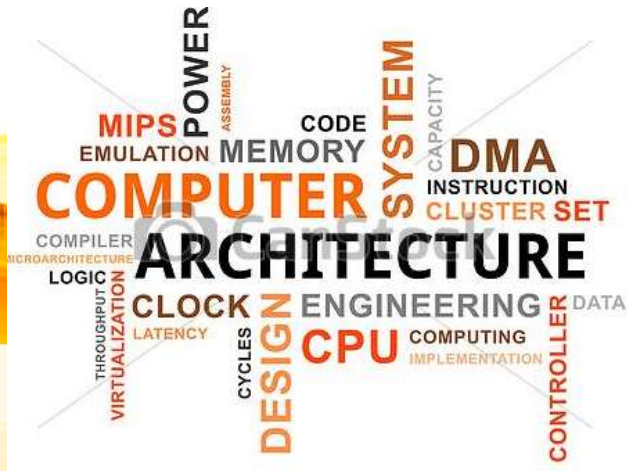
# UNIT IV
# MEMORY SYSTEM

**Basic concepts of Semiconductor RAMs - ROMs** – Speed, Size and Cost – Cache memories – Performance consideration – Virtual memory – Memory Management requirements – Secondary storage.

Case Study: Memory Organization in Multiprocessors



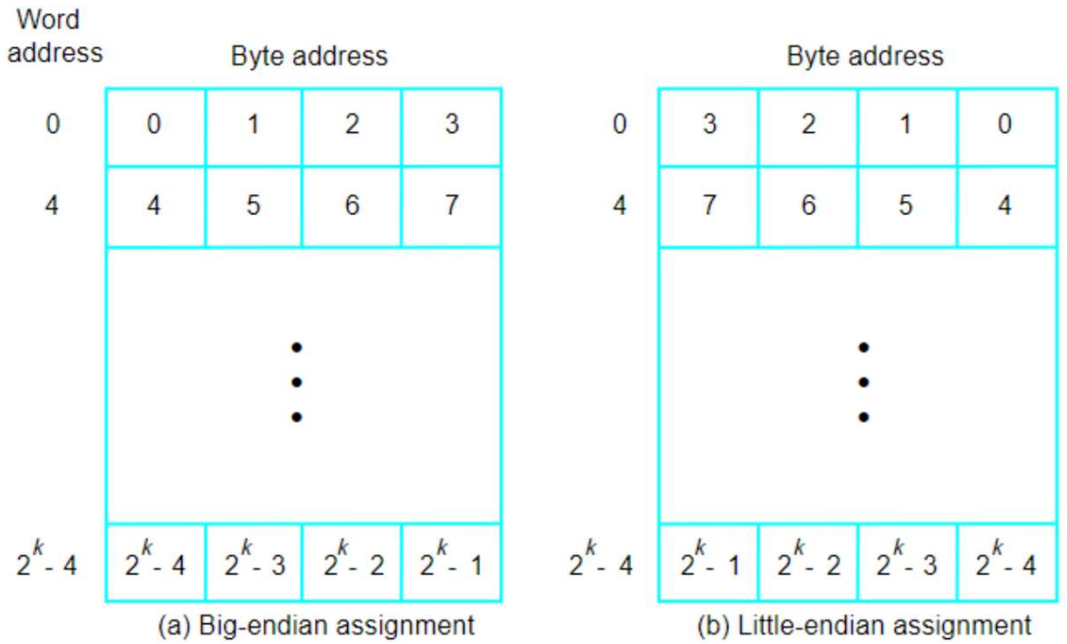
# Recap the previous Class

# Overview

- Basic Memory Circuit
- Organization of Main Memory
- Cache Memory
- Virtual Memory
- Secondary Storage

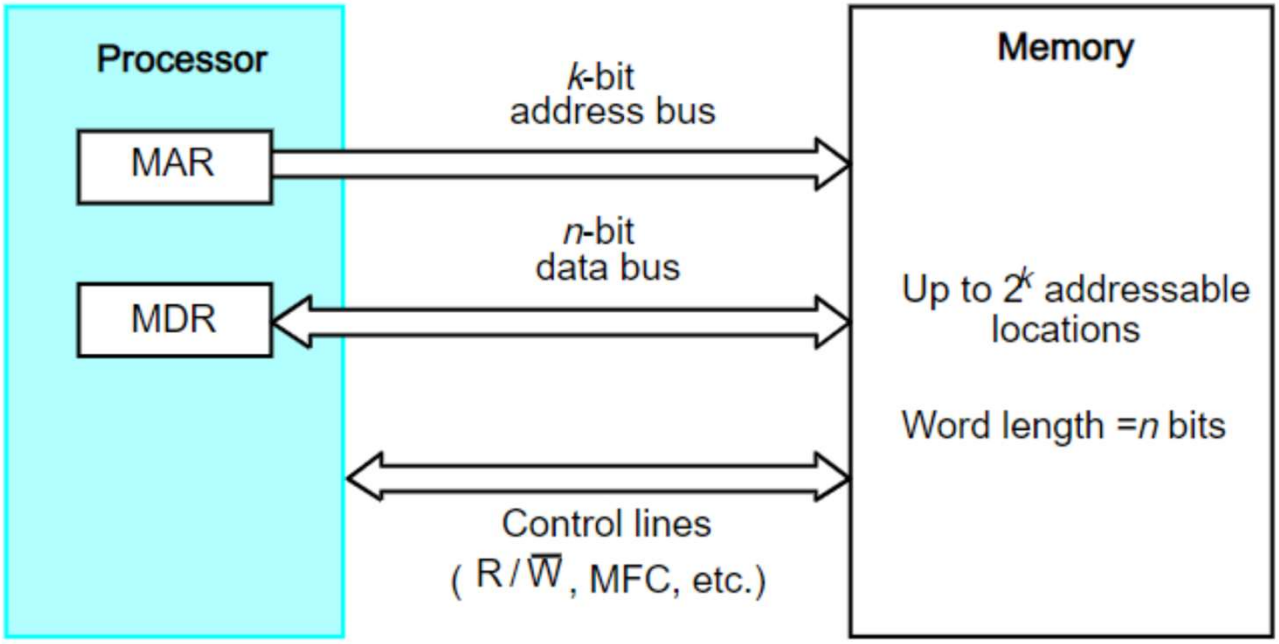
# Basic Concepts

- Maximum size of the memory that can be used in any computer determined by the addressing scheme.
  - 16 bit addresses = $2^{16}$ = 64 k Memory Location
- Modern Computers are byte addressable.



# Traditional Architecture

Connection of the memory to the processor



# Basic Concepts

- Block transfer – Bulk data transfer
- Memory access time
- Memory Cycle time
- RAM – 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
- Cache Memory
- Virtual Memory, Memory management Unit

# Semiconductor RAM Memories

| Parameter | SRAM | DRAM |
|---|---|---|
| **Full Form** | Static Random Access Memory | Dynamic Random Access Memory |
| **Read & Write speed** | Faster | Slower than SRAM |
| **Storage component** | Uses transistor to store single bit of information | Uses separate capacitor to store each bit of data |
| **Price** | Expensive than DRAM | Economical than SRAM |
| **Power consumption** | More | Less |
| **Refresh** | No need to refresh for maintaining data | Needs to be refreshed thousands of time every second |
| **Used in** | Cache memory | Main memory |
| **Internal structure** | Complex | Simpler than SRAM |
| **Density** | Less dense | Highly dense |
| **Storage per bit** | Can store many bits per chip | Cannot store many bits per chip |

# Internal Organization of Memory Chips

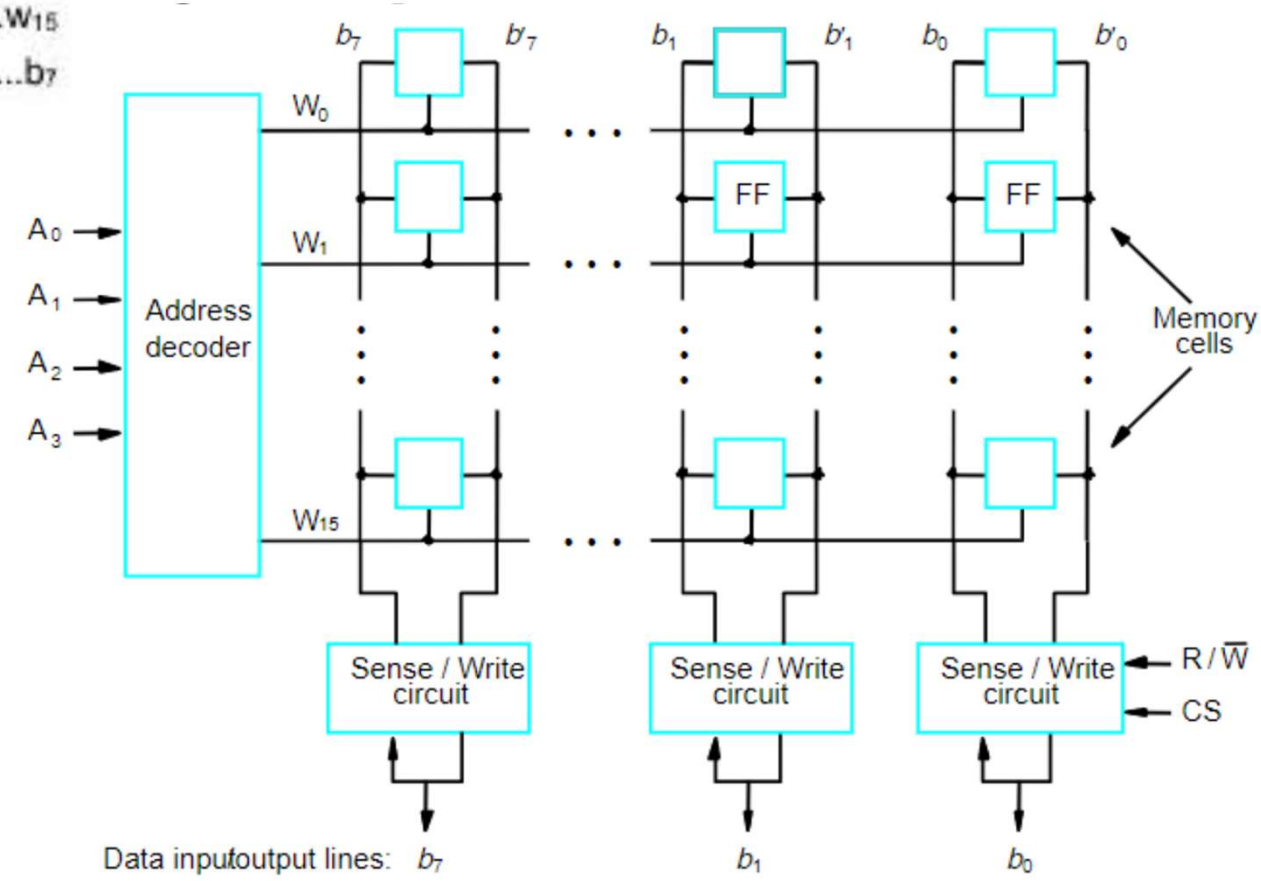
16 memory location $w_0$, $w_1$,.....$w_{15}$

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

16 words of 8 bits each: 16x8 memory org.. It has 16 external connections: addr. 4, data 8, control: 2, power/ground: 2

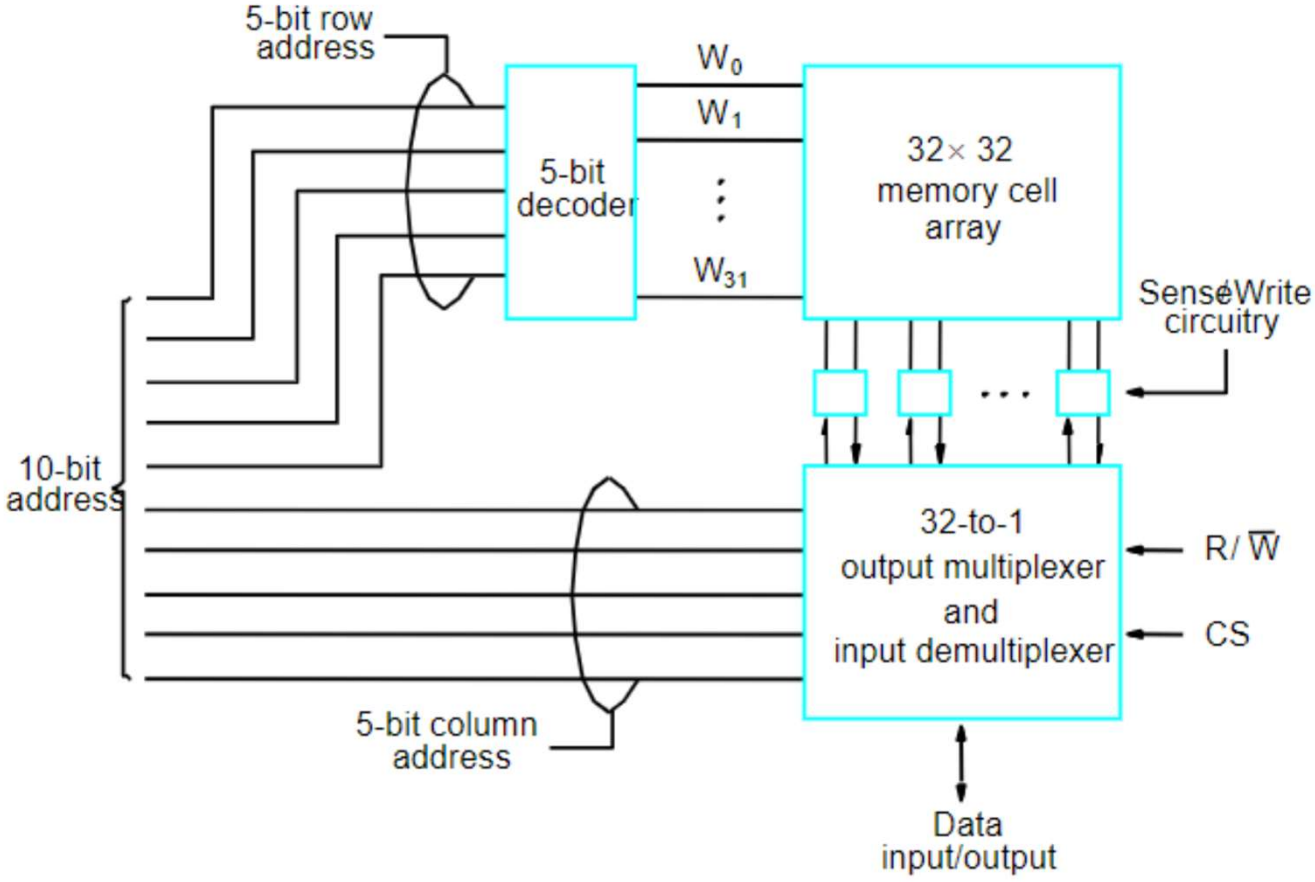
1K memory cells: 128x8 memory, external connections: ? 19(7+8+2+2)

1Kx1:? 15 (10+1+2+2)



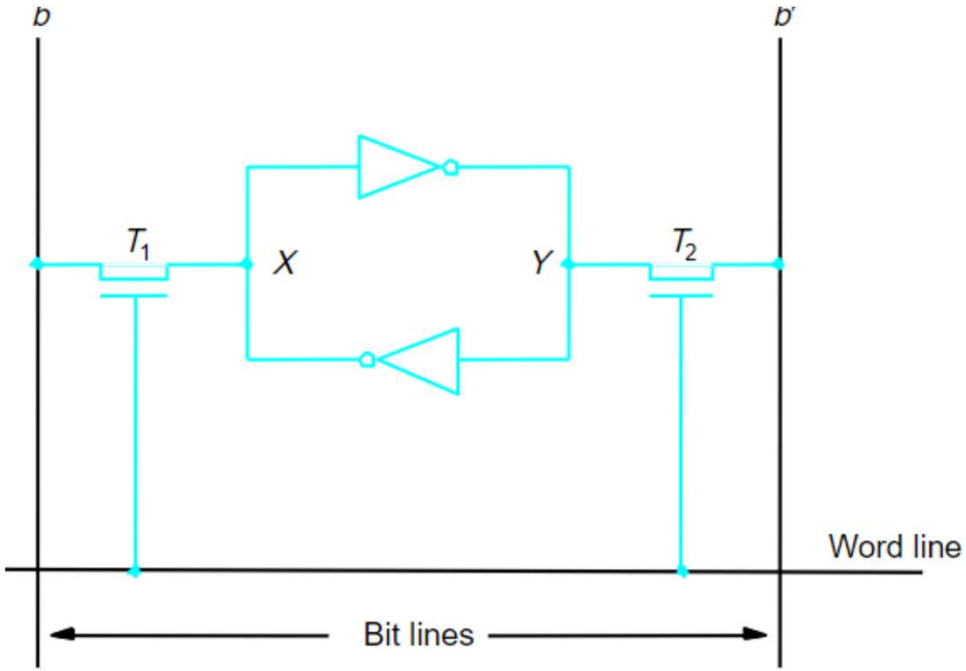
# A Memory Chip

Organization of a 1K x 1 Memory Chip

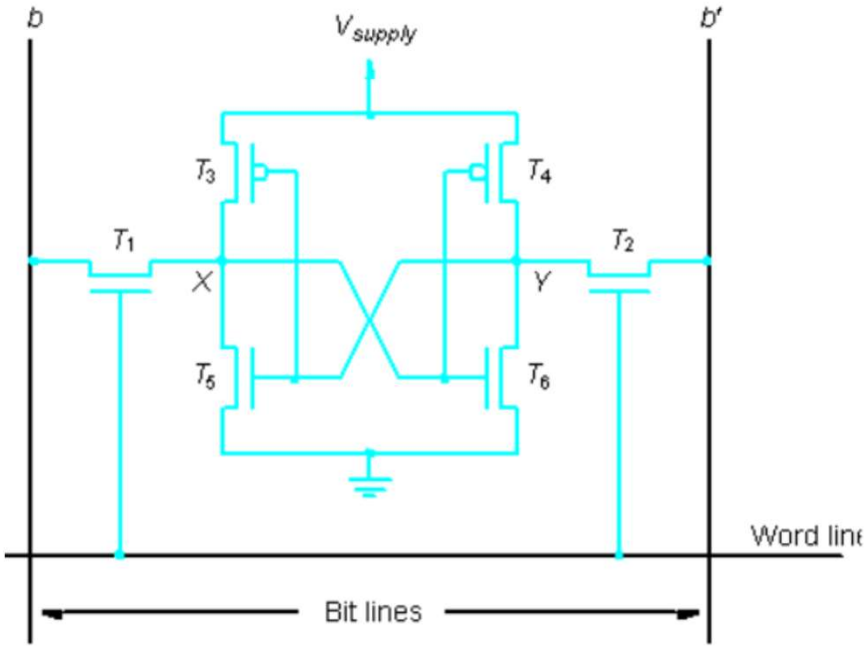


# Static Memories

- The circuits are capable of retaining their state as long as power is applied

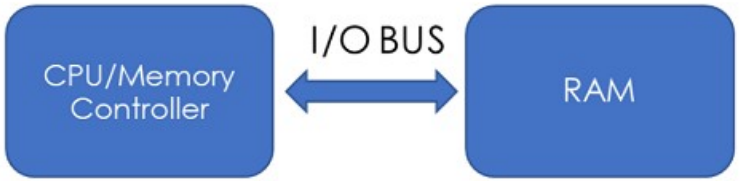



CMOS Memory Cell

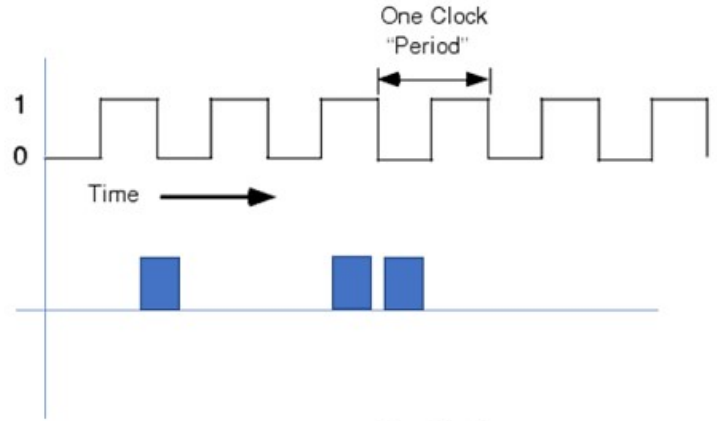


- **Asynchronous**
  - The processor timing and the memory timing (refreshing schedule) were independent. Thus the processor might have to wait until the memory "window" was open for access.
- **Synchronous (SDRAM)**
  - The processor and memory timing are linked. This allows for more efficient processor-memory interaction.
  - *Note SDRAM is NOT static dynamic RAM – that doesn't make sense*

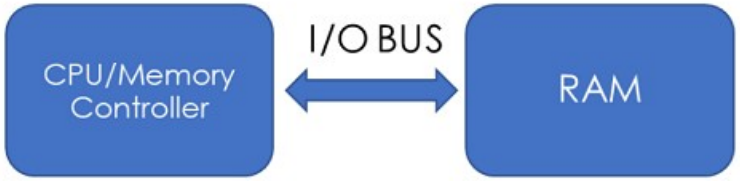
Asynchronous DRAM



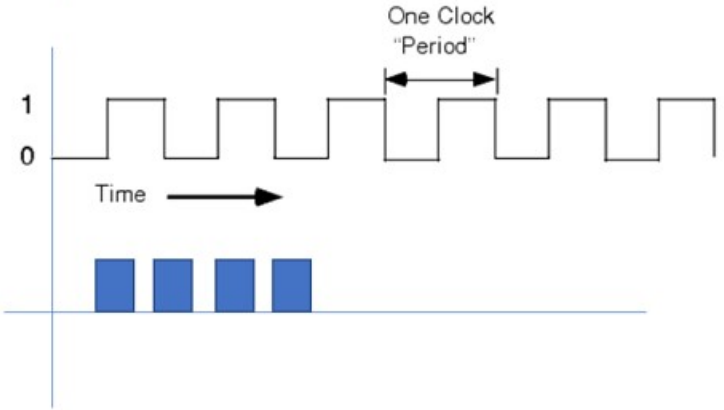

RAM is not synchronized with CPU Clock




Synchronous DRAM

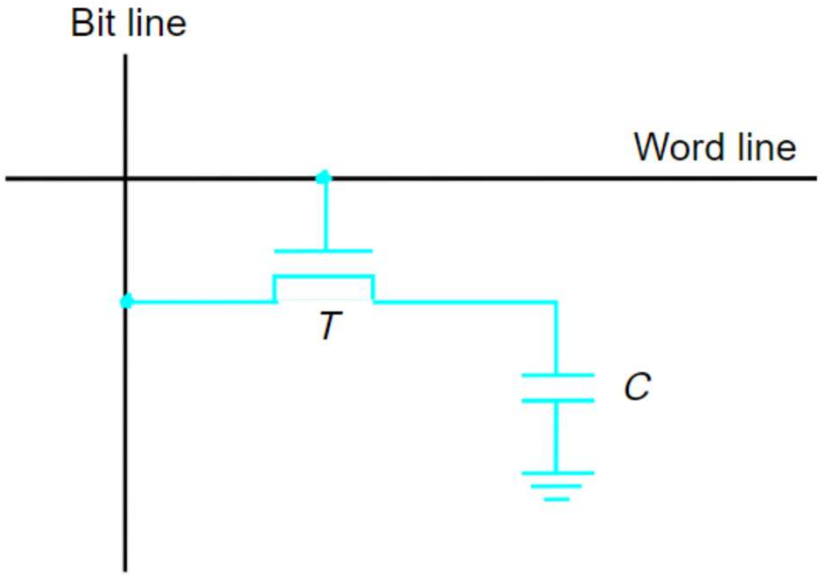


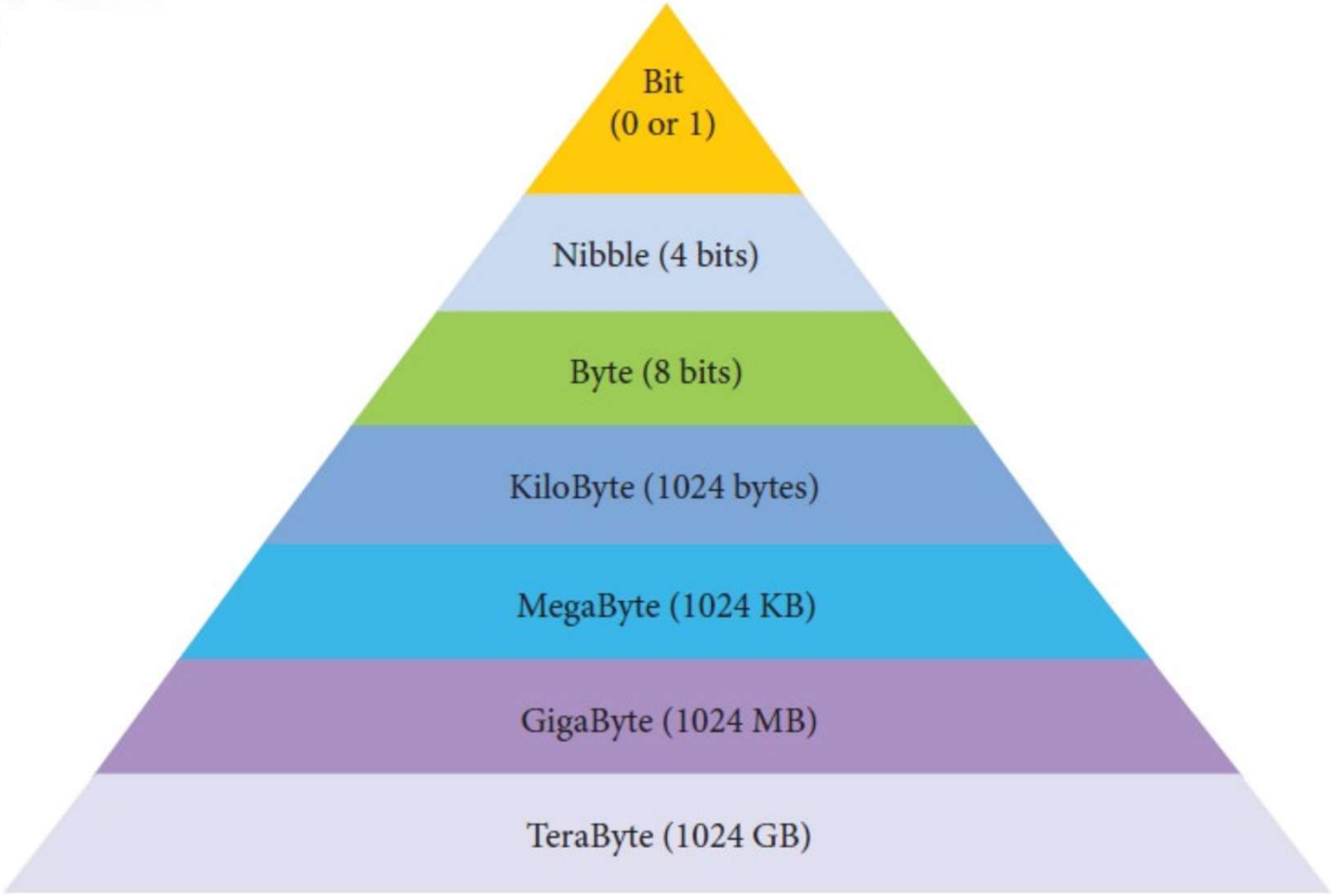

RAM is synchronized with CPU Clock



# DRAM

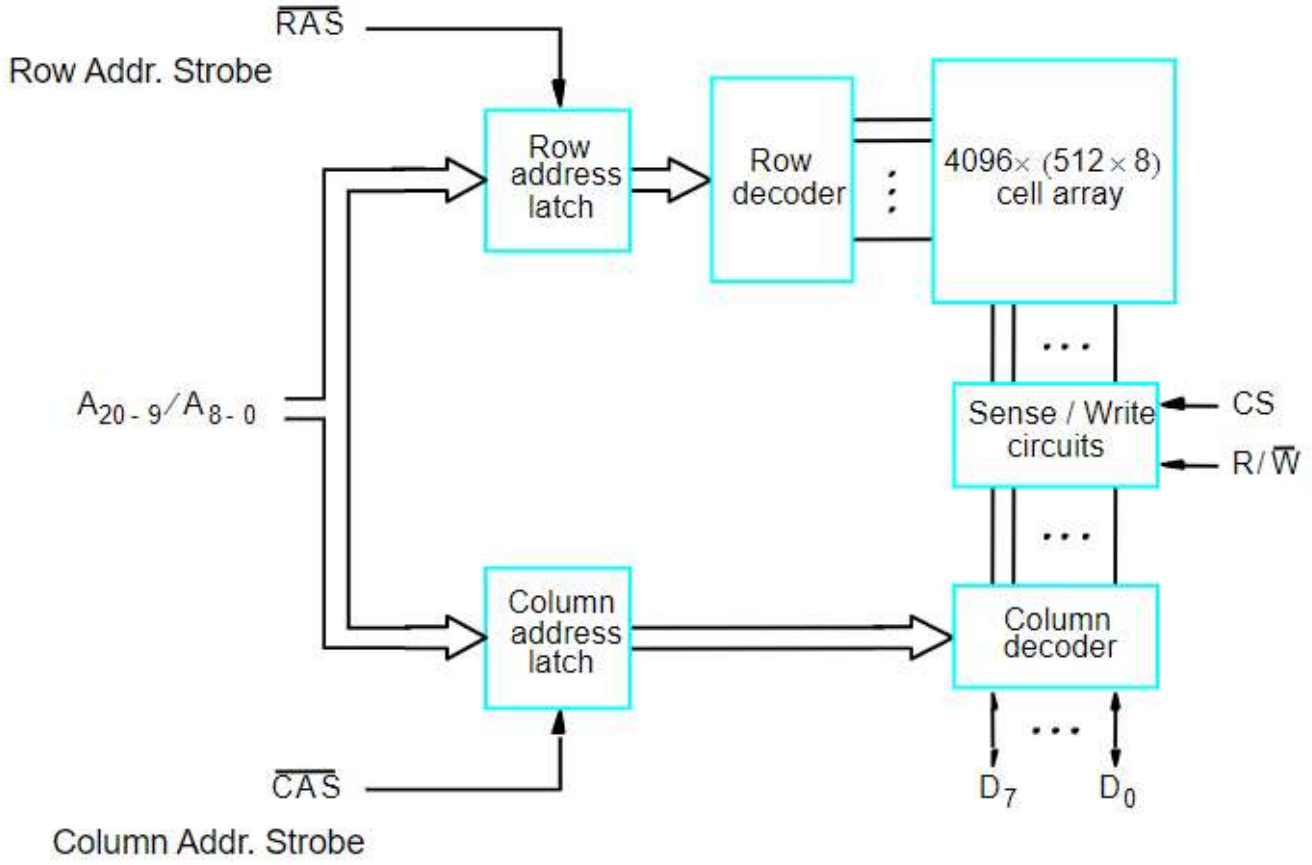
- Static RAMs are fast, but they cost more and more expensive
- Dynamic RAMs (DRAMs) are Cheap but they cannot retain their state – need periodically refreshed




Bit
(0 or 1)
Nibble (4 bits)
Byte (8 bits)
KiloByte (1024 bytes)
MegaByte (1024 KB)
GigaByte (1024 MB)
TeraByte (1024 GB)

# Dynamic Memory Chip

Internal organization of a 2M × 8 dynamic memory chip.

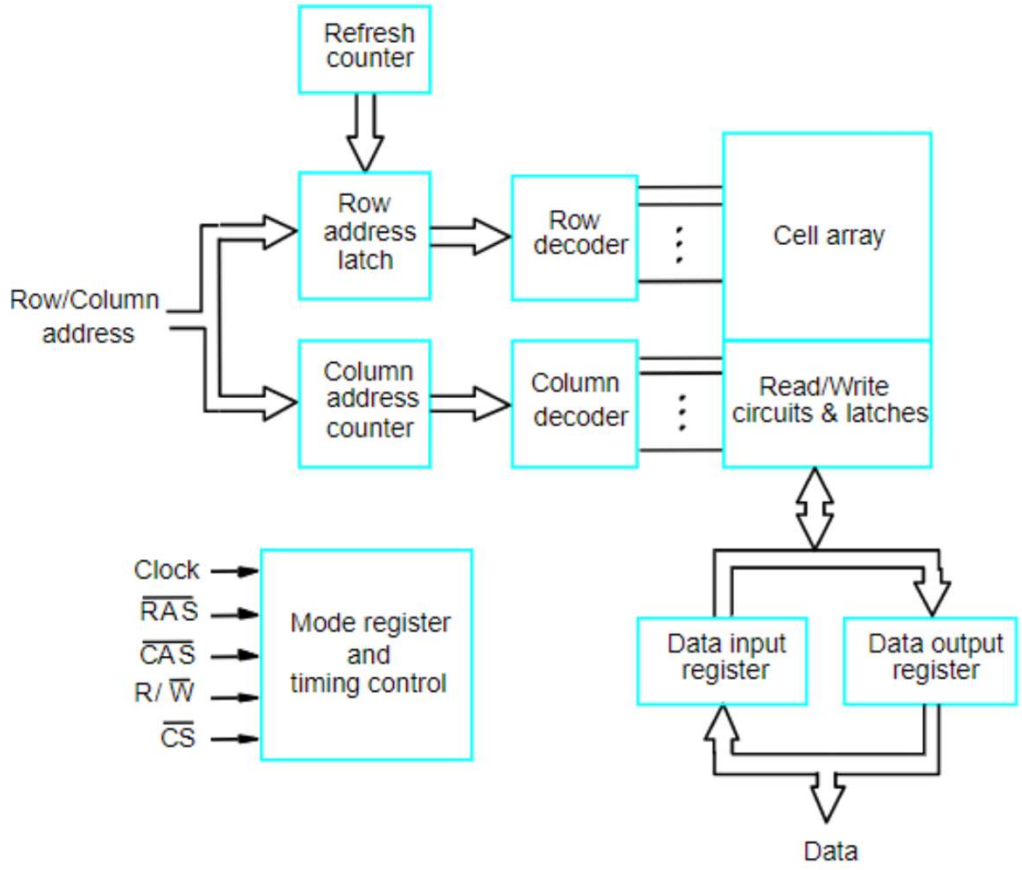


# Fast Page Mode

- When the DRAM in last slide is accessed, the content of all 4096 cells are selected row are sensed, but only 8bits are placed on the data lines $D_{7-0}$ , as selected by $A_{8-0}$
- Make it possible to access the other bytes in the same row without having to reselect the row
- A latch is added at the output of the sense amplifier in each column
- Gook for bulk transfer

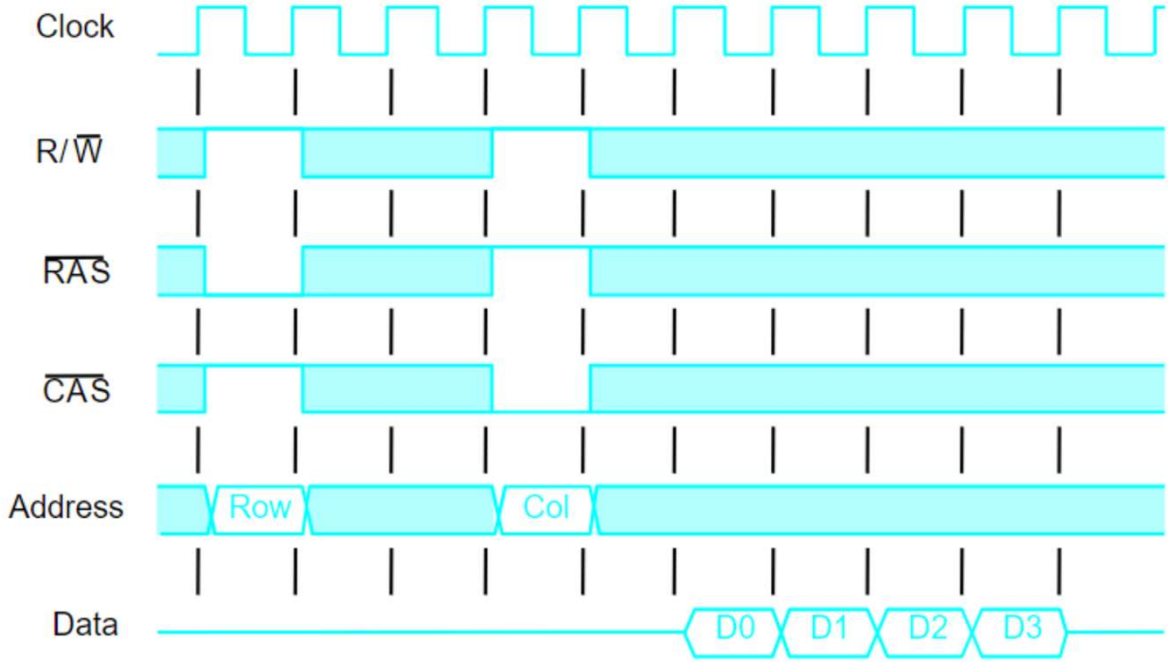
# Synchronous DRAM

- The operations of SDRAM controlled by a clock signal



# Synchronous DRAM

- No CAS pulses is needed in burst operation.
- Refresh circuits are included (every 64ms).
- Clock frequency > 100 MHz
- Intel PC100 and PC133



# Latency and Bandwidth

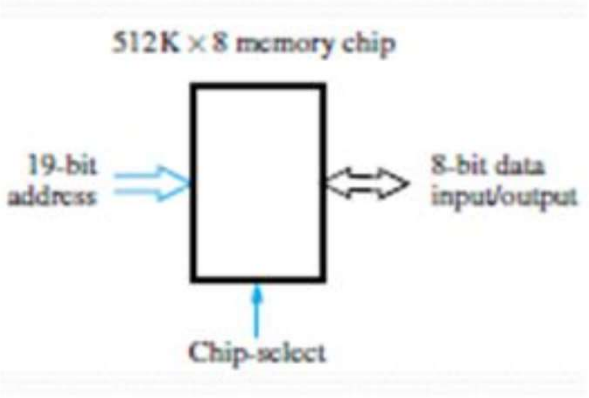
- Memory Latency – Time taken to transfer a word to or from the memory
- Memory bandwidth – the number of bits or bytes that can be transferred in one seconds
- It is used to measure how much time is needed to transfer an entire block of data

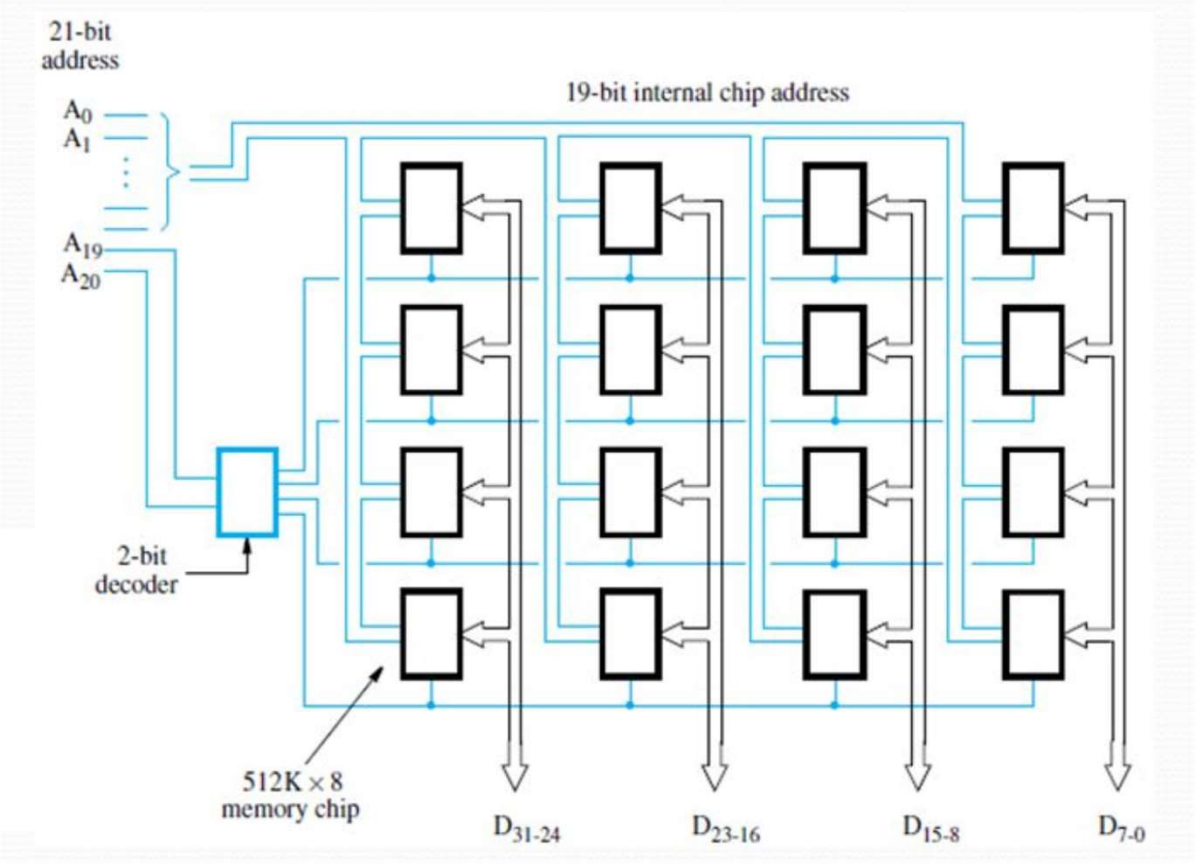
# DDR SDRAM

- Double-Data-Rate SDRAM
- Standard SDRAM performs all actions on the rising edge of the clock signal.
- DDR SDRAM accesses the cell array in the same way, but transfers the data on both edges of the clock.
- The cell array is organized in two banks. Each can be accessed separately.
- DDR SDRAMs and standard SDRAMs are most efficiently used in applications where block transfers are prevalent.

# Structures of large Memories

Organization of a 2M × 32 memory module using 512K × 8 static memory chips.



# Memory System Consideration

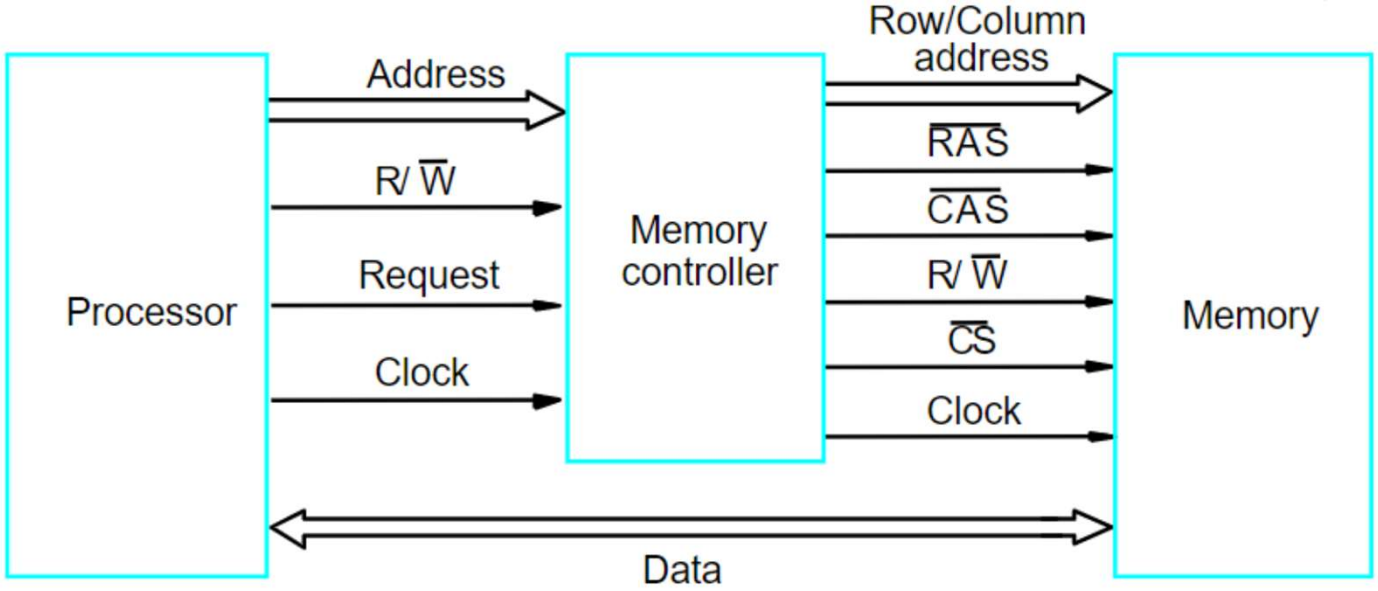
- The choice of RAM Chip of an application depends on
  - Cost, Speed, Size, Power, etc.,
- SRAM – Faster, Less Expensive and Smaller
- DRAM – Slower, Cheaper, Larger

# Memory System Consideration

Which one for cache and main memory, respectively?

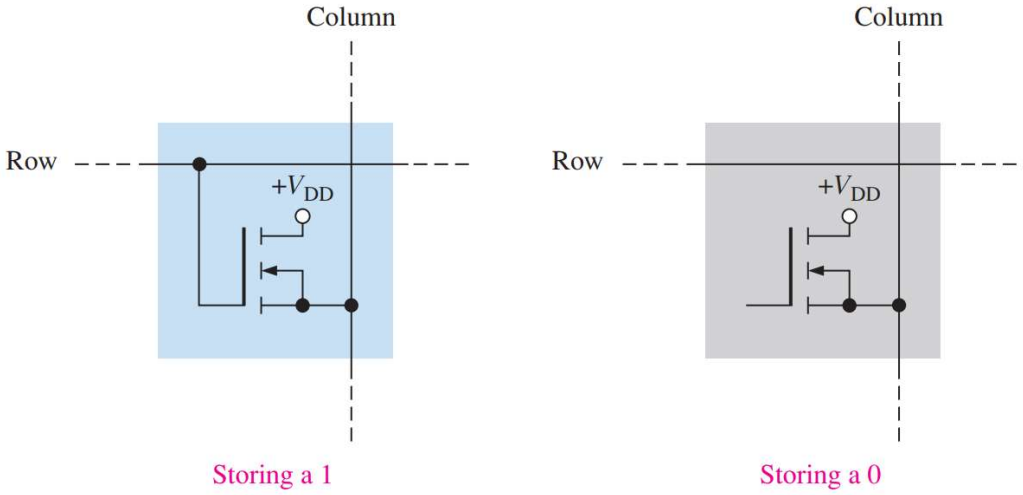
Refresh overhead – suppose a SDRAM whose cells are in 8K rows; 4 clock cycles are needed to access each row; then it takes 8192×4=32,768 cycles to refresh all rows; if the clock rate is 133 MHz, then it takes $32,768/(133\times10^{-6})=246\times10^{-6}$ seconds; suppose the typical refreshing period is 64 ms, then the refresh overhead is 0.246/64=0.0038<0.4% of the total time available for accessing the memory.

# Memory Controller

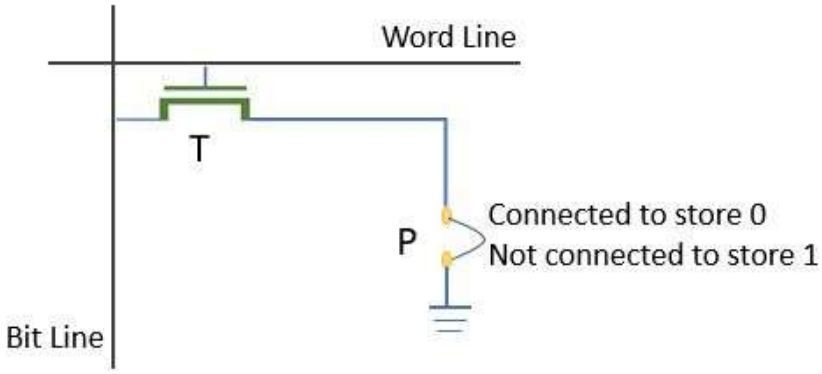


# ROM

- Volatile / Non Volatile ?
- Read only Memory
- PROM – Programmable ROM
- EPROM – Erasable , Reprogrammable ROM
- EEPROM – Can be Programmed and Erased Electrically




ROM cells.

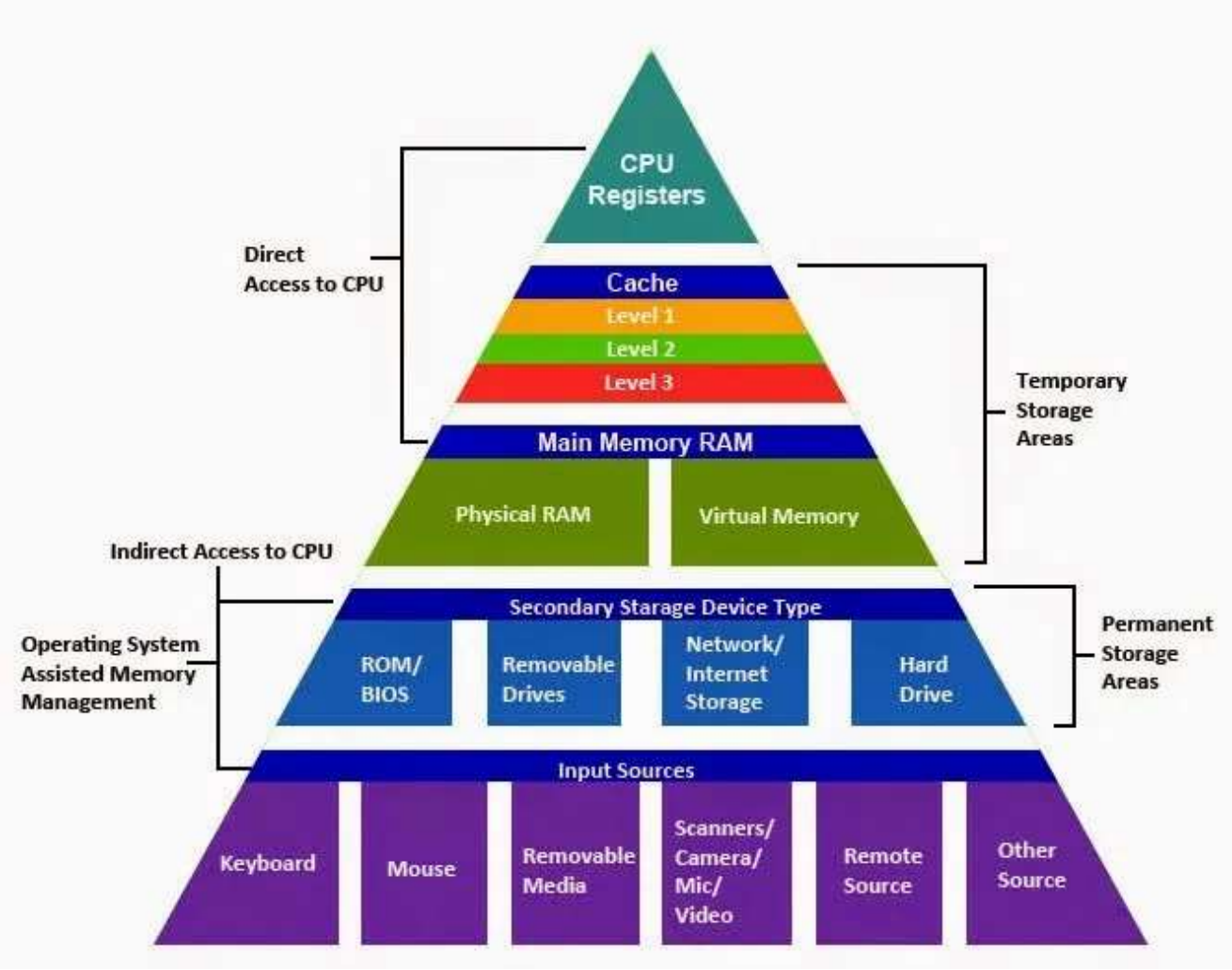


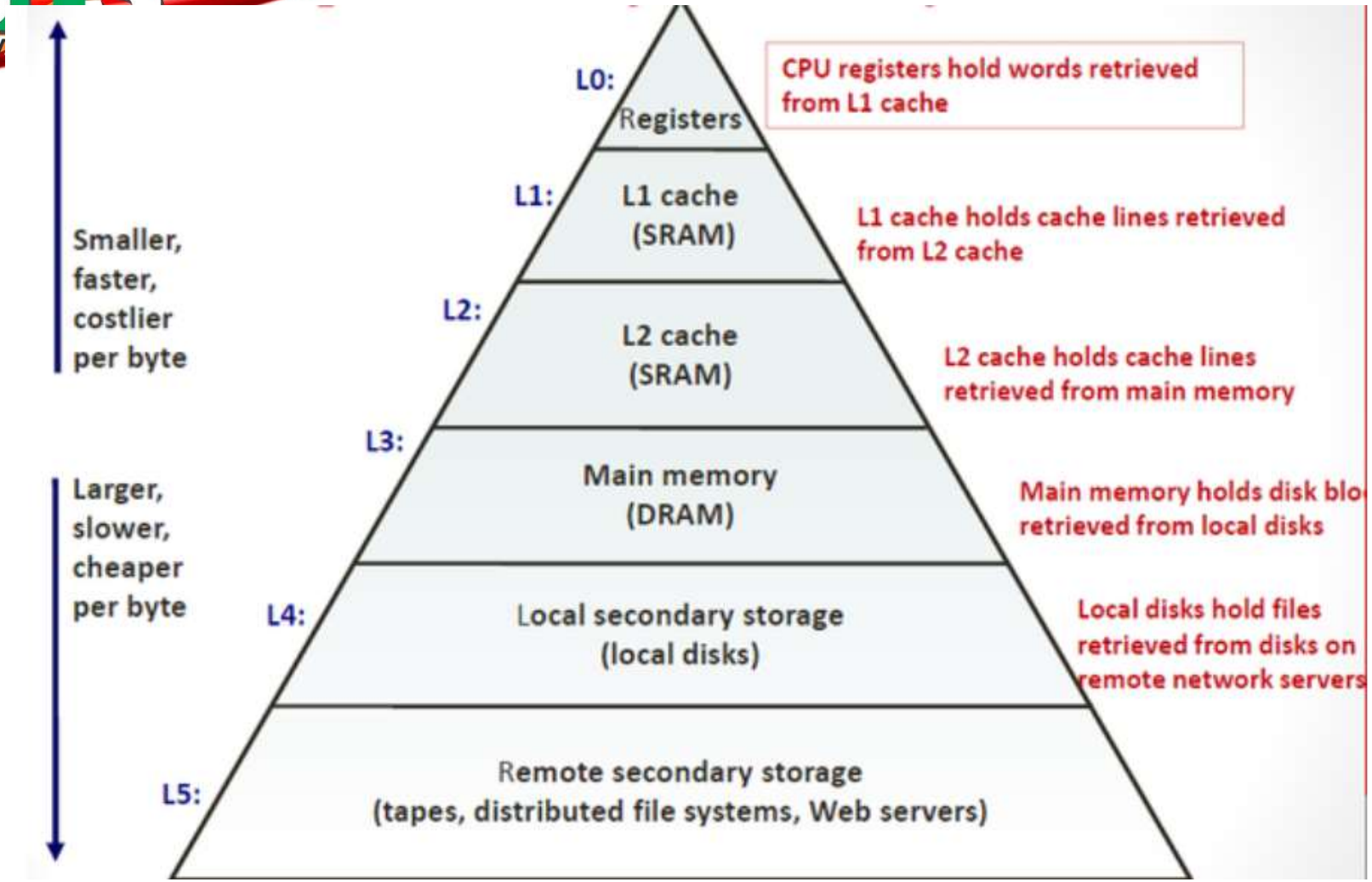

A ROM Memory Cell

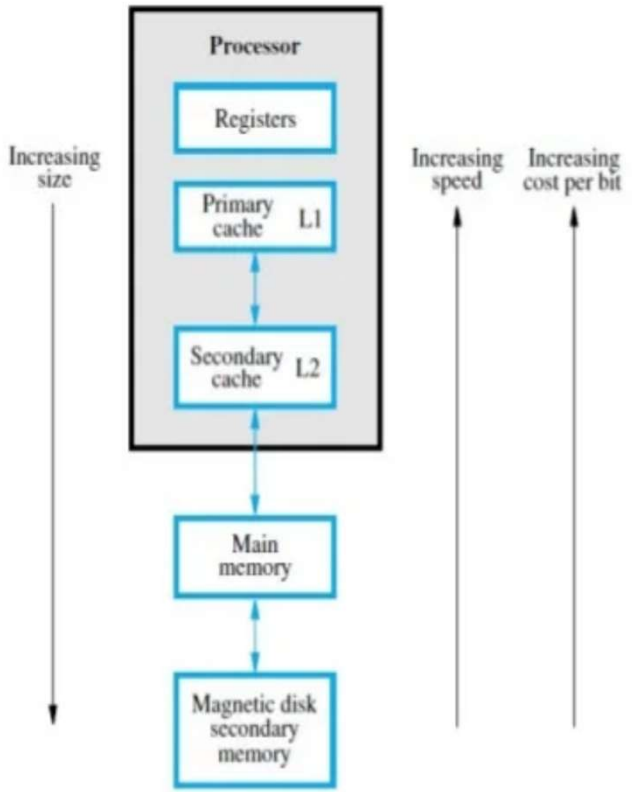
# Flash Memory

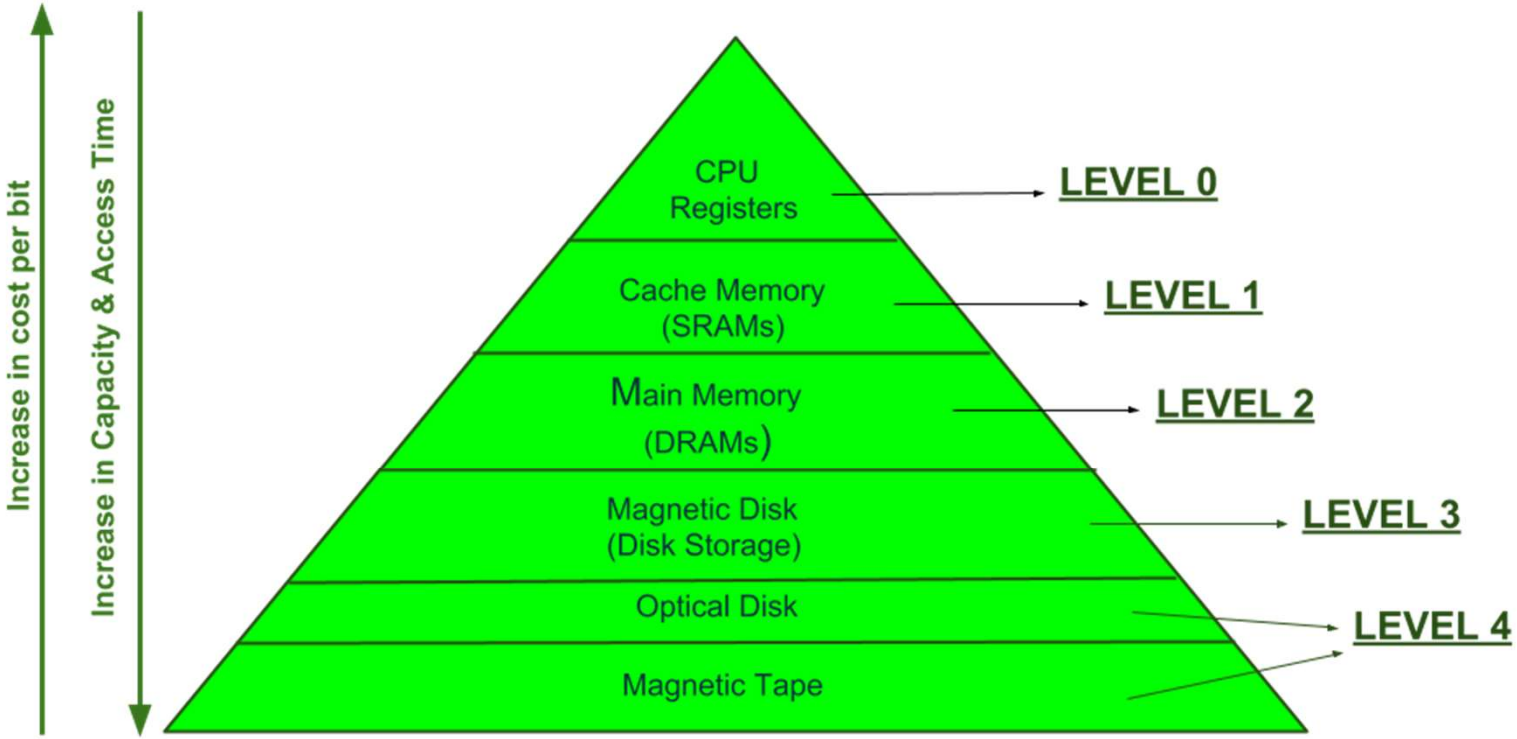
- Similar to EEPROM
- Only possible to write an entire block of  Cell instead of single block Cell
- Low Power
- Used in Portable Equipments
- Modules -  Flash Cards and Drivers

# Memory Hierarchy




Smaller, faster, costlier per byte
Larger, slower, cheaper per byte
L0:
Registers
CPU registers hold words retrieved from L1 cache
L1:
L1 cache (SRAM)
L1 cache holds cache lines retrieved from L2 cache
L2:
L2 cache (SRAM)
L2 cache holds cache lines retrieved from main memory
L3:
Main memory (DRAM)
Main memory holds disk blo
retrieved from local disks
L4:
Local secondary storage (local disks)
Local disks hold files retrieved from disks on remote network servers
L5:
Remote secondary storage (tapes, distributed file systems, Web servers)




## MEMORY HIERARCHY DESIGN


Thank You