Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Algorithms for deriving crystallographic space-group information. II. Treatment of special positions

R. W. Grosse-Kunstleve and P. D. Adams

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 27 July 2001
Accepted 8 October 2001

Algorithms for deriving crystallographic space-group information. II. Treatment of special positions

R. W. Grosse-Kunstleve* and P. D. Adams

Lawrence Berkeley National Laboratory, One Cyclotron Road, Mail Stop 4-230, Berkeley, CA 94720, USA. Correspondence e-mail: rwgrosse-kunstleve@lbl.gov

Algorithms for the treatment of special positions in three-dimensional crystallographic space groups are presented. These include an algorithm for the determination of the site-symmetry group given the coordinates of a point, an algorithm for the determination of the exact location of the nearest special position, an algorithm for the assignment of a Wyckoff letter given the site-symmetry group, and an alternative algorithm for the assignment of a Wyckoff letter given the coordinates of a point directly. All algorithms are implemented in ISO C++ and are integrated into the Computational Crystallography Toolbox. The source code is freely available.

1. Introduction

When space-group symmetry is used in the description of crystal structures, the problem of how to treat atomic sites in special positions arises. By definition (Hahn, 1983, Section 2.11), a point in a special position is one that is mapped onto itself by symmetry. This leads to the notion of the *multiplicity* of a point, which is defined as the number of distinct equivalent points per unit cell. If a point is not mapped onto itself by any of the symmetry operations of the given space group, the point is said to be in a *general position*, and the multiplicity is equal to the order of the space group. For a point in a special position, the multiplicity is a factor of the order of the space group. This follows from Lagrange's theorem because the symmetry operations that map a point onto itself form a point group, the *site-symmetry group*, which is a subgroup of the space group.

To determine the multiplicity of a point, the symmetry operations are applied to the coordinates of the point. For each symmetrically equivalent point, the distance to the original point is computed. Since the atomic coordinates are real numbers, these calculations are universally implemented with floating-point arithmetic. To compensate for the rounding errors that are inevitably associated with floating-point arithmetic, a tolerance has to be admitted when determining if a point is mapped onto itself by a given symmetry operation. Necessarily, this tolerance must be very small. Otherwise the results of subsequent calculations, such as structure-factor calculations, would significantly violate the space-group symmetry.

If the determination of the multiplicity of a point is exclusively based on floating-point distance calculations, it is unfortunately impossible to devise a numerically stable algorithm. Owing to the inevitable rounding errors, it may happen that one symmetrically equivalent point is just within the chosen tolerance and another one just outside. Consequently, the multiplicity of the point will not be computed correctly and results of subsequent calculations will be incorrect.

A simple method for ruling out incorrectly determined multiplicities is to define an exclusion radius and to raise an exception if a symmetrically equivalent point is within this radius around the original point but not within the tolerance. This approach does not silently lead to incorrect results but manual intervention is required if a problem is detected.

If a point is close to a special position but outside a meaningful tolerance, in some contexts (*e.g.* Grosse-Kunstleve *et al.*, 1997) it is necessary to determine the exact location of the nearest special position. An obvious approach is to determine this location by averaging all symmetrically equivalent points that are within a given radius around the original point. However, the rounding errors associated with floating-point arithmetic pose the same problem as discussed above: one equivalent point may just be inside and another one just outside the chosen radius. This may lead to an incomplete group of symmetrically equivalent points (corresponding to an incorrectly determined multiplicity), and the average of the coordinates of such an incomplete group will result in incorrect coordinates for the exact location of the nearest special position.

To address these problems, in this paper we present numerically robust algorithms for the determination of the multiplicity of a point. Explicitly (§3) or implicitly (§6), these algorithms involve the determination of the site-symmetry group, which can then be used also for other purposes, such as deriving a point-group symbol or moving a site to the exact location of the nearest special position. We also present two alternative algorithms for the assignment of Wyckoff letters.

One of these algorithms (§6) is probably similar to the algorithm used by Le Page & Raymond (2001) but has the additional advantage of facilitating the determination of the exact location of the nearest special position.

All the algorithms presented in this paper lend themselves to a fully automatic treatment of atomic coordinates and are therefore suitable for integration in highly automated software systems, such as multisolution systems for structure determination (*e.g.* Grosse-Kunstleve *et al.*, 1997; Falcioni & Deem, 1999; Weeks & Miller, 1999; Cerny & Favre-Nicolin, 2000) or the automatic processing and validation of large databases where the need for human intervention is prohibitive.

2. Notation

*IT*83: *International Tables for Crystallography*, Vol. A (Hahn, 1983).

Space-group type: See definition in Section 8.2.1 of *IT*83 (Wondratschek, 1983). Two space groups belong to the same space-group type if they correspond to the same entry in *IT*83.

Space-group representation: A particular group of symmetry operations is a representation of the corresponding space-group type. For example, the space-group representations denoted by the Hermann–Mauguin symbols *Pmna*, *Pnmb*, *Pbmn*, *Pcnm*, *Pncm* and *Pman* all correspond to the same space-group type (space group No. 53 in *IT*83). In general, each space-group type has an infinite number of representations.

Matrices are enclosed by square brackets. However, matrices shown in *Jones–Faithful* notation (for example x, y, z) are not enclosed by brackets. For convenience, column vectors are shown transposed and are enclosed by parentheses to distinguish them from matrices [for example $(0.1, 0.2, 0.3)$].

The $(\boldsymbol{R}, \boldsymbol{T})$ formalism similar to that of *IT*83 is used. $\boldsymbol{R}$ is the (3×3) rotation part or rotation matrix and $\boldsymbol{T}$ the (3×1) translation part or translation vector of the operation.

Sets are enclosed in curly braces (for example $\{-1, 0, 1\}$).

Half-open intervals are denoted as $]x, y]$ (x is not included, y is included) or $[x, y[$ (x is included, y is not included).

A square matrix $\boldsymbol{A}$ is called unimodular if the absolute value of the determinant of $\boldsymbol{A}$ is equal to one [$|\det(\boldsymbol{A})| = 1$].

$\mathbb{Z}$ = set of integer numbers, $\mathbb{R}$ = set of real numbers.

3. Algorithm for the determination of the site-symmetry group

The input parameters of the algorithm for the determination of the site-symmetry group $\boldsymbol{G}_{\boldsymbol{X}}$ of a point $\boldsymbol{X}$ are: (i) unit-cell parameters (for distance calculations); (ii) symmetry operations of the space group $\boldsymbol{G}$; (iii) coordinates $\boldsymbol{X}$ of a point; (iv) minimum distance $\Delta_{\text{equiv\_min}}$ between symmetrically equivalent points of $\boldsymbol{X}$ (tolerance).

The symmetry operations of the space group $\boldsymbol{G}$ are represented by matrices of the form $(\boldsymbol{R}, \boldsymbol{T})$, where $\boldsymbol{R}$ is the rotation part and the translation part $\boldsymbol{T}$ is given modulo $\mathbb{Z}$.

The algorithm for the determination of the site-symmetry group consists of the following steps:

1. Determination of the symmetry operations $\boldsymbol{S} = (\boldsymbol{R}, \boldsymbol{T})$ that have a translation part $\boldsymbol{T}$ without an intrinsic (screw or glide) component and that lead to close contacts between $\boldsymbol{X}$ and the symmetrically equivalent point $\boldsymbol{SX}$. Pairs of $\boldsymbol{S}$ and squared distances $(\boldsymbol{X} - \boldsymbol{SX})^2$ are stored in a list of candidate symmetry operations.

Details: The trial symmetry operations $\boldsymbol{S}$ are generated by a loop over all symmetry operations $\boldsymbol{S_i} \in \boldsymbol{G}$. The symmetrically equivalent point $\boldsymbol{X_S} = \boldsymbol{S_i X}$ is computed, followed by the determination of the difference $\Delta\boldsymbol{X_S} = \boldsymbol{X_S} - \boldsymbol{X}$. The unit translations of the space group are applied to $\Delta\boldsymbol{X_S}$ in the form of the modulus operation to yield a vector $\Delta_{\text{short}}\boldsymbol{X_S}$ with all elements in the range $]-\frac{1}{2}, \frac{1}{2}]$. The symmetry operation that corresponds to the mapping of $\boldsymbol{X}$ to $\Delta_{\text{short}}\boldsymbol{X_S}$ is defined as $\boldsymbol{S}_{\text{short}} = \boldsymbol{S_i X} + \text{nearest\_int}(\boldsymbol{X} - \Delta_{\text{short}}\boldsymbol{X_S})$, where nearest_int($\boldsymbol{Y}$) is a function that determines the vector $\boldsymbol{J} \in \mathbb{Z}^3$ that is closest to $\boldsymbol{Y} \in \mathbb{R}^3$.

In most cases, $\boldsymbol{S}_{\text{short}}$ will correspond to the final $\boldsymbol{S}$, but in the general case it is necessary to loop over 27 additional combinations of unit translations $\boldsymbol{U}$ (this is a 3-deep loop, each over the three integers $\{-1, 0, 1\}$) because the translation part of $\boldsymbol{S}_{\text{short}}$ could have an intrinsic component. The trial symmetry operations $\boldsymbol{S}$ are determined as $\boldsymbol{S} = \boldsymbol{S}_{\text{short}} + \boldsymbol{U}$. The intrinsic part of the trial $\boldsymbol{S}$ is determined with the algorithms in §4 of Grosse-Kunstleve (1999). If the squared distance $(\boldsymbol{X} - \boldsymbol{SX})^2$ is less than $(\Delta_{\text{equiv\_min}})^2$, $\boldsymbol{S}$ is stored in the list of candidate operations for the site-symmetry group.

It should be noted that the determination of the squared distance $(\boldsymbol{X} - \boldsymbol{SX})^2$ involves the unit-cell parameters, for example in the form of the metrical matrix, or alternatively an orthogonalization matrix (see *e.g.* ch. 2 of Giacovazzo, 1992).

2. Sorting of the list of candidate symmetry operations by squared distance, smallest distance first.

3. Generation of the site-symmetry group $\boldsymbol{G_X}$ starting with the symmetry operation that has the smallest distance. The symmetry operations are added to the group in the order in which they appear in the sorted list and group multiplication is carried out. If the group multiplication leads to a symmetry operation with a pure translation, the operation that was added last is discarded.

The group multiplication in step 3 ensures that the algorithm is not negatively affected by rounding errors in the distance calculations. The elimination of operations that lead to pure translations ensures that the site-symmetry group is a point group. If $\Delta_{\text{equiv\_min}}$ is large compared to the unit-cell parameters, it can happen that $\boldsymbol{X}$ is close enough to more than one distinct special position and that the whole set of candidate operations therefore generates a space group.

Given the operations of the site-symmetry group, the point-group type can easily be determined with the algorithm described in §5.2 of Grosse-Kunstleve (1999). This gives immediate access to a point-group symbol for the site-symmetry group. Using the algorithms for the characterization of symmetry operations as described in §4 of Grosse-Kunstleve (1999), it is straightforward to also derive an oriented

site-symmetry symbol from the knowledge of the point-group type and the operations of the site-symmetry group.

4. Algorithm for the determination of the exact location of the nearest special position

The main steps of the algorithm for the determination of the exact location of the nearest special position are:

1. *Determination of the site-symmetry group $\boldsymbol{G_X}$ according to the algorithm in §3.*

2. *Determination of the special-position operator.*

The special-position operator $\boldsymbol{P_X}$ is the average of the symmetry operations $\boldsymbol{S}$ of the site-symmetry group $\boldsymbol{G_X}$:

$$\boldsymbol{P_X} = \frac{1}{O(\boldsymbol{G_X})} \sum_i \boldsymbol{S}_i, \tag{1}$$

where $O(\boldsymbol{G_X})$ is the order of the site-symmetry group and the $\boldsymbol{S_i}$ are the $O(\boldsymbol{G_X})$ symmetry operations.

Example:

space group: $P6$

unit-cell parameters: $a = b = 10$, $c = 13$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$.

Using a tolerance of 0.5 Å, the symmetry operations of the site-symmetry group $\boldsymbol{G_X}$ of $\boldsymbol{X} = (0.35, 0.65, 0.1234)$ are:

$$x, y, z;\ -y+1, x-y+1, z;\ -x+y, -x+1, z.$$

The sum of these operations is $(1, 2, 3z)$, corresponding to the special-position operator $\boldsymbol{P_X} = (1/3, 2/3, z)$.

3. *Determination of the exact location of the nearest special position.*

The exact location $\boldsymbol{X_P}$ of the special position nearest to $\boldsymbol{X}$ is defined by the average of all symmetrically equivalent points of $\boldsymbol{X}$ in the site-symmetry group $\boldsymbol{G_X}$:

$$\boldsymbol{X_P} = \frac{1}{O(\boldsymbol{G_X})} \sum_i \boldsymbol{S}_i \boldsymbol{X}. \tag{2}$$

From this it follows that $\boldsymbol{X_P}$ can be obtained as the product

$$\boldsymbol{X_P} = \boldsymbol{P_X X} \tag{3}$$

Example:

$$\begin{aligned}\boldsymbol{P_X X} &= (1/3, 2/3, z)(0.35, 0.65, 0.1234)\\ &= \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1/3 \\ 2/3 \\ 0 \end{bmatrix} \right) \begin{bmatrix} 0.35 \\ 0.65 \\ 0.1234 \end{bmatrix}\\ &= (1/3, 2/3, 0.1234)\end{aligned}$$

The distance between $\boldsymbol{X}$ and $\boldsymbol{X_P}$ is 0.29 Å.

5. Algorithm for the determination of the Wyckoff letter for a given special-position operator

In general, the assignment of Wyckoff letters is ambiguous. This is discussed in detail by Fischer & Koch (1983*a*). When two assignments of Wyckoff positions are compared, operations of the affine normalizer (Fischer & Koch, 1983*b*) have to be taken into account. The algorithm presented here will produce one possible assignment of Wyckoff letters that is determined by (i) the assignments for a reference setting and (ii) the selection of a 'canonical' change-of-basis matrix for the transformation of symmetry operations from that reference setting to a given space-group representation.

International Tables for Crystallography (Hahn, 1983) (*IT*83) define a Wyckoff position $\boldsymbol{W_G}$ as the set of all points $\boldsymbol{X}$ for which the site-symmetry groups $\boldsymbol{X_G}$ are conjugate subgroups of $\boldsymbol{G}$.

Two subgroups $\boldsymbol{X_G}$ and $\boldsymbol{Y_G}$ of $\boldsymbol{G}$ are conjugate subgroups if there exists a symmetry operation $\boldsymbol{M} \in \boldsymbol{G}$ such that

$$\boldsymbol{Y_G} = \boldsymbol{M}^{-1} \boldsymbol{X_G} \boldsymbol{M}. \tag{4}$$

Equation (1) defines the special-position operator $\boldsymbol{P}$ as the average of the symmetry operations of a site-symmetry group. From this it follows that, if two subgroups of $\boldsymbol{G}$ are conjugate subgroups, the condition

$$\boldsymbol{P_Y} = \boldsymbol{M}^{-1} \boldsymbol{P X M} \tag{5}$$

must hold, where $\boldsymbol{P_X}$ is the average of the operations of $\boldsymbol{X_G}$ and $\boldsymbol{P_Y}$ is the average of the operations of $\boldsymbol{Y_G}$.

To find the Wyckoff letter for a given special-position operator $\boldsymbol{P_X}$, a table is needed for each of the 230 space-group types. These tables correspond to the tables labeled *Positions* in ch. 7 of *IT*83. However, there are two important differences between the tables needed for the algorithm that is described here and the tables in ch. 7 of *IT*83:

(i) Only one *representative* operator per Wyckoff position is needed from the *Coordinates* column of ch. 7 of *IT*83. The other operators are simply products of the symmetry operations and the representative operator. [The idea of representative operators is also found in Altermatt & Brown (1987).]

(ii) The entries in the *Coordinates* columns of ch. 7 of *IT*83 are re-parameterized special-position operators. For example, the first entry for Wyckoff position *f* of space group $P42_12$ (No. 90) is listed as $(x, x, \frac{1}{2})$ (see also Table 1). However, the operations of the site-symmetry group of, *e.g.*, $(0.1, 0.1, \frac{1}{2})$ are

Table 1
Table of representative special-position operators for space group $P42_12$ (No. 90).

The representative special-position operators are the strict averages of the symmetry operations of the corresponding site-symmetry groups. For comparison, the first entry in the *Coordinates* table of *IT*83 is also shown.

Multiplicity, Wyckoff letter	First entry in *IT*83	Representative special-position operator
8 *g*	*x*, *y*, *z*	*x*, *y*, *z*
4 *f*	*x*, *x*, $\frac{1}{2}$	$\frac{1}{2}\,x{+}\frac{1}{2}y$, $\frac{1}{2}x{+}\frac{1}{2}y$, $\frac{1}{2}$
4 *e*	*x*, *x*, 0	$\frac{1}{2}\,x{+}\frac{1}{2}y$, $\frac{1}{2}x{+}\frac{1}{2}y$, 0
4 *d*	0, 0, *z*	0, 0, *z*
2 *c*	0, $\frac{1}{2}$, *z*	0, $\frac{1}{2}$, *z*
2 *b*	0, 0, $\frac{1}{2}$	0, 0, $\frac{1}{2}$
2 *a*	0, 0, 0	0, 0, 0

$$x, y, z; y, x, -z + 1.$$

This leads to the special-position operator $(\frac{1}{2}x + \frac{1}{2}y, \frac{1}{2}x + \frac{1}{2}y, \frac{1}{2})$. Comparison with the entry in *IT*83 shows that x in ch. 7 of *IT*83 corresponds to $\frac{1}{2}x + \frac{1}{2}y$ in the special-position operator.

The parameterization of the special-position operators in ch. 7 of *IT*83 is convenient for a human reader but not in general useful for the purposes of the algorithm presented in this section. Therefore new two-column tables were derived from the tables as found in ch 7 of *IT*83. As an example, the table for space group $P42_12$ (No. 90) is shown in Table 1. The first column of each table lists the Wyckoff letters along with the position multiplicities as listed in ch. 7 of *IT*83. The second column lists the corresponding representative special-position operators $\boldsymbol{P_W}$ that are the strict averages of the symmetry operations of the corresponding site-symmetry groups.

A table of representative special-position operators $\boldsymbol{P_W}$ is provided for a reference setting for each of the 230 space-group types. To obtain the $\boldsymbol{P_W}$ for an arbitrary space-group representation, the $\boldsymbol{P_W}$ are transformed using a change-of-basis matrix that transforms the given symmetry operations to the setting for which the reference table was computed. This transformation property follows immediately from the fact that the $\boldsymbol{P_W}$ are averages of symmetry operations. The change-of-basis matrices are determined with the algorithm for the determination of the space-group type that is described in §5 of Grosse-Kunstleve (1999) and then combined with the operations of the affine normalizer. Each product of a given change-of-basis matrix and an operation of the affine normalizer is an alternative change-of-basis matrix. Since the assignment of Wyckoff letters is in general ambiguous and depends on the exact choice of the change-of-basis matrix, a 'canonical' change-of-basis matrix is selected. The selection is based on a set of rules that ensure that the selected matrix is independent of the order in which the alternative matrices are generated. This ensures a reproducible assignment of Wyckoff letters for any given space-group representation.

To determine the Wyckoff letter for a given special-position operator $\boldsymbol{P_X}$, and given a table of representative $\boldsymbol{P_W}$ for the given space-group representation $\boldsymbol{G}$, the $\boldsymbol{M} \in \boldsymbol{G}$ has to be found that solves the equation

$$\boldsymbol{P_X} = \mathbf{M}^{-1}\boldsymbol{P_W}\boldsymbol{M}. \tag{6}$$

Since both the multiplicity of $\boldsymbol{X}$ and the multiplicity corresponding to each $\boldsymbol{P_W}$ are known, comparison of these is used as a pre-selection.

For the remaining $\boldsymbol{P_W}$ in the table, the solution of (6) is implemented as a trial-and-error algorithm with a predictably finite number of trials that is less than or equal to the order of $\boldsymbol{G}$, per candidate $\boldsymbol{P_W}$. Trial matrices for $\boldsymbol{M}$ are derived from the symmetry operations of $\boldsymbol{G}$. Let $\boldsymbol{R}$ be the rotation part of a symmetry operation of $\boldsymbol{G}$ and $\boldsymbol{T}$ be the corresponding translation part modulo 1. A trial matrix $\boldsymbol{M}$ is defined as $\boldsymbol{M} = (\boldsymbol{R}, \mathbf{T} + \boldsymbol{U})$, where $\boldsymbol{U}$ is an unknown integer vector of unit translations. Rearrangement of (6) and substitution of $\boldsymbol{M} = (\boldsymbol{R}, \boldsymbol{T} + \boldsymbol{U})$, $\boldsymbol{P_X} = (\boldsymbol{R_X}, \boldsymbol{T_X})$ and $\boldsymbol{P_W} = (\boldsymbol{R_W}, \boldsymbol{T_W})$ leads to:

$$\boldsymbol{M}\boldsymbol{P_X} = \boldsymbol{P_W}\boldsymbol{M} \Leftrightarrow (\boldsymbol{R}, \boldsymbol{T} + \boldsymbol{U})(\boldsymbol{R_X}, \boldsymbol{T_X}) = (\boldsymbol{R_W}, \boldsymbol{T_W})(\boldsymbol{R}, \boldsymbol{T} + \boldsymbol{U}). \tag{7}$$

$\boldsymbol{U}$ is the only unknown in this equation. Elementary rearrangement leads to

$$(\boldsymbol{R}\boldsymbol{R_X}, \boldsymbol{R}\boldsymbol{T_X} + \boldsymbol{T} + \boldsymbol{U}) = (\boldsymbol{R_W}\boldsymbol{R}, \boldsymbol{R_W}(\boldsymbol{T} + \boldsymbol{U}) + \boldsymbol{T_W}). \tag{8}$$

The condition

$$\boldsymbol{R}\boldsymbol{R_X} = \boldsymbol{R_W}\boldsymbol{R} \tag{9}$$

is used as a second pre-selection for the candidate $\boldsymbol{P_W}$ from the table of Wyckoff letters. For the remaining $\boldsymbol{P_W}$, the unknown $\boldsymbol{U}$ is determined as the solution of

$$\begin{aligned}\boldsymbol{R}\boldsymbol{T_X} + \boldsymbol{T} + \boldsymbol{U} &= \boldsymbol{R_W}(\boldsymbol{T} + \boldsymbol{U}) + \boldsymbol{T_W} \Leftrightarrow (\boldsymbol{R_W} - \boldsymbol{I})\boldsymbol{U} \\ &= \boldsymbol{R}\boldsymbol{T_X} + \boldsymbol{T} - \boldsymbol{R_W}\boldsymbol{T} - \boldsymbol{T_W},\end{aligned} \tag{10}$$

where $\boldsymbol{I}$ is the (3×3) unit matrix. Equation (10) is of the general form

$$\boldsymbol{M}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \in \mathbb{Z}. \tag{11}$$

Comparison with (10) leads to $\boldsymbol{M} = (\boldsymbol{R_W} - \boldsymbol{I})$, $\boldsymbol{x} = \boldsymbol{U}$, $\boldsymbol{b} = \boldsymbol{R}\boldsymbol{T_X} + \boldsymbol{T} - \boldsymbol{R_W}\boldsymbol{T} - \boldsymbol{T_W}$. A procedure for solving (11) is given by, for example, Domenjoud (1991). Let r be the rank of $\boldsymbol{M}$. There exist then two unimodular matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that:

$$\boldsymbol{P}\boldsymbol{M}\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & 0 \end{bmatrix}. \tag{12}$$

$\boldsymbol{D}$ is an $(r \times r)$ diagonal matrix and the 0's stand for null matrices of suitable sizes.

The determination of $\boldsymbol{P}$ and $\boldsymbol{Q}$ amounts to computing the *Smith normal form*. A procedure for this is given, for example, in §5.5 of Grosse-Kunstleve (1999).

Any solution of $\boldsymbol{x}$ is then of the form $\boldsymbol{Q}\boldsymbol{x}'$, where $\boldsymbol{x}'$ is a solution of

$$\begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{x}' = \boldsymbol{P}\boldsymbol{b}. \tag{13}$$

Checking the satisfiability of equation (13) and computing the solution is straightforward. If equation (13) has a solution for a given $\boldsymbol{P_W}$, both the Wyckoff letter and the matrix $\boldsymbol{M} = (\boldsymbol{R}, \mathbf{T} + \boldsymbol{U})$ are determined simultaneously.

An $\boldsymbol{M}$ that solves (4) is necessarily a solution of (5) but it is not immediately clear that the converse is also true. A rigorous proof for this is currently not available. However, we were able to show empirically that (5) combined with the condition that the multiplicity of the site-symmetry group of $\boldsymbol{X}$ and the multiplicity of the Wyckoff position must be equal can be used as a necessary and sufficient condition. This was performed by comparing the Wyckoff letters assigned with the algorithm in this section with the letters assigned by the alternative algorithm that is presented in the next section. The two algorithms produce consistent results for a very large number of test cases that include all conventionally used space-group representations and an additional large number of unconventional settings.

6. Algorithm for the determination of the Wyckoff letter given the coordinates of a point

The input parameters for the algorithm for the determination of the Wyckoff letter given the coordinates of a point are: (i) unit-cell parameters (for distance calculations); (ii) symmetry operations of the space group $\boldsymbol{G}$; (iii) coordinates $\boldsymbol{X}$ of a point; (iv) minimum distance $\Delta_{\rm near}$ to the nearest special position of $\boldsymbol{X}$ (tolerance).

The algorithm consists of the following steps:

1. Construction of a table of representative special-position operators $\boldsymbol{P_W}$ as explained in §5.

As before, the $\boldsymbol{P_W}$ are strict averages of the operations of a site-symmetry group. To facilitate the algorithm presented in this section, the locations $\boldsymbol{X_W}$ to which the $\boldsymbol{P_W}$ apply are chosen such that all elements of the vector $\boldsymbol{X_W}$ are in the range $]-\frac{1}{2},\frac{1}{2}]$.

2. Application of the unit translations to $\boldsymbol{X}$:

$$\boldsymbol{X_0} = \text{mod\_short}(\boldsymbol{X}). \tag{14}$$

mod_short is a function that applies the unit translations to the elements of $\boldsymbol{X}$ such that the elements are in the range $]-\frac{1}{2},\frac{1}{2}]$. The unit shifts applied are defined as $\boldsymbol{U_0} = \boldsymbol{X_0} - \boldsymbol{X}$.

3. For each $\boldsymbol{P_W}$, starting with Wyckoff position *a* and working upwards in the alphabet, the shortest distance squared

$$\Delta\boldsymbol{P_W} = (\boldsymbol{P_W M'} \boldsymbol{X_0} - \boldsymbol{M'} \boldsymbol{X_0})^2, \tag{15}$$

with $\boldsymbol{M'} = (\boldsymbol{R}, \boldsymbol{T} + \boldsymbol{U'})$, is computed in a loop over all symmetry operations $(\boldsymbol{R}, \boldsymbol{T}) \in \boldsymbol{G}$ and unit shifts $\boldsymbol{U'}$ with vector elements in the set of integers $\{-1, 0, 1\}$ (*i.e.* the number of loop iterations is 27 times the order of $\boldsymbol{G}$, per $\boldsymbol{P_W}$). The symmetry operations of $\boldsymbol{G}$ are manipulated such that the elements of $\boldsymbol{T}$ are in the range $]-\frac{1}{2},\frac{1}{2}]$.

4. The Wyckoff letter for $\boldsymbol{X}$ is determined by the first $\boldsymbol{P_W}$ for which the smallest $\Delta\boldsymbol{P_W} < \Delta_{\rm near}$. The exact location $\boldsymbol{X_P}$ of the special position nearest to $\boldsymbol{X}$ is determined as

$$\boldsymbol{X_P} = \boldsymbol{M}^{-1} \boldsymbol{P_W M X}. \tag{16}$$

$\boldsymbol{M}$ is defined as $\boldsymbol{M} = \boldsymbol{M'} + \boldsymbol{R U_0} = (\boldsymbol{R}, \boldsymbol{T} + \boldsymbol{U'} + \boldsymbol{R U_0})$, with the $\boldsymbol{M'}$ that corresponds to the smallest $\Delta\boldsymbol{P_W} < \Delta_{\rm near}$.

In this algorithm, the problem of determining $\boldsymbol{U}$ is solved in two steps, instead of one as in the algorithm of §5. The first step is to bring $\boldsymbol{X}$ within ± 1 unit translations of the space where the $\boldsymbol{P_W}$ are known to apply [equation (14)]. The remaining unit shifts $\boldsymbol{U'}$ are determined with a trial-and-error approach.

7. Discussion

The central algorithm presented in this paper is that for the determination of the site-symmetry group (§3). This algorithm determines the multiplicity of a point in a numerically robust way, and at the same time provides the information from which the special-position operator can easily be computed (§4). This special-position operator can then be used to move a site to the exact location of the nearest special position.

In §5, we show that the Wyckoff positions as listed in *IT*83 are re-parameterized special-position operators. We have derived tables of *representative special-position operators* for the Wyckoff positions of 230 reference space-group representations (one for each of the 230 space-group types) where the operations listed in *IT*83 are replaced by the corresponding special-position operators. This table facilitates the determination of a Wyckoff letter for a given point. The Wyckoff letter can then be used in turn to obtain the special-position operator $\boldsymbol{P_W}$ from the table. To explain how the special-position operator $\boldsymbol{P_W}$ that is produced by either of the alternative algorithms in §§5 or 6 can be used, we will analyze the combination of (3) and (6), which is identical to (16):

$$\boldsymbol{X_P} = \boldsymbol{M}^{-1} \boldsymbol{P_W M X}. \tag{17}$$

$\boldsymbol{MX}$ is a symmetrically equivalent point of $\boldsymbol{X}$ for which $\boldsymbol{P_W}$ is a special-position operator. $\boldsymbol{M}^{-1}$ is another symmetry operation of $\boldsymbol{G}$ that maps the exact location of the special position as defined by $\boldsymbol{P_W}$ back to the special position that is nearest to $\boldsymbol{X}$. In applications where it is not important which symmetrically equivalent point of $\boldsymbol{X}$ appears in a symmetry-unique set of sites, $\boldsymbol{X}$ in this set could be replaced by $\boldsymbol{MX}$ and the Wyckoff letter for this site could be stored. In computing intensive calculations, such as structure-factor calculations, the Wyckoff letter could then be used as a reference into a table of lists of non-redundant symmetry operations for each Wyckoff position. This leads to implementations that are both robust and fast.

The alternative algorithms for assigning Wyckoff letters and computing special-position operators are largely interchangeable. The question of which one is best selected has to be answered for each application individually. In most cases, it is expected that the combination of the algorithms in §§3 and 5 is slightly faster than the algorithm in §6, especially if the number of sites in general positions is large. If the operations of the site-symmetry group are needed, for example to derive a symbol for the point-group type, clearly the algorithm in §3 has to be used. In such a case, the special-position operator is already known and it is always best to use the algorithm in §5 for the assignment of Wyckoff letters.

The algorithms presented in this paper are building blocks that are intended for integration into a highly automated software system where a fully automatic and robust treatment of atomic coordinates is essential. The choice of a modern programming language (ISO C++, International Standardization Organization *et al.*, 1998) that supports a modular system design through object orientation and exception handling ensures both ease of use and a high degree of flexibility.

8. Source-code availability

All algorithms that we present in this paper are implemented in ISO C++ and are integrated into the Computational Crystallography Toolbox (cctbx). The cctbx source code is freely available under an Open Source license for both non-profit and commercial use at http://cctbx.sourceforge.net/. The tables

of representative special-position operators $\boldsymbol{P}_{\boldsymbol{W}}$ (§5) are included in the source code. They may also be accessed *via* an end-user friendly web interface at http://cci.lbl.gov/servers/ ('Explore symmetry' example).

We would like to thank Y. LePage and E. Kroumova for providing electronic copies of the Wyckoff tables as published in *IT*83. Our work was funded in part by the US Department of Energy under contract No. DE-AC03-76SF00098 and NIH/NIGMS under grant No. 1P50GM62412.

References

Altermatt, U. D. & Brown, I. D. (1987). *Acta Cryst.* A**43**, 125–130.

Cerny, R. & Favre-Nicolin, V. (2000). http://crystsun1.unige.ch/cerny/Project_structures_from_powders.htm.

Domenjoud, E. (1991). Proceedings 16th International Symposium on Mathematical Foundations of Computer Science, Kazimierz Dolny, Poland, edited by A. Tarlecki. Vol. 520 of *Lecture Notes in Computer Science*, pp. 141–150. Berlin: Springer-Verlag. http://citeseer.nj.nec.com/domenjoud91solving.html.

Falcioni, M. & Deem, M. W. (1999). *J. Chem. Phys.* **110**, 1754–1766.

Fischer, W. & Koch, E. (1983*a*). *International Tables for Crystallography*, Vol. A, ch. 14. Dordrecht: Kluwer.

Fischer, W. & Koch, E. (1983*b*). *International Tables for Crystallography*, Vol. A, ch. 15. Dordrecht: Kluwer.

Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*. International Union of Crystallography/Oxford University Press.

Grosse-Kunstleve, R. W. (1999). *Acta Cryst.* A**55**, 383–395.

Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, Ch. (1997). *J. Appl. Cryst.* **30**, 985–995.

Hahn, Th. (1983). Editor. *International Tables for Crystallography*, Vol. A. Dordrecht: Kluwer.

International Standardization Organization (ISO), International Electrotechnical Commission (IEC), American National Standards Institute (ANSI) & Information Technology Industry Council (ITI) (1998). International Standard ISO/IEC 14882, 1st ed., Information Technology Industry Council, 1250 Eye Street NW, Washington, DC 20005, USA (also available at http://webstore.ansi.org/).

Le Page, Y. & Raymond, S. (2001). *ALTWYK* program, http://ylp.icpet.nrc.ca/altwyk/default.htm.

Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.

Wondratschek, H. (1983). *International Tables for Crystallography*, Vol. A, ch. 8. Dordrecht: Kluwer.