Acta Cryst A FOUNDATIONS ADVANCES

ISSN 2053-2733

Computational analysis of thermal-motion effects on the topological properties of the electron density

J. Robert Michael and Tibor Koritsanszky*

Center for Computational Science, Middle Tennessee State University, 1301 E. Main Street, Murfreesboro, TN 37128, USA. *Correspondence e-mail: tibor.koritsanszky@mtsu.edu

Received 5 December 2014
Accepted 20 January 2015

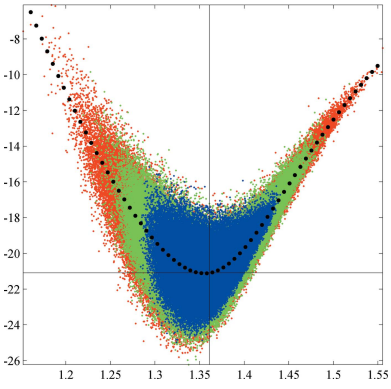
Keywords: electron density; bond topological properties; mean-square displacement amplitudes; probability distribution function; standard uncertainty; equilibrium geometry.

The distributions of bond topological properties (BTPs) of the electron density upon thermal vibrations of the nuclei are computationally examined to estimate different statistical figures, especially uncertainties, of these properties. The statistical analysis is based on a large ensemble of BTPs of the electron densities for thermally perturbed nuclear geometries of the formamide molecule. Each bond critical point (BCP) is found to follow a normal distribution whose covariance correlates with the displacement amplitudes of the nuclei involved in the bond. The BTPs are found to be markedly affected not only by normal modes of the significant bond-stretching component but also by modes that involve mainly hydrogen-atom displacements. Their probability distribution function can be decently described by Gumbel-type functions of positive (negative) skewness for the bonds formed by non-hydrogen (hydrogen) atoms.

1. Introduction

Bond topological properties (BTPs), such as the location of the bond critical point (BCP, $\mathbf{r}_{\rm BCP}$), the electron density (ED, $\rho_{\rm BCP}$), the Laplacian ($\nabla^2\rho_{\rm BCP}$) and the eigenvalues of the Hessian of the ED at the BCP, are well established indicators of the nature of the chemical bond and atomic interaction in general (Bader, 1990). They are widely regarded to be accessible not only *via* quantum-chemical calculations but also through analysis of high-resolution X-ray Bragg structure factors (SFs). In fact, indispensable parts of modern X-ray charge-density reports are lengthy tables posed to match the BTPs of 'the static' ED extracted from a set of diffraction data (experimental ED) with those computationally derived from an approximate solution to the Schrödinger equation of the isolated molecule or the periodic system (theoretical ED). Obviously, such a collation lacks the *bona fide* verification power due to the lack of a well established link between what is claimed to be 'the' experimental density (a particular least-squares solution of SF fitting utilizing a particular ED model) and what is elected to be 'the' theoretical density (a certain quantum-chemical method utilizing a certain basis set). Spuriously, the majority of X-ray charge-density studies reveal from good to excellent agreement between the BTPs obtained by the two methods, especially for molecular crystals (Koritsanszky & Coppens, 2001; Gatti, 2005). It is most likely due to these findings that the crystallographic community tends to misuse these properties, to the extent that comparison between theoretical and experimental BTP values has become an ultimate and almost mandatory means to validate experimental EDs.

Such assertions for verification purposes should be convincingly substantiated by standard uncertainties (s.u.'s) of the experimental BTPs. These quantities are, however,



notoriously underestimated – typically rounded to the hundredth (e Å^{-3})/tenth (e Å^{-5}) place for the ED/its Laplacian at the BCP – based on the popular pseudoatom model (Hansen & Coppens, 1978). Uncertainties output by *XD* (Volkov *et al.*, 2006), for example, are calculated on the basis of a linear error-propagation scheme that includes only the least-squares (LS) variances of multipole parameters but ignores the ED's (non-linear) dependence on other refined parameters, such as nuclear positions and radial screening factors. A more delicate issue of this over-simplified error estimation stems from the often-overlooked fact that the target of the LS fit is, of course, a set of SFs associated with the space–time average ED (a dynamic ED at best) and not the static ED. However, for the evaluation of the uncertainty of the latter property, the correlations between static (nuclear positions, multipole populations and radial screening parameters) and dynamic (mean-square displacement amplitudes, MSDAs) variables are completely ignored which makes more of an intuitive than common sense, since these two sets of parameters are jointly refined against experimental SFs and, thus, their LS estimates are inevitably linked. In the course of a pseudoatom refinement, for example, the correlations between quadrupole populations and anisotropic displacement parameters (ADPs) typically settle in the range of 70–80%, which is a lucid warning about significant linear dependence of 'basis functions' comprising the scattering model fitted to the X-ray data (putting it simply, the 'knowledge' of one set of parameters fixes the other set within a 70–80% certainty). In the presence of such high correlations, even a modest attempt to distinguish between static and dynamic variables and their properties is quite an obscure endeavour. Another distinct issue is the error estimation of the location of a BCP, since it is commonly derived *via* an iterative numerical procedure and thus analytic error propagation schemes cannot be utilized. It is noteworthy to mention in this respect that even a narrow dispersion of the BCP location can transform a high uncertainty in the Laplacian at the BCP [especially in polar bonds for which the Laplacian can exhibit dramatic change in the vicinity of the BCP (Gatti *et al.*, 1992)]. In summary, error estimation of these 'experimental' properties is far from being trivial and the issue deserves more attention today than it has ever received.

The quantum-chemical approach designates the equilibrium molecular geometry to the global minimum of the electronic energy as an approximate function of the nuclear geometry that is numerically but self-consistently evaluated. According to the Hohenberg–Kohn (HK) theorem, the nuclear configuration (external potential) uniquely determines the ED (Hohenberg & Kohn, 1964). This also implies that to each nuclear configuration visited during nuclear vibrations there belongs one and only one ED. The intrinsic connection between the nuclear potential and the ED is only formally, but neither directly nor self-consistently, imposed on X-ray data fitting. The commonly used SF models all subsume the harmonic convolution approximation (one-centre partitioning and rigid following) (Debye, 1913; Waller, 1923), according to which the thermally averaged ED is deducible from that of a hypothetical stationary ED corresponding to the equilibrium nuclear configuration, which is however not known *a priori*, but to be derived simultaneously with the ED in the course of data fitting, and thus it is contingent upon the adequacy of the ED model in use. While it appears to be unavoidable, this approach obviously violates both the Born–Oppenheimer (BO) approximation (Born & Oppenheimer, 1927) and the HK theorem. Consequently, the experimental nuclear geometry, the static ED and the nuclear ADPs associated with it have a rather fragile connection. The diffraction geometry, corresponding to the converged LS data fit, is usually taken for granted as the equilibrium geometry, that is, the thermally averaged locations of the nuclei are assumed to form the stationary nuclear state characterized by vanishing forces acting on each nucleus. Quite obviously, this criterion cannot be met even if the scattering model applied perfectly accounts for the diffraction data. Simply putting it, X-ray-diffraction-based molecular geometries cannot trivially be identified with stationary isolated-molecule geometries, and this is not purely due to intermolecular interactions governing solid-phase formation.

This computational study is tailored for estimating the s.u.'s of BCP locations and the BTPs of the molecular ED due to internal harmonic nuclear vibrations, or in other words, due to the distribution of nuclear positions of known s.u.'s. Since the ED is a unique functional of the external potential (electron–nuclei Coulomb potential) and since the nuclear geometry can be identified as the collection of (3, −3) critical points of the ED, the analysis presented here is simply about revealing correlations between different types of critical points of the ED of a single molecule in thermal equilibrium with its environment and treated within the framework of stationary quantum mechanics. From the crystallographer's point of view, this computational experiment is aimed at elucidating correlations of errors in local ED properties in the hope of triggering progress in our error-estimation practices. While our most recent study validates the convolution approximation (Michael & Koritsanszky, 2014), the current contribution is devoted to an exploratory statistical analysis of how the local topology of the ED is affected by intramolecular nuclear vibrations that are not directly accessible *via* Bragg scattering.

2. Density sampling

Within the harmonic approximation to thermal motion, the nuclear displacements follow normal distributions, thus providing a mathematically straightforward and a physically plausible way to generate large quantum-chemical samples of local properties of static EDs from approximate solutions of the isolated-molecule clamped Hamiltonians (with well defined nuclear arrangements). In pursuit of this goal, we generate a large ensemble of geometries (M = 500 000 nuclear configurations) consistent with the harmonic internal vibrations of the formamide molecule. To each nuclear configuration of this ensemble, an approximate 'single-point' wavefunction is calculated and the topology of the corresponding static ED is derived (Volkov *et al.*, 2009) to obtain an

Table 1
Internal normal modes, wavenumbers (ν) and the corresponding mean-square displacements (σ) of the formamide molecule.

Units for ν, cm^{-1}, and σ, 10^6 Å^2.

Mode	1 NH_2 wagging	2 NCO bend	3 NH_2 torsion	4 CH OP bend
ν	216.8	567.6	650.8	1048.3
σ	63695	12176	21368	10340
Mode	5 NH_2 IP bend	6 CN stretch	7 CH bend	8 NH_2 scissor
ν	1055.3	1266.8	1422.4	1620.1
σ	8948	4743	9852	8434
Mode	9 CO stretch	10 CH stretch	11 NH_2 S stretch	12 NH_2 AS stretch
ν	1817.0	2921.7	3582.1	3720.0
σ	1247	5320	4503	4101

ensemble of static BTPs, which forms the basis of the subsequent statistical analysis.

All calculations were performed at the B3LYP/6-311G** level of theory (Becke, 1988) using the *Gaussian09* program package (Frisch *et al.*, 2009) and locally developed computer codes. First, the MSDA matrix (**U**) was generated *via* harmonic vibration analysis of the optimized planar molecule (*xy* plane),

$$\mathbf{U} = \langle \bar{\mathbf{u}}\mathbf{u} \rangle_T = \mathbf{L}\boldsymbol{\sigma}\bar{\mathbf{L}}, \tag{1}$$

where $\mathbf{u} = \mathbf{R} - \mathbf{R}^0$ is the $3N$ row vector ($\bar{\mathbf{u}}$ being its transpose column vector) of Cartesian nuclear displacements relative to the equilibrium positions $\mathbf{R}^0$ and $\boldsymbol{\sigma} = \{\sigma_{j=1,3N}\}$ is the diagonal (eigenvalue) matrix of the temperature-dependent normal-mode mean-square displacements,

$$\sigma_j = \frac{h}{8\pi^2 \nu_j} \coth\left(\frac{h\nu_j}{2k_B T}\right), \tag{2}$$

with ν_j being the frequency of the jth mode (Higgs, 1955). Table 1 lists the harmonic wavenumbers and the normal-mode mean-square displacements calculated for T = 23 K. This analysis leads to a $3N$-multivariate normal distribution of nuclear displacements:

$$N(0, \mathbf{U}) = (2\pi)^{-(3N/2)} |\mathbf{U}|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{u}\mathbf{U}^{-1}\bar{\mathbf{u}}\right). \tag{3}$$

The block-diagonal elements of the Cartesian MSDA matrix **U** represent the ADPs for each atom. A distorted geometry is expressed as a linear combination of nuclear displacements, $\mathbf{d}_j$, added to the equilibrium geometry, $\mathbf{R}^0$. For each normal mode of vibration, we have, from the *Gaussian09* calculations, the unit displacements in Cartesian coordinates as well as the amplitude of displacement (the standard deviation of a normal mode). Each perturbed geometry, $\mathbf{R}_i$, is to be considered as one of the many members of the ensemble $\{\mathbf{R}_i\}$ exemplifying harmonic oscillation about the equilibrium geometry resulting in $\mu(\{\mathbf{R}_i\}) = \mathbf{R}^0$ and $\sigma^2(\{\mathbf{R}_i\}) = \mathbf{U}$ (μ and σ denoting, respectively, the mean and standard deviation of the normal distribution). To create a perturbed geometry, $\mathbf{R}_i$, it is sufficient to sample each normal mode independently as the $3N$ vector of nuclear positions is a linear combination of displacements from the equilibrium geometry. Every displacement is associated with one of the $3N - 6$ normal modes, each defined by a normalized vector of displacement, $\mathbf{d}_j$, and the related mean-square amplitude, σ_j^2:

$$\mathbf{R}_i = \mathbf{R}^0 + \sum_{j=1}^{3N-6} r_\sigma^{(j)} \mathbf{d}_j, \tag{4}$$

where $r_\sigma^{(j)}$ is an independent random variable drawn from the normal distribution $N(0, \sigma_j)$.

It is important to note that all the molecular geometries generated in this way are checked whether or not they are topologically equivalent to the stationary (optimized) geometry, that is, if their ED had the same molecular graph (the same number and type of critical points). For the atom-numbering scheme we refer to the *ORTEP* (Burnett & Johnson, 1996) plot displayed in Fig. 1, which visualizes the nuclear (and BCP) thermal ellipsoids at the 90% probability level.

3. Validation of the sample

To check if the geometries are truly representative of harmonic vibrations and if the sample size is large enough to faithfully represent the population, it is sufficient to show that the ensemble mean (statistical average, $\mathbf{Q} = M^{-1}\sum_k^M \mathbf{R}_k$) reflects the equilibrium geometry and that the sample covariance $[\Gamma_{ij} = (M-1)^{-1}\sum_k^M (\mathbf{R}_i^{(k)} - \mathbf{Q}_i)(\mathbf{R}_j^{(k)} - \mathbf{Q}_j)]$ returns the input MSDA calculated from normal-mode analysis. In Table 2 the distances between the statistical mean and the equilibrium

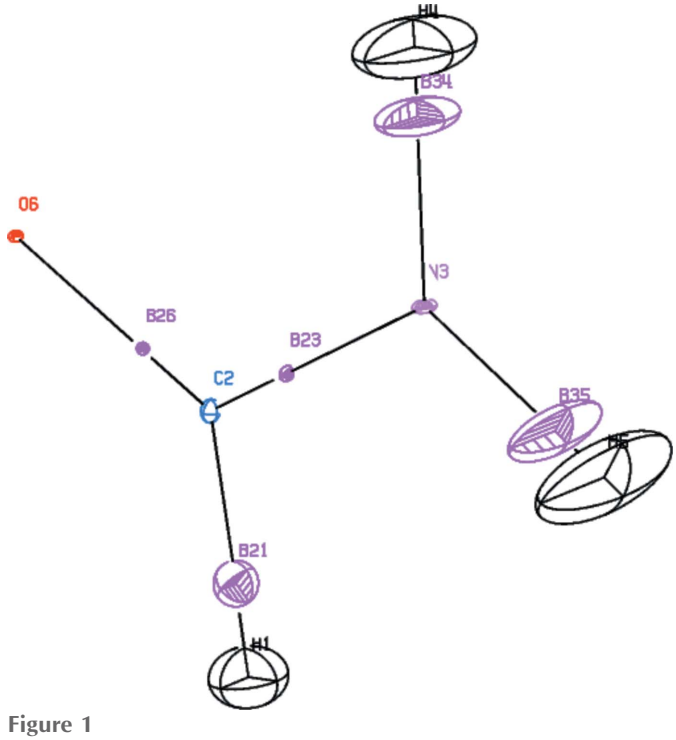


Figure 1
ORTEP plot of 90% probability thermal ellipsoids due to internal modes of vibrations of the formamide molecule. Ellipsoids drawn on the bond lines between the nuclei correspond to the covariance in the location of bond critical points (BNM, between nuclei N and M) of the electron density calculated over an ensemble of half a million nuclear configurations.

Table 2
Distances between static (geometry-optimized) and the sample-mean nuclear positions (units Å).

Nucleus	Distance
H1	4.46e-04
C2	2.94e-05
N3	3.37e-05
H4	3.67e-04
H5	1.71e-04
O6	2.89e-05

Table 3
Anisotropic displacement parameters for the nuclei from analytic (first row) and statistical (second row) frameworks (units 10^6 Å^2).

Atom	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
H1	10401	7150	10636	1993	0	0
	10395	7146	10604	1972	17	7
C2	592	997	693	−72	0	0
	594	998	691	−71	0	−1
N3	710	348	1081	200	0	0
	710	349	1081	202	−3	−1
H4	14314	5750	30386	3475	0	0
	14302	5751	30402	3478	−30	−8
H5	5334	12758	52542	1920	0	0
	5316	12758	52551	1911	−13	−92
O6	567	403	125	236	0	0
	567	403	125	236	0	0

Table 4
Distances between the equilibrium ($\mathbf{r}_{\rm BCP}$) and the sample-mean ($\bar{\mathbf{r}}_{\rm BCP}$) BCP locations (units Å).

BCP	Distance
C2—H1	1.15e-03
C2—N3	6.78e-03
C2—O6	1.45e-03
N3—H4	3.35e-03
N3—H5	5.09e-03

nuclear positions are listed, while Fig. 2 shows the Cartesian displacement components binned and plotted together with the fitted normal distribution for the N3 and O6 nuclei. The error for all coordinates of both fits remains under 0.5%, indicating that the sample distribution is a reliable representation of the population distribution. Similarly, low relative errors were obtained for all nuclei. Table 3 compares the ADPs and the sample variances of the nuclei; the largest difference (2.7×10^{-5}) represents the correlated motion of H4 and H5 nuclei in the z direction with a relative error of only 0.147%.

4. Distribution of bond critical points

A quite remarkable observation, undoubtedly emerging from the statistical analysis, is that the BCP locations follow normal distributions. Moreover, they appear to correlate in an elementary fashion with the distributions of the nuclei. The distances between the equilibrium ($\mathbf{r}_{\rm BCP}$) and the sample mean ($\bar{\mathbf{r}}_{\rm BCP}$) BCP positions are listed in Table 4. The standard deviations of these displacements are found by fitting

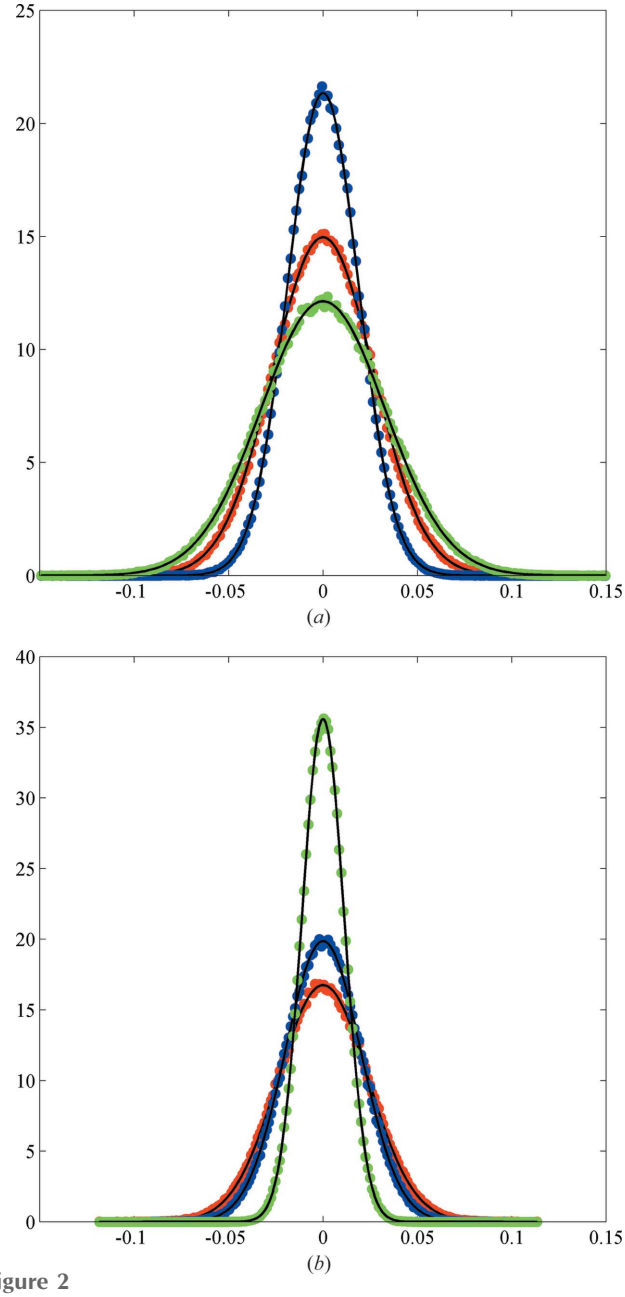


Figure 2
Cartesian displacements $(X, Y, Z = R, B, G)$ binned and plotted together with fitted Gaussian distributions (black) for the N3 (*a*) and O6 (*b*) nuclei.

univariate Gaussians to the Cartesian components of the $\mathbf{r}_{\rm BCP}$ ensemble. The largest per cent error of the fit (0.77%) appears for the X coordinate of the C2—N3 BCP location. Fig. 3 shows the comparison of the fitted Gaussians to the C2—N3 and the C2—O6 BCP distributions, while the sample-based covariance matrices (the 'ADPs of BCPs') are listed in Table 5. The corresponding thermal ellipsoids (also drawn in Fig. 1) show obvious correlations with those of the nuclei involved in the bond.

To find a relationship between the statistical ADPs (sample covariance) of BCPs and the ADPs of nuclei, we test the simplest scheme that assumes a linear dependence for a

general riding ED position (two-centre basis functions) suggested earlier (Scheringer, 1977*a*,*b*). Accordingly, we estimate the ADP of a BCP (U^*) as a linear combination of nuclear ADPs, $U^{i,i}$, and their correlations, $U^{i,j}$ ($U^{i,i}$ representing the ADPs of the ith nucleus and $U^{i,j}$ representing the correlation between the ith and jth nucleus):

$$U^* = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j U^{i,j} \qquad (5)$$

subject to the constraint of $\sum_{i=1}^{N} \alpha_i = 1$. The simplest approximation is restricted to $N = 2$, that is, only the nuclei of the atoms involved in the bond. We test three methods, each involving a single parameter, α_1 (since $\alpha_2 = 1 - \alpha_1$): method 1 utilizes the entire sample to find the least-squares estimate for α_1, method 2 defines α_1 as the distance of the BCP from one of the nuclei relative to the bond length,

$$\alpha_1 = \frac{|\mathbf{r}_{\rm BCP} - \mathbf{R}_1|}{|\mathbf{R}_2 - \mathbf{R}_1|}, \qquad (6)$$

while method 3 resembles method 2, but ignores the correlation term in equation (5). The difference between the latter two approaches can illuminate the importance of the off-diagonal elements of the MSDA (not accessible from Bragg diffraction data) in the construction of U^*. Based on the errors, listed in Table 6, both methods 2 and 3 provide a reasonable estimation for the BCP ADPs (though method 3 seems to fail for the polar C2—O6 bond).

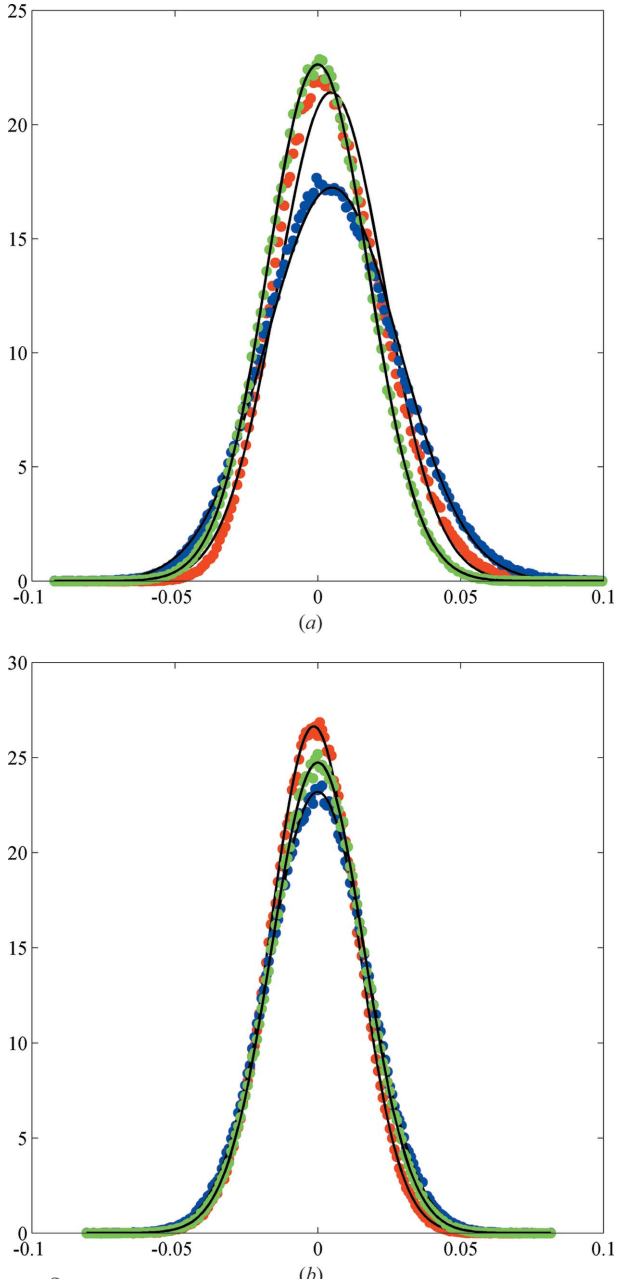


Figure 3
Cartesian displacements ($X, Y, Z = R, B, G$) of bond critical point locations binned and plotted alongside the fitted Gaussian distributions (black) for C2—N3 (*a*) and C2—O6 (*b*) bonds.

Table 5
Anisotropic displacement parameters of bond critical points (units 10^6 Å^2).

BCP	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
C2—H1	3579	3869	3160	155	5	2
C2—N3	347	535	310	113	0	0
C2—O6	224	296	260	17	0	0
N3—H4	7223	2370	14169	2027	12	−3
N3—H5	2332	5919	24867	1082	−3	−42

Table 6
Per cent errors in reproduction of the anisotropic displacement amplitudes of bond critical points using various methods described in the text.

Bonds	Method 1 error	Method 2 error	Method 3 error
H1—C2	3.22	4.75	10.02
C2—N3	7.79	17.05	9.23
C2—O6	2.12	3.43	22.20
N3—H4	0.26	0.37	2.06
N3—H5	0.09	0.18	1.56

5. Bond topological properties do not follow normal distributions

Fig. 4 plots the distribution of $\rho_{\rm BCP}$ and the $\nabla^2\rho_{\rm BCP}$ data for the C2—O6 and the C2—N3 bonds against the corresponding bond distances. To make these plots more comprehensible, the BTP sample of half a million is divided into three sets colour-coded depending on whether both (blue), only one (green) or none (red) of the nuclei involved in the bond are situated inside their associated 50% probability ellipsoids. An ellipsoid is defined as $\mathbf{u}\mathbf{U}^{-1}\overline{\mathbf{u}} = c^2$, where c = 1.538 for a probability of 50%. In general, the probability of finding an atom inside an ellipsoid is described as (Burnett & Johnson, 1996)

$$P(c) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^c r^2 \exp(-r^2/2)\,\mathrm{d}r$$
$$= \left(\frac{2}{\pi}\right)^{1/2} \left[-ce^{-c^2/2} + \left(\frac{\pi}{2}\right)^{1/2} \mathrm{Erf}\left(\frac{c}{2^{1/2}}\right)\right]. \quad (7)$$

The variation of the C2—N3 $\nabla^2\rho_{\rm BCP}$ sample distribution *versus* the bond distance is quite compelling. Fig. 5 also shows how each normal mode, acting in isolation, affects the distribution of the Laplacian data (black dotted curves labelled by integers corresponding to the normal-mode designation used in Table 1). Curve i displays $\nabla^2\rho_{\rm BCP}$ for nuclear positions corresponding to the ith normal-mode displacement vector ($\mathbf{d}_i$):

$$\mathbf{R}_i = \mathbf{R}^0 + s_i\mathbf{d}_i, \quad (8)$$

where

$$1 - 6\sigma_i \leq s_i \leq 1 + 6\sigma_i. \quad (9)$$

In Fig. 5 the overall $\nabla^2\rho_{\rm BCP}$ distribution for the C2—N3 bond is superimposed by distributions due to normal modes that have a significant (5*a*) and negligible (5*b*) bond-stretching component. It is obvious, and in line with earlier observations, that $\nabla^2\rho_{\rm BCP}$ is sensitive to bond-distance change (Gibbs *et al.*, 1998, 2008). What is however non-trivial and quite compelling is the degree to which $\nabla^2\rho_{\rm BCP}$ is affected by modes having minor bond-stretching contribution (mode numbers: 1, 2, 7 and 10). For example, the NH_2 wagging or the C—H stretching

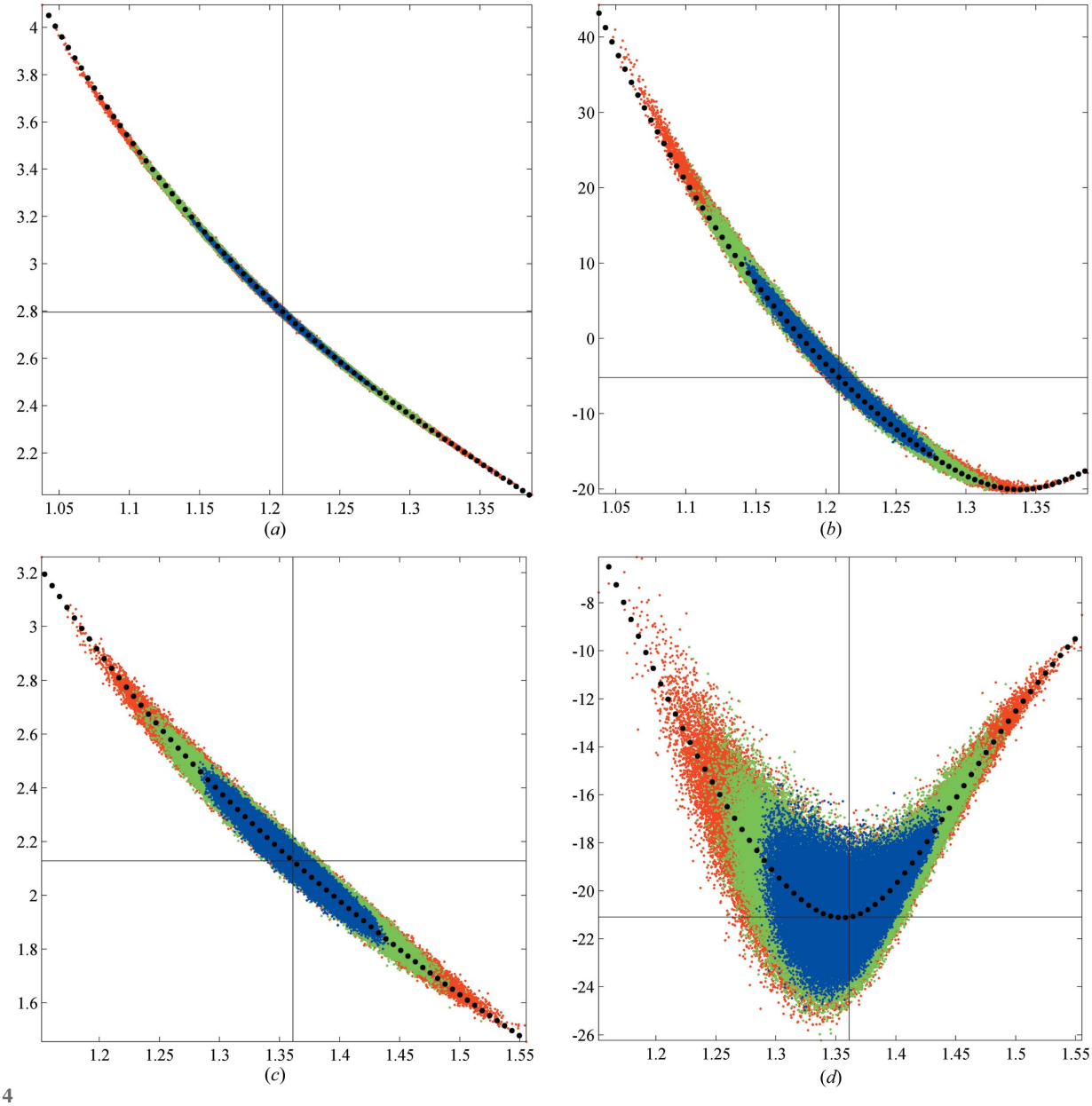


Figure 4
Ensemble distributions of $\rho(\mathbf{r}_{\rm BCP})$ (*a*), (*c*) and $\nabla^2\rho(\mathbf{r}_{\rm BCP})$ (*b*), (*d*) for the C—O (*a*), (*b*) and C—N (*c*), (*d*) bonds. Each point represents one of the 500 000 members of the ensemble and is classified by how many (0 – red, 1 – green, 2 – blue) of the nuclei are inside their associated 50% probability ellipsoid. The black dotted lines show the effect that bond-length alternation, in isolation, has on the BTPs.

modes hardly alter the C2—N3 bond length, yet the $\nabla^2\rho_{BCP}$ shows a crucial dependence on these modes; it varies in the range of 0.77 and 1.20 e Å^{-5} for the NH_2 and C—H modes, respectively, if the sample size is limited to geometries for which both the C2 and N3 nuclei are inside their respective 50% ellipsoid. On the other hand, the NH_2 torsion (3), the C—H bending (4) and the two N—H stretching (11, 12) modes have almost no effect on the C2—N3 Laplacian at the BTP.

For a detailed classification of the sampling distributions, we bin the centred ensemble (deviations of each BTP from the mean BTP where ρ_i is an element of the sample and ρ_Ψ is the stationary ED),

$$\delta = \{\delta_i\} = \{\nabla^2\rho_i(\mathbf{r}^i_{BCP}) - \nabla^2\rho_\Psi(\mathbf{r}_{BCP})\}, \quad (10)$$

and fit these binned data with appropriately parameterized (analytic) distributions to estimate the statistical moments up to fourth order. A visual inspection of these graphs (displayed in Fig. 6 for different bonds) leads to the conclusion that the BTPs are not normally distributed about their mean. This is quantitatively supported by the entries in Table 7 containing the first four standardized statistical moments and the per cent errors of fitting the binned data to normal distributions.

To model the binned BTP data, we adopt the Gumbel probability distribution function (PDF) (Gumbel, 1935):

$$P(z) = \frac{1}{\tau}\exp\{-[z + \exp(-z)]\}, \quad (11)$$

where $z = \beta[(q - \varepsilon)/\tau]$, with ε and τ being the shape parameters to be fitted, and

$$\beta = \text{Sign}[\text{Skewness}(P)]$$

is chosen so that the PDF is skewed positively (tails to the right) or negatively (tails to the left) for $\beta = 1$ or $\beta = -1$, respectively. The maximum-likelihood estimations of the Gumbel parameters (ε and τ) are listed in Table 8. BTPs (δ) for bonds formed by non-H atoms (C2—N3 and C2—O6) are shown in Fig. 6 to be leptokurtic distributed (tailed to the right) following the 'Gumbel-Max' ($\beta = 1$) distribution (maximum extreme value type I). The $\nabla^2\rho_{BCP}$ for bonds including H atoms are however platykurtic (tailed to the left) following the 'Gumbel-Min' ($\beta = -1$) function (minimum extreme value type I).

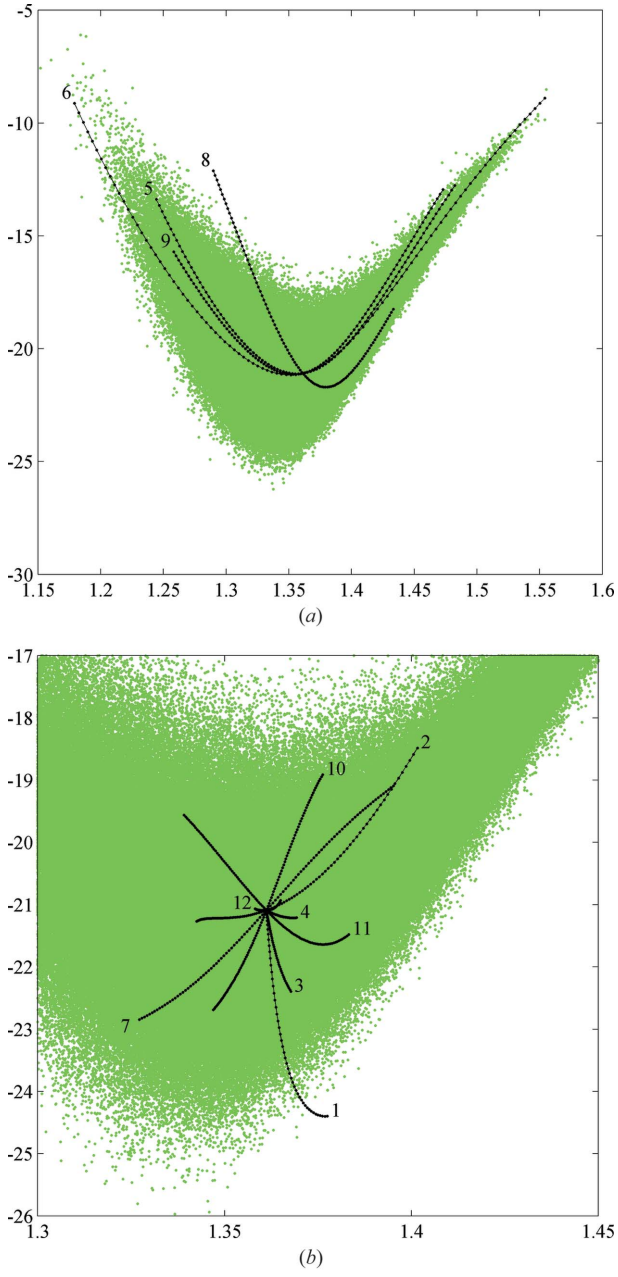


Figure 5
The ensemble distributions of $\nabla^2\rho(\mathbf{r}_{BCP})$ for the C2—N3 bond (green dots) superimposed by curves (black dots) representing alternations upon nuclear geometries due to individual normal modes numbered according to Table 1; (*a*) modes with notable bond-stretch contribution, (*b*) modes that have relatively small bond-stretch components.

Table 7
The first four statistical moments of distributions of bond topological parameters and per cent errors for fitting the binned data to normal distributions (units e and Å).

Bond	BTP	Mean	Variance	Skewness	Kurtosis	% Error
H1—C2	ρ_{BCP}	−0.02	0.09	0.55	3.51	1.67
	$\nabla^2\rho_{BCP}$	−0.06	53.11	−0.99	4.28	6.36
C2—N3	ρ_{BCP}	0.01	0.03	0.24	3.17	0.30
	$\nabla^2\rho_{BCP}$	0.95	3.44	0.93	4.57	4.81
C2—O6	ρ_{BCP}	0.00	0.04	0.37	3.28	0.74
	$\nabla^2\rho_{BCP}$	0.66	50.01	0.58	3.50	1.65
N3—H4	ρ_{BCP}	−0.11	0.16	0.60	3.74	1.81
	$\nabla^2\rho_{BCP}$	2.64	229.69	−1.16	4.98	8.55
N3—H4	ρ_{BCP}	−0.16	0.17	0.49	3.63	1.29
	$\nabla^2\rho_{BCP}$	4.33	228.25	−1.15	4.96	7.99

Table 8
Gumbel parameters fitted to the Laplacian distributions for each bond critical point.

BCP	β	ε	τ
C2—H1	−1	3.34	5.78
C2—N3	+1	0.20	1.46
C2—O6	+1	−2.05	6.33
N3—H4	−1	10.08	11.46
N3—H5	−1	11.58	11.48

To obtain a less demanding yet realistic uncertainty estimate for the BTPs, we rerun the simulation to generate a more feasible sample of only 2500 members using standard deviations of $10^{-3}/10^{-4}$ Å for the location of hydrogen/non-hydrogen nuclei. While the previously mentioned, larger, ensemble represents s.u.'s in nuclear positions due to thermal motion, this smaller ensemble may be regarded as s.u.'s due to the inability to perfectly fit the location of the nuclei to the data. Corresponding to this nuclear geometry ensemble, BTP samples were created *via* topological analyses of individual EDs to estimate also the s.u. of BCP locations. The $\rho_{\rm BCP}$ and the $\nabla^2\rho_{\rm BCP}$ values for each ED sample element were also evaluated at the BCP locations corresponding to the population means, so that we can estimate the accuracy (trueness) of the BTPs or, at least, find the range of their absolute error. Despite the much smaller sample size, the deviation of the sample mean from that of the population is found to be within the range of the nuclear s.u.'s for each BTP figure.

Table 9 shows that the s.u.'s of the BCP locations are on a par with the s.u.'s of the nuclear positions. It is noted again that the former quantities can be safely estimated from the latter ones using the method presented above [equations (5) and (6)]. The statistical standard deviations (σ) for $\rho_{\rm BCP}$ and $\nabla^2\rho_{\rm BCP}$ are also listed together with their range corresponding to the argument of the 'true' (population mean) BCP locations ($\mathbf{r}_{\rm BCP}$). For both properties, the standard uncertainty is almost always about an order of magnitude smaller than the range.

6. Conclusion

Statistical analyses of sufficiently large samples can lead to unbiased estimates of population parameters, as found in this study using a sample of half a million wavefunction-based ED topological figures. One of the results of stunning significance is that the BCP locations are normally distributed subject to harmonic oscillations of the nuclei about their mean (equilibrium) positions. Furthermore, it was shown that a simple formula, which combines the ADPs of nuclei involved in the bond, can quite adequately estimate the ADP of the BCP location. For a non-polar bond, for example, the BCP ADP is equal, to a fair approximation, to the average of the nuclear ADPs participating in the bond. This relationship has been applied to estimate the ADPs of the two-centre basis function products needed for density matrix fitting of structure factors (Coppens *et al.*, 1971). However, it is far from being trivial why a simple 'riding' model can predict the s.u. of such a 'complex'

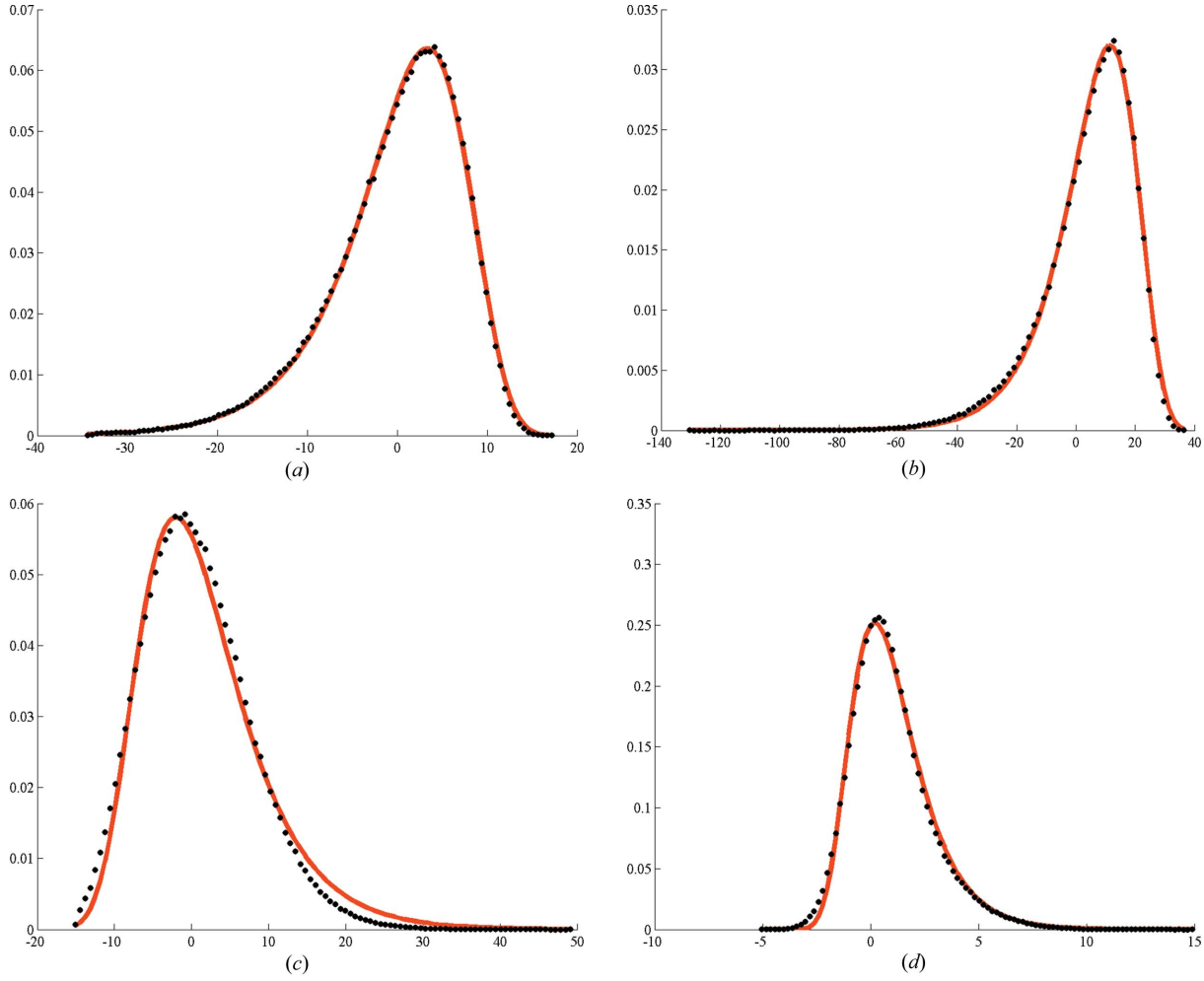


Figure 6
$\nabla^2\rho(\mathbf{r}_{\rm BCP})$ sample data binned and fitted with Gumbel-type distributions for the C—H (*a*), N—H (*b*), C—O (*c*) and C—N (*d*) bonds.

Table 9
Standard uncertainties of the bond critical point locations ($\sigma_x, \sigma_y, \sigma_z$), the electron density [$\sigma(\rho_{BCP})$] and the Laplacian at these points [$\sigma(\nabla^2\rho_{BCP})$].

The range of the values [$\Delta(\rho_{BCP})$, $\Delta(\nabla^2\rho_{BCP})$] of bond topological indices are also listed (units e and Å, σ_x, σ_y, σ_z 10^4 Å).

BCP	σ_x	σ_y	σ_z	$\sigma(\rho_{BCP})$	$\Delta(\rho_{BCP})$	$\sigma(\nabla^2\rho_{BCP})$	$\Delta(\nabla^2\rho_{BCP})$
C2—H1	6.02	6.79	6.21	0.0038	0.0266	0.0905	0.6343
C2—N3	9.65	9.01	8.94	0.0006	0.0040	0.0095	0.0709
C2—O6	7.20	7.87	7.87	0.0008	0.0058	0.0565	0.4367
N3—H4	6.97	5.95	7.59	0.0057	0.0400	0.3754	2.7113
N3—H5	5.93	6.95	7.41	0.0057	0.0386	0.3761	2.5147

numerical property as the $\mathbf{r}_{BCP}$ so well. While the change in the topology of the ED due to nuclear vibrations is tractable within the BO-approximation-based quantum chemistry, it is certainly not predictable. Our findings discussed above have direct applications to X-ray charge-density studies; indeed, we suggest supplementing experimental ED reports with the s.u. estimation of the BCP locations and BTPs. Sample-based statistics, as described in this report using theoretical ED, could be quite efficient for estimating 'experimental s.u.'s', given an analytic function of the ED (such as the pseudoatom model) at the 'experimental stationary nuclear geometry'.

Another significant observation is that the $\nabla^2\rho_{BCP}$ distribution is far from being 'univariate', which has in fact been implied by previous studies emphasizing only the effect of bond-length alternation on bond topology (Gibbs *et al.*, 1998, 2008). The computational protocol followed in this study allows one to track in detail how a particular internal normal mode affects the local topology of the ED. In line with earlier observations, we find a drastic change in the BCP properties upon changing the bond distance (Gibbs *et al.*, 2008). This alternation can be quite well described by a power law suggested earlier (Gibbs *et al.*, 1998). However, the $\nabla^2\rho_{BCP}$ distributions, especially for the C2—N3 bond, have emerged to be unexpectedly and profoundly sensitive to geometry changes not altering the corresponding bond length. In a related recent study, the topology of the dynamic ED was found to be incompatible with that of the static for bonds involving H atoms (no BCPs could be located for the N—H bonds for the dynamic ED) (Michael & Koritsanszky, 2014), indisputably suggesting that the static ED topology of these bonds is essentially determined by the mean nuclear position and s.u.'s (ADPs) of H atoms that are unreliably accessible from X-ray data. The results of this study also press the need for cautious modelling of H-atom sites, as the uncertainty of the H-atom parameters is shown to transform to over 1 e Å^{-5} uncertainty in the $\nabla^2\rho_{BCP}$ value for the nearby C2—N3 bond. This uncertainty alone is almost an order of magnitude higher than the experimental uncertainty typically estimated on the basis of linear error-propagation schemes using the standard pseudoatom model.

The BTP distributions are shown to depart from normal distributions. This observation makes sense, because the relationship between the BTP and the bond distance is non-linear and because the bond distance is not normally distributed. To interpret the non-Gaussian nature of the $\nabla^2\rho_{BCP}$ samples, we find it convenient to use the Gumbel statistics, since this PDF seems to be an appropriate parametric model with fixed complexity. Our choice of the Gumbel distribution is due entirely to its capability to represent the sample sufficiently (good figures of fit) and efficiently (with only two parameters to be fitted). As of writing this report, we have no straightforward explanation why extreme value statistics – associated with the Gumbel distribution – can adequately represent the BTP sample. We expect that convolution of PDFs of the Hessian eigenvalues could also well account for the data. Additional statistical models for the latent source of variations/co-variations are being targeted in our ongoing research.

The skewness and kurtosis are measures of asymmetry and outlier-describing behaviour of a distribution. Being location- and scale-free, these statistical figures reflect only the shape of the distribution. Based on their skewness, none of the ρ_{BCP} distributions is markedly different from a normal distribution; however, the $\nabla^2\rho_{BCP}$ distributions of the bonds including H atoms are consistently negatively skewed (tailed to the left), producing extreme values lower than the mode with a greater probability. Conversely, the $\nabla^2\rho_{BCP}$ distributions for the C2—O6 and C2—N3 bonds are positively skewed (tailed to the right), indicating greater tendency for yielding extreme values on the high side of the mode. Since all the PDFs possess positive kurtosis, they can be considered to be more receptive to outliers (or have 'fatter' tails) than the normal distribution. Yet, because the kurtoses are close to the value characteristic for normally distributed data (3), none of these PDFs should be marked as strikingly different from a normal distribution.

References

Bader, R. F. W. (1990). *Atoms in Molecules – a Quantum Theory*. New York: Oxford University Press.
Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.
Born, M. & Oppenheimer, R. (1927). *Ann. Phys.* **389**, 457–484.
Burnett, M. N. & Johnson, C. K. (1996). *ORTEPIII.* Report ORNL-6895. Oak Ridge National Laboratory, Tennessee, USA.
Coppens, P., Willoughby, T. V. & Csonka, L. N. (1971). *Acta Cryst.* A**27**, 248–256.
Debye, P. (1913). *Ann. Phys.* **348**, 49–92.
Frisch, M. J. *et al.* (2009). *Gaussian09*, Revision A.1. Wallingford: Gaussian, Inc.
Gatti, C. (2005). *Z. Kristallogr.* **220**, 339–487.
Gatti, C., Bianchi, R., Destro, R. & Merati, F. (1992). *J. Mol. Struct. THEOCHEM*, **255**, 409–433.
Gibbs, G. V., Downs, R. T., Cox, D. F., Rosso, K. M., Ross, N. L., Kirfel, A., Lippmann, T., Morgenroth, W. & Crawford, T. D. (2008). *J. Phys. Chem. A*, **112**, 8811–8823.
Gibbs, G. V., Hill, F. C., Boisen, M. B. & Downs, R. T. (1998). *Phys. Chem. Miner.* **25**, 585–590.
Gumbel, E. J. (1935). *Ann. Inst. Henri Poincaré*, **5**, 115–158.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Higgs, P. W. (1955). *Acta Cryst.* **8**, 99–104.
Hohenberg, P. & Kohn, W. (1964). *Phys. Rev.* **136**, B864–B871.
Koritsanszky, T. S. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1628.
Michael, J. R. & Koritsanszky, T. (2014). *J. Math. Chem.* doi: s10910-014-0425-y.
Scheringer, C. (1977*a*). *Acta Cryst.* A**33**, 426–429.
Scheringer, C. (1977*b*). *Acta Cryst.* A**33**, 430–433.

Volkov, A., Koritsanszky, T., Chodkiewicz, M. & King, H. F. (2009). *J. Comput. Chem.* **30**, 137–1391.
Volkov, A., Macchi, P., Farrugia, L., Gatti, C., Mallinson, P., Richter, T. & Koritsanszky, T. (2006). *XD2006. A computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structure factors*.
Waller, I. (1923). *Z. Phys.* **17**, 398–408.