Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 3 February 2010
Accepted 13 April 2011

Using molecular-dynamics simulations to understand and improve the treatment of anharmonic vibrations. I. Study of positional parameters

Anthony M. Reilly,[a,b] Carole A. Morrison[a]* and David W. H. Rankin[a]

[a]School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland, and [b]Chair for Process Systems Engineering, Technische Universität München, D-85350 Freising, Germany. Correspondence e-mail: carole.morrison@ed.ac.uk

Molecular-dynamics-derived numerical probability density functions (PDFs) have been used to illustrate the effect of different models for thermal motion on the parameters refined in a crystal structure determination. Specifically, anharmonic curved or asymmetric PDFs have been modelled using the traditional harmonic approximation and the anharmonic Gram–Charlier series treatment. The results show that in cases of extreme anharmonicity the mean and covariance matrix of the harmonic treatment can deviate significantly from physically meaningful values. The use of a Gram–Charlier anharmonic PDF gives means and covariance matrices closer to the true (numerically determined) anharmonic values. The physical significance of the maxima of the anharmonic distributions (the most probable or mode positions) is also discussed. As the data sets used for the modelling process are theoretical in origin, these most probable positions can be compared to equilibrium positions that represent the system at the bottom of its potential-energy surface. The two types of position differ significantly in some cases but the most probable position is still worthy of report in crystal structure determinations.

1. Introduction

One of the most important considerations in determining crystal structures is the fact that the atoms in the crystal of interest are not actually static: zero-point energy and thermal excitations lead to a variety of atomic and lattice vibrations, which have a big impact on the information obtained using diffraction methods. The vibrations occur on short timescales compared to the data-collection time with two important consequences. First, the raw structure determined by diffraction is time averaged over these motions and represents the mean atomic configuration. Secondly, the condition for maximally constructive interference and Bragg scattering is only fulfilled for the time that the mean position is occupied. The result of the latter is that thermal motion attenuates the diffraction intensities. This attenuation is an unwelcome problem for data collection but the necessary modelling of thermal motion required to rectify it can lead to very useful information on the displacements of atoms in a crystal. Such information is naturally complementary to spectroscopic methods, which give information on energetics.

Thermal motion is accounted for in the structure factor using the Debye–Waller factor (Debye, 1913; Waller, 1923). A plethora of forms for the Debye–Waller factor have been proposed in the literature, ranging from simple one-parameter harmonic treatments to complex statistical models. The different treatments of thermal motion can lead to a variety of different parameters that can be used to define and assess a crystallographic structure and its dynamic behaviour. Understanding the meaning of and the differences between these parameters is difficult because of the need for high-quality data, limitations in anharmonic models and the fact that not all of the desired information is accessible by diffraction experiments alone.

In the present work, we have used molecular-dynamics (MD) simulations to provide a theoretical viewpoint on the crystallographic interpretation of thermal motion. In an MD simulation the atomic forces in a system are calculated (either from an empirical force field or some form of quantum mechanics) and are used to follow the dynamics of the system over time. We show that the resulting trajectories can be used to carry out theoretical investigations of the different treatments of thermal motion and their effect on the determined parameters. This analysis is performed in real space only and therefore additional issues that arise in experimental structure determinations (*i.e.* the need to perform analysis in reciprocal space) are not present here. The relationship between the different structural parameters is the focus of this paper, while the use of MD data sets in assessing and developing new Debye–Waller factors

will be shown in the second paper in this series (Reilly *et al.*, 2011).

2. Modelling thermal motion in crystallography

2.1. Debye–Waller factors

The Debye–Waller factor can be derived from a number of different starting points (Kuhs, 1992). Indeed, as well as describing thermal motion it can account for static disorder in a crystal structure (Kuhs, 1983) and can have contributions from systematic errors (Trueblood *et al.*, 1996). In the present work we are interested solely in its ability to model thermal motion, in which case the structure-factor equation takes the form

$$\langle F(\mathbf{Q})\rangle = \sum_{n=1}^{N} f_n \hat{\boldsymbol{P}}_n(\mathbf{Q}) \exp(i\mathbf{Q}\cdot\mathbf{r}_{n0}), \qquad (1)$$

where $\hat{\boldsymbol{P}}(\mathbf{Q})$ is the Debye–Waller factor. The real-space Fourier transform of the Debye–Waller factor, $P(\mathbf{u})$, describes the three-dimensional thermal motion of an atom in the mean field of all the other atoms in the lattice, and thus attenuates the scattering intensity in line with the fact that an atom spends time away from the mean position that optimally satisfies the condition for Bragg scattering.

The functional form of $\hat{\boldsymbol{P}}(\mathbf{Q})$, and hence $P(\mathbf{u})$, plays an important role in the structure factor. The better the representation of thermal motion the better the quality of a fit to experimental intensities will be, ensuring that the determined structure is physically reasonable. $P(\mathbf{u})$ is most often approximated as being a trivariate Gaussian distribution, thereby allowing harmonic motion along three separate axes. The mathematical expression for this is given by (Kuhs, 1992; Johnson, 1969)

$$P(\mathbf{u}) = \frac{[\det(\mathbf{U}^{-1})]^{1/2}}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2}\mathbf{u}^{T}\mathbf{U}^{-1}\mathbf{u}\right), \qquad (2)$$

where $\mathbf{U}$ is the symmetric, 3×3 variance–covariance matrix of the distribution, whose elements U^{ij} equal $\langle u_i u_j\rangle$. Subscript indices are used on the components of $\mathbf{u}$, as in the present work a Cartesian basis is always used. See Trueblood *et al.* (1996) for detailed recommendations on the notation of probability functions and Debye–Waller factors in crystallography. If $i = j$ then U^{ii} represents the variance (the mean-square displacement from the average position) or extent of the thermal motion along a particular axis. When $i \neq j$, U^{ij} gives the covariances of the thermal motion, which give the orientation of the distribution relative to the axis system of choice. Taking the Fourier transform leads to

$$\hat{\boldsymbol{P}}(\mathbf{Q}) = \exp\left(-\frac{1}{2}\mathbf{Q}^{\mathbf{T}}\mathbf{U}\mathbf{Q}\right). \qquad (3)$$

The six unique elements of $\mathbf{U}$ usually have units of Å^2 and are referred to as the anisotropic displacement parameters (ADPs) (Trueblood *et al.*, 1996), although other names and forms for them are given in some older papers. The trivariate or anisotropic approximation can lead to significantly better R factors than a simpler isotropic model and is routinely used in small-molecule X-ray and neutron diffraction.

The trivariate Gaussian probability density function (PDF) corresponds to a harmonic model of thermal motion where atomic motions are resolved into three components along orthogonal axes. Such an approximation represents a good balance between the quality of the model and the number of parameters needed to describe the system. However, for some systems the approximation can be inadequate. One of the earliest examples of this is provided by Cruickshank (1956), who showed that the librational or curvilinear motion of an atom relative to another can lead to apparent shortening of bond lengths as the mean position of an atom is shifted from a physically reasonable location towards the centre of rotation (Schomaker & Trueblood, 1998). Intramolecular distances can be corrected *a posteriori* for this effect using the TLS method (Schomaker & Trueblood, 1968). The source of the problem is the harmonic nature of the Gaussian distribution. As the mean position is also the most likely or probable one, the PDF is incapable of properly modelling a distribution that is skewed or bent or otherwise has mean and most likely positions that do not coincide.

A variety of anharmonic Debye–Waller factors have been proposed in the past, using methods such as the one-particle potential approach (Willis, 1969; Tanaka & Marumo, 1983), the Edgeworth series (Johnson, 1969) and the Gram–Charlier (GC) series (Zucker & Schulz, 1982). These approaches generally involve some form of Taylor expansion of the harmonic approximation. The most widely used and recommended method is based on the GC series (Trueblood *et al.*, 1996). This involves an expansion of the harmonic PDF using high-order quasi-moments (Kuznetsov *et al.*, 1960), c^{ijkl}, and Hermite polynomials, H_{ijkl}:

$$P_{\mathrm{GC}}(\mathbf{u}) = P(\mathbf{u})_{\mathrm{harm}} \times \left[1 + \frac{1}{3!}c^{ijk}H_{ijk}(\mathbf{u}) + \frac{1}{4!}c^{ijkl}H_{ijkl}(\mathbf{u}) + \ldots\right], \qquad (4)$$

where $P(\mathbf{u})_{\mathrm{harm}}$ is a standard trivariate Gaussian PDF and all indices that appear twice are implicitly summed over. The corresponding Debye–Waller factor is

$$\hat{P}_{\mathrm{GC}}(\mathbf{Q}) = \hat{P}_{\mathrm{harm}}(\mathbf{Q}) \times \left(1 - ic^{ijk}Q_iQ_jQ_k + c^{ijkl}Q_iQ_jQ_kQ_l + \ldots\right). \qquad (5)$$

The odd-order terms skew the distribution, while the even-order terms affect its 'peakedness' or physical extent.

The use and implementation of anharmonic Debye–Waller factors have been limited for a number of reasons. Anharmonic Debye–Waller factors require many more parameters: equations (4) and (5) feature ten unique third-order and 15 unique fourth-order coefficients. In addition, the interpretation of the quasi-moments *etc.* that are derived from the anharmonic models is difficult. Visualization and other subtleties of their use complicate matters further. The rather generic approach to obtaining anharmonic forms of $\hat{\boldsymbol{P}}(\mathbf{Q})$ is a result of the fact that little independent information is available on the true nature and, perhaps more importantly, the

effect of anharmonic motion on refined structures. The focus of the present work is to illustrate how MD simulations can bridge this gap in our knowledge of anharmonic effects, providing useful test data sets to assess and develop the various models used to describe anharmonic motion and to increase our understanding of them.

2.2. Positional parameters

The position $\mathbf{r}_{n0}$, which appears in the structure factor is, by definition, the mean position, $\mathbf{r}_a$, and represents the average position adopted by the atom over the course of time. In turn, the refined value of the position is dependent on the functional form of the PDF used to describe the thermal motion of an atom. The PDF and Debye–Waller factor are centred on this mean position so that

$$\langle u_i \rangle = \int_{-\infty}^{\infty} u_i P(\mathbf{u})\, \mathrm{d}\mathbf{u} = 0. \quad (6)$$

Only when the approximate PDF correctly models the true PDF is the refined mean guaranteed to be the true mean position.

The mean is not the only parameter that we might determine in a diffraction experiment. In a harmonic description of thermal motion the 'most probable' position that represents the maximum of the PDF coincides with the mean position. In an anharmonic distribution this is not necessarily the case and a separate most probable position, denoted here $\mathbf{r}_p$, can be defined (Johnson, 1969). While such a position is not directly determined from the refinement process it can be obtained from the numerical or analytical maximum (or mode) of the anharmonic PDF. The most probable position also represents the minimum of the three-dimensional effective potential-energy surface that describes the mean-field behaviour of an atom. It should be a more reasonable description of the molecular structure, particularly with respect to bond lengths.

The final type of position is the equilibrium position, $\mathbf{r}_e$. This represents the vibrationless structure of the molecule at the minimum of the $(3N-3)$-dimensional potential-energy surface of the crystal, corresponding to the maximum of the $(3N-3)$-dimensional PDF that completely describes thermal motion. The three-dimensional PDFs determined in a diffraction experiment are marginal distributions of this PDF; for a given atom they represent the total PDF integrated over all of the other atoms in the unit cell (Scheringer, 1986). In integrating out the information on other atoms the probable structures will not necessarily correspond to the equilibrium one because the correlations between atoms cannot be obtained from a single experiment.

The equilibrium structure is the most desirable one, as it permits the direct comparison of structural parameters of different molecules, phases, polymorphs *etc.* in the absence of any thermal-motion effects. In addition, equilibrium structures are easily obtained from theoretical calculations. The growth of *ab initio* and density functional theory (DFT) structural studies, particularly in the solid state, makes comparison enviable for validating the various computational approaches. Average structures can be distorted by thermal motion, which will be different for different systems and temperatures. The most probable structure ignores the correlations between atoms and may also give spurious values as a result.

2.3. MD simulations

Relating the different structures that we might determine for a molecule is difficult because experiment alone cannot access all of the information. In particular, the equilibrium geometry cannot be readily obtained because the PDFs determined in an experiment represent only the marginal distributions of the total $(3N-3)$-dimensional potential-energy surface upon which the equilibrium structure lies. An alternative approach would be to use theory to estimate this information. We have recently developed a method of determining experimental equilibrium structures using MD simulations (Reilly *et al.*, 2007). The simulations determine the theoretical or computed time-averaged structure. Comparison of this with a computed equilibrium structure yields corrections that can be applied to an experimental time-averaged structure to estimate the experimental equilibrium structure. The true equilibrium structure of a crystal is represented by a motionless system at 0 K. While, in principle, for some systems experimental results could be extrapolated to 0 K, for some phases or polymorphs it will be impossible to experimentally estimate 0 K behaviour. Therefore the experimental equilibrium structures determined using MD simulations (including those discussed here) represent effective equilibrium structures for unit cells at particular temperatures.

The MD simulations also provide a picture of where each atom spends its time in space. Such information can be used to determine the full $(3N-3)$-dimensional PDF describing the thermal motion in the system and therefore leads to the information on diffuse scattering and interatomic correlation, which is lost in a normal diffraction experiment. Our focus in the present work is on the relationship between the various different structure parameters ($\mathbf{r}_a$, $\mathbf{r}_p$ and $\mathbf{r}_e$) as far as they can be determined with Bragg scattering experiments.

By numerically 'binning' the MD trajectory of each atom we can determine the three-dimensional atomic PDFs that the structure factor aims to model in a standard refinement. These PDFs can be modelled with any function and can be used to understand how the different structural parameters relate to one another and how the nature of the approximate PDF/Debye–Waller factor affects the fitted parameters. As we can also compare the 'refined' data with the theoretical equilibrium structure, we can, for the first time, compare time-averaged harmonic and anharmonic structures with equilibrium parameters without any approximation or assumptions being required as to how theory and experiment relate to one another. Thus, molecular dynamics provides the ideal data sets for assessing the various approximations that can be made in a crystallographic refinement. Even where theory cannot replicate the true experimental behaviour exactly, the MD-simulated results will still represent a physically realistic and

consistent data set for understanding the refinement process and the different parameters it can produce.

3. Simulation and analysis methods

3.1. Nitromethane and ammonia

The crystal structure of d_3-nitromethane has been the subject of numerous structural studies (David *et al.*, 1992; Jeffrey *et al.*, 1985; Trevino *et al.*, 1980) that have focused on the large-amplitude curvilinear motion of the deuterium atoms in its methyl group. We have recently employed MD simulations to determine the experimental equilibrium structure of this molecule (Reilly *et al.*, 2010*b*), showing highly anharmonic PDFs not only for the D atoms, as expected, but also for the O atoms at higher temperatures. The comparison between the simulation results and neutron diffraction experiments at low temperatures was favourable, with the ADPs in many cases agreeing within error with the experimental values. The simulations were also able to reproduce the low-temperature TLS vibrational corrections to the C—D distances. These MD simulations used a classical force field to determine the interatomic forces. To ensure that the primarily low-temperature quantum behaviour was modelled correctly the path-integral method (Feynman & Hibbs, 1965) was used to propagate the trajectory. In the present work the results of the simulations at 15 and 228 K are used as model data sets for comparing the different parameters refineable from crystallographic data. See Reilly *et al.* (2010*b*) for more details about the simulations, and comparison of the experimental and simulated means and variances.

The D atoms of nitromethane are an extreme example of curvilinear motion, whereas phase-1 deuterated ammonia, ND_3, is a more typical demonstration of librational motion and has also been studied using MD simulations (Reilly *et al.*, 2007, 2010*a*). The results of our previous DFT simulations are employed here. A detailed comparison of DFT and force-field-derived results with experimental data is presented elsewhere (Reilly *et al.*, 2010*a*).

3.2. Urea–phosphoric acid (1:1)

The 1:1 adduct of urea and phosphoric acid (UPA) has attracted considerable interest in the literature owing to the short, strong hydrogen bond between the urea oxygen proton and an oxygen of phosphoric acid (Fig. 1). UPA has been studied using neutron diffraction (Wilson, 2001; Wilson *et al.*, 2001) and computational methods, including plane-wave DFT (Wilson & Morrison, 2002; Morrison *et al.*, 2005; Fontaine-Vive *et al.*, 2006). The experimental and theoretical results show that the proton migrates from the urea to the phosphoric acid as a function of temperature, and that its effective potential/PDF is likely to be skewed, particularly at lower temperatures.

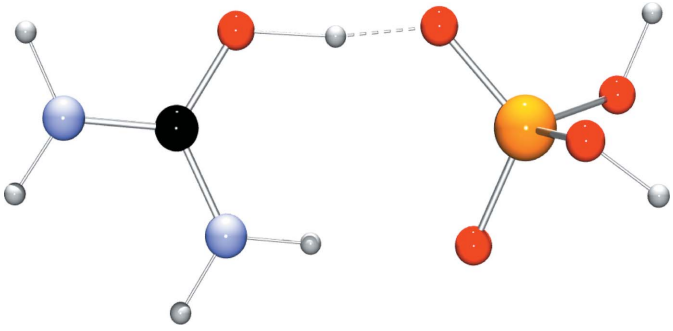

Figure 1
Molecular structure of the 1:1 adduct of urea and phosphoric acid.

MD simulations of UPA were performed at 150 and 350 K. The simulations were carried out on single unit cells (fixed at experimental cell vectors) with full periodic boundary conditions used to replicate the long-range order and interactions of the real crystal. The Car–Parrinello method (Car & Parrinello, 1985) was used as incorporated into the *CPMD* simulation code (CPMD, 2008). The time step for the MD simulation was 0.0725 fs. The Car–Parrinello method models not only the dynamics of the nuclei but also the faster electron dynamics and therefore a smaller time step than would normally be used in a force-field-type MD simulation is required. The electronic structure of UPA was represented using a plane-wave basis set with a cutoff energy of 1400 eV. This value represents the convergence of the plane-wave basis in terms of its highest-energy wavevector. A larger value leads to negligible changes in the total energy and individual atomic forces. The core electrons were represented using Troullier–Martins norm-conserving pseudopotentials (Troullier & Martins, 1991). The Perdew–Burke–Ernzerhof functional (Perdew *et al.*, 1996) was used to model the exchange and correlation energies.

The temperature was regulated using a chain of Nosé–Hoover (Nosé, 1984; Hoover, 1985) thermostats for the nuclei and the electrons. The 150 K MD simulation was run for a total of 24 ps, while the 350 K simulation was run for 30 ps, with data being collected every fifth Car–Parrinello time step in both simulations. The short running time is a constraint of using computationally expensive DFT-derived forces. However, even with such short simulation lengths we can resolve low-frequency vibrations (see for example Reilly *et al.*, 2008) and the use of the crystal symmetry (as detailed below) for averaging of the data ensures we maximize the available information.

3.3. Trajectory analysis and fitting of probability density functions

The true anharmonic mean, $\langle \mathbf{u} \rangle$, can be calculated directly from the trajectory using a numerical version of equation (6),

$$\langle \mathbf{u} \rangle = (\langle u_1 \rangle, \langle u_2 \rangle, \langle u_3 \rangle), \tag{7}$$

given that

$$\langle u_i \rangle = \frac{1}{N} \sum_{n=1}^{N} u_{i,n}, \tag{8}$$

where N is the total number of steps the average is calculated over and $u_{i,n}$ is the value of u_i at step number n. The true

Table 1
Harmonic and anharmonic mean of selected atomic PDFs with origin, (0, 0, 0), representing the numerical mean.

The indices of the nitromethane atoms refer to those used in the previous crystallographic studies of nitromethane (Jeffrey *et al.*, 1985). The PDFs have been orientated such that $U^{ij} = 0$ for $i \neq j$.

		Harmonic mean (Å)			Anharmonic mean (Å)		
Atom	T (K)	x	y	z	x	y	z
D-ND_3	180	−0.0037 (2)	−0.0056 (2)	−0.0027 (1)	−0.0021 (4)	−0.0005 (3)	0.0010 (3)
C-$MeNO_2$	15	−0.0001 (2)	−0.0007 (2)	−0.0002 (1)	0.0000 (3)	0.0000 (2)	0.0001 (2)
D1-$MeNO_2$	15	0.0077 (3)	−0.0031 (1)	−0.0087 (1)	−0.0005 (4)	0.0006 (2)	0.0009 (1)
O1-$MeNO_2$	15	0.0004 (2)	−0.0006 (2)	0.0013 (1)	0.0000 (3)	−0.0001 (3)	−0.0002 (2)
C-$MeNO_2$	228	−0.0029 (1)	−0.0006 (1)	−0.0015 (1)	−0.0002 (2)	0.0001 (2)	−0.0008 (2)
D1-$MeNO_2$	228	0.0072 (5)	0.0085 (3)	−0.0151 (1)	−0.0088 (7)	0.0243 (3)	0.0259 (3)
O1-$MeNO_2$	228	0.0051 (1)	−0.0048 (1)	0.0071 (1)	−0.0005 (3)	0.0000 (2)	0.0000 (1)
H-UPA	150	−0.0011 (1)	−0.0028 (1)	0.0009 (1)	0.0018 (3)	−0.0007 (2)	−0.0006 (1)
H-UPA	350	−0.0070 (3)	0.0046 (3)	−0.0005 (2)	0.0043 (6)	0.0033 (5)	−0.0007 (4)

covariance matrix, $\mathbf{U}_{\rm anharm}$, can be calculated in a similar fashion:

$$U^{ij}_{\rm anharm} = \langle (u_i - \langle u_i \rangle)(u_j - \langle u_j \rangle) \rangle = \langle \Delta u_i \Delta u_j \rangle. \qquad (9)$$

The sampling error in these quantities can be estimated using the central limit theorem (Allen & Tildesley, 1989), while correlations in the data can be accounted for using the blocking method (Flyvbjerg & Petersen, 1989). Fortran code was written to calculate the true, anharmonic values of the mean and variance using equations (8) and (9); the parameters determined in this fashion will be referred to as the 'numerical' mean and variance. In performing the analysis the full space-group and translational symmetry of the simulation cell (which in some cases represented a supercell of the crystallographic unit cell) were used to maximize the data set.

The MD trajectories were used to determine numerical PDFs by 'binning' each position adopted by an atom over the course of the MD simulation. This procedure results in three-dimensional histograms, which when normalized yield a PDF. The distribution consisted of 100^3 points for ammonia and UPA. Two-dimensional and one-dimensional PDFs that represent marginals of the three-dimensional PDF with one or two variables integrated out were determined in a similar fashion. Where the three-dimensional distribution was determined 'on the fly', the one-/two-dimensional distributions were determined by numerical integration of the three-dimensional distribution (over the full range of the numerical distribution) and re-normalization. For nitromethane the 'on-the-fly' distributions were determined with a histogram of 200^3 points, which was then orientated and re-binned to produce a coarser PDF (100^3 points), which was found to be more than sufficient for fitting the PDFs. An artefact of this process was the introduction of some high-frequency noise into the data sets. As the analytical PDFs can only model the low-frequency features of the distribution, the quality of the fit of these re-orientated data sets is not affected relative to the un-orientated data sets. However, to aid visualization, a low-pass Fourier filter may be applied to the numerical distribution.

In most cases the histograms were determined in the coordinate system of the harmonic approximation so the distributions were centred on the numerical mean and orientated so that $U^{ij}_{\rm anharm} = 0 \quad (i \neq j)$, with the longest principal axis directed along the x axis and the shortest along the z axis. Visualization and manipulation of the data are greatly simplified in this coordinate system.

The versatile program *Mathematica* (Wolfram Research Inc., 2007) was used to determine the best fits of the harmonic and anharmonic models to the numerical PDFs in real space. In addition, it was also used to determine the maxima (*i.e.* $\mathbf{r}_{\rm p}$) of the anharmonic distributions and the uncertainties and correlations (analytically) for the refined parameters.

4. Results and discussion

4.1. Harmonic and anharmonic means and variances

The Gaussian approximation of thermal motion has remained the most widely used method in crystallography because of its simplicity. The functions involved can be readily implemented in refinement software and the values of parameters determined can be interpreted with ease. It determines the mean (or time-averaged) position of the atom and the variance of the PDF. The variance is in part a measure of the size or extent of the PDF and therefore how far the atom moves from the mean position. In contrast, the statistical methods for incorporating anharmonicity involve much more complex equations and yield parameters such as quasi-moments that are difficult to relate to the physics or chemistry of the crystal in question. However, it is important to note that the mean and variance determined from a harmonic fit may differ from the true mean, $\langle u \rangle$ (the average position adopted by the atom over time), or variance, $\langle u^2 \rangle$ (the average square displacement from the mean), that is obtained from the true probability function, P, by integration:

$$\langle u^n \rangle = \int_{-\infty}^{\infty} u^n P(u)\, {\rm d}u. \qquad (10)$$

This is the case both for real-space (as done here) and reciprocal-space fitting of data. If the distribution is centred on the mean then $\langle u^1 \rangle = 0$ and $\langle u^2 \rangle$ is the variance. Some forms of anharmonic PDF are formulated so that the mean and variance are determined directly as the parameters of the harmonic part of the PDF.

Table 2
Harmonic, anharmonic and numerical values of U^{11} (in Å^2) for a series of atomic PDFs.

The indices of the nitromethane atoms refer to those used in the previous crystallographic studies of nitromethane (Jeffrey *et al.*, 1985).

Atom	T (K)	Harmonic	Anharmonic	Numerical
D-ND_3	180	0.03443 (6)	0.03575 (17)	0.03555 (3)
C-$MeNO_2$	15	0.00980 (3)	0.00978 (6)	0.00968 (1)
D1-$MeNO_2$	15	0.08342 (10)	0.07626 (20)	0.08065 (4)
O1-$MeNO_2$	15	0.01046 (3)	0.01049 (7)	0.01041 (1)
C-$MeNO_2$	228	0.04646 (5)	0.04692 (10)	0.04699 (5)
D1-$MeNO_2$	228	0.38550 (60)	0.25840 (40)	0.28068 (15)
O1-$MeNO_2$	228	0.07803 (7)	0.07871 (10)	0.07929 (8)

There have only been a few studies that have considered the difference between the harmonic and anharmonic values. Scheringer (1986) has shown that the differences between the harmonic and anharmonic mean bond lengths (*i.e.* the distance between the mean position of two atoms) of urea were of the order of 0.002–0.01 Å at room temperature. Assessing the significances of such differences is difficult in experiments owing to the non-convergent nature of many of the anharmonic methods used to determine such mean positions. By fitting the numerical MD-derived data sets in real space we can compare the anharmonic and harmonic means to the true value for the first time, without resorting to any approximations or assumptions.

4.1.1. Means. The difference between the mean obtained from harmonic and anharmonic models can be assessed by fitting the MD-derived PDFs to a harmonic PDF and to a GC PDF, respectively. These values can then be compared to the 'true' numerically determined values. Because of the centring of the numerical data sets the mean position refined by the harmonic and anharmonic PDFs corresponds to the difference between the different models and the 'true' numerically evaluated value. The GC series PDF is of the same form as equation (4). The third-order Hermite polynomials, H_{ijk}, used in this equation are centred on the mean position (where $\mathbf{u} = 0$ by definition). This expression results in the correct definition of the Debye–Waller factor for use in crystallographic refinements but for refinements of the numerical PDFs in real space it is necessary to include parameters for the mean in the polynomials.

Table 1 lists the mean positions (x, y, z) obtained by fitting the numerical PDFs for ammonia, nitromethane and UPA to a Gaussian PDF and a third-order GC series PDF. In general, the difference between the numerical (*i.e.* the raw MD) and harmonic means is less than 0.01 Å. Only for the D1 atom of nitromethane at the high temperature of 228 K is the value larger than this. For all but one case, the use of the GC series improves the agreement between the fitted mean and the numerical value. For the D1 atom at 228 K the agreement is made worse but the strong, curvilinear anharmonicity of the D atoms in nitromethane is difficult to represent adequately with just a third-order GC series (see Reilly *et al.*, 2011, for more details on the importance of the fourth-order terms). The number of values in Table 1 is not enough to draw any wide conclusions but it clearly illustrates that differences between the parameters can occur and they can be more significant than typical experimental uncertainties. It is encouraging that the differences are relatively small (even at high temperatures) and that, in general, the anharmonic model improves the agreement. This means that the harmonic model is adequate for getting a good picture of the structure and estimating the mean positions. Even the values for the D1 atom of nitromethane are small in comparison to the equilibrium positional corrections that are discussed elsewhere (Reilly *et al.*, 2010*b*).

4.1.2. Variances. The variance of the PDF is, in principle, affected only by even-order terms in a polynomial expansion. A harmonic PDF models the atom as having a quadratic energy surface. Higher-order polynomial terms permit the potential to be broader or steeper than this. For a nearly Gaussian distribution the three H_{iiii} $(i = 1, 2, 3)$ polynomials will have the most important contributions if these polynomials are directed along the principal axes of the system. This is because in this case they are functions of only a single variable and will therefore affect the 'peakedness' or kurtosis of the three orthogonal directions separately, in keeping with the approximation of independent variables implicit in the Gaussian approximation. A GC series with only the fourth-order diagonal, H_{iiii}, terms was used to fit a number of PDFs; the harmonic, anharmonic and numerical values of the largest of the three variances, U^{11}, are given in Table 2. In the case of the C and O atoms of nitromethane, small differences are seen between the harmonic and numerical values. The anharmonic fit improves the agreement in some cases but taking the uncertainties into account the discrepancies are not significant. However, for the D atoms of ND_3 and nitromethane large and significant differences are found between the harmonic and numerical values. For the D atom of ND_3 the anharmonic model reduces this difference to within error. The disagreement is particularly serious at 228 K for the D1 atom of CD_3NO_2 with the harmonic value being nearly 33% larger than the numerical value. The inclusion of the H_{1111} parameter does improve the agreement significantly. A full third- and fourth-order GC series would improve things further but higher-order terms might still be important. Fig. 2 shows the

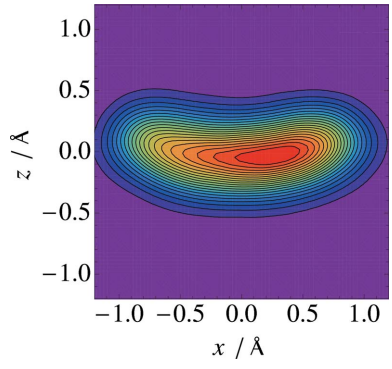

Figure 2
The xz two-dimensional PDF of the D1 atom of nitromethane at 228 K, showing the large degree of anharmonicity. (The x axis represents the longest principal axis of thermal motion, while z is the shortest axis.)

Table 3
Magnitudes of selected probable and equilibrium positional corrections to the numerical mean positions of nitromethane atoms along with the migratory proton of UPA, as determined from MD-derived numerical three-dimensional PDFs.

The GC series used are all third-order expansions except for the D atoms at 228 K, where it was necessary to add the diagonal H_{iiii} terms to the expansion (see Reilly *et al.*, 2011, for more details of why this is necessary).

Atom	T (K)	$\lvert\mathbf{r}_{p,GC}\rvert$ (Å)	$\lvert\mathbf{r}_{e}\rvert$ (Å)
Nitromethane			
D1	15	0.0238	0.1153
D1	228	0.1951	0.2569
D2	15	0.0314	0.0562
D2	228	0.2151	0.0570
C	15	0.0016	0.0365
C	228	0.0059	0.0963
UPA			
H	150	0.0225	0.0338
H	350	0.0190	0.0677

numerical two-dimensional xz PDF of the D1 atom at 228 K, which clearly illustrates the high degree of anharmonicity that leads to both the harmonic and anharmonic fits deviating from the true mean and variance values.

The data in Tables 1 and 2 highlight the importance of remembering that the mean and variance determined by the standard harmonic model are still primarily fitting parameters. While they may be and often are close to or within the error of the true mean or variance, their physical significance cannot be taken for granted. The experimental temperature dependence of the ADPs (Bürgi *et al.*, 2000) may give an indication of where this might be an issue, as an anharmonic dependence of the ADPs with temperature indicates the need for anharmonic refinement of the data.

4.2. Most probable and equilibrium structures

A single diffraction experiment tells us about the three-dimensional effective potential-energy surface that an atom experiences in the mean field of the motions and interactions of the other atoms in the system. The MD simulations allow us to determine each atom's mean and probable positions on this three-dimensional surface, while geometry optimizations yield the overall equilibrium geometry of the system on the full $(3N-3)$-dimensional potential-energy surface. In the following subsections the most probable and equilibrium positions determined for a few molecules are compared.

4.2.1. Nitromethane. Nitromethane is again an ideal system to compare and contrast the most probable and equilibrium positions. The large extent of the thermal motion and its strongly anharmonic nature ensures that limitations in the anharmonic model (*i.e.* how well the GC series fits the distribution) will not prevent comparison of the two types of position. Table 3 lists the magnitudes of the most probable and equilibrium corrections to the D1, D2 and C atoms at 15 and 228 K.

Considering only the most probable positions first, it can be seen that the correction increases with temperature as would be expected. This is also seen for the other atoms in nitromethane. The magnitude of the corrections is, in all cases, significant, even at the low temperature of 15 K. Comparing these values to the equilibrium corrections we can see that for the D1 atom and the C atom (and indeed the other atoms not presented in Table 3) the equilibrium correction is much larger than the probable correction. Only in the case of D2 at 228 K is the probable correction larger than the equilibrium one.

The difference between the equilibrium and probable positions is surprisingly large. Fig. 3 shows the numerical two-dimensional xz PDF of the D1 atom at 15 K with equilibrium and third-order GC probable positions marked. It is clear that the discrepancy is not a result of the probable position being wrong: the GC-determined maximum is clearly very close to the maximum of the numerial PDF, which in turn is far from the equilibrium position. A similar situation is found for the other D atoms as well, with the majority of the equilibrium correction being along the longest principal axis in each case. This implies that the methyl group rotates going from the equilibrium to probable structure. Corrections to the heavy atoms (*i.e.* the C atom) indicate that the whole molecule shifts its position when going from the equilibrium structure to the three-dimensional PDF average/most probable structure. The complexity of the $(3N-3)$-dimensional potential surface makes it hard to rationalize this behaviour and the analysis and evaluation of the $(3N-3)$-dimensional numerical PDF cannot easily be performed. The discrepancy of the D2 equilibrium correction being approximately the same at both high and low temperatures ($\sim$0.057 Å) may be explained by the 'static' equilibrium correction cancelling the part that arises from thermal motion.

While the most probable and equilibrium position corrections differ, this is mainly an issue of the whole molecule shifting or translating its position. The C—D1 bond length obtained from the most probable positions of the C and D atoms at 15 K is 1.0886 Å. This compares very well to the equilibrium value of 1.0880 Å, showing that while we do not get the absolute equilibrium position of the atom it is possible that the relative 1,2 and 1,3 positions, and therefore the most important structural information, may be correctly modelled by the most probable structure.

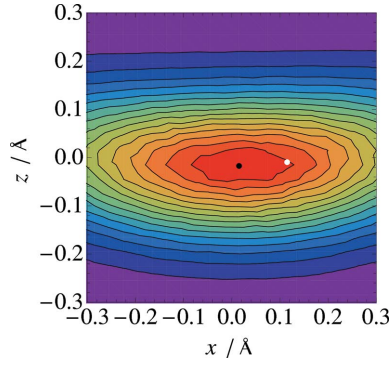

Figure 3
The xz two-dimensional PDF of the D1 atom of nitromethane at 15 K. The black dot indicates the third-order GC probable position, while the white dot is the equilibrium position.

4.2.2. Ammonia. For the D atom of ammonia the difference between the probable and equilibrium corrections is much better, with $|(\mathbf{r}_e - \mathbf{r}_p)| = 0.0046$ Å. The magnitudes of the corrections are much smaller than in nitromethane, with $|\mathbf{r}_e|$ being only 0.0176 Å. Taking the uncertainties in the positions and the nature of the fitting process into account, the differences between the probable and equilibrium positions are not that significant. The small size of the corrections may stem from the much higher symmetry of the ammonia crystal structure. This makes the potential- and free-energy surfaces less complicated as there are only four independent degrees of freedom.

4.2.3. UPA. The shape and temperature dependence of the free-energy surface of the migratory proton have been of considerable interest in recent publications (Wilson & Morrison, 2002; Morrison *et al.*, 2005; Fontaine-Vive *et al.*, 2006). The MD simulations confirm, as we might expect from its behaviour, that the migratory proton has a non-Gaussian PDF. At 150 K the distribution is asymmetric or egg shaped, whilst at 350 K the distribution appears to be more Gaussian-like (see Fig. 4). Note that the distributions feature far more high-frequency noise compared to the nitromethane and ammonia distributions because the size and nature of the nitromethane and ammonia simulations permitted longer trajectories and larger data sets. In general, DFT–MD simulations will produce noisier distributions than those of force-field simulations because of the much larger data sets available with the faster force-field methods. It is evident from Fig. 4 that the effective potential-energy surface does have some temperature dependence. This suggests that the full free-energy surface will have some dependence too. It is interesting to consider whether the three-dimensional PDF appears more Gaussian-like at higher temperature owing to a fundamental change in the underlying energy surface. As before, a third-order GC series was used to model the atomic PDFs.

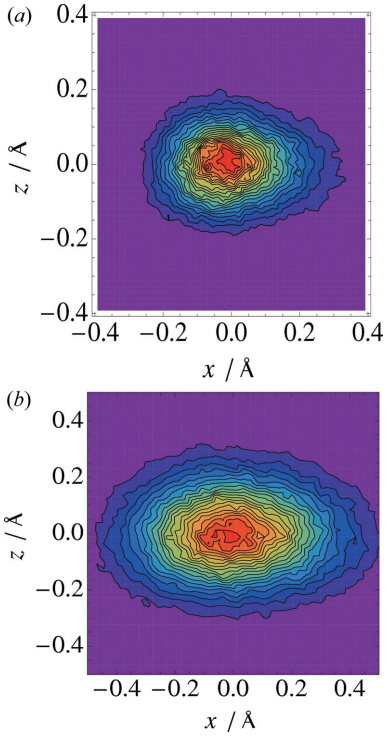

Figure 4
The *xz* two-dimensional numerical PDFs of the migratory proton of UPA at (*a*) 150 K and (*b*) 350 K. The proton migrates towards the left-hand side of the figures. (The *x* axis represents the longest principal axis of thermal motion, while *z* is the shortest axis.)

Considering first the H-atom positional corrections, the magnitude of the equilibrium correction is 0.0338 Å, while the magnitude of the most probable correction position is 0.0225 Å. At 350 K the equilibrium correction for the H atom is 0.0677 Å but the most probable correction is only 0.0190 Å. Just as for nitromethane, the most probable and equilibrium corrections do not produce the same structure. The slightly smaller, most probable correction at the higher temperature is possibly an indication of less anharmonicity in the high-temperature PDF.

A more direct measure of how the system changes would be to compare the different bond lengths at the two temperatures. As the two simulations were performed with different cell vectors, the effective equilibrium positions may be different. However, the equilibrium O—H and O···H distances are not significantly affected by this; the former is 1.0990 Å at 150 K and 1.0968 Å at 350 K. In contrast, the most probable bond lengths do change with r_p O—H being 1.1089 Å at 150 K and 1.1468 Å at 350 K. Similarly, r_p O···H is 1.3289 Å at 150 K but 1.2981 Å at the higher temperature. The small difference in the equilibrium distances with temperature indicates that the potential-energy surface does not change significantly as a result of the cell expansion. If the same were true for the free-potential surface we would expect the most probable interatomic distances to be very similar at the two temperatures. However, they are not. Thus an increase in temperature must result in a flattening of the free-energy potential surface such that r_p O—H lengthens by almost 0.04 Å. This is in accordance with previous findings (Fontaine-Vive *et al.*, 2006).

4.3. Experimental equilibrium structures

It is always important to ask how much information we can extract from experimental methods. The current approach to determining experimental equilibrium structures relies on MD simulations, which can have a significant cost in terms of computational time and effort. While this might change with better theoretical methods and computational hardware and software, for the present it would be highly desirable to determine the equilibrium geometry directly from the experiment without the need for time-consuming MD simulations.

Reconstructing the full $(3N - 3)$ PDF or potential-energy surface from diffraction experiments would be very difficult to achieve. The missing information in a single diffraction experiment is the correlations between the three-dimensional marginal PDFs. Bürgi & Capelli (2000) have shown that some of the required correlations between the atomic PDFs, namely the intramolecular ones, can be obtained by multi-

temperature experiments but for a truly anharmonic system the number of parameters and potential for non-trivial couplings would greatly limit such an approach.

The results of the MD simulations suggest that the most probable position will not be a reliable indication of the equilibrium position of an atom. In some cases the relative positions of the atoms may give reliable estimates of equilibrium bond lengths, as was seen for the r_p C—D distances in nitromethane. In UPA this is not the case and it is clear that more studies will be required to understand the types of systems where large differences are likely to be found. Nevertheless, the use of anharmonic Debye–Waller factors and the determination of most probable positions should be encouraged. Not only will they reduce *R* factors and improve reliability of structures but they can also shed more light on the complex behaviour of atoms in some systems. In using anharmonic PDFs to study the proton's migration in UPA we learn a considerable amount of information. Previous work on UPA has used theory to shed further light on this behaviour (Wilson & Morrison, 2002; Morrison *et al.*, 2005; Fontaine-Vive *et al.*, 2006). However, if sufficient quality data are collected then the anharmonicity and temperature dependence of the potential could be studied experimentally.

5. Conclusion

There are a variety of different types of atomic positions that can be determined by a refinement or are otherwise of interest to structural scientists. The comparison of different types of structure parameters has been made possible for the first time using MD-derived data. From MD simulations of three different molecules, numerical probability density functions have been determined and then used to assess the meaning of the different parameters. The effect of different models of thermal motion on the value of the mean atomic positions and the refined covariance matrix has been studied. For systems with strong anharmonicity the values fitted to harmonic models deviate significantly from the 'true' statistical value evaluated numerically from the model data sets, affirming that the physical meaning of the fitting parameters is not always assured.

In addition, the simulations show large differences between the equilibrium and fitted most probable positions of atoms for some molecules. This suggests that experimentally determined probable positions will not give suitable estimates of equilibrium positions. However, the most probable C—D distances in nitromethane are far closer to their equilibrium values than their mean bond lengths. In UPA the most probable and equilibrium bond lengths do not agree but their comparison does provide more insights into the nature of the proton migration in this system.

It should be stressed that the present study is limited to only a few compounds. A broader study is necessary for firm conclusions, and the greater use of anharmonic Debye–Waller factors in experimental refinements would also shed further light on the differences between anharmonic and harmonic means and variances. From a theoretical standpoint, more insights into the differences between the probable and equilibrium positions could be garnered by combining the three-dimensional MD-derived PDFs to obtain the full $(3N-3)$ PDF, whose maximum should correspond better with the equilibrium geometry. However, this would be quite difficult to achieve in practice for nitromethane and UPA but for systems like UPA the nine-dimensional PDF of the O—H···O fragment may prove just as useful. This sort of approach and application of the method to a wider range of systems is the focus of our continuing work in this area.

The EaStCHEM Research Computing Facility (http://www.eastchem.ac.uk/rcf) is acknowledged for its provision of computational resources. This facility is partially supported by the eDIKT initiative (http://www.edikt.org.uk). The DFT–MD simulation of UPA made use of the Edinburgh Parallel Computing Centre (http://www.epcc.ed.ac.uk). AMR acknowledges the School of Chemistry, Edinburgh, for funding a studentship, and CAM the Royal Society for the award of a University Research Fellowship.

References

Allen, M. P. & Tildesley, D. J. (1989). *Computer Simulation of Liquids.* Oxford University Press.
Bürgi, H. B. & Capelli, S. C. (2000). *Acta Cryst.* A**56**, 403–412.
Bürgi, H. B., Capelli, S. C. & Birkedal, H. (2000). *Acta Cryst.* A**56**, 425–435.
Car, R. & Parrinello, M. (1985). *Phys. Rev. Lett.* **55**, 2471–2474.
CPMD (2008). *CPMD*. Version 3.13.2. IBM Corp. and MPI für Festkörperforschung, Stuttgart.
Cruickshank, D. W. J. (1956). *Acta Cryst.* **9**, 757–758.
David, W. I. F., Ibberson, R. M., Jeffrey, G. A. & Ruble, J. R. (1992). *Phys. B*, **180–181**, 597–600.
Debye, P. (1913). *Ann. Phys.* **348**, 49–92.
Feynman, R. P. & Hibbs, A. R. (1965). *Quantum Mechanics and Path Integrals.* New York: McGraw-Hill Book Company.
Flyvbjerg, H. & Petersen, H. G. (1989). *J. Chem. Phys.* **91**, 461–466.
Fontaine-Vive, F., Johnson, M. R., Kearley, J., Howard, J. A. K. & Parker, S. F. (2006). *J. Am. Chem. Soc.* **128**, 2963–2969.
Hoover, W. G. (1985). *Phys. Rev. A*, **31**, 1695–1697.
Jeffrey, G. A., Ruble, J. R., Wingert, L. M., Yates, J. H. & McMullan, R. K. (1985). *J. Am. Chem. Soc.* **107**, 6227–6230.
Johnson, C. K. (1969). *Acta Cryst.* A**25**, 187–194.
Kuhs, W. F. (1983). *Acta Cryst.* A**39**, 148–158.
Kuhs, W. F. (1992). *Acta Cryst.* A**48**, 80–98.
Kuznetsov, P. I., Stratonovich, R. L. & Tikhonov, V. I. (1960). *Theory Probab. Appl.* **5**, 80–97.
Morrison, C. A., Siddick, M. M., Camp, P. J. & Wilson, C. C. (2005). *J. Am. Chem. Soc.* **127**, 4042–4048.
Nosé, S. (1984). *J. Chem. Phys.* **81**, 511–519.
Perdew, J. P., Burke, K. & Ernzerhof, M. (1996). *Phys. Rev. Lett.* **77**, 3865–3868.
Reilly, A. M., Habershon, S., Morrison, C. A. & Rankin, D. W. H. (2010*a*). *J. Chem. Phys.* **132**, 134511.
Reilly, A. M., Habershon, S., Morrison, C. A. & Rankin, D. W. H. (2010*b*). *J. Chem. Phys.* **132**, 094502.
Reilly, A. M., Middlemiss, D. S., Siddick, M. M., Wann, D. A., Ackland, G. J., Wilson, C. C., Rankin, D. W. H. & Morrison, C. A. (2008). *J. Phys. Chem. A*, **112**, 1322–1329.
Reilly, A. M., Morrison, C. A., Rankin, D. W H & McLean, K. R. (2011). *Acta Cryst.* A**67**, 346–356.
Reilly, A. M., Wann, D. A., Morrison, C. A. & Rankin, D. W. H. (2007). *Chem. Phys. Lett.* **448**, 61–64.

Scheringer, C. (1986). *Acta Cryst.* A**42**, 356–362.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Schomaker, V. & Trueblood, K. N. (1998). *Acta Cryst.* B**54**, 507–514.
Tanaka, K. & Marumo, F. (1983). *Acta Cryst.* A**39**, 631–641.
Trevino, S. F., Prince, E. & Hubbard, C. R. (1980). *J. Chem. Phys.* **73**, 2996–3000.
Troullier, N. & Martins, J. L. (1991). *Phys. Rev. B*, **43**, 1993–2006.
Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.
Waller, I. (1923). *Z. Phys. A*, **17**, 398–408.
Willis, B. T. M. (1969). *Acta Cryst.* A**25**, 277–300.
Wilson, C. C. (2001). *Acta Cryst.* B**57**, 435–439.
Wilson, C. C. & Morrison, C. A. (2002). *Chem. Phys. Lett.* **362**, 85–89.
Wilson, C. C., Shankland, K. & Shankland, N. (2001). *Z. Kristallogr.* **216**, 303–306.
Wolfram Research Inc. (2007). *Mathematica*. Version 6.0. Wolfram Research Inc., Champaign, Illinois.
Zucker, U. H. & Schulz, H. (1982). *Acta Cryst.* A**38**, 563–568.