BIO / COM*: Mixing Water with Fire

Dennis Tsichritzis

GMD-German National Research Center for Information Technology
Schloss Birlinghoven, D-53754 Sankt Augustin

1. Introduction

As we have reached a new century and a new millenium it is appropriate to comment on the past and predict the future. Undoubtedly the past century was shaped by material products and industrialization. Physics and Chemistry gave the foundation for new products that changed completely the economic activity and our lives. The scientific results of the nineteenth century were expanded and applied through engineering to create an industrial society. We influenced our environment at an unprecedented scale both for the good by providing technological wonders and the bad by creating pollution and using technology for war. Will the new century be just a continuation? There are already many signs pointing to a new direction. Material goods consumption is reaching its limits. In addition, it is debatable if the world can go on indefinitely consuming as it does today. Economic expansion needs, however, new products and new markets. Sustainable development points in that direction. What will be the basis in the future for scientific and economic development? In the next sections we will argue that BIO (Biology, Biotechnology) and COM (Computer, Communication) will provide that basis not only in isolation but in combination.

It is already widely accepted that the future economic development will be based on services and not so much on material products. Terms such as Information Society and Knowledge Society point in that direction. Computers and Communications have become indispensable in our daily lives. Finally, the expansion of the Internet and the stock market valuation of its related companies show that at least the investors are betting heavily on such a future. COM related services, companies and jobs are expanding like fire.

At the same time our lives are subject to another revolution. Biology is making tremendous progress. It is chartering new territory at an unprecedented scale. Biotechnology is already delivering economically significant results. Pharmacology and medicine are already beginning to reap the benefits of our increasing ability to understand life itself. BIO is the essence, the water, of this knowledge. As the application of science and technology increasingly changes our environment it is vital to our survival and well-being for us to understand how we function ourselves.

It is perhaps controversial but not so surprising to claim that the future will be shaped by these two areas, BIO and COM. Many scientific and general articles have already discussed these issues. The discussions, however, run in parallel as if these two areas are not only independent but are somehow competing in claiming the

* BIO stands for Biology, Biochemistry, Biotechnology, etc.
COM stands for Computer, Communication, Computability, Complexity, etc.

B. Wangler, L. Bergman (Eds.): CAiSE 2000, LNCS 1789, pp. 1-6, 2000

future. We believe that on the contrary many future developments will be based on their combination.

At first glance everything separates these areas. Biology studies life, its diversity of material and organizational forms, their natural evolution and interdependencies. Computers and Communications are based on abstractions, completely virtual and rationally optimized to function in the artificial world of our information problems. Whereas Biology is predominantly experimental, Computer Science is more mathematical and abstract. The favorite support for materialization of information abstractions is crystalline silicon, a very dead substance known for its stability, whereas biological organisms continually reconstruct themselves from unstable molecules based on carbon. Biotechnology is using fluids, catalysts and processes. Computer Engineering is using the solid state, electronics and programming. Even in their applications they differ. The BIO economic sector and the COM sector seem to be disjoint, at least so far. Needless to say, university departments and research projects are also usually far apart. We claim that they should get close very quickly because they need each other.

2. Computational Biology (COM-BIO)

We first look at what COM can do for BIO. The customers are in Biology and Biotechnology. The methods and the tools come from computers and communications. Biology in the past, in particular molecular and developmental biology, has had to focus on unraveling specific mechanisms, small pieces from complex biological organisms and ecosystems. How the whole process was happening, and why, were problems too difficult to attack. Lately, however, a wealth of knowledge on entire organisms and processes down to the molecular level is becoming available which enable us to address the questions of how and why. There is an avalanche of facts, of structures, of models which have to be studied, organized and disseminated. In short, there is an emergence of extreme complexity which needs to be mastered. If we view computer science not in terms of technology but in terms of substance everything we do serves to master complexity. Because our computer systems have become extremely complex we introduced tools to deal with complexity. Similar tools can be used to deal with the complexity of Biological systems.

- Algorithms of different sorts can be combined to approximate and simulate complex biological phenomena.
- Data Bases of different kinds are needed to organize rapidly accumulated knowledge and to make this knowledge readily available.
- Data mining techniques can be very important.
- Visualization tools can be used to explain structures and grasp behavior.
- Modeling concepts can be applied to abstract from known facts and produce general theories.
- Communication Networks and Digital Libraries can be introduced to support the wild expansion of scientific activity and economic development.

Take as an example bioinformatics which develops algorithms for the structural analysis of proteins and molecular complexes involving proteins. These methods can be applied to two very important problems in the pharmaceutical industry. The first problem is finding appropriate target proteins for the design of new drugs. These proteins have to be selected from a large set of hundreds or thousands of proteins turned up by high-tech genomics currently being developed. The second is to find the right drug binding to a given target protein. The key to solving both problems are pattern matching methods coming from computer science and statistics and being carefully taylored to this challenging application.

Information Technology and methodology can also help to develop a Theory of Biology. If the structure of proteins are known, their interaction is explained and their cycles of metabolism understood we still need to combine all of these in some sort of abstract framework. We claim that many areas of Computer Science from Algorithms, to knowledge modeling to AI can be very useful to underpin such a theory. We hope that Computer Science can play a similar role for Biology as Mathematics plays in Physics. In return, the challenge of both new paradigms and enormous complexity will force us to develop even more and better theoretical tools. Both computer science and biology are grappling for a theory of complex design processes, and this will be essential both for future computer engineering and biotechnology.

Computer science is playing a significant role in our attempts to understand the cognitive processes of our minds. The major activity of the field of Artificial Intelligence is not to build an artificial brain but to understand the workings of intelligence and learning through the analysis of model constructs within the context of the conceivable. Computer science is also already playing a significant role in understanding the abstract self-reproducing and self-organizing nature of biological organisms. The field of Artificial Life deals with the analysis of model constructs to understand life. This approach has a tradition dating back to von Neumann. We don't strive to produce artificial life. We only hope to produce model artifacts which behave as life and which can be used to explain and predict life.

The overall hope and promise to bring biologists together with computer scientists is that using analytical and synthetical means in an interleaving, coordinated, and systematic way we may be able to sort out basic principles in the organization of living systems and the processes of life such that it covers more than one level of description. This certainly will rise the quality of theories about living systems.

If we had a Theory of Biological systems we can predict behavior. For example, from a protein solution, we can obtain its structure. From its structure we can make hypotheses about behavior. Going backward from abnormal behavior we could suggest causes and remedies.

3. Biological Computing (BIO-COM)

We turn our attention on what BIO can do for COM. The customers and the applications are in Computing and Communications. The tools and the methods come from Biology and Biotechnology. Computers and Communications evolved over the years and are currently mainly based on digital electronic technology. In the future we may need to exploit novel ways of computing using perhaps quantum, optical or biological computing. In this paper we will concentrate on biological computing which has already been demonstrated in principle.

Beginning in the early '80s, physical molecular biologists began to build machines to deal with natural information processing in biological organisms. DNA sequencing machines, DNA synthesis machines and DNA amplification machines have become widespread. Even evolution machines were constructed. In the '90s, in the quest for ever higher levels of integration and following a market pressure, such machines have migrated to a silicon technology with extensive overlaps with the electronics industry. DNA Chips with arrays of different sequences bound to the surface of wafers are an example of such an approach

Biological communicators have also been studied. Directed diffusion and signal amplification is the usual mode of biological communication at the molecular level. The directionality is achieved mainly by selective transport across membranes or selective attachment to static or dynamic structures. The global signaling starting with antibodies in the clonal response of the immune system to foreign substances provides an impressive communication network, components of which are already finding application in biotechnology (e.g. antibody chips). The role of transport proteins in cells is getting increasing attention. Cell receptors, phosphorylation cascades, electron transfer reactions and intracellular calcium waves provide further examples of complex biomolecular signaling.

If we achieve substantial progress in biological computing and communication it does not necessarily imply that we will use these technologies in the same way as modern Computers and Communications are used. We do not expect that biological computers will control elevators or that we will base Internet on biological sea communications. There are many applications in controlling biological phenomena where we see tremendous possibilities. There we can directly interface real life biological processes with artificial biological computers and communications. The applications can be endless from medicine to pollution control. We need to concentrate on the applications that nature does better and avoid those that we already know how to do well with electronic computers. It is probably a mistake to try to prove the superiority of biological computers by using them to solve complex mathematical problems. Maybe they will never be as good as their electronic counterparts.

The most interesting, influence of biological computing and communication is potentially in the areas of Theory and Programming of computers.

The Theory of Computation evolved this century in a couple of phases before settling down to its current stable version. In the beginning there was an attempt to define computability in a constructive manner. Starting with computations we know (like arithmetic) and using composition and iteration we obtained some notion of computability (primitive recursive functions). It proved to be too narrow of a scope although much of what is computed is of that kind. The widely accepted notion of computability was obtained in a different way. Models of different kinds were defined (Church, Turing, Post) and then proved to be of equal power. Church then declared in his thesis that the informal notion of computability corresponds to the formal notion defined by any of these models. Many problems were proved to be not computable in their general from. In restricted form they are computed every day. The study of algorithms gave rise to different notions of complexity in time and space. A wide class of difficult problems were defined as NP complete, that is we can compute them in reasonable time (polynomial) if we could deal with non determinism (a form of uncontrolled parallelism). Specific constrained NP complete problems are of course computed often but we still cannot handle the non determinism in general (i.e. for large problems).

Biological computing can provide other paradigms which could unsettle the current state of computability and complexity theory. First, the constructive approach did not go far enough because only limited forms of combinations were used. Biological computing offers naturally, evolution, replication, even mutation which could prove very successful. The constructive way of defining computability will get a new boost. Second, the interaction between models was strictly controlled. A Turing Machine could only work on what was on its tape. A Turing machine with so called "oracles" could do more. Biological computing offers a paradigm of a variety of unrestricted interactions which may expand the scope of what is effectively computable. The so called Heuristics which were always used to solve real problems in practice could possibly be inside the model. Third, non determinism in the form of uncontrolled parallelism is built-in in biological computing. This implies that non-determinism is not simulated but it is effectively computed. The implications can be that a host of very hard problems could be effectively computable.

The biomolecular paradigm will also influence programming. Internet programming is already far removed from early programming techniques. It is component based, running in a distributed fashion on an architecture which is not even known a priori. Biological computing programming will be even more flexible. In what ways will we be able to influence and externally control Biological Computing? We are still in the initial phases.

We can provide input (for instance in the form of biomolecular sequence libraries) in a known environment and we can read the corresponding output or send it as input to a further such processor. A stored program approach can be designed using a dataflow-like hardware architecture, produced using microstructured silicon, with the program specified by a sequence of photomasks which determine the spatial location of specific biomolecules in the dataflow architecture. Aspects of message passing and interprocess communication can also be realized in this approach but clearly we are at a very early stage in this development. The programming languages in Biological computing may be similar or very different from what we have today. Biological computing will force us to redefine our notions of programming or reactivate old ones which are not in wide use.

In these new programming paradigms replication and evolution will play a very significant role. Biomolecular systems process enormous amounts of information in parallel while mutually constructing themselves. When specific subsequences on RNA are translated to make components of the translation apparatus or polymerases which can copy RNA, we perceive a partly programmed and partly autonomous processing of sequence information to make self-assembling construction components. This is a bootstrapping process involving both hardware (molecules and reactions) and software (specific sequences and folding algorithms). A tight coupling of information processing and construction was foreseen as a vital aspect of computer science by John von Neumann. Up until recently however, the technology has caused computers to be built with a complete temporal and spatial separation of hardware construction and information processing. Now self-reproducing biomolecular systems are revealing exciting new paradigms for massively parallel, self-repairing, robust, adaptive and constructive computation which break down this distinction between hardware and software and between a computer and its execution.

Biological organisms are so perfectly adapted to their environment as a result of an iterative optimization process, rather than of directly programmed design. Evolutionary optimization is a key feature of such biomolecular information processing. Molecular evolution can be harnessed to solve large difficult problems where physical simulation in software is intractable. Biomolecular systems occuring in biological organisms have evolved sophisticated ways of selecting fleeting structural patterns, to be converted into permanent structures which themselves become triggers for further structural pattern selection. Recently this basic mechanism has been harnessed to allow artificial *in vitro* self-replicating systems to be constructed. Adleman's demonstration that externally specified optimization problems could also be programmed and solved has now encouraged an exchange of ideas between computer scientists and molecular biologists working on molecular selection systems.

4. Concluding Remarks

We claimed that not only BIO and COM will shape the next century but their mutual support and cooperation will open new perspectives for both areas. We see already a flurry of activity but we also notice many barriers.

First, Biology and Computer Science are very far apart and there are too few scientists grasping both. Second, most Research organizations are concentrating on one or the other but seldom mix both. Third, Research Budgets are usually separate and combination projects are looked with suspicion in both areas. Fourth, each area has a wild expansion and it is natural for each one to be self centered. Fifth, senior persons in both areas may recognize the tremendous possibilities but there is massive inertia in their Institutions and Societies. Finally, companies and economic sectors are far apart. Biotechnology is being pushed by the Chemical and Pharmaceutical sector while Computers and Communications have their own sector, the Information Technology economic sector.

The most severe difficulties are in terms of scientific language. The two communities talk different languages. They initially don't understand each other. It is worthwhile to overcome all these barriers, Science seems to develop faster at the boundaries of areas rather than at their core. We expect BIO / COM to expand like fire and be as essential as water. The expected results can have major economic and societal implications. Mixing water with fire will be difficult. It is, however, worthwhile.