

# The Use of Punched Cards in Molecular Structure Determinations

## I. Crystal Structure Calculations

P. A. SHAFFER, JR., VERNER SCHOMAKER, AND LINUS PAULING

*Gates and Crellin Laboratories,\* California Institute of Technology, Pasadena, California*

(Received July 22, 1946)

The application of punched cards and electrical accounting machines (of the International Business Machines Corporation) to crystal structure calculations is described. The machines may be used to evaluate Fourier series with greater accuracy, greater speed, and greater density of calculated points than is possible by the methods used heretofore. A method for the punched-card calculation of structure factors is also set forth, and the use of the machines for the reduction of observational equations to normal equations in least squares is discussed.

### INTRODUCTION

THE extensive applications of Fourier methods of crystal structure analysis in recent years, including the representation both of electron-distribution functions<sup>1</sup> and of vector-distribution functions,<sup>2</sup> has emphasized the need of rapid and accurate means of making Fourier projections and carrying out related calculations. The increased complexity of structures now being investigated requires that many calculations of the Fourier type and many structure factor computations be made during a single structure determination. These two types of calculation are the most laborious ones in this field of work, and the time which they require largely determines the period of an investigation. The efficiency and precision with which such calculations may be made is, therefore, a factor of considerable importance in determining the success of a structural investigation. We have found that these calculations and other calculations involved in the processes of determining the structure of molecules and crystals can be carried out very easily and effectively by the use of punched-card methods, with the machines of the International Business Machine Corporation.

\* Contribution No. 1019 from Gates and Crellin Laboratories.

<sup>1</sup>W. H. Bragg, *Trans. Roy. Soc. (London)* **215**, 253 (1915); W. Duane, *Proc. Nat. Acad. Sci.* **11**, 489 (1925); R. J. Havinghurst, *Proc. Nat. Acad. Sci.* **11**, 502, 507 (1925); A. H. Compton, *X-rays and Electrons* (Van Nostrand Company, Inc., New York, 1926); W. L. Bragg, *Proc. Roy. Soc. (London)* **A123**, 537 (1929).

<sup>2</sup>A. L. Patterson, *Phys. Rev.* **46**, 372 (1934), *Zeits. f. Krist.* **90**, 517, 543 (1935); D. Harker, *J. Chem. Phys.* **4**, 381 (1934).

### EVALUATION OF FOURIER SERIES

The Fourier calculations made in this work are the evaluation of one, two, or three-dimensional Fourier series; the one-dimensional case is represented by

$$f(x) = \sum_{h=0}^{\infty} (A_h \cos 2\pi hx + B_h \sin 2\pi hx). \quad (1)$$

The series is evaluated at the points  $x = n\Delta x$ , where  $\Delta x$  is small enough to give the desired detail in the complete curve of the function. Before the application of punched-card electrical accounting machines to these problems, series were most easily evaluated by use of an adding machine and the Beevers-Lipson Strips.<sup>3</sup> These are cardboard strips on which are printed values of  $A \sin 2\pi hx$  and  $A \cos 2\pi hx$  with  $A = \pm 1, 2, 3, 4, \dots, 100$ ,  $h = 0, 1, 2, 3, \dots, 20$ , and  $x = 0$  to  $\frac{1}{4}$ ,  $\Delta x = 1/60$ . The numerical values are given to  $\pm \frac{1}{2}$  in the unit position. The strip corresponding to each term of the series is taken from a file and the strips are then arranged to permit the convenient addition of terms for constant  $x$  from all strips.

The use of the Beevers-Lipson strips greatly facilitates the evaluation of Fourier series, but the method has several unsatisfactory features. First, there is no check on the terms used or on the operations performed; the wrong strips may be drawn or errors may be made in summing them. Second, the method requires for long

<sup>3</sup>H. Lipson and C. A. Beevers, *Proc. Phys. Soc. (London)* **48**, 772 (1936). An electrical machine for the summation of Fourier series has been described by C. A. Beevers, *Proc. Phys. Soc. (London)* **51**, 660 (1939), and D. Macewan and C. A. Beevers, *J. Sci. Inst.* **19**, 150 (1942).

periods of time the constant attention of the operator putting numbers into the adding machine and recording the results; the fatigue due to these operations contributes greatly to the probability of error. Third, the tabulated functions are not given to a number of places sufficient for all present-day needs. Fourth, the interval  $\Delta x$  is too large and the range of  $h$  too small for the resolution of structural features in large unit cells. Further, the extension of the computed range 0 to  $\frac{1}{4}$  of the argument  $x$  to a larger one can be made only by summing the odd and even sines and cosines separately and combining the four terms with proper changes of sign; the method of doubling or quadrupling the frequencies of the terms cannot be used because the multiplied frequencies soon exceed the range of the tables. Because of these difficulties we have sought during the last few years to devise a mechanical means of carrying out Fourier calculations in a precise and efficient manner.

A machine for making the calculations for crystal structure work was first designed. While the construction of this machine was under consideration, Dr. W. J. Eckert suggested that the punched-card machines of the International Business Machines Corporation might be used for evaluating Fourier series, and proposed a method of using these machines. We then worked out somewhat different methods of using the IBM machines for evaluating not only the Fourier series required for crystal structure investigations but also the Fourier sums and integrals occurring in electron diffraction work; the latter calculations are discussed in the following paper. Other important uses of the machines in molecular structure work have also been found.

### The Machines Used in the Calculations

The machines used in punched-card calculations have been described in detail elsewhere.<sup>4</sup> Specially prepared cards of insulating paper are passed under several sets of reading brushes in

the machine. When holes punched in a card come into position, the brushes are permitted to touch an electrified roller beneath the card. The card feed is synchronized with the other operating units of the machine and the timing of the impulses picked up by the reading brushes is thus given numerical significance. The brushes are connected to plug positions in a removable plugboard. Almost all of the electrical circuits of the tabulating machine have terminals in the plugboard, and these circuits may be plugged together to cause the machine to perform in a variety of ways. For instance, number impulses read by the card brushes can be transmitted to counters, through the counter entrance hubs, and the counter exit hubs can be plugged to print bar hubs in order to print totals of numbers represented by the punches in the cards. In this way holes punched in the cards activate sorting, printing, and accumulating mechanisms as desired. In the calculations at hand only three pieces of machine equipment are used. These are an electric key punch (type 11), an electric sorting machine (type 80), and an alphabetic direct-subtraction tabulating machine (type 405). The punch is used to replace damaged cards and to punch special cards used in the arrangement of the printed result of the calculation. The sorter is used for arranging the cards in order before filing or machine tabulation. The tabulator accumulates totals of numbers punched on the cards, prints the results, and then continues with the next step of the calculation, thus performing the duties of an adding machine and its operator at the same time.

The alphabetic tabulator used is a standard machine except for the addition of seven relays, besides the five which are standard equipment, for the switching of counters from addition to subtraction; twelve of these relays, called "X-distributors," are needed for the controlled addition and subtraction of numbers in twelve independent six-place counters. With modification of the card files to permit addition of complements rather than direct subtraction, the standard machine could be used to give identical results; but the standard machine without modification cannot be used with our present card files.

The character of the calculations requires that essentially all of the circuits of the machine be

<sup>4</sup>W. J. Eckert, *Punched Card Methods in Scientific Computation* (Thomas J. Watson Astronomical Computing Bureau, January 1940); *Practical Applications of the Punched Card Method in Colleges and Universities* edited by G. W. Baehne (Columbia University Press, New York, 1935); *Machine Methods of Accounting* (International Business Machines Corporation, 1936); Harry P. Hartke-meier, *Principles of Punched Card Machine Operation* (Thomas Y. Cromwell Company, New York, 1942).

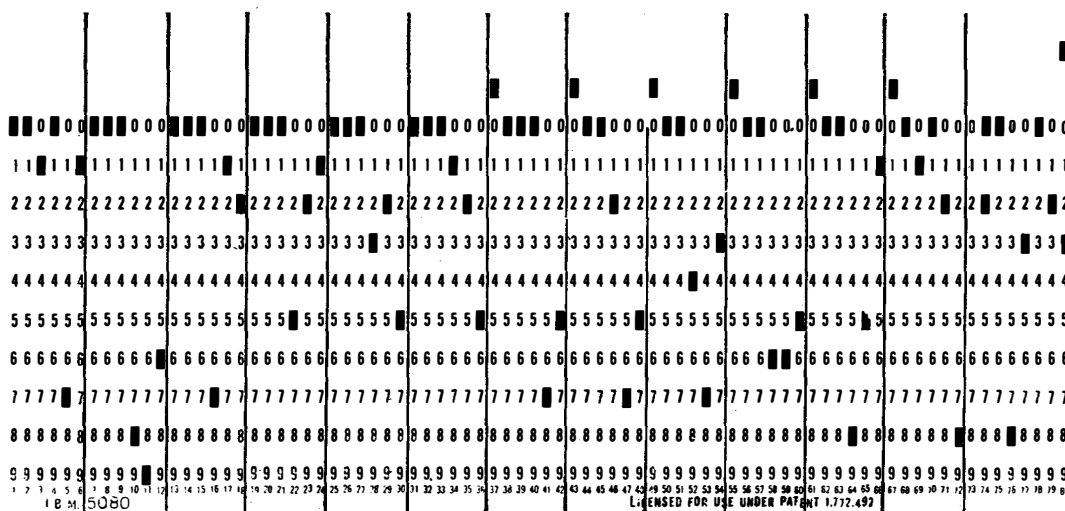


FIG. 1. A typical punched card. The vertical lines indicate the twelve function fields and the classification field for the card design described in the text.

used. The "automatic plugboard," a board which may be removed, completely wired, from the machine and replaced by another, is almost completely plugged, and several toggle switches have been attached to the cover of the board to increase the flexibility of the wiring arrangement. All of the problems to be discussed in the papers of this series with the exception of the least-squares calculations and tabulation of "reflection cards" are performed with use of one plugboard. The external toggle switches permit also the running of electron-diffraction calculations and verification of the files with the same board. A second board is kept wired for least-squares use, and a third is used for the tabulation of "reflection cards" and for incidental calculations of a temporary or experimental nature. The wiring of the plugboard and extra toggle switches for the electron diffraction and crystal structure calculations is described in a set of supplementary notes which may be obtained from the Gates and Crellin Laboratories on request. Additional details of card design and procedure are also described in these notes.

#### Arrangement of the Punched-Card Tables

Four equivalent files, each of about 20,000 cards, have been prepared to represent the functions  $A \sin 2\pi hx$  and  $A \cos 2\pi hx$ . The argument  $x$  runs from 0.000 to 0.250 cycle with a tabular

interval of 0.002, the frequency  $h$  has the values 0, 1, 2,  $\dots$ , 30, and the amplitude values  $A$  are  $\pm 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 500$ .

The eighty-column tabulating card (see Fig. 1) is divided for this use into twelve six-column fields for punches which represent the numerical values of the tabulated functions for twelve values of the argument, and an eight-column field for punches which identify the card and control the sorting and tabulating. Eleven of these cards, which we call "function cards," are required for the representation of the 126 tabulated values of the function from zero to one-quarter cycle. The six empty fields are symmetrically placed—two on the left of the first card, two in the middle of the sixth card, and two at the end of the eleventh card.

As an example, the values of the function punched in successive cards of the packet  $100 \cos 2\pi x$  are shown in the upper half of Fig. 2. The corresponding values of the argument are listed in a similar array at the bottom of the figure. The number 100 listed at (A) indicates that the amplitude of the function represented is 100; the (C) on the extreme left of the same row shows that the function is a cosine. The frequency of the function is determined by the position of the amplitude entry in the table. The serial order of the cards is printed in the column indicated by (B). The numbers printed in row 5 are the figures

contained in the card of serial number 5, etc. The card-count indicates how many cards have been summed to give the printed totals; the count is printed in the column marked by (C). Since only a single packet has been tabulated in this example, the card count is 1 throughout the table.

In addition to the eleven function cards there is an index card for each packet. The index cards have a filing-tab on the upper left-hand corner on which are printed the amplitude, frequency, and phase of the function represented by the packet. The amplitude of the packet is also punched in a numerical field of the index card which corresponds to the units digit of the frequency. The purpose of the index cards is to permit easy identification and separation of the packets in the files and to print, during tabulation, a record of the amplitudes and frequencies of the terms. Index cards are recognized by an "x-punch" in card-column 72. This punch activates relays during the tabulation of the index cards and brings about the changes of wiring necessary for the special tabulation of the data punched in them.

The packet labeled  $20 \cos 16\pi x$ , for example, thus consists of twelve cards: the index card (serial number 00) and eleven function cards (serial numbers 01 through 11);  $x$  runs across the cards, 0.000 through 0.018 on the first function card, 0.020 through 0.042 on the second, and so on up to 0.250. When negative values of the function occur, they are indicated by an  $x$ -punch in the first column of the field in which the value is punched. Examples of this punching appear in the seventh and following fields in Fig. 1. The eight-place classification field of the cards of the packet are the same except for the serial number (columns 78, 79). The remaining six columns of

this field give the sign (74), frequency (75, 76), amplitude code (77), phase (80), and file number (74). The filing arrangement in order of decreasing detail is amplitude, sign, frequency, and phase (sine and cosine).

### Calculation Procedure

In carrying out a calculation the operator first pulls from the files the packets corresponding to the terms of the series; a term  $+26C12$  ( $+26 \cos 2\pi 12x$ ) would require the packets  $+30C12$  and  $-4C12$ , or  $+20C12$  and  $+5C12$  and  $+1C12$ . The drawing of the required amplitudes is facilitated by the amplitude color-coding of the cards; all the functions of a given amplitude are punched into cards of the same color. The cards drawn from the file are then sorted on the columns containing the serial punches. Special cards for controlling the form of the report are introduced automatically during the sorting operation, and title cards are placed on the top of the sorted stack. The entire set of cards, sometimes as many as 8000, is fed continuously into the hopper of the tabulator. The series is summed and the results, together with all data necessary for assuring the correctness of the calculation, are printed by the tabulator in a continuous and automatic operation, at a rate of about 140 cards per minute for typical summations. Several calculations may be tabulated in a single running of the machine, since the card-activated control mechanisms of the tabulator permit spacing, printing, and clearing of totals, or listing of numerical or alphabetic information upon the occurrence of punches or changes in the sequence of punches specified by the wiring of the tabulator plugboard. After a calculation or group of calculations has been tabulated, the cards are

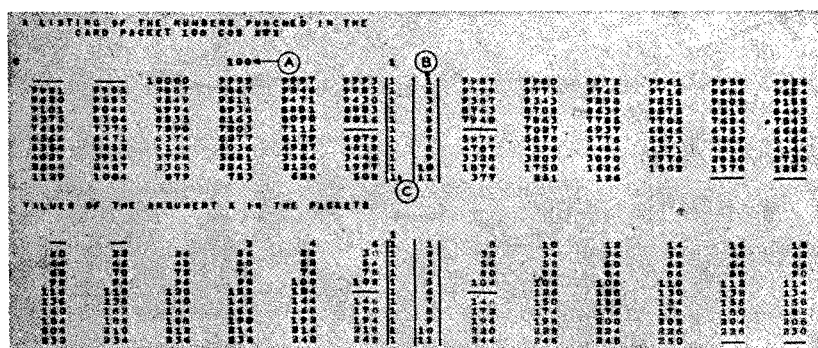


FIG. 2. Tabulator records illustrating the card design.

FORWARD PROJECTION ON THE CORE FACE  
OF THE LATTICE DENSITY OF  
A CRYSTAL STRUCTURE  
DATA OF WILL SHINN HERRMANN AND PETERS  
FOUR-DIMENSIONAL SUMMATION  
ALL 100 PER. CORRECT  
1 000

| C | 000 | 010    | 020    | 030    | 040    | 050    | 060    | 070    | 080    | 090    | 100    | 110    | 120    | 130    | 140    | 150    | 160    | 170    | 180    | 190    |
|---|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |

1 180

| C | 000 | 010    | 020    | 030    | 040    | 050    | 060    | 070    | 080    | 090    | 100    | 110    | 120    | 130    | 140    | 150    | 160    | 170    | 180    | 190    |
|---|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| S | 139 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |
| C | 13  | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 | 000000 |

FIG. 3. Two-dimensional Fourier series. Sample second-step summations for two values of the parameter  $z$ .

gathered into the original packets by machine sorting on the columns containing the amplitude, sign, frequency, phase, and file number. The operator then restores the cards to their proper places in the files.

### One-Dimensional Series

The printed results of two one-dimensional series calculations are shown in Fig. 3. At the top of the form are the title of the calculation and a description of special aspects of the problem. Following the title field is the index field (printed from the punches contained in index cards), which contains the amplitudes and frequencies of the terms summed. The amplitudes are printed explicitly (for example the number 537 indicated by the arrow at (A) represents the term  $537 \cos 2\pi 2x$ ) and the frequencies are indicated by the position of the numbers in the field. The unit of frequency is given by the number of the column in which the amplitude is printed (numbering column 3 "0," column 12 "9") and the ten of frequency by the number of the row ("00," "10," "20" at the arrow marked (C)); sines and cosines are printed in separate rows labeled (S) and (C) on the extreme left of the figure. Thus the figure 85 in column "6" and row 10-C of the first calculation indicates that a term  $85C16$  has been included in the summation. Beneath the index field are the eleven rows of totals which give the results of the calculation. Each row of the report is automatically designated by the corresponding serial number (D) of the cards to prevent accidental interchange of rows by the

operator. A card-count (B) which gives the units place of the number of cards tabulated in each row is printed also; the card-count is essential for the detection of extra or missing cards. (The card-count is one less than the true count because the card count is accumulated in the extreme left-hand position of, the eight-place counter used for the sixth field.)

When the calculation has been completed the index field is checked against the original data sheet and the card count is examined for deviations. If there is no indication of error the calculation is presumed to be correct. The only possible errors not checked in this manner are those arising from extreme coincidence or machine error. Machine errors may be brought to negligible incidence by running test cards prior to the calculation. The index-field tabulation and the card count make it possible for the calculation to be carried out by a person not acquainted with the significance of the calculation; the crystal structure investigator can check the correctness of the results in a few minutes at any time thereafter.

The ability of the tabulator to detect changes in the sequence of cards and to perform a variety of special operations when activated by specially punched cards is the principal factor in establishing the high efficiency of the method; this coupled with the high running speed of the machine and its capacity to carry out twelve summations at once permits the reduction of operating time and the great extension of the number of calculated points.

**Two-Dimensional Series**

The automatic control feature of the tabulator is of great convenience in extended calculations; as an example two-dimensional summations will now be considered. It is possible to reduce a two-dimensional Fourier series to a product of one-dimensional series.<sup>5</sup> For example the function

$$f(x, y) = \sum_{k=0}^{\infty} \left( \sum_{h=0}^{\infty} A_{hk} \cos 2\pi hx \right) \cos 2\pi ky - \sum_{k=0}^{\infty} \left( \sum_{h=0}^{\infty} B_{hk} \sin 2\pi hx \right) \sin 2\pi ky, \quad (2)$$

which is symmetric about the origin, may be rewritten as

$$f(x, y) = \sum_{k'=0}^{\infty} C(k', x) \cos 2\pi k'y - \sum_{k'=0}^{\infty} S(k', x) \sin 2\pi k'y, \quad (3)$$

with

$$C(k', x) = \sum_{h=0}^{\infty} A_{hk'} \cos 2\pi hx$$

and

$$S(k', x) = \sum_{h=0}^{\infty} B_{hk'} \sin 2\pi hx. \quad (4)$$

The first step in the calculation is to determine  $C(k', x)$  and  $S(k', x)$  for  $k'=0, 1, 2, \dots$ . To determine  $f(x', y)$  one carries out the summation indicated above for  $x=x'$ . The result is a Fourier function of  $y$  only, for given  $x'$ ; this summation is then carried out, giving the function  $f(x', y)$ , which is a section across the function  $f(x, y)$ . The contour map for  $f(x, y)$  is developed from several sections (up to as many as 500 per cycle in an extreme case) calculated in this manner.

The first step of the machine calculation might be the tabulation of the many one-dimensional series separately, as it is carried out by other methods; the values of the  $C$ 's and  $S$ 's would then be picked out of the individual series. The tabulator is able to carry out the first step in one operation, however, and to arrange the  $C$ 's and  $S$ 's of given  $x$  in a column, listing with each item the frequency to be associated with it in subsequent summations. Thus to obtain the slice

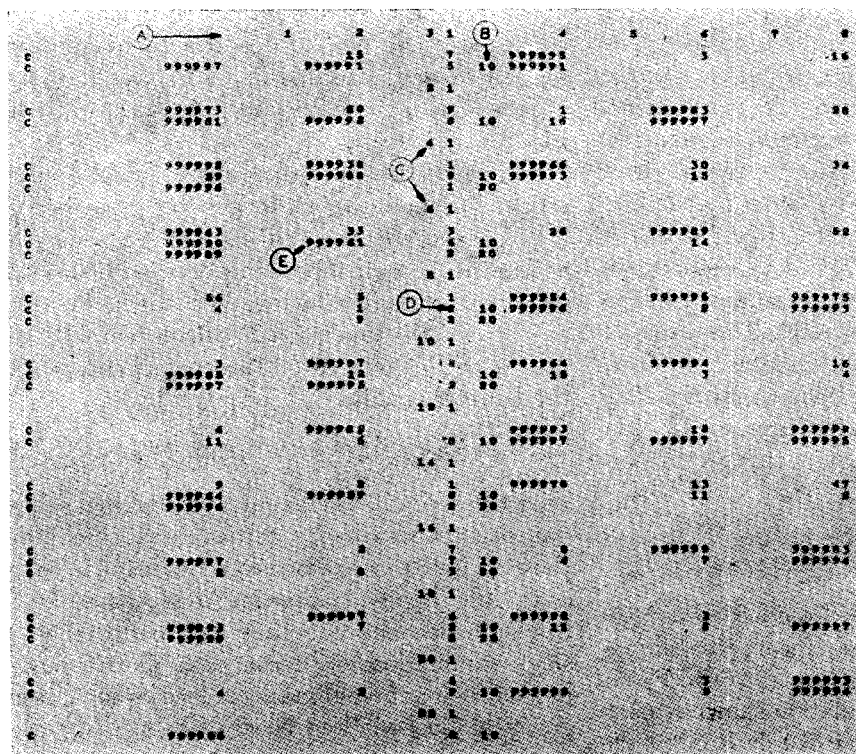


FIG. 4. Two-dimensional Fourier series. Sample tabulation of index cards for part of a first-step summation. (Column 9 is not included in this figure because it contains no entries. This column did appear in the original report, however.)

<sup>5</sup> C. A. Beevers and H. Lipson, *Phil. Mag.* 17, 855 (1934).

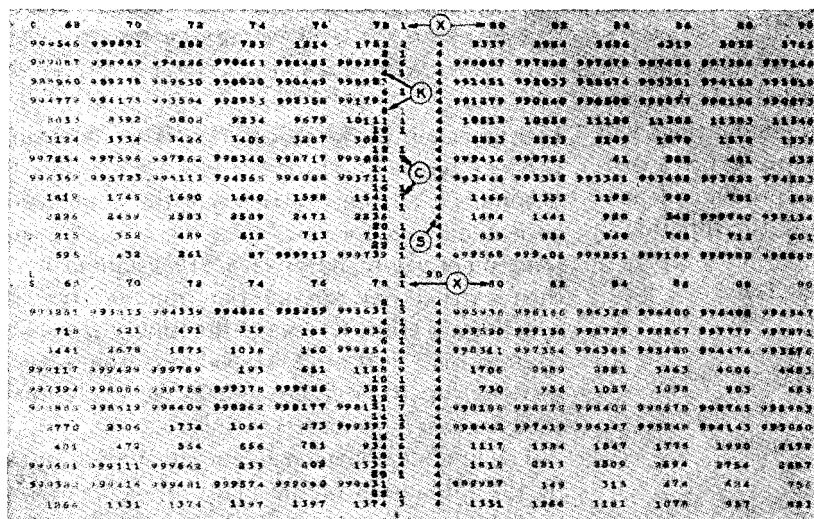


FIG. 5. Two-dimensional Fourier series. Sample first-step summation giving the coefficients of the sine and cosine terms of the second summation for  $z=0.068, 0.070, \dots, 0.090$ .

$f(0.120, y)$ , the operator uses the terms listed in order in a single column labeled 0.120.

Portions of a first-dimensional step of a two-dimensional Fourier projection are shown in Figs. 4 and 5. Figure 4 is a reproduction of a part of the index field tabulation. The C's at the extreme left indicate that the terms listed are cosine terms. In the center of the figure are three columns of one-place or two-place figures. The first of these (C) gives the index  $k$  for each of the tabulated groups. The second (D) is the card-count column. The third (B), containing only 00 (not printed), 10, and 20, gives the tens digit of the index of frequency  $h$ . The units digit of  $h$  is given by the numbers in the top row (A) of the figure. Thus the entry 999961 at (E) represents a term  $-39C_{12}$  which has been summed in the  $k=6$  group; in short  $A_{12,6} = -39$ . The terms of the group  $k=6$  are given (in abbreviated form) for comparison with the tabulated record. The corresponding figures are to be found in Fig. 4 in the section headed "6" at (C); this section includes the term marked by (E):

$$C(6, x) = -37 + 33C_2 + 26C_4 - 11C_6 + 52C_8 - 10C_{10} - 39C_{12} + 14C_{16} - 11C_{20}.$$

The order of the entries read from left to right is the order of the original data sheet; this simple correspondence makes it very easy to check the terms of the series. The tabulation shown is part of a Fourier projection using  $hk0$  data in which both  $h$  and  $k$  are even; thus every other column is vacant.

Figure 5 is a copy of a portion of the tabulation record of function cards, both cosine and sine, for values of the argument from 0.068 to 0.090 cycle. The cards tabulated in this section have the serial number 4. The rows marked (X) give the values of  $x$  in thousandths. Some of the listed values of  $k$  are indicated by (K); other items marked are the serial number (S) and the card-count (C). The tabulations of subsequent summations over the second index  $k'$  have the form of the one-dimensional calculation shown in Fig. 3.

The calculation of a three-dimensional Fourier series, although a much longer job, possesses no new features. A three-dimensional series may be written as the sum of products of one-dimensional series and the procedure is then much the same as for a two-dimensional calculation. If the ranges of the indices of the coefficients of the three-dimensional series are given by

$$0 \leq h < r, \quad 0 \leq k < s, \quad \text{and} \quad 0 \leq l < t, \quad (5)$$

then the series is calculated by summing  $st$  one-dimensional series of  $r$  terms each, combining these to give  $t$  two-dimensional series of  $s$  terms each, and finally computing  $n$  sections of the three-dimensional function by performing  $n$  single summations of  $t$  terms each. Details of the procedure for three-dimensional series have been given by Goodwin and Hardy.<sup>6</sup>

<sup>6</sup> T. H. Goodwin and R. Hardy, *Phil. Mag.* **25**, 1096 (1938).

### Preparation of the Master Fourier-Series File and Duplicate Files

The master Fourier-series file of 21,960 cards was prepared in the following way. A four-place table of values of  $\cos 2\pi x$  for  $x=0, 0.002, 0.004, \dots, 1.000$  was taken from standard trigonometric tables. The first 126 of these numbers are those punched on the set of cards 1C1. The numbers for other frequencies were obtained by suitable selection from these numbers; thus for  $\cos 2\pi 2x$  the values for  $x=0, 0.002, 0.004, \dots$ , are the first, third, fifth, etc., numbers in the original sequence. In this way tables were prepared for  $100 \cos 2\pi hx$ ,  $-100 \cos 2\pi hx$ ,  $100 \sin 2\pi hx$ , and  $-100 \sin 2\pi hx$ ,  $h=0$  to 30. These were then transferred to punched cards by an experienced operator, using a motor-driven key punch. They were checked by sorting into alternate sequence the complete set for amplitude  $+100$  and that for amplitude  $-100$  and adding in pairs on the tabulator; each non-zero sum tabulated indicated an error.

The complete set of cards in alternate sequence was then tabulated in pairs by subtraction of the second card of each pair from the first, to give the numbers for amplitude  $+200$ , which were then transferred to cards by a summary punch. Cards for amplitude  $-200$  were similarly prepared, and the two sets were then checked by the same method. The process was repeated for amplitudes  $\pm 300$ ,  $\pm 400$ , and  $\pm 500$ .

The cards for amplitudes  $\pm 1$  to  $\pm 50$  were reproduced from these automatically by the Los Angeles Service Bureau of the IBM Corporation. The working files were similarly reproduced automatically from the master file. Each working file contains 19,032 cards; the cards for amplitudes  $\pm 300$  and  $\pm 400$ , which are not of great use, are not included in these files. One working file permits the satisfactory evaluation of Fourier series. If a large amount of work is being done, however, it is convenient to have several files at hand.

The values of the function entries in the cards are accurate to  $\pm 1$  in the last (hundredth's) place except for the cards of amplitude 200 and 500, which, as a result of their preparation from the cards of amplitude 100, are accurate only to within  $\pm 2$  and  $\pm 5$  hundredths, respectively. In

consequence of the method of generation of the lower amplitudes (1,  $\dots$ , 50) by diagonal reproduction, proper rounding could not be accomplished, and the entries on these cards have an average error of  $-0.005$  in the *magnitude* of the function superimposed on the random error (which does not exceed  $\pm 0.005$ ). This degree of non-randomness of the errors will significantly increase the resulting average error of the sums for values of the argument at which most of the Fourier terms happen to be in phase.

From time to time the files are checked for errors caused by card damage or incorrect replacement of damaged cards by tabulating the cards in their file order with control on frequency using a special control circuit which is set up by a toggle switch mounted on the tabulator plugboard. If a non-zero total appears, indicating an error, the frequency group in which the error exists is sorted into groups of corresponding cards of the same amplitude (positive or negative), and the cards are tabulated again with control on column 77 using a second special wiring. The error is located as to amplitude by a non-vanishing total in the printed record. The faulty amplitude pack is now sorted according to serial number and tabulated in the normal way. The error is thus limited to one of two cards—the positive and negative pair of the last non-zero total—where it is finally located by direct inspection and comparison with the corresponding cards of the master file. It is important that this procedure of verification also serves to reveal by non-zero totals and an incorrect card count when, through sorting or filing error, cards are missing from a frequency group or misplaced or extra cards are present. The identification of missing or extra cards can readily be completed by inspection, after the tabulation by amplitude groups, with much less trouble than is the lot of the operator if a missing or extra card error turns up in a lengthy normal calculation.

Our master file is available for the production of working files for use by investigators in universities and research laboratories.

### REFLECTION CARDS AND THEIR USE

It has been found convenient to punch a card for each x-ray reflection obtained in a given investigation; a reflection card may contain the



indices, the interplanar distance, the observed intensity, the product of the Lorentz and polarization factors, the observed and calculated structure amplitudes, the multiplicity factor, and other information which it may be desired to list, tabulate, or sort in any order. Such cards are of use particularly in the tabulation of data sheets for the calculation of two- or three-dimensional series. For example, the Fourier series representing the electron density distribution in the plane  $z = \frac{1}{2}$  of a crystal cell is

$$\rho(x, y, \frac{1}{2}) = \frac{1}{VN} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \times \cos 2\pi(hx + ky + l/2), \quad (6)$$

where  $V$  is the volume of the cell,  $F_{hkl}$  is the amplitude of the reflection  $hkl$ , and  $N$  is the total number of electrons in one unit cell. This may be written

$$\rho(x, y, \frac{1}{2}) = \frac{1}{VN} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \times \left\{ \sum_{l=-\infty}^{\infty} F_{hkl} (-1)^{|l|} \right\} \cos 2\pi(hx + ky). \quad (7)$$

With use of the control facilities of the tabulator the summation over  $l$  can be made in a few minutes. In a problem requiring extensive use of Fourier methods the time saved in the arrangement of data prior to calculation may greatly exceed the few hours required for the punching of the reflection cards.

#### STRUCTURE-FACTOR CALCULATIONS

The evaluation of structure factors generally consists in forming the quantities  $a = (hx + ky + lz)$ , looking up in tables the atomic scattering factors for the reflections and the values of the trigonometric functions of  $(hx + ky + lz)$  for each atomic term, multiplying these in pairs, and summing these products over all atoms in the unit.

#### Parameter Cards

A considerable fraction of the time used in calculating structure factors is consumed in computing quantities of the type  $(hx + ky + lz) = a$  for each of the several atoms in the unit of structure. The calculation of these quantities is

greatly facilitated by the use of parameter cards; the parameters  $x_1, x_2, \dots$  of the atoms are punched into the successive fields of a card, and the  $y$  and  $z$  parameters in the same fields of other cards, one field being used for each independent atom in the unit. Several copies of each of the cards are made; differently colored cards are used for  $x, y$ , and  $z$  so that they may be distinguished visually. The evaluation of  $(5x + 3y + 7z)$ , for example, is then accomplished by tabulating together 5  $x$ -cards, 3  $y$ -cards, and 7  $z$ -cards. The result is printed together with the indices of the reflection.

#### Atomic Scattering Cards

Most of the schemes for the computation of structure factors which we have considered require either the preparation of very extensive card files or the use of machines in existence but not available in these Laboratories at present. A relatively simple application of the machines to this problem is described below. The values of  $f_j \cos 2\pi a$  are tabulated in card files,  $f_j$  being the scattering factor of the atom of kind  $j$ . There is one card for each value of  $a$  from 0.000 to 0.500 in steps of 0.002. The twelve numerical function-fields of each card contain  $f_j \cos 2\pi a$  for  $\sin \vartheta/\lambda = 0.05, 0.10, \dots, 0.60$ . The values of  $a$  and  $1-a$  are stamped on the face of each card. Thus far files for carbon, nitrogen, and oxygen have been prepared using Pauling-Sherman<sup>7</sup> and Hartree<sup>8</sup> scattering factors; each file contains only 250 cards. If these files are used in conjunction with parameter cards only the cosines are needed, since sines are obtained by summing  $(hx + ky + lz - 0.250)$ , and negatives of either function by adding 0.500 during the parameter summation.

The complete procedure for structure-factor calculation is as follows. Parameter cards for  $x, y$ , and  $z$  are punched and tabulated, as described in the preceding section, in small hand-picked groups to give a quadruple-spaced tabular report of the required parameter sums  $a$ . This report sheet, generally 6 to 12 feet long, is run

<sup>7</sup> Linus Pauling and J. Sherman, *Zeits. f. Krist.* **81**, 1 (1932).

<sup>8</sup> The "Hartree scattering factors" have been calculated by R. W. James and G. W. Brindley, *Phil. Mag.* **12**, 81 (1932), using electron density functions computed by D. R. Hartree, *Proc. Cambridge Phil. Soc.* **24**, 89, 111 (1928).

back to its starting position on the platen of the machine. Reading from the parameter-sum list on this report sheet, the operator draws from the small  $f_j \cos 2\pi a$  files the terms indicated on the report sheet and tabulates them on the report sheet just under the appropriate parameter sums. The used cards are allowed to collect until the files have been so depleted as to be no longer usable; then the files are refreshed by sorting in the used cards by machine. For continued use several copies of each file could be prepared, their size and cost being very small. When the desired set of structure factors has been tabulated the investigator interpolates the true structure factor from those given over the range of  $\sin \vartheta/\lambda$ , with use of the value of  $\sin \vartheta/\lambda$  for the reflection. For crystals of small scattering power, such as those containing only carbon, nitrogen, and oxygen, visual interpolation or simple proportion is sufficient for most needs, errors running well under 5 percent in all the calculations which have been made. This method of calculation seems to be sufficiently accurate for all but the most exacting investigations, and it is certainly of great use in the calculation of structure factors for many preliminary models.

#### LEAST-SQUARES CALCULATIONS

The evaluation of crystal parameter values by the least-squares method as applied by Hughes<sup>9</sup> has been greatly expedited by the use of the punched-card machines. The numerical factors of the observational equations are punched into cards, one card being used for each observational equation containing not more than 23 parameters. It is convenient to punch in an additional field of each card the sum of all of the elements of the observational equation. The pack of cards is then sorted on the units-digit of the first field; the cards punched "9" in this column are passed through the tabulator, then the "8" and "9" packs are put through, etc., until all the cards except the "0" pack have been added the indicated number of times. The complete pack is then sorted on the tens-digit column of the first field, and the small packs added as before with a one-place displacement of the counter wiring. If only two-place figures have been used, the operations mentioned suffice to compute the numerical

factors of the first normal equation. The results are printed by pressing the hand total key of the machine. Continued sorting and totaling of the rest of the fields result in the printing of the complete matrix representing the set of normal equations. The symmetry of the matrix of the set provides a check of the correctness of the calculation except for the diagonal terms, which may be checked by forming the sum of the elements in each row or column of the matrix and comparing it with the corresponding values which the machine computes from the sums of the elements of the observational equations. The solution of the set of normal equations is then carried out by any of the various methods for the solution of simultaneous equations. In the reduction of 80 observational equations to 12 normal equations in 12 variables the following times were required in the computational steps:

|  |          |
|--|----------|
| Punching of observational equations into cards | 1½ hours |
| Verification and correction of cards           | ½ hour   |
| Tabulation of cards to give normal equations   | 3 hours  |

The normal equations were solved by methods not using punched cards in about one and one-half days. The total time required was thus about two days. The same calculation made with an adding machine would have taken about five days. The time reduction in the step carried out with the IBM machines is from 3½ days to 5 hours. The procedure just described is longer than the progressive digitizing method described by A. E. Brandt of Iowa State College in "Practical Applications of the Punched Card Method in Colleges and Universities."<sup>4</sup> This method, however, requires the use of a summary punch, which is not available to us at the present time. Although the saving in time is great in the longer method just described, still greater efficiency may be attained with the progressive digitizing method.

#### THE SAVING OF TIME

The time factor is in all cases greatly in favor of the punched-card method relative to summation procedures used in the past. Fourier projections which by the Beevers-Lipson method required several days of calculation can now be made in 5 to 7 hours. At the same time the density of calculated points is much greater and the accuracy of the computation is assured. The machine steps in the least-squares calculations require

<sup>9</sup>E. W. Hughes, *J. Am. Chem. Soc.* **63**, 1737 (1941).

only a few hours, as compared to one or two days with use of an adding machine, and again the accuracy of the work is assured. With the use of parameter cards and the structure-factor files the calculation of structure factors can be accomplished in about one-eighth of the time previously required.

During the past five years the punched-card files and machines have been used by a dozen different investigators in the study of the structures of more than a score of crystals. Among the crystals whose investigation has been facilitated by this method of calculation are melamine,<sup>9</sup> the addition compound of urea and hydrogen peroxide,<sup>10</sup> dl-alanine,<sup>11</sup> biphenylene,<sup>12</sup> hexamethylenetetramine,<sup>13</sup> and nickel glycine dihydrate.<sup>14</sup>

<sup>10</sup> Chia-Si Lu, E. W. Hughes, and Paul A. Giguère, *J. Am. Chem. Soc.* **63**, 1507 (1941).

<sup>11</sup> Henri A. Lévy and Robert B. Corey, *J. Am. Chem. Soc.* **63**, 2095 (1941).

#### ACKNOWLEDGMENTS

We wish to express our gratitude to the International Business Machines Corporation, which has made available the machines for this work, and to the Committee on Scientific Aids to Learning, which made the California Institute of Technology a grant for the preparation of sets of punched cards and for incidental expenses. We are also grateful to Dr. W. J. Eckert for suggesting the use of the IBM machines, and to Mr. A. A. Burgoyne, Jr., of the Los Angeles Office of the IBM Corporation, Dr. A. J. Stosick, Dr. H. Lévy, Dr. S. Weinbaum, and Dr. J. H. Sturdivant for their assistance in the development of the methods which have been described.

<sup>12</sup> Jürg Waser and Chia-Si Lu, *J. Am. Chem. Soc.* **66**, 2035 (1944).

<sup>13</sup> Philip A. Shaffer (to be published).

<sup>14</sup> A. J. Stosick, *J. Am. Chem. Soc.* **67**, 365 (1945).