

A Procedure for the Automatic Choice of Dynamical Coordinates for Semirigid Molecules

BY MICHAEL MONKENBUSCH

*Institut für Festkörperforschung, Kernforschungsanlage Jülich GmbH, D-5170 Jülich,
Federal Republic of Germany*

(Received 15 May 1986; accepted 26 August 1986)

Abstract

A fully automatic procedure for the generation of constraint-conserving coordinates especially suited for the lattice dynamics of molecular crystals has been developed. Molecules with arbitrary constraints from rigid, semirigid to completely flexible may be treated. The procedure generates a mapping from new dynamical coordinates $\{\zeta_p\}$ to Cartesian atomic displacements. The new coordinates are defined by this mapping. It is accurate up to second order in ζ_p , thus allowing for the inclusion of force-constant contributions due to curved paths of atoms whose displacements are subject to some constraints. The procedure has been implemented in a lattice dynamics program currently under development. Exemplifying calculations on an oxalic acid molecule with several sets of constraints are presented.

Introduction

In the modelling of molecular crystals for refinement or dynamical calculations one often wishes not to deal with all $3N$ Cartesian coordinates of the individual atoms. It is usually more economical and clearer to take account only of a smaller number of degrees of freedom that are relevant for the actual problem. The most extreme and widely used form of such a reduction in the number of degrees of freedom is the assumption of a completely rigid molecule (Pawley, 1972). There are, however, many molecules that cannot be considered as completely rigid. Nevertheless, most of them still contain rigid groups that allow a considerable reduction of the number of relevant coordinates. In general such a semirigid molecule may be specified by a number of geometrical constraints, e.g. fixed bond lengths, fixed bond or torsional angles or confinement of four atoms into one plane. Because of the curvilinear nature of coordinates fulfilling such constraints, a linear relation of the Cartesian atomic displacements with the new coordinates is not sufficient. Neto gives a metric tensor formalism for the calculation of higher-order coefficients for analytically specified curvilinear coordinates (Neto, 1984).

In the following a numerical algorithm is described that allows for a completely automatic choice of an appropriate set of coordinates for molecules with any

geometrical constraints. There is no need to predefine any new coordinates analytically. This algorithm has been programmed in Fortran 77 as part of a general lattice-dynamics program for molecular crystals currently under development in our laboratory.

Statement of the problem

Let $\{x_i^\nu | i = 1, \dots, 3, \nu = 1, \dots, N\}$ denote the Cartesian coordinates of atom ν of the molecule under consideration and $\{u_i^\nu\}$ the deviations of the $\{x_i^\nu\}$ from their equilibrium values by any type of molecular motion. For a system without constraints this implies $3N$ degrees of freedom per molecule. However, for a system with a certain number N_c of independent constraints (semirigid), the total number of degrees of freedom is reduced to $M = 3N - N_c$. Therefore one has to specify M new coordinates, $\{\zeta_p | p = 1, \dots, M\}$, for the description of the dynamics of the semirigid molecule. This means that one has to look for a mapping T :

$$\{u_i^\nu\} = T(\{\zeta_p\}),$$

such that the constraints are not violated but all allowed displacements $\{u_i^\nu\}$ may be generated by some vector $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_M)$. If T is used in lattice dynamics, in general it has to be specified up to second order in ζ :

$$u_i^\nu = \sum_{p=1}^M T_{ip}^\nu \zeta_p + \frac{1}{2} \sum_{p,q=1}^M {}^2 T_{ipq}^\nu \zeta_p \zeta_q + O(\zeta^3).$$

The need for the second-order term may easily be seen by looking at the expression for force constants, C_{pq} , with respect to the new coordinates $\{\zeta_p\}$ in a molecular crystal. Let Φ be the total lattice energy; then

$$\begin{aligned} C_{pq} &= \partial^2 \Phi / \partial \zeta_p \partial \zeta_q \\ &= \sum_{\nu i \mu j} (\partial^2 \Phi / \partial u_i^\nu \partial u_j^\mu) (\partial u_i^\nu / \partial \zeta_p) (\partial u_j^\mu / \partial \zeta_q) \\ &\quad + \sum_{\nu i} (\partial \Phi / \partial u_i^\nu) (\partial^2 u_i^\nu / \partial \zeta_p \partial \zeta_q) \\ &= \sum_{\nu i \mu j} (\partial^2 \Phi / \partial u_i^\nu \partial u_j^\mu) {}^1 T_{ip}^\nu {}^1 T_{jq}^\mu \\ &\quad + \sum_{\nu i} (\partial \Phi / \partial u_i^\nu) {}^2 T_{ipq}^\nu. \end{aligned}$$

A very simple example may further clarify the above statement: imagine a rigid dipolar dumbbell in a homogeneous electric field E turning around its centre C as depicted in Fig. 1. Here the new dynamical coordinate ξ is proportional to the librational angle around the centre C . The linear part of a coordinate generating the mapping T is only able to approximate the displacement paths of the + and - 'atoms' by straight lines parallel to the equipotential lines of the assumed electric field. Therefore no restoring torque aligning the dipole along the electric-field vector E can be derived from this description. Only the second-order part of T yields parabolic paths that approximate the exact circular motion of the charged 'atoms' to a high degree of accuracy as long as ξ does not become too large. If a realization of T including the second-order terms is used, the right restoring force may easily be computed.

The goal of our procedure will be the automatic specification of the coefficients $\{T_{ip}^\nu\}$ and $\{T_{ipq}^\nu\}$ for arbitrary molecules with arbitrary constraints.

Computational method

First-order coefficients

For a given molecule, $\{x_i^\nu\}$, the constraints are specified in terms of nonzero values for the corresponding intramolecular valence force constants k . A pseudo-distortion energy, Φ_d , is defined as

$$\begin{aligned} \Phi_d = & \frac{1}{2} \sum_{\nu\nu'} k_{\nu\nu'}^{\text{stretch}} (d_{\nu\nu'} - d_{\nu\nu'}^0)^2 \\ & + \frac{1}{2} \sum_{\nu\nu'\nu''} k_{\nu\nu'\nu''}^{\text{bend}} (\varphi_{\nu\nu'\nu''} - \varphi_{\nu\nu'\nu''}^0)^2 \\ & + \frac{1}{2} \sum_{\nu\nu'\mu\mu'} k_{\nu\nu'\mu\mu'}^{\text{torsion}} (\psi_{\nu\nu'\mu\mu'} - \psi_{\nu\nu'\mu\mu'}^0)^2 \\ & + \frac{1}{2} \sum_{\nu\nu'\nu''\mu} k_{\nu\nu'\nu''\mu}^{\text{out of plane}} (\gamma_{\nu\nu'\nu''\mu} - \gamma_{\nu\nu'\nu''\mu}^0)^2 \end{aligned}$$

with $d_{\nu\nu'}$ = distance between atoms ν and ν' , $\varphi_{\nu\nu'\nu''}$ = bond angle of the bond $\nu - \nu' - \nu''$, $\psi_{\nu\nu'\mu\mu'}$ = torsion angle around the bond $\nu' - \mu$, and $\gamma_{\nu\nu'\nu''\mu}$ = angle of bond $\nu - \mu$ with the plane containing ν , ν' and ν'' .

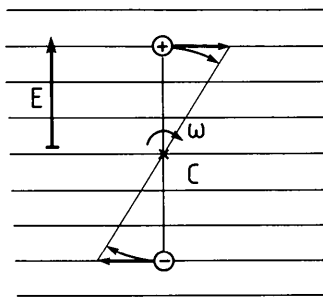


Fig. 1. Dipolar dumbbell in a homogeneous electric field to illustrate the need to specify T up to second order in ξ .

d^0 , φ^0 , ψ^0 and γ^0 are the corresponding equilibrium values.

For every bond distance d , bond angle φ , torsion angle ψ and out-of-plane displacement γ that should be considered as rigid, the corresponding coefficient k is set to an arbitrary nonzero value, all others being set to zero. Once Φ_d is given, the intramolecular pseudo-force-constant matrix $[C_{ij}^{\nu\mu}]$ can be computed:

$$C_{ij} = \partial^2 \Phi_d / \partial u_i^\nu \partial u_j^\mu.$$

Because of the required rigidity of the geometrical bond properties contributing to $[C_{ij}^{\nu\mu}]$, any allowed distortion $\{^*u_i^\nu\}$ of the molecule must satisfy the following equation:

$$0 = \sum_{\nu i \mu j} C_{ij}^{\nu\mu} {}^*u_i^\nu {}^*u_j^\mu.$$

A set of M linearly independent orthogonal distortion patterns $\{^*u_i^\nu\}$ may easily be found by diagonalizing $[C_{ij}^{\nu\mu}]$. The diagonalization yields $3N$ eigenvalues, $\{\lambda_s | s = 1, \dots, 3N\}$, M of which are zero within the numerical accuracy. The corresponding eigenvectors, $\{t_{ip}^\nu | \lambda_p = 0, p = 1, \dots, M\}$, represent one specific set of M linearly independent distortion patterns that fulfil the condition

$$\sum_{\mu j} C_{ij}^{\nu\mu} t_{jp}^\mu = 0.$$

Therefore

$$\sum_{\mu j \nu i} C_{ij}^{\nu\mu} t_{jp}^\mu t_{ip}^\nu = 0.$$

The t_{jp}^ν may be identified with the coefficients T_{jp}^ν we are looking for. Of course, this choice is not unique; any other set of linearly independent vectors $\{\tau_p = [\tau_j^\nu]_p\}$ that span the same M -dimensional linear subspace U of \mathbb{R}^{3N} that is spanned by the $\{T_p\}$ may be chosen instead. For example, in lattice-dynamical problems it is convenient to have a diagonal mass tensor, $\tilde{\mu} = [\mu_{pp'} \delta_{pp'}]$, yielding the simple expression for the kinetic energy:

$$E_{\text{kin}} = \frac{1}{2} \sum_p \mu_p \dot{\xi}_p^2.$$

This can be achieved by a simple rotation of the original basis $\{T_p\}$ to a new basis $\{T'_p\}$ within the subspace U and taking the T'_{ip}^ν as linear coefficients of the modified mapping T' . Another possibility is to align the rotated basis $\{T'_p\}$ such that the first three coordinates ($p = 1, \dots, 3$) correspond to the centre-of-mass motions and the next three coordinates correspond to the rotation of the whole molecule.

As a variation of the procedure outlined above one may also employ a realistic valence force field and enter atomic masses m_ν , then diagonalize $[C_{ij}^{\nu\mu} / (m_\nu m_\mu)^{1/2}]$ and take all eigenvectors $\{t_{ip}^\nu | \lambda_p < \lambda_{\text{lim}}\}$

that correspond to vibration frequencies below a limiting frequency ν_{lim} as basis for the subspace U . However, in that case no second-order coefficients of T may be computed by the procedure described in the next section.

Second-order coefficients

Now that we have a procedure to compute the first-order coefficients of T for a molecule in a specific orientation and conformation $\{x_i^r\}$ it is possible to devise a method to obtain the second-order coefficients, ${}^2T_{ipq}^\nu$. Whereas in general an analytical type of solution seems exceedingly complicated, the following numerical procedure is quite feasible. Let us denote by ${}^1T_{ip}^{\nu 0}$ the first-order coefficients derived using the original molecular coordinate set $\{x_i^r\}$. To compute ${}^2T_{ipq}^\nu$ now generate first a slightly distorted coordinate set, $\{x_i^r + {}^1T_{iq}^{\nu 0} \Delta\zeta_q\}$, by displacing the system along the new coordinate ζ_q employing the already known linear part of the mapping. With the new atomic positions thus generated ${}^1T_{ip}^\nu$ may be calculated again; the result is denoted ${}^1T_{ip}^{\nu q}(\Delta\zeta_q)$. Owing to the arbitrariness of the selection of the basis $\{{}^1T_p^q(\Delta\zeta_q)|p=1, \dots, M\}$ of the subspace $U^q(\Delta\zeta_q)$, before comparison of this result with the $\{T_p^0\}$ vectors an adapted set, $\{{}^1T_p^{\nu q}(\Delta\zeta_q)\}$, has to be constructed by projecting the $\{T_p^0\}$ onto the subspace $U^q(\Delta\zeta_q)$. The corresponding second-order coefficients are then obtained from

$${}^2T_{ipq}^\nu = \lim_{\Delta\zeta \rightarrow 0} [{}^1T_{ip}^{\nu q}(\Delta\zeta_q) - {}^1T_{ip}^{\nu 0}] / \Delta\zeta_q.$$

This procedure has to be repeated for each q with $1 \leq q \leq M$ to get all second-order coefficients.

Implementation

The procedure outlined above has been programmed in IBM VS-Fortran (Fortran 77) employing double-precision arithmetic as part of a larger general lattice-dynamics program for molecular crystals. Constraints may be specified redundantly or non-redundantly by specifying the numbers ν, ν' (ν'', μ, \dots) of the atoms concerned together with the corresponding interaction constants k . The pseudo-distortion energy is calculated according to the formula given above (also including stretch-stretch, bend-stretch and bend-bend interactions). The geometric parameters d, φ, ψ, γ are computed from the distance vectors connecting the atoms ν, ν' (ν'', μ, \dots). The intramolecular force-constant matrix $C_{ij}^{\nu\mu}$ is computed by numerical differentiation (incremental atomic displacements are $\Delta x = 5 \times 10^{-7}$). The diagonalization is performed by the IMSL routine *EIGRS* (International Mathematical and Statistical Library, 1979). All eigenvectors corresponding to eigenvalues $\lambda_p > 2 \times 10^{-3}$ [or if realistic vibration frequencies are computed $\lambda_p >$

$(2\pi\nu_{\text{lim}})^2$] are selected to span the subspace U . The second-order coefficients are derived as described in the previous section using a value of $\Delta\zeta = 0.05$. Before storage they are symmetrized with respect to p and q .

The transformation properties for transferring the T coefficients from one molecule to those of another symmetry-related molecule in the unit cell are given in the Appendix.

Example

As a still comparatively simple but instructive example some computational results on the oxalic acid molecule (Derissen & Smit, 1974) are presented here. As far as realistic valence force constants have been employed these were taken from De Villepin, Novak & Bougeard (1982); see Fig. 2. Three different constraint models are illustrated by the example calculations.

The first model is generated by setting all valence force constants less than or equal to $0.15 \text{ aJ } \text{\AA}^{-2}$ equal to zero. The remaining force constants are used to compute Φ_d and thus give rise to the constraints. The

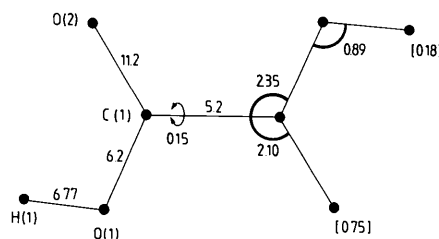


Fig. 2. Example: oxalic acid. The numbers indicate the valence force constants for bond stretching, bond-angle bending, torsion and out-of-plane motion (numbers in square brackets), respectively (De Villepin *et al.*, 1982), units are $\text{aJ } \text{\AA}^{-2}$ ($= 1 \text{ mdyne } \text{\AA}^{-1}$).

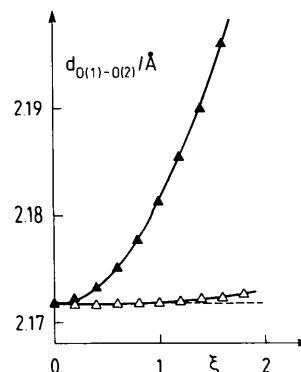


Fig. 3. Distance $O(1) \cdots O(2)$ for an oxalic acid molecule with the second model set of constraints as a function of the distortion parameter ξ (see text). The solid triangles correspond to a pure first-order T , the open triangles correspond to a mapping including second-order coefficients, the broken line indicates the constant value.

largest eigenvalue λ_p that has been considered as zero within the numerical accuracy was $\lambda_7 = 0.13 \times 10^{-8}$, the lowest λ_p considered as nonzero was $\lambda_8 = 0.23$. That is to say, the torsion around the C(1)–C(1') bond is free; all other elements are still rigid. This yields a total of $M = 7$ degrees of freedom to be compared with the $3N = 24$ degrees of freedom of the completely flexible molecule. Six of the seven remaining degrees of freedom stem from the translational and rotational motion of the whole molecule.

The second model allows for additional out-of-plane motions of the two hydrogen atoms by setting equal to zero all force constants less than or equal to $0.18 \text{ aJ } \text{Å}^{-2}$. The two additional degrees of freedom increase M to $M = 9$.

Finally, the third model includes the out-of-plane motions of the O(2), O(2') atoms and bending of the C(1)–O(1)–H(1), C(1')–O(1')–H(1') angles by setting all force constants less than or equal to $0.83 \text{ aJ } \text{Å}^{-2}$ to zero ($M = 13$).

In Figs. 3 and 4 some tests of the derived coordinate-generating mappings T for the three models are presented (first: circles, second: triangles, third: squares). The atomic displacements $\{u_i^v\}$ are computed for a distortion generated by the M -component constrained coordinate vector $(0.1, 0.1, \dots, 0.1)\xi$. With model 2 the distance O(1)⋯O(2) should ideally be constant irrespective of any allowed distortion. Fig. 3 demonstrates that this is valid only to first order

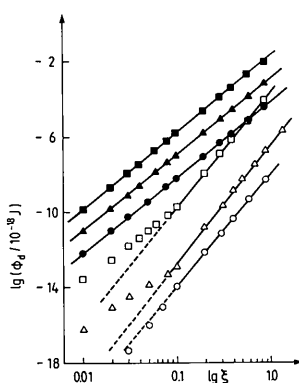


Fig. 4. Energy Φ_d owing to constraint-violating distortions as a function of the distortion parameter ξ . Solid symbols: first-order mapping only, open symbols: second-order mapping. Different symbol shapes correspond to the three different models explained in the text.

in ξ for a pure first-order T , whereas the mapping including the second-order terms meets this constraint much better. In Fig. 4 the logarithm of the distortion energy Φ_d owing to constraint-violating distortions is plotted *vs* $\log(\xi)$. As expected for the pure first-order mapping the slope of the straight lines connecting the computed values for all three models implies $\Phi_d \sim \xi^4$. The slope of the lines through the points from the calculation with the second-order mapping show a slope indicating $\Phi_d \sim \xi^6$. The deviations of the computed values at small ξ are due to the increasing influence of numerical truncation errors during the evaluation of Φ_d . These results do not depend on the specific choice of the distortion vector but may also be obtained with any other distortion of the form $(a_1, a_2, \dots, a_M)\xi$.

APPENDIX

Transformation properties of the T coefficients

For cases where several identical molecules are in one unit cell the first- and second-order coefficients of the mapping T may be transferred from one molecule A to another molecule B by employing the appropriate proper or improper rotation matrix, \tilde{R} :

$$x_i^{\nu B} = \sum_j R_{ij} x_j^{\nu A} + s_i,$$

where s denotes a shift of the whole molecule. Then

$${}^1 T_{ip}^{\nu B} = \sum_j R_{ij} {}^1 T_{jp}^{\nu A}$$

$${}^2 T_{ipq}^{\nu B} = \sum_j R_{ij} {}^2 T_{jpq}^{\nu A}.$$

For any orthogonal \tilde{R} the diagonal mass tensor $\tilde{\mu}$ is not influenced, *i.e.* $\mu_p^A = \mu_p^B$.

References

- DERISSEN, J. L. & SMIT, P. H. (1974). *Acta Cryst.* B30, 2240–2242.
 DE VILLEPIN, J., NOVAK, N. & BOUGEARD, D. (1982). *Chem. Phys.* 73, 291–312.
 International Mathematical and Statistical Library (1979). Edition 7, IMSL LIB-0007, revised Jan 1979. IMSL, 7500 Bellaire Boulevard, Houston, Texas.
 NETO, N. (1984). *Chem. Phys.* 91, 89–100.
 PAWLEY, G. S. (1972). *Phys. Status Solidi B*, 49, 475–488.