REPORT

OF THE

PANEL ON LARGE-SCALE COMPUTING

IN SCIENCE AND ENGINEERING

Peter D. Lax, Chairman

Under the sponsorship of

Department of Defense (DOD)
National Science Foundation (NSF)

In cooperation with

Department of Energy (DOE)
National Aeronautics and Space Administration (NASA)

December 26, 1982

This report does not necessarily represent either the policies or the views of the sponsoring or cooperating agencies.

EXECUTIVE SUMMARY

Report of the Panel on Large Scale Computing in Science and Engineering

Large scale computing is a vital component of science, engineering, and modern technology, especially those branches related to defense, energy, and aerospace. In the 1950's and 1960's the U. S. Government placed high priority on large scale computing. The United States became, and continues to be, the world leader in the use, development, and marketing of "supercomputers," the machines that make large scale computing possible. In the 1970's the U. S. Government slackened its support, while other countries increased theirs. Today there is a distinct danger that the U.S. will fail to take full advantage of this leadership position and make the needed investments to secure it for the future.

Two problems stand out:

> Access. Important segments of the research and defense communities lack effective access to supercomputers; and students are neither familiar with their special capabilities nor trained in their use.

Access to supercomputers is inadequate in all disciplines. Agencies supporting some disciplines such as fusion energy, atmospheric sciences, and aerodynamics have funded National computing facilities through which their remote users have limited networking capabilities. In those disciplines that attempt to fund computing through individual research grants, access to large scale computing remains minimal.

> Future Supercomputers. The capacity of today's supercomputers is several orders of magnitude too small for problems of current urgency in science, engineering, and technology. Nevertheless, the development of supercomputers, as now planned in the U.S., will yield only a small fraction of the capability and capacity thought to be technically achievable in this decade.

Significant new research and development effort is necessary to overcome technological barriers to the creation of a generation of supercomputers that tests these technical limits. Computer manufacturers in the U. S. have neither the financial resources nor the commercial motivation in the present market to undertake the requisite exploratory research and development without partnership with government and universities.

Unless these barriers are overcome, the primacy of U. S. science, engineering, and technology could be threatened relative to that of other countries with national efforts in supercomputer access and development. Although the Federal Government is the first and by far the largest customer for supercomputers, there are no national plans to stimulate the development and use of advanced computer technology in the U. S.

Recommendations:

The Panel recommends the establishment of a National Program to stimulate exploratory development and expanded use of advanced computer technology. The Program has four principal components, each having short- and long-term aspects. Underlying them all is the establishment of a system of effective computer networks that joins government, industrial, and university scientists and engineers. The technology for building networks that allow scientists to share facilities and results is already developed and understood; no time should be lost in connecting existing research groups and computing facilities.

The four components of the recommended program are:

1. Increased access for the scientific and engineering research community through high bandwidth networks to adequate and regularly updated supercomputing facilities and experimental computers;

2. Increased research in computational mathematics, software, and algorithms necessary to the effective and efficient use of supercomputer systems;

3. Training of personnel in scientific and engineering computing; and

4. Research and development basic to the design and implementation of new supercomputer systems of substantially increased capability and capacity, beyond that likely to arise from commercial requirements alone.

The Panel recommends that this program be coordinated within the Federal Government by an interagency policy committee, and that an interdisciplinary Large Scale Computing Advisory Panel be established to assist in its planning, implementation, and operation.

The Panel believes that current funding levels are insufficient to maintain the Nation's leadership in large scale computing. Federal agencies that depend on large scale computing to fulfill their missions must work together to reexamine priorities and to create a coherent program responsive to their individual missions. The Panel has set forth policy and planning issues and has outlined some options for implementation.

Members of the
Panel on Large Scale Computing
in Science and Engineering

William F. Ballhaus, Jr.	NASA-Ames Research Center
James C. Browne	University of Texas, Austin
Brice Carnahan	University of Michigan
Michael Creutz	Brookhaven National Laboratory
Richard H. Gallagher	University of Arizona
Herbert Keller	California Institute of Technology
John Killeen	Lawrence Livermore National Laboratory
Peter D. Lax, Chairman	New York University, National Science Board
Walter M. Macintyre	National Center for Atmospheric Research
Steven Orszag	Massachusetts Institute of Technology
Werner Rheinboldt	University of Pittsburgh
Allan Robinson	Harvard University
Jacob Schwartz	New York University
Edward Wilson	University of California, Berkeley
Kenneth Wilson	Cornell University

CONTENTS

I. Introduction

Supercomputers are the fastest and most powerful scientific computing systems[1] available at any given time: they offer speed and capacity, or special characteristics, significantly greater than on the most widely available machines built primarily for commercial use.[2] Large scale scientific computing is the application of supercomputers to the solution of a model or simulation of a scientific or engineering problem through the appropriate use of numerical algorithms and techniques.

The availability of supercomputers during the past thirty years has been crucial to the Nation's advances in science, engineering, national security, and industrial productivity. Supercomputers have been essential to scientific and engineering investigations in areas such as atmospheric research, astrophysics, molecular biology, integrated circuit design, and fusion research. The weapons programs of DOE, cryptographic analysis, and weather forecasting are dependent on the availability of computational facilities. The use of supercomputers in the aerospace, petroleum, semiconductor, and nuclear industries contributes substantially to the nation's productivity. The development of supercomputers has significant spinoffs for all the technologically based components of the national economy. Research and development in semiconductor technology and in computer research has directly supported and expanded the defense, industrial, medical, and consumer segments of the economy.

The U.S. is the acknowledged leader in the development and use of supercomputers. In 1970 this Nation was preeminent in all aspects of electronic, computer, and computational technology. However, America's present leadership in supercomputers is challenged in the areas of components, development, and applications. Recently, Hitachi has begun marketing what is claimed to be the first 16k bipolar ECL RAM[3]; this device, representative of the continuing advances of Japanese microelectronic manufacturers, is designed for applications in today's scientific computers. Fujitsu, Nippon Electric Company, and Hitachi have each developed supercomputers, which are claimed to compare favorably with the available, or announced, American systems. American universities and research centers, which have historically created new applications of

----

1 These include specialized machines, such as array processors, that are equal to or, for some problems, more powerful than general purpose mainframes.

2 The term supercomputer, as used in this report, encompasses hardware, software, supporting peripherals, and the facilities and personnel needed for their appropriate use. Appendix I reproduces three papers presented to the Panel: NCAR Computing Capabilities and Services, by W. Macintyre; Magnetic Fusion Energy and Computers, by J. Killeen; and The Potential of Los Alamos National Laboratory to Provide Large Scale Computational Capabilities to the Research Community, by B. Buzbee and D. Sparks. These papers present descriptions of supercomputer facilities and of their access from remote facilities through networking.

3 See Electronic Engineering Times, December 6, 1982.

supercomputers and sustained research in computational mathematics, have lagged behind their Japanese and European counterparts in the installation of supercomputers.

Significant national thrusts in supercomputing are being pursued by the governments of Japan, West Germany, France, and Great Britain. Some of these, notably the Japanese effort, center on the development of supercomputers; others, on the provision of supercomputers, or access to them through networks, to the scientific and engineering research community. The British program, for example, is designed to provide research scientists and engineers in academic and government laboratories access to supercomputers through modern workstations connected to a high-speed national network. These agressive foreign national initiatives provide a striking contrast to the current state of planning in the U.S.. The domestic computer industry continues its vigorous research and development efforts in the supercomputer field; however, it is felt that these efforts, necessarily dictated by commercial conditions, are less than they could be and far less than should be for the national scientific and technical capability as a whole. The U.S. research community does not have sufficient access to supercomputing facilities, as is documented in numerous studies, papers, and reports directed toward specific disciplines and specific agencies. A partial bibliography of these studies is included in this report.

Expressions of concern that the U.S. is failing to exploit its position of leadership in supercomputing are being voiced from many quarters. Reflecting this concern, the NSF/DOD Coordinating Committee requested, in April of 1982, that a workshop be organized to explore the problems, needs, and opportunities in large scale computing. This Workshop, sponsored by NSF and DOD with the cooperation of DOE and NASA, was led by a panel of fifteen scientists and engineers from a broad spectrum of disciplines. It took place at the NSF on June 21-22, 1982, and was attended by over one hundred participants. Experts in the use, design, and management of large scale computers from the computing, defense, and other industries, government laboratories, universities, and research funding agencies were included. The lists of the participants in this Workshop are contained in the Supplement.

The Panel assessed the role of supercomputing in scientific and engineering research; surveyed the current use, availability, and adequacy of supercomputers; and considered near- and long-term needs. Subsequent to the June 21-22 Workshop, numerous meetings of smaller groups of participants have taken place; in particular, experts on computer development (Group 3 of the Lax Panel) met at Bellaire, Michigan, on August 23-24, 1982, to further explore avenues for assuring the development of future supercomputers. From these meetings a large number of suggestions and position papers have been directed to the Panel and to the Organizing Committee. This report is an attempt, on the part of the Organizing Committee and the Panel, to outline both the results of the Workshop and the subsequent discussions and contributions. The Panel has chosen not to repeat all the detailed technical arguments or examples of the use of supercomputers found in the literature. A bibliography and appendices are included.

Overall, this report outlines the issues and options for the U.S. to maintain its leadership in supercomputers and supercomputing. Because the issues involve many Federal agencies, government laboratories, universities, private sector companies, and scientific disciplines, they need to be addressed on a National basis and require Federal study, planning, and support. The Panel's report attempts to bring the fundamental issues to the attention of policymakers; however, it deliberately avoids details of an organizational, programmatic, or budgetary nature.

II Summary of Findings and Recommendations

Summary of Findings

Large scale computing is a vital component of science, engineering, and technology[4], bringing together theory and applications. It is essential for the design of many technologically sophisticated products, and is making possible for the first time the analysis of very complex scientific and engineering problems which to date have defied analytical and experimental techniques. Examples of the importance of supercomputing are briefly noted below[5].

Renormalization group techniques[6] are a major theoretical breakthrough that provide a new framework for the understanding of a number of unsolved scientific and engineering problems ranging from problems in quantum field theory, the onset of phase transitions in materials, the development of turbulence, propagation of cracks in metals, and the exploitation of oil reservoirs. Only a minute fraction of these problems can be solved analytically. Large scale computational techniques have been essential to the use of renormalization group methods, and even today's largest computational machines are not sufficiently powerful to address most of these problems[7].

Aerodynamic design using a supercomputer has resulted in the design of an airfoil with 40% less drag than that developed by previous experimental techniques[8]. The solution of the Navier-Stokes equations with sufficient

------

4 See, for example, R & D for National Strength, Center for Strategic and International Studies, Georgetown University, Washington, D.C., 1982. See also The Defense Science Board Summer Study, 1981.

5 In the Supplement of this Report, H. B. Keller and J. R. Rice describe in some detail scientific and engineering areas in need of supercomputers. Appendix II contains a number of examples of scientific and engineering problems successfully addressed on supercomputers as well as additional examples requiring supercomputing capabilities not yet available.

6 These techniques were devised to handle the description of phenomena with interactions spanning an extremely wide scale. K. G. Wilson was awarded the 1982 Nobel Prize in Physics for his contributions to the theoretical development and application of renormalization group techniques to critical phenomena. See his contributions to this report in the Supplement, as well as in Appendices II and III.

7 See, in Appendix II, several problems posed by K. G. Wilson.

8 For a detailed description see "Trends and Pacing Items in Computational Aerodynamics", [40].

resolution to represent faithfully fluid behavior became practical with the current (Class VI) generation of supercomputers. The wings of the Boeing 767 and of the European Airbus 310 were designed by computational methods on such machines, resulting in this most significant improvement.

The aerodynamic design of an entire aircraft is not feasible with today's supercomputers; it is estimated that machines 100 times more powerful are needed for this purpose. The design of jet engines, involving the simulation of complex three-dimensional fluid flows and associated chemical reactions, also requires significantly increased computational capability and capacity.

In design, especially of advanced weapons systems, large scale computational modeling is an essential substitute for experimentation. Similarly, the design of future generations of nuclear power plants, and their operation--relying on real-time simulation for their control--require computational facilities several orders of magnitude greater than those available today.

Perhaps the most significant applications of scientific computing lie not in the solution of old problems but in the discovery of new phenomena through numerical experimentation; the discovery of nonergotic behavior, such as the formation of solitons, and the presence of strange attractors as universal features common to a large class of nonlinear systems are examples of this scientific process.

Current and feasible supercomputers are extremely powerful scientific and engineering tools. They permit the solution of previously intractable problems, and motivate scientists and engineers to explore and formulate new areas of investigation. They will surely find significant applications not yet imagined. For these reasons, the Panel believes that it is in the National interest that access to constantly updated supercomputing facilities be provided to scientific and engineering researchers, and that a large and imaginative user community be trained in their uses and capabilities.

The U.S. has been and continues to be the leader in supercomputer technology and in the use of supercomputers in science and engineering. The present position of leadership is evidenced by the dominance of the supercomputer market by American producers and by the successful exploitation of supercomputing at national laboratories. However, the Panel finds that this position of leadership is seriously undermined by the lack of broad scale exploration, outside of a few national laboratories, of the scientific and engineering opportunities offered by supercomputing, and by a slowdown in the introduction of new generations of supercomputers. This threat becomes real in light of the major thrust in advanced supercomputer design that is being mounted by the Japanese Government and industry, and by vigorous governmental programs in the United Kingdom, West Germany, France, and Japan to make supercomputers available and easily accessible to their research and technological communities.

American preeminence in large scale computing has been a result of the confluence of three factors: the vitality of the U.S. computer industry, the far-sighted policies of the Federal government, and the leadership of scientists and engineers from universities and government laboratories. The Atomic Energy Commission, on the urging of John von Neumann, initiated the use of large scale computation in research and weapons design; NASA, prodded by Hans Mark, advanced the use of supercomputing in its scientific programs. American universities and government laboratories conducted the research that formed the basis for constructing and applying computers, trained the needed scientific and engineering personnel, and made computers and computing an essential tool in scientific and engineering research.

The Federal government vigorously implemented policies that supported these efforts, granted generous funds for computation, and, through its role as the major purchaser of scientific computers, provided the incentives and insured the market for these unique machines. Forward-looking corporations exploited the scientific and engineering opportunities, developed an advanced industrial technology, and created this most vital component of the American economy.

During the 1970's the Federal government retreated from its support of large scale computing in universities. The NSF program to provide and expand university computing facilities for scientific and engineering research was terminated in 1972; at about the same time IBM discontinued its generous discounts for the purchase of computing equipment by academic institutions as a result of pressures from the Justice Department and competitors. Since then large scale university computing facilities have withered while the action shifted to national laboratories and to industrial users. The most advanced scientific computer of the early seventies, the CDC 7600, was not installed on a single American campus, although it was available to researchers at several foreign universities and research institutes.

This continues today. With the exception of two universities and a few government laboratories, either dedicated to special tasks or specific disciplines, universities and government research installations lack supercomputers.[9]

Within the Government, fully integrated supercomputer facilities are found exclusively at dedicated national laboratories such as Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory (LLNL) in support of weapons and fusion programs; and NASA installations, the Geophysical Fluid Dynamics Laboratory, and the National Center for Atmospheric research (NCAR) in support of aerospace, oceanographic, and atmospheric research programs. The National Magnetic Fusion Energy Computer Center (NMFECC) of LLNL is accessible in interactive mode to researchers at remote locations by a high speed network.[10] On the other hand, the National Bureau of Standards and most DOD laboratories do not have supercomputers and far too few universities have the specialized computational equipment needed for scientific computing (e.g., array processors). As a result of limited access to supercomputing facilities by the broad research community, significant research opportunities[11] have

----

9 See, in Appendix I, <u>Partial Inventory and Announced Orders of Class VI Machines</u>.

10 See Appendix I for a description of some of these facilities.

11 See, for example, the "Prospectus for Computational Physics," [2], "Future Trends in Condensed Matter Theory and the Role and Support of Computing" [4], "Report by the Subcommittee on Computational Capabilities for Nuclear Theory," [28], and "An Assessment of Computational Resources Required for Ocean Circulation Modeling," [35].

been missed, and the younger generation of researchers is inadequately trained in large scale computing.

The need for access to large scale computational facilities has become so critical[12] that several universities, assuming significant financial risks, have felt it essential to acquire supercomputers. Several more are seriously considering doing so, and others are in the process of forming consortia for this purpose. Some of these endeavors have applied for Federal funding without which they may have financial difficulties. Other groups are pressing funding agencies to expand or replicate highly successful facilities, such as those at NCAR, NMFECC, and NASA-Ames, at universities or at national laboratories. Class VI scientific remote computing services are available from a few commercial service bureaus, but neither the academic nor the government research communities make extensive use of this resource. This seems due to a combination of lack of funds for computing services, the perceived high cost associated with these services, and a lack of sophisticated high-speed networking facilities. It is an indication of the absence of a national plan that a substantial number of leading scientists are clamoring for access to supercomputers at the same time that some supercomputing facilities are underutilized.[13]

A supercomputer is a general purpose scientific instrument serving a broad and diverse base of users. The decline of supercomputing at universities is analogous to the decline of instrumentation; neither large scale computing nor instrumentation can be sustained in a stable manner through funding of individual research grants, where their costs must compete with that of scientific personnel.[14]

The findings of the Panel regarding the development of supercomputers are as alarming as the findings on their access and availability. The U.S. supercomputing market is, at this time, dominated by Cray Research (CRAY-1) and Control Data Corporation (CYBER 205).[15] The Japanese vendors, Hitachi (S-210/20) and Fujitsu (VP-200), have announced the delivery of supercomputers in the near future, and these machines appear to be comparable to the available American systems. The Japanese are striving to become serious competitors of domestic manufacturers, and U.S. dominance of the supercomputer market may soon be a thing of the past. The Japanese Government-sponsored National Super Computer Project[16] is aimed at the development, by 1989, of a machine one thousand times faster than current machines. There is no comparable technical program in the U.S.. The Panel notes that in the case of the NASA Numerical Aerodynamic Simulator, a very high performance supercomputer, no acceptable proposals for its development

----

12 See, in Appendix III, The Supercomputer Famine in American Universities, by L. L. Smarr.

13 Supercomputer cycles are available at the University of Minnesota CRAY-1 and at Colorado State University CYBER 205.

14 See, in Appendix III, the paper by R. G. Gillespie. Most supercomputer facilities are funded directly by the Federal Government.

15 See, in Appendix I, Partial Inventory and Announced Orders of Class VI Machines.

16 See, in Appendix III, Japan's Initiatives in Large Scale Computing, by L. Lee.

were received. Neither of the two competing vendors could assure NASA that they would meet the performance requirements. Rather than developing new products, the vendors attempted to fit all NASA requirements to their existing product line.

Upon review of previous studies, the Panel also finds that the power of current and projected supercomputers is insufficient to meet existing needs in science, engineering, and technology, both military and civilian.[17] Research at universities and in the computer industry has indicated that future generations of very high performance computer systems may have parallel architectures radically different from the conceptually sequential architectures of today's supercomputers. There are many candidate architectures that must be evaluated before commercial feasibility can be established.[18] Simultaneously, the rapid and continuing advance of microelectronic technology makes it feasible to build such parallel machines. There is also a need for improvement of component performance.

The Panel believes that under current conditions there is little likelihood that the U.S. will lead in the development and application of this new generation of machines.[19] Factors inhibiting the necessary research and advanced development are the length and expense of the development cycle for a new computer architecture, and the uncertainty of the market place. Very high performance computing is a case where maximizing short-term return on capital does not reflect the national security or the long-term national economic interest. The Japanese thrust in this area, through its public funding, acknowledges this reality.

The Panel estimates[20] that the present annual investment in basic research on algorithms, software, and architecture is between 5 and 10 million dollars, while the annual advanced development expenditures for supercomputers (beyond Class 6 machines) are between 20 and 40 million dollars. This is contrasted with the development cost for a new high-speed conventional architecture system of approximately 150 million dollars, as well as the estimated 200 million dollars national superspeed computer project in Japan. The panel considers current levels of United States investments insufficient to maintain leadership in supercomputers.

The Panel believes that U.S. leadership in supercomputing is crucial for the advancement of science and technology, and therefore, for economic and national security.

----

17 See, for some illustrative examples, Appendix II. Also notable is "Trends and Pacing Items in Computational Aerodynamics", by D.R. Chapman, [40].

18 See, in Appendix III, a series of contributions to the Workshop. In particular, the papers by Dennis, Gajski, et al., Ris, and Fernbach; also the report of Group 3 of the Lax Panel in the Supplement.

19 See, in Appendix III, Why the U.S. Government Should Support Research on and Development of Supercomputers, by B. Buzbee.

20 Members of the Panel and of the Organizing Committee have conducted a survey to estimate the current total national investment in research and advanced development for supercomputer and supercomputing, both public and private. This survey included research and development costs but excluded funding for the acquisition, maintenance, and operation of supercomputer facilities.

Recommendations

The Panel recommends that the present needs and challenges to U. S. leadership in scientific computing and supercomputer technology be addressed with high priority. To this end, the Panel has set forth the background for planning and policy issues, outlined some options, and noted that current total funding in this area is insufficient to maintain the Nation's leadership in large scale computing. The Panel has avoided recommendations of a programmatic and organizational nature; these, and their implementation, are best left to the appropriate government agencies. These agencies must work together to respond to the issues raised and put together a detailed coherent program whose components are responsive to their individual missions. The program plan should contain a clear statement of goals, directions, and roles for the academic, industrial, and Federal government segments; responsibilities of the participating Federal agencies; and funding required.

The Panel recommends that a long-term National Program on Large Scale Computing should be initiated immediately, with the participation of the appropriate Federal agencies, the universities, and industry. The goals of this National Program should be:

1. Increased access for the scientific and engineering research community through high bandwidth networks to adequate and regularly updated supercomputing facilities and experimental computers;

2. Increased research in computational mathematics, software, and algorithms necessary to the effective and efficient use of supercomputer systems;

3. Training of personnel in scientific and engineering computing; and

4. Research and development basic to the design and implementation of new supercomputer systems of substantially increased capability and capacity beyond that likely to arise from commercial sources.

This Program should be coordinated by an interagency policy committee consisting of representatives of the appropriate Federal agencies, including DOC, DOD, DOE, NASA, and NSF. A Large Scale Computing Advisory Panel, with representatives from Government, the universities, and industry, should be established to assist in the planning, implementation, and operation of the Program.

The Panel finds two points that require emphasis:

As the few successful facilities amply demonstrate, solution of the problem of access to supercomputing facilities, on a national basis, is possible.

Secondly, the domestic computer industry must allocate its scarce research and development funds to meet all the commercial opportunities and competitive challenges. Supercomputing enjoys a priority within the computer industry. But this priority, which reflects competitive commmercial conditions, does not reflect the entire national scientific and security interest. It is not reasonable to rely solely on industry's own initiatives and resources in this area.

Possible Approaches for the National Program

The Panel has received many suggestions[21] for carrying out the thrusts of the proposed National Program. We outline here those considered most promising.

1. Access: There appear to be three approaches to provide reliable and efficient access to supercomputers to the research and development community. Common to all these is the development of a nation-wide interdisciplinary network[22] through which users will have access to facilities. This network should connnect all supercomputer facilities (except those dedicated to very special tasks), including commercial supercomputing centers and experimental machines.

   - The most expedient and perhaps least expensive way to provide supercomputer access to the broad range of scientific and engineering researchers is to enhance supercomputer capacity and staff at existing centers which have demonstrated sophisticated capabilities for providing large scale computing.

   - Provide supercomputers to selected government laboratories without such facilities and make them available to the broad research and development community through networking. In addition, there should be sharing and enhancement of current supercomputer facilities located at universities and government laboratories.

   - Establish additional regional centers at selected universities, interconnected with existing facilities at other universities and government laboratories.

   The existence of a national network would permit combinations of these nonexclusive options, as well as the appropriate use of commercial services. The mechanisms for funding these facilities and allocating access should be carefully studied.[23]

----

21 see position papers in Appendix III.

22 The NMFECC network, described in Appendix I, is the example repeatedly mentioned for emulation because of its high bandwidth. The ARPANET network is often mentioned because of its interdisciplinary nature.

23 See the position paper by R. G. Gillespie in Appendix III.

The above recommendations are related to the problems of access to general purpose supercomputer facilities. It should be noted, however, that there are scientific and engineering problems that can perhaps be better and more economically attacked by specialized supercomputing facilities and by sophisticated array processors.[24] The Panel recommends that, as part of the National Program, significant emphasis be placed on providing this specialized equipment to the research community. Finding the proper balance between investments on these two types of facilities requires a careful analysis, at the multidisciplinary and interagency level.

2. Research in Software and Algorithms

Today's supercomputers are a major departure from traditional sequential machines. Future significant improvements may have to come from architectures embodying parallel processing elements - perhaps several thousands of processors. In order to exploit today's vector processors and future parallel processors, entirely new algorithms must be conceived. Research in languages, algorithms, and numerical analysis will be crucial in learning to exploit these new architectures fully. The contributions of numerical analysis, computational mathematics, and algorithm design to the practice of large scale computing is as important as the development of a new generation machines.

3. Training

Another important component of this National Program is the development of an imaginative and skilled user community in supercomputing. There is a considerable shortage of appropriately trained personnel and of training opportunities in this area. Forms of institutional encouragement, such as NASA's special fellowships in the area of numerical fluid mechanics, special summer schools, and special allocation of access time to supercomputers for those projects that involve graduate students, should be considered. Some of the more mathematical aspects of these activities can be accomplished independently of the machines

----

[24] Some of these problems arise in a number of areas associated with experimental physics. See, in Appendix III, the letter from A. E. Brenner, Jr. See, also in the same Appendix, the position paper by K. Wilson on the use of array processors.

on which the actual calculations are done; however, the true integration of methods and their implementation cannot be done without access to supercomputers. The nature of the machine architecture has a very profound effect on the numerical methods, on the algorithms, and of course on the software. Thus, while being trained, students must have access to state-of-the-art computers. Today such training is virtually nonexistent; yet the skills gained from such training are essential to science and engineering.

4. Research and Development for New Supercomputers

There are serious component and architectural problems that must be solved as part of the development of future generations of supercomputers. The unique strengths of industry, universities, and Government laboratories should be brought together for this purpose. A group of panelists from this workshop, with the aid of a number of experts from industry and universities, has produced a report[25] which describes one such program.

Since a great deal of careful analysis and detailed planning is required before the proposed National Program can be implemented, the Panel urges that its recommendations be acted upon as soon as possible.

----

25 See, in the Supplement, A Program for Development of Very High Performing Computer Systems, by J. C. Browne and J. T. Schwartz.

III. SELECTED BIBLIOGRAPHY

Selected Bibliography

[1] "A Push for Major Improvements in Both Large-Scale Computing Capability and Simulation Techniques for Physical Processes," memo to Users of Large Scale Computing for Simulation in Government, Industry and Universities from Kenneth Wilson, James A. Weeks Professor of Physical Science, Cornell University, July 15, 1981.

[2] "Prospectus for Computational Physics," Report by the Subcommittee on Computational Facilities for Theoretical Research to the Advisory Committee for Physics, Division of Physics, National Science Foundation, March 15, 1981, William H. Press, Chairman.

[3] "A Proposal Submitted to the United States Department of Energy for a National Nuclear Computer Center at the University of Maryland," and Appendix A., "A Report on Computer Needs for Nuclear Theory," William MacDonald, Chairman, July 15, 1979.

[4] "Future Trends in Condensed Matter Theory and the Role and Support of Computation," report by the Condensed Matter Theory Panel, published under the auspices of the council on Materials Science for the Division of Materials Sciences, U.S. Department of Energy, J. D. Joannopoulos and A. N. Berker, Co-Chairman, September 22, 1981.

[5] "A Preliminary Proposal to the Department of Energy submitted by Colorado State University for the Support of a Class VI Computer," by Robert W. Hayman and Gerold R. Johnson, Colorado State University, Fort Collins, Colorado, August 24, 1981.

[6] "NAS Project Status Report for the NAS User Steering Group," prepared by the Numerical Aerodynamic Simulator Project Office, National Aeronautics and Space Administration, Ames Research Center, July 1, 1980.

[7] "The Numerical Aerodynamic Simulator - Description, Status, and Funding Option;" a report prepared at the request of the Committee on Commerce, Science, and Transportation, United States Senate; Dec. 1981, NASA.

[8] Numerical Aerodynamic Simulator; Background, Requirements, and Justification; NASA, Ames Research Center, Revised January 1982.

[9] "Proposal for Research on Parallel Applicative Programming for Multiprocessor Systems," Rodney R. Oldehoeft, Stephen J. Allan, Colorado State University, and Burton J. Smith, Denelcor, Inc., October 1981.

[10] "Development Program, Scientific Computation Center," Florida State University Computing Center, Tallahassee, Florida.

[11] "Second Generation S-1 Systems: Computing and Memory Capacity for Demanding Navy Applications," A Review Presentation for Members of the Navy Technical Community, Phys. Brief 81-103, Lowell Wood, et al., S-1 Project, Lawrence Livermore National Laboratory, January 16, 1980.

[12] "Simulation of Turbulent Shear Flows: What can we do and what have we learned?" presented by invitation at the Seventh Biennial Symposium on Turbulence, Rolla, MO, by W. C. Reynolds, Department of Mechanical Engineering, Stanford University, September 21, 1981.

[13] "A Commentary on Computational Mechanics," J. T. Oden and K. J. Bathe, Applied Mechanics Reviews 31 (1978) pp. 1053-1058.

[14] "Draft Proposal to the National Science Foundation for support of Fellowship Program to Institute for Computation, the Center for Advanced Technology at Colorado State University," George G. Olson, June, 1981.

[15] "Report on the Utilization of Large Scale Computers," Committee on the Utilization of Large Scale Computers in ASGA, Bennett Miller, chairman, July 1, 1977.

[16] "Computing Division Two Year Operational Plan, FY 1981-1982," Robert H. Ewald, W. Jack Worlton, Margery McCormick, Los Alamos National Laboratory, LA-8725-MS, February 1981.

[17] Letter to Dr. Alvin W. Trivelpiece, Director, Office of Energy Research from Dr. David A. Shirley, Director, Lawrence Berkeley Laboratory with Conceptual Proposal for Energy Research Network, September 14, 1981.

[18] "Nuclear Weapons and Computers, Summary: The Role of Large Scale Computers in the DOE Nuclear Weapons Program," the Gold Report, prepared for the Director of Military Applications, U. S. Department of Energy, March 1978.

[19] "DOE Research in Utilization of High-Performance Computers," B.L. Buzbee, W. Jack Worlton, G.A. Michael, G. Rodrigue, LA-8609-MS, Los Alamos National Laboratory, December 1980.

[20] "A Philosophy of Supercomputing," Jack Worlton, LA-8849-MS, Los Alamos National Laboratory, June 1981.

[21] "Department of Energy FY 1983-1987 ADP Long Range Plan," Office of ADP Management, December 1981.

[22] "An Assessment of Trends in the Solid State Sciences," National Research Council, July 1980.

[23] "Future of Nuclear Science," by Ad Hoc Panel on the Future of Nuclear Science, 1977.

[24] "Long Range Plan for Nuclear Science," excerpts from a report by the Nuclear Science Advisory Committee, July 30-August 4, 1979.

[25] "Draft II - Astronomy Survey Committee: Report of the Panel on Data Processing and Computational Facilities," May 19, 1980.

[26] "Workshop on Computational Astrophysics," NASA Ames Research Center, February 1980.

[27] "Magnetic Fusion Energy and Computers: The Role of Computing in Magnetic Fusion Energy Research and Development," Office of Fusion Energy, DOE, October 1979.

[28] "Report by the Subcommittee on Computational Capabilities for Nuclear Theory to the DOE/NSF Nuclear Science Advisory Committee," December 1981.

[29] "Purdue Center for Parallel and Vector Computing," Dennis Gannon, John Rice, Saul Rosen, April 23, 1982.

[30] "Universities and the Future of Computer Technology Development," Kenneth G. Wilson, Laboratory of Nuclear Studies, Cornell University, Ithaca, New York.

[31] "Present vs. Future Supercomputers," notes by Kenneth Wilson, Newman Laboratory, Cornell University, May 1982.

[32] "A Data Flow Supercomputer," Jack B. Dennis, G-R Gao, Kenneth W. Todd, Computation Structures Group Memo 213, MIT Laboratory for Computer Science, Cambridge, Massachusetts, March 1982.

[33] "Foreign Travel Trip Report: Visits with Japanese Computer Manufacturers," R. Ewald, Bill Buzbee, J. Worlton, CDO/82-6782A, Computation Division, Los Alamos National Laboratory, April 5, 1982.

[34] "Supercomputers," Ronald D. Levine, Scientific American, January 1982.

[35] "An Assessment of Computational Resources Required for Ocean Circulation Modeling," Ad Hoc Committee on Computing Resources and Facilities for Ocean Circulation Modeling, Ocean Sciences Board, Commission on Physical Sciences, Mathematics and Resources, National Research Council of the National Academy of Sciences, 1982.

[36] "Nearest Neighbor Concurrent Processor," E. Brooks, G. Fox, R. Gupta, O. Martin, S. Otto, E. De Benedictis; California Institute of Technology, DOE Research and Development Report, CALT-68-867, September 1981.

[37] "Why the U.S. Government Should Support Research On And Development Of Supercomputers," Bill Buzbee, Los Alamos Report LA-UR-82-1968, July 1982.

[38] Los Alamos SCIENCE, Los Alamos National Laboratory, Vol. 2, No. 2 (1981).

[39] Justification for Class VI Scientific Computing Capability at Sandia Laboratories, Albuquerque; Sandia Laboratories, February 1979.

[40] "Trends and Pacing Items in Computational Aerodynamics," D. R. Chapman, Lecture notes in Physics, 141, Springer-Verlag, 1981.

IV. SUPPLEMENT

Organization of the Workshop on Large Scale Computing for Science and Engineering

Members of the Panel on Large Scale Computing in Science and Engineering

William F. Ballhaus	NASA-Ames
James C. Browne	University of Texas, Austin
Brice Carnahan	University of Michigan
Michael Creutz	Brookhaven National Laboratory
Richard H. Gallagher	University of Arizona
Herbert Keller	California Institute of Technology
John Killeen	Lawrence Livermore National Laboratory
Peter D. Lax, Chairman	New York University, National Science Board
Walter M. Macintyre	National Center for Atmospheric Research
Steven Orszag	Massachusetts Institute of Technology
Werner Rheinboldt	University of Pittsburgh
Allan Robinson	Harvard University
Jacob Schwartz	New York University
Edward Wilson	University of California, Berkeley
Kenneth Wilson	Cornell University

Organizing Committee of the Workshop on Large Scale Computing in Science and Engineering

Donald Austin	Department of Energy
Melvyn Ciment	National Bureau of Standards
Kent Curtis	National Science Foundation
David Fox	Air Force Office of Scientific Research
J. M. Greenberg	National Science Foundation
Carl Hall	National Science Foundation
E. F. Infante	National Science Foundation
Peter D. Lax, Chairman	New York University, National Science Board
Marshall Lih	National Science Foundation
Lawrence Lee	National Science Foundation
James Pool	Office of Naval Research
Paul Schneck	National Aeronautics and Space Administration

June 14, 1982

AGENDA FOR
WORKSHOP ON LARGE-SCALE COMPUTING FOR SCIENCE AND ENGINEERING

Location

National Science Foundation
1800 G Street, N.W., Rm. 540
Washington, D.C. 20550

June 21, 1982

Time	Topic	Leader
8:30 a.m.	Preamble	Lax
9:00 a.m.	Large-Scale Computing Needs	Keller, Robinson
10:00 a.m.	BREAK	
10:15 a.m.	Arrangements for Making Computing Power Available (Ballhaus, Hayes, Killeen, Macintyre)	Orszag
12:00	LUNCH	
1:00 p.m.	Computer Technology in the U.S. (Presentations) (Ballhaus, Buzbee, Michael, Schneck, Wilson)	Schwartz
3:00 p.m.	BREAK	
3:15 p.m.	Discussion (Fernbach, Patton, Olson, Rice)	Lax
6:00 p.m.	Adjournment (except for Panel Members)	

June 22, 1982

Time	Topic	Leader
8:30 a.m.	Charge to Panel	Lax
9:00 a.m.	Group Working Sessions	
12:00	LUNCH	
1:30 p.m.	Plenary Session I	
2:30 p.m.	BREAK	
2:45 p.m.	Group Working Session	
5:00 p.m.	Closing Remarks	Lax

List of Participants of the
Workshop on Large-Scale Computing
in Science and Engineering
Washington, DC
June 21-22, 1982

W. Richards Adrion - National Science Foundation
Carl Amthor - Associated Universities, Inc.
Lloyd Armstrong - National Science Foundation
Donald Austin - Department of Energy
Steven Bachman - Control Data Corporation
William F. Ballhaus - National Aeronautics and Space Administration, Ames
Bruce H. Barnes - National Science Foundation
F. Brett Berlin - Cray Research Incorporated
William Bernard - National Science Foundation
Alfred Brenner - Fermi Laboratory
James C. Browne - University of Texas, Austin
Alphonse Buccino - National Science Foundation
Bill Buzbee - Los Alamos National Laboratory
Andrew Callegari - Exxon Research and Engineering Company
P. Cannon - Star Technology
Jagdish Chandra - Army Research Office
Alan Charlesworth - Floating Point Systems, Incorporated
John C. Cherniavsky - National Science Foundation
Y. T. Chien - National Science Foundation
Melvyn Ciment - National Bureau of Standards
Andrew Collier - Time Magazine
Curtis Collins - National Science Foundation
Michael Creutz - Brookhaven National Laboratory
Kent K. Curtis - National Science Foundation
Carl G. Davis - BMD Advanced Technology Center
Dorothy Deringer - National Science Foundation
A. Erisman - Boeing Computer Service
Sidney Fernbach - Consultant, Information Systems
Judith Flippen-Anderson - Naval Research Laboratory
David Fox - Air Force Office of Scientific Research
J. Gaylor - Boeing Computing Services
Robert Gillespie - University of Washington
Paul Gilmore - Goodyear Aerospace Corporation
Randolph Graves, Jr. - National Aeronautics and Space Administration
James M. Greenberg - National Science Foundation
Lloyd Griffiths - University of Colorado
Carl Hall - National Science Foundation
W. F. Hanrahan - CBEMA
Allen J. Hansen - Burroughs Corporation
Richard Harris - Cray Research Incorporated
W. C. Harris - National Science Foundation
Edward Hayes - National Science Foundation
Kenneth Hoffman - Massachusetts Institute of Technology
E. F. Infante - National Science Foundation
Glenn Ingram - National Bureau of Standards

R. F. Johnson - Floating Point Systems, Incorporated
Ronald E. Kagarise - National Science Foundation
Robert Kahn - Advanced Research Projects Agency
James S. Kane - Office of Energy Research, Department of Energy
Herbert Keller - California Institute of Technology
John Killeen - Lawrence Livermore National Laboratory
Ronald Larsen - National Aeronautics and Space Administration
Peter Lax - New York University - National Science Board
Lawrence Lee - National Science Foundation
Stephen F. Lundstrom - Burroughs Corporation
Walter M. Macintyre - National Center for Atmospheric Research
J. V. Martinez - Department of Energy
Paul Mathews - Office of Military Applications, DOE
George Michael - Lawrence Livermore National Laboratory
K. Mohan - Office of Management and Budget
Thomas Nash - Fermi Laboratory
David B. Nelson - Department of Energy
Robert Numrich - Control Data Corporation
C. E. Oliver - Air Force Office of Scientific Research
George Olson - Colorado State University
Barbara Onestak - National Science Foundation
Steven Orszag - Massachusetts Institute of Technology
A. R. Osborne - Johns Hopkins University
Seymour Parter - University of Wisconsin
Peter Patton - University of Minnesota
Howard Penn - National Science Foundation
Victor Peterson - National Aeronautics and Space Administration, Ames
A. H. Pettifor - Office of Science and Technology Policy
Elliot Pierce - Department of Energy
Irwin Pikus - National Science Foundation
James Pool - Office of Naval Research
Jesse Poore - Georgia Institute of Technology
B. Prosswimmer - Boeing Computing Services
Lawrence Rapagnani - Air Force Weapons Laboratory
Werner Rheinboldt - University of Pittsburgh
John Rice - Purdue University
Fred Ris - International Business Machines
Allan Robinson - Harvard University
Paul N. Romani - National Science Foundation
William G. Rosen - National Science Foundation
Sherwin Ruben - Goodyear Aerospace
John Ryff - National Science Foundation
Russ Saunders - Denelcor, Incorporated
Paul Schneck - National Aeronautics and Space Administration
Jacob Schwartz - New York University
Burton J. Smith - Denelcor, Incorporated
W. R. Smith - Cray Research Incorporated
P. Sorianic, Jr. - Star Technology
Dale Sparks - Los Alamos National Laboratory
Paul G. Spyropoulos - National Science Foundation
Leonard Stans - Air Force Weapons Laboratory
Martin T. Storma - Cray Research Incorporated
Jane Stutsman - National Science Foundation
Alvin Thaler - National Science Foundation

Eric Thompson - National Science Foundation
D. Triantaphyllou - Mitre Corporation
Jean M. Vocar - Goodyear Aerospace
Kenneth Wallgren - National Aeronautics and Space Administration
Hugh Walsh - International Business Machines
Frederick Warther - Denelcor, Incorporated
Edward Wilson - University of California, Berkeley
Kenneth Wilson - Cornell University
Robert Woods - Department of Energy

List of Participants of the
Follow-up Workshop on Large Scale Computing for
Science and Engineering
Shanty Creek, Michigan
August 23-24, 1982

Jim Browne - University of Texas
Kent Curtis - National Science Foundation
Jack Dennis - Massachusetts Institute of Technology
David Fox - Air Force Office of Scientific Research
David Kuck - University of Illinois, Urbana
George Michael - Lawrence Livermore National Laboratory
Ed Oliver - Air Force Office of Scientific Research
Fred Ris - IBM Research
Paul Schneck - NASA/Goddard Space Flight Center
Jack Schwartz - New York University
Burton Smith - Denelcor, Inc.
Larry Snyder - Purdue University
Marty Storma - Cray Research
Bob Voigt - NASA/Langley Research Center
Ken Wallgren - NASA/Goddard Space Flight Center

ON SUPERCOMPUTING

K. Wilson
Cornell University

Summary and Recommendations

The lack of large scale scientific computing* resources for basic university research has become a major problem for U.S. science. The largest need for these resources is in the theoretical science community, where they are required for computer simulation of complex physical processes that cannot be studied by traditional analytic means.

The immediate national needs are the following:

1) A national network linking all scientists involved in open basic research, vastly generalizing the existing Arpanet and Plasma fusion energy networks.
2) A development program in support of large scale scientific computing, encompassing hardware, systems software, and algorithm development and carried out as a collaboration between knowledgeable members of the scientific community, the computer science and electrical engineering community, and the computing industry.
3) Building an adequate equipment base (computers, peripherals, and network access) for training and theoretical research in universities.

---

* "large scale scientific computing" includes any problem requiring hours or more on a superminicomputer and any problem requiring major software development. It excludes routine evaluation of integrals or ordinary differential equations, and any problem solved on a pocket calculator.

4) Providing adequate access by researchers on the network to special and general purpose facilities at the National Laboratories and elsewhere, for computing needs that go beyond the base level.

An interagency committee should be established to channel support for scientific computing into relevant agency programs funding basic research. The interagency committee requires a strong scientific advisory board, thoroughly knowledgeable about both scientific computing and developments in computing technology, to set priorities and make sure that long term as well as short term needs are met.

Background

A major change has taken place in the role of computers in basic scientific research. The use of large scale scientific computing for simulation is growing very rapidly in universities in very many areas of basic theoretical science. Previously, the principal uses of computers have been for data collection and analysis (such as in high energy experimental physics and space physics) and in very specialized areas of theory (for instance in specific areas of fluid dynamics or in plasma fusion studies), where there are major simulation efforts centered at the national laboratories.

Many of the simulations that have been performed recently in universities are of a preliminary nature. Numerical approximations used have been very rough; important aspects of the science involved have been simplified or neglected altogether. The preliminary results have been promising, but no more; the preliminary work must be followed up by much more careful computations taking into account all important scientific principles and including careful determination of numerical truncation errors. In a number of areas, further progress in both analytic and numerical studies is dependent on reliable completion of simulations that are in this preliminary stage. Specific examples of such simulations in elementary particle theory, statistical mechanics, and astrophysics are given in Appendix

A major instrumentation program, outlined above, is required to enable completion of the existing simulation efforts. The nature and magnitude of computing capacity that will be needed varies considerably from case to case. A number of computations in the very general area of fluid dynamics could be completed if computing access in universities was made comparable to that presently available at national laboratories. Other computations, in elementary particle theory and statistical mechanics or turbulence will require computers to be built that are very much more powerful than any current supercomputers. See Appendix for examples of computing support needs.

Several other changes have taken place that underscore a need for a computing initiative for theory.

Science is becoming more interdisciplinary than ever before and faces new demands from the high technology economy. To meet these needs a new level of interchange must take place between scientists in widely different disciplines. For example, renormalization group ideas are being pursued in subjects ranging from quantum field theory to oil propagation in reservoirs to polymer folding problems; scientists in all these fields must have access to the latest developments in all the other areas. Computer networks provide extraordinarily powerful means of communication that cannot be duplicated otherwise. A researcher on Arpanet can broadcast a request for information about a subject totally outside the researcher's own field of expertise and expect replies from experts on the subject asked about. These communication capabilities will be greatly enhanced by use of the emerging technology of "desk top work stations". The newest generation of scientists, who see this technology coming and understand its extraordinary capabilities for improving scientific productivity and making science more rewarding, are becoming demoralized because of continued delays or inaction in making network access generally available nationally or even within single universities.

Another recent development is the soaring enrollment in computer science courses nationwide as students realize that they must be prepared to work with computers regardless of their field of study or later career. Estimates at Cornell are that 75% of all undergraduates will shortly be taking the basic computer science course. The time has come to provide training in large scale scientific computing throughout undergraduate and graduate programs for all science majors. A considerable upgrade of current facilities will be required to support the training effort. There can be no flinching from introducing students to very powerful computing, since by the time they graduate and have entered the work force, the most powerful computers available while they were students will have become cheap and commonplace, and graduates will have to be able to work with them effectively to be competitive in the work force.

The large scale computing scene and the role of universities in this scene is changing. A great variety of limited purpose computing systems are emerging from the computing industry which will be as important as today's mainframes and supercomputers. Even a listing of current or shortly to be delivered systems illustrates this diversity the Floating Point Systems attached processors, the ICL DAP and the Goodyear MPP, the Denelcor HEP, and the CDC Advanced Flexible Processor all have unique architectures and each has strengths for specific application areas that mainframes or vector supercomputers cannot match.

Each new large scale computing system with a new architecture and new software faces a long breaking in period before the hardware and software work with sufficient reliability to support large user communities. Universities provide the most effective framework for the breaking in of a new computer. University faculty carrying out large scale simulations need access to the most cost-effective computing systems coming on the market in order to complete their work; graduate students and undergraduate work study students help overcome the innumerable minor problems new systems present; computing support personnel use this experience to help the computer manufacturer find and correct problems early in the product cycle. Computer science students can work with computing support personnel to flesh out inadequate software; new architectures can provide inspiration for major projects in computer science and electrical engineering. A university installation showcases a new computing system for the entire world. The Array Processor project at Cornell has already illustrated most of these capabilities of universities.

Despite the enormous benefit an early university installation provides to a computing manufacturer, the manufacturer rarely can afford to provide the system for free. However, there are many ways a manufacturer can reduce the costs to a university for a new system, and this in turn can provide major cost reductions for a governmental computing initiative. At the same time that the government saves money, the U.S. computing industry can be strengthened against major international competition by the help universities can provide in bringing new systems to market.

Because of the great variety of new computing systems that need to be placed in universities in support of the simulation effort, a national computer network is essential to enable individual researchers to work remotely on the computer most suited to their problems, and to equalize access among different university and national laboratory researchers to the computing systems that exist.

A strong basic research effort involving computer simulation is essential to the health of industrial research and development. Industrial products and processes are developing rapidly in response to increasing international competition, changing raw material mixes, etc. Computer simulation, when feasible, is usually the least expensive and most rapid method available for prototyping and optimization of new products and processes. Use of computer simulation must grow rapidly in industrial research and development if industry is to stay competitive. Unfortunately the bulk of U.S. scientists today are both ignorant and highly skeptical of large scale computer simulation, and they bring this attitude to industry, whether as employees, consultants, or members of industrial advisory boards. It takes many years of practical experience with large scale computing to be able to make sensible decisions or provide sound advice on simulation matters; very few U.S. scientists have this experience. The leaders of advanced research groups in industry who must build up aggressive computer simulation programs are often very isolated; they typically face higher level managers with little computing experience, computing support personnel with more concern for business data processing than scientific computing, and scientific support staff with little experience especially in large scale software.

The current simulation effort in universities sometimes involves subjects (such as turbulence) of major industrial engineering importance. Even in areas of no practical significance, such as general relativity, numerical simulation experience is invaluable training for students moving to industry and helps build a reservoir of university faculty with competence in large scale computing matters.

A number of the university simulation efforts if allowed to continue will encounter the same software complexity barriers that plague industry and governmental laboratories. Industry cannot do much about the software complexity problem in simulations because there is too much pressure from the backlog of production software to maintain and incessant deadlines to meet. The computing industry itself cannot do much because they have little expertise in the area of software for scientific simulation and the scientific simulation market is not large enough for its software problems to be given high priority.

The computing support initiative should include very strong encouragement for university scientists engaged in simulation to encounter software complexity problems and then collaborate with computer scientists to find long term solutions to this problem. Extraordinary ideas are circulating within the computer science community that can help deal with software complexity issues, but scientists are needed to help package the computer scientist's ideas in working software tools that will provide maximum benefit in the scientists' work. The computing resources presently available to university theorists are not powerful enough to allow major software difficulties to develop. Some of the major data processing efforts in experimental science have encountered software problems, as well as major productivity losses due to software development hassles, but the problems have never been severe enough to motivate the experimentalists to address the software complexity problem directly. There is already an awareness among a few members of the theory community that software problems are imminent and must be dealt with.

Justification for the Recommendations

1) The need for a national network linking all scientists has already been justified. The network must support computer mail, bulletin boards, newsletters, etc. which provide the basis for interdisciplinary communication. The network must support remote interactive computing sessions and computer-to-computer file transfer. Many researchers already have informal access to Arpanet, which provides these facilities but many others do not.
2) A major development program is needed in conjunction with any computing support initiative. The continued rapid development of VLSI (very large scale integrated circuits) is opening up a bewildering variety of computer architectures and designs, mostly of a special purpose nature. High performance computer design (beyond current

mainframe levels) will necessarily become a collaborative effort between a computing manufacturer and the scientists who will use the computer or their representatives, illustrated for example by the collaboration between NASA and Goodyear Aerospace Corporation on the design of the MPP image processing system. The architectures of these systems pose major software problems that the computing industry is ill-equipped to handle, especially for special purpose systems with limited markets. Scientists will have to seek help from computer scientists to get advanced software systems developed for new special purpose hardware, as NASA has already done for the MPP; this will provide a major source for computer science thesis projects. Even the most powerful computers modern technology can produce will not be sufficient for some of the most demanding simulations unless major advances are achieved in the algorithms used to represent problems numerically; scientists must collaborate with numerical analysts and applied mathematicians to seek the most efficient and reliable numerical techniques for their simulations.

3) The base level of computing support in universities must be continually modernized and built up in order to provide students with training in large scale computing that is not hopelessly obsolete before they graduate and to provide an informal atmosphere for exploratory computations. In the near future an up-to-date university computing support system would begin with a desk-top work station on every faculty member's desk and other work stations easily accessible to every student (graduate or undergraduate). These work stations would be linked via high speed networks to form clusters around a central "server" with disk storage space, printers,

and modest graphics facilities. The "server" would itself be linked to other campus facilities (including centralized very high performance graphics) and the national network. Inexpensive, high performance computers (such as attached processors) should be provided locally in sufficient quantity to minimize the remote computing load on the national network, while supporting heavy student and faculty use.

4) The National Laboratories have a long record of providing high performance computing. Their non-secret facilities need to be strengthened and made accessible to university researchers on a national network. A variety of high performance supercomputers and more specialized systems need to be provided to universities and likewise made available on the network. University sites for these systems should have effective centralized management and adequate interest and support to bring a new system into full operation quickly and economically. The most important function of these facilities, once in full operation, should be to enable completion of the scientifically most urgent simulations, in preference to work of lesser significance. Finishing a simulation, with reliable error analysis, can require orders of magnitude more computing power than the initial exploratory computations; careful and unbalanced resource allocation procedures will be needed to provide maximum scientific return from supercomputers and other expensive systems. It is not possible on any reasonable budget to allow all scientists unlimited access to these systems and still avoid total saturation and long job turn-around times which reduce user's productivity; severe allocation procedures will be required to prevent this saturation.

NEEDS AND OPPORTUNITIES FOR SUPERCOMPUTERS

H.B. Keller and J.R. Rice

1. Introduction.

It has been widely observed that a new discipline - Large Scale Scientific Computing - has been born. Many make the case that science and engineering is no longer divided into just two parts - theoretical and experimental; there is now a third equal part - computational. There is no national policy recognition of this new and basic scientific and engineering discipline nor of the importance that supercomputers, its main tool, will play in the future development of technology. The needs and opportunities that we shall address are concerned with long range and basic research goals. There are any number of short range, practical and very important accomplishments that can and will be made. But we are here concerned with the future of one of the basic developments in twentieth century science and technology and we are determined that the United States should continue to lead the world in these developments.

Supercomputers are a new kind of instrument for science and engineering. Telescopes and particle accelerators come to mind as analogies and, indeed, such instruments will always continue to have a basic influence on scientific developments. Such devices enhance the ability to observe the consequences of basic laws of physics on the largest and smallest scales possible. However, supercomputers enhance man's ability to reason rather than to observe. This is a completely different type of activity - it cuts across all fields - and we have not yet begun to see its ultimate implications. The possibilities are so profound that we must ensure that we are at the forefront in these developments.

The next section describes eight areas where supercomputers will have major impacts that are critically important to the economic and military health of the country. There are many more opportunities in the areas (see Section III) which are now or soon will be involved in large scale scientific computing. These areas represent the specific needs and opportunities for large scale scientific computing. The general needs and opportunities have not been properly recognized before. This is surprising because they are common to all users of large scale scientific computing; in the past the support has been justified in each specific area. These general needs center on (a) the training of researchers, (b) the development of algorithms and software and (c) the design and organization of supercomputers. These needs cannot be met without adequate access to Class VI (and beyond) computers. There is not now and never has been any organized national program in large scale scientific computing. One of our basic thrusts is that we must develope some form of a National Program in Large Scale Scientific Computing.

The computing needs for large scale scientific computing have never been met and existing technology will not meet them for the foreseeable future. The profit potentials of individual industrial and commercial applications are not sufficient, at present, to spur the required supercomputer development. Thus, an important part of any national program must include serious efforts in developing the supercomputers themselves.

2. AREAS OF MAJOR IMPACT

We give a sample where the supercomputers will have a major impact on areas of critical national importance.

* Aircraft design. Aircraft are now designed in "pieces" because no computer can simulate the entire aircraft and the flow of air around it. The wings, the tail, the landing gear, etc. are designed individually in detail (by using computers, of course). The engineers then build the plane from these pieces and the test pilot sees how well they work together. The first company (or country) to have computers powerful enough to design an aircraft as a whole will undoubtedly produce planes with superior performance.

* Submarine design. The scientific problems and current state of submarine design are very similar to those of aircraft design. It's very possible that submarines should look much like mackerel or sharks in order to have optimum efficiency.

* Geophysical exploration. Current computers can handle only crude, simplified, models of geological formations of existing or potential oil fields. More accurate simulations would increase secondary and tertiary recovery by well over 10-20%, equivalent to the discovery of a field the size of the Alaska north slopes or the North Sea.

* Atmospheric models. Current computers are so inadequate that national weather predictions can not even include the effects of the Rocky Mountains. More detailed models will dramatically improve short range predictions and allow for meaningful studies of long

term weather patterns.

* <u>Nuclear weapons</u>. The analysis and design of nuclear weapons currently consumes enormous amounts of computer power and yet drastic simplifications are made in most situations.

* <u>Electronic devices</u>. Supercomputers are now being used to design both the circuit layouts and the individual components of electronic chips. Yet current designs are all two-dimensional and it is clear that much greater performance (i.e. much cheaper chips) can be obtained with three dimensional designs and manufacturing. The computational requirements are so formidable that such chips are only in science fiction - for the moment.

* <u>Command and control</u>. Chemical refineries and power plants are now controlled by rather modest computers. Automated assembly lines with many robots require total computing power equal to that of the current biggest supercomputers. The analysis and response to an intense attack will require a Navy ship to have a supercomputer more powerful than any that exists now. Computational requirements increase dramatically as the complexity of the situation and response speed increases. We are just beginning to imagine what supercomputers can do in this area.

* <u>Disease control</u>. Even simple viruses are enormously complex molecules, it is a major computational project just to determine their structure. In principle, we can use simulation to determine their chemical behavior, test their reaction to various drugs and finally understand how to control them. This (and many other medical advances) must await computers that are hundreds of times more powerful than anything that exists now.

We have omitted many equally important areas (e.g. nuclear power plant accidents, circulation of the ocean, magnetic fusion energy, satellite photo analysis), but the message is the same: a country that wants to be at the forefront scientifically, militarily and economically, must have access to the best computers, the supercomputers.

3. DISCIPLINES AND PROJECTS USING LARGE SCALE SCIENTIFIC COMPUTATION

Several disciplines and projects have already demonstrated the value of Large Scale Scientific Computation. Common to all of these areas is the need to solve extremely complex problems; so complex that, to date, only gross simplifications of the full problems have been attacked with the aid of supercomputers. Included in this group of disciplines and projects which already have dedicated supercomputers are:

A1. Nuclear weapons research

A2. Atmospheric sciences

A3. Magnetic fusion energy research

A4. Aeronautical research and development

A5. Nuclear reactor theory and design

A6. Petroleum engineering

A7. Geophysics

A8. Intelligence (classified)

There are numerous other disciplines and projects that are in need of the same or greater computing power and that do not yet have it in any organized way. These areas are for the most part actively engaged in Large Scale Scientific Computation and have begun to try to acquire supercomputers. They are grouped as follows:

B1. Computational physics

B2. Computational mechanics and structural design

B3. Ocean sciences and underwater acoustics

B4. Computational chemistry and chemical engineering

B5. VLSI and circuit design

B6. Nonlinear optics and electromagnetic theory

B7. Computational fluid dynamics

In addition there are many areas that will shortly realize that Large Scale Scientific Computation is either vital to their continued development or can play a large role in solving some of their basic problems. These areas include:

C1. Astrophysics, planetary science and astronomy

C2. Economic modelling and operations research

C3. Biosciences

C4. Computational statistics and graphics

The above listed disciplines and projects are frequently interrelated, they are not merely of academic interest and include work in government laboratories and industry. Thus, the interest in and applications of Large Scale Scientific Computation clearly cut across all the standard scientific, technological, industrial and governmental lines.

A brief bibliography of documents addressing the need and opportunities in the areas covered above are contained in Supplement. In essentially all of these discussions the case is made that new knowledge and understanding of phenomena ranging from the basic laws of physics to the behavior of a nuclear power plant during an accident can be attained if more powerful computing equipment is available. Each significant increase in computing power can lead to a host of significant advances in each discipline or project. Thus, the effect of improved computing power can be multiplicative if the opportunity is taken of making this power broadly available.

4. LARGE SCALE SCIENTIFIC COMPUTING: A New Discipline

The basic use of computing in almost all of the above areas is either:

(i) to approximate the solution of complicated systems of nonlinear partial differential equations,

(ii) to model or simulate complicated physical phenomena or systems in terms of interacting simple systems.

These two procedures are identical in many cases. As the phenomena becomes more complex or the models more realistic, the demands on computing power increase rapidly. For example, in typical equilibrium fluid dynamics problems (e.g. simulation of a plane in steady flight) going from one to two or from two to three space dimensions increases the complexity (as measured by the number of unknowns) by a factor of 20 to 200. The operational count to solve equilibrium fluids problems is proportional to the cube of the number of unknowns and thus it increases by a factor of from 8×10^3 to 8×10^6. It is thus abundantly clear that the need for continual improvements in computing power will be with us for some time to come. In all fields using Large Scale Scientific Computation, the indicated estimate or very similar ones are given to show that increases of the order of 10^6 or greater in computing power are needed and will be extremely beneficial.

The payoff in being able to solve two or three space dimensional problems (compared to the previous one or two dimensional cases) can be enormous. Drag reduction calculations over full three dimensional ships or aircraft could alone save billions of dollars in fuel costs. Accurate simulation of three dimensional two-phase flows in oil fields could increase secondary and tertiary recovery by well over 10% - 20%. This is equivalent to finding oil worth over $100 billion.

The entire Large Scale Scientific Computation community agrees that future supercomputers will operate in a parallel manner. They also agree that it is extremely difficult to plan software, algorithms and numerical methods that will take full advantage of this parallelism. Furthermore people will just not devote themselves to such difficult tasks unless they can have access to the supercomputers to test their results.

The role of software in current supercomputers already shows the striking but disturbing features to be expected. On one of the most basic algorithms (Gaussian elimination) which pervades Large Scale Scientific Computation , the use of different Fortran compilers and some "tweaking" with handwritten code produced running programs with a maximum to minimum speed ratio of 60. All the compilers used are considered quite good! Such gross variations do not occur on the standard software for serial machines. Thus, even with the currently existing minimal parallelism (or rather pipelining) we do not yet know how to produce near optimal software. It will be even more difficult for the coming generations of supercomputers.

The contributions of numerical analysis and algorithms design to the practice of Large Scale Scientific Computation is, in a very real sense, as important as the development of new generation machines. That is the speed up or improvement on what can be effectively computed comes as much from the numerical methods and their algorithmic implementation as it does from hardware improvements. Indeed a recent study (J.R. Rice, Numerical Methods, Software and Analysis, McGraw-Hill, 1982) of algorithms for solving elliptic problems in three dimensions found a speed up of 5×10^{10} from 1945 to 1975! This is much greater than the improvement in going from in I.B.M. 650 to a Cray-I. These estimates do not include mulit-grid methods, currently in active development, and it is reasonable to assume another factor of 10^3 will have been achieved during 1975-85. It is clear that Large Scale Scientific Computation is the most significant and powerful scientific instrument developed in the 20th century, the development is not nearly complete and that very shortly it will not be possible to do first rate scientific research in many areas without the best supercomputers and methodology of Large Scale Scientific Computing.

It seems generally to be the case that discipline or project oriented scientists do not use the latest or best numerical methods or algorithms. Also, it is rare, but not unknown, that improvements in numerical methods or algorithms are made by such project scientists using Large Scale Scientific Computation. Thus, an important aspect of our program must be an attempt to close this gap between development and use of new ideas.

5. GENERAL NEEDS

The general needs in large scale scientific computing are in four categories:

1. Trained people
2. Software systems
3. Algorithm design and numerical analysis
4. Supercomputer hardware and systems

Some view that the last category as part of electrical engineering and/or computer science hardware. Where it is included is not nearly so important as that it be recognized and adequately supported.

Some of the more mathematical aspects of these activities can be accomplished independently of actual machines. However, the complete integration of methods and their implementation cannot be done without access to supercomputers. The nature of the machine architecture has a very profound effect on the numerical methods, the algorithms and, of course, the software. One does not devise new methods or learn to think "in parallel" overnight, so a whole range of scientists and engineers must be introduced to and kept up to date with Large Scale Scientific Computing as the power of supercomputer increases and the nature of methods change.

One of the tasks in the software area will be to invent methods so that discipline or project oriented scientists have access to the latest and best numerical methods as well as to the supercomputers. And history shows that some important improvements in numerical methods and algorithms are made by project scientists using Large Scale Scientific Computing. Thus, an important aspect of this program is to bridge between the development and use of new ideas.

A PROGRAM FOR DEVELOPMENT OF VERY HIGH PERFORMANCE COMPUTER SYSTEMS

REPORT FROM

GROUP 3
OF
THE
LAX COMMITTEE

SEPTEMBER 2, 1982

TABLE OF CONTENTS

A PROGRAM FOR DEVELOPMENT OF VERY HIGH PERFORMANCE COMPUTER SYSTEMS

1.Ø EXECUTIVE SUMMARY

The premise upon which this report is based is that continued leadership in very high performance computing and its applications are crucial to the United States both for national security and for economic development. The rapid and continuing advance of microelectronic technology is opening a path to the development of a new generation of supercomputers which can potentially obtain computation rates two or three orders of magnitude faster than today's most powerful machines. Development and application of these very high performance computers is crucial for both national security and long-term economic development.

Research has established that future generations of very high performance computer systems will have parallel architectures radically different from the conceptually sequential architectures of today's supercomputers. There are many candidate architectural proposals which must be evaluated before commercial feasibility can be established. There is, at present, little likelihood that this research will be developed and exploited in this country given the length and expense of the development cycle for a radical new computer architecture and the numeric size of the marketplace. Immediate action on these opportunities is essential if U.S. leadership in very high performance computing is to continue.

The program presented herein has as its first goal acceleration of the development of current research concepts for very high performance computer architectures into commercial products. The goal can be accomplished by a research and development program which brings very high performance computer technology to a state where the normal venture capital mechanism will select it for commercial development.

This program is required because the unfavorable short-term risk-to-return ratio for development of a radically architectured class of supercomputers totally inhibits operation of the normal venture capital mechanism. Very high performance computing is a case where maximizing short-term return on capital does not reflect the national security and long term national economic interests.

The program is founded on a development model based on the concept of university/industry/government laboratory collaboration. This development model is appropriate to the U.S. economic system and culture and may be a procedural prototype for acceleration of application of research in other high technology areas with unfavorable short-term risk/return ratios but where long-term national significance development is concerned.

The Very High Speed Integrated Circuit (VHSIC) program, sponsored by the Department of Defense, is developing the component technology base for high performance parallel computing. The application for which the VHSIC technology is intended will require very high performance parallel architectures in order to meet their objectives. This program focuses on accelerated development of the architectures and systems which will utilize the VHSIC device technology and is thus a natural successor of the VHSIC program. The absence of such a program may lead to the Japanese being the principal beneficiaries from commercial application of the enhancement of component technology driven by the VHSIC program. It should be noted that the VHSIC program does not address several component technology issues crucial to development of general purpose very high performance computer systems.

2.0 PROGRAM JUSTIFICATION

The current applications of super computers include

- design and simulation of VLSI chips
- design of nuclear weapons
- design and analysis of nuclear reactors
- fusion energy research
- directed energy weapons
- intelligence applications
- aerodynamics, structural design and evaluation
- structural mechanics
- electric power distribution
- atmospheric science
- oceanography
- fluid dynamics
- automobile design
- design of manufacturing systems
- geophysics
- petroleum exploration and reservoir management

Potential but not yet exploitable applications include real time image processing and fast robotics. The list of potential applications does not include artificial intelligence oriented applications such as expert and/or knowledge based systems. These problem domains may require very high performance parallel architectures for effective application. The software algorithms and user interfaces are so different from the essentially numerical applications as to justify separate consideration. (See Appendices B and H for additional information.)

Electronics, nuclear weapons, intelligence, aerodynamics and energy production are currently major components of U.S. defense. Progress in each of these areas will be paced, in fact may be bounded, by progress in high performance computing. The same is true for most of the other applications and potential applications. Thus, leadership in supercomputers is fundamental to U.S. defense and to U.S. leadership in crucial areas of technology. If another country should assume leadership in supercomputers, U.S. defense and technology will depend upon access to computers of foreign manufacture. This presents three risks.

1. Currency of Access. If other countries consume the first two years of production then U.S. scientists and engineers will be denied use of these machines in U.S. technology and weapon system design and defense application could lag theirs by that amount. (Hitachi has recently announced a very

high performance array processor which will _not_ be marketed in this country.)

2. _Denial of Access_. In this worst case U.S. technology development could be handicapped until domestic sources are developed.

3. _Computing Technology Lag_. Development of supercomputers has always driven the development of other computer systems and been an important driver of electronics technology development. If another country assumes leadership we may lose these benefits.

These risks are sufficiently crucial to both national security and national economic development to be unacceptable.

The United States currently leads in the development of supercomputers and will be expected to do so for the next two to three years. Continuing American leadership in this area is threatened by two factors. One factor is the unfavorable short-term risk/return ratios which attach to the major innovations required to the development of new supercomputers and block the normal venture capital path for the development of research into commercial products. The factors which yield the unfavorable risk/return ratio include

1. There are many alternative courses for development of parallel architectures which must be evaluated.

2. Parallel architecture will require an entirely new generation of software and peripherals.

3. The long (5-8 year) development cycle

4. The large capital requirement

5. The existence in the computer system marketplace of many low risk, short term, high return opportunities

6. The relatively small market. There are now approximately fifty current generation "super computers" now installed.

Short-term risk/return assessment for product development does not reflect long term national economic and security interests.

The second factor is that the Japanese government and computer industry, having recognized the crucial nature and economic significance of the development of supercomputers, are very strongly committed to a research and development program in this area. The Japanese are now engineering for commercial exploitation machines based upon the research done in this country in the 1970's. This last statement is particularly significant. It points to the crucial element which must be supplied to continue U.S. leadership. That is, commercial exploitation of research concepts whose development cycle is too long and complex for operation of the venture capital avenue which has been so dramatically successful with short term pay-out concepts. The program proposed has as its goal lowering of market risk and pay-out cycle to where the normal venture capital mechanism will operate effectively.

3.Ø DESCRIPTION OF PROPOSED PROGRAM

This section outlines a program designed to lower the market risk and development costs of a new generation of very high performance supercomputers to a level where development and marketing by private computer vendors will become not only feasible but attractive. The program mobilizes the present strength of U.S. research groups, government laboratories and computer vendors. The program has three major focuses. The key concept for accelerating development is to combine the creativity of the university research programs, the engineering expertise of computer vendors and the practical concerns of users and potential users of very high performance computers. It is also important to note that there are now available limited capability systems which will support concept evaluation and application development for parallel systems. The functions of the program will be to:

1. accelerate the development of current and future generations of parallel architectures to be the bases for new very high performance computers,

2. establish a knowledge base for application of parallel computing concepts to problems of national security and economic interests and

3. broaden and enlarge the base of research activities to ensure the continuation of the flow of innovative architectures in very high performance computing concepts while at the same time diverting a significant fraction of the research community in very high performance computing to development.

The steps in the development and application of a major new computer architecture are shown in Figure 1. The proposed program has as its major thrust acceleration of the product development and application cycle represented by steps 2, 3 and 4 of Figure 1. It also specifies strengthening basic research in the area of very high performance computing.

```
+---------->---------+
|                    |
|                    V
|              Basic Research in      (1)
|       High Performance Computing
|                    |
+--------+           |
|        |           |
|        V           V
|      scalable, feasibility          (2)
|      demonstrations
|                    |
+---------+          |
|         |          |
|         V          V
|      commercial prototyping         (3)
|                    |
+----------+         |
|          |         |
|          V         V
|       System Application and        (4)
|       Evaluation   |
|                    |
+-----------<--------+
```

Figure 1: Stages in Development of Very High Performance Computer Systems

There are four activities in the proposed program, supporting all of the development phases of Figure 1. These activities are:

a. <u>Accelerated Development of Scalable Feasibility Demonstrations</u>.

Proposals for exploratory development of significant research architectures will be received and considered for funding as accelerated development projects. It is anticipated that there will be formed university/industry/user consortia to develop scalable feasibility demonstrations of the most promising of the architectural concepts established by current research. The scalable feasibility demonstration must include development of significant applications to demonstrate the capabilities of the system.

b. <u>Commercial Prototyping</u>.

Projects whose scalable feasibility demonstrations establish significant commercial potential will become candidates for construction of a full scale commercial prototype. This will involve a supported development effort with focus shifting to industry and which is expected to lead to a product that industry will subsequently carry forward on a commercial basis.

c. <u>Evaluation and Application</u>.

There will be established application and evaluation laboratories based upon the existing products which provide prototype products that illustrate the concepts of parallel computing.

d. Basic Research.

The funding for basic research in very high performance computing will be enhanced in order to continue the flow of new research architectures and concepts. A crucial component of basic research which requires immediate emphasis is development of algorithms for effective utilization of parallel architectures.

This development program may serve as a model for shortening the research to product cycle in other high technology/high risk areas which have poor short-term risk/return ratios, but which have long term national interest significance. The panel believes that the recommended program will, within the context of the U.S. economic system, dramatically shorten the duration between research and product in the critical area of very high performance computer systems.

All of the candidate projects for accelerated development exhibit some form of parallel architecture. The projects selected for accelerated development must include not only hardware architecture development but also research and development on algorithms, software systems, languages and the peripheral and supporting facilities which are necessary to make supercomputers usable.

The accelerated development proposals may focus on general purpose systems with very high performance or on special purpose systems which have broad application across a spectrum of disciplines such as real time image processing.

4.0 ORGANIZATIONAL AND ADMINISTRATIVE STRUCTURE

The panel proposes an administrative structure consisting of three groups: one to set policy, one to coordinate and administer the program and a technical advisory body to work with the coordination and review groups. Figure 2 shows the relationship of the proposed groups one to another as well as their relationship to the federal agencies which will make ultimate funding decisions concerning projects to be supported.

```
                 +---------------------+
                 | Computational       |
                 | Science Policy      |------------+
                 | Board               |            |
                 +---------------------+            |
                   |                                |
                   |                                |
       +---------------------+            +-----------------+
       | Interagency         |------------| Technical       |
       | Working Group       |            | Board           |
       +---------------------+            +-----------------+
          /      |      |      \
         /       |      |       \
      DOD       DOE   NASA      NSF
```

Figure 2

The Computational Science Policy Board (CSPB) will oversee program and policy issues. The board will review resource allocations among the agencies for consistency with the overall policy.

The Inter-Agency Working Group (IWG) will allocate projects and budgetary resources among the agencies. The membership of the group will come from the participating agencies and will also include the Chair of the policy board. The IWG will coordinate budget allocations, requirements and plans among the agencies.

The Technical Board (TB) will assist the CSPB and IWG in technical matters. It will not have direct budgetary responsibility. It will, however, be used to assist in evaluation of proposals and advise on distribution of agency resources for project execution.

It is anticipated that up to five accelerated development projects may be active at any time after the first year of the program. The average cost of each development project will not exceed five million dollars per year. The first commercial prototype development will be initiated in year three or four of the program. Each prototype construction will take three to five years. No more than two prototypes should be under development at any one time.

The budget given in Table 1 provides an additional ten million dollars to fund evaluation of current and future prototypes, to begin development of parallel formulations of applications of national security interest and for additional funding for research in VHP computing.

	Basic Research and System Evaluation	Scalable Feasibility Demonstrations	Commercial Prototypes	Total
1984	10	15	-	25
1985	12	20	-	32
1986	14	25	5	44
1987	15	25	15	55
1988	15	25	30	70

Table 1: (figures are millions of dollars)

The basis for selection on program activities a, c and d will be unsolicited proposals. It is anticipated that the accelerated development proposals will originate with the university/industry/user consortia and prototype construction proposals from vendors. Proposals of any structure will, however, be considered on merit.

APPENDIX A

Statement of Problems

The technological objective for long-range supercomputer development, we believe, should be the availability by 1990 of computers 1000 times more powerful than the current generation of supercomputers. There are serious problems that must be solved to reach this objective, but the perception of the research community is that solutions to the problems can be attained within this time frame. The difficulties of designing systems of such extraordinary power can be broken down into a matrix which is shown in Figure A-1. Going across the three principal parts of a computer system are listed. The "engine" is the CPU and memory; the "system" is the set of high speed peripherals (disks and tapes) directly attached to the engine, and the "satellites" consist of graphics, communications and slow-speed peripherals.

Reading down, the basic issues of computer development and use are listed: software, components, architecture and algorithms. "Components" is understood to include all aspects of hardware such as chip design, packaging and cooling, while architecture refers to global aspects of the computer design, such as vector vs. parallel organization. We have listed the degree of difficulty faced for each entry of the matrix. "Critical need" means there are problems that must be solved to reach the objectives. "Marginally acceptable" means that it should be possible to live with normal technological development but there are problems that ought to be attacked. "Satisfactory" means we anticipate no serious problems.

The really crucial software issues of language, programming environments, etc. are subsumed under the entry for "engine".

	ENGINE	SYSTEM	SATELLITES
Software	critical need	critical need	satisfactory
Components	critical need, device aspects being considered by VHSIC, attention to packaging and cooling is required	critical need for a majority of problems	some development needed
Architecture	critical need	critical need for some applications	satisfactory
Algorithms	critical need	critical need for some problems	critical need for some problems

Figure A-1

APPENDIX B

The Japanese Supercomputer Program

The Japanese government and computer industry have noted the tremendous opportunities and leverage potentially available through development of such computers, and have established two major projects in this area: a scientifically oriented "supercomputer" project and a "fifth generation computer" project with a strong artificial intelligence orientation. These projects are coordinated and funded by the Japanese government industrial agency MITI, and involve all of the major Japanese computer companies.

An extended visit to Japan was recently made by visitors from Los Alamos Scientific Laboratories and Lawrence Livermore National Laboratories. We quote a report of these trips to illustrate the commitment of the Japanese government and computer industry. "The supercomputer project calls for the development of supercomputers for scientific and technological applications... Funds cover software and hardware development... The supercomputer association (established to manage this project) has a contract with the Japanese government. Its six member companies are Hitachi, Fujitsu, Mitsubishi, NEC and Oki Electric... The purpose of this project is to 'develop an ultra high speed computing facility for scientific and technological applications that is more than one thousand times faster than "high-end" conventional machines.'"

MITI has also established a "fifth generation computer" project which is controlled by another association whose membership comprises the same six Japanese computer companies plus others. It is important to note that the "fifth generation" project is an entirely distinct effort from the "supercomputer" project. The fifth generation project was formally initiated in April 1982 and will run approximately ten years.

More immediate signs of the speed of Japanese development come from the very recent Hitachi and Fujitsu announcements of machines which are expected to compete directly with the yet to be announced Cray 2 computer which, when it arrives, will be the fastest U.S. machine and the announcement by Hitachi of a 640 MFLOP array processor.

APPENDIX C

Candidate Projects for Accelerated Development

Several architecture candidates for accelerated development can be identified from among the recent research efforts into parallel computation. The following list is representative rather than exhaustive. The entries represent varying levels of maturity and differing degrees of completeness with respect to the spectrum of hardware/software and application problems.

Blue CHiP Project - Purdue

Data Flow Project - MIT

HEP2 - Denelcor

High Speed Multiprocessor Project - Illinois

Homogeneous Machine Project - Cal Tech

Los Alamos PUPS Machine

PASM Project - Purdue EE

Systolic Arrays Project - CMU

Texas Reconfigurable Array Computer (TRAC) - Texas

Ultra Computer Project - NYU

In addition, there is at least one candidate for an applications laboratory -- the Purdue Center for Parallel and Vector Computing.

The Blue CHiP Project, headed by Lawrence Snyder of Purdue (CS), has as its focus the Configurable, Highly Parallel (CHiP) Computer. The CHiP architectures are composed of a lattice of homogeneous computers each with local program and data memory that are connected together by programmable switches. By programming the switches the physical machine is configured to match the topology of algorithms. A preprototype CHiP computer is being built, a programming language and environment have been implemented, and the feasibility of wafer scale VLSI implementation has been worked out.

Two projects to develop data flow computer architecture are in progress at MIT. In a data flow computer instructions are activated by data instead of an incrementea program counter. The static architecture for a data flow supercomputer is being developed under Professor Jack Dennis. Machine level code structures for several large application codes have been constructed by hand to prove performance potential, and an engineering model is in operation with eight processing units. The next phase requires development of custom LSI devices and an optimizing compiler for the VAL programming language.

The tagged-token data flow architecture, being developed by MIT Professor Arvind, is a form of data flow multiprocessor using colored tokens to distinguish values associated with different procedure invocations. This architecture is ready for detailed simulation to determine good choices of architectural parameters and to evaluate program behavior and potential performance for realistic application codes.

HEP 1 is a commercially available, scalar MIMD computer system. HEP 2 is an upward compatible enhancement of the HEP 1 architecture which will offer both increased speed and increased parallelism. HEP 2 will be very appropriate for general purpose scientific computing, and will be especially cost effective for those applications which exhibit parallelism of a non-vector kind. The logic and packaging technologies used in HEP 2 will be state-of-the-art in all respects. The projected completion date for HEP 2 is early 1986.

High Speed Multiprocessor Project -- A University of Illinois group has been engaged for the past ten years in automatically restructuring ordinary Fortran programs for high speed machines. Currently their system can do this effectively for parallel, pipeline and multiprocessor systems. These results have been demonstrated on a number of existing commercial machines. As a consequence of this work they are now designing a multiprocessor architecture that is aimed at providing high speed processing over a wide spectrum of applications. The design is based on measurements of over 1000 application programs and as such is an example of a project that has developed an optimizing compiler hand in hand with a high speed architecture.

The Homogeneous Machine Project, headed by Charles Seitz of Cal Tech, is focused around an array of processors for solving high energy physics codes. The processors are connected into a binary n-cube. A prototype built with

Intel 8086's, each with 65K of memory, is under construction.

The Los Alamos Parallel Microprocessor Systems (Pups) is designed to accommodate up to sixteen computational processors and two communication processors. It can be configured as either a shared memory machine, a distributed machine, or as a collection of clusters of shared memory machines. It will support Fortran, have floating point, and a relatively large memory.

The PASM Project, headed by H. J. Seigel of Purdue (EE), has as its focus the Partitionable Array SIMD/MIMD Computer. The PASM machine is composed of processing elements and common memory that can operate in SIMD mode or MIMD mode. The architecture is motivated by image processing tasks, and a design for an MC68000 based prototype has been completed.

Systolic Array Project -- These are computing structures which attain extreme efficiency by making use of designs in which data moves in very regular fashion through a sequence of processing nodes laid out in a dense pattern on the two-dimensional surface of a VLSI chip. Designs which keep all data paths short and process many items of data in parallel allow remarkable processing rates to be attained. Designs of this kind have been proposed by groups at Carnegie-Mellon University and elsewhere, and it is already clear that they can be very important for such significant special applications as signal processing and manipulation of matrices having a favorable band structure. The ability of this approach to produce "miracle chips" of great importance for new application areas is still far from exhausted, so that rigorous development of the systolic approach is appropriate.

Texas Reconfigurable Array Computer (TRAC) -- The TRAC system is based on coupling processors, memories and I/O's through a dynamically reconfigurable banyan network. TRAC implements multiple models of parallel computing. The TRAC project integrates consideration of hardware, software and applications of parallel computing. The TRAC architecture can be scaled up in number of processors and memory elements at a cost growth rate of n log2 n. TRAC is being developed as a laboratory for parallel computing. A 4 processor - 9 memory configuration of TRAC is now operating and a Pascal compiler for programming of applications is available.

The New York University Ultracomputer -- This is a highly parallel MIMD machine which aims to combine hundreds or thousands of small, relatively conventional processing elements, all communicating with a large shared memory, to attain very high performance. In the present concept for this machine, processors communicate with memory through a very high bandwidth switching network which executes a few operations vital to ultraparallel inter-process synchronization in addition to its basic data-routing function. Various advantageous operating systems software structures have been worked out for this machine, as has a fairly detailed design for the switching chip control for its communication network concept. Various scientific application simulations carried out by the NYU group show that this machine can be programmed using relatively conventional techniques (essentially in a slightly extended version of the widely used Fortran programming language) to attain high processor utilization.

Purdue Center for Parallel and Vector Computing -- The purpose of the center at Purdue University will be to advance the state of the art in the use of parallel and vector computers by engaging in research and development in areas such as the design and implementation of algorithms, performance analysis, modern language design and related computer development, appropriate software tools and the design and development of software modules. This work would be driven by specific application areas arising from the simulation of physical systems and would be accomplished by small multi-disciplinary teams consisting of Purdue staff and visitors from other universities, private industry and government laboratories.

Crystal Project - University of Wisconsin -- The University of Wisconsin Crystal project, headed by D. Dewitt, has been funded by the NSF CER program to design and construct a multicomputer with 50 to 100 nodes (processors). The processors are to be interconnected using broadband, frequency agile local network interfaces. Each processor will be a high performance 32 bit computer with approximately 1 megabyte of memory and floating point hardware. The total communications bandwidth is expected to be approximately 100 Mbits/sec. The multichannel capabilities of the frequency agile interfaces, along with the Crystal support software, provide researchers a number of unique capabilities. First, the multicomputer can be divided into multiple partitions enabling researchers to share the facility in a manner analogous to a timesharing machine. In addition, the processors within a partition can utilize the frequency agile interfaces to efficiently emulate a number of interconnection topologies. This will permit different groups of researchers to use the interconnection topology that is best suited for their application. Applications of interest include distributed operating systems, programming languages for distributed systems, tools for debugging distributed systems, multiprocessor database machines, and evaluation of alternative protocols for high performance local network communications. The system will support experimentation with parallel algorithms for solving computation intensive problems in the areas of mathematical programming, numerical analysis and computer vision.

APPENDIX D

Applications and Benchmarks

The list of applications given here are initial subjects for application development studies on the currently available parallel architectures.

Monte Carlo techniques for simulating fusion processes. Specific programs in both time dependent and time independent models exist.

Two- and three-dimensional hydrodynamic calculations in weapons design.

Specific model computations include

Particle-in-Cell for plasma models

Adaptive Mesh Refinement for fluid processing

These represent two algorithmic extremes and also the most current thinking that covers both the physical model and how to fit onto parallel processors.

Another class of computations is the Many Body problem used to study atomic and molecular interactions.

Real time image processing is an integral component of many potential military applications.

Two- and three-dimensional aerodynamics calculations for aircraft-like bodies. Techniques should include both explicit and implicit methods; the former is known to vectorize well, the latter has theoretical advantages but does not lend itself to parallelism as readily.

Petroleum exploration and reservoir management are applications with very high potential economic payoffs. Reservoir management applications include elements of both fluid dynamics and heat transfer. Exploration studies deal with signal processing techniques.

In all these examples the need is to encourage early interaction in order to influence future computer architectures.

APPENDIX E

Relationship to VHSIC

The VHSIC program is focused on the development of very high performance device technology. The ultimate goal of this technology is to support military applications which require very high performance parallel computer systems. The purpose of the program proposed here is to accelerate the development of the architectures, systems and algorithms necessary to effectively apply this VHSIC technology.

The VHSIC developed technology will begin to arrive next year. It is desirable to have in place a program of evaluation and utilization of this technology in appropriate architectures in order to provide feedback in the later stages of the VHSIC program. The program proposed in this report is a natural follow on to VHSIC. It has a broader spectrum of applications than only direct military applications such as real-time image processing. It extends to other crucial national security areas such as weapons development and energy production. It also includes economically critical applications such as weather modeling, geophysics, computer aided design and high performance robotics.

The absence of such a program as is proposed here may lead to the Japanese being a principal beneficiary of the enhancement of high performance components generated by the VHSIC program through their effective system development of products which require a substantial engineering investment and have a high market risk.

APPENDIX F

Relationship to Industry Programs

The semiconductor industry through the Semiconductor Research Association (SRA) and the computer industry through the Microelectronics and Computer Technology Corporation (MCC) have begun to establish collaborative programs to support research in universities and to accelerate product development from university research programs in both the semiconductor and computer system fields. These initiatives by the computer vendors and semiconductor manufacturers strongly suggest that the role model projected here for cooperation between consortia of users, vendors and university research projects is one whose time has come and which will be accepted by industry as an effective means of accelerating product development in difficult and/or high risk areas.

It is also possible that partial funding support can be obtained for these projects from these industrial consortia.

APPENDIX G

Feasibility of Target Attainment

It is the best reasoned judgement of computer architects, microelectronics researchers and software experts that the goal of a system approximately a thousand times faster than today's fastest computers in the early 1990 time frame is viable. The requirements are a) component technology, b) architectural and organizational structure, c) the software systems, and d) algorithms to exploit the architecture for significant applications.

The VHSIC program is accelerating the development of very high performance component technology. There is need for a strong program in the development of packaging and cooling technologies in order to be able to exploit integrated circuit chips in very high performance computer systems, whether they be dedicated and special purpose systems or general purpose systems.

A number of architectural concepts for exploiting parallelism which have promise of delivering very high performance computing systems have been defined and evaluated. These architectures will clearly be capable of delivering the required performance levels if they can be implemented in appropriate technology, be supported by appropriate peripherals, have the appropriate software available, and have the algorithms for application established.

The necessary peripherals such as large scale primary memories, high performance secondary memories, interconnection networks, etc. are the least developed of the architectural elements of a full system. It is clear that development paths do exist for these system elements.

The development of basic concepts of parallel computing and the necessary software systems to support exploitation of these architectures is proceeding along with the development of the architectural concepts. A great deal of work is needed in this area, particularly with respect to applicability.

The final requirement is the development of applications which are formulated in parallel concepts and which can exploit the very high performance parallel architectures. A very great deal of work is needed here. However, the initial work which has been done has been amazingly successful.

In summary, in each element of the requirements for very high performance computer systems, there is substantial reason to believe that a sustained research and development program supported by effective engineering can approach the goal of 1000 times today's supercomputers.

APPENDIX H

Numeric and Non-Numeric Computing: Relationship and Status

Numeric and non-numeric computing have traditionally been regarded as different discipline areas. The two disciplines use different software, interfaces and algorithms even when sharing hardware environments. Numeric computing applications have traditionally driven the development of very high performance computer systems. This is because the problems to be solved are relatively well understood and thus attention could be focused on problem solution, while in the non-numeric areas (all basically some form of automation of the reasoning process), relatively few problems have been sufficiently well understood to justify large scale application. Those areas which have been systematically approached such as theorem proving and program verification have been found to require enormous computing resources for substantial applications. Thus practical application to non-numeric problems will probably require very high performance computer systems.

The Japanese government and computer industry have recognized this state of affairs with the Fifth Generation Computer Project which aims at producing "supercomputer systems" for artificial intelligence applications.

The "Fifth Generation" and "Supercomputer" projects of MITI preserve the traditional separation of numeric and non-numeric computing. We believe that there are major overlaps in technology for both areas. Development of the parallel generation of "supercomputers" will have a major accelerating effect on high performance systems for non-numeric applications.

We further believe that there is a strong need for coupling automated reasoning with mathematical modeling and data base techniques to produce effective control processes for complex systems and for effective modeling of complex systems involving human and mechanical components.

It should be noted that the current generation of non-numeric applications are being developed on systems of power approximately one to two orders of magnitude below the current generation of supercomputers. There is need to consider scaling of non-numeric computing to the current generation of supercomputers. It may be desirable to organize a study panel to determine the national security and economic development impacts of non-numeric computing and the probable effectiveness of the entrepreneurial capital system of meeting these national security requirements as the research programs bring the concepts and applications to fruition. The Japanese government and computer industry have already established their commitment to economic importance with the Fifth Generation project. It is worthy of note that there are now being marketed small scale systems specialized for non-numeric applications.

A P P E N D I X   I

INFORMATION ON SELECTED FACILITIES

Partial Inventory and Announced Orders of Class VI Machines

Country	Site	Number	Purpose	Computer
U.S.	Los Alamos Nat. Lab.	5	Weapons Research	Cray-1
	Lawrence Livermore Nat. Lab.	4	Weapons Research	Cray-1
		2	Magnetic Fusion Energy Research	Cray-1
	Sandia Nat. Lab.	2	Weapons Research	Cray-1
	KAPL	1	Reactor Research	Cyber 205
	Bettis	1	Reactor Research	Cyber 205
	Kirtland Air Force Base	1	Military	Cray-1
	National Center for Atmospheric Research	1	Atmospheric Science	Cray-1
	NSA	2	Intelligence	Cray-1
	NASA-Ames	1	Aerodynamics	Cray-1
	NASA-Goddard	1	Atmospheric Science	Cyber 205
	NASA-Lewis	1	Fluid Dynamics	Cray-1
	FNOC-Monterey	1	Oceanography	Cyber 205
	National Environmental Satellite Service (NOAA)	1	Research	Cyber 205
	Colorado State Univ.	1	Engineering Research	Cyber 205
	Univ. of Minnesota	1	Research	Cray-1
	Geophysical Fluid Dynamics Laboratory	2	Geophysics	Cyber 205
	Purdue University	1	Research	Cyber 205
	Univ. of Georgia	1	Research	Denelcor HEP
	CHEVRON	1	Petroleum	Cray-1
	Bell	1	Research	Cray-1
	ARCO	1	Petroleum Engineering	Cray-1
	EXXON	1	Petroleum Engineering	Cray-1
	Grumman Corp.	1	Jet Engine Simulation	Cray-1
	Westinghouse Corp.	1	Nuclear Power Plant Design	Cray-1
	TEXACO	1	Petroleum Engineering	Cyber 205
	SOHIO	1	Petroleum Engineering	Cyber 205
	Digital Production, Inc.	1	Graphics	Cray-1
	Boeing Computing Serv.	1	Timesharing	Cray-1
	Control Data Corp.	1	Timesharing	Cyber 205
	United Information Serv.	1	Timesharing	Cray-1
Germany	Max Planck	1	Research	Cray-1
	Bochum	1	Research	Cyber 205
	PRAKLA	1	Research	Cyber 205
	Univ. of Karlsruhe	1	Research	Cyber 205
	Univ. of Stuttgart	1	Research	Cray-1
	Deutch Forschungs und Versuchsanstalt fur Luft Raumfahrt	1	Aerospace Research	Cray-1

Partial Inventory and Announced Orders of Class VI Machines (Continued)

Country	Site	Number	Purpose	Computer
France	GETIA	1	Electric Power Institute	Cray-1
	Commissariat A'Lenergie Atomique	1	Nuclear Energy	Cray-1
	Compagnie International De Services En Informatique	1	Timesharing	Cray-1
	Ecole Polytechnique	1	Research	Cray-1
England	European Centre for Medium Range Forecasting	1	Weather	Cray-1
	Brit Met	1	Weather	Cyber 205
	Daresbury	1	Physics Research	Cray-1
	AWRE Harwell	1	Nuclear Energy, Weapons	Cray-1
	Shell Oil, U.K.	1	Petroleum	Cray-1
	Univ. of London	1	Research	Cray-1
	Univ. of Manchester	1	Research	Cyber 205
Japan	Mitsubishi	1	Research	Cray-1
	Century Research	1	Research	Cray-1

From W. F. Ballhaus, Jr.
NASA/Ames Research Center

August 13, 1982

For the past decade, NASA-Ames has conducted pioneering research in the rapidly advancing areas of computational fluid dynamics and computational aerodynamics as well as other aerospace disciplines. This research, conducted both on site and remotely at university and industrial sites, has been serviced by a continually expanding computational capability. Currently this capability includes access to a Class VI computer housed onsite. NASA has plans to further augment its computational capability by means of the Numerical Aerodynamic Simulator (NAS), a project to be conducted by the Ames Research Center. This facility will provide by 1988 a computation rate of a billion floating-point operations per second (sustained) with a memory sufficiently large to support this speed. The facility will provide interactive access, both remote and on-site, to a large number of users from NASA, DOD, academia, and industry. It will be used to solve previously intractable problems of national concern in the aerospace disciplines of interest to the DOD and NASA. Limited operational capability is expected by 1985 with full service in 1988. The NAS facility will include graphics and work stations, satellite and telecommunications interfaces for remote access, a large data base mass store, and a fast network linking these elements with a high-speed computing engine. This year a new computer science effort has been initiated at Ames to seek innovative ways to apply advanced computational concepts to the solution of the Agency's technical problems. This activity will be complemented by a new institute at Ames under the auspices of the University Space Research Association. The Research Institute for Applications of Computer Science will be operating by early 1983.

NCAR COMPUTING CAPABILITIES AND SERVICES *

Walter Macintyre

The missions of NCAR as set by the UCAR Board of Trustees and endorsed by the National Science Foundation, are as follows:

- In cooperation with university research groups and other organizations, to identify, develop and make accessible selected major research services and facilities of the outstanding quality required by the universities and NCAR for effective progress in atmospheric research programs. NCAR will be responsible to assure the most effective use of these facilities and services by scientists in the universities and NCAR.

- In cooperation with universities and other organizations, to plan and carry out research programs of highest quality on selected scientific problems of great national and international importance and scope...It is appropriate that most of the research at NCAR be on problems that are characterized by their central importance to society, scientific interest, and by the requirement for large-scale, coordinated thrusts by teams of scientists from a number of institutions....

The overall pace of progress in the atmospheric and related sciences is limited and modulated by the nature and power of available computers. Hence both of NCAR's missions require the Center to make available to the university-NCAR community computing capabilities that represent the current state of the art in speed, capacity, architecture, and software, as well as a wide variety of effective services.

NCAR established its computing facility two decades ago. Throughout its history, NCAR has pioneered in making advanced hardware, software and services readily accessible to the university-based and NCAR research community. The NCAR Scientific Computing Division is widely acknowledged as a leader in the development of new capabilites for the university atmospheric sciences community. During the past year, more than 900 individuals from 80 institutions used the facility, and scientists at 75 locations had access via the NCAR Remote Job Entry System. In an analysis of existing large-scale operations for atmospheric research modeling, done by the National Advisory Committee on Oceans and Atmosphere for the President's Science Advisor (see Table 1), NCAR is shown to be the only facility fully open to the overall university community on a first-priority basis.

History of NCAR Computing Capacity

Table 2 presents a history of NCAR computing system milestones since 1963, and Figure 1 shows the growth of NCAR's computing capability over the years in terms of units of computer resource delivered to users. The acquisition of successive generations of computers has made

* This is a chapter in the December 1982 NCAR document entitled "Scientific Justification for An Advanced Vector Computer."

Table 1

Summary of existing large-scale operations in support of atmospheric modeling research, fiscal year 1980*

Agency/Location	Computers Type	Users	Percentage use for atmospheric research Percent
Department of Commerce			
NOAA			
GFDL, Princeton University	TI/ASC	in-house	70-80
Environmental Research Laboratories, Boulder, Colo.	CDC Cyber 170/750	in-house	50-60
Suitland, Maryland	(3) IBM 360/195	in-house	15
Department of Defense			
AFGL, Bedford, Massachusetts	CDC 6600	in-house	n/a
U.S. Army, White Sands	U 1180	in-house	n/a
U.S. Navy, (FNOC), Monterey, California	CDC 6500, Cyber 170/175, Cyber 170/720, Cyber 203	in-house	10
NASA			
GISS, New York, N.Y.	IBM 360/195	in-house	75
GSFC, Greenbelt, Maryland	Amdahl	in-house 80% universities 20%	80
	(2) IBM 360/195	in-house 90%	20
Hampton, Virginia	CDC Cyber 203, Cyber 170/173, (2) Cyber 173 (2) Cyber 175 (2) 6600	in-house 95% other 5%	20
Pasadena, California	(3) U 1108	in-house	1.5
NSF			
NCAR, Boulder, Colo.	Cray CDC 7600	NCAR & universities	100
EPA			
Research Triangle Park, N.C.	(2) U 1100, IBM 360/165	in-house	15
DOE			
Argonne National Laboratory Argonne, Illinois	IBM 370/195, IBM 370/75, IBM 370/50, (2) 30/33	in-house	5
Brookhaven National Laboratory Upton, N.Y.	CDC 7600, (2) CDC 6600 DEQ PDP/10, Sigma 7	in-house	3
Battelle, Northwest Laboratory Richland, Wash.	Not available	in-house	2
Idaho Nat'l Engineering Lab. Idaho Falls, Idaho	IBM 360/75, Cyber 76	in-house	n/a
Los Alamos Scientific Lab. Los Alamos, N.M.	(2) CDC 6600, (4) CDC 7600, Cray 1, (2) Cyber 73	in-house	0.4
Lawrence Berkeley Laboratory Berkeley, California	CDC 7600, CDC 6600 CDC 6400	in-house	1
Lawrence Livermore Laboratory Livermore, California	Cray 1, (4) CDC 7600 (2) Star/100, CDC 6600	in-house	n/a
Oak Ridge National Laboratory Oak Ridge, Tennessee	IBM 360/75, PDP/10, SEL 8108	in-house	3
Sandia Laboratories Albuquerque, N.M.	U 1108.U1100-82, (3) CDC6600, CDC 6400, CDC 7600, Cyber 76, PDP10	in-house	0.5
Savannah River Laboratory Aiken, S.C.	IBM 360/115	in-house	0.3

* From: A Review of Atmospheric Science Research Facilities, National Advisory Committee on Oceans and Atmosphere, Washington, June, 1981, p. 20.

Table 2. Milestones of Major Computing Systems at NCAR

Year	Capability	Characteristics
Late 1963	CDC 3600	Memory = 32,700 words, 48 bits each Overall speed = .06 of CDC 7600
December 1965	CDC 6600	Memory = 65,000 words, 60 bits each Overall speed = 1.5 x CDC 3600 when new = 3.0 x CDC 3600 by 1968 (after software system changes) or = .2 of CDC 7600
July 1971	CDC 7600	Memory = 65,500 x 60 bits small core = 512,000 x 60 bits large core Basic cycle time = 27 ns (speed 10-15 MIPS) Disks = 2 x 5.072 billion bits at first Early 1976 = 8 x 2.4 billion bits on disks
June 1972	RJE Capability	Dialup from remote sites to Modcomp computer at NCAR Mesa Lab
February 1976	TBM Mass Store (used on-line Feb 1977)	Oct 78 = 9300 active TBM volumes Aug 79 = 15000 active TBM volumes Aug 80 = 27000 active TBM volumes (Average volume size 72 million bits) Apr 82 = 43170 active TBM volumes (Total data volume 7.6 trillion bits)
February 1978	CRAY-1A	Memory = 1,048,000 x 64 bits Basic cycle time = 12.5 ns (and pipeline) (Speed 40-80 MIPS) Overall speed = 4.5 x CDC 7600 Disks = 16 DD 19's x 2.4 billion bits each
March 1981	IBM 4341 front end computer	Used by SCD and other divisions for program job preparation. Also used for selected I/O tasks. Gives 6250 BPI tape capability for first time. Provides interactive access to the general user for the first time.
September 1981	NCAR internal network completed	Facilitates communication among the various hardware components.
October 1982	Connection of NCAR system to a common-carrier pocket switching network	Gives all users of NCAR computers inexpensive interactive access to the system.

Figure 1

History of Computing Output Units at NCAR

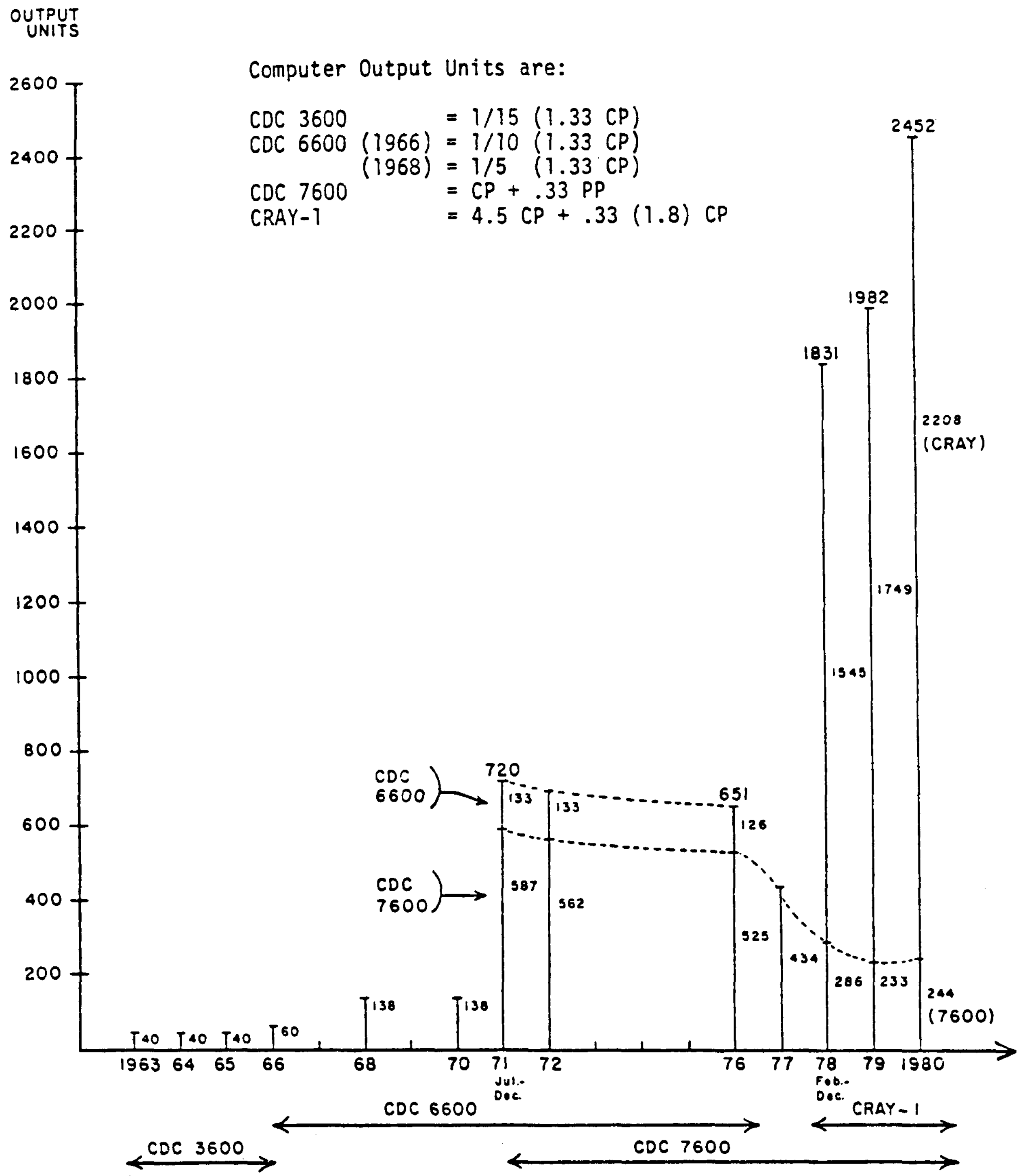

Figure 1. Computer output units used at NCAR. The calculation of power units is based on monthly hours of central processor (CP) time and peripheral (PP) time (on 7600), averaged over each year. When peripheral time is not available, a factor is applied to CP time to estimate I/O production. If a machine is not in production for the complete first or last year, the data shown is the average for the months used during the given calendar year.

it possible for the science to address qualitatively different, previously intractable problems in many areas of atmospheric research. Table 3 shows the distribution of total NCAR CRAY computing resources to the university-NCAR community by major disciplinary area during the period February 1978 to May 1982.

NCAR puts high importance on peripherals and systems to maximize the usefulness of the computer output and to minimize the time required of user scientists. It has the best developed and proven system for community participation in the process that governs the allocation of its resources NCAR's computing plans and priorities are set with active participation by the user community. In developing future plans for the facility, the assignment of first priority to the acquisition of additional main-frame power has been made with widespread community consultation and support.

Current Capabilities of the NCAR Scientific Computing Division (SCD)

1. Hardware capabilities. The major NCAR computing hardware capabilities are a CRAY-IA, capable of executing 80 million instructions per second for very high speed computations; a Control Data Corporation (CDC) 7600 for fast data analysis and file manipulation; an IBM 4341 computer system for providing interactive access to the major computing engines; an AMPEX Terabit Memory System (TBM), which provides mass storage for archived data as well as significant on-line data storage; and a remote job entry system that communicates with 75 locations across the country. The SCD also has a DICOMED graphics systems, a Network Systems Corporation (NSC) high-speed data network to connect the various systems, and a GANDALF port contention device which permits selective use access to the various machines.

2. Services. The services provided to the community by the NCAR SCD include the following:

a. Data Support. The Scientific Computing Division's Data Support Section maintains a large archive of computer-readable research data and provides assistance to users in locating data appropriate to their research needs, interfacing their programs with the data sets, and accessing utility routines for manipulation of the data. Users can access the data from remote terminals, in addition to using the data at NCAR or receiving tapes. This group has achieved a worldwide reputation in acquiring, formatting, updating, and making accessible atmospheric observational data sets.

The Data Support Section maintains many large sets of analyzed grid data and observed data from the National Meteorological Center, the National Climatic Center, the U.S. Navy, and the U.S. Air Force. Other countries and laboratories also provide data. Supporting data such as land elevation and ocean depth are included. The archives are largely describd in Data Sets for Meterological Research, by Jenne (NCAR-TN/IA-111, 1975).

b. User Services. The User Services Section offers consulting services to users, provides information on all services and operational procedures of the division, and provides software libraries of numerical and

Table 3

TOTAL CCU USE BY AREA OF INTEREST

February 1, 1978 through July 31, 1982

AREA OF INTEREST	CCU's	% TOTAL USE
1. Cloud Physics	1,225.7	12
2. Weather Prediction	2,147.2	21
3. Solar Physics	1,213.7	12
4. Chemistry and Upper Atmosphere	949.2	9
5. Climate	2,979.7	29
6. Oceanography	765.8	7
7. Basic Fluid Dynamics & Miscellaneous	1,092.4	10
Total:	10,373.7	100

utility tools. This section provides an effective interface to the hardware and operating system software for user scientists.

Within the User Services Section, the Communication group manages the current communication links for remote job entry (RJE) and interactive access. The Library Group manages and distributes the large numerical library (4700 routines) already available from SCD and develops new tools that facilitate program development and maintenance in a multi-machine environment. The User Interfaces Group provides documentation and consultation services and continuously improves methods of access to the facilities. The Multi-User Software Group is now completing a multi-year project for GENPRO2, a generalized software package for signal processing and data analysis.

c. Systems. The Systems Section develops network, communication and mass storage systems, as well as maintaining operating systems and their related software and language compilers supplied by vendors for several types of computers, ranging from small- to large-scale machines.

The highest priority is the maintenance of the integrity of the current operating systems, including the software running the internal NCAR network. The staff are involved in the development of interface software and non-vendor provided systems software as well as the provision of data management software. They assist with hardware and software planning and acquisitions.

d. Operations and Maintenance. The Operations and Maintenance Section operates and maintains the hardware systems of the facility and provides digital data library services, statistics on system use, and microfilm/microfiche and movie production.

e. Advanced Methods. The activities of the Advanced Methods Group include research, consulting and the production of advanced mathematical software in such areas as thermospheric physics, computational fluid dynamics and spherical vector harmonics.

3. Algorithm and Software Development. The NCAR computing staff and various users have contributed to the development of algorithms and software that are valuable in many areas of science. Examples are vectorized fast Fourier transform software, non-linear second-order partial differential equation algorithms and software, and the NCAR graphics package. This work continues and is effective disseminated throughout the scientific community.

Figure 2 gives the current organization and staff levels of the NCAR Scientific Computing Division.

Allocation Procedures

Allocations of NCAR computer resources are governed by the UCAR Board of Trustee policy, which requires that 45% of NCAR computing resources be available to university users, 45% to NCAR research and 10% to joint NCAR-university projects. Use of the computers by postdoctorals and other long-term visitors and processing of data from aircraft, radars, etc., is counted in the NCAR allocation, unless the investigator involved has applied for resources in the university category. This seldom occurs.

SCD ORGANIZATION CHART

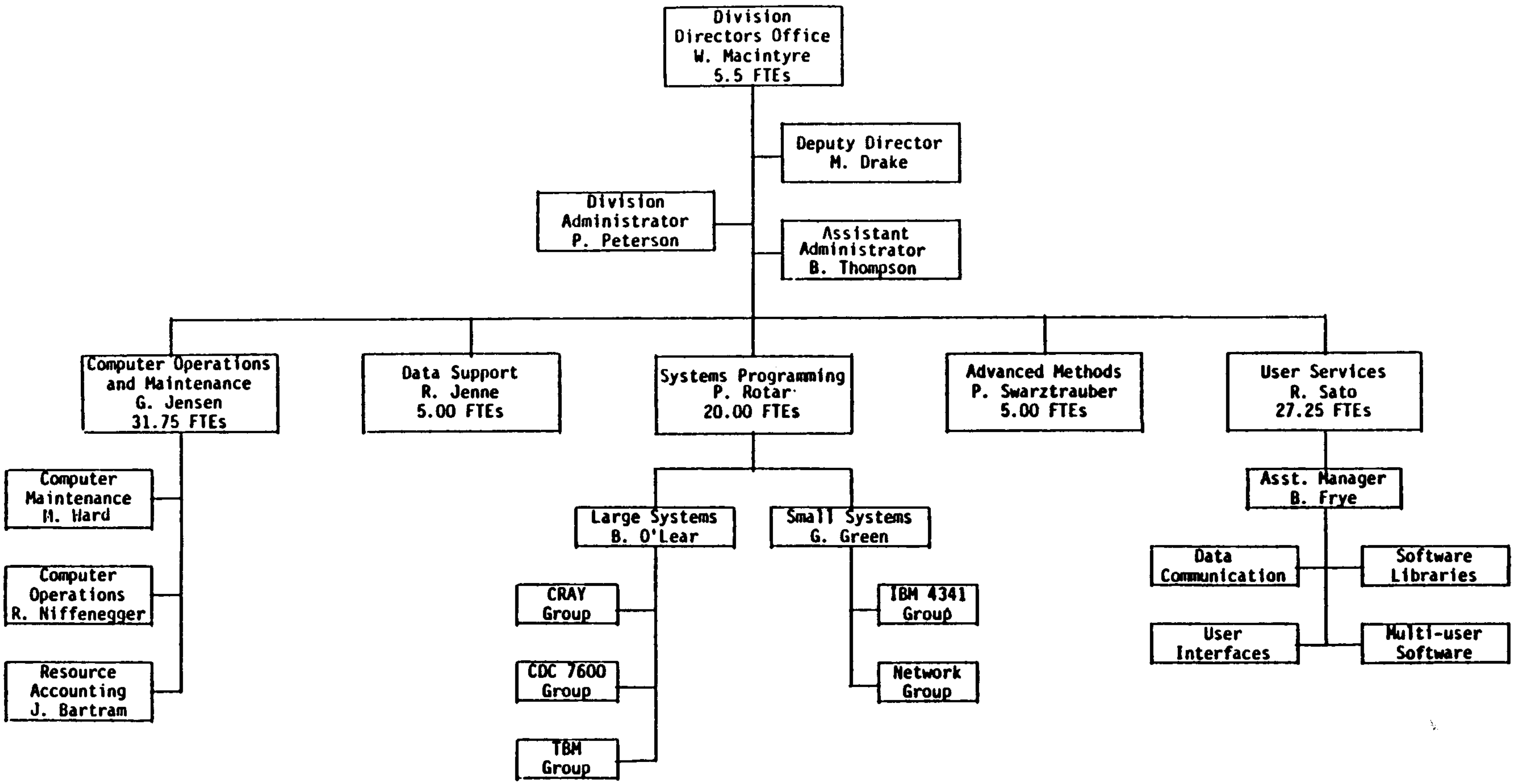

Figure 2

1) Computer Operations - Operates and maintains all SCD computers, peripherals, and related equipment; accounts for the usage of this equipment; processes all film and fiche output.

2) Data Support - Acquires, prepares, and maintains archives of meteorological, oceanographic, and other necessary data; provides counseling services regarding these archives; fulfills requests for data sets for use elsewhere.

3) Systems Programming - Maintains the operating systems on all SCD computers; develops interfaces to these systems for user programs.

4) Advanced Methods - Research, consulting, and software development for problem solving in scientific computing.

5) User Services - Provides the services required to allow users access to SCD computers, documents user interfaces to SCD computers, consults with users on problems, trains users, develops and maintains multi-user software, and acquires and maintains program libraries for general use.

6) Division Director's Office - Overall responsibility for delivering large-scale computing services to users; for allocation of computing resources to non-NCAR users; for administering the activities of SCD including planning, budgeting, and general supervision; for delivering support services to division personnel.

Non-NCAR requests for a total of more than five hours on the CRAY-1 or CDC 7600 are reviewed by the SCD Advisory Panel after preliminary review by two or more individual reviewers. The Panel meets twice a year, in the spring and fall. Panel members are selected from the community at large on the basis of established competence in the atmospheric sciences, computer science, and related fields. Normally, three-quarters of the Panel members are from outside NCAR.

The panel assesses as the merit of proposals for the use of the SCD facilities and recommends action with respect to a prospective user's request on the basis of scientific merit, computational effectiveness and need. The following specific questions are asked about each proposal:

- What contribution is the project likely to make to the advancement of the atmospheric sciences?
- Is the work original?
- Are the scientific approaches and techniques appropriate?
- Will the project make efficient use of computing resources? Will it be I/O bound? Are current mathematical and numerical methods used? Are appropriate algorithms employed?

The Panel may recommend that the request for resources be fully granted, that only a portion of the request be granted (which will mean the project must be scaled back or stretched out in time), or on occasion, that the request be denied. In cases where a request for computing time at NCAR is an integral part of a proposal to NSF, NCAR coordinates its review process with that of the NSF.

Requests for university use of less than 5 hours of CRAY-1 or CDC 7600 time are evaluated by the Director of the Scientific Computing Division, with guidance from scientific reviewers selected from the NCAR staff or external community. These evaluations are reviewed as a group twice a year by the SCD Advisory Panel.

NCAR use of the computer is allocated by the Director of NCAR as part of the overall NCAR budget and resource allocation systems. The Director presents and defends his proposals for NCAR programs and associated resource allocations before the UCAR Board of Trustees and its Budget and Program Committee. The efficient use of resources by NCAR staff is also reviewed through the Scientific Programs Evaluation Committee (SPEC) process, in which scientists drawn from the community at large review all aspects of NCAR programs on a triennial basis.

The joint use allocation is also administered by the Director of NCAR.

MAGNETIC FUSION ENERGY AND COMPUTERS*

John Killeen

National Magnetic Fusion Energy Computer Center
Lawrence Livermore National Laboratory
Livermore, California 94550

The application of computers to magnetic fusion energy research is essential. In the last several years the use of computers in the numerical modeling of fusion systems has increased substantially. There are several categories of computer models used to study the physics of magnetically confined plasmas. A comparable number of types of models for engineering studies are also in use. To meet the needs of the fusion program, the National Magnetic Fusion Energy Computer Center has been established at the Lawrence Livermore National Laboratory. A large central computing facility is linked to smaller computer centers at each of the major MFE laboratories by a communication network. In addition to providing cost effective computing services, the NMFECC environment stimulates collaboration and the sharing of computer codes among the various fusion research groups.

INTRODUCTION

In June 1973 an Ad Hoc Panel on the Application of Computers to Controlled Thermonuclear Research[1] was convened at the USAEC in order

(a) to survey and summarize the existing level-of-effort in the application of computers to CTR.

(b) to identify important CTR physics and engineering questions that are or may be soluble by the use of computers, and to evaluate the benefits that would accrue to the CTR program if such solutions were obtained, and

(c) to survey, summarize and evaluate the status of present and anticipated computer technology for the purpose of accurately forecasting the type, size, scope, and composition of a facility that would realize the benefits identified in (b) above, and the lead time necessary for assembling it.

The wide range of the questions posed to the panel dictated that its composition be as diverse as the issues it was asked to address. Therefore, besides AEC Headquarters personnel, the panel had three different groups of participants--plasma physicists, plasma engineers, and computer scientists--who were organized along functional lines into three sub-panels.

A recommendation of the Ad Hoc Panel was a significant expansion in the development and use of computer models in the fusion program. The following plasma physics models were identified and their importance to the fusion program discussed.[1]

1. Time-dependent magnetohydrodynamics
2. Plasma transport in a magnetic field
3. MHD and guiding-center equilibria
4. MHD stability of confinement systems
5. Vlasov and particle models
6. Multi-species Fokker-Planck codes
7. Hybrid codes

Engineering models needed in fusion reactor design studies include

1. Plasma engineering-burning plasma dynamics
2. Nucleonics
3. Mechanical design
4. Magnetic field analysis
5. Systems studies
6. Thermal hydraulics
7. Tritium handling
8. Safety and environmental studies

Another recommendation of the Ad Hoc Panel was the establishment of a computing facility dedicated to the magnetic fusion program. The National Magnetic Fusion Energy Computer Center (NMFECC) was organized in 1974 at Lawrence Livermore Laboratory and service began in late 1975. From the above lists of computer models in plasma physics and engineering it is clear that the fusion program requires the most advanced scientific computer available. In September 1975, NMFECC installed a new CDC 7600 dedicated to fusion physics calculations.

A review of requirements at the National MFE Computer Center was conducted in the spring of 1976. The results of this review led to the procurement of the CRAY 1, the most advanced scientific computer available, in the spring of 1978. The utilization of this computer in the MFE program has been very successful.[2]

A new study of computer requirements for the MFE program was conducted during 1979.[2] The MFE program has grown considerably since 1973, and major directions have emerged, e.g., TFTR, MFTF, and the proposed Fusion Engineering Device (FED). The computing requirements are necessarily much larger and the models more ied and demanding.[2] The increased emphasis usion technology has led to a substantial growth in the use of the NMFECC for engineering calculations. The 1979 study[2] led to the procurement by the NMFECC of a second CRAY 1 computer in September 1981. The study[2] also recommended the addition of a more advanced computer during 1984.

THE NATIONAL MFE COMPUTER NETWORK

The purpose of the MFE Computer Network is to provide to all fusion researchers the full range of available computational power in the most efficient and cost effective manner. This is achieved by using a network of computers of different capability tied together and to the users via dedicated data lines and dial up telephone lines. The existence of this nationwide computer network allows projects to be sited anywhere in the country without regard to local computer availability, and therefore increases enormously the flexibility of the fusion program.

The Center began first operations using a CDC 6600 loaned by the Lawrence Livermore National Laboratory Computer Center (LCC). In September 1975 the Center installed its own CDC 7600 computer, and by March 1976 significant transmission over the dedicated Data Communications Network was taking place. By that time the new CDC 7600 was saturated with a calculation workload which had been implemented, until then, solely by dial-up telephone communications between the Center and the user community.

Levels of Computer Capability in NMFECC

The concept of the NMFECC is that different levels of computer capability are provided at the various remote locations according to research priorities and anticipated computational demand. At the national center, providing high level capability to the entire community, is the original CDC 7600 plus two high-speed CRAY 1 computers with one and two million words of memory, respectively. Additional equipment at the national center includes processors and other ADP equipment for communications, file management, and data storage. (Figure 1)

At the next level of capability are User Service Centers (USC's): DEC-10 computer systems with direct high-speed access to the national center through PDP-11/40 remote communications control processors. There are now five operational USCs (Figure 2) in the field located at Princeton Plasma Physics Laboratory (PPPL), the Los Alamos ional Laboratory (LANL), the Oak Ridge onal Laboratory (ORNL), General Atomic (GA), and LLNL (for the mirror confinement program). A sixth USC, used in center operations, is located at the NMFECC itself.

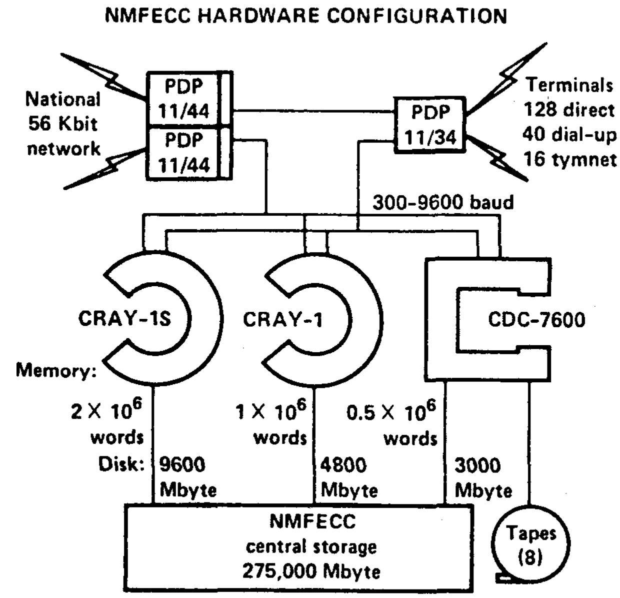

Figure 1

A third level of capability is provided through the Network Access Port (NAP). MFECC designed the NAP to permit remote computers to be connected to the MFE network as remote hosts. There are currently five NAPs installed, all of them connecting VAX 11 series computers into the network.

A fourth level of capability is provided by Remote User Service Stations (RUSS) at selected MFE research sites. RUSS stations are currently installed at 18 remote locations (Figure 2). RUSS stations provide users with the capability of printing output files locally on a 1000 line/minute printer and act as a terminal concentrator for up to 16 interactive terminal users. RUSS stations are connected to the nearest MFE-NETWORK communications processor over a 4800 baud dedicated line. (Figure 2)

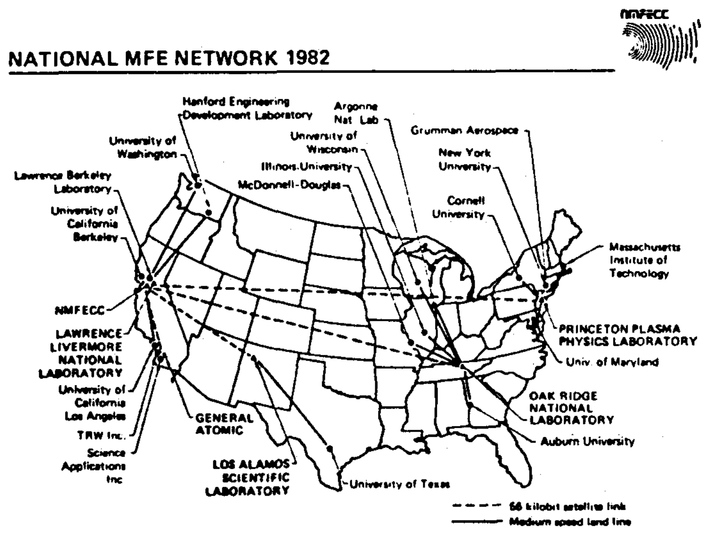

Figure 2

A fifth level of capability is dial-up access. MFE researchers at some 35 localities access the center using computer terminals equipped with acoustic modems or couplers.

Data-Communications Systems

Data Communications service to the National MFE Computer Center is provided on a 24 hours/7 day basis. Three types of service are provided to NMFECC users as outlined below:

1. Wide band (56 KB/sec) Satellite Network Service. Users at Major USC's on the MFE net may log on to their local DEC-10 system and interact with the computing resources at the Central facility in Livermore. Currently four major network satellite links are in service from LLNL to Princeton, N.J., Oak Ridge, Tenn., Los Alamos, N.M., and San Diego, Calif. These are dedicated dual channels and modems which are connected to LLNL owned communication control processors (DEC 11/40's). (Figure 3)

MFENET TYPICAL SITE CONFIGURATIONS

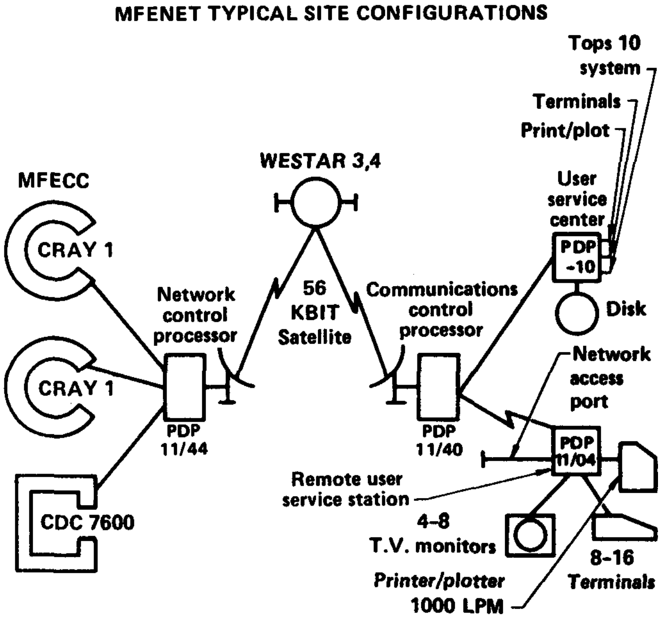

Figure 3

2. Dedicated 4800 Baud Service. Remote User Service Stations on the MFE Net are served by dedicated leased 4800 baud lines which terminate either at the Center (LLNL) or at the nearest MFE Communications Control Processor. Location's served by 4800 baud dedicated service are shown in Figure 2.

3. Dial Up Service. Users not at major fusion laboratories may dial-up the Center using one of the following services:

(a) TYMNET--A Tymnet owned CP-16A/1200 processor is installed at the NMFECC in Livermore. NMFECC users have access to the six unlimited 300 baud ports and two unlimited 1200 baud ports. TYMNET calls are also routed through the LLNL owned terminal concentrator.

(b) ARPANET ACCESS--NMFECC users with access to an ARPA IMP may use the ARPANET as a means communicating with the Center.

(c) DIRECT DIAL COMMERCIAL or FTS--Thirty 300 baud and ten 1200 baud ports are available through auto-answering modems which connect dial up users to MFECC's terminal concentrator.

NMFECC COMPUTING ENVIRONMENT

The NMFECC computing environment reflects the needs of computer users in the Magnetic Fusion Energy research community. Both interactive timesharing and batch processing are available. A summary of some service follows:

Timesharing Services

The fusion community has always found that interactive computing, even with the largest codes, is by far the most efficient use of physicists efforts. The 5% overhead in swapping codes in and out of the machines provides fast debugging, immediate turn around on key results, and the capability to interact with codes which need user control. The Livermore Time Sharing System (LTSS) developed for the CDC 7600 by the LCC was adapted by the NMFECC for the CRAY 1 computer in about six months. CTSS was available as the first CRAY 1 was delivered and the final bugs were removed within a couple of months. CTSS is supported by libraries of FORTRAN callable subroutines which enable a user to issue almost every system call, giving access to every part of the hardware. A typical physics code can be run from a terminal, display graphics as it runs, be interrupted or interrogated at any time. The ability to start or stop a code at any point and inspect the results provides debugging at least 100 times faster than older methods.

File Storage Services

NMFECC has designed a multi-level file storage system called FILEM. FILEM is a highly versatile system which allows users to store and retrieve programs and data files in the central computing facility at Livermore for an indefinite period of time. FILEM has been designed to accommodate the needs of users at remote sites. The CDC 7600 has been programmed to assume virtually all of the tasks associated with file custodianship including indexing, storage, retrieval, and efficient management of the file storage media. The NMFECC file storage media currently consist of three levels of storage (Figure 4).

User Services

It is the policy of NMFECC to make all computer documentation available on line so that users may provide themselves with up-to-date system documentation by simply printing out the document at their local printer or terminal. Any part of any document may be displayed on remote terminals and the routine DOCUMENT is capable of scanning text for the user to locate a specific topic of interest. NMFECC has provided the user with two routines called MAIL and NEWS which allow users to send a message or question to any other user on the NMFECC network. NEWS and MAIL are also commonly used by users to ask NMFECC staff about specific problems they have encountered. NMFECC systems programmers and documentarians use NEWS to broadcast all system or documentation changes. Users who are unable to solve a computational problem by consulting DOCUMENT or inquiring through NEWS may seek assistance from the software consultation staff at the Center. Depending on the user's needs the staff may diagnose problems, recommend solutions, and follow through to insure that a satisfactory solution is realized. Specialized assistance in the areas of mathematical libraries, graphics, engineering analysis and symbolic manipulation is also available through the Center.

NMFECC CENTRAL STORAGE

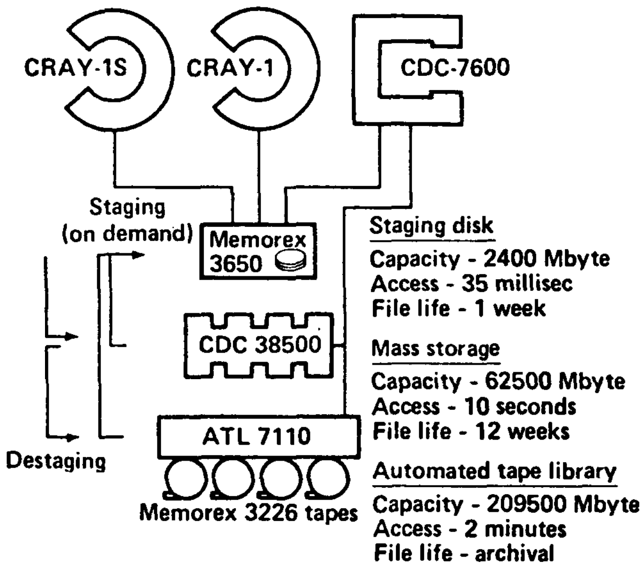

Figure 4

LIBRIS Computational Plasma Physics and Engineering Data Base

NMFECC's policy is to encourage the exchange of information about application codes within the MFE computational community. To facilitate sharing, a common data base of computational physics and engineering codes is available through a routine called LIBRIS, which allows any user to abstract a code that he wants to make available to the MFE community or to interactively interrogate the abstracts in the LIBRIS data base.

NMFECC PLANNING PROCESS

The need for a more advanced computer in 1984 should properly be discussed from two perspectives. The first perspective has to do with capability and can be discussed in the context of the following question: Are there important fusion problems that cannot be solved with todays most capable computers? The second perspective has to do with capacity and suggests the question: With two CRAY machines in operation does NMFECC still fall short of the projected requirements of the fusion program? The answer to both questions is yes.

The Need for More Powerful Computers

The central focus of magnetic fusion theory is the behavior of plasma as it interacts with magnetic fields. The growing body of physics theory that pertains to this interaction is extremely complex. Computers such as the CRAY 1 have permitted code designers to simulate plasma interactions more accurately, but improved models are needed and the improvements will require machines more capable than a CRAY-1S. It should be emphasized that increases in memory size should be matched by increases in CPU cycle time. Expanding the parameters of a problem to use a larger memory size might otherwise result in unreasonably long run times to obtain a result. Thus we look to a machine which will demonstrate significant improvements in both memory size and execution speed over the CRAY-1S.

We can examine the need for more powerful computers with respect to some of the major types of codes which are used in the MFE program.

(a) MHD Normal Mode Codes - One approach to the determination of MHD behavior is the calculation of linear normal modes; that is, natural modes of oscillation of small perturbations away from equilibrium. This has been carried out in very detailed fashion by the PEST and ERATO codes, which calculate an equilibrium and then obtain the structure and frequencies of these normal modes. These codes have made vital contributions to the understanding of the dangerous kink and ballooning instabilities, as well as other MHD modes in tokamaks. This has permitted the much more accurate determination of limits on the plasma beta, the ratio of the plasma pressure to the magnetic field pressure. Both codes run approximately ten times faster on the CRAY 1 than on the CDC 7600. This substantial decrease in running time results not only from the fact that the CRAY 1 is a much faster machine, but also because its fast access memory is substantially larger, thereby permitting more efficient coding. Even so, both PEST and ERATO and other codes of their type require a great deal of computer time. Each computer run takes about thirty minutes on the CRAY 1; several runs are required to analyze one equilibrium; and hundreds are required to discover trends. Such stability calculations will be vital in the investigation of future experiments and reactor configurations.

(b) Time-dependent MHD Codes - The other technique for determining plasma instabilities along with their growth rates is through the solution of the time dependent MHD equations of motion. The full set of MHD equations comprise a coupled system of eight nonlinear partial differential equations, the solution of which is a formidable task on any computer system. In order to make these computations tractable, approximations have often been made, including reduction in dimensionality, linearization, restriction to a particular geometry, ordering, or regime, and the assumption of no transport or resistivity. To explain complex phenomena such as the major disruption in a tokamak requires a three dimensional resistive code. It is of utmost importance to understand the major disruption, for if it occurs it will cause the tokamak walls to vaporize.

The rate of progress of resistive MHD calculations in tokamaks has been strongly dependent on two factors: advances in numerical techniques and increases in computer capability. These calculations have produced a succession of results which have given increasingly detailed comparison with tokamak experiments. However, to obtain results on a CRAY 1, it has been necessary to impose numerous simplifications which, if removed, might materially increase our understanding of fusion devices. We expect that further advances in numerical techniques will be hard to achieve and will offer even smaller speed advantages. Thus, further progress in this area is likely to be strongly coupled to computer capability.

The theory is at a stage now where one can construct a computer program to include "exact" treatment of tokamak geometry and pressure effects with a realistic level of resistivity and some kinetic effects. However, it is doubtful that the CRAY 1 has the capability to run such cases. It is even more doubtful that such calculations can be made for stellerator configurations. The primary limitation is CPU speed. A secondary consideration is increased fast memory size.

The recent advance in three-dimensional MHD calculations for tokamaks has depended crucially on obtaining a reduced set of equations by expanding the original MHD equations in a small parameter, which is on the order of the inverse aspect ratio. This is possible because of the strong and almost uniform toroidal magnetic field in tokamaks. Since the field components in the Reversed Field Pinch (RFP) are all of the same order, and since these devices possess finite beta, there exists no universally small parameter in which to expand the basic equations. Additionally, the computational speed of the codes based on the tokamak reduced equations is greatly enhanced by the assumption of incompressibility, which eliminates the compressional Alfvén wave. Because of the strong field in a tokamak, the fastest remaining mode evolves on a time scale on the order of the major circumference divided by the Alfvén velocity. This time scale may be more than an order of magnitude longer than that of the compressional Alfvén wave. In the RFP, on the other hand, even the assumption of incompressibility does not provide much of an advantage, since now a shear Alfvén wave propagating near the field reversal point evolves on a time scale on the order of the minor circumference divided by the Alfvén velocity, a reduction of a factor of only 2 or 3 from that of the compressional Alfven wave. Thus the next meaningful step in MHD simulations for the RFP will tax even the next generation of computers.

To make these three-dimensional codes applicable to more general geometries (e.g., stellarators) and to simultaneously include enough effects to ensure a complete description of the important physics effects (e.g., parallel heat transport, compressibility, finite larmor radius effects, and smaller values of resistivity) will require a machine with about 100 times the CPU speed of the CRAY 1 in order to keep the run times about the same i.e., tens of hours. Finally, an increase of a factor of 10 in memory size would allow a factor of 2 increase in each direction of a three-dimensional calculation, while keeping the entire code resident in fast memory. This increase would provide a significant improvement in resolution.

(c) Particle and Hybrid Codes - In many cases fluid models are not adequate to describe plasma behavior, for it is necessary to consider microscopic effects, i.e., the effects of the way particles are distributed in velocity. Numerically this is most often accomplished through particle codes. Fully nonlinear kinetic ion and electron simulations in 2-D Cartesian geometry were carried out over the last decade. In the past, Cartesian geometry was not a major physics limitation even with the obvious cylindrical and toroidal nature of experiments, because these models necessarily dealt with length and time scales on the order of the electron gyroradius and plasma oscillation period for stability. Resolving such length and time scales meant that any realistic macroscopic dimension could be considered infinite in light of the huge number of computational time steps required for any information to travel such a distance. Such models are primarily useful for plasma transport studies, which are made computationally accessible by studying relatively sharp gradients. Even so, present 2-D computational methods still require an unrealistically small mass ratio M_i/M_e and other artifical compression of disparate time scales to retain acceptable run times. Implicit and orbit-averaged methods improve the situation somewhat, but routine 3-D simulation with realistic parameters is simply not practical with present computational resources. In addition, with the increase of grid resolution allowed improved computers and methodology, the scope of particle simulations has grown to encompass nonlocal effects and more realistic geometries. This further adds to the complexity of codes and has lead to renewed demand for more memory and speed of the computer.

At present computers, large scale particle simulations in 2-1/2D and 3D are mainly limited by the size of the maximum fast memory the CRAY 1 can handle (of the order of 1 M words, or 2 M for the CRAY-1S). With necessary diagnostics this amount roughly corresponds to two-dimensional grids of 128 x 32 or 64 x 64 for electromagnetic particle codes and to a three-dimensional grid of 32 x 16 x 8 for MHD particle codes. For example, in order to have relevant mode-conversion physics for Ion Cyclotron Resonance Heating (ICRH) in tokamak geometry, a minimum grid of 128 x 128 was necessary.

An enhanced Class VI computer with 2.5 times more memory and speed than a CRAY-1S will permit a grid of 128 x 64 and perhaps 128 x 128 in an electromagnetic particle code and use more realistic parameters. Experimentally relevant physics problems in magnetic confinement have important three-dimensional aspects, such as in the multiple-helicity interaction of collisionless tearing modes and in the drift wave turbulence in sheared magnetic fields; the increased memory and speed will increase the practicality of 3-D simulations. It is, nevertheless, clear that such an increase in memory will not be enough. The advent of a Class VII computer with memory of the order of 30 M words, for example, will be able to tackle a 512 x 128 system in electromagnetic codes, corresponding to a plasma of the size of several tens to a hundred collisionless skin depths.

In addition to this, one should consider the acquisition of a large fast solid-state peripheral memory. This will allow users to double or triple-buffer the grid and particles without resorting to much slower discs. It would be helpful to have a solid-state memory of around a few hundred million words for large scale 2D and 3D calculations. When a large fast solid-state peripheral memory is attached to the fast core, it is very important to have fast, efficient, large scale I/O between the core and memory. It may be preferable to arrange the solid-state memory as virtual memory with some paging capabilities.

Particle-fluid hybrid models have become important in the last five years. A typical hybrid model represents the ion components as kinetic species and the electrons as a fluid in order to eliminate some or all fast electron frequencies and short length scales. Without these electron-imposed limitations, the kinetic ion effects can be modeled on macroscopic, almost MHD, time and length scales—making experimental relevance much easier to establish. The ion temperature gradient drift instability in a tokamak was studied with such a model. Sometimes, without sacrificing the parallel dynamics of electrons, the guiding-center particle model is used. This method has been actively used particularly for tokamak plasmas.

Recent progress with hybrid models is impressive but is still quite computationally expensive (typically taking roughly two to four times more CRAY CPU time than does an MHD code of equal dimensionality). Further, 2D meshes of size 128 x 128 with 20 particles per cell lead to memory requirements of order 2 x 10^6 words. Results obtainable with present 2-D codes, as well as progress on 3-D codes, are presently hampered by lack of CPU speed and memory capacity. For example, typical 2D quasi-neutral hybrid simulations of rotational instabilities in the Field Reversed Experiment at Los Alamos use 50,000 particles, require approximately 3 hours of time on the CRAY 1, and use half of the CRAY 1 active memory.

In order to resolve ion spin-up effects in rotational instabilities of compact toroids, it would require at least 10 hours of CPU time on the CRAY 1 for a simulation run with a present 2D hybrid code. It would also be desirable to increase the number of particles used in simulations in order to more properly represent the ion velocity distribution in low density regions; however, this is not practical on the CRAY 1, without resorting to disks and buffering and hence motivates the use of a Class VII computer.

Another new particle simulation technique that has led to much more realistic simulation of fusion experiments on transport time scales is orbit averaging. In orbit-averaged simulations, time-splitting has been combined with temporal averaging to allow the self-consistent solution of Maxwell's equations on a slow time scale using a long time step, but the particle dynamics are followed on the natural time scale of the orbit. Orbit averaging in a two-dimensional, magneto-inductive algorithm has been successful. A natural separation of time scales allows orbit averaging over the particle trajectories (in the cases studied so far, this means averaging over many ion-cyclotron and axial-bounce periods in a magnetic mirror), and great reductions in the number of required particles have been achieved. Simulations have been performed with parameters that directly correspond to the Lawrence Livermore National Laboratory 2XIIB experiment. These simulations were able to simultaneously resolve the ion cyclotron time scale in the vacuum magnetic field $\omega_{ci} = 2.8 \times 10^7 s^{-1}$, and the ion-electron slowing down rate $\nu_s i/e = 3 \times 10^2 s^{-1}$ without artificial distortion. The new simulations required 500 to 1000 ions rather than the ~20,000 previously needed, had a correspondingly smaller memory requirement (were contained in core), and were able to run 10 to 100 more steps in the same amount of computer time (two to three hours on the CRAY 1).

Extension of the orbit-averaged simulation model to include fluid electrons and provide a self-consistent implicit calculation of the ambipolar potential has been undertaken to simulate tandem mirror configurations. The inclusion of electron effects requires that an even finer time scale than the ion cyclotron period be resolved, viz, the electron transit time in the mirror or mirror end-plug of the tandem. This time scale is roughly a factor of ten faster than the ion gyration and will require that a realistic simulation span the additional spread in time scales. Because of the complexity of tandem mirrors with thermal barriers a significant increase in axial grid resolution is also required. Accommodating the first few azimuthal Fourier modes to incorporate quadrupole and elliptical flux-tube effects will further strain present capability. A 2 x 10^6 word memory and a machine 2 to 4 times faster than a CRAY 1 will not suffice without continued artificial compression of time scales.

Particle simulations of microinstabilities in mirror devices also strain the limits of capabilities of a CRAY 1 computer. Loss-cone simulations with a stretched one-dimensional code were performed with 0.5-20 x 10^4 particles, 64-256 grid points, $\Delta x \approx (0.05 - 0.2)\rho i$ and $\omega_{ci}\Delta t$ = 0.05 or 0.1. To accommodate the disparate time scales of lower hybrid oscillations, ion cyclotron and ion bounce motion, the linear growth of microinstability, neutral beam charge-exchange, and ion drag, in the simulations, the rates of the slower processes have been artificially accelerated (by as much as 10^2); and either $\omega_{pi}^2/\omega_{ci}^2$ or m_i/m_e is at least a factor of ten smaller in the simulations than it is in mirror plasmas like 2XIIB and TMX end-plugs. Nevertheless, even with somewhat artificial parameters and very simple models of the aspects of a neutral-beam driven mirror machine, the simulations of loss-cone modes give many results in agreement with experimental data and quasi-linear theory. The cost of these simulations scales directly with $(m_e/m_i + \omega_{ci}^2/\omega_{pi}^2)^{-1/2}$ and the number of particles. Typical simulations require less than one hour on the CRAY 1 with particle data stored on disks and input/output overlapped, but a few of the simulations were as long as two to three CRAY CPU hours. More realistic microstability simulations of plasmas with parameters more closely approaching 2XIIB, TMX, TMX-U, and MFTF-B will be possible on an enhanced Class VI computer and easier still on a Class VII; and the code could be contained in core. Increasing the dimensionality and incorporating important electron and electromagnetic effect await a Class VII machine.

The recent advent of the implicit particle codes, which allow large time step without compromising microscopic physics, makes it possible to run a particle code to examine slow phenomena such as drift waves within a reasonable computational time. This means that the particle simulation technique can describe an enriched and enlarged field of physics. Because implicit particle codes require storage of additional grid or particle data over that stored in conventional explicit codes, memory requirements are increased. Furthermore, the desire to simulate experiments using fully realistic parameters in 2D or 3D continues to demand computer capabilities beyond Class VII.

(d) Fokker-Planck Codes - In the simulation of magnetically confined plasmas where the ions are not Maxwellian and where a knowledge of the distribution functions is important, kinetic equations must be solved. At number densities and energies typical of mirror machines, end losses are due primarily to the scattering of charged particles into the loss cones in velocity space by classical Coulomb collisions. The kinetic equation describing this process is the Boltzmann equation with Fokker-Planck collision terms. The use of this equation is not restricted to mirror systems. The heating of plasmas by energetic neutral beams, the thermalization of alpha particles in DT plasmas, the study of runaway electrons and ions in tokamaks, and the performance of two-energy component fusion reactors are other examples where the solution of the Fokker-Planck equation is required.

The problem is to solve a nonlinear, time-dependent partial differential equation for the distribution function of each charged species in the plasma, as functions of six phase space variables (three spatial coordinates and three velocity coordinates). Such an equation, even for a single species, exceeds the capability of any present computer, so several simplifying assumptions are required to treat the problem.

The most advanced state-of-the-art time-dependent Fokker-Planck code assumes that the distribution functions depend on one spatial coordinate (radius r) and two velocity coordinates (speed v and pitch angle θ). Moreover, the collision operator at each radius depends only on the distribution functions at that radius, so that a zero-spatial-dimensional, two-velocity-space-dimensional Fokker-Planck solver may be utilized.

Such a solver requires 11.5 μs per meshpoint on the CRAY 1 to compute the Fokker-Planck operator and time-advance the distribution function (using an alternating direction method) for one species. For a typical 101 (v) by 81(θ) mesh this comes to 0.094 seconds per timestep per species. Allowing for 10 radial meshpoints, two species and 1000 time steps, the total amount of computer time is 31 minutes.

The preceeding is not intended to give the idea that any zero-spatial-dimensional, two-velocity-space dimensional Fokker-Planck problem can be easily handled by the CRAY 1. Fusion efficiency studies of the D-D fuel cycle require the solution of Fokker-Planck equations for five ionic distribution functions, including very high energy protons. Using a state-of-the-art multi-species Fokker-Planck code, the maximum allowable mesh size on the CRAY 1 is approximately 161(v) by 30(θ), and the computer time per timestep (using a fully implicit method) is

12 sec. A typical problem requires anywhere from 1 to 4 hours. Moreover, there are many regions of parameter space which require two to five times as many meshpoints. Some of these problems can be attacked on the CRAY-1S, at a premium cost, and others will require an extended Class VI or a Class VII machine.

There are many situations in which the charged particles execute regular orbits on a timescale much faster than their collision time. In such cases the Fokker-Planck equation need not be solved everywhere in space; instead bounce-averaging can be employed. Toward that end a two-velocity-space dimensional zero-banana-width Fokker-Planck solver has been developed. In its first approximation the ambipolar potential in the bounce-direction is ignored, thereby simplifying the orbit equations.

For a 101(v) by 81(θ) mesh with 25 axial (z) positions (over which the Fokker-Planck coefficients are averaged) 0.32 sec per timestep are required. The Fokker-Planck-related storage requirement is about 400,000 words. When placed in a radial transport code with 25 radial points, 8 sec per timestep will be required, so that a typical 700 timestep problem will take 93 minutes.

Let it be emphasized that the above 93-minute problem is a gross simplification of what is needed. In mirror-like devices such as tandem mirror plugs and EBT, the axial electric field cannot be ignored. Since the midplane transformation $(v,\theta)\rightarrow(v_0,\theta_0)$ now depends on v, to use a similar algorithm requires, for a 101-point v-mesh, 101 times as much storage, which is clearly prohibitive. It will therefore be necessary to recompute the transformation arrays each timestep; this should result in a factor of at least 2 increase in computer time.

More important, the zero-banana-width assumption is totally invalid for modeling neutral beam injection into a tokamak. Including a finite banana width will not only greatly increase the storage, but it will also result in a factor of at least 5 increase in computer time, so that a typical one-species 700 timestep problem on a 101(v) by 81(θ) by 25(r) mesh with 25(z) positions will take, on a Class VII machine, 465 minutes.

An example of an important 3-D (r, v, θ) calculation which is beyond the capabilities of the CRAY 1 and which taxes the limits of an extended Class VI is the modeling of the transport of electron energy out of a tokamak due to the combined effects of a stochastic magnetic field and a radial ambipolar field coupled to a Fokker-Planck model for Coulomb collisions. This problem is both nonlinear and essentially 3-D. Using an implicit scheme employing a 3-D ICCG matrix inversion package, assuming a mesh of about 120,000 points (a minimum for a physically reasonable 3-D calculation), and a cost of 1.5×10^{-3} seconds per time step per mesh point on the CRAY 1, and assuming that a calculation requires 200 time steps, the amount of CRAY 1 computer time required is about 10 hours, generally an unacceptable amount of time for a single run. Incidentally, the total of storage required would be about 50% greater than the matrix size or about 3.4×10^6 words. This could be accommodated only on an extended Class VI or on a Class VII machine. Assuming that these machines are respectively 4 and 10 times faster than the CRAY 1 implies that this calculation would require 2.5 hours on the extended Class VI and 1 hour on the Class VII machine. Consequently the extended Class VI would be only marginally adequate both in terms of speed and storage for this calculation, whereas on a Class VII computer the problem would be tractable.

Summary

In summary, as the fusion program has advanced rapidly in the last few years with the development of more sophisticated theory and experiment, computational requirements for accuracy and realism have increased to the point that Class VII capabilities and beyond are urgent.

It is not possible to define a performance level that represents the ultimate capability for fusion studies. Each successive generation of supercomputers has been eagerly awaited by the user community. Codes to exploit the new hardware capabilities are typically under development before the hardware is actually installed. It is safe to assert that the fusion computing community can effectively use the best performance that the supercomputer manufacturer's are capable of providing for the foreseeable future.

ACKNOWLEDGEMENT

The section on The Need for More Powerful Computers contains contributions from R. Hicks, ORNL; D. Monticello, PPPL: A. Sgro, D. Hewett, D. Harned, LANL; B. Cohen, LLNL; T. Tajima, IFS Texas; A. Mirin, M. McCoy, MFECC. The earlier sections contain contributions from J. Fitzgerald, MFECC. All of these contributions are gratefully acknowledged.

REFERENCES

1. "The Application of Computers to Controlled Thermonuclear Research" edited by Bennett Miller, WASH-1296 July 1973, USAEC.

2. "Magnetic Fusion Energy and Computers" edited by John Killeen, DOE/ER-0033 October 1979, USDOE.

*Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48.

National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE THE POTENTIAL OF LOS ALAMOS NATIONAL LABORATORY TO PROVIDE LARGE-SCALE COMPUTATIONAL CAPABILITIES TO THE RESEARCH COMMUNITY

AUTHOR(S) Bill Buzbee and Dale Sparks

MITTED TO Workshop on Large-Scale Computing for Science and Engineering
National Science Foundation, Washington, DC
June 21-22, 1982

os Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

THE POTENTIAL OF LOS ALAMOS NATIONAL LABORATORY TO PROVIDE LARGE-SCALE COMPUTATIONAL CAPABILITIES TO THE RESEARCH COMMUNITY

by

Bill Buzbee and Dale Sparks

Los Alamos National Laboratory
Computing Division
P. O. Box 1663
Los Alamos, New Mexico 87545
(505) 667-1449

I. INTRODUCTION

In the Spring of 1982, the Department of Defense, National Science Foundation, the Department of Energy, and the National Aeronautics and Space Administration organized a "Workshop on Large-Scale Computing for Science and Engineering." The charter of the workshop is to determine needs and examine methods of providing large-scale computing capabilities to science and engineering; also to study the future of computer technology in the United States. Further details on the workshop can be found in Appendix A. The Computing Division of Los Alamos National Laboratory was invited to participate and to prepare this position paper on how it would provide large-scale computational capability to the research community.

A. What Los Alamos Can Provide.

Los Alamos operates one of the most powerful scientific computing facilities in the world. Our mission is to provide the computing resources required in scientific research and large-scale numerical simulation. Thus we have assembled a wide variety of hardware, software, communication facilities, and services. In carrying out our mission, we seek to maximize first the productivity of people and second the productivity of hardware. Currently, we have approximately 3000 validated users within Los Alamos County and another 500 users distributed throughout the United States. Although our users are engaged in a wide spectrum of applications ranging from document preparation to large-scale scientific simulation, they have the following basic needs in common:

- suitable hardware/software,
- convenient access,
- mass storage,
- a menu of output options,
- support services, and
- ease of use.

1. Suitable Hardware/Software.

The nucleus of our computing facility is a collection of computers designed primarily for scientific computation. They are described in Table I.

TABLE I

COMPUTERS USED FOR SCIENTIFIC COMPUTATION

Quantity	Description	Operating System
4	Cray 1	Interactive (CTSS)
4	CDC-7600	Interactive (LTSS)
1	CDC-6600	Interactive (NOS)
2	CDC Cyber 73	Interactive (NOS)
20	DEC VAX 11/780	Interactive (VMS and UNIX)

The Cray 1s and CDC 7600s are used for number crunching. The CDC-6600 and Cybers provide continuity with the past while supporting administrative computing. The VAXs are incorporated into a distributed network called XNET. They are used to support experimental facilities controlling experiments, collecting and analyzing data, while at the same time providing a modern software-rich environment, for example, screen editing, virtual memory, Fortran 77.

Note that an interactive operating system is provided on every computer. Experience shows that interactivity maximizes the productivity of users.

2. Convenient Access.

The typical user must frequently refer to notes, books, documentation, etc.; thus, the most convenient point of access is from the employee's office. We provide this through an integrated network exploiting our interactive operating systems. A functional diagram of the network is shown in Figure 1. Our network is partitioned into three security partitions: Secure, Administrative, and Open. For the purposes of this discussion, we need consider only the Open partition. That partition contains the following computers:

- 3 CDC 7600s,
- 1 CDC Cyber 73, and
- 9 VAX 11/780s.

A user in the Open partition can sign onto any of the above computers. Further, we would like to acquire a Cray 1 computer for the Open partition and are eager to work with this workshop and other interested parties to make the appropriate arrangements.

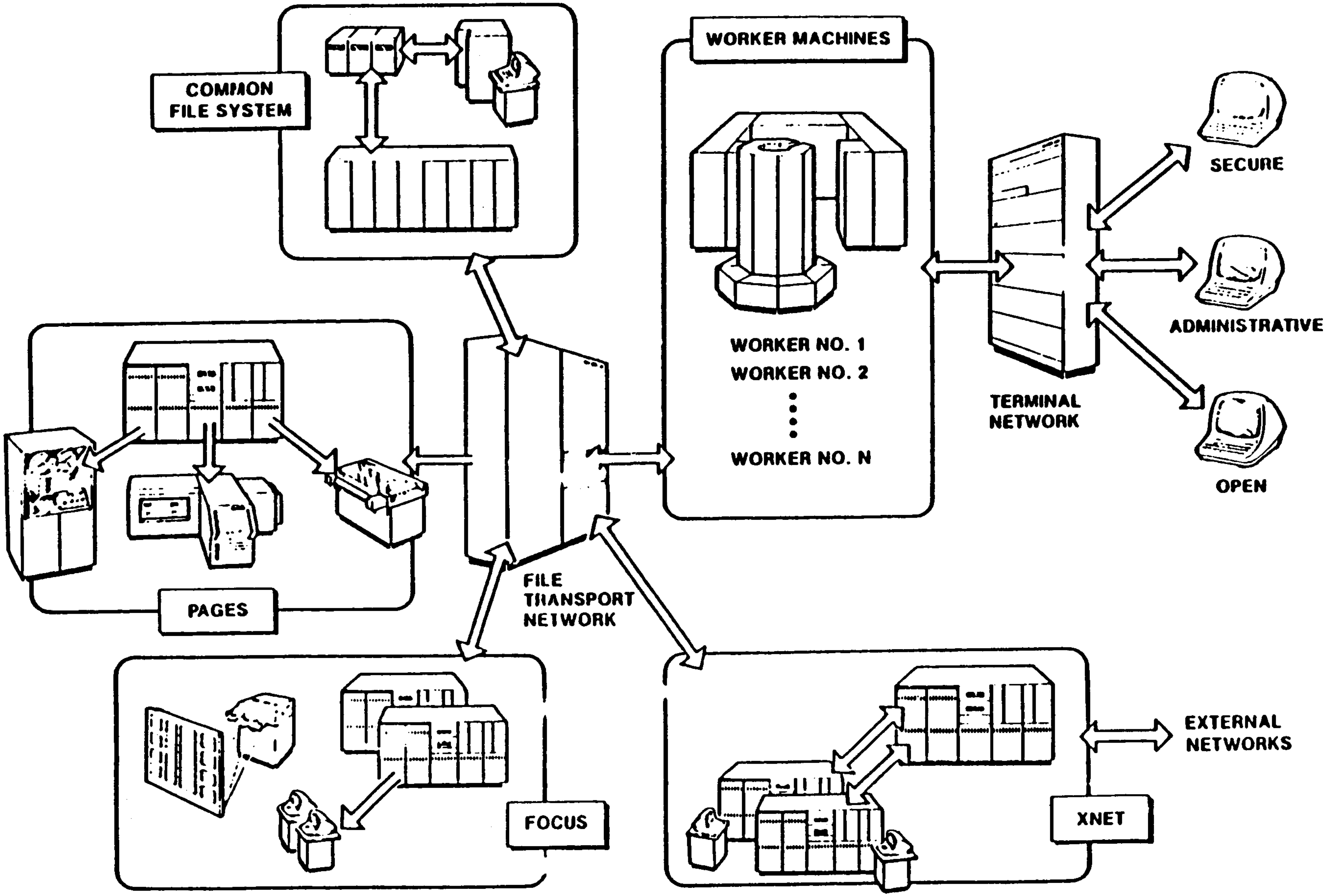

Figure 1.
Functional diagram of the Los Alamos Integrated Computing Network.

3. Mass Storage.

Supercomputers can consume and generate enormous amounts of data in a short time. Thus every large-scale computing facility needs a mass storage facility of considerable capacity. Most large-scale computing facilities in this country recognize this need; however, few of them have mass storage facilities with sufficient efficiency and reliability to be quality systems. Our system is a quality system using IBM equipment and is shown in Figure 1 as the Common File System (CFS). This equipment has a large market among banks and insurance companies and thus there is significant customer pressure to make the equipment work and keep it working. Currently, our system has an online capacity of 2.7 trillion bits and unlimited offline capacity. To date, its availability averages 98-99%. The file organization within it is tree structured much like UNIX. One of its novel features is automatic file migration. Storage within the system is hierarchical. The 3350 disks provide the first level of storage, the IBM 3850 cartridge store provides the second level and offline cartridges provide the third level. Recent performance data from this system are given in Table II.

TABLE II

RECENT CFS PERFORMANCE DATA

Level of Storage	% of Total Data Contained in this Level	% of Requests Satisfied From this Level	Average Response Time From this Level
Disk	1	82	10 seconds
3850	17	17	1 minute
Offline	82	1	5 minutes

<u>We believe this is one of the finest mass storage systems in the world. It is accessible from all of our computers and utilization of it would be included in our provision of large-scale computing facilities.</u>

4. Output Options

In code development and exploratory computations, the scientist often needs a variety of output options. For example, during code development and problem checkout, graphical display at the terminal can be extremely valuable. During the course of a parameter study, one may wish to put copies of the source and numerical results on microfiche for efficient archival storage. If one is dealing with a time-dependent problem that consumes several hours of supercomputer time, then one may wish to produce a movie to show time dependencies. Finally, in code development, there is the time-tested axiom "when in doubt, count out." All of these options are available in the Los Alamos Computing

Facility through the equipment listed in Table III (shown in Figure 1 as PAGES, Print and Graphical Express Station).

TABLE III

OUTPUT OPTIONS AND EQUIPMENT

Output Option	Equipment
8 1/2- x 11-in paper (double sided)	2 Xerox 9700s
11-in roll (electrostatic)	Versatec
36-in roll (electrostatic)	Versatec
36-in vellum (electrostatic)	Versatec
16-mm color film	FR80 film recorders
35-mm color and black and white film	FR80 film recorders
105-mm microfiche	FR80 film recorders

PAGES is accessible online from all of our computers and utilization of it could be included in our provision of large-scale computing.

5. Support Services

Requisite support services include

- operations,
- documentation,
- education,
- consulting,
- accounting, and
- research.

We are particularly proud of the efficiency in our operations. Excepting the VAX 11/780s, all of our computers, the Common File System, and the output equipment in PAGES are operated 24 hours a day, 363 days a year by a total of 76 people. As evidenced by CFS and PAGES, part of our objective is to automate where possible. Thus, our Facility for Operations Control and Utilization Statistics (FOCUS in Figure 1) is a recently added node to our network from which we load level production jobs across the supercomputers. Because of automation, we anticipate no growth in our operation staff in the next few years despite plans to significantly increase our total computing capacity.

Documentation is the Achilles heel of computing. The Los Alamos Computing Division is organized into eight working groups. Because of the breadth of our network and unique facilities such as the Common File System, one of those groups has responsibility for developing and maintaining documentation. This is a measure of the importance we attribute to documentation. Much of our documentation is available online and our goal is to put all of it online in the near future.

Education is helpful to new users and we already make considerable use of video cassettes and computer-aided instruction. Should we offer service to a national computing community of users, these would probably be the primary media for education.

Consultation is required when users encounter difficult bugs or when they seek information about how to accomplish sophisticated tasks. Today we have a staff of approximately 10 consultants; this staff would have to be increased should we offer service to a larger community. Since the introduction of interactive systems, the consultants' primary mode of communications with users has been by telephone and it would likewise be the primary mode should we provide service to a national community.

The objective of our accounting system is to charge equitably for all resources consumed. We charge for CPU utilization, memory utilization, storage of information in the Common File System, number of pages printed, frames of film generated, etc. A summary of projected charges for FY 83 are attached in Appendix B.

Our research divides into three areas:

- parallel processing,
- person/machine interface, and
- modeling support.

The next generation of supercomputers will likely incorporate parallel processing and will be available by 1985. Preparation for them is driving much of our research. Because of the processing power soon to be available in desktop personal workstations, we believe that the person/machine interface will undergo radical changes in this decade. This will generate significant new requirements in networking and perhaps change the way we do scientific computing. Good computer modeling includes development and analysis of mathematical models, implementation of mathematical models into software, and validation of both. These are important areas of research for any scientific computing organization.

6. Ease of Use

Ease of use includes

- specialized supercomputer software,
- software-rich systems, and
- common software across the network.

Our specialized supercomputer software includes symbolic debugging tools and highly efficient machine-language subroutines for vector operations. Also optimizing compilers are provided. Software richness is achieved by a variety of computing systems, a menu of programming languages, and a variety of vendor-supplied applications packages. Commonality is provided across all computing systems through common libraries of graphics and mathematical software and through common utilities for communication with the Common File System and with PAGES.

II. NATIONWIDE ACCESS TO THE LOS ALAMOS CENTRAL COMPUTING FACILITY

We are providing a variety of communication services into the Los Alamos Central Computing Facility. Today, we have telephone dialup access at 300/1200 bit/s to provide asynchronous terminal service. We also have telephone access at 1200/2400 bit/s for 200 UT user stations to access the NOS systems. There are several telephone dedicated leased lines that provide service. Leased line service provides 9600/56k bit/s. In addition, the telephone company can provide Direct Digital Service (DDS) at 9600/56k bit/s.

An experimental software system is running in one VAX at Los Alamos that is accessible through Telenet. That service provides 300/1200-bit/s service. We are in the process of extending Telenet access into the Computing Facility with the expectation that terminal access through Telenet will be available before the end of the summer of 1982.

The next step in increasing communications capacity is to acquire equipment to put the Central Computing Facility onto ARPANET. Purchase orders and other administrative matters are complete, and we are awaiting the delivery of the C30 processor from Bolt, Beranek & Newman to complete this task. The ARPANET service should be available at Los Alamos by the beginning of 1983.

Los Alamos is also part of the magnetic fusion energy network. We are currently served by American satellite with dual 56k channels on the magnetic fusion net.

The Department of Energy has an ambitious project to acquire and place in service wideband satellite communications nationwide. This project is called Operational Model (OPMODEL). The intent of OPMODEL is to provide wideband data service, voice, and full-motion video-conferencing capability. The first services on OPMODEL for DOE sites will become available about October 1983. We expect that there will be local telephone access into some of the OPMODEL satellite ground stations, but that is a longer term development, probably in the 1984-85 timeframe.

We believe that the integrated network that has been developed and is in place and operational at Los Alamos is among the finest in the world. The network is stable, it has a high reliability, and we can provide service as required. Of course, there is lead time between the time of request for service and the time that we can provide it. Today that lead time varies between three and six months.

III. WILL THESE FACILITIES SERVE EDUCATIONAL NEEDS?

The Los Alamos Computing Facility provides supercomputers, mass storage facilities, a menu of output options, and interactive access plus requisite support services. A recent user satisfaction survey by a professional firm shows that our computing facility is meeting the needs of the scientific user We believe that it will also meet the needs of educational institutions.

IV. HOW SHOULD THESE FACILITIES BE MANAGED AND FINANCED?

These facilities should be managed by the Computing Division of the Los Alamos National Laboratory with an advisory panel of at most six people representing the interests of the research community. Financing should be by recharge for resources consumed and services rendered.

V. PLANNING AND POLICY STRUCTURES

The Los Alamos Computing Division produces annually a two-year operational plan. The operational plan is driven by programmatic requirements, user requirements, and technology trends. Copies of the FY 82-83 Two-Year Plan are available on request from the authors and the reader is encouraged to consult Chapter 2 of it for discussion of the planning process.

Policy would have to be negotiated between Laboratory management and an administrative body representing the research community.

APPENDIX B
RATES FOR CCF SERVICES

CCF Service	FY 1983
Computing Services	
CTSS (Recorded Hour)	422.50
LTSS (Recorded Hour)	295.00
NOS (Recorded Hour)	407.50
Tape Mount (Each)	3.75
ICN Services	
300-bit/s Port/Month	90.00
1200-bit/s Port/Month	135.00
9600-bit/s Port/Month	180.00
150-kbit/s Port/Month	270.00
Intelligent Workstation/Month	270.00
200 UT Port/Month	720.00
CBT Port/Month	1080.00
56-kbit/s XNET Port/Month	1620.00
256-kbit/s XNET Port/Month	3240.00
Connect Hour (300 bit/s)	2.70
Connect Hour (1200 bit/s)	4.15
Connect Hour (9600 bit/s)	5.50
Connect Hour (150 kbit/s)	8.25
Connect Hour (200 UT)	7.20
Network Transmission (megaword)	3.75
CFS Services	
Online Access (Each)	0.32
Offline Access (Each)	3.75
Online Storage (megaword Month)	13.75
Offline Storage (megaword Month)	2.62
PAGES Services	
Printed Output (Pages)	0.20
Megabytes Processed	4.05
105-mm Fiche (Each)	4.05
35-mm Film (Frames)	0.22
16-mm Film (Frames)	0.11
Plotter (Sheets)	0.33
VMS Services	
Monthly Fee	262.00

Facility through the equipment listed in Table III (shown in Figure 1 as PAGES, Print and Graphical Express Station).

TABLE III

OUTPUT OPTIONS AND EQUIPMENT

Output Option	Equipment
8 1/2- x 11-in paper (double sided)	2 Xerox 9700s
11-in roll (electrostatic)	Versatec
36-in roll (electrostatic)	Versatec
36-in vellum (electrostatic)	Versatec
16-mm color film	FR80 film recorders
35-mm color and black and white film	FR80 film recorders
105-mm microfiche	FR80 film recorders

PAGES is accessible online from all of our computers and utilization of it ould be included in our provision of large-scale computing.

5. Support Services

Requisite support services include

- operations,
- documentation,
- education,
- consulting,
- accounting, and
- research.

We are particularly proud of the efficiency in our operations. Excepting the VAX 11/780s, all of our computers, the Common File System, and the output equipment in PAGES are operated 24 hours a day, 363 days a year by a total of 76 people. As evidenced by CFS and PAGES, part of our objective is to automate where possible. Thus, our Facility for Operations Control and Utilization Statistics (FOCUS in Figure 1) is a recently added node to our network from which we load level production jobs across the supercomputers. Because of automation, we anticipate no growth in our operation staff in the next few years despite plans to significantly increase our total computing capacity.

A P P E N D I X I I

APPLICATIONS OF SUPERCOMPUTING

Supercomputers and the Equations of State of a Classical One Component Plasma

Bill Buzbee

Following the pioneering work of Brush, Hansen, and Teller(1), scientists have been attempting to compute the equation-of-state (EOS) of a classical one-component plasma (OCP). OCP is an idealized system of ions immersed in a uniform sea of electrons such that the whole system is electrically neutral. Though OCP is a "simple material", if its EOS could be computed, a framework would be available from which to address the EOS of more complicated substances. Prior to the advent of the Cray-1, scientists had not been able to compute the EOS of OCP throughout the energy interval of interest.

The EOS can be calculated with a Monte Carlo technique. The density and temperature in the system are fixed, and the system energy is computed. Then a particle (ion) is selected at random and subjected to a small, random perturbation (movement). The system energy is recalculated, and the move is accepted if the energy decreases. If the energy increases, the move is accepted with a probability described by the Boltzmann distribution. Now another particle is chosen at random, ... ad infinitum. Once the energy stabilizes, a point of the EOS is in hand. New values of density and temperature are chosen, etc.

Calculation of the system energy requires evaluation of an interparticle potential. The potential is not available in closed (finite) form, so an approximation must be made and the approximation must be computationally affordable on the computer. With the availability of the Cray-1, scientists (2) developed an improved (more complex) approximation to the interparticle potential. By using very efficient vector algorithms and software developed under the Applied Mathematics Program of DOE's Office of Basic Energy Science, it was possible for the Cray-1 to execute the new calculation at about 90 million floating point operations per second. Even so, some seven hours are required to carry out the EOS computation. Most important, for the first time, the EOS for a OCP was accurately calculated throughout the interval of interest. Further, the melting temperature was found to be 20% lower than the widely accepted previous estimates. This advance is a direct result of the high level of performance available from the Cray-1.

This example also illustrates an important point about algorithms for vector computers. Throughout the era of electronic computation, our research has sought to develop algorithms that have a minimum number of arithmetic operations. Indeed, this is the correct approach for scalar computers, but it is not necessarily the correct approach for vector computers. Evaluation of the interparticle potential approximation requires evaluation of the complimentary error function. The complimentary error function contains an integral over a semi-infinite interval. Thus it must be approximated for efficient computation on a computer. On sequential (scalar) computers, scientists use a very simple and sensible approximation ---, namely, they tabulate the function and evaluate it via table lookup and interpolation. Unfortunately, table lookup is inefficient on a vector computer. A better approach on a vector machine is to use a polynomial approximation thereby eliminating table lookup. This approach involves many more arithmetic operations than the tabular approach, but they are fully vectorizable and can be executed at a very high rate. So in this case, the best algorithm for a vector computer is quite different from that for a conventional computer.

This example of computing the equation-of-state for a classical one-component plasma illustrates how increases in supercomputer performance make possible advances in fundamental knowledge. It also shows how changes in computer architecture can impact computer simulations and the payoff of research to find suitable algorithms for them.

1. S. G. Brush, H. L. Sahlin, and E. Teller, J. Chem. Phys. 45, 2102 (1966).

2. W. L. Slattery, G. D. Doolan, and H. E. DeWitt, Physical Rev A 21 (#6), 2087 (1980).

Aerospace Structural Design and Supercomputers

Richard H. Gallagher

The products of the aerospace industry -- commercial and military aircraft and aerospace vehicles -- represent extremely challenging structural design problems. They must be of the lightest possible weight, perform safely, and should strive for the lowest cost of manufacture. They are of complicated form and in each succeeding decade they are planned to operate in more severe environments and to incorporate new materials and design concepts. Because these challenges have been met up until now, the United States has been and continues to be foremost in the aerospace industry and that industry is a major factor in our economy and balance of payments.

From the onset of electronic computation, the aerospace design community has employed the largest computers available to commerical enterprise at any given time. At first, in the early 1950's, the existing computers were applied to static analysis in terms of a few unknowns. The scale of analyses advanced in tandem with the development of Class II - V computers. Even today the unknowns of the real structure outdistance the capacities of class V machines. If design optimization, computer graphics, and dynamics phenomena are treated, then there must be tradeoffs such that the computer description falls substantially short of the real structure. Class VI machines, which are imminent participants in the aerospace structural design process, will make inroads into this discrepancy but will not resolve it fully.

The above remarks refer to structural design in isolation. Major strides are being made in design for aerodynamics, especially in the definition of shapes for drag reduction. As noted in Section III of the report, the computations resources that are needed for more effective aerodynamic design are 100 times greater that those of Class VI machines. Inevitably, aeroelastic interaction must be taken into account, wherein the structural and aerodynamic behaviors are dealt with in an integrated computational exercise. The computational needs will tax even supercomputers.

The requirements increased fuel efficiency, reduction of material weight, greater range, and reduced manufacturing costs are quite evident. The international aerospace industry is aware that these goals can be achieved through realizable, expanded computer power. The U.S. segment of that industry is a prime candidate for loss of its pre-eminent position if the computational tools available to it do not move ahead with the sate-of-the art.

To date, Class VI machines have not been available to a broad spectrum of engineering researchers, particularly academic researchers. Despite this, these machines have already had a significant impact on research in several engineering disciplines. The problems tackled so far are characterized less by the particular discipline than by the mathematical formulations involved. Broadly speaking, the following appear to be the most important:

(1) Solution of linear and nonlinear PDE describing transient (unsteady state, dynamic) behavior using finite difference methods. Typically, the solutions involve of the order of a million nodes (say 100 per space dimension). Major applications are the solution of problems in fluid mechanics (e.g., Navier-Stokes equations) in aeronautical engineering, heat transfer in mechanical engineering, and flow through porous media (reservoir simulation) in chemical and petroleum engineering.

(2) Solution of PDE describing static (steady-state) behavior using finite element methods. Typical problems involve of the order of 10,000 element nodes and direct solution methods are used to solve the nodal equations. Principal applications have been in the structures area (particularly for aerospace structures). Some fluid mechanics and heat transfer problems have also been solved using this approach.

(3) Monte-Carlo methods involving of the order of 1000 particles. A major application involves solution of the radiation transport equation in nuclear reactor shielding analyses.

(4) Real time image processing. Extremely fast filtering, interpolation, and back-projection calculations are required for analysis of high-resolution video signals, such as are generated in tomographic equipment.

There are problems in each of these areas that could profitably use machines 100 to 1000 times faster that the current Class VI machines. For example, the finite element method could be used to solve transient problems involving perhaps 100,000 element nodes. This would require the solution of thousands of simultaneous ordinary differential equations in addition to the direct solution of large numbers of (typically linear) equations. Faster, larger machines would allow the use of much larger sample sizes in Monte-Carlo simulations with attendant improvements in the accuracy of problem statistics. Many of the image processing problems, particularly those associated with real-time nondestructuve testing, will be difficult to solve without much faster machines.

Probably the greatest impact that the supercomputer of tomorrow will have on engineering will be the solution of systems problems that simply cannot be tackled at all with current equipment. For example, in aerospace engineering, the fluid mechanical, structural, and control systems analyses are treated in isolation. In the future, models of the integrated systems will be studied, and will undoubtedly lead to far better, more efficient aerospace vehicles.

Similarly, in chemical engineering, it should be possible to study the dynamic behavior of entire chemical processing complexes (currently almost all chemical systems designs are based on steady-state models and any associated control systems are designed using much simplified dynamic models of plant subsystems) requiring the solution of hundreds of thousands of simultaneous algebraic and differential equations. For the first time, it should be possible to use nonlinear programming algorithms to optimize the design and operation of such nonlinear systems, which incorporate a significant number of design variables. Automatic process synthesis (given only specifications for raw materials and desired products) algorithms may well be feasible, with sufficient computing power.

Undoutedly, similar examples involving the design, simulation, optimization, and even operation of complex systems can be found in virtually any engineering discipline.

SOME EXAMPLES OF SUPER COMPUTER PROJECTS

L. L. SMARR

UNIVERSITY OF ILLINOIS

* High Resolution Gas Dynamics - The large memory and high speed of supercomputers allows one to simulate dynamic gas flows with high resolution. Astrophysical examples are the flows that occur in quasars and radio galaxies. These can involve the central powerhouse of active galactic nuclei - nonspherical accretion onto supermassive black holes [J. Hawley and L. Smarr (Univ. of Ill.) with J. R. Wilson (LLL) - calculations performed on CRAY-1 at LLL]. Or they can involve the formation, propagation, and stability of supersonic and relativistic jets in these objects [calculations by MUNAC (Munich Numerical Astrophysics Coalition) - M. Norman (MPI), K.-H. Winkler (MPI) and L. Smarr (Univ. of Ill.) on the MPI CRAY-1 and by J. R. Wilson (LLL) on the LLL 7600]. The use of large numbers of grid points ($\sim 5 \times 10^4$) allow theoretical resolution of the flows comparable to the observational resolution of the Very Large Array of radio telescopes or the Einstein orbiting x-ray telescope.

* Numerical Relativity - Here one solves the fully time-dependent Einstein equations of general relativity to study the strongly curved spacetime near black holes or in the early universe. Again, many grid points ($> 10^4$) are needed to cover both regions near the strong gravitational field as well as far away. Recent projects computed include formation of nonspherical black holes and neutron stars, generation of gravitational waves, and distribution of the dark matter in the universe [J. Centrella (Univ. of Ill.), L. Smarr (Univ. of Ill.), P. Dykema (LLL), C. Evans (Univ. of Texas), and J. R. Wilson (LLL) - calculations on LLL CRAY-1]. Detailed calculations of these nonlinear processes are needed by the experimentalists designing gravitational wave detectors.

* Solving Quantum Chromodynamics - Exciting progress is being made in the program, started by Kenneth Wilson and Alexander Polyakov, to gain insight into the nature of the strong interactions by applying statistical methods to a formulation of quantum chromodynamics on a discrete spacetime lattice. For example, a recent breakthrough investigating the range of forces responsible for breaking chiral symmetry using a lattice of fermions was carried out numerically at Los Alamos using CDC 7600's [J. Kogut, M. Stone, and H.W. Wyld (all of Univ. of Ill.), J. Shigemitsu (Brown), S.H. Shenker (U of CA) and D.K. Sinclair (Stanford]. The work must be limited to spacetime lattices of 5 x 5 x 5 x 10 points because of memory and spacetime speed limitations. Clearly a supercomputer would enable much more detailed work to be done immediately.

* Severe Storms, Tornados, and Downbursts - The basic characteristic of storms which spawn tornados can now be simulated using CRAY-1 level computers [e.g. R. Wilhelmson (Univ. of Ill.) and others - calculations on NCAR CRAY]. The development, 3-dimensional structure, and movement of these storms can be modeled and compared with observed wind fields from Doppler radar. With enough supercomputer time, these studies can be extended to understand the detailed formation of tornados, hazardous downbursts, hail growth, and damaging surface winds associated with convective storms.

* Intramolecular Dynamics Calculations - Realistic calculations of intramolecular energy flow in even moderate size molecules (e.g. five to eight atoms) requires large computers. Using small computers (i.e. CYBER 175 class) many groups have done classical calculations comprising a few trajectories at unrealistically high energies. Some work was done in the mid 1970's on the largest available computer [Illiac IV by J.D. McDonald (Univ. of Ill.)]. This remains the only work with a statistically significant number of trajectories.

What is really needed are quantum calculations on intramolecular energy transfer. These calculations are completely hopeless without the latest generation supercomputers. Even on these computers they count as large scale work.

EXAMPLES AND NEEDS OF SUPERCOMPUTERS

K. Wilson

Cornell University

Thermodynamic Properties of Fluids

A research group consisting of Keith Gubbins, William Street, and S. Thompson, in the Chemical Engineering department at Cornell, use both a shared Floating Point Systems Array Processor and their own minicomputer for simulations of liquids at the molecular level. At the present time these simulations can only supplement information obtained from analytic theory and experiment, due to the limited sample sizes (numbers of molecules) that can be studied in a practical simulation. Availability of computers thousands of times more powerful than current supercomputers would encourage this group to give heavy emphasis to the determination of thermodynamic properties of fluids directly from molecular potentials via computer simulation. Such a capability would be of enormous significance for many engineering applications.

Theory of Nuclear Forces

One of the most spectacular developments of basic theoretical research in the 1960's was the discovery of gauge theories providing a unified description of nuclear interactions, weak interactions and electromagnetic interactions. The Glashow-Weinberg-Salam gauge theories of electromagnetic and weak interactions are well understood theoretically and have received impressive experimental confirmation. In contrast, all that can be said so far about the gauge theory of nuclear (strong) interactions--called Quantum Chomodynamics--is that it is not obviously wrong. (It is however virtually the only theory of strong interactions with this property.) The reason for the poorly understood state of Quantum Chromodynamics is theoretical: no analytic techniques are known for solving this theory, not even for approximately solving it. A major international effort has been underway for several years to try to solve quantum chromodynamics numerically, to obtain its predictions for fundamental quantities such as the masses and magnetic moments of the proton and neutron. Theoretical groups in the U.S. at M.I.T., Cornell, Brookhaven National Laboratory, Princeton, Columbia, Cal. Tech., the Institute for Theoretical Physics at Santa Barbara, Stanford Linear Accelerator Center, University of Chicago, University of Illinois, University of Indiana, etc. have engaged in this effort; mostly using borrowed computer time on minicomputers and array processors.[1] Present day computers have been grossly inadequate for a complete, reliable computation; the current results involve simplifying assumptions and are unreliable even after the simplifications.

As a result of computations done so far, it is clear that a complete, creditable computation will stretch the very limits of future computing technology. Such a computation will probably require thousands or millions of specially designed, very high speed processors embedded in a highly parallel network for data interchange. (The basic computational problem here is that sparse matrices of size roughly $10^6 x 10^6$ must be inverted at every step of a Monte Carlo numerical integration; the Monte Carlo calculation involves of order 10^6 integration variables; the answer must give high accuracy for some very small long range correlation functions.) While many of the quantities to be computed are already known experimentally, the success of the calculation is of fundamental importance to the understanding of nuclear forces and will also lay the basis for new, nonperturbative studies of particle interactions at super high energies. More generally the technology established for this calculation is likely to lead to equally important computations in many other areas of basic and applied scientific research.

The Renormalization Group and Statistical Mechanics

A major breakthrough in theoretical research in the last two decades was the discovery of a new framework for understanding a whole class of previously unsolved problems.[2] The problems addressed by renormalization group techniques includes critical phenomena (the onset of phase transitions), quantum field theory, polymer problems, various problems in solid state physics, and turbulence in classical fluid flow. Analytic expansion methods were found to make realistic renormalization group analyses in special cases of some of the above problems--for example, an expansion about four space dimensions for critical phenomena. Unfortunately only a minute fraction of the problems addressed by the renormalization group framework can be solved with current analytic renormalization group techniques. For example, relatively little progress has been made in its application to turbulence. Attention is now turning to the use of computer simulation to support the renormalization group approach: the "Monte Carlo Renormalization Group method" of Swendsen and Wilson has provided accurate numerical results for some models of specific critical phenomena on surfaces that were previously intractable.[3] However, full flowering of the renormalization group approach requires vastly greater computing power than is available today. Even for the prototype system for all renormalization group studies--the three dimensional Ising model of a ferromagnetic transition--current Renormalization group research is blocked by inadequate computing power.

Wilson and Swendsen have already spent hundreds of hours of computing time on an array processor in an effort to study the numerical truncation errors in the Monte Carlo Renormalization Group technique. Unfortunately, this effort has failed so far; the calculations to date show that a thousand or even million fold increase in computing power will be required to identify sources of error and make sure each error decreases at the expected rate as the truncations are removed. This effort is critical to establishing the credibility of numerical Renormalization Group methods and encouraging its use in many applications.

The Ising model is best handled through special purpose VLSI chips. An example of such a chip, with spectacular performance, was partially designed at the Institute for Theoretical Physics at Santa Barbara[4] using the Cal. Tech. chip design system. To make full use of such a chip requires thousands of chips be operated in parallel with high speed data paths linking the chips, and parallel general purpose computers to carry out supporting computations. It has been impossible under present funding patterns to reproduce the chip or set up this support framework; as a result the Santa Barbara chip has been more of a curiosity than a productive scientific instrument. Experience with a network of Ising chips would lay the groundwork for a far more multi-purpose, highly parallel, system needed for the nuclear force calculations that could handle a large range of statistical mechanical simulations, predicting the properties of bulk matter from basic interactions of molecular constituents. In constrast, some two dimensional statistical mechanical problems are already feasible; researchers at M.I.T. and the University of Georgia have carried out Monte Carlo Renormalization Group computations on two dimensional statistical mechanical models (in collaboration with Swendsen)[5] with considerable success.

1. See, e.g. H. Hamber, Brookhaven preprint (1982) for a recent list of references.

2. See, e.g. K. Wilson in Scientific American, Vol. 241, No. 2, p.158.

3. See D. P. Landau and R. H. Swendsen, Phys. Rev. Lett. 46, 1437 (1981).

4. R. B. Pearson, J. L. Richardson, and D. L. Toussaint, Santa Barbara ITP report ITP-81-139 (unpublished).

5. See Ref. 3 and R. H. Swendsen, P. Andelman, and A. N. Berker, Phys. Rev. B24, 6732 (1981).

Role and Needs for Large Scale Computing for Theoretical Physics

Kenneth G. Wilson

Cornell University

I. Statement of Need for Theoretical Physics as a Whole

At the present time the principal trend in theoretical physics is the broadening of the scope (in practice) of theoretical ideas by the inclusion of ever more complex problems within the theorists' purview. This trend is taking place uniformly throughout every subfield of theory. This broadening is taking place through re-examination of old problems (such as turbulence), the development of new areas (such as supersymmetry or grand unified theories of elementary particle physics) or the recognition of fundamental questions in applied areas such as metallurgy. The current trend is making possible for greater contact between fundamental theory and complex practical applications than ever before, and this is reflected in a new interest in hiring of theorists by industry, for example at Schlumberger-Doll Research and at Exxon.

In consequence of the move towards complexity, theorists have outrun the capabilities of traditional analytic theoretical methods. Complications dealt with today include multiple degrees of freedom, irregular geometries, multiple length scales, and random irregularities, any one of which can defeat analytic techniques. Modern theory has conceptual, qualitative methods to treat complexity: the "renormalization group" is one very general framework that has emerged over the last thirty years with an enormous sweep. However, the qualitative concepts such as the renormalization group provides must be backed up by more quantitative results. In many cases, numerical simulations are the only approach available for providing quantitative analysis.

Unfortunately, the problems which are too complicated for analytic analysis can place very severe demands on a numerical approach. For example, each new degree of freedom that is included in a numerical treatment can easily cause a factor of 100 or 1000 increase in computing requirements (both for cycles and memory); a standard example is the change from a nonrotating, spherically symmetric star to a rotating star where the angle to the rotation axis is an extra degree of freedom. Problems with randomness, thermal fluctuations or quantum fluctuations typically have thousands or millions of degrees of freedom, which can only be treated by statistical Monte Carlo or statistical dynamics methods. Computing time needed for these problems increases as the square of the accuracy desired; the isolation of small errors and establishment of numerical reliability for these statistical calculations demands very high accuracy and hence very painful computing requirements.

The building of a simulation capability for a complex problem can take many years of effort. In early stages there comes the pursuit of various algorithms to achieve both convergence and efficiency, along with supporting studies to make sure all necessary physical principles have been correctly taken into account in the numerical procedures. At later stages come the study of errors; in a complex situation there can be many sources of error with lots of cancellations. In this case naive estimates based for example on two calculations with modestly different levels of truncation can be totally misleading and years of lengthy calculations and careful study may be required before the numerical simulation can be performed with a reliable error estimate. The computing resources needed for all stages of a simulating effort usually vastly outruns (by a factor of 100 to 1,000,000 or more) the computing time needed for a single production run once a reliable production code is established.

A crucial practical aspect of large scale simulation is that when the computing demands of a simulation increase, the increase is not by a factor of 2 or less: the increase is usually a factor of 10 or 100 or more. For example, simulations often involve a multidimensional grid approximating a continuum; to achieve a better approximation a minimal improvement involves doubling the number of grid points along each dimension. In a three dimensional grid this means a factor of eight increase in the number of grid points overall. Typically, further complications develop so that the increased computing demand is considerably more than a factor of eight.

In conclusion, it is normal and reasonable that the computing demands for scientific simulation are prodigious, even by comparison with today's most powerful supercomputers. In fact, the whole area of scientific simulation in support of modern theoretical physics has barely been scratched; providing full access to theorists on the best of today's computers, can only be viewed as the beginning of many further stages of providing new levels of computing capability in support of ever more complex simulations. The complexity in simulation that can be handled numerically must keep pace with the growth in complexity that theorists are able to handle conceptually.

It is of major economic importance to the U.S. that the current broadening trend in theoretical physics be encouraged, and in particular, that the computing support problem be rectified. The high technology economy itself is facing much more complex problems than ever before. Oil used to come out of gushers; now it must be coaxed out and in future much more recalcitrant raw materials will replace oil. This fact underlies the new interest in hiring of theorists by the oil industry. Simple industrial materials like steel are being replaced by ever wider ranges of composites; mastering their properties and designing new ones quickly (to stay abreast of competition from abroad) can be helped by theoretical studies, if these studies can be comprehensive and accurate. Simulations of products and processes throughout industry can be cheaper and faster than experimentation and prototyping and provide more opportunity for optimization or for meeting conflicting requirements. In all these applications new problems requiring the attention of theorists will arise; theory will grow and grow in importance as the capabilities of theory become more powerful. It would be folly to block the expanding capability of theory by continued denial of access to computing.

II. Practical Problems of Computing Support

There are three issues which I think are not adequately dealt with in the body of the workshop report.

1) Obtaining data that establish in detail the scientific importance and quality of scientific simulations.

2) Providing adequate capacity despite rapidly growing demands for computing.

3) Management and Manpower issues.

1) None of the committees I have served on, including this workshop,

have solved the problem of documenting the importance of computing support for theorists. Members bring lists of needs within their fields, but these lists seldom provide any overview of the importance of the needs. There must be a more ongoing effort to collect data establishing the scientific importance of theorists' computing needs, especially given the skepticism of many eminent scientists towards large scale computing. This should be done by collecting reports on computing needs on a continuing basis from workshops and conferences and other meeting places where the impact of scientific computing on science is manifest and emphasis can be given easily to computations with the greatest impact.

2) Theorists can saturate any computer, no matter how powerful with their simulation programs; once a computer is saturated, computing job queues lengthen and user productivity diminishes, which leads to pressure for more powerful computers. This fact makes it difficult to determine the level of computing capacity to provide. For example, one may ask: why bother providing $100 million dollars worth of capacity if, in the end, as many complaints will occur as if only $10 million dollars of capacity is provided?

I propose the following strategy. The computing capacity made available to theory computing should be divided into two parts. A base level of computing support should be established, which every theorist is entitled to by virtue of being a qualified theorist. The base level would be supplied by computers which are cheap — so cheap that as many of them would be supplied as necessary to prevent over-saturation of them. At the present time a national program of reasonable cost could supply enough super minicomputers and attached processors to keep even with any level of demand that theorists could generate on them — if a particular theory group saturates one attached processor with legitimate research and teaching use they would be given one or two or four more, as needed. This procedure has the enormous benefit that computing time does not have to be allocated at the base level. As a result, students are free to try things out on base level systems without fear of exhausting an entire research allocation (this experimentation capability is exceedingly important for effective student training in computational physics) and faculty and management time is not wasted on the allocation process. There should however be flexibility in the provision of base level computing — for example where existing research areas are already extensively tied to supercomputers by existing software such as the Plasma Fusion energy system based at Livermore, there is no point in forcing changes.

Many computing projects, when they are first begun, can be targeted equally effectively for a supercomputer or for an attached processor despite the factor of ten or more advantage in speed of the supercomputer. The reason is that when the factor of ten in computing power is converted into a grid spacing or the potential accuracy of a statistical simulation, it is much less of a difference. Unfortunately, once the decision has been made to use the supercomputer level of power it is much more difficult for a user to retrospectively shrink a program from the supercomputer to a less powerful machine. It is important that theorists not yet using supercomputers be pointed towards attached processors from the start — if this is done there will be very little loss in scientific productivity due to using the much cheaper attached processors instead of supercomputers for the base level computing.

There is still an extremely important role for supercomputers. I would reserve them for a few key programs that require the entire capacity of a supercomputer. For example, a Cray X-MP could be 40 times faster than an Attached processor. This speed is most useful if there is just one program using the entire Cray, say for a week. These wholesale lots of supercomputer time would have to be allocated; a possible mode of operation would be as follows. Programs accepted for supercomputer runs would have to meet three criteria:

a) only programs ready to run on the supercomputer would be considered, and only if all necessary preparations that could be done at the based level are complete.

b) any program accepted for runs on the supercomputer should have a reasonable chance of producing truly useful results in the allotted time

c) any program accepted should have scientific priority; for example, a workshop might establish that a particular simulation is blocking progress and needs to be completed in preference to other projects in the same area.

These criteria should be kept strict enough to avoid saturation of available supercomputers, so that once a program is accepted it is scheduled quickly, while the whole problem is still fresh in the user's mind. The advantage of this approach is that the scientifically most important computations can be given the factor of 100 or more increase in computing power above base level that they often require for satisfactory completion.

No method of supplying computing capacity will work unless computing capabilities at the base level increase rapidly. Simulation projects, once started require growing capabilities especially to allow determination of errors; new levels of capability must be reached so that new levels of complexity can begin to be explored by simulation. The growth of theoretical physics will be stunted if computer capabilities do not grow.

The workshop report proposes a major effort in computer development. I heartily endorse that effort; it is so important that I believe it should receive roughly half of all funds provided for the large scale scientific simulation program. The computer development effort should have two parts. One part is to develop extremely powerful systems that would be vast upgrades for existing supercomputers. The other part should be to expand the capability of the base level of computing supplied for scientific use. The incredibly rapidly growing capabilities of VLSI technology should be used as far as possible for base level systems. The reason for this is that by heavy investment in state-of-the-art VLSI design and equally heavy investment in large scale manufacturing facilities, the unit costs of VLSI components can be reduced to almost 0, as the plunging prices of memory chips illustrate. For the base level of computing the most urgent need is to reduce the unit cost of single systems so that there is a maximum ability to keep on providing additional systems to the scientific community to prevent saturation and attendant loss of productivity.

The development of very powerful systems that are factors of 100 and

1000 and even 10^6 above the base level are equally important; in the plan I outline these systems would be used to enable completion of only the most urgent scientific simulations, even when factors of 1000 or 10^6 are needed to finish these simulations with adequate reliability. As scientific simulation grows generally in importance both for basic research and practical applications, the most urgent simulations will become far more important than the cost of the computers they require.

III. Management and Manpower Issues

The draft workshop report does not address two important practical issues. One is the problem of computer management, the other is the impact of a major computing support program on the national technical manpower pool in computing.

Good medium, and high level management for computing systems is extremely scarce, as most university people know from local experience. The most difficult role in a computing center is that of the middle level manager, who must cope with major problems coming from both above and below. In addition, whenever any large community of users and support personnel is organized, the community as a whole tends to become very resistant to change. Over the next ten years there are likely to be extraordinary changes taking place — from vector supercomputers to parallel systems to a great variety of special purpose systems, from FORTRAN to a great variety of higher level programming systems. Any program to deal with large scale scientific computing needs must continuously adapt and respond to these changes.

In the plan I presented, the base level of computing would be provided mostly by self-contained systems serving small, localized theory communities, often within a single university department. The only support staff required is one system manager, perhaps aided by some undergraduate work-study students. This system manager's task is easily performed by eager young people: not much prior experience is required. These small units can adapt to change fairly easily when it is necessary to do so.

The supercomputers, in the plan I described, would also have small user communities; these systems should also be easy and inexpensive to manage and to adapt to change.

Strong management would be required, I think, for the following areas:

a) bringing up initial production versions of major new computing systems — this involves innumerable software and consulting headaches which makes good management essential.

b) a national high performance network

c) major archives of a permanent or semi-permanent nature

d) very major software packages receiving widespread use.

I strongly urge that centralized management skills available to large scale scientific computing be targeted whenever possible to the areas of greatest need; in particular the base level of computing should be distributed and informally managed to the extent that local conditions permit.

Any major program involving computing must be scrutinized for its impact on the nation's trained manpower pool, since any reduction in this pool can result in reduced economic activity, especially within the computing industry itself. The professional support personnel used as system managers or in central computing centers are drained from the overall manpower pool. The impact will be less if these personnel move on to industry fairly quickly, but in the meantime train their successors. In contrast, undergraduate work-study programs, higher level software projects which help support computer science graduate students, and supported computer development projects in computer science departments help increase the manpower pool. In addition a basic function of the large scale scientific computing support effort is the training of graduate students in computational aspects of science.

A P P E N D I X I I I

POSITION PAPERS

Japan's Initiatives in Large Scale Computing

Lawrence A. Lee
National Science Foundation

1. Japanese Supercomputer Project

In July of 1980 Japan announced a supercomputer project aimed at creating a machine by 1989 capable of speeds at least three orders of magnitude higher than currently attainable, that is, a 10 gigaflop computer. Six major Japanese computer vendors (Fujitsu, Hitachi, NEC, Mitsubishi, Oki, and Toshiba), led by the government's Electro-Technical Laboratory (ETL), have joined in this project entitled National Super-Speed Computer Project (NSCP). The joint effort is organized as the Scientific Computer Research Association (SCRA). The project, started in January 1982, is scheduled for completion by 1989 and is funded by the Ministry of International Technology and Industry (MITI), with additional support from each of the six vendors. Total funding for this project is expected to be about $200M.

The research is to be carried out in two subprograms simultaneously:

(a) Development of new types of high-speed logic and memory elements. Researchers will consider a variety of possible alternatives to silicon semiconductor technology--including Josephson Junction elements, High Electron-mobility Transistors (HEMT) and Gallium Arsenide Field-effect Transistors--with the goal of producing processors with speeds in the 40-100 mflops range.

(b) Development of parallel processing systems incorporating arrays of as many as a thousand processor elements.

Finally, the results of the two subprograms are to be combined, resulting in large-scale parallel systems which use 40-mflops to function at speeds of 10 gflops or even faster.

For the sake of comparison, the following are operating speeds and in-service dates of several existing and planned U. S. supercomputers:

The CRAY-1: 88 mflops (1976)
The CDC Cyber 205: 400 Mflops (early 1981)
The CRAY-2: about 800 mflops (late 1983 or early 1984)

2. Japanese 5th Generation Computers

MITI is funding another project called the National Fifth-Generation Computer Project; this project will start in April 1982 and run for at least ten years. MITI and vendors will provide some $500M for this project. Whereas the NSCP is based on extensions of current technology, the fifth-Generation Project is to be a revolutionary approach to computer design based on artificial intelligence. The overall project goal is to design information processing systems to deal with the basic social problems Japan foresees for itself in the 1990's (such as low productivity in primary and tertiary industries; international competition; energy and resource shortages; and a rapidly aging population).

Among the specific applications set as eventual goals for the project are:

(1) A machine translation system automating 90% of natural-language translation (e.g., between Japanese and English).

(2) A consultation system capable of functioning as a small interactive reference library for workers in various specialized fields.

(3) A system capable of automatic or nearly-automatic creation of software from specified requirements, with little or no need for computer sophistication on the part of the user.

This is a highly ambitious attempt to change the whole domain of operation of computers in society from "Data"--abstract bits of information stored in the form of electrical impulses and manipulated by machines without regard for what they might signify--to "Knowledge"--something much more like people's everyday use of the word 'information,' and much more accessible to people who do not have special training in computer/science.

The project is described in terms of certain "key words," including: artificial intelligence, parallel processing, inference mechanization, knowledge-base, relational data-base, relational algebra, software-development systems, VLSI-CAD, machine translation, and Prolog-like languages.

One approach to building such new machines is to carry on the work of the past three decades in increasing storage capacity and operating speed. The fifth generation project in its final approved form contains little if any work of this kind; rather, administrative mechanisms will be set up to allow the fifth generation researchers to take advantage of exploitable new hardware developments as they are achieved in, for example, the supercomputer project, and the next generation basic industries semiconductor project. The hardware approach that will be taken in the fifth generation project will focus on exploring new computer architectures designed specifically for efficient operation of the "knowledge base" fifth generation software.

3. Japan's Vector Processors

In addition to the supercomputer and 5th generation projects, each of the vendors has one or more internal projects that will affect future supercomputers in the areas of architecture, software, algorithms, and device technology. The company project that appears closest to completion is the Fujitsu Vector Processor (VP); this computer is based largely on the same technology as the Fujitsu M380/382 (the M380/382 is to be delivered this year and will exceed the performance of the announced IBM 3081 products by about a factor of 1.5). The Fujitsu Vector Processor is intended to compete with the Cray-1 and will (if all goes well for Fujitsu) exceed the performance of the Cray-1 by up to a factor of 5.

WORKSHOP ON LARGE SCALE COMPUTING FOR SCIENCE AND ENGINEERING

POSITION PAPER

PREPARED BY: Robert G. Gillespie
University of Washington

June 18, 1982

Introduction

The following issues dealing with the Workshop On Large Scale Computing for Science and Engineering were developed during discussion with a group of faculty members from the University of Washington. The disciplines involved included Chemistry, Physics, Atmospheric Sciences, Oceanography and Computer Science.

Issues

1. ACCESS

A strategic option to address the fundamental question of providing access to large scale computational facilities would be the linkage of the existing facilities with major research universities. It was assumed that high band width communication facilities, such as can be provided from satellite stations, would be provided. The establishment of these nodes would significantly aid the resource sharing of the class VI facilities and also would have other secondary effects by making the communication effort transparent. Essentially, the nodes could reduce the friction that limits use of existing information resources. This option should be viewed as a valuable alternative when compared to adding one more CLass VI computer.

2. COMMUNITIES OF INTEREST

Scientific communities tend to be organized around discipline areas. The opportunities for resource sharing---ideas, software, databases---along with improved effectiveness associated with computer support for reasonably homogeneous communities, should command attention when options for approach are considered. The most successful of the shared scientific large scale systems have been organized along discipline lines. This approach simplifies the allocation because, with a homogeneous community, allocation decisions can be made by members of the scientific community.

A clearly perceived need should exist before a new facility if funded and built. This is another reason that discipline oriented centers (e.g. NCAR) have been successful. While other arrangements such as regional centers might work effectively, the need should be clearly seen.

3. FUNDING

Funding for large scale computing should not be approached by choosing only to provide additional funds for grants. This choice risks fragmenting the resources available nationally as well as converting the problem of providing the computing facilities to a strictly entrepreneurial model.

4. PILOT PROJECT RISKS

Pilot project approaches, are attractive for many projects, and can be an effective way to develop support. The use of large scale computing systems involves more resources than the funding or access to successfully achieve effective use. The investigator must undertake personal investment in learning new systems, redeveloping existing codes which took, in some cases, years to stabilize. Thus a pilot (sometimes a synonym for underfunded) project may never achieve the success. Some of the elements of this concern were generated by observations of the experience with the National Resource for Computational Chemistry. Thus any programs suggested that have a pilot phase should be carefully evaluated to insure that they will not be self fulfilling prophesies of failure.

5. FEDERAL DISINCENTIVE PROGRAMS

The administration of implicit Federal policies affects the strategy and approaches to computing at universities by individual investigators. The lack of a funding mechanism for long term, ongoing computer use and operating costs tends to push acquisition of small and intermediate computer systems. Essentially the capital funds in a grant or contract are a mechanism for prepaying for computational resources. Unfortunately, this suboptimization does not necessarily either meet investigator's goals or draw together Federal resources where resource sharing of facilities would be most appropriate.

6.PROCUREMENT CYCLES AND TECHNOLOGICAL CHANGE

The Workshop should consider the long Federal procurement cycles when alternatives for acquisition of very large scale systems are considered. Cycles of over five years have been observed and the present state of obsolescence at some Federal Laboratories could mean the acquisition of out of date equipment.

THE SUPERCOMPUTER FAMINE IN AMERICAN UNIVERSITIES

by

Larry L. Smarr

Associate Professor

Departments of Astronomy and Physics

University of Illinois at Champaign-Urbana

I. Supercomputers

By "supercomputer" I mean the computer with the largest memory and the fastest CPU currently available commercially. At the moment (September 1982), the two major American supercomputers are the CRAY-1 and the CYBER 205. Such machines possess 1-4 million 64-bit words in the central memory with a relative capacity of more than 4 CDC 7600's (the equivalent of 6 CYBER 175's). [See Appendix B for rough conversion figures.]

The CRAY X-MP machine, which becomes available in one year, will yield a throughput increase of ~ 3-5 over the CRAY-1. It also offers a 10 gigabits/sec transfer rate to 32 million (64-bit) words of external solid state memory, in addition to the 4 million words in central memory. This would make it equivalent to 30-50 CYBER 175s. The above figures are used to "set the scale" for what universities should be thinking about in the next few years, and do not constitute an endorsement of any manufacturer.

The current distribution of supercomputers (Appendix A) virtually excludes American universities. As of mid-1982 there are ~ 50 supercomputers in the world. In this country the national DOE weapons labs have the largest concentrations of them. Lawrence Livermore National Laboratory (LLNL) and Los Alamos Scientific Laboratory (LASL) each have four CRAY-1's, or 1/5 of the world supply. Specialized scientific groups have access to CRAY's at National Centers, e.g. the National Center for Atmospheric Research (NCAR) and the Magnetic Fusion Energy Center (MFECC). Finally, a number of major corporations have acquired CRAY's, e.g. Bell Labs and Exxon. European scientific centers have also begun using supercomputers, e.g. the Max-Planck Insitute (MPI) for Physics and Astrophysics in Garching-bei-Munchen, West Germany.

The first arrival of supercomputers at American universities has occurred in the last year. Colorado State University and Purdue University have each bought a CYBER-205, while the University of Minnesota has obtained a CRAY-1. At least a dozen universities are presently considering whether to purchase supercomputers either individually or in regional groups.

II. Flight From the Universities

The drought of computing power in the universities over the last five years has resulted in a a major drain of highly talented researchers from the universities to the sites with supercomputers. I perceive two stages in this process: first, a flight to the weapons labs (LLNL and LASL), and second, a flight to foreign computing centers. Most of this section is anecdotal knowledge and represents my narrow sphere of expertise. I imagine similar stories can be told by others in different areas of research.

Much of the work in computational astrophysics has been done at LLNL and LASL during the last decade. Areas of numerical research include: supernovae, general relativity (black holes, neutron stars, gravitational waves, cosmology), nonspherical accretion onto compact objects, jet formation, star formation, and radiation astrophysics. Many scientists from universities who wanted to work on these frontier fields were forced to go to the national labs to gain access to the largest computers. As a result of this limited access, as well as the standard shortcomings of American universities (i.e. low salaries, lack of job openings, etc.), many of these people became lab employees. Others remained consultants to the labs so they could do their computing while remaining in the universities.

More recently, computational astrophysicists have been going to MPI in Munich, West Germany to do their computing. This is because the MPI CRAY-1 is devoted to physics and astrophysics, not to weapons design. Even scientists employed at the weapons labs can often get more CRAY time for their astrophysics projects at MPI than they can at LLNL or LASL. In the last two years the following astrophysicists have spent extended visits at MPI: David Arnett (U. of Chicago), Peter Bodenheimer (U. of CA.- Santa Cruz), Dick Miller (U. of Chicago), Michael Newman (LASL), Michael Norman (Ph.D. from U. of CA. -

Davis), Larry Smarr (U. of IL.), and Paul Woodward (LLNL). A similar list exists for theoretical chemistry. There are about 100 visitors to MPI per year, of which ~ 20 are heavy CRAY users.

Currently, the Japanese are furiously publishing work in the area of numerical hydrodynamics. Their major fields of attack so far are protostars, stellar evolution, accretion onto compact objects, radio jets, and general relativity. They are limited by their currently available computers, but within one to two years supercomputers will come on-line. At that point it is highly likely that Japan and Germany (already on-line) will take over from the U.S. the scientific leadership in numerical astrophysics.

Besides the obvious loss of talent from the universities to the weapons labs and foreign institutions, what are the other costs American universities must pay for lacking supercomputers? The primary problem is that numerical science must be done in the "crash program" mode. A visit to a supercomputer site is limited in duration: typically a few days to a few months. Thus one is under great pressure to complete a project in a short period of time. This means long hours (100 hours per week is not uncommon), great fatigue, little time for thinking or literature searching, and no time to discuss the project with colleagues. Furthermore, there is not the diversity of colleagues to confer with that one has nurtured in one's university. In many cases, one can't even look at the output of the calculation until one has left the site. This is not good scientific methodology.

Clearly, it would be much better to have the supercomputer (or access to one) at the home institution of the scientist. Then he could spend all year in "compute, think, read, talk, recompute" mode. That is, he could do a calculation, produce output graphically, thoughtfully look through it, realize he needed different functions plotted, compare his work with previous efforts, discuss his ideas with other American-based collegues, do an analytic side-calculation, understand more deeply what the physics is, and then decide what to compute next. He has access to a friendly and comfortable local computing system instead of having to learn an entirely new, foreign system in a few days' time. He has graduate students, secretaries, his own office and library, and a familiar circle of scientific colleagues to talk with. All of these factors are important for good scientific research. Finally, only the most hardy souls with the most forgiving families can socially, psychologically, and physiologically withstand the constraints imposed by computing sessions at distant supercomputer sites. Having supercomputers available to universities would bring many more top scientists into the numerical science revolution.

III. Frontier Science and Supercomputers

The above arguments do not pinpoint why supercomputers, rather than say many small computers, are essential in areas of frontier science. Perhaps the most succinct statement is that attributed to Edward Teller. "Teller's Rule" is that the most challenging frontier problems in science require at least 100 hours on the fastest computer available. As computers become faster, the knowledge gained on previous machines enables one to fruitfully attack the next level of complexity. Currently, there is virtually nowhere in the world where one could get this much CRAY time in a year as on off-site visitor.

One could make the argument that such problems could be attacked on small host computers attached to dedicated array processors. Indeed, fundamental work in quantum field theory and astrophysics has been done in this mode by some groups (e.g. the Cornell array processor users group). However, to date most computational astrophysicists prefer to have the large memory, high speed, fast turnaround of a supercomputer. The point is that the scientist should be able to choose the tool which best enables him to do his science on his timescale. Currently the answer to this choice is being masked by the physical difficulty of gaining access to supercomputers. Let me turn to a few examples of the type of fundamental science that is currently being done on supercomputers. Again let me stress that I can only relate areas of research I have personal familiarity with. The Press NSF Report "Prospectus for Computational Physics," 1981, details many more areas. Also, some of this work is work that will have to be done on supercomputers to get a reliable answer, but which is currently forced to be done on lesser computers (e.g. CDC 7600 or CYBER 175) because of lack of access to supercomputers.

My set of examples are a list of ~ 60 scientists at the University of Illinois whom I have talked to about the desirability of obtaining a supercomputer at the University of Illinois. These are researchers who would either actively use the supercomputer or bring in associates who would. The one-line description of areas of potential supercomputer research is augmented by more detailed statements of research intent in Appendix C. The clear conclusion is that there is a tremendous breadth of areas represented at only one major university.

POTENTIAL SCIENTIFIC PROJECTS
University of Illinois
CRAY X-MP Facility
(as of 11/1/82)

Physics Department

* Richard Brown (Prof.) + 9 other faculty	- Experimental Particle Physics
* Haldan Cohn (Res. Asst. Prof.)	- Evolution of Stellar Systems - Magnetohydrodynamical Accretion
* John Dow (Prof.)	- Electronic Materials
* Hans Frauenfelder (Prof.)	- Dynamics of Biological Molecules
* John Kogut (Prof.)	- Solving Quantum Chromodynamics
* Barry Kunz (Prof).	- Quantum Chemistry of Solids
* Fred Lamb (Prof.)	- Radiation Hydrodynamics of Neutron Stars
* V.R. Pandharipande (Prof.)	- Quantum Liquids and Solids
* David Ravenhall (Prof.)	- Nuclear Structure and Scattering
* John Stack (Assoc. Prof.)	- Quantum Field Theory
* Michael Stone (Assist. Prof.)	- Simulating Quantum Field Theory
* H. W. Wyld (Prof.)	- Lattice Quantum Chromodynamics and Turbulence Simulation

Astronomy Department

* Joan Centrella (Post. Doc.)	- Large Scale Structure of the Universe
* Helene R. Dickel (Res. Assoc. Prof.)	- Radio Interferometric Spectral Line Analysis - Radiation Hydrodynamics of Molecular Cocoons
* John Gallagher (Prof.)	- Optical Astronomy Instrumentation
* Icko Iben (Prof.)	- Stellar Evolution
* Steven Langer (Res. Assoc.)	- Accretion Onto Neutron Stars
* Telemachos Mouschovias (Assoc. Prof.)	- Star Formation
* Larry Smarr (Assoc. Prof.)	- Numerical General Relativity - High Resolution Finite Differencing of Hydrodynamics
* Lewis Snyder (Prof.)	- Chemical Astrophysics
* James Truran (Prof.)	- Hydrodynamics of Novae - Nucleosynthesis Networks

Chemistry Department

* David Chandler (Prof.)	- Quantum Dynamics of Molecules
* Clifford Dykstra (Assist. Prof.)	- Electronic Structure Theory
* Douglas McDonald (Prof.)	- Intramolecular Dynamics of Molecules
* Donald Secrest (Prof.)	- Molecular Scattering Theory
+ Peter Wolynes (Assoc. Prof.)	- Monte Carlo and Renormalization Group in Quantum Chemistry - Simulation of Solitons in Polymers

Computer Sciences

+ Roy Campbell (Assoc. Prof.) - Software Engineering
* Charles W. Gear (Prof.) - Numerical Software
* William Kubitz (Assoc. Prof.) - VLSI Design Automation
* Daniel Gajski (Assoc. Prof.) - "
* David Kuck (Prof.) - Supercomputer Software & Algorithm Development
* Ahmed Sameh (Prof.) - "
* Paul Saylor (Assoc. Prof.) - Iterative Solution of Large Systems of Linear Algebraic Equations
* Dan Slotnick (Prof.) - Supercomputer Design & Application
* Daniel Watanabe (Assoc. Prof.) - Simulation of Semiconductor Devices

Atmospheric Sciences

* Stan Kidder (Assist. Prof.) - Forecasting with Satellite Data
+ Yoshimitsu Ogura (Prof.) - Simulation of Convection & Turbulence
+ Su-Tzai Soong (Assoc. Prof.) - Simulation of Mesoscale Convection Systems
+ Kevin Trenberth (Prof.) - Climate Modeling
+ John Walsh (Assoc. Prof.) - Sea Ice Modeling
* Robert Wilhelmson (Assoc. Prof.) - Dynamics of Severe Storms

Chemical Engineering

* Thomas Hanratty (Prof.) - Turbulence
* Mark Stadtherr (Assoc. Prof.) - Chemical Process Design

Electrical Engineering

* Jacob Abraham (Assoc. Prof.) - VLSI System Design
* Sid Bowhill (Prof.) - Three-Dimensional Atmospheric Modeling
* Edward Davidson (Prof.) - Computer Architecture & Reliability
* Karl Hess (Prof.) - Simulation of 3-D Submicron Semiconductor Device
* Nick Holonyak (Prof.) - Quantum-well Heterostructures
* W. Kenneth Jenkins (Assoc. Prof.) - Inverse Problems in Tomography, Synthetic Aperture Radar Images, and Sonar Beam Forming
* Michael Pursley (Prof.) - Performance Evaluation of Communication Systems
* Timothy Trick (Prof.) - Two-Dimensional Simulation of Electronic Devices

Mechanical and Industrial Engineering

+ A. L. Addy (Prof.) - Gas Dynamics
+ Richard Buckius (Assoc. Prof.) - Radiation Heat Transfer
* Herman Krier (Prof.) - Hydrodynamic Shock Physics
* Frederick Leckie (Prof.) - Damage Mechanics

Theoretical & Applied Mechanics

+ Ronald Adrian (Prof.) - Comparison of Simulated Turbulence with Experiment
* Lawrence Bergman (Assist. Prof.) - System Dynamics
* Robert McMeeking (Assoc. Prof.) - Fracture Mechanics

Civil Engineering

* Robert Haber (Assist. Prof.) - Structural Optimization
* Leonard Lopez (Prof.) - Structural Mechanics and CAD/CAM

Aeronautical & Astronautical Engineering	
* S. M. Yen (Prof.)	- Rarefied Gas Dynamics - Computational Fluid Dynamics
Anatomical Sciences	
* Jay Mittenthal (Assist. Prof.)	- Biomechanics of Cell Sheets
Genetics and Development	
* Robert Futrelle (Research Scientist)	- Simulation of Living Cells
Physiology and Biophysics	
* Eric Jakobsson (Assoc. Prof.)	- Emergent Oscillatory Properties of Neural Networks
Accountancy	
* James McKeown (Prof.)	- Multiple Time Series Analysis

* = spoken to by L. Smarr
+ = spoken for by * people

Floyd R. Newman
Laboratory of Nuclear Studies
Cornell University

Newman Lab. Cornell Univ.
Ithaca, New York 14853
607-256- 5169

July 12, 1982

To: The Lax Panel

From: K. Wilson K G Wilson

Re: Priorities for Funding Current Needs

I suggest that the priorities for funding current needs, the subject of panel 2's discussion, be stated as follows. The basis for this plan should be the "Prospectus for Computational Physics" report (generalized to all areas of scientific simulation). The priorities proposed below are intended to implement the plan of that report for an effective national network of national, regional, and local computing systems. There should be heavy emphasis on the importance of the network as opposed to any particular site on the network. There should be minimum standards set for access and ability to use remote capabilities on the network, which most governmentally-funded researchers should be entitled to. The first priority should be to achieve this minimum standard for any researcher seriously involved in large scale scientific computation.

Secondly, it should be stated that none of the proposed budgets can come close to providing adequate cycles for scientific computing in universities, if only present-day or announced future large scale systems are used. It is not possible within this funding to provide all university researchers with the computing capabilities presently available (per researcher) at Los Alamos and Livermore. Besides which the capabilities even at Los Alamos and Livermore are grossly inadequate for current needs in basic research. They are even more inadequate to train students for the systems they will use professionally when they graduate (because of rapid technology development, the systems our present students will use after graduation will make the Cray-1 look like a desk calculator). Hence it is essential that rapid technology development be given very high priority; the proposals for providing cycles on current systems are only a very inadequate stopgap measure.

Therefore, I suggest the funding priorities stated below. This proposal is in addition to the proposals to supporting technology development, which I do not address here because panel 3 already has a good proposal worked out. The funding for technology development should have equal top priority with minimum network access.

1) Minimum standard network access from universities should include

   a) hardware and software to attach to network (to the extent not provided by DOE or other sources; also access must not be restricted to specific areas of basic research)

   b) Terminals in adequate number

Contd./

c) Local graphics capabilities – there is no centralized substitute for local graphics

d) modest dollar grants for local operating expenses and remote access computing charges

For this purpose, grants should be available for individual research groups or consortia, in the range of $10 to $50 K per principal investigator. Unless this need for access is met, the concept of the network is a joke.

2) A class VII (not class VI) national facility, with heavy emphasis on selecting a site that can achieve full production operation of the system rapidly. This is essential because of the rapid obsolescence of supercomputers that we are likely to face for the forseeable future.

3) Support for specialized capabilities at specific network modes (for example – special hardware and software offerings, or centers for specific disciplines).

With a full budget I would give equal weight to the three types of support, but with less than a full budget I would strictly adhere to the priorities even if that meant no funding at all for low priority items. Recall also that technology development has (in my view) equal priority with minimum network access.

KW/jmm

Yale University

Department of Chemistry • P.O. Box 6666 • New Haven • Connecticut • 06511

Kenneth B. Wiberg
203-436-2443

May 26, 1982

Dr. Richard Nicholson
Division of Chemistry
National Science Foundation
Washington, D.C. 20550

Dear Rich:

I recently returned from a JSPS sponsored 4 week trip through Japan. As you well know, their research work in chemistry has gotten to be quite impressive. I was surprised by the quality of their instrumental facilities. Every place I visited had a high field nmr spectrometer in the order of 300-500 MHz, and except for FT-IR, their other instruments were comparably good. At the present time, I think Japan and the U.S. are at about the same level with regard to generally used instruments.

There is one aspect of instrumentation in which they appear to be well ahead of us. At the Institute of Molecular Science, there are two CDC-7600 computer and Morokuma is planning for a new computer on the order of a Cray 1 in 1-2 years from now. At the National Chemistry Laboratory for Industry, there is a large Japanese computer with fiber optics links to the laboratories for high speed data transfer. The chemists at Tsukuba University (located a short distance away) have access to this computer. The University of Tokyo is planning to get a computer with the capabilities of a Cray-1 or Cyber-205. Tohuki University and Hokkardo University (on different islands) have their computers connected via a fiber optics link.

Every university I visited has a fairly large computer operated with a rather low user charge (~ 3¢/s.or $100/hr. with discounts for large users - overnight runs, etc.). In the U.S., no university can match the computing power available to Japanese chemists. Even LBL and BNL have only one 7600 each, where as The Institute of Molecular Science has two - for a smaller group of in-house users.

Japanese chemists have not, for the most part, come to realize the potential impact of computational methods on chemistry. Experience so far indicates that they can learn quickly - and they could essentially take over computational chemistry. NRCC was a feeble effort in comparison to what Japanese chemists already have - and even that was abolished. I am concerned about the future of U.S. chemistry.

Sincerely,

KBW:jl

Brenner

Fermilab

Fermi National Accelerator Laboratory
P.O. Box 500 • Batavia, Illinois • 60510

June 23, 1982

Professor Peter P. Lax
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012

Dear Peter:

I was pleased to be able to participate, at least in part, in the NSF and DOD sponsored Workshop on Large Scale Computing for Science and Engineering. Unfortunately, I had to leave for another meeting so that I was not able to attend Tuesday's Plenary Session. I did, however, have an opportunity in the morning to contribute a bit to the working group chaired by Jack Schwartz.

From my observations and efforts Tuesday morning, I'm convinced it's not going to be easy to put together a report that properly projects all of the concerns and the urgency of these concerns that were expressed by many people participating in the Workshop. Without the benefit of seeing the body of the report, it will be difficult, but as you requested I'll make an attempt at a few sentences which might become part of a preamble to define some of the areas of computing whose needs are not included in the scope of the Workshop's attention.

"Since the advent of the modern computer, many science and engineering disciplines, taking advantage of the new capabilities that these tools make possible, have made enormous strides in advancing their fields. The breadth and depth of functions to which these computing tools, in their various embodiements, have been put to use are enormously varied. Furthermore, even with the rapid improvement in performance and increasing diversity of these computing tools, the computational needs of the science and engineering communities have grown faster than industry is able to satisfy."

"One of the most important areas where the needs have far outstripped the capability of commercially available tools is in the large scale mathematical computing area, typically simulation of phenomena in science and engineering. Here there is a need for a two or three orders of magnitude improvement in current capability. It is primarily this need which has been addressed by this Workshop."

"It should be noted, however, that there is yet another class of problems, with needs for large scale computers, likely to require different solutions. These needs also are not met by commercially available systems by two and three orders of magnitude in computing power. These problems, associated with a number of areas of experimental physics, typically have relatively large amounts of input and output in association with a major amount of computational requirements, characterized by a large number of branching operations. Examples of these problems are found in the high energy physics and the magnetic energy fusion communities when they analyze their data. Although some of these problems are suited to special purpose-type processing machines, in most cases large scale general-purpose computing engines of appropriate architecture (very likely different from that which will serve the simulation class of problems) are more practical. In any case, these problems which may require different solutions than those under consideration here are not further discussed in this report."

Although I am sure that these three paragraphs will require much editing, I do hope they give you a framework from which may evolve an appropriate preamble.

If there is anything else that I can do to help in the report of the Workshop, please feel free to call upon me. I do hope that I will see a copy of the end product.

Sincerely yours,

Al Brenner

A. E. Brenner, Head
Computing Department

AEB:lp

WHY THE U.S. GOVERNMENT SHOULD SUPPORT RESEARCH ON AND DEVELOPMENT OF SUPERCOMPUTERS

Bill Buzbee
Computing Division
Los Alamos National Laboratory

OVERVIEW

Supercomputers are the largest and fastest computing systems available at any point in time; thus, the term supercomputer encompasses hardware, software, and supporting equipment. This document asserts that supercomputing is important to U.S. Defense and advocates government action to maintain U.S. self-sufficiency and leadership. Consequently, we will

- o review the role of supercomputers in science and engineering;
- o assess their importance to the U.S. and, thus, the risk incurred if another country assumes leadership; and
- o discuss why the private sector is unlikely to commit the investment necessary to maintain U.S. leadership.

The proposition is that government-supported research and development is required to lower technical/market risks to a point where private industry will commit to product development.

ROLE OF SUPERCOMPUTERS IN SCIENCE AND ENGINEERING

Basically, science and engineering consists of three activities: observation, hypothesis, and prediction. Observation involves the collection of data about physical phenomena, often under controlled conditions. Hypothesis is the formulation of theory to explain observation and typically yields complex mathematical models. Prediction is achieved by applying these models to real world problems. In all three cases, computers enable the scientist engineer to treat complexity that is not otherwise tractable, and, in so doing, the scientist engineer gains insight and understanding. Appendix I illustrates this point. Thus, acquisition of knowledge is the single most important benefit obtained from computers.

The amount of complexity that a computer can treat is limited by its speed and storage. If the computer cannot handle the desired complexity, the model is simplified until it is tractable (Appendix II contains an example of simplification). This is why scientists constantly press for bigger and faster computers--so that they can add complexity (realism) to their calculations. When a new supercomputer becomes available, it is analogous to having a new telescope, because with it the scientist engineer will see things not seen before and understand things not understood before.

In general, many important problems cannot be treated analytically. Computer modeling is the only way to cope with them.

IMPORTANCE OF SUPERCOMPUTERS TO THE U.S.

State-of-the-art supercomputers are manufactured by two companies, both U.S. Fifty of their supercomputers are in operation worldwide--38 in the U.S., 10 in Europe, and 2 in Japan. This is a small market, even worldwide. Of the 38 in the U.S., 25 are in government laboratories.

Current applications of supercomputers include

design and simulation of very large-scale integrated (VLSI) circuits,
design of nuclear weapons,
nuclear reactor research,
fusion energy research,
directed energy weapons,
intelligence,
aerodynamics,
atmospheric science,
geophysics,
petroleum engineering,
electric power distribution,
oceanography, and
fluid dynamics.

Potential applications include automotive design and manufacturing.

Electronics, nuclear weapons, and energy production are major components of U.S. defense. Design and simulation of state-of-the-art VLSI devices, nuclear weapons, fusion reactors, and nuclear (fission) reactors require state-of-the-art supercomputers. As discussed in Appendix III, progress in each of these areas may be paced, even bounded, by progress in supercomputing. The same is true for many of the above applications. Further, other important applications, such as information analysis and software engineering, may be paced by progress in fifth generation technology (see Appendix IV). Thus, leadership in supercomputing is fundamental to U.S. defense and to U. S. technology in general.

RISK IF ANOTHER COUNTRY ASSUMES LEADERSHIP

Should another country assume leadership in supercomputing, U.S. defense and technology will depend on access to computers of foreign manufacture. This presents three risks:

1. Currency of Access. If other countries consume the first two years of production, then U.S. scientists will be denied use of these machines and U.S. technology could eventually lag theirs by that amount.

2. Denial of Access. In this worst case, U.S. technology could be handicapped until domestic sources are developed.

3. Computing Technology. Development of supercomputers has always contributed to development of other computer systems. If another country assumes leadership, we may lose these benefits.

These risks are unacceptable.

WILL THE PRIVATE SECTOR MAINTAIN LEADERSHIP?

The growth rate in the speed of supercomputers is diminishing. In the '60s their speed increased by about 100, in the '70s by about 10; in the '80s it may increase by less than 10. The problem is that conventional computer architecture is approaching fundamental limits in speed imposed by signal propagation and heat dissipation. Continued progress requires revolutionary developments in computer architecture that will affect all aspects of computer technology [1]. Basic research is required before full commercial development will be possible. We must discover the right kind of architectures, ways to exploit VLSI in their construction, and suitable software for them. This research will necessitate collaboration between industry, government laboratories, and academia--industry knows how to build computers, government laboratories understand the computational requirements of large scale models, and academia has innovative people.

Historically, the cost of developing a supercomputer in conventional architecture has been $100 million (hardware and software). This cost has combined with the relatively small market to keep capital and resource-rich U.S. companies out of supercomputer development. It must be recognized that U.S. dominance in this market has been partially due to a lack of foreign competition. The cost of developing a new supercomputer in a new architecture will be substantially greater; also the risk of market failure will be greater. Simply put, there are more attractive avenues of investment. Thus, the private sector is unlikely to carry out the research and development necessary to maintain U.S. leadership. The proposition advocated here is that government-sponsored research and development is required to identify and isolate avenues of development that are likely to yield technical and market success. This will lower the investment risk to a point where the private sector will take over.

CONCLUSION

Computers are fundamental to the advancement of technology. This is particularly true of supercomputers because we use them to do investigations not previously doable. As we enter an era where other countries recognize their value and undertake their development and manufacture, the U.S. will be wise to maintain self-sufficiency and leadership in supercomputing technology.

SUMMARY

1. Supercomputers aid in the acquisition of knowledge.

2. Progress in U.S. defense technology is often paced, even bounded, by progress in supercomputing technology.

3. Dependence upon foreign sources for supercomputer technology poses unacceptable risk.

4. Conventional supercomputers are approaching their maximal performance level. Continued progress depends on development of new architectures.

5. The private sector may not assume the cost and risk of developing new architectures.

6. Because self-sufficiency and leadership in supercomputing are important to U.S. defense, the U.S. Government should fund research and development that will lower the technical/market risk to a point where private product development becomes feasible

---

APPENDIX I

COMPUTER SIMULATION

After the Three Mile Island accident, the Presidential (Kemeny) Commission and the Nuclear Regulatory Commission asked Los Alamos to analyze the early stages of the accident using computer simulation. Excellent agreement between computed and measured temperatures and pressures provided confidence in the computed results. But, of particular significance to this report, the computations provided information not otherwise available [2]; namely, they revealed the source of hydrogen that eventually caused so much concern. Thus, the use of computers provided new understanding about the accident.

Because of this demonstrated ability to simulate reactor behavior, people have suggested that this computational tool be made available to reactor operators, so that in future accidents it can be used to predict the merits of remedial alternatives. Unfortunately, these calculations require hours to perform on state-of-the-art supercomputers. For practical use by a reactor operator, the computation would have to be done in at most a few minutes and that necessitates a computer at least 100 times faster than any available today.

APPENDIX II

SIMPLIFICATION OF COMPUTER MODELS

One approach to producing energy from fusion uses magnetic confinement. Many of the associated devices contain a torus (a "doughnut" shaped container). Plasma is confined in it and then magnetically compressed and heated to the density and temperature required for fusion. Plasma scientists have long used computer simulation to model this approach. However, the toroidal representation is complex, so for many years they have simplified it. First the torus is sliced and "bent" into a cylinder (of course, doughnuts don't "bend," but this is a mathematical doughnut, so it does whatever we want). Periodic boundary conditions are imposed across the ends of the cylinder and radial symmetry is assumed within it. The result is a two-dimensional, rectangular approximation to a "doughnut." This simplification is used because the computed parameters vary widely in space and time. State-of-the-art supercomputers can handle only crude approximations to the torus, whereas the simplification is tractable. To handle well the complexity of the torus requires computers 10 to 100 times faster than any available today.

APPENDIX III

PROGRESS IN TECHNOLOGY

This appendix shows how progress in technology can be led by progress in computational capability. Appendixes I and II are also relevant to this issue.

DESIGN OF NUCLEAR WEAPONS

Since the beginning of the Manhattan project, computational models have played a central role in understanding and designing nuclear weapons. For many years ability to produce smaller and more efficient devices has been dependent on steady increases in computational power. The complexity of these designs has now reached a point where computational models have serious shortcomings. New developments in hydrodynamics and transport theory offer the possibility to correct this situation, but they require considerably more compute power than the techniques they replace. For example, implicit Monte Carlo techniques can require from 10 to 50 times the compute power of current techniques [3].

PRODUCTION OF ENERGY FROM FUSION

A second approach to production of energy from fusion involves inertial confinement. Construction of associated experimental facilities is expensive and time-consuming. In an effort to select designs that are most likely to succeed, computer simulation is used to study design alternatives. In some cases we are able to compute only a few percent of the phenomena of interest, even after running state-of-the-art supercomputers for as long as 100 hours. A major impact on fusion energy research could occur if we had new supercomputers that are 100 times faster than what we have today [4].

VLSI

Colleagues in the semiconductor industry indicate that supercomputers may assume an important role in the design and simulation of VLSI devices. Confirmation should be obtained from experts in this area.

APPENDIX IV

FIFTH GENERATION COMPUTERS

"Fifth generation" computers are computers designed to support application of artificial intelligence. These machines are also "super" in their level of performance, and like supercomputers necessitate new architectures. The Japanese are pursuing both areas and are spending about $100 million per year on the combined effort. Artificial intelligence has potential for broad application; for example, information analysis, intelligence based weaponry, and software engineering. Leadership in this area is essential to the security of the U.S. If the U.S. government initiates an R&D program in supercomputing, fifth generation capability should be included.

ACKNOWLEDGMENTS

This report benefited by comments from

Donald Austin, Department of Energy;
James Browne, University of Texas at Austin;
Robert Ewald, Los Alamos National Laboratory;
Sidney Fernbach, Control Data Corporation;
Paula Hawthorn, Lawrence Berkeley Laboratory;
George Michael, Lawrence Livermore National Laboratory;
James Pool, Office of Naval Research;
Dale Sparks, Los Alamos National Laboratory; and
Jack Worlton, Los Alamos National Laboratory.

REFERENCES

1. "R & D for National Strength," edited by David Williamson, Jr., The Center for Strategic and International Studies, Georgetown University, Washington, D.C. (1982), pp. 21-23.

2. "Los Alamos National Laboratory Annual Report 1980," Los Alamos National Laboratory report LAC-81-15, pp. 28-29.

3. "ADP and Data Communications Long-Range Plan - FY 1984," Computing Division, Los Alamos National Laboratory, January 1982.

4. David Forslund, "Large Scale Simulation Requirements for Inertial Fusion," presented at the conference on High Speed Computing, Gleneden Beach, Oregon, 1981.

Massachusetts Institute of Technology
Laboratory for Computer Science
Cambridge, Ma. 02139

Notes for the Shanty Creek Meeting

Jack Dennis
13 August 1982

Organization

I suggest that major funding for high performance computer research in universities be overseen by a Review Board, much in the style used by the British Science Research Council. Membership in the Board would be selected from universities and from industry to represent competence in conceptual, analytical, manufacturing and application aspects of high performance computer systems. Such an arrangement is needed to avoid the haphazard nature of current funding which results from the different review and evaluation procedures used by different agencies, and from the difficulty of obtaining comparable reviews of competing projects.

It might be best to have the evaluations of the Board advisory to the agencies so that funding decisions are left in the agency's hands.

Scope of the Area

The area covered by the Review Board should be projects that involve funds for substantial equipment purchase and/or substantial use of non-academic research staff for pursuing relevant goals. Projects involving one faculty and one or two graduate students and no industrial cooperation should not come within the purview of the Board, although the relevant work of such groups should be brought to the attention of the Board.

Goals

The goal of research projects supported should be to further the understanding of how concurrency may be exploited in future computer systems to achieve systems that have greatly improved performance (absolute), performance per unit cost, reliability/availability, and programmability. Systems designed to support computational physics, general-purpose computation, or signal processing, or data base processing stand equally to benefit from this organization of research support. It would be an unfortunate error to restrict the purview of the Board to computer systems capable of supercomputer performance on computations of mathematical physics.

An Opinion

A small but growing group of experts in computer science research believes that future computer systems must be designed to provide a closer and better match of the structure of the hardware to the structure of programs to be run on the proposed system and the design of the language in which these programs are expressed.

This aspect should be an important criterion in judging the suitability of projects for support. It is reasonable to expect that the feasibility of a complete system of software support be established -- or even that it be implemented -- before major funding for hardware construction is committed.

Methods of Evaluating Architectures

Given a proposed computer systems architecture, it must pass certain tests of viability. One test (functionality) is to assure that programs expressed in an appropriate source language can, in fact, be translated into correct machine code. Simulation by means of programmed translators and interpreters is the standard method for doing this.

The second test (performance) must evaluate the performance of a specific machine configuration for a suitable set of realistic application benchmark problems. There are four approaches to carrying this out:

- Build: A full or scaled down version of the proposed system can be built, scaled down problems run on it, and the results extrapolated to full size problems and systems.
- Emulate: Multiple hardware units or small microprogrammed machines can be built to perform the intended function of full-scale units. This is hard to distinguish from "Build" except that cruder implementation techniques may be used to save on design time and commitment, and cost. This is distinct from simulation in that timing information is not directly available, and performance levels must be ascertained from analysis of emulation results.
- Simulation: Simulation would most likely be done using conventional techniques with traditional high-level languages and machines of conventional architecture. Higher performance may be achieved by running simulations on multi-processor systems, but true simulation is not easy to do on a multi-processor.
- Analysis: In some cases the applications have a sufficiently regular structure and the mapping of problem onto machine level code is sufficiently understood that manual analysis is adequate to justify a claim of high performance.

Evaluation is not complete without consideration of cost of fabrication for the proposed machine. Some proposals, which have significant merit otherwise, will fall down when judged by the cost of their construction.

Thoughts on Very High Speed Computer Systems for the 1980s

- by -

D. D. Gajski, D. J. Kuck, and D. H. Lawrie

Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801-2987

May 1982

Background

Consider the goal of producing several types of broad-range, multi-gigaflop computer systems in the next decade. The basic problems to be solved are (as always) what hardware to use, what architectures to use, and what combination of software and algorithms to use in achieving the goals. We will avoid here the specifics of the details of performance goals, exact application areas or breadth of application. But it is assumed that we are concerned with more-or-less general-purpose systems, at least general-purpose within an application area. Several general areas may be under consideration though.

We will discuss some of the strengths and weaknesses of universities proceeding alone and companies proceeding alone. This leads to the conclusion that some kind of joint effort is the best approach.

Universities have traditionally been good at proposing architectural ideas, implementing research software projects, and carrying out theoretical and experimental performance evaluations. On the other hand, they have been poor at technology innovation. While the silicon foundry approach makes it possible for them to produce new chips, they will not be able to exploit the state of the art in performance or reliability. Another recent trend in academic computer architecture research seems to be toward the "one big idea" approach to computer architecture. This seems to produce architectures with a few simple, easy to understand ideas, that lead to a lot of related theoretical work, but too often without enough real-world applicability.

The focus in companies is shifting somewhat these days. The traditional "system houses" are continuing in some cases with products in which they have a long-term investment (e.g., Cray, CDC), others seem to have sharply retreated (e.g., Burroughs), while some are trying to enter for the first time with new ideas (e.g., Denelcor). Some will attempt to copy the successes of others (e.g., the Japanese, Trilogy). An interesting shift of focus, however, is to the semiconductor houses themselves. Several are now producing interesting, innovative, low-speed VLSI products (e.g., the Intel 432). Several are breaking into the one megaflop (32-bit) chip or chip-set arena (e.g., TI and HP).

It would seem that a great moment of opportunity is near in the possible wedding of high-speed architectures and VLSI components. In fact, we have been approached by one semiconductor manufacturer with precisely the question, "How can we produce a chip set that OEM customers can fabricate into high performance multiprocessors of various types?" It is clear that companies are frequently forced to announce new products without enough time to think through the architectural, software, and applications ramifications of their work. Preliminary evaluations of the Intel 432 seem to be rather mixed, indicating a product that was perhaps rushed to the marketplace too quickly. In the area of high-speed systems, it is probably even more difficult to get a set of components that will perform well in a variety of systems and applications.

A joint attack on the problem by industry/university cooperation would seem at this moment to have a high probability of success for several reasons.

1) Several universities have well-developed efforts in

architecture, software, and algorithms for high-performance computers.

2) Several companies have already produced high-performance component prototypes and are exploring the systems-level implications of these.

3) The VHSIC program provides an example of a successful prototype program for this type of cooperation.

As indicated above, the chances for success seem low if university and company efforts proceed independently. However, by properly managing a joint effort, some rather important breakthroughs should be possible.

Goals

The goal of the project would be to have one or more very high-performance computers built with advanced architectures and advanced technologies. Furthermore, these computers would be built by industrial organizations that would be likely to continue product-oriented development of the systems.

We would argue that certain key conditions should be met to ensure the success of such a joint venture. It would almost certainly make sense to carry out two or three parallel projects which would be of different natures and which could cross-fertilize each other. Each such project should follow the outline below:

1) Prototype. The architecture should be worked out and a (scaled down) prototype built using off-the-shelf parts in a short term (e.g., two-year) project. This milestone would include the development of register-transfer level diagrams together with timing diagrams that would be

delivered to the industrial partner. By concentrating here on the architecture, the university partner would not be drawn into the fatal position of trying to push architecture and technology at the same time.

2) Software and Performance Evaluation. In order to demonstrate the success of the project, software and algorithms should be produced by the university for the prototype so that it can be used on a variety of applications and its performance can be measured. This milestone is a validation of the prototype and simulation of the final product. These efforts should go hand-in-hand with the architecture design and not be afterthoughts (as we have seen in some university projects).

3) Implementation. The initial design should have the property that it may be speeded up by (one or more of) several techniques.

a) The design should be architecturally scalable, e.g., by replicating hardware parts it should perform faster.

b) The design should be amenable to cost/performance improvement through hardware enhancements by the industrial partner. For example, the use of faster circuits or the use of custom chips instead of gate arrays would allow a faster clock. Custom VLSI chips would give a cost and size reduction as well.

The software developed by the university partner on the prototype would be able to run on the final implementation. Further software development by the university in parallel with the final product implementation should relieve the industrial partner of long software development delays.

Summary

A broad-based approach in which several industry-university teams were involved in separate projects would seem to be an ideal approach to reach the goal of multi-gigaflop systems in the 1980s. If the groups were competing with respect to system speed, quality of results and schedule (not budget dollars), they would also be able to contribute to one another by exchanging ideas, serving on a joint advisory panel, etc.

The recent DOE-sponsored meeting in New York demonstrated that there are a number of university groups working in the same general direction of high-performance computer system design. Some have narrower application goals than others, some have less software strength than others, and some have rather narrowly focused architectural ideas (i.e., the "one-issue" architectures). But, overall there is a lot of vitality in academic architecture research.

In practice, most system manufacturers have a large investment in one or two architectural styles and are constantly seeking new hardware components and new software that will allow their architectural styles to gain a factor of two-to-four speed advantage over their competitors. The merchant semiconductor manufacturers have just recently arrived at this system level and are currently trying to determine what to do next. While their highest volume products are clearly in other areas, the diversity of 32-bit products and projects they have revealed in the past year indicate that the time is exactly right to forge bonds between these three communities to accelerate and focus progress on superspeed systems for the 1980s.

Position Paper for NSF Workshop
Bellaire, Michigan
23-24 August 1982

Fred Ris
IBM Research
Yorktown Heights, NY

The views expressed herein are strictly my personal views and should not be construed as being shared by other individuals or any organization, including my employer, IBM.

We are here because we share a fundamental premise that there may be grounds for establishment of a national-interest activity in high-speed scientific computing systems for which the coordination (technical and financial) of the Federal government is indispensable. The end goal that seems the most appropriate is a computing system providing sustained (not instantaneous peak) computing power two to three orders of magnitude in excess of that available today. General purpose computing systems delivering 10 to 100 gigaflops are extremely unlikely to appear in the market in the 1990's without government intervention at this point. I therefore take the design and development of such a system to be the aim of the enterprise we postulate.

Certain component technologies will not be accelerated any further by the existence of such a project and may therefore present potential unavoidable implementation constraints. While there will be some dispute as to precisely which component technologies fall into this category, I believe that high-speed circuit technology, levels of circuit integration, and capacity and access time of rotating storage media all have maximal impetus arising from the needs of commercial data processing. These technologies will therefore not be accelerated by the existence of a national project for high speed scientific computing, no matter how potent. Further, I do not expect these technologies to produce three orders of magnitude performance improvement in the next decade.

If I assume that such a system can nevertheless be built in ten years, I would conclude that the technologies in which leverage must be applied are processor architecture and machine organization, system architecture, and application partitioning methodology. These are challenging research areas, of which many organizations are currently in pursuit. There is no particular national coherence in these pursuits, and it is therefore in those areas that a Federal initiative might make a substantial difference.

Although I claim that silicon technology will not be significantly accelerated by scientific computing initiatives, it is nonetheless true that processor and system architecture and organization questions are not independent of this technology. The VHSIC program, for example, may be quite significant, and its progress must therefore be tracked to understand the implications of its results on the areas in which we will be interested.

NSF position paper F. N. Ris

Since there is no reason to believe we will be able to build large uniprocessors with sub-nanosecond cycle times in any known technology, multiprocessing will be inevitable. Most multiprocessing research presently focuses on communications among processors, and many of the communications network problems thus raised have lives of their own as interesting research topics. While understanding the fundamental pros and cons about particular network topologies is important, of greater importance for this endeavor is understanding the degree to which applications can be fit on multiprocessing implementations without compromising severely the domain of applicability of the implementation and without imposing inordinate human-intensive, front-end partitioning loads.

To this end, I think it crucial to understand a few key representative applications for such computing systems, to understand whether the balance among computing power, memory capacity, backing store capacity/latency, and communications is reasonable, and also to provide the earliest possible indication that the technologies' courses of evolution will be consistent with the requirements of these applications, as they exist today or as they can reasonably be projected to exist in a decade.

A fundamental dilemma is hereby presented in that some of the applications that motivate such a national project in the first place are classified, particularly at the level of detail that is necessary for evaluating the suitability of prospective system configurations. In this case, knowing gross characteristics does not suffice; details of the inner workings of the applications are crucial for an understanding of how processing, memory, and communications requirements can be partitioned.

One area that can absorb unlimited quantities of computing resources and that is fairly easy to grasp is atmospheric modeling. Weather models, while in some cases subject to constraints relating to academic proprietary interests, are for all intents and purposes freely accessible for the purposes of gaining the understanding suggested here. Some NASA applications may also have much the same properties. Classified national laboratories application requirements, while perhaps strongly motivating a Federal initiative, will not be of practical use in helping make design decisions about the systems that are ostensibly to be built for their benefit.

It was repeatedly stressed at the June meeting that the probability of success of such an endeavor would be seriously compromised absent joint participation from the outset of government laboratories, universities, and industry. National laboratory involvement is required to provide application understanding and the ability to project it into the future, including knowing how today's implementations can be restructured if necessary to fit a different balance of component technological capabilities tomorrow. The universities are a source of invention and should also provide a mechanism for ensuring the integrity and consistency of experiments and the evaluation methodology of alternative implementations. Industry must bridge this gap as well as bringing the realities of engineering and the ability to produce prototypes.

These ingredients have been an integral part of the operation of the successful VLSI and PIPS projects sponsored by MITI, as well as in the ongoing Fifth Generation and High-Speed Scientific Computer projects. What the MITI development model adds to these components is a mechanism for achieving vital early consensus on technical direction. The Japanese government provides not only financing, but also technical leadership (often co-opted from national laboratories and universities) to achieve a nationally optimized balance between competition and co-operation. My concern is that without some similar mechanism in the U.S., there will be too much local (institutional) optimization, leading both to subcritical implementation mass for good ideas, and possibily precluding a more potent synthesis. A Federal government role is the obvious (but perhaps not the only) answer, which would be effective only if financial support is the mechanism for enforcing this discipline.

I do not believe that an exact MITI model can work in the west because there are fundamental cultural forces in Japan that help to cement these relationships which are substantially weaker elsewhere. There are large-scale projects in the past twenty years that we might look to for retrospective insights about where organizational strengths and weaknesses lie. Among these are the Alaska pipeline, Apollo program, Concorde, the Space Shuttle program, and VHSIC. Similarly, we may want to contemplate the direct initiatives the French are taking in semiconductors and personal computing. Finally, we should consider private-sector initiatives recently launched in the U.S.; notably the Semiconductor Research Cooperative of the Semiconductor Industry Association and the proposed Microelectronics and Computer Technology Enterprises.

In summary, while there are plenty of exciting technical challenges ahead of us, I believe the political and social problems will be equally hard and will appear on the critical path sooner rather than later.

SIDNEY FERNBACH, Ph.D.
CONSULTANT
2020 RESEARCH DR. LIVERMORE, CA 94550
(415) 443-1300

September 9, 1982

Dr. Ettore F. Infante
National Science Foundation
1800 G Street, NW
Room 339
Washington, DC 20550

Dear Dr. Infante:

The draft report on Large Scale Computing in Science and Engineering is fairly good, but fails to make what I believe to be the essential point. We do not lag the Japanese in general access or applications know how for supercomputers. We are falling behind in the plans to design and actually construct the supercomputers for the 1990's. Fujitsu and Hitachi have already announced their supercomputers for delivery in 1983 which are on the average superior to the Cray 2 (not only the X-MP) and possibly the next CDC computer. Besides the Japanese government is supporting the component research and construction of a next generation (for the 1990's). We must stress not only the "long-range research basic to the design of a new supercomputer systems", but we must also stress that the government should procure by competitive bid as early as possible a 100 BFlop machine--hopefully before 1990.

Without this aim the paper is worthless. It just asks for more money to do what is already being done and which will not match the Japanese competition. We need a real goal to make this a new exciting venture for the government to support as well as to provide the U.S. with the continued leadership it may lose (most certainly will) in this area.

Sincerely,

Sidney Fernbach

SF:kw

SIDNEY FERNBACH, Ph.D.
CONSULTANT
2020 RESEARCH DR. LIVERMORE, CA 94550
(415) 443-1300

October 7, 1982

Dr. Ettore F. Infante
National Science Foundation
1800 G Street, NW
Room 339
Washington, DC 20550

Subject: Comments on the report from Group 3 of the Lax Committee.

Dear Dr. Infante:

Distinguishing between the various ongoing university efforts in "Advanced" computer architecture is like distinguishing between Illiac's, Maniac's, and Johnniac's in the 1950's. Most efforts are very similar, the chief differences among them being in how to switch from one unit to another. One is led to think that putting a large number of small processors together is the only way to create a super-scale computing engine.

There should be no question that this "research" should be continued, but does it bear still greater expenditures? They should be pushed to completion, if possible. There are some techniques that are still so far from reality that a great deal more work is necessary (such concepts as data flow and the systolic array, for example). This work should be supported at higher levels and also pushed into reality.

More "Traditional" systems with special features, such as vector boxes or functional processors can be exploited now. Let us set our goals and go. If we start researching the parallel processor approach it will be years before we start moving. Even with this architecture we can move fast if we take advantage of what has already been done, e.g. Denelcor's HEP, CDC's AFP, and LLNL's S-1.

It is interesting that all the talk is about a CPU-memory system. Little or no mention is made of the essential peripheral equipment that can help make or break the system. Not only do we need a 100 GFLOP machine, but we also need matching peripherals. We must define our goals and let the manufacturers tell us how to get there. They will if there is any incentive.

Our immediate need is a target performance and target date. The cost may be high but we can reach both goals if we pay the price. Certainly there should be a cooperative effort between industry, government and university. Only one, however, should be in charge. This role should be played by industry--the manufacturer which takes the lead in the development. Government committees are not good for anything but offering criticism. They should be used in an advisory capacity. They may have to sort out alternatives and be prepared to make compromises, but no changes.

Sincerely,

Sidney Fernbach

SF:kw