National Science Foundation
Washington, D.C. 20550

Official Business
PENALTY FOR PRIVATE USE, $300

Postage and Fees Paid
National Science Foundation

THIRD CLASS
Bulk Rate

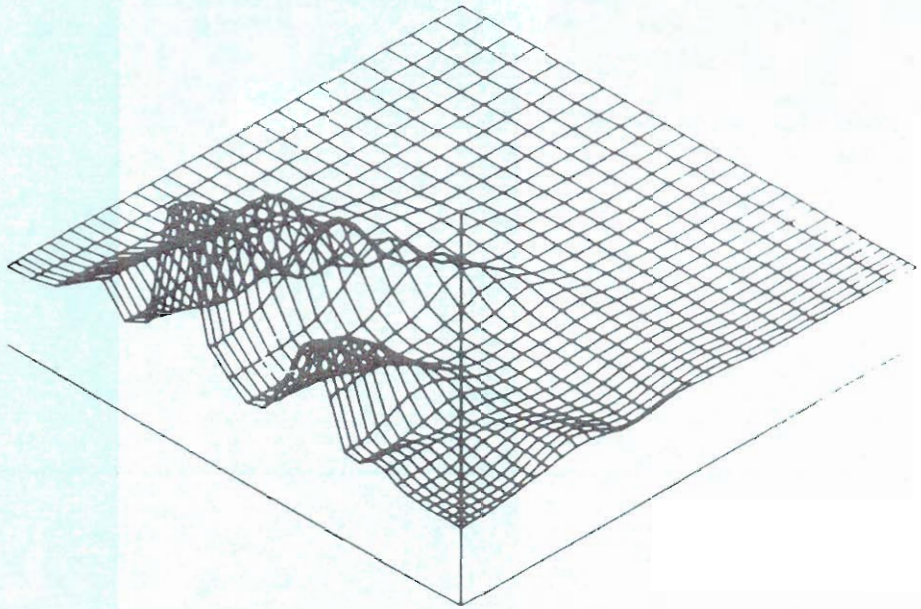
U.S. MAIL

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



NATIONAL SCIENCE FOUNDATION

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

Subcommittee Membership:

William H. Press, Chairman (Harvard University)
J. Norman Bardsley (University of Pittsburgh)
Arthur J. Freeman (Northwestern University)
Edward J. Groth (Princeton University)
John Killeen (Lawrence Livermore Laboratory)
Oliver McBryan (The Courant Institute, New York University)
William M. MacDonald (University of Maryland)
Claire E. Max (Lawrence Livermore Laboratory)
Steven Orszag (Massachusetts Institute of Technology)
Larry L. Smarr (University of Illinois - Urbana)
Kenneth G. Wilson (Cornell University)

CONTENTS

ABSTRACT AND EXECUTIVE SUMMARY

Computational Physics extends theoretical physics beyond the limitations of analytic techniques. This extension has become essential to the advance of many different subfields of physics as systems of interest have become more complex, moving from a few degrees of freedom to many degrees of freedom. In this report, specific examples are discussed from the subfields of quantum field theory, statistical mechanics, condensed matter theory, atomic and molecular physics, plasma physics, nuclear physics, physics of fluids, astrophysics, and gravitation theory. In all subfields, computational physics studies inevitable complexity.

Present NSF support of computational physics is inadequate to the present needs and state of development of the field, and woefully inadequate to the field's future development. Direct expenditure on theoretical physics computing in FY80 was less than $400K. The actual level of effort was, we estimate, in excess of $10M, obtained through special arrangements outside of the NSF grant system. While one might wish for this "invisible" funding to continue, it is in fact drying up rapidly at just the time that the intellectual prospects of the field are expanding explosively. It is time for the NSF to take on major responsibility for computational physics, as a substantial new initiative within the Division of Physics.

The field is at an immediate crisis stage. This report proposes the establishment of a networked computational system of national scope to bring necessary capacity and capability to the computational physicist (with allocations according to established peer-review procedures.) Necessary components include: a managed communications network; a principal node offering the mature capability of a large-scale super-computer; a variety of networked regional and local nodes incorporating computing capacity complementary to that provided at the principal node; augmentation of some of these local nodes with state-of-the-art array processors, available to all network users; provision of the ability to handle large amounts of data in graphic form.

Some rough cost data are given. The *immediate* needs of the field are at the level of about $10M per year; substantially larger funding will be needed for the long term health of this vigorous new mode for physics research. Computational physics extends theoretical physics, but its support requirements (and, in fact, intellectual styles) are in many ways more directly comparable to experimental physics. There are also prospects for close interaction and cross-fertilization with industry-based physicists, and with university-based scientists in allied fields.

I. INTRODUCTION

At its October 1980 meeting, the NSF Advisory Committee for Physics expressed concern about the present state and future prospects of computational physics. As a result, a Subcommittee on Computational Facilities for Theoretical Research was appointed with a four-point charge: 1) to examine present and future trends for computer usage in university-based theoretical physics research, including a review of studies by other committees; 2) to review the present NSF response to scientific opportunities for computational physics and to form an opinion as to its adequacy; 3) to consider, as alternative approaches to meeting future needs, a range of possibilities from smaller computers and networks to large supercomputers; and 4) to recommend an appropriate strategy for strengthening the computational capability of the NSF Physics community over the coming decade.

The Subcommittee met by teleconference on November 14, 1980, and for three days at NSF on January 15-17, 1981. R. Deslattes (Director, Division of Physics) addressed the Subcommittee, and a presentation on present NSF support of computational physics was given by R. Isaacson, with additional NSF staff representation by J. Mandula, L. P. Bautz, E. Hayes (Chemistry), J. Connolly (Materials Science), and K. Curtis (Computer Sciences). Available to the subcommittee were previous reports and technical documents which are listed in Appendix A.

The Subcommittee is unanimous in its finding that NSF support of computational physics is inadequate to the present needs and state of development of the field, and woefully inadequate to the field's future development.

Current total support for theoretical physics (including condensed matter) by NSF is at the level of $13.2M (FY80 expenditures). Of this amount, $22,010 was budgeted for computer equipment, and $340,101 for purchase of computer services (representing 140 awards). As a FY80 budget category, then, direct expenditure on computing commands 2.8% of theoretical support.

However, it is clear that this figure has no relationship whatsoever to the actual level of computing effort by university-based theoretical physicists. We estimate that the "real" funding level is on the order of $10M per year or larger. The use of special arrangements for obtaining computer resources is not merely endemic; it has of necessity become the norm in this field. Sources of the computer capacity that is actually used include the Livermore and Los Alamos laboratories; underutilized (and usually obsolescent) university-owned computers; small and medium-scale computers bought for experimental efforts; informal arrangements

with industry. Many researchers are having to go abroad for their computing, since resources there are more freely available and are rapidly increasing.

One might wish for computational physics to continue in its mode of "invisible" funding. This is flatly impossible. Resources are drying up at just the time that computational needs are growing rapidly. The need to train students, both as the next generation of leaders in the field, and for industry (where computational science is expanding explosively) cannot be met under the present mode. Most important, as we discuss in detail in the next section, the significance of computational physics to the advance of physical theory in all areas has become too important to leave to a haphazard system under which large (but now insufficient) resources are obtained informally and without reasonable time-stability. The problem is at an immediate crisis stage; we attempt in this report to suggest what needs to be done.

The report is structured as follows: Section II deals with the intellectual content of computational physics as a mode for extending theory well beyond analytic limitations; examples from the various subfields of physics are cited. Section III focuses on the special nature of computational physics as an activity: its subject matter is theory, but its methods, styles and needs are in many ways more closely akin to experimental science.

Section IV addresses the question of what is to be done—soon. It focuses on the components of a system of national scope required to bring necessary capacity and capability, with proper balance, to the computational user. Necessary components include: a managed communications network, the capability of a large-scale supercomputer, building up of a system of compatible medium scale, virtual memory machines at local and regional levels, augmentation of some of these local nodes with array processors running state-of-the-art software, and provision of the ability to handle large amounts of data in graphic form.

Section V suggests some additional guidelines for the development of the expanding and vigorous field of computational physics. If we do not continue beyond "today's necessities" (Section IV) we will not keep up with the intellectual imperatives of the field. Necessary directions that we discuss include: supercomputers based on array processor technologies, operating systems, program languages, data management, and data bases.

II. COMPUTATIONAL PHYSICS: THE INVESTIGATION OF COMPLEX SYSTEMS

As the fields of physics advance, theoretical investigations naturally evolve from a stage where the most important problems can be solved analytically, to where numerical solutions become essential. At any one time, different sub-fields are in different stages of this process. However

one can identify—even across the boundaries of different subfields—common features that drive this evolution. In general, large-scale computation becomes necessary as the systems of interest become more complex. One moves from a few degrees of freedom to many degrees of freedom.

Complexity emerges in different ways in the different subfields of physics. In some cases, it is in moving from ordinary differential equations to partial differential equations for the solution of interesting problems. This has happened, long since, in fluid dynamics and for the Schrodinger equation in non-central or many-body conditions, for radiative transfer problems in astrophysics, and in other areas. Alternatively, in areas such as nuclear reaction theory, many coupled integro-differential equations naturally arise and must be solved self-consistently.

Complexity also increases dramatically when one moves from simple or one-dimensional models of physical processes, to realistic simulations. One sees this in research on the phases of fluids and solids and lattice models, in plasma simulations, in many-body calculations of galactic dynamics, and in quantum field theory.

Complexity can come from the impetus to move from low order to high order expansions. This is the case in quantum electrodynamics, quantum chromodynamics; in high temperature and other expansions for liquids, solids, and lattice systems; and in high partial wave expansions for nuclear reactions. When the expansion techniques are algebraic, one moves from hand-algebra to algebraic symbol-manipulation on the computer.

Complexity arises in moving from scalar systems to vector or tensor systems, and from linear systems to nonlinear systems. A striking example of this is General Relativity, whose partial differential equations are both tensor and highly nonlinear. The fluid dynamics of many-component systems is another example.

One sees, then, that complexity arises not from "bad taste in the choice of problems", but inevitably as theory advances. As one surveys the scope of subfields of physics, as we shall now do, one finds over and over again that (i) there are important problems whose solutions must be found by computational techniques, and (as we shall have more to say about in following sections) (ii) that even the best investigators are resource-starved, without the facilities for accomplishing even the tasks already at hand.

QUANTUM FIELD THEORY

In quantum electrodynamics it is possible to extend perturbative solutions to high order, and large scale computing has played an essential role. The calculations of the sixth-order magnetic moment of the electron,

and now the eighth-order moment by Kinoshita, and its impressive agreement with experiment to 1 part in 10^8, has provided one of the most accurate verifications of a fundamental theory in all of physics.

It has now become necessary to carry out equally complex calculations for quark theory (quantum chromodynamics) in order to make realistic comparisons of theory with experiments. The algebra of these calculations, in particular, is extremely complex, and computer-based algebra programs such as MACSYMA have become crucial. In addition, the integrals of sixth and eighth order calculations can only be done by lengthy Monte Carlo integrations.

In continuum quantum field theory, there is no known method for numerical approximation of the non-perturbative regime. However, the formulation of lattice versions of currently popular gauge theories has led to substantial new insights into the theory of quark confinement, and has also opened up the application of statistical mechanical high-temperature expansion techniques to study the behavior of gauge theories for large values of the gauge coupling constant. The Monte Carlo calculations of Creutz and others for the pure gauge theory have had a substantial impact on elementary particle theory and have been repeated and extended in many laboratories. However the Creutz calculations are just a beginning. Very substantial increases in computing power are needed to allow more realistic calculations which use lattice sizes large enough to show independence of the lattice itself, to incorporate quarks in a realistic way into the calculations, and to use the physically relevant color SU(3) group in place of the simpler SU(2) model mostly studied to date.

One of the most fundamental and baffling problems facing elementary particle theorists is to understand dynamical breaking of chiral symmetry, namely a breaking which does not require Higgs fields. This occurs in strong interactions and is an essential part of a major class of unified theories of weak and electromagnetic interactions. It occurs only in strongly interacting field theories, so it cannot be studied perturbatively. To date, no lattice approximation has been found which maintains continuous chiral symmetry. An important research effort at the moment is the search for a framework allowing the numerical study of dynamical symmetry breaking. If such a framework is found, it is certain that the computing requirements to study this problem (which involves both fermions and gauge fields) will be immense.

STATISTICAL MECHANICS

An important chapter in the history of the theory of critical phenomena in matter (magnets, fluids, etc.) was the use of high-temperature expansions, to very high order, to compute critical exponents. Large scale computing

was required to push the series beyond the level computable by hand; this was needed to check the consistency of extrapolations to the critical temperature from very high temperatures. This work established the failure of mean field theories of critical behavior and paved the way for the renormalization group approach to describing critical behavior. This effort continues both through higher order calculations for simple systems, and through new, more complex, applications. Some new applications (e.g. multicritical points) involve expansions in several variables which require considerable computing power.

Another very fundamental result in computational statistical mechanics was the calculation by Alder and Wainwright of the equation of state of the hard sphere liquid using molecular dynamics methods. This has been the basis for many detailed studies of the liquid state by perturbing around this model. Molecular dynamics and Monte Carlo methods are now being used to tackle increasingly complex liquid systems. The simulation of liquid water is currently a topic of great interest and difficulty.

RENORMALIZATION GROUP

Theoretical physics at its most fundamental level involves simple insights which unify vast areas of physics. For example, Maxwell's equations unified electricity and light, and recent developments in gauge theories promise to unify strong, weak and electromagnetic interactions. A major achievement of the 1960's and early 1970's was the development of the renormalization group approach. This approach showed that quantum field theory, the theory of critical phenomena, the behavior of large polymers and the problem of fully-developed turbulence all reflected different realizations of the same fundamental problem, namely the treatment of systems with a large or infinite range of length scales. Through the work of Fisher, Kadanoff, Wilson and many others, a framework was developed to characterize the behavior of these systems, and new computational techniques were found to solve a number of problems of this type. However the quantitative methods developed so far have been applicable only to a small fraction of the problems involving many length scales, and in the case of turbulence, it is not even clear whether the formal ideas truly apply. At the present time investigators are trying to apply very large scale computational methods (Monte Carlo methods in a renormalization group framework) to try to obtain reliable numerical calculations for a broader class of problems than possible hitherto, and in particular to verify and quantify some of the renormalization group formalism applied to these problems. For example, one hopes to calculate a range of critical exponents (including subdominant exponents), or to follow the effective coupling of quarks in quantum chromodynamics as it "crosses over" from weak coupling at short distances to strong coupling

at long distances. Further development of numerical procedures based on renormalization group ideas will require very extensive computing resources, but could have a profound impact on very many areas of physics and applied physics research. The renormalization group ideas have a very broad reach.

CONDENSED MATTER THEORY

Condensed matter theory includes such diverse subdisciplines as the electronic and geometric structure of perfect and imperfect solids, disordered and amorphous materials, liquid metals and alloys, surfaces, interfaces, and artificial (laboratory made) exotic materials such as superlattices. It ranges from direct applications of principles for device fabrication to complete formal theories to explain and predict new properties of matter. This is one of the rare areas where the science is as basic as science gets, yet the ideas developed are often quickly applicable to practical problems.

It is widely recognized that the best research in these areas is being carried out in the U.S.; the principal reason for this is the close coupling between theory and experiment. Collaborative theoretical and experimental research have been responsible for many important technological developments such as the transistor and the laser. The close, accurate, and sophisticated application of fundamental theoretical concepts to experimental data has been an essential element in the phenomenal growth of condensed matter physics, both as a basis for technology and as basic science. Condensed matter theory also interacts strongly with other areas of theoretical physics. For example, the areas of many-body physics and statistical mechanics have strong connections with field theory, nuclear and astrophysical research. At present, areas such as surfaces, polymers, biomaterials, liquid crystals, phase transitions, highly non-equilibrium phenomena such as turbulence, and problems of electrons in random systems are highly active.

New computational methods have been developed recently which allow, for the first time, accurate solutions of what are essentially complex many-body problems. Many of these developments have been sparked by the increasingly urgent demands of experimentalists for theoretical interpretation of data obtained with new, sophisticated techniques. In the July 1981 "Report of the Solid State Sciences Committee, Assembly of Mathematical and Physical Science, National Research Council," one of the two main problems identified is the inadequate funding in making the new generation of scientific computers available to carry out these innovative and ground-breaking studies. That report cites as a prime example the area of "real materials and effects" where techniques are now available to calculate answers to many vital questions in condensed

matter science. These include electronic structure determinations of complex bulk solids, polymers and biomaterials, impurities and vacancy defects, superfine particles, disordered solids, amorphous materials such as modulated (superlattice) structures and sandwiches. Several of these problems focus on the unexplored regime which lies between molecules and solids which is important for the future development of electronic devices.

Problems in which increased computational capacity and innovative numerical techniques are having wide impact include: the static and dynamic structure of quantum liquids and their surfaces, the static and dynamic structure of classical liquids and their surfaces, the nature of homogeneous systems (the microscopic description of liquid-solid interfaces, for example), the problem of solitary excitations in extremely nonlinear systems, the vast area of nonequilibrium properties of solids and liquids (for example, the microscopic simulation of shock-fronts in condensed matter), the problem of turbulence in both classical and quantum fluids, the simulations of lattice defects, the electronic structure of bounded systems such as thin films, slabs, and absorbed layers.

To give some idea of the impact of increased computational capacity, it may be worth expanding on the last item by way of example. Only with the advent of the linearized, augmented plane-wave method did realistic, self-consistent calculations of the electronic levels of thin metal films, including films with ordered overlayer absorbates, become possible. Even so, very modest systems strain the resources of most computers. While further improvement in calculational schemes can be expected, these will come only from the understanding of extensive calculations of related systems, if the experience with bulk systems is any guide. Of greater interest is the construction of potential surfaces for molecules outside of a transition metal surface, a project requiring a ten to 100-fold improvement in computational capacity. Such surfaces can be the input for studying (i) the sticking of atoms on surfaces, (ii) the adsorbtion and disassociation of molecules, and (iii) the subsequent motion and reaction of atoms on the surfaces. All of the above have been the subject of primitive calculations, which could be improved if the requisite computational facilities were available. For the more distant future, there is the problem of the simulation of finite temperature surfaces, where the entropy aspect of reactions must be included.

ATOMIC AND MOLECULAR PHYSICS

Atomic and molecular science today is characterized by rapid advances in experimental technique, especially the ability to prepare and control a wide variety of atomic and molecular states. Highly ionized species, atoms in strong external fields, states with many electrons

excited, states with dimensions approximating the macroscopic, and high angular momentum states are but a few examples. The properties of such states and their interactions play a central role in atomic physics research.

The availability of powerful computers has enabled theorists to make significant contributions to the rapid growth of atomic and molecular physics during the past fifteen years. Old methods have been applied to more complex problems, and new methods have been developed for the study of many of the processes that are now amenable to experimental study or are of interest to applied physicists. In the determination of electronic structure, calculations of wave functions for atoms and small molecules have progressed well beyond the Hartree-Fock level. However, present computing power and theoretical techniques are insufficient for accurate multi-configuration calculations for heavy atoms in which relativistic effects are important. Such calculations will be required in the study of heavy-ion fusion and are needed, for example, for the analysis of the experiments searching for parity-violating atomic transitions. Further development of radiation physics and laser optics will require broader and more detailed studies of photon-atom interactions, often with highly ionized or perturbed atoms. Recent investigations of bremsstrahlung, Rayleigh scattering, Compton scattering and the photo-effect have revealed interesting new phenomena that have been explored by only a few groups with extraordinary access to fast computers. At lower energies better calculations on photoionization will be necessary to interpret the wealth of new data generated with synchrotron light (for ground-state atoms or molecules) and with infra-red or visible lasers (for highly excited states). With respect to larger systems, self-consistent-field calculations can be carried out using the local density or local spin-density approximations on polyatomic molecules, including polymers and weakly-bound clusters, and for molecules adsorbed on surfaces. Calculations by better methods will facilitate the assessment of the accuracy of these approaches, and further applications of these methods should encourage greater collaboration among atomic physicists, quantum chemists, solid-state physicists and biochemists.

In the study of atomic collisions, theory is now capable of verifying and augmenting experimental measurements on many processes in electron-atom collisions. There have been some notable successes in the theory of electron-molecule and ion-atom collisions at both high and low energies. Tremendous problems remain, particularly at intermediate energies and for collisions involving molecules in which electronic or vibrational excitation is important. Useful calculations on rearrangement collisions, energy transfer, excited-state reactions and break-up processes will require new methods and increased computing power. The successful methods should be extended to treat collisions with atoms or molecules

on surfaces. Many of the new diagnostic techniques for studying plasmas and solid surfaces involve atomic collisions, and more detailed calculations of the energy, angular distribution and polarization of scattered particles will be needed if these techniques are to be fully utilized. Studies of electron-atom and atom-atom collisions in the presence of a laser field give information not otherwise obtainable. The calculations are necessarily difficult, however, and require extensive computational effort.

Monte Carlo techniques have been introduced into the study of the electronic structure and interactions of atoms and molecules, within quantum, semi-classical and purely classical theories. Simulations are also being used to relate the macroscopic behavior of ionized gases to the properties of the individual atoms and molecules. These simulations have led to significant improvements in transport theory and to a better understanding of swarm measurements of the reactions of atomic ions. However, further studies of energy exchange between molecular ions and neutral ions and molecules are needed. Better-designed simulations would be valuable in the exploration of the many body effects that occur in dense gases, about which very little is currently understood. For example, computer simulations of three-body recombination should help to clarify many of the mysteries concerning combustion at atmospheric pressure. Sir David Bates has already used many hours on the super-computer at Daresbury in England on a preliminary analysis of this problem, but there are U.S. physicists who think they could make forceful advances if given the computer resources.

PLASMA PHYSICS

Theoretical plasma physics today includes several active subfields. Their common theme is the importance of collective processes, which dominate the interaction of ionized gases with electric and magnetic fields. Basic plasma theory addresses the fundamental natures of plasma turbulence, statistical mechanics, and magnetic-field topology. Equally important are the major applications of plasma theory to controlled fusion research, solar system and magnetospheric plasmas, and astrophysical plasmas.

Most plasma physics theory to date has evolved within the DOE-sponsored fusion energy programs. It is largely as a result of the controlled fusion programs that much of our basic understanding of plasma dynamics, statistical mechanics, and kinetic theory has developed. Through generous support for computational plasma physics, the DOE has encouraged the development of a variety of computational techniques which have become a vital part of both basic and applied plasma theory today.

Areas of basic plasma physics theory which have shown rapid and exciting advances in the last few years include strong turbulence, soliton

formation and dynamics, magnetic reconnection, plasma heating by intense beams, and single-species plasmas.

Developments in the theory of plasma turbulence illustrate a typical situation in contemporary plasma theory, the necessity of understanding strongly nonlinear collective behavior in two and three spatial dimensions. In one-dimensional turbulence, interacting solitons can play a distinctive role. However, understanding the analogous role of persistent nonlinear structures in two- and three-dimensional turbulence has proven a major challenge. Large scale computer experiments are a crucial tool in developing the required theoretical framework.

Magnetic reconnection and tearing modes are a second active area of basic plasma theory. Here, the theoretical challenge arises from the existence of two distinct spatial scales, since resistivity dominates small scale regions where magnetic field lines can reconnect. Computer simulations using the resistive magnetohydrodynamic description are playing an important role in understanding reconnection.

New mechanisms for plasma heating by intense electromagnetic waves and by charged particle beams are a third active research topic. Nonlinear wave-particle and wave-wave interactions seem to dominate the energy transfer process. Extensive use is made of computer models which follow the microscopic physics of particle heating and acceleration.

Single species plasmas (either the pure electron or pure ion plasma) are proving to have unique and exciting properties. These include the existence of unusual thermodynamic laws and the possibility of an approach to the crystalline state of a plasma under readily accessible laboratory conditions. The statistical mechanics of single-species plasmas can be studied by Monte Carlo computational methods similar to those developed for strongly coupled charge-neutral plasmas.

Numerical techniques for computational plasma physics can be grouped into two classes:

(i) Techniques which follow the microscopic electromagnetic fields and particle distribution functions are ideal for providing detailed information on the effect of turbulence, and on the growth and saturation of strong plasma instabilities. However, they have the disadvantage of being computationally slow, so that they cannot be used to model long timescale phenomena. Approaches that describe the microscopic nonlinear physics include (a) particle in cell computer simulations with up to 400,000 individual particles, and (b) codes which solve plasma kinetic equations. The latter involve partial differential equations which are functions of seven independent variables (3 space dimensions, 3 velocity dimensions, and time). The capability of present computers places a strong limit on the number of space and velocity variables that can be used in practical kinetic-equation calculations.

(ii) For a macroscopic and long-timescale description of the plasma, one uses a variety of fluid codes (which solve moment equations), or else hybrid codes which can average over fast timescales and short spatial scales. State-of-the-art in fluid computations is represented by the three-dimensional, resistive MHD calculations presently being done on supercomputers. These require many hours of computing time per run. Hybrid codes are a more recent and rapidly-evolving computational method, and represent an area of major research effort at present. It seems likely that the next few years will see hybrid physics codes synthesized into comprehensive and realistic computational models. The computing requirements are expected to be enormous.

NUCLEAR PHYSICS

The vigorous state of nuclear theory in recent years can be seen in the development of new approaches to nuclear structure and reactions which represent substantial progress on some of the most fundamental problems in nuclear physics. Examples of these developments are large basis shell model calculations, hypernetted-chain calculations of nuclear matter binding energies with comparisons to Brueckner theory, the time-dependent Hartree-Fock theory of heavy ion collisions, calculations of high energy scattering with multiple-scattering corrections, few-body methods for three and more body processes, and "fast" coupled-channel reaction formalisms. New work in intermediate energy physics has included the effects of virtual mesons and baryon resonances on nuclear structure and reactions, predictions of the pion-nucleon interaction from field theory, studies of the pion-nucleus optical potential, and derivations of properties of the nucleon-nucleon interaction from meson theory or quark models based on quantum chromodynamics.

Large basis shell model calculations continue to be important for understanding nuclear structure and reactions and as a microscopic model for developing and testing new approaches to treat finite nuclei. For example, the interacting boson approximation (IBA) has had some remarkable successes in describing the complicated spectra of rotational and transitional nuclei, and it promises to play a major role in explaining the spectra of heavy nuclei far from stability and at intermediate excitation energies. The full exploration of this model and its relation to microscopic theories of nuclei requires computing facilities not generally available in this country and much of the research on the IBA is being carried out in Europe.

Statistical moment methods are a new approach to finding quantum averages of variables in a many-body system by calculating the traces of the relevant operators in large Hilbert spaces. There is some evidence that very accurate calculations of non-average properties of nuclei can

be made by using traces in superlarge spaces of relatively low powers (4-5) of the Hamiltonian. Even so these calculations require many hours on the fastest computers available.

The coupled channel reaction (CRC) theory describes nuclear reactions by using a large number of coupled integro-differential equations. It appears to be a fruitful approach for describing reactions which proceed through a number of different channels. The computers generally available are inadequate for serious calculations of reactions on intermediate to heavy nuclei.

The theory of few-body (3-5) reactions has been developed considerably in recent years generalizing the pioneering work of Fadeev and others. Several approaches now exist which provide well-posed sets of integral or differential equations for describing three and four body systems. Very approximate and preliminary calculations have been successful almost beyond expectations, but access to the fastest existing computers would be necessary for the full testing and further exploration of these methods in nuclear structure and reaction calculations.

A variety of approaches have been developed for the theoretical investigation of heavy-ion collisions, currently one of the two major areas of experimental research in the world. These approaches deal both with the low energy and the high energy regions: (1) statistical multistep calculations using distorted wave Born approximation matrix elements, (2) coupled-channel equations, (3) time-dependent Hartree-Fock (TDHF), (4) cascade theory, and (5) hydrodynamical calculations. Although extensive calculations have been made using some of these methods, the results point to the need for computationally more demanding investigations not only to exploit these methods, but to test new theoretical ideas.

Cascade calculations using Monte Carlo methods provide virtually the only means for explaining the global features of reactions initiated by energetic hadrons. Gross features can be obtained from sampling a relatively small number of cascades, perhaps 5000, but recent experiments have probed rarer and therefore more significant events. Calculations of these events requires large statistical samples, more than 100,000 cascades, and long computer runs.

Quantum chromodynamics is emerging as the fundamental theory of elementary particles. Despite its wide acceptance, mathematical details of the theory are still in a very rudimentary stage. Not only is QCD important for nuclear processes, but also the methods of nuclear physics can be brought to bear on such fundamental problems as quark confinement, nucleon structure, nuclear forces, etc. The experience of theoretical nuclear physicists with variational techniques, Hartree-Fock, and other complex computational schemes, lends itself to QCD calculations. Some calculations have begun on simple models based on QCD, but it is quite

clear that more sophisticated explorations will require state-of-the-art computer facilities.

PHYSICS OF FLUIDS

The frontier problems of fluid dynamics can be divided into two broad areas: basic physical understanding of nonlinear dynamics, and the phenomenology of complex flows. Both kinds of problems challenge the capabilities of modern computers.

The three-dimensional character of many flows is essential to their proper understanding. Many important effects just never appear in two space dimensions. Among important three-dimensional problems, transition and turbulence involve the calculation of flows with a wide range of excited scales of motion, When the range of excited scales increases by only a factor of two (which occurs when the Reynolds number of the flow increases by about a factor of two), the computer power necessary to calculate this flow increases by an order of magnitude.

The calculation of real fluid flows usually involves several other difficulties, including complicated geometries and topologies, multi-phase flows, thin boundary and internal layers, interacting shocks, nonlinear instabilities, material interfaces, and so on. All these problems are exacerbated when extra physics, like chemical reactions, magnetohydrodynamic effects, and radiation physics, are included. Numerical simulation of such flows requires large computer memories and high computer speed to solve systems of coupled partial differential equations in several space dimensions and time. Computational methods for fluid dynamics have now matured to a point where it is realistic to expect major breakthroughs in the 80's, provided advanced computer resources are available. Some problems of modern fluid dynamics that can be effectively attacked, and possibly solved, with a new major national computational resource are:

(i) Turbulence—The highest resolution three-dimensional turbulence code now uses 512x512x512 (or over 100 million) modes to describe each velocity component; this simulation can barely be used to simulate three-dimensional inertial-range dynamics. These simulations are performed now at nearly the highest Reynolds number that can be obtained in a high-quality, low turbulence wind tunnel. With more advanced computational capability, a numerical fluid dynamics laboratory is a realistic expectation.

(ii) Transition to turbulence—The phenomenon of transition is essentially three-dimensional and so it requires great computational resources to solve. The problems of pipe flow, thermal convection, boundary layer transition, etc., can be effectively solved and analyzed numerically. Numerical simulation offers the great advantage over

laboratory experiment for these problems in that perturbations can be controlled accurately so nonlinear effects can be isolated.

(iii) High speed flows—Shocks are central to the dynamics of many physical systems. Examples include gas dynamics and implosion physics. Applications range from laser fusion physics to sonic boom propagation to astrophysics. Many features of these flows are not well understood, such as shock stability, interaction of shocks with vortical and acoustical disturbances, and multi-dimensional shock interactions.

(iv) Free-surface flows—Applications include random surface waves, nonlinear interactions of surface and internal waves, Rayleigh-Taylor instability and secondary recovery of oil. These problems involve the calculation of flows in highly complicated and convolved geometries.

One important aspect of fluid dynamic calculations in the 80's and beyond will be the presentation and analysis of multi-dimensional flow fields. Sophisticated techniques for the graphical presentation of results, including high-speed interactive graphics and movie-making capability, will be required.

ASTROPHYSICS

Astrophysics problems are characterized by the extreme dynamical range of their physical situations. For example, in the observed jets of giant active galaxies, causally coupled hydrodynamical structure is seen over a range of 10^6 in linear scale, from parsecs to megaparsecs; and from time variations, it is deduced that the actual scale of the "machine" powering the hydrodynamics is another factor of 10^3 smaller. The range of relevant densities in calculations of accretion onto a neutron star exceeds fifteen orders of magnitude. The range of relevant times in stellar evolution calculations ranges from billions of years (for the lifetime of a star of medium mass) down to milliseconds (for the relevant dynamics of that star's supernova phase).

Astrophysics problems are also characterized by the range of different physical theory that they must include. Understanding the convective solar interior involves not just hydrodynamics, but also magnetohydrodynamics, radiative transport, atomic physics (since the structure of ionization zones is of crucial importance), and probably also new understanding of the transition from order to chaotic behavior (which has recently become a unifying theme in many different areas of theoretical physics). The physics of supernovae requires an intimate melding of shock hydrodynamics, nuclear physics, and (in some cases) general relativity.

One might list other problems at the frontier of theoretical astrophysics that are particularly suited to numerical attack: star formation, accretion disk hydrodynamics, evolution of supernova remnants, galaxy formation

(with the possible importance of massive neutrinos), spiral structure in galaxies, inhomogeneous baryon number creation in the early universe, formation of the solar system, common envelope binary stars and their evolution, shock dynamics of the interstellar medium.

It should be evident that there is a great commonality of technique connecting almost all of these problems: they generally require extensive numerical integration of the partial differential equations of hydrodynamics (almost always compressible hydrodynamics), or coupled radiative transport and hydrodynamics, in two—and sometimes three—space dimensions. The hydrodynamic computer codes that perform these integrations will be a key theoretical tool of the 1980's and beyond.

GRAVITATION THEORY

Only very recently has it become possible to contemplate numerical integration of the full dynamical equations of General Relativity. Already, the field of numerical relativity has answered important questions about the behavior of full relativistic systems such as black holes, neutron stars, and gravitational waves. The nonlinear, tensor nature of the Einstein equations in several dimensions makes obtaining relevant new analytic solutions virtually impossible. Hence, this is a field where numerical solutions have a tremendous impact. Important problems include the formation and collision of black holes, the "central relativistic engine" of quasars, and highly chaotic conditions in the early universe. The solution of these problems requires not only large-scale computational power, but also the availability of high-speed vector graphics facilities (to make sense of tensor functions over two-dimensional grids).

III. COMPUTATIONAL PHYSICS: EXPERIMENTAL THEORY

THE CHARACTER OF COMPUTATIONAL PHYSICS

The computational physicist leads a triple life. First, and foremost, he or she must be fully conversant with conventional theoretical physics and the analytical techniques used by theorists. Secondly, the computational physicist must have many of the practical skills of an experimentalist. Thirdly, the computational physicist faces the unique problems of computer programming.

Specialized physics knowledge must be combined with a general theoretical background in order to 1) formulate problems to be studied on the computer; 2) have the physical insight to guide the project through the myriad of decisions that must be made, and; 3) be sure that the results have a real impact on the understanding of the problem being studied.

Equally important, tremendous analytical ability may be required to develop and understand the test cases and other tricks used to establish

the correctness of a computational procedure (both the algorithm used and the computer program which realizes the algorithm). Since the solution space of discretized equations is very different from the solution space of partial differential equations, great care must be exercised in interpreting the solutions.

Furthermore, like an experimentalist, the computational physicist has an apparatus (the computer) whose recalcitrance must be overcome and whose idiosyncracies must be separated from the object being studied. The physicist must understand all sources of error in the calculations and design the code to keep all the errors from the approximations and numerical procedures under control. These include the problems of numerical instabilities, round-off errors and, in the most extreme cases, the possibility of random hardware errors. He or she must deal with the practical limitations of limited funding, pushing the capabilities of the apparatus to its very limits to enable carrying out computations on the forefront of physics research. Often considerable entrepreneurial skills are required to obtain the resources needed to carry out the research.

Finally, computer programming introduces problems. Many experimentalists also face this problem, but for the computational theorist the programming problems have led to special difficulties, including a great deal of misunderstanding and underestimation of the role and intellectual quality of computational physics. Computer programming and debugging is, in large part, a mind-dulling, menial task, where hours and days and weeks are spent making trivial changes in response to trivial errors, or figuring out how to format the output. Yet one must be able at any moment to apply the deepest analytical skills in order to understand an unexpected result or to track down a subtle bug. The computational physicist lives in perpetual terror that some unexpected combination of circumstances will cause the failure of a program in a real-life calculation despite the best efforts at debugging and testing the program. The computational physicist may need a detailed understanding of how the computer works, and the nature of the compiler compiling the program, in order to know how to test the program thoroughly and get it to run as fast as possible.

Underlying the practical problems of computer programming is the fundamental problem of readability of programs. An analytic theorist communicates his or her results by writing a paper in which the results of his work are *derived.* Another theorist can read this paper, rederive the results in it and use both the results and the derivation in further work. In contrast, computer programs, the means by which computational results are derived, are by their very nature unpublishable, and complex programs can only be read, if at all, at enormous cost in time and effort. Furthermore, reading a large program is not enough; to achieve the understanding of

another theorist's program that is comparable to the understanding of another theorist's analytic result requires building a totally new program to carry out the same computation, running both programs, and tracking down all discrepancies.

PRACTICAL CONSIDERATIONS OF COMPUTATIONAL PHYSICS

A basic need in any computation is to be able to increase its accuracy or expand its scope at the cost of increased computing time. Unfortunately, in virtually all the problems requiring large scale computing there is a slow convergence problem: a very substantial increase in computer time is required to achieve only a marginal increase in accuracy and scope. For example, one might want to decrease the grid spacing by a factor of two in a 4-dimensional space-time calculation. This requires at least a factor of 16 increase in computing time. To increase the accuracy of a statistical simulation by a factor of 2 requires at least a factor of 4 increase in computing time. When the size of a matrix is increased by a factor of 2, the diagonalization time goes up by a factor of 8. These rapid increases often force computational physicists to search out the most powerful computing facilities available for their problems, even if this means going far from their home base and expending an enormous effort on program development. The computational physicist learns to live with inadequate accuracy in his or her approximations and with an inability to explore many problems because the computing time needed is out of reach. Nevertheless, a great deal of excitement is felt at the present time among computational physicists: there are extraordinary developments taking place (or expected) in all parts of the computing world. If these are made accessible to physicists, they can open up whole new vistas for attack by computational methods despite the barrier of slow convergence.

In order to realize the potential benefits of greatly expanded opportunities for computing, there are a number of problems to be addressed. First, computational physicists must have access to a very large range of computational facilities. The reason for this is that a single most cost-effective general purpose computer does not exist anymore. Access to the largest of supercomputers is required for problems involving very large memory and/or very fast disk access combined with vectorizable programs. Local access to medium-scale, virtual memory computers is required to enable medium size computations to be made efficiently and without a large investment of expertise, so as to encourage students, post-docs and even the most skeptical of faculty into the use of computers. (Some major and important problems can, in fact, best be run on these medium-scale machines.) Access to array processors is required because they are, in some cases, the only way to give a single user the capability

to run enormous CPU-intensive calculations which even a supercomputer cannot handle at a reasonable cost. Very much better graphics facilities are required to manage numerical output. A network is required to allow easy interchange of programs and data between researchers at different sites, and to provide remote access to the required mix of machines.

The fundamental problem of program readability must be faced. This is the problem which causes the most serious credibility problems of computational physics; it is the problem which ultimately limits the productivity of a computational physicist. A central effort of computer science has been to address this problem. The "structured programming" framework and the more recent work on "proving programs correct" developed by computer scientists can provide a very substantial first step towards dealing with the readability problem.

Physicists will have to take advantage of these and other computer science ideas and, in fact, go far beyond the present state of the art in order to cope with the problem of communicating to other physicists (and themselves) through computer programs. Documented black-boxed programs in a library cannot substitute for readable programs because it is impossible to explain what a complex, steadily evolving program does, except through the program itself.

In the past, the central campus facilities of many colleges and universities have provided inexpensive computing and software support for computational physics. With a few notable exceptions, this is no longer the case. The trend towards minicomputer systems for experimental science, financial pressures on universities and their computing centers, and the very much increased demand for social science, administrative and student computing have all downgraded the service and increased the expense for physics computing. In addition, most central facilities are becoming increasingly inadequate technologically for computational physics. Thus, in most cases computational physicists will need either separate facilities or shared departmental facilities for their local computing needs. However, there are special cases where good service is available from a central facility, and, in these cases, adequate support is needed to enable physicists to use these facilities and still be in communication with other physics-based sites.

RELATIONS TO INDUSTRY

The availability of high speed computers has been particularly important in applied physics. The increasing ability to study complex phenomena makes it feasible for academic physicists to make major contributions to industrial problems. For example, the recent advances in statistical mechanics, that grew from ideas generated in part by

elementary particle physicists, may be applicable to percolation problems associated with tertiary recovery of oil.

Within industry and the Federal laboratories computer use is growing more rapidly than in academe. For example, in aerodynamics, computers have already supplanted wind tunnels in the intermediate stages of airfoil design, at significantly reduced cost. In the development of new and expensive devices, computational physics is playing an increasing role in the design of components, such as target pellets for laser fusion, and in the prediction of the behavior of full scale systems by scaling the known properties of small models or prototypes. In some areas simulations are essential in safety studies and the training of operators. In these applications of computers it is important that the basic physical processes are described accurately and that theoretical physicists work together with engineers and mathematicians to improve the quality and efficiency of the computer codes.

All this implies an increasing cross-fertilization between basic research by university-based computational physicists and physicists and engineers working on simulation of practical problems in national laboratories and industry. Students trained by university physicists move into positions in national laboratories and industry, bringing with them the training in the present state-of-the-art in tackling the complexity of problems being simulated. A number of the problems now being studied in industry are challenging and broad enough to be legitimate subjects for basic research; industrial workers in these problems should be integrating themselves into the basic research community through exchange of seminar speakers, publication in common journals, etc., as well as being integrated at some level into a computational physics network.

CAPACITY AND CAPABILITY

The distinction between *capacity* and *capability* of a computing network (or aggregation of computer resources) is a vital one. Capacity is a measure of the volume of work or the number of jobs a specified configuration can accomplish in a unit of time. Increased computer capacity is obtainable by the addition of similar computers, the acquisition of additional peripherals, or even by improvements in operating system performance or in communications between different computers on a network.

Capability, on the other hand, measures the highest degree of complexity or largest size of *one job* that a computer or network can process in some specified time frame. Capability must be referenced to a time frame, since even a minicomputer (with e.g., an infinite supply of magnetic tapes) can perform the most complex jobs if given years of time. The standard reference time-frame is usually taken as "overnight

turnaround"; for shorter times, computer utilization would not be averaged over the 24-hour cycle, while for longer times human efficiencies quickly go to zero. (Of course turnaround should be very much less than overnight, e.g., minutes, for jobs much smaller than the capability limit. Conversely, a few jobs on array processors might usefully run for many days.)

Unlike capacity, capability cannot be increased by the addition of a similar computer. Capability can be increased by the addition of a more powerful computer, by the addition of larger or faster memory to an existing computer, by adding a specialized processor (array or attached processor), or (in some cases) by adding peripherals such as advanced graphics devices.

From a funding agency's point of view, scientific research productivity behaves very differently as a function of capacity than it does as a function of capability. As a function of capability, it is strongly thresholded: a certain level of capability is necessary even to begin to work on the frontier problems in a given area of theoretical research. (In Section IV we define a useful threshold for the short-term.) Above threshold, as capability is increased, there is a gradual increase in scientific output as new problems become accessible to numerical solution.

Productivity as a function of capacity, on the other hand, should be regarded as almost linear over a very large range. As capacity increases, more researchers can be supported in computation-based theoretical research, or more resources can be allocated to the best researchers, or any such combination. Since capacity enters so linearly, proposed increases in capacity are highly susceptible to cost-effectiveness optimization. There is a very large dynamical range of computer sizes and configurations which can provide increased capacity, and there are large dollar savings to be made by a sensible optimization plan.

It is essential that any new thrust in computational facilities provide both increased capacity for the very large number of problems to be addressed,and increased capability, to attack the frontier problems in computational physics. But this is not all. Both capacity and capability must be delivered to the user and there must be sufficient flexibility that users can select the optimum resources for the problem at hand. Further, the facilities and their operation must promote increased communication among users, sharing of software, and ease of use.

IV. WHAT IS TO BE DONE NOW?

As stated in Section I, the state of NSF support for computational physics is dismal. In this section we outline steps that can be taken immediately and that will lead to substantial near-term improvement. These steps must also form the basis for a stable long-term development of computational physics.

A THRESHOLD FOR USEFUL ACTION

Now, in 1981, one can identify a threshold level of access to computer facilities which is required for a meaningful attack on some fraction of the important problems for which techniques of solution already exist. This threshold will not be sufficient five, or even three, years from now, but it would be adequate today for a start.

A typical large code might require 1-4 million words of memory (4-32 megabytes); given overnight turnaround on a major run of such a code, an "active" researcher might require 2 to 20 CDC 7600-equivalent-hours per week to run and develop codes in a "cleverness-limited" mode, where human effort and machine effort are in rough balance. Counting the time to formulate problems, write up results, or simply think, a theoretical investigator might spend 20 weeks per year in this "active" mode.

It is straightforward to estimate the resources required to support one such researcher today. A rough estimate, valid on both medium-scale computers and supercomputers (e.g., at DOE's Magnetic Fusion Energy Computer Center [MFECC]) is $200 per NSU (nominal service unit; the NSU-per-year equivalent of various computer systems is given in Appendix B). The requirements given above correspond to a range of 60-600 NSU's per investigator per year. The cost is thus in the range $12K - $120K. This is not out of line with present funding patterns in other areas of physics, especially experimental physics. The bottleneck, we see, is not just lack of money, but also lack of an efficient mode of delivery of computer resources to the theoretical user. It is for this reason, in part, that we feel that much can be done *right now* towards providing for the future health of computational physics.

We propose a distributed computational physics system in which a number of computational facilities are tied together on an information network. The principal node of the network will contain a supercomputer providing the greatest possible capability. The other nodes will contain medium scale computers to provide capacity, advanced graphics devices for an effective man-machine interface and in some cases array processors for additional capability. The central facility will also play a major role in developing software and software standards, the design and expansion of the system, and system management and resources allocation.

THE NETWORK

The Subcommittee stresses the extreme importance of networking of computer facilities and users (i.e., providing the user with fast and transparent communication links to computer capacity and capability that may be geographically remote).

The difference between a computer network and a disconnected set of computers must not be underestimated. A network (and its associated software and human components) provides economies of scale in software development; it allows scientific collaborations on computational projects between participants who are physically remote from each other; it gives the user a choice among a variety of facilities that may be more suitable to his problem than the facility he happens to be physically closest to; it allows researchers at smaller universities to participate on an equal basis with larger institutions. We would not seriously contemplate the establishment of a new initiative in computing for theoretical physics if it did not include the establishment of a supported communication network.

There are a number of specific advantages to the distributed picture. Communication costs are, on the average, reduced by placing the physical facilities near the greatest source of their use. The network's total capacity degrades more gracefully under occasional hardware failure, and the probability of catastrophic loss of capacity is reduced almost to zero. There are often observed to be large efficiencies which occur when services such as operator time or software development are provided informally at local nodes at little or no cost (and can then filter out to other nodes). Local management and consulting services, especially when informal and on a small scale, are almost invariably more responsive to special user requirements or situations. Certain requirements (such as graphics processing, three-dimensional displays, and convenient editing on a modern screen editor) require very high baud rates, and so can *only* be implemented efficiently at a local site.

The single most important advantage of a distributed network is that it can be changed gradually. It can be increased or decreased in capacity without severe dislocation of most users. It can accommodate changing technological developments without becoming "locked-in" to a super-computer of some one particular technological state.

The goals in establishing a network for the Computational Physics System are (1) to provide uniform access by all federally-funded, university-based scientists to all levels of computational resources including a supercomputer, (2) to facilitate exchange of programs and allow for the development of software libraries, (3) to provide access to data bases, and (4) to reduce wasteful duplication of effort and encourage collaboration on significant computationally demanding problems.

The operational requirements for a communications system to accomplish these goals are the following: (1) accessibility by public telephone using any terminal at speeds from 110 baud to at least 1200 baud; provision for nodes consisting of a large number of terminals and fast line printers; (2) support for computer to computer communications using the internationally accepted X.25 protocol and communication bandwidths of at least 56 kilobaud for connections between major nodes; (3)

reliable and accurate transmission of data; (4) controlled access by authorized users with an accurate accounting of connect time and traffic charges by user and by site; (5) international links to major computing centers throughout the world and accessibility by authorized users on a world-wide scale.

These requirements may be satisfied in a number of ways from lease of commercial services (such as Telenet or Tymnet) to lease of dedicated lines and connection equipment. All computational physicists will then be able to access the network at least via dial-up access on Tymnet or Telenet. The minimum equipment required is a display terminal, printer, and modem, all able to operate at 1200 baud.

THE PRINCIPAL NODE

One of the main functions of the principal node of the Computational Physics System is the provision of the most advanced scientific computer available. Sophisticated computer models of physical phenomena require the most capable machines, i.e., the so-called "supercomputers." If the System starts out with a 1970 vintage machine, e.g., CDC 7600, it will have failed in its objective. Physicists with the most advanced computer models will not make use of the System, and will try to meet their needs elsewhere.

On the other hand the principal node should not be a laboratory for developing and testing new concepts in processors. Such an activity is important and has its place on the network, but the principal node must provide a stable environment on an established state-of-the-art computer with mature software i.e., a stable operating system and efficient compilers.

At the present time there are at least two computers that may meet these objectives. One has been on the market for three years, and is now available in 0.5, 1.0, 2.0, or 4.0 million words of 64 bit central memory. Another manufacturer's machine is just being released this year, with memory also available up to 4 million words. To either of these machines (or ones of similar specifications) must be added a suitable amount of fast, on-line storage. The one million word supercomputers at the Magnetic Fusion Energy Computer Center (MFECC) and at the National Center for Atmospheric Research (NCAR) have 16 fast disks and a similar number will be required for the Computational Physics System principal node.

At a centralized facility, adequate archival file storage capacity must be provided. The network communication links are generally used only for input and output of data. Codes, restart data, data bases, etc., should reside at the principal node. Disk storage should be provided for on-line file retrieval for files used frequently. Longer term storage is provided by conventional magnetic tapes, but a mass store device should be considered for intermediate term storage.

Other important functions of the principal node are software coordination and software development. The principal node should have a staff which promulgates software standards and guidelines and which assists users of other locations on the network to access the principal node and to develop codes which will take maximum advantage of the capabilities of the principal node supercomputer. The staff will also play a major role in determining overall architecture of the network and upgrades in the network required to keep pace with problems to be solved. By concentrating these functions at the central node, one obtains economies of scale which will lead to much more cost effective implementation of these functions.

THE LOCAL NODES

Distribution of balanced computational capacity at the local nodes is as important as the provision of maximum capability of the principal node. The number of computational physics problems to be addressed is such that the capacity of the principal node can quickly be oversubscribed if it is the only resource provided. Medium scale computers with virtual memory operating systems and provision for at least 8 Mbytes of physical memory can provide capacity (though not capability) with roughly the same cost effectiveness as the supercomputers. The decentralization of this capacity provides many other benefits. As mentioned earlier, the overall system performance degrades gradually when equipment fails. The system provides flexibility for future upgrades and allows such improvements to occur gradually without major impacts on a given budget year.

To the scientific user, the local nodes make possible modes of operation and scientific investigations not possible via remote access to a central node. Since a typical local node is shared among a much smaller number of users than a central node, there is much greater ability to adapt computer operations to the requirements of the user. The machine may be dedicated or adapted to a special project for some period of time with much less disruption than would be the case with supercomputers at the central node. Further, the allocation of resources at the local node can be much more flexible than at the central node; some of the computation time can be allocated without formal and lengthy justification. Thus, there can be a mechanism for pursuing innovative projects to the point of determining the need for a thorough investigation, possibly then using the supercomputer of the central node.

A local node also provides for much higher input/output rates to the user than can be provided with remote access to a central facility. This is important when one is using a modern screen oriented editor and even

more important for examining and manipulating the results of computations (performed either at the local node or the central node).

Any commitment to increased computational resources must be accompanied by an expansion of graphics facilities to help manage the increased data flow. The human mind is not capable of scanning piles of computer output in any meaningful way. However, if the same functions are presented graphically (as lines or surfaces), the eye can quickly scan for smoothness, global structure, and spatial relationships. The graphics format also acts as an index to the data. If one sees an area of interest, one can then go back to the tabular output for more precise data.

ARRAY (ATTACHED) PROCESSORS AND OTHER SPECIAL FACILITIES

Local nodes also provide the ability to obtain cost-effective increases in capability through the installation of array or attached processors. Supercomputers and medium-scale machines each have about the same capacity per dollar, as we have already mentioned. For presently existing systems with thoroughly debugged software, these are the principal options available. However, it has already been demonstrated (e.g. at Cornell) that an attached processor can provide a very much increased capacity per dollar, by one or even several orders of magnitude, than either a supercomputer or a medium-scale machine. Furthermore, the capability of attached processors will increase very substantially when deliveries of powerful 64-bit attached processors begin. Attached processors can serve two functions. First, they can provide a single researcher with a capacity for considerably exceeding our "initial threshold" (section IV). Secondly, as the attached processor software matures, such machines can be used to augment the basic capacity of the network.

There needs to be a range of attached processor capabilities available on the network. These should range from very inexpensive systems with small memory size (64,000 words) and short word lengths to much more deluxe systems with large memory (1.5 million words or more) and a 64 bit word-length.

To realize the full power of these array processors, it is essential that compilers for these processors with full optimization be completed and debugged as rapidly as possible. It is also essential that users of the network be provided adequate documentation and consultation concerning the use of array processors.

There are other special facilities that should be made available to the network from an appropriate local node. These should include some specific large programs, such as the computer algebra program MACSYMA, which may require a specially configured system. There will also be special purpose libraries, particularly libraries of programs which can be and have been permanently debugged and completely

tested, for example, a library of programs to compute Clebsch-Gordan coefficients and other group-theoretical constructs.

IMPLEMENTATION AND MANAGEMENT

The establishment of a national system for computational physics including a communication network, computing facilities at a principal node and local nodes will require a strong, dedicated organization in order to achieve success. The emphasis in designing the system should be meeting the needs of the user community—the computational physicists.

The computational physics network that we propose has much in common with the highly effective Magnetic Fusion Energy network (the MFECC). That network also has a principal node with a supercomputer, regional nodes supplying resources complementary to the principal node, and an effective management which delivers computing capacity and capability in a cost-effective manner. We take this as an "existence proof" that a computational physics network is technically feasible.

The institution selected for managing the central computing facility should be one with proven experience in serving a scientific computing community. The quickest and most economical way of establishing such a center would be to put an additional supercomputer and file storage capacity at an existing center which is already serving remote users, e.g., NCAR, MFECC. At the same time, however, one must be careful not to place existing management in an impossible conflict between an established user group and a new one.

The institution which manages the principal facility should also manage the communication network. The management of the local nodes will vary from one university to another, but there will have to be uniformity in responsibilities since the computers at these nodes must interface effectively to the entire network. The managers of these nodes should meet periodically to plan hardware and software developments.

The central facility should have a user's advisory committee to advise NSF on hardware/software needs.

The allocation of time on the central supercomputer to the various user groups should be the responsibility of NSF. These allocations should be made on the basis of written proposals which are refereed by peer review procedures. The implementation of the allocations should be the responsibility of the center managment.

SUMMARY

In summary, a balanced, effective Computational Physics System requires that adequate attention and funding be provided to the local nodes as well as the central node. One cannot have one or the other, but

must have both, tied together in a centrally managed network, in order to make effective progress in computational physics.

PROJECTED COSTS

The subcommittee was charged with estimating the costs of the expansion of computational facilities we recommend. The following figures are crude estimates in 1981 dollars.

The capitalization costs for a central site, not including the supercomputer (which will be leased), is about $5M. A regional node's capitalization is about $0.5M. The annual operating budget can be estimated as follows:

Central Site	*$K*
Lease of Supercomputer	2800
Maintenance of Supercomputer	350
Maintenance of File Storage	200
Communication Network	875
Supplies and Services	500
Manpower Cost (35 FTE)	2000
Building Lease	100
Total	6025

One Regional Node	*$K*
Staff, Maintenance, Supplies	300

V. GUIDELINES FOR DEVELOPMENT

Once the initial computational physics system is in place, its capacity and capability must increase rapidly. Computational physicists will be driven to machines with vastly increased capacity and capability under the double pressures of slow convergence and the coming within reach of new threshholds. To maintain the health and vigor of computational physics will require both substantial funding increases and rapid incorporation of new technology in the network.

Computing systems at all nodes of the network will undergo a life cycle. To start with, immature software will be tested by those few users who must use the capacity or capability of the new system regardless of cost in programming effort or frustration. As the software matures, there must be a fairly long period where the system operates in a high reliability 24-hour access mode, providing stability for long term program development and an accumulation of software designed for the system. This middle period is likely to extend well past the time when the system

is technologically obsolete, except where totally software-compatible upgrades can keep the technology of the system current while maintaining a mature and stable system. Most systems will eventually die due to obsolete technology and discontinued maintenance.

As different nodes will be in different stages of life cycles of different technologies there may be shifts in the location of the largest capacity and capability from time to time. Thus, it is essential that the same standards of service demanded of the principal node be required also of other major nodes in the mature stage of their life cycles. It is also important that time allocation schemes allow users a choice of several entry points in the balance between mature software and cheap CPU cycles.

As computational physics expands its capability, the requirement for efficient interaction of the physicist and the computational system becomes more and more essential. Experience has shown that certain features of the "computational environment" crucially affect a scientist's productivity. These include operating systems, utilities (editors, text processors, debug, file management, etc.), "instant" graphics, sophisticated languages, data bases, and transparent access to other computers in the network.

We feel strongly that continued funding in these areas be adequate to achieve a reasonable balance between raw increases of computer power, and increased abilities to put that power to work. We have some detailed recommendations in these areas.

SYSTEM SOFTWARE AND COMPATIBILITY

Systems software should be designed to minimize the time needed to develop the software for a given project and to maximize the ease of extending and modifying existing software. This is vitally important to attract high quality physicists to computational physics and to make it possible to carry out much more ambitious projects than can be tackled in non-optimal program development systems.

Portability, documentability and quality of design are the basic requirements for systems software. Portability means that all computers should look the same to the user as far as possible. This simplifies cooperation between physicists at different institutions and minimizes expense of software development costs. In the context of a network of computers, as many machines as possible should run the same operating system. Machines not running the standard operating system should at least have the same software utilities available—e.g., text editors or text formatting systems. To achieve this level of portability requires that all systems software be written in a portable language. Utility programs can

then be compiled on any machine with a compiler for that language. An example of a nearly-portable operating system is the UNIX system.

Systems software documentation should be short, complete, easy to read and to reference. A physicist should not need to become a systems expert in order to take full advantage of the system. The power of systems software is achieved through generality, regularity and simplicity, not by complexity. Languages, including the command language, should be as self-documenting as possible—user supplied documentation is frequently out of date. Standards should be specified to prevent the proliferations of incompatible versions.

The operating system should be structured to minimize the difficulties of managing multi-file programs or sets of programs. Network requirements involve dealing with multiple machines. Almost all commands should be executable across networks. Program updating and source code control facilities are essential to the orderly development and maintenance of large projects. Utility programs greatly increase the effectiveness of an operating system. Most user time is devoted to file editing so that a high quality screen editor can greatly improve productivity. Associated tools are required for text processing, particularly for reducing the output of programs to manageable size. Text formatting programs allow effective use of the computer for writing reports and monographs.

DATA MANAGEMENT FACILITY

As the capacity of the Computational Physics System increases, it will be essential to increase both the capacity and the capabilities of graphics devices at the local nodes, in order to manage the ever rising data flow. An efficient method is to establish a Data Management Facility (DMF) at each local node which has more than a certain threshold of capacity. A typical DMF would consist of four subsystems: Controller, Display, Storage, and Hard Copy. The Controller is a small computer which acts as a data receiver from the other nodes or local computer and as a switchyard within the DMF. The Display subsystem contains as many graphics display devices as required by the research group. The Storage subsystem consists of disks or tapes on which images or numerical data are stored. Hard Copy devices include line printers, plotters, or screen copying devices.

Such a loosely coupled system can be upgraded as demand requires. Adding capability can be achieved by updating only one subsystem at a time. We stress that one critical feature of the system is data rate to the screens in the DMF. These must be high enough (greater than 9600bps) that a picture appears instantly upon request. This is because managing data graphically may require the scanning of hundreds of images per session at the console.

A scientist would typically use the DMF for post-processing of calculations run at either the local node or the principal node. The numerical output from a computation at the principal node must be sent to the local node where it is stored for high speed output to the graphics terminal. The user then uses the terminal to produce various graphical representations of the data. As he scans through his results, he can quickly make hard copies of those pictures he finds most interesting or generate different, more informative, representations of the data based on the results that were actually generated by the computation. The goal is a deeper understanding of the physics and the elucidation of the most promising ideas for subsequent computations and investigations.

Finally, we come to the issue of archiving images. One approach is to use microfiche. However, large capacity microfiche production devices are very expensive. Therefore, it may be most cost effective to have only one or a few regional production centers to which output is shipped electronically across the network. The hard copy microfiche could then be mailed back to the user. Another concept is "electronic microfiche" in which the images are stored on disk to be retrieved when desired. Which system will be preferable is an issue to be studied further. It will also be important to have movie or video-cassette producing capability at some node. This is one of the most natural ways to represent time dependent solutions to differential equations.

PROGRAMMING LANGUAGES

FORTRAN has been the standard language for computational physics for over 20 years. The size and complexity of computer applications has increased enormously during that time, yet FORTRAN has remained almost static. As a result, FORTRAN seems inadequate as a future language for forefront problems in computational physics. Fundamental gaps are in the areas of control-flow, user-defined data structures and pointers, recursion, mechanisms for sharing information between sets of routines and for hiding unnecessary information. Portability of FORTRAN codes across machine architecture is difficult. Because of the enormous economic investment in FORTRAN codes, FORTRAN will continue to be an important language in computational physics. However, modern computing languages are beginning to be used in an increasingly significant way by physicists. The efficiency requirements of computational physics limit the list of likely languages. Possible candidates might be ADA or C. In the long run, languages will evolve much further, adapting more and more of the flexibility of human languages to the precision required for computer codes.

DATA BASES

Many calculations require input from experimental data or tables of quantities which are derived from the analysis of experimental data. In nuclear physics, for example, phase shifts or other experimentally-determined amplitudes for nucleon-nucleon scattering, photopion reactions, etc. are needed for calculations of nuclear reactions or structure. Progress in a number of areas of theoretical physics would be accelerated by access through a network to a well-maintained data base of such quantities. The data provided in this way can be more timely, more complete, and more correct than those published in journals. In fact, journals are reluctant to publish more than the most summary experimental results, and these are often not sufficient for further theoretical investigations. The scale of experimental research in this country demands that considerably more attention be given to the establishment of data bases accessible by remote users.

THE FUTURE OF SUPERCOMPUTERS

The continued rapid development of computer power through the end of this century seems assured. The natural evolution of circuits and chips from LSI to VLSI to ULSI (ultra-large scale integration) concurrent with changing technologies from silicon to GaAs to cryogenic Josephson junction implies that impressive new circuit speeds will be achieved.

New architectures may play an important role in computers to be developed over the next decade. Pipeline machines (like the CRAY-I, TI ASC, CDC Cyber 203) are already playing an important role. Here arithmetic is done as on an automobile assembly line in which major operations are segmented into smaller steps and several operations proceed at once. Parallel machines (like the ILLIAC and DAP) have several (or even many) arithmetic units all controlled by a single control unit, much as an orchestra is led by its leader, thereby breaking up a large computation into subunits.

Other architectures under study include data-flow machines, many-element processers, etc. The data-flow architecture, as developed by J. Dennis, is characterized by asynchronous program control in which data words carry the information as to what operations are to be performed on them.

Many-element processors (such as DENELCOR's HEP or the NYU ultracomputer project) consist of separate processors, each having its own arithmetic and control unit, that are linked to a common set of memory modules via a high-throughput network. The architecture might possibly emerge as a framework for achieving very high capability, by combining large numbers of mass-produced VLSI processors with very large, cheap, semiconductor memories.

The structure of resources employed in computational physics in the early 1980's must be flexible enough to allow for the exploitation of the major advances in architecture which are expected for the late 1980's and beyond. Computational physicists who are at the forefront of their field must be intimately involved in the design of the computer architecture and of the accompanying system software. They have a detailed understanding of the computers that they use, and will know from practical experience what new directions ought to be fruitful for the areas of pure and applied computation.

The spectrum of needs of theoretical physicists for enhanced computing facilities overlaps to a considerable extent with those of applied mathematicians and other physical scientists. As seen in Section II and III, the growth of computational physics will increase the amount of cross-disciplinary research and assist the experimental physicist in the analysis and interpretation of data. Thus, the possibility of jointly-planned facilities should be part of the long-term planning. However, the crisis within theoretical physics is so serious that an initiative of rapid improvement should begin at once.

APPENDIX A

BIBLIOGRAPHY

(1) The National Resource for Computation in Chemistry, Phase 1: Evaluation and Recommendations (Report of an Ad Hoc Review Committee, published May 1980).

(2) Department of Energy FY1981-1985 ADP Long Range Plan (Office of ADP Management, November 1979).

(3) "An Assessment of Trends in the Solid State Sciences" (National Research Council, July 1980).

(4) "Future of Nuclear Science" by Ad Hoc Panel on the Future of Nuclear Science, 1977.

(5) "The Application of Computers to Controlled Thermonuclear Research" (United States Atomic Energy Commission, July 1973).

(6) "Long Range Plan for Nuclear Science" (excerpts from a report by the Nuclear Science Advisory Committee, July 30-August 4, 1979).

(7) "Draft II - Astronomy Survey Committee: Report of the Panel on Data Processing and Computational Facilities (May 19, 1980)."

(8) "A Report on Computer Needs for Nuclear Theory" (William M. MacDonald et al., Seminar Workshop at the University of Maryland [SWUM], April 27, 1979).

(9) "Report on the Results of a Questionnaire on Computer Facilities Needs for Nuclear Physics Research" (Richard D. Koshel and William M. MacDonald, SWUM, May 14, 1979).

(10) Workshop on Computational Astrophysics (NASA-Ames Research Center, February 5-6, 1980).

(11) "Magnetic Fusion Energy and Computers: The Role of Computing in Magnetic Fusion Energy Research and Development" (Office of Fusion Energy, DOE, October 1979).

APPENDIX B

COMPUTER EQUIVALENCE TABLE*

Manufacturer	*Model*	*Relative Capacity*	*Nominal Capacity NSU's)*
CRI	CRAY-1	4.0	40,000
CDC	STAR-100	1.8	18,000
CDC	7600	1.0	10,000
IBM	3033	.95	9,500
IBM	360/370/195	.95	9,500
UNI	1100/82	.88	8,800
UNI	1100/44	.88	8,800
CDC	CYBER 175	.70	7,000
IBM	360/168-3	.58	5,800
IBM	370/168-1	.53	5,300
IBM	360/91	.51	5,100
UNI	1108	.29[1]	2,900
IBM	3031	.24	2,400
CDC	6600	.22	2,200
CDC	CYBER 174	.22	2,200
IBM	370/158	.20	2,000
IBM	360/75	.20	2,000
CDC	CYBER 74	.18	1,800
IBM	IBM 360/65	.16	1,600
DEC	PDP KL10	.14	1,400
CDC	CYBER 73	.12	1,200
DEC	VAX 11/780	.12	1,200
CDC	6400	.11	1,100
HIS	66/40	.11	1,100
IBM	360/44	.07	700
DEC	PDP KI10	.06	600
DEC	PDP KA10	.04	400
IBM	360/50	.036	360
DEC	11/70	.036	360
DEC	11/45	.033	330

[1] Specifically UNI 1108 complex at SLA.

*SOURCE: Department of Energy FY1981-1985 ADP Long Range Plan