A Grand Challenge for Computing: Full Reactive Modeling of a Multi-Cellular Animal

(Position paper)

David Harel

Faculty of Mathematics and Computer Science
The Weizmann Institute of Science, Rehovot, Israel

(With many thanks for help from J. Hubbard, M. Stern, N. Kam and S. Efroni)

Submitted to the UK Workshop on Grand Challenges in Computing Research
October 5, 2002

One of the central issues in software and system engineering over the last couple of decades has been to develop languages, methods and tools for the reliable construction of *reactive systems*. This term, coined in 1985, denotes systems whose complexity stems not necessarily from complicated computation but from complicated reactivity over time. Such systems are most often highly concurrent and time-intensive, and exhibit hybrid behavior that is predominantly discrete in nature but has continuous aspects too. Their structure consists of many interacting, often distributed, components. Very often the structure itself is dynamic, with its components being repeatedly created and destroyed during the system's life span. The heart of the problem is the need for good approaches to modeling and analyzing the dynamic behavior of such systems.

The most widely used frameworks for developing such systems feature *visual formalisms*, which are both graphically intuitive and mathematically rigorous, supported by powerful tools that enable full model executability, linkable to GUIs and other structural descriptions of the system under development. This enables realistic simulation prior to actual implementation. At present, such languages and tools --- often based on the *object-oriented* paradigm --- are being strengthened by verification modules, making it possible not only to execute and simulate the system models (test and observe) but also to verify dynamic properties thereof (prove). They are also being linked to tools for dealing with the system's continuous aspects in a full hybrid fashion. Most of the tools are state-based, encouraging intra-object style specification, but there are also recent scenario-based approaches, which enable executable inter-object specification.

The Grand Challenge proposed here has three premises: (i) that these currently used frameworks represent a reasonable solution to the problem of reactive system design, so that the area can be said to have reached a stage where it is satisfactorily based on a number of acceptable and workable ideas; (ii) that biological research is preparing for an extremely significant transition from analysis (reducing experimental observations to elementary building blocks) to synthesis (integrating the parts into a comprehensive whole), a transition that requires mathematics and computation; and (iii) that the true complexity of biological systems, specifically multi-cellular living organisms, stems from their reactivity.

As far as the third of these goes, biological systems exhibit the characteristics of reactive systems remarkably, and on many levels; from the molecular, via the cellular, and all the way up to organs, full organisms, and even entire populations. It doesn't take much to observe within such systems the heavy concurrency, the event-driven discrete nature of the behavior, the chain-reactions and cause-effect phenomena, the time-dependent patterns, etc.

From these three premises there follows a thesis: that biological systems can be modeled beneficially as reactive systems, using languages and tools developed for the construction of man-made systems. In two items of previous work, one on T-cell reactivity (the cellular level) and one on T-cell behavior in the thymus (uniform multi-cellular level), we have addressed the feasibility of this thesis on a very small scale and in a preliminary fashion, and the results are very encouraging.

In contrast, the sole use of differential equations --- the main vehicle of classical control theory and dynamical systems --- to model biological phenomena, is limited by the continuous, non-discrete nature of DEs, and the difficulty of comprehending even small sets thereof. Differential equations (or any kind of equations, for that matter) are also unfit as a medium for the biologists themselves to carry out parts of the modeling.

The Grand Challenge we propose is to model a full multi-cellular animal as a reactive system.

We actually have a particular organism in mind, the *Caenorhabditis elegans* nematode worm, better known simply as *C. elegans*, a suggestion that is in line with the extraordinarily insightful 30-year old proposal of Sydney Brenner, a venerated UK scientist, who chose this organism to challenge biologists with the task of discovering the entire development and neurobiology of a living creature. (For this proposal and the tremendously influential work that followed it, Brenner and two others received the 2002 Nobel prize.) C. elegans is complex, but very well-defined in terms of anatomy and genetics. In fact, its anatomy is better defined than any other multicellular organism, and it is the only creature for which the entire somatic cell lineage is known from the one-cell embryo to the adult worm. This was possible since the worm is transparent, its cell lineage is invariant (i.e., does not change from worm to worm), and the total cell number is relatively small (959 for the hermaphrodite and 1031 for the male). Moreover the wiring diagram for its 302 neurons is known. Genetically, the genome is sequenced; the functions of many genes are known, and this data set is expanding rapidly.

We have recently started a project, joint with two teams of biologists who are experts on the relevant parts of C. elegans, to model the development of its egg-laying system. But the Challenge is to construct a full, true-to-all-known-facts 4-dimensional model of this worm (or of a comparable multi-cellular animal), which is easily extendable as new facts are discovered. The front end would be an anatomically correct, animated graphical rendition, tightly linked to a reactive system model of the development and behaviour of the creature. The model would be fully executable, flexible, interactive, comprehensive and comprehensible. It would enable full and realistic simulation of the worm's development and behaviour (the fourth dimension), which would help uncover gaps, correct errors, suggest new experiments and help predict unobserved phenomena. It would be zoomable, enabling easy switching between levels of detail, and allowing researchers to see and understand the organism and its behavior in ways not otherwise possible. The underlying computational framework would be not only rigorous and realistic, but would be set up in such a way that biologists would be able to enter new data themselves as it is discovered, and even plug in varying theses about aspects of behaviour that are not yet known, in order to see their effects.

Achieving this goal would require enormous amounts of work, both in setting up the computational framework --- with its reactive system modeling and execution engine, the analysis and verification tools, and, of course the animation mechanisms --- and in gathering the biological data, translating it into reactive system terminology and incorporating it into the model. This would have to involve many teams of researchers, of multi-faceted backgrounds on both sides of bioinformatics.

The result would be an unprecedented tool for the research community --- not only those working on the particular animal being modeled. It would allow us to ask questions and perform cyber experiments before taking them to the lab bench, and to compare the bench results with what we think we know. It would help confront open questions, which, because of their complexity, cannot be addressed by standard laboratory techniques alone. "One of the beauties of mathematical modeling" writes one researcher, "is that it raises questions that may not have been addressed before". A model of this depth and flexibility would also serve as a basis for comparisons not only within a species, but with evolutionarily related species that have different forms and behaviors.

Also, if this challenge can be met for a worm, for which we have lots of information, it could actually tell us a lot about what we would need to make reasonable models for other less-defined organisms (much in the way that the genome project used the worm as a test case for sequencing the human genome).

Obviously, this kind of grand-scale work will impact not only biology, but also computer science, by revealing limitations of the computational tools and motivating further development, in order to modify, extend and adapt them for such use.

We feel that research will be going very much this way in the future: grand collaborations between computing --- including the particularly tantalizing coordination of reactive system modeling and animation --- with sophisticated biology.

A 1996 quote by Bruce Alberts, president of the US National Academy of Sciences, is particularly apt:

"When I stand back and consider what I have learned from the study of C. elegans, what do I remember most? … It is as if there were a time-counting 'computer' inside each cell, that inputs present events, stores them in memory along with the inputs from past events, and then performs the calculations that make the cell behave appropriately with regard to its subsequent behavior."

"Worms are so much easier to study than humans…. We can say with confidence that the fastest and most efficient way of acquiring an understanding of ourselves is to devote an enormous effort to trying to understand such relatively 'simple' organisms".