

# Reminiscences of Least Squares

Robert F. Stewart, Chemistry Department  
Carnegie Mellon University, 4400 Fifth Avenue  
Pittsburgh, PA 15213-2617

Since the invention and applications of the method of least squares by Carl Friedrich Gauss in the early 19th century an extremely useful tool for analysis of scientific data has remained in the scientists' arsenal of analytical methods for the last 200 years.



The present paper will reveal my own experiences with least squares methods applied to a variety of studies in chemical physics. The earliest study was the determination of the solubility of  $\text{PbSO}_4(s)$  as a function of  $[\text{HClO}_4(aq)]$  in mixtures of  $\text{LiClO}_4(aq)$  or  $\text{NaClO}_4(aq)$  at an ionic strength of unity. Least squares was employed to determine the principal values and directions in the polarized absorption spectra of a single crystal of a 1:1 complex of 1-methylthymine and 9-methyladenine (the Hoogsteen dimer). A functional equation for extracting the H atom form factor from the molecular form factor of  $\text{H}_2$  was a solution to a least squares problem. Small gaussian expansions of SCF atomic orbitals and of Slater type orbitals (STO) are an example of highly non-linear least squares equations. Generalized X-ray scattering factors (gsf's) are a rotationally invariant basis set for projection onto simple functions of measured X-ray crystal structure factors by the method of least squares. A three-parameter equation of state for strong electrolytes was successfully fit, by least squares, to osmotic pressure measurements at concentrations in excess of 6 M. A rather accurate construction of the density of phonon states in  $\text{Si}(s)$  was obtained from a least squares fit of 21 force constants (Born-von Kármán model) to coherent, inelastic neutron scattering data, crystal elastic constants and an optical Raman line. Inclusion of

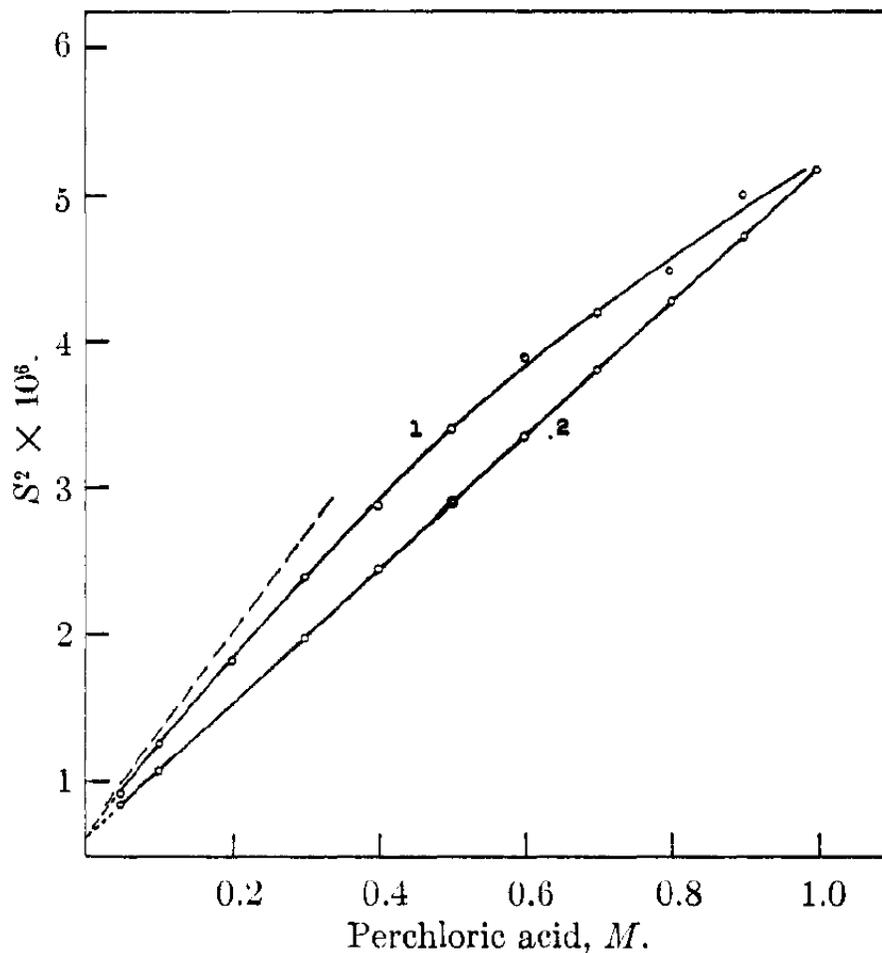
correlation coefficients among the errors of the force constants was essential for reliable error estimates of thermodynamic properties of Silicon. A final example will illustrate the use of least squares to establish a non-existent surface harmonic in the valence electron structure of a bonded F atom.

## 1 The solubility of $\text{PbSO}_4(\text{s})$ as a function of $[\text{HClO}_4(\text{aq})]$ in mixtures of $\text{LiClO}_4(\text{aq})$ or $\text{NaClO}_4(\text{aq})$ at unit ionic strength

The solubility of lead sulfate at 25 °C in varying acidity mixtures of perchloric acid and sodium perchlorate, and of perchloric acid and lithium perchlorate was determined. The strong electrolyte solutions were fixed at an ionic strength of unity. The dominant ionic species, coupled by equilibria, in the acidic aqueous solutions saturated with the sparingly soluble salt  $\text{PbSO}_4(\text{s})$  are  $\text{Pb}^{+2}(\text{aq})$ ,  $\text{HSO}_4^-(\text{aq})$  and  $\text{SO}_4^{-2}(\text{aq})$ . From elementary considerations of ionic equilibria, the solubility squared should be a linear function of  $[\text{H}_3\text{O}^+(\text{aq})]$

$$S^2 = Q_{sp} + \frac{Q_{sp}}{Q_2}[\text{H}_3\text{O}^+(\text{aq})]$$

if the solubility quotient,  $Q_{sp}$ , for lead sulfate and the bisulfate dissociation quotient,  $Q_2$ , are constant over the acidity range fixed at an ionic strength of one.



A least squares fit to the solubility data confirmed our expectations for the  $\text{Li}^+(aq)$  salt mixtures, but deviations in the  $\text{Na}^+(aq)$  mixtures indicated the probable formation of the ion pair  $\text{NaHSO}_4(aq)$  as a significant player in the equilibrium mixtures.

R. W. Ramette and R. F. Stewart, *J. Phys. Chem.*, **65**, 243-246 (1961).

## 2 Polarized absorption spectra of purines and pyrimidines

Dichroism measurements of thin (>100 nm) single crystals of 1-methylthymine, 9-methyladenine and the Hoogsteen dimer were used to determine the tran-

sition dipole moment directions in the UV wavelength range of 300 - 230 nm for several  $\pi^* \leftarrow \pi$  absorption bands.

The Hoogsteen dimer crystallizes in space group  $P2_1/m$ . The crystal is dichroic for polarized light propagating along the  $b$  axis, but is not fixed by crystal symmetry. If  $T_1$  and  $T_2$  are the transmittances corresponding to the principal directions  $P_1$  and  $P_2$ , respectively, then the measured transmittance as a function of  $\phi$ , the angle between the electric vector of the light and a crystal edge, can be written in the linear form,

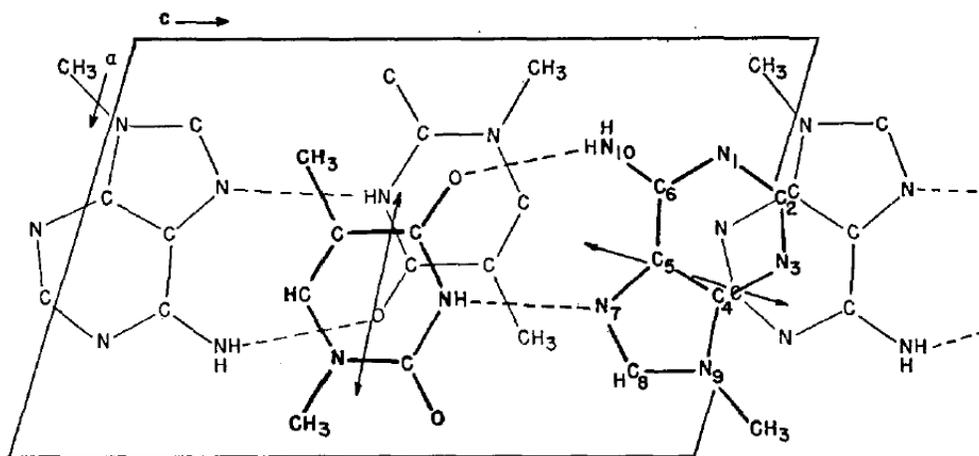
$$T(\phi) = \beta \xi + \gamma \eta + \delta$$

where  $\xi$  is  $\cos 2\phi$ ,  $\eta$  is  $\sin 2\phi$  and  $\delta$  is  $(T_2 + T_1)/2$ . The least squares variables,  $\beta$ ,  $\gamma$ , and  $\delta$  are related to the principal transmittances,  $T_1$ , and  $T_2$  by trigonometric relations and the angle from a crystal edge to the principal direction  $P_2$  is,

$$\theta = \frac{1}{2} \arctan\left(\frac{\gamma}{\beta}\right)$$

The weights used were  $1/T^2$ . The least squares fits were applied to transmission measurements for light wavelengths from 290 to 230 nm every 5 nm.

The final determination of the transition dipole moment directions for thymine and adenine was completed after the crystal structure of 9-methyladenine was elucidated.



R. F. Stewart and N. Davidson, *J. Chem. Phys.*, **39**, 255-266 (1963).  
 R. F. Stewart and L. H. Jensen, *J. Chem. Phys.*, **40**, 2071-2075 (1964).

### 3 A bonded H atom form factor

Lyle Jensen pointed out to me that something was amiss with the H atom form factor. He was getting temperature factors that were systematically less than the atom to which the H was bonded. The results were based on least squares fits to a product of an isotropic Debye-Waller factor and a standard form factor for the ground state H atom. I knew that Ms. Wang's valence bond wavefunction for molecular dihydrogen indicated a contracted 1s atomic orbital from a variation treatment. Weinbaum's  $H_2$  wavefunction (a mix of molecular orbital and valence bond  $\Psi$ ) also pointed to contraction in the 1s bases. If  $H_2$  could serve as a model for a bonded hydrogen atom, then perhaps an H atom form factor could be extracted from an accurate molecular form factor of  $H_2$ . During my post-doctoral stay at the University of Washington, Ernie Davidson was a newly appointed assistant professor in the chemistry department. He had recently derived the first ten natural orbitals from the Kolos-Roothaan wavefunction for  $H_2$  at its equilibrium internuclear distance  $R_e$ . The resulting electron density distribution function was a 91 term polynomial in confocal elliptical coordinates. By expanding the plane wave in spheroidal wavefunctions, it was possible to compute a molecular form factor for  $H_2$ .

The task was to construct an approximate molecular form factor,  $G(\mathbf{K})$  and fit it by least squares to the molecular form factor,  $F(\mathbf{K})$  for  $H_2$ . We define the mean square error function,

$$\epsilon = \int W(K) |F(\mathbf{K}) - G(\mathbf{K})|^2 d^3\mathbf{K}, \quad W(K) \geq 0$$

$W(K)$  is a positive weighting function and depends only on the magnitude of  $K$ . Within the spherical approximation of the bonded atom,  $G(\mathbf{K})$  may be written as

$$G(\mathbf{K}) = f(K) [e^{i\mathbf{K}\cdot\delta} + e^{-i\mathbf{K}\cdot\delta}]$$

where  $\pm|\delta|$  are the distances from the midpoint of the bond to the centers of bonded H atoms (not necessarily the sites of the two protons).  $G(\mathbf{K})$  must transform as the totally symmetric representation in point group  $D_{\infty h}$ . With  $\eta$  the direction cosine of  $\mathbf{K}$  with respect to the molecular axis which contains  $\mathbf{R}$ ,  $\epsilon$  is a minimum for any  $W(K) \geq 0$  if

$$\epsilon(K) = \int_{-1}^1 |F(\mathbf{K}) - 2f(K) \cos(g\eta)|^2 d\eta$$

is a minimum for each  $K$ .  $g$  is some fraction between 0 and 1 and  $c$  is  $K R_e/2$ . Accordingly, for a given  $g$ , the minimum condition is satisfied when

$$f(K) = \int_{-1}^1 F(\mathbf{K}) \cos(gc\eta) d\eta / 2 \int_{-1}^1 \cos^2(gc\eta) d\eta.$$

When  $g$  is 1, the form factor is situated at the protons. The optimum value for  $g$  was 0.81 which gave the lowest value for

$$\epsilon = \int_0^\infty \epsilon(K) K^2 dK.$$

Both the best floated and proton centered form factor curves were more extended in  $K$  space than was the form factor for an isolated H atom. The best floated sphere scattering factor was tabulated as a function of  $\sin \theta/\lambda$ . This curve usually led to larger, positive values for the isotropic displacement parameter for H atoms that are covalently bonded in organic molecules.

Robert F. Stewart, Ernest R. Davidson, and William T. Simpson, *J. Chem. Phys.*, **42**, 3175-3187 (1965).

## 4 Small gaussian expansions of SCF atomic orbitals and Slater type orbitals

In 1950 Boys published a paper pointing out that molecular integrals over gaussian functions were much easier to calculate than corresponding ones with Slater-type orbitals. The reason for this is that a product of two gaussians on separate centers is also a gaussian function at a different site. A two-center orbital product in gaussian bases may be easily expressed as a form factor with a gaussian dependence on  $K$ . A least squares expansion of STF's (Slater-type functions) in gaussian bases should provide one with an optimal density distribution function, but by no means would such expansions be optimal for minimizing the SCF energy of an atom.

I originally expanded SCF-AO's in a small number, two to five, of gaussian functions. The atomic orbitals were those of Clementi and were made up of STF's. For a single SCF atomic orbital the error function to be minimized was

$$\epsilon = \int |\phi - \chi|^2 d\tau + \lambda \left( 1 - \int \chi^* \chi d\tau \right).$$

where  $\phi$  is the SCF orbital of interest,  $\chi$  is a linear combination of GTO's (gaussian-type orbitals), and  $\lambda$  is a Lagrange multiplier. The minimization

of  $\epsilon$  is linear in the expansion coefficients,  $\{d\}$ , but highly nonlinear in the exponential parameters,  $\{\alpha\}$ . We put these expansion parameters, including  $\lambda$  into an array  $\mathbf{P}$ . In matrix notation, the equations to solve are,

$$\mathbf{V} = \mathbf{W}\Delta\mathbf{P}$$

where

$$V_j = \left( \frac{\partial \epsilon}{\partial P_j} \right)_{\mathbf{P}_0}$$

and

$$W_{jk} = - \left( \frac{\partial^2 \epsilon}{\partial P_j \partial P_k} \right)_{\mathbf{P}_0}$$

The equations are solved iteratively until  $\Delta\mathbf{P}$  is an effective null. In the space spanned by  $\mathbf{P}$ , the matrix  $\mathbf{W}$  frequently had negative eigenvalues, at which point a search along the direction of the eigenvector with the most negative eigenvalue was undertaken to seek a lower value in  $\epsilon$ . The expansions were used to generate two-center form factors, but were never really used in applications to X-ray diffraction data. Other constraints were imposed for quantum chemical calculations and became the bases in minimal basis STO-3G calculations.

Robert F. Stewart, *J. Chem. Phys.*, **50**,2485-2495,(1969).

Robert F. Stewart, *J. Chem. Phys.*, **52**,431-438,(1970).

W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**,2657-2664(1969).

W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **52**,2769-2773(1970).

## 5 Generalized X-ray scattering factors

The expansion of a plane wave in spherical wavefunctions is a thing of beauty:

$$e^{i\mathbf{K}\cdot\mathbf{r}} = \sum_{n=0}^{\infty} \sum_{k=0}^n i^n (2 - \delta_{0k}) (2n+1) \frac{(n-k)!}{(n+k)!} \times \cos[k(\phi_{\mathbf{K}} - \phi_{\mathbf{r}})] P_n^k(\eta_{\mathbf{K}}) P_n^k(\eta_{\mathbf{r}}) j_n(Kr)$$

Note that both  $\mathbf{K}$  and  $\mathbf{r}$  are represented with spherical polar coordinates. The surface harmonics in the plane wave expansion make up a complete set

in the space spanned by the solid angle,  $\Omega$ , of a sphere. We write them as real functions,

$$y_{nk\pm}(\Omega) = P_n^k(\eta) \begin{cases} \cos k\phi \\ \sin k\phi \end{cases}$$

where  $P_n^k$  is an *unnormalized* Associated Legendre function.

The  $y_{nk\pm}(\Omega_{\mathbf{K}})$  transform directly to  $\mathbf{r}$  space as

$$(-i)^n y_{nk\pm}(\Omega_{\mathbf{r}})/4\pi.$$

The spherical surface harmonics may be incorporated into a structure factor equation that employs *generalized X-ray scattering factors* (gsf),

$$F_c = \sum_{sym} \sum_p f_p(\mathbf{K}) T_p(\mathbf{K}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_p)$$

The  $\mathbf{x}_p$  are atomic (or nuclear) positions of the atoms in the unique part of the crystallographic unit cell. The function  $T_p(\mathbf{K})$  characterizes the mean thermal motion of the nuclei and its associated electron density, which is represented by  $f_p(\mathbf{K})$ . The product  $f_p(\mathbf{K}) T_p(\mathbf{K})$  is an X-ray scattering model for a *rigid pseudoatom*. The function  $f_p(\mathbf{H})$  is a generalized X-ray scattering factor (*gsf*) for pseudoatom  $p$ ,

$$f_p(\mathbf{K}) = \sum_{nk\pm} C_{p,nk\pm} f_{p,l}(|\mathbf{K}|) i^n y_{nk\pm}(\Omega_{\mathbf{K}}).$$

The  $C_{p,nk\pm}$  are determined by least squares projection of the *gsf* bases onto the reduced X-ray scattering observations.

The  $y_{nk\pm}$  are real spherical surface harmonics with the angular components of the Bragg reflections as arguments.

Since the  $y_{nk\pm}(\Omega_{\mathbf{K}})$  span all the irreducible representations of the rotation groups,  $f_p(\mathbf{K})$  is a rotationally invariant function.

The radial factor for pole  $n$  of pseudoatom  $p$  is,

$$f_{p,n}(|\mathbf{K}|) = \int_0^\infty \rho_{p,n}(r) j_n(Kr) r^2 dr$$

The  $\rho_{p,n}(r)$  are usually based upon quantum chemical calculations. The usual procedure is to employ products of SCEAO's that were used to determine the atomic form factor of an isolated atom.

In practice the  $f_{p,n}(|\mathbf{K}|)$  are radially scaled with the scaling parameters adjusted to afford an optimal fit to the diffraction data.

The electronic parameters are determined most precisely from the low to intermediate angles of the scattering measurements.

The  $f_{p,n}(|\mathbf{K}|)$ , for  $n > 0$ , have an appreciable amplitude of scattering for  $|\mathbf{K}|$  up to about  $9.5 \text{ \AA}^{-1}$  ( $\sin \theta/\lambda$  equal to  $0.76 \text{ \AA}^{-1}$ ). The sundry parameters are components of a vector  $\mathbf{P}$  which spans the  $\chi^2$  error function,

$$\chi^2 = \sum_{\mathbf{H}} w_{\mathbf{H}} (|F_{\mathbf{H},rel}|^n - |F_c(\mathbf{P})|^n)^2, \quad n = 1 \text{ or } 2$$

and

$$w_{\mathbf{H}} = 1/\sigma_{\mathbf{H}}^2$$

where  $\sigma_{\mathbf{H}}^2$  is the estimated variance of the reduced measurement  $|F_{\mathbf{H},rel}|^n$ . For the notation here,

$$|\mathbf{H}| = |\mathbf{K}|/2\pi$$

The components of  $\mathbf{P}$  are varied to render  $\chi^2$  a minimum.

In order to precisely determine the parameters in the  $F_c$  model the ratio of observations to variables should in general exceed twenty. When a minimum in  $\chi^2$  has been found, the inverse elements of the Hessian of  $\chi^2$  can be used to estimate the variance of some direct space property,  $G(\mathbf{r})$ , that is derived from the  $F_c$  model.

$$\sigma^2(G(\mathbf{r})) = \frac{\chi^2}{N_o - N_v} \sum_{j \geq k} \frac{\partial G}{\partial P_j} (2 - \delta_{jk}) (w^{jk}) \frac{\partial G}{\partial P_k}$$

where  $w^{jk}$  is an inverse matrix element of the Hessian of  $\chi^2$ .

The  $\{P_j\}$  used for a variance estimate may be a subset of the components of  $\mathbf{P}$ . The set of electronic parameters used to construct the *gsf* is an example.

Robert F. Stewart, *J. Chem. Phys.*, **51**,4569-4577(1969).

## 6 A three parameter equation of state for strong electrolytes

A three parameter equation of state for strong electrolytes was used to analyze osmotic coefficient data of chloride salts in water at 25°C and up to concentrations in excess of 6 M.

The observed osmotic coefficient  $\phi$  is defined by

$$\Pi = \nu c RT\phi(c)$$

where  $\Pi$  is the observed osmotic pressure,  $\nu$  is the number of ions per electrolyte  $c$  is the molar concentration in moles/liter and  $T$  is the absolute temperature. A trial function,  $\phi_{\text{trial}}$ , which represented our best guess to the ionic interactions that influences the osmotic pressure was constructed. The trial function contains one or more adjustable parameters, each of which has a precise physical interpretation. The parameters were adjusted to obtain the best mean square fit of  $\phi_{\text{trial}}$  to  $\phi$  over a pertinent concentration range.

The deviation  $\delta\phi$  to  $\phi$  from unity is rather well represented with an electrostatic term,  $\delta\phi_a$ , and an excluded volume due to packing of hard spheres,  $\delta\phi_b$ . A primitive trial function accounted for much of the observed behavior of  $\phi$ ,

$$\phi_p = 1 + \delta\phi_a(a_0) + \delta\phi_b(b_0c)$$

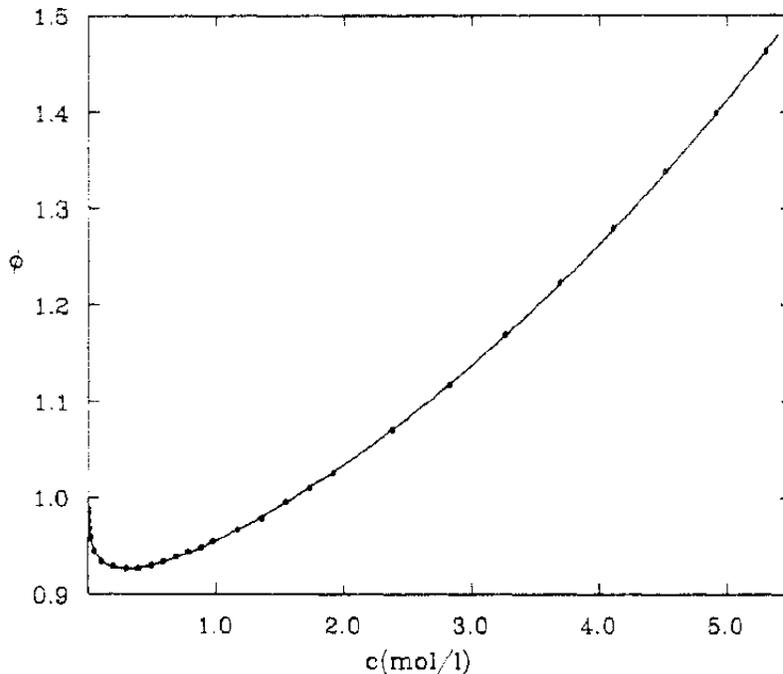
The explicit form for  $\delta\phi_a$  is given by the theory of Debye and Hückel; the hard sphere pressure term  $\delta\phi_b$  is computed with a 2-2 Padé approximate. The adjustable parameters in the primitive function were  $a_0$  (in Å) and  $b_0$  (in L/mole). At concentrations greater than 1 M it was clear that deviations from the measured osmotic coefficients were due to the association of ions into dimers or clusters of higher aggregates. The dominant change an association has upon  $\phi$  is a reduction in the concentration of Brownian particles contributing to the osmotic pressure. The expanded trial function

$$\phi_{\text{trial}} = (1 - F)\phi_p$$

with the new function

$$F = \delta c_B(c)/\nu c$$

$F$  has two parameters. One is an association constant  $K$ ; the other is the number of ions,  $\tau$ . Most of the alkaline metal chlorides had a  $\tau$  value of two. Both HCl and LiCl had  $\tau$  values of 6 and 8, respectively. A typical result was for NaCl with a  $\tau$  of 2,  $a_0$  of 4.82(6)Å,  $b_0$  0.1131(3)L/mol and  $K$  .210(5).



**Figure 1.** Osmotic coefficient data of NaCl in water at 25 °C. Solid curve is the three-parameter fit to the data.

Robert F. Stewart and Clarence Zener, *J. Phys. Chem.* **1988**, *92*,1981-1985.

## 7 Phonon dispersion curves for Si(s) at 293K

Ordinarily the mean square amplitudes of motion for atoms (atomic displacement parameters) are determined from least squares fits of structure factor models to neutron or X-ray diffraction data. An alternate route to an experimental mean square amplitude of vibration may be carried out by a lattice dynamical treatment of simple crystals. In particular, coherent inelastic neutron scattering measurements on a monatomic crystal can be fit with a Born-von Kármán force constant model from which the density of phonon states may be derived.

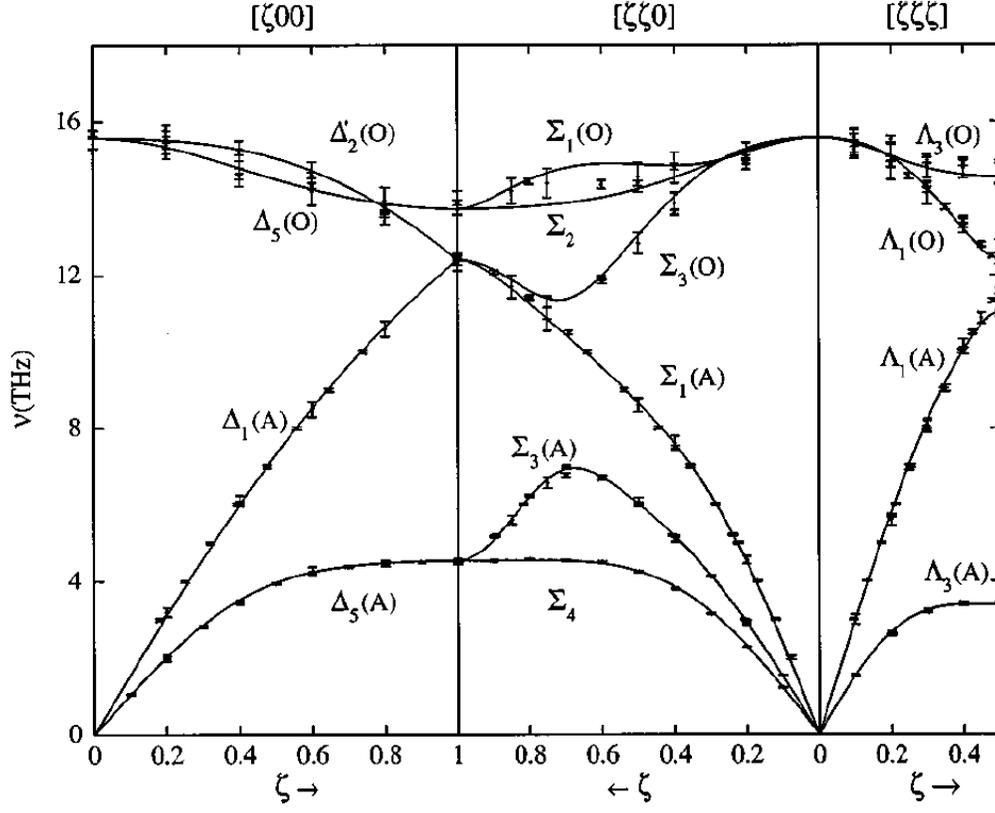
Dispersion data sets from inelastic neutron scattering measurements for the symmetry directions  $\Delta$ [100],  $\Sigma$ [110] and  $\Lambda$ [111] in the Brillouin zones of

Si( $s$ ) were fit by least squares to the BvK force constants up to sixth nearest neighbors. This makes up a total of 21 force constants. A least-squares program was written to fit the BvK force constants to the squares of the measured cyclic frequencies,  $\omega^2$ , for the  $\Delta$ ,  $\Sigma$ , and  $\Lambda$  symmetry directions. The elastic constants and an optically measured Raman frequency were included as observational constraints. (The elastic constants can be related to the force constants by the method of long waves.) The mean square error surface is,

$$\epsilon = \sum_{n=1}^{N_o} w_n [O_n - F(\mathbf{q}; \mathbf{P})]^2$$

$F$  is an explicit algebraic expression which is a function of  $\mathbf{q}$ , a vector in the Brillouin zone and  $\mathbf{P}$  is a vector that consisted of the BvK force constants.  $N_o$  was 193.

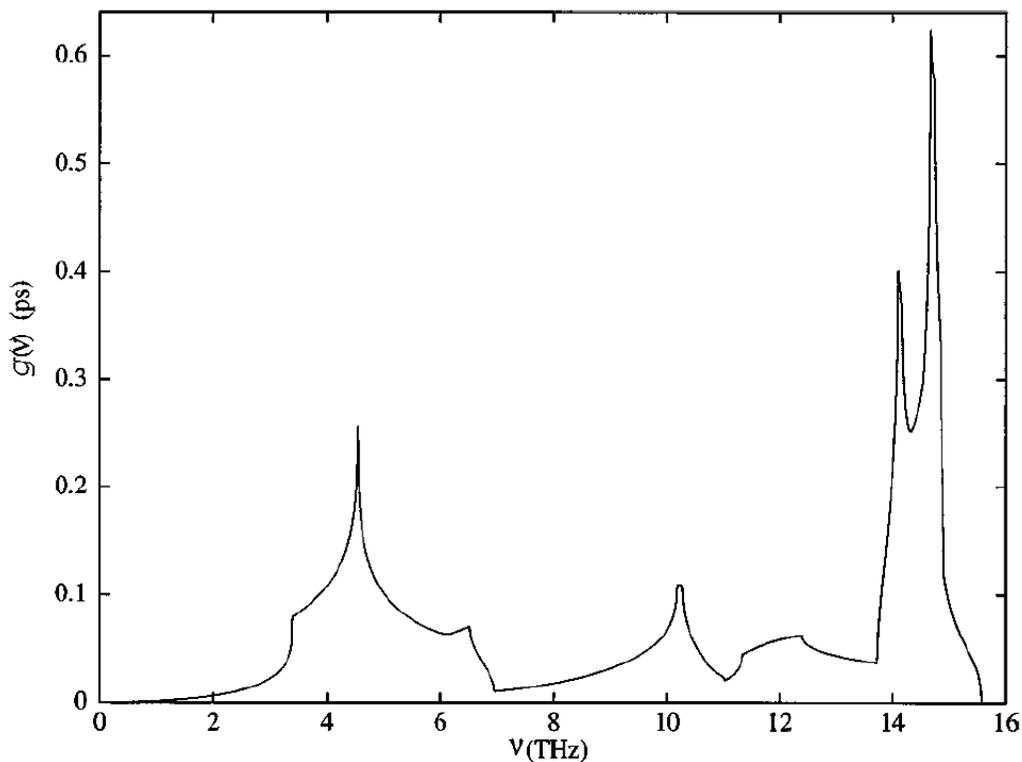
The nonlinear least squares equation was solved for the 21 force constants. The solution seemed to be unique.



A density of phonon states,  $g(\nu)$ , was derived from the force constants. A full propagation of error, based on all of the inverse matrix elements of the Hessian of  $\epsilon$ , allowed us to determine moments and their associated error, of  $g(\nu)$ .

$$B = \frac{2h^2}{mk_B T} \int_0^\infty g(\nu) \frac{1}{x} \left( \frac{1}{2} + \frac{1}{e^x - 1} \right) d\nu$$

where  $x = h\nu/k_B T$ .



The value of  $B$  for  $\text{Si}(s)$  at 293 K was found to be  $.4691 \pm .0016 \text{ \AA}^2$  or equivalently a  $\langle u^2 \rangle$  of  $0.005941(21) \text{ \AA}^2$ . The result was essentially the same as derived from a structure factor fit to Pendellösung X-ray data.

Claus Flensburg and Robert F. Stewart, Phys. Rev. B **60**, 284-291(1999).

## 8 Valence electron structure of F in tetra-fluoroterephthalonitrile

Most least squares programs assume that the basis functions employed have a physical relevance to the experimental data. Sometimes, inclusion of yet another parameter in the Hilbert space of least squares leads to a “divergence” in the iterative process of seeking a solution. Introduction of another coordinate should only lower the mean square error or keep it the same. If  $\epsilon$  increases, then the least squares algorithm is failing.

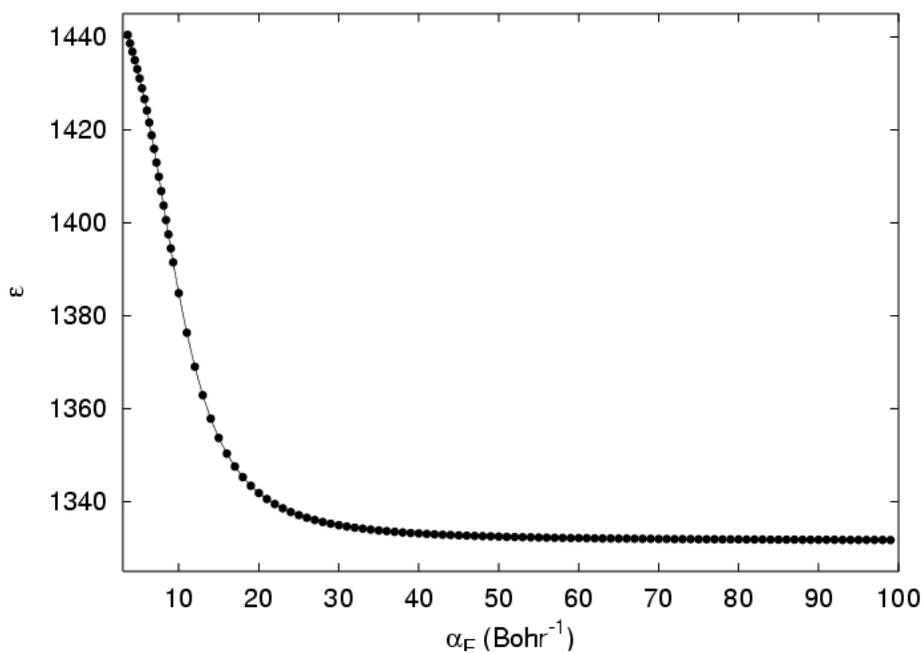
An interesting example of an “unwanted” parameter occurred in an analysis of tetrafluoroterephthalonitrile X-ray diffraction data with a gsf model. In an effort to optimize an octopole radial form factor for the bonded F atom, the minimum in  $\epsilon$  occurred when the radial parameter,  $\alpha_3(\text{F})$  became indefinitely large. At this stage there were 100 variables in the analysis and 3293 observations. The radial form factor was

$$f_3(K) = \frac{K^3}{7!![1 + (K/\alpha)^2]^5}.$$

At large  $\alpha$  this form factor becomes

$$\frac{K^3}{7!!}$$

The  $f_3$  form factor had smoothly transformed itself into a third degree Gram-Charlier scattering function. The inference to be made is that the F atom has virtually no valence structure of third degree symmetry and the model displayed anharmonicity in the motion of the F nucleus.



A plot of  $\epsilon$  versus  $\alpha_3(\text{F})$  (with variation of the other 99 parameters) shows the minimum comes from below and that a Newton-Raphson type algorithm will clearly fail.

H. Sørensen, R. Stewart, G.J. McIntyre and Sine Larsen, *Acta Cryst.* (2003). **A59**,540-550.