8 February 1985, Volume 227, Number 4687 SCIENCE

Perspectives on Supercomputing

B. L. Buzbee and D. H. Sharp

Summary. This article provides a brief look at the current status of supercomputers and supercomputing in the United States. It addresses a variety of applications of supercomputers and the characteristics of a large modern supercomputing facility, the radical changes in the design of supercomputers that are impending, and the conditions that are necessary for a conducive climate for the further development and application of supercomputers.

For many years scientific computing has been a major factor in maintaining U.S. leadership in several areas of science and technology and in the application of science to defense needs. Electronic computers were developed, in fact, during and after World War II to meet the need for numerical simulation in the design and control of military technology. Today, scientists and engineers are using supercomputers to tackle ever more complex and demanding problems (*1*), and the availability of supercomputers 10^4 times faster than the first electronic devices is having a profound impact on all branches of science and engineering—from astrophysics to elementary particle physics, from fusion energy research to automobile design. The reason is that supercomputers extend enormously the range of problems that are effectively solvable. However, several factors have combined to place the discipline of scientific computing in a critical position. These factors involve both scientific and societal issues.

Although the past 40 years have seen a dramatic increase in computer performance, the number of users and the difficulty and range of applications have been increasing at an even higher rate, to the point that demands for greater performance now far outstrip the improvements in hardware. For example, to achieve an increase in resolution by an order of magnitude in a three-dimensional, time-dependent problem would require an increase of speed by a factor of 10^4—far in excess of the increases projected to be available in the next few years. The result is that difficult problems are typically undercomputed relative to the scientific requirements. Of comparable importance is the fact that the broadened community of users is demanding improvements in software that will permit a more comfortable interface between user and machine.

Ways to meet these needs are on the horizon. But to exploit these opportunities will require scientific progress on several fronts. Computers as we have known them for the past 20 to 30 years are about as fast as we can make them. The demand for greatly increased speed can be met only by a radical change in computer architecture—from a single serial processor, whose logical design goes back to Turing and von Neumann, to a computer which is an aggregation of many parallel processors that can perform independent operations concurrently. Substantial changes and improvements in programming languages and other software will also be necessary to make use of large-scale parallelism in scientific computation.

While massively parallel machines hold great promise for the future, the importance of measures to more effectively utilize available hardware cannot be overemphasized. These measures include the development of better numerical algorithms and better software for their implementation. Improved software could also greatly reduce the time needed to translate difficult problems into machine-computable form. Kemeny (*2*) has observed that since the introduction of computers, improvements in software have increased the productivity of people by at least two orders of magnitude. If another order of magnitude can be realized within the next decade, then practice in many fields of science and technology will be revolutionized.

Supercomputing is unusual in the degree and rapidity with which progress in the field will affect vital national needs in the areas of security and economics. For example, the skill and effectiveness with which supercomputers can be used to design new and more economical civilian aircraft will help to determine whether there is employment in Seattle or in a foreign city. Computer-aided design of automobiles is playing an important role in Detroit's effort to recapture its position in the automobile market. The speed and accuracy with which information can be processed will bear on the effectiveness of our national intelligence services. It is thus a matter of concern that traditional U.S. leadership in supercomputer technology is being challenged by robust foreign competition. What strategies and policies are available to the United States for maintaining its leadership in supercomputing?

In the next section we briefly examine some of the ways that supercomputers are used in applications. We then describe the requirements for supercomputing that can arise from a large and diverse group of users, and how the computing network of the Los Alamos National Laboratory copes with them. The following section discusses recent trends in computer performance and architecture, and a final section outlines some conditions that are necessary for a conducive climate for the development and application of supercomputers.

B. L. Buzbee is the deputy leader of the Computing and Communications Division and D. H. Sharp is a fellow at the Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

Uses of Supercomputers

The term "supercomputer" refers to the most powerful scientific computer available at a given time (*3*). The power of a computer is measured by its speed, storage capacity (memory), and precision. Today's commercially available supercomputers, the Cray-1 from Cray Research and the CYBER 205 from Control Data Corporation, have a peak speed of over 100 million operations per second, a memory of about 4 million 64-bit "words," and a precision of at least 64 bits.

There are at present about 75 of these supercomputers in use throughout the world. They are being used at national laboratories to solve complex scientific problems in weapon design, energy research, meteorology, oceanography, and geophysics. In industry they are being used for design and simulation of very large scale integrated circuits, aircraft design, and oil and mineral exploration. To demonstrate why we need even greater speed, we will examine some of these uses in more detail.

The first step in scientific research and engineering design is to model the phenomena being studied. In most cases the phenomena are complex and are modeled by equations that are nonlinear and singular and, hence, refractory to analysis. Nevertheless, one must somehow discern what the model predicts, test whether the predictions are correct, and then (invariably) improve the model. Each of these steps is difficult, time-consuming, and expensive. Supercomputers play an important role in each step, helping to make practical what would otherwise be impractical. The demand for improved performance of supercomputers stems from our constant desire to bring new problems over the threshold of practicality.

As an example of the need for supercomputers in modeling complex phenomena, consider magnetic fusion research. Magnetic fusion requires heating and compressing a magnetically confined plasma to the extremes of temperature and density at which thermonuclear fusion will occur. During this process unstable motions of the plasma may occur that make it impossible to attain the required final conditions. In addition, state variables in the plasma may change by many orders of magnitude (*4*). Analysis of this process is possible only by means of large-scale numerical computation, and scientists working on the problem report the need for computers 100 times faster and with larger memories than those currently available.

Table 1. Computers used for scientific computation.

Description	Quantity	Operating system
Cray XMP	2	Interactive
Cray-1	4	Interactive
CDC-7600	4	Interactive
CDC Cyber 176	1	Interactive
CDC Cyber 176	3	Interactive
DEC VAX 11/780	35	Interactive

Another example concerns computing the equation of state of a classical one-component plasma. A one-component plasma is an idealized system of one ionic species immersed in a uniform sea of electrons such that the whole system is electrically neutral. If the equation of state of this "simple" material could be computed, it would provide a framework within which to study the equations of state of more complicated substances. Before the advent of the Cray-1, the equation of state of a one-component plasma could not be computed throughout the parameter range of interest.

Using the Cray-1, scientists developed a better but more complex approximation to the interparticle potential (*5*). Improved software and very efficient algorithms made it possible for the Cray-1 to execute the new calculation at about 90 million operations per second. Even so, the calculation took some 7 hours to run; but, for the first time, the equation of state for a one-component plasma was calculated throughout the interesting range.

There are numerous ways in which supercomputers play a crucial role in testing models. Models of the circulation of the oceans or the atmosphere or the motion of tectonic plates cannot be tested in the laboratory but can be simulated on a large computer. Many phenomena are difficult to investigate experimentally in a way that will not modify the behavior being studied. An example is the flow of reactants and products within an internal-combustion engine (*6*). Supercomputers are currently being used to study this problem. Running a modern wind tunnel to test airfoil designs costs $150 million a year. Supercomputers can explore a much larger range of designs than can actually be tested, and expensive test facilities can be reversed for the most promising designs.

Supercomputers are also needed to interpret test results. For example, in underground nuclear weapons tests many of the crucial physical parameters are not accessible to direct observation, and the measured signals are only indirectly related to the underlying physical processes. Henderson (*7*) of the Los Alamos National Laboratory characterized the situation well: "The crucial linkage between these signals, the physical processes, and the device design parameters is available only through simulation."

Finally, supercomputers permit scientists and engineers to circumvent some real-world constraints. In some cases environmental considerations impose constraints. Clearly, the safety of nuclear reactors is not well suited to empirical study. But with powerful computers and reliable models, scientists can simulate reactor accidents, either minor or catastrophic, without endangering the environment.

Time can also impose severe constraints on the scientist. Consider, for example, chemical reactions that take place within microseconds. One of the advantages of large-scale numerical simulation is that the scientist using it can "slow the clock" and, with graphic display, observe the associated phenomena in slow motion. At the other end of the spectrum are processes that take years or decades to complete. Here numerical simulation can provide an accelerated picture that may help, for example, in assessing the long-term effects of increasing the percentage of carbon dioxide in our atmosphere.

Requirements for Supercomputing

When computers first appeared, they were used as stand-alone systems. Productivity of the machine was generally given priority over the productivity of people using it. The recent fusion of communications and computation has combined with a steady improvement in cost performance of computer hardware to reverse this situation. Today, provision of a modern computing facility requires the assembly of a wide variety of hardware, software, and communications equipment. Further, efficient operation of such a facility requires expertise in such areas as hardware engineering, software engineering, communications engineering, preparation and management of massive amounts of information, computer-aided as well as classroom education, and financial and technical management. In this section we review some of the requirements that arise in providing computing capability to a large and diverse group of scientists.

Since the beginnings of the Manhattan Project, large-scale computations have played a central role in the design of nuclear weapons. The Los Alamos Na-

tional Laboratory had its genesis in this project, and has maintained state-of-the-art supercomputing throughout its history. Thus, our discussion of the requirements for supercomputing will draw on the example of the Los Alamos computing facility. We will first discuss the work load and then the computing facilities that support it.

Work load. The Los Alamos work load divides into two classes: production and development. Production takes place on weeknights and weekends and involves the execution of the large-scale simulations required in the study of nuclear weapons, the design of devices for achieving controlled thermonuclear fusion, the analysis of the safety of fission reactors, and other difficult and complex engineering problems. These simulations often require from 1 million to 4 million words of memory (1 word = 8 bytes, 1 million bytes = 1 Mbyte.) They may run from 1 hour to 100 hours on the Cray-1, and a single execution may require as much as 20 million words (160 Mbytes) of local disk storage. Development occurs primarily on weekdays, during which Los Alamos scientists are engaged in interactive code development, interactive setup of simulations for the production shift, analysis of results obtained from previous production runs, and execution of smaller jobs. The associated tasks typically use up to 300,000 words (2.4 Mbytes) of memory and run for as long as 30 minutes on a Cray-1.

Objective. The charter of the Los Alamos Computing Division is to provide the computing resources required for both production and development activities. In deploying these resources, the guiding objective has been, first, to maximize the productivity of people and, secondly, to make efficient use of the hardware. Although the activities supported by the computing facility are diverse, ranging from document preparation to large-scale computations, they turn out to have the following common needs: suitable hardware and software, convenient access, mass storage, a varied menu of output options, support services, and ease of use. We shall briefly review each of these points in turn.

Suitable hardware and software. The nucleus of the computing facility is a collection of computers designed primarily for scientific computation. They are listed in Table 1. The Cray XMP, Cray-1's, and CDC-7600's are used for large-scale simulation. The Cybers support administrative computing as well as computer-aided design and manufacturing. The VAX's are incorporated into a distributed network called XNET.

Table 2. Recent CFS performance data.

Level of storage	Percent of total data contained in this level	Percent of requests satisfied from this level	Average response time from this level
Disk	1	82	10 seconds
3850	17	17	1 minute
Off-line	82	1	5 minutes

They are used for controlling experiments and collecting and analyzing data, while also providing a modern software-rich environment that includes screen editing and virtual memory. An interactive operating system is provided on every computer, since experience shows that interactivity increases the productivity of people.

Convenient access. The scientist's tools include notes, books, blackboards, and computers. Access to a computer is needed from the same location as the other tools; namely, the scientist's office or laboratory. This is provided through an integrated network exploiting interactive operating systems. A functional diagram of the network is shown in Fig. 1. It is divided into three security partitions: secure, administrative, and open. From each of these partitions the scientist has a menu of computers and services. The Los Alamos network includes terminal concentrators (terminal network), computational servers (worker machines), a file server (common file system), an output station (print and graphical express station, PAGES), a network production controller (FOCUS), and a gateway to other networks (XNET). This distribution and concentration of functionality enhances efficiency.

Mass storage. Supercomputers can consume and generate enormous amounts of data in a short time. Thus every large-scale computing center needs a mass storage facility of considerable capacity. Most large-scale computing centers have recognized this need; however, few have mass storage facilities which are both efficient and reliable. The Los Alamos system is shown in Fig. 1 as the common file system. Files may be stored on it and accessed from it interactively from all computers. Currently, it has an on-line capacity of about 3 trillion bits (375,000 Mbytes) and unlimited off-line capacity. Los Alamos scientists have stored over 1 million files containing a total of about 13 trillion bits (1.5 million Mbytes) of information. To date, its availability averages over 99 percent. Its file organization is tree-structured, much like UNIX. One of its novel features is automatic file migration. Storage within the system is hierarchical: disks provide the first level of storage, an IBM 3850 cartridge store provides the second level, and off-line cartridge storage provides the third level. The file migration system monitors the frequency with which a file is accessed and situates it in the hierarchy as a function of usage frequency (*8*). Recent performance data are given in Table 2.

The quick response, reliability, and ease of use of this system have made it effective. Also, its on-line accessibility has greatly reduced the magnetic tape

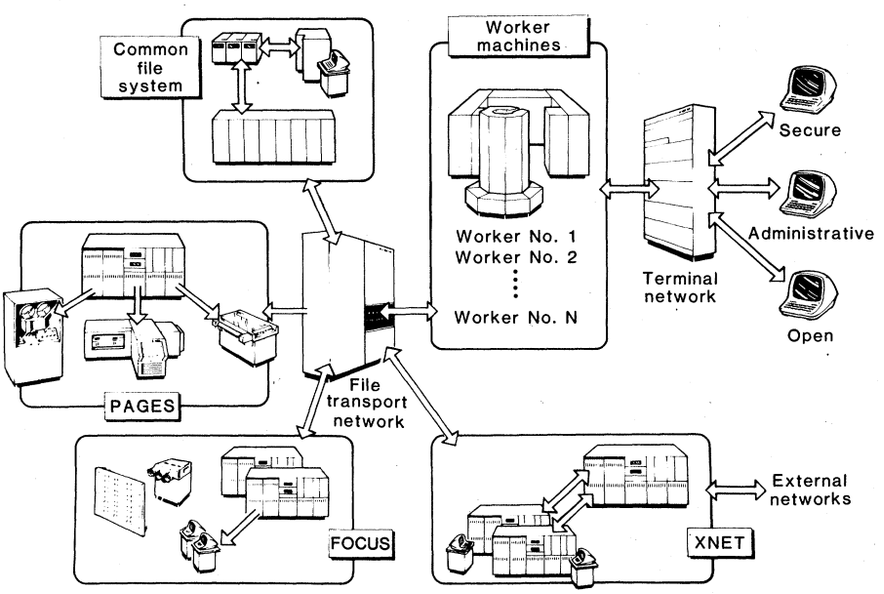

Fig. 1. The Los Alamos computer network.

Table 3. Output options and equipment.

Output option	Equipment
8½- by 11-inch (double-sided)	High-speed printers
11-inch roll (electrostatic)	Plotter
36-inch roll (electrostatic)	Plotter
36-inch vellum (electrostatic)	Plotter
16-mm color film	Film recorders
35-mm color and black-and-white film	Film recorders
105-mm microfiche	Film recorders

inventory and increased the productivity of the operations staff.

Output options. The scientist who uses a computer needs a variety of output options. For example, during code development and problem checkout, the scientist may need to generate graphics on his terminal before getting hard copy. In the course of a parameter space search, a scientist may wish to save copies of the program and numerical results on microfiche for efficient archival storage. When dealing with a time-dependent problem that consumes several hours of supercomputer time, the scientist may wish to produce a movie to show the time history. All of these options are available through the equipment listed in Table 3 and located in a single network node (shown in Fig. 1 as PAGES).

PAGES is accessible on-line from all computers. Relative to its off-line predecessor, PAGES significantly reduced turnaround time for graphics and significantly increased the graphics capacity and quality at Los Alamos.

Support services. Maintenance and modernization of the Los Alamos computer network, its components, and the associated software require people skilled in various disciplines. About half of the Computing Division's staff of 300 is required to provide these "basic services." In addition to basic services, supercomputing requires "support services" such as operations, documentation, education, consulting, accounting, and research.

Ease of use. Ease of use is achieved through specialized supercomputer software, software-rich systems, and common software across the network. The Los Alamos specialized supercomputer software includes symbolic debugging tools and efficient machine-language subroutines for vector operations and optimizing compilers. Software richness is achieved through a variety of computing systems, a menu of programming languages, and a variety of vendor-supplied applications packages. Commonality is provided across all systems through common libraries of graphics and mathematical software and through communication with common utilities such as the common file system and with PAGES.

Some critical issues. The Los Alamos Integrated Computing Network constitutes one of the largest scientific computing facilities in the world. Yet, steady growth in the effectiveness and application domain of computers have necessitated continuing efforts to increase the computational capability and capacity of our network. This has far-reaching consequences, and, in general, the requirement to maintain a responsive, reliable, and secure network adds substantial complexity to our hardware and software projects. We must manage improvements such that most of the functions that work today will work tomorrow. As is the case with many scientific laboratories, we have a large inventory of applications software that constitutes an important part of the laboratory's technology base. This base must be carried into the future. The cost in time and resources to make alterations to this inventory can quickly get out of hand, so preservation of the associated investment is a major consideration. At the same time, as we discuss in the next section, rapid change in computing technology will continue unabated into the foreseeable future. Powerful work stations have the potential to significantly increase the productivity of people through new and rich programming environments. Integrating them into our network and merging their functionality into supercomputer applications is nontrivial. Further, we must soon integrate and utilize supercomputers with radically new architectures.

Trends in Supercomputer Performance and Architecture

Improvement in the performance of supercomputers in the 1950's and the 1960's was rapid. First came the switch from cumbersome and capricious vacuum tubes to small and reliable semiconductor transistors. Then in 1958 a method was invented for fabricating many transistors on a single silicon chip a fraction of an inch on a side, the so-called integrated circuit. In the early 1960's computer switching circuits were made of chips each containing about a dozen transistors. This number increased to several thousand (medium-scale integration) in the early 1970's and to several hundred thousand (very large scale integration, or VLSI) in the early 1980's. Furthermore, since 1960 the cost of transistor circuits has decreased by a factor of about 10,000.

The increased circuit density and decreased cost has had two major impacts on computer power. First, it became possible to build very large, very fast memories at a tolerable cost. Large memories are essential for complex problems and for problems involving a large data base. Second, increased circuit density reduced the time needed for each cycle of logical operations in the computer.

Until recently a major limiting factor on computer cycle time has been the gate, or switch, delays. For vacuum tubes these delays are 10^{-5} second, for single transistors 10^{-7} second, and for integrated circuits 10^{-9} second. With gate delays reduced to a nanosecond, cycle times are now limited by the time required for signals to propagate from one part of the machine to another. The cycle times of today's supercomputers are between 9 and 20 nanoseconds and are roughly proportional to the linear dimensions of the computer, that is, to the length of the longest wire in the machine.

Figure 2 summarizes the history of computer performance, and the data have been extrapolated into the future by approximation with a modified Gompertz curve. The asymptote to the curve, which represents an upper limit on the speed of a single-processor machine, is about 3 billion operations per second. Is this an accurate forecast in view of developments in integrated circuit technology? Estimates (*9, 10*) are that a supercomputer built with Josephson-junction technology would have a speed of at most 1 billion operations per second, which is greater than the speed of the Cray-1 or the CYBER 205 by only a factor of 10.

Thus, supercomputers appear to be close to the performance maximum based on our experience with single-processor machines. However, most scientists engaged in solving complex problems of the kind outlined above feel that an increase in speed of at least two orders of magnitude is required. If we

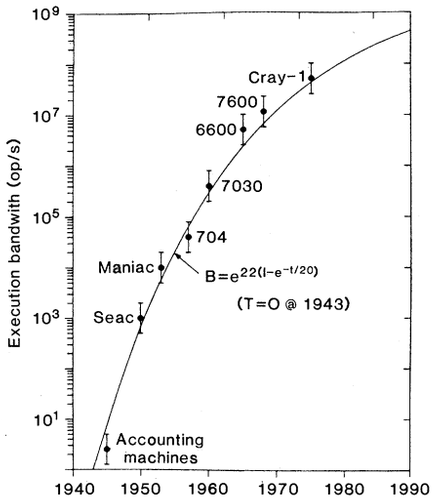

Fig. 2. Computer performance increases since the 1940's.

are to achieve an increase in speed of this size, we must look to machines with multiple processors arranged in parallel architectures, that is, to machines that perform many operations concurrently. Three types of parallel architecture hold promise of providing the needed hundredfold increase in performance: lockstep vector processors, tightly coupled parallel processors, and massively parallel machines.

Vector processors may be the least promising. It has been shown that to achieve maximum performance from a vector processor requires vectorizing at least 90 percent of the operations involved, but a decade of experience with vector processors has revealed that only about 50 percent of the average problem can be vectorized. However, vector processing may be ideal for those special cases that are amenable to high vectorization.

The second type of architecture employs tightly coupled systems of a few high-performance processors. The so-called asynchronous systems that use a few tightly coupled high-speed processors are a natural evolution from high-speed single-processor systems. Indeed, systems with two to four processors are becoming available (for example, the Cray X-MP, the Cray-2, the Denelcor HEP, and the Control Data Cyber 2XX). Systems with 8 to 16 processors are likely to be available by the end of this decade.

What are the prospects for using the parallelism in such systems to achieve high speed in the execution of a single application? Experience with vector processing has shown that plunging forward without a precise understanding of the factors involved can lead to disastrous results. Such understanding will be even more critical for systems now contemplated that may use up to 1000 processors.

A key issue in the parallel processing of a single application is the speedup achieved, especially its dependence on the number of processors used. We define speedup (S) as the factor by which the execution time for the application changes, that is,

$$S = \frac{\text{(execution time for one processor)}}{\text{(execution time for } p \text{ processors)}} \quad (1)$$

To estimate the speedup of a tightly coupled system on a single application, we use a model of parallel computation introduced by Ware (*11*). We define α as the fraction of work in the application that can be processed in parallel, and consider for simplicity a two-state machine, that is, one in which at any instant either all p processors are operating or only one processor is operating. Then, if we normalize the execution time for one processor to unity

$$S(p,\alpha) = \frac{1}{(1-\alpha) + \alpha/p} \quad (2)$$

Note that the first term in the denominator is the execution time devoted to that part of the application that cannot be processed in parallel, and the second term is the time for that part that can be processed in parallel. How does speedup vary with α? In particular, what is this relationship for $\alpha = 1$, the ideal limit of complete parallelization? Differentiating S, we find that

$$\frac{\partial S\,(p,\alpha)}{\partial \alpha}\Big|_{\alpha=1} = p^2 - p \quad (3)$$

Figure 3 shows the Ware model prediction of the speedup as a function of α for a 4-processor, an 8-processor, and a 16-processor system. The quadratic dependence of the derivative on p results in low speedup for α less than 0.9. Consequently, to achieve significant speedup, we must have highly parallel algorithms. It is by no means evident that algorithms in current use on single-processor machines contain the requisite parallelism, and research will be required to find suitable replacements for those that do not. Further, the highly parallel algorithms available must be implemented with care. For example, it is not sufficient to look at just those portions of the application amenable to parallelism because α is determined by the entire application. For α close to 1, changes in those few portions less amenable to parallelism will cause small changes in α, but the quadratic behavior of the derivative will

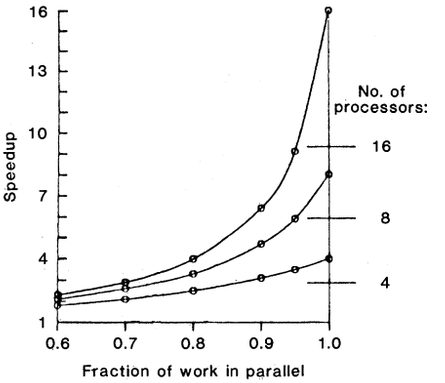

Fig. 3. Speedup as a function of parallelism and number of processors.

translate those small changes in α into large changes in speedup.

Those who have experience with vector processors will note a striking similarity between the Ware curves and plots of vector processor performance versus the fraction of vectorizable computation. This similarity is due to the assumption in the Ware model of a two-state machine, since a vector processor can also be viewed in this manner. In one state it is a relatively slow general-purpose machine, and in the other state it is capable of high performance on vector operations.

Ware's model is inadequate in that it assumes that the instruction stream executed on a parallel system is the same as that executed on a single processor. This seldom is the case because multiple-processor systems usually require execution of instructions dealing with synchronization of the process and communications between processors. Further, parallel algorithms may inherently require additional instructions. To correct for this, one may add a term $\sigma(p)$ to the execution time for parallel implementation that is nonnegative or perhaps monotonically increasing with p. Actually, σ is a function not only of p but of the algorithm, the architecture, and even of α. Let $S(p,\alpha,\sigma)$ denote speedup for this modified model. Then

$$S(p,\alpha,\sigma) = \frac{1}{(1-\alpha) + \alpha/p + \sigma(p)} \quad (4)$$

If the application can be put in completely parallel form, we find

$$S(p,\alpha,\sigma)|_{\alpha=1} = \frac{p}{1 + p\sigma(p)}$$

In other words, the maximum speedup of a real system is less than the number of processors p, and it may be significantly less. Also note that, whether or not α equals 1, S will have a maximum for sufficiently large p because α/p becomes

insignificant while $\sigma(p)$ continues to increase.

Thus the research challenge in parallel processing involves finding algorithms, programming languages, and parallel architectures that, when used as a system, yield a large amount of work processed in parallel (large α) at the cost of a minimum number of additional instructions (small σ).

Recent work (*12*) on tightly coupled parallel processors has concentrated on systems with two to four vector processors sharing a large memory. Such machines have been used successfully for parallel processing of scientific computations. Logically, the next steps are systems with 8, 16, or even 64 processors; however, scientists may not be able to find sufficient concurrent tasks to achieve high parallelization with 64 processors. The problem lies in the granularity of the task—that is, the size of the pieces into which the problem can be broken. To achieve high performance on a given processor, granularity should be large. However, to provide a sufficient number of concurrent tasks to keep a large number of processors busy, granularity will have to decrease, and high performance may be lost.

The situation is even more challenging when we consider a massively parallel system with thousands of processors communicating with thousands of memories (for example, data flow systems, the Ultracomputer project at New York University, and the Cedar project at the University of Illinois). In general, scientists cannot find and manage parallelism for such large numbers of processors. Rather, the software must find it, map it onto the architecture, and manage it. Therein lies a formidable research issue.

In summary, the architecture of supercomputers is likely to undergo fundamental changes within the next few years, and these changes will affect many aspects of large-scale computation. To make this transition successfully, a substantial amount of basic research and development will be needed.

Conclusion

We have taken a brief look at the current status of supercomputers and supercomputing, touching on a variety of applications of supercomputers, on the characteristics of a large modern supercomputing facility, and on the radical changes in the design of supercomputers that are impending. What are the future prospects for supercomputing in the United States, and how does the U.S. effort compare with those in other countries?

Although several countries have undertaken substantial efforts to develop the technology of supercomputing, the most vigorous and advanced projects are under way in Japan. A detailed comparison of the U.S. and Japanese efforts was recently given in (*13*). Some of the salient points are these. The Japanese leadership have thoroughly grasped the importance of supercomputing to their economic and scientific objectives. Under the auspices of the Ministry for International Trade and Industry, they have embarked on two ambitious projects—the National Super-Speed Computer Project, intended to produce high-performance computers for scientific computation, and the Fifth Generation Project, intended to exploit applications of artificial intelligence.

It is interesting to consider the emphasis the Japanese are giving to artificial intelligence in their projects. As the technology develops, efforts to produce superfast machines and efforts to build intelligent machines may draw closer together: this is because the software necessary to use superfast computers effectively may require artificial intelligence capabilities, while to run advanced artificial intelligence algorithms may require the fastest machines.

Japanese supercomputing technology is beginning to be comparable in its overall capabilities to that of the United States. If their projects succeed fully, they will have machines far more advanced than anything now available. We do not know if this will happen, but even if they are only partially successful the results should be impressive. In any event, past experience has shown that it is wise to respect Japan's technological capabilities.

Technological development efforts in Japan make use of several interesting organizational methods: vertical integration of industries, cooperation between different manufacturers, and close collaboration between the industrial and academic sectors and the concerned government ministries. While few would suggest that Japan can provide an organizational model for the United States, this does raise the question of what can be done to improve the climate for the development of supercomputers in this country.

This question was extensively discussed at the recent conference on "Frontiers of Supercomputing" held at Los Alamos and sponsored jointly by the Los Alamos National Laboratory and the National Security Agency (*14*). A broad spectrum of interests and points of view were represented, but there was wide agreement on several points. One of these was that the country would benefit if there were closer cooperation among the industrial, academic, and government sectors.

The computer manufacturers made the point that the current market for supercomputers is thin, making it difficult to raise private capital for major supercomputing R&D projects. It is thought that the potential market for supercomputers is quite large, but to access this market will require better software and more people trained in the use of supercomputers.

Universities have a very important role to play. With financial help they can themselves provide a market for state-of-the-art supercomputers, train people in their use, and develop new applications for supercomputers and university researchers can devise sophisticated software. To carry out these functions, the university community needs ready access to supercomputers. How best to do this raises some difficult questions. We have seen that the operation of a large center for supercomputing requires many people and a substantial amount of specialized equipment. Locating several supercomputers in a single center is a way to reduce costs by economies of scale and thus make these costs more reasonable. However, large centers tend to lose some of the flexibility which is desirable for experimenting with new computer systems.

Thus, we may see a trend toward the development of a few large centers whose purpose is to provide machine cycles to a national user community which accesses the centers by an elaborate communications network, supplemented with a program to locate smaller computers and experimental machines at many individual sites. In considering this suggestion, the technical challenge and cost of developing a network to serve the large centers should not be underestimated. We also remark that in referring to "smaller" computers above, we mean something like a VAX 780 with an array processor. A great deal of computing can be done with such equipment, which in many applications can be cost-effective compared to a supercomputer.

Government can help the industrial and academic sectors by providing tax incentives to computer manufacturers, by assisting universities to access supercomputers, and by increasing the level of

support for basic research on supercomputers and their applications. Also, new antitrust legislation could make more secure new companies such as SRC (Semiconductor Research Corporation) and MCC (Microelectronics and Computer Technology Corp.), which want to pool scarce talent and resources in the semiconductor and computer technology fields. Finally, multiyear authorization bills would greatly aid the planning of research in supercomputing, as in other fields of science.

We believe that measures such as those outlined above will produce a climate conducive to progress in supercomputing. They will permit the United States to marshal its impressive strengths in this area—an entrepreneurial supercomputer industry, a robust and innovative academic research establishment, and a large and growing base of experience in supercomputer applications—to maintain its leadership in supercomputer design, and to make further impressive strides in the application of supercomputing to scientific and engineering problems.

References and Notes

1. *Physics Today* **36** (1983), special issue on "Doing Physics with Computers."
2. C. Morgan, *On Computing* (fall 1981), p. 36.
3. P. D. Lax, chairman, *Report of the Panel on Large Scale Computing in Science and Engineering* (National Science Foundation, Washington, D.C., 26 December 1982).
4. *Magnetic Fusion Energy and Computers: The Role of Computing in Magnetic Fusion Energy Research and Development* (Office of Fusion Energy, Department of Energy, Washington, D.C., October 1979).
5. W. L. Slattery, G. D. Doolan, H. E. DeWitt, *Phys. Rev. A* **21**, 2087 (1980).
6. T. D. Butler *et al.*, *Proc. Energy Combust.* **7**, 293 (1981).
7. D. Henderson, *Computation: The Nexus of Nuclear Weapons Development*, Proceedings of the Conference on Frontiers of Supercomputing (Univ. of California Press, Berkeley, in press).
8. T. McLarty, W. Collins, M. Devaney, in *Proceedings of the Sixth IEEE Symposium on Mass Storage Systems* (IEEE, New York, in press).
9. A. L. Robinson, *Science* **201**, 602 (1978).
10. J. Matisoo, *Sci. Am.* **242**, 50 (May 1980).
11. W. Ware, *IEEE Spectrum* 84 (March 1972).
12. O. Lubeck, paper presented at the 1983 SIAM (Society for Industrial and Applied Mathematics) Conference on Parallel Processing, Norfolk, Va.
13. B. L. Buzbee, R. H. Ewald, W. J. Worlton, *Science* **218**, 1189 (1982).
14. The proceedings of this conference are to be published by the University of California Press. Highlights of the conference are reported in B. L. Buzbee, N. Metropolis, and D. H. Sharp, *Los Alamos Science* **9**, 64 (1983).

15. Supported by the U.S. Department of Energy. We thank R. H. Ewald, J. Glimm, and W. J. Worlton for their helpful comments on this article.

Galileo, Planetary Atmospheres, and Prograde Revolution

G. D. Parker

Summary. Early in March 1610 Galileo was preoccupied with curious brightness variations of the newly discovered satellites of Jupiter. In formulating an incorrect explanation he advanced important generalizations about the existence of planetary atmospheres and counterclockwise circulation within the solar system.

The first serious scientific conjecture that planetary atmospheres exist quite generally was made by Galileo in March 1610 in the concluding paragraphs of his *Starry Messenger*. Analysis of Galileo's data shows that he developed this hypothesis from his misinterpretation of an optical illusion and from an unmentioned assumption about the counterclockwise sense of revolution for motions within the solar system.

The hypotheses of planetary atmospheres and prograde revolution were employed by Galileo to account for certain appearances of the system of four satellites that he had recently discovered around Jupiter. Records of his early telescopic observations of the Galilean satellites consist of their apparent configurations and relative brightnesses. Because these data are of good quality, their detailed analysis elucidates the manner in which Galileo reached his correct but unfounded conclusions.

Gary D. Parker is associate professor of physics at Norwich University, Northfield, Vermont 05663.

Orbital Brightness Variations

A striking feature of the relative brightness information recorded in the *Starry Messenger* is the frequent dimness of the satellite nearest to Jupiter. Galileo recognized this variation of apparent brightness with orbital position and was preoccupied with its explanation in his conclusion of the *Starry Messenger*. Reconstruction of the satellite configurations shows that Galileo regularly underestimated satellite brightness at small separation angles (*1*). In Fig. 1 records from the *Starry Messenger* are reproduced for some of the observations in which the brightness of the innermost satellite is underestimated. This brightness diminution occurs in approximately half of Galileo's brightness data.

Reconstruction of the satellite configurations shows that the probability of Galileo's underestimating the relative brightness of the nearest satellite varies inversely with separation angle and approaches 100 percent for the smallest separations detected (*1*). Figure 2 shows this probability compared with the judgments of eight contemporary scientists viewing a simulated Jovian star field in the laboratory. In the simulation, a viewer in a darkened room directly views three colinear light sources that mimic Galileo's telescopic view of Jupiter and two satellites (*2*). The viewer controls the brightness of the closer "moon." The separation of this moon is varied by the experimenter, while the viewer matches the brightness of the two moons. The electrical power which a viewer delivers to the inner moon is measured as a function of its separation. Two viewers can differ considerably in the power provided at a given separation, and it is impossible to know which viewer's perception would be most similar to Galileo's. Nevertheless, a clear pattern emerges: all viewers decrease the supplied power as separation is increased. For each viewer there is a separation beyond which the inner satellite is reduced no further. The fraction of viewers who underestimate the inner satellite