Algorithmic Models for Evolution

M. Ausloos, N. Vandewalle

SUPRAS, Institut de Physique, Université de Liège

Statistical physics combined with computation modelling together provide new techniques for studying evolutionary systems.

Biologists are confronted in Nature by many patterns including DNA sequences, cells, organisms, phylogenetic trees, the collective behaviour of insects and swarms, animal territories, *etc.* ranging from the microscopic to the macroscopic. In their quest for a better understanding of life, scientists and humans in general are captivated by the high level of complexity and diversity of patterns built from a finite (but sometimes very high) number of elements. In 1859, Darwin introduced a theory of evolution in which gradual complexity derives from a stochastic natural selection of species: "God plays with dice" in contrast to some *"Deus ex machina"* admitted since antiquity. This concept has dominated the theory of evolution this century. Recently, Gould and coworkers introduced the notion of self-organization in evolution by suggesting that "God plays with loaded dice".

Physicists up to a few decades ago tried mainly to solve the question of the origin(s) of the Universe but rarely attacked the problem of the origin(s) of life on Earth and perhaps elsewhere. The crucial question asked of physicists relates to the physical principles that can be carried over into concepts describing evolution. Many attempts to model life on a physical basis have been presented. The great majority represent a mere mixing of computer modelling and statistical physics owing to the great versatility of these two scientific domains. The models and most findings have much in common with critical phase transitions and the geometry of fractals.

Models based on algorithms have several advantages because the mechanisms under study can be completely controlled. Moreover, today's computers can now simulate biological systems having sizes which can be stored and manipulated. This is possible for the human genome (containing about 10^6 elements), for the human body (10^9 cells), and for the whole of humanity (some 5×10^9 individuals). By contrast, most physical systems contain an Avogadro number of elements. It goes without saying that an adequate understanding of Nature's rules and laws requires *a priori* much more than the identification of mathematical laws.

Marcel Ausloos (ausloos@gw.unipc.ulg.ac.be) is on the research staff of the Institut de Physique B5, Université de Liège, Sart Tilman, B-4000 Liège. **Nicolas Vandewalle** is a research fellow in the same institute.

Moreover, one should note that theoretical approaches such as a mean field or a re-normalization group approximation are not lacking, even though there is no Hamiltonian formalism. However, a continuum-like approach has not been proposed so differential equations are not considered, only simple (binary) rules to drive the systems.

Several Models

The most significant models [1] are based on cellular automata, ageing and co-evolution. Introduced by Conway as a mathematical toy, the *deterministic cellular automaton* "game of life" is studied extensively as a paradigm for modelling the evolution of a society of living organisms. Particular rules simulate biological-like loneliness and over-crowding processes as well as the spreading of life. Growth rules incorporating irreversibility generate numerous local stationary configurations indicating the emergence of a variety of cyclic or locally stable societies that follow each other in a complex sequence. The process reaches a particular steady state after an "avalanche" of life activity, where the avalanche events are distributed according to power laws. Among the interesting physical ideas put forward, those pertaining to "damage spreading" owing to inner perturbations or various boundary conditions are the most fascinating.

Kauffman studied another cellular automaton in order to simulate gene activity in the DNA of a biological cell. Assuming that each cell can be "on" or "off", evolutionary changes could then be understood as resulting from a small modification of the "switching" systems. For an infinite connectivity between switches, the process was found to be in a chaotic regime, but for a finite connectivity a transition occurred between order and chaos at some critical connectivity value. There are also numerous models for RNA, DNA replication and cell differentiation.

The problem of *ageing* is concerned with population dynamics that accounts for hereditary survival and reproduction rates. Computer models seem to date back to 1985. Relatively recent Monte-Carlo simulations show that a population suffering bad mutations, repairing mutations and/or food restriction can reach a steady state depending on the parameters selected. The dynamics of the population in approaching its

self-organized steady state again exhibits power-law behaviour.

The *co-evolution model* introduced recently by Bak and Sneppen considers a set of interacting species through the mutations of the less-adapted ones to the ecosystem. On an arbitrary, geological-like timescale, the model leads to punctuated equilibrium, *i.e.*, rapid bursts of life activity separated by long periods of stasis. This is in agreement with Gould's theories, but is in opposition to Darwin's gradualism for which speciation takes place continuously and slowly. Moreover, the size distribution of the bursts of activity and the dynamics of the evolving ecosystem follow power-law behaviour.

Investigations of self-organized critical properties based on studying power laws are of crucial interest (see insert). Simulations should be carried out in order to obtain power laws over at least two or three decades for events in both space and time. For good statistics, the typical size of the system should be several tens of thousands of individuals (species, genes or cells in the above models) and the minimum required time should be on the order of 10 000 evolution steps (mutations, switches or life spans in the models).

Self-organized Criticality

Phylogenetic trees are used by paleontologists to describe evolution. We have proposed extensions to phylogenetic-like tree-growth processes that allow interesting studies such as those dealing with the screening of species from evolution and the effects of the genetic range in competition-correlation prone populations [2]. Computer simulations demonstrate a well-punctuated equilibrium in such a process. (*e.g.*, the cover illustration showing four steps in the growth of a phylogenetic tree).

Computer models of this type do not aim to explain the mechanisms of biological evolution from a biological point of view. They

The top quark reconstructed mass distribution for the b-tagged $W + \geq$ 4-jet events (solid line). Also shown are the background shape (dotted line) and the sum of the background plus $t\bar{t}$ Monte Carlo calculations (dashed line) for a top quark mass of 175 GeV/c^2, with the background constrained to the calculated value of $6.9^{+2.5}_{-1.9}$ events.
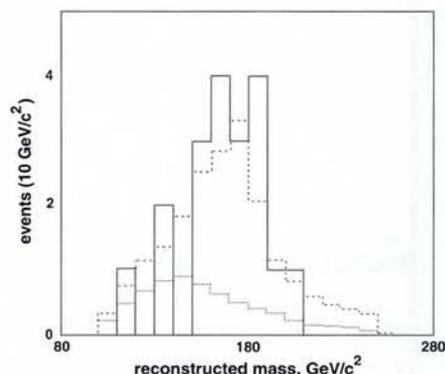
mimic, instead, the complex features of biological evolution based on simple rules so that interesting generalisations can be developed. The main finding is that most models demonstrate self-organized critical behaviour, as in sand piles, earthquakes, *etc.* Indeed, numerous models lead to an unstable critical state. When perturbing this quasi-steady state, the evolution process always self-organizes into the same critical steady state having, as a consequence, the same appearance. The dynamics of the perturbations show power-law behaviour. Self-organized criticality of this type is sometimes said to be on "the edge of chaos" since a small perturbation can cause an accidental and unpredictable endless dissipation.

Postulating the self-organization between chaos and order as the basic rule for life evolution is plausible because this behaviour can only evolve if small mutations have small but finite effects. If they have no effect (*i.e.*, they are deep in the ordered region), mutations cannot drive evolution; for catastrophic effects (*i.e.*, deep in the chaotic region), a single mutation can destroy almost everything. This leads one to speculate whether the physical and universal rules for quarks and atoms which lead to self-organization into a critical state also

apply to the emergence of complexity and living organisms.

The physics community's contribution to understanding evolution will bear much upon the study of self-organized criticality presented by "artificial" (but biologically motivated) computer systems. The concept of critical self-organization is recent and a general formalism of such behaviour is still lacking. Ideas will take time to mature, hopefully not on a geological timescale.

One should, for example, look whether additional biological constraints lead to the emergence of evolution process(es) from self-organized criticality. Fundamental and philosophical questions about evolution and the origins of life, as well as how many types of life can exist, are raised as a result. Computers can search for some answers. It is also necessary to identify which "perturbations" force natural systems to evolve towards a particular "steady-state". Finally, it will be important, but very difficult, to study transient effects that today defy a systematic description.

[1] Vandewalle N. & Ausloos M., "Evolution motivated computer models", in *Ann. Rev. Comp. Phys.*, Ed.: D. Stauffer (World Scientific, 1996) Vol. 3; in press.
[2] Vandewalle N. & Ausloos M., *J. Phys. I* (1995), in press.

Top Quark Confirmed

On 24 February 1995, both the CDF and D0 Collaborations at the Fermilab Tevatron Collider submitted papers [1] detailing new and convincing evidence for the existence of the top quark. These papers confirm with significantly increased statistics evidence from CDF announced in April 1994 [2].

Top quarks are produced as particle-antiparticle pairs in $\bar{p}$-p collisions, and because of their large mass (>170 times that of the proton) are produced very rarely. From $\approx 10^8$ events recorded on tape by the experiments, the CDF and D0 Collaborations reported respectively 43 and 17 events consistent with t-quark production. The new results correspond to a data sample corresponding to an integrated luminosity of ≈ 117 pb^{-1}, as compared with the ≈ 19 pb^{-1} used for the previous CDF analysis. This huge increase in statistics has been made possible by continued and higher intensity operation of the Fermilab collider.

Top quarks decay to a W-boson and a b-quark, and the W-boson decays into either leptons or quarks. Events for which both W-bosons decay into leptons (e or μ) result in a striking high transverse momentum P_T opposite-sign dilepton (e, μ) signature with associated missing transverse energy from undetected neutrinos and at least 2 b-quark jets. Both experiments are able to detect this production and decay chain with good efficiency. The CDF experiment reports a

signal of 6 events with an estimate background from non-$t\bar{t}$ sources of 1.3 ± 0.3 events. Of the 12 possible b-quark tags in this event sample, 5 b-tags are observed after applying CDF identification criteria, compared with 3.5 expected, and 0.5 from a background sample (see below). The D0 experiment reports 3 dilepton events with an expected background of 0.65 ± 0.15 events.

Events for which one W-boson decays into leptons and the other into quarks are more plentiful (30% of all the events decay in this way). However, the background from QCD W-production processes with associated jet activity is large, and in both experiments b-quark tagging is used to reduce the W-background. This is done in each experiment by identifying low-P_T charged leptons near the jet resulting from semileptonic b-quark decay, and more importantly in CDF by using a high-resolution silicon tracking detector (SVX) to directly reconstruct the offset vertices (up to several mm) from b-decays.

The performance and acceptance of the SVX was significantly upgraded prior to the current run. From 21 (W + ≥ 3 jets) having at least one SVX tag (a total of 27 tags are recorded), the expected background from non-$t\bar{t}$ processes is estimated to be 6.7 ± 2.1 tags. For the soft lepton analysis, CDF records 23 tags in 22 events, with 15.4 ± 2.0 tags expected from background. The equi-

valent result in D0 is 17 lepton + ≥ 3 jet events with a background estimate of 3.9 ± 0.6 events (this sample includes 6 low-P_T muon tags from semileptonic b-decay, with a background of 1.2 ± 0.2 events).

The event samples and associated background uncertainties noted above reflect the differing capabilities of the two experiments; the CDF experiment uses to advantage its excellent secondary vertex reconstruction capability (SVX), while the D0 experiment benefits from its outstanding muon-detector and calorimeter coverage.

From these data the probability of a fluctuation of the background to produce the observed signal is estimated as < 2 x 10^{-6}, corresponding to 4.6 (4.8) standard deviations for the D0 and CDF experiments, respectively. Also in agreement, the t-$\bar{t}$ production cross-section in $\bar{p}$-p interactions at the centre-of-mass energy $\sqrt{s} = 1.8$ TeV is measured to be respectively 6.4 ± 2.2 pb and $6.8^{+3.6}_{-2.4}$ pb, for the quoted t-quark mass values. The results are in reasonable agreement with theoretical expectations.

Several concerns of the previous CDF analysis, *e.g.*, the measured cross-section and the excess of W+jet events compared with W+jet QCD predictions, have been resolved by the increased statistics.

From the sample of W+4-jet events, each experiment has performed a t-quark mass reconstruction. The CDF experiment estimates the t-quark mass to be 176± 8(stat)±10(sys) GeV/c^2, to be compared with an estimated mass given as 174±10(stat) ±13(sys) GeV/c^2 in the 1994 data. The t-quark reconstructed mass is shown in the figure. The significant systematic uncertainty results primarily from uncertainties on the mass-dependence of the background, and on uncertainties on the effect of gluon radiation. The challenge within CDF is now to reduce both the statistical and systematic uncertainties on the mass measurement to a level that constrains the allowed values of the Higgs mass (assuming validity of the Standard Model). To this end, each experiment is likely to double its data sample in the current collider run late in 1995. The D0 experiments estimates the mass to be 199^{+19}_{-20}(stat)±22(sys) GeV/c^2; this result is not inconsistent with the CDF result.

[1] Abe F. *et al.*, *Phys. Rev. Lett.* **74** (1995) 2626; Abachi S. *et al.*, *Phys. Rev. Lett.* **74** (1995) 2632. [2] Abe F. *et al.*, *Phys. Rev. D* **50** (1994) 2966; Abe F. *et al.*, *Phys. Rev. Lett.* **73** (1994) 22. See also: Abachi S. *et al.*, *Phys. Rev. Lett.* **72** (1994) 2138; Bellettini G. & Clark A.G., *Europhys. News* **25** (1994) 86.

A.G. Clark, *Geneva University*