The Proposed National Resource for Computation in Chemistry: A User-Oriented Facility (1975)

Pages
61

Size
5 x 8

ISBN
0309338190

Office of Chemistry and Chemical Technology; Assembly of Mathematical and Physical Sciences; National Research Council

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

Order from
National Technical
Information Service,
Springfield, Va.

PB243-812

The Proposed National Resource for Computation in Chemistry: A User-Oriented Facility

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

Planning Committee for a National Resource for Computation in Chemistry

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

William Spindel, *Staff Director*
Martin A. Paul, *Consultant*

Office of Chemistry and Chemical Technology

Richard B. Bernstein, University of Texas at Austin, *Chairman*
William Spindel, Executive Secretary
Peggy J. Posey, Administrative Assistant

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
Computation in Chemistry

June 8, 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 nonumerical 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) non-numerical 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

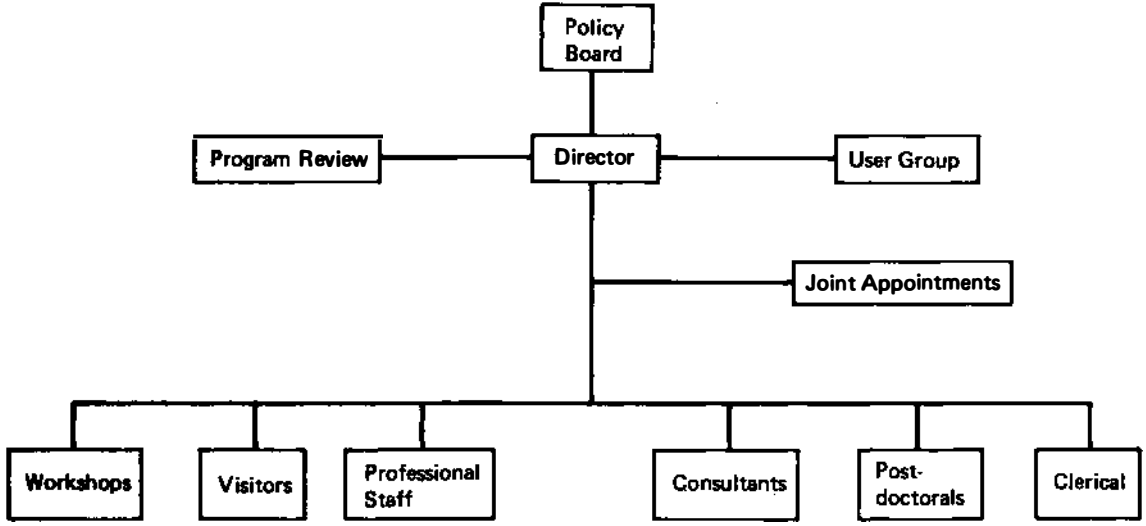

FIGURE 1 Administrative Structure of the NRCC

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.

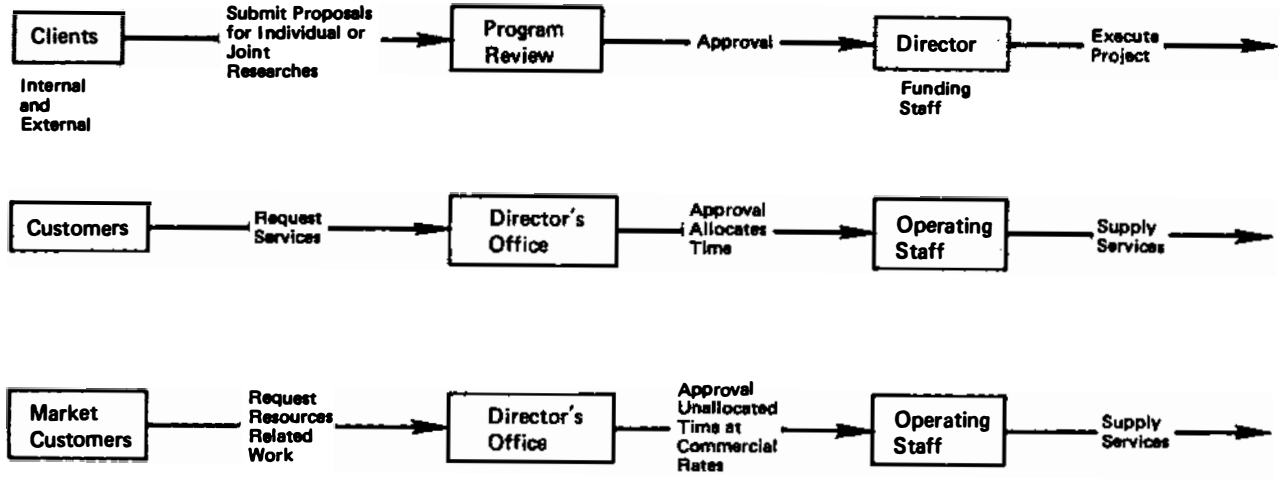

FIGURE 2 Proposed Mode of Operation of the NRCC

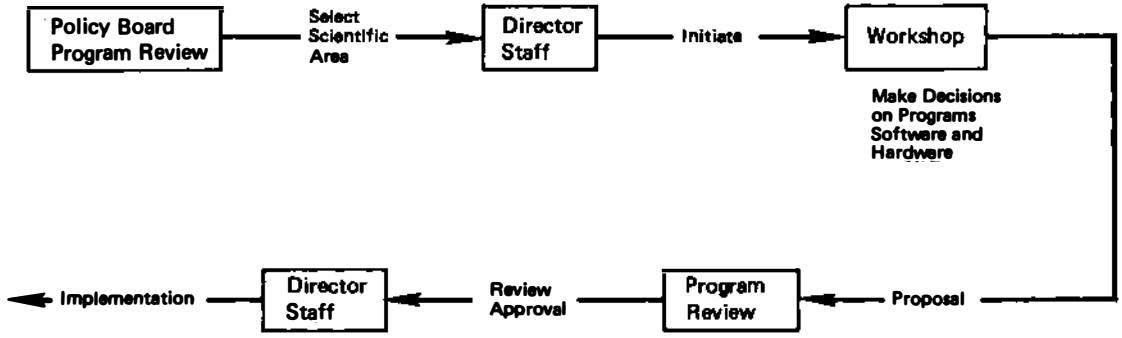

FIGURE 3 Major Program Implementation for the NRCC

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 own 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).