AACS3064
Computer Systems Architecture

Chapter 12: Performance Enhancement

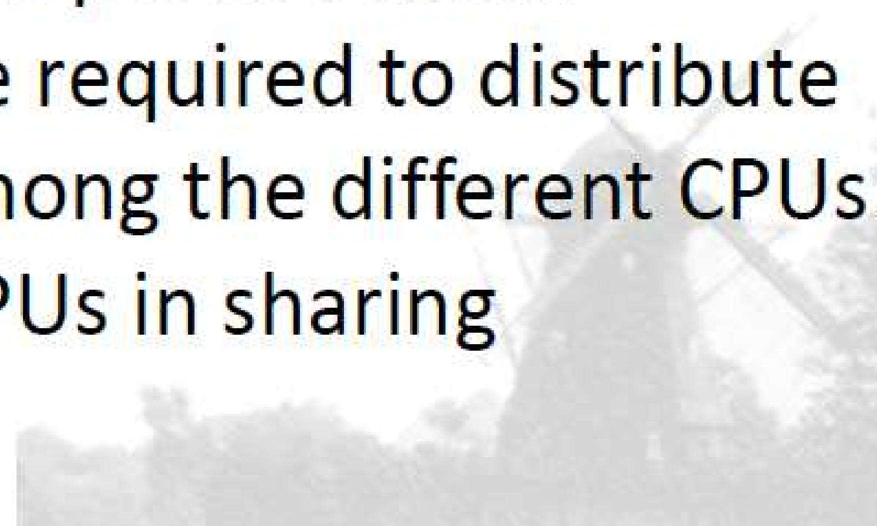
Chapter Overview

1) System Performance Enhancement and Optimization
2) Multiprocessing
3) Memory Enhancement
4) Virtual storage
5) Parallel Computers

1. System Performance Enhancement & Optimization

1. System Performance Enhancement & Optimization

- **Multiple CPUs**
  - Computers that have a multiple CPUs within a single system, are called **multiprocessors**.
  - Increasing the numbers of CPUs not necessary increase the performance.
    - ✓ Overheads that are required to distribute the instructions among the different CPUs.
    - ✓ Conflicts among CPUs in sharing resources.



1. System Performance Enhancement & Optimization (Continued)

- **Faster system clock speed**
  - Speed up the instruction cycle of the system and directly affect the performance of the system
  - Design of faster CPU circuits and buses will increase the clock speed.

1. System Performance Enhancement & Optimization (Continued)

- **Wider instruction and data paths**
  - A wider interface between CPUs and memory bus allows the CPUs to carry more instructions in a single operation and reduce disk accesses.
  - It also allows the CPUs to perform pipelining or instruction pools.

1. System Performance Enhancement & Optimization (Continued)

- **Faster disk access**
  - CPU is able to perform thousands of instructions in the time required for a single disk access. Thus, Small improvements in the disk access speeds will increase the system performance
  - Designers tend to develop disk drives with smaller size and more compactly packed, which greatly increase storage capacity.

1. System Performance Enhancement & Optimization (Continued)

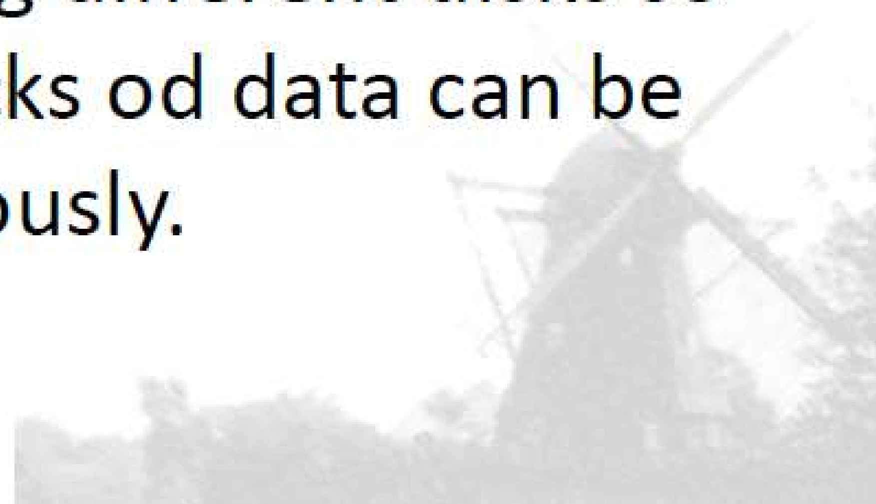
- **Memory access time**
  - Memory accesses can be reduced by providing more registers (to reduce memory access instructions) or apply extremely fast memory (known as cache memory).

1. System Performance Enhancement & Optimization (Continued)

- **Increased amounts of memory and faster memory**
  - Increased amounts of memory provide larger buffer spaces to hold additional data to enable system to continue processing during disk accesses.
  - Faster memory access time refers to memory speed. It reduces the time required to access to the instructions and data operand in memory.

1. System Performance Enhancement & Optimization (Continued)

- **Disk access**
  - Minimize the number of disk accesses is the best way in using memory.
  - In order to reduce disk access time, data can be placed among different disks so that multiple of blocks od data can be accessed simultaneously.



2. Multiprocessing

2. Multiprocessing

- Basically, Multiprocessing can be done in two different approaches:

  i. Tightly coupled systems

  ii. Loosely coupled systems

2. Multiprocessing (Continued)

Tightly coupled systems

- Also known as multiprocessor systems
- All the computers are in a tightly coupled configuration and have access to the same programs & data.
- Program executions are divided among CPUs
- Each CPU operates independently. No communication channel is required since each CPU accesses to the same memory and I/O devices.

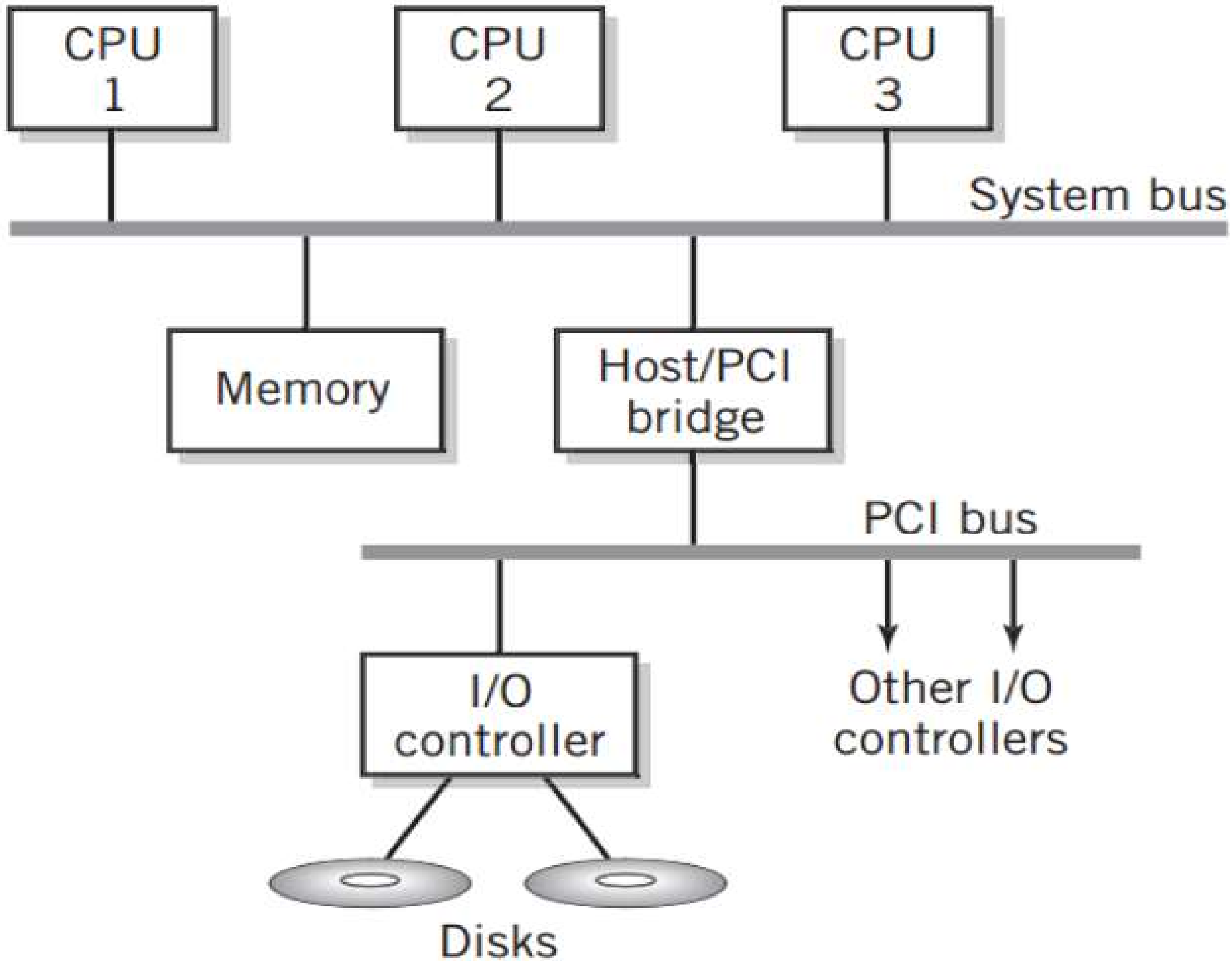
2. Multiprocessing (Continued)

Tightly coupled systems

- Theoretically, any program inside the memory can be executed by the CPUs as long as the CPU is free. But this raises the question of control of the system. Two ways to configure :

  i. Master-slave multiprocessing
  ii. Symmetrical multiprocessing (SMP)

2. Multiprocessing (Continued)



Tightly coupled systems

2. Multiprocessing (Continued)

Master-slave multiprocessing

- One CPU is set as the master, manages the system, control all resources and scheduling.

- Only master can control the operating system while the other CPUs (slave) perform the work assigned by the master.

2. Multiprocessing (Continued)

Master-slave multiprocessing

- **Advantages:**
  - Simple
  - Better protection :
    System is tightly control by the master CPU
- **Disadvantages:**
  - Low system reliability
  - Delay in job scheduling
  - Heavy burden for master

2. Multiprocessing (Continued)

Symmetrical multiprocessing (SMP)

- Each CPU has equal access to resources
- Each CPU determines what to run using a standard algorithm
- **Advantages :**
  - High reliability
  - Fault tolerant support is straightforward
  - Balanced workload
- **Disadvantages :**
  - Resource conflicts – memory, i/o, etc.
  - Complex implementation

2. Multiprocessing (Continued)

Loosely coupled systems

- A.k.a. clusters or multi-computer systems.
- Each system is complete in itself, each has its own CPU, memory, and I/O facilities.
- Each system is known as a **node of the cluster**.
- Data communications provide the link between the different computers.
- **Advantages:** Fault-tolerant, scalable, well balanced, distance is not an issue.

3. Memory Enhancement

3. Memory Enhancement

- Memory in the modern computers is usually made up of **dynamic random access memory (DRAM).**
- DRAM is inexpensive and has the ability to hold millions of bits of data but the access time of DRAM is too slow to keep up with today's CPU.

3. Memory Enhancement (Continued)

- Hence in order to overcome this bottleneck, **static random access memory (SRAM)** is introduced as alternative.
- SRAM is two or three times faster than the DRAM but it is more expensive and can only hold small amount of memory. So it is not practical solution for large amounts of memory.

3. Memory Enhancement (Continued)

- **3 approaches:**
  - **i. Wider Path Memory Access**
    - Retrieve multiple bytes instead of 1 byte at a time
  - **ii. Memory Interleaving**
    - Partition memory into subsections, each with its own address register and data register
  - **iii. Cache Memory**

3. Memory Enhancement (Continued)

Wider path memory access

- several bytes or words can be read/written between CPU and memory with each access.
- these bytes can be separated as required and processed in the usual way.
- increased the complexity in locating memory addresses, hence fast memory access become more difficult. Widely used because of its simplicity.

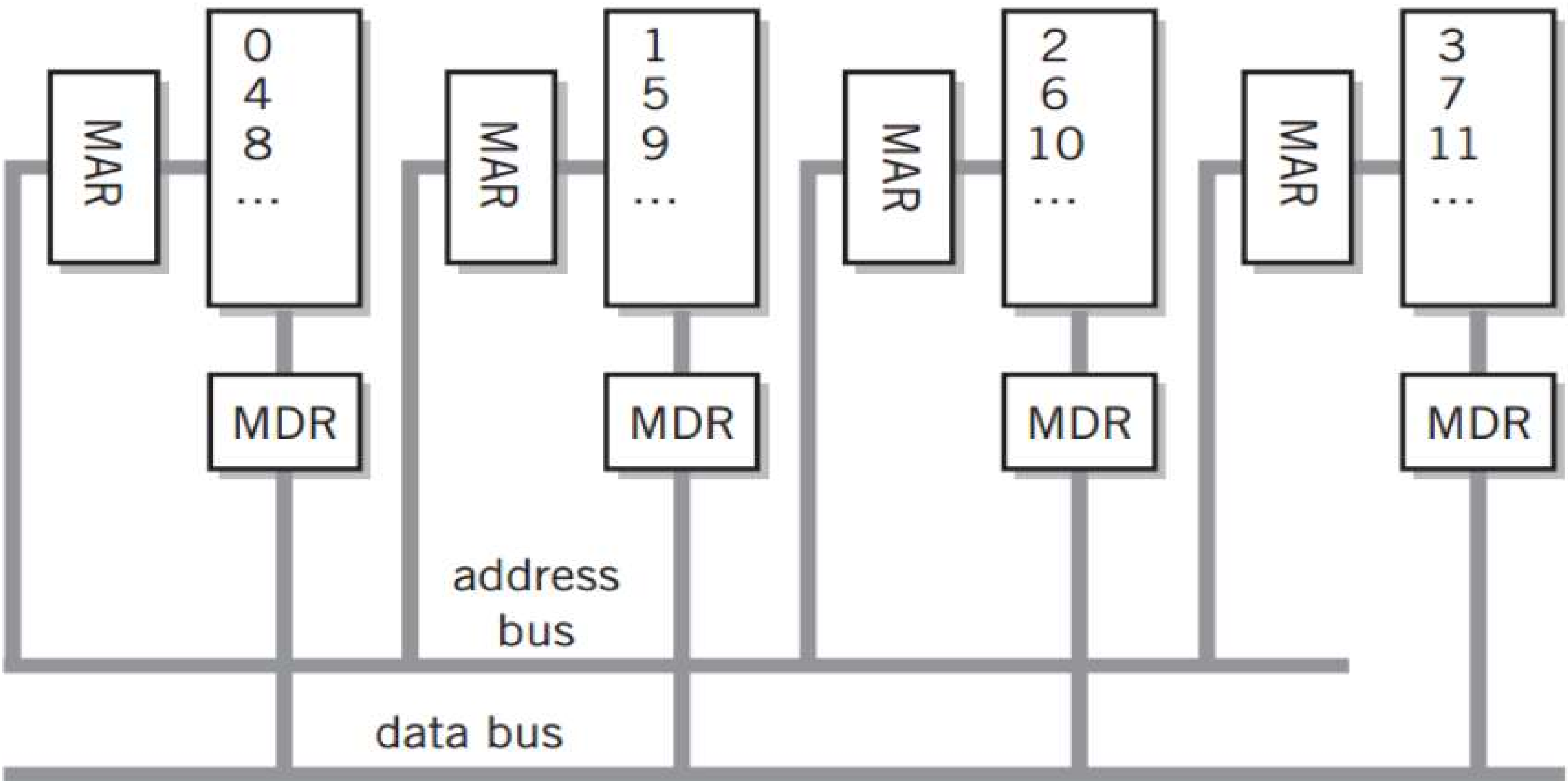
3. Memory Enhancement (Continued)

Memory interleaving

- A technique that divide memory into parts, make it possible to access more than one location at a time.
- Each part would have its own MAR and MDR. Each part is independently accessible.
- The memory cannot be accessed simultaneously if the locations are in the same block but it can be accessed if the memory locations are of different blocks.

3. Memory Enhancement (Continued)

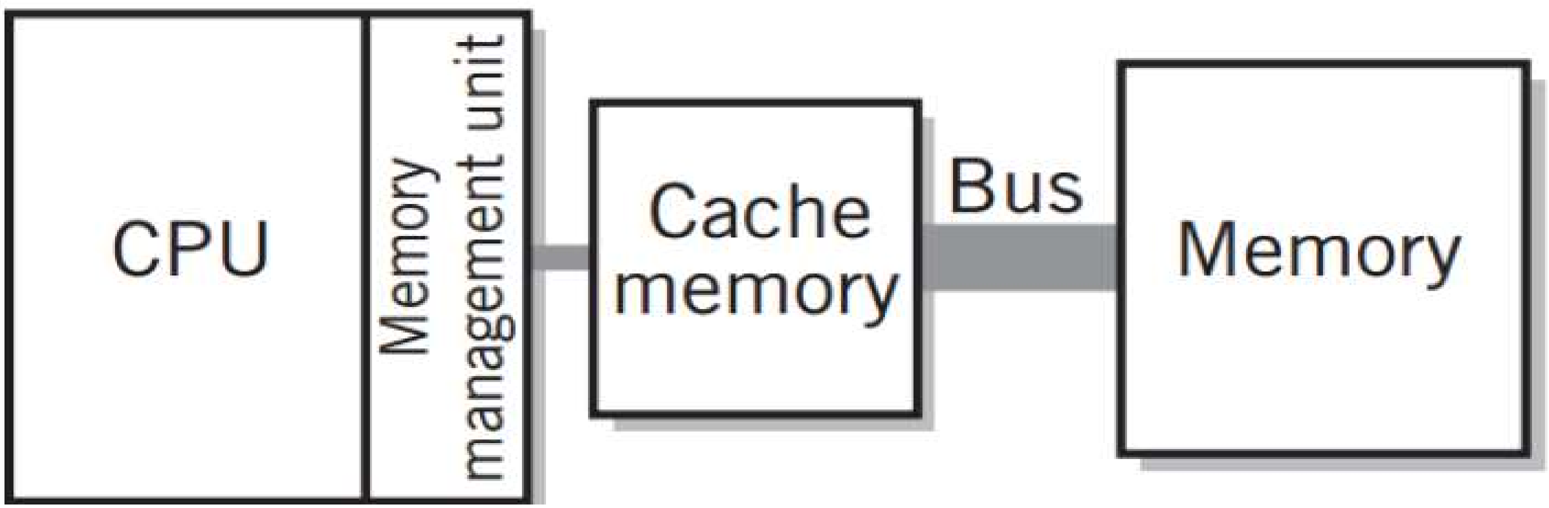
Memory interleaving



3. Memory Enhancement (Continued)

Cache Memory

- A small amount of high speed memory between CPU and main memory storage.
- It is invisible to the programmer & cannot be directly addressed in the usual way.



3. Memory Enhancement (Continued)

Cache Memory

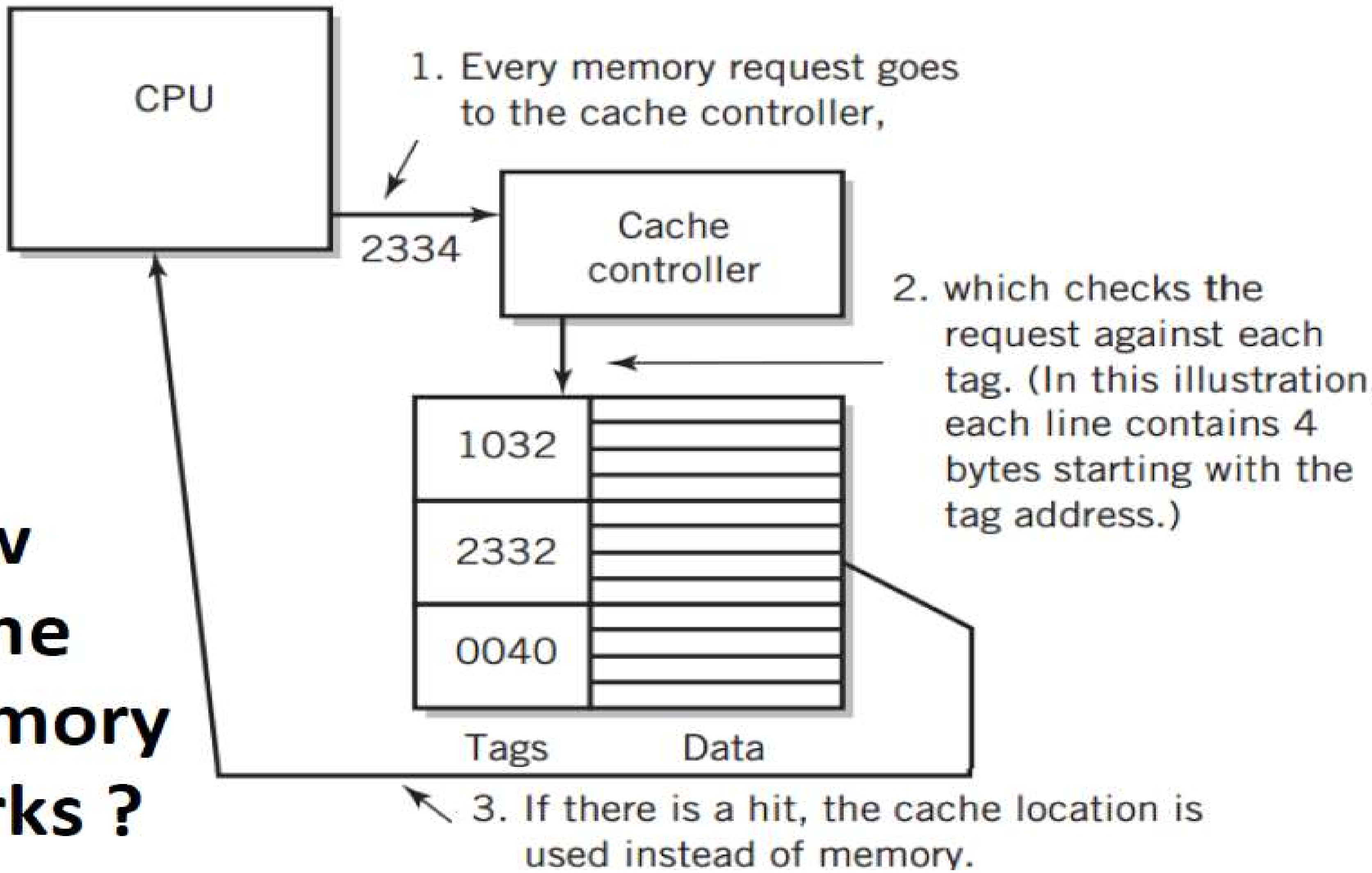
- Cache memory is organized into blocks. These blocks are used to hold an exact reproduction of a corresponding amount of storage somewhere in the main memory.
- Each block holds a tag which identifies the corresponding location of data in the main memory.

3. Memory Enhancement (Continued)

Cache Memory

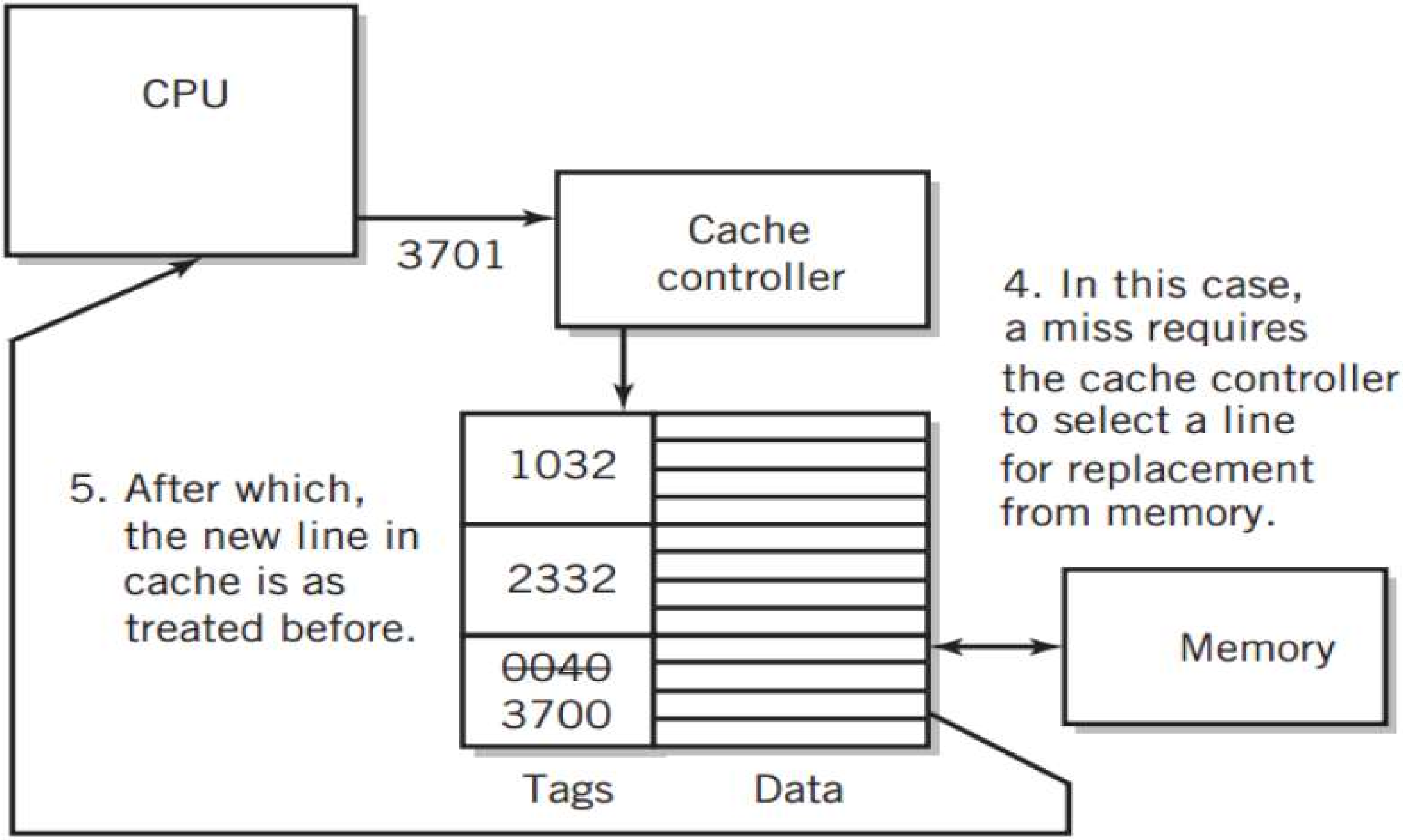
- The tags act as directory that can be used to determine exactly which storage location from main memory
- **Cache controller :** hardware that checks tags
- **Cache Line :** Unit of transfer between storage and cache memory
- **Hit Ratio**: ratio of hits out of total requests

3. Memory Enhancement (Continued)



How cache memory works ?

3. Memory Enhancement (Continued)



3. Memory Enhancement (Continued)

Expanded Memory

- Used to provide additional RAM storage.
- Slower than main memory but larger size
- Normally seen only on large computer sys.
  - Example: mainframe computer.
- Acts as a buffer between main memory and hard disk storage which reduce the disk accesses.
- It works independently, → not affect the operations of CPU or even main memory.

4. Virtual storage

4. Virtual storage

- A.k.a. Virtual memory
- Com is able to conceptually split the address used in a program from the addresses that identify physical locations in the memory.
- The program addresses are referred to as **logical address.** They do not have any reality outside the program itself.
- The actual memory addresses are called **physical addresses.**

4. Virtual storage (Continued)

- Virtual storage creates a correspondences between logical & physical addresses, so that each logical address is automatically convert into physical address during the execution. This method is called **mapping**.

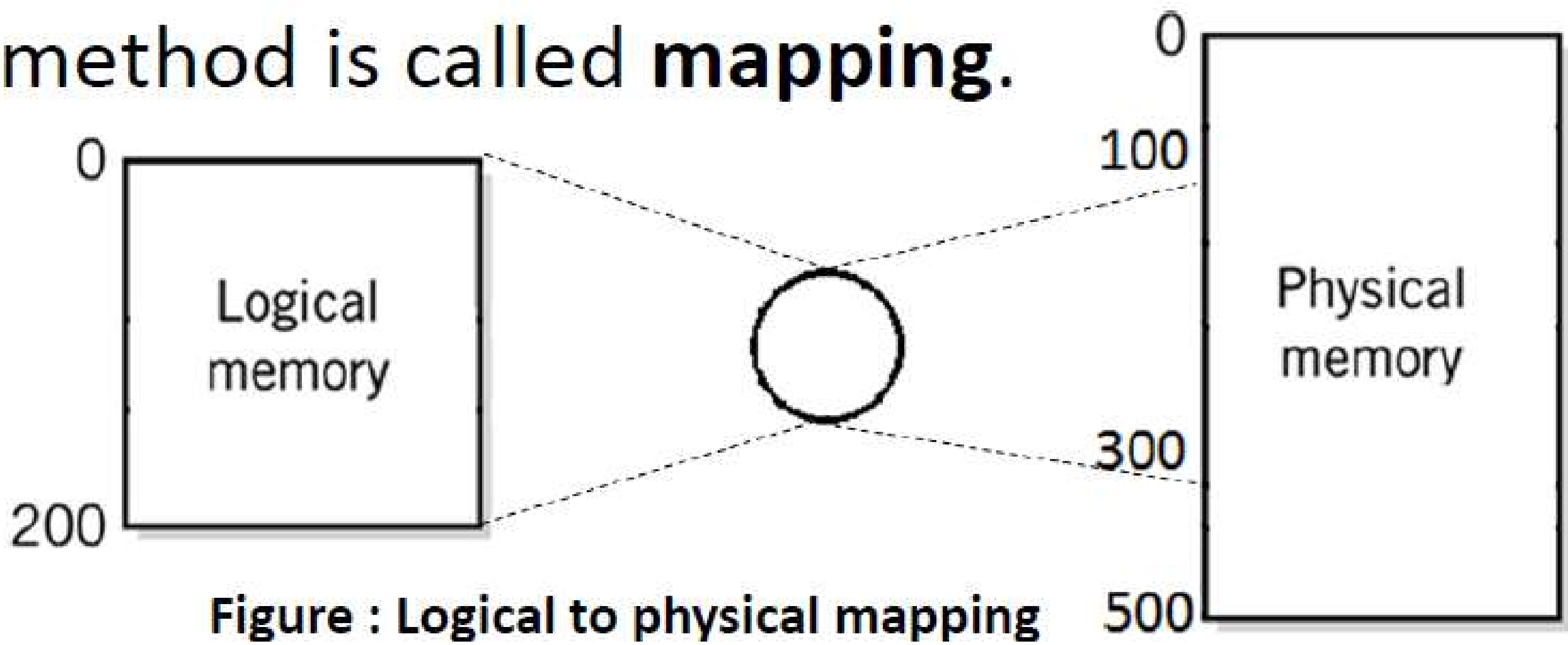
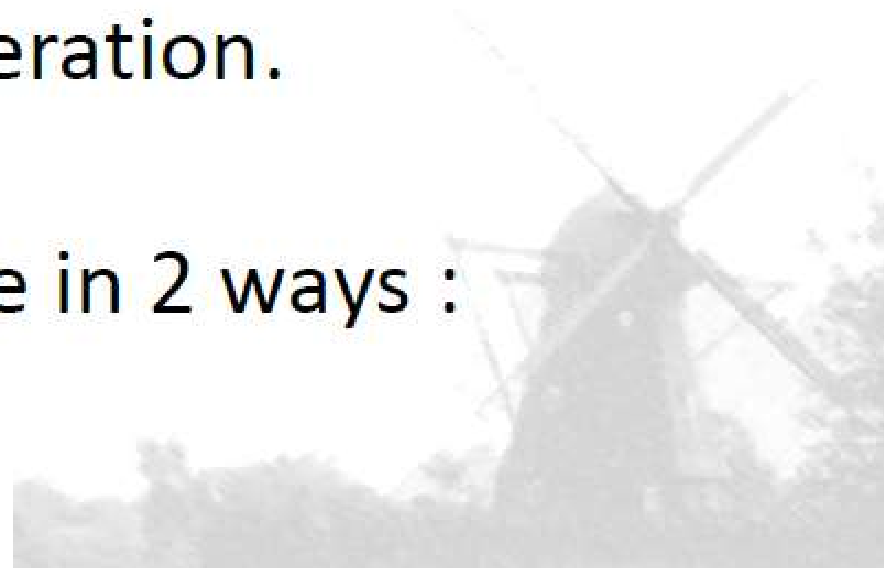


Figure : Logical to physical mapping

4. Virtual storage (Continued)

- The mapping of logical and physical address is dynamic. It can be changed any time by the OS to meet the different conditions within the system. A memory management unit performs this operation.

- Virtual storage is done in 2 ways :
  - Segmentation
  - Paging



4. Virtual storage (Continued)

Segmentation

- With segmentation, it has a lookup tables which contains the starting addresses for logical and physical memory and the size of the block.
- The physical location of a particular memory cell is located by adding the offset from the starting of corresponding physical address in the table.

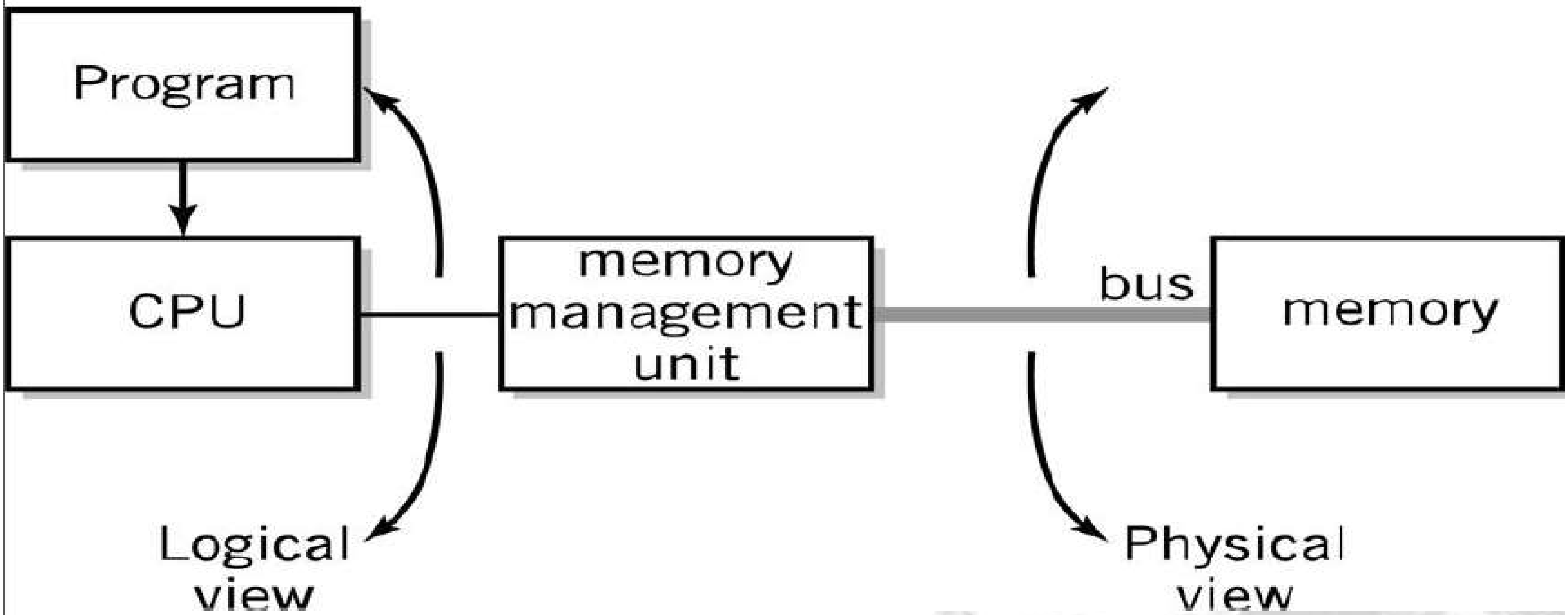
4. Virtual storage (Continued)

Paging

- Managed by the operating system
- Built into the hardware
- Independent of application
- The size of the block is defined for the system, and is identical for every block.
- Each ***logical block*** is called a **page** and the corresponding ***physical block*** is called **frame**

4. Virtual storage (Continued)

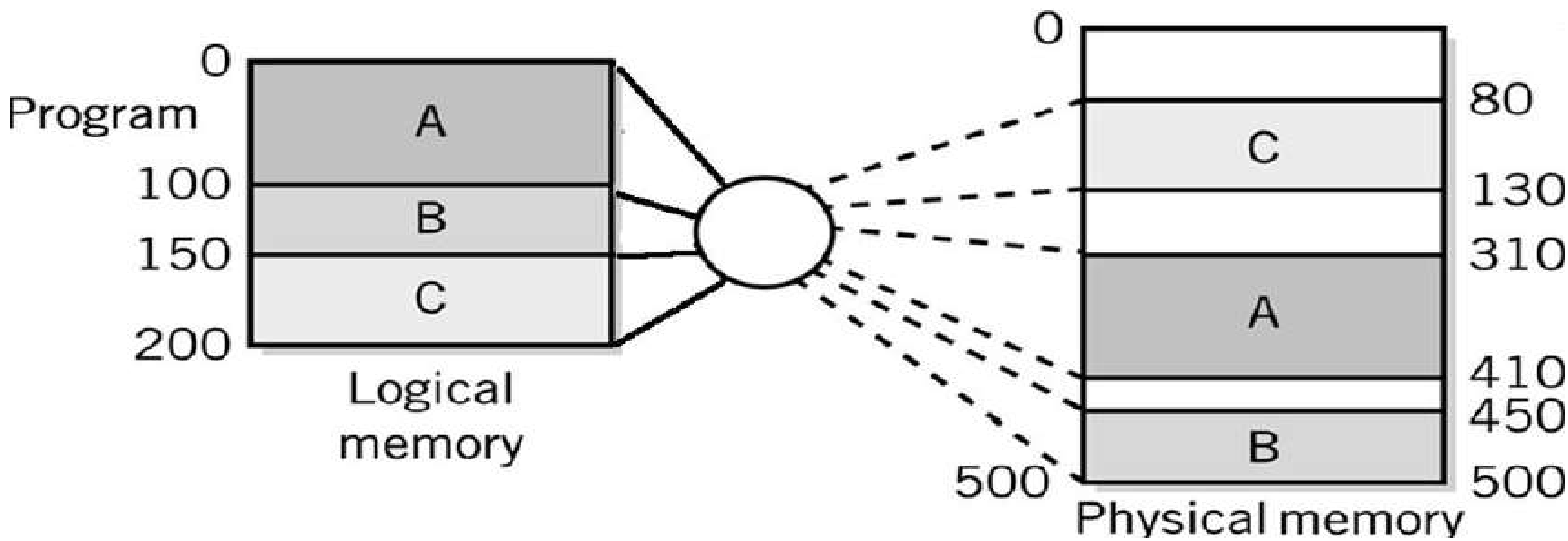
Virtual storage : Paging



4. Virtual storage (Continued)

Logical vs. Physical addresses

- Logical addresses are **relative** locations of data, instructions and branch target & are **separate** from physical addresses



4. Virtual storage (Continued)

Logical vs. Physical addresses

- The block size is typically 4KB, is chosen so that the bits of a memory address can be naturally divided into a page number and an offset. Thus each line of the page table only consists of a page number and its corresponding frame number.
- The goal of the virtual storage is to increase the perceived amount of memory available in a system.

4. Virtual storage (Continued)

Logical vs. Physical addresses

- Provides important capabilities.
  - The ability of relocating programs easily from one part of memory to another. This ability is very important to the multitasking.
  - The ability to split program into smaller pieces that can be loaded into different part of memory.

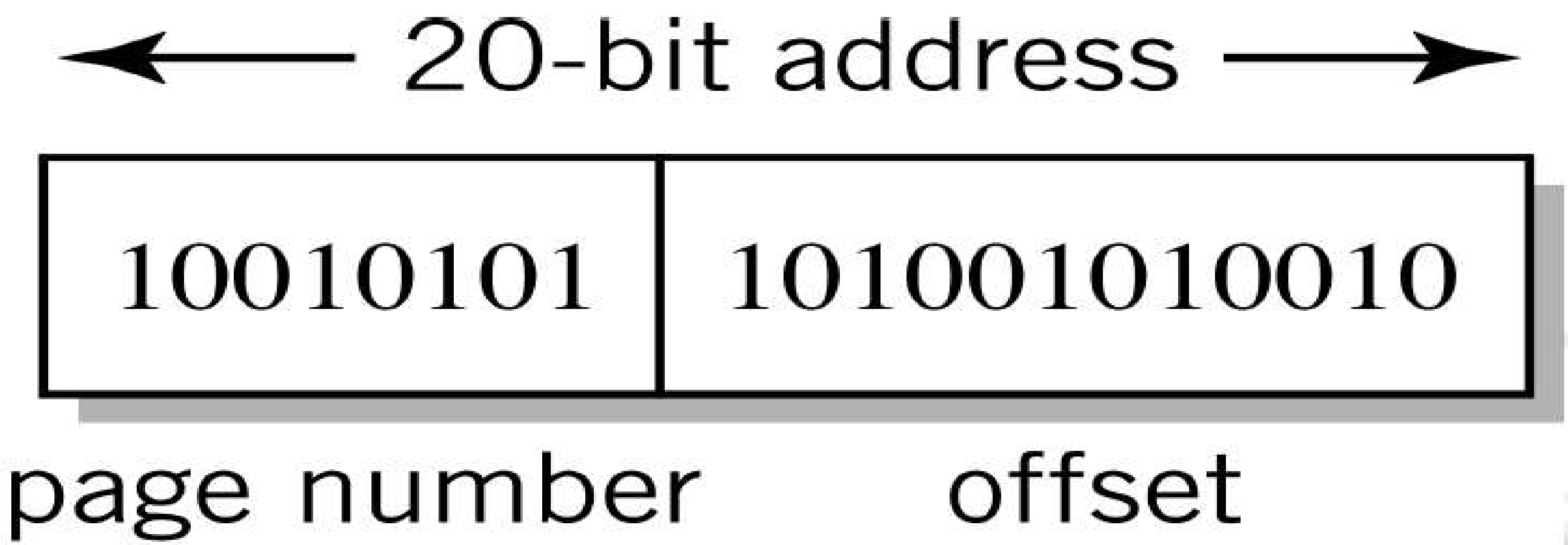
4. Virtual storage (Continued)

Logical vs. Physical addresses

- Provides important capabilities.
  - Same logical address of two different programs transformed into different physical location or vice versa.
  - Makes a program think itself has more memory than actually exist on the computer.

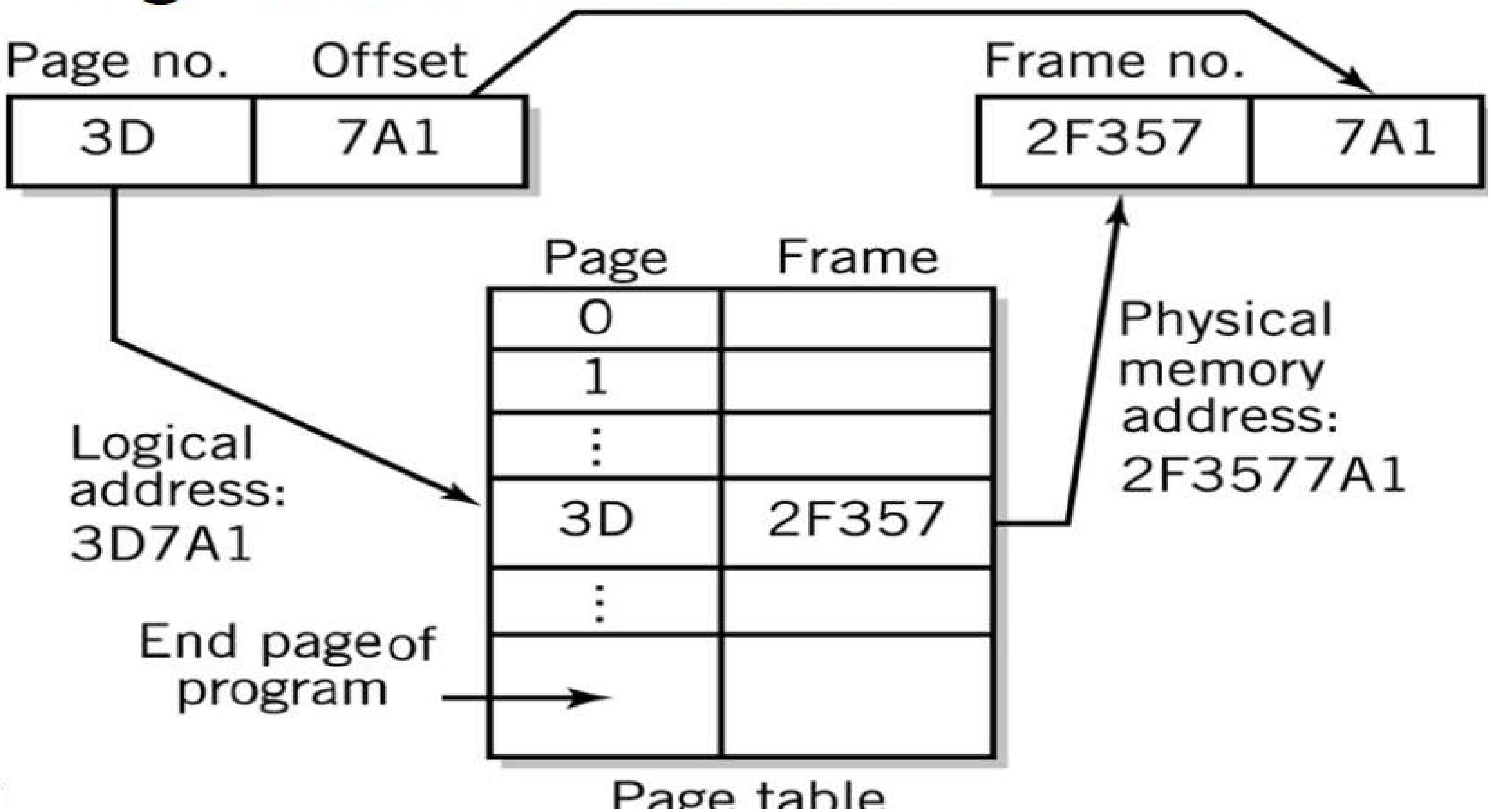
4. Virtual storage (Continued)

Page Address Layout



4. Virtual storage (Continued)

Page Translation Process



4. Virtual storage (Continued)

Page Translation Process

- **Advantages:**
  - Programs can share memory space
  - More programs run at the same time
  - Programs run even if they cannot fit into memory all at once.
  - Process separation
- **Disadvantages:**
  - SWAP file takes up space on disk
  - Paging takes up resources of the CPU

5. Parallel Computers

5. Parallel Computers

Parallel Computers

- A system that uses a large numbers of CPUs that are connected together.
- Also called massively parallel architectures.
- one of architectural for the high performance computing / supercomputing.
- used to solve complex problem such as prediction of the economic and social effects of a particular political policy, analysis of weather patterns.

5. Parallel Computers (Continued)

Parallel Computers

- CPUs have small amounts of local memory
- All CPUs have access to global shared memory
- Pipelined CPUs
  - results from one CPU immediately flow to the next CPU for further/additional processing
- **Examples of application**: Image processing, data searching in a massive database.

Chapter Review

1) System Performance Enhancement and Optimization
2) Multiprocessing
3) Memory Enhancement
4) Virtual storage
5) Parallel Computers