

Chapter 4

XDLSM - Least Squares Program for Multipole Refinement

4.1 Overview

XDLSM is a full-matrix least squares program based on the generalized scattering model detailed in the Introduction. Its present version includes multipole expansion up to $l=4$ and anharmonic treatment of the thermal motion up to 4th order of the Gram-Charlier expansion. XDLSM, being based on the Hansen-Coppens formalism [1], necessarily has many common elements with MOLLY, the algorithm of which has been rebuilt and extended to allow for further developments. XDLSM supports sophisticated density modelling, and features of previous refinement programs have been incorporated (LSEXP [2]). Further important aspects of XDLSM provide methods to locate inadequacies in the model, to control the refinement and to monitor the results.

4.1.1 The method of least squares

In this chapter some aspects of the method of least squares are discussed, whose knowledge are necessary for the user to handle the input and output of XDLSM. This introduction is based on reference [3], to which the reader is referred for more details.

Consider a given set of m observations $y_o\{y_{o1}, y_{o2}, y_{o3}, \dots, y_{om}\}$ represented by the corresponding set of model functions $y_c\{y_{c1}, y_{c2}, y_{c3}, \dots, y_{cm}\} = y_c(\mathbf{x})$, where \mathbf{x} is the n -component vector of the parameters $x\{x_1, x_2, x_3, \dots, x_n\}$. The best unbiased estimates of x can be obtained by minimizing the square of the residual:

$$R^2 = (\mathbf{y}_o - \mathbf{y}_c)' \mathbf{W} (\mathbf{y}_o - \mathbf{y}_c) = (\mathbf{y}_o - \mathbf{y}_c)' \mathbf{Q}' \mathbf{Q} (\mathbf{y}_o - \mathbf{y}_c) \quad (\text{Eq. 4.1})$$

where \mathbf{W} , the weight matrix, is to be chosen as the inverse of the variance-covariance matrix of the observations (in practice, it is taken to be diagonal), and \mathbf{Q} is an upper triangular matrix, i.e. $\mathbf{Q}'\mathbf{Q}$ is the Cholesky decomposition of \mathbf{W} . If \mathbf{y}_c can be expanded about \mathbf{x}_o in Taylor's series retaining only the first order terms, then

$$\mathbf{y}_c = \mathbf{y}_c(\mathbf{x}_o) + \mathbf{D}(\mathbf{x} - \mathbf{x}_o) \quad (\text{Eq. 4.2})$$

with $D_{ij} = dy_{ci} / dx_j$ being the design matrix. Eq. (4.1) becomes:

$$R^2 = [\Delta \mathbf{y} - \mathbf{Z} \Delta \mathbf{x}]' [\Delta \mathbf{y} - \mathbf{Z} \Delta \mathbf{x}] \quad (\text{Eq. 4.3})$$

where $\Delta \mathbf{y} = \mathbf{Q}(\mathbf{y}_o - \mathbf{y}_c(\mathbf{x}_o))$, $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_o$ and $\mathbf{Z} = \mathbf{Q}\mathbf{D}$.

The n conditions

$$\left(\frac{dR^2}{dx_i} \right)_{x_i = x_{oi}} = 0 \quad \text{for } i = 1, 2, 3, \dots, n \quad (\text{Eq. 4.4})$$

lead to the system of normal equations

$$\mathbf{Z}'\mathbf{Z}\Delta \mathbf{x} = \mathbf{Z}'\Delta \mathbf{y} \quad (\text{Eq. 4.5})$$

whose solution vector is

$$\mathbf{x} = \mathbf{x}_o + \mathbf{B}^{-1}\mathbf{Z}'\Delta\mathbf{y} \quad (\text{Eq. 4.6})$$

with $\mathbf{B}=\mathbf{Z}'\mathbf{Z}$.

An alternative solution of the least-squares problem is provided through the singular value decomposition of the standardized design matrix \mathbf{Z} . Let

$$\mathbf{Z} = \mathbf{U}\mathbf{G}\mathbf{V}' \quad (\text{Eq. 4.7})$$

where \mathbf{U} is an $m \times n$ column orthogonal matrix, \mathbf{G} is a diagonal matrix of the singular values and \mathbf{V} is an $n \times n$ orthogonal matrix.

A solution of the over-determined system of equations

$$\mathbf{Z}\Delta\mathbf{x} = \Delta\mathbf{y} \quad (\text{Eq. 4.8})$$

can be given as

$$\Delta\mathbf{x} = \mathbf{Z}^{-1}\Delta\mathbf{y} \quad (\text{Eq. 4.9})$$

where

$$\mathbf{Z}^{-1} = \mathbf{V}\mathbf{G}^{-1}\mathbf{U}' \quad (\text{Eq. 4.10})$$

This solution can be proved to be the best possible solution in the least-squares sense as $\Delta\mathbf{x}$ is the vector which minimizes the residual:

$$R = |\mathbf{Z}\Delta\mathbf{x} - \Delta\mathbf{y}| \quad (\text{Eq. 4.11})$$

The solution of the least-squares problem through the system of normal equations (4.6) has the disadvantage that it fails if \mathbf{B} is singular or ill-conditioned. A clear distinction should be made between ill-conditioning of an *analytical* and that of a *numerical* nature. The former case is likely to occur for an *over-parametrized* model, when some combination of basis functions are irrelevant to the fit. The normal equation matrix has zero or nearly zero eigenvalues and the inversion gives no or only a formal solution. This problem manifests itself in undesirable correlations among the variables. The method used for establishing hidden indeterminacies in the model is the singular value decomposition of the matrix of observation-equations (4.8). This procedure gives a diagnosis of the degeneracies and provides a solution minimizing the residual. The matrix can be considered ill-conditioned if its inverse condition number, the ratio of the smallest to the largest eigenvalue, is comparable with the machine precision. The components of the eigenvector (a row or column vector of \mathbf{V}) corresponding to the smallest eigenvalue define a linear dependence among the variables (orthonormal basis for the null-space) which leads to the singularity. Zeroing an eigenvalue in the calculation of the inverse matrix (4.10) means introducing the constraint given by the corresponding eigenvector. The term *numerical ill-conditioning* refers here to an unbalanced least-squares matrix which is due to the fact that the model function is simply not equally sensitive to the changes of the different parameters, *i.e.* the components of the design matrix can differ by many orders of magnitude. A condition number of the order of 10 is typical for the multipole-model based structure factor least-squares matrix. This number indicates that a small change (error) in an element of the design matrix (\mathbf{Z}) can cause large changes in the elements of \mathbf{B}^{-1} . That is why the solution *via* the inversion of the normal equation matrix is susceptible, to a considerable extent, to roundoff errors and requires double precision arithmetic. This problem can be overcome if the elements of the \mathbf{Z} (or \mathbf{B}) matrix are brought to a common scale. In XDLSM the normal equation matrix is analyzed and its conditioning is accomplished by a similarity transformation:

$$\mathbf{B}_c = \text{diag}(\mathbf{B})^{-1/2} \mathbf{B} \text{diag}(\mathbf{B})^{-1/2} \quad (\text{Eq. 4.12})$$

Inversion, based on the Gauss-Jordan elimination method [4], is the default option to solve eq. (4.5). If the matrix inversion fails or if diagonalization has been selected as the method of

solution, the eigenvalues are calculated by the Householder reduction [4], and the singularities are reported and eliminated. The eigenvalue filtering is based on the inverse condition number. The lowest eigenvalues are rejected (zeroed) from the inverse calculation until the inverse condition number reaches a user specified limit. While this procedure gives a mathematically correct solution, its indiscriminate application does not necessarily reveal the physical meanings of the indeterminacies that made the least-squares equations singular or nearly singular in the first place.

4.1.2 Model ambiguities

The XD package will be available for a wide scientific community. This Section tries to help those who have not yet been involved in charge density research. In particular, it tries to help those users who have not yet had the uncomfortable feeling of getting stuck at a certain stage of the refinement. This happens when decisions need to be made as to which parametrization is preferable among several alternative ones which perform equally well in fitting the data.

The scattering model described in the Introduction formally allows 66 parameters per atom (in the present implementation of XD) to be included in the refinement. However, any interpretation of the data set using an "all-parameter" fit is hardly feasible, nor is it appropriate. Even if one could afford it (*i.e.* even if enough data points were available) and even if convergence was reached with a satisfactory fit, the physical significance of the results would certainly be doubtful. While the total dynamic ED obtained could account for the data very well, any property which is a function of a subset of the variables could well be meaningless. As mentioned above, the reason for this is that many basis functions of the structure factor expansion have a similar dependence on the components of the scattering vector. Consequently the data cannot differentiate between them. A typical example of this type of bias is that introduced into the static density deformations by the inadequate decomposition of the thermal smearing. This is caused by the formal similarity between density basis functions and pdf's of the nuclear displacements. Strong correlations, as high as 80-90%, are likely to occur between quadrupole populations and second order displacement parameters. The Gram-Charlier model has been shown to be as adequate as the multipole expansion in accounting for static density asphericities [5]. Such indeterminacies can appear especially pronounced for non-centrosymmetric structures.

The flexibility of the model and the limited number of observations forces one to limit the optimization to a subset of parameters or to their combinations. The variables are usually selected on the basis of simple chemical arguments or preconceptions. The outcomes must be tested in order to judge their physical significances. A careful study should not neglect an independent analysis of static and dynamic parameters.

4.1.2.1 Testing the results

The most important test to judge the success of the model and the quality of the fit is to evaluate the residual ED through a Fourier summation ($F_{\text{obs}} - F_{\text{model}}$). This provides a direct-space representation of the extent to which the model accounts for the observations. A featureless residual map is a necessary condition for the adequacy of a model, but is far from being a sufficient one for judging its physical significance. Another usual procedure is to compare the static deformation density obtained from X-ray data with that calculated theoretically. Deformation peak-shapes and peak-heights are subject to specific conditions that are characteristic for the different methods to be compared. The ab-initio ED depends on the level of the theory applied and on the quality of the basis sets. Both factors place severe limitations on any direct comparisons, especially for larger systems. However, without such comparisons, the interpretation of the results in terms of the deformation ED remains only of a qualitative nature. This is because of the arbitrariness in selecting the reference state and the sensitivity of the ED to the structural parameters.

We suggest that the experimental ED is tested through its local and global topological characteristics and by evaluating its integrated properties. XDPROP makes it feasible to trace the refinement process almost "continuously" by inspecting the different stationary points of the total ED and related scalar properties. In this respect the Laplacian of the ED, as a sensitive measure of charge concentrations, should play an important role. A static ED which fails to reproduce the characteristic topological features of a typical covalent bond, e.g. (3,-1) CP's, bond charge concentrations shown by the Laplacian, is likely to be suspect.

One-electron properties are directly obtainable from the ED and their comparison with the outcomes of independent measurements and/or theoretical results are of great importance. The molecular dipole moment and the electrostatic potential are the quantities most frequently evaluated from the experimental ED. Such applications are being explored with a promising success.

One way to gain information on the physical significance of the thermal parameters is to test them against the rigid-body motion model [6] which is based on the observation that in molecular crystals the external (lattice) vibrations make the major contribution to the atomic motion. Satisfactory agreement between observed and calculated anisotropic displacement parameters may suggest that the molecule is rigid to a good approximation or the thermal parameters are uniformly affected by systematic errors. Significant residuals after the rigid-body fit may indicate either the importance of soft internal modes or simply a bias in the atomic displacements. A directly applicable test for the correctness of the atomic displacement parameters is the rigid-bond test [7].

If $z_{A,B}^2$ denotes the mean square displacement amplitude of atom A in the direction of atom B, then for every covalently bonded pair of atoms A and B

$$\Delta_{A,B} = z_{A,B}^2 - z_{B,A}^2 = 0$$

Conversely, if in parts of the molecule this *rigid bond postulate* is not fulfilled, one may deduce that the structural model is insufficient. Hirshfeld estimated that for atoms at least as heavy as carbon $\Delta_{A,B}$ should normally be smaller than 0.001 \AA^2 . Verification of the model and the anisotropic displacement parameters by this test strengthens confidence in the experimentally determined ED.

A very useful visualization of the atomic displacement parameters is provided by the computer-graphics program PEANUT [8], developed recently to analyze observed (fitted to diffraction data), calculated (as given by a model) or residual (observed- calculated) thermal parameters in terms of closed surfaces defined by the root-mean-squares displacements ($\langle u(\mathbf{n}) \rangle^{1/2} = (\mathbf{n}' U \mathbf{n})$, where \mathbf{n} is a unit vector in any direction). Applications are given in reference [9].

A plausible approach to reduce ambiguities in the model is to introduce constraints into the refinement. It is desirable to replace external checks on one of the possible, mathematically equivalent solutions by internal constraints applicable to support the physically most relevant solution. An advanced feature of XDLSM is to allow for general linear restrictions on any set of variables. Efforts are being made to further develop this option in order to incorporate more 'physics' into the refinement model.

4.1.2.2 Constraints in XDLSM

The treatment of constraints in XDLSM is based on the technique of *direct elimination*. Consider a system of nc linear equations, each of which defines a constraint among nv variables:

$$\mathbf{C}(nc, nv) \Delta \mathbf{x}(nv) = \mathbf{a}(nc) \quad (\text{Eq. 4.13})$$

By decomposing the matrix \mathbf{C}

$$\mathbf{C} = \mathbf{P}\mathbf{S}\mathbf{R}' = \mathbf{P}_{(nc,nv)} \begin{pmatrix} \mathbf{S}(nr,nr) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}'_1(nr,nv) \\ \mathbf{R}'_2(nv-nr,nv) \end{pmatrix} \quad (\text{Eq. 4.14})$$

with \mathbf{S} being a diagonal matrix of nr non-zero singular values ($nr \leq nc$), two sets of new variables can be introduced:

$$\Delta \bar{\mathbf{x}}_1(nr) = \mathbf{R}'_1 \Delta \mathbf{x} \quad \Delta \bar{\mathbf{x}}_2(nv-nr) = \mathbf{R}'_2 \Delta \mathbf{x} \quad (\text{Eq. 4.15})$$

where the first set can be eliminated by means of eq. (4.13) and (4.14):

$$\Delta \bar{\mathbf{x}}_1 = \mathbf{S}^{-1} \mathbf{P}' \mathbf{a} \quad (\text{Eq. 4.16})$$

This leads to a decomposition of the unconstrained variables

$$\Delta \mathbf{x} = \mathbf{R} \Delta \bar{\mathbf{x}} = \mathbf{R}_1 \Delta \bar{\mathbf{x}}_1 + \mathbf{R}_2 \Delta \bar{\mathbf{x}}_2 = \mathbf{R}_1 \mathbf{S}^{-1} \mathbf{P}' \mathbf{a} + \mathbf{R}_2 \Delta \bar{\mathbf{x}}_2 \quad (\text{Eq. 4.17})$$

The equations of observations 4.2 becomes

$$\Delta \mathbf{y} - \mathbf{D} \mathbf{R}_1 \mathbf{S}^{-1} \mathbf{P}' \mathbf{a} = \mathbf{D} \mathbf{R}_2 \Delta \bar{\mathbf{x}}_2 \quad (\text{Eq. 4.18})$$

and the system of normal equations is reduced to

$$\mathbf{B}_2 \Delta \bar{\mathbf{x}}_2 = \mathbf{R}'_2 \mathbf{Z}' \Delta \mathbf{y}_2 \quad (\text{Eq. 4.19})$$

where

$$\mathbf{B}_2 = \mathbf{R}'_2 \mathbf{Z}' \mathbf{Z} \mathbf{R}_2 \quad \text{and} \quad \Delta \mathbf{y}_2 = \Delta \mathbf{y} - \mathbf{D} \mathbf{R}_1 \mathbf{S}^{-1} \mathbf{P}' \mathbf{a} \quad (\text{Eq. 4.20})$$

The elimination through the singular-value decomposition of the constraints matrix has two advantages;

1. the dimension of problem is reduced by the number of independent constraints
2. the restrictions can be formulated in an automatic way as all accidental redundancies are easily filtered out.

Some of the constraints mentioned below have already been implemented in a user-friendly way, others will be available in subsequent releases of XDLSM.

4.1.2.3 Restrictions on the multipole populations

Monopole population constraint. The sum of the monopole populations, by definition, gives the number of valence electrons in the molecule (unit cell). This statement is part of the multipole expansion formalism which involves "atomic" partitioning and thus provides a particular assignment of the atomic charge to the corresponding monopole population. The monopole population constraint (**KEEP charge group1**, previously referred to as the electroneutrality constraint, see section 4.6.5 below) exactly maintains the input charge on the ASU. If this charge is zero on input, it keeps the unit cell neutral. In XDLSM it is possible to define any subset of atoms (*i.e.* any functional group) for which the total number of valence electrons is kept constant. This option then precludes any charge transfer between the group(s) selected and the rest of the atoms in the unit cell.

Local pseudo symmetry, "chemical" symmetry. Preconceptions based on chemical intuition can also be applied to reduce the number of multipole populations to be refined. One can assume a simple hybridization scheme which corresponds to the actual geometrical arrangement of the atoms. This is usually achieved by imposing site symmetry in a properly chosen local Cartesian frame and using symmetry adapted angular functions. The symmetry restrictions for real spherical harmonics are given in Table 4-5. Another feasible restraint is to keep the valence density of chemically equivalent or similar atoms to be the same during the refinement. This is a widely accepted practical approach in studies on larger molecules. The real question is how to judge the actual applicability and success of our chemical expectations implemented in such a way. Static equivalences might be

hampered in an unconstrained refinement by dynamic non-equivalences of the atoms considered. Another important aspect is that in crystals, the 'chemical symmetries' characteristic of the isolated molecules may not be preserved. Any subsequent enforcement of static equivalencies may result in the effects of the crystal field becoming unobservable.

4.1.2.4 Restrictions on the radial functions

The shape of R_l 's are controlled by $n(l)$ and a_l (see eq. 1.19), the latter being estimated from the Hartree-Fock-optimized single- ξ values. In case of quadrupolar atoms (which have only *ss*, *sp*, and *pp* type orbital products) the selection of a_l for $l > 2$ is not straightforward. The corresponding "virtual" density basis functions are shown to account for bond densities [10]. The usual practice is to keep $a_l = a$ for all l and optimize κ' scaling of a . Even under this severe restriction κ' becomes highly correlated with the populations and convergence can be troublesome. In this respect, κ' is by far the most critical parameter of the formalism. This may indicate that the constraint implemented is not adequate. Model studies on di-atomic molecules showed that a satisfactory fit of the HF ED with one-center multipole densities requires, in certain cases, highly structured radial functions while in other cases, depending on the level of expansion, simple Slater functions are sufficient [11]. The extent to which this statement applies to many-atom molecules remains to be examined. A trivial choice for improving the situation is the use of radial functions corresponding to extended basis HF atomic orbitals. In studies on transition metal complexes, the HF radial scattering factors were shown to be superior to those of single Slater functions [12].

4.1.2.5 Restrictions on the vibrational parameters

Rigid-body or segmented-rigid-body models could be incorporated into the structure factor refinement. Both approaches require a linear transformation of the design matrix leading to a reduction in the number of dynamic variables. Severe indeterminacies, depending on the formalism, can be introduced.

A more elegant alternative procedure applied in XDLSM is to define rigid molecules or segments by invoking rigid-bond and rigid-link constraints. This is a very efficient way to define the degree of flexibility, but a full control requires a detailed knowledge of the intramolecular motion. Normal coordinate analysis, if a suitable force-field is available, provides the MSDA matrix associated with any normal mode. For molecules of first row elements, standard force fields are readily available and procedures are in general use to refine them against spectroscopic data. Frequencies at the HF level are typically 10% larger than those of measured, and even semi-empirical methods can provide fair estimations. Incorporation of calculated ADP's for hydrogen atoms into charge density refinements has been recently reported by a number of authors, using differing methodologies [30-33]. See Chapter 11.1 concerning the program XDVBIB for inclusion of calculated ADP's.

An easy to handle approach is to apply constraints of the rigid-bond (rigid-link) type to the shift of the ADP's calculated from an intramolecular force field. Such shifts give only rigid-body type contributions to the ADP's and the procedure preserves atomic displacements due to intramolecular vibrations. The success of such applications depends on the extent to which the mean-field approximation is valid. Another difficulty is that the optimized molecular geometry needed to calculate the harmonic force field can considerably differ from that found in the crystal. Another approach is to start from a set of ADP's predicted by the TLS model. These ADP's satisfy the Hirshfeld condition for all internuclear separations. By invoking the rigid-body constraint to all covalent bonds between atoms of comparable mass the bias in the ADP's can be reduced significantly.

4.1.2.6 Restrictions on the reflection phases

There are well established problems [28,29] which arise when refining a multipole model for a non-centrosymmetric crystal structure. These arise because of the phase ambiguity and

can result in poorly determinations of the odd-order multipole populations, which are invariant under certain crystal-class symmetry operations. In essence, some combinations of odd-order multipoles may make very small contributions to the structure factor amplitudes, but have significant contributions to the phases. In such cases, great care needs to be taken to ensure that physically meaningful parameters are obtained from the least squares procedure. As has been shown [28,29], the eigenvalue filtering method (used by default in XDLSM for singular or near-singular normal matrices) greatly alleviates these problems, though sometimes a 'slack constraint' on the scale factor (e.g. by including $F(000)$ as an additional observation with unit weight) may also be necessary [29] to obtain accurate parameters.

An alternative solution is to use fixed phases, determined from, say, a fully periodic *ab initio* calculation. In XD2024, a phase-constrained refinement is implemented through a special format for the reflection file XD.HKL, see **Table 2-2**. The NDAT entry number must be specified as -7, and the phase angle (in radians) must be supplied as the seventh data item for each reflection. No extra instructions in the master file are required. This phase-constrained strategy may be useful when refining against theoretical structure factors calculated on a non-centrosymmetric crystal structure.

4.2 Refinement strategy

A general rule, it is strongly advised that the complexity of the model should be increased in a stepwise manner. Each stage of the refinement could provide a hypothesis for the next step. In this respect it is difficult to suggest a specific scheme, in advance, according to which one should proceed. The spherical-atom refinement could serve as a reference for comparison during the whole fitting procedure. This could be followed by a restricted multipole refinement in which all possible chemical constraints and atomic pseudo-symmetries are applied. As argued above, the extent to which these restrictions should be applied depends on many factors. In most cases the number of observed intensity data limits the number of free variables. The ratio of the number of reflections to the number of variables should not fall considerably below 10. Atoms with the same valence and first coordination sphere should always be considered chemically equivalent at this stage of the fit. The spherical HF radial screening parameters (κ) can already be included. These variables, in contrast to those scaling the Slater exponents (κ'), are much more stable and their changes should stay below 5-10%. If the resolution and accuracy of the observations allows, the different restrictions can be released in subsequent refinement cycles, in the hope of testing the adequacy of the assumed chemical equivalences. In this way, 'second order effects' (crystal field, conformation differences, second neighbours, etc.) on the valence density might become visible. To decide if a new variable contributes significantly to the fit, the ratio of its value to its standard deviation and the change in the goodness of fit are to be checked. More sophisticated statistical tests will be available in follow-up versions of XDLSM.

Because of their low scattering power and intense thermal motion, hydrogen atoms should be treated with a special care. A poor model for their static density manifests itself in unreliable dynamic parameters and conversely, no reasonable estimate of the charge transfer can be obtained without meaningful displacement parameters. In organic molecules a considerable amount of the charge transfer occurs at the expense of charge on the hydrogen atoms. Due to the electro-neutrality constraint these uncertainties can seriously affect the result. To overcome this difficulty, the following strategies can be applied. The position and thermal parameters of the hydrogen atoms should be fixed at the values obtained by neutron diffraction, when such data are available. An overall scaling of the neutron displacement tensor components should be applied to account for the temperature difference (or rather the difference in the diffuse scattering) between the two data collections [26]. In the absence of neutron data, the parameters of the hydrogen atoms could be obtained from spherical-atom refinement using the contracted scattering factors of Stewart *et al.* [13]. The isotropic displacement parameters can then be fixed during the

multipole refinement. The correctness of this estimation can be judged by the distance of the bonds to the corresponding hydrogen atoms and by their net charges obtained in such a way. The ADP's of the hydrogen atoms can also be estimated by fitting the rigid-body or segmented rigid-body model to the motion of the non-hydrogen atoms. A simple riding model could also be feasible ($U(H) = 1.5 * U_{eq}(\text{non-H})$). Such a constraint can easily be incorporated. The density asphericities of the hydrogen atoms can be represented by a bond-directed dipole. For those involved in a strong hydrogen bond an additional quadrupole can also be introduced. The RESET BOND command (Section 4.6.5) is very useful here to constrain X-H distances to neutron determined standard values.

4.3 Dimensioning

The parameters in **Table 4-1** are used in certain DIMENSION and COMMON statements. These are the current limits for XDLSM. Future versions will have full dynamic memory allocation.

4.4 Variable names and order numbers

See **Table 4-2** for a list of symbols and code numbers to be used as variable identifications.

4.5 Files used and created by XDLSM

Input:	xd.mas, xd.inp, xd.hkl, xd.bnk_*
Output:	xd_lsm.out, xd.res
Optional output:	xd.fou, xd.der, xd.mat, xd.cov

4.6 Input instructions for XDLSM

The next section describes those commands which are interpreted by the program. All of these instructions must be placed between the MODULE *XDLSM and the END XDLSM lines in the xd.mas file.

4.6.1 Control instructions

4.6.1.1 SELECT

SELECT (*)model *m1 m2 m3 m4* **based_on** (*f|f²*) **(*)test** **verbose** *verbose_level*
SELECT cycle *cycles dampk dampk cmin cmin cmax cmax eigcut r convcrit convcrit*

(*)model *m1 m2 m3 m4*

This option provides a global control over certain parameters which characterize the structure factor formalism applied in the refinement. These parameters are shown in **Table 4-3**

The values given after the **model** option are applied for all atoms *only as an upper limit*. The option has only limited applications but can provide an easy way to reduce the complexity of the scattering formalism without having to modify all necessary parameters one by one. Note that certain combinations of the control parameters are meaningless, which might not be recognized by the program. If **model** is not starred, then it is assumed that $m3 = 1$, i.e. that the reflection data contain anomalous dispersion effects. If theoretical data (or

experimental data with anomalous dispersion removed) are being used, then ***model** *xx 0 x* must be used. An alternative approach in this case (if a global model control is not desired) is to set all the values of f' and f'' in the **SCAT** table (see **6.4.2.7** below) to zero.

based_on (f|f^2) The refinement is based on structure factors or on their squares. The data in the reflection file *xd.hkl* are transformed accordingly.

(*)test If flagged an input test is performed. This includes calculation and printing of

1. the scattering factor tables,
2. the local coordinate systems,
3. the variable-parameter list,
4. the matrix of constraints, together with the result of its singular value decomposition
5. a file *xd_scatter_atom.out* is printed for each atom type read in scat table in order to check two different calculations of the scattering factors (from the analytical expansion and from the wave function databank selected)

verbose verbose_level Setting *verbose_level* greater than one switches on extra printout to the log files.

cycle cycles [0]

>0	The number of least squares cycles requested.
=0	Structure factor calculation.
<0	Scale factor refinement.

dampk dampk [1.0]. This is a damping parameter applied to refinement of kappa's.

cmin cmin [0.6] cmax cmax [1.] Lower and upper limit used as a criteria for printing the correlation matrix elements.

eigcut r [1.e-10] If the solution of the system of normal equations are obtained through diagonalization, r is used as a cutoff limit for the singularity test. Eigenvalues are considered to be zero and omitted from the calculation of the inverse matrix until the inverse condition number is smaller than r :

$$\min(\text{eigenvalue}) / \max(\text{eigenvalue}) < r$$

This test is applied to the eigenvalues of the reduced matrix (derived from the constraints) and the conditioned matrix (see Introduction). The same parameter is used as a criteria for eliminating linear dependencies among the constraints. This is singular value decomposition of the matrix of the constraints, also known as eigenvalue filtering.

convcrit convcrit [0.0] If the maximum shift/su in the current cycle of refinement is less than *convcrit*, the program immediately proceeds to the final cycle and finishes the requested XDLSM job.

4.6.1.2 SAVE

SAVE (*)deriv (*)lsqmat (*)cormat

deriv If starred, the structure factor derivatives for each reflection (design matrix) in the last cycle are saved in the binary file *xd.der* (see Section 2.7).

lsqmat If starred, the least squares matrix and vector in the last cycle are printed to the binary file *xd.mat* (see Section 2.8).

cormat If starred, the variance-covariance matrix is written to the binary file *xd.cov*. This file is needed for estimating the standard deviations of different properties. The structure and the content of the file is given in **Table 2-4**.

4.6.1.3 SOLVE

SOLVE [*]inv (*)diag [*]cond

The solution of the least squares normal equation can be obtained through inversion or diagonalization.

inv If starred, the Gauss-Jordan for inversion method is implemented. The program will automatically switch to diagonalization if the matrix is found ill-conditioned (or singular) during the inversion in the first cycle. The matrix inversion is the default option.

diag If starred, the Householder reduction and QL algorithms are used for the calculation of the eigenvalues and eigenvectors. If an eigenvalue fails the test based on the condition number (see **eigcut**), the corresponding eigenvector is eliminated and printed.

cond If starred, the normal equation matrix is conditioned via the transformation 4.12, irrespective of the method of solution selected.

4.6.1.4 SKIP

SKIP (*)obs obsmin obsmax [*]sigobs sigmin sigmax (*)sinthl snlmin snlmax

The SKIP instruction defines criteria for rejecting observations from the refinement (not from the structure factor calculation). It is important to note that the rejection criteria are *always* applied, starring these options merely allows the use to change the values of the default criteria. If more than one are starred, the "AND" logic is applied. The rejection criteria and default values are:

obs [0.0, 1.0e10] all observations for which $obsmin > obs$ or $obs > obsmax$

sigobs [3.0, 1.0e10] observations for which $sigmin * sigobs > obs$ or $obs > sigmax * sigobs$

sinthl [0.0, 2.0] lower and upper limit in $\sin\theta/\lambda$

IMPORTANT The **obs** and **sigobs** cutoffs are applied to the data in the XD.HKL reflection file, *after any conversion* implied by the **based_on** criterion (Section 4.6.1.1). This action (which is more logical) is different from that previously applied in XDLSM, where the cutoffs were applied *before* conversion.

4.6.1.5 PRINT

PRINT (*)sinthl snlmin snlmax (*)obs obsmin obsmax (*)delta dmin dmax [*]del% min% max% (*)extcn extmin extmax (*)abssc

The **PRINT** instruction defines criterion for printing observations. After the last cycle the following quantities can be printed:

no h k l sinthl scgrp obs calc delta (del%|flag) extcn code

where

<i>no</i>	the order number of a reflection
<i>h k l</i>	reciprocal-lattice components of the scattering vector
<i>sinthl</i>	$\sin(\theta)/\lambda$
<i>scgrp</i>	scale group number
<i>obs</i>	F_o or F^2_o
<i>calc</i>	F_c or F^2_c
<i>delta</i>	$F_o - F_c$ or F^2_o or F^2_c
<i>flag</i>	a flag based on $f = 100 * (obs - calc) / obs $ It is a four character long string as follows:

```

'      '      for 0 < f < 5
*      '      for 5 < f < 10
**     '      for 10 < f < 15
***    '      for 15 < f < 20
****   '      for 20 < f < 25
?????  '      for 25 < f < 30
del%    f is printed instead of a flag
extcn   the extinction correction in percentage
code    0    included in the refinement
        -1   rejected based on criterion obs
        -2   rejected based on criterion sigobs
        -4   rejected based on criterion sinthl
        -3   rejected based on criteria obs and sigobs
        -5   rejected based on criteria obs and sinthl
        -6   rejected based on criteria sigobs and sinthl
        -7   rejected based on criteria obs and sigobs and sinthl

```

The options, if flagged, serve as a lower and an upper limit applied for printing. Again, the 'AND' logic applies.

(*)sinthl *snlmin snlmax* [0 2]

(*)obs *obsmin obsmax* [0 10]

(*)delta *dmin dmax* [-50 50]

(*)del% *min% max%* [80 100]

(*)extcn *extmin extmax* [80 100]

(*)abssc if flagged the observations are printed on an absolute scale

4.6.2 The SCAT table

The **SCAT** table provides a compact format for defining different scattering factors or modifying the entries in the databank file *xd.bnk_**. In contrast with previous versions of the program, the SCAT table now includes all atomic orbitals. If an old *xd.mas* file is used, this table *must* be modified otherwise it will not be read correctly. The heading of the SCAT table is:

SCAT core sphv defv 1s 2s 3s 4s 2p 3p 4p 3d 4d 4f 5s 5p 6s 6p 5d 7s 6d 5f Δf' Δf'' nsctl

core	core scattering factor
sphv	spherical valence scattering factor
defv	scattering factors due to valence deformation functions
1s 2s 3s ...	occupations of HF atomic orbitals
Δf'	real part of anomalous dispersion correction
Δf''	imaginary part of anomalous dispersion correction
nsctl	neutron scattering length

This **SCAT** line has to be followed by as many input lines or subsegments as atom types are present in the unit cell. Each row should begin with the element name that must be identical to one of the atom types stored in *xd.bnk_** (see element naming convention in section 2.5). If the element name is the only string in the input line, the data on the corresponding segment of the databank file will be used to create the scattering factors. The databank file can be extended by introducing new segments assigned to dummy atom names. In this way considerable freedom is provided for designing scattering factors from atomic wavefunctions expanded over Slater-type basis functions. All the entries indicated above have default assignments. To change the default assignment of a particular entry all preceding entries in the list have to be given. For example, to change the default values

for the anomalous dispersion corrections ($\Delta f'$ and $\Delta f''$) all three types of scattering factors as well as the occupations have to be input.

For the scattering factors the following options are available:

```
core [chfw]  rdtb
sphv [chfw]  rdtb  rhft
defv chfw    rdtb  [cszd] rdsd
```

4.6.2.1 chfw - Clementi's Hartree-Fock Wavefunction

This is the default option for the core and sphv scattering factors and it means that the Slater-type atomic orbitals stored on the `xd.bnk_*` file are used. The user has the freedom to decide what to consider core and what valence density. This is done by specifying the orbital occupations, which have to be *positive* or *negative* integers for *core* or *valence* orbitals respectively. If they are omitted, the default configuration in `xd.bnk_*` is taken. The order of the orbitals is given in the heading of the **SCAT** table. For example, the default configuration of the ground state carbon atom is (1s²), (2s²,2p²) and the corresponding line in the input table is:

```
C    chfw chfw cszd    2 -2  0  0 -2
```

In this case $2\langle j_0 \rangle(1s1s)$ and $(2\langle j_0 \rangle(2s2s) + 2\langle j_0 \rangle 2p2p)/4$ is calculated, respectively, for the core and the spherical valence scattering factors. Note, that the sphv scattering factor is normalized, but not the core.

A 'frozen' spherical atom (only core or spherical atom scattering) could be defined as

```
C    chfw chfw cszd    2  2  0  0  2
```

while that of with radial screening (only valence or spherical atom scattering)

```
C    chfw chfw cszd   -2 -2  0  0 -2
```

Another application of the orbital occupations is to form spherical valence scattering factors corresponding to an assumed hybridization. For example, one can 'generate' an sp³ type carbon atom with the following input

```
C    chfw chfw cszd    2 -1  0  0 -3
```

which assigns $(\langle j_0 \rangle(2s2s) + 3\langle j_0 \rangle(2p2p))/4$ to the spherical valence scattering factors. Since more than one sets of scattering factors can be generated from the same wavefunction, the multiple use of an element name is allowed.

4.6.2.2 rdtb - Read table

This option is available for all three types of scattering factors. It indicates that the corresponding scattering factor table is to be read from the master file. For an unknown element (not stored in the `xd.bnk_*` databank file) the **rdtb** option *must* be specified. The input should consist of *ngrd* values (8 entries/lines) of the function taken at an equidistant grid of $\sin\theta/\lambda$ with a step size of *grd*. *ngrd* and *grd* are parameters with default values of 40 and of 0.05, respectively, in the present version of XDLSM (See **Table 4.1**). The first grid point must be zero. The default setup requires the table to be given up to 1.95 in $\sin\theta/\lambda$. The parameters *ngrd* and *grd* should be adjusted to the wavelength of the radiation used for the data collection. The scattering factor at an arbitrary scattering angle is interpolated and the derivatives with respect to the expansion-contraction parameters are numerically obtained. Accurate evaluation require a considerably fine grid size (not exceeding 0.06 Å⁻¹).

Example:

```
C    RDTB RDTB CSZD
```

2.00000	1.99642	1.98575	1.96816	1.94394	1.91349	1.87726	1.83581
1.78973	1.73965	1.68621	1.63006	1.57183	1.51212	1.45148	1.39046
1.32950	1.26904	1.20944	1.15100	1.09400	1.03863	0.98506	0.93343
0.88381	0.83628	0.79085	0.74754	0.70632	0.66717	0.63004	0.59488
0.56163	0.53021	0.50055	0.47258	0.44621	0.42137	0.39798	0.37597
1.00000	0.93697	0.77692	0.58120	0.40061	0.25845	0.15714	0.08962
0.04686	0.02103	0.00626	-0.00155	-0.00512	-0.00622	-0.00596	-0.00502
-0.00381	-0.00256	-0.00140	-0.00037	0.00048	0.00118	0.00173	0.00216
0.00247	0.00269	0.00283	0.00291	0.00294	0.00294	0.00291	0.00285
0.00278	0.00269	0.00260	0.00250	0.00240	0.00230	0.00220	0.00210

4.6.2.3 cszd, rdsd - Single-zeta density parameters for defv

By default, the radial functions of the valence deformation density are of single Slater-type (**cszd**). The parameters of the radial functions ($n(l)$, $\xi(l)$) are obtained from the corresponding single- ξ wavefunctions of Clementi & Roetti [14] stored also in `xd.bnk_*` files. In previous versions of XD, the option **cszd** in the SCAT table computed the exponents for the radial deformation functions by simply averaging the valence exponents of the "best" single- ξ orbitals (Clementi and Raimondi [15]) of the default configuration. The new version of the program now computes ξ 's by weighting the orbitals by their occupation. For noble gases and closed shell ions, ξ is computed thus :

- from the (weighted) exponents of the outermost shell of the core for noble gases and anions (e.g. the 2s and 2p orbitals for F^- , Ne etc.)
- from the exponents of the first empty orbital(s) for closed-shell cations (3s for Na^+ , Mg^{2+} etc.; 3s and 3p for Al^{3+} , Si^{4+} etc.; 4s for K^+ , Ca^{2+} ; 3d for Sc^{3+} , Ti^{4+} etc.).

The closed-shell configurations recognized are those of the noble gases (thus, 2, 10, 18, 36, 54 electrons), and those of some cations of the 4th and 5th row (like Cu^+ , Ga^{3+} , Sb^{5+}), which may have 28 or 46 electrons (single- ξ exponents considered are those of 3d and 4d orbitals, respectively). All other configurations missing the valence electrons are not recognized by the program, which then stops.

The orbitals used to compute the average are directly linked to the SCAT table configuration. Thus, if the user modifies the number or the type of valence electrons (at his own risk!) in the SCAT table, then Z will change. Note that in the previous versions of XD, the SCAT table was intended to modify just the SPHV monopole, evaluated by the multi-exponent HF wave functions of Clementi and Roetti [14].

Warning messages will appear in the output if the configuration chosen is unusual or dangerous and severe stops are applied if the requested orbitals are not stored for a given atom.

The default values can be modified by using the option **rdsd** which makes it possible to input all $n(l)$ and $\xi(l)$ in atomic units:

```
C      chfw chfw rdsd
n(0) zeta(0) n(1) zeta(1) n(2) zeta(2) n(3) zeta(3) n(4) zeta(4)
```

4.6.2.4 The **chfw** option for **defv**

An advanced feature of XDLSM is to allow for the use of HF radial functions for the deformation density. Such application needs each $\langle J_l \rangle$ to be attributed to a proper combination of orbital products. The Table given in the Introduction (Section 1.9) summarizes the different order of Fourier-Bessel transforms that occur for the different orbital products. An orbital product is given by the names of the comprising orbitals in brackets: (2s2s), (2p2p), (3d3d), etc. If more than one product contributes to $\langle J_l \rangle$ they should be connected by the plus "+" sign. The character string composed in such a way must contain no embedded blanks : (2s2s)+(2p2p) ... etc. A product or a sum of products contributing to $\langle J_l \rangle$ has to be specified for each l . Note that not all radial densities can be constructed from a given wavefunction. To satisfy Poisson's equation [27], $2(n-1)$ must be $\geq l$, so for example a 2s orbital cannot be used to construct octupole or hexadecapole radials parts. To make a complete set, all options available for **defv** can be combined, as explained below. If neither **cszd** nor **rdsd** is specified for **sphv** the program expects additional input lines with one of the following contents:

```
l [cszd]
l chfw conf
l rdsd n(l) zeta(l)
l rdtb
```

NOTE : The **chfw** option in XD means that the radial part (and only the radial part) of the specified orbital product will be used to construct a density function. The XD program allows such a radial function to be associated with any multipole, so that a **chfw** directive such as "l chfw <3d3d>" is permissible, despite having limited physical meaning.

For each l , an option can be selected which determines any further input. If no line is given for certain l values the default (**cszd**) applies. For **chfw** the configuration (*conf*) is to be given in terms of orbital products or their sum. For **rdsd** the parameters of the radial functions are needed. After **rdtb** a scattering factor table is to be read as described above.

Example:

```
C      chfw chfw chfw  2 -2  0  0 -2
0 chfw  (2s2s)+(2p2p)
1 chfw  (2s2p)
2 chfw  (2p2p)
3 rdsd  3 3.71
4 rdtb
0.00000  5.99918  4.95113  3.64245  2.42954  1.49816  0.87092  0.48586
0.26409  0.14158  0.07557  0.04045  0.02182  0.01190  0.00658  0.00370
0.00211  0.00122  0.00072  0.00043  0.00026  0.00016  0.00010  0.00006
0.00004  0.00003  0.00002  0.00001  0.00001  0.00001  0.00000  0.00000
0.00000  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000  0.00000
```

In the above example $\langle J_0 \rangle$ is the sum of the 0-th order transforms of ss and pp type radial functions and irrespective of the normalization it is equivalent to **sphv**. The $\langle J_1 \rangle$ and $\langle J_2 \rangle$ functions are related to sp and pp type orbital products, respectively. The scattering factor for octupoles is created from single- ζ radial functions while that for hexadecapoles is read in.

4.6.2.5 **rhft** - Relativistic Hartree-Fock scattering factors for **sphv**

If a spherical atom model is selected, the RHF scattering factors, as given in the International Tables [16] or in references [17,18] in the form of an expansion over Gaussian functions, can also be used. The **rhft** option for hydrogen selects the contracted scattering factors of Stewart, Davidson & Simpson [13].

4.6.2.6 Current Limitations

The calculation of the static electron density and of electronic properties requires the evaluation of the radial functions within an accuracy that can hardly be reached by numerical inverse Fourier transform of the scattering factors. The default choice, the use of Slater-type HF wavefunctions (**chfw**, **cszd** or **rdsd**), means analytical representation of both direct and reciprocal space functions. For a refinement to be consistent with the property calculation, it *must* be based on the wavefunctions stored in `xd.bnk_*`. The corresponding scattering factors are certainly not the best available ones and can slightly differ from those found in the *International Tables*. Relativistic effects are important only for heavier elements - this can be seen by comparing the total chfw spherical scattering factors with those based on relativistic numerical wavefunctions. Efforts are being made to eliminate this limitation.

4.6.2.7 Anomalous scattering

delf' delf''

The defaults correspond to Mo radiation. XDINI now computes the correct values for any arbitrary X-ray wavelength.

4.6.2.8 Neutron Scattering Length

nsctl

The last entry of a **SCAT** line is the neutron scattering length. **NOTE** Compounds containing elements with very large absorption cross-sections (B, Cd, In, Sm, Eu, Gd, Dy), have an imaginary component of the scattering length, and cannot be dealt with in the current version of XDLSM. The current databases have an entry for deuterium (element D) which is identical to H except for the neutron scattering length and atomic weight. **NOTE** In previous versions of XD, the values of the neutron scattering lengths listed in the master file were in units of 10^{-14} m, but are now given in the more usual Fermi units of 10^{-15} m.

4.6.3 The ATOM table

For each atom included in the structure factor calculation the following entries are to be given:

ATOM atom0 ax1 atom1 atom2 ax2 r/l tp tbl kap lmx sitesym chemcon

4.6.3.1 Atom name conventions

The atom name is a continuous string of up to 8 characters, starting with a correct, case sensitive chemical symbol (*e.g.* 'Na' and not 'NA') used in the **SCAT** table and followed by further characters enclosed in parentheses (). Legal atom names are:

Cu(3) Ti3+(1a)

4.6.3.2 The local coordinate system

The entries in the first seven columns define the local coordinate systems. *atom*, *atom0*, *atom1* and *atom2* are atom names from the **ATOM** table list. *ax1* and *ax2* stand for different axis assignments, each being either x or y or z. The first axis (*ax1*) is given by the internuclear vector from atom to atom0 (\mathbf{v}_1). This together with the second vector from *atom1* to *atom2* (\mathbf{v}_2) define the (*ax1*,*ax2*) plane. The third vector (\mathbf{v}_3) is taken perpendicular to this plane.

$$\mathbf{v}_1 = (\mathbf{r}_0 - \mathbf{r})$$

$$\mathbf{v}_2 = (\mathbf{r}_2 - \mathbf{r}_1)$$

$$\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$$

Finally, an orthonormal vector triplet (\mathbf{e}_{ax1} , \mathbf{e}_{ax2} , \mathbf{e}_{ax3}) is formed which can be chosen to be either right (**R**) or left (**L**) handed

$$\mathbf{e}_{ax1} = \mathbf{v}_1 / |\mathbf{v}_1|$$

$$\mathbf{e}_{ax2} = (\mathbf{v}_3 \times \mathbf{v}_1) / |(\mathbf{v}_3 \times \mathbf{v}_1)|$$

$$\mathbf{e}_{ax3} = \mathbf{v}_3 / |\mathbf{v}_3|$$

where \mathbf{r} , \mathbf{r}_0 , \mathbf{r}_1 and \mathbf{r}_2 are the position vectors of atom, atom0, atom1 and atom2, respectively (MOLLY).

4.6.3.3 tp - the Order of the Atomic Displacement Tensor

- 0 no thermal parameter is applied (static scattering model)
- 1 isotropic U
- [2] anisotropic U_{ij}
- 3 anharmonic 3rd order Gram-Charlier expansion C_{ijk}^{ijk}
- 4 anharmonic 4th order Gram-Charlier expansion D_{ijkl}^{ijkl}

If this value differs from that in the input file the thermal displacement parameters will be converted:

xd.mas	xd.inp	
1	2	from anisotropic to isotropic
2	1	from isotropic to anisotropic

4.6.3.4 tbl - The Core Scattering Table

Refers to the core scattering table. It is the order number of the corresponding element given in the **SCAT** table.

4.6.3.5 kap - the Kappa Set

Defines the kappa set applied to the valence radial functions. *If a new set is introduced or the previous arrangement is redefined, the corresponding changes must also be made in the parameter file xd.inp and/or to the **KEY** table.* If the values in the parameter file are not changed the refinement will start from the default value (1.0) for all kappa sets.

4.6.3.6 lmx - the maximal level of multipole expansion

- 0 Monopoles (sphv and defv)
- 1 Dipoles (default for hydrogen atoms)
- 2 Quadrupoles
- 3 Octupoles
- 4 Hexadecapoles (default for non-hydrogen atoms)

4.6.3.7 sitesym [1]

The point group symbol of the atomic site symmetry. These symbols initialised by XDINI are for information only and are not used by any XD program. The ISYM codes in the parameter file refer to the crystallographic site symmetries of atoms in special positions (codes 1-77) and the user imposed model pseudo-symmetries based on the 32 crystallographic point groups (codes 80-112). See tables 3-1 & 3-2 in the Appendix for further details.

4.6.3.8 chemcon

Refers to the atom to which the valence deformation density of the atom considered is constrained. Atoms with different site occupancies cannot be linked by chemcon constraints. If it is desired to link the multipole populations of atoms having different site occupancies, this must be done using an explicit set of CON instructions (see Section 4.6.10), ensuring that the different occupancies are properly taken into account. *If the same set of multipole populations are to be shared by two or more atoms, the definition of the local coordinate systems of the corresponding atoms must be consistent.*

4.6.3.9 Dummy Atoms

To enable one to define a local system of arbitrary orientation, dummy atoms can be used. These are to be specified after the atom list but within the atom table by giving a name, composed of the string "DUM" and a number, followed by the three coordinates (free format) in the crystal system.

Example:

```

ATOM      atom0      ax1 atom1      atom2      ax2 r/l tp tbl kap lmx sitesym  chemcon
O(1)      O(2)       X  O(1)      DUM0       Y  R  2  1  1  4  NO
O(2)      O(1)       X  O(1)      DUM1       Y  R  2  1  1  4  NO
.
.
DUM0      0. 0. 0.
DUM1     -0.4800  0.5335  0.0973
END ATOM

```

4.6.4 GROUPn

GROUP*n* atom(1)... with *n* > 1

The **GROUP** command selects a set of atoms to be considered as a unit for special applications. Such applications available presently are the electroneutrality and rigid-body, rigid-link type constraints see (**KEEP** instruction). The first group (**GROUP1**) is, by definition composed of the atoms in the asymmetric unit. An atom is allowed to be part of more than one group. The atom list defining a group can be on more than one input line, but each line must start with the same **GROUP***n* command.

4.6.5 KEEP

The KEEP instruction simplifies the application of certain constraints.

KEEP kappa set(1) ...

For each set defined κ'_l , the expansion-contraction parameter of defv, is kept the same for all *l*. This is a default constraint that is suggested to apply, at least in the initial stages of a refinement.

IMPORTANT The fit is always very sensitive to κ'_l , even if a single parameter is refined for all *l* values. The results of κ'_l refinement should *always* be critically examined and compared to those obtained with $\kappa'_l = 1.0$. For HF radial functions, the chance of obtaining convergence with reliable estimates of different κ'_l parameters is expected to be better than for single Slater orbitals. A separate κ'_l refinement is worth trying for transition metals.

KEEP charge [group1] group*n* ...

Each group defined by the **GROUP** command can be treated as a closed unit for which the total charge is kept fixed during the refinement. The total charge of the group is given by the sum of the starting monopole populations of the comprising atoms. The user is free to define any subset of atoms (even having common elements) which are excluded from charge transfer. Each group fixed in this way adds one new equation to the system of constraints. A zero singular value of the matrix of constraints means inadequate grouping and the redundancy found will be rejected. The **KEEP charge group1** directive implements the monopole population constraint, and in normal situations this constraint should always be present.

KEEP rigid [group1] group*n* ...

Each group defined by the **GROUP** command is kept rigid in the sense that the shifts in the ADP's of the atoms comprising the group are constrained to satisfy Hirshfeld's rigidity postulate. To make such a restriction work, all ADP's of all atoms in the group must be refined. The equation of constraint is set for all internuclear connections in the group and the linearly dependent equations are eliminated leading to the necessary reduction in the number of restrictions.

RESET BOND atom H-atom distance (Å)

The distance between an H-atom and its bonded atom is reset to the designated (usually neutron determined) distance at the end of each cycle. The H-atom *must* be listed second, and its coordinates should not be refined.

RESET KAPPA (*)all (*)onlykp 1 1.0 2 1.0 3 1.0

The kappa values used in XDLSM may be reset prior to refinement using this new directive to avoid manually editing the xd.inp file. This is especially useful when the kappas have refined to unrealistic values, perhaps because of inappropriate refinement strategies or applied constraints. It is a global command which resets all the kappa's in the specified kappa set *kset* to the specified value *val*. The instruction has pairs of numbers, the first in the pair (*kset*) is an integer referring to the kappa set, the second (*val*) to its desired value. Either specific kappa sets may be modified, or if the **all** directive is selected (starred) then all kappa values will be reset to the value specified in the first *kset/val* pair or if that is missing then the default value of 1.0 will be used. If the **onlykp** directive is activated (starred), then only the deformation kappa's K0-K4 are affected by this directive. Unlike the **RESET BOND** directive, it is only sensible to use this directive in an initial xdlsm refinement and then switch it off (i.e. comment it out) for subsequent refinements.

4.6.6 The Weighting Scheme

WEIGHT a b c d e f

a [0.0] *b* [0.0] *c* [0.0] *d* [0.0] *e* [0.0] *f* [1/3]

For refinement on F^2 , the weighting scheme as implemented in SHELXL is used:

$$w_2 = q / [s_2^2 + (ap)^2 + bp + d + e \times \sin(\theta)]$$

where

$$\begin{aligned} s_2 &= \sigma(F_o^2) \\ p &= f \times F_o^2 + (1-f) \times F_c^2 \\ q &= 1.0 && \text{if } c = 0 \\ \text{or } q &= \exp[c \times (\sin\theta/\lambda)^2] && \text{if } c > 0 \\ \text{or } q &= 1.0 - \exp[c \times (\sin\theta/\lambda)^2] && \text{if } c < 0 \end{aligned}$$

For refinement on F , the weight (w_1) is calculated as follows:

$$w_1 = [F_o \times \sqrt{w_2} + \alpha]^2$$

where

$$\begin{aligned} \alpha &= 0.0 \text{ if } \beta < 0.0 ; \quad \alpha = \sqrt{\beta} \text{ if } \beta > 0.0 \\ \beta &= (F_o^2 \times w_2) + \sqrt{w_2} \text{ where } w_2 \text{ is calculated as above.} \end{aligned}$$

IMPORTANT This general weighting scheme has been developed for refinement based on a conventional, spherical atom model and thus may not be adequate for multipole refinement.

There are two special weighting schemes ; when $a = -1.0$ and when $a < -1.0$. These two schemes are applied regardless of the input values of the other weighting parameters $b - f$. For a multipole refinement on F , it is suggested to set $a < -1.0$, when statistical weights [i.e. $w_2 = 1/\sigma^2(F_o^2)$ and $w_1 = 1/\sigma^2(F_o)$] will be applied. The default xd.mas written by XDINI has $a = -2.0$, which sets this condition for refinement on either F or F^2 . Note that the values of zero

for the parameters $a - e$ given above also imply statistical weights for refinement on F^2 , but NOT for refinement on F !

To apply units weights, set $a = -1.0$. The parameters of the weighting form cannot be refined. Two goodness of fit parameters are printed in `xd_lsm.out`, one (GOFw) based on the weighting scheme used in refinement, and one (GOF) based on statistical weights $w = 1/s^2$.

4.6.7 DMSDA

DMSDA *rmin rmax*
rmin [1.1] *rmax* [1.8]

The difference of the projections of the mean square amplitude tensors of two atoms to the corresponding internuclear vector are calculated if the interatomic distance falls in the range given by *rmin* and *rmax*. As discussed in the Introduction, Hirshfeld's rigid-bond test [7] can help to reveal model inadequacies and should always be a part of a careful analysis. The positional coordinates and the anisotropic displacement parameters in an orthogonal system are also printed.

4.6.8 Extinction refinement

The following instruction line for extinction correction appears in `xd.mas`:

EXTCN (*)**iso** (*)**aniso** (*)**type_1** (*)**type_2** (*)**type_3** (*)**distr_g** (*)**distr_1** (*)**msc_0** (*)**msc_1**

For any extinction refinement, the absorption weighted path length (*tbar*) should be stored in the reflection file `xd.hkl`. For an anisotropic extinction refinement, an additional six data items (the direction cosines - see Table 2-2) for each observation are also required. If *tbar* is missing from `xd.hkl`, then for an isotropic extinction only, an identical value of 0.5 mm is assumed for all reflections. This assumption will, in general, lead to incorrectly scaled extinction parameters, but this is not often of major concern. For anisotropic extinction refinement, *tbar* values in the reflection file are mandatory.

By default the EXTCN command appears as a comment in `xd.mas`. To get it activated the exclamation mark (!) has to be removed. The extinction correction is based on the models proposed by Becker and Coppens [19-21], which can be summarized as follows:

4.6.8.1 Isotropic extinction (*iso, default):

extinction type:

- **type 1** (*type_1, default): mosaic spread, the g' coefficient is refined (variable EXT11);
- **type 2** (*type_2): particle size, the ρ coefficient is refined (variable EXT11);
- **type 3** (*type_3, *type_g is also accepted): generalized type, mosaic spread and particle size (g' and ρ) are simultaneously refined (variables EXT11 and EXT22, respectively);

mosaic spread distribution (active only for type_1 and type_3):

- Gaussian (*distr_g): a Gaussian distribution is assumed;
- Lorentzian (*distr_1, default) : a Lorentzian distribution is assumed;

4.6.8.2 Anisotropic extinction (*aniso):

extinction type:

- **type 1** (*type_1, default): mosaic spread, the **Z** tensor is refined (variables EXT11-EXT23) ($g(\mathbf{D}) = (\mathbf{D}^t \mathbf{Z} \mathbf{D})^{1/2}$; **D** is a unit vector perpendicular to the diffraction plane);

- **type 2** (*type_2): particle size, the **W** tensor is refined (variables EXT11-EXT23) ($\rho(\mathbf{N}) = \lambda(\mathbf{N}^t \mathbf{W} \mathbf{N})^{-1/2}$; **N** is a unit vector in the diffraction plane, perpendicular to the incident beam);
- **type 3** (*type_3, *type_g is also accepted): generalized type (mosaic spread and particle size), **Z** tensor and ρ simultaneously refined (variables EXT11-EXT23 for **Z**; variable RHOEX for ρ);

mosaic spread distribution (active only for type_1 and type_3):

- Gaussian (*distr_g): a Gaussian distribution is assumed;
- Lorentzian (*distr_l, default) : a Lorentzian distribution is assumed;

mosaic orientation (active only for type_1 and type_3):

- Coppens and Hamilton (*msc_0): the distribution proposed by Coppens and Hamilton [22];
- Thorney and Nelmes (*msc_1, default): the distribution proposed by Thorney and Nelmes [23]. In this case the **Y** tensor is refined instead of **Z** ($g(\mathbf{D}) = (\mathbf{D}^t \mathbf{Y} \mathbf{D})^{-1/2}$).

Given the expression for tensors **W**, **Y**, and **Z**, a switch from isotropic to anisotropic correction requires the following variables in `xd.inp`:

- (1) g' to **Z**:
 $\text{EXT11} = \text{EXT22} = \text{EXT33} = (g'_{\text{iso}})^2$
 $\text{EXT12} = \text{EXT13} = \text{EXT23} = 0.0$
- (2) g' to **Y**:
 $\text{EXT11} = \text{EXT22} = \text{EXT33} = 1/(g'_{\text{iso}})^2$
 $\text{EXT12} = \text{EXT13} = \text{EXT23} = 0.0$
- (3) ρ to **W**:
 $\text{EXT11} = \text{EXT22} = \text{EXT33} = 1/(\rho_{\text{iso}})^2$
 $\text{EXT12} = \text{EXT13} = \text{EXT23} = 0.0$

When a non-positive definite tensor (**W**, **Y** or **Z**) is obtained, the program stops if the automatic resetting of the tensor fails.

In the output file `xd_lsm.out`, the following parameters are reported:

- for isotropic extinction: the mosaic spread, η (η is proportional to $1/g'$, units in seconds) and the domain size r ($r = g'\lambda 10^{-4}$, units in centimeters) are given, as derived from the refined g' and/or ρ .
- for anisotropic extinction: The principal axes of mosaic spread $\eta(\mathbf{D})$ (if *type_1 or *type_3) or domain size $r(\mathbf{N})$ (*type_2) distribution are given; the corresponding 'equivalent' η and r scalars are computed (or the refined r is printed if *type_3 is applied).

4.6.9 FOUR - Structure factor calculation

FOUR fmod1 m1.1 m1.2 m1.3 m1.4 fmod2 m2.1 m2.2 m2.3 m2.4

The **FOUR** command ensures that a Fourier file `xd.fou` is written after the last cycle. Structure factors based on two models but on the same set of parameters are calculated and saved together with F_{obs} and the phases. The latter quantities may have the effects of anomalous dispersion removed [34] as follows.

$$A_{\text{obs}}^0 = \frac{F_{\text{obs}} A_{\text{calc}}}{F_{\text{calc}}} - (A_{\text{calc}} - A_{\text{calc}}^0)$$

$$B_{obs}^0 = \frac{F_{obs} B_{calc}}{F_{calc}} - (B_{calc} - B_{calc}^0)$$

$$F_{obs}^0 = \sqrt{(A_{obs}^0)^2 + (B_{obs}^0)^2}$$

where A and B are the real and imaginary parts of the structure factor F and a zero superscript (0) designate terms without anomalous dispersion. The standard deviation of F_{obs}^0 is estimated as

$$\sigma(F_{obs}^0) = \sigma(F_{obs}) \frac{\sqrt{(A_{obs}^0 A_{calc})^2 + (B_{obs}^0 B_{calc})^2}}{F_{obs}^0 F_{calc}}$$

The phases are based on the model applied in the refinement (see instruction **MODEL**).

Each of the structure factor models (**fmod1**, **fmod2**) is specified with four integers, in the same way as described above. The combination of **fobs**, **fmod1** and **fmod2** makes it possible to generate six different Fourier maps (see XDFOUR & XDFFT). By default, the **FOUR** command appears as a comment line in `xd.mas`. In the example below, a Fourier file is created with two calculated structure factors. The first one based on a multipole model ($lmax=4$), the second one on a neutral spherical atom model ($lmax=-1$). Both are free of anomalous dispersion and extinction.

```
FOUR  fmod1 4 2 0 0  fmod2 -1 2 0 0
```

IMPORTANT In contrast to earlier versions of the program, it is the values of *m1.3* and *m1.4* for **fmod1** which determine the treatment of F_{obs} . Anomalous dispersion effects will only be removed from F_{obs} if *m1.3* is given as zero. Likewise, F_{obs} will only be corrected for extinction effects (if any) when *m1.4* is given as zero. Since it is normally desired that Fourier syntheses should represent the electron density, and be independent of the wavelength used for the experiment, the effects of anomalous dispersion must be removed from experimental F_{obs} [34,35]. The values of *m1.3* and *m1.4* must therefore be given as zero, as in the above example.

Another new feature in version 5.3 is the addition (at the end of the `xd.fou` file) of an entry for reflection $F(000)$. The inclusion of $F(000)$ in Fourier summations ensures a correct absolute scale for the electron density.

Only those reflections used in the refinement will be written to the XD.FOU file.

4.6.10 CON - General linear constraint

CON *a1 var1 a2 var2 a3 var3 ... = c*

The **CON** instruction defines a linear combination among a set of variables used as a constraint. A list of the coefficients (*a1,a2,a3,...*) and variable-symbols (*var1, var2, var3,...*) are to be given terminated by the equal sign ("="), which is followed by the last entry (*c*) to define the right side of the equation. The coefficients and the variable names are read as numeric and character fields, respectively. A variable name is composed from the corresponding symbol described before (**Table 4.2**) and from a number (if needed) referring to the atom (κ set or scale group) to which the variable is assigned. The two components of the name are divided by the slash ("/"). The resulting character string must *not* contain blanks. The following are correct variable names:

X/1, U12/12, C333/1, M1/2, H4+/11, KS/3, K2/2, SCALE/1, EX12

The list of coefficients and variable names *must* be terminated by the equal sign. More than one line can be input with the same **CON** command. A new line is read until the equal sign is found. *Note, that here, what is meant by the term "variable" is actually the shift in that variable and not the variable itself.*

Applications of great practical importance are the constraints due to crystallographic site symmetries. *The correct constraints applicable to those atoms in special positions are now given in the xd.mas file written by the updated XDINI. These should never be modified unless you know what you are doing !* The violation of a symmetry restriction leads to singular least squares matrix. If the solution is obtained via diagonalization, the singularity can be eliminated (the corresponding constraint is introduced). Although this procedure might work in most cases, it is not advisable to let a numerical procedure handle the symmetry. After several cycles, round-off errors are likely to break the symmetry in the shifts. Symmetry restrictions can easily be formulated. Either the variable itself or the combination of two variables are fixed. The former does not need extra **CON** card since the corresponding variable is simply not refined (see **KEY** table).

Example: Formula KHF_2 , space group $I\ 4/mcm$ (No. 140). The three atoms in the asymmetric unit occupy the following special positions:

No.	atom	Wyckoff letter	x	y	z
1	K	a	0	0	1/4
2	F	h	x	x+1/2	0
3	H	d	0	1/2	0

The corresponding **CON** cards are:

```
CON 1 X/2 -1 Y/2 = -0.5
CON 1 U11/1 -1 U22/1 = 0
CON 1 U11/2 -1 U22/2 = 0
CON 1 U11/3 -1 U22/3 = 0
```

Another example of practical importance is in the use of Kubic harmonics. Table 4-4 shows the density normalized Kubic harmonics as linear combinations of density normalized spherical harmonics, and also indicates which Kubic harmonics are symmetric with respect to the cubic site symmetries (and hence are allowed). Suppose that atom 1 is at a symmetry site of $-43m$ in a cubic space group. Table 4-4 shows that the only symmetry-allowed Kubic harmonics (of relevance to the current version of XD) are $K_{0,1}$, $K_{3,1}$ and $K_{4,1}$. The first two simply correspond to M1,M2 and O2- (y_{00} and $y_{3,-2}$) while $K_{4,1}$ is a linear combination of the hexadecapoles H0 and H4+ ($0.78245\ \text{H0} + 0.57939\ \text{H4+}$). The corresponding constraint is thus

```
CON 0.57939 H0/1 -0.78245 H4+/1 = 0
```

or more simply

```
CON 1 H4+/1 -0.74048 H0/1 = 0
```

IMPORTANT If three parameters are involved in a cyclic constraint, *e.g.* if $U11=U22=U33$ (a common occurrence in cubic space groups), the user should note that all three constraints are always explicitly listed in the master file written by XDINI. Their linear dependency is noticed and removed by the SVDCMP routine, but the user should be aware that this routine is not infallible, as rounding errors may affect the logic. If the SVDCMP routine fails to determine the correct number of non-redundant constraints, the user is advised to remove the redundant constraints by commenting them out in the master file. This procedure normally solves any issues. If such a cyclic constraint is being applied in a non-mandatory situation, *e.g.* to impose some desired pseudo-symmetry, then only the two non-redundant constraints need normally be given.

The mandatory site symmetry restrictions are now handled automatically in XDINI. Tables 4-5 – 4-9, previously in this manual are now given in the Appendix. Site symmetry restrictions on spherical harmonics are given in Table 4-5 [25]. Tables 4-6 to 4-9 (simplified versions of Tables 5.5A - 5.5D from reference 24) are reproduced with kind permission of the IUCr. These Tables provide the site symmetry restrictions on the U^{ij} anisotropic thermal motions tensors and on the Gram-Charlier anharmonic coefficients C^{ijk} and D^{ijkl} . The order of indices for C^{ijk} and D^{ijkl} in these Tables follows the order of symbolic names used by XD (see Table 4-2). Table 4-6 should be consulted first, to obtain the appropriate cross reference key for the crystallographic site symmetry in question.

4.6.11 The KEY table

KEY xyz -U2- ----U3-----U4----- M- -D- -Q- -O- ----H----

This input segment is to specify which parameter is to be refined and which is not. It is done by giving the **KEY**-integer array with values 0 or 1 for a fixed or for a refined parameter, respectively. The order of the parameters is as defined before in **Table 4-2**. First the atomic parameters ($x, y, z, U^{ij}, C^{ijk}, D^{ijkl}, P_{lm}$) have to be given for all of the atoms included in the structure model. These are followed by the "shared" ($\kappa; \kappa'$) and by the "global" parameters (extinction, overall thermal parameter, scale factors). The heading of the **KEY** table helps in keeping account of the variables. The different abbreviations are as follows:

xyz	3 positional coordinates
U_n, C_n, D_n	n th order displacement tensor components. There are 6, 10 and 15 components for $n=2, 3$ and 4, respectively
M	2 monopole populations; the first for sphv and the second for defv
D	3 dipole
Q	5 quadrupole
O	7 octupole
H	9 hexadecapole populations
KAPPA	1 for sphv and 5 for defv ($\kappa', l=0, lmx$). It should be given for each KAPPA set defined in the ATOM table
EXTCN	1, 6 or 7 extinction parameters
OVTHP	1 overall thermal parameter
SCALE	NQ scale factors

As many atom entries are to be given as in the atom table. The atom names used here have to be identical to those in the atom table otherwise the program terminates with error message. Similarly, the number of kappa entries must be equal to the maximal number used in the atom table to refer to kappa sets (see 4.6.3.5) If the command **KEEP kappa** is applied to a set then all but the first kappa-integers for the corresponding devf ($\kappa', l=1, 4$) should be zero. The number of key integers for the scale factors should be less than or equal to the number given in `xd.inp` (NQ). The key integers are interpreted according to the maximal level of the thermal motion and multipole expansion defined in the atom table for each atom by the parameters tp and lmx , respectively.

Important! The multipole populations of the atoms involved in "chemical constraints" (those which are constrained) may either be fixed or free in the KEY table. On the other hand, parameters involved in any other constraint *must* be made variables (*i.e.* set to "1" in the KEY table). The program will issue a notice and ignore the constraint if this is not the case. An example is :

```
KEY      xyz --U2-- ----U3-----U4----- M- -D- --Q-- --O-- ----H----
O(1)     111 111111 0000000000 0000000000000000 10 110 10011 0110011 100110011
.
KAPPA    110000
EXTCN    0000000
OVTHP    0
SCALE    111
        END KEY -----
```

4.7 Appendix

4.7.1 Treatment of the reflection data by XDLSM

Regardless of the format of Y_{obs} (*i.e.* F_{obs} or F^2_{obs}) supplied in the reflection file XD.HKL, all the values F_{obs} , $\sigma(F_{obs})$, F^2_{obs} and $\sigma(F^2_{obs})$ are calculated. The standard interconversions, given below, are used. The refinement procedure minimises either $w(\Delta F)$ or $w(\Delta F^2)$, depending on the **based_on** record in `xd.mas`, but R -indices for both F and F^2 are always reported. Since a zero or negative value for $\sigma(Y_{obs})$ leads to physically meaningless weights in a least-squares refinement, any record in XD.HKL which has a zero or negative $\sigma(Y_{obs})$ is *ignored* (*i.e.* it is immediately skipped and not counted). A warning message is issued on the console, and such reflections should be removed from XD.HKL

The expressions for deriving the quantities F_{obs} , $\sigma(F_{obs})$, F^2_{obs} and $\sigma(F^2_{obs})$ from each other are :

If F_{obs} and $\sigma(F_{obs})$ are supplied in XD.HKL then

$$F^2_{obs} = (F_{obs})^2 \text{ but } F^2_{obs} \text{ retains the sign of } F_{obs} \text{ (see note 1)}$$

$$\sigma(F^2_{obs}) = 2 \times \sigma(F_{obs}) \times |F_{obs}| \text{ for the case where } |F_{obs}| > 0$$

$$\sigma(F^2_{obs}) = 2 \times \sigma(F_{obs}) \text{ for the case where } F_{obs} = 0 \text{ (see note 2)}$$

If F^2_{obs} and $\sigma(F^2_{obs})$ are supplied in XD.HKL then

$$F_{obs} = \sqrt{|F^2_{obs}|} \text{ but } F_{obs} \text{ retains the sign of } F^2_{obs}$$

$$\sigma(F_{obs}) = \frac{0.5 \times \sigma(F^2_{obs})}{|F_{obs}|} \text{ for the case where } |F_{obs}| > 0 \text{ (see note 2)}$$

$$\sigma(F_{obs}) = 0.5 \times \sigma(F^2_{obs}) \text{ for the case where } F_{obs} = 0$$

These conventions ensure that $\sigma(F_{obs})$ and $\sigma(F^2_{obs})$ are always > 0 , but F_{obs} or F^2_{obs} may be either negative, zero or positive.

4.7.2 Definitions of the Residual indices used in XDLSM

$$R(F) = \frac{\sum |F_{obs} - k_1 F_{calc}|}{\sum |F_{obs}|}$$

$$R(F^2) = \frac{\sum |F^2_{obs} - k_2 F^2_{calc}|}{\sum |F^2_{obs}|}$$

$$wR(F) = \sqrt{\frac{\sum w(F_{obs} - k_1 F_{calc})^2}{\sum w |F^2_{obs}|}}$$

$$wR(F^2) = \sqrt{\frac{\sum w(F^2_{obs} - k_2 F^2_{calc})^2}{\sum w (F^2_{obs})^2}}$$

The scale factors k_1/k_2 place Y_{obs} and Y_{calc} on the same scale. The moduli signs ($| |$) are only used here to indicate the *absolute value* of the appropriate quantity and *not* the commonly used alternative meaning of $|F|$ as the phaseless modulus of a structure factor. This distinction is important in summations where Y_{obs} may have a negative value (see note 3). The indices $R(F)$, $R(F^2)$, $wR(F)$ and $wR(F^2)$, reported in the file XD_LSM.OUT, are computed *only* for those reflections used in the refinement (*i.e.* those passing the rejection criteria on the SKIP directive). The indices $R_{all}(F)$ and $R_{all}(F^2)$ have the same definition as $R(F)$, $R(F^2)$, but also include those reflections rejected by the rejection criteria - this will mean *all* reflections in XD.HKL, except any ignored because of zero or negative $\sigma(Y_{obs})$.

4.7.2 Definitions of the Goodness of Fit indices used in XDLSM

The Goodness of Fit (*GOF*) indices are defined thus :

$$GOF = \sqrt{\frac{\sum (\Delta f_s)^2}{(N_{ref} - N_{var})}} \quad GOF_w = \sqrt{\frac{\sum w(\Delta f)^2}{(N_{ref} - N_{var})}}$$

where N_{ref} is the number of observations and N_{var} the number of independent variables in the least-squares refinement. The definitions of Δf and Δf_s depend of whether refinement was based on F or F^2 , and unlike the R -indices, the *GOF* indices are only reported for the actual refinement mode.

For refinement on F

$$\Delta f = (F_{obs} - k_1 F_{calc}) \quad \text{and} \quad \Delta f_s = (F_{obs} - k_1 F_{calc}) / \sigma(F_{obs})$$

For refinement on F^2

$$\Delta f = (F^2_{obs} - k_2 F^2_{calc}) \quad \text{and} \quad \Delta f_s = (F^2_{obs} - k_2 F^2_{calc}) / \sigma(F^2_{obs})$$

GOF_w uses the reflection weights w as defined on the WEIGHT directive in the XDLSM section of `xd.mas` (see Section 4.6.6). *GOF* is the Goodness of Fit using statistical weights *i.e.* $1/\sigma^2(Y_{obs})$, so the two *GOF* indices will be identical if this weighting scheme is used (*i.e.* if the a parameter on the WEIGHT directive has a value less than -1.0). Note that the default values of zero given for parameters $a - e$ in Section 4.6.6 will also imply statistical weighting if refinement is based on F^2 , but NOT if refinement is based on F !

If the general form of the XD weighting scheme is used, the definitions of the weighted R indices given above have the unfortunate property of being dependent on the overall scale of the data being summed. The weighting scheme used for refinement on F^2 (which is the same scheme as used in SHELXL), in its most commonly used simple form, where only the a and b parameters are non-zero, is given by :

$$w = 1 / (\sigma^2(F^2_{obs}) + a^2 P^2 + bP)$$

where $P = (2F^2_{calc} + F^2_{obs})/3$ for the default value of parameter f .

If the observations used to accumulate the R -indices are on some arbitrary scale compared with the absolute scale, then F^2_{obs} , F^2_{calc} and $\sigma(F^2_{obs})$ are all effectively multiplied by a constant s , so we are dealing with sF^2_{obs} , sF^2_{calc} and $s\sigma(F^2_{obs})$. The expression for $wR(F^2)$ becomes

$$wR(F^2) = \sqrt{\frac{\sum w s^2 (F^2_{obs} - k_2 F^2_{calc})^2}{\sum w s^2 (F^2_{obs})^2}}$$

If unit weights are used, then clearly the wR indices are independent of s . The same pertains if statistical weights are used. In this case $w = 1/s^2 \sigma^2(F^2_{obs})$ and the constant s^2 is eliminated. *However* if the general form of the weighting scheme given above is used, the weight is now given by

$$w = 1 / (s^2 \sigma^2(F^2_{obs}) + s^2 a^2 P^2 + s b P)$$

and a dependence on s remains. Since the current version of XDLSM accumulates the summations on the *observational* scale, while the structure factors in XD.FCO are on an *absolute scale*, it is impossible to correctly compute the weighted R -indices if a general form of the weighting function is used. It would be more logical to accumulate the indices in XDLSM on an absolute scale.

4.7.3 NOTES

Some minor changes in the above definitions in XD for version 6.02 may result in small changes to reported R values, compared with earlier versions.

1. Previously F_{obs}^2 was simply taken as $(F_{obs})^2$, but this has been changed to distinguish between positive and negative F_{obs} . Since the majority of data reduction programs will never report negative F_{obs} , this change is unlikely to have any consequences.
2. For this case $\sigma(F_{obs}^2)$ or $\sigma(F_{obs})$ were previously either undefined, or arbitrarily set to the value of the input $\sigma(Y_{obs})$. For refinements which use any type of sigma cut-off, this change will not have any consequences.
3. Previously, the denominator for $R(F^2)$ was simply $\sum F_{obs}^2$, while that for $R(F)$ was $\sum |F_{obs}|$. For consistency, both summations now use the *absolute* value of the summand, as indicated above. In cases where data sets contain negative F_{obs} , this change may mean small differences in $R_{all}(F^2)$ and possibly in $R(F^2)$, compared with previous versions.

4.8 XDBLOCK - A Shell Program for Blocked Least-Squares Refinement using XDLSM

A common problem in least-squares refinement is high parameter correlations, which may lead to slow convergence or even unstable refinement. One way round this issue is to refine those parameters which are highly correlated in separate blocks, the so-called block refinement. A new utility in XD called XDBLOCK implements this approach.

XDBLOCK is a shell program which launches XDLSM using four specially created master files called `xd_1.mas`, `xd_2.mas` etc. These are a set of edited versions of `xd.mas` where only a selected subset of the parameters are flagged to be refined. If these special master files are not present, then default versions are created. It is more sensible to create these by using the command line `xdblock -create` (or corresponding GUI option in Windows) since this option just creates the files and then halts, allowing them to be examined and possibly edited. The default versions of these files are :

- `xd_1.mas` - where only the positional & thermal parameters are set to refine
- `xd_2.mas` - where only the multipole parameters are set to refine
- `xd_3.mas` - where only the kappa parameters are set to refine
- `xd_4.mas` - where as many as possible of the parameters are set to refine

The blocked refinement is useful in the preliminary stages of refinement and in cases where there are severe correlations between parameters. The user may of course edit these special master files and choose other subsets to be refined in each block. The only limitation at present is that only four such master files can be created and used.

When XDBLOCK is invoked, `xd_1.mas` file is copied to `xd.mas` and a copy of this file is kept. The XDLSM program is launched and at the termination of this program, the `xd.res` file is copied to `xd.inp` and a second round of blocked refinement commences when `xd_2.mas` is copied to `xd.mas`. The whole process is repeated for four cycles.

After each block cycle the stability of the refinement is carefully assessed. All the R values are examined and the refinement is only deemed to be stable (or converging) if the majority of R values are decreasing or remaining steady. If the majority of R values are rising, or if there is a more pathological behaviour (for example if the resultant `xd.res` file becomes

unreadable because of fields of *****) then the refinement is deemed unstable, the `xd.res` file is not copied and the whole process is halted. The results of the refinements are kept in the files `xd_lsm-1.out` etc and so may be examined.

Table 4-1 : Dimensioning of XDLSM

Name	Values	Description
<i>nat</i>	2000	maximum number of atoms in the asymmetric unit
<i>ntx</i>	31	maximum number of displacement tensor components: $6 U_{ij} + 10 C_{ijk} + 15 D_{ijkl} = 31$
<i>lmx</i>	4	maximum level of multipole expansion
<i>nzz</i>	30	maximum number of kappa sets
<i>nto</i>	1	currently not used
<i>nsc</i>	99	maximum number of scale factors
<i>ntb</i>	20	maximum number of core, valence scattering factor tables
<i>nov</i>	2500	maximum number of variables allowed
<i>ncst</i>	200	maximum number of constraints
<i>nao</i>	18	maximum number of atomic orbitals allowed in the wavefunction input for the scattering factors: 1s, 2s, 3s, 4s, 2p, 3p, 4p, 3d, 4d, 4f, 5s, 5p, 6s, 6p, 5d, 7s, 6d, 5f
<i>mgrd</i>	40	maximum number of grids used to store scattering factors
<i>grd</i>	0.05	Step size in $\sin\theta/\lambda$

Related to these the following parameters are also in use:

Name	Value	Description
<i>npop</i>	$lmx * lmx + 2 * lmx + 2$	maximum number of multipole populations
<i>nap</i>	$3 + ntx + npop$	maximum number of atomic parameters
<i>npp</i>	$nap * nat + (lmx + 2) * nzz + nsc + 8$	total number of parameters

Note – some of these dimensions (*nat*, *nzz*, *nov*, *ncst*) are determined dynamically.

Table 4-2 : Variable names and order numbers

Parameter	Symbolic name	Order number
Fractional Coordinates	X, Y, Z	1-3
Displacement Tensor Components		
2nd order U_{ij}	U11, U22, U33, U12, U13, U23	4-9
3rd order C_{ijk}	C111, C222, C333, C112, C122, C113, C133, C223, C233, C123	10-19
4th order D_{ijkl}	D1111, D2222, D3333, D1112, D1222, D1113, D1333, D2223, D2333, D1122, D1133, D2233, D1123, D1223, D1233	19-34
Multipole Populations		
Monopoles	M1, M2	35-36
Dipoles	D1+, D1-, D0	37-39
Quadrupoles	Q0, Q1+, Q1-, Q2+, Q2-	40-44
Octupoles	O0, O1+, O1-, O2+, O2-, O3+, O3-	45-51
Hexadecapoles	H0, H1+, H1-, H2+, H2-, H3+, H3-, H4+, H4-	52-60
Radial Screening		
$\kappa, \kappa'(l)$	KS, K0, K1, K2, K3, K4	61-66
Isotropic and Anisotropic Extinction	EX11, EX22, EX33, EX12, EX13, EX23, RHOEX (for isotropic case ISOEX = EX11)	67-73
Overall U	OVTHP	74
Scale Factor	SCALE(l), I=1, MAXSCAL	75-

IMPORTANT NOTE : Following normal conventions, the magnitudes of the Gram-Charlier coefficients C_{ijk} and D_{ijkl} , which are reported in XD_LSM.OUT and XD_LSM.CIF, are multiplied by 10^3 and 10^4 respectively.

Table 4-3 : The model limits

<i>m1</i> static scattering models	
-4	neutron
-3	core
-2	conventional, spherical-atom promolecule model with RHF scattering factors taken from the International Tables
-1	neutral, spherical-atom model with HF scattering factors generated from Slater-type wavefunctions [14]
lmax	aspherical-atom model: frozen-core, spherical valence, multipolar deformation density up to <i>lmax</i> in the expansion over spherical harmonics [1]
<i>m2</i> thermal motion models	
-1	overall-isotropic-harmonic
0	static
1	isotropic – harmonic
2	anisotropic – harmonic
tmax	anharmonic model: Gram-Charlier expansion up to 4 th order [24]
<i>m3</i> anomalous dispersion	
0	excluded
1	included
<i>m4</i> extinction	
0	excluded
1	included

Table 4-4 : Density normalised Kubic harmonics K_{lj} **(a) as linear combinations of spherical harmonics**

<i>mp</i>					
Even <i>l</i>		0+	2+	4+	6+
<i>l</i>	<i>j</i>				
0	1	1.0			
4	1	0.78245		0.57939	
6	1	0.37790		-0.91682	
6	2		0.83848		-0.5
Odd <i>l</i>		2-	4-	6-	8-
<i>l</i>	<i>j</i>				
3	1	1.0			
7	1	0.73145		0.63290	

(b) site symmetry

<i>l j</i>	23 <i>T</i>	m-3 <i>T_h</i>	432 <i>O</i>	-43m <i>T_d</i>	m-3m <i>O_h</i>
0 1	yes	yes	yes	yes	yes
3 1	yes	no	no	yes	no
4 1	yes	yes	yes	yes	yes
6 1	yes	yes	yes	yes	yes
6 2	yes	yes	no	no	no
7 1	yes	no	no	yes	no

Tables 4-5 – 4-9 which are needed to ascertain the necessary site-symmetry restrictions on the positional, thermal and multipole parameters and which were previously at this point in the manual are now in the Appendix. The new version of XDINI automatically determines these restrictions, so that a manual inspection of tables 4-5 to 4-9 will normally no longer be necessary. They are retained in the XD manual for archival purposes.

Bibliography

1. N.K. Hansen and P. Coppens, *Acta Cryst.* **A34**, 909-921 (1978).
2. (a) F.L. Hirshfeld, *Acta Cryst.* **B27**, 769-781 (1971). (b) F.L. Hirshfeld, *Isr. J. Chem.* **16**, 198-201 (1977).
3. E. Prince and P.T. Boggs, *International Tables for X-ray Crystallography*, Vol. C, pp 594-603, Kluwer Academic Publishers, 1992.
4. W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes*, Cambridge University Press, 1992.
5. P.R. Mallinson, T. Koritsanszky, E. Elkaim, N. Li and P. Coppens, *Acta Cryst.* **A44**, 336-342 (1988).
6. V. Schomaker and K.N. Trueblood, *Acta Cryst.* **B24**, 63-77 (1968).
7. F.L. Hirshfeld, *Acta Cryst.* **A32**, 239-244 (1976).
8. W. Hummel, J. Hauser and H.-B. Bürgi, *J.Mol.Graphics* **8**, 214-220 (1990).
9. W. Hummel, A. Raselli and H.-B. Bürgi, *Acta Cryst.* **B46**, 683-692 (1990).
10. K. Ruedenberg, *J. Chem. Phys.* **19**, 1433-1434 (1951).
11. J. Epstein, J. Bentley and R.F. Stewart, *J.Chem. Phys.* **66**, 5564-5567 (1977).
12. N. Li, P. Coppens and J. Landrum, *Inorg. Chem.* **27**, 482-488 (1988).
13. R.F. Stewart, E.R. Davidson and W.T. Simpson, *J. Chem. Phys.* **42**, 3175 (1965).
14. E. Clementi and C. Roetti, *Atomic Data and Nuclear Data Tables* **14** 177-478 (1974).
15. E. Clementi and D. L. Raimondi, *J. Chem. Phys.* **38**, 2686-2689 (1963)
16. *International Tables for X-ray Crystallography*, Vol. IV, pp 103-144, Kynoch Press: Birmingham, 1974.
17. Z. Su and P. Coppens *Acta Cryst.* **A54**, 646- (1998).
18. P. Macchi and P. Coppens *Acta Cryst.* **A57**, 656- (2001).
19. P.J. Becker and P. Coppens, *Acta Cryst.* **A30**, 129-147 (1974).
20. P.J. Becker and P. Coppens, *Acta Cryst.* **A30**, 148-153 (1974).
21. P.J. Becker and P. Coppens, *Acta Cryst.* **A31**, 417-425 (1975).
22. P. Coppens, and W. C. Hamilton, *Acta Cryst.* **A26**, 71-83 (1970).

23. F. R. Thorney and R. J. Nelves, *Acta Cryst.*, **A30**, 748-757 (1974).
24. C. K. Johnson and H. A. Levy *International Tables for X-ray Crystallography*, Vol. IV, pp 324-336, Kynoch Press: Birmingham, 1974.
25. K. Kurki-Suonio, *Israel J. Chem.* **16**, 115-123 (1977).
26. R. H. Blessing, *Acta Cryst.* **B51**, 816-823 (1995).
27. R.F. Stewart, *Israel J. Chem.* **16**, 124-131 (1977).
28. A. El Haouzi, N. K. Hansen, C. Le Henaff and J. Protas *Acta Cryst.*, **A52**, 291-301 (1996)
29. M. A. Spackman and P. G. Byrom *Acta Cryst.*, **B53**, 553-564 (1997).
30. A. Ø. Madsen, H. S. Osholm. R. F. Stewart, C. Flensburg and S. Larsen *Acta Cryst.*, **A60**, 550-561 (2004).
31. A. Ø. Madsen *J. Appl. Cryst.* **39**, 757-758 (2006).
32. P. Roversi and R. Destro *Chem. Phys. Lett* **386**, 472-478 (2004).
33. A. E. Whitten and M. A. Spackman *Acta Cryst.*, **B62**, 875-888 (2006) and refs therein.
34. J. A. Ibers and W. C. Hamilton *Acta Cryst.*, **17**, 781-782 (1964).
35. (a) A. L. Patterson *Acta Cryst.*, **16**, 1255-1256 (1963). (b) A. C. Larson in *Crystallographic Computing Techniques* (ed A Ahmed), Munkgaard, (1976), pp 396-398.