Microarticle

Efficient methods of X-ray diffraction pattern inversion

Ian Gregory Shuttleworth

School of Science and Technology, Nottingham Trent University, Nottingham NG11 8QN, United Kingdom

ARTICLE INFO

ABSTRACT

An innovative, computational approach to the problem of X-ray diffraction pattern inversion is presented. The technique approximates the electron density of the trial structures encountered during the structure-fitting part of the inversion process with the superposition of the gas-phase electron densities of the component atoms of those structures. These latter electron densities were obtained using Density Functional Theory (DFT). The approach is shown to be computationally efficient, and to be effective when applied to both existing single-crystal x-ray diffraction data and to a series of simulated data sets obtained from periodic crystals. The effects of thermal fluctuation of the ion cores have been quantified, and the useful range of the technique has been discussed.

Introduction

Crystal structure prediction (CSP) is a fundamental challenge in both experimental and theoretical physics, and chemical physics. CSP [1] has been seen for many years as an accompanying process to x-ray diffraction pattern analysis, which itself is one of the main sources of structural information about molecules and molecular solids. Conventionally, two separate calculations are performed during CSP; a total energy search which determines the most likely structure or polymorph, followed by a second calculation which predicts the intensities and phases of the individual peaks in the diffraction pattern.

The actual tools of CSP are diverse but fall broadly into two main categories: force-field or semi-empirical methods [2] and less empirical techniques such as density functional theory (DFT) [3]. Though computationally less expensive, force field methods are limited by their accuracy. For a force field simulation to be believable, or rather to ensure that the predictions of the force field are an accurate reflection of what the experimental teams will have seen in their laboratories, is extensive 'training' of the field using more computationally demanding techniques is generally necessary [4–6].

Screening of possible trial structures would therefore seem to be preferable if performed using a less empirical tool such as DFT. However, for large and complex crystals the screening process could potentially require thousands, or more, estimated structures each of which would need to be relaxed. The computational demands of such a process are well known to the DFT community [7] and require significant computing power to be overcome.

Problems also arise once a structure has been relaxed under any scheme as the electron density throughout the crystal is required before

it can be Fourier transformed and the resulting structure factors can be compared with an experimental diffraction pattern [8]. The electron density throughout a crystal can be directly calculated once locations of the ion cores and its computational expense is comparable to that of a single 'step' of the relaxation. In addition, fast Fourier transform (FFT) algorithms are widely developed particularly in parallel environments and can be implemented to scale extremely well with problem size.

The current work presents a key innovation to the problem of x-ray diffraction pattern inversion. Central to the technique is the approximation of the crystals electron density using a superposition of the gas-phase electron densities of the individual ion cores. The paper is presented in the following way: a description of the theoretical and computational methods used in the current work are first presented followed by a quantitative study of the application of the technique to a sequence of ten known crystallites. An estimation of the error involved in the gas-phase approximation and the significance of isotropic and non-isotropic deformations of the gas phase electron density when the ion cores are incorporated into a bulk phase are made for each of the trial crystallites. A comparative discussion of this error and the effect of thermal noise is made, and the application of the technique to an experimental single-crystal x-ray diffraction pattern is then performed.

Theory and computational methods

Using fractional position coordinates, the generalised structure factor $F(hkl)$ for x-ray diffraction is given by

$$F(hkl) = V \int_0^1 \int_0^1 \int_0^1 \rho(xyz) exp\left[i2\pi(hx + ky + lz)\right] dxdydz \tag{1}$$

Eq. (1) shows explicitly the role that the electron density $\rho(xyz)$ plays in determining the structure factor. In this technique, the integrand in Eq. (1) replaces the conventional scattering factor which is usual in diffraction pattern analysis. This quantity can be estimated in a large number of ways; however, here DFT will be used both to generate $\rho(xyz)$ for the crystal, and also for the isolated gas-phase ion cores. The ion cores are the nuclei of the individual atoms within the material under investigation. DFT practitioners routinely use pseudopotentials to replace the electron density very close to that ion core as this reduces the very high number oscillations in the electron density that are seen in this region, and concomitantly reduces that amount of computational work required to accurately perform the DFT calculation [9]. This divides the materials charge into a 'core' term composed of the ion core itself and the pseudopotential region, and the valence charge. The charge density plots presented in this work are of the valence charge, and show little actual charge density close to the ion core.

The DFT calculations in this paper were performed using the LCAO method implemented in the SIESTA package [10]. The exchange-correlation interaction was estimated using the PBE functional [11]. Norm conserving pseudopotentials were generated using the improved Troullier-Martins method [12,13] and included non-linear core corrections [14]. A double-ζ split-valence basis set [15] was used together with polarisation orbitals, and single Kleinman-Blylander projector [16] was used to represent each angular momentum channel. The DFT code was compiled in an Intel environment using OpenMPI (1.8.1) and NetCDF (4.3.3.1) to ensure optimum parallelisation and efficient data flow.

The DFT calculations performed in this work were used to generate electron densities in two different contexts. In the first context, the electron density throughout the unit cell of a bulk-phase crystal was simulated using a single, fully self-consistent DFT simulation. This approach to generating the crystalline electron density $\rho(xyz)$ is denoted Full DFT in this work and the electron density obtained in this way is denoted by $\rho_F(xyz)$. In the second context, SIESTA was used to generate the electron density surrounding a single isolated ion core. This approach is common amongst the DFT community and is used, for example, when determining the formation energies of bulk crystalline samples. Because of the isolated nature of the ion core it is often referred to as being in the gas-phase and this terminology will be used in the current work. For this part of the investigation a $20 \times 20 \times 20 \, \text{Å}^3$ box was used with a single ion core located at the centre of the box. The same dimensions were used to generate electron density profiles for each ion cores species, and each line profile was consequently a function of only the radial distance from the ion core centre and independent of the azimuthal and polar angle.

The algorithm then generates the crystalline electron density $\rho(xyz)$ by summing the gas phase electron densities ρ_G for each of the component ion cores, with an appropriate weighting for distance. Because of the additive nature of the algorithm it will be termed Sum DFT, and the electron density obtained in this way is denoted by $\rho_S(xyz)$. Mathematically

$$\rho_S(xyz) = \sum_j \rho_G([x - x_j][y - y_j][z - z_j])$$

(2)

$\rho_G([x - x_j][y - y_j][z - z_j])$ is the gas phase electron density of the ion core located at $(x_j y_j z_j)$ and the summation across the unit cell and its nearest neighbours to ensure that the contribution of ion cores in adjacent unit cells are correctly added to total crystalline electron density of the central unit cell. Only the central unit cell is used in the subsequent Fast Fourier Transform used to generate the structure factors. Because the ρ_G were generated using the SIESTA DFT package they are necessarily quantised; therefore, an interpolation was performed using a polynomial algorithm [17] to deduce the contribution to ρ_S at intermediate points. Up to 5 interpolating points were used between each DFT sampling point, which themselves were separated by 0.02 Å.

An appropriate metric to quantify the effectiveness of the approach

summarise in Eq. (2) should compare the structure factors obtained with $\rho_S(xyz)$ with those obtained $\rho_F(xyz)$. The metric used is given in Eq. (3)

$$\chi_F^2 = \frac{1}{N} \sum_{j=1}^{N} F_{FS}^2 \times 100\%$$

(3)

N is the number of diffraction peaks. The F_{FS} are defined by Eq. (4)

$$F_{FS} = 2 \frac{F_F(hkl) - F_S(hkl)}{F_F(hkl) + F_S(hkl)}$$

(4)

$F_F(hkl)$ and $F_S(hkl)$ are the structure factors obtained using the Full DFT and Sum DFT approaches respectively, and which were outlined previously in this paper. χ_F^2 plays the role of the R-factor in the current work where the Full DFT diffraction intensities are playing the role of the experimental data. A further metric is useful in determining the accuracy of the summation performed in Eq. (2). The metric given in Eq. (5) compares the electron densities generated using the Full DFT and Sum DFT approaches

$$\chi_\rho^2 = \frac{1}{N_\rho} \sum_{j=1}^{N_\rho} \rho_{FS}^2 \times 100\%$$

(5)

N_ρ is the number of points in the electron density mesh. The ρ_{FS} are defined by Eq. (6)

$$\rho_{FS} = 2 \frac{\rho_F(xyz) - \rho_S(xyz)}{\rho_F(xyz) + \rho_S(xyz)}$$

(6)

$\rho_F(hkl)$ and $\rho_S(hkl)$ are the electron densities obtained using the Full DFT and Sum DFT approaches respectively. This comparative metric, though not as directly applicable to the analysis of actual experimental diffraction data as χ_F^2 is, is of use in analysing the quality of the approximations used to define $\rho_S(xyz)$ in Eq. (2). In this way, it quantifies the quality of the approximations used to define Eq. (2). These include the assumption that ρ_G is isotropic and that both isotropic and non-isotropic deformations of the electron density occur as the ion core transforms from the gas phase to a bulk phase. These approximations will be discussed in more detail in the Results and Discussion section of this paper.
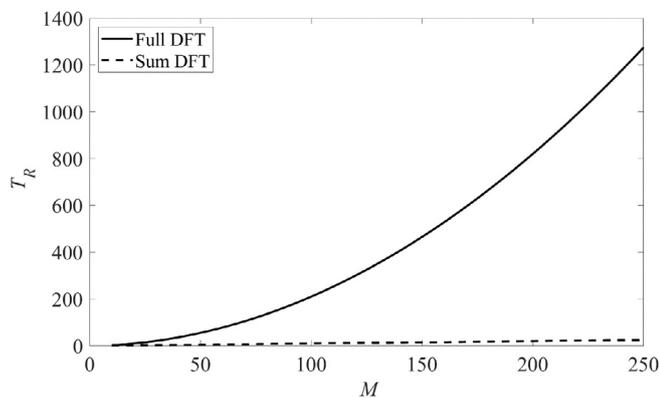
In applications of the Sum DFT technique to sets of experimental data the following quality-of-fit function was used

$$\chi_E^2 = \frac{1}{N_E} \sum_{j=1}^{N_E} \left(\frac{|F_E(hkl)|^2 - |F_S(hkl)|^2}{|F_E(hkl)|^2} \right)^2$$

(7)

N_E is the number of experimental data points, and the $|F_E(hkl)|^2$ are the experimental diffraction intensities. As this metric effectively amplifies low intensity diffraction peaks, only experimental diffraction peaks which had a value of greater than 0.01% of the maximum reported experimental diffraction intensity were used in the summation shown in Eq. (7). This was done to reduce the effect of effectively corrupting the metric with experimental noise.

Results and discussion

Fig. 1 shows the relative execution time T_R versus problem size, or number of atoms in the unit cell of the simulated crystal, M for a typical crystalline system. To generate the crystalline electron densities $\rho(xyz)$ up to 48 nodes with 8 processors per node were used in the SIESTA simulations presented in this work, whereas the $\rho(xyz)$ generated using the novel summation algorithm described in this work were performed on a single node. In both cases the CPU time required to perform a Fast Fourier Transform (FFT) to transform the electron density to diffraction space was identical as both real space grids had identical dimensions. Consequently, this component of the CPU time was ignored in any subsequent timing studies of the two algorithms and T_R was defined as the ratio of the CPU times required to generate the $\rho(xyz)$ for a system
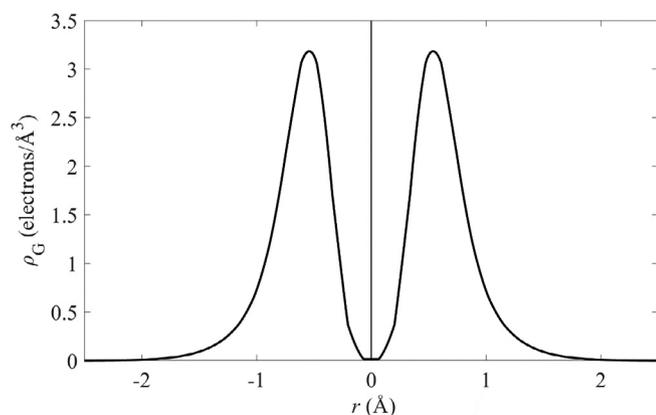
Fig. 1. Comparison of the relative execution time T_R versus problem size M for inversion algorithms that generate the crystal's electron density using a full self-consistent density functional theory simulation (Full DFT) or the novel summative approach (Sum DFT) presented in this work.

with M atoms to that required for a system with M_{ref} atoms. For the example shown in Fig. 1 an FCC Cu crystal was used and M_{ref} was 10. FCC Cu was used partly because of its well-known behaviour in DFT, and partly because the primitive unit cell has only a single atom allowing a large number of data to be generate.

Fig. 1 shows the well-understood non-linear divergence with problem size routinely encountered when DFT is used whereas the interpolative algorithm shows a more linear and lower magnitude increase in T_R with M. A similar behaviour was observed for each of the other crystalline systems investigated in this study though the density of data in the T_R curves varied relatively largely because of the different numbers of ion cores in each unit cell. The conventional question of how the execution time of the DFT algorithm scales with problem size is discussed in the original manifesto for the SIESTA code [10].

Fig. 2 shows the gas-phase electron density lineshape $\rho_G(r)$ as a function of the radial distance r from the centre of the ion core for an Au ion core. The features that appear in this line profile will affect the physicality of the inversion and merit highlight. However, all the features that appear in this profile have been developed within the DFT community to improve the quality of the model of the system and are therefore generally approved of by the DFT community for the systems under investigation.

The nuclear core exists in the $r \in [-0.07, +0.07]$Å region which itself is surrounded by a semi-core region with occupies the $r \in [0.07, 0.20]$Å region in the positive interval of r and a similar region in the negative interval of r. Non-linear core corrections [14] were applied during these simulations and these give rise to the transition in gradient seen at $r = \pm 0.20$Å. Application of the non-linear core

Fig. 2. Electron density $\rho_G(r)$ versus radial distance r from the centre of the ion core, for a gas phase Au ion core.

Table 1
Summary of the structure factor and electron density quality metrics for each of the ten periodic crystalline systems used to benchmark the inversion algorithm.

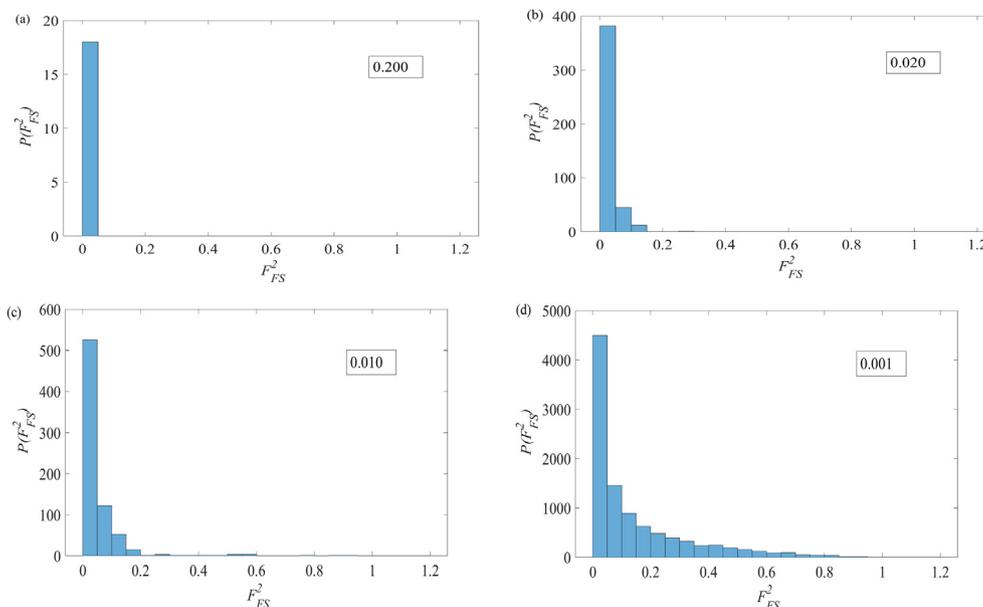| Mineral	Stoichiometric formula	χ_F^2 (%)	χ_ρ^2 (%)
Oxide	Goethite (FeO$_2$H)	3.89	4.07
Native element	Gold (Au)	1.12	1.21
	Copper (Cu)	2.65	2.94
	Diamond (C)	2.12	2.23
Silicate	Quartz (SiO$_2$)	2.65	2.73
Sulphate	Barite (BaSO$_4$)	3.82	4.03
Sulphide	Galena (PbS)	0.76	0.94
	Pyrite (FeS$_2$)	2.56	2.88
Halides	Halite (NaCl)	0.87	0.95
	Fluorite (CaF$_2$)	2.79	2.86

correction is generally seen to improve the accuracy of the DFT simulation though at the added computational expense. This expense is acceptable as the correction needs only to be applied once, and then the potential can be repeatedly estimated using the interpolation routines discussed earlier in this paper. The high electron density region extends out to maxima at $r = \pm 0.53$Å before starting to taper to zero at large magnitudes of r.

Table 1 shows the crystalline systems used for benchmarking. The breadth of the survey – which is defined by the anion of the mineral or by the atomic number of the native element – was chosen to provide a reasonable survey of elements and structural classes across the periodic table. This challenges the method to perform in a range of environments where the features of the line profile of the gas-phase atom, outlined in the previous paragraph, might be expected to show sensitivity to the local stoichiometries and lattice type. The electron density around individual ion cores will be deformed when the atoms are part of a crystal compared to be when they are in their isolated, gas phase state. Consequently, the metrics presented in Table 1 quantify the effects of these deformations.

The structure of each crystal was taken from the American Mineralogist Crystal Structure Database [18]. For each crystal the equation of state was determined using DFT and the inversion algorithm benchmarked using the crystal dimensions which correspond to the minimum of the equation of state. The metrics shown a correlation between χ_F^2 and χ_ρ^2 and this behaviour is reasonable as the two quantities are based on a Fourier transform pair.

Though the correlation is useful and the scaling and normalisations used in the definitions of χ_F^2 and χ_ρ^2 provide an approximate aligning of the actual magnitudes of the two quantities, some more insight into how the functions actually develop is required. Fig. 3 shows the histograms of the F_{FS}^2 which contribute to χ_F^2 for a sequence of different thresholds. It can be seen that for the larger diffraction intensities $F_F(hkl) > 0.010$ the difference between the Full DFT and Sum DFT estimates of the diffraction intensities are close to zero and certainly lower than any nominal experimental error of, for example, 1–5%. This would indicate that for the larger diffraction intensities the method is limited by experimental error rather than errors in the approximation of the crystals electron density using Eq. (2). For lower intensities, however, the range of F_{FS}^2 becomes larger and certainly within the $F_F(hkl) < 0.010$ the analysis has become limited by the analysis method rather than experimental error. The explicit analysis presented in histograms shown in Fig. 3 is very much one of the worst cases. It can be seen by looking down the χ_F^2 column in Table 1 that barite is one of the poorer performing crystals. It might be anticipated based on the χ_F^2 data in Table 1 that for other crystals the transition from experimental error limiting to algorithm error limiting would occur at lower values of $F_F(hkl)$.
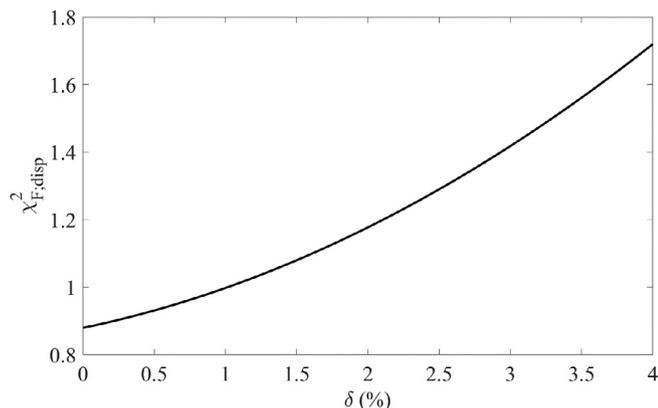
The final stage in benchmarking the inversion algorithm was to estimate the effect of displacing ion cores from their equilibrium positions within the crystal. This estimate simulates two effects. The first is the physical phenomenon of thermal agitation of the ion cores and is a

Fig. 3. Histograms of the normalised structure factor amplitude differences $F_{FS}(hkl)$ defined in Eq. (4) for the barite crystal. The structure factor differences are defined in Eq. (4) in terms of the structures determined using the Full DFT and Sum DFT methods. The inset shows the threshold value for $F_F(hkl)$ used to generate each of the four histograms presented in this figure; only diffraction channels where $F_F(hkl)$ is greater than the threshold value are included in each panel.

standard part of x-ray diffraction analysis. The second effect is movement of the ion during a search routine. When fitting of a diffraction pattern to a structural model, the ion cores are moved, the resulting electron density is calculated using various levels of estimation and the structure factors are obtained using a Fourier transform. The 'best fit' structure is then determined when χ_F^2 is minimised and suite of multi-dimensional search routines exist to perform this fitting [17].

It is also useful to get some idea of how much effect changes in the position of the ion cores will affect χ_F^2. A secondary metric affect $\chi_{F;disp}^2$ was defined which has the same formal definition as χ_F^2 which was given in Eq. (3). However the F_{FS} are defined differently, and use $F_F(hkl)$ from simulations with the all the ion cores in their equilibrium positions whereas the $F_S(hkl)$ are obtained with atoms displaced from their equilibrium positions. Fig. 4 shows the degradation of $\chi_{F;disp}^2$ with a systematic change in the position of one of the Cl ion cores in a halite (NaCl) crystal. The normalised displacement δ was defined as the ratio of the absolute displacement of the Cl ion core away from its equilibrium position to the equilibrium lattice constant of the halite crystal.
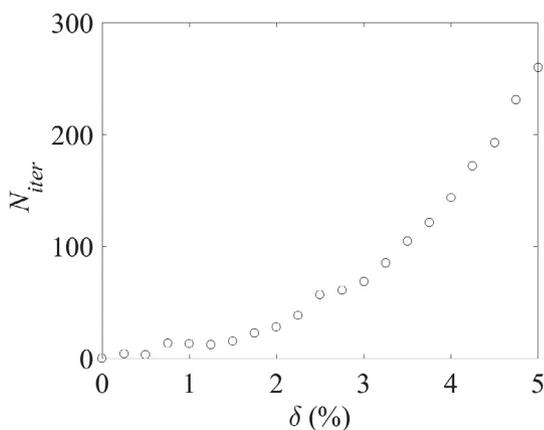

Fig. 4. Effect on $\chi_{F;disp}^2$ of displacement δ of the ion cores in the halite (NaCl) crystal from their equilibrium positions. $\chi_{F;disp}^2$ is the quality of fit between the Full DFT and Sum DFT simulations for crystals which have ion cores displaced from the equilibrium positions. In each simulation the Cl ion core was displaced along $\langle 100 \rangle$.

A similar trend-line was seen when the Na ion core was displaced, and when ion cores in each of the other crystalline systems were displaced along either high- or low-symmetry directions. This suggests that the effect of displacement of ion cores away from their equilibrium positions would significantly alters the quality-of-fit parameter in a fitting process even for displacements δ of just a fraction of a percent of the equilibrium lattice parameter. Though no attempt to generalise this conclusion will be made in this paper a similar set of trend-lines were obtained when this process was repeated across all of the other crystals. However, instead a proposed methodology of using the Sum DFT technique to locate the ion cores to within a few percent of their equilibrium positions followed by use of Full DFT to find the exact final position coordinates would seem highly appropriate.

The technique was also tested against a set of experimental x-ray diffraction data obtained from a single LaZn$_5$ crystal [19]. From the original reference [19] the location of the ion cores in the crystal are known, so the search space was be defined by randomly displacing these ion cores to simulate a trial structure and then measuring the number of iterations required by a search engine to re-find the original equilibrium positions of the ion cores. The quality of fit function χ_E^2 was defined in Eq. (7) and the Fast Inertial Relaxation Engine (FIRE) [20] was used to perform the search. Fig. 5 shows the number of iterations N_{iter} required by the FIRE algorithm to determine equilibrium positions of the ion cores as a function of their fractional displacement δ from their equilibrium position. The value of χ_E^2 obtained with the ion cores at their equilibrium ($\delta = 0$) positions was 0.04 using Sum DFT to estimate the diffraction intensities, and 0.03 when Full DFT was used. In the original work [19] the experimental diffraction intensities were compared with those obtained using tight-binding linear muffin tin calculations, and a value of $\chi_E^2 = 0.08$ was obtained. The improvement in χ_E^2 between the original work [19] and the current work is due to the level of approximation used in the level of approximation used in the simulated diffraction intensities.

Conclusions

A novel approach to the problem of single-crystal x-ray diffraction pattern inversion has been proposed which is based on the approximating the crystalline electron density with a superposition of the

Fig. 5. Effect on the number of search iterations N_{iter} of displacement δ of the ion cores in the LaZn$_5$ crystal from their equilibrium positions during simulations of experimental diffraction intensities. In each simulation, the ion cores were displaced isotropically.

isotropic electron densities that surround the constituent ion cores in the gas phase. The work has shown that DFT can be used to generate a set of benchmark electron densities, which by using simple interpolative techniques, can be used to then survey a very large range of trial structures in a computational extremely cost-efficient way. The method has been shown to correctly invert the x-ray diffraction images generated by a sequence of minerals and native elements drawn from a wide range of the periodic table, and an approximate range for the error that arises from both the gas phase approximation and displacement of the ion cores from their equilibrium positions has been determined. The systems that have been looked at in the current work have been relatively easily accessible to DFT as they have been well-ordered, defect-free and required small unit cells to be simulated. Real samples are far more diverse, with defects and step edges, for example, requiring significantly larger unit cells to effectively survey. This, however, is when the Sum DFT procedure really will be able to come into its own; because of its very slow scaling with system size, which was evidenced in Fig. 1, large unit cells are very easily accessible to the technique, far more so than using a conventional, Full DFT type approach. Because of its computationally efficient form, the Sum DFT is not the limiting part of

the overall fitting process. Once the equilibrium ion core positions have been determined using the Sum DFT approach, a single Full DFT calculation is required to qualify the approximations used in the Sum DFT algorithm. Using modern DFT codes these Full DFT calculations can be applied to systems of even, for example, 6 000–14 000 atoms [21] but they may be challenged if applied to significantly larger systems.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rinp.2019.102605.

References

[1] Gu T, Luo W, Xiang H. WIREs Comput. Mol. Sci. 2017;7:e1295.
[2] Harrison JA, Schall JD, Maskey S, Mikulski PT, Knippenberg MT, Morrow BH. Appl. Phys. Rev. 2018;5:031104.
[3] Mardirossian N, Head-Gordon M. Mol. Phys. 2017;115(19):2315–72.
[4] Monticelli L, Tieleman DP. Force Fields for Classical Molecular Dynamics. Monticelli L, Salonen E, editors. Biomolecular Simulations. Methods in Molecular Biology (Methods and Protocols), vol 924. Totowa, NJ: Humana Press; 2013.
[5] González MA. Collect. SFN 2011;12:169–200.
[6] Trnka T, Tvaroška I, Koča J. J. Chem. Theory Comput. 2018;14(1):291–302.
[7] Woodley SM, Catlow CRA. Phys. Chem. Chem. Phys. 2014;16:21001.
[8] Ladd M, Palmer R. Structure Determination by X-ray Crystallography. Springer; 2013.
[9] Martin RM. Electronic Structure: Basic Theory and Practical Methods. (Cambridge University Press); 2004.
[10] Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, et al. J. Phys.: Condens. Matt. 2002;14:2745–79.
[11] Perdew JP, Burke K, Enzerhof M. Phys. Rev. Lett. 1996;77:3865–8.
[12] Troullier N, Martins JL. Phys. Rev. B 1991;43(3):1993–2006.
[13] Troullier N, Martins JL. Phys. Rev. B 1991;43(11):8861–9.
[14] Louie SG, Froyen S, Cohen ML. Phys. Rev. B 1982;26(4):1738–42.
[15] Junquera J, Paz Ó, Sánchez-Portal D, Artacho E. Phys. Rev. B 2001;64(23). 235111-1/9.
[16] Kleinman L, Bylander DM. Phys. Rev. Lett. 1982;48(20):1425–8.
[17] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C. The Art of Scientific Computing. Cambridge University Press; 2002.
[18] Downs RT, Hall-Wallace M. Am. Mineral. 2003;88:247–50.
[19] Oshchapovsky I, Zelinska O, Rozdzynska-Kielbikb B, Pavlyuk V. Acta Crystallogr Sect. E: Struct. Rep. Online 2012;68(1):i1.
[20] Bitzek E, Koskinen P, Gahler F, Moseler M, Gumbsch P. Phys. Rev. Lett. 2006;97:170201.
[21] Erba A, Baima J, Bush I, Orlando R, Dovesi R. J. Chem. Theory Comput. 2017;13:5019–27.