

## 1. Introduction

SYMSEARCH is a computer program which finds the space-group symmetry of a crystal, given the dimensions of a unit cell and the positions of atoms inside that unit cell. SYMSEARCH does not require the unit cell to be primitive. It finds the primitive unit cell.

SYMSEARCH was specifically written for use with molecular dynamics simulations, where we often want to find the space-group symmetry of the crystal. Here, however, the data is often “noisy”. The atomic positions are not known exactly, and sometimes even the dimensions of the unit cell are not known exactly. The space-group symmetry is only approximate. In order for the crystal to exhibit a given space-group symmetry exactly, we may need to move some of the atoms a little and even introduce some long-range strain. SYMSEARCH does this automatically, looking for space-group symmetries which meet some predetermined criteria of how well the data should fit the symmetry.

## 2. Conventional unit cell

The conventional unit cell is chosen so that axes of rotation are parallel to its edges, and planes of reflection are parallel to its faces. The lattice which defines this cell is called the conventional lattice and is not necessarily primitive.

Let  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  be the basis vectors of the conventional lattice. These vectors are usually defined by six lattice parameters:  $a, b, c, \alpha, \beta, \gamma$ , where  $a, b, c$  are the lengths of  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , respectively, and  $\alpha$  is the angle between  $\mathbf{a}_2$  and  $\mathbf{a}_3$ ,  $\beta$  is the angle between  $\mathbf{a}_1$  and  $\mathbf{a}_3$ , and  $\gamma$  is the angle between  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . We can define a cartesian coordinate system such that

$$\begin{aligned}\mathbf{a}_1 &= a \hat{\mathbf{i}}, \\ \mathbf{a}_2 &= b \cos \gamma \hat{\mathbf{i}} + b \sin \gamma \hat{\mathbf{j}}, \\ \mathbf{a}_3 &= c \cos \beta \hat{\mathbf{i}} + c \frac{\cos \alpha - \cos \gamma \cos \beta}{\sin \gamma} \hat{\mathbf{j}} + c \left[ \sin^2 \beta - \left( \frac{\cos \alpha - \cos \gamma \cos \beta}{\sin \gamma} \right)^2 \right]^{1/2} \hat{\mathbf{k}},\end{aligned}\tag{1}$$

chosen so that

$$\begin{aligned}|\mathbf{a}_1| &= a, \\ |\mathbf{a}_2| &= b, \\ |\mathbf{a}_3| &= c, \\ \mathbf{a}_2 \cdot \mathbf{a}_3 &= bc \cos \alpha, \\ \mathbf{a}_1 \cdot \mathbf{a}_3 &= ac \cos \beta, \\ \mathbf{a}_1 \cdot \mathbf{a}_2 &= ab \cos \gamma.\end{aligned}\tag{2}$$

We put the cartesian components of  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  into a matrix,

$$\mathbf{a}_i = \sum_{\mu} \hat{\mathbf{e}}_{\mu} T_{CX\mu i}, \quad (3)$$

where  $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$  are  $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ , respectively, the unit vectors along the three cartesian coordinate axes. As a convention, we will use greek subscripts  $\mu, \nu$  for cartesian components of vectors.

An atomic position is defined by a vector  $\mathbf{r}$  from the origin. This vector can be written both in terms of the cartesian coordinate system and in terms of the basis vectors of the conventional lattice:

$$\mathbf{r} = \sum_{\mu} x_{\mu} \hat{\mathbf{e}}_{\mu} = \sum_i x'_i \mathbf{a}_i. \quad (4)$$

Using Eq. (3), we have

$$\sum_i x'_i \mathbf{a}_i = \sum_{i,\mu} x'_i \hat{\mathbf{e}}_{\mu} T_{CX\mu i} = \sum_{\mu} \hat{\mathbf{e}}_{\mu} \sum_i T_{CX\mu i} x'_i, \quad (5)$$

so that we obtain

$$x_{\mu} = \sum_i T_{CX\mu i} x'_i. \quad (6)$$

$T_{CX}$  is a transformation matrix that takes vector components in terms of the basis vectors of the conventional lattice to cartesian components, i.e.,  $\mathbf{x} = T_{CX} \mathbf{x}'$ . Conversely, we also have  $\mathbf{x}' = T_{XC} \mathbf{x}$  where  $T_{XC} = T_{CX}^{-1}$ .

### 3. Unit cell

We define a unit cell of the crystal which may be smaller than the conventional unit cell but still may not be primitive. We give the position  $\mathbf{r}_i$  of every atom in the unit cell. The lattice which defines the unit cell is determined by three basis vectors which we call the basis vectors of the unit cell. We write the basis vectors of the unit cell in terms of the basis vectors of the conventional lattice:

$$\mathbf{t}_i = \sum_j \mathbf{a}_j T_{PCji}. \quad (7)$$

$T_{PC}$  is the transformation matrix which takes vector components in terms of the basis vectors of the unit cell to components in terms of the basis vectors of the conventional lattice. Also,  $T_{PX} = T_{CX} T_{PC}$  is the transformation matrix which takes vector components in terms of the basis vectors of the unit cell to cartesian components. Note that since the conventional lattice points are a subset of the unit cell lattice points, every element of the matrix  $T_{CP} = T_{PC}^{-1}$  is an integer.

#### 4. Primitive lattice

Now, suppose that the lattice defined by  $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$  is not primitive. If this is the case, then there must be one or more lattice vectors which cannot be written as an integer combination of  $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ . If such a lattice vector  $\mathbf{v}$  exists, then for each atom at  $\mathbf{r}_i$ , there must exist an atom of the same type at  $\mathbf{r}_i + \mathbf{v}$ . We can find these additional lattice vectors using the following method. We choose any atom  $i$ . We then form vectors from this atom to every other atom of the same type, i.e.  $\mathbf{v} = \mathbf{r}_j - \mathbf{r}_i$  for all  $j \neq i$  where atoms  $i$  and  $j$  are of the same type. We test each of these vectors  $\mathbf{v}$  to determine whether or not it is a lattice vector by adding  $\mathbf{v}$  to the position  $\mathbf{r}_k$  of each atom  $k$  in the unit cell and then looking to see if there is an atom of the same type at  $\mathbf{r}_k + \mathbf{v}$ .

We must also take noise into account. The position of each atom may only be approximate. This means that  $\mathbf{v}$  is only approximate, and, even if it were exact, the position of an atom relative to  $\mathbf{r}_k + \mathbf{v}$  is also only approximate. We therefore look for the atom *nearest* to the position of  $\mathbf{r}_k + \mathbf{v}$ . This means that we try every atom  $m$  in the unit cell which is the same type as atom  $k$  and choose the atom for which the value of

$$d = |[\mathbf{r}_k + \mathbf{v} - \mathbf{r}_m]_{\min}| \quad (8)$$

is the smallest. The function  $[\mathbf{r}]_{\min}$  is obtained by adding a lattice vector to  $\mathbf{r}$  so that the result is as close to the origin as possible. This is done by first obtaining the coordinates of  $\mathbf{r}$  in terms of the basis vectors of the unit cell, then adding an integer to each coordinate so that the result is between  $\pm 0.5$ , and finally obtaining the cartesian coordinates of the resulting vector. We must perform this operation since  $\mathbf{r}_k + \mathbf{v}$  may actually be near atom  $m$  in another unit cell. The lattice vector

$$\mathbf{u}_k = \mathbf{r}_k + \mathbf{v} - \mathbf{r}_m - [\mathbf{r}_k + \mathbf{v} - \mathbf{r}_m]_{\min} \quad (9)$$

points to the unit cell containing atom  $m$  near  $\mathbf{r}_k + \mathbf{v}$  so that

$$\mathbf{r}_k + \mathbf{v} \approx \mathbf{r}_m + \mathbf{u}_k. \quad (10)$$

We also define a mapping function  $M(k) = m$ , i.e., the action of the translation  $\mathbf{v}$  maps atom  $k$  onto atom  $m$ . We test the properties of the mapping functions. If we examine the series,  $M(k), M^2(k), M^3(k), \dots$ , where  $M^2(k) = M(M(k))$ ,  $M^3(k) = M(M^2(k))$ , etc., we find that  $M^n(k) = k$  for some minimum value of  $n$ . This defines a period of a cycle. All cycles among the atoms must have the same period. Also, the inverse mapping function  $M^{-1}(k)$  must be single-valued.

The atoms  $k, M(k), M^2(k), \dots, M^{n-1}(k)$  are all connected by the new lattice vector  $\mathbf{v}$ . Since after  $n$  applications of  $\mathbf{v}$  we return to the original atom  $k$  (but in a different unit cell),  $n\mathbf{v}$  must be a vector of the original primitive lattice. We can now find the exact value of  $\mathbf{v}$ :

$$\mathbf{v} = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{u}_{M^i(k)}. \quad (11)$$

Now let us adjust the atomic positions so that  $\mathbf{r}_k + \mathbf{v} = \mathbf{r}_{M(k)} + \mathbf{u}_k$  exactly. The atoms can be divided into cycles of  $n$  atoms so that the atoms within a single cycle are connected by the new lattice vector  $\mathbf{v}$ . Only a single cycle of atoms need to be considered at a time. To simplify the notation, let us rearrange the numbering of the atoms so that the atoms in the cycle under consideration are numbered 1 through  $n$  such that

$$M(k) = \begin{cases} k+1, & \text{if } k < n, \\ 1, & \text{if } k = n. \end{cases} \quad (12)$$

Let  $\Delta\mathbf{r}_k$  be the adjustment to  $\mathbf{r}_k$  so that

$$\mathbf{r}_k + \Delta\mathbf{r}_k + \mathbf{v} = \mathbf{r}_{k+1} + \Delta\mathbf{r}_{k+1} + \mathbf{u}_k. \quad (13)$$

We thus have

$$\begin{aligned} \Delta\mathbf{r}_2 &= \Delta\mathbf{r}_1 + \mathbf{r}_1 + \mathbf{v} - \mathbf{r}_2 - \mathbf{u}_1, \\ \Delta\mathbf{r}_3 &= \Delta\mathbf{r}_2 + \mathbf{r}_2 + \mathbf{v} - \mathbf{r}_3 - \mathbf{u}_2 \\ &= \Delta\mathbf{r}_1 + \mathbf{r}_1 + 2\mathbf{v} - \mathbf{r}_3 - \mathbf{u}_1 - \mathbf{u}_2. \\ \Delta\mathbf{r}_k &= \Delta\mathbf{r}_1 + \mathbf{r}_1 + (k-1)\mathbf{v} - \mathbf{r}_k - \sum_{i=1}^{k-1} \mathbf{u}_i. \end{aligned} \quad (14)$$

We note that  $\Delta\mathbf{r}_1$  is an independent variable in the above equations. Its value is arbitrary. We choose its value so that the rms average of the adjustments to the atomic positions is a minimum. In other words, we move the atoms as little as possible. The rms average is given by

$$\Delta r_{\text{rms}} = \frac{1}{n} \left( \sum_{i\mu} \Delta r_{i\mu}^2 \right)^{1/2}. \quad (15)$$

Setting  $d\Delta r_{\text{rms}}/d\Delta r_{1\mu} = 0$ , we obtain

$$\Delta r_{1\mu} + \sum_{k=2}^n \left( \Delta r_{1\mu} + r_{1\mu} + (k-1)t_\mu - r_{k\mu} - \sum_{i=1}^{k-1} u_{i\mu} \right) = 0 \quad (16)$$

for  $\mu = 1, 2, 3$ . Rearranging terms, we obtain the adjusted position of atom 1,

$$\mathbf{r}_1 + \Delta\mathbf{r}_1 = \frac{1}{n} \sum_{k=1}^n \left( \mathbf{r}_k - (k-1)\mathbf{v} + \sum_{i=1}^{k-1} \mathbf{u}_i \right), \quad (17)$$

which is simply the average position of each atom brought back to the position of atom 1.

Once we have the adjusted position of atom 1, we find the adjusted position of each of the other atoms from Eqs. (14). Then we calculate  $\Delta r_{\text{rms}}$  from Eq. (15). If its value is less than some predetermined maximum allowed value, we accept  $\mathbf{v}$  as a member of the primitive lattice. We repeat the above test for every possible candidate  $\mathbf{v} = \mathbf{r}_j - \mathbf{r}_i$ , where, as stated above,  $\mathbf{r}_i$  is the position of some fixed atom  $i$ , and  $\mathbf{r}_j$  is the position of every other atom in the unit cell of the same type as atom  $i$ . To make this search as brief

as possible, we choose atom  $i$  to be the least abundant type of atom. When we are finished with this search, we have a set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  which represent  $n$  points of the primitive lattice. (For completeness, we add  $\mathbf{v}_1 = 0$  to the list.)

The set of vectors  $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  (the basis vectors of the unit cell plus each of the new lattice vectors found) generate the complete set of primitive lattice vectors of the crystal. We find the basis vectors  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  of the primitive lattice by trying any three vectors from the generating set and then testing them for the following property. We must be able to write each vector from the generating set as an integer combination of the basis vectors. Let us define a matrix  $T_{PX}$  by

$$\mathbf{p}_i = \sum_j \hat{\mathbf{e}}_j T_{PXji} \quad (18)$$

and its inverse,  $T_{XP} = T_{PX}^{-1}$ . Then we must require that every element of  $T_{XP}\mathbf{t}_i$  and  $T_{XP}\mathbf{v}_i$  be an integer. (Here, the notation  $\mathbf{t}_i$  and  $\mathbf{v}_i$  means a column matrix containing the cartesian components of the vector.)

The atoms in the unit cell can be grouped into sets of  $n$  atoms. Let  $M_j(i)$  be the atom onto which  $\mathbf{v}_j$  maps atom  $i$ , i.e.,

$$\mathbf{r}_i + \mathbf{v}_j \approx \mathbf{r}_{M_j(i)} \text{ modulo } \mathbf{t}. \quad (19)$$

Then atoms  $i, M_2(i), M_3(i), \dots, M_n(i)$  form a set of  $n$  atoms closed with respect to these mapping functions. If atom  $k$  is a member of the set, then atom  $M_j(k)$  is also a member of the set. Let us adjust the atomic positions in each set so that

$$\mathbf{r}_i + \Delta\mathbf{r}_i + \mathbf{v}_j = \mathbf{r}_{M_j(i)} + \Delta\mathbf{r}_{M_j(i)} \text{ modulo } \mathbf{t} \quad (20)$$

exactly for every atom  $i$  in the set and for every primitive lattice vector  $\mathbf{v}_j$ . To simplify the notation, let us rearrange the numbering of the atoms so that the atoms in the set under consideration are numbered 1 through  $n$  such that  $M_j(1) = j$ . Then the condition becomes

$$\mathbf{r}_1 + \Delta\mathbf{r}_1 + \mathbf{v}_j = \mathbf{r}_j + \Delta\mathbf{r}_j \text{ modulo } \mathbf{t} \quad (21)$$

or

$$\Delta\mathbf{r}_j = \Delta\mathbf{r}_1 + [\mathbf{r}_1 + \mathbf{v}_j - \mathbf{r}_j]_{\min}, \quad (22)$$

where the function  $[\mathbf{r}]_{\min}$  minimizes  $\mathbf{r}$  with respect to vectors of the primitive lattice. This gives us  $n - 1$  equations ( $j = 2, 3, \dots, n$ ). Similar to Eqs. (14), we see that  $\Delta\mathbf{r}_1$  is an independent variable in the above equations. We choose its value so that the rms average of the adjustments to the atomic positions is a minimum. We obtain

$$\Delta\mathbf{r}_1 = -\frac{1}{n} \sum_{j=2}^n [\mathbf{r}_1 + \mathbf{v}_j - \mathbf{r}_j]_{\min}. \quad (23)$$

We repeat this for every set of  $n$  atoms in the unit cell.

When finished, we have a set of atomic positions which exactly exhibit the translational symmetry required by vectors of the primitive lattice. Each set of  $n$  atoms can be mapped onto a single atom in the primitive unit cell. The  $N$  atoms in the unit cell are mapped onto  $N/n$  atoms in the primitive unit cell. To simplify notation, we will now let  $N$  be the number of atoms in the primitive unit cell, and  $\mathbf{r}_i$  be the adjusted position of one of the atoms in the  $i$ th set.

## 5. Space group elements

Each element of the space group consists of a point operation  $R$  followed by a translation  $\mathbf{f}$ , usually denoted by  $\{R|\mathbf{f}\}$ . The translation  $\mathbf{f}$  is often called the “fractional.” The point operation is represented by a matrix such that when  $R$  operates on a vector  $\mathbf{r} = r_1\hat{\mathbf{e}}_1 + r_2\hat{\mathbf{e}}_2 + r_3\hat{\mathbf{e}}_3$ , the result is  $\mathbf{r}' = r'_1\hat{\mathbf{e}}_1 + r'_2\hat{\mathbf{e}}_2 + r'_3\hat{\mathbf{e}}_3$ , where

$$r'_\mu = \sum_\nu R_{\mu\nu} r_\nu. \quad (24)$$

There are 48 possible point operations in a cubic crystal and 24 in a hexagonal crystal. Since 8 of these are in common (eg., identity operation, 2-fold rotation about the  $z$  axis, etc.), the total number of different point operations is 64. Note that the  $x, y, z$  axes are defined by the basis vectors of the conventional unit cell. We will only try the 64 point operations defined with respect to these axes. This means that the conventional unit cell must be chosen carefully so that all possible symmetry elements of the crystal will be found.

We test each of the 64 point operations, first with respect to the primitive lattice. Any point operation which is a part of a space-group element must bring the primitive lattice back into itself again. These means that  $R\mathbf{p}_i$  must be an integer combination of  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ , i.e., every element of  $T_X P R T_{PX}$  must be an integer. Because of noise, this may not be exactly satisfied. (The lattice parameters of the conventional unit cell may not be known exactly.) In general, then, some strain will be required for the primitive lattice to exhibit exactly the symmetry required by  $R$ .

Let us define  $R' = T_X P R T_{PX}$ .  $R'$  is the point operation matrix with respect to the basis vectors of the primitive lattice, i.e.,

$$R\mathbf{p}_i = \sum_j \mathbf{p}_j R'_{ji}. \quad (25)$$

We want to make a small adjustment  $\Delta\mathbf{p}_i$  to each basis vector  $\mathbf{p}_i$  of the primitive lattice so that every element of  $R'$  is an integer. When this adjustment is made, every element in  $R'$  will be replaced by its nearest integer. We thus obtain

$$R(\mathbf{p}_i + \Delta\mathbf{p}_i) = \sum_j (\mathbf{p}_j + \Delta\mathbf{p}_j) [R'_{ji}]_{\text{nint}}, \quad (26)$$

where  $[x]_{\text{nint}}$  is the nearest integer to  $x$ . Writing the  $\mu$ th cartesian component of the equation, we have

$$\sum_{\nu} (R_{\mu\nu} \Delta p_{i\nu} - \sum_j [R'_{ji}]_{\text{nint}} \Delta p_{j\mu}) = - \sum_{\nu} R_{\mu\nu} T_{PX\nu i} + \sum_j T_{PX\mu j} [R'_{ji}]_{\text{nint}}. \quad (27)$$

There are 9 equations (all possible pairs of integers  $i, \mu$ ) and 9 variables  $\Delta p_{i\mu}$ . However, the equations are not likely to be linearly independent, so we need an additional condition: the rms average of the fractional adjustments  $\Delta \mathbf{p}_i / |\mathbf{p}_i|$  to the basis vectors is minimized (see Appendix).

The unit basis vectors of the cartesian coordinates can be written in terms of the basis vectors of the primitive lattice:

$$\hat{\mathbf{e}}_{\mu} = \sum_i \mathbf{p}_i T_{Xi\mu}. \quad (28)$$

Therefore the strain tensor is given by

$$\epsilon_{\mu\nu} = \sum_j T_{Xpj\mu} \Delta p_{j\nu}. \quad (29)$$

We symmetrize  $\epsilon$  (to remove pure rotations), diagonalize the result, and calculate the rms average of the diagonal elements. This gives us a single number which characterizes the amount strain necessary for the primitive lattice to exhibit the symmetry required by  $R$ . If the value of this number is less than some predetermined maximum allowed value, we accept  $R$  as a possible symmetry element of the lattice. We adjust the basis vectors of the primitive lattice so that they now have the symmetry required by  $R$ , i.e., we use new basis vectors  $\mathbf{p}'_i = \mathbf{p}_i + \Delta \mathbf{p}_i$  and new matrices  $T'_{PX}$  and  $T'_{XP}$ . We also adjust the cartesian coordinates of each atom,  $\mathbf{r}'_i = T'_{PX} T_{XP} \mathbf{r}_i$ , so that the strain is uniform throughout the primitive unit cell.

Now we try to find an element  $\{R|\mathbf{f}\}$  of the space group. If such an element exists, then for each atom at  $\mathbf{r}'_i$ , there must exist an atom of the same type at  $R\mathbf{r}'_i + \mathbf{f}$ . We can find the translational part  $\mathbf{f}$  using the following method. We choose any atom  $i$ . We then form vectors from  $R\mathbf{r}'_i$  to every atom the same type (including atom  $i$  itself), i.e.  $\mathbf{f} = \mathbf{r}'_j - R\mathbf{r}'_i$  for every atom  $j$  of the same type as atom  $i$ . We test each of these vectors  $\mathbf{f}$  by adding it to  $R\mathbf{r}'_k$  of each atom  $k$  in the unit cell and then looking to see if there is an atom of the same type at  $R\mathbf{r}'_k + \mathbf{f}$ .

Again, we must take noise into account. There may not be an atom at exactly  $R\mathbf{r}'_k + \mathbf{f}$ . We therefore find the atom *nearest* to the position of  $R\mathbf{r}'_k + \mathbf{f}$ . This means that we try every atom  $m$  in the unit cell which is the same type as atom  $k$  and choose the atom for which the value of

$$d = |[R\mathbf{r}'_k + \mathbf{f} - \mathbf{r}'_m]_{\min}| \quad (30)$$

is the smallest. The lattice vector

$$\mathbf{u}_k = R\mathbf{r}'_k + \mathbf{f} - \mathbf{r}'_m - [R\mathbf{r}'_k + \mathbf{f} - \mathbf{r}'_m]_{\min} \quad (31)$$

points to the unit cell containing atom  $m$  near  $R\mathbf{r}'_k + \mathbf{f}$  so that

$$R\mathbf{r}'_k + \mathbf{f} \approx \mathbf{r}'_m + \mathbf{u}_k. \quad (32)$$

We also define a mapping function  $M(k) = m$ , i.e., the action of  $\{R|\mathbf{f}\}$  maps atom  $k$  onto atom  $m$ . We test the properties of the mapping functions. If  $R^n = E$ , the identity operation, then  $M^n(i) = i$  for each atom  $i$ . Also,  $M^{-1}(k)$  must be single-valued. If we examine the series,  $M(k), M^2(k), M^3(k), \dots$ , where  $M^2(k) = M(M(k))$ ,  $M^3(k) = M(M^2(k))$ , etc., we find that  $M^n(k) = k$  for some minimum value of  $n$ . (This value of  $n$  may be less than the value for which  $R^n = E$ .) This defines a period of a cycle.

Now, we adjust the position of every atom and also the fractional  $\mathbf{f}$  so that the crystal has the exact symmetry required of  $\{R|\mathbf{f}\}$ :

$$R(\mathbf{r}'_k + \Delta\mathbf{r}'_k) + \mathbf{f} + \Delta\mathbf{f} = \mathbf{r}'_{M(k)} + \Delta\mathbf{r}'_{M(k)} + \mathbf{u}_k. \quad (33)$$

The  $\mu$ th cartesian component of that equation is

$$\sum_{\nu} R_{\mu\nu}(r'_{k\nu} + \Delta r'_{k\nu}) + f_{\mu} + \Delta f_{\mu} = r'_{M(k)\mu} + \Delta r'_{M(k)\mu} + u_{k\mu}. \quad (34)$$

This gives us  $3N$  equations (all possible pairs of values for  $k, \mu$ ) for the  $3N + 3$  variables,  $\Delta r'_{k\mu}$  and  $f_{\mu}$ . We find the solution for which the rms average adjustment to the atomic positions is a minimum (see Appendix). Then we calculate this rms average. If its value is less than some predetermined maximum allowed value, we accept  $\{R|\mathbf{f}\}$  as a element of the space group. (We use the adjusted value of  $\mathbf{f}$  for the translational part.) If not, we try another value for  $\mathbf{f}$ . We try every candidate  $\mathbf{f} = R\mathbf{r}'_j - \mathbf{r}'_i$ , where, as stated above,  $\mathbf{r}'_i$  is the position of some fixed atom  $i$ , and  $\mathbf{r}'_j$  is the position of every atom in the unit cell of the same type as atom  $i$ .

After trying to form space-group elements from each of the 64 point operations, we finally obtain a list of  $n$  space-group elements,  $\{R_1|\mathbf{f}_1\}, \{R_2|\mathbf{f}_2\}, \dots, \{R_n|\mathbf{f}_n\}$ , where  $R_1 = E$  and  $\mathbf{f}_1 = 0$ . Each element  $\{R_i|\mathbf{f}_i\}$  maps atom  $k$  onto atom  $M_i(k)$ . The mapping must obey the same multiplication table as the space-group elements, i.e., if  $R_i R_j = R_m$ , then  $M_i(M_j(k)) = M_m(k)$  for every atom  $k$ .

Now, the basis vectors of the primitive lattice must exhibit the symmetry of every point operation  $R_i$  which is part of some space-group element. We obtain an equation similar to Eq. (27),

$$\sum_{\nu} (R_{m\mu\nu} \Delta p_{i\nu} - \sum_j [R'_{mj\mu}]_{\text{nint}} \Delta p_{j\mu}) = - \sum_{\nu} R_{m\mu\nu} T_{PX\nu i} + \sum_j T_{PX\mu j} [R'_{mj\mu}]_{\text{nint}}, \quad (35)$$

where  $R'_m = T_{XP} R_m T_{PX}$ . We only need to include the equations involving the point operation part of the generating elements of the the space group. We solve these equations under the condition that the rms average of the fractional adjustments  $\Delta\mathbf{p}_i/|\mathbf{p}_i$  to the basis vectors be a minimum (see Appendix). We adjust the basis vectors of the primitive lattice so that they now have the symmetry required by  $R$ , i.e., we use new



basis vectors  $\mathbf{p}'_i = \mathbf{p}_i + \Delta\mathbf{p}_i$  and new matrices  $T'_{PX}$  and  $T'_{XP}$ . We also adjust the cartesian coordinates of each atom,  $\mathbf{r}'_i = T'_{PX}T_{XP}\mathbf{r}_i$ , so that the strain is uniform throughout the primitive unit cell.

Now, we adjust the position of every atom and also the translational part  $\mathbf{f}_i$  of every space group element so that the crystal has the exact symmetry required of the space group.

$$R_m(\mathbf{r}'_k + \Delta\mathbf{r}'_k) + \mathbf{f}_m + \Delta\mathbf{f}_m = \mathbf{r}'_{M_m(k)} + \Delta\mathbf{r}'_{M_m(k)} \text{ modulo } \mathbf{p}'. \quad (36)$$

The  $\mu$ th cartesian component of that equation is

$$\sum_{\nu} R_{m\mu\nu} \Delta r'_{k\nu} - \Delta r'_{M_m(k)\mu} + \Delta f'_{m\mu} = - \left[ \sum_{\nu} R_{m\mu\nu} r'_{k\nu} - r'_{M_m(k)\mu} + f_{m\mu} \right]_{\min}. \quad (37)$$

We only need to include the equations involving the generating elements of the the space group. We find the solution for which the rms average adjustment to the atomic positions is a minimum (see Appendix). We make the adjustments to the atomic positions and to the translational parts of the space-group elements.

We next adjust the origin of the space group so that the translational part of each space-group element is as simple as possible. If we move the origin to the position  $\boldsymbol{\tau}$ , then each space-group element  $\{R_i|\mathbf{f}_i\}$  becomes  $\{R_i|\mathbf{f}_i - \boldsymbol{\tau} + R_i\boldsymbol{\tau}\}$ . Thus we have for the new translational parts,

$$\mathbf{f}'_i = \mathbf{f}_i - \boldsymbol{\tau} + R_i\boldsymbol{\tau}. \quad (38)$$

The  $\mu$ th cartesian component of this equation is

$$f'_{i\mu} + \tau_{\mu} - \sum_{\nu} R_{i\mu\nu} \tau_{\nu} = f_{i\mu}. \quad (39)$$

We only need to include the equations involving the generating elements of the the space group. We solve these equations under the conditions that we set as many of the values of  $f'_{ij}$  to zero as possible. The adjusted fractionals  $\mathbf{f}'$  can be written as linear combinations of the basis vectors of the primitive lattice with coefficients which are simple rational fractions, i.e., every element of  $T_{XP}\mathbf{f}'_i$  is a simple rational fraction. The denominators of these fractions are probably restricted to the values, 2,3,4,6,8.

With each space-group element written in simple form, we can now identify the space group in the *International Tables for Crystallography*. This process has been described elsewhere [D. M. Hatch and H. T. Stokes, Phys. Rev. B **31**, 2908 (1985)]. The identification will include the space-group number, the new conventional basis vectors, and the new origin.

## 6. Subgroups

When the data is noisy, it is sometimes useful to see how well we can fit the data to subgroups of the space group found above. For example, we may find that certain space-group elements fit the data much worse than others, and certain subgroups may therefore fit the data much better than the space group found above.

Let us start with the lattice points found inside the unit cell. These are represented by a set of vectors  $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n(V)}\}$ , where  $n(V)$  is the number of lattice points found. We include the point at the origin,  $\mathbf{v}_1 = 0$ , so that we can form translation groups from these vectors. Group multiplication is defined to be vector summation modulo  $\mathbf{t}$ , a lattice vector of the unit cell. Note that it is possible that  $V$  itself is not a group. (For example, maybe two vectors barely meet the criteria to be included but the sum of these two vectors does not.) We can however, form all subsets of  $V$  which are groups. We will call these sublattices and denote them by  $V_i = \{\mathbf{v}_1^{(i)}, \mathbf{v}_2^{(i)}, \dots, \mathbf{v}_{n_i(V)}^{(i)}\}$ , where  $n_i(V)$  is the number of elements in the sublattice. Each sublattice is closed under group multiplication, i.e., if  $\mathbf{v}_j^{(i)}$  and  $\mathbf{v}_k^{(i)}$  are elements of  $V_i$ , then  $\mathbf{v}_j^{(i)} + \mathbf{v}_k^{(i)}$  modulo  $\mathbf{t}$  is also an element of  $V_i$ .

We consider each sublattice at a time. We consider the sublattice to be the primitive lattice of the crystal. We find the basis vectors  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  of this primitive lattice using the method described in a previous section. We also find coset representatives  $\mathbf{s}_j$  of the largest supergroup  $V'$  of  $V_i$  in  $V$  with respect to  $V_i$ , i.e.

$$V' = \mathbf{s}_1 V_i + \mathbf{s}_2 V_i + \dots + \mathbf{s}_{n'_i(V)} V_i, \quad (40)$$

where no two cosets  $\mathbf{s}_j V_i, \mathbf{s}_k V_i$  have any common elements and where every element of each coset  $\mathbf{s}_j V_i$  is in  $V$ .  $n'_i(V)$  is the number of cosets, where  $n'_i(V)n_i(V) \leq n(V)$ . If  $V$  is a group, then  $V' = V$  and  $n'_i(V) = n(V)/n_i(V)$ .

Given this primitive lattice, we proceed to find possible space-group elements using the same methods described in a previous section. We obtain a set of such operators:  $W = \{\{R_1|\mathbf{f}_1\}, \{R_2|\mathbf{f}_2\}, \dots, \{R_{n(W)}|\mathbf{f}_{n(W)}\}\}$ , where  $n(W)$  is the number of elements in the set. Note that there is one element for each point operation. We only look for one fraction  $\mathbf{f}_j$  for each point operation  $R_j$ . However, if  $\{R_j|\mathbf{f}_j\}$  is a possible symmetry element, then  $\{R_j|\mathbf{f}_j + \mathbf{v}_k\}$  is also, where  $\mathbf{v}_k$  is some element of  $V$ . These additional forms of the space-group element only become significant when  $\mathbf{v}_k$  is not an element of  $V_i$ , since elements of  $V_i$  are just vectors of the primitive lattice. We can cover all unique possibilities if we add each of the coset representatives  $\mathbf{s}_m$  to  $\mathbf{f}_j$ .

We thus combine the set  $W$  with the coset representatives  $\mathbf{s}_m$  to form the following set:

$$\begin{aligned} G = \{ & \{E|0\}, \\ & \{R_2|\mathbf{f}_2\}, \{R_2|\mathbf{f}_2 + \mathbf{s}_2\}, \dots, \{R_2|\mathbf{f}_2 + \mathbf{s}_{n'_i(V)}\}, \\ & \{R_3|\mathbf{f}_3\}, \{R_3|\mathbf{f}_3 + \mathbf{s}_2\}, \dots, \{R_3|\mathbf{f}_3 + \mathbf{s}_{n'_i(V)}\}, \\ & \dots \\ & \{R_{n(W)}|\mathbf{f}_{n(W)}\}, \{R_{n(W)}|\mathbf{f}_{n(W)} + \mathbf{s}_2\}, \dots, \{R_{n(W)}|\mathbf{f}_{n(W)} + \mathbf{s}_{n'_i(V)}\} \} \end{aligned} \quad (41)$$

Note that we have excluded any elements of the form  $\{E|\mathbf{s}_j\}$  for  $j \neq 1$ . We form all subsets of  $G$  which are groups. In the group multiplication, vectors are always considered modulo  $\mathbf{b}$ , vectors of the primitive lattice. Each group will contain no more than one element containing a particular point operation. The elements of each group form the coset representatives of a space group with respect to the translation group of the

primitive lattice. As described in a previous section, we find the strain and atomic displacements required so that the crystal exhibits the exact symmetry of the space group. We compare the rms averages of these values to predetermined maximum values. If they pass the test, we identify this space group and list it in the output. We repeat this procedure for every subgroup of  $G$  and for every subgroup of  $V$ .

## Appendix: Minimization of variables

Suppose we have  $n$  linearly independent equations. The  $i$ th equation has the following form:

$$x_i + \sum_{j=1}^m A_{ij} y_j = B_i. \quad (42)$$

$y_1, y_2, \dots, y_m$  are independent variables, and  $x_1, x_2, \dots, x_n$  are dependent variables.  $A_{ij}$  and  $B_i$  are constants. We want to find a solution to these equations for which the rms average of the first  $n'$  dependent and  $m'$  independent variables is a minimum. We minimize

$$\begin{aligned} S &= \sum_{i=1}^{n'} x_i^2 + \sum_{j=1}^{m'} y_j^2 \\ &= \sum_{i=1}^{n'} \left( B_i - \sum_{j=1}^m A_{ij} y_j \right)^2 + \sum_{j=1}^{m'} y_j^2. \end{aligned} \quad (43)$$

Minimizing with respect to each of the independent variables,

$$\sum_{i=1}^{n'} A_{ik} x_i = \begin{cases} y_k, & 1 \leq k \leq m', \\ 0, & m' + 1 \leq k \leq m. \end{cases} \quad (44)$$

This adds  $m$  equations to the set of equations, so that we now have  $n + m$  equations and  $n + m$  variables. Each variable can now be uniquely determined.