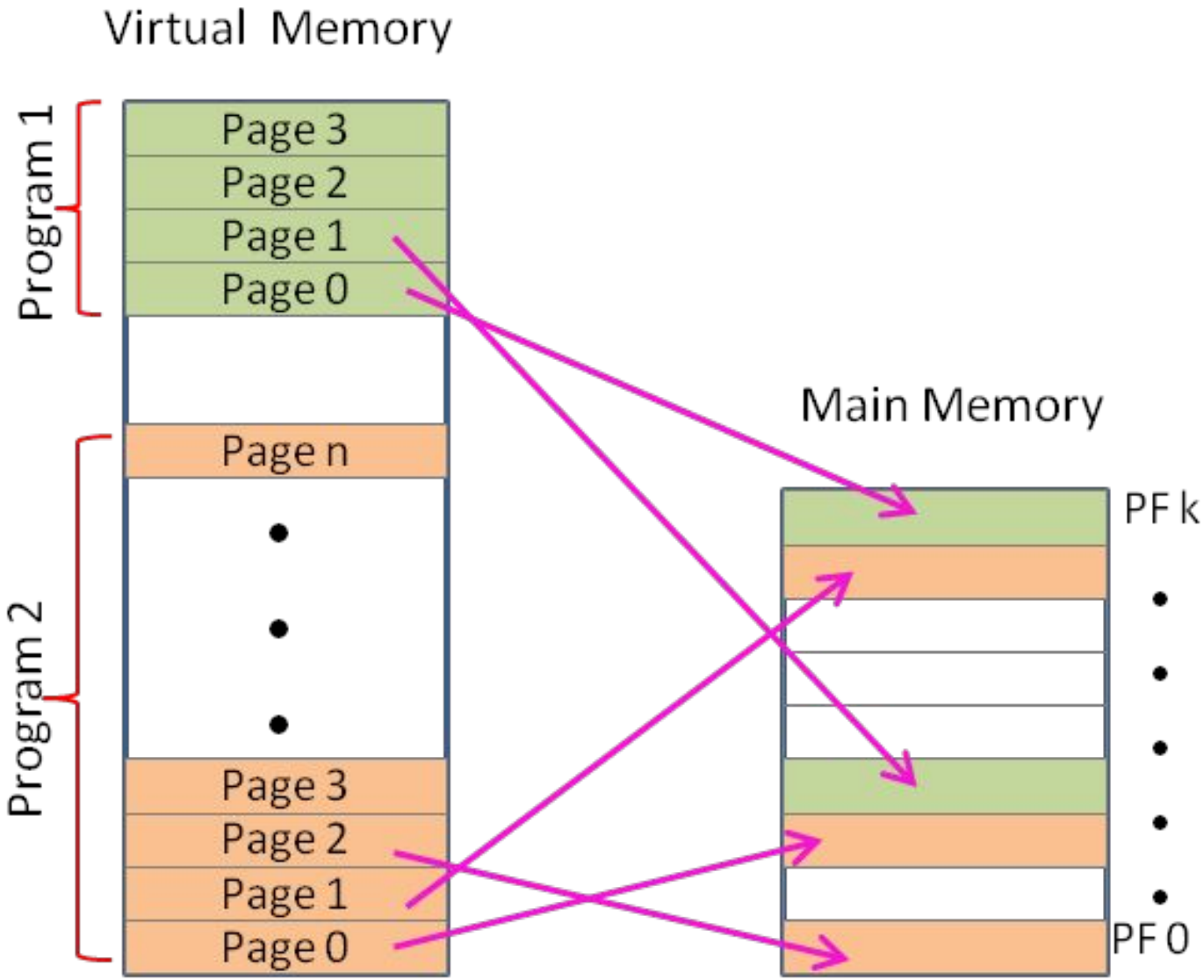
Operating Systems
Course Code: 7120300200 4
Introduction to Virtual Memory

by -
Asst. Prof. Minal Rajwar

What is Virtual Memory?

- It is a technique that makes your computer think it has more memory (RAM) than it actually does.
- It is done by using part of your hard drive or SSD as extra “memory” when the real RAM runs out.
- This allows you to run bigger programs or more programs at the same time, even if computer doesn’t have enough physical RAM.

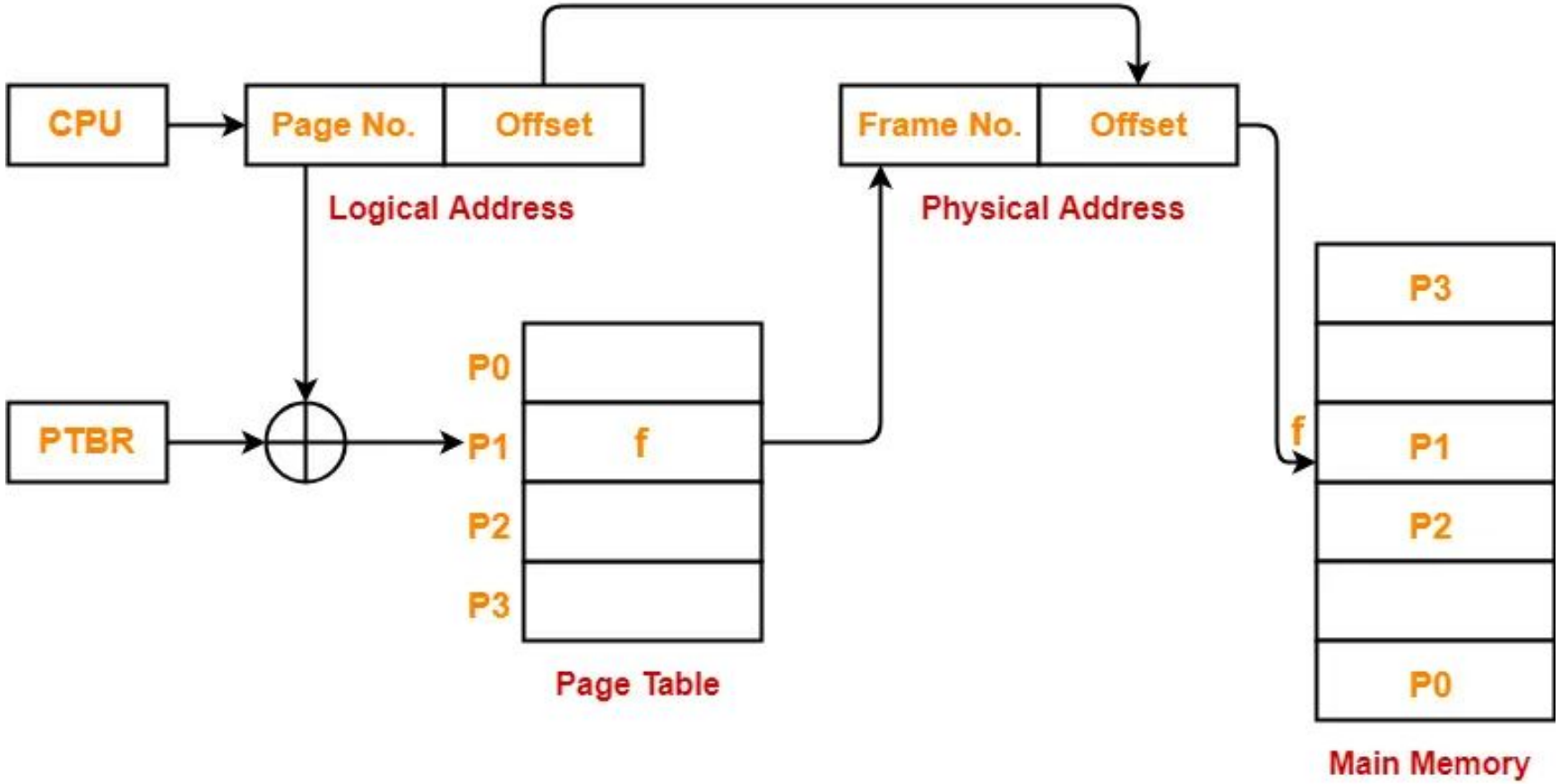

Virtual Memory
Program 1
Page 3
Page 2
Page 1
Page 0
Page n
Program 2
Page 3
Page 2
Page 1
Page 0
Main Memory
PF k
PF 0

History of Virtual Memory

- earlier memory was expensive and limited.
- programs had to fit entirely into RAM, which was a big problem as software grew larger.
- Programmers used method called - *overlaying,* moving parts of a program **in** and **out** of memory manually - but this was slow and complicated.
- virtual memory was invented in 1950s & 1960s to solve this.
- First working system was developed for the Atlas Computer in 1962, using a method called paging.

How Does Virtual Memory Work?

- Virtual memory uses both hardware and software to manage memory. Here's how it works:
  - **Paging** - Memory is split into small fixed-size block called pages. If RAM is full, unused pages are moved to the hard drive (into a swap file). When needed again they are brought back into RAM.



Translating Logical Address into Physical Address

- **Segmentation** - Memory is divided into variable-sized segments (like parts of a program). Only the needed segments stay in RAM, while other are stored in disk.

- **Combination** - some systems use both paging and segmentation for better efficiency.

If a program tries to access data that isn't in RAM (page fault), the system fetched it from the disk, updates the memory map, and continues running.

Thrashing : When Virtual Memory Goes Wrong

- If the system spends too much time moving data between RAM and disk (instead of running programs), performance drops drastically - this is called *thrashing*. It happens when:
  - Too many programs are running at once (high multiprogramming).
  - There's not enough RAM for active processes.

How to Fix Thrashing?

- ★ Reduce the number of running programs.
- ★ Increase physical RAM.
- ★ Adjust virtual memory settings.

Pros of Virtual Memory

- ★ Lets you run large programs even with limited RAM.
- ★ Allows multiple programs to run at once.
- ★ Makes memory management easier for developers.
- ★ Prevents one program from crashing others by isolating memory.

Cons of Virtual Memory

- ★ Slower than real RAM (since disk access is slower.)
- ★ Can cause thrashing if overused.
- ★ Uses up disk space for swap files.

How to Manage Virtual Memory?

- ★ **Adjust Swap File Size** - Let the OS manage it automatically, or set a custom size.
- ★ **Use a Fast Drive (SSD)** - Improves speed when swapping data.
- ★ **Monitor Usage** - Check if your system is relying too much in virtual memory (indicating low RAM).
- ★ **Disable if You have Enough RAM** - If you have 16GB+ RAM, you might not need much virtual memory.

What is Swapping?

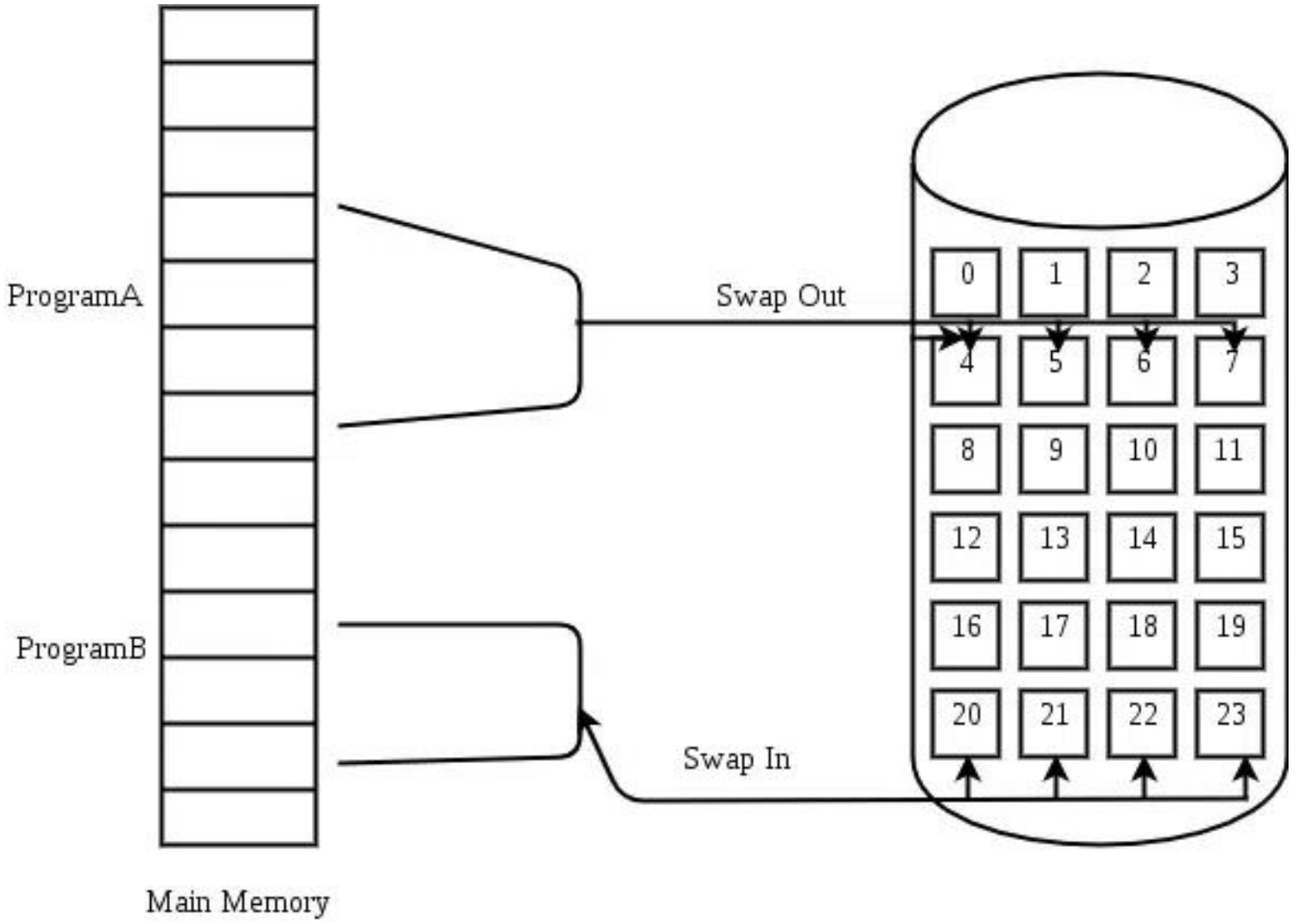
★ Is a process when the OS moves an entire program (or parts of it) out of RAM (physical memory) and onto the hard drive/SSD (secondary storage) to free up space for other programs. Later when needed the OS brings it back into RAM.

★ **How Swapping Works:**

- **Swap Out** - The OS removes a program (or parts of it) from RAM and saves it to disk.
- **Swap In** - When the program is needed again, the OS loads it back into RAM.

What happens during Swapping?

- ★ The program is paused (suspended) while swapped out.
- ★ The OS keeps track of where the swapped data is stored.
- ★ When swapped back in, the program resumes where it left off.


ProgramA
ProgramB
Swap Out
Swap In
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Main Memory

Page Fault

- ★ happens when a program tries to access data that is not currently in RAM but stored on hard drive/SSD (virtual memory).
- ★ The OS must then fetch the missing data from disk and load it into RAM before the program can continue.

Why Page Faults Happen?

- ★ **Normal Behavior :** The OS intentionally keeps only frequently used data in RAM to save space.
- ★ **High Memory Demand:** If too many programs run at once, RAM fills up, causing more page faults.
- ★ **Slow Performance:** Too many page faults lead to thrashing.

Page Replacement

★ When your computer runs out of free RAM, it needs to swap out some old data to make space for new data. This is called page replacement.

★ **How page Replacement works:**

- **a program needs data** > CPU check RAM.
- **RAM is full** > the OS must remove a page from RAM to free space.
- **New page is loaded** > The requested data is brought in from disk.
- **Program continues** > The CPU can now access the data.

Key Terms

- ★ **Page** = A fixed-size block of memory.
- ★ **Frame** = A slot in RAM where a page is stored.
- ★ **Page fault** = When a needed page is not in RAM. (must be loaded from disk).
- ★ **Page Replacement Algorithm** = Decided which page to remove when RAM is full.

Discussion & Revision

- ★ What is the fixed size block of memory called?
- ★ What happens when RAM is full and data must be loaded from disk?
- ★ What is the slowdown caused by excessive swapping called?