Chapter 2

Computer Organization and Architecture

The architecture of a computer defines the manner in which the hardware components of the system work to meet functional objectives, and enabling these hardware components to work in a harmonized manner is called computer organization.

INTRODUCTION

Just as a tall building has different levels of detail, including the number of storeys, the size of rooms, door and window placement, etc., each computer has a visible structure, which is referred to as its *architecture*. when we talk about architecture in terms of a computer, it is defined as the science of selecting and interconnecting hardware components to create computers that meet functional, performance and cost goals.

Extending the concept of the architecture and making these hardware components work in a harmonized manner to achieve a common objective in an environment is known as *computer organization*. Computer organization is like a group of people who work together to achieve a common goal, for example, a football team where different members of the team play specific roles in the game. The role of the goalkeeper is to safe guard the goal. Similarly, other members of the team (striker, mid-fielder and defender) have their respective roles to perform. However, the ultimate objective of the team as a whole is to defeat the opponent by scoring more goals than the opponent. Hence, the study of computer organization focuses more on the collective contribution from the hardware peripherals than individual electronic components.

CENTRAL PROCESSING UNIT

The central processing unit (CPU) is referred to as the "**brain**" of a computer system and it converts data (input) into meaningful information (output). It is a highly complex, extensive set of electronic circuitry, which executes stored program instructions. It controls all internal and external devices, performs arithmetic and logic operations, and operates only on binary data, that is, data composed of 1s and 0s. In addition, it also controls the usage of the main memory to store data and instructions, and controls the sequence of operations (Figure 2.1).

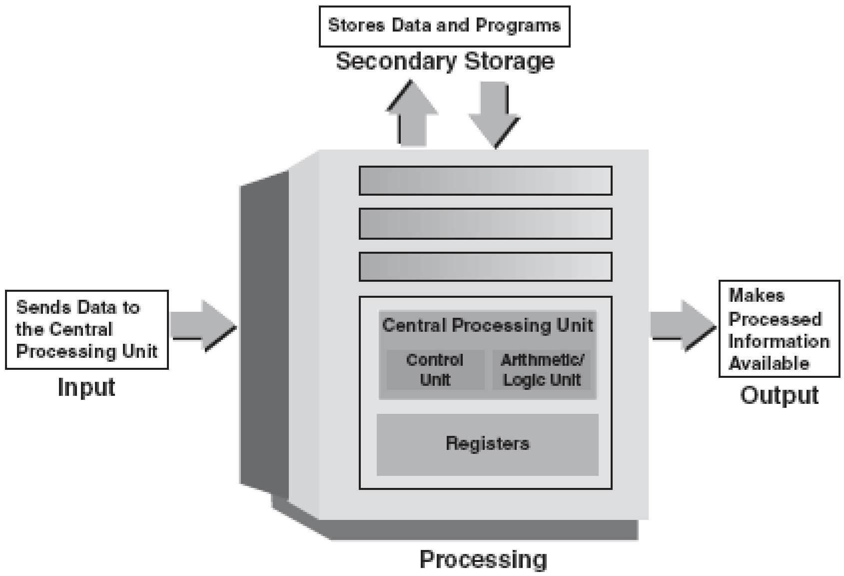

Figure 2.1 The Central Processing Unit

The CPU consists of three main subsystems, the *Arithmetic/Logic Unit (ALU)*, the *Control Unit (CU)* and the *registers*. The three subsystems work together to provide operational capabilities to a computer.

Arithmetic/Logic Unit

The arithmetic/logic unit (ALU) contains the electronic circuitry that executes all arithmetic and logical operations on the data made available to it. The data required to perform the arithmetic and logical functions are inputs from the designated registers. The ALU comprises two units: Arithmetic Unit (AU) and Logic Unit (LU).

Arithmetic Unit: The AU contains the circuitry that is responsible for performing the actual computing and carrying out arithmetic calculations, such as addition, subtraction, multiplication, and division. It can perform these operations at a very high speed.

Logic Unit: The LU enables the CPU to make logical operations based on the instructions provided to it. These operations are logical comparisons between data items. The LU can compare numbers, letters or special characters and then take action based on the result of the comparison. Logical operations of the LU test for three conditions:

- **Equal-to Condition**: In a test for this condition, the ALU compares two values to determine if they are equal. For example, if the number of tickets sold equals the number of seats in a cinema hall, then housefull is declared.
- **Less-than Condition**: To test this condition, the ALU compares values to determine if one is less than the other; for example, if a person purchases less than two tickets, then no discount rate is given.

- **Greater-than Condition**: In this type of comparison, the computer determines if one value is greater than the other; for example, if a person purchases greater than two tickets, then the discount rate is 5 per cent.

Registers

The registers are special-purpose, high-speed temporary memory units. These are temporary storage areas for holding various types of information, such as data, instructions, addresses and intermediate results of calculations. Essentially, they hold the information that the CPU is currently working on. The registers can be thought of as the CPU's working memory, a special additional storage location that offers the advantage of speed. The registers work under the direction of the CU to accept, hold, and transfer instructions or data, and perform arithmetic or logical comparisons at high speed. The CU uses a data-storage register in the similar way a store owner uses a cash register as a temporary, convenient place to store the transactions. As soon as a particular instruction or piece of data is processed, the next instruction immediately replaces it, and the information that results from the processing is returned to the main memory. Figure 2.2 reveals the types of registers present inside the CPU.

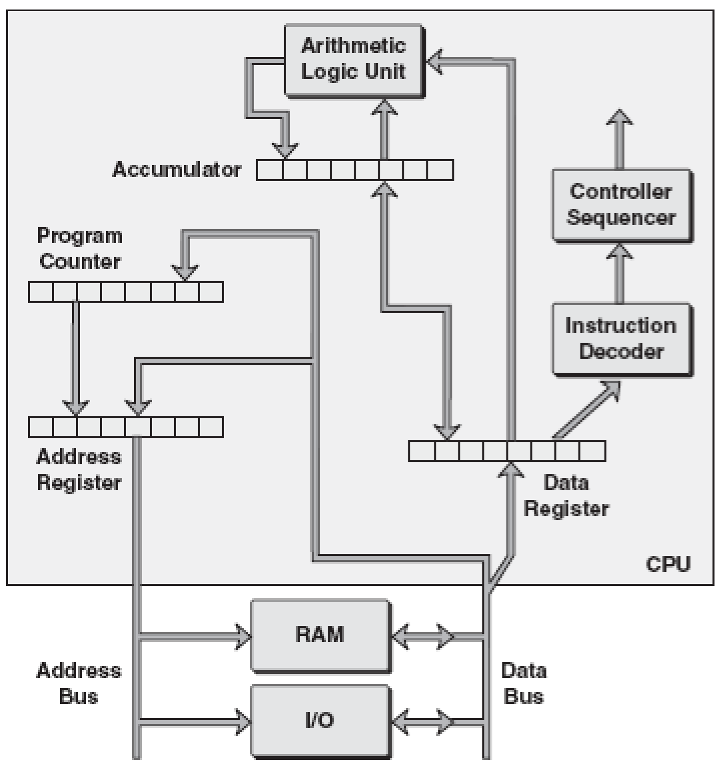

Figure 2.2 Register in the CPU

Instruction addresses are normally stored in consecutive registers and are executed sequentially. The CU reads an instruction from the memory by a specific address in the register and executes it. The next instruction is then fetched from the sequence and executed, and so on. This type of instruction sequencing is possible only if there is a counter to calculate the address of the instruction that has been executed. This counter is one of the registers, which stores intermediate data used during the execution of the instructions after they are read from the memory. Table 2.1 lists some of the important registers used in the CPU.

Table 2.1 Registers and Functions

Register Name	**Function**
Program Counter (PC)	The PC keeps track of the next instruction to be executed
Instruction Register (IR)	The IR holds the instruction to be decoded by the CU
Memory Address Register (MAR)	The MAR holds the address of the next location in the memory to be accessed
Memory Buffer Register (MBR)	The MBR is used for storing data either coming to the CPU or data being transferred by the CPU
Accumulator (ACC)	The ACC is a general-purpose register used for storing temporary results and results produced by the ALU
Data Register (DR)	The DR is used for storing the operands and other data

Control Unit

The control unit (CU) of the CPU contains circuitry that uses electrical signals to direct the entire computer system to carry out, or execute, stored program instructions. This resembles an orchestra leader who himself does not play a musical instrument but directs other people to play the musical instrument in a harmonized manner. The CU also does not execute program instructions; rather, it directs other parts of the system to do so by communicating with both the ALU and the memory.

The CU controls the I/O devices and transfer of data to and from the primary storage. The CU itself is controlled by individual instructions in programs located in the primary storage. Instructions are retrieved from the primary storage, one at a time. For this, the CU uses the IR for holding the current instruction and an instruction pointer to hold the address of the next instruction. Each instruction is interpreted (decoded) so that it can be executed. Based on the instructions, the CU controls how other parts of the CPU and, in turn, the rest of the computer system should work so that the instructions are executed in a correct manner. An analogy can be considered between the CU and the traffic police; the control unit decides which action will occur just as the traffic police takes decisions on which lanes the traffic will move or stop.

Figure 2.3 illustrates how the CU instructs other parts of the CPU (ALU and registers) and the I/O devices on what to do and when to do. It also determines what data are needed, where they are stored and where to store the results of the operation, as well as sends the control signals to the devices involved in the execution of the instructions. It administers the movement of a large amount of instructions and data used by the computer. In order to maintain the proper sequence of events required for any processing task, the CU uses clock inputs.

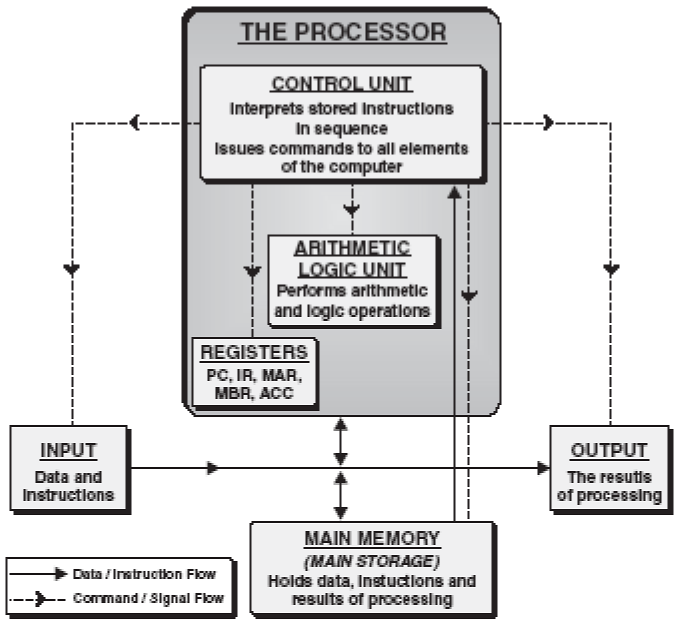

Figure 2.3 Control Unit

System Bus

A *bus* is a set of connections between two or more components/devices, which are designed to transfer several/all bits of a word from the source to the destination. It is a shared medium of information transfer. A bus consists of multiple paths, which are also termed *lines*, and each line is capable of transferring one bit at a time. Thus, to transmit 8 bits simultaneously over a bus, eight lines are required. In addition, some other lines are needed to control this transfer. A bus can be unidirectional (transmission of data can be in only one direction) or bi-directional (transmission of data can be in both directions). In a shared bus, only one source can transmit data at one time, while one or more than one can receive that signal. A bus that connects to all three components (CPU, memory, I/O devices) is called a *system bus* (see Figure 2.4). A system bus consists of 50–100 separate lines. These lines are broadly categorized into three functional groups.

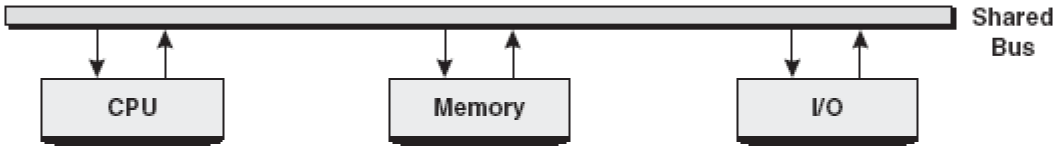

Figure 2.4 System Bus

Data Lines: Data lines provide a path for moving data between the system modules. Data lines are collectively known as a *data bus*. Normally, a data bus consists of 8, 16 or 32 separate lines. The number of lines present in a data bus is called the *width* of the data bus. Data bus width limits the maximum number of bits, which can be transferred simultaneously between two modules. The width of a data bus helps in determining the overall performance of a computer system.

Address Lines: Address lines are used to designate the source of data for a data bus. As the memory may be divided into a linear array of bytes or words, for reading or writing any information on to the memory, the CPU needs to specify the address of a particular location. This address is supplied by the address bus (address lines are collectively called an *address bus*). Thus, the width of the address bus specifies the maximum possible memory supported by a system; for example, if a system has a 16-bit-wide address bus, it can support memory size of equal to 2^{16} (or 65,536) bytes.

Control Lines: Control lines are used to control the access to data and the address bus; this is required as a bus is a shared medium. The control lines are collectively called a *control bus* (see Figure 2.5). These lines are used for transmission of commands and timing signals (which validate data and addresses) between the system modules. Timing signals indicate whether data and address information is valid, whereas command signals specify which operations are to be performed. Some of the control lines of a bus are required to provide clock signals to synchronize operations and reset signals to initialize the modules. Control lines are also required for reading/writing to I/O devices or the memory. Control lines if used as a bus request indicate that a module needs to gain control of the bus. The bus grant control line is used to indicate whether the requesting module has been granted control of the bus.

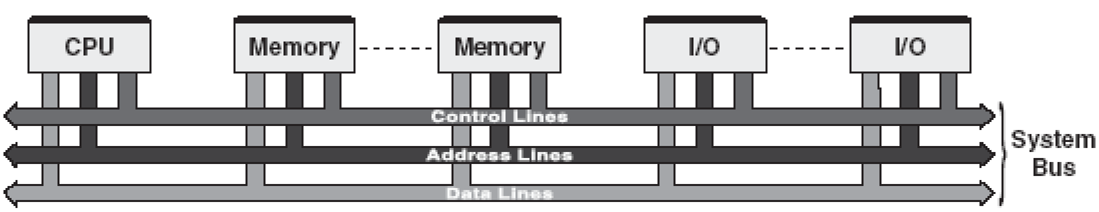

Figure 2.5 Bus Interconnection

Main Memory Unit

The memory is that part of the computer that holds data and instructions for processing. Logically, it is an integral component of the CPU, but, physically, it is a separate part placed on the computer's motherboard. The memory stores program instructions or data for only as long as the program they pertain to is in operation. The CPU accesses the main memory in a random manner, that is, the CPU can access any location of this memory to either read information from it or store information in it. The primary memory is of two types: the first is called the *Random Access Memory (RAM)* and the second is the *Read Only Memory (ROM)*.

The RAM directly provides the required information to the processor. It can be defined as a block of sequential memory locations, each of which has a unique address determining the location and those locations contain a data element. Storage locations in the main memory are addressed directly by the CPU's instructions. It is volatile in nature, which means the information stored in it remains as long as the power is switched on. As soon as the power is switched off, the information contained in the RAM is lost.

The ROM stores the initial start-up instructions and routines in the Basic Input/Output System (BIOS), which can only be read by the CPU, each time it is switched on. The contents of the ROM are not lost even in the case of a sudden power failure, thus making it non-volatile in nature. The instructions in the ROM are built into the electronic circuits of the chip, called *firmware*. The ROM is also randomly accessible in nature, which means the CPU can randomly access any location within the ROM. Improvement in technology for constructing a flexible ROM comes in various types, namely Programmable Read Only Memory (PROM), Erasable Programmable Read Only Memory (EPROM) and Electrically Erasable Programmable Read Only Memory (EEPROM).

Cache Memory

The cache is a very high speed, expensive piece of memory, which is used to speed up the memory-retrieval process. Due to its higher cost, the CPU comes with a relatively small amount of cache compared with the main memory. Without the cache memory, every request from the CPU to access data would be sent to the main memory, which in response sends the data back to the CPU through the system bus. This is a slow process in computing terms. The idea of introducing cache is that this extremely fast memory would store data that are frequently accessed and, if possible, the data that are around it. This is to achieve the quickest possible response time to the CPU.

Let us consider an example of a library system for a better understanding of the cache concept. When a person comes into the library and asks for a book, the librarian searches for the requested book in the bookshelves, retrieves the book and delivers it to the person. After the person reads the book, it is returned to the library. At the same time, if any other person comes in requesting the same book, which is still placed on the librarian's desk, the librarian does not have to go through the process of searching the book in the bookshelves. This results in saving of precious time in retrieving the requested book and thus improving the efficiency of the work.

The above analogy can be related to the computer's cache system. The computer uses logic to determine which data are the most-frequently accessed and keeps them in the cache. A cache is a piece of very fast memory made from a high-speed static RAM that reduces the access time of the data. It is very expensive and generally incorporated in the processor, where valuable data and program segments are kept. The cache memory can be categorized into three levels: *L1 cache*, *L2 cache* and *L3 cache* (see Figure 2.6).

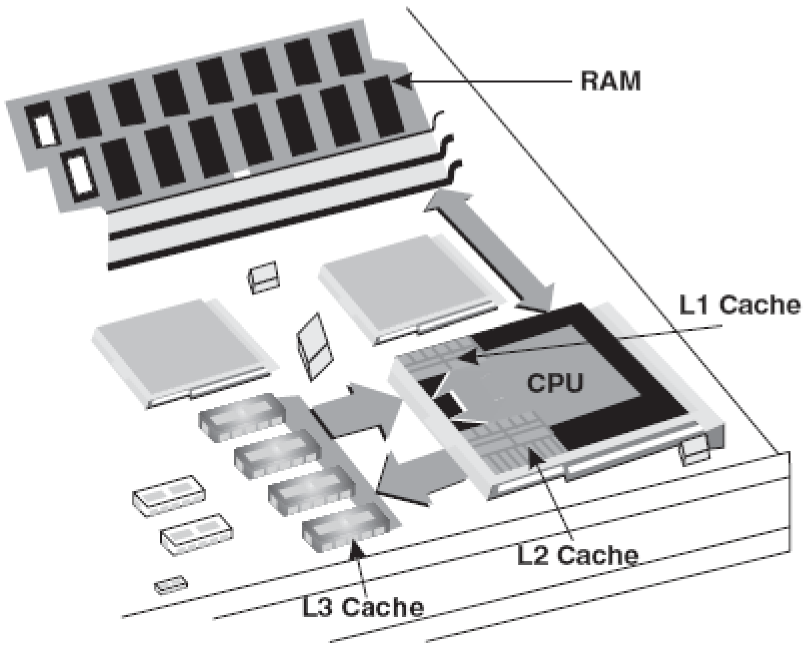

Figure 2.6 Various Levels of Cache

L1 Cache: This cache is closest to the processor and hence is termed as the primary or L1 cache. Each time the processor requests information from the memory, the cache controller on the chip uses special circuitry to first check if the requested data are already in the cache. If they are present, then the system is spared from time-consuming access to the main memory. In a typical CPU, the primary cache ranges in size from 8 to 64 KB, with larger amounts on the newer processors. This type of cache memory is very fast because it runs at the speed of the processor since it is integrated into it. There are two different ways that the processor can organize its primary cache: first, some processors have a single cache to handle both command instructions and program data (called a *unified cache*), while others have a separate data and instruction cache called *split cache*. However, the overall performance difference between integrated and separate primary cache is not significant.

L2 Cache: The L2 cache is larger but slower in speed than L1 cache. It is used to see recent accesses that is not picked by the L1 cache and is usually 64 to 2 MB in size. The L2 cache is also found on the CPU. If the L1 and L2 cache are used together, then the missing information that is not present in the L1 cache can be retrieved quickly from the L2 cache.

L3 Cache: The L3 cache memory is an enhanced form of memory present on the motherboard of the computer. It is an extra cache built into the motherboard between the processor and main memory to speed up the processing operations. It reduces the time gap between the request and retrieval of the data and instructions, thereby accessing data much more quickly than the main memory. The L3 cache is being used with processors nowadays having more than 3 MB of storage in it.

COMMUNICATION AMONG VARIOUS UNITS

All units in a computer system work in conjunction with each other to formulate a functional computer system. To have proper coordination among these units (processor, memory and I/O devices), a reliable and robust means of communication is required. Let us discuss one of the most important functions in the computer system, that is, the communication between these units.

Processor to Memory Communication

The whole process of communication between the processor and memory can be divided into two steps, namely, information transfer from the memory to processor and writing information in the memory. The following sequence of events takes place when information is transferred from the memory to the processor (see Figure 2.7 also):

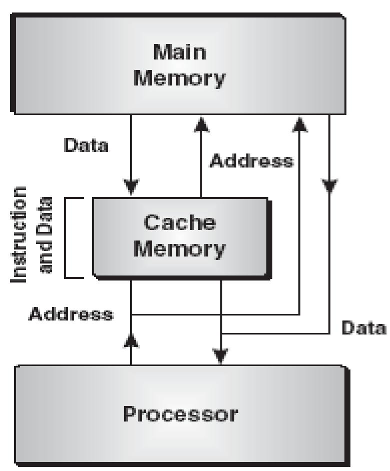

Figure 2.7 Processor-Memory

The processor places the address in the MAR through the address bus.

1. The processor issues a READ command through the control bus.
2. The memory places required data on the data bus, which are then transferred to the processor.

Based on the read time of the memory, a specific number of processor clock intervals are allotted for completion of this operation. During this interval, the processor is forced to wait. Similarly, the following sequence of events takes place when information is written in the memory:

1. The processor places the address in the MAR through the address bus.
2. The processor places the data to be written in the memory on the data bus.
3. The processor issues a WRITE command to the memory by the control bus.
4. The data are written in the memory at the address specified in the MAR.

The main concern in the processor-memory communication is the speed mismatch between the memory and processor. The memory speed is slower than the CPU's speed. Hence, the CPU is forced to wait for the data. As already discussed, this speed mismatch is reduced by using the small fast memory (called *cache*) as an intermediate buffer between the processor and memory.

2.3.2 Processor—I/O Devices Communication

The I/O units are connected to the computer system through the system bus. Each I/O device in a computer system is first met with the controller, called the *Direct Memory Access* (*DMA*) controller, which controls the operation of that device. The controller is connected to the buses to perform a sequence of data transfers on behalf of the CPU (see Figure 2.8). It is capable of taking over control of the system bus from the CPU, which is required to transfer data to and from the memory over the system bus. The DMA controller can directly access the memory and is used to transfer data from one memory location to another or from an I/O device to the memory and vice versa. The DMA controller can use the system bus only when the CPU does not require it or it should suspend the operations currently being processed by the CPU.

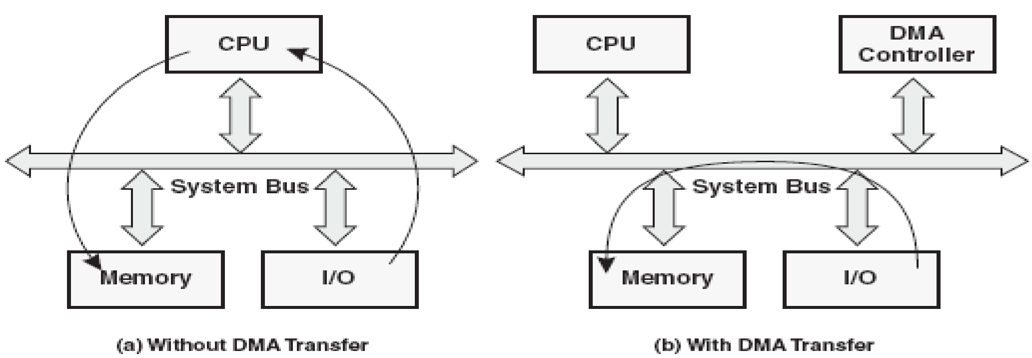

Figure 2.8 Processor—I/O Communication

With the DMA controller, a dedicated data-transfer device reads the incoming data from a device and stores that data in a system memory buffer for later retrieval by the CPU. The DMA controller allows peripheral devices to access the memory for both read and write operations without affecting the state of the computer's central processor. As a result, the data-transfer rate is significantly increased, thus improving the system efficiency. When a large amount of data is to be transferred from the CPU, the DMA controller can be used. The DMA controller allows the I/O unit to exchange data directly with the memory without going through the CPU except at the beginning (to issue the command) and at the end (to clean up after the command is processed). While the I/O operation is being performed by the DMA controller, the CPU can start execution of some other part of the same program or can start executing some other program. Thus, the DMA controller increases the speed of I/O operations by taking over buses and thus eliminating the CPU's intervention.

INSIDE A COMPUTER

Computing machines are complex devices made from numerous electronic components. Many of these components are small, sensitive, and expensive and operate with other components to provide better performance to the computing machines. Therefore, to ensure better performance and increase in the life of these components, they are placed inside a metal enclosure called the ***system case* or *cabinet*** (see Figure 2.14). The system case is a metal and plastic box that houses the main components of the computer. It protects electronics hardware against the heat, light, temperature, etc. It serves important roles in the functioning of a properly designed and well-built computer. Several areas where the system case plays an important role are:

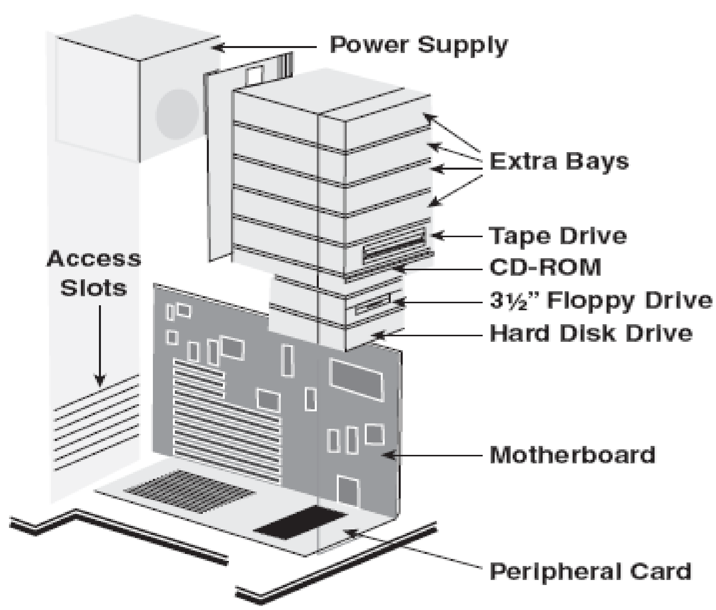

Figure 2.14 System Case

- **Structure**: The system case provides a rigid structural framework to the components, which ensures that everything fits together and works in a well-organized manner.
- **Protection**: The system case protects the inside of the system from physical damage and electrical interference.

- **Cooling:** The case provides a cooling system to the vital components. Components that run under cool temperature last longer and are less troublesome.
- **Organization and Expandability**: The system case is a key to the organization of physical systems. If a system case is poorly designed, upgradation or expansion of a peripheral is limited.
- **Status Display**: The system case contains lights or LEDs that provide varied information pertaining to the working of the system to the user.

The system case encloses all the components, which are essential in running the computer system. These components include the motherboard, processors, memory, power supply, expansion slots, cables, removable drives, etc.

Power Supply (SMPS)

Power supply or *Switched Mode Power Supply* (*SMPS*) is a transformer and voltage control device in a computer that furnishes power to all the electronic components by converting incoming AC into the low-voltage DC supply. When a computer is turned on, the power supply allows the converted electricity to travel to other components inside the computer. The modern-day power supply provides protection against surge and spikes in the power, which could damage vital components of the computer. Nowadays, the computer power supply is capable of providing different voltages, at different strengths, and manages additional signals for the motherboard. The power supply plays an important role in the following areas of the computer system (see Figure 2.15):

Figure 2.15 Power Supply

- **Stability**: A high-quality power supply with sufficient capacity to meet the demands of the computer provides years of stable power for the computer.

- **Cooling**: The power supply contains the main fan that controls the flow of air through the system case. This fan is a major component in the computer cooling system.
- **Expandability**: The capacity of the power supply determines the ability to add new drives to the system or upgrade to a more powerful motherboard or processor.

Motherboard

The *motherboard*, also known as the *system board*, is a large multilayered printed circuit board inside a computer. The motherboard contains the CPU, the BIOS ROM chip and the CMOS Setup information. It has expansion slots for installing different adapter cards like the video card, sound card, network interface card and modem. The circuit board provides a connector for the keyboard as well as housing to the keyboard controller chip. It possesses RAM slots for the system's RAM chips, and enables the system's chipset, controllers and underlying circuitry (bus system) to tie everything together. In a typical motherboard, the circuitry is imprinted on the surface of a firm planar surface and is usually manufactured in a single piece. The most common design of the motherboard in today's desktop computers is the ATX design. In ATX designs, the computer components included are processor, co-processors (optionally), memory, BIOS, expansion slot and interconnecting circuitry. Additional components can be added to a motherboard through its expansion slot. Nowadays, they are designed to put peripherals as integrated chips directly onto the motherboard. Initially, this was confined to audio and video chips, but in recent times, the peripherals integrated in this way include SCSI, local area network (LAN) and RAID controllers. Though, there is cost benefit to this approach, the biggest downside is the restriction of future upgrade options. Figure 2.16 provides an insight into various components on motherboards.

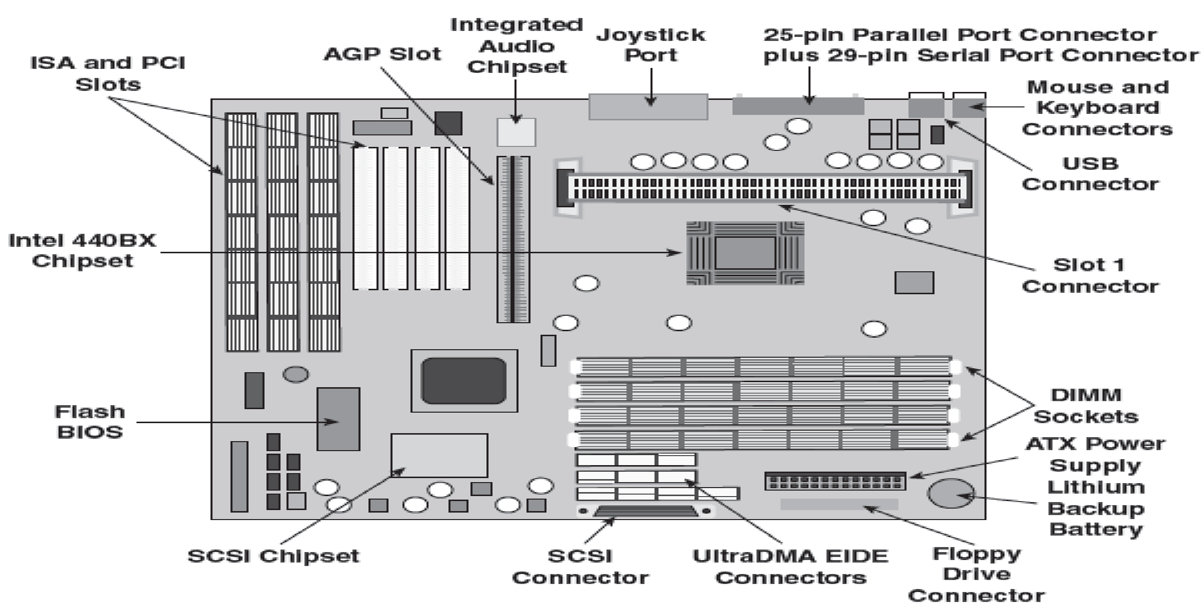

Figure 2.16 Motherboard

BIOS: The BIOS comprises a set of several routines and start-up instructions inside the ROM. This gives two advantages to the computer. First, the code and data in the ROM BIOS need not be reloaded each time the computer is started. Secondly, they cannot be corrupted by wayward applications that are accidentally written into the wrong part of the memory. As the machine is switched on, the control is transferred to the bootstrap procedure of the BIOS. This procedure inspects the computer to determine what hardware is fitted and then conducts a simple test (power-on self-test) for normal functionality. If all the tests are passed, the ROM then determines the drive to boot the machine. Most computers have the BIOS set to check for the presence of an operating system in the primary hard disk drive. Once the machine is booted, the BIOS serve a different purpose by presenting the DOS with a standardized Application Program Interface (API) for the computer hardware.

CMOS: The motherboard includes a separate block of memory called the *complementary metal oxide silicon* (*CMOS*) chip that consumes very low power. This chip is kept alive by a battery even when the computer's power is off. The function of the CMOS chip is to store basic information about the computer's configuration, such as the type of hard disks and floppy drives, memory capacity, etc. The other important data, which is kept in the CMOS memory, is the system time and date. The clock, CMOS chip and batteries are usually all integrated into a single chip.

Ports and Interfaces

Ports and interfaces are a generic name for the various “holes” (and their associated electronics) found at the back of the computer, using which external devices are connected to the computer's motherboard. Different interfaces and ports run at varying speeds and work best with specific types of devices (see Figure 2.17).

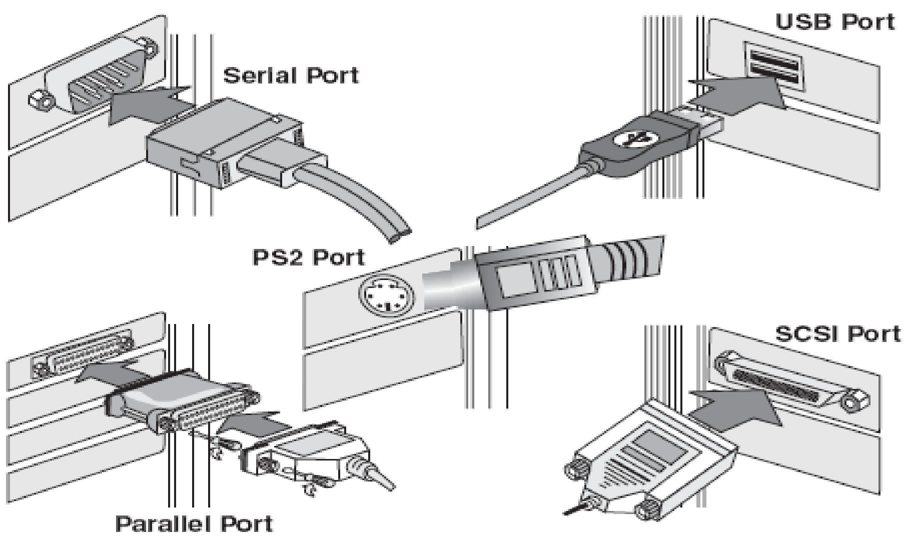

Figure 2.17 Different Types of Ports

- **PS/2 Ports**: It is a standard serial port connector used to plug the computer mouse and keyboards into the personal computer. It consists of six pins in a small, round-shape socket.
- **Serial Ports**: It is a general-purpose communications port through which data are passed serially, that is, one bit at a time. These ports are used for transmitting data over long distances. In the past, most digital cameras were connected to a computer's serial port in order to transfer images to the computer. However, because of its slow speed, these ports are used with the computer mouse and modem.
- **Parallel Port**: It is an interface on a computer, which supports transmission of multiple bits of data (usually 8 bits) at the same time. This port transmits data faster than a serial port and is exclusively used for connecting peripherals such as printers and CD-ROM drives.
- **SCSI Port:** These ports are used in transmitting data up to seven devices in the "daisy chain" fashion and at a speed faster than serial and parallel ports (usually 32 bits at a time). In the daisy chain fashion, several devices are connected in series to each other, so that data for the seventh device need to go through the entire six devices first. These ports are hardware interfaces, which include an expansion board that plugs into the computer called the *SCSI host adapter* or *SCSI* controller. Devices which can be connected to SCSI ports are hard-disk drives and network adapters.
- **Universal Serial Bus (USB) Port**: The USB port is a plug-and-play hardware interface for connecting peripherals such as the keyboard, mouse, joystick, scanner, printer and modem. It supports a maximum bandwidth of 12 MB per second and has the capability to connect up to 127 devices. With the USB port, a new device can be added to the computer without adding an adapter card. These ports are the replacement for the parallel and serial ports.

Expansion Cards

An *expansion card*, also called an *adapter card*, is a circuit board that provides additional capabilities to the computer system. Adapter cards are made up of large-scale integrated circuit components Sound Card installed on it. The cards are plugged into the expansion sockets present on the computer's motherboard to provide the computer an added functionality. Common available expansion cards connect monitors (for enhanced graphics) and microphones (for sound), each having a special purpose to perform (see Figure 2.18). However, nowadays most of the adapters come in-built on the motherboard and no expansion card is required, unless the need for high performance is required.

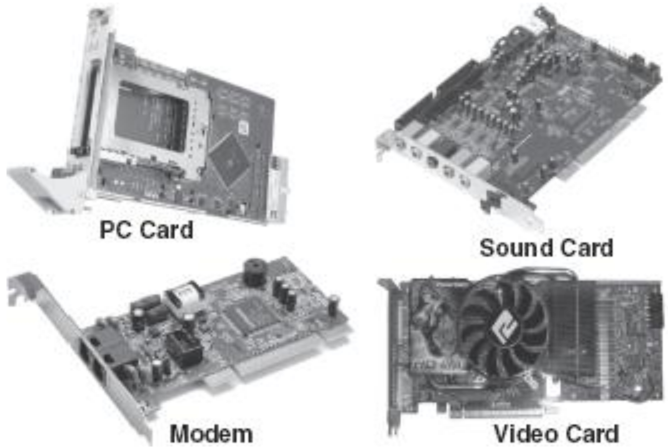

Figure 2.18 Expansion Card

- **Sound Cards**: An expansion card that allows the computer to output sound through connected speakers, to record sounds from a microphone and to manipulate sounds stored on the computer is called a sound card. It contains special circuits for operating the computer's sound and allows playback and recording of sound from the CD-ROM.
- **Video Cards**: A video card, also called the display adapter, is used for enhancing graphics images that are seen on the computer's monitor. The card converts the images created in the computer to electronic signals required by the monitor. Generally, a good card with a graphics accelerator is preferred for editing digital videos. There are different video cards with varying capabilities related to the size of the monitor and total number of displayable colours.
- **Network Interface Card**: A network interface card is a computer circuit board that is installed in a computer so that it can be connected to other computers in a network. Personal computers and workstations on a LAN contain a network interface card specifically designed for transmitting data across LANs. Network interface cards provide a dedicated, full-time connection to a network.
- **Modem**: Modem is an expansion card that allows two computers to communicate over ordinary phone lines. It converts digital data from computers into analog data, transmits over the telephone lines and also converts incoming analog signals back to digital signals for the receiving computer. Modems do not provide high bandwidth for the data communication and as a result, they do not support high-speed Internet access as current modems can run up to 56 KBps.
- **PC Card**: A PC card is a removable device, approximately the size of a credit card, which is designed to plug into a Personal Computer Memory Card International Association (PCMCIA) slot. It is a standard formulated by the PCMCIA for providing expansion capabilities to computers. The PCMCIA standard supports input-output

devices, memory, fax/modem, SCSI and networking products. The card fits into a notebook or laptop computer.

Ribbon Cables

Ribbon cables are wide, flat and insulated cables, which are flexible enough to fit into areas with little space. These cables are made up of numerous tiny wires (traces and electronic pathways) called the bunch, where one *bunch* carries data/information around to different components on the motherboard and another bunch connects these components to the various devices attached to the computer. These cables connect the hard drive, floppy drive and CD-ROM drive to the connectors on the motherboard and control the drives by getting and sending data from and to them. These cables connect different external devices, peripherals, expansion slots, I/O ports and drive connections to the rest of the computer (see Figure 2.19)

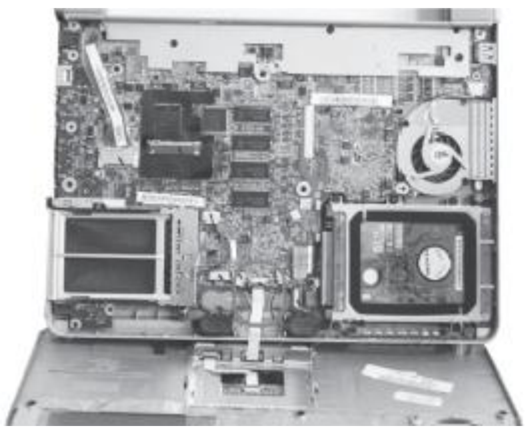

Figure 2.19 Ribbon Cables

Memory Chips

The memory is the place where the computer holds programs and data that are currently in use. The system memory 4.25” 72-pins SIMM on the motherboard is arranged in groups called the *memory banks*. The number of the memory banks and their configurations vary from computer to computer because these are determined by the CPU and the 168-pins DIMM way it receives information. The speed of Fi220 SIMM and DIMM RAM Chips the CPU determines the number of memory sockets required in a bank. For the main memory, either of the two types of memory chips is used: *Single In-Line Memory Modules* (*SIMM*) or *Dual In-Line Memory Modules* (*DIMM*) (see Figure 2.20).

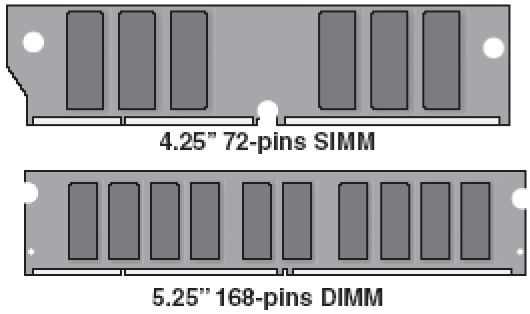

Figure 2.20 SIMM and DIMM RAM Chips

SIMM: SIMM are small circuit boards designed to accommodate surface-mount memory chips. A typical SIMM chip comprises a number of RAM chips on a printed circuit board (PCB), which fits into a SIMM socket on a computer's motherboard. These chips are packed into small plastic or ceramic dual inline packages (DIPs), which are assembled into a memory module. A typical motherboard offers four SIMM sockets capable of taking either single-sided or double-sided SIMMs with module sizes of 4, 8, 16, 32 or even 64 MB. When 32-bit SIMM chips are used with processors, they have to be installed in pairs, with each pair of modules making up a memory bank. These chips support 32-bit data paths and are originally used with 32-bit CPUs. The CPU then communicates with the memory bank as one logical unit. SIMM chips usually come in the following two formats:

- A 30-pin SIMM used in older system boards, which deliver 1 byte of data.
- A larger 72-pin SIMM used in modern PCs, which deliver 4 bytes of data (plus parity) in every memory request.

DIMM: With the increase in speed and bandwidth capability, a new standard for the memory was adopted called *dual in-line memory module* (*DIMM*). These chips have 168 pins in two (or dual) rows of contacts; one on each side of the card. With the additional pins, a CPU retrieves information from the DIMM chip at 64 bits as compared to a 32- or 16-bit transfer with SIMMs. Some of the physical differences between 168-pin DIMMs and 72-pin SIMMs include the length of the module, the number of notches on the module and the way the module is installed. The main difference between the two is that on a SIMM, opposing pins on either side of the board are tied together to form one electrical contact, while on a DIMM, opposing pins remain electrically isolated to form two separate contacts. DIMMs are often used in computer configurations that support a 64-bit or wider memory bus (like Intel's Pentium 4).

Storage Devices

Disk drives are the important components present inside the system case. These drives are used to read and write information to and from the processor. The three most common disk drives located inside a system case are the hard drive, floppy disk drive and CD-ROM. These drives are high-storage devices, which enable the user to store large amount of data. Out of these drives, the hard disk drive provides the largest storage space for saving. All the vital applications ranging from the operating system to word processor are stored in the hard disk drive. The hard disk drive is costly and not robust enough to transfer data physically; therefore, CD-ROMs and floppy disks are used as an alternative means to transfer data physically.

Processors

Processor, often called the CPU, is the central component of the computer. It is referred to as the brain of a computer responsible for carrying out operations in an efficient and effective manner. A processor holds the key for carrying out all the processing and computational work. Every work that is done by the user on the computer is performed either directly or indirectly by the processor. The following factors should be considered while choosing a processor of a computer system:

- **Performance**: The processor's capabilities dictate the maximum performance of a system. It is the most important single determinant of system performance (in terms of speed and accuracy) in the computer.
- **Speed**: The speed of a processor defines how fast it can perform operations. There are many ways to indicate speed, but the most obvious way to measure is through the internal clock speed of the CPU. The faster the speed of the internal clock of the processor, the faster the CPU will work and, therefore, hardware will be more expensive.
- **Software Support**: New and faster processors support resource-consuming software in a better manner. For example, new processors such as the Pentium 4 enable the use of specialized software, which were not supported on earlier machines.
- **Reliability and Stability**: The reliability of the computer system directly depends on the type and quality of the processor.
- **Energy Consumption and Cooling**: Although processors consume relatively little power compared to other system devices, a newer processor consumes a great deal of power resulting in the impact on everything from the cooling method selection to overall system reliability.
- **Motherboard Support**: The type of processor used in the system is a major determining factor of the chipset used on the motherboards. The motherboard, in turn, dictates many facets of the system's capabilities and performance.

LET US SUMMARIZE

1. The *CPU* or the *processor* is a chip inside the system plugged onto the motherboard and controls all internal and external devices as well as performs arithmetic and logic operations. The unit consists of three main subsystems, the CU, the ALU and the registers.
2. The functions of a processor include carrying out arithmetic and logic functions, controlling the use of the main memory to store data and instructions, and controlling the sequence of operations and all the parts of the computer system.

3. The set of wires used for interconnection of various units is known as the *system bus*. The system bus is divided into three logical units, namely the address bus, the data bus and the control bus.
4. The *registers* are special-purpose, high-speed temporary memory units. They hold various types of information pertaining to data, instructions, addresses and the intermediate results of calculations. Some of the important registers within the CPU are the PC, IR, MAR, MBR, ACC and DR.
5. The ALU carries out arithmetic and logical operations on the data made available to it. The ALU contains the circuitry that is responsible for performing the actual computing and carrying out the arithmetic calculations. On the other hand, the importance of the ALU is to facilitate the CPU to make logical operations based on the instructions provided to it.
6. The CPU controls the I/O devices and transfers data to and from the primary storage. It repeats a set of four basic operations: fetching, decoding, executing and storing.
7. The main memory unit is a collection of registers, logically integrated to the processor but physically separate from the processor. It is implemented by two types of memory technologies, namely *RAM* and *ROM*.
8. A cache memory, sometimes called a *cache store*, is a portion of the memory made up of the high-speed static RAM (SRAM) instead of the slower and comparatively cheaper dynamic RAM (DRAM) used for the main memory. It is very expensive and generally incorporated in the processor, where valuable data and program segments are kept.
9. In the processor to memory communication, the whole process can be divided into two steps, namely information transfer from the memory to the processor and writing information in the memory.
10. In the processor to I/O devices communication, the I/O units are connected to the computer through the system bus. In addition, it buffers the flow of data from the device to the processor and vice versa.
11. The *instruction cycle* is sequence of well-defined instructions in the form of programs, which consist of the following steps: fetch cycle (fetching the instruction from the memory), decode cycle (decoding the instruction), execute cycle (executing the instruction) and store cycle (storing the result back to the memory).
12. The *instruction set* of a processor is a limited set of basic operations built into the processor. Based upon the instruction sets, there are two common types of architectures: CISC and RISC.
13. A *system case* or *cabinet* is a metal and plastic box that houses the main components of the computer. It protects all the vital electronic components against heat, light and temperature.
14. A *power supply* or *SMPS* is a transformer and voltage control device in the computer that furnishes power to all the electronic components by converting incoming AC supply into the low-voltage DC supply. When a computer is turned on, the power

supply allows the converted electricity to travel to other components inside the computer.

15. A *motherboard* is a large multilayered printed circuit board inside a computer. It contains the CPU, BIOS ROM chip, CMOS setup information and is equipped with expansion slots for installing different adapter cards, the connector for the keyboard, slots for the system's RAM, etc.
16. *Ports* and *interfaces* are a generic name for the various sockets, found at the back of the computer, using which external devices are connected to the computer's motherboard.
17. An *expansion card* is a circuit board that provides additional capabilities to the computer system. These cards are made up of large-scale integrated circuit components installed on it. The cards are plugged into the expansion sockets present in the computer's motherboard to give the computer an added functionality.
18. *Ribbon cables* are wide, flat and insulated cables, which are flexible to fit into areas with little space. These cables connect the hard drive, floppy drive and CD-ROM drive to the connectors on the motherboard and control the drives by getting and sending data from and to them.
19. A *number system* defines a set of values used to represent a “quantity”. These are of two types: non-positional and positional number systems.
20. In a non-positional number system, special symbols or characters are used to indicate the values. It is very difficult to perform arithmetic operations with such a number system, as it has no symbol for zero.
21. In a positional number system, the value of each digit in a number is not only defined by the symbols, but also defined by the symbol's position. These symbols are called digits.
22. The positional number system, which is being used nowadays, is called as the *decimal number system*. Apart from this number system, there are some other positional number systems, such as the binary number system, octal number system and hexadecimal number system.
23. The *base* or *radix* of the number system tells about the number of symbols or digits used in the system. The base of the decimal number system is 10, of binary number system is 2, of octal number system is 8 and of hexadecimal number system is 16.
24. In computers, the coding scheme is made up of fixed-size groups of binary positions, where each binary position in a group is assigned a specific value. Some of the most commonly used coding systems are the *ASCII* code and *Unicode*.

EXERCISES

Fill in the Blanks

1. The .... controls the I/O devices and transfer of data to and from the primary storage.
2. .... enables the processor to access data quickly whenever they are needed.
3. The .... informs the CPU about the location of the data residing in the memory.
4. The .... is responsible for recognizing which operation the bit pattern represents.
5. Decoding and executing instructions are complicated and time-consuming in the .... processor architecture.
6. A metal and plastic box that houses the main components of the computer and protects against the heat, light, temperature and other means is called ....
7. .... is a large multilayered printed circuit board inside a computer that contains CPU, BIOS ROM chip and the CMOS setup information
8. .... is a general-purpose communications port through which data are passed slowly over long distances.
9. .... is an expansion card that allows two computers to communicate over ordinary phone lines.
10. The number systems can be categorized into two broad categories: .... and ....

Multiple-choice Questions

1. The processor is a .... chip plugged onto the motherboard in a computer system.

    1. LSI
    2. VLSI
    3. ULSI
    4. XLSI

2. The set of wires, which carry information in a controlled manner, is called ....

0. System bus
    1. Public bus
    2. Private bus
    3. None of these

3. The ALU works on the instructions and data held in the ....

    0. Notebook
    1. Registers
    2. Copy Pad

3. I/O devices

4. A register that keeps track of the next instruction to be executed is called a ....

0. Program Counter
1. Instruction register
2. Accumulator
3. Data register

5. The fastest memory in a computer system is ....

0. ROM
1. RAM
2. Cache
3. None of these

6. In the sequence of events that takes place in an instruction cycle, the first cycle is ....

0. Store cycle
1. Execute cycle
2. Fetch cycle
3. Decode cycle

7. .... is a type of the processor architecture that utilizes a small, highly optimized set of instructions.

0. CISC
1. RISC
2. VISC
3. LISC

8. RISC processors are ideal for embedded applications, such as mobile phones and PDAs, because ....

0. They are smaller in size and consume less power.
1. They are large in size and consume less power.
2. They are smaller in size and consume more power.
3. They are larger in size and consume large amount of power.

9. The concept of the CISC architecture is to accomplish the task in ....

0. As longer lines of code as possible
1. As few lines of code as possible

2. Both (a) and (b)
3. None of these

10. An expansion card that allows the computer to output sound through connected speakers is ....

0. Video Card
    1. Network Interface Card
    2. PC Card
    3. Sound Card

State True or False

1. A processor only operates on binary data, that is, data composed of 1s and 0s.
2. The address bus informs the CPU about the location of the data residing in the memory.
3. Registers are special-purpose, high-speed permanent memory units.
4. A program counter is a general-purpose register used for storing temporary results and results produced by the ALU.
5. The L2 cache is usually 64 KB–2 MB in size.
6. A power supply is a transformer that converts power into the low-voltage AC supply.
7. A USB port is a plug-and-play hardware interface that supports maximum bandwidth of 12 Mbit per second.
8. The PC card is a removable device, approximately the size of a credit card.
9. SIMM chips are faster than DIMM chips.
10. The Unicode is a character-encoding standard developed by the Unicode Consortium.

Descriptive Questions

1. What do you understand by the CPU? Describe in details the various units of the CPU.
2. What are registers in a CPU? Name five registers with their functions.
3. Write a detailed note on the instruction cycle describing the various steps involved.
4. Compare the RISC and CISC architecture briefly. Also, discuss their advantages and disadvantages.
5. What is a system bus? Name the various units of the system bus.
6. With an appropriate example, explain the conversion of:

    1. Binary to octal and vice versa
    2. Binary to hexadecimal and vice versa
    3. Octal to hexadecimal and vice versa

7. What are expansion cards? How many types of expansion cards can be used in a computer system?
8. Write short notes on:

0. Motherboard
    1. Power supply
    2. Ports
    3. Ribbon cables

9. Perform the following conversions:

0. Convert binary 00011011 to decimal
    1. Convert decimal 278 to binary
    2. Convert decimal 0.625 to binary

10. What is a radix or base of the system? With the help of this system, brief the various types of number systems.

ANSWERS

Fill in the Blanks

1. CU
2. Cache
3. Address bus
4. Decode cycle
5. CISC
6. System case
7. Motherboard
8. Serial port
9. Modem
10. Non-positional number system, positional number system

Multiple-choice Questions

1. (b)
2. (a)
3. (b)
4. (a)
5. (c)
6. (c)
7. (b)

8. (a)
9. (b)
10. (d)

State True or False

1. True
2. True
3. False
4. False
5. True
6. False
7. True
8. True
9. False
10. True