Mathematics of the Digital World

Attila Egri-Nagy

v17.09.05

How do computers work? What makes them fast? What makes digital communication possible? How come that we can send text messages, images to remote places almost instantly in a reliable and secure way? These questions have several answers depending on the context: in mathematics, in engineering, or in physics. The mathematical answers are the easiest to understand, due to their abstract nature. Here is a brief overview of digital computation and communication, mentioning the relevant parts of Mathematics and some historical references on the side.

Our world is increasingly digital: computing devices mediate our communication, facilitate our learning and thinking, and store our memories. They changed how we work and our entertainment. Behind all the diverse applications, there is a simple idea.

We can break down information processing tasks into steps so tiny that dumb physical processes can carry them out.

Historically, this decomposition of some of our thought processes took some time. Arithmetic calculation was the first to be automated, but it was initially done with a less suitable basic unit: the decimal digit. While the 10 digits $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ are perfectly suitable to carry out numerical computations, the physical realization is cumbersome. The best attempt is the cogwheel with ten teeth, where the rotations of the wheel represented addition and subtraction, depending on the direction.

Difference Engine by Charles Babbage
polynomial functions
method of finite differences

Since we mainly use decimal numbers for representing quantities and doing calculations, we thought for long that arithmetic is the only possible usage of numbers. However, they are universal in terms of descriptive power. We can associate numbers to letters, then a sequence of numbers can represent text. They can encode colours and their shades, so numbers can describe images as well. Numerical values can also be interpreted as frequencies, thus a soundwave can be represented by numbers.

number systems: decimal, binary, hexadecimal

What is the minimal number of symbols that still allow efficient calculations? Two symbols are enough, for example 0 and 1, or an answer to a yes-or-no question.

The smallest unit of information is the binary digit, the bit.

We can describe any number by zeros and ones, by the so-called binary representation. Since numbers are already universal, we can describe any piece of information with a sequence of bits.

The biggest advantage of the binary representation if that its physical realization is the easiest. We need to distinguish between two physical states only, e.g. high-voltage and low-voltage in electricity.

We showed that any type of information can be written as (a possibly very long) sequence of bits, like this.
11000100 01101110 00101100 01100110 00100110 10011011 00111101
11100111 10000001 11101111 00101011 00001010 11111100 11111111
10101111 11111010
But what do these bits mean? It depends on what they encode, i.e. their meaning is determined by a set of rules how the bits should be interpreted. The bits can stand for numbers, and thus they can encode text through the codes of the letters and punctuation symbols. Another simple interpretation is to have zeros denoting white pixels and ones black pixels. Thus, dividing the bit sequence into equal sized rows, we can encode a black and white image. Sometimes these rules are used to hide the content and we talk about encryption. The encoding rules can also protect the content, by correcting errors in the bit sequence.

digital geometry

cryptography
number theory
probability & statistics

coding theory
combinatorics
linear algebra

The most important aspect of the encoding rules that give meaning to bit sequences is that they can also encode operations. We can store numbers 19 and 23 in binary representation, and we can use their bits to add them together, there is no need convert them to their decimal form. The computer can add binary numbers directly. Again, this is true in general, for any type of data. Information processing is simply just 'flipping' bits.

A computer processes information by changing bits of a sequence of bits (its memory content).

This bit flipping is of course extremely organized and systematical: it mainly follows a hierarchical structure.

Understanding how computers work is about this organization of how computers are built. The first level is organized by logic. We can look at our everyday usage of our natural language, make it more precise. In our everyday thinking, using the logical connectives (like 'AND', 'OR' and negation) is rather different from doing arithmetic. But they happen to be the same. An exciting part of the theory of computers is to see how complicated operations are built of very simple building blocks.

set theory
propositional logic
logic gates & circuit design

Summary: in order to automate, we break information processing down to bit flips, then in reverse, we interpret bits on higher and higher levels.