

# Taking on the ITER Challenge, Scientists Look to Innovative Algorithms, Petascale Computers

By Michelle Sipics

The promise of fusion as a clean, self-sustaining and essentially limitless energy source has become a mantra for the age, held out by many scientists as a possible solution to the world's energy crisis and a way to reduce the amounts of greenhouse gases released into the atmosphere by more conventional sources of energy. If self-sustaining fusion reactions can be realized and maintained long enough to produce electricity, the technology could potentially revolutionize energy generation and use.

ITER, initially short for International Thermonuclear Experimental Reactor, is now the official, non-acronymic name (meaning "the way" in Latin) of what is undoubtedly the largest undertaking of its kind. Started as a collaboration between four major parties in 1985, ITER has evolved into a seven-party project that finally found a physical home last year, when it was announced that the ITER fusion reactor would be built in Cadarache, in southern France. (The participants are the European Union, Russia, Japan, China, India, South Korea, and the United States.) In May, the seven initialed an agreement documenting the negotiated terms for the construction, operation, and decommissioning of the ITER tokamak, signifying another milestone for both the project itself and its eventual goal of using fusion to facilitate large-scale energy generation for the world.

Problems remain, however—notably the years, and perhaps decades, of progress needed to attain such a goal. In fact, even *simulating* the proposed ITER tokamak is currently out of reach. But according to David Keyes, a computational mathematician at Columbia University and acting director of the Institute for Scientific Computing Research (ISCR) at Lawrence Livermore National Laboratory, the ability to perform such simulations may be drawing closer.

## Hardware 3, Software 9

"Fusion scientists have been making useful characterizations about plasma fusion devices, physics, operating regimes and the like for over 50 years," Keyes says. "However, to simulate the dynamics of ITER for a typical experimental 'shot' over scales of interest with today's most commonly used algorithmic technologies would require approximately  $10^{24}$  floating-point operations." That sounds bleak, given the 280.6 Tflop/s ( $10^{12}$  flops/s) benchmark performance of the IBM BlueGene/L at Lawrence Livermore National Laboratory—as of June the fastest supercomputer in the world. But Keyes is optimistic: "We expect that with proper algorithmic ingenuity, we can reduce this to  $10^{15}$  flops."

Optimizing the algorithms used, in other words, could lower the computing power required for some ITER simulations by an astounding nine orders of magnitude. Even more exciting, those newly feasible simulations would be at the petascale—ready to run on the petaflop/s supercomputers widely expected within a few years.

The ingenuity envisioned by Keyes even has a roadmap. Together with Stephen Jardin of the Princeton Plasma Physics Laboratory, Keyes developed a breakdown that explains where as many as 12 orders of magnitude of speedup will come from over the next decade: 1.5 from increased parallelism, 1.5 from greater processor speed and efficiency, four from adaptive gridding, one from higher-order elements, one from field-line following coordinates, and three from implicit algorithms.

The increased parallelism, processor speeds, and efficiency that together account for three orders of magnitude are hardware rather than software improvements; Keyes and Jardin based their roadmap estimates simply on trends over the past decade that are extrapolated in a document known as the semiconductor industry roadmap. Keyes quickly gets to the software side of things by explaining the potential four orders of magnitude improvement from adaptive gridding.

Conceptually, he says, "adaptive gridding means not putting observers where they're not needed." In a uniform grid whose resolution is dictated by the smallest feature, "start taking away points where the solution is smooth, until further loss would endanger its resolution." In certain cases, he says, the volume of the domain that dictates the finest resolution may be as little as 1% and the refinement may be required in only one or two dimensions.

Using high-order elements, Keyes says, could also lower the performance requirements for some ITER simulations—in essence, a given accuracy can be obtained with fewer discretization variables and less overall work. He gives an example of a fourth-order method in which error decreases like  $h^4$  ( $h$  being a typical mesh spacing), compared with a second-order method in which the error decreases like  $h^2$ .

"Suppose we can tolerate an error of about  $10^{-4}$ ," he says. "We can get this with about 10 points in the high-order method, but it takes about 100 points with the low-order method [per dimension]." In higher spatial dimensions, he says, "this ratio of ten in savings powers up."

Keyes identifies an additional advantage: "Higher-order methods link together fewer points with greater coupling." As a result, "there are fewer memory loads and stores to do, and more flops to do between loads and stores, which plays to the advantage of today's computer architectures." However, he cautions: "The best algebraic solution methods we have today have a much greater difficulty with high-order discretizations than with the simplest low-order discretizations."

Discretizing the domain with field-line-following coordinates can contribute an order of magnitude by lowering the resolution requirements in certain directions, says Keyes. In a tokamak the most rapid changes in the fields tend to occur across magnetic flux surfaces rather than along them.

For a final three orders of magnitude, Jardin and Keyes turned to implicit algorithms. Explaining the basis for the potential improvement, Keyes cites the CFL (Courant–Friedrichs–Lewy) criterion, in particular the finding from the landmark 1928 paper that “small errors can be amplified to arbitrarily large errors in timestepping algorithms in which the computational timestep exceeds a certain size.” A large number of computational timesteps are required to get to a certain physical time limit, Keyes explains. And the size of the maximum computational timestep relates to two other parameters: the smallest physical mesh step and the fastest wavespeed supported by the differential equation.

“For magnetohydrodynamic problems typical of tokamaks, we should not have to resolve the fastest waves, which for ITER are as much as a million times faster than time scales of interest,” he says. “To avoid the catastrophic effects of numerical instability, however, we need either implicit methods, or frightfully many explicit timesteps.” Implicit methods are expensive, he continues, but not nearly as expensive as timestepping beneath the CFL stability limit. If an implicit timestep costs a thousand times as much computation as an explicit timestep, but a million fewer implicit timesteps are required, three orders of magnitude can be saved.

But why focus on algorithms? Why not concentrate on hardware improvements to achieve faster supercomputers? Pointing to the breakdown of the 12 orders of magnitude of speedup to be gained, Keyes gives a final score: hardware 3, software 9.

“Since the beginning of computing, algorithmic innovation has been much, much more important than improvements in computer hardware engineering,” Keyes says. “Pick any starting point, any year you want in any simulation field, and I bet someone can make the case that algorithmic advances have outpaced Moore’s law in getting more science done per second.”

### **Moore’s Law for the Petascale Era**

Hardware advances are clearly a driving force behind what scientists call “the dawn of the petascale era.” But in some ways, says Caltech computer scientist Thomas Sterling, hardware improvements are a double-edged sword.

“It’s ironic that the improvements garnered through Moore’s law also make it increasingly difficult to benefit from the same technology for high-end computing,” he says, citing the exponentially increasing memory capacity in DRAM (dynamic random access memory) chips. “Even with improved data access to these components, the amount of time required, measured in processor clock cycles, to touch every byte on a memory chip once continues to increase,” he says, “in essence slowing down at least those simulations that are memory constrained.”

Moore’s original assertion, in a 1965 article in *Electronics Magazine* titled “Cramming More Components onto Integrated Circuits,” was that the density of transistors on a chip doubles approximately every two years. Frequent casual references in the media have directly linked the “law” with performance, but, as Sterling’s statement reflects, the number of transistors on a chip is not directly related to the number of flop/s a system can perform. The number of transistors is important, but the overall performance of the system also depends on how efficiently each transistor can be utilized.

The memory constraints that Sterling refers to, often called the “memory wall,” are a case in point of the good news/bad news element of Moore’s law. Current DRAM chips can accommodate huge numbers of transistors and capacitors, giving them enormous memory capacity—but researchers have yet to devise a rapid method for accessing this data. The resulting memory wall is one of the largest obstacles to high efficiency in large-scale systems.

“It takes approximately 100 times longer to access an 8-byte word from main memory than to perform an arithmetic operation on it in the CPU,” Keyes explains. “Hence, most CPUs operate at only slightly over 1% efficiency on average, with much higher efficiencies—close to 100%—on codes structured to exploit cache locality. Accessing cache typically takes only a few processor cycles.”

Nanotech memory, which would store information in carbon nanotubes, should make all memory appear as close as cache once it becomes available for on-chip integration, Keyes says. But “nanotech memory is not even on the drawing board of the BlueGene designers at IBM.”

When it comes to addressing such problems, Sterling sees a few possibilities as worth consideration. One specific change that he describes as not popular but probably inevitable is a move to merge CMOS logic and DRAM memory on the same chip.

“That could greatly expand the available local memory bandwidth while reducing the access latency,” Sterling says. He also cites significant reductions in power consumption as a benefit of such a change, through the avoidance of going off-chip and the use of simpler processors. (Because of the energy required to operate and cool a machine’s components, power consumption is an extremely important issue in high-end computing.)

### **Achieving a Hardware/Software Balance**

Sterling doesn’t disagree with Keyes’s assertion that software can greatly improve the performance of high-end computers—at least once current systems have been redesigned to give programmers a better hardware basis from which to start.

“Many consider the parallel computing problem to be largely a software problem, and there are good reasons for that viewpoint,” he says. “However, in spite of a decade of investment by government and industry in system software and tools for parallel processing, the big breakthrough has yet to be realized.”

During an interview at the 2006 NEC User Group meeting in Toronto, Sterling stressed that nearly all current high-performance computers are designed with essentially no scalability, forcing programmers to address latency, contention for resources, and similar problems in excruciating detail.

“We are not, in fact, programming parallel computers but rather attempting to domesticate large herds of sequential processors, not designed for the scalable domain,” he says. “Without components designed by intent to enable parallel computing, it’s no wonder that our current processor ensembles are such a challenge to coerce into effective parallel computing.”

Advances to be made in architecture design, according to Sterling, will address some of the major difficulties with software development for large-scale systems.

“It’s the balance between hardware and software that needs to be addressed, and from such an improved relationship can come the derivation of both,” he says.

In the race to deploy the world's first relatively general-purpose petascale system, Sterling's current best guess is that the U.S. will get there first. "But I would not be disappointed should the honors go to Japan, the EU, or to anyone else for that matter," he says. "The world needs such systems, and many of them."

The ITER project will certainly be among the beneficiaries of such systems. Keyes points out that if he is right and algorithmic improvements can bring ITER's simulation needs down to the petascale, "we could simulate one ITER operating cycle in about one second. That's probably enough for making reasonable progress in designing and operating ITER."

Still, he admits, although specific applications can be suggested for petascale machines, there is nothing particularly special about petaflop/s. "It is just one milepost along a long path that has stretched from one Gflop/s in 1998 to 100 Tflop/s in 2005. People get inordinately excited when the prefix changes, as when a car turns 100,000 miles," he says. "But each mile is the same as the one before."

Regardless, as high-end machines roll toward 1,000,000,000,000, you can be sure mathematicians will be watching.

*Michelle Sipics spent the summer as an intern at SIAM News and is now a contributing editor.*

---

## Getting the Most Out of a High-End Computer

In a recent interview\*, Pittsburgh Supercomputing Center co-director Michael Levine addressed a discrepancy in the performance of two of the center's supercomputing systems. One of them, a 10-teraflop/s Cray XT3, has a processor clock speed just about two and a half times faster than the other, a 6-teraflop/s LeMieux system—yet the XT3 outperforms LeMieux by as much as ten times for some applications. How does the PSC account for the gap in the performance of the two systems in such cases—and if the XT3 does indeed perform so much better, why even bother using LeMieux?

Levine listed three elements that contribute to the discrepancy: the differences in interconnect systems, memory bandwidth, and software between the two systems. In particular, he pointed out that the LeMieux operating system is responsible for supporting each processor individually. In the XT3, by contrast, the OS isn't designed to support the machine as a whole—that task is mostly assigned to certain processors within the system, leaving the OS free to "concentrate" on calculating.

But that doesn't mean that LeMieux is inferior to the XT3. LeMieux lags significantly behind in some situations, but might equal or outperform the XT3 for problems best solved by a system in which each processor operates independently.

Levine's comments and the differences between the two supercomputers reflect a larger issue in high-performance computing: What defines performance? With terascale comput-

ing already a reality and petascale on the horizon, have flop/s-based performance metrics outlived their usefulness?

While the Top500 project still ranks the performance of supercomputing systems based on flop/s using a specialized version of the LINPACK benchmark, LINPACK creator and Top500 VIP Jack Dongarra is among the experts who no longer consider the benchmark a useful real-world test. (In fact, he says, "I was probably the first one to say that.") The benchmark used in the Top500 ranking measures performance (in flop/s) in the solution of a dense system of linear equations. The version adopted by the Top500 project allows system users to optimize software and scale the problem to run efficiently on their machines, thus leveling the playing field somewhat.

Yet, as Dongarra points out, "at the end of the day, it's not the speed of the computer, it's the ability of scientists to get out of that machine what they're trying to accomplish [that matters]. So the machines have to make it easy for scientists to gain insight into the problems they're facing." If a system rates high on the Top500 list for solving a dense system of linear equations but is inefficient at solving the main problems for which it was built, the performance numbers produced by the LINPACK benchmark would certainly not be the most important metric for rating the machine's overall worth.

"LINPACK is something that has evolved over the last 32 years, and at one time it was perhaps a good thing, but systems have changed quite a bit in those 32 years," says Dongarra. He concedes that the benchmark can still provide potentially useful information, "but it shouldn't be used alone."

So how should the performance of large-scale systems be assessed? Dongarra stresses balance among a computer's components.

"You're going to build a machine that's going to cost hundreds of millions of dollars, and whoever's building that system has to build in a balance in terms of the memory, the amount of disks it has, and its capability of really solving petascale problems," he says. "And the environment is more than just the hardware. The environment itself relates to the software, to the algorithms, to the applications. There's sort of an ecosystem in place, and one needs to understand how all of those components interact in order to give the full picture."

Could that full picture somehow provide a new metric for performance? Maybe. If the system isn't balanced—if the ecosystem isn't well designed or maintained—the efficiency of the machine will suffer.

"The ecosystem depends on the components working well together. So the best system would have a matched environment where everything fits in a way that they could all relate to each other very easily," Dongarra says. "Unfortunately, we have a situation in most cases where the hardware is out of balance—far exceeding the capabilities of the rest of the system. We struggle with that aspect—we understand that [computer systems] have tremendous capabilities, but we struggle with ways of coming up with that performance."

Efforts to improve efficiency in large-scale systems are nothing new, and scientists are increasingly focused on addressing the imbalances between system components as petascale computing moves closer to reality. Will the next generation of high-performance systems require fresh metrics to provide meaningful performance data? That's an issue up for debate among computer scientists as the petascale era approaches.—MS

---

\*HPCWire, week of June 9, 2006; <http://www.hpcwire.com/hpc/686730.html>.

## Selling the Public on the Value of Large-Scale Simulations

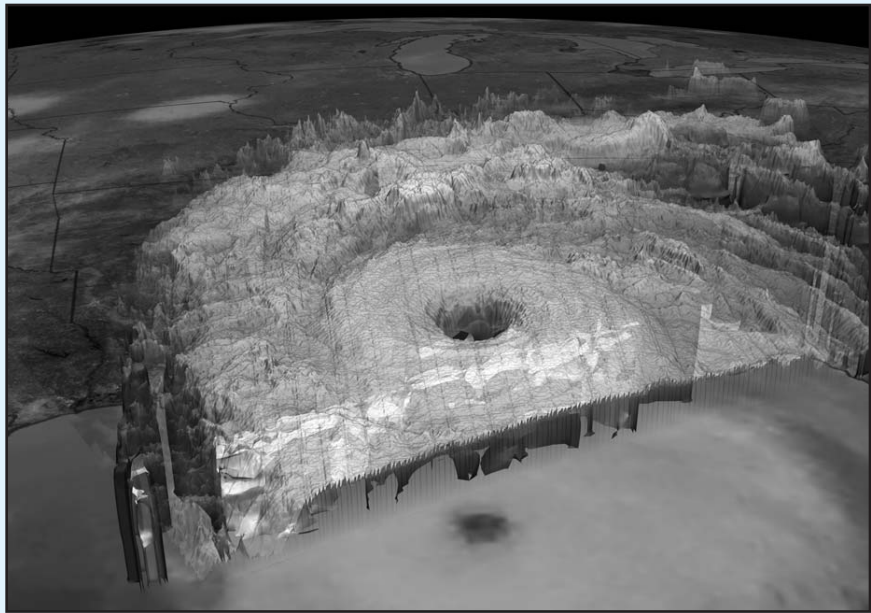
Scientists have long stressed the importance of large-scale simulations for weather prediction. High-end computing systems can be used to predict the likely behavior of a hurricane as it approaches landfall, for example. Such predictions, however, require enormous amounts of computing power, and researchers have scrambled for ways to improve the performance and accuracy of their simulations on today's terascale systems.

Clint Dawson of the University of Texas at Austin believes that petascale computing could be an answer to researchers' prayers. While current systems can produce reasonably accurate hind-casts, he says, petascale computing could allow for the continuous operation of storm-forecasting systems along the Gulf Coast and the eastern seaboard of the U.S. Thus, he says, instead of having to start a new simulation and re-run a minimum of 30 days' worth of data before getting to the present time, a system would always be up to date and able to produce results nearer to real time, in greater detail.

With petascale systems, Dawson says, "we could obtain far more detail within our domains. We could get to the level of putting streets or buildings in. Right now our meshes are on the order of 100 to 1000 meters spacing, but if we could get to tens of meters or below that, we could really model fine details that we just can't capture right now."

Dawson refers to simulations done before Hurricane Katrina struck New Orleans in August of last year. "By Saturday, before the storm hit on Monday, [researchers] had released a storm surge prediction that showed New Orleans was going to get flooded, and it ran in the newspaper on Sunday," he says. "But in a city the size of New Orleans, we can't yet predict which neighborhoods are going to flood precisely because we can't get that detail into the model. I think that if we had petascale, we could do that in the future."

Dawson doesn't want to downplay the value of current simulations. He emphasizes that even at the terascale, simulations



*With petascale simulations should come storm surge predictions more accurate than those produced with today's terascale computers. The image of Hurricane Katrina shown here was captured by precipitation radar, visible infrared scanning, microwave imaging, and geostationary environmental satellites on August 28, 2005. From NASA/Japan Aerospace Exploration Agency (JAXA).*

can prove vital in averting the loss of life that often accompanies major hurricanes. "If there had been good emergency planning" in New Orleans, he says, the storm surge predictions issued on the Saturday before Katrina hit the city might have been put to much better use. Part of the problem, he says, is that the public is unaware of how useful large-scale simulations can be in a real event. Dawson thinks that the improvements introduced with petascale computing could change that.

"We could sell the science to the general public in a much more convincing way," he says. "I think we're getting to the point now where people understand a little bit, because after Katrina, the results of our model were suddenly known to the media, when before they hadn't paid any attention to it." But with even more accuracy and detail in simulations, he says, "we could demonstrate on a broader scale to the public that we have the tools to study these problems."

Getting that point across could have great

societal impact. "I think it could, for one, maybe force people to come to grips with the fact that we can't build everywhere along the coast," says Dawson. "I think it would open people's eyes."

Increased public support for climate research could help speed improvements in the field as well, he says, with those improvements eventually benefiting the public in the form of better long-term forecasting and more accurate weather prediction.

"If we could get to the point where we could couple together weather prediction, the effect of the weather prediction on the general atmosphere, maybe even the temperature, how that affects the ocean, and then how that affects actual storm surge, that I think is a grand challenge that we will eventually tackle somewhere down the road," says Dawson. "Sometime in the future, we'll be able to handle that problem."

With petascale computing approaching, that future may not be far off.—MS