UNIT-IV

The Memory Organization

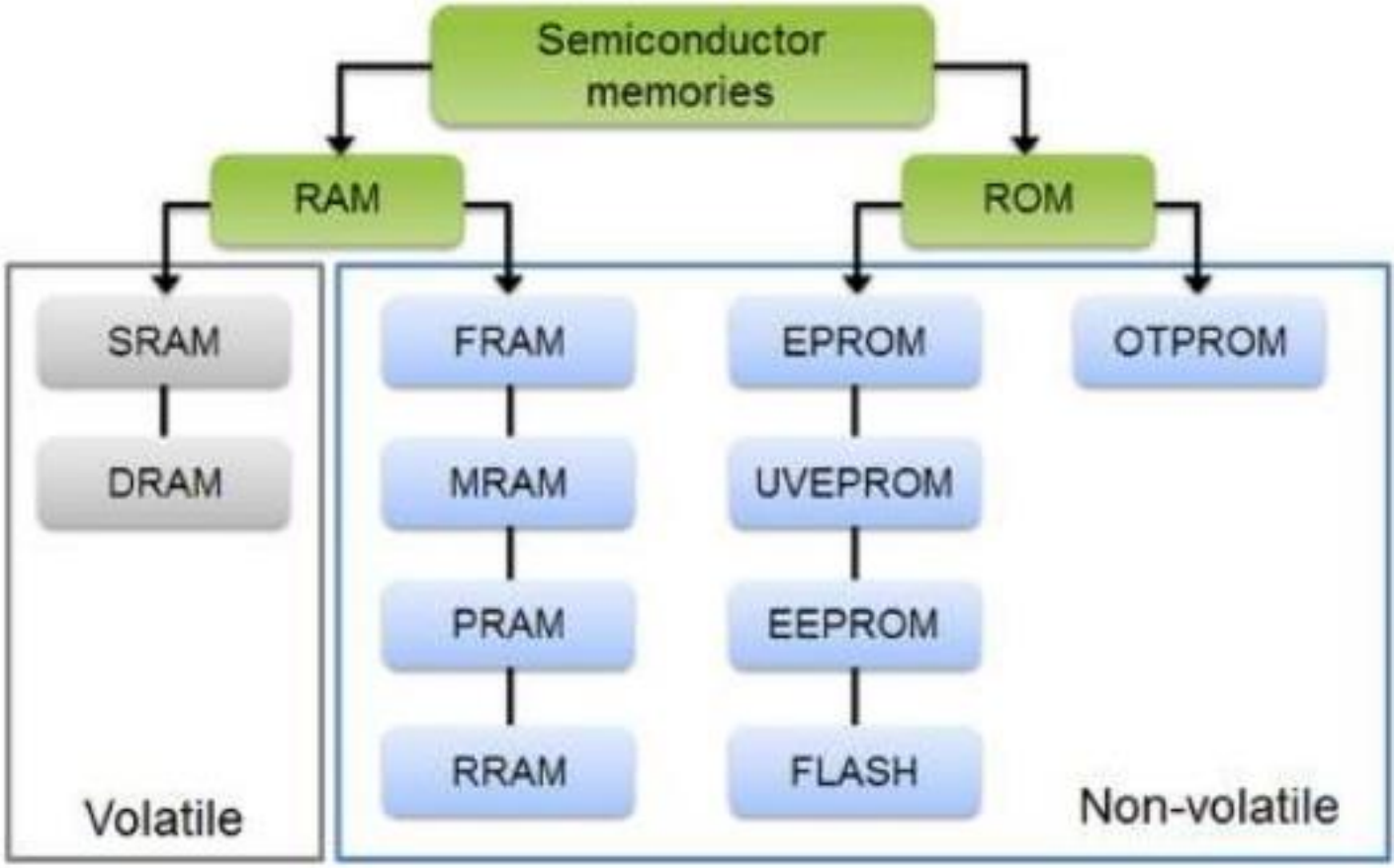
SEMICONDUCTOR RAM MEMORIES

BY
R.LIKITHA

CONTENTS:

- Semiconductor memories
- Definition
- Uses
- Types of RAM
- Advantages and disadvantages

SEMICONDUCTOR MEMORIES



RANDOM ACCESS MEMORY

- **Definition:**
  - It is a form of data storage that can be accessed randomly at any time in any order and form of any physical location.
- **USE:**
  - It allows the computer to read data quickly to run applications.
  - It allows reading and writing .

RANDOM ACCESS MEMORY

- **VOLATILITY:**
  - It is a voitile it contents are lost when the device is powered-off
- TYPES:
  - The two main types RAM are
    - static RAM
    - dynamic RAM

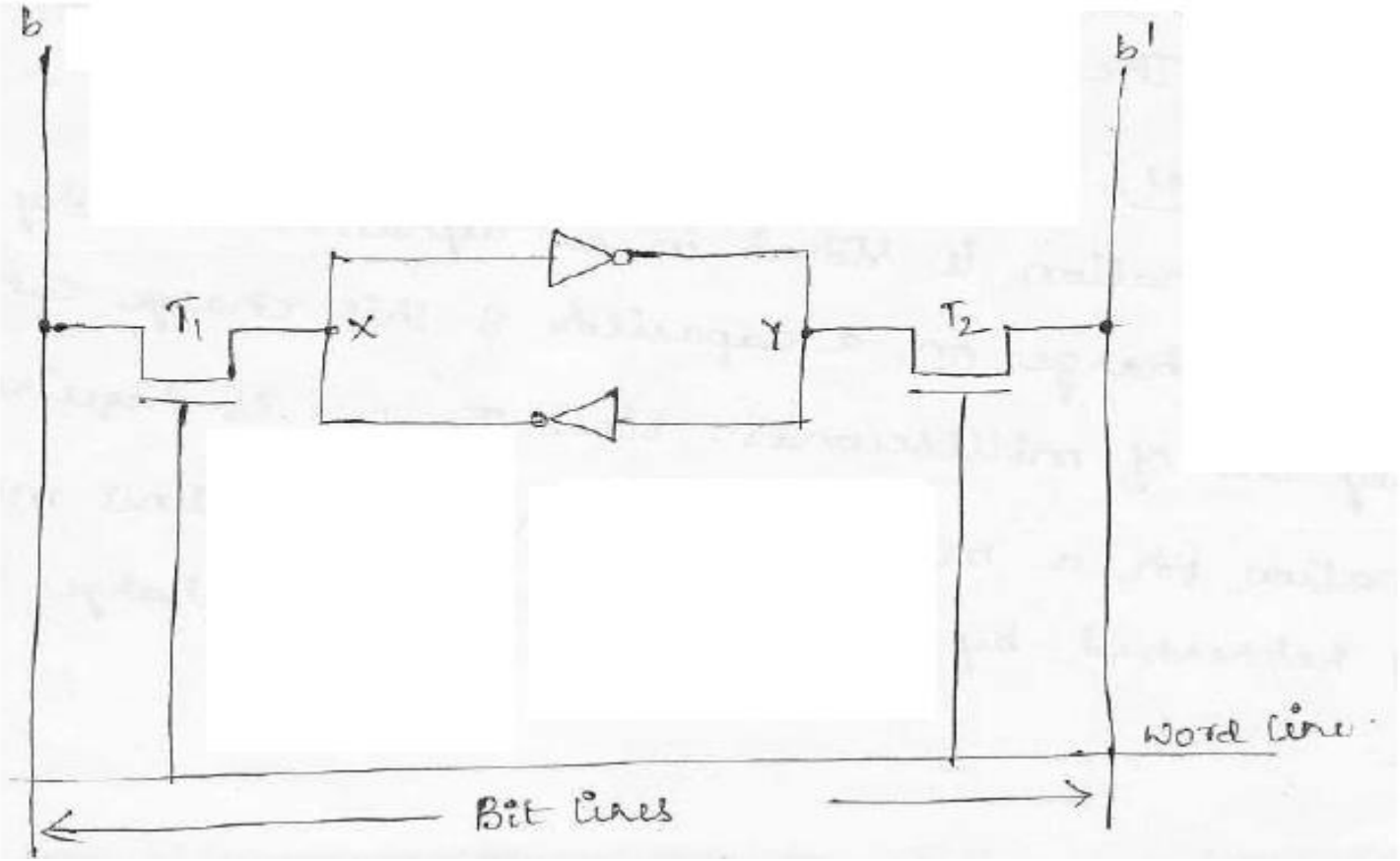
STATIC RAM

- In a static RAM binary values are stored using traditional flip-flop logic gate configurations.
- A SRAM will holds its data as long as power is supplied to it.
- SRAM is used as CACHE memory.

DYNAMIC RAM

- A DRAM is made with cells that store data as charge on capacitors.
- DRAM's are among the VLSI cirrcuit interms of transistor perchip.

THE BELOW FIG SHOWS THE SRAM CAN BE IMPLEMENTED AS



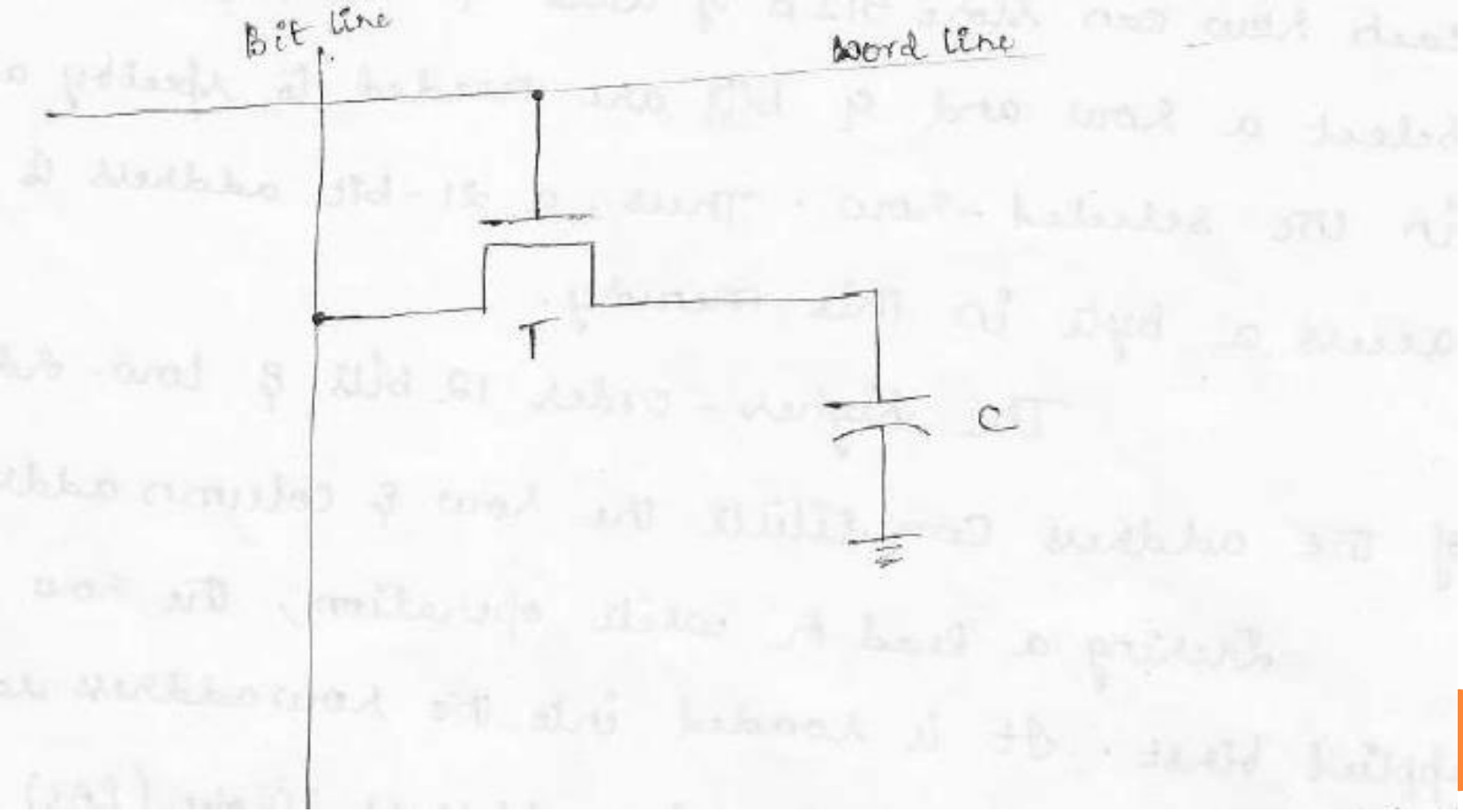
- The two invertors are cross-connected to form a latch.
- The latch is connected to two bit lines by transistors T1 and T2.
- These transistors act as switches that can be opened or closed under the control of wordline.
- When the wordline is ground level and the transistor are turned-off and the latch retains its state.
- eg: let us assume that cell is state1 if the logic value at point X is 1 and at point Y is 0.

Read operation:

- To read in SRAM cell,the wordline is activated to close switches T1 & T2.
- If the cell is in state1, the signal on bitline b is high and the signal on bitline b^1 is low.

- **Write opration:**
  - The state of the cell is set by placing the appropriate value on bitline ‘b’ and its complement on b^1 and the activating the wordline.

The below fig shows the dynamic memory cell consist of transistor and capacitor



- In order to store information in this cell the transistor T is turned on and an appropriate voltage is applied to the bitline.
- This causes a known amount of charge to be stored in the capacitor .
- After the transistor is turned off the capacitor begins to discharge

Advantages of SRAM:

- Fastest type of memory
- Low density

Disadvantage of SRAM:

- More expensive

- Generally used in small amounts(L2 cache) or expensive servers.

Advantages of DRAM:

- It has high density and low cost.
- These are widely used.
- Available chips are of size1M to 256M bits and large chips are developed.
- A DRAM chip is organised to read or write a number of bits in parallel.
- It provides flexibility in designing memory system.

Read only memory

- Introduction
- Meaning of Read Only Memory
- Types of ROM
- Features of ROM
- Use of ROM
- Basic Structure of ROM

Read-Only Memory

Introduction

- Computer memory is the storage space in computer where data is to be processed and instructions required for processing are stored.
- Memory is divided into large number of small parts called cells.

Meaning of ROM

- **Read-only memory** (ROM), also known as **firmware(software)**, is an integrated circuit programmed with specific data when it is manufactured.
- ROM chips are used not only in computers, but in most other electronic items as well.

ROM Types

There are FIVE basic ROM types:

- PROM
- EPROM
- EEPROM
- EAROM
- Flash memory

- Each type has unique characteristics, but they are all types of memory with two things in common:

Read-Only Memory - Features

- Data stored in these chips is **nonvolatile** -- Data is not lost when power is removed.
- Data stored in these chips is either **unchangeable** or requires a special operation to change.
- This means that removing the power source from the chip will not cause it to lose any data.

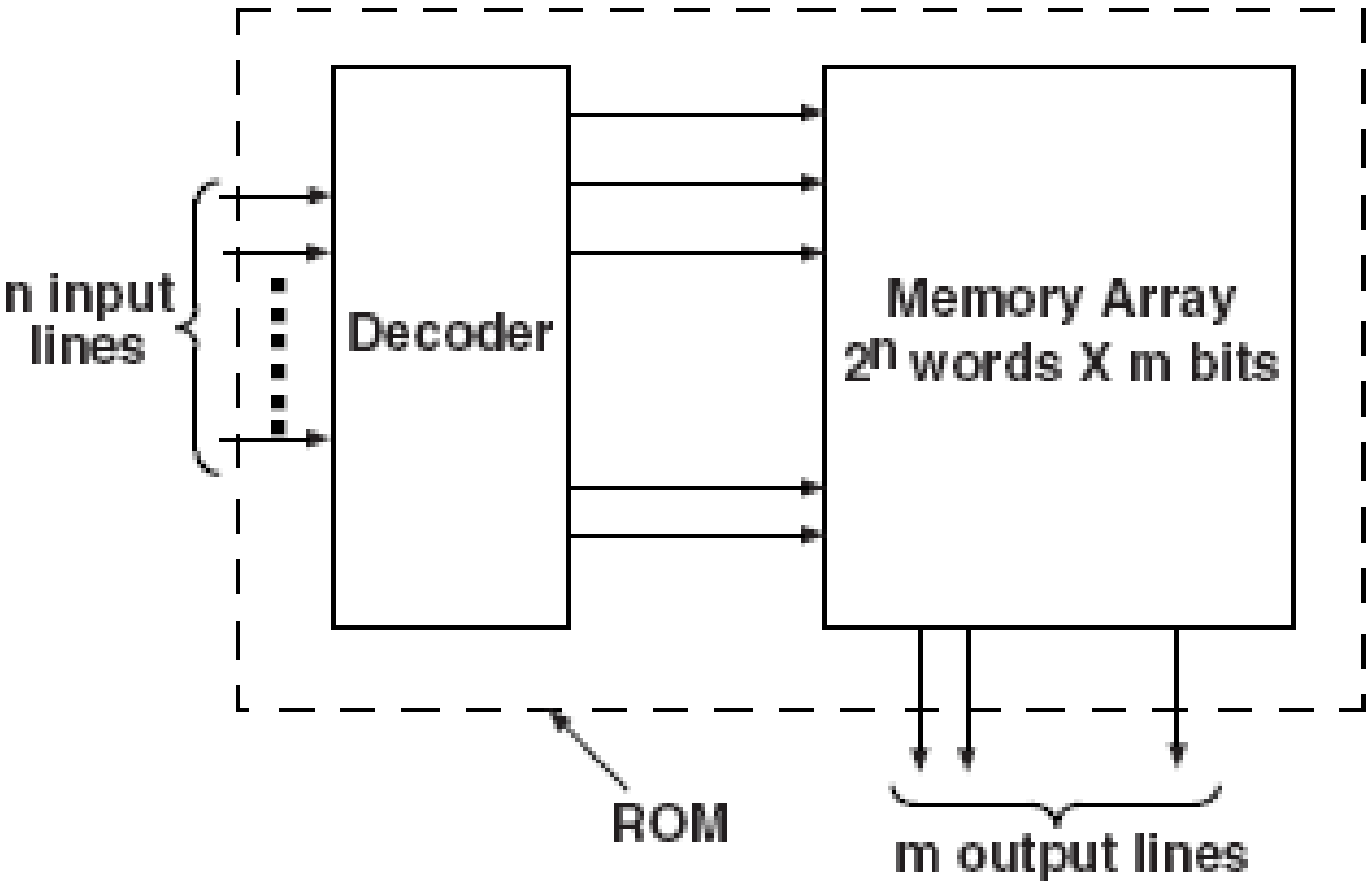
CONTD.

- ROM-Memory does not allow neither for deletion, nor can it be overwritten by the user via program commands.
- Therefore, all programs and data that is in the ROM have to be installed through the production process

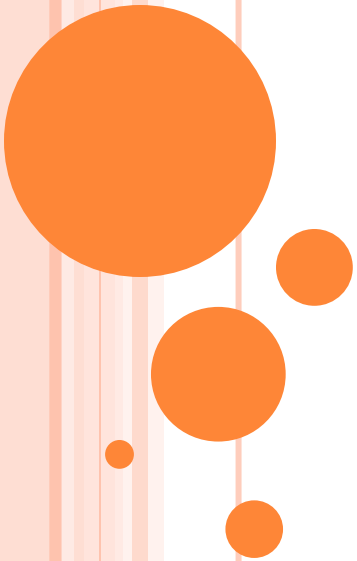
Use of Rom

- Used to store permanent data which are very essential to run the hardware correctly.
- A good example for Rom is BIOS chip(Basic Input Output Set).
- Generally used for updates of firmware.

Basic ROM Structure



- From left to right there are three stages.
- First stage is usually called as address decoder(de-multipluxer) in memory circuits.
- Second stage is memory arrays.
- For each word line there may be a connection to those bitlines that should be activated for corresponding memory word.
- An additional stage is an amplifier or output buffer used to generate a strong output signal.



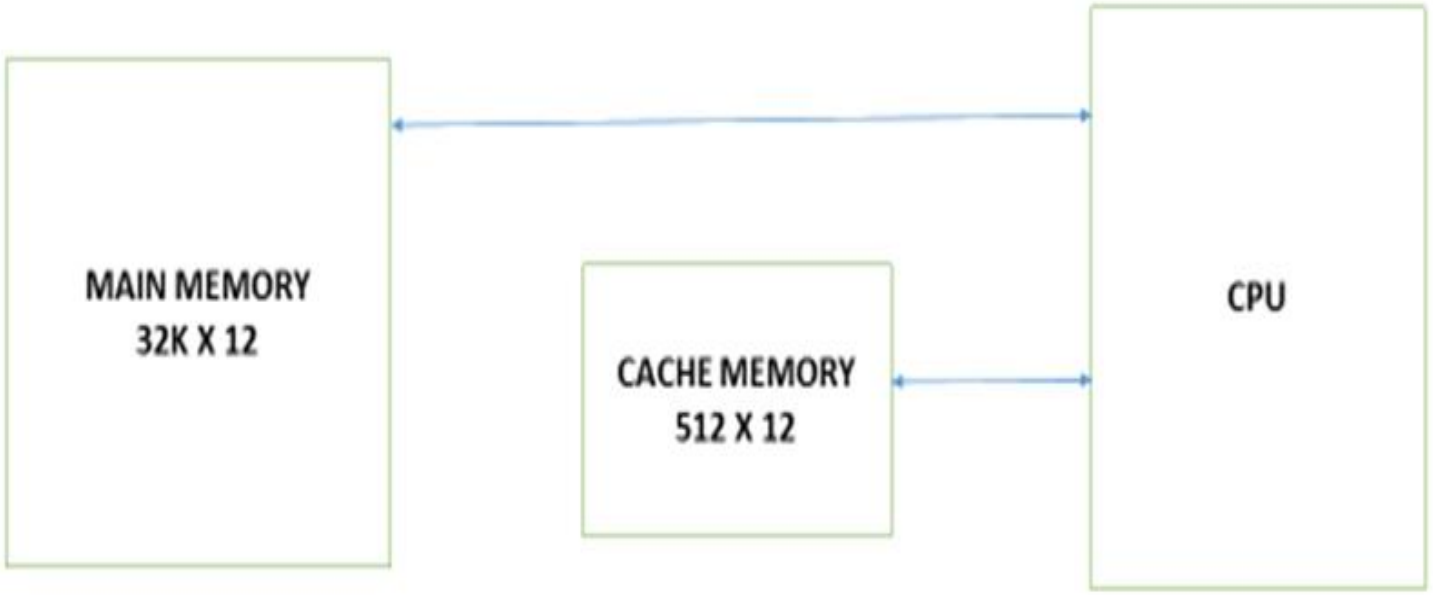
Cache Memory

Introduction:

- When a program instruction is executed, the CPU repeatedly refers to the set of instructions in the memory.
- Every time a subroutine is called, its instructions are fetched from the memory.
- Over a short interval of time, the address generated by a program refers to few localized areas of memory repeatedly.
- That is ,only some portion of memory will be accessed repeatedly at a particular time while the remaining memory in the main memory is accessed less frequently.

- It takes more time for the CPU to access the main memory each and every time .
- If the active portions of the program and data are placed in a fast small memory, the average memory access time can be reduced, which reduces the total execution time of the program.
- Such a fast small is called *"Cache Memory"*.
- Cache memory is placed between CPU and main memory.
- Cache memory access time is less than that of main memory by a factor 5 to 10.
- Cache is the fastest components in the memory hierarchy.

Example of a cache memory:



Basic Operation of Cache memory:

- When the CPU needs to access some memory, first the cache is examined.
- If the word to be accessed is found in the cache memory, then it is read from the fast memory.
- If the word is not found in the cache memory, then it is read from the main memory.
- A block of words just accessed in the main memory is automatically transferred to cache memory.
- Cache memory contains a replica or the copy of the main memory.

- When the CPU refers to a memory and finds that word in the cache, it is called "*hit*".
- When the word is not found in the cache and it is present in the main memory, then it is counted as "*miss*".
- The ratio of no.of hits to the no.of total CPU references to memory(hit+ misses) is called as "*hit ratio*".
- The performance of the cache memory is frequently measured in terms of hit ratio.
- Hit ratios are found to be more than 0.9.
- Hit ratio = (hit)/(hit+miss).

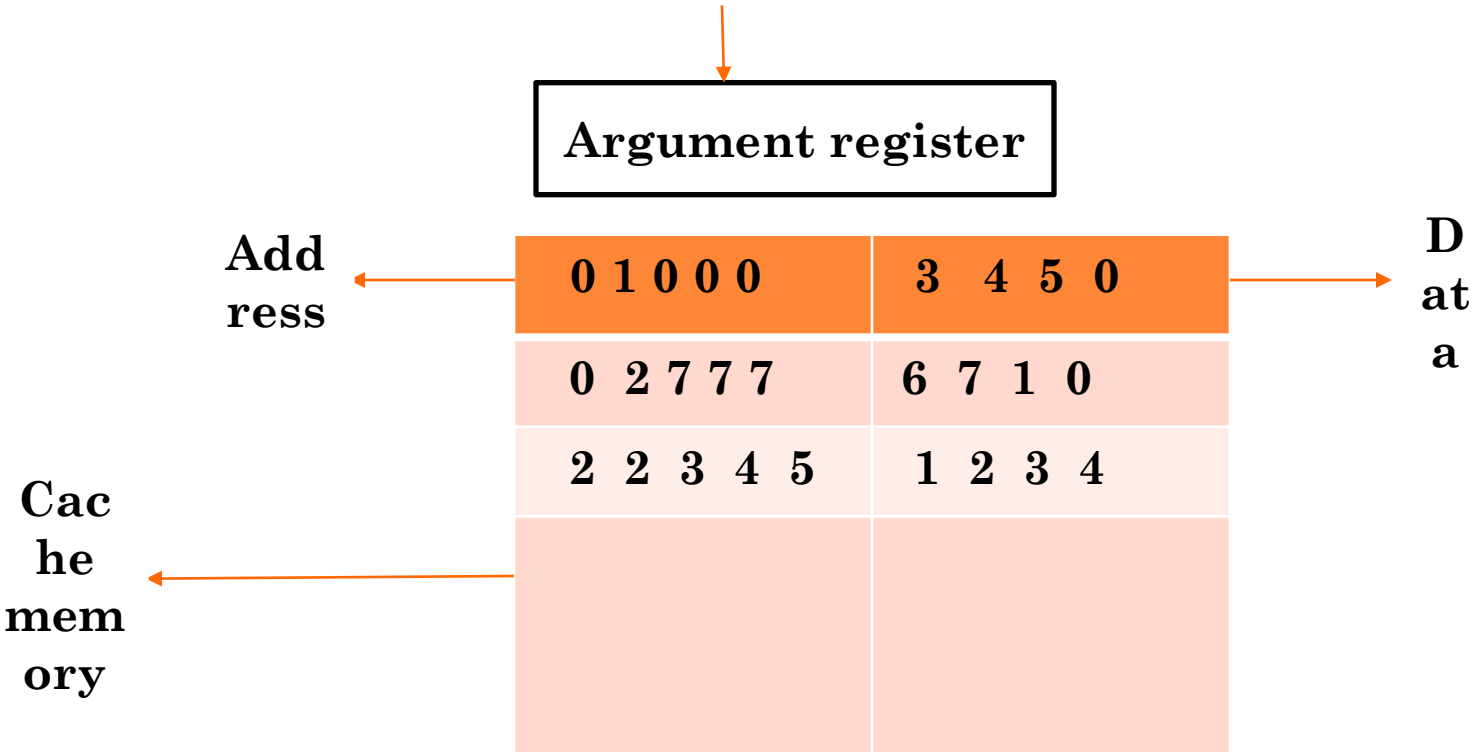
Mapping process:

- The transformation of data from main memory to cache memory is called “mapping process”.
- There are three types of mapping procedures in the organization of cache memory:

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

Associative Mapping:

- The fastest and most flexible cache organization uses an associative memory.
- The associative cache memory stores both the address and the content i.e., data of the memory word.



- The diagram shows three words presently stored in the cache.
- The address value is of 15 bits which is shown in 5 digit octal number.
- Its corresponding 12-bit word is show in 4-digit octal number.
- A CPU address of 15-bits is stored in the argument register and the associative memory is searched for the matching address .
- If the address is found, it is *hit* and the word is read into the CPU,if the address is not found, then it is miss and the word is read from the main memory and the address data pair is then transferred to the cache memory.
- If the cache is full, then the words in the cache are to be displaced to make room for new words.

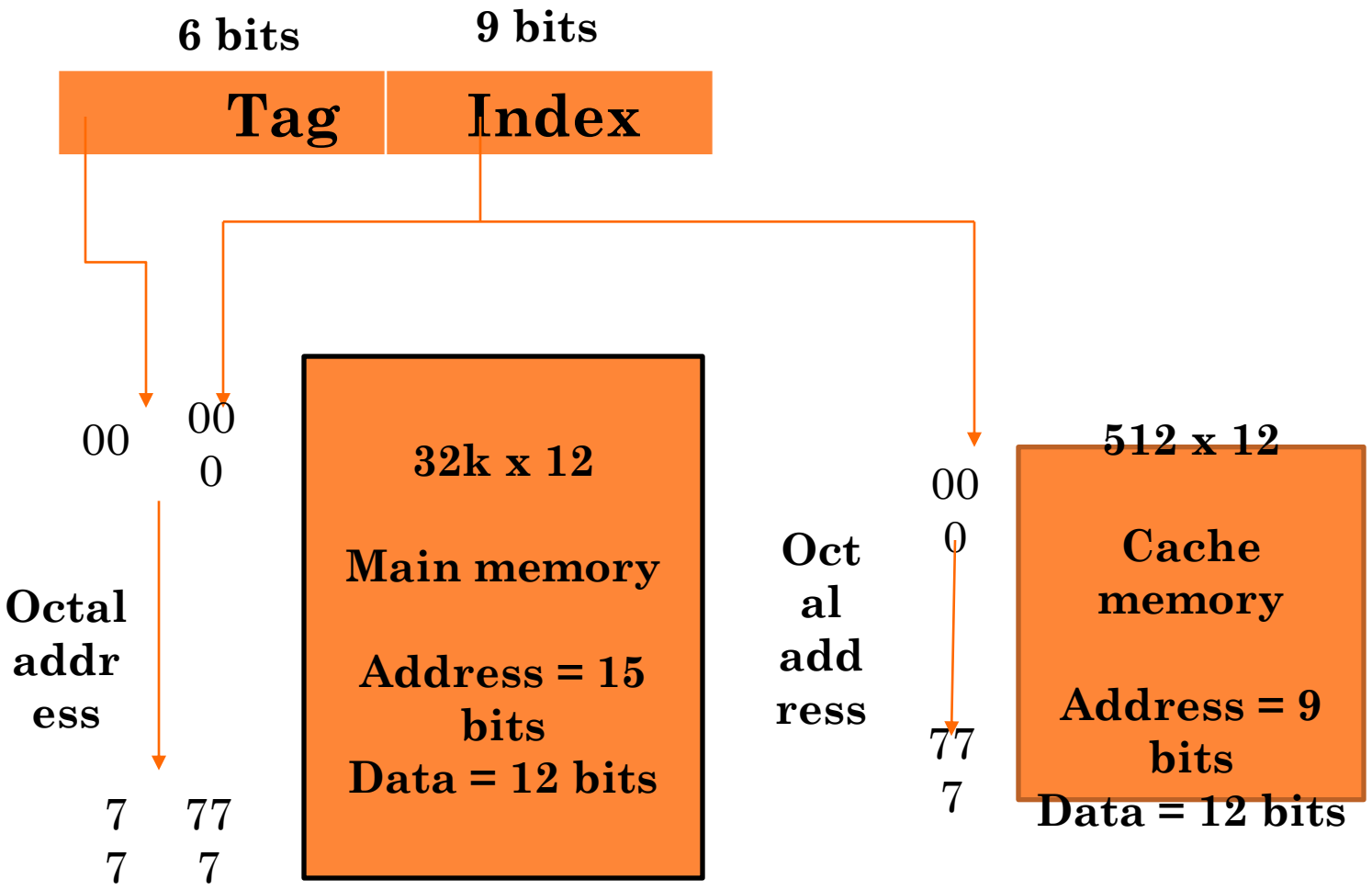
Direct Mapping:

- The drawback of Associative mapping is it is more expensive as it has an added logic associated with each cell.
- To overcome this drawback ,we move to Direct Mapping.
- In direct mapping we use the random access memory.
- To organize direct mapping, the CPU address of 15 bits is divided into two fields:

1. the nine least significant bits constitute the *index* field.

2. the remaining six bits constitute the tag field.

ADDRESSING RELATIONSHIPS BETWEEN MAIN AND CAHE MEMORIES:



Direct mapping cache organization:

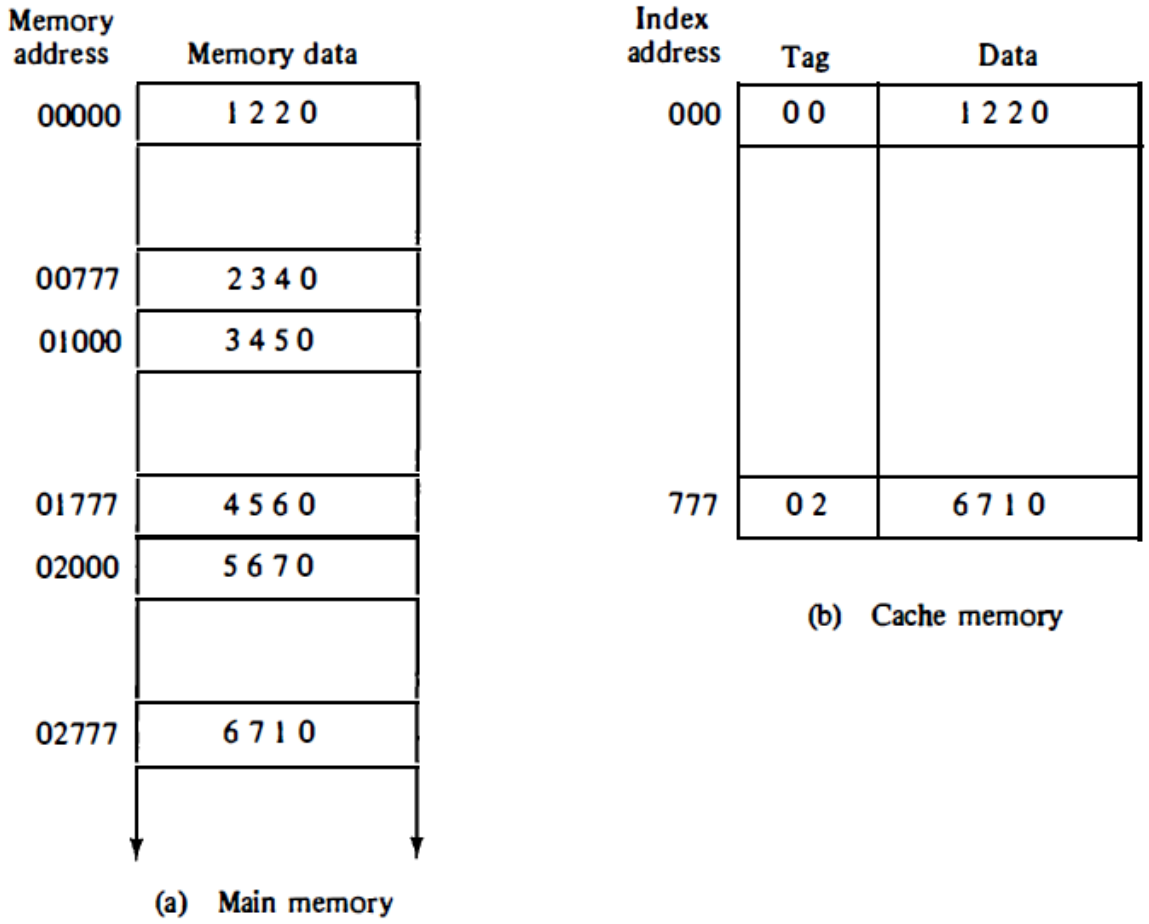


Figure 12-13 Direct mapping cache organization.

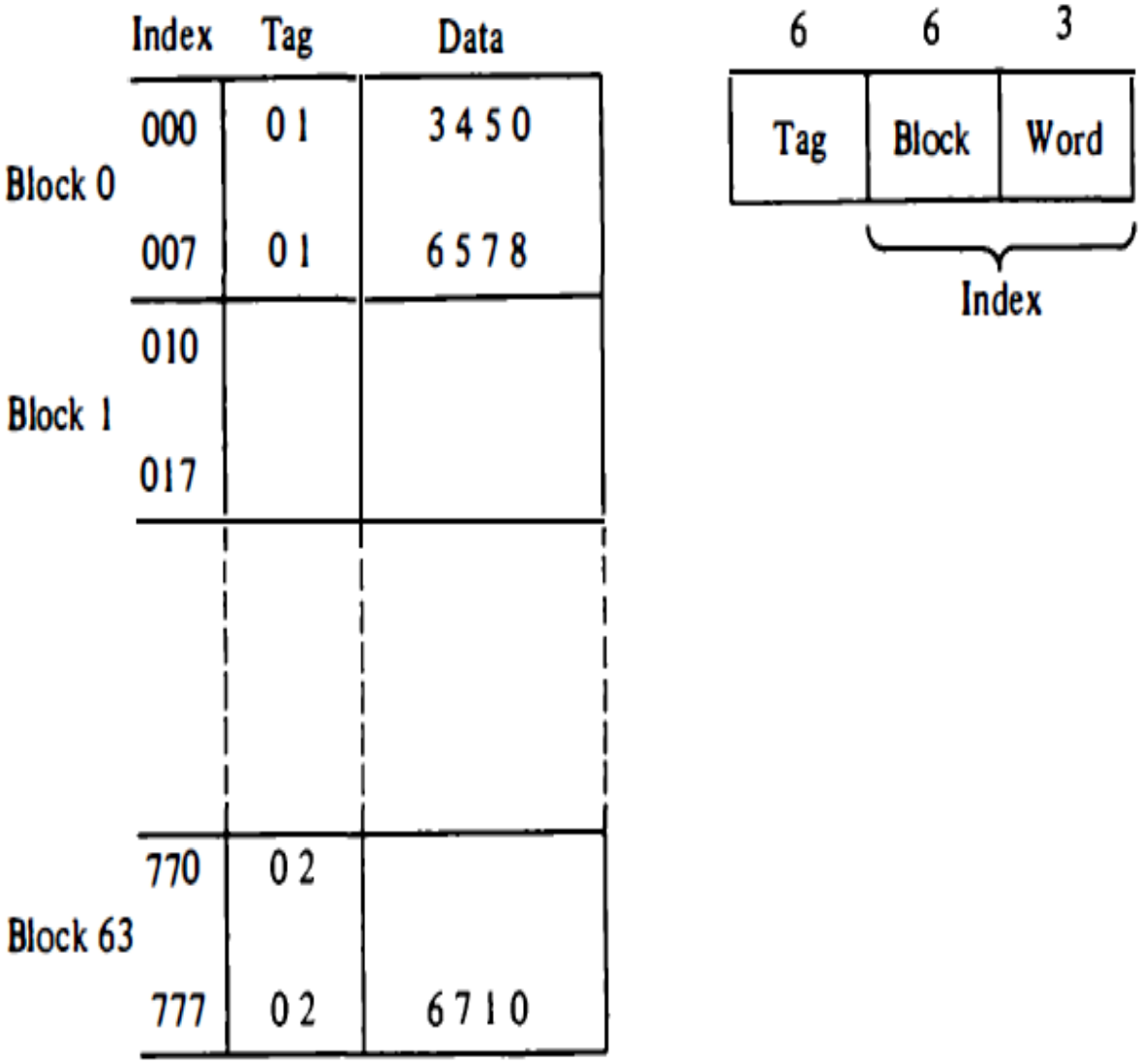


Figure 12-14 Direct mapping cache with block size of 8 words.

Set-Associative Mapping:

- The drawback of direct mapping is two words is two words with same index but with different tag values cannot reside in the cache memory for the same time.
- Set-Associative mapping is an improvement of direct mapping in which 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 word of cache is said to form a set.

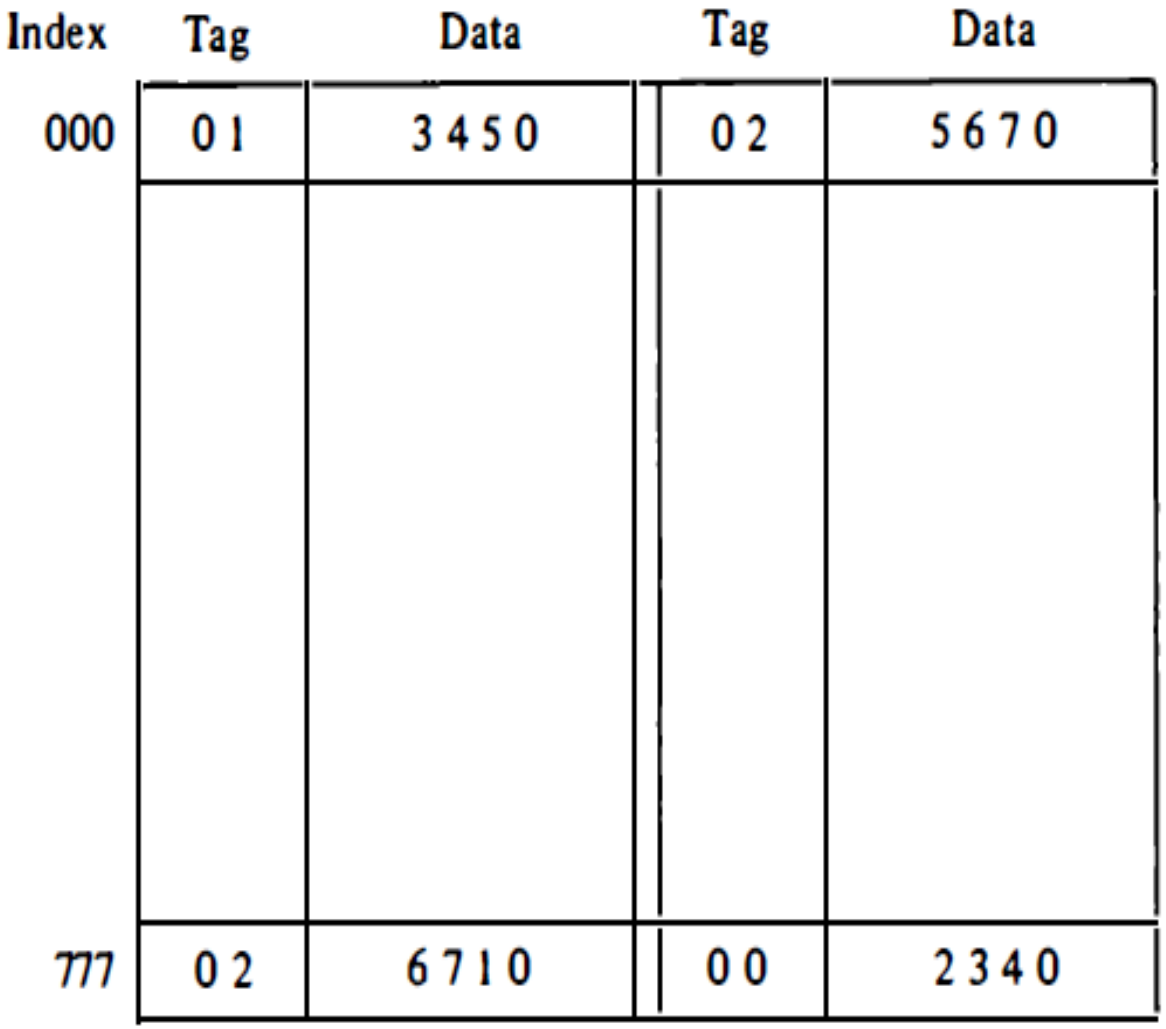


Figure 12-15 Two-way set-associative mapping cache.

WRITING INTO CACHE:

There are two methods:

1. Write-through method.
2. Write-back method.

1.Write-through method:

In this method , main memory is updated with every memory write operation and the cache memory is also updated in parallel if it contains that word in the specific address.

It has an advantage that main memory always contains the same data as that of the cache memory.

2.Write-Back method:

In this method only the cache memory is updated.

The location where the memory is updated is marked by a flag

So that when the word is removed from the cache , it can be added to main memory easily.

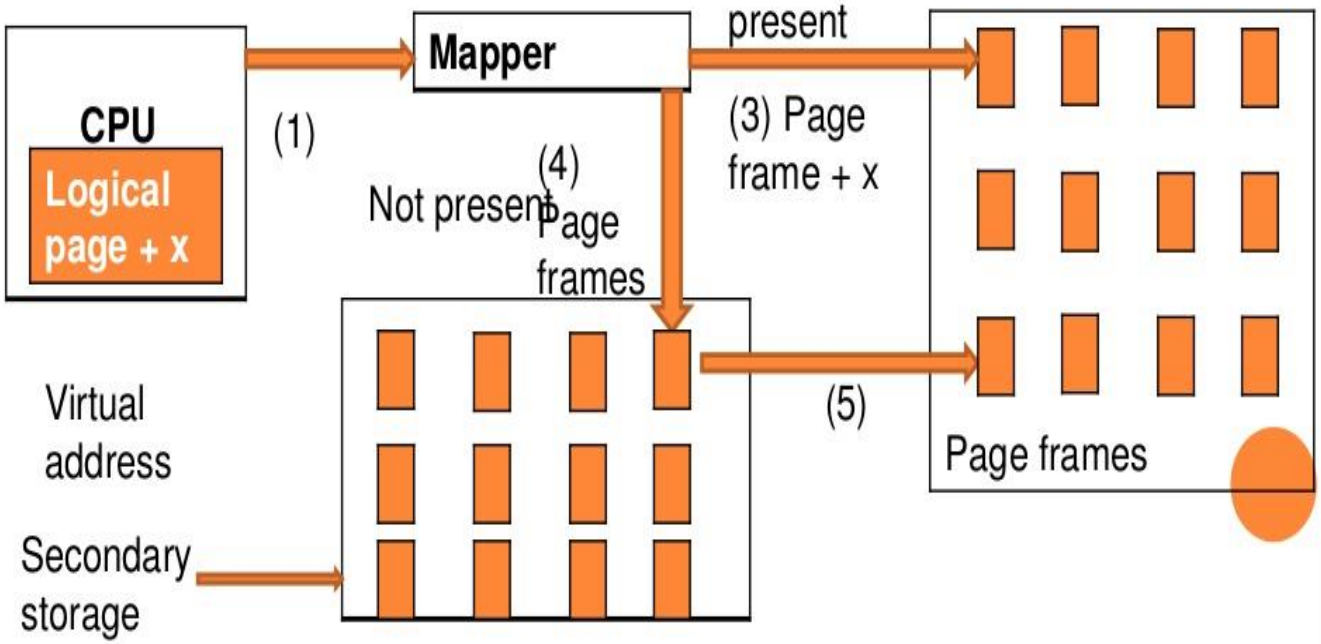
The advantage of this method is we can update the memory several times easily during the program execution and when the word id removed from the cache the accurate copy is re-written into the main memory.

Cache Initialization:

- The cache memory is initialized when the power is supplied to the computer or when the main memory is loaded with a complete set of programs from the auxiliary memory.
- After initialization , ideally the cache has to be empty , but it instead it contains some invalid data.
- Therefore, a *"valid bit"* is included with each word in cache to check whether the word contains valid data or not.
- Firstly , all the valid bits are set to 0 when the cache is initialized.
- When a new word is loaded from main memory to cache , the valid bit is set to 1.
- 1 indicates valid data and 0 indicates invalid data.

VIRTUAL MEMORY

- Virtual memory is a common part of operating system on desktop computers.
- The term virtual memory refers to something which appears to be present but actually it is not.
- The virtual memory technique allows users to use more memory for a program than the real memory of a computer.



Need of Virtual memory

- Virtual memory is a imaginary memory which we are assuming. If we have a material that exceed your memory at that time we need to use the concept of virtual memory.
- virtual memory is temporary memory which is used along with the ram of the system.

Importance of Virtual memory

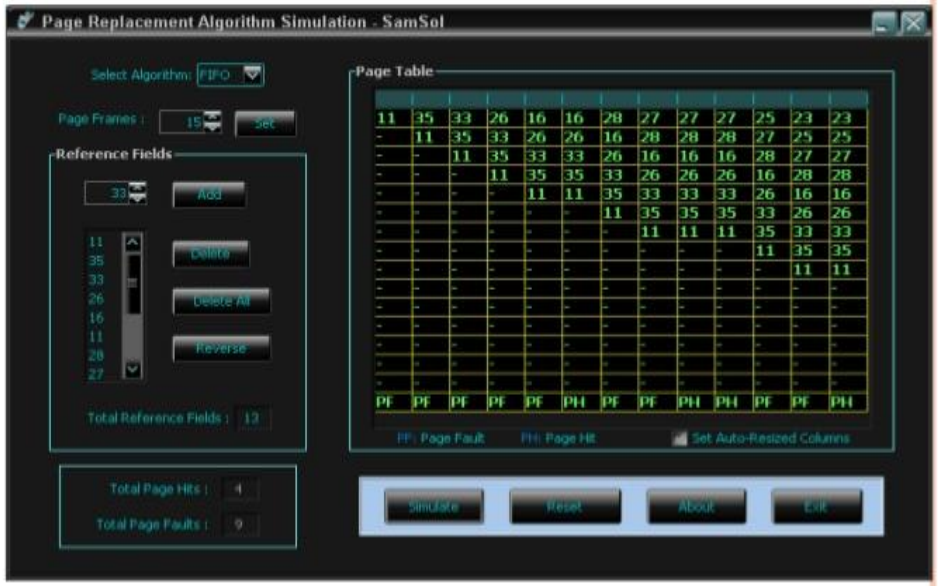
- When your computer runs out of physical memory it writes what it needs to remember to the hard disc in a swap file as virtual memory.
- If a computer running Windows requires more memory/RAM then there is installed in the system to run a program, etc, it uses a small section of the hard drive for this purpose

Address space and memory space

- Virtual memory is the address used by the programmer and the set of such addresses is called address space.
- An address in main memory is called a physical address.
- The set of such locations in main memory is called the memory space.
- Thus the memory space consist of the actual main memory locations directly addressable for processing.

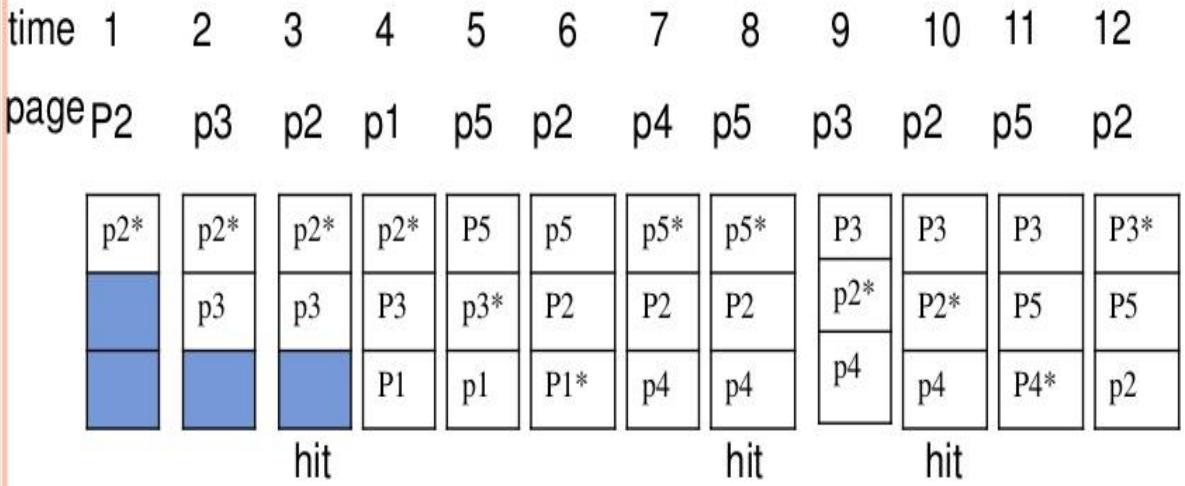
PAGE REPLACEMENT ALGORITHMS

- In a computer operating system that uses paging for virtual memory management, page replacement algorithm decide which memory pages to page out. When a page of memory need to be allocated.
- FIFO (first in first out)
- LRU (Least Recently used)
- OPT (Optimal)



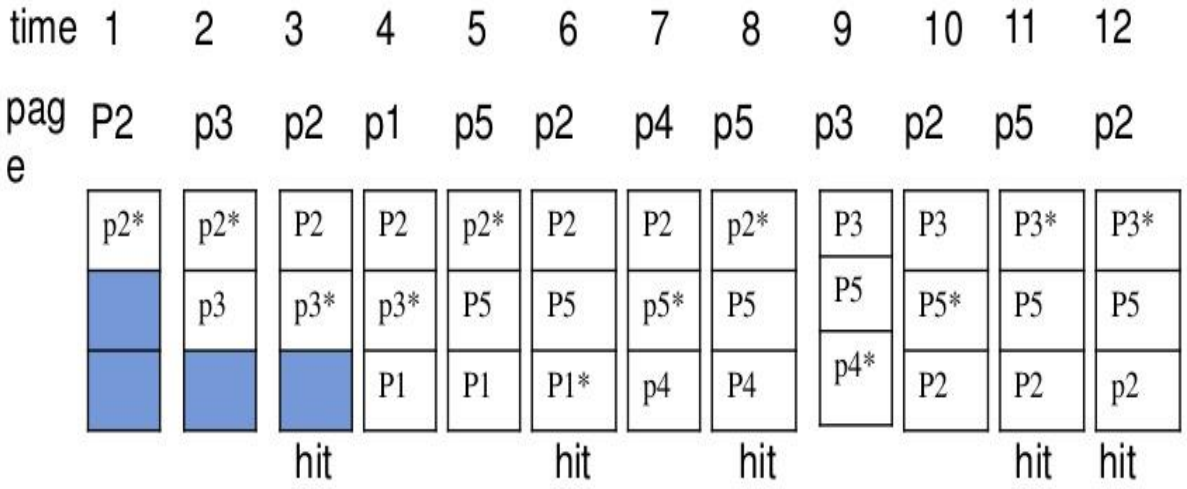
FIFO

First in first out is very easy to implement.
The fifo algorithm select the page for replacement that has been in memory the longest time.



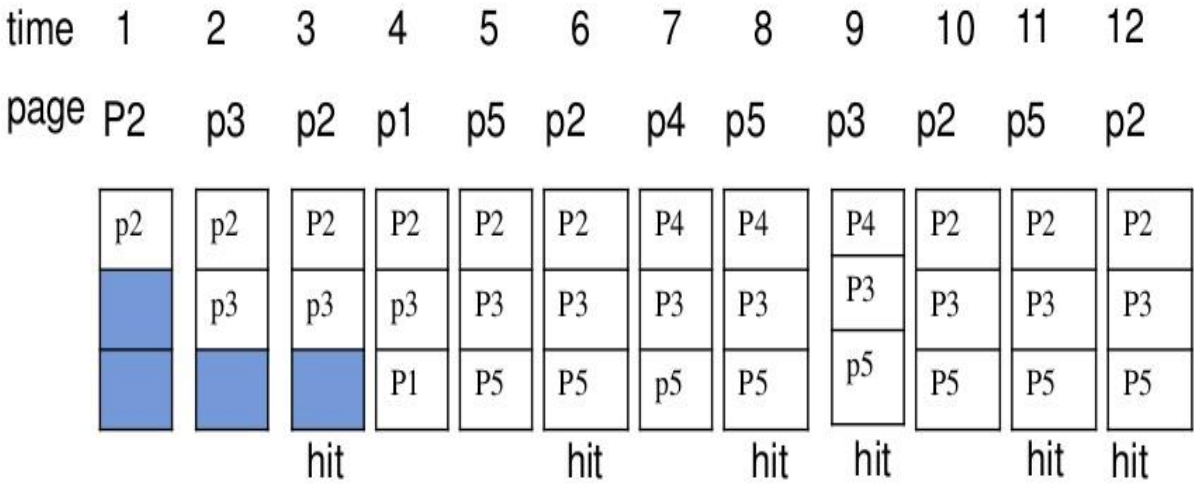
LRU

- The least recently used page replacement algorithm keeps track page uses over a short period of time.
- The LRU algorithm can be implemented by associating a counter with every page that is n main memory.



OPT

- The optimal policy selects that page for replacement for which the time to the next reference is longest.
- This algorithm result is fewest number of page faults.



Advantages of Virtual Memory

- Allows processes whose aggregate memory requirement is greater than the amount of physical memory, as infrequently used pages can reside on the disk.
- Virtual memory allows speed gain when only a particular segment of the program is required for the execution of the program.
- This concept is very helpful in implementing multiprogramming environment.

Disadvantages of Virtual memory

- Applications run slower if the system is using virtual memory.
- It Takes more time to switch between applications.
- Less hard drive space for your use.
- It reduces system stability.

Reference:

1. **Computer Organization and Architecture, William Stallings, 11thEdition, Pearson, 2022.**

Prepared by:

Mr A Srinivasan,

Associate Professor,

Dept of CSE-DS