LABORATORY FOR COMPUTER SCIENCE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The Future of Scientific Computing for Grand Challenge Problems

Computation Structures Group Memo 383
June 11, 1996

Arvind

Andrew Shaw

John Marshall

Submitted to *Science*

This report describes research done at the Laboratory for Computer Science and the Department of Earth, Atmospheric, and Planetary Sciences of the Massachusetts Institute of Technology. Funding for the Laboratory for Computer Science is provided in part by the Advanced Research Projects Agency of the Department of Defense under the Office of Naval Research contract N00014-92-J-1310.

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

The Future of Scientific Computing for Grand Challenge Problems

Arvind, Andrew Shaw and John Marshall*
Massachusetts Institute of Technology

In the past twenty years, high performance computing has fundamentally changed the practice of science in many fields, including physics, chemistry, biology, atmospheric and oceanic sciences. During this era, the extremely high cost of supercomputers and the wide performance gap between supercomputers and mass-produced computers motivated a centralized, shared resource approach exemplified by government-funded supercomputer centers.

Today, increases in the performance of mass-produced computers are leading towards a more community-based, distributed approach to scientific computing. Despite this trend, there will always exist important scientific problems which require computational power beyond that of even the most powerful mass-produced computers. For reasons we will explain, the computers to solve these "Grand Challenge" problems will be rarer and more expensive than current supercomputers. We call these computers "Jumbocomputers", and make some policy recommendations to maximize their benefit to the scientific community.

Commodity Supercomputing

Within the past few years, many American supercomputer manufacturers have either folded or been taken over by more mainstream computer companies.[1] At the same time, mainstream computer manufacturers such as DEC, HP, IBM, Silicon Graphics, and Sun are building and selling Symmetric Multiprocessors (SMPs) whose performance overlaps the lower-end of supercomputer performance. These SMPs, more commonly called *servers*, have become very popular with the corporate world which uses them to run large databases and high-volume World Wide Web servers. In addition to their high performance, servers are much cheaper, easier to maintain, more reliable, and also offer a better software environment than traditional supercomputers for application development and everyday use.

These new machines have seriously changed the landscape of high-performance computing. Almost all of the 500 individual most powerful computers in the world today are used for scientific computing.[2] By next year, the bottom half of the top 500 computers will be taken over completely by servers which will be used primarily for non-scientific computing by the commercial world. More significantly, these servers will be built in the tens of thousands, rendering a list of the 500 individual most powerful computers almost as meaningless as a list of the 500 individual most powerful PCs.

Community Supercomputing

There is already some evidence that this shift in high performance computing is having an impact on the government-funded supercomputer centers. In each of the past four years, the number of active users of NSF supercomputer centers has declined; furthermore, the majority of users are from the centers themselves or other institutions in the same state.[3]

What is motivating this localization of usage? Users at the centers find it easy to interact, share information, and thus form a natural community in real life. This community lowers the barriers for new users, making them more likely to use computing as part of their research, increasing the size of the local community.

With the advent of servers, other scientists can now have supercomputer-level performance within their own local communities, at their own institutions. Greater control over their computing environment gives them more opportunities to treat servers as "personal", dedicated computers, mitigating some of the gap in performance between local computing resources and those at the centers. Users are also willing to trade off some degree of performance for comfort – for instance, being able to deal with local system administrators, having faster, more responsive, more secure network connections, working within a familiar, local software setup, and having colleagues with whom they can discuss their problems and experiences with the system.

Jumbocomputing Economics

While servers are creating a movement towards community supercomputing, there are still important scientific problems which require performance beyond that of single servers. Unfortunately, building and selling computers which satisfies those needs is a money-losing proposition, as evidenced by the recent bankruptcies and takeovers in the supercomputing industry. There are three interacting reasons for this: first, computer performance increases roughly 50% per year, second, up-front development

*Arvind is Johnson Professor of Computer Science and Engineering; Andrew Shaw is a PhD candidate in the Laboratory for Computer Science; John Marshall is Professor of Physical Oceanography. E-mail may be directed to arvind@lcs.mit.edu.

[1] Thinking Machines and Kendall Square Research went bankrupt. Intel has transformed its supercomputing division into a server group. Cray was taken over by Silicon Graphics; portions of Cray were then sold to Sun Microsystems; Convex was taken over by Hewlett-Packard. NEC, Fujitsu and Hitachi, the Japanese companies, are still in the business, but by all accounts are not making any money in supercomputers.

[2] *TOP500 Supercomputer Sites*, Jack Dongarra, Hans W. Meuer, and Erich Strohmaier. http://www.netlib.org/benchmark/top500/

[3] *National Supercomputers Usage Database*, Quantum Research Corporation, http://www.qrc.com/nsf/cise/asc/usagedb/start.htm

costs are very high in the computer industry, and third, the market for these machines is, by definition, measured in tens of computers rather than thousands per year.

These facts combine to make the supercomputing market a big gamble for companies. They must spend an enormous amount of money to develop computers on a very tight schedule. A six-month delay increases development costs significantly, and produces a product which is 25% slower than the competition (which may be other supercomputers or servers). A one-year delay is usually catastrophic for the product, and in the worst case, fatal for the company. Even if all deadlines are met, development costs can only be amortized over a few sales, raising prices on the product, which must compete with relatively inexpensive servers at the bottom end. A supercomputing company can easily go broke if a few sales don't go through, and in fact, many have for these reasons.

Jumbocomputers, while necessary for the solution of important problems, are not good business propositions. Scientists working on these problems will be increasingly hard-pressed to find pre-packaged hardware/software solutions. *In the near future, we believe that all Jumbocomputers will be custom-made, like other big-ticket science items, such as space telescopes and particle accelerators.*

A $46 million Supercomputer at a PC Price?

Under the DoE's Accelerated Strategic Computing Initiative (ASCI) program, Sandia Laboratory recently bought a 9,000 processor computer from Intel for $46 million.[4] By most measures, this will be the most powerful computer in the world, and the first to break the Teraflop barrier. The cost-per-processor comes to about $5,000, which is about the same price as a PC based on the same P6 microprocessor. Even without accounting for software and other development costs, it is difficult to imagine how Intel will make a direct profit on this one-of-a-kind machine. On the other hand, publicity may generate additional sales of P6-based PCs, generating indirect profits for Intel. In the end, these circumstances garnered the DoE a Jumbocomputer at a bargain price, but such circumstances are unlikely to arise again.

Although ASCI has a large budget for computing ($100 million per year), it it is still difficult for companies to make profits catering to customers like ASCI for the reasons explained above. Furthermore, ASCI's budget is dwarfed by the computing budgets of typical large commercial enterprises such as Citibank, which by some estimates spends about $1 billion per year on computing. *It is the money of the corporate world which is driving the leading edge in computing.*

Some Assembly Required

To leverage off the massive investments of the corporate world, scientific users should consider assembling Jumbocomputers as collections of servers connected with high-performance networks. The networks can either be based on standard fast local area network (LAN) technology, or non-standard network technologies being tested in the marketplace today which have been derived from massively parallel supercomputers. Because LANs aren't meant to be used in this way, the former option would require completely re-written networking software for good performance, while the latter option would result in faster, but more experimental machines, because hardware standards are not yet set in place.

Although operating systems, compilers and application software for *individual* servers are mature, the software for task and resource management and for programming an *ensemble* of servers is at best in an embryonic stage of development. For scientific computing, such a software layer is absolutely necessary for ease of application development, and for the end-user to view the ensemble as a single machine. This software is unlikely to be developed for the corporate world until there is a compelling commercial application which requires it.

Jumbos, not White Elephants

The hardware and software characteristics described above will make Jumbos both expensive and unique, and there is a real danger of a Jumbo becoming a white elephant unless a larger community of users exists to iron out hardware and software problems. A larger community also provides more opportunity for algorithmic innovations necessary for good utilization of such an unusual computer.

One solution is to build and distribute mini-Jumbos which have identical hardware characteristics and software environment to the main Jumbo. For example, rather than a single 9,000 processor machine, it makes more sense to have an 8,000-processor machine and ten 100-processor machines distributed to other institutions. The cost in dollars may be comparable to the cost of the single 9,000-processor machine, and the payback in terms of usability, reliability and even performance (from better algorithms and software) of the main machine will almost certainly more than offset the lower peak performance.

Mini-Jumbos can also serve a role as testbeds for hardware and software development prior to the delivery of the main Jumbo. The huge costs and short lifespans of Jumbos are strong motivation to make them immediately usable on arrival and based on the most recent hardware. Mini-Jumbos allow development at much lower cost, and give enough breathing space so that the next generation of hardware technology can be used in the construction of the main Jumbo.

In summary, we consider mini-Jumbocomputers infrastructure necessary to the success of any Jumbocomputer – funding for Jumbocomputers should reflect costs for this infrastructure.

[4] ASCI's mission is to verify the safety, reliability and performance of the US nuclear stockpile using computer simulation as a replacement for underground nuclear testing. *ASCI Executive Summary*. http://www-c8.lanl.gov/asci/summary.html