THE PHYSICAL BASIS OF COMPUTABILITY

Simulations work in practice because they exploit higher-level organizing principles in nature. Good code writing requires faithfulness to these principles and the discipline not to exceed their limits of validity. An important exception is the use of simulation to search for new kinds of emergence.

Once, as a graduate student attending an international conference, I made the mistake of admitting during my talk that I could not calculate the optical absorption spectrum of a certain structural defect in an insulator. I meant that I could not calculate *accurately*, of course, but that did not matter. Before I had time to think or qualify my statement, an ambitious young assistant professor in the back leaped to his feet and yelled, "Maybe *you* can't calculate this spectrum, but *I can*!" His words ring in my ear to this day; one tends not to forget such things. I got through my presentation somehow, retired, and then later went about finding out how this guy had managed to do a computation I had found impossible.

It turned out to be a wild goose chase. He could no more do an honest calculation of that spectrum than I could. He had simply redefined "calculation" to mean a *post*diction of a complicated model with lots of parameters fit to the data. He had also hidden these weaknesses in a large, proprietary, poorly documented computer program, so they could not be discovered without considerable work. But I was motivated, had time on my hands, and was eventually able to get to the truth—and truth it was. To this day, no one has ever calculated these spectra correctly from first principles. Their shape is dominated by Franck-Condon broadening (correlated electronic and nuclear motion) and complex multielectron shake-off effects.[1] This calculation, like many others I have encountered in my career, is just too hard.

It is understandably difficult for any of us to admit that a calculation is just too hard. Large segments of our society now view the understanding of natural phenomena, in the sense of Bohr and Einstein, to be a quaint anachronism rendered obsolete by computers. I strongly disagree with this view, and can defend this position with sobering accounts of its failure, but it is part of our culture at the moment and something with which we have to live. From this perspective, a computer code or computational strategy unable to produce some essential result is just outdated technology, something to be supplanted shortly by either a market challenge or the next upgrade. None of us can afford to be uncompetitive, so none of us is anxious to admit weakness, even when it is true and even when the reason for the weakness is fundamental.

The fundamental limits to simulating physical phenomena by computer are real.[2] In contrast to

1521-9615/02/$17.00 © 2002 IEEE

ROBERT B. LAUGHLIN
Stanford University

the situation in business, economics, or law, which are inherently fuzzy disciplines, this is quite easy to demonstrate in physics. Consider, for example, the celebrated spin glass problem—a set of N half-integral quantum spins interacting by random Heisenberg exchanges. Because this system's configuration space has dimension 2^N, a straight solution of the quantum mechanics requires the diagonalization of a $2^N \times 2^N$ matrix, something we know how to do algorithmically. However, even for the case of $N = 200$, this matrix has $2^{400} = 2.6 \times 10^{120}$ elements, a number vastly larger than all the atoms in the visible universe. The computational task would obviously exceed the memory capacity of any conventional digital computer that could ever be built and is therefore fundamentally impossible. This example is not merely academic. Spin glasses exhibit an array of important thermodynamic behaviors—the glass transition, remanence, huge low-temperature specific heats—that do not occur in ordered antiferromagnets. Most of the fundamental limits to physical simulations we care about—fully developed turbulence, reaction chemistry, cuprate superconductivity, life, and so forth—are like this. They proceed from the configurational nature of matter, particularly quantum matter, to the conclusion that the size of the computation must increase exponentially with the number of degrees of freedom and rapidly become *qualitatively* larger than the resources of any imaginable computer. The importance and universality of this effect is beautifully articulated by P.W. Anderson in his famous paper, "More is Different."[3]

Most of us have an intuitive understanding of emergent laws and respect them even though we rarely talk about them.

A formally correct but physically problematic response to this argument is the quantum computer.[4] One way to "compute" the motion of the spins in the previous example is simply to initialize them physically and measure their behavior as a function of time using conventional spectroscopy. This could be done without employing all the atoms in the universe and would constitute one example of simulation by quantum computer—namely, time evolution of the system itself. The quantum computer evades the problem of thermodynamic size by being itself fundamentally different from conventional computers. Indeed, the unique ability of quantum computers to solve such problems is what makes them so interesting. However, the very property that allows quantum computers to solve such problems also makes them more noise-prone and unpredictable. This effect is the microscopic origin of the second law of thermodynamics and is fundamental. Whether it can be overcome technically is an important and deep question, but one that can be answered only by experiment. I am aware of no such technology on the immediate horizon.

In light of the configurational difficulty and the stupendous variety and unpredictability of the natural world, it is remarkable that anything can be simulated reliably. One is moved to ask why simulation is possible at all, and what distinguishes a system that can be simulated from one that cannot.

The physical world is simple because it is regulated by higher-level physical law.[5] The laws in question are collective in nature and emergent—meaning that they are exact only in the thermodynamic limit, encoded only indirectly by the underlying laws of quantum mechanics, and in a deep sense independent of them. The laws of hydrodynamics are a good example of this, as are the laws of crystalline elasticity, the laws of plasticity, the laws of superfluidity, the laws of magnetism, and the laws of thermodynamics. When hydrodynamics is working, say, in the collision or two very large nuclei, a hydrodynamic simulation can predict the scattering experiment's essential features without taking into account the equations of motion of individual nuclei.[6] When the laws of plastic flow are working, say, in the motion of a glacier, we can predict where the ice will flow without knowing where each atom goes.[7] It is the existence of these higher-level laws—the simple mathematical relationships among measured quantities created and enforced by emergent physical phenomena—that makes meaningful computer simulation possible.

Most of us have an intuitive understanding of emergent laws and respect them even though we rarely talk about them. In praising a good code we often say that it "captures the physics." We mean by this that the author skillfully exploited one or more legitimate collective organizational principles in writing the code, disciplined himself to stay within the limits of validity of those principles, and did not just make things up. The latter is centrally important because computer simulations, like the equations of mathematical physics on which they are based, are symbolic representations of physical law and are meaningful only insofar as they are faithful to that law.

The reliance of physical simulations on organizational principles for their validity means that they must be judged by higher standards than those we use for other software. Even if they work well and produce breathtaking graphics, simulations can be, and often are, wrong. Sometimes this does not matter, as in the recreational F-22 simulator my teenage son flies, and sometimes it does. When I first heard about the plan to design airplanes with electronic wind tunnels, for example, a happy vision flashed into my mind of sending the people responsible for this decision up on the first test flight.

When the validity of a simulation matters, one must be profoundly uncomfortable with its central physical principles being proprietary or relying on "tests" designed by the owner. We are only human, and there are just too many incentives and opportunities for misrepresenting the truth. These are sometimes subtle and difficult to detect. For example, one strategy is to write a complex code based on bogus principles, adjust its parameters to fit a handful of experiments, and then use comparison with one of these experiments as a "test." Another is to "test" the code in some extreme limit where it cannot fail and does not matter. Another is to define the difference between what the code actually produces and what it should have produced as acceptable error. Another is to benchmark the code against another code that is wrong in the same way. Another is to dismiss as wrong the experiment with which the code should have agreed. Another is to declare the code valid only in regimes that are experimentally inaccessible. I have had so much unpleasant personal experience with these and other failure modes that I have now come to automatically mistrust physical simulations not based on principles I fully understand and that are rigorously tested against experiment in their entire range of validity.

An important exception to this rule is the class of simulation that aims to identify new emergent principles. There are legendary successes of this approach in physics: Fermi, Pasta, and Ulam discovered the soliton effect in simulations of lattice thermization in the late 1940s.[8] Alder and Wainwright discovered the Kosterlitz-Thouless transition,[9] long-time tails, and atomic-scale hydrodynamics[10] in molecular dynamics simulations of fluids. Mitchell Feigenbaum discovered scaling on the period-doubling path to chaos in turbulence studies.[11] Edward Lorentz discovered the strange attractor and the principle of chaos while trying to model the weather.[12] The vast majority of commercial applications, such as urban growth,[13] war strategy,[14] crop forecasts,[15] and drug design searches,[16] are in this category. In these and similar applications the underlying physical basis is not fully understood, and simulation aims only for "some" quantitative analysis on the grounds that it is better than none. However, for such a simulation to be correct, as opposed to merely useful, it must predict one or more contentful experimental facts. *It is not generally possible to start from the wrong equations and get the right result*. This can only happen if the answer is insensitive to details, and therefore reliable, because of some as-yet-unknown higher-level emergent law. The identification of such laws by means of simulation is still in its infancy, and I believe it to be one of the great outstanding challenges for computer science. It is, however, potentially vulnerable to abuse, especially since the tests are hard to quantify and administer.

It is not generally possible to start from the wrong equations and get the right result.

I believe that the long-term health of computing as a branch of physical science is invested in the opposite approach—the repudiation of market-based science and a demand of total fidelity with established physical law. Mark Twain captured the problem well when he said that truth is always stranger than fiction because fiction is forced to stick to possibilities, while truth is not. Real science always begins with careful observations of nature and thoughtful consideration of facts that "ought not to be true" but nonetheless are. Studying only a simulated world based on what one *thinks* is true rather than what actually is true automatically precludes rethinking the facts, and therefore automatically precludes making a fundamental discovery. Klaus von Klitzing's discovery of the quantum hall effect is a beautiful case in point.[17] No one before von Klitzing had ever bothered to measure the hall conductance accurately because it was not supposed to be quantized.

It has become clear to many of us that the central task of physics in our time is the identification and enumeration of as-yet-undiscovered higher organizing principles of nature. This is a vast frontier and an incredibly exciting one, especially in light of developments in the life sciences. It is my great hope that simulation will play the important role in this drama that its nature and traditions suggest it should. Computers are enormously powerful tools, and they can do great good when used properly and wisely. Many lives are spared each year because strong winter storms can now be predicted a few days in advance. Computer modeling of the sun has led to the discovery of the neutrino flux deficit[18] and the likely detection of neutrino oscillations.[19] Modern quantitative hydrodynamic simulations have vastly decreased the cost of airplane design.[20] However, there is a serious danger of this power being misused, either by accident or through deliberate deception. All of us trained in science and concerned about the integrity of physical law must be committed to preventing this. As caretakers of a tradition that is perhaps the greatest contribution to humanity western civilization has ever made, we are obligated to remind our fellow citizens and the younger generations that real science deals with truth and is as different from falsehood as night is from day. Nothing that has happened, or will happen, in government or the economy can change this. After all of us are memories and the last grant proposal, program review, and IPO have passed into history, there will still be truth, reason, experimental discipline, and the majesty of physical law.

References

1. G.D. Mahan, *Many-Particle Physics*, 2nd ed., Plenum, New York, 1990.
2. N. Orsekes, K. Shrader-Frechette, and K. Belitz, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science*, vol. 263, no. 5147, 1994, pp. 641–646.
3. P.W. Anderson, "More is Different," *Science*, vol. 177, no. 4047, 1972, pp. 393–396.
4. M.A. Nielson and I.I. Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press, Cambridge, UK, 2000.
5. R.B. Laughlin and D. Pines, "The Theory of Everything," *Proc. Nat'l Academy Science,* vol. 97, no. 1, Jan. 2000, pp. 28–31.
6. U. Ornik, F.W. Pottag, and R.M. Weiner, "High-Energy Reactions and Relativistic Hydrodynamics in Three Dimensions," *Physical Rev. Letters*, vol. 63, no. 24, 1989, pp. 2641–2644.
7. G.A. Olyphant, "Computer Simulation of Rock-Glacier Development under Viscous and Pseudo-Plastic Flow," *Bull. Geological Soc. Am.*, vol. 94, 1983, pp. 499–505.
8. E. Fermi, J.R. Pasta, and S.M. Ulam, "Studies of Nonlinear Problems," *Sets, Numbers, and Universes*, S. Ulam et al., eds., MIT Press, Cambridge, Mass., 1974, pp. 491–501.
9. B.J. Alder and T.E. Wainwright, "Phase Transition in Elastic Disks," *Physical Rev.*, vol. 127, no. 2, 1962, pp. 359–361.
10. B.J. Alder and T.E. Wainwright, "Decay of the Velocity Autocorrelation function," *Physical Rev. A*, vol. 1, no. 1, 1970, pp. 18–21.
11. M.J. Feigenbaum, "Quantitative Universality for a Class of Nonlinear Transformations," *J. Statistical Physics*, vol. 19, no. 1, 1978, pp. 25–52.
12. E.N. Lorentz, "Deterministic Nonperiodic Flow," *J. Atmospheric Science*, vol. 20, 1963, pp. 130–141.
13. R. Liggett, "Automated Facilities Layout: Past, Present, and Future," *Automation in Construction*, vol. 9, no. 2, 2000, pp. 197–215.
14. S.I. Irwin, "Commanders Want Realistic Simulations," *National Defense Magazine*, Apr. 2001; www.nationaldefensemagazine.org.
15. Y.W. Jame and H.W. Cutforth, "Crop Growth Models for Decision Support Systems," *Canadian J. Plant Science*, vol. 46, no. 1, Jan. 1996, pp. 9–19.
16. R.S. Dewitte, A.V. Ischenko, and E.I. Shakhnovich, "De Novo Method Based on Simple, Fast, Accurate Free Energy Estimates 2: Case Studies in Molecular Design," *J. Am. Chemical Sciences*, vol. 119, no. 20, 1997, pp. 4608–4617.
17. K. von Klitzing, G. Dorda, and M. Pepper, "New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Resistance," *Physical Rev. Letters*, vol. 45, no. 6, 1980, pp. 494–497.
18. J.N. Bahcall, "Solar Neutrinos: Where We Are, Where We Are Going," *Astrophysical J.*, vol. 467, 1996, pp. 475–484.
19. Y. Fukuda et al., "Evidence for Oscillation of Atmospheric Neutrinos," *Physical Rev. Letters*, vol. 81, no. 8, 1998, pp. 1562–1567.
20. R. Agarwal, "Computational Fluid Dynamics of Whole-Body Aircraft," *Ann. Rev. Fluid Mechanics*, vol. 31, 1999, pp. 125–169.

Robert B. Laughlin is a professor of physics and applied physics at Stanford University. His research interests include electron motion in mott insulators, principles of emergence, connections between the quantum mechanics of ordinary matter and the vacuum, and bioinformatics. He received an AB in mathematics from the University of California, Berkeley, and a PhD in physics from the Massachusetts Institute of Technology. He is a member of the National Academy of Sciences and the winner of numerous prizes and awards, including the E.O. Lawrence Award for Physics, the Oliver E. Buckley Prize, the Franklin Medal for physics, and the Nobel Prize. Contact him at the Geballe Lab. for Advanced Materials, Rm. 342, McCullough Bldg., 476 Lomita Mall, Stanford, CA 94305-4045; rbl@large.stanford.edu.