Paper Source
Subject.....
Date :.....Time:.....

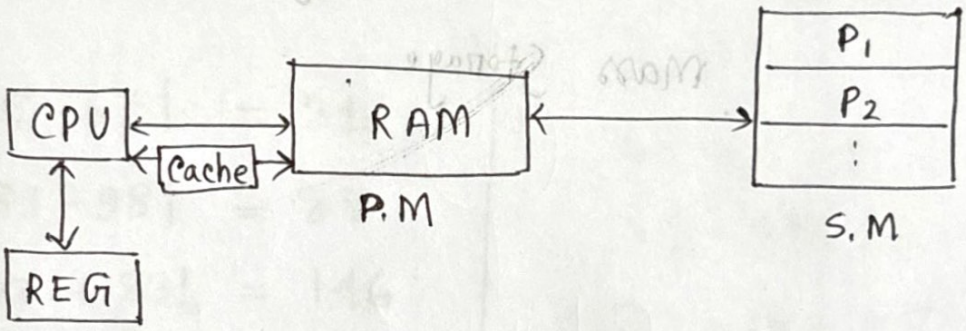
Chapter-9

Main memory

RAM - Primary memory

CPU directly Secondary Memory থেকে কোন Program কে execute করতে পারে না।

| Size
| Accestime
| Per unit cost

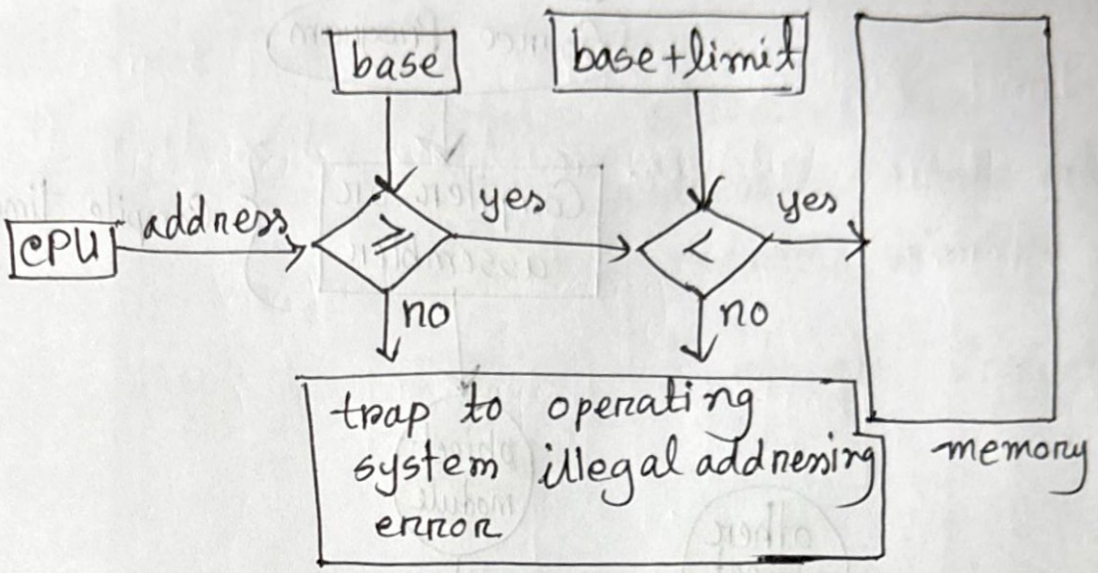


Main memory and registers are only storage CPU can access directly

Protection : It ensures memory protection ~~by~~ so that processes cannot interfere or access unauthorized resource. It ensures that a process can ~~only~~ access only those addresses in its address space.

We can provide protection by using pair of base and limit registers, define logical adress space of a process. It is a hardware adress protection.

CPU must check every memory access generated in user mode to be sure it is between base and limit for that user



Address Binding: Address binding is the process of mapping a program's logical address to physical addresses. It ensures that a program can access memory properly while running.
It can happen in 3 different stages:-

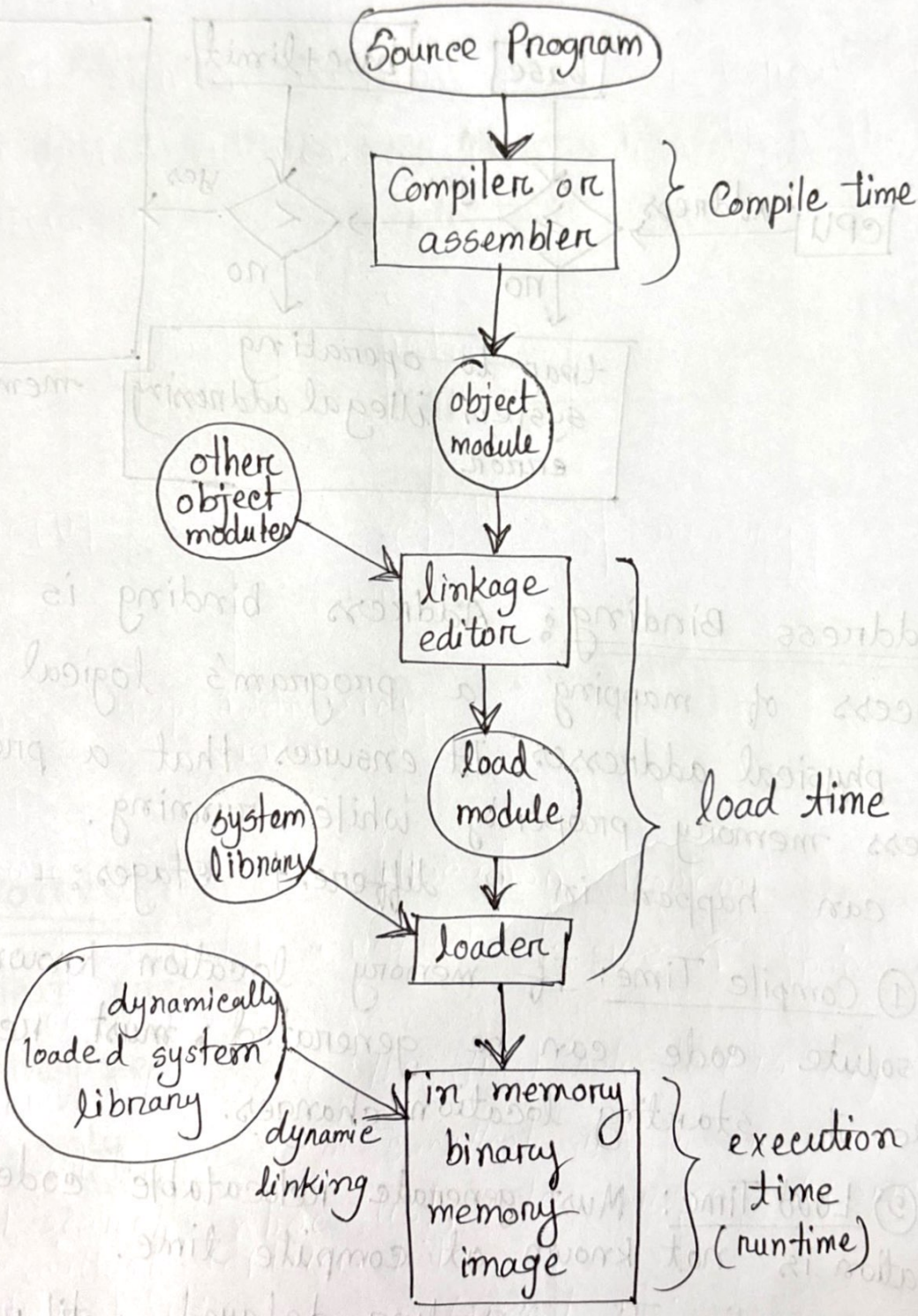
① Compile Time: If memory location known a priori, absolute code can be generated; must recompile code if starting location changes.

② Load Time: Must generate relocatable code if memory location is not known at compile time.

③ Execution Time: Binding delayed until runtime if the process can be moved during its execution from one memory segment to another. Then it needs hardware

support for adress maps. (base and limit registers)

Multistep Processing of a User Program :-



Paper Source
Subject.....
Date :.....Time:.....

Logical Vs Physical Address Space:-

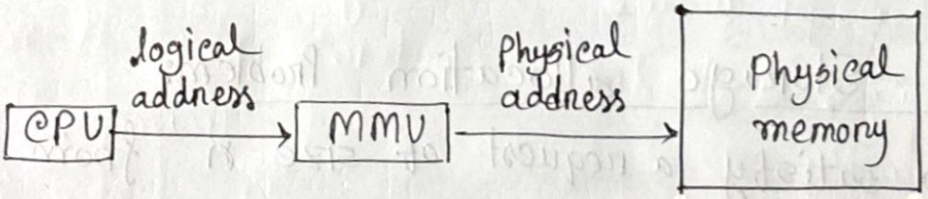
• Logical Address:— generated by CPU, also referred to as virtual address.

• Physical Address:— address seen by the memory unit.

Logical address and physical addresses are same in compile time and load time address-binding-schemes; Both differ in execution time address binding scheme;

- logical address space - is a set of all logical adresses generated by a program.
- physical address space - is a set of all physical addresses generated by a program.

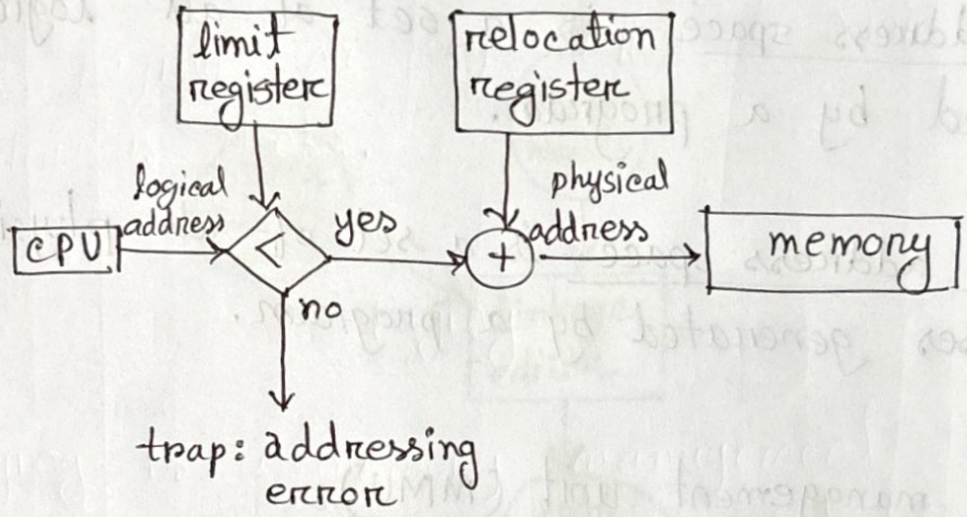
Memory management unit (MMU):-



It's like a base register that helps adjust addresses. The program only works with logical addresses. When the program accesses memory, the relocation register converts the logical address into a physical address by adding relocation value. It allows program to run in any memory location, without needing to know where there are in physical memory

Paper Source
Subject.....
Date :.....Time:.....

- MMU maps logical address dynamically.
- Base register contains value of smallest physical address.
- Limit Register contains range of logical addresses — each logical address must be less than the limit register.



Dynamic Storage allocation Problem:

How to satisfy a request of size n from a list of free holes?

better than worst fit
- First fit: Allocate the first hole that is big enough
- Best fit: " " smallest " " " " " "

must search entire list, unless ordered by size. Produces the smallest leftover hole.

Worst fit: " " largest "

" " " " Produces the largest leftover hole.

Paper Source
Subject.....
Date :.....Time:.....

Fragmentation: Fragmentation happens when memory (RAM/storage) is not used efficiently, causing gaps or "holes" in the memory. These gaps prevent the system from fully using available memory, even though there's still free space.

Types of fragmentation:

① External Fragmentation — total memory space exists to satisfy a request, but its not contiguous. It happens when free memory is split into small, scattered chunks. Even if total memory is enough for a program, it cannot be used because the memory chunks are not contiguous (next to each other)

② Internal Fragmentation: It occurs inside allocated memory blocks. It's allocated memory may be larger than requested memory. The unused space inside the block cannot be used by other programs.

*** Internal fragmentation wastes space inside memory blocks.

*** External " makes it hard to use free space effectively.

How to fix fragmentation?

⇒ For internal fragmentation, use better memory allocation techniques, like allocating memory closer to the program's exact need.

For external fragmentation, use compaction, where the system rearranges memory to group free space together. Use memory management techniques like paging or segmentation to avoid needing contiguous blocks.

Paging

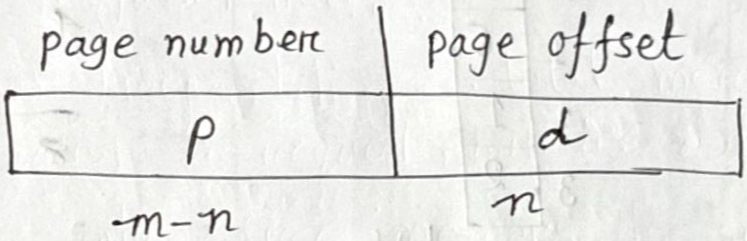
Paging is a memory management technique that divides a program's memory into fixed-size blocks called pages and the computer's physical memory into equally sized blocks called frames. This allows program to use memory efficiently, even if the blocks are not contigous (next to each other)

- Avoids external fragmentation.
- Avoids problem of varying sized memory chunks.

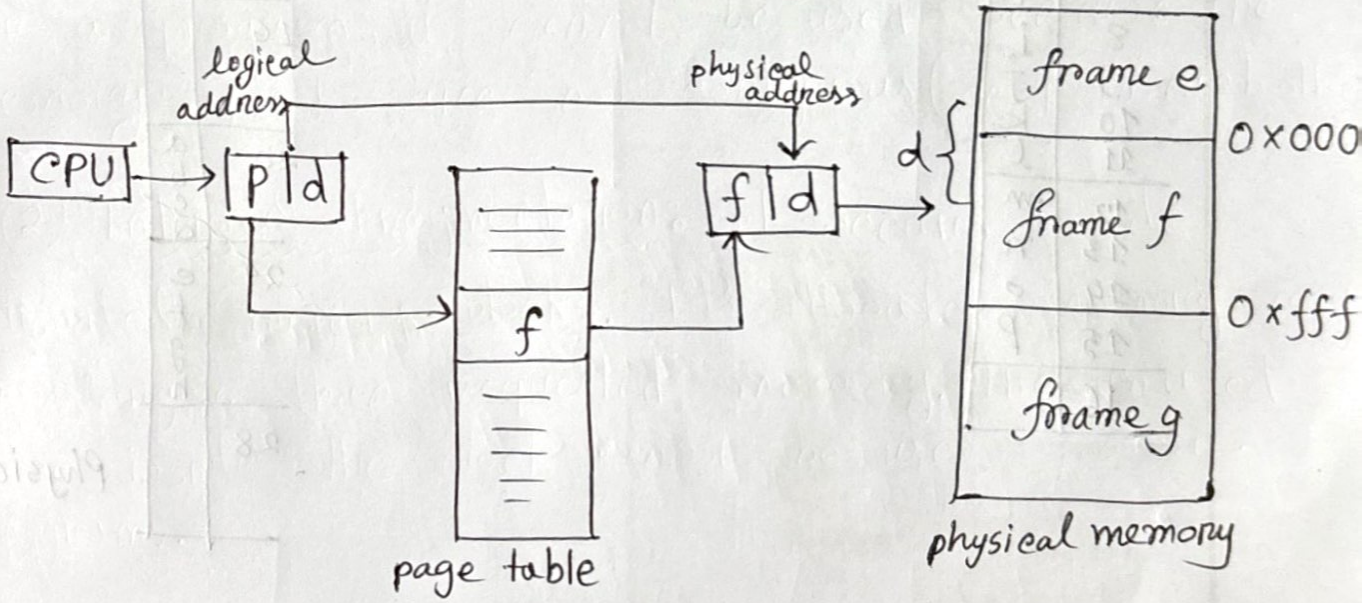
Divide logical memory into block of same size called pages. It keeps track of all free frames.

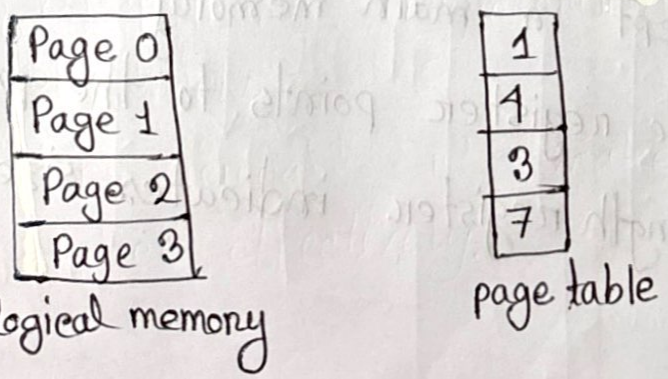
Paper Source
Subject.....
Date :.....Time:.....

- Page number (p) – used as an index into a page table which contains base address of each page in physical memory.
- Page offset (d) – combined with base address to define the physical memory address that is sent to the memory unit.



For given, logical address space 2^m and page size 2^n





Paper Source
Subject.....
Date :.....Time:.....

□ **Logical address:** $n = 2$
$m = 4$

using a page size of 4 bytes and a physical memory of 32 bytes (8 pages)

0	a
1	b
2	c
3	d
4	e
5	f
6	g
7	h
8	i
9	j
10	k
11	l
12	m
13	n
14	o
15	p

0	5
1	6
2	1
3	2

page table

0	
4	i j k l
8	m n o p
12	
16	
20	a b c d
24	e f g h
28	

Physical memory

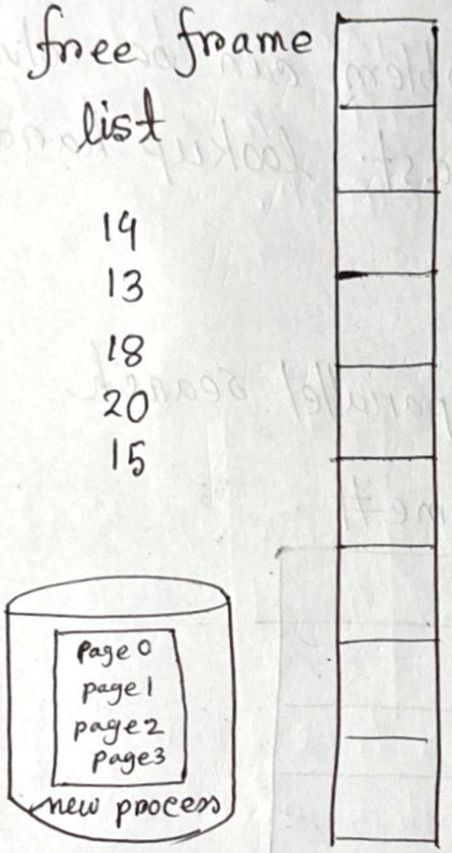
Page table is kept in main memory

(PTBR) Page Table base register points to the page table.

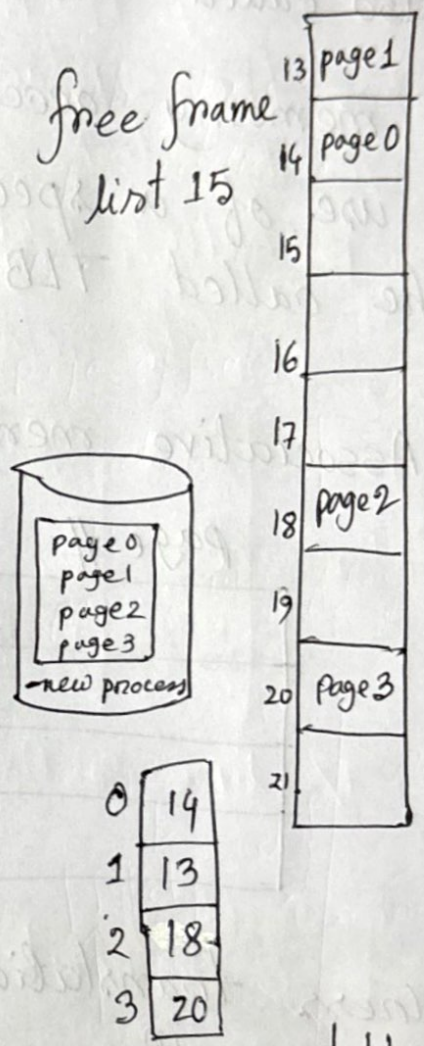
(PTLR) Page Table Length register indicates size of the page table.

Paper Source
Subject.....
Date :.....Time:.....

Free frames



(a) before allocation



new page table

(b) After Allocation

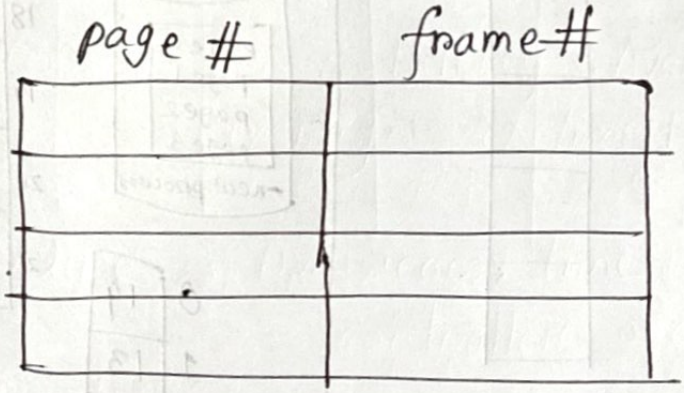
Paper Source
Subject.....
Date :.....Time:.....

TLB → Translation look aside buffer

also called associative memory.

- two memory access problem can be solved by the use of a special fast lookup hardware cache called TLB.

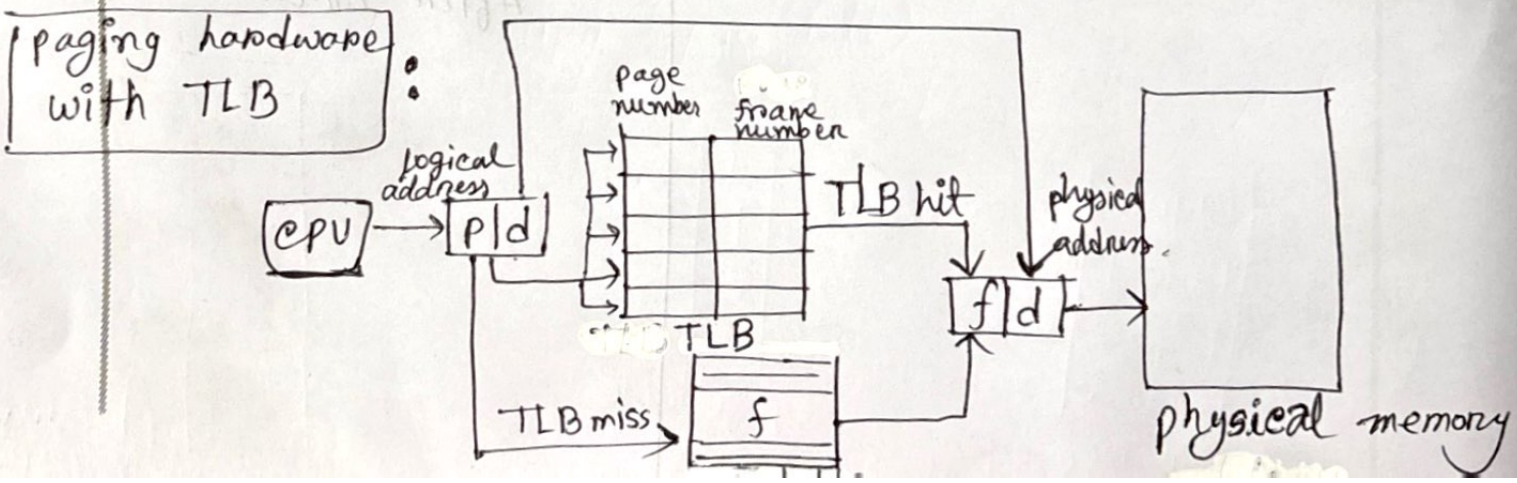
Associative memory - parallel search



Address translation (p, d)

- if p is in associative register, get frame # out
- otherwise get frame # from page table in memory

Paging hardware with TLB:



Paper Source
Subject.....
Date :.....Time:.....

Chapter - 11

Mass Storage Systems

A mass storage structure refers to the organization and management of large volumes of data on storage devices such as hard drives, solid state drives (SSD), magnetic tapes, or cloud storage systems.

~~It focuses on~~

Modern computers often use hard disk drives (HDDs) and non volatile memory (NVM) devices for storing large amount of data.

HDD:

HDD has spinning disks (called platters) coated with a magnetic material to store data. A small part called 'read-write-head' moves over the spinning disk to read or write data.

- Drives rotate 60 - 250 times per second.
- Transfer rate is a rate which data moves between HDD and the computer.
- Positioning time (Random-Access-Time) :- is time to move the read-write head to the right track on the disk (seek time) and the spin desired sector

Paper Source
Subject.....
Date :.....Time:.....

rotate under the disk head (notational latency).

- **Head Crush:** If the read-write head accidentally touches the surface of spinning disk, it can damage the disk and result in data loss. This is called ~~dat~~ head crush.

- Disks can be removable.
- Range from 30GB to 3TB per drive.

- **Performance:** Access latency = Average acces time
  = average seek time + average latency

- ∴ Avg I/O time = Avg access time + transfer rate + controller overhead.

- **First Commercial Disk Drive**

  1956
  IBM RAMDAC computers
  IBM Model 350 disk storage system

Paper Source
Subject.....
Date :.....Time:.....

Non - volatile Memory devices :- NVM devices are storage technologies that retain data even when the power is turned off. They are essential for long term storage and are widely used in modern computers, smartphones and other devices.

Examples ___ SSD, USB Flash drives, Memory cards (SD cards), flash drives, thumb drives, DRAM disk, sunface mountain on motherboads.

- Can be relible than HDDs
- Less capacity but much faster
- Shorter life span, no moving parts, So, no seek time or rotational latency.

- Read & write in Pages.
- Can't overwrite directly,
- life span measured in DWPD (Drive writes per day).

~~NAND~~

Paper Source
Subject.....
Date :.....Time:.....

NAND Flash Controller Algorithms

It manage how data is written, read and erased in NAND flash memory.

- With no overwrite, pages end up with mix of valid and invalid data

Valid page	Valid page	invalid page	invalid page
invalid page	Valid page	invalid page	Valid page

NAND block with valid and invalid page

- Wear leveling — needed to write equally to all cells.
- Garbage collection — to free invalid page space
- Over provisioning — to provide working space for GC
- flash translation layer (FTL) — ~~needed to write equally to all cells~~ to track which logical blocks are valid.

Paper Source
Subject.....
Date :.....Time:.....

Volatile Memory

Volatile memory is a type of computer memory that requires power to ~~drain~~ retain data. When the power is turned off, all the data stored in volatile memory is lost.

- It is power dependent, temporary storage

Example — RAM (DRAM, SRAM), Cache memory, VRAM (Video RAM).

Magnetic Tape (Non-Volatile)

mainly used for backup, as a medium ~~of~~ for transferring information from one system to another. It is slow compare to HDD's and ~~RAM~~ SSDs.

Disk Structure

Disk Structure is a combination of physical components ~~(tracks, sector~~ and logical layouts, that work together to store and retrieve data efficiently.