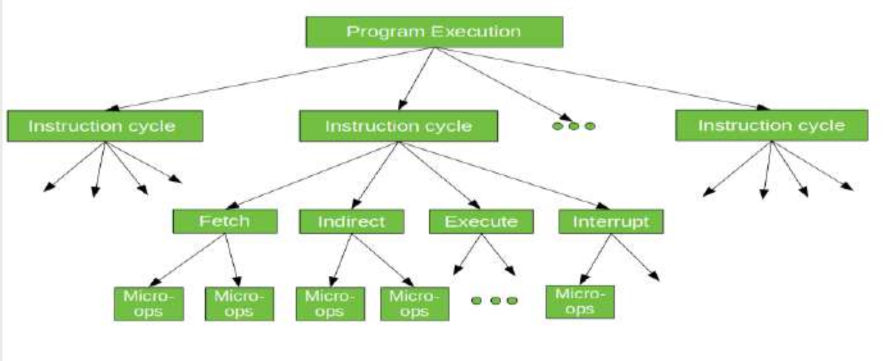
TOPIC 1:CPU control unit design:

A.Computer Organization | Micro-Operation

In computer central processing units, **micro-operations** (also known as micro-ops) are the functional or atomic, operations of a processor. These are low level instructions used in some designs to implement complex machine instructions. They generally perform operations on data stored in one or more registers. They transfer data between registers or between external buses of the CPU, also performs arithmetic and logical operations on registers.
In executing a program, operation of a computer consists of a sequence of instruction cycles, with one machine instruction per cycle. Each instruction cycle is made up of a number of smaller units – *Fetch, Indirect, Execute and Interrupt cycles*. Each of these cycles involves series of steps, each of which involves the processor registers. These steps are referred as micro-operations. the prefix micro refers to the fact that each of the step is very simple and accomplishes very little. Figure below depicts the concept being discussed here.



Summary: Execution of a program consists of sequential execution of instructions. Each instruction is executed during an instruction cycle made up of shorter sub-cycles(example – fetch, indirect, execute, interrupt). The performance of each sub-cycle involves one or more shorter operations, that is, *micro-operations.*
In my next article I will give detailed information of each Instruction Cycle.

Microarchitecture and Instruction Set Architecture

In this article we look at what an *Instruction Set Architecture (ISA)* is and what is the difference between an **'ISA'** and *Microarchitecture*. An **ISA** is defined as the design of a computer from the *Programmer's Perspective*.
This basically means that an **ISA** describes the **design of a Computer** in terms of the **basic operations** it must support. The ISA is not concerned with the implementation specific details of a computer. It is only concerned with the set or collection of basic operations the computer must support. For example the AMD Athlon and the Core 2 Duo processors have entirely different implementations but they support more or less the same set of basic operations as defined in the x86 Instruction Set.
Let us try to understand the Objectives of an ISA by taking the example of the **MIPS ISA**. MIPS is one of the most widely used ISAs in education due to its simplicity.

1. The ISA defines the **types of instructions** to be supported by the processor.
   Based on the type of operations they perform MIPS Instructions are classified into 3 types:

- **Arithmetic/Logic Instructions:**
  These Instructions perform various Arithmetic & Logical operations on one or more operands.

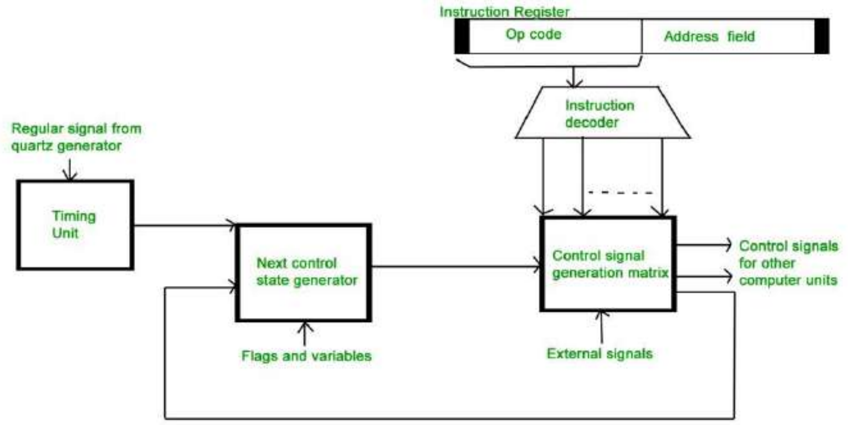
- **Data Transfer Instructions:**
  These instructions are responsible for the transfer of instructions from memory to the processor registers and vice versa.
- **Branch and Jump Instructions:**
  These instructions are responsible for breaking the sequential flow of instructions and jumping to instructions at various other locations, this is necessary for the implementation of *functions* and *conditional statements*.

B.Computer Organization | Hardwired v/s Micro-programmed Control Unit

To execute an instruction, the control unit of the CPU must generate the required control signal in the proper sequence. There are two approaches used for generating the control signals in proper sequence as Hardwired Control unit and Micro-programmed control unit.

HardwiredControlUnit –
The control hardware can be viewed as a state machine that changes from one state to another in every clock cycle, depending on the contents of the instruction register, the condition codes and the external inputs. The outputs of the state machine are the control signals. The sequence of the operation carried out by this machine is determined by the wiring of the logic elements and hence named as "hardwired".

- Fixed logic circuits that correspond directly to the Boolean expressions are used to generate the control signals.
- Hardwired control is faster than micro-programmed control.
- A controller that uses this approach can operate at high speed.
- RISC architecture is based on hardwired control unit

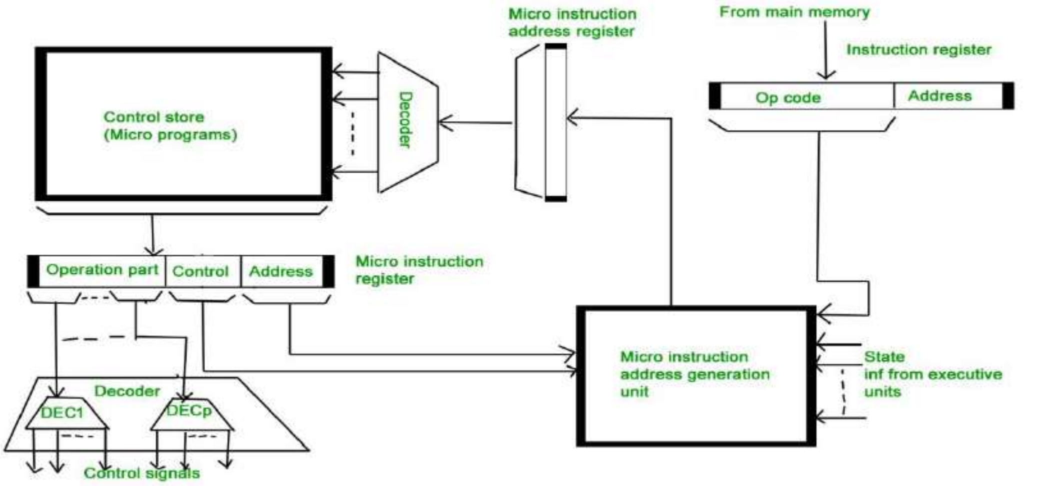


Micro-programmed Control Unit –

- The control signals associated with operations are stored in special memory units inaccessible by the programmer as Control Words.
- Control signals are generated by a program are similar to machine language programs.
- Micro-programmed control unit is slower in speed because of the time it takes to fetch microinstructions from the control memory.

Some Important Terms –

1. **Control Word :** A control word is a word whose individual bits represent various control signals.
2. **Micro-routine :** A sequence of control words corresponding to the control sequence of a machine instruction constitutes the micro-routine for that instruction.
3. **Micro-instruction :** Individual control words in this micro-routine are referred to as microinstructions.
4. **Micro-program :** A sequence of micro-instructions is called a micro-program, which is stored in a ROM or RAM called a Control Memory (CM).
5. **Control Store :** the micro-routines for all instructions in the instruction set of a computer are stored in a special memory called the Control Store.



Types of Micro-programmed Control Unit – Based on the type of Control Word stored in the Control Memory (CM), it is classified into two types :

1. Horizontal Micro-programmed control Unit :
The control signals are represented in the decoded binary format that is 1 bit/CS. Example: If 53 Control signals are present in the processor than 53 bits are required. More than 1 control signal can be enabled at a time.

- It supports longer control word.
- It is used in parallel processing applications.
- It allows higher degree of parallelism. If degree is n, n CS are enabled at a time.
- It requires no additional hardware(decoders). It means it is faster than Vertical Microprogrammed.
- It is more flexible than vertical microprogrammed

2. Vertical Micro-programmed control Unit :
The control signals re represented in the encoded binary format. For N control signals- Log2(N) bits are required.

- It supports shorter control words.
- It supports easy implementation of new conrol signals therefore it is more flexible.
- It allows low degree of parallelism i.e., degree of parallelism is either 0 or 1.
- Requires an additional hardware (decoders) to generate control signals, it implies it is slower than horizontal microprogrammed.
- It is less flexible than horizontal but more flexible than that of hardwired control unit.

Computer performance is the amount of work accomplished by a computer system. The word performance in computer performance means "How well is the computer doing the work it is supposed to do?". It basically depends on response time, throughput and execution time of a computer system.

Response time is the time from start to completion of a task. This also includes:

- Operating system overhead.
- Waiting for I/O and other processes
- Accessing disk and memory
- Time spent executing on the CPU or execution time.

Throughput is the total amount of work done in a given time.

CPU execution time is the total time a CPU spends computing on a given task. It also excludes time for I/O or running other programs. This is also referred to as simply CPU time. Performance is determined by execution time as performance is inversely proportional to execution time.

Performance = (1 / Execution time)

And,

(Performance of A / Performance of B)

= (Execution Time of B / Execution Time of A)

If given that Processor A is faster than processor B, that means execution time of A is less than that of execution time of B. Therefore, performance of A is greater than that of performance of B.

Example –

Machine A runs a program in 100 seconds, Machine B runs the same program in 125 seconds

(Performance of A / Performance of B)

= (Execution Time of B / Execution Time of A)

= 125 / 100 = 1.25

That means machine A is 1.25 times faster than Machine B.

And, the time to execute a given program can be computed as:

Execution time = CPU clock cycles x clock cycle time

Since clock cycle time and clock rate are reciprocals, so,

Execution time = CPU clock cycles / clock rate

The number of CPU clock cycles can be determined by,

CPU clock cycles

= (No. of instructions / Program) x (Clock cycles / Instruction)

= Instruction Count x CPI

Which gives,

Execution time

= Instruction Count x CPI x clock cycle time

= Instruction Count x CPI / clock rate

The units for CPU Execution time are:

$$\text{CPU time} = \frac{\text{Seconds}}{\text{Program}} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Cycles}}{\text{Instruction}} \times \frac{\text{Seconds}}{\text{Cycle}}$$

TOPIC 1: Semiconductor memory:

- Semiconductor memory is used in any electronics assembly that uses computer processing technology. Semiconductor memory is the essential electronics component needed for any computer based PCB assembly. The use of semiconductor memory has grown, and the size of these memory cards has increased as the need for larger and larger amounts of storage is needed. Terms like DDR3, DDR4, DDR5 and many more are seen and these refer to different types of SDRAM semiconductor memory. Semiconductor memory: main types
- There are two main types or categories that can be used for semiconductor technology. These memory types or categories differentiate the memory to the way in which it operates:
- RAM - Random Access Memory: As the names suggest, the RAM or random access memory is a form of semiconductor memory technology that is used for reading and writing data in any order - in other words as it is required by the processor. It is used for such applications as the computer or processor memory where variables and other stored and are required on a random basis. Data is stored and read many times to and from this type of memory.

  Random access memory is used in huge quantities in computer applications as current day computing and processing technology requires large amounts of memory to enable them to handle the memory hungry applications used today. Many types of RAM including SDRAM with its DDR3, DDR4, and soon DDR5 variants are used in huge quantities.
- ROM - Read Only Memory: A ROM is a form of semiconductor memory technology used where the data is written once and then not changed. In view of this it is used where data needs to be stored permanently, even when the power is removed - many memory technologies lose the data once the power is removed.
- Semiconductor memory technologies
- There is a large variety of types of ROM and RAM that are available. Often the overall name for the memory technology includes the initials RAM or ROM and this gives a guide as to the overall type of format for the memory.
- With technology moving forwards apace, not only are the established technologies moving forwards with SDRAM technology moving from DDR3 to DDR4 and then to DDR5, but Flash memory used in memory cards is also developing as are the other technologies.
- In addition to this, new memory technologies are arriving on the scene and they are starting to make an impact in the market, enabling processor circuits to perform more effectively.

- The different memory types or memory technologies are detailed below:
- DRAM: Dynamic RAM is a form of random access memory. DRAM uses a capacitor to store each bit of data, and the level of charge on each capacitor determines whether that bit is a logical 1 or 0.

  However these capacitors do not hold their charge indefinitely, and therefore the data needs to be refreshed periodically. As a result of this dynamic refreshing it gains its name of being a dynamic RAM. DRAM is the form of semiconductor memory that is often used in equipment including personal computers and workstations where it forms the main RAM for the computer. The semiconductor devices are normally available as integrated circuits for use in PCB assembly in the form of surface mount devices or less frequently now as leaded components
- EEPROM: This is an Electrically Erasable Programmable Read Only Memory. Data can be written to these semiconductor devices and it can be erased using an electrical voltage. This is typically applied to an erase pin on the chip. Like other types of PROM, EEPROM retains the contents of the memory even when the power is turned off. Also like other types of ROM, EEPROM is not as fast as RAM.
- EPROM: This is an Erasable Programmable Read Only Memory. These semiconductor devices can be programmed and then erased at a later time. This is normally achieved by exposing the semiconductor device itself to ultraviolet light. To enable this to happen there is a circular window in the package of the EPROM to enable the light to reach the silicon of the device. When the PROM is in use, this window is normally covered by a label, especially when the data may need to be preserved for an extended period.

  The PROM stores its data as a charge on a capacitor. There is a charge storage capacitor for each cell and this can be read repeatedly as required. However it is found that after many years the charge may leak away and the data may be lost.

  Nevertheless, this type of semiconductor memory used to be widely used in applications where a form of ROM was required, but where the data needed to be changed periodically, as in a development environment, or where quantities were low.
- Flash memory: Flash memory may be considered as a development of EEPROM technology. Data can be written to it and it can be erased, although only in blocks, but data can be read on an individual cell basis.
- F-RAM: Ferroelectric RAM is a random-access memory technology that has many similarities to the standard DRAM technology. The major difference is that it incorporates a ferroelectric layer instead of the more usual dielectric layer and this provides its non-volatile capability. As it offers a non-volatile capability, F-RAM is a direct competitor to Flash. MRAM: This is Magneto-resistive RAM, or Magnetic RAM. It is a non-volatile RAM memory technology that uses magnetic charges to store data instead of electric charges. PROM: This stands for Programmable Read Only Memory. It is a semiconductor memory which can only have data written to it once - the data written to it is permanent. These memories are bought in a blank format and they are programmed using a special PROM programmer.

  Typically a PROM will consist of an array of fuseable links some of which are "blown" during the programming process to provide the required data pattern.
- SDRAM: Synchronous DRAM. This form of semiconductor memory can run at faster speeds than conventional DRAM. It is synchronised to the clock of the processor and is

capable of keeping two sets of memory addresses open simultaneously. I/O Interface (Interrupt and DMA Mode)

•

TOPIC 3 : Input-output subsystems

Mode of Transfer:

The binary information that is received from an external device is usually stored in the memory unit. The information that is transferred from the CPU to the external device is originated from the memory unit. CPU merely processes the information but the source and target is always the memory unit. Data transfer between CPU and the I/O devices may be done in different modes.

Data transfer to and from the peripherals may be done in any of the three possible ways

1. Programmed I/O.
2. Interrupt- initiated I/O.
3. Direct memory access(DMA).

Now let’s discuss each mode one by one.

1. **Programmed I/O:** It is due to the result of the I/O instructions that are written in the computer program. Each data item transfer is initiated by an instruction in the program. Usually the transfer is from a CPU register and memory. In this case it requires constant monitoring by the CPU of the peripheral devices.

Example of Programmed I/O: In this case, the I/O device does not have direct access to the memory unit. A transfer from I/O device to memory requires the execution of several instructions by the CPU, including an input instruction to transfer the data from device to the CPU and store instruction to transfer the data from CPU to memory. In programmed I/O, the CPU stays in the program loop until the I/O unit indicates that it is ready for data transfer. This is a time consuming process since it needlessly keeps the CPU busy. This situation can be avoided by using an interrupt facility.

2.Interrupt- initiated I/O: Since in the above case we saw the CPU is kept busy unnecessarily. This situation can very well be avoided by using an interrupt driven method for data transfer. By using interrupt facility and special commands to inform the interface to issue an interrupt request signal whenever data is available from any device. In the meantime the CPU can proceed for any other program execution. The interface meanwhile keeps monitoring the device. Whenever it is determined that the device is ready for data transfer it initiates an interrupt request signal to the computer. Upon detection of an external interrupt signal the CPU stops momentarily the task that it was already performing, branches to the service program to process the I/O transfer, and then return to the task it was originally performing.

3.Direct Memory Access (DMA) in Computer Architecture

For the execution of a computer program, it requires the synchronous working of more than one component of a computer. For example, Processors – providing necessary control information, addresses…etc, buses – to transfer information and data to and from memory to I/O devices…etc.

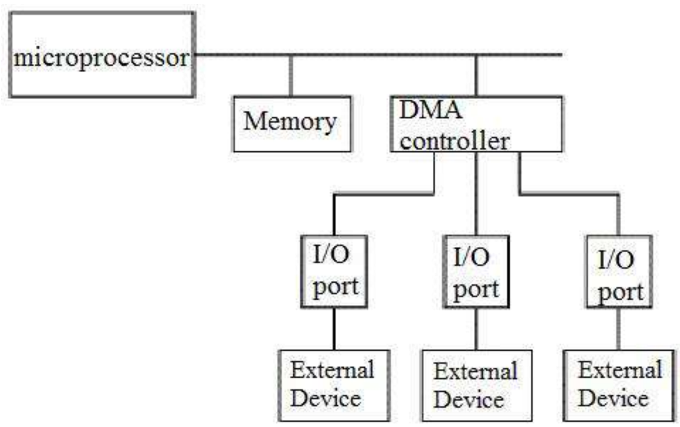
What is a DMA Controller?

The term DMA stands for direct memory access. The hardware device used for direct memory access is called the DMA controller. DMA controller is a control unit, part of I/O device's interface circuit, which can transfer blocks of data between I/O devices and main memory with minimal intervention from the processor.

DMA Controller Diagram in Computer Architecture

DMA controller provides an interface between the bus and the input-output devices. Although it transfers data without intervention of processor, it is controlled by the processor. The processor initiates the DMA controller by sending the starting address, Number of words in the data block and direction of transfer of data .i.e. from I/O devices to the memory or from main memory to I/O devices. More than one external device can be connected to the DMA controller.

DMA controller contains an address unit, for generating addresses and selecting I/O device for transfer. It also contains the control unit and data count for keeping counts of the number of blocks transferred and indicating the direction of transfer of data. When the transfer is completed, DMA informs the processor by raising an interrupt. The typical block diagram of the DMA controller is shown in the figure below.

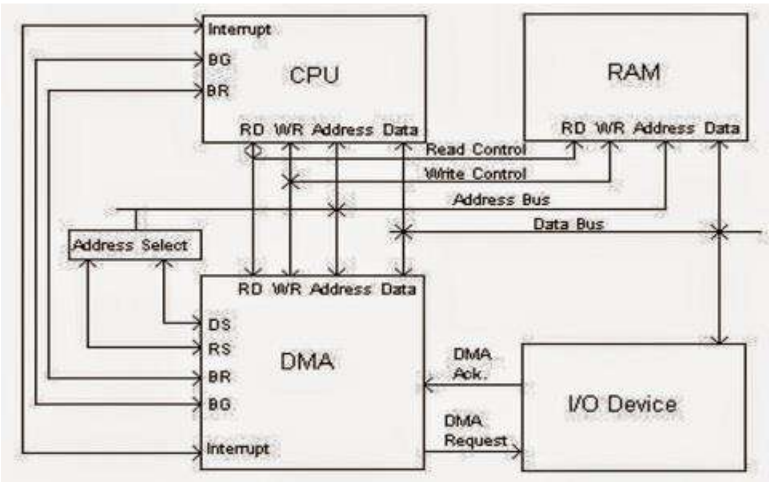


Typical Block Diagram of DMA Controller

Working of DMA Controller

DMA controller has to share the bus with the processor to make the data transfer. The device that holds the bus at a given time is called bus master. When a transfer from I/O device to the memory or vice verse has to be made, the processor stops the execution of the current program, increments the program counter, moves data over stack then sends a DMA select signal to DMA controller over the address bus.

If the DMA controller is free, it requests the control of bus from the processor by raising the bus request signal. Processor grants the bus to the controller by raising the bus grant signal, now DMA controller is the bus master. The processor initiates the DMA controller by sending the memory addresses, number of blocks of data to be transferred and direction of data transfer. After assigning the data transfer task to the DMA controller, instead of waiting ideally till completion of data transfer, the processor resumes the execution of the program after retrieving instructions from the stack.



Transfer Of Data in Computer By DMA Controller

DMA controller now has the full control of buses and can interact directly with memory and I/O devices independent of CPU. It makes the data transfer according to the control instructions received by the processor. After completion of data transfer, it disables the bus request signal and CPU disables the bus grant signal thereby moving control of buses to the CPU.

When an I/O device wants to initiate the transfer then it sends a DMA request signal to the DMA controller, for which the controller acknowledges if it is free. Then the controller requests the processor for the bus, raising the bus request signal. After receiving the bus grant signal it transfers the data from the device. For n channeled DMA controller n number of external devices can be connected.

The DMA transfers the data in three modes which include the following.

a) **Burst Mode**: In this mode DMA handover the buses to CPU only after completion of whole data transfer. Meanwhile, if the CPU requires the bus it has to stay ideal and wait for data transfer.

b) **Cycle Stealing Mode**: In this mode, DMA gives control of buses to CPU after transfer of every byte. It continuously issues a request for bus control, makes the transfer of one byte and returns the bus. By this CPU doesn't have to wait for a long time if it needs a bus for higher priority task.

c) **Transparent Mode:** Here, DMA transfers data only when CPU is executing the instruction which does not require the use of buses.

•

TOPIC 4 : INTERRUPTS

Interrupts

Data transfer between the CPU and the peripherals is initiated by the CPU. But the CPU cannot start the transfer unless the peripheral is ready to communicate with the CPU. When a device is ready to communicate with the CPU, it generates an interrupt signal. A number of input-output devices are attached to the computer and each device is able to generate an interrupt request

---

Priority Interrupt

A priority interrupt is a system which decides the priority at which various devices, which generates the interrupt signal at the same time, will be serviced by the CPU. The system has authority to decide which conditions are allowed to interrupt the CPU, while some other interrupt is being serviced. Generally, devices with high speed transfer such as *magnetic disks* are given high priority and slow devices such as *keyboards* are given low priority.

When two or more devices interrupt the computer simultaneously, the computer services the device with the higher priority first.

Types of Interrupts:

Following are some different types of interrupts:

A.**Hardware Interrupts**

When the signal for the processor is from an external device or hardware then this interrupts is known as **hardware interrupt**.

Let us consider an example: when we press any key on our keyboard to do some action, then this pressing of the key will generate an interrupt signal for the processor to perform certain action. Such an interrupt can be of two types:

B.Maskable Interrupt: The hardware interrupts which can be delayed when a much high priority interrupt has occurred at the same time.

C.Non Maskable Interrupt The hardware interrupts which cannot be delayed and should be processed by the processor immediately.

D.Software Interrupts

The interrupt that is caused by any internal system of the computer system is known as a **software interrupt**. It can also be of two types:

E.Normal Interrupt: The interrupts that are caused by software instructions are called **normal software interrupts**.

F.Exception: Unplanned interrupts which are produced during the execution of some program are called **exceptions**, such as division by zero.

TOPIC 6 : I/O device interfaces – SCII, USB.

SCSI device

SCSI originally stood for Small Computer System Interface, but it's really outgrown the "small" designation. It's a fast bus that can connect lots of devices to a computer at the same time, including hard drives, scanners, CD-ROM/RW drives, printers and tape drives. Other technologies, like serial-ATA (SATA), have largely replaced it in new systems, but SCSI is still in use. This article will review SCSI basics and give you lots of information on SCSI types and specifications.Common SCSI components.

There are several components used in SCSI storage systems:

- **Initiator.** An initiator issues requests for service by the SCSI device and receives responses. Initiators come in a variety of forms and may be integrated into a server's system board or exist within a host bus adapter. ISCSI connectivity typically uses a software-based initiator.
- **Target.** A SCSI target is typically a physical storage device (although software-based SCSI targets also exist). The target can be a hard disk or an entire storage array. It is also possible for non-storage hardware to function as a SCSI target. Although rare today, it was once common for optical scanners to be attached to computers through the SCSI bus and to act as SCSI targets.
- **Service delivery subsystem.** The mechanism that allows communication to occur between the initiator and the target; it usually takes the form of cabling.
- **Expander.** Only used with serial-attached SCSI (SAS); allows multiple SAS devices to share a single initiator port.

USB

- Stands for "Universal Serial Bus." USB is the most common type of computer port used in today's computers. It can be used to connect keyboards, mice, game controllers, printers, scanners, digital cameras, and removable media drives, just to name a few. With the help of a few USB hubs, you can connect up to 127 peripherals to a single USB port and use them all at once (though that would require quite a bit of dexterity).
- USB is also faster than older ports, such as serial and parallel ports. The USB 1.1 specification supports data transfer rates of up to 12Mb/sec and USB 2.0 has a maximum transfer rate of 480 Mbps. Though USB was introduced in 1997, the technology didn't really take off until the introduction of the Apple iMac (in late 1998) which used USB ports exclusively. It is somewhat ironic, considering USB was created and designed by Intel, Compaq, Digital, and IBM. Over the past few years, USB has become a widely-used cross-platform interface for both Macs and PCs.