Simulation and Beyond: Computing in the 21st Century

John Guckenheimer

Computers have a remarkable ability to simulate natural and artificial phenomena. Industry, science, economic affairs and national security have come to rely upon simulation as an essential technology. We bet our lives on devices like fly-by-wire aircraft and digitally controlled pacemakers that require digital computation to function properly. We have replaced testing of nuclear weapons with an aggressive program of computer simulation, and international treaties on carbon dioxide emissions are based upon computer predictions of long term effects of human activity on global climate. The fidelity and reliability of computer simulation is a critical issue for many endeavours. As our computers become faster and cheaper, simulations become larger, more complex and more difficult to evaluate. The process of simulation itself becomes more diverse in ways described here. A conventional view of simulation is presented briefly. This is followed by a discussion of issues that extend beyond this traditional view.

The starting point for fluid simulation – and many other problems – is frequently expressed as the solution of systems of partial differential equations. We assume that these equations give a correct "first principles" description of a fluid flow. The key mathematical problems are the creation of algorithms that approximate the solutions of PDEs as the scale of discretizations tends to zero. As the resolution of the calculations becomes finer and more computational resources become available, we are able to solve these problems more accurately. This viewpoint is very prevalent, but it applies fully to only a limited set of problems. These are problems in which a precise description can be given with a data set of fixed size, in which the problem behavior is stable to perturbations of the magnitude of the errors inherent in the calculations and in which the simulation results are readily compared with test data. The issues that extend beyond these problems are considered below in terms of five dichotomies, using an example for each:

- First principles vs. phenomenonology

- Determinism vs. indeterminisms

- Continuous vs. discrete models

- Special structure vs. genericity

- Complexity vs. aggregation

We then discuss issues of numerical analysis that arise in simulation.

First principles vs. phenomenonology

Computational neuroscience is growing rapidly. Our efforts to model the brain are rooted firmly in biophysical models of membranes and the channels they contain, but there is a vast difference between compartmental Hodgkin-Huxley models of neurons and the Navier-Stokes equations of fluid dynamics. The Hodgkin-Huxley models are based upon sound biophysical principles, but these principles do not constrain the models to a definite set of equations in the same manner that a few assumptions about fluid properties lead to the Navier-Stokes equations. Instead, many aspects of the models depend upon approximations, choices of parametric forms of relationships among quantities and measurements that fit data to these parametric forms. The assumptions oversimplify and distort information that we know about these systems. However, when we try to increase the resolution of the models, then we increase the number of parameters that must be measured to fit finer models. This creates the need for more measurements, many of which are unlikely to be feasible. The measurements that can be made create voluminous data sets that need to be analyzed to extract useful information for parametrizing and initializing the models. Thus, it is hardly clear that efforts to increase resolution in these models will to better fits. Uncertainty about the values of additional parameters may prevent us from obtaining the improved fidelity that we expect to obtain from finer resolution models.

When data for initializing high resolution models can be collected, it is usually expensive. Some databases are maintained by the Federal government at great cost and made partially available for scientific purposes. However, large scientific collaborations like the human genome project or those centered around the NSF Long Term Ecological Research Sites are required to lay the substrate for detailed simulation of many natural processes. Building high fidelity models from data using phenomenological models requires planning and coordination greater than has been customary in most research areas. Successful efforts will require data standards and computational tools that are accepted by the researchers contributing to modeling efforts directly or indirectly. Agreement on such standards is likely to increase scientific orthodoxy, perhaps at the risk of thwarting individual creativity. Thus simulation may bear cultural costs as well as those of money and effort. This is

not an argument that we should avoid a focus on simulation, but rather a plea that we should examine the enterprise. As with large telescopes in astronomy, many areas of science stand to reap enormous benefits from thoughtful planning and investment in technology that will help simulation efforts.

Determinism

Is the weather predictable? There is abundant evidence that global forecast models display sensitive dependence to initial conditions. Perturbations within the measurement error of current observations lead to different forecasts. It is likely that this sensitivity is not an artifact, but an inherent property of atmospheric circulation. Let us assume here that the atmospheric flow is turbulent and that disturbances grow to effect global circulation. What are the implications for simulation? Dynamical systems analysis of chaotic systems gives insight into the answer to this question. It clarifies how unpredictability can arise in a deterministic, clockwork universe. Sensitive dependence to initial conditions is a property of many dynamical systems, from simple nonlinear mechanical linkages to population models. The chaotic nature of such systems has been studied extensively, to the point that there are solid mathematical foundations for quantifying their unpredictabiity using such quantities as invariant measures, entropy, Lyapunov exponents and fractal dimensions. Still, the application of these concepts to the simulation of complex systems remains problematic.

Faced with systems that display sensitive dependence to initial conditions, long term prediction of the full state of the system as a function of time is simply impossible. Weather forecasts five weeks in advance, let alone five years, can specify at best average properties of the weather. Operational weather forecasts have begun to employ ensemble forecasting based on simulations of several initial states. The usefulness of dynamical systems methods is problematic for simulations with attractors whose dimensions are large. Probabilistic methods about system behavior seem more suited to these situations, but assumptions about underlying statistical distributions are hard to verify. For example, the statistical properties of turbulent fluid flows related to coherent geometric structures remain a controversial subject. Monte Carlo methods embody systematic approaches to the modeling of stochastic systems. Producing high fidelity simulations in complex systems with many parameters using Monte Carlo methods is an even larger task than it is for deterministic models with stable asymptotic states. When faced with noisy,

unpredictable systems like the brain, we have only a primitive understanding of which aspects of dynamical behavior of complex systems we can hope to simulate.

From Model Architecture to Dynamics

Modern molecular biology has invested great effort to determine the reaction pathways of living organisms. This research produces graphs that show how important biological molecules are produced, modified and used to regulate physiological processes. Frequently missing from this work is an understanding of the kinetics of the pathways. We see information about the structure of the pathway, but we do not understand how it works. In particular, we have difficulty predicting the functional consequences of drugs, mutations or other modulatory effects on the network. For example, the deleterious side effects of calcium blockers used as cardiac drugs illustrate the importance of understanding biological kinetics. Molecular biologists have assumed that function will be evident from structure. They have made the elucidation of structure the principal goal of the subject. However, the case of neural networks demonstrates that a single morphological network may support a diverse set of behaviors. Molecular biologists increasingly recognize that systems modeling will facilitate their understanding of complex biological processes.

Apart from considerations of symmetry, we have had little success in relating system structure to function. Dynamical systems theory provides a context for guiding our intuition of dynamical phenomena that we expect to observe in generic systems. Using results from differential topology and singularity theory, we have a coherent view of phenomena that we regard as typical and phenomena that are exceptional. When there is symmetry in a system, we know how to modify the theory. We believe that there are architectural principles which are important in building robust complex systems, but there is little theory to support our intuition. In the context of specific applications such as electronic circuit design, we build hierarchical systems of astounding complexity with millions of elements. Concepts such as hierarchy and feedback control have not been incorporated into a general theory of nonlinear dynamical systems. Our lack of insight into how system architecture constrains dynamical behavior limits the power of simulation as a tool for studying complex systems.

4

Continuous, Discrete and Hybrid Models

High fidelity simulation of human walking is a demanding task. We perceive small variations in gait, and readily distinguish departures from normality. Attempts to build two legged locomotion machines have floundered over the issue of maintaining balance in dynamically unstable states. Computational models of walking must contend with impacts. These break stride into phases in which the primary physical forces acting on the body differ. Dynamical systems in which there are discrete events in the phase space that result in discontinuous changes of the underlying model are called hybrid systems. Hybrid models can have continuous and discrete components in space and time. For example, the engagement of gears changes the dimension of the phase space of a model of two rotating shafts. Most of the machines that we build are hybrid systems. The construction of full system simulations from simulations of components frequently introduces the need for hybrid models.

Theoretical models and computational tools for studying dynamical systems are framed in terms of continuous or discrete time, but seldom both. Theory and numerical methods that apply to discontinuous or singular systems are more limited than those that treat analytic or smooth systems. Consequently, the foundations for simulations of hybrid systems are shaky. Dynamical systems theory has sharpened our intuition about what types of phenomena we should expect to see in continuous phase spaces of systems that operate in continuous or in discrete time. Extending that understanding to hybrid systems is a barrier to confident simulation of hybrid systems. We are left with weaker intuition to guide design of machines and industrial processes that meet our desired specifications.

Scales and Aggregation

Decomposition of natural systems into different "scales" is one of the central tasks in producing high fidelity simulations. We seek to understand how macroscopic behavior results from physical laws that operate on smaller scales. For example, we would like to understand fracture in terms of atomistic properties of materials. Reducing all complex phenomena to atomic interactions is clearly a hopeless task. It is preposterous to model the effects of global climate change on natural populations and agriculture on an atomic scale. Recognizing when we can separate physical scales and encapsulate smaller scale information in models that operate at larger scales has

been a central issue within condensed matter physics. In studying population biology, economies, or the brain the issue of aggregating small scales is fuzzier and more challenging.

Only recently have we acquired the computing resources required to simulate detailed multi-scale models. One of the areas that is growing rapidly is the simulation of models with large numbers of components. In applications as varied as molecular dynamics, battlefield simulations and traffic flow, we build stochastic models from which we seek to observe emergent behavior at the system level. Such simulations form the basis for a whole set of computer games. The challenge with these efforts is to obtain results that fit the real world. We seldom know which details of component behavior are most significant for determining system properties, so it would be prudent to have systematic ways of evaluating the effects of uncertainty in model components upon system behavior. Such methods hardly exist at this time.

Numerical Analysis

Numerical implementation of dynamical systems models depends upon approximations that themselves are subtle. The simplest, most direct numerical integration algorithm (the Euler method) is subject to substantial errors. These errors can accumulate to give qualitatively incorrect predicitions about long time dynamics, as happens with the harmonic oscillator. Historically, numerical solution methods for differential equations addressed stringent limitations on the speed and cost of performing arithmetic. The dramatic improvements of digital computers during the past fifty years have completely transformed these parts of mathematics. The speed of computation in simulation of physical systems as a limiting factor has largely been replaced by issues such as memory hierarchies, round-off errors inherent in floating-point arithmetic, and extracting useful information from very large data sets.

Consider the problem of simulating electric power systems. Reliable electric power produced with minimal environmental impact is vital to the world today. Adequate capacity to handle anticipated loads and real-time monitoring of operations are essential for these systems. Simulating network models at first sight seems like a straightforward task in numerical integration. However, when we look a bit closer we find technical issues that are bothersome. One issue is that the equations for a network are naturally expressed as differential-algebraic equations rather than as ordinary differential equations. The mathematical theory of DAE's is more complex than that of

ODE's. For DAE's, not all initial conditions in the phase space are consistent with the equations. Moreover, there are points in phase space where the algebraic constraints inherent in the equations are satisfied, but there still are no solutions (or multiple solutions) with these initial conditions. The mathematical underpinnings of the theory of DAE's remain incomplete. Restrictive assumptions on models are required to guarantee that simulation algorithms will work. Still DAE's are common in engineering applications and cannot be ignored. In addition to power systems, DAE's arise naturally as models of constrained mechanical systems.

Beyond Simulation

Simulation viewed as the evolution of specific initial conditions for a dynamical computer model is unlikely to directly answer many of the questions that we ask. For example, consider the problem of fitting parameters to experimental data. In Hodgkin-Huxley models for neurons, a typical vector field may have a ten dimensional phase space and forty parameters. The model of a network of ten neurons built from single compartment Hodgkin-Huxley neurons will have a phase space of dimension approximately one hundred with several hundred parameters. If there are many model parameters that cannot be measured directly, then we are left with a complex "inverse" problem of using simulation data to optimize the parameters. In the case of the neural network, the problem is further complicated by the distortion of the primary voltage measurements that occurs due to unmodeled spatial effects in the system. This means that the most useful data for model comparisons is likely to be related to qualitative properties like the period of oscillations or the stability boundaries for different dynamical states as physical parameters are varied. Obtaining this information by sampling trajectories can only be done for a small number of parameters because the number of required trajectories grows exponentially with the number of parameters. Thus, solution of these parameter identification problems seems to require algorithms that go beyond simulation. The problems of fitting model parameters inhibit the creation of high fidelity models. As described previously, increasing model resolution to include smaller scales in a problem may increase the number of parameters that must be determined faster than the fidelity of the models improve.

Bifurcation theory for dynamical systems provides a framework for direct determination of information about how system behavior changes qualita-

tively with parameter variations. Implementation of algorithms based on this theory is a step towards computing parameter ranges that produce desired behavior. For bifurcations of equilibria, stability boundaries can be determined without numerical integration by formulating defining equations for bifurcations from the derivatives of the vector field. These methods have been implemented in a continuation setting to compute curves of codimension one Hopf and saddle-node bifurcations in two-parameter families of vector fields and curves of codimension two bifurcations in three-parameter families of vector fields. We need better algorithms to compute multi-dimensional continuation of submanifolds of bifurcations and to treat bifurcations of periodic orbits reliably.

There are additional geometric questions about dynamical systems that are important for varied applications. Computation of mixing properties of fluid flows has been greatly facilitated by regarding the instanteous velocity fields as generating a dynamical system of streamlines and computing invariant manifolds of these dynamical systems. The stagnation points of the fluid flow are saddle points of the dynamical system and their stable and unstable manifolds give separation boundaries for the fluid flow. Computing the intersections of these manifolds and the evolution of their turnstile structures gives approximations to the mixing properites of these fluids. These techniques have been used to investigate the design of industrial reactors, chemical reaction rates and fluid transport of ocean eddies. Because invariant manifolds become highly convoluted with sharp bends, substantial care is need to compute them accurately.

We desire the ability to robustly and routinely compute far more about the qualitative properties of dynamical systems than we can today. There are phenomena that play a prominent role in the qualitative theory that occur on very fine scales in many examples. Developing consistent, converged calculations requires that these scales be resolved. Since the phenomena often involve singularities and bifurcations, classical algorithms need to be modified and extended to work with these problems. Mathematical theory has guided this work, leading to the creation of algorithms that solve challenging problems. Unlike prevailing trends in computational science, the problems have often been small and the computing highly interactive. The interplay between classical and modern mathematics, geometry and numerical analysis and computational science will continue to be important to progress in the use of simulation as a powerful scientific tool.

8