Souping up Supercomputing

Retooling the underpinnings of high-performance computing

By JANET RALOFF

Computers and the information industry that they've spawned drive more and more of the U.S. economy. During the past 5 years, production of computers, semiconductors, and communications equipment in the United States has quadrupled—a rate of growth more than 10 times that of the industrial sector in general.

"Over the past 3 years, information technology alone has accounted for more than one-third of America's economic growth," noted Vice President Al Gore in an address at the American Association for the Advancement of Science (AAAS) annual meeting last month in Anaheim, Calif.

Most recent advances in this industry derive from investments in fundamental computer science made 30 or 40 years ago, according to a report prepared last August by the President's Information Technology Advisory Committee (PITAC).

This panel of industry leaders and academic scientists noted that, over the intervening years, the federal government and the information industry have made steady investments in computer research. However, PITAC concludes that both sectors have "compromised" the return on those investments by their continuing shift toward applied research—efforts that focus on explicit, near-term goals.

A host of other researchers chorused similar concerns at a series of workshops last year. Fundamental computer-science research has not been keeping pace with the growth of the industry, they argued, or with its ability to churn out ever-faster computer chips.

Already, several U.S. national laboratories have systems that can perform a trillion operations per second (SN: 12/12/98, p. 383). These systems are known as teraflops (for trillion floating point, or real number, operations). The next decade promises "petaflops" machines, which will crunch numbers 1,000 times faster than teraflops (SN: 4/15/95, p. 234).

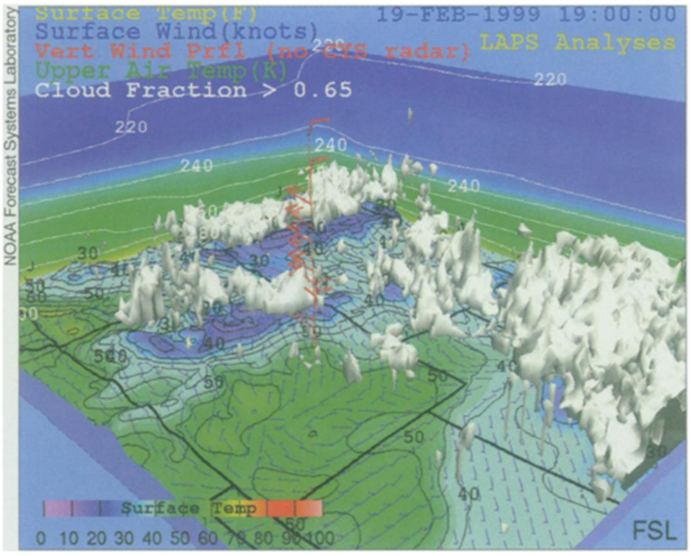
It's clear that this hardware "has gotten well ahead of the software and of the ways that we organize information-storage capacity," says James Langer, the University of California, Santa Barbara physicist who chaired a national workshop on advanced scientific computing last July.

The gap between hardware and the software that runs it has reached a point where "we don't understand at a scientific level many of the things that we're now building," observes George Strawn, acting director for computer and information science and engineering at the National Science Foundation (NSF) in Arlington, Va.

To better understand these machines and what it will take to build and effectively use even more powerful computers, Gore unveiled plans for a federal initiative. Called Information Technology for the 21st Century—or IT^2—it would boost federal support for fundamental computer science. Of the roughly \$1.8 billion in federal spending on computer research slated for the coming fiscal year, \$366 million in new funding would support IT^2, according to the President's proposed budget (SN: 2/6/99, p. 87).

The program has strong support in the powerful information-technology industry and the research community at large, according to presidential science adviser Neal Lane. At a AAAS press briefing immediately following Gore's announcement of IT^2, Lane noted that the initiative's broad outline had been drafted to deal with specific problems that have been identified both by PITAC and researchers who depend on high-performance computing. They span the disciplines from particle physics to pharmacology.

Indeed, IT^2 has had broader input from the scientific community than any research initiative in history, according to NSF Director Rita Colwell. "That's good," she adds, "because this is the most important initiative, in my view, that will be launched in the 21st century."



NOAA Forecast Systems Laboratory

Supercomputer-compiled 3-D view of weather over Colorado, created by the National Oceanic and Atmospheric Administration's Local Analysis Prediction System. The image combines data from surface stations, aircraft, wind profilers, Doppler radar, and satellites. Horizontal surface and vertical back walls are color coded to denote temperature. Barbs show wind direction and speed. White areas are clouds. To effectively model regional climate changes, the Department of Energy says the resolution of such simulations must improve by a factor of 10.

Computers have traditionally tackled problems serially, as a sequence of related steps. However, Strawn notes, "you just can't build a serial computer that's big enough and fast enough to attack some of the huge supercomputing problems that we're addressing now, such as good prediction of tornadoes or the simulation of combustion."

Supercomputer designers have therefore been making a general transition to parallel computers. These lightning-quick systems divide a huge computing problem into small elements. Linked computers then work on them simultaneously, eventually integrating the results.

"Today's really big computers are being put together from assemblages of desktop-type systems," Strawn says. Their hundreds to thousands of networked computers don't even have to share the same address. When the software uniting them works effectively, such a distributed supercomputer can span the globe (SN: 2/21/98, p. 127).

"For the past 5 years," he says, "we've been experimenting with and developing such distributed, high-performance computing facilities." What those efforts have driven home is how hard it is to make them work as one, Strawn says. "Clearly, there are still plenty of fundamental understandings that elude us on how to do highly parallel programming."

He notes that PITAC, recognizing this, said that "the first three issues it wanted us to focus on are software, software, and software."

The demand for software far exceeds the nation's ability to produce it, PITAC found. It attributed this "software gap" to a number of issues, including labor shortages, an accelerating demand for new programs, and the difficulty of producing new programs—which PITAC described as "among the most complex of human-engineered structures."

When a software program is released, PITAC found, it tends to be fragile—meaning it doesn't work well, or at all, under challenging conditions. Programs often emerge "riddled with errors," or bugs, and don't operate reliably on all of the machines for which they were designed.

Contributing to all of these problems is the tendency for the complexity of a software program to grow disproportionately to its size. "So if one software project is 10 times bigger than another, it may be 1,000 times more complicated," notes Strawn. Huge programs therefore "become increasingly harder to successfully implement."

The solution, he and many others now conclude, is that the writing of software codes "has to be transformed into a science" from the idiosyncratic "artsy-craftsy activity" that characterizes most of it today. If that can be achieved, he says, "we should be able to create a real engineering discipline of software construction."

Establishing such a science will be among the primary goals of IT^2, Lane says. One dividend of that pursuit, he believes, will be the emergence of software modules—large, interchangeable, off-the-shelf chunks of computer code that can be selected and shuffled to reliably achieve novel applications. Automakers can today order standard nuts, bolts, mufflers, spark plugs, and pistons to build a new car. "We don't have that in software," Lane says, "but we're going to."

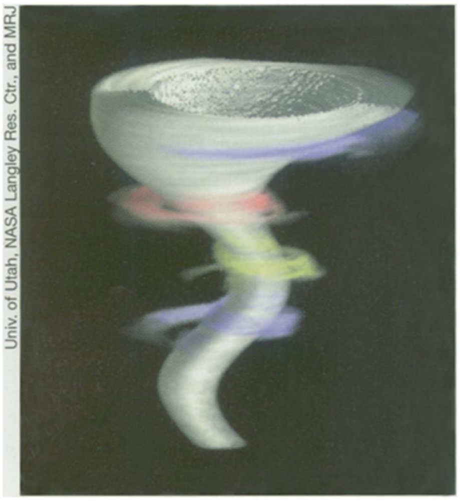
San Diego Supercomputer Center

The San Diego Supercomputer Center hosts the world's largest data-storage system. This peek inside one of its three tape libraries reveals the robot arm (center) used to retrieve 10-gigabyte cartridges of filed data.

At the same time, IT^2 will be probing new means to test software, notes Jane Alexander, acting deputy director of the Defense Department's Advanced Research Projects Agency in Arlington, Va. At issue, she says, is how quality-control engineers can debug programs that may contain many millions of lines of software code. Such debugging will prove vital if huge codes come to determine the safety of flying in a jumbo jet or the ability to reliably direct missiles away from civilian centers.

The IT^2 initiative will also spur research in other areas integral to harnessing supercomputers, such as the development of technologies to manage and visually represent data.

Like software, these technologies lag far behind today's sophisticated computer chips. The shortcomings already threaten to hobble Department of Energy programs. That department plays a lead role in modeling complex phenomena including climate, nuclear detonations, and chemical reactions unleashed by burning fuels.



Univ. of Utah, NASA Langley Res. Ctr., and MRJ

Supercomputer animation of a tornado, based on a new wind-vector visualization process. It allows researchers to probe aspects of the turbulence by injecting colored "dyes" that then flow with the local winds.

The mind-numbing complexity of these simulations has pushed DOE to the forefront of supercomputing—and up against the field's data-management limits—notes Michael Knotek, program adviser for science and technology for DOE.

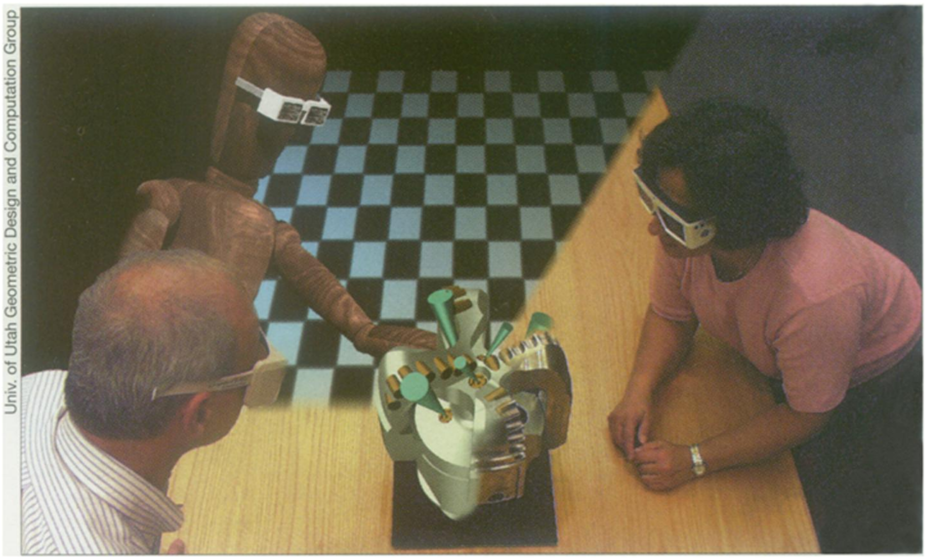
Today's supercomputers spit out files of gigantic size. The new teraflops machines will bump up data-storage needs even more. Computer scientists expect the machines to generate terabytes of data per hour, Knotek says—or enough daily to fill the equivalent of 1 million desktop-computer hard drives.

The largest archival storage system in existence holds just 86 terabytes of data. "We're going to need to hold tens or hundreds of petabytes," Knotek says. Without question, "this will require new technology."

Storing all of these data will be pointless, however, if it isn't cataloged so that it can be easily retrieved. New techniques and software will have to be developed for managing these data libraries and mining nuggets of useful information from them.

Even this challenge pales, however, when compared with figuring out how to display such massive amounts of data in a way that humans can meaningfully comprehend. For instance, a high-density computer monitor can display 1 million pixels, or display elements, on its screen. Attempting to depict a terabyte of data would require assigning 1 million data points to each pixel—a fruitless exercise, Knotek explains.

One virtual-reality display technology being developed to cope with large data sets goes by the name of CAVE, for Cave



With "telepresence" technology, depicted here, computers portray off-site, telecommunicating colleagues as avatars—virtual-reality figures with synchronized facial expressions, hand gestures, and audio responses. Today's telepresence systems remain quite primitive and can host only a few avatars.

Univ. of Utah Geometric Design and Computation Group

Automatic Virtual Environment (SN: 11/12/94, p. 319). It projects into a room a three-dimensional image of data from a computer simulation. A viewer wears special goggles to see the projection in 3-D and a headset to tell the system where the individual is relative to the depicted scene. These gadgets allow the viewer to walk within a CAVE to examine the data from many angles and probe different variables.

While studying gases swirling in a combustion chamber and chimney, for instance, the viewer might alter the flame temperature or the position of baffles and then watch how this changes gas eddies or the generation of pollutants.

Renderings of such scenes in today's CAVEs look cartoonish, and the views are limited. The ultimate goal is a realistic rendition of some simulated environment—akin to scenes depicted by the Holodeck in *Star Trek: The Next Generation* television series. Ideally, such a system should simultaneously afford many linked viewers a full-sensory Holodeck experience, including the sounds, feel, and smell of a simulated environment.

The new initiative will also tackle a host of other challenges, Lane says, such as the development of new computer hardware architectures, language-translation strategies (SN: 3/8/97, p. 150), and technologies that make computing easier. The last might include better programs to recognize voice commands or programs that better hide from the user's view the complexity of a computer's activities. At the touch of a button, for instance, programs might not only surf the Internet to find desired information but also assemble it into an easy-to-understand report.

Langer advocates that developers of these new technologies should work hand-in-hand with the scientists who will use them. This should ensure "that we focus on the right science problems, the right engineering problems, and the right computer problems." In the absence of such cooperation, he argues, a lot of money could be spent "to make a toy—something that makes pretty pictures but doesn't advance our science."

Similarly, there is always the risk that quantitative changes in computing won't bring along important advances—that "we might just use our new teraflops computers as big gigaflops machines"—observes Steven Koonin, a particle physicist and provost of the California Institute of Technology. "And until about 6 months ago, we were," he says. "Now, people are starting to understand the capabilities of these machines and to use them in qualitatively different ways."

One example, he says, is that "we're finally starting to get some real science from some simulations in the [nuclear-weapons stewardship] program that you could never have gotten with a gigaflops machine."

Part of what it takes to make that leap in effectively harnessing a new generation of supercomputers is the assembling of cadres of specialists, much the way hospitals now bring together teams of experts to consult on thorny medical cases, Koonin says. The day of the general-purpose computer scientist is gone. No individual has the vision to take in and comprehend all the vistas these computers are now presenting, he argues.

Such new collaborations will be necessary, Lane and Colwell agree, to deliver the type of novel research that PITAC called for—"groundbreaking, high-risk–high-return research . . . that will bear fruit over the next 40 years."

Colwell concludes, "When people ask, Why invest in the IT^2? I say it's absolutely a must . . . a national imperative." □