JOURNAL OF COMPUTATIONAL PHYSICS 2, 417–438 (1968)

COMPUTERS IN CHEMISTRY CONFERENCE

Computers, Chemists, and Crystallography[1,2]

WALTER C. HAMILTON

Chemistry Department, Brookhaven National Laboratory, Upton, Long Island, New York 11973

ABSTRACT

The impact of the development of fast digital computers has been felt in all phases of crystallography, and the crystallographer has been among the most sophisticated users of computers in chemistry. The important interaction between man, machine, and chemical problem has been exhibited in the areas of data acquisition and presentation of the results in a concise and meaningful manner. In addition, the speed of the modern digital computer has made possible the rapid solution and refinement of crystal structures which at one time would have been thought to be near-impossible.

The use of the Brookhaven Multiple Spectrometer Control System, an on-line, time-sharing system for computer-controlled data acquisition, is described in detail. Some examples of direct methods of crystal structure determination are discussed briefly. The techniques of three-dimensional illustration of crystal and molecular structures and computer analysis of thermal motions to present the wealth of material available in a crystal structure determination are discussed with examples. Finally, the importance of improved literature and data retrieval techniques are discussed.

INTRODUCTION

Crystallography is an area of physical chemistry where the increasing use of computers has had an important impact in many different ways. It is also an area where the acceptance and use of the computer has come more rapidly than in many other fields because of the fact that for many years the crystallographer has been faced with time-consuming computational problems that have made familiarity with the computer necessary for survival. Crystal structure determination continues to grow in importance as a tool in the understanding of inorganic, organic and biological chemistry, and computers have provided important aids in presenting the results of the crystallographic investigation to the noncrystallographer in easily understandable forms.

[1] Work performed under the auspices of the United States Atomic Energy Commission.

[2] Paper presented at the Symposium on Computers in Chemistry, sponsored by the American Chemical Society Division of Physical Chemistry and the American Physical Society Division of Chemical Physics, University of California at San Diego, June 1967.

One important feature of a crystal structure is that it is complex: an average-size interesting organic molecule containing sixty atoms. In order to specify the positions of all the atoms, and hence the bond lengths and angles, it is necessary to determine the x, y, and z coordinates of each atom—180 parameters in this example. Furthermore, an interesting part of the description of a molecular structure is the description of the dynamics of the structure. If the motion is assumed to be simple harmonic motion, each atom has associated with it six independent components of a

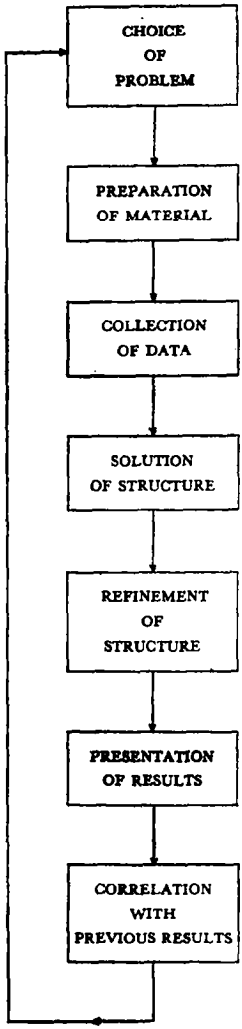

FIG. 1. The progress of a crystallographic structure determination.

vibrational tensor. For our hypothetical sixty-atom molecule, there are thus an additional $6 \times 60 = 360$ parameters. Thus a complete molecular structure determination in this case involves the determination of 540 parameters.

This complexity has the following implications:

(a) Many data are required.

(b) Much calculation is needed to obtain the structural parameters.

(c) Much detailed information is available to be visualized.

(d) Much information is available to be correlated with other structures and other chemical facts.

The chemist is more interested in items (c) and (d), the professional crystallographer in (a) and (b). The complete chemical crystallographer will not neglect any of these areas. The situation is perhaps best viewed as a closed loop, illustrated in Fig. 1. The words *many* and *much* in points (a)–(d) suggest immediately that there are broad possibilities for the application of the digital computer, for the computer is able to do what we are able to do ourselves—only very much faster and without becoming bored. The remainder of this paper discusses some of these areas of application in greater detail.

Computer-Controlled Data Acquisition

The determination of the 540 parameters necessary to describe a 60-atom molecular structure obviously requires at least 540 independent observations. In practice we wish to have many more observations so that we may obtain more precise values for the parameters as well as for the estimates of probable errors in the parameters. A figure of four times as many observations as parameters is typical in crystal structure work today.

Each of the observations made in a crystal structure study is a measurement of the intensity in one of the Bragg peaks corresponding to the reflection of the radiation employed in the experiment from a set of rational planes in the crystal. (Such a single observation is often called simply a *reflection.*) Each observation consists first of the orientation of the crystal in three-space such that the normal to the reflecting planes bisects the angle between the incident and diffracted beams. This is accomplished by setting three Eulerian angles to prescribed positions, usually by specifying the angular position of the shafts on three motors. This is followed by setting the shaft of another motor to define the angle between the incident beam and the radiation counter which detects the diffracted beam. The geometry of the experiment is indicated in Fig. 2. Once the four angles has been set, one or more of the angles may be scanned through a small range and the intensity of the

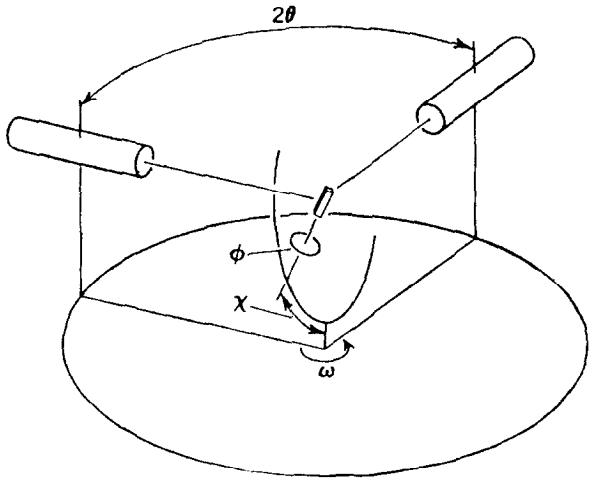

FIG. 2. The geometry of a single crystal diffraction experiment. 2θ is the angle between the incident and diffracted beams. ϕ, χ, and ω define the orientation of the crystal in space.

diffracted beam measured as a function of this scanning angle. The integral under the resulting curve (See Fig. 3) provides a measure of the square of the structure-factor magnitude $|F|^2$ for the reflection. The experiment thus consists of shaft

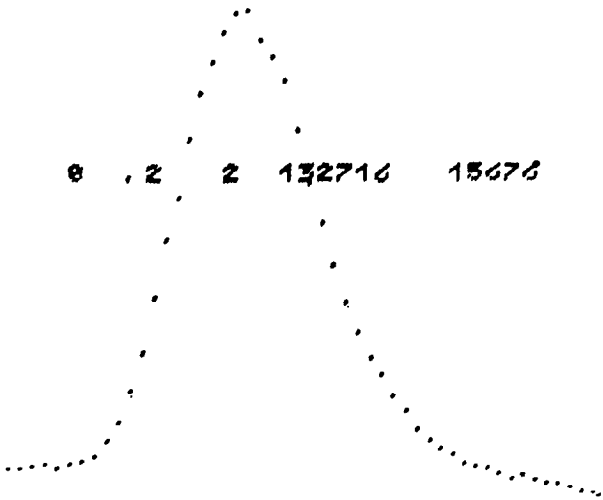

FIG. 3. Profile of a Bragg reflection from a single crystal as displayed on the storage oscilloscope of the Brookhaven MSCS. The Miller indices, peak, and background counts are also displayed. The square of the structure factor is proportional to the area under this profile.

setting, turning counters on and off, and somehow recording the results. This procedure must be repeated thousands of times during one complete experiment; it is a procedure ideally suited to automation.

Many crystallographic experiments have been automated in the past fifteen years. The early instruments involved control of the diffractometer by punched paper tape or cards which provided the necessary information for shaft setting, count timing, and other experimental parameters. We will refer to such instruments as *fixed-program instruments*. One such instrument has been described by Abrahams [1]; another very sophisticated design has been in use for several years at the Brookhaven Graphite Research Reactor [2]. At least three such instruments of commercial design have been widely and successfully used in this country.[3–5] Two commercial analog machines, using mechanical specification of the input parameters, have also been widely used [3].[6]

As a logical next step in instrument development, the tape- or card-controlled instrumentation has been replaced by computer control. The computer is able to perform all the functions of the special-purpose machines and is of course extremely flexible in its operation. This flexibility is one of the main advantages of computer-controlled automation over fixed-program automation. When it is desirable to change the sequence of steps that make up the experiment, it is often possible to do so by changing a few words in a computer program rather than by making hardware modifications to the equipment. The second important advantage of computer control is that feedback is facilitated; the results of previous operations may be used to determine the future course of the experiment.[7]

The kinds of operations that can be carried out include refinement of cell constants and orientation as the experiment progresses, adjustment of counting times for individual reflections as a function of intensity above background at the peak position, and control of temperature and magnetic fields. Although some of these functions could be carried out with fixed-program instruments, the importance of the computer lies in the ease with which new control functions can be added and old ones changed when the need arises. Contrary to some expectations, the experimenter tends to feel closer to his experiment operating through a computer than with a fixed-program wired control box. The first computer-controlled x-ray diffractometer was built by Okaya, Cole and Chambers at IBM [4]. Several other systems have been developed during the past few years including that by

[3] Picker Four Angle Programmer System, Picker Nuclear Co., White Plains, New York.

[4] General Electric-Datex Programmed X-ray Diffractometer, General Electric Co., X-ray Division, Milwaukee, Wisconsin.

[5] Supper-Pace Diffractometer, Charles-Supper Co., Newton, Massachusetts.

[6] PAILRED, Philips Electronic Instruments, Mount Vernon, New York.

[7] Feedback is of course present in any scientific experiment. The computer removes the human from the feedback loop and thus can provide feedback with a much smaller time constant.

Busing at the Oak Ridge National Laboratory [5], and by three commercial organizations.[8–10] One of the most interesting systems is that which was developed at the Brookhaven National Laboratory during 1963–1965 and which has been operating routinely for about two years.

The Brookhaven Multi-Spectrometer Control System

Among the experiments at the Brookhaven High-Flux Beam Reactor are eight neutron-diffraction and inelastic-scattering experiments, each of which involves shaft positioning and counting of the type described above. In addition, control of the physical environment (temperature, magnetic field) of the scattering sample is important in many of the experiments. When the hardware design of these experiments was initiated, it was thought that a fixed-program controller would be used with each experiment. An examination of the cost involved (at least $30,000 and probably $50,000 per machine) led us to consider the possibility of using a single digital computer with time-sharing software to control all eight experiments. It was estimated that the cost of development of a computer system to do the job would be comparable to the total cost of the fixed-program controllers. Further-

TABLE I

Computer and Control Equipment Associated with the Brookhaven Multiple Spectrometer Control System

(a)	Scientific Data Systems 920 computer; 8-μ sec memory access time, 16-μsec fixed-point add time; 656 μsec floating-point multiply (48 bits); 16 kilowords (24-bit) core memory. interlaced buffer for input/output; 24-bit parallel input/output
(b)	32 kiloword random-access magnetic drum
(c)	2 magnetic tape units (200 bits/in.)
(d)	Paper-tape reader and punch
(e)	CRT (Tektronix storage oscilloscope) display with light pen
(f)	Two Teletype typewriters (will be expanded to one per experiment in near future)
(g)	96 priority-interrupt lines
(h)	Scientific Data Systems Model SCS-1 Spectrometer Control Station. This is the interface—one per experiment. Includes switches for manual and computer control of experiments, digital displays which are programmable, breakpoint switches which may be interrogated for program control.

[8] Picker Computer Controlled Diffractometer, Picker Nuclear Co., White Plains, New York.
[9] XRD490 Crystallographic System, General-Electric Co. X-ray Division, Milwaukee, Wisconsin.
[10] Hilger-Watts, Ltd., London, England.

more, the system would be very much more versatile than the fixed-program controllers contemplated.

At the time the system was being planned, an inexpensive digital computer capable of operating a single experiment was not on the market. Thus a system to operate all experiments, sharing time on a more expensive computer, seemed to be the best answer for us. The most serious objection to a shared-time system is that the experiments may interfere with each other and with the systems program in such a way as to cause lost experimental time or invalid data. This is a problem that can be partially solved by memory-protect features which are available on many computers today, but this option was not conveniently available to us in 1963. We considered that the similarity of the several experiments was great enough so that some of the difficulties of time-sharing might be reduced in seriousness. Furthermore, many difficulties can be avoided by careful systems programming. It was concluded that the advantages that the shared time system offered to us

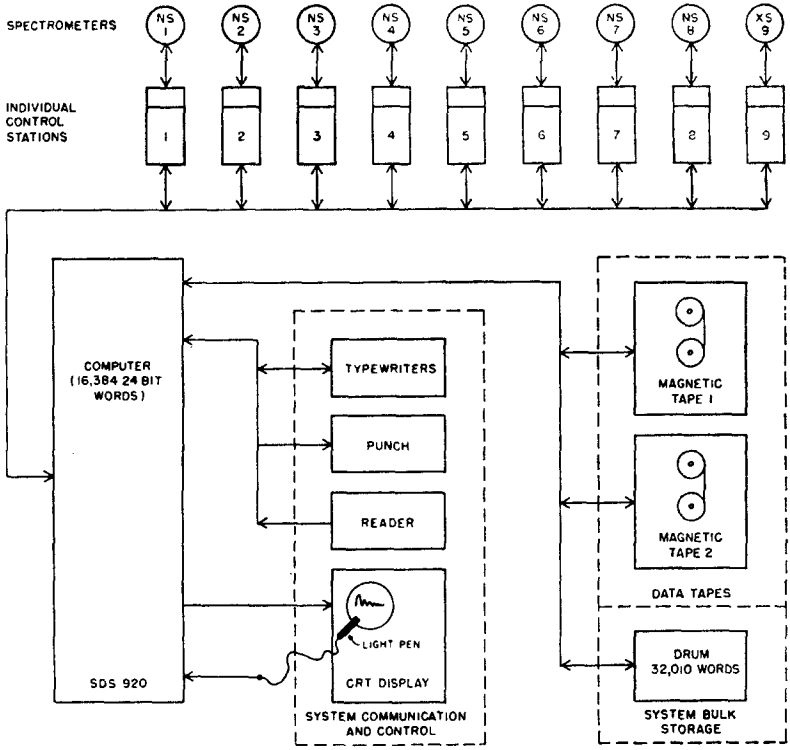

FIG. 4. Components which make up the Brookhaven MSCS. There are two typewriters at the present time, but there will be shortly one for each of the nine experiments.

outweighed the possible disadvantages, and the decision was made to go ahead with such a system.

The computer-controlled system was developed by the Instrumentation Division and the Applied Mathematics Department of Brookhaven; a complete description of the technical details of the hardware and basic software has been published [6]. The hardware associated with the computer system is summarized in Table I. (See also Fig. 4.) The total cost of the system was approximately $500,000; of this about $350,000 was allocated to hardware and about $150,000 to personnel costs for systems design, programming, and technical services necessary for the installation. The present hardware costs for an equally powerful system would probably be considerably less because of the continuing decrease in the price of digital computers, specifically in high-speed memories. There are now nine experiments attached to the system—the nine neutron experiments and a single-crystal x-ray diffractometer. The total cost per experiment was thus comparable to the cost of the commercially available fixed-program instruments. Since the capacity of the system is not completely saturated, additional instruments can be added for the price of a single interface (less than $10,000).

Program Structure of the Brookhaven MSCS

The basic concept of the MSCS systems program may be outlined as follows. Each experiment (user) is allocated a relocatable region of core storage which may be used for storage of a program controlling the experiment (the *user's program*) and for temporary data storage. The systems program (the distributor) calls upon each experiment in turn and questions whether there are any computations to be carried out or operations to be performed; if so, control is transferred to that user's program. The user either returns control to the distributor or control is wrested from him by the distributor if a time limit—currently 10 seconds—is exceeded. The distributor then sequences to the next user. During the time that a user's program has control, the user has access to a large block of common core storage in which he can carry out extensive computations. The subroutines for carrying out these computations are stored on the 32,000 word drum and are called into core as needed. When the user relinquishes control to the distributor, the large block of common storage becomes available to the next user in sequence.

On top of the distributor-sequencing of users' programs, real-time operations come into play. The SDS 920 computer is ideally suited for such operation by the availability of a priority-interrupt system. When a signal appears on a line from an external device, the computation being carried out is immediately interrupted, and transfer is made to a subroutine which must be executed in order to properly control the external device. When the execution of this *interrupt subroutine* is

completed, control returns to the program which was previously being executed. Each such external line is assigned a priority, and a higher-priority interrupt will delay the execution of the routine corresponding to a lower-priority interrupt that has been activated.

The following categories of programs thus exist:

(A) The *systems programs* are permanently resident in core and occupy approximately 4000 words of memory. These include the distributor program, the typewriter program for semi-automatic control discussed below, the basic programs for motor and counter control (mainly interrupt subroutines), and a number of commonly used subroutines for handling input/output for performing floating point arithmetic and for calculation of the elementary functions.

(B) The *users' main programs* are called by the distributor program. They are resident in core but may be easily changed by loading through the typewriter control program. Each user is allocated approximately 1000 words of core; he will typically use about 200 words for temporary data storage and about 800 words for the program controlling his experiment. Such a program will consist mainly of calling sequences to subroutines which are stored on the drum or to systems programs included in (A).

(C) The *library subroutines* are called by the users' programs. They are stored on the drum and brought into the common area of core when needed. The common area consists of about 3000 words and has proved to be adequate for each of our separate computational and control functions. Each routine on the drum is available to any user at the time he has control, so that the user must store all necessary parameters in his own area. In a technical sense, *our procedures are pure.*

(D) The *interrupt subroutines* provide all the real-time servicing of the experiments and are included in the 4000 words allocated to the systems programs above. These routines are primarily involved with shaft positioning, counter control, and error indication.

The following categories of experimental control are available:

(A) *Local control.* A switch on each interface rack switches the experiment from computer to local control. In local control, the motors can be driven and counters can be stopped and started using switches on the control panel. Three banks of "nixie" tubes can display digital information; two of these display neutron or x-ray counts, while the other is computer-programmable and typically displays a motor position. A selector switch allows display of up to eight different variables under program control.

(B) *Typewriter control.* In this semi-automatic mode of operation, the user can talk to his experiment via a small number (about two dozen) of mnemonic

```
CCC2
          TIME:  02533  SEC
TYS

          STATION NUMBER  2  STATUS
          TIME:  02543  SEC
          UNDER COMPUTER CONTROL
          COUNTERS ON
          MONITOR              000
          DETECTOR             000143
          BREAKPOINT           A SET     B  SET     C  SET     D  SET
          PROGRAM    INACTIVE
          MOTOR POSITIONS    0 OFF     1 OFF     2 OFF     3 OFF
                           -031.25   +025.10   +015.37   -077.64


          DONE

STA 2 + 01032
          DONE

TYH
          ST 2
          MOTOR POSITIONS    0 OFF     1 OFF     2 ON      3 OFF
                           -031.25   +025.10   +014.05   -077.64


          DONE

RSC
          DONE

STC 0 200000
          DONE

LDP 0000AXDA
**        PT ERROR
          TIME:  02763  SEC


CCC2
          TIME:  02839  SEC
LDP0000AXDA
          DONE

LDP1222VECT
          DONE

TYW1034
            00000001
          DONE

ALW1034  12345671A
          DONE

TYW1034
            12345671
          DONE

ALW1034  00000001A
          DONE

TYW1034
            00000001
          DONE

STP0001
          DONE

$
```

```
CCC2
          TIME:  03058  SEC
TYS

          STATION NUMBER  2  STATUS
          TIME:  03066  SEC
          UNDER COMPUTER CONTROL
          COUNTERS ON
          MONITOR              000
          DETECTOR             000008
          BREAKPOINT           A SET     B  SET     C  SET     D  SET
          PROGRAM ACTIVE
          MOTOR POSITIONS    0 OFF     1 OFF     2 OFF     3 OFF
                           -031.25   +025.10   +010.32   -077.64


          DONE

HLP
          DONE

LDD 1235
**        PT ERROR
          TIME:  03170  SEC


CCC2
          TIME:  03181  SEC
LDI 1236
**        PT ERROR
          TIME:  03201  SEC


RSP3
          WRONG IDEN
CCC2
          TIME:  03232  SEC
RSP
          DONE
```

FIG. 5. Sample typewriter control sequence for the Brookhaven MSCS, as discussed in the text. The letters CCC2 identify the experimenter (CCC) and the number of the experiment he wishes to control (2). The user has typed in the instructions on the left. The control program has typed out the information that is tabulated to the right.

instructions which are available in the system through use of the typewriter. (Logically, the typewriter is another station in the distributor sequencing.) This operational mode has contributed greatly in making the scientist feel intimately involved with his experimental equipment. The typewriter commands available are those that might be given to a competent technician. Figure 5 includes some typical scientist-experiment dialog from the typewriter pages. Some of the commands available include the following:

TYS
: TYpe the Status of the experiment, including information regarding motor positions, counter contents and program activity.

STA 2 + 02037
: SeT Angle 2 to + 20.37 degrees.

TYM
: TYpe out the Motor positions only.

RSC
: ReSet Counters

STC 0 200000
: STart the Counters and continue counting until a monitor counter or clock has reached 200000 counts with 0 prescales of 10.

LDP 0000 AXDA
: LoaD the Program called AXDA into location 0000 relative to the beginning of my user's area. (A paper tape containing the program must have been placed in the paper-tape reader; otherwise a paper-tape error results.)

TYW 1034
: TYpe the Word in my location 1034.

ALW 1034 12345671A
: ALter the Word in my location 1034 to the value 12345671. (A indicates that the numbers in the address portion of this word are to be interpreted as absolute rather than relative.)

STP 0001
: STart Program in my location 0001.

HLP
: HaLt Program

LDD 1235
: Load a tape with Data in Decimal format into location 1235.

LDI 1236
: Load a tape with Data in Integer format into location 1236.

(C) *Automatic Control* by computer program. If a program has been loaded with the typewriter command LDP and initiated with the command STP, the experimental sequence will be controlled by the user's program which was on the loaded paper tape. This category of control is used for all routine data collection. The typewriter control program may be used simultaneously to ask for information but not to move motors or otherwise control the experiment.

The data from all the experiments, after temporary buffering in the computer core, is output onto a single magnetic tape. This tape is regularly processed at the laboratory central computing facility to separate the various users' data, which they are then free to process as they will. Limited amounts of data can also be output on the typewriters and the CRT display device as well as on strip-chart recorders or digital printers at the experiment. These auxiliary outputs are used mainly in setting up experiments with the bulk of the data going onto magnetic tape.

Most of the users' programs have actually been written by scientists rather than by computer personnel. The programs are written in a symbolic assembly language which does however include many pseudo-operations. The assemblies are done off-line on a CDC 924 computer; a paper tape is produced which is ready to use with the system.

Our experience with this system over a two-year period puts us in a position to answer the following questions:

(1) *What is the most important feature of any computer-controlled experiment (ours in particular)?* The ease of carrying on a natural dialog between the non-computer-oriented scientist and his experiment.

(2) *What advantages does computer control have over special control devices designed for each experiment?* There is the capability of rapid feedback from experimental results to experimental design—on-line redefinition of the experiment. There is enormous flexibility in designing one-shot experiments for special purposes. Experimental parameters may be rapidly and conveniently changed.

(3) *What is the greatest defect in the present system?* Lack of a memory-protect feature. This has caused us little down time but has forced us to be more careful than we would like to be. Programs must be carefully debugged before putting onto the system. The off-line assembler on the CDC 924 and a 920 simulator which is being developed for the CDC 6600 are designed to alleviate this problem.

(4) *What recommendations would we make for future systems of this type?* Foolproof memory protection. Even better machine–man interaction through a use of a more extended users' language. More flexibility in output devices. A greatly expanded fast memory (by the use of magnetic disc-packs for example).

(5) *What are the merits of a shared-time computer over small individual computers for this type of experimental control?* Systems of both types obviously work. We have been happy with ours. The chief advantage of the large shared-time system is that it allows even the small experimenter access to a large program library and a higher-level language through which he can communicate with the machine. Any system is no better than its software. (Hardware bugs are less frequent.) Both types of system will undoubtedly continue to be developed according to need.

The most important fact for the experimenter and the laboratory administrator to bear in mind is that with any system the time, and hence the personnel costs, in software development will be relatively large. It is appropriate that they should be if the system is to be a really useful laboratory tool which frees the scientist from some of his labors rather than by adding to them.

Crystal Structure Determination and Refinement

The greatest impact of the digital computer in crystallography has not been in the area of data collection and processing but in the area of structure determination and refinement. As recently as fifteen years ago, the determination of an organic crystal structure of any complexity was an undertaking that required many months of undivided attention. Even then the amount of computation involved to make the most of the available data was so large that the results were often less than satisfactory by present-day standards. The problem may be divided into two: finding a trial structure and refining the parameters of this structure.

The solution of a crystal structure requires a solution of the crystallographic *phase problem.* The intensity of a Bragg reflection is proportional to the square of the structure factor—which is in general a complex number. The scattering density, which gives a picture of the crystal structure, is given by a Fourier series with the structure factors as coefficients

$$\rho(x, y, z) = \sum_{h,k,l} F(h, k, l) \exp\{2\pi i(hx + ky + lz)\}$$

The structure factor can be expressed as $|F|\, e^{i\alpha}$ where the magnitude $|F|$ is available from experiment but the phase α is not.

By applying criteria of chemical reasonableness and demanding a density that is everywhere positive, at least partial solutions to the phase problem exist. One can develop probability formula for the phases that are functions of the intensities; these encompass the so-called *direct methods* of crystal structure solution. Other methods involve interpretation of the Patterson function—a Fourier synthesis

with the $|F|^2$ values as coefficients. For complex structures, either approach demands an enormous number of arithmetic operations. It is only with the computing speeds that have become available in the last few years that it has proved possible to determine many difficult structures by direct methods and by computer interpretation of the Patterson function. [7]

For centrosymmetric structures, the phases are all either $+\pi$ or $-\pi$, and the phase problem becomes a problem of *sign* determination. One of the most popular methods for the solution of centrosymmetric structures has been the symbolic addition method of Karle and Karle [8] and the related relationships proposed by Sayre [9] and Zachariasen [10]. The basic concept is that for structure factors of large relative magnitude there is a high probability that the sign of $F(h_1 + h_2)$ is equal to the product of the signs of $F(h_1)$ and $F(h_2)$. Successful programs for application of these methods have been written by Bednowitz [11], Fleischer [12], and Long [13], among others. The solution of centrosymmetric equal-atom structures containing up to 100 atoms per molecule no longer seems to be much of a problem with these programs in hand.

The end result of the sign determining methods are at most a few probable sets of signs for the large structure factors. The most probable sets are used in the calculation of scattering density syntheses by Fourier summation. The one synthesis that makes chemical sense is used as a starting point for refinement. The calculation of Fourier syntheses is also an important step in the refinement of a trial structure, particularly in locating light atoms in the presence of heavy atoms whose positions are already known. A feeling for the order of magnitude of the time required for these calculations may be gained by considering that a recent calculation (on a CDC 6600 computer) required about eight minutes for evaluation of a Fourier series containing 350 terms at 143,000 points. A significant portion of this time was spent in input/output, particularly in writing a tape to plot the results on a x–y plotter. Fifteen years ago, such a calculation would have required many days of hard work on punched card machines and would probably have been attempted by a crystallographer no more than once a year. A few years prior to that it would have been impossible.

The impact of the computer in being able to rapidly carry out Fourier syntheses has been of primary importance in the solution of the several protein structures which have been reported in recent years [14]–[16]. In addition, the immense data reduction problems involved in comparing hundreds of thousands of reflections from heavy-metal substituted compounds to arrive at phases could not have been easily carried out before the advent of the modern computer.

For the refinement of structures containing fewer atoms than do proteins, the iterative application of a nonlinear least-squares technique has been the preferred method. For the refinement of a structure containing 250 parameters, a 250×250 matrix must be calculated and inverted. Each element in the matrix is a sum over

all reflections of products of derivatives of the structure factors with respect to the parameters being refined. Again large computing times are required. A typical refinement cycle now requires several minutes. Some years ago it would have been hours or days, and a few years prior to that would not have been attempted. It is worth noting that the complete refinement of our hypothetical sixty-atom problem with 540 parameters would exceed the fast memory of the largest computer available today just for storage of the matrix of the least-squares normal equations. The increasing availability of high-speed mass memory devices will certainly increase our ability to refine large structures; perhaps detailed least squares refinement of protein structures (which are handled now almost entirely by Fourier techniques) will soon become feasible.

Presentation of the Results

The items discussed in the previous sections of this paper—data acquisition and crystal structure solution and refinement—are areas primarily of concern to the professional crystallographer. More important and interesting are the chemical data that emerge from these techniques. How are the 540 parameters in our structure to be presented to the chemist in the most useful way? Here again the computer has come to our aid.

The chemist is interested in a crystal structure study because of the fact that it may illuminate some point of stereochemistry—perhaps involving only the topology or gross shape of a molecule, perhaps involving detailed consideration of accurate bond distances and angles.

With regard to the calculation of numerical values for bond lengths and angles, a fast computer again allows us to calculate many such interesting quantities taking complete account of the correlations between errors which are contained in the matrix of the least-squares normal equations. The assessment of reliable error estimates may be fully as important as determination of the best parameter values.

Most chemists feel most comfortable with a molecule when they can draw a picture of it, and the pictorial representation of a molecular structure may convey more information than a thousand words. Crystallographers and others interested in molecular structure have of course illustrated their papers with elegant drawings for many years. Complex molecules are however often difficult to visualize from a two-dimensional drawing, and the laboratory of any crystallographer is complete with three-dimensional models which better illustrate the stereochemical points to be made. The development of computer graphics, particularly the use of high resolution x–y plotters, has now made it possible to bring the three-dimensional model to the printed page by the production of accurate, detailed stereoscopic drawings. One of the major contributions to this area has been a computer pro-

gram written by C. K. Johnson of the Oak Ridge National Laboratory [17]. This versatile program builds a molecular model in the memory of the computer, rotates it, scales it, and writes a magnetic tape which will produce beautiful stereo

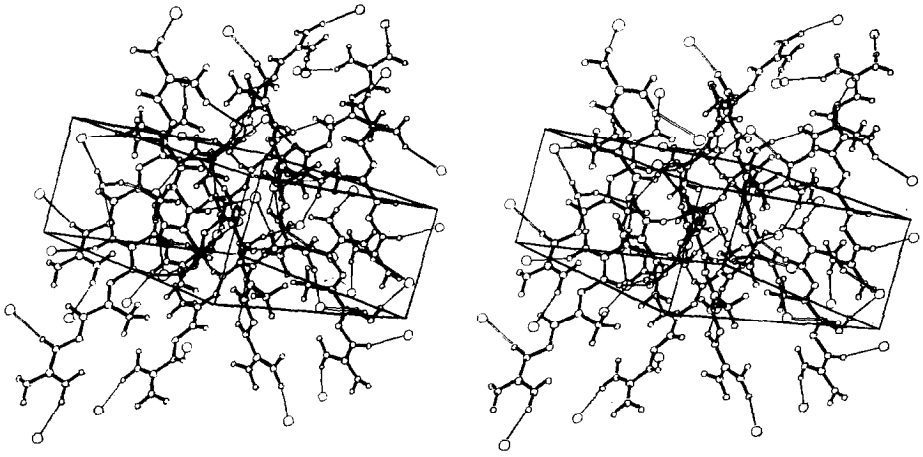

FIG. 6. Hydrogen bonding in methylglyoxalbisguanylhydrazone. The intricacies of the hydrogen-bonded network structure are much easier to see in three dimensions. (This is a stereoscopic pair.)

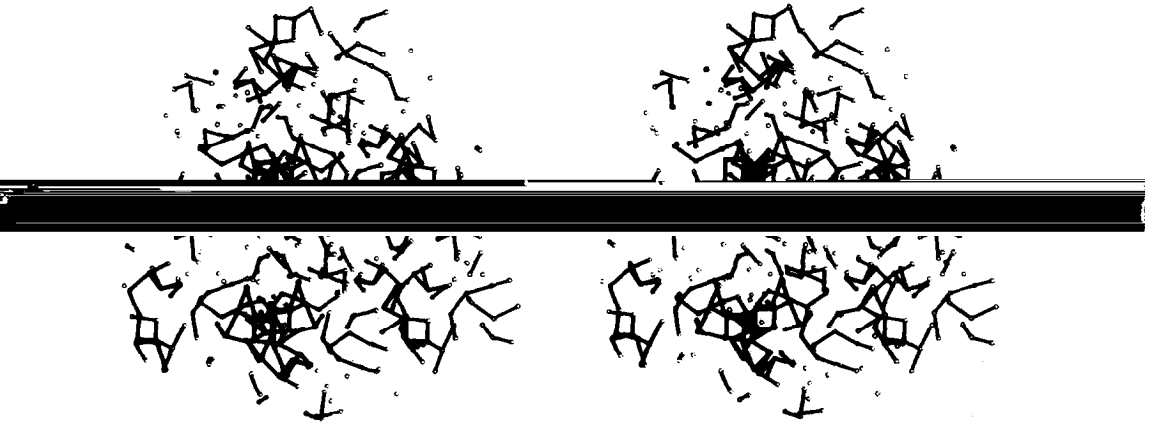

FIG. 7. Stereoscopic view of "structure" defined by the largest 60 peaks in an electron density map resulting from an incorrect assignment of phases. Bonds are drawn between all atoms less than 2.0 apart, and all atoms within 3.0 A of any other are drawn.

drawings on an x–y plotter. Some examples of these drawings are given in Figs. 6–9.[11]

In addition to producing drawings showing the final results of a crystal structure determination, the program has also proved useful to us in the early stages of such a study. In attempting to solve a structure by the symbolic addition procedure, we may calculate Fourier syntheses for several equally probable sets of phases. The Fourier program punches out cards containing positions of the highest peaks. These may be used directly as input to the structure-drawing program. One can view the resulting pictures stereoscopically and immediately ascertain whether the possible structure makes chemical sense. One such result is shown in Fig. 7. The presence of only small molecular fragments probably indicates a wrong choice of signs.

Among the results of a complete crystal structure determination are the six parameters describing the harmonic motion of each atom in the structure. The interesting information for each of the atoms includes the lengths and directions of the principal axes of the vibrational ellipsoids. Although such information can be presented in tabular form, again a picture tells all, and the computer graphic techniques can again present these ellipsoidal shapes in a way that would not be easily comprehended by study of a table of principal axes (Fig. 8).

Computer analysis of many accurate crystal-structure determinations in recent years has revealed that the motions of the individual atoms in a molecule in the

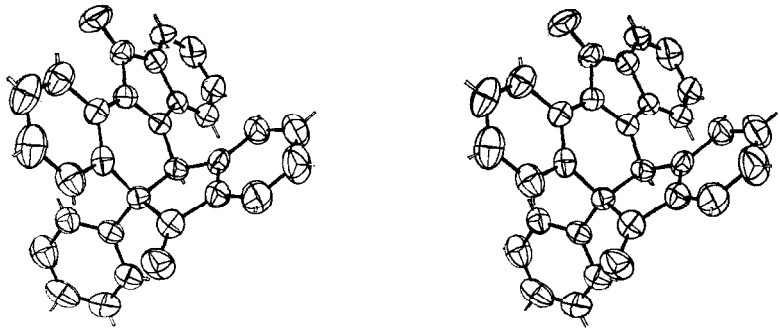

FIG. 8. The molecular structure of phenylcinnamalone. The principal axes of the ellipsoids are equal to the root-mean-square amplitudes of vibration in these directions. The larger amplitudes at the periphery of the molecule and the wagging motions of the quinone oxygens are especially obvious. The stereoscopic view allows dramatic presentation of the geometry of the molecule in three dimensions.

[11] These drawings may be viewed with a small hand-held stereo viewer. Most people can also obtain the stereo effect by looking at one picture with each eye and causing the images to merge; this process may be assisted by placing a sheet of cardboard between the two views.

solid are often mainly reflections of the overall motions of the molecule. The thermal ellipsoids of the methyl-group hydrogens in potassium dihydrogen aspirinate (Fig. 9), for example, clearly indicate that the motions are not random but that what one sees is a wide-amplitude hindered rotation of the methyl group. Similarly, entire molecules undergo motions as rigid bodies on which may be

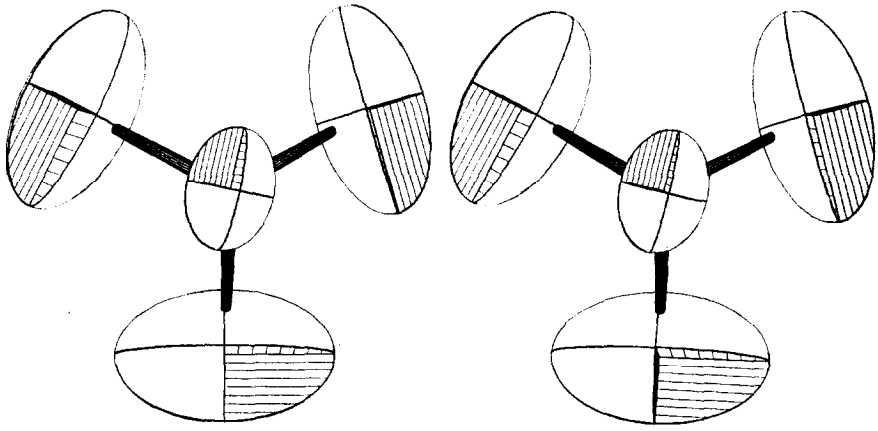

FIG. 9. The methyl group in potassium hydrogen diaspirinate. The amplitudes of thermal motion clearly indicate that the group is undergoing a hindered rotation.

superimposed internal vibrations. The rigid body motions in crystals at room temperature do, however, account for about 90% of the total atomic motions. The rigid-body motion of a molecule may be described in terms of 20 independent parameters [18]: three principal components of rigid-body translation; three parameters giving the directions of these rigid-body translations; three principal components of rigid-body libration; three parameters giving the directions of the libration axes; six parameters specifying the displacement of the libration axes from an origin (they need not intersect in the molecular center of mass); and two independent parameters specifying the correlation between translation and libration to give a screw motion. In a molecule with thirty atoms, for example, we may attempt to explain the $6 \times 30 = 180$ vibrational tensor components for the individual atoms by obtaining the best possible twenty rigid-body motion parameters. Least-squares programs for carrying out this fit on a computer are available [18] and lead, in a very short time (a few seconds of computing time on a CDC 6600), to a determination of the rigid-body components which can afford a most concise picture of the largest contribution to the molecular motions. The results of such an analysis for the molecule illustrated in Fig. 8 are given in Fig. 10. The next logical step in a complete description is to superimpose

internal rotations of groups on the rigid-body motions and to carry out a least-squares refinement of a larger parameter set. Such calculations have been carried out for a number of molecules by Johnson [19].

The end aim of these computations is a molecular description—both static and dynamic—which will be quickly comprehended by the chemist or chemical physicist interested in a particular molecule.

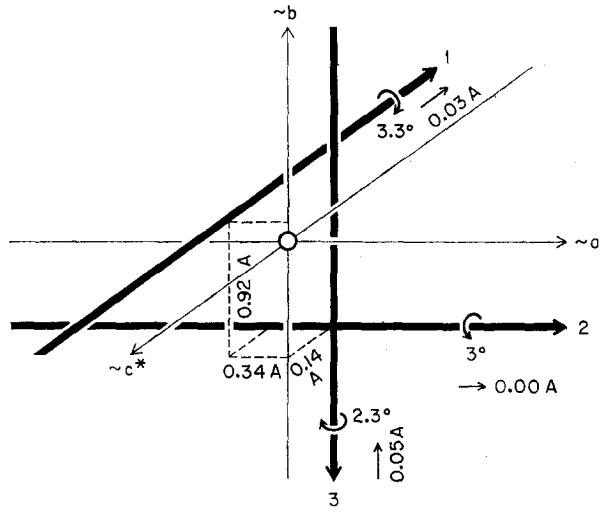

FIG. 10. Results of rigid-body motional analysis of phenylcinnamalone (Figure 9). The motion can be described as a superposition of independent translational motions (not shown) which are isotropic with root-mean-square amplitude 0.20 A and three screw librations about three-nonintersecting axes. The root-mean-square amplitudes of libration and the associated translations giving rise to the screw motion are indicated in the figure, as are the locations of the screw-rotation axes with respect to the center-of-mass (O) of the molecule. Axes 2 and 3 nearly intersect, but 1 is significantly displaced from the intersection of 2 and 3; it does however lie closer to the center of mass. The three screw-libration axes are approximately (and accidentally) parallel to the unit cell axes, a, b, and c*.

LITERATURE AND DATA RETRIEVAL

An isolated crystal or molecular structure may be of great interest. A comparison with other structures may be more interesting. Now that the techniques of data collection and crystal structure refinement have become greatly speeded up, it becomes easily possible for a chemist to say, "I will do twelve crystal structures that shed some light on the problem of interest to me at the moment." Of course

he will be far happier if someone else has done the structures and if he can find all the information he needs in the literature. Clearly, literature-searching techniques are as important in crystallography as in any other field, and computers are becoming very important in literature searching. We are all enthusiastic about the new computer-based abstracting and searching aids that Chemical Abstracts Service in planning for the future; the searching for keywords in abstracts and titles is a very useful technique.

Crystallographers often have to go a step further than just literature retrieval; it is often necessary for them to search for and correlate numerical data. Again crystallographers have had long experience in this area. One of the oldest and most-used data retrieval systems is the ASTM file of powder diffraction patterns which has allowed the analytical chemist to rapidly identify unknown specimens from their powder diffraction patterns. In a related area, information on unit cell constants for all crystals which have been studied have been printed in two editions in the very useful compendium "Crystal Data" [20], [21]. The information in "Crystal Data" is now being placed on magnetic tape so that it may be rapidly updated as well as searched.

These two examples—the ASTM file and "Crystal Data"—are again of more interest to the professional crystallographer than to the average chemist. More pertinent to the problem posed by the average chemist who wishes to avoid doing twelve crystal structures, even though it is now within his capability to do so, is a project which is presently underway at Cambridge University under the direction of Dr. Olga Kennard. The crystal data and the atomic coordinates of all organic crystal structures which have been done since 1960 are being put onto magnetic tape. A software system is being developed that will allow one to search the file for substructures and to compute bond distances and angles of interest. One may for example ask for the P—O distances in all organic phosphorus compounds in which the phosphorus is tetrahedrally bonded. Such systems are of great importance in making rapidly available the information which may be buried in many papers in many sources—and where the calculation of interest may not have been performed by the author. With appropriate financial support, we may expect to see such systems greatly expanded, not only in crystallography, but in all of chemistry.

The better we are able to find and correlate existing information, the more rapidly we are able to pose more interesting problems to return full circle to the point of doing more experiments.

PROGNOSIS

One may ask whether we are likely to soon reach the day when the chemist may toss a crystal in one end of a machine and have a typewritten description and

pictorial representation of the structure come out the other end. Probably not. It is true, however, that the increased power of computers and the increasingly sophisticated uses of them will allow the crystal-structure determination of many complex compounds to become a reasonably routine laboratory procedure. A load is thus lifted from the crystallographer's back, hopefully allowing him more time to think about interesting chemistry and to decide which are the most important structures to solve in the coming year.

A final word of warning is perhaps not out of place. When an experimental technique becomes routine, great care must be taken to see that it does not degenerate into a sloppy experiment. A shiny diffractometer does not produce good data without care, and a number printed out by a computer is not necessarily meaningful—despite the fact that ten million iterations have been used to produce it. In fact, the opposite is likely to be the case.

Finally it is sobering to note the following lines from one of the greatest contributors to structural chemistry

> "This method [*qualitative intensity comparison*] permitted one or two parameters to be determined rather easily, and more with considerable difficulty... My luck was not so good... The space group [of $K_2Ni_2(SO_4)_3$] is T^4. This space group has positions for four equivalent atoms, with one parameter, or twelve equivalent atoms, with three parameters. Hence there is one parameter for every four atoms in the unit cube. There are 76 atoms in the unit cube, and the structure is accordingly determined by 19 parameters. I at once decided to work on another crystal." (Ref. [22].)

We have come a long way in technique in fifty years but perhaps not so much further in understanding.

ACKNOWLEDGMENTS

The MSCS data collection system was developed by Dr. Robert Spinrad of the Brookhaven National Laboratory Instrumentation Division in collaboration with an able group of co-workers in that Division and in the Brookhaven Applied Mathematics Department. The users' programs for the collection of diffraction data were written almost single-handedly by Dr. Allan Bednowitz of the Brookhaven Chemistry Department. Since the system has gone into operation, able assistance has also been provided by David Hartmann and Bruce Martin. In part of this paper, I have taken the liberty of drawing on the remarks of Dr. William Busing of Oak Ridge made in his presentation at the Conference on Computers in Chemistry.

References

1. S. C. Abrahams, *Rev. Sci. Instr.* **33**, 973 (1962).
2. J. Tillinger (unpublished).
3. Hilger-Watts Linear Diffractometer; U. W. Arndt and D. C. Phillips, *Acta Cryst.* **14**, 807 (1961).
4. H. Cole, Y. Okaya, and F. W. Chambers, *Rev. Sci. Instr.* **34**, 872 (1963).
5. W. R. Busing, R. D. Ellison, C. K. Johnson, and H. A. Levy, *Abstr. Am. Cryst. Assoc.* D7 (1965). Suffern, New York.
6. D. R. Beaucage, M. A. Kelley, D. Ophir, S. Rankowitz, and R. J. Spinrad, *Nucl. Instr. Methods* **40**, 26 (1966).
7. *Trans. Am. Cryst. Assoc.* **2**, 1–78 (1966).
8. I. L. Karle and J. Karle, *Acta Cryst.* **17**, 1356 (1964).
9. D. Sayre, *Acta Cryst.* **5**, 60 (1952).
10. W. H. Zachariasen, *Acta Cryst.* **5**, 68 (1952).
11. A. L. Bednowitz and B. Post, *Acta Cryst.* **21**, 566 (1966).
12. E. Fleischer (private communication).
13. R. E. Long, Unpublished doctoral dissertation, University of California, Los Angeles, California, 1965.
14. G. Bodo, H. M. Dintzis, J. C. Kendrew, and H. W. Wyckoff, *Proc. Roy. Soc. (London)* **A253**, 70 (1959).
15. C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Nature* **206**, 757 (1965).
16. G. Kartha, J. Bello, and D. Harker, *Nature* **213**, 862 (1967).
17. C. K. Johnson, "A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations." ORNL-3794, revised 1965.
18. V. Schomaker and K. Trueblood, *Acta Cryst.* **B24**, 63 (1968).
19. C. K. Johnson, as quoted by W. Busing in a talk at this conference.
20. J. D. H. Donnay and W. Nowacki "Crystal Data," Geological Society of America, Washington, D. C. (1954).
21. J. D. H. Donnay (Ed.), "Crystal Data," ACA Monograph Number 5. American Crystallographic Association, Washington, D. C. (1963).
22. Linus Pauling in "Fifty Years of X-ray Diffraction" (P. P. Ewald, Ed.). International Union of Crystallography, Utrecht (1962).