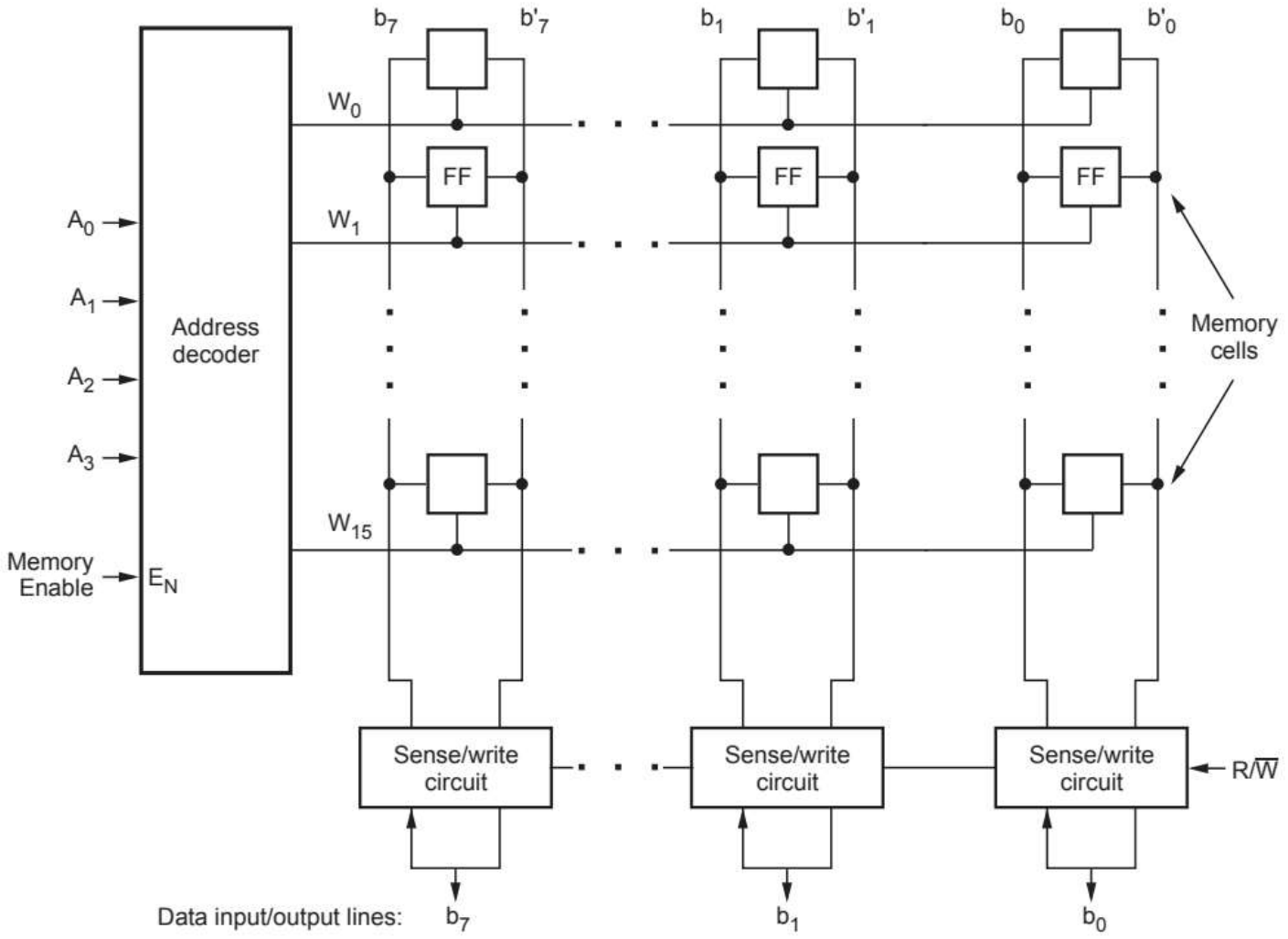

Data to/from CPU instruction and execution units
Registers
In CPU
In CPU
L1
Cache
L2
L3
Memory (DRAM)
Storage (Drives)
Performance (latency/bandwith)
Cost per byte

Semiconductor memory

Semiconductor memory is a digital electronic semiconductor device used for digital data storage. It typically refers to as MOS memory, where data is stored within metal–oxide–semiconductor (MOS) memory cells on a silicon integrated circuit memory chip.



INTERNAL ORGANIZATION OF MEMORY CHIPS

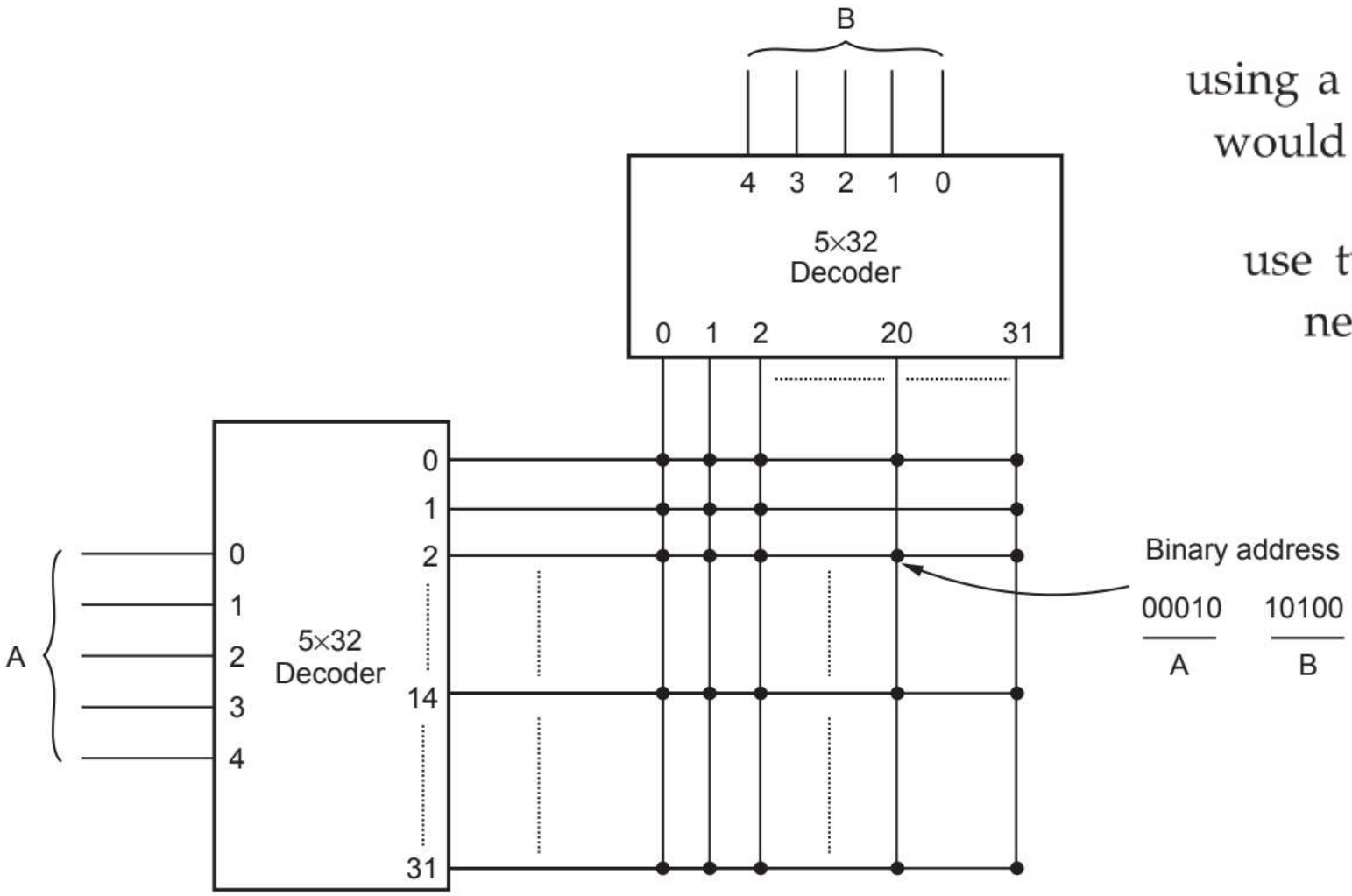


Organization of bit cells in a memory chip

Memory cells are usually organized in the form of an array, in which each cell is capable of storing one bit of information.

Each row of cells constitutes a memory word, and all cells of a row are connected to a common line referred to as the *word line*, which is driven by the address decoder on the chip. The cells in each column are connected to a Sense/Write circuit by two *bit lines*. The Sense/Write circuits are connected to the data input/output lines of the chip. During a Read operation, these circuits sense, or read, the information stored in the cells selected by a word line and transmit this information to the output data lines. During a Write operation, the Sense/Write circuits receive input information and store it in the cells of the selected word.

Organization using Two Dimensional Decoding

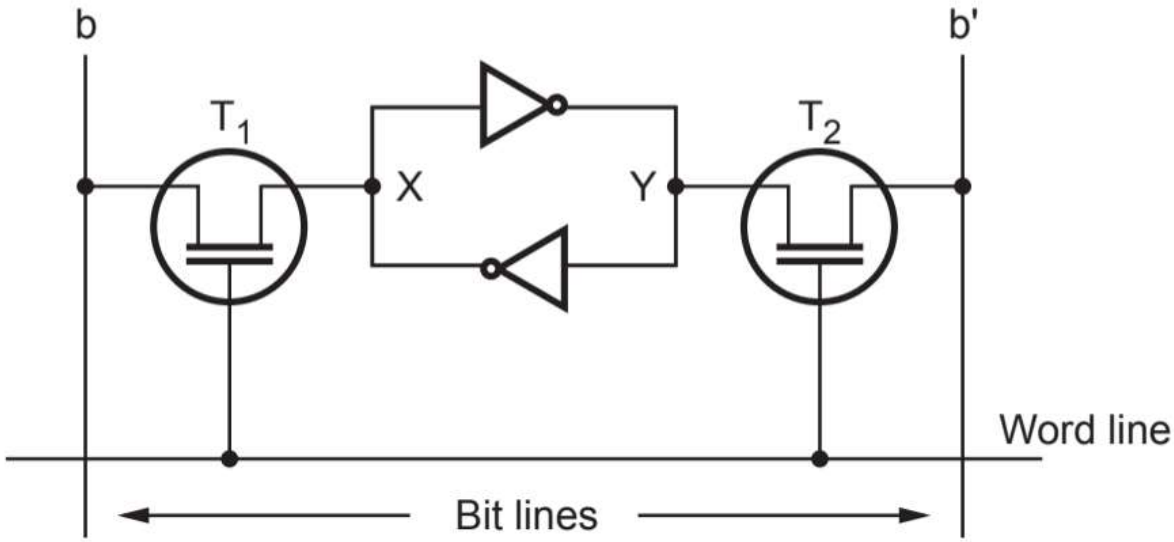


using a single 10×1024 decoder, would need 1024 AND gates

use two 5×32 decoders. need 64 AND gates

Memories that consist of circuits capable of retaining their state as long as power is applied are known as *static memories*.

Static RAM Cell



Static RAM Cell

It consists of two cross-coupled inverters as a latch and two transistors T_1 and T_2 which act as a switches.

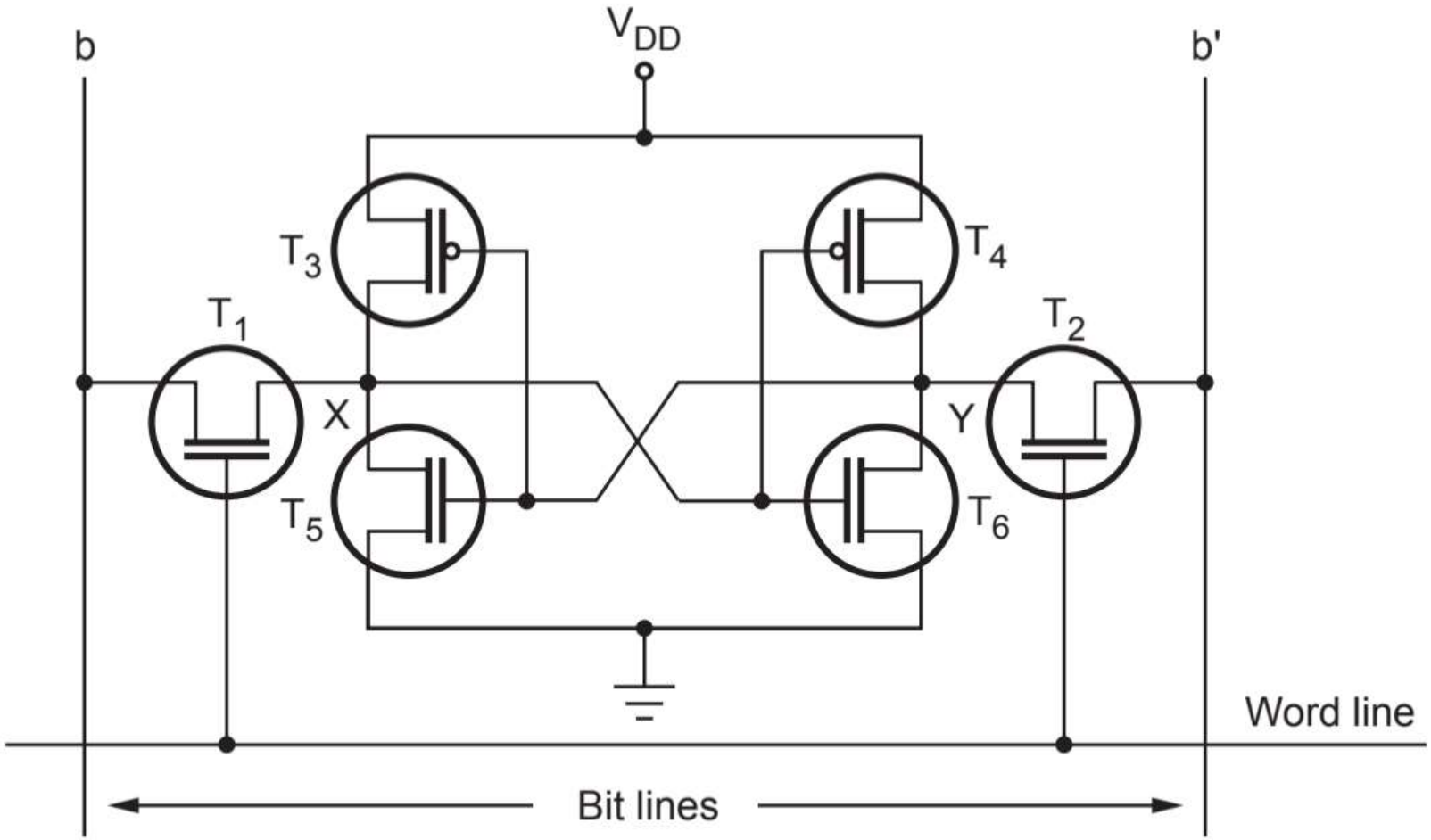
The latch is connected to two bit lines by transistors T_1 and T_2. The word line controls the opening and closing of transistors T_1 and T_2. When word line is at logic 0 level (Ground level), the transistors are off and the latch retains its state.

Read operation

For read operation, word line is made logic 1 (high) so that both transistors are ON. Now if the cell is in state 1, the signal on bit line b is high and the signal on bit line b′ is low. The opposite is true if the cell is in state 0. The b and b′ are complements of each other. The sense/write circuits connected to the bit lines monitor the states of b and b′ and set the output accordingly.

Write operation

For write operation, the state to be set is placed on the line b and its complement is placed on line b′ and then the word line is activated. This action forces the cell into the corresponding state and write operation is completed.

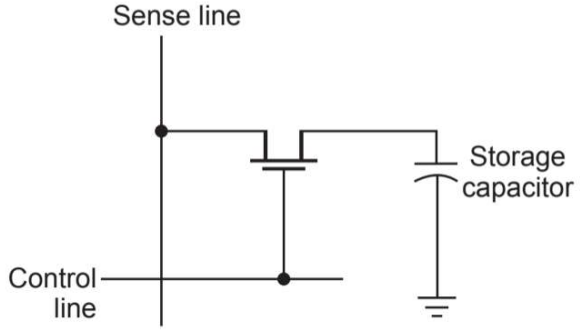


CMOS memory cell

DRAM (ASYNCHRONOUS)

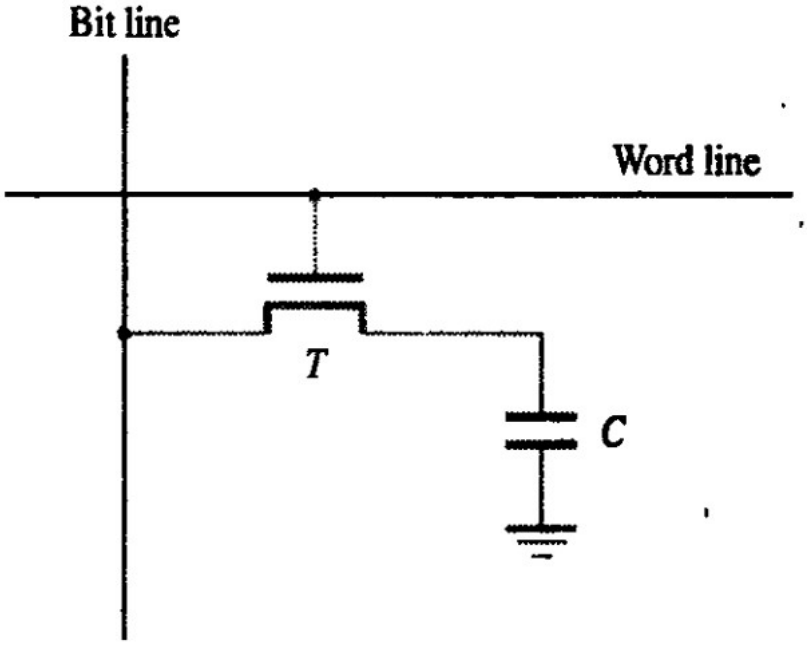
Static RAMs are fast, but they come at a high cost because their cells require several transistors. Less expensive RAMs can be implemented if simpler cells are used. However, such cells do not retain their state indefinitely; hence, they are called *dynamic RAMs* (DRAMs).

Information is stored in a dynamic memory cell in the form of a charge on a capacitor, and this charge can be maintained for only tens of milliseconds. Since the cell is required to store information for a much longer time, its contents must be periodically refreshed by restoring the capacitor charge to its full value.



DRAM (ASYNCHRONOUS)

An example of a dynamic memory cell that consists of a capacitor, C, and a transistor, T, is shown in Figure 5.6. In order to store information in this cell, transistor T is turned on and an appropriate voltage is applied to the bit line. This causes a known amount of charge to be stored in the capacitor.



A single-transistor dynamic memory cell.

After the transistor is turned off, the capacitor begins to discharge.

the information stored in the cell can be retrieved correctly only if it is read before the charge on the capacitor drops below some threshold value.

During a Read operation, the transistor in a selected cell is turned on.

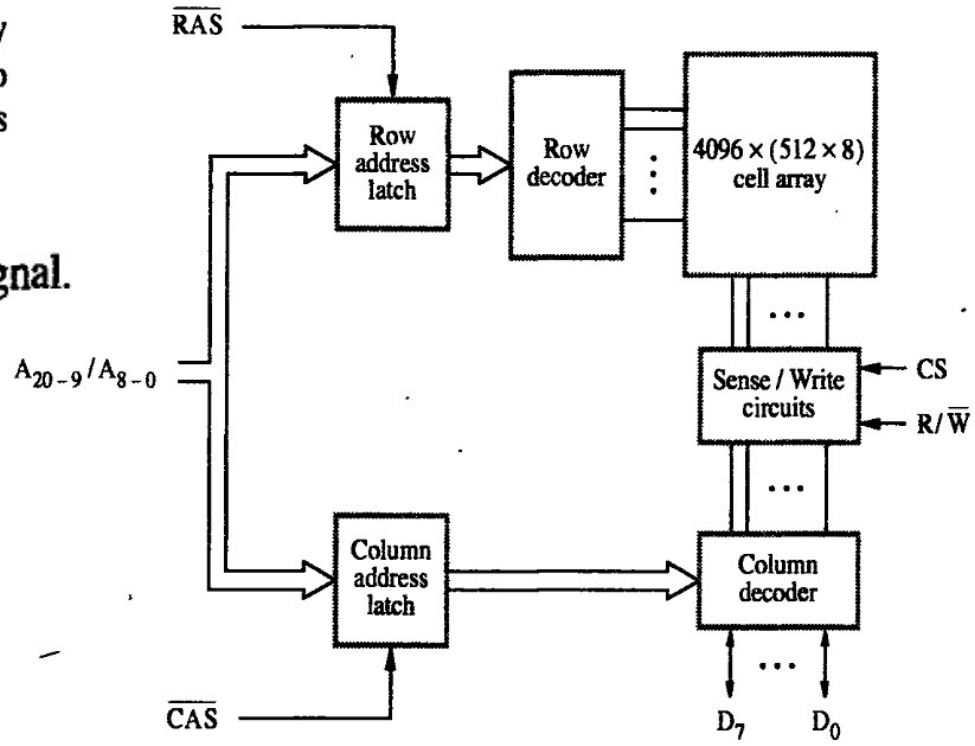
A sense amplifier connected to the bit line detects whether the charge stored on the capacitor is above the threshold value. If so, it drives the bit line to a full voltage that represents logic value 1. This voltage recharges the capacitor to the full charge If the sense amplifier detects that the charge on the capacitor is below the threshold value, it pulls the bit line to ground level, Thus, reading the contents of the cell automatically refreshes its contents.

To ensure that the contents of a DRAM are maintained, each row of cells must be accessed periodically. A *refresh circuit* usually performs this function automatically.

DRAM (ASYNCHRONOUS)

In the DRAM described in this section, the timing of the memory device is controlled asynchronously. A specialized memory controller circuit provides the necessary control signals, RAS and CAS, that govern the timing. The processor must take into account the delay in the response of the memory. Such memories are referred to as *asynchronous DRAMs*.

DRAMs whose operation is directly synchronized with a clock signal. are known as *synchronous DRAMs* (SDRAMs).



SYNCHRONOUS DRAM (SDRAM)

cell array is the same as in asynchronous DRAMs.

The address and data connections are buffered by means of registers.

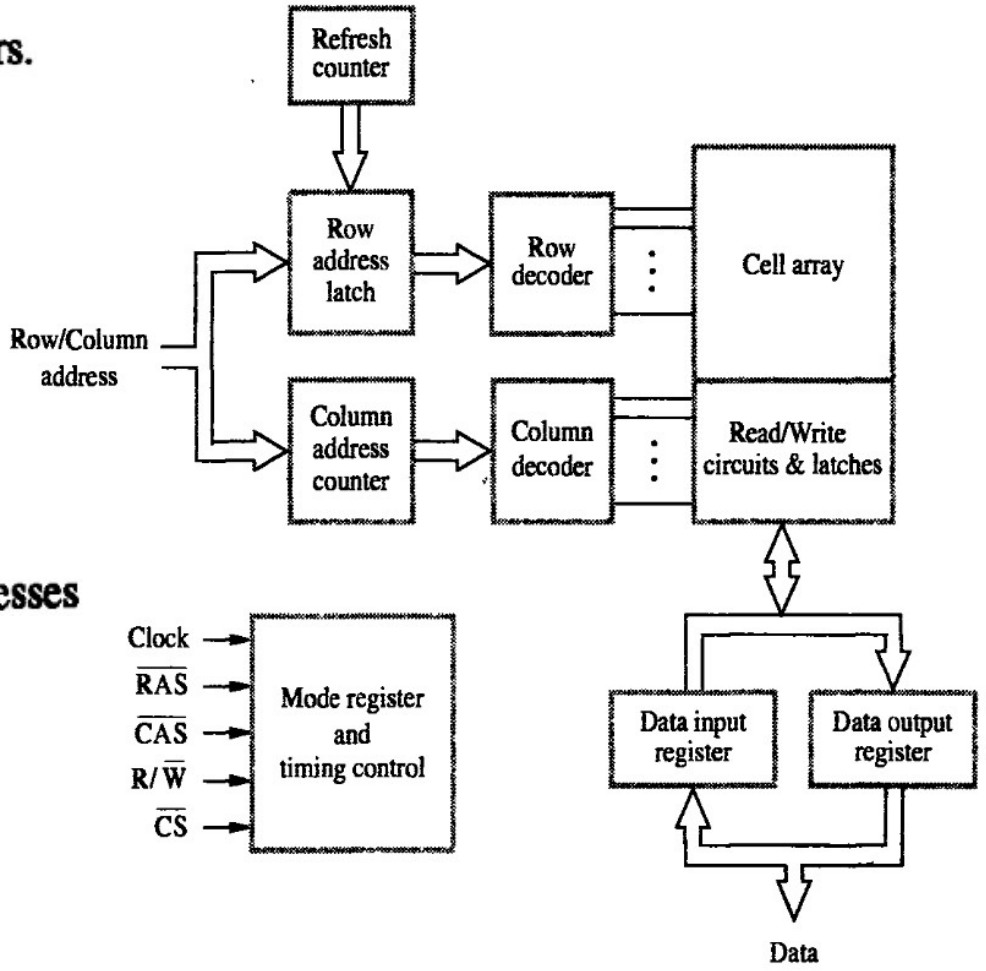
SDRAMs have several different modes of operation, which can be selected by writing control information into a *mode* register.

All actions are triggered by the rising edge of the clock.

SDRAMs have built-in refresh circuitry.

A part of this circuitry is a refresh counter, which provides the addresses of the rows that are selected for refreshing.

each row must be refreshed at least every 64 ms.



DOUBLE DATA RATE SDRAM (DDR)

The standard SDRAM performs all actions on the rising edge of the clock signal.

A similar memory device is available, which accesses the cell array in the same way, but transfers data on both edges of the clock. The latency of these devices is the same as for standard SDRAMs.

But, since they transfer data on both edges of the clock, their bandwidth is essentially doubled for long burst transfers.

Such devices are known as *double-data-rate SDRAMs* (DDR SDRAMs).

To make it possible to access the data at a high enough rate, the cell array is organized in two banks. Each bank can be accessed separately. Consecutive words of a given block are stored in different banks. Such *interleaving* of words allows simultaneous access to two words that are transferred on successive edges of the clock.

DDR SDRAMs and standard SDRAMs are most efficiently used in applications where block transfers are prevalent. This is the case in general-purpose computers in which main memory transfers are primarily to and from processor caches,

Different Types of DDR RAM

There are five types of DDR SDRAM is given below.

1. **DDR1 SDRAM**

2. **DDR2 SDRAM**

3. **DDR3 SDRAM**

4. **DDR4 SDRAM**

5. **DDR5 SDRAM**

	SDRAM	DDR	DDR2	DDR3	DDR4	DDR5
Prefetch	1 - Bit	2 - Bit	4 - Bit	8 - Bit	Bit per Bank	16 - Bit
Date Rate (M	100 - 166	266 - 400	533 - 800	1066 - 1600	2133 - 5100	3200 - 6400
Transfer Rate GB/s	0.8 - 1.3	2.1 - 3.2	4.2 - 6.4	8.5 - 14.9	17 - 25.6	38.4 - 51.2
Voltage (V)	3.3	2.5 - 2.6	1.8	1.35 - 1.5	1.2	1.1

SRAM VS DRAM

Basis For Comparision	SRAM	DRAM
Speed	Faster	Slower
Size	Small	Large
Cost	Expensive	Cheap
Used In	Cache Memory	Main Memory
Density	Less Dense	Highly Dense
Construction	Complex and uses transistors and Latches	Simple and Uses Capacitors and very few transistors
Single block of memory Requres	6 Transistors	Only one Tranistors
Power Consumption	Low	High

The performance of the memory can be measured using two typical performance parameters : memory latency and bandwidth.

Memory latency : It is the term used to refer to the amount of time it takes to transfer a word of data to or from the memory. In case of single word transfer memory latency gives proper indication of its performance. In case of burst data transfer, number of data words are transferred and time needed to complete the operation depends also on the rate at which successive words can be transferred and on the rate at which successive words can be transferred and on the size of the block. In block transfers, the term memory latency is used to indicate the time it takes to transfer the first word of the block data. This time is usually substantially longer than the time needed to transfer each subsequent word of a block. Therefore, memory latency can not be a proper parameters to measure the perform of memory under block transfer of data.

Memory bandwidth : We know that the block size in block data transfer is not fixed, hence it is useful to define a performance measure in terms of the number of bits or bytes that can be transferred in one second. This performance measure is often referred to as the **memory bandwidth**. The bandwidth of a memory unit depends on the speed of memory access and on the number of bits that can be accessed in parallel. The number of bits that can be accessed in parallel depends on the data bus width of the processor. Therefore, the bandwidth is the product of the rate at which data are transferred (and accessed) and the width of the data bus.

Read-Only Memories (ROMs)

- SRAM and SDRAM chips are volatile:
  - Lose the contents when the power is turned off.
- Many applications need memory devices to retain contents after the power is turned off.
  - For example, computer is turned on, the operating system must be loaded from the disk into the memory.
  - Store instructions which would load the OS from the disk.
  - Need to store these instructions so that they will not be lost after the power is turned off.
  - We need to store the instructions into a non-volatile memory.
- Non-volatile memory is read in the same manner as volatile memory.
  - Separate writing process is needed to place information in this memory.
  - Normal operation involves only reading of data, this type of memory is called Read-Only memory (ROM).

The Fig. shows the typical configuration of a ROM cell. It consists of a transistor T and switch P. The transistor T is driven by the word line. The contents of cell can be read from the cell when word line is logic 1. A logic value 0 is read if the transistor is connected to ground through switch P. If switch P is open, a logic value 1 is read. The bit line is connected through a resistor to the power supply. A sense circuit at the end of the bit line generates the proper output value. Data is stored into a ROM when it is manufactured.

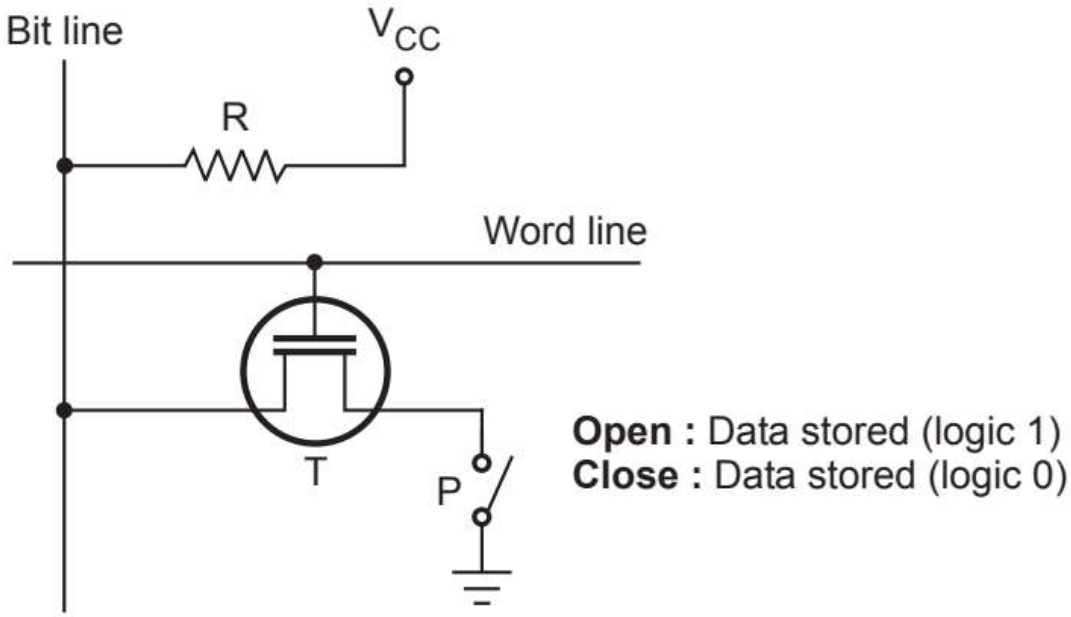
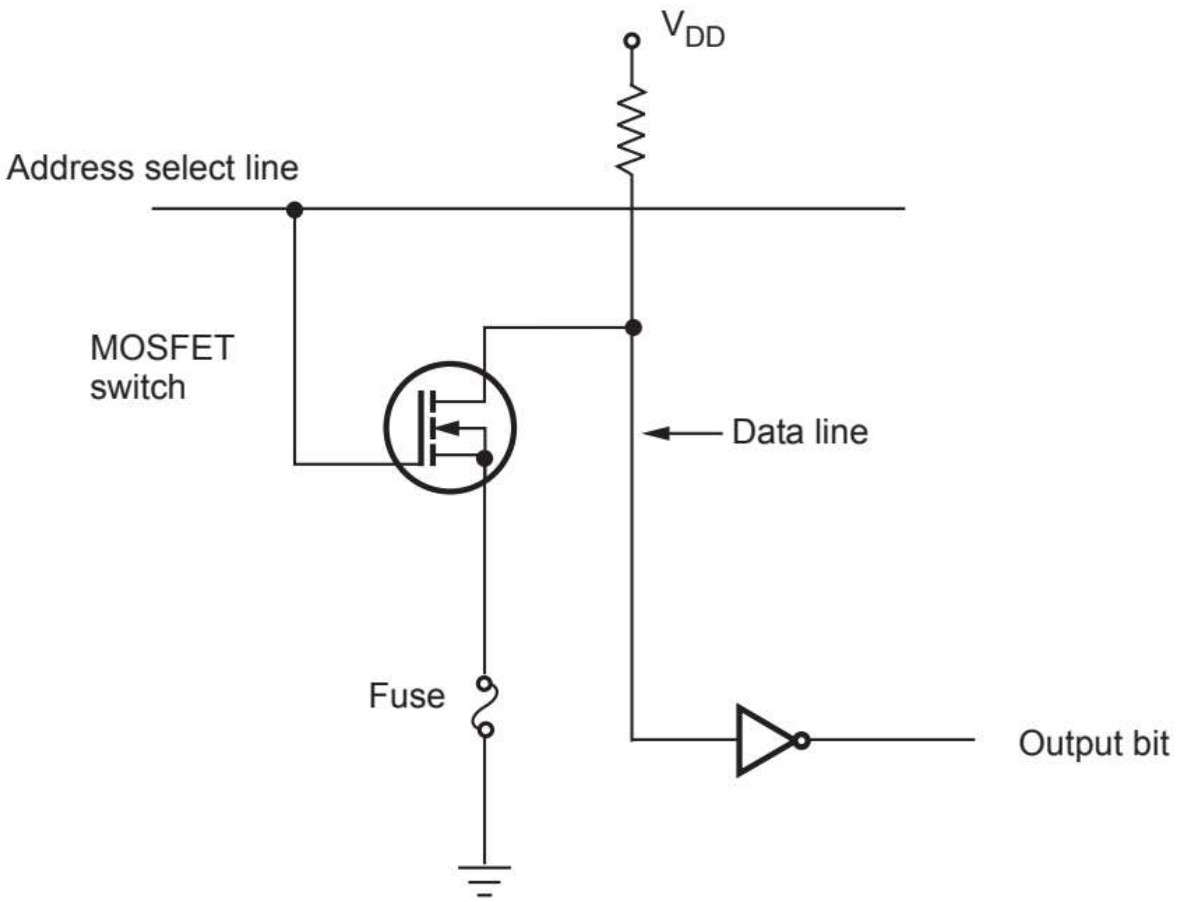


Fig. ROM cell

There are four types of ROM : Masked ROM, PROM, EPROM and EEPROM or E^2PROM.

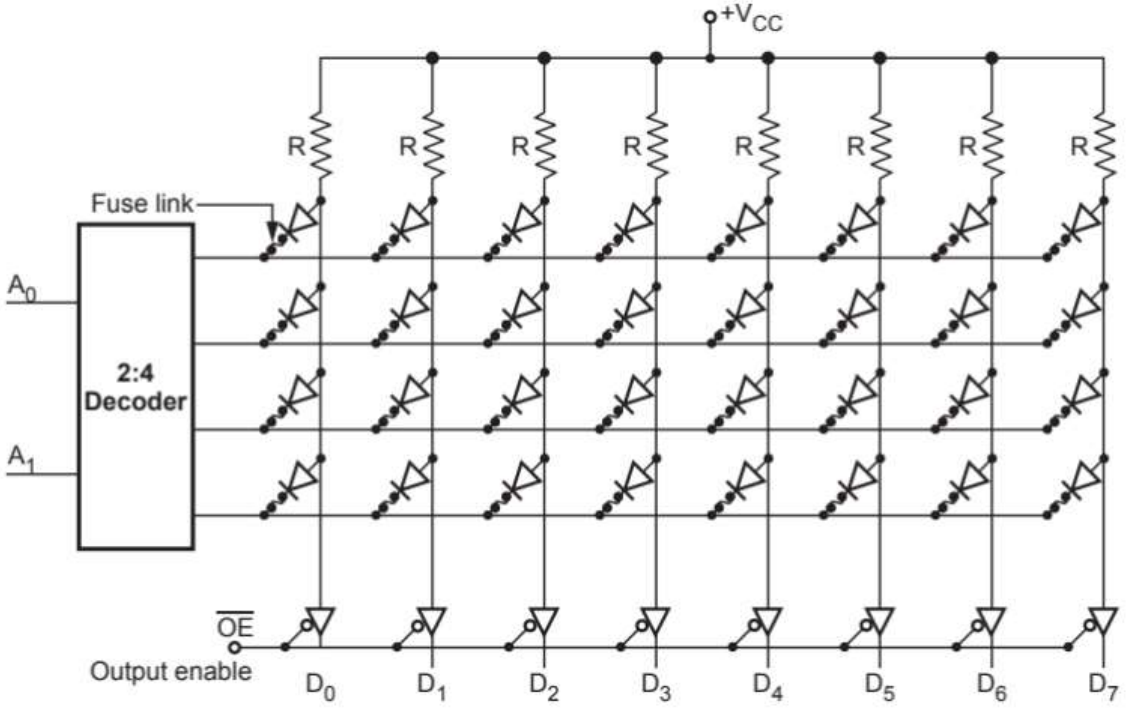
PROM (Programmable Read Only Memory)

PROMs are programmed by user. To provide the programming facility, each address select and data line intersection has its own fused MOSFET or transistor. When the fuse is intact, the memory cell is configured as a logic 1 and when fuse is blown (open circuit), the memory cell is logical 0. Logical 0s are programmed by selecting the appropriate select line and then driving the vertical data line with a pulse of high current.



Single fused PROM cell

Fig. shows four byte PROM. It has diodes in every bit position; therefore, the output is initially all 0s. Each diode, however has a fusible link in series with it. By addressing bit and applying proper current pulse at the corresponding output, we can blow out the fuse, storing logic 1 at that bit position. The fuse uses material like nichrome and polycrystalline. For blowing the fuse it is necessary to pass around 20 to 50 mA of current for period 5 to 20 μs. The blowing of fuses according to the truth table is called programming of ROM. The user can program PROMs with special PROM programmer. The PROM programmer selectively burns the fuses according to the bit pattern to be stored. This process is also known as burning of PROM. The PROMs are one time programmable. Once programmed, the information stored is permanent.



EPROM (Erasable Programmable Read Only Memory)

Erasable programmable ROMs use MOS circuitry. They store 1's and 0's as a packet of charge in a buried layer of the IC chip. EPROMs can be programmed by the user with a special EPROM programmer. The important point is that we can erase the stored data in the EPROMs by exposing the chip to ultraviolet light through its quartz window for 15 to 20 minutes, as shown in the Fig.

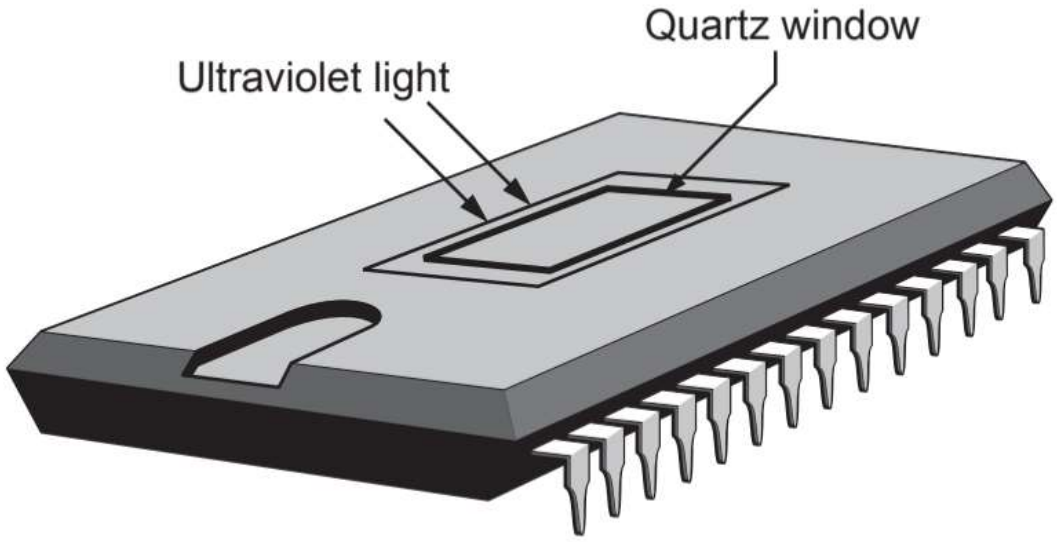


Fig. EPROM

It is not possible to erase selective information, when erased the entire information is lost. The chip can be reprogrammed. This memory is ideally suitable for product development, experimental projects and college laboratories, since this chip can be reused many times.

EEPROM (Electrically Erasable Programmable Read Only Memory)

Electrically erasable programmable ROMs also use MOS circuitry very similar to that of EPROM. Data is stored as charge or no charge on an insulated layer or an insulated floating gate in the device. The insulating layer is made very thin (< 200 Å). Therefore, a voltage as low as 20 to 25 V can be used to move charges across the thin barrier in either direction for programming or erasing. EEPROM allows selective erasing at the register level rather than erasing all the information since the information can be changed by using electrical signals. The EEPROM memory also has a special chip erase mode by which entire chip can be erased in 10 ms. This time is quite small as compared to time required to erase EPROM and it can be erased and reprogrammed with device right in the circuit. However, EEPROMs are most expensive and the least dense ROMs.

Flash Memory

Flash memories are read/write memories. In flash memories it is possible to read the contents of a single cell, but it is only possible to write an entire block of cells. A flash cell is based on a single transistor controlled by trapped charge.

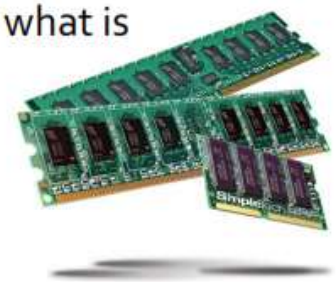
Flash devices have greater density than the EEPROM memory. Due to this flash devices have higher capacity and a lower cost per bit. They require a single power supply voltage and consume less power in their operation. The low power consumption of flash memory makes it suitable for portable equipments such as hand-held computers, cell phones, digital cameras, MP3 music players and so on. In hand-hold computers and cell phones, flash memory is used to store the software needed to operate the equipment. In digital cameras, flash memory is used to store picture image data. In MP3 players, flash memory is used to store the sound data.

The flash memories are available in modules. These modules are implemented in two types : flash cards and flash drives.

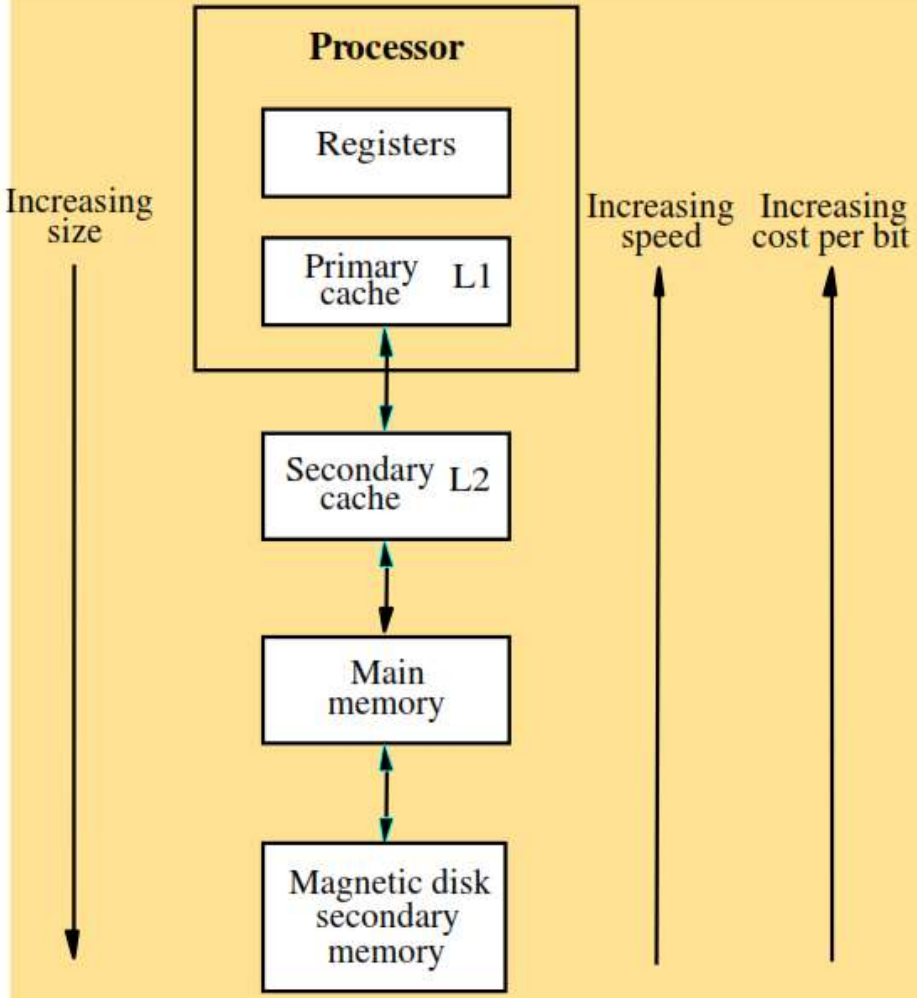
Sr. No.	Flash drives	Hard-disk drives
1.	These are solid state electronic devices.	These are magnetic devices.
2.	They don't have movable parts and hence insensitive to vibration.	They have movable parts. They are sensible to vibrations.
3.	They have shorter seek and access times which results in faster response.	They have comparatively larger seek and access times.
4.	They have lower power consumption and hence suitable for battery driven applications.	They have comparatively large power consumption.
5.	They are available in smaller storage capacity.	Storage capacity is larger.
6.	Higher cost per bit.	Lower cost per bit.
7.	Deterioration rate is high.	Deterioration rate is low.

Speed, Size, and Cost

- A big challenge in the design of a computer system is to provide a sufficiently large memory, with a reasonable speed at an affordable cost.
- Static RAM:
  - Very fast, but expensive, because a basic SRAM cell has a complex circuit making it impossible to pack a large number of cells onto a single chip.
- Dynamic RAM:
  - Simpler basic cell circuit, hence are much less expensive, but significantly slower than SRAMs.
- Magnetic disks:
  - Storage provided by DRAMs is higher than SRAMs, but is still less than what is necessary.
  - Secondary storage such as magnetic disks provide a large amount of storage, but is much slower than DRAMs.



Memory Hierarchy



- *Fastest access is to the data held in processor registers. Registers are at the top of the memory hierarchy.*
- *Relatively small amount of memory that can be implemented on the processor chip. This is processor cache.*
- *Two levels of cache. Level 1 (L1) cache is on the processor chip. Level 2 (L2) cache is in between main memory and processor.*
- *Next level is main memory, implemented as SIMMs. Much larger, but much slower than cache memory.*
- *Next level is magnetic disks. Huge amount of inexepensive storage.*
- *Speed of memory access is critical, the idea is to bring instructions and data that will be used in the near future as close to the processor as possible.*

Cache Memories

In a computer system the program which is to be executed is loaded in the main memory (DRAM). Processor then fetches the code and data from the main memory to execute the program. The DRAMs which form the main memory are slower devices. So it is necessary to insert wait states in memory read/write cycles. This reduces the speed of execution. To speed up the process, high speed memories such as SRAMs must be used. But considering the cost and space required for SRAMs, it is not desirable to use SRAMs to form the main memory. The solution for this problem is come out with the fact that most of the microcomputer programs work with only small sections of code and data at a particular time. In the memory system small section of SRAM is added along with main memory, referred to as **cache memory**. The program which is to be executed is loaded in the main memory, but the part of program (code) and data that work at a particular time is usually accessed from the cache memory. This is accomplished by loading the active part of code and data from main memory to cache memory. The cache controller looks after this swapping between main memory and cache memory with the help of DMA controller. If processor finds that the addressed code or data is not available in cache, then processor accesses that code or data from the

main memory (DRAM). The percentage of accesses where the processor finds the code or data word it needs in the cache memory is called the **hit rate**. The hit rate is normally greater than 90 percent.

$$\text{Hit rate} = \frac{\text{Number of hits}}{\text{Total number of bus cycles}} \times 100\ \%$$

Example 4.5.1 *The application program in a computer system with cache uses 1400 instruction acquisition bus cycle from cache memory and 100 from main memory. What is the hit rate? If the cache memory operates with zero wait state and the main memory bus cycles use three wait states, what is the average number of wait states experienced during the program execution ?*

Solution : $\text{Hit rate} = \frac{1400}{1400+100} \times 100 = 93.3333\ \%$

$$\text{Total wait states} = 1400 \times 0 + 100 \times 3 = 300$$

$$\text{Average wait states} = \frac{\text{Total wait states}}{\text{Number of memory bus cycles}} = \frac{300}{1500} = 0.2$$

Cache Memories

Program Locality

In cache memory system, prediction of memory location for the next access is essential. This is possible because computer systems usually access memory from the consecutive locations. This prediction of next memory address from the current memory address is known as **program locality**. Program locality enables cache controller to get a block of memory instead of getting just a single address.

The principle of program locality may not work properly when program executes JUMP and CALL instructions. In case of these instructions, program code is not in sequence.

Cache Memories

Locality of Reference

We know that program may contain a simple loop, nested loops, or a few procedures that repeatedly call each other. The point is that many instructions in localized area of the program are executed repeatedly during some time period and the remainder of the program is accessed relatively infrequently. This is referred to as **locality of reference**. It manifests itself in two ways : **temporal** and **spatial**. The temporal means that a recently executed instruction is likely to be executed again very soon. The spatial means that instructions stored near by to the recently executed instruction are also likely to be executed soon.

The temporal aspect of the locality of reference suggests that whenever an instruction or data is first needed, it should be brought into the cache and it should remain there until it is needed again. The spatial aspect suggests that instead of bringing just one instruction or data from the main memory to the cache, it is wise to bring several instructions and data items that reside at adjacent address as well. We use the term block to refer to a set of contiguous addresses of some size.

Cache Memories

Block Fetch

Block fetch technique is used to increase the hit rate of cache. A block fetch can retrieve the data located before the requested byte (look behind) or data located after the requested byte (look ahead), or both. When CPU needs to access any byte from the block, entire block that contains the needed byte is copied from main memory into cache.

The size of the block is one of the most important parameters in the design of a cache memory system. The following section describes the pros and cons of small block size and large block size.

Cache Memories

PROS and CONS : size of the block

1. If the block size is too small, the look ahead and look-behind are reduced and therefore the hit rate is reduced.

2. Larger blocks reduce the number of blocks that fit into a cache. As the number of blocks decrease, block rewrites from main memory becomes more likely.
3. Due to large size of block the ratio of required data and useless data is less.
4. Bus size between the cache and the main memory increases with block size to accommodate larger data transfers between main memory and the cache, which increases the cost of cache memory system.

Elements of Cache Design

The cache design elements include cache size, mapping function, replacement algorithm write policy, block size and number of caches.

Cache Size : The size of the cache should be small enough so that the overall average cost per bit is close to that of main memory alone and large enough so that the overall average access time is close to that of the cache alone.

Mapping function : The cache memory can store a reasonable number of blocks at any given time, but this number is small compared to the total number of blocks in the main memory. Thus we have to use mapping functions to relate the main memory blocks and cache blocks. There are two mapping functions commonly used : direct mapping and associative mapping.

Replacement Algorithm : When a new block is brought into the cache, one of the existing blocks must be replaced, by a new block.

There are four most common replacement algorithms :

- Least-Recently Used (LRU)
- First-In-First-Out (FIFO)
- Least-Frequently-Used (LFU)
- Random

Cache design change according to the choice of replacement algorithm.

Write Policy : It is also known as cache updating policy. In cache system, two copies of the same data can exist at a time, one in cache and one in main memory. If one copy is altered and other is not, two different sets of data become associated with the same address. To prevent this the cache system has updating systems such as : write through system, buffered write through system and write-back system. The choice of cache write policy also change the design of cache.

Mapping Functions

Usually, the cache memory can store a reasonable number of blocks at any given time, but this number is small compared to the total number of blocks in the main memory. The correspondence between the main memory blocks and those in the cache is specified by a **mapping function.**

There are two main mapping techniques which decides the cache organisation :

1. Direct-mapping technique
2. Associative-mapping technique

The associative mapping technique is further classified as fully associative and set associative techniques.