CS1004: COMPUTER ORGANIZATION AND ARCHITECTURE

CSE, SNU Chennai

Course Objective

- To make students understand the basic structure and operation of digital computer
- To understand the hardware-software interface
- To familiarize the students with arithmetic and logic unit and implementation of fixed point and floating-point arithmetic operations
- To expose the students to the concept of pipelining
- To familiarize the students with hierarchical memory system including cache memories and virtual memory
- To expose the students with different ways of communicating with I/O devices and standard I/O interfaces

Course Outcome

SHIV NADAR UNIVERSITY CHENNAI

- CO1. Explain processor fundamentals
- CO2. Design arithmetic and logic unit
- CO3. Evaluate performance of memory systems
- CO4. Extend the learning to parallel processing architectures
- CO5. Explain interconnection structures

Textbooks / References

- ***TEXTBOOK:***
- Carl Hamacher, Zvonko Vranesic, Safwat Zaky, “Computer Organization”, McGraw Hill Education, Fifth Edition, 2011

Computer

It is an electronic machine that takes input, process the input, and gives back the result in human understandable format.

Organization vs. Architecture

Computer Architecture

- Layout/framework of machine which holds hardware components on which software is built
- Deals with functional behavior of computer system
- Design implementation for the various parts of the computer

Computer Organization

- Integrating all hardware components in a single unit
- Deals with structural relationship
- Operational attributes are linked together and contribute to realize the architectural specification

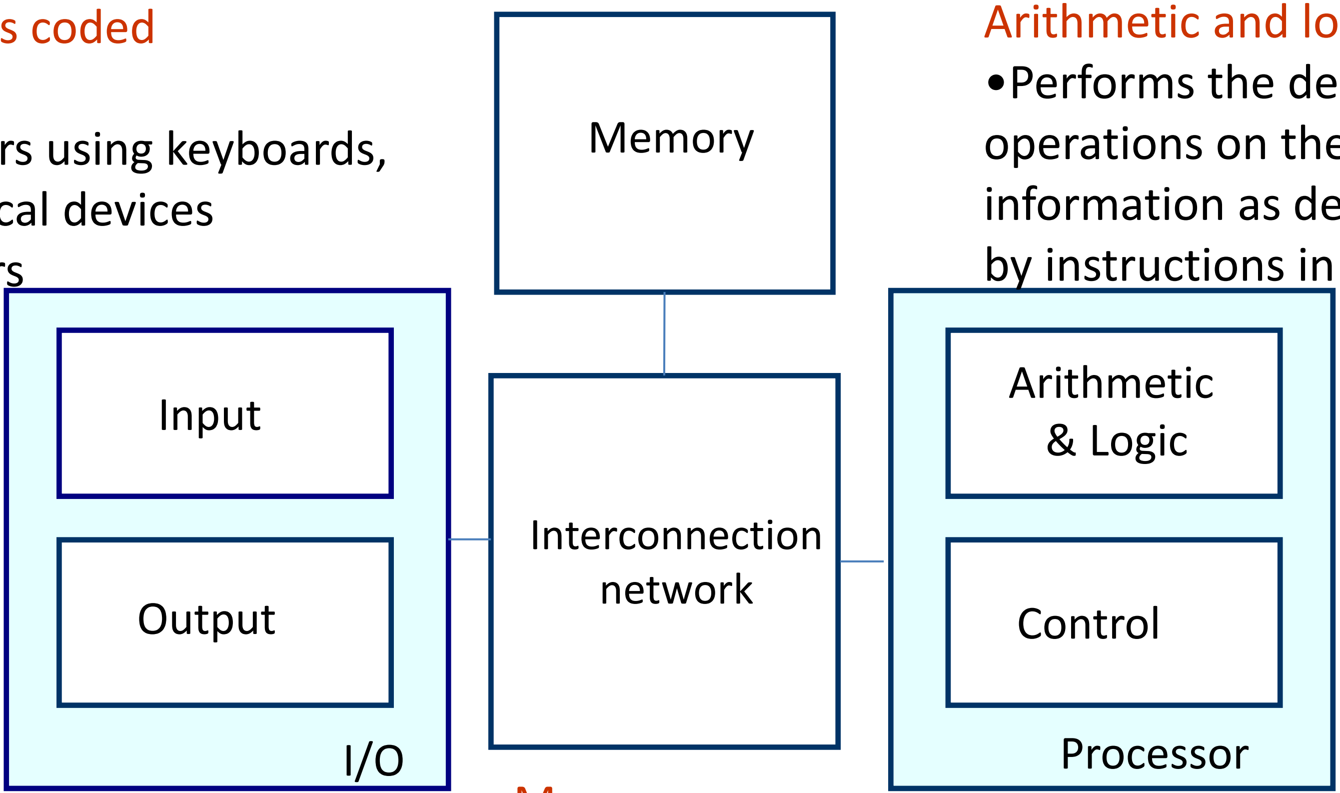
Computer

- Electronic device perform a given task
  - Input Devices (Keyboard)
  - Output Devices (Monitor)
  - Memory Unit (Primary, Secondary)
  - Processor
    - ALU (Arithmetic and Logical Unit)
    - CU (control Unit)

Functional units of a computer

Input unit accepts coded information:
- Human operators using keyboards,
- Electromechanical devices
- Other computers

Arithmetic and logic unit(ALU):
- Performs the desired operations on the input information as determined by instructions in the memory



Interconnection Network:
- means for the functional units to
- (a) Exchange information
- (b) Coordinate their actions

Output unit sends results of processing:
- To a monitor display,
- To a printer

Memory
- Stores information:
- Instructions,
- Data

Control unit coordinates various actions
- Input,
- Output
- Processing

Information in a computer

- Data are the “operands” upon which instructions operate.
- Data could be:
  - Numbers
  - Encoded characters
- Data, in a broad sense means any digital information.
- Computers use data that is encoded as a string of binary digits called bits.

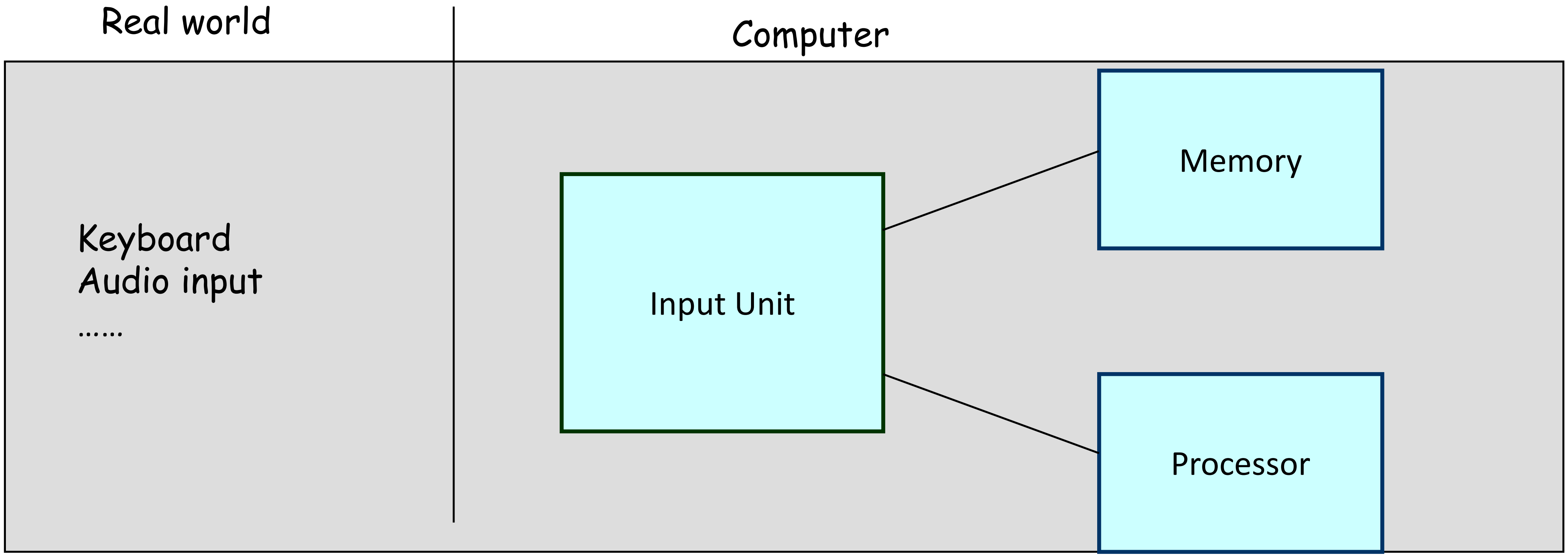
- Instructions are explicit commands that:
- (a) Govern the transfer of information within a a computer as well as b/w the computer and its I/O devices.
- (b) Specify the arithmetic and logic operations to be performed.

- Program: List of instructions which performs a task.

Input unit

Binary information must be presented to a computer in a specific format. This task is performed by the input unit:

- Interfaces with input devices.
- Accepts binary information from the input devices.
- Presents this binary information in a format expected by the computer.
- Transfers this information to the memory or processor.



Memory unit

- Memory unit stores instructions and data.
  - Recall, data is represented as a series of bits.
  - To store data, memory unit thus stores bits.
- Processor reads instructions and reads/writes data from/to the memory during the execution of a program.
  - In theory, instructions and data could be fetched one bit at a time.
  - In practice, a group of bits is fetched at a time.
  - Group of bits stored or retrieved at a time is termed as "word"
  - Number of bits in a word is termed as the "word length" of a computer.
  - Typical word length ranges from 16 to 64 bits.
- In order to read/write to and from memory, a processor should know where to look:
  - "Address" is associated with each word location.

Arithmetic and logic unit (ALU)

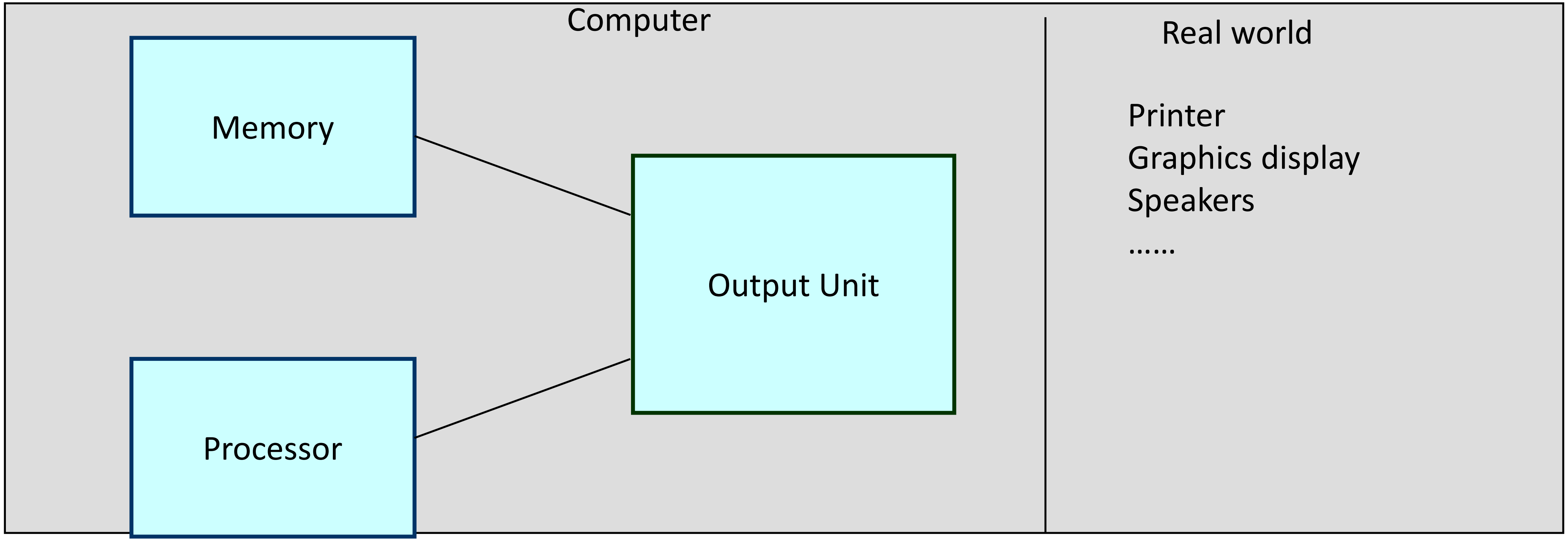
SHIV NADAR UNIVERSITY CHENNAI

- Operations are executed in the Arithmetic and Logic Unit (ALU).
  - Arithmetic operations such as addition, subtraction.
  - Logic operations such as comparison of numbers.
- In order to execute an instruction, operands need to be brought into the ALU from the memory.
  - Operands are stored in general purpose registers available in the ALU.
  - Access times of general purpose registers are faster than the cache.
- Results of the operations are stored back in the memory or retained in the processor for immediate use.

SHIV NADAR UNIVERSITY CHENNAI

Output unit

- Computers represent information in a specific binary form. Output units:
  - Interface with output devices.
  - Accept processed results provided by the computer in specific binary form.
  - Convert the information in binary form to a form understood by an output device.



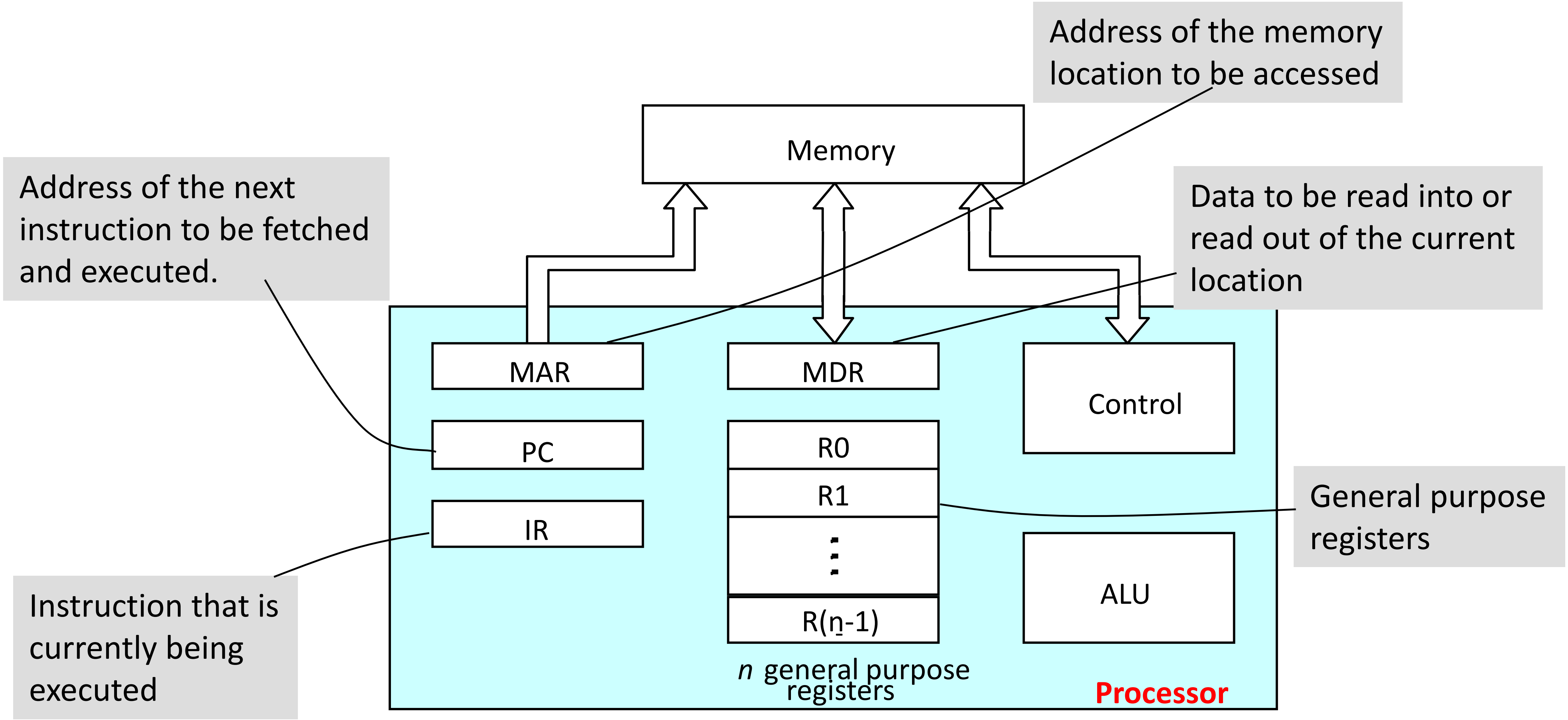
Control unit

Operations of Input unit, Memory, ALU and Output unit are coordinated by Control unit.

- Nerve center that sends control signals to other units and senses their states.
- Instructions control “what” operations take place (e.g. data transfer, processing).
- Control unit generates timing signals which determines “when” a particular operation takes place.

- Operation of a computer can be summarized as:
  - Accepts information from the input units (Input unit).
  - Stores the information (Memory).
  - Processes the information (ALU).
  - Provides processed results through the output units (Output unit).

Basic processor architecture



Registers in the control path

- Instruction Register (IR):
  - Instruction that is currently being executed.
- Program Counter (PC):
  - Address of the next instruction to be fetched and executed.
- Memory Address Register (MAR):
  - Address of the memory location to be accessed.
- Memory Data Register (MDR):
  - Data to be read into or read out of the current memory location, whose address is in the Memory Address Register (MAR).

Interrupt signal: Request from an I/O device for service by processor.

- ❑Normal execution of programs may be preempted if some device requires urgent servicing.
- ❑For example: monitoring device detects dangerous condition.
- ❑Normal execution of the current program is interrupted using **interrupt** signal.

Execution of an instruction

- Recall the steps involved in the execution of an instruction by a processor:

  – Fetch an instruction from the memory.
  – Fetch the operands.
  – Execute the instruction.
  – Store the results.

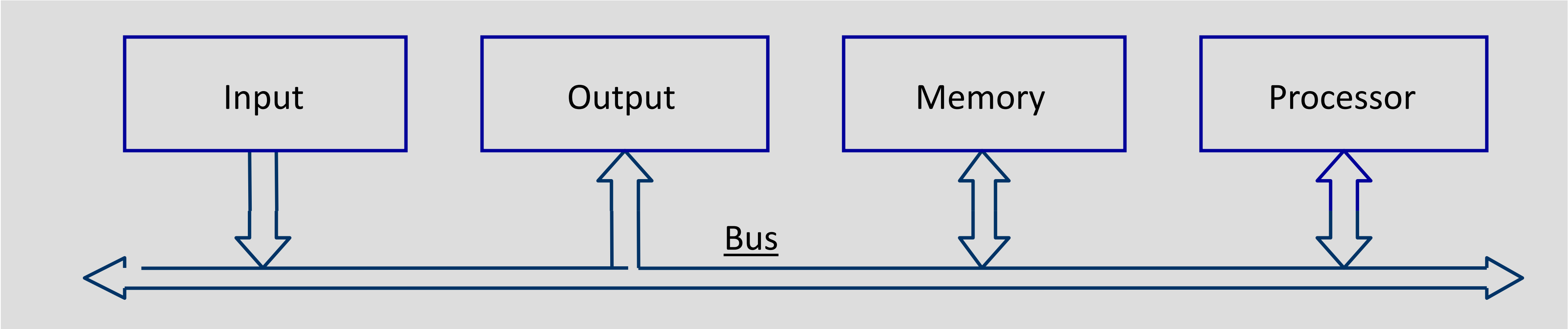
- Basic processor architecture has several registers to assist in the execution of the instructions.

Fetch/Execute cycle

- Execution of an instruction takes place in two phases:
  - Instruction fetch.
  - Instruction execute.
- Instruction fetch:
  - Fetch the instruction from the memory location whose address is in the Program Counter (PC).
  - Place the instruction in the Instruction Register (IR).
- Instruction execute:
  - Instruction in the IR is examined (decoded) to determine which operation is to be performed.
  - Fetch the operands from the memory or registers.
  - Execute the operation.
  - Store the results in the destination location.
- Basic fetch/execute cycle repeats indefinitely.

How are the functional units connected?

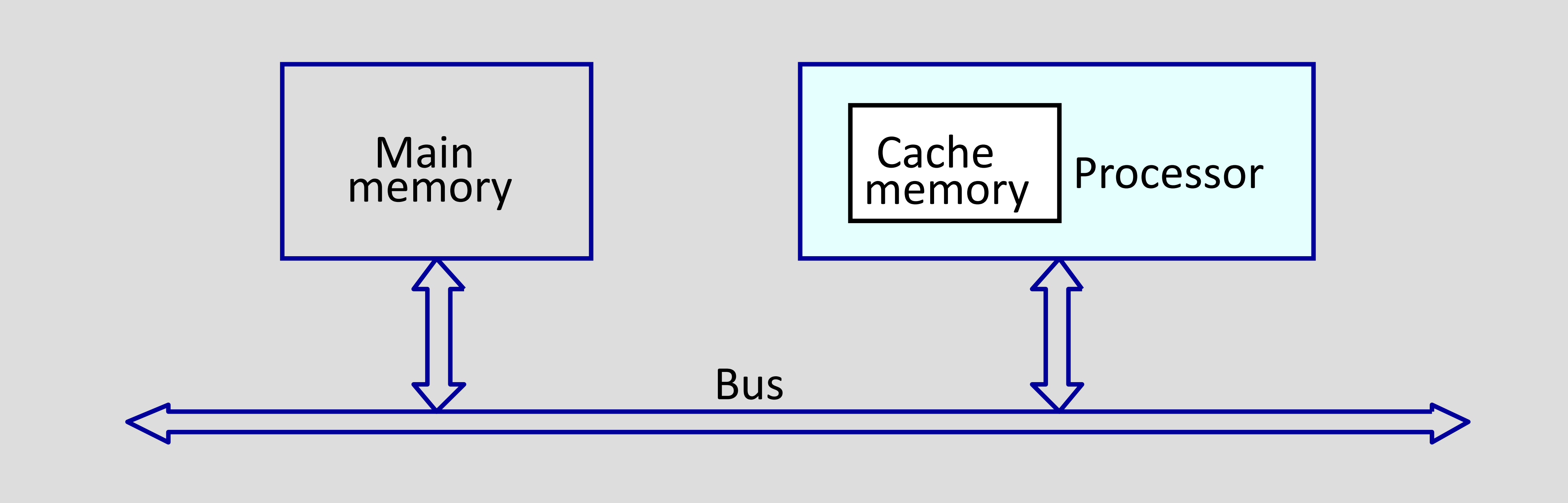
- For a computer to achieve its operation, the functional units need to communicate with each other.
- In order to communicate, they need to be connected.



- Functional units may be connected by a group of parallel wires.
- The group of parallel wires is called a bus.
- Each wire in a bus can transfer one bit of information.
- The number of parallel wires in a bus is equal to the word length of a computer.
- Devices connected to bus vary in speed of operation
- Keyboards and printers are slow ; optical, magnetic disks are faster- So prefer using BUFFER REGISTERS!!

Organization of cache and main memory

SHIV NADAR UNIVERSITY CHENNAI



- All program instructions and data are stored in main memory.
- As execution proceeds, instructions are fetched over the bus into the processor, and copy is placed in cache.

Contd.

- When execution of an instruction calls for data located in the main memory, the data is fetched and copy is placed in the cache.
- Later, if the same instruction or data is needed second time, it is read directly from cache.

- The processor and a small cache memory can be fabricated in a single integrated circuit chip !!

- If repeatedly used instructions are placed in the cache, they can be fetched quickly.

- Program will be executed faster

Performance

- Measure of performance of a computer is how quickly it can execute programs.

- Speed with which a computer executed programs is affected by :
  - ✓ Design of its hardware
  - ✓ Machine language instructions
  - ✓ Compiler that translates high-level language into machine language

- For best performance, coordination is required b/w:
  - ✓ Compiler design
  - ✓ Machine instruction set
  - ✓ Compiler design

Parameters affecting the processor time

1. Processor clock

- Processor circuits are controlled by a timing signal called clock.
- Clock defines regular time intervals, called clock cycles.
- To execute a machine instruction, processor divides the action to be performed in steps.
- Each step completed in one clock cycle.

P= length of one clock cycle (Affects processor performance)

Clock rate, R= 1/P (cycles per second or hertz)

- Today's processors have few hundred million to billion cycles per second.

SHIV NADAR UNIVERSITY CHENNAI

2. Basic performance equation

.

- Compiler generates machine language program corresponding to source program.
- Complete execution requires N machine language instructions.

N- actual number of instruction executions

- Average number of basic steps to execute one machine instructions is, S in one clock cycle.
- If clock rate is R cycles per second,

Processor time required to execute a program that is in high-level language is

$$T=(N \times S)/R$$

- To achieve high performance, reduce T by reducing N and S and increasing R.

SHIV NADAR UNIVERSITY CHENNAI

3. Pipelining and superscalar operation

- Performance can be improved by overlapping the execution of successive instructions using pipelining

- Consider the instruction

  Add R1,R2,R3

  Adds the contents of R1 and R2 and result is placed in R3

- The processor can read the next instruction while addition being performed.

- Higher degree of concurrency can be achieved if multiple instruction pipelines are implemented.

- Superscalar execution:

Multiple functional units are used creating parallel paths for different instructions.

4. Clock rate

Possibilities for increasing the clock rate, R

1. Improving the integrated-circuit (IC) technology

- to make logic circuits faster by reducing the time.
- allows clock period, P to reduce and clock rate, R is increased.

2. Reducing the amount of processing done in one basic step reduces clock period, P.

5. Instruction Set: CISC and RISC

SHIV NADAR UNIVERSITY CHENNAI

- Simple instructions require a small number of basic steps to execute.
  - Here, large number of instructions are needed
  - N increases and S decrease.

- Complex instructions involve large number of steps.
  - Fewer instructions are needed.
  - Lower value of N and S increases.

- Design of the instruction set of a processor and the options include:

  - Reduced Instruction Set Computers (RISC) - Processors with simple processors.

  - Complex Instruction Set Computers (CISC) - Processors with complex instructions.

6. Complier

- Compiler translates high-level language program into a sequence of machine instructions.
- To reduce N, suitable machine instruction set and compiler
- *Optimized compiler* reduces the product N X S,

  which is total number of clock cycles needed to execute a program.

- Compiler may rearrange program instructions to achieve better performance.
- High-quality compiler must be closely linked to the processor architecture.
- Objective is to reduce the total number of clock cycles needed to perform a required programming task

7. Performance measurement

- System Performance Evaluation Corporation (SPEC) rating is,

  SPEC rating= Running time on the reference computer/Running time on the computer under test

- The test is repeated for all the programs in the SPEC suite and geometric mean is computed.

- The overall SPEC rating=

$$\text{SPEC rating} = \frac{\text{Running time on the reference computer}}{\text{Running time on the computer under test}}$$

$$= (\pi \, SPEC_i)$$

8. Multiprocessors and multi-computers

- Large computer systems may contain a number of processor units, they are *multiprocessor systems.*

- These systems either execute a number of different application tasks in parallel, or they execute subtasks of single large tasks in parallel.

- An interconnected group of complete computers to achieve high computational power- *multi-computers*