A Study of a National Center for Computation in Chemistry (1974)

Pages
90

Size
5 x 9

ISBN
0309338093

Division of Chemistry and Chemical Technology; National Research Council; National Science Foundation

A STUDY OF A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

National Research Council
Division of Chemistry and Chemical Technology

with support from the

National Science Foundation

National Academy of Sciences
Washington, D. C.
1974

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.

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

Division of Chemistry and Chemical Technology

Cheves Walling, University of Utah, *Chairman*
Virgil C. Boekelheide, University of Oregon, *Past Chairman*
Martin A. Paul, Executive Secretary
Peggy J. Posey, Administrative Assistant

Planning Committee for the Study

Kenneth B. Wiberg, Yale University, *Chairman*
L. S. Bartell, University of Michigan
Jacob Bigeleisen, University of Rochester
R.B.K. Dewar, Illinois Institute of Technology
Frank E. Harris, University of Utah
F. A. Matsen, University of Texas at Austin
Harrison Shull, Indiana University
L. C. Snyder, Bell Laboratories

Martin A. Paul, Staff Director

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

Dr. Philip Handler
President
National Academy of Sciences

Dear Dr. Handler:

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

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

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

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

Sincerely yours,

Melvin Calvin
Chairman
Committee on Science and Public Policy

December 1973

OPPORTUNITIES AND NEEDS FOR LARGE COMPUTING FACILITIES IN CHEMICAL RESEARCH

Statement to the Committee on Science and Public Policy

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

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

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

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

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

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

[1] Recommendation (1), Chapter X.

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

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

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

20 October 1973

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

Summary

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

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

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

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

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

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

RECOMMENDATIONS

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

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

I Background of the Report

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

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

[1] See Appendix information.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

II Computing Needs in Chemistry

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

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

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

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

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

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

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

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

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

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

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

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

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

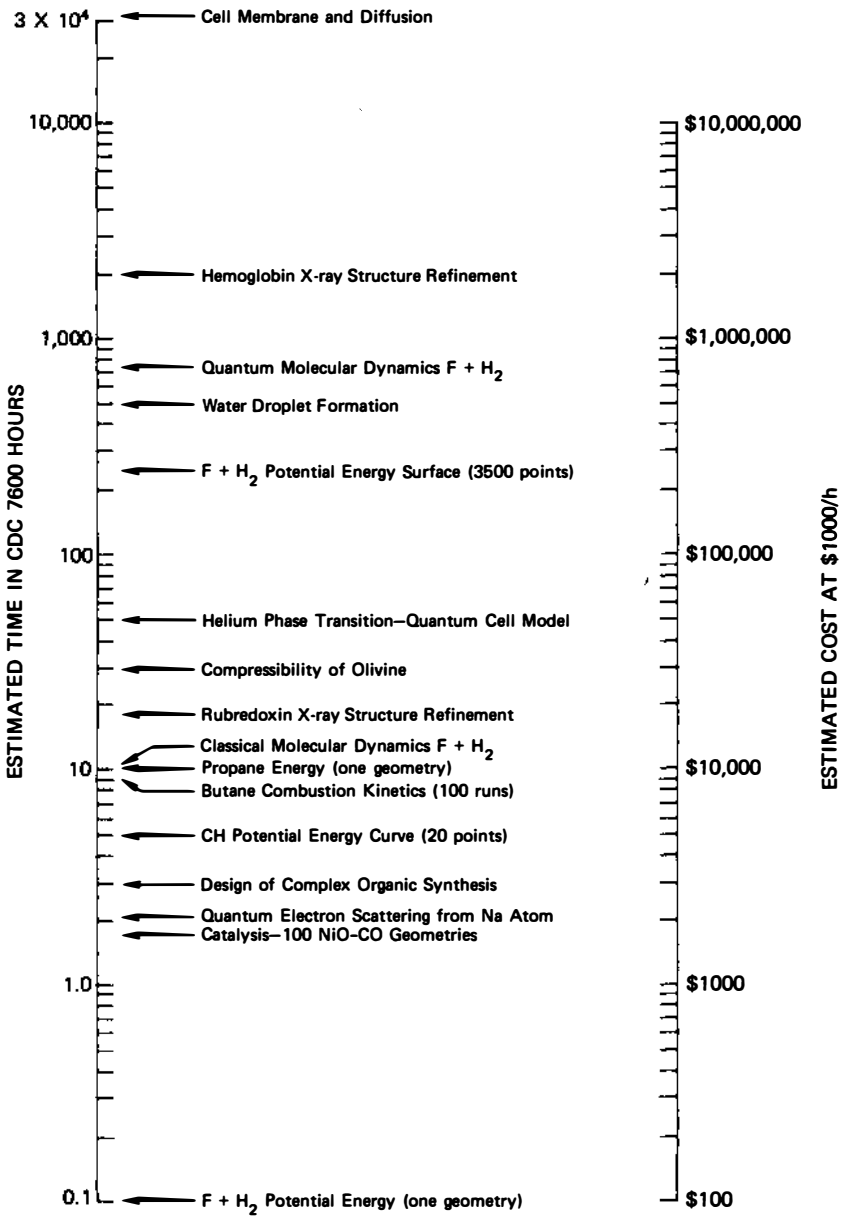

FIGURE 1. Computer requirements estimated for solving representative chemical problems. The times for the systems generally in the lower third of the figure represent actual experience. For the other systems, the times have been estimated by extrapolation of methods currently in practice. The times are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

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

III The Opportunity

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

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

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

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

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

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

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

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

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

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

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

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

IV Computation in the Advancement of Chemistry

THEORETICAL CHEMISTRY TODAY

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

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

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

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

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

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

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

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

5. Van der Waals well depths.[6]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

For Mode 1

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

For Mode 2

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

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

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

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

For Mode 3

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

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

AREAS OF SCIENTIFIC OPPORTUNITY

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

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

POTENTIAL-ENERGY SURFACES

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

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

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

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

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

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

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

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

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

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

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

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

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

COLLISION PROCESSES

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

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

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

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

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

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

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

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

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

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

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

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

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

STATISTICAL MECHANICS

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

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

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

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

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

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

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

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

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

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

OTHER AREAS

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

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

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

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

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

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

ROLE OF MINICOMPUTERS

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

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

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

minicomputers can play a significant role in computational chemistry.

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

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

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

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

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

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

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

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

V Alternatives in Meeting Computing Needs

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

1. TOTAL RELIANCE ON LOCAL COMPUTER CENTERS

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

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

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

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

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

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

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

2. USE OF FACILITIES AVAILABLE IN GOVERNMENT LABORATORIES

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

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

Serious drawbacks, however, include the difficulty of obtaining guarantees or predictions of available computer capacity. It is not in the interest of a mission-oriented facility to admit that it has computer capacity in excess of that needed to support its mission, even when such excess occurs as a result of intelligent planning for future needs. Moreover, because computer needs tend to increase continuously while computer capacity must be changed discontinuously, there will necessarily be large fluctuations in the computer capacity chemists will be able to obtain from any individual government laboratory. It is also unrealistic to expect that the chem-

[1] See p. 00 for relevant figures.

ists will enjoy priorities of use comparable to those available to mission-oriented users. Finally, this alternative shares with several others the disadvantage of not contributing to software development of computer use by a wider group of chemists.

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

3. USE OF COMMERCIAL SOURCES OF COMPUTER TIME

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

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

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

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

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

[a]Using standard operating systems. These rates would ordinarily include free use of peripheral equipment including tape drives and a mass storage disk or drum, and would be on a "wall clock basis." Prime-shift rates would be higher by a factor of two or three.
[b]The first figure is for smaller amounts of time (up to 5 or 10 hours).
[c]Indicates that large time blocks might be negotiated at a somewhat lower rate than the figure shown.
[d]Limited availability.

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

4. CONTRIBUTIONS OF COMPUTER TIME BY THE COMPUTER INDUSTRY

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

5. REGIONAL GENERAL-PURPOSE COMPUTER CENTERS

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

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

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

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

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

6. A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

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

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

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

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

7. A NATIONAL CENTER FOR COMPUTATION OF BROADER SCOPE

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

probably have to be resolved by a more or less arbitrary partitioning of the computer capacity among disciplines.

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

8. ACCELERATED DEVELOPMENT OF COMPUTER PROGRAM EXCHANGES

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

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

9. NO ACTION

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

VI Mission and Priorities for a National Center

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

SCIENTIFIC SCOPE

For a number of practical reasons, the scientific scope of the center should be limited to an area in which the practitioners have to some extent a common language and set of interests. It is proposed that the scope should be *chemistry* in a fairly broad sense. An exact delineation of the scope should not be made in advance, and should in any case be subject to continual adjustment through the evolution of patterns for the approval of requests to use the facilities of the center. At present it is contemplated that the scope should include electronic structure computations, whether done by chemists, molecular physicists, solid-state physicists, astrophysicists, molecular biologists, or others, and should include statistical mechanical and statistical thermodynamic studies of both real systems and models, including such diverse applications as investigations of reacting chemical systems and magnetic phase transitions in model lattices. Also included should be the full range of kinetics and molecular dynamics studies involving atoms, ions, molecules, surfaces, and solids, the interpretation of spectra of various kinds, and the reduction of data of most of the sorts cur-

rently obtained by chemists. Further applications included would be programs for classifying molecules, chemical properties, and chemical reactions, and the use of this information to search for such items as synthesis paths, molecules suitable for particular purposes, or new techniques for chemical analyses. In short, with a few exceptions noted below, the scientific domain of the center should include most of the computational activities it may be expected chemists will wish to pursue.

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

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

ACTIVITIES

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

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

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

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

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

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

PRIORITIES

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

VII Interaction with the Chemistry Community

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

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

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

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

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

COMMENT

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

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

COMMENT

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

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

COMMENT

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

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

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

COMMENT

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

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

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

COMMENT

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

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

COMMENT

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

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

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

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

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

COMMENT

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

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

COMMENT

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

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

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

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

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

COMMENT

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

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

COMMENT

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

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

GENERAL REMARKS

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

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

VIII Organizational Structure

SUPERSTRUCTURE

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

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

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

[1] See also Chapter VII, Comment 5.

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

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

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

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

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

STRUCTURAL FEATURES OF A NATIONAL CENTER FOR COMPUTATION IN CHEMISTRY

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

USERS

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

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

STAFF

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

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

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

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

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

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

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

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

SCIENTIFIC ADVISORY BOARD

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

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

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

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

COSTS

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

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

IX Conclusions

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

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

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

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

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

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

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

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

X Recommendations

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

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

Appendix Computer Use In Theoretical Chemistry 1973: Questionnaire and Data

COMPUTER USE IN THEORETICAL CHEMISTRY 1973: QUESTIONNAIRE

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

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

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

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

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

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

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

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

   Seriously _____
   Moderately _____
   Slightly _____

Name: _____
Institution: _____

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

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

COMPUTER USE IN THEORETICAL CHEMISTRY 1973: DATA

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

APPENDIX TABLE 1 Primary Research Interest

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

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

APPENDIX TABLE 2 Method of Handling Major Computation

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

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

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

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

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

APPENDIX TABLE 4 Computer Used at External Facility

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

APPENDIX TABLE 5 Rates Charged[a]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Participants

Members of the Planning Committee

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

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

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

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

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

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

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

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

Panel Members

Panel I Scientific Objectives

1. Electronic States of Atoms and Molecules

Chairman: F. A. Matsen

Charles F. Bender, Chemist, Lawrence Livermore Laboratory
Richard F. Fenske, Professor of Chemistry, University of Wisconsin
Frank E. Harris
Hugh P. Kelly, Professor of Physics, University of Virginia
*J. A. Pople, Professor of Chemistry, Carnegie-Mellon University
A. C. Wahl, Senior Scientist and Group Leader, Chemistry Division, Argonne National Laboratory

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

2. Inorganic Chemistry, Organic Chemistry, and Spectroscopy

Chairman: Kenneth B. Wiberg

R. E. Christoffersen, Professor of Chemistry, University of Kansas
F. A. Cotton, Professor of Chemistry, Texas A & M University
Donald S. McClure, Professor of Chemistry, Princeton University
Andrew Streitwieser, Jr., Professor of Chemistry, University of California, Berkeley

3. Solid-state and Surface Chemistry

Chairman: Frank E. Harris

*Frank Herman, San Jose Research Laboratory, IBM Corporation
Keith H. Johnson, Professor of Metallurgy and Materials Science, Massachusetts Institute of Technology

4. Statistical Mechanics and Macromolecules

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

H. A. Scheraga, Professor of Chemistry, Cornell University
*Frank H. Stillinger, Member of the Research Department, Bell Laboratories

5. Structural Studies

Chairman: Lawrence C. Snyder

Lawrence S. Bartell
William D. Gwinn, Professor of Chemistry, University of California, Berkeley
*Walter C. Hamilton, Senior Chemist, Brookhaven National Laboratory (deceased)

6. Molecular Dynamics and Scattering

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

R. Stephen Berry, Professor of Chemistry, University of Chicago
Don L. Bunker, Professor of Chemistry, University of California, Irvine
I. Shavitt, Staff Member, Theoretical Chemistry, Battelle Memorial Institute

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

7. Atmospheric Science

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

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

PANEL II CHARACTERIZATION OF COMPUTATIONAL FACILITIES

Chairman: Robert B. K. Dewar

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

PANEL III ADMINISTRATIVE AND FINANCIAL STRUCTURE AND POLICY

Chairman: Harrison Shull

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

OTHER PARTICIPANTS IN THE STUDY

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

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

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

OBSERVERS FROM FEDERAL AGENCIES

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

*STAFF DIRECTOR: Martin A. Paul

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