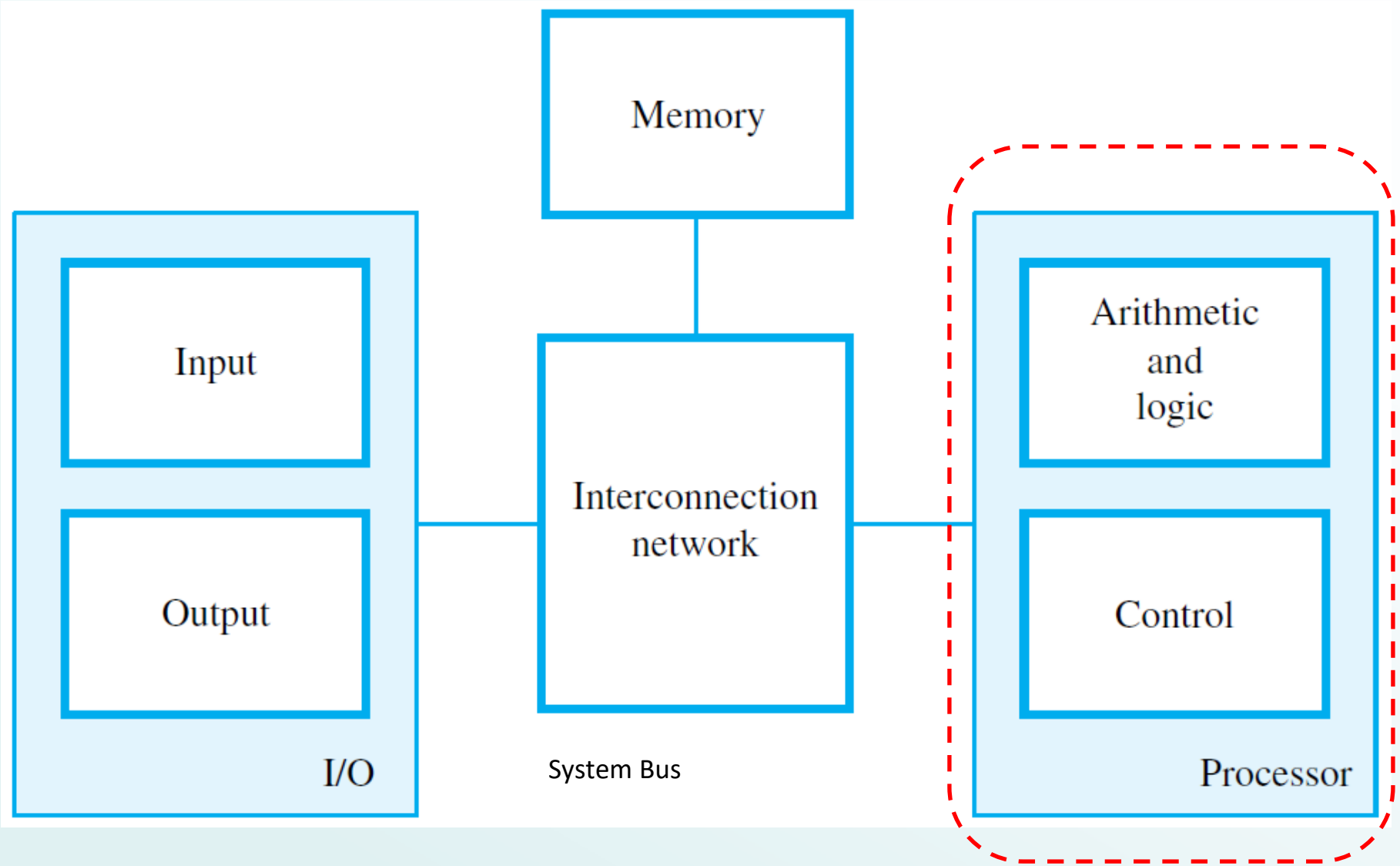
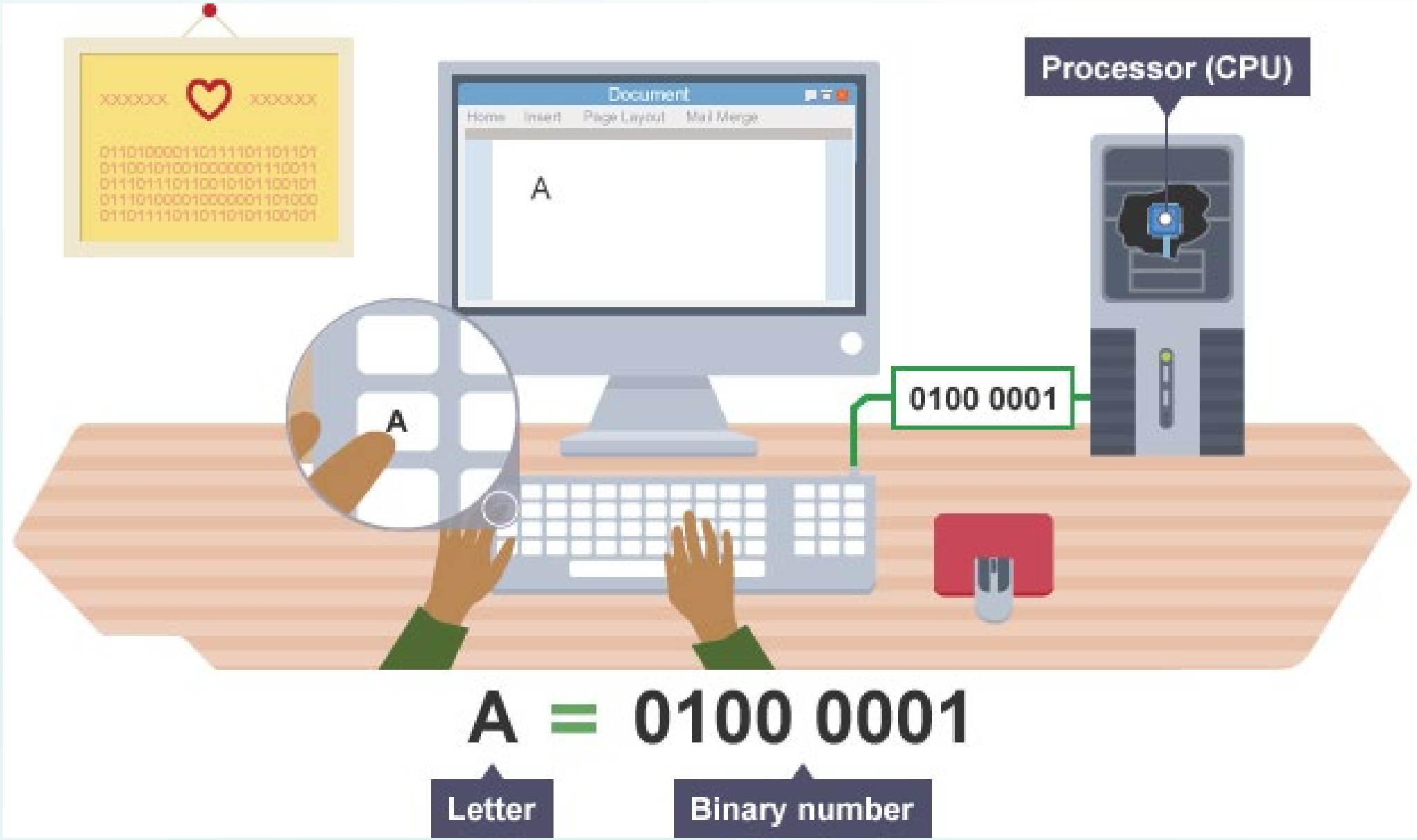
Topic 01B
Fetch, Execution and IO

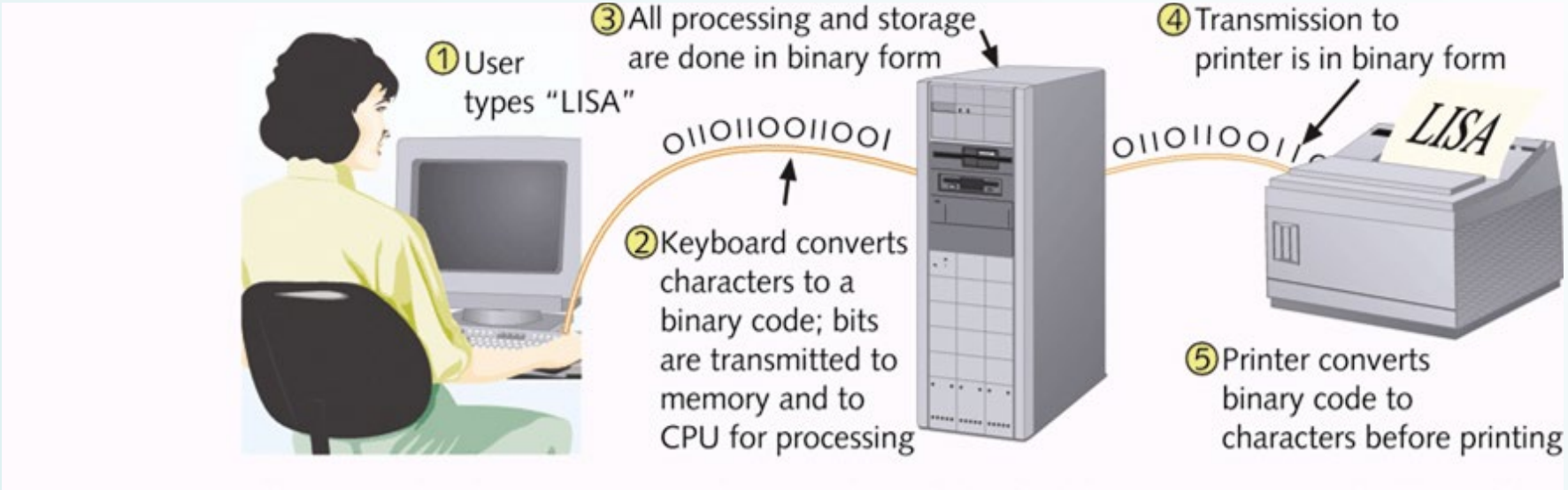


The “heart” of the computer is the CPU or Central Processing Unit, also called the Processor

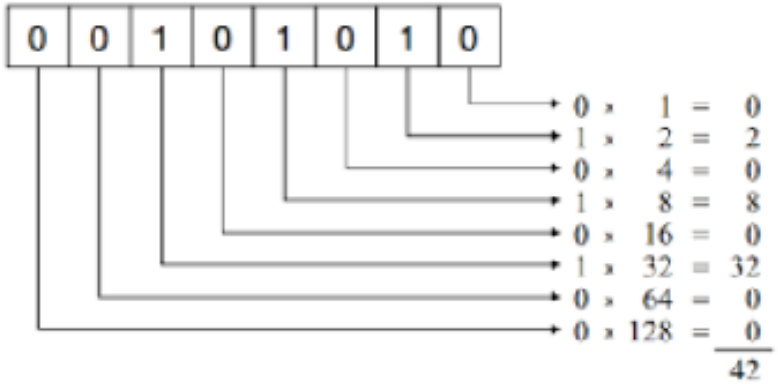

Processor (CPU)
Document
Home
Insert
Page Layout
Mail Merge
A
0100 0001
A
A = 0100 0001
Letter
Binary number

Data Is Stored in Bits

Data on a computer is stored as binary digits (“bits” for short)
A bit holds a 1 or 0 value
A pulse of 5 volts of electricity can represent a 1 bit and a pulse of 0 volts (the absence of voltage) can represent a 0 bit
With fiber-optic cable, a 1 bit is represented by the presence of light and a 0 bit by the absence of light
A “byte” is a collection of 8 bits


1 User types "LISA"
2 Keyboard converts characters to a binary code; bits are transmitted to memory and to CPU for processing
3 All processing and storage are done in binary form
4 Transmission to printer is in binary form
5 Printer converts binary code to characters before printing
LISA

Bits and Bytes



Binary	Decimal	Character	Binary	Decimal	Character
01000001	65	A	01100001	97	a
01000010	66	B	01100010	98	b
01000011	67	C	01100011	99	c
01000100	68	D	01100100	100	d
01000101	69	E	01100101	101	e
01000110	70	F	01100110	102	f
01000111	71	G	01100111	103	g
01001000	72	H	01101000	104	h
01001001	73	I	01101001	105	i
01001010	74	J	01101010	106	j

ASCII – American Standard Code for Information Interchange

Functional units

Primary memory (also called Main memory)

1) Organized into words of typically 32 bits
2) A 32-bit word contains four 8-bit bytes
3) A personal computer memory might have 4 Gigabytes or more

   (4 Gbyte = $2^2 * 2^{30}$ bytes)

4) Programs and their data must be in this memory to be executed

Functional units

Cache memory

1) An adjunct to the main memory, fabricated on the processor chip
2) Much smaller and faster than the main memory
3) Holds sections of the program and data currently/frequently being executed

Functional units

Processor

1) Logic circuits for performing arithmetic and logic operations on word-size data operands

2) Timing and control circuits for fetching program instructions and data from memory, one after another

3) Registers (typically 16 or 32), each of which hold one word of operand data

Processor - Arithmetic and Logic Unit

Most computer operations are executed in the ALU of the processor

Performs arithmetic or logic operation

Processor - Control Unit

Memory, ALU and I/O units store and process information and perform input and output operations

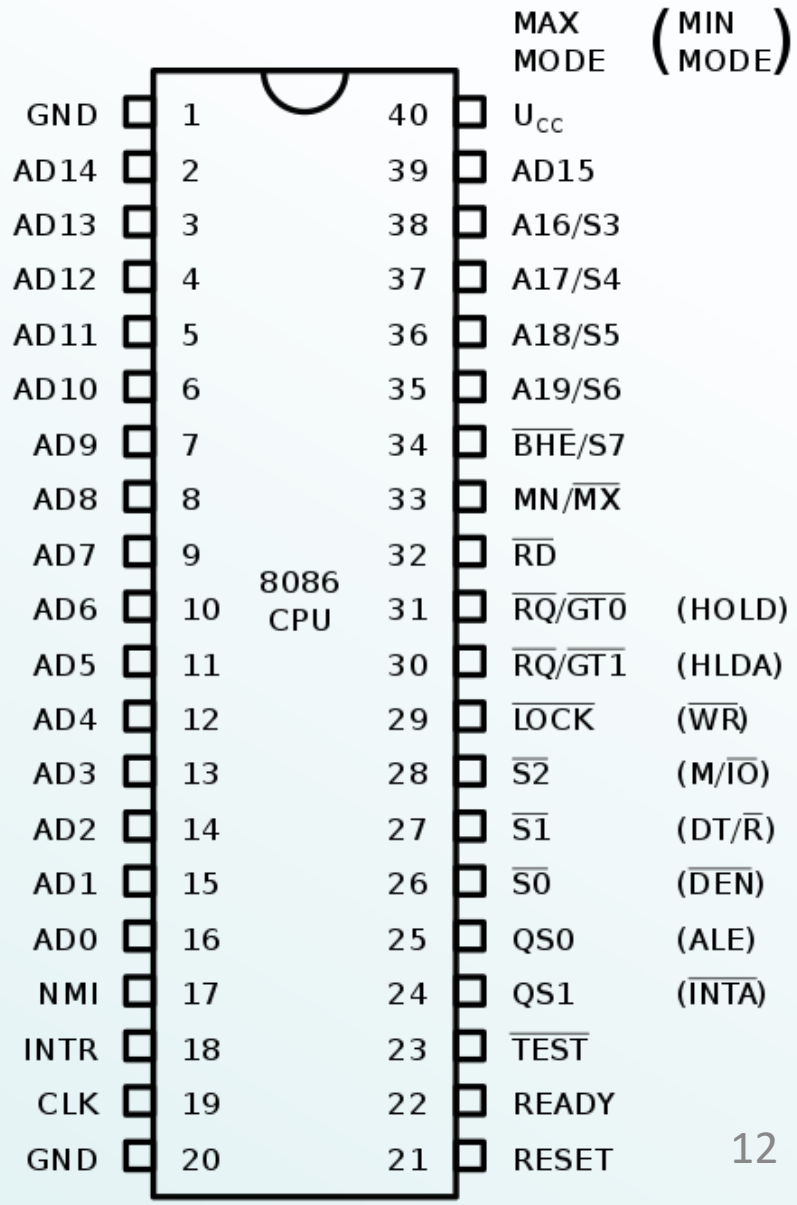
The operation of these units must be coordinated (this is the responsibility of the control unit)

Case Study:Intel CPU 8086

All internal registers, as well as internal and external data buses, are 16 bits wide, which firmly established the "16-bit microprocessor" identity of the 8086.

A 20-bit external address bus provides a 1 MB physical address space (2^{20} = 1,048,576).

This address space is addressed by means of internal memory "segmentation".



Computer

Operation of a computer can be summarized as:

1) Computer accepts information in the form of programs and data through an input unit and stores it in the memory
2) Information stored in the memory is fetched under program control into an ALU, where it is processed
3) Processed information leaves the computer through an output unit
4) All activities in the computer are directed by the Control Unit

Instruction Cycle

Although a computer system is able to run very complex programs; its own basic operation is very simple.

It derives its power by being able to repeat the basic operations billions of times per second.

A single cycle of operation, also called the instruction cycle or machine cycle, consists of Fetch, Decode, Execute, Memory Access and Write Back phases

Instruction cycle operations

Step	Action
1	Fetch an instruction and increment the program counter.
2	Decode the instruction and read registers from the register file.
3	Perform an ALU operation.
4	Read or write memory data if the instruction involves a memory operand.
5	Write the result into the destination register, if needed.

Instructions and Programs

1) An instruction specifies an operation and the locations of its data operands
2) A 32-bit word typically holds one encoded instruction
3) A sequence of instructions, executed one after another, constitutes a program
4) Both a program and its data are stored in the main memory

Instruction types

Three basic instruction types:

1) Load - Read a data operand from memory or an input device into the processor

1) Store - Write a data operand from a processor register to memory or an output device

1) Operate - Perform an arithmetic or logic operation on data operands in processor registers

Example program

A, B, and C, are labels representing memory word addresses;
Ri are processor registers

A program for the calculation

C = A + B is:

```
Load        R2, A
Load        R3, B
Add         R4, R2, R3
Store       R4, C
```

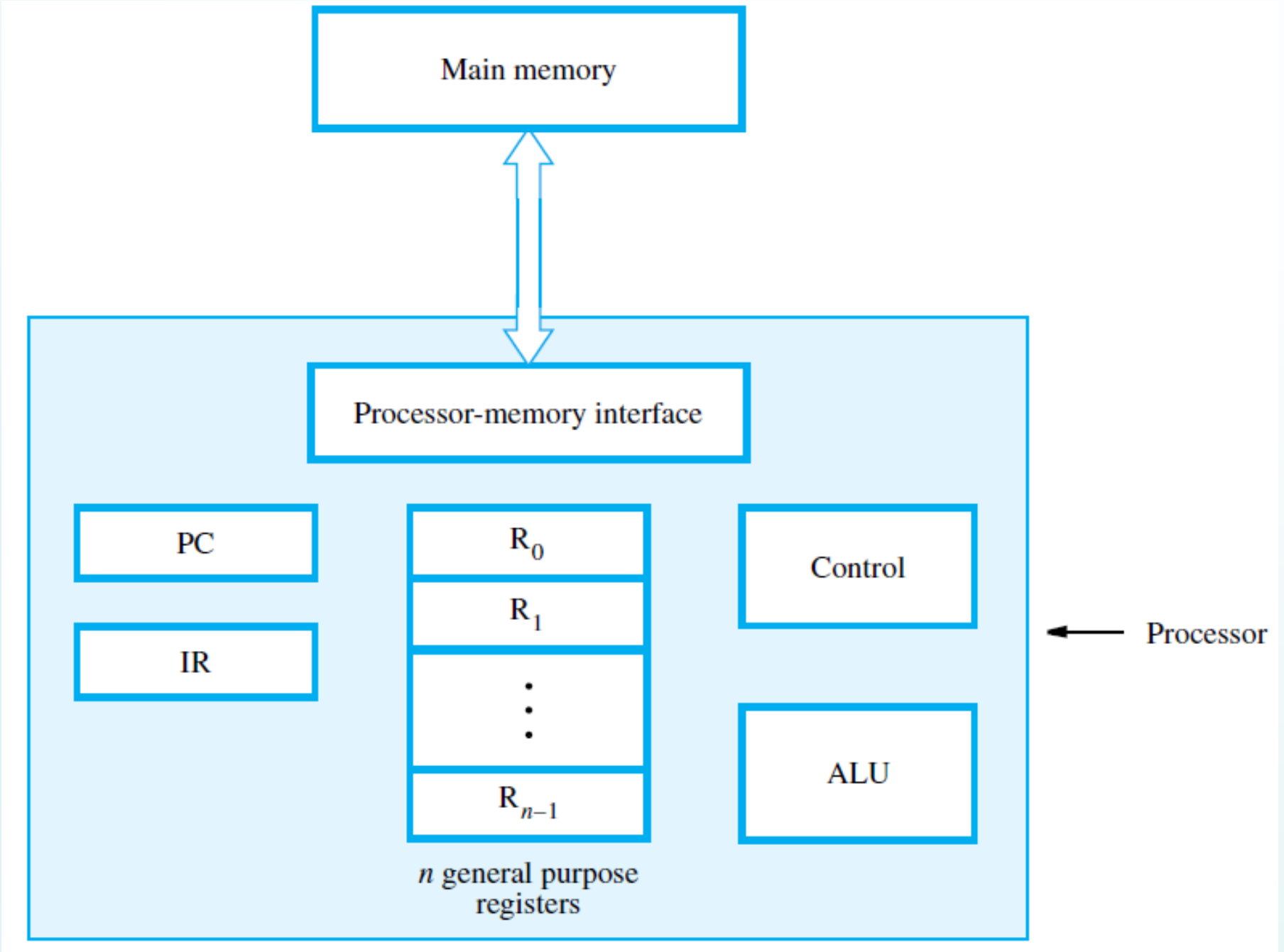
Processor components

1) The program counter (PC) register holds the memory address of the current instruction

1) The instruction register (IR) holds the current instruction

1) General-purpose registers hold data and addresses

1) Control circuits and the arithmetic and logic unit (ALU) fetch and execute instructions


Main memory
Processor-memory interface
PC
IR
R_0
R_1
R_{n-1}
n general purpose registers
Control
ALU
Processor

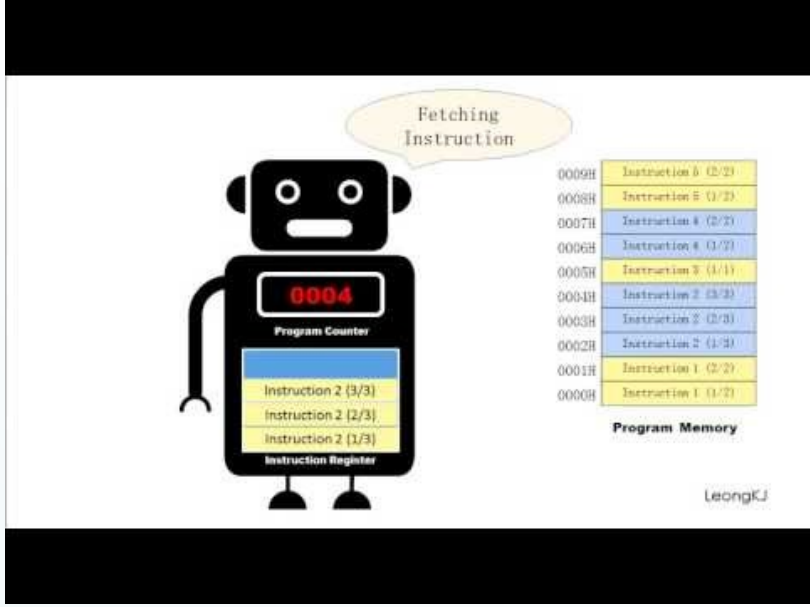
Fetching and executing instructions

Example: Load R2, LOC

The processor control circuits do the following:

1) Send address in PC to memory; issue Read
2) Load instruction from memory into IR
3) Increment PC to point to next instruction
4) Send address LOC to memory; issue Read
5) Load word from memory into register R2

PC: Program Counter



A program counter is a register in a computer processor that contains the address (location) of the instruction being executed at the current time.

As each instruction gets fetched, the program counter increases its stored value by 1.
After each instruction is fetched, the program counter points to the next instruction in the sequence.

When the computer restarts or is reset, the program counter normally reverts to 0.

Handling I/O devices

An application program can:

1) Read data (such as a keyboard character) from an input device

1) Write data (such as letter character) to an output display screen

1) Sense the readiness of an input or output (I/O) device to perform a transfer

Performance

How quickly can a program be executed ?

Some factors:

1) Speed of electronic circuits in the processor
2) Access times to the cache and main memory
3) Design of the instruction set
4) Number of operations that can be done at the same time (parallelism)

Technology

Improvement in the materials used of transistors means instructions can be executed faster

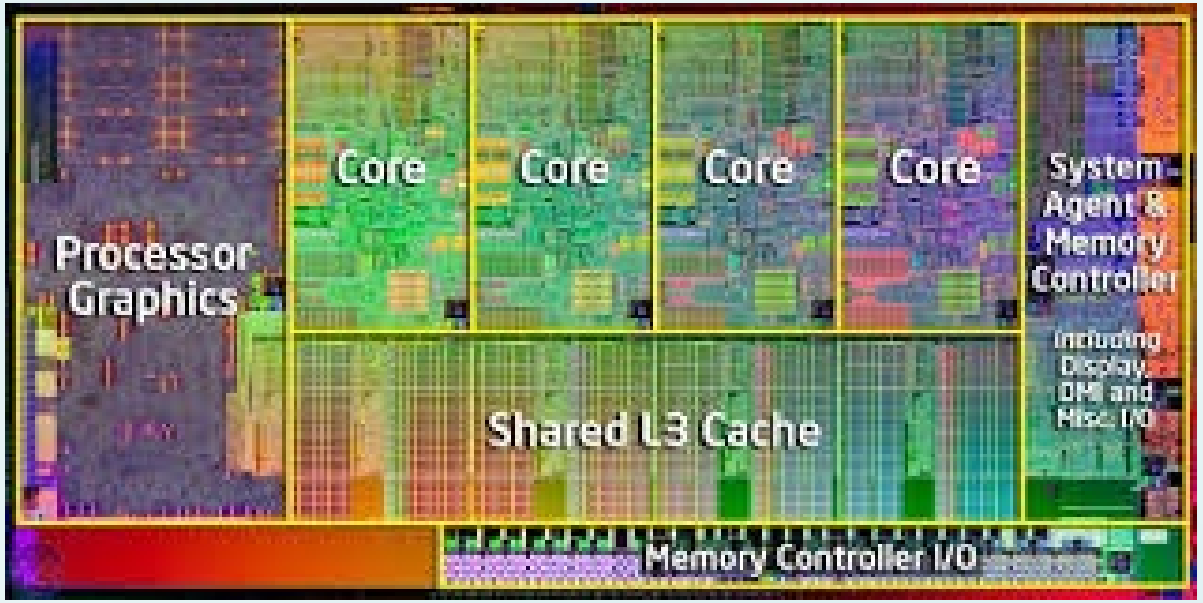
Improvements in process technology means more transistors can be placed on a chip (see next slide) leading to

- More logic functionality
- More memory storage capacity

Performance - Parallelism

Multicore processors (across multiple cores)

- Multiple processing units can be fabricated on a single chip.
- core is used for each of these processors
- the term processor is then used for the complete chip
- *dual-core*, *quad-core* and *octo-core* processors for chips



You want a fast CPU for the new computer you are building and your motherboard only has one CPU socket. To increase the processing power, you need a CPU that provides more than one processor. What is this CPU called?

A. multicore
B. multiple
C. multifaceted
D. muliticast

Summary

1) Basic structure of computers - I/O, Memory, Processor with interconnection network.
2) Instruction cycle operations - Fetch and Execute
3) Performance - technology, parallelism