A NEW EXTENSION OF BIOINFORMATICAL THEORY AND THE PRINCIPLES OF BIOCOMPUTATION

Author: István Szentesi (Institute for Staregic Research)

Bioinformatical theory

In laying the foundations of information theory in the 40's of the 20th century, Claude Shannon, Warren Weaver, and John von Neumann also determined the direction of progress in the field. Partly because of the simultaneous emergence of a number of parallel research initiatives, this perspective also set the general trend of the emerging disciplines of computation and computer science in that the initially pursued analog approaches were soon given up in favor of thinking along the lines of a digital philosophy. A-mong other things, major technological circumstances justified this shift. Simultaneously with the researchers mentioned above, Warren McCulloch and Walter Pitts presented a model offering formal description for the behavior of neural networks. It was the pressure of these first three simultaneous initiatives that primarily acted on ongoing research, forcing it, then as well as today, to follow the track that still predominantly determines how informatics and computation science work both in terms of the frontlines of basic research and in the widest range of applications.

In 1931 Kurt Gödel used f o r m a l argumentation to define the limitations of formal mathematics. Starting out of Gödel's incompleteness theorem, Alan Turing shifted the focus of his studies from the examination of axiomatic systems to first clarifying the concept of algorithms and then to studying algorithms themselves /Turing's Machine/. In the meantime, Neumann recommended – not very fortunately – that Shannon use the term „entropy" for „the extent of uncertainty" within the field of information theory. The subtile of Norbert Wiener's 1948 book *Cybernetics* - „Control and communication in the world of animals and machines" – is also fairly thought provoking. Finally, the following thought in the closing chapter /"The brain does not use the language of mathematics"/ of Neumann's last book, *The Computer and the Brain* may be seen as a theoretical synthesis of this trend of critical scientific thinking: „Thus the outward forms of our mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language truly used by the cent-ral nervous system is." Having written these lines half a century ago, Neumann proposed a germ of a thought and a prospective direction of research

that, at the time, only he could have carried through, thereby showing the right way to others as well.

Insummary: We have still not offered a full exploration, one built on the widest possible basis, of an exact information theory that, in all probability, combines analog and digital principles and that defines the control principles of the genuine operation, behavior, and communication of the /more or less/ complex systems and networks of systems found in living beings, including us, humans. In other words, we still do not have a bioinformatical theory.

The extension of bioinformatical theory

In his book *Molecular Information Processing in the Nervous System*, Michael Conrad offers descriptions for processes taking place in the nervous system, each of such descriptions being basically models built on specific algorithms. Conrad's constructions, which are suitable for describing more tangible operations, are often labeled „non-algorithmic" or „con-formationally changing". This mostly goes back to epistemological reasons. The partial results of the calculation process specify the calculation scheme to be used in the next step, which therefore implies that what we have hereis indeed an algorithm after all. In his 1989 publication *The Brain - Machine Disanalogy*, Conrad emphasizes that the differences between the brain and the computer are more fundamental than their similarities. He thinks that the brain's ability for evolution and the programmability of computers are a pair of concepts that are complementary to one another.

In 1990, Theodore Roszak writes: „...First it was biolo-gists who borrowed from cybernetics in order to explain genetics as a mechanism of information transfer. Now it is computer scientists that borrow from biology to prove that data processing technologies are evolutionary in nature. Such is human culture: it often progresses by one area of thought borrowing from another certain metaphors perceived to offer meaningful insights. Beyond a certain point, however, metaphoral explanations will simply lead to wrong thinking. This is the point where the metaphors stop offering meaningful insights and start operating in their verbatim meaning. Nevertheless, this is the type of superficial thinking that is so characteristic of certain computer experts, even our own thinking about technology issaturated with the feeling of human inferiority and existential failure..."

In his 2002 book A New Kind of Science, which indeed may easily become one of the key elements in the birth of exact bioinformatical theory, Stephen Wolfram offers a comprehensive and detailed algorithmic description of reality that creates an entirely new situation and sheds light on a plausible new synthesis.

It also gives an entirely new, unusual, and complete interpretation of the concept of interdisciplinarity, which also happens to be a vital element of our own approach. The fusion of the above-mentioned three platforms – Con-rad, Roszak, and Wolfram – may be the starting point from which the road to an extended bioinformatical theory may be paved.

The principles of biocomputatin From an analytical viewpoint, the foundations of biocomputation may only be laid down with a solid background in exact b ioinformatical theory. Intuitively, however, we can already formulate certain fundamental assumptions, which are as follows:

/1/ biocomputation organized on the basis of bioinformatical theory is, fundamentally, in essence, and in terms of its content, analog information processing, at the level of their realization, bioinformatical and biocomputation systems occurring in nature /that is, living systems, and especially the human being/ involve combined analog and digital information processing, in the case of the bioinformatical and biocomputationsystems occurring in nature /that is, living beings/ the primarily mechanism is genetic information processing /whether at the molecular and submolecular level/ while the nervous system and the brain are only „secondary" /consequential/ and as such, represent a type of interface, as a piece of plausible evidence, it should be taken considered that while all living beings have genetic information mechanisms, no nervous system and brain is present in prokaryotes, fungi, and plants, the genetic mechanisms of the bioinformatical and biocomputation systems occurring in nature /that is, living beings/ represent double genetic information processing that consists of a „digital" /coding /genomic region /as combined „self-programming software" and „quasi-hardware"/ and of an „analog" /non-coding/ genomic region /as „genuine" or „pure software"/, and actually contain the two functionally so different genomial fields as structurally integrated into each other, the „gene pools" of living systems – understood as the sets of genetic information processing mechanisms of the bioinformatical and biocomputation systems occurring in nature, that is, living beings – represent double genetic information processing: one is directed at „digital" /somatic/ and the other at „analog" /intellectual and mental/ endpoints.

/ 6 / bioinformatical and biocomputation information processing is, in and of itself, fundamentally analog in the living world: however, on the level of concrete realization in real systems, these, much as though they were „parasites", are inserted into the digital „hosts" in combined formations and with combined /digital or analog/ consequences – and not only at the level of basic consequences but also at the level of endpoints,

/7/ „artificial computer systems" utilizing bioinformatical and biocomputing information processing have quite a prospect, as the hardware may operate in a digital, the interfaces in a combined analog and digital, and the software in an analog organizational mode: and as far as their performance is

concerned, their efficiency may achieve, or potentially exceed, that of the human brain.

Preliminary possibilities of testing /bioinformatical theory and the principles of biocomputation/

As far as we currently know, biological evolution on Planet Earth goes back to a history of some four billion years. George G. Simpson estimates that ever since life has been present on the globe, some 500 million species may have made their appearance on the stage of evolution. Based on up-to-date estimates, the species currently in existence and known to science /approximately 2.5 million species/ and the ones currently in probable existence but not yet known to science /approximately 1.5 million species/ amount to a total of four million, which – although it may be less than one percent of the total number of species Simpson estimates to have ever appeared in the course of the evolution of life on Earth – is still an enormously great number. It should also not be forgotten that the pros and cons of the gre-at „genetic experiments" of the other 496 million species are still around – as concentrated into the total pool of genetic material currently to be found in the four million species now in existence. This is our bioinformatical and biocomputing reservoir.

At the end of 1984, the A l t a   S u m m i t set a new objective that was to become known as the Human Genome Initiative. Upon the conclusion of the first stage of the consequent Human Genome Project in the summer of 2000, today we have what is considered the first version of the base series of the human genome, which, approximately, consists of 3 billion nucleotides. Almost 97 % of the nearly completed human genomic DNA base sequence includes „non-coding" sequences that are often termed „genetic scrap". Almost exclusively, this „evolutionary trash" consists of various so-called „repetitive sequences" and so-called „intronic sequences" spliced from the „coding sequences" during transcription and /finally/ so-called „specific unique sequences with unknown functions".

The repetitive sequences, intronic regions and unique sequences comprising this 97 % of the total human DNA matter may be interpreted as a „distorted" /degenerated/ linear sequence of digital signals where /"A" complementary „T" = 1/ and /"G" complementary „C" = 0/. Transformed into an „n"-dimensional structure, this linear sequence of digital signal is to be treated as an one-dimensional Turing's Machine that serves as the primary information base of the higher-level operation of the nervous system of Homo sapiens, while also implying the species' genetic evolutionary history going back to billions of years.

The first preliminary testing of the bioinformatical theory and of the biocomputation principle, currently both in their embryonic stage of development, might commence by analyzing the computer database of the Human Genome Project through the development and application of special

purpose-built software making it possible to study the significant differences between „non-coding” and „coding” genome compartments in detail. Such software development and application may be expec-ted to have two central elements:

/ 1 / treating a complete haploid „non-coding” genome compartment as an intact and homogeneos entity and describing the analog code function of the genome compartment by way of fractal representation,

/ 2 / treating the fractal representation of the analog code function of a complete haploid „non-coding” genome compartment as an intact and homogeneous entity and completely decoding, in an analog way, the fractal representation by way of holographic imaging.

Once / 1 / and / 2 / have been achieved successfully, an attempt may be made on the basis of the results to construct a type of „primitive” biocomputer in a computing and biological laboratory, and to commence its elementary testing. This biological information technology project would involve the cooperation of mathematicians, information technology experts and molecular and population geneticists, and rely heavily on the application of high-capacity /”cluster”/ super computers and high level molecular genetical laboratory technical possibilities too.

F i n a l l y : An attempt may be made to offer analog versus digital transformational descriptions of the bioinformatical theory and biocomputation principle in their relation to „orthodox” informatics and to develop the expected/potential equivalence principle between the bioinformatical theory /biocomputing principle and quantum computation.

The derived databases define the operating principles of the primary information network of human consciousness /PAI = Powerful Artificial Intelligence/.

I have also paid due attention to THE POTENTIAL STRATEGIC IMPORTANCE of the bioinformatical theory and of the biocomputation principle.