Information Technology for Digital Humanities

Lecture 3

Mario Verdicchio

Università degli Studi di Bergamo

Academic Year 2023-2024

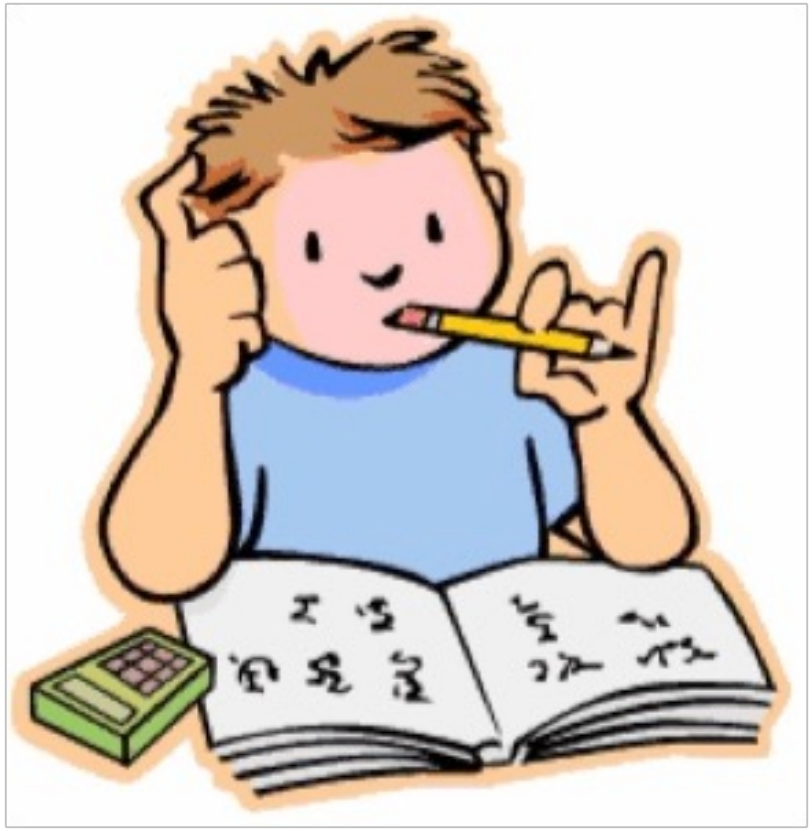
Lecture 3 (October 3 2023)

- Fundamental concepts: technology
  - Previous lecture: software
  - Now: hardware

Now let's go inside the computer to study its main components and how they work.

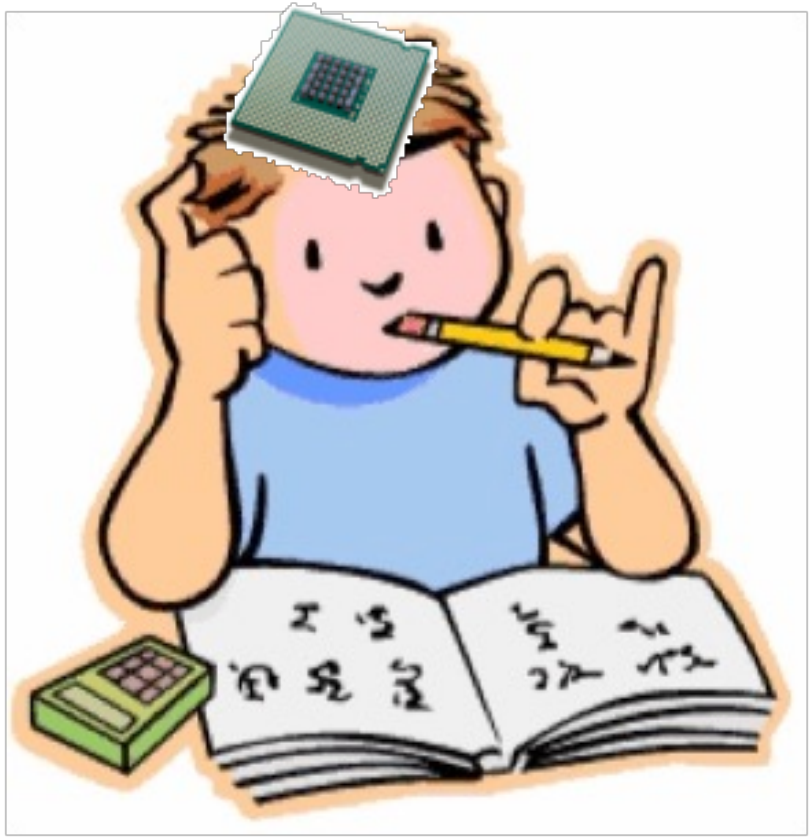
Hardware

Components of a computer



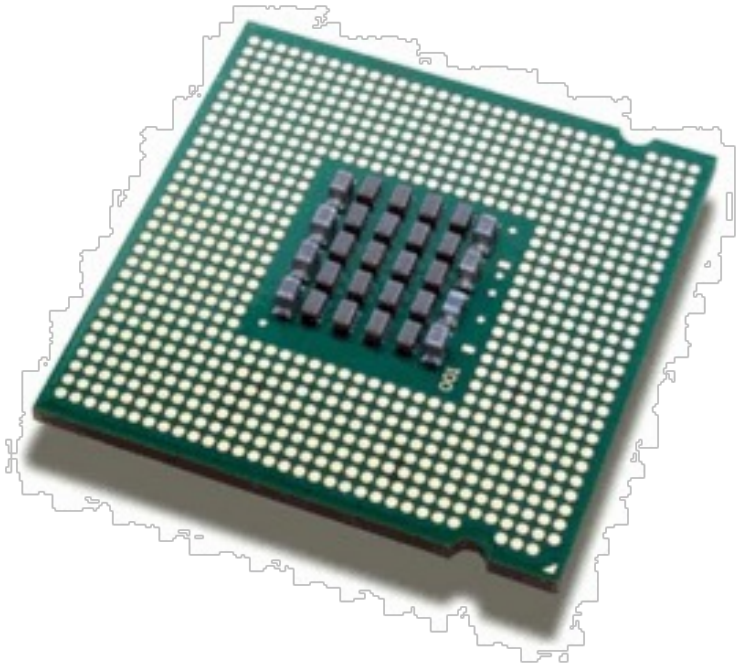
We will illustrate the components of a computer by comparison with what a human being uses to work on data.

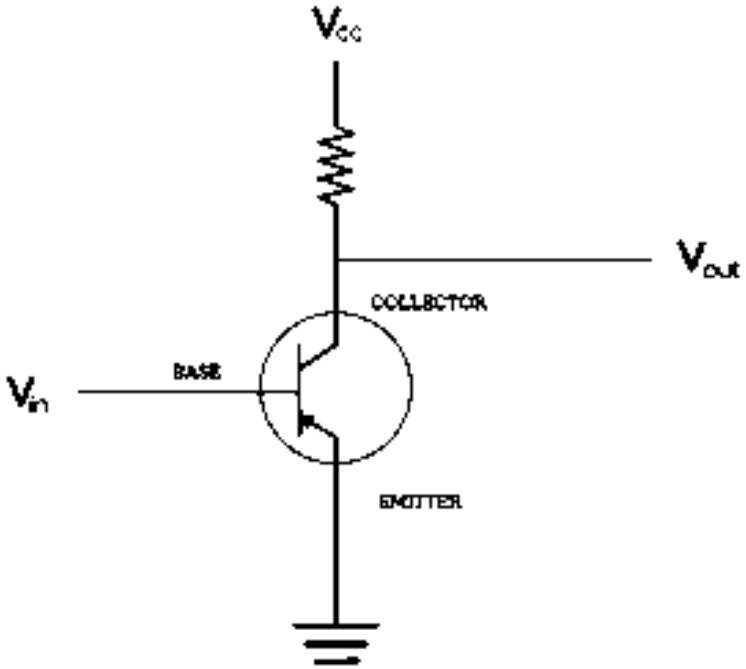
The most important component is the CPU (Central Processing Unit). In this human/computer comparison, the CPU corresponds to the brain, because it is the part of the computer that performs operations.



CPU

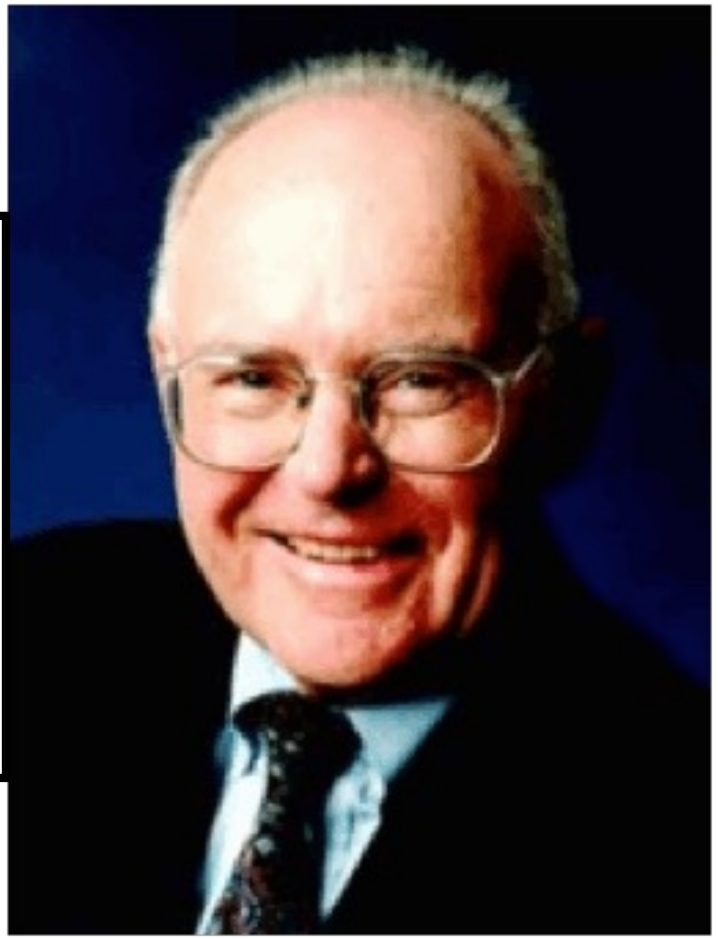
The CPU is an electronic circuit made up of a significant number of transistors. Each transistor is built in a way that it reacts to electrical input stimuli (V_{in} in the figure) with a specific output electrical voltage (V_{out} in the figure). When we say that the CPU performs operations, we mean that it receives electrical input signals that represent the data and indicate the operation to be performed.





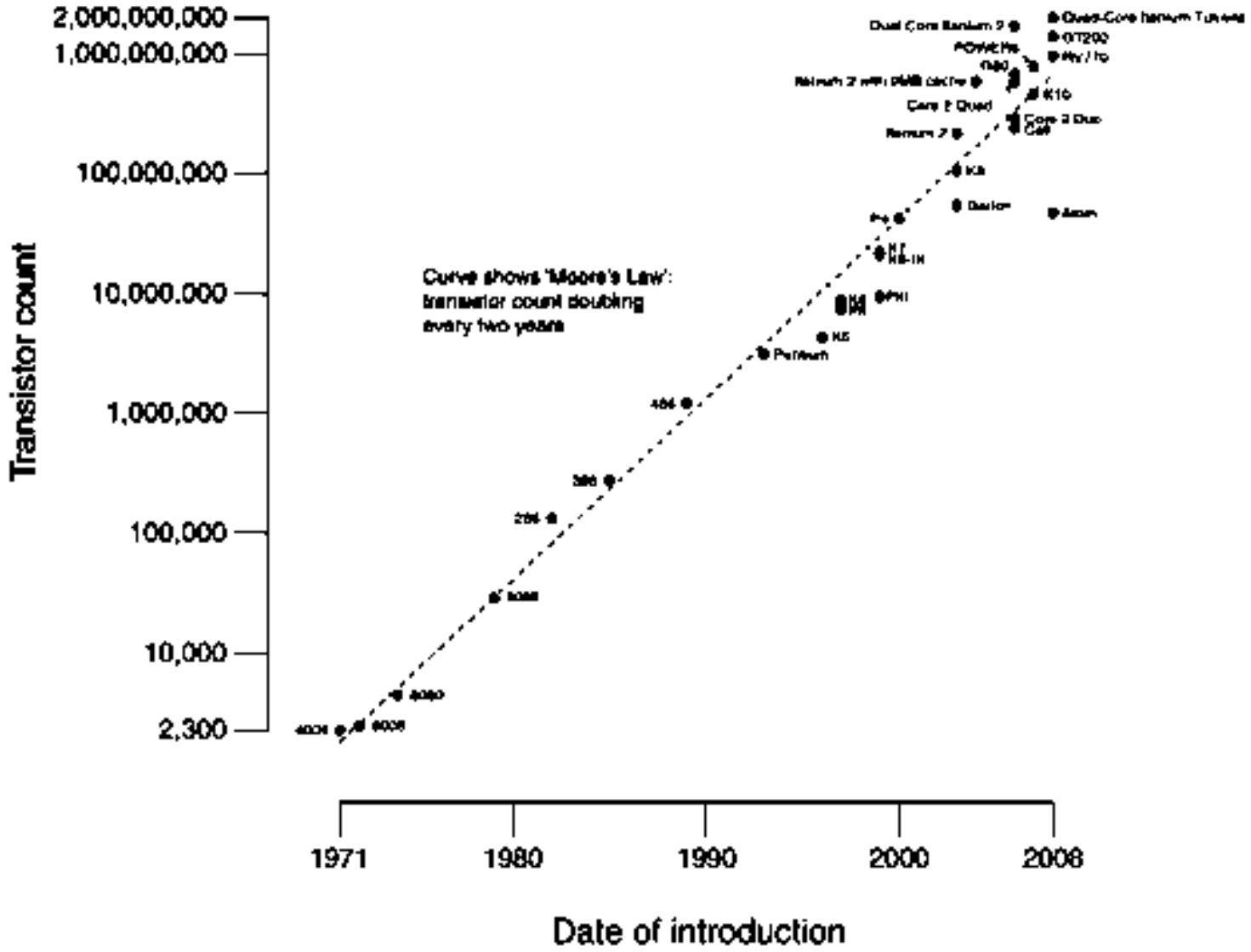
The CPU transistors react to stimuli and output other electrical signals, which represent the result of the operations performed.

The transistors in a CPU are very numerous (today they reach 2 billion per processor), and have continued to grow in number since the first CPUs were built. The trend is curiously constant, and predicts a doubling every year and a half. The first to notice this trend was Gordon Moore, co-founder of Intel, one of the largest CPU manufacturing companies. This phenomenon, known as “Moore's law”, is named after him.



Gordon E. Moore

CPU Transistor Counts 1971-2008 & Moore's Law

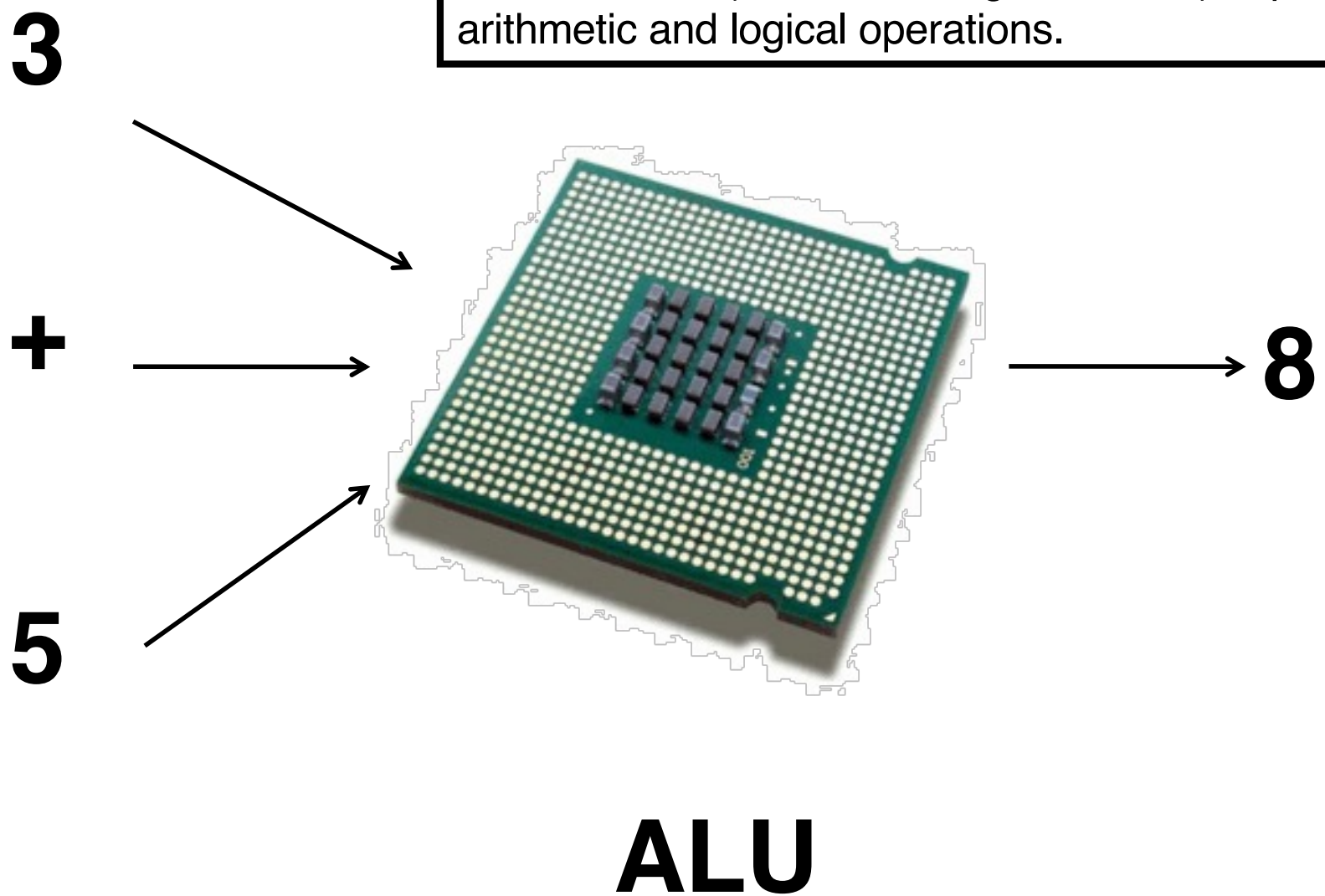


Despite some adjustments in the following decades, the pattern detected by Moore seems to hold. What can we make of this? There are at least two observations that should prevent us from jumping to apocalyptic or utopistic conclusions about the future of IT.

Firstly, there are physical limits to hardware given by the laws of physics all material entities are subject to. It is true that transistors can become smaller and smaller, but they cannot be smaller than one atom. The curve of the number of transistors per chip can be modeled as an exponential, but there is a cap. More generally, we must not take a mathematical model as a realistic depiction in all its parts, so even if Information Technology has shown exceptional growth in the past decades, this may not be the case in the future.

Secondly, we must not forget about the quantity vs quality (signs vs meanings; syntax vs semantics) dichotomy: an increase in transistor density surely leads to faster computing machines, which in turn means that a greater quantity of computational operations can be carried out per unit of time, but this does not entail that certain tasks will become amenable to machine operation.
There is a distinction between unfeasible and impossible tasks: an unfeasible task is one for which there is a computational solution, but it requires so many computational resources that it is not reasonable to tackle it; an impossible task is one for which there is no (known) computational solution.
Breaking a cryptography-based protection is currently unfeasible, but it may become much easier once quantum computing, i.e. computation exploiting quantum mechanics phenomena, becomes available thanks to a technological breakthrough.
Computing what happens in our brains when we understand a word is, instead, an impossible task, since we do not know how consciousness is produced in the brain, nor whether that mechanism can be obtained via computation. Increasing the number of operations that a computing machine performs in a unit of time will not change this.

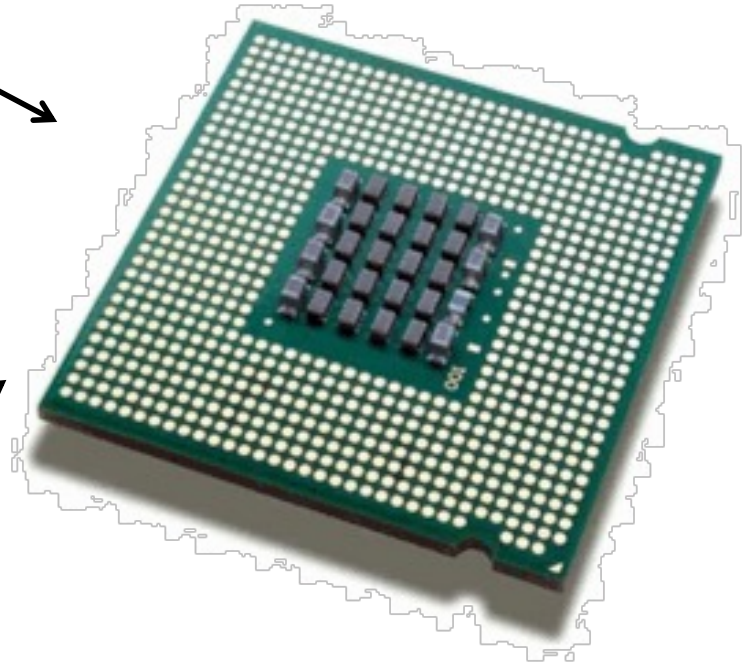
The CPU is actually not a unitary block, but we can distinguish two parts based on their function.
The ALU (Arithmetic-Logical Unit) performs arithmetic and logical operations.



ALU

The CU (Control Unit) is the part that controls the transfer of data from one part of the computer to another. In fact, everything that happens in the computer happens because electrical signals have been emitted by the CU: even the input of data into the ALU, for example, is controlled by the CU which executes the instructions written in the programs given to the computer.

print



“ciao”

CU

But where do the data that are processed by the CPU come from?

3

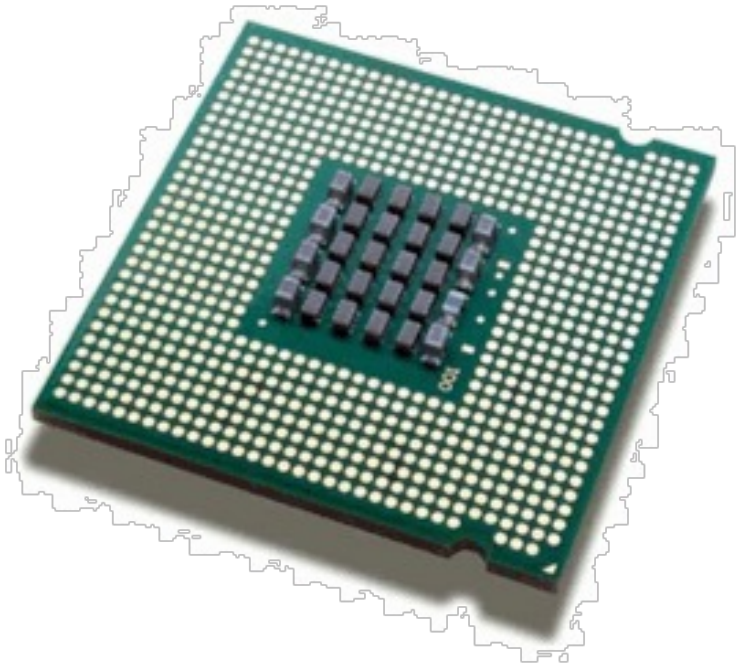
print

+

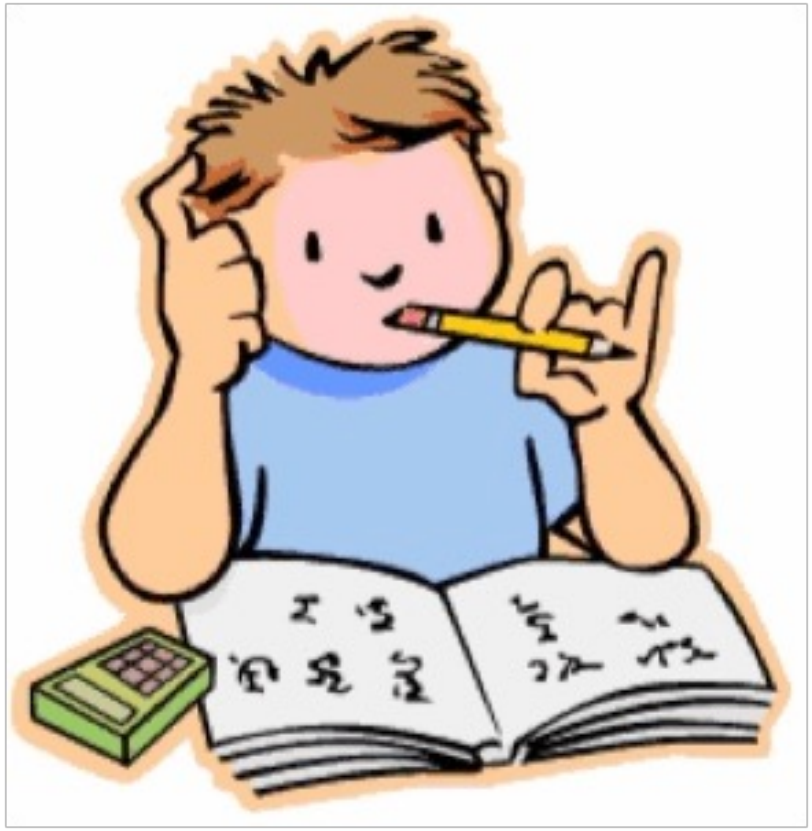
5

“ciao”

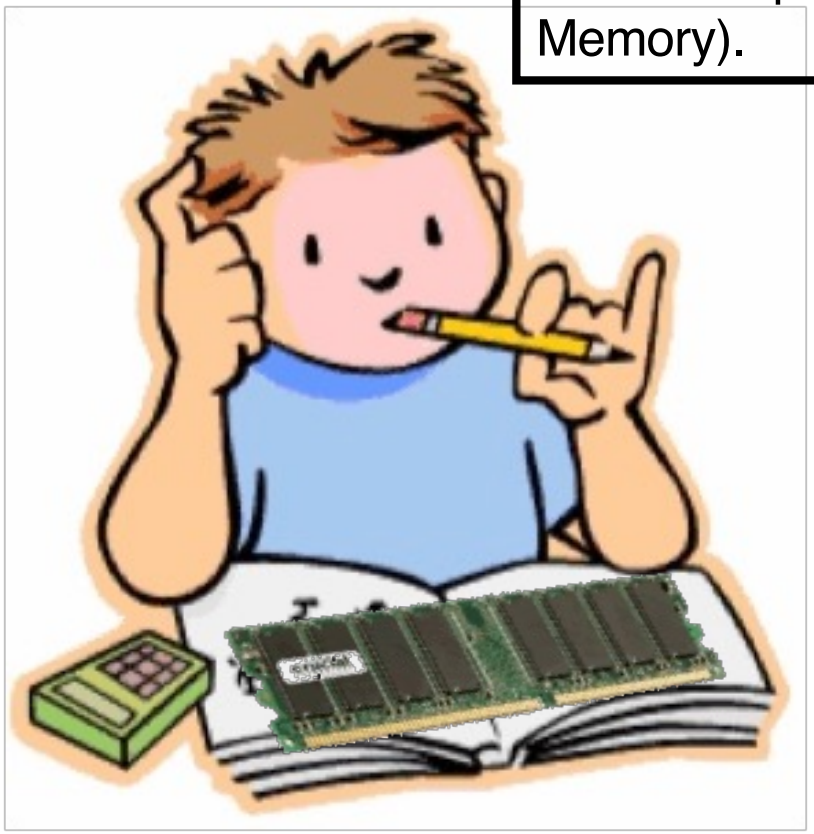
?



Let's go back to the comparison with the person who is processing data.
The data of a problem are written in their notebook.

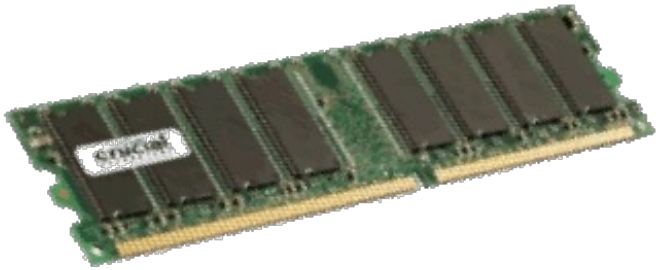


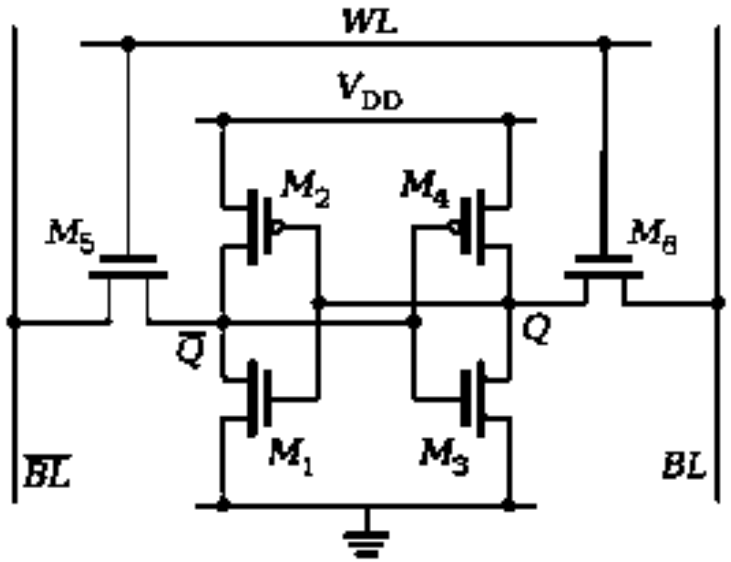
Here is another component of the computer, which corresponds to the notebook from which the person takes the data, and in which they write down the intermediate steps and the final results.
This component is called RAM (Random Access Memory).



RAM

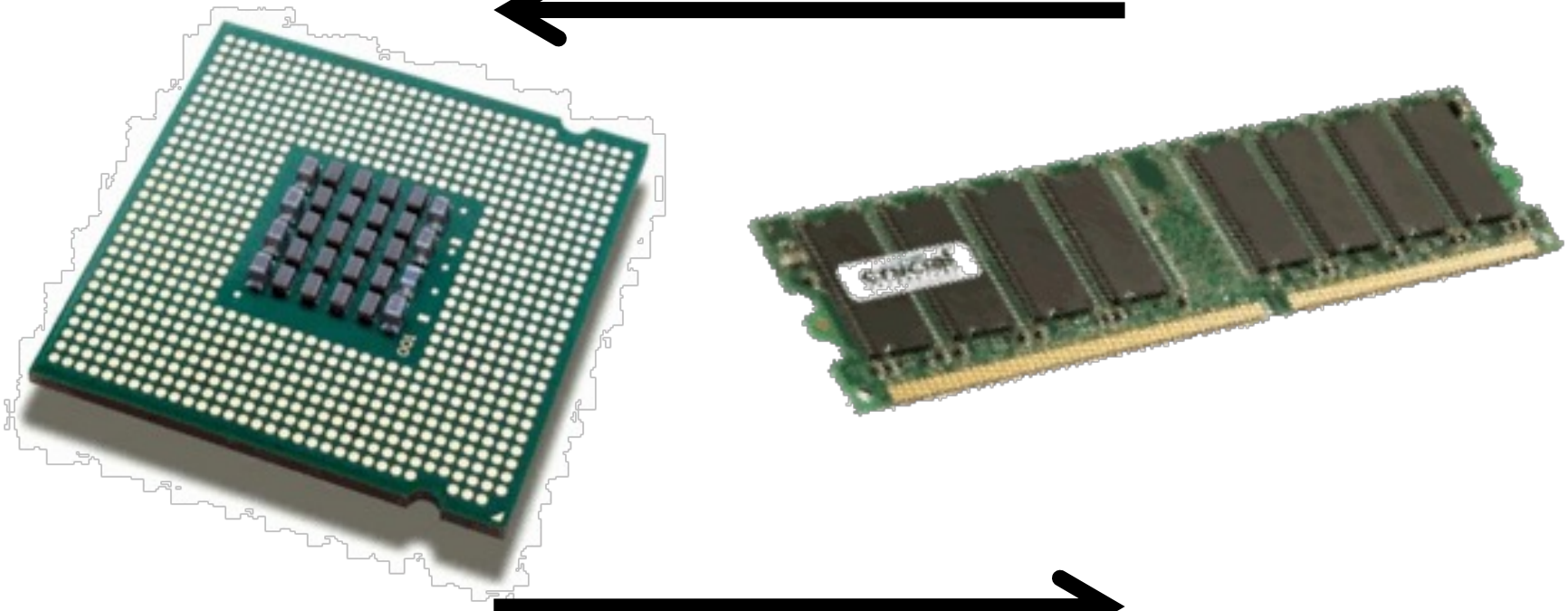
The RAM is also made up of electronic circuits with numerous transistors. The configuration appears more complex because the purpose of the RAM is not to respond to electrical stimuli by sending output signals, but to store data internally. Building a memory requires a more complex composition of transistors.





The name “RAM (Random Access Memory)” is meant to highlight a specific characteristic of this device: it allows access to data contained in any position of the device. You can choose a random access point, and the RAM is able to provide the data contained at that point. The name in Italian (“memoria centrale”), instead, highlights the central role played by the RAM inside the computer: the RAM works very closely with the CPU.

operations, operands



results

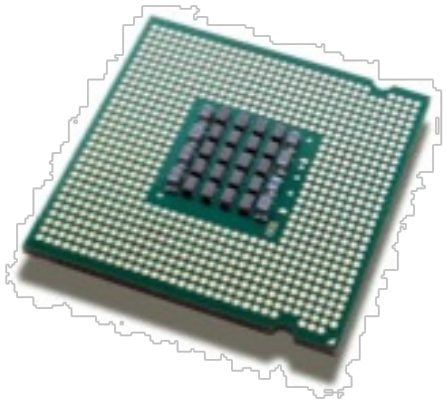
The operands and the operations that the CPU must process come from the RAM.
The results obtained in output from the CPU are, in turn, sent to and stored in the RAM.

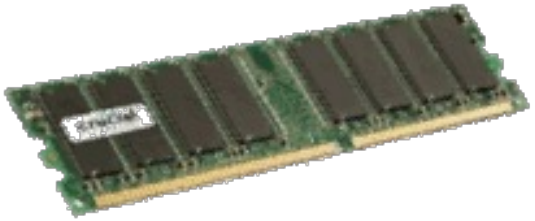
Let's now see the other components that participate in the computer's operations: clearly with only the CPU and RAM we cannot display the result of an operation on the screen.

3

+

5



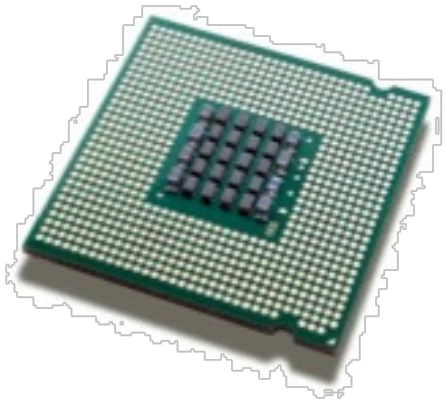


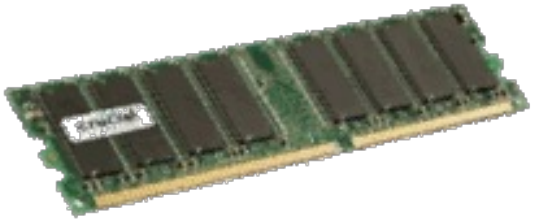
We said that the data and operations to be performed are stored in RAM.

3

+

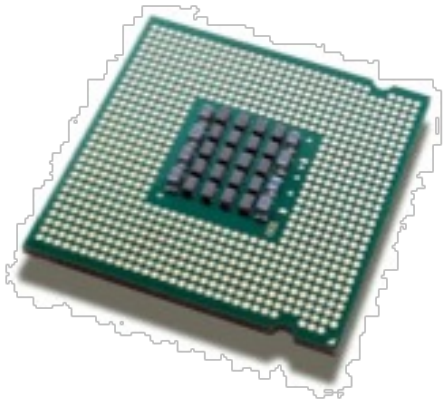
5

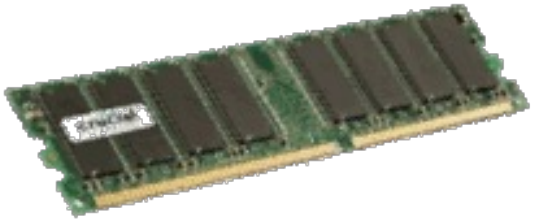




From the RAM, they are transferred to the CPU, by order of the CPU itself (or rather, of the CU) which takes instructions to be executed from the RAM continuously, from the moment the computer is turned on up to when it is switched off.

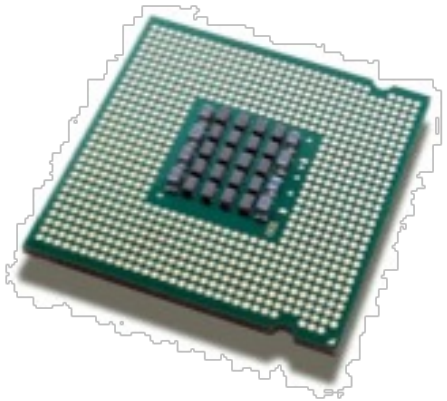
8

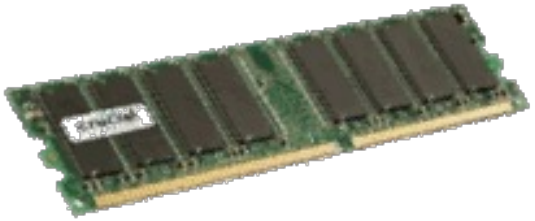




Once the output result from the ALU is obtained, it is sent to the RAM...

8

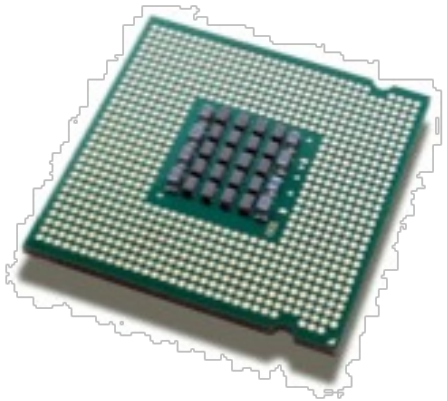
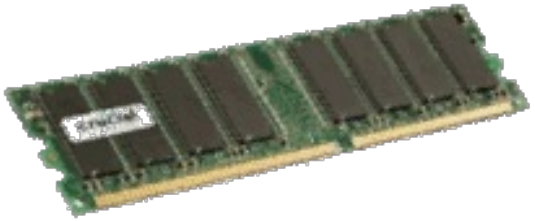




…where it is stored.

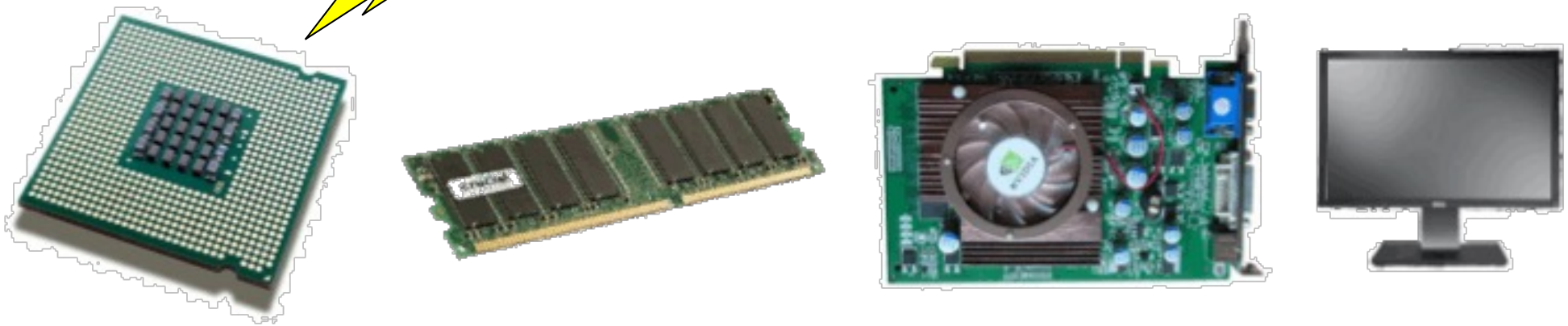
⟵ print

8

The CPU continues to take instructions from the RAM and let's suppose that the next instruction commands a printout of the result of the previously performed operation on the screen.

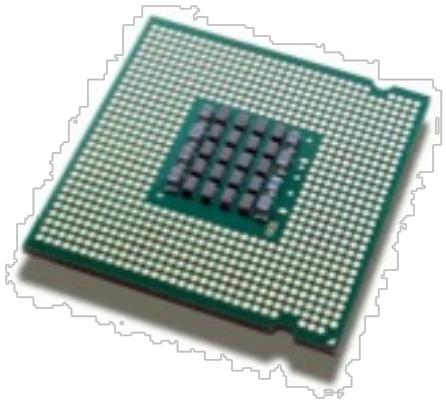
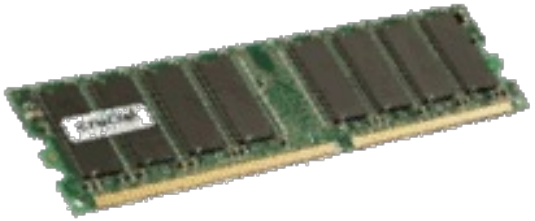
8



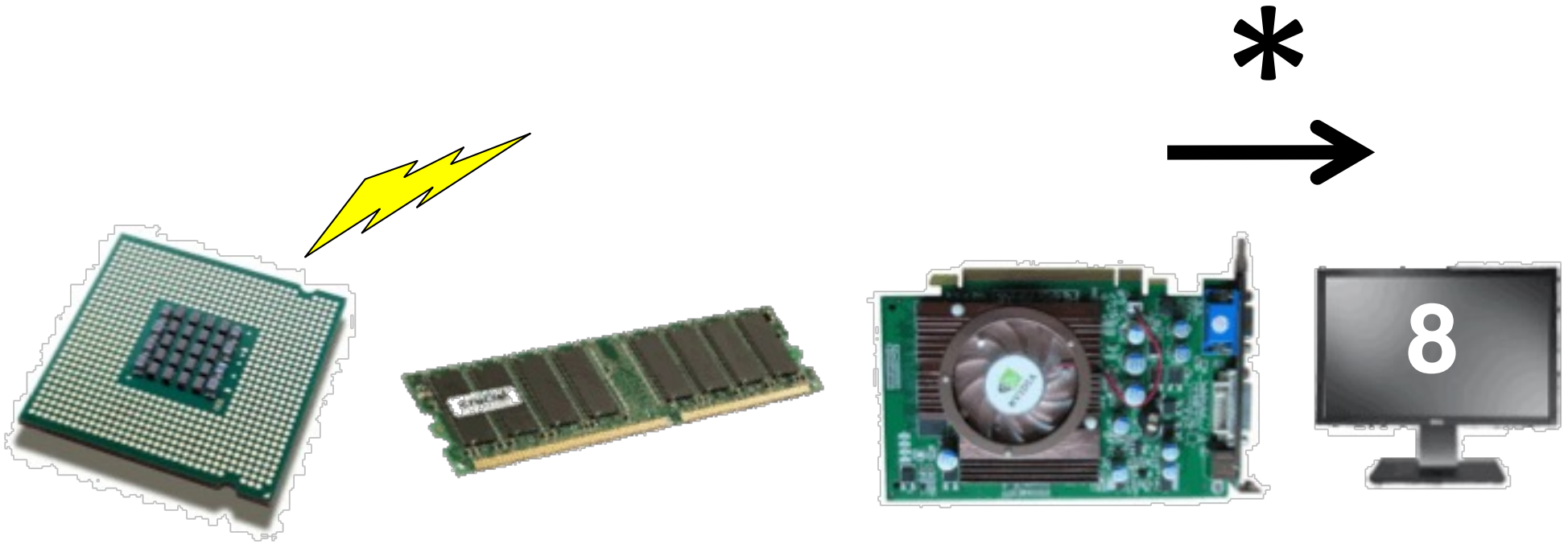
The CU in the CPU receives the command and sends appropriate signals to carry out the transfer of the result from the the RAM to the screen.

Graphics card

→ 8

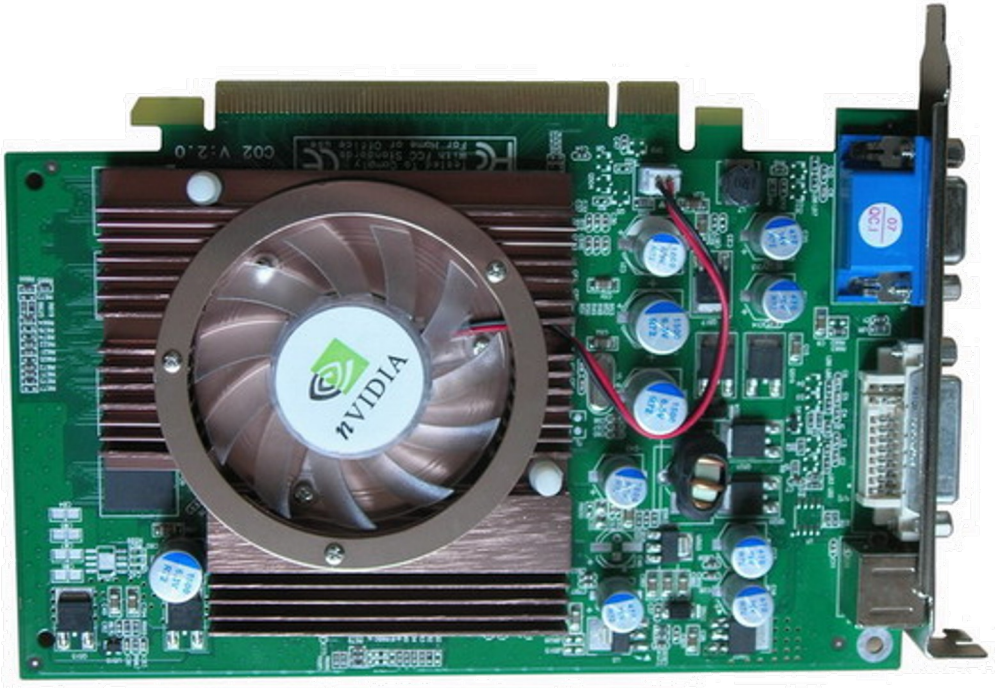



The signals from the CU ensure that the data is transferred from the RAM to the component that manages the signals to be sent to the screen: the graphics card...

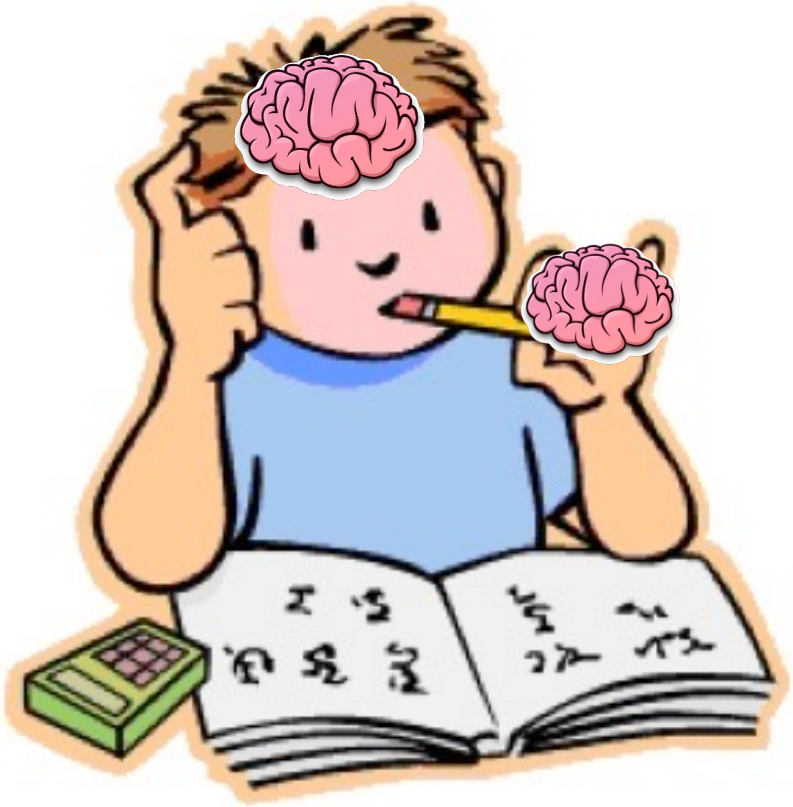


…and from the graphics card to the screen. This last transfer is more complex than the others because we have to shift from the data inside a computer to physical phenomena suitable for the human beings in front of the screen: no longer electrical signals, but rather light waves that represent a sign.

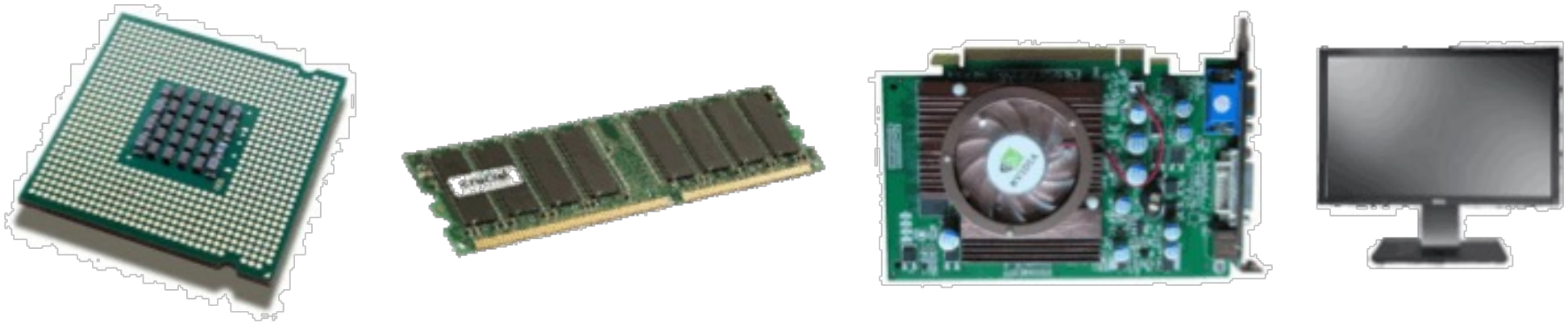
The graphics card shows that it has a fan. In reality, even the CPU is always accompanied by a fan, which turns on as soon as the computational work raises the temperature of the circuits, bringing it closer to an alert threshold. Without ventilation, each processor would melt due to its own activity (moving electrons from one side to the other produces heat). In fact, the graphics card has its own processor, the GPU (Graphical Processing Unit), to which the central CPU delegates the execution of instructions related to display of data on the screen. The CPU can handle displaying a simple '8' easily, but in the case of more demanding graphics applications such as electronic games, the CPU relies on the GPU to be able to work on other things.



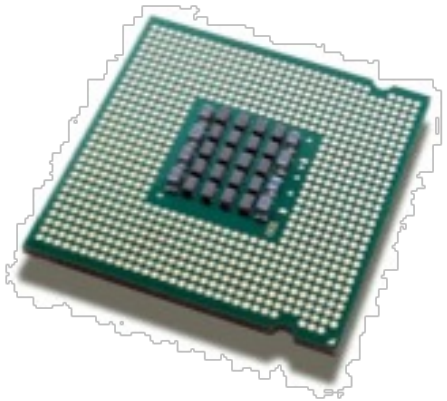
GPU

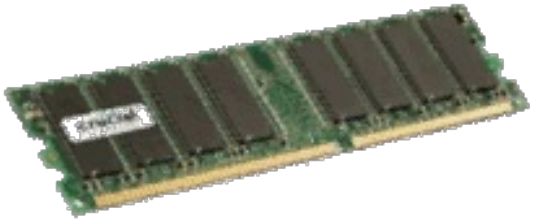


It is a bit as if the human’s hand had a small local brain that guides it in writing, without engaging the main brain.



There is another reason to place an asterisk on the transfer of data from the graphics card to the screen.





This is the first transfer that leaves the confinement of the box that encloses the computer's internal components.

Peripheral

The external components of the computer are called peripherals.
The screen is, in fact, a computer peripheral.

Input peripheals

Input peripherals are those that allow the flow of information from the outside to the inside of the computer.

Output peripherals

Output peripherals, vice versa, are those that allow the flow of information from inside the computer to the outside.

Did you notice that the Wii controller was in both peripheral groups? This is because there are peripherals that allow the flow of information in both directions. In the case of the Wii controller, it allows the user to move the cursor in the Wii menu or control the game (input), but, in certain game situations or when the cursor passes over a selectable menu item, it vibrates, providing information to the user (output).

In the following slide, the Wii controller is shown with other peripherals that are both input and output.

Wii controller (Nintendo)

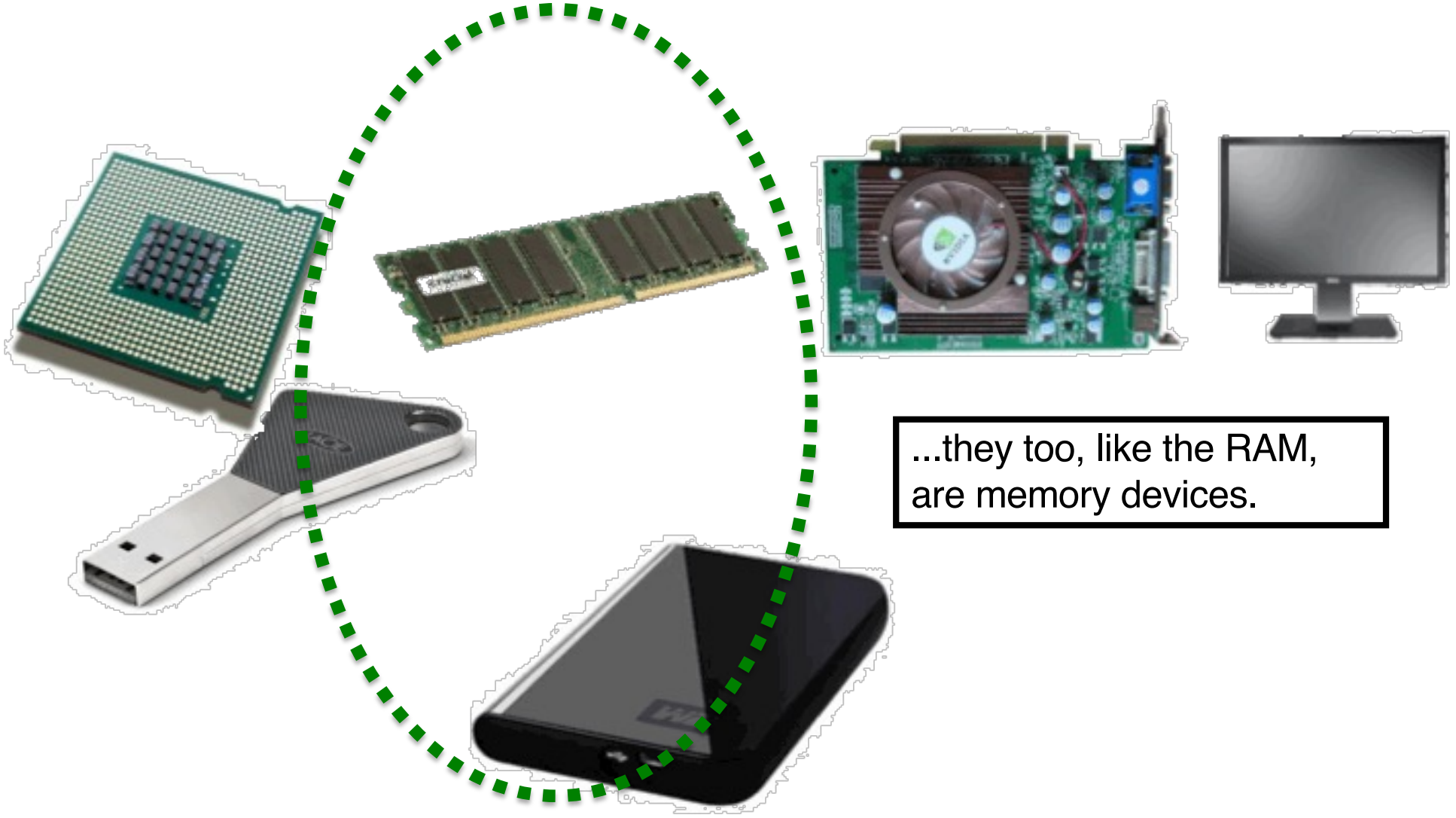
PS3 Dual Shock controller (Sony)

USB stick (Lacie)

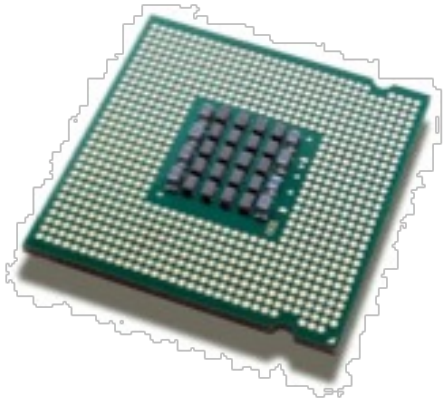
External Hard Disk (WD)

iPhone touch screen (Apple)

In particular, let's focus on the USB stick and the external hard disk...



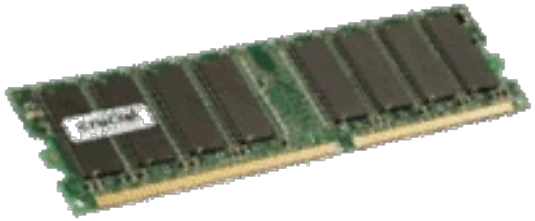
...they too, like the RAM,
are memory devices.



HD

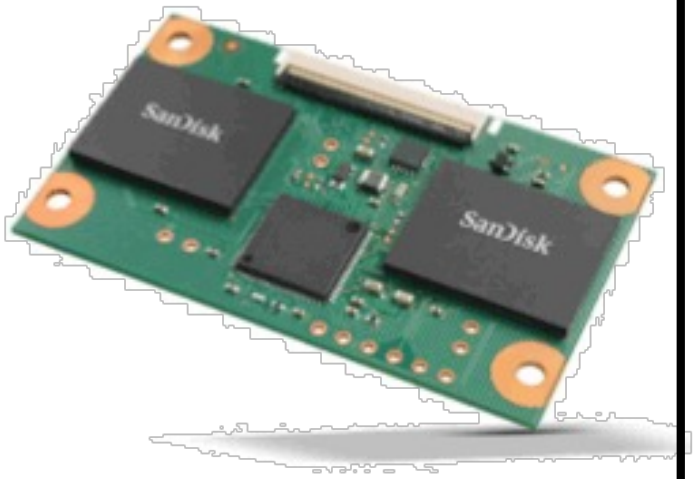
There is another memory device internal to the computer: the hard disk, alkso known as the hard drive.

Let's compare the RAM with the HD. They have complementary characteristics that make them both necessary for the proper functioning of a computer.

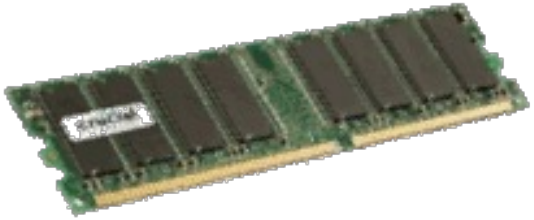


It is an electronic device made of transistors.	It is a magnetic device made of disks and a read-and-write head.
It is volatile: it retains its stored contents only while it is powered by electricity. Once the computer is turned off, it loses its contents without electricity	It stores data in the form of oriented magnetic domains in its disks. Like magnets, they never run out, and retain information even without electricity. It is non-volatile.
It is very fast and therefore manages to work in sync with the CPU.	It is very slow, so it is used to retain information when the computer is turned off.

SSD



Recently, a new type of HD is available as an alternative to the classic magnetic HD, called SSD (Solid State Disk). It has the same non-volatility characteristic as the magnetic device, but it is based on electronic technology, therefore without mechanical parts (disks, heads) which can fail, and it is faster.



One wonders why the SSD is not proposed as an alternative to volatile RAM. In reality, the SSD, like other non-volatile electronic memories (USB sticks, "Secure Digital" SD cards typical of digital cameras), although faster than magnetic HD, are still much slower than RAM, and not they would be able to work in sync with the CPU (which can work at frequencies of 3 GHz, i.e. perform 3 billion operations per second).

Magnetic HDs still remain a widespread solution for storing information with the computer swicthed off, because it is definitely the most convenient from an economic point of view. SSD technology is relatively new, and still very expensive.

In fact, the HD is where all the information must be stored before the computer is turned off. When you “save” a document, what you are doing is ordering it to be stored on the HD. The HD is the origin of the data and instructions that go from the RAM to the CPU. First of all they are transferred from HD to RAM. This is what happens when you turn on a computer: you have to wait for some time because this transfer is happening.

All components of a computer are connected to each other through a channel that carries electrical signals representing data from the CPU to the various parts of the computer. The channel is called bus. When you connect a peripheral to your computer (a printer or a “Universal Serial Bus” USB stick), you are actually connecting it to the bus to make it part of the system.