Chapter 4 Microprocessors

The central processing unit, or CPU, is at the core of every computer. Functioning as the "brain" of the computer, the CPU performs basic mathematical and logical functions as instructed by a computer program. The CPU is located on the motherboard and is connected to it through a special port called the CPU socket. When the CPU is in use it generates heats, which must be transferred away from CPU chip so that it is not damaged. This is accomplished through the use of a heat-sink and fan which draws the heat out of the CPU chip and transfers it into the case instead.

A CPU has a maximum speed at which it can safely process instructions. This speed is called the *clock rate* and is measured in hertz (Hz). Some CPUs will allow a user to raise the clock rate beyond the speed at which the CPU is safely rated. This is called *overclocking*, and can provide increased computer performance at the cost of potential system instability or damage. Due to the complexity of overclocking and the potentially destructive side effects it can cause, overclocking is a technique not recommended for average or professional computer users.

A relatively recent development in the field of CPUs is the multi-core processor. Having many processing cores in the CPU effectively multiplies its potential by allowing the CPU to simultaneously process as many instructions as it has cores. However, the OS must be designed to make use of multiple cores for this ability to take effect. The CPU of a modern PC will likely have anywhere from two to eight cores

CPU Core Components

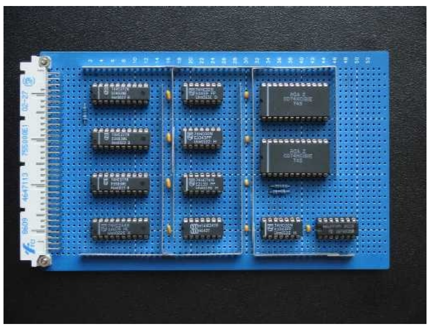
1. Arithmetic Logic Unit (ALU)

The ALU is the part of a CPU that performs all arithmetic computations including addition, subtraction, multiplication, and division. The Arithmetic Logic Unit also performs all logical operations. The ALU is a literally the fundamental building block of the CPU, and even the simplest processors contain an ALU.

Today's modern CPUs and graphics processing units (GPUs) in graphic cards have very complex ALUs, and some contain a number of ALUs.

In some CPUs an individual ALU is further divided into two units called an arithmetic unit (AU) and a logic unit (LU). Some processors even contain more than one AU. Normally the ALU has direct input and output access to the processor controller, main system memory (RAM), and input/output devices.

In a nutshell an ALU works by loading data from what is called input registers, then an external Control Unit tells the ALU what operation to perform on that data, and finally the ALU stores its result into an output register. At this point the data is moved between the registers and the memory via a data path called a bus.



2. Control Unit

The CPUs control unit is responsible for executing or storing the results coming out of the ALU. Within the CPU, the control unit performs the functions of fetch, decode, execute, and store.

The control unit communicates with both the arithmetic logic unit (ALU) and memory, and literally directs the entire computer system to carry out, or execute, stored program instructions.

In a nutshel here is how a control unit works. Basically a control unit fetches or retrieves an instruction from memory and then analyzes the instruction it fetched before deciding how it should be processed. Depending on the action required, the control unit will then send segments of the original instruction to the appropriate section of the processor.

3. Registers

Registers are the temporary storage areas for instructions or data within the processor. Registers are basically special storage locations somewhat similar to a computer's memory though contained within the processor and exceptionally faster.

Registers work under the direction of the control unit to accept, hold and transfer instructions or data and perform arithmetic or logical comparisons at a high rate of speed.

Metaphorically speaking, the control unit uses the CPUs data storage registers similar to the way a cashier at a local market would conveniently store money away in a cash register to be used temporarily for transactions.

4. Clock Speed

The CPUs clock speed/rate is the rate in cycles per second that the processors internal clock operates. A CPUs clock speed is a frequency that is measured in hertz. For example, you may see a processor's speed measured in megahertz (MHz) or gigahertz (GHz). 1000 MHz equals 1 GHz. A modern processors clock speed is rated in gigahertz.

The clock speed actually used to be the main determining factor of how fast a processor is, and generally a faster clock speed equaled better performance, but in today's digital age there are many other factors to consider such as cache size, number of internal processing cores, etc.

As a general rule of thumb if all other processor specifications are equal, (number of cores, cache size, bus speed, etc.) then a CPU with a higher internal clock speed will perform better than a slower clocked CPU.

5. Cores

A multi-core processor is a processing system composed of two or more independent processing cores. More processing cores mean there is literally more than one processing engine within the same CPU package, essentially like a sports car with more than one turbo charger.

More cores allow computer systems to be more responsive while multi-tasking and multiple cores also help numerous applications specifically coded for multiple core processors to perform much faster.

When shopping for a computer look for a processor that contains the most cores, especially if you are a power user or into digital media applications, you may just notice a big performance increase. Many digital media apps like Adobe Photoshop, Adobe Premiere Pro, and others will benefit from multiple processing cores.

6. Cache

Cache is very important processor feature that you should pay attention to when shopping for a modern CPU for it can dramatically speed up the performance of a processor. Cache acts as a buffer between your RAM and the CPU to speed up the processors internal processing engine. Cache is also sometimes referred to as a CPUs buffer so the terms are used interchangeably.

Cache is essentially a super fast memory technology that resides on the CPU; it holds recently accessed or pre-fetched instructions so the processor can grab them quickly to operate more efficiently. Pulling data and instructions from the systems RAM in comparison takes much longer than pulling it from a cache.

Processors can contain multiples levels and different sizes of cache. When looking at a processors specification sheet, you will often see cache referenced as a level 1 or level 2 cache along with an accompanying cache size. When selecting a processor, a larger cache size as well as multiple levels of cache will improve overall CPU performance.

Motherboard

- *Motherboard* – The CPU and memory are mounted on a larger printed circuit board called the motherboard. It connects all of the hardware in the computer and provides the electrical connections. It also contains expansion slots to connect other hardware.

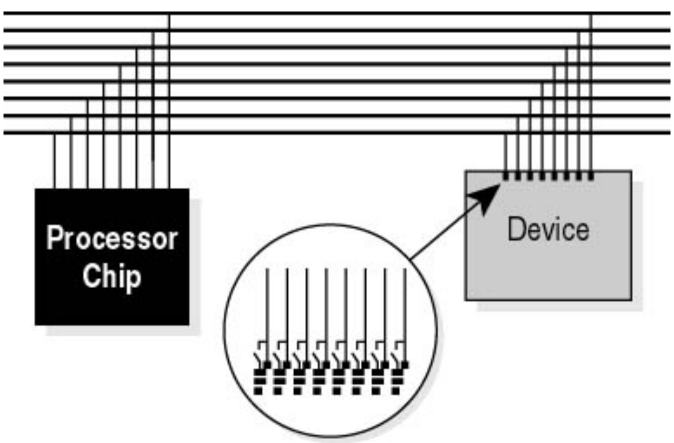
-

External Data Bus

Is a bus that connects a computer to peripheral devices. Two examples are the Universal Serial Bus (USB) and IEEE 1394.

- Determines the clock speed for the CPU
- The CPU uses the external data bus to address RAM
- The external data bus provides a channel for the flow of data and commands between the CPU and RAM>
- The CPU uses the external data bus to access registers

You can think of a bus as a highway on which data travels within a computer. The data bus transfers actual data whereas the address bus transfers information about where the data should go. The size of a bus, known as its *width,* is important because it determines how much data can be transmitted at one time. For example, a 16-bit bus can transmit 16 bits of data, whereas a 32-bitbus can transmit 32 bits Every bus has a clock speed measured in MHz. A fast bus allows data to be transferred faster, which makes applications run faster. Nearly all PCs made today include a local bus for data that requires especially fast transfer speeds, such as video data. The local bus is a high-speed pathway that connects directly to the processor.

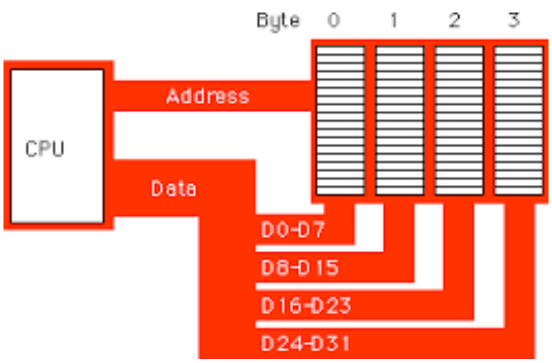


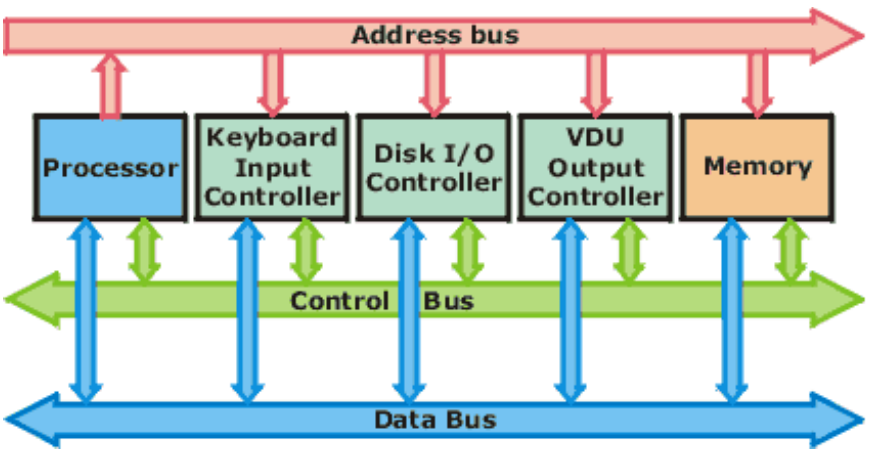
External Data Bus

Address Bus

An address bus is a computer bus architecture used to transfer data between devices that are identified by the hardware address of the physical memory (the physical address), which is stored in the form of binary numbers to enable the data bus to access memory storage.

The address bus is used by the CPU or a direct memory access (DMA) enabled device to locate the physical address to communicate read/write commands. All address busses are read and written by the CPU or DMA in the form of bits.





Address, Control and Data Bus

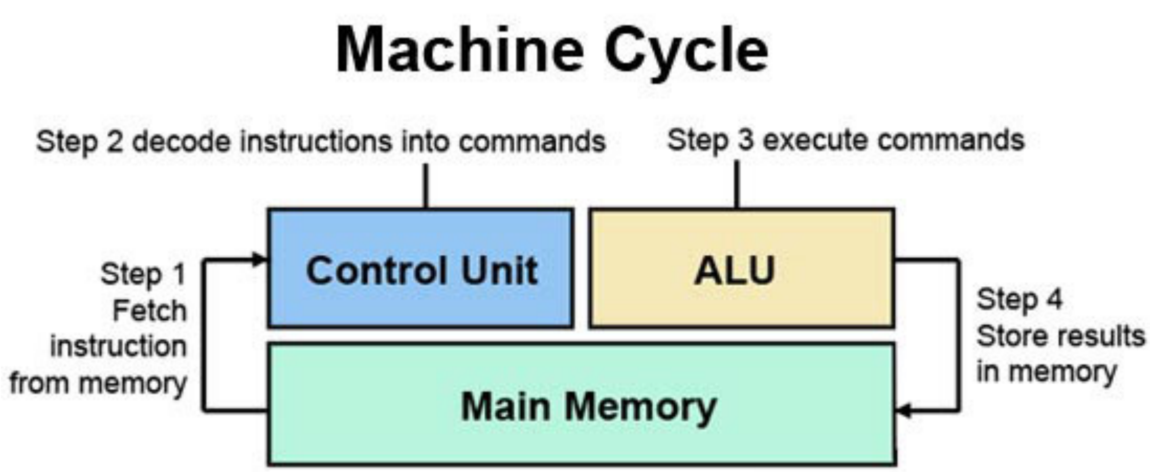
Storage

- ***Storage* – is where data is kept by the computer so the information can be viewed,, played or otherwise used. Data can be stored on hard disk drive, flash memory cards, thumb drives or other external drive.**

- ***Storage – Memory*** **is part of the computer that stores information for *immediate* processing. It can be volatile (erased when power is turned off) or involatile (kept when power is turned off). There are two types of memory:**

1. **Random-access memory**
2. **Read-only memory**

- ***Random-access memory (RAM) -*** **is memory that can be changed. Hardware hold the instruction that the processor can immediately use. When programs are used, constantly changing programs are loaded into RAM. Ram is volatile.**
- **Read-only memory (ROM) – is memory that cannot be changed. It contains static information the computer will always need to operate.**
- **It contains instructions that will tell the system what steps to take during start-up**
- **It holds information even when power is off (involatile)**



Pipelining is an implementation technique where multiple instructions are overlapped in execution. The computer pipeline is divided in **stages**. Each stage completes a part of an instruction in parallel. The stages are connected one to the next to form a pipe - instructions enter at one end, progress through the stages, and exit at the other end.

Pipelining does not decrease the time for individual instruction execution. Instead, it increases instruction throughput. The **throughput** of the instruction pipeline is determined by how often an instruction exits the pipeline.

Because the pipe stages are hooked together, all the stages must be ready to proceed at the same time. We call the time required to move an instruction one step further in the pipeline ***a machine cycle*** *.* The ***length*** of the machine cycle is determined by the time required for the slowest pipe stage.

The pipeline designer's *goal is to balance the length of each pipeline stage* . If the stages are perfectly balanced, then the time per instruction on the pipelined machine is equal to

Cache

Pronounced *cash*, a special high-speed storage mechanism. Cache can be either a reserved section of main memory or an independent high-speed storage device. Two types of caching are commonly used in personal computers: *memory caching* and *disk caching*.

Memory Caching

A memory cache, sometimes called a *cache store* or *RAM cache*, is a portion of memory made of high-speed static RAM (SRAM) instead of the slower and cheaper dynamic RAM (DRAM) used for main memory. Memory caching is effective because most programs access the same data or instructions over and over. By keeping as much of this information as possible in SRAM, the computer avoids accessing the slower DRAM.

Multicore Processing

A multi-core processor is an integrated circuit (IC) to which two or more processors have been attached for enhanced performance, reduced power consumption, and more efficient simultaneous processing of multiple tasks (*see* parallel processing). A dual core set-up is somewhat comparable to having multiple, separate processors installed in the same computer, but because the two processors are actually plugged into the same socket, the connection between them is faster. Ideally, a dual core processor is nearly twice as powerful as a single core processor. In practice, performance gains are said to be about fifty percent: a dual core processor is likely to be about one-and-a-half times as powerful as a single core processor. Multi-core processing is a growing industry trend as single-core processors rapidly reach the physical limits of possible complexity and speed. Most current systems are multi-core. Systems with a large number of processor core -- tens or hundreds -- are sometimes referred to as many-core or massively multi-core systems.

Integrated Memory Controller

Most microprocessors have the IMC moved from the motherboard to the CPU to optimize the flow of information into and out from the CPU. It enables faster control over things like the large L3 cache shared among multiple cores.