Supercomputing Moves to Universities and Makes Possible New Ways to Organize Computational Research

PAGE 30, 33

Too cheap to meter? Surely not. But microprocessors have come a long way in delivering inexpensive computing power.

The development is eroding the distinction between computers made from low-cost microprocessors, and more conventional supercomputers operated—for example—at the U.S. National Science Foundation’s Supercomputer Centers. It is also shifting the balance in high-performance computing (HPC) away from the NSF’s Supercomputer Centers, to powerful, local computer facilities run by academic departments and universities.

These local facilities tend to have user communities with well-defined computing needs, often centered around a few key applications, such as climate modeling or “ab initio” material simulations. In other words, their computing needs are topical. Topical supercomputing is unusually effective. It holds the promise of increasing our modeling capacity substantially more and more “grand challenge” calculations are performed locally.

Since genuine supercomputing resources—defined to lie within 10% of the NSF’s top supercomputers—can now be purchased for well under $1 million, new and innovative forms of organizing scientific computing should be explored. For example, dedicated support for software development and hardware acquisition should be tailored specifically around individual scientific communities. Likewise, faculty who rely on HPC resources for their research and teaching should be supported by local campus facilities. The result is “democratization” of numerical modeling should serve large and small institutions alike, and it should go a long way toward achieving a true integration of education and research in the geosciences.

Microprocessors on the Rise

The speed of microprocessors has improved dramatically over the past 10 years. Measured in million floating point operations per second (MFlops), processing power increased by more than two orders of magnitude, from about 1 MFlop on Sun Microsystems’ venerable SPARC 1 in the early 1990s, to about 500 MFlops on Intel or AMD processors today. Measured in terms of price-performance—that is, processing speed per dollar—the development is even more breathtaking. On that account, today’s microprocessors exceed their earlier cousins by a factor of 10,000. In other words, since the early 1990s, the cost of 1 MFlops plummeted from about $10,000, to less than $1 today.

There is deeper significance to this leap forward. It stems from our ability to harness hundreds of microprocessors at once through a parallel programming technique known as “active message passing.” Developed in the late 1980s, the method relies essentially on e-mail to communicate among a group of individual programs. Comparing the technique to e-mail is, of course, an analogy. But it helps us to better understand how groups of programs can tackle demanding computational tasks in a coordinated way across computer networks. The key is that the messages carry information from one program to the other, while providing at the same time an effective way to synchronize the entire group; hence the term “active” message passing. Message passing programs require a considerable development effort. But what makes it worthwhile is that they perform exceedingly well across networks of computers, and parallel efficiencies of better than 90% are not uncommon.

The development bodes well for Earth scientists, who have long yearned for affordable parallel computers to satisfy their insatiable appetite for computing power. Hunger for massive computation is, of course, a consequence of the natural systems we tend to model. These systems are often characterized by vastly different time and length scales, which demands the simultaneous resolution of physical processes over a wide range of temporal and spatial scales. Most of us are familiar, for example, with the oceanographers’ need to resolve narrow western boundary currents, such as the Gulf Stream, in global models of the general circulation. Nowadays, hydrologists, space physicists, mineral physicists, geodynamicists, and seismologists are joining their lot with sophisticated groundwater models, simulations of solar wind–magnetosphere interaction, first principles calculations of matter under exotic conditions, fluid dynamic simulations of mantle and core convection, or numerical studies of seismic wave propagation, all of which consume vast computing resources.

Responding to growing computing needs seems easy enough: network a group of inexpensive PCs into a so-called “cluster” and run message-passing programs across the system. There is a term for this innovative approach to parallel computation that has struck a cord with the HPC community: “Beowulf” clustering, named after the original Beowulf cluster project initiated by Jim Fisher at NASA’s Goddard Space Flight Center some 10 years ago. A growing number of Beowulfs is now in use. At Princeton’s...
Supercomputers is based on the same off-the-press model. More important, the comparison in terms of raw CPU hours becomes even more striking. What is new, however, is that this highly effective approach to supercomputing can now be exploited by far smaller scientific communities, as well as by academic departments and universities seeking to enable ambitious computer modeling efforts at their institution.

Organizing Next-generation Community Modeling Infrastructures

Computational geodynamicists comprise a small scientific community that has drawn on NCAR and NOAA's computer facilities are topical in terms of raw CPU hours. The potential impact of this work is that top-tier of supercomputing will be more affordable to small research groups.

Topical Computers: One Size Does Not Fit All

Beowulf work best when they are targeted at a few key applications; that is, when its usage is topical. The topical approach to supercomputing is effective. It is flexible enough to tailor a machine's processors, network, memory, or disc space specifically to the application at hand. But it also works well because it is relatively unburdened by bureaucratic overhead. Seen in this light, clusters are an attractive supercomputing platform, and one that provides computer resources of surprising magnitude. Simply put, a Beowulf is a computing block grant in line with NSF's largest supercomputer allocations. Taking a 256-processor Beowulf as an example, 256 x 2 x 365 = 2,242,560 CPU hours are available each year. By comparison, NSF's highest class of supercomputer awards must be renewed each year. NSF's supercomputer allocations are available for about 90,000-1,000,000 CPU hours per year, or about 5-50% of this amount. While NSF's supercomputer awards must be renewed each year, Beowulf resources are available over the lifetime of the machine, so the simple comparison in terms of raw CPU hours becomes even more striking. More important, the comparison is relevant because today the majority of commercial supercomputers is based on the same off-the-shelf microprocessors that form the core of Beowulfs. Thus, the actual computer power per CPU hour is, in fact, quite similar across all machines.

Topical supercomputing has a long and successful tradition in the Earth sciences. NCAR's and NOAA's computer facilities are topical in nature, with technical staff and hardware dedicated to support atmospheric and oceanographic researchers across the country. What is new, however, is that this highly effective approach to supercomputing can now be exploited by far smaller scientific communities, as well as by academic departments and universities seeking to enable ambitious computer modeling efforts at their institution.

At the same time, ambitious new data acquisition initiatives, such as USAarray, will gather information on Earth's interior in far greater detail than ever before. These new initiatives will necessitate modeling efforts, directed for example at efficient data assimilation methods, on a massive scale. Such efforts will be difficult to meet at a reasonable cost within the traditional infrastructure of the NSF's Supercomputer Centers. Arguably, the cost-effective approach is to support a number of comprehensive modeling frameworks aimed at efficiently utilizing low-cost topical Beowulf clusters.

Japan's Earth Simulator

One computer in the world is about 10 years ahead of Beowulf machines. This remarkable computer, the Earth Simulator (ES) in Japan (www.es.jamstec.go.jp), has 640 8-processor compute nodes, for a total of 5120 processors. Each node has 16 GB of shared memory, for a total of 10 TB of memory. The peak performance per node is 64 GFLOPS, and the total peak performance is 40 TFLOPS, about a factor of 100 larger than what can be accomplished today on a large Beowulf. Access to the ES is still highly restricted, and computer resources are allocated based on strict parallelization and vectorization benchmarks. Recently we were able to test software on the ES.

Jeroen Tromp, in collaboration with Seiji Tsuobi and Dimitri Komatitsch, for example, has been running a global seismic wave propagation code on the ES. On a 2-year-old Beowulf with 75 dual processor nodes and 75 GB of memory, one can simulate global wave propagation at periods greater than 18 s, and a typical simulation lasts 48 h. On 48 nodes of the ES, one can model periods of 8.75 s, and a simulation lasts 10 h. On 432 ES nodes, one can perform simulations at periods of 3.75 s, and a simulation is expected to last about 24 h.

To put these numbers in perspective, typical convolutional summation codes that compute semi-analytical synthetic seismograms for one-dimensional Earth models are accurate up to 6 s. In other words, the ES allows us at this point to simulate global seismic wave propagation in fully three-dimensional Earth models at periods shorter than current seismological practice for one-dimensional spherically symmetric models.

In our view, this experience suggests a dual strategy for NSF supercomputing over the next few years. On the one hand, modeling capacity is probably best served by powerful local supercomputers based on cost-efficient Beowulf clusters that can be upgraded frequently to keep pace with the rapid advance of modern microprocessors. On the other hand, NSF centers should focus on enabling modeling capabilities that can only be met on ES-class machines. The necessary advances in geophysical modeling software, including seismic wave propagation codes and geodynamic forward and inverse models, are already being developed to assist this effort.

---Peter Bunge, Princeton University, N.J. and Jeroen Tromp, California Institute of Technology, Pasadena