

Varinkas notes on Computer Architecture

Computer architecture defines how hardware and software interact to execute instructions. The basic structure includes three main parts - CPU, memory, and an input/output system. The CPU, or the Central processing unit, performs calculations and controls data flow. Memory stores data and instructions, while I/O systems handle communication with the outside world, humans particularly.

Inside the CPU there are 3 main units - the Arithmetic Logic Unit (ALU), the Control Unit (CU), and registers. The ALU does the math and logic, The CU conducts the execution of instructions, registers hold small, temporary pieces of quick data. Little buffers. Does that remind you of RAM? We're gonna talk about this later.

Now say a homo sapiens pressed a key. The signal travels through wires, hits the mother board and eventually lands at the CPU. The CU picks it up and figures out what the signal means in machine language, it basically extracts an instruction out of the signal. Then it sends that instruction to the ALU if something needs to be calculated, or to registers if it just needs to hold data for a moment. The result goes back through the system, out to memory, and eventually ends up as a pixel or letter on your screen.

So now we know what CPU units do. Now let's see memory. Memory isn't a one thing either. There's a whole hierarchy. You remember the registers? They are the fastest. Next comes cache (slightly larger but slower cells). It keeps copies of data the CPU might need soon. Then comes the RAM. The storer

of active programs and data while the computer is on. Finally there's SSD/hard drives/ where information stays even after power-off.

Usually computers don't run single atomic instructions, they run tasks - groups of instructions bundled together, to do something meaningful. When a task is running, the CPU doesn't take the whole thing at once. It takes small chunks of instructions from memory, pulls them into cache or registers and executes them step by step.

Modern CPU's can also do multiple instructions at the same time. This is called parallelism. because it's its name. Some of this happens inside the core / different parts of the CPU working on different pieces of the same task. And some of it happens across multiple cores: several tasks running side by side.

So to wrap it up: a computer needs the Control Unit, ALU and registers to compute. Memory stores instructions and data, feeding them in chunks to the CPU. Modern features like parallelism make things faster. Add in RAM, cache, storage and you have a full system designed to take electricity and turn it into logic. To serve humans. To calculate.