Wonderful

GC 1

In Vivo ⇔ In Silico:

High fidelity reactive modelling of development and behaviour in plants and animals

A Grand Challenge for Computer Science

Ronan Sleep[123]

Computers play an increasingly dominant role in the process by which the natural Scientist records, explores and models natural phenomena. This has potentially enormous consequences particularly for the Life Sciences, as they move from their traditional largely descriptive role to one involving accurate modelling and prediction[4]:

- There is a growing mass of biological data, now computer accessible and viewable either statically or as movies.
- Theories are beginning to emerge that give causative explanations of this data and predictions of future observations.
- Many of these theories can be expressed directly as computer simulation programs - active models with both discrete and continuous abstractions or approximations.

We believe that the state of the art in Computing Science for specifying, modelling and realising complex systems has advanced sufficiently to realise fully detailed, accurate and predictive models of some of the most studied life forms used as models in biology, such as Aribidopsis, bakers yeast (S. Cerivisiae) or the Nematode worm (C. elegans). This would build on partial computer models that are already under development many laboratories, to create a complete, consistent, integrated representation of all that is known about a particular plant or animal. This representation should be accessible to

---

[1] School of Computing Sciences, University of East Anglia, Norwich NR47TJ, m.r.sleep@btconnect.com

[2] **Acknowledgements**: the proposal incorporates aspects of the following submissions to the Nov. 2002 Edinburgh Grand Challenge Workshop: *Full Reactive Modelling of a Multi-Cellular Animal* (David Harel of the Weizmann Institute) *Computational Gastrulation* (Ronan Sleep of UEA) *Cell Model for S. Cerevisiae* (Ross King of Aberystwyth and Ashwin Srinivasan of Oxford); *The Challenge for 21st century Computer Science is Biology* (Ray Paton of Liverpool; *Building a Synthetic sensory-motor system*; *The Virtual Organism and its Community* (Mike Holcombe of Sheffield); *How can we build human scale complex systems?* (Julian F. Miller and Catriona Kennedy of Birmingham); *Why Computer Scientists need to put the 'science' back into 'Computing Science'* (Ajit Narayanan of Exeter). Andrew Bangham, Robin Forrest, David May, Enrico Coen and John Fox engaged in interesting and helpful discussions. Tony Hoare gave insightful and rapid feedback throughout.

[3] **Apologies**: to those already engaged in this challenge who are not mentioned in this document. A proper background study is needed to identify you all.

[4] see for example BBSRC's vision document 'Towards Predictive Biology' (http://www.bbsrc.ac.uk/about/strategic_plan/bbsrc_vision.pdf).

humans via extensible view selection mechanisms that include the interaction modes possible between an experimenter and the real life form, and also between the life forms themselves. Thus for an animal model, such as C. elegans (the Nematode worm) it should be possible to model phenomena such as:

a. DEVELOPMENT from an initial fertilized cell to a full adult, at various resolution levels. An accurate model will respect knowledge about, for example: cell lineage, cell differentiation, cell lifetime, morphology, size and relation between major cellular sub-systems. Virtual experiments (e.g. moving a virtual cell during development, or making an incision) should lead to the same outcomes as real life.

b. CELL FUNCTION and INTERACTION: the specific functions of cells should be captured in appropriate detail together with principal modes of interaction.

c. MOTILITY and SENSORY aspects of behaviour: types of reaction to various stimuli, including neighbouring life forms; speed and nature of movement.

d. ENVIRONMENTAL INTERACTION: interactions between organisms and the surrounding environment should be captured.

The outcome of the Challenge will take the following forms:

a. Showpiece demonstrators for a small number of selected life forms. These demonstrators will be a medium of exchange between Computer Science and Biological approaches to complex systems, reflecting simultaneously the state of the art in both fields.

b. Joint publications of experiments describing the principles which have emerged from experience gained with the showpiece demonstrators, and evaluating their expressive and predictive power.

c. For the Biologists, new conceptual frameworks for representing and reasoning about complex developmental patterns will emerge, built from existing computer science theory modulated by detailed biological knowledge of living organisms. These frameworks will provide the fundamental tools necessary to turn an ever growing mountain of micro data about life forms into a trusted knowledge resource enabling us to build and test advanced theories of complex reactive systems. Bioinformatics is already exploiting many Computer Science techniques and tools (for example Perl, hidden Markov models) in genetics. Developmental models need more advanced data structures than simple strings, and consequently more advanced Computing tools and models. Elements of the necessary framework such as distributed process formalisms and rewriting systems already exist in Computer Science, and are already sufficiently successful at describing some aspect of living systems to offer a basis for modelling for complex reactive system development.

π-calc

d. Although it by no means certain that Computer Science must follow Biology in the way it does things, the Computer Scientists may take inspiration from Biology to construct new ways of specifying complex reactive systems that construct and maintain themselves from small initial and perhaps sketchy specifications. Perhaps we can uncover some fundamental system design principles which nature uses to realise an effective 'SYSTEM = NATURE + NURTURE' paradigm, creating a new generation of system design methodologies for complex adaptive self-maintaining systems. Theories and methods arising could give us important handles on large emerging complex systems such as the web..

e. Complex Object Design and Manufacture. From a computing viewpoint, we might consider that in realising living creatures which develop from a single cell through interactions with their environment, nature has developed some rather effective 3-D volumetric rendering algorithms. If we understand how nature does this, we can develop a new generation of design and manufacture processes in which a small partial specification grows into a complex object suited to its environment. This would have applications at many different scales from nanotechnology, where it could be used to orchestrate the development of complex nano components, to systems for managing the aircraft manufacture or shipbuilding. It is notable that such systems begin with a very small specification that grows dramatically, so much so that the weight of paper needed to describe the final system can outweigh the final product even in the case of a ship.

Scientific success of the project will be judged by:

- *The extent to which Biologists incorporate the modeling principles generated*: an early success would involve widespread use of at least some components of the demonstrator models. A major milestone would be reached when a series of in-silico predictions made with the demonstrator models were subsequently confirmed in-vivo.
- *The extent to which the Computer Scientists incorporate the principles in new approaches to designing and constructing complex systems.* An early success might involve deriving some key properties of an exemplar organism from its developmental parameters, for example the constancy (or otherwise) of the cell lineage. Software engineering may learn how to build software systems that can identify and adapt to changing user needs without never-ending surgical patching.. Nano-technology will benefit from a 'lazy specification' paradigm that leaves the system to work out at the last moment many of the details needed to develop in a specific environment. And such an approach may be essential if realistic volumetric representation of complex objects is to be realised in advanced, possibly holographic, display devices.

The coping stone of a successful challenge would be a generic approach to modelling of complex systems which becomes the standard medium for

expressing and reasoning in the life sciences and which also has major applications in the design of man made complex distributed reactive adaptive systems. Even partial success would open major commercial opportunities: there should be clear signs of this towards the final phases of this Grand Challenge exercise.

How the proposal meets the Criteria for a Grand Challenge

Scientific Significance.

IS IT DRIVEN BY CURIOSITY ABOUT THE FOUNDATIONS, APPLICATIONS OR LIMITS OF BASIC SCIENCE?

The relationship between machines and living things is an ancient but ongoing preoccupation of man, crystallised by von Neumann, Turing, Wiener and others.

IS THERE A CLEAR CRITERION FOR THE SUCCESS OR FAILURE OF THE PROJECT AFTER FIFTEEN YEARS?

Yes: wide use of models by Biologists to create new theories and results; new approaches to complex systems engineering and systems maintenance, and commercial activity with respect to both.

DOES IT PROMISE A REVOLUTIONARY SHIFT IN THE ACCEPTED PARADIGM OF THINKING OR PRACTICE?

By giving primary place to accurate computer modelling of known and future data, we can aspire to convince biologists of the effectiveness and relevance of computational ideas; theories, analogies, abstractions and structures, and thereby open the doors to applying existing and future analytic tools to complex biological problems. Conversely, Computer Scientists can hope to develop theories and techniques for the analysis and synthesis of massively complex reactive systems with profound repercussions for industries ranging from systems engineering to manufacturing.

DOES IT AVOID DUPLICATING EVOLUTIONARY DEVELOPMENT OF COMMERCIAL PRODUCTS?

The value of a successful out come is clear: even as a common dynamic framework for knowledge representation, a successful challenge would be of considerable use to life science. If the model were predictive as well, the value would be immensely enhanced.

But a profit-generating business model is not so obvious. The right choice from an individual company's point of view is to be second, that is, to be ready to adopt the fruits of the challenge as soon as someone else has done (and paid for) the work and ironed out the bugs.

Impact on Practice.

WILL ITS PROMOTION AS A GRAND CHALLENGE CONTRIBUTE TO THE PROGRESS OF SCIENCE?

From a purely computing perspective, the challenge can be thought of as a domain specific exploration of issues such as: massively distributed control;

knowledge representation and navigation; hi-fi volumetric representation; autonomic computing; sensory data integration architectures. For these and other issues, life form modelling provides a clear and concrete target, which complements the many theoretical and pragmatic 'free' explorations underway of topics such as ubiquitous computing, artificial life, emergent properties of complex systems. More speculatively, success may produce a quite revolutionary new development-as-computing paradigm.

DOES IT HAVE THE ENTHUSIASTIC SUPPORT OF ESTABLISHED SCIENTIFIC COMMUNITIES?

There are over 300 laboratories working on C. elegans, so there is considerable interest in modelling it. There is a similar level of activity on Aribidopsis and S. Cerevisiae. A number of groups are already active in computer modelling, for example:

- Dept. of Medical and Biological Informatics at the German Cancer Research Centre Heidelberg: sophisticated virtual models of C. elegans components http://mbi.dkfz-heidelberg.de/mbi/research/cellsim/
- The Dutch Silicon Cell project, Free University Amsterdam: http://www.bio.vu.nl/hwconf/Silicon/index.html
- L-systems for Plant Growth modelling: Prusenkiewicz and others, Calgary.
- Snapdragon Development: E. Coen and A. Bangham IFR/UEA Norwich.
- Institute for Systems Biology, Seattle. Software tools for Systems Biology. http://www.systemsbiology.org/research.html
- Japanese E-Cell project: http://www.e-cell.org/index.htm
- Armand M. Leroi, Imperial College at Silwood Park Ascot. Detailed modelling of growth of C. elegans. http://www.bio.ic.ac.uk/research/amleroi

DOES IT APPEAL TO THE IMAGINATION OF THE GENERAL PUBLIC?

A complete simulation of a living creature appeals to the public interest in the creation of artificial life, and even a limited predictive ability could ease the need for live animal use in science.

WHAT KIND OF LONG-TERM BENEFITS TO SCIENCE, INDUSTRY, OR SOCIETY MAY BE EXPECTED?

For Science: significant acceleration of progress due to:

- Massively improved shared distributed computational observation engines based on standard models for recording, representing and accessing knowledge about life forms
- Effective frameworks for incorporating knowledge about complex systems with applications in both biology and systems engineering
- Inspiration to create radically new models of computation

For Society: just as accurate computer models of hydrodynamics have almost entirely replaced live nuclear testing, so we might hope that sufficient investment in accurate biological models might remove or at least considerably weaken many of the arguments for experimenting with live animals. Bringing together our rapidly expanding knowledge about the mechanics of cell lifetime control in the form of a predictive model may give us an indispensable tool for sharing and creating new knowledge about cancer.

Scale and Distribution.

DOES IT HAVE INTERNATIONAL SCOPE?

A large number of labs are beginning to build serious models of parts of plants or animal systems. By making a leading contribution to the construction of a faithful computational models, the UK can have enormous influence on the development of research in both biology and computer science.

HOW DOES THE PROJECT SPLIT INTO SUB-TASKS OR SUB-PHASES, WITH IDENTIFIABLE GOALS AND CRITERIA, SAY AT FIVE-YEAR INTERVALS?

A possible (partial) breakdown is: development; reactivity; front end; object modelling; accessing existing and future data; exemplars; cell virtual machine; inter cell protocol; intra cell protocol; view control; version control; information provenance; concurrent development control;

5y: accurate development models of exemplar organisms beginning to predict the result of experiments.

10y: models with partially accurate sensory responses beginning to emerge.

15y: first complete models.

WHAT CALLS DOES IT MAKE FOR COLLABORATION OF RESEARCH TEAMS WITH DIVERSE SKILLS?

Successful completion of the challenge obviously requires efficient distributed interactive simulation coupled with a good visual / haptic front end extended to model the desired experimental environments. Specifying and realising accurate models even with just hundreds of cells will stretch the current technology, and achieving the necessary realism at the interface is beyond the current state of the art.

We can expect continued advances in the general capability for distributed simulation and graphics to emerge from industry, driven for example by the games market. The real challenge will be in determining the right design choices for representing cells and cell interactions. There are many groups already working in this difficult area (see list of references for a very small sample), and it looks as if there is enough evidence on which to base some initial design decisions for the challenge. However, to achieve the best chance of success will require an inspired mix of knowledge about development and computing.

It is tempting to assume that some parts of computer science, for example the parallel rewriting models, massively parallel architectures, cellular automata etc, have – if not the answer – then at least the right sort of questions. But we should be very careful indeed in making such assumptions. There are many

bits of computing science which *might* contribute to this challenge, but we may be more likely to achieve success if we assume until we can prove otherwise that our theories and concerns are not as relevant as we might like to think, and that we need *new* questions. Perhaps an analogy with thermodynamics is relevant here: it gives the practical engineer limits on what he can achieve, but it doesn't tell him how to realise a practically effective engine.

HOW CAN IT BE PROMOTED BY COMPETITION BETWEEN TEAMS WITH DIVERSE APPROACHES?

Good science is about good questions. A critical step towards a successful challenge is to identify sharp sub-questions that different teams can attack simultaneously, perhaps in the form of a competition. A good recent example of the use of a competition to advance science is the Abbadingo One DFA (Deterministic Finite Automata) Learning Competition[5]. This involved careful attention to establishing a proving ground consistent with but just beyond the state of the art, and precise rules for submission and victory.

Some preliminary competition regarding the architecture for the whole design might be attempted, and this might also help draw in people with the necessary vision and other qualities to drive the challenge.

Timeliness.

WHEN WAS IT FIRST PROPOSED AS A CHALLENGE? WHY HAS IT BEEN SO DIFFICULT SO FAR?

In the sense of simulating life aspects of life, the challenge is very old. If the challenge is regarded in a more down-to-earth fashion, it can be seen as attempting to extend the reach of computer simulation to include increasingly accurate models of simple life forms. In this view, computer simulation and visualisation is a new medium for knowledge representation rather than a way of creating life, and in this sense the challenge is about as old as computer simulation.

Its difficulty arises from the need to develop a dynamic representation of knowledge that can accommodate both dynamic and static aspects of living creatures over a wide range of resolution levels.

WHY IS IT NOW EXPECTED TO BE FEASIBLE IN A TEN TO FIFTEEN YEAR TIMESCALE?

Some ad-hoc computer modelling techniques are already making headway by concentrating on specific aspects. We can expect at the very least that specialist computer models will become a de-facto set of standards for working with particular life forms. The real challenge is to do a small number of whole organisms very, very well, and to synthesise a generic framework capable of dealing with a wide range of knowledge about a wide range of creatures.

---

[5] Kevin J. Lang, Barak A. Pearlmutter and Rodney A. Price, *Results of the Abbadingo One {DFA} Learning Competition and a New Evidence-Driven State Merging Algorithm*, Lecture Notes in Computer Science, V1433, 1998, also citeseer.nj.nec.com/article/lang98results.html

WHAT ARE THE FIRST STEPS?

1. Identify stakeholders, building on existing BioModelling communities (e.g. MIPNETS) with EPSRC/BBSRC assistance.
2. Establish a series of meetings to flesh out the skeletal challenge into a much more precise form, and construct a roadmap for the challenge.
3. Preparation of more detailed outline plan and presentation to major stakeholders.

It is suggested that the initial focus is on modelling accurately the development of a small number of life forms. This focus is proposed because:

a. There is considerable detail about some well-studied models such as C. elegans, arabidopsis, yeast etc.

b. The number of cell divisions modelled gives a logarithmic measure of the computational cost, so enabling the modelling work to concentrate on establishing detail for some number of divisions which is easily within reach of fairly standard computers.

c. Getting even the first division right requires establishing the key attributes of cells and their interactions.

d. Last but not least, the initial divisions of many classes of life show similarities.

e. The exact life forms chosen should emerge from detailed consideration of the issues to be tackled and the work already underway. For example, work is already underway on C. elgans [].

WHAT ARE THE MOST LIKELY REASONS FOR FAILURE?

Vested interests are likely to cause tensions within and between existing scientific cultures. Strong and visionary leadership from both communities will be needed to establish an effective framework and to attract the strongest rather than the weaker scientists to a risky interface area.

Biologists may need to adapt to new approaches to representing knowledge that cuts across existing biological compartments, and to take on board some sophisticated notions from distributed computing.

Computer scientists may have to suppress a natural desire to see the challenge as a way of justifying yet more work on some existing research topic that may be of only peripheral relevance to the challenge.

We continually underestimate the rate of progress in computing, and this frequently leads to investment in research efforts that address problems, which, by the end of the research, have largely vanished. A *technology intercept* strategy is needed which works towards a model which can be realised with technology which will be around in 5,10,15 years time.

References which indicate the state of the art

The following references illustrate the state of the art in the detailed computer modelling of life forms. Activity is widespread, and although it is generally

diffuse some common approaches are beginning to emerge, and some ambitious groups are beginning to attempt serious integrated models, although not yet to the extent proposed in this challenge.

Agarwal95 P.Agarwal. The Cell Programming Language. Artificial Life, 2(1), pp 37-77, 1995.

AGB+01 azevedo:01, Azevedo, R. B. and Gumbel, M. and Braun, V. and Agapow, P. M. and Houthoofd, W. and Jacobsen, K. and Platzer, U. and Meinzer, H. P. and Borgonie, G. and Leroi, A. *Nematode embryonic cell lineages are computationally efficient*, Proceedings of the 1st Computational Embryology Workshop, editor = {Braun, V. and Gumbel, M. and Leroi, A M. and Meinzer, H P, 2001, Abteilung Medizinische und Biologische Informatik, Technical Report 125

AmsSS The Dutch Silicon Cell Project, Free University Amsterdam, http://www.bio.vu.nl/hwconf/Silicon/index.html

BBSRCvision http://www.bbsrc.ac.uk/about/strategic_plan/bbsrc_vision.pdf

Bong00 Josh Bongard, Chandana Paul (2000). *Investigating Morphological Symmetry and Locomotive Efficiency using Virtual Embodied Evolution.* To appear in J.-A. Meyer et al (eds.), From Animals to Animats: The Sixth International Conference on the Simulation of Adaptive Behaviour http://citeseer.nj.nec.com/bongard00investigating.html

D-BS94 Duvdevani-Bar, S. and L. Segel. 1994. "On Topological Simulations in Developmental Biology." Journal of theoretical Biology 166: 33-50.

Eggen97 Peter Eggenberger. *Evolving morphologies of simulated 3d organisms based on differential gene expression.* In [10], page in press, 1997. http://citeseer.nj.nec.com/article/eggenberger97evolving.html

FI95 Kurt W. Fleischer, david H. Laidlaw, Bena L. Currin and Alan H. Barr, *Cellular Texture Generation*, ACM SIGRAPH 1995

FIPhD95 Fleischer, Kurt W. (1995) *A Multiple-Mechanism Developmental Model for Defining Self-Organizing Geometric Structures*. PhD thesis, Caltech 1995, http://resolver.caltech.edu/CaltechCSTR:1995.cs-tr-95-14

FrPr92 deBoer, M. J. M., F. D. Fracchia and P. Prusinkiewicz. 1992. "Analysis and Simulation of the Development of Cellular Layers." In *Artificial Life II.*, edited by C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen. 465-483. Reading, MA: Addison-Wesley.

GCweb02 http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/Grand_Challenges/background.html

Heidl Sophisticate virtual models of C. elegans components. Department of Medical and Biological Informatics, German Cancer Research Centre Heidelberg. http://mbi.dkfz-heidelberg.de/mbi/research/cellsim

KHKMPH03 N. Kam, D. Harel, H. Kugler, R. Marelly, A. Pnueli, E.J.A Hubbard, and M.J. Stern: "Formal Modeling of *C. elegans* Development: A Scenario Based Approach". Proc. 1st International Workshop on Computational Methods in Systems Biology (CMSB03), LNCS 2602, pp. 4-20, Rovereto, Italy, February 2003. © Springer-Verlag Berlin Heidelberg 2003

Leroi Armand M. Leroi, Imperial College at Silwood Park Ascot. *Detailed modelling of growth of C. elegans*. http://www.bio.ic.ac.uk/research/amleroi

LiTeng01 X.-Y. Li and S.-H. Teng.. Proc. 12th Annu. ACM-SIAM Sympos. Discrete Algorithms, 28- *Generating well-shaped Delaunay meshes in 3D* 37, 2001. http://citeseer.nj.nec.com/li01generating.html

MF79	Raymond J Matela and Robert J Fletterick, *A Topological Exchange Model for Cell Self-sorting*, J. theor. Biol. (1979) V76, pp.403-414
Paton02a	Paton, R.C. Process, structure and context in relation to integrative biology, Biosystems 2002 Jan V64, pp63-72
Paton94	Paton R.C. Computing with Biological Metaphors
Pru99	Przemyslaw Prusinkiewicz: *Modeling Plants and Plant Ecosystems: Recent Results and Current Open Problems*. 1999 Computer Graphics International (CGI '99), June 7-11, 1999 - Canmore, Alberta, Canada, Proceedings. IEEE Computer Society, 1997, ISBN 0-7695-0185-0
Vernon	Plant Embryogenesis. Dr. Daniel M. Vernon, Department of Biology, Witman College, Washington. http://marcus.whitman.edu/~vernondm/embryo/plant.html

END