Committee 1
The Information Revolution,
Higher Education and Research

DRAFT--Oct. 15, 1997
For Conference Distribution Only

COMPUTER SIMULATIONS AS THE SCIENCE OF FALSE REALITY

by

Itamar Procaccia
Department of Chemical Physics
The Weizmann Institute of Science
Rehovot, Israel

The Twenty-first International Conference on the Unity of the Sciences
Washington, D.C.   November 24-30, 1997

[illegible]

assessment of numerical simulations cannot be achieved without a preliminary analysis of the difference between theories and models. Theories are based on symmetries and invariants, and the degree of arbitrariness in them is hence immensely reduced. Models that are not respecting symmetries and invariants can have much freedom in them, and in particular freedom to make mistakes that simulations cannot possibly detect. The urgency of an assessment of these issue is enhanced by the recent developments in the biological sciences, developments that convince many researchers that they are ripe for quantitative modeling. Are they? Do we have a real understanding of the invariants of biological sciences? We need to remember always that numerical simulations are [illegible] translation into code lines of some proposed model description of nature. Is it then possible to be surprised by the results of simulations? Can one make discoveries about Nature using simulations? Can simulations replace laboratory experiments? These and related problems are the subject of this note.

2. REALITY, SCIENCE AND MATHEMATICS

Is Science discovering the secrets of Nature? The present author believes that Science is more creative than that. Science is not [illegible] Nature the way we see it. Scientific research is a creative process that is concerned with Nature out there, but is not limited to the passive role of "discovery". After all, Nature is an [illegible] complicated reality, too detailed and too amorphous to succumb to mere intellectualization. Rather, the process of scientific progress invariably has to do with an [illegible], almost a caricaturization, of a piece of Nature, and with the process of learning what [illegible] questions have precise answers.

Take as an example a subject that is close to the heart of the present author, turbulent flows of fluids. Turbulence is a ubiquitous phenomenon in nature, in the atmosphere, in the oceans, in [illegible] straits and rivers, in the wake of a car and around the wings of aircrafts, to name just [illegible] examples. The motion of the fluid particles in turbulent flows is [illegible] complex, chaotic and unpredictable. The way that fundamental physics deals with such [illegible] is to focus on those aspects for which there are questions that can be answered. Firstly, the physicist idealizes turbulence to an isotropic and homogeneous state, which means in [illegible] velocity, and considers the flow of a continuous medium rather than that of particles. The flow [illegible] are described in mathematical language using the so-called Navier-Stokes equations of fluid mechanics, in which sound, heat, humidity etc. are totally neglected. Then, rather than asking questions about the detailed aspects of the flow, the physicist focuses on statistical objects, [illegible] averages, correlation functions and structure functions. If the forcing that creates turbulence is statistically stationary, the statistical objects are describable in terms of invariants: time invariants and invariants to rotation, translation, and reparametrization of the spatial coordinates. Finally, the physicist declares *ex cathedra* that "understanding turbulence" is equivalent to being able to compute, using the mathematical description, those characteristics of the statistical objects that can be measured in controlled experiments. In this way a scientific discipline is formed, with all [illegible] details which are responsible for lifting airplanes in the air or for improving our gasoline [illegible] being delegated to the applied sciences, in which it is less important to "understand", but [illegible] construct machines that "work".

[illegible] this description seems to belittle the process of Science, this is not our aim. On the contrary, we [illegible] the creative ingenuity of the process in which a proper set of icons is being identified in order to discuss aspects of complex phenomena that otherwise just form a mysterious mess. The [illegible] involves the isolation of appropriate icons, idealization of the problem such that the icons [illegible] identifiable relations between them, creating the mathematical description that allows the discussion of these relations, and finally choosing what questions to ask for which there are answers, and the discovery of such questions *is* what is meant by scientific progress. Once a "good question" has been asked, (as every scientist would agree) one is more than half the way to resolving a scientific issue. Of course, to ask a good question one needs a set of iconic representations with relations between them to allow for such a question. Progress and understanding are intimately related to the development of a language in the precise sense that we outline above.

There is of course a bit more to this. After all, the magic of language is also at the basis of Botany and Zoology, but this is different. When Adam was invited by God to name the animals, humanity began its ascendancy to a position of dominion over the beasts. Naming is controlling. Even now a multitude of people draw extreme pleasure from naming plants and trees with complicated Latin names, feeling that by doing so they know these plants better and form some kind of ownership on them. A pine is a pine even if you do not call it anything, and calling it *Pinaceae pinus* does not add anything to the understanding of this species of trees. Yet humans have a fascination with language even on this dumb and mechanistic level. The language of Science is more powerful, since it offers (when it is successful) an ability to predict the outcome of yet untried experiments. This is indeed one of the greatest prides of science, bringing to mind Albert Einstein's remark that "the most incomprehensible thing about the Universe is that it is comprehensible". By the creative process of our brain we develop a language with which we discuss natural phenomena, but then the same language happens to be so self consistent (*grammatically perfect?*) that it allows us to predict results of new experiments! The deep reason of this success is of course what we stressed above, i.e. that the basic objects of the scientific discourse are the invariants for which we create iconic representations. If our discussion has some consequences about invariants, these consequences are guaranteed to agree with future observations, simply because invariants are invariant.

We note in passing that the deep role of a language is not limited to the scientific creative process. In fact, other fields of human creativity are also marked by a similar predominance of language. Every great artist creates a new language, which is often not understood, some times even rejected, by the same professionals that exalt the already accepted languages of previous generations. Many such examples are known, and in recent memory we have the famous event of the "Salon des Refusés" in which Impressionist art staged the first onslaught of a new language on 19th century academic art. Picasso's art is one of the brilliant examples of this century's ever developing new language of art, where each period of his creativity took some time to be accepted by a public that reacted slowly to the turns and jolts of Picasso's ingenious creativity. Once a new Picasso dialect was accepted it was quickly hailed as "great art" and was assigned to the halls of rich museums and important private collections. Literature and poetry are languages that use standard language as their vehicle, but one observes the tremendous changes in the way that this is done over the centuries. Music is a language of sounds, and the rules of harmony and counterpoint its grammar. Every creative activity is finally connected to the creation of a language, with rules, self consistency, and a communicative inherent value. Humans interact through language, in all the myriad forms that it may take. Science is similar, but science boasts the unique added value of being "precise" and being "right". Being precise is nothing more than rooting the scientific discourse in a mathematical apparatus, an apparatus that is solidly built on logic and commonly accepted rules. Being right in science means specifically that predictions can be made and tested against "objective reality". It is important to assess now what is the status of numerical simulations in the context of the scientific endeavour.

3. SOME HISTORICALLY SUCCESSFUL NUMERICAL SIMULATIONS

The availability of fast digital computers, and consequently of numerical simulations of complex phenomena, has already changed forever the way scientific research is conducted. I will limit my comments to subjects of my expertise, which is the physics of complex systems, or the physics of macroscopic phenomena on the human scale. These phenomena include macroscopic systems out of equilibrium, in which the dynamics of many-body systems play a crucial role. In my memory the first ever new discovery made with the help of digital computers in this field was the simulation performed in the early 1950s by Fermi, Pasta and Ulam using the newly available MANIAC-I computer in Los Alamos [1]. Fermi suggested that it would be highly instructive to integrate the equations of motion numerically for a judiciously chosen, one-dimensional, harmonic chain of mass points weakly perturbed by nonlinear forces. In particular Fermi had in mind to test some of the most widely believed assertions of equilibrium statistical mechanics, such as equipartition of energy, ergodicity, and the like. After some back and forth, Fermi, Pasta and Ulam decided to numerically integrate the weakly nonlinear, fixed end, one dimensional chain of (N [illegible]) [illegible] mass points having the Hamiltonian

$$H = \sum_{k=1}^{N-1} \frac{P_k^2}{2} + \sum_{k=0}^{N-1} \frac{(Q_{k+1} - Q_k)^2}{2} + \alpha \sum_{k=0}^{N-1} \frac{(Q_{k+1} - Q_k)^3}{3} ,$$

with $Q_0 \equiv Q_N \equiv 0$, and Q_k and P_k are the coordinate and momentum of the kth particle, and α is a small nonlinear coupling parameter. It is natural to discuss the results of the simulation in terms of the normal modes coordinates A_l specified by

$$A_l = \sqrt{(2/N)} \sum_{k=1}^{N-1} Q_k \sin(kl\pi/N) .$$

The simulation, performed by a programmer by the name of Mary Tsingou, employed $N = 32$ particles, with initial conditions in which only the fundamental mode A_1 was excited and given amplitude $A_1 = 1$.

The results of the simulations were very surprising. Whereas the initial behavior seemed to be as expected, i.e. the energy of the fundamental mode was sucked into higher modes until the energy in the fundamental reduced to about 15% of its initial value, at later time the energy in the fundamental modes returns to within 3% of its value at t=0! Fermi, Pasta and Ulam immediately recognized that these results were simply astounding. First, they appear to violate the canons of statistical mechanics, which assert that this nonlinear system should exhibit an approach to equilibrium with energy being shared equally among all degrees of freedom. But even more astonishing, they seem to invalidate Fermi's theorem regarding ergodicity in nonlinear systems. Fermi is said to have remarked that these results might be one of the most significant discoveries of his career.

Another early example of successful numerical simulations was the discovery of long-time tails in hydrodynamic systems by Alder and Wainwright in 1967 [2]. As mentioned above, in hydrodynamics one attempts to describe the motion of a large number N of interacting particles using a continuous description of "fields" which vary in space and time on scales much larger than the typical scales of individual particles. Thus instead of recording, say, the momentum p_i of every particle with position $r_i(t)$, one considers the momentum "field" $P(r,t)$ at some point r in space at time t:

$$P(r,t) = \sum_i p_i \delta(r - r_i) \quad (2)$$

Next one introduces the Fourier representation

$$P_k = \int dr\, e^{-ik\cdot r} P(r,t) \,. \quad (3)$$

and then focuses attention only to those k-components that are much smaller than the interparticle distance. Such a reduced description is based on the notion of separation of time scales. The idea is that even a complicated system that is composed of many particles is characterized, when it is isolated from the rest of the world, by a few conserved variables, like the total energy, the total momentum and the total number of particles. These quantities are invariants, not changing at all as a function of time. Consequently, if one is interested in changes that occur on large scales, scales much larger than the typical size of a single particle, the densities of these quantities are expected to vary very slowly compared to molecular motion, and one then hopes that it is possible to construct a mathematical language that provides a description of the large scale dynamics of the densities of the conserved fields in closed form. One of the aims of Statistical Physics in the 50's and 60's was to achieve such a description based on first principles, and on the assumption that once the "slow" variables are taken into account in full, all the rest decays on much faster time scales, maybe exponentially, and quickly gets out of the way of the large- scale long-time description that people hoped to get. The mathematics seemed fine, and a major result was achieved, i.e. microscopic formulae for transport coefficients like the viscosity and diffusion coefficient. The simulations of Alder and Wainwright indicated that there was a fly in the ointment. They performed on the computer what the theorists cannot do in their brain, they followed the detailed dynamics of a bunch of classical particles (hard spheres) and observed that objects that were supposed to decay fast had long time tails. They computed the time correlation functions of the velocity of a tagged particle. The correlation decayed rapidly to zero as expected, but then, almost wickedly, overshot zero and remained finite for macroscopically long times until they settled, as power laws, to zero again. This simulation was the first indication that the mathematical language that was developed by the statistical physicists was not rich enough, and had to be augmented. Later it turned out that the ways to enrich the theory had an exciting relevance to the dynamics of systems near the critical point of phase transitions, in which the assumptions of relative slowness of the conserved variables fails in an even grander manner. These developments led to the theory of dynamical critical phenomena, but this is already a different story, which is told in many articles and text-books [3].

The trivial moral of this story is that simulations can be very useful. The deep moral of the story is that one needs to think about levels of description and levels of abstractions. The simulations of Alder and Wainwright translated into code lines the mathematical description of classical mechanics as it was known since Newton for the motion of a bunch of particles. Then it asked questions about the mathematical description on the level of hydrodynamics, the continuum description that was attempted by current research. By doing so the simulations could shed important light on deficiencies in the assumptions that were used in the development of the macroscopic new language, and led to improvements that closed the gap. This is a good example of the power of simulations, and unfortunately it is less common than one would like to believe.

Another beautiful example of the abilities of numerical simulations is the discovery of chaos in simple dynamical systems by E. Lorenz [4]. Lorenz was trying to understand the mathematics of weather prediction, and constructed a model weather by integrating a set of coupled differential equations that were supposed to simulated a simple weather. His machine, a Royal McBee, was the best thing that he could have in his office in 1960. He examined the printouts of this simulations, and devised a primitive graphic output of one of the variables as a function of time. One day in the winter of 1961, wanting to examine one sequence of greater length, Lorenz took a shortcut. Instead of starting the whole run over, he started midway through, typing the numbers for initial conditions straight from an earlier output. To his surprise, the resulting orbit diverged rapidly from the printed version of the earlier run. Lorenz immediately grasped the reason. In the computer's memory, six decimal places were stored: .506127. On the printout, to save space, just three appeared: .506. Lorenz had entered the rounded-off numbers, assuming that the difference - one part in a thousand - was inconsequential. This was the discovery of the butterfly effect and of sensitivity to initial conditions in chaotic systems. After the later demonstration of metric universality in the period doubling scenario describing the *onset* of chaos by Feigenbaum, the road opened to an avalanche of mathematical and physical progress in understanding chaos in dynamical systems.

Let us analyze why these examples were so successful. In the first two cases one asked the computer to integrate presumably correct equations to assess *asserted* or *assumed* facts. It turned out that these facts were only alleged facts, and new thinking was called for. In the third example the model equations were probably more arbitrary, but there was a *generic* phenomenon to be discovered, a phenomenon that exists in any dynamical system that displays chaotic behaviour, and it was therefore less crucial to begin with "correct" equations. None of these simulations did what is becoming more common these days, which is asking questions about the specific behaviour of made up models; no amount of fancy computer graphics may save such simulations from being highly questionable, and we turn next to discuss these issues.

4. MODELS, THEORIES, AND UNIVERSALITY

In the applied sciences it is very customary to construct models which are endowed with enough parameters to be able to fit a particular situation, with the aim of constructing a device, a machine or a structure that can operate successfully under the conditions that the model pertains to. It is less important that the model were "right" in the deep sense, and the decisive criterion is whether it provides enough constraints for a successful construction of the said device. In the fundamental sciences one is more ambitious: the professed aim is to provide a *theory* that says something *correct* about the world.

Let us take an example to underline the difference. Suppose that we want to derive equations of motions that describe the dynamics of a given physical system. If these equations are not Galilean invariant, or, in other words, if the equations fail to describe the correct dynamics if we put the physical system on a train that is moving with a fixed velocity, we say that the equations are wrong, even stupid. In applications one may be less demanding. If we offer an engineer a *model* that works as long as the system is *not* put on a train, and if the application *does not* require the system to ever be put on a train, the engineer may be satisfied with such a model. Another example: suppose that we have a hamiltonian system and we write its equations of motion. If these equations do not have a simplectic structure, we say that the equations are wrong, even if they can be fit approximately to the dynamics of the system in some limited range of parameters. In other words, in the fundamental sciences one learned to respect fundamental symmetries and invariances and one demands absolute adherence to some requirements of correctness.

These general requirements have been turned into a powerful machine in the physical sciences, a machine that provides "correct" models that work because they have to. There are many examples, and I chose one from the work of the French school of pattern dynamics, led theoretically by P. Coullet. The example pertains to Rayleigh-Benard convection experiment with a static spatially periodic forcing, achieved by adding regular arrays of small bumps on both the upper and lower plates of the convection cell [5]. It was discovered that above the convection threshold, the spatial forcing gave rise to a parity-symmetry-breaking steady state. The heat currents are no longer vertical as in usual convection, but tilted either to the left or to the right, making an angle θ with the vertical line. When further increasing the Rayleigh number, the system reaches time-dependent regimes characterized by an oscillation of the direction of the hot fluid current around the mean tilted position. To the eye, the observed patterns are rather complex and time dependent.

Obviously one can develop a model for such a phenomenon by writing the equations of fluid mechanics, trying to solve for the post-threshold dynamics, satisfying rather unpleasant non-uniform boundary conditions. But we have learned to do this differently. One observes that the description of both parity transition and the breaking of the time translational symmetry (to give rise to oscillations) requires two order parameters A and B slowly varying with respect to space and time. The tilt angle θ should then be expressed in terms of A and B

$$\theta(x,t) = A(x,t) + \mathrm{Re}\left[B(x,t)e^{i\omega t}\right] .$$

The time evolution equations for A and B can then be derived from symmetry considerations. The global invariance of the set of solutions with respect to parity ($x \to -x$, $\theta \to -\theta$) and time translation ($t \to t + t_0$) implies that the equations are invariant under the following transformations:

$$x \to -x \quad A \to -A \quad B \to -B ,$$
$$t \to t + \phi/\omega \quad A \to A \quad B \to Be^{i\phi}$$

Assuming that the two order parameters are of the same order of magnitude leads to equations of the form

$$\frac{\partial A}{\partial t} = i\alpha\left(B\frac{\partial \bar{B}}{\partial x} - \bar{B}\frac{\partial B}{\partial x}\right) + (\mu + \nu|B|^2)A - \beta A^3 + a\frac{\partial^2 A}{\partial x^2} + \ldots$$
$$\frac{\partial B}{\partial t} = z_1 A\frac{\partial B}{\partial x} + \left(z_{11} + z_{12}A^2 + z_{13}\frac{\partial A}{\partial x}\right)B - z_2|B|^2 B + z_3\frac{\partial^2 B}{\partial x^2} + \ldots .$$

where the coefficients in the first and second equations are real and complex respectively. The numerical simulations of these equations, after a careful choice of the numerical values of the coefficients, agreed very well with the experimental images [5].

We note that this way of deriving a model is not free of assumptions. For one thing, one assumed that a series expansion in the magnitude of the amplitudes of the order parameters A and B and their gradients converges. This is not guaranteed, and in every case calls for careful reasoning. Also, a full "understanding" of the physics calls also for derivation (from first principles) of the numerical value of the coefficient, that in this approach can be only fitted by comparison with the experimental results. Nevertheless, I believe that this random example is a convincing reminder of the immense power of using symmetries, or invariances, to write down models that are *likely* to give the correct answers, because in some deep sense they have a good chance to be right. Once we write down all the terms that are allowed by symmetry, and convergence is guaranteed by some deep reason, the equations of motion above *must* reveal the observed dynamics in some corner of the parameter space. If they did not it would only signal that we either blew a symmetry, or forgot an order parameter, or that for some reason one of the order parameters were much smaller than the other. Otherwise we wrote down a model that is in fact a theory which can be trusted to deliver proper results.

This ideology reaches its utmost flowering in those situations when a universal behaviour exists. Critical phenomena related to second order phase transitions are a case in point. To the degree that the critical exponents depend only on the dimension of the system, its symmetries, and the number and tensorial nature of the order parameters, the simplest possible model that shares these properties with any arbitrarily complicated physical system will have the same critical exponents. Thus a ridiculously simple model like the Ising model becomes relevant not only for describing the behaviour of realistic ferromagnets in the vicinity of the critical point, but even for describing other, non-magnetic, physical systems (like a simple fluid!) with the same symmetries: one says that they exhibit an "Ising model fixed point". Similarly, a 1-parameter family of dynamical systems that undergo a series of period-doubling bifurcations on their way to chaos can be quantitatively understood in terms of an absurdly simplified quadratic map of the interval. The power of universality, when it applies, is immense in offering us the possibility of writing down simple models that are theories, or establishing beautiful theories by analyzing simple models.

What can we do when we do not have a sure understanding of the symmetries of a given problem, as in most biological applications? To the present author it is far from obvious. Is it likely that a model of a complex biological phenomenon can be simulated on the computer with useful consequences? All that we can do at this point is to shed additional light on the difficulties inherent in numerical simulations when there are no sure principles to guide our steps.

5. MACHINE PRECISION AND THE INABILITY TO USEFULLY SIMULATE

In this final section we make some comments on the implications of machine imprecision regarding the ability to simulate dynamics. In this discussion we deal with both non-chaotic as well as chaotic dynamical systems.

A. non-chaotic dynamics

The following example was proposed by M.J. Feigenbaum [6], motivated by Burger's equation for one-dimensional "fluid" without pressure

$$\frac{\partial u(x,t)}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} \tag{4}$$

Feigenbaum proposed to look at "modes" $\partial u(x,t)/\partial t \to \lambda u$. Denoting spatial derivatives with a dot and rescaling, we get

$$\ddot{u} + u\dot{u} + u = 0 . \tag{5}$$

The solution of this equation is depicted in Fig.x. Yet, no numerical integration may reveal this solution. The problem is that $\dot{u} = -1$ is a solution of this equation, whereas for large enough u, during the long linear ramp, $\dot{x}$ exponentially approaches -1, with higher derivatives exponentially diminishing. The computer throws away the exponentially small information that is essential in order to turn the ride back, and falls forever on the linear ramp.

Of course, the problem can be overcome. All the we need is to observe that (4) derives from a Lagrangian

$$L = (1 - \dot{u})\ln(1 + \dot{u}) - \dot{u} - u^2/2 . \tag{6}$$

of Hamiltonian

$$H = P^2/2 - \alpha(e^x - 1 - x) \qquad (7)$$

With a known energy and an appropriate code, Fig. can be reproduced. But this is exactly the point - we need to know in advance something *exact* about this equation of motion to know how to simulate it. If we do not, we are doomed to fail.

B. Chaotic dynamics

The problem is even more vexing when the dynamics is chaotic. On the face of it, chaos spells disaster, since local sensitivity to small errors is the hallmark of a chaotic systems. In other words, we know from the very beginning that in trying to simulate a chaotic system, any minute change in the initial conditions results in a macroscopic change in finite time. So what is the meaning of the simulation?

In one class of mathematically pleasant (and physically rare) dynamical systems the situation is saved by a remarkable theorem. The class of systems is "hyperbolic", and the theorem is known as the "shadowing lemma" [7,8]. A dynamical system is hyperbolic if phase space can be spanned locally by a fixed number of independent stable and unstable directions which are consistent under the operation of the dynamics. The shadowing lemma states that when the dynamical system is hyperbolic, then the pseudotrajectory that is generated numerically is very close to a true trajectory of the dynamical system, with adjusted initial conditions, which is called the shadowing trajectory. Thus locally sensitive trajectories are in fact globally insensitive, and a numerical simulation reveals behaviour that is very close to a "possible" behaviour of the true system.

But there's the rub - almost any natural dynamical system is not hyperbolic. So we run head-on to questioning how useful simulations may be? When we model a natural system it differs from its theoretical model by at least modeling errors. Generically we do not have to perbolic structure, and *global* sensitivity may lead to trajectory mismatch, in particular when long times are considered. The bottom line is that no trajectory of the theoretical model matches, even approximately, the true system outcome over long time spans [9].

6. CONCLUSIONS

This section will be written at a later stage, probably after hearing the reactions of the ICTS committee.

1. E. Fermi, *Collected Papers*, Vol II, U. of Chicago Press, Chicago, 1965, p.978.
2. B. J. Alder and T. E. Wainwright, Phys. Rev. Lett. 18, 988 (1967).
3. For an early reference see: K. Kawasaki, "Mode Coupling and Critical Dynamics" in *Phase Transitions and Critical Phenomena* Vol 5a, eds. C. Domb and M.S. Green (Academic, London, 1976).
4. E. Lorenz, J. of Atmospheric Sciences 20, 130 (1963).
5. J. Gil, C. Pfister, P. Coullet, M. Dubois and P. Berge, Phys. Rev. Lett. 66, [illegible] (1991).
6. M. J. Feigenbaum in *Trends and Perspectives in Applied Mathematics*, L. Sirovich ed. (Springer, NY, 1994).
7. D. V. Anosov, Proc. Steklov Inst. Math. 90, 1 (1967).
8. R. Bowen, J. Differential Equations 18, 333 (1975).