

ERROR MATRICES IN GAS-ELECTRON DIFFRACTION

I. EFFECTS OF SYSTEMATIC ERRORS IN INTENSITIES*

L. S. BARTELL AND HSIUKANG YOW

Department of Chemistry, The University of Michigan, Ann Arbor, Mich. 48104 (U.S.A.)

(Received 22 August 1972)

ABSTRACT

Electron diffractionists have largely ignored the effect of systematic errors in intensities upon derived molecular parameters. The present investigation proposes a treatment of such errors in terms of certain coefficients characterizing systematic errors which are included as variables interacting with molecular parameters in the information matrix. A simple method is devised to limit coefficient variances to reasonable values. Simple analytical approximations are derived to account for parameter regression slopes. Model systems with poorly resolved internuclear distances are tested. It is found that systematic errors may have a substantial effect, particularly in the case of amplitudes of vibration and the resolution of closely similar internuclear distances. Different types of vibrational averages [*e.g.*, $r_s(0)$ versus $r_s(1)$] are influenced to different extents. The resolution of internuclear distances is influenced less by systematic errors if the internuclear distances correspond to atomic scattering factors with distinctly different angular dependencies. Investigations in which a knowledge of highly correlated parameters is crucial should include a treatment analogous to that proposed above.

INTRODUCTION

A number of papers on error analyses in gas-phase electron diffraction structure determinations have appeared since the introduction of automated least-squares procedures into structure refinements¹⁻⁷. Virtually all the treatments

* Based, in part, on a dissertation by Hsiukang Yow in partial fulfillment of requirements for the degree of Doctor of Philosophy, The University of Michigan, 1972; this work was supported by a grant from the National Science Foundation.

are based on the assumptions that least-squares residuals are statistically distributed and that systematic errors are not revealed in the residuals and influence only the scale factor of internuclear distances. Another type of systematic error can occur, however, which may interfere with the resolution of closely spaced distances and, to some extent, with amplitudes of vibration. This latter type of systematic error arises from imperfections in emulsion calibration curves or emulsion development techniques, from extraneous scattering by stops, apertures, and residual gases, from inaccurate scattering factors, and from other flaws introducing distortions into the envelope modulating the molecular intensity oscillations. Such errors may often be absorbed by shifts in derived molecular parameters and thereby escape recognition. It is the purpose of the present study to propose a method for treating this neglected type of systematic error and to investigate the magnitude of its effect.

METHOD OF PROCEDURE

Scheme to treat systematic errors in intensity

An elaborate and rigorous theory of error analysis has been developed for the case in which errors are statistically distributed⁸. No such general treatment exists when systematic errors are present although some useful admonitions concerning the reporting of such errors have been published⁹ and some suggestions for handling them in structure studies have appeared¹⁰. The difficulty with systematic errors of the type considered here, whether they occur in the experimental data or in the model used to interpret the data, is that they are associated with a definite pattern of misfit in the data points instead of with a random scatter. The only completely satisfactory way to treat such errors is to have an adequate measure of the pattern of misfit. But such patterns are different for each different type of experiment and, furthermore, if an adequate measure of the misfit were known, the misfit would not have to be tolerated.

In the absence of a satisfactory measure of misfit we must strive for some plausible gauge of the effects of these elusive errors. Experience has taught us rough and ready limits of such patterns of error and these limits can be used to estimate the possible influence on derived parameters in individual molecular cases.

Let us assume that systematic errors due to mismeasurements of intensity, extraneous scattering, or inaccurate scattering theory can be lumped together into an "envelope function" $E(s)$ such that the observed reduced intensity $M(s)_{\text{obs}}$ is related to the ideal reduced intensity by

$$M(s)_{\text{obs}} = E(s) M(s)_{\text{ideal}} \quad (1)$$

where $E(s)$ can be regarded (in the notation of electron diffraction) as a variable "index of resolution". To make the analysis tractable it is necessary to adopt an explicit form for $E(s)$. For sake of calculation, it is reasonable to assume that $E(s)$ can be expressed as

$$E(s) = R[1 + \sum_k E_k \cos(n\pi s/s_M)] \quad (2a)$$

or, probably less satisfactorily,

$$E(s) = R[1 + \sum_k E_k \cos(n\pi s/2s_M)] \quad (2b)$$

in which R is the conventional index of resolution, s_M is the maximum value of the angular variable s , and considerations of smoothness would suggest that k be limited to a range from unity to perhaps 3 or 4. It is probably more realistic to use a separate $E(s)$ for each camera distance but it may be acceptable to apply eqn. (2) simply to the composite data from all camera ranges. In current careful work the coefficients E_k might be expected to be at least several hundredths and values of 0.1 are not implausible; values several-fold higher due to extraneous scattering and poor scattering factors have been encountered in some of our work and must not be ruled out unless a critical analysis of the work justifies it.

The basic idea of our treatment is that the E_k parameters can be regarded as parameters characterizing intensity errors which might, in some circumstances, be derivable from the experimental intensity function. The theoretical function $E(s, E_k)M_{\text{calc}}(s, \theta_m)$ could, in principle, be fitted to $M(s)_{\text{obs}}$ by adjusting the envelope coefficients E_k simultaneously with R and with the molecular structure parameters θ_m by a least-squares procedure. It is evident at a glance that $E_1 \cos \pi s/2s_M$ modifies the molecular interference terms $M(s)$ in the same manner as do amplitudes of vibration, l_m , for example. If the E_k and θ_m turned out to be virtually uncorrelated in the least-squares analysis, the θ_m values could be adopted with some confidence. On the other hand, damaging correlations between systematic errors and molecular parameters would be revealed by high correlation coefficients in the error matrix and would lead to appropriately increased computer standard deviations.

Presumably, when the data have been processed and corrected as well as possible, the best guess of the E_k coefficients is zero. It is not necessary in actual structure refinements to include the E_k and risk having them drift away from zero, carrying the θ_m with them to distorted values. The error matrix, augmented by the additional parameters E_k , can be computed after the final structure refinement based on $E_k = 0$, as we shall see below.

The treatment of poorly characterized systematic errors, then, consists of breaking the errors up into components, each characterized by one coefficient, and treating the coefficients by random error theory. In defence of this seeming contradiction, we note that, to the best of our knowledge, the coefficients E_k are

as likely to be negative as positive after correction of the data and the end result seems to give a good qualitative, physically reasonable guideline if not a rigorous, quantitative uncertainty.

Reduction to formalism suitable for computation

Standard least-squares procedures will be followed. Calculated values of n observations at points s_i in terms of m independent parameters can be expressed as

$$F(s_i, E_k, \theta_l) = AX = M - V \quad (3)$$

where

F = the column matrix with n elements corresponding to the n observations, and with functional form based on eqn. (33) of Bartell and Kuchitsu¹¹.

X = the column matrix with m elements representing the $m - N$ displacements $(\theta_l - \theta_l^0)$ and N displacements $(E_k - E_k^0)$ from initial parameter values θ_l^0 and E_k^0 .

A = the design matrix with elements $A_{it} = \partial F_i / \partial X_t$.

M = the column matrix of the n observations which we shall hereafter assume represents measurement of the reduced intensity function $sM(s)$.

V = the column matrix of residuals.

If the observations are weighted by the weight matrix W , the least-squares solution for X is

$$\hat{X} = B^{-1} A' WM, \quad (4)$$

where B stands for the product $A'WA$, and the corresponding matrix of errors in derived parameters is

$$M_x = B^{-1} V'WV / (n - m), \quad (5)$$

provided W is the optimum weight matrix*. In the following illustrations, for simplicity, we have taken W to be the identity matrix. This implies that observations are uncorrelated with random errors uniformly distributed in $sM(s)$ *. The simplifications embodied in the model are unlikely to distort the influence of the systematic error coefficients E_k upon the molecular parameters θ_l .

In order to keep track of systematic effects we shall use the conventions

M_x^0 = the error matrix calculated by deleting the N envelope coefficients E_k from the A and X matrices

M_x^s = the error matrix calculated by including the coefficients E_k in the A and X matrices and by setting $E_k^0 = 0$.

Diagonal elements $(M_x)_{it}$ of the error matrices correspond to $\sigma^2(\theta_l)$ or $\sigma^2(E_k)$, the squares of the desired standard deviations in the derived parameters θ_l or E_k .

* For discussions of weight matrices and the effects of data correlations, see refs. 3-7.

So far, nothing has been included in the treatment to incorporate our knowledge of practical limits in the variance of the systematic error coefficients E_k . As we shall see, it is easily possible to get such a high correlation between some θ_i and E_k that the standard deviations in θ_i and E_k become absurdly high. The simplest way to build our previous knowledge of reasonable values of $\sigma^2(E_i)$ into the treatment and thereby also to constrain the $\sigma^2(\theta_i)$ to reasonable values is to introduce a new error matrix M_x^b based entirely on prior experience and not at all upon the observations M , where

M_x^b = the error matrix with diagonal elements $(M_x^b)_{kk}$ corresponding to what are believed to be plausible values of $\sigma^2(E_k)$, with elements $(M_x^b)_{ii}$ corresponding to $\sigma^2(\theta_i)$ taken as infinity inasmuch as M_x^b is devoid of information about the structure parameters, and with off-diagonal elements taken as zero because the prior information does not correlate the parameters.

A new error matrix M_x^{ab} can now be calculated to correspond to the merging of data set a (with observations M and error matrix M_x^a) and the independent "data set b " (with its error matrix M_x^b). The optimum mixing of sets a and b leads to the simple combination rule

$$M_x^{ab} = [(M_x^a)^{-1} + (M_x^b)^{-1}]^{-1}. \quad (6)$$

Model systems to test new error theory

The type of molecule most likely to give unrealistically low standard deviations if systematic errors are ignored is one with several internuclear distances differing from each other by less than their amplitudes of vibration. We assume for these badly behaved cases that the principal information about the poorly resolved distances resides in the radial distribution peak of the distances in question. In many cases, severely overlapping bond lengths can be resolved easily by taking into account nonbonded distances and the constraints provided by geometric self-consistency. Often such constraints are illusory, however, because the nonbonded distances may be subject to appreciable uncertainties by virtue of inadequate shrinkage corrections¹². Error matrices neglecting errors in shrinkage effects may be unduly optimistic also, but we shall ignore this problem for the present.

Two hypothetical molecules, C_4 and OS_2 , were postulated in order to test the influence of systematic errors on resolving power in particularly simple examples. In each case there are only two internuclear distances. The molecule C_4 , taken to have a nearly tetrahedral C_{3v} structure with basal C-C bond lengths shorter than apical bond lengths by the amount ε , was selected to represent a case with two distances of exactly equal scattering powers. The molecule OS_2 , assumed to be cyclic with C_{2v} symmetry and an S-S bond exceeding the two S-O bonds by ε , was chosen to provide a case with two distances of nearly equal scattering power

but significantly different electron scattering factors¹³. The Morse asymmetry constant¹¹ a was assumed to be zero. Results proved to be sensitive to the particular type of average bond length considered and both $r_x(0)$ [frequently written as r_x] and $r_x(1)$ [frequently written as r_a] were studied¹¹.

The three different error matrices M_x^0 , M_x^a and M_x^{ab} were calculated for both C_4 and OS_2 for assumed values of the bond separations ε of 0.001 Å, 0.025 Å, 0.050 Å, and, in a few instances of 0.0001 Å and 0.1 Å. Four parameters were freely varied in generating M_x^0 in both cases, including:

- (1) \bar{r} , the average of the two postulated bond distances, a value of 1.54 Å being taken for C_4 and 1.675 Å for OS_2 ,
- (2) ε , the difference between the two assumed bond distances,
- (3) l , the amplitude of vibration, assumed to be identical for each component of the two postulated bond distances; l was given a value of 0.05 Å, and
- (4) R , the index of resolution.

An arbitrary scale factor, $V'WV/(n-m)$, was adopted to make the standard deviation in $M(s)$ weighted by s^2 equal to 0.0012, a common experimental value. The standard deviations adopted for the E_k 's in calculating M_x^{ab} were 0.1 for E_1 and 0.05 for E_2 and E_3 , respectively. The scattering variable s ranged from 5.7 to 38.6 with uniform increments of $10/\pi$ Å⁻¹. Variant b of eqn. (2) was used for $E(s)$.

RESULTS

Standard deviations as a function of ε

The ε dependencies of the standard deviations for the various derived parameters as computed from the three different types of error matrices, M_x^0 (neglecting systematic errors), M_x^a (including unconstrained systematic errors) and M_x^{ab} (including constrained systematic errors) are plotted in Figs. 1-8. Although the standard deviations of the individual distances, $r_A = \bar{r} - \frac{1}{2}\varepsilon$ and $r_B = \bar{r} + \frac{1}{2}\varepsilon$, are not shown, they are easily found in the ε regions of interest from the two relations

$$\begin{aligned}\sigma^2(\bar{r}) &\equiv \sigma^2(\frac{1}{2}r_B + \frac{1}{2}r_A) \\ &= \frac{1}{4}[\sigma^2(r_A) + \sigma^2(r_B) + 2\rho_{AB}\sigma(r_A)\sigma(r_B)],\end{aligned}\quad (7)$$

and

$$\begin{aligned}\sigma^2(\varepsilon) &\equiv \sigma^2(r_B - r_A) \\ &= \sigma^2(r_A) + \sigma^2(r_B) - 2\rho_{AB}\sigma(r_A)\sigma(r_B).\end{aligned}\quad (8)$$

At small ε where $\rho_{AB} \rightarrow -1$, it is apparent that

$$\sigma(r_A) \approx \sigma(r_B) \approx \frac{1}{2}\sigma(\varepsilon).$$

At large ε where $\rho_{AB} \rightarrow 0$, it can be seen that

$$\sigma(r_A) \approx \sigma(r_B) \approx \sqrt{2} \sigma(\bar{r}).$$

Except for Fig. 1 where ε was not allowed to vary in the least-squares analysis, the figures show results of analyses in which all parameters \bar{r} , ε , l , R , E_1 , E_2 , and E_3 were taken as free variables in M_x^a . The portrayals of $\sigma(\bar{r})$ corresponding to analyses in which ε is fixed (Fig. 1) and freely varied (Fig. 2) illustrate the markedly different behaviors of the different types of distance averages, $\bar{r}_z(0)$ and $\bar{r}_z(1)$.

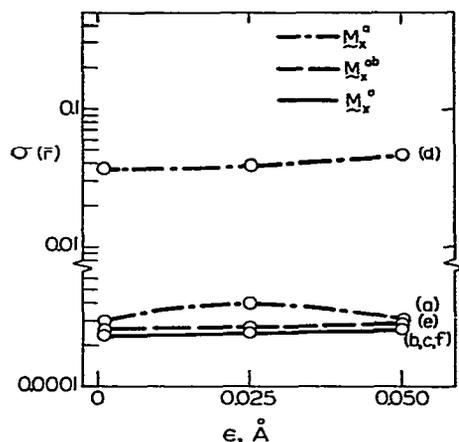


Fig. 1. Standard deviations of \bar{r} versus ε for the hypothetical molecule C_4 . Parameters \bar{r} , l , and R were varied simultaneously but ε was treated as a constant in calculating M_x^o . Curves (a), (b), and (c): $r_z(1)$ basis; curves (d), (e) and (f): $r_z(0)$ basis.

Case of two internuclear distances of identical scattering power

Calculations for the hypothetical molecule C_4 embodied in Figs. 2–5 confirm that:

- (i) the standard deviations of all molecular parameters (*i.e.* r_A , r_B , $\bar{r}_z(1)$, ε , and l increase precipitously as $\varepsilon \rightarrow 0$, except for $\bar{r}_z(0)$ which is insensitive to ε ,
- (ii) the effect of unconstrained systematic errors, as diagnosed from M_x^a , is to increase strikingly the standard deviations of all parameters, especially $r_z(0)$ and l ,

(iii) when systematic errors are unconstrained, their strong correlation with the molecular parameters inflates the $\sigma(E_k)$ values to absurd values,

(iv) the inclusion of “data set b ”, via M_x^b , to limit the variance of the $\sigma(E_k)$ leads to much more moderate and entirely plausible values of standard deviations of the other parameters, and

(v) the influence of systematic errors may be an important source of uncertainty in derived molecular parameters.

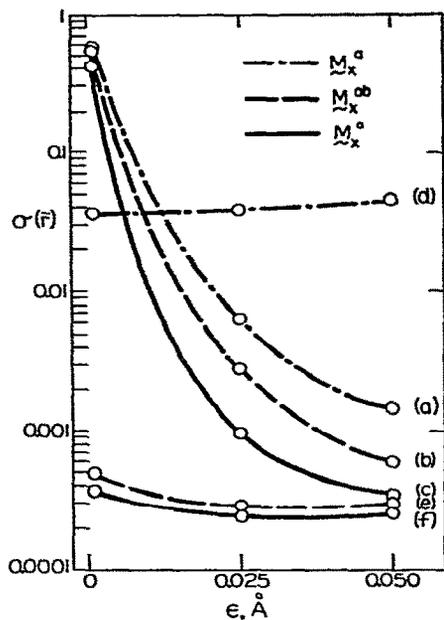


Fig. 2. Standard deviations of \bar{r} versus ϵ for the hypothetical molecule C_4 . In this and the remaining figures, \bar{r} , ϵ , l and R varied simultaneously in calculating M_x^0 . Curves (a), (b) and (c): $r_g(1)$ basis; curves (d), (e) and (f): $r_g(0)$ basis.

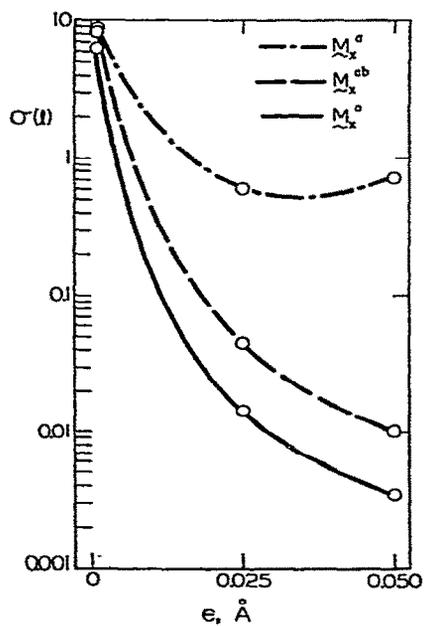


Fig. 3. Standard deviations of l versus ϵ for the hypothetical molecule C_4 .

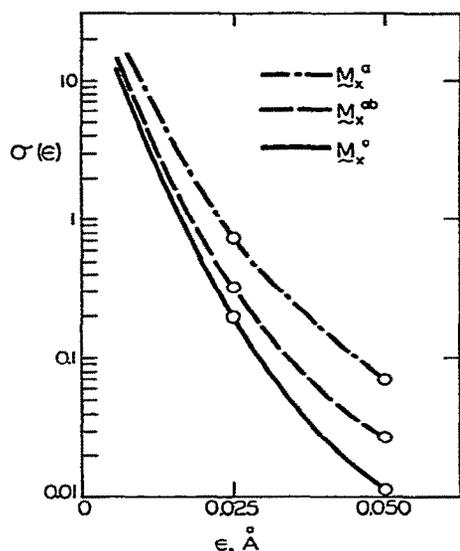


Fig. 4. Standard deviations of ε versus ε for the hypothetical molecule C_4 .

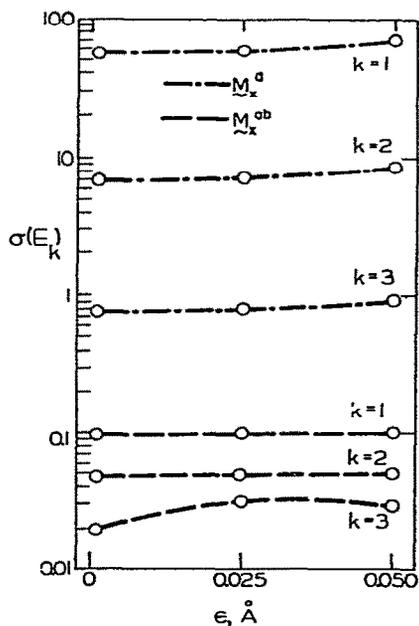


Fig. 5. Standard deviations of E_k 's versus ε for the hypothetical molecule C_4 .

When ε is a small fraction of l , huge, physically unreasonable uncertainties occur as mathematical artifacts due to the differential nature of the error formalism. When good data are available, it is only when ε is comparable to or smaller than l that difficulty in resolving distances occurs. Therefore, standard deviations in ε based on ε - l correlations are meaningless when they greatly exceed the magnitude of l itself.

The characteristic properties of the functions displayed in Figs. 1–8 are dictated primarily by two interactions. If ε is not a small fraction of l , the influence of systematic errors is through the strong correlations between the amplitude of vibration, l , and the parameters of $E(s)$. When ε becomes very small, the correlation between ε and l becomes overriding, breaking the $E(s)$ – l coupling to some extent, and mushrooming standard deviations of parameters dependent on ε and l . One seemingly surprising result may be noted. Recall that, for a single internuclear distance, the length $r_z(1)$ is the parameter governing the period of the interference terms in the scattered intensity¹⁴. The true mean length $r_z(0)$ is very nearly $[r_z(1) + l^2/r_z(1)]$. At first glance it is curious, then, that when the average \bar{r} of distances r_A and r_B is examined, the large uncertainty in l as $\varepsilon \rightarrow 0$ inflates $\bar{r}_z(1)$ but leaves $\bar{r}_z(0)$ unaffected. The reason for this is outlined in the next section. In any event, in studies for which extreme accuracy in a mean distance is important, it is well to bear in mind that the two different distance parameters have different susceptibilities to error.

It might be wise to include a constraint in M_x^b on R as well as on the E_k parameters, for the constraints on E_k in our calculations still allowed a value of $\sigma(R) \approx 0.1$ in M_x^{ab} (or higher if $\varepsilon \ll l$). Experience shows that R is often stable to 0.06 or so, though capricious shifts exceeding 0.1 can occur in a given series of plates.

Case where scattering factors are distinctly different

Calculations for the hypothetical molecule OS_2 reveal a very different pattern of behavior than that encountered in the case of C_4 . Figures 6–8 demonstrate that

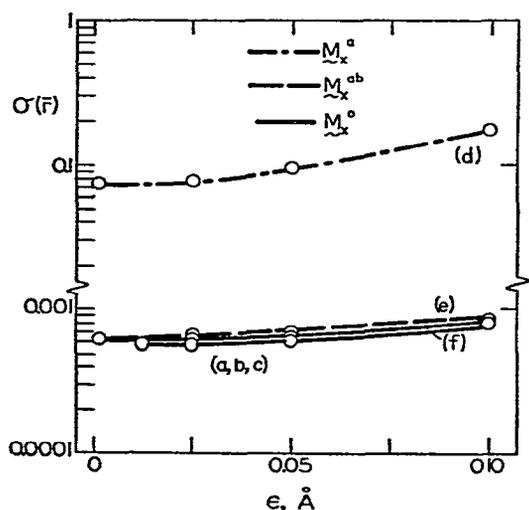


Fig. 6. Standard deviations of \bar{r} versus ε for the hypothetical cyclic molecule OS_2 . Curves (a), (b) and (c): $r_z(1)$ basis; curves (d), (e) and (f): $r_z(0)$ basis.

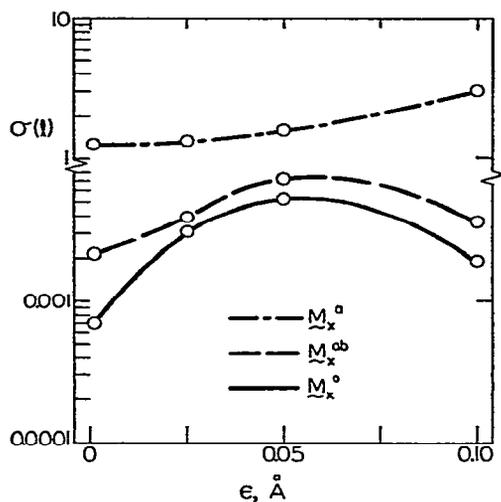


Fig. 7. Standard deviations of l versus ϵ for the hypothetical cyclic molecule OS_2 .

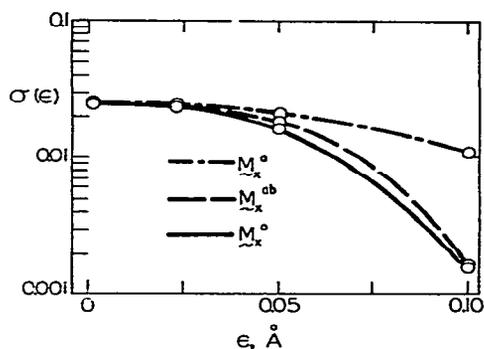


Fig. 8. Standard deviations of ϵ versus ϵ for the hypothetical cyclic molecule OS_2 .

the resolving power (as reflected in $\sigma(\epsilon)$) is worsened as ϵ becomes smaller. On the other hand, the curves display no steeply rising standard deviations as ϵ goes to zero. This shows that, even at ϵ near zero, the S–O and S–S interference terms are sufficiently different in their s -dependency, owing to their different scattering factors¹³, that they can be distinguished from each other. Another way of looking at the problem is in terms of the Fourier sine transforms of $sM(s)$ which correspond to radial distribution peaks distorted by the scattering by planetary electrons. The distorted radial distribution peak for S–O can be discriminated in shape from that of S–S and, hence, a superposition of the two can be resolved into the individual components. Such a resolution is not possible for C_4 at small ϵ .

Regression slopes

Off-diagonal elements in the error matrices, as well as the diagonal elements discussed in the foregoing, are of concern to structural chemists. Parameter correlation coefficients ρ_{kl} are simply $(M_x)_{kl}[(M_x)_{kk}(M_x)_{ll}]^{-\frac{1}{2}}$. Regression slopes⁸, which

we shall designate as $\delta x_k/\delta x_l$, are given by $(M_x)_{kl}/(M_x)_{ll}$ and represent the variation in parameter x_k with forced changes in parameter x_l subject to satisfying the least-squares condition $V'WV = \text{minimum}$ at the various preset values of x_l . Regression slopes calculated from M_x^0 , M_x^{ab} , from M_x^a , and from simplified formulas derived in the Appendix, are listed in Tables 1-3 for C_4 for several combinations of parameters.

The zero-th order approximation for understanding the regression slopes can be derived from the trigonometric identity related to the molecular interference terms

$$e^{-I^2s^2/2} (\sin sr_A + \sin sr_B) = 2e^{-I^2s^2/2} (\sin s\bar{r}) (\cos sr/2). \quad (9)$$

TABLE 1

REGRESSION SLOPES $\delta\epsilon/\delta l$ AND $[\delta l/\delta\epsilon]$ CALCULATED FROM M_x^0 , M_x^{ab} , M_x^a , AND SIMPLIFIED FORMULA, EQN. (10), $r_g(0)$ BASIS*

ϵ	M_x^0	M_x^{ab}	M_x^a	Eqn. (10a)
0.001	-199.7, $[-199.7]^{-1}$	-199.7, $[-199.7]^{-1}$	-198.0, $[-198.8]^{-1}$	-200.0
0.025	-7.6, $[-7.6]^{-1}$	-7.3, $[-7.3]^{-1}$	-0.3, $[-5.7]^{-1}$	-8.0
0.050	-3.2, $[-3.2]^{-1}$	-2.6, $[-2.7]^{-1}$	-0.02, $[-0.40]^{-1}$	-4.0

* Results for $r_g(1)$ basis virtually identical.

TABLE 2

REGRESSION SLOPES $(\delta\bar{r}/\delta l) \times 10^3$ CALCULATED FROM M_x^0 , M_x^{ab} , M_x^a , AND FROM SIMPLIFIED FORMULAE

Basis	$\epsilon(\text{\AA})$	M_x^0	M_x^{ab}	Eqn. (14)	M_x^a	Eqn. (12)
$r_g(0)$	0.001	0.0	0.0	0	0.6	0.7
	0.025	0.2	2.3	0	63	63
	0.050	1.5	12.1	0	64	65
$r_g(1)$	0.0001	-65	-65	-65	-64	-64
	0.025	-65	-63	-65	-2.5	-2.4
	0.050	-63	-53	-65	-0.5	-0.6

TABLE 3

REGRESSION SLOPES $(\delta\bar{r}/\delta\epsilon) \times 10^3$ CALCULATED FROM M_x^0 , M_x^{ab} , M_x^a , AND FROM SIMPLIFIED FORMULAE

Basis	ϵ	M_x^0	M_x^{ab}	Eqn. (14a)	M_x^a	Eqn. (13)
$r_g(0)$	0.001	-0.000	-0.000	0	-0.002	-0.002
	0.025	-0.024	-0.30	0	-2.9	-3.3
	0.050	-0.26	-3.8	0	-141	-145
$r_g(1)$	0.0001	0.033	0.033		0.033	0.033
	0.025	8.6	8.6		8.6	8.6
	0.050	20	20		20	20

Taking the variation of the interference terms with respect to l and ε we see, for $\varepsilon \ll l$, that the sum $2l \delta l + \varepsilon \delta \varepsilon / 2$ must vanish if a change in ε is to compensate a change in l and leave the interference terms unchanged. That is, if ε is small, the ε , l regression slopes must be

$$\delta \varepsilon / \delta l \approx -4l / \varepsilon \quad (10a)$$

and

$$\delta l / \delta \varepsilon \approx -(4l / \varepsilon)^{-1}. \quad (10b)$$

Table 1 confirms that is the case, approximately, even out to $\varepsilon \approx l$ as long as there are no other parameters correlating strongly with l . The regression slopes from M_x^a verify that eqn. (10) holds for $\varepsilon \ll l$ even in the face of the systematic error coefficients E_k , but that with increasing ε , the l - ε correlation becomes less perfect and is broken by the stronger ε - E_k correlations.

The different behavior of $\bar{r}_\varepsilon(0)$ and $\bar{r}_\varepsilon(1)$ is evident in the regression slopes as well as the standard deviations. As shown in the Appendix, the first-order approximations relating the response of \bar{r} to forced variations in l and ε are

$$\delta \bar{r}_\varepsilon(0) \approx (2l/r) \delta l + (\varepsilon/2r) \delta \varepsilon \quad (11a)$$

and

$$\delta \bar{r}_\varepsilon(1) \approx 0 \delta l + (\varepsilon/2r) \delta \varepsilon. \quad (11b)$$

It follows that the \bar{r} - l and \bar{r} - ε regression slopes are roughly

$$\delta \bar{r}_\varepsilon(0) / \delta l \approx 2l/r + (\varepsilon/2r) \delta \varepsilon / \delta l, \quad (12a)$$

$$\delta \bar{r}_\varepsilon(1) / \delta l \approx (\varepsilon/2r) \delta \varepsilon / \delta l, \quad (12b)$$

$$\delta \bar{r}_\varepsilon(0) / \delta \varepsilon \approx (2l/r) \delta l / \delta \varepsilon + \varepsilon/2r, \quad (13a)$$

and

$$\delta \bar{r}_\varepsilon(1) / \delta \varepsilon \approx \varepsilon/2r. \quad (13b)$$

If ε is so small that the ε - l correlation overrides the E_k - l correlations, we may simplify by inserting eqn. (10) into eqns. (12)-(13) to get

$$\delta \bar{r}_\varepsilon(0) / \delta l \approx \delta \bar{r}_\varepsilon(0) / \delta \varepsilon \approx 0 \quad (14a)$$

and

$$\delta \bar{r}_\varepsilon(1) / \delta l \approx -2l/r \quad (14b)$$

where the insensitivity of $\bar{r}_\varepsilon(0)$ to ε is a consequence of a strong l - ε correlation and the attendant cancellations of terms in eqn. (12a). We may understand the surprisingly large $\sigma(r_\varepsilon(0))$ values from M_x^a most easily in terms of a huge uncertainty in l stemming from the large correlation between l and the systematic error coefficients, together with the inference from eqn. (11) that

$$\sigma(\bar{r}_\varepsilon(0)) \approx (2l/r) \sigma(l) \quad (15)$$

except when $\varepsilon \ll l$ and the ε - l correlation effects are pervasive. In the $\varepsilon \geq 0.025 \text{ \AA}$ calculations for C_4 and OS_2 , whether ε was fixed or freely varied, the ratio of $\sigma(r_s(0))/\sigma(l)$ from M_x^a was found to be within a percent of the value of $2l/r$.

CONCLUSIONS

A simple, apparently successful procedure has been devised to treat the contribution of systematic intensity errors to the total uncertainty in molecular parameters derived from electron diffraction data. Such intensity errors may be appreciable and are most conspicuous in amplitudes of vibration l and, especially, in the resolution of poorly resolved internuclear distances. Mean bond lengths are also significantly influenced by the errors considered in the foregoing, and the influence is different for different types of vibrational averages [*e.g.*, $r_s(0)$ versus $r_s(1)$]. In the model systems considered the resolving power was appreciably better if the internuclear distances involved corresponded to dissimilar atomic scattering factors than if the scattering factors were identical. The examples considered should serve as useful guidelines in performing practical error analyses. In those cases where accurate values of ε and l are crucial, it would be prudent to carry out an analysis parallel to the treatment outlined in the foregoing.

APPENDIX

Distinction between $r_s(0)$ and $r_s(1)$ parameters

Significantly different patterns of behavior of the different internuclear distance parameters $r_s(0)$ and $r_s(1)$ were observed in the various error matrices. Some insight into the differences can be gained by considering the physical distinctions involved without going through the least squares formalism of eqns. (3)–(5). The interference terms $sM(s)$ in electron diffraction are given, for a given internuclear distance, r_k , by

$$[sM(s)]_k \propto \int [P(r)/r]_k \sin sr \, dr. \quad (\text{A.1})$$

The fact that the probability distribution $P(r)$ is multiplied by $1/r$ leads to the consequence that the most natural distance parameter in electron diffraction is $r_s(1)$, the distance averaged over the function $[P(r)/r]$. This distance differs from $r_s(0)$, the true vibrational average, by an amount depending upon the breadth of $P(r)$ according to¹⁴

$$r_s(1) \approx r_s(0) - l^2/r_s \quad (\text{A.2})$$

neglecting higher order terms. If we have two internuclear distances r_A and r_B

corresponding to identical amplitudes but differing in length by ε , we can write, following eqn. (A.2),

$$r_z(1)_A = \bar{r}_z(0) - \varepsilon/2 - l^2/r \quad (\text{A.3a})$$

and

$$r_z(1)_B = \bar{r}_z(0) + \varepsilon/2 - l^2/r. \quad (\text{A.3b})$$

For small ε the two distances blend into one distribution peak, and this composite peak generates nodes in $sM(s)$ corresponding to a natural length we denote as $\langle r_z(1) \rangle$. In order to determine the relation between the effective mean $\langle r_z(1) \rangle$ and the components $r_z(1)_A$ and $r_z(1)_B$, we note that

$$\langle r_z(1) \rangle = \frac{w_A r_z(1)_A + w_B r_z(1)_B}{w_A + w_B} \quad (\text{A.4})$$

where the natural weights are seen from eqn. (A.1) to be proportional to $1/r$. Accordingly, if the two internuclear distances have identical scattering factors, we may replace w_k by $1/r_z(1)_k$ in eqn. (A.4) and obtain

$$\langle r_z(1) \rangle = 2r_z(1)_A r_z(1)_B / [r_z(1)_A + r_z(1)_B], \quad (\text{A.5})$$

a quantity which differs from the arithmetic average. Substituting eqns. (A.3a) and (A.3b) into eqn. (A.5) yields, to first order in the correction terms,

$$\langle r_z(1) \rangle \approx \bar{r}_z(0) (1 - l^2/r^2 - \varepsilon^2/4r^2), \quad (\text{A.6})$$

or

$$\bar{r}_z(0) \approx \langle r_z(1) \rangle (1 + l^2/r^2 + \varepsilon^2/4r^2). \quad (\text{A.7})$$

Presumably $\langle r_z(1) \rangle$ is established by the diffraction nodal positions and, hence, is constant in a least-squares analysis. Therefore, the best least-squares value of $\bar{r}_z(0)$ must vary, when l and ε are forced to vary, as the differential of eqn. (A.7), or

$$\delta \bar{r}_z(0) \approx 2l \delta l/r + \varepsilon \delta \varepsilon / 2r. \quad (\text{A.8})$$

The corresponding relation for $\bar{r}_z(1)$ can be obtained from the definition

$$\bar{r}_z(1) = [r_z(1)_A + r_z(1)_B] / 2 \approx \bar{r}_z(0) - l^2/r, \quad (\text{A.9})$$

or

$$\begin{aligned} \delta \bar{r}_z(1) &\approx \delta \bar{r}_z(0) - 2l \delta l/r \\ &\approx (2l \delta l/r + \varepsilon \delta \varepsilon / 2r) - 2l \delta l/r \\ &\approx \varepsilon \delta \varepsilon / 2r. \end{aligned} \quad (\text{A.10})$$

Equations (A.8) and (A.10) are introduced into the text as eqns. (11a) and (11b) and regression slopes are calculated from them. If regression slopes are desired for an analysis in which ε is held at a fixed value, the above equations can be used with $\delta \varepsilon$ set equal to zero.

ACKNOWLEDGMENTS

We are pleased to acknowledge a generous allowance of computing time from the Michigan Computing Center and the partial support of one of us (H.Y.) by a Barbour fellowship.

REFERENCES

- 1 O. BASTIANSEN, L. HEDBERG AND K. HEDBERG, *J. Chem. Phys.*, 27 (1957) 1311.
- 2 R. A. BONHAM AND L. S. BARTELL, *J. Chem. Phys.*, 31 (1959) 702;
L. S. BARTELL, K. KUCHITSU AND R. J. DE NEUI, *J. Chem. Phys.*, 35 (1961) 1211.
- 3 Y. MORINO, K. KUCHITSU AND Y. MURATA, *Acta Crystallogr.*, 18 (1965) 549;
Y. MURATA AND Y. MORINO, *Acta Crystallogr.*, 20 (1966) 605.
- 4 L. S. BARTELL, *Acta Crystallogr.*, A25 (1969) S76.
- 5 H. M. SEIP, T. G. STRAND AND R. STÖLEVIK, *Chem. Phys. Lett.*, 3 (1969) 617;
H. M. SEIP AND R. STÖLEVIK, in S. J. CYVIN (editor), *Molecular Structures and Vibrations*, Elsevier, Amsterdam, 1972, p. 171.
- 6 M. A. MACGREGOR AND R. K. BOHN, *Chem. Phys. Lett.*, 11 (1971) 29.
- 7 L. S. BARTELL AND M. G. ANASHKIN, Part II of this series (submitted to *J. Mol. Structure*).
- 8 W. C. HAMILTON, *Statistics in Physical Science*, The Ronald Press Co., New York, 1964.
- 9 C. EISENHART, *Science*, 160 (1968) 1201.
- 10 E. J. JACOB, H. B. THOMPSON AND L. S. BARTELL, *J. Mol. Structure*, 8 (1971) 383.
- 11 L. S. BARTELL AND K. KUCHITSU, *J. Chem. Phys.*, 35 (1961) 1945.
- 12 Y. MORINO, S. J. CYVIN, K. KUCHITSU AND T. IJIMA, *J. Chem. Phys.*, 36 (1962) 1109.
- 13 L. SCHAFER, A. C. YATES AND R. A. BONHAM, *J. Chem. Phys.*, 55 (1971) 3055;
C. TAVARD, D. NICOLES AND M. ROUAULT, *J. Chim. Phys.*, 64 (1967) 540, 555;
D. T. CROMER, *J. Chem. Phys.*, 50 (1969) 4857. The aforementioned references provided the scattering factors adopted in our analyses.
- 14 L. S. BARTELL, *J. Chem. Phys.*, 23 (1955) 1219.