Needs and Opportunities for the National Resource for Computation in Chemistry (NRCC): Report of a Workshop (1976)

Pages
167

Size
8.5 x 10

ISBN
0309334667

Planning Committee for a National Resource for Computation in Chemistry; Assembly of Mathematical and Physical Sciences; National Research Council

Needs and Opportunities for the National Resource for Computation in Chemistry (NRCC)

Report of a Workshop

Planning Committee for a
National Resource for Computation in Chemistry
Assembly of Mathematical and Physical Sciences
National Research Council

NATIONAL ACADEMY OF SCIENCES
Washington, D. C. 1976

-0209

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

Support for this project was provided by the National Science Foundation under Contract No. NSF-C310, Task Order 303, the U.S. Energy Research and Development Administration under Contract No. E(11-1)-2518, and the National Institutes of Health under Contract No. SB 3208(a)-75-C590.

Available from
Office of Chemistry and Chemical Technology
National Research Council
2101 Constitution Avenue
Washington, D.C. 20418

PREFACE

The National Research Council's Planning Committee for a National Resource for Computation in Chemistry (NRCC) of the Assembly of Mathematical and Physical Sciences organized a workshop supported jointly by the Energy Research and Development Administration (ERDA), the National Institutes of Health (NIH), and the National Science Foundation (NSF). It was held at the National Academy of Sciences Summer Study Center in Woods Hole, Massachusetts, from July 13 to 16, 1976. The main purpose of the workshop was to identify computational problems of high priority for progress in chemical research to which a NRCC could initially provide unique computational support not now widely available to the chemical community.

To accomplish this task, panels of five to six scientists in each of seven fields of research were selected. The members were chosen to be broadly representative of their individual fields of chemistry and of the constituencies (governmental, industrial, and academic) that would use the NRCC once it is established. Nominations were solicited from a number of sources, including members of the NRCC Planning Committee, the Executive Committee of the Assembly of Mathematical and Physical Sciences, the Chairman of the Assembly's Office of Chemistry and Chemical Technology, members of the Chemistry Section of the National Academy of Sciences, and the chairmen of the topical area panels. Each of the participants is expert in computer usage in at least one field of chemistry.

Advice from the chemical community was sought in advance of the workshop. First, the purpose of the workshop was publicized in a news article in *Chemical and Engineering News* (May 10, 1976). Second, letters were distributed by the panel chairmen to a broad representation of colleagues, soliciting topics for possible discussion and consideration at the workshop on the role that the NRCC might play in their specific fields of chemistry.

This report is intended as a resource for the prospective NRCC Policy Board and Director for use in the planning and implementation of the NRCC's initial program. Equally importantly, it should serve to catalyze additional ideas and proposals generated by individual scientists before the NRCC gets under way.

The initial plan of the workshop included parallel working groups in each of eight fields identified in previous reports (Appendixes A and B) as pertinent to the NRCC. It proved impossible to organize a representative five-member panel on the newly emerging field of energy systems and still meet the

time schedule for the workshop. Nevertheless, in the reports on quantum chemistry, statistical mechanics, chemical kinetics, and nonnumerical methods, recommendations are made relevant to research on energy systems. The Planning Committee recommends that a workshop in this field be organized early in Phase I of NRCC operations.

From the inception of the NRCC, the importance to the chemical community not only of the hardware and software but of the human-resource functions of such an organization has been emphasized. As we focus more specifically on detailed operation, it becomes apparent, as seen in the chapters prepared independently by the seven working panels, that the human-resource function is highly valued. The NRCC is more than a network of large and small computers, minicomputers, and microcomputers.

Each of the panels has emphasized the need for interdisciplinary workshops of a magnitude not currently being satisfied by national society meetings, symposia, Gordon Research Conferences, or seminars. Each has emphasized the benefit that would accrue to chemistry as a whole and to research specialists in the various fields through new avenues of communication that could be established by the NRCC. The establishment of means of communication across fields would be particularly significant in an age of increasing specialization in science.

Although the NRCC Planning Committee assumes overall responsibility for this report, the individual chapters are mainly the work of the panels representing the respective fields of chemistry. These chapters present an abundant selection of recommended services and scientific activities. In order that each retain its own flavor, they have not been edited by the Committee with respect to scientific language or content. Neither has any attempt been made to blend the reports nor to eliminate areas of overlapping interest. Consequently, the panel reports vary widely in organization, technical language, and style.

The Committee expresses its appreciation to the panel chairmen and members for their contributions to this report. It also acknowledges assistance from many corresponding contributors (listed in Appendix E), who provided suggestions prior to the workshop or comments on the draft report. Their assistance in the Committee's task does not necessarily imply that they endorse all aspects of the report.

PLANNING COMMITTEE FOR A NATIONAL RESOURCE FOR COMPUTATION IN CHEMISTRY

*Jacob Bigeleisen, University of Rochester, *Chairman*
Bruce J. Berne, Columbia University
F. Albert Cotton, Texas A&M University
Harold A. Scheraga, Cornell University
*Howard E. Simmons, E. I. du Pont de Nemours & Company, Incorporated
Lawrence C. Snyder, Bell Laboratories
*Kenneth B. Wiberg, Yale University
W. Todd Wipke, University of California, Santa Cruz

William Spindel, *Executive Secretary*
Martin Paul, *Staff Consultant*

*Unable to attend the workshop

WORKSHOP COMMITTEE

Lawrence C. Snyder, Bell Laboratories, *Chairman*

Panel on Chemical Kinetics

James T. Muckerman, Brookhaven National Laboratory, *Chairman*
Thomas F. George, University of Rochester
James C. Keck, Massachusetts Institute of Technology
John C. Light, University of Chicago
Robert E. Wyatt, University of Texas at Austin

Panel on Crystallography

David H. Templeton, University of California, Berkeley, *Chairman*
Wayne A. Hendrickson, Naval Research Laboratory
*James T. Ibers, Northwestern University
Carroll K. Johnson, Oak Ridge National Laboratory
Keith D. Watenpaugh, University of Washington

Panel on Macromolecular Science

Frederick T. Wall, Rice University, *Chairman*
John C. Curro, Sandia Laboratories
Marshall Fixman, Yale University
Robert L. Jernigan, National Institutes of Health
Roy A. Scott III, Ohio State University

Panel on Nonnumerical Methods

Charles F. Bender, Lawrence Livermore Laboratory, *Chairman*
Richard J. Feldmann, National Institutes of Health

*Unable to attend the workshop

Charles E. Klopfenstein, California Institute of Technology
Jacob T. Schwartz, Courant Institute of Mathematical Sciences
Dennis H. Smith, Stanford University

Panel on Physical Organic Chemistry

Gilles Klopman, Case Western Reserve University, *Chairman*
Kendall N. Houk, Louisiana State University
David Pensak, E. I. du Pont de Nemours & Company, Incorporated
Calvin D. Ritchie, State University of New York at Buffalo
Martin Saunders, Yale University

Panel on Quantum Chemistry

Frank E. Harris, University of Utah, *Chairman*
Thomas H. Dunning, Los Alamos Scientific Laboratory
Charles W. Kern, Battelle Memorial Institute
Keiji Morokuma, University of Rochester
Howard S. Taylor, University of Southern California
Arnold C. Wahl, Argonne National Laboratory

Panel on Statistical Mechanics

John A. Barker, IBM Research Laboratory, San Jose, *Chairman*
Harold L. Friedman, State University of New York at Stony Brook
Keith E. Gubbins, University of Florida
William G. Hoover, Lawrence Livermore Laboratory
Theodore A. Litovitz, Catholic University
William W. Wood, Los Alamos Scientific Laboratory

Observers from Federal Agencies

O. William Adams, National Science Foundation
William R. Baker, Jr., National Institutes of Health
Edward F. Hayes, National Science Foundation
Jerome Hudis, Energy Research and Development Administration

NRC Staff

William Spindel, *Workshop Director*
Martin A. Paul, *Workshop Consultant*
Norman Metzger, *Liaison Representative, Office of Information*
Peggy J. Posey, *Administrative Assistant*

1
INTRODUCTION AND OVERVIEW

The purpose of the workshop was stated by Jacob Bigeleisen, Chairman of the NRCC Planning Committee, in his letter of June 17, 1976, to the participants:

> Our main purpose is to identify computational problems of high priority to progress in chemical research that should be undertaken with support provided in the initial program of NRCC and that could not feasibly be undertaken with the kinds of support presently available or in prospect elsewhere. This assembled information will be a valuable resource to those charged with planning and implementing NRCC's initial scientific program. It will be issued in the form of a report from the National Academy of Sciences, to be made broadly available to the chemical and related scientific research community as well as to the funding agencies concerned.

As background material, panelists were provided in advance of the workshop with copies of previous Academy reports (Appendixes A and B) relating to the need for, and providing a recommended organizational structure and proposed initial budget for, the NRCC. Worksheets outlining general and specific questions proposed as guidelines (Appendix C) were enclosed in the same mailing to the participants, with the suggestion that they be replicated and distributed to interested colleagues for completion and later discussion at the workshop.

In a final charge to the panels at the opening session, Lawrence Snyder, Chairman of the Workshop, suggested that each panel identify five early activities of high priority in scientific research or service for the NRCC. He asked the panels to prepare for each such activity:

1. A description of the activity, as specific as possible;
2. An estimate of the associated requirements in staff, computer time, and special hardware;
3. A statement of the consequent benefits to chemistry;
4. A discussion of why NRCC is uniquely appropriate for the activity.

He emphasized to the panelists that their suggestions of scientific studies would be understood as examples, set forth to aid in the planning for the NRCC and to provide models to stimulate the generation of additional proposals from individual scientists. The actual scientific program will be established at a later date by NRCC management: its Policy Board, Program Committee, and

Director. The strength of that program will depend on an abundant flow of imaginative proposals from the chemical research community.

The following chapters of the report were largely prepared by the respective panels working separately. Several plenary sessions attended by all the participants provided an opportunity for exchange of ideas and for the emergence of some consensus on perceived needs common to all areas. In this sense, the workshop represented an initial approximation to a users-group meeting for the NRCC. In this chapter, these common needs are briefly reviewed, along with several other matters generally bearing on computational aspects of chemical research.

During the past decade, the deficiences perceived in computational resources have changed, and the concept of an NRCC has evolved accordingly in several successive studies. At the outset, the major problem was financing larger-scale computations. The cost of computing has decreased and become less pressing, except for the largest computational tasks; but the computer programs used by chemists have increased vastly in complexity and in the concomitant costs of software creation, documentation, and maintenance. An analysis of the programs most in demand through the Quantum Chemistry Program Exchange (QCPE) maintained at Indiana University illustrates this trend. Richard Counts of QCPE has provided a list of the 16 programs (see Appendix D) requested most often during the past 20 months from the QCPE library of about 300 programs. These programs are complex major systems; the most popular of them, Gaussian 70 (QCPE #236), developed by John Pople and collaborators, consists of over 13,000 symbolic cards.

The workshop participants agreed that the NRCC, in addition to providing major computing hardware, should provide the needed manpower to ensure that existing chemical computing software be made more efficient, accessible, and widely usable. Moreover, there was a commonly expressed desire to cooperate in the creation of needed software and to set up means to facilitate such cooperation among fields of chemistry and within them, as well as between chemists and computer scientists. Through workshops and other interactions among users, the NRCC should promote agreement, without stifling innovation, on unification, standardization, and clarification of codes used in chemical computation, thereby making them more readily comprehensible and transferable. The NRCC is expected to function catalytically to increase the productivity and long-term benefits of our present large investment in computing methods of chemistry.

Data bases of various kinds are essential to many of the computations utilized in chemical research. Specialized existing data bases should be readily accessible to NRCC users, but the NRCC should not undertake, with its limited resources, the maintenance, updating, and editing of data except for those generated by its own projects. The NRCC should require that the most reliable input data be utilized for all computations carried out under its auspices. In this regard, advice and cooperation from the Office of Standard Reference Data of the National Bureau of Standards and other authoritative sources should be sought.

The panel reports show that there are many topics of common scientific interest. Potential energy curves and surfaces, both intramolecular and intermolecular, are deemed basic to several areas, as are improved programs for classical molecular dynamics and force-field calculations. Studies of large molecular systems are challenging to quantum chemists, macromolecular scientists,

crystallographers, and investigators in statistical mechanics. For many of the problems, real-time interactive graphics facilities can greatly improve the productivity of the investigators; several panel reports highlight the need for a versatile graphics facility at NRCC.

The importance of access to the NRCC by geographically remote users was stressed in the earlier report of the Planning Committee for a NRCC (Appendix B). The potential for access to a large central computing facility via a national computer network was demonstrated at the workshop through several portable terminals connected by ordinary telephone lines, through the Boston node on the TYMNET network, to the SUMEX-AIM system at Stanford University. The workshop participants strongly endorsed the recommendation that the NRCC develop ample facilities for remote user access.

The scientific projects presented in the following panel reports, taken together, exceed the recommended initial operational level of the NRCC severalfold. This abundance presages a flow of imaginative proposals from the chemical community from which the management of NRCC can select a truly strong research program.

Written suggestions received from a large number of correspondents (Appendix E), before the workshop and also in response to a first draft, contributed materially to shaping this report. Although all suggestions were not incorporated, the documents themselves will be made available to the NRCC's management and to others interested.

As stated in the earlier report of the Planning Committee (Appendix B), a computer having the speed and memory of a CDC 7600, an IBM 370/195, or their equivalent is the minimum with which the NRCC could fulfill its function. Throughout this document, these are defined as fourth-generation computers.

2
CHEMICAL KINETICS

INTRODUCTION

An understanding of chemical kinetics, ranging from such elementary processes as inelastic excitation of a molecule in a bimolecular collision to complicated reactions of polyatomic molecules in condensed phases, is important to many areas of science, technology, and industry. Within this field, a number of problems are well enough understood theoretically to permit the generation of important data computationally. Some are in the forefront of theoretical research, with theoretical knowledge currently being reduced to practical computational algorithms. Many are so complex that only simple, approximate, theoretical approaches have yet been developed.

Within those areas of chemical kinetics that depend significantly on computation, the NRCC can make an extremely important contribution by fulfilling the following functions:

1. It should survey those areas of kinetics in which computer codes to solve certain classes of problems already exist (e.g., classical trajectories, coupled rate equations, and semiempirical potential energy surfaces) and acquire, test, evaluate, clean up, document, and make available to both experimental and theoretical users standard codes of high utility and reliability
2. It should promote via the mechanisms of workshops, visiting scientists, and active collaboration with workers in the field the creation of new codes for computation of key rate processes. Examples of areas in which new computational codes are needed are electronic excitation in collisions, and inelastic and reactive quantum scattering processes, both exact and approximate.
3. It should coordinate and support, by means of workshops and transfer of technology, computational efforts in evolving areas of kinetics.
4. It should provide resources for benchmark calculations of selected dynamic processes.
5. It should acquire and make available selected compilations of rate constants.

Listed below are four areas of kinetics in which the existence and activities of the NRCC during Phase I, with relatively moderate resources, can have appreciable effect on a large segment of the chemistry research community:

Although the subjects listed are assigned highest priority and are discussed here in detail, a number of others deserve computational support, such as gas-solid interactions and collisions, reactions in condensed phases, and electron-molecule scattering processes.

The total requirements estimated to initiate work in the four high-priority research fields at the NRCC are four staff members, four workshops per year, and about 600 hours of computer time per year. With the cooperation of outside scientists, this small research effort can have an effect out of proportion to its magnitude.

I. POTENTIAL-ENERGY SURFACES USABLE FOR DYNAMICAL CALCULATIONS

Since the starting point for any calculation of the nuclear motions of a system of atoms, whether classical or quantum mechanical, is some convenient and accurate representation of the potential-energy surface (or set of surfaces), one project for the NRCC in chemical kinetics should be to provide users with the tools for obtaining such representations. A satisfactory representation will allow the accurate and facile evaluation of the potential energy and, ideally, also the gradient of the potential energy, at arbitrary nuclear geometries. Clearly, a small set of tabulated points from *ab initio* electronic-energy calculations, no matter how accurate, does not satisfy this criterion. One possible approach to a satisfactory representation is to fit some (assumed) flexible function to the computed points; another is to employ a semiempirical electronic-energy calculation at any nuclear geometry for which the potential is required. Although the latter approach would benefit from the existence of accurate *ab initio* results for comparison and/or calibration purposes, its implementation does not depend on such comparisons, and, therefore, it has applications to a much wider variety of chemical systems. The fitting methods might also be used to represent semiempirical surfaces if such an approach would increase the speed of evaluating the potential. Other advantages of the semiempirical surface approach for dynamical purposes are that both ground- and excited-state surfaces are obtainable from solutions of the same secular equation and coupling elements between them are readily calculable.

Considerable progress both in fitting tabulated points on a potential-energy surface and in generating semiempirical surfaces has been made recently by a relatively small number of workers, each with a unique methodology. The tools that they have developed are not widely available to other workers and in most cases have been developed only to solve some specific problem. One major

impediment to the widespread use of semiempirical surfaces of the valence-bond diatomics-in-molecules variety is the amount of algebraic labor required to derive matrix elements of the electronic Hamiltonian (in terms of atomic-state energies and diatomic potential-energy curves) from an input set of valence atomic orbitals on each atom. This labor could be circumvented by the application of nonnumerical methods for the computer solution of algebraic problems.

The NRCC should bring together the various experts in this field to identify the best elements in each of their approaches to representing or calculating surfaces semiempirically, and develop general programs incorporating these best features. The NRCC should also attempt to catalyze improvements in existing methods through the transfer of technology from other fields. More specifically, the NRCC should not only make available to users programs for multidimensional spline interpolation and other polynomial fits of points on potential-energy surfaces but should also sponsor a critical evaluation of the relative merits of such procedures and develop recommendations as to which approach should be used for representing various types of potential-energy surfaces. The NRCC should also bring together in a workshop investigators with experience in calculating semiempirical valence-bond diatomics-in-molecules surfaces for various specific systems, to develop a general program for calculating such surfaces for any system of three or four atoms having a total of up to eight or nine explicitly treated valence electrons. Such systems include $O + CO$, $O + OH$, and $CO + H_2$, which are of considerable importance in both experimental and theoretical chemical kinetics and in related technological fields such as combustion and air pollution.

Clearly the kinetics-oriented activity in potential-energy surfaces should be coordinated with the NRCC's activities in quantum chemistry, especially to take advantage of the NRCC's unique capability for the cross-fertilization of fields. The NRCC's activities in nonnumerical methods should also prove to be an extremely valuable resource in programming for computer solution the algebraic procedures previously mentioned for the diatomics-in-molecules method.

The project of representing potential-energy surfaces for dynamical use is envisaged to be completed within Phase I of the NRCC. Its demands are primarily for staff time to participate in and to coordinate the development of computer codes, and only secondarily for computer time. Annual staff requirements are estimated to be one full-time member, one full-time equivalent visiting scientist (actually about five different individuals), and one workshop. About 20 hours of fourth-generation computer time would be required to debug, test, and document the software produced.

Summary of Resources Required:

Duration of Project:	2 years
Annual Resources:	
NRCC Staff	1 full-time
Visiting Scientist	1 full-time equivalent
Workshops	1
Fourth-Generation Computer Time	10 h

II. SMALL-SYSTEMS DYNAMICS

A. Classical Trajectories

The integration of classical trajectories to describe a chemical reaction process assumes that one or more of the degrees of freedom obeys Newtonian mechanics. The most popular and successful use of the classical trajectory approach is based on the separation of nuclear and electronic degrees of freedom, whereby nuclear degrees of freedom are represented by classical trajectories propagating on a single electronic (ground-state) potential-energy surface. The simulation of a particular experiment is achieved through an appropriate Monte Carlo selection of initial conditions, such that a trajectory is integrated for each initial condition. As in most of chemical kinetics, this type of calculation depends on the availability of a potential-energy surface. However, given a suitable surface, a number of excellent program packages exist that carry out Monte Carlo-averaged classical trajectory calculations. The systems studied have included inelastic atom (ion)-diatom processes ($He + H_2$, $Li^+ + H_2$, etc.), reactive atom (ion)-diatom processes ($H + H_2$, $F + H_2$, $H + I_2$, etc.), more complicated exchange reactions ($K + CH_3I$, etc.), and dissociative and associative processes ($M + H_2 \rightarrow M + H + H$, etc.).

The role of the NRCC in single-surface classical trajectory calculations will have five aspects, of which four are service-oriented and the fifth relates to the initiation of new research efforts. First, the NRCC could conduct a survey of existing software, followed by detailed comparisons and critical evaluations of the different codes in use. The ultimate result will be a program library with comprehensive instructions for each program. Second, the NRCC would investigate possible hardware for optimizing the efficiency of trajectory calculations, such as hybrid digital-analog simulators. It would also establish a graphics display capability to depict trajectories as functions of suitable variables (time, coordinates, etc.). Third, the NRCC would establish a data bank for rate constants. This objective can be accomplished in collaboration with other existing banks. Fourth, it would provide staff time equivalent to one-half man per year, preferably a PhD chemist with experience in trajectory calculations and an interest in computer software development. While the staff would be responsible for developing the three aspects of the role of the NRCC described above, it would also be available in an advisory capacity to outside users, particularly to experimentalists who have had little experience in integrating classical trajectories. In this way the NRCC would provide a service to outside users who need results of trajectory calculations for certain systems but do not have the time nor the need to learn all the input operations for running a trajectory program. It is estimated that 10 hours of fourth-generation computer time will be required during the first year for the foregoing projects.

The fifth aspect of the role of the NRCC would be the promotion of new research efforts. This could include trajectory studies of new systems involving multivalent species (such as $O + O_2$) and species present in high-temperature gas dynamics, combustion, and flames. The testing of new ideas will also be encouraged, such as semiclassical approaches that use classical trajectories (or analogous quantities) as input to theories that include

quantum effects (tunneling and resonances). This fifth aspect will require an estimated five hours of fourth-generation computer time.

Extension of the classical trajectory approach to include the dynamic coupling of two or more potential-energy surfaces is an important area of research in chemical kinetics. Any process involving the transfer of electronic energy or ionization must proceed through the interaction of an excited-state surface with the ground-state surface or with another excited-state surface. The simplest and most straightforward application of classical trajectories to multisurface problems is through the assumption of localized interactions, where a trajectory makes a localized transition between surfaces. Such an approach has been carried out for a few systems ($H^+ + H_2$, $Ar^+ + H_2$, $F + H_2$, etc.). Since the various theories behind these calculations are recent and still in the developmental stages, the NRCC can play a valuable role by organizing a workshop in which the several outside research groups (four or five) engaged in this research can meet to test and categorize the most accurate and useful techniques with respect to particular chemical systems. This workshop can also provide incentive for extending the theories in a computationally attractive form to situations for which transitions between surfaces are not localized, such as may arise because of a particular manner of coupling between electronic and nuclear angular momenta. An estimated 10 hours of fourth-generation computer time will be required during the first year, along with the remaining half-time of the staff member from the single-surface project.

In addition to the testing and categorizing of various theories, production runs will be encouraged by outside users on selected new processes, such as

$$O(^3P,^1D) + H_2(^1\Sigma_g^+) \rightarrow OH(^2\Pi) + H(^2S),$$

which involves the coupling between a $^1A'$ and a $^3A''$ surface when spin-orbit interaction is included. During the first year, this production running will require an estimated 20 hours of fourth-generation computer time. For all the work at the NRCC on classical trajectories for the dynamics of small systems, the estimated initial annual requirements are summarized as follows:

Annual Resources:

NRCC Staff	1
Workshops	1
Fourth-Generation Computer Time:	
In-house Development	25 h
External Productions	20 h

B. Quantum-Mechanical Studies of Collision Dynamics

A significant role can be played by the NRCC in quantum-mechanical studies of both nonreactive and reactive collisions in systems containing no more than four interacting atoms. For these few-particle collisions, the theory of inelastic (nonreactive) energy transfer is now rather well developed, and computer programs to integrate the quantum-mechanical equations of motion (in both

the coupled-differential-equation and integral-equation formulations) are being developed by several different groups. Generally, these programs are designed to treat vibration-rotation-translation energy transfer in atom-diatomic molecule (and in a few cases diatomic molecule-diatomic molecule) collisions on a single electronic potential surface. The NRCC can provide an important user service by testing and evaluating the available inelastic scattering programs and by making them available to the user community. Access to these codes would be extremely valuable to many scientists who need to interpret experimental data (or who want to predict results) of energy transfer collisions, which occur in many laser, atmospheric, and combustion problems. The current state of the quantum theory and computer codes necessary for studying inelastic phenomena on multiple interacting electronic surfaces is not nearly so well developed as for the adiabatic electronic processes discussed above. The NRCC can function as a catalyst to promote the development of efficient general programs for studying the quantum dynamics of simultaneous electronic-vibrational-rotational-translational energy transfer.

Several groups are currently working hard to develop both the theory and associated computer codes required in quantum studies of atom-diatomic molecule reactions on a single electronic surface. (Accurate results have recently been obtained for the hydrogen atom-hydrogen molecule exchange reaction.) The desired results include tables of reaction cross sections (or probabilities) from specific vibration-rotation states in reactants to specific vibration-rotation states in products. (These state-to-state reaction cross sections could in turn serve as input to calculations of thermal rate constants for use in kinetics modeling studies listed below.) The computer programs being developed are based on several alternate formulations of the reactive scattering problem, including the numerical integration of close-coupled differential or integro-differential equations and numerical application of R-matrix theory (developed originally for nuclear reactions). Application of these codes to the $H + H_2$, $F + H_2$, and F + HD (electronically adiabatic) reactions are currently in progress. An essential early activity for the NRCC in this area would be to sponsor workshops to enable collaboration among the several independent groups in developing and comparing their respective numerical technologies.

Because of the large number of equations that must be simultaneously integrated in the close-coupling formulation, fast computers with extensive high-speed core are required. Here the NRCC can play a unique and highly significant role by providing access for external users to such computers. Benchmark quantum calculations on several important reactions (e.g., $F + H_2$, F + HD, $O + H_2$) will be vital in assessing the accuracy of model dynamical theories, including classical mechanics and approximate quantum models. Quantum reaction models (e.g., centrifugal decoupling approximations) are currently being developed which require the integration of a much smaller number of coupled equations than in the "exact" quantum approach. It may thus be possible within the next several years to perform reasonably accurate quantum calculations (without excessive computational effort) on a variety of reactive systems. Computational hardware at the NRCC will permit quantum studies that range in scope far beyond the small number of benchmark studies mentioned above. In addition, as exact and approximate quantum scattering codes are developed for these single electronic surface processes, rapid extension of these numerical

techniques to multisurface processes should be possible, e.g., the chemical laser processes

$$F(^2P_{3/2}, {}^2P_{1/2}) + H_2 \rightarrow HF + H$$

and the flame reactions

$$O(^1D, {}^3P) + H_2 \rightarrow OH + H.$$

It is expected that 500 hours of computer time per year on the CDC-7600 or its equivalent will be required for the quantum studies mentioned in this section, and one staff scientist will be required to maintain this program. The overall initial annual requirements for this work at the NRCC on quantum-mechanical studies of collision dynamics are estimated as follows:

Annual Resources:

NRCC Staff	1
Outside Consultant	0.25
Workshops	1
(5 Research Groups)	
Fourth-Generation Computer Time:	
In-house Development	10 h
External Production	500 h

III. COLLISION DYNAMICS IN INTENSE RADIATION FIELDS

In recent years, the coupling between gas-phase dynamical processes and lasers has generated keen scientific and technological interest. The development of gas-phase lasers "pumped" by chemical reactions, the recognition of the importance of collisional relaxation processes in gas lasers, the demonstration that selective excitation by lasers may lead to very large differences in reaction rates, and the utilization of lasers in double-resonance experiments to study collisional relaxation, all demonstrate how an understanding of both the collision dynamics and the interaction of the atomic and molecular species with the radiation field is necessary. Possible technological applications of such knowledge abound, ranging from laser fusion and laser isotope separation to spectroscopic detection of air pollutants and "directed" chemical synthesis. For research in this area, state-to-state cross sections for inelastic and reactive collisions and radiative transition probabilities are required, and the accumulation of such information has been proposed in the preceding section.

Still more recently, it has been recognized that strong enough laser fields exist to require the simultaneous consideration of collision dynamics and the radiation field. Examples are multiphoton dissociation and radiation-enhanced inelastic or reactive collisions. In the latter, for example, one can enhance the cross section for a collisional energy transfer by a laser field of the appropriate frequency, e.g.,

$$A^* + B + h\nu \rightarrow A + B^*.$$

The theoretical description of such processes is difficult, since neither the collision nor the radiation field can be treated as a small perturbation. Thus one must deal with the coupled dynamics of the particles in the laser field. Formulation of the problem in terms of collision in quantized radiation fields is currently an active topic of research. The description of such processes must start with accurate and convenient representations of the photon field and electronic-energy surfaces involved, including matrix elements for both radiative- and kinetic-energy coupling. One must then have accurate methods of solving the resulting dynamical problem, using either quantum close-coupling methods or multisurface classical trajectory methods.

Since a number of experimental and theoretical research groups are interested in this general problem and have devoted considerable resources to its elucidation, the appropriate role for the NRCC would seem to be supportive and catalytic rather than in a direct, but necessarily minor, research effort. The NRCC would have both expertise and programs for solving the types of coupled equations arising from this research, as well as expertise, experience, and facilities for computing the required potential surfaces and matrix elements, and should play an extremely valuable role as a resource for the research groups involved. This area in particular, involving as it does chemists, physicists, and specialists in laser optics, should benefit immensely by the transfer of computational technology from the area of collision dynamics--methods of solution of close-coupled equations or semiclassical trajectory problems--to the problems of dynamics in the radiation field.

To this end we recommend that the NRCC organize a workshop devoted to the computational problems in this area and assign one member of the scientific staff part-time to monitor progress and facilitate the creation and transfer of required programs. A specific example for detailed study might be the laser-induced electronic-energy transfer in the collision:

$$\mathrm{Br}(^2P_{3/2}) + M \xrightarrow{h\nu} \mathrm{Br}(^2P_{1/2}) + M.$$

Annual Resources:

NRCC Staff	1
Outside Consultant	1
Workshops	1 or 2
Fourth-Generation Computer Time	10 h

IV. KINETICS MODELING

The modeling of transport processes and chemical reactions in gases and liquids is important for a wide variety of practical problems related to energy generation and conversion, materials processing, and environmental pollution. Among systems of current interest are laminar and turbulent combustion in

engines and burners, the chemistry of NO and O_3 in the atmosphere, lasers and laser-induced chemical reactions, the aerodynamics of shock waves and fast expansions; magnetohydrodynamics (MHD) and thermionic energy conversion, isotope separation, gas-surface interactions and catalysis, and oscillating reactions in chemical and biological systems.

The theoretical treatment of such problems typically involves two types of calculation for which large-scale computers and sophisticated programming are required. The first is the determination of the local equilibrium properties of complex chemical systems undergoing slow reversible changes as in a slow expansion. The second is the integration of large sets of coupled nonlinear differential rate and transport equations governing the evolution of systems undergoing fast irreversible changes such as in a laminar flame.

Equilibrium Calculations

At present, several reasonably general and fairly reliable programs exist for calculating equilibrium properties in complex systems containing both gaseous and condensed phases. Such programs are an essential element in virtually all calculations involving chemically reacting systems. The NRCC could perform a valuable service by assembling, evaluating, and documenting already existing programs and making them available in easily usable form to interested investigators. In addition, the NRCC could take the lead in developing needed faster, more flexible equilibrium programs. Many investigators have stressed the importance of access to a continually updated base of reliable, computer-compatible thermodynamic data for such equilibrium calculations.

Integration of Coupled Rate and Transport Equations

Although several programs now exist for solving the systems of ordinary differential equations describing the kinetic behavior of chemical reactions *neglecting* diffusion, turbulence, etc., they are not widely available to the chemical community, e.g., reaction-mechanism chemists. In addition, most of the programs for solving the coupled partial differential equations describing the more general rate *and* transport problems encountered in a wide variety of flow, combustion, and laser systems are highly specialized and have not been sufficiently well documented to make them easily transferable. Both kinds of problems are plagued by "stiffness" of the differential equations, so that solving them by brute force is frequently inefficient and expensive. There is a need to collect, evaluate, document, generalize, and distribute the existing programs, as well as to develop improved methods for dealing with the problem of "stiffness."

Since such calculations frequently involve 50 or more species and a corresponding number of rate and transport equations, they can easily strain the capacity of the largest computers. A great deal of attention has therefore been devoted to the development of simpler approximate methods. The methods most commonly used are the global models and the steady-state approximation. A constrained-equilibrium approximation, which has many advantages over the steady-state approximation, has also been recently proposed. The review and

critical evaluation of the accuracy and efficiency of these methods relative to one another and to techniques for integrating the full set of rate and transport equations would be an excellent activity for the NRCC.

An extremely important fundamental problem that might be supported by the NRCC in this area is the determination of the laminar flame structure for a general hydrocarbon. This problem involves simultaneous consideration of both chemical reactions and diffusion and, depending on the complexity of the hydrocarbon and the number of species included, can easily take up the capacity of the largest computers. The calculations could be modeled on those already performed for the oxidations of H_2 and CH_4; the results would be of considerable value not only for practical applications in combustion but also for obtaining insight into the important rate-limiting steps in hydrocarbon combustion. Such insight would facilitate the development of much needed approximate models of hydrocarbon combustion similar to those based on the constrained-equilibrium assumption or global-kinetic schemes.

Widespread interest in kinetic modeling is anticipated on the part of not only the scientific groups active in developing the field but also a user community interested in applications. Upward of 10 outside research groups might become actively involved with the NRCC. Most of the original research would be done by these groups. One full-time NRCC staff member assisted by the equivalent of a quarter-time consultant would be required to carry on supporting activities at the NRCC. These would include the assembly, evaluation, refinements, and documentation of outside programs; the establishment and maintenance of communications between research groups and the user community; and the organization of an annual workshop on kinetics modeling. The estimated annual requirements for a meaningful initial effort are as follows:

Annual Resources:

NRCC Staff	1
Outside Consultant	1/4
Workshops	1
Fourth-Generation Computer Time:	
Within NRCC	10 h
Outside NRCC	100 h

3
CRYSTALLOGRAPHY

I. INTRODUCTION

The analysis of chemical structure by diffraction techniques, applied mostly to materials in the crystalline state, is producing an enormous volume of information fundamental to the understanding of all sorts of chemical phenomena. Those working in this field need day-to-day access to computer facilities of appreciable power and capacity. Many of them find that the progress of their work is limited by lack of access to the largest computers or lack of sufficient budget for computation. The contemplated initial scale of the NRCC cannot satisfy any substantial fraction of this need, and one should recognize the continued importance of other sources of support. The Resource can have an impact by supporting some advanced calculations that could not be done with the facilities otherwise available to the investigator.

Crystallography is a broad and active science, and this brief report can be neither comprehensive nor exhaustive concerning meritorious projects. Below are listed four types of calculation that we consider to be representative of the most urgent needs. A fifth item is an information-exchange service, which would be of great value to the chemical community.

II. CRYSTALLOGRAPHIC REFINEMENT OF MACROMOLECULAR STRUCTURES

Crystal structures of molecules with up to about 50 atoms may be refined to yield models with precision in bond lengths of 0.001 Å to 0.003 Å. The refinement techniques employed, while quite routinely applied to the analysis of such small molecules, cannot yet be applied routinely to macromolecular structures. As a result, accuracies in atomic coordinates of, for example, proteins have generally been no better than 0.5 Å. Better accuracy is essential if one is to make detailed geometrical interpretations of the mechanisms of biochemical reactions of these molecules. The successful refinement of a protein (e.g., rubredoxin) illustrated that models and their accuracy can be vastly improved by careful application of classical refinement techniques. The refinement problem, however, is not trivial, and various methods of refining

macromolecular structure are under active investigation. The computing requirement for this research is large. As a result, most refinement methods in current use reflect efforts to minimize computer time at the expense of failing to extract all the chemical information from the available experimental data. Many macromolecular structures have been or are currently being refined by differential-synthesis or real-space refinement techniques that are effective to a degree. However, powerful least-squares methods have not generally been utilized because of the huge computational power required.

The least-squares refinement of even a very small protein (500 atoms or 2000 parameters and 10,000 observed reflections) currently requires about 30 minutes of CPU time on a CDC-7600 or its equivalent per cycle of refinement, with a highly optimized program neglecting the vast majority of the off-diagonal terms of the normal matrix. While it is difficult to forecast the number of cycles required to reach convergence, ten cycles of refinement starting with a reasonably correct initial model would probably be the minimum. These cycles must be interspersed with periods of reassessing the model, a time-consuming procedure calling for value judgments that probably could be effectively aided by interactive graphics. The need for the reassessment periods could be reduced and the speed of convergence increased also by proper incorporation of other physicochemical information into the refinement process. This information could consist of constraints on interatomic distances and angles and inclusion of interatomic potential functions. Inclusion of such auxiliary information will be essential in the refinement of most protein structures.

While the methods for refinement will probably be devised outside of the NRCC, the Resource could provide a unique service by affording (1) the computational facilities to test the methods and (2) the opportunity for interaction with individuals more familiar with the theoretical aspects of interatomic forces and numerical methods. The computational time required for program development would be small, while the computational time for testing methods on one or more macromolecules could be sizable. A test on one macromolecule consisting of 2000 nonhydrogen atoms with data at 2.0 Å resolution might require 20 hours of CPU time on the CDC-7600 or its equivalent. The auxiliary manpower requirement could be small, consisting primarily of consultation on numerical analysis or computational methods. A workshop of about 10 persons combining experience in the areas of crystallographic refinement techniques, interatomic forces, and constrained optimizational procedures would have great value.

If more efficient and powerful crystallographic refinement methods could be developed and proven out, the impact would be substantial not only on the macromolecular crystallographic community but on the entire field of structural chemistry. This program might require 50 to 100 hours per year on the CDC-7600 or its equivalent in Phase I of the NRCC.

III. GRAPHICAL INTERPRETATION FACILITIES

A major feature of crystallographic computing is the mass of data that must be examined throughout the course of the crystallographic analysis. The many calculations leading from the experimental diffraction data to the final structural model involve partial, trial structures which may turn out to be incorrect or only partially correct. The usual algorithmic numerical refinement procedures inherently are capable of making only relatively minor changes in the trial solutions. The major changes must be introduced by other means involving recognition of the imperfect portions of the trial solution and substitution of a more appropriate trial model. Ideally, the computer should have enough "machine intelligence" to recognize and correct all possible problems, but in reality these recognitions and modifications must be made by the crystallographer. Unlike the computer, the human cannot mentally store, process, and comprehend more than a few numbers but has the capacity for recognizing patterns and pattern anomalies in two- and three-dimensional visual presentations.

An essential feature for the NRCC is an efficient and versatile device for converting numbers into images, which can serve as an interface between the numerical computations and the user. Graphic presentations have proven indispensable in at least two specific levels of a crystal-structure analysis. The first is the contoured three-dimensional electron-density map, and the second is the ball-and-stick molecular structure model. One can visualize other levels in the structure analysis where graphical presentations could lead to better-informed judgments.

Graphical displays of electron-density maps have generally been contoured sections of the three-dimensional functions. In protein crystallography, these contoured sections are often viewed in a Richards optical comparator, which, by means of a half-silvered mirror, optically superimposes images of the density map and the wire model under construction. The electronic replacement for this device displays portions of the density-map contours and the protein model simultaneously on a cathode-ray screen and has real-time interactive rotational hardware for adjusting the model to fit the density contours. The graphical display system must be capable of drawing a sufficient number of line segments to depict adequately the contoured map and molecular model without distractive flicker. It must also be capable of making rotational transformations of the images. The current generation of graphical displays is adequate for this task. Software development is in progress at a number of sites, and the programs could be transferred to the NRCC to serve as a nucleus for further development.

This facility could operate in such a manner that a protein crystallographer could visit for a few weeks with a set of experimentally phased diffraction data and a known amino acid sequence. The person would calculate a density map and then use the interactive graphics device to fit a model accurate enough to be subjected to refinement by means of the main NRCC computer. Such a procedure could save several months of conventional model building.

The computer graphics facility envisaged would require one full-time professional staff member with expertise in this type of hardware and software. The minimum cost of special hardware is estimated to be $120,000 in 1976 dollars. Such a facility would be essential for the anticipated crystallographic computations at the NRCC and would have applications in many other chemical computations.

IV. PHASING METHODS

The so-called "phase problem" has been central to crystallographic structure analysis since the inception of this science. Only the amplitudes of individual x-ray reflections (structure factors) can be directly observed; the phases are lost in the process of observation. Both amplitudes and phases are needed before an image of the crystal structure can be synthesized by Fourier transformation from the diffraction pattern. Phases can often be determined by taking advantage of the dominance that heavy-atom scattering can have in the diffraction pattern, or by other methods. However, a more general solution derives from phase relationships among the structure factors. These relationships stem from physical constraints implicit in the nonnegativity of the electron density and the atomicity of the scattering matter. Direct methods exploiting this type of phase information have been developed to the point where they are routinely successful in many crystal-structure determinations. However, the effectiveness of current methods decreases as the size of the structure increases. What is needed now are methods that can reliably provide phase information for the important class of structures that contain more than about 100 nonhydrogen atoms.

New attacks on this problem are in prospect with the availability of large computing power. For example, new joint probability distributions have been derived for the values of certain linear combinations of phases. These are made conditional on the observed amplitudes associated with the reflections involved in the chosen relationship and in a "nested" series of neighboring reflections. Thus, probability distributions have been derived in space groups P1 and $P\bar{1}$ for quartets of phases dependent on seven amplitudes. Related distributions are being developed for additional space groups, for more extended neighborhoods, and for higher-order relationships (quintets, sextets, etc.). As the order of a relationship increases, the number of such possible relationships also increases but the accuracy of estimates of individual phase combinations tends to decrease. The computational problem is one of systematically locating all possible phase relationships and computing the probability distributions for each one. Estimates of expected values and associated variances can then be derived from these distributions. Provided that a sufficient number of sufficiently reliable relationships can be found, the values of the desired set of individual phases will then be overdetermined and can be extracted by least-squares techniques.

This is only one of several possible new advances in phasing methods that can be envisaged for testing with the aid of a powerful computer. Among the others are multisolution techniques, more sophisticated symbolic-addition

techniques, and much fuller use of the systematic expression of the positivity criterion found in the Karle-Hauptman determinants (J. Karle and H. Hauptman, Acta Cryst. 3:181, 1950). A principal goal in this area would be the demonstration of feasibility. Efficiency in computational algorithms could be sought later, once effectiveness in solving large structures has been demonstrated. Structure determinations at the intermediate level of complexity, between 50 and 300 atoms, are often onerous or even totally intractable with present methods. Those that have been solved have usually demanded many months or even years of work. Success in such a project would have immediate impact in a number of areas of structural chemistry. Problems affected include such important molecules as oligopeptides, antibiotics, and hormones. It is even a worthy goal to consider the possibility of extending the applicability of direct phasing methods to molecules the size of small proteins.

Another application of phasing methods is possible, if the necessary computational resources are available, in protein crystallography. Methods have recently been proposed for refining and extending a limited set of phases determined by conventional techniques such as isomorphous replacement. The objective is to produce a clearer image of the structure. Phase refinements of this kind have already been performed on rubredoxin and insulin by using electron-density modification techniques, Sayre's equation, (D. Sayre, Acta Cryst. 5:30, 1952) and the Karle-Hauptman determinants as the source of new information. The results look promising but have not been evaluated on an unknown structure. If such a procedure could be convincingly shown to enhance the ease of interpretation or perhaps even to render previously inscrutable features understandable, protein structure determination could be appreciably facilitated.

Any anticipated proposal for research on phasing methods would make little demand on the NRCC beyond computer time. The investigator would most likely provide his own programming, although interaction with numerical analysts may be beneficial. Work could be carried out either remotely or on site. It is difficult to predict the computer requirements, but a total of about 25 hours per project should provide a fair preliminary test of a particular method.

V. DIFFRACTION PHYSICS

Large computational problems of a nonstandard nature are encountered in the interpretation of electron-microscope images of crystals at the highest possible resolution and in other applications of x-ray or electron diffraction when dynamical theory is necessary. Electron-microscopy imaging is an increasingly important method of structural research for everything from viruses and proteins to inorganic oxides and simple metals. Low-energy electron diffraction can give information on the atomic structure of surfaces and, in particular, on the initial stages of reaction of gas atoms with solids, with important consequences for the understanding of surface processes including catalysis and oxidation. The necessity of including multiple scattering effects in the diffraction calculations implies that very lengthy computer cal-

culations are needed to provide adequate interpretation of observed diffracted beam intensities.

For n-beam dynamical calculations of images of crystals having large unit cells and complicated structures, as many as 400 beams commonly have been used. Extensions to larger numbers of beams are needed, and extensive calculation on trial cases may be needed before satisfactory methods can be established. Dynamical calculations of images of defects in crystals, with atomic resolution, is another example of a job likely to require access to the largest available computer. Three-dimensional reconstruction from two-dimensional images, including the practical case (not yet considered) for which the image intensity is a nonlinear function of the projected density, is another problem in this field for which extensive calculation may be required to find a good procedure.

These problems are appropriate candidates to be considered because the individual investigator may lack the combination of computer power, computer budget, and knowledge of alternative methods that might be provided or exchanged through the NRCC. The computer time needed on individual projects might be from a fraction of an hour upward. A significant impact could be made with an initial total of 10 hours per year of CPU time on the CDC-7600 or its equivalent. No need is foreseen for a permanent staff member in this area. The scientific personnel would be outside investigators, who might or might not need to work on site.

VI. PROGRAM LIBRARY AND ACCESS TO DATA BASE

The NRCC, a national facility with good communications and a staff expert in programming and in data retrieval and transmission, is a natural location for exchange of computer programs and for access to large data files of general interest. Like many other scientists, crystallographers have produced and exchanged computer programs in an uncoordinated manner. The NRCC would perform a great service in improving access to well-documented and tested programs. Many of these already exist and could be made available with nominal concerted effort. At least some should be maintained in operating condition at the NRCC to facilitate the work of crystallographic clients or customers. The NRCC could be a means of encouraging improved documentation, standardization, and generality of other existing or new programs of general value.

Working programs for both small molecules and macromolecules should be kept operational at the NRCC. They could be maintained by a competent programmer.

Distribution, updating, and correspondence, including a newsletter, for the programs in the library should be carried on by the staff of the NRCC. All distributed programs should be accompanied by a notice that the user is expected to acknowledge the author of the program in publications utilizing the results. A simple acknowledgment that the program came from the NRCC would not be adequate.

A number of crystallographic data bases of general scientific interest are currently being maintained, and access to them is available or anticipated via a commercial timesharing network. These include (a) the Cambridge Crystallographic Data File on solved crystal structures, (b) the Brookhaven Protein Data

Bank, (c) the NBS Crystal Data File on unit cell parameters and space groups, and (d) the Joint Committee on Powder Diffraction Standards file on powder diffraction data. Although these files should be made accessible early by computer network through the NRCC, there is no present need for them to be maintained by the NRCC. If access by commercial network proves to be inadequate, or if the locations of the files must be changed, the NRCC should be considered as the new repository. There is high priority for continuation of this activity, in one way or another. One full-time staff member would be required at the minimum anticipated level. A larger staff, including one or more professional crystallographers, would be necessary if responsibility were assumed for compilation or revision of such files.

VII. OTHER WORTHY PROJECTS

The preceding projects reflect only a few of the possible computational activities that the NRCC might consider initially in the area of crystallography. Other worthy projects can readily be envisaged. For example, extremely accurate experimental determinations of electron-density distributions can provide a test of quantum-chemical calculations. Intermolecular potentials also can be derived from the analysis of crystal structures, and the results can be applied to the prediction of crystal structures and even macromolecular structures. Statistical-mechanical procedures for model building of liquid and glass structures can be tested by matching predicted against observed radial distribution functions derived from diffraction data. The structure of the active site of an enzyme derived from protein crystallography could be studied as a model for quantum-chemical calculations of the dynamics of enzyme activity. Such a study would be particularly aided by crystallographic results at low temperature for enzyme-substrate intermediates.

Many of these possible crystallographic projects bear strongly on other areas of computational chemistry. A cross-fertilization of computational techniques and scientific content can be expected in the NRCC environment. Significant opportunities for advances on these interdisciplinary problems thus may be presented, which in the absence of the NRCC would be lost.

4
MACROMOLECULAR SCIENCE

I. INTRODUCTION

In recent years, considerable progress has been made in the theoretical description of macromolecules in terms of models that have steadily become more realistic. Nevertheless, fundamental problems still remain--problems involving interactions between different parts of the polymer chain and interactions between different polymer molecules as well as polymer and solvent. Certainly, analytic solutions of the mathematical problems associated with the description of large macromolecules are not yet in sight.

Although single polymer molecules or dilute polymer solutions can be described theoretically with some success, we do not yet have the means for dealing with concentrated systems. Nevertheless, reasonable pictures of limited applicability have been advanced. For example, we can take cognizance of entanglements, temporary networks, and tunnel diffusion, but a quantitative analytic theory of any one of these effects, not to mention their combinations, does not exist even for simple schematic models. Because of the unlikelihood of obtaining analytic solutions for the important macromolecular problems, it is clear that more work requiring numerical calculations will be necessary in the years ahead.

II. REPRESENTATIVE MACROMOLECULAR PROBLEMS THAT REQUIRE EXTENSIVE NUMERICAL CALCULATIONS

There are countless problems in theoretical macromolecular chemistry calling for computations, but we shall list here only a few. Following the list, we describe the problems briefly, after which we indicate the computer requirements for each:

1. Computation of average properties of isolated flexible polymer molecules;
2. Monte Carlo simulation of concentrated polymer solutions;

3. Dynamical properties of polymers;
4. Self-assembly of protein molecules and supramolecular structures.

These problems are of such complex character that significant advances beyond what has already been accomplished will require extensive computational facilities in conjunction with opportunities for organized interaction between participants in the research.

III. PROBLEM DESCRIPTION

A. Computation of Average Properties of Isolated Flexible Polymer Molecules

An important theoretical problem in macromolecular chemistry is that of chain conformation. A polymer molecule can be represented by a string of beads and simulated by a self-avoiding random walk on a lattice. As a first approximation, the individual beads can be regarded as hard spheres, but more realistically, one should introduce intrachain potential energies that are functions of the distances involved.

Exact solutions for the distribution of sizes and shapes of flexible strings of beads do not appear to be forthcoming at this time. However, useful statistical information concerning polymer dimensions can be obtained by Monte Carlo methods. Such methods call for a great deal of computation, especially as non-hard-sphere intramolecular potentials are introduced. Although semi-empirical interatomic potential energy functions are not fully refined at present, a variety of approximate forms exist. In cases where comparisons have been made, the results do not appear to be very sensitive to the specific functions. In any event, programs utilizing several of these potential forms might be useful for calculating conformational energies.

A modified version of the complete self-avoiding-chain problem is that corresponding to chains with interactions of limited order. Under limited order, we consider only those intramolecular interactions that correspond to chain contour separations equal to or less than some stipulated distance. Two beads separated along the chain contour by more than that specified distance shall be deemed for theoretical purposes not to interact, even if they should actually come into close proximity.

The problem of limited-order chains can be set up in matrix form and, from a knowledge of certain eigenvalues of such matrices, one can learn much more about the distribution of end-to-end separations of the chains. The matrices involved are huge, of the order of 10^5 square, for chains with as few as a dozen beads. For chains of 20 or more links, the matrix size becomes so exceedingly large as to prohibit any direct calculation of eigenvalues. However, the necessary eigenvalues, corresponding to eigenvectors possessing appropriate symmetry characteristics, can still be obtained by Monte Carlo methods, provided substantial computer capabilities are available.

The problem of macromolecular configurations is important because the rheological properties of polymers are related thereto, as are the properties of biologically important materials such as proteins.

B. Monte Carlo Simulation of Concentrated Polymer Solutions

Past work in theoretical polymer chemistry has been largely devoted to the problem of the isolated polymer chain, including analytical descriptions of equilibrium and nonequilibrium properties of dilute polymer solutions. Where analytical techniques have failed, Monte Carlo simulations of isolated chains have answered many of the questions concerning chain conformation. Although much remains to be done, our knowledge of the isolated chain and its relationship to the macroscopic properties of dilute polymer solutions is well under way. Unfortunately, in most practical applications of polymers we rarely deal with dilute solutions; rather we deal with concentrated solutions or bulk polymers. Our knowledge of interacting polymer chains and the effect of such interactions on macroscopic properties is not so far advanced as our understanding of dilute polymer solutions.

With the advent of larger and faster computers it is now possible to consider the simulation of multiple-polymer-chain systems by Monte Carlo techniques. Of paramount importance is the effect that neighboring chains have on the average conformation of a given chain (intermolecular excluded volume). Some preliminary work in this direction has been done on small systems of short chains. This has brought out the need for two particular areas of investigation. First, the effect of sample size must be investigated; in other words, how many chains (using periodic boundary conditions) need to be considered to obtain a reasonable approximation to a macroscopic system. Second, efficient sampling techniques need to be developed to simulate systems of high density. Methods thus far used suffer from serious attrition problems as the density is increased, resulting in the inefficient use of valuable computing time.

As with simple fluids, work in this area should be directed first to the equilibrium description of hard-sphere chains. More sophisticated potentials could be introduced later, either by a perturbation approach or by direct Monte Carlo simulation. One goal of this equilibrium study would be to obtain a hard-sphere-chain radial distribution function for polymer dimensions that could be compared with experimental results and provide a basis for developing an analytical theory. Such a radial distribution function could also be used to obtain equilibrium thermodynamic properties of concentrated polymer solutions and bulk polymers in terms of the molecular details of the chain. The theoretical work described above would be particularly appropriate in the near future because it would complement the recent experimental work in neutron-scattering techniques designed to study chain conformations in concentrated solutions and in bulk.

C. Dynamical Properties of Polymers

In this section we suggest several problems concerning the time-dependent behavior of polymeric systems. The basic mode of attack involves obtaining computer solutions of the classical equations of motion for one or more polymer chains. For some of the systems to be studied, a solvent must be introduced, so problems are proposed involving computer simulations.

Dilute polymer solutions, in contrast to concentrated solutions and melts, are relatively susceptible to analytic attack if the polymers are simple and the questions semiquantitative. But with a few exceptions, such as the reduction of turbulent friction and characterization of macromolecular weights and dimensions, technological interest concerns concentrated solutions of overlapping polymer domains, melts, and amorphous or crystalline solids. One wants to know the relation between the mechanical and dielectric relaxation of such systems and their chemical composition, molecular weight, branching, and cross-linking. Liquid crystals, which consist usually of elongated rodlike molecules, if not actually polymeric, exhibit orientational relaxation; the rate is significant in the design of display devices.

The activity of biopolymers depends, of course, on their secondary and tertiary structure; the rates of formation of these structures, and their deactivation by thermal, chemical, ionizing, and other agents, are problems in the relaxation of polymer conformations. An understanding of the rates and mechanisms of these conformational adjustments should aid in their control.

The experimental study of polymer-relaxation processes is, of course, fundamental but is severely hindered by two broad difficulties. First, the preparation of macromolecules of precisely known molecular weight and primary structure is often prohibitively difficult. The interpretation of experimental properties in terms of these polymer characteristics is accordingly uncertain. Second, experiments rarely afford a direct insight into molecular pathways. Rather the experiments must be understood by incorporation into a theoretical framework descriptive of molecular processes. We briefly review here this framework of polymer-relaxation theory.

The Verdier-Stockmayer calculations (P. H. Verdier and W. H. Stockmayer, J. Chem. Phys. 36:227, 1962) dealt with the relaxation of the end-to-end distance of an isolated polymer chain. The model was structurally realistic to some extent but allowed only a limited type of conformational motion that later turned out to invalidate the application to excluded-volume effects. More realistic representations of the interaction between polymer and solvent would permit a greater variety of chain motions. Possibly the dynamical interaction could be simulated by a Langevin random force; a few such attempts have been published in the past year. A comparison would be made between these two models (Langevin and molecular) of solvent forces on a polymer molecule, with specific attention to time-correlation functions of polymer bond vectors. Preferably the solvent would be simulated by realistic models of small molecules, but no attempt has yet been made along these lines, nor to simulate a nonequilibrium chain-polymer melt. A persuasive calculation would have to deal with a system of dimensions comparable with polymer contour lengths, and this system would contain almost as many degrees of freedom (10^3 or more) as a small-molecule system of comparable size.

A large-scale computer simulation is required for investigation of realistic models, especially those that account explicitly for solvent interactions. But it is probable that simpler models could be verified by such simulations. For example, we can raise these questions: Are random impacts from the solvent adequately simulated by a Langevin force, or can rotational energy barriers to conformational rearrangements be simulated by an internal viscosity?

Dielectric and mechanical response functions can be calculated from time-correlation functions involving bond vectors and bond tensors. Computer simulations should be employed to obtain such correlations.

The reduction of turbulent friction by extremely small concentrations of polymer molecules is of interest to polymer theorists and fluid dynamicists and is presumably a problem adequately modeled by a single macromolecule in a continuous solvent that obeys the Navier-Stokes equations.

The dynamics of secondary and tertiary structure formation can be simulated. The theoretical study of entanglements has begun with a single chain that loops around a fixed barrier. Since no experiment exists to test these calculations, further theoretical work is difficult to motivate, but computer simulation could bridge the gap between tractable models and experimental systems. Quite simple polymers, such as polymethylmethacrylate, form helical structures and could be studied on a computer to understand hydrophobic interactions in polypeptides.

Models of entanglements in concentrated solutions and melts exist but are analytically intractable. Experimental lifetimes for such structures are not available, but these would be directly accessible by simulation. A tunnel model has been proposed for the escape of a polymer chain from entanglements. The occurrence of this mechanism and its generalization to relaxing entanglements could be obtained by simulation. The orientational relaxation of liquid crystals, of importance to display devices, is a closely related problem.

The diffusion of imperfections, holes, and solutes in polymeric domains is of great importance in polymerization kinetics, membrane diffusion, and mechanical and dielectric relaxation. The effects of branching, swelling, and local order on diffusion could be investigated by simulation.

New problems of great difficulty and interest arise when side chains and charged groups are attached to a polymer backbone. How do orientational relaxations of the side chains couple with the slower modes of the backbone? What is the mobility of bound charges? The dielectric behavior of polymers, the mechanical properties of polyelectrolytes, the passage of ions through membranes, and so on require for their understanding this kind of elaboration of the basic model. Likewise, the interaction of charges on a polyelectrolyte in an aqueous medium is often discussed in terms of an effective dielectric constant. Quantitative simulation of a charged polymer in a dipolar medium could determine the validity of an effective dielectric constant and yield its magnitude. Another problem that depends on detailed representation of the solvent is the pairwise additivity of the average-force potential between different parts of the polymer molecule.

D. Self-assembly of Protein Molecules and Supramolecular Structure

Understanding the molecular basis of protein conformations is a prominent problem; detailed treatment by any method requires extensive computation. The major portion of previous work has dealt with secondary structure of globular proteins; in other words, are sections of the chain in a helix, extended sheet or in a more random form? Characteristically, these methods yield secondary-structure probabilities for individual amino acids, or groups of acids, from

the frequencies of forms observed in crystals. Secondary structures are then predicted for longer regions by combining such probabilities. As more protein crystal structures are reported, the methods for obtaining secondary structures should improve. Lengthy computations are then required to combine the secondary structures into an overall three-dimensional form, regardless of the method of accounting for intramolecular and solvent interactions. The assembly of these subunits into the overall protein could be treated by combining subunits of known three-dimensional form.

Tertiary and quaternary structure predictions require an extensive investigation of the free-energy surface in order to find the global minimum. One of the difficulties still to be overcome is to find, among the many local minima on the surface, the potential well that contains the global minimum. Because of the large number of independent degrees of freedom involved, these calculations require large amounts of computer time as well as ingenious programming and computational techniques. The possibility remains that inclusion of higher orders of structure might modify lower-order structures; such possibilities must be investigated at each step.

The theoretical study of helix-coil transitions in polypeptides and polynucleotides is an intrinsically interesting problem and is of special importance because it serves as a first step toward a detailed understanding of the rates and mechanisms of protein folding and unfolding. Analytical theory yields the general mathematical form of the solution to this problem as a sum of normal modes of relaxation, each with a characteristic amplitude and relaxation time. Carrying out detailed numerical calculations for particular models often requires Monte Carlo kinetic simulations using large amounts of computer time. Also, classical and quantum energy calculations will be required to estimate the transition probabilities or rate constants for the conversion of individual amino acid residues between helical and coiled states. Currently, both the analytical and the simulation work has been limited to homopolymers with neglect of excluded-volume effects. Inclusion of sequence heterogeneity and excluded-volume effects could increase the computational times by one or two orders of magnitude.

Computer simulations of pathways of protein folding should provide results that could be compared directly with experimental denaturation kinetics. Success could lead to conformational identification of observed intermediate states, as well as to the final native form. If the native state is not the conformation of lowest free energy, then specification of the pathway may be required in order to predict protein conformations successfully.

Continuing developments in studies of conformations of isolated biopolymers make it feasible to begin a study of the assembly of these molecules into larger structures. Above, we mentioned the combination of small numbers of protein subunit chains to form the final protein. A similar but more difficult problem is the arrangement of much larger numbers of molecules into a membrane. Understanding the molecular details of ion transport through membranes is critical to the comprehension of many biological processes. Organization of molecules into functional cellular organelles such as ribosomes is even more complex.

A theoretical consideration of binding between large and small molecules is of fundamental importance for understanding the action of antibodies. Such a basic understanding would make it possible to design drugs and scavenger

molecules to remove specific harmful molecules of environmental origin. Crystal structures of several antibodies are now known. Calculations of the free energies of binding interactions with small molecules is feasible, although it would require substantial expenditures of computer time.

IV. SUMMARY OF REQUIREMENTS

The estimated requirements for initiating studies in macromolecular science at the NRCC during Phase I are as follows (on an annual basis):

Personnel (Full-Time Equivalents)		
Resident scientists	1.0	
Numerical analysts	0.5	
Technical assistants	0.5	
TOTAL	2.0	
Workshops	0.5	
Computing Time (Fourth-Generation Computer)		
Isolated macromolecules	100	h
Concentrated solutions	150	
Dynamical properties	200	
Proteins	350	
TOTAL	800	h

V. NEED FOR THE NRCC

The NRCC could, of course, benefit theoretical research on macromolecules by making machine time available, but there are also many other significant and unique services to be provided. A catalytic influence on the work of scientists from different backgrounds who could come together through the NRCC would be most helpful. They would be attracted to the NRCC not only by hardware, such as computers of special utility, but also by scientific libraries, the presence of experts, and workshops serving to identify the most significant problems and the most practical approaches to their solutions. The problems are of such magnitude that they are not likely to be pursued in the absence of the nucleating effect provided by an organization like the NRCC. For any individual group to duplicate these services would cost far more than a common effort under the management of the NRCC.

At the NRCC, programs could be documented, standardized, and made available to other users. Data banks of equilibrium configurations and dynamical trajectories could also be assembled for use by the polymer-research community.

As soon as feasible, the NRCC should develop contacts with individuals or groups working in allied fields requiring computations. For example, dynamical simulation is carried on not only in chemical kinetics, statistical mechanics, and polymer theory but also in meteorology, fluid dynamics, and areas of physics. Advances in one area should be made available to all, and the NRCC could well facilitate such exchanges.

5
NONNUMERICAL METHODS

OVERVIEW

Nonnumerical methods are finding application in many fields of chemistry to solve important problems. Simply defined, these methods employ symbolic logic and such techniques as pattern recognition and interactive graphic display to constrain or simplify a complex problem to a point where a solution is possible. They may employ symbolic descriptions and manipulations under known constraints to analyze chemical structure and reactions. They may be used to sort through large collections of potentially related chemical information in search of correlations germane to the solution of a problem.

Computer programs in nonnumerical methods are now reaching sufficient sophistication and reliability that applications to practical problems of synthesis and analysis of molecular structure are possible and are being carried out. These problems are of types routinely confronted by literally thousands of practicing chemists. If only a fraction of these scientists had occasional access to such programs established and made available at the NRCC, then a substantial user community would exist and would derive benefit from the power of the methods. Although new as a generic class of techniques used by the chemist, nonnumerical computational methods have been around for many years. For example, symbolic logic was used, in part, to generate such computational codes as POLYATOM, and symbolic procedures have been used in processing numerical and nonnumerical (e.g., bibliographic) data bases.

During the course of our deliberations, many uses for nonnumerical methods were discussed, including (a) computer-assisted planning of chemical synthesis, (b) computer-assisted elucidation of chemical structure, (c) computer graphics, (d) pattern recognition, (e) data enhancement, (f) data bases and information-management systems, and (g) uses of minicomputers and microcomputers. Programs in these areas are being used to design complex chemical syntheses, assist structure elucidation in the field of natural products, provide graphic displays of chemical information meaningful to the chemist, and uncover interrelationships among complex chemical data and process data to bring out features relevant to a particular investigation. Many large data bases are available and can be searched with a variety of nonnumerical computer-based techniques. New data-management systems have been implemented, and new languages have been developed, to codify input and manipulation of chemical information in ways that simplify interfacing to small computers.

Our conclusions, and the ranking of priorities among the uses listed above and projects outlined below, were derived in the light of information and suggestions received by the committee from a large number of scientists. A particularly common request was that the NRCC provide both bibliographic and spectral-data searching services on existing data bases. A conclusion in an earlier report of the NRCC Planning Committee (Appendix A) was that this area is adequately served by existing federal and commercial agencies, and that the personnel required to manage a large data base would be excessive in relation to the NRCC's primary mission. This conclusion should be kept under continuous review by NRCC management. Furthermore, special facilities to implement real-time services have not been recommended except where such services might be provided by the communications and monitor systems of the NRCC. Initially, the NRCC could well provide information on where such facilities exist and could consider establishing access to them via network communications.

II. GENERAL FACILITIES OF THE NRCC

The Planning Committee has identified a number of services that the NRCC should carry on in relation to outside users. Some of these are general to all the potential activities of the NRCC; others are more specific to nonnumerical techniques and closely related areas with common interests.

A. General Requirements

Mechanisms for review of major programs should be established, aimed at ensuring code quality and transportability or facile access at the NRCC. The NRCC should provide important programs, and advice on their adaptation and use, to chemists working at other sites or having access to the NRCC from remote locations. Some attention should be devoted to the establishment of good programming techniques, machine-independent languages, and interlanguage translation techniques.

The NRCC should help to promote educational activities apart from program documentation and workshops. It could initially offer staff assistance via telephone, on-site visits, or linked terminals over a network. However, other opportunities will be made possible by the existence of the NRCC, such as fostering communication and collaboration among the user community by establishing a system for automated message-handling. On-line documentation and self-help facilities would greatly assist novice users in learning about specific programs, and the burden on the staff would decrease commensurately.

The system must tie in to established national computer networks in order to provide chemists with remote access at minimum expense in communications. Network access would also allow for intercomputer transfer of data and programs from other systems supporting computational chemistry (e.g., SUMEX and DCRT/NIH). Among possible networks are ARPANET, TYMNET, and TELENET; the latter two in particular would provide a broad geographic distribution of access points.

To encourage broad use of the NRCC facilities, "guest" access to programs should be possible. This category of access would permit a chemist to explore program availability and to try certain programs to determine their usefulness for specific research projects. Simplified channels of access to the NRCC will help chemists to explore the utility of the programs on a trial basis.

B. Specific Requirements

Many of the programs in nonnumerical methods have been designed to be interactive so that the chemist can examine intermediate results and use chemical knowledge to help guide the procedure through its next step. Such interaction is absolutely essential to the utility of these programs. The NRCC must be capable of supporting real-time interactive programs--programs where interactive activity is interspersed with short bursts of heavy computation requiring large core images and where the results of such computations can be displayed within a short time--so as to preserve the interactive nature of the program and to enable the next phase to be carried out. To ensure the utility of such programs, the NRCC must provide interactive computing of high quality, either on-site or by distribution among different hardware facilities; networks can be used to accomplish this distribution.

C. Specific Proposals

The following five major projects are considered feasible for development under Phase I of the NRCC and will contribute importantly to the uniqueness and usefulness of the facility:

I. Implementation of certain existing nonnumerical programs
II. Development of a standard for representation of molecular structure
III. General facilities for manipulation and presentation of molecular structure
IV. Access to Chemical Abstracts Service (CAS) structure files
V. On-site graphics facility

Success achieved by the NRCC in these areas will lead not only to greater awareness in the chemical research community of the power of nonnumerical computational methods but will stimulate further creative research on the methods and on their applications to the solution of chemical problems.

I. IMPLEMENTATION OF CERTAIN EXISTING NONNUMERICAL PROGRAMS

A. Purpose and Justification

Representative programs from the following areas should be implemented and supported by the NRCC:

Computer-assisted design of chemical synthesis and elucidation of structure
Pattern recognition
Cross assemblers, compilers, and special languages
Remote graphics support
Data enhancement
Information-management systems

Synthesis and structure elucidation form the foundation for much of experimental chemistry. Synthesis is used to confirm structure and to create new chemicals of importance to chemistry; it encompasses both research and development in both academic and industrial settings. Structure elucidation represents in many ways a complementary activity. Structures of unknown compounds, whether derived from natural products or from synthetic chemical procedures, must be characterized to ensure the continued progress of research. Synthesis frequently represents a way of verifying a hypothesis for an unknown structure. Both of these inextricably linked procedures are now being simulated by interactive computer programs, which act as assistants to the chemist, allowing exhaustive exploration of alternative pathways or solutions. The interactive nature of the procedures allows the chemist to restrict the alternatives to a most plausible set at every step.

A number of general-purpose pattern-recognition programs useful for the interpretation of chemical information are readily available. Applications cover all areas of chemistry, particularly drug design and structure-activity relationships. Because these programs will be of greatest value to the non-computer-oriented chemist, considerable documentation, HELP packages, and other educational devices will be required.

Recent advances in large-scale-integrated (LSI) technology allow construction of low-cost hardware to perform a variety of laboratory data-acquisition and computational functions. Development of such systems using microprocessors, for instance, is under way in nearly every major laboratory where experimental data are acquired from chemical instrumentation. The most difficult aspect is preparation of software to support the desired functions, since the microprocessors themselves are usually too limited in capabilities to provide for program generation. Therefore, well-documented and maintained cross assemblers, compilers, and other software-generating programs would be welcomed and used by a large number of experimental chemists. In addition, special languages exist for symbolic manipulation of potential interest to the chemist: LISP and BALM (list processing), SAIL, CSMP (continuous systems modeling), SIMULA and GPSS (discrete system modeling), SNOBOL (reduction and analysis of textual data), PASCAL, MAINSAIL (machine-independent language), SETL (general symbolic calculation), as well as FORMAC and other general-purpose systems for symbolic manipulation.

Several of the programs described in this report require some sort of interactive graphics to display intermediate and final results of computations and, in some cases, for input of structural information. We propose that the same run-time package be used by all such programs in order to promote standardization and to simplify interfacing new graphics devices to existing programs. OMNIGRAPH, an assembly language program written for the PDP-10 computer, is an excellent example of such a run-time system, in that virtually all commercial graphics devices are supported by it.

A large number of sophisticated computational methods are now available for extracting information from experimental data. Among them are digital filtering, deconvolution, correlation, and other methods, some of which require a substantial computational effort to accomplish. The main benefit of supporting these methods at the NRCC is that they would be readily available to a large number of experimental chemists who would not have to maintain a large program library separately. Furthermore, an expert in the application of these methods should be available at the NRCC to give users advice as to which of the methods might be best suited for particular types of data.

If information generated at the NRCC is to be accessible to the chemistry community, at least one information-management system should be installed and supported. The ability to manipulate and merge information from a number of data bases is often required. If all data generated and/or stored at the NRCC utilize one information-management system, such cross-communication is greatly simplified.

B. Technical Plan

In all areas mentioned above, programs exist and, with a minimum effort, could be installed at the NRCC. In some cases, a number of different programs exist, and either workshops or review papers are required to determine which are the most useful and transferable to the NRCC. In many cases, the availability of a PDP-10 or equivalent computer, either on site or via network, would greatly reduce the cost of implementation and would allow ready access to a broad class of programs.

C. Resource Requirements

Plan A--Without access to PDP-10 or equivalent
- NRCC staff: 4 full-time equivalents per year
- Outside collaboration: 2-4 full-time equivalents per year
- Consultants: 2 full-time equivalents
- Fourth-generation computer time: 100 hours per year
- Workshops: 3

Plan B--With access to PDP-10 or equivalent
- NRCC staff: 1 full-time equivalent per year
- Outside collaboration: 2 full-time equivalents per year
- Consultant: 0.5 full-time equivalent
- Computer time on PDP-10 or equivalent: 30 hours per year

II. DEVELOPMENT OF A STANDARD FOR REPRESENTATION OF MOLECULAR STRUCTURE

A. Purpose and Justification

The exchange of molecular data between programs and/or program systems is hampered by the lack of standards. An NRCC workshop should explore possible standards for representing structure and establish one for programs operating at the NRCC. A standard format should be implemented in existing programs and enforced in new programs installed at the NRCC. Such a standard would allow optimal communication of structural information between different computational efforts and facilities within the NRCC. In addition, the NRCC should encourage the National Bureau of Standards to define a national standard for exchange of molecular structural information.

B. Technical Plan

The standard for exchange will be a language that will permit a scientist to generate a valid file format for a new or existing program. This facility would be implemented through a subroutine that the user would automatically use when specifying structural information and transferring such information among various programs.

C. Resource Requirements

NRCC staff: 0.2 full-time equivalent per year for workshop organization, writing standards, retrofitting existing standards, new code development
PDP-10 or equivalent time: 20 hours per year
Workshop to achieve a consensus on the standard: 1

III. GENERAL FACILITIES FOR MANIPULATION AND PRESENTATION OF MOLECULAR STRUCTURE (CHEMLAB)

A. Purpose and Justification

The NRCC should establish the capability of carrying out on the computer a variety of experiments commonly done manually in the chemistry laboratory. Realization of this capability should be accomplished by bringing together in a unified package a collection of programs (called "CHEMLAB") for simulation of experiments and/or exploration of hypotheses in areas involving manipulation and presentation of molecular structure. CHEMLAB would include programs for planning of syntheses, elucidation of molecular structure, interpretation and correlation of chemical data and molecular modeling, along with graphics capabilities for facile input, examination, and editing of molecular structures.

CHEMLAB would have a tremendous impact on chemical investigations. The capabilities mentioned above are not currently accessible or available together at any site, nor are they likely to be, without the NRCC. Although many of the programs will exist separately (see Section I), an effort to unify them (a) to take advantage of their complementary capabilities; (b) to standardize input and output of chemical information; and (c) to provide automatic transfer of information between separate programs, will enable many chemists with little or no computer expertise to use them productively.

B. Technical Plan

CHEMLAB will be feasible only if the activities outlined in Sections I and II (above) are accomplished. An initial, reasonably broad set of capabilities within CHEMLAB can be established in the first two years of the NRCC by forging interfaces among many currently available programs that will be implemented at the NRCC (see Section I), using as a common mode of communication the representation of molecular structure discussed in Section II. This initial effort would bring together a variety of problem-solving programs that can be used sequentially under a common format to solve a complex chemical problem. For example, given an unknown compound, one might use: (a) spectral interpretation programs to determine structural features, (b) a structure generator to suggest candidate structures, and (c) molecular-mechanics methods to construct a model for the suspected structure, to be used as (d) a target for proposing synthetic routes to the compound.

CHEMLAB would require a monitor program that would handle initial communication with an outside chemist/user, help to determine the most appropriate problem-solving program, start up that program, and store relevant output automatically to pass on to the next computational step. CHEMLAB would require the capability of accepting input from a variety of computer terminals, ranging from teletypes to graphics terminals (e.g., those supported by OMNIGRAPH), to ensure the widest possible access. Some of the input and output functions of current programs would have to be modified slightly to fit comfortably in that superstructure. Those programs with limited interactive capabilities will require further development to be useful to a broad community of users. These tasks will be carried out by the NRCC staff in collaboration with authors of programs, with the understanding that collaboration is essential for guaranteeing access at the NRCC to a particular author's program.

Longer-term development would include (a) interfacing CHEMLAB to programs in other topical areas of the NRCC to provide a truly general-purpose method of manipulating and transferring structural information; (b) upgrading existing facilities within CHEMLAB as on-going chemical research provides more powerful programs; and (c) adding new capabilities for structure manipulation in any of the areas mentioned previously, as they are designed and tested. The eventual goal would be to develop a general manipulator of representations of chemical structure, accessible and utilizable by noncomputationally oriented chemists who would use a chemical vocabulary to instruct the program.

C. Resource Requirements

NRCC staff: 1 full-time equivalent per year
(Activities outlined in Section III deferred to 2nd year if the same person is involved in activities outlined in Sections I and II)
Outside assistance and collaboration with program authors: 2 full-time equivalents per year
Development and testing: 40 hours of PDP-10 or its equivalent time
Use: 100 hours of PDP-10 or its equivalent time (for first year; development time decreases and use time increases dramatically after initial development)
Workshop to establish initial plans and goals: 1

IV. ACCESS TO CHEMICAL ABSTRACTS SERVICE (CAS) STRUCTURE FILES

A. Purpose and Justification

The NRCC should provide the capability of access to the Chemical Abstracts Service (CAS) structure files in two ways: (a) structure matching for complete structures and (b) substructure searching. Although many data files are available at local sites or over commercial networks, including much of the CAS literature information, there is currently no way to determine rapidly whether a particular structure has been reported previously or what known structures possess particular substructures or combinations of substructures. The impact of this capability on general chemical research would be impressive. Molecular structure is the common language of all chemists, and it is important to obtain access to the chemical literature by asking questions phrased in terms of structure rather than by much less informative molecular formulas or *ad hoc* nomenclature schemes.

B. Technical Plan

A two-step plan would be carried out.

(a) *Structure Search.* A common set of questions asks about complete structures. Given a structure (a target for synthesis or quantum-chemical calculation or a possible structure for an unknown), has this structure been reported previously, and if so, where? Although the results provided by part (b) of this plan (below) will help to answer this question, there is a more efficient method. We propose that, by an algorithmic procedure, the CAS structure file be converted to an ordered list of unique "names" that can be searched quickly, given any "new" structure and its name derived from the same algorithm. We further propose that the name so derived include all relevant information on stereochemistry of the molecule, i.e., a name such as that generated by the SEMA algorithm published by Wipke and Dyott (W. T. Wipke and T. M. Dyott, J.

Am. Chem. Soc. 96:4834, 1974). The only technical problem in this procedure are in determining stereochemistry automatically from the CAS representation of structure. Some effort must be devoted to make the procedure efficient. Programs to derive the name are available; additional programs must be written to order and search the resulting list.

C. Resource Requirements for Part (a)

NRCC Staff: 0.25 full-time equivalent per year
Outside Assistance: 0.25 full-time equivalent per year
Fourth-Generation Computer Time for Conversion: 50 hours (one-time job with periodic updates)
Workshop, to be shared with Part (b), below: 1

(b) *Substructure Search.* A more general approach to the problem of deriving information on structure from the CAS files would be to provide capabilities for interactive substructure search. This can be accomplished by available programs to derive substructures from a given structure and to construct an inverted file of information based on those substructures. Modifications to currently available programs might be used to search the resulting file interactively for rapid answers to questions on the presence of substructures up to and including complete structures. There are, in principle, no technical problems in providing this capability.

Resource Requirements

NRCC Staff: 0.5 full-time equivalent per year for 3 years
Outside Collaboration: 0.25 full-time equivalent per year
Fourth-Generation Computer Time: 50 hours (one-time job with a few hours periodically for updates)

V. ON-SITE GRAPHICS FACILITY

A. Purpose and Justification

The NRCC should establish an on-site graphics facility capable of supporting diverse needs. Graphical representation of input or output of various calculations can be accomplished in many cases on ordinary graphics terminals if a unifying code is available (see Project I). However, interactive situations exist where moderate amounts of computation are followed by interpretive analysis requiring a higher level of graphics. The modeling of protein and nucleic acid structures in the so-called "Electronic Richard's Box" is an example. Similar meaningful applications are envisaged in a number of the topical areas covered in this report. Systems and hardware for these graphical purposes have already been developed and coupled to a CDC-7600-class machine,

resulting in a practical and useful facility. The visiting scientist would benefit considerably from such a facility. For example, a protein crystallographer using a few hours per day could complete a stage of structure refinement in two to three weeks. Although sophisticated graphics systems exist at various sites in the United States, there is no single site that all investigators may utilize.

B. Technical Plan

There are two different approaches to such graphics capabilities: software-dominated or hardware-dominated. Sophisticated, developed systems exist. The first step in developing the facility would be a workshop or review to determine (a) the graphics needs of the NRCC and (b) the graphics system(s) best satisfying those needs. Implementation would require minimal programming support, as a number of software systems are available.

C. Resource Requirements

NRCC Staff: 0.1 full-time equivalent per year
Workshop: 1
Fourth-Generation Computer Time: 50 hours
Hardware Costs:
Software-dominated facility: $125,000-$175,000
Hardware-dominated facility: $250,000-$325,000

6
PHYSICAL ORGANIC CHEMISTRY

I. INTRODUCTION

Computers have considerably aided the development of physical organic chemistry over the past ten years. The processes of data gathering have been revolutionized by interfacing measuring instruments with computers; and the interpretation of complex data, such as spectral and kinetic information, has been greatly facilitated by the development of sophisticated computational techniques. However, the difficulties associated with the use of these computational routines have prevented many organic chemists from taking full advantage of the theoretical tools that are being developed. The creation of a National Resource for Computation in Chemistry (NRCC) was seen by the present panel as an opportunity to ameliorate this situation and, in so doing, to fulfill several important needs of physical organic chemistry.

A small but significant fraction of the interests of physical organic chemists is coincident with the specific interests of theorists in quantum chemistry and in statistical mechanics, and with the techniques developed by specialists in nonnumerical methods. For these interests, the services and facilities provided by the NRCC should be identical with those outlined by the quantum chemistry, statistical mechanics, and nonnumerical methods panels.

However, the committee believed that the major contribution of the NRCC to research in physical organic chemistry would be not in providing large-scale computational facilities but rather in delivering high-quality human services. The resource should disseminate computational knowledge of relevance to chemistry and provide the necessary help to match the interests of individual researchers with available computational techniques. Furthermore, it must provide the facilities for initial testing of such investigations without requiring a major commitment of time or money until the probable success of the project can be assessed. In this report, we first discuss (Section II) the kinds of services that will fulfill (at least in part) the general computational needs of the physical organic chemists, and in Sections III-V we identify representative areas of research that would be particularly well served by the creation of the NRCC.

II. PROPOSED SERVICE FUNCTIONS OF THE NRCC TO PHYSICAL ORGANIC CHEMISTRY

The services itemized below could best be provided by the presence on the staff of a full-time professional scientist with a PhD in physical organic chemistry and experience in computing, who would be available for consulting and who would supervise the additional personnel required, as described.

A. Creation of a Library of Programs

Requirements:

Personnel: 0.5 staff member at B.S. level
Computer Time: 10 hours per year

1. Standardization

In view of no previous coordination in the development and presentation of programs, a national center should establish, through a workshop, a set of guidelines for programming style. It should further establish input/output compatibility among existing and future analogous and related programs.

2. Testing and Comparisons

The state-of-the-art programs to be maintained in the library, or for distribution to customers, should be properly tested. Similar programs should be tested on similar problems for reliability and consistency.

3. Advice and Information on Use

A qualified staff member should be on hand to advise potential users on the available programs, their uses, and unique features of applicability.

4. Documentation

Each library program, with properly standardized I/O, should be associated with appropriate detailed information such as its history, some information about the algorithm used, a few typical test solutions (see 2), and relevant comments about its usefulness and limitations (see 3). Reference to a responsible staff member also would provide the user with a contact for solving difficulties encountered in the use of the programs.

5. *Awareness*

The NRCC should use every means at its disposal (such as newsletters and an electronic post office) to keep the chemical community at large aware of ongoing projects and of program improvements contemplated or implemented at the center. The newsletter could also serve individual researchers by advertising their need for specialized routines that may exist in other laboratories.

6. *Library Content*

We recommend that the NRCC maintain a library of state-of-the-art operational programs with standardized I/O to cover the various needs of physical organic chemistry. These include:

(a) Multivariate regression analysis and curve-fitting programs, including plotting and graphical display.
(b) Molecular mechanics programs, such as force-field programs and semiempirical and simple *ab initio* quantum-mechanical programs.
(c) Programs relevant to kinetics and spectroscopy.

B. Hardware and Interfacing Information

Requirements:

Personnel: 1 full-time staff member at B.E.E. level
Computer Time: None
Equipment: Fully stocked electronics shop, including measurement and testing instruments. (Much of this could be made available by the host institution for the NRCC.)

Much of the modern research in physical organic chemistry can be aided by analog/digital conversion of experimental data followed by calculations using programmable calculators or minicomputers. The computational requirements dictated by specific experiments are quite variable, ranging from 8- to 16-bit accuracy and from microseconds to hours per data point. At present, it is virtually impossible for anyone not familiar with electronics to obtain information on, for example, available equipment and interfacing requirements.

We propose that the NRCC provide information on state-of-the-art digital equipment, including a library of schematics and descriptions of equipment and of interfaces adapted to specific experimental techniques. In addition, a staff electronics engineer should be available for consultation on interfacing electronic instrumentation with digital equipment.

C. Data Bases and Data Base Information

Requirements:

Personnel: 0.25 staff member at B.S. or M.S. level
Computer Time: 1 hour per year

Access to three types of information would clearly aid physical organic chemists:

(a) Rate and equilibrium constants for various organic reactions.
(b) Spectral information (NMR, ultraviolet, visible, infrared, mass) on organic compounds.
(c) Results of completed calculations on potential-energy surface, molecular orbitals, and Monte Carlo and molecular dynamics.

Only in (c) is it necessary that the NRCC maintain data files and search routines. In (a) and (b), data bases already exist at other institutions, and the NRCC should be able to provide information on location and accessibility of the desired data.

D. Workshop and Awareness Conferences

At least twice each year, the NRCC should organize and convene conferences for interested physical organic chemists. These conferences should inform the participants of the services and facilities available at the NRCC and of the latest technical and theoretical capabilities of hardware and software in their field. The conferences should also provide feedback to the NRCC on developing needs in physical organic chemistry and form a focus for collaborative efforts in significant new fields of research.

III. MOLECULAR MECHANICS

Requirements:

Personnel: 0.25 staff member collaborating with up to three outside research groups
Computer Time: 8 hours per year

The computational prediction of structures and energies of large organic molecules could be carried out by empirical force-field programs. A number of different force-fields have been developed based on different algorithms, using different parameterizations, and intended for different classes of molecules. This classical mechanical technique has been found capable of providing results of chemical accuracy, with computational efficiency 100 to 1000 times that of quantum-mechanical calculations. An effort should be made to collect, test

standardize, and reparameterize the best of these systems for all important atoms. The expected benefits of such a research program are many: the reliable prediction of molecular structure and energy is a goal worthy in itself, and the results can be incorporated into other programs for spectral prediction, graphical representations of molecules, and quantum-mechanical calculations.

Extensions of the technique are worthwhile from a scientific point of view and practically attainable in terms of demands on personnel. Thus, inclusion of solvent associations would be of considerable value. Theoretical examinations of this problem are sufficiently advanced that first-order approximations (utilizing only small amounts of computation) would give results of useful accuracy. In principle, a range of solvent molecules could be treated, not just spherical approximations to an idealized solvent.

Another worthy extension would be to include intermolecular interactions, to analyze, for example, how two polymer chains interact either by rotation or translation. The results should shed new light on the behavior of macromolecules. Still another area of utility would be the modeling of transition states for organic reactions.

Molecular mechanics is the only currently known technique that will permit calculation of the structures of macromolecules in a reasonable amount of computer time. Interdisciplinary collaboration in this area could be fostered by the NRCC. Molecular mechanics calculations could also be used as a first-level screen to assist quantum chemists in choosing problems of general importance to the chemical community. A data base of structures for which complete geometries are known would be of considerable value in facilitating parameterization and the testing of calculated results.

IV. KINETICS AND REACTION MECHANISMS

Requirements:

Personnel: 1 staff member in collaboration with 2 outside research groups
Computer Time: 30 hours per year

Physical organic chemists bring all the tools of chemistry to bear on the study and understanding of organic reactions. Because of the diversity of needs, the NRCC can be an important focus of communication on the computational aspects of the studies in which experimentally determined rates are correlated with reaction mechanisms.

A. Experimental Studies of Chemical Reactions

In experimental studies of chemical reactions, whether in solution, in the gas state, or in mixed phases or involving ordinary or excited states, the raw experimental data must be translated into detailed rate expressions. Although a number of programs have been developed to simulate rate data from given kinetic rate laws or to fit experimental data to trial kinetic expressions

these programs are often designed for a special purpose or are insufficiently well known or well tested to be of general utility. The computational techniques are under active investigation by mathematicians and engineers, but for the most part the programs are not available to chemists. Acquisition and modification of such programs for general application to chemical systems would be an appropriate activity of the NRCC. Programs are particularly needed for systems of many coupled reactions, such as oscillating reactions (both biological and otherwise) and those encountered in environmental studies.

The NRCC should have one staff member experienced in surface-fitting programs, able to communicate both with experimental chemists who wish to use programs of this type and with graphics experts who can be consulted on development of interactive graphics for eludicating the results. During the initial operation of the NRCC, this staff member would collect available programs and consult with their users, standardize and consolidate useful codes and make them machine-independent, and collaborate with outside scientists on any major computation in chemical kinetics. This latter project would serve as a testing ground for the creation of a general-purpose system of programs for kinetic computations, to be made available to experimental chemists; it would demonstrate the utility of such programs through the solution of an important chemical problem.

B. Interpretations of Reactions

The preceding section describes that area of research combined with service having greatest impact on the treatment of experimental data. The interpretation of the results and the development of theories of reaction rates require a different effort, outlined as follows. Even though potential-energy-surface, trajectory, and statistical-mechanical calculations show some promise of unlocking the secrets of reaction dynamics, such treatments are impossible for reacting systems of the complexity with which organic chemists are generally concerned. However, the description of organic reactivity in terms of quantum-mechanical rather than empirical concepts is having a major influence on the course of experimental organic chemistry. The establishment of the NRCC will be of immediate value because of the increased communication it will foster between experimental and theoretical chemists. Furthermore, as discussed in more detail in Section II, the increased dissemination of quantum-mechanical programs and the expected greater efficiency with which they can be used will enable organic chemists to carry out theoretical calculations relevant to their needs and obtain the results in meaningful form.

There are several areas in which projects unlikely to be carried out in the absence of the NRCC, but feasible with its support, can have a major impact on physical organic chemistry. One of the major unsolved problems in organic chemistry is the general influence of solvents on the rates of reactions. Present theoretical techniques are unsatisfactory either because of excessive computational time required or because of inherent unreliability. New approaches must be developed. Other areas where major collaborative efforts should be concentrated are the theoretical treatment of both homogenous and heterogeneous catalysis and the theoretical study of the interactions

between electronically excited molecules. It is almost certain that these problems will call for the development of new computational techniques on a scale that can only be provided by a major computational facility.

V. SIMULATION AND ANALYSIS OF SPECTRA

Requirements:

Personnel: 0.5 staff members collaborating with 3 outside research groups
Computer Time: 30 hours per year with graphical display capabilities

A. Nuclear Magnetic Resonance (NMR)

NMR is undoubtedly the most powerful spectroscopic technique employed by modern organic chemists. It is used for a great variety of tasks ranging from the routine to detailed conformational analysis of molecules as different as complex natural products, proteins, nucleotides, and synthetic polymers. Anything the NRCC can do to foster and extend the applications of this technique will have an important impact on organic, biological, and physical chemistry.

The following tasks, directed toward aiding chemists who use NMR spectroscopy, can be accomplished largely by seeking, collecting, standardizing, and improving programs already written; however, assembling and improving these programs and making them widely available will make a great difference in the extent to which the various techniques involved are used.

1. Simulation of Spectra

(a) In the simplest form, the input is chemical shifts and coupling constants and the output is a list of peaks and a graph of the spectrum.

(b) If the input is a structural formula, the program must then look up chemical shifts and coupling constants from data tables containing spectral information about similar compounds and produce spectral data that can be utilized in various ways, including matching or prediction of an observed spectrum.

2. Analysis of Experimental Spectra

(a) Obtaining the best (least-squares) shifts and coupling constants from a spectrum.

(b) Providing structural possibilities that would fit a spectrum (using any other information available). For this task an interactive program operating at the NRCC would be highly desirable.

3. Obtaining Fast Reaction Rates from NMR or EPR Line Shapes

(a) Simulation of spectra of systems undergoing fast reactions: (1) uncoupled or with first-order coupling or (2) with non-first-order coupling.
(b) Best (least-mean-squares) fitting of experimental data.

B. Mass Spectrometry

Mass spectra contain valuable structural information. Modes of fragmentation in the spectrometer are sufficiently understood to enable prediction of which fragments will form and thus to select among various structures for unknown compounds. Programs for simulating spectra and deducing structures from mass spectroscopic data have been written. It would be the task of the NRCC to acquire, standardize, update, maintain, and distribute programs and the relevant data bases. These functions should be carried out in collaboration with existing facilities.

C. Infrared and Raman Spectroscopy

Infrared spectroscopy is used to identify and characterize certain groups within molecules, such as carbonyls. Reference books listing group frequencies in different molecules have long been available.

Detailed vibrational analysis provides a connection between the three-dimensional structure and force constants of a molecule and the infrared and Raman spectra. Programs are available that simulate spectra, allow frequencies in experimental spectra to be assigned, and then use these assignments to refine the structures and force constants. This procedure has a natural and obvious connection with the project described in Section III. The molecular-mechanics model can produce initial structures and force constants for the vibrational-analysis routine, which in turn produces data to aid the parameterization involved in molecular mechanics. The task of the NRCC would be to combine various programs now separated, in order to implement these procedures.

D. Combined Methods of Structural Elucidation

An ideal system of spectroscopic programs would be capable of using NMR, mass spectroscopy, infrared, ultraviolet, and other data to perform structural elucidation. Combining and evaluating structural inferences from all these sources is a nontrivial task but would be materially aided by the variety of expertise and programs at the NRCC.

7
QUANTUM CHEMISTRY

I. INTRODUCTION

Among the chemical fields to be served by the NRCC, quantum chemistry has pioneered in the use of state-of-the-art computational methods; the field today is characterized by a large volume of computer-intensive activity. Methods have been identified and tested for the carrying out of a variety of potentially highly useful calculations, and some such calculations have been carried out. However, the magnitudes of the hardware and software resources required are such that relatively few definitive major studies have yet been completed.

Progress in quantum-chemical studies is hampered by three factors, of which the first is a lack of transferability and general availability of flexible and efficient software, diverse components of which exist scattered among various laboratories. There is a pressing need for such software to be standardized, validated, improved, and made available on a basis not requiring expertise in computer science as well as chemistry. A second unfulfilled need is for better cooperation and coordination among at least the more expensive quantum-chemical studies. This coordination should involve not only choices of systems for study but also output-data management and preservation, so that the results of calculations can be brought into manageable forms, both for validation of the work of others and for the use of their results in further studies. A third and obvious need is for a reliable source of computer time sufficient for a few key calculations that would unambiguously exploit the present capacities of quantum chemistry, would be at the levels of accuracy and detail currently achievable, and would significantly advance the theoretical understanding of chemical science. These studies are expensive but should not be compromised by insufficient support.

In light of the foregoing recital of needs, we believe that the activity at the NRCC of highest priority for the furtherance of quantum chemistry would be to provide those services that would improve the capabilities and efficiency of all workers in the field. This would expand by severalfold the productivity of the hardware available to the NRCC (at least during Phase I), while enhancing the effectiveness of many scientists who use computing as a tool rather than as a primary research activity. Our second priority is the coordination of software and hardware in key areas, so as to better define directions of future high-yield cooperative activity. Our third priority is the carrying out of at least one major research project of great intrinsic scientific significance.

We believe that it should be possible to support fully the services outlined in more detail below, some substantial portion of the coordination activities, and at least one major computing project. Because of the great breadth and diversity of the scope of quantum chemistry, it was not practical to limit our presentation to a volume of activity consistent with the expected size of the NRCC. We therefore present a large number of research possibilities, grouped in six general areas (some of which overlap areas of primary interest to other topical groups), from which we realize that but a few can be initially chosen, and that others, not listed, may be equally meritorious. We urge that whatever research is undertaken be supported at a sufficiently high level to make a real impact, and that the NRCC not be divided in so many ways that major benefits fail to result. In this report, we consider specifically Phase I of operations; some of our recommendations may also apply thereafter to Phase II.

The specific activities outlined below have been chosen because they are uniquely suited to and can be most effectively implemented at a computational resource organized along lines recommended in earlier Academy reports (see Appendixes A and B and *Computational Support for Theoretical Chemistry*, National Academy of Sciences, Washington, D.C., 1971). An overview of the recommended activities and their associated resource requirements is presented in Table 1. The mutually supportive interrelations expected between these proposed activities and those of other topical groups are summarized in Table 2.

It may be noted that except as implicit in the proposed service activity, there is no specific prescription in our report for the development of new computational methods in quantum chemistry. This is not because we believe such activity to be unimportant; on the contrary, we believe it to be of the highest importance. However, such innovation cannot be programmed in detail, and we do not feel competent to predict its nature or extent. In any event, the planning for the NRCC should include the expectation that innovative projects leading toward methods development would be conceived and encouraged and, as they appear, would be included in the NRCC program. We do not anticipate that the development and initial testing of proposed new methods will require a large proportion of the facilities of the NRCC.

II. SERVICE FUNCTIONS

A. NRCC Service Functions

An essential NRCC function must be to develop informational and educational processes that will make the benefits of the NRCC known and available to the scientific community and will promote the transfer of state-of-the-art programs and methodologies to the computer centers of universities and research laboratories. The following six kinds of service are identified as supporting the activities of the NRCC in quantum chemistry. These services and the expertise uniquely assembled to provide them are a major component of the NRCC concept, distinguishing it from a computing center. While some of the services have broader applicability than to quantum chemistry, the resource levels given in Table 1 have been based for the most part on quantum-chemistry needs.

TABLE 1. Resource Requirements for Proposed Initial Activities in Quantum Chemistry (on an Annual Basis Except where Otherwise Noted)

	NRCC Personnel					Outside[a]						
	PhD Chemist	PhD Molecular Biologist	Computer Scientist	Computer Programmer	Technician	Client User Groups	Nonclient Groups	Individual Collaborators	Computer Time (4th Generation Computer Hours) 1st Year	Computer Time (4th Generation Computer Hours) 3rd Year	No. Workshops	Priority[b]
Service Activities[c]	2	-	1	1	2	-	-	4.5	100	100	4	1
Potential Energy Surfaces												
Coordination	0.5	-	-	-	-	-	10	-	20	20	2	2
Production[d]	0.5	-	-	-	1	3	-	-	200	400	1	3
Excited States												
Coordination	0.5						10		20	20	1.5	2
Production[d]	0.5					3			100	300	0.5	3
Molecular Biology												
Coordination		1						4	50	100	2	2
Surfaces and Clusters												
Coordination	0.5						6		20	20	1	2
Production[d]	0.5				1	3			200	500	1	3
Solvation												
Coordination	0.5						6		20	20	2	2
Production[d]	0.5					3			100	250		3
Transition Metals												
Coordination	0.5						6		20	20	2	2
Production[d]	0.5					4			200	425		3

[a] Clients carry out projects approved by the NRCC for support including computer time; customers are also served by the NRCC but are charged for computer time.

[b] Priority 1--essential; Priority 2--selection of several essential; Priority 3-- selection of at least one essential.

[c] A portion of these service activities will be directed toward the quantum-chemistry research projects involved in the NRCC program. These allocations will also suffice to provide program exchange service to all NRCC activities and program maintenance service for the quantum-chemistry activities in other areas, e.g., physical organic chemistry.

[d] In addition to the coordination resource requirements for the same activity.

TABLE 2. Interrelationships between Projects in Quantum Chemistry and in Other Areas (Arrows Indicate Most Significant Directions of Information Flow)

Quantum-Chemistry Projects	Chemical Kinetics	Physical Organic Chemistry	Crystallography	Statistical Mechanics	Macromolecular Science	Nonnumerical Methods
Potential energy surfaces	↕	↕		↕	↑	←
Surfaces and clusters	↕	↕	↕			
Excited states	↕	↕				
Biologically active molecules					↑	←
Molecular interactions, solvation	↕	↑	↕	↑	↕	
Transition-metal complexes		↑	↕			

However, the requested level of support will provide program exchange in all topical areas. With appropriate expansion of the resource allocation, all the services outlined here could be provided as needed in all topical areas.

1. Program Development

This service would involve the design and testing of new computer codes and their unification into master systems. It would require a core of regular NRCC staff members with the necessary expertise in state-of-the-art methodologies in quantum chemistry, augmented by temporary staff members and supplemented by visiting scientists, their assistants, and students. Members of the program development staff would work on projects under the direction of scientists who are not part of the program development staff and would also provide consultation and assistance for NRCC users who develop their own programs. A roster of consultants would be available to the program-development staff for assistance in computational areas where the NRCC resident scientists are not expert. The service would foster the development of major new or modified codes through resources that cannot be marshaled by individual research groups but would be uniquely available at the NRCC.

2. Systems Development

This service would be concerned with development of discipline-oriented hardware and systems software that will optimize the performance of installed scientific codes. It would not be large enough to undertake major projects in systems development; such work when needed might be carried out under contract with outside vendors. The systems-development staff would supervise contract performance, provide consultation on acquisition of systems hardware and software, and carry out minor projects in its field.

3. Program Certification and Maintenance

Staff members assigned to this function would be charged with responsibility for certifying programs; for comparing, maintaining, and standardizing programs in active use; for providing expert advice and assistance to users of these programs; for making minor modifications and improvements to the benefit of users; and for documentation. Close cooperation with the program-development staff and consultants will obviously be necessary. The program certification and maintenance staff would also prepare a catalog of available programs, with appropriate documentation.

4. Program Applications

The program-applications staff would be the main contact with users interested primarily in obtaining theoretically calculated values of physical observables. Existing codes would be applied to fulfill requests from either experimentalists

or theoreticians (e.g., for excited states, potential energy curves and surfaces, or other properties of molecular systems). Visitors and other outside investigators would receive instruction in the use of these codes. Staff duties include also production runs in connection with major scientific projects approved by the Program Committee. Although the program-applications staff would work closely with the program maintenance staff, they would be more discipline-oriented. They would have to be knowledgeable in the limitations of the programs from the viewpoint of the underlying theory as well as the mathematical algorithms, and they must be able to advise users on the best method and program to use and the probable limits of error of the calculated results. Past experience suggests that the most fruitful interaction between experiment and theory, and between the NRCC and scientific community, would occur through the program applications service. This service could be permitted to grow with demand if provided on a cost-recovered basis, within constraints established by the Policy Board to assure proper balance of the total NRCC program.

5. *Scientific Data Bank*

This would consist of a computerized library of the results of calculations by users of the NRCC, selected results of major calculations done elsewhere, and also relevant experimental data. It is needed in order to avoid unnecessary repetition of calculations and to make results of previous work conveniently available in an accepted standard format. The responsibilities of the data-bank staff would include evaluation of the quality of the data and, in cooperation with the program and systems development staff, the design of efficient and standardized means of storage and retrieval. Many important bank topics come to mind, and specific areas to be covered would be delineated through proposals approved by the Program Committee.

6. *Program Exchange*

Essential services provided by the NRCC would include the collection (with consultant advice), routine testing, packaging, and dissemination of computer programs and program documentation. This activity is to be carried out by technicians; necessary program development, certification, and maintenance are provided for as concrete service functions. It will be important to coordinate this activity with the services currently provided by the Quantum Chemistry Program Exchange at Indiana University in order to avoid duplication.

B. Resource Requirements

The service functions outlined above would be carried out in part by NRCC personnel and in part by outside personnel interested in these areas, with coordination by the NRCC staff. Coordination can be facilitated by workshops where service needs are identified and software approaches to meeting these needs are discussed and agreed upon. Some workshops would be designed also

TABLE 3. Resource Requirements for Services to Quantum Chemistry (on an Annual Basis)

Service	Personnel[a]		Computer Hours	Workshops
	NRCC Staff	Outside		
Program development	1.0[b]	1.0	50	2 of 3[i]
Systems development	0.5[c]	0.5	10	1
Program certification and maintenance	1.0[d]	1.0	10	2 of 3[i]
Program applications	1.0[e]	1.0[h]	At cost	0[j]
Data banks	0.5[f]	1.0[h]	10	1
Program exchange	2.0[g]	0	20	2 of 3[i]

[a] Numbers in full-time equivalents.

[b] PhD chemist with major interest in code development.

[c] PhD computer scientist.

[d] Computer scientist with chemistry background.

[e] PhD chemist with major interest in code application.

[f] Computer scientist.

[g] Technician.

[h] These classes of outside personnel might well be self-supporting and accordingly contain more members.

[i] Workshops in program development, certification, maintenance, and exchange should be held as a series, annually, with two of the three annual workshops devoted to determining priorities in code activities at the NRCC.

[j] It is assumed that there will be at least one Users Group meeting per year at which the program applications activity of the NRCC will be discussed (and explained).

to provide environments for intensive development of service software by experts external to the NRCC.

The NRCC resources needed are outlined in detail in Table 3.

C. Off-Site Activities

The center of all the above service functions would be the NRCC. As indicated previously, however, most of the systems development would be let out on contract. A major goal of the NRCC would be to aid in the transfer of its software and expertise to user institutions when possible. Users would be encouraged to move software to their local sites after they have become sufficiently knowledgeable in its structure and use. Contributions, in standardized format, to the program library and to the selected data banks would be sought from the widest possible list of contributors. When appropriate, these services would be set up for remote access.

D. Justification

The great promise of modern quantum chemistry has in no way come to full realization, primarily because of the lack of the coordinated services and functions envisaged herein. These services would make available to the entire chemistry community the state-of-the-art resources of computational quantum chemistry and greatly facilitate its further development.

Experience has shown that the most significant interactions between experimental and theoretical quantum chemistry have taken place when a major subset of these service functions has existed at a single site. Usually this advantageous state of affairs has been short-lived, and the continuity of services required for effective distribution of codes and expertise has not endured long enough to have a major effect on the total chemistry community. The NRCC services can remedy this situation and permit computational quantum chemistry to build continuously upon the vast but so far uncoordinated structure and techniques now available. With the above services, the NRCC would provide a fertile and flexible environment in which research requiring its unique computational resources can be efficiently carried out, new ideas can be feasibly tested in a reasonable time span, and valuable new techniques are sure to emerge. Finally, the services can be expected to be cost-effective because of sharing common to many projects.

III. RESEARCH ACTIVITIES

A. Potential Energy Surfaces

A detailed knowledge and compendium of the potential-energy surfaces for small-molecule systems is of primary importance in many branches of chemistry and physics. Such knowledge underlies the qualitative understanding of chemical

reactions, the study of molecular energy transfer and reaction dynamics, and the interpretation of spectra. Such diverse fields as atmospheric and environmental chemistry, astrochemistry, photochemistry, combustion chemisty, and spectroscopy all vitally need potential-energy surfaces of ground and of excited electronic states, together with the associated wavefunctions and transition matrix elements. The paucity of such data not only hinders progress but is frustrating because growing evidence points to quantum chemistry as a mature field with an ample variety of reliable procedures for the study of chemical phenomena.

The NRCC can stimulate progress by selecting, after appropriate participation by prospective users through workshops and other means, two or more of the existing large electronic structure codes for standardization and preparation for use on a variety of computers. (This should be done carefully to avoid stifling method development.) Molecular systems the potential-energy surfaces of which would be particularly useful would also be identified by quantum chemists, dynamicists, and experimentalists; investigators would be encouraged to contribute to our knowledge of these surfaces. Examples might include systems of importance in various contexts, such as atmospheric chemistry (H_2O, CO_2, NO_2, O_3, etc.), combustion ($H + O_2$, $OH + CO$, etc.), ion-molecule reactions ($N_2 + O^+$, etc.), astrochemistry ($CH + O$, $OH + He$, etc.), reactive systems of interest for laser or other applications ($Mg + F_2$, $Li + HF$, CH_2O, etc.). However, it would be premature to identify conclusively the most fruitful systems at this time. Through cooperative efforts, it is expected that the reliability and accuracy of the calculations will become more fully established and their utility in chemical dynamics enhanced.

A few prototypical surfaces should be determined carefully and completely at the NRCC. For a typical triatomic system this task might take as much as 250 hours of computer time. A desirable survey of the quality required and attainable would be provided by studies of four systems over a three-year period. In addition to *ab initio* calculations carried out as accurately as is required and is practical, the adequacy and calibration of less accurate semi-empirical methods such as that of diatomics-in-molecules should be studied.

The specific studies proposed in this project should be selected and planned jointly by quantum chemists and molecular dynamicists. Particular attention should be given to the system chosen, the position and number of energy points computed, the surface accuracy required, and the ultimate fitting of the surface. The procedure will be iterative, involving successive refinements of the surface computation as increasingly detailed studies of the dynamics are made thereon. It is estimated that the code development, validation, and user-coordination will require about 0.5 full-time equivalent of NRCC personnel in addition to the assistance provided from the recommended service operations; personnel requirements are a PhD chemist with an interest in computer code applications. The coordination phase will require 20 hours of computer time annually, and can best be carried out if at least two workshops are convened. To enter the production phase, 0.5 equivalent of an NRCC PhD chemist plus the services of one technician will be required. Computer time requirements are estimated at 200 hours for the first year, increasing to 400 hours the third year. This estimate allows for production calculations on four molecules during the three-year period. One additional workshop involving dynamicists will be needed to discuss and coordinate the production activities.

B. The Excited States of Polyatomic Molecules

The excited states of molecules are of importance in many chemical processes. These include laser development and applications, atmospheric chemistry (especially in the formation of photochemical smog), biological problems (such as photosynthesis and the chemistry of vision), and solar-energy conversion and storage. For example, within recent years the lowest triplet state of molecules has attracted much attention among chemists and has been the subject of a number of conferences aimed at exploring the structure and reactions of this state. Theoretical studies of excited electronic states of molecules are expected to be particularly fruitful since the experimental study of such states is hampered by their transient nature and by the difficulty in preparing the molecule in the state of interest.

While an adequate description of the excited states of diatomic molecules is at hand, the situation is far from satisfactory for polyatomic molecules. Theoretical attack involves three major areas of research: the electronic-vibrational-rotational structure of the excited state, the photophysical processes involved (photodissociation, predissociation, internal conversion, and intersystem crossing), and photochemical processes (such as the quenching and reactions of the excited state and photofragments). The problems of description of the electronic structure for molecular excited states differ significantly from those for the ground state. In many cases the lack of a variational principle seriously restricts the application of self-consistent-field techniques such as the Hartree-Fock, multiconfiguration Hartree-Fock, and generalized valence bond; in such cases the configuration-interaction method currently provides the only viable approach, and great care is necessary to ensure equivalent descriptions of both the ground and the excited states. The theoretical description of the photophysics and photochemistry of excited states is even less advanced: very simplified models are presently used to study both photodissociation and nonadiabatic transition in polyatomic systems, and potential-energy surfaces for the reactions of electronically excited atoms and molecules are not available for any system of chemical interest.

Problems involving molecular excited states that would be appropriate in the initial program of the NRCC include:

1. Study of the excited states of selected organic molecules, with special attention on resolving the question concerning the nature of the $\pi \rightarrow \pi^*$ singlet states in organic systems; examples of molecules that would yield desired information are C_2H_4, H_2CO, $(HCO)_2$, C_4H_6, and C_6H_6.
2. The study of the excited states of molecules of importance in atmospheric chemistry and combustion, such as CO_2, NO_2, SO_2, HO_2, HCO, and CH; an objective of the study would be an advancement of our understanding of the photochemistry of the atmosphere and the spectra of flames.
3. The study of the reactions of electronically excited atoms and molecules, such as the reactions of $O(^1D)$, $O_2(^1\Delta_g)$, and $CH_2(^1A_1)$.
4. The study of highly excited states or ionizing states in relation to radiation chemistry and radiation damage as related to energy technologies (such as fission and fusion reactors).

A commitment of 100 hours of fourth-generation computer time during the first year, increasing to 300 hours or more during the third year, is estimated as the minimum conducive to efficient development of the project during Phase I of the NRCC. Non-NRCC personnel directly associated with the project might come from three collaborating research groups; we envisage another 10 research groups interacting in this program through workshops.

A study of the electronic excited states of molecules in detail sufficient to yield results of undoubted chemical significance would require commitment of 4-5 scientists and programmers and 300-700 hours of computer time. At least one of the scientists associated with the project should be a PhD chemist on the staff of the NRCC. In addition to carrying on appropriate research, the NRCC staff member should be charged with coordination of the overall project, which should include workshops.

In this latter study, programs must be developed to compute spin-orbit and nuclear derivative matrix elements in order to facilitate the study of the magnetic properties of triplet states and to treat curve crossings in reactions. Software to calculate the electronic structure and potential surfaces is largely in hand, although updating and refinement of the programs would be advisable before the program is initiated.

C. Surfaces, Clusters, and Catalysis

Progress in understanding heterogeneous catalysis would be greatly facilitated by an improved knowledge of the electronic structures of solid surfaces and clusters and of interactions with atoms or molecules adsorbed thereon. The surfaces to be examined should include not only regular, planar surfaces but also the less regular structures obtained by considering clusters of appropriate atoms. In fact, such clusters may be the better models for heterogeneous catalytic sites. Studies of the nucleation of solid phases would be similarly aided by electronic-structure information on clusters of atoms. Both the above types of studies would benefit from comparison of properties of the surface or cluster systems with those of the corresponding bulk solid phase.

Calculations of the electronic structures of surface and cluster systems are so time-consuming that only a few sketchy studies have so far been carried out. By providing computer time and coordinating efforts and programs, the NRCC has an opportunity to improve the technology in this developing field and to speed the generation of practical results greatly.

The specific activity proposed is a detailed examination of some prototype catalytic and noncatalytic systems. A possible set of systems might consist of bulk nickel and copper crystals, Ni and Cu surfaces with and without adsorbed atoms (e.g., H, O, etc.) and molecules (H_2, O_2, C_2H_4, etc.), and clusters of Ni and Cu atoms, both clean and with absorbed species. The calculations should be designed to elucidate the mechanisms of binding of atomic and molecular species to the surfaces/clusters and of chemical reactions among adsorbed species, and should also be relevant to a description of the process of nucleation and growth of Ni and Cu crystals.

The calculations should be carried out in part by *ab initio* methods, but pseudopotential or effective-exchange approximate methods might be studied as well. It will be necessary to build up software for crystal and surface

integrals, for Hartree-Fock studies of such systems, and for appropriate approximations to their correlation energies. Since translational periodicity must be considered in the surface calculations, software and codes developed for typical atomic and molecular calculation will not suffice.

To carry out this activity at a meaningful rate (completing the sample studies within two years), the following resources will be needed:

1. One or two workshops per year to coordinate planned activities of the NRCC and outside personnel.
2. One NRCC scientific staff member, probably with a PhD in chemistry, with expertise in quantum mechanics of solid-state or surface structure.
3. Ancillary support of three non-NRCC collaborating research groups.
4. One technician to help carry out code development and production.
5. 200 hours (first year) to 500 hours (third year) of computer time.

D. Biologically Active Molecules

As the speed and core size of computers increase and as theoretical techniques are developed to include larger numbers of atoms, it is clear that computational resources such as those at the NRCC can make important contributions to biology as well as to chemistry. Quantum chemists have already been engaged in the study of significant biological problems. These include conformational analysis of proteins, phospholipids, sugar phosphates, and drugs; binding of oxygen, carbon monoxide, and metal ions to hemoglobin and to other proteins; correlation of molecular properties with pharmacological activity; electron transfer mechanisms in enzymes; hydrogen bonding between nucleotide base pairs; electronic and geometric structural studies of chlorophylls and other molecules related to photosynthesis; and spin-label characterizations. Definitive computations in all of these areas are limited by hardware and computer time, even at the Hartree-Fock level, and especially where geometry searches are needed to locate structures of minimum energy.

If the computational resources were available, it would be possible to undertake in a comprehensive and systematic fashion other studies not presently feasible. For example, a fragment approach leading to comparisons between the charge distributions and energy levels of all 24 essential amino acids might provide insight into the sequencing of polypeptide chains. Other theoretical projects that could be undertaken include problems central to vision (e.g., *cis-trans* isomerization of retinal), photosynthesis (e.g., transformation of ADP to ATP by phosphate removal), and neural transmission (e.g., action of acetylcholine). The interaction between certain drugs (e.g., actinomycin D) and various receptors (e.g., the nucleic acid backbone) might also be examined by quantum-chemical calculations on appropriate model fragments of the total system. Studies are also recommended of single- and double-stranded RNA interactions and cholorophyll oligomeric and photosynthetic reaction-center characterizations.

Such prototypical projects, and the concepts extracted from them, can lead not only to a better understanding of the mechanisms involved in biological processes but also to more accurate empirical potentials for use in thermodynamic studies, in the interpretation of spectra, and as a guide to x-ray structure refinements. Because of the complexity of the problems, all quantum biology

undertaken at the NRCC should be exceptionally well coordinated with, and guided by, on-going experimental work elsewhere, at universities and at government and industrial research laboratories. Isolated molecules, or even clusters of molecular fragments, can be modified substantially in a biological environment; therefore, the model systems must be chosen for study with great care.

Workshops, to be held preferably on a semiannual basis, would be essential to inform the NRCC staff and the quantum-chemistry community at large about important biological problems. Although the entire area of quantum biology is still in its infancy, it will expand in future years as the field matures. To initiate work in this area during Phase I of the NRCC's operations, the following requirements should be met (on an annual basis):

0.5 staff molecular quantum biologist at the PhD level
0.5 staff quantum chemist at the PhD level
Two workshops on core topics (such as structure and functions of proteins, membranes, nerve, muscle, cell recognition, the genetic code, photosynthesis, vision, and drug-receptor interactions)
Two outside client user groups
Four visiting scientists, consultants, and/or joint staff appointments
50 hours (first year) to 300 hours (third year) of computer time, in anticipation of further expansion in Phase II

E. Solute-Solvent Interactions and Solvation

Solute-solvent interactions have long been a central problem of chemistry. Molecules in solution behave quite differently from those in the gas phase. Most organic, inorganic, and biochemical reactions and interactions take place in solution, and numerous solvent effects have been observed. Quantum chemistry has been reasonably successful in determining the structure of inorganic and organic complexes and in interpreting the nature of the interactions, although the full power of currently available methodologies has yet to be applied. The NRCC should, therefore, coordinate an extensive study of the solute-solvent interactions and solvation. The topics to be included in the study are the following.

1. Structure and Electronic Properties of Molecular Clusters

The NRCC should coordinate a quantum-chemical study of molecular clusters. Optimization of the geometry of clusters in terms of the number of solvent molecules would be extremely useful in interpreting and guiding gas-phase experiments on molecular clusters and in shedding light on the molecular interpretation of solvation. For example, one may study clusters consisting of a metal ion with one, two, and more water molecules in order to determine accurately two-, three-, and many-body interaction energies. Or one may examine how many solvent molecules are needed to solvate an electron.

2. *Development of Alternative Schemes*

In order to discuss meaningfully the electronic structure of a molecule in the bulk solvent, one has to consider its interaction with many solvent molecules. To facilitate a calculation, various alternative models have been proposed in which the solute molecule and a limited number of solvent molecules are explicitly considered, and the effect of the bulk solvent is taken into account as either a dielectric continuum, an electrostatic potential, or a point charge or dipole distribution. The NRCC should coordinate an effort to compare these existing methods, to encourage further development, and to apply promising methods to various solvation problems discussed below.

3. *Study of Spectroscopic and Other Physical Properties of Solvated Molecules*

Such properties should include solvent shifts and hyperchromism and hypochromism in electronic, vibrational, photoelectron, and magnetic-resonance spectra.

4. *Study of Change in Bonding and Reactivity in Solution*

Solvation and molecular-complex formation often cause drastic changes in bonding characteristics and reactivities in solution. Efforts should be encouraged to develop understanding of such solvent effects on bonding and reactivities. A candidate project would be a study of alkyl substituent effects on the proton affinity of amines in solution. A study of selected transition states and reaction paths for some solvolysis reactions of organic compounds would be another. During this research, close collaboration with spectroscopists and physical organic chemists would be essential.

Staff and computing time requirements in Phase I of the NRCC's operations are estimated on an annual basis as follows:

One PhD chemist with computer and quantum chemistry experience
Two workshops on selected topics
Three visiting scientists or external collaborating groups
100 hours (first year) to 250 hours (third year) of computing time

F. Transition Metal Complexes and Organometallic Compounds

Transition metal complexes and organometallic compounds offer a substantial challenge to quantum chemistry because of the large numbers of electrons and orbitals involved and the varieties of modes of bonding. A better understanding of electronic structures of these compounds would afford a theoretical background for their practically important application as homogeneous catalysts. Though quantum-chemical studies in this area have been scattered and limited in scope, recent developments in computational methods and technology have brought a coordinated study within reach. The topics to be included are the following.

1. *Critical Evaluation and Development of Model Potential Methods*

Model potential methods within an *ab initio* framework, which consider explicitly only the valence electrons and replace the core electrons by a model or pseudopotential, have shown promise in several cases both in reproductibility and in economy. The NRCC should coordinate a critical evaluation and comparison of existing model potential methods and stimulate further development.

2. *Study of the Structure and Bonding*

A systematic comparison of the electronic structures should be carried out among complexes of various transition metals with many ligands from "hard" (F, NH_2R, CN) to "soft" (CO, NO, O_2). This study will elucidate semiquantitatively the electronic nature of the bonding, including the factors determining the coordination number and structure of the complex, the stereochemistry of coordination (*trans* or *cis*, axial or equatorial, and linear or bent), and the role of charge transfer and back charge transfer in stabilization of the ground state. The ligand-field theory can be critically evaluated in light of the results.

3. *Study of Spectroscopic and Associated Excited-State Properties*

Included should be studies of electronic, vibrational, photoelectron and x-ray electron, and electron and nuclear magnetic resonance spectra. The emphasis in the electronic and electron spectrum computations should be in obtaining reliable ordering of various excited and ionized states and in identifying their charge-transfer and charge-localization characteristics.

4. *Study of Chemical Reactivities and Catalytic Activities*

The NRCC should coordinate a quantum-chemical study of select groups of chemical reactions to locate the transition states and reaction paths. The reactions should include electron-transfer reactions and various types of elimination, addition, and substitution reactions. Particular attention should be given to multicenter interactions as a source of stereospecificity and stereoselectivity. Theoretical studies should be encouraged on catalytic activities of complexes, particularly as applied to the activation of hydrogen, nitrogen, oxygen, and small hydrocarbon molecules.

All of these studies call for close collaboration between inorganic and quantum chemists, which can be promoted through workshops on individual topics.

Estimated requirements for staff and computing time during Phase I of NRCC's operations in this area are estimated on an annual basis as follows:

One PhD chemist with experience in inorganic chemistry and computation
Two workshops on selected topics
Four external collaborating groups
200 hours (first year) to 425 hours (third year) of computing time

8
STATISTICAL MECHANICS

I. INTRODUCTION

A. Statistical Mechanics and Chemistry

Statistical mechanics plays a central role in the interpretation of experiments on chemical systems. The basic aim is to use information on molecular and intermolecular forces (ideally, provided by quantum-mechanical calculations) to provide a detailed description of the dynamics, structure, and bulk properties of chemical systems. Even when full information on intermolecular forces is not available, the construction of model potentials fitted to experimental data can provide invaluable assistance in the interpretation of spectroscopic experiments (NMR and optical), neutron scattering, light scattering, diffraction experiments (x-ray and neutron), and reaction rate measurements, particularly in condensed phases. Statistical mechanics is particularly effective in dealing with isotope and solvent effects. In the case of monatomic gases, statistical mechanics has led to a detailed synthesis in which the same potential functions correlate phenomena as diverse as gas imperfection and transport properties, molecular-beam scattering data, vibrational levels of van der Waals dimers, liquid- and solid-state structure factors and thermodynamics, and phonon dispersion curves derived from inelastic neutron scattering. Similar work on polyatomic systems has made progress and will be a developing interest of the next few years. The most important aspect of statistical mechanics is the ability to compare and correlate experiments of quite different kinds.

Statistical mechanics overlaps quantum chemistry in its use of potential-energy functions. The empirical determination of such functions by statistical mechanics applied to experimental data provides a valuable cross-check on quantum-chemical methods. When the potential-energy functions can be determined accurately from quantum mechanics (10^{-3} to 10^{-4} eV for intermolecular forces), statistical mechanics can use the results to provide a wealth of information, some of which may be very difficult to obtain by experiment.

One of the most important applications of statistical mechanics is to mixtures, with technological implications for separation processes such as distillation, solvent extraction, and chromatography. Thermodynamic properties of such mixtures as liquefied natural gas are of obvious importance in the utilization of energy. Isotopic mixtures are particularly important for nuclear energy. Aqueous solutions are of special interest, particularly for biology.

Other important applications lie in the area of materials science. For example, recent studies based on molecular dynamics have begun to elucidate the process of fracture. Statistical mechanics also has been invaluable in analyzing the behavior of macromolecules. Developments in computational statistical mechanics can therefore contribute greatly to progress in macromolecular science.

B. Statistical Mechanics and the NRCC

In the following we describe several projects in statistical mechanics recommended for consideration in the initial program of the NRCC. For each project we have pointed out those features that make it uniquely suited for study at the NRCC. In several cases this rests on the capability of the NRCC to focus the activites of several investigations into a unified approach to different aspects of the problem and to ensure continuity of effort over a long period. In other cases, the choice is based on the ability of the NRCC to provide a unique service (as in the centralization of programs and data, and the provision of computer access and services for experimentalists). We believe that these service contributions (subjects C and D listed below) should be regarded as an essential component of the NRCC's operation.

We emphasize that the actual projects to be undertaken will depend on the research proposals received by the NRCC and that we have not listed exhaustively all important areas and possibilities. In the case of the development of a program library and data base for statistical mechanics, initiative from the NRCC will probably be required; individual scientists should be invited to contribute their programs and help to get them running, tested, and as fully documented as possible. The question of providing incentives for this relatively arduous task is important; to some extent, it will be answered naturally as projects are carried through on the NRCC computer.

We note that projects will differ in the extent to which the resulting programs are of wide interest and usefulness. Projects should be specially favored if the resulting programs would be widely applicable (e.g., are "Hamiltonian-independent") in contrast to being useful for only one computation. Programs of the former kind will make a lasting contribution when added to the NRCC library. Of course, program versatility is only one of the considerations in judging whether a project is to receive support; scientific merit is the primary one.

An important function of the NRCC should be to promote contacts and collaborations, through workshops and other modes of communication, among scientists in different disciplines working on related or overlapping problems. Contacts among quantum chemists, statistical mechanicians, chemical kineticists, and macromolecular chemists should be fostered. In the particular case of research on nonspherical molecules discussed below, it would clearly be desirable to standardize as far as possible the model potentials studied by different methods, so as to develop a unified body of knowledge with clear intercomparisons rather than a set of fragments. An NRCC workshop would be invaluable in this area.

II. REPRESENTATIVE PROBLEMS IN STATISTICAL MECHANICS

A. Properties of Polyatomic Fluids and Solids: A Study of Intermolecular Force Models and Properties in Polyatomics

Purpose: (a) To understand the nature of the differences between gases, liquids and solids composed of spherical molecules and those composed of nonspherical molecules that are rigid (in the sense that rotational-vibrational coupling and internal rotation effects are unimportant); for example, determination of the relationship between the underlying intermolecular forces and such observed phenomena as phase transitions (liquid crystal, rotational melting, liquid-liquid separation in mixtures, azeotropes), spectra (NMR, infrared, Raman), and diffraction (neutrons, x-ray).

(b) To determine accurately the intermolecular forces in polyatomics (for example, C_6H_6, HCl, CO_2, CH_4, and C_2H_4).

Background: Systems of spherical molecules (notably the monatomic noble gases) have been studied by experimental, theoretical, and computer-simulation methods. They show only translational melting, exhibit simple spectra and diffraction patterns, and give rise to simple phase diagrams for their mixtures (the excess free energies are small, and no azeotropes occur). Statistical mechanics has provided detailed understanding of the systems, both by analytic solution and by computer simulation. Thus, computer simulation of systems composed of uniform hard spheres early established the existence of a fluid-solid transition, and theories have been developed that accurately describe the fluid properties of such systems. For more realistic potential models (e.g., Lennard-Jones) several perturbation theories were proposed: it was only when accurate simulation results were used to test these theories that it became clear which procedure was best. A similar situation occurred for mixtures, where conformal solution theories were first proposed in 1951 by Longuet-Higgins but were not developed fully until the advent of simulation results for mixtures. More recently, accurate intermolecular-force models have been developed for the monatomic gases.

An extension of such methods is proposed to the more complex but much more numerous polyatomic fluids and solids. Elongated molecules show *rotational* melting, which may occur, as the temperature is raised, before (plastic crystals) or after (liquid crystals) translational melting. Spectra and diffraction results for such substances contain a wealth of information about the intermolecular forces, including the dependence on molecular orientations, which cannot now be extracted because theoretical interpretations are lacking. Mixtures of polyatomics show a wide variety of phase diagrams, including various types of gas-solid, gas-liquid, and liquid-liquid equilibria. An understanding of these phenomena would be of great chemical and technological interest. A few scattered molecular dynamics studies are under way at laboratories in the United States, England, and France using simple model potentials. Extensive studies of water are also being carried out. At present, there is no agreement as to the most suitable intermolecular potential forms or the relative importance of steric (repulsive) and attractive forces.

Procedure: Because of its magnitude and the diversity of the techniques needed, the study would involve several principal investigators, with strong emphasis on the interaction between statistical mechanics and experiments. There would be some overlap with the quantum chemistry group in the selection of intermolecular force models, and also with the chemical kinetics group in the design and interpretation of experiments. The work would consist of two parts, corresponding to the two aims listed above under Purpose.

Part 1: Initially, two kinds of intermolecular potential-energy functions would be studied. The first are models, such as that of Berne-Pechukas (B. Berne and P. Pechukas, J. Chem. Phys. 56:4213, 1972), that account for the departure of the overlap forces from spherical symmetry and have three parameters (energy, diameter, and length-to-diameter ratio). The second is the site-site model, in which potential interactions between sites placed in different molecules are summed (this model has been used by Steele, Chandler, Streett, Powles, Singer, Levesque, and others). The influence of added multipolar forces (dipole, quadrupole, etc.) on these potentials would also be studied. Equilibrium and dynamical properties would be studied by molecular dynamics (MD), Monte Carlo (MC, particularly for phase diagrams), perturbation theories, and integral equation methods, with the aim of determining the relations between observed phenomena and the underlying intermolecular forces. Properties and methods to be studied would include:

(i) Phase diagram and thermodynamic properties of pure fluids, including phase transitions of oriented crystals, plastic (rotating) crystals, and liquid crystals.

(ii) Phase diagram and thermodynamic properties for mixtures: gas-liquid, gas-solid, liquid-liquid, liquid-solid equilibria, and azeotropes.

(iii) Dynamics of molecules, orientational relaxation, time correlation functions and their implications for NMR, infrared, and Raman spectra, neutron and light scattering, and other related phenomena.

(iv) Effect of periodic boundary conditions on conservation of angular momentum in molecular dynamics and size of system. The effect of other boundary conditions would be studied (e.g., propagation of the effect of molecules constrained to lie flat at one boundary).

(v) Simple, inexpensive theories (such as scaled particle and perturbation theories, Mayer-Saupe theory) would be tested against MC and MD results.

Part 2: A detailed comparison would be made between experimental and theoretical (including MC and MD) results for specific compounds, using particular intermolecular force models; the object is to determine accurately the intermolecular forces, to provide methods for interpreting experimental data, and to suggest new experimental studies.

Initially, a study of a compound such as benzene is proposed with one of the intermolecular potential models used in Part 1 as a starting point. MD and MC results would be compared in detail with experiment, first for the gas (pressure and dielectric constant, second virial coefficients, infrared and Raman band increments) and then for the liquid and solid (thermodynamic properties, infrared and Raman spectra, light and neutron scattering, solid-state properties).

The potential would be modified as necessary by changing the form of the overlap model, adding multipolar and anisotropic dispersion terms, considering three-body effects, etc.

Analogous studies that would be of great value include (a) extension to other pure polyatomics, (b) aqueous mixtures (ionic and nonionic), (c) nonaqueous mixtures, (d) *flexible* polyatomics.

Uniqueness: The degree of collaboration needed for this study (see under Procedure above) would only be possible under the auspices of a center such as the NRCC. Carying this project out at the NRCC would ensure continuing benefits, and later work could build on the foundation so laid. The availability of computer time is a crucial element in bringing this about.

Resource Requirements: Molecular dynamics runs for these potential models are expected to require at least two hours of CPU time for 10,000 time steps; more complex potentials will require larger running times. We estimate an annual need of 200 hours of fourth-generation computer time, sufficient for 40–80 runs.

One workshop would be required during the first year. Personnel requirements would be met by 0.5 staff scientist at the NRCC (besides scientists in other collaborating groups).

Molecular dynamics programs are available for the following potentials: (a) Lennard-Jones atoms plus dipolar, quadrupolar, and anistropic overlap terms; (b) site-site Lennard-Jones potential for diatomics; (c) the potential of Berne *et al.*

B. Brownian Dynamics

Purpose: Brownian dynamics is a simulation technique like molecular dynamics except that the particles are assumed to obey Langevin's equation rather than Newton's laws. It is appropriate for simulating solutions at a level at which only the solute particles are represented explicitly, the solvent entering through the frictional force and random force acting on each solute particle and through the solute-solute forces. The latter must be introduced as appropriate solvent-averaged forces.

For the molecular interpretation of solution properties, a description at the level of solvent-averaged forces is appropriate, just as a description at the level of intermolecular forces is proper for the systems described in Section A. However, to determine these solvent-averaged forces by the "inverse process," a reliable and flexible method is needed, such as Brownian dynamics, for calculating the "measurable" properties of model systems in which the solvent-averaged forces are specified. The other known theoretical tools are at best satisfactory for calculating the model behavior in only a single class of measurements, so that different theories must be used, for example, for thermodynamic data, and/or sound absorption data.

The use of Brownian dynamics to calculate various "measurable" properties of solution models can contribute to the solution of important problems such as the following: How does solvation contribute to interionic forces in solution? What is the nature of the hydrophobic bond? How do specific interactions of

ions contribute to salting out, to double-layer phenomena, to properties of polyelectrolytes, to the conformation of biopolymers in solution, and to the rates of diffusion-controlled reactions?

Background: The equilibrium solvent-averaged forces between solute particles are given by the McMillan-Mayer theory (W. B. McMillan and J. E. Mayer, J. Chem. Phys. 13:276, 1945) in terms of the ordinary intermolecular potentials, but the latter are seldom (if ever) well enough known, and there are no reliable techniques for this application of the theory. (The reliability of simulation techniques is limited, in this application, by the boundary problems referred to in Section E.)

The "inverse process" of finding solvent-averaged forces that are consistent with the thermodynamic excess functions has already been applied, using integral equation techniques to calculate the properties of the models. As one might expect, the thermodynamic data turn out to provide insufficient information to permit a distinction between various reasonable models.

Procedure: It is important to apply Brownian dynamics to simulate the properties of the primitive model, for comparison of the results with earlier theories of electrical conductivity (Onsager, Bjerrum, Fuoss, etc.) and other measurements. Simulation of the properties of more refined models that incorporate the effects of solvation would help to determine the most realistic model for fitting the thermodynamic data.

In either case, careful attention needs to be given to the problem of representing the hydrodynamic forces among the ions and other solute particles. These velocity-dependent interactions have in the past been derived for "brass balls in a bathtub" models (i.e., semicontinuum models) that may not be completely self-consistent solvent-averaged models. Since the hydrodynamic interactions are long-range, the reliability of simulation techniques again is limited by the boundary problems referred to in Section E. On the other hand, the Brownian dynamics simulation can equally well be done with replacement of the strict Langevin equation by the generalized Langevin equation, which includes memory effects in the solute-solvent interaction.

It also is desirable to test the underlying assumptions of Brownian dynamics by determining the solvent-averaged forces among small numbers of solute and solvent particles, using direct intermolecular potentials obtained from other work.

Uniqueness: To make the Brownian dynamics method available to the many chemists who could use it in their research, the programs should be documented and distributed by the NRCC. Such programs will be more widely useful in proportion to the degree to which the users can introduce their own concepts regarding the solvent-averaged forces. The availability of the programs can be enhanced by supporting part of the program development work at the NRCC so that it can be coupled to the process of documentation and making the programs relatively machine-independent. The participation of the NRCC in program development is also needed because certain aspects of this process will require very large and fast computers.

Resources Needed:

Time: 2 years
Fourth-Generation Computer Time per Year: 50 hours
Staff Programmer at the NRCC: 0.5

C. Program Library for Equilibrium and Nonequilibrium Statistical-Mechanical Calculations

Purpose: To make available for general use convenient and reliable statistical-mechanical programs.

Background: Scattered among a large number of independent groups of investigators there exist programs in the following areas:

(a) Monte Carlo programs for the calculation of equilibrium properties of single-component and multicomponent systems for specified intermolecular interactions. Techniques are known for the grand canonical and isothermal-isobaric ensembles, in addition to the familiar canonical ensemble procedures. Also extant are procedures for the calculation of free-energy differences between the system of interest and a simpler system of known free energy.

(b) Molecular dynamics programs for the calculation of both equilibrium and nonequilibrium properties of the systems mentioned in (a) above. These include programs that evaluate equilibrium time averages to obtain equilibrium thermodynamic properties, and equilibrum time-correlation functions to obtain transport coefficients and other nonequilibrium quantities. Programs exist also for various nonequilibrium molecular dynamics models in which actual transport processes are simulated.

(c) Phase equilibria. Using free-energy data obtained from (a) or (b), the phase diagram of the system can be studied.

(d) Programs for the calculation of equilibrium properties from perturbation theories. Included here are the effects of many-body contributions to the potential energy upon the properties of interest.

(e) Programs for the calculation of virial coefficients and collision integrals for specified interactions.

(f) Programs for the calculation of solid-phase properties, using techniques such as: (i) quasi-harmonic lattice dynamics, including the use of Rayleigh-Schrödinger perturbation theory for anharmonic effects; (ii) self-consistent phonon theory; (iii) cell models.

(g) Programs for the calculation of liquid solution and dense gas properties by integral equation techniques (e.g., Percus-Yevick, hypernetted-chain, various self-consistent procedures).

(h) Programs for calculating line shapes derived from various combinations of spin Hamiltonians and dynamical models.

(i) Programs for the calculation of isotope chemistry from molecular and intermolecular potentials.

A considerable quantity of results from such calculations is also dispersed among the same research groups. Those obtained by the programs mentioned under (a) and (b) often involve rather large amounts of personnel time and computer time. Their utility to the scientific community would be greatly increased if they were organized into carefully evaluated and easily accessible data bases. Examples of such data are equation-of-state and other thermodynamic properties, tables of pair and triplet correlation functions, and various time correlation functions of interest in the theory of transport processes, as well as certain correlation functions that appear in the perturbation theories mentioned under (d) above.

Procedure: The work involved consists of collecting, standardizing insofar as is practical, and maintaining a coherent library of programs. Assistance must be provided in converting the program for use at other computing facilities. It is essential that the work be done by persons expert in the design and use of the programs and familiar with the limitations of the techniques in question. Programs under categories (a) and (b) must include appropriate procedures for estimating the statistical reliability of the results.

Uniqueness: The activities described here fall among the service functions of the NRCC. As emphasized elsewhere in this report, individual investigators are unlikely to undertake sustained responsibilities for the documentation and consultation required to make their programs suitable for wide dissemination without substantial incentives.

Requirements: This work should begin under Phase I of the NRCC and will require the equivalent of one full-time staff member at the doctoral level in statistical mechanics. The computer time requirement is likely to be quite nominal, perhaps 50 hours per year on a fourth-generation computer.

D. Molecular Dynamics Facility as an Aid to the Experimental Chemist

Background: Over the past two decades many new techniques have been developed for the study of molecular motions in condensed phase. These include, among others, neutron and light-scattering spectroscopy, infrared absorption spectroscopy, ESR, and NMR. The field is no longer equipment-limited--the available techniques cover time and distance scales that can in principle lead to an understanding of the detailed molecular motions involved in the many transport properties of interest to the chemist. It is, however, interpretation-limited. With a plethora of readily available experimental techniques rapidly yielding data, the important question becomes that of extracting from the data the desired understanding of the motions involved.

Fortunately, computer-simulated molecular dynamics has developed to the point where it can be of enormous use to the experimentalist. It can (1) assist in the interpretation of existing data, (2) aid in the design of future experiments to avoid waste of experimental effort, and (3) extend the range or capability of the experimentalist. For example, the expense and effort of observation over large ranges of density (or pressure) can be reduced by a judicious combination of real and simulated experiments.

Unfortunately, these significant advantages of theoretical molecular dynamics are not currently available to a large majority of experimental chemists.

Proposal: The NRCC should make available validated molecular dynamics programs. In order to evaluate, generate, and update such programs, experts in this area should be at the NRCC. These experts, in addition to performing research in molecular dynamics, would also be available to help the experimentalist with the programs. MD programs should ultimately be available for spherical, nonspherical, and flexible molecules with rather general potential energy functions.

Uniqueness: Through NRCC research and communications (workshops, newsletters, etc.) experimentalists would rapidly become aware of the accessibility and capabilities of molecular dynamics as well as its limitations. Widespread interaction between real and computer experiments could be achieved. In this effort, the NRCC would not only provide a service but also act as a catalyst for encouraging cooperative efforts.

Requirements:

NRCC Staff: 1 full-time equivalent staff scientist experienced in statistical mechanics and methods of molecular dynamics
Workshop: 1
Fourth-Generation Computer Time: 150 hours per year equipped with 5 ports capable of interactive remote access by computers, minicomputers, or teletypes

E. Dynamics of Dense Coulombic and Dipolar Systems

Purpose: Enhanced understanding of the thermal equilibrium and transport properties of systems with long-range Coulombic forces. Ultimately, dipole and quadrupole systems also would be included. The new knowledge so obtained would impinge on several specialized fields: modeling the extraction of deuterium's fusion energy; modeling the dynamical structure of light stars and heavy-planet cores; predicting properties of simple liquid metals (e.g., sodium) and molten salts; improving the simulation of water and aqueous solutions.

Procedure: Use of molecular dynamics for a careful analysis of mixed plasma systems, with two positive-ion species immersed in neutralizing baths of electrons. The goals are, first, to evaluate the relative accuracy of the several approaches now in use for treating (1) representation of the long-range forces, (2) motion of the ions, (3) electronic structure; and, second, to develop an approach for treating boundaries in Coulombic systems to make possible nonequilibrium molecular-dynamic treatments of wall collisions, diffusion, viscosity, and conductivity. The plasma conditions and composition would be chosen to facilitate the comparison of results with those of existing integral-equation, cell-model, and Monte Carlo studies.

Uniqueness: The difficulties and complexities in developing accurate treatments for long-range-force systems require an approach by a team with broad experience in quantum statistical mechanics, electricity and magnetism, hydrodynamics, and large-system many-body computing.

Requirements: In addition to two cooperating visiting specialists (0.5 man-year), the help of the NRCC computer staff and consultants to the extent of one-sixth man-year would be required. A workshop is desirable.

A basic computer program, such as that of J. P. Hansen (University of Paris) or W. B. Hubbard (University of Arizona), could be improved to study the three representational questions (forces, motion, and structure). Trial calculations to answer these questions, with 4, 32, and 108 ions, would offer checks against previous work on the one-component plasma. This work would require 10 hours of fourth-generation computer time.

Development work with more general boundaries should follow and would require more time, for this field is novel. An additional 40 hours of computer time would afford a significant impact. Detailed specific systematic studies could then be based on these developmental findings and could be carried out at more generously endowed computing centers of the mission-oriented agencies.

F. Other Studies in Statistical Mechanics

The following problems are singled out for brief mention, though budgetary constraints make their study at the NRCC unlikely during Phase I.

(a) The Liquid-Gas Interface. Satisfactory simulation techniques for determining surface structure and surface tension of simple liquids have recently been developed. Application of these techniques to more complicated liquids with nonspherical, anisotropic molecules is now timely and important. The questions of partial alignment of molecules at the surface, of surface potentials, and of surface absorption in mixtures are all amenable to treatment by computer simulation. Theoretical treatment of these questions should be developed and will be greatly assisted by the results of computer simulations.

(b) Solid-Liquid and Solid-Gas Interfaces. Theoretical and computer studies of adsorption at solid surfaces, whether from gas or liquid phases, are now timely. Detailed information on gas-surface interactions will be available from new surface scattering measurements as well as from absorption measurements and quantum-chemical calculations. In conjunction with computer studies this should lead to advances in the understanding of adsorption equilibria and dynamics of value to colloid science as well as to the study of catalysis.

(c) Chemical Reactions in Solution. Molecular dynamics can be used to make a detailed study of the relative diffusion of pairs of solute atoms in a liquid; in particular the recombination of atoms produced by photodissociation of molecules could be studied in this way, a problem of great interest in the interpretation of results obtained by picosecond spectroscopy.

(d) Quantum Dynamics and Statistical Mechanics of Many-Body Systems. This is a field in which new advances are required. For nearly classical systems, quantum corrections can be calculated by the well-known Wigner-Kirkwood

TABLE 4. Summary of Requirements in Statistical Mechanics (on an Annual Basis)

	Hours of Fourth-Generation Computer Time	Resident Scientists Experienced in Statistical Mechanics (Full-Time Equivalent)	Staff Programmers (Full-Time Equivalent)	Workshops
Properties of polyatomics	200	1/2	0	1
Brownian dynamics	100	0	1/2	0
Program library for statistical mechanics	50	1	0	0
Molecular dynamics for experimentalists	150	1	0	1
Dynamics of dense Coulombic systems	50	0	1/6	1

expansion. For highly quantized systems such as liquid helium, this treatment is unsatisfactory. Development of semiclassical and path-integral techniques for such systems at nonzero temperatures is an important activity for the future; methods for the ground state of liquid helium are already available.

(e) Kinetics of Phase Transformations. Important computational contributions to the kinetics of phase transitions (nucleation theory, spinodal decomposition, etc.) have been made and will continue to be made; this research should be followed with interest by the NRCC.

(f) Critical Phenomena, Magnetic Transitions. A large area of statistical mechanics is devoted to the study of critical points, using Ising and other similar models, renormalization group theory, etc. The NRCC may be expected ultimately to participate in such research.

(g) Developing a Scheme for Relativistic Classical Molecular Dynamics. This would have an impact on several areas of research. Theoreticians could

settle outstanding questions concerning the definition of temperature in relativistic systems. Experimentalists could better understand the results of high-energy beam nuclear collisions.

G. General Comments

It is to be assumed that the programs in the NRCC library will often be made available for use on other computers. Since the constraints imposed by computer systems vary enormously from one installation to another, depending on hardware and system architecture, storage capacity, job mix, charging algorithms, channel capacity, and many other factors, several different versions of a given program may be needed to meet at least approximately the needs of all users.

We would expect that the NRCC would maintain and provide information on subroutines and algorithms for general numerical procedures such as matrix operations, fast Fourier transform, least-squares procedures, and other statistical packages. The provision of graphics facilities ranging from, at one end of the scale, the making of color and stereo movies, and at the other, simple remote graphic terminal support, are regarded as essential.

Table 4 is a summary of the annual requirements in statistical mechanics at the NRCC.

A STUDY OF A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

National Research Council
Division of Chemistry and Chemical Technology
with support from the
National Science Foundation

National Academy of Sciences
Washington, D. C.
1974

Available from
Division of Chemistry and Chemical Technology
National Academy of Sciences–National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Support for this study was provided by the National Science Foundation under Contract No. NSF-C310, Task Order No. 221.

Dr. Philip Handler
President
National Academy of Sciences

Dear Dr. Handler:

The Committee on Science and Public Policy is pleased to forward the attached report, *A Study of a National Center for Computation in Chemistry*, for publication and distribution by the National Academy of Sciences. This report is the most recent and comprehensive of several which delineate a growing potential for significant and exciting scientific and social contributions from chemical computations of great complexity. The problem is to identify appropriate means of deriving the greatest possible benefit from this great potential.

In forwarding the earlier (1970) report, *Computational Support for Theoretical Chemistry*, our Committee encouraged further study by chemists of this whole matter. At the same time, we identified a number of troublesome questions that merited attention. We believe that the present report discusses many of these questions carefully and objectively.

The report recommends proceeding with a computational center dedicated to the advancement of chemistry and related sciences. Phase I of this development would consist principally of assembling the software expertise and documentation necessary for the exploitation of the new opportunities in theoretical chemistry made possible by high-speed, high-capacity computers. Our Committee endorses this Phase I effort.

In discussing Phase II of this project, the report looks ahead to the problem of access to appropriate computational capability. The nature of the access needed and its cost are not at all clear at this time. Networking, communications systems, and computer technology itself are still in a fast-moving developmental mode. The report emphasizes that access to computing capacity is needed, and we agree, but we wish to underscore the possibility that this essential computational capability need not necessarily require a new complex of physical facilities: that remains to be seen. The proposal recommended guards against premature commitment to such action, while providing a mechanism for meeting immediate needs and continuing study of longer-range alternative approaches.

Sincerely yours,

Melvin Calvin
Chairman
Committee on Science and Public Policy

December 1973

OPPORTUNITIES AND NEEDS FOR LARGE COMPUTING FACILITIES IN CHEMICAL RESEARCH

Statement to the Committee on Science and Public Policy

In recent years we have seen great progress in the computation of molecular energies, wave functions, dynamics of reacting systems, and liquid structure by high-speed computers. Some very important work has been done in university centers by semiempirical methods on computers such as the IBM 360/65, CDC 6400, and UNIVAC 1108 on energy states of molecules with as many as 10 - 12 atoms. The work of Pople in prediction of geometries and stabilities of small molecules is a noteworthy example. Reliable methods have been developed for the calculation of potential energy surfaces for three-atom systems, starting from $H + H_2$ and going to the $F + H_2$ system—methods that have been crucial to the advancement of our understanding of reactive collisions. There have been semiquantitative calculations of the strength of hydrogen bonds and protein conformation. The structure of liquids as complex as H_2O has come under attack as a result of the collaborative work of Stillinger and Rahman. These problems require appreciable time on computers of the size of CDC 6600 and 7600. Such facilities are presently available only to a small group of chemists, principally those on the staffs of the national laboratories.

It is not difficult to envision that the development of software oriented to problems in quantum chemistry, statistical mechanics, and structural chemistry could have a major impact on problems of interest to chemists with present state-of-the-art hardware. The present study of a national center for computation in chemistry has brought this point home not only to those engaged in the study but to a wide audience of chemists—not just theoretical chemists—to whom we have distributed preliminary drafts for comment and from whom we now have a wide range of support. Typical problems that could be attacked with our present state of theoretical knowledge include reactions of NO_x in the stratosphere, energy states of molecules in surface catalysis, energy and reactivity of excited electronic states (i.e., photochemistry), molecular distribution functions, and autocorrelation functions near discontinuities (phase transitions and critical phenomena)—a spectrum of important and solvable problems of interest to groups studying them.

To undertake such work we need to do more than just add to present budgets for computing in chemistry. We need to bring in to-

gether people who will work on software development–scientists who will participate in this development and attack important problems in chemistry. For this, "we recommend the establishment, as a national resource, of a computational center with facilities and personnel dedicated to advancing chemistry and related sciences through widespread, innovative, and intensive use of high-speed computational equipment. This mission is to be accomplished by making appropriate facilities available to a wide group of scientists, by providing and developing software to expedite and upgrade computer use, by encouraging and supporting research efforts to build new and more effective computational methods, and by carrying out an informational and educational program to bring the benefits created through the center to the widest possible scientific public."[1] The establishment of such a resource can be expected to lead to new developments in both theoretical and experimental chemistry.

At this time we are not suggesting bricks and mortar, nor are we suggesting major hardware acquisition. Rather, we believe that the crucial need is to assemble an inhouse staff of specialists in chemical research computation and make available a resource to the scientific community. This group could lease hardware, e.g., through the ARPA network, using unutilized time on large computers at national laboratories and elsewhere. The resource envisioned is not to replace what is more readily done at home institutional computing centers. We do envisage university and other groups tying into the resource from their home center and utilizing developed software to provide output to particular problems of the working experimental chemist. The cost of this Phase I operation for personnel and leased computer time is estimated initially at $0.5 - 1 million per year, two to three times the present NSF budget for computing in quantum chemistry. If this phase proved successful, we might go to Phase II, which could involve hardware and require a capital investment on the order of $10 - 20 million with an annual operating budget of $3 - 5 million. Clearly, any planning on Phase II would come only out of demonstrated need from Phase I.

We would like to obtain an endorsement of our recommendation that a committee responsible to an appropriate contracting agency (e.g., the National Academy of Sciences, or one of the several existing university consortia) be commissioned to bring this national resource into being.

[1] Recommendation (1), Chapter X.

In making our recommendation we have addressed ourselves to the following considerations: (1) the impact of such a specialized center on the funding of university computing centers; (2) the impact on the funding of regional computing centers; (3) the impact on the development and utilization of computer networks; (4) plans for a quantum chemistry program exchange; and (5) the interfacing of the center with the scientific community.

These matters have been discussed at length in the drafting of this report, which has received widespread endorsement by many who have worked on or have been consulted about the study.

Dr. Harrison Shull, who has long investigated quantum chemistry and who has fostered the Quantum Chemistry Program Exchange at Indiana University, and Dr. Martin A. Paul, Executive Secretary of the Division of Chemistry and Chemical Technology of the National Research Council, who has been the staff officer for our study, are with me to answer any questions you may have and receive suggestions.

20 October 1973

Jacob Bigeleisen
Member of the Executive Committee
NRC Division of Chemistry
and Chemical Technology

Contents

Summary

During the past decade, the impact of the electronic digital computer has revolutionized the conduct of research in chemistry, a pervasive, basic science with applications of great importance to human welfare. The computing needs of chemists now extend over an exceedingly broad spectrum, encompassing among major categories the need for automated control of instrumentation and the associated analysis of experimental data, the need for mechanized storage and retrieval of information, and the need for direct numerical solution of complex systems of differential equations such as are encountered in theoretical approaches to chemical problems by the methods of quantum chemistry and statistical mechanics. Meeting this latter category of need is the concern that has led to the study described in this report.

Pioneers in the use of computers to solve theoretical chemical problems usually had access to computers that had been procured by their institutions for other, more general purposes, and that were in many cases subsidized wholly or in part. Because of changes in the support policy for the university computing centers, and because of increasing demand on the part of other users, such subsidized time is becoming increasingly less available, and chemical investigators dependent on large-scale computation are finding themselves priced out of the market. Support for chemical computation is at best unevenly available, and access to the more powerful computers is restricted by conditions extraneous to scientific merit. Few universities can afford the largest computers now in commercial production and functioning in mission-oriented research institutions.

At the same time, chemical computation has come of age. Theoretical methods of established reliability are on hand, and in prospect, for solving important chemical problems inaccessible to or too costly for experimental approach. Of perhaps even greater significance, a close coupling of theoretical and experimental techniques affords for many problems a more powerful and more reliable mode of attack than either experiment or theory alone. At present these possibilities for exploiting theoretical and computational advances are severely underutilized, both because the necessary computational facilities are not sufficiently available and because the necessary

coordinated effort has not yet been made to provide for practical and easy access to tested computational programs and equipment by the wide group of potential users in the chemical research community. Recent developments in communication techniques and software make feasible a major increase in the utilization and exploitation of computational resources by the scientific community.

These facts suggest that the time is opportune for the creation, as a national resource, of a national center that will take full advantage of progress in theoretical methods applicable to chemistry, provide appropriate facilities for theoretical and associated computational research, and make these methods and the knowledge derived therefrom accessible to all who have scientifically cogent uses for them. Alternative methods of administering and operating such a center are explored in this report, and models exist in national facilities supported in various specialized areas of research by the National Science Foundation, the Atomic Energy Commission, and other federal agencies.

Detailed planning for the center is left to a committee that should be appointed for the purpose, responsible to an appropriate contracting organization that could initially be under the aegis of the National Academy of Sciences.

RECOMMENDATIONS

1. We recommend the establishment, as a national resource, of a computational center with facilities and personnel dedicated to advancing chemistry and related sciences through widespread, innovative, and intensive use of high-speed computational equipment. This mission is to be accomplished by making appropriate facilities available to a wide group of scientists, by providing and developing software to expedite and upgrade computer use, by encouraging and supporting research efforts to build new and more effective computational methods, and by carrying out an informational and educational program to bring the benefits created through the center to the widest possible scientific public.

2. We recommend that a committee responsible to an appropriate contracting organization (perhaps the National Academy of Sciences) be commissioned to bring this national resource into being.

I Background of the Report

Chemistry, the science and technology of materials, pervades almost all apsects of science and human welfare—medicine, agriculture, production of energy, population control, protection of the environment, conservation and effective use of natural resources, and development of new materials with specialized properties and uses. The conduct of chemical research and its applications have been revolutionized during the past decade by the impact of the high-speed electronic digital computer. Chemists actually were among the earliest groups of scientists to develop extensive use of computers, not only to facilitate ordinary calculations related to the reduction and interpretation of experimental data, and for automated control of measurement itself, but for the essentially large-scale computations required to solve the equations of quantum mechanics applied to atoms and molecules containing more than a single electron, and to solve equations for model systems of many interacting molecules by the methods of statistical mechanics.

Pioneers in the application of computers to chemistry were able to gain access to computers usually by capitalizing on time made available either free of charge or through heavy subsidization on an institutional computer procured for other purposes. Many institutions continue to provide their chemists, among other research investigators on their staffs, with "free" time up to the limit of the capacity of their computing systems.[1] Increasingly, however, with changes in governmental support policies, institutionally subsidized computer time is becoming less available, and chemical investigators dependent on large-scale computation are finding themselves priced out of the market in competition with other users of the same computers. Few individual universities will be able to afford the economies of scale inherent in such high-speed, high-capacity modern computer systems as Control Data Corporation's Model 7600 or International Business Machines Corporation's Model 360/91.[2] Meanwhile, chemical com-

[1] See Appendix information.

[2] Three universities presently operate IBM 360/91 computers: Columbia, Princeton, and UCLA (which has a second one utilized by the university's Health Sciences Computing Facility). No university has for general use a CDC 7600, though access is available to the facility at Lawrence Berkeley Laboratory for members of the faculty of the University of California.

putation is flourishing.[3] Specialists in computer science and technology have found a place in chemistry, and increasingly reliable theoretical methods are coming into use as a practical means of solving important chemical problems, many of which are inaccessible to direct experimental solution or are accessible only at unacceptable cost.

The growing awareness of constraints on the efficient use of computers in chemical research has been voiced particularly by quantum chemists, though the growing impact of computers in general was noted in the National Academy of Sciences (NAS) 1965 survey of chemistry.[4] In November 1965 the Division of Chemistry and Chemical Technology of the National Research Council (NRC) convened with support from the National Science Foundation (NSF) a conference on Uses of Electronic Computers in Chemistry. The conference, chaired by Harrison Shull, Indiana University, and attended by 25 invited participants, was in part a response to a proposal generated informally at the January 1965 Sanibel Island Symposium on Quantum Chemistry and dealt with problems in the financing of computation, in publication of computational methods and programs, in education, and in the organization and administration of computing centers. In May 1970 another 2-day conference was convened under the chairmanship of Peter G. Lykos, Illinois Institute of Technology, on Computational Support for Theoretical Chemistry. The 44 participants included quantum chemists and representatives of governmental agencies funding chemical research. Among the focal points for discussion was the concept of a national center for theoretical chemistry. A consensus was reached, with some reservations as to detail, that a case exists for a national computing facility devoted to theoretical chemistry, having the following attributes: its primary mission should be to evolve sophisticated hardware - software computer systems and perform computations vital to the further development of chemistry; consistent with this primary mission, it should provide facilities for research on new quantum-mechanical methods and develop systems for optimum use of these methods by specialists in many areas of chemistry; its scientific policies should be designed to have major impacts not only on chemistry

[3] For a nontechnical review, see A. C. Wahl, *Scientific American*, 222 (4), 54 (1970).

[4] *Chemistry: Opportunities and Needs*, Committee for the Survey of Chemistry, National Academy of Sciences, Washington, D.C., 1965.

but also on related fields such as molecular biology, solid-state physics, and atomic physics.

A covering letter to the President of the National Academy of Sciences from Harvey Brooks, then Chairman of the Academy's Committee on Science and Public Policy who reviewed the conference report, stated that the Committee had found the report interesting, stimulating, and worthy of attention by the scientific community, though not fully persuasive in its documentation of the need for a national center. He identified several important questions having to do with specialized computing centers in general, not with chemistry specifically, that needed detailed study before the Committee was prepared to endorse the feasibility and desirability of proceeding with such a national center for quantum chemistry. The Committee encouraged quantum chemists to explore these questions with the help of knowledgeable individuals in other areas of chemistry as well as in related disciplines including computer and communication technology, the study to be in the context of the much larger problem confronting the universities in the cost-effective delivery of computing services.

The NRC Division of Chemistry and Chemical Technology accordingly prepared a proposal to NSF for the support of the present study. This study aimed at exploring in depth the feasibility and desirability of a national center for theoretical chemistry with sophisticated hardware - software systems dedicated to the solution of important chemical problems involving computational research, such problems to be restricted in general to those not capable of being treated with currently available computer systems. The task assigned to the study group was originally conceived as follows:

1. to identify important chemical problems susceptible to solution by such a center;
2. to characterize appropriate quantum chemical strategies and tasks;
3. to identify special problems of computer science involved;
4. to specify hardware and software requirements; and
5. to explore and evaluate alternative models of operation for the center.

In addition to these substantive matters relating particularly to chemistry, the study group was to be concerned with:

1. the impact of such a specialized center on the funding of university computing centers;
2. the impact on the funding of regional computing centers;
3. the impact on the development and utilization of computer networks;
4. plans for a quantum chemistry program exchange; and
5. the interfacing of the center with the scientific community.

Kenneth B. Wiberg, Yale University, was appointed by the President of NAS as chairman of the study, with a planning committee consisting originally of L. S. Bartell, University of Michigan; R.B.K. Dewar, Illinois Institute of Technology; Frank E. Harris, University of Utah; F. A. Matsen, University of Texas at Austin; Harrison Shull, Indiana University; and L. C. Snyder, Bell Laboratories. The last five members had participated in the previous conference.

The planning committee met in December 1971 and commissioned panels involving some 40 additional specialists in various areas of chemistry and related disciplines to prepare position papers on scientific objectives, proposed computational facilities, and alternative methods of administering and financing a national center. In May 1972 the committee then convened a study group of some 25 invited participants with diverse scientific interests (including members selected from the panels but also others who had had no previous connection with the study) to discuss the position papers and identify problem areas still to be resolved. The group met again in August 1972 and reached a consensus that a national center for computation in chemistry was feasible and desirable, with certain constraints on how it should be administered.

As this study has progressed, it has become increasingly apparent to the participants that it is now technically feasible not only to tackle advanced computational problems relevant to theoretical chemistry but also to make existing computational tools and programs practically accessible and usable by a much wider circle of users on a nationwide basis. The concept of the center here proposed therefore has been broadened to include within its scope not only advanced research by specialists in theoretical and computational methods but also the development of facilities and services available to a wide variety of potential users.

Meanwhile, parallel developments have provided information bearing on some of the issues the study group was originally charged to consider. The concept of a research and educational

computer network linking the universities has been reviewed in a series of three seminars organized late in 1972 by EDUCOM with support from NSF.[5] The National Science Foundation has also initiated an expanded program of research relative to a national science computer network. An experimental computer network developed by the Advanced Research Projects Agency (ARPA) of the Department of Defense is in operation. This project, linking some 24 existing regional computing centers, was aimed at providing an efficient and reliable communications system (utilizing message switching techniques) in which computer resources such as programs, data, storage, and special hardware could be shared among computers and among many users.[6] It is particularly timely, therefore, to consider the development of a national center for computation in chemistry as a specialized national resource, access to which could be provided to users remotely through a computer network.

With regard to the impact of such a specialized center on the funding of university computing centers, a survey of university computers conducted for NSF by John W. Hamblen is illuminating.[7] The total expenditures for computing in higher education in the United States in 1971 were $500 million, increasing at a rate of about $60 million per year. These expenditures include all educational, research, and administrative uses. Of the total, the overall

[5] *Networks for Research and Education: Sharing Computers and Information Resources Nationwide*, M. Greenberger, J. Aronofsky, J. L. McKenney, and W. F. Massy, Eds., MIT Press, Cambridge, Mass., 1973

[6] L. G. Roberts and B. Wessler, *Computer Network Development to Achieve Resource Sharing*, American Federation of Information Processing Societies Conference Proceedings, May 1970; H. Frank, R. E. Kahn, and L. Kleinrock, *Computer Communications Network Design – Experience with Theory and Practice*, AFIPS Conference Proceedings, May 1972. Telenet Communications Corporation, a majority-owned subsidiary of Bolt Beranek and Newman, Inc., with principal offices in Washington, D.C., filed in October 1973 an application with the Federal Communcations Commission for authorization to establish and operate a commercial nationwide communications network specifically designed for computer-to-computer and terminal-to-computer communications based on the "packet-switching" technology developed for the ARPA network.

[7] John W. Hamblen, *Inventory of Computers in U.S. Higher Education, 1969 - 1970*, Office of Computing Activities, National Science Foundation, Washington, D.C., 1972; see particularly Figure III-12 on p. III-17.

support from NSF, according to a computation by Peter G. Lykos[8], accounts for only about 2 - 3%. Therefore, it is clear that the funding of a specialized computing center dedicated to chemical research will have a negligible effect on the university computing centers as whole, although it could affect a few individual centers where research grants for chemical computation are now a substantial source of support.

Besides the members of the planning committee, a large number of other individuals have contributed to the actual writing of this report, particularly, John C. Light of the University of Chicago, Felix T. Smith of Stanford Research Institute, and A. C. Wahl of Argonne National Laboratory. A list of all those who participated on the panels and in the study group appears at the conclusion of the report. These individuals do not necessarily subscribe to all details of the conclusions reported here, but it is hoped that the report fairly addresses each controversial issue raised in the course of the meetings and in correspondence with other scientists who could not otherwise participate. Deliberate effort has been made to enlist the interest and elicit the opinions of a broad spectrum of the chemical community by distributing widely a preliminary draft for comment among several hundred informed critics in university chemistry departments and in other institutions, public and private, where chemical research is conducted.

[8]Presented in *Computational Needs and Resources in Crystallography*, National Academy of Sciences, Washington, D.C., 1973.

II Computing Needs in Chemistry

The computing needs of chemists are distributed over an exceedingly broad spectrum. A majority of research chemists now use computers in their investigations, but both the type of utilization and the computing requirements in terms of time and machine power vary widely. In terms of type of utilization, three main areas may be identified. First, there is the use of computers as direct adjuncts to experimental operations, as in computer-controlled apparatus and in other applications where direct, short-term interrelationships exist between experiments in progress and the immediate analysis of the data being produced. Automated x-ray and neutron diffraction studies of crystals are a case in point.[1] A second area of computer use is in the range of activities encompassing data retrieval and searching. Functions in this area include the use of data banks, bibliographic searches, various sorting processes, and other specialized services associated with libraries. The third main area of computer use is in theoretical calculation, both for purely theoretical projects and in theoretical support of experimental programs.

Many of the chemist's computing needs in the area closely coupled to experimental apparatus are best met by local minicomputers, and the acquisition and operation of such computers is becoming recognized as an integral part of many experimental research programs. This report is not concerned with further analysis in this area, nor is extensive further consideration given here to the data-bank and library area. Library services involve problems, the solutions of which call for different expertise than is represented in our study group, and it is not clear that the opportunities presently recognizable in the chemical-research computation area yet extend to the library-service area. This chapter is therefore restricted

[1]See *Computational Needs and Resources in Crystallography*, National Academy of Sciences, Washington, D.C., 1973. This report of a conference organized under the chairmanship of Walter C. Hamilton in April 1972, by the NRC Division of Chemistry and Chemical Technology with support from NSF, surveyed crystallographic computing from a standpoint corresponding to that of the report, *Computational Support for Theoretical Chemistry*, mentioned in Chapter I.

mainly to an analysis of computing needs for chemical-research calculations.

Even within a scope restricted to chemical calculations, there still remains a range of computing needs, and it is probably fair to observe that an optimum development of computer hardware and software would develop new needs for many chemists in addition to stimulating needs for computations already identified as desirable. At one end of this computing spectrum lie the activities of a large class of chemists who, by choice or necessity, have small computational requirements (say, up to the equivalent of a few hours per year of time on a CDC 7600 computer). The actual calculations made by these users vary considerably, and include the routine use of packaged quantum-mechanical programs of limited accuracy, simple analysis of experimental data, elementary x-ray analysis, and solution of simple differential and integral equations. Often these users are producing results that are "intermediate" in the sense that they are a guide to further experimentation, perhaps after a choice between simple theoretical models. If computation were made sufficiently convenient, many users in this class would carry out many more theoretical calculations than they do at present, and the expansion of computer use in this area would constitute an example of the new needs a good computer development could generate. The key to meeting the needs of this class of users is *convenience*; the convenience of making a computation will often determine whether it will be done.

The foregoing discussion indicates that the real need for users with small requirements would be support for an increased variety of easily used computer programs. This support could take the form of better libraries of computer programs [as would result, for example, from an expansion of activities of the present Quantum Chemistry Program Exchange (QCPE)], or from new activities designed to produce the desired results. With the growth in effectiveness of communcations facilities, a viable alternative to program libraries is the maintenance of sets of programs on machines that can be accessed remotely by a wide community of users. For example, it is now practical to maintain sets of quantum-chemistry programs on a computer accessible over telephone lines with teletype equipment, and to organize these programs so that input and output can be conveniently handled by such equipment. Whatever methods are used to provide access to programs, the small users would also benefit from an increased scope of activity directed

systematically at increasing the supply of useful programs, at improving their reliability and documentation, and at providing educational and informational services to facilitate the use of these programs.

Those chemists with moderately larger computational requirements form another class of users. They require on the order of 10 hours equivalent CDC 7600 time per year, machines of current capacity being adequate, and computing being an important, if not dominant, aspect of their research. In this category current inequities in the availability of computational facilities begin to have a significant impact. Although third-generation computers (such as CDC 6400 and IBM 360/65) are widespread, located at most universities, industrial laboratories, and government laboratories it is only at those institutions actively subsidizing computer use by chemists that the adequate facilities are *available*. The scope and direction of research for many chemists in this class is severely limited not by their desires nor by the quality of their research, but by the fact that adequate computational facilities are not available and cannot be predicted to become available in the foreseeable future. This situation can apply both to experimentalists who need considerable computer time for reduction of data and comparison with established theory, and to theoreticians who wish to attack more complex problems or to develop improved methods.

This class of users must be assured access to adequate computational facilities for scientifically qualified research programs. This assurance would be provided to most moderate-scale users if they had funds for the purchase of needed computing services, but a number of alternative approaches can be identified and are analyzed in later chapters. Moderate-scale users also share the needs outlined above for the small-requirements users, as extensive computer studies are usually preceded by smaller, preliminary investigations or rely in part on existing computational methods for which library programs could be available.

The moderate-scale user stands to benefit greatly from research on the development of new computational methods applicable to his field. Many users need improved support in this area. These users also might be helped greatly by further development of *algebraic* (as opposed to *arithmetic*) programming systems and by other improvements in computer languages. There is a particular need for improved methods of handling large amounts of data.

Finally, there are a few chemists whose planned or current calculational activities require access to the largest and most sophisticated equipment available, and whose projects could consume hundreds of hours or more on such equipment. Figure 1 presents a range of problems for which the computational requirements have been estimated according to methods currently in practice. It emphasizes the logarithmically increasing scale of the computational capability required with increasing complexity of the chemical system. The mere possibility of undertaking truly large-scale calculations presently depends upon free or heavily subsidized computer access, and only a handful of fortunately located chemists can expect to carry out such calculations.[2] It is unreasonable to claim that there is a "need" to satisfy all planned projects in this category, but some represent benchmarks or milestones that will have great value in answering key questions or in helping to determine directions for future progress. The need in this area is for a few major projects, chosen on the basis of scientific merit rather than through accidental factors, to be given access to the computing facilities they require. The feasibility of solving problems calling for calculations on a large scale depends, furthermore, not only on available computing power but also on organization of that power through intensively designed systems software and computational algorithms that are beyond the capabilities of uncoordinated systems designers. The increase in computational capability from this type of advance is likely to be many powers of ten greater than that to be expected from improvements in hardware and in the economics of computation.

The foregoing analysis suggests a two-phase approach, spelled out further along in this report, to meeting the foreseeable computing needs of chemistry: (1) immediate attention should be given to the problem of systematic development, maintenance, and improvement of software for computational chemistry, in readily accessible form, including remote access by telephone line or by more sophisticated means such as a computer network; (2) longer-range planning should be instituted on the provision of access to

[2]Compare the scale of Figure 1 with the present pattern of support for computation in theoretical chemistry summarized in Appendix Table 6.

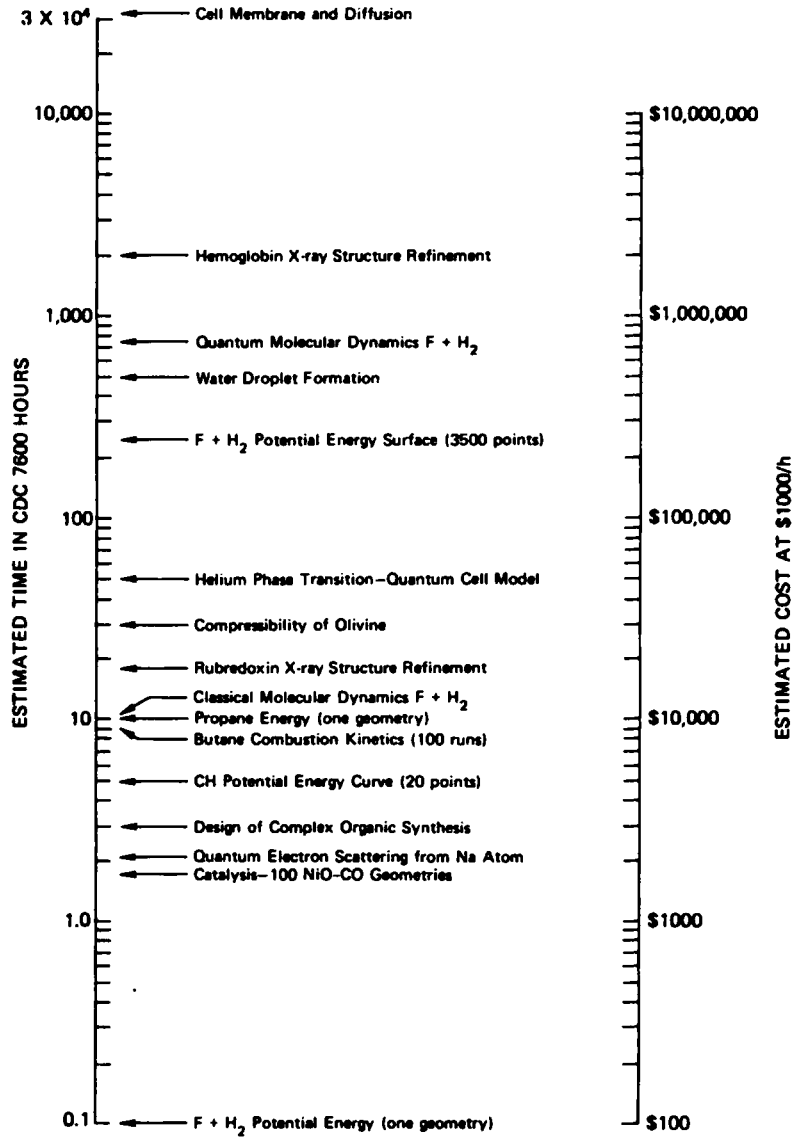


FIGURE 1. Computer requirements estimated for solving representative chemical problems. The times for the systems generally in the lower third of the figure represent actual experience. For the other systems, the times have been estimated by extrapolation of methods currently in practice. The times are for computer operation only and do not include associated program development.

large computers, with specialized hardware and supporting services, foreseeably beyond the resources of most individual generalized research institutions, including the universities.

III The Opportunity

Chemists are now at an impasse in applying the available knowledge in attacking a variety of problems of great national importance. Substantial and rapid progress on these problems cannot be achieved without bold and decisive action to forge a new mechanism for organizing computational capability for chemists. The opportunity is not to make easier and cheaper those kinds of research that can now be accomplished, or to provide funds for currently ongoing projects. The opportunity is to create the appropriate mechanism for doing what otherwise cannot or will not be done. Access to computational facilities beyond those currently available to the chemical community at large actually can be expected to lead to new approaches in theoretical and experimental chemistry.

The following are examples of research areas where theoretical approaches are needed and cannot be undertaken without computational resources now unavailable to most investigators.

Photochemical processes have proven to be of value in the synthesis of medicinally important chemical products, and of course are involved in the basic process of photosynthesis from which all food is derived. They are important also in the chemistry of the atmosphere, the design of materials for the capture and storage of solar energy, the effects of ultraviolet radiation on living organisms, and the degradation of materials generally when exposed to light or ultraviolet radiation. The control of photochemical processes is a major goal of chemistry. These processes involve molecules or ions in transient, energetically excited states that cannot be studied in detail by currently available experimental methods. It is now clear that theoretical methods are approaching the capability of dealing with such processes.

Heterogeneous catalysis is involved in the synthesis of many commercial products, the reforming of petroleum intermediates, and the destruction of noxious by-products of combustion. Virtually no unequivocal information about the structures and dynamic properties of the substrate-surface molecular species in these catalytic processes has been generated by the existing methods of experimental chemistry. Presently feasible applications of theory have

given answers of marginal validity. The important chemical systems are large, involving many-electron atoms, and may require computation of many configurations of the reactants. The potential value of clues to the design of catalysts to be gained through insights derived from basic theoretical and semiempirical approaches is very great. These approaches cannot be systematically explored without access to the necessary computational resources.

The design of *organic conductors* is currently based on model calculations, e.g., the Hubbard model. However, Hubbard calculations have been performed on systems of such small numbers of atoms that prediction of the properties for practical systems is ambiguous. The resolution of such ambiguity is well within the capability of organized effort to develop the necessary computational resources.

High-temperature superconductors open vast new technological possibilities. The fact that high-temperature superconductivity occurs near a phase transition makes it plausible that theoretical studies of intermolecular forces and phase transitions may contribute to the design of higher-temperature superconductors. Such studies cannot be undertaken without access to the necessary computational resources.

The atmospheres of the earth and sun are relatively inaccessible to experiment. Theoretical cross sections for the interaction of atoms and molecules with electrons and light are required to assay the compositions and evolutions of these atmospheres. Systematic research in this area calls for organized effort to develop the necessary computational resources. As a specific example of a directly related important problem, concern has been publicly expressed that emission of nitrogen oxides from a fleet of supersonic transport planes could reduce the concentration of ozone in the earth's atmosphere to such an extent that increased solar ultraviolet radiation reaching the earth's surface would be detrimental to health. A definite response to this concern is well within the capability of chemists if the requisite organized computational support were available.

Protein conformation and *enzyme - substrate interactions* are related to function in living organisms. Theoretical studies on a major scale could provide the potential functions and statistical descriptions of the conformations of these systems. The structure and function of *biological membranes* also can be modeled in calculations of molecular dynamics. The computational requirements

are exceedingly large (see Figure 1 in the preceding chapter), but the rewards in terms of understanding nerve conduction and other vital processes are potentially great.

These research areas are characterized not only by the fact that a necessary component of the means identified for their exploration involves extensive computation but also by the fact that their computational requirements cannot be met without a degree of coordination and organization that is presently lacking. Despite the trend towards decreasing unit costs of computation, these problems will be opened to solution only when the available hardware is coupled with effective development of new computational methods and software. The scope of this needed development is far too great to be achieved without planning that transcends individual research efforts.

The opportunity, then, is to develop a new mechanism that will facilitate attack on large and important chemical problems by those investigators who are committed to their solution.

IV Computation in the Advancement of Chemistry

THEORETICAL CHEMISTRY TODAY

In chemistry, as in other sciences and indeed in many other areas of life, the rapid development over the past two decades of high-speed computers at ever lower cost per item of information processed has effected enormous changes. In theoretical chemistry one need look back only 20 years to recall when the hydrogen molecule was the only chemical system whose electronic structure was quantitatively successfully understood, and even then the excited vibrational and electronic states were poorly and erroneously represented. At that time many intuitive approaches to the theory of electronic structure of molecules were competing with each other, but it was well recognized that most agreements of theoretical calculations with experiment were fortuitous and depended on an accidental cancellation of errors or of omitted terms. Fortunately, the theoretical work in chemistry then existing had brought the profession to the point where it recognized its frustrations and was sufficiently sophisticated to exploit promptly and effectively the new computational possibilities opened up by the development of high-speed computers. The electronic structures and other properties of many diatomic and a few triatomic systems are now solidly established on a calculational basis, and the magnitude and nature of the remaining uncertainties are well understood.[1] Among specific examples of reliable calculations with well-understood error limits are the following:

1. The binding energies and potential-energy curves of many diatomic molecules, typified by the results obtained for NH, AlO and F_2.[2] In these cases experimental values were in doubt. Similar

[1] See recent reviews by H. F. Schaefer III, *The Electronic Structure of Atoms and Molecules. A Survey of Rigorous Quantum Mechanical Results*, Addison-Wesley Publishing Company, Reading, Mass., 1972, and A. C. Wahl, "The Calculation of Energy Quantities for Diatomic Molecules." *MTP (Med. Tech. Publ. Co.) Int. Rev. Sci.: Phys. Chem.*; Ser. 1, I, 41 (1972).

[2] NH: W. J. Stevens, *J. Chem. Phys.*, 58, 1264 (1973). AlO: P. S. Bagus and B. Liu, to be published. F_2: G. Das and A. C. Wahl, *J. Chem. Phys.*, 56, 3532 (1972)

calculations of spectroscopic accuracy have been carried out for OH.[3]

2. Electron affinities of several species important in the atmosphere.

3. Infrared intensities predicted *a priori* for OH and NO^+, both important emitters in the atmosphere.[4]

4. Prediction of a low-lying bound triplet state of O_3 and excitation energies to other states.[5]

5. Van der Waals well depths.[6]

6. Geometries successfully predicted for transient organic molecules and ions, such as CH_2 (where the computed nonlinear geometry was later verified by spectroscopic observation after an earlier reported incorrect linear assignment),[7] and $C_3H_5^+$.[8]

7. Accurate energy surfaces governing prototypical chemical reactions, such as for the systems H_3, FH_2, and LiHF.[9]

8. Heats of chemical reactions involving closed-shell systems.[10]

9. Nuclear magnetic resonance shielding of first-row elements.[11]

[3] G. C. Lie and J. Hinze, *J. Chem. Phys.*, 57, 625 (1972).

[4] OH: W. J. Stevens, G. Das, A. C. Wahl, M. Krauss, and D. Neumann, to be published. NO^+: F. Billingsley, *Chem. Phys. Lett.* (in press).

[5] P. J. Hay and W. A. Goddard, *Chem. Phys. Lett.*, 14, 46 (1972); C. W. Wilson and A. C. Wahl, *J. Chem. Phys.* (in press).

[6] H. F. Schaefer III, D. R. McLaughlin, F. E. Harris, and B. J. Alder, *Phys. Rev. Lett.*, 25, 988 (1970); P. J. Bertonini and A. C. Wahl, *Phys. Rev. Lett.*, 25, 991 (1970).

[7] J. M. Foster and S. F. Boys, *Rev. Mod. Phys.*, 32, 303 (1960); J. F. Harrison and L. C. Allen, *J. Am. Chem. Soc.*, 91, 807 (1969); S. V. O'Neil, H. F. Schaefer III, and C. F. Bender, *J. Chem. Phys.*, 55, 162 (1971).

[8] L. Radom, P. C. Hariharan, J. A. Pople, and P. V. R. Schleyer, *J. Am. Chem. Soc.*, 95, 6531 (1973).

[9] H_3: I. Shavitt, R. M. Stevens, F. L. Minn, and M. Karplus, *J. Chem. Phys.*, 48, 2700 (1968); I. Shavitt, *J. Chem. Phys.*, 49, 4048 (1968); B. Liu, *Intern. J. Quantum Chem.*, 5S, 123 (1971). FH_2: C. F. Bender, P. K. Pearson, S. V. O'Neil, and H. F. Schaefer III, *J. Chem. Phys.*, 56, 4626 (1972); C. F. Bender, S. V. O'Neil, P. K. Pearson, and H. F. Schaefer III, *Science*, 176, 1412 (1972). LiHF: W. A. Lester, Jr., and M. Krauss, *J. Chem. Phys.*, 52, 4775 (1970).

[10] P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta.*, 28, 213 (1973).

[11] R. Ditchfield, "Studies of Molecular Properties Pertaining to Electronic Charge Distribution: A Comparison between Theory and Experiment," in *Critical Evaluation of Chemical and Physical Structural Information*, National Academy of Sciences, Washington, D.C., 1974.

10. Conformations and bond lengths in carbohydrates.[12]
11. Vertical excitation energies of urea and other polyatomic species.[13]

Although the ability to evaluate fundamental electronic properties of atoms and molecules and the forces between them represents currently the most sophisticated and well-developed numerical machinery, a parallel development has taken place in the ability to use this central and fundamental atomic and molecular information to predict spectra, collisional and dynamic phenomena,[14] and the structures of liquids and solids.[15] In addition, the increasing success of computationally intensive theoretical chemistry in prediction and in complementing experimental information is forming a new bridge of communication between theorists and other members of the chemical community. This new development promises to usher in an era in which modern calculations are viewed and used as a viable tool, alternative or complementary to experiment, and deeply interwoven with the current traditional inventory of techniques and equipment used by the chemist.

The desirable integration between theory and experiment has taken place due to five contributory factors:

1. development of reliable theoretical methods;
2. translation of these methods into *usable, reliable, general* computer codes;

[12]L. Radom, G. A. Jeffrey, and J. A. Pople, *Carbohydrate Research*, 25, 117 (1972). This study illustrates how a careful investigation of a simple model compound (methanediol) can be used as a guide to understanding properties of large molecules.

[13]S. T. Elbert, Ph.D. thesis, University of Washington (1973); manuscript in preparation for publication by S. T. Elbert and E. R. Davidson. The point of particular interest is that the first excited singlet state for urea is predicted to be the $n\rightarrow\pi^*$ $^1A_2$ state. For urea and all other amides, empirical and semiempirical assignments had been based on the assumption that this state was a $\pi\rightarrow\pi^*$ $^1A_1$ state. The potential for such contributions to correct assignment of electronic bands for polyatomic molecules is enormous.

[14]J. R. Krenos, R. K. Preston, R. Wolfgang, and J. C. Tully, to be published (molecular beam and trajectory studies of the reaction of H^+ with H_2).

[15]F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, 57, 1281 (1972). This study of water structure is based on computation of the simultaneous interactions of 216 water molecules.

3. awareness of important chemical problems by creators or users of codes;
4. awareness by experimentalists of theoretical capabilities;
5. performance of the relevant calculation.

These requisites, however, have been present simultaneously in relatively few instances, at relatively few places, and are far from optimal as the following considertaions will make clear.

As indicated in (1) and (2), fundamental to and indispensably interwoven with this new capability of theoretical chemistry has been the development of "master" computer codes that can be applied to a wide variety of chemical systems. The most widely used are IBMOL, BISON, POLYATOM, MOLE, ATMOL, and more recently. ALCHEMY. Each of these codes represents from 10 to 20 man-years of analytical, numerical, and program development. Although considerable effort has gone into making them machine-independent and reliable, it still constitutes major effort for the scientist unfamiliar with them to implement and use these highly complex computing systems effectively. Further, as implied in (3) and (4), the full potential of these powerful computing systems can be realized only if they are put into the hands of scientists trying to solve *real* problems on which the computation can have impact.

Finally, once an important problem for computation has been identified as well as defined, it is seldom that personnel or machine resources are available to the conceiver of the physical problem. This combination of circumstances exists today only at a handful of centers, and to a very limited extent there. These centers are the national laboratories supported by the Atomic Energy Commission (Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, Brookhaven); the National Bureau of Standards; the Centre Européen de Calcul Atomique et Moléculaire at Orsay; and in a more restricted way the facilities at several universities and private research institutes.

There are three basic ways to ensure the coincidence of the five essential factors previously identified. The first is to create and distribute "canned," well-documented, easily machine-transferable, master codes which are transparent to use and impossible to abuse. The second is to bring the user physically or by electronic means to a center where the desired code is routinely operating, transparent to use, and impossible to abuse. The third is for the compu-

tation to be done by an expert as a service, as collaboration, or out of personal interest. To date, the outstanding examples of theoretical capabilities have been accomplished through the third mode. The second has been used to some extent on an informal, inconsistently funded, and often "buddy" basis, primarily at the centers mentioned above, while the first has been explored at a minimal level through the QCPE and through personal program distribution.

To promote effective use of each of these modes, the following would be necessary.

For Mode 1

An order of magnitude increase must go into the process of making major programs machine-transferable, reliable, and foolproof. Some consideration has gone into making such programs as MOLE, POLYATOM, IBMOL, ALCHEMY, and BISON satisfy these conditions. However, they all involve extensive effort to implement on a new machine and to be used intelligently. A great need exists also to develop software interfaces between such major codes and the nonexpert user, such as those suggested by the *Interview* and *Procedural Executive* modules of the BISON code.[16] Such interfaces can educate the user, elicit proper and unambiguous input from him, and then direct a reliable calculation to answer him. Another way of implementing mode 1 would be to send an installer - instructor with the code. This again is often done on an informal basis.

For Mode 2

Primarily long-term support must be provided at existing installations for program development and maintenance, for computer cost, and for travel or communication of the user to the center. The establishment at all of the current major centers of a permanent staff dedicated to facilitating use at that center of desired codes, coupled with computer support for the user, and travel or communication support, would accomplish many of the goals of the proposed national center. The use of long-range communication for many problems in mode 2 is particularly attractive since it is a characteristic of much chemical computation that the questions asked and the results desired can both be expressed concisely,

[16] A. C. Wahl, P. Bertoncini, K. Kaiser, and R. Land, *Intern. J. Quantum Chem.*, 3S, 499 (1970).

even though the operations required to obtain the desired answers may be exceedingly lengthy. For example, for a given molecule composed of known atoms and with a known number of electrons, one may wish to know what the energy levels are at several specified locations of the various nuclei. In practice, one may also need to begin the problem with a specification of the principal electron orbitals that are expected to be involved in computing the structure. Obviously, such simple questions are easily adapted to being asked and answered via long-range communications systems.

In order to achieve this successfully, however, the computational programs should be presented in the simplest possible way for convenience of address by a comparatively inexperienced user. Much progress has already been made in adapting some chemical programs to comparatively general use in this way, but clearly much more activity of the sort would be valuable. The developmental problem involved is in some ways similar to the problem of developing an instrument from an experimental model, usable by the expert who invented or designed it, to a commercial device capable of being operated by an inexperienced technician for routine measurements, as stressed by BISON.

For Mode 3

The presence of a service-oriented staff at existing installations as discussed above, coupled with computer support, would be one way to implement mode 3. The alternative is the establishment of a physically new national center.

A major opportunity is thus at hand for the exploitation of existing programs by a potentially wide circle of users, provided that access to the appropriate program and computer can be facilitated. It is practically possible to provide such access in a way that will diminish enormously the cost in time and effort now severely limiting the alternative mode of procedure, namely, the use of a program exchange and the adaptation, debugging, and testing of the program on a different computer.

AREAS OF SCIENTIFIC OPPORTUNITY

It is important that this study identify areas of chemistry where significant advances will be facilitated by access to excellent computational facilities. Several discipline-oriented panels of scientists

were charged to consider such areas and came up with a large number of problems, some with absolutely enormous computational requirements. The more typical requirements are discussed here, as well as the extraordinary. In what follows, one should recognize a distinction between the requirements for and utility of "production" calculations and those of "benchmark" calculations, which may push current techniques and facilities toward their limits. These latter calculations can establish standards against which approximate calculations may be compared, theoretical limitations or predictions tested, or numerical answers provided to specific problems of high scientific value.

Potential-Energy Surfaces

In terms of current and future value to chemistry as a whole, one of the most important goals of theoretical chemists clearly is to be able to compute with an appropriate degree of accuracy, the energy and other properties of ground and excited electronic states of atomic and molecular systems. This impinges upon essentially every aspect of chemistry, from spectroscopy to crystal structure, from thermodynamics to reaction rates.

The properties of the ground electronic states of polyatomic systems are of particular interest for reactive systems, for intermolecular potentials of stable systems, for microwave and infrared spectroscopy, etc. Two problems are associated with the computational requirements: the accuracy of the wave functions required and the number of nuclear configurations needed. The value of an approximate wave function or potential-energy surface for small systems even over a limited region is today of significant value. It allows the determination of equilibrium molecular geometry and also the estimation of saddle-point geometry and shape for reactive systems. Since high absolute accuracy for these purposes is not demanded, the calculations can be extended by approximate or semiempirical quantum chemistry methods to a large variety of systems. Although such calculations do not put excessively high demands on computational facilities, the volume of computation could be high. Thus economy, availability, and convenience of computation in using these methods of quantum chemistry would be of significant value to a large class of chemists, and are objectives that should be vigorously pursued.

As the required accuracy of the calculation or the size of the system increases, the computational requirements escalate rapidly (see Table 1). One can infer from the proliferation of approaches to the problem of calculating electronic energies to high accuracy—extended configuration-interaction (CI), multiconfiguration self-consistent-field (MCSCF), Bethe - Goldstone approach, Green's function techniques—that the field is still in a rapid stage of development and that general optimal approaches have not yet been determined. Since this type of computation often strains the currently available computers and can consume large amounts of time, it is likely that "developmental" aspects will dominate "production" aspects for some time. It is the goal of many computational chemists to make such accurate computations convenient, reasonable in cost, and generally applicable.

TABLE 1 Comparative Times to Solve the Structure of the Ethanol Molecule by Computational Approaches of Increasing Refinement in a Single-Configuration SCF Study[a]

Computational Method	Time (s)
Extended Hückel	1
CNDO/INDO	5
Ab initio minimal basis (STO-3G)	150
Ab initio split valence basis (4-31G)	1000
Ab initio polarized split valence basis (6-31G*)	6000

[a] J. A. Pople, personal communication. The times, representing seconds of central processing on a UNIVAC 1108, are extremely approximate.

Examples of calculations near current limits of accuracy are found in two studies previously noted. One is the computation of the $F + H_2$ potential-energy surface by standard CI methods.[17] The calculation required about 5 minutes of CDC 7600 time per point and the surface of 150 points thus required about 12 hours. The relative accuracy was apparently very good (estimated to be about 0.03 eV or 4 kj/mol), such that the results could be used with some confidence in dynamics calculations. The other study is a computation of the potential-energy curve of the CH molecule by MCSCF methods to high relative accuracy,[18] reproducing the

[17] S. V. O'Neil, P. K. Pearson, H. F. Schaefer III, and C. F. Bender, *J. Chem. Phys.*, 58, 1126 (1973); C. F. Bender, S. V. O'Neil, P. K. Pearson, and H. F. Schaefer III, *Science*, 176, 1412 (1972).

[18] G. C. Lie and J. Hinze, *J. Chem. Phys.*, 57, 625 (1972).

spectroscopically observed vibrational-energy levels within a few reciprocal centimeters, and providing reliable information on the long-range potentials. The computation required roughly 15 minutes per point (5 hours total per curve), but the excited-state potential-energy curves were obtained as well. Although these calculations are significant "benchmark" calculations, it should not be inferred that large-scale production of potential-energy surfaces of such accuracy will be forthcoming in the near future. As noted previously, extension to systems containing more and heavier atoms would dwarf these already large computational requirements. Thus, the impetus to develop more efficient theories and computational methods will remain strong.

Properties of excited electronic states of atoms and molecules are required for the qualitative understanding and quantitative treatment of many processes, including absorption and emission of radiation, photochemical reactions, and ionization. The type of information needed may often require the evaluation of "difficult" operators between different electronic wave functions, such as the nuclear-kinetic-energy and spin-orbit coupling operators. The variation in the accuracy required of such wave functions or potential-energy surfaces is great—often the equilibrium geometry and ordering of excited states is most important, but for some problems (in collision processes, for instance), the locations and energies of crossings or psuedo-crossings of surfaces may be needed with high accuracy. With current approaches it is usually more difficult to obtain comparable accuracy for both ground-state and excited-state surfaces, although the more sophisticated CI and MCSCF calculations do yield some information on ground and excited states from the same calculation. The improvement, extension, and utilization of these methods will definitely require extensive computation with excellent facilities.

In summary, the demand for moderately accurate quantum calculations of molecular ground-state wave functions, energies, and geometries seems likely to grow, and the need can be met by a center to provide economy, reliability, and convenience in these calculations. An important further goal of computational chemistry is the production of highly accurate methods that are also relatively economical. Current methods for computing highly accurate surfaces require 10^1 to 10^3 hours of CDC 7600 time per surface (10 electrons, 3 nuclei)—a costly enough undertaking so that

routine production of such surfaces at a center would seem undesirable. Development of better methods and the computation of a few "benchmark" calculations would, however, be appropriate, and the center should stimulate research aimed at improving our ability to calculate accurate surfaces at modest cost in time and effort.

Collision Processes

Included in the dynamics of gas-phase processes are chemical reactions, energy transfer processes, and electronically non-adiabatic events. Knowledge about these processes, and in particular rather detailed information about them, are fundamental to the understanding of a number of important problems. Although the importance of the shape of the potential-energy surface of a reacting system in determining the detailed outcome of the reactive event was realized in the 1920's, it is only in the last decade, with the advent of high-speed computers, that detailed theoretical investigation has become possible. In addition, the development of molecular-beam, laser-excitation, and chemiluminescence techniques has permitted the experimental investigation of a variety of systems in unparalleled depth. There has been, of course, a concomitant rise in the theoretical effort in this area, increasing by about an order of magnitude in the decade from 1960 to 1970. It is likely that efforts at the current high level will be sustained or increased in the 1970's and, as theory catches up with experiment, the area will become a mature discipline of great value.

As computational capabilities have improved, close interactions between conceptual theory, calculation, and experiment have become possible. This mode of operation will surely become more common in the future. An instructive example is provided by a current study of the scattering of excited (metastable) atoms by ground-state atoms, specifically $He(2^3S)$ by He. For collision energies between 5 and 30 eV this scattering was expected to be controlled by the two adiabatic excited potential curves of symmetry $^3\Sigma_g$ and $^3\Sigma_u$ that separate to $He(2^3S) + He(1^1S)$. Electronic structure calculations had already provided reliable potential curves for these excited states, from which features of the expected elastic scattering pattern such as rainbow structures can be predicted quantitatively by scattering calculations. To the surprise of the experimenters, only one set of these features was seen—the rainbow structure associated with the $^3\Sigma_u$ curve—and none of the expected features of the $^3\Sigma_g$ curve appeared.

Since the $^3\Sigma_g$ curve was known to be intersected by a $^1\Pi_g$ state, it was suspected that rotational $\Sigma - \Pi$ coupling was responsible for destroying the expected simplicity of the $^3\Sigma_g$ scattering. With the help of molecular electronic structure codes at the University of Texas, this hypothesis could be tested. Accordingly, *ab initio* calculations were made of the rotational coupling matrix element and potential curves as a function of distance, a close-coupled scattering calculation was carried through, and the results confirmed the suggested explanation. In addition, it was possible to predict the precise angular location and the expected amplitude of the inelastic excitation into the 2^3P state. A time-of-flight measurement enabled this inelastic process to be identified experimentally, but only because the theory predicted just where the comparatively weak signal should be looked for. The experimental results agreed in both location and amplitude with the prediction, thus providing a convincing confirmation of the initial intuitive explanation, and a gratifying validation of both the electronic structure and collision calculations.[19] This was a clear-cut example of the mutual reinforcement all these approaches can provide when focused together on the same problem.

Other areas for which studies in molecular collision dynamics may have significant value include upper-atmosphere physics, combustion problems, and theory of arc discharges. In these areas, as for lasers, the systems are often complex, with many chemical components, nonthermal distributions, and spatial inhomogeneities. Therefore, coupled rate equations as well as elementary events may have to be considered in such studies.

The dynamics of atomic and molecular collisions on electronically adiabatic potential-energy surfaces can be followed by solving the classical equations of motion or by solving an appropriate form of the Schrödinger equation. In addition, semiclassical methods appear promising for markedly improving the accuracy of classical trajectory results for highly quantized systems (those with relatively few accessible states). The area should be of significant scientific value and a large amount of work will continue.

The method one uses and the computational requirements depend strongly on the type of information desired. As a rough guide, requirements have been assessed as follows for computation

[19] R. E. Olson, R. Morgenstern, D. C. Lorents, J. C. Browne, and L. Lenamon, *Phys. Rev.*, **A8**, 2387 (1973).

of those quantities that can be measured experimentally for some systems—rate constants, envelopes of internal-energy distributions, angular distributions, and individual elastic, inelastic, and reactive cross sections for systems of three and four light atoms. (It is not implied that computations should be restricted to reproducing experimentally measured quantities, but rather that they should be aimed towards interpreting experimentally unavailable data.)

Classical computations can conveniently be done on medium-size computers, with time the only serious consideration. The times for classical trajectory studies of three-atom systems are estimated to be 0.1 hour for rate constants, 1 hour for averaged energy and angular distributions, and perhaps 10 hours for all individual cross sections for light systems. Classical calculations appear to be the method of choice for the less-detailed information, except perhaps for light systems at low temperatures (where tunneling may be important), and even there the semiclassical extensions may prove adequate with little extra effort.

In the past few years, numerical methods have been developed that have reduced computational requirements for the quantum calculations by two orders of magnitude, and current research may provide accurate approximations that will reduce requirements even further. Nevertheless, the accurate computation of many collision processes remains a large, unsolved problem. The requirements to compute by current methods cross sections over a reasonable energy range for a simple reactive system such as $F + H_2$ are still roughly 10^3 hours of CDC 7600 time. However, given the incentive of access to good computational facilities, rapid progress can be expected in this area with the development of much more efficient theories and techniques.

In summary, highly accurate studies of the dynamics of molecular collision processes can use large amounts of computer time and power. However, techniques for obtaining less-detailed information with modest computational requirements exist and are useful. Development of more efficient accurate methods should continue and will, occasionally, make heavy demands on computational power.

The broad area of electron - atom - molecule - ion scattering, including inelastic collisions (with electronic, vibrational, and rotational excitation), ionization processes, and dissociation is under intensive investigation both experimentally and theoretically. Such processes are of great significance to chemists, physicists, atmos-

pheric scientists, and other research investigators in related fields. Although high-energy processes (exceeding 1 keV) can generally be treated adequately by perturbation theory, the interesting phenomena that occur at lower energies require much more complex treatments. In particular, electron - atom and electron - molecule collisions can be reduced to manageable $N + 1$ electron problems dominated, in large measure, by correlation effects. Thus, the more sophisticated quantum chemistry methods must be used, requiring significant computational power. The alternative approach of solving the Schrödinger equation by close-coupled integro-differential equations also requires extensive computation. In view of the importance of these processes, it seems likely that the impetus for development of methods for computing cross sections will continue for some time.

STATISTICAL MECHANICS

Large-scale computations are playing a dominant role in statistical mechanics. By simulation techniques it has been possible to get more detailed information than experimentalists have been able to obtain. By computing exactly (in a numerical sense) properties of some well-characterized simple models, it has been possible to check on detailed theoretical predictions and the underlying hypotheses. In this way it has been possible to find qualitatively new effects just as in real experiments, such as the solid - fluid phase transition for hard spheres and the existence of a hydrodynamic vortex mode contributing significantly at all and especially at long times to transport coefficients.

These significant achievements have come about through developments of techniques uniquely applicable through large-scale computers. The Monte Carlo method of sampling configuration space was developed specifically for computer evaluation of the thermodynamic properties of any system of molecules for which the potential of interaction between two such molecules is given. Similarly, the molecular dynamics method of following the trajectories of a few hundred particles interacting through a given pair potential can give not only the thermodynamic properties but the transport properties as well.

These unique computer methods have overcome the immense mathematical difficulties that had restrained progress in the field, and have led to new insights as well as new specific results. In the

latter category it is important to point out the prediction of properties of materials on which experimentation would be extremely difficult to carry out, such as superdense matter in the interior of stars.

Unfortunately, the computer time requirements for the application of these methods are enormous if any property is to be evaluated with reasonable accuracy. Long runs are required to reduce the statistical error. The machine time requirements are such that no present university computing center can support such an effort. For all practical purposes these calculations are hence presently carried out only at industrial and primarily at mission-oriented government computer centers. A national computing center could provide an essential service not only by providing computer time to a few significant problems, but also by eliminating much duplication. For example, considerable time is spent in many cases in generating the trajectories of particles. These could be stored on tapes, and a subsequent analysis, which frequently does not require massive computer time, could be made by different investigators for different problems using the same trajectories.

Such a service can assume even more important proportions in the future as more complex systems will be studied. The effort so far has been primarily confined to the study of relatively simple model systems that have spherically symmetric interaction potentials. Investigations involving nonspherical interaction potentials, such as the behavior of diatomic molecules, are just in their infancy. The generation of their trajectories is an order of magnitude more time consuming. More ambitious projects such as the predictions of the properties of water and of ions in water have been initiated. For such important systems it eventually may be necessary to remove the pair-wise additive interaction approximation, which would further enormously increase the demands on computer time.

In order to deal with water adequately, it will be necessary ultimately to remove the classical description and find ways to deal with quantum mechanical many-body systems. It has been possible already, by Monte Carlo methods, to deal with a few hundred quantum mechanical particles obeying Bose statistics at absolute zero. Much effort will be required to find methods of extending these calculations to finite temperatures and Fermi statistics. Of still greater difficulty would be the development of a method for

quantum molecular dynamics calculations to predict, for example, the unusual transport coefficients of helium.

Better access to computers would enable many more investigators to consider such problems. It would stimulate the use of the calculations in more areas because the same trajectories can be used for several different purposes. It would encourage applications of these methods to numerous more realistic systems. Finally, by cooperative effort, new numerical methods could be developed not only to deal more efficiently with existing calculations but also to develop entirely new ones for situations as yet too complex.

In another context, the accurate evaluation of many of the analytic approaches in statistical mechanics depends heavily on computational power available. We may cite the evaluation of higher virial coefficients, the solution of the Percus - Yevick or hypernetted chain equations for realistic potentials, and the evaluation of the high-temperature expansion for magnetic systems as examples in which the analytic theory has been pushed to maturity and extensive computation can yield most valuable results.

OTHER AREAS

The development of a number of other areas of chemistry during the next decade may depend significantly on the availability of good computational facilities. In most of these areas the use of computers to date has not been both widespread and intensive, but growing interest in the field will certainly increase the demands. The areas mentioned below are, of course, only a sampling, and major computer utilization in areas not mentioned is to be expected as well.

Surface chemistry and heterogeneous catalysis are emerging as areas in which a significantly increased theoretical and experimental effort will be forthcoming in the next few years. The theoretical problems are complex, involving not only the electronic structure of the surface but also those of the molecular species. Although qualitative models of these interactions are available and a rudimentary understanding of the nature of the controlling electronic processes is emerging, relatively few attempts at a quantitative theoretical understanding have been made. However, in addition to the actual bonding properties of the surface and molecular species involved, the problem of accommodation coefficients is

increasingly a subject of theoretical research. Quantitative approaches based on the classical dynamics of the molecule and the atoms of the local surface involved have been attempted (with an assumed potential), but the area is still under intensive theoretical investigation from a different point of view, treating the interaction of the molecule with the surface via the phonon spectrum of the semi-infinite solid.

Sets of nonlinear coupled differential equations are used to describe a variety of physical phenomena, from the kinetics of multicomponent chemical systems (as in combustion, smog formation, and chemical lasers) to electrical networks. Since the mathematical analysis of such systems far from steady-state solutions is so general as to be almost useless from a quantitative point of view, important problems can now be resolved only numerically by computation. Although considerable theoretical analysis is continuing, the quantitative resolution of such problems will continue to require extensive computation.

There are many other areas of chemistry in which computation will become increasingly important. Analysis of experimental data, development of possible synthetic routes to large organic and biologically significant molecules, and computer simulation of the performance of different equipment designs, are only a few of the necessary and creative uses which are now in their infancy.

Although the need for computational facilities in the various areas we have mentioned is well recognized, it is also clear that in many areas the current state of theory and computational methods is far from optimal. A primary goal of a computation center for chemistry should be the stimulation of theoretical progress leading toward both better conceptual understanding of the physical phenomena involved and more efficient and innovative use of computers in obtaining quantative information about these phenomena.

ROLE OF MINICOMPUTERS

Work in the field of x-ray crystallography[20] and a study under way in the area of quantum chemical calculations[21] indicate that

[20] R. Shiono, in *Crystallographic Computing*, Proceedings of the 1969 International Summer School on Crystallographic Computing, Munksgaard, Copenhagen, 1970.

[21] W. H. Miller and H. F. Schaefer III, "Large Scale Scientific Computation via Minicomputer," project supported at the University of California, Berkeley, by the National Science Foundation, 1973.

minicomputers can play a significant role in computational chemistry.

In this respect, the significant limitations of minicomputers are as follows:

1. limited main memory;
2. limited external memory;
3. unavailability of sophisticated operating system software;
4. slow execution speed.

Limits on main memory are currently significant but this limitation may be expected to disappear in the near future. There is now available one minicomputer that has essentially unlimited memory and addressing capability.

The limits on external memory are more significant. This report envisions that 10^{12}-bit memory units are required for manipulation of certain large-scale data bases, archival functions, etc. Such memories are likely to be relatively expensive for some time. Thus, such applications are not likely to be suited to minicomputers for the foreseeable future.

Operating-system software is of critical importance in implementation of large-scale software systems involving interactive access and the accommodation of very large programs with associated segmentation and loading problems. The investment required here is too large for manufacturers of minicomputers. For example, the PDP 11/45 is distributed without any operating software matching its hardware capability. The effort of implementing a system such as BISON on a minicomputer would be greatly hampered by such limitations.

For a large range of small problems (<4 hours CDC 7600 time) slow execution speed is less important than it might seem since the time on a minicomputer is still manageable (<256 hours) and the cost may actually be less. However, truly large computations become impractical on minicomputers merely because the elapsed time becomes too great.

In summary, even though minicomputers have an important, and as yet largely unexplored, role in computational chemistry, the impact is in the area of smaller problems and the need remains for major facilities of some kind. It can be anticipated that improvements in computational methods made possible by the activities of a national center will further increase the range of practical applications of minicomputers and make development of their capabilities even more important.

The use of minicomputers is also significant from the point of view of their established communication potential and the possibility of using them as part of an integrated system involving task distribution. For example, an appropriate "terminal" for accessing a large central computer may consist of a powerful minicomputer capable of handling locally many smaller-scale problems or parts of large problems.

V Alternatives in Meeting Computing Needs

A number of alternatives for meeting the computing needs in chemistry are identified and discussed in this chapter. These alternatives differ in cost, in the manner and extent to which they do (or do not) meet the computing needs, in their possible effects on existing institutions, and in their possible influence on future programs for the distribution of computer services.

1. TOTAL RELIANCE ON LOCAL COMPUTER CENTERS

Most chemists currently obtain computational services through this means. Typically, the chemist is served by an institutional computer center (university, government, or industrial laboratory) that has the responsibility for providing a broad range of services to a large number of users, and where computation in chemistry accounts for a small fraction (say, 2 to 5 percent) of the users but considerably more of the computer time. Historically, these centers have usually attempted to provide for all the needs of all their users, and until fairly recently many centers have been able to provide hardware appropriate to the needs of the chemists. Such hardware, designed for large-volume rapid calculation, is typically needed by the chemists and a few other research users, and is quite different from that which is optimum for time-shared systems handling large numbers of small users or for the usual run of data-processing problems. It has become increasingly difficult for individual centers to meet completely these disparate needs, and a significant number of individual centers have abandoned the attempt to satisfy the class of users that includes the chemists. Other centers have encountered severe economic difficulties, and some universities have even discontinued their centers. Therefore, if this alternative is to be effective, it must include the allocation of sufficient funds to cover the costs of maintaining the necessary capacity in individual computer centers.

This alternative has some attractive features. It provides a good degree of flexibility and responsiveness to the needs of the chemists, partly because it is controlled locally and partly because the

chemists form a sizable percentage of the knowledgeable users who exert a disproportionate influence on the computer operations. This alternative also supports local institutions in the maintenance of a facility that can perform a variety of other educational and research tasks, thereby tending to strengthen the institution's computational program.

Unforfunately, there are also several important negative features. This alternative is a costly solution to the computing problem, requiring far more money than has been available to support computing in chemistry, and even far more than the national center that is considered as another alternative. This alternative would also be contrary to an emerging governmental policy that is diverting direct computational support away from local centers. Further, this alternative will do little to accelerate the cooperative development of interchangeable software, which may hamper the rapid growth of new applications of theoretical techniques by a wider body of chemists.

If this alternative is not selected, there will be two main effects upon the existing local centers. First, there will be diminished pressure to expand facilities to accommodate a small number of specialized chemist users. This will presumably be accompanied by a decrease in revenue from such users. Some observers have expressed the view that for many local centers, in their present financial plight, such a development could be catastrophic. However, an informal survey of a number of chemists who are large computer users and their respective local centers indicates that (1) few such users provide enough revenue to offset the costs associated with extending the local center capacity to meet their relatively unique needs, and (2) considerable use will still be made of a local facility to run smaller programs and to check out programs under development. No large user who was contacted claimed to use an overwhelming preponderance of his computer time for large-scale production. Large-scale users tend to carry out large numbers of compilations and smaller runs, along with production work which does make extreme demands on the capacity of whatever equipment is available.

The magnitude of the effects just mentioned will depend upon the scale of the alternative selected. For example, assume selection of an alternative involving the setting up of a center containing the equivalent of one CDC 7600 available to accept work currently at

local centers. One might expect perturbations on local center usage in approximately the proportion that the work bears to total capacity, probably of the order of a few percent.[1]

2. USE OF FACILITIES AVAILABLE IN GOVERNMENT LABORATORIES

This alternative considers supplementing the computational services currently available to chemists by making available to them computer facilities located in various government laboratories. It could include arrangements whereby capacity on government computers is deliberately created for use by chemists, as well as arrangements whereby chemists obtain access to "unused" or "excess" capacity on government computers. Deliberately created computer capacity will require funding, and many obvious difficulties are attendant on adjoining computing for chemists to specific mission-oriented programs that would logically not be funded from the same sources. From a practical viewpoint, then, this alternative deals with possibilities for free or low-cost distribution of computer services not consumed by the mission for which the computer facilities were obtained.

One favorable aspect of this alternative is the substantial time that may be available on equipment suitable for chemical calculations. Such equipment is located at a number of AEC installations, including Argonne National Laboratory, Los Alamos Scientific Laboratory, Lawrence Livermore Laboratory, and Lawrence Berkeley Laboratory, and at several Department of Defense installations. Many of these installations have personnel sympathetic to the computing needs of chemists, and who have some understanding of their computing problems and expertise in solving them.

Serious drawbacks, however, include the difficulty of obtaining guarantees or predictions of available computer capacity. It is not in the interest of a mission-oriented facility to admit that it has computer capacity in excess of that needed to support its mission, even when such excess occurs as a result of intelligent planning for future needs. Moreover, because computer needs tend to increase continuously while computer capacity must be changed discontinuously, there will necessarily be large fluctuations in the computer capacity chemists will be able to obtain from any individual government laboratory. It is also unrealistic to expect that the chemists will enjoy priorities of use comparable to those available to mission-oriented users. Finally, this alternative shares with several others the disadvantage of not contributing to software development of computer use by a wider group of chemists.

[1] See p. 00 for relevant figures.

It should be noted that the aforementioned drawbacks would not be serious if facilities in government laboratories were used to supplement more stable solutions to the chemists' computing problems. If temporarily excess government computer capacity were coupled to networks, it might be possible to obtain substantial benefits not only for the chemists, but also for a far wider class of prospective users.

3. USE OF COMMERCIAL SOURCES OF COMPUTER TIME

A number of companies sell computer time, and some have made at least limited attempts to reach large-scale scientific users. University Computer Corporation contributed to the development of the MOLE quantum chemistry system, making at least parts of it available to users of its computers. Computer Science Corporation, IBM, CDC, and other companies sell time on computer systems large enough to handle the research requirements of theoretical chemists. The rates available depend critically on whether prime shift or priority use is involved. The lowest rates are for block time in nonprime shifts, using standard operating systems, as shown in Table 2.

The main argument in favor of this alternative involves the availability of service on fairly short notice. At least one of the commercial vendors maintains a national network with convenient regional tie-line points. Also supporting this alternative is the possibility of purchasing time for an individual use on a computer that is optimum for that use.

One of the strongest arguments against this alternative is based on cost: The above quoted figures lie considerably above the average cost of providing comparable services on well-utilized hardware controlled by nonprofit organizations, and nonprofit organizations could afford to offer the lower rates during both prime and nonprime shifts. Commercial vendors would probably even be unwilling to make large blanket commitments of prime-shift time

TABLE 2 Typical Rates for Block Time in Nonprime Shifts[a]

Computer Model	Operating System	Cost, $/h
UNIVAC 1108	Exec 8; $200{,}000_8$ 36-bit words core	500[b] - 250
IBM 360/65 or 370/155	IBM; $1{,}000{,}000_8$ bytes core	325 or less[c]
CDC 6600	SCOPE; $400{,}000_8$ 60-bit words core	600 or less[c]
CDC 7600	SCOPE; $200{,}000_8$ 60-bit words small core memory, $1{,}000{,}000_8$ words large core memory	2200[d]

[a] Using standard operating systems. These rates would ordinarily include free use of peripheral equipment including tape drives and a mass storage disk or drum, and would be on a "wall clock basis." Prime-shift rates would be higher by a factor of two or three.

[b] The first figure is for smaller amounts of time (up to 5 or 10 hours).

[c] Indicates that large time blocks might be negotiated at a somewhat lower rate than the figure shown.

[d] Limited availability.

unless they received guarantees of usage, and it might therefore be even more expensive to obtain truly convenient service. Moreover, this alternative would not be likely to be effective in contributing to software development or more widespread computer use.

4. CONTRIBUTIONS OF COMPUTER TIME BY THE COMPUTER INDUSTRY

This alternative has frequently been suggested, but no information is at hand to suggest a willingness on the part of any computer manufacturer to contribute significant amounts of appropriate computer resources under circumstances it cannot control in detail.

5. REGIONAL GENERAL-PURPOSE COMPUTER CENTERS

This alternative involves the supplementation of local computer centers by general-purpose centers operated for the institutions in a geographical area by a designated institution or by a management agreed upon by a group of institutions. Such centers currently exist; examples are Triangle University Computing Center (TUCC),

serving all institutions of higher learning in North Carolina; Northeastern Regional Computing Center (NERComp), with headquarters at Massachusetts Institute of Technology, serving a number of institutions throughout New England; the University of Georgia Computing Center, serving institutions throughout Georgia; the University of Colorado Computing Center; and many others.

When the institutions involved lie within a geographically compact region, a regional center can act much as a local computer center. A center serving a more extended region will be effective only if accessible via a suitable communications network. Most chemical computations can be carried out at a regional center by accessing it over ordinary telephone lines, if adequate provisions exist at the regional center for longer-term storage of large amounts of data. There appear to be no significant unsolved technical problems associated with the operation of regional centers. However, a considerable body of evidence indicates that the management of such centers necessitates the solution of a number of sensitive political and economic problems involving the interrelationship of the institutions and users.

On the positive side, this alternative may provide extended computing capacity at reasonable cost and, in addition, may provide a means for securing a satisfactory economic condition for the computers at the institution furnishing the regional services.

On the negative side are the political and organizational problems already mentioned, as well as the obvious fact that for the near future many chemists cannot expect to be served by suitable centers of this type. Even if a reasonably complete system of regional centers were to become established, the fact would still remain that these centers would not focus upon chemical computations. They might not be extremely responsive to chemist users, and they could not be expected to catalyze specialized software development or more widespread use of computers by chemists.

6. A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

Such a center, with an organization, a mission, and priorities as suggested elsewhere in this report, could provide, in common with some of the other alternatives, extended computing capacity to chemists at reasonable cost. However, it would provide several additional unique advantages, including those of coordinating and

accelerating software development, of providing a computational environment solely responsive to the needs of chemists, and of stimulating the growth of computation and computational methods for chemistry. As a pilot project serving computing needs in a specialized area of science, it could serve as an attractive vehicle for obtaining funding for increased support of these activities over that available under various other alternatives.

Negative factors possibly associated with such a center fall into two categories, denoted here as direct and indirect. The direct category includes the possible effects on local computer centers, which were discussed under the first alternative. Also in this category is the possible effect upon trends and practices among computational chemists. It has been suggested that the existence of widely available software and the pattern of activities by center personnel might work against diversity of effort and innovation in computational chemistry. The indirect negative factors have potentially significant effects whose importance is difficult to assess. They would be coupled to the success of the center and would relate to any changes thereby generated in patterns of funding for individual chemists, for chemistry generally, and for computational service.

Creation of such a center would call for consideration of a number of issues, some of which are discussed elsewhere in this report. Among them are the optimum arrangements for ownership and management of the center, the development of mechanisms for controlling and distributing access, and the identification of an optimum communications network to provide access to geographically remote users.

7. A NATIONAL CENTER FOR COMPUTATION OF BROADER SCOPE

This alternative could have the advantages and disadvantages already identified for regional general-purpose centers, with the additional difficulty that no one center of reasonable size would be capable of coming even close to meeting the demands that would immediately be placed upon it. This difficulty, in turn, means that it would be necessary to consider the allocation of priorities among an array of disciplines so broad that few, if any, people would have the expertise to do so competently. This dilemma would probably have to be resolved by a more or less arbitrary partitioning of the computer capacity among disciplines.

One unique advantage can be identified for a general-purpose national center: It might serve as a pilot project for the development of a set of regional centers.

8. ACCELERATED DEVELOPMENT OF COMPUTER PROGRAM EXCHANGES

This alternative contemplates limiting new activity to an expansion of efforts designed to facilitate the exchange of computer programs among chemist users. At present there are two significant activities of this type in chemistry: the Quantum Chemistry Program Exchange (QCPE), and the journal *Computer Physics Communications* (CPC). QCPE accepts programs and distributes them to prospective users; it undertakes no software development and performs a severely limited testing and documentation function. CPC referees submitted programs and their accompanying documentation and, for a price. distributes the programs published therein.

This alternative could, if pursued aggressively, provide better and more interchangeable software, and thereby encourage an expanded use of computers by chemists. However, it would not contribute to the expansion of computer facilities for meeting the needs of chemists.

9. NO ACTION

If no action is taken, a major opportunity will be missed or postponed. Possibly more money would then be devoted to experimental science and to the support of individual researchers, but even this conclusion is questionable since the federal science budget is hardly a zero-sum game.

VI Mission and Priorities for a National Center

The primary mission here proposed for a national center for computation in chemistry is to produce advances in chemistry and related sciences through more widespread, innovative, and intensive use of high-speed computational equipment. This mission is to be accomplished by making appropriate facilities available to a wide group of scientists, by providing and developing software to expedite and upgrade computer use, by encouraging and supporting research efforts that build new and more effective computational methods, and by carrying out an information and education program to bring the benefits created through the center to the widest possible scientific public.

SCIENTIFIC SCOPE

For a number of practical reasons, the scientific scope of the center should be limited to an area in which the practitioners have to some extent a common language and set of interests. It is proposed that the scope should be *chemistry* in a fairly broad sense. An exact delineation of the scope should not be made in advance, and should in any case be subject to continual adjustment through the evolution of patterns for the approval of requests to use the facilities of the center. At present it is contemplated that the scope should include electronic structure computations, whether done by chemists, molecular physicists, solid-state physicists, astrophysicists, molecular biologists, or others, and should include statistical mechanical and statistical thermodynamic studies of both real systems and models, including such diverse applications as investigations of reacting chemical systems and magnetic phase transitions in model lattices. Also included should be the full range of kinetics and molecular dynamics studies involving atoms, ions, molecules, surfaces, and solids, the interpretation of spectra of various kinds, and the reduction of data of most of the sorts currently obtained by chemists. Further applications included would be programs for classifying molecules, chemical properties, and chemical reactions, and the use of this information to search for such items as synthesis paths, molecules suitable for particular purposes, or new techniques for chemical analyses. In short, with a few exceptions noted below, the scientific domain of the center should include most of the computational activities it may be expected chemists will wish to pursue.

Certain areas, though clearly involving chemistry, probably should be excluded for the most part from the planned scope of the center. These are characterized by fairly well-defined computational requirements or by extensive data bases, *and* by the existence, at least in the planning stages, of special-purpose facilities as advanced as for the center under discussion here. One such area in which specialized development is already taking place is health-related computing. Clinical data, data on drugs, and other related biochemical and biomedical information, together with the software facilitating use of this class of data are being centralized with equipment set up with support from the National Institutes of Health. The Health Services Computing Facility of the University of California, Los Angeles, conducts an extensive program of research and resource sharing on computational methods in biomedical research. Another well-cultivated area is that of chemical documentation; it requires unusually large data-storage and data-searching capacity. Efforts are in progress at Chemical Abstracts Service and other organizations, public and private, to find optimum methods of dealing with the problems in this area.

There are other areas more or less related to chemistry that should probably not be automatically included. Various agencies, including AEC, NASA, and units within the Department of Defense, study complex problems involving large amounts of chemical data. Examples are modeling studies of the atmosphere, hydrodynamic studies in both normal and exotic temperature and pressure ranges, and studies of plasmas in various environments. These problems normally arise under circumstances that make practical answers of immediate importance, and therefore they are pursued by methods that can be extremely costly in consumption of computational resources. However, such studies would be appropriate to this center when they have promise of contributing to the development of better computational methods in areas clearly within the center's scientific scope.

ACTIVITIES

The provision of large-scale computational facilities to enable the carrying out of calculations that otherwise could not be attempted. These facilities need not necessarily be owned or directly managed by the center, at least in its initial stages; they could be made available on contract with an existing public or private organization having excess computing capacity of the desired scope.

The development of software to improve the effectiveness of computation by chemists, for use both at the center and at other locations. Substantial gains could be expected both from increased reliability of the software and from increased ease of use. It is contemplated that software development would include not only numerical methods for carrying out computations of types presently or in the future identified as useful, but also packages permitting effective algebraic and general analytical manipulations. This activity would not take large amounts of computer time, but would involve significant amounts of effort by the staff of the center. In this development the center should devote considerable attention to appropriate arrangements and conditions for optimum and flexible communications, such as would allow its programs and software to be addressed from a distance by the widest possible class of potential users in the field of chemistry. Included in these considerations should be such interactions as may seem appropriate with the ARPA and other special computer and information networks, as well as use of the national telephone system.

The provision of service to chemists seeking to broaden their use of appropriate computational techniques. This service should include consultation with respect to the center software and the provision of computer time for approved projects. The main thrust of this activity should be to enlarge the group of chemists using computation as a routine adjunct to the remainder of their efforts. This will in turn generate feedback indicating areas where computational research will have immediate practical value. This activity will be one of the main ways in which the center can provide something of value to a large population of chemists. It is not contemplated, however, that the service provided under this activity should expand to the point where it hampers the ability of the center to pursue the projects that it will be uniquely able to handle.

The development of new computational methods in chemistry. This activity will not take a disproportionate amount of the computer resources, but in the long term may be of extreme importance because of the changes it can make in the limits of computational feasibility. This activity can be pursued by providing research opportunities to the scientific staff of the center, a majority of which should be on short-term or visiting appointments. The center should be able to provide a stimulating and productive environment for computational research, and the short-term nature of most of its appointments will enable these benefits to be distributed over a substantial group of prospective research contributors. A significant part of this activity should be carried out by nonresident investigators at the center, and the center's communications facilities and access policies should make remote use as effective as possible.

Programs for education and for the dissemination of information about computing in chemistry. The center should present short courses in computational techniques for chemistry, publish descriptions of available software, take over and expand the activities of the QCPE, and issue materials designed to facilitate both on-site and remote use of its facilities.

Other activities. Among potential activities that have been discussed for the center is maintenance of data banks for various classes of chemical information—thermodynamic, kinetic, spectroscopic, and structural. Majority opinion of the study group is opposed to such data-banking as a primary function of the center. While data-banking on a large scale is, of course, an area in which modern computers are outstandingly useful, it involves a different type of usage from numerical computation. Furthermore, the personnel appropriate to data collection and information retrieval have quite different duties from those involved in the support of numerical computation. Thus, data-banking would involve a significant addition of staff devoted to what should surely be a secondary function. Storage and retrieval of data relevant to computations that have been carried out at the center, on the other hand, should be a useful and appropriate function.

PRIORITIES

It is inevitable that a center with scope and activities as described above will rapidly reach a point where demands upon its facilities exceed their capacity. It will then become necessary to make allocations of time to various activities, and to the various projects proposed by individual investigators. It is premature to decide such questions at this time, but it is clear that there must be a suitable mechanism for reaching such decisions. As indicated in Chapter VIII on organizational structure, there should be a broadly constituted scientific advisory board for making policy decisions as to scientific priorities as well as individual decisions on larger projects. Such an advisory board should be representative of chemists in general, rather than only of computational chemists, and it should not be dominated by scientists on the resident staff of the center. It should give due regard to the computing needs of the center staff and of outside users, and to the distribution of resources among activities and scientific areas. Further, it should function in a way designed to keep the bulk of the computer time assigned to projects of approved scientific merit. In its decision making, it should give particular attention to worthy projects that cannot be carried out elsewhere and to projects that give promise of developing new computational methods. Through its decisions, the advisory board will define the scope and function of the center.

VII Interaction with the Chemistry Community

To be of maximum value both to chemical science and as a national resource, a center for computation in chemistry must be so organized and managed as to be outward-looking and to interact to the maximum degree with other areas of the science and profession of chemistry. Not only is this outward interaction essential to the health and success of the center on a continuing basis, but a belief in the possibility of accomplishing this interaction is essential if chemists in general are to be induced to make effective use of its resources. Indeed, it is hardly possible to consider establishing such a national resource without a comparatively widespread conviction in the chemistry community that this center can provide an important service to a large segment of chemical science.

Unfortunately, it is also easy to imagine the converse possibility, i.e., that a center for chemical computations might be entirely inward-looking and devoted exclusively to the computational needs of a comparatively small circle of specialists. Indeed, a number of the members of the study group responsible for this report started with a considerable feeling of opposition to the concept of a national center because of a fear that it would be likely to take such an inward-looking course. Most of these individuals have found that their opinions have changed greatly as a result of this study, and they are now convinced that it is entirely feasible to ensure that a center will serve a wide class of users and to set up institutional safeguards against its becoming dominated by a narrow in-group.

These concerns are natural ones and occur in many forms to anyone who seriously contemplates the concept of a major computational center. Individuals may reasonably differ in their evaluation of the various arguments involved and in their opinion of the probable outcome of a given course of action. A number of specific concerns are reviewed here and comments of the study group are offered.[1]

[1] Further discussion of expressed concerns is presented in Chapter VIII.

1. *The center is likely to become the captive of a small in-group of theoretical chemists who will use it exclusively for their own purposes.*

COMMENT

This possibility exists, of course, but it is also possible, and indeed important, to set up appropriate institutional devices to guard against such an outcome. Other national facilities have developed various solutions to the problem. Possible safeguards include (a) broad representation on the ultimate governing body (the board of directors); (b) strong participation of a broadly representative scientific advisory board, with rotating membership, in the choice of projects to be undertaken by the center; (c) an open policy of inviting and accepting competitive proposals from scientists throughout the chemistry community; (d) limitations on the fraction of the center's activities and computing time that can be devoted to projects generated by the resident staff; (e) requirements that each member of the center's resident staff devote a specified portion of his work to serving the needs of external users; (f) a policy of rotation for a large fraction of the center's staff; (g) definitely established policies regarding the allocation of a certain fraction of the center's resources to the needs of the larger chemistry community other than to experts in large-scale computations.

2. *The availability of a large amount of computing time at the center will make possible extensive, wasteful, and unimaginative computations on a large scale using existing programs.*

COMMENT

Open competition for the center's resources, coupled with a discriminating selection procedure and high visibility, should provide considerable protection against such wasteful use. If these conditions are effective, the more pedestrian computations of large scale should be driven back to other sources of support than the center's. The aim of the center should be to set a standard of excellence and efficiency.

3. *The existence of the center is apt to lead to freezing of calculational techniques and their exploitation, at the expense of the development of new and more efficient methods.*

COMMENT

(a) The assignment of the center should include devoting a considerable fraction of its resources to the development of new computational techniques and programs. At the same time it should make existing and standard computations available to a variety of users who need the information they can produce. Provided that an allocation of time and resources between these two users is deliberately established and maintained, both are clearly compatible with each other.

(b) Experience thus far has shown that the development of computer capacity has indeed made possible increases in computational scale, but has also stimulated the continued development of new techniques, programs, and possibilities to extract more and more information through the use of a given amount of computer time and capacity. To some degree, improvement of any system is in conflict with exploitation of what one already has; this phenomenon is not confined to computations but exists also in experimental science, in engineering, and elsewhere. There will always be both exploiters and developers. Both are needed and can be accommodated.

4. *A center for computation will only enhance the unfortunate tendency for theoretical chemists to talk only to each other and increase their isolation from the rest of science.*

COMMENT

A major aim of the center clearly will be to make chemical computations and their results available to a much wider circle of chemists, including experimentalists as well as theoreticians, and to increase the effective communication between theoreticians and experimentalists. A visible, open, and accessible center where information and service can be readily obtained should catalyze much more active and effective interactions, and accelerate the process of breaking down barriers between theoreticians and experimentalists. Clearly, however, special efforts and safeguards are needed to

ensure the development of the interactions desired and to ensure that remedies are sought if progress tends in an undesirable direction.

5. *The establishment of a computation center will provide an undesirable competition with university computer centers which are already suffering major economic stresses.*

COMMENT

It should be the aim of the center to provide for needs that cannot now be met by the university computer centers. These include (a) access to unique and specialized computational programs for chemical purposes that are not available locally; (b) providing major economies in cost and efficiency for a limited and specialized class of computations; (c) providing the possibility of carrying out computations of a comparatively expensive nature (equivalent to hours of IBM 360/195 or CDC 7600 time) beyond the resources available to most chemists except those fortunate enough to have access to essentially free or heavily subsidized computer time; (d) providing for extensive and continuing software development, both to develop new capabilities and to make the programs available to a wide class of users. We have mentioned elsewhere in this report that university computing in the United States is now a large enterprise, costing over $500 million per year. The operations of the center will come to less than 1 percent of that amount.

6. *The proposed center will compete for funds otherwise available to support research in chemistry and endanger the support for other lines of chemical research.*

COMMENT

(a) In a time of restricted or dimishing research budgets, this possibility cannot be denied. However, theoretical and computational chemistry is currently one of the principal growing points of chemical science, and an investment in it should maintain or improve the competitive standing of chemistry vis-à-vis other sciences and other areas for governmental investment.

(b) The concept of devoting the center to fostering increased interaction between theoretical and other portions of chemistry,

and to exploiting the advantages of specialization in computational facilities, is an innovation that, if successful, may serve as a model to be followed in other sciences as well. As such, it could attract support that would not otherwise come to chemical science at all.

(c) By improving in a major way the cost efficiency of chemical computations, the center should increase the productivity of many research grants and contracts. This may reduce a heavy burden of computational costs and free some research funds to be used on other aspects of a research program.

7. *It might be wiser to allow increased computing funds to be included in ordinary research grants to individual investigators to be spent as they choose rather than to make an investment in a new computing center.*

COMMENT

Such a course of action lacks focus and does not lead to systematic program and software development that would enable broader use of advance computing techniques by ordinary chemists, nor would it achieve the cost economies attainable by a center that can concentrate on the specific needs of the chemistry community. In addition, this solution is unlikely to be practical in a time of restrictive budgets.

8. *The software development needed to make the research computing programs accessible to a comparatively large group of users can be done without establishing a major center. All that is needed is a convincing proposal to NSF by an individual researcher.*

COMMENT

(a) Some work of this kind is already being done. However, most research scientists have no motivation at present to undertake this kind of development for their own computer programs, since they understand perfectly well how to use them. Furthermore, the problem of maintaining and updating programs is not solved by this approach. The specialized scientific users in chemistry do not provide a broad market to attract commercial software development with remunerative sales in mind. Therefore, the job will not be done on a sufficient scale to provide the real advantages that could

be obtained unless it is stimulated by some agency with that particular function in mind.

(b) Considerable advantages to the unsophisticated user will arise only if the development is aimed particularly at him by formatting programs in a uniform style, providing a standard family of programs of varying levels of accuracy for different purposes, and supplying a standard catalogue of the programs available, with instructions for calling them out and providing the input data. Unless these programs are all available and accessible at a single center, with easy access through long-distance communications from any research laboratory, there will be no possibility of attaining anything near the comprehensive service to potential users that can be envisaged.

(c) Without such a center, the situation is bound to be afflicted by confusion, overlapping and competing services, inefficiency, and high cost if, indeed, any useful development of this type occurs at all.

9. *Much of the necessary facilities and software development are probably already part of the ARPA network activities and need not be duplicated.*

COMMENT

The ARPA network, an experimental nationwide computer network developed by the Advanced Research Projects Agency of the Department of Defense, links some 24 existing computer centers and is now being used by some theoretical chemists to provide access to a computer at a distant location. The most significant need that appears for other chemists is the development and provision of the necessary software to make an accessible and appropriate family of computing programs. The ARPA network was not established for the purpose of carrying out such development, but is available for the communication function if desired. A center devoted specifically to the computing needs of chemists could of course provide access through the ARPA network as well as through other communications networks, including standard telephone channels, and each route might provide advantages for different purposes.

10. *The differential equations to be solved in chemical electronic structure problems are similar to equations that are solved for many purposes in physics, aerodynamics, and related sciences. The suggestion has been made that a center should be focused on the broadest group of sciences including fluid mechanics, physics, engineering, involving mathematical problems of the same basic form.*

COMMENT

(a) While the techniques involved in solving the differential equations applicable to many chemical problems overlap those for solving the problems of a number of other sciences, similarities in computing techniques already are being taken advantage of in many ways. On the other hand, the idea of focusing the activities of a single center on the needs and conveniences of a particular group of potential users in chemistry would be undermined if the center had to provide services simultaneously for the activities of a number of sciences with problems formulated in different terms, even if they superficially use similar equations. The source of the computational difficulty is in fact quite different for structural chemistry, where extremely large numbers of variables may require simultaneous solution. A clear choice needs to be made between focusing on similarities in basic computing techniques or on similarities in practical user needs.

(b) Once the operation is running smoothly and has shown its potential for satisfying the needs of the chemistry community, it may serve as an example to be followed in other sciences. At that point it would be possible to examine the question of whether this center should be expanded for the broader purpose, or whether it is better to establish separate centers for each scientific discipline.

GENERAL REMARKS

The chemistry community is generally not well informed as to the role large-scale computation legitimately plays in theoretical research. Furthermore, the feasibility of putting important, well-tested chemical computation codes at the service of relatively unsophisticated users through interactive long-distance communications systems is not widely realized. Many members of the chemistry community are simply not aware of the possibilities for advancing their science by a concerted effort to make accessible the

best computational technology now practically available. It is hoped that this report has fairly addressed the real problems in bringing a national center for computation in chemistry into being, and that it will be received with an open mind by those who have not given much previous thought to the impact of computation on chemical research.

VIII Organizational Structure

SUPERSTRUCTURE

The superstructure of a center for chemical computations is obviously of great importance. Its role should be to set general policy and to choose the principal officers, as well as to ensure that the organization keeps on the course broadly desired by the users. It should not become too closely identified with the existing structure of the center and its staff, but should be responsive to expressions of need or concern from the chemical research community experienced in what can be done through computation.

A popular device to achieve independent ultimate control of a national or regional scientific center has been a consortium of universities. The members or their representatives in this ultimate governing body, which is commonly chartered as a nonprofit corporation, are then responsible for selecting a board of directors or trustees for the center. These in turn may select an executive director or manager of the center, and may find it desirable also to appoint a scientific advisory board, broadly representative of the user community. This type of organization has its attractive features, but conflicts can conceivably arise, particularly in later years, if institutional interests of individual member universities come to be perceived as conflicting to some degree with the scientific interests of the center and the purposes for which it was founded. Conceivably, at a center devoted to chemical computation, such conflicts might come to focus on competing needs of university computing centers. If this possibility is recognized at the outset, appropriate safeguards can be designed against its arising, such as a balance between user representatives and administrative representatives in the governing body of the consortium.[1]

Several existing national facilities are managed by university consortia. Argonne National Laboratory is operated by the Argonne Universities Association and the University of Chicago. Brookhaven National Laboratory was established and is operated by Associated Universities, Inc., a nonprofit corporation representing 9 universities, under contract with AEC. The National Accelerator

[1] See also Chapter VII, Comment 5.

Laboratory has been planned, constructed, and managed under contract with AEC by Universities Research Association, Inc., which includes 52 universities. The National Center for Atmospheric Research has been built and operated under contract with NSF by the University Corporation for Atmospheric Research, a consortium of 37 universities, including 2 in Canada. Kitt Peak National Observatory is managed under contract with NSF by the Association of Universities for Research in Astronomy, Inc., representing 12 member universities.

Although these consortia were founded originally in connection with establishing a specific facility, their charters are broad enough to encompass other activities of a generally similar character. Associated Universities, Inc., thus accepted responsibility under contract with NSF for the establishment and operation of the National Radio Astronomy Observatory, and also of the Very Large Array Radio Telescope now under construction. Universities Space Research Association, a consortium of 51 universities founded to operate the Lunar Science Institute at Houston under contract with the National Aeronautics and Space Administration, has recently (1972) undertaken to operate a national Institute for Computer Applications in Science and Engineering at NASA's Langley Research Center.

Some national facilities are managed by a single university. Lawrence Berkeley Laboratory and other AEC laboratories are managed by the University of California, Berkeley. The Francis Bitter National Magnet Laboratory, formerly supported at Massachusetts Institute of Technology by the Air Force Office of Scientific Research, is now operated by MIT under contract with NSF.

An experimental Computer Research Center for Economics and Management intended to develop packaged software, available to all researchers, for quantitative methods of potential value in the social sciences, is operated under contract with NSF by the National Bureau of Economic Research, Inc., a nonprofit organization founded in 1920 that conducts and supports research on topics of national importance in the field of economics. The center is supported jointly by the Office of Computing Activities and Division of Social Sciences of NSF.

This brief survey indicates that there is no dearth of tested kinds of organizational structures capable of contracting in a responsible way for funds and overall management of a national center such as the one envisioned in this report.

STRUCTURAL FEATURES OF A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

Detailed planning and the drafting of specific proposals for funding are necessarily left to a planning committee to be assigned those responsibilities by the agency contracting to build and operate the center. We shall do no more here than suggest guidelines for a structure that will fit the functions envisioned. The organization should be left sufficiently flexible to benefit from actual operating experience.

When fully operational, the center will provide a high-level computer and software system available to users for research in chemistry and related computational methods. The center will also provide a variety of services, including state-of-the-art programs for standard calculations, modular subunits that can be combined for broad-ranging research programs, high-level technical consultation, access to data bases accumulated from calculations previously carried out, and documented catalogues of computational methods. Basic supporting services will also be provided, including offices for visitors, generous quantities of key punches and terminal services, various output devices, remote access, and a good selective library.

The center is foreseen as developing in two phases. Phase I should begin as soon as is feasible and should include the following projects:

1. Developmental work towards making available in a convenient and effective form software for computational chemistry. This important function is not now carried on systematically and continuously by any other organization.
2. Assessment and evaluation of standard routines for projects in computational chemistry, with a view toward incorporating them where possible in the work under project 1.
3. Generation of a data base consisting of the programs developed under projects 1 and 2 and by QCPE, and the collected output of current computational programs in the form of wave functions, potential curves and surfaces, transition matrix elements, etc.
4. Developmental work toward participation in a communications network for the dissemination and utilization of the output of projects 1, 2, and 3 possibly based on the ordinary telephone system, the ARPA network, the educational and research

4. The foregoing requirements suggest a complex of computers: one large machine with high CPU rate and high input/output capacity, coupled to a machine designed to cope with high-volume interactive use and batch remote-terminal work. Such a combination might include a CDC 6400 coupled to a 7600, or an IBM 360/67 coupled to a 370/195. The key consideration is that the machines should be compatible in terms of hardware, so that algorithms found to be stable in the debug stage can have guaranteed stability in the production machine.

5. A necessary component of the hardware is a random-access store of extremely high capacity (10^{12} bit), which could contain the output of approximately 300 workers over a 10-year period. This data base should be available on a demand basis, with records kept as to utilization.

6. In general, the design of the first system should be conservative, with respect to both hardware and software. The operating system should stress convenience and offer high protection against failures and losses of information. It is likely to be 5 years before the next generation of computers is stabilized with respect to architecture and operating systems. Therefore, to provide a highly reliable and productive service, the architecture and performance capabilities of the CDC 7600 and IBM 370/195 generation of computers should be anticipated as available in practice for the next 5 years.

7. Price quotations should be solicited for specialized hardware, either added to the standard central processors or as add-on boxes. These would include fast Fourier-transform boxes, matrix boxes, scalar-product boxes, and perhaps an eigenvalue box, as well as the usual elementary mathematical functions such as logarithms, exponentials, and sines and cosines.

8. The first system installed should have provision for the development of software for support of algebraic work on the computer. In the longer term, additional hardware capabilities in this area should be explored.

No systems currently available are adequate to meet the projected long-term demands and to produce the returns that can reasonably be expected from applications of computational chemistry. Systems with vast inherent capabilities such as the Texas Instruments ASC, the CDC STAR, and the ILLIAC IV will surely be developed and become stabilized over the next 3 to 5 years. Much

network under consideration by EDUCOM, or a commercial computer network such as Tymenet.

5. Study of the effects of computer architecture on the algorithms of computational chemistry.

6. Study of the utility of computer-supported analytical and symbolic mathematics in computational chemistry.

7. Establishment of an inventory of computers, and their rate structures, available for computational chemical research.

Phase I need not engage the center in direct management of a large computer. The necessary computational facilities could be leased from or time-shared at an existing computer center. It will be noted that the activities identified for Phase I have substantial value independent of the development of a computer center with its own facilities. This consideration affords flexibility of future decision making and timing, and Phase I can be undertaken independent of a full commitment to Phase II.

Phase II should be developed over a 3- to 5-year period from the commencement of Phase I. Because of the rapid advance of computer technology, it is impractical at this time to specify in any detail the initial choice of hardware and software system. Among specific points to be kept in mind in designing and establishing the equipment are the following:

1. There should be provision for convenient and rapid debugging of programs that require large resources in the developmental stage.

2. There should be capability of coping with large production jobs, which may require both enormous data bases at a high level of activity and large amounts of central-processing-unit (CPU) service.

3. Remote access should be available, both individually interactive and serving a high volume of data. High-volume terminals would be established at major research laboratories for theoretical chemistry across the nation. The individual terminals should be available where research workers can demonstrate need. The speed of operation of the high-volume terminals is envisioned as being in the range of 40 - 100 kilobits per second, allowing for transmission of high-volume data bases. The interactive terminals can be serviced through an ordinary, voice-grade telephone network. The center should be designed for use as one node in the development of a national scientific or educational communications and computer network.

software development at basic and applied levels is needed to allow fully effective utilization of the potential capabilities of these systems. Parallel processors require a wholly new approach to algorithm design in order that the potential can be realized. Architectures of systems of perhaps an order of magnitude greater power than even these systems are currently in design stages at major manufacturers.

USERS

Use of the center should be available not only to investigators pursuing advanced research in theoretical chemistry, but also to a wide range of experimental chemists and scientists in fields allied to chemistry who are interested in using the techniques of theoretical chemistry for the study of their problems. These latter classes of users should not be expected to have a high degree of familiarity with computational chemistry, but should be able to use the center's facilities with the aid of consultation services to assist them in making effective use of it to solve their problems. For research investigators in quantum chemistry and other theoretical chemistry, the center should facilitate the development of new techniques and computational methods by removing unnecessary constraints on the efficient use of all types of existing techniques.

Space would be maintained for casual visitors who might stay for from a day to a month or so. It should be the policy of the center to accept longer-term visitors, if they contribute to the center's mission. Every effort should be made, however, to provide convenience for the short-term visitor in the way of computer access, local transportation, housing, travel grants, etc.

STAFF

The basic staff functions are to carry on research in computational chemistry and to support the users of the center's facilities. This latter function will include not only systematic research and program development across a fairly broad base of computational methods but also study of the interaction between computer architecture and the conduct of computational chemistry. It may include considerable work on the study and development of new

hardware concepts and systems for the kinds of computations of interest to the center.

The resident scientific staff should include several theoretical research chemists of varying backgrounds, one of whom will be the director of the center. It should include also several computer scientists generally familiar with computational problems in chemistry. The work of the resident staff should be supported by a research budget with provisions for program, postdoctoral, and technical support. The resident members should also have responsibility for the administration of scientific policy.

Concern has been expressed that members of the resident staff will gain an excessive advantage over other research investigators not so well equipped with computing facilities, and also that the scientific research carried on at the center could become isolated and lacking in innovation. Such concern has been specifically recognized in the development of other national facilities, and various countermeasures have been devised to allay it. In the first place, a policy of limited terms of appointment and of rotation of the resident staff should be instituted. Then, recognizing that a substantial amount of computer time must be spent on inhouse systems development, the operating time should be allocated to visitors, remote users, and resident staff alike by a user's committee on the basis of the scientific merit of the proposed research.

The location of the facility need not be physically detached from an institution where other chemical research is being carried on. Furthermore, a policy of encouraging short-term and longer-term visitors, including postdoctorals, will maintain a flux of ideas through the center from many institutions.

To allay another expressed concern, the size of the center when fully operational will not be so large as to attract away a significant fraction of the theoretical research chemists from other institutions. The bulk of theoretical research will continue to be conducted elsewhere. The national center will be available for attack on problems where the computational resources required are not otherwise accessible, and also to serve as an active focus for the collection, generation, and dissemination of computing programs and information of interest to chemists.

The operation of the computer should be under the supervision of an operating manager, who would have responsibility for the technical support staff, purchase of supplies, relations with vendors, and supervision of routine maintenance and development of

standard system software. The operating manager would be under the authority of the director. The technical support staff would include programmers, operators, and computer technicians. The operating manager would have general responsibility also for the secretarial and clerical support staff.

SCIENTIFIC ADVISORY BOARD

It is obvious that a major national resource must have a policy board to see that the center's performance is directed in the interests of its scientific mission, with due consideration of scientific, social, and technological relevance. The scientific advisory board should have built-in turnover of its membership, hold regular meetings, and be directly responsible to the trustees of the organization contracting to operate the center. Responsibility for the selection of the director should rest with the board, and he should sit with the board as an ex officio member.

An important function for the management of the center and its scientific advisory board is to establish rules and mechanisms for the allocation of resources, both of finances and of computer time. In regard to general resources in money and personnel, it will be necessary to set up a proper balance between three main functions: (1) the development of software and communications for the sake of the general chemical user; (2) the development of advanced programs for chemical computations of new types or large scale; and (3) research activities, including those of the research staff, visiting or resident.

In allocating computer resources for research, it will be important to establish an agreed distribution of the computer time to various classes of projects as measured in terms of their size or expense. For each class, procedures will have to be established for allocating priorities and choosing which proposed projects to support by assigned computing time. Naturally, the most stringent requirements of choice will apply to requests for large blocks of time, and the scientific advisory board itself may be involved in the review of such proposals. For smaller tasks, criteria and methods will have to be established to prevent diversion to the center of jobs that should be done at university computer centers or elsewhere. The board might designate a user's committee to

maintain continuing responsibility for allocating priorities and computing time to such tasks.

COSTS

The level of effort and funding for Phase I is flexible, but we recommend energetic initiation of it in view of its timeliness and its innovative character in bringing effectively the fruits and techniques of computational chemistry into the hands of a wide circle of scientists who are not experts in computing. Expenditures in the range of $0.5 - 1 million per year would be needed and justifiable.

The initial capital investment for the facilities envisaged under Phase II, in terms of today's dollars and market for equipment, would be between $10 million and $20 million. Annual operating costs of the center should then be expected to run between $3 million and $5 million per year.

IX Conclusions

1. Current developments in communications and software make resource sharing feasible to expand greatly the utility of many modern programs for theoretical chemical computation, and to adapt them for ready use at distant locations, by chemists relatively inexperienced in computation. The systematic development of software to make this possibility a reality will not occur without careful organization and financial support from the government.

2. Efficient and economical exploitation of these possibilities will best be achieved if a large family of tested computational programs, systematically updated, is accessible at a single dynamic center, equipped with reliable computing facilities and providing remote access by chemists throughout the nation.

3. Large-scale computational facilities have enabled the successful prediction of many chemical structures, properties, and reactions from theoretical first principles, and computational chemistry has become a major research front of chemical science, increasingly coupled with experimental chemistry. The possibility of access to a modern high-speed, high-capacity computer greatly extends the range of theoretical ideas capable of being tested and opens new areas of investigation central to chemistry. A national computing center dedicated to the advancement of chemistry and open to all qualified investigators can provide unique opportunities for the exploitation of these possibilities, and this chemical-research function is naturally complementary to the software-development function identified in item 2.

4. Of great importance to the success of such a national computational center is dynamic interaction with the rest of the chemical community. To achieve maximum value in stimulating and contributing to the growth and applications of chemical science, the center must be actively oriented outward toward the interest of chemists and their allied scientists in general and must attract both support and use by chemists who are not necessarily experts in computation.

5. A number of institutional safeguards are available, and in practice at other national research centers, to ensure responsiveness

to the broad scientific needs of the supporting research community, and to guard against the possibility that a narrow in-group will usurp control.

6. It is recognized that many influential members of the chemistry community are not completely informed as to the nature and scientific possibilities of a computational center with functions such as have evolved through the course of the discussions leading to this report. It is hoped that continued discussion will resolve areas of misunderstanding and will unite the chemical community behind a major effort to solve the real problems of meeting its computational needs for the future.

7. The algorithmic implementation of problem-solving methods is a powerful means of diffusing new techniques and fostering intercommunication. A national center for computation in chemistry would provide a forum for such diffusion and intercommunication of new techniques found useful in solving chemical problems.

X Recommendations

1. We recommend the establishment, as a national resource, of a computational center with facilities and personnel dedicated to advancing chemistry and related sciences through widespread, innovative, and intensive use of high-speed computational equipment. This mission is to be accomplished by making appropriate facilities available to a wide group of scientists, by providing and developing software to expedite and upgrade computer use, by encouraging and supporting research efforts to build new and more effective computational methods, and by carrying out an informational and educational program to bring the benefits created through the center to the widest possible scientific public.

2. We recommend that a committee responsible to an appropriate contracting organization (perhaps the National Academy of Sciences) be commissioned to bring this national resource into being.

Appendix Computer Use In Theoretical Chemistry 1973: Questionnaire and Data

COMPUTER USE IN THEORETICAL CHEMISTRY 1973: QUESTIONNAIRE

1. Primary research interest (please check):
   (a) Quantum chemistry _____
   (b) Simulation of bulk matter _____
   (c) Statistical mechanics _____
   (d) Other _____ (specify: _____)

2. Method of handling your major computation:
   (a) Institution's computer center _____
   (b) External computer center _____
   Location: _____
   Mode of access: _____
   (c) Other _____ (description:_____)

3. Model of computer used: _____
   Rate charged to you (or your institution): _____
   (check here if not known_____) Differentiating features of rate, if any known to you, for your use of computer (e.g., for CPU time, for I/O, for prime time, etc.):
   _____
   _____
   _____

4. Extent and type of computer use:
   (a) Number of hours (or $ cost) per year charged to your computer use: _____
   (b) Estimated percentage of use (if known to you) in:
   Central processing _____%
   Input/Output _____%
   Other _____% (please specify: _____)

5. Average number of associates, postdoctorals, and graduate students in your research group, besides yourself, whose computer use is included in answer of item 4: _____

6. Sources of support for your computation:
   (a) Support from own institution _____%
   (b) Research grant, NSF _____%
   (c) Research grant, other government agency _____%
   (d) Research grant, other outside agency _____%
   (e) Other
   (please specify _____)

7. Estimated number of hours per year you and your research group spend on programming: _____

8. Do you feel that limitations on the quality or quantity of the computing available to you are affecting your research?
   Yes _____ No _____

   Seriously _____
   Moderately _____
   Slightly _____

Name: _____
Institution: _____

Returns will be kept privileged and statistical summaries, only, will be released. Your reporting your name and institution is optional, but will be helpful to us in assaying the reliability of the sample of returns received as indicative of the total situation. Please duplicate this questionnaire and distribute it, where appropriate, to other theoretical chemists whom we may have missed.

Comments (if any) on computational needs for theoretical chemical research:

COMPUTER USE IN THEORETICAL CHEMISTRY 1973: DATA

Returns were received from 155 investigators out of 195 to whom the questionnaire was sent. The recipients were selected from the QCPE membership list, with some additions, identified as active theoretical chemists at academic, industrial, governmental, and private research institutions in the United States. Of the returns, three were not germane. The 152 valid returns summarized here represent 682 individual investigators, including the principals, their associates, postdoctorals, and graduate students. That the returns are reasonably representative is indicated in the fact that the total NSF support reported is consistent with estimated support budgeted for computation in proposals funded by NSF for research in theoretical chemistry (about $0.3 million in the Quantum Chemistry Program of the Chemistry Section and $0.1 million elsewhere in NSF).

APPENDIX TABLE 1 Primary Research Interest

Research Interest	No. Groups[a]	No. Groups with Other Shared Major Interest[a]
Quantum chemistry	106	16
Simulation of bulk matter[b]	7	11
Statistical mechanics	8	7
Other major interests[c]	13	4

[a]The distribution of the 152 groups with some overlap.
[b]Includes theoretical research on molecular collisions and trajectories.
[c]Includes solid-state chemistry and crystallography, microwave and optical spectra, magnetic resonance, normal coordinate analysis, experimental organic chemistry, laser-induced transient effects, polynucleotide structure, neuronal networks, pattern recognition applied to chemical problems.

APPENDIX TABLE 2 Method of Handling Major Computation

Location	No. Groups
At own institutions's center	134 + 5 with partial use elsewhere
At external computer center[a]	12 + 5 with partial use of home center
Other	1[b]

[a]Includes one user of a centralized facility of his university located on another campus.
[b]Group using tabulated numerical data.

APPENDIX TABLE 3 Computer Used at Home Institution Center[a]

Model		No. Respondent Groups
IBM	7094	1
	360/50, 65, 67, 70, 75	37
	370/155, 165, 175	21
	360/91,[b] 360/195,[c] 370/195[d]	13
CDC	3300, 3600, 3800	3
	6400, 6500, CYBER 72	21
	6600	11
	7600[e]	4
UNIVAC	1108, 1110	15
DEC PDP-10		7
Honeywell	635, 6000, 6035	3
Burroughs	6700, 85500	2
XDS Sigma 7		1
		139

[a]Includes 5 groups with partial use elsewhere.
[b]Columbia University; Princeton University; University of California, Los Angeles.
[c]IBM Research Laboratory.
[d]Argonne National Laboratory.
[e]Lawrence Berkeley Laboratory; Lawrence Livermore Laboratory; Los Alamos Scientific Laboratory.

APPENDIX TABLE 4 Computer Used at External Facility

Model		No. Respondent Groups
IBM	7044	1
	370/165	1
	360/91	4
CDC	6600	5
	7600/6600	1
		12

APPENDIX TABLE 5 Rates Charged[a]

Model of Computer		No. Respondent Groups: No charge for computer use	$75 - 200/h	$201 - 500/h	$501 - 1000/h	Rate not known to respondent
IBM	7044, 7094, 360/50, 65, 67, 70, 75	19	8	4	5	3
	370/155, 165, 175	11	2	5	1	1
	360/91, 360/195, 370/195	11	3	1	1	1
CDC	3300, 3600, 3800	1	0	2	0	0
	6400, 6500, 6600, CYBER 72	18	4	11	3	2
	7600	5	0	0	1	0
UNIVAC	1108, 1110	4	3	3	3	2
DEC PDP-10		3	2	0	0	2
Other		5	0	1	0	0
Totals		77	22	27	14	11

[a]The figures are uncertain because (a) approximately half of the institutions included make no real charge to research users on their own staffs; (b) many others provide partial subsidies; and (c) some use complex rate structures that take into consideration the nature of the computation and whether prime time is a requisite. The figures represent real charges to the user groups against research support external to that provided by the computing center, taking into consideration total or partial subsidies by the home institution. The great diversity in rates charged to the respondent users is evident, and is more attributable to differences in institutional policy than in the power of the computer.

APPENDIX TABLE 6 Amount and Sources of Support for Computation[a]

Range of Total Value of Computer Time Per Year 1972 - 1973 ($ in thousands)	No. Respondent Groups	Total Value of Computational Support ($ in thousands)	Part Supported By Own Institution ($ in thousands)	Part Supported From NSF Research Grants ($ in thousands)
0 - 5	44	121	67	36
6 - 10	25	209	162	34
11 - 20	31	516	316	118
21 - 50	20	679	487	54
51 - 150	21	1932	1440	104
160 - 600	7	2533[b]	2511[b]	0
	148	5990	4983	346

[a]The figures are uncertain for the reasons mentioned in Table 5. In the above table, the dollar value of the computer time spent by the respondent groups in one year is either that directly reported by the respondent group or that computed from the institution's rate structure, including any subsidy of the institution's own research computation as reflected in its internal rate structure (which may include a varying allowance from institution to institution for amortization and maintenance not otherwise supported.) Four of the 152 groups did not provide data and are not included.

[b]$1000/h has been arbitrarily assigned to cost of computer time for subsidized use of IBM 360/91, IBM 360/195, and CDC 7600, and $250/h to subsidized use of CDC 6600. For respondent groups in this bracket, computer time was either entirely or at least 90% subsidized by their institutions. Only one of these groups is located at a university.

If one can assume that the respondents are a representative sample of approximately 200 active research groups in theoretical chemistry in the United States, the estimated dollar value of computation in this field is about $8.0 million per year, of which 83% is presently subsidized by the home institutions. Direct support of this activity by NSF is about $0.4 million per year. An unknown amount of indirect support is provided in how the various computers and computer centers are themselves funded.

APPENDIX TABLE 7 Estimated Time Spent on Program Development and Testing[a]

Range of Estimated Hours per Year	No. Respondent Groups	% Respondent Groups
0 - 50	19	14
51 - 100	14	10
101 - 500	43	32
600 - 2000	37	27
2100 - 5000	17	13
> 5000	5	4
	135[b]	100

[a]In this tabulation, one man-year was counted as 2000 hours. The sizes of the research groups ranged from 1 to 15 individuals, with a mean of 4.5. The total program development and testing effort corresponds to $2 - 4 million per year, of the same order of magnitude as the estimated cost of computing time.

[b]Other respondents were unable to answer this question.

APPENDIX TABLE 8 Response to Question: Do You Feel that Limitations on the Quality or Quantity of the Computing Available to you are Affecting your Research?

Response	No.	%
Yes	80	53
Seriously affecting research	27	18
Moderately affecting research	33	22
Slightly affecting research	20	13
No[a]	72	47

[a]Several of these respondents stated that they foresaw increasing dissatisfaction in the near future.

Participants

Members of the Planning Committee

*Kenneth B. Wiberg, Professor of Chemistry, Yale University, *Chairman*

*Lawrence S. Bartell, Professor of Chemistry, University of Michigan

*Jacob Bigeleisen, Chairman, Department of Chemistry, University of Rochester

*Robert B. K. Dewar, Associate Professor of Computer Science, Information Science Center, Illinois Institute of Technology

*Frank E. Harris, Professor of Physics and Chemistry, University of Utah

*F. A. Matsen, Professor of Chemistry, University of Texas at Austin

*Harrison Shull, Vice Chancellor, Research and Development, Indiana University

*Lawrence C. Synder, Member of the Technical Staff, Bell Laboratories

Panel Members

Panel I Scientific Objectives

1. Electronic States of Atoms and Molecules

Chairman: F. A. Matsen

Charles F. Bender, Chemist, Lawrence Livermore Laboratory

Richard F. Fenske, Professor of Chemistry, University of Wisconsin

Frank E. Harris

Hugh P. Kelly, Professor of Physics, University of Virginia

*J. A. Pople, Professor of Chemistry, Carnegie-Mellon University

A. C. Wahl, Senior Scientist and Group Leader, Chemistry Division, Argonne National Laboratory

*Present at one or both general meetings of the study group.

2. Inorganic Chemistry, Organic Chemistry, and Spectroscopy

Chairman: Kenneth B. Wiberg

R. E. Christoffersen, Professor of Chemistry, University of Kansas

F. A. Cotton, Professor of Chemistry, Texas A & M University

Donald S. McClure, Professor of Chemistry, Princeton University

Andrew Streitwieser, Jr., Professor of Chemistry, University of California, Berkeley

3. Solid-state and Surface Chemistry

Chairman: Frank E. Harris

*Frank Herman, San Jose Research Laboratory, IBM Corporation

Keith H. Johnson, Professor of Metallurgy and Materials Science, Massachusetts Institute of Technology

4. Statistical Mechanics and Macromolecules

Chairman: Stuart A. Rice, Professor of Chemistry, University of Chicago

H. A. Scheraga, Professor of Chemistry, Cornell University

*Frank H. Stillinger, Member of the Research Department, Bell Laboratories

5. Structural Studies

Chairman: Lawrence C. Snyder

Lawrence S. Bartell

William D. Gwinn, Professor of Chemistry, University of California, Berkeley

*Walter C. Hamilton, Senior Chemist, Brookhaven National Laboratory (deceased)

6. Molecular Dynamics and Scattering

**Chairman*: John C. Light, Professor of Chemistry, University of Chicago

R. Stephen Berry, Professor of Chemistry, University of Chicago

Don L. Bunker, Professor of Chemistry, University of California, Irvine

I. Shavitt, Staff Member, Theoretical Chemistry, Battelle Memorial Institute

*Present at one or both general meetings of the study group.

7. Atmospheric Science

*Chairman: M. Krauss, Physical Chemistry Division, National Bureau of Standards

Alexander Dalgarno, Chairman, Department of Astronomy, Harvard University
James N. Bardsley, Associate Professor of Physics, University of Pittsburgh
*H. Harvey Michels, Senior Theoretical Physicist, United Aircraft Research Laboratories

PANEL II CHARACTERIZATION OF COMPUTATIONAL FACILITIES

Chairman: Robert B. K. Dewar

*James C. Browne, Professor of Physics and Computer Science, University of Texas at Austin
*Anthony C. Hearn, Professor of Physics, University of Utah

PANEL III ADMINISTRATIVE AND FINANCIAL STRUCTURE AND POLICY

Chairman: Harrison Shull

W. O. Baker, President, Bell Laboratories
*Sidney Fernbach, Head, Computation Division, Lawrence Livermore Laboratory

OTHER PARTICIPANTS IN THE STUDY

Berni J. Alder, Theoretical Physicist, Lawrence Livermore Laboratory
*William A. Goddard, Professor of Chemistry, California Institute of Technology
*Carl Moser, C.N.R.S., Centre Européen de Calcul Atomique et Moléculaire
*Robert G. Parr, Professor of Chemistry, The Johns Hopkins University
*John S. Perry, Program Manager, Advanced Research Projects Agency
*Norman F. Ramsey, President, Universities Research Association, Inc.

*Present at one or both general meetings of the study group.

*Richard J. Royston, Director, Applied Mathematics Division, Argonne National Laboratory
*Ronald Rutledge, Director, Computer Center, Carnegie-Mellon University
*Felix T. Smith, Manager, Molecular Physics Group, Stanford Research Institute

OBSERVERS FROM FEDERAL AGENCIES

*Murray Aborn, Special Projects Program, Division of Social Sciences, NSF
*O. William Adams, Chemistry Section, NSF
*D. Don Aufenkamp, Office of Computing Activities, NSF
*William H. Cramer, Chemistry Section, NSF
*Robert P. Epple, Nuclear, Structural and Inorganic Chemistry Branch, AEC
*Peter G. Lykos, Office of Computing Activities, NSF, on leave from Illinois Institute of Technology
*Michael A. Oxman, Special Research Resources Branch, NIH
*John R. Pasta, Head, Office of Computing Activities, NSF
*M. Kent Wilson, Head, Chemistry Section, NSF

*STAFF DIRECTOR: Martin A. Paul

*Present at one or both general meetings of the study group.

The Proposed National Resource for Computation in Chemistry: *A User-Oriented Facility*

National Research Council
Assembly of Mathematical and Physical Sciences
Office of Chemistry and Chemical Technology

NATIONAL ACADEMY OF SCIENCES
Washington, D.C.
1975

Contents:

NOTICE: The Project which is the subject of this report was approved by the Governing Board of the National Research Council, acting in behalf of the National Academy of Sciences. Such approval reflects the Board's judgment that the project is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the committee selected to undertake this project and prepare this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. Responsibility for the detailed aspects of this report rests with that committee.

Each report issuing from a study committee of the National Research Council is reviewed by an independent group of qualified individuals according to procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved, by the President of the Academy, upon satisfactory completion of the review process.

Support for this project was provided by the National Science Foundation under Contract No. NSF-C310, Task Order 303, and by the U.S. Energy Research and Development Administration under Contract No. E(11-1)-2518.

Available from
Office of Chemistry and Chemical Technology
National Research Council
2101 Constitution Avenue
Washington, D.C. 20418

Dr. Philip Handler
President
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

Dear Dr. Handler:

I am pleased to forward the attached report, *The Proposed National Resource for Computation in Chemistry: A User-Oriented Facility* prepared by the National Research Council committee appointed by you in response to a joint request from the U.S. Energy Research and Development Administration and the National Science Foundation. This report completes a series of comprehensive studies by the National Research Council on the needs for computational resources for research in chemistry. The initial impetus for these studies came from the needs of academic chemists in one research area. The National Research Council's studies have shown that the needs go beyond the initial focus -- quantum chemistry research in the university sector. The most recent report *A Study of a National Center for Computation in Chemistry* addressed itself to the scientific needs in all of chemistry. The present report spells out how the proposed Resource would meet these needs for research in the university, in industry, and in the Federal laboratories. It proposes an organization, facilities, site selection criteria, funding, and an administrative structure to serve needs which are currently unfulfilled, and a mechanism for bringing the Resource into being without the acquisition of new major hardware.

The National Research Council has now provided appropriate agencies of the Federal Government the necessary justification and guidelines for the implementation of this user-oriented national Resource. It will require new funds to implement the proposed Resource; these can be justified by the new scientific programs which will become possible through the proposed Resource. In addition, the proposed Resource will offer services beyond those currently available by our present practices. The projected budgets are small compared with the present total national expenditures for research in academic chemistry alone. Our surveys show a strong interest from academic and industrial scientists, and from the Federal laboratories. There is wide interest in participation in the proposed Resource and all of these diverse interests recognize the benefits such a Resource could bring

to their individual needs. Implementation of our recommendations should, therefore, not impact adversely on meritorious research programs in chemistry, particularly those in the academic sector from which this proposal originated. The present report should enable planning to proceed during FY 76 and the initiation in FY 77 of the Phase I effort endorsed by the Committee on Science and Public Policy in December, 1973.

As part of the work of the Committee, we have discussed the substance of this report with a wide cross section of interested and knowledgeable persons in chemical research, computer science, communications, and scientific administration. The organization and structure of a national Resource for computation proposed in this report has met with wide acceptance. It could readily become the basis for planning similar facilities in other areas of the physical, engineering, and life sciences.

The work of a National Research Council committee is usually completed with the acceptance of its report. In the present case, this Committee has been requested and has agreed to remain available for possible future service to the sponsoring agencies.

The task assigned to this Committee has been accomplished in a short time as a result of the dedication of all of its members. It has been a pleasure to work with this Committee and bring the proposed Resource close to realization.

Sincerely,
Jacob Bigeleisen
Chairman
Planning Committee for a
National Resource for
June 8, 1975 Computation in Chemistry

Dr. Edward C. Creutz
Assistant Director for Research
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Dr. John M. Teem
Assistant Administrator for Solar,
Geothermal, and Advanced Energy
Systems
U.S. Energy Research and Development
Administration
Washington, D.C. 20545

Gentlemen:

I am pleased to forward the report of the Planning Committee for a National Resource for Computation in Chemistry, *The Proposed National Resource for Computation in Chemistry: A User-Oriented Facility*.

This report is the latest in a series of comprehensive studies by the National Research Council on the needs for computational resources for research in chemistry. Starting from an initial focus on research needs of university chemists in the area of quantum chemistry, these studies have shown that such a resource would meet scientific needs in all of chemistry. In its report, the Planning Committee has made thoughtful recommendations regarding organization, facilities, site selection criteria, funding, and administrative structures designed to serve the user community.

Hopefully, the guidelines provided by the Committee will be of assistance to the appropriate Federal agencies in the implementation of the Phase I effort recommended for such a resource in the March 1974 National Research Council report, *A Study of a National Center for Computation in Chemistry* with endorsement by the Academy's Committee on Science and Public Policy.

I understand that your agencies have requested that the Planning Committee not be disbanded upon submission of this report, in anticipation of the likelihood of requests for further services of the Committee at a later date. The Committee will be pleased to accede to these requests.

Sincerely yours,
Philip Handler
President

June 11, 1975

1 SUMMARY

In response to recommendations on page 2 of the report* of the National Research Council study chaired by Kenneth B. Wiberg calling for "the establishment, as a national resource, of a computational center with facilities and personnel dedicated to advancing chemistry and related sciences," the Atomic Energy Commission (AEC) and the National Science Foundation requested President Handler of the National Academy of Sciences to initiate a study of certain questions involved in the detailed planning of the organizational structure, management, and scientific policy of such a proposed resource. The questions (see page 5) were outlined in a letter to Dr. Handler from Dr. John M. Teem, Director of the Division of Physical Research of the AEC (now the Energy Research and Development Administration). The Committee authoring this report was appointed to respond to that request.

The Planning Committee recommends that a National Resource for Computation in Chemistry (NRCC) be organized as a user-oriented facility, with hardware and personnel dedicated toward serving the needs of the broadest chemical community. The prime function of the Resource would be to give impetus to the solution of important chemical problems by providing enhanced computational opportunities and capabilities not presently available to individual research investigators throughout the nation.

The Committee envisages a centralized Resource serving the community by making available the potential to be derived from systematic, collaborative attention to software development, documentation, and improvements in

*A Study of a National Center for Computation in Chemistry (National Academy of Sciences, Washington, D.C., March 1974).

computational procedures as applied to chemistry, as well as by making available the benefits of increased computational hardware.

The Committee is convinced that the enhanced computational efficiency of such a National Resource will be more cost-effective than a further increase in the application of computers as presently practiced and will contribute to the solution of important current national problems. Illustrations of such problems are detailed on pages 10-13.

The Committee believes that an NRCC can uniquely contribute to the enhancement of chemical research in several ways. The availability of a human resource of computer scientists and chemists working together will permit the documentation, testing, and improvement of existing programs; the development of more efficient algorithms; the generalization of computer programs for recurring chemical problems; the development of new computational methods; the design of specialized hardware, software, and languages particularly suited to use by the chemical community; and the establishment and updating of important data bases. A dedicated staff would bring about improvements in means of remotely accessing computers, including the possibility of efficiently utilizing a nationwide computer network. The NRCC would stimulate interaction among groups and individuals sharing common computational interests via workshops to establish priorities for important and yet unsolved computational problems and through visitor programs that will bring outstanding scientists together at the NRCC for periods ranging from weeks to months, so that they may interact with each other and with the in-house staff.

In order that the NRCC make a unique contribution to chemical computation and software development, it must have access to the most advanced and powerful computational equipment. A computer having the speed and memory of a CDC 7600, an IBM 370/195, or their equivalent, is the minimum with which the NRCC could fulfill its function. In order to permit optimal participation by and benefits to chemists throughout the nation, communications facilities must be available for convenient and equitable remote access. A good scientific library readily available to staff, visitors, and users is an essential part of the site requirement. The Committee recommends that the NRCC be established at one of the federally supported institutions where the needed major computational hardware is currently available for utilization by the NRCC during Phase I (first three years) of its operation.

This recommendation makes unnecessary a major capital investment at this time and in addition would immediately provide the NRCC with a rich scientific environment.

The operation of the NRCC would present some unique organizational problems. It would be a relatively small organization committed to serving its external user community and would require identity and independence.

If the NRCC is to function effectively when folded into a large, existing organization, the management structure must be such as to nurture the growth and utilization of the Resource without detriment to the functioning of the host institution. Phase I should be capable of fostering the growth of the NRCC to Phase II, when the Resource could independently justify acquisition of its own major hardware or when it would be desirable to affiliate with more than one federally supported institution. Operation under the management of two institutions would clearly be impossible.

A careful examination of the above considerations has led the Committee to conclude that the NRCC must function in one of two modes: either as a division within a currently existing institution or as an independent nonprofit organization. In either case it must have a large degree of autonomy in establishing its scientific goals, priorities, and budget, and its policy should be set by a cross section of individuals from the disciplines of chemistry, computational science, and research management. Cogent arguments are presented in Chapter 5 that the needs of the NRCC can best be met in the latter operational mode.

In either mode, the NRCC will require a Policy Board to establish scientific and management policies to be carried out by the Director. The latter would be the executive officer of the organization and would serve at the pleasure of the Policy Board. A Program Committee is necessary to review the scientific content of all major activities proposed to the NRCC and to recommend relative priorities among competing programs and proposals within guidelines established by the Policy Board. A User Committee, representing outside users of the Resource, as well as the in-house scientific staff, would provide feedback to the Resource to match NRCC operations to user needs. Details of this recommended administrative organization are described in Chapter 9.

Recommended personnel levels and budget projections for the first three years of operation are presented in Chapter 7, Tables 2 and 3, respectively.

2
Background and Charge to the Committee

In March 1974, the National Academy of Sciences published a report, A Study of a National Center for Computation in Chemistry, prepared with support from the National Science Foundation by a study group chaired by Kenneth B. Wiberg. The study was the culmination of a series of earlier discussions and conferences on problems relating to the impact of the electronic digital computer on the conduct of chemical research. While the initial focus was on requirements in theoretical chemistry, the Wiberg study showed that the needs extended to many areas of chemistry, including many aspects of experimental research. It was addressed to the feasibility and desirability of a national computing center that would include personnel and facilities dedicated to solving important chemical problems for which the necessary computational technology existed, or could be developed, but was not accessible to the majority of well-qualified investigators.

The two-year study was widely publicized in the chemical community, and approximately 50 individuals from various areas of chemistry, computer science, and research administration directly participated in it. It led to the following recommendations:

"1. We recommend the establishment, as a national resource, of a computational center with facilities and personnel dedicated to advancing chemistry and related sciences through widespread, innovative, and intensive use of high-speed computational equipment. This mission is to be accomplished by making appropriate facilities available to a wide group of scientists, by providing and developing software to expedite and upgrade computer use, by encouraging and supporting research efforts to build new and more effective computational methods, and by

carrying out an informational and educational program to bring the benefits created through the center to the widest possible scientific public.

2. We recommend that a committee responsible to an appropriate contracting organization (perhaps the National Academy of Sciences) be commissioned to bring this national resource into being."

This study was followed up by public discussion of the questions at issue in a symposium organized by the new American Chemical Society Division of Computers in Chemistry at the September 1974 national meeting of the ACS in Atlantic City. The symposium, in which members of the NRC study group participated, was well attended and evoked much favorable comment.

Meanwhile, in the spring of 1974, the Los Alamos Scientific Laboratory had drafted a proposal to implement the recommendations of the Wiberg report by utilizing the Laboratory's existing computational facilities. In June 1974, Argonne National Laboratory (ANL) and Argonne Universities Association (AUA) jointly sponsored a two-day workshop for interested individuals and representatives from interested organizations to develop recommendations that would provide a basis for writing proposals on behalf of ANL/AUA to establish such a national resource, hereafter designated as a National Resource for Computation in Chemistry (NRCC). Later, additional national laboratories supported by the Atomic Energy Commission (now the Energy Research and Development Administration) indicated a strong interest in making available their expertise in computing to this Resource and drafted specific proposals concerning their roles in such an organization.

Consequently, on July 1, 1974, Dr. John M. Teem, Director of the AEC's Division of Physical Research (now Assistant Administrator of ERDA), addressed a letter to Dr. Philip Handler, President of the National Academy of Sciences, in which he expressed the clear conviction of the AEC that the NRCC would serve important national goals. He requested that the Academy follow up the implications of the above recommendations of the Wiberg report. In particular, he requested detailed recommendations on the following questions:

1. The scientific policy and management of a proposed Resource.

2. The appropriate structure under whose auspices the Resource will operate. The organization should be capable of contracting in a responsible way for funds and overall management of the Resource, but it should be left sufficiently flexible to benefit from actual operating experience.

3. The composition, size, and responsibilities of a policy board of prominent scientists drawn from a wide spectrum of interests in chemistry and computing, which will be responsible for ensuring that the Center's performance is directed in the interests of its scientific mission, with due consideration of scientific, social, and technological relevance to national needs.

4. The relationships of a policy board to the operating structure of the Resource, to the Atomic Energy Commission and possibly other Federal funding agencies, and to the user groups.

5. Desirable priorities, growth rate, and levels of funding and operation for the first several years.

6. The issue of charge structure as it relates to academic, government, and industrial users, to the health of regional and university computing centers, and to the general fiscal policies of the Atomic Energy Commission.

7. Facilities, access, and site requirements for such a Resource.

The National Science Foundation through its Assistant Director for Scientific Research, Dr. Edward C. Creutz, has also expressed interest in such further detailed recommendations concerning the NRCC. The present study

is in response to these expressions of interest on the part of ERDA and NSF and is jointly supported by these two agencies.

The present Planning Committee was appointed by the Chairman of the National Research Council upon recommendation from the Executive Committee of the Assembly of Mathematical and Physical Sciences. Its composition was selected to be broadly representative of all areas of chemistry with an interest in computation.

In carrying out the task assigned to this Committee, the members as a whole, in smaller groups, and individually have sought out the opinions of a wide range of users and organizations interested in chemical computation. An extensive questionnaire was prepared requesting input information on computing hardware, communications, communications environment, software, system performance, rate structures, availability of hardware and software to an NRCC, and administrative organization. This questionnaire (Appendix A) was distributed to five ERDA laboratories and two university-based centers. After compilation of the returns from six of these institutions, site visits were made to those institutions that had responded in time to meet the Committee's operating schedule. Each site visit was by no fewer than three Committee members. A questionnaire concerning user needs was distributed to approximately 500 college and university chemistry departments, 340 industrial research laboratories, and 600 individual research scientists who make extensive use of large computers.

Much time and energy have been devoted to the planning of an NRCC. During the planning period, the Committee learned of the existence of the Atlas Computer Laboratory, which is operated by the Scientific Research Council of Great Britain. The Atlas Laboratory has played the same role in Britain that an NRCC would play here and seems to have contributed much to the growth and quality of scientific research in several important areas of chemistry. The Director of Atlas met with members of our Committee to alert them to problems encountered by Atlas in its formative years. In addition, one of the Committee members made a site visit to the Atlas Laboratory.

The present report incorporates recommendations based in part on the responses to these questionnaires, in part on information gained from the site visits, and in part on information regarding the experiences of the Atlas Computer Laboratory. (The user-need questionnaire and an analysis of the responses received are attached as Appendix B.)

3
Need and Possibilities for the NRCC

The report of the Wiberg Committee addressed the expanded role of computational methods facilitated by the high-speed electronic digital computer, for solving important experimental and theoretical problems. A brief outline of the needs and possibilities of the NRCC may be presented by quoting from the summary of that report:

"During the past decade, the impact of the electronic digital computer has revolutionized the conduct of research in chemistry, a pervasive, basic science with applications of great importance to human welfare. The computing needs of chemists now extend over an exceedingly broad spectrum, encompassing among major categories the need for automated control of instrumentation and the associated analysis of experimental data, the need for mechanized storage and retrieval of information, and the need for direct numerical solution of complex systems of differential equations such as are encountered in theoretical approaches to chemical problems by the methods of quantum chemistry and statistical mechanics. Meeting this latter category of need is the concern that has led to the study described in this report.

"Pioneers in the use of computers to solve theoretical chemical problems usually had access to computers that had been procured by their institutions for other, more general purposes, and that were in many cases subsidized wholly or in part. Because of changes in the support policy for the university computing centers, and because of increasing demand on the part of other users, such subsidized time is becoming increasingly less available, and chemical investigators dependent on large-scale computation are finding themselves priced out of the market. Support of chemical computation is at best unevenly available, and access to the more powerful computers is

restricted by conditions extraneous to scientific merit. Few universities can afford the largest computers now in commercial production and functioning in mission-oriented research institutions.

"At the same time, chemical computation has come of age. Theoretical methods of established reliability are on hand, and in prospect, for solving important chemical problems inaccessible to or too costly for experimental approach. Of perhaps even greater significance, a close coupling of theoretical and experimental techniques affords for many problems a more powerful and more reliable mode of attack than either experiment or theory alone. At present these possibilities for exploiting theoretical and computational advances are severely underutilized, both because the necessary computational facilities are not sufficiently available and because the necessary coordinated effort has not yet been made to provide for practical and easy access to tested computational programs and equipment by the wide group of potential users in the chemical research community. Recent developments in communication techniques and software make feasible a major increase in the utilization and exploitation of computational resources by the scientific community.

"These facts suggest that the time is opportune for the creation, as a national resource, of a national center that will take full advantage of progress in theoretical methods applicable to chemistry, provide appropriate facilities for theoretical and associated computational research, and make these methods and the knowledge derived therefrom accessible to all who have scientifically cogent uses for them."

The Argonne workshop identified the potential users of the Resource in two major classes with distinct, complementary interests. One class, called "clients," consists of users who require large amounts of computational resources in the solution of important new problems. The other class, called "customers," includes a wide range of chemists whose primary needs are for well-documented operational programs that can be immediately applied to the solution of specific chemical systems. These programs should be transferable to home institutional facilities, as presently provided in a limited scope by the Quantum Chemistry Program Exchange (QCPE), but preferably should be available for direct remote access from a home terminal.

To accomplish the above objectives it is insufficient merely to assemble a library of computer programs.

Extensive software development is necessary to be able to extend the scope of problems under current investigation with moderate computational facilities. Extensive documentation is necessary to make computer programs readily transferable.

This effort will make computation in chemistry more efficient than it is at present; it will open new frontiers; it will bring to the individual investigator computational resources he cannot afford to develop on his own. A centralized effort will not replace much work currently in progress. Through the development of a centralized resource, not only will increased computational hardware become available to many users, but equally important, there will exist the potential to be derived from the availability of a human resource dedicated to software improvement, documentation, and collaboration in the advancement of the art and science of computing as applied to chemistry. This collaborative approach will clearly be more cost-effective than a further increase in the application of computers as presently practiced.

The establishment of a central resource will contribute in time to the solution of important current national problems. One must always realize the lead time between research, development, and application. For a number of the problems illustrated below, one can anticipate a time span of the order of a decade between inception of research at the NRCC and applications.

Studies in chemical kinetics. Applications of reaction-rate data to complex, interacting chemical systems are central to the study of fossil fuel combustion, coal gasification, and the processing of oil shale. Furthermore, they are the cornerstone to the understanding and control of the formation and dispersion of atmospheric pollutants. For many of the specific problems in these areas, there does not even exist a direct method of solution, but one must devise realistic models for which numerical solution techniques can be applied and the results compared with experiment. Besides the need of direct computation in the pursuit of solutions to the problems, one may also need statistical information on the effects of uncertainties in the input variables, such as component reaction-rate coefficients and diffusion coefficients. The complexity of the chemical systems involved in such problems calls for a high order of skill in designing appropriate models and resourceful computational support in testing them.

2. *Studies of macromolecular structure and conformation.* Determinations of macromolecular structure, such as of proteins, from crystallographic x-ray or neutron diffraction data would be greatly facilitated if it were possible to employ the classical least-squares method of structure refinement so successfully used for smaller molecules. While such calculation is a tour de force today, it is well within the capability of a modern vector computer. The NRCC, by making this greatly improved tool available to the protein crystallographer, could help bring on a dynamic period of great potentiality in an important area of crystallographic research. Furthermore, much information is now available about intramolecular and intermolecular forces in polypeptides so that it is now possible to try to understand how a polypeptide chain folds up into the conformation of a native protein. With the available computer facility of the NRCC, it would become feasible to compute the most stable conformations of model peptides and, also, of proteins. In addition, the complexes formed by interactions between proteins and other molecules of biological interest can be treated by similar computational procedures, thereby contributing importantly to our understanding of meaningful biological phenomena.

3. *Studies of the structure of liquids.* Molecular dynamic and Monte Carlo computational techniques have already provided insight into hitherto unknown phenomena in the condensed phases of matter. Much remains to be done in this area; it requires fast computers of large capacity not available at most university computing centers. The experimental and the computational studies go hand in hand. Since most chemical and most biochemical processes occur in the condensed phase and since our present understanding of how these processes take place is almost nonexistent, new insights into the theory of liquid structure can be exceedingly fruitful for the development of chemistry. A promising beginning has been made on an adequate dynamic model for water based on statistical mechanical computations, but it cannot be further developed within present constraints on the availability of computer time. Computer-simulated experiments, moreover, are capable of providing valuable insight into the mechanisms of nucleation processes and chemical dynamics in condensed phases.

4. *Other theoretical studies of ground and excited states of molecules.* Such studies relate not only to the further development of tunable lasers, and of lasers operating in regions of the electromagnetic spectrum not

presently covered, but also to the highly important problems of energy transfer between molecules and the productive utilization of excited-state chemistry.

5. *Theoretical studies of photochemical reactions.* Such studies are of importance for input information in modeling transport processes and chemical reactions that are fundamental in understanding the role of free radicals and common atmospheric pollutants (e.g., nitrogen oxides from internal combustion engines and halogen compounds from aerosol dispensers and other sources) in controlling the concentration of ozone in the upper atmosphere. They are needed more generally for designing practical new photochemically induced organic and biochemical synthetic pathways.

6. *Studies of the chemistry of surfaces.* A large variety of important chemical processes occur at the surface of catalysts. Particularly noteworthy among them are the processes by which petroleum is transformed into fuel products of specified quality and into the petrochemicals on which the synthetic plastics industry is based. Yet, most of the information about how these processes occur is empirical, with little theoretical foundation for making predictions regarding catalytic specificity and effectiveness. Surface chemistry is now entering a new era. Ultra-high-vacuum technology combined with such techniques as low-energy electron diffraction and Auger spectroscopy now permit us to prepare and control adsorbents on surfaces, determine their composition, and study their chemistry. These new experimental advances need to be coupled with progress in theoretical computation of the energy states of molecules interacting with specific surfaces.

7. *Nonnumerical methods.* This is a developing area that includes applications of artificial intelligence, pattern recognition, symbol manipulation, graph theory, and computer graphics. These methods can often simplify a complex problem to the point where a numerical solution is finally possible. For example, graph theory may be used to represent systems of coupled kinetic equations, artificial intelligence may be used to find productive sequences of reactions for synthesis of a compound, and pattern recognition may be used to predict biological activity and even to guide energy minimization calculations. Another application is the computer modeling of complex organic structures such as coal and wood, which have a direct bearing both on the national energy problem and on our national resources for synthetic organic

chemistry. The coupling of nonnumerical methods with numerical methods is an important research area for the NRCC.

8. *Studies of biologically active molecules.* Promising results have been reported in the correlation of specific kinds of pharmacological activity with the electronic structures of molecules. The theoretical computation of structures can become a highly valuable and cost-effective adjunct of experimental screening programs in the search for new pharmaceuticals having desired biological activity, if the computational resources can be made available for the needed developmental studies.

These are just a few illustrations of the types of problem that can in time be attacked with computational support such as will be made available at the NRCC. A much more detailed list is outlined in the Wiberg report. The specific studies to be undertaken, of course, will be determined by the imagination, intuition, and needs of the research scientists throughout the nation who will call on the resources of the facility. We envisage as major areas where computational support will be made available initially through the NRCC those listed in Table 1.

TABLE 1 Major Program Areas

Energy Systems
Quantum Chemistry
Statistical Mechanics
Chemical Kinetics
Physical Organic Chemistry
Macromolecular Science
Crystallography
Nonnumerical Methods

4 Unique Role of the NRCC

Computation in chemistry has had a curious history. Many chemists recognized early the potential of computers for solving important chemical problems and in the 1950's and early 1960's were able to obtain significant allocations of computer time from their home institutions. This was possible because research funding organizations such as the National Science Foundation supplied funds to the universities for the purchase of computers. The advances in chemical research through the use of computers during that period were remarkable at those institutions where computing time was available at little or no cost. During this period, university computing centers were not extensively used in undergraduate instruction.

During the past ten years, the development of theory in chemistry and its application to important experimental problems have reached a fruitful stage. But at the same time, the availability of computational facilities to chemists has not increased correspondingly. The new developments require in many cases third- or fourth-generation computers*; but few universities have such facilities available. Chemists at the universities are now in competition for computer time with a wide and expanding variety of other scientists, plus the educational and administrative demands for time. Extensive subsidization of computational chemistry by the university has necessarily been reduced.

Thus we have the paradox of increased ability to solve important problems by computation, along with a constant or decreasing level of facilities available to

*CDC-6600, IBM 360/91, and Univac 1108 are third-generation computers; CDC-7600 and IBM 370/195 are fourth-generation computers.

make use of this ability. The NRCC would make an important start in reversing this trend and maintain a better balance between what is possible with current computer technology and what we can actually do. Without the NRCC, the present undesirable condition can only deteriorate further.

The NRCC in its function as a user-oriented facility can implement uniquely the possibilities and opportunities described in Chapter 3 in several different ways.

HUMAN RESOURCE FUNCTIONS

1. Development of efficient algorithms and studies of possible hardware implementation. Computer time is a limiting factor in the solution of many important problems. The development of more efficient ways to deal with the computations would make it possible to solve larger and more significant problems within the constraint of the available resources.

2. Documentation, testing, and improvement of existing computer programs. Many potentially important computer programs have been written to solve specific problems. They are often not useful to other investigators because of inadequate documentation or because some parts have been designed for the specific problem rather than for the general case. The individual investigator has little incentive to develop and document his program further. An NRCC in-house staff could make a major contribution to the development of chemical science by doing the work required to make such programs efficient, broadly useful, and well-documented.

3. Computation of specific systems. It is frequently a major effort to adapt a computer program from one university center to another. Often the effort in making it operational is not justified for application to a specific chemical research problem under investigation by the experimentalist. There is wide interest in the user community in having available operational programs that could be used at the central facility to calculate many properties, e.g., energies, rates, thermodynamic properties, of specific chemical systems. In general, the computational effort for such work is small once a computer program has been made operational.

4. Development of needed computational methods. Areas urgently needing improved computational methods would be identified through workshops organized by leaders in various areas of chemical research. An in-house staff

would then proceed to implement the recommendations from the workshops, either alone or in collaboration with interested users.

5. Development of specialized hardware, utilizing the rapidly developing technology of mini-computers and microprocessors. Such technology has been usefully exploited in high-energy physics and should be similarly applicable to efficiently solving problems in molecular dynamics, crystal structure, and other areas of chemical research.

6. Development of specialized software and languages particularly suited for the chemical community, as well as standards for machine-independent software. This effort will reduce the cost of moving software from machine to machine and the cost of chemical computing in general.

7. Development of provisions for remote access, including the possibility of utilizing a nationwide computer network.

8. Establishment and updating of a data base. A unique resource of the NRCC will be the accumulated results of computations, such as of molecular wavefunctions and integrals, which could be of considerable value to other scientists. Such valuable and expensive data should be retained in a form accessible to other scientists. An in-house staff can ensure this at a central site such as is envisaged for the NRCC.

INTERACTION WITH THE CHEMICAL COMMUNITY

The NRCC is conceived as an organization serving the needs of both experimental and theoretical chemists throughout the nation. It will do so by carrying out the activities described above and by making computing facilities available. But this in itself is not enough. It is also important to provide for interaction among parties sharing common computational interests. This would be done via the following:

1. Workshops. These may be devoted, for example, to (a) the question of what are the important and yet unsolved computational problems; (b) cooperative efforts to develop software; (c) the most effective usage of major computer programs available at the facility; (d) nonnumerical applications of computational methods to chemical research, such as pattern recognition and graphic displays. Such activities would better define the role of

the NRCC, would help to minimize duplicative effort by different groups in solving a given problem, and would help to inform experimental chemists in the use of complex programs designed for theoretical calculations. The workshops would benefit the research of many investigators who would not necessarily require the use of hardware made available by the NRCC.

2. Visitor programs. One effective way to assist the development of computational methods in chemistry is to bring interested scientists to the NRCC for periods ranging from weeks to months so that they may interact with the in-house staff and, even more important, so that they may interact with each other. Communication is a vital and often neglected element in the development of science. A visitor program, along with the workshops, would greatly facilitate communication among individuals and among research groups.

3. Modes of utilizing computational facilities. Users of the NRCC may conveniently be divided into two classes: clients and customers. The former would generally be working on major projects calling for large use of computer time and would submit proposals to the Program Committee requesting a portion of the computer time available to the NRCC. In return for the computer time given to them on the basis of the scientific merit of their proposals, they would be expected to make the programs they use and develop at the NRCC available to other users. This is an efficient method to further the development of computation in chemistry and at the same time to disseminate programs to the chemical community. Customers would generally be smaller users who would not be expected to contribute new software. They would pay for services from their own funds, whether derived from individual research grants or from operating budgets of their institutions. The services so rendered, consistent with the availability of computer time, as allocated and approved by the Policy Board, could range from providing an on-line program and computer time to taking the customer's data, running the necessary computer programs, and sending the results to the customer. Such provision for customers would enable relatively unsophisticated (in the computational sense) investigators, who now are unable to use computational techniques in solving their chemical problems, to have access to the NRCC facility and to utilize programs and associated services not available elsewhere.

Customers would be encouraged to use their own facilities, when the home facilities are appropriate to the

needed function. Part of the service provided by the NRCC would be the improvement of software to make their use more efficient at home computer centers, thus enhancing the ability of home centers to service both customers and other users more effectively at lower cost. These customer services would make the resources of the NRCC broadly accessible, and thus benefit chemical research accordingly. Rather than compete with university computing centers, the NRCC would increase their effectiveness by making specialized programs for chemical computation available to them.

IN-HOUSE STAFF

A small in-house staff is necessary to process and coordinate the work of the Resource. Expertise is needed in various areas of chemistry, e.g., quantum chemistry, statistical mechanics, chemical kinetics, crystallography, macromolecular science. In addition, programmers will be needed. The Resource should have its own small staff of computer scientists and be located and organized in such a way as to provide interaction with strong viable groups in applied mathematics and computer science. Some of these professionals might hold joint appointments between the national Resource and some other institution. A number of specialized tasks would be best fulfilled by paid consultants in particular areas.

The Resource staff would participate in workshops, carry out program improvement and documentation, develop software, carry on their own research programs as well as research in cooperation with the user community, consult with customers, and perform the customer services. The research done by the in-house staff would be a small fraction of the total operation of the Resource and subject to the same restrictions imposed on external users. To attract the quality of staff necessary for the Resource, the in-house staff would devote on the average about 30 percent of their effort to research.

5
THE NRCC -- AN ORGANIZATION TO SERVE THE USER COMMUNITY

During the study conducted by the Wiberg Committee, a number of federal laboratories and institutions were identified that could supply the major computing hardware requirements of the NRCC under Phase I of the recommendations of that Committee. Phase I envisaged an organization that would initiate and carry on a program of systematic software development, research in chemical computation, negotiations for computer time, and other related functions, without a commitment to the purchase or management of a third- or fourth-generation computer. Working groups of that Committee also gave consideration to the possibility of utilizing existing consortia to provide the management of the NRCC.

In the present study, we have further explored in detail the utilization of existing consortia. Six federally supported laboratories have submitted proposals to NSF, to ERDA, or to both to operate the NRCC as a division, or under an existing division, within their present operating structures. These have been reviewed as part of the present study. We have availed ourselves of consultation with leaders in the management of federal laboratories, consortia, and other nonprofit institutions. Among the persons interviewed were the following:

E.F. MacNichol, Jr., Assistant Director for Research Services, Marine Biological Laboratory
Paul McDaniel, President, Argonne Universities Association
Norman F. Ramsey, President, University Research Associates
Gerald F. Tape, President, Associated Universities, Inc.

We have also had extensive private discussions with individuals prominent and interested in the application of high-speed computers to chemistry.

Many of the organizational and management requirements of the NRCC are comparable and similar to those that have been developed in the use of high-energy accelerators and astronomical facilities. These include a policy board, a program review, a user community group, and an administrative structure as part of the facility. In each of these cases the organization has grown with the facility. In the case of the NRCC, a small operation requiring independence and identity will have to be folded into a large, existing, operating organization without detriment to the latter and in a way that would nurture the growth and utilization of the NRCC. It must be an organization committed to serving the external user community. Since the organization will utilize hardware at one or possibly more federally supported institutions in Phase I, it should be capable of negotiating and collaborating with those institutions. Operation under the management of two or more institutions would be impossible. The management structure in Phase I should be capable of fostering the growth of the NRCC to Phase II, when it could independently justify acquisition of its own major hardware. It should be sufficiently flexible to grow to the responsibilities associated with Phase II operation. The work of the NRCC at any federal laboratory should not impinge on work in related areas already under way or on future expansion at federal laboratories. Furthermore, the goals of the NRCC will not be achieved by dominance of the effort by a local in-house staff, however competent, if subordinate to the broader mission of the host institution.

The above considerations, then, rule out operation of the NRCC under an existing department of chemistry or of computer science at a federal laboratory. Among the difficulties in operation as a division of a given laboratory or institution we call attention to the fact that personnel associated with the NRCC must be subject to the same appointment policies as other comparably trained personnel of the institution. Some of these institutions have a ladder of professional appointments that include tenure. It is difficult to see how tenure appointments could be made during Phase I operation. If the decision is made to operate the NRCC as a division of one of the present ERDA facilities, the organizational structure must ensure independence of program and policy through guidelines mutually satisfactory to (1) a policy board of the user community, (2) the management of the ERDA facility, and (3) the federal funding agencies. It should be so constituted as to be able to contract with and accept support from the private sector.

The organization must have a Policy Board, which would have responsibility for appointment of the Director of the NRCC, the establishment of overall policy of the organization, and all the usual functions of a corporate board. The appointment of the Director would be made in consultation with interested parties, including the management of federal laboratories cooperating with the NRCC, the program and user groups, and the funding agencies.

Although we are reluctant to recommend the establishment of yet another organization, we do not see that operation of the NRCC as a subunit under any of the more promising existing consortia is optimal for achieving the goals set for the NRCC. We are cognizant of the successes achieved by Argonne Universities Association, Associated Universities, Inc., and University Research Associates, each of which was organized to operate national facilities in neutron physics, high-energy physics, radio astronomy, and other areas of science. The fact that the NRCC must serve all of chemistry, not just the university sector, makes participation in management by public-spirited individuals from all sectors desirable. We are encouraged by the successes achieved by such independent organizations as the Gordon Research Conferences and the Marine Biological Laboratory. *The needs of the NRCC can best be met by a nonprofit organization whose policy is set by a cross section of individuals from the disciplines of chemistry, computational science, and research management. The long-range advantages of an independent organization could more than offset the initial effort required to establish a new administrative organization.*

6 Computing, Communications, and Site Requirements

In order that the NRCC make a unique contribution to chemical computation and software development, it must have access to the most advanced and powerful computational equipment consistent with reliable operation. In order that the functions be performed with participation and benefits by chemists throughout the nation, communications facilities must be available for convenient and equitable remote access. A good scientific library readily available to the NRCC staff, visitors, and users is an essential part of the site facilities.

Further, in making a choice between various existing sites as a location for the NRCC, it will be useful to keep the following criteria in mind:

1. The present staff at the site should already have wide experience with outside users.
2. The site should already have a high-quality staff of applied mathematicians, computer scientists, and computer hardware engineers with a traditional interest in the user community and with a clear interest in having the NRCC.
3. It would be useful if some of the computer scientists are involved in (a) designing and implementing machine-independent languages and (b) developing systems for artificial intelligence.

COMPUTING FACILITY

A computer having the speed and memory of a CDC 7600, an IBM 370/195, or their equivalent is the minimum with which the NRCC can fulfill its function. We anticipate that a minimum of 50 million words of mass storage in drums, disks, or data cells will be required. The availability

at the Resource of an Extended Core Storage or Large Core Storage device would enhance computer performance on many of the problems encountered in chemical computation. Adequate peripheral equipment such as tapes, card readers, punches, and line printers obviously will be needed and provision should be made also for a variety of forms of graphical output.

COMPUTING SOFTWARE

The computer operating system must include time sharing. It must have provision for submission of remote batch jobs and employ a simple job control language. The facility must provide an adequate program library and program documentation, in machine readable form if possible.

Standard (manufacturer supplied) operating systems and compilers should be available and usable at the Resource. This would facilitate program transferability and reduce the effort needed for program adaptation and system maintenance.

SYSTEM PERFORMANCE

The computer system and its facilities must be reliable and operate at close to 95 percent of the scheduled time. Batch turnaround time should not exceed the computation time by more than 1/2 hour in most cases. Response of the time-sharing system should average 2 seconds or less.

COMMUNICATIONS HARDWARE

The computer system must provide an ample supply of ports for remote access by teletype and remote job entry. We suggest an initial capacity of 40 dial-up teletype ports, preferably of programmed speed up to 4800 bits per second (bps). We suggest provision of at least 10 dial-up ports for remote job entry with speeds in the 2000 to 9600 bps range. It is important that most ports accept the American Standard Code for Information Interchange (ASCII) code configuration.

COMMUNICATIONS ENVIRONMENT AND NETWORKING

The site must be serviced by a phone system with adequate circuits and alternate routes. It is desirable that the site have access to the Advanced Research Projects Agency Network (ARPANET) through an interface message processor (IMP), or a terminal interface message processor (TIP) with a high-capacity circuit. Research should be initiated on the feasibility of utilizing existing computer networks, and of creating a dedicated network for the NRCC that could incorporate several different operating systems. Such development would facilitate interactive remote access to the NRCC by users trained in different systems.

SITE ACCESS AND HOUSING

Office space will be required for a staff of 20 persons. Housing at moderate cost must be available to temporary staff and visitors. The site should have convenient access by air, rail, and/or highway. Access to the site and computer area by staff and visitors must not be unduly hampered or delayed by site security restrictions.

7 Proposed Phase I Operation and Funding Level

The levels of operation proposed by this Committee are a result of our review of the various ERDA laboratory proposals, information obtained in the course of site visits, our own assessment of the scope and nature of the NRCC programs, and the response of the interested user community to our questionnaire. The Committee also reviewed prior surveys summarized in the Wiberg report, which includes estimates of the current levels in chemical computation. Since a critical complement of staff is needed to support a viable scientific program, the Committee recommends that Phase I be implemented at a single site rather than be assigned to two or more groups in collaboration at different sites.

MANPOWER

The total scientific staff including visiting scientists should begin at a first-year level of 15 and be augmented to 20 in the second and 25 in the third years of operation, distributed as shown in Table 2.

COMPUTING

We propose that computing capacity equivalent to 1000 hours of CDC 7600 time should be provided in the first year of operation. We anticipate an increased demand to 1625 hours in the second and 2500 hours in the third years. A cost of the order of $400/hour is anticipated for this computing capacity.

TABLE 2 Proposed Personnel Levels

Personnel	Year First	Second	Third
Scientific staff[a]	8	10	12
Visiting scientists[b]	2	4	6
Workshop leaders[b]	1	1.5	2
Joint appointments[b]	1	1	1
Consultants[b]	1	0.5	0.5
Secretarial-clerical	2	3	3
TOTAL MAN-YEARS	15	20	24.5

[a] In-house chemists, computer scientists, programmers.
[b] Full-time equivalent.

OTHER REQUIREMENTS

We expect that additional expenditures will be required for workshops to define the initial software and scientific objectives of the center, for consultants with specialized skills in software development, for communications to supply modems and lines to distant centers of chemical computations, for specialized equipment and terminals at the center, and for expenses of the program review and Policy Board functions.

FUNDING

A proposed level of support for the basic activities of the NRCC is itemized in Table 3 for the first three years of operation. A more detailed budget will be possible after the center begins operations, but the proposed levels and categories should be sufficient to support an effective Phase I period of development. An initial commitment must be made for annual funding, if a three-year initial commitment is impossible.

TABLE 3 Budget Projections[a]

	Year First	Second	Third
Staff[b]	440	575	700
Visiting scientists	80	160	240
Consultants	40	25	25
Workshops	50	75	100
Program review and Policy Boards	40	40	40
Equipment	150	200	200
Communications	100	100	100
Computing[c]	400	650	1000
TOTAL	1300	1825	2405

[a] In thousands of 1974 dollars, including indirect costs estimated at 50 percent of salaries.
[b] Includes joint appointments, post-doctorals, and all scientific, professional, administrative, secretarial, and clerical staff.
[c] Calculated at $400/hour for a fourth-generation computer.

8 Funding of Computer Charges and Software Development

Experience in the operation of major facilities shared by a scientific community, e.g., high-energy particle accelerators, optical and radio telescopes, and research ships, shows that the only feasible method of operating the facility is to have the operating funds for the facility as a single budget item allocated to the operating organization. The utilization of the Resource and allocation of the time available by the facility is determined by policy and program review committees. One of the functions to be served by the NRCC (service to clients) fits exactly into this mode. We conclude, therefore, that software development and computer operating time for programs authorized by the Director, based on the recommendations of the policy and program committees, be funded through the central budget of the NRCC and be distributed to the clients. There are already precedents in the funding of computer usage of university investigators in atmospheric science at the National Center for Atmospheric Research (NCAR) through this mode and similar precedents for the astronomical community at the national observatories.

An important function to be served by the NRCC is the upgrading of existing programs to carry out computations in many areas of quantum chemistry, statistical mechanics, scattering theory, and physical organic chemistry and to make these operational for the calculation of properties of specific systems. An example of the latter could be the calculation of the energy states of some proposed new drug or an intermediate in a chemical reaction. Such energies can be calculated by well-established principles to various degrees of sophistication. The effort involved to develop the software and carry out such a computation for a single chemical system cannot usually be justified. As part of a routine service, the cost becomes trivial in comparison with the overall cost of the research program.

This is a service function for which full operative costs should be recovered from the consumer. As interests develop, various aspects of this function could be transferred to the private sector, if suitable commercial organizations show interest.

Chemistry is a science practiced in federal laboratories, universities, and other nonprofit institutions and in private industry. While much of the basic research is carried on in universities, a significant component exists both in federal laboratories and in private industry. The federal laboratories in general have outstanding computer facilities, almost unparalleled in the world. This proposal does not exclude participation of federal laboratories in the program of the NRCC when established. In fact, it encourages it. The major thrust of the NRCC, however, is toward the needs of the university, nonprofit, and private industry sectors. In all cases, the work done at the NRCC must be of nonproprietary nature and be in the public domain.

The federal government has adopted a policy of full cost recovery including capital amortization for work performed for profit-making organizations using federal government facilities. We recommend that this policy be adopted in the operation of the NRCC, again with the provision that the services requested are not available in the private sector.

Not uncommonly, research programs in private industry requiring the use of major computing facilities are similar in nature to those in universities or federal government laboratories. Collaborative efforts on important scientific problems of national interest exist between scientists in private industry and scientists in the universities and federal laboratories and should be encouraged. We recommend, therefore, that nonsalaried appointments to the NRCC be made to qualified scientists from private industry. The work to be performed must be reviewed, like other proposals, by the Program Committee. If it meets the criteria established by the Policy Board, this work could be supported either from the central funding of the NRCC or at the operating rate applicable to federal grant and contract work. Such a determination would accord with a practice of nondiscrimination to all users.

We anticipate a small market for computer services related to the mission of the NRCC in areas not previously established by the client and customer modes of operation. We have called this mode the "market customer" (see Figure

2 in Chapter 9). Typically this could be a request for a computer run by a profit-making organization that would require resources uniquely available at the NRCC (e.g., data base, specialized software). These services would be supplied subject to overall management policies established by the Policy Board and would require full cost recovery. In no case should work of this nature compete unfairly against services available in the private sector.

Patent policies operating at the NRCC should be such as to protect the interest of the federal government but not to inhibit access to the facility by research scientists in the employ of industrial organizations.

9 Administrative Organization

The administrative structure proposed is applicable either to operation of the NRCC as part of an existing organization, or as an independent organization. It is shown graphically in Figure 1.

The responsibility for management will reside in a Policy Board and an executive officer, the Director, who will serve at the pleasure of the Board. As an independent organization, the Policy Board would report and be responsible to the federal agency or agencies with which it contracts for funds in support of the Resource. The Policy Board would be the main unit that establishes policy for the NRCC and would represent the user community on policy matters.

The Policy Board will act as the governing agency of the Resource. It might appropriately consist of seven members of the scientific community elected for three-year overlapping terms. The selection of Policy Board members will follow recommendations made jointly by the Program Committee, the User Committee, and the Chemistry Section of the National Academy of Sciences. If the NRCC is part of an existing organization, the Policy Board will be selected in consultation with the management of the facility. In either case, the Policy Board will have the responsibility of establishing the evolving scientific policy and mission of the Resource, approval of projected budgets, and hiring of the Director. The Board will also review the priorities established by the Program Committee and will see that a justifiable distribution of work in the various areas of chemistry is maintained. In the early phase of the Resource, the Policy Board might meet as often as once per month (less frequently thereafter), and it may be appropriate for some of its members to spend part time in residence. The Policy Board should approve

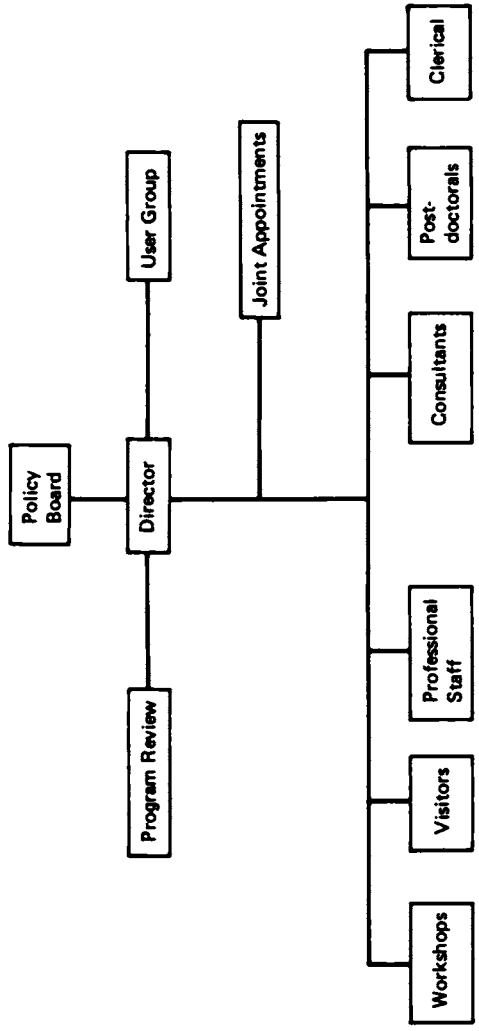


FIGURE 1 Administrative Structure of the NRCC

appointments to the scientific staff, including visiting appointments, recommended by the Director.

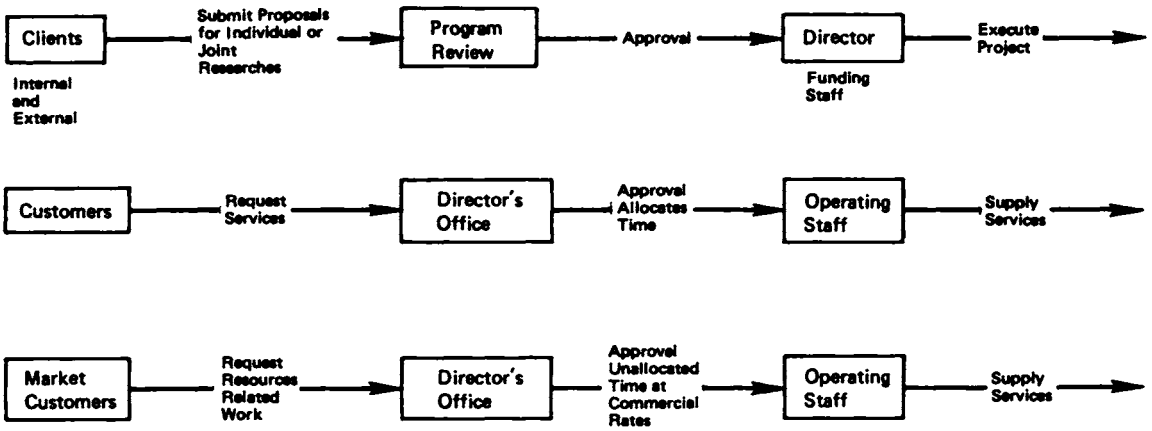
The Director will be responsible for the routine site management, subject to recommendations by the Program Committee and approval by the Policy Board. The Director or his designate will act as Chairman of the Program Committee and attend Policy Board meetings at the Board's pleasure. Under specified conditions the Director may approve new proposals with provision for subsequent review by the Program Committee. The Director may be authorized also to enter into contractual agreements, subject to the approval of the Policy Board, for the lease or purchase of facilities needed in the operation of the Resource, and to negotiate grants and contracts in support of the Resource's normal operations.

The Program Committee might appropriately consist of twelve members chosen by nomination from the Policy Board, and the User Committee. The Program Committee will be so constituted so as to maintain a wide representation of important areas in chemical computation such as quantum chemistry, statistical mechanics, macromolecular science, scattering theory, physical organic chemistry, crystallography, and computer science. The main responsibilities of the Program Committee will be to review the scientific content of both in-house and outside user activities and recommend scientific priorities. In addition, it will advise the Director in recruiting scientific staff, visitors, and workshop leaders.

The User Committee will consist of representatives from the group of outside users of the Resource and from the in-house scientific staff. It will determine its own structure and will presumably be patterned after similar groups associated with the national laboratories. It will provide feedback to the Resource to match NRCC operations to user needs. The Committee will also provide a means for exchange of information between scientists both in and outside the Resource.

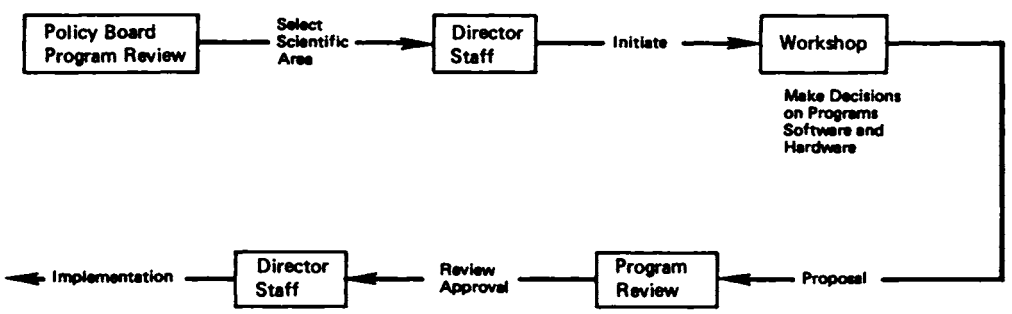
The proposed mode of operation is indicated in Figure 2.

A plan for implementation of major programs at the Resource is shown graphically in Figure 3.



34

FIGURE 2 Proposed Mode of Operation of the NRCC



35

FIGURE 3 Major Program Implementation for the NRCC

APPENDIX A

DECEMBER 1974

QUESTIONS TO PROPOSED HOST INSTITUTIONS
FOR THE NATIONAL RESOURCE FOR COMPUTATION IN CHEMISTRY

I. COMPUTING HARDWARE

1. Give an overall description of your computing hardware system. Would any of its features be particularly suited to the needs of NRCC? Mini- and midi-computers are pertinent.

2. For each CPU specify:
   a. Make and model
   b. Word length
   c. Cycle time
   d. Addressable memory size
   e. Floating point hardware
   f. Double precision hardware
   g. Floating addition time
   h. Floating multiply time
   i. Partial word addressing
   j. Virtual Memory: Can I/O devices access virtual core while CPU is operating?

3. For the memory specify:
   a. The memory size for each CPU
   b. The total memory
   c. The maximum possible memory for the system
   d. The memory cycle
   e. Storage protection
   f. Word length
   g. Minimum byte size
   h. Are the memories interleaved?
   i. Do devices other than the central processor directly access the memories?

4. For the mass storage specify for drums, disks, and data cells:
   a. The name
   b. The number of units
   c. The capacity of a unit in 32-bit words
   d. The average access time
   e. The transfer rate

5. For peripheral equipment specify:
   a. For magnetic tape handlers: the name, number of units, number of tracks, the density (bpi), the start/stop time, the read/write speed (ips)
   b. For printers: the name, number of units, lines per minute, character set
   c. For card readers: the name, number of units, cards per minute
   d. For card punches: the name, number of units, cards per minute
   e. Paper tape reader
   f. Can you write and read IBM labeled and unlabeled tapes?
   g. Is DEC tape available?
   h. Do you have cassette tape drives?
   i. Do you have an electrostatic printer/plotter?

6. For special purpose hardware specify:
   a. What plotters are available; their paper width, their step size or resolution, their speed.
   b. Can you provide microfilm output, of what type?
   c. Is there provision for making color films?
   d. Can you provide interactive graphics, of what type, point plotting, vector drawing, or other scan displays? Are they interfaced easily to your system?
   e. Do you have specialized interfaces such as A/D, or D/A converters, or flying spot scanners?

7. What expansions of your hardware system have been funded and are scheduled to be installed?

8. Describe how your hardware is maintained. Do you have your own hardware staff? If so, describe their capabilities. Can you down units and keep running?

9. Briefly mention any significant hardware development at your center.

II. COMMUNICATIONS HARDWARE

1. How many ports of all types are available? For each type give the number, speed, code configuration (e.g., ASCII, EBCDIC, BCD), protocol required (synchronous, asynchronous, bisynchronous). Does your system support full duplex and half duplex?
   a. How many ports are reserved for system functions, e.g., operator consoles?
   b. How many ports are available for users?
   c. How many ports would be available for NRCC?

2. What is the total port capacity of the communication controllers?

3. What is the highest speed remote transmission you now have operational on a regular basis?

4. Are there any restrictions on the various combinations of ports?

5. What equipment would a user need to access your system? What kinds of terminals are now being used by your users? What is the approximate cost of these terminals?

6. Can a Texas Instruments 30 character per second acoustically coupled terminal (teletype compatible, ASCII) access your system remotely over standard dial-up lines now? If not, what would be involved to permit that?

7. Can your software and hardware support the Vadic 3400 series modems (1200 baud asynchronous full duplex-103 type protocol) or other high-speed modems?

III. COMMUNICATIONS ENVIRONMENT

1. What telephone company services your site?

2. Name the central office from which you are billed.

3. What is the major city through which your calls are routed?

4. Characterize the telephone service in your area with respect to reliability and noise.

5. Comment on data services provided by your local phone company.

6. Are you now regularly part of any computer network? Please describe in detail.

7. Is your system attached to ARPANET? Do you have an IMP?

8. What are your plans to join networks in the future?

9. Would you provide facility for rapid mailing of large-scale computer output?

IV. SOFTWARE

1. Describe the operating system or systems and indicate the maximum size of user programs, taking into account the region occupied by the operating system or enhanced capabilities due to virtual memory schemes. Please specify in bits.

2. What is the job control language: is it easy to use?

3. Describe the program size limitations when the system is running in its normal state. What are the partition sizes if relevant? If these sizes vary on some schedule, describe the schedule. Describe the procedure for running the largest possible job your system can handle.

4. Does the system run unattended during off-hours?

5. Can batch jobs be submitted remotely? How?

6. Can series of jobs be run serially dependent on successful completion?

7. What facilities are there for file manipulation, indexing of files, program library formation, and format conversions? Are files device independent?

8. How are tape and/or disk files protected from being written on or read by unauthorized persons?

9. Do you have tape labeling?

10. Does your linking loader provide overlaying? Is this automatic call overlaying? Can the overlay structure be tree-structured with many levels? Does it have explicit include and exclude of elements?

11. Please describe your procedures for backing up your on-line mass storage devices.

12. Describe the procedure for retrieving a single lost or damaged file, and the procedure for restoring an entire disk or drum.

13. Does the system provide time sharing? If so, what type of time-sharing monitor is used, i.e., virtual memory, paging, partitioning, or swapping? Is the sharing algorithm simple time slicing or a more complex algorithm involving the size of the job and I/O activity? What is a typical response time? How much core is available to a time-sharing job? Can a time-sharing job submit a batch job? How many jobs are active at one time during prime time, and during off-prime time? What are the limitations on a time-sharing job?

14. Does time sharing have the same language as batch? Does it support interactive diagnostics?

15. What languages are supported on the system: compilers, assemblers, interpreters, etc.? What editing facilities are there for data and source program files? Do you have debug and optimization versions of compilers? Do you have dump control; when, if, what?

16. What subroutine libraries are available to NRCC users? Are these for numerical calculations, statistics, pattern recognition, integration, optimization, graphics, nonnumerical computing, symbol manipulation, compiler writing, and simulation?

17. How do you maintain your software? Is specialized application software upgraded to be consistent with system changes by your staff?

18. How is your system documented? How is the documentation maintained?

19. How would you maintain communication with the user group (and chemistry community at large) to inform them of changes in rates, operating procedures, library acquisitions, etc.? How are manuals, updates, and documentation for software made available to users?

20. How do you supply documentation to remote users?

21. What documentation is maintained on-line for remote terminal users to access at all times? How does a new user educate himself?

22. When something goes wrong for a remote user, what facilities are now available to help him solve his problems? Will the machine operator help? Are program counselors available for more detailed questions?

V. SYSTEM PERFORMANCE

a. Number of hours/day that CPU is available.

b. In the case of interconnected devices such as shared disks, memories, communication controllers; indicate the nature of the connection, whether the connections are maintained continually, on schedule or by special arrangement.

c. What is the maximum memory available to any given job?

d. Are your teleprocessing systems, terminal systems, job entry, and time-sharing systems available at all times?

e. Specify the current down statistics, normal turnaround time for jobs of your several classes, and response time for terminal users. Give the mean time between failures (MTBF), and the mean time to repair (MTTR). How long is your system out of service for preventative maintenance (hours/week), or is this done while the system is running?

VI. RATE STRUCTURE

1. What are the details of the present charging algorithm as a function of service and priority class:
   a. For CPU time
   b. For core residency time
   c. For core region used
   d. For input-output transactions
   e. For physical unit record transactions (cards read, lines printed, cards punched, and pages printed)
   f. For special mounting of tape, disks, printer forms, printer frames

2. What is the charging schedule for remote terminal services:
   a. For connect time
   b. For CPU usage
   c. For guaranteed remote access to ports

3. For general purposes list the following:
   a. Charges for on-line storage
   b. Charges for backup and retrieval of information

4. Are users charged for consulting? If so, what are the rates?

5. If the rate structure for NRCC is to be different, please explain the differences in detail.

6. a. Is remote access for outside users now operational?

   b. If so, provide names and addresses of users during the past three years and indicate rate charged. Give the name and address of your user group (if one exists).

   c. Is the rate the same as charged to in-house users?

VII. LIMITATIONS ON NRCC USE

1. What part of the computing hardware will be available to NRCC? Specify the reasons for restricting use.

2. What is the current availability of on-line mass storage and how much can be dedicated to NRCC?

3. What other limitations may be placed on NRCC use of hardware?

4. Describe the operations schedule and time during which there are limits on system use.

5. How much core would be available to an NRCC time-sharing job?

6. Would you permit NRCC to attach special-purpose hardware to your computer?

7. What impact would security classifications have on NRCC with respect to personnel and use? Please be as specific as possible.

VIII. ADMINISTRATIVE QUESTIONS

1. Who would appoint the director of NRCC, how would he be appointed, and to whom will he be responsible?

2. Would both your chemistry and computer science people be involved in NRCC?

3. a. Which of your current staff members are interested in NRCC? (Names, research interests, background, etc.).
   b. How would you allocate their time to NRCC?
   c. Would the above staff members be interested in a joint appointment with NRCC?
   d. Would you make additional (part- or full-time) appointments to your computing staff?

4. List your current research areas that make use of computers.

5. How would you attract first-rate chemists and computer scientists to work on the staff of NRCC?

6. How would you allocate computer time between NRCC, outside users, and the users of your <u>own</u> laboratory?

7. a. How would you cooperate with a stand-alone NRCC facility?

   b. How would you arrange to cooperate with another affiliated facility if more than one is required?

8. How would you handle the scheduling of NRCC batch jobs by local NRCC staff? What would be the policy and mechanisms?

9. How would you handle the scheduling of NRCC batch jobs submitted remotely?

10. How would you handle the scheduling and priority of NRCC remote time-sharing jobs?

11. What problems do you foresee in the above areas?

12. Do you currently have any users from "profit-making" organizations? If so, what rates and policies pertain to them? Are these policies and rates the same as for all others?

13. What would be your policy regarding the attachment to your computer of special hardware for numeric or symbolic computation? What would be your policy on the installation of special equipment for handling remote access by NRCC users? How would you maintain this hardware?

IX. ACCOMMODATIONS

1. Describe the air and rail access to your site.

2. Describe the housing available at your site for short- and long-term use.

3. What work space or office space could be provided for outside users?

4. a. What charges would be made to NRCC for the use of these facilities?
   b. Will overhead be charged on salaries, and if so, how much?

5. How much storage area would be made available to outside users for card decks and for tapes?

APPENDIX B

NATIONAL RESEARCH COUNCIL, DIVISION OF CHEMISTRY AND CHEMICAL TECHNOLOGY

Questionnaire on Potential Use of a National Resource for Computation in Chemistry

January 1975

Please indicate average number of full-time scientists (graduate students, postdoctorals, research associates, staff scientists, etc., exclusive of technicians) including yourself in the research group covered by this reply. _____

Please base your reply on the assumption that a National Resource for Computation in Chemistry (NRCC) will have several staff scientists experienced in chemical computing, an allocation of time on a large computer (equivalent to CDC 7600 or IBM 370/195), and a library of tested and well-documented programs for computing in chemistry.

1. Indicate by numeral 1 which of the following possible functions of the NRCC would be most helpful to your research group. Rank by 2, 3, etc. any other functions that you might also make use of.
   (a) Access to large computer remotely _____
   (b) Access to large computer by short visits _____
   (c) Interaction with other scientists _____
   (d) Access to staff specialists in chemical computing and computer science _____
   (e) Access to operational computer programs for chemical calculations _____
   (f) Other _____

2. Remote Access
   (a) Are you currently accessing a remote computer by phone? _____
   (b) If so, what computer are you accessing?
   Make and Model _____
   Location and Institution _____
   What kind of remote terminal are you using? _____
   Do you use normal telephone lines? _____
   Leased line? _____
   Other (specify) _____
   (c) Has your use of a remote computer been satisfactory? Indicate in order of importance, three factors that would improve the efficiency of your remote computing:

1. _____
2. _____
3. _____

3. Modes of Access
   (a) Indicate by numeral 1 which mode of access to NRCC would be most useful for your present research computational needs:
   1. Personal visit to site _____
   2. Submission of problem or data by mail _____
   3. Remote batch entry _____
   4. Interactive terminal local to your own site _____
   (b) Estimate the total man-days per year you anticipate that all members of your research group might spend on-site at NRCC _____
   (Your reply to this question will assist us in estimating on-site housing requirements.)

4. Estimated Usage
   (a) How much has your research group spent for computer time ($) during the past twelve months? (Please do not include any allocation of either time, or dollars, from your institution specifically earmarked for computing at your own institutional computing facility.)
   $ _____

5. Memory Requirement
   Indicate the memory size required (in 32-bit words) for
   your largest program _____;
   your typical program _____

Please feel free to append comments and suggestions.

Name _____
Institution _____

Returns will be kept privileged and statistical summaries, only, will be released. Your reporting your name and institution is optional, but will be helpful to us in assaying the reliability of the sample of returns received.

SUMMARY OF RESPONSES

The questionnaire was designed to elicit present (1975) information from potential users on the use they would make of a National Resource for Computation in Chemistry. Naturally the respondents in thinking about an institution not yet in being, with facilities and scientific policy yet to be established, were unable to be precise in formulating their interest in and anticipated use of the NRCC. Nevertheless the returns are instructive and will be helpful in further planning.

The questionnaire was sent to the ACS-approved chemistry departments of 522 colleges and universities, of which returns were received from 196 (38%). The returns are fairly representative; 82 were from the 190 PhD-granting departments (43% of these departments), including 8 from the top 19 in PhD production (42% of these departments). The questionnaire was sent also to research directors of 340 industrial companies engaged to a significant extent in chemical research; returns were received from 85 (25%). Finally, the questionnaire was sent to 600 individuals recognized as contributors to chemical research involving large computers in university, industrial, governmental, and nonprofit institutional laboratories, and returns were received from 248 (41%) of this group. The returns from these three categories of respondents are treated separately in the following analysis. Some overlap exists in the statistical information obtained in the third category with that obtained in the other two.

A. College and University Departments of Chemistry

Returns from the 196 responding institutions represent 3725 full-time scientists, including faculty members, postdoctorals, graduate students, and other research associates (the figure is uncertain because of variation in the interpretation of the question -- some reported number in entire department while others reported number involved particularly in large-scale chemical computation).

1. *Preference as to NRCC functions* (185 informative replies normalized to 100):

	1st choice	2nd choice	3rd choice
(a) Remote access to large computer	45	15	9
(b) Access to large computer for short visits	4	14	14
(c) Interaction with other scientists	4	14	21
(d) Access to staff specialists	13	24	22
(e) Access to operational programs	34	29	17
	100	96*	83*

*Some respondents indicated 1st choice, or 1st and 2nd choices, only.

2. *Remote access by phone*

(a)

	Number of institutions
Yes, to in-house university computer system	45
Yes, to external computer center or service	51
No	100

Of the 96 users of remote access by phone, 77 report satisfaction, 10 report partial satisfaction, and 9 report dissatisfaction with service.

(b) Computer system utilized by remote access:

	Number of institutions
CDC Cyber 70, 72, 73, 74	6
3300, 3700	5
6400, 6500, 6600	10
6600/7600, 7600	4
DEC PDP-8	1
PDP-10	11
PDP-11/40, 11/45	3
HP 2100	1
Honeywell 645 (Dartmouth system)	6
Interdata 7/50	1
IBM 1130	1
360/51, 65, 67	10
370/145, 155, 158, 165, 168	24
360/75-91; 370/195	3
UNIVAC 70-3	1
1108, 1110	5
XEROX Sigma 9	2
DIALCOM system, Rapiddata Corp.	2

3. *Preferred mode of access*

(a) 196 returns normalized to 100:

	Percentage of returns (%)
0. No access anticipated	5
1. Personal visit to site	8
2. Submission by mail	18
3. Remote batch entry	29
4. Interactive local terminal	38
No choice indicated	2
	100

(b) Estimated number of man-days at NRCC from institution:

	Number of institutions
None	74
1-5	55
6-10	23
11-25	14
26-50	15
51-150	3
200-400	2
Unable to anticipate	10
	196

Total number estimated man-days (196 institutions): 2000

4. *Estimated external usage of computer*
   (a) Dollars spent externally for computer time by 69 departments (including 4 in range $100,000 to $350,000) during past 12 months: $1,060,000.
   Note: of the 127 departments reporting no external use, 81 report significant chemical computation funded entirely by own institution; 46 indicate no significant research based on chemical computation.
   (b) Dollars anticipated to be spent during coming year outside institution's own computing facility by 92 departments anticipating such expenditure: $576,000.

5. *Memory requirements* (32-bit words)

Largest program	Percentage of 151 replies (%)
0-50 k	36
51-100	23
101-500	36
501-1000	4
1001-2000	1

Typical program	Percentage of 151 replies (%)
0-10 k	34
11-50	36
51-100	14
101-500	16

B. Industrial Companies

Of the 85 companies responding:

13 do not use a computer for large-scale computational chemistry; 9 have adequate in-house support and have no interest in NRCC; 63 make up the analysis following.

The 63 included companies report 8700 full-time research scientists (the figure is uncertain because some companies reported their entire research staff while others reported only their major users of chemical computation; probably no more than 5% of the 8700 are in this latter category).

1. *Preference as to NRCC functions* (63 replies normalized to 100):

	1st choice	2nd choice	3rd choice
(a) Remote access to large computer	30	5	10
(b) Access to large computer for short visits	2	6	5
(c) Interaction with other scientists	11	11	35
(d) Access to staff specialists*	11	52	19
(e) Access to operational programs	46	22	16
	100	96**	85**

*One expressed interest in visits by staff specialists to company laboratory.
**Some respondents indicated 1st choice, or 1st and 2nd choices, only.

2. *Remote access by phone*
   (a)

	Number of companies
Yes, to in-house corporate computer	21*
Yes, to external computer service	30
No	12**

*Several of these companies use external service besides own computer.
**One of these uses computer only for literature searches.
   (b) Computer system utilized by remote access

	Number in use*
CDC 3600	1
6400, 6500, 6600	5
7600	1
DEC PDP-10	3
Honeywell 1642, 1648, 6000, 6060, 6068	5
IBM 360/65	2
370/125, 145, 155, 158, 165, 168	23
370/195	1
UNIVAC 1108	5
XEROX Sigma 9	1
GE Timeshare system	11
Other timeshare systems	3
Remote access effected by*:	
normal telephone line	43
leased line	12
other (dial-up, point-to-point, dedicated band)	3

*Numbers include duplications (companies utilizing more than one computer), therefore add up to more than 51.
   (c) Of 51 returns, 48 report satisfaction with use of remote computer; 3 report partial satisfaction.

3. *Preferred mode of access*
   (a) 85 returns normalized to 100:

	Percentage of returns (%)
0. No access anticipated	28
1. Personal visit to site	7
2. Submission by mail	13
3. Remote batch entry	15
4. Interactive local terminal	37
	100

   (b) Estimated number of man-days per year at NRCC from company:

	Number of companies
None	45
1-5	22
6-10	11
11-25	4
25-50	3
	85

Total number estimated man-days (85 companies): 350

4. *Estimated external usage of computer*
   (a) Dollars spent externally for computer time during past 12 months by 40 companies responding: $1,490,000.
   (b) Dollars estimated to be spent during coming year outside company's own computing facility by 32 companies responding: $1,060,000.
5. *Memory requirements* (32-bit words)

Largest program	Percentage of 59 replies (%)
0-50 k	46
51-100	20
101-500	32
1500	2

Typical program	Percentage of 59 replies(%)
0-10 k	39
11-50	44
51-100	10
101-500	7

C. Individual Investigators

The 248 respondents represent 1397 full-time scientists, including affiliated postdoctorals, graduate students, research associates, and others. Some of these individuals are employed at institutions included in sections A and B of this survey. Of the 248 respondents, 195 are in colleges or universities, 31 in industry, 19 in U.S. Government laboratories, and 3 in independent research institutions.

1. *Preference as to NRCC functions* (236 informative replies* normalized to 100):

	1st choice	2nd choice	3rd choice
(a) Remote access to large computer	43	10	8
(b) Access to large computer for short visits	6	18	11
(c) Interaction with other scientists	13	12	22
(d) Access to staff specialists	8	25	24
(e) Access to operational programs	30	31	11
	100	96**	76**

*12 indicated no choice (no interest in utilizing NRCC).
**Some indicated 1st choice, or 1st and 2nd choices, only.

2. *Remote access by phone*

	Number of research groups
Yes, to in-house computer or network	67
Yes, to external computer or service	50
No	131

Of the 117 users of remote access by phone, 102 report satisfaction with service, 7 report partial satisfaction, and 8 report dissatisfaction.

3. *Preferred mode of access*
   (a) 248 returns normalized to 100:

	Percentage of returns (%)
0. No access anticipated	5
1. Personal visit to site	10
2. Submission by mail	14
3. Remote batch entry	38
4. Interactive local terminal	33
	100

   (b) Estimated number of man-days per year at NRCC from individual's research group:

	Number of research groups
None	80
1-5	55
6-10	40
11-25	25
26-50	20
51-150	7
200-400	4
Unable to anticipate	17
	248

Total number estimated man-days (248 individual research groups): 3600

4. *Estimated external usage of computer*
Dollars spent externally for computer time during past 12 months by 118 individual groups (130 others reported that all their computer costs were borne internally, in full, by their institutions): $1,505,000 (includes three large users in range $200,000 to $250,000).

APPENDIX C: WORKSHOP GUIDELINES

In this appendix we present the guideline questions that were mailed to participants in advance of the workshop for distribution to interested colleagues.

Workshop Guidelines, July 1976

I. General Questions

I.1 In summary, what considerations limit the amount, productivity, and innovation of chemical computing in your area (for example, in computing hardware, communications facilities, computing software, community functions, and personnel specializations)?

I.2 What computing hardware, communications, or software developments would make an order-of-magnitude change in your present computational capabilities?

I.3 What projects could be carried on uniquely at NRCC to advance computational methods and/or answer scientific questions:

a) in your specific area?

b) in general scientific computation?

I.4 What projects, if any, do you regard as of major importance in your field that could *not* be carried at an NRCC of the scope specified in the Bigeleisen report?

I.5 Given the growing reliance on dedicated minicomputers and microcomputers in chemistry, in what way can NRCC be useful to that part of the user community?

II. Specific Questions

II.1 Suggest a specific scientific study and system (chemical, physical or ideal) for which computations or software development at NRCC would be appropriate. Estimate facilities required to execute the study (for example, existing software, new software, computing hardware, communications, personnel specializations, and other).

II.2 What existing programs should be made operational at NRCC? For each program supply the following information:

a) Name of program:

b) Names of authors:

c) Language in which written:

d) Computers on which it now operates:

e) Identify present users by name if possible:

f) Identify classes and numbers of potential users:

g) Does it require additional documentation?

h) Does it require testing?

i) What extensions or modifications should be undertaken?

j) Should it be modified to operate on other computers?

k) What portions of the above tasks should be undertaken at NRCC?

l) What facilities and how many persons of what specializations will be required?

m) Would the author favor and cooperate in such an effort?

n) Estimate time reasonably required to complete the above tasks.

II.3 What new software tasks should be undertaken by NRCC? For each provide the following information:

a) Describe program and purpose:

b) Name persons who might work on it:

c) Identify classes and numbers of potential users:

d) Is a workshop or other mechanism required for the planning?

e) What portions of the program writing or planning should be done at NRCC?

f) What facilities and how many persons of what specializations will be required?

g) Estimate time reasonably required to complete the above task.

II.4 For each existing or new program proposed above, compile:

a) Core storage requirements:

b) Mass storage requirements:

c) I/O equipment needed:

d) Whether batch or interactive?

e) Graphics requirements:

II.5 Are there significant data bases of computational results that NRCC should maintain? What kind and how much storage would probably be required?

II.6 Given the enormous advance in LSI technology and microprocessors, what specific computing hardware studies or developments should be undertaken by NRCC? What resources of funds, personnel, and facilities would be required?

II.7 What specific communications studies or developments, such as networks, should be undertaken at NRCC? What resources of funds, personnel, and facilities would be required?

II.8 Suggest topics for early workshops to define activities at NRCC or to uncover new opportunities for scientific computing in chemistry.

_____

Name of respondent (optional)

APPENDIX D: COMPUTATIONAL PROGRAMS FREQUENTLY REQUESTED FROM THE QUANTUM CHEMISTRY PROGRAM EXCHANGE

Table D.1 lists the frequency of distribution for the 16 most ordered QCPE programs for the past 20 months (up till June 1976).

Table D.2 lists the titles of the most ordered QCPE programs.

TABLE D.1 Frequency of Distribution for the 16 Most Ordered QCPE Programs for the Past 20 Months

Program No.	Total Orders	U.S.	Foreign	Academic	Industrial	Government
236	160	90	70	131	14	15
279	110	54	56	85	17	8
174	109	45	64	88	11	10
256	81	31	50	63	12	6
220	77	24	53	54	12	11
261	76	25	51	58	9	9
247	60	28	32	43	12	5
141	58	34	24	47	9	2
176	53	20	33	46	4	3
249	51	18	33	36	8	7
238[a]	51	34	17	38	9	4
232	47	23	24	41	2	4
64	46	25	21	35	6	5
199[a]	45	22	23	33	4	8
286	35	20	15	31	3	1
165	35	20	15	30	2	3

[a]These are actually the same program; one is for the CDC-6600, the other for the IBM 360/370.

TABLE D.2 Titles of Most-Ordered QCPE Programs

236	GAUSSIAN 70:	Ab Initio SCF-MO Calculations on Organic Molecules, 13,779 cards
279	MINDO/3:	Modified Intermediate Neglect of Differential Overlap, 4644 cards
174	CNDO/S:	CI Molecular Orbital Calculations with the Complete Neglect of Differential Overlap and Configuration Interaction, 6000 cards
256	EHT/SPD:	Extended Huckel Program for Atoms through Fourth Row, 2112 cards
220	PCILO:	(Perturbation Configuration Interaction using Localized Orbital) Method in the CNDO Hypothesis, 2410 cards
261	CNDO/2-3R:	CNDO for Third Row Elements, 1803 cards
247	QCFF/PI:	A Program for the Consistent Force Field Evaluation of Equilibrium Geometries and Vibrational Frequencies of Molecules, 6953 cards
141	CNINDO:	CNDO and INDO Molecular Orbital Program (Fortran IV), 1982 cards
176	NORCRD:	Short XYZ Version, 688 cards
249	VSS:	Isoenergy Curve of Electronic Distributions, 277 cards
238	POLYATOM:	Version II (IBM 360), 25,095 cards
232	NMR-LAOCN-4A:	NMR Analysis by Least Squares Fit, 1651 cards
64	EXTHUC:	Extended Huckel Theory Calculations, 1107 cards
199	POLYATOM:	(Version 2) System of Programs for Quantitative Theoretical Chemistry, 18,000 cards
286	ECEPP:	Empirical Conformational Energy Program for Peptides, 4302 cards
165	DNMR3:	A Computer Program for the Calculation of Complex Exchange-Broadened NMR Spectra. Modified Version for Spin Systems Exhibiting Magnetic Equivalence or Symmetry, 1387 cards

APPENDIX E: CORRESPONDING CONTRIBUTORS

Sidney C. Abrahams, Solid State Chemistry Research Department, Bell Laboratories
Norman L. Allinger, Department of Chemistry, The University of Georgia
Alfred B. Anderson, Department of Chemistry, Yale University
V. Austel, Thomae Chemisch-pharmazeutische Fabrik, Biberach an der Riss, Federal Republic of Germany
A. T. Balaban, Institute for Atomic Physics, Bucharest, Romania
J. Norman Bardsley, Department of Physics and Astronomy, University of Pittsburgh
Rodney J. Bartlett, Pacific Northwest Laboratory, Battelle Memorial Institute
Albert L. Batik, American Society for Testing and Materials
Simon H. Bauer, Department of Chemistry, Cornell University
Sidney W. Benson, Department of Chemistry, University of Southern California
Helen Berman, The Institute for Cancer Research
Richard B. Bernstein, Department of Chemistry, The University of Texas at Austin
Michael J. Bielefeld, System Sciences Division, Computer Sciences Corporation
Lesser Blum, Department of Physics, University of Puerto Rico
Russell A. Bonham, Department of Chemistry, Indiana University
Ronald Breslow, Department of Chemistry, Columbia University
Bernard R. Brooks, Department of Chemistry, University of California, Berkeley
William R. Busing, Chemistry Division, Oak Ridge National Laboratory
Joseph Callaway, Department of Physics and Astronomy, Louisiana State University
Felix Chayes, Geophysical Laboratory, Carnegie Institution of Washington
Olav H. J. Christie, Laboratory of Mass Spectrometry, University of Oslo, Oslo, Norway
Ralph E. Christoffersen, Department of Chemistry, The University of Kansas
Gary Christoph, Department of Chemistry, The Ohio State University
Dwight C. Conway, Department of Chemistry, Texas A&M University
Philip Coppens, Department of Chemistry, State University of New York at Buffalo
Charles N. Costakos, Research and Development Division, Chemical Abstracts Service
Richard W. Counts, Quantum Chemistry Program Exchange, Indiana University
John M. Cowley, Department of Physics, Arizona State University
Richard D. Cramer, Research and Development, Smith Kline & French Laboratories
Stanley N. Deming, Department of Chemistry, University of Houston

DeLos F. DeTar, Department of Chemistry, The Florida State University
Paul L. DeVries, Department of Chemistry, The University of Rochester
Michael J. S. Dewar, Department of Chemistry, The University of Texas at Austin
Hugh E. DeWitt, University of California, Livermore
David Edelson, Physical Chemistry Research and Development Department, Bell Laboratories
John G. Edwards, Research and Development, Norwich Pharmaceutical Company
R. N. Euwema, Columbus Laboratories, Battelle Memorial Institute
Jack H. Freed, Department of Chemistry, Cornell University
Gordon A. Gallup, Department of Chemistry, The University of Nebraska-Lincoln
William C. Gardiner, Department of Chemistry, University of Texas at Austin
Johann Gasteiger, Organisch-Chemisches Laboratorium, Technische Universität München, München, Federal Republic of Germany
J. Gether, Central Institute for Industrial Research, Oslo, Norway
T. L. Gilbert, Solid State Science Division, Argonne National Laboratory
Ernest Grunwald, Department of Chemistry, Brandeis University
Keith E. Gubbins, Department of Chemical Engineering, University of Florida
Peter Gund, Research Laboratories, Merck Sharp & Dohme
Howard J. M. Hanley, Cryogenics Division, National Bureau of Standards
Corwin Hansch, Seaver Science Center, Pomona College
Jonathan C. Hanson, The Thomas C. Jenkins Department of Biophysics, The Johns Hopkins University
Herbert Hauptman, Medical Foundation of Buffalo Research Laboratories
Warren J. Hehre, Department of Chemistry, University of California, Irvine
James B. Hendrickson, Department of Chemistry, Brandeis University
Jürgen Hinze, Fakultät für Chemie, Universität Bielefeld, Bielefeld, Federal Republic of Germany
Joseph O. Hirschfelder, Theoretical Chemistry Institute, University of Wisconsin-Madison
Roald Hoffmann, Department of Chemistry, Cornell University
Anton J. Hopfinger, Macromolecular Science Department, Case Western Reserve University
Philip K. Hopke, Institute for Environmental Studies, University of Illinois at Urbana-Champaign
Ward N. Hubbard, Chemical Engineering Division, Argonne National Laboratory
Heinrich E. Hunziker, International Business Machines Corporation, San Jose
Takanobu Ishida, Department of Chemistry, Brooklyn College
Richard L. Jaffe, Ames Research Center, National Aeronautics and Space Administration
Quintin Johnson, Chemistry & Materials Science Department, Lawrence Livermore Laboratory
Peter C. Jurs, Department of Chemistry, The Pennsylvania State University
Malvin H. Kalos, Courant Institute of Mathematical Sciences, New York University
Jerome Karle, Laboratory for the Structure of Matter, U.S. Naval Research Laboratory
C. W. Kern, Department of Chemistry, The Ohio State University
Ralph D. Kern, Jr., Department of Chemistry, University of New Orleans
Thomas F. Koetzle, Department of Chemistry, Brookhaven National Laboratory
Charles E. Kolb, Aerodyne Research, Incorporated

Andrew Komornicki, Department of Chemistry, The University of Rochester
Bruce R. Kowalski, Department of Chemistry, University of Washington
T. M. Krygowski, Department of Chemistry, University of Warsaw, Warsaw, Poland
John R. Laing, Department of Chemistry, The University of Rochester
Kai-Shue Lam, Department of Chemistry, The University of Rochester
G. N. Lance, Division of Computing Research, Commonwealth Scientific and Industrial Research Organization, Canberra, Australia
Joel L. Lebowitz, Department of Physics, Yeshiva University
David Lefkovitz, Department of Computer Information Science, University of Pennsylvania
William A. Lester, Jr., International Business Machines Corporation, San Jose
Paul J. Lewi, Research Laboratoria, Janssen Pharmaceutica, Beerse, Belgium
John W. Lewis, Biomedical Computer Laboratory, Barnes Hospital
David R. Lide, Jr., Chief, Office of Standard Reference Data, National Bureau of Standards
Denis Lisle, Laboratory of the Government Chemist, London, England
Peter Lykos, Department of Chemistry, Illinois Institute of Technology
John C. MacDonald, Department of Chemistry, Fairfield University
Cornelius Maher, Department of Chemistry, University of Michigan
Harold Martens, Norwegian Food Research Institute, Oslo, Norway
J. Doyle McClure, Pulsed Chemical Laser Group, Boeing Aerospace Company
E. A. McCullough, Jr., Department of Chemistry, Utah State University
J. D. McDonald, School of Chemical Sciences, University of Illinois at Urbana-Champaign
F. W. McLafferty, Department of Chemistry, Cornell University
David A. Micha, Department of Chemistry, University of Florida
Josef Michl, Department of Chemistry, The University of Utah
Paulette B. Middleton, National Center for Atmospheric Research
William H. Miller, Department of Chemistry, University of California, Berkeley
Robert K. Nesbet, International Business Machines Corporation, San Jose
Marshall D. Newton, Department of Chemistry, Brookhaven National Laboratory
A. K. Oppenheim, College of Engineering, University of California, Berkeley
Christopher Alan Parr, Department of Chemistry, The University of Texas at Dallas
Jerome K. Percus, Courant Institute of Mathematical Sciences, New York University
Donald R. Petersen, Dow Chemical Company, U.S.A.
Russell M. Pitzer, Department of Chemistry, The Ohio State University
C. Quannes, C.N.R.S., Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, France
John D. Roberts, Division of Chemistry and Chemical Engineering, California Institute of Technology
Robert L. Rowell, Department of Chemistry, University of Massachusetts
Henry F. Schaeffer III, Department of Chemistry, University of California, Berkeley
Barry Schneider, Los Alamos Scientific Laboratory, University of California
Martin H. Schultz, Computer Science Department, Yale University
Isaiah Shavitt, Department of Chemistry, The Ohio State University

David P. Shoemaker, Department of Chemistry, Oregon State University
Ven Shui, AVCO Everett Research Laboratory, Inc.
Martin Sichel, Department of Aerospace Engineering, The University of Michigan
Jack Simons, Department of Chemistry, The University of Utah
Felix T. Smith, Stanford Research Institute
Robert L. Snyder, Division of Engineering and Science, New York State College of Ceramics at Alfred University
Walter H. Stockmayer, Department of Chemistry, Dartmouth College
William B. Streett, Science Research Laboratory, United States Military Academy, West Point
William C. Stwalley, Department of Chemistry, University of Iowa
William C. Swope, Department of Chemistry, University of California, Berkeley
William J. Taylor, Department of Chemistry, The Ohio State University
Alan E. Tonelli, Polymer Chemistry Department, Bell Laboratories
J. Turkstra, Chemistry Division, Atomic Energy Board, Pretoria, South Africa
C. W. von Rosenberg, Jr., AVCO Everett Research Laboratory, Inc.
Albert Wagner, Chemistry Division, Argonne National Laboratory
Fa-Mei Wang, Department of Chemistry, University of California, Irvine
Frank A. Weinhold, Department of Chemistry, Stanford University
Jerry L. Whitten, Department of Chemistry, State University of New York at Stony Brook
Randolph C. Wilhoit, Thermodynamics Research Center, Texas A&M University
Charles L. Wilkins, Department of Chemistry, The University of Nebraska-Lincoln
Jack M. Williams, Argonne National Laboratory
E. Bright Wilson, Department of Chemistry, Harvard University
John Winn, Department of Chemistry, University of California, Berkeley
Max Wolfsberg, Department of Chemistry, University of California, Irvine
Svante Wold, Department of Organic Chemistry, Umea University, Umea, Sweden
George Wolken, Department of Chemistry, Battelle Memorial Institute
Kurt L. Wray, Physical Sciences Incorporated
Jiam-Min Yuan, Department of Chemistry, The University of Rochester
Bruno H. Zimm, Department of Chemistry, University of California, San Diego
Howard E. Zimmerman, Department of Chemistry, University of Wisconsin-Madison
I. Harold Zimmerman, Department of Chemistry, The University of Rochester