



Linearization routines for the parameter space concept to determine crystal structures without Fourier inversion

Muthu Vallinayagam,^{a,b,‡} Melanie Nentwich,^{c,*‡} Dirk C. Meyer^{b,d} and Matthias Zschornak^{a,b,d,*}

Received 4 November 2024

Accepted 28 February 2025

Edited by A. Barty, DESY, Hamburg, Germany

Dedicated to our revered teacher Professor Karl Fischer.

‡ These authors contributed equally to this work.

Keywords: crystal structure determination; centrosymmetric structures; two- and three-dimensional parameter space; linearization; parameter space concept; Monte Carlo simulation; one-dimensional projection; X-ray diffraction.

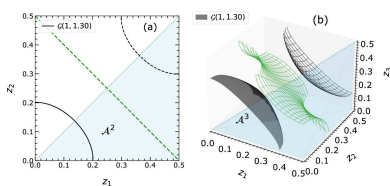
Supporting information: this article has supporting information at journals.iucr.org/j

^aTechnical Physics, University of Applied Sciences, Friedrich-List-Platz 1, 01069 Dresden, Germany, ^bCenter for Efficient High Temperature Processes and Materials Conversion ZeHS, TU Bergakademie Freiberg, Winklerstr. 5, 09596 Freiberg, Germany, ^cDeutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany, and ^dInstitute of Experimental Physics, TU Bergakademie Freiberg, Leipziger Str. 23, 09596 Freiberg, Germany. *Correspondence e-mail: melanie.nentwich@desy.de, matthias.zschornak@physik.tu-freiberg.de

We present the elaboration and first generally applicable linearization routines of the parameter space concept (PSC) for determining one-dimensionally projected structures of m independent scatterers. This crystal determination approach does not rely on Fourier inversion but rather considers all structure parameter combinations consistent with available diffraction data in a parameter space of dimension m . The method utilizes m structure-factor amplitudes or intensities represented by piecewise analytic hyper-surfaces to define the acceptable parameter regions. The coordinates of the point scatterers are obtained through the intersection of multiple isosurfaces. This approach allows for the detection of all possible solutions for the given structure-factor amplitudes in a single derivation. Taking the resonant contrast into account, the spatial resolution achieved by the presented method may exceed that of traditional Fourier inversion, and the algorithms can be significantly optimized by exploiting the symmetry properties of the isosurfaces. The applied one-dimensional projection demonstrates the efficiency of the PSC linearization approach based on fewer reflections than Fourier sums. Monte Carlo simulations, using the projections of various random two- and three-atom structure examples, are presented to illustrate the universal applicability of the proposed method. Furthermore, ongoing efforts aim to enhance the efficiency of data handling and to overcome current constraints, promising further advancements in the capabilities and accuracy of the PSC framework.

1. Introduction

Solving crystal structures from diffraction intensities plays a vital role in many areas of solid-state research, *e.g.* physics, chemistry, mineralogy, materials sciences, biology and pharmacy, as it forms the fundamental basis for understanding the properties of materials as well as their effects and functionalities. The corresponding databases grow by tens of thousands of structures every year. The state-of-the-art structure determination methodology is based on Fourier inversion (FI) of the scattering density (*e.g.* electron density for X-ray diffraction, nuclear density for neutron diffraction). In the early days, the developments in crystallography were mainly based on computationally efficient FI techniques either directly or indirectly, such as the charge flipping method (Oszlányi & Süto, 2004), the algebraic method (Rothbauer, 1994; Rothbauer, 1995; Rothbauer, 1998), geometrical methods (Navaza & Silva, 1979), analytical function methods (Cervellino & Ciccariello, 2005), fit methods such as Rietveld refinement (Toby, 2024) and matching learning algorithms (Shi, 2022; Munteanu *et al.*, 2024; Billinge & Proffen, 2024).



The currently used FI techniques to construct electron density systematically introduce noise and errors in the calculation due to series termination. Therefore, a large number of terms in the FI series are required, which in turn necessitates a substantial set of experimental observations. Furthermore, the quality of X-ray diffraction intensities greatly influences the FI series, with weaker observations contributing less. However, the experimental database is in most cases incomplete since only the quadratic amplitudes of the Fourier coefficients (*i.e.* the structure factors via the reflection intensities) can be determined, and the well known phase problem of crystallography makes solving the structure more challenging (Harrison, 1993; Fischer *et al.*, 2005; Fischer *et al.*, 2008; Kirfel & Fischer, 2009).

In order to overcome the demerits of FI techniques, alternative methods have been developed. In this study, we examine the relationship between the structure factor and the atomic positions in crystal structures by considering the geometrical correlations. In general, an m -atom structure has $3m$ free positional parameters to be determined, which include the x , y and z coordinates of all m atoms. To simplify the structure-solving process with the parameter space concept (PSC), the task is split into several independent one-dimensional projections (in real space), each providing the solution of m parameters within an m -dimensional parameter space (PS; space of atomic coordinates with the orthogonal basis in \mathbb{R}^m) (Zschornak *et al.*, 2024; Fischer *et al.*, 2005; Knop, 1989; Pilz, 1996; Ott, 1928). Each point in the PS corresponds to a possible combination of atomic coordinates (*e.g.* projected onto the z axis) and generates a unique X-ray diffraction intensity for a predefined reflection. Vice versa, the set of points that reproduce the experimentally observed intensity of a particular reflection defines a manifold called an isosurface [see Figs. 1(a)–1(c)]. The intersection point from all isosurfaces of different reflections expresses the intended structure [see Fig. 1(d)].

The theory of PSC has been developed during the past two decades, mainly focusing on equal atoms (Knop, 1989; Pilz, 1996; Ott, 1928), aiming to achieve higher spatial resolution than the FI techniques while at the same time using a reduced

number of available intensity data sets. Apart from these advantages, PSC will always recognize all possible solutions that can reproduce the given experimental intensities, although at the cost of parameterizing functions in continuous high-dimensional spaces. Hence, PSC provides an elegant but computationally expensive method to solve crystal structures in a stepwise approach, splitting the full structure into one-dimensional projections.

In the present work, we develop further theoretical approaches to treat all aspects of m -dimensional PSs, in particular generally applicable linearization routines to parameterize the isosurfaces for efficient functional handling as well as computational storage and determination of intersections. The PSC algorithms are enhanced to treat the artificial values of the atomic scattering factor of scatterers in the m -dimensional PS. The implemented capability to overcome the previously employed equal point atom (EPA) model (see Section 2.2) improves PSC in the direction of a generally applicable structure determination approach. However, to test the developed algorithms and code, in this article, we determine the centrosymmetric structures only in two- and three-dimensional PS. Nevertheless, on the first estimation, the derived equations may be straightforwardly enhanced towards higher dimensions, which will be the focus of our continuous research efforts.

This study is structured as follows: The subsequent sections provide an in-depth exploration of our approach, commencing with the fundamental theory underlying the PSC-based framework. Next, the steps involved in solving a maximum of three parameters in the one-dimensional projection are discussed, including a general description of the linear approximation of isosurfaces. Finally, we present noteworthy generalizations based on Monte Carlo simulations for structure determination within the two- and three-dimensional centrosymmetric PS.

2. Theoretical methods

Solving a crystal structure is a meticulous process that involves the precise determination of the position of each atom within

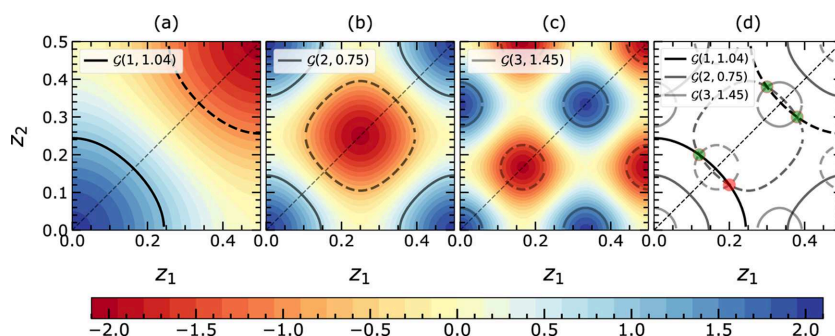


Figure 1

Basic explanation of the parameter space concept. (a)–(c) Two-dimensional parameter space for the projection of a crystal structure of two equal scatterers onto the z axis for reflections $00l$, $l = 1, 2, 3$. The color map represents the calculated amplitude for each combination of atomic coordinates. The isosurfaces for positive and negative amplitudes generated by the arbitrary atomic coordinates $[0.2, 0.12]$ are highlighted with solid and dashed lines, respectively. (d) Overlay of the isosurfaces from (a)–(c) with their intersecting points highlighted by red and green circles. The intersection point at the red circle is the intended structure to be found, and alternative solutions are at the green circles.

the crystal. In the context of PSC analysis, an effective strategy is employed, making use of a linear approximation of the isosurfaces defined by intensities derived from X-ray diffraction data within the corresponding PS. Careful analysis of the experimentally measured intensity of each reflection facilitates the reconstruction of the potential combinations of atomic coordinates by utilizing sophisticated piecewise analytic hyper-surfaces (Fischer *et al.*, 2005; Zschornak *et al.*, 2024).

2.1. Parameter space and isosurfaces

The complete solution of a structure with m atoms in the unit cell consists of $3m$ parameters, accounting for the three distinct coordinate components (x , y and z) of each atom. The $3m$ -dimensional space containing all possible combinations of those parameters is called the parameter space. By employing a one-dimensional projection of the crystal structure onto the main axes, the complex task of solving $3m$ coordinates is separated into several distinct steps, each solving m coordinates in the m -dimensional PS \mathcal{P}^m . Then, assuming oblique reflections (more than one non-zero component), the relationship between the x , y and z components of atomic coordinates can be assigned (Fischer *et al.*, 2008). However, the necessary steps to combine the projections into a complete structure solution will be the topic of a further article.

Here, we adopt the projection onto the z axis as a representative projection for all axes, while employing the Miller index l to abbreviate the relevant X-ray reflections $00l$. In general, the PS \mathcal{P}^m consists of m orthonormal axes, which we assign to the z components z_1, \dots, z_m of the atomic coordinates. When considering a specific reflection, each point $\mathbf{z} = (z_1, \dots, z_m)$ in \mathcal{P}^m has a uniquely defined amplitude, which is directly related to the intensity (see Section 2.2). However, a specific amplitude value can be achieved through several \mathbf{z} , resulting in an ‘isosurface’. The isosurface embodies the entirety of possible combinations of atomic coordinates that generate the same amplitude for a specific reflection within the intricate landscape of \mathcal{P}^m . Each reflection provides insights into the possible configurations of the crystal structure: only the well defined combination of atomic coordinates described by all the isosurfaces can create the experimentally determined amplitude. The precise structure solution is, thus, defined as the intersection of the isosurfaces corresponding to different reflections [see Fig. 1(d)]. Through detailed and precise analysis, we unravel the regions where these isosurfaces intersect, which are ultimately interpreted as the coordinates of the atoms within the crystal structure. For error-free intensity values and using the isosurfaces directly, only m different reflections are required to solve for m coordinates, realized as the intersection of m isosurfaces. However, in the cases of linearized isosurfaces (see Section 2.4) as well as experimental and thus naturally erroneous values, the isosurfaces gain volume, and, therefore, so does the intersection region. An exact point solution cannot be determined; however, adding more reflections will improve the accuracy of the solution.

2.2. The structure factor

The basic requirement of structure determination is information about the Miller index and its corresponding intensity I , which can be expressed as $I \propto |F|^2$, with the structure factor F . The complete structure factor is expressed as a weighted sum over atomic scattering amplitudes f_j of all atoms present in the crystallographic unit cell. However, the natural disorder due to lattice vibrations and defects consequently leads to averaging over all unit cells using the parameters site occupancy o_j and Debye–Waller (DW) factor $\exp(-M_j)$. The DW factor accounts for the reduction in scattering amplitude caused by the positional uncertainty of atom j at \mathbf{r}_j , leading to the expression (Richter *et al.*, 2018)

$$F(E, \mathbf{Q}) = \sum_{j \in \text{u.c.}} o_j f_j(E, \mathbf{Q}) \exp(-M_j) \exp(i\mathbf{Q} \cdot \mathbf{r}_j), \quad (1)$$

where u.c. is the unit cell. As a simple treatment, temperature-induced lattice vibrations are approximated in the harmonic limit by the mean square displacement \mathbf{u}_j of atom j in the direction of the momentum transfer vector \mathbf{Q} via $M_j = (1/2)\langle(\mathbf{Q} \cdot \mathbf{u}_j)^2\rangle$. The anisotropic nature of atomic displacements is typically described using the anisotropic displacement parameter tensor U^{ji} , which provides a more comprehensive representation of the mean square displacement (Trueblood *et al.*, 1996). Nevertheless, as a first attempt to develop the linearization method, equation (1) is adapted by assuming that the atomic sites have full occupancy ($o_j = 1$) and are not influenced by temperature (*i.e.* the DW factor is neglected). These additional degrees of freedom can be easily implemented as additional dimensions within the PSC approach. They will specifically affect the topology and values of the respective isosurfaces. The precise application and validation are subjects for further developments of the code during the next phase of the PSC project. Also, we will constrain the discussion to centrosymmetric cases, which provides the advantage that the solution is not a complex value but lies in the real number space. Thus, we can express the structure factor, equation (1), as (Fischer *et al.*, 2005; Kirfel *et al.*, 2006)

$$F(l) = 2 \sum_{i=1}^m f_i(l) \cos(2\pi l z_i) = 2s(l) \left| \sum_{i=1}^m f_i(l) \cos(2\pi l z_i) \right|, \quad (2)$$

where l is the Miller index of reflection $00l$, m is the number of atoms in the unit cell, $s(l)$ is the sign of the expression, f_i is the atomic structure factor (a real number, considering only Thomson scattering without resonance corrections) and z_i is the coordinate of the atom with index i .

If all atoms are considered to be equal (and point-like), then their scattering factors are identical, $f_i(l) = f(l)$, and can be set to unity; hence, equation (2) becomes

$$G(l) = 2s(l) \left| \sum_{i=1}^m \cos(2\pi l z_i) \right| = 2s(l)g(l), \quad (3)$$

where

$$g(l) = \left| \sum_{i=1}^m \cos(2\pi lz_i) \right| \quad (4)$$

is the modulus of the geometric structure factor G . Detailed relationships for atoms on special positions are discussed by Fischer *et al.* (2005). The case of identical point-like atoms is referred to as the EPA model hereafter. The case of realistic scattering factors will be referred to as the non-EPA model.

Note the difference between the isosurface, *i.e.* the manifold, and the structure factor in terms of parameter dependencies. The isosurface is a subspace of dimension $m - 1$ in \mathcal{P}^m , fulfilling a boundary condition of coordinates z_i with a certain amplitude or intensity. For experimentally observed or theoretically calculated amplitudes $2g_0(l)$ and $|F_0(l)|$, the isosurface of the geometric structure factor is mathematically expressed for the EPA case as

$$\mathcal{G}[z_i, l, g_0(l)] \text{ satisfying } g(l) - g_0(l) = 0, \quad (5)$$

and that of the structure factor for the non-EPA case as

$$\mathcal{F}[z_i, l, |F_0(l)|/2] \text{ satisfying } |F(l)| - |F_0(l)| = 0. \quad (6)$$

For better readability we use the expressions $\mathcal{G}[l, g_0(l)]$ or $\mathcal{F}[l, |F_0(l)|/2]$ for EPA or non-EPA cases, respectively, where only l and $g_0(l)$ or $|F_0(l)|/2$ are specified explicitly. Further, the (geometric) structure factors $G(l)$ or $F(l)$ represent fixed structures z_i and only depend on l . Otherwise, simply \mathcal{G} or \mathcal{F} are used for EPA or non-EPA cases. The isosurface can be defined from given amplitudes $2g(l)$, $|F_0(l)|$ or intensities $|2g_0(l)|^2$, $|F_0(l)|^2$, which is indicated with the respective index A or I on \mathcal{G} or \mathcal{F} .

The value of the amplitude $g(l)$ for a given reflection l is visualized as a function of the atomic coordinates z_1 and z_2 as color maps in Fig. 1. Also, Fig. 1 shows an isosurface corresponding to an arbitrary set of amplitudes $g_0(l)$, represented by black lines. To account for the complexities of realistic

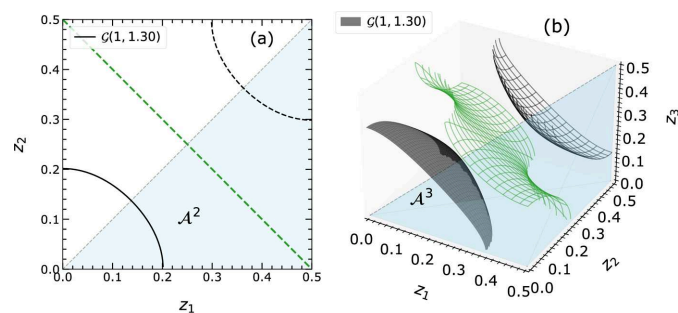


Figure 2

Isosurfaces \mathcal{G} for reflection $l = 1$ in two- and three-dimensional PS in (a) and (b), respectively, with highlighted asymmetric parts \mathcal{A}^2 and \mathcal{A}^3 . The black continuous line in (a) and solid surface in (b) represent the instance of the isosurface with the positive sign [$s(l) = +1$]. The black dashed line in (a) and wireframe in (b) represent the negative instance [$s(l) = -1$]. The blue shaded regions represent \mathcal{A}^2 and \mathcal{A}^3 of \mathcal{P}^2 and \mathcal{P}^3 , respectively, and contain the contributions from both $s(l) = \pm 1$. However, the knowledge or selection of $s(l)$ of the amplitude further restricts \mathcal{A}^2 and \mathcal{A}^3 , reducing the volume by a factor of two via the zero $\mathcal{G}[l, g(l) = 0]$ isosurface represented by the green dashed line in \mathcal{P}^2 and wireframe in \mathcal{P}^3 .

applications including different atoms and their scattering behavior, we will also consider various combinations of scattering factors. Although f_i are complex quantities influenced by factors like the energy of the incoming X-rays and inter-lattice distances (Woolfson & Hai-Fu, 1995), we have carefully chosen appropriate real numbers representing heavy, medium and light atoms.

We realized that varying scattering factors f_i affect the topology of the isosurfaces [*cf.* equation (2)]. In EPA, f_i are set to unity, and hence the isosurfaces \mathcal{G} with large g_0 values are almost a circle in \mathcal{P}^2 and a sphere in \mathcal{P}^3 (*cf.* Figs. 1 and 2). In the non-EPA model, the isosurface \mathcal{F} exhibits a variety of topologies, that require careful and independent handling (Section 2.4.1). Furthermore, the overall appearance of the isosurfaces in the PS is influenced not only by the ratios of scattering factors but also by the reflection index l . As the index l increases, the number of disjoint isosurface regions also increases (see Fig. 1). Section 2.4 gives more details on the systematic approaches.

The detailed analysis of intrinsic symmetries can contribute to a further reduction of the computational effort, by reducing the possible solution space \mathcal{P}^m to the asymmetric PS \mathcal{A}^m . Those symmetries comprise (i) the centrosymmetry of the structure, (ii) the permutation symmetry for equal or partially equal atoms, and (iii) the choice of origin. The assumed centrosymmetry leads to the spatial limitation of the PS to the range $[0, 0, 0, 0.5]^m c$, where c is the lattice constant of the crystal in the specified projection.

Furthermore, following the discussion by Fischer *et al.* (2005), in the case of EPA, the full PS \mathcal{P}^m can be reduced by permutation of atomic coordinates, which encompasses both positive and negative instances of the isosurfaces. This reduction can be visualized using the corner points fixed at coordinates $[0, 0, \dots, 0]$, $[0.5, 0, \dots, 0]$, \dots , $[0.5, 0.5, \dots, 0.5]$ (*cf.* entire shaded area in Fig. 2). Even in the non-EPA cases, the occurrence of n equal atoms will induce limited permutation symmetry within the respective subspace of the PS, for which a reduction of \mathcal{P}^{m-n} towards \mathcal{A}^{m-n} can be obtained. The possible cases for $m = 3$ are shown in Fig. 3. As an example, if f_1 and f_2 are equal, then \mathcal{P}^3 is halved and \mathcal{A}^3 is a triangular prism as shown in Fig. 3(b).

In addition to the permutation of atomic coordinates, the choice of the origin of the crystal system at one of the two centers of inversion (at $[0, 0, \dots, 0]$ and $[0.5, 0, \dots, 0]$) provides another intrinsic symmetry to reduce the possible solution space where the given structure resides. The composition of the set of f_i defines a specific zero isosurface, one boundary of the asymmetric unit given by the isosurface $\mathcal{G}[1, g(1) = 0]$ [*cf.* Fig. 3(d) as an example for \mathcal{P}^3]. Once the choice of origin is utilized only the blue region below or above the zero isosurfaces needs to be analyzed. These boundaries simultaneously define the linearization boundaries and the allowed solution space for the structure investigation.

Additionally, \mathcal{P}^m can be reduced using the sign of the isosurfaces if available; this is represented for $l = 1$ by the separation of the shaded region in Figs. 2(a) and 2(b) via the zero isosurfaces $\mathcal{G}[l, g(l) = 0]$.

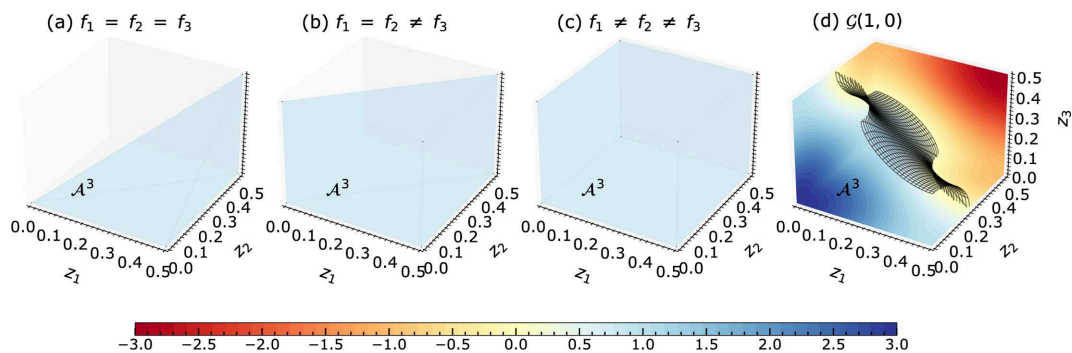


Figure 3

The definition of asymmetric units upon specific sets of f values is shown in (a)–(c). The \mathcal{A}^3 in (a) is defined for the EPA framework where all f_i are equal. The \mathcal{A}^3 in (b) and (c) are defined for the non-EPA framework. In (b) it is assumed that two f s (along the z_1 and z_2 directions) are the same and f_3 is different. In (c) it is assumed that all f s are different and hence \mathcal{A}^3 is equivalent to \mathcal{P}^3 . In (d) $\mathcal{G}[l = 1, g_0(l)]$ are given for the EPA case representing the magnitude of $g_0(l)$ from -3 to 3 in color code. The zero isosurface $\mathcal{G}[l = 1, g_0(l) = 0]$ separates the PS into two halves containing isosurfaces $\mathcal{G}[l = 1, g_0(l)]$ of positive or negative signs. When applying the choice of origin symmetry, the origin is fixed at $[0, 0, 0]$. Then the asymmetric unit reduces to half under the zero isosurface, containing only the positive signs $s(l = 1)$, for the solution search. The color bar gives the magnitude of $g_0(l = 1)$ with the applied sign.

2.3. General linearization approach within PSC

The primary objective of structure determination in \mathcal{P}^m is to identify the intersection of isosurfaces \mathcal{G} or \mathcal{F} corresponding to different reflections (see Fig. 4). However, finding the intersection point directly is challenging as it involves complex trigonometric functions [cf. equation (2)]. One possible approach to reduce complexity, the focus of this article, consists of employing linearization techniques, which enable the replacement of intricate trigonometric expressions within linear boundaries. This approximation can simplify the problem and utilizes set theory to explore all potential solution regions. Note that the linear approximation may yield multiple solution regions as a result of step-wise intersections of linear approximants of different reflections. The approximation process generally involves the following steps, depicted in Fig. 4.

Initialization. In the first step, we implement specific measures to ensure the flawless execution of the linearization algorithm. This important step involves arranging the atomic structure factors f_i in a descending order, which helps us to determine the general curvature of the isosurface \mathcal{F} , as well as to apply the permutation symmetry. The larger f_i values correspond to heavier atoms and exert maximum control over the behavior of \mathcal{F} , while the smaller f_i values pertain to lighter atoms and thus smaller contributions in the interference dependencies. The influence of f_i on the curvature of \mathcal{F} is depicted in Section 2.4.1. In this work, we distinguish between different types of topologies: they are called closed if \mathcal{F} intersects all parameter axes and open if \mathcal{F} does not intersect with at least one axis.

Linear approximation. The simplest approximation method is linearization. We bound the curved isosurfaces of each reflection by two straight, parallel lines in \mathcal{P}^2 or two planes in \mathcal{P}^3 . The normal vector and distance from the origin are the required descriptors of the boundaries and are being determined in this step. The equations and routines for these parameters are described in detail in Section 2.4. Notably, the

linear approximation invokes the mean-value theorem at the core (Lozada-Cruz, 2020; Sahoo & Riedel, 1998). The challenge is the development of a new non-EPA linearization framework, which we present here for the first time, giving algorithms to linearize the isosurfaces governed by different atomic scattering factors. This non-EPA framework elevates the capability of PSC to handle realistic X-ray diffraction data.

Solution finding. Subsequently, the boundary descriptors obtained from the previous step are used to construct polytopes, described by a system of linear inequality equations with the number of variables identical to the dimension m of the PS. These polytopes are reflecting the PS region that matches the observed intensity for the specific reflection in linear approximation. The goal is to find the common regions that are enclosed by all polytopes of every reflection and that represent the projected atomic coordinates of the structure under investigation. To achieve this final solution, we search for the intersection regions of the polytopes created for successive reflections. In ideal cases, *i.e.* error-free amplitudes for a full set of reflections starting from $l = 1$, a single polytope solution region that uniquely represents the given structure may be identified. However, in many cases, PSC can yield multiple polytope solution regions in a consecutive intersection step for an arbitrary reflection l , or when taking into account intensities as observable without the knowledge of the structure factor's phase.

Data writing to HDF. After the solution-finding process, we obtain crucial information such as the volume of solution regions, the coordinates of the edges of these regions, the number of solutions and the computation time. These data are stored in a file. Additionally, details about processed reflections, the atomic scattering factor, the structure factor, and experimental or theoretical intensity/amplitude are also saved for future reference. We have chosen to use the hierarchical data format (HDF) due to its versatility as a data model, making it ideal for managing large and complex data sets (The HDF Group & Koziol, 2020). HDF allows for the storage of

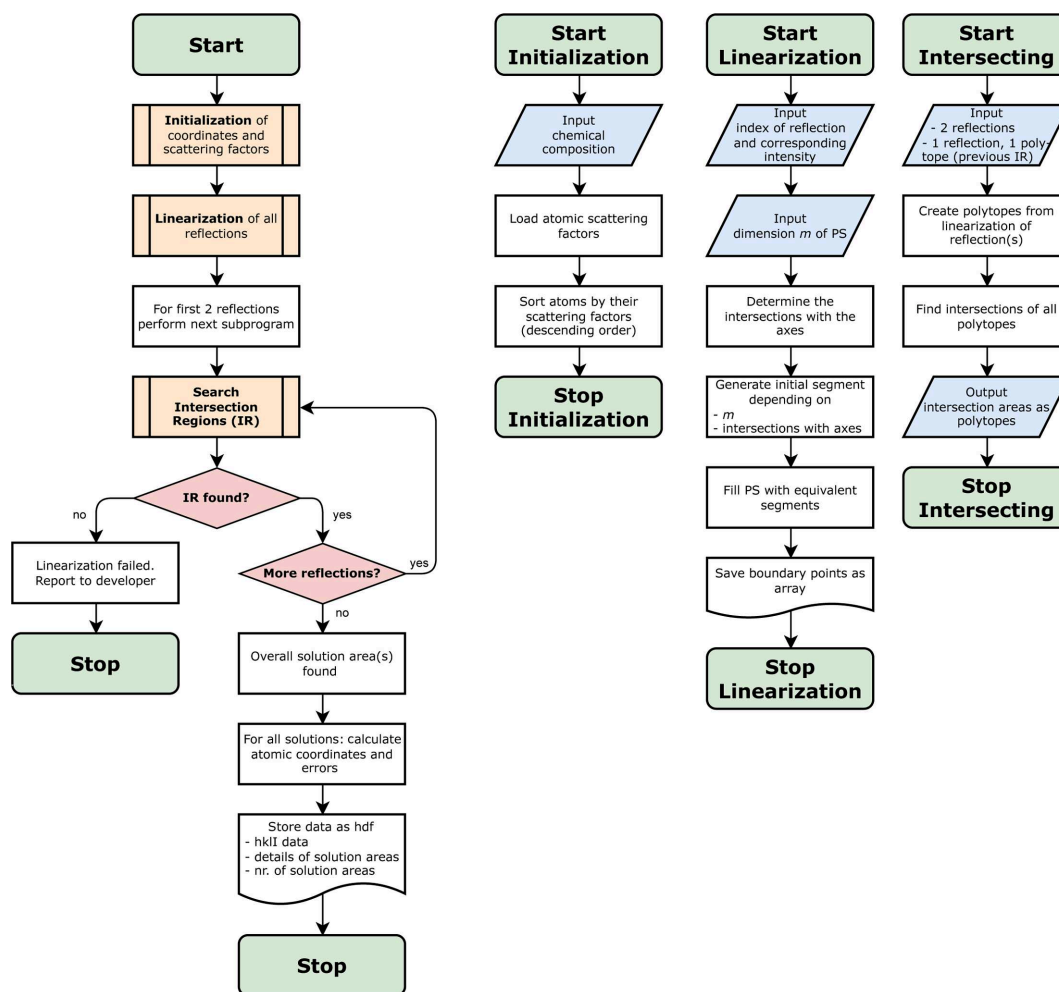


Figure 4

The scheme of operations flow in PSC includes the initialization, linearization and solution-finding process. Also, many intermediate decisions are made to verify the inclusion of all reflections, reaching intersection regions, imposing an accuracy limit on the area/volume of intersection regions *etc.* At the end of the structure determination, detailed outputs including reflections and their polytopes, intermediate intersection results, found solutions, the volume of each polytope, and the error on computed z_i are written in HDF-formatted files.

various types of data within a single file. There is an efficient Python library available that facilitates the integration of this specific file format into our routines. This format guarantees that the data are easily accessible and portable across different platforms, thereby enabling seamless analysis and sharing of results, making it a superior choice for data storage and retrieval. Furthermore, data stored in HDF files can be easily visualized using the *HDFView* tool, enhancing the practicality of this choice (Group & Koziol, 2020). More details about *HDFView* are given in the supporting information (see Section S2).

2.4. Linear approximation routines

Solving for the intersection point $\mathbf{z} = (z_1, z_2, \dots, z_m)$ on a dense grid in \mathcal{P}^m will become computationally expensive, challenging and sometimes even impossible for higher dimensions. However, we can reduce the computational effort significantly by solving for linearized approximations of the isosurfaces. Unfortunately, instead of solution points, linear-

ization introduces expanded solution regions, which ideally should be kept as small as possible. Here, we apply a linear approximation as described by Fischer *et al.* (2005): we develop a basic unit called a 'segment' to linearize a well defined part of the isosurface and extend the linearization by shifting and rotating the segment to create a full cover-up of the isosurfaces within \mathcal{P}^m . The segments consist of a set of parallel boundaries, the inner and outer limits of the isosurface with respect to the PS origin, and limiting boundaries perpendicular to each z_i direction. We have discovered that, in addition to the circle-like topologies considered by Fischer *et al.* (Fig. 2), band-like isosurfaces can also occur. The origin of these and also the different handling within the linearization procedure will be discussed in Section 2.4.1.

Examples of the process of linearizing \mathcal{G} for the EPA case are shown in Fig. 5. The complete procedure to accomplish linearization is explained in the following sections; steps include the generation of segments around the origin of the PS as well as the repetition of segments in the PS.

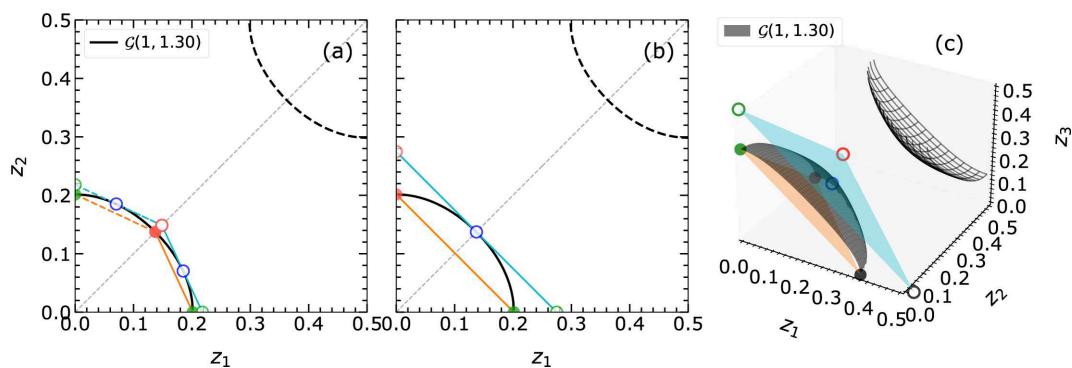


Figure 5 Linearization of an isosurface \mathcal{G} with (a) two segments in \mathcal{P}^2 , (b) one segment in \mathcal{P}^2 and (c) one segment in \mathcal{P}^3 . The filled and open circles represent the coordinates of the inner and outer boundary line/plane, respectively. The tangent point is represented by a blue open circle. The gray dashed line represents the mirror plane in \mathcal{P}^2 along the $[1, 1]$ direction created by the EPA model. An arbitrary value of 1.3 for the magnitude g_0 is selected and the isosurface is shown for reflection $l = 1$ in \mathcal{P}^2 and \mathcal{P}^3 .

2.4.1. Topologies of isosurfaces

In previous literature, the isosurfaces were always described as circle- or sphere-like, resulting in closed loops in \mathcal{P}^2 and \mathcal{P}^3 . Fischer *et al.* have not introduced the concept of topology as it was not a requirement, since only the EPA models were applied, in which most of the hyper-surfaces are closed circle- or sphere-like. However, during our extensive investigations using Monte Carlo simulations (see Sections 3.2 and 3.3), we learned that this is not always true, and open, band-like structures can appear.

In general, the values of the atomic scattering factors are found to influence the curvatures of the isosurfaces. We learned that an anisotropic \mathcal{F} may show cases with an open topology along the direction where the less contributing f_i , *i.e.* the smaller f_i , are assigned. For cases of similar atoms, where all structure factors f_i are alike (*e.g.* EPA model), closed topologies appear. Some of these isosurfaces have been examined and are presented in Figs. 6 and 7. The different shapes result from varying ratios of the atomic scattering factors.

A complete overview of the different isosurfaces \mathcal{F} in \mathcal{P}^2 is realized via keeping f_1 at a constant value of 10 and varying f_2 . As f_2 increases, the curvature of the isosurface changes, resulting in two categories of \mathcal{F} : those that cut both main axes (closed topology) and those intersecting with only one main axis (open topology). Fig. 6 shows that \mathcal{F} bends and tends to intersect both axes z_1 and z_2 as f_2 increases and becomes more similar to f_1 .

A similar study is carried out in \mathcal{P}^3 by assuming different f_i combinations. Fig. 7 summarizes the observed isosurfaces upon varying the ratios of the scattering factors f_i . As in \mathcal{P}^2 , the isosurface in \mathcal{P}^3 may exhibit an open topology along the z_i direction, which is associated with a low scattering factor f_i [*cf.* the open topology along z_3 in Fig. 7(b), or along z_2 and z_3 in Fig. 7(d)]. The isosurface intersects all axes when all f_i are similar.

In order to control the direction where the open topology may appear and to simplify the cases that may occur during the PSC approach, we sort the atoms according to their scattering strength in the initialization step such that the atom

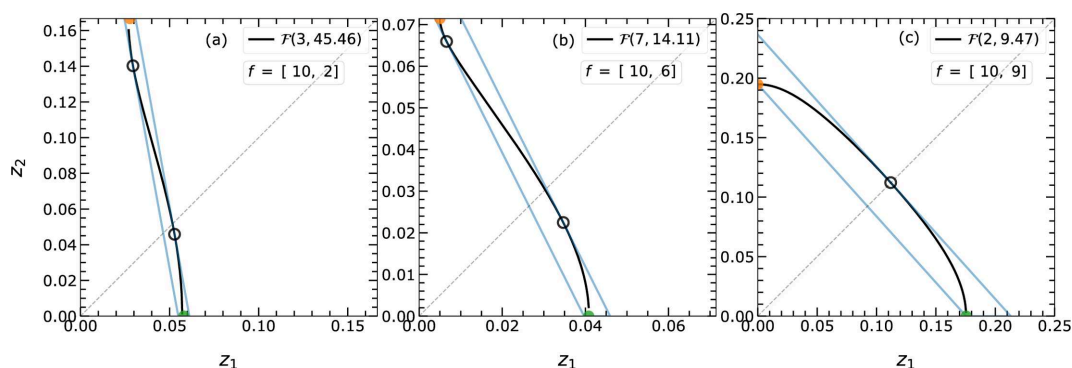


Figure 6 The different possible topologies of isosurface \mathcal{F} in \mathcal{P}^2 on varying atomic scattering factors f_i in equation (2). The \mathcal{F} can be categorized according to their intersection along the z_1 and z_2 directions. The \mathcal{F} in (a) and (b) exclusively intersect the z_1 axis (open topology) and that in (c) intersects both the z_1 and z_2 axes at different distances from the origin (closed topology). The developed algorithm handles and approximates these \mathcal{F} s alike. The boundaries from the approximation are shown by blue lines. The schematic demonstrates the effect of the ratio between f_i on the curvature of \mathcal{F} . The very first segment of the linearization around the origin is shown and all repeated segments are avoided for better visibility. The filled green and orange points are used to get the slope and the found tangent points are shown by open black circles.

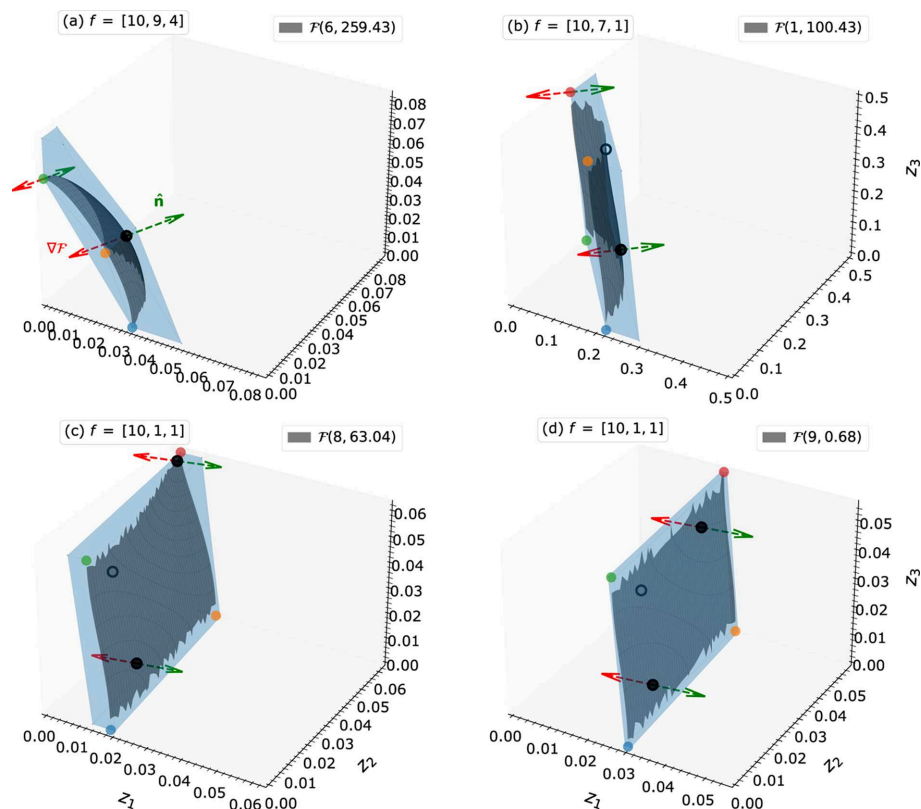


Figure 7
 The different possible topologies of \mathcal{F} in \mathcal{P}^3 on varying atomic scattering factors f_i in equation (2). The \mathcal{F} s are differentiated according to their intersection along the z_i directions. They can intersect (a) with all three axes (closed topology), (b) only with the two axes z_1 and z_2 , or (c) and (d) only with axis z_1 (open topologies). The very first segment of the linearization around the origin is shown and all repeated segments are avoided for better visibility. Also, the computed gradient of \mathcal{F} and the normal vector are depicted at different tangent points. The filled, colored dots are used to define the required normal vector. The tangent points representing planes nearest to and farthest from the origin are represented by solid black dots, while the open black circle indicates an additionally computed tangent point.

with the smallest scattering factor is associated with the highest index of z .

2.4.2. Generating the segments

The linearization of any complete isosurface \mathcal{G} or \mathcal{F} starts with defining a segment in the vicinity of the origin. The concept of linearization can be easily understood in \mathcal{P}^2 and can be systematically extended for higher dimensions. The isosurfaces for a given l have a period of $1/l$, i.e. each isosurface is repeated in each complete PS in translations of $1/l$ along each axis. Hence 1, 1/2 and 1/3 are the periods for $l = 1, 2$ and 3, respectively [cf. Figs. 1(a)–1(c)]. This characteristic period assists in the linearization.

The first task is to find the coordinates \mathbf{z} , where the isosurface intersects with the axes in order to define the inner and outer boundaries. In \mathcal{P}^2 , this coordinate can be determined by rearranging equation (2) to z_2 , while z_1 is set to zero:

$$z_2 = k^{-1} \arccos \left[\frac{s(l)|F_0(l)|}{2f_2} - \frac{f_1 \cos(k0)}{f_2} \right], \quad (7)$$

where $k = 2\pi l$ hereafter. In the case of \mathcal{P}^3 , setting two of z_i (for example z_1 and z_2) to zero yields the remaining z_j ($i \neq j$), e.g.

$$z_3 = k^{-1} \arccos \left[\frac{s(l)|F_0(l)|}{2f_3} - \frac{f_1 \cos(k0)}{f_3} - \frac{f_2 \cos(k0)}{f_3} \right]. \quad (8)$$

These intersections with the main axes are represented by the filled points in Figs. 5(b) and 5(c). The determination of the boundaries differs significantly for closed and open topologies (see Section 2.4.1). Therefore, we will discuss them separately.

Closed topologies. For closed topologies, all intersections with the axes exist and the inner boundary is devised by forming a line (in \mathcal{P}^2) or a plane (in \mathcal{P}^3) from those determined intersection points. The filled points in Fig. 5 define the inner boundary described by a unique normal vector $\hat{\mathbf{n}}$, which assists in determining the outer boundary.

The outer boundary is defined by employing a parallel shift of the inner boundary. Therefore it is crucial to determine the corresponding tangent point (open blue circles in Fig. 5) on the isosurface with respect to the normal of the inner boundary. The line or plane formed at this juncture delineates the outer parallel boundary, marking the point beyond which \mathcal{F} ceases to exist. The existence of this specific tangent point is assured by the mean-value theorem (Hobson, 1909), which – for the general non-EPA case – states that a normalized gradient of the isosurface equivalent to the unit normal exists. With knowledge of $\hat{\mathbf{n}}$ from the inner boundary, a tangent point

can be found by solving $\nabla\mathcal{F}/|\nabla\mathcal{F}| = -\hat{\mathbf{n}}$; for derivations see Section 2.4.3.

Open topologies. For a given set of atoms with varying scattering contributions f_i and a specific intensity or amplitude, the isosurfaces may show an open topology and not intersect with all the axes (see Section 2.4.1). In such cases, the curvature of the isosurface is not convex over the full period but switches from convex to concave in the local vicinity of the open topology. In consequence, lines and planes derived from the intersections (closed circles in Fig. 5) will cut \mathcal{F} . Thus, for these cases, the determined lines/planes must be shifted both towards the origin and away from it to define limiting but not restricting boundaries. Appropriate anchor points for these shifts are defined with a strategy similar to that used for the closed topologies: we search for the tangent points at the isosurface with respect to the normal vector of the line/plane defined by the intersection points. Due to the change between concave and convex behavior, at least two such tangent points exist, which serve as anchor points for inner and outer boundaries, respectively.

Finalizing the segment creation. Once the inner and outer boundaries are defined, the segment can be created using the boundary distances between the origin and the inner and outer boundaries. For a given l , the boundary distances and $\hat{\mathbf{n}}$ are unique and together they provide segment descriptors. This segment linearizes only a part of \mathcal{F} in the vicinity of the origin (cf. Section 2.4.1). The remaining parts of \mathcal{F} are linearized by utilizing the translational and rotational symmetries as described in Section 2.4.5. The routines are generally applicable, including the EPA case with geometric structure factor $G(l)$ and the corresponding isosurfaces $\mathcal{G}[l, g_0(l)]$.

2.4.3. Finding tangent points

In the linear approximation, finding the tangent point is vital as well as critical. To find the tangent point, the least-squares (LS) refinement (Dekking *et al.*, 2005; Lawson & Hanson, 1974) is employed for the isosurface \mathcal{F} [equation (5)]. Within the refinement, the deviation between a desired normal unit vector $\hat{\mathbf{n}}$ and the direction of the gradient of \mathcal{F} is minimized to identify the parallel hyperplane tangent to \mathcal{F} . Illustrated for \mathcal{P}^2 , a selected initial point is shifted along \mathcal{F} to meet the requirement

$$-\hat{\mathbf{n}} = \frac{\nabla\mathcal{F}}{|\nabla\mathcal{F}|} \equiv \nabla\hat{\mathcal{F}}, \quad (9)$$

where $\hat{\mathbf{n}} = (a_1, a_2)$ has the fixed components a_1 and a_2 along the z_1 and z_2 axes, respectively. At the required tangent point, the conditions

$$\nabla\hat{\mathcal{F}} \cdot \hat{\mathbf{n}} = \pm 1 \quad \text{and} \quad |F(l)|^2 - I = 0 \quad (10)$$

must be satisfied, with the sign +1 and -1 of $\nabla\hat{\mathcal{F}} \cdot \hat{\mathbf{n}}$ denoting the convex and concave curvature of \mathcal{F} , respectively. These conditions define the error function to be solved numerically. The points \mathbf{z} where \mathcal{F} cuts the axes [calculated from equation (8), cf. the colored points in Fig. 7] and the point \mathbf{z} along the main diagonal [*i.e.* the z_i values of \mathbf{z} are equal, determined

from $z_i = k^{-1} \arccos[|F_0(l)|/(2 \sum_{j=1}^m f_j)]$ with $s(l) = +1$] are considered as the different initial guesses for the possible tangent points to start the iterative process. Depending on the curvature of \mathcal{F} , each initial guess can yield similar or different tangent points, which are dealt with differently for open and closed topologies.

The isosurfaces with closed topologies have a single convex curvature and will have a single tangent point [cf. Fig. 7(a)]. Isosurfaces with open topologies may also show additional tangent points due to concave-convex curvature change [cf. Figs. 7(b)–7(d)]. These tangent points as well as the intersection points of \mathcal{F} with the axes are used to determine the distance of the outer and inner boundaries. For the assumed case of $f = [10, 7, 1]$, the single open topology results in two different tangent points [cf. Fig. 7(b)]. The tangent point closer to the origin gives the inner boundary distance, and the point further away is used to fix the outer boundary. The isosurfaces with double open topology result in more than two tangent points due to further changes in the curvature near all intersecting points [cf. Figs. 7(c) and 7(d)].

The above formulations are followed for a PS of dimension 3 or higher. However, the parallel boundary lines in \mathcal{P}^2 offer an alternative elegant solution to solve for the tangent point. The normal vector $\hat{\mathbf{n}}$ in \mathcal{P}^2 can be replaced by the slope ζ of the inner (and outer) boundary, which is defined as

$$\zeta = \frac{dz_2}{dz_1} = \frac{d \left\{ k^{-1} \arccos \left[\frac{|F_0(l)|}{2f_2} - \frac{f_1 \cos(kz_1)}{f_2} \right] \right\}}{dz_1}. \quad (11)$$

Using the identity $d[\arccos(x)]/dx = -1/\sqrt{1-x^2}$ and for simplicity applying $s(l) = +1$, the above equation becomes

$$\zeta = \frac{-k^{-1} \left[k \frac{f_1 \sin(kz_1)}{f_2} \right]}{\sqrt{1 - \left[\frac{|F_0(l)|}{2f_2} - \frac{f_1 \cos(kz_1)}{f_2} \right]^2}}. \quad (12)$$

After resolving the root and rearranging the above equation, we obtain

$$(1 - \zeta^2) \left[\frac{f_1}{f_2} \cos(kz_1) \right]^2 + 2\zeta^2 \left[\frac{|F_0(l)|}{2f_2} \right] \left[\frac{f_1}{f_2} \cos(kz_1) \right] + \zeta^2 \left\{ 1 - \left[\frac{|F_0(l)|}{2f_2} \right]^2 \right\} - \left(\frac{f_1}{f_2} \right)^2 = 0. \quad (13)$$

With $\xi = (f_1/f_2) \cos(kz_1)$ the above equation becomes

$$(1 - \zeta^2)\xi^2 + 2\zeta^2 \left[\frac{|F_0(l)|}{2f_2} \right] \xi + \zeta^2 \left\{ 1 - \left[\frac{|F_0(l)|}{2f_2} \right]^2 \right\} - \left(\frac{f_1}{f_2} \right)^2 = 0. \quad (14)$$

Solving this equation for ξ and z_1 on the basis of the identified slope ζ from the inner boundary gives the linear function for the outer boundary as well. In particular, the two roots of equation (14) can probe all possible maxima or minima of \mathcal{F} . If both roots are valid, *i.e.* real and smaller than 1, then \mathcal{F} will have two tangent points [cf. Figs. 6(a)–6(b)], defining at the

Table 1

Comparison of theoretical (from the atomic structure) and calculated (from the tangent point) intensities for the atomic structure [0.349, 0.362, 0.1615].

The found normal vector $\hat{\mathbf{n}}$, the tangent point for the given atomic scattering factors f_i and the deviation angles are also listed. A large value of the difference in angle indicates a larger deviation. The tangent point column lists two sets of coordinates in two rows corresponding to the inner (first row) and outer (second row) boundaries. The respective \mathcal{F} s along with the listed tangent points (filled black points) are given in Fig. 7.

Reflection	f	I		$\hat{\mathbf{n}}$	Tangent points	Deviation angle (°)
		Theoretical	Calculated			
6	[10, 9, 4]	259.427	259.427	[0.680, 0.639, 0.358]	[0.000, 0.000, 0.063]	0.0005
					[0.018, 0.018, 0.027]	0.0005
1	[10, 7, 1]	100.431	100.431	[0.777, 0.626, 0.055]	[0.000, 0.226, 0.500]	0.0005
					[0.144, 0.179, 0.093]	0.0005
8	[10, 1, 1]	63.042	63.042	[0.992, 0.088, 0.088]	[0.002, 0.060, 0.060]	0.0004
					[0.017, 0.015, 0.015]	0.0004
9	[10, 1, 1]	0.68	0.68	[0.996, 0.064, 0.064]	[0.023, 0.044, 0.044]	0.0004
					[0.029, 0.012, 0.012]	0.0004

same time inner and outer boundaries. If only one root is valid or the two roots are identical, then \mathcal{F} will have merely one tangent point defining the outer boundary [cf. Fig. 6(c)].

2.4.4. Precision of the tangent point

Irrespective of topology, the exact linearization of the isosurfaces is important to reduce false structure predictions or the number of pseudo-solutions. The quality of the linearization of \mathcal{F} , particularly in higher dimensions $m \geq 3$, can be inferred from the metrics: (a) the intensity at the tangent point and (b) the angle between the normal $\hat{\mathbf{n}}$ and the gradient $\nabla\hat{\mathcal{F}}$ of the isosurface. The found tangent point must be on the isosurface and hence must result in the same intensity value as that of \mathcal{F} . The incorrect prediction of the tangent point may lead to differences between the structure-immanent solution space and the approximation, which may, on the one hand, unnecessarily increase the volume of the linearization segment and, on the other hand, even more severe, exclude valid solution volume. The angle between $\hat{\mathbf{n}}$ and $\nabla\hat{\mathcal{F}}$ should be 180° , due to the anti-parallel condition of equation (9) and we monitor the discrepancy by the deviation angle. However, inherited from the mixed concave–convex curvatures of isosurfaces with an open topology, the intersecting points of \mathcal{F} on the main axes are no longer co-planar, introducing the need for approximation on defining $\hat{\mathbf{n}}$. We apply the singular value decomposition (SVD) method (Campbell *et al.*, 2021) to determine the initial plane with the respective $\hat{\mathbf{n}}$ from the corner points [e.g. the blue, green, red and orange colored points in Figs. 7(b)–7(d)]. The found $\hat{\mathbf{n}}$ from SVD defines again the tangent point, the open black points in Fig. 7.

In Table 1, both intensity and deviation angle are given to estimate the quality of linearization of respective isosurfaces for the challenging cases shown in Fig. 7. For the given examples, the theoretical intensity (from the atomic structure) is identical to the one calculated from the tangent point; additionally the deviation angle is on the order of 10^{-4} in all cases.

The current implementation still offers significant potential for enhancing the linearization method, specifically in resolving both inner and outer tangent points for cases involving open topologies and mixed concave–convex curvatures in higher-dimensional PS. One such enhancement could involve

the development of a double-segment approach tailored for higher dimensions. Additionally, it is crucial that the determined boundaries fully encapsulate the entirety of the isosurface. To mitigate the risk of overlooking any valid solution spaces, we have introduced a cross-checking routine based on the grid-based method that verifies the complete enclosure of \mathcal{F} by the polytope. In the fail-safe case, the portion of the uncovered isosurface is returned to the momentary solution space, and the outer boundary distance is recalibrated, allowing for the continuation of the linearization process. However, this grid-based approach may demand substantial computational time, especially in higher-dimensional PS. Therefore, the development of a more efficient algorithm that guarantees the complete enclosure of \mathcal{F} while also validating the accuracy of the determined normal vectors remains a priority for future research and development efforts.

2.4.5. Completing the linearization

The segment obtained in the previous step only linearizes a part of the isosurfaces with $s(l) = +1$. The complete \mathcal{P}^m encompasses $2l^m$ copies of this unique segment, due to l -fold mirror symmetry along each z_i direction: l copies in all m directions, maintaining a modulo of $1/l$ [see Figs. 1(a)–1(c)]. Before performing this translational repetition in PS, the formed first segment is rotated around and mirrored about the origin to enclose further parts of the isosurface \mathcal{G} or \mathcal{F} .

The replication of the segment in our code is carried out using the mirror planes that pass through the origin. In \mathcal{P}^2 there are two different mirror planes represented using the vectors [1, 0] and [0, 1] for \mathcal{G} and \mathcal{F} and two additional mirrors perpendicular to the main and secondary diagonals exclusively for \mathcal{G} . The mirror planes represented by the vectors [1, 0, 0], [0, 1, 0] and [0, 0, 1] are used for \mathcal{G} and \mathcal{F} , as well as additional mirrors perpendicular to the main and secondary diagonal exclusively for \mathcal{G} . The segment descriptors are then multiplied by these vectors to form further rotated segments in both closed and open topologies (cf. Fig. 8 as an example for four/eight segments in \mathcal{P}^2). This initial set of equivalent segments is then repeated in PS with translation vectors \mathbf{R} as follows:

$$\mathbf{R} = (\Delta v_1, \dots, \Delta v_m), \quad (15)$$

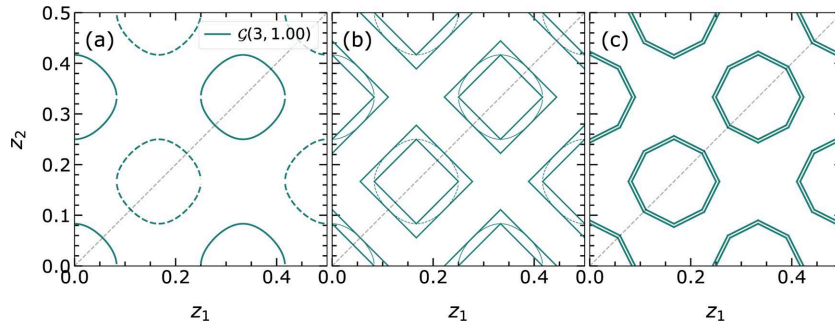


Figure 8
An example \mathcal{G} generated for $l = 3$ with $g(l) = 1.00$ in complete \mathcal{P}^2 . (a) $\mathcal{G}(3, 1.00)$ and their linearization (b) with single-segment and (c) with double-segment approaches. The continuous and dashed lines in (a) represent $s(l) = +1$ and $s(l) = -1$, respectively. They are centered around the points with $+$ or $-$ maximum amplitude.

where Δv_i are the translations along the z_i direction in \mathcal{P}^m , varying between 0 and 0.5 in increments of $1/(2l)$. Those translations are applied only if the centers of polarity condition at maximum amplitude is fulfilled, demanding that all signs of the individual contributions are equal. The centers of polarity condition is defined as

$$[\cos(k\Delta v_1), \dots, \cos(k\Delta v_m)] = \begin{cases} (+1, +1, \dots, +1), & \text{valid,} \\ (-1, -1, \dots, -1), & \text{valid,} \\ \text{any other result} & \text{invalid.} \end{cases} \quad (16)$$

A $+$ ($-$) center is found if all elements in equation (16) are positive (negative). Any other combination is denoted as mixed translation centers and describes positions where there is no presence of maximum or minimum amplitude (cf. Fig. 8). For example, the combination $(\Delta v_1, \Delta v_2) = (0, 0)$ in \mathcal{P}^2 has $+$ polarity, since $[\cos(0), \cos(0)] = [+1, +1]$. The combination $(1/6, 1/6)$ has $-$ polarity because $[\cos(2\pi \times 3 \times 1/6), \cos(2\pi \times 3 \times 1/6)] = [-1, -1]$. However, combinations such as $(1/2, 0)$ have mixed polarity, i.e. $[\cos(2\pi \times 3 \times 1/2), \cos(0)] = [-1, 1]$, and are invalid translation centers in \mathbf{R} construction. Once the reflection l is defined, we can derive the translation vectors \mathbf{R} to ensure proper selection of Δv_m values. These

Δv_m values are then utilized to replicate the segments and enclose the $2l^m$ copies of \mathcal{G} or \mathcal{F} unique segments within \mathcal{P}^m . This iterative procedure effectively completes the linearization process.

2.4.6. Improving the linearization

The explained procedure of linearization for \mathcal{P}^2 and \mathcal{P}^3 involves the definition of a single segment that fully encompasses the isosurface in the vicinity of the origin. However, the defined segment includes a rather large solution space and, thus, has the potential not only to minimize the unwanted solution space but also to suggest a new linearization technique. This can be achieved by increasing the number of used segments. We will refer to this approach as the double-segment linear approximation. Figs. 5(a) and 5(b) depict a comparison of these different linearization variations in \mathcal{P}^2 for closed topologies.

The transition from the single- to the double-segment approximation is achieved by dividing the PS into two parts in the most advantageous way: at the main diagonal, where the components of \mathbf{z} are identical. The intersection of the isosurface with the main diagonal, which is represented by the filled orange point in Fig. 5(a), i.e. $z_1 = z_2 = k^{-1} \arccos[s(l)g(l)/\sum f_i]$, serves as the first point to determine

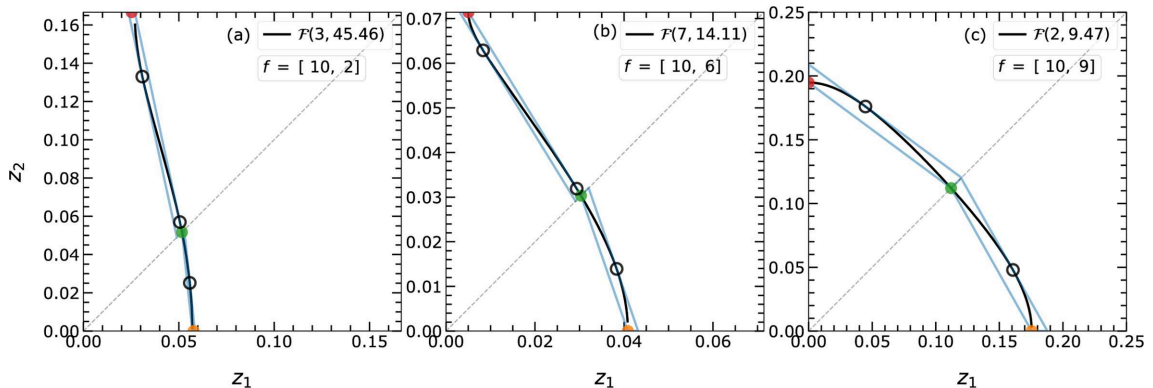


Figure 9
Application of the double-segment approach to the isosurfaces from Fig. 6. The filled green–orange (green–red) points are used to calculate the slope for the part of \mathcal{F} below (above) the diagonal line. The found tangent points are shown by open black circles. When the part of \mathcal{F} has a concave–convex mixed curvature, it possesses multiple tangent points for a given slope [cf. (a) and (b)] and equation (14) is capable of exploring all points in a single analysis. In the case of multiple tangent points, all points are used to define inner and outer boundaries.

the inner boundaries of the two desired segments. Then setting z_1 or z_2 to zero results in the respective z_2 or z_1 component of the second point [filled green points in Fig. 5(a)]. The two straight lines formed from these two points with the orange point define the inner and outer boundaries, similarly to the single-segment approach. Following the same methodology, the double-segment approximation is successfully developed also for non-EPA, only in \mathcal{P}^2 PS. As depicted in Fig. 9, the isosurfaces from Fig. 6 are approximated using double segments.

This improvement in linearization results in a significantly reduced solution space at the cost of additional computational load for handling a larger number of polytopes (*cf.* Sections 3.2.1 and 3.4.1). However, the variation in the approximation is exclusively developed for \mathcal{P}^2 at the moment, while the single-segment linearization is employed for three- or higher-dimensional PS by default. Developing optimized variations other than single-segment linearization in higher-dimensional PS is an important future goal and will be considered in future work.

2.5. Intersection of linearized isosurfaces: the solution

Once the linearization is completed, the polytopes are created by collecting all segments for a given reflection to search for the solution region. The set of polytopes contains 2^l segments for a given reflection l . In principle, exactly one of the manifolds of an isosurface of reflection l always contains the crystal's structure solution, and therefore the corresponding segments also do. In turn, we need to find that region that is common to all the isosurfaces of different reflections; this process corresponds to determining the intersection between the polytopes.

The solution is found systematically through the following procedure. The solution space is reduced by intersecting the polytopes of the different reflections. Hence, with the addition of more reflections to the calculation, the solution space and the positional errors are in general gradually decreasing. This process is repeated until the last available reflection. Often, the final solution region consists of several very small, disjoint areas that form clusters. Subsequently, the final solution regions are used to calculate the coordinates of atoms in the structure of interest; these coordinates are presented as a list of atomic positions with their errors, as usual in crystallographic information file format. To determine the atomic position and the error in the analysis, we utilize the centroid of the solution regions and the extension in the z_i directions. The extension of each component can be quantified as

$$\Delta z_i = z_i^{\max} - z_i^{\min} \quad (17)$$

using the minimum z_i^{\min} and maximum z_i^{\max} values. These set operations are carried out by means of the *polytope* (Filippidis *et al.*, 2016) and *shapely* packages (Gillies *et al.*, 2022). The polytopes shown in all figures are generated with the *IntvalPy* package (Androsov & Shary, 2022).

3. Results

In the upcoming sections, we present an introductory example (Section 3.1) and the results of Monte Carlo (MC) simulations of PSC to scrutinize the theory's performance through a large number of simulations in \mathcal{P}^2 (Section 3.2) and \mathcal{P}^3 (Section 3.3). The examples will cover EPA as well as non-EPA calculations as the variation of the structure factor can have a severe impact on the shape of the isosurface and thus on the linearization process and the validity of a determined solution concerning real data.

The MC simulations are analyzed to gain insights into the overall solution landscape. The volume of the solution regions and the error are monitored in each addition of reflections to pinpoint any artifacts. For the given coordinate, we calculate and sum the volumes of all obtained solution regions as simple performance descriptors. This cumulative volume is then used to construct a virtual, representative, m -dimensional sphere. Ideally, we receive a singular final solution identified within the overall solution landscape. However, in the case of numerous solutions, all are equally probable within the limits of linearized polytope regions, highlighting the significance of considering all possible solutions in the analysis.

3.1. Explanatory example in \mathcal{P}^2

The effectiveness of the linear approximation is evaluated by solving an example structure consisting of two atoms, with coordinates $z_1 = 0.151$ and $z_2 = 0.138$. The structure is solved by considering the reflections 1 to 4 whose intensity is converted into the associated \mathcal{G} within the EPA framework, *i.e.* the scattering factor of the two atoms is set to 1. Figs. 10(a) and 10(b) illustrate the superimposition of all \mathcal{G} and their linear approximations, respectively.

The obtained isosurfaces \mathcal{G} exhibit a predominantly smooth curvature and have a nearly circular topology in \mathcal{P}^2 [Fig. 10(a)]. The behavior is trivial due to the application of the EPA model for intensities. The determined solution in \mathcal{A}^2 is $(0.1489 \pm 0.0021, 0.1397 \pm 0.0022)$, shown in Fig. 10(c). The error is on the order of $\sim 10^{-3}$, which can be reduced further by increasing the number of reflections (for example, *cf.* Section 3.2.1). The solution is unique in \mathcal{A}^2 [as shown in Fig. 10(c)]; however, the mirror symmetries along the main diagonal lead to three additional, but equivalent, solution regions in \mathcal{P}^2 .

If experimental errors are considered, they enlarge the width of the linearization polytopes and further distinct solution regions may appear (Zschornak *et al.*, 2024). The results are summarized in Fig. 10(d) with the centroid (green circle), area (orange circle) and extent of the solution region (cross). Including more reflections in the calculation may merge the green and orange circles, reducing error and further decreasing the area of the solution region. The schematic as in Fig. 10(d) is used to outline the results from the MC simulation in the sections below.

3.2. MC simulations in \mathcal{P}^2

The performance of the developed linearization technique and the accuracy of the PSC approach are analyzed through

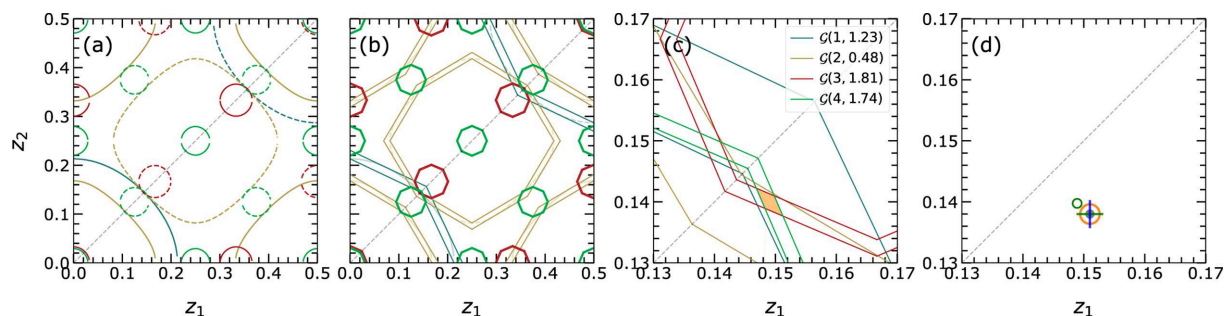


Figure 10

As a test case the coordinates $z_1 = 0.151$ and $z_2 = 0.138$ are solved by double-segment linear approximation and using the first four reflections for the EPA case and intensities with (a) the \mathcal{G} for $l = 1, \dots, 4$, (b) the boundaries of linearization repeated in complete \mathcal{P}^2 , and (c) the common intersection region from the linearized polytope with a total area of 7.99×10^{-6} . The dashed gray lines denote the mirror symmetry in \mathcal{P}^2 for EPA cases. The green open circle in (d) denotes the centroid of the solution region [the shaded area shown in (c)] which is predicted to be $(z_1, z_2) = (0.1489 \pm 0.0021, 0.1397 \pm 0.0022)$. The green and blue bars represent the calculated errors in z_1 and z_2 and are attached to the test structure's assumed coordinates for reasons of comparability. A radius corresponding to the polytope area is calculated and represented as an open orange circle in (d).

MC simulations. Within all PSC routine variations, the same set of random atomic coordinates has been used to ensure comparability. The trend of area or volume (in \mathcal{P}^2 or \mathcal{P}^3 , respectively) and extension of the solution region is monitored to track the deviations in structure prediction and identify weaknesses of the routines. Finally, the computing time is measured and presented for each MC simulation. The results are presented in the same style as in Fig. 10(d) with an additional open yellow circle highlighting the total number of solutions. This yellow circle is centered in the same way as the orange one and its thickness changes according to the number of identified total solutions. Hence, the thicker the yellow circles, the greater the number of solutions obtained. Also, timings and code performance for the MC simulations are evaluated in detail in Section 3.4.1.

3.2.1. MC simulation within the EPA framework

At first, randomly generated coordinates are solved using the EPA model, *i.e.* the atomic scattering factors $f_i = 1$. Since the sign and magnitude of the amplitude \mathcal{G} are calculated explicitly, the given atomic structures can be solved using the following two different approaches. For the amplitude approach [$2g(l), |F_0(l)|$], both sign and magnitude are utilized, and only the polytopes corresponding to the correct sign are considered in the calculation. In contrast, for the intensity approach [$|2g_0(l)|^2, |F_0(l)|^2$], all polytopes are considered without knowledge of the sign. The amplitude approach is particularly beneficial when the experimental data contain the sign.

Figs. 11(a) and 11(b) illustrate the progression of the intersection region using both single- and double-segment approaches. As more reflections are added incrementally the positional uncertainties and the area of the solution region are continuously decreased. Fig. 11 shows that the double-segment linearization reduces the error more than the single-segment linearization. The double-segment approach provides higher resolution by a factor of 2.7 with just the first two reflections and 3.6 with eight reflections. The main reason is that the polytopes generated in the single-segment approach

include a larger area; thus the intersection regions tend to remain larger than those from the double-segment approach. Consequently, the positional uncertainties on the computed z_1 and z_2 are higher in single segments. However, the area and positional uncertainties of the solution region exhibit steady improvement as more reflections are considered. Remarkably, with only four reflections, the double-segment linearization reduces the average area of the solution region below 10^{-3} , while the single-segment linearization achieves a comparable result after considering as many as eight reflections. These findings highlight the advantages of employing multiple segments in enhancing the accuracy and precision of the computed atomic coordinates.

The amplitude and intensity approaches solve the coordinates, resulting in a similar real solution. In most cases, the amplitude approach results in a unique solution in \mathcal{P}^2 . However, the intensity approach produces many equivalent mirror solutions due to ambiguity in the sign [*cf.* the difference in thickness of the yellow circles between Figs. 11(a) and 11(b)]. Hence, as known from conventional X-ray diffraction refinements, it is beneficial to measure both magnitude and sign for a unique structure prediction.

3.2.2. MC simulation within the non-EPA framework

The interesting question for realistic structures and diffraction data is to investigate the applicability of PSC for realistic atomic scattering factors f_i , which influence the solution-finding process significantly. The results including the heavy-light atom combination $f = [10, 2]$ as well as the combination of similarly weighted atoms $f = [10, 9]$ are presented in Fig. 12. For all the non-EPA cases, we focus exclusively on the more complex intensity approach, since it is the general case of experimental diffraction data with a higher multiplicity of solution regions and thus more demanding for the code.

The area of the solution region is set as the common scale across all subfigures in order to compare the effect of increasing the number of reflections in the calculation. Fig. 12

demonstrates the crucial role of the f_i ratios in defining the size of the solution region. The different f_i cause anisotropically deformed \mathcal{F} (cf. Fig. 6). In some cases, the linearization of such deformed \mathcal{F} for a given l may contain a significantly larger area than that from other l . This is directly reflected in the count of observed solution regions, denoted by yellow circles in Fig. 12. Irrespective of these effects, the area of the solution region is again continuously reduced by increasing the reflection index in the calculations.

The solutions of the calculations with $l \leq 2$ exhibit large positional uncertainties and areas, as indicated by the size of the orange circles in Fig. 12. When considering up to eight reflections, all atomic coordinates are successfully determined with positional uncertainties below 10^{-3} regardless of the specific combinations of f_i . Additionally, Fig. 12 indicates that the PSC implementation effectively handles any possible combination of f_i in \mathcal{P}^2 . Overall, the findings demonstrate the robustness and reliability of the PSC method in accurately determining atomic coordinates, even when dealing with diverse f_i combinations.

3.3. MC simulations in \mathcal{P}^3

Following the MC simulations in \mathcal{P}^2 , the analogous investigations are carried out in the three-dimensional PS \mathcal{P}^3 . The polytopes are analyzed with their volume and extension in the z_1 , z_2 and z_3 directions. Again, the solutions identified by the code are counted and their volumes are summed to construct a virtual sphere. So far, no double-segment linearization has been established for \mathcal{P}^3 , and thus we focus only on single-segment linearization. Timings and code performance for the MC simulations are evaluated in detail in Section 3.4.2.

3.3.1. MC simulations within the EPA framework

Fig. 13 presents the results of the MC simulations in \mathcal{P}^3 within the EPA framework, which uses both amplitude and intensity approaches. The conclusions drawn from the \mathcal{P}^2 case apply similarly to \mathcal{P}^3 . Since the single-segment linearization includes a volume around the isosurface, the intersection process potentially reveals more than one solution. By gradually

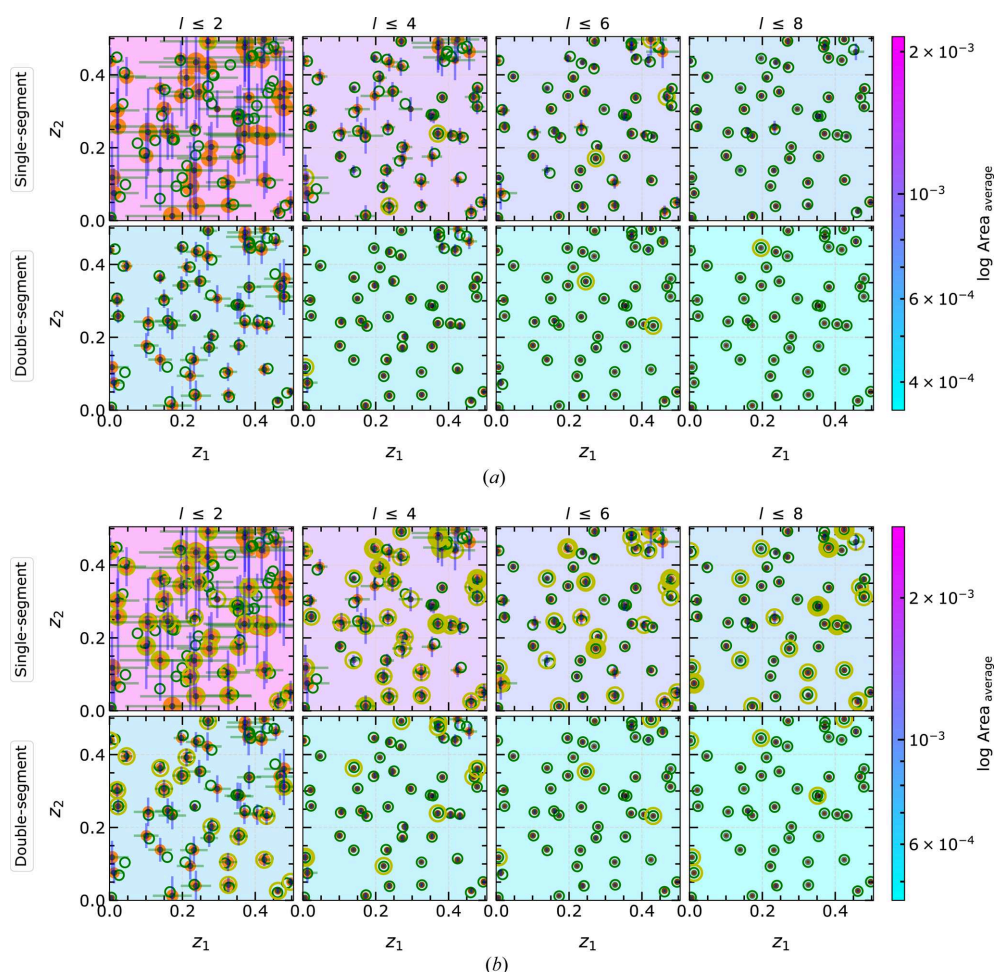


Figure 11

Results of MC simulation using single- and double-segment linearization with amplitude (a) and intensity (b) approaches. The green circles surround the centroid of the solution region within the EPA framework. When more than one solution region is available, the number is represented by the yellow circle (increased thickness for a larger number, cf. Fig. 15). The green and blue bars are the positional uncertainties on the computed z_1 and z_2 , respectively. The area of all solution regions is summed to calculate the radius of the virtual circle as $\sqrt{\text{area}/\pi}$, shown by the orange circle. The background color is defined by the average area of the solution regions of all random pairs, normalized to the overall minimum and maximum for all presented settings.

adding further reflections, the positional uncertainties and the size of the solution regions are progressively reduced.

Our findings demonstrate that our code efficiently reproduces the given coordinates for the three structural degrees of freedom, regardless of the specific set of structural positions. Again, as observed in \mathcal{P}^2 , the intensity approach results in more possible solutions than the amplitude approach due to

the ambiguity in the sign, and only a few solutions are identified uniquely.

3.3.2. MC simulation within the non-EPA framework

The MC simulations are repeated for general f_i values (non-EPA) in \mathcal{P}^3 . As in \mathcal{P}^2 non-EPA, different combinations for atomic scattering factors are considered in \mathcal{P}^3 , covering the

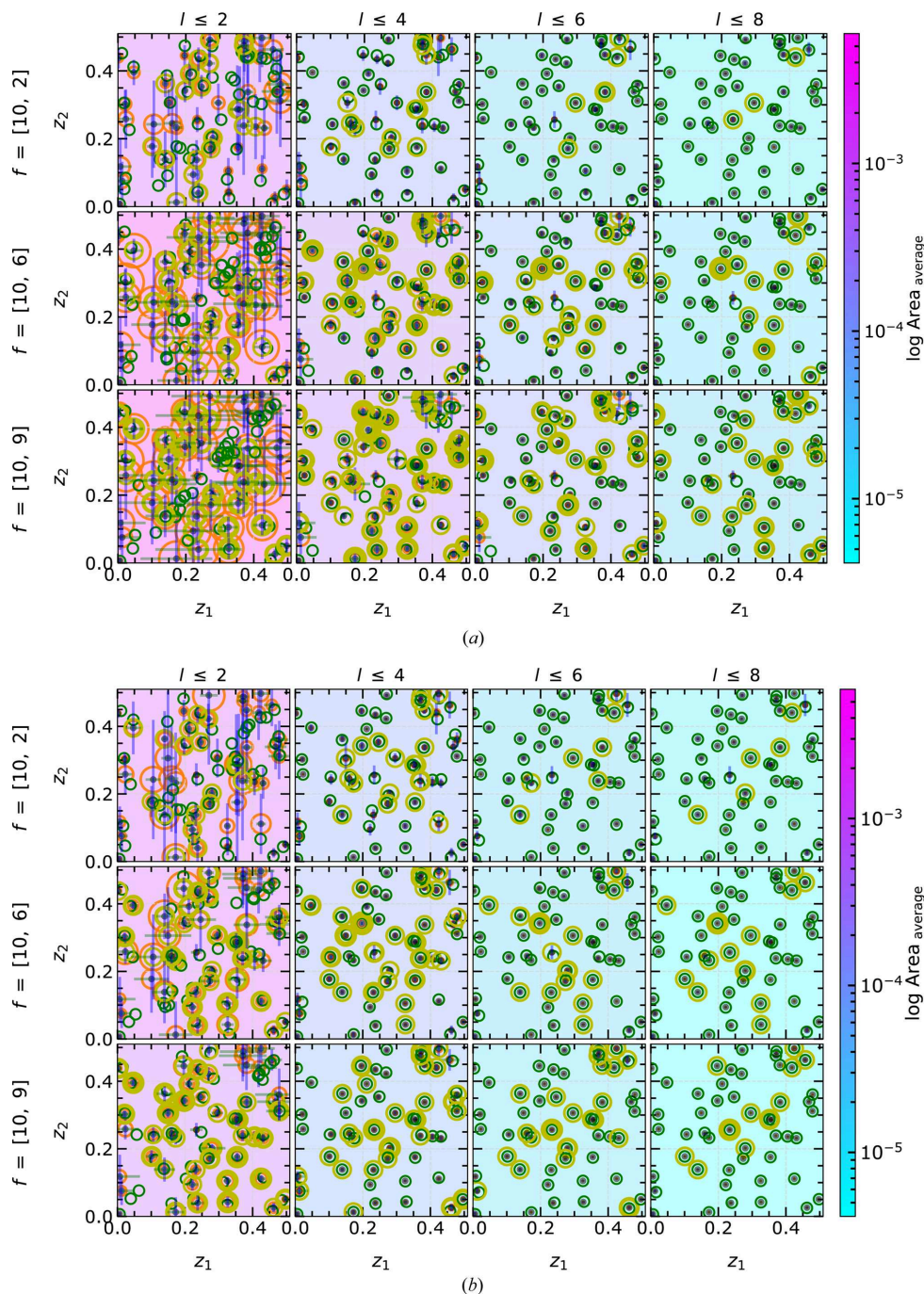


Figure 12 Solved atomic coordinates for (a) single- and (b) double-segment approaches using intensities within the non-EPA framework. The atomic scattering factors $[f_1, f_2]$ in equation (2) are set to [10, 2] (top row), [10, 6] (middle row) and [10, 9] (bottom row) to represent heavy–light, heavy–medium and similar atom combinations along the z_1 and z_2 directions. The results are shown for the increasing number of reflections involved in the solution-finding process. The black dots, green circles and yellow circles (increased thickness for a larger number, cf. Fig. 16) represent the generated random atomic coordinates, the centroid of the solution region and the number of solution regions obtained, respectively.

three scenarios of differently weighted atoms: heavy–light–light, heavy–medium–light and heavy–heavy–light. We fixed f_1 and f_3 to 10 and 1, respectively, to represent heavy and light atoms, and only changed the contribution of f_2 . These f_i

combinations are treated within the intensity approach. For comparability of the results, the initial atomic coordinates are kept identical for all scenarios; only the scattering factors f_i are varied (see Fig. 14).

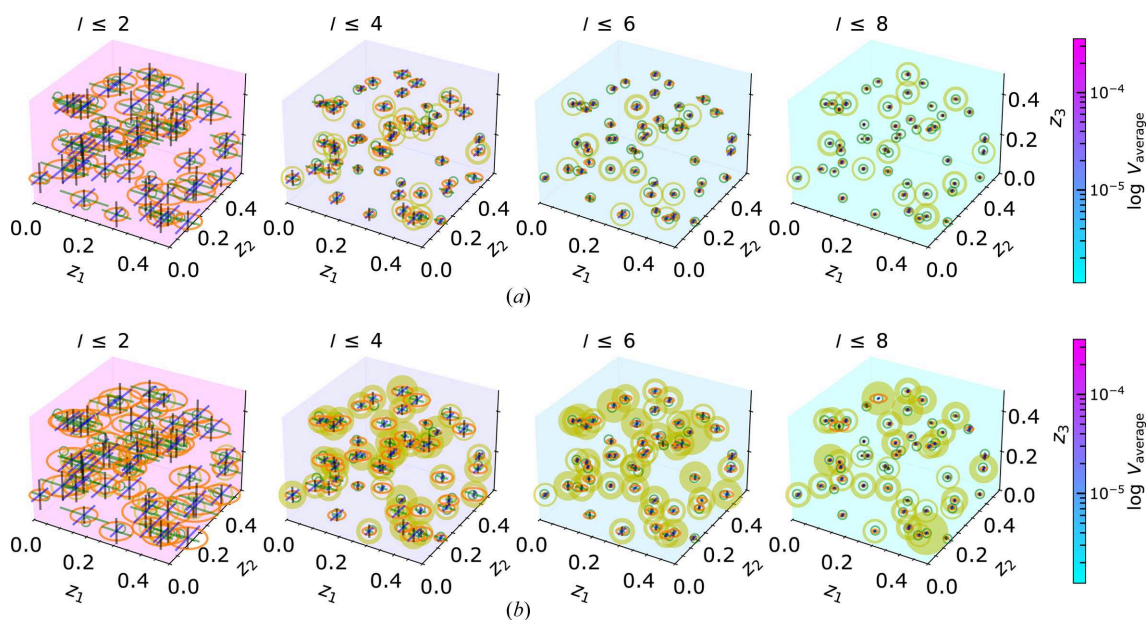


Figure 13 Results of MC simulations for $m = 3$ within the EPA framework. The generated atomic coordinates are solved using amplitude (a) and intensity (b). The green circle represents the centroid of the polytope that encompasses the given atomic coordinates. The cross on each point indicates the error along each z_i direction. When more than one solution region is available, the number is represented by the yellow circle (increased thickness for a larger number, cf. Figs. 17 and 18). The volume of all found solutions V is summed to calculate the radius of the virtual sphere as $\mathcal{R} = \sqrt[3]{3V/(4\pi)}$, which is represented by orange circles. The background color represents the average polytope volume of all solution regions of all considered coordinates.

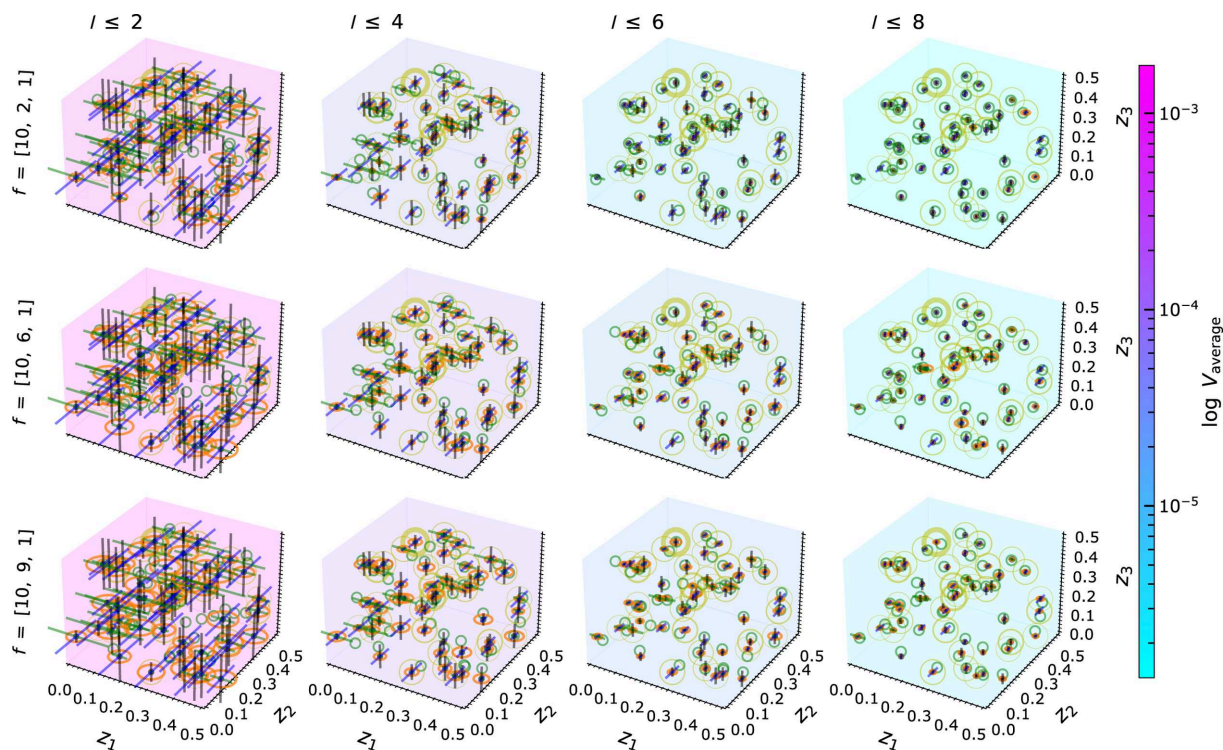


Figure 14 Results of MC simulations for $m = 3$ within the non-EPA framework with the intensity approach. The structures are solved for different settings of f_i . For the definition of different colors and symbols see the caption of Fig. 13.

For each scenario of f_i combinations, the expected reduction in volume and positional uncertainty of the solution region is visible, in analogy to the \mathcal{P}^2 case. The variation in the f_i ratios affects mainly the number of solutions, as shown by the size of the virtual sphere in Fig. 14. The curvature of \mathcal{F} varies strongly depending on f_i values. These multiple solutions can be further minimized or eliminated by increasing the number of reflections in the calculation and by reducing the linearization volume. If the newly considered reflection does not result in a polytope with a smaller intersection volume than the previous reflection, the intersection process will yield the same outcome as before. The presented non-EPA results demonstrate the capability of PSC in generally handling crystal systems with three structural degrees of freedom and different f_i combinations, which can be utilized further for realistic diffraction data.

3.4. Timing benchmark

As explained in Section 2.3, the structure determination process consists of four distinct steps: initialization, linearization, intersecting and writing. We monitor the time consumption of each step to benchmark the performance of our developed algorithm. The initialization step consumes a small amount of time, taking less than a millisecond. During

the linearization step, we identify the first segment, which is then repeated in the complete $\mathcal{P}^2/\mathcal{P}^3$ with the constraint given by equation (16), as explained in Section 2.4.5. This step consumes significantly more time. The details of the first segment are stored in a variable for later purposes. To visualize the dimensional scaling, two timings are separately captured, the time for linearization $t_{\text{Linearization}}$ and the time required to fill the complete PS with the first segment t_{Polytope} for each reflection l .

After linearizing and filling the PS, the subsequent intersection step is carried out to find the solution region; this step represents the most time-consuming part of the structure determination process, measured by $t_{\text{Intersection}}$. At the end, the routine creates an HDF file and writes the information about the processed reflections l , the first segments, found solutions, the error on each solution and the volume of each solution within the time t_{Writing} . In the case of MC simulations, we additionally write the information about the generated artificial atomic coordinates and the exact solution region that encloses the given structure. The total time t_{total} for the structure determination processes includes all four contributions separately for an increasing number of considered reflections l . The individual times of the MC simulation in \mathcal{P}^2 and \mathcal{P}^3 are analyzed in detail in the supporting information (Sections S3 to S6); below we only give t_{total} .

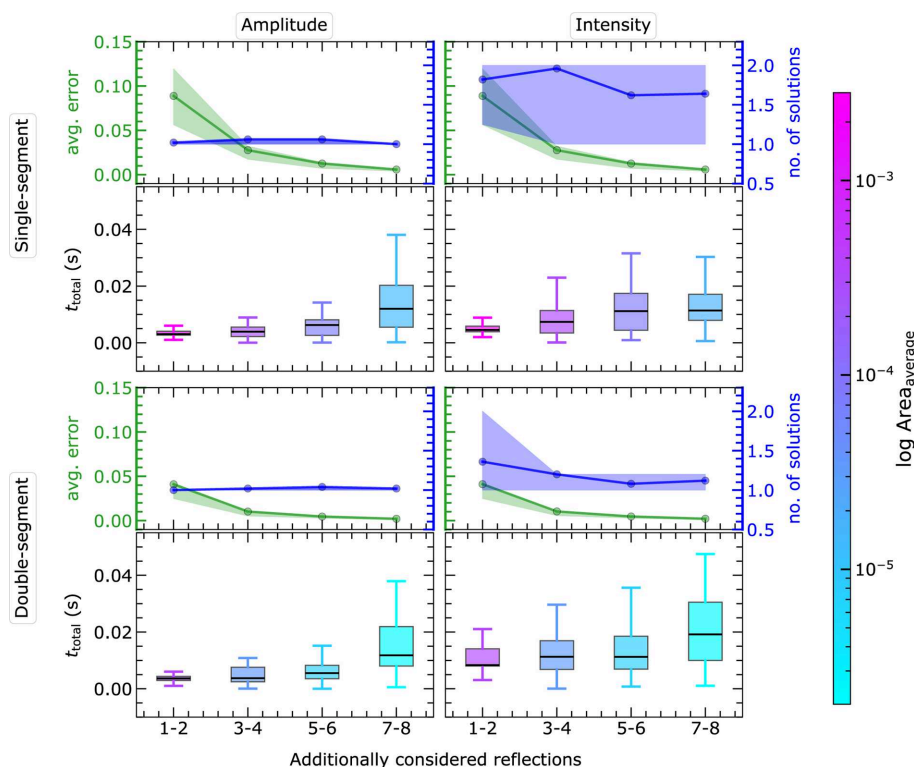


Figure 15

Incremental increases of t_{total} consumed for MC simulation within the EPA framework in \mathcal{P}^2 , including time for linearization, polytope creation, repeating in the complete PS, intersection between successive reflections and writing found solution details to an HDF file. The timing information is presented as boxplots, highlighting the median (black horizontal line), quartiles (ends of the color box) and whiskers (here, complete data range) (for details see Section S1). In addition, the average number of solutions and average maximum possible error on computed \mathbf{z} coordinates are also summarized. These two characteristics are described by the median (circle) and the whisker position (envelope). The color of the boxes represents the calculated average area of the solution regions, which can be compared with the color code given in Fig. 11. The time information for the individual process is analyzed in detail in the supporting information (see Sections S3.1 and S3.2).

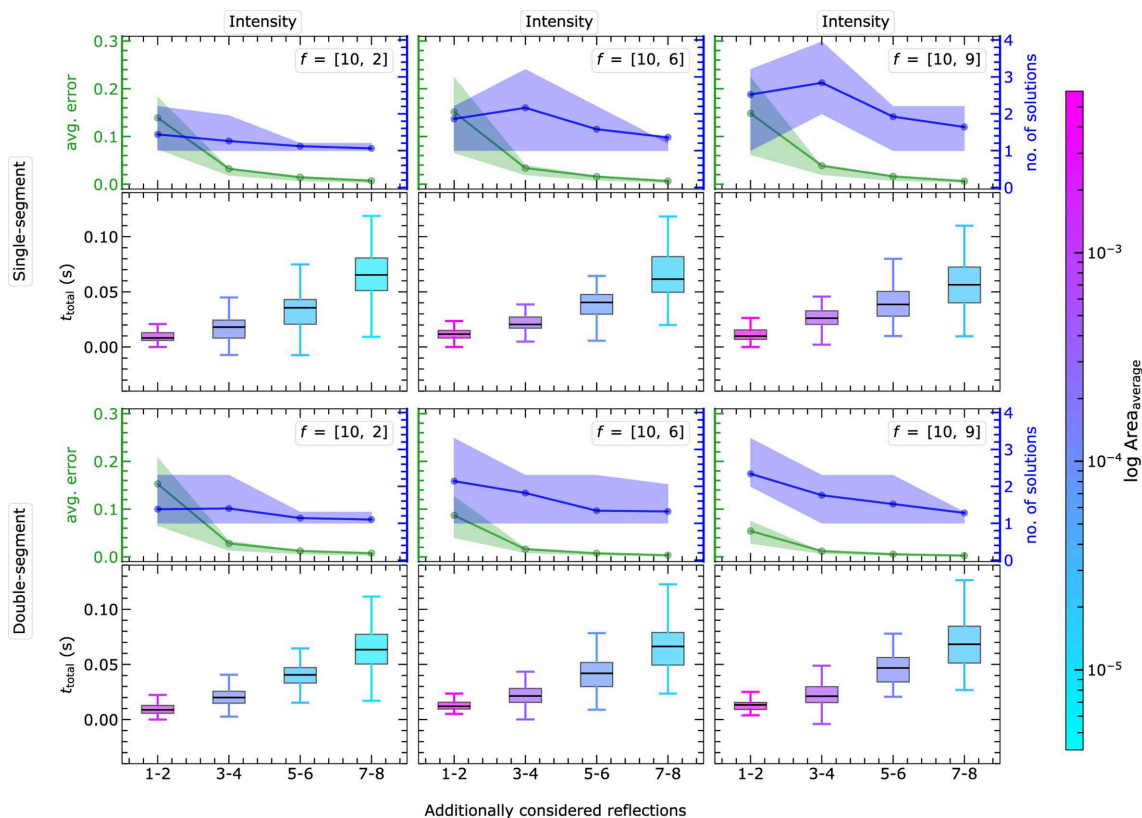


Figure 16 Incremental increases of t_{total} consumed for MC simulation within the non-EPA framework in \mathcal{P}^2 . The color of the boxes represents the calculated average area of solution regions which can be compared with the color code given in Fig. 12. For further explanation see the caption of Fig. 15. The time information for the individual process is analyzed in detail in the supporting information (see Sections S4.1 and S4.2).

Along with t_{total} , the resulting average number of solutions and maximum error on \mathbf{z} as defined in equation (17) are presented using a box plot analysis (see Section S1 for more details). Here, the average error on \mathbf{z} is obtained from the error on individual components, *i.e.* $\text{avg}(\Delta z_1, \Delta z_2, \dots, \Delta z_m)$, and subsequent averaging over all MC instances. In addition, the area/volume of the solution region is given by the

bar. For the purpose of readability, we display the timings for adding two consecutive reflections l .

3.4.1. Timing benchmark for two-dimensional MC simulation

The MC results show the performance of the developed code and algorithm with respect to an increasing number of considered reflections. The individual timings are analyzed for

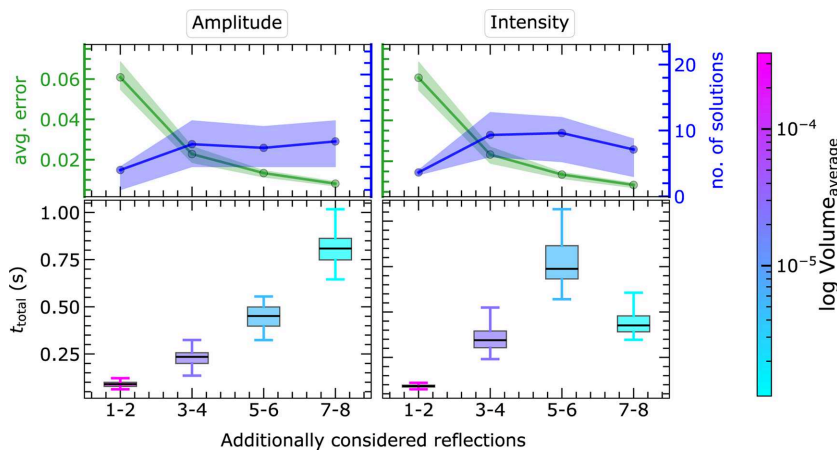


Figure 17 Incremental increases of t_{total} consumed in PSC simulation within the EPA framework in \mathcal{P}^3 , including linearization, polytope creation and repeating in the complete PS, intersection between successive reflections, and writing found solution details to an HDF file. The time information for the individual process is analyzed in detail in the supporting information (see Section S5).

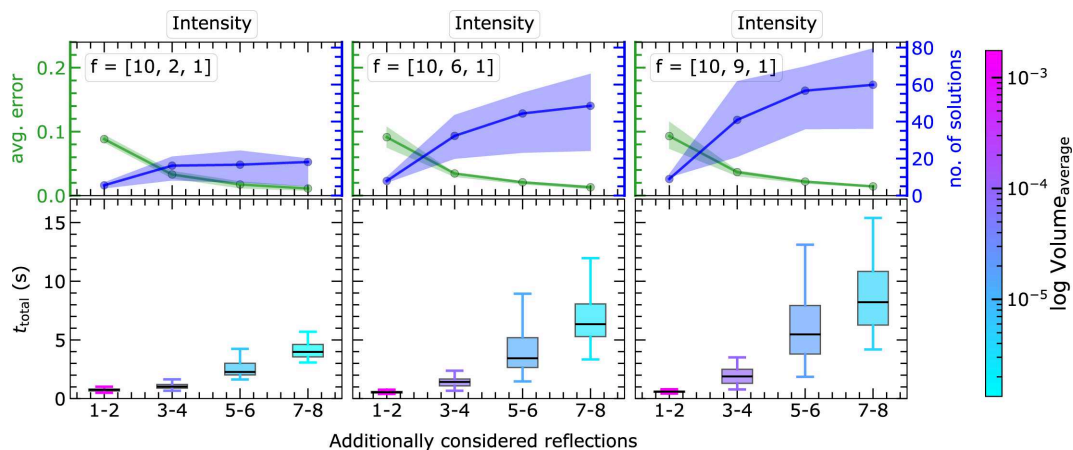


Figure 18 Incremental increases of t_{total} consumed in PSC simulation under the non-EPA framework in \mathcal{P}^3 , including linearization, polytope creation and repeating in the complete PS, intersection between successive reflections, and writing found solution details to an HDF file. The time information for the individual process is analyzed in detail in the supporting information (see Section S6).

the MC simulation in \mathcal{P}^2 and presented in the supporting information (Sections S3 to S4) for both intensity and amplitude approaches within the EPA and non-EPA frameworks.

As presented in Figs. 15 and 16 for EPA and non-EPA frameworks, a maximum of 12 and 200 ms, respectively, is spent solving a structure with eight reflections irrespective of f_i combinations. These timing observations are the result of executing the PSC code in the serial configuration. By parallelization of PSC routines, particularly the routine to fill the complete PS with segments, the time consumption can be severely reduced. Parallelization can be easily implemented in the future.

Further, the time taken for each step in the code reveals that the two parts of completing the segments in the PS and the intersection process dominate (*cf.* Sections S3 to S4). The repetition of the segment (Section 2.4.5) can also be parallelized, which is a future task and has not yet been implemented in the code. The computational time consumption is expected to increase following the same trends in higher dimensions.

3.4.2. Timing benchmark for three-dimensional MC simulation

As already observed for the two-dimensional cases, the repetition of segments through symmetry application and the intersection of polytopes consume more time than the linear approximation and the writing of solution details (see Sections S5 and S6). The total time for the entire process is given in Fig. 17 for the EPA framework. The structure solution within the EPA model using eight reflections takes a maximum of 10 and 2 s for the intensity and amplitude approach, respectively. The intensity approach is slower as the number of polytopes is significantly larger than for the amplitude approach.

In the case of the non-EPA framework, *cf.* Fig. 18, t_{total} strongly varies depending on the individual structure to be solved, which is visible from the outliers of the boxplots (https://en.wikipedia.org/wiki/Box_plot). Due to the open topology, the number of possible solutions generally increases,

and hence t_{total} increases. Also, the computational workload increases continuously from the heavy–light–light to the heavy–heavy–light configuration.

In certain cases, the number of solutions may decrease when adding more reflections in the calculation, which may reduce the computational load. Again, for each considered structure, we observed all possible solutions in one go, *cf.* the orange spheres in Fig. 14. Hence, the PSC routines are robust for the different f_i combinations in \mathcal{P}^3 .

4. Conclusion

Thanks to the advancements in computational resources, we are now able to apply and implement the PSC approach proposed and developed by Fischer *et al.* within the past 15 years. In the presented work, PSC has been enhanced to handle a broad spectrum of combinations of atomic scattering factors, making it suitable for realistic X-ray data analysis. In this study, a concrete workflow is developed to initialize the obtained experimental/theoretical data, linearize the amplitude or intensity, span the PS with polytopes, carry out the intersection process, and perform the solution-finding routine. A stable algorithm has been developed to linearize the amplitudes and intensities under EPA and non-EPA schemes. It is observed that the developed algorithm can handle the open topologies in isosurfaces for all general cases. So far, the developed program effectively handles the structures in \mathcal{P}^2 and \mathcal{P}^3 .

The linearization of the isosurfaces starts with defining the inner boundary using either the intersecting points of the isosurface with each axis for closed topologies or the period for open topologies. These intersecting points are utilized by invoking the SVD method to find the required normal vector $\hat{\mathbf{n}}$. Then the respective tangent points on the isosurface are found by solving the parallel condition $\nabla \hat{\mathcal{F}} \cdot \hat{\mathbf{n}} = \pm 1$ numerically using the LS method. The found tangent points are used to calculate the distance from the inner and outer boundaries to the origin. This implementation facilitates the general-

ization of PSC as well as the computational scaling of this step for the m -dimensional PS. In the next step, the required information (normal vector and boundary distances) can be converted into polytopes using efficient third-party Python libraries. The intersection process is also generalized to handle polytopes of any dimension. This step is carried out sequentially between successive reflections to scrutinize the feasible solution space for the given intensity or amplitude information.

The implemented PSC routines have been consecutively tested by employing MC simulations. Artificial atomic structures have been generated randomly and treated with EPA and non-EPA models. A comprehensive analysis is conducted in both \mathcal{P}^2 and \mathcal{P}^3 , allowing not only for a visual exploration of the combined effects of different atomic scattering factors but also for qualitative analysis by evaluating the average values of area/volume of the polytopes and errors on the computed structures for each simulation step. A key observation from the simulations is that the curvature of the isosurface (*cf.* EPA versus non-EPA cases) is predominantly decreased for heavy–light atom combinations and for reflections with small structure amplitudes.

The presented results show that, with a limited number of reflections, a solution space with volume as small as 10^{-6} and extension on each structural degree of freedom z_i of the order of 10^{-4} can be computed. A further advantage of PSC is that all possible structure solutions in accordance with the diffraction data appear in the PSC-determined solution volume.

In summary, the derived linearization-based PSC approach presents the means to solve crystal structures with equivalent and non-equivalent scattering factors with up to $m = 3$ degrees of freedom in the complete PS, to handle experimental diffraction data, and to explore all possible solutions in a single analysis. Additionally, the integration of third-party Python libraries for further data post-processing introduces new strategies for determining unknown structure parameters, offering an alternative to conventional FI refinement. The routines presented here as well as informative examples are publicly available on GitHub (Vallinayagam *et al.*, 2024).

5. Outlook

For the generalization of the PSC routines to an m -dimensional PS, efficient data handling becomes crucial and calls for further development of the current methodologies. The current centrosymmetric constraint is expected to be resolved in the upcoming phase of PSC development, which should allow a further advancement in PSC code applicability.

Overall, PSC opens up new possibilities for improving the resolution and accuracy of atomic structure solutions employing resonant contrast (Zschornak *et al.*, 2024). The comparison between single- and double-segment linearization shows that precise linearization would reduce the error from linearization approximation at the cost of computational effort and the number of polytopes handled. Therefore, further development of the linearization schemes is needed.

In addition to the aforementioned challenges, the technical implementation of set theory operation critically requires major computation time. The algorithm for the intersection of polytopes is kept in step-wise operation, recursively reducing the PS with each additional reflection l . The themes for the next improvements are a sound parallel algorithm and innovative strategies to save the user time in handling the data. The sequence of processing the reflections may provide options to achieve the intended accuracy on atomic coordinates faster, and the search for such a processing scheme is under way. Further enhancements will increase the possible degrees of freedom and aid in handling non-centrosymmetric structures.

The implemented linearization algorithms can in principle also handle chemically complex or disordered structures, provided they maintain centrosymmetry. Point defects, including substitutional and vacancy defects, may be managed efficiently by the linearization method through variations in site occupancy and atomic scattering amplitudes [*cf.* equation (2) and Figs. 12 and 14 for the effects on structure determination due to varying atomic scattering amplitudes]. Interstitial defects can be addressed by treating the defect site as an additional degree of freedom. However, displacive and dynamic disorder requires consideration of a span of atomic positions, presenting more challenges that are beyond the scope of the current PSC routines. Well ordered extended defects may be modeled using supercells with rational modulation vectors, requiring further degrees of freedom depending on the structural details. Overall, addressing these defect scenarios will be a task for future work and is not part of the present study.

Moreover, predicting the exact structure from multiple possible solutions is an open issue. A significant number of non-unique solutions will disappear once the sign of intensity is known and utilized. More may be identified as non-valid with sufficiently small intensity errors or the use of resonant contrast. The linearization method leverages the advantages of modern synchrotron X-ray sources, particularly through the use of multiple wavelengths and identified phases of the amplitudes, along with high-quality intensities [narrow isosurfaces, see *e.g.* Figs. 4–6 of Zschornak *et al.* (2024)] as well as high resolution in reciprocal space and high q values. In any case, the remaining non-unique solutions may be further analyzed and refined with modern state-of-the-art theoretical simulations like density functional theory (Blöchl, 1994; Kresse & Joubert, 1999; Perdew *et al.*, 1996) calculations, where the ground-state energies of the predicted structures can be assessed to discriminate chemically unstable solutions.

In the present implementation, all routines are in principle generalized in such a way that they can be extended to handle arbitrary non-centrosymmetric structures in m -dimensional PS. Performance and reliability tests for $m \geq 4$ are currently in progress and will be the scope of further work.

6. Related literature

The following references are cited in the supporting information: Komorowski *et al.* (2008), Rousseeuw *et al.* (1999),

Benjamini (1988); *Matplotlib* (https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.box.html, https://www.tutorialspoint.com/matplotlib/matplotlib_box_plot.htm).

Acknowledgements

This article is dedicated to the enduring legacy of the respected Professor Dr Karl Fischer (deceased in 2023), whose profound contributions to crystallography continue to inspire us. He was an invaluable part of the PSC project and we shall remember the valuable lessons learned under his insightful guidance and unwavering passion for knowledge. Professor Fischer actively participated in weekly video conferences from his retirement home until shortly before the end of his life. His intellectual freshness, professional and expressly human advice, and constant joy in exchanging ideas, even with young representatives of the next generation, were highly enriching. MZ and DCM acknowledge the work of their common teacher, Professor Peter Paufler (TU Dresden), and are very grateful for his continued support. MV thanks the Department of Information Service and Computing at Helmholtz-Zentrum Dresden-Rossendorf and the Center for Information Services and High-Performance Computing (ZIH), Technical University Dresden, for providing extensive computing facilities. MV acknowledges and thanks Professor Dr Ralf Hielscher (Faculty of Mathematics and Informatics, TU Bergakademie Freiberg, Germany) for the fruitful discussion on the linearization of manifolds. Open access funding enabled and organized by Projekt DEAL.

Funding information

MV, DCM and MZ acknowledge funding by the DFG within the project DFG 442646446, ZS 120/51.

References

- Androsov, A. S. & Shary, S. P. (2022). *Vestn. NSU Ser. Inf. Technol.* **20**, 5.
- Benjamini, Y. (1988). *Am. Stat.* **42**, 257–262.
- Billinge, S. J. L. & Proffen, Th. (2024). *Acta Cryst.* **A80**, 139–145.
- Blöchl, P. E. (1994). *Phys. Rev. B*, **50**, 17953–17979.
- Campbell, B. J., Stokes, H. T., Averett, T. B., Machlus, S. & Yost, C. J. (2021). *J. Appl. Cryst.* **54**, 1664–1675.
- Cervellino, A. & Cicciariello, S. (2005). *Acta Cryst.* **A61**, 494–500.
- Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. & Meester, L. E. (2005). *A modern introduction to probability and statistics: understanding why and how*. Springer Texts in Statistics. Springer.
- Filippidis, I., Dathathri, S., Livingston, S. C., Ozay, N. & Murray, R. M. (2016). *2016 IEEE conference on control applications (CCA)*, p. 1030. IEEE.
- Fischer, K. F., Kirfel, A. & Zimmermann, H. (2008). *Croat. Chem. Acta*, **81**, 381–389.
- Fischer, K. F., Kirfel, A. & Zimmermann, H. W. (2005). *Z. Kristallogr.* **220**, 643.
- Gillies, S. *et al.* (2022). *Shapely user manual*, <https://shapely.readthedocs.io/en/latest/index.html>.
- Harrison, R. W. (1993). *J. Opt. Soc. Am. A*, **10**, 1046.
- Hobson, E. W. (1909). *Proc. London Math. Soc.* **s2.7**, 14–23.
- Kirfel, A. & Fischer, K. F. (2009). *Z. Kristallogr.* **224**, 325–339.
- Kirfel, A., Fischer, K. F. & Zimmermann, H. W. (2006). *Z. Kristallogr.* **221**, 673.
- Knop, W. (1989). Dissertation, Universität Saarbrücken, Germany.
- Komorowski, M., Marshall, D. C., Saliccioli, J. D. & Crutain, Y. (2008). *Exploratory data analysis*. Springer.
- Kresse, G. & Joubert, D. (1999). *Phys. Rev. B*, **59**, 1758–1775.
- Lawson, C. L. & Hanson, R. J. (1974). *Solving least squares problems*. Prentice-Hall.
- Lozada-Cruz, G. (2020). *Int. J. Math. Educ. Sci. Technol.* **51**, 1155–1163.
- Munteanu, V., Starostin, V., Greco, A., Pithan, L., Gerlach, A., Hinderhofer, A., Kowarik, S. & Schreiber, F. (2024). *J. Appl. Cryst.* **57**, 456–469.
- Navaza, J. & Silva, A. M. (1979). *Acta Cryst.* **A35**, 266–275.
- Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* **A60**, 134–141.
- Ott, H. (1928). *Z. Kristallogr.* **66**, 136.
- Perdew, J. P., Burke, K. & Ernzerhof, M. (1996). *Phys. Rev. Lett.* **77**, 3865–3868.
- Pilz, K. (1996). Dissertation, Universität Saarbrücken, Germany.
- Richter, C., Zschornak, M., Novikov, D., Mehner, E., Nentwich, M., Hanzig, J., Gorfman, S. & Meyer, D. C. (2018). *Nat. Commun.* **9**, 178.
- Rothbauer, R. (1994). *Z. Kristallogr.* **209**, 578.
- Rothbauer, R. (1995). *Z. Kristallogr.* **210**, 255.
- Rothbauer, R. (1998). *Z. Kristallogr.* **213**, 195.
- Rousseeuw, P. J., Ruts, I. & Tukey, J. W. (1999). *Am. Stat.* **53**, 382–387.
- Sahoo, P. K. & Riedel, T. (1998). *Mean value theorems and functional equations*. World Scientific.
- Shi, H. L. (2022). *J. Appl. Cryst.* **55**, 669–676.
- The HDF Group & Koziol, Q. (2020). *Hdf5-version 1.12.0*, <https://doi.org/10.11578/dc.20180330.1>.
- Toby, B. H. (2024). *J. Appl. Cryst.* **57**, 175–180.
- Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramacchioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* **A52**, 770–781.
- Vallinayagam, M., Nentwich, M. & Zschornak, M. (2024). *pypsc: parameter space concept for determining 1D-projected crystal structures*, <https://github.com/mvnayagam/pypsc.git>.
- Woolfson, M. M. & Hai-Fu, F. (1995). *Physical and non-physical methods of solving crystal structures*. Cambridge University Press.
- Zschornak, M., Wagner, C., Nentwich, M., Vallinayagam, M. & Fischer, K. F. (2024). *Crystals*, **14**, 684.