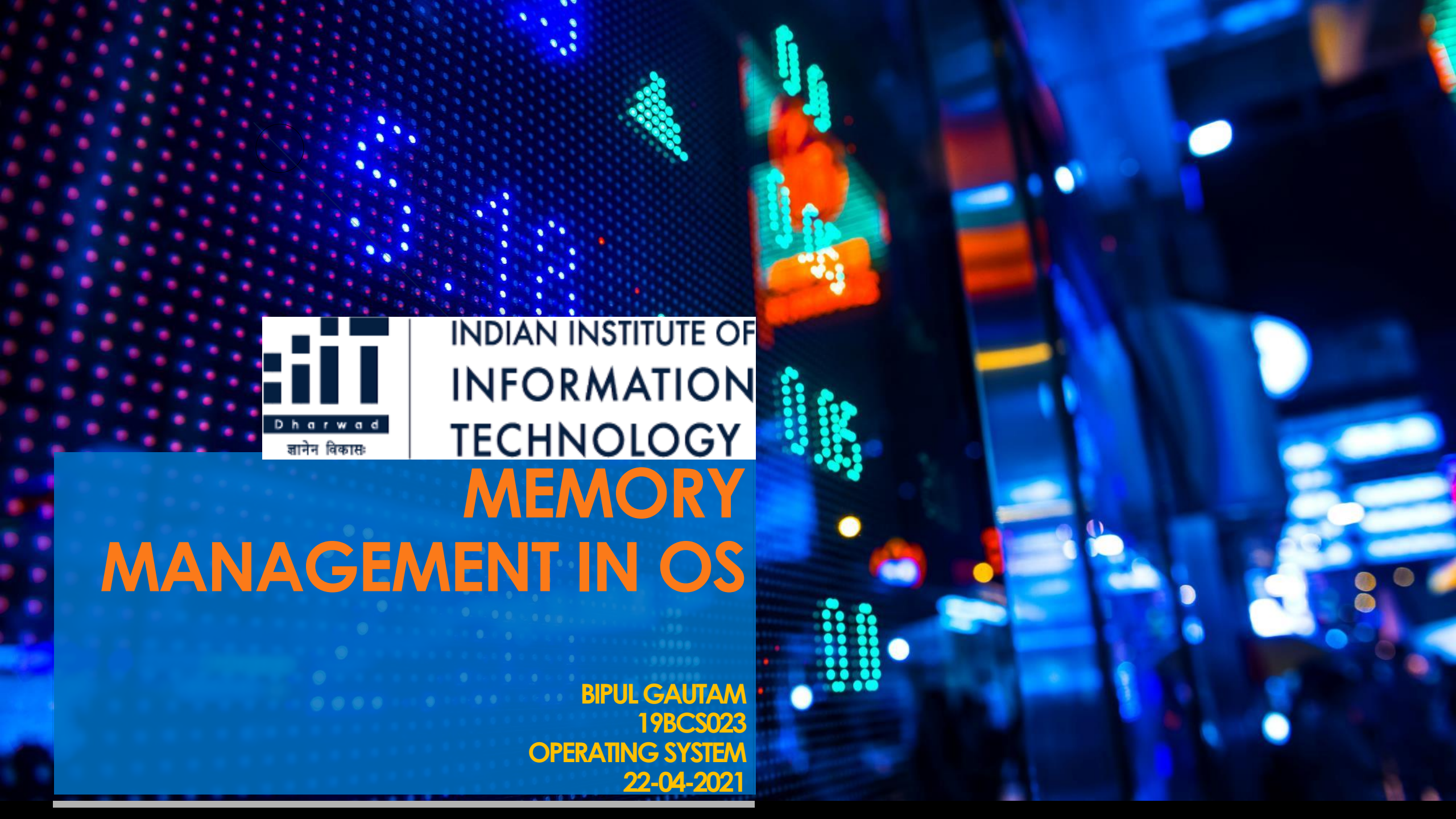


INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

Dharwad

ज्ञानेन विकासः

MEMORY MANAGEMENT IN OS

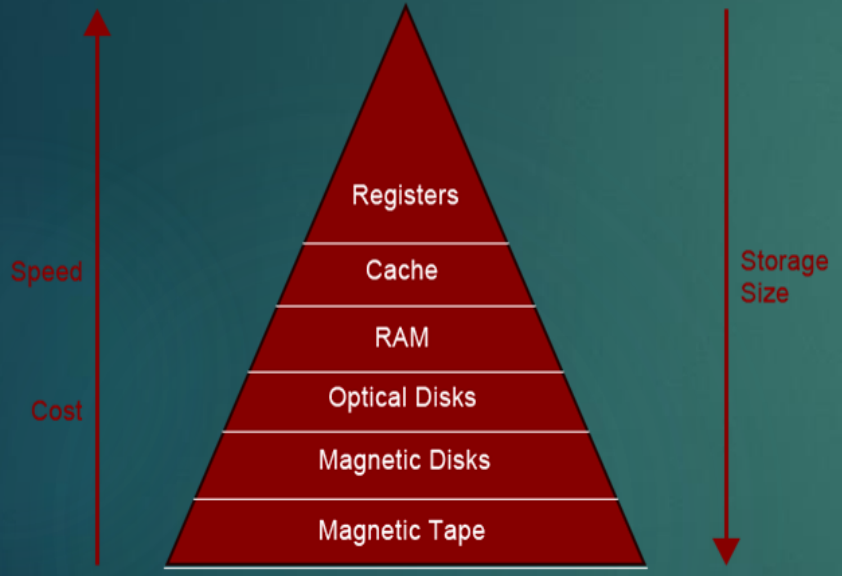
BIPUL GAUTAM
19BCS023
OPERATING SYSTEM
22-04-2021

What is Memory Management?

- MEMORY MANAGEMENT IS THE PROCESS OF CONTROLLING AND COORDINATING COMPUTER MEMORY, ASSIGNING PORTIONS KNOWN AS BLOCKS TO VARIOUS RUNNING PROGRAMS TO OPTIMIZE THE OVERALL PERFORMANCE OF THE SYSTEM.

- IT IS THE MOST IMPORTANT FUNCTION OF AN OPERATING SYSTEM THAT MANAGES PRIMARY MEMORY.

Memory Hierarchy and Observations

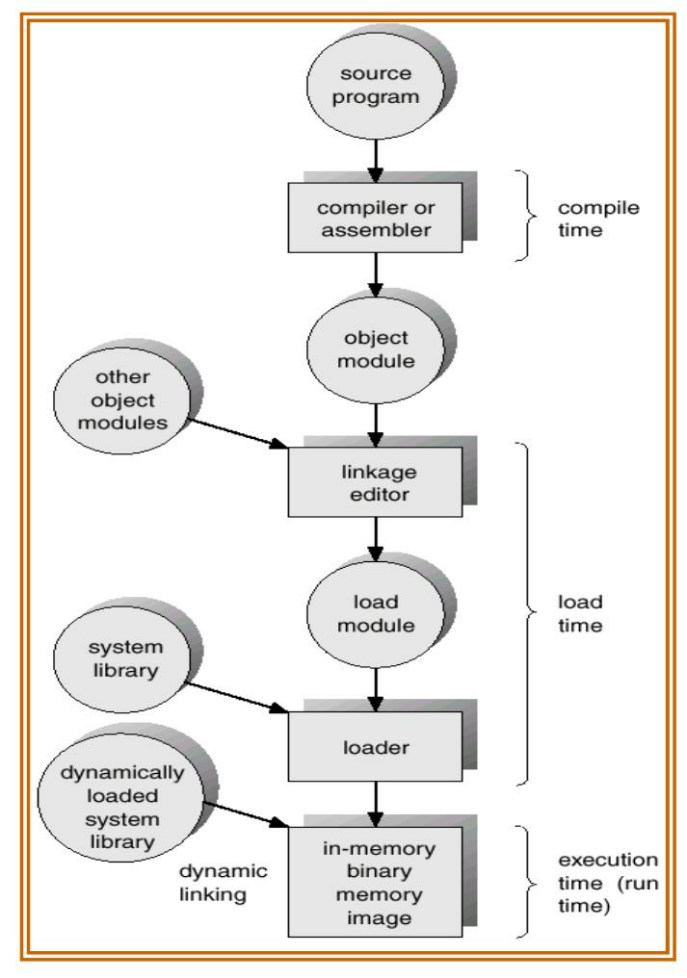


- Program must be brought (from disk) into memory and placed within a job queue for it to be run
- Main memory and registers are only storage elements which CPU can access directly
- Register access in one CPU clock (or less)
- Cache sits between main memory and CPU registers

Why Use Memory Management?

- It allows you to check how much memory needs to be allocated to processes that decide which processor should get memory at what time.
- Tracks whenever inventory gets freed or unallocated. According to it will update the status.
- It allocates the space to application routines.
- Helps protect different processes from each other
- It places the programs in memory so that memory is utilized to its full extent.

Address binding



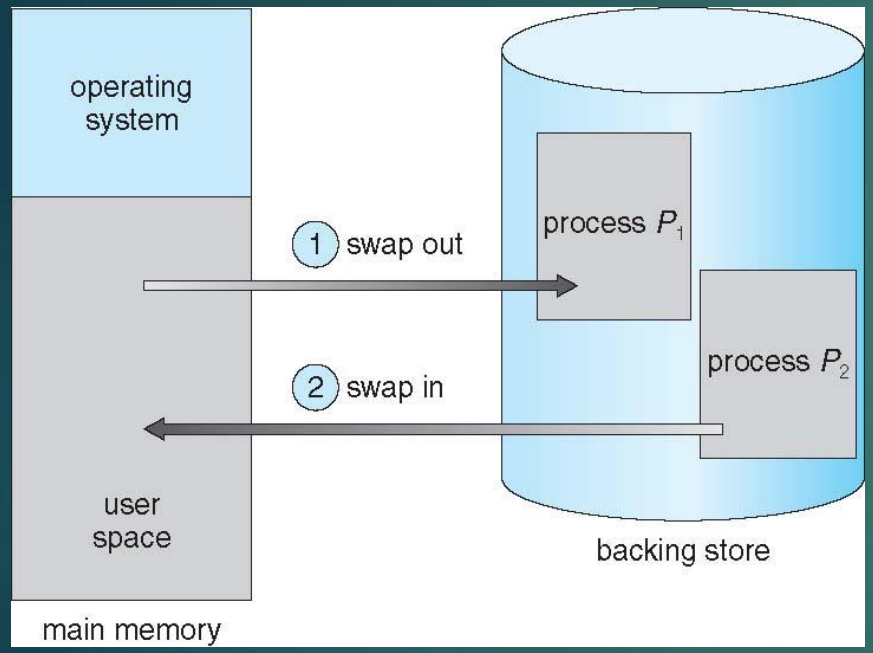
Address binding of instructions and data to memory addresses can happen at three different stages.

1. Compile time
2. Load time
3. Execution time

Logical Versus Physical Address Space

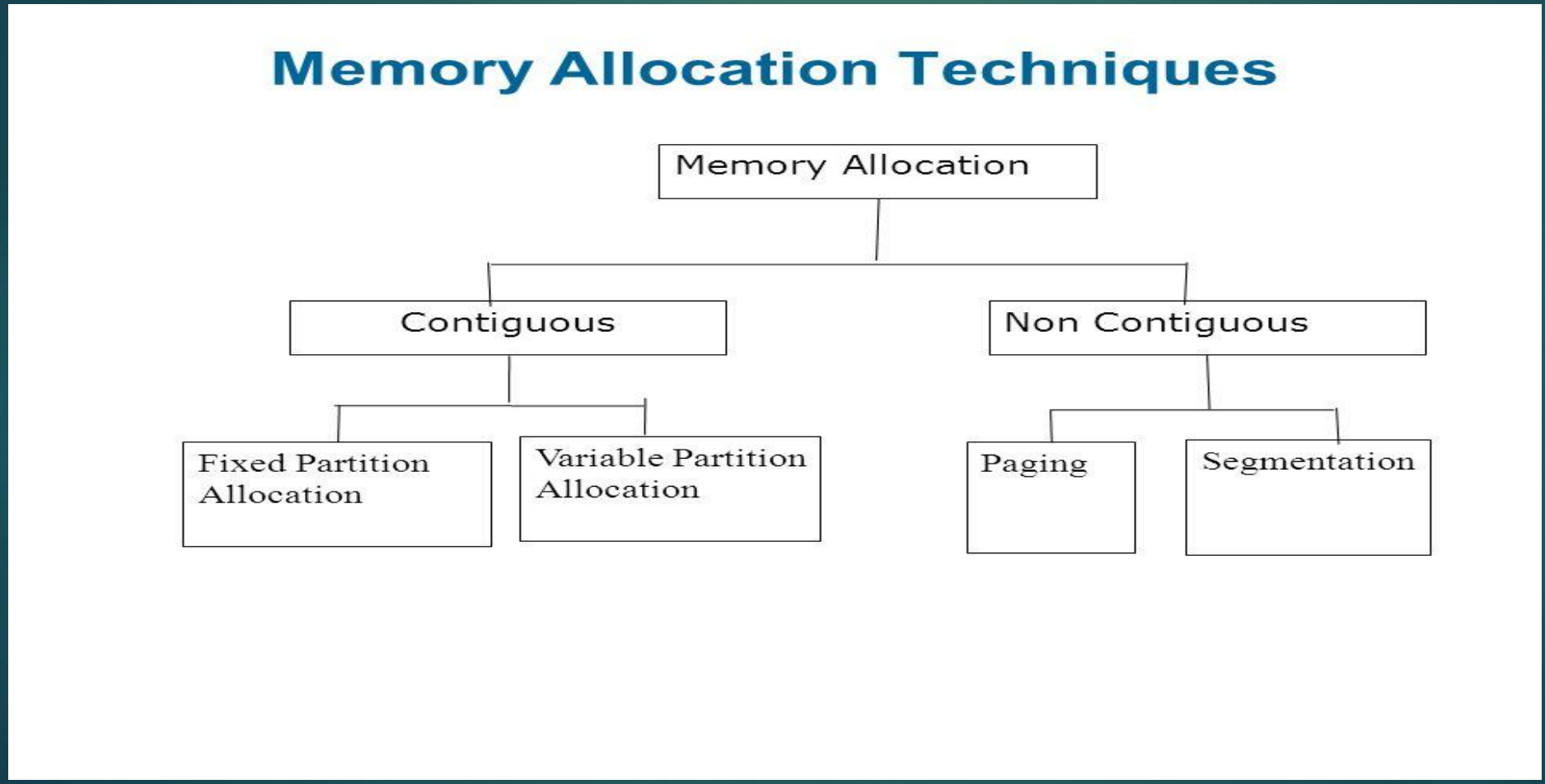
- Logical address – address generated by the CPU; also referred to as virtual address.
- Physical address – address seen by the memory unit.
- Same in *compile time and load time* address binding schemes.
- The set of all logical addresses generated by a program is a logical address space; the set of all physical addresses corresponding to these logical addresses are a physical address space.
- The run-time mapping from virtual to physical addresses is done by a hardware device called the memory management unit (MMU).

Swapping



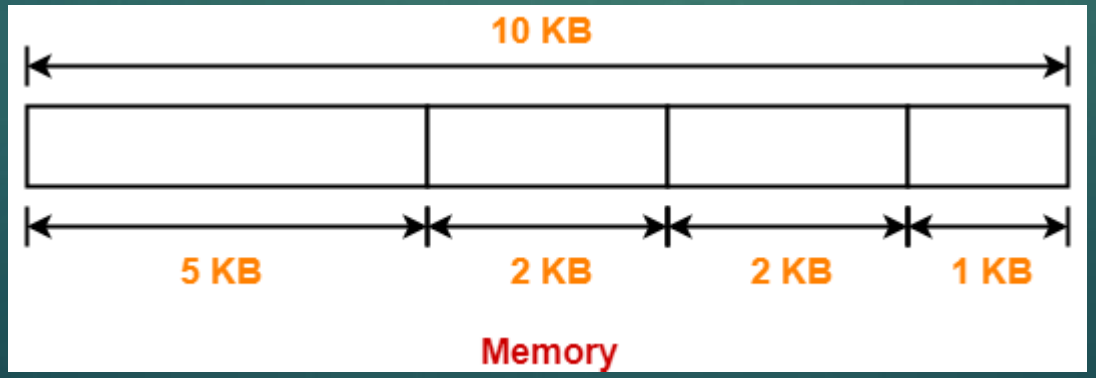
- A process can be swapped temporarily out of memory to a backing store, and then brought back into memory for continued execution.
- Backing store – fast disk large enough to accommodate copies of all memory images for all users; must provide direct access to these memory images
- Roll out, Roll in – swapping variant used for priority-based scheduling algorithms; lower-priority process is swapped out so higher-priority process can be loaded and executed
- Major part of swap time is transfer time; total transfer time is directly proportional to the amount of memory swapped
- System maintains a ready queue of ready-to-run processes which have memory images on disk

Storage allocation and Management Techniques



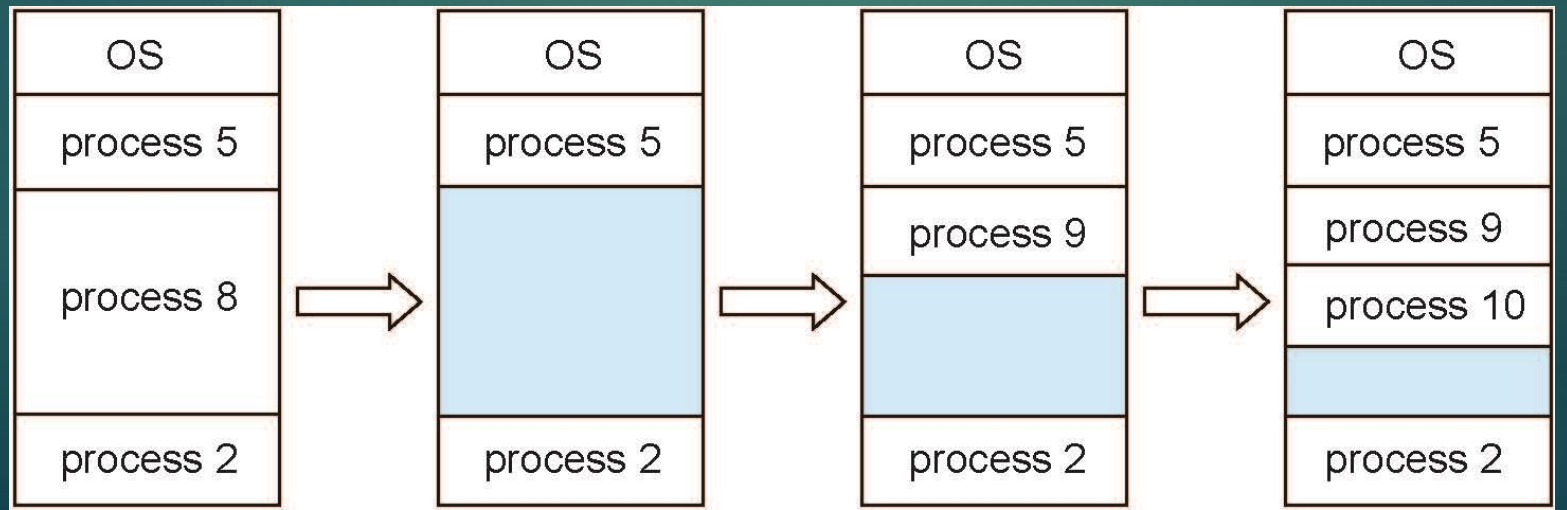
Fixed-sized partition allocation

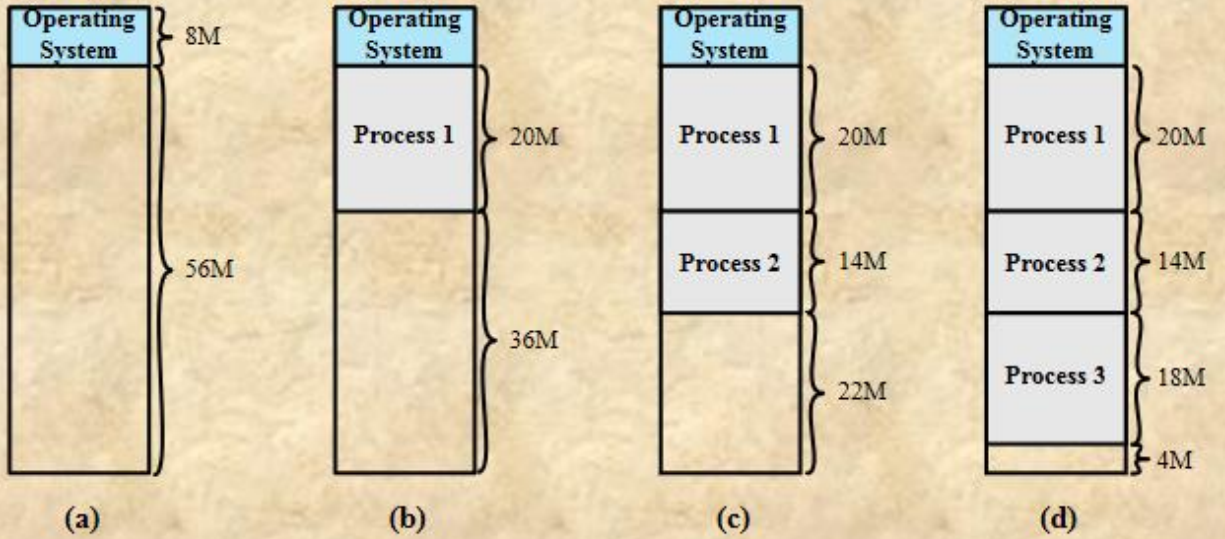
- Static partitioning is a fixed size partitioning scheme.
- In this technique, main memory is pre-divided into fixed size partitions.
- The size of each partition is fixed and can not be changed.
- Each partition is allowed to store only one process.

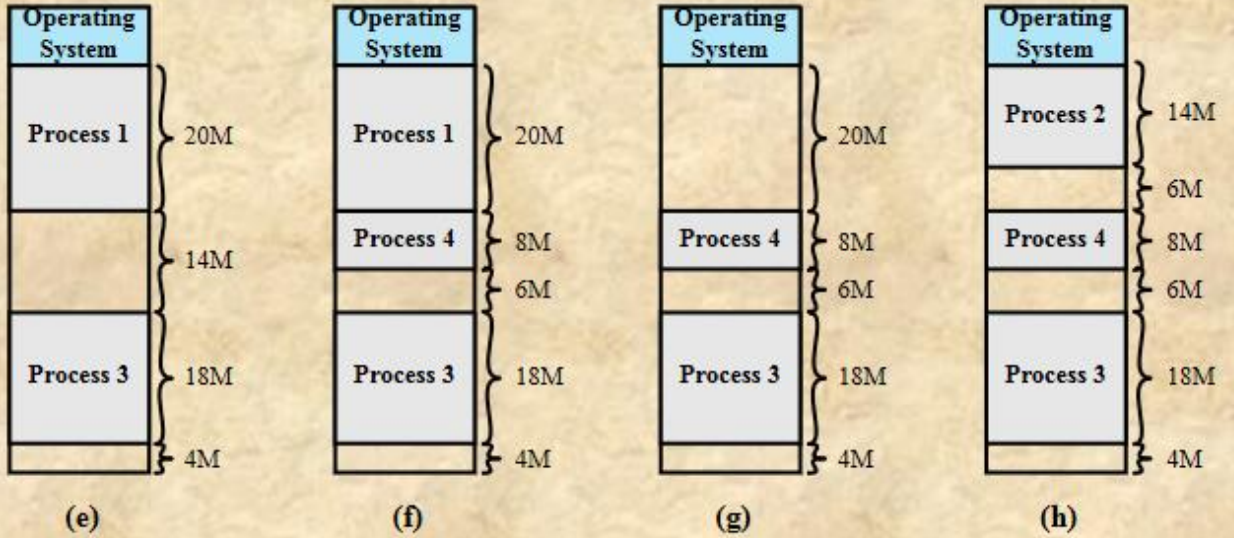


Variable-sized partition allocation

- Dynamic partitioning is a variable size partitioning scheme.
- It performs the allocation dynamically.
- When a process arrives, a partition of size equal to the size of process is created.
- Then, that partition is allocated to the process.







The effect of Dynamic Partitioning

External Fragmentation

- This phenomenon of entering and leaving the memory can cause the formation of unusable memory holes (like the unused space between two vehicles). This is known as External Fragmentation.

- What is Internal Fragmentation then?
- Internal Fragmentation- allocated memory may be slightly larger than requested memory; this size difference is memory internal to a partition, but not being used

Best-Fit, First-Fit, Worst-Fit and Next-Fit

•Four strategies that can be used to allocate memory to Variable – partition. They are the following:
•Best - Fit - chooses a partition that is smallest and whose size is greater than equal to k. It leaves small-sized unusable partitions or holes.
•First - fit - chooses the first partition whose size is greater than equal to k.
•Worst - fit - chooses the largest partition and allocates it to process p. It can leave bigger unusable partitions.
•Next-fit- same as first fit but start search always from last allocated hole

Advantages of Contiguous allocation

Static partitioning

- It is simple and easy to implement.
- It supports multiprogramming since multiple processes can be stored inside the main memory.
- Only one memory access is required which reduces the access time.

Dynamic partitioning

- It does not suffer from internal fragmentation.
- Degree of multiprogramming is dynamic.
- There is no limitation on the size of processes

Disadvantages of Contiguous allocation

Static partitioning

- It suffers from both internal fragmentation and external fragmentation.
- It utilizes memory inefficiently.
- The degree of multiprogramming is limited equal to number of partitions.
- There is a limitation on the size of process since processes with size greater than the size of largest partition can't be stored and executed.

Dynamic partitioning

- It suffers from external fragmentation.
- Allocation and deallocation of memory is complex.

Non- Contiguous allocation

- To resolve the problem of external fragmentation and to enhance the degree of multiprogramming to a greater extent, it was decided to sacrifice the simplicity of allocating contiguous memory to every process. It was decided to have a non-contiguous physical address space of a process so that a process could be allocated memory wherever it was available.

- 1. Paging
- 2. Segmentation

Paging

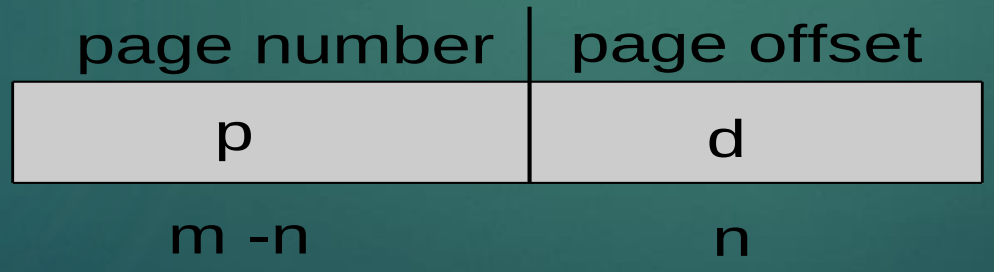
- Paging is a memory management scheme that permits the physical address space of a process to be non contiguous.
- Divide physical memory into fixed-sized blocks called frames.
- Divide logical memory into blocks of same size called pages. When a process is to be executed, its pages are loaded into any available memory frames from the backing store.
- The backing store is divided into fixed sized blocks that are of the same size as the memory frames.

Address Translation Scheme and Hardware

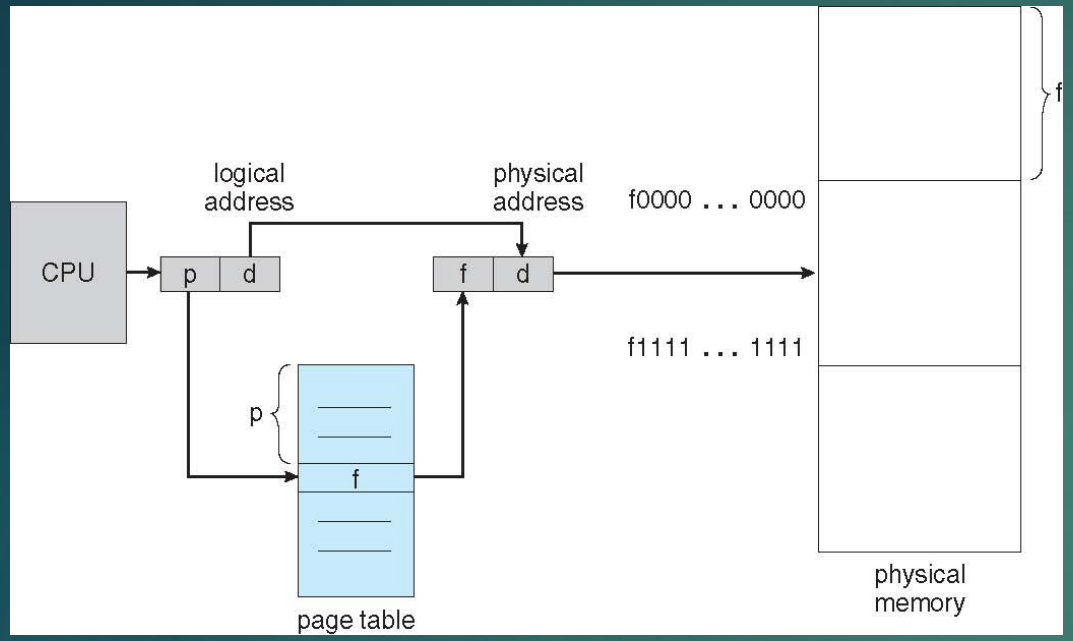
- Every address generated by the CPU is divided into two parts:

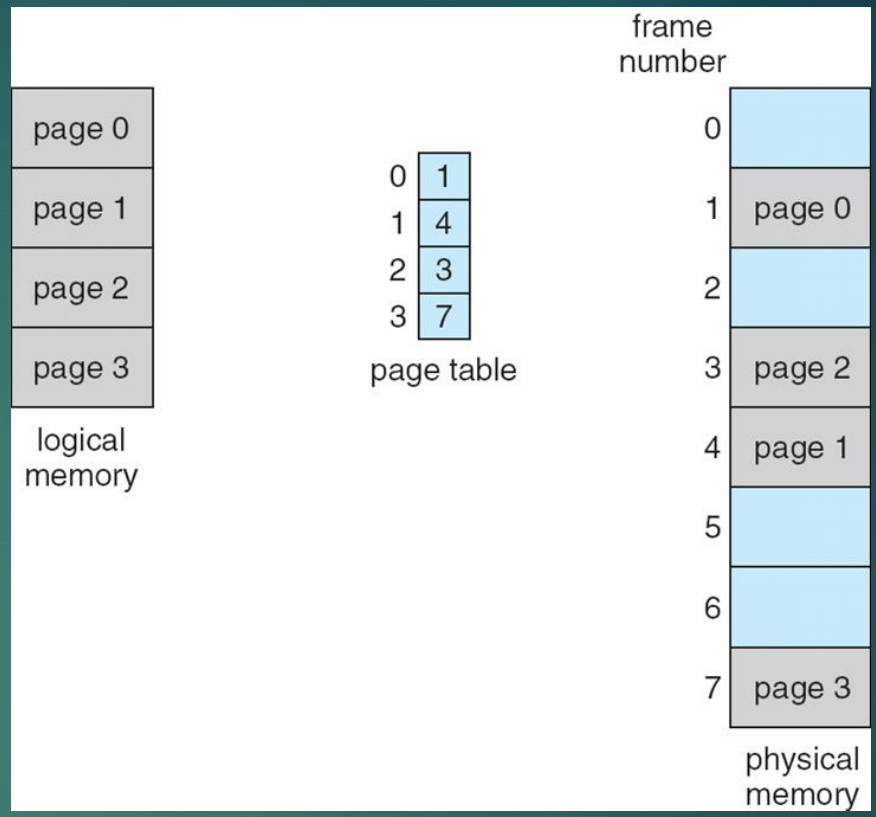
  a page number (p) and a page offset (d).

The page number is used as an index into a page table. The page table contains the base address of each page in physical memory. This base address is combined with the page offset to define the physical memory address that is sent to the memory unit.



Hardware



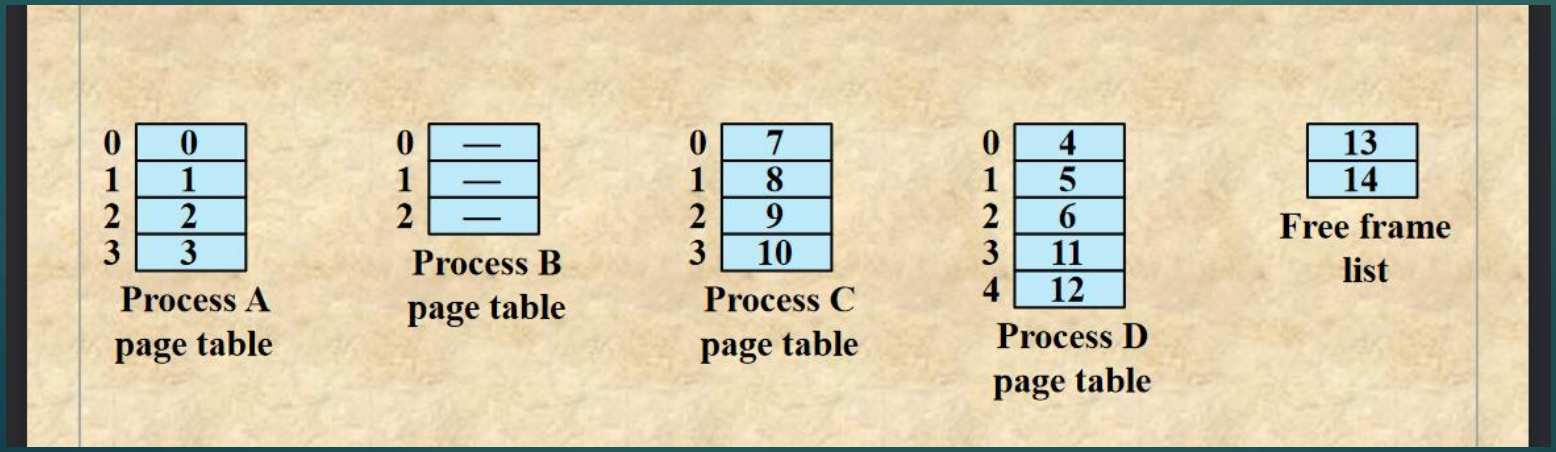


The page size is defined by the hardware. The size of a page is typically of a power of 2, varying between 512 bytes and 16 MB per page, depending on the computer architecture. The selection of a power of 2 as a page size makes the translation of a logical address into a page number and page offset particularly easy. If the size of logical address is 2m, and a page size is 2n addressing units, then the high order m-n bits of a logical address designate the page number, and the n low order bits designate the page offset.

Page Table

- Maintained by operating system for each process
- Contains the frame location for each page in the process
- Processor must know how to access for the current process
- Used by processor to produce a physical address

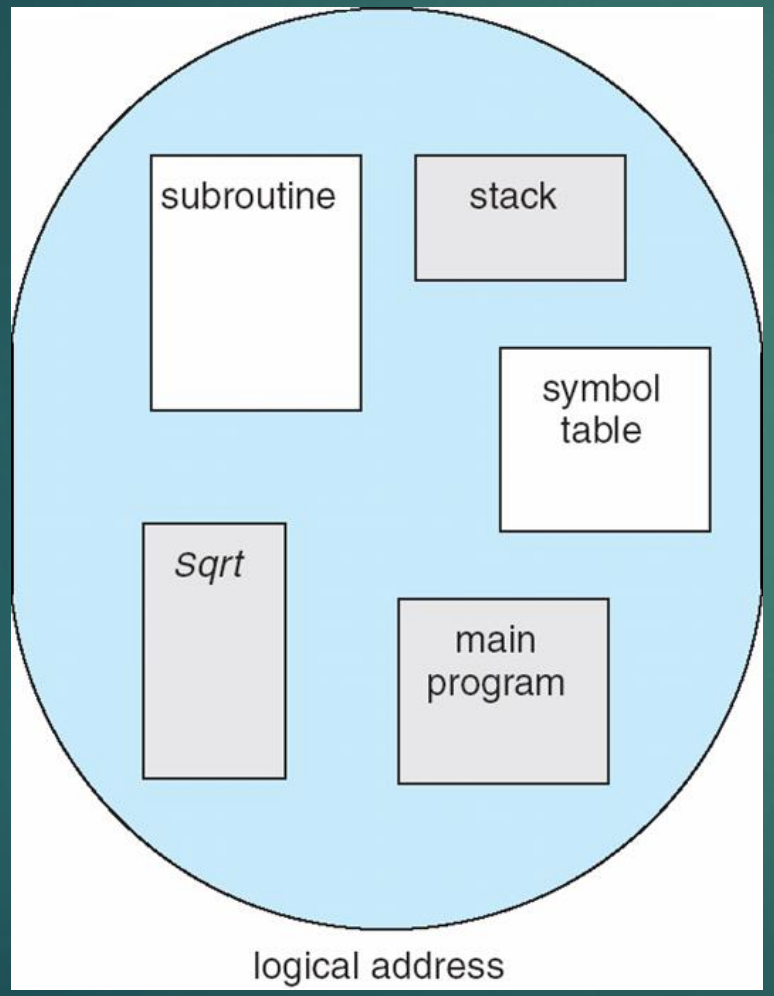
-Hierarchical Paging -Hashed Page Tables -Inverted Page Tables



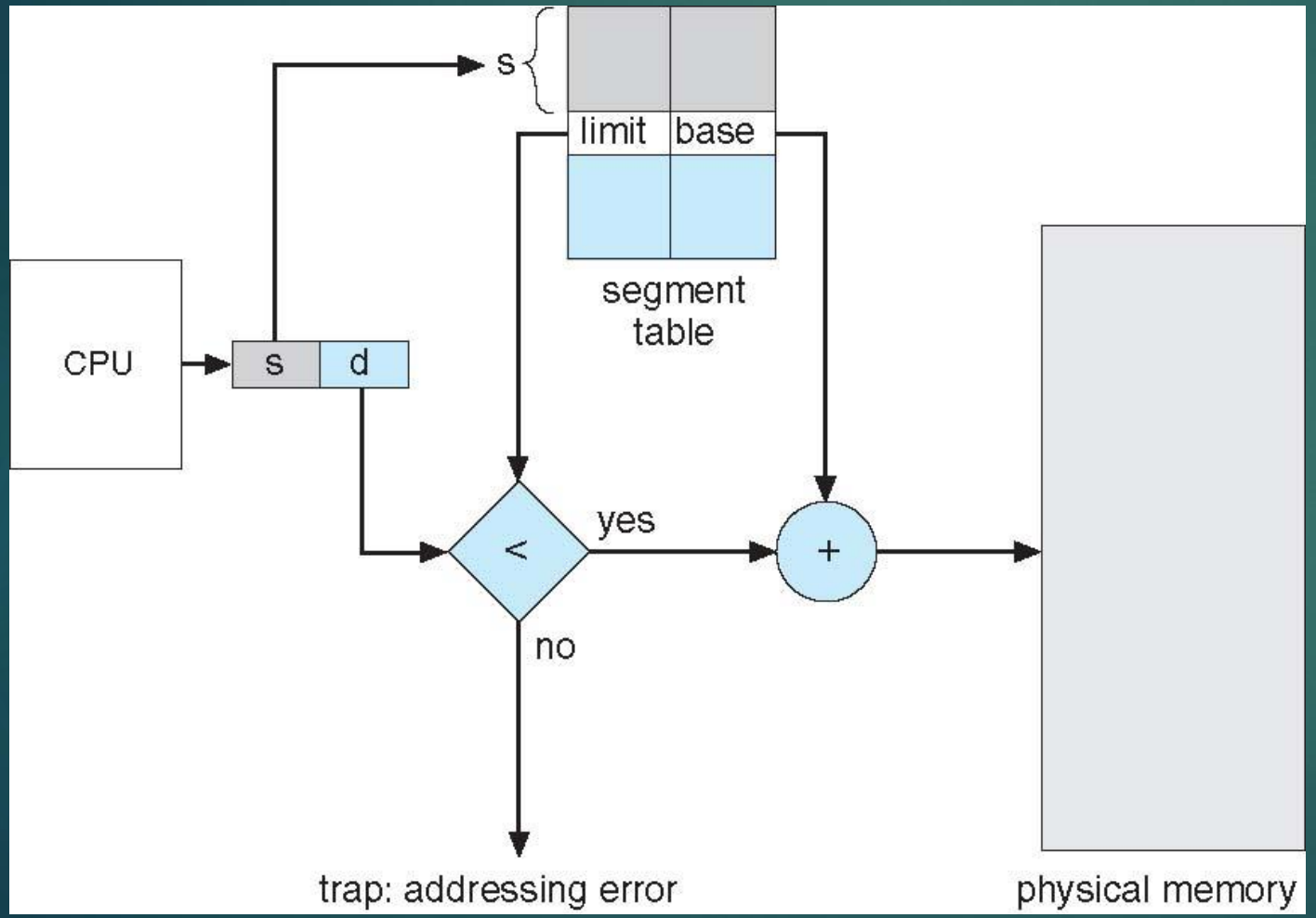
Segmentation

- Segmentation is another technique for the noncontiguous storage allocation.
- It is different from paging as it supports users' view of his program.
- For a programmer it might be more relevant to divide the logical address space of his program into variable sized segments (with respect to his view of main program, subroutines, data, etc.) than to divide it into fixed size pages.
- Such variable sized segments, which are a collection of logically related information, are the basis of segmentation technique.

Users view of program



Segmentation Hardware



Advantages of Non-Contiguous Allocation

- It increases the processing **speed**
- External Fragmentation is further reduced
- Internal Fragmentation is further reduced
- Degree of multiprogramming increases

THANK YOU