Supercomputers, Workstations and Visualization

Dr. Larry Smarr, Director, National Center for Supercomputing Applications

Superconductivity is certainly exciting, and one of its most promising applications is in supercomputers. These ultra-powerful machines will also have an impact on superconductors: People such as Arthur Freeman at Northwestern use them to calculate superconductor electron structure.

I'd like to take you through some of the key technologies driving the revolution in scientific computing, as opposed to the data-processing computing world still dominating 90 to 95 percent of this country's computing investment. We are talking about an emerging new market.

The three items I want to focus on are: distributed computing (you use a series of interconnected computers, connected via networking from the personal computer, for instance, to the supercomputer); replacement of numbers by imaging as the primary computer output; and the critical need for a new revolution in software to interpret between the computer and the human. Computers already compute too fast for us; they approach a billion floating-point operations per second.

I now want to define a few terms; I will use the Cray computer we received in 1985 to do that. First, to do advanced scientific computing, one needs a scientist. It is surprising how many managers leave scientists out of the system architecture. One such scientist is John Kogut, who exemplifies the phase transition that is occurring. He is one of the leading theorists in how quarks and gluons, the fundamental constituents of matter that hold themselves together inside protons and neutrons. He theorizes how the universe came into being when it was all one big quark-gluon soup. His group has used over 12,000 Cray supercomputer hours in the last year (there are 8,000 hours in a year). Remember that a Cray hour, a single processor hour, is about 200 times a VAX-780 hour.

Using a supercomputer as a sledge hammer isn't seen often in American industry, and it is mind-boggling to consider what will happen when industry decides to do that. It is amazing that you have university professors doing this rather than $10-billion-a-year corporations. You can see the potential.

Kogut uses a Macintosh. This is important, because the real revolution is not in supercomputers. We have always had supercomputers; they are simply the fastest, largest memory machine available at any given time. Some $35 wristwatches have the computational power of the supercomputer of 30 years ago. So, a definition of a supercomputer is a moving target.

Five years ago, when IBM announced the personal computer, following up Apple, we began to move from dumb terminals connected to mainframes, and now we work with some 20 to 30 million personal computers. If the PC is only 1/1,000 of the Cray in speed, 25,000,000 divided by 1,000 is 25,000 Cray equivalents sitting on desk tops in this country compared to only 200 actual supercomputers out there. The revolution has come and gone. Most of the computing power is now on desks and must be hooked up, and that will totally change the world into a distributed environment.

I keep mentioning Cray because Cray Research, Inc., is the leading supercomputer manufacturer in the world. The Cray X-MP/24 in the Kogut slide contains two proces-



John Kogut uses a Macintosh in conjunction with Cray supercomputer.

sors. Cray Research, Inc. supports 1 to 8 processor supercomputers. The top IBM mainframe now has six processors. The move to multiple processors will continue; ETA from CDC will utilize a four- and an eight-processor machine.

The object at the left in the slide is the Cray Solid State Storage Device (SSD). It is a large pile of RAM chips that costs several million dollars. Our current SSD holds 1 billion bytes of fast memory, compared to the 1 million bytes on a high-end PC. Memory is very cheap per chip, but it takes many chips to handle the power of a supercomputer.

The processor box on our Cray X-MP/48 supercomputer, one year newer, is the same size as the one on the Cray X-MP/24 but holds four processors and has twice the central memory. VLSI has shrunk the size of the chips, so more memory and speed goes into the computer. Supercomputers of the future will shrink, or you get more in the same size box, and that process is going to accelerate dramatically with the Cray 2 and the Cray 3.

A supercomputer always seems to cost $20 million in constant dollars -- one of these laws of the marketplace -- but you get much more for your buck. At our center alone, the upgrade policy, using this parallelism, creates a power increase of something like a 60-percent annual compounded growth rate. That is; in '85, the two-processor came in; in '86, the four-processor; in '88, it will be followed by an eight-processor; in '89, probably a 16-processor. The Cray 4, which Seymour Cray will produce by the early '90s, will have 64 processors made of gallium arsenide.

Anybody in the business of buying supercomputers has to manage this. You have a $20-million investment. Double it, at least, to get the total investment in terms of people and the peripherals that can handle a fire hydrant like this, and it goes obsolete every two years. That experience is totally outside of current management concepts. Yet if this is really the machine tool of the 1990s, economic competitiveness depends on the ability to utilize it. We have to learn how to manage a fundamental engine of development whose power grows at 60 percent compounded annually.

Managing memory growth is a worse problem, because the enormous installed base of personal computers is driving memory costs to zero; it decreases exponentially. So, the amount of memory that you can get with these machines increases at a staggering rate. Most places still think of the Cray 1 as a supercomputer. Remember that the Cray 1 had 1 to 2 million words of memory. Now the smallest memory that can be purchased on Cray 2 is 128 million words. By the early 1990s, you will get 2 to 4 gigawords of memory -- multiply by 8 to get gigabytes -- on these machines. That is a qualitative change; it is not just a quantitative change. Exponentials have that property.

We live in a multi-vendor environment, which will not change for a long time in scientific and engineering computing. We have to build an "ecology" around the supercomputer to take advantage of the enormous speed and memory. For instance, the Alliant mini-supercomputer, an eight-processor, can go about a third the speed of a Cray X-MP single processor. That has become the center, in our case, of a visualization facility that hooks many workstations together. We create images and link them to a full video-production facility. Vendors represented in such a subsystem might include: Apple, Sun, IBM, Silicon Graphics, DEC, Raster Technology, Abacus, Alliant, Cray. The challenge today is to be able to manage a multi-vendor environment, all parts of which are constantly changing.

One of the most useful developing software tools is the

user interface. The runaway success in the market (it even converted IBM) is the legacy of the Xerox PARC experiment in the early '70s, in which Xerox tried to figure how humans would interact with computers when the machines would be cheap enough to be put on every desk. Nobody believed that would happen in 1970, only five years before the personal computer was invented. Sun and Apple are direct descendants from that program, and now IBM has said that it will come out with the Windows system from MicroSoft that emulates this windowed environment. Basically, you replace the operating systems and the machines by icons.

We would like this software, which is now being developed, to manage the distributed system of computers which each user accesses. You want, for instance, to open up a file on a VAX front end. In our proposed user interface, you double-click on the VAX and list your files. If you want to transfer a file from the VAX to the Cray, which has a different way of listing and moving files, you hit the mouse on "file 2" on the VAX list until it turns black, pull it over to the Cray icon, and when the Cray turns black, the file has been transferred.

If we do not do something like this with the user interface, human productivity will be totally wiped out, because we will spend all our time learning operating systems. This has become a real crisis.

Let's look at how we can now do science and engineering with computers. Fracture analysis is an example. Robert Haber and colleagues in the Department of Civil Engineering at the University of Illinois wanted to test a block of aluminum containing a hidden flaw, by stressing the outside impulsively. A wave of stress will run through until it hits that crack, which absorbs energy and propagates until the material fails. They wanted to see how that happens.

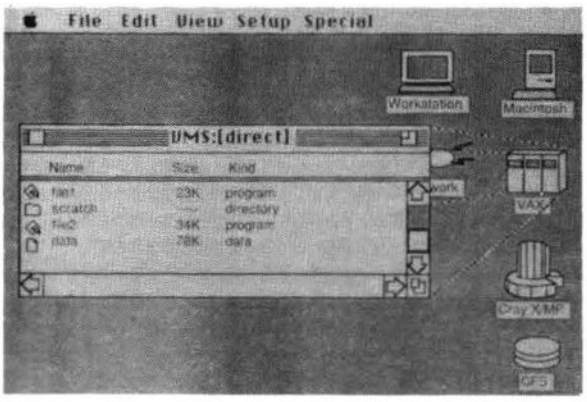
They laid down a grid on the material in the computer. This is called a finite element approximation. The supercomputer simulates the solid-body motion in the aluminum block and calculates stress and wave velocity at every point in space and time. These numbers are then carefully stored.

Most scientists today can access only a tiny fraction of these stored numbers. They say, "There are just too many data to look at, so we will look at one point, like the end of the crack tip versus time." Line graphs like this are typically the kind of data that have been published.

Far better to visualize those numbers. Treat them with modern computer graphics: Plot a two dimensional surface whose height above the plane is one variable (the specific velocity in this case). You can color that surface to represent the local stress intensity. In a sequence of time "snapshots", you see the wave coming in, hitting the crack, and then sending out ripples of waves through the material. You have edge effects which cause reflected waves from the end. Those run over the outgoing ripples, and you see the whole story unfold before your eyes.

At NCSA, you can watch the action on a video tape or directly at a workstation. You are inside the aluminum watching the wave phenomenon. The switch from line graphics to full visualization is only possible because of the advanced visualization facilities that Haber had access to at NCSA.

Another example demonstrates the usefulness of collaboration between scientists and artists. In a study of supersonic jet flow, NCSA Research Scientist Michael Norman used about 250,000 numbers to simulate the supersonic flow of gas at one moment in time. To understand the results of flow field we must provide a visual index to the numbers. As you go across the little boxes that were the 250,000 grid zones in this calculation, you use a color map to replace each number by a color. With that, you



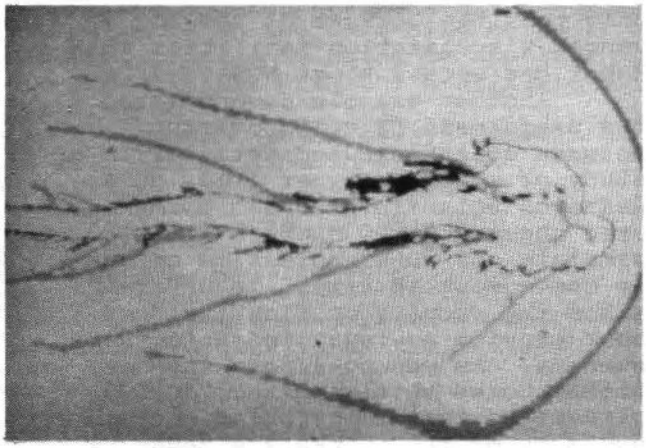
Transferring VAX files to the Cray.

have a visual pattern, you have painted by numbers.

After bringing together a computer artist, Donna Cox, who is an assistant professor of Art and Design, and a scientist, Michael Norman, they were able to devise a very complicated color map that brings out all the details that were computed. You can sit at a workstation and tune this map to pull out features in the flow that would not otherwise be visible. That kind of tuning ("image processing") is important.

Visualization should become a part of the whole code-development environment because it can locate errors. In some work of Michael Norman, Karl-Heinz Winkler, and myself visualization was most helpful. It showed a shock wave bouncing off the bottom corner of the grid; the code was supposed to simulate an open environment. That error would be difficult to find in your code. It is not as though you had dropped a comma; it is an algorithmic error that only shows up in strong shock waves, but does not show up in weaker reflections. Here visualization is needed from the beginning, as you develop the code.

The ability to see, to go from a byte (a single number), up to a megabyte (which is an image) -- a factor of a million -- by using images rather than numbers, is critical to the opening of a whole new methodology that we call computational science. Through visualization as a primary means of interfacing with the computer, one can calculate cases of realistic complexity and perform theoretical experiments on the supercomputer.



Shock wave bounces off grid.

Key areas of our economy are already beginning to be impacted by computational science. One is the biotechnology of proteins. Proteins are basically everything in the body, outside of nucleic acids; proteins are long polymers of amino acids that wind themselves up into complex three-dimensional shapes.

The ability to take one of these protein molecules and do quantum mechanical calculations was impossible before this year. One can now model quantum tunneling and look at the electrons in that molecule that jump essentially instantaneously from one side of the molecule to the other. This required computing at NCSA the Feynman-path integral with 500,000 path integrals to study the statistics of whether this transition was possible. The work was published in Science by Peter Wolynes and colleagues; this was the first quantum tunneling event that had ever been calculated for a biomolecule, and yet presumably these quantum effects underlie the role of macromolecules in the life process. Until supercomputer power was coupled with visualization, we could not approach this important problem.

Another example deals with the protein that allows photosynthesis to occur, which is the fundamental life process on this planet. This work was done by Karplus at Harvard, Shulten in Munich, who came to Urbana to do this calculation, and their colleagues. Five thousand atoms in the molecule were allowed to oscillate. The various functional subpieces of the molecule were then cut apart by the visualization team. We can see in the movie how the molecule is working. This sort of CAD/CAM of molecules will be increasingly important; the design of new drugs, the ability of drugs to dock with viruses and other such biotechnology applications are underway.

Another key area of the economy depends on understanding material properties. Quasi-crystals are a new form of material discovered recently. Wolynes and his colleagues used the NCSA supercomputer to build up such a material, atom by atom. We have visualized the computation of magnesium-aluminum quasi-crystals.

In gallium arsenide, the new material for semiconductors, Karl Hess and his co-workers are com-

puting electron transport by following 50,000 electrons. As chips get smaller and smaller, the ability to simulate the quantum effects along with the classical effects inside these submicron structures is going to be necessary, or we are not going to be able to make them.

Such work is not abstract and theoretical; Business Week's September 1987 issue reported gallium arsenide and quasi-crystals as two of six recent developments to watch. Computational science will be an important partner with laboratory work in such fields.

I want to end with a few points about the national infrastructure required to take advantage of this explosion in computer power. Robert Leach spoke earlier about 1970 as a key year for industry. At that time, federal support for university activities was cut, particularly for large-scale computing, and for the next 15 years that pattern held. During those 15 years, students going into industry had no experience with supercomputers. This is the major reason corporate America is making so little use of them.

The National Science Foundation has worked to move supercomputing resources back into the university community. Through NSF's national competition, five supercomputer centers have been established in San Diego, Champaign-Urbana, Pittsburgh, Princeton, and Ithaca. A modest beginning of a national network to link these centers with each other and with universities in the country has begun.

Our center, the National Center for Supercomputing Applications, is a case in point. We opened our doors about 18 months ago. Some 2,000 scientists and 135 universities now access our system, all of whom had to write proposals receiving peer-review at the national level. We are growing at something like twice the rate of the early days of Alliant Computer, which is one the fastest start-up companies in the country.

There are now about 6,000 users among the five centers who have access to the state-of-the-art supercomputers. Our center offers a comprehensive multi-vendor computing environment, that is tightly networked. We spend considerable time on creating strategic alliances with the vendors, making five-year agreements with each and deciding how we will upgrade to remain at the state-of-the-art. With Cray, Alliant, IBM, Apple, and Sun, we are far along in that process.

At NCSA, we have created an interdisciplinary supercomputer research center. Such a national research center exists at Fermilab for particle physics, which is a tiny fraction of all science and engineering and has very little impact on the national economy. Yet for advanced computing, which has national impact, researchers have had no such place to come together, no national watering hole. Establishing such a center must be a critical national priority. As complicated as it is to hook up these computers, there are far fewer computers than there are applications of them. So there is an enormous economy of scale to be gained by placing enough resources in one center that can figure out how to network computers, how to use them efficiently, how to develop visualization software, and how to foster training and inter-action in the national community.

Where we should go by the early '90s? We should achieve the sort of economy of scale that the national nuclear weapons labs have found: That is, once staff and buildings have been paid for, it is very cheap to just keep adding the best new computer into that environment. The machine should be networked, so that it operates in a real-world environment with the best users of that technology, the end customer, providing feedback. By creating such a national facility, we can test out new technologies such as massively paralled architectures as they are introduced. We need to put them into an environment where we can shorten the time to users such as corporations.

NCSA has also started a corporate partners program. Eastman Kodak has been there a year, Amoco just announced, and we will have several more announcements this year. [Since this speech, Motorola and Eli Lilly & Company have joined.] The corporation makes a commitment for 3 to 5 years to interact with the center, to bring their own researchers up to speed on our computing environment. More than 40 Kodak researchers came to NCSA in the first year to work on

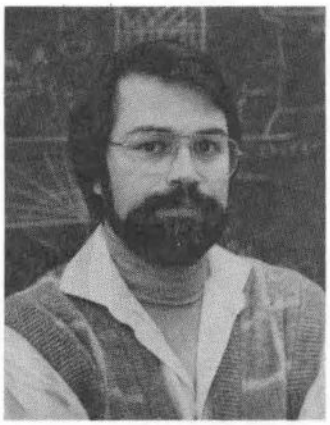
proprietary projects. Once they have their codes up, participants can continue their research remotely over a high-speed link to our center.

Our goal is to establish 10 partners whose projects span the economy and bring them into this advanced computational environment where they work with academic researchers. The industrial partners have early looks at new technologies, and they learn the new methodology of computational science. The idea is to use the national resource we've become and leverage that expertise into the national economy.

NSF had made a five-year commitment to national computer centers. The state of Illinois and the University of Illinois have made a major commitment as well. The goal of the industrial program is to bring industrial America in as an equal player with federal and state governments and universities.

The real challenge now is to maintain the centers at the state-of-the-art, to keep from suffering cash-flow starvation the way most entrepreneurs do. We are, in this year, going to make that decision in this country. We will see what the response is to this challenge is from the various players, the federal government, the state, universities, and industry.

The chance will not come again. I am very hopeful that our country will pull together and build the national infrastructure in advanced computing that will provide a foundation for economic growth for decades. Stay tuned.



Larry L. Smarr

Dr. Smarr is professor of physics and of astronomy at the University of Illinois at Urbana-Champaign and Director of the National Center for Supercomputing Applications (NCSA). He is an internationally recognized astrophysicist and a UI faculty member since 1979. His 1982 paper on "The Supercomputer Famine at American Universities" is regarded as a landmark in the movement to bring supercomputing power to U.S. universities.

DISCUSSION

DR. YOUNG: Larry, thank you very much. I think we have time for one or two quick questions.
DR. HILL: Chris Hill with the Congressional Research Service. You described a very exciting opportunity for the future, but it sounds like you have described something where there ought to be a lot of private money to be made in providing the service. Is it not time to transition this sort of opportunity out of the university-federally suported environment into the private sector if it is as pregnant with possibility as you suggested?
DR. SMARR: I think it is time to start the discussions that see how you might do that. I am afraid that the success at getting so many of our researchers hooked up, and so forth, has blinded some of the people, particularly in the federal government, from the fact that the job is not done yet, that this whole thing could die within the next couple of years because of a lack of just a little bit more increasing funding necessary to keep these things at the state-of-the-art. In other words, you have started the centers, but you have got to manage that 60 percent annually compounded growth rate. That problem has not yet been tackled, the NSF has not yet made a commitment to any of its centers to provide for upgrades. I will be talking with Erich Bloch, the director of the NSF, today, on just that topic. But they have not made the decision. It is not that much money that would allow them to

do that. If they can do that, so that the next five years are more or less set, then the platform is in place to do exactly what you talk about.

DR. DIMAN: What are some of the organizational implications of this technology in the science and engineering world now that this tool is available and real? Where does it lead organizationally?

DR. SMARR: What you are seeing here is a need for a more holistic or a synthetic vision of management to handle this. It is clear that you have to manage this very wide range of technologies that are changing on just a very scary time scale, going obsolete every two years. That is not the way businesses are run. So, there is that need to it. But the network that is going to be hooking up all of our researchers in the country will create a whole new scientific entity, one in which even electronic mail processes becomes important. For instance, I have daily conversations with my colleagues all over the country whom I normally used to see at meetings like this once a year. Now, how does that affect productivity? For instance, in the gallium arsenide case, as a result of the success that this one group in Illinois had, they have now got a proposal in to the NSF to set up a national center for computational electronics. They have gotten IBM, Texas Instruments, all of the players in the corporate world, signed up that do large-scale simulation of electron devices, most of all the major universities. They are all going to electronically hook into our center, they are going to put their models up there, they are going to compare each other, they are going to standardize on the algorithms and on the input parameters. That is being done right now. There are such teams in France, in Italy, now in Britain, and certainly in Japan. Until this happened, there were not any in this country. When those players start working together, what is going to happen and how is that going to be managed? How is that going to be a challenge to the "not invented here," and if we do not do it in-house, we do not do it? I have had 50 corporations, 200-odd industrial people come through our center for presentations in the last year, and I can tell you that these companies each take over a year to come to grips with whether they want to be part of this program or not, even though they are unanimous in saying early on they want to. There is an organizational barrier to making that commitment -- as we heard Leach talk this morning about the difficulties of the big companies dealing with the small companies. Those barriers have got to be broken if this country is going to have any future.

DR. YOUNG: I am afraid time has run out. All I can say is I wish I were 30 years younger and could start all over again. It is fascinating.