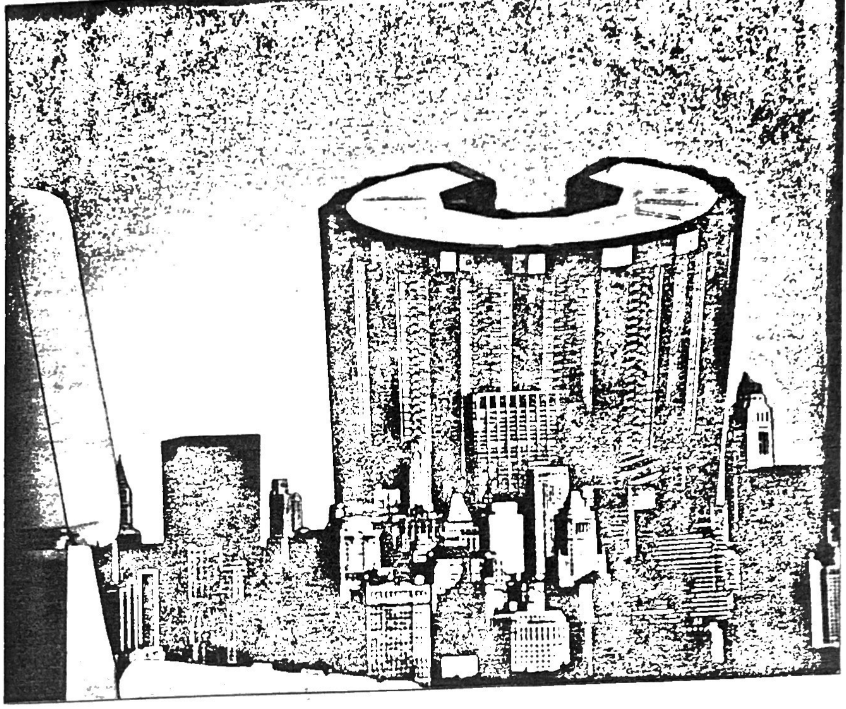
from Joss

Government must back development

Supercomputer design is best done by small engineering teams uncluttered by bureaucracy


USTRATION BY CARL WESLEY

BY WILLIAM NORRIS
Chairman Emeritus, CDC

Defining a supercomputer is complicated, and a definition today is less precise than it was even a couple of years ago. In technical publications and in the general news media, we read such terms as "near," "entry" and "affordable" supercomputers and variants such as "Crayettes" and "mini-supers." Those terms generally apply to large-scale scientific machines designed primarily for scientists and engineers who deal with complex mathematically oriented problems. There are possibly as many as 40 such products being developed and even delivered today.

My definition is restricted to the breed commonly known as very large-scale general-purpose scientific computers and, more specifically, "today's most powerful general-purpose computer." That definition implies that there can be only one supercomputer at any one time, but we all know that any computer's power varies for different applications. So it may well be that today perhaps only two or three computers deserve to be called supercomputers.

In their day, the Eniac, the EDVAC, the ERA 1103, the CDC 6600, the CDC 7600, the Cray 1 and the CDC Cyber 205 could be legitimately called supercomputers. I believe the term first came into widespread use in the 1960s with the introduction of Control Data Corp.'s CDC 6600.

Each of these computers was designed and developed by small teams of highly capable technologists, usually in an environment of high risk. But none of them would have succeeded without the underpinning of advanced technology that resulted from years of federal government research and development funding nurtured in an environment of cooperation among the federal government, universities and industry.

In the '40s and early '50s, direct federal government funding of supercomputer development leaned heavily on the vast reservoir of electronic technology developed under government auspices during World War II.

In the late '50s and 60's, the pattern changed to government-supported base technology development. The government placed orders in advance of development by small companies, which risked their future on technical success with supercomputers.

Since the 1970s, the federal government's approach has been to buy its supercomputers at arm's length from established suppliers who take all the risk for their development. It is no coincidence that in the past decade, the rate of progress has slowed.

'Try before buy'

This "try before buy" policy, which is now common in U.S. government procurements, superficially seems equitable in protecting government interest, but it places too much of the risk for developing new supercomputer technology on the manufacturers, without any assurance of an adequate market.

In addition, since the market for supercomputers is relatively small, the major manufacturers are increasingly unable to justify the necessary investments to achieve the best performance in the time frame needed. This funding risk can be greatly alleviated if the U.S government would commit to purchasing quantities of supercomputers in advance of their development.

Relevant history of government funding shows that the Aberdeen Proving Grounds funded John Brainard, J. Presper Eckert and John Mauchly to develop Eniac at the University of Pennsylvania. They started their own company in 1944 to design and market Univac, the first commercial electronic computer. Early on, the government contracted to buy three machines, but the onus was on Eckert and Mauchly to achieve specifications before payment was made — a significant change to the prior procurement pattern. In 1950, they sold out to Remington Rand, which subsequently merged with Sperry in 1955.

From then on, while Eckert and Mauchly continued to introduce improvements, their creative genius atrophied in the environment of a large company.

After World War II, I helped form Engineering Research Associates (ERA) in St. Paul, Minn., in order to retain a technical resource of vital national importance to the U.S. Navy. We had been developing special-purpose electronic systems for cryptologic applications. It was suggested that instead of forming our own company, we should attach ourselves to a public institution, such as a university or research foundation. However, we felt the need to better control our destiny by running our own company to avoid the lethargic and often unsupportive environment of a large institution.

ERA was formed in 1946, and in August 1947, it received a cost-plus-fixed-fee development contract for the design of a general-purpose stored computer, called Atlas, which was shipped in December 1950. I am convinced that if ERA had been part of a large organization, that development, if it could have been completed at all, would have taken several more years.

After ERA was acquired by Remington-Rand, the Atlas development spawned the ERA 1101 and 1103 computers, the first of the Univac 1100 series. In those days, government customers accepted responsibility for software development. They knew their problems far better than we, they had the resources and they made a tremendous contribution to the software art during that period and for many years after.

The National Security Agency (NSA) was a vital catalyst to supercomputer development in those days. A few of many computer firsts at NSA, all funded under cost-based contracts in the '40s and '50s, were the following:

- Demon, the first practical use of magnetic drums.
- Atlas I, the first parallel electronic computer with drum memory.
- Atlas II, delivered by ERA in October 1953, with vastly enhanced I/O capabilities compared with Atlas I.
- Lightning, high-speed circuitry research aimed at a 1,000-megacycle computer.
- Solo, the first completely transistorized computer.

It is clear that the '50s and '60s were a most fertile time for the advancement of supercomputers. The environment was characterized by enlightened self-interest and financial support from knowledgeable government agencies working with small entrepreneurial teams of computer engineers focused on the creation of

a single product. The work was underpinned by a vast reservoir of base technologies derived largely from government funding in the national labs, universities and major company basic and applied research organizations.

In the late '50s and '60s, assistance from the national laboratories was very important. The orders from Livermore for the first 6600 and CDC 7600 computers were of enormous help in leavening the risk for CDC.

The 6600 and 7600 developments followed the same pattern of success with a small development team. Seymour Cray's development group never exceeded 30 people, and of course, at the time of the 6600 development, CDC was still a small company.

In addition, the availability of risk capital for small companies **since World War II — a unique feature of the U.S. economy — was vital to spawning the entrepreneurial enterprises that have done most to accelerate the state of the art in supercomputers. Eckert and Mauchly, ERA, CDC and Cray Research are salient examples.**

Limited market

Although good growth is occurring, the market for supercomputers is still only capable of supporting a few competitors worldwide on its own merits. The market can be entered with the expectation of a reasonable return, but most companies opt for larger markets with lower risk and the potential for greater profitability.

Of course, a company can enter the market because of national or corporate prestige with the hope that the beneficial image it carries **will provide better-than-average profits in other product and services lines of the company. History also shows that significant technological fallout occurs from supercomputer development to the benefit of less powerful computer products.**

The supercomputer designer's never-ending struggle to balance computation and I/O capability has also been a constant spur to the evolution of peripheral devices.

Because of the high cost of supercomputers, the only reasonable way to provide access to them for the engineers, scientists and researchers who need to use them is by way of compatible, easy-to-use and reliable networks. This need, again, has been a constant spur to the evolution of network architectures.

Supercomputer developments **have been a catalyst for enhancements of the technology in both system and applications software.** Compilers such as Fortran and Pascal are being enhanced to exploit supercomputer hardware architectural features, such as vector processors, multiprocessors and parallel processors.

Preprocessor software is also appearing in the industry to help users gain greater benefits from their existing codes when they are run on supercomputers. New algorithms are being developed, and old algorithms are being researched and modified to optimize supercomputer applications performance.

In the area of reliability, today's supercomputer developments are advancing the state of the art in such areas as circuit maintenance, fault-tolerant logic, error detection and correction and remote technical assistance.

Nevertheless, the prestige and technology fallout motivations for a company to participate in the supercomputer market are propositions too tenuous on which to bet the national survival and international competitiveness of the U.S.

Direct support

Direct government support is necessary. It should take two forms. The first is funding for national labs and universities to buy and use supercomputers. This funding increases the size of the market so that more competitors can stay in the game. Second, it helps enormously in developing base technology within the national labs and universities and that technology can, in part, be the catalyst for new advances in supercomputers.

The funding available for national laboratories and universities to buy supercomputers declined seriously between 1970 and 1983. For example, during that time, CDC delivered 85 Cyber 7600 and Cyber 176 computer systems, not one of which was procured for a university in this country.

In contrast, we delivered seven 7600 systems to universities abroad. Two Cyber 205 systems were installed in U.S. universities by year-end 1983, while four were installed in European universities. And we wondered why the rest of the world was gaining on us so rapidly in basic research and advanced technology!

There were only three supercomputers installed in U.S. universities by the end of 1983 — at Colorado State, University of Minnesota and Purdue University. Each of them was underutilized, and other university researchers were unable to take advantage of their research benefits because of lack of funds.

Fortunately, the U.S. government has finally recognized its need to help assure the continued leadership of the U.S. in supercomputing. The National Science Foundation decision, prompted by the Lax Report to resume its policy of providing grants for supercomputing support at U.S. universities, is beginning to prove very beneficial.

The second needed form of support is for the government to assume more of the funding risk for supercomputer development and the advanced technology on which it relies, particularly in the area of component research. Industrial co-

operative enterprises, such as the Microelectronics and Computer Technology Corp. and Semiconductor Research Corp., can bear some of that advanced technology burden but not enough of it.

Another lesson from history is that product development is best done by small engineering teams working in an environment uncluttered by bureaucracy. In the U.S., at least, the small company is undoubtedly the most conducive environment for such development. It is entrepreneurial. It is dedicated. Its personnel have fortunes to gain from success and bankruptcy to face from failure. There is no better motivation for hard and creative work. The small company lacks deep pockets and is therefore forced to focus on the most direct route to success.

Government agencies and national laboratories that place advanced orders can look forward to a far closer working relationship with a small company than is possible with a large corporation. That relationship can be the stimulus for more beneficial cooperation between the parties if for no other reason than that the CEO is inevitably more accessible — and if the small company gets out of line, the government's kick will be felt far more in a small company than in a big one.

The government is also less vulnerable to the accusation of providing unfair competitive advantage to the small company than would be the case if the same contract were awarded to an established competitor.

Direct government R&D funding of commercial supercomputers beyond basic advanced technology should be concentrated in the area of applications. Learning to fully use the power of new architectures is a painfully slow process. But it could be much faster if there were more directly sponsored government work in universities and laboratories as a means of better applying these machines to important classes of problems.

By the end of this century, the architectural improvements over the next one or two decades may be implementable in optics rather than electronics.

The trend in supercomputers is undoubtedly toward more parallelism. The massively parallel supercomputers of the future will provide enough simulation speed and accuracy to provide a profound analytical resource to help us deal effectively with the complexities of our world. No longer will it be necessary, in most cases, to rely on approximate solutions based on costly empirical experiments or analog analyses.

Examples abound in education, in three-dimensional imaging, medicine, chemistry, genetics, fluid dynamics, destructive testing and weather modeling and forecasting, as well as a wide range of applications in the economic and social sciences. Significant results are being achieved today, but there are far more opportunities waiting to be explored.

The trend toward greater parallelism will permit truly scalable computing to be realized for the first time, with the resultant evolution of more powerful minicomputers that are truly compatible with their supercomputer brethren.

By the end of this century, the architectural improvements over the next one or two decades may be implementable in optics rather than electronics. Optical computers capable of operating hundreds of times faster than their electronic counterparts may well be realized. When this happens, the devices on which these optical computers are based will open a wide spectrum of novel approaches to computer architectural design.

The heart of computer networks

Inexpensive communications, based on fiber-optic technology, will be at the heart of highly reliable new computer networks that are completely transparent to computers of different origins.

Application software, created with the assistance of artificially intelligent computers, will be more precise, more adaptable, more reliable and dramatically less expensive to develop than ever before.

And the interaction between the human and the computer will become profoundly simpler and more integrated as future machines learn, increasingly, to become articulate in human languages. But above all, progress in artificial intelligence will be the key to the full integration of the computer with its human partners.

Already, by concentrating on knowledge representation — the machine counterpart to human memory — researchers are delivering practical results, especially in the field of expert systems. Although machine intelligence, as Alan Turing defined it, has not yet been achieved, limited but commercially viable expert systems are beginning to emerge.

Relatively little has been achieved so far in machine vision — not because there has been little research, but because the obstacles are immense. While assembly-line robots now have some ability to discern shapes, the problem of eliciting keen discrimination from machines is proving very difficult. Machine reasoning and natural language processing remain largely the subject of the laboratory, not commerce. But because natural language can be used in applications where vocabulary needs are limited, natural front-ends are beginning to find their way into a few application areas.

In my opinion, AI will eventually become so architecturally embedded in systems and products that it will cease to exist as a separately identifiable entity, just as the microprocessor is invisible in many domestic appliances today. It is probable that expert systems that truly rival the capabilities of the human expert in a wide variety of fields will be available in the mid-1990s.

Expert systems will become prevalent in application areas where knowledge bottlenecks are present, where job performance is inconsistent, where a process must be performed more rapidly than is currently possible, where adverse working conditions and tedious or repetitive tasks make human involvement unpleasant, where rapid change is being experienced and where knowledge-intensive tasks are key.

Take the matter of education in our public schools. By any measure, the public schools in the U.S. are failing badly at their most important tasks. A good deal of attention has been paid to scores on the Scholastic Aptitude Tests in recent years and on the failure of many students to deal with even the simplest writing and reading tests.

As the schools face this crisis with increasing enrollment, declining numbers of teachers and the persistent problems of motivating a group of students with widely diverse learning backgrounds, they need help.

Fortunately, the development of computer-based education systems has already reached the point where computer-assisted instruction can provide high-quality educational experiences for all youngsters.

Similarly, the instruction management and student testing components of these systems, which include embryonic expert systems characteristics in some cases, have reached a level of sophistication well beyond anything imagined just a few years ago.

As a result, teachers can be far more effective than before because they have the capability to analyze, diagnose and prescribe for each individual student's particular learning needs, assisted by the computer-based education system.

In summary, supercomputers will continue to be critically important in helping to maintain the nation's well-being and catalyze major advances in computer technology. Progress in artifical intelligence will be the key to the full integration of the computer with its human partners, and expert systems in a wide variety of fields will be available in the mid-'90s.

There are exciting days ahead in the design and application of computers.

HANDS-ON

Surgical robot performs biopsies

Ole has performed 18 biopsies of brain tumors, and he's less than 3 years old. Eventually, when he has gained a little more sophistication and learned to interact better with strangers, his inventor expects even greater things of the surgical robot.

Inventor and electrical engineer Yik San Kwoh directs CAT scan research at Memorial Medical Center in Long Beach, Calif. "I don't really know the end potential, but I think we've just barely scratched the surface," Kwoh says. "There are any number of situations that demand the kind of precision and stability that a robot like this offers." The hospital is planning to expand Ole's sphere of responsibility to include assists in brain stimulations, Kwoh says. A few other possibilities include eye surgery, spinal surgery and knee joint replacements.

Kwoh wrote Ole's software, which works in conjunction with a CAT scanner, during a period of three years in consultation with eight other researchers. Ole was named after a benefactor of the hospital, Sven Olsen, who underwrote the robot's purchase and much of the development expense.

The robot wields its 29-lb., six-jointed aluminum hand with accuracy that even a brain surgeon could envy. In fact, Kwoh says, the robot is so exact that it has reduced the necessary size of a biopsy incision from one-half or five-eighths of an inch down to one-eighth of an inch.

Other advantages the robot brings to this kind of work include memory precision, mobility and relative immunity. Ole is able to return to the same spot with an accuracy within two thousandths of an inch, can be moved easily from one location to another and does not suffer ill effects from X-ray exposure.

That last characteristic could be particularly valuable in reducing time spent in operations where surgeons could benefit from having continuous CAT scan readings. According to Kwoh, Ole might someday be used to perform actual surgical procedures while the human surgeons direct its hand from another room, outside the X-ray's reach.

COMMENT

Americans do great things when they can stand on a foundation of technology as they respond to perceptive leadership. Our response to the Soviet challenge in space and President John F. Kennedy's stated goal of landing men on the moon and returning safely to Earth stood principally on a growing understanding of how to process and use information rapidly and accurately.

Although primitive in comparison with the computer of today, the *Apollo* flight, launch and mission control computers made possible this first great adventure by humankind in space.

The next great adventure in space, the human settlement of Mars, will depend even more heavily on computer technology. The largest remaining question is whether it will be Soviets or Americans who lead in the application of computers to this pivotal project of the third millennium.

HARRISON SCHMITT
Apollo astronaut and former U.S. Senator