Least-Squares Determination of Molecular Structures from Gaseous Electron-Diffraction Data. II. Polynomial Expression of Background

Shuzo Shibata
Department of Chemistry, Shizuoka University, Oya, Shizuoka, Japan

(Received October 6, 1971)

A new least-squares method has been devised for interpreting electron-diffraction data by comparison of an experimental background with a calculated background. The former is derived from an experimental total intensity by assuming a molecular model. The latter is approximated by a polynomial function. A computer program is written to refine structural parameters of a molecule so as to yield a smooth background. The characteristic of the present method is that it does not need any bias depending upon human judgement throughout an analysis. The distinctive features of the procedure are discussed, and the application of this method to structure determination of several molecules is also described.

Since a molecular intensity is not directly observable in electron-diffraction experiment, the procedure of drawing a background through an experimental intensity has been used as a bias to obtain a molecular intensity. In the past this background has often been frozen throughout an analysis; consequently, an uncertainty in the background probably leads to errors in final values of molecular parameters. Therefore, a method with no operator bias is desirable. In a previous paper[1] the criteria for smoothness of a background were presented and a least-squares method was given, that refines molecular parameters to make the background as smooth as possible. The purpose of this paper is to report another method which is based on quite the same viewpoint as the previous method. In the present method a background was expressed by a polynomial function, of which the parameters were allowed to vary freely in every iteration of the least-squares calculation. Bartell *et al.*[2] presented recently a least-squares analysis of total intensity, where a background was presented in terms of a polynomial with an exponential function. The present method does not essentially differ from it except that the best fit between experimental and calculated backgrounds, instead of that between total intensities, is sought by the refinements of molecular parameters and background parameters. Moreover, this method is mathematically rather simple and works without any initial guesses of background parameters.

Procedure

Criterion. An experimental total intensity is divided by an atomic background calculated by means of the scattering theory. Such leveling does not usually yield a horizontal background which cleaves molecular oscillations, but yields a considerably curved background because of imperfections of theory and experiment. However, it was found that this curved background can be approximated satisfactory in terms of a polynomial function, as described below.

A background I_b is provided from both a leveled experimental total intensity I_t and a molecular intensity M which is calculated for an assumed model by use of the relation (5) in Ref. 1. If a molecular model is strictly correct, the background should have no sudden fluctuation over the whole range of scattering angles. On the other hand, if the model is not correct, the background may have some fluctuations with a frequency similar to that of the molecular intensity. Thus, molecular parameters should be adjusted so as to make the background smooth. In the present method, the criterion for smoothness of a background is expressed as follows:

$$P_n = \sum_i w_i[I_b(q_i) - \sum_n a_n q_i^n]^2 = \text{minimum} \qquad (1)$$

Here w_i is a suitable weighting function, and $I_b = I_t/(1+RM)$, where R is an index of resolution. The equation means the minimization of the sum of the squares of the difference between an experimental and a calculated background over experimental points. Smoothness of the background is guaranteed by the degree of the polynomial. This method which we refer to as the *smooth-background method* differs from the usual method in that a background is never frozen and is refined with molecular parameters, and therefore it can determine values of molecular parameters without accompanying the uncertainties arising from a fixed background. The mathematical formulation of this method is omitted here since the details were described in other papers on the least-squares method.[3,4]

Sectional Smooth-background Method. The number of unknowns in a normal equation is generally equal to $(2m+1)+(n+1)$, if molecular parameters of interatomic distances and mean amplitudes with an index of resolution are varied independently to determine the structure of a molecule having m atomic pairs by use of a polynomial of nth degree. When a molecule consists of many atoms, the number of molecular parameters to be determined is large even if the symmetry of a molecule is taken into consideration. In the structure analysis of such a molecule, it seems difficult to solve the normal equation and get a good convergence of a standard deviation. This may be

---

[1] S. Shibata and L. S. Bartell, *J. Mol. Structure*, **9**, 1 (1971).

[2] L. S. Bartell, D. A. Kohl, B. L. Carroll, and R. M. Gavin, Jr., *J. Chem. Phys.*, **42**, 3079 (1965).

[3] K. Hedberg and M. Iwasaki, *Acta Crystallogr.*, **17**, 529 (1964).

[4] Y. Morino, K. Kuchitsu, and Y. Murata, *ibid.*, **18**, 549 (1965).

mainly attributable to the unreasonable expansion of parameters at the initial values and to the correlation among the parameters. To avoid such difficulty, a program has been written so as to first solve the normal equation for a part of molecular parameters with all of background parameters and other molecular parameters being fixed, and then to repeat this procedure alternatively with respect to another part of molecular parameters. After the initial values of the molecular parameters are improved sufficiently by the iteration of such block-diagram calculations, it is possible to obtain reasonable solutions of the complete normal equation with a good convergence. Correlation between molecular parameters should be taken into consideration with respect to the division of molecular parameters into groups. This sectional method can determine a structure of the molecule involving more than thirty interatomic distances as independent parameters, as described in the Application.

In contrast to the above analysis in which all interatomic distances are refined independently, a few parameters are allowed to vary with background parameters in consideration of geometric constrains. The geometrical consistency is maintained by the procedure in Ref. 5 at all stages through the analysis.

Fitting of Background. Some of the most probable backgrounds used for radial distributions are fitted in terms of various analytical functions, and a polynomial was found to be a good approximate function. In general, a background for a short camera range of $q \gtrsim 20$ is leveled fairly well; and it can be approximated satisfactory with the polynomials of low degree, *e.g.*, $n=5$, within an error of 0.3%. Significance test[6] shows that a fitting by lower degree polynomials may be rejected at 0.005 significance level. Nevertheless, a background for a long camera range of $q \gtrsim 10$ has often a hump in the range of small scattering angles and falls off rapidly inside this range; a fit of the jump by polynomials sometimes disturbs the remainder of the background curve. In such a case, however, polynomials of higher degree, *e.g.*, $n=7$, can fit the background within the above error. Moreover, the use of μ background cited in Ref. 1 is probably a better method because the fitting by the polynomial of higher degree causes a large correlation between molecular parameters, and the final parameters determined are usually accompanied by large errors.

Correlation Between Data Points. Correlation coefficients ρ_k between data points q_i and q_{i+k} are computed by the usual way with use of the residuals of backgrounds I_b with respect to data of various molecules, assuming that all the correlation coefficients for constant interval are equal and that all the standard deviations for data points are also equal. The result indicates that they can be expressed approximately by the function $\exp(-\gamma_n|k|\Delta q)$, where Δq is the interval between observations.[7] The correlation matrix can therefore be expressed as follows:

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \cdots\cdots\cdots\cdots & \\ \rho_1 & 1 & \rho_1 & \rho_2 \cdots\cdots\cdots\cdots & \\ \rho_2 & \rho_1 & 1 & \rho_1 \cdots\cdots\cdots\cdots & \\ \cdots\cdots\cdots\cdots & & & \rho_2 \quad \rho_1 & 1 \end{vmatrix} \qquad (2)$$

where all off-diagonal elements are included. The values of the parameter γ_n is slightly dependent on the degree n of a polynomial used for a fit of the background. A higher degree polynomial gives a smaller correlation coefficient, but it can be estimated to be about 0.4 for ρ_1, *i.e.*, $\gamma \sim 1$ Å. However, if a background is manually drawn, a correlation between data points is often significantly increased to about 0.8 for ρ_1, *i.e.*, $\gamma \sim 0.2$ Å.

The inverse of the correlation matrix for which the i j elements are $\exp(-\gamma_n|i-j|\Delta q)$ can be simply expressed as,

$$\begin{vmatrix} 1 & p & 0 & 0 & 0 \cdots\cdots\cdots\cdots 0 \\ p & 1 & p & 0 & 0 \cdots\cdots\cdots\cdots 0 \\ 0 & p & 1 & p & 0 \cdots\cdots\cdots\cdots 0 \\ 0 & 0 & 0 & 0 & 0 \cdots\cdots\cdots 0 \quad p \quad 1 \end{vmatrix} \qquad (3)$$

where

$$\begin{aligned} p &= -\exp(-\gamma_n \Delta q)/[1+\exp(-2\gamma_n \Delta q)] \\ &= -\rho_1/(1+\rho_1^2) \end{aligned} \qquad (4)$$

This form is identical with the relation (13) in Ref. 8, where the above form of the inverse matrix is assumed. The inverse matrix corresponds to the weight matrix w_{ij} in the following relation:

$$\begin{aligned} P_n' &= \sum_i \sum_j w_{ij}[I_b(q_i) - \sum_n a_n q_i^n][I_b(q_j) - \sum_n a_n q_j^n] \\ &= \text{minimum} \end{aligned} \qquad (5)$$

where the diagonal weight matrix elements w_i in the Eq. (1) are taken to be unity for simplicity.

For example, in the data of water[9] for the camera distance of 11 cm, the values of correlation parameter γ_n are 0.90 and 1.05 Å for $n=5$ and 7, respectively. Table 1 lists the molecular parameters and their errors for water obtained by the smooth-background method. A consideration of the correlation between data points increased the errors by a factor of about 1.4, if γ_n of 0.9 Å, *i.e.*, p of -0.35 is used, as seen in the table.

Applications

An application of the smooth-background method for the determination of molecular structures is described, as examples, for carbon tetrachloride, water[9] and bis-(acetylacetonato)palladium(II).[10] The analysis of CCl_4 is illustrated in Fig. 1. The backgrounds were expressed by means of a polynomial of the 7th degree for both long and short camera distance data. The convergence of molecular parameters is shown in Fig. 2 for both data. The C–Cl and Cl–Cl distances were refined independently of each other.

In the analysis of H_2O by both long and short camera distance data, polynomials of various degree were used

---

5) K. Kuchitsu, T. Fukuyama, and Y. Morino, *J. Mol. Structure*, **1**, 463 (1967—1968).

6) W. C. Hamilton, *Acta Crystallogr.*, **18**, 502 (1965).

7) L. S. Bartell, *ibid.*, **A25**, Supplement, 76 (1969).

8) Y. Murata and Y. Morino, *ibid.*, **20**, 605 (1966).

9) S. Shibata and L. S. Bartell, *J. Chem. Phys.*, **42**, 1147 (1965).

10) S. Shibata, K. Ozaki, and K. Iijima, to be published.

to fit the background; however, it was found that the converged values of molecular parameters did not significantly depend upon the degree of the polynomial used and were in agreement with one another within

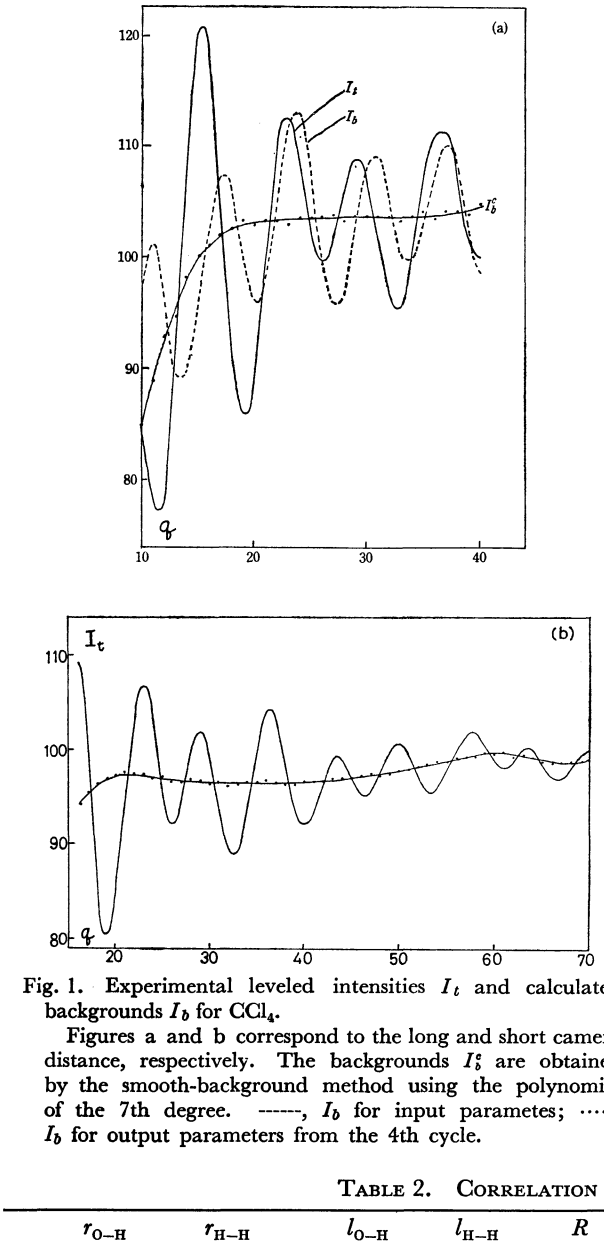

Fig. 1. Experimental leveled intensities I_t and calculated backgrounds I_b for CCl_4.

Figures a and b correspond to the long and short camera distance, respectively. The backgrounds I_b^s are obtained by the smooth-background method using the polynomial of the 7th degree. ------, I_b for input parametes; ······ I_b for output parameters from the 4th cycle.

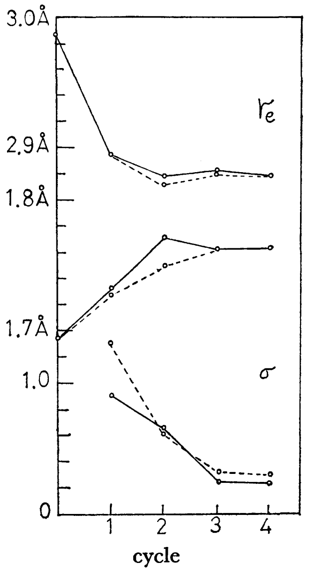

Fig. 2. Determination of the C–Cl and Cl–Cl distances o CCl_4 by the smooth-background method.

——, the short camera distance data; ------, the long camera distance data.

Table 1. Derived parameters for O–H in water from the smooth-background method.[a)]

(units; Å)

	P_5*(L)	P_7(L)	P_5(S)	P_7(S)	P_5'(S)	P_7'(S)
r_e	0.9609	0.9640	0.9605	0.9581	0.9605	0.9575
$r_g(1)$	0.9718	0.9765	0.9713	0.9704	0.9712	0.9702
$r_g(0)$	0.9765	0.9818	0.9760	0.9758	0.9758	0.9757
l_m	0.0672	0.0718	0.0669	0.0717	0.0666	0.0727
$\sigma(r)$	0.0004	0.0064	0.0004	0.0008	0.0005	0.0010
$\sigma(l)$	0.0008	0.0077	0.0006	0.0016	0.0009	0.0020

a) L and S mean long (21 cm) and short (11 cm) camera distance data, respectively. The ranges covered are seen in Fig. 3. The results for P_n were obtained by Eq. (1) by use of polynomials of the nth degree. For P_n* the method of μ background was used. The results for P_n' were obtained by Eq. (5) by use of −0.35 for p in Eq. (4). The Morse asymmetry constant of O–H was assumed to be 2.3 $Å^{-1}$, and w_i in Eq. (1) was taken to be unity. The errors σ do not include systematic errors.

Table 2. Correlation matrix of parameters for water[a)]

r_{O-H}	r_{H-H}	l_{O-H}	l_{H-H}	R	a_0	a_1	a_2	a_3	a_4	a_5
1.000	−0.047	−0.360	0.080	−0.269	0.485	−0.473	0.459	−0.442	0.425	−0.408
	1.000	0.291	0.014	0.445	0.313	−0.287	0.265	−0.247	0.233	−0.221
		1.000	0.279	0.885	0.046	−0.033	0.026	−0.025	0.027	−0.031
			1.000	0.304	0.119	−0.115	0.113	−0.111	0.111	−0.111
				1.000	0.057	−0.032	0.016	−0.005	0.000	−0.000
					1.000	−0.997	0.991	−0.982	0.971	−0.958
						1.000	−0.998	0.992	−0.984	0.974
							1.000	−0.998	0.993	−0.986
								1.000	−0.998	0.994
									1.000	−0.998
										1.000

a) Matrix elements are given by $\rho_{ij}=B^{-1}{}_{ij}/(B^{-1}{}_{ii}\,B^{-1}{}_{jj})^{1/2}$. This matrix corresponds to the analysis P_5(S) in Table 1. The notation in the above relation corresponds to that in reference 11.

---

11) O. Bastiansen, L. Hedberg, and K. Hedberg, *J. Chem. Phys.*, **27**, 1311 (1957).

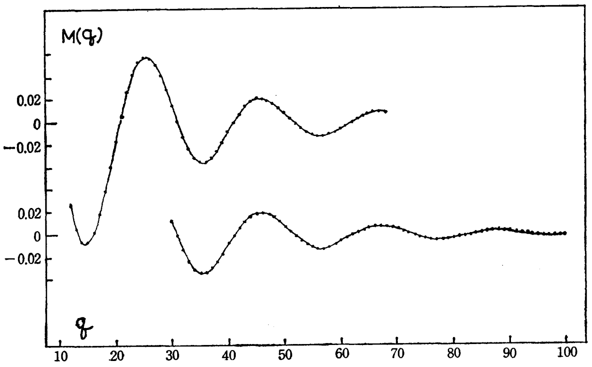

Fig. 3. Experimental and calculated molecular intesities M for H_2O.

The solid curves are the calculated molecular intensities obtained by the smooth-background method using the polynomials of the 5th degree. The upper curve corresponds to the long camera distance and the lower corresponds to the short camera distance. The dots are observations. Both curves correspond to P^*_5 (L) and P_5 (S) in Table 1, respectively.

the limits of errors. The derived parameters for O–H in H_2O are listed in Table 1, where the results obtained by use of the polynomials of 5th and 7th degrees are compared with each other. Table 2 shows the correlation matrix of parameters for water obtained from the analysis using the 5th degree polynomial as the background intensity. The use of higher degree polynomial resulted in larger correlation between molecular parameters, and consequently, increased errors in molecular parameters as seen in Table 1. This seems to be enhanced in the case of a long camera distance data, and therefore, the use of μ background is desirable, as shown in Table 1. Figure 3 shows the experimental and calculated molecular intensities.

As an example of the structure determination of a complex molecule, the structure analysis of Pd-$(C_5H_7O_2)_2$ is described. The three following calculations were compared with respect to the convergence of the standard deviation: 1) all of thirty-two interatomic distances were varied independently, 2) these distances were divided into four sets of eight distances and the least-squares calculation was performed successively by the use of the sectional method described above, 3) four interatomic distances and three bond angles were varied independently and other distances were changed according to the geometrical constraint remains. In these calculations it was assumed that the molecule takes a planar form and the ligand of acetylacetone $C_5H_7O_2$ has the symmetry of C_{2v}. The first case could not give any convergence of the standard deviation. The second and third cases gave the good convergence given in Fig. 4; moreover, the computing time required for the second case was about twice for the third case. Generally, the present smooth-background method gave a relatively slow convergence of solutions in comparison with the previous method using the slope and curvature criteria.[1]

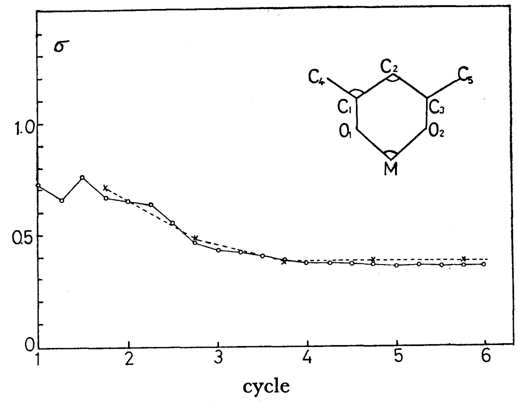

Fig. 4. The standard deviations at each cycle of the calculation for the analysis of $Pd(C_5H_7O_2)_2$.

——, case 2; ------, case 3 in the text. Independent parameters in the case 3 are as follows; interatomic distances of $M\text{-}O_1$, $M\text{-}C_2$, $C_1\text{-}C_2$, $C_1\text{-}C_4$ and bond angles of O_1MO_2, $C_1C_2C_3$, and $C_2C_1C_4$.

Conclusion

The present method seems to be useful for the determination of interatomic distances and mean-amplitudes of vibrations because it is basically simple and straightforward. Electron-diffraction data are handled successively by a digital computer without manual processing of data. Numerical input data to the program routine are digital microphotometer readings. The present method does not need initial guess of background parameters, and thus it removes an operator bias from the analysis of molecular structure by the electron-diffraction method.

We are indebted to the Ministry of Education of Japan for a Scientific Research Grant supported this study. We are also indebted to Professor L.S. Bartell for his kind advice on this study.