LIMITS ON COMPUTATIONS

E ach year, computers grow more powerful, and we use them to solve increasingly complex and important problems. In this issue, we explore some limits to this growth.

The steady improvement of computer capability that Moore's Law characterizes—speed or storage increase as time$^{1.7}$(years)—has resulted in the wide availability of large, powerful computers. For example, a computer with a 1.6-GHz processor, 1 Gbyte of RAM, and a 100-Gbyte hard disk, delivered overnight, costs less than US$1,200—approximately 0.01 percent of the cost of a 1990 supercomputer. Extrapolating this performance, computers should be able to handle problems that involve all the universe's particles—10^{80}—in about 300 years. What are some of the limits to this exponential growth?

Research

The explosion in computational power has driven the widespread use of computational analysis and prediction in almost all areas of engineering and the hard and soft sciences. We now do engineering design, analysis, and product manufacturing control almost entirely with computers. Theoretical physicists and chemists have forsaken pencil and paper to develop computational models that describe, analyze, and predict the physical world's behavior. Examples abound of "first principles" models for atomic structure, material behavior, and turbulence. Researchers are using complex multiphysics simulations to predict the weather and design airplanes, jet engines, and nuclear weapons.

Experimental research also relies on computerized data collection and analysis. Social scientists—from economists to biologists—use computers to mine statistical data on human and animal behavior, develop models to predict the economy's behavior, and simulate battlefield conditions, traffic patterns, and social behavior. But how close can we really come to computational models that capture a full understanding of the complex world of natural phenomena or human events?

In addition to finite computer resources, many

DOUGLASS E. POST
Los Alamos National Laboratory
FRANCIS SULLIVAN
IDA Center for Computing Sciences

aspects of the models themselves limit computational simulation accuracy and fidelity. We can only carry forward simulations that amplify errors in initial conditions to the point where the growth in the errors becomes comparable to the solution. Longer-running solutions require finer resolution in initial conditions. Simulations such as ground water transport, where it is impossible to obtain accurate, high-resolution initial conditions, are not highly predictive. We can't class as truly predictive simulations in which we cannot accurately specify the interactions between the calculation's elements, such as traffic flow, economic behavior, and so on, even if they often give insights and predict trends.

This issue

Three articles in this issue explore a few of the intrinsic limits of computational tools for the hard and soft sciences. At some future point, Moore's Law will saturate. Higher component density on processor chips helps achieve higher clock speeds and increased processor complexity and memory size. However, the continuing reduction of feature sizes on integrated circuits—now close to 100 nanometers—continually increases the lithographical challenge for chip manufacture and the difficulty of keeping the heating power density on the chip within acceptable bounds. The computer industry will undoubtedly continue to overcome these near-term technological limitations, but there are some fundamental limits. For example, feature sizes will likely never be smaller than an atom (0.1 nanometers).

Michael Frank explores several basic limits in his article, "The Physical Limits of Computing." In a particularly clear exposition of the thermodynamic and quantum limits of information storage, he concludes that a storage density of approximately 1 bit/Å^3—40 years away, following Moore's Law—will likely be an upper bound. Following Moore's Law, in 2035 we should reach the thermodynamic limit of 0.7 kT for heat generation involved in storing a bit of information in memory and destroying the information already there. We can avoid this limit if we develop storage techniques that don't involve destroying previously stored information. Processor clock rates will not exceed the limit set by the maximum rate for atomic transitions. A realistic limit is about 10^{15} Hz—roughly 10^6 above present rates and 30 years away. It is interesting that these limits are all about 35 years away.

The other two articles explore some limits due to algorithmic complexity and the inherent limitations of computer models applied to the natural world. In "The Physical Basis of Computability," Robert Laughlin puts forward the view that "first principles" calculations of physical phenomena are fundamentally impossible. Detailed, correct calculations of complicated physical systems are almost always computationally intractable. The only simulations that have the potential to be predictive are those that exploit higher-level organizing principles in nature such as thermodynamics or general conservation laws. For example, one could never compute the flow of the roughly 10^{40} water molecules in a river by simulating the interaction of each molecule with the other molecules. However, the usual equations of hydrodynamics that embody conservation of particles, momentum, and energy do a good job until the flow becomes turbulent. He also distinguishes between simulations based on physical laws and the simulations typical in the social and biological sciences. Laughlin pays particular attention to the dilemma that we face with research papers based on computational results. Unless we reproduce the calculation—usually not realistically possible—we must rely on the author's competence for the results' validity. All too often, they are erroneous or overstated. Laughlin stresses the special responsibility of computational scientists to ensure that their results are correct.

"Computational Complexity for Physicists," by Stephan Mertens, introduces the concepts of tractable and intractable computational problems where the complexity due to interactions among the problem elements determines the calculation's scale. If the required size or time for a computation scales as n^k, where n is the problem size and k is a real number—usually less than 10—the problem is tractable. If, on the other hand, the scaling is exponential, the problem is intractable. Mertens amplifies on the characteristics of tractable and intractable problems, how they are related, and how we can sometimes make intractable problems tractable. ⬛

Douglass E. Post works in the Applied Physics Division at Los Alamos National Laboratory. He is also an associate editor in chief of *Computing in Science & Engineering*. Contact him at post@lanl.gov.

Francis Sullivan is the editor in chief of *Computing in Science & Engineering*. He is also the director of the IDA Center for Computing Sciences. Contact him at fran@super.org.