ICS3U: Introduction to Computer Science

# How Computers Work

J. Garvin





## Computers and Programs

Originally, computing machines were hard-wired to perform specific tasks.

Almost all personal computers, workstations, minicomputers, and mainframes are based on the design principle of stored programs.

In 1945, John Von Neumann proposed that the program to control the computer should be stored in the memory of the computer. Changing the program in memory allowed the computer to perform a completely different computation.

This allowed computers to become general-purpose problem solving machines.





## Computers and Programs

What is a computer program?

- Sequence of instructions
- Performs some task

Typically, a program involves some input and processes some output, but this may not always be true.





## Program Execution

When a program is created, it may be run in one of two ways:

- Compiled: a program is converted into a machine-readable format using a binary encoding.
- Interpreted: The program requires a separate program, called an *interpreter*, to translate program instructions on-the-fly.

Some programming languages, like C/C++, are compiled. Others, like Java or Python, are interpreted.





## Program Execution

Steps in program execution:

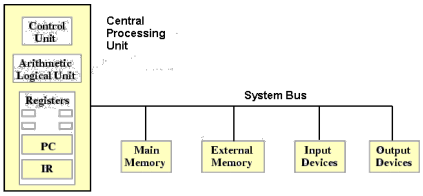
1. A program is stored in memory (RAM).
2. CPU fetches instructions and data from there via the system *bus*.
3. CPU decodes, and executes the stored instructions of the program sequentially, inputting data as needed, and outputting results via the bus.
4. Steps 2-3 repeated until the program has completed.

Von Neumann's architecture describes a computer with four main sections: the *Arithmetic and Logic Unit* (ALU), the *control circuitry*, the *memory*, and the *input* and *output* (I/O) devices. These parts are interconnected by wires (referred to as the *bus*).





## Computers and Programs



## Central Processing Unit (CPU)

Arithmetic-Logic Unit (ALU)

- Performs arithmetic operations (add, sub, mult, div).
- Performs logical operations ($<, >, =$).
- Allows computer to calculate and compare.

Control Unit

- Directs the movement of electronic signals between memory and the ALU.
- Coordinates control signals between CPU and input/output devices.
- Tells the computer system *how* to execute a program.





## Registers

Registers are temporary storage areas for instructions or data.

- Located inside of the Control Unit and ALU.
- Work under the direction of the control unit to accept, hold, and transfer instructions or data.
- High speed.

Certain registers have special roles:

- *Accumulator* – collects the result of computations.
- *Address register* – keeps track of where a given instruction or piece of data is stored in memory.
- *Storage register* – temporarily holds data taken from or about to be sent to RAM.
- *General-purpose register* – miscellaneous functions.





## Cache Memory

Cache memory is a specialized high-speed, high-performance memory that is integrated in the CPU.

Because it is fast and efficient, recent program instructions are stored in cache so that can be executed quickly.

Cache memory is expensive, so most computers have a small cache.





## Main Memory

Programs to be executed by the computer are placed in main memory.

The CPU fetches each instruction in turn from memory and executes it.

Main memory is fast and limited in capacity.

The CPU can *only* directly access information in main memory.

Main memory consists of a series of locations, each of which is associated with a numerical address by which it can be accessed.





## Paging

Recall that secondary storage includes hard drives, floppy disks, flash media, etc.

Sometimes, there is not enough free memory to hold an entire program.

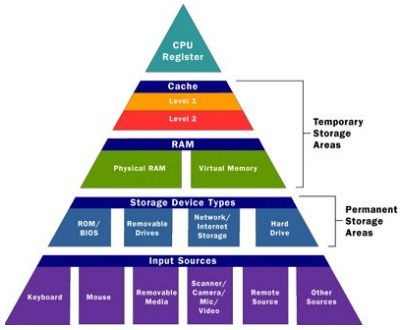
But information on external memory can only be accessed by the CPU if it is first transferred to main memory.

To remedy this, a computer uses *paging* – a system whereby blocks of a program are transferred into main memory. All blocks are the same size, and are referred to as *pages*.





## Memory Heirarchy



## System Bus

Recall that all communication between the major components of the computer occurs via the system bus.

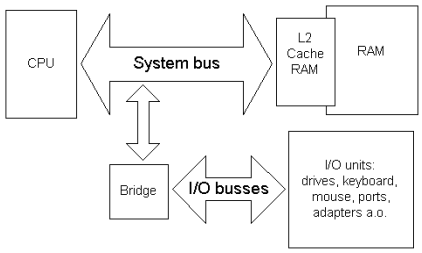
As a program is executed, each instruction is transferred from main memory to the CPU via the bus.

If any input or output are required, data to and from peripheral devices also travels along the bus.





## System Bus








## Power-On Self-Test (POST)

When a computer is first started, it reads instructions from a read-only memory chip (ROM).

These instructions check the hardware, to ensure that it is working properly, and thet all necessary components are connected.

The system also checks the system clock, to ensure that timing works correctly.

Once all tests have passed, the computer moves on to another chip.





## Basic Input/Output System (BIOS)

After POST, the system reads instructions from the BIOS.

These provide the computer with instructions on how to communicate with system devices, how to process basic data, etc.

BIOS is typically stored on a programmable chip on the motherboard, but older systems may have non-programmable chips that use ROM (Read-Only Memory).





## Operating Systems and Applications

Once the computer has successfully POSTed, and has completed performing any necessary instructions specified by the BIOS, the computer launches its *operating system*.

An operating system is a collection of programs that allow the user to perform certain tasks.

In the early days of personal computers, the operating system was a Command Line Interface (CLI), in which the user typed commands.

Nowadays, most operating systems use Graphical User Interfaces (GUIs), which offer text, graphics, animations, etc.

Major operating systems today are Windows, Mac OS, Linux, and Android, but there are many more than the ones listed here.