Chapter 08: The Memory System

Lesson 01:
Basic Concepts

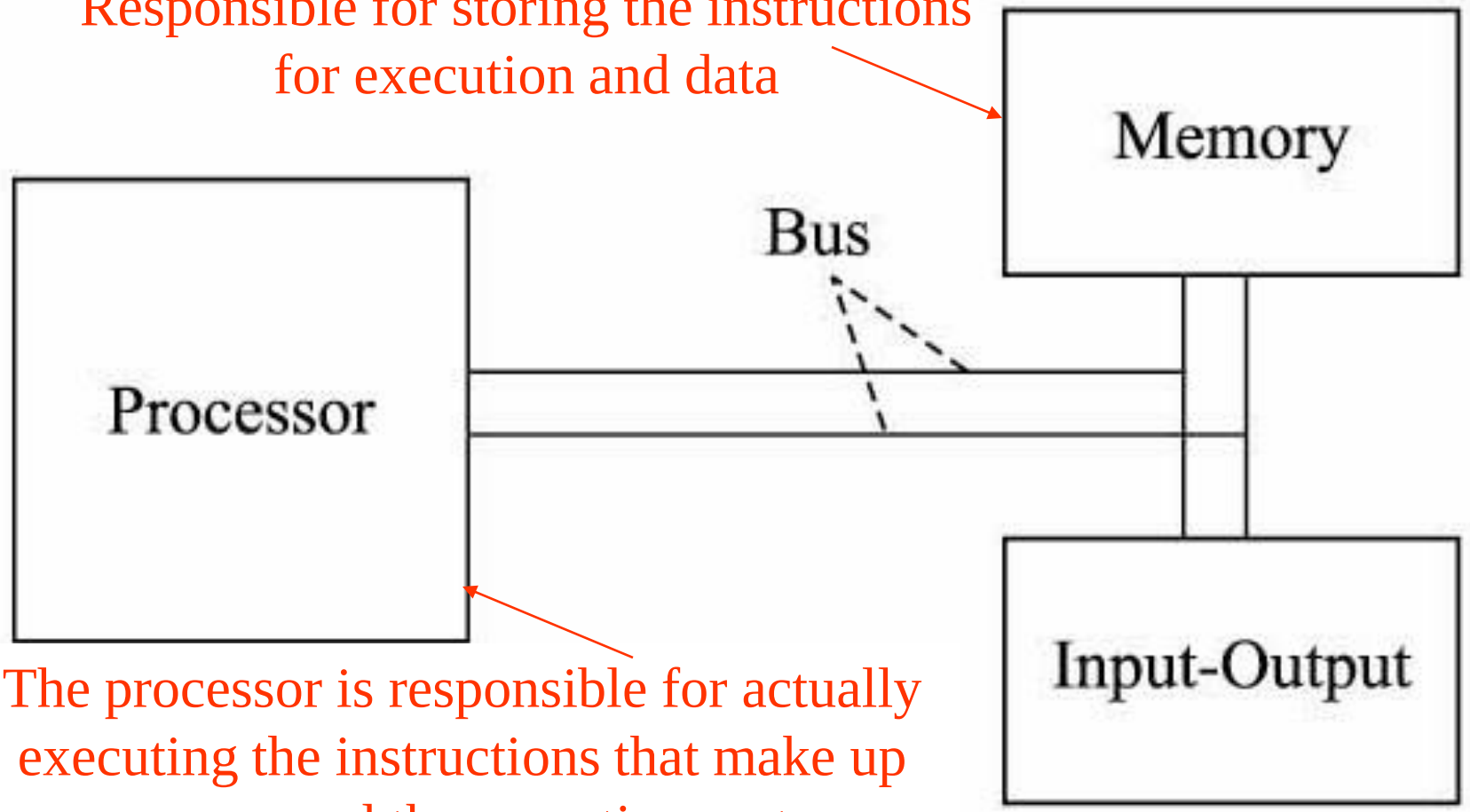
Objective

- Understand the concepts of interconnecting processor to memory devices
- Understand the speed of access of memory-devices, latency and bandwidth
- Learn how they relate in a memory system

Buses of Memory System

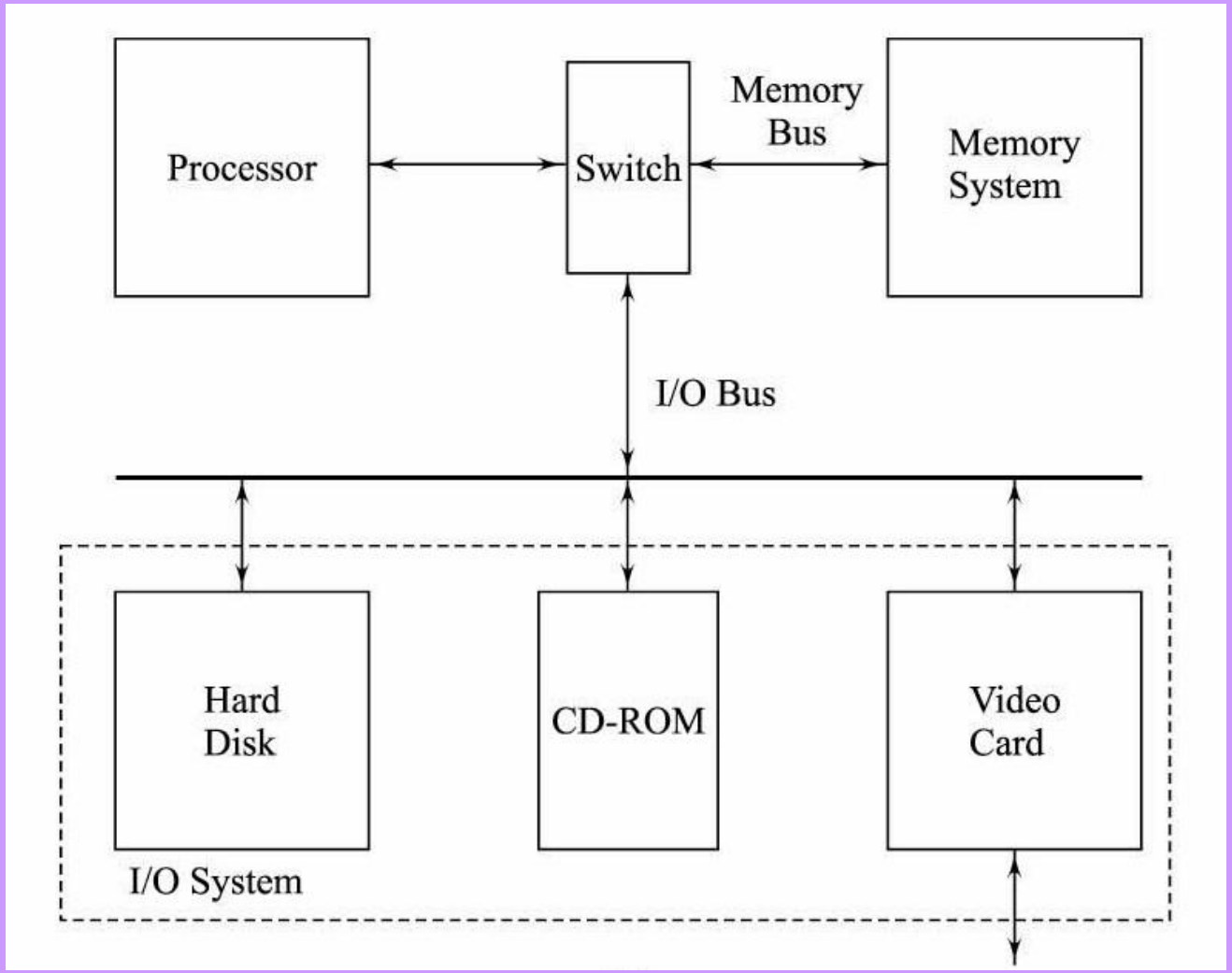
A simple view of Computer Organisation

Responsible for storing the instructions for execution and data



The processor is responsible for actually executing the instructions that make up programs and the operating system

A view of Computer Organisation



The Von Neumann stored-program computer architecture

- Memory stores the data and program, and processor fetches the instruction and data from the memory and stores the data to memory

Previous Assumptions for Memory System

Assumption so far

- Until now, treated the memory system as a "black box", which stores the program and data, from which the processor fetches the instructions to execute at the processing subunits, and into which the processor could place data for retrieval later

Assumption so far

- All memory operations take the same amount of time to complete, and that each memory operation finishes before the next one begins
- As processing logic circuit performance increases, the instruction throughput also increases proportionately in instruction throughput

Not Considered so far

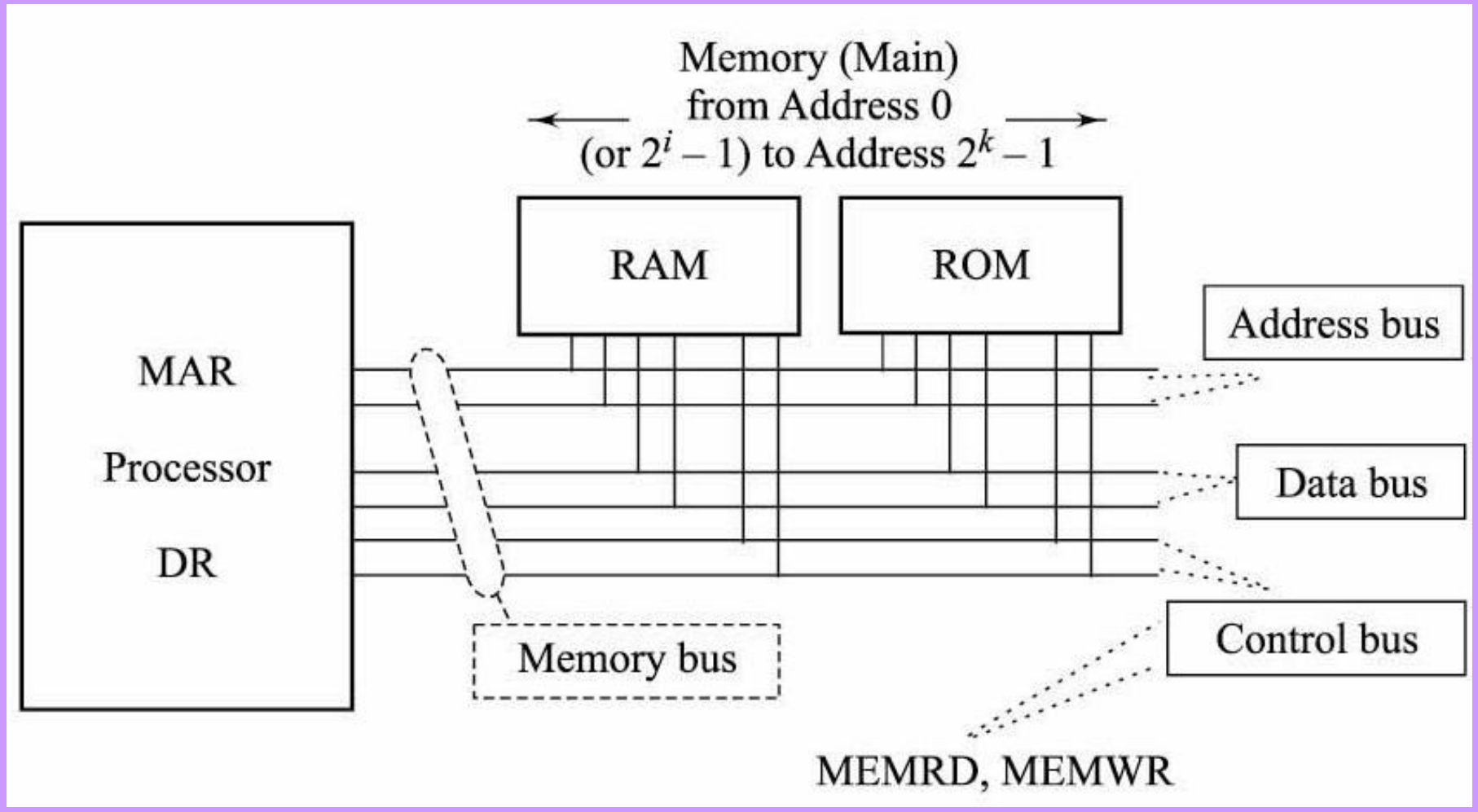
- How does the memory system affects the instruction throughput in our discussions thus far?

Buses for Memory Interconnection

Memory Interconnection

- Memory system interconnection with the processor through three buses
- Address bus, data bus, and control bus

Interconnection of the processor and memory system



Addresses

Addresses

- Memory units have addresses
- Can store one byte at one address
- A word has 1 or 2 or 4 bytes in an organisation
- A word stores as either little-endian or big-endian

Addresses

- The address bus from processor is of 32 bits
- There are 4 G addresses at memory for 4 GB
- 4 G addresses─ Between 0 and $(4 \times 1024 \times 1024 \times 1024 - 1)$
- We assume by convention that 1 K memory = $2^{10} = 1024$
- 1 M memory = $2^{20} = 1024 \times 1024$
- 1 G = $2^{30} = (4 \times 1024 \times 1024 \times 1024) = 1K \times 1K \times 1K$

Fetch or Store

Instruction-fetch

- When a PC (program counter) writes the MAR (memory address register), which connects address bus to memory system

Transfer on fetch of instruction or data word

- Transfer to MDR
- MDR─ a register in the processor, which interconnects to the data bus of the memory system

The control signals from the processor

- Read (MEMRD)
- Write (MEMWR) to the memory system
- Active during transfer from the memory-address to MDR or MDR to memory-address else inactive

Number of words accesses per cycle

Memory access cycle

- Starts from the program counter PC or other operand transferring the address to MAR
- Issuing of control signals
- Data-bus transferring to or from MDR
- Finally, deactivating control signals to memory

Number of Words/Cycle

- Early computers only one word (= 32 in 32-bit data bus) can be fetched in one cycle of fetch

More Number of Words/Cycle

- Modern computers, however, have more complexity

RAM, ROM and Cache

ROM

- **ROM**— the read-only memory
- Cannot be modified by the computer but may be read
- In general, the ROM— to hold a program that is executed automatically by the computer every time it is turned on or reset

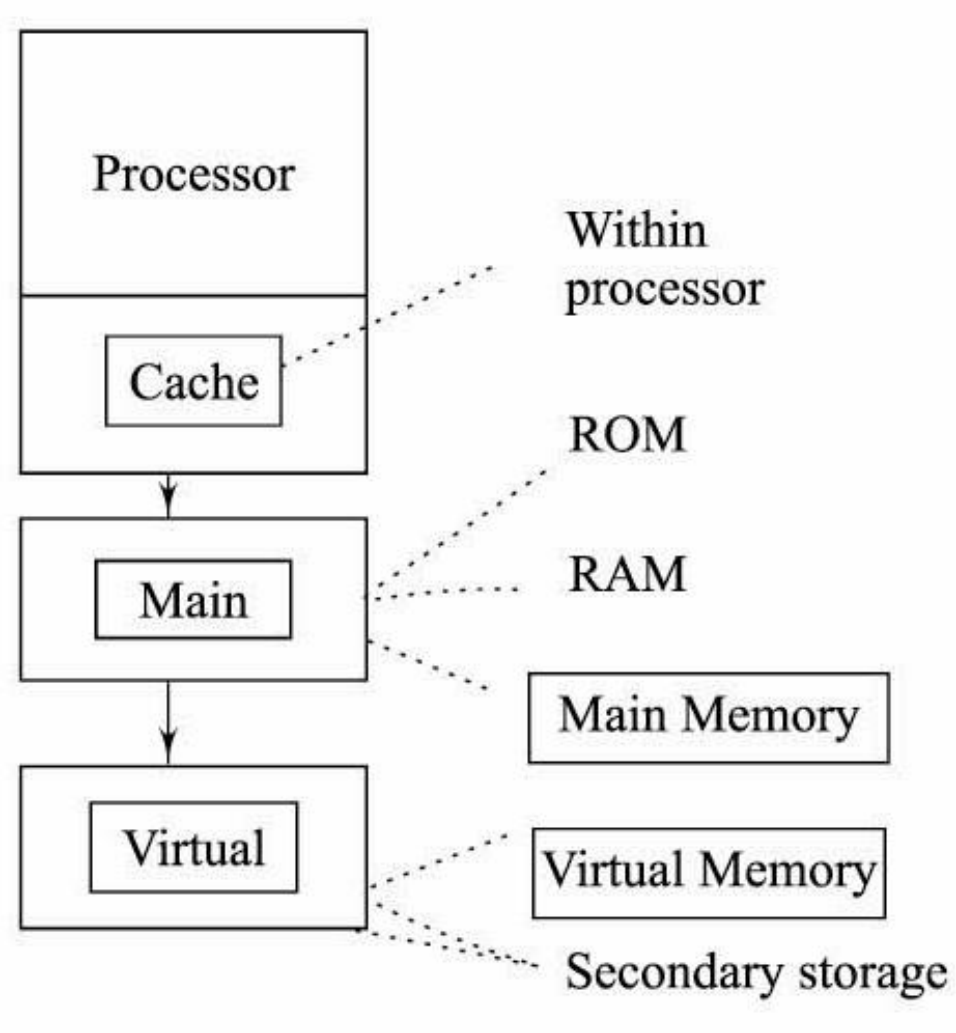
RAM

- Random-access memory─ can be both read and written
- Used to hold the programs, operating system, and data required by the computer
- Volatile─ does not retain the data stored in it when the computer's power is turned off

Cache

- Memory access cycle time > instruction execution time taken by the processing unit
- Cache memory inside the processor used for issuing instructions to the execution unit at a fast rate

Memory System



Virtual memory

Virtual Memory

- Physical addresses— Memory addresses issued by the processor for the semiconductor memory RAM or ROM
- A program specified addresses— could be different from the physical location of the addresses in the RAM or ROM
- A program can have a large number of addresses, as the program size is unlimited

Memory size and Virtual Memory

- Memory size─ , however limited by the actual presence of the addresses
- Virtual addresses─ the specified addresses in stored program
- Virtual addresses─ translated and mapped to physical addresses when a program or program section loads into the memory chips
- Also called logical addresses

Secondary memory

Secondary Memory

- Stores the OS, application programs, and the programs' data
- Processor executes the instructions after these or part of these is loaded into RAM

Secondary Memory

- ROM has a program called the bootstrap, or "boot" loader, which instructs the computer to load its operating system (OS) off of its hard disk or other secondary storage IO device

Summary

We learnt

- Basic concepts of addresses
- Number of words/cycle
- RAM, ROM, Cache, Virtual and secondary memory

End of Lesson 01 on

Basic Concepts