

The Use of Punched Cards in Molecular Structure Determinations

II. Electron Diffraction Calculations*

P. A. SHAFFER, JR., VERNER SCHOMAKER, AND LINUS PAULING
Gates and Crellin Laboratories, California Institute of Technology, Pasadena, California

(Received July 22, 1946)

Special files of punched cards for use with the electrical accounting machines of the International Business Machines Corporation in evaluating the Fourier sums and integrals used in the determination of the structure of gas molecules from electron diffraction photographs are described.

INTRODUCTION

IN this article we discuss briefly the evaluation of the intensity and radial distribution functions which are involved in the determination of the structures of gas molecules by the electron diffraction method.¹ Heretofore, these calculations were by far the most time-consuming part of the entire investigation; however, with International Business Machines Corporation equipment and the punched card files described here the calculations typically require about the same amount of time as the preparation and visual examination of the diffraction photographs. In addition to speeding up the ordinary course of the work, the rapid punched card methods make feasible the use of certain powerful and generally useful numerical methods which would have proved too time-consuming to be practical with the earlier computational procedures. These new methods may be expected to facilitate the study of relatively complicated molecules.

THE THEORETICAL INTENSITY FUNCTIONS

The intensity functions and radial distribution functions which are utilized in the interpretation of the observed diffraction pattern are related to the theoretical expression

$$I(s) = K \sum_{i,j} \psi_i(s) \psi_j(s) \frac{\sin sl_{ij}}{sl_{ij}}, \quad (1)$$

* Contribution No. 1020 from Gates and Crellin Laboratories.

¹ Comprehensive discussions of the methods of interpretation of electron diffraction photographs of gas molecules are contained in the following papers: L. Pauling and L. O. Brockway, *J. Chem. Phys.* **2**, 867 (1934); L. Pauling and L. O. Brockway, *J. Am. Chem. Soc.* **57**, 2684 (1935); L. O. Brockway, *Rev. Mod. Phys.* **8**, 231 (1936); V. Schomaker, Thesis, California Institute of Technology, 1938.

for the intensity of coherent scattering of a beam of electrons at angles ϕ by randomly oriented molecules. In this expression K is a constant, ψ_i is the scattering factor (for electrons) of the i th atom, l_{ij} is the distance between the i th and j th atoms in the molecule, and

$$s = \frac{4\pi \sin \phi/2}{\lambda}, \quad (2)$$

λ being the wave-length of the electrons. The summation is extended over all the atoms in the molecule. The factor ψ_i is related to the x-ray form factor $f_i(s)$:

$$\psi_i = (Z_i - f_i(s))/s^2, \quad (3)$$

where Z_i is the atomic number of atom i .

Because of the contrast-sensitive properties of the eye the appearance of the electron diffraction photographs is more closely related to some derived expression in which the very rapid decline of $I(s)$ with increasing s is suppressed. Graphs of what we call the simplified intensity function

$$\begin{aligned} & \frac{1}{\sum_i \psi_i^2} \{I(s) - I_{at}(s)\} \\ &= \frac{K}{\sum_i (Z-f)_i^2} \sum_{\substack{i,j \\ i \neq j}} (Z-f)_i (Z-f)_j \frac{\sin l_{ij}s}{l_{ij}s} \end{aligned} \quad (4)$$

and the reduced intensity function

$$\begin{aligned} I^0(s) &= \frac{s[I(s) - I_{at}(s)]}{\sum_i \psi_i^2} \\ &= \frac{K}{\sum_i (Z-f)_i^2} \sum_{\substack{i,j \\ i \neq j}} \frac{(Z-f)_i (Z-f)_j}{l_{ij}} \sin l_{ij}s, \end{aligned} \quad (5)$$

both represent the appearance of the photographs well enough for our purposes. In both functions the atomic scattering, I_{at} (given by the terms of (1) for $i=j$, which are independent of the molecular structure), has been omitted. Equations (4) and (5) represent the diffraction pattern for essentially rigid molecules, such as carbon dioxide or carbon tetrachloride, in which the relative displacements of the atoms caused by intramolecular vibration are negligibly small. For non-rigid molecules Eqs. (1), (4), and (5) must be suitably modified¹ to represent, essentially, the appropriate average over all the instantaneous configurations assumed by the molecules in the course of their vibrations. Each term in the intensity function must accordingly be replaced by an integral over all values of the interatomic distance l , with a radial distribution function $l^2 D_{ij}(l)$ which gives the probability of finding a molecule with the atoms i and j separated by a distance l :

$$I^0(s) = \frac{K}{\sum_i (Z-f)_i^2} \sum_{\substack{i,j \\ i \neq j}} (Z-f)_i (Z-f)_j \times \int_0^\infty l_{ij} D(l_{ij}) \sin l_{ij} s dl_{ij} \quad (6)$$

It usually happens that the molecule is significantly non-rigid with respect to only a few of the interatomic distances.

It is possible to write Eq. (6) in a formally simpler fashion as

$$I^0(s) = K \int_0^\infty l D(l) \sin ls dl, \quad (7)$$

in which $D(l)$ is a radial distribution function. If all the atoms of the molecule are of the same kind, $l^2 D(l)$ gives the relative probability of finding two of them a distance l apart and peaks of $l^2 D(l)$ obviously correspond to interatomic distances, but if the atoms are not all of the same kind, the precise meaning of $D(l)$ may become rather abstruse, and will not be discussed here. For any assumed molecular model, however, $D(l)$ can readily be found; moreover, the interpretation of a known $D(l)$ in terms of interatomic distances is usually not difficult.

THE RADIAL DISTRIBUTION INTEGRAL

Equation (7) is a Fourier integral, which, like the Fourier series discussed in Part I, can be inverted. The resulting expression

$$lD(l) = K' \int_0^\infty I^0(s) \sin ls ds \quad (8)$$

is called the radial distribution integral. It does not matter that s^2 is not defined for values greater than $4\pi/\lambda$ because with the usual values of λ the amplitude of variation $I^0(s)$ always decreases essentially to zero at a value of s smaller than $4\pi/\lambda$, but it is important when $lD(l)$ is to be evaluated that $I^0(s)$ cannot be determined throughout the range in which it is supposed to have an appreciable amplitude. The devices mentioned in this connection in the next paragraph constitute only one of the possible ways of making use of the radial distribution integral in spite of our incomplete knowledge of $I^0(s)$.

² Three items concerning this calculation need mention here: (1) Because the first maximum of $I^0(s)$ (corresponding to the central maximum of the simplified intensity function) is essentially unobservable, it is either estimated entirely from past experience or is taken from a theoretical intensity curve, and suitable (but altogether minor) allowances for error and arbitrariness ultimately have to be made. (2) $I^0_{\text{visual}}(s)$ is multiplied by an attenuation or convergence factor in order that the requirements of Eq. (8) may practically be satisfied by an integration extending only to the largest angle s_{max} for which the diffraction pattern is observed. We use the factor $\exp(-as^2)$ (generally with a chosen so that $\exp(-as^2_{\text{max}}) = 0.1$), which has the effect of transforming any potentially sharp peak of $lD(l)$ into a normal error curve distribution function with $\langle \delta l^2 \rangle_M = 2a$. Since this is just the distribution function which is ordinarily assumed (reference 1), for example in Eq. (6), to represent the distributions of interatomic distances which arise from the thermal and zero-point vibrations of molecules, $\exp(-as^2)$ is sometimes called the artificial temperature factor. (3) It is necessary to approximate the integration in Eq. (8) by a summation over selected values of s . Heretofore (reference 1), the summation has been restricted to a set of terms in $I(s_i) \sin ls_i$ corresponding to the apparent maxima, or to the apparent maxima and minima, of the diffraction pattern. This summation is neither very satisfactory nor easy to comprehend, although it does give fairly useful results; yet for even a simple molecule it may be rather time consuming, by the usual methods, involving the accumulation of twenty or thirty products for each calculated point. With the help of punched cards, however, it is not difficult to sum a series of even a hundred terms. Accordingly, we now usually evaluate a sum over evenly spaced s values (reference 3) that is loosely called the radial distribution integral, the spacing being chosen so small that the summation is a fully satisfactory approximation to the integration for the range of l which is of interest. Ordinarily we take a spacing $\Delta s = (\pi/10)A^{-1}(\Delta q = 1, \text{ see below})$ that in practice is satisfactory even for $l = 7A$.

The Interpretation of the Electron Diffraction Pattern

The interatomic distances that occur in a molecule, and determine its structure, can be deduced from the diffraction photograph in two different ways. In the correlation procedure theoretical intensity curves (computed according to Eqs. (4), (5), or (6)) are drawn for all the likely molecular models. Any model whose curve adequately represents the details of the appearance of the photographs is compatible with the electron diffraction experiment; if satisfactory models of only one distinct type can be found, the structure determination is unique, at least within the realm of what have been considered likely models, and the parameters of the structure are determined with a precision that depends on how widely they may be varied without destroying the agreement between calculated curves and photographs. In the radial distribution method an approximation to the radial distribution function (Eq. (8)) is calculated with the use of a curve, representing a reduced intensity function, $I^0_{\text{visual}}(s)$, which has been drawn in accordance with the appearance of the diffraction photographs and the expected general behavior of $I^0(s)$ for the type of molecule being investigated.^{2,3} The plotted radial distribution function shows immediately what interatomic distances occur in the molecule, and from these the molecular structure can usually be deduced. Mathematically the two methods are almost strictly equivalent, but in practice the two procedures are almost always used together in order to take advantage of the greater directness of the radial distribution method and the greater speciality and somewhat greater sensitivity of the correlation method.

The Calculations

The general form of the calculations described in the foregoing sections is seen to be the evaluation of

$$F(x) = \sum_i G_i(x) \sin y_i x. \quad (9)$$

For the simplest intensity functions there may be only two sine terms with constant coefficients; for more complicated molecules there may

³ R. Spurr and V. Schomaker, *J. Am. Chem. Soc.* **64**, 2693 (1942).

be ten or twenty or more terms in the intensity function and many of the $G_i(x)$ may be variable. For non-rigid molecules even more terms are required to approximate adequately the integrals which appear in Eq. (6). The adequate representation of the radial distribution integral² typically requires 75 to 100 terms with constant coefficients, while the older type of radial distribution summation involves from 10 to 30 or more terms. In all cases it is necessary to evaluate Eq. (1) for about 100 values of x .

In the beginning these calculations were made with the aid of only the usual trigonometric tables and an adding machine, so that only the simplest types could be carried out in a reasonable time. Later, with the Sherman-Cross strips described below and a motor-driven adding-calculator, a ten-term intensity function with constant coefficients could be evaluated by a skilled operator in two to four hours, but longer calculations such as the radial distribution integrals were too difficult to be used at all. With our present punched-card equipment the filing, sorting, and tabulating operations for a ten-term intensity calculation with constant coefficients require only ten to fifteen minutes while a hundred-term approximation to a radial distribution integral can be evaluated in an hour and a half or two hours. For the longer calculations the work can be so arranged as to provide useful subtotal tabulations without appreciably increasing the time required. Although the transition to variable coefficients is relatively very much more troublesome and time-consuming for the punched-card methods than with the Sherman-Cross strips, in our work a considerable advantage for the punched-cards remains. Fortunately, however, most of the calculations can be carried out with constant coefficients by means of a suitable device.

The Sherman-Cross Tables

The Sherman-Cross tables consist of a set of paper strips, each an inch wide and two feet long, on which are printed the values of $\sin y_i x / y_i x$ for values of y_i at intervals of 0.01 from 0.80 to 4.00 for values of x up to 40, and for values of y_i from 4.02 to 8.00 at intervals of 0.02 for values of x up to 20. The tabular interval in x is 0.2 throughout. Each strip corresponds to a given

value of y_i (indicated at the top of the strip) and a range of x values, the values of $(\sin y_i x / y_i x)$ being arranged along the length of the strip in correspondence to the successive values of x . For computation with these strips Eq. (9) is written in the form

$$\frac{F(x)}{x} = \sum_i y_i G_i(x) \frac{\sin y_i x}{y_i x} \quad (10)$$

and the required strips are arranged side by side on a suitable holder so that entries corresponding to the same value of x lie in the same horizontal row. Then for each value of x , the sum of the products of the entries read from the strips by the corresponding coefficients $(y_i G_i(x))$ is accumulated on an adding-calculator and recorded.

THE PUNCHED-CARD FILES AND THEIR USE IN MAKING ELECTRON DIFFRACTION CALCULATIONS

In design, preparation, and use the punched-cards for electron diffraction calculations are so similar to those already described in the preceding paper that only the important differences need be given here.

For all the cards, which happen to be designated by three different file numbers, the tabulated function is $A \sin yx$. In file No. 1, the amplitudes A (indicated by the same color and punch codes as for the crystal structure cards) are +1, -1, +2, -2, +3, -3, +4, -4, +5, -5, +10, -10, +20, -20, +30, -30, +40, -40, +50, -50, +100, -100, +200, -200, +500, and -500; and in file No. 3 they are +300, -300, +400, -400; file No. 2 contains the same amplitudes as file No. 1 up to and including -100. The cards of file No. 1 of amplitudes +100, -100, +200, -200, +500, -500 and the cards of file No. 3 are regarded as constituting a master set which is never used in routine work, but is held in reserve for use in preparing further working sets by high speed machine reproduction. The "frequency" y , which in intensity calculations is identified with l and in radial distribution calculations with s , ranges from 0.01 to 5.00 in steps of 0.01.⁴ Each ampli-

⁴ In the radial distribution integral calculation the interval in s is usually identified with an interval of 0.05 in y , and only moderate amplitudes are required for suitable accuracy in the representation of $I^0_{\text{vis}}(s) \exp(-as^2)$. It has

tude-frequency pack contains, besides the index card, 12 function cards which give $(A \sin yx)$ for 144 values of x ranging from 0.0 to 143 ($\pi/10$) in steps of $(\pi/10)$.⁵ In our electron diffraction files there are accordingly altogether 312,000 function cards and 26,000 index cards.

The classification field of the electron diffraction cards consists of columns 73-80, just as for the Fourier cards, and the same plugboard is used for both. In column 73 is punched the file number and an x or a 0 to indicate, respectively, whether the amplitude is negative or positive. The frequency y is punched in columns 74, 75, and 76, the amplitude code in column 77, the serial number in columns 78 and 79 (the serial number ranges from 20 to 32), and E (12-5) is punched in column 80. The transition from crystal structure to electron diffraction cards thus requires that the wirings to columns 73 and 74 on the tabulator be interchanged; this is done by means of a toggle switch mounted on the plugboard cover. The index cards are similar to those of the crystal-structure set with the addition of a second punching of the amplitude in columns 66 (sign), 67, 68, and 69. A second toggle switch controlling a class selector which is activated by the X-72 punch of the index cards makes it possible to list and tabulate the index cards either after the crystal structure fashion or in another way, with a control on all digits of the frequency, such that the amplitude, the frequency, and an "E" are printed (the card count is suppressed) at the extreme left-hand side of the record for each term of the summation. A sample calculation, a radial distribution function for hexamethylenetetramine, in which this second method of listing and tabulating the index

been convenient, therefore, to keep separately for this calculation the cards of file No. 1 with amplitudes +1, -1, ..., +20, -20 and frequencies 0.05, 0.10, ..., 5.00.

⁵ There are somewhat fewer entries (72,000 $x-y$ combinations) in the punched card tables than in the Sherman-Cross tables (84,200 $x-y$ combinations); we find that the arrangement of our tables makes them more convenient to use in electron diffraction work while their content ($\sin z$, instead of $\sin z/z$) often makes them useful for problems outside the field of electron diffraction. The inconvenience of the odd tabular interval in x , which becomes s in making intensity calculations, is entirely eliminated by dealing always with the new quantity

$$q = \frac{10}{\pi} s = \frac{40}{\lambda} \sin \phi / 2. \quad (11)$$

In terms of q the tabular interval becomes unity.

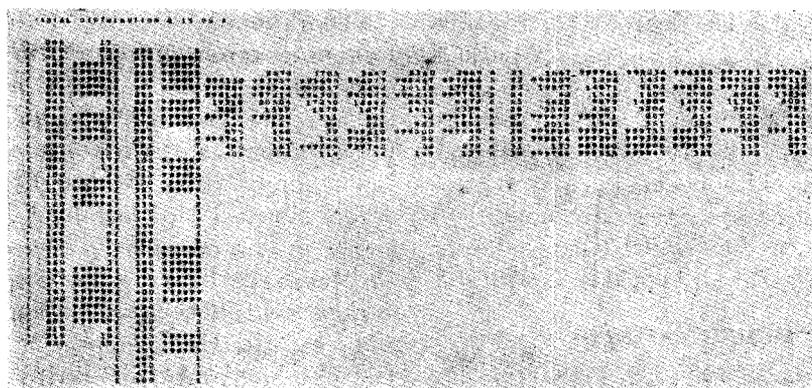


FIG. 1. Sample tabulation of electron diffraction cards. A radial distribution function for hexamethylenetetramine.

cards has been used, is shown in Fig. 1. For electron diffraction calculations this special treatment of the index cards is usually to be preferred to the normal one because it gives a much more compact and legible record—only rarely does an electron diffraction calculation involve frequencies which are spaced closely enough to favor the normal treatment.

The function to be evaluated by use of the punched cards is written in the form of Eq. (9). If the coefficients $G_i(x)$ are constants, the appropriate packs of cards are then withdrawn from the files, sorted, tabulated, resorted into filing order, and returned to the files in essentially the manner described in the preceding paper for the crystal structure calculations. The printed report is like that obtained in the crystal structure work except (usually) for the tabulation of the index cards. The only significant difference in procedure is in the resorting for filing where the sequence of operations is to sort on columns 77, 73 (zone), 76, 75, 74, and 73 (normal). Very frequently, however, some of these sorting operations can be omitted or performed manually, or the order may be changed to advantage, as, for example, in the evaluation of a series of intensity functions which have one or several terms in common. It is then useful to take the common packs from one file, and the rest from the other, so that the cards used in one calculation which are required also for the next may be recovered after one sorting, that on column 73 (normal).

If the coefficients $G_i(x)$ of the summation as

first written down are not constants it is possible to carry out a mathematical transformation which makes them so, as is suggested by Eq. (7) which can indeed be approximated by a sum of sine terms with constant coefficients. Sometimes this is the expeditious thing to do, but often it is better to adopt some other procedure. Because our practical methods of handling variable coefficients are still being developed, it seems best not to give them in detail here, although two devices which are well established may be mentioned. The most common distribution of interatomic distances $l^2 D_{ij}(l)$ which occurs as a result of the vibrations of molecules is a normal error curve distribution about l_0 , the equilibrium value of l_{ij} ($l^2 D_{ij}(l) = \text{Const. exp} [-\alpha(l_{ij} - l_0)^2]$). In our present work this distribution is best represented directly by the use of Eq. (6) (in the actual calculation the integration is adequately approximated by a summation over three to seven terms for values of l_{ij} spaced at intervals of about 0.05Å in the neighborhood of l_0); heretofore, it was customary to make use of the mathematical transformation according to which the integral over l_{ij} is replaced by the single, approximately equivalent term $\exp(-as^2) \sin l_0 s$, with $a = 1/4\alpha$.

The second device is useful whenever a group of terms of the summation appear unchanged in a series of calculations. These terms are tabulated separately and the results are transferred to a special pack of cards which is then used for each calculation of the series. In particular, terms with variable coefficients may be handled

in this way by evaluating them for the required values of x and punching them onto a pack of cards. These cards have to be used enough times to justify the fifteen or twenty minutes required by an unskilled operator for their preparation. This procedure would of course be used to great advantage in almost any series of computations if a summary punch were available for automatic punching of the special card packs.

CONCLUDING REMARKS

A typical electron diffraction study involves the calculation of one or more radial distribution summations or integrals and five to thirty or more theoretical intensity functions, depending on the complexity of the molecule and the thoroughness of the investigation. By affording a threefold to tenfold reduction of the time required for these calculations the punched card procedure makes feasible the study of more complicated molecules for which the labor of making intensity calculations would otherwise be too great and makes practical the use of more powerful methods than could heretofore be used (e.g., the radial distribution integral). It reduces fatigue, for the operations of filing, sorting, and tabulating make less exacting demands on the operator than those of multiplication and adding involved with the Sherman-Cross strips. Finally, it is easier to check the correctness of all parts of the calculations, and the incidence of errors is decreased.

Illustrations of the use of the punched card files in electron diffraction investigations are provided by papers published in the *Journal of the American Chemical Society* from these Laboratories during the past few years.⁶

The preparation of the electron diffraction sets of punched cards involved the help of many people at various times in the period June 1940 to March 1941. We wish to acknowledge the various contributions to various phases of the work by Mr. H. Feldman, Dr. and Mrs. Paul Giguère, Mr. W. J. Howell, Jr., Dr. H. A. Lévy, Dr. R. E. Rundle, Dr. M. T. Rogers, Dr. W. Shand, Jr., Dr. R. A. Spurr, Dr. S. M. Swingle, Dr. A. L. Wahrhaftig, Dr. J. Waser, Dr. S. Weinbaum, and Dr. W. West. We are also greatly indebted to the International Business Machines Corporation, which 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.

⁶ The propargyl halides: L. Pauling, W. Gordy, and J. H. Saylor, *J. Am. Chem. Soc.* **64**, 1753 (1942); perylene: reference 3; spiropentane: J. Donohue, G. L. Humphrey, and V. Schomaker, *ibid.* **67**, 332 (1945); methylisocyanide, W. Gordy and L. Pauling, *ibid.* **64**, 2952 (1942); ozone, W. Shand, Jr., and R. Spurr, *ibid.* **65**, 179 (1943); biphenylene, J. Waser and V. Schomaker, *ibid.* **65**, 1451 (1943); methylenecyclobutane and 1-methylcyclobutene, W. Shand, Jr., V. Schomaker and J. R. Fischer, *ibid.* **66**, 636 (1944); sulfur nitride, arsenic disulfide, arsenic trisulfide, and sulfur, Chia-Si Lu and J. Donohue, *ibid.* **66**, 818 (1944); hexamethylenetetramine, V. Schomaker and P. A. Shaffer, Jr. (to be published).